Active Learning: Any Value for Classification of Remotely Sensed Data?

M. M. Crawford, *Fellow, IEEE*, D. Tuia, *Member, IEEE*, H. L. Yang, *Student Member, IEEE*

Abstract

Active learning, which has a strong impact on processing data prior to the classification phase, is an active research area within the machine learning community, and is now being extended for remote sensing applications. To be effective, classification must rely on the most informative pixels, while the training set should be as compact as possible. Active learning heuristics provide capability to select unlabeled data that are the "most informative" and to obtain the respective labels, contributing to both goals. Characteristics of remotely sensed image data provide both challenges and opportunities to exploit the potential advantages of active learning. We provide an overview of active learning methods, then review the latest techniques proposed to cope with the problem of interactive sampling of training pixels for classification of remotely sensed data with support vector machines. We discuss remote sensing specific approaches dealing with multi-source and spatially and time-varying data, and provide examples for high dimensional hyperspectral and very high resolution multispectral imagery.

Index Terms

Active learning, classification, support vector machines, hyperspectral, high resolution multispectral, multi-view, spatial learning, adaptation

MMC and HLY are with the School of Civil Engineering and the Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN 47907- 1971 USA. E-mail: mcrawford@purdue.edu, hhyang@purdue.edu. Phone: +1 765 496 3224, Fax: +1 765 496 1466

DT is with Laboratoire des Systèmes d'Information Géographique, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. Email: devis.tuia@epfl.ch, Web: http://lasig.epfl.ch/ , Phone: +4121 693 57 85, Fax: +4121 693 57 90.

I. INTRODUCTION

Remote sensing technologies now acquire enormous quantities of data from ground based, airborne, and space-based sensors, providing unprecedented capability for mapping and monitoring dynamic processes over extended areas. The new generation of sensing technologies generates very large data sets, and applications of very high spatial [1], [2] and spectral [3], [4] resolution data are growing rapidly. Supervised classification methods that require labeled reference data for learning models are among the most popular approaches for analyzing these data. These approaches provide capability to generate representations of land use/land cover, map disasters, and monitor areas of risk due to climate change, to name a few, thus providing information to analysts and decision makers that is obtained in a non-intrusive way. However, these supervised models rely on observed data, and their performance is strongly dependent on the availability of representative labeled data for training [5]. For hyperspectral data, the problem is exacerbated by the large number of parameters that must be estimated in traditional parametric classification methods [6].

Unfortunately, obtaining labeled data is an enormous challenge for both researchers and users of supervised methods and has led to overuse of a limited number of data sets in the remote sensing classification literature. The large spatial extent and accessibility to sites often make acquisition of appropriate training sets via field surveys difficult and expensive, necessitating careful planning of field campaigns to collect reliable, informative class labels and spectra. Thus, resulting training sets typically contain a relatively small number of samples compared to the extended coverage of the scene over which class signatures may vary with local conditions. Labeled data may also be obtained via visual interpretation of high resolution images, although the process is subjective and dependent on the knowledge of the analyst. Pixels selected by photo interpretation occur in spatially contiguous groups, so the training set is typically highly redundant and may include noisy data and outliers that impact class statistics, leading to poor performance of the classifier. Even if spatially contiguous pixels can be useful in contextual classifiers, they are typically over-represented, and only a subset of these pixels contributes effectively to the development and performance of the classifier. Therefore, it is worthwhile to search for a small data set with high training utility, whereby both the human annotation cost and the computational load for training a classifier are reduced, but performance of the classifier

is not sacrificed. Also, given the abundance of unlabeled data, it is advantageous to exploit the spectral information in both the labeled and unlabeled data to yield improved classification results.

Active learning (AL) provides unique capability for addressing these issues. Like bagging strategies, active learning is a form of resampling [7]. It has been studied extensively in the field of machine learning and applied to problems in document retrieval and natural language processing [8]–[10]. Unlike bagging, which draws samples from the data pool (usually labeled) to create multiple versions of the training set with the goal of describing the statistical population, AL selects samples from the unlabeled data pool in a biased manner via query strategies that are designed to exploit properties of the classifier and the current labeled and auxiliary unlabeled data. This interactively constructed training set does not necessarily represent the entire data space, but is otherwise assumed to consist of samples that are the most informative and useful for the learner. The overall goal of AL is to obtain satisfactory classification performance with fewer labeled samples than those of conventional passive learning, where the training set is often selected randomly or manually without interaction with the classifier [11]. Thus, it potentially leads to greater information exploitation for the data and significant reduction of the annotation cost. Recently, AL has gained attention for classification of remotely sensed data (a survey in [12]), and has also been investigated for image segmentation [13], target detection [14] and regression [15]. The focus of this paper is to provide an overview of active learning strategies for supervised classification, summarizing some of the most popular approaches, then providing more details on methods that have been developed recently to address specific challenges and opportunities in analysis of remote sensing images. These include spectral redundancy in hyperspectral data, spatial redundancy in high spatial resolution data and spatial drift in class signatures. We illustrate these methods using two example hyperspectral data sets, chosen to show the potential contribution of active learning in these scenarios. Our results are based on the SVM classifier, which is natural for many active learning strategies, although other classifiers can be used effectively with many active learning methods.

The remainder paper is organized as follows: the active learning framework is presented in Section II. Methods focused on advances in active learning that are particularly relevant to classification of remotely sensed image data are summarized with examples in Section III (multi-view active learning), Section IV (spatial information in active learning), and Section V

(adaptation in active learning). Section VI includes a summary of observations and a look to future opportunities. The appendix contains a description of the data and the inputs to the classification experiments.

II. A BRIEF INTRODUCTION TO ACTIVE LEARNING

This section contains an overview of the main elements of a general purpose active learning algorithm. The notions of output-space optimization (Section II-A) and the concepts of active learning heuristics (Section II-B) and uncertainty and diversity (Section II-C) are discussed. The aim of this work is to outline problems specific to remote sensing that have been readdressed via AL, rather than presenting and comparing specific base heuristics. Readers interested in details of the methods can consult the work of Settles [11] or the survey by Tuia *et al.* [12], that provides a review of several active learning heuristics proposed for remote sensing. In the following, we summarize these methods for the sake of completeness.

A. Optimal network design in the output space

Finding a relevant training set for image classification (or retrieval of biophysical parameters) can be considered as the task of designing an optimal monitoring network [16], [17]: given a network, (the current training samples), we want to add new measurements in order to improve the current performance of the algorithm. In a remote sensing setting, this reduces to the task of finding new locations where the output can be measured either by a user or by a sensing device. In geostatistics, a large body of work deals with space-filling methods, aiming to fill the input space, often characterized by the spatial location of training samples [18]. In remote sensing, the focus has often been on systematic methods in which the samples are acquired on a regular grid or on stratified methods, where the number of samples is balanced according to an estimate of the abundance of the classes present in the image or to another relevant parameter for which greater variability corresponds to a larger number of required samples [19]–[21]. This last strategy in particular has facilitated improvement in results where data were obtained by random sampling, but it still requires prior knowledge of the relevant parameter on which to base the stratification. Moreover, the problem of spatial autocorrelation between samples is often disregarded [22]. These strategies correspond to an exploration phase, where there is no attempt to control the predictive power of the model directly, which would correspond to an exploitation phase. Active

learning seeks to fill this gap. Instead of optimizing the coverage of the input space, active learning considers the output space, i.e. the model's predictions, and ranks the potential new training samples according to the prediction confidence of the current model on each potential sampling location [23]. Roughly speaking, active learning answers the question *"which samples should be added to improve the generalization of a given model?"* Training sets designed with active learning are therefore model-specific and do not enhance exploration naturally. Samples which are expected to have the most impact on the current model are selected, and those samples for which the current model provides a prediction with high certainty are ignored.

B. Recipe for an active learner

Following the terminology of Li and Sethi [24], an active learner can be summarized as a quintuple (C, L, S, U, Q), where C is a classifier, L the labeled set used for training, S is a user who can discover the label of the samples queried in a pool of unlabeled candidates U. Samples in U are ranked by a criterion (or *heuristic*) Q. In a remote sensing setting, the sets L and U are composed of $d-$dimensional pixels, d being the total number of spectral bands and/or contextual filters used by the model. For pixels in L, the labels are known ($L = \{\mathbf{x}_i, y_i\}_{i=1}^l$), while for pixels in U, only the input vector is known ($U = \{\mathbf{x}_j\}_{j=1}^u$). Together, these sets cover the entirety of the n pixels of the image ($n = l + u$). Unlike systematic or stratified methods, C and the user S interact constantly in the active learning model: the first by predicting the output variable for the pixels in U and the second by providing labels for the samples highly ranked by the criterion Q. For this reason, active learning processes are naturally iterative, in the sense that for a given state ϵ and a corresponding training set L^ϵ, the classifier response C^ϵ will be different and will produce a different ranking of the candidates. Based on such ranking, the user S provides the labels, and the newly labeled pixels are transferred from U^ϵ to L^ϵ, thus creating the next training and candidate sets $L^{\epsilon+1}$ and $U^{\epsilon+1}$. The lower part of Fig. 1 illustrates the active learning component of the classification process.

