

SMALL-VARIANCE ASYMPTOTICS OF HIDDEN POTTS-MRFs: APPLICATION TO FAST BAYESIAN IMAGE SEGMENTATION

Marcelo Pereyra*

School of Mathematics
University of Bristol
Bristol, UK
marcelo.pereyra@bristol.ac.uk

Steve McLaughlin†

School of Engineering and Physical Sciences
Heriot-Watt University
Edinburgh, UK
s.mclaughlin@hw.ac.uk

ABSTRACT

This paper presents a new approximate Bayesian estimator for hidden Potts-Markov random fields, with application to fast K -class image segmentation. The estimator is derived by conducting a small-variance-asymptotic analysis of an augmented Bayesian model in which the spatial regularisation and the integer-constrained terms of the Potts model are decoupled. This leads to a new image segmentation methodology that can be efficiently implemented in large 2D and 3D scenarios by using modern convex optimisation techniques. Experimental results on synthetic and real images as well as comparisons with state-of-the-art algorithms confirm that the proposed methodology converges extremely fast and produces accurate segmentation results in only few iterations.

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

1. 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 conveniently 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 propose a small-variance asymptotics estimator for hidden Potts-Markov random fields (MRFs), a powerful class of statistical models that is widely used in Bayesian image segmentation methods (see [1–3] for applications to hyperspectral, ultrasound and fMRI imaging). Despite their wide application,

*Marcelo Pereyra gratefully acknowledges funding from the French Ministry of Defence, from the SuSTaIN program - EPSRC grant EP/D063485/1 - at the Department of Mathematics of the University of Bristol, and from the DynBrain project supported by the 2011 STIC-AMSUD program.

†Steve McLaughlin gratefully acknowledges funding from EPSRC platform grant EP/J015180/1

performing inference on hidden Potts MRFs remains a computationally challenging problem and thus most image processing methods compute approximate estimators obtained via Monte Carlo approximations, mean-field approximations, local optimisation approximations, etc.

Small-variance asymptotics estimators were introduced in [4] 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 [5]. Here we exploit these same techniques for the hidden Potts MRF to develop a fast converging image segmentation methodology that delivers accurate segmentation results in very few iterations. 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 description of our proposed method and a series of experiments where the method is applied to some standard example images. Finally some brief conclusions are drawn in Section V.

2. 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. For simplicity we only consider univariate Gaussian mixture models. However, the results presented hereafter can easily 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 [6].

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

$$\mathbf{y}_n | n \in \mathcal{C}_k \sim \mathcal{N}(\mu_k, \sigma^2) \quad (1)$$

where $\mu_k \in \mathbb{R}$ represents the mean intensity of class \mathcal{C}_k . We

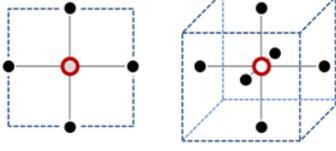


Fig. 1. 4-pixel (left) and 6-voxel (right) neighborhood structures. The pixel/voxels considered appears as a void red circle whereas its neighbors are depicted in full black and blue.

assume that $\mu_k \neq \mu_j$ for all $k \neq j$.

To perform segmentation, a label vector $\mathbf{z} = (z_1, \dots, z_N)^T$ is introduced to map or classify observations \mathbf{y} to classes $\mathcal{C}_1, \dots, \mathcal{C}_K$ (i.e., $z_n = k$ if and only if $n \in \mathcal{C}_k$).

Assuming that observations are conditionally independent given \mathbf{z} and given the parameter vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)$, the likelihood of \mathbf{y} is given by

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

with $\mathcal{S}_k = \{n : z_n = k\}$.

