Fast unsupervised Bayesian image segmentation with adaptive spatial regularisation

Fast Unsupervised Bayesian Image Segmentation With Adaptive Spatial Regularisation

Marcelo Pereyra and Steve McLaughlin, *Fellow, IEEE*

***Abstract*— This paper presents a new Bayesian estimation technique for hidden Potts–Markov random fields with unknown regularisation parameters, with application to fast unsupervised K-class image segmentation. The technique is derived by first removing the regularisation parameter from the Bayesian model by marginalisation, followed by a small-variance-asymptotic (SVA) analysis in which the spatial regularisation and the integer-constrained terms of the Potts model are decoupled. The evaluation of this SVA Bayesian estimator is then relaxed into a problem that can be computed efficiently by iteratively solving a convex total-variation denoising problem and a least-squares clustering (K-means) problem, both of which can be solved straightforwardly, even in high-dimensions, and with parallel computing techniques. This leads to a fast fully unsupervised Bayesian image segmentation methodology in which the strength of the spatial regularisation is adapted automatically to the observed image during the inference procedure, and that can be easily applied in large 2D and 3D scenarios or in applications requiring low computing times. Experimental results on synthetic and real images, as well as extensive comparisons with state-of-the-art algorithms, confirm that the proposed methodology offer extremely fast convergence and produces accurate segmentation results, with the important additional advantage of self-adjusting regularisation parameters.**

***Index Terms*— Image segmentation, Bayesian methods, spatial mixture models, Potts Markov random field, convex optimisation.**

I. Introduction

IMAGE segmentation is a canonical inverse problem which involves classifying image pixels into clusters that are spatially coherent and have well defined boundaries. It is widely accepted that this task can be formulated as a statistical inference problem and most state-of-the-art image segmentation methods compute solutions by performing statistical inference (e.g., computing penalized maximum likelihood or maximum-a-posteriori estimates). In this paper we focus on new Bayesian computation methodology for hidden Potts-Markov random fields (MRFs) [1], a powerful class of statistical models that is widely used in Bayesian image segmentation methods (see [2]–[5] for recent examples in hyperspectral, non destructive testing, ultrasound, and fMRI imaging).

Despite the wide range of applications, performing inference on hidden Potts MRFs remains a computationally challenging problem. In particular, computing the maximum-a-posteriori (MAP) estimator for these models is generally NP-hard, and thus most image processing methods compute approximate estimators. This has driven the development of efficient approximate inference algorithms, particularly over the last decade. The current predominant approaches for approximate inference on MRFs are based on convex models and convex approximations that can be solved efficiently by convex optimisation [6]–[8], and on approximate estimators computed with graph-cut [9], [10] and message passing algorithms [11]–[13]. In a similar fashion, modern algorithms to solve active contour models, the other main class of models for image segmentation, are also principally based on convex relaxations and convex optimisation [14], [15] and on Riemannian steepest descent optimisation schemes [16]–[19].

An important limitation of these computationally efficient approaches is that they are supervised, in the sense that they require practitioners to specify the value of the regularisation parameter of the Potts MRF. However, it is well known that appropriate values for regularisation parameters can be highly image dependent and sometimes difficult to select a priori, thus requiring practitioners to set parameter values heuristically or by visual cross-validation. The Bayesian framework offers a range of strategies to circumvent this problem and to design unsupervised image segmentation inference procedures that self-adjust their regularisation parameters. Unfortunately, the computations involved in these inferences are beyond the scope of existing fast approximate inference algorithms. As a consequence, unsupervised image segmentation methods have to use more computationally intensive strategies such as Monte Carlo approximations [20], [21], variational Bayes approximations [22], and EM algorithms based on mean-field like approximations [23], [24].

In this paper we propose a highly efficient Bayesian computation approach specifically designed for performing approximate inference on hidden Potts-Markov random fields with unknown regularisation parameters, with application to fast unsupervised K-class image segmentation. A main original contribution of our development reported here is to use a small-variance-asymptotic (SVA) analysis to design an

Manuscript received February 2, 2016; revised July 27, 2016, November 19, 2016, and January 20, 2017; accepted February 19, 2017. Date of publication March 15, 2017; date of current version April 11, 2017. This work was supported in part by the SuSTaIN Program - EPSRC at the Department of Mathematics, University of Bristol, under Grant EP/D063485/1, in part by the Post-Doctoral Fellowship from the French Ministry of Defence, and in part by EPSRC under Grant EP/J015180/1. Part of this work was conducted when MP held a Marie Curie Intra-European Fellowship for Career Development at the School of Mathematics, University of Bristol. Part of this work was presented at EUSIPCO'14 in Lisbon 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David Clausi.

M. Pereyra is with the School of Mathematical and Computer Sciences, Heriot Watt University, Edinburgh EH14 4AS, U.K. (e-mail: m.pereyra@hw.ac.uk).

S. McLaughlin is with the School of Engineering and Physical Sciences, Heriot Watt University, Edinburgh EH14 4AS, U.K. (e-mail: s.mclaughlin@hw.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2675165

approximate MAP estimator in which the spatial regularisation and the integer-constrained terms of the Potts model are decoupled. The evaluation of this SVA Bayesian estimator can then be relaxed into a problem that can be computed efficiently by iteratively solving a convex total-variation denoising problem and a least-squares clustering (K-means) problem, both of which can be solved straightforwardly, even in high-dimensions, and with parallel computing techniques.

Small-variance asymptotics estimators were introduced in [25] as a computationally efficient framework for performing inference in Dirichlet process mixture models and have been recently applied to other important machine learning classification models such as the Beta process and sequential hidden Markov models [26], as well as to the problem of configuration alignment and matching [27]. Here we exploit these same techniques for the hidden Potts MRF to develop an accurate and computationally efficient image segmentation methodology for the fully unsupervised case of unknown class statistical parameters (e.g., class means) and unknown Potts regularisation parameter.

The paper is organised as follows: in Section II we present a brief background to Bayesian image segmentation using the Potts MRF. This then followed by a detailed development of our proposed methodology. In Sections IV and V the methodology is applied to some synthetic and real test images and compared to other image segmentation approaches from the state of the art. Finally some brief conclusions are drawn in Section VI.

II. Background

We begin by recalling the standard Bayesian model used in image segmentation problems, which is based on a finite mixture model and a hidden Potts-Markov random field with known regularisation parameter β. For simplicity we focus on univariate Gaussian mixture models. However, the results presented hereafter can be generalised to all exponential-family mixture models (e.g., mixtures of multivariate Gaussian, Rayleigh, Poisson, Gamma, Binomial, etc.) by following the approach described in [28].

Let $y_n \in \mathbb{R}$ denote the nth observation (i.e. pixel or voxel) in a lexicographical vectorized image $\boldsymbol{y} = (y_1, \ldots, y_N)^T \in \mathbb{R}^N$. We assume that $\boldsymbol{y}$ is made up by K regions $\{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$ such that the observations in the kth class are distributed according to the following conditional marginal observation model

$$y_n | n \in \mathcal{C}_k \sim \mathcal{N}(\mu_k, \sigma^2), \tag{1}$$

where $\mu_k \in \mathbb{R}$ represents the mean intensity of class $\mathcal{C}_k$. For identifiability we assume that $\mu_k \neq \mu_j$ for all $k \neq j$.