C. The Evaluation Criterion Q

The heuristic Q differentiates active learning from traditional sampling strategies: Q is based on the output of the current classifier C^ϵ and relates only indirectly to the input space design (unless the heuristic is designed to enforce this relation. This is related to multi-view and spectral/spatial
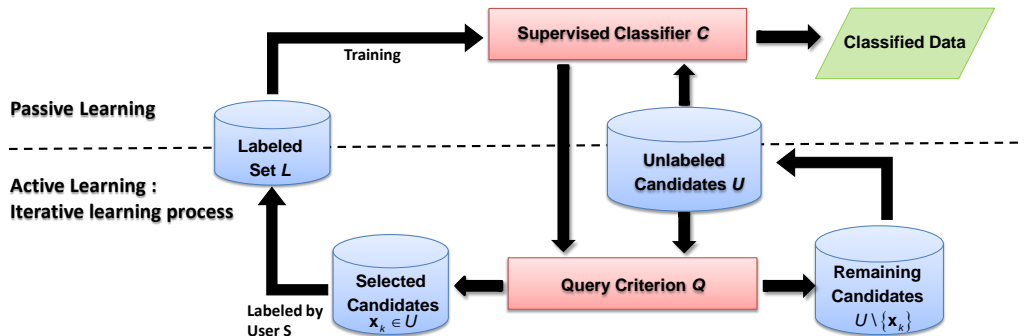
Fig. 1. General active learning flowchart using the quintuplet terminology of [24].

strategies and will be considered in Sections III and IV). The criterion is intended to provide information for ranking the candidate pixels relative to their potential contribution, based on the current classifier C^ϵ. Such a value can be assessed according to: the uncertainty/confidence of a pixel and its diversity in a batch or view perspective.

1) Uncertainty of pixel labels: unlabeled pixels are not equal in terms of "informativeness" for the current classifier C^ϵ. Considering support vector machine classifiers as an example, only pixels that have a chance of becoming support vectors are informative, because others would be disregarded if added to the training set $L^{\epsilon+1}$. In this sense, a good criterion Q must be capable of assigning high rankings to pixels that have a great chance of becoming support vectors. This is strongly related to the concept of pixel uncertainty: a pixel that can be correctly handled by the current model has almost no chance of becoming a support vector, while a pixel located in the margin, i.e. close to the support vectors, is uncertain, and therefore highly informative. This is a major difference from stratified approaches, where the pixels are balanced by a variability measure, but still chosen randomly, and therefore in a suboptimal way for the model.

Several heuristics are based on this concept: Margin Sampling (MS [25]) minimizes the distance to the closest hyperplane, while the Breaking Ties (BT, [26]) and the Multiclass-Level Uncertainty (MCLU [27]) methods consider the confidence of the two most probable classes. This

family of heuristics is the most studied in the active learning community (see, for example, works in [28], [29]). Approaches based on a committee of models have also been considered, either based on models trained on sample subsets of L^ϵ (e.g. Entropy Query-by-Bagging (EQB) [30]) or on subsets of the d-dimensional feature space (Multi-view [31], [32]). See Fig. 2 for a comparison of these last two types of AL architectures. When selecting a single pixel per iteration, these heuristics return the most uncertain (or informative) pixel in U^ϵ.

Heuristics based on committees have the advantage of being independent of the classifier (see Section III). The BT strategy can also be applied to all models outputting posterior probabilities (see examples in [26], where the heuristic is used with linear discriminant analysis or in [33], where it is used with multinomial logistic regression). The same holds for the approach proposed in [34], where divergence on the posterior probabilities is used either with the Maximum Likelihood or the BHC classifier.

2) Diversity in active learning: selecting a single pixel per iteration is not computationally efficient, since the classifier C^ϵ must be retrained at every loop with the new training set $L^{\epsilon+1}$. Moreover, pixels queried by only their uncertainty may be redundant relative to each other. Therefore, many studies have been devoted to the question of sample diversity [35]: if selecting many pixels at once (this selection is hereafter called a *batch*), the set of pixels must be as diverse as possible to avoid redundancy. In this way, for each state ϵ, the most effective batch of pixels is included in $L^{\epsilon+1}$. Diversity measures seek the pixels that are most dissimilar among those highly ranked by Q. The resulting batch contains diverse, uncertain pixels for the current classifier. Diversity is also important for active learning strategies that involve multiple inputs or "views ", where the goal is to exploit differences in information in the input space relative to the selected criterion, and to exploit potential parallelism in computation. Many methods have been proposed to achieve this diversity evaluation: they are either iterative (adding the most diverse pixel with respect to the current batch [27], [30], [36]), based on clustering of the uncertain pixels of U^ϵ [27], [37], or on various diversity metrics related to view similarity [32]). In these works, only the spectral diversity of the pixels or views was considered. Later, we consider the image-specific characteristic of diversity.

The remote sensing data classification problem has some unique and specific features, related to the fact that data are intrinsically multisource, spatial, and evolving in time. The next sections
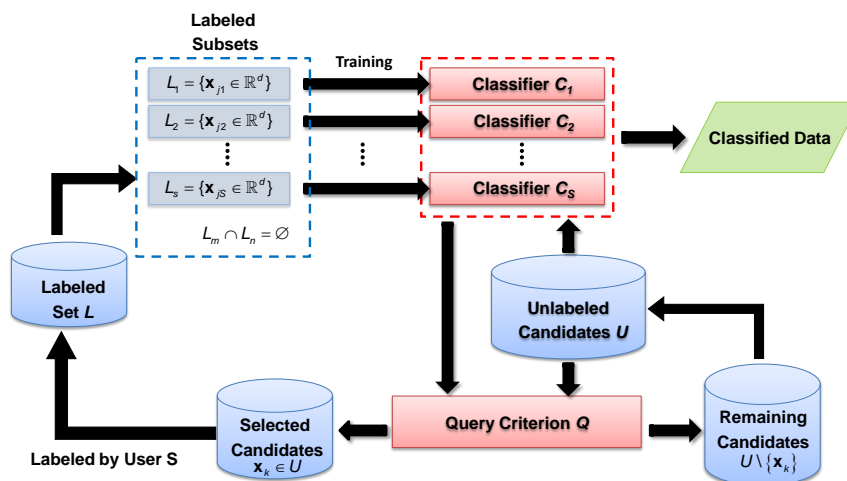
consider these specific characteristics in an active learning perspective and present approaches for exploiting these unique properties. In the following section, we introduce the concept of multi-view active learning, where subsets of disjoint inputs are provide a natural means of using ensemble strategies to increase classification accuracy and improve computational performance.
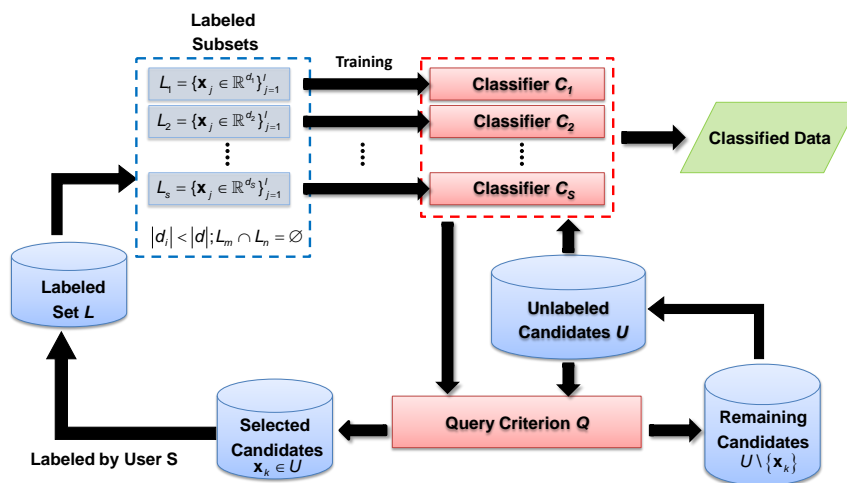
III. MULTI-VIEW ACTIVE LEARNING

Multi-view strategies, which provide different "views" of input data or alternative models, are particularly attractive for classification problems with high dimensional input spaces. In remote sensing image based applications, committees may be comprised of inputs from different sensors, subsets of the spectral bands, spectral and textural features derived from a given sensor, or different classification models. When the input space is large, random subspace feature selection can potentially provide improved classifier diversity, while stabilizing parameter estimates. This is related to the reduction of the number of inputs to each classifier and to the construction of multiple classifiers in the resulting input space [38]. The inconsistency of outcomes from the classifiers that comprise an ensemble or committee is exploited by popular approaches such as "query by committee" (QBC) [39] and its variants, as noted in Section II, and many active learning strategies utilize the value of agreement [30], [40], [41] of a candidate point relative to the current classifier as a criterion for selection (see Fig. 2a). This idea is naturally extended to multi-view approaches, as disagreement among ensemble members is an indicator of the samples with the most confusion about outcomes, which implies high value for the next query.

Appropriate selection of views is key to multi-view strategies: individual views should be capable of learning the correct outcome (i.e. sufficient), they should provide complimentary information [42], contribute to diversity, and provide accurate information relative to the decision. Multi-view approaches are characteriized both in the way they generate the committee members and in the combination strategy of their respective outputs. This implies different implementations that should be designed relative to the characteristics of the particular data and classification problem.

