

EXPLOITING SPATIAL INFORMATION IN SEMI-SUPERVISED HYPERSPECTRAL IMAGE SEGMENTATION

Jun Li , José M. Bioucas-Dias

Instituto de Telecomunicações,
Instituto Superior Técnico, TULisbon,
1900-118, Lisboa, Portugal

Antonio Plaza

Department of Technology of Computers and
Communications, University of Extremadura,
E-10071 Caceres, Spain

ABSTRACT

We present a new semi-supervised segmentation algorithm suited to hyperspectral images, which takes full advantage of the spectral and spatial information available in the scenes. We mainly focus on problems involving very few labeled samples and a larger set of unlabeled samples. A *multinomial logistic regression* (MLR) is used to model the posterior class probability distributions, whereas a *multilevel logistic level* (MLL) prior is adopted to model the spatial information present in class label images. The multinomial logistic regressors are learnt using an *expectation maximization* (EM) type algorithm, where the class labels of the unlabeled samples are dealt with as unobserved random variables. The expectation step of the EM algorithm is computed using *belief propagation* (BP). In the maximization step of the EM algorithm, we compute the *maximum a posteriori estimate* (MAP) estimate of the multinomial logistic regressors. For the segmentation, we compute both the MAP solution and the *maximizer of the posterior marginal* (MPM) provided by the belief propagation algorithm. We show, using the well-known AVIRIS Indian Pines data, that both solutions exhibit state-of-the-art performance.

Index Terms— Semi-supervised classification, belief propagation, expectation maximization, hyperspectral segmentation, integer optimization.

1. INTRODUCTION

The wide availability of hyperspectral images leads to new developments in the fields of image segmentation and classification [1]. The detailed information about spectral signatures provided by hyperspectral sensors has fostered the development of new algorithms capable of properly handling the high dimensionality of the data. In hyperspectral remote sensing image classification, the acquisition of labeled training data is costly and time consuming. In other words, we are usually given an insufficient set of labeled pixels to develop the classifier. Thus, the classification problem is challenging as the size of the labeled pixels is typically very small. The difficulties in learning high dimensional densities from a limited number of training samples, known as the Hughes phenomenon, is one of the major problems related with the analysis of this type of data and, although progresses have been made, it remains an active area of research [2, 3].

In classification problems, discriminative approaches (which learn the class probability distributions in high dimensional spaces by inferring the boundaries between classes in the feature space [4]),

have been proved to be successful in dealing with small class distances, high dimensionality, and limited training sets [5]. The *support vector machines* (SVMs) [6] and the MLR [7] are among the state-of-the-art discriminative techniques to classification. Due to their ability to deal with large input spaces efficiently and to produce sparse solutions, SVMs have been successfully used for supervised and semi-supervised classification of hyperspectral data [5, 8]. The MLR, compared with SVMs, has the advantage of being able to learn the class probability distributions. The *sparse multinomial logistic regression* (SMLR) [9] and the *fast sparse multinomial logistic regression* (FSMLR) [10] are effective algorithms to estimate multilogistic regressors. More recently, the introduction of the *logistic regression via splitting and augmented Lagrangian* (LORSAL) algorithm [11] has fostered processing of larger data sets and number of classes. These ideas have been applied to hyperspectral image classification problems [12–14].

Recently, the incorporation of spatial information has been an active area of research in order to improve the accuracy of hyperspectral classification and segmentation techniques [8, 10, 15, 16]. Available approaches exploit, in a way or another, the continuity (in probability sense) of neighboring labels, i.e. in hyperspectral images, it is likely that two neighboring pixels have the same label. However, in most hyperspectral imaging applications the acquisition of labeled samples is very hard and/or expensive, while unlabeled samples are easier to obtain and use. This observation is currently fostering active research towards the integration of labeled and unlabeled samples in hyperspectral data classification [8, 12–14, 17].

In this paper, we present a new semi-supervised algorithm which exploits both the spatial and the spectral information in the interpretation of remotely sensed hyperspectral data. In our proposed technique, an MLR is used to model the posterior class probability distributions, while an MLL prior is adopted to describe the class labels. The main contributions of this work to the state-of-the-art in hyperspectral classification can be summarized as follows:

- the integration of both spectral and spatial information to infer the posterior class probability distributions.
- the integration of both labeled and unlabeled samples to learn the parameters of the complete model.
- the development of an EM algorithm using BP to estimate marginals, which enables the inclusion of the unlabeled samples and also provides a segmentation output.

