SUPERVISED HYPERSPECTRAL IMAGE SEGMENTATION USING ACTIVE LEARNING

Jun Li , José M. Bioucas-Dias

Antonio Plaza

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

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

ABSTRACT

This paper introduces a new supervised Bayesian approach to hyperspectral image segmentation. The algorithm mainly consists of two steps: (a) learning, for each class label, the posterior probability distributions, based on a multinomial logistic regression model; (b) segmenting the hyperspectral image, based on the posterior probability distribution of the image of class labels built on the learned pixelwise class distributions and on a multi-level logistic prior encoding the spatial information. Aiming at reducing the costs of acquiring large training sets, we use active label selection based on the *the posterior marginals of the complete model* provided by Belief propagation. A comparison of the proposed method with state-of-the-art competitors shows its effectiveness.

Index Terms— Hyperspectral image segmentation, multinomial logistic regression, spatial information, Markov random field, active label selection, belief propagation.

1. INTRODUCTION

In recent years, with the development of remote sensing sensors, hyperspectral images are widely available. The special characteristics of hyperspectral data sets bring difficult processing problems. For example, the Hughes phenomenon [1] comes out as the data dimensionality increases. In order to get an acceptable classification accuracy, large amount of training samples are required, which may be quite difficult, expensive, or sometimes impossible to get. These difficulties have fostered the development of new classification methods, which are able to deal with ill-posed classification problems, in particular, high dimensional datasets and limited training samples [1]. For instance, several machine learning techniques have been applied to extract relevant information from hyperspectral data sets [2,3]. However, although many progresses have been made, the difficulty in learning high dimensional densities from a limited number of training samples is still an active area of research.

Possible approaches which are capable to circumvent this kind of difficulties are the discriminative approach, which learns the class distributions in high dimensional spaces by inferring the boundaries between classes in the feature space [4, 5]. For instance, the support vector machines (SVMs) [6] are among the state-of-the-art discriminative techniques in ill-posed classification problems. Due to their ability to deal with large input spaces efficiently and to produce sparse solutions, SVMs have been successfully used for hyperspectral supervised and semi-supervised classification with limited training samples [7–11]. The multinomial logistic regression (MLR) [12] also shows high quality while dealing with ill-posed problems, with

the advantage over the SVMs of learning the class probability distributions themselves. Effective Sparse multinomial logistic regression (SMLR) methods are available [13, 14]. These ideas have been applied to hyperspectral image classification [3, 15–17] leading to state-of-the-art performance.

In order to improve the classification accuracies obtained by SVMs and MLR-based techniques, a recent trend is to integrate spatial contextual information with spectral information in hyperspectral data interpretation [3, 8, 10]. These methods exploit, in a way or another, the continuity, in probability sense, of neighboring labels: it is very likely that, in an hyperspectral image, two neighboring pixels have the same label.

In this paper, we introduce a new Bayesian segmentation approach which exploits the spatial contextual information and implements active learning. The algorithm implements two main steps: a) learning the posterior class probability distributions by estimating, with the LORSAL algorithm [18], the parameters of an MLR model; b)segmenting the hyperspectral image by inferring the image class labels from a posterior distribution built on the learned MLR model and on a multi-level logistic (MLL) prior. Active label selection [19, 20] based on the posterior marginals of the complete model, provided by the Belief propagation (BP) algorithm, is implemented. In comparison with our previous method [21], where the active learning depends only on the spectral information, our present active leaning approach uses both spectral and spatial information leading to better performances, as shown in Section 4.

The remainder of the paper is organized as follows. Section 2 formulates the problem. Section 3 describes the proposed approach. Section 4 illustrates the active selection approach. Section 5 reports segmentation results based on real hyperspectral datasets; in comparison with state-of-the-art competitors are also included. Finally, section 6 concludes with some remarks.

