

Semi-supervised hyperspectral classification using active label selection

Jun Li, José Bioucas-Dias¹ and Antonio Plaza²

¹Instituto de Telecomunicações, Instituto Superior Técnico, Technical University of Lisbon, 1049-001, Lisboa, Portugal;

²Department of Technology of Computers and Communications, University of Extremadura, E-10071 Cáceres, Spain

ABSTRACT

This paper introduces a new semi-supervised Bayesian approach to hyperspectral image segmentation. The algorithm mainly consists of two steps: (a) semi-supervised learning, by using the LORSAL algorithm to infer the class distributions, followed by (b) segmentation, by inferring the labels from a posterior density built on the learned class distributions and on a Markov random field. Active label selection is performed. Encouraging results are presented on real AVIRIS Indiana Pines data set. Comparisons with state-of-the-art algorithms are also included.

Keywords: hyperspectral classification, graph cut, multinomial logistic regression, MLL Markov-Gibbs prior, active label selection

1. INTRODUCTION

In the recent years, the wide availability of hyperspectral images led to new developments in the fields of image classification and segmentation.¹ The high resolution of spectral signatures provided by hyperspectral sensors requests new techniques capable of properly handling the high dimensionality of the data. The difficulties in learning high dimensional densities from a limited number of training samples (Hughes phenomenon) is one of the major problems related with this type of data, which has fostered the development of new algorithms, and although many progresses have been made, it is still an active area of research.

In classification problems, the discriminative approach circumvents this kind of difficulties by directly inferring the boundaries between classes in the feature space.^{2,3} Discriminative approaches have been proved to be successful in dealing with small class distances, high dimensionality, and limited training sets. Discriminative classifiers hold the state-of-the-art performance in supervised hyperspectral image classification (see, *e.g.*⁴).

The support vector machines (SVMs)⁵ and multinomial logistic regression (MLR)⁶ are among the most consolidated discriminative supervised classification tools in hyperspectral classification. Due to their ability to deal with large input spaces efficiently and to produce sparse solutions, SVMs have been successfully used for hyperspectral supervised classification.^{4,7,8} The multinomial logistic regression has the advantage of learning the class distributions themselves. Effective sparse multinomial logistic regression (SMLR) methods are available.⁹ The fast sparse multinomial logistic regression algorithm¹⁰ (FSMLR) has been introduced in hyperspectral image classification to extend the capabilities of the SMLR algorithm in term of sample size and of number of classes. More recently, the introduction of the logistic regression via splitting and augmented Lagrangian algorithm^{10,11} (LORSAL) has open the door to deal with even larger data sets and with a higher number of classes. These ideas have been applied to hyperspectral image classification problems.^{10,11}

Jun Li and José Bioucas-Dias: Address: Instituto Superior Técnico, Instituto de Telecomunicações, Av. Rovisco Pais, Torre Norte, Piso 10, 1049-001, Lisbon, Portugal, E-mail: {jun, bioucas}@lx.it.pt, Telephone: +351.21.8418466, Fax: +351.21.8418472

Antonio Plaza: Address: Computer Architecture and Technology Area, Department of Technology of Computers and Communications, Escuela Politecnica de Cáceres, University of Extremadura, Avda. de la Universidad S/N, E-10071 Cáceres (SPAIN) Telephone: +34.927.257000, E-mail: aplaza@unex.es

Recently, a trend to improve the classification accuracy in hyperspectral classification is to integrate spatial and spectral information.^{7,8,10} Markov Random Field (MRF) models allow contextual constraints to be incorporated and have been used extensively for various segmentation applications, including hyperspectral data classification.^{8,10,11} These methods exploit, in a way or another, the continuity, in a probabilistic sense, of neighboring labels: it is very likely that, in a hyperspectral image, two neighboring pixels have the same label. More recently, graph-based methods have been applied to hyperspectral classification with good results.^{12,13}

In this paper, we present a new semi-supervised approach to hyperspectral images. It mainly implements two steps. First, the learning step computes the class densities by using the LORSAL algorithm; in order to improve the classification performance of this step, we include the contextual information by adopting a multi-level logistic (MLL) Markov-Gibbs prior. The maximum a posteriori (MAP) segmentation is computed via a min-cut based integer optimization algorithm. Second, labels are actively selected. The learning step signs the importance of the labeled and unlabeled samples, which is used to evaluate the unlabeled samples for active query selection^{14,15}

The paper is organized as follows. Section 2 formulates the problem. Section 3 briefly reviews the multinomial logistical regression, the MLL prior and the active selection approach. Section 4 reports classification results a real hyperspectral data set in comparison with state-of-the-art competitors. Finally, Section 5 includes some concluding remarks and future directions.