To perform segmentation, a label vector $\boldsymbol{z} = (z_1, \ldots, z_N)^T$ is introduced to map or classify observations $\boldsymbol{y}$ to classes $\mathcal{C}_1, \ldots, \mathcal{C}_K$ (i.e., $z_n = k$ if and only if $n \in \mathcal{C}_k$). Assuming that observations are conditionally independent given $\boldsymbol{z}$ and given the parameter vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$, the likelihood of $\boldsymbol{y}$ can be expressed as follows

$$f(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\mu}) = \prod_{k=1}^{K} \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n | \mu_k, \sigma^2), \tag{2}$$

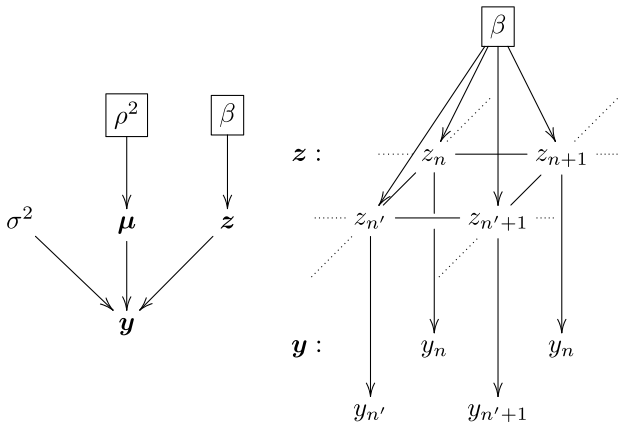


Fig. 1. [Left:] Directed acyclic graph of the standard Bayesian model for image segmentation (parameters with fixed values are represented using black boxes). [Right] Local hierarchical representation of the hidden Potts MRF and the observed image for 4 neighbouring pixels.

with $\mathcal{S}_k = \{n : z_n = k\}$ (to simplify notation the dependence of distributions on σ^2 is omitted). A Bayesian model for image segmentation is then defined by specifying the prior distribution of the unknown parameter vector $(\boldsymbol{z}, \boldsymbol{\mu})$. The prior for $\boldsymbol{z}$ is the homogenous K-state Potts MRF [29]

$$f(\boldsymbol{z}|\beta) = \frac{1}{C(\beta)} \exp[\beta H(\boldsymbol{z})], \tag{3}$$

with regularisation hyper-parameter $\beta \in \mathbb{R}^+$, Hamiltonian

$$H(\boldsymbol{z}) = \sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \delta(z_n == z_{n'}), \tag{4}$$

where $\delta(\cdot)$ is the Kronecker function and $\mathcal{V}(n)$ is the index set of the neighbors of the nth voxel (most methods use the 1st order neighbourhoods depicted in Fig. 2), and normalising constant (or partition function)

$$C(\beta) = \sum_{\boldsymbol{z}} \exp[\beta H(\boldsymbol{z})]. \tag{5}$$

Notice that the Potts prior (3) is defined conditionally to a given value of β. Most image segmentation methods based on this prior are supervised; i.e., assume that the value of β is known and specified a priori by the practitioner. Alternatively, unsupervised methods consider that β is unknown and seek to adjust its value automatically during the image segmentation procedure (this point is explained in detail in Section III).

In a similar fashion, the class means are considered prior independent and assigned Gaussian priors $\mu_k \sim \mathcal{N}(0, \rho^2)$ with fixed variance ρ^2,

$$f(\boldsymbol{\mu}) = \prod_{k=1}^{K} p_{\mathcal{N}}(\mu_k | 0, \rho^2). \tag{6}$$

(to simplify notation the dependence of distributions on the fixed quantity ρ^2 is omitted).

Then, using Bayes theorem and taking into account the conditional independence structure of the model (see Fig. 1), the joint posterior distribution of $(\boldsymbol{z}, \mu)$ given $\boldsymbol{y}$ and β can be expressed as follows

$$f(\boldsymbol{z}, \boldsymbol{\mu} | \boldsymbol{y}, \beta) \propto f(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\mu}) f(\boldsymbol{z}|\beta) f(\boldsymbol{\mu}), \tag{7}$$

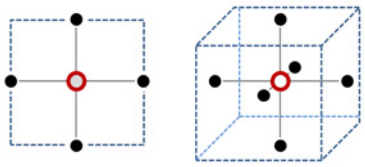


Fig. 2. 4-pixel (left) and 6-voxel (right) neighborhood structures. The pixel/voxels considered appears as a void red circle whereas its neighbors are depicted in full black and blue.

where $\propto$ denotes proportionality up to a normalising constant that can be retrieved by setting $\int f(z, \boldsymbol{\mu}|\boldsymbol{y}, \beta)\, \mathrm{d}z\mathrm{d}\boldsymbol{\mu} = 1$. The graphical structure of this Bayesian model is summarised in Fig. 1 below. Notice the Markovian structure of z and that observations y_n are conditionally independent given the model parameters z, $\boldsymbol{\mu}$ and σ^2.

Finally, given the Bayesian model (7), a segmentation of $\boldsymbol{y}$ is typically obtained by computing the MAP estimator

$$\hat{z}_1, \hat{\boldsymbol{\mu}}_1 = \underset{z,\boldsymbol{\mu}}{\operatorname{argmax}}\ f(z, \boldsymbol{\mu}|\boldsymbol{y}, \beta), \tag{8}$$

which can also be obtained by solving the equivalent optimisation problem

$$\hat{z}_1, \hat{\boldsymbol{\mu}}_1 = \underset{z,\boldsymbol{\mu}}{\operatorname{argmin}} -\log f(z, \boldsymbol{\mu}|\boldsymbol{y}, \beta). \tag{9}$$

Unfortunately these optimisation problems are known to be NP-hard due to the combinatorial nature of the Potts Hamiltonian $H(z)$ defined in (4). As mentioned previously, modern image segmentation methods based on (7) typically address this issue by using approximate (local) integer optimisation algorithms (e.g., graph-cut, message passing) [10]–[12], and more recently with convex relaxations of the Potts model (see for instance [6], [7]).

III. Proposed Method

This section presents a highly computationally efficient approach for performing approximate inference on z when the value of the regularisation parameter β is unknown. The approach is based on a small-variance asymptotics (SVA) analysis combined with a convex relaxation and a pseudo-likelihood approximation of the Potts MRF. Our development has three main steps. In the first step we adopt a hierarchical Bayesian approach to remove β from the model by marginalisation; because marginalising w.r.t. β requires knowledge of the intractable Potts partition function (5) we use a pseudo-likelihood approximation. However, performing inference with the resulting marginalised model is still NP-hard. In the second part of our development we address this difficulty by using auxiliary variables and an SVA analysis to decouple the spatial regularisation and the integer-constrained terms of the Potts model. The evaluation of the resulting SVA Bayesian estimator is then relaxed into a problem that can be computed efficiently by iteratively solving a convex total-variation denoising problem and a least-squares clustering problem, both of which can be solved straightforwardly, even in high-dimensions, with parallel implementations of Chambolle's optimisation algorithm [30] and of K-means [31].

A. Marginalisation of the Regularisation Parameter β

Following a hierarchical Bayesian approach, we address the fact that the value of β is unknown by modelling it as an additional random variable of the Bayesian model. Precisely, we assign β a prior distribution $f(\beta)$ and define an augmented model that includes β within its unknown parameter vector. By using Bayes' theorem we obtain the joint posterior distribution

$$f(z, \boldsymbol{\mu}, \beta|\boldsymbol{y}) \propto f(\boldsymbol{y}|z, \boldsymbol{\mu}) f(\boldsymbol{\mu}) f(z|\beta) f(\beta) \tag{10}$$

which includes β as an unknown variable. The rationale for replacing the fixed regularisation parameter β of (7) by a random variable with prior $f(\beta)$ is that it is often possible to specify this prior distribution such that the amount of regularisation enforced by the Potts MRF is driven by data and the impact of $f(\beta)$ on the inferences is minimal. At the same time, experienced practitioners with knowledge of good values of β can specify $f(\beta)$ to exploit their prior beliefs. In this paper we use a gamma (hyper-)prior distribution

$$f(\beta) = \gamma^{\alpha} \beta^{\alpha-1} \exp(-\gamma\beta) \mathbf{1}_{\mathbb{R}^+}(\beta) / \Gamma(\alpha)$$

because it has favourable analytical tractability properties that will be useful for our development (appropriate values for the fixed parameters α and γ will be derived later through a small-variance asymptotics analysis).

Moreover, in order to marginalise β from the model we notice that β is conditionally independent of $\boldsymbol{y}$ given z; to be precise, that $f(z, \boldsymbol{\mu}, \beta|\boldsymbol{y}) = f(\beta|z) f(z, \boldsymbol{\mu}|\boldsymbol{y})$. Therefore, integrating $f(z, \boldsymbol{\mu}, \beta|\boldsymbol{y})$ with respect to β is equivalent to redefining the posterior distribution (12) with the marginal prior $f(z) = \int_{\mathbb{R}^+} f(z, \beta)\mathrm{d}\beta$. Evaluating this marginal prior exactly is not possible because it requires computing the normalising constant of the Potts model $C(\beta)$ defined in (5), which is a reputedly intractable problem [20]. To obtain an analytically tractable approximation for this marginal prior we adopt a pseudo-likelihood approach [32] and use the approximation $C(\beta) \propto \beta^{-N}$, leading to

$$\begin{aligned} f(z) &= \int_{\mathbb{R}^+} f(z, \beta)\mathrm{d}\beta \\ &\propto \int_{\mathbb{R}^+} \beta^N \exp(\beta H(z)) \beta^{\alpha-1} \exp(-\gamma\beta)\mathrm{d}\beta \\ &\propto [\gamma - H(z)]^{-(\alpha+N)}, \end{aligned} \tag{11}$$

and to the following (marginal) posterior distribution

$$\begin{aligned} f(z, \boldsymbol{\mu}|\boldsymbol{y}) \propto &\left[\prod_{k=1}^{K} \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}\left(y_n|\mu_k, \sigma^2\right)\right] \\ &\times f(\boldsymbol{\mu}) (\gamma - H(z))^{-(\alpha+N)}, \end{aligned} \tag{12}$$

that does not depend on the regularisation parameter β.

