BAYESIAN IMAGE SEGMENTATION USING HIDDEN FIELDS: SUPERVISED, UNSUPERVISED, AND SEMI-SUPERVISED FORMULATIONS

José M. Bioucas-Dias and Mário A. T. Figueiredo

Instituto de Telecomunicações and Instituto Superior Técnico,
Universidade de Lisboa, Portugal

ABSTRACT

Segmentation is one of the central problems in image analysis, where the goal is to partition the image domain into regions exhibiting some sort of homogeneity. Most often, the partition is obtained by solving a combinatorial optimization problem, which is, in general, NP-hard. In this paper, we follow an alternative approach, using a Bayesian formulation based on a set of hidden real-valued random fields, which condition the partition. This formulation yields a continuous optimization problem, rather than a combinatorial one. In the supervised case, this problem is convex, and we tackle it with an instance of the *alternating direction method of multipliers* (ADMM). In the unsupervised and semi-supervised cases, the optimization problem is nonconvex, and we address it using an *expectation-maximization* (EM) algorithm, where the M-step is implemented via ADMM. The effectiveness and flexibility of the proposed approach is illustrated with experiments on simulated and real data.

Index Terms— Image segmentation, hidden fields, expectation maximization, alternating direction method of multipliers (ADMM).

1. INTRODUCTION

The goal of image segmentation is to partition an image into regions that are "homogeneous". Since the notion of homogeneity is highly problem-dependent, image segmentation is a vast area of research, which has been the focus of a very large amount of work by the computer vision and image analysis communities. Moreover, image segmentation is almost invariably an ill-posed inverse problem, requiring some form of regularization (prior knowledge, in Bayesian terms) to be imposed on the solution, with the objective of promoting "desirable" solutions. Naturally, the definition of desirable solutions is highly problem-dependent. Image segmentation plays a central role in many applications, such as, remote sensing [1], computer vision [2], and medical imaging.

If approached with Bayesian tools, image segmentation is often formulated as a *maximum a posteriori* (MAP) estimate of the partition, *i.e.*, that which maximizes the product of the likelihood function (the probability of the observed image given the partition) and the prior probability the partition, usually expressed via a *Markov random field* (MRF) [3]. In a variational framework (*e.g.*, active contours/snakes, geodesic active contours, level sets [4, 5]), images are segmented by minimizing the sum of data misfit terms and regularization terms. In graph-based methods [6], image segmentation is formulated as a graph partition problem, where the regularization is implicit in the definition of the partition.

A natural representation of a discrete image segmentation is as an image of labels, each indicating to which segment the corresponding pixel belongs. With this representation, MAP and variational segmentation correspond to combinatorial optimization problems, which, apart from a few exceptions, are NP-hard, thus impractical to solve exactly. In the last decade, several powerful approximations have been introduced, such as those based on graph cuts or convex relaxations of the original combinatorial problems (see [5] for a comprehensive review).

1.1. Contributions

In this paper, inspired by the "hidden Markov measure fields" introduced by Marroquin *et al* [7], we sidestep the hurdles raised by the combinatorial approach to image segmentation by: (a) adopting a Bayesian framework, (b) introducing a set of hidden real-valued fields, conditioning the probability of the partitions, and (c) adopting a suitable prior for the hidden fields. Armed with this model, we compute the *marginal MAP* (MMAP, by marginalizing with respect to the segmentation) estimate of the hidden fields. For the prior, we adopt *vectorial total variation* (VTV) [8], which promotes piece-wise smooth vector fields, with coordinated preservation of discontinuities. We consider supervised, unsupervised, and semi-supervised scenarios. The supervised case leads to a convex program, which we tackle using an instance of ADMM called CSALSA [9]. In the un/semi-supervised cases, the resulting problem is non-convex and we address it using EM. From the MMAP estimate of the hidden fields, both soft or hard segmentations may be trivially obtained.

1.2. Related work

The work in [10, 11] also approaches image segmentation using the hidden fields paradigm, but with a key difference with respect to [7]. Whereas [7] uses hidden measure fields, *i.e.*, each element is a probability distribution over segments, thus

This work was partially supported by the (Portuguese) *Fundacão para a Ciência e Tecnologia* (FCT), grant UID/EEA/5008/2013.

under non-negativity and sum-to-one constraints, [10, 11] use a collection of unconstrained real-valued fields, which express the local probability distribution of the segments via a logistic link; the drawback is that, even in the supervised case, the resulting optimization problem is non-convex.

