POTTS MODEL PARAMETER ESTIMATION IN BAYESIAN SEGMENTATION OF PIECEWISE CONSTANT IMAGES

Roxana-Gabriela ROSU †, *Jean-François GIOVANNELLI* †, *Audrey GIREMUS* †, *Cornelia VACAR* ‡

† Univ. Bordeaux, IMS, UMR 5218, F-33400 Talence, France
‡ FEI Visualization Sciences Group

ABSTRACT

The paper presents a method for estimating the parameter of a Potts model jointly with the unknowns of an image segmentation problem. The method addresses piecewise constant images degraded by additive noise. The proposed solution follows a Bayesian approach, that yields the posterior law for all the unknowns (labels, gray levels, noise level and Potts parameter). It is explored by means of MCMC stochastic sampling, more precisely, by Gibbs algorithm. The estimates are then computed from these samples. The estimation of the Potts parameter is challenging due to the intractable normalizing constant of the model. The proposed solution is based on pre-computing the value of this normalizing constant for different image dimensions and number of classes, this being the novelty of this paper. The segmentation results are as satisfying as those obtained when tuning the parameter by hand.

Index Terms — Bayes, Potts model, normalizing constant, unsupervised segmentation, stochastic sampling.

1. INTRODUCTION

The paper addresses the problem of image segmentation of piecewise constant images in a Bayesian framework. A common approach relies on the Potts model that takes into account the spatial correlation between neighbouring pixels. This model is driven by the β parameter, also called the granularity coefficient. The value of this parameter dictates the amount of spatial correlation introduced by the model. More precisely, it affects the size and number of the regions obtained as a result of image segmentation. A too small value for this parameter leads to the formation of small regions in the image. As a consequence, the image can be over-segmented. On the other hand, a too large value of β might lead to under-segmentation since the image is segmented in a small number of large size regions.

Hand adjusting the parameter's value might be a very time-consuming task, especially in the case of big data sets. For this reason, the estimation of the granularity coefficient is of major interest. However, its estimation is challenging since the likelihood requires the normalizing constant or partition function of the Potts model. This term, defined in section 2, is a sum over all possible label configurations in the image, which makes its evaluation prohibitive [1]. As a result, in the majority of existing methods, the value of β parameter is fixed by hand. However, a few approaches estimate β jointly with all the unknown quantities. This is achieved by sampling the a posteriori conditional law for the parameter by using methods such as Likelihood free Metropolis Hastings [2], where the exact computation of the normalizing constant is avoided by means of approximation (see also [3]).

The novelty of the paper is in pre-computation of the partition function for several image sizes and numbers of classes. An additional advantage of the approach, compared to the existing ones, is a significant reduction of computational cost.

In order to compute the estimates, the posterior law for all the quantities is calculated by the Bayes' rule. However, given its complex expression and the intricate posterior dependence between the unknowns, its handling cannot be directly achieved and MCMC stochastic sampling is employed, more specifically a Gibbs loop. Thus, by making use of the pre-computed values of the partition function, the conditional posterior for β is sampled jointly with all the unknowns, namely the labels, the gray levels of each class and the noise precision. Finally, the estimates for the labels are computed as marginal posterior maximizers and the other parameters as posterior means based on the simulated samples.

2. HIERARCHICAL MODEL

The observed image $\boldsymbol{y}$ is modelled as the addition of a noise $\boldsymbol{n}$ over the input image $\boldsymbol{x}$. All the images are vectorised, of size P. The observation model is then:

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n} \tag{1}$$

The segmentation's goal is splitting the image in different regions according to a homogeneity criterion. Here, the input image is considered to be composed of K classes of constant gray level, so that each class k is represented by a unique value ν_k. The proposed segmentation approach is based on a hidden label field, denoted by $\boldsymbol{z}$ so that each pixel z_p, $p = 1, \ldots, P$, receives a discrete label $k \in \{1, 2...K\}$. As a result, the estimation of this label field gives the image segmentation.

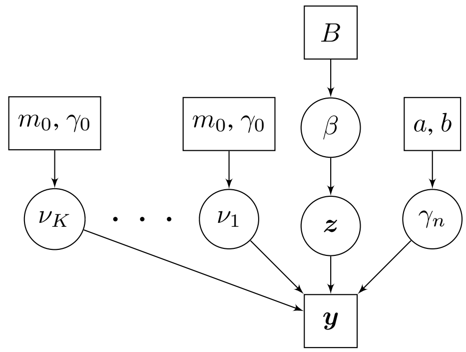

Fig. 1: Hierarchical model: the round nodes / the square ones show the estimated / fixed quantities.

