Gustavo Camps-Valls, Devis Tuia, Lorenzo Bruzzone, and Jón Atli Benediktsson

Advances in Hyperspectral Image Classification

Bridging the Gap Between Remote Sensing and Signal Processing

© ISTOCKPHOTO.COM/MIPAN
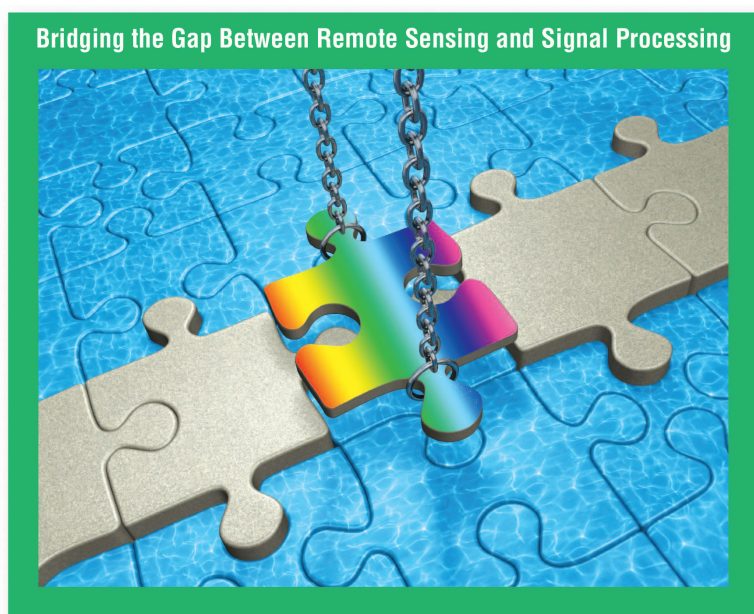
[Earth monitoring with statistical learning methods]

The technological evolution of optical sensors over the last few decades has provided remote sensing analysts with rich spatial, spectral, and temporal information. In particular, the increase in spectral resolution of hyperspectral images (HSIs) and infrared sounders opens the doors to new application domains and poses new methodological challenges in data analysis. HSIs allow the characterization of objects of interest (e.g., land-cover classes) with unprecedented accuracy, and keeps inventories up to date. Improvements in spectral resolution have called for advances in signal processing and exploitation algorithms. This article focuses on the challenging problem of hyperspectral image classification, which has recently gained in popularity and attracted the interest of other scientific disciplines such as machine learning, image processing, and computer vision. In the remote sensing community, the term classification is used to denote the process that assigns single pixels to a set of classes, while the term segmentation is used for methods aggregating pixels into objects and then assigned to a class.

One should question, however, what makes HSI so distinctive. Statistically, HSI are not extremely different from natural grayscale and color photographic images (see [1, Ch. 2]). Grayscale images are spatially smooth: the joint probability density function

(PDF) of the luminance samples is highly uniform, the covariance matrix is highly nondiagonal, the autocorrelation functions are broad, and have generally a $1/f$ band-limited spectrum. In the case of color images, the correlation between the tristimulus values of the natural colors is typically high. While the three tristimulus channels are equally smooth in generic red/green/blue (RGB) representations, opponent representations imply an uneven distribution of bandwidth between channels. Despite all these commonalities, the analysis of HSI turns out to be more difficult, especially because of the high dimensionality of the pixels, the particular noise and uncertainty sources observed, the high spectral redundancy, and the typically nonlinear relations observed between spectral channels as well as with the corresponding material. Such nonlinearities can be related to a plethora of factors, including the multiscattering in the acquisition process, the heterogeneities at subpixel level, as well as the impact of atmospheric and geometric distortions. These characteristics of the imaging process lead to distinct nonlinear feature relations, i.e., pixels lie in high-dimensional complex manifolds. The high spectral sampling of HSI (the bands usually cover narrow portions of the electromagnetic spectrum, typically 5–10 nm) also leads to strong collinearity issues. Finally, the spatial variability of the spectral signature increases the internal class variability. All of these factors, in conjunction to the few labeled examples typically available, make HSI image classification a very challenging problem. As a result, the accuracy obtained with standard parametric classifiers commonly used for multispectral image classification is typically compromised when applied to HSI [2].

Many of these limitations have been recently addressed under the framework of statistical learning theory (SLT) [3]. SLT is a general framework for learning functions from data, which reduces to finding a linear function defined in a high- (eventually infinite) dimensional Hilbert feature space $f \in \mathcal{H}$ that learns the relation between observed input-output data pairs $(x_1, y_1), \ldots, (x_\ell, y_\ell) \in \mathcal{X} \times \mathcal{Y}$, and that generalizes well. Generalization is the capability of a method to extrapolate to unseen situations, i.e., the function f should accurately predict the label $y^* \in \mathcal{Y}$ for a new input example $x^* \in \mathcal{X}$. Generalization has recurrently appeared in statistics literature for decades under the names of *bias-variance dilemma, capacity control*, or *complexity regularization tradeoff*. The underlying idea is to constrain too flexible functions to avoid overfitting the training data.

The SLT framework formalizes this intuition [3] and seeks for prediction functions that optimize a functional $\mathcal{L}_{reg}$ that takes into account both an empirical estimation of the training error (loss), $\mathcal{L}_{emp}$, and an estimate of the complexity of the model (or regularizer), $\Omega(f)$:

$$\mathcal{L}_{reg} = \mathcal{L}_{emp} + \lambda \, \Omega(f)$$
$$= \sum_{i=1}^{\ell} V(x_i, y_i, f(x_i)) + \lambda \, \Omega(f),$$

where V is a loss function acting on the ℓ labeled samples, and λ is a tradeoff parameter between the cost and the regularization. Different losses and regularizers can be adopted for solving the problem, involving completely different families of models

and solutions. To ensure unique solutions, many SLT algorithms use strictly convex loss functions. The regularizer $\Omega(f)$ limits the capacity of the classifier to minimize $\mathcal{L}_{emp}$ and favors smooth functions.

The hyperspectral image processing community has contributed to the design of specific loss functions and regularizers to take the most out of the acquired images. For example, regularization appears explicitly in many HSI classifiers when trying to impose the spatial homogeneity of images, when including the wealth of user's labeling in active learning (AL), or when exploiting the information contained in the unlabeled pixels to better describe the image manifold in semisupervised learning. Classifiers should also be robust to changes in the image representation: small perturbations of pixels and objects in the image manifold should not produce big differences in the classification. This is why the inclusion of proper image representations and invariances is also an active field.

ADVANCED REGULARIZED IMAGE CLASSIFICATION
Before HSI, most of the classifiers used in remote sensing were parametric, such as Gaussian maximum likelihood or linear discriminant analysis. These methods, based on the estimate of the covariance matrix, were successful when dealing with early multispectral images, whose dimensionality was usually comprised between four and ten bands. HSI changed the rules, as the increased dimensionality of pixels raised to hundreds. Standard parametric methods became either unfeasible or unreliable, since estimating the class-covariance matrices requires many labeled samples, which are usually not available. For that reason, research turned to include regularization, either explicitly through Tikhonov's terms in the involved covariance matrices, or by perform classification in a subspace of reduced dimensionality [4], [5].

