Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:
http://oatao.univ-toulouse.fr/25033

Official URL

DOI : https://doi.org/10.1109/ACCESS.2019.2936541

To cite this version: Sodjo, Jessica and Giremus, Audrey and Dobigeon, Nicolas and Caron, François *A Bayesian nonparametric model for unsupervised joint segmentation of a collection of images.* (2019) IEEE Access, 7. 12017-12018. ISSN 2169-3536

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Digital Object Identifier 10.1109/ACCESS.2019.2936541

A Bayesian Nonparametric Model for Unsupervised Joint Segmentation of a Collection of Images

JESSICA SODJO[1], AUDREY GIREMUS[2], NICOLAS DOBIGEON[3], (Senior Member, IEEE), AND FRANÇOIS CARON[4]

[1]Department of Science and Technology, University of French Guiana, 97337 Cayenne, France
[2]Department of Sciences and Technology, University of Bordeaux—CNRS—BINP, 33400 Talence, France
[3]IRIT/INP-ENSEEIHT, University of Toulouse, 31071 Toulouse, France
[4]Department of Statistics, University of Oxford, Oxford OX1 3LB, U.K.

Corresponding author: Jessica Sodjo (jessica.bechet@univ-guyane.fr)

The research was funded by ANR-BNPSI 2014-2018.

ABSTRACT Jointly segmenting a collection of images with shared classes is expected to yield better results than single-image based methods, due to the use of the shared statistical information across different images. This paper proposes a Bayesian approach for tackling this problem. As a first contribution, the proposed method relies on a new prior distribution for the class labels, which combines a hierarchical Dirichlet process (HDP) with a Potts model. The latter classically favors a spatial dependency, whereas the HDP is a Bayesian nonparametric model that allows the number of classes to be inferred automatically. The HDP also explicitly induces a sharing of classes between the images. The resulting posterior distribution of the labels is not analytically tractable and can be explored using a standard Gibbs sampler. However, such a sampling strategy is known to have poor mixing properties for high-dimensional data. To alleviate this issue, the second contribution reported in this paper consists of an adapted generalized Swendsen-Wang algorithm which is a sampling technique that improves the exploration of the posterior distribution. Finally, since the inferred segmentation depends on the values of the hyperparameters, the third contribution aims at adjusting them while sampling the posterior label distribution by resorting to an original combination of two sequential Monte Carlo samplers. The proposed methods are validated on both simulated and natural images from databases.

INDEX TERMS Bayesian models, Bayesian nonparametrics, hierarchical Dirichlet process, image segmentation, potts model.

I. INTRODUCTION

Image segmentation has received a lot of attention in the literature since it is a key step in image analysis. Biomedical imaging based diagnosis [1] or remote sensing [2] are examples of archetypal applications. Segmentation aims at identifying K homogeneous areas, referred to as classes, in an image or a collection of images. Within a statistical framework, region-based segmentation can be formulated as the estimation of class labels univocally assigned to pixels sharing the same statistical properties. Adopting a Bayesian formulation, one has to elicit a prior distribution over the class labels. A classical choice is the Potts model, which promotes spatially piecewise homogeneous regions by yielding a higher prior probability for configurations where neighboring pixels are in the same class. However, a major limitation of the Potts model lies in the fact that the number of classes K must be set in advance. As an alternative, K can also be estimated jointly with the class labels. The resulting posterior distribution is highly intractable but can be explored using sampling methods, e.g., reversible jump Markov Chain Monte Carlo (RJ-MCMC) algorithms. However, the efficiency of RJ-MCMC may be limited by the dimension of the problem and is highly driven by the ability of designing

The associate editor coordinating the review of this article and approving it for publication was Diego Oliva.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

efficient moves across spaces of different dimensions. Conversely, Bayesian nonparametric methods offer a scalable solution by allowing K to increase with the dimension of the data. Following this idea, a prior combining the Dirichlet process (DP) and the Potts model has been proposed in [3] for automatic image segmentation. In this case, the DP ensures the automatic inference of the number of classes while the Potts model enforces spatial homogeneity.

When considering the segmentation of a collection of J images, the most straightforward approach consists in partitioning separately each of the images in a set of $m_j \cdot (j = 1, \ldots, J)$ regions and then matching the regions among the different images to identify common classes. However, this suboptimal two-step processing does not fully take the shared information into account to characterize the classes. As an alternative, a joint analysis can be carried out: such an approach deals jointly with all the images, fully exploiting the fact that a given class can be present in several of them, but possibly in different proportions. For instance, when analyzing remote sensing images, vegetation can be encountered in countryside but also in urban and suburban areas. Nevertheless, its density is far less significant in the last categories of images. Thus, the joint processing is expected to improve the detection and recognition of the vegetation class in images where it is quite rare. In such scenarios, the model in [3] may not be suitable because the DP cannot depict a sharing of classes between images.

A first nonparametric approach for shared segmentation of image databases is proposed in [4]. A hierarchical model is considered so that the different regions of an image appear with proportions given by Pitman-Yor (PY) processes; such processes are well-suited to capture the power-law behavior of natural objects. These regions are assigned to global classes whose prior distribution is also driven by a PY process. To favor smooth regions, thresholded correlated Gaussian processes are used to generate the class labels. Keeping up with the random field formalism, in this work, we propose an extension of the nonparametric model in [3]. The joint segmentation is made possible using a hierarchical Dirichlet process (HDP) [5] to model the labels of the regions and classes as well as their relationship, coupled with a Potts model that ensures spatial smoothness. The use of the HDP allows both the number of regions $m_{j\cdot}$, the overall number of classes K and the shared classes to be inferred automatically.

The posterior distribution of the labels can be written up to a normalization constant as the product of the prior and the likelihood of the set of observations. Since Bayesian estimators associated to this posterior distribution cannot be derived analytically, a Markov chain Monte Carlo (MCMC) approximation is considered. We presented preliminary results based on a Gibbs sampler in [6]. However, that sampling scheme was not suitable for the segmentation of real images due its high computational cost. Moreover, the sampling procedure in [6] did not allow the model hyperparameters to be estimated. This paper addresses both limitations.

First, adopting the strategy followed by most image segmentation approaches, the images are decomposed into super-pixels, here obtained using the simple linear iterative clustering algorithm (SLIC) [7]. Then, to fully characterize the super-pixels, both color and texture histograms are considered as descriptive features. As for the segmentation algorithm, a generalized Swendsen-Wang (SW) [8] based sampler is proposed to accelerate the convergence and overcome the mixing issues of the conventional Gibbs sampler. The SW algorithm proposed in [9] introduces a set of *bond* latent variables that allow the class labels of subsets of neighboring super-pixels to be jointly updated; as these class labels are typically highly correlated, this joint update improves the convergence rate of the sampler. The last contribution of this paper is to estimate the hyperparameters jointly with the partition. For that purpose, they could be also assigned a prior distribution and sampled within the SW procedure. However, due to the intrication of the Potts and HDP models, the presence of unknown normalization constants precludes the closed-form computation of the conditional distribution of the hyperparameters. To overcome this issue, we propose to embed the MCMC algorithm into a sequential Monte Carlo (SMC) sampler [10] which sequentially explores the posterior label distribution for a sequence of hyperparameter values. At each step, the likelihood of the current set of hyperparameters is also obtained up to an unknown constant, such that the optimal value can be selected afterwards. An original approach is proposed to compensate for this hyperparameter-dependent constant which requires to run an additional SMC algorithm that targets the prior instead of the posterior label distribution. Finally, the best partition must be selected based on the posterior label samples. To avoid label switching issues, it is chosen in this work as the one minimizing the Dahl's criterion [11].

The paper is organized as follows. The statistical model of the images to be segmented and the proposed prior distribution over the labels are presented in Section II. Section III describes the single-site Gibbs sampling procedure for posterior inference and the more efficient generalized SW-based sampler is introduced in section IV. In section V, a sequential Monte Carlo sampler is derived to jointly estimate the model hyperparameters and the image partition. Finally, after deriving the selection procedure of the best partition, experimental results are reported in section VII. Section VIII concludes the paper and proposes perspectives to this work.

II. PROPOSED IMAGE MODEL

A. LIKELIHOOD

Let us consider a set of J images to be jointly segmented. To reduce the data dimension and hence save computational complexity at the processing stage, each image j ($j \in \{1, \ldots, J\}$) is first divided into N_j super-pixels. The presegmentation algorithm used in this paper is the simple linear iterative clustering (SLIC) presented in [7]. It should be noted that the super-pixels are irregularly shaped but a

neighborhood system can still be defined. Indeed, two super-pixels are said to be neighbors if they share connected pixels.

Super-pixel n ($n \in \{1, \ldots, N_j\}$) in image j is characterized by the observation vector y_{jn} and its associated distribution denoted f is assumed to be parameterized by a latent vector θ_{jn}, i.e., $y_{jn}|\theta_{jn} \sim f(y_{jn}|\theta_{jn})$. To make the segmentation more efficient, both colors and textures of the super-pixels can be taken into account. More precisely, in this case, the observation vector is defined as a set of two descriptors, namely y_{jn}^{t} and y_{jn}^{c} representing the texture information and the red-green-blue (RGB) color histogram, respectively. As a consequence, the vector of observations writes $y_{jn} = (y_{jn}^{\mathrm{t}}, y_{jn}^{\mathrm{c}})$, inducing $\theta_{jn} = (\theta_{jn}^{\mathrm{t}}, \theta_{jn}^{\mathrm{c}})$. Assuming conditional independence, the joint likelihood associated with super-pixel n in image j writes $f(y_{jn}|\theta_{jn}) = f(y_{jn}^{\mathrm{t}}|\theta_{jn}^{\mathrm{t}})f(y_{jn}^{\mathrm{c}}|\theta_{jn}^{\mathrm{c}})$. This composite model will be considered in the experiments conducted on real natural images (see Section VII-B). It should be noted that preliminary tests on synthetic images of small size will be based on a simpler model. More precisely, no pre-segmentation will then be required and each class will be characterized only by the gray level of its pixels.