Although multi-view active learning is common in applications such as text classification, it has received limited attention in the remote sensing community. The inherent redundancy of hyperspectral data facilitates the construction of compatible views, and there is little risk of under-representation since all the spectral bands can be included. However, accuracy of views

Labeled Subsets

$L_1 = \{\mathbf{x}_{j1} \in \mathbb{R}^d\}$

$L_2 = \{\mathbf{x}_{j2} \in \mathbb{R}^d\}$

$L_s = \{\mathbf{x}_{jS} \in \mathbb{R}^d\}$

$L_m \cap L_n = \varnothing$

Training

Classifier C_1

Classifier C_2

Classifier C_S

Classified Data

Labeled Set L

Unlabeled Candidates U

Labeled by User S

Selected Candidates $\mathbf{x}_k \in U$

Query Criterion Q

Remaining Candidates $U \setminus \{\mathbf{x}_k\}$

(a) Flow chart for Query-by-Committee Active Learning

Labeled Subsets

$L_1 = \{\mathbf{x}_j \in \mathbb{R}^{d_1}\}_{j=1}^l$

$L_2 = \{\mathbf{x}_j \in \mathbb{R}^{d_2}\}_{j=1}^l$

$L_s = \{\mathbf{x}_j \in \mathbb{R}^{d_s}\}_{j=1}^l$

$|d_i| < |d|; L_m \cap L_n = \varnothing$

Training

Classifier C_1

Classifier C_2

Classifier C_S

Classified Data

Labeled Set L

Unlabeled Candidates U

Labeled by User S

Selected Candidates $\mathbf{x}_k \in U$

Query Criterion Q

Remaining Candidates $U \setminus \{\mathbf{x}_k\}$

(b) Flow chart for Multi-view Active Learning

Fig. 2. Comparison between Query-by-committee (a) and Multi-view (b) approaches.

on unlabeled data cannot be known, and research on view validation is limited. Interestingly, the "curse of dimensionality" [43] that plagues supervised classification of hyperspectral data is further exaggerated in the context of active learning, but can be alleviated by multi-view approaches. Strategies that decompose the feature space into low-dimensional, mutually exclusive subsets are also computationally advantageous for large scale problems, as they can be
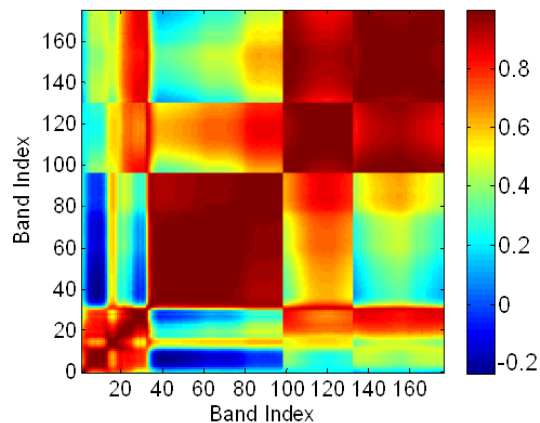
Fig. 3.   Correlation coefficient matrix of KSC hyspectral data on Area 1.

implemented in parallel architectures.

Di and Crawford [32], conducted an extensive study of view generation (VG) strategies for a maximum disagreement based active learning strategy for classification of hyperspectral data. They considered generation of views based on correlation of contiguous bands, k-means based band clustering, deterministic selection of every "kth" band (band slicing), and random sampling. Views generated by correlation and clustering are diverse, but may differ in their discriminative ability for individual classes, so there is a risk of insufficiency, while views obtained from "band slicing" may be redundant, but are sufficient. Finally, random sampling provides diverse views, but they are not guaranteed to be either sufficient or accurate. Figure 3 illustrates multi-view subsetting of the input space of the KSC Area 1 data based on interband correlation (see Appendix for details about the data). For these data, blocks of contiguous bands along the diagonal that are highly correlated define the correlation based subsets. Figure 4 illustrates progression of active learning in a multi-view scenario when using 5 views of the KSC Area 1 data generated by correlation (C_r), clustering (C_k), uniform band slicing (U_s), and randomly generated (R_G), compared to Random Sampling (RS) and Margin Sampling (MS), where the latter computed by using the minimum distance among the One-Against-All (OAA) hyperplanes.

All methods except random sampling converge to a high accuracy outcome with fewer than 100 queries. For this data set, margin sampling has the highest overall accuracy, although multi-view methods converge to approximately the same overall accuracy, reduce impact of high
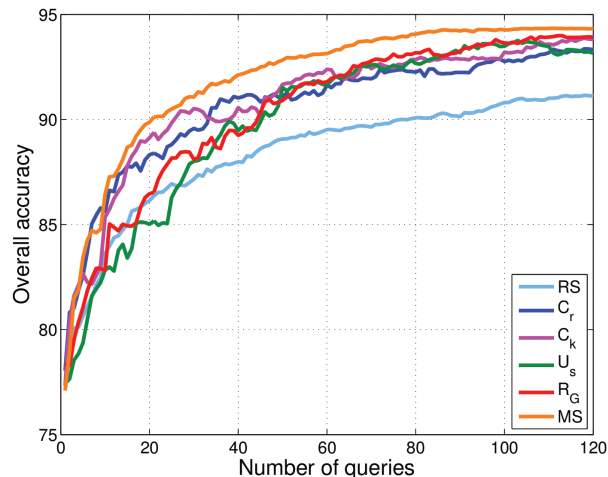
Fig. 4. Overall classification accuracy for SVM classification of KSC Area 1 with multi-view active learning based on band correlation, k-means clustering, uniform band slicing, and random view generation compared to random sampling and margin sampling.

dimensionality when training samples are limited, and can be parallelized to reduce computational overhead. Even randomly generated views outperform random sampling. Fig. 5 shows the detail for correlation-based view generation: as learning progresses, views tend to agree with each other, except for View 1 (which primarily covers the blue portion of the spectrum for these data), leading to reduced rate of learning, as exhibited in Fig 4. Noting this, [32] explored approaches to mitigate this problem, including increasing the number of views, switching to different view generation methods, and using bagging mode random view generation. The use of multiple strategies coupled with thresholds for switching based on learning rate was found to be quite effective.

Trade-offs between view diversity, sufficiency, and accuracy associated with different strategies influence performance of multi-view active learning. Selection of diverse, reliable views for a given problem is the most important factor relative to the overall success of these methods. Application of domain knowledge by the analyst can also be important to this end.

IV. INCLUDING SPATIAL INFORMATION IN ACTIVE LEARNING

Compared to other sources of information, remote sensing images are intrinsically spatial on a grid and have geospatial coordinates. Spatial relationships have been exploited in remote
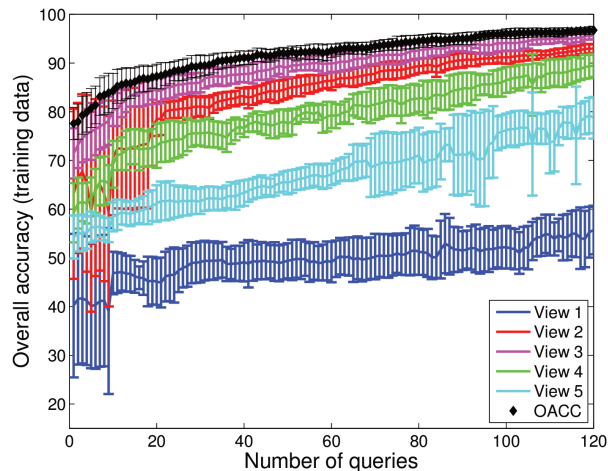
Fig. 5.   Overall classification accuracy (OACC) and view performance derived from correlation based view generation for KSC AVIRIS data.

sensing research to improve classifiers by incorporating spatial information via neighborhoods or contextual inputs [44], [45], or by segmenting images into spatially coherent and spectrally similar segments [46]–[49]: in both cases, constraining the model with spatial information results in increased performance. As for traditional classification, active learning strategies can benefit from the inclusion of spatial information: most heuristics presented in the literature deal with spectral data only (even if the manifold is considered, regularization is performed in the spectral space), and few take into account the *position* of the samples, either in terms of pixel location in the image or the clusters to which they belong. Recently, this issue has been considered, mainly for two reasons:

- Active learning is a set of techniques aimed at building training sets: when new acquisitions are scheduled, AL can be used as a tool for planning the measurement campaign. In this case, the spectral criterion is used to sample pixels that are useful and different from each other, but a spatial criterion can be useful for distributing measurements in the geographical space and handling variations in local conditions, which can also be source of spectral signature drift. (see Section V).

- When acquiring new samples, a spectral criterion is not spatially informative (see Fig. 6). Adding spatial information i) further differentiates the samples and ii) permits selection of samples that are not spatially adjacent, thus promoting more robust field sampling.

