12

Hyperspectral Image Processing: Methods and Approaches

Jun Li
Sun Yat-Sen University

Antonio Plaza
University of Extremadura

Acronyms and Definitions

EMPs	Extended morphological profiles
EMPs	Extended morphological profiles
LDA	Linear discriminant analysis
LogDA	Logarithmic discriminant analysis
MLR	Multinomial logistic regression
MLRsubMRF	Subspace-based multinomial logistic regression followed by Markov random fields
MPs	Morphological profiles
MRFs	Markov random fields
PCA	Principal component analysis
QDA	Quadratic discriminant analysis
RHSEG	Recursive hierarchical segmentation
ROSIS	Reflective optics spectrographic imaging system
SVMs	Support vector machines
TSVMs	Transductive support vector machines

12.1 Introduction

Hyperspectral imaging is concerned with the measurement, analysis, and interpretation of spectra acquired from a given scene (or specific object) at a short, medium, or long distance, typically, by an airborne or satellite sensor [1]. The special characteristics of hyperspectral data sets pose different processing problems, which must be necessarily tackled under specific mathematical formalisms [2], such as classification and segmentation [3] or spectral mixture analysis [4]. Several machine learning and image-processing techniques have been applied to extract relevant information from hyperspectral data during the last decade [5,6]. Taxonomies of hyperspectral image-processing algorithms have been presented in the literature [3,7,8]. It should be noted, however, that most recently developed hyperspectral image-processing techniques focus on analyzing the spectral and spatial informations contained in the hyperspectral data in simultaneous fashion [9]. In other words, the importance of analyzing spatial and spectral information simultaneously has been identified as a desired goal by many scientists devoted to hyperspectral image analysis. This type of processing has been approached in the past from various points of view. For instance, several possibilities are discussed by Landgrebe [10] for the refinement of results obtained by spectral-based techniques through a second step based on spatial context. Such contextual classification [11] accounts for the tendency of certain ground cover classes to occur more frequently in some contexts than in others. In certain applications, the integration of high spatial and spectral information is mandatory to achieve sufficiently accurate mapping and/or detection results. For instance, urban area mapping requires sufficient spatial resolution to distinguish small spectral classes, such as trees in a park or cars on a street [12] (Figure 12.1).

However, there are several important challenges when performing hyperspectral image classification. In particular, supervised classification faces challenges related with the unbalance between high dimensionality and the limited number of training samples or the presence of mixed pixels in the data (which may compromise classification results for coarse spatial resolutions). Specifically, due to the small number of training samples and the high dimensionality of the hyperspectral data, reliable estimation of statistical class parameters is a very challenging goal [13]. As a result, with a limited training set, classification

