# SEMI-SUPERVISED HYPERSPECTRAL IMAGE CLASSIFICATION BASED ON A MARKOV RANDOM FIELD AND SPARSE MULTINOMIAL LOGISTIC REGRESSION

*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

## 1. INTRODUCTION

The Bregman iteration-based SMLR method (Bregman-SMLR) recently introduced in [1] shows good performance in supervised learning. It allows to replace a difficult non-smooth convex problem with a sequence of quadratic plus diagonal $l_2$-$l_1$ problems, which are very easy to solve. In previous studies, the application of Bregman-SMLR to large datasets, such as hyperspectral imagery, shows state-of-the-art results in terms of overall accuracy, computational cost, size of the training set, and robustness to noise[2].

This paper presents a semi-supervised classification/segmentation approach based on the discriminative Bregman-SMLR classifier and on a Markov random filed (MRF) prior, more precisely, a Multi-Level Logistic (MLL) Markov-Gibbs prior. The effectiveness of the proposed methodology is demonstrated by classifying simulated data sets and real AVIRIS hyperspectral imagery. A comparison with the Camps-Valls state-of-art method [3] is also included.

## 2. PROPOSED METHOD

### 2.1. Supervised Segmentation

Let $y = \{y_i\}_{i \in \mathcal{S}}$, where $y_i \in \mathcal{L} = \{1, 2, ..., K\}$ and $i \in \mathcal{S} = \{1, \ldots, n\}$, denotes an image of labels and $x = \{x_i \in \mathbb{R}^d, i \in \mathcal{S}\}$ denotes the observed hyperspectral image. The goal of classification is to estimate $y$ from the observed image $x$. We adopt the Bayesian maximum a posterior probability (MAP) criterion, which amounts to maximize, with respect to $y$, the posterior density

$$p(y|x) \; \propto \; p(x|y)p(y), \tag{1}$$

where $p(x|y)$ is the probability of the feature image giving labels, or the likelihood function, and $p(y)$ is the prior over the classes. In the vein of work [4], we learn $p(x|y)$ via the fast sparse multinomial logistic regression algorithm [5] and adopt as spatial prior a MLL Markov-Gibbs distribution [6]; *i.e.*, we have $p(y) = \frac{1}{Z}e^{\beta n(y)}$, where $Z$ is the normalizing constant, $\beta$ is a nonnegative constant, and $n(y)$ denotes the number of pixel neighbors having the same label. Then, the MAP segmentation is given by

$$
\begin{aligned}
\hat{y} &= \; \arg \max_{y \in \mathcal{L}^n} p(x|y)p(y) \\
&= \; \arg \max_{y \in \mathcal{L}^n} \sum_{i \in \mathcal{S}} \log p(x_i|y_i) + \beta n(y) \\
&= \; \arg \min_{y \in \mathcal{L}^n} \sum_{i \in \mathcal{S}} - \log p(x_i|y_i) - \beta n \sum_{i \sim j, \; i,j \in \mathcal{S}} \delta(y_i - y_j),
\end{aligned}
\tag{2}
$$

where $i \sim j$ denotes neighboring sites.

The minimization (2) is a hard combinational optimization problem. However, given that the pairwise interaction term on the right hand side of (2) is a metric, it's possible to achieve a very good approximation using the alpha-expansion graph cut algorithm [7].

### 2.2. Semi-supervised Segmentation

Very often, the acquisition of labeled samples is very hard and/or expensive. In these cases, we resort to semi-supervised classification/segmentation approaches, which are based on very few training samples. In our setup, we run a two-step iterative procedure: in the first step, we apply the Bregman-SMLR algorithm recently introduced in [1], to learn the logistic regressors; in the second step we compute the MAP segmentation (2) and expand the set of labeled sampled with a subset of the just obtained labeled samples. Different active [8] strategies are possible to choose this kind of samples, such as as random, neighboring, based of mutual information , etc.

The pseudo-code for the semi-supervised algorithm is presented below, where $\beta$ is the prior strength parameter, $T$ is the initial training set, $t$ is the selected labeled sets from the segmentation result of last iteration, $Q$ is the selection criterion, $l$ is the log-likelihood, $MMiter$ is the maximum number of iterations, and $tol$ is the tolerance to control the log-likelihood.

