A generalized Swendsen-Wang algorithm for Bayesian nonparametric joint segmentation of multiple images

Jessica Sodjo, Audrey Giremus, Nicolas Dobigeon, Jean-François Giovannelli

A GENERALIZED SWENDSEN-WANG ALGORITHM FOR BAYESIAN NONPARAMETRIC JOINT SEGMENTATION OF MULTIPLE IMAGES

Jessica Sodjo[(1)]*, Audrey Giremus*[(1)]*, Nicolas Dobigeon*[(2)] *and Jean-François Giovannelli*[(1)]

[(1)] IMS (Univ. Bordeaux, CNRS, BINP). F-33400 Talence, France
[(2)] IRIT/INP-ENSEEIHT Toulouse. F-31071 Toulouse, France

ABSTRACT

A generalized Swendsen-Wang (GSW) algorithm is proposed for the joint segmentation of a set of multiple images sharing, in part, an unknown number of common classes. The labels are a priori modeled by a combination of the hierarchical Dirichlet Process (HDP) and the Potts model. The HDP allows the number of regions in each image and classes to be automatically inferred while the Potts model ensures spatially consistent segmentations. Compared to a classical Gibbs sampler, the GSW ensures a better exploration of the posterior distribution of the labels. To avoid label switching issues, the best partition is estimated using the Dahl's criterion.

Index Terms— Image segmentation, Bayesian nonparametrics, Dirichlet Process, Swendsen-Wang algorithm, Potts model.

1. INTRODUCTION

In various computer vision applications ranging from medical engineering to Earth observation, image classification has been shown to be a crucial processing which still motivates numerous research works. When analyzing a collection of J images, the information shared among these images can be exploited by conducting a joint segmentation. It is expected to provide more reliable classification results than J individual classifications operated on each image separately. More precisely, a joint segmentation consists in dividing each image into $m_{j\cdot}(j = 1, \ldots, J)$ homogeneous regions and grouping the regions that share common characteristics in K classes. The number of classes is mostly considered known, but, for more flexibility, the estimation of K can be also of interest. Estimating the optimal number of classes can be formulated as a model order selection. This issue has been addressed following various approaches in the literature. One popular approach conducted within a Bayesian framework consists in sampling the joint posterior distribution of the labels and the number of classes by resorting to reversible jumps between spaces of different dimensions [1].

More recently, Bayesian nonparametric models have been advocated to overcome the computational burden required by reversible jump algorithms. In particular, the Dirichlet process (DP) [2] has been shown to be well-suited for segmenting images without requiring the prior knowledge of the number of classes. However, the DP cannot model shared classes between the images to be segmented. As an alternative, the hierarchical Dirichlet process (HDP) introduced by Teh [3] can be considered. Benefiting from this formalism, the number $m_{j\cdot}$ of regions in each image ($j = 1, \ldots, J$) and the number K of classes can be estimated in a unsupervised Monte Carlo sampling procedure.

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

Beyond this automatic selection, a key feature which should be ensured when designing a segmentation procedure is to promote the homogeneity of the considered images. Within a statistical framework, Markov random fields (MRF) [4, 5] have been a popular modeling to ensure that neighboring pixels have higher probability to be assigned to the same class. To address both order selection and spatial smoothness, we proposed in [6] a prior model combining the HDP and the Potts-MRF model to jointly segment of a collection of several images. By adopting this approach, Bayesian inference of the parameters of interest cannot be performed analytically and, consequently, a Gibbs sampler has be derived. The method has been applied on a toy example. However, resorting to this so-called HDP-MRF model to analyze images of significantly higher size, e.g., extracted from the LabelMe database[1], leads to severe computational issues when using the crude instance of Gibbs sampler developed in [6]. This paper specifically proposes an algorithmic strategy to alleviate this difficulty by a threefold contribution. First, it implements a pre-segmentation into super-pixels which reduces the complexity of the problem. Then, it derives a generalized Swendsen-Wang (GSW) [7] based algorithm for the HDP-MRF model. It consists in introducing link variables between pixels of the same regions; this link variables do not modify the posterior distribution but they can be efficiently sampled jointly with the variables of interest, which speeds up convergence. Finally, a Dahl's criterion is considered to infer the optimal partition within the sampled ones.

