

VARIATIONAL BAYES WITH GAUSS-MARKOV-POTTS PRIOR MODELS FOR JOINT IMAGE RESTORATION AND SEGMENTATION

Hacheme AYASSO and Ali MOHAMMAD-DJAFARI

*Laboratoire des Signaux et Systèmes, UMR 8506 (CNRS-SUPELEC-UPS)
SUPELEC, Plateau de Moulon, 3 rue Joliot Curie, 91192 Gif-sur-Yvette Cedex, France
Firstname.Lastname@lss.supelec.fr*

Keywords: Variational Bayes Approximation, Image Restoration, Bayesian estimation, MCMC.

Abstract: In this paper, we propose a family of non-homogeneous Gauss-Markov fields with Potts region labels model for images to be used in a Bayesian estimation framework, in order to jointly restore and segment images degraded by a known point spread function and additive noise. The joint posterior law of all the unknowns (the unknown image, its segmentation hidden variable and all the hyperparameters) is approximated by a separable probability laws via the variational Bayes technique. This approximation gives the possibility to obtain practically implemented joint restoration and segmentation algorithm. We will present some preliminary results and comparison with a MCMC Gibbs sampling based algorithm

1 INTRODUCTION

A simple direct model of image restoration problem is

$$g(\mathbf{r}) = h(\mathbf{r}) * f(\mathbf{r}) + \varepsilon(\mathbf{r}) \quad (1)$$

where $g(\mathbf{r})$ is the observed image, $h(\mathbf{r})$ is a known point spread function, $f(\mathbf{r})$ is the unknown image, and $\varepsilon(\mathbf{r})$ is the measurement error. A discretized form of this relation is

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \varepsilon \quad (2)$$

where \mathbf{g}, \mathbf{f} , and ε are vectors containing samples of $g(\mathbf{r}), f(\mathbf{r})$, and $\varepsilon(\mathbf{r})$, and \mathbf{H} is a huge matrix whose elements are determined using $h(\mathbf{r})$ samples.

In a Bayesian framework for such an inverse problem (B.R. Hunt, 1977), one start by writing the expression of the posterior law:

$$p(\mathbf{f}|\mathbf{g}, \mathbf{g}; \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}, \theta_1; \mathcal{M}) p(\mathbf{f}|\theta_2; \mathcal{M})}{p(\mathbf{g}|\mathbf{g}; \mathcal{M})} \quad (3)$$

where $p(\mathbf{g}|\mathbf{f}, \theta_1; \mathcal{M})$, called the *likelihood*, is obtained using the forward model (2) and the assigned probability law $p_\varepsilon(\varepsilon)$ of the errors, $p(\mathbf{f}|\theta_2; \mathcal{M})$ is the assigned prior law for the unknown image \mathbf{f} and

$$p(\mathbf{g}|\mathbf{g}; \mathcal{M}) = \int p(\mathbf{g}|\mathbf{f}, \theta_1; \mathcal{M}) p(\mathbf{f}|\theta_2; \mathcal{M}) d\mathbf{f} \quad (4)$$

is the evidence of the model \mathcal{M} with hyperparameters $\theta = (\theta_1, \theta_2)$. Assigning Gaussian priors

$$\begin{aligned} p(\mathbf{g}|\mathbf{f}, \theta_\varepsilon; \mathcal{M}) &= \mathcal{N}(\mathbf{H}\mathbf{f}, (1/\theta_\varepsilon)\mathbf{I}), \\ p(\mathbf{f}|\theta_f; \mathcal{M}) &= \mathcal{N}(\mathbf{0}, \Sigma_f) \\ \text{with } \Sigma_f &= (1/\theta_f)(\mathbf{D}'\mathbf{D})^{-1} \end{aligned} \quad (5)$$

It is easy to show that the post law is also a Gaussian

$$\begin{aligned} p(\mathbf{f}|\mathbf{g}, \theta_\varepsilon, \theta_f; \mathcal{M}) &\propto p(\mathbf{g}|\mathbf{f}, \theta_\varepsilon; \mathcal{M}) p(\mathbf{f}|\theta_f; \mathcal{M}) \\ &= \mathcal{N}(\hat{\mathbf{f}}, \hat{\Sigma}_f) \end{aligned} \quad (6)$$

with

$$\begin{aligned} \hat{\Sigma}_f &= [\theta_\varepsilon \mathbf{H}'\mathbf{H} + \theta_f \mathbf{D}_f' \mathbf{D}_f]^{-1} \\ &= \frac{1}{\theta_\varepsilon} [\mathbf{H}'\mathbf{H} + \lambda \mathbf{D}'\mathbf{D}]^{-1}, \end{aligned} \quad (7)$$