A number of convex relaxations for the combinatorial optimization formulations have been introduced in recent years (see [5] for a comprehensive review). In these relaxations, the objective is to obtain solutions close to those of the original discrete optimization. As seen later, this contrasts with our approach, where the objective is to compute probabilities of the partitions, expressed in the hidden field. This opens the door to statistical inference, *e.g.*, of the model parameters in the unsupervised scenario, which are not directly accessible in the above relaxations.

1.3. Paper Organization

Section 2 presents the proposed formulation and inference criterion. Sections 3 and 4 present the algorithms proposed for the supervised and un/semi-supervised cases, respectively. Section 5 reports experimental results, and Section 6 presents concluding remarks and pointers to future work.

2. PROBLEM FORMULATION

Let $\mathcal{S} \equiv \{1, \cdots, n\}$ index the n pixels of an image and $\mathbf{x} \equiv [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ be a $d \times n$ matrix of d-dimensional feature vectors. Given $\mathbf{x}$, image segmentation aims at finding a partition $P \equiv \{R_1, \dots, R_K\}$ of $\mathcal{S}$ such that the feature vectors with indices in a given set R_i, for $i = 1, \dots, K$, are *similar* in some sense. A partition P is equivalent to an image of labels $\mathbf{y} \equiv (y_1, \cdots, y_n) \in \mathcal{L}^n$, where $\mathcal{L} \equiv \{1, \dots, K\}$, such that $y_i = k$ if and only if $i \in R_k$.

2.1. MAP Segmentation

The MAP segmentation is given by

$$\widehat{\mathbf{y}}_{\text{MAP}} \in \arg\max_{\mathbf{y} \in \mathcal{L}^n} \big(\log p(\mathbf{x}|\mathbf{y}) + \log p(\mathbf{y})\big), \tag{1}$$

where $p(\mathbf{x}|\mathbf{y})$ is the observation model (*i.e.*, probability of observing the collection of features $\mathbf{x}$, given segmentation $\mathbf{y}$) and $p(\mathbf{y})$ is the prior probability of segmentation $\mathbf{y}$. A common assumption is that of *conditional independence* [3], *i.e.*,

$$p(\mathbf{x}|\mathbf{y}) = \prod_{i \in \mathcal{S}} p(\mathbf{x}_i|y_i) = \prod_{k=1}^{K} \prod_{i \in R_k} p_k(\mathbf{x}_i), \tag{2}$$

assuming (as is also common) that the probability function of a feature vector depends only on the segment to which it belongs: $p_k(\mathbf{x}_i) \equiv p(\mathbf{x}_i|y_i = k)$. For now, consider the *supervised* scenario: the class-conditional probability functions p_k are known (maybe previously learned from a training set). Later, we discuss the *unsupervised* and *semi-supervised* cases, where these functions depend on unknown parameters to be learned from the image being segmented.

Various forms of MRF have been used as prior $p(\mathbf{y})$. A paradigmatic example is the *multilevel logistic* (MLL) [3], which promotes coherent segmentations, *i.e.*, such that neighboring labels are more probably of the same class than not.

The MAP criterion in (1) is a combinatorial optimization problem. For MLL priors and $K = 2$, the problem can be mapped into that of computing a minimum cut (min-cut) on a suitable graph [12], for which efficient algorithms exist. However, for $K > 2$, the problem is NP-hard, thus intractable to solve exactly. In the past decade, several algorithms have been proposed to approximate $\widehat{\mathbf{y}}_{\text{MAP}}$, of which we highlight the graph-cut-based α-expansion, *sequential tree-reweighted message passing* (TRW-S), *loopy belief propagation* (LBP), and various convex relaxations; see [13], for a comprehensive review and comparison of these methods.

2.2. Hidden Fields and MMAP Segmentation

MAP image segmentation in terms of class labels $\mathbf{y}$ raises difficulties, namely: the computational complexity resulting from the combinatorial nature of the problem; learning unknown model parameters (both of the prior and of the observation model). These difficulties have stimulated research in several fronts.