The sequel of the paper is organized as follows. Section 2 introduces the proposed prior model. The HDP and Potts model are described and the prior distributions are provided. In Section 3, the GSW algorithm is detailed and the sampling equations are derived. Results obtained on a set of several images are presented in Section 4 and concluding remarks are reported in Section 5.

2. BAYESIAN NONPARAMETRIC MODEL

2.1. Notations and observation model

Let us consider a set of J images $\mathcal{I}_j$ to be jointly segmented ($j = 1, \ldots, J$). To reduce the computational cost due to sampling-based exploration, a common approach consists in dividing each image $\mathcal{I}_j$ ($j = 1, \ldots, J$) into N_j super-pixels[2]. The observation associated with the nth super-pixel ($n = 1, \ldots, N_j$) in image j is denoted y_{jn} and assumed to be distributed according to a distribution f parameterized by θ_{jn}, i.e., $y_{jn}|\theta_{jn} \sim f(y_{jn}|\theta_{jn})$.

The region label associated with the nth super-pixel in the jth image is denoted c_{jn}. In an given image, a set of super-pixels that

[1] http://labelme.csail.mit.edu/Release3.0/

[2] Note that the proposed algorithm is also valid if directly applied to the pixels.

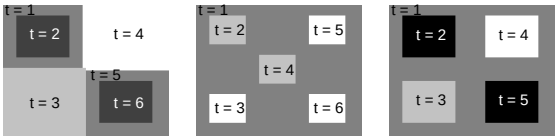

Fig. 1. Simplistic example of joint segmentation with $J = 3$ and $K = 5$. The regions in each image are numbered and the colors identify the classes. It can be noticed that there is a different number of regions in each image and some regions in these images can be assigned to the same class, such as region 5 in image 2 and region 4 in image 3, both assigned to class "white".

are assigned the same region label value is referred to as a region. At a higher level of the modeling, the class label of the tth region in the jth image is denoted d_{jt}. The images are assumed to share at most K distinct classes where the kth class is defined as the collection of all regions assigned the class label value k. An example is shown on figure 1.

All super-pixels (j, n) assigned to the kth class $(k = 1, \ldots, K)$ share the same parameter vector $\theta_{jn} = \phi_k$. Thus, assuming prior independence between the classes, the marginal distribution of the super-pixels $\boldsymbol{y} = \cup_{k=1}^{K} \boldsymbol{y}_{A_k}$ can be written as

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

where $\boldsymbol{y}_{A_k} = \{y_{jn}|(j, n) \in A_k\}$ is the set of super-pixels assigned the kth class label with $A_k = \{(j, n)|d_{jc_{jn}} = k\}$ and $h(\cdot)$ is the prior distribution of the parameters ϕ_k $(k = 1, \ldots, K)$.

2.2. Hierarchical Dirichlet process

Let $\mathbb{G}_j$ denote the unknown probability distribution of the parameter vectors θ_{jn} of the jth image $(j = 1, \ldots, J)$. Since the number of classes is assumed to be unknown, the parameters θ_{jn} can take a priori an infinite number of values. This naturally induces a nonparametric prior modeling for $\mathbb{G}_j$. Here, several parameter vectors can take the same value, hence $\mathbb{G}_j$ should be discrete. A solution is to assume that $\mathbb{G}_j$ is distributed according to a Dirichlet process (DP). The latter depends on a scalar parameter α_0 and a base measure $\mathbb{G}_0$:

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

with ψ_{jt} the parameter vector of the tth region in the jth image. More precisely, for all super-pixels n such that $c_{jn} = t$, we have $\theta_{jn} = \psi_{jt}$. In (1), $\mathbb{G}_j$ is an infinite sum of Dirac measures on the ψ_{jt}, weighted by τ_{jt}. To allow classes to be shared, all the distributions $\mathbb{G}_j$ should have common atoms ϕ_k. The adopted solution consists of defining $\mathbb{G}_0$ as a discrete measure centered on these atoms ϕ_k. These latter are unknown and assumed independently distributed according to a probability measure H with probability density function h as introduced in (2.1). Since the number of classes K is supposed unknown, a DP is chosen as prior, i.e.,

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

An interesting property with the above described model is that the probability that the nth super-pixel in the jth image is assigned to the tth region is proportional to the number ν_{jt} of super-pixels in that region. It can also be assigned to a new region t^{new} proportionally to α_0:

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

with $\boldsymbol{c}_j^{-n} = \{c_{jn'}|n' = 1, \ldots, N_j, n' \neq n\}$. When considering the tth region in the jth image, two cases are also possible: it can either be assigned to an existing class k proportionally to $m_{\cdot k}$ or to a new one proportionally to γ, where $m_{\cdot k}$ is the number of regions of all the images assigned to class k, i.e.,

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

where $\boldsymbol{d}^{-jt} = \{d_{j't'}|j' = 1, \ldots, J; t' = 1, \ldots, m_{j'\cdot}; (j', t') \neq (j, t)\}$. The prior φ induced by the HDP for the set of region labels $\boldsymbol{c} = \{c_{jn}|j = 1, \ldots, J; n = 1, \ldots, N_j\}$ and the set of class labels $\boldsymbol{d} = \{d_{jt}|j = 1, \ldots, J; t = 1, \ldots, m_{j\cdot}\}$ depends on the size of the regions, the number of regions per class and the overall number of regions denoted $m_{\cdot\cdot}$. It can be written [3]:

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

2.3. Potts model

The Potts model is a prior on the class labels [8]. With the Potts model, the image is redefined using a neighboring system on the pixels. This model allows the homogeneity of the classes to be preserved by favoring that a given pixel and its neighbors share the same class. It can be noticed that, within a super-pixel formalism, two super-pixels are defined as neighbors if they have a common ridge. The Potts prior writes

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

where $n \sim q$ means that super-pixel q is a neighbor of n and $\delta(\cdot)$ is the Kronecker symbol.

2.4. Joint prior distribution

The proposed prior distribution is

$$\Pr(\boldsymbol{c}, \boldsymbol{d}) \propto \varphi(\boldsymbol{c}, \boldsymbol{d}) \rho(\boldsymbol{c}, \boldsymbol{d}). \tag{4}$$

It consists of a combination of a global penalization φ and a local one ρ where φ ensures that the number of regions and classes do not need to be a priori fixed and ρ favors spatial homogeneity.

3. SAMPLING SCHEME

In part due to the nonparametric nature of the posterior distribution, no closed-form expressions of the Bayesian estimators can be derived. To approximate these estimators, a Gibbs sampler was derived in [6]. However, as noticed earlier, the Gibbs algorithm has poor mixing properties, which motivates the proposed contribution. The generalized Swendsen-Wang (GSW) algorithm described in what follows aims at improving the exploration of the posterior distribution [9, 10].

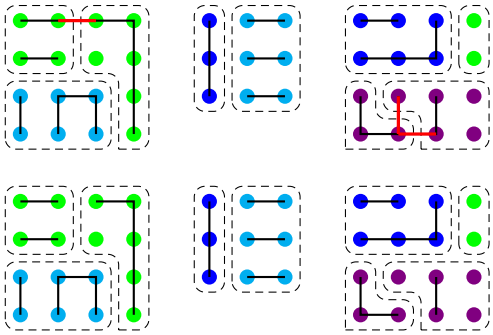

Fig. 2. Example of links sampled within the GSW algorithm. The dashed lines delimit the regions in the images, the thick lines are links and the colors represent the classes. On first line the partition that can be obtained by assigning links only based on classes and the last the links obtained based on the region labels.