Although successful, parametric models make strong assumptions about the normality of the class conditional PDFs or about the linearity of the problem. Due to the complexity of HSI, such assumptions rarely hold, which encouraged research towards nonparametric and nonlinear models. Several approaches have been introduced in the last decade in the field of hyperspectral image classification: kernel methods and support vector machines (SVMs) [2], sparse multinomial logistic regression [6], neural networks [7], and Bayesian approaches like relevance vector machines [8] and Gaussian processes classification [9]. Nevertheless, the SVM has undoubtedly become the most widely used method in HSI classification research [10]. Unlike other nonparametric approaches, such as regularized radial basis function (RBF) neural networks, SVM naturally implements regularization through the concept of maximum margin: given a linear classification function $f(x) = w^\top x + b$, maximizing the linear separability between classes is equivalent to minimize the ℓ_2-norm of model weights w used as regularizer, $\Omega(f) = \|w\|_2^2$. Nonlinearity is also implemented via reproducing kernels, which allows the ability to work in high-dimensional Hilbert spaces implicitly, while still resorting to linear algebra operations [2].

In spite of those desirable properties, the effectiveness of SVM rapidly turned out to be insufficient to exploit the rich information

contained in HSI. In all methodologies that follow, the functional to be optimized will consider additional information such as the one contained in unlabeled samples, ancillary data, or distinct signal characteristics. Such heterogeneous information is typically included in the HSI classifiers through additional regularizers.
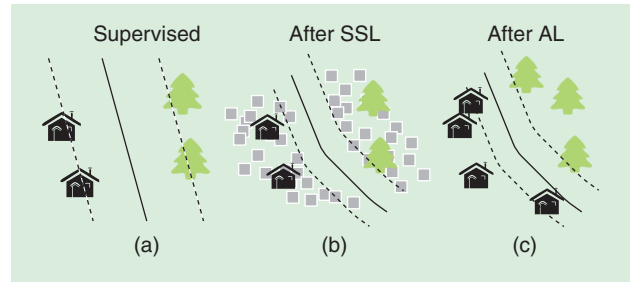
REGULARIZATION WITH UNLABELED SAMPLES

Recently, researchers started to exploit the abundant unlabeled information contained in the image itself, and new forms of regularization and priors were introduced. This is the field of semisupervised learning (SSL) [11]–[13] [see Figure 1(b)], where the minimization functional is modified to take into account the structure of the hyperspectral image manifold. Usually, semisupervised algorithms modify the decision function of the classifier by adding an extra regularization term Ω_u that acts on both labeled and unlabeled examples

$$\mathcal{L}_{reg} = \mathcal{L}_{emp} + \lambda \, \Omega(f) + \lambda_u \, \Omega_u(f).$$

Several strategies to design the regularizer have been presented. One may use the graph Laplacian as a metric on the predictions to build Ω_u. Since regularization is performed on a proximity graph, the assumption enforced is that decisions on neighboring pixels in the data manifold should be similar [14]. Another possibility considers regularizers that enforce wide and empty SVM margins [11]. Other strategies deform the kernel function by changing the metric induced using the unlabeled samples [13], [15]. Table 1 summarizes the different approaches in semisupervised learning.

We evaluate the performance of semisupervised algorithms in an Airborne Visible Infrared Imaging Spectrometer (AVIRIS) image acquired over the Kennedy Space Center (KSC), in Florida in 1996, with a total of 224 bands of 10-nm bandwidth with center wavelengths from 400 to 2,500 nm. The data was acquired from an altitude of 20 km and has a spatial resolution of 18 m. After removing low signal-to-noise ratio (SNR) bands and water absorption bands, a total of 176 bands remains for the analysis. We merged some subclasses to deal with a more balanced problem of ten classes. The high dimensionality and number of classes and subclasses pose challenging problems for the classifiers, especially when few labeled examples are available.

Figure 2 illustrates classification results for cluster kernels, probabilistic mean map kernel, label propagation, Laplacian SVM (LapSVM), and semisupervised neural networks (SSNNs).

[FIG1] Regularization of models with unlabeled samples: (a) purely supervised solution, (b) semisupervised solution exploiting low-density areas, and (c) active learning solution, where three new samples are labeled by a user.

We used $\ell = 200$ labeled pixels (20 per class) and $u = 1,000$ unlabeled pixels. LapSVM, cluster kernels, and mean map kernels perform similarly, and all improve the results of the label propagation whose training was particularly difficult in this high-dimensional setting. More homogeneous areas and better classification maps are observed in general for the mean map and bag kernels, and particularly for the SSNN, which efficiently deals with complex marsh areas [bottom right part of the classification map in Figure 2(g)] and cope with large scale data sets.

REGULARIZATION VIA USER'S INTERACTION

Another possibility to cope with small sample problems is to provide additional labeled examples. This is possible since HSI represent land surfaces, usually physically reachable or that can be displayed in an image processing software. Therefore, the new samples can be collected either by photointepretation of the images (only if the classes can be recognized on screen) or by organizing field campaigns. However, since providing additional samples is costly, the samples to be labeled must be selected carefully. To this end, AL [16], [17] [see Figure 1(c)] has gained popularity in recent years: rather than proceeding by random sampling or stratification (i.e., sampling according to a measure of the expected variability within a class), AL uses the outcome of the current model to rank the unlabeled pixels according to their expected importance for future labeling. The aim is to detect the most difficult (and diverse) pixels for the current classifier. The top-ranked pixels are then screened by a human operator, who provides the labels, enlarging the training set. With the enlarged training set, a new improved classifier is built and the process is

ASSUMPTION	MODEL	IDEA
[TABLE 1] SUMMARY OF SEMISUPERVISED ALGORITHMS USED IN HSI CLASSIFICATION.		
LOW DENSITY	TSVM [11]	LOOK FOR THE EMPTIEST MARGIN.
MANIFOLD	LABEL PROPAGATION [12]	SPREAD CLASS INFORMATION ON THE GRAPH OF LABELED/UNLABELED (NEARBY SAMPLES ARE CLASSIFIED IN THE SAME CLASS).
	LapSVM [14]	SVM HINGE LOSS PLUS LAPLACIAN EIGENMAPS FOR MANIFOLD REGULARIZATION: PIXELS CLOSE IN THE INPUT SPACE ARE ALSO CLOSE IN THE GRAPH (NEARBY SAMPLES ARE MAPPED CLOSE TOGETHER).
	SSNN [7]	NEURAL NETWORK TRAINED WITH GRADIENT DESCENT REPLACES SVM, GRAPH REGULARIZATION WITH LOSS THAT FORCES SIMILAR PIXELS TO BE MAPPED CLOSELY AND DISSIMILAR ONES TO BE SEPARATED.
CLUSTER	CLUSTER KERNEL [13]	INCREASE THE SIMILARITY MEASURE (KERNEL) IF SAMPLES FALL IN THE SAME CLUSTER, THEN RUN STANDARD SVM.
	MEAN MAP KERNEL [15]	INCREASE SIMILARITY IF SAMPLES ARE MAPPED CLOSE TO CENTROIDS IN HILBERT SPACE, THEN RUN STANDARD SVM.