In a Bayesian framework, the information on the unknowns is modelled by means of a priori laws. The following paragraphs present the chosen prior laws for each unknown.

2.1. Label field model

As mentioned before, a Potts field models the label $\boldsymbol{z}$:

$$p(\boldsymbol{z}|\beta) = C(\beta)^{-1} \exp\Big\{\beta \sum_{p\sim a} \delta(z_p; z_a)\Big\} \tag{2}$$

where $\sim$ stands for the neighbour relation between pixels in a first order system and δ is the Kronecker function that is to say $\delta(u; v)$ is 1 if $u = v$ and 0 if not.

The model depends on β (called Potts parameter or granularity coefficient) tuning the spatial correlation. It includes a normalizing coefficient also named the partition function:

$$C(\beta) = \sum_{\boldsymbol{z}\in\{1,\dots K\}^P} \exp\Big\{\beta \sum_{p\sim a} \delta(z_p; z_a)\Big\} \tag{3}$$

Its knowledge is naturally crucial in order to estimate β, since it is involved in the likelihood of β attached to a given configuration. Its analytical expression is unknown, except for the Ising field[1]. Numerically, it is a colossal summation over the K^P possible configurations and the exhaustive exploration of these configurations is impossible (except for minuscule images). However, based on stochastic simulation, we have precomputed the partition function for several image sizes and numbers of classes, as explained in Appendix A. It remains a huge task but it is attainable: it required several weeks of intensive computation (on a standard PC), but it is done once for all. This is the keystone for the estimation of β here.

2.2. Image and class levels model

Here, we formalize the rearrangement of the ν_k in a piecewise constant image for a label configuration $\boldsymbol{z}$.

[1]For the Ising field, i.e. $K = 2$, the partition has been explicitely known for a long time [4] (but yet a certain part of the literature seems unaware). It has been included in several papers [5, 6].

For each pixel p consider $t_k(z_p) = \delta(z_p; k)$ that is to say $t_k(z_p) = 1$ if the pixel p is in the class k and 0 otherwise. Then collect the $t_k(z_p)$ in K binary vectors with size P: $\boldsymbol{t}_k(\boldsymbol{z}) = [t_k(z_1), \dots t_k(z_P)]$.

Starting from the vectors $\boldsymbol{t}_k(\boldsymbol{z})$ and from the class levels ν_k, the rearrangement as a piecewise constant image writes:

$$\boldsymbol{x} = \sum_{k=1}^{K} \nu_k \boldsymbol{t}_k(\boldsymbol{z}) \tag{4}$$

When it comes to the a priori law definition, a Gaussian distribution is considered for modelling the class levels. Starting from the hypothesis that the levels are independent, and considering that each one has the same parameters m_0 et γ_0, the prior density for the K levels is written as:

$$\begin{aligned} p(\boldsymbol{\nu}) &= \prod_{k=1}^{K} \mathcal{N}(\nu_k; m_0, \gamma_0^{-1}) \\ &= (2\pi)^{-K/2} \gamma_0^{K/2} \exp\Big\{ -\gamma_0 \sum_{k=1}^{K} (\nu_k - m_0)^2/2\Big\} \end{aligned} \tag{5}$$

where $\boldsymbol{\nu} = [\nu_1, \dots, \nu_K]$ denotes the vector containing the gray levels for all the classes.

2.3. Noise model and likelihood

Regarding the noise, a Gaussian, white, homogenous and zero-mean with precision γ_n model is considered:

$$\begin{aligned} p(\boldsymbol{n}|\gamma_n) &= \mathcal{N}(\boldsymbol{n}; 0, \gamma_n^{-1}\boldsymbol{I}) \\ &= (2\pi)^{-P/2} {\gamma_n}^{P/2} \exp\{-\gamma_n ||\boldsymbol{n}||^2/2\} \end{aligned}$$

Starting from this model together with the observation model (1) and the rearranging equation (4), the likelihood is:

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{z}, \gamma_n, \beta) &= \mathcal{N}(\boldsymbol{y}; \sum_k \nu_k \boldsymbol{t_k}(\boldsymbol{z}); \gamma_n^{-1}\boldsymbol{I}) \\ =&(2\pi)^{-P/2}\gamma_n^{P/2} \ \exp\Big\{ -\gamma_n ||\boldsymbol{y} - \sum_k \nu_k \boldsymbol{t_k}(\boldsymbol{z})||^2/2\Big\} \end{aligned} \tag{6}$$

It is the probability of the data given the unknowns.