2. PROBLEM FORMULATION

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

$\mathcal{S} \equiv \{1, \dots, n\}$	Set of integers indexing the n pixels of an image
$\mathcal{L} \equiv \{1, \dots, K\}$	Set of K labels
$\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{R}^{d \times n}$	Image in which the pixels are d -dimensional feature vectors
$\mathbf{X} \in \mathbb{R}^{d \times L}$	Training set
$\mathbf{y} = (y_1, \dots, y_n) \in \mathcal{L}^n$	Image of labels
$\mathbf{y}_i = [y_i^{(1)}, \dots, y_i^{(K)}]^T$	"1-of- K " encoding of the K classes

Note that the variables \mathbf{y}_i and y_i have different structure but are equivalent, *e.g.*,

$$(\mathbf{y}_i = [0, 1, 0, 0]^T) \Leftrightarrow (y_i = 2).$$

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

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

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

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

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) estimate is then given by

$$\hat{\mathbf{y}} = \arg \max_{\mathbf{y} \in \mathcal{L}^n} \left\{ \left(\sum_{i=1}^n p(y_i|\mathbf{x}_i) \right) + \log P(\mathbf{y}) \right\}. \tag{2}$$

In the present approach, the densities $p(y_i|\mathbf{x}_i)$ are learned by using the LORSAL algorithm. The prior $P(\mathbf{y})$ follows a MLL MRF prior. The next section summarizes the main aspects of the proposed approach.

3. PROPOSED APPROACH

3.1 Learning the MLR regressors with LORSAL algorithm

The multinomial logistic regression is formally given by⁶

$$p(y_i^{(k)} = 1|\mathbf{x}_i, \boldsymbol{\omega}) = \frac{\exp(\boldsymbol{\omega}^{(k)}\mathbf{h}(\mathbf{x}_i))}{\sum_{k=1}^K \exp(\boldsymbol{\omega}^{(k)}\mathbf{h}(\mathbf{x}_i))}, \quad (3)$$

where $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \dots, h_l(\mathbf{x})]^T$ is a vector of l fixed functions of the input, often termed features, $\boldsymbol{\omega} = [\boldsymbol{\omega}^{(1)T}, \dots, \boldsymbol{\omega}^{(K-1)T}]^T$, and, because the density function (3) does not depend on translations on the regressors $\boldsymbol{\omega}^{(k)}$, we take $\boldsymbol{\omega}^{(K)} = \mathbf{0}$.

Usual choices for the function $\mathbf{h}(\mathbf{x})$ are 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) and 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 $K_{\mathbf{x}, \mathbf{x}_j} = K(\mathbf{x}_i, \mathbf{x}_j)$ and $K(\cdot, \cdot)$ is some symmetric kernel function, is a relevant example of the nonlinear case. Kernels have been largely used because they tend to improve the data separability in the transformed space. In this paper, we use the Gaussian Radial Basis Function (RBF) kernel

$$K(\mathbf{x}, \mathbf{z}) = -\exp(-\|\mathbf{x} - \mathbf{z}\|^2)/2\sigma^2,$$

which is widely used in hyperspectral image classification.⁴ From now on, d denotes the dimension of $\mathbf{h}(\mathbf{x})$.

The MAP estimate of $\boldsymbol{\omega}$ is

$$\hat{\boldsymbol{\omega}} = \arg \max_{\boldsymbol{\omega}} [l(\boldsymbol{\omega}) + \log p(\boldsymbol{\omega})], \quad (4)$$

where $l(\boldsymbol{\omega})$ is the log-likelihood function and $p(\boldsymbol{\omega})$ is the sparsity promoting prior aimed at controlling the classifier complexity and thus enhancing its generalization capacity. As in,^{10,11} we adopt the Laplacian prior

$$p(\boldsymbol{\omega}) \propto \exp(-\lambda\|\boldsymbol{\omega}\|_1),$$

where λ is a regularization parameter controlling the degree of sparseness of $\hat{\boldsymbol{\omega}}$. Problem (4), although convex, is hard. It is difficult to solve because the term $l(\boldsymbol{\omega})$ is non-quadratic and the term $\log p(\boldsymbol{\omega})$ is non-smooth. The SMLR algorithm presented in⁹ solves this problem with $O((dK)^3)$ complexity, which is frequently unbearable, namely when dealing with kernels and large datasets. The LORSAL algorithm introduced in¹⁶ is able to handle these difficulties. It solves (4) with a much lighter computational complexity. The LORSAL algorithm 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 with a complexity cost of $O(d^2K)$. Compared with the SMLR algorithm, the reduction of computational complexity is $O(dK^2)$. More details about this algorithm can be seen in.^{16,17}