The J images are assumed to be composed of K, possibly shared, homogeneous classes. The joint segmentation has two levels: *i*) first, each image j is divided into $m_{j\cdot}$ ($j = 1, \ldots, J$) regions and *ii*) each region is assigned to a class. Thus, the classification task consists in recovering the region label $c_{jn} \in \{1, \ldots, m_{j\cdot}\}$ of super-pixel n in image j as well as the class label $d_{jt} \in \{1, \ldots, K\}$ of region t in image j.

Pixels assigned to the same class k share the same latent parameter vector, denoted ϕ_k. By denoting A_k the set of super-pixels assigned to class k, i.e., $A_k \triangleq \{(j, n)|d_{jc_{jn}} = k\}$, we have $\theta_{jn} = \phi_k$. Assuming the observations are conditionally independent given the parameter vector, their marginal likelihood is $f(\boldsymbol{y}) = \prod_{k=1}^{K} f(\boldsymbol{y}_{A_k})$ where

$$f(\boldsymbol{y}_{A_k}) = \int \left[\prod_{(j,n)\in A_k} f(y_{jn}|\phi_k) \right] h(\phi_k) \mathrm{d}\phi_k \tag{1}$$

with $\boldsymbol{y}$ the collection of all observations, $\boldsymbol{y}_{A_k} = \{y_{jn}|(j, n) \in A_k\}$ the observations associated to the pixels with indexes in A_k and $h(\cdot)$ the prior distribution on ϕ_k.

This model is conditional upon the class and region assignments. In the following section, we describe the considered prior model for the partition, i.e for the labels c_{jn} and d_{jt}.

B. PRIOR DISTRIBUTION

The proposed prior modeling for the set of class and region labels $(\boldsymbol{c}, \boldsymbol{d})$ combines two potential functions that promote sharing of the classes and spatial homogeneity, respectively. They are described in what follows.

1) HIERARCHICAL DIRICHLET PROCESS

Let $\mathbb{G}_j$ be the distribution of the hyperparameter vector θ_{jn} ($j \in \{1, \ldots, J\}$ and $n \in \{1, \ldots, N_j\}$). For the sake of generality, this distribution is assumed unknown and not enforced to belong to a specific parametric family. Within a Bayesian framework, it is assigned a prior distribution which has the specificity to be defined on the infinite dimensional space of the probability measures. In this work, a DP is chosen as a prior since its discrete realizations favor several pixels to share the same class labels [12]. Under this assumption, $\mathbb{G}_j$ writes:

$$\mathbb{G}_j \sim \mathrm{DP}(\alpha_0, \mathbb{G}_0) \quad \text{with } \mathbb{G}_j = \sum_{t=1}^{\infty} \tau_{jt} \delta_{\psi_{jt}}. \tag{2}$$

The DP depends on two parameters, a base measure denoted here $\mathbb{G}_0$ and a scale parameter α_0. The support points ψ_{jt} are independent and identically distributed according to $\mathbb{G}_0$ while the probabilities τ_{jt} indirectly depend on α_0 through the stick-breaking generative process [13].

To induce class sharing among the images, the base distribution $\mathbb{G}_0$ should be the same for all the images but also discrete to ensure that the distributions $\mathbb{G}_j$ ($j \in \{1, \ldots, J\}$) are defined with common atoms. As the number of atoms is unknown, a DP is also selected as a prior distribution for $\mathbb{G}_0$. In the sequel, its base distribution is denoted H, with probability density function h and scale factor γ, leading to

$$\mathbb{G}_0 \sim \mathrm{DP}(\gamma, H) \quad \text{with } \mathbb{G}_0 = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}. \tag{3}$$

In the literature, this model composed of nested layers of DPs is referred to as the hierarchical DP (HDP) [5]. To better understand it, we can resort to the well-known metaphor of the Chinese restaurant franchise (CRF) [5], based on the Pólya urn scheme. First, since the value of the hyperparameter vector θ_{jn} is uniquely determined by the region c_{jn} and class $d_{jc_{jn}}$ labels, the models (2) and (3) can be completely rewritten in terms of the latter. Furthermore, the distributions $\mathbb{G}_0$ and $\mathbb{G}_j$ can be marginalized out to focus on the sequential generation of the labels. Under the DP assumption, they are exchangeable so that each label can be sampled as if it were the last one. The procedure is the following: given the remainder of the super-pixels, super-pixel n in image j has a positive probability of being assigned to an existing region t which is proportional to the number ν_{jt} of super-pixels in that region whereas the probability of being assigned to a new one t^{new} is proportional to α_0, i.e.,

$$\Pr(c_{jn} = t|\boldsymbol{c}_j^{-n}) \propto \begin{cases} \nu_{jt} & \text{if } t \leq m_{j\cdot} \\ \alpha_0 & \text{if } t = t^{\mathrm{new}} \end{cases}$$

with $\boldsymbol{c}_j^{-n} = \{c_{jn'}|n' = 1, \ldots, N_j, n' \neq n\}$.

Similarly, region t in image j can be assigned to an existing class k proportionally to the overall number $m_{\cdot k}$ of regions assigned to class k, or to a new one proportionally to γ, i.e.,

$$\Pr(d_{jt} = k|\boldsymbol{d}^{-jt}) \propto \begin{cases} m_{\cdot k} & \text{if } k \leq K \\ \gamma & \text{if } k = k^{\mathrm{new}} \end{cases} \tag{4}$$

where $\boldsymbol{d}^{-jt} = \{d_{j't'}|j' = 1, \ldots, J; t' = 1, \ldots, m_{j'\cdot}; (j', t') \neq (j, t)\}$. Thus, the prior potential on the partition induced by the

HDP writes

$$\varphi(\boldsymbol{c},\boldsymbol{d}) = \prod_{j=1}^{J}\left\{\left[\prod_{n=1}^{N_j}\frac{1}{(\alpha_0+n-1)}\right]\alpha_0^{m_{j\cdot}}\left[\prod_{t=1}^{m_{j\cdot}}\Gamma(v_{jt})\right]\right\} \times\left[\prod_{t=1}^{m_{\cdot\cdot}}\frac{1}{(\gamma+t-1)}\right]\gamma^K\left[\prod_{k=1}^{K}\Gamma(m_{\cdot k})\right]. \quad (5)$$

2) POTTS MODEL

The prior (5) is complemented by a Potts-Markov random field which promotes spatial homogeneity of the class labels [14]. It is known to be particularly suitable for image segmentation [15] and thus has been widely used for various applications, ranging from remote sensing [2] to medical imaging [16]. Given a description of the image through a neighboring system, the probability for a super-pixel to be assigned to a class k depends on the number of its neighbors assigned to the same class. In our context, two super-pixels are considered to be neighbors if they have at least two pixels that are themselves neighbors according to a conventional 8-pixel neighborhood structure. The prior potential can thus be written

$$\rho(\boldsymbol{c},\boldsymbol{d}) \propto \prod_{j=1}^{J}\exp\left(\sum_{q\in\mathcal{V}_n}\beta\,\mathbf{1}_{d_{jc_{jq}},d_{jc_{jn}}}\right) \quad (6)$$

where $\mathcal{V}_n$ denotes the set of super-pixels neighbors of super-pixel n, $\mathbf{1}_{\cdot,\cdot}$ is the Kronecker delta function and β is the so-called granularity parameter. For high values of this parameter, the image exhibits few classes. On the contrary, for small values of β, the images tend to be over-segmented. The parameter β is assumed to be known in this paper. It could be estimated using a procedure similar to [17].

3) PROPOSED PRIOR MODEL

The proposed composite prior combines a global penalization of the number of classes provided by the HDP and a local penalization coming from the Potts model. The HDP induces an implicit prior on the number of regions in each image, the number of classes and the shared classes. With the Potts model, a spatial proximity is taken into account. The joint prior model of the region and class label is given by

$$\Pr(\boldsymbol{c},\boldsymbol{d}) \propto \varphi(\boldsymbol{c},\boldsymbol{d})\rho(\boldsymbol{c},\boldsymbol{d}). \quad (7)$$

III. GIBBS SAMPLING

The posterior distribution $\Pr(\boldsymbol{c},\boldsymbol{d}|\boldsymbol{y}) \propto \Pr(\boldsymbol{c},\boldsymbol{d})f(\boldsymbol{y}|\boldsymbol{c},\boldsymbol{d})$ is not analytically tractable, i.e., the optimal partition cannot be derived analytically. Thus, a Gibbs sampler is derived to draw samples asymptotically distributed according to the posterior. The region and class labels are iteratively sampled according to the posterior probabilities $\Pr(c_{jn}|\boldsymbol{c}^{-jn},\boldsymbol{d},\boldsymbol{y})$ and $\Pr(d_{jt}|\boldsymbol{c},\boldsymbol{d}^{-jt},\boldsymbol{y})$, respectively [18], with $\boldsymbol{c}^{-jn} = \{c_{j'n'}|j' = 1,\ldots,J; n' = 1,\ldots,N_{j'}; (j',n') \neq (j,n)\}$. This sampling scheme is inspired from the one initially proposed in [5].