An alternative approach, pioneered in [7, 10], reformulates the original segmentation problem by introducing a hidden field $\mathbf{z}$ of continuous variables, conditioning $\mathbf{y}$. The *marginal MAP* (MMAP) estimate of $\mathbf{z}$ is then computed, which corresponds to a *soft segmentation*. This approach avoids combinatorial optimization, replacing it with an unconstrained continuous problem [10, 11], or a constrained convex one (thus efficiently solvable [14]).

Let $\mathbf{z} = [\mathbf{z}_1, \dots, \mathbf{z}_n] \in \mathbb{R}^{K \times n}$ denote a matrix of (hidden) random vectors, such that each label y_i depends on the corresponding $\mathbf{z}_i$ (in a conditional independent way),

$$p(\mathbf{y}|\mathbf{z}) = \prod_{i \in \mathcal{S}} p(y_i|\mathbf{z}_i). \tag{3}$$

Combining (3) with the conditional independence assumption in (2), allows writing

$$p(\mathbf{x}|\mathbf{z}) = \prod_{i \in \mathcal{S}} p(\mathbf{x}_i|\mathbf{z}_i), \tag{4}$$

where each $p(\mathbf{x}_i|\mathbf{z}_i)$ is obtained by marginalizing $p(\mathbf{x}_i, y_i|\mathbf{z}_i) = p(\mathbf{x}_i|y_i)\, p(y_i|\mathbf{z}_i)$, with respect to the segment label y_i:

$$p(\mathbf{x}_i|\mathbf{z}_i) = \sum_{y_i \in \mathcal{L}} p(\mathbf{x}_i|y_i)\, p(y_i|\mathbf{z}_i). \tag{5}$$

Finally, the MMAP estimate of the hidden field $\mathbf{z}$ is given by

$$\widehat{\mathbf{z}}_{\text{MMAP}} \in \arg\max_{\mathbf{z} \in \mathbb{R}^{K \times n}} \left(\log p(\mathbf{z}) + \sum_{i \in \mathcal{S}} \log p(\mathbf{x}_i|\mathbf{z}_i)\right), \tag{6}$$

from which the *soft segmentation* $p(\mathbf{y}|\widehat{\mathbf{z}}_{\text{MMAP}})$ may be obtained. Hard segmentation may be done (pixel-wise) via

$$\widehat{y}_i \in \arg\max_{y_i \in \mathcal{L}} \; p\big(y_i|(\widehat{\mathbf{z}}_{\text{MMAP}})_i\big),$$

where $(\mathbf{a})_i$ denotes the i-th component of vector $\mathbf{a}$.

2.3. Link Between Hidden Field and Labels

The conditional probabilities $p(y_i|\mathbf{z}_i)$ play a central role in hidden field approaches. As in [7], we adopt the simple model

$$p(y_i = k|\mathbf{z}_i) = \big(\mathbf{z}_i\big)_k, \quad i \in \mathcal{S}, \quad k \in \mathcal{L}. \tag{7}$$

Given that each $\mathbf{z}_i$ is a probability distribution, it satisfies the non-negativity constraint $\big(\mathbf{z}_i\big)_k \geq 0$, for $k \in \mathcal{L}$, and the sum-to-one constraint $\mathbf{1}^T\mathbf{z}_i = \mathbf{1}$, where $\mathbf{1}$ denotes generically a column vector of ones with appropriate dimension.

Inserting (7) into (5) allows writing $p(\mathbf{x}_i|\mathbf{z}_i) = \mathbf{p}_i^T\mathbf{z}_i$, where $\mathbf{p}_i \equiv [p(\mathbf{x}_i|y_i = 1), \ldots, p(\mathbf{x}_i|y_i = K)]^T$. Finally, denoting $\phi(\mathbf{z}) = -\log p(\mathbf{z})$, MMAP estimation of $\mathbf{z}$ corresponds to solving the problem

$$\begin{aligned} \min_{\mathbf{z}\in\mathbb{R}^{K\times n}} \quad & \phi(\mathbf{z}) - \sum_{i\in\mathcal{S}} \log(\mathbf{p}_i^T\mathbf{z}_i) \\ \text{subject to:} \quad & \mathbf{z} \geq 0, \;\; \mathbf{1}^T\mathbf{z} = \mathbf{1}^T, \end{aligned} \tag{8}$$

which is convex, if ϕ is convex.

