SUPERVISED HYPERSPECTRAL IMAGE SEGMENTATION: A CONVEX FORMULATION USING HIDDEN FIELDS

Filipe Condessa [a,b,c,e], José Bioucas-Dias [a,b], and Jelena Kovačević [c,d,e]

[a]Instituto de Telecomunicações, Lisboa, Portugal
[b]Instituto Superior Técnico, University of Lisbon, Lisboa, Portugal
[c]Department of ECE, Carnegie Mellon University, Pittsburgh, PA, USA
[d]Department of BME, Carnegie Mellon University, Pittsburgh, PA, USA
[e]Center for Bioimage Informatics, Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

Image segmentation is fundamentally a discrete problem. It consists of finding a partition of the image domain such that the pixels in each element of the partition exhibit some kind of similarity. The optimization is obtained via integer optimization which is NP-hard, apart from few exceptions. We sidestep from the discrete nature of image segmentation by formulating the problem in the Bayesian framework and introducing a hidden set of real-valued random fields determining the probability of a given partition. Armed with this model, the original discrete optimization is converted into a convex program. To infer the hidden fields, we introduce the Segmentation via the Constrained Split Augmented Lagrangian Shrinkage Algorithm (SegSALSA). The effectiveness of the proposed methodology is illustrated with hyperspectral image segmentation.

Index Terms— Image segmentation, hidden Markov measure fields, hidden fields, alternating optimization, Constrained Split Augmented Lagrangian Shrinkage Algorithm (SALSA).

1. INTRODUCTION

Image segmentation plays a crucial role in many hyperspectral imaging applications [1]. The image segmentation problem consists in finding a partition of the image domain such that the image properties in a given partition element, expressed via image features or cues, are similar in some sense. Because image segmentation is almost invariably an ill-posed inverse problem, some form of regularization (a prior in Bayesian terms) is usually imposed on the solution with the objective of promoting solutions with desirable characteristics.

The type of regularization and the estimation criteria used to infer a partition are relates issues. In the Bayesian framework, the segmentation is often obtained by computing the *maximum a posteriori probability* (MAP) estimate of the partition, which maximizes the product of likelihood function (*i.e.*, the probability of the observed image given the partition) with the prior probability for the partition, usually a Markov Random Field (MRF) [2].

Images of integers are natural representations for partitions. With this representation, the MAP segmentation is an integer op-

timization problem that, apart from a few exceptions, is NP-hard and thus impossible to solved exactly. In the last decade, a large class of powerful integer minimization methods based on graph cuts [3] and based on convex relaxations[4] has been proposed to solve approximately MAP estimation problems of discrete MRFs.

In this paper, inspired by the "hidden Markov measure fields" introduced by [5], we sidestep the hurdles raised by the discrete nature of image segmentation by (a) formulating the problem in the Bayesian framework and (b) introducing a hidden set of real-valued random fields conditioning the probability of a given partition. Armed with this model, we compute the marginal MAP (MMAP) estimate of the hidden fields, which is, under suitable conditions, a convex program. From the MMAP estimate of the hidden fields and the conditional probability of the partition, we obtain a soft and a hard estimate of the partition.

In the hidden field model, the prior on the partition is indirectly expressed by the prior on the hidden fields. In this paper, we use a form of vectorial total variation (VTV) [6, 7], which promotes piecewise smooth segmentations and promotes sharp discontinuities in the estimated partition.

In [8] the image segmentation problem is approached by closely following the "hidden Markov measure fields" paradigm [5], being the main difference the statistical link, based on the multinomial logistic model, and the prior on the hidden fields, based on wavelets. In [9] a multi-class labeling is approximately solved using tools from convex optimization. The approach proposed there has links with ours in that it also uses a VTV regularizer and the optimization imposes constraints similar to ours. However, the data terms are different: ours is derived in under the a Bayesian framework whereas theirs is introduced heuristically. In addition, our optimization algorithm exploits the SALSA algorithm [10] splitting flexibility to avoid double loops as those shown in [9]. Finally, we mention [11], which also uses non-isotropic total variation as a regularizer and imposes constraints similar to ours.

The main contributions of the paper are the proposal of the VTV as prior on the hidden fields, and the introduction of an instance of the SALSA [10] algorithm, which we term SegSALSA, to compute the exact MMAP estimate of the partition with $O(Kn \ln n)$ complexity, where K is the cardinality of the partition and n the number of image pixels.