and

$$\hat{\mathbf{f}} = \theta_\varepsilon \hat{\Sigma}_f \mathbf{H}' \mathbf{g} = [\mathbf{H}'\mathbf{H} + \lambda \mathbf{D}'\mathbf{D}]^{-1} \mathbf{H}' \mathbf{g} \quad (8)$$

which can also be obtained as the solution that minimises:

$$J_1(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda \|\mathbf{D}\mathbf{f}\|^2 \quad (9)$$

where we can see the link with the classical regularization theory (Tikhonov, 1963).

For more general cases, using the MAP estimate:

$$\hat{\mathbf{f}} = \arg \max_{\mathbf{f}} \{p(\mathbf{f}|\mathbf{g}, \mathbf{g}; \mathcal{M})\} = \arg \min_{\mathbf{f}} \{J_1(\mathbf{f})\} \quad (10)$$

We have

$$\begin{aligned} J_1(\mathbf{f}) &= -\ln p(\mathbf{g}|\mathbf{f}, \theta_\epsilon; \mathcal{M}) - \ln p(\mathbf{f}|\theta_2; \mathcal{M}) \\ &= \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\Omega(\mathbf{f}) \end{aligned} \quad (11)$$

where $\lambda = 1/\theta_\epsilon$ and $\Omega(\mathbf{f}) = -\ln p(\mathbf{f}|\theta_2; \mathcal{M})$. Two family of priors could be distinguished: separable:

$$p(\mathbf{f}) \propto \exp \left[-\theta_f \sum_j \phi(f_j) \right] \quad (12)$$

and Markovian

$$p(\mathbf{f}) \propto \exp \left[-\theta_f \sum_j \phi(f_j - f_{j-1}) \right] \quad (13)$$

where different expressions have been used for the potential function $\phi(\cdot)$, (Bouman and Sauer, 1993; Green, 1990; Geman and McClure, 1985) with great success in many applications.

Still, this family of priors can not give a precise model for the unknown image in many applications, due to global image homogeneity assumption. For this reason, we have chosen in paper to use a non-homogenous prior model which takes into account the assumption that the unknown image is composed of finite number of homogenous materials. This implies the introduction of a hidden image $\mathbf{z} = \{z(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ which associates each pixel $f(\mathbf{r})$ with a label (class) $z(\mathbf{r})$, \mathcal{R} represent the whole space of the image surface. All pixels with the same label $z(\mathbf{r}) = k$ share the same properties. Indeed, we use Potts model to represent the dependence between hidden variable pixels, as we will see in the next section. Meanwhile, we propose two models for the unknown image \mathbf{f} , independent mixture of Gaussian and a Gauss-Markov model. However, this choice of prior makes it impossible to get analytical expression for the maximum a posterior (MAP) or posterior mean (PM) estimator. Consequently, we will use the variational Bayes technique to calculate an approximated form of this law.

The rest of this paper is organized as follows. In section 2, we give more details about the proposed prior models. In section 3, we employ these priors using the Bayesian framework to obtain a joint posterior law of the unknowns (image pixels, hidden variable, and the hyperparameters including the region statistical parameters and the noise variance). Then in section 4, we will use the variational Bayes approximation in order to have a tractable approximation of joint posterior law. In section 5, we show an image restoration example. Finally, we conclude this work in section 6.

2 Proposed Gauss-Markov-Potts prior models

As we introduced in the previous section, the main assumption here is the piecewise homogeneity of the restored image. This model corresponds to number of application where the studied image is composed of finite number of materials, as example, muscle and bone or gray-white materials in medical images. Another application, is the non-destructive imaging testing (NDT) in industrials applications, where studied materials are, in general, composed of air-metal or air-metal-composite. This prior model have already been used in several works for several application (Mohammad-Djafari, Humblot and Mohammad-Djafari, 2006; Féron et al., 2005).

In fact, this assumption permits to associate a label (class) $z(\mathbf{r})$ to each pixel of the image \mathbf{f} . The ensemble of this labels \mathbf{z} form a K color image, where K corresponds to the number of materials, and \mathcal{R} represents the entire image pixel area. We call this discrete value variable a hidden field, which represents the segmentation of the image.