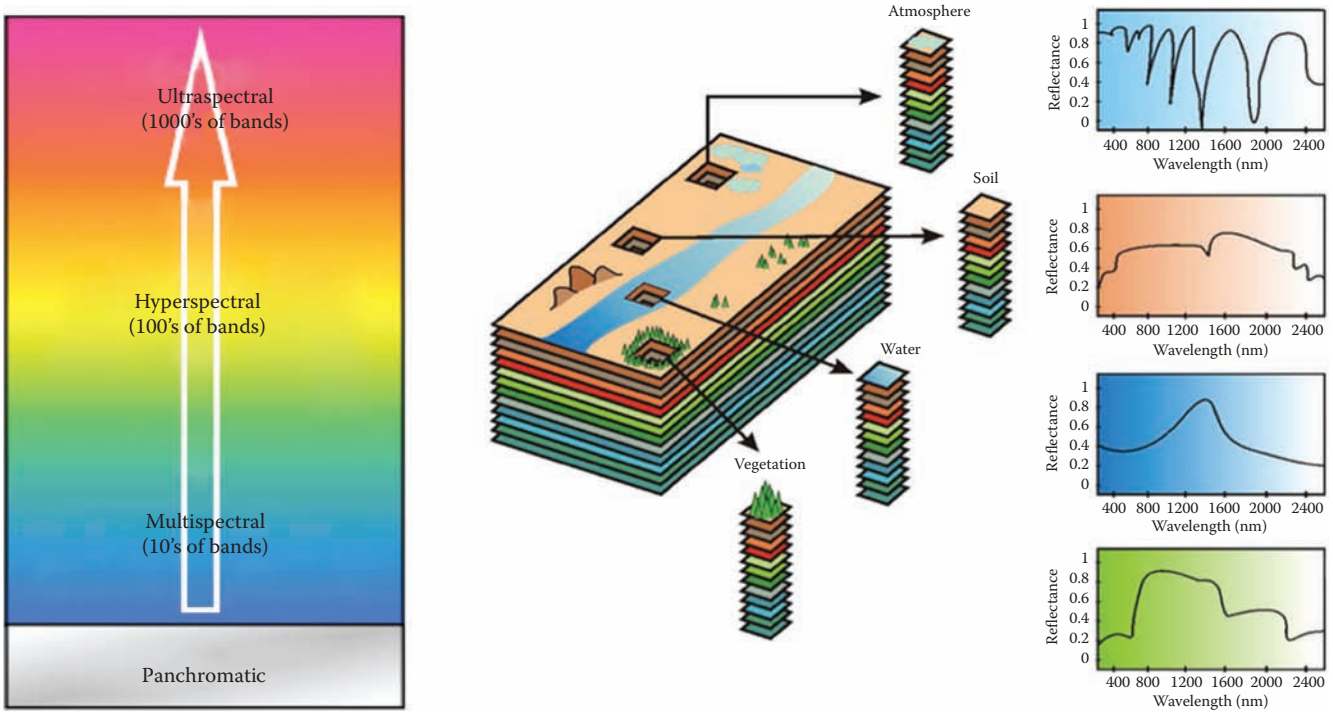


FIg u r e 1 2.1 The challenges of increased dimensionality in remote-sensing data interpretation.

accuracy tends to decrease as the number of features increases. This is known as the Hughes effect [14]. Another relevant challenge is the need to integrate the spatial and spectral information to take advantage of the complementarities that both sources of information can provide. These challenges are quite important for future developments and solutions to some of them have been proposed. Specifically, supervised [15] and semisupervised [16–18] techniques for hyperspectral image classification, strategies for integrating the spatial and the spectral information [19–22], or subspace classifiers [23] that can better exploit the intrinsic nature of hyperspectral data have been quite popular in the recent literature.

Our main goal in this chapter is to provide a seminal view on recent advances in techniques for hyperspectral image analysis that can successfully deal with the dimensionality problem and with the limited availability of training samples *a priori*, while taking into account both the spectral and spatial properties of the data. The remainder of the chapter is structured as follows. Section 12.2 discusses available techniques for hyperspectral image classification, including both supervised and semisupervised approaches, techniques for integrating spatial and spectral information and subspace-based approaches. Section 12.3 provides an experimental comparison of the techniques discussed in Section 12.2, using a hyperspectral data set collected by the ROSIS over the University of Pavia, Italy, which is used here as a common benchmark to outline the properties of the different processing techniques discussed in the chapter. Finally, Section 12.4 concludes the paper with some remarks and hints at the most pressing ongoing research directions in hyperspectral image classification.

12.2 Classification Approaches

In this section, we outline some of the main techniques and challenges in hyperspectral image classification. Hyperspectral image classification has been a very active area of research in recent years [3]. Given a set of observations (i.e., pixel vectors in a hyperspectral image), the goal of classification is to assign a unique label to each pixel vector so that it is well defined by a given class. In Figure 12.2, we provide an overview of a popular strategy to conduct hyperspectral image classification, which is based on the availability of labeled samples. After an optional dimensionality reduction step, a supervised classifier is trained using a set of labeled samples (which are often randomly selected from a larger pool of samples) and then tested with a disjoint set of labeled samples in order to evaluate the classification accuracy of the classifier.