(a) Spectral criterion   (b) Spatial criterion   (c) Combined criterion
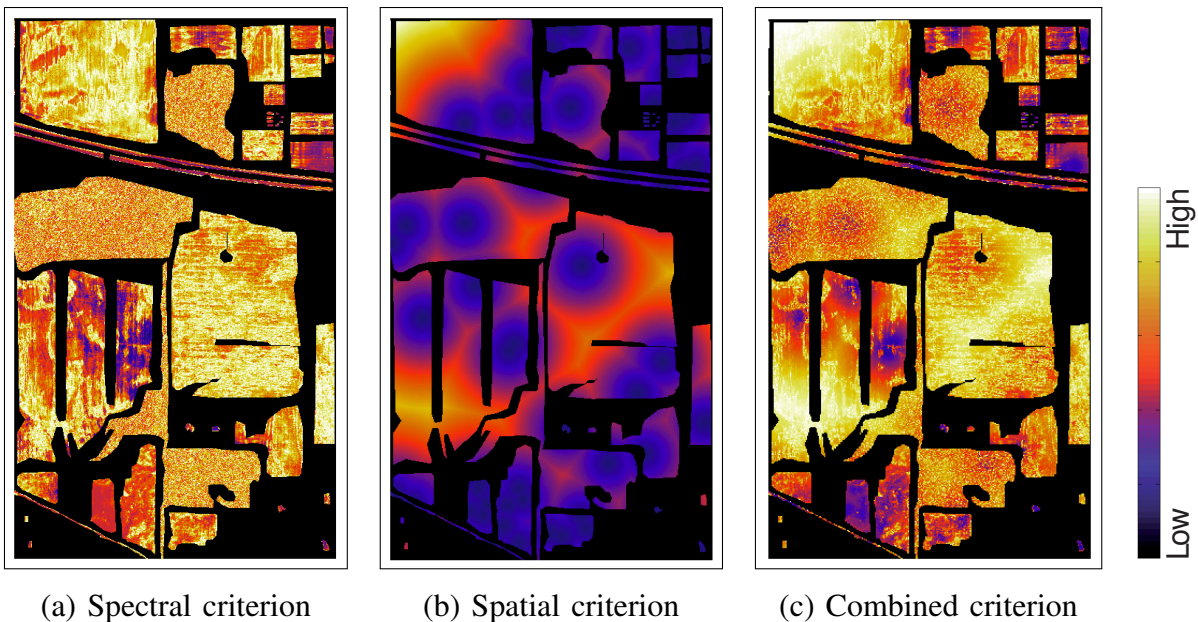
Fig. 6.   Spectral/spatial active learning with spatial/spectral information: (a) spectral MCLU criterion [27]; (b) distance to current support vectors in the spatial domain; (c) combined criterion exploiting non-dominated solution between (a) and (b) [54]. In all three maps, bright tones highlight interesting regions for new samples. (see Appendix for details about the ProSpecTIR data)

For these reasons, a spatially differentiating criterion can be of interest. In recent literature, three types of spatial enforcing criteria have been proposed:

a) Spatial information seen as the minimization of the "travel" distance to cover: in [50], the authors encode a spatial criterion associated with overall travel time as a function to be minimized. Starting from a known location, they solved the traveler's salesman problem [51], [52], in which information is maximized (through the spectral criterion), while the travel distance between locations is minimized. This setting could be useful when the team in the field has limited resources for traveling. The approach operates on an unrestricted geographical space; more work is needed to convert it into an effective campaign planning tool that considers other inputs, such as existing road networks and local terrain. Another recent work [53] explores these possibilities including information about topography and road networks.

b) Spatial information seen as a metric to minimize selection of spatially collocated samples: research in sample diversity [27], [30] deals with the selection of batches whose pixels are not spectrally redundant. An alternative approach is to track variability with respect to samples selected at previous iterations, as in [37]. The key idea is that important samples are selected

during the active learning process, and the efficiency of the system decreases if at the next iteration, pixels are similar to those selected previously (even if they are spectrally different). This intuition can be extended in spatial (geographical) terms: pixels that are geographically very close to current support vectors (initial training samples or pixels selected in previous iterations) have a higher risk of providing the same information as those already in the model. Moreover, avoiding a well characterized region also provides the opportunity to discover isolated patches and new data structures, where the spectral distribution of a given class may have shifted spatially or temporally (see Section V). In [54], the distance to current support vectors is used as a spatial criterion that competes with the spectral heuristic. By identifying the set of non-dominated solutions (the Pareto front [55]) between the spectral and spatial criteria, the authors seek to perform queries that are simultaneously spectrally informative (Fig. 6a) and spatially distinct from previously sampled areas (Fig. 6b). Fig. 6c shows the non-dominated solutions for the ProSpecTIR image (see Appendix), from the first front (bright tones, interesting locations) to the last (dark tones, corresponding to areas either spectrally clear or too close to current support vectors).

c) Spatial information incorporated in a segmentation problem: a different problem setting would be the parallel definition and labeling of a segmented image [13]. Given a spatially coherent and hierarchical partitioning of an image (a hierarchical segmentation [46], [48]), the aim is to find the level of clustering (i.e. number of clusters, as well as their spectral coherence) for which every cluster belongs to a single class. Consider the clustering algorithm proposed by [46]: using this algorithm, a hierarchy of possible partitionings of the data is retrieved, and then an attempt is made to label the clusters using class distribution information. Active learning can be used to minimize the number of queries to provide a single label to each cluster considered pure enough to belong to a single class. For the strategy to be effective, the segments in the cluster hierarchy should be homogeneous, in order to retrieve labels that are coherent for the cluster with few queries. Fig. 7 illustrates these principles for the ProSpecTIR agricultural data of Indiana (see Appendix): without using spatial information (top row of Fig. 7), the clustering of this high resolution image is spatially ambiguous, and the segments considered as pure by the algorithm are either overestimated (at the beginning, where only very few labels are available) or overfit (when more labels are available) resulting in complex areas being over-segmented. By using spatial information (middle and bottom row of Fig. 7), coherent clusters

are determined, and few labeled pixels are necessary to label the fields correctly. Using an active learning strategy (the $s_4 d_2$ strategy of [13]; see the bottome row of Fig. 7), an accurate classification map is retrieved with just 240 queries, – although it should be noted that this is related in part to the regular spatial geometry of agricultural scenes. Another interesting application of active learning for (supervised) segmentation is presented in [33], where active learning heuristics are applied to increase the training set used to perform image segmentation based on the pixel's neighborhood and multinomial logistic regression. Finally, a last recent approach [56] queries spatial segments within regions of maximal uncertainty and diversity: this strategy allows querying multiple segments simultaneously from the region.

All three approaches for incorporating spatial information into the active learning process assume that the class distributions are stationary. For remote sensing data, this may not be valid, even over small areas. For these problems, the areas of knowledge transfer and adaptive strategies have been investigated recently in conjunction with active learning. Research dealing with the application of active learning to adaptation is presented in the next section.

V. ACTIVE LEARNING FOR KNOWLEDGE TRANSFER AND ADAPTATION PROBLEMS

Remote sensing data are impacted by local conditions during acquisitions, which result in variations in class signatures across a scene, or between images acquired at different times. In a time series of images acquired by a given sensor, factors such as the acquisition date/time, atmospheric conditions, and incidence angle impact the observed spectra of vegetation and materials. These changes, or *shifts*, reduce the portability of classification models. By portability, we mean the possibility of re-applying a model, which is optimal for an image (or subset of an image), to another area of a large scene or to another image, without obtaining labeling information on the second (or with very limited labeled information). Hereafter, these images for which knowledge transfer is needed are referred to as *source* and *destination domains,* and the general problem of model portability is referred to as *(domain) adaptation*.

Direct application of a classifier to class-specific spectral data that have drifted yields low classification accuracies. The top row of Fig. 8 illustrates this principle for two classes of the KSC dataset, which consists of two spatially disjoint areas (see Appendix for details): on the left plot, the average Graminoid marsh signature is shown for the two regions, while the right plot contains the Salt marsh signatures. In both cases, the average signature of these classes is
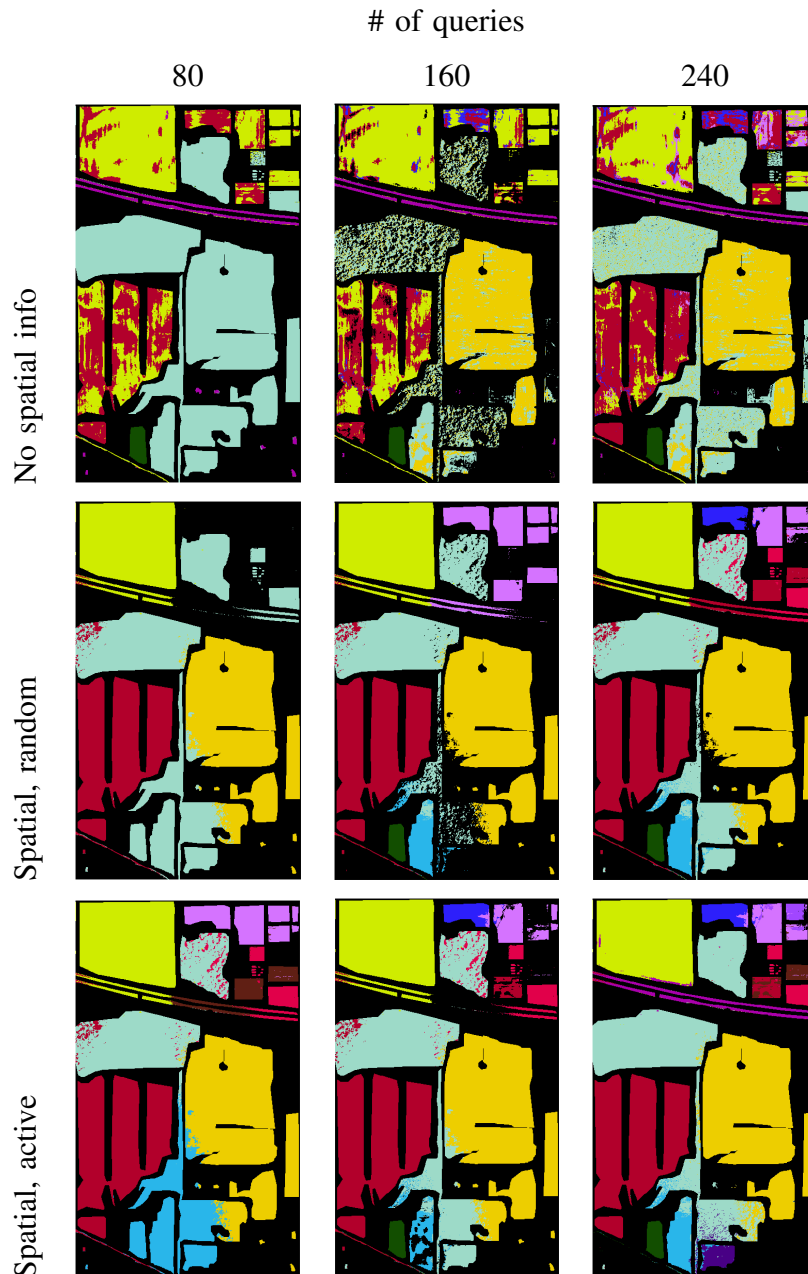
Fig. 7.   Active learning for labeling a segmentation hierarchy. Best classification per number of queries using bisecting k-means [57]. (top) result after building the clustering without spatial constraints; (middle and bottom) results using constraint on cluster contiguity using (middle) random queries and (bottom) active queries [13]. For legend colors, please refer to Table II