The paper is organized as follows. Section 2 formulates the problem, introduces the hidden fields, the MMAP of the hidden fields, the statistical link between the class labels and the hidden fields, and the VTV prior. Section 3 presents the SegSALSA algorithm, which is an instantiation of SALSA to the problem in hand.

Section 4 presents a number of experimental results with hyperspectral images. Finally, Section 5 presents as few concluding remarks and pointers to future work.

2. PROBLEM FORMULATION

To formulate the segmentation problem in mathematical terms, we start by introducing notation. Let $\mathcal{S} \equiv \{1, \cdots, n\}$ denote a set of integers indexing the n pixels of an image and $\mathbf{x} \equiv [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ a $d \times n$ matrix holding the d-dimensional image feature vectors. Given $\mathbf{x}$, the goal of image segmentation is to find a partition $P \equiv \{R_1, \ldots, R_K\}$ of $\mathcal{S}$ such that the feature vectors with indices in a given set R_i, for $i = 1, \ldots, K$, be *similar* in some sense[1]. Associated with a partition P, we introduce the image of class labels, also termed segmentation, $\mathbf{y} \equiv (y_1, \cdots, y_n) \in \mathcal{L}^n$, where $\mathcal{L} \equiv \{1, \ldots, K\}$, such that $y_i = k$ if and only if $i \in R_k$. We remark that there is a one-to-one correspondence between partitions and segmentations.

2.1. Maximum *a posteriori* probability segmentation

We adopt a Bayesian perspective to the segmentation problem. The MAP segmentation is given by

$$\widehat{\mathbf{y}}_{MAP} = \arg \max_{\mathbf{y} \in \mathcal{L}^n} p(\mathbf{y}|\mathbf{x}) = \arg \max_{\mathbf{y} \in \mathcal{L}^n} p(\mathbf{x}|\mathbf{y}) p(\mathbf{y}), \quad (1)$$

where $p(\mathbf{y}|\mathbf{x})$ is the posterior probability[2] of $\mathbf{y}$ given $\mathbf{x}$, $p(\mathbf{x}|\mathbf{y})$ is the observation model, and $p(\mathbf{y})$ is the prior probability for the labeling $\mathbf{y}$. Under assumption of *conditional independence*, we have

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

where $p_k(\mathbf{x}_i) = p(\mathbf{x}_i|y_i = k)$. We assume that the class densities p_k, for $k \in \mathcal{L}$ are known or learned from a training set in a supervised fashion.

Various forms of Markov random fields (MRFs) have been widely used as prior probability for the class labels $\mathbf{y}$. A paradigmatic example is the multilevel logistic/Potts model (MLL) [2]. These models promote piecewise smooth segmentations, *i.e.*, segmentations in which it is more likely to have neighboring labels of the same class than the other way around.

The maximization in (1) is an integer optimization problem, impossible to solve exactly for more than two classes. Various algorithms to approximate $\widehat{\mathbf{y}}_{MAP}$ have been introduced in the last decade of which we highlight the graph cuts based α-expansion [3], the sequential tree-reweighted message passing (TRW-S) [12], the max-product loopy belief propagation (LBP) [13], and convex relaxations [14].

2.2. Hidden fields

The MAP formulation regarding the class labels $\mathbf{y}$ raises a series of difficulties regarding (a) the high computational complexity involved in computing the solution of the integer optimization problem (1), (b) the selection of prior $p(\mathbf{y})$ and (c) the learning of unknown

---

[1] We recall that a partition of a set $\mathcal{S}$ is a collection of sets $R_i \subset \mathcal{S}$, for $i = 1, \ldots, K$, where $\cup_{i=1} R_i = \mathcal{S}$ and $R_i \cap R_j = \emptyset$, $i \neq j$.

[2] To keep the notation light, we denote both probability densities and probability distributions with $p(\cdot)$. Furthermore, the random variable to which $p(\cdot)$ refers is to be understood from the context.

parameters $\boldsymbol{\theta}$ parameterizing the model $p(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta})$. In [5], the original segmentation problem is reformulated in terms of real-valued hidden fields conditioning the random field $\mathbf{y}$ and endowed with a Gaussian MRF prior promoting smooth fields. The segmentation is obtained by computing the *marginal MAP* (MMAP) estimate of the hidden fields, which corresponds to a soft segmentation. This approach converts an integer optimization problem into a smooth constrained convex problem, simpler to solve exactly using convex optimization.