A. SAMPLING OF THE CLASS LABELS

Sampling the class label d_{jt} of the region t in image j is achieved by deriving the corresponding conditional posterior probability, which is proportional to the prior model (7) and to the distribution of the observations associated with the pixels of interest, conditionally on the other observations and current partition, i.e.,

$$\Pr(d_{jt}|\boldsymbol{c},\boldsymbol{d}^{-jt},\boldsymbol{y}) \propto p(\boldsymbol{y}_{jt}|\boldsymbol{c},d_{jt},\boldsymbol{d}^{-jt},\boldsymbol{y}^{-jt})\Pr(d_{jt}|\boldsymbol{c},\boldsymbol{d}^{-jt}) \quad (8)$$

with, using (7),

$$\Pr(d_{jt}|\boldsymbol{c},\boldsymbol{d}^{-jt}) = \varphi(d_{jt}|\boldsymbol{c},\boldsymbol{d}^{-jt})\rho(d_{jt}|\boldsymbol{c},\boldsymbol{d}^{-jt}). \quad (9)$$

The term $\varphi(d_{jt}|\boldsymbol{c},\boldsymbol{d}^{-jt})$ can be computed by resorting to the CRF metaphor. By assuming that K classes are active at the current iteration of the Gibbs sampler, two cases should be considered: the region t in image j can be assigned to an existing class, i.e., $d_{jt} = k$ ($k \in \{1,\ldots,K\}$) or to a new class, i.e., $d_{jt} = k^{\text{new}}$ ($k^{\text{new}} \notin \{1,\ldots,K\}$). These two distinct cases are detailed in what follows.

1st Case ($d_{jt} = k$ With $k \in \{1,\ldots,K\}$: The first term in the right-hand side of (8) exhibits the distribution of the observations associated to pixels in the considered region conditionally on the observations associated to pixels assigned to the class $d_{jt} = k$ omitting those in the considered region, i.e., $p(\boldsymbol{y}_{jt}|\boldsymbol{c},d_{jt} = k,\boldsymbol{d}^{-jt},\boldsymbol{y}^{-jt}) = f(\boldsymbol{y}_{jt}|\boldsymbol{y}_{A_k^{-jt}}) = f(\boldsymbol{y}_{jt},\boldsymbol{y}_{A_k^{-jt}})/f(\boldsymbol{y}_{A_k^{-jt}})$. Moreover, the CRF allows the first prior factor to be evaluated using (4) as $\varphi(d_{jt} = k|\boldsymbol{c},\boldsymbol{d}^{-jt}) = \varphi(d_{jt} = k|\boldsymbol{d}^{-jt}) \propto m_{\cdot k}^{-jt}$. The second factor $\rho(d_{jt}|\boldsymbol{c},\boldsymbol{d}^{-jt})$, corresponding to the Potts model, depends on the class assignment of the set of neighbors of the super-pixels in the region t that do not belong to this region, herein denoted $\mathcal{V}^{[t]}$.

2nd Case ($d_{jt} = k^{\text{new}}$ With $k^{\text{new}} \notin \{1,\ldots,K\}$): In this case, since a new class is sampled, the conditional distribution of the observations $\boldsymbol{y}_{jt}$ becomes the marginal distribution, i.e., $p(\boldsymbol{y}_{jt}|\boldsymbol{c},d_{jt} = k^{\text{new}},\boldsymbol{d}^{-jt},\boldsymbol{y}^{-jt}) = f(\boldsymbol{y}_{jt}) = \int[\prod_{n|c_{jn}=t}f(y_{jn}|\phi)]h(\phi)\mathrm{d}\phi$. Furthermore, $d_{jt} = k^{\text{new}}$ induces that none of the remaining super-pixels could have been assigned to that class, i.e., $\rho(d_{jt} = k^{\text{new}}|\boldsymbol{c},\boldsymbol{d}^{-jt}) \propto 1$, and, according to (4), the CRF gives $\varphi(d_{jt} = k^{\text{new}}|\boldsymbol{c},\boldsymbol{d}^{-jt}) \propto \gamma$.

To summarize the two cases detailed above, the conditional posterior probabilities of d_{jt} writes

$$\Pr(d_{jt} = k|\boldsymbol{c},\boldsymbol{d}^{-jt},\boldsymbol{y}) \propto \begin{cases} m_{\cdot k}^{-jt}\exp\left(\sum_{q\in\mathcal{V}^{[t]}}\beta\mathbf{1}_{d_{jc_{jq}},k}\right)f(\boldsymbol{y}_{jt}|\boldsymbol{y}_{A_k^{-jt}}) & \text{if } k \leq K \\ \gamma f(\boldsymbol{y}_{jt}) & \text{if } k = k^{\text{new}}. \end{cases} \quad (10)$$

B. SAMPLING OF THE REGION LABELS

The conditional distribution of the region label c_{jn} associated with super-pixel n in image j is proportional to the prior model of this region label and to the likelihood of the observation

vector y_{jn}

$$\Pr(c_{jn}|\boldsymbol{c}^{-jn},\boldsymbol{d},\boldsymbol{y}) \propto p(y_{jn}|c_{jn},\boldsymbol{c}^{-jn},\boldsymbol{d},\boldsymbol{y}^{-jn})\Pr(c_{jn}|\boldsymbol{c}^{-jn},\boldsymbol{d}) \tag{11}$$

with

$$\Pr(c_{jn} \mid \boldsymbol{c}^{-jn},\boldsymbol{d}) = \rho(d_{jc_{jn}}|c_{jn},\boldsymbol{c}^{-jn},\boldsymbol{d}^{-jc_{jn}})\varphi(c_{jn}|\boldsymbol{c}^{-jn},\boldsymbol{d}^{-jc_{jn}}) \tag{12}$$

In (12), as for the class labels, the term $\varphi(c_{jn}|\boldsymbol{c}^{-jn},\boldsymbol{d})$ resulting from the Bayesian nonparametric prior implies that super-pixel n in image j can be assigned either to an existing region or to a new one. These two cases are discussed below.

1st Case ($c_{jn} = t$ With $t \in \{1,\ldots,m_{j\cdot}\}$): The first term of the conditional distribution (11) is the distribution of the observation vector y_{jn} conditionally on the observations attached to pixels assigned to region t in image j omitting the nth one: $p(y_{jn} \mid c_{jn} = t,\boldsymbol{c}^{-jn},\boldsymbol{d},\boldsymbol{y}^{-jn}) = f(y_{jn} \mid y_{A_{d_{jt}}^{-jn}})$. Using the CRF paradigm (4), $\varphi(c_{jn} = t|\boldsymbol{c}^{-jn},\boldsymbol{d}) = \varphi(c_{jn} = t|\boldsymbol{c}^{-jn}) \propto \nu_{jt}^{-jn}$. Moreoever, the Potts term writes $\rho(d_{jc_{jn}}|c_{jn},\boldsymbol{c}^{-jn},\boldsymbol{d}^{-jc_{jn}}) = \exp\left(\sum_{q\in\mathcal{V}_n}\beta\mathbf{1}_{d_{jc_{jq}},d_{jc_{jn}}}\right)$.

2nd Case ($c_{jn} = t^{\text{new}}$ With $t^{\text{new}} \notin \{1,\ldots,m_{j\cdot}\}$): The CRF metaphor (4) leads to $\varphi(c_{jn} = t^{\text{new}} \mid \boldsymbol{c}^{-jn}) \propto \alpha_0$ and the likelihood function associated with the observation vector y_{jn} is obtained by summing over the different possibilities for the class label $d_{jt^{\text{new}}}$ to be assigned to the new region

$$p(y_{jn}|c_{jn} = t^{\text{new}},\boldsymbol{c}^{-jn},\boldsymbol{d},\boldsymbol{y}^{-jn}) \propto \left[\sum_k m_{\cdot k}\exp\left(\sum_{q\in\mathcal{V}_n}\beta\mathbf{1}_{d_{jc_{jq}},k}\right)+\gamma\right]^{-1} \tag{13}$$

$$\times\left[\sum_{k=1}^{K} m_{\cdot k}\exp\left(\sum_{q\in\mathcal{V}_n}\beta\mathbf{1}_{d_{jc_{jq}},k}\right) f\left(y_{jn}|\boldsymbol{y}_{A_k^{-jn}}\right)+\gamma f(y_{jn})\right] \tag{14}$$

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

Finally, sampling the region label c_{jn} can be conducted from the conditional posterior probabilities

$$\Pr(c_{jn} = t|\boldsymbol{c}^{-jn},\boldsymbol{d},\boldsymbol{y}) \propto \begin{cases} \nu_{jt}^{-jn}\exp\left(\sum_{q\in\mathcal{V}_n}\beta\mathbf{1}_{d_{jc_{jq}},d_{jt}}\right) f(y_{jn}|\boldsymbol{y}_{A_{d_{jt}}^{-jn}}) & \text{if } t \le m_{j\cdot} \\ \alpha_0 p(y_{jn}|c_{jn} = t^{\text{new}},\boldsymbol{c}^{-jn},\boldsymbol{d},\boldsymbol{y}^{-jn}) & \text{if } t = t^{\text{new}}. \end{cases} \tag{15}$$

Note that, once a new region label t^{new} has been chosen, the corresponding class label $d_{jt^{\text{new}}}$ should be sampled according to

$$\Pr(d_{jt^{\text{new}}} = k|\boldsymbol{c},\boldsymbol{d}^{-jt^{\text{new}}}) \propto \begin{cases} m_{\cdot k}\exp\left(\sum_{q\in\mathcal{V}_n}\beta\mathbf{1}_{d_{jc_{jq}},k}\right) f(y_{jn}|\boldsymbol{y}_{A_k^{-jn}}) & \text{if } k \le K \\ \gamma f(y_{jn}) & \text{if } k = k^{\text{new}}. \end{cases} \tag{16}$$

IV. A GENERALIZED SWENDSEN-WANG BASED ALGORITHM

In high-dimensional problems, the efficiency of the proposed Gibbs sampler can be impaired by the slow convergence of the generated Markov chain. In particular, to speed-up convergence, one opportunity is offered by the SW algorithm, specifically designed to sample according to a Potts model [9], one of the key ingredient in the proposed prior model (7). It considers an augmented model, introducing a set of latent binary variables $\boldsymbol{r}$ linking pairwise pixels to form so-called spin-clusters. This latent variable $\boldsymbol{r}$ is defined such that its sampling only depends on the current partition of the pixels, while preserving the marginal distribution of the labels. The major interest of this strategy results from the fact that the labels of linked pixels are jointly updated. We propose here a generalized counterpart of the conventional SW algorithm presented in [19].