B. Small-Variance Approximation

The next step of our development is to conduct a small-variance asymptotics analysis on (12) and derive the asymptotic MAP estimator of $z, \boldsymbol{\mu}$. We begin by introducing a carefully selected auxiliary vector $\boldsymbol{x}$ such that $\boldsymbol{y}$ and $(z, \boldsymbol{\mu})$

are conditionally independent given $\boldsymbol{x}$, and that the posterior $f(\boldsymbol{x}, z, \boldsymbol{\mu}|\boldsymbol{y})$ has the same maximisers as (7) (after projection on the space of $(z, \boldsymbol{\mu})$). More precisely, we define a random vector $\boldsymbol{x} \in \mathbb{R}^N$ with degenerate prior

$$f(\boldsymbol{x}|z, \boldsymbol{\mu}) = \prod_{k=1}^{K} \prod_{n \in \mathcal{S}_k} \delta(x_n - \mu_k), \tag{13}$$

and express the likelihood of $\boldsymbol{y}$ given $\boldsymbol{x}, z$ and $\boldsymbol{\mu}$ as

$$f(\boldsymbol{y}|\boldsymbol{x}, z, \boldsymbol{\mu}) = f(\boldsymbol{y}|\boldsymbol{x}) = \prod_{n=1}^{N} p_{\mathcal{N}}(y_n|x_n, \sigma^2).$$

The prior distributions for z and $\boldsymbol{\mu}$ remain as defined above. The joint posterior distribution of $\boldsymbol{x}, z, \boldsymbol{\mu}$ is given by

$$\begin{aligned} f(\boldsymbol{x}, z, \boldsymbol{\mu}, \beta|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{x}|z, \boldsymbol{\mu}) f(z|\beta) f(\boldsymbol{\mu}) \\ &\propto \left[\prod_{k=1}^{K} \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n|x_n, \sigma^2) \delta(x_n - \mu_k) \right] \\ &\quad \times f(\boldsymbol{\mu}) \left[\gamma - H(z) \right]^{-(\alpha+N)}. \end{aligned} \tag{14}$$

Notice that from an inference perspective (14) is equivalent to (12), in the sense that marginalising $\boldsymbol{x}$ in (14) results in (12).

Moreover, we define $H^*(z)$ as the "complement" of the Hamiltonian $H(z)$ in the sense that for any $z \in [1, \ldots, K]^N$

$$H(z) + H^*(z) = \sum\nolimits_{n=1}^{N} |\mathcal{V}(n)|,$$

where $|\mathcal{V}(n)|$ denotes the cardinality of the neighbourhood structure of the nth pixel. For the Potts MRF this complement is given by

$$H^*(z) \triangleq \sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \delta(z_n \neq z_{n'}). \tag{15}$$

Replacing $H(z) = \sum_{n=1}^{N} |\mathcal{V}(n)| - H^*(z)$ in (14) we obtain

$$\begin{aligned} &f(\boldsymbol{x}, z, \boldsymbol{\mu}, \beta|\boldsymbol{y}) \\ &\quad \propto \left(\prod_{k=1}^{K} \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n|x_n, \sigma^2) \delta(x_n - \mu_k) \right) \\ &\qquad \times f(\boldsymbol{\mu}) \left[H^*(z) + (\gamma - \textstyle\sum_{n=1}^{N} |\mathcal{V}(n)|) \right]^{-(\alpha+N)}. \end{aligned} \tag{16}$$

Furthermore, noting that $H^*(z)$ only measures if neighbour labels are identical or not, regardless of their values, it is easy to check that the posterior (14) remains unchanged if we substitute $H^*(z)$ with $H^*(\boldsymbol{x})$

$$\begin{aligned} &f(\boldsymbol{x}, z, \boldsymbol{\mu}, \beta|\boldsymbol{y}) \\ &\quad \propto \left[\prod_{k=1}^{K} \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n|x_n, \sigma^2) \delta(x_n - \mu_k) \right] \\ &\qquad \times f(\boldsymbol{\mu}) \left[H^*(\boldsymbol{x}) + (\gamma - \textstyle\sum_{n=1}^{N} |\mathcal{V}(n)|) \right]^{-(\alpha+N)}. \end{aligned} \tag{17}$$

Finally, we make the observation that for 1st order neighbourhoods (see Fig. 2) we have $H^*(\boldsymbol{x}) = 2||\nabla \boldsymbol{x}||_0$, where $||\nabla \boldsymbol{x}||_0 = ||\nabla_h \boldsymbol{x}||_0 + ||\nabla_v \boldsymbol{x}||_0$ denotes the ℓ_0 norm of the horizontal and vertical components of the 1st order discrete gradient of $\boldsymbol{x}$, and therefore

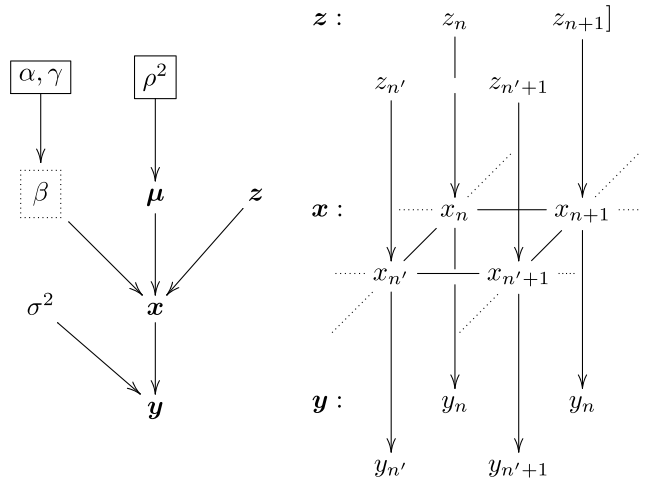


Fig. 3. [Left:] Directed acyclic graph of the proposed Bayesian model, augmented by the auxiliary variable $\boldsymbol{x}$ decoupling $\boldsymbol{\mu}$ and z from $\boldsymbol{y}$, and with marginalisation of the regularisation parameter β (parameters with fixed values are represented using solid black boxes, marginalised variables appear in dashed boxes). [Right] Local representation of three layers of the model for 4 neighbouring pixels.

$$\begin{aligned} &f(\boldsymbol{x}, z, \boldsymbol{\mu}, \beta|\boldsymbol{y}) \\ &\quad \propto \left[\prod_{k=1}^{K} \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n|x_n, \sigma^2) \delta(x_n - \mu_k) \right] \\ &\qquad \times f(\boldsymbol{\mu}) \left[||\nabla \boldsymbol{x}||_0 + (\gamma - \textstyle\sum_{n=1}^{N} |\mathcal{V}(n)|)/2 \right]^{-(\alpha+N)}. \end{aligned} \tag{18}$$

The graphical structure of this equivalent hierarchical Bayesian model is summarised in Fig. 3 below. Notice that in this model $\boldsymbol{x}$ separates $\boldsymbol{y}$ and σ^2 from the other model parameters, that the regularisation parameter β has been marginalised, that the MRF is now enforcing spatial smoothness on $\boldsymbol{x}$ not z, and that the elements of z are prior independent.