2.4. The Prior

An important aspect of hidden field approaches is that they allow using state-of-the-art priors/regularizers for images, such as those based on wavelet frames [10]. In this paper, we use a form of *vector total variation* (VTV) [8] defined as

$$\phi(\mathbf{z}) = \lambda \sum_{i\in\mathcal{S}} \sqrt{\big\|(\mathbf{D}_h\,\mathbf{z})_i\big\|^2 + \big\|(\mathbf{D}_v\,\mathbf{z})_i\big\|^2} + C^{\text{te}} \tag{9}$$

where $\lambda \geq 0$ is a regularization parameter, $\|\cdot\|$ is the Euclidean norm, and $\mathbf{D}_h, \mathbf{D}_v : \mathbb{R}^{K\times n} \to \mathbb{R}^{K\times n}$ are operators computing horizontal and vertical first order backward differences, respectively. *i.e.*,

$$(\mathbf{D}_h\,\mathbf{z})_i \equiv \mathbf{z}_i - \mathbf{z}_{h(i)}, \quad (\mathbf{D}_v\mathbf{z})_i \equiv \mathbf{z}_i - \mathbf{z}_{v(i)},$$

where $h(i)$ and $v(i)$ denote, respectively, the horizontal and vertical backward neighbors of pixel i, on the image lattice, assuming cyclic boundaries. We also define $\mathbf{D}$ as the linear operator such that $(\mathbf{D}\mathbf{z})_i = [(\mathbf{D}_h\,\mathbf{z})_i, (\mathbf{D}_v\,\mathbf{z})_i]^T$.

VTV regularization has a number of desirable properties: as standard TV, it promotes piecewise smoothness, but preserves strong discontinuities; the coupling among the field components in (9) tends to align the discontinuities among these components [8]; finally, it is convex and amenable to optimization via proximal methods [15].

3. SUPERVISED SEGMENTATION

The optimization (8) was addressed in [14] by converting it into the equivalent unconstrained form

$$\min_{\mathbf{z}\in\mathbb{R}^{K\times n}} \sum_{j=1}^{4} g_j(\mathbf{H}_j\mathbf{z}) \qquad \begin{array}{ll} \mathbf{H}_j : & \mathbb{R}^{K\times n} \to \mathbb{R}^{n_j} \\ g_j : & \mathbb{R}^{n_j} \to \mathbb{R}, \end{array} \tag{10}$$

where the g_j are closed, proper, and convex functions, and the $\mathbf{H}_j$ are linear operators, given by

$$\begin{array}{lll} g_1(\boldsymbol{\xi}) = -\sum_{i\in\mathcal{S}} \log\big(\mathbf{p}_i^T\boldsymbol{\xi}_i\big)_+ & \mathbf{H}_1 = \mathbf{I} & n_1 = Kn \\ g_2(\boldsymbol{\xi}) = -\sum_{i\in\mathcal{S}} \|\boldsymbol{\xi}_i\| & \mathbf{H}_2 = \mathbf{D} & n_2 = 2Kn \\ g_3(\boldsymbol{\xi}) = \iota_{\mathbb{R}_+^{K\times n}}(\boldsymbol{\xi}) & \mathbf{H}_3 = \mathbf{I} & n_3 = Kn \\ g_4(\boldsymbol{\xi}) = \iota_{\{\mathbf{1}\}}(\mathbf{1}^T\boldsymbol{\xi}) & \mathbf{H}_4 = \mathbf{I} & n_4 = Kn. \end{array}$$

