

A New Semi-Supervised Classification Strategy Combining Active Learning and Spectral Unmixing of Hyperspectral Data

Yanli Sun^{*a,b}, Xia Zhang^a, Antonio Plaza^b, Jun Li^c, Inmaculada Dópido^b, and Yi Liu^b
^a State Key Laboratory of Remote Sensing Science, Institute of Remote sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China; ^b Hyperspectral Computing Laboratory (Hypercomp), Department of Technology of Computers and Communications, University of Extremadura, Cáceres, E-10003, Spain; ^c the School of Geography and Planning and Guangdong Key Laboratory for Urbanization and Geo-Simulation, Sun Yat-sen University, Guangzhou, China

ABSTRACT

Hyperspectral remote sensing allows for the detailed analysis of the surface of the Earth by providing high-dimensional images with hundreds of spectral bands. Hyperspectral image classification plays a significant role in hyperspectral image analysis and has been a very active research area in the last few years. In the context of hyperspectral image classification, supervised techniques (which have achieved wide acceptance) must address a difficult task due to the unbalance between the high dimensionality of the data and the limited availability of labeled training samples in real analysis scenarios. While the collection of labeled samples is generally difficult, expensive, and time-consuming, unlabeled samples can be generated in a much easier way. Semi-supervised learning offers an effective solution that can take advantage of both unlabeled and a small amount of labeled samples. Spectral unmixing is another widely used technique in hyperspectral image analysis, developed to retrieve pure spectral components and determine their abundance fractions in mixed pixels. In this work, we propose a method to perform semi-supervised hyperspectral image classification by combining the information retrieved with spectral unmixing and classification. Two kinds of samples that are highly mixed in nature are automatically selected, aiming at finding the most informative unlabeled samples. One kind is given by the samples minimizing the distance between the first two most probable classes by calculating the difference between the two highest abundances. Another kind is given by the samples minimizing the distance between the most probable class and the least probable class, obtained by calculating the difference between the highest and lowest abundances. The effectiveness of the proposed method is evaluated using a real hyperspectral data set collected by the airborne visible infrared imaging spectrometer (AVIRIS) over the Indian Pines region in Northwestern Indiana. In the paper, techniques for efficient implementation of the considered technique in high performance computing architectures are also discussed.

Keywords: Hyperspectral remote sensing, image classification, semi-supervised learning, spectral unmixing

1. INTRODUCTION

Hyperspectral imaging has been a widely used technique in remote sensing that collects hundreds of images, at different wavelength channels, for the same area on the surface of the Earth [1]. The special characteristics of hyperspectral data sets bring difficult processing problems [2] [3]. Classification [4] and spectral unmixing [5] are two important techniques for hyperspectral data exploitation.

Hyperspectral image classification has been a very active area of research in the last few years [6], [7]. Given a set of observations (i.e., possibly mixed pixel vectors), the goal of classification is to assign a unique label to each pixel vector so that it is well defined by a given class [8]. Supervised classification techniques, such as the support vector machine (SVM) [9] [10] or multinomial logistic regression (MLR) [11], can deal effectively with the Hughes phenomenon [12]. However, supervised classification is generally a difficult task due to the unbalance between the high dimensionality of the data and the limited availability of labeled training samples in real analysis scenarios [13]. While the collection of labeled samples is generally difficult, expensive, and time-consuming, unlabeled samples can be generated in a much easier way [14]. This observation has fostered the idea of adopting semisupervised learning techniques in hyperspectral image classification.