A. CONDITIONAL DISTRIBUTION OF THE LINKING VARIABLE

Within the proposed hierarchical segmentation scheme, two sets of labels are introduced, namely, the labels of region $\boldsymbol{c}$ and the labels of class $\boldsymbol{d}$. One can think of sampling $\boldsymbol{r}$ conditionally to the class labels, as it is classically conducted when handling a Potts model. Unfortunately, in this case, a problem arises: under this configuration, two pixels in different regions can be linked, which makes the sampling scheme of $\boldsymbol{c}$ and $\boldsymbol{d}$ challenging since several regions can be assigned to the same class. As an alternative, we propose to sample the linking variable conditionally to both the region and class labels. Note however that when two pixels belong to the same region, they also share the same class. Thus, the linking variable can be equivalently sampled conditionally on the region labels $\boldsymbol{c}$ only. As a consequence, the linking variable $r_{j_{nq}}$ associated with two super-pixels n and q in the image j can be sampled according to

$$r_{j_{nq}}|c_{jn},c_{jq} \sim \text{Ber}\left(1-\exp\left(-\beta\lambda\mathbf{1}_{c_{jn},c_{jq}}\right)\right) \tag{17}$$

where $\text{Ber}(p)$ is the Bernoulli distribution with parameter p and λ governs the behavior of the generalized SW (GSW) algorithm: the greater its value, the bigger the spin-clusters.

B. EMBEDDING THE GENERALIZED SW INTO THE GIBBS SAMPLER

Finally, at iteration i of the algorithm, the sampling scheme targeting the posterior distribution of interest is

1) $\boldsymbol{r}^{(i)} \sim \Pr(\boldsymbol{r}|\boldsymbol{c}^{(i-1)},\boldsymbol{d}^{(i-1)},\boldsymbol{y})$
2) $\boldsymbol{c}^{(i)} \sim \Pr(\boldsymbol{c}|\boldsymbol{d}^{(i-1)},\boldsymbol{r}^{(i)},\boldsymbol{y})$
3) $\boldsymbol{d}^{(i)} \sim \Pr(\boldsymbol{d}|\boldsymbol{c}^{(i)},\boldsymbol{r}^{(i)},\boldsymbol{y})$

In step 1, since the distribution of $\boldsymbol{r}$ only depends on the partition in regions $\boldsymbol{c}$, the conditional probabilities reduce to $\Pr(\boldsymbol{r}|\boldsymbol{c},\boldsymbol{d},\boldsymbol{y}) = \Pr(\boldsymbol{r}|\boldsymbol{c})$. As a consequence, the linking variables are sampled according to (17). Moreover, in step 3, the class labels are conditionally independent of the linking variable $\boldsymbol{r}$. Thus this step is equivalent to the one described in paragraph III-A.

In all and for all, only step 2, i.e., the sampling of the region labels $\boldsymbol{c}$, needs to be reconsidered carefully. As previously said, the SW algorithm allows the region labels associated with observations belonging to the same spin-cluster to be simultaneously updated. Let $\mathcal{C}_{jl}$ be the set of super-pixels in the spin-cluster l of image j. We denote $\boldsymbol{y}_{jl}, \boldsymbol{r}_{jl}$ and $\boldsymbol{c}_{jl}$ the associated set of observations, linking labels and region labels, respectively, and $\boldsymbol{y}^{-jl}$, $\boldsymbol{r}^{-jl}$ and $\boldsymbol{c}^{-jl}$ the set of observations, linking labels and region labels when omitting the super-pixels in $\mathcal{C}_{jl}$. The conditional posterior probability writes

$$\begin{aligned}\Pr(\boldsymbol{c}_{jl} = t|\boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{r}, \boldsymbol{y}) \\ \propto \Pr(\boldsymbol{c}_{jl} = t|\boldsymbol{c}^{-jl}, \boldsymbol{d}) \Pr(\boldsymbol{r}_{jl}|\boldsymbol{r}^{-jl}, \boldsymbol{c}_{jl} = t, \boldsymbol{c}^{-jl}) \\ \times p(\boldsymbol{y}_{jl}|\boldsymbol{c}_{jl} = t, \boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{y}^{-jl}) \end{aligned} \tag{18}$$

The first term in (18) can be rewritten as: $\Pr(\boldsymbol{c}_{jl} = t|\boldsymbol{c}^{-jl}, \boldsymbol{d}) \propto \varphi(\boldsymbol{c}_{jl} = t|\boldsymbol{c}^{-jl}, \boldsymbol{d})\rho(\boldsymbol{c}_{jl} = t|\boldsymbol{c}^{-jl}, \boldsymbol{d})$. The Potts field involves the number of neighboring super-pixels of the spin-cluster that are in region t. The set $\mathcal{V}_{jl}$ of that neighboring super-pixels is the set of super-pixels that are neighbors of super-pixels in $\mathcal{C}_{jl}$ but are not in $\mathcal{C}_{jl}$. The conditional probability for the super-pixels in $\mathcal{C}_{jl}$ to be in an existing region then writes

$$\begin{aligned}\Pr(\boldsymbol{c}_{jl} = t \leq m_{j.}^{-jl}|\boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{r}, \boldsymbol{y}) \\ \propto \frac{\Gamma(\nu_{jt}^{-jl} + |\mathcal{C}_{jl}|)}{\Gamma(\nu_{jt}^{-jl})} f(\boldsymbol{y}_{jl}|\boldsymbol{y}_{A_{d_{jt}}^{-jl}}) \\ \times \exp\left(\sum_{q \in \mathcal{V}_{\mathcal{C}_{jl}}} \beta\left[\mathbf{1}_{d_{jc_{jq}}, d_{jt}} - \lambda \mathbf{1}_{c_{jq}, t}\right]\right) \end{aligned} \tag{19}$$

Moreover, the probability that the super-pixels in $\mathcal{C}_{jl}$ are in a new region is

$$\begin{aligned}\Pr(\boldsymbol{c}_{jl} = t^{\text{new}}|\boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{r}, \boldsymbol{y}) \\ \propto \alpha_0 \ \Gamma(|\mathcal{C}_{jl}|) \ p(\boldsymbol{y}_{jl}|\boldsymbol{c}_{jl} = t^{\text{new}}, \boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{y}^{-jl}) \end{aligned} \tag{20}$$

where

$$\begin{aligned}p(\boldsymbol{y}_{jl}|\boldsymbol{c}_{jl} = t^{\text{new}}, \boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{y}^{-jl}) \\ \propto \left[\sum_{k=1}^{K} m_{\cdot k} \exp\left(\sum_{q \in \mathcal{V}_{\mathcal{C}_{jl}}} \beta \mathbf{1}_{d_{jc_{jq}}, k}\right) + \gamma\right]^{-1} \\ \times \left[\sum_{k=1}^{K} m_{\cdot k} \exp\left(\sum_{q \in \mathcal{V}_{\mathcal{C}_{jl}}} \beta \mathbf{1}_{d_{jc_{jq}}, k}\right) f\left(\boldsymbol{y}_{jl}|\boldsymbol{y}_{A_k^{-jl}}\right) + \gamma f(\boldsymbol{y}_{jl})\right] \end{aligned}$$

with $f(\boldsymbol{y}_{jl}) = \int [\prod_{n \in \mathcal{C}_{jl}} f(y_{jn}|\phi_{k^{\text{new}}})] h(\phi_{k^{\text{new}}}) \mathrm{d}\phi_{k^{\text{new}}}$.

V. ESTIMATION OF THE HYPERPARAMETERS

The quality of the segmentation highly depends on the fine adjustment of the hyperparameter values. As for the parameter of the Potts model β, its estimation has already been investigated in the literature and can be performed for instance using an approximate Bayesian computation as proposed in [17]. Thus, this paper rather focuses on the parameters of the HDP prior which are denoted in the sequel $\chi = \{\alpha_0, \gamma\}$.

Adopting a fully Bayesian framework, they could be assigned a prior distribution and jointly estimated with the parameters of interest, namely the class partition [5]. Within the Gibbs sampler presented in Section III, sampling according to the conditional posterior distribution $\mathrm{p}(\chi|\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{y})$, e.g., using a Metropolis-Hastings step, would require to compute the normalization constant associated to the prior probabilities $\Pr(\boldsymbol{c}, \boldsymbol{d}|\chi)$ defined by (7). However, mainly due to the Potts potential included in the prior distribution (7), this normalization is not analytically tractable. Moreover, as it involves a sum over all possible partitions of (7), its pre-computation for realistic segmentation problems is not possible.

The proposed alternative consists in estimating χ under a maximum likelihood paradigm while exploring the posterior distribution of the class partition by resorting to a sequential Monte Carlo (SMC) sampler [10]. This algorithm, which combines MCMC and importance sampling, is intended to sequentially sample from a sequence of probability distributions defined on the same measurable space. In our case, this sequence corresponds to the posterior distributions of the class and region labels conditionally upon a pre-defined grid of values for χ. The interest of the SMC is then twofold. Firstly, it only requires the target distributions to be known up to a normalizing constant. Secondly, it calculates at each iteration an estimate of this normalizing constant. Since the latter is proportional to the likelihood of the hyperparameters, it makes it possible to select afterwards their most likely value.