2.3. Marginal MAP estimate of the hidden fields

To formulate the hidden field concept, and following closely [5], let $\mathbf{z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] \in \mathbb{R}^{K \times n}$ denote a $K \times n$ matrix holding a collection of hidden random vectors, $\mathbf{z}_i \in \mathbb{R}^K$, for $i \in \mathcal{S}$, and define the joint probability $p(\mathbf{y}, \mathbf{z}) = p(\mathbf{y}|\mathbf{z}) p(\mathbf{z})$, with $p(\mathbf{y}|\mathbf{z}) = \prod_{i=1}^{n} p(y_i|\mathbf{z}_i)$. The joint probability of $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is given by

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{x}|\mathbf{y}) p(\mathbf{y}|\mathbf{z}) p(\mathbf{z}), \quad (3)$$

from which we may write the marginal density with respect to $(\mathbf{x}, \mathbf{z})$ as

$$p(\mathbf{x}, \mathbf{z}) = \prod_{i=1}^{n} \left\{ \sum_{y_i \in \mathcal{L}} p(\mathbf{x}_i|y_i) p(y_i|\mathbf{z}_i) \right\} p(\mathbf{z}). \quad (4)$$

The MMAP estimate of the of the hidden field $\mathbf{z}$ is then given by

$$\widehat{\mathbf{z}}_{MMAP} = \arg \max_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}). \quad (5)$$

From $\widehat{\mathbf{z}}_{MMAP}$, we obtain the soft segmentation $p(\mathbf{y}|\widehat{\mathbf{z}}_{MMAP})$. A hard segmentation may be then obtained by computing the labelling that maximizes the soft segmentation.

2.4. Statistical link between class labels and hidden fields

The conditional probabilities $p(y_i|\mathbf{z}_i)$, for $i \in \mathcal{S}$, play a central role in our approach. As in [5], we adopt the following model

$$p(y_i = k|\mathbf{z}_i) \equiv [\mathbf{z}_i]_k \quad i \in \mathcal{S}, \quad k \in \mathcal{L}, \quad (6)$$

where the $[\mathbf{a}]_k$ is the k-th element of vector $\mathbf{a}$. Given that $[\mathbf{z}_i]_k$, for $k \in \mathcal{L}$, represents a probability distribution, then the hidden vectors $\mathbf{z}_i$, for $i \in \mathcal{S}$, satisfies the component-wise nonnegativity constraint $\mathbf{z}_i \geq 0$, and the sum-to-one constraint $\mathbf{1}_K^T \mathbf{z}_i = 1$.

2.5. The prior

We adopt form of vector total variation (VTV) [6, 7] regularizer,

$$- \ln p(\mathbf{z}) \equiv \lambda_{TV} \sum_{n \in \mathcal{S}} \sqrt{\left\| \mathbf{D}_h \mathbf{z}[n] \right\|^2 + \left\| \mathbf{D}_v \mathbf{z}[n] \right\|^2} + c^{te}, \quad (7)$$

where $\lambda_{TV} > 0$ is a regularization parameter, and $\mathbf{D}_h$, $\mathbf{D}_v$ are linear operators computing horizontal and vertical first order differences, respectively. The regularizer (7) has a number of desirable properties: (a) it promotes piecewise smooth hidden fields; (b) it tends to preserve discontinuities aligning them among classes; (c) it is convex, although not strictly, allowing optimization via proximal methods relying on Moreau proximity operators [15].