3.1. Generalized Swendsen-Wang algorithm

The Swendsen-Wang algorithm introduced in [11] and generalized in [12] is a sampler derived from the Potts model to ensure a more efficient exploration of the posterior distribution of the labels. First, groups of super-pixels are formed using a latent variables. They are defined as follows: $r_{j_{nq}} = 1$ if super-pixels n and q in image j are linked and $r_{j_{nq}} = 0$ otherwise. Then, linked super-pixels are grouped into spin-clusters and their labels are simultaneously updated. Moreover, the introduction of this latent variables should not modify the corresponding marginalized posterior distribution of the labels, i.e., $\sum_{\boldsymbol{r}} p(\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{r}|\boldsymbol{y}) = p(\boldsymbol{c}, \boldsymbol{d}|\boldsymbol{y})$ with $\boldsymbol{r} = \{r_{j_{nq}}|j = 1, \ldots, J; n, q = 1, \ldots, N_j\}$. Since two sets of labels have been introduced in the context of joint segmentation considered in this paper, namely $\boldsymbol{c}$ and $\boldsymbol{d}$, an important issue to addressed is to identify conditionally to which of them the links should be sampled.

A first solution would consist of introducing the links with respect to the class assignment. However, since different regions can be assigned to the same class in an image, super-pixels of different regions could be sampled jointly which is not desirable. As an alternative, the links can also be based on the region labels. Conditionally to the partition, the probability that two super-pixels are not linked is written $\Pr(r_{j_{nq}} = 0|\boldsymbol{c}, \boldsymbol{d}) \propto \exp\left(-\beta\lambda\delta(c_{jn}, c_{jq})\delta(d_{jc_{jn}}, d_{jc_{jq}})\right)$, where λ is a parameter to be adjusted. Since pixels in the same region necessarily belong to the same class, $\delta(c_{jn}, c_{jq})\delta(d_{jc_{jn}}, d_{jc_{jq}}) = 1$ only when $c_{jn} = c_{jq}$. It follows $p(\boldsymbol{r}|\boldsymbol{c}, \boldsymbol{d}) = p(\boldsymbol{r}|\boldsymbol{c})$ and

$$\Pr(r_{j_{nq}} = 1|\boldsymbol{c}) = 1 - \exp\left(-\beta\lambda\delta(c_{jn}, c_{jq})\right) \tag{5}$$

As long as only the partition matters, a GSW-based Gibbs sampling consists in first sampling the links $\boldsymbol{r} \sim \Pr(\boldsymbol{r}|\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{y})$, then the region labels, $\boldsymbol{c} \sim \Pr(\boldsymbol{c}|\boldsymbol{r}, \boldsymbol{d}, \boldsymbol{y})$ and finally the class labels $\boldsymbol{d} \sim \Pr(\boldsymbol{d}|\boldsymbol{c}, \boldsymbol{r}, \boldsymbol{y})$. In the following, these conditional distributions are detailed.

3.2. Sampling of the links

Regarding the conditional distribution of the links, it can be noticed that conditionally to the partition, the links are independent of the observations, $\Pr(\boldsymbol{r}|\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{y}) = \Pr(\boldsymbol{r}|\boldsymbol{c})$. Thus, for all super-pixels, the links can be independently sampled according to (5).

3.3. Sampling of the region label

Once the links have been sampled, linked super-pixels are grouped into spin-clusters. Thanks to the Swendsen-Wang algorithm, the labels of the regions are sampled simultaneously for all pixels in the same spin-cluster. In the next paragraphs, the following notations are adopted: for the lth spin-cluster in the jth image, the corresponding set of super-pixels, the set of region labels and the set of observations are denoted C_{jl} with size $|C_{jl}|$, $\boldsymbol{c}_{jl}$ with $\boldsymbol{c}_{jl} = \{c_{jn}|n \in C_{jl}\}$ and $\boldsymbol{y}_{jl}$, respectively.