[TABLE 2] SUMMARY OF AL ALGORITHMS [16].

| CRITERION	CLASSIFIER	UNCERTAINTY	DIVERSITY	MODELS TO TRAIN
EQB	ALL	AGREEMENT OF A COMMITTEE	×	p MODELS
AMD	ALL	AGREEMENT OF A COMMITTEE	×	p MODELS
MS	SVM	DISTANCE TO SVM MARGIN	×	ONE SVM
cSV	SVM	DISTANCE TO SVM MARGIN	SPECTRAL DISTANCE TO CURRENT SVs	ONE SVM + DISTANCES TO SVs
MOA	SVM	DISTANCE TO SVM MARGIN	ANGULAR DIFFERENCES	ONE SVM + DISTANCES TO ALREADY SELECTED SAMPLES
MCLU-ECBD	SVM	DISTANCE TO SVM MARGIN	DIFFERENT CLUSTER ASSIGNMENT	ONE SVM + NONLINEAR CLUSTERING OF c SAMPLES
KL-Max	PROB. OUTPUT	DIVERGENCE OF PDF IF ADDING THE CANDIDATE	×	$(c-1)$ MODELS
BT	PROB. OUTPUT	DIFFERENCE IN POSTERIOR OF MOST CONFIDENT CLASSES	×	ONE MODEL

c : Number of candidates.
p : Members of a committee of learners.

iterated. Since AL focuses on difficult areas, it boosts the performances with fewer samples than those required by random sampling. Table 2 includes a taxonomical classification of the field of active learning. The section "Adapting the Classifier" will illustrate an example of active learning model regularization for the specific task of adapting classifiers to multiple scenes.

REGULARIZATION THROUGH SPARSITY PROMOTION

Despite their high dimensionality, hyperspectral pixels belonging to the same class typically lie in a low-dimensional subspace. This observation has been recently used in sparse signal representations. Here, the assumption is that pixels can be represented accurately as a linear combination of a few training samples from a structured dictionary.

Note the connection with SVMs that embed the dictionary (the training samples) into a high-dimensional feature space $\mathcal{H}$. The use of the hinge loss in the SVM functional induces a sparse solution, i.e., few training examples are selected. Recently, sparse kernel methods have been presented, such as the kernel matching pursuit, the ℓ_1-SVM, the kernel basis pursuit, or the generalized LASSO. In all of these, the dictionary functions are the kernels centered around the selected "support vectors." Alternatively, in [18], several sparse kernel approaches have been presented with a different philosophy: the target pixel is the test pixel itself, not a similarity evaluation, and the dictionary is composed by the training pixels in the feature space. In this article, a basis projection (BP) approach is used to promote sparsity with $\Omega = \| w \|_1$, as a relaxation of the more computationally demanding problem
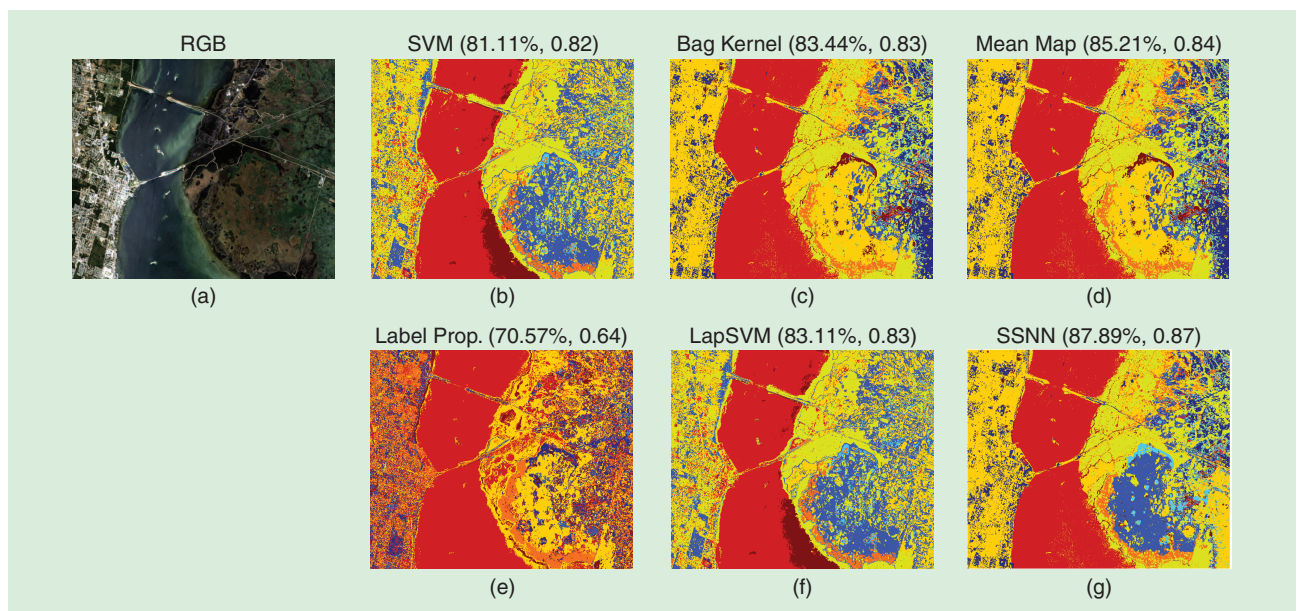
[FIG2] (a) RGB composition and classification maps with (b) SVM, (c) cluster (or bag) kernels, (d) probabilistic mean-map kernel, (e) label propagation, (f) LapSVM, and (g) SSNN for the KSC image ($\ell = 200$, $u = 1,000$). Overall accuracy and kappa statistic are given in brackets.

induced by using the ℓ_0-norm. To solve the BP problem, greedy algorithms, such as the orthogonal matching pursuit (OMP) and the subspace pursuit (SP) methods, can be used. The dictionary can be obtained offline or from the same image. The classification can be additionally improved by incorporating the contextual information from the neighboring pixels into the classifier (see the next section).