*sunyanli1601@gmail.com; phone +34 615 096 317

The area of semisupervised learning has experienced a significant evolution in terms of the adopted models, which comprise complex generative models [15]–[18], self-learning models [19], [20], multiview learning models [21], [22], transductive support vector machines (TSVMs) [23], [24], and graph-based methods [25]. A survey of semisupervised learning algorithms is available in [26]. These algorithms generally assume that a limited number of labeled samples are available a priori, and then enlarge the training set using unlabeled samples, thus allowing these approaches to address ill-posed problems [27]. However, in order for this strategy to work, several requirements need to be met. First and foremost, the new (unlabeled) samples should be generated without significant cost/effort. Second, the number of unlabeled samples required in order for the semisupervised classifier to perform properly should not be too high in order to avoid increasing computational complexity in the classification stage. In other words, as the number of unlabeled samples increases, it may be unbearable for the classifier to properly exploit all the available training samples due to computational issues. Further, if the unlabeled samples are not properly selected, these may confuse the classifier, thus introducing significant divergence or even reducing the classification accuracy obtained with the initial set of labeled samples [28]. In order to address these issues, it is very important that the most highly informative unlabeled samples are identified in computationally efficient fashion, so that significant improvements in classification performance can be observed without the need to use a very high number of unlabeled samples.

An important problem for hyperspectral image classification is the presence of mixed pixels [29]. If a pixel is highly mixed, it is very difficult to determine its class label as it is not easily separable from other classes. A common way for dealing with mixed pixels has been spectral unmixing [5], which aims at estimating the abundance fractions of a set of pure spectral signatures (endmembers) that can be then used to determine endmember proportions within the pixel [30]. In this literature[31], A predefined labeled and unlabeled samples design protocol was proposed in the context of active learning architecture to improve hyperspectral image classification with smaller in size but consist of more valuable of information. In this literature, the values of mixed and pure pixels for AL heuristics were investigated comparatively. The experimental results showed that mixed pixels are always capable of providing more value information than pure pixels, either by initial labeled samples or unlabeled samples.

In order to take advantage of the complementary properties of spectral unmixing and classification, several approaches have been studied in the recent literature for the integration of these two techniques. For instance, in [32], spectral unmixing was used as a feature extraction strategy prior to classification. It was found that spectral unmixing can be used to derive suitable features for classification purposes. This idea was further explored in [33], in which several spectral unmixing chains (unsupervised and supervised) were used to derive suitable features for classification purposes. More recently, the synergistic nature of spectral unmixing and classification has been explored in the context of a semisupervised framework [34]. In this approach, a weight parameter was investigated to combine concepts of spectral unmixing and classification, which bridge the gap between the two most important techniques for hyperspectral interpretation.

Another important issue of semisupervised classification is how to active learn the most highly informative unlabeled samples. Samples with high uncertainty are generally preferred as they are usually more informative. For instance, the mutual information (MI-based) criterion focuses on the most complex regions (i.e., regions with the largest number of boundaries). Breaking ties (BT) criterion focuses on the boundary region between two classes, with the goal of obtaining more diversity in the composition of the training set. And the modified BT (MBT) intends to guarantee unbiased samplings among the classes [35]. In literature [34], BT algorithm was used to select the most informative samples by calculating the difference between the two highest probabilities, which finds the samples minimizing the distance between the first two most probable classes.

In this work, we further improved the hybrid strategy combining semisupervised classification and unmixing proposed in [34] by considering both the boundary region between two classes and the regions with the largest number of boundaries instead of only focusing on the former for the active learning process. Since the classification probabilities and the abundance fractions are combined to select the most informative unlabeled samples, we have two implications by this active learning. From the classification probabilities viewpoint, it means to consider both the boundary region between two classes and the regions with the largest number of boundaries. From the spectral unmixing viewpoint, it means to find the samples contain two main endmembers, which are easily misclassified, and the samples stay in the most mixed area. Both of them are supposed to be the most informative samples.

2. METHOD

The proposed approach can be summarized by the flowchart given in Fig.1. As shown in the figure, the proposed approach consists of three main steps: 1) semisupervised learning; 2) spectral unmixing; 3) active learning.

First, set up a potential unlabeled samples candidate set by a semisupervised classification using spatial information. For the semisupervised part of our approach, we use the MLR probabilistic classifier to model the class posterior density. The considered semisupervised approach increment the unlabeled training set from 4-connected neighboring samples combined the maximum a posteriori (MAP) estimate from the MLR classifier. This increment is reasonable due to the following considerations: first, from a global viewpoint, samples which have the same spectral structure likely belong to the same class and second, from a local viewpoint, it is very likely that two neighboring pixels also belong to the same class. Therefore, the newly included samples are reliable for learning the classifier. In this work, we run an iterative scheme to increment the training set, as this strategy can refine the estimates and enlarge the neighborhood set such that the set of potential unlabeled training samples is increased.