Moreover, all pixels $\mathbf{f}_k = \{f(\mathbf{r}), \mathbf{r} \in \mathcal{R}_k\}$ which have the same label k , share the same probabilistic parameters (class means μ_k , and class variances v_k), $\bigcup_k \mathcal{R}_k = \mathcal{R}$. Indeed, these pixels have a spatial structure while we assume here that pixels from different class are a priori independent, which is natural since they belong to different materials. This will be a key assumption when introducing Gauss-Markov prior model of source later in this section.

Using the former assumption, we can give the prior probability law of a pixel knowing the class as a Gaussian (homogeneity inside the same class).

$$p(f(\mathbf{r})|z(\mathbf{r}) = k, m_k, v_k) = \mathcal{N}(m_k, v_k) \quad (14)$$

This will give a Mixture of Gaussians (MoG) model for the pixel $p(f(\mathbf{r}))$. It can be written as follows:

$$p(f(\mathbf{r})) = \sum_k a_k \mathcal{N}(m_k, v_k) \text{ with } a_k = P(z(\mathbf{r}) = k) \quad (15)$$

Another important point is the prior modeling of the spatial interaction between different elements of prior model. This study is concerned with two interactions, pixels of images within the same class $\mathbf{f} = \{f(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$ and elements of hidden variable $\mathbf{z} = \{z(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$. In this paper, we assign Potts model for hidden field \mathbf{z} in order to obtain more homogeneous classes in the image. Meanwhile, we present two models for the image pixels \mathbf{f} ; the first is independent, while the second is Gauss-Markov model. In the following, we give the prior probability of image pixels and hidden field elements for the two models.

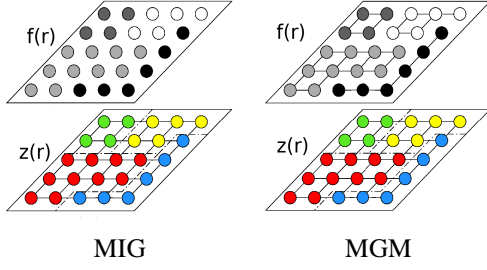


Figure 1: Proposed *a priori* model for the images: the image pixels $f(\mathbf{r})$ are assumed to be classified in K classes, $z(\mathbf{r})$ represents those classes (segmentation). In MIG prior, we assume the image pixels in each class to be independent while in MGM prior, image pixels these are considered dependent. In both cases, the hidden field values follows Potts model in the two cases

Case 1: Mixture of Independent Gaussians (MIG) :

In this case, no prior dependence is assumed for elements of image $\mathbf{f}|z$:

$$\begin{cases} p(f(\mathbf{r})|z(\mathbf{r})=k) &= \mathcal{N}(m_k, v_k), \quad \forall \mathbf{r} \in \mathcal{R} \\ p(\mathbf{f}|\mathbf{z}, m_z(\mathbf{r}), v_z(\mathbf{r})) &= \prod_{\mathbf{r} \in \mathcal{R}} \mathcal{N}(m_z(\mathbf{r}), v_z(\mathbf{r})) \end{cases} \quad (16)$$

with $m_z(\mathbf{r}) = m_k, \forall \mathbf{r} \in \mathcal{R}_k, v_z(\mathbf{r}) = v_k, \forall \mathbf{r} \in \mathcal{R}_k$, and

$$\begin{aligned} p(\mathbf{f}|\mathbf{z}, \mathbf{m}, \mathbf{v}) &= \prod_{\mathbf{r} \in \mathcal{R}} \mathcal{N}(m_z(\mathbf{r}), v_z(\mathbf{r})) \\ &\propto \exp \left[-\frac{1}{2} \sum_{\mathbf{r} \in \mathcal{R}} \frac{(f(\mathbf{r}) - m_z(\mathbf{r}))^2}{v_z(\mathbf{r})} \right] \\ &\propto \exp \left[-\frac{1}{2} \sum_k \sum_{\mathbf{r} \in \mathcal{R}_k} \frac{(f(\mathbf{r}) - m_k)^2}{v_k} \right] \end{aligned} \quad (17)$$

Case 2: Mixture of Gauss-Markovs (MGM) :

We present here a more sophisticated model for interaction between image pixels, where we keep the independence between different classed pixels. The pixels a region are assumed Markovian with the four nearest neighbours.