To further detail the proposed approach, let us denote $\{\chi_m\}_{m=1,\ldots,M}$ a predefined grid of possible hyperparameter values, Ψ_m the target distribution up to a normalizing constant and Z'_m this normalizing constant at the m^{th} iteration of the SMC. More precisely, the latter are defined as follows:

$$\Pr(\boldsymbol{c}, \boldsymbol{d}|\boldsymbol{y}, \chi_m) = \frac{\Psi_m(\boldsymbol{c}, \boldsymbol{d})}{Z'_m}, \tag{21}$$

with $\Psi_m(\boldsymbol{c}, \boldsymbol{d}) = \varphi_m(\boldsymbol{c}, \boldsymbol{d}) \, \rho_m(\boldsymbol{c}, \boldsymbol{d}) f(\boldsymbol{y}|\boldsymbol{c}, \boldsymbol{d})$, where φ_m and ρ_m write as in equations (5) and (6), respectively, with the hyperparameters set to χ_m. The proposed SMC proceeds by running in parallel several MCMC samplers, while varying at each iteration the value of the hyperparameters according to the grid. The generated chains are inhomogeneous but an importance sampling step is conducted to guarantee that a Monte Carlo approximation of $\Pr(\boldsymbol{c}, \boldsymbol{d}|\boldsymbol{y}, \chi_m)$ is obtained. More precisely, if $\boldsymbol{c}_m^{(i)}$ and $\boldsymbol{d}_m^{(i)}$ are the samples generated by the ith ($i \in \{1, \ldots, I\}$) chain at the mth iteration, this probability writes

$$\Pr(\boldsymbol{c}, \boldsymbol{d}|\boldsymbol{y}, \chi_m) \simeq \sum_{i=1}^{I} W_m^{(i)} \delta_{\boldsymbol{c}_m^{(i)}}(\boldsymbol{c}) \delta_{\boldsymbol{d}_m^{(i)}}(\boldsymbol{d}), \tag{22}$$

where $\left\{\boldsymbol{c}_m^{(i)}, \boldsymbol{d}_m^{(i)}\right\}_{i=1,\ldots,I}$ and $\left\{W_m^{(i)}\right\}_{i=1,\ldots,I}$ are called the particles and the importance weights, respectively. The latter are introduced to correct for the discrepancy between the actual and the target distributions of the particles. It should be noted that different implementations of the SMC can be considered depending on the choice of the MCMC samplers: either the Gibbs or the generalized Swendsen Wang algorithms described in Sections III and IV, respectively, could fit. In the first case, the particles should be updated based on (8)–(16) whereas, in the second one, (18)–(20) should be used. The exact computation of the importance weights involves intricate high dimensional integrals and cannot be carried out in practice. It is shown in [10] that a relevant approximation of the optimal importance weights, in the sense that they have the minimum variance, can be sequentially computed as

$$W_m^{(i)} = W_{m-1}^{(i)} w_m^{(i)} / \sum_{\ell=1}^{I} W_{m-1}^{(\ell)} w_m^{(\ell)}, \tag{23}$$

where the incremental weights $w_m^{(i)}$ reduce to:

$$w_m^{(i)} = \frac{\Psi_m(\boldsymbol{c}_{m-1}^{(i)}, \boldsymbol{d}_{m-1}^{(i)})}{\Psi_{m-1}(\boldsymbol{c}_{m-1}^{(i)}, \boldsymbol{d}_{m-1}^{(i)})}. \tag{24}$$

It is well-known that sequential importance sampling suffers from weight impoverishment: that is, after a few iterations, all the weights except one are close to 0. To prevent this degeneracy, the particles are classically resampled according to (22) when the effective sample size (ESS) [20] is below a given threshold classically chosen as ϵI, with $0.5 < \epsilon < 1$. Then, their weights are reset to $1/I$.

A remarkable property of the SMC is that the sum of the unnormalized importance weights provides an estimate of the ratios of consecutive normalization constants, i.e.,

$$R'(m) = \frac{\widehat{Z'_m}}{Z'_{m-1}} = \sum_{\ell=1}^{I} W_{m-1}^{(\ell)} w_m^{(\ell)}. \tag{25}$$

Then, by computing the product of the sequences of estimates of the form (25) up to the current iteration, one can obtain an estimate of Z'_m/Z'_1 which is directly related to the likelihood of χ_m as follows

$$\frac{Z'_m}{Z'_1} = \prod_{p=2}^{m} R'(p) = \frac{f(\mathbf{y}|\chi_m) Z_m}{f(\mathbf{y}|\chi_1) Z_1} \tag{26}$$

where Z_m and Z_1 are the normalizing constants of the HDP-Potts prior distributions for the 1st and mth value of χ, respectively. Using notations introduced in Section II, we have:

$$\mathrm{P}(\boldsymbol{c}, \boldsymbol{d}|\chi_m) = \frac{\varphi_m(\boldsymbol{c}, \boldsymbol{d})\ \rho_m(\boldsymbol{c}, \boldsymbol{d})}{Z_m}. \tag{27}$$

The difficulty inherent to the considered setting is that these normalization constants also depend on the hyperparameters and cannot be elicited. To overcome this issue, we propose to run an additional SMC defined with the same grid of hyperparameter values but targeting the prior instead of the posterior distribution of the partitions. At each iteration, the particles are updated by sampling from equations (8)–(16) or (18)–(20), but removing all the terms depending on the measurements $\boldsymbol{y}$. As for the incremental weights, they are obtained as:

$$w_m^{(i)} = \frac{\varphi_m(\boldsymbol{c}_{m-1}^{(i)}, \boldsymbol{d}_{m-1}^{(i)}) \rho_m(\boldsymbol{c}_{m-1}^{(i)}, \boldsymbol{d}_{m-1}^{(i)})}{\varphi_{m-1}(\boldsymbol{c}_{m-1}^{(i)}, \boldsymbol{d}_{m-1}^{(i)}) \rho_{m-1}(\boldsymbol{c}_{m-1}^{(i)}, \boldsymbol{d}_{m-1}^{(i)})}. \tag{28}$$

The trick is that this SMC yields at each iteration an estimate of the ratio Z_m/Z_{m-1}:

$$R(m) = \frac{\widehat{Z_m}}{Z_{m-1}} = \sum_{\ell=1}^{I} W_{m-1}^{(\ell)} w_m^{(\ell)}. \tag{29}$$

These ratios can be combined to obtain an approximation of Z_m/Z_1 which can be used to compensate this term in (25) and directly estimate $f(\mathbf{y}|\chi_m)/f(\mathbf{y}|\chi_1)$ in the end. Thus, $f(\mathbf{y}|\chi)$ can finally be computed for each value of the hyperparameter χ chosen on the grid. The maximum likelihood estimate of the hyperparameters $\hat{\chi}$ is finally derived as:

$$\hat{\chi} = \arg\max_{\chi_m} \frac{f(\mathbf{y}|\chi_m)}{f(\mathbf{y}|\chi_1)}. \tag{30}$$

The different steps of both SMC are detailed in Algo. 1 as well as the combination of their outputs to estimate the likelihoods of the hyperparameter values. Last but not least, it should be noted that it is not necessary to resample a partition conditional on the optimal χ. Indeed, the posterior distribution of partition is directly given by (22) taking $\chi_m = \hat{\chi}$.

VI. DERIVING THE OPTIMAL PARTITION

After a burn-in period, the algorithms described in Sections III and IV generate I samples from the posterior distribution of interest which are denoted $(\boldsymbol{c}^{(1)}, \boldsymbol{d}^{(1)}), \ldots, (\boldsymbol{c}^{(I)}, \boldsymbol{d}^{(I)})$. Similarly, if the SMC is considered, I samples from the target distribution can be directly obtained by resampling the empirical distribution (22) corresponding to the most likely values of the hyperparameters. Then, within a Bayesian segmentation framework, the optimal partition is commonly chosen as the marginal maximum a posteriori estimator, i.e., the realization with maximum order of multiplicity. However, under the considered model, deriving this estimator is not straightforward not only because of the well-known label switching problem [21, Chap. 13] but also due to the varying number of classes induced by the hierarchical Dirichlet process. To overcome label switching, each sample would need to be re-labeled, resulting in a prohibitive computation cost. As an alternative, this paper proposes to follow the approach introduced in [11] and detailed in what follows. Let

$$\boldsymbol{\kappa} \triangleq \left\{d_{jc_{jn}},\ j \in \{1, \ldots, J\},\ n \in \{1, \ldots, N_j\}\right\} \tag{31}$$

denote the set of class labels assigned to each super-pixel in the images with $\kappa_{jn} \triangleq d_{jc_{jn}}$. The best partition $\hat{\boldsymbol{\kappa}}$ is then

