HOGMep: VARIATIONAL BAYES AND HIGHER-ORDER GRAPHICAL MODELS APPLIED TO JOINT IMAGE RECOVERY AND SEGMENTATION

Aurélie Pirayre[1]*, Yuling Zheng*[3]*, Laurent Duval*[1]*, and Jean-Christophe Pesquet*[2]

[1] IFP Energies nouvelles, Rueil-Malmaison, France ; [2] CentraleSupélec, France ; [3] IBM Research, China

ABSTRACT

Variational Bayesian approaches have been successfully applied to image segmentation. They usually rely on a Potts model for the hidden label variables and a Gaussian assumption on pixel intensities within a given class. Such models may however be limited, especially in the case of multicomponent images. We overcome this limitation with HOGMep, a Bayesian formulation based on a higher-order graphical model (HOGM) on labels and a Multivariate Exponential Power (MEP) prior for intensities in a class. Then, we develop an efficient statistical estimation method to solve the associated problem. Its flexibility accommodates to a broad range of applications, demonstrated on multicomponent image segmentation and restoration.

Index Terms— variational Bayes, higher-order graphical models, multicomponent images, segmentation, restoration

1. INTRODUCTION

Variational Bayesian Approximation (VBA) has been widely used in various applications such as in graphical model learning [1], image processing [2, 3, 4, 5], source separation [6] or super-resolution [7]. Its popularity is mainly due to its ability to generate efficient approximations of the posterior distribution of the variables to be estimated. In image segmentation, VBA-based approaches traditionally uses a Potts (resp. Gaussian) model for hidden label variables (resp. pixel intensities corresponding to a given label) [8, 9, 10].

We propose a more generic Bayesian model, combining two main priors on the sought data. The first one concerns the latent label variables for which an HOGM is employed for clustering or classification [11, 12]. The second one adopts a MEP distribution [13, 14] for pixel intensities in a given class. To the best of our knowledge, these priors have not been jointly used for image recovery and segmentation tasks. The novel model is suitable for applications dealing with multicomponent or multivariate images (e.g. multi/hyperspectral data) [15, 16, 17]. One of the advantages of the proposed approach is that it allows us to estimate the associated hyperparameters by defining suitable priors. While alternatives could solve segmentation [18, 19, 20] or reconstruction [21, 22] problems, VBA benefits from low computational cost while providing high quality results.

Section 2 introduces the HOGMep Bayesian formulation for joint restoration and segmentation. The VBA methodology solving the related problem is described in Section 3. The proposed joint recovery and segmentation approach is favorably evaluated and compared within a deconvolution context in Section 4.

2. HOGMep BAYESIAN FORMULATION

2.1. Inverse problem model

We concentrate on the standard inverse problem consisting of recovering an unknown signal $\mathbf{x}$ from a degraded one $\mathbf{y}$. We consider a linear model with additive noise formulated as $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where, for $(M, N, B) \in (\mathbb{N}^*)^3$, $\mathbf{y} \in \mathbb{R}^M$ is the observed data, $\mathbf{x} \in \mathbb{R}^{NB}$ corresponds to the unknown signal to be recovered, $\mathbf{H} \in \mathbb{R}^{M \times NB}$ represents a linear degradation operator and $\mathbf{n}$ is a noise (statistically independent of $\mathbf{x}$). We are interested in B-component images where $\mathbf{x} = [\mathbf{x}_1^\top, \ldots, \mathbf{x}_N^\top]^\top$ and, for every pixel $i \in \{1, \ldots, N\}$, $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,B})^\top$. Assuming a zero-mean white Gaussian noise with inverse variance (or precision) γ, the likelihood $p(\mathbf{y}|\mathbf{x}, \gamma)$ is given by the normal distribution $\mathcal{N}(\mathbf{H}\mathbf{x}, \gamma^{-1}\mathbf{I})$. In conjunction with the recovery task, our objective is to perform a classification of the components of $\mathbf{x}$. For this purpose, L being the number of expected classes, the label field is encoded by a vector of hidden variables $\mathbf{z} \in \{1, \ldots, L\}^N$.

2.2. Prior model on the unknown data and hyperpriors

HOGM for label variables [11, 12] deals with cliques of arbitrary size instead of being limited to pairwise interactions between variables. The chosen prior is:

$$p(\mathbf{z}) \propto \exp\left(\sum_{s=1}^{S} \sum_{(i_1,\ldots,i_s)\in\mathcal{N}_s} V_s(z_{i_1}, \ldots, z_{i_s})\right), \qquad (1)$$

where S is the size of the maximal clique and, for every $s \in \{1, \ldots, S\}$, the function V_s is a potential function of order s, and $\mathcal{N}_s$ is the set of cliques of size s. The model contains a prior weighting parameter $\boldsymbol{\lambda}$, not explicitly written in (1).