2. PROBLEM FORMULATION

First, let us define the following notations used in this paper:

$$
\begin{array}{ll}
\mathcal{S} \equiv \{1, \ldots, n\} & \text{Set of integers indexing the } n \text{ pixels of an image} \\
\mathcal{L} \equiv \{1, \ldots, K\} & \text{Set of } K \text{ class labels} \\
\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{d \times n} & \text{Image of feature } d\text{-dimensional vectors} \\
\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{L}^n & \text{Image of class labels}
\end{array}
\tag{1}
$$

With the above definitions in place, the goal of classification is to assign a label $y_i \in \mathcal{L}$ to each $i \in \mathcal{S}$, based on the vector $\mathbf{x}_i$, resulting in an image of class labels $\mathbf{y}$. We call this assignment a *labeling*. The goal of segmentation is, based on the observed image $\mathbf{x}$, to compute a partition $\mathcal{S} = \cup_i \mathcal{S}_i$ of the set $\mathcal{S}$ such that the pixels in each element of the partition share some common property, for example to represent the same type of land cover. Notice that, given a labeling $\mathbf{y}$, the collection $\mathcal{S}_k = \{i \in S \mid y_i = k\}$, for $k = 1, \ldots, K$

is a partition of $\mathcal{S}$. On the other way around, given the segmentation $\mathcal{S}_k$, for $k = 1, \ldots, K$, the image $\{y_i \mid y_i = k \text{ if } i \in \mathcal{S}_k, \ i \in \mathcal{S}\}$ is a labeling. There is, therefore, a one-to-one relation between labelings and segmentations. Nevertheless, in this paper, we use the term classification when there is no spatial information and segmentation when the spatial prior is being considered.

Inference in a Bayesian framework is often carried out by maximizing the posterior distribution[1]

$$p(\mathbf{y}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{y})p(\mathbf{y}), \tag{2}$$

where $p(\mathbf{x}|\mathbf{y})$ is the likelihood function (*i.e.*, the probability of feature image given the class labels) and $p(\mathbf{y})$ is the prior over the image of labels. Assuming conditional independency of the features given the class labels, *i.e*, $p(\mathbf{x}|\mathbf{y}) = \prod_{i=1}^{i=n} p(\mathbf{x}_i|y_i)$, then the posterior probability $p(\mathbf{y}|\mathbf{x})$, as a function of $\mathbf{y}$, may be written as

$$
\begin{aligned}
p(\mathbf{y}|\mathbf{x}) &= \frac{1}{p(\mathbf{x})} p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \\
&= \alpha(\mathbf{x}) \prod_{i=1}^{i=n} \frac{p(y_i|\mathbf{x}_i)}{p(y_i)} p(\mathbf{y}),
\end{aligned}
\tag{3}
$$

where $\alpha(\mathbf{x}) \equiv \prod_{i=1}^{i=n} p(\mathbf{x}_i)/p(\mathbf{x})$ is a factor not depending on $\mathbf{y}$. In this paper we assume, without loss of generality, that $p(y_i) = 1/K$. The maximum a posteriori (MAP) segmentation is then given by

$$\widehat{\mathbf{y}} = \arg \max_{\mathbf{y} \in \mathcal{L}^n} \left\{ \left(\sum_{i=1}^{n} p(y_i|\mathbf{x}_i) \right) + \log p(\mathbf{y}) \right\}. \tag{4}$$

3. PROPOSED APPROACH

In the present approach, the probability distributions $p(y_i|\mathbf{x}_i)$ are modeled with the MLR [12], which writes as

$$p(y_i = k|\mathbf{x}_i, \boldsymbol{\omega}) = \frac{\exp(\boldsymbol{\omega}^{(k)}\mathbf{h}(\mathbf{x}_i))}{\sum_{k=1}^{K} \exp(\boldsymbol{\omega}^{(k)}\mathbf{h}(\mathbf{x}_i))}, \tag{5}$$