Above, $(a)_+ = \max\{a, 0\}$ and $\iota_A(\mathbf{x}) = 0$, if $\mathbf{x} \in A$, and $\iota_A(\mathbf{x}) = \infty$, otherwise. Next, we write (10) in constrained form (with $\mathbf{u}_j \in \mathbb{R}^{n_j}$ and $\mathbf{u} = [\mathbf{u}_1^T, \mathbf{u}_2^T, \mathbf{u}_3^T, \mathbf{u}_4^T]^T \in \mathbb{R}^{5Kn}$),

$$\min_{\mathbf{u},\mathbf{z}} \sum_{j=1}^{4} g_j(\mathbf{u}_j), \quad \text{subject to:} \quad \mathbf{u} = \mathbf{G}\mathbf{z}$$

where $\mathbf{G} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \mathbf{H}_3^T, \mathbf{H}_4^T]^T$. We then apply the *constraint split augmented Lagrangian shrinkage algorithm* (CSALSA) [9], which is an instance of ADMM [16]. The pseudocode for the resulting algorithm, termed SegSALSA, is shown below.

Algorithm *SegSALSA*
1. Set $t = 0$, choose $\mu > 0$, $\mathbf{u}^0 = (\mathbf{u}_1^0, \mathbf{u}_2^0, \mathbf{z}_3^0, \mathbf{z}_4^0)$
2. Set $\mathbf{d}^0 = (\mathbf{d}_1^0, \mathbf{d}_2^0, \mathbf{d}_3^0, \mathbf{d}_4^0)$
3. **repeat**
4. $\quad \mathbf{z}^{t+1} \leftarrow \arg\min_{\mathbf{z}} \big\|\mathbf{G}\mathbf{z} - \mathbf{u}^t - \mathbf{d}^t\big\|_F^2$
5. $\quad$ (* update $\mathbf{u}$ *)
6. $\quad$ **for** $i = 1$ **to** 4
7. $\qquad$ **do** $\boldsymbol{\nu}_i \leftarrow \mathbf{H}_i\mathbf{z}^{t+1} - \mathbf{d}_i^t$
8. $\qquad\quad$ (* apply Moreau proximity operators *)
9. $\qquad\quad \mathbf{u}_i^{t+1} \leftarrow \arg\min_{\mathbf{u}_i} g_i(\mathbf{u}_i) + \frac{\mu}{2}\big\|\mathbf{u}_i - \boldsymbol{\nu}_i\big\|_F^2$
10. $\qquad\quad$ (* update Lagrange multipliers $\mathbf{d}$ *)
11. $\qquad\quad \mathbf{d}_i^{t+1} \leftarrow \mathbf{u}_i^{t+1} - \boldsymbol{\nu}_i$
12. $\quad t \leftarrow t + 1$
13. **until** stopping criterion is satisfied.

Given that the functions g_j are closed, proper, and convex, $\mathbf{G}$ has full column rank, and SegSALSA is an instance of ADMM, then the sequence $\mathbf{z}^t$, for $t = 0, 1, \ldots$ converges to a solution of (8) if $\mu > 0$. A few comments about the main steps of SegSALSA are in order: the quadratic problem in line 4 can be solved efficiently in the frequency domain, using the FFT, with complexity $O(Kn\log n)$ [9]; the *Moureau proximity operators* (MPO) in line 9 are pixelwise decoupled and have complexity $O(n)$ (see [14] for details on these particular MPOs). The stopping criterion is based in the primal and dual residuals [16]. In all examples shown in Section 5, we use $\mu = 1$ and SegSALSA converges in less than 200 iterations.

4. UNSUPERVISED AND SEMI-SUPERVISED SEGMENTATION

In Section 3, we assumed that all the class-conditional probability functions $p(\mathbf{x}_i|y_i = k)$ are known. In unsupervised or semi-supervised scenarios, this is not the case and those functions include unknown parameter(s) $\boldsymbol{\theta}_k$ to be learned from $\mathbf{x}$, *i.e.*, we write $p_k(\mathbf{x}_i) = p(\mathbf{x}_i|\boldsymbol{\theta}_k)$. We begin by extending to the unsupervised case the approach proposed in Section 3, by computing the joint MMAP estimate of the couple $(\mathbf{z}, \boldsymbol{\theta})$, where $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_1 \ldots, \boldsymbol{\theta}_K)$, which is a solution of