In addition to a prior on hidden label variables, we assume that the conditional distribution of $\mathbf{x}$ given label variables $\mathbf{z}$ is a Multivariate Exponential Power (MEP) distribution denoted by $\mathcal{M}$. Such a prior is well suited to multicomponent images [23]. We denote by $z_i \in \{1, \ldots, L\}$ the label of the class to which the pixel indexed by i belongs, and by $\mathbf{z} = (z_1, \ldots, z_N)^\top$ the label vector for all pixels of the image. Furthermore, for every class labeled by $l \in \{1, \ldots, L\}$, $\mathbf{m}_l$, $\boldsymbol{\Omega}_l$ and β_l denote the parameters of the MEP distribution associated with label value l. By assuming that $\beta_l \in (0, 1]$, the MEP distribution can be expressed as a Gaussian Scale Mixture (GSM) [24]:

$$\mathcal{M}(\mathbf{x}_i; \mathbf{m}_l, \boldsymbol{\Omega}_l, \beta_l) = \int_{\mathbb{R}^+} \mathcal{N}(\mathbf{x}_i; \mathbf{m}_l, u_i^{-1}\boldsymbol{\Omega}_l^{-1}) p(u_i|\beta_l) \mathrm{d}u_i,$$

where $p(u_i|\beta_l)$ denotes the probability density function (pdf) of u_i with shape parameter β_l and $\mathbf{u} = (u_1, \ldots, u_N)^\top$ gathers all the latent variables. When $\beta_l < 1$, this pdf can be expressed as a function

of a positive alpha-stable distribution. When $\beta_l = 1$, it degenerates into a Dirac distribution [24]. As a result, we have

$$p(\mathbf{x}|\mathbf{z}, \mathbf{m}, \mathbf{\Omega}, \boldsymbol{\beta}) = \prod_{i=1}^{N} p(\mathbf{x}_i|z_i, \mathbf{m}, \mathbf{\Omega}, \boldsymbol{\beta}),$$

where $p(\mathbf{x}_i|z_i = l, \mathbf{m}, \mathbf{\Omega}, \boldsymbol{\beta}) = \mathcal{M}(\mathbf{x}_i; \mathbf{m}_l, \mathbf{\Omega}_l, \beta_l)$. Our Bayesian model involves four types of hyper-parameters: inverse noise variance γ, mean variables $(\mathbf{m}_l)_{1\leq l\leq L}$, inverse covariance matrices $(\mathbf{\Omega}_l)_{1\leq l\leq L}$ and shape parameters $(\beta_l)_{1\leq l\leq L}$. In this work, we assume that, for all the classes $l \in \{1, \ldots, L\}$, the shape parameters of the MEP distribution are identical, i.e. $\beta_1 = \cdots = \beta_L = \beta$. This single parameter is fixed in advance, according to prior knowledge. If $\mathcal{G}$ and $\mathcal{W}$ denote Gamma and Wishart distributions, to jointly estimate all the remaining parameters, we define a hyperprior on each of them:

$$p(\gamma) = \mathcal{G}(\bar{a}, \bar{b}), \quad p(\mathbf{m}_l) = \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\mathbf{\Lambda}}), \quad p(\mathbf{\Omega}_l) = \mathcal{W}(\bar{\mathbf{\Gamma}}, \bar{\nu}),$$

Figure 1 summarizes the dependency relationships between variables involved in this hierarchical graphical model.

2.3. Joint posterior distribution

Using Bayes' rule, the joint posterior distribution reads

$$\begin{aligned} &p(\mathbf{x}, \mathbf{u}, \mathbf{z}, \gamma, \mathbf{m}, \mathbf{\Omega}|\mathbf{y}, \beta) \\ &\propto \; p(\mathbf{y}|\mathbf{x}, \gamma) \prod_{i=1}^{N} \Big(p(\mathbf{x}_i|z_i, u_i, \mathbf{m}, \mathbf{\Omega}) p(u_i|\beta) \Big) p(\mathbf{z}) p(\gamma) \\ &\quad \times \prod_{l=1}^{L} p(\mathbf{m}_l) p(\mathbf{\Omega}_l) \end{aligned} \tag{2}$$

Its intricate form is due to the dependence between the unknown variables. VBA can circumvent this issue [8, 9, 10], and is indeed shown in the following section to provide an elegant solution to the Bayesian inference.