significantly different. For Gramonoid marsh, the change appears to involve linear scaling of the signature, while for Salt marsh, the difference is wavelength-dependent: the average signature decreases for low wavelengths, but increases for longer wavelengths, necessitating alternative

approaches. Considering classes showing similar signatures (bottom row of Fig. 8) yields a similar conclusion: class signatures change from one area of the flightline to the other. For instance, Oak/broadleaf hammock has the highest response around $1000nm$ in the first area, while in the second area, the highest signal is provided by Cabbage/oak hammock, highlighting the difficulty in analyzing complex natural scenes. When considering the data manifolds for all the classes (Fig. 9a-b), the effect of the shift on a classifier becomes evident: either the centers of gravity of the different classes or the border patterns (those that will become support vectors) are displaced by the spectral shift, and direct application of a model trained in Area 1 will result in many misclassified pixels if applied to Area 2 (see also Fig. 9c, where 100 training examples from Area 1 are superimposed on the labeled examples of Area 2). These differences between the local conditions can result in low accuracies if transferred directly to the second area: a model trained on 500 randomly selected pixels from Area 1 achieved an overall accuracy of 90.6% on

(a) Graminoid marsh

(b) Salt marsh
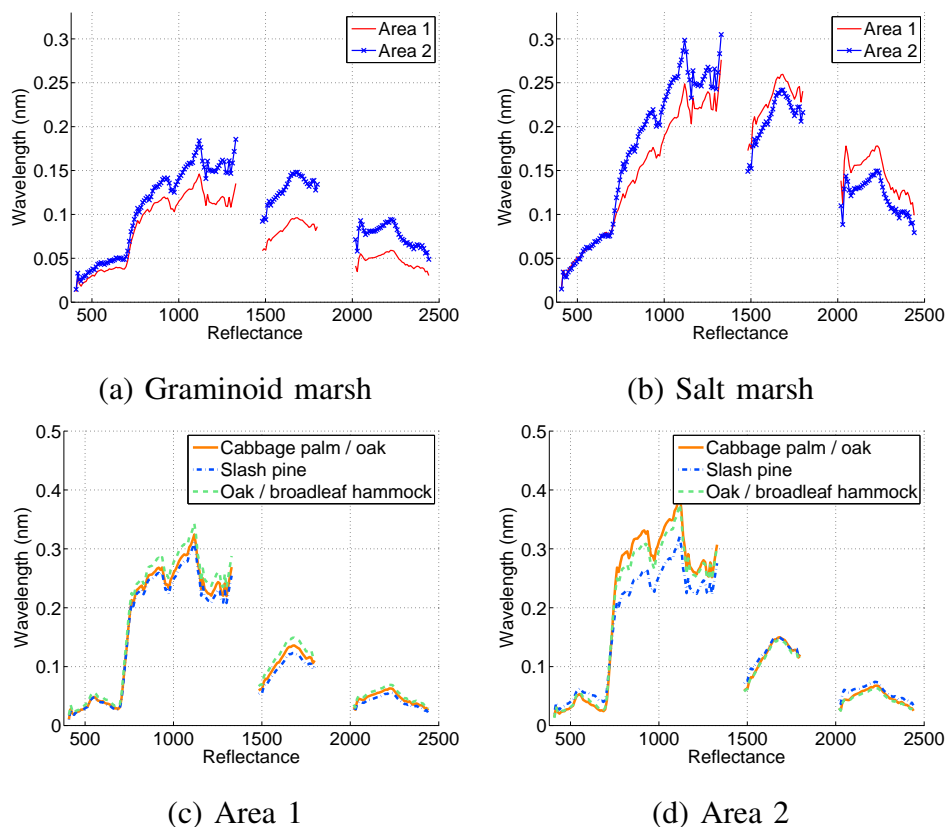
(c) Area 1

(d) Area 2

Fig. 8. Top: average spectral signature for classes (a) Graminoid marsh and (b) Salt marsh found in the two disjoint areas of the KSC image. Bottom: signatures of three similar classes for (c) Area 1 and (d) Area 2.

(a) Area 1   (b) Area 2   (c) Combined
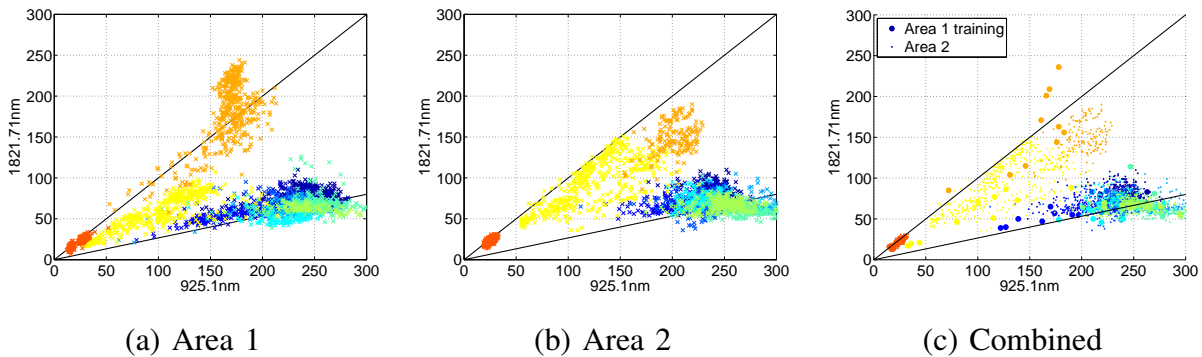
Fig. 9.   (a)-(b) Feature space visualization of the ten classes for the two areas of the KSC image. (c) Superposition of 10 training pixels per class taken from Area 1 (large dots) on the Area 2 samples (small dots). Each color represents a landcover class.

the test set of the same area (average of 10 experiments). When applied to Area 2, it achieved only 62.5% overall accuracy on the test set (see the black bullet in Fig. 10). A model trained with pixels from Area 2 achieved 89.2% on the same test data (blue line in Fig. 10). Model portability has been considered by many researchers for analysis of data over extended spatial areas and slowly varying multi-temporal scenarios using various approaches, including signature extension through clustering [58], spatially invariant features [59] obtained by spatial detrending with Gaussian processes in [60], and by manifold alignment [61]. The resulting classifiers were more robust to local shift in areas where training samples were unavailable. Finally, approaches inspired by semi-supervised learning have been adapted to the domain adaptation problem: by considering the data distribution on the target domain as the unlabeled samples, classifiers are modified to be more robust when applied to the target domain. A precursor of this research direction was the work of [62], where class statistics are recomputed and updated using the target domain for a Gaussian Maximum Likelihood classifier. Later, an adaptation strategy based on SVM (the DA-SVM [63]) was proposed: in this model, the SVM is constrained by adding and removing support vectors from both domains in an interactive way. Another semi-supervised adaptation strategy [64] is based on an ensemble of classifiers which are pruned using a measure of diversity of predictions in the destination domain. In [65], a manifold regularization approach was used to adapt the classifier to new domains. All these solutions assume that no labels can be acquired in the target domain. Relaxing this assumption by allowing *some* pixels to be sampled in the target domain, we can also conjecture that the most useful pixels to be sampled are those
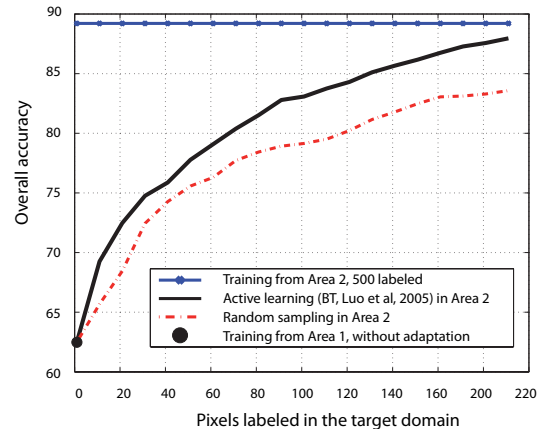
Fig. 10. Active learning used to correct the shift between class distributions of the two disjoint areas of the KSC dataset. Curves are averages over ten experiments using SVM. The initial training set is composed by 500 pixels randomly selected from Area 1.

that are located in the areas where signatures changed, which are, with high probability, close to the classification boundaries among classes. This intuition justifies the use of active learning methods to learn the shift, in order to sample pixels in spatially disjoint areas and then to modify the models accordingly [66]. Fig. 10 illustrates this principle, where active learning is used to add pixels from the target domain to the training set: the accuracy increased consistently and was higher than random selection, almost achieving the overall accuracy of a model developed using only 200 pixels from Area 2. Recent research extends the understanding of how active learning can be used to learn and correct the shift:

- Migration of samples from source to destination domains [67], [68]: in this case, the most relevant samples from the target domain are added to the training set (as in the previous example), while the less relevant samples for the source domain (estimated by the loss of confidence in classification for source pixels) are removed. Thus, the training set becomes less and less dependent on the source domain. In [67], the samples are removed from the training set, while in [68], they are deweighted using a sample-weighted SVM. In both cases, active learning strongly improves adaptation of the model. This type of approach can be very effective when using classifiers based on class statistics, since they eliminate samples that impact the class statistics in the target domain.