According to the HDP prior, the super-pixels in C_{jl} can be assigned to an existing region or a new one. The conditional probability of having $\boldsymbol{c}_{jl} = t \leq m_{j\cdot}$ is proportional to the probability that the first super-pixel in C_{jl} is assigned to the tth region (ν_{jt}^{-jl}), then the second ($\nu_{jt}^{-jl} + 1$) till the last one ($\nu_{jt}^{-jl} + |C_{jl}| - 1$). It is also proportional to the distribution of the observations attached to C_{jl} conditionally to the observations attached to super-pixels in class d_{jt}, $f(\boldsymbol{y}_{jl}|\boldsymbol{y}_{A_{d_{jt}}^{-jl}}) = f(\boldsymbol{y}_{jl}, \boldsymbol{y}_{A_{d_{jt}}^{-jl}})/f(\boldsymbol{y}_{A_{d_{jt}}^{-jl}})$ with $A_{d_{jt}}^{-jl}$ the set of super-pixels in class d_{jt} except the ones in spin-cluster C_{jl}. Similarly, the probability of having $\boldsymbol{c}_{jl} = t^{\text{new}}$ is proportional to $\alpha_0 \times 1 \times \cdots \times (|C_{jl}| - 1)$ and $p(\boldsymbol{y}_{jl}\,|\,\boldsymbol{c}_{jl} = t^{\text{new}}, \boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{y}^{-jl})$ obtained by integrating out within all possibilities for the class that can be assigned to the new region. It follows

$$\begin{aligned}&\Pr(\boldsymbol{c}_{jl} = t \leq m_{j\cdot}|\boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{r}, \boldsymbol{y}) \\ &\propto \frac{\Gamma(\nu_{jt}^{-jl} + |C_{jl}|)}{\Gamma(\nu_{jt}^{-jl})} \exp\left(-\sum_{q \in \mathcal{V}_{C_l}} \beta\,\lambda\,\delta(t, c_{jq})\right) f(\boldsymbol{y}_{jl}|\boldsymbol{y}_{A_{d_{jt}}^{-jl}})\end{aligned} \tag{6}$$

where $\mathcal{V}_{C_l}$ is the set of super-pixels neighbors of the super-pixels in spin-cluster C_l and

$$\begin{aligned}&\Pr(\boldsymbol{c}_{jl} = t^{\text{new}}|\boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{r}, \boldsymbol{y}) \\ &\propto \alpha_0\,\Gamma(|C_{jl}|)\,p(\boldsymbol{y}_{jl}|\boldsymbol{c}_{jl} = t^{\text{new}}, \boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{y}^{-jl})\end{aligned} \tag{7}$$

with

$$\begin{aligned}&p(\boldsymbol{y}_{jl}\,|\,\boldsymbol{c}_{jl} = t^{\text{new}}, \boldsymbol{c}^{-jl}, \boldsymbol{d}, \boldsymbol{y}^{-jl}) \\ &\propto \left\{\sum_{k=1}^{K} m_{\cdot k} \exp\left(\sum_{q \in \mathcal{V}_{C_l}} \beta\delta(d_{jc_{jq}}, k)\right) + \gamma\right\}^{-1} \\ &\left\{\sum_{k=1}^{K} m_{\cdot k} \exp\left(\sum_{q \in \mathcal{V}_{C_l}} \beta\delta(d_{jc_{jq}}, k)\right) f\left(\boldsymbol{y}_{jl}\,|\,\boldsymbol{y}_{A_k^{-jl}}\right) + \gamma f(\boldsymbol{y}_{jl})\right\}\end{aligned} \tag{8}$$

where $f(\boldsymbol{y}_{jl}) = \int[\prod_{n \in C_{jl}} f(y_{jn}|\phi_{k^{\text{new}}})]h(\phi_{k^{\text{new}}})\mathrm{d}\phi_{k^{\text{new}}}$. In the case of a new region, the assigned class label is sampled following

$$\begin{aligned}&\Pr(d_{jt^{\text{new}}} = k\,|\,\boldsymbol{c}, \boldsymbol{d}^{-jt^{\text{new}}}) \\ &\propto \begin{cases} m_{\cdot k} \exp\left(\sum_{q \in \mathcal{V}_{t^{\text{new}}}} \beta\,\delta(d_{jc_{jq}}, k)\right) f(\boldsymbol{y}_{jl}\,|\,\boldsymbol{y}_{A_k^{-jl}}) & \text{if } k \leq K \\ \gamma\, f(\boldsymbol{y}_{jl}) & \text{if } k = k^{\text{new}} \end{cases}\end{aligned} \tag{9}$$