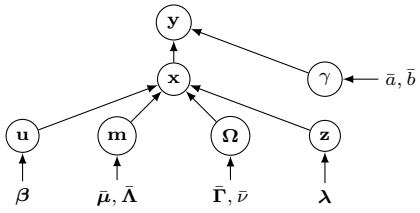


Fig. 1: Dependency relationships between variables.

3. VARIATIONAL BAYESIAN APPROACH

Let $\Theta = (\Theta_j)_{1\leq j\leq J}$ be a vector containing all the variables $(\mathbf{x}, \mathbf{u}, \mathbf{z}, \gamma, \mathbf{m}, \mathbf{\Omega})$ to be estimated. As previously mentioned, we want to estimate the true posterior distribution $p(\Theta|\mathbf{y})$. The core idea of VBA is to approximate $p(\Theta|\mathbf{y})$ by a separable distribution, denoted by $q(\Theta)$, such that $q(\Theta) = \prod_{j=1}^{J} q(\Theta_j)$. The minimization of the Kullback-Leibler ($\mathcal{KL}$) divergence between $q(\Theta)$ and $p(\Theta|\mathbf{y})$ allows us to determine the optimal separable approximation:

$$\mathcal{KL}(q(\Theta)||p(\Theta|\mathbf{y})) = \int q(\Theta) \ln \frac{q(\Theta)}{p(\Theta|\mathbf{y})} \mathrm{d}\Theta. \tag{3}$$

The optimal approximate distribution can be computed [25]:

$$(\forall j \in \{1, \ldots, J\}) \quad q(\Theta_j) \propto \exp\Big(\langle \ln p(\mathbf{y}, \Theta) \rangle_{q_{-\Theta_j}} \Big), \tag{4}$$

where $q_{-\Theta_j} = \prod_{i\neq j} q(\Theta_i)$, and $\langle . \rangle_q$ denotes the expectation value with respect to a probability distribution q. Implicit relations between pdfs $\big(q(\Theta_j)\big)_{1\leq j\leq J}$ generally prevent analytical expressions for $q(\Theta)$. These distributions are determined in an iterative way, by updating one of the separable components $\big(q(\Theta_j)\big)_{1\leq j\leq J}$ while fixing the others. In this work, the approximation takes the form:

$$q(\Theta) = q(\gamma) \prod_{i=1}^{N} \Big(q(\mathbf{x}_i, z_i) q(u_i) \Big) \prod_{l=1}^{L} \Big(q(\mathbf{m}_l) q(\mathbf{\Omega}_l) \Big), \tag{5}$$

with $q(\mathbf{x}_i, z_i) = q(\mathbf{x}_i|z_i) q(z_i)$. Hence, using (4), for every $i \in \{1, \ldots, N\}$ and $l \in \{1, \ldots, L\}$, the optimal solutions for $q(\mathbf{x}_i|z_i)$, $q(z_i)$, $q(\mathbf{m}_l)$, $q(\mathbf{\Omega}_l)$ and $q(\gamma)$ are such that

$$\begin{aligned} &q(\mathbf{x}_i|z_i = l) = \mathcal{N}(\boldsymbol{\eta}_{i,l}, \mathbf{\Xi}_{i,l}), \; q(z_i = l) = \pi_{i,l}, \\ &q(\mathbf{m}_l) = \mathcal{N}(\boldsymbol{\mu}_l, \mathbf{\Lambda}_l), \quad q(\mathbf{\Omega}_l) = \mathcal{W}(\mathbf{\Gamma}_l, \nu_l), \quad q(\gamma) = \mathcal{G}(a, b). \end{aligned} \tag{6}$$

Except for $q(u_i)$, the optimization of these distributions can be performed iteratively, since they belong to known parametrized distribution families. The MEP prior distribution can be expressed as a GSM. Hence, the mean value of u_i can be determined without an analytical expression for $q(u_i)$ [26, 27]. This allows us to derive an approximate distribution of $\mathbf{x}_i$, which depends on the mean value of u_i.

In the following, assuming that $k \in \mathbb{N}$ designates the iteration number, we describe how to update these distributions by deriving closed form expressions for their parameters.