The prior for \mathbf{z} is the homogenous K -state Potts MRF [7]

$$f(\mathbf{z}) = \frac{1}{C(\beta)} \exp[\beta H(\mathbf{z})] \quad (3)$$

with fixed hyper-parameter $\beta \in \mathbb{R}^+$ and Hamiltonian

$$H(\mathbf{z}) = \sum_{n=1}^N \sum_{n' \in \mathcal{V}(n)} \delta(z_n \neq z_{n'}) \quad (4)$$

where $\mathcal{V}(n)$ is the index set of the neighbors of the n th voxel and $\delta(\cdot)$ is the Kronecker function. Here 2D MRFs are considered as prior distribution for \mathbf{z} for single-slice (2D) images and 3D MRFs for multiple-slice (3D) images (the corresponding neighborhoods $\mathcal{V}(n)$ are depicted in Fig. 1). Similarly, the class means are assigned Gaussian priors $\mu_k \sim \mathcal{N}(0, \rho^2)$ with fixed variance ρ^2 independent of \mathbf{z}

$$f(\boldsymbol{\mu}) = \prod_{k=1}^K p_{\mathcal{N}}(\mu_k | 0, \rho^2). \quad (5)$$

Then, using Bayes theorem, the posterior distribution of $(\mathbf{z}, \boldsymbol{\mu})$ given \mathbf{y} can be expressed as follows

$$f(\mathbf{z}, \boldsymbol{\mu} | \mathbf{y}) \propto f(\mathbf{y} | \mathbf{z}, \boldsymbol{\mu}) f(\mathbf{z}) f(\boldsymbol{\mu}) \quad (6)$$

where \propto means “proportional to”.

A segmentation of \mathbf{y} is typically obtained by computing the joint maximum-a-posteriori (MAP) estimator

$$\hat{\mathbf{z}}_1, \hat{\boldsymbol{\mu}}_1 = \underset{\mathbf{z}, \boldsymbol{\mu}}{\operatorname{argmax}} f(\mathbf{z}, \boldsymbol{\mu} | \mathbf{y}) \quad (7)$$

which can also be obtained by solving the equivalent optimisation problem

$$\hat{\mathbf{z}}_1, \hat{\boldsymbol{\mu}}_1 = \underset{\mathbf{z}, \boldsymbol{\mu}}{\operatorname{argmin}} -\log f(\mathbf{z}, \boldsymbol{\mu} | \mathbf{y}). \quad (8)$$

Unfortunately these optimisation problems are known to be NP-hard due to the combinatorial nature of the Potts Hamiltonian $H(\mathbf{z})$. Modern image segmentation methods typically address this issue by using stochastic simulation or optimisation algorithms (e.g., MCMC, simulated annealing) [8], local optimisation algorithms (e.g., EM, iterated conditional models) [9].

3. PROPOSED METHOD

This section presents a new approach for performing inference on \mathbf{z} based on a small-variance asymptotic analysis combined with a convex relaxation of the Potts MRF. This approximation will lead to a new estimator of \mathbf{z} that can be computed very efficiently using modern convex optimisation techniques.

We begin by introducing a carefully selected auxiliary vector \mathbf{x} such that \mathbf{y} and $(\mathbf{z}, \boldsymbol{\mu})$ are conditionally independent given \mathbf{x} , and that the posterior $f(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu} | \mathbf{y})$ has the same maximisers as (6) (after projection on the space of $(\mathbf{z}, \boldsymbol{\mu})$). More precisely, we define a random vector $\mathbf{x} = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ with degenerate prior

$$f(\mathbf{x} | \mathbf{z}, \boldsymbol{\mu}) = \prod_{k=1}^K \prod_{n \in \mathcal{S}_k} \delta(x_n - \mu_k) \quad (9)$$

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

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

The prior distributions for \mathbf{z} and $\boldsymbol{\mu}$ remain as defined in (3) and (5). The posterior distribution of $\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}$ is given by

$$\begin{aligned} f(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu} | \mathbf{y}) &\propto f(\mathbf{y} | \mathbf{x}) f(\mathbf{x} | \mathbf{z}, \boldsymbol{\mu}) f(\mathbf{z}) f(\boldsymbol{\mu}) \\ &\propto \left(\prod_{k=1}^K \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n | x_n, \sigma^2) \delta(x_n - \mu_k) \right) \\ &\quad \times \exp[\beta H(\mathbf{z})] f(\boldsymbol{\mu}). \end{aligned} \quad (10)$$