Algorithm 1 Sequential Monte Carlo Samplers for the Hyperparameter Estimation

Input: $\boldsymbol{y}$, $\{\chi_m\}_{m=1,\cdots,M}$

% SMC targeting at the posterior distribution
for $m = 1$ **to** M **do**
  %Sampling conditionally to the hyperparameter χ_m
  for $i = 1$ **to** I **do**
    %Sampling along the I particles
    Sample $\boldsymbol{c}_m^{(i)}, \boldsymbol{d}_m^{(i)}$ according to (8)–(16) or (18)–(20);
    Compute $w_m^{(i)}$ according to (24);
  end
  Compute $W_m^{(i)}$ according to (23);
  Compute the ratio $R'(m)$ according to (25);
  if ESS $< \varepsilon I$ **then**
    Resample the particles:
    $\forall i \in [\![1, I]\!]$, set $W_m^{(i)} = 1/I$;
  end
end
% SMC targeting at the prior distribution
for $m = 1$ **to** M **do**
  %Sampling conditionally to the hyperparameter χ_m
  for $i = 1$ **to** I **do**
    %Sampling along the I particles
    Sample $\boldsymbol{c}_m^{(i)}, \boldsymbol{d}_m^{(i)}$ from their prior distribution;
    Compute $w_m^{(i)}$ according to (28);
  end
  Compute $W_m^{(i)}$ according to (23);
  Compute the ratio $R(m)$ according to (29);
  if ESS $< \epsilon I$ **then**
    Resample the particles:
    $\forall i \in [\![1, I]\!]$, set $W_m^{(i)} = 1/I$;
  end
end
% Compute the marginal likelihood
$$\frac{p(\boldsymbol{y}|\chi_m)}{p(\boldsymbol{y}|\chi_1)} \simeq \prod_{\ell=2}^{m} \frac{R'(\ell)}{R(\ell)};$$
Output: $\hat{\chi}$

defined as the one minimizing the Dahl's criterion corresponding to the loss function associated to the Bayesian risk

$$\hat{\boldsymbol{\kappa}} = \underset{i\in\{1,\ldots,I\}}{\operatorname{argmin}} \sum_{j=1}^{J} \sum_{n,q=1}^{N_j} \left(\mathbf{1}_{\kappa_{jn}^{(i)}, \kappa_{jq}^{(i)}} - \zeta_{jnq} \right)^2 \tag{32}$$

where $\boldsymbol{\kappa}^{(i)}$ is the partition generated at iteration i and

$$\zeta_{jnq} = \frac{1}{I} \sum_{i=1}^{I} \mathbf{1}_{\kappa_{jn}^{(i)}, \kappa_{jq}^{(i)}}$$

can be interpreted as an average pairwise probability. Thus, the optimal partition is chosen as the closest to the average one in the least-squares sense. An alternative would have consisted in thresholding $\boldsymbol{\zeta}$ to define the optimal partition $\hat{\boldsymbol{\kappa}}$. However this choice would have promoted an over-estimated number of classes, in particular for pixels located in edges of classes. Moreover, choosing the optimal partition among the sampled ones amounts to select a fairly probable partition.

VII. EXPERIMENTS

The proposed model has been first applied to toy-images so as to have at our disposal a ground truth to validate the obtained segmentation. Then, it has been tested on natural images from the LabelMe[1] database.

A. RESULTS ON TOY IMAGES

A set of $J = 3$ images is generated as follows: synthetic classification maps of size 64×64 pixels are simulated using a Potts model with parameter $\beta = 1.2$ and a number of classes chosen as $K_{\text{real}} = 3$. To each class is associated a gray level: -25, 0 or 25. These piecewise constant images are then degraded by a centered white Gaussian noise of variance $\sigma_y^2 = 15^2$ to form the observed images. Noise-free and noisy images are depicted in figure 1.

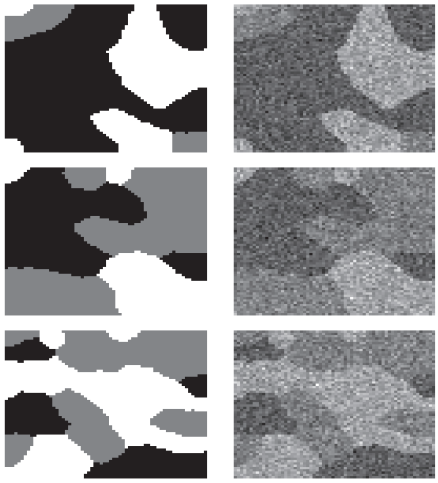

FIGURE 1. (a) Noise-free toy-images and (b) Noisy toy-images ($\sigma_Y = 15$).

1) MARGINAL LIKELIHOOD

To segment the image set, a particular instance of the generic model introduced in Section II is considered. Since the pre-segmentation in super-pixels is not required for small size images, the observations that are taken into account for the segmentation directly consist of the noisy gray levels of the pixels. The base distribution H of the HDP, according to which the class parameters ϕ_k are assumed to be drawn, is the following Gaussian distribution:

$$H \equiv \mathcal{N}(\mu_0, \sigma_0^2),$$

with $\mu_0 = 0$ and $\sigma_0^2 = 75^2$. Moreover, conditionally upon these parameters ϕ_k, the distribution $f(\cdot|\phi_k)$ of the pixel

[1] labelme.csail.mit.edu/

values is assumed to be Gaussian, i.e.,

$$y_{jn}|\phi_k \sim \mathcal{N}(y_{jn}; \phi_k, \sigma_y^2).$$

Note that the distributions H and f are conjugate, which allows the marginal likelihood to be analytically computed. Indeed, the observation model can be rewritten as

$$y_{jn} = \phi_k + \varepsilon_{jn}, \quad \varepsilon_{jn} \sim \mathcal{N}(0, \sigma_y^2). \tag{33}$$

The marginal distribution of the set of observations associated to pixels in a class k is then a Gaussian distribution $\mathcal{N}(\boldsymbol{y}_{A_k}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ with mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$ defined as

$$\boldsymbol{\mu}_k = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \mu_0$$

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} \sigma_0^2 + \sigma_y^2 & \sigma_0^2 & \cdots & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 + \sigma_y^2 & \cdots & \sigma_0^2 \\ & & \ddots & \\ \sigma_0^2 & \sigma_0^2 & \cdots & \sigma_0^2 + \sigma_y^2 \end{bmatrix}.$$

The conditional distribution of the nth observation associated to the kth class in the jth image is also Gaussian, i.e.,

$$y_{jn}|\boldsymbol{y}_{A_k^{-jn}} \sim \mathcal{N}(y_{jn}; \mu_n, \sigma_y^2 + \sigma_n^2)$$

with

$$\sigma_n^2 = \text{var}\left[\phi_k|\boldsymbol{y}_{A_k^{-jn}}\right] = \left(\frac{1}{\sigma_0^2} + \frac{|A_k| - 1}{\sigma_y^2}\right)^{-1}$$

$$\mu_n = \mathbb{E}\left[\phi_k|\boldsymbol{y}_{A_k^{-jn}}\right] = \sigma_n^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum\limits_{(j',n') \in A_k^{-jn}} y_{j'n'}}{\sigma_y^2}\right)$$

Similarly, the conditional distribution of the observations associated to pixels in the tth region is

$$\boldsymbol{y}_{jt}|\boldsymbol{y}_{A_k^{-jt}} \sim \mathcal{N}(\boldsymbol{y}_{jt}; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$$

with

$$\boldsymbol{\mu}_t = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \mu_p$$

$$\boldsymbol{\Sigma}_t = \begin{bmatrix} \sigma_y^2 + \sigma_p^2 & \sigma_p^2 & \cdots & \sigma_p^2 \\ \sigma_p^2 & \sigma_y^2 + \sigma_p^2 & \cdots & \sigma_p^2 \\ & & \ddots & \\ \sigma_p^2 & \sigma_p^2 & \cdots & \sigma_y^2 + \sigma_p^2 \end{bmatrix}$$

where

$$\mu_p = \sigma_p^2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum\limits_{(j',n') \in A_k^{-jt}} y_{j'n'}}{\sigma_y^2}\right)$$

and

$$\sigma_p^2 = \left(\frac{1}{\sigma_0^2} + \frac{|A_k| - \nu_{jt}}{\sigma_y^2}\right)^{-1}.$$

The same equation holds for a set of observations belonging to the same spin cluster. These quantities are involved in the different steps of the Gibbs algorithm and the GSW algorithm introduced in Sections III and IV, respectively.

2) RESULTS

The proposed HDP-Potts model-based segmentation, implemented by using either the Gibbs or the GSW algorithms, is compared to the DP-Potts [3] for which the considered images have been concatenated into a unique one beforehand. We have enforced that pixels that originate from different images cannot be neighbors. The parameters are set as

- scale parameter of the DP in the DP-Potts model: $\alpha = 1$,
- Potts model granularity parameter in the DP- and HDP-Potts model: $\beta = 1.2$,
- scale parameters of the DPs in the HDP-Potts model: $\alpha_0 = 1, \gamma = 1$.

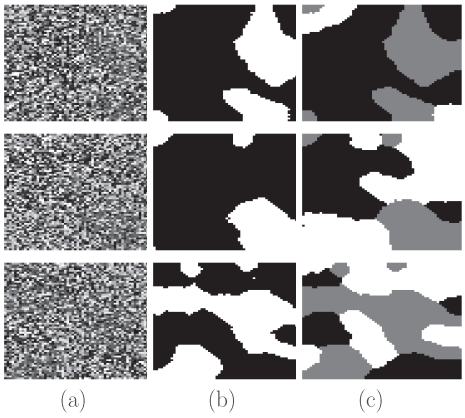

FIGURE 2. Initialization for the DP on (a) and results of segmentation for (b) the DP-Potts and (c) the DP-Potts (GSW, $\lambda = 0.5$) algorithms.

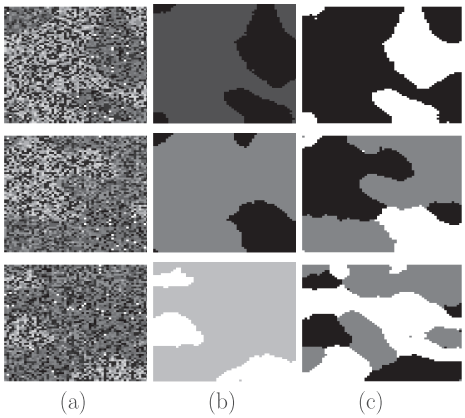

FIGURE 3. Initialization for the HDP on (a) and results of segmentation for (b) the HDP-Potts and (c) the HDP-Potts (GSW, $\lambda = 0.025$) algorithms.