## 3. EXPERIMENT RESULTS

### 3.1. Simulated Dataset

Simulated scenes of dimension $100 \times 100 \times 221$ ($100 \times 100$ is the spatial size and 221 is the number of bands), and 10 classes are generated. Each scene is composed of spectral vectors $x = z + n$, where $x \in \mathbb{R}^{221}$, $z$ are spectral signatures obtained from the USGS library and $n$ is

**Algorithm 1** Semi-supervised Segmentation based on Bregman-SMLR

---

**Require:** $\beta, T$
  **while** $k \leq MMiter$ or $\|l - l_{k-1}\| \geq tol$ **do**
    $t \quad\quad \Leftarrow Q(y^{(k)})$
    $T \quad\quad \Leftarrow T \cup t$
    $\hat{y}^{(k+1)} = \arg\max_{y} p(x|y)p(y)$
  **end while**

---

zero-mean white Gaussian. The signal-to-noise ratio is defined as SNR$= E[\|z\|^2]/E[\|n\|^2]$. Fig.1, left, shows the overall accuracy (OA) as a function of the SNR. The initial training set is just $0.1\%$ of the whole scenario (10 pixels) randomly selected; the remaining samples were used as test set. Fig. 1, middle, shows OA results as a function of the initial training set. The selection criterion is random. Each value of OA was obtained from 10 Monte Carlo runs. In side of using a much smaller training set, the semi-supervised algorithm outperforms the supervised counterparts in both plots.

### 3.2. Real AVIRIS Image

Fig. 1, right, illustrates the results on the AVIRIS spectrometer image, the Indian Pines 92 from Northern Indiana, taken on June 12, 1992, with $145 \times 145$ pixels and 220 spectral bands. Experiments were carried out without 20 noisy bands.For comparison with [3] purposes, seven classes were discarded, leaving a dataset with 9 classes distributed by 9345 elements. This dataset was randomly partitioned into a set of 4757 training samples and 4588 validation samples. During the experiments, different number of initial training samples is randomly selected from the training set, as shown in Fig. 1, right. Random selection and neighborhood method criterion is used. Very good results are obtained using both methods. While the training samples is 52, around 5 pixels per class, the accuracies of both method are higher than $76\%$,, while the best accuracy presented in [3] is $66.04\%$, around $10\%$ less than our algorithm .
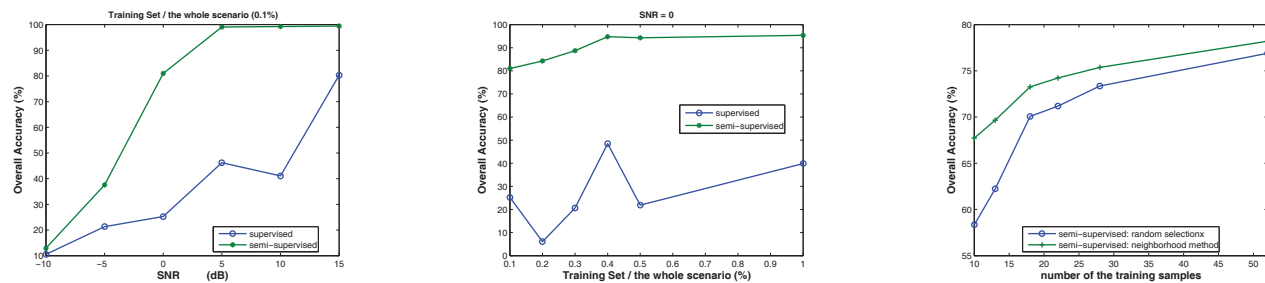


**Fig. 1**. Semi-supervised segmentation results.

## 4. CONCLUDING REMARKS

In this paper, a semi-supervised classification/segmentation version of the supervised MRF and Bregman-SMLR algorithm is introduced. The information of the unlabeled samples is actively included in the supervised version. In this way the class densities are iteratively inferred by adding samples labeled in the previous step, thus, improving the classification overall accuracy. More experiments and results will be shown at the conference.

## 5. REFERENCES

[1] J. Bioucas-Dias, "Bregman-smlr. a fast sparse logistic regression algorithm," Tech. Rep., Instituto Superior Técnico, TULisbon, 2008.

[2] J. Li, J. Bioucas-Dias, and Antonio Plaza, "Hyperspectral image classification based on a fast bregman sparse multinomial logistic regression algorithm," in *6th EARSeL SIG IS Workshop*, 2009.

[3] 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, 2007.

[4] J. Borges, J. Bioucas-Dias, and A. Marçal, "Bayesian hyperspectral image segmentation with discriminative class learning," in *Pattern Recognition and Image Analysis: 3rd Iberian Conference, IbPRIA*, Girona, Spain, 2007.

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

[6] S. Geman and D. Geman, "Stochastic relaxation, gibbs distribution, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[7] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1222–1239, 2001.

[8] David J. C. MacKay, "Information-based objective functions for active data selection," *Neural Computation*, vol. 4, pp. 590–604, 1992.