Moreover, we define as $H^*(\mathbf{z})$ the “complement” of the Hamiltonian $H(\mathbf{z})$ in the sense that for any \mathbf{z}

$$H(\mathbf{z}) + H^*(\mathbf{z}) = N|\mathcal{V}|$$

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

$$H^*(\mathbf{z}) \triangleq \sum_{n=1}^N \sum_{n' \in \mathcal{V}(n)} \delta(z_n \neq z_{n'}). \quad (11)$$

Replacing $H(\mathbf{z}) = N|\mathcal{V}|-H^*(\mathbf{z})$ in (10) we obtain

$$f(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}|\mathbf{y}) \propto \left(\prod_{k=1}^K \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n|x_n, \sigma^2) \delta(x_n - \mu_k) \right) \times \exp[-\beta H^*(\mathbf{z})] f(\boldsymbol{\mu}). \quad (12)$$

Note that (12) produces the same MAP segmentation of \mathbf{y} as (8) in the sense that the estimates

$$\hat{\mathbf{x}}_2, \hat{\mathbf{z}}_2, \hat{\boldsymbol{\mu}}_2 = \underset{\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}}{\operatorname{argmin}} -\log f(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}|\mathbf{y})$$

verify $\hat{\mathbf{z}}_2 = \hat{\mathbf{z}}_1$ and $\hat{\boldsymbol{\mu}}_2 = \hat{\boldsymbol{\mu}}_1$, and that $\hat{\mathbf{x}}_2$ is perfectly determined by $\hat{\mathbf{z}}_2, \hat{\boldsymbol{\mu}}_2$ through (9).

Furthermore, noting that $H^*(\mathbf{z})$ only measures if neighbour labels are identical or not, regardless of their values, it is easy to check that the posterior (12) remains unchanged if we substitute $H^*(\mathbf{z})$ with $H^*(\mathbf{x})$

$$f(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}|\mathbf{y}) \propto \left(\prod_{k=1}^K \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n|x_n, \sigma^2) \delta(x_n - \mu_k) \right) \times \exp[-\beta H^*(\mathbf{x})] f(\boldsymbol{\mu}). \quad (13)$$

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

$$f(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}|\mathbf{y}) \propto \left(\prod_{k=1}^K \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n|x_n, \sigma^2) \delta(x_n - \mu_k) \right) \times \exp[-2\beta \|\nabla \mathbf{x}\|_0] f(\boldsymbol{\mu}). \quad (14)$$

We are now ready to conduct a small-variance asymptotic analysis on (14) and derive the MAP asymptotic estimator of $\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}$, which is defined as [4]

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

First, we use the fact that $\delta(s) = \lim_{\tau^2 \rightarrow 0} p_{\mathcal{N}}(s|0, \tau^2)$ to express (14) as follows

$$\begin{aligned} f(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}|\mathbf{y}) &\propto \lim_{\tau^2 \rightarrow 0} \left(\prod_{k=1}^K \prod_{n \in \mathcal{S}_k} p_{\mathcal{N}}(y_n|x_n, \sigma^2) p_{\mathcal{N}}(x_n|\mu_k, \tau^2) \right) \\ &\quad \times \exp[-\beta \|\nabla \mathbf{x}\|_0] f(\boldsymbol{\mu}) \\ &\propto \lim_{\tau^2 \rightarrow 0} \left(\prod_{k=1}^K \prod_{n \in \mathcal{S}_k} \exp\left(-\frac{(x_n - y_n)^2}{2\sigma^2} - \frac{(x_n - \mu_k)^2}{2\tau^2}\right) \right) \\ &\quad \times \exp[-\beta \|\nabla \mathbf{x}\|_0] f(\boldsymbol{\mu}) \end{aligned} \quad (15)$$

Then, in a manner akin to Broderick et al. [4], we allow the model's hyper parameters to scale with σ in order to preserve the balance between the prior and the likelihood and avoid a trivial limit. Precisely, we set $\beta = \beta'/2\sigma$ and assume that σ vanishes at the same rate as τ . Then, the limit of $-\sigma^2 \log f(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}|\mathbf{y})$ as $\sigma^2 \rightarrow 0$ is given by

$$\lim_{\sigma^2 \rightarrow 0} -\sigma^2 \log f(\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}|\mathbf{y}) = \sum_{k=1}^K \sum_{n \in \mathcal{S}_k} \frac{1}{2} (x_n - y_n)^2 + \frac{1}{2} (x_n - \mu_k)^2 + \beta' \|\nabla \mathbf{x}\|_0 \quad (16)$$