3.1. Determination of model pdf $q(\mathbf{x}_i, z_i)$

As mentioned in (6), for $l \in \{1, \ldots, L\}$, $q^{k+1}(\mathbf{x}_i|z_i = l)$ is a Gaussian distribution. Using (4), its covariance matrix $\mathbf{\Xi}_{i,l}$ and mean $\boldsymbol{\eta}_{i,l}$ are given, at the iteration $k+1$, by

$$\mathbf{\Xi}_{i,l}^{k+1} = \Big(\widehat{\gamma}^k \mathbf{H}_i^\top \mathbf{H}_i + \widehat{u}_i^k \widehat{\mathbf{\Omega}}_l^k \Big)^{-1} \tag{7}$$

$$\begin{aligned} \boldsymbol{\eta}_{i,l}^{k+1} = \mathbf{\Xi}_{i,l}^{k+1} \Big(&\widehat{\gamma}^k \mathbf{H}_i^\top (\mathbf{y} - \sum_{j<i} \mathbf{H}_j \widehat{\mathbf{x}}_j^{k+1} - \sum_{j>i} \mathbf{H}_j \widehat{\mathbf{x}}_j^k) \\ &+ \widehat{u}_i^k \widehat{\mathbf{\Omega}}_l^k \boldsymbol{\mu}_l^k \Big). \end{aligned} \tag{8}$$

$\mathbf{H}_i \in \mathbb{R}^{M\times B}$ is the columnwise decomposition of $\mathbf{H}$. Furthermore, for an arbitrary variable $\mathbf{a}$, $\widehat{\mathbf{a}}^k$ is its expectation at iteration k with respect to the pdf $q^k(\mathbf{a})$. We can then derive the expressions of the probabilities $q^{k+1}(z_i = l)$:

$$\begin{aligned} &q^{k+1}(z_i = l) = \pi_{i,l}^{k+1} \\ &\propto \Big(|\mathbf{\Xi}_{i,l}^{k+1}| |\mathbf{\Gamma}_l^k| \Big)^{1/2} \exp\Bigg(\frac{1}{2} (\boldsymbol{\eta}_{\boldsymbol{i,l}}^{\boldsymbol{k+1}})^\top (\mathbf{\Xi}_{i,l}^{k+1})^{-1} \boldsymbol{\eta}_{i,l}^{k+1} \\ &+ \frac{1}{2} \sum_{b=1}^{B} \psi\left(\frac{\nu_l^k + 1 - b}{2} \right) \\ &- \frac{1}{2} \widehat{u}_i^k \mathrm{tr}[(\mathbf{\Lambda}_l^k + \boldsymbol{\mu}_l^k (\boldsymbol{\mu}_l^k)^\top) \widehat{\mathbf{\Omega}}_l^k] + \tilde{V}_l^k \Bigg), \end{aligned} \tag{9}$$

where ψ is the digamma function and

$$\widehat{V}_l^k = V_1(l) + \sum_{s=1}^{S-1} \Big\langle \sum V_{s+1}(l, z_{i_1}, \ldots, z_{i_s}) \Big\rangle_{\prod_{j \neq i} q^k(z_j)}.$$

For every $j \in \{1, \ldots, N\}$, we have $\widehat{\mathbf{x}}_j^{k+1} = \sum_{l=1}^{L} \pi_{j,l}^{k+1} \boldsymbol{\eta}_{j,l}^{k+1}$.

Instead of computing the $q(u_i)$ distribution, we focus on the expectation of u_i which can be expressed as follows, thanks to the integral form of GSM [26]:

$$\widehat{u}_i^{k+1} = \beta \left(\sum_{l=1}^{L} \pi_{i,l}^{k+1} \mathrm{tr}[\mathbf{A}_k \widehat{\boldsymbol{\Omega}}_l^k] \right)^{\beta - 1}, \tag{10}$$

where $\mathbf{A}_k = (\boldsymbol{\eta}_{i,l}^{k+1} - \boldsymbol{\mu}_l^k)(\boldsymbol{\eta}_{i,l}^{k+1} - \boldsymbol{\mu}_l^k)^\top + \boldsymbol{\Xi}_{i,l}^{k+1} + \boldsymbol{\Lambda}_l^k$.

3.2. Finding of hyperprior pdfs $q(\mathbf{m}_l)$, $q(\boldsymbol{\Omega}_l)$ and $q(\gamma)$

As shown by (6), $q^{k+1}(\mathbf{m}_l)$ is a Gaussian distribution with mean $\boldsymbol{\mu}_l^{k+1}$ and covariance matrix $\boldsymbol{\Lambda}_l^{k+1}$ expressed as

$$\boldsymbol{\Lambda}_l^{k+1} = \left(\bar{\boldsymbol{\Lambda}}^{-1} + \widehat{\boldsymbol{\Omega}}_l^k \sum_{i=1}^{N} \pi_{i,l}^{k+1} \widehat{u}_i^{k+1} \right)^{-1}, \tag{11}$$

$$\boldsymbol{\mu}_l^{k+1} = \boldsymbol{\Lambda}_l^{k+1} \left(\bar{\boldsymbol{\Lambda}}^{-1} \bar{\boldsymbol{\mu}} + \widehat{\boldsymbol{\Omega}}_l^k \sum_{i=1}^{N} \pi_{i,l}^{k+1} \widehat{u}_i^{k+1} \boldsymbol{\eta}_{i,l}^{k+1} \right). \tag{12}$$

Then, the mean value of $q^{k+1}(\mathbf{m}_l)$ is $\widehat{\mathbf{m}}_l^{k+1} = \boldsymbol{\mu}_l^{k+1}$.