The remainder of the paper is organized as follows. Section 2 formulates the problem. Section 3 presents the proposed approach. Section 4 reports segmentation results on a real hyperspectral dataset, in comparison with state-of-the-art competitors. Finally, Section 5 concludes with some remarks and future work.

This work was supported by Marie Curie training Grant MEST-CT-2005-021175 and MRTN-CT-2006-035927 from the European Commission.

2. PROBLEM FORMULATION

Let $\mathcal{L} \equiv \{1, \dots, K\}$ denote a set of K class labels, $\mathcal{S} \equiv \{1, \dots, n\}$ a set of integers indexing the n pixels of an image, $\mathbf{x} \equiv (\mathbf{x}_1, \dots, \mathbf{x}_n)$ an image of d -dimensional feature vectors, $\mathbf{y} \equiv (y_1, \dots, y_n)$ an image of labels, $\mathcal{D}_L \equiv \{(y_{i_1}, \mathbf{x}_{i_1}), \dots, (y_{i_L}, \mathbf{x}_{i_L})\}$ a set of labeled samples, $\mathcal{Y}_L \equiv \{y_{i_1}, \dots, y_{i_L}\}$ the set of labels in \mathcal{D}_L , $\mathcal{X}_L \equiv \{\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_L}\}$ the set of feature vectors in \mathcal{D}_L , $\mathcal{X}_U \equiv \{\mathbf{x}_{i_{L+1}}, \dots, \mathbf{x}_{i_{L+U}}\}$ the set of unlabeled feature vectors, and $\mathcal{Y}_U \equiv \{y_{i_{L+1}}, \dots, y_{i_{L+U}}\}$ the set of random variables associated with the unlabeled class labels.

With the above definitions in mind, the goal of image segmentation (and of classification) is to estimate \mathbf{y} having observed \mathbf{x} . In a Bayesian framework, this estimate is usually carried out by maximizing the posterior distribution¹

$$p(\mathbf{y}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{y})p(\mathbf{y}),$$

where $p(\mathbf{x}|\mathbf{y})$ is the likelihood function (*i.e.*, the probability of feature image given of class label image \mathbf{y}) and $p(\mathbf{y})$ is the prior over \mathbf{y} . Assuming conditional independency of the features given the labels, *i.e.*, $p(\mathbf{x}|\mathbf{y}) = \prod_i p(\mathbf{x}_i|y_i)$, then the posterior $p(\mathbf{y}|\mathbf{x})$, as a function of \mathbf{y} , may be written as

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \frac{1}{p(\mathbf{x})} p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \\ &= \alpha(\mathbf{x}) \prod_i \frac{p(y_i|\mathbf{x}_i)}{p(y_i)} p(\mathbf{y}), \end{aligned} \quad (1)$$

where $\alpha(\mathbf{x}) \prod_{i=1}^{i=n} p(\mathbf{x}_i)/p(\mathbf{x})$ is a factor not depending on \mathbf{y} . Herein, we assume, without loss of generality, that $p(y_i) = 1/K$.

3. PROPOSED APPROACH

In this section we go briefly through the MLR details, describe the MLL prior, present the EM algorithm to infer the multilogistic regressors, and present the MPM and MAP algorithms.

3.1. The multinomial logistic regression

In this paper, the posterior class probability distribution $p(y_i|\mathbf{x}_i)$ is modeled with the MLR [7]

$$p(y_i = k|\mathbf{x}_i, \boldsymbol{\omega}) \equiv \frac{\exp(\boldsymbol{\omega}^{(k)} \mathbf{h}(\mathbf{x}_i))}{\sum_{k=1}^K \exp(\boldsymbol{\omega}^{(k)} \mathbf{h}(\mathbf{x}_i))}, \quad (2)$$