We are now ready to conduct a small-variance asymptotics analysis on (18) and derive the asymptotic MAP estimator of $\boldsymbol{x}$, z, $\boldsymbol{\mu}$, which is defined for our model as [25]

$$\operatorname*{argmin}_{\boldsymbol{x}, z, \boldsymbol{\mu}} \lim_{\sigma^2 \to 0} -\sigma^2 \log f(\boldsymbol{x}, z, \boldsymbol{\mu}|\boldsymbol{y}).$$

First, we use the fact that $\delta(s) = \lim_{\tau^2 \to 0} p_{\mathcal{N}}(s|0, \tau^2)$ to express (18) as follows

$$\begin{aligned} &f(\boldsymbol{x}, z, \boldsymbol{\mu}|\boldsymbol{y}, \beta) \\ &\propto \lim_{\tau^2 \to 0} \left(\prod_{k=1}^{K} \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n|x_n, \sigma^2) p_{\mathcal{N}}(x_n|\mu_k, \tau^2) \right) \\ &\qquad \times f(\boldsymbol{\mu}) \left[||\nabla \boldsymbol{x}||_0 + (\gamma - \textstyle\sum_{n=1}^{N} |\mathcal{V}(n)|)/2 \right]^{-(\alpha+N)}, \\ &\propto \lim_{\tau^2 \to 0} \left(\prod_{k=1}^{K} \prod_{n \in \mathcal{S}_k} \exp\left(-\frac{(x_n - y_n)^2}{2\sigma^2} - \frac{(x_n - \mu_k)^2}{2\tau^2} \right) \right) \\ &\qquad \times f(\boldsymbol{\mu}) \left[||\nabla \boldsymbol{x}||_0 + (\gamma - \textstyle\sum_{n=1}^{N} |\mathcal{V}(n)|)/2 \right]^{-(\alpha+N)}. \end{aligned} \tag{19}$$

Then, in a manner akin to Broderick et al. [25], we allow the model's hyper parameters to scale with σ^2 in order to preserve the balance between the prior and the likelihood and avoid a trivial limit. More precisely, we set $\alpha = N/\sigma^2$ and assume that σ^2 vanishes at the same speed as τ^2. Then, the limit of $-\sigma^2 \log f(\boldsymbol{x}, z, \boldsymbol{\mu}|\boldsymbol{y})$ as $\sigma^2 \to 0$ is given by

$$\begin{aligned}\lim_{\sigma^2\to 0} & -\sigma^2 \log f(\boldsymbol{x}, z, \boldsymbol{\mu}|\boldsymbol{y}) \\ &= \sum_{k=1}^{K} \sum_{n\in\mathcal{S}_k} \frac{1}{2}(x_n - y_n)^2 + \frac{1}{2}(x_n - \mu_k)^2 \\ &\quad + N \log(||\nabla \boldsymbol{x}||_0 + (\gamma - \textstyle\sum_{n=1}^{N} |\mathcal{V}(n)|)/2), \end{aligned} \tag{20}$$

and the MAP asymptotic estimators of $\boldsymbol{x}$, z, $\boldsymbol{\mu}$ by

$$\begin{aligned}\underset{\boldsymbol{x},z,\boldsymbol{\mu}}{\operatorname{argmin}} & \sum_{k=1}^{K} \sum_{n\in\mathcal{S}_k} \frac{1}{2}(x_n - y_n)^2 + \frac{1}{2}(x_n - \mu_k)^2 \\ &\quad + N \log(||\nabla \boldsymbol{x}||_0 + 1), \end{aligned} \tag{21}$$

where we have set $\gamma = 2 + \sum_{n=1}^{N} |\mathcal{V}(n)|$ such that the penalty $\log\left[||\nabla \boldsymbol{x}||_0 + (\gamma - \sum_{n=1}^{N} |\mathcal{V}(n)|)/2\right] \geq 0$.

C. Convex Relaxation and Optimisation

Computing the estimator (21) is still NP-hard due to $\log(||\nabla \boldsymbol{x}||_0 + 1)$. To address this difficulty we use a convex relaxation of $||\nabla \boldsymbol{x}||_0$ and exploit the concavity of the logarithmic function. Precisely, we replace $||\nabla \boldsymbol{x}||_0$ by the convex approximation $\mathrm{TV}(\boldsymbol{x}) = ||\nabla \boldsymbol{x}||_{1-2}$, (i.e., the isotropic total-variation pseudo-norm of $\boldsymbol{x}$ [33]), and obtain the following optimisation problem

$$\begin{aligned}\underset{\boldsymbol{x},z,\boldsymbol{\mu}}{\operatorname{argmin}} & \sum_{k=1}^{K} \sum_{n\in\mathcal{S}_k} \frac{1}{2}(x_n - y_n)^2 + \frac{1}{2}(x_n - \mu_k)^2 \\ &\quad + N \log(TV(\boldsymbol{x}) + 1), \end{aligned} \tag{22}$$

which can be very efficiently computed by iterative minimisation w.r.t. $\boldsymbol{x}$, z and $\boldsymbol{\mu}$. The minimisation of (22) w.r.t. z (with $\boldsymbol{x}$ and $\boldsymbol{\mu}$ fixed) is a trivial separable integer problem that can be formulated as N independent (pixel-wise) minimisation problems over $1, \ldots, K$ (these unidimensional integer problems can be solved by simply checking the value $z_n = 1, \ldots, K$ that minimises (22) for each pixel $n = 1, \ldots, N$). Similarly, the minimisation with respect to $\boldsymbol{\mu}$ is a trivial quadratic least squares fitting problem with analytic solution (i.e., by setting $\mu_k = \frac{1}{|S_k|}\sum_{n\in\mathcal{S}_k} x_n$ for each $k = 1, \ldots, K$, where $|S_k|$ denotes the cardinality of S_k). Also note that iteratively minimising (22) with respect to z and $\boldsymbol{\mu}$, with fixed $\boldsymbol{x}$, is equivalent to solving a least squares clustering problem with the popular K-means algorithm [31]. Moreover, the minimisation of (22) w.r.t. $\boldsymbol{x}$ (with z and $\boldsymbol{\mu}$ fixed) is achieved by solving the non-convex optimisation problem

$$\begin{aligned}\underset{\boldsymbol{x}}{\operatorname{argmin}} & \sum_{k=1}^{K} \sum_{n\in\mathcal{S}_k} \frac{1}{2}(x_n - y_n)^2 + \frac{1}{2}(x_n - \mu_k)^2 \\ &\quad + N \log\left[TV(\boldsymbol{x}) + 1\right], \end{aligned} \tag{23}$$

Algorithm 1 Unsupervised Bayesian Segmentation Algorithm

1: Input: Image $\boldsymbol{y}$, number of maximum outer iterations T and inner iterations L, tolerance ϵ.
2: Initialise $\boldsymbol{x}^{(0)} = 2\boldsymbol{y}$, $z = [1, \ldots, 1]^T$, $\boldsymbol{\mu} = [0, \ldots, 0]^T$.
3: **for** $t = 1 : T$ **do**
4: Set $\boldsymbol{v}^{(0)} = \boldsymbol{x}^{(t-1)}$.
5: **for** $\ell = 0 : L$ **do**
6: Set $\lambda_\ell = N/\{TV[\boldsymbol{v}^{(\ell)}] + 1\}$.
7: Compute $\boldsymbol{v}^{(\ell+1)}$ using (24), with fixed $z = z^{(t-1)}$ and $\boldsymbol{\mu} = \boldsymbol{\mu}^{(t-1)}$, using Chambolle's algorithm [30].
8: **if** $(N/\{TV[\boldsymbol{v}^{(\ell+1)}] + 1\} - \lambda) \geq \epsilon\lambda$ **then**
9: Set $\ell = \ell + 1$.
10: **else**
11: Exit to line 14.
12: **end if**
13: **end for**
14: Set $\boldsymbol{x}^{(t)} = \boldsymbol{v}^{(L)}$.
15: Compute $z^{(t)}$ and $\boldsymbol{\mu}^{(t)}$ by least-squares clustering of $\boldsymbol{x}^{(t)}$ using the K-means algorithm [31].
16: **if** $z^{(t)} \neq z^{(t-1)}$ **then**
17: Set $t = t + 1$.
18: **else**
19: Exit to line 22.
20: **end if**
21: **end for**
22: Output: Segmentation $z^{(t)}$, $\boldsymbol{\mu}^{(t)}$, $\lambda = N/(TV[\boldsymbol{x}^{(t)}] + 1)$.