2.4. Hyperparameters prior laws

The hyperparameters are the Potts parameter β and the noise precision parameter γ_n. A classical choice is based on a conjugate prior, that will ease the further computations. With this in mind, a Gamma prior density is defined for γ_n:

$$p(\gamma_n) = \mathcal{G}(\gamma_n; a, b) = \Gamma(a)^{-1}\, b^a\, \gamma_n^{a-1} \ \exp\{-b\gamma_n\} \tag{7}$$

When it comes to β, a conjugate choice is not an obvious one, given the expression of the Potts model. A uniform prior on an interval $[0, B]$ is considered as a reasonable choice:

$$p(\beta) = \mathcal{U}_{[0,B]}(\beta) \tag{8}$$

where B is defined as the maximum possible value of β.

2.5. Posterior law

The posterior distribution is the law for all the unknowns given the observation. It is first based on the prior distribution for the unknown parameters (2), (5), (7) and (8). It is naturally also based on the likelihood (6). Finally, its construction relies on conditional independance properties encoded in the hierarchical model given in Fig. 1.

$$\begin{aligned} p(\boldsymbol{\nu}, \boldsymbol{z}, \gamma_n, \beta | \boldsymbol{y}) &\propto \gamma_n^{P/2+\alpha_b-1}\, \mathcal{U}_{[0,B]}(\beta) \\ &\times \exp\Big\{ -\gamma_n ||\boldsymbol{y} - \sum_k \nu_k \boldsymbol{t}_k(\boldsymbol{z})||^2/2 \Big\} \\ &\times \exp\Big\{ -\gamma_0 \sum_k (\nu_k - m_0)^2/2 \Big\} \\ &\times \exp\Big\{ \sum_{p\sim a} \beta \delta(z_p; z_a) \Big\} \exp\{-b\gamma_n\} \end{aligned} \tag{9}$$

The intricate expression of this law stems from the a posteriori dependence between the problem's unknowns, in spite of their a priori independence. As a result, the estimations cannot be inferred directly. To this purpose, MCMC stochastic sampling is employed, as presented in the next section.

3. SAMPLING ALGORITHM

To explore the posterior, we resort to a Gibbs loop that splits the global sampling problem in four easier sub-problems. More precisely, the conditional posterior for each unknown is successively sampled in an iterative way. The samples form a Markov chain whose distribution converges to the posterior. The pseudo-code is given in algorithm 1.

First of all, the posterior for each unknown conditionally on the other unknowns must be obtained. Each one is obtained by keeping from the expression of the posterior (9) only the factors containing that parameter. It should be noted that certain parameters will be simplified so that, eventually, the conditional posterior for a given unknown will depend only on some other unknowns, in accordance with the hierarchical model given in Fig. 1. The conditional posterior laws are given below.

Algorithm 1 Gibbs sampling algorithm

```
Initialize γ_n^(0), ν^(0), z^(0), β^(0)
for q = 1, 2, ...Q do
    sample γ_n^(q) under p(γ_n | ν^(q-1), z^(q-1), β^(q-1), y)
    for k = 1, 2, ...K do
        sample ν_k^(q) under p(ν_k | γ_n^(q), z^(q-1), β^(q-1), y)
    end for
    sample z^(q) under p(z | ν^(q), γ_n^(q), β^(q-1), y)
    sample β^(q) under p(β | ν^(q), z^(q), γ_n^(q), y)
end for
```

At this point, we can see the advantage of a conjugate prior for γ_n and the ν_k: their posterior distributions remain in the same family as the prior ones. Here, the conditional posterior for γ_n is a Gamma law with parameters a' and b':

$$\begin{cases} \widetilde{a} = a + P/2 \\ \widetilde{b} = b + 1/2||\boldsymbol{y} - \sum \nu_k \boldsymbol{t}_k(\boldsymbol{z})||^2 \end{cases}$$

Secondly, the conditional posterior for each gray level ν_k is Gaussian with mean $\widetilde{\nu}_k$ and precision parameter $\widetilde{\gamma}_k$:

$$\begin{cases} \widetilde{\gamma}_k = \gamma_n N_k + \gamma_0 \\ \widetilde{\nu}_k = \widetilde{\gamma}_k^{-1}(\gamma_n \boldsymbol{t}_k(\boldsymbol{z})^t \boldsymbol{y} + \gamma_0 m_0) \end{cases}$$

where N_k denotes the number of pixels having the label k.