3.2 The Multi-Level Logistic spatial prior

In order to include the spatial contextual information, we adopt the Markov random field (MRF) isotropic Multi-Level Logistic (MLL) prior, which models the piecewise continuous nature of the image elements, considering that the neighboring pixels are likely to belong to the same class. According to the Hammersly-Clifford theorem, the density associated with a MRF is a Gibbs's distribution,¹⁸ which can be written as

$$P(\mathbf{y}) = \frac{1}{Z} e^{\mu \sum_{i \sim j} \delta(y_i - y_j)}, \quad (5)$$

where Z is a normalizing constant, $i \sim j$ denotes first order neighboring sites, $\delta(y)$ is the unit impulse function*, 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.

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

3.3 Energy minimization via graph cuts

Using the LORSAL algorithm to learn $p(y_i|\mathbf{x}_i)$ and the MLL prior $P(\mathbf{y})$, and according to (2), the MAP segmentation is finally given by

$$\hat{\mathbf{y}} = \arg \min_{\mathbf{y} \in \mathcal{L}^n} \left\{ \sum_{i \in \mathcal{S}} -\log p(y_i|x_i) - \mu \sum_{i \sim j} \delta(y_i - y_j) \right\}. \quad (6)$$

The minimization (6) is a hard combinatorial optimization problem. However, given that the pairwise interaction term on the right hand side of (2) is a metric, it is possible to achieve a very good approximation using the α -Expansion graph cut based algorithm.¹⁹⁻²²

3.4 Active label selection

In order to improve the segmentation performance, active query selection is performed in this paper. The basic idea of active learning is that of iteratively enlarging the training set by requesting an expert to, in each iteration, label samples from the unlabeled set. The relevant question is, of course, what set should be chosen. In this paper, we use a mutual information (MI) based criterion^{14,15} that maximizes the MI between the MLR regressors and the labels. Furthermore, we implement active label selection just based on the class densities $p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\omega})$ and not on the spatial prior $P(\mathbf{y})$. Although the resulting scheme is suboptimal, it yields excellent results.

The approach uses a Laplace approximation of the posterior $p(\boldsymbol{\omega}|\mathbf{X}) \simeq \mathcal{N}(\boldsymbol{\omega}|\hat{\boldsymbol{\omega}}, \mathbf{H}^{-1})$, where \mathbf{H} is the posterior precision matrix, *i.e.*, the Hessian of minus the log-posterior $\mathbf{H} = \nabla^2(-\log p(\boldsymbol{\omega}|\mathbf{X}))$. Let \mathbf{x}_* be an unlabeled sample and \mathbf{y}_* be its label. Assume that the MAP estimate $\hat{\boldsymbol{\omega}}$ remains unchanged after including \mathbf{y}_* . This assumption is clearly not true at the beginning of the active learning procedure. However, it was empirically proved that it is a very good approximation.¹⁵ As shown in,¹⁴ the MI $I(\boldsymbol{\omega}; \mathbf{y}_*)$ between the MLR regressors and the label \mathbf{y}_* is given by

$$\begin{aligned} I(\boldsymbol{\omega}; \mathbf{y}_*) &= (1/2) \log\{|\mathbf{H}'|/|\mathbf{H}|\} \\ &= (1/2) \log \left(1 + \prod_{i=1}^K p_*(y_i|\mathbf{X}, \hat{\boldsymbol{\omega}}) \mathbf{x}_*^T \mathbf{H}^{-1} \mathbf{x}_* \right). \end{aligned}$$

The MI is maximized when $p_i \approx 1/K$, *i.e.*, for samples near the classifier boundaries, corresponding to the higher entropy of the class distribution of \mathbf{y}_* . More details can be found in.¹⁵

The pseudo-code for the proposed algorithm is presented below

Algorithm 1 Semi-supervised segmentation algorithm using active label selection

while The stop criterion is not fulfilled **do**
 1. Learn the MLRs $\boldsymbol{\omega}$ parameterizing $p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\omega})$ by using LORSAL algorithm according to (3).
 2. Use the MLL prior $P(\mathbf{y})$ according to (5).
 3. Estimate the MAP solution using α -Expansion graph cut based algorithm.
 4. Label \mathbf{x}_* using the active query selection method.
end while

4. EXPERIMENTAL RESULTS

In this section, experimental results on a real hyperspectral image will be presented. 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 $\mu = 1.5$. All though these are not optimal choices, they lead to very good results. The initial training set were randomly selected from the ground truth image. Active query selection approach was used to label new samples. Each value of overall accuracy (OA) was obtained from 10 Monte Carlo runs.