In the spectral unmixing step, MTMF-based unmixing is conducted using the resulting endmember after k-means clustering. To solve the problems highlighted by endmember extraction algorithms which are sensitive to outliers and pixels with extreme values of reflectance, an unsupervised clustering method k-means is used on the available labeled samples. The endmembers extracted (from class centers) are expected to be more spatially significant. Study suggested that partial unmixing using mixture-tuned matched filtering (MTMF) [36] could be an effective solution to deal with the likely fact that not all pure spectral constituents in the scene (needed for spectral unmixing purposes) are known a priori, but a more exhaustive investigation of partial unmixing (in combination with spatial information) is desirable. In [33], it is shown that MTMF can outperform other techniques for abundance estimation such as FCLSU, since it can provide meaningful abundance maps by means of partial unmixing in case not all endmembers are available a priori. In this work, we consider the MTMF unmixing strategy addressing the aforementioned issues as discussed in [33].

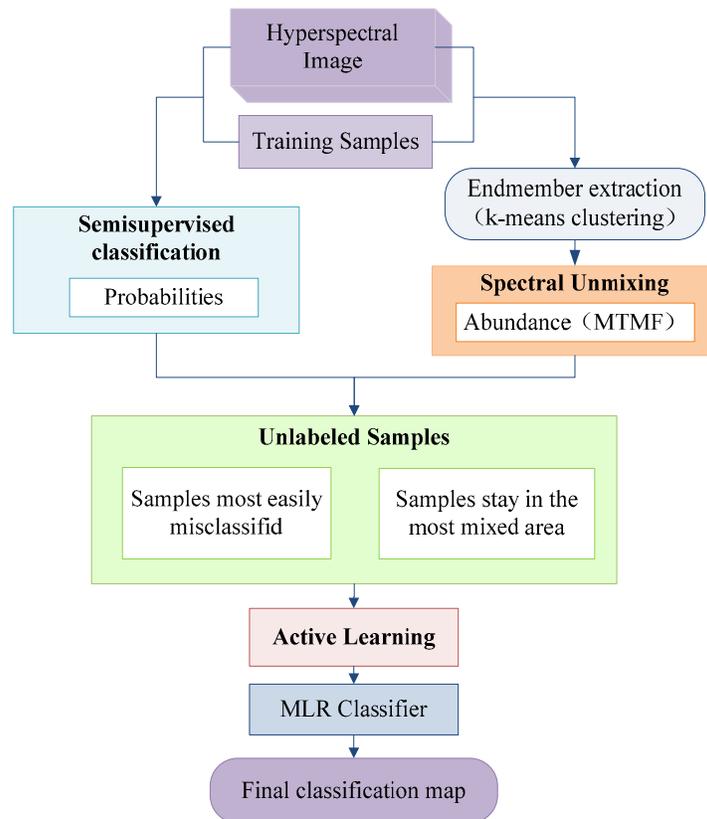


Figure 1. General flowchart described the proposed semisupervised classification method.

In the active learning step, the classification probabilities and spectral unmixing abundance are combined to learn the most informative unlabeled samples. Literature [34] has investigated the balance between the classification probabilities and the abundance fractions. A weigh parameter of 0.75 is suggested to be a reasonable compromise for the combination (where 1 means only classification probabilities are considered, 0 means only spectral unmixing is considered), which indicates that classification generally needs more weight than unmixing in order to obtain the best analysis results. Therefore, in this work, we used the parameter 0.75 for the combination of classification and spectral unmixing, such that the new unlabeled samples selected are more informative in comparison with those samples selected only from classification or unmixing methods. On the other hand, in order to select the most informative unlabeled samples, we considered two kinds of criteria in the active learning. From the classification probabilities viewpoint, it means to consider both the boundary region between two classes and the regions with the largest number of boundaries. From the spectral unmixing viewpoint, it means to find the samples contain two main endmembers, which are easily misclassified, and the samples stay in the most mixed area.