where $\mathbf{h}(\mathbf{x}_i) \equiv [h_1(\mathbf{x}_i), \dots, h_l(\mathbf{x}_i)]^T$ is a vector of l fixed functions of the input, often termed features; $\boldsymbol{\omega}$ denotes the regressors and $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}^{(1)T}, \dots, \boldsymbol{\omega}^{(K-1)T}]^T$; because the density (2) does not depend on translations on the regressors $\boldsymbol{\omega}^{(k)}$, we take $\boldsymbol{\omega}^{(K)} = \mathbf{0}$. It should be noted that the function \mathbf{h} may be linear (*i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, x_{i,1}, \dots, x_{i,d}]^T$, where $x_{i,j}$ is the j -th component of \mathbf{x}_i) or nonlinear. Kernels, *i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, K_{\mathbf{x}, \mathbf{x}_1}, \dots, K_{\mathbf{x}, \mathbf{x}_l}]^T$, where $K_{\mathbf{x}_i, \mathbf{x}_j} = K(\mathbf{x}_i, \mathbf{x}_j)$ and $K(\cdot, \cdot)$ is some symmetric kernel function, are a relevant example of the nonlinear case. Kernels have been largely used as they tend to improve the data separability in the transformed space. Herein, we use a Gaussian Radial Basis Function (RBF) $K(\mathbf{x}, \mathbf{z}) \equiv -\exp(-\|\mathbf{x} - \mathbf{z}\|^2 / (2\sigma^2))$ kernel, which is widely used in hyperspectral image classification [5]. From now on, d denotes the dimension of $\mathbf{h}(\mathbf{x})$.

¹To keep the notation simple, we use $p(\cdot)$ to denote both probability densities or distributions of random variables.

3.2. The Multi-Level Logistic spatial prior

In the vein of [10, 12, 18], we assume that $p(\mathbf{y})$, the prior on the image class labels is the isotropic MLL

$$p(\mathbf{y}) = \frac{1}{Z} e^{\frac{\mu}{Z} \sum_{i \sim j} \delta(y_i - y_j)}, \quad (3)$$

where Z is a normalizing constant, the notation $i \sim j$ stands for neighboring pixels, δ is the unit impulse function², and $\mu > 0$ is a parameter controlling the likelihood that two neighboring pixels have the same class label. Note that the pairwise interaction terms $\delta(y_i - y_j)$ attach higher probability to equal neighboring labels than the other way around. In this way, the MLL prior promotes piecewise smooth segmentations.

3.3. EM type algorithm to learn the multilogistic regressors

The proposed approach uses the information of both labeled and unlabeled samples to learn the vector of regressors $\boldsymbol{\omega}$. Since our approach is semi-supervised, this estimation is based on a small set of labeled samples ($\mathcal{Y}_L, \mathcal{X}_L$) and a larger set of unlabeled samples \mathcal{X}_U . Assuming a prior $p(\boldsymbol{\omega})$ for the logistic regressors, we have

$$p(\boldsymbol{\omega}, \mathbf{y}|\mathbf{x}) = p(\mathcal{Y}_L, \mathcal{Y}_U|\mathbf{x}, \boldsymbol{\omega})p(\boldsymbol{\omega}). \quad (4)$$

Having in mind expression (1), we may write

$$p(\boldsymbol{\omega}, \mathbf{y}|\mathbf{x}) = \alpha(\mathbf{x})p(\mathcal{Y}_L|\mathcal{X}_L, \boldsymbol{\omega})p(\mathcal{Y}_U|\mathcal{X}_U, \boldsymbol{\omega})p(\boldsymbol{\omega})c(\mathbf{y}) \quad (5)$$

where $c(\mathbf{y})$ depends only on \mathbf{y} .

To compute the MAP estimate of $\boldsymbol{\omega}$, we set forth an EM algorithm, where the set of random variables \mathcal{Y}_U is treated as missing data. At iteration t , the E-step and M-step are then given by

E-step:

$$\begin{aligned} Q(\boldsymbol{\omega}|\hat{\boldsymbol{\omega}}_t) &\equiv E[\log p(\boldsymbol{\omega}, \mathbf{y}|\mathbf{x})|\hat{\boldsymbol{\omega}}_t] \\ &= l_L(\boldsymbol{\omega}) + E[\log p(\mathcal{Y}_U|\mathcal{X}_U, \boldsymbol{\omega})|\hat{\boldsymbol{\omega}}_t] + \log p(\boldsymbol{\omega}) \\ &\quad + \log \alpha(\mathbf{x}) + \log c(\mathbf{y}) \end{aligned} \quad (6)$$

M-step:

$$\hat{\boldsymbol{\omega}}_{t+1} \equiv \arg \max_{\boldsymbol{\omega}} Q(\boldsymbol{\omega}|\hat{\boldsymbol{\omega}}_t), \quad (8)$$

where

$$l_L(\boldsymbol{\omega}) = \sum_{i=1}^L \left(\sum_{k=1}^K y_{i_j}^{(k)} \boldsymbol{\omega}^{(k)} \mathbf{x}_{i_j} - \log \sum_{k=1}^K \exp(\boldsymbol{\omega}^{(k)} \mathbf{x}_{i_j}) \right). \quad (9)$$