$$p(f(\mathbf{r})|z(\mathbf{r}), f(\mathbf{r}'), z(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) = \mathcal{N}(\mu_z(\mathbf{r}), v_z(\mathbf{r})) \quad (18)$$

with

$$\begin{cases} \mu_z(\mathbf{r}) &= \frac{1}{|\mathcal{V}(\mathbf{r})|} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \mu_z^*(\mathbf{r}') \\ \mu_z^*(\mathbf{r}') &= \begin{cases} m_z(\mathbf{r}) & \text{if } z(\mathbf{r}') \neq z(\mathbf{r}) \\ f(\mathbf{r}') & \text{if } z(\mathbf{r}') = z(\mathbf{r}) \end{cases} \\ v_z(\mathbf{r}) &= v_k \quad \forall \mathbf{r} \in \mathcal{R}_k \end{cases} \quad (19)$$

We may remark that $\mathbf{f}|\mathbf{z}$ is a non homogeneous Gauss-Markov field because the means $\mu_z(\mathbf{r})$ are functions of the pixel position \mathbf{r} .

For both cases, a Potts Markov model will be used to describe the hidden field prior law for both image models:

$$\begin{aligned} p(\mathbf{z}|\gamma) &\propto \exp \left[\sum_{\mathbf{r} \in \mathcal{R}} \Phi(z(\mathbf{r})) \right] \\ &\times \exp \left[+\frac{1}{2} \gamma \sum_{\mathbf{r} \in \mathcal{R}} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(z(\mathbf{r}) - z(\mathbf{r}')) \right] \end{aligned} \quad (20)$$

Where $\Phi(z(\mathbf{r}))$ is the energy of singleton cliques, and γ is Potts constant.

The hyperparameters of the model are class means m_k , variances v_k , and finally singleton clique energy $\alpha_z(\mathbf{r}) = \Phi(z(\mathbf{r}))$.

3 Bayesian joint reconstruction, segmentation and characterization

So far, we have presented two prior models for unknown image based on the assumption that studied object is composed of known number of materials. That led us to the introduction of hidden field which assigns each pixel a label corresponding to its material. Thus, each material can be characterized by statistical properties (m_k, v_k, α_k) . Now in order to estimate the unknown image and its hidden field, we have to use the Bayesian framework to calculate the joint posterior law.

$$p(\mathbf{f}, \mathbf{z}|\theta, \mathbf{g}; \mathcal{M}) = \frac{p(\mathbf{g}|\mathbf{f}, \theta_1) p(\mathbf{f}|\mathbf{z}, \theta_2) p(\mathbf{z}|\theta_3)}{p(\mathbf{g}|\theta)} \quad (21)$$

This demands the knowledge of $p(\mathbf{f}|\mathbf{z}, \theta_2)$, and $p(\mathbf{z}|\theta_3)$ which we have already provided in the previous section, and the model likelihood $p(\mathbf{g}|\mathbf{f}, \theta_1)$ which depends on the error model. Normal choice for it is the zero mean Gaussian with variance $\frac{1}{\theta_\epsilon}$, which gives:

$$p(\mathbf{g}|\mathbf{f}, \theta_\epsilon) = \mathcal{N}(\mathbf{H}\mathbf{f}, \frac{1}{\theta_\epsilon} \mathbf{I}) \quad (22)$$

In fact the previous calculation assumes that we have the hyperparameters values. This is not true in many practical applications. Consequently, these parameters have to be estimated jointly with the unknown image. This is possible using the Bayesian framework. We just need to assign a prior model for each of the hyperparameters and write the joint posterior law

$$\begin{aligned} p(\mathbf{f}, \mathbf{z}, \theta|\mathbf{g}; \mathcal{M}) &\propto p(\mathbf{g}|\mathbf{f}, \theta_1; \mathcal{M}) p(\mathbf{f}|\mathbf{z}, \theta_2; \mathcal{M}) \\ &\times p(\mathbf{z}|\theta_3; \mathcal{M}) p(\theta|\mathcal{M}) \end{aligned} \quad (23)$$

Where θ regroup all the hyperparameters that need to be estimated which are means m_k , variances v_k , energy singleton α_k , and error inverse variance θ_ϵ . While, Potts constant γ is chosen to be fixed due to the difficulty of finding conjugate prior to it. We choose an Inverse Gamma for the model of the error variance θ_ϵ , Gaussian for the means m_k , Inverse Gamma for variances v_k , and finally a Dirichlet for α_k .

$$\begin{cases} p(\theta_k | a_{e0}, b_{e0}) &= \mathcal{G}(a_{e0}, b_{e0}), \forall k \\ p(m_k | m_0, v_0) &= \mathcal{N}(m_0, v_0), \forall k \\ p(v_k^{-1} | a_0, b_0) &= \mathcal{G}(a_0, b_0), \forall k \\ p(\alpha | \alpha_0) &= \mathcal{D}(\alpha_0, \dots, \alpha_0) \end{cases} \quad (24)$$

where $a_{e0}, b_{e0}, m_0, v_0, a_0, b_0$ and α_0 are fixed for a given problem. In fact, the previous choice of conjugate priors is very helpful for the calculation which we are going to perform in the next section.