Supervised classification has been widely used in hyperspectral data interpretation [2], but it faces challenges related with the high dimensionality of the data and the limited availability of training samples, which may not be easy to collect in pure form. However, mixed training samples can also offer relevant information about the participating classes [10]. In order to address these issues, subspace-based approaches [23,24] and semisupervised learning techniques [25] have been developed. In subspace approaches, the goal is to reduce the dimensionality of the input space in order to better exploit the (limited) training samples available. In semisupervised learning, the idea is to exploit the information conveyed by additional (unlabeled) samples, which can complement the available labeled samples with a certain degree of confidence. In all cases, there is a clear need to

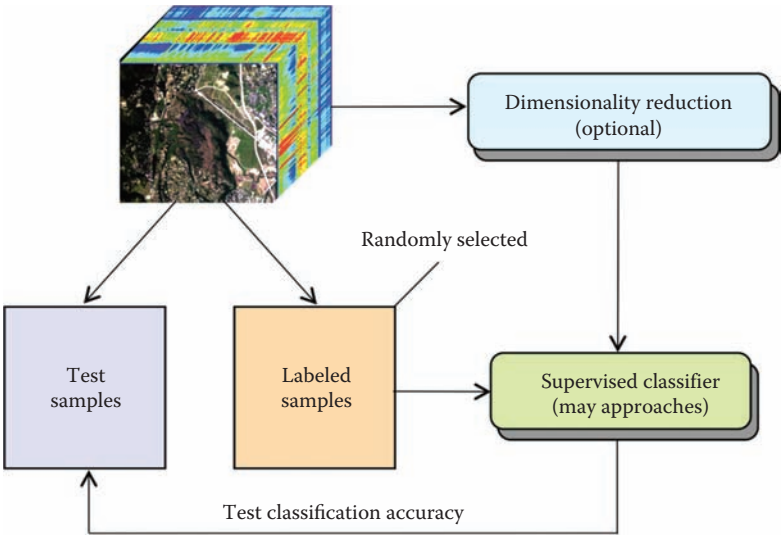


FIg u r e 1 2.2 Standard approach for supervised hyperspectral image classification.

integrate the spatial and spectral information to take advantage of the complementarities that both sources of information can provide [9]. An overview of these different aspects, which are crucial to hyperspectral image classification, is provided in the following subsections.

12.2.1 Supervised Classification

Several techniques have been used to perform supervised classification of hyperspectral data. For instance, in discriminant classifiers, several types of discriminant functions can be applied: nearest neighbor, decision trees, linear functions, or nonlinear functions (see Figure 12.3). In linear discriminant analysis (LDA) [26], a linear function is used in order to maximize the discriminatory power and separate the available classes effectively. However, such a linear function may not be the best choice, and nonlinear strategies such as quadratic discriminant analysis (QDA) or logarithmic discriminant analysis (LogDA) have also been used. The main problem of these classic supervised classifiers, however, is their sensitivity to the Hughes effect.

In this context, kernel methods such as the support vector machine (SVM) have been widely used in order to deal effectively with the Hughes phenomenon [27,28]. The SVM was first investigated as a binary classifier [29]. Given a training set mapped into a Hilbert space by some mapping, the SVM separates the data by an optimal hyperplane that maximizes the margin (see Figure 12.4). However, the most widely used approach in hyperspectral classification is to combine soft margin classification with a kernel trick that allows separation of the classes in a higher-dimensional space by means of a nonlinear transformation (see Figure 12.5). In other words, the SVM used with a kernel function is a nonlinear classifier, where the nonlinear ability is included in the kernel and different kernels lead to different types of SVMs. The extension of SVM to the multiclass cases is usually done by combining several binary classifie s.

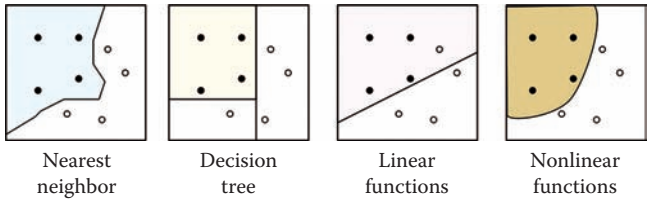


FIg ur e 12.3 Typical discriminant functions used in supervised classifie s.