In the example illustrated in Figure 3, we compare the performance of the linear (SP, OMP) and kernel (KSP, KOMP) sparse HSI classifiers introduced in [18]. The sparseness factor was tuned for best performance. We also included a linear SVM and the ν-SVM using an RBF kernel, in which the parameter $\nu \in (0, 1)$ controls the degree of sparsity. The RBF σ parameter was tuned by standard tenfold cross-validation. Figure 3 shows the results for these methods and different number of training samples in the standard AVIRIS Indian Pines hyperspectral image (220 spectral channels and spatial resolution 20 m, shown in Figure 4). This is the standard benchmark hyperspectral image, which is used here to allow comparison with results in [18]. We split the data into a training set (20% of the available labeled pixels) and a test set (80%). We trained the classifiers for different rates {1, 5, 10, 15, 20, 25, 30}% of the training set, and show results for the test set that remained constant. Nonlinear methods show a much better performance over linear approaches. In the linear case, SP clearly outperforms the rest, but when the nonlinearity is included all methods perform very similarly.

SPATIAL-SPECTRAL IMAGE CLASSIFICATION

HSI live in a geographical manifold, in the sense that spatially neighboring pixels carry correlated information and that images are usually smooth in the spatial domain [1, Ch. 2]. Accounting for spatial smoothness 1) provides less salt-and-pepper classification maps, 2) reveals the size and shape of the structure the pixel belongs to, and 3) allows the discrimination between structures made of the same materials, but belonging to different land-use types. Spatial regularization has been widely used to improve classification [21]. The joint exploitation of both spectral and spatial information considers that either the loss, the regularizer, or both depend on the spatial neighborhood of a pixel [22]. In the next subsections we review the field of spatial-spectral classification, whose major developments are summarized in Table 3.

SPATIAL FEATURE EXTRACTION

A simple yet effective way to regularize for spatial smoothness is to enrich the input space with features accounting for the neighborhood of the pixels. This is usually done by using moving windows or adaptive filters applied to the spectral bands. These filtered images are then used to learn the classifier. Standard filters based on occurrence or co-occurrence, morphological operators, Gabor filters, or wavelets decompositions generally provide significant improvements over purely spectral classifiers. Among them, morphological filters are the most promising. In [23], filtering was performed at many scales and an extended morphological profile (EMP) was used for classification. Proceeding in a multiscale fashion enables the adaptive

definition of the neighborhood of a pixel according to the structure it belongs to. Filtering in HSI is more challenging than in multispectral images, and one typically resorts to compute the EMP based on only a few principal components (PCs) using morphological reconstruction operators. All of the features are then fed to a classifier, either alone [23] or combined with the original spectral information [24]. Furthermore, feature selection [25] or extraction [23] can be used to find the relevant features. Recently, connected tree-based morphological operators have been investigated for the analysis of HSI [26]. These so-called attribute filters extract thematic attributes of the connected components of an image which are thresholded according to their geometry (area, length, shape factors), or texture (range, entropy). The multiscale version, extended morphological attribute profile (EMAP), has been also introduced.

In Figure 5, the approaches based on mathematical morphology (EMP and EMAP extracted from the first four PCs) are used for classification of ROSIS-03 data from an urban area in Pavia, Italy. We selected this image to illustrate the capabilities of several spatial-spectral classifiers since urban areas monitoring at very high resolution typically requires the extraction of directional, rotational, and scale features from objects. A significant improvement in terms of classification accuracies with respect to the spectral SVM was achieved by applying the EMAP with four different attributes on the image (area and diagonal of the bounding box of connected components, moment of inertia, and standard deviation; see [26]). On the other hand, some redundancy was observed in the original 144-dimensional filter vector of EMAP. Therefore, decision boundary feature extraction (DBFE) was applied on it. After extraction with DBFE, the accuracies improved significantly (+5%), thus confirming the importance of feature extraction routines. Finally, comparison with composite kernels (see the section "Advanced Spatial-Spectral Classifiers" on the next page and in [19] and [20]) yielded improved classification accuracy, but with a strong change of response in the right part of the image, where much more soil is predicted.
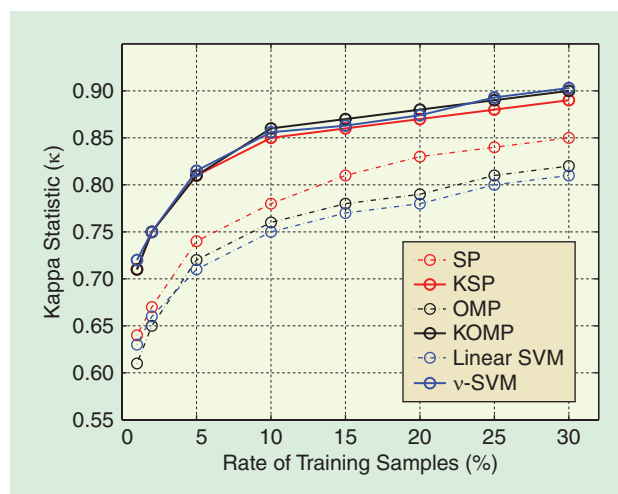
[FIG3] The performance measure with the estimated Cohen's kappa statistic, κ, for different sparsity-promoting classifiers.

RGB

SVM (78.2%, 0.75)

(a)

(b)

Majority Voting (90.8%, 0.90)

Markers (91.8%, 0.91)

(c)

(d)
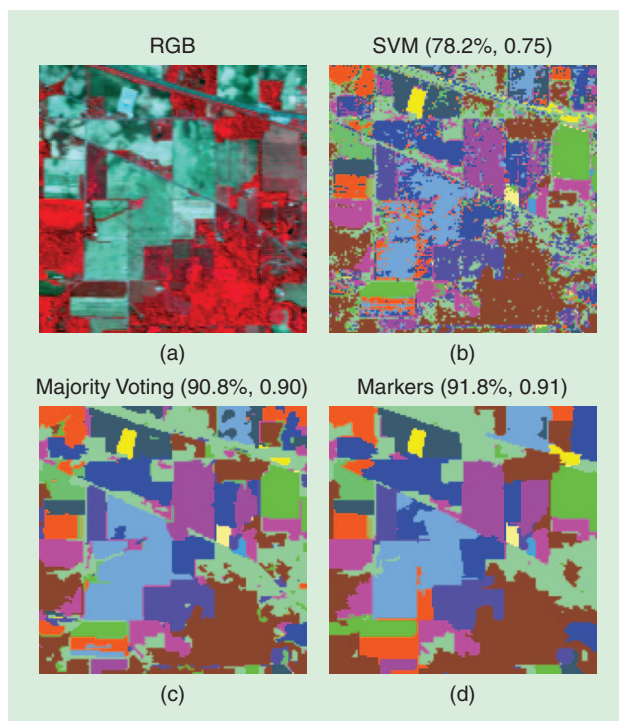
[FIG4] RGB composition of the standard AVIRIS Indian Pine data set (200 spectral channels and spatial resolution 20 m). Classification maps are shown for spatiospectral classification with the segmentation and classification with majority voting and segmentation with markers against SVM on the original image only. Overall accuracies and the kappa statistic for each method are reported in parentheses [21].