Regarding the label set, its conditional posterior is a discrete one. In order to sample it, the probabilities π_k^p for the pixel p to be in the class k need to be computed. It is:

$$\pi_k^p = \widetilde{\pi}_k^p \Big/ \sum_l \widetilde{\pi}_l^p$$

where $\widetilde{\pi}_k^p$ is a non-normalized probability deduced from (9):

$$\widetilde{\pi}_k^p = \exp\{-\gamma_n (y^p - \nu_k)^2/2\} \exp\{\beta N_k^p\}$$

where y^p denotes the p-th pixel in the observed image and N_k^p the number of neighbours of pixel p having the label k.

Finally, the conditional posterior for β is deduced:

$$p(\beta|\boldsymbol{z}) \propto C(\beta)^{-1} \exp\Big\{ \beta \sum_{p\sim a} \delta(z_p; z_a) \Big\} \mathcal{U}_{[0,B]}(\beta)$$

By taking advantage of the fact that β is scalar with values in the finite interval $[0, B]$ and that the partition function $C(\beta)$ is pre-computed for a fine grid[2] of $[0, B]$, we can easily compute the conditional cumulative density function (cdf) $F(\beta)$ by standard numerical integration techniques. The obtained values are then interpolated to yield an approximation of the cdf denoted $\widetilde{F}(\beta)$. Then, at iteration q, it suffices to inverse the cdf to generate the sample under $p(\beta|\boldsymbol{z})$. More precisely, the procedure is the following:

Sample $u \sim \mathcal{U}_{[0,1]}(u)$; Compute $\beta^{(q)} = \widetilde{F}^{-1}(u)$.

4. RESULTS

The ground truth $\boldsymbol{x}^\star$ is a 256×256 piecewise constant image with $K = 3$ classes, shown in Fig. 2-a and the true labels $\boldsymbol{z}^\star$ are shown in Fig. 2-b. The true gray levels are -50, 50 and 150 and the true precision is $\gamma_n^\star = 0.005$ (see Tab. 1). The observed noisy image $\boldsymbol{y}$ is given in Fig. 2-c.

In different cases, the Gibbs sampler has been run several times from identical and different initializations, and it shows stable qualitative and quantitative behaviours similar to those in Fig. 3. In the present case, the algorithm has been iterated 500 times and it can be seen that the sample distributions are

(a) True image $\boldsymbol{x}^\star$ (a) True label $\boldsymbol{z}^\star$ (b) Observation $\boldsymbol{y}$ (c) Denoised $\hat{\boldsymbol{x}}$ (d) Estimated label $\hat{\boldsymbol{z}}$

Fig. 2: Input image re-patching and segmentation result

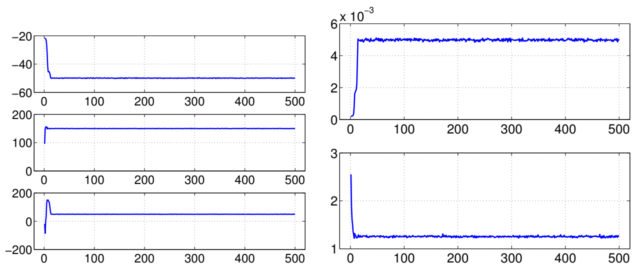

Fig. 3: Simulated chains. Left: for the ν_k and right: for γ_n (top) and β (bottom).

stable after about 50 iterations (burn-in period). These samples have been removed for computing the estimates.

Regarding the labels, the estimates are computed as the empirical marginal posterior maximizers from the samples. Besides, the estimates for the other parameters are given in terms of the empirical posterior mean. As shown in Tab. 1: the estimated values are very close to the true ones, the relative error being less than 1%.

When it comes to β, a hand-tuned value in the range going from 0.7 to 1.5 leads to satisfying segmentations results, the best ones being obtained for a value around 1. The estimated value is slightly higher, but, it should be noted that the corresponding segmentation results are perfect, meaning that 100% of the image's pixels are correctly labelled. This aspect is confirmed by the results displayed in Fig. 2.