Table 1. OA [%] results over both images. In total, 20% of the ground truth labels were used as the training set. Best results (Bold) are highlighted for each problem.

Classifier	Subset	Whole
Euclidean ²⁴	67.43	48.23
BLOOC+DAFE+ECHO ²⁴	93.50	82.91
Composite Kernel ²³	98.86	96.53
Composite Kernel using Wavelet smoothing ¹³	98.96	97.85
Composite Kernel using PDE smoothing ¹³	98.83	93.62
LORSAL	96.18	84.51
Supervised algorithm: LOSAL + MLL ¹¹	98.70	94.36
Semi-supervised LORSAL	99.69	92.99
Proposed semi-supervised algorithm	99.97	98.77

The well-known AVIRIS Indian Pines scene was used to evaluate the proposed algorithm. This image was collected over Northwestern Indiana in June of 1992.¹ This scene is available online[†], containing 145×145 pixels and 220 spectral bands in the range of 400-2500nm. Two scenarios were considered in our experiments. In the first experiment, the whole image of 145×145 pixels and 224 spectral bands was considered, as in¹³ and.²³ 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 semi-supervised algorithm over both images in comparison with equivalent results from other researchers.^{11, 13, 23, 24} In these experiments, we used 50 samples per class as the initial training set. The total size of the labeled samples used was 20% of the ground truth image. The remaining samples were used as the validation set. The proposed semi-supervised algorithm obtained better results than other classifiers in both scenarios. With active query selection, both the semi-supervised LORSAL (LORSAL algorithm with active query selection) and the proposed semi-supervised algorithm outperformed the LORSAL algorithm and the supervised algorithm in,¹¹ which is equivalent to LORSAL plus a MLL spatial prior, respectively. The classification maps shown in figure 1 and figure 2. Effective results can be seen from these maps.

Table 2 presents the classification results as functions of the number of labeled samples over the subset image respectively. The results are compared with state-of-the-art classifiers.^{11-13, 16} The proposed algorithm yields very good results, which outperformed the compared classifiers in all cases. In order to show the good performance of the proposed semi-supervised algorithm, we run the algorithm with 50 labeled samples per class, an OA of 99.17% was obtained. The result is better than those of the other classifiers used 20% of ground truth as the training sets, which is around 220 labeled samples per class.

5. CONCLUSIONS

This work has presented a new semi-supervised approach for hyperspectral classification problems, which combines the spectral and spatial information by using the MRF graph-based strategy. Active query selection is considered. The results obtained showed very good performance over real AVIRIS Indiana Pines dataset, both in the whole scene and in the subset, which outperformed state-of-the-art algorithms.^{12, 13, 23}

ACKNOWLEDGMENTS

This work was supported by Marie Curie Research Training Networks Program Grants MEST-CT-2005-021175 [European Doctoral Program in Signal Processing (SIGNAL)] and MRTN-CT-2006-035927 [Hyperspectral Imag-

[†]<http://cobweb.ecn.purdue.edu/biehl/MultiSpec/>

Table 2. OA [%] results as a function of the number of labeled samples in the subset Indiana Pines image. The number of initial training set used for the proposed semi-supervised algorithm was 5 samples per class. Best results (Bold) are highlighted for each problem.

Algorithms	number of labeled samples per class							
	3	5	10	15	20	25	30	100
Proposed algorithm	-	-	94.69	96.47	96.95	97.84	98.09	99.70
Semi-supervised LORSAL	-	-	85.64	88.68	91.56	92.62	94.10	98.57
Supervised algorithm ¹¹	82.80	87.51	92.83	95.52	96.59	97.64	97.34	97.82
LORSAL	74.01	77.51	85.42	88.38	89.78	90.32	91.90	94.67
Wavelet ¹³	73.65	78.78	82.90	85.74	86.85	87.69	88.68	92.59
PDE ¹³	84.89	86.89	90.03	90.51	91.33	92.67	93.74	94.20
Semi-supervised algorithm ¹²	66.73	67.13	71.32	79.49	82.04	83.12	84.99	86.44



(a) Ground truth



(b) Classification map

Figure 1. Classification maps obtained by the proposed algorithm over the whole scenario along with the ground truth image.