3. EXPERIMENTAL RESULTS

In this section, we evaluate the new methodology presented in this paper using the AVIRIS Indian Pines hyperspectral image. In our experiments, we apply the MLR classifier (with Gaussian RBF kernel) to a normalized version of the considered hyperspectral data sets. In all cases, the reported figures of overall accuracy (OA), average accuracy (AA), and kappa statistic (κ) are obtained by averaging the results obtained after conducting 10 independent Monte Carlo runs with respect to the labeled training set from the ground-truth data, where the remaining samples are used for validation purposes. In order to illustrate the good performance of the proposed approach, we use very small labeled training sets on purpose (5, 10, and 15 labeled samples per class).

3.1 Hyperspectral Data Set

The hyperspectral data set used in our experiments was collected by the AVIRIS sensor over the Indian Pines region in Northwestern Indiana in 1992. This scene, with a size of 145 lines \times 145 samples, was acquired over a mixed agricultural/forest area, early in the growing season. The scene comprises 220 spectral channels in the wavelength range from 0.4 to 1.5 μ m, nominal spectral resolution of 10 nm, moderate spatial resolution of 20 m by pixel, and 16-b radiometric resolution. After an initial screening, bands 104 to 108, 150 to 163 and band 220 were removed from the data set due to noise and water absorption phenomena, leaving a total of 200 radiance channels to be used in the experiments. For illustrative purposes, Fig. 2(a) shows a false color composition of the AVIRIS Indian Pines scene, whereas Fig. 2(b) shows the ground-truth map available for the scene, displayed in the form of a class assignment for each labeled pixel, with 16 mutually exclusive ground-truth classes.

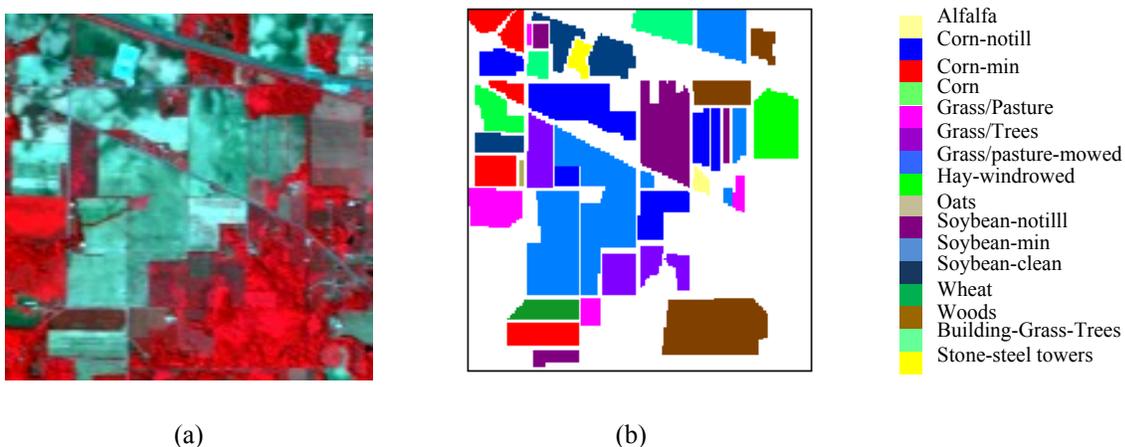


Figure 2. (a) False color composition of the AVIRIS Indian Pines scene. (b) Ground-truth map containing 16 mutually exclusive land-cover classes.