3. OPTIMIZATION ALGORITHM

Considering the model (6) and the prior (7), we may write the MMAP estimation of $\mathbf{z}$ as

$$\widehat{\mathbf{z}}_{MMAP} = \arg \min_{\mathbf{z} \in \mathbb{R}^{K \times n}} \sum_{i=1}^{n} - \ln \left(\mathbf{p}_i^T \mathbf{z}_i \right) +$$
$$+ \lambda_{TV} \sum_{n \in \mathcal{S}} \sqrt{\left\| \mathbf{D}_h \mathbf{z}[n] \right\|^2 + \left\| \mathbf{D}_v \mathbf{z}[n] \right\|^2} \quad (8)$$
$$\text{subject to: } \mathbf{z} \geq 0, \quad \mathbf{1}_K^T \mathbf{z} = \mathbf{1}_n^T,$$

where $\mathbf{p}_i \equiv \left[p(\mathbf{x}_i|y_i = 1), \ldots, p(\mathbf{x}_i|y_i = K) \right]^T$ and it was assumed that $\mathbf{p}_i^T \mathbf{z}_i > 0$ for $\mathbf{z}_i$ in the feasible set. As the Hessian matrix of $- \ln \left(\mathbf{p}_i^T \mathbf{z}_i \right)$ is a semipositive definite matrix, (8) is convex. In this section, we develop an instance of the Split Augmented Lagrangian Shrinkage (SALSA) methodology introduced in [10] to compute $\widehat{\mathbf{z}}_{MMAP}$. We start by rewriting the optimization (8),

$$\min_{\mathbf{z} \in \mathbb{R}^{K \times n}} \sum_{i=1}^{4} g_i(\mathbf{H}_i \mathbf{z}), \quad (9)$$

where g_i, for $i = 1, \ldots, 4$, denote, closed, proper, and convex functions, and $\mathbf{H}_i$, for $i = 1, \ldots, 4$, denote linear operators. The particular definitions of these entities for our problem are as follows:

$$\mathbf{H}_1 = \mathbf{I}, \ \mathbf{H}_2 = \begin{pmatrix} \mathbf{D}_h \\ \mathbf{D}_v \end{pmatrix}, \ \mathbf{H}_3 = \mathbf{I}, \ \mathbf{H}_4 = \mathbf{I},$$
$$g_1(\boldsymbol{\xi}) = \sum_{n \in \mathcal{S}} - \ln \left(\mathbf{p}_n^T \boldsymbol{\xi}_n \right)_+,$$
$$g_2(\boldsymbol{\xi}) = \lambda_{TV} \sum_{n \in \mathcal{S}} \sqrt{\left\| \boldsymbol{\xi}^h[n] \right\|^2 + \left\| \boldsymbol{\xi}^v[n] \right\|^2}, \quad (10)$$
$$g_3(\boldsymbol{\xi}) = \iota_+(\boldsymbol{\xi}), g_4(\boldsymbol{\xi}) = \iota_1(\mathbf{1}_K^T \boldsymbol{\xi}),$$

where $\mathbf{I}$ denotes the identity operator, $\boldsymbol{\xi}$ are dummy variables whose dimensions depend on the functions g_i, for $i = 1, 2, 3, 4$, $(x)_+$ is the positive part of x, and $\ln(0) \equiv +\infty$. The function ι_+ denotes the indicator in the set in $\mathbb{R}_+^{K \times n}$, i.e., $\iota_+(\boldsymbol{\xi}) = 0$ if $\boldsymbol{\xi} \in \mathbb{R}_+^{K \times n}$ and $\iota_+(\boldsymbol{\xi}) = \infty$ otherwise. By the same token $\iota_1(\boldsymbol{\xi})$ is the indicator in the set $\{\mathbf{1}_n\}$.

We now introduce the variable splitting $\mathbf{u}_i = \mathbf{H}_i \mathbf{z}$, for $i = 1, 2, 3, 4$, in (9) and convert the original optimization into the equivalent constrained form

$$\min_{\mathbf{u}, \mathbf{z}} \sum_{i=1}^{4} g_i(\mathbf{u}_i) \quad \text{subject to} \quad \mathbf{u} = \mathbf{G} \mathbf{z}, \quad (11)$$

where $\mathbf{G} : \mathbb{R}^{K \times n} \mapsto \mathbb{R}^{5K \times n}$ is the linear operator obtained by columnwise stacking the operators $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{H}_3$, and $\mathbf{H}_4$.