4 Bayesian computation

In the previous section, we have found the necessary ingredient to obtain the expression of joint posterior law. However, calculating the joint maximum posterior (JMAP)

$$(\hat{f}, \hat{z}, \hat{\theta}) = \arg \max_{(f, z, \theta)} \{p(f, z, \theta | g; \mathcal{M})\} \quad (25)$$

or the Posterior Means (PM):

$$\begin{cases} \hat{f} &= \sum_z \int \int f p(f, z, \theta | g; \mathcal{M}) df d\theta \\ \hat{\theta} &= \sum_z \int \int \theta p(f, z, \theta | g; \mathcal{M}) df d\theta \\ \hat{z} &= \sum_z \int \int z p(f, z, \theta | g; \mathcal{M}) df d\theta \end{cases} \quad (26)$$

can not be obtained in an analytical form. So, we explore here two approaches to solve this problem, which are Monte Carlos technique and variational Bayes approximation.

Numerical exploration and integration via Monte Carlos techniques:

This method aims to solve the previous problem by generating a great number of representing samples of the posterior law and then calculate the desired estimators numerically using these samples. The main difficulty lays in the generation of those samples. Markov Chain Monte Carlos (MCMC) samplers are used generally in this domain and they have a great interest because of the exploration of the whole space of the joint posterior. Though, the major drawback of this non parametric approach is the computational cost, where a great number of iterations is needed to reach the convergence then lots of samples should be generated to obtain a good estimation of the parameters.

Variational or separable approximation techniques :

One of the main difficulties to obtain an analytical estimator is the posterior dependence between the searched parameters. For this reason, we propose, in this kind of methods, a separable form of the joint

posterior law, and then we try to find the closest posterior to the original posterior under this constraint.

The idea of approximating a joint probability law $p(x)$ by a separable law $q(x) = \prod_j q_j(x_j)$ is not new (Ghahramani and Jordan, 1997; Penny and Roberts, 1998; Roberts et al., 1998; Penny and Roberts, 1999). The way to do and the particular choices of parametric families for $q_j(x_j)$ for which the computations can be done easily have been addressed more recently in many data mining and classification problems (Penny and Roberts, 2002; Roberts and Penny, 2002; Penny and Friston, 2003; Choudrey and Roberts, 2003; Penny et al., 2003; Nasios and Bors, 2004; Nasios and Bors, 2006; Friston et al., 2006; Choudrey and Roberts, 2003). However, the use of these techniques for Bayesian computation for the inverse problems in general and in image restoration in particular, using this class of prior model, is the originality of this paper.

To give a synthetic presentation of the approach we consider the problem of approximating a joint pdf $p(x | \mathcal{M})$ by a separable pdf $q(x) = \prod_j q_j(x_j)$. The first step to do this approximation is to choose a criterion. A natural criterion is the Kullback-Leibler divergence:

$$\begin{aligned} \text{KL}(q : p) &= \int q(x) \ln \frac{q(x)}{p(x | \mathcal{M})} dx \\ &= -H(q) - \langle \ln p(x | \mathcal{M}) \rangle_{q(x)} \\ &= -\sum_j H(q_j) - \langle \ln p(x | \mathcal{M}) \rangle_{q(x)} \end{aligned} \quad (27)$$

So, the main mathematical problem to study is finding $\hat{q}(x)$ which minimizes $\text{KL}(q : p)$.

Using property of the exponential family, this functional optimization problem can be solved as follows

$$q_j(x_j) = \frac{1}{C_j} \exp \left[-\langle \ln p(x | \mathcal{M}) \rangle_{q_{-j}} \right] \quad (28)$$

where $q_{-j} = \prod_{i \neq j} q_i(x_i)$ and C_j are the normalizing factors.