- Discovering new classes [26]: in this work, an unsupervised criterion is proposed to highlight

dense and unsampled areas of the target domain, where new classes could be hidden. These studies also show the potential of active learning for the adaptation of classifiers on multitemporal sequences. Selecting the samples of the destination domain drastically reduces the number of labels to be acquired to ensure the efficacy of a model on a new image of similar properties. In this context, model portability, which is critical for a discipline where massive quantities of data (images) acquired exceeds the labeling effort achievable by operators is greatly enhanced.

VI. OBSERVATIONS AND FUTURE DIRECTIONS FOR AL IN REMOTE SENSING

The goal of this paper is to introduce Active Learning (AL) in the context of remote sensing, particularly with respect to classification. We have concentrated on adaptations of classic AL strategies that focus on issues that are relevant to supervised classification and have used hyperspectral and very high resolution (VHR) imagery to illustrate these heuristics. AL provides a means to more effectively exploit information from labeled samples that may have been acquired previously without regard to their potential contribution to training a classifier. It can also be useful for planning ground surveys to collect reference. Although the assumption that pixel labels can always be obtained is not always practical or even possible in remote sensing applications, AL can still be valuable in supervised classification of remotely sensed data. A study of the effectiveness of AL in returning pixels that analysts are able to label by photointerpretation has recently been published in [69]. Results show that, to be effective, AL strategies must include a measure of the confidence of the user in labeling, as well as the pixel's uncertainty/diversity. Highly uncertain areas that often correspond to shadow/border areas cannot be easily labeled by an operator, and correspond to wasted queries. As a consequence, an AL heuristic often needed twice as many screened pixels as random sampling to provide the batch of labeled pixels. This is not desirable, since it leads to fatigue of the operator, frustration, and potentially increases in mislabeling. To address this, authors in [69] trained a second model, learning which examples were easily recognizable by the user, thus creating a filter for difficult pixels and minimizing the number of screened pixels to obtain the batch. This approach combined the class uncertainty of pixels selected and the recognition capacity of the user.

We used the state-of-the art SVM classifier with a One-Against-All (OAA) strategy to illustrate these concepts. Distance measures in the OAA framework were more easily implemented for

some active learning schemes, although One-Against-One (OAO) and hierarchical strategies can be more effective in discriminating among pairs of overlapping classes and for implementation of some methods. Although strategies related to margin sampling are naturally linked to SVM classifiers, other classifiers or ensembles of classifiers can also employed in conjunction with AL. Our experiments demonstrate that multi-view active learning can potentially exploit information from remotely sensed data in a variety of ways, including use of multi-sensor platforms. It can also be implemented in parallel computing architectures to address the inherent computational burden of AL. However, the effectiveness of multi-view AL is dependent on the reliability of each view relative to the classification problem. Unfortunately, this cannot generally be known in advance, although knowledge of class characteristics can guide view specification. Preliminary results also illustrate the potential of view switching for maintaining high learning rates throughout the sampling process, reducing the number of iterations. Future work could explore constraints on view accuracy via joint optimization of the sampling and view selection strategies.

Spatially focused AL is particularly effective for extended homogeneous areas (e.g. agriculture or large water bodies) and in VHR data, where class boundaries are associated with a large portion of image pixels. Incorporation of spatial information in AL was shown to reduce spatial redundancy and utilize non-contiguous clusters of labeled data effectively in choosing training samples. Published strategies have focused primarily on inclusion of contextual information from texture transforms and local spatial neighborhoods, either directly or via segmented images. These approaches also provide capability to mitigate the impact of within-class variability on classification observed in traditional spectral-based classification methods. Future studies could explore use of hierarchical multi-resolution classification strategies and inclusion of constraints to introduce additional realism into models (e.g. terrain and infrastructure related limitations). Semi-supervised approaches focused on reducing the user supervision of hierarchical segmentation tasks or easing the choice of unlabeled pixels for semi-supervised classifiers are also new promising directions, where AL interacts with spatially consistent information.

Classification of large scenes and multi-temporal data necessitates adaptation of classifiers to local conditions, while maintaining some degree of generalization. Detection of change, including identification of new classes, is also very important. Traditional methods that are based on labeled data from the original subset of an image do not provide capability to "learn" about these changes

quickly. Semi-supervised approaches have traditionally been used to accommodate these issues, although they are traditionally implemented when labeled data are limited, but the spectral distributions of classes are stationary. If labels can be obtained, appropriate implementations of AL also provide an effective tool for classification of data in dynamic environments. Algorithms that exploit both AL and semi-supervised approaches could be extremely useful, but are largely unexplored to date.

We have explored AL strategies for supervised classification problems in remote sensing, but the framework is useful for other problems as well, motivating further research in several areas. Because of its capability for learning "on the fly", AL provides capability for near real time analysis via onboard processing. Investigations of new approaches for handling multi-sensor problems are also needed, particularly for SAR, where speckle would impact the value of obtaining a label for a given pixel. Classification results for most investigations, including here, are reported in terms of overall accuracies. In reality, AL can be most beneficial for the "difficult" classes, as it provides a rational approach for investigation of "what is unknown". Class specific accuracies indicate that the greatest gain is often in terms of the those classes that are hard to discriminate [70]. More directed research in the area of "critical class active learning" may also be merited.

As noted in the Introduction, AL has also been investigated for segmentation and unmixing in a few studies, and additional efforts toward these applications could be fruitful. Moreover, regression and function estimation could largely benefit from the AL framework, but, for now, only a few attempts have been published on the topic [15]. Another key application where active learning is gaining popularity is change detection, where co-registered images are analyzed to detect changed regions: specific algorithms have been proposed recently, that either rely on training sets obtained by transferring the labels of unchanged samples from the source to target domain [71], on joint conditional probabilities among acquisitions [72], or on target detection models and exploration of the feature space [73]. Finally, it must be acknowledged that AL methods are problem-driven heuristics which are often the most advantageous for two very dissimilar problems: "reconnaissance" or operational implementation [74] where the algorithms are specifically designed to address the unique characteristics of the application and computational environment. Ample opportunities remain for advancing the state of both.

VII. ACKNOWLEDGEMENTS

APPENDIX

EXPERIMENTAL DESIGN AND REMOTELY SENSED DATA SOURCES

Experiments in this paper are designed to illustrate unique characteristics of the various active learning strategies, rather than for comparison of absolute accuracies. In practice, the performance of active learning for classification is strongly dependent on the characteristics of the data, the classifier, and details of implementation. Our goal is to provide results of a set of experiments conducted under consistent conditions using data which illustrate both the opportunities and challenges of active learning methods for remote sensing applications.

All experiments utilized a radial basis function (RBF) SVM [1] classifier implemented in the ALToolbox [12] available at `http://code.google.com/p/altoolbox/`. Parameters were obtained by cross validation via grid search[2]. Results for each method are reported in terms of overall performance of ten-fold cross validation and compared to baseline methods of random sampling and margin sampling. The methods were implemented in a multi-class One-Against-All (OAA) design for all experiments, except the adaptation experiments, where a One-Against-One (OAO) strategy was employed to obtain posterior probabilities. It should be noted that in preliminary experiments using these data, the OAA strategy provided somewhat higher overall classification accuracies than the OAO approach, while the standard deviation of the overall accuracies for the OAO implementation was lower. OAO strategies were also more robust for discriminating "difficult" pairs of classes [70]. All experiments were based on the

---

[1]SVM is one of the most popular classifiers for hyperspectral data [75]. For an introduction to the mathematical formulation, the reader is referred to [76].