$$\max_{\mathbf{z},\boldsymbol{\theta}} \quad p(\mathbf{z})\, p(\boldsymbol{\theta}) \prod_{i\in\mathcal{S}} \sum_{y_i\in\mathcal{L}} p_{y_i}(\mathbf{x}_i|\boldsymbol{\theta}_{y_i})\, p(y_i|\mathbf{z}_i),$$
$$\text{subject to:} \quad \mathbf{z} \geq 0, \;\; \mathbf{1}^T\mathbf{z} = \mathbf{1}^T,$$

where $p(\boldsymbol{\theta})$ is a prior on $\boldsymbol{\theta}$. This problem could be tackled via alternating optimization w.r.t. $\mathbf{z}$ and $\boldsymbol{\theta}$. The optimization with respect to $\mathbf{z}$ is as in (8), thus may solved using SegSALSA. However, the optimization w.r.t. $\boldsymbol{\theta}$ may be rather involved.

To circumvent the above difficulties, and as in [10,11], we propose an EM algorithm, by treating $\mathbf{x}$ as observed data, $\mathbf{y}$ is the missing/latent data, and the pair $(\mathbf{z}, \boldsymbol{\theta})$ are the optimization variables. At the t-th iteration, the E-step and M-step of the EM algorithm are as follows:

$$\text{E-step:} \quad Q(\mathbf{z}, \boldsymbol{\theta}; \mathbf{z}^t, \boldsymbol{\theta}^t) = \mathbb{E}_{\mathbf{y}}\big[\log p(\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\theta})|\mathbf{x}, \mathbf{z}^t, \boldsymbol{\theta}^t\big],$$
$$\text{M-step:} \quad (\mathbf{z}^{t+1}, \boldsymbol{\theta}^{t+1}) \in \arg\max_{\mathbf{z},\boldsymbol{\theta}} Q(\mathbf{z}, \boldsymbol{\theta}; \mathbf{z}^t, \boldsymbol{\theta}^t).$$

Given that the complete likelihood has the form

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) = p(\mathbf{z})\, p(\boldsymbol{\theta}) \prod_{i\in\mathcal{S}} p_{y_i}(\mathbf{x}_i|\boldsymbol{\theta}_{y_i})\, p(y_i|\mathbf{z}_i)$$

and that the link $p(y_i|\mathbf{z}_i)$ is given by (7), and after simple but lengthy manipulation, we obtain

$$Q(\mathbf{z}, \boldsymbol{\theta}; \mathbf{z}^t, \boldsymbol{\theta}^t) = Q(\boldsymbol{\theta}; \mathbf{z}^t, \boldsymbol{\theta}^t) + Q(\mathbf{z}; \mathbf{z}^t, \boldsymbol{\theta}^t),$$

where

$$Q(\boldsymbol{\theta}; \mathbf{z}^t, \boldsymbol{\theta}^t) = \log p(\boldsymbol{\theta}) + \sum_{i\in\mathcal{S}} \sum_{k\in\mathcal{L}} w_{i,k}^t \log p_k(\mathbf{x}_i|\boldsymbol{\theta}_k)$$
$$Q(\mathbf{z}; \mathbf{z}^t, \boldsymbol{\theta}^t) = \log p(\mathbf{z}) + \sum_{i\in\mathcal{S}} \sum_{k\in\mathcal{L}} w_{i,k}^t \log\big((\mathbf{z}_i)_{y_i}\big),$$

and

$$w_{i,k}^t \equiv p(y_i = k|\mathbf{x}_i, \mathbf{z}_i^t, \boldsymbol{\theta}^t) = \frac{p_k(\mathbf{x}_i|\boldsymbol{\theta}_k^t)\, p(y_i = k|\mathbf{z}_i^t)}{\sum_{l=1}^K p_l(\mathbf{x}_i|\boldsymbol{\theta}_l^t)\, p(y_i = l|\mathbf{z}_i^t)}.$$

The function $Q(\mathbf{z}, \boldsymbol{\theta}; \mathbf{z}^t, \boldsymbol{\theta}^t)$ is decoupled w.r.t. $\mathbf{z}$ and $\boldsymbol{\theta}$, and the term $Q(\boldsymbol{\theta}; \mathbf{z}^t, \boldsymbol{\theta}^t)$ is decoupled w.r.t. $\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_K$, if $\ln p(\boldsymbol{\theta})$ is also similarly decoupled. The structure of the proposed EM algorithm for unsupervised segmentation, termed U-SegSALSA, is shown below.