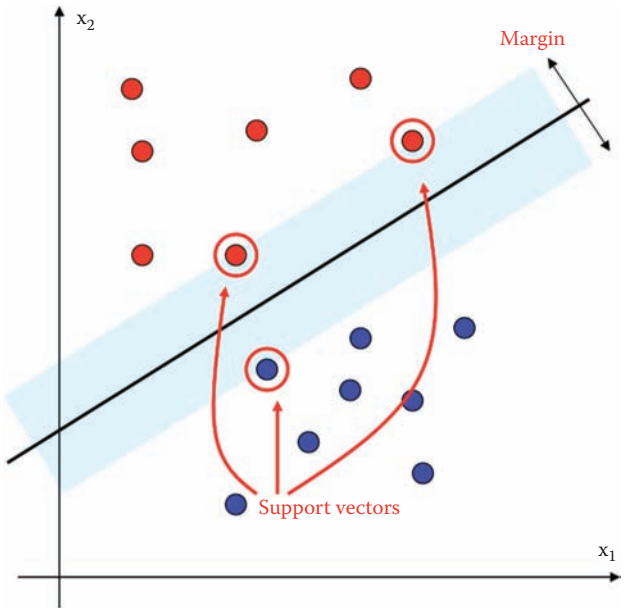


FIg u r e 1 2.4 Soft argin classification with slack variables.

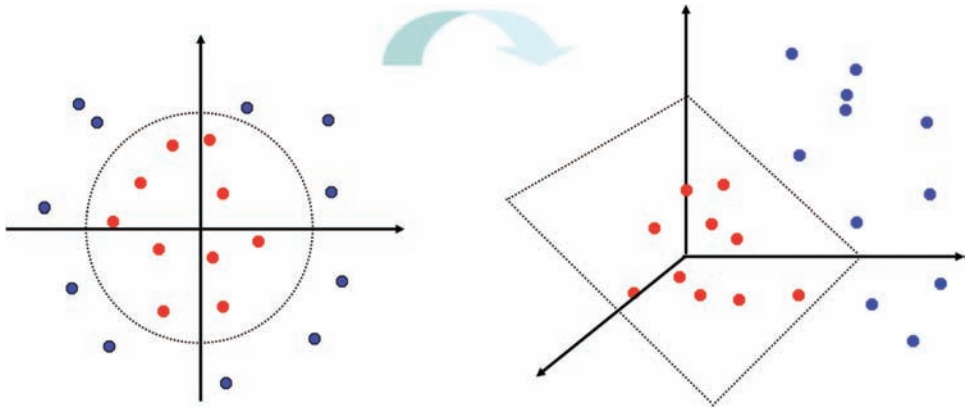


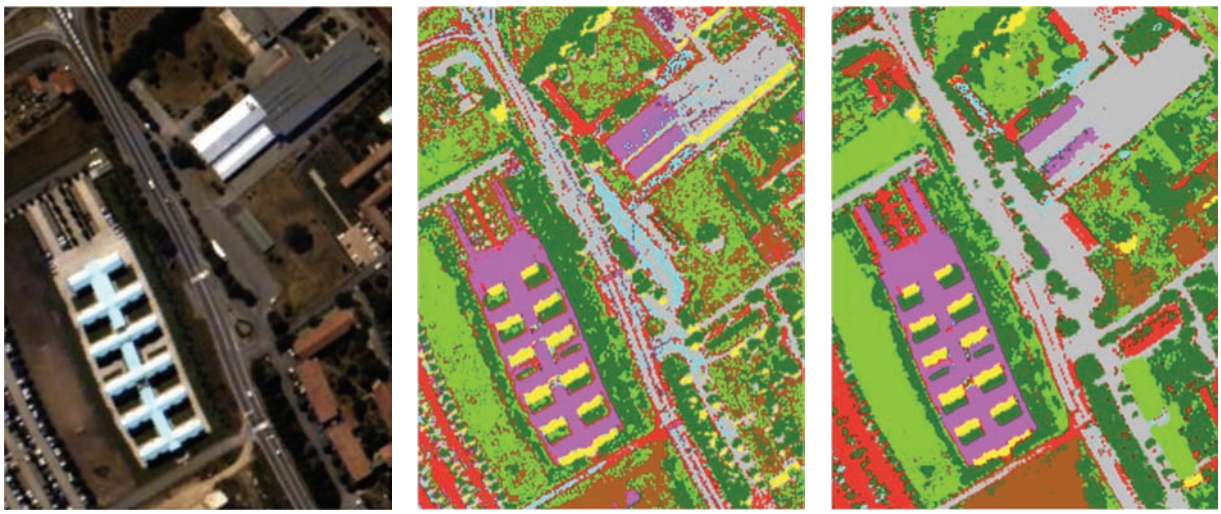
FIg u r e 1 2.5 The kernel trick allows separation of the classes in a higher-dimensional space by means of a linear or nonlinear transformation.