3.4. Sampling of the class label

As for the region labels, the probability that the tth region in the jth image can be assigned to an existing kth class is proportional to the number $m_{\cdot k}^{-jt}$ of regions assigned to the kth class, omitting the

considered one. Conversely, it can be assigned to a new class with probability proportional to γ, leading to

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

3.5. Deriving the Bayesian estimators

In a Bayesian framework, the best partition is generally estimated using the marginal maximum a posteriori estimator. However, when facing to the nonparametric Bayesian models, not only the well-known label-switching problems may occur but the number of classes varies within the exploration routine. A re-labeling is thus needed, which may be computationally prohibitive. An alternative consists of directly choosing the partition that maximizes the posterior distribution. However, this strategy does not take into account all the richness of the information described by the distribution of interest. Motivated by numerous works in the statistical community [13, 14], the approach adopted in this paper consists of computing an optimal label assignment by selecting the best partition in the sense that it minimizes a given loss function.

Let us denote $\boldsymbol{\kappa} = \{d_{jc_{jn}}; j = 1, \ldots, J; n = 1, \ldots, N_j\}$ the set of classes assigned to each super-pixels in the images with $\kappa_{jn} = d_{jc_{jn}}$. The *optimal* estimate $\hat{\boldsymbol{\kappa}}$ is chosen here as the one minimizing an appropriate loss function associated with partitions and defined[3] as [16].

$$\hat{\boldsymbol{\kappa}} = \underset{\boldsymbol{\kappa}^{(i)} \in \{\boldsymbol{\kappa}^{(1)}, \ldots, \boldsymbol{\kappa}^{(I)}\}}{\operatorname{argmin}} \sum_{j,n,q} \left(\delta(\kappa_{jn}^{(i)}, \kappa_{jq}^{(i)}) - \zeta_{j_{nq}} \right)^2 \tag{11}$$

with

$$\zeta_{j_{nq}} = \frac{1}{I} \sum_{i=1}^{I} \delta(\kappa_{jn}^{(i)}, \kappa_{jq}^{(i)})$$

and $\boldsymbol{\kappa}^{(i)}$ the vector $\boldsymbol{\kappa}$ at the ith iteration of the HDP-GSW algorithm and I the total number of iterations.

4. RESULTS

The algorithm has been applied on three images ($J = 3$) of the LabelMe database of size 256×256. Each image has been pre-segmented in approximately 500 super-pixels using the SLIC algorithm [17]. An observation for a super-pixel is defined as a histogram of 120-bins and $f(y_{jn}|\theta_{jn}) = \text{Mult}(\theta_{jn})$ where $\text{Mult}(.)$ is the multinomial distribution.

For the prior model, h is chosen as a Dirichlet density, $\text{Dir}(\varpi\tilde{\pi})$ with $\tilde{\pi}$ taken as the normalized sum of the histograms and ϖ a scalar parameter, here, $\varpi = 10^4$. It should be noted that ϖ influences the inference of the number of classes, the greater it is, the less classes will be proposed. The hyperparameters are chosen as: $\alpha = 1, \gamma = 1$, $\beta = 0.25$ and $\lambda = 10$.

The figure 3 represents the logarithm of the non-normalized posterior distribution of the labels of region and class corresponding to each sample of the HDP-GSW algorithm (blue) and the standard Gibbs for the HDP-Potts model (black). It can be seen that the exploration derived by the classical Gibbs algorithm is stuck in a local maximum, contrary to the HDP-GSW one.