If the components of \mathbf{y} are independent, the term $\alpha(\mathbf{x})$ present in (7) does not depend on $\boldsymbol{\omega}$ and thus can be discarded in the M-step. It happens that, in hyperspectral image segmentation, the components of \mathbf{y} are dependent and therefore $\alpha(\mathbf{x})$ does depend on $\boldsymbol{\omega}$. Nevertheless, we still discard it, as it is impossible to compute exactly and we have heuristic evidence that its inclusion does not yield noticeable improvements in the estimates of the vector $\boldsymbol{\omega}$.

By noting that $p(\mathcal{Y}_U|\mathcal{X}_U, \boldsymbol{\omega})$ has the structure of (9) with the sum defined over the unlabeled samples and that the mean value in (6) is with respect to the random variables \mathcal{Y}_U , then we have

$$\begin{aligned} l_U(\boldsymbol{\omega}) &\equiv E[\log p(\mathcal{Y}_U|\mathcal{X}_U, \boldsymbol{\omega})|\hat{\boldsymbol{\omega}}_t] \\ &= \sum_{i=L+1}^{L+U} \left(\sum_{k=1}^K \bar{y}_{i_j}^{(k)} \boldsymbol{\omega}^{(k)} \mathbf{x}_{i_j} - \log \sum_{k=1}^K \exp(\boldsymbol{\omega}^{(k)} \mathbf{x}_{i_j}) \right) \end{aligned} \quad (10)$$

²*i.e.*, $\delta(0) = 1$ and $\delta(y) = 0$, for $y \neq 0$

where

$$\bar{y}_{i_j}^{(k)} \equiv E[y_{i_j}^{(k)} | \mathbf{x}, \hat{\omega}_t] = p(y_{i_j} = k | \mathbf{x}, \hat{\omega}_t). \quad (12)$$

The computation of the conditional marginals $\bar{y}_{i_j}^{(k)}$ is a combinatorial task and thus very hard solve exactly. Herein, we use BP to compute an approximation. BP is an iterative algorithm in which local messages are passed in graphical models [19]. For details of the BP implementation in our graphic model, which is a first order Markov random field, see [18, 19].

As in [9], we adopt the Laplacian density $p(\omega) \propto \exp(-\lambda \|\omega\|_1)$, which enforces sparseness on the elements of ω and therefore generalization capability. The degree of sparseness is controlled by the regularization parameter λ .

Finally, we note that, with the above approximations, the M-step of the proposed EM algorithm is

$$\hat{\omega}_{t+1} = \arg \min_{\omega} \{L_L(\omega) + L_U(\omega) - \lambda \|\omega\|_1\}. \quad (13)$$

which we solve very efficiently with LORSAL [11].

3.4. Energy minimization via graph cuts

As a result of the EM algorithm introduced in the previous section, we get the posterior marginals $p(y_i | \mathbf{x}, \hat{\omega}_t)$ from which we compute, for each pixel, the maximum, thus obtaining the MPM segmentation. In addition, we compute the MAP segmentation given by

$$\hat{\mathbf{y}} = \arg \min_{\mathbf{y} \in \mathcal{L}^n} \left\{ \sum_i -\log p(y_i | \mathbf{x}_i, \hat{\omega}) - \mu \sum_{i,j} \delta(y_i - y_j) \right\}. \quad (14)$$

Optimization (14) is a combinatorial problem, involving unary and pairwise interaction terms, which we solve to a very good approximation using the graph cut α -Expansion based algorithm [20].

4. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed algorithms with a real hyperspectral data set. In all experiments, the spectral vectors are normalized and the RBF scale parameter is set to $\sigma = 0.6$. The prior regularization parameter is set to $\mu = 4$. For the BP algorithm, we consider the first-order neighborhood system [18]. Although these are not optimal choices, they lead to very good results. The training samples are randomly selected from the ground truth image. The remaining samples are used as validation set. Each value of overall accuracy (OA) is obtained from 10 Monte Carlo runs.