[2]This procedure finds the best SVM parameters by grid search. For each combination, the labeled data (limited in an AL process) are split into k sets. The classifier is trained k times, where the $k-1$ of the sets are used to predict the classes of the remaining validation set. After predicting each set once with models trained on the rest, an average training error is estimated. The best parameters are those related to minimal error.
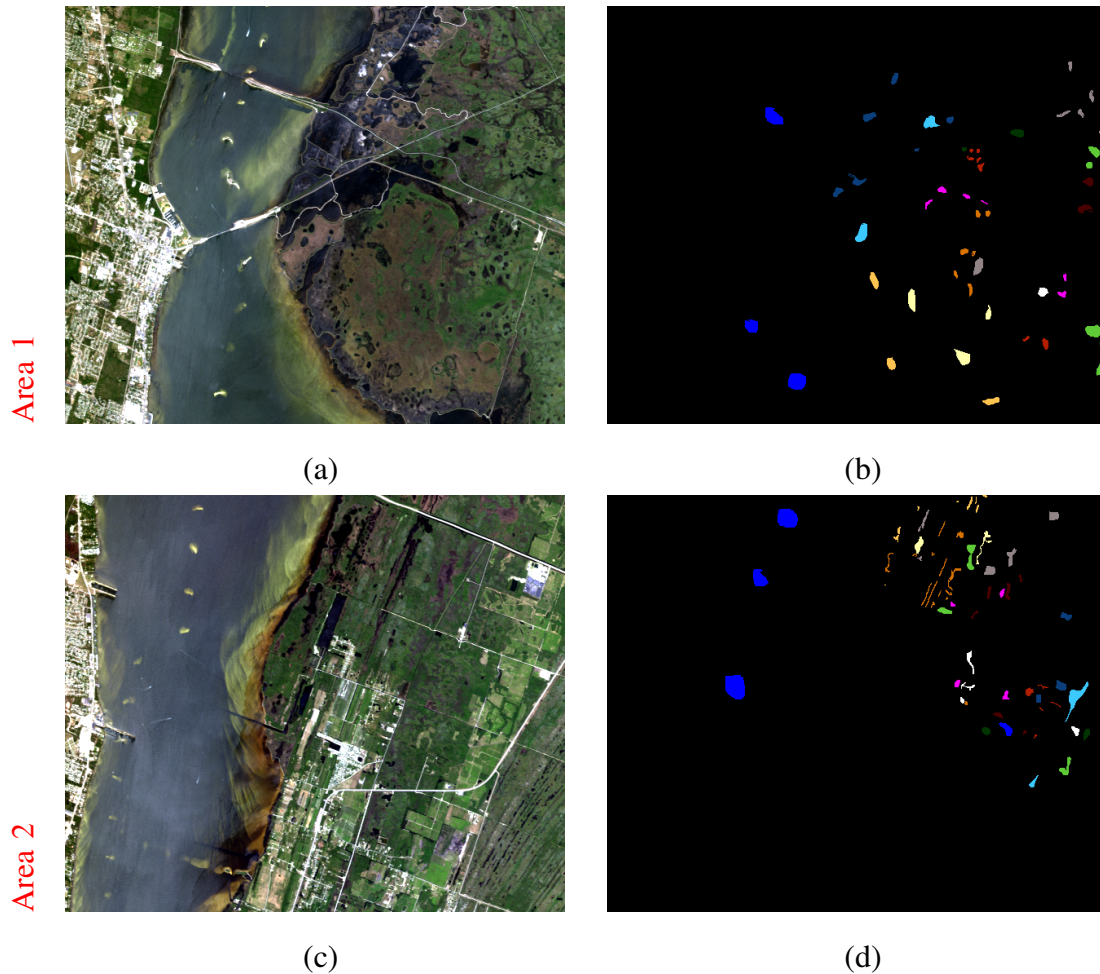
Fig. 11.   (a), (c) RGB images and (b), (d) Class label images of KSC Area 1 and Area 2, respectively.

same labeled data, which were sampled to provide 50% of the data for training (that compose the initial training set L^0 and candidate set U^0) and 50% for validation.

Two airborne hyperspectral remote sensing data sets were used in the experiments: Kennedy Space Center (KSC) and Indian Pines 2010. These data were selected to illustrate the characteristics of the various active learning strategies.

- *Kennedy Space Center* (KSC)

  The NASA AVIRIS instrument collected data at 18m spatial resolution over the Kennedy Space Center (KSC), Florida, on March 23, 1996. The data consist of 224 bands of 10-nm width from 400 to 2500 nm. After removing water absorption and low signal-to-noise (SNR) bands, 176 bands of reflectance data were used for the analysis. Training data were

selected using land cover maps derived from color infrared photography provided by KSC and Landsat Thematic Mapper (TM) imagery. The vegetation classification scheme was developed by KSC personnel in an effort to define functional types that are discernible at the spatial resolution of Landsat and this AVIRIS data set. The scene is comprised of water, uplands woodland classes, and lowlands marsh classes, which occur in patches on the landscape. Discrimination of land cover for this environment is difficult because many classes are mixed, and the spectral signatures for certain vegetation types are similar. Also, the coastal waters are very clear, so bottom signatures are visible near shore. For classification purposes, 10 classes representing the various land cover types that occur in this environment were defined for the site. Two spatially disjoint subsets of a flightline, referred to as Area 1 and Area 2 were analyzed in the experiments. Fig. 11 contains RGB images of the data and the classes that are common to both areas. The classes and number of training samples per class are listed in Table I. Active learning was initiated with 3 points per class, and batch sizes of 5 were selected.

TABLE I

CLASS LABEL OF KSC DATA

ID	Color	Class name	Area 1 No. of samples	Area 2 No. of samples
1		Scrub	761	422
2		Willow swamp	243	180
3		Cabbage palm hammock	256	431
4		Cabbage palm/oak	252	132
5		Slash pine	161	166
6		Oak/broadleaf hammock	229	274
7		Hardwood swamp	105	248
8		Graminoid marsh	431	453
9		Salt marsh	419	156
10		Water	927	1392

- *Indian Pine 2010*

  The ProSpecTIR system acquired multiple flightlines of data over agricultural areas near Purdue University, West Lafayette, Indiana, on May 24-25, 2010 for a study focused on estimating residue cover to evaluate tillage practices. The image subset analyzed in this

study contains 445×750 pixels at 2-m spatial resolution, with 360 spectral bands of 5-nm width. Sixteen land cover classes were identified, which included fields of different crop residue covers, vegetated areas, and man-made structures. Many classes have regular geometry associated with fields, while others are associated with woodlands, roads, and isolated man-made structures. An RGB image of reflectance data and a class map of the area are shown in Fig. 12, and Table II contains class labels and number of training samples per class. Active learning was initiated with 5 points per class, and batch sizes of 10 were selected.

(a)   (b)

Fig. 12.   (a) RGB image and (b) Class label image of Indian Pine 2010 SpecTIR data

REFERENCES

[1] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2587–2600, 2006.

[2] F. Pacifici, M. Chini, and W. J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 1276–1292, 2009.

[3] D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," *IEEE Signal Process.*, vol. 19, no. 1, pp. 17–28, 2002.

[4] G. Camps-Valls, L. Gómez-Chova, J. Calpe-Maravilla, J. D. Martín-Guerrero, E. Soria-Olivas, L. Alonso-Chordá, and J. Moreno, "Robust support vector method for hyperspectral data classification and knowledge discovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1530–1542, 2004.

TABLE II

CLASS LABEL OF INDIAN PINE 2010 DATA

ID	Color	Class name	No. samples
1		Corn-high	3387
2		Corn-mid	1740
3		Corn-low	356
4		Soybean-high	1365
5		Soybean-mid	37865
6		Soybean-low	29210
7		Resideus	5795
8		Wheat	3387
9		Hay	50045
10		Grass/Pasture	5544
11		Cover crop 1	2746
12		Cover crop 2	2164
13		Woodlands	48559
14		Highway	4863
15		Local road	502
16		Houses/Buildings	546

[5] G. M. Foody, A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd, "Training set size requirements for the classification of a specific class," *Remote Sens. Env.*, vol. 104, no. 1, pp. 1–14, 2006.

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

[7] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, NJ: Wiley-Interscience, 2nd edition, 2000.

[8] G. Schohn and D. Cohn, "Less is more: Active learning with support vector machines," in *Proc. 17th Int. Conf. Mach. Learn.*, Stanford, CA, Jun. 20-Jul.2 2000, pp. 839–846.

[9] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, 2002.

[10] F. Olsson, "A literature survey of active machine learning in the context of natural language processing," Tech. Rep. SICS Tech. Rep. T2009:06, 2009, Swedish Inst. Comput. Sci., Kista, Sweden, 2009.

[11] B. Settles, "Active learning literature survey," Computer Sciences Technical Report 1648, University of Wisconsin-Madison, 2010.

[12] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Muñoz-Marí, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Proc.*, vol. 5, no. 3, pp. 606–617, 2011.

[13] J. Muñoz-Marí, D. Tuia, and G. Camps-Valls, "Semisupervised segmentation of remote sensing images with active queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3751–3763, 2012.

[14] E. Dura, Yan Zhang, Xuejun Liao, G.J. Dobeck, and L. Carin, "Active learning for detection of mine-like objects in side-scan sonar imagery," *IEEE J. Oceanic Eng.*, vol. 30, no. 2, pp. 360 – 371, 2005.

[15] E. Pasolli, F. Melgani, N. Alajlan, and Y. Bazi, "Active learning methods for biophysical parameter estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 4071–4084, 2012.

[16] J. de Gruijter, D. Brus, M. Bierkens, and M. Knotters, *Sampling for Natural Resource Monitoring*, Springer, 2006.

[17] W. G. Müller, *Collecting Spatial Data. Optimum Design of Experiments for Random Fields*, Springer - Physica Verlag, 2007.

[18] W. Müller, L. Pronzato, and H. Waldl, "Beyond space-filling: an illustrative case," Tech. Rep. 2011-53, Institut für Angewandte Statistik (IFAS), 2011.

[19] F. Achard, H.D. Eva, H.-J. Stibig, P. Mayaux, J. Gallego, T. Richards, and J.-P. Malingreau, "Determination of deforestation rates of the world's humid tropical forests," *Science*, vol. 5583, pp. 999–1002, 2002.

[20] F. Javier Gallego, "Stratified sampling of satellite images with a systematic grid of points," *ISPRS J. Photogramm. Rem. Sens.*, vol. 59, no. 6, pp. 369–376, 2005.