Algorithm *U-SegSALSA*
1. Set $t = 0$, choose $\mathbf{z}^0$, $\boldsymbol{\theta}^0$
2. **repeat**
3. $\quad$ (∗ E-step ∗)
4. $\quad w_{i,k}^t \leftarrow p(y_i = k|\mathbf{x}_i, \mathbf{z}_i^t, \boldsymbol{\theta}^t), \; i = 1, ..., n, \; k = 1, ..., K$
5. $\quad \boldsymbol{\theta}^{t+1} \leftarrow \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \mathbf{z}^t, \boldsymbol{\theta}^t)$ $\quad$ (∗ M-step w.r.t. $\boldsymbol{\theta}$ ∗)
6. $\quad \mathbf{z}^{t+1} \leftarrow \arg\max_{\mathbf{z}} Q(\mathbf{z}; \mathbf{z}^t, \boldsymbol{\theta}^t)$ $\quad$ (∗ M step w.r.t. $\mathbf{z}$ ∗)
7. $\quad t \leftarrow t + 1$
8. **until** stopping criterion is satisfied.

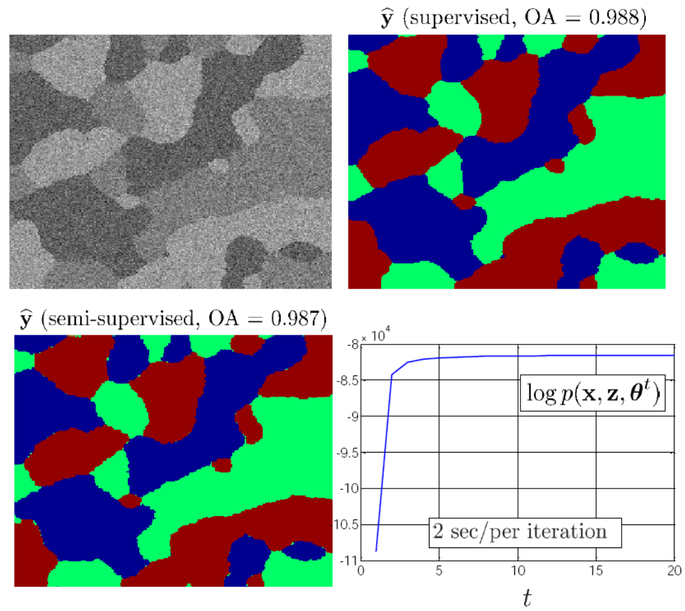

Fig. 1. Top left: observed image of features ($p_k(x_i) = \mathcal{N}(k, 0.8)$, for $k \in \{1, 2, 3, 4\}$). Top right: supervised segmentation by SegSALSA. Bottom left: semi-supervised segmentation (10 labeled samples per class). Bottom right: evolution of the log-likelihood.

Notice that $Q(\mathbf{z}; \mathbf{z}^t, \boldsymbol{\theta}^t)$ is convex w.r.t. $\mathbf{z}$ and that the optimization in line 6 of U-SegSALSA is similar to (8). The only difference is that instead of $-\log(\mathbf{p}_i^T\mathbf{z}_i)$, we now have $-(\mathbf{w}_i^t)^T \log(\mathbf{z}_i)$, where $\mathbf{w}_i^t = [w_{i,1}^t, \ldots, w_{i,K}^t]^T$. The corresponding MPO (which is required to use SegSALSA to solve line 6 of U-SegSALSA) is

$$\arg\min_{\boldsymbol{\xi}} \frac{\mu}{2}\|\mathbf{v} - \boldsymbol{\xi}\|^2 - \mathbf{w}^T \log\big((\boldsymbol{\xi})_+\big) = \frac{\mathbf{v} + \sqrt{\mathbf{v}^2 + 4\mathbf{w}/\mu}}{2},$$

component-wise, thus with $O(Kn)$ cost.

Finally, in the semi-supervised case, we have access to the class/segment labels of some of the image locations. The only difference w.r.t. the fully unsupervised case just described is that the $w_{i,k}^t$ variables corresponding to the labeled samples are kept frozen at the given labels.