Table 1: Parameters' true vs. estimated values

parameter	ν_1	ν_2	ν_3	γ_n	β
true values	-50	50	150	0.005	–
estimates	-49.88	49.86	150.03	0.00503	1.26

5. CONCLUSIONS

In Bayesian image segmentation based on Potts model, the granularity parameter is critical since an inappropriate value might lead to poor segmentation. Its automatic-tuning represents a major stake since for large data sets hand-made tuning might be extremely time-consuming, even impossible. A new method is proposed for its estimation jointly with the other unknowns (labels, gray levels, noise level). It is based on pre-computing the partition function for different image sizes and numbers of classes. An advantage of the approach compared to other ones is that the estimation relies on a precise computation of the likelihood based on the partition function. Furthermore, the computational cost is significantly reduced. The simulations showed promising results.

As a perspective, we plan to tackle deconvolution-segmentation and other reconstruction-segmentation problems [7, 8], still including the Potts parameter estimation. We also plan to pay specific attention to texture models [9, 10].

[2]Practically $B = 3$ and the grid step is 0.01.

A. PARTITION AS AN EMPIRICAL MEAN

Here we describe the pre-computation of the partition function based on well-known relation [11, 12].

Let us note $S(\boldsymbol{z}) = \sum_{p\sim a} \delta(z_p; z_a)$ the number of pair of adjacent pixels with identical label and $\bar{C}(\beta) = \log C(\beta)$ the log-partition. Relation (3) rewrites:

$$C(\beta) = \sum \exp\left\{\beta S(\boldsymbol{z})\right\}$$

where the summation runs over all the configuration of the field $\boldsymbol{z} \in \{1, ...K\}^P$. By derivation, we clearly have:

$$C'(\beta) = \sum S(\boldsymbol{z}) \exp\left\{\beta S(\boldsymbol{z})\right\}$$

then dividing by $C(\beta)$ we have the log-partition derivative:

$$\bar{C}'(\beta) = \sum S(\boldsymbol{z})\, C(\beta)^{-1} \exp\left\{\beta S(\boldsymbol{z})\right\} = \mathrm{E}[\, S(\boldsymbol{Z})\,]$$

and it writes as an expectation that can be approximated by an empirical mean

$$\bar{C}'(\beta) \simeq \frac{1}{N} \sum S(\boldsymbol{z}_n)$$

where the $\boldsymbol{z}_n$ are realizations of the field (given β).

B. REFERENCES

[1] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Tokyo: Springer-Verlag, 2001.

[2] 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 Processing*, vol. 22, no. 6, pp. 2385–2397, 2013.

[3] L. Risser, T. Vincent, P. Ciuciu, and J. Idier, “Robust extrapolation scheme for fast estimation of 3D Ising field partition functions. Application to within-subject fMRI data analysis,” LondonEngland, Sep. 2009.

[4] L. Onsager, “A two-dimensional model with an order-disorder transition,” *Phys. Rev.*, vol. 65, no. 3 & 4, pp. 117–149, Feb. 1944.

[5] J.-F. Giovannelli, “Estimation of the Ising field parameter thanks to the exact partition function,” in *Proc. IEEE ICIP*, Hong-Kong, Sep. 2010, pp. 1441–1444.

[6] ——, “Estimation of the Ising field parameter from incomplete and noisy data,” in *Proc. IEEE ICIP*, Brussels, Belgium, Sep. 2011, pp. 1893–1896.

[7] O. Féron, B. Duchêne, and A. Mohammad-Djafari, “Microwave imaging of piecewise constant objects in a 2D-TE configuration,” *International Journal of Applied Electromagnetics and Mechanics*, vol. 26, no. 6, pp. 167–174, IOS Press 2007.

[8] H. Ayasso and A. Mohammad-Djafari, “Joint NDT image restoration and segmentation using Gauss-Markov-Potts prior models and variational Bayesian computation,” *IEEE Trans. Image Processing*, vol. 19, no. 9, pp. 2265–2277, 2010.

[9] C. Vacar, J.-F. Giovannelli, and Y. Berthoumieu, “Langevin and Hessian with Fisher approximation stochastic sampling for parameter estimation of structured covariance,” in *Proc. IEEE ICASSP*, Prague, Czech Republic, May 2011, pp. 3964–3967.

[10] ——, “Bayesian texture and instrument parameter estimation from blurred and noisy images unsing MCMC,” *IEEE Signal Proc. Let.*, vol. 21, no. 6, pp. 707–711, 2014.

[11] D. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[12] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Springer Verlag, Berlin Germany, 2003.