which was studied in detail in [34]. Essentially, given some initial condition $\boldsymbol{v}^{(0)} \in \mathbb{R}^N$, (23) can be efficiently minimised by majorisation-minimisation (MM) by iteratively solving the following sequence of trivial convex problems,

$$\begin{aligned}\boldsymbol{v}^{(\ell+1)} = \underset{\boldsymbol{x}}{\operatorname{argmin}} & \sum_{k=1}^{K} \sum_{n\in\mathcal{S}_k} \frac{1}{2}(x_n - y_n)^2 + \frac{1}{2}(x_n - \mu_k)^2 \\ &\quad + \lambda_\ell TV(\boldsymbol{x}) \quad \text{with } \lambda_\ell = \frac{N}{TV[\boldsymbol{v}^{(\ell)}]+1}, \end{aligned} \tag{24}$$

in which λ_ℓ plays the role of a regularisation parameter, and where we have used the majorant [34]

$$\begin{aligned} q(\boldsymbol{x}|\boldsymbol{v}^{(\ell)}) &= \frac{\left(TV(\boldsymbol{x}) - TV(\boldsymbol{v}^{(\ell)})\right)}{(TV(\boldsymbol{v}^{(\ell)}) + 1)} + \log\left(TV(\boldsymbol{x}) + 1\right) \\ &\geq \log\left(TV(\boldsymbol{v}^{(\ell)}) + 1\right). \end{aligned} \tag{25}$$

Notice that each step of (24) is equivalent to a trivial convex total-variation denoising problem that can be very efficiently solved, even in high-dimensional scenarios, by using modern convex optimisation techniques (in this paper we used a parallel implementation of Chambolle's algorithm [30]).

The proposed unsupervised segmentation algorithm based on (22) is summarised in Algo. 1 below. We note at this point that because the overall minimisation problem is not convex the solution obtained by iterative minimisation of (22) might depend on the initial values of $\boldsymbol{x}, z, \boldsymbol{\mu}$. In all our experiments we have used the initialisation $\boldsymbol{x}^{(0)} = 2\boldsymbol{y}$, $z = [1, \ldots, 1]^T$, $\boldsymbol{\mu} = [0, \ldots, 0]^T$ that produced good estimation results.

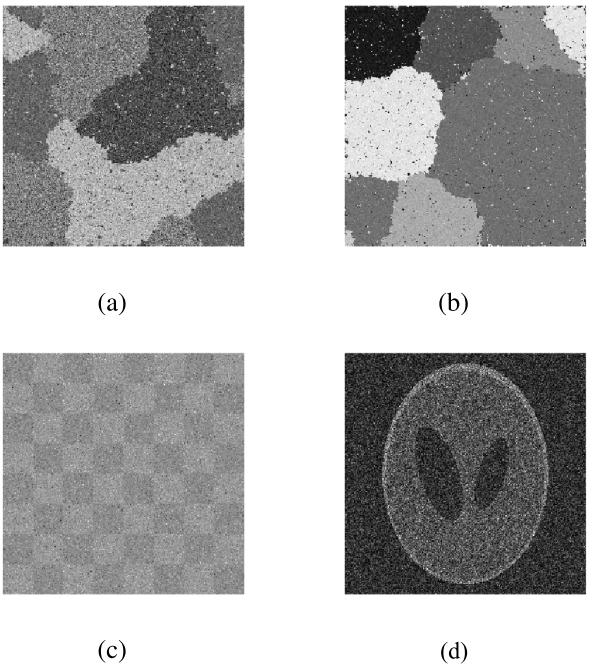


Fig. 4. The four synthetic datasets used to benchmark the proposed image segmentation methodology. Segmentation accuracy and computing times reported in Tables I-II. (a) GMM4. (b) GMM8. (c) LMM2. (d) PMM3.

IV. Validation With Synthetic Data

In this section we validate the proposed Bayesian image segmentation methodology with a series of experiments on synthetic data for which we have ground truth available. To assess the accuracy of our method we compare the results with the estimates produced by the Markov chain Monte Carlo algorithm [20], which estimates the marginal posterior of the segmentation labels $f(\boldsymbol{z}|\boldsymbol{y})$ with very high accuracy. For completeness, we also report comparisons with the Iterated Conditional Modes (ICM) method [32], which is the predominant approach to perform approximate inferences with the hidden Potts MRF model. We consider two fully unsupervised instances of this method. The first is a non-iterative algorithm in which $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{z}$ are initialised by K-means clustering, followed by β estimated from $\boldsymbol{z}$ by pseudo-likelihood estimation [35], and finally $\boldsymbol{z}$ estimated by ICM conditionally on the values of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and β. The second instance is an iterative algorithm in which we update alternatively $\boldsymbol{z}$ by ICM, β by pseudo-likelihood estimation, and $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ by maximum-likelihood estimation, until the estimate of $\boldsymbol{z}$ stabilises (this algorithm is also initialised by K-means clustering). The iterative instance is generally more accurate than the non-iterative one because the estimates of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and β are refined in each iteration, however it is also more computationally expensive.

We tested the algorithms with the four synthetic datasets displayed in Figure 4, which we have designed to represent a range of challenging segmentation conditions related to high-noise, large numbers of classes, and model misspecification (i.e., deviations from the model such as heteroscedasticity and non-gaussianity):

1) GMM4: Gaussian mixture model with $K = 4$ regions with parameters $\boldsymbol{\mu} = \{0, 1, 2, 4\}$, $\boldsymbol{\sigma} = \{1, \sqrt{2}, \sqrt{3/2}, \sqrt{2}\}$, and spatial organisation according to a Potts MRF with $\beta = 1.2$ and size 256×256 pixels, resulting in a signal-to-noise ratio (SNR) of 7.7dB. This dataset is challenging because there is strong overlap between the distribution of the mixture components (i.e., low SNR) as well as heteroscedasticity, which our method does not take into account and hence represents a case of mild likelihood misspecification.
2) GMM8: Gaussian mixture model with $K = 8$ regions with parameters $\boldsymbol{\mu} = \{1, 2, \ldots, 8\}$, $\boldsymbol{\sigma} = \{0.3, \ldots, 0.3\}$, and spatial organisation according to a Potts MRF with $\beta = 1.5$ and size 256×256 pixels. The main challenge here is the large number of mixture components, which is further complicated by the fact that the distributions overlap partially (the SNR for this dataset is 24.1dB).
3) LMM2: Laplace mixture model with $K = 2$ components with parameters $\boldsymbol{\mu} = \{1, 2\}$, $\boldsymbol{\sigma} = \{1, 1\}$, and checkerboard spatial organisation (size 256×256 pixels), resulting in a very low SNR value of 1.0dB. This dataset is challenging because it strongly deviates from the Bayesian model considered, which is misspecified both at the level of the prior and the likelihood (note that deviations from the model can degrade significantly segmentation performance [36]). Also, both mixture components overlap significantly, making the segmentation even more difficult.
4) PMM3: Poisson mixture model with $K = 3$ components with parameter $\boldsymbol{\mu} = \{1, 6, 11\}$, and spatial organisation according to the three main structures of the Shepp-Logan phantom. This dataset has a low SNR value of 4.7dB. Again, the challenges here are the strong misspecification in the likelihood and prior, and that the mixture components overlap.

TABLE I
Segmentation Accuracy (Pixels Correctly Classified) for the Four Data Displayed in Fig. 4

	GMM4	GMM8	LMM2	PMM3
Proposed	88.9%	89.6%	96.0%	93.1%
MCMC	94.4%	98.6%	96.1%	98.1%
Iterative ICM	79.5%	18.0%	96.0%	81.5%
Non-iterative ICM	78.4%	60.4%	94.8%	95.6%

All experiments have been conducted using a MATLAB implementation of Algo. 1 with parameters $T = 50$, $L = 25$, $\epsilon = 10^{-3}$, and computed on an Intel i7 quad-core workstation running MATLAB 2014a. For the ICM algorithms we have used the MATLAB implementation of [36]. The implementation of the MCMC algorithm [20] is written in MATLAB with specific functions in C, so it has an advantage in terms of computational performance.