12.2.2 Spectral–Spatial Classification

Several efforts have been performed in the literature in order to integrate spatial–contextual information in spectral-based classifiers for hyperspectral data [3,9]. It is now commonly accepted that using the spatial and the spectral information simultaneously provides significant advantages in terms of improving the performance of classification techniques. An illustration of the importance of integrating spatial and spectral information is given in Figure 12.6. As shown in this figure, spectral–spatial classification (obtained using morphological transformations) provides a better interpretation of classes such as urban features, with a better delineation and characterization of complex urban structures.

Some of the approaches that integrate spatial and spectral information include spatial information prior to the classification, during the feature extraction stage. Mathematical morphology [30] has been particularly successful for this purpose. Morphology is a widely used approach for modeling the spatial characteristics of the objects in remotely sensed images. Advanced morphological techniques such as morphological profiles [31] have been successfully used for feature extraction prior to classification of hyperspectral data by extracting the first few principal components of the data using principal component analysis [13] and then building so-called extended morphological profiles (EMPs) on the first few components to extract relevant features for classification [32].

Another strategy in the literature has been to exploit simultaneously the spatial and the spectral information. For instance, in order to incorporate the spatial context into kernel-based classifiers, a pixel entity can be redefined, simultaneously both in the spectral domain (using its spectral content) and also in the spatial domain, by applying some feature extraction to its surrounding area, which yields spatial (contextual) features, for example, the mean or standard deviation per spectral band. These separated entities lead to two different kernel matrices, which can be easily computed. At this point, one can sum spectral and textural dedicated kernel matrices and introduce the



True color image | Spectral classification | Spectral–spatial classification

FIg u r e 1 2.6 The importance of using spatial and spectral information in classification.

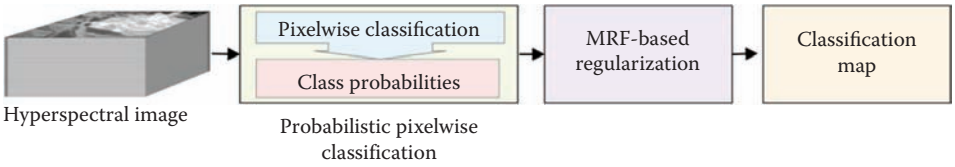


FIg u r e 1 2.7 Standard processing framework using pixel-wise probabilistic classification followed by MRF-based spatial postprocessing.

cross information between textural and spectral features in the formulation. This simple methodology yields a full family of new kernel methods for hyperspectral data classification, defined in [33] and implemented using the SVM classifier, thus providing a composite kernel-based SVM.

Another approach to jointly exploit spatial and spectral information is to use Markov random fields (MRFs) for the characterization of spatial information. MRFs exploit the continuity, in probability sense, of neighboring labels [19,34]. In this regard, several techniques have exploited an MRF-based regularization procedure, which encourages neighboring pixels to have the same label when performing probabilistic classification of hyperspectral data sets. An example of this type of processing is given in Figure 12.7, in which a pixel-wise probabilistic classification is followed by an MRF-based spatial postprocessing that refines the initial probabilistic classification output.