We use the well-known AVIRIS Indian Pines scene to evaluate the proposed algorithm. This image was collected over Northwestern Indian in June of 1992 [1]. The scene is available online³, and contains 145×145 pixels and 220 spectral bands in the range of 400-2500nm. Following [16, 21], A subset scene (consisting of pixels in columns [27-94] and rows [31-116]) with size of 68×86 and 4 classes was considered.

In Table 1, we present experimental results for following algorithms: semi-supervised composite kernels [8]; semi-supervised composite PDE [16]; semi-supervised composite wavelet [16]; MLR-EM, multilogistic regression parameterized with the regression vector obtained with the proposed EM algorithm; MPM-EM, proposed MPM segmentation algorithm; MAP-EM, proposed MAP segmentation algorithm; MLR, multilogistic regression obtained by the LORSAL algorithm [11]. All of these algorithms use spectral

³<http://cobweb.ecn.purdue.edu/biehl/MultiSpec/>

and spatial information except the MLR classifier, which only uses spectral information. From Table 1, we conclude the following:

- The inclusion of spatial information largely improves the classification results just based on spectral information as those with [12].
- The inclusion of unlabeled samples significantly improves the segmentation results for very small sizes of the training set. These gains decrease as the size of the training set increases.
- MAP segmentation is marginally better than the MPM segmentation.

An important observation is that the advantages introduced by the proposed semi-supervised algorithm are less relevant as the size of the training set increases. This is expected, since the uncertainty of the classifier boundaries decreases as the training set size increases. For example, using 100 labeled samples per class ($L = 400$), we get an OA=98.87% for $U = 0$ and an OA=98.89% for $U = 7L$, *i.e.*, there is no noticeable gain. This is to be contrasted with the use of just 5 labeled samples per class ($L = 20$), in which we get an OA=84.23% for $U = 0$ and an OA=91.54% for $U = 7L$, *i.e.*, a gain of 7.31%. To give a broader picture of the gains in using unlabeled samples, for small training set sizes, we plot in Figure 1 overall accuracy (OA) results as a function of the number of unlabeled samples using just 3 labeled samples per class.

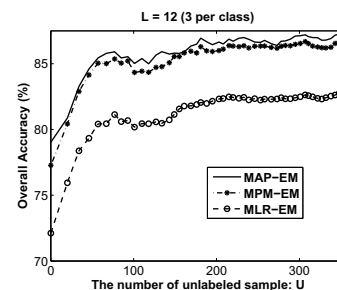


Fig. 1. Overall accuracy (OA) as a function of the unlabeled samples with $L = 12$, that is, 3 labeled sample per class.

Meanwhile, in order to shown the high performance of the proposed supervised algorithms, we report the minimum (min), maximum (max), and average (mean) over 10 runs for the case of 3 labeled samples per class. As we can see, even the minimum OA of the proposed MAP and MPM algorithm is comparable with those obtained by the competitors. This property is definitely relevant when the acquisition of medium to large training sets is costly. The true fact is that the proposed approach yields very competitive results with regards to other state-of-the-art approaches. Either we obtain comparable results with less training samples or we obtain better results with the same size of the training set. For instance, with just 10 labeled samples per class ($L = 40$) and $U = 7L$ unlabeled samples, both of the proposed MAP segmentation and MPM algorithm obtained OAs comparable with those presented in [8, 16] using 100 labeled samples per class ($L = 400$). For illustrative purposes, Figure 2 show the classification and segmentation maps obtained using $L = 12$ (3 per class), and $U = 102$.

5. CONCLUSIONS

This paper introduced a new semi-supervised hyperspectral image segmentation algorithm in which the final segmentation is inferred

Table 1. OA [%] results as a function of the number of labeled samples in the subset Indiana Pines image. L denotes the number of labeled samples. U denotes the number of unlabeled samples. The best results (in bold typeface) are highlighted for each problem.

Algorithms	Number of labeled samples per class									
	3			5	10	15	20	25	30	100
	min	max	mean							
MPM-EM ($U = 0$)	59.99	84.04	73.14	84.06	89.49	93.91	94.61	95.85	96.05	98.82
MPM-EM ($U = 7L$)	78.13	88.47	84.59	91.25	93.01	95.19	96.36	97.02	97.21	98.88
MAP-EM ($U = 0$) [12]	59.21	84.84	73.49	84.23	90.09	95.07	94.88	96.24	96.34	98.87
MAP-EM ($U = 7L$)	80.32	88.81	85.32	91.54	93.23	96.08	96.52	97.08	97.46	98.89
MLR [12] ($U = 0$)	55.42	79.23	69.07	77.99	83.97	88.85	88.91	90.42	90.52	94.91
MLR-EM ($U = 7L$)	70.85	85.71	80.74	86.66	89.13	92.88	91.96	94.19	94.43	96.46
Composite Wavelet [16]		73.65		78.78	82.90	85.74	86.85	87.69	88.68	92.59
Composite PDE [16]		84.89		86.89	90.03	90.51	91.33	92.67	93.74	94.20
Composite kernels [8]		66.73		67.13	71.32	79.49	82.04	83.12	84.99	86.44