(a) Ground truth



(b) Classification map

Figure 2. Classification maps obtained by the proposed algorithm over the subset image along with the ground truth image.

ing Network (HYPER-I-NET)] from the European Commission.

REFERENCES

- [1] Landgrebe, D. A., [*Signal Theory Methods in Multispectral Remote Sensing*], John Wiley, Hoboken, NJ (2003).
- [2] Ng, A. Y. and Jordan, M. I., “On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes,” in [*Proc. 16th Annual Conference on Neural Information Processing Systems*], (2002).
- [3] Vapnik, V., [*Statistical Learning Theory*], John Wiley, New York (1998).
- [4] Camps-Valls, G. and Bruzzone, L., “Kernel-based methods for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing* **43**, 1351–1362 (2005).
- [5] Scholkopf, B. and Smola, A., [*Learning With Kernels Support Vector Machines, Regularization, Optimization and Beyond*], MIT Press Series, Cambridge, MA (2002).
- [6] Böhning, D., “Multinomial logistic regression algorithm,” *Annals of the Institute of Statistical Mathematics* **44**, 197–200 (1992).
- [7] Fauvel, M., Benediktsson, J., Chanussot, J., and Sveinsson, J., “Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles,” *IEEE Transactions on Geoscience and Remote Sensing* **46**(11), 3804–3814 (2008).
- [8] Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C., and Trianni, G., “Recent advances in techniques for hyperspectral image processing,” *Remote Sensing of Environment* **113**, 110–122 (September 2009).
- [9] Krishnapuram, B., Carin, L., Figueiredo, M., and Hartemink, A., “Sparse multinomial logistic regression: Fast algorithms and generalization bounds,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(6), 957–968 (2005).
- [10] Borges, J., Bioucas-Dias, J., and Marçal, A., “Evaluation of Bayesian hyperspectral imaging segmentation with a discriminative class learning,” in [*Proc. IEEE International Geoscience and Remote sensing Symposium*], (2007).
- [11] Li, J., Bioucas-Dias, J., and Plaza, A., “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).
- [12] Camps-Valls, G., Bandos, T., and Zhou, D., “Semi-supervised graph-based hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing* **45**, 3044–3054 (Oct 2007).
- [13] Velasco-Forero, S. and Manian, V., “Improving hyperspectral image classification using spatial preprocessing,” *IEEE Geoscience and Remote Sensing Letters* **6**, 297–301 (2009).
- [14] Mackay, D., “Information-based objective functions for active data selection,” *Neural Computation* **4**, 590–604 (1992).
- [15] Krishnapuram, B., Williams, D., Xue, Y., Hartemink, A., Carin, L., and Figueiredo, M., “On semi-supervised classification,” in [*Proc. 18th Annual Conference on Neural Information Processing Systems*], (2004).
- [16] Bioucas-Dias, J. and Figueiredo, M., “Logistic regression via variable splitting and augmented lagrangian tools,” tech. rep., Instituto Superior Técnico, TULisbon (2009).
- [17] Li, J., Bioucas-Dias, J., and Plaza, A., “Hyperspectral image classification based on a fast bregman sparse multinomial logistic regression algorithm,” in [*6th EARSeL SIG IS Workshop*], (2009).
- [18] Geman, S. and Geman, D., “Stochastic relaxation, gibbs distribution, and the bayesian restoration of images,” *IEEE TPAMI* **6**, 721–741 (1984).
- [19] Boykov, Y., Veksler, O., and Zabih, R., “Efficient approximate energy minimization via graph cuts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, 1222–1239 (November 2001).
- [20] Boykov, Y. and Kolmogorov, V., “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision,” *IEEE transactions on Pattern Analysis and Machine Intelligence* **26**, 1124–1137 (September 2004).

- [21] Kolmogorov, V. and Zabih, R., “What energy functions can be minimized via graph cuts?,” *IEEE transactions on Pattern Analysis and Machine Intelligence* **26**, 147–159 (February 2004).
- [22] Bagon, S., “Matlab wrapper for graph cut,” (December 2006).
- [23] Camps-Valls, G., Gomez-Chova, L., Muoz-Mar, J., Vila-Francis, J., and Calpe-Maravilla, J., “Composite kernels for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters* (Jan 2006).
- [24] Mrazek, P. and Weickert, J., *Classification of high dimensional data with limited training samples*, PhD thesis, Purdue University (1998).