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

$$\underset{\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}}{\operatorname{argmin}} \sum_{k=1}^K \sum_{n \in \mathcal{S}_k} \frac{1}{2} (x_n - y_n)^2 + \frac{1}{2} (x_n - \mu_k)^2 + \beta' \|\nabla \mathbf{x}\|_0. \quad (17)$$

Unfortunately (17) is still NP-hard due to $\|\nabla \mathbf{x}\|_0$. However, unlike (8), (17) can be easily approximated by a convex optimisation problem (in \mathbf{x}) and thus efficiently solved using state-of-the-art convex optimisation methods. More precisely, we replace $\|\nabla \mathbf{x}\|_0$ by its convexification $\|\nabla \mathbf{x}\|_1$ and propose the following estimators of $\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}$

$$\hat{\mathbf{x}}_3, \hat{\mathbf{z}}_3, \hat{\boldsymbol{\mu}}_3 = \underset{\mathbf{x}, \mathbf{z}, \boldsymbol{\mu}}{\operatorname{argmin}} \sum_{k=1}^K \sum_{n \in \mathcal{S}_k} \frac{1}{2} (x_n - y_n)^2 + \frac{1}{2} (x_n - \mu_k)^2 + \beta' \operatorname{TV}(\mathbf{x}) \quad (18)$$

where $\operatorname{TV}(\mathbf{x}) = \|\nabla \mathbf{x}\|_1$ is the total-variation norm of \mathbf{x} . These estimates can be very efficiently computed by iteratively minimising (18) w.r.t. \mathbf{x}, \mathbf{z} and $\boldsymbol{\mu}$. The minimisation w.r.t. \mathbf{x} is equivalent to a 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 FISTA on the dual problem [10]). Similarly, the joint minimisation w.r.t. \mathbf{z} and $\boldsymbol{\mu}$ is equivalent to performing a K-means clustering on \mathbf{x} that can also be (approximately) solved very efficiently in high dimensions.

4. RESULTS

This section presents two experiments conducted to assess the performance of the proposed image segmentation methodology. All experiments have been computed on an Intel i7 quad-core workstation running MATLAB 2013a.

In the first experiment we compare our method with three algorithms from the state of the art: the Chan-Vese active contour by natural gradient descent [11], the Chan-Vese active contour by generalised Newton descent, and the two-stage

convex optimisation and thresholding method [12]. To guarantee that the comparisons are fair we have used a synthetic shape image from [13], which has also been used in [11], and the accompanying MATLAB codes. This image of size 216×187 pixels contains 3 objects with well defined boundaries and is contaminated with white additive Gaussian noise of 5.36dB SNR. For this experiment, we implemented our method using $K = 2$ and $\beta' = 0.25$, and the Two-stage algorithm using $K = 2$ and $\lambda = 40$ as these values produced the best results. For the natural gradient and generalised Newton methods we have used $\sigma = 0.75$ and $\sigma = 1$ respectively as in [11] and [13]. The methods from the state of the art require initialising an active contour function which we have set to a right circular cone as recommended in [13] and [11]. Our method has been initialised by setting $\mathbf{x}^{(0)} = \mathbf{y}$.

Fig. 2 depicts in red the contours estimated with each method (for the methods based on active contours the initialisation is depicted in blue). Fig. 2(a) shows the segmentation obtained with the proposed method. The results obtained with the natural gradient descent [11], the Two-stage algorithm [12] and the generalised Newton method [13] are presented in Figures 2(b), 2(c) and 2(d). We observe that all four methods produced very accurate segmentation results (note however that many other methods from the state-of-the-art produce segmentation results that are significantly less accurate, see the experiments in [11, 13]). More importantly, Table 1 shows that the proposed method converged in only 2 iterations (0.22 seconds), closely followed by the natural gradient descent (2 iterations, 0.24 seconds), which is the fastest state-of-the-art method for this type of two-class image segmentation problems [11]. These two methods were 3 times faster than the Two-stage algorithm, which required 13 split-Bregman iterations and 0.71 seconds to converge, and 25 times faster than generalised Newton method, which required 30 iterations and 5.77 seconds. Additional segmentation results obtained with other state-of-the-art methods are provided in [11, 13].