However, we may note that, first the expression of $q_j(x_j)$ depends on the expressions of $q_i(x_i), i \neq j$. Thus the computation can only be done in an iterative way. The second point is that to be able to compute these solutions we must be able to compute $\langle \ln p(x | \mathcal{M}) \rangle_{q_{-j}}$. The only family for which these computations can be done in an easy way is the conjugate exponential families. And here we see the importance of our choice of priors in the previous section;

In fact there is no rule for choosing the appropriate separation; nevertheless, this choice must conserve the strong dependence between variables and break the weak ones, keeping in mind the computation complexity of posterior law. In this work, we propose a

strongly separated posterior, where only dependence between image pixels and hidden fields is conserved. Notably, we can obtain the posterior law more easily.

$$q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) = \prod_{\mathbf{r}} [q(\mathbf{f}(\mathbf{r})|\mathbf{z}(\mathbf{r}))] \prod_{\mathbf{r}} [q(\mathbf{z}(\mathbf{r}))] \prod_l q(\boldsymbol{\theta}_l) \quad (29)$$

Applying the approximated posterior expression (eq.28) on $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\mathbf{g}; \mathcal{M})$, we see that optimal solution for $q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})$ have the following form

$$\left\{ \begin{array}{ll} q(\mathbf{f}|\mathbf{z}) &= \prod_{\mathbf{r}} \mathcal{N}(\tilde{\mu}_z(\mathbf{r}), \tilde{v}_z(\mathbf{r})) \\ p(\mathbf{z}) = \prod_{\mathbf{r}} \hat{\alpha}_k &= \prod_{\mathbf{r}} p(\mathbf{z}(\mathbf{r})|\tilde{z}(\mathbf{r}'), \mathbf{r}' \in \mathcal{V}(\mathbf{r})) \\ p(\mathbf{z}(\mathbf{r})|\tilde{z}(\mathbf{r}')) &\propto \tilde{c} \tilde{d}_1 \tilde{d}_2(\mathbf{r}) \tilde{e}(\mathbf{r}) \\ q(\boldsymbol{\theta}_\varepsilon|\tilde{\alpha}_e, \tilde{\beta}_e) &= \mathcal{G}(\tilde{\alpha}_e, \tilde{\beta}_e) \\ q(m_k|\tilde{m}_k, \tilde{v}_k) &= \mathcal{N}(\tilde{m}_k, \tilde{v}_k), \forall k \\ q(v_k^{-1}|\tilde{a}_k, \tilde{b}_k) &= \mathcal{G}(\tilde{a}_k, \tilde{b}_k), \forall k \\ q(\boldsymbol{\alpha}) &\propto \mathcal{D}(\tilde{\alpha}_1, \dots, \tilde{\alpha}_K) \end{array} \right. \quad (30)$$

where all tilted quantities are defined later.

For the approached posteriors of the unknown image we distinguish two different results according to the used prior model.

Case 1: MIG :

$$\begin{aligned} \tilde{\mu}_z(\mathbf{r}) &= \tilde{v}_z(\mathbf{r}) \frac{\tilde{m}_k}{\tilde{v}_k} \\ &+ \tilde{v}_z(\mathbf{r}) \boldsymbol{\theta}_\varepsilon \sum_{\mathbf{s}} H(\mathbf{s}, \mathbf{r}) (g(\mathbf{s}) - \hat{g}_{-\mathbf{r}}(\mathbf{s})) \\ \tilde{v}_z(\mathbf{r}) &= (\tilde{v}_k^{-1} + \boldsymbol{\theta}_\varepsilon \sum_{\mathbf{s}} H^2(\mathbf{s}, \mathbf{r}))^{-1} \\ \hat{g}_{-\mathbf{r}}(\mathbf{s}) &= \sum_{\mathbf{t} \neq \mathbf{r}} H(\mathbf{s}, \mathbf{t}) \tilde{f}(\mathbf{t}) \\ \tilde{v}_k &= \langle v_k \rangle = (\tilde{a}_k \tilde{b}_k)^{-1} \end{aligned} \quad (31)$$

Case 2: MGM :

$$\begin{aligned} \tilde{\mu}_z(\mathbf{r}) &= \tilde{v}_z(\mathbf{r}) \frac{\tilde{\mu}_z^*(\mathbf{r})}{\tilde{v}_k} \\ &+ \tilde{v}_z(\mathbf{r}) \boldsymbol{\theta}_\varepsilon \sum_{\mathbf{s}} H(\mathbf{s}, \mathbf{r}) (g(\mathbf{s}) - \hat{g}_{-\mathbf{r}}(\mathbf{s})) \\ \tilde{\mu}_z^*(\mathbf{r}) &= \frac{1}{|\mathcal{V}(\mathbf{r})|} \sum_{\mathbf{r}' \in \mathcal{V}(\mathbf{r})} \delta(\mathbf{z}(\mathbf{r}') - \tilde{z}(\mathbf{r})) \tilde{f}(\mathbf{r}') \\ &+ (1 - \delta(\mathbf{z}(\mathbf{r}') - \tilde{z}(\mathbf{r})) \tilde{m}_z(\mathbf{r}) \\ \tilde{v}_z(\mathbf{r}) &= (\tilde{v}_k^{-1} + \boldsymbol{\theta}_\varepsilon \sum_{\mathbf{s}} H^2(\mathbf{s}, \mathbf{r}))^{-1} \\ \hat{g}_{-\mathbf{r}}(\mathbf{s}) &= \sum_{\mathbf{t} \neq \mathbf{r}} H(\mathbf{s}, \mathbf{t}) \tilde{f}(\mathbf{t}) \\ \tilde{v}_k &= \langle v_k \rangle = (\tilde{a}_k \tilde{b}_k)^{-1} \end{aligned} \quad (32)$$