[21] H. Eva, S. Carboni, F. Achard, N. Stach, L. Durieux, J.-F. Faure, and D. Mollicone, "Monitoring forest areas from continental to territorial levels using a sample of medium spatial resolution satellite imagery," *ISPRS J. Photogramm. Rem. Sens.*, vol. 65, no. 2, pp. 191–197, 2010.

[22] G. J. Fitzgerald, S. M. Lesch, E. M. Barnes, and W. E. Luckett, "Directed sampling using remote sensing with a response surface sampling design for site-specific agriculture," *Comput. Electron. Agr.*, vol. 53, no. 2, pp. 98–112, 2006.

[23] D. Tuia, A. Pozdnoukhov, L. Foresti, and M. Kanevski, "Active learning for monitoring network optimization," in *Spatio-Temporal Design: Advances in Efficient Data Acquisition*, J. Mateu and W. G. Müller, Eds., Statistics in Practice. Wiley, 2013.

[24] M. Li and I. K. Sethi, "Confidence-based active learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1251–1261, 2006.

[25] P. Mitra, B. Uma Shankar, and S.K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recogn. Lett.*, vol. 25, no. 9, pp. 1067–1074, 2004.

[26] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, 2011.

[27] B. Demir, C. Persello, and L. Bruzzone, "Batch mode active learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1032, 2011.

[28] S. Patra and L. Bruzzone, "A fast cluster-assumption based active-learning technique for classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1617–1626, 2011.

[29] E. Pasolli, F. Melgani, and Y. Bazi, "Support vector machine active learning through significance space construction," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 431 – 435, 2011.

[30] D. Tuia, F. Ratle, F. Pacifici, M. Kanevski, and W.J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, 2009.

[31] W. Di and M. M. Crawford, "Active learning via multi-view and local proximity co-regularization for hyperspectral image classification," *IEEE J. Sel. Topics Signal Proc.*, vol. 5, no. 3, pp. 618–628, 2011.

[32] W. Di and M.M. Crawford, "View generation for multiview maximum disagreement based active learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1942 –1954, 2012.

[33] J. Li, J.M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947 –3960, 2011.

[34] S. Rajan, J. Ghosh, and M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, 2008.

[35] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Intl. Conf. Mach. Learn. (ICML)*, Washington D.C., USA, 2003, pp. 59–66.

[36] M. Ferecatu and N. Boujemaa, "Interactive remote sensing image retrieval using active relevance feedback," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 818–826, 2007.

[37] M. Volpi, D. Tuia, and M. Kanevski, "Memory-based cluster sampling for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3096–3106, 2012.

[38] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 20, no. 8, pp. 832–844, 1998.

[39] H. S. Seung, M. Opper, and H. Smopolinsky, "Query by committee," in *Proc. 5th Ann. ACM Workshop Comput. Learn. Theory*, Pittsburgh, PA, Jul. 27-29 1992, pp. 287–294.

[40] B. Leskes and L. Torenvliet, "The value of agreement: A new boosting algorithm," *J. Comput. Syst. Sci.*, vol. 74, no. 4, pp. 557–586, 2008.

[41] Z. Zhou and M. Li, "Semi-supervised learning by disagreement," *Knowl. Inf. Syst.*, vol. 24, no. 24, pp. 415–439, 2009.

[42] V. R. de Sa, "Learning classification with unlabeled data," in *Proc. Adv. Neural Inf. Process. Syst.*, 1994, vol. 6, pp. 112–119.

[43] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," in *Aide-Memoire Invited Lecture Conf. "Mathematical Challenges of the 21st Century"*, Los Angeles, CA, Aug. 2000.

[44] J.A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–490, 2005.

[45] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804 – 3814, 2008.

[46] S. Lee and M.M. Crawford, "Unsupervised multistage image classification using hierarchical clustering with a bayesian similarity measure," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 312–320, 2005.

[47] A. Marcal and L. Castro, "Hierarchical clustering of multispectral images using combined spectral and spatial criteria," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 1, pp. 59–63, 2005.

[48] J. Tilton, "Image segmentation by region growing and spectral clustering with a natural convergence criterion," in *Proc. IEEE Geoscience and Remote Sensing Symposium*, Seattle, WA, July 6-10 1998, vol. 4, pp. 1766–1768.

[49] A. J. Plaza and J. C. Tilton, "Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images," in *Proc. IEEE Geoscience and Remote Sensing Symposium*, Seoul, Korea, July 25-29 2005, vol. 7, pp. 4946–4949.

[50] A. Liu, G. Jun, and J. Ghosh, "Active learning with spatially sensitive labeling costs," in *NIPS Workshop on Cost-sensitive Learning*, Whistler, Canada, 2008.

[51] S. Lin and B. W. Kernighan, "An effective heuristic algorithm for the traveling-salesman problem," *Operations Res.*, vol. 21, pp. 498–516, 1973.

[52] E. L. Laweler, J. K. Lenstra, A. H. G. R. Kan, and D. B. Shmoys, *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, Wiley, 1985.

[53] B. Demir, L. Minello, and L. Bruzzone, "A cost-sensitive active learning technique for the definition of effective training sets for supervised classifiers," in *Proc. IEEE Geoscience and Remote Sensing Symposium*, Munich, Germany, July 22-27 2012, pp. 1781–1784.

[54] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, "Improving active learning methods using spatial information," in *Proc. IEEE Geoscience and Remote Sensing Symposium*, Vancouver, Canada, July 24-29 2011, pp. 3923–3926.

[55] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002.

[56] A. Stumpf, N. Lachiche, J.P. Malet, N. Kerle, and A. Puissant, "Active learning in the spatial-domain for landslide mapping in remote sensing images," in *European Conference on Machine Learning (ECML), Workshop "Active Learning in Real Applications"*, Edinbrough, Scotland, September 24 - 28 2012.

[57] R. Kashef and M.S. Kamel, "Enhanced bisecting k-means clustering using intermediate cooperation," *Pattern Recogn.*, vol. 42, no. 11, pp. 2257–2569, 2009.

[58] M. D. Fleming, J. S. Berkebile, and R. M. Hoffer, "Computer-aided analysis of LANDSAT-I MSS data: a comparison of three approaches, including a "modified clustering" approach," LARS information note 072475, Purdue University, 1975.

[59] L. Bruzzone and C. Persello, "A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3180–3191, 2009.

[60] G. Jun and J. Ghosh, "Spatially adaptive classification of land cover with remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2662–2673, 2011.

[61] D. Tuia, J. Muñoz-Marí, L. Gómez-Chova, and J. Malo, "Graph matching for adaptation in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, in press.

[62] L. Bruzzone and D. Fernandez-Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 456–460, 2001.

[63] L. Bruzzone and M. Marconcini, "Toward the automatic updating of land-cover maps by a domain-adaptation SVM classifier and a circular validation strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1108–1122, 2009.

[64] S. Rajan, J. Ghosh, and M. Crawford, "Exploiting class hierarchy for knowledge transfer in hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3408–3417, 2006.

[65] W. Kim and M.M. Crawford, "Adaptive classification for hyperspectral image data using manifold regularization kernel machines," *IEEE Trans. Geosci. and Remote Sens.*, vol. 48, no. 11, pp. 4110 –4121, 2010.

[66] G. Jun and J. Ghosh, "An efficient active learning algorithm with knowledge transfer for hyperspectral remote sensing data," in *Proc. IEEE Geoscience and Remote Sensing Symposium*, Boston, MA, July 6-11 2008, pp. I–52 I–55.

[67] C. Persello and L. Bruzzone, "Active learning for domain adaptation in the supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4468–4483, 2012.

[68] G. Matasci, D. Tuia, and M. Kanevski, "SVM-based boosting of active learning strategies for efficient domain adaptation," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 5, pp. 1335–1343, 2012.

[69] D. Tuia and J. Muñoz-Marí, "Learning user's confidence for active learning," *IEEE Trans. Geosci. Remote Sens.*, in press.

[70] W. Di and M. Crawford, "Critical class oriented active learning for hyperspectral image classification," in *Proc. IEEE Geoscience and Remote Sensing Symposium*, Vancouver, Canada, July 22-27 2011, pp. 3899–3902.

[71] B. Demir, F. Bovolo, and L. Bruzzone, "Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach," *IEEE Trans. Geosci. Remote Sens.*, in press.

[72] B. Demir, F. Bovolo, and L. Bruzzone, "Detection of land-cover transitions in multitemporal remote sensing images with active learning based compound classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1930–1941, 2012.

[73] M. Furlani, D. Tuia, J. Muñoz-Marí, F. Bovolo, G. Camps-Valls, and L. Bruzzone, "Discovering single classes in remote

sensing images with active learning," in *Proc. IEEE Geoscience and Remote Sensing Symposium*, Munich, Germany, July 22-27 2012, pp. 7341–7344.

[74] J. Michel, J. Inglada, and J. Malik, "Object based and geo-spatial image analysis: a semi-automatic pre-operational system," in *SPIE Remote Sensing Europe*, Toulouse, France, September 20-22 2010, vol. 7830.

[75] G. Mountrakis, J. Ima, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Rem. Sens.*, vol. 66, no. 3, pp. 247–259, 2011.

[76] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *5th ACM Workshop on Computational Learning Theory*, Pittsburgh, PA, July 27 - 29 1992, pp. 144–152.