where $\mathbf{h}(\mathbf{x}_i) = [h_1(\mathbf{x}_i), ..., h_l(\mathbf{x}_i)]^T$ is a vector of l fixed functions of the input feature vectors, often termed features vectors as well; $\boldsymbol{\omega}$ is the regressors, and $\boldsymbol{\omega} = [\boldsymbol{\omega}^{(1)^T}, ..., \boldsymbol{\omega}^{(K-1)^T}]^T$; because the density (5) does not depend on translations on the regressors $\boldsymbol{\omega}^{(k)}$, we take $\boldsymbol{\omega}^{(K)} = \mathbf{0}$. Function $\mathbf{h}$ may be linear (*i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, x_{i,1}, ..., x_{i,d}]^T$, where $x_{i,j}$ is the j-th component of $\mathbf{x}_i$) or nonlinear. Kernels, *i.e.*, $\mathbf{h}(\mathbf{x}_i) = [1, K_{\mathbf{x}, \mathbf{x}_1}, ..., K_{\mathbf{x}, \mathbf{x}_l}]^T$, 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, which are largely used because they tend to improve the data separability in the transformed space. In this paper, we use a Gaussian Radial Basis Function (RBF) $K(\mathbf{x}, \mathbf{z}) = -\exp\left(\|\mathbf{x} - \mathbf{z}\|^2/(2\sigma^2)\right)$ kernel, which is widely used in hyperspectral image classification [9]. From now on, d denotes the dimension of $\mathbf{h}(\mathbf{x})$.

We formulate the inference of vector $\boldsymbol{\omega}$ parameterizing the MLR (5) as in [13]. Given $\mathcal{D}_L \equiv \{(\mathbf{x}_{i_1}, y_{i_1}), \ldots, (\mathbf{x}_{i_L}, y_{i_L})\}$, a training set, we compute a MAP estimate of the vector $\boldsymbol{\omega}$ based on a Laplacian prior. This prior promotes sparseness on the components of $\boldsymbol{\omega}$, forcing many components to be zero. In this way the machine complexity is controlled thus ensuring generalization capability. To compute the MAP estimate of $\boldsymbol{\omega}$, we use the LORSAL algorithm [18],

which is able to solve problems far from the reach of the SMLR algorithm introduced in [13].

The prior probability distribution $p(\mathbf{y})$ is the MLL MRF [17]

$$p(\mathbf{y}) = \frac{1}{Z} e^{\displaystyle \mu \sum_{i \sim j} \delta(y_i - y_j)}, \tag{6}$$

where Z is a normalizing constant, $i \sim j$ denotes first order neighboring sites, $\delta(y)$ is the unit impulse function[2], and $\mu > 0$ is a parameter controlling the likelihood that two neighboring pixels belong to the same class. 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.

The MAP segmentation is finally given by

$$\widehat{\mathbf{y}} = \arg \min_{\mathbf{y} \in \mathcal{L}^n} \left\{ \sum_{i \in \mathcal{S}} - \log p(y_i|\mathbf{x}_i, \boldsymbol{\omega}) - \mu \sum_{i \sim j} \delta(y_i - y_j) \right\}. \tag{7}$$

The minimization (7) is a hard combinatorial optimization problem. However, given that the pairwise interaction term on the right hand side of (4) is a metric, we apply the α-Expansion graph cut based algorithm [22], which yields exact results in binary problems and very good approximations otherwise.

4. ACTIVE LEARNING

In order to reduce the size of the training set, we implement active query selection. The basic idea of active learning is that of iteratively enlarging the training set by requesting an expert to, in each iteration, label feature vectors from the set of unlabeled feature vectors $\{\mathbf{x}_i, i \in \mathcal{S}_U\}$, where $\mathcal{S}_U$ is the set of unlabeled image pixels. The relevant question is, of course, what samples should be chosen. In this paper, following [19], we iteratively select the label which contains the maximum information with respect to the actual random vector of MLR regressors $\boldsymbol{\omega}$:

$$i^* = \arg \max_{i \in \mathcal{S}_U} I(\boldsymbol{\omega}; y_i), \tag{8}$$

where $I(\mathbf{X}; \mathbf{Y})$ stands for the mutual information between the random vectors $\mathbf{X}$ and $\mathbf{Y}$. Using the Laplace approximation $p(\boldsymbol{\omega}|\mathbf{x}) \simeq \mathcal{N}(\boldsymbol{\omega}|\widehat{\boldsymbol{\omega}}, \mathbf{H}^{-1})$, where $\mathbf{H}$ is the posterior precision matrix, and assuming that the MAP estimate $\widehat{\boldsymbol{\omega}}$ remains unchanged after including $\mathbf{y}_i^*$, then we have (see [19, 20] for more details)

$$I(\boldsymbol{\omega}; \mathbf{y}_{i^*}) \simeq \frac{1}{2} \log \left(1 + \prod_{i=1}^{K} p(y_{i^*}|\mathbf{x}, \mathcal{D}_L, \widehat{\boldsymbol{\omega}}) \, \mathbf{x}_{i^*}^T \mathbf{H}^{-1} \mathbf{x}_{i^*} \right). \tag{9}$$
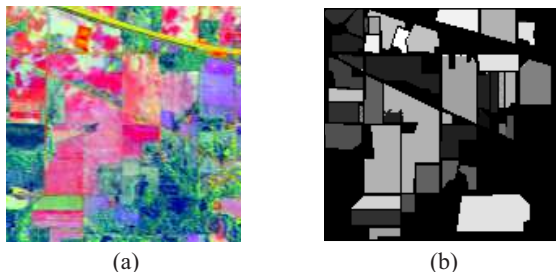
A straightforward calculus leads us to the conclusion that $I(\boldsymbol{\omega}; \mathbf{y}_i)$ is maximized when $p_i = 1/K$, *i.e.*, for class labels y_i with maximum entropy, which correspond to those near the classifier boundaries. In order to find the maximum entropy labels, we use the belief propagation (BP) algorithm [5, 23] to compute the marginal probability distributions $p(y_i|\mathbf{x}, \widehat{\boldsymbol{\omega}})$ from the joint probability distribution $p(\mathbf{y}|\mathbf{x}, \widehat{\boldsymbol{\omega}})$. In this way, we are implicitly including the spatial information, what produces considerable improvements with respect to a scenario in which the active learning is based only on spectral information, as we have done in our previous work [21].

The pseudo-code for the proposed algorithm is presented below.

---

[1]To keep the notation simple, we use $p(\cdot)$ to denote both continuous probability densities and discrete probability distributions of random variables. The meaning should be clear from the context.

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

Algorithm 1 Supervised segmentation algorithm using active label selection

1: **while** The stop criterion is not fulfilled **do**
2:   Learn the MLRs $\boldsymbol{\omega}$ parameterizing $p(\mathbf{y}_i|\mathbf{x}_i,\boldsymbol{\omega})$ by using LORSAL algorithm according to (5).
3:   Use the MLL prior $p(\mathbf{y})$ according to (6).
4:   Estimate the MAP solution using α-Expansion graph cut based algorithm.
5:   Compute the marginals $p(y_i|\mathbf{x},\widehat{\boldsymbol{\omega}})$ by using BP.
6:   Order the set $\mathcal{S}_U$ by decreasing entropy of labels y_i for $i \in \mathcal{S}_U$ and label the first p feature vectors.
7: **end while**

(a)   (b)

Fig. 1. (a) False color composition of the AVIRIS Indian Pines scene. (b) Segmentation map with OA = 98.69%.

The active selection criterion described above considers just one labeling per iteration. Since we set $p > 1$ in line 6 of the proposed algorithm, we are labeling more than one sample per iteration. This is, of course, a sub-optimal procedure. Nevertheless, we found out experimentally that it still leads to very good results with the advantage of being p times faster. More, " The stop criterion" mentioned in Algorithm 1 stands for the criterion to exhaust the supervised algorithm, *i.e.*, the maximum number of iterations. In this paper, the maximum size of the training set considered was used as the stop criterion.