The Wishart distribution $q^{k+1}(\boldsymbol{\Omega}_l)$ is parametrized by

$$\nu_l^{k+1} = \bar{\nu} + \sum_{i=1}^{N} \pi_{i,l}^{k+1}, \tag{13}$$

$$\boldsymbol{\Gamma}_l^{k+1} = \left(\sum_{i=1}^{N} \pi_{i,l}^{k+1} \widehat{u}_i^{k+1} \tilde{\mathbf{A}}_k + \bar{\boldsymbol{\Gamma}}^{-1} \right)^{-1}, \tag{14}$$

where $\tilde{\mathbf{A}}_k = (\boldsymbol{\eta}_{i,l}^{k+1} - \boldsymbol{\mu}_l^{k+1})(\boldsymbol{\eta}_{i,l}^{k+1} - \boldsymbol{\mu}_l^{k+1})^\top + \boldsymbol{\Xi}_{i,l}^{k+1} + \boldsymbol{\Lambda}_l^{k+1}$. The mean value of $q^{k+1}(\boldsymbol{\Omega}_l)$ is

$$\widehat{\boldsymbol{\Omega}}_l^{k+1} = \nu_l^{k+1} \boldsymbol{\Gamma}_l^{k+1}. \tag{15}$$

The $q^{k+1}(\gamma)$ distribution is a Gamma distribution with parameters $a^{k+1} = \bar{a} + \frac{M}{2}$ and

$$b^{k+1} = \bar{b} + \frac{1}{2} \|\mathbf{y} - \mathbf{H}\widehat{\mathbf{x}}^{k+1}\|^2 + \sum_{i=1}^{N} \sum_{l=1}^{L} \pi_{i,l}^{k+1} \mathrm{tr}[\mathbf{H}_i^\top \mathbf{H}_i \boldsymbol{\Xi}_{i,l}^{k+1}]. \tag{16}$$

From standard properties of the Gamma distribution, the expectation of γ at the iteration $k+1$ is equal to

$$\widehat{\gamma}^{k+1} = \frac{a^{k+1}}{b^{k+1}} = \frac{2\bar{a} + M}{2b^{k+1}}. \tag{17}$$

3.3. Resulting algorithm

Altogether, our algorithm can be summed up as follows:

Set initial values: $\boldsymbol{\eta}_{i,l}^0$, $\boldsymbol{\Xi}_{i,l}^0$, $\widehat{u}_i^0$, $\pi_{i,l}^0$, $\boldsymbol{\mu}_l^0$, $\boldsymbol{\Lambda}_l^0$, $\boldsymbol{\Gamma}_l^0$, ν_l^0, b^0, and set $a^k \equiv \bar{a} + \frac{M}{2}$. Compute $\widehat{\boldsymbol{x}}_i^0 = \sum_{l=1}^{L} \pi_{i,l}^0 \eta_{i,l}^0$, $\widehat{\boldsymbol{\Omega}}_l^0 = \nu_l^0 \boldsymbol{\Gamma}_l^0$, and $\widehat{\gamma}^0 = 1/b^0(\bar{a} + M/2)$.
For $k = 0, 1, \ldots$

1. Update parameters $\boldsymbol{\Xi}_{i,l}^{k+1}$ and $\boldsymbol{\eta}_{i,l}^{k+1}$ of $q^{k+1}(\boldsymbol{x}_i | z_i = l)$ using (7) and (8). Compute $\pi_{i,l}^{k+1}$ from (9).
2. Update mean values $\widehat{u}_i^{k+1}$ of $q^{k+1}(u_i)$ using (10).
3. Update parameters $\boldsymbol{\Lambda}_l^{k+1}$ and $\boldsymbol{\mu}_l^{k+1}$ of $q^{k+1}(\mathbf{m}_l)$ using (11) and (12).
4. Update parameters ν_l^{k+1} and $\boldsymbol{\Gamma}_l^{k+1}$ of $q^{k+1}(\boldsymbol{\Omega}_l)$ using (13) and (14). Compute $\widehat{\boldsymbol{\Omega}}_l^{k+1}$ from (15).
5. Update parameter b^{k+1} of $q^{k+1}(\gamma)$ using (16). Compute $\widehat{\gamma}^{k+1}$ from (17).