3.2 Experimental results with the hyperspectral data set

In this experiment, we consider different numbers of labeled samples per class: $L_n = \{5, 10, 15\}$. Table I shows the OA, AA, and the kappa statistic (κ) obtained by different classification technique. To verify the proposed semisupervised approach, the supervised strategy (based on the MLR classifier), the random sampling (RS) unlabeled sample selection approach, the BT semisupervised approach, and the semisupervised strategy proposed in literature [34] were performed for comparison. In all the semisupervised approach, 300 unlabeled samples were selected. As a result, in Table I, BT denotes the semisupervised classifier with BT for unlabeled sample selection. RS denotes the semisupervised classifier with unlabeled samples selected using RS. Strategies 4 denote the semisupervised hybrid classifier integrating classification and spectral unmixing proposed in literature [34] (with a weigh parameter of 0.75), where unlabeled samples are selected using BT.

As shown in Table I, the inclusion of unlabeled samples significantly improved the classification results in all cases, which indicate that semisupervised techniques can significantly help to improve the accuracy results. The results of Strategy 4 proposed in [38] indicated the effectiveness of combining classification and spectral unmixing. Our proposed approach further improved the classification accuracy on the basis of Strategy 4.

Table I. OA, AA(%), and Kappa Statistic (κ) Obtained Using Different Classification Technique When Applying to the AVIRIS Indian Pines Hyperspectral Data Set.

Ln=5 labeled samples per class					
	Supervised	BT	RS	Strategy 4 [34]	Proposed approach
OA	51.78%	65.25%	60.03%	70.45%	71.62%
AA	63.82%	69.98%	66.31%	72.32%	72.47%
κ	0.4626	0.6065	0.5481	0.6625	0.6754
Ln=10 labeled samples per class					
	Supervised	BT	RS	Strategy 4 [34]	Proposed approach
OA	60.12%	68.51%	65.59%	75.29%	78.09%
AA	71.74%	75.66%	73.49%	79.05%	80.11%
κ	0.5543	0.6440	0.6117	0.7184	0.7495
Ln=15 labeled samples per class					
	Supervised	BT	RS	Strategy 4 [34]	Proposed approach
OA	66.20%	70.46%	70.29%	76.68%	78.92%
AA	77.39%	79.47%	78.76%	81.94%	82.79%
κ	0.6209	0.6665	0.6641	0.7353	0.7591

(Some results are directly taken from [34])

For illustrative purposes, Fig.3 shows some of the classification maps obtained for the AVIRIS Indian Pines scene. These classification maps correspond to one of the 10 Monte-Carlo runs that were averaged in order to generate the classification scores reported in Table I. On the one hand, the advantages obtained by adopting a semisupervised learning approach, which combines classification and unmixing concepts, can be clearly appreciated in the classification maps displayed in Fig. 3. On the other hand, the modified active learning method by considering two kinds of the most informative samples was proved to effectively improve the classification accuracy.

4. CONCLUSIONS AND FUTURE RESEARCH LINES

In this paper, we developed a semisupervised method by combining concepts of spectral unmixing and classification and learning the two kinds of most informative unlabeled samples. The combination of spectral and classification has been done by injecting spectral unmixing information in the semisupervised classification process with a relative weight. The active learning considers two kinds of criterion to achieve the most informative unlabeled samples. Experiments with the AVIRIS Indian Pines data have been performed in order to investigate effectiveness of the proposed method. Results

indicate that the proposed method can effectively improve the classification accuracy. In our future work, we will explore the balance of the two kinds of criterion in the active learning. We are also planning on developing high-performance computing implementations that can take advantage of the inherent parallel nature of some of the steps of the presented methodology.