SPATIAL-SPECTRAL SEGMENTATION

Another approach for the inclusion of spatial information is through image segmentation, typically using watershed, mean shift, and hierarchical segmentation [21]. After segmentation, a supervised scheme assigns the pixels in the segments to the classes. Two approaches are the mostly used: in the first, the regions are treated as input vectors in a supervised classifier. In the second, regions are considered as basins to postprocess the class memberships attributed by a pixel-based classifier within each segment.

The reverse view on the problem is proposed in [27], where a supervised classifier is used to produce confidence values for each pixel. Then, pixels with maximal confidence are used as seeds for a region growing algorithm. In [28] segmentation and classification are linked by user-provided labels: working with a hierarchical segmentation of the data, the labels provided are used to isolate coherent clusters, both spatially and thematically, thus ending with the good segmentation and the labels of the segments. The number of queries is minimized with AL.

In Figure 4, two approaches based on segmentation and classification with majority voting and markers are applied to a 220-bands AVIRIS data set over Indian Pines (Indiana, United States). Significant improvement in terms of overall classification accuracies and kappa statistic were achieved over the pixel-based SVM classifier. Using markers provided the best accuracies: a +1% improvement over the simple majority voting and more than +13% over the traditional pixel-based SVM classifier.

Furthermore, as seen in Figure 4, it is clear that the classification and segmentation approaches provide a significantly more uniform classification map when compared to the purely spectral SVM classification map.

ADVANCED SPATIAL-SPECTRAL CLASSIFIERS

The main problem with spatial-spectral feature extraction approaches is the possibly high dimensionality of the feature vectors used to feed the classifier. This was alleviated in [19] where dedicated kernels for the spectral and spatial information were combined. The framework has been recently extended to deal with convex combinations of kernels through multiple-kernel learning [25] and generalized composite kernels [20]. In both cases, however, the methodology still relies on performing an ad hoc spatial feature extraction before kernel computation. Other alternatives in the literature considered the definition of graph kernels that capture multiscale higher-order relations in a neighborhood without computing them explicitly [29], and the modification of the SVM to seek for the spatial filter that maximizes the margin [30].
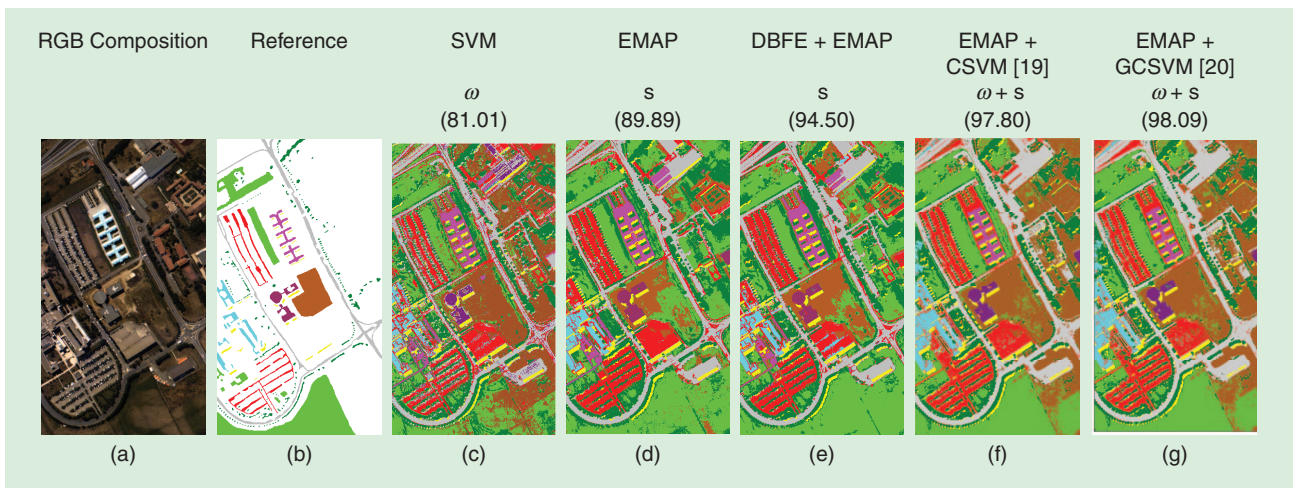
A final alternative is to include contextual information with Markov random fields (MRFs), which naturally include a spatial term on class smoothness in the energy function. However, in the high-dimensional context of HSI, the standard application of the neighbor system definition makes the problem computationally intractable, and therefore recent works have focused on joining MRF spatial priors and discriminative models in HSI classification [31], [32]. An excellent review of MRF spatial-spectral methods can be found in [33].

ADAPTATION AND INVARIANCES

One of the greatest challenges of modern HSI classification is the adaptation of classifiers between acquisitions that differ either by the zone they represent and/or the acquisition conditions such as illumination, angle, and season, among other effects. Adaptation is a central issue in HSI classification. For example, the increase in revisit time of recent satellites has improved multitemporal analysis of scenes. Nevertheless, algorithms must be able to adapt to changing situations. Generally, the direct application of classifiers trained on one image to new images leads to poor results: even if the objects represented in the images are roughly the same, differences in acquisition induce significant local changes in the PDF. These changes must be modeled and introduced in the classifiers. The concept of adaptation can be implemented at the levels of image preprocessing, robust and invariant feature extraction, or in the design of the classification algorithm.

PREPROCESSING

The preprocessing phase can address adaptation through the use of radiometric correction techniques applied to the images. Absolute corrections aim at transforming the radiance measured at the sensor into surface reflectance. Relative calibration techniques adapt the radiometric properties between portions of an image or between images. Generally, absolute correction techniques require additional ground reference data that in many

RGB Composition	Reference	SVM ω (81.01)	EMAP s (89.89)	DBFE + EMAP s (94.50)	EMAP + CSVM [19] ω + s (97.80)	EMAP + GCSVM [20] ω + s (98.09)
(a)	(b)	(c)	(d)	(e)	(f)	(g)

[FIG5] (a) RGB composition along with (b) the available reference data for the ROSIS-03 Pavia University area data set (103 spectral channels and spatial resolution 1.3 m). Classification maps are shown for (c) SVM on the original image only against spatiospectral classification with the (d) EMAP and (e) EMAP after feature extraction by DBFE (81 data channels), (f) composite kernels with cross-kernels and SVM [19], and (g) generalized composite kernels with multinomial logistic regression [20]. Overall accuracies [%] are reported in parentheses (ω: using spectral bands, s: using spatial filters from PCA).

cases are not available or are difficult to collect. Relative calibration methods are often considered as a pragmatic alternative for adaptation. Among these methods, we recall histogram matching, relative radiometric normalization of time series [34], and multivariate histogram matching [35].