Several other approaches include spatial information as a postprocessing, that is, after a spectral-based classification has been conducted. One of the first classifiers with spatial postprocessing developed in the hyperspectral imaging literature was the well-known extraction and classification of homogeneous objects [10]. Another one is the strategy adopted in [35], which combines the output of a pixel-wise SVM classifier with the morphological watershed transformation [30] in order to provide a more spatially homogeneous classification. A similar strategy is adopted in Ref. [36], in which the output of the SVM classifier is combined with the segmentation result provided by the unsupervised recursive hierarchical segmentation (RHSEG)* algorithm.

12.2.3 Subspace-Based Approaches

Subspace projection methods [23] have been shown to be a powerful class of statistical pattern classification algorithms. These methods can handle the high dimensionality of hyperspectral data by bringing it to the right subspace without losing the original information that allows for the separation of classes. In this context, subspace projection methods can provide competitive advantages by separating classes that are very similar in spectral sense, thus addressing the limitations in the classification process due to the presence of highly mixed pixels. The idea of applying subspace projection methods to improve classification relies on the basic assumption that the samples within each class can approximately lie in a lower-dimensional subspace. Thus, each class may be represented by a subspace spanned by a set of basis vectors, while the classification criterion for a new input sample would be the distance from the class subspace [24].

Recently, several subspace projection methods have been specifically designed for improving hyperspectral data characterization, with successful results. For instance, the subspace-based multinomial logistic regression followed by Markov random fields (MLRsubMRF) method in Ref. [23] first performs a learning step in which the posterior probability distributions are modeled by a multinomial logistic regression (MLR) [37] combined with a subspace projection method. Then, the method infers an image of class labels from a posterior distribution built on the learned subspace classifier and on a multilevel logistic spatial prior on the image of labels. This prior is an MRF that exploits the continuity, in probability sense, of neighboring labels. The basic assumption is that, in a hyperspectral image, it is very likely that two neighboring pixels will have the class same label. The main contribution of the MLRsubMRF method is therefore the integration of a subspace projection method with the MLR, which is further combined with spatial–contextual information in order to provide a good characterization of the content of hyperspectral imagery in both spectral and the spatial domains. As will be shown by our experiments, the accuracies achieved by this approach are competitive with those provided by many other state-of-the-art supervised classifiers for hyperspectral analysis (Figure 12.8).

12.2.4 Semisupervised Classification

A relevant challenge for supervised classification techniques is the limited availability of labeled training samples, since their collection generally involves expensive ground campaigns [38]. While the collection of labeled samples is generally difficult, expensive, and time-consuming, unlabeled samples can be generated in a much easier way. This observation has fostered the idea of adopting semisupervised learning techniques in hyperspectral image classification. The main assumption of such techniques is that new (unlabeled) training samples can be obtained from a (limited) set of available labeled samples without significant effort/cost. This can be simply done by selecting new samples from the spatial neighborhood of available labeled samples, under the principle that it is likely that the new unlabeled samples will have similar class labels as the already available ones.

In contrast to supervised classification, semisupervised algorithms generally assume that a limited number of labeled samples are available *a priori* and then enlarge the training set using unlabeled samples, thus allowing these approaches to address ill-posed problems. However, in order for this strategy to work,

* http://opensource.gsfc.nasa.gov/projects/HSEG/.