Meanwhile, the posterior law of the hidden field remains the same, and given by the following relation

$$\begin{aligned} \tilde{c} &= \exp[\Psi(\tilde{\alpha}_k) - \Psi(\sum_{\mathbf{z}} \tilde{\alpha}_{\mathbf{z}})] \\ \tilde{d}_1 &= \exp\left[\frac{1}{2} (\Psi(\tilde{b}_k) + \ln(\tilde{a}_k))\right] \\ \tilde{d}_2(\mathbf{r}) &= \exp\left[\frac{1}{2} \left(\frac{\tilde{\mu}_k^2(\mathbf{r})}{\tilde{v}_k(\mathbf{r})} - \frac{\tilde{m}_k^2}{\tilde{v}_k} - \ln(\tilde{v}_k(\mathbf{r}))\right)\right] \\ \tilde{e}(\mathbf{r}) &= \exp\left[-\frac{1}{2} \gamma \sum_{\mathbf{r}'} \Phi_m(\mathbf{r}, \mathbf{r}')\right] \end{aligned} \quad (33)$$

where $\Phi_m(\cdot, \cdot)$ class projection function see (Ayasso and Mohammad-Djafari, 2007).

Finally the hyperparameters posterior variables are,

$$\begin{aligned} \tilde{\alpha}_e &= [a_{e0}^{-1} + \frac{1}{2} \sum_{\mathbf{r}} \mathcal{E}((g(\mathbf{r}) - \tilde{g}(\mathbf{r}))^2)]^{-1} \\ \tilde{\beta}_e &= b_{e0} + \sum_{\mathbf{r}} \frac{1}{2} \\ \tilde{m}_k &= \tilde{v}_k \left(\frac{m_0}{v_0} + \tilde{v}_k \sum_{\mathbf{r}} \hat{\alpha}_k(\mathbf{r}) \tilde{\mu}_k(\mathbf{r}) \right) \\ \tilde{v}_k &= (v_0^{-1} + \sum_{\mathbf{r}} \hat{\alpha}_k(\mathbf{r}))^{-1} \\ \tilde{a}_k &= [a_0^{-1} + \frac{1}{2} \sum_{\mathbf{r}} \hat{\alpha}_k(\mathbf{r}) \tilde{u}_k(\mathbf{r})]^{-1} \\ \tilde{u}_k &= \tilde{\mu}_k^2(\mathbf{r}) + \tilde{v}_k(\mathbf{r}) + m_k^2 - 2\tilde{m}_k \tilde{\mu}_k(\mathbf{r}) \\ \tilde{\alpha}_k &= \alpha_0 + \sum_{\mathbf{r}} \hat{\alpha}_k(\mathbf{r}) \end{aligned} \quad (34)$$

Several observations can be made on the these results. The most important is that the problem of probability law optimization turned into simple parametric computation, which reduces significantly the computational burden. Indeed, although of the strong chosen separation, posterior mean value dependence between image pixels and hidden field elements is present in the equations, which justifies the use of spatially dependent prior model with this independent approximated posterior. On the other hand, the obtained values are mutually dependent. One difficulty still remains which is an appropriate choice of an stopping criterion A subject on which we are working.

5 Numerical experiment results

In this section, we apply the proposed methods on a synthesized restoration problem. The original image, which is composed of two classes, is filtered by a square shape point spread function(PSF). Then white Gaussian noise is added in order to obtain the distorted image. We have used variational approximation method with Gauss-Markov-Potts priors to jointly restore and segment it. In comparison with MCMC method, we note that image quality is approximately the same. However, the computation time is incredibly less in the proposed method in comparison with MCMC.