Images can be casted as point clouds in a geometrical space endorsed with an appropriate distance measure. Such a view is quite convenient because it allows us to move from image adaptation to manifold adaptation. Recent methods have explicitly considered the distortions occurring between image manifolds. In [36], multitemporal sequences for each pixel were aligned based on a measure of similarity between sequences barycenters, thus consisting into a global measure of alignment. In [37], spectra of the pixels are spatially detrended using Gaussian processes to avoid shifts related to geometrical differences or to localized class variability. A recent principled approach tries to match graphs representing the data manifolds [38]. There, the graphs of the two domains are matched using a procedure aiming at maximizing their similarity, while at the same time preserving the original structure of the graphs.

ADAPTING THE CLASSIFIER

Learning a transformation between domains may be insufficient to handle all the perturbing factors, so alternative approaches are concerned with the adaptation of the classifier itself. From a machine-learning perspective, the problem of classifier adaptation is studied in the framework of transfer learning, and in particular, of domain adaptation. Domain adaptation reduces to learning from data in a source domain $\mathcal{S}$ (e.g., a portion of an image) to extrapolate to a different target domain $\mathcal{T}$ (another portion of the image or to another image). The problem has been given attention in HSI classification lately [39]. In this setting, source and target domains are assumed to share the same set of information classes (exceptions to this constraint in [40] and

[TABLE 3] SUMMARY OF SPATIAL-SPECTRAL ALGORITHMS.		
TYPE OF APPROACH	**MODEL**	**IDEA**
SPATIAL FILTERS EXTRACTION	CO-OCCURRENCE	EXTRACT TEXTURE BASED ON STATISTICS OF PAIRS OF PIXELS IN A NEIGHBORHOOD.
	EMP	MULTISCALE MATHEMATICAL MORPHOLOGY (BASED ON SIZE).
	EMAP	MULTISCALE MATHEMATICAL MORPHOLOGY (VARIETY OF ATTRIBUTE TYPES).
SPATIAL-SPECTRAL SEGMENTATION	SEGMENTATION AND CLASSIFICATION BASED ON MAJORITY VOTING	ALL PIXELS ARE ASSIGNED TO THE MOST FREQUENT CLASS INSIDE A SEGMENTED REGION.
	SEGMENTATION AND CLASSIFICATION BASED ON MARKERS	MOST RELIABLY CLASSIFIED PIXELS ARE SELECTED AS "REGION MARKERS" FOR SEGMENTATION.
	SEMI-SUPERVISED HIERARCHICAL CLUSTERING TREE	RETURNS BOTH CLASSIFICATION AND CONFIDENCE MAPS. ACTIVE LEARNING USED TO SELECT INFORMATIVE SAMPLES.
ADVANCED SPATIAL-SPECTRAL CLASSIFICATION	COMPOSITE AND MULTIPLE KERNELS	BALANCES BETWEEN SPATIAL AND SPECTRAL INFORMATION WITH DEDICATED KERNELS.
	GRAPH KERNELS	TAKES INTO ACCOUNT HIGHER ORDER RELATIONS IN EACH PIXEL NEIGHBORHOOD.
	MRF	MARKOV RANDOM FIELD MODELING (PROBABILISTIC).
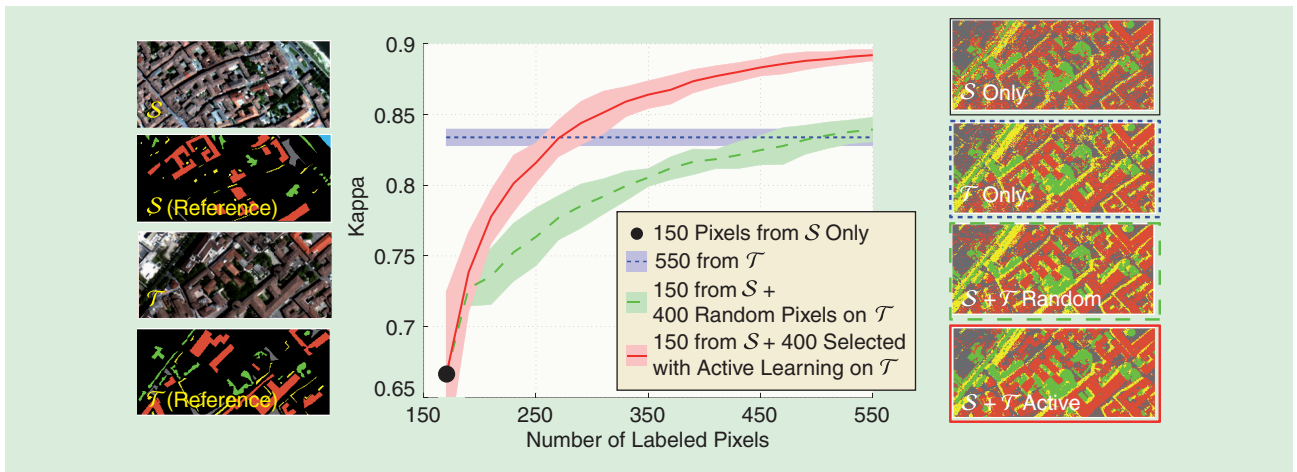
[FIG6] The use of AL to adapt a maximum likelihood classifier [40]. (a) Given an image and a set of reference pixels in a first source area ($\mathcal{S}$), we want to classify another spatially disconnected target area ($\mathcal{T}$), by adding labels chosen actively in the reference of $\mathcal{T}$. (b) Learning curves for the active (solid red line) and random sampling (dashed green line), evolving between the extremes of a model without adaptation (black dot at 150 samples), and another actively adapted that uses 550 samples randomly selected from $\mathcal{T}$ (blue dashed line). (c) Classification maps in the target domain $\mathcal{T}$. The color of the bounding box in each map refers to the legend in the central plot.

[41]) and to follow similar (but not the same) class distributions. Domain adaptation problems in remote sensing have been mainly addressed with semisupervised techniques, which exploit the labeled samples from $\mathcal{S}$ and the unlabeled samples from $\mathcal{T}$ to derive a classification rule suitable for the target domain. The most recent developments in this sense consider semisupervised and domain adaptation SVMs [39], Gaussian processes [37], [41], and the mean-map kernel methods [15].