---

4. SIMULATION RESULTS

The performance of HOGMep is assessed through segmentation and recovery tasks in a deconvolution context. Experiments are performed on both synthetic and benchmark color images. Channels are degraded by distinct blur operators and additive Gaussian noises. A natural comparison is made with VB-MIG [8], in which our MEP prior is restricted to a Gaussian one. Restoration results are also compared to those obtained with the recent state-of-the-art variational approach 3MG [21, 28] and segmentation with the iterated conditional mode (ICM) algorithm [29]. For the 'synth' image with ground truth, results in SNR are given in Table 1a, and displayed in Figure 2 (restored channels) and 3 (color image). HOGMep consistently yields the best objective results, with crisper images.

Channel	Initial	3MG	VB-MIG	HOGMep
Red	10.73	11.21	22.62	**24.25**
Green	12.42	15.17	16.41	**17.14**
Blue	3.92	2.60	18.39	**19.75**
Color	6.98	6.08	19.41	**20.63**

(a) Channel and color restoration results in terms of SNR. Best performers are in bold.

	ICM	VB-MIG	HOGMep
VI	0.344	0.13	**0.006**
RI	92.72	98.24	**99.98**
% misclass.	41.10	4.70	**0.05**

(b) Segmentation: variation of information [30], Rand index [31].

Table 1: Objective restoration and segmentation performance ('synth').

Segmentation performance, summarized in Table 1b, is numerically assessed through variation of information (VI, [30]) and Rand index (RI, [31]) measures. Segmented images and their binary difference to the original segmentation are displayed in Figure 4. HOGMep offers a very low rate of wrongly assigned pixels, and globally correctly labels all regions. For the 'peppers' image, Table 2 gathers objective SNR and SSIM performance, while restoration is depicted in Figure 5, and segmentation results in Figure 6. Through the obtained results, we conclude that HOGMep outperforms state-of-the-art methods in both recovery and segmentation problems.

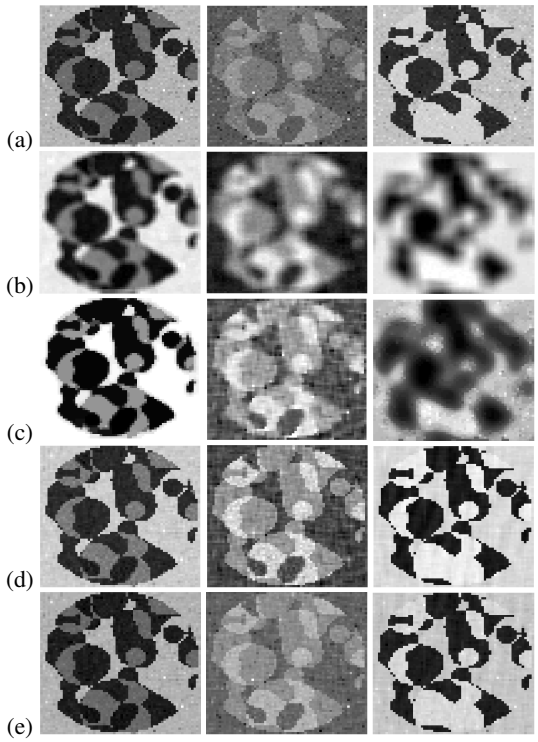


Fig. 2: RGB channels ('synth'): (a) original (b) blurred and noisy. Restored with (c) 3MG, (d) VB-MIG, (e) HOGMep.

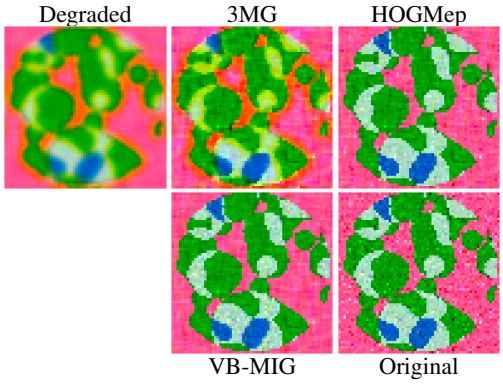


Fig. 3: Restoration results for the color image ('synth').