Table 1. Experiment 1: Convergence and computing times.

	Iterations	Comp. time (sec)
Proposed method	2	0.22
Natural gradient [11]	2	0.24
Two-stage algorithm [12]	13	0.71
Generalised Newton [13]	30	5.77

An important property of the proposed method is that it can be applied to problems with more than two classes (as opposed to the methods from [11, 13] which have been specifically designed for $K = 2$). The second experiment applies the proposed method to one slice of a 3D in-vivo MRI image of a human brain, which is depicted in Fig. 3(a). Note that this image is composed by three biological tissues (cerebro spinal fluid, white matter, and grey matter) with complex shapes and

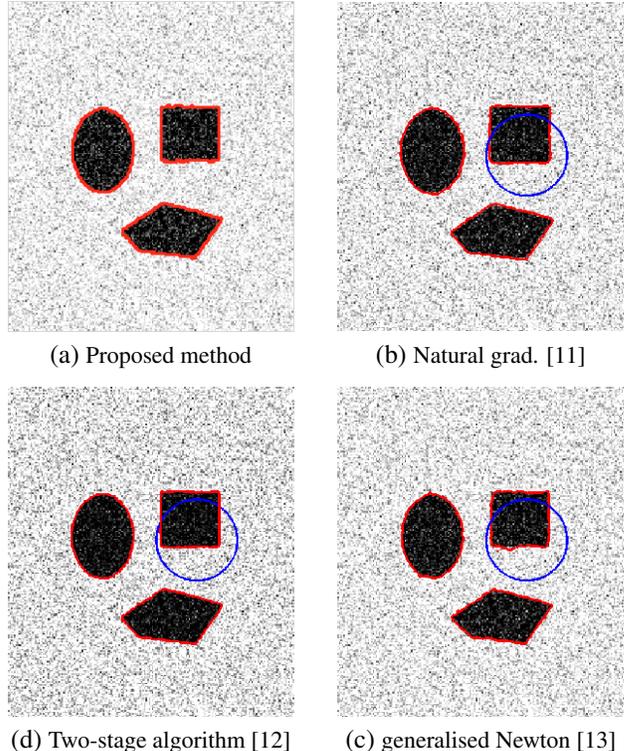


Fig. 2. Comparison with the state of the methods [11], [12] and [13] using the synthetic shape image from [13] (216×187 pixels, additive Gaussian noise, SNR 5.36 dB).

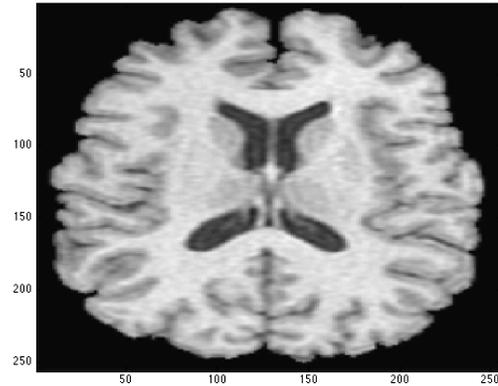
textures, making the segmentation problem challenging. The segmentation obtained with our method using $\beta' = 1.0$ and $K = 2$ is depicted in Fig. 3(b), and with $\beta' = 1.0$ and $K = 3$ in Fig. 3(c). This result has been computed in only 2 iterations (0.23 seconds) for $K = 2$ and 2 iterations (0.26 seconds) for $K = 3$, confirming that the proposed algorithm can produce accurate segmentation results in very few iterations.