Recently, AL has been also used for adaptation assuming that some samples (as few as possible) from the target domain can be labeled by the user and added to the existing training set (defined on $\mathcal{S}$) to adapt the classifier to the target domain $\mathcal{S}$ [40], [42]. This makes the adaptation process more robust than in the case of semisupervised learning at the cost of requiring additional labeled samples. Figure 6 illustrates this principle for a 102-bands image of the city center of Pavia acquired by the ROSIS-03 airborne sensor. The whole image (only a portion is shown) has a size of $1,400 \times 512$ pixels and spatial resolution is 1.3 m. Five classes of interest (buildings, roads, water, vegetation, and shadows) are considered, and a total of $206,009$ labeled pixels are available. We explore the potential of migrating a classifier built on a source area $\mathcal{S}$ with as few labeled pixels as possible from the rest of the image $\mathcal{T}$. First, the model trained with 150 labeled pixels randomly drawn from $\mathcal{S}$ was directly applied to $\mathcal{T}$ and yielded a $\kappa_{\mathcal{S}}$ of 0.67. Noticeably, a classifier built on 550 samples randomly selected on $\mathcal{T}$ reached $\kappa_{\mathcal{S}} = 0.84$. This suggests that different regions of the same scene follow very different statistics for classification. To improve the first classifier, we enlarged the first training set with labeled samples drawn from $\mathcal{T}$, either taken through random sampling (RS, green dashed line) or with AL (AL, red solid line). Using AL allows for concentrating efforts in areas where the first model is suboptimal, so performance is improved with respect to RS. After 400 queries (thus, a model using 550 training samples in total), random sampling yields similar performance to a model

using 550 randomly drawn pixels ($\kappa_{RS} = 0.84$), while AL improves the results with $\kappa_{AL} = 0.89$. To reach the performance of the model using 550 random pixels, AL requires only 120 active queries (thus, a total of 270 samples in the model).

ENCODING INVARIANCES IN THE CLASSIFIER

Image classifiers must be robust to changes in the data representation within each land cover class. The property of such mathematical functions is called *invariance*. A classifier should be invariant to object rotations, changes in illumination, the presence of shadows, and the spatial scale of the objects to be detected. Extracting robust features (invariants) for classification and domain adaptation has been traditionally pursued by looking at the spatial or the spectral signal characteristics. On the one hand, scale invariants aim to make classifiers invariant to perturbations of object scales. In HSI classification, a single spatial scale is typically suboptimal because different classes exhibit diverse sizes, shapes, and internal variations. Multiscale classification schemes may alleviate these problems. Also, translation invariants have been recently explored by means of scale invariant feature transform descriptors and wavelet-based representations. On the other hand, spectral invariants are considered the fundamental descriptors of object structure and are commonly employed to characterize canopy structure. Spectral invariance to daylight illumination allows, e.g., the improvement of multitemporal image classification.

Incorporating invariances in SVM can be achieved by designing particular kernel functions that encode local invariance under transformations, or to generate artificial examples for training to which the model must be invariant. In the following example, we consider the latter possibility, with the virtual SVM (VSVM) method, which has been successfully exploited to encode scale, rotation, translation, and shadow invariance in HSI classification [43].

In Figure 7, we illustrate the use of the VSVM encoding (spectral) shadow-invariance for image classification. We use the same data acquired by the ROSIS-03 optical sensor of the city center of Pavia (Italy) used before. We perform patch-based classification using only 50 training patches of size $w = 5$. The classes to be detected are, again, buildings, roads, water, vegetation, and shadows. Virtual support vectors (VSVs) were generated according to the observed exponential behavior of the ratio shadow/sunlit as a function of the wavelength [43]. Numerical results, as well as the zoom on a detail of the classification map, show that VSVM leads to more accurate results than the standard SVM: encoding shadow invariance reduces misclassifications on the bridge area and an overall more homogeneous classification over flat areas [see, e.g., the crossroads in Figure 7(b) and (c)].

## CONCLUSIONS AND DISCUSSION

This article reviewed and analyzed the recent developments in hyperspectral image classification. Even though HSI follow similar spatial, spectral, and spatial-spectral image statistics to those conveyed by conventional photographic images, the hyperspectral signals impose additional challenges related to their high dimensionality and heterogeneity. Therefore, even though standard techniques in image processing and computer vision may be transported directly, HSI impose important constraints to develop efficient and effective classifiers.

The use of methods derived from SLT has been a driving factor in recent years. SLT constitutes a proper framework to tackle the problems posed by hyperspectral remote sensing images, which typically involve scenarios with high-dimensional data and few training samples. SLT permits the embedding of numerical regularization in nonlinear classifiers, and also the design of alternative forms of conditioning and incorporation of prior knowledge. Additionally, classification is often improved by including spatially based and manifold-based regularizers. SLT also allows the ability to design sparse methods that are able to work in relevant feature subspaces, where compact and computationally efficient methods can be run. Finally, the SLT framework has allowed the inclusion of prior knowledge in a very fruitful way, e.g., classifiers can now incorporate spatial and spectral invariances that disentangle ambiguities present in land-cover classification.

The field is fast moving, and it is attracting researchers from the computer vision and machine-learning communities. New approaches are introduced regularly, which tackle new scenarios issued from high-resolution imaging (e.g., multitemporal, multiangular), while learning the relevant features via robust classifiers. It should be also noted that, with upcoming satellites, efficient algorithms for dimensionality reduction before classification and fast/parallel computing solutions will be necessary to accelerate the interpretation and efficient exploitation of HSI.

ACKNOWLEDGMENTS
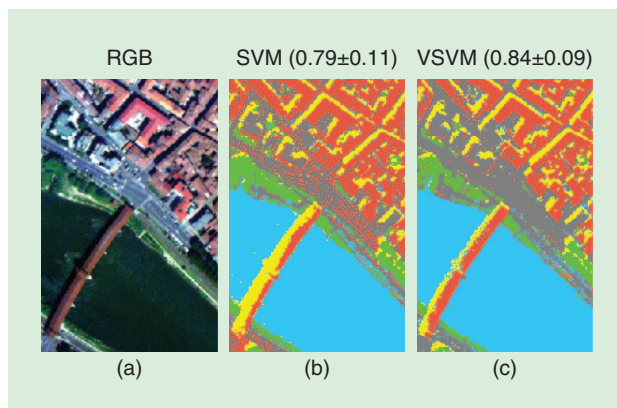
[FIG7] An experiment of patch-based classification with the virtual SVM encoding shadow invariance. (a) True-color composite and classification maps using (b) the standard SVM and (c) the VSVM obtained using 50 training pixels. Results are shown in parentheses in the form of (mean $\pm$ standard deviation of κ in 20 realizations).

RGB   SVM (0.79±0.11)   VSVM (0.84±0.09)

(a)   (b)   (c)

AUTHORS

Gustavo Camps-Valls (gustavo.camps@uv.es) received a Ph.D. degree in physics (2002) from the Universitat de València, Spain, where he is currently an associate professor in the Electrical Engineering Department and head of the Image and Signal Processing Group. His research is in machine learning for signal and image processing with a special focus on remote sensing data analysis. He has coauthored many journal papers and edited several books on kernel methods. The Institute for Scientific Information lists him as a highly cited researcher. He is a member of the IEEE Signal Processing Society's Machine Learning for Signal Processing Technical Committee and is an associate editor of *IEEE Transactions on Signal Processing, IEEE Signal Processing Letters*, and *IEEE Geoscience and Remote Sensing Letters*. He is a Senior Member of the IEEE.