6 conclusion

A variational Bayes approximation is proposed in this paper for image restoration. We have introduced a hidden variable to give a more accurate prior model of the unknown image. Two priors, independent Gaussian and Gauss-Markov models were studied with Potts prior on the hidden field. This method was applied to a simple restoration problem, where it gave promising results.

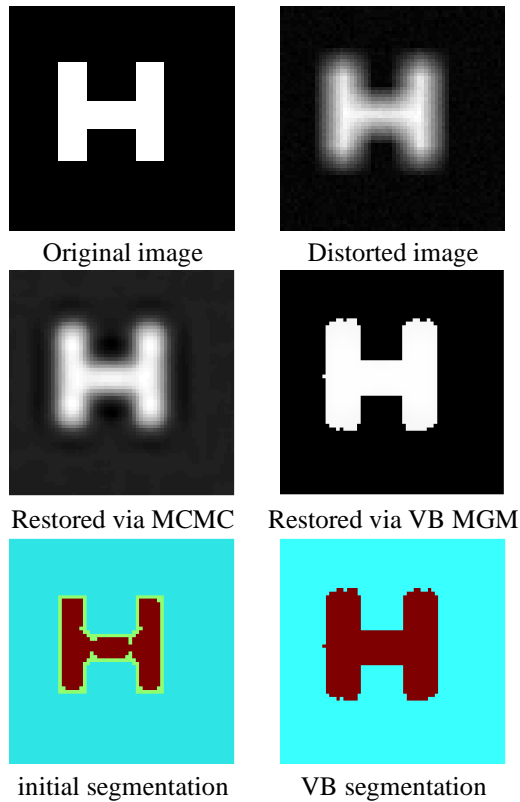


Figure 2: The proposed method is tested by a synthesized 2 classes distorted image. Number of classes in each method were set to 3. There is no great difference in quality between the two methods, VB MGM, and MCMC. Meanwhile, variational bayes algorithm is noticeably faster than MCMC sampler used here

Still, a number of the aspects regarding this method have to be studied, including the convergence conditions, choice of separation and the estimation of Potts parameter.

REFERENCES

- Ayasso, H. and Mohammad-Djafari, A. (2007). Approche bayésienne variationnelle pour les problèmes inverses. application en tomographie microonde. Technical report, Rapport de stage Master ATS, Univ Paris Sud, L2S, SUPELEC.
- Bouman, C. and Sauer, K. (1993). A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Transaction on image processing*, pages 296–310.
- Choudrey, R. and Roberts, S. (2003). Variational Mixture of Bayesian Independent Component Analysers. *Neural Computation*, 15(1).
- Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2006). Variational free energy and the laplace approximation. *Neuroimage*, (2006.08.035). Available Online.
- Geman, S. and McClure, D. (1985). Bayesian image analysis: Application to single photon emission computed tomography. *American Statistical Association*, pages 12–18.
- Ghahramani, Z. and Jordan, M. (1997). Factorial Hidden Markov Models. *Machine Learning*, (29):245–273.
- Green, P. (1990). Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transaction on Medical Imaging*, pages 84–93.
- Nasios, N. and Bors, A. (2004). A variational approach for bayesian blind image deconvolution. *IEEE Transactions on Signal Processing*, 52(8):2222–2233.
- Nasios, N. and Bors, A. (2006). Variational learning for gaussian mixture models. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 36(4):849–862.
- Penny, W. and Friston, K. (2003). Mixtures of general linear models for functional neuroimaging. *IEEE Transactions on Medical Imaging*, 22(4):504–514.
- Penny, W., Kiebel, S., and Friston, K. (2003). Variational bayesian inference for fmri time series. *NeuroImage*, 19(3):727–741.
- Penny, W. and Roberts, S. (1998). Bayesian neural networks for classification: how useful is the evidence framework ? *Neural Networks*, 12:877–892.
- Penny, W. and Roberts, S. (1999). Dynamic models for nonstationary signal segmentation. *Computers and Biomedical Research*, 32(6):483–502.
- Penny, W. and Roberts, S. (2002). Bayesian multivariate autoregressive models with structured priors. *IEEE Proceedings on Vision, Image and Signal Processing*, 149(1):33–41.
- Roberts, S., Husmeier, D., Penny, W., and Rezek, I. (1998). Bayesian approaches to gaussian mixture modelling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1133–1142.
- Roberts, S. and Penny, W. (2002). Variational bayes for generalised autoregressive models. *IEEE Transactions on Signal Processing*, 50(9):2245–2257.
- Tikhonov (1963). Solution of incorrectly formulated problems and the regularization method. *Sov. Math.*, pages 1035–8.