Figures 2 and 3 respectively show the segmentation results obtained with the DP-Potts and HDP-Potts models.

Each sampler has been run for 10000 iterations but only the last 5000 sampled partitions were used to estimate the optimal one. With the considered value of σ_y, the images are quite noisy, making the segmentation problem harder.

To speed-up the convergence, a *k-means* algorithm has been applied on the noisy images for the initialization. For the DP-Potts (Gibbs and GSW) algorithms, wherein all the images are concatenated to form a unique one, the *k-means* is applied with $K_{\text{init}} = 30$. For the HDP-Potts (Gibbs and GSW) algorithms, the *k-means* algorithm is separately applied on each image of the set with $K_{\text{init}} = 15$.

As for the proposed results, one should keep in mind that the colors are arbitrary and are automatically set depending on the number of classes. The relevance of the segmentation should be studied by comparing the set of pixels that are in the same class initially and also in the estimated partition. It can be seen that the Gibbs algorithm detects only 2 classes. The interest of the SW technique then appears: for both models, applying it improves the results. However, by comparing the required time to converge, it takes far less iterations for the HDP-Potts than for the DP-Potts. Indeed, while more than a hundred iterations are necessary for the DP-Potts (GSW), less than ten are required for the HDP-Potts (GSW). It should be noted that it is difficult to define a convergence criterion that is independent of the label-switching problem: we chose to simply monitor the stabilization of the posterior distribution. However, this is not optimal since the chain can be trapped in a local minimum.

In order to quantitatively evaluate the obtained segmentations, we have also computed the V-measure introduced in [22]. Its value, comprised in the interval $[\![0, 1]\!]$, has the advantage of being independent of the label-switching issues. The closer it is to 1, the better the classification. The V-measure yields similar values for both models on the toy-images, ranging between 0.87 to 0.95 depending on the considered images. These high values confirm the validity of the proposed method.

B. RESULTS ON NATURAL IMAGES

1) MARGINAL LIKELIHOOD

Each image is pre-segmented in super-pixels with the SLIC [7] method. The descriptor that we choose for a super-pixel is a histogram. In this case, the likelihood of the observations is defined as a multinomial distribution. To be able to compute the marginal distribution of the observations, the prior distribution on the parameters is set as a Dirichlet distribution. Let ϕ_k be the vector of parameters, the model then writes:

$$\phi_k|H \sim H \equiv \text{Dir}(\varphi_0)$$
$$y_{jn}|\phi_k \sim f(\cdot|\phi_k) \equiv \text{Mult}(y_{jn}; \phi_k)$$

Thanks to the conjugacy of the Dirichlet and multinomial distributions, the marginal distribution of the observations associated to pixels in class k writes:

$$f(\boldsymbol{y}_{A_k}) = \left[\prod_{(j,n)\in A_k} \frac{1}{\mathcal{M}(y_{jn})}\right] \frac{\mathcal{D}\left(\varphi_0 + \sum_{(j,n)\in A_k} y_{jn}\right)}{\mathcal{D}(\varphi_0)}$$

with $\mathcal{D}$ and $\mathcal{M}$ respectively the normalizing constants of the Dirichlet and multinomial distributions [23]. As for the conditional distributions, they are expressed as:

$$f(y_{jn}|\boldsymbol{y}_{A_k^{-jn}}) = \frac{f(y_{jn}, \boldsymbol{y}_{A_k^{-jn}})}{f(\boldsymbol{y}_{A_k^{-jn}})}$$

$$f(y_{jn}|\boldsymbol{y}_{A_k^{-jn}}) = \frac{1}{\mathcal{M}(y_{jn})} \times \frac{\mathcal{D}\left(\varphi_0 + y_{jn} + \sum_{(j',n')\in A_k^{-jn}} y_{j'n'}\right)}{\mathcal{D}\left(\varphi_0 + \sum_{(j',n')\in A_k^{-jn}} y_{j'n'}\right)}$$

$$f(\boldsymbol{y}_{jt}|\boldsymbol{y}_{A_k^{-jt}}) = \frac{f(\boldsymbol{y}_{jt}, \boldsymbol{y}_{A_k^{-jt}})}{f(\boldsymbol{y}_{A_k^{-jt}})}$$

$$f(\boldsymbol{y}_{jt}|\boldsymbol{y}_{A_k^{-jt}}) = \frac{1}{\prod_{q|c_{jq}=t} \mathcal{M}(y_{jq})} \times \frac{\mathcal{D}\left(\varphi_0 + \sum_{q|c_{jq}=t} y_{jq} + \sum_{(j',n')\in A_k^{-jt}} y_{j'n'}\right)}{\mathcal{D}\left(\varphi_0 + \sum_{(j',n')\in A_k^{-jt}} y_{j'n'}\right)}$$

In the sequel, it should be noted that the type of considered histogram depends on the processed images: it can be gray level histograms or RGB histograms. They can be used in combination with a histogram of oriented gradients (HOG) [24] to address textured classes and changes of illumination.

2) RESULTS

The parameters have been set as:

- scale parameter of the DP in the DP-Potts model: $\alpha = 1$
- Potts model granularity parameter in the DP- and HDP-Potts model: $\beta = 1$
- number of bins in the histograms: $N_{\text{bi}} = 16$
- scale parameters for the DPs in the HDP-Potts model: $\alpha_0 = 1, \gamma = 1$
- parameter of the prior Dirichlet distribution on the observed gray level histograms: $\varphi_0 = 10^4 \times \bar{y}$, where $\bar{y}$ is the mean of the observed gray level histograms and for the RGB combined to HOG histograms: $\varphi_0^{\text{c}} = 3 \times 10^4 \times \bar{y}^{\text{c}}$, with $\bar{y}^{\text{c}}$ the mean of the observed RGB histograms.

Figures 4, 5, 6 and 7 give the inferred segmentation for a set of images with a road and a set of images of roads and buildings. On each figure are first represented the set of original images and the pre-segmented images into super-pixels. The results obtained while the descriptor of the super-pixel is a RGB color histogram combined with the HOG are also shown.

On figures 4 and 5 the models DP-Potts and HDP-Potts yield good segmentation results. The main classes are the

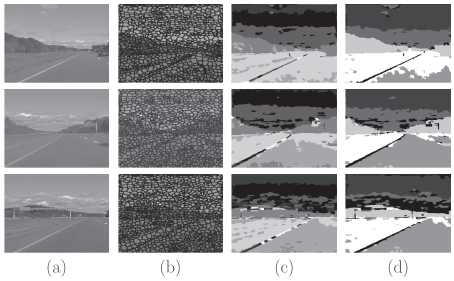

FIGURE 4. On (a) the original images for the first set with images of road; on (b) the pre-segmented images with the SLIC method; on (c) the results of segmentation using the DP-Potts algorithm applied to gray-level histograms and on (d) to RGB histograms combined to HOGs.

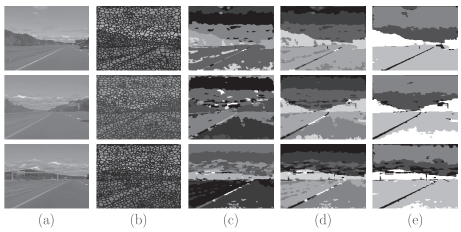

FIGURE 5. On (a) the original images for the first set with images of road; on (b) the pre-segmented images with the SLIC method; on (c) the segmentation in regions and on (d) in classes using the HDP-Potts with the gray-level histograms and on (e) in classes with the RGB histograms combined to the HOGs.

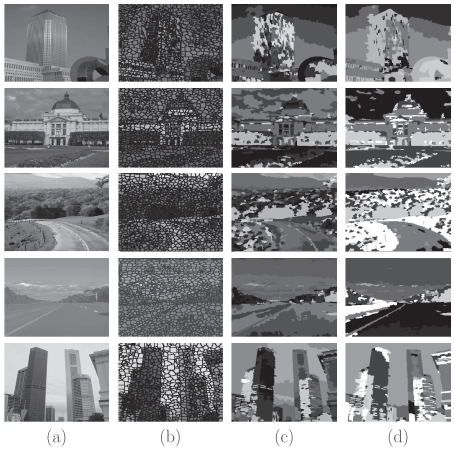

FIGURE 6. On (a) the original images for the second set with a mix of images with roads and buildings; on (b) the pre-segmented images with the SLIC method; on (c) the results of segmentation using the DP-Potts algorithm applied to gray-level histograms and on (d) to RGB histograms combined to HOGs.

road, the vegetation and the sky. The road is quite well-recognized in the two last images while for the first, due to a different illumination setup, it is only partially identified. The vegetation is also well detected. As for the class sky, with the histogram of gray levels, not only the clouds are separated, but also, it is divided regarding the changes of illumination. With the HOG, the problems of different illuminations are tackled. It should be noted that the HDP-Potts model provides valuable additional information: the partition in regions of each image.

Figures 6 and 7 show the results of segmentation for a mixed set of images with, as main classes, road, vegetation

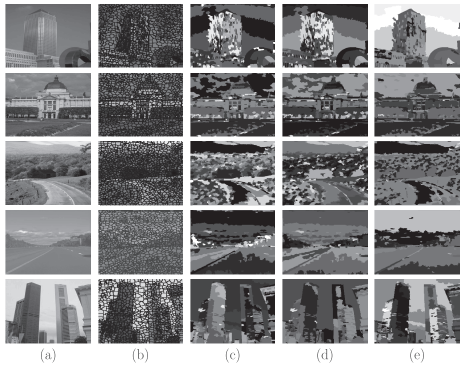

FIGURE 7. On (a) the original images for the second set with a mix of images with roads and buildings; on (b) the pre-segmented images with the SLIC method; on (c) the segmentation in regions and on (d) in classes, both using the HDP-Potts with the gray-level histograms and on (e) with the RGB histograms combined to the HOGs.