5. EXPERIMENTAL RESULTS

This section shows the effectiveness of the proposed method in hyperspectral remote sensing image segmentation. 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 $\mu = 10$. Although these values for the parameters are not optimal choices, they lead to very good results and, of course, leave room for more improvements. In each experiment, the initial labeled set, randomly selected from the ground truth image, is set to half of its final value. The active selection procedure takes 4 iterations. Each value of overall accuracy (OA) was obtained from 10 Monte Carlo runs.

The well-known AVIRIS Indian Pines scene was used to evaluate the proposed algorithm. This image was collected over Northwestern Indiana in June of 1992 [24]. This scene is available online[3], containing 145×145 pixels and 220 spectral bands in the range of 400-2500nm. Following [2, 21, 25, 26], two scenarios were considered in our experiments. In the first experiment, the whole image of 145×145 pixels, 16 classes and 224 spectral bands was considered, as shown in Figure 1 (a). The second scenario is a subset scene (consisting of pixels in columns [27-94] and rows [31-116]) with size of 68×86 and contains 4 classes.

Table 1 shows the OA results from the proposed supervised algorithm over both images in comparison with the results published

Table 1. OA [%] results over the subset and the complete AVIRIS image. Size of the training set: 2073 labeled samples (20% of the ground truth containing 10366 samples samples) for the whole image; 878 labeled samples (20% of the ground truth containing 4393 samples samples) for the subset. Best results (Bold) are highlighted for each problem.

Classifier	Subset	Whole
Euclidean [27]	67.43	48.23
BLOOC+DAFE+ECHO [27]	93.50	82.91
Composite Kernel [2]	98.86	96.53
Composite Kernel using Wavelet smoothing [26]	98.96	97.85
Composite Kernel using PDE smoothing [26]	98.83	93.62
LORSAL	96.05	84.51
Supervised sementation: LOSAL + MLL [17]	98.11	94.36
LORSAL with active learning [21]	97.56	86.83
LORSAL + MLL with active learning [21]	98.70	97.89
Proposed LORSAL with active learning	97.62	86.94
Proposed LORSAL + MLL with active learning	**99.06**	**98.58**

in [2, 17, 21, 26, 27] for a final training set with 20% (2073 for the whole image and 878 for the subset) of the ground truth. The proposed algorithm outperforms all the competitors. We would like to stress the gains over our previous work presented in [21]. As expected, the active learning based on the marginals of the complete probability distribution $p(\mathbf{y}|\mathbf{x},\boldsymbol{\omega})$ is more informative than that just based on the $p(y_i|\mathbf{x}_i,\boldsymbol{\omega})$, which includes only spectral information. For illustration purpose, Figure1 (b) shows the segmentation map of the full image with an OA of 98.69%. Effective result can be see from this figure.

Table 2 presents the classification results as functions of the number of labeled samples over the subset image. We have not applied active selection for 3 and 5 labeled samples per class because, the initial training set would be very small, just 1 and 2 samples per class, respectively, what would lead to a poor initialization of the active selection procedure. Anyway, the results produced without active selection are those of [17]. The results are compared with state-of-the-art classifiers [17, 18, 21, 25, 26]. Again, the proposed method outperforms the competitors in all cases, the advantage increasing as the size of the training set decreases. This is a relevant property when the acquisition of large training sets is costly.

We stress that the performance of the proposed algorithm depends on the size of samples which are actively selected and on the number p of samples actively selected per iteration. For instance, we run experiments over the subset with 50 labeled samples in total, in which 45 samples are actively selected. An OA of 99.32% is obtained by the proposed segmentation algorithm with active selection. In regarding to number of samples actively selected per iteration, we made experiments over the whole image with 20% of the ground truth labels used as the training set, half of which were considered for active selection, which is the same as experiment 1. The difference is, we used 9 iterations to exhaust the training set. An OA of 98.74% was obtained, which is a little better than 98.58% obtained in experiment 1 using 4 iterations to exhaust the training set.