Table I reports the segmentation accuracy for the four test data and each method (we measure accuracy as the percentage of correctly classified pixels with respect to the ground truth). We observe that the MCMC method produced the most accurate segmentation results, with a remarkable accuracy of the order of $95\% - 99\%$, followed by the proposed

TABLE II
COMPUTING TIMES (SECONDS) FOR THE FOUR DATA DISPLAYED IN FIG. 4

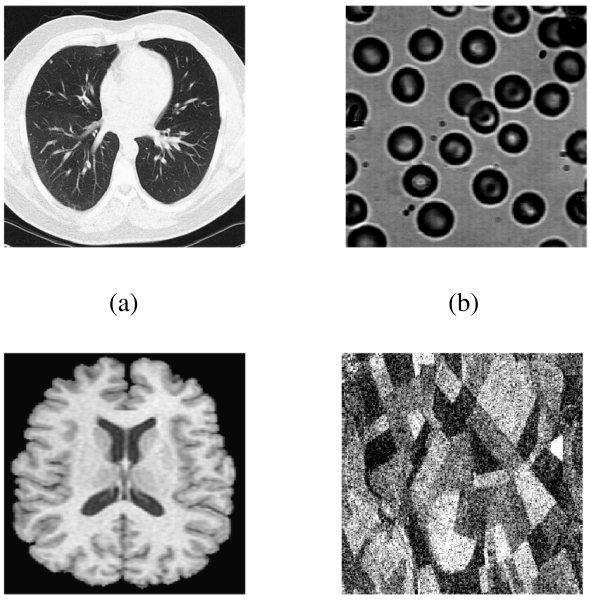
	GMM4	GMM8	LMM2	PMM3
Proposed	1.41	1.45	1.75	1.54
MCMC	208	252	76.5	73.0
Iterative ICM	113	144	30	108
Non-iterative ICM	20.4	60.4	7.37	21.0

method which achieved $90\% - 96\%$ of correctly classified pixels. Moreover, we also observe that the ICM algorithms were less accurate on average, and struggled particularly with `Data 2` because they were not able to estimate correctly the mixture model parameters. More importantly, Table II reports the computing times associated with these experiments. Observe that the proposed method is very computationally efficient and was one or two orders of magnitude faster than the ICM and MCMC approaches in all experiments (notice that these low computing times are in agreement with a large body of literature reporting that convex relaxations, combined with convex optimisation algorithms, lead to state-of-the-art computational performance). In conclusion, these experiments with synthetic data indicate that the proposed methodology offer extremely fast and accurate segmentation results. Finally, for completeness, we note that in all cases Algo. 1 converged in $t = 2$ iterations, and determined the following values for the regularisation parameter λ_ℓ: `GMM4`, $\lambda_\ell = 5.5603$; `GMM8`, $\lambda_\ell = 5.6314$; `LMM2`, $\lambda_\ell = 6.5282$; and `PMM3`, $\lambda_\ell = 4.8355$.

V. EXPERIMENTAL RESULTS AND OBSERVATIONS

In this section we demonstrate empirically the proposed methodology with a series of experiments with real data and comparisons with state-of-the-art algorithms. Similarly to Section IV, to assess the accuracy of our method we compare the results with the estimates produced by a Markov chain Monte Carlo algorithm [20], which estimates the marginal posterior of the segmentation labels $f(\boldsymbol{z}|\boldsymbol{y})$ with very high accuracy. We also report comparisons with four supervised fast image segmentation techniques that we haven chosen to represent different efficient algorithmic approaches to image segmentation (e.g. MRF energy minimisation solved by graph-cut, active contour solved by Riemannian gradient descent, and two convex models solved by convex optimisation). The specific methods used in the comparison are as follows:

- The two-stage smoothing-followed-by-thresholding algorithm (TSA) [15], which is closely related to a semi-supervised instance of Algo. 1 with a single iteration (TV-denoising followed by K-means), and with a fixed regularisation parameter λ specified by the practitioner.
- Hidden Potts MRF segmentation (7) with fixed β, solved by graph max-flow/min-cut approximation [37].
- Chan-Vese active contour by natural gradient descent [16] (to our knowledge this method is currently the fastest approach for solving active contour models).
- The fast global minimisation algorithm (FGMA) [14] for active contour models. In a similar fashion to our method, this algorithm also involves a model with a TV convex relaxation that is solved by convex optimisation.



(a) (b) (c) (d)

Fig. 5. The `Lungs` (336 × 336 pixels), `Bacteria` (380 × 380 pixels), `Brain` (256 × 256 pixels), and `SAR` (256 × 256 pixels) images used in the experiments. (a) `Lung`. (b) `Bacteria`. (c) `Brain`. (d) `SAR`.

We emphasise that, unlike the proposed method, all these efficient approaches are supervised, i.e., they require the specification of a regularisation parameters. In the experiments reported hereafter we have tuned and adjusted the parameters of each algorithm to each image by use of visual cross-validation to ensure we produce the best results for each method on each image.

To guarantee that the comparisons are fair we have applied the six algorithms considered in this paper to four images with very different characteristics: the `Lungs` and `Bacteria` images from the supplementary material of [14]; one slice of a 3D in-vivo MRI image of a human brain composed of biological tissues (white matter and grey matter) with complex shapes and textures, making the segmentation problem challenging; and a SAR image of an agricultural region in Bourges, France, containing three types of crops (note that SAR image segmentation is challenging because of the presence of strong non-Gaussian noise). The four test images are depicted in Figure 5. These images have been selected as they are composed of different types and numbers of objects; objects which have different shapes, (regular and irregular); noise characteristics; and a range of potential segmentation solutions. Again, all experiments have been conducted using a MATLAB implementation of Algo. 1 with parameters $T = 50$, $L = 25$, $\epsilon = 10^{-3}$, and computed on an Intel i7 quad-core workstation running MATLAB 2014a. With regards to the algorithms used for comparison, when possible we have used MATLAB codes made available by the respective authors. It should be noted that these are mainly MATLAB scripts, however the graph-cut method is written in C++, (the [38] implementation was used here), and the MCMC method is partly written in C, so they have an advantage in terms of computational performance.

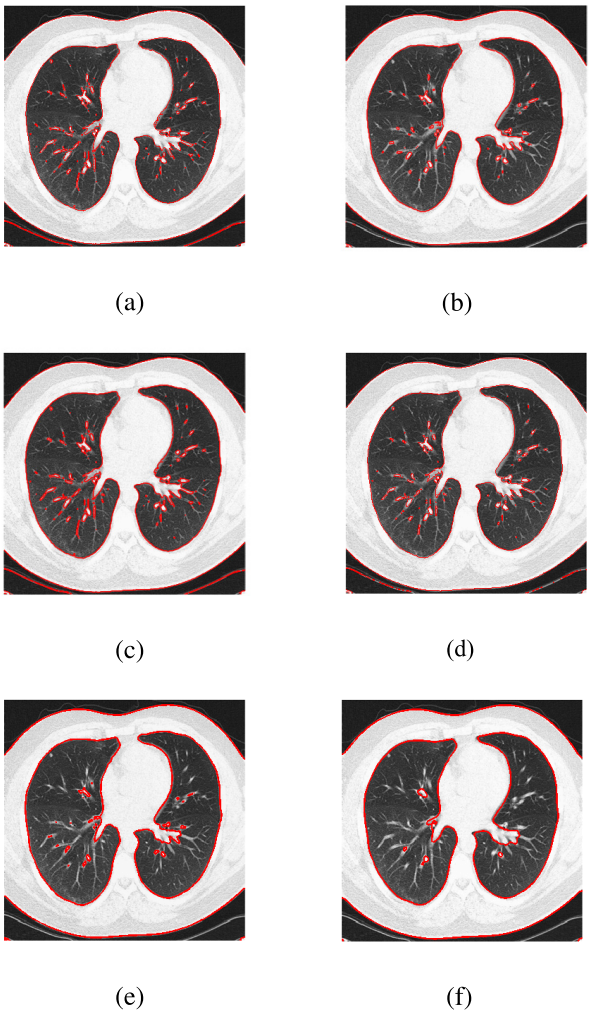


Fig. 6. Comparison with the state-of-the-art methods [14], [15], [16], and [38] using the lung image (336 × 336 pixels) from the supplementary material of [14]. (a) Proposed. (b) MCMC [20]. (c) TSA [15]. (d) Graph-Cut [38]. (e) Natural grad. [16]. (f) FGMA [14].