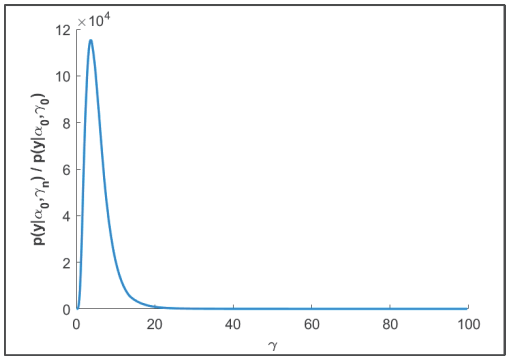

FIGURE 8. Estimation of the marginal likelihood on a grid of values $\{\gamma_m\}_{m\geq 1}$ for the set of road images and for a pre-segmentation using the SLIC method.

and buildings. The road and vegetation are quite well recognized in the different images but the buildings are not always identified as belonging to the same class. Indeed, they are very textured, therefore the variability intra-class is high.

C. ESTIMATION OF HYPERPARAMETERS

To illustrate the relevance of the SMC algorithm proposed in section V, it is applied to the set of road images for the estimation of the scalar parameter of the HDP-Potts model γ. The algorithmic setting is the following:

- Potts model granularity parameter in the HDP-Potts model: $\beta = 1$
- scale parameter for the first DP in the HDP-Potts model: $\alpha_0 = 1$
- a grid of $M = 2000$ test values $\{\gamma_m\}_{1\leq m\leq M}$ for the HDP scale parameter decreasing from 100 to 0.1 in logarithmic scale
- number of particles in parallel: $I = 100$
- parameter of the prior Dirichlet distribution on the observed gray level histograms: $\varphi_0 = 10^4 \times \bar{y}$

Figure 8 shows the estimation of the marginal likelihood with the SMC. As intended, it exhibits a maximum for a value of γ here equal to 3.65. It is taken as the estimation of the HDP hyperparameter in the maximum likelihood sense. Using this estimate of γ, the HDP-Potts method achieves

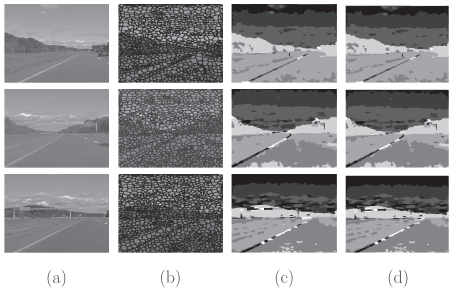

FIGURE 9. On (a) the true images; on (b) the pre-segmented images into roughly 1000 super-pixels; on (c) the results of segmentation with the SMC applied to the HDP-Potts for the estimation of γ with $N_{\text{bi}} = 16$; on (d) the results of segmentation with the HDP-Potts applied with the hyperparameters set as in section VII-B but with γ set to its optimal value.

satisfying segmentation results for the set of road images, as depicted in figure 9.

To validate the SMC, one could think of generating a set of images directly with the prior model and then recovering the hyperparameters values with the SMC algorithm. However, the images thus obtained are quite irregular and therefore useless for our purpose. Indeed, the proposed algorithms aim at segmenting quite homogeneous images.

VIII. CONCLUSION

In this article, we propose a new method for the joint segmentation of a set of images with shared classes, combining the hierarchical Dirichlet process and the Potts model. The Bayesian non parametric approach allows us to automatically infer the number of classes from the data, while the Markov field classically ensures spatial smoothness. The posterior distribution is intractable and a Gibbs sampling is therefore derived to explore it. Then, to speed up the convergence of the chain, a generalized Swendsen-Wang algorithm is described. Finally, an original approach consisting of two interacting sequential Monte Carlo samplers is presented for the estimation of the model hyperparameters. Results of segmentation on both toy and natural images show the usefulness of the proposed algorithms.

As a perspective of this work, the prior model can be rewritten using the stick-breaking construction so that a variational method can be used for approximate posterior inference. It could be computationally more efficient. Also, the results show that it may be of interest to choose an appropriate texture representation for very textured images. Finally, it would be worth theoretically studying the influence of the hyperparameters (scale factors and granularity parameter) on the mean number of classes.

REFERENCES

[1] M. Albughdadi, L. Chaari, J.-Y. Tourneret, F. Forbes, and P. Ciuciu, "A Bayesian non-parametric hidden Markov random model for hemodynamic brain parcellation," *Signal Process.*, vol. 135, pp. 132–146, Jun. 2017.

[2] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.

[3] P. Orbanz and J. M. Buhmann, "Nonparametric Bayesian image segmentation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 25–45, May 2007.

[4] E. B. Sudderth and M. I. Jordan, "Shared segmentation of natural scenes using dependent Pitman–Yor processes," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 1, Dec. 2008, pp. 1585–1592.

[5] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1566–1581, Dec. 2006.

[6] J. Sodjo, A. Giremus, F. Caron, J.-F. Giovannelli, and N. Dobigeon, "Joint segmentation of multiple images with shared classes: A Bayesian nonparametrics approach," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Jun. 2016, pp. 1–5.

[7] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels," EPFL, Lausanne, Switzerland, Tech. Rep. 149300, Jun. 2010.

[8] D. M. Higdon, "Auxiliary variable methods for Markov chain Monte Carlo with applications," *J. Amer. Stat. Assoc.*, vol. 93, no. 442, pp. 585–595, Jun. 1998.

[9] R. H. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Phys. Rev. Lett.*, vol. 58, pp. 86–88, Jan. 1987.

[10] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 68, no. 3, pp. 411–436, 2006.

[11] D. Dahl, "Model-based clustering for expression data via a Dirichlet process mixture model," in *Bayesian Inference for Gene Expression and Proteomics*, K.-A. Do, P. Müller, and M. Vannucci, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[12] T. S. Ferguson, "Prior distributions on spaces of probability measures," *Ann. Statist.*, vol. 2, no. 4, pp. 615–629, Jul. 1974.

[13] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statist. Sinica*, vol. 4, pp. 639–650, Mar. 1994.

[14] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[15] F. Y. Wu, "The potts model," *Rev. Mod. Phys.*, vol. 54, pp. 235–268, Jan. 1982.

[16] M. Pereyra, N. Dobigeon, and H. Batatia, "Segmentation of skin lesions in 2-D and 3-D ultrasound images using a spatially coherent generalized rayleigh mixture model," *IEEE Trans. Med. Imag.*, vol. 31, no. 8, pp. 1509–1520, Aug. 2012.

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

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

[19] R. G. Edwards and A. D. Sokal, "Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm," *Phys. Rev. D, Part. Fields*, vol. 38, no. 6, no. 442, pp. 2009–2012, Sep. 1988.

[20] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *J. Amer. Stat. Assoc.*, vol. 93, no. 443, pp. 1032–1044, Aug. 1998.

[21] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models* (Springer Series in Statistics). Springer, 2005. [Online]. Available: https://www.springer.com/fr/book/9780387402642

[22] A. Rosenberg and J. Hirschberg, "V-measure: A conditional entropy-based external cluster evaluation measure," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn.*, 2007, pp. 410–420.

[23] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006. [Online]. Available: https://www.springer.com/fr/book/9780387310732

[24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

JESSICA SODJO received the M.Sc. degree in signal and image processing from the University of Rennes, France, in 2014, and the Ph.D. degree in image processing from the University of Bordeaux, France, in 2018. She is currently a Lecturer with the University of French Guiana, France. Her research interests include statistical image processing and Monte Carlo sampling methods.

AUDREY GIREMUS received the engineering and Ph.D. degrees in signal processing from the École Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, France, in 2002 and 2005, respectively. She is currently an Associate Professor with the University of Bordeaux, France. Since 2006, she has been with the IMS Laboratory, Signal and Image Research Group. Her research interests include statistical signal processing and optimal filtering techniques applied to navigation, tracking, mobile communications, and biomedical engineering.

NICOLAS DOBIGEON (S'05–M'08–SM'13) was born in Angoulême, France, in 1981. He received the engineering degree in electrical engineering from ENSEEIHT, Toulouse, France, in 2004, the M.Sc. degree in signal processing from Toulouse INP, in 2004, and the Ph.D. degree and the Habilitaion à Diriger des Recherches degree in signal processing from Toulouse INP, in 2007 and 2012, respectively.

From 2007 to 2008, he was a Postdoctoral Research Associate with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. Since 2008, he has been with INP-ENSEEIHT Toulouse, University of Toulouse, where he is currently a Professor. He conducts his research within the IRIT Laboratory, Signal and Communications Group, and he currently holds an AI Research Chair with the Artificial and Natural Intelligence Toulouse Institute (ANITI). His recent research activities include statistical signal and image processing, with a particular interest in Bayesian inverse problems and applications to remote sensing, and biomedical imaging and microscopy. He is also a Junior Member of Institut Universitaire de France (IUF).

FRANÇOIS CARON was born in France, in 1980. He received the M.Eng. degree from the Institut Superieur d'Electronique du Nord (ISEN), France, in 2003, and the Ph.D. degree from the University of Lille, France, in 2006. He was a Postdoctoral Researcher with the Department of Computer Science, University of British Columbia, Vancouver, BC, Canada, and a full-time Researcher with the INRIA Bordeaux, France. He is currently an Associate Professor with the Department of Statistics and a Tutorial Fellow with the Keble College, University of Oxford. His research interest includes Bayesian/Monte Carlo methods. He is also a Fellow of the Alan Turing Institute, London, and the Co-Director of the Oxford-Warwick EPSRC-MRC Centre for Doctoral Training in modern statistical Science (OxWaSP).

• • •