Devis Tuia (devis.tuia@epfl.ch) received the Ph.D. degree in environmental sciences at the University of Lausanne (Switzerland) in 2009. He was a postdoctoral researcher at both the University of València, Spain, and the University of Colorado at Boulder. He is now a senior research associate at the Laboratory of Geographic Information Systems, at the Ecole Polytechnique Fédérale de Lausanne. His research interests include the development of algorithms for information extraction and classification of very high-resolution remote sensing images using machine-learning algorithms.

Lorenzo Bruzzone (lorenzo.bruzzone@ing.unitn.it) received the Ph.D. degree in telecommunications engineering from the University of Genoa, Italy, in 1998. He is currently a full professor at the University of Trento, Italy, where he is the founder and head of the Remote Sensing Laboratory. His current research interests are in the areas of remote sensing, radar and synthetic aperture radar, signal processing, and pattern recognition. He is the author (or coauthor) of many highly cited scientific publications in referred international journals, conference proceedings, and book chapters. He is the cofounder of the IEEE International Workshop on the Analysis of Multitemporal Remote-Sensing Images series.

Since 2003, he has been the chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. Since 2013, he has been the founding editor-in-chief of *IEEE Geoscience and Remote Sensing Magazine*. He is a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) and a Distinguished Speaker of the IEEE GRSS. He is Fellow of the IEEE.

Jón Atli Benediktsson (benedikt@hi.is) received the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, Indiana, in 1990. He is currently the prorector for academic affairs and a professor of electrical and computer engineering at the University of Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing. He was a cofounder of Oxymap, Inc. He was the 2011–2012 president of the IEEE Geoscience and Remote Sensing Society and was editor-in-chief of *IEEE Transactions on Geoscience and Remote Sensing* (2003–2008). He is a Fellow of the IEEE and SPIE.

REFERENCES

[1] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jimenez, and J. Malo, *Remote Sensing Image Processing, Synthesis Lectures on Image, Video, and Multimedia Processing*. San Rafael, CA: Morgan and Claypool, 2011.

[2] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.

[3] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[4] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 47, no. 3, pp. 862–873, 2009.

[5] A. C. Jensen, M. Loog, and A. H. Schistad Solberg, "Using multiscale spectra in regularizing covariance matrices for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 4, pp. 1851–1859, 2010.

[6] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 3, pp. 809–823, 2012.

[7] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 5, pp. 2271–2282, 2010.

[8] F. A. Mianji and Y. Zhang, "Robust hyperspectral classification using relevance vector machine," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 6, pp. 2100–2112, 2011.

[9] Y. Bazi and F. Melgani, "Gaussian process approach to remote sensing image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 1, pp. 186–197, 2010.

[10] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*. Hoboken, NJ: Wiley, 2009.

[11] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for semisupervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 11, pp. 3363–3373, 2006.

[12] G. Camps-Valls, T. Bandos, and D. Zhou, "Semisupervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 45, no. 10, pp. 3044–3054, 2007.

[13] D. Tuia and G. Camps-Valls, "Semi-supervised remote sensing image classification with cluster kernels," *IEEE Geosci. Remote Sensing Lett.*, vol. 6, no. 1, pp. 224–228, 2009.

[14] L. Gómez-Chova, G. Camps-Valls, J. Muñoz-Marí, and J. Calpe, "Semi-supervised image classification with Laplacian support vector machines," *IEEE Geosci. Remote Sensing Lett.*, vol. 5, no. 4, pp. 336–340, 2008.

[15] L. Gómez-Chova, G. Camps-Valls, L. Lorenzo Bruzzone, and J. Calpe-Maravilla, "Mean map kernels for semi-supervised cloud classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 1, pp. 207–220, 2010.

[16] 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. Select. Topics Signal Processing*, vol. 5, no. 3, pp. 606–617, 2011.

[17] M. M. Crawford, D. Tuia, and L. H. Hyang, "Active learning: Any value for classification of remotely sensed data?" *Proc. IEEE*, vol. 101, no. 3, pp. 593–608, 2013.

[18] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sensing*, vol. 51, no. 1, pp. 217–231, 2013.

[19] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sensing Lett.*, vol. 3, no. 1, pp. 93–97, 2006.

[20] J. Li, P. R. Marpu, A. Plaza, J. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 51, no. 9, pp. 4816–4829, 2013.

[21] 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, 2013.

[22] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing Environ.*, vol. 113, Suppl. 1, pp. S110–S122, Sept. 2009.

[23] 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 Sensing*, vol. 43, no. 3, pp. 480–490, 2005.

[24] 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 Sensing*, vol. 46, no. 11, pp. 3804–3814, 2008.

[25] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple kernel classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 10, pp. 3780–3791, 2010.

[26] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 10, pp. 3747–3762, 2010.

[27] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst. Man, Cybern. B*, vol. 40, no. 5, pp. 1267–1279, 2010.

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

[29] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio-spectral remote sensing image classification with graph kernels," *IEEE Geosci. Remote Sensing Lett.*, vol. 7, no. 4, pp. 741–745, 2010.

[30] R. Flamary, D. Tuia, B. Labbe, G. Camps-Valls, and A. Rakotomamonjy, "Large margin filtering," *IEEE Trans. Signal Processing*, vol. 60, no. 2, pp. 648–659, 2012.

[31] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sensing Lett.*, vol. 7, no. 4, pp. 736–740, 2010.

[32] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 51, no. 5, pp. 2734–2752, 2013.

[33] G. Moser, S. B. Serpico, and J. A. Benediktsson, "Land-cover mapping by Markov modeling of spatial-contextual information," *Proc. IEEE*, vol. 101, no. 3, pp. 631–651, 2013.

[34] M. J. Canty, A. A. Nielsen, and M. Schmidt, "Automatic radiometric normalization of multitemporal satellite imagery," *Remote Sensing Environ.*, vol. 91, no. 3–4, pp. 441–451, June 2004.

[35] S. Inamdar, F. Bovolo, L. Bruzzone, and S. Chaudhuri, "Multidimensional probability density function matching for pre-processing of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 4, pp. 1243–1252, 2008.

[36] F. Petitjean, J. Inglada, and P. Gancarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, 2011.

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

[38] D. Tuia, J. Muñoz-Marí, L. Gómez-Chova, and J. Malo, "Graph matching for adaptation in remote sensing," *IEEE Trans. Geosci. Remote Sensing*, vol. 51, no. 1, pp. 329–341, 2013.

[39] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A DASVM classification technique and a circular validation strategy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 770–787, 2010.

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

[41] G. Jun and J. Ghosh, "Semisupervised learning of hyperspectral data with unknown land-cover classes," *IEEE Trans. Geosci. Remote Sensing*, vol. 51, no. 1, pp. 273–282, 2013.

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

[43] E. Izquierdo-Verdiguier, V. Laparra, L. Gómez-Chova, and G. Camps-Valls, "Encoding invariances in remote sensing image classification with SVM," *IEEE Geosci. Remote Sensing Lett.*, vol. 10, no. 5, pp. 981–985, 2013.

[SP]