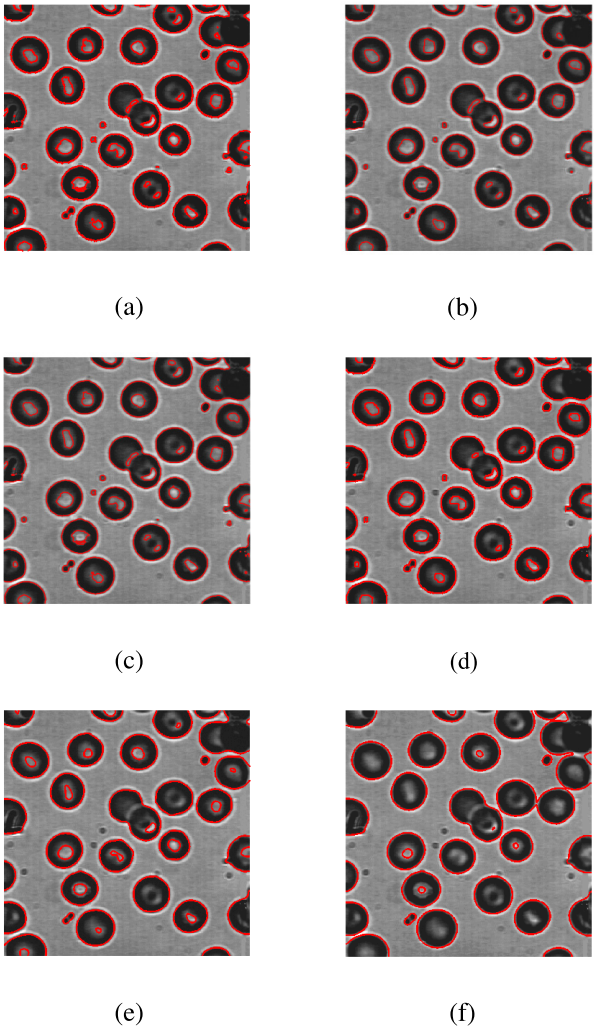


Fig. 7. Comparison of the supervised and unsupervised methods with the state of the algorithm [14], [15], [16], and [38] using the bacteria image (380 × 380 pixels) from the supplementary material of [14]. (a) Proposed. (b) MCMC [20]. (c) TSA [15]. (d) Graph-Cut [38]. (e) Natural gradient [16]. (f) FGMA [14].

We emphasise at this point that we do not seek to explicitly compare the accuracy of the methods because: 1) there is no objective ground truth; 2) the "correct" segmentation is often both subjective and application-specific; and 3) the segmentations can often be marginally improved by fine tuning the regularisation parameters. What our experiments seek to demonstrate is that our method performs similarly to the most efficient deterministic approaches of the state-of-the-art, both in terms of segmentation results and computing speed, with the fundamental advantage that it does not require specification of the value of regularisation parameters (i.e., it is fully unsupervised).

Figures 6, 7, 8 and 9 respectively show the segmentation results obtained for the `Lungs`, `Bacteria`, `Brain` and `SAR`test images with each method. The segmentations of the `Lungs` and `Bacteria` images have been computed using $K = 2$ classes to enable comparison with the natural gradient method [16] and FGMA [14] (these methods are based on an active contour model that only supports binary segmentations), whereas the `Brain` image has been computed using $K = 3$ classes to produce a clear segmentation of the grey matter and the white matter. Similarly, the `SAR` image has also been computed using $K = 3$ to identify the three crops. The computing times associated with these experiments are reported in Table III. Observe that all six methods produced similar segmentation results that are in good visual agreement with each other. In particular, we observe that the proposed method successfully determined the appropriate level of regularisation for each image and produced segmentations that are very

TABLE III

COMPUTING TIMES (SECONDS) FOR THE `Lungs`, `Bacteria` AND `Brain` IMAGES DISPLAYED IN FIGS. 6, FIGS. 7 AND FIGS. 8

	`Bacteria`	`Bacteria`	`Brain`	`SAR`
Proposed	0.65	0.80	0.23	0.23
TSA [15]	0.20	0.21	0.17	0.18
Graph-Cut [38]	0.30	0.30	0.21	0.23
Nat. grad. [16]	0.20	0.18	n/a	n/a
FGMA [14]	0.32	0.47	n/a	n/a
MCMC [16]	900	1 150	533	535

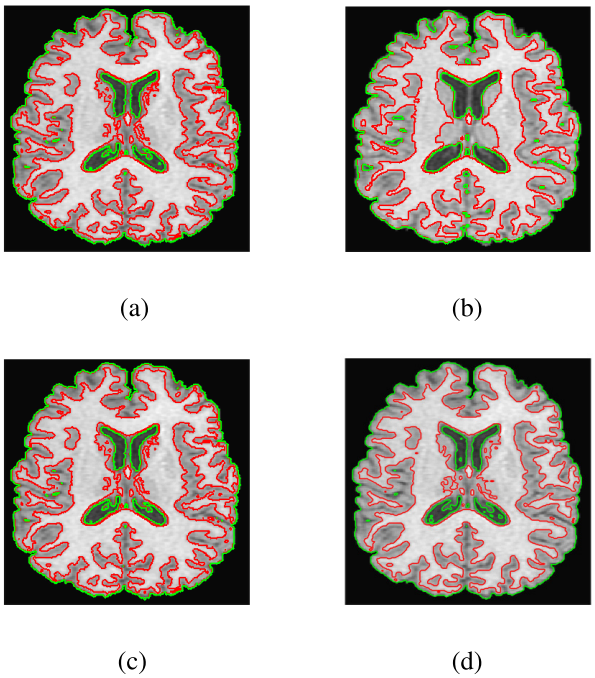


Fig. 8. Segmentation of a brain MRI image (256×256 pixels). (a) Proposed. (b) MCMC [20]. (c) TSA [15]. (d) Graph-Cut [38].

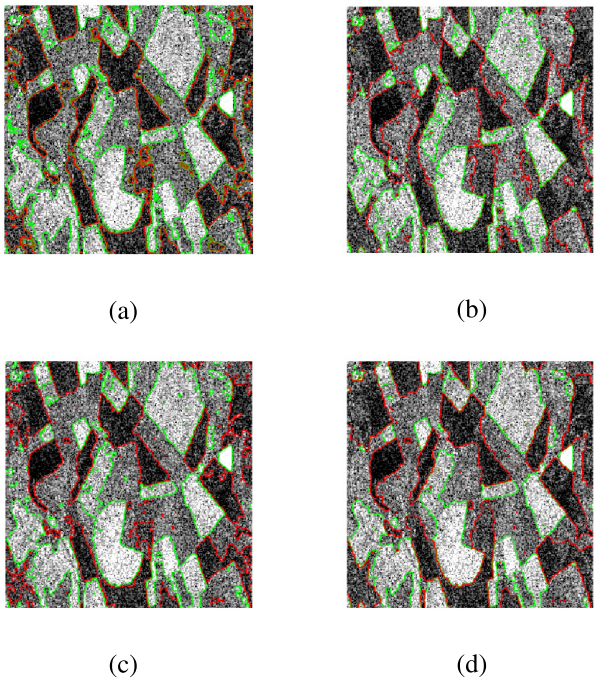


Fig. 9. Segmentation of a SAR image of an agricultural region (256×256 pixels). (a) Proposed. (b) MCMC [20]. (c) TSA [15]. (d) Graph-Cut [38].

similar to the results obtained with the supervised methods graph-cut [37] and TSA [15], and with the unsupervised MCMC algorithm [16] that in a sense represents a benchmark for these approximate inference methods. Moreover, Table III shows that the proposed method was only 2 or 3 times slower than state-of-the-art supervised approaches, which is an excellent performance for a fully unsupervised method. This additional computing time is mainly due to the additional computations related to the non-convex program (23); however, we emphasise that this algorithm has the property of adapting automatically the level of regularisation to the image, and that the computing times reported in Table III do not take into account the time involved in running the supervised algorithms repeatedly to adjust their regularisation parameters. Finally, for completeness, we note that in all cases Algo. 1 converged in $t = 2$ iterations, and determined the following values for the regularisation parameter λ_ℓ: `Lung`, $\lambda_\ell = 0.065$; `Bacteria`, $\lambda_\ell = 0.110$; `Brain`, $\lambda_\ell = 0.095$; and `SAR`, $\lambda_\ell = 6.6314$. Observe the large range of values of λ_ℓ, which highlights the fact that different images do require very different amounts of regularisation, and that the capacity of the proposed methodology to self-adjust λ_ℓ represents an important advantage over supervised approaches.