The next step consists in applying the SALSA methodology [10] to (11). SALSA is essentially an instance of the alternating method of multipliers (ADMM) designed to optimize sums of an arbitrary number of convex terms. Solving (11), becomes equivalent to solving the following decoupled problem

$$\mathbf{z}^{k+1} = \arg \min_{\mathbf{z}} \left\| \mathbf{G} \mathbf{z} - \mathbf{u}^k - \mathbf{d}^k \right\|_F^2, \quad (12)$$
$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} f(\mathbf{u}) + \frac{\mu}{2} \left\| \mathbf{G} \mathbf{z}^{k+1} - \mathbf{u} - \mathbf{d}^k \right\|_F^2, \quad (13)$$
$$\mathbf{d}^{k+1} = \mathbf{d}^k - \left[\mathbf{G} \mathbf{z}^{k+1} - \mathbf{u}^{k+1} \right], \quad (14)$$

with $\mathbf{d}$ denoting the scaled Lagrange multipliers. Solving the quadratic optimization problem is dominated by computing independent cyclic convolutions on each image of $\mathbf{z}$ (corresponding to the VTV prior), which can be efficiently done in the frequency domain using the fast Fourier transform (FFT) with $O(Kn \ln n)$ complexity. A distinctive feature of SALSA is that optimization with respect to $\mathbf{u}$ is decoupled into optimization problems with respect to the blocks $\mathbf{u}_i$, for $i = 1, 2, 3, 4$, whose solutions are the so-called Moreau proximity operators (MPOs) [15] for the respective convex functions g_i, for $i = 1, 2, 3, 4$. Solving for $\mathbf{u}_i$ can be done efficiently by computing the proximity operators for g_i, which have a closed form (finding a root for g_1, soft thresholding for g_2, projection on positive orthant for g_3, and projection on a simplex g_4), corresponding to a $O(Kn)$ complexity.

We term the resulting algorithm Segmentation via Augmented Lagrangian Shrinkage Algorithm (SegSALSA). SegSALSA converges for any $\mu > 0$, having a complexity of $O(Kn \ln n)$. Regarding the stopping criterion, we impose that the primal and dual residuals be smaller than a given threshold. We have observed, that a fixed number of iterations of the order of 200 provides excellent results.

4. RESULTS

We use the SegSALSA algorithm to classify the ROSIS Pavia scene (Fig. 1). This hyperspectral image was acquired by the ROSIS optical sensor on the University of Pavia, Italy. It is a 610×340 image with a spatial resolution of $1.3m/$pixel, and 103 spectral bands. The image contains nine exclusive land-cover classes, with the accuracy of the classification being measured on those nine classes. The class models are learned using the LORSAL algorithm [16]. We run the SegSALSA algorithm for four training sets of different dimensions (20, 40, 200, and 500 samples per class randomly selected). The accuracy is computed from 10 Monte Carlo runs. We present in Table 1 the overall accuracy and average accuracy. The value of overall accuracy $= 96.23\% \pm 0.65\%$ obtained with 200 samples per class is considered state-of-the-art [1].

Table 1. Accuracy on ROSIS Pavia scene after 10 Monte-Carlo runs with varying training set size.

Training samples per class	Overall accuracy	Average accuracy
20	$89.10\% \pm 3.67$	$91.48\% \pm 2.51$
40	$92.85\% \pm 2.27$	$93.15\% \pm 1.01$
200	$96.23\% \pm 0.65$	$94.96\% \pm 0.53$
500	$97.19\% \pm 0.52$	$95.77\% \pm 0.51$