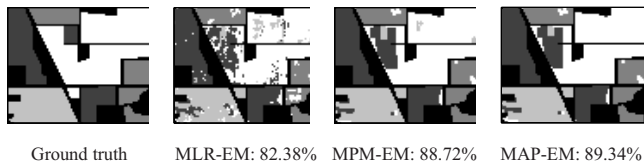


Fig. 2. Classification results obtained by the proposed algorithm for the subset image with $L = 12$ (3 per class), $U = 102$.

from a posterior probability distribution, taking into account both the spectral information (through a multinomial logistic regression), and the spatial information (through a multilevel logistic Markov random field). The parameters of the complete model are inferred, from labeled and unlabeled samples, via Bayesian inference using expectation maximization and belief propagation tools. The reported experimental results exhibit state-of-the-art performance. We highlight the excellent segmentation result obtained with just a few training samples per class. As future work, we will address inference on the the model hyperparameters aiming at more adaptiveness, and will produce a more comprehensive experimental characterization.

6. REFERENCES

- [1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley, Hoboken, NJ, 2003.
- [2] B. Krishnapuram, D. Williams, Y. Xue, A. Hartemink, L. Carin, and M. Figueiredo, "On semi-supervised classification," in *Proc. 18th Annual Conference on Neural Information Processing Systems*, Vancouver, Canada, 2004.
- [3] O. Chapelle, M. Chi, and A. Zien, "A continuation method for semi-supervised svms," in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 185–192, ACM Press.
- [4] V. Vapnik, *Statistical Learning Theory*, John Wiley, New York, 1998.
- [5] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, pp. 1351–1362, 2005.
- [6] B. Scholkopf and A. Smola, *Learning With Kernels Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press Series, Cambridge, MA, 2002.
- [7] D. Böhning, "Multinomial logistic regression algorithm," *Annals of the Institute of Statistical Mathematics*, vol. 44, pp. 197–200, 1992.
- [8] G. Camps-Valls, T. Bandos, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, pp. 3044–3054, Oct 2007.
- [9] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 957–968, 2005.
- [10] J. Borges, J. Bioucas-Dias, and A. Marçal, "Evaluation of Bayesian hyperspectral imaging segmentation with a discriminative class learning," in *Proc. IEEE International Geoscience and Remote sensing Symposium*, Barcelona, Spain, 2007.
- [11] J.M. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented lagrangian tools," Tech. Rep., Instituto Superior Técnico, TULisbon, 2009.
- [12] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image classification based on a markov random field and sparse multinomial logistic regression," in *Proc. IEEE International Geoscience and Remote sensing Symposium*, 2009.
- [13] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral classification," in *First IEEE GRSS Workshop on Hyperspectral Image and Signal Processing-WHISPERS'2009*, 2009.
- [14] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Transactions on Geoscience and Remote Sensing (submitted)*, 2009.
- [15] M. Fauvel, J.A. Benediktsson, J. Chanussot, and J.R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, 2008.
- [16] S. Velasco-Forero and V. Manian, "Improving hyperspectral image classification using spatial preprocessing," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, pp. 297–301, 2009.
- [17] Devis Tuia and Gustavo Camps-Valls, "Semi-supervised hyperspectral image classification with cluster kernels," *IEEE Geoscience and Remote Sensing Letters. Accept for publication*, 2009.
- [18] S. Geman and D. Geman, "Stochastic relaxation, gibbs distribution, and the bayesian restoration of images," *IEEE TPAMI*, vol. 6, pp. 721–741, 1984.
- [19] J. Yedidia, W. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," in *Proceedings of International Joint Conference on Artificial Intelligence*, 2001.
- [20] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1222–1239, November 2001.
- [21] Gustavo Camps-Valls, Luis Gomez-Chova, Jordi Muoz-Mar, Joan Vila-Francis, and Javier Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, Jan 2006.