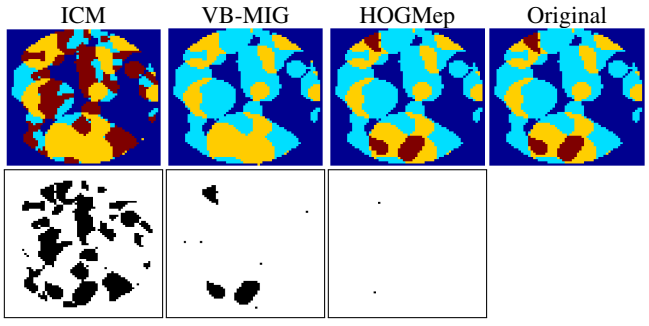


Fig. 4: Segmentation results for 'synth' and binary difference to original ground truth (wrongly assigned pixels are in black).

Channel		Initial	3MG	VB-MIG	HOGMep
Red	SNR	24.30	28.99	19.94	**33.43**
	SSIM	0.938	**0.972**	0.939	**0.972**
Green	SNR	18.36	22.82	21.13	**23.96**
	SSIM	0.824	**0.903**	0.875	**0.904**
Blue	SNR	13.77	14.19	14.12	**14.30**
	SSIM	0.688	0.748	0.779	**0.792**
Color	SNR	19.70	22.25	21.77	**23.00**

Table 2: Objective restoration performance ('peppers'). Best performers are in bold.

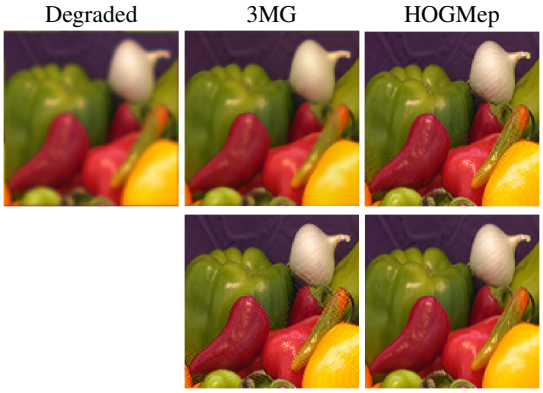


Fig. 5: Restoration results for the color images ('peppers').

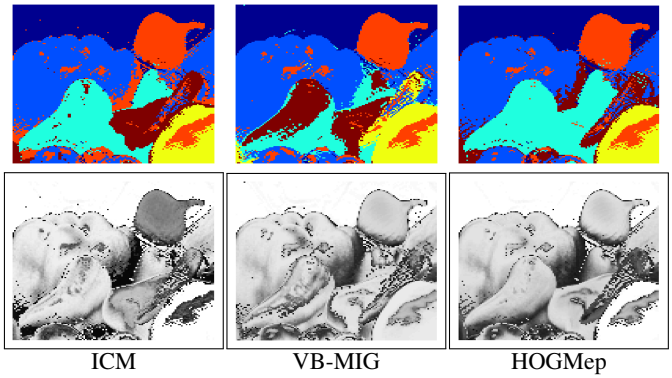


Fig. 6: Segmentation results for 'peppers' and the corresponding silhouette. Silhouette image: blackish pixels reflect silhouette indices close to -1 while whitish pixels correspond to scores near 1. Using this color code, brighter zones reflect satisfying clustering.

5. CONCLUSIONS

In this work, we have performed a joint image recovery and segmentation *via* a Bayesian approach. This formulation combines two main priors on the sought data: an HOGM for the latent label variables and a MEP distribution for the unknown signal within a given class. We have proposed an original VBA method for solving the related problem where most of the hyper-parameters are automatically estimated. The provided simulation results show that HOGMep delivers good results for both recovery and segmentation tasks.

6. REFERENCES

[1] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol. 37, no. 2, pp. 183–233, Nov. 1999.

[2] C. L. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2222–2233, Aug. 2004.

[3] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders, "Variational Bayesian image restoration based on a product of t-distributions image prior," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1795–1805, Oct. 2008.

[4] Z. Chen, S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian methods for multimedia problems," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 1000–1017, Jun. 2014.

[5] Y. Zheng, A. Fraysse, and T. Rodet, "Efficient variational Bayesian approximation method based on subspace optimization," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 681–693, Feb. 2015.

[6] R. A. Choudrey, *Variational Methods for Bayesian Independent Component Analysis*, Ph.D. thesis, University of Oxford, 2002.

[7] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian super resolution," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 984–999, Apr. 2011.

[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 Process.*, vol. 19, no. 9, pp. 2265–2277, Sep. 2010.