VI. Conclusions

We have presented a new fully unsupervised approach for computationally efficient image segmentation. The approach is based on a new approximate Bayesian estimator for hidden Potts-Markov random fields with unknown regularisation parameter β. The estimator is based on a small-variance-asymptotic analysis of an augmented Bayesian model and a convex relaxation combined with majorisation-minimisation technique. This estimator can be very efficiently computed by using an alternating direction scheme based on a convex total-variation denoising step and a least-squares (K-means) clustering step, both of which can be computed straightforwardly, even in large 2D and 3D scenarios, and with parallel computing techniques. Experimental results on synthetic and real images, as well as extensive comparisons with state-of-the-art algorithms showed that the resulting new image segmentation methodology performs similarly in terms of segmentation results and of computing times as the most efficient supervised image segmentation methods, with the important additional advantage of self-adjusting regularisation parameters. A detailed analysis of the theoretical properties of small-variance-asymptotic estimators in general, and in particular of the methods described in this paper, is currently under investigation. Potential future research topics include the extension of these methods to non-Gaussian statistical models from the exponential family, taking into consideration linear degradation effects such as blurring and missing pixels [39], model choice techniques to address segmentation problems where the number of classes K is unknown, applications to ultrasound and PET image segmentation, and comparisons with other Bayesian segmentation methods based on alternative hidden MRF models that can also be solved by convex optimisation, such as [8].

References

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

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

[3] 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.

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

[5] T. Vincent, L. Risser, and P. Ciuciu, "Spatially adaptive mixture modeling for analysis of fMRI time series," *IEEE Trans. Med. Imag.*, vol. 29, no. 4, pp. 1059–1074, Apr. 2010.

[6] M. P. Kumar, V. Kolmogorov, and P. H. S. Torr, "An analysis of convex relaxations for MAP estimation of discrete MRFs," *J. Mach. Learn. Res.*, vol. 10, pp. 71–106, Jan. 2009.

[7] N. Komodakis, N. Paragios, and G. Tziritas, "MRF energy minimization and beyond via dual decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 531–552, Mar. 2011.

[8] J. Bioucas-Dias, F. Condessa, and J. Kovačević, "Alternating direction optimization for image segmentation using hidden Markov measure field models," *Proc. SPIE*, vol. 9019, p. 90190P, Feb. 2014, doi: 10.1117/12.2047707.

[9] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[10] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[11] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.

[12] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 41–54, Oct. 2006.

[13] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.

[14] X. Bresson, S. Esedoğlu, P. Vandergheynst, J.-P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *J. Math. Imag. Vis.*, vol. 28, no. 2, pp. 151–167, Jun. 2007.

[15] X. Cai, R. Chan, and T. Zeng, "A two-stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 368–390, Aug. 2013.

[16] M. Pereyra, H. Batatia, and S. McLaughlin, "Exploiting information geometry to improve the convergence properties of variational active contours," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 4, pp. 700–707, Aug. 2013.

[17] M. Pereyra, H. Batatia, and S. McLaughlin, "Exploiting information geometry to improve the convergence of nonparametric active contours," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 836–845, Mar. 2015.

[18] L. Bar and G. Sapiro, "Generalized Newton-type methods for energy formulations in image processing," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 508–531, 2009.

[19] G. Sundaramoorthi, A. Yezzi, A. C. Mennucci, and G. Sapiro, "New possibilities with Sobolev active contours," *Int. J. Comput. Vis.*, vol. 84, no. 2, pp. 113–129, May 2009.

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

[21] M. Pereyra, N. Whiteley, C. Andrieu, and J.-Y. Tourneret, "Maximum marginal likelihood estimation of the granularity coefficient of a Potts–Markov random field within an MCMC algorithm," in *Proc. IEEE Workshop Statist. Signal Process. (SSP)*, Jun. 2014, pp. 121–124.

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

[23] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognit.*, vol. 36, no. 1, pp. 131–144, Jan. 2003.

[24] F. Forbes and G. Fort, "Combining Monte Carlo and mean-field-like methods for inference in hidden Markov random fields," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 824–837, Mar. 2007.

[25] T. Broderick, B. Kulis, and M. I. Jordan, "MAD-Bayes: MAP-based asymptotic derivations from Bayes," *J. Mach. Learn. Res.*, vol. 28, no. 3, pp. 226–234, 2013.

[26] A. Roychowdhury, K. Jiang, and B. Kulis, "Small-variance asymptotics for hidden Markov models," in *Advances in Neural Information Processing Systems*, vol. 26. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds. Cambridge, MA, USA: MIT Press, 2013, pp. 2103–2111.

[27] P. J. Green, "MAD-Bayes matching and alignment for labelled and unlabelled configurations," in *Geometry Driven Statistics*, I. L. Dryden and J. T. Kent, Eds. Chichester, U.K.: Wiley, 2015, ch. 19, pp. 365–375.

[28] K. Jiang, B. Kulis, and M. I. Jordan, "Small-variance asymptotics for exponential family Dirichlet process mixture models," in *Advances in Neural Information Processing Systems*, vol. 25, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. Cambridge, MA, USA: MIT Press, 2012, pp. 3167–3175.

[29] F. Y. Wu, "The Potts model," *Rev. Mod. Phys.*, vol. 54, no. 1, pp. 235–268, Jan. 1982.

[30] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1, pp. 89–97, 2004.

[31] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1. 1967, pp. 281–297.

[32] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B (Methodol.)*, vol. 48, no. 3, pp. 259–302, 1986.

[33] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.

[34] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Adaptive total variation image deblurring: A majorization–minimization approach," *Signal Process.*, vol. 89, no. 9, pp. 1683–1693, 2009.

[35] A. L. M. Levada, N. D. A. Mascarenhas, and A. Tannüs, "Pseudolikelihood equations for Potts MRF model parameter estimation on higher order neighborhood systems," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 522–526, Jul. 2008.

[36] J. Gimenez, A. C. Frery, and A. G. Flesia, "When data do not bring information: A case study in Markov random fields estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 1, pp. 195–203, Jan. 2015.

[37] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[38] S. Bagon. (Dec. 2006). *MATLAB Wrapper for Graph Cut*. [Online]. Available: http://www.wisdom.weizmann.ac.il/~bagon

[39] X. Cai, "Variational image segmentation model coupled with image restoration achievements," *Pattern Recognit.*, vol. 48, no. 6, pp. 2029–2042, Jun. 2015.

Marcelo Pereyra was born in Buenos Aires, Argentina, in 1984. He studied electronic engineering and received a double M.Eng. degree from ITBA, Argentina, and INSA Toulouse, France, together with a M.Sc. degree from INSA Toulouse, in June 2009. In July 2012 he obtained a Ph.D. degree in Signal Processing from the University of Toulouse. From 2012 to 2016, he held a Brunel Post-Doctoral Research Fellowship in Statistics, a Post-Doctoral Research Fellowship from the French Ministry of Defence, and a Marie Curie Intra-European Fellowship for Career Development at the School of Mathematics, University of Bristol, U.K. In 2017, he joined the School of Mathematical and Computer Sciences, Heriot-Watt University, as an Assistant Professor. His research interests include Bayesian statitical analysis and computation for high-dimensional inverse problems related to mathematical and computational imaging. He was awarded the Leopold Escande Ph.D. Thesis Award from the University of Toulouse (2012), an INFOTEL R&D Award from the Association of Engineers of INSA Toulouse (2009), and an ITBA R&D Award from the Buenos Aires Institute of Technology (2007).

Steve McLaughlin (F'11) was born in Clydebank, Scotland, in 1960. He received the B.Sc. degree in electronics and electrical engineering from the University of Glasgow in 1981, and the Ph.D. degree from The University of Edinburgh in 1990. From 1981 to 1984, he was a Development Engineer of Industry, where he was involved in the design and simulation of integrated thermal imaging and fire control systems. From 1984 to 1986, he was involved in the design and development of high frequency data communication systems. In 1986, he joined the Department of Electronics and Electrical Engineering, The University of Edinburgh, as a Research Fellow, where he studied the performance of linear adaptive algorithms in high noise and nonstationary environments. In 1988, he joined the Academic Staff with Edinburgh, and from 1991 to 2001 he held a Royal Society University Research Fellowship focused on nonlinear signal processing techniques. In 2011, he joined Heriot-Watt University, as a Professor of Signal Processing and Head of the School of Engineering and Physical Sciences. His research interests include the fields of adaptive signal processing and nonlinear dynamical systems theory and their applications to biomedical, energy, and communication systems. He is a fellow of the Royal Academy of Engineering, the Royal Society of Edinburgh, and the Institute of Engineering and Technology. In 2002, he was awarded the Personal Chair of Electronic Communication Systems with The University of Edinburgh.