5. CONCLUSIONS

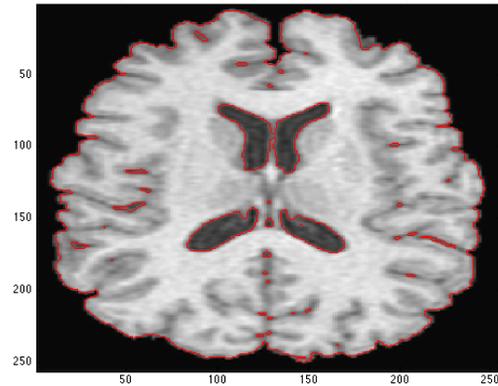
We presented a new approximate Bayesian estimator for hidden Potts-Markov random fields based on a small-variance-asymptotic analysis of an augmented Bayesian model and a convex relaxation. The estimator can be very efficiently computed by using an alternating direction scheme based on a total-variation denoising step and a K-means clustering. This leads to a new image segmentation methodology that converges extremely fast and produces accurate segmentation results in only few iterations. A detailed analysis of the proposed estimator and of the proposed optimisation scheme is currently under investigation. Perspectives for future work include the estimation of β' jointly with the other unknown parameters of the model as in [8], as well as a comparison with other Bayesian segmentation methods based on convex models and graph-cut algorithms [14].

References

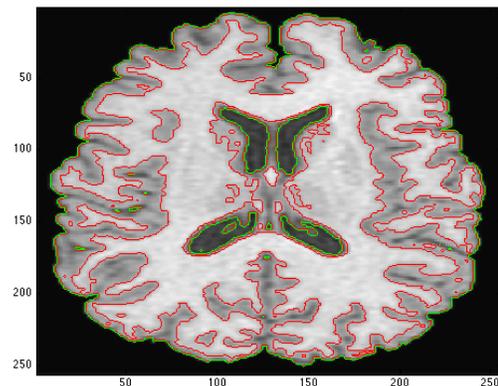
- [1] O. Eches, N. Dobigeon, and J.-Y. Tournet, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011.
- [2] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tournet, "Segmentation of skin lesions in 2D and 3D ultrasound images using a spatially coherent generalized Rayleigh mixture model," *IEEE Trans. Med. Imag.*, vol. 31, no. 8, pp. 1509–1520, Aug. 2012.
- [3] 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, April 2010.
- [4] T. Broderick, B. Kulis, and M. I. Jordan, "MAD-Bayes: MAP-based Asymptotic Derivations from Bayes," *ArXiv e-prints*, Dec. 2012.
- [5] A. Roychowdhury, K. Jiang, and B. Kulis, "Small-variance asymptotics for hidden markov models," in *Advances in Neural Information Processing Systems 26*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds., 2013, pp. 2103–2111. [Online]. Available: http://media.nips.cc/nipsbooks/nipspapers/paper_files/nips26/1045.pdf
- [6] 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 25*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., 2012, pp. 3167–3175.
- [7] F. Y. Wu, "The Potts model," *Rev. Mod. Phys.*, vol. 54, no. 1, pp. 235–268, Jan. 1982.
- [8] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tournet, "Estimating the granularity parameter of a Potts-Markov random field within an MCMC algorithm," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2385–2397, June 2013. [Online]. Available: <http://pereyra.perso.enseiht.fr/pdf/PereyraIEETIPtr2012.pdf>
- [9] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131 – 144, Jan. 2003.
- [10] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Img. Sci.*, vol. 2, no. 1, pp. 183–202, March 2009.
- [11] M. Pereyra, H. Batatia, and S. McLaughlin, "Exploiting information geometry to improve the convergence properties of variational active contours," *IEEE J. Sel. Topics Signal Processing*, vol. 7, no. 4, pp. 1–8, Aug. 2013.
- [12] T. Z. Xiaohao Cai, Raymond H. Chan, "A two-stage image segmentation method using a convex variant of the mumford-shah model and thresholding," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 368–390, Aug. 2013.
- [13] L. Bar and G. Sapiro, "Generalized newton-type methods for energy formulations in image processing," *SIAM J. Imaging Sciences*, vol. 2, no. 2, pp. 508–531, 2009.
- [14] M. Figueiredo, D. Cheng, and V. Murino, "Clustering under prior knowledge with application to image segmentation," in *Neural Information Processing Systems - NIPS'2006*, 2006.



(a) Brain MRI



(b) Proposed method



(c) Proposed method

Fig. 3. Segmentation of a brain MRI image (256×256 pixels).