[9] L. Chaari, F. Forbes, T. Vincent, M. Dojat, and P. Ciuciu, "Variational solution to the joint detection estimation of brain activity in fMRI," in *Proc. Medical Image Computing and Computer-Assisted Intervention Conf.*, 2011, pp. 260–268.

[10] C. A. McGrory, D. M. Titterington, R. Reeves, and A. N. Pettitt, "Variational Bayes for estimating the parameters of a hidden Potts model," *Stat. Comput.*, vol. 19, no. 3, pp. 329–340, Sep. 2009.

[11] E. Marinari and R. Marra, "Cluster algorithms for the generalized $3d$, $3q$ Potts model," *Nucl. Phys. B*, vol. 342, no. 3, pp. 737–752, Oct. 1990.

[12] E. Zheleva, L. Getoor, and S. Sarawagi, "Higher-order graphical models for classification in social and affiliation networks," in *NIPS Workshop on Networks Across Disciplines: Theory and Applications*, 2010.

[13] E. Gómez-Sánchez-Manzano, M. A. Gómez-Villegas, and J. M. Marín, "Sequences of elliptical distributions and mixtures of normal distributions," *J. Multivar. Anal.*, vol. 97, no. 2, pp. 295–310, Feb. 2006.

[14] Y. Marnissi, A. Benazza-Benyahia, E. Chouzenoux, and J.-C. Pesquet, "Generalized multivariate exponential power prior for wavelet-based multichannel image restoration," in *Proc. Int. Conf. Image Process.*, Melbourne, Australia, Sep. 15-18, 2013, pp. 2402–2406.

[15] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet, "A nonlinear Stein based estimator for multichannel image denoising," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3855–3870, Aug. 2008.

[16] G. Verdoolaege and P. Scheunders, "Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination," *Int. J. Comput. Vis.*, vol. 95, no. 3, pp. 265–286, 2011.

[17] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet, "A nonlocal structure tensor-based approach for multicomponent image recovery problems," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5531–5544, Dec. 2014.

[18] M. Pereyra, N. Dobigeon, H. Batatia, and J. 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.

[19] 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 Workshop Stat. Signal Process.*, Palma de Majorca, Spain, Jun. 26-29, 2016.

[20] J. Bioucas-Dias, F. Condessa, and J. Kovačević, "Alternating direction optimization for image segmentation using hidden Markov measure field models," in *Proc. SPIE Image Process. Algorithms Syst.*, K. O. Egiazarian, S. S. Agaian, and A. P. Gotchev, Eds., San Francisco, CA, USA, Feb. 3-5, 2014, SPIE.

[21] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A majorize-minimize subspace approach for ℓ_2-ℓ_0 image regularization," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 563–591, Mar. 2013.

[22] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, "Proximal algorithms for multicomponent image processing," *J. Math. Imaging Vision*, vol. 41, no. 1, pp. 3–22, Sep. 2011.

[23] Y. Marnissi, E. Chouzenoux, J.-C. Pesquet, and A. Benazza-Benyahia, "An auxiliary variable method for Langevin based MCMC algorithms," in *Proc. IEEE Workshop Stat. Signal Process.*, Palma de Majorca, Spain, Jun. 26-29, 2016.

[24] E. Gómez-Sánchez-Manzano, M. A. Gómez-Villegas, and J. M. Marín, "Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications," *Commun. Stat. Theory Methods*, vol. 37, no. 6, pp. 972–985, 2008.

[25] V. Šmídl and A. Quinn, *The Variational Bayes Method in Signal Processing*, Springer, 2006.

[26] J. A. Palmer, D. P. Wipf, K. Kreutz-Delgado, and B. D. Rao, "Variational EM algorithms for non-Gaussian latent variable models," in *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, 2005, pp. 1059–1066.

[27] Y. Zheng, A. Fraysse, and T. Rodet, "Wavelet based unsupervised variational Bayesian image reconstruction approach," in *Proc. Eur. Sig. Image Proc. Conf.*, Nice, France, Aug. 31-Sep. 4, 2015.

[28] E. Chouzenoux, F. Zolyniak, E. Gouillart, and H. Talbot, "A Majorize-Minimize memory gradient algorithm applied to X-ray tomography," in *Proc. Int. Conf. Image Process.*, Melbourne, Australia, Sep. 15-18, 2013.

[29] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 48, no. 3, pp. 259–302, 1986.

[30] M. Meilă, "Comparing clusterings by the variation of information," in *Learning Theory and Kernel Machines*, Bernhard Schölkopf and Manfred K. Warmuth, Eds., vol. 2777 of *Lect. Notes Comput. Sci.*, pp. 173–187. Springer, 2003.

[31] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Am. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, Dec. 1971.