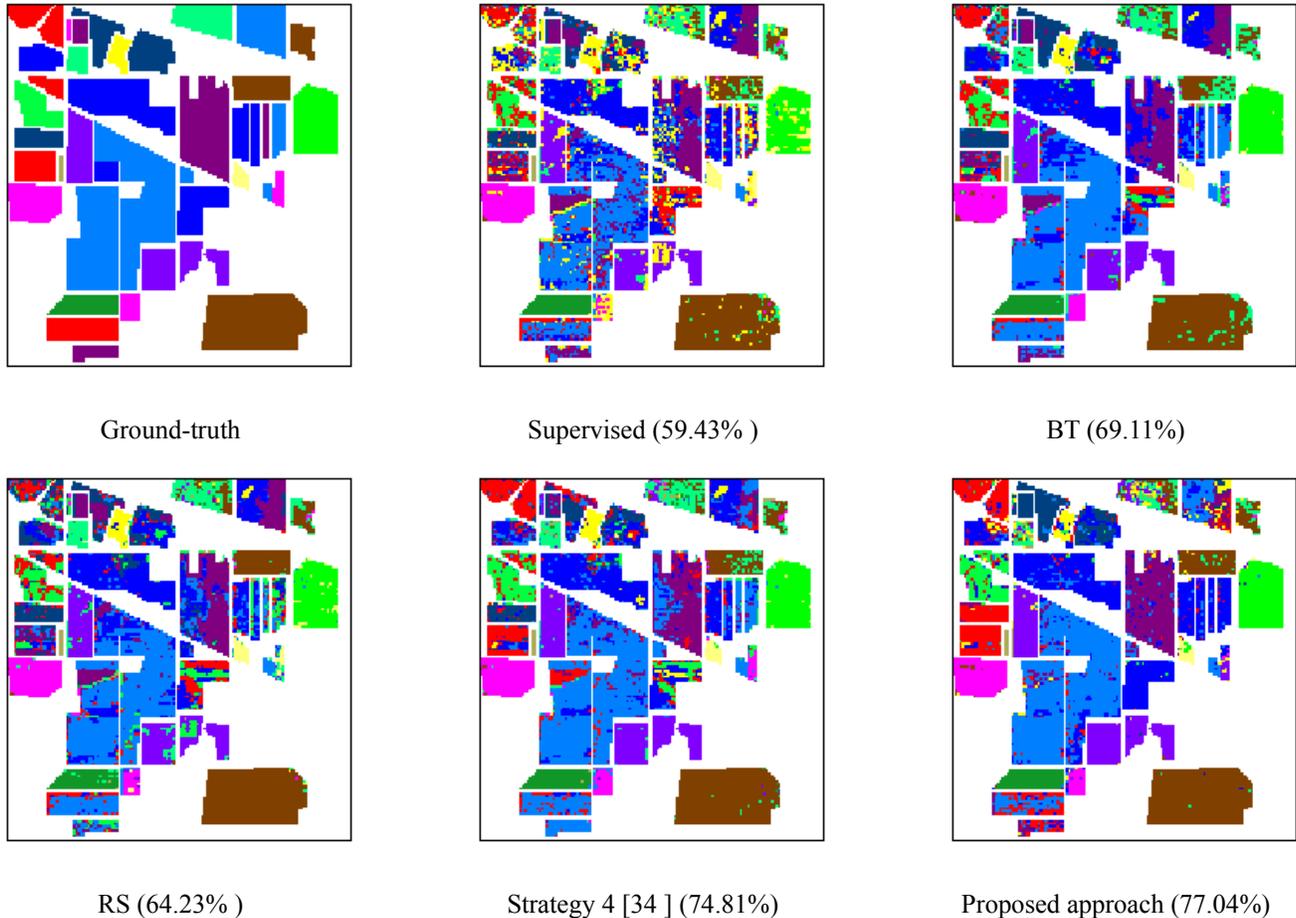


Figure 3. Classification maps and OAs obtained after applying classification techniques to the AVIRIS Indian Pine data set. In all cases, the number of labeled samples was $L_n = 10$ and the number of unlabeled samples 300.