[3]Note that this definition is equivalent to the minimization-driven procedure proposed in [15]

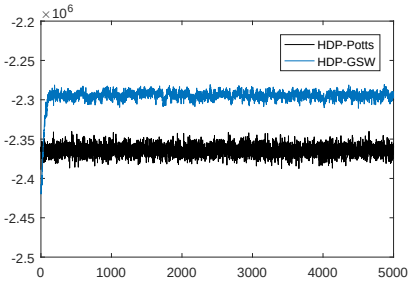

Fig. 3. Logarithm of the non-normalized posterior distribution of the labels corresponding to the samples obtained with the classical Gibbs and HDP-GSW algorithms depending on the number of the iteration.

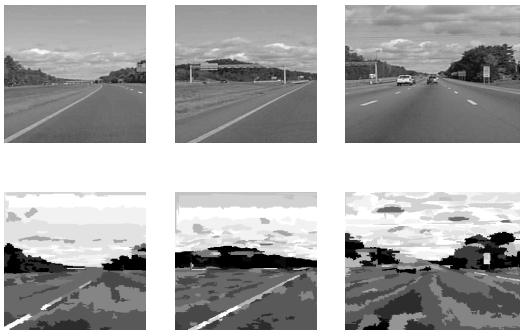

Fig. 4. On first line, the true images and on second line, the partition obtained with the HDP-GSW algorithm

The best segmentation in terms of Dahl's criterion is shown in figure 4. The resulted segmentation take into account the shared classes as expected. However, the images are over-segmented ($K = 18$), this may be due to the fact that the images have not been taken in the same physic conditions (brightness, sensor, ...) and the histogram may not be the best attribute to characterize the classes.

5. CONCLUSION

An algorithm for segmenting jointly multiple images is proposed to overcome the computational issues encountered while using the HDP-Potts model introduced in [6]. It combines the HDP and the GSW algorithm. While the HDP allows the number of classes to be derived automatically, the Potts model ensures a spatial homogeneity in each image and the GSW improves the exploration scheme of the posterior distribution of the labels. Obtained results shows that the HDP-GSW algorithm exploration is more efficient than the classical HDP-Potts one, yet, the inference is sensitive to not only the value of the hyperparameters but also to the way of describing the observations. To overcome these issues, we are currently investigating on the one hand the use of sequential Monte Carlo samplers [18] to adjust the hyperparameters while sampling the labels with the best hyperparameters and on the other hand the use of more relevant and robust descriptors of the classes (e.g. the texture).

6. REFERENCES

[1] 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.

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

[3] 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.

[4] 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.

[5] 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.

[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," *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, 2016.

[7] D. Higdon, "Auxiliary variable methods for Markov chain Monte Carlo with applications," *Journal of the American statistical Association*, vol. 93, no. 442, pp. 585–595, 1998.

[8] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields," *IEEE Transactions on geoscience and remote sensing*, vol. 41, pp. 675–686, 2003.

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

[10] 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.

[11] R. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Physical Review Letters*, vol. 58, no. 2, pp. 86–88, 1987.

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

[13] F. Caron, Y. W. Teh, and T. B. Murphy, "Bayesian nonparametric Plackett-Luce models for the analysis of preferences for college degree programmes," in *The Annals of Applied Statistics*, vol. 8, no. 2, 2014, pp. 1145–1181.

[14] A. Fritsch and K. Ickstadt, "Improved criteria for clustering based on the posterior similarity matrix," in *Bayesian Analysis*, vol. 4, no. 2, 2009, pp. 367–392.

[15] D. Binder, "Bayesian cluster analysis," *Biometrika*, vol. 65, no. 1, pp. 31–38, 1978.

[16] D. Dahl, *Model-based clustering for expression data via a Dirichlet process mixture model in Bayesian inference for gene expression and proteomics, Kim-Anh Do, Peter Müller, Marina Vannucci*. Cambridge University Press, 2006.

[17] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels," *EPFL Technical report 149300*, June 2010.

[18] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *J. R. Statistic Soc. B*, vol. 68, pp. 411–436, 2006.