5. CONCLUDING REMARKS

In this paper, we introduce a new approach to supervised image segmentation that avoids the discrete nature of problem present in many formulations. This is achieved by leveraging on the "hidden Markov measure field" introduced by [5] in 2003. The proposed approach relies on four main ingredients: (a) formulating the image segmentation in the Bayesian framework; (b) introducing a hidden set of real-valued random fields determining the probability of a given partition; (c) adopting an form of isotropic vector total variation; and (d) introducing the Segmentation via the Constrained Split Augmented Lagrangian Shrinkage Algorithm (SegSALSA) to effectively solve the convex program which constitutes the marginal MAP inference
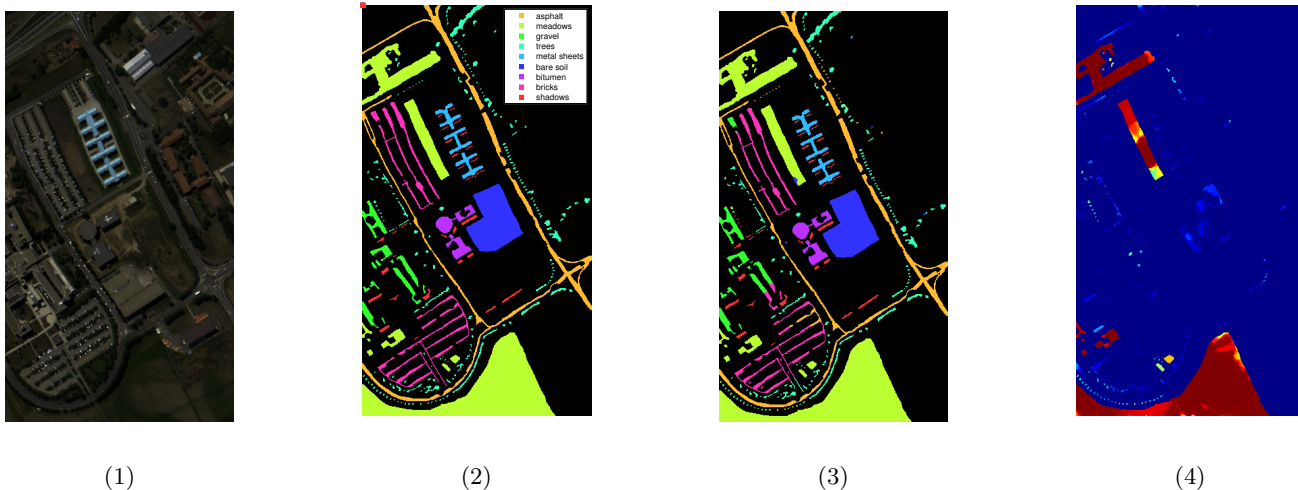
(1)   (2)   (3)   (4)

Fig. 1. Classification of the ROSIS Pavia scene with varying dimension of the training set. Top row: (1) False color composition of the ROSIS Pavia scene, (2) ground truth containing the 9 mutually exclusive land-cover classes, (3) classification for 200 training samples per class (98.5% accuracy), (4) latent probabilities (hidden field) for "meadow" class for 200 training samples per class.

of the hidden field. Future work will focus on extending the proposed methodology to unsupervised and semi-supervised scenarios.

6. 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," *Geoscience and Remote Sensing Magazine, IEEE*, vol. 1, no. 2, pp. 6–36, 2013.

[2] S. Li, *Markov random field modeling in computer vision*, Springer-Verlag New York, Inc., 1995.

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

[4] A. Martins, M. Figueiredo, P. Aguiar, N. Smith, and E. Xing, "An augmented Lagrangian approach to constrained MAP inference," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 169–176.

[5] J. Marroquin, E. Santana, and S. Botello, "Hidden Markov measure field models for image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 11, pp. 1380–1387, 2003.

[6] 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, no. 4, pp. 455–484, 2008.

[7] B. Goldluecke, E. Strekalovskiy, and D. Cremers, "The natural vectorial total variation which arises from geometric measure theory," *SIAM Journal on Imaging Sciences*, vol. 5, no. 2, pp. 537–563, 2012.

[8] M. Figueiredo, "Bayesian image segmentation using wavelet-based priors," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 437–443.

[9] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr, "Convex multi-class image labeling by simplex-constrained total variation," in *Scale Space and Variational Methods in Computer Vision*, X-.C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, Eds., vol. 5567 of *Lecture Notes in Computer Science*, pp. 150–162. Springer Berlin Heidelberg, 2009.

[10] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *Image Processing, IEEE Transactions on*, vol. 20, no. 3, pp. 681–695, 2011.

[11] L. Sun, Z. Wu1, J. Liu, and Z. Wei, "Supervised hyperspectral image classification using sparse logistic regression and spatial-tv regularization," in *IEEE Geoscience and Remote Sensing Symposium (IGARSS'13), Melbourne, Australia, 2013*, 2013.

[12] Vladimir Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 10, pp. 1568–1583, 2006.

[13] P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," *International journal of computer vision*, vol. 70, no. 1, pp. 41–54, 2006.

[14] M. Kumar, V. Kolmogorov, and P. Torr, "An analysis of convex relaxations for MAP estimation of discrete MRFs," *The Journal of Machine Learning Research*, vol. 10, pp. 71–106, 2009.

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

[16] J. Li, J. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new bayesian approach with active learning," *Geoscience and Remote Sensing Magazine, IEEE*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.