REFERENCES

- [1] J. Bioucas-Dias et al., "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [2] A. Plaza, J. M. Bioucas-Dias, A. Simic, and W. J. Blackwell, "Foreword to the special issue on hyperspectral image and signal processing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 347–353, Apr. 2012.
- [3] Q. Du et al., "Foreword to the special issue on hyperspectral remote sensing: Theory, methods, and applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 459–465, Apr. 2013.
- [4] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral–spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.
- [5] J. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

- [6] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New York, NY, USA: John Wiley & Sons, 2003.
- [7] J. Chanussot, M. M. Crawford, and B.-C. Kuo, "Foreword to the special issue on hyperspectral image and signal processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 3871–3876, Nov. 2010.
- [8] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York, NY, USA: Springer-Verlag, 2006.
- [9] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [10] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
- [11] J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
- [12] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.
- [13] B. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1087–1095, Sep. 1994.
- [14] F. Bovolo, L. Bruzzone, and L. Carlin, "A novel technique for subpixel image classification based on support vector machine," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2983–2999, Nov. 2010.
- [15] S. Baluja, "Probabilistic modeling for face orientation discrimination: Learning from labeled and unlabeled data," in *Proc. NIPS*, 1998, pp. 854–860.
- [16] A. Blum and T. Mitchell, *Combining Labeled and Unlabeled Data With Co-Training*. San Mateo, CA: Morgan Kaufmann, 1998, pp. 92–100.
- [17] T. Mitchell, "The role of unlabeled data in supervised learning," in *Proc. 6th Int. Colloq. Cognit. Sci.*, 1999, pp. 1–8.
- [18] A. Fujino, N. Ueda, and K. Saito, "A hybrid generative/discriminative approach to semi-supervised classifier design," in *Proc. 20th Nat. Conf. Artif. Intell.*, 2005, vol. 2, pp. 764–769.
- [19] D. Yarowsky, "Unsupervised word sense disambiguation rivaling supervised methods," in *Proc. 33rd Annu. Meeting Assoc. Comput. Linguist., ser. ACL*, 1995, pp. 189–196.
- [20] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised selftraining of object detection models," in *Proc. 7th IEEE Workshop Appl. Comput. Vis.*, Jan. 2005, pp. 29–36.
- [21] V. R. de Sa, "Learning classification with unlabeled data," in *Proc. NIPS*, J. D. Cowan, G. Tesauro, and J. J. Alsppector, Eds., 1993, pp. 112–119.
- [22] U. Brefeld, T. Gärtner, T. Scheffer, and S. Wrobel, "Efficient co-regularised least squares regression," in *Proc. 23rd ICML*, 2006, pp. 137–144.
- [23] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [24] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proc. 16th ICML*, 1999, pp. 200–209.
- [25] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proc. 18th ICML*, 2001, pp. 19–26.
- [26] X. Zhu, "Semi-supervised learning literature survey," *Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, Tech. Rep. 1530*, 2005.
- [27] J. Muñoz-Marí, D. Tuia, and G. Camps-Valls, "Semisupervised classification of remote sensing images with active queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3751–3763, Oct. 2012.
- [28] Dópido, I., et al. (2013). "Semisupervised self-learning for hyperspectral image classification." *Geoscience and Remote Sensing, IEEE Transactions on* 51(7): 4032-4044.
- [29] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, Sep. 2009.
- [30] W.-K. Ma et al., "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Sep. 2014.
- [31] Samat, A., et al. (2016). "Improved hyperspectral image classification by active learning using pre-designed mixed pixels." *Pattern Recognition* 51: 43-58.

- [32] I. Dopido, M. Zorteza, A. Villa, A. Plaza, and P. Gamba, "Unmixing prior to supervised classification of remotely sensed hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 760–764, Jul. 2011.
- [33] I. Dopido, A. Villa, A. Plaza, and P. Gamba, "A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 421–435, Apr. 2012.
- [34] I. Dopido, J. Li, P. Gamba, and A. Plaza, "A new hybrid strategy combining semi-supervised classification and unmixing of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, Aug. 2014, Art. ID 2322143.
- [35] Li, J., et al. (2011). "Hyperspectral image segmentation using a new Bayesian approach with active learning." *Geoscience and Remote Sensing, IEEE Transactions on* 49(10): 3947-3960.
- [36] J. Boardman, "Leveraging the high dimensionality of AVIRIS data for improved subpixel target unmixing and rejection of false positives: Mixture tuned matched filtering," in *Proc. 5th JPL Geosci. Workshop*, 1998, pp. 55–56.