6. CONCLUSIONS

This work has presented a new supervised approach for hyperspectral classification, which combines the spectral information, modeled with multinomial logistic regression, and spatial information, mod-

Table 2. OA [%] results as a function of the number of labeled samples per class in the subset. Best results (bold) are highlighted for each problem.

Algorithms	number of labeled samples per class							
	3	5	10	15	20	25	30	100
Propose segmentation with active selection	-	-	**94.25**	**96.58**	**97.01**	**97.39**	**97.49**	**98.70**
LORSAL with Proposed active selection	-	-	85.94	89.47	90.92	91.29	92.49	96.32
Segmentation with active selection [21]	-	-	92.17	95.08	96.71	96.98	97.37	98.26
LORSAL with active selection [21]	-	-	84.05	88.01	90.56	91.15	92.11	96.14
Supervised segmentation [17]	82.80	87.51	92.33	94.37	95.51	96.09	96.75	97.55
LORSAL	74.01	77.51	83.43	86.88	88.71	90.10	91.28	94.67
Wavelet [26]	73.65	78.78	82.90	85.74	86.85	87.69	88.68	92.59
PDE [26]	84.89	86.89	90.03	90.51	91.33	92.67	93.74	94.20
Semi-supervised algorithm [25]	66.73	67.13	71.32	79.49	82.04	83.12	84.99	86.44

eled with a multi-level logistic prior. Active query selection is considered. The results obtained in set of experiments using the AVIRIS Indiana Pines data set are state-of-the-art, outperforming the competitor algorithms [2, 21, 25, 26].

7. REFERENCES

[1] G.F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory, IT*, vol. 14, no. 1, pp. 55–63, 1968.

[2] Gustavo Camps-Valls, Luis Gomez-Chova, Jordi Muoz-Mar, Joan Vila-Francs, and Javier Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, Jan 2006.

[3] 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.

[4] V. Vapnik, *Statistical Learning Theory*, John Wiley, New York, 1998.

[5] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 1st edition, 2007.

[6] B. Scholkopf and A. Smola, *Learning With KernelsSupport Vector Machines, Regularization, Optimization and Beyond*, MIT Press Series, Cambridge, MA, 2002.

[7] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive svm for the semisupervised classification of remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3363–3373, 2006.

[8] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. 110–122, September 2009.

[9] 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.

[10] 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.

[11] M. Chi and L. Bruzzone, "A semi-labeled-sample driven bagging technique for ill-posed classification problems," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 1, pp. 69–73, 2005.

[12] D. Böhning, "Multinomial logistic regression algorithm," *Annals of the Institute of Statistical Mathematics*, vol. 44, pp. 197–200, 1992.

[13] 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.

[14] Kwangmoo Koh, Seung-Jean Kim, and Stephen Boyd, "An interior-point method for large-scale l1-regularized logistic regression," *J. Mach. Learn. Res.*, vol. 8, pp. 1519–1555, 2007.

[15] J. Borges, J. Bioucas-Dias, and A. Marçal, "Fast sparse multinomial regression applied to hyperspectral data," in *International Conference on Image Analysis and - ICIAR*, 2006.

[16] 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.

[17] 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.

[18] J.M. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented lagrangian tools," Tech. Rep., Instituto Superior Técnico, TULisbon, 2009.

[19] D. Mackay, "Information-based objective functions for active data selection," *Neural Computation*, vol. 4, pp. 590–604, 1992.

[20] 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.

[21] J. Bioucas-Dias J. Li and Antonio Plaza, "Semi-supervised hyperspectral classification and segmentation with discriminative learning," in *SPIE Europe Remote Sensing*, 2009.

[22] 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.

[23] J. Yedidia, W. Freeman, and Y.Weiss, "Understanding belief propagation and its generalizations," in *Proceedings of International Joint Conference on Artificial Intelligence*, 2001.

[24] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley, Hoboken, NJ, 2003.

[25] 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.

[26] 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.

[27] P. Mrazek and J. Weickert, *Classification of high dimensional data with limited training samples*, Ph.D. thesis, Purdue University, 1998.