5. EXPERIMENTAL RESULTS

The effectiveness of the proposed algorithms is now illustrated on simulated and real data. Fig. 1, top left, shows a simulated 256×256 image of real-valued ($d = 1$) features: $p_k(x_i) = \mathcal{N}(k, 0.8)$, for $k \in \{1, 2, 3, 4\}$. The label image $\mathbf{y}$ is a sample of a first-order MLL-MRF with parameter 1.5. The significant overlap of the four class-conditional densities suggests a difficult segmentation problem, which is confirmed by the *overall accuracy* (OA) of the *maximum likelihood* (ML) segmentation, which is 0.66. Fig. 1, top right, shows the SegSALSA segmentation, which has OA = 0.998,

Fig. 2. Left: 256×256 RGB image. Right: U-SegSALSA background/foreground segmentation.

corresponding to an almost perfect segmentation. The semi-supervised segmentation obtained with only 10 labeled samples per class is shown in Fig.1, bottom left. Since the optimization problem is nonconvex, we run the EM algorithm 10 times with independent noise samples and initialization of the class means set to the respective sample mean and unit variance, and hidden vectors $\mathbf{z}_i$ uniformly distributed. The semi-supervised algorithm achieved mean OA equal to 0.987 (± 0.0015). We highlight that the supervised and semi-supervised algorithms produced identical segmentations and that the semi-supervised version converges in less than 20 iterations (roughly 40 seconds, on a standard PC running MATLAB). In both algorithms the regularization parameter was set to $\lambda = 1.4$; values between 1 and 2 yield very similar results.

Fig. 2 shows a 256×256 RGB image of two horses on a grass background and its semi-supervised segmentation, using two Gaussian class densities and 4 labeled samples per class. The two Gaussians are initialized with the sample means of the labeled samples and identity covariances, while the hidden vectors are initialized with uniform distributions. The segmentation is qualitatively very good, with the two horses accurately separated from the background. The computation time was about 40 seconds.

6. CONCLUDING REMARKS

This paper avoids the integer optimization problems that usually appear in image segmentation, by resorting to the hidden field approach pioneered by [7]. We revisited the SegSALSA algorithm introduced in [14] for supervised scenarios and extended it to unsupervised and semi-supervised scenarios. The proposed method is an EM algorithm, where the E-step is similar to that of a finite mixture, and the M-step is similar to SegSALSA. The effectiveness of the proposed method was illustrated with simulated and real images.

REFERENCES

[1] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.

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

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 6, pp. 721–741, 1984.

[4] L. Vese and T. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *Intern. Jour. Computer Vision*, vol. 50, pp. 271–293, 2002.

[5] C. Nieuwenhuis, E. Töppe, and D. Cremers, "A survey and comparison of discrete and continuous multi-label optimization approaches for the Potts model," *Intern. Jour. of Computer Vision*, pp. 223–240, 2013.

[6] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 22, pp. 888–905, 2000.

[7] J. Marroquin, E. Santana, and S. Botello, "Hidden Markov measure field models for image segmentation," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 25, pp. 1380–1387, 2003.

[8] X. Bresson and T. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems and Imaging*, vol. 2, pp. 455–484, 2008.

[9] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Processing*, vol. 20, pp. 681–695, 2011.

[10] M. A. T. Figueiredo, "Bayesian image segmentation using wavelet-based priors," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 437–443.

[11] ——, "Bayesian image segmentation using Gaussian field priors," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer, 2005, pp. 74–89.

[12] D. Greig, B. Porteous, and A. Seheult, "Exact maximum a posteriori estimation for binary images," *Jour. of the Royal Statistical Society (B)*, pp. 271–279, 1989.

[13] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Analysis Machine Intell.*, vol. 30, pp. 1068–1080, 2008.

[14] J. Bioucas-Dias, F. Condessa, and J. Kovačević, "Alternating direction optimization for image segmentation using hidden Markov measure field models," in *IS&T/SPIE Electronic Imaging*, 2014, pp. 90 190P–90 190P.

[15] P. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, 2011, pp. 185–212.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, pp. 1—122, 2011.