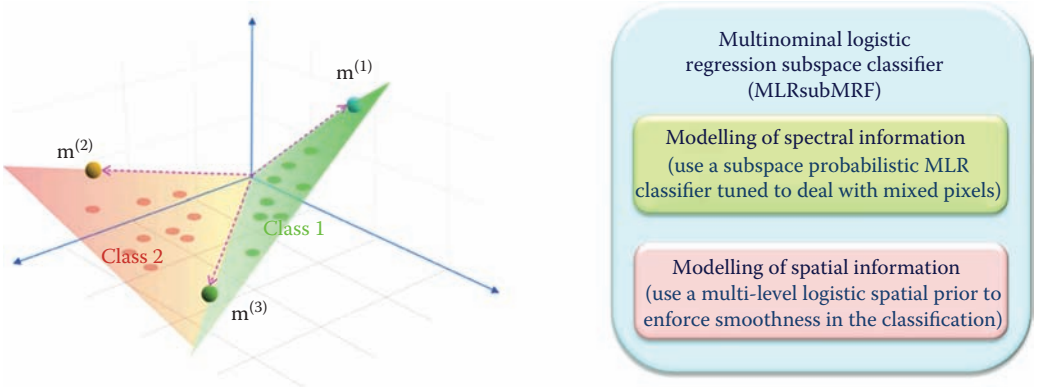


FIg u r e 1 2.8 Multinomial logistic regression subspace classifier.

several requirements need to be met. First and foremost, the new (unlabeled) samples should be obtained without significant cost/effort. Second, the number of unlabeled samples required in order for the semisupervised classifier to perform properly should not be too high in order to avoid increasing computational complexity in the classification stage. In other words, as the number of unlabeled samples increases, it may be unbearable for the classifier to properly exploit all the available training samples due to computational issues. Further, if the unlabeled samples are not properly selected, these may confuse the classifier, thus introducing significant divergence or even reducing the classification accuracy obtained with the initial set of labeled samples. In order to address these issues, it is very important that the most highly informative unlabeled samples are identified in computationally efficient fashion, so that significant improvements in classification performance can be observed without the need to use a very high number of unlabeled samples.

The area of semisupervised learning for remote-sensing data analysis has experienced a significant evolution in recent years. For instance, looking at machine learning–based approaches, in Ref. [39] transductive support vector machines (TSVMs) are used to gradually search a reliable separating hyperplane (in the kernel space) with a transductive process that incorporates both labeled and unlabeled samples in the training phase. In [40], a semisupervised method is presented that exploits the wealth of unlabeled samples in the image and naturally gives relative importance to the labeled ones through a graph-based methodology. In [41], kernels combining spectral–spatial information are constructed by applying spatial smoothing over the original hyperspectral data and then using composite kernels in graph-based classifiers. In [18], a semisupervised SVM is presented that exploits the wealth of unlabeled samples for regularizing the training kernel representation locally by means of cluster kernels. In [42], a new semisupervised approach is presented that exploits unlabeled training samples (selected by means of an active selection strategy based on the entropy of the samples). Here, unlabeled samples are used to improve the estimation of the class distributions, and the obtained classification is refined by using a spatial multilevel logistic prior. In [16], a novel context-sensitive semisupervised SVM is presented that exploits the contextual information of the pixels belonging to the neighborhood system of each training sample in the learning phase to improve the robustness to possible mislabeled training patterns.

In [43], two semisupervised one-class (SVM-based) approaches are presented in which the information provided by unlabeled samples present in the scene is used to improve classification accuracy and alleviate the problem of free-parameter selection. The first approach models data marginal distribution with the graph Laplacian built with both labeled and unlabeled samples. The second approach is a modification of the SVM cost function that penalizes more the errors made when classifying samples of the target class. In [44], a new method to combine labeled and unlabeled pixels to increase classification reliability and accuracy, thus addressing the sample selection bias problem, is presented and discussed. In [45], an SVM is trained with the linear combination of two kernels: a base kernel working only with labeled examples is deformed by a likelihood kernel encoding similarities between labeled and unlabeled examples and then applied in the context of urban hyperspectral image classification. In [46], similar concepts to those addressed before are adopted using a neural network as the baseline classifier. In [17], a semiautomatic procedure to generate land cover maps from remote-sensing images using active queries is presented and discussed.

Last but not least, we emphasize that the techniques summarized in this section only represent a small sample (and somehow subjective selection) of the vast collection of approaches presented in recent years for hyperspectral image classification. For a more exhaustive summary of available techniques and future challenges in this area, we point interested readers to [47].

12.3 Experimental Comparison

In this section, we illustrate the performance of the techniques described in the previous section by processing a widely used hyperspectral data set collected by ROSIS optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union. The image size in pixels is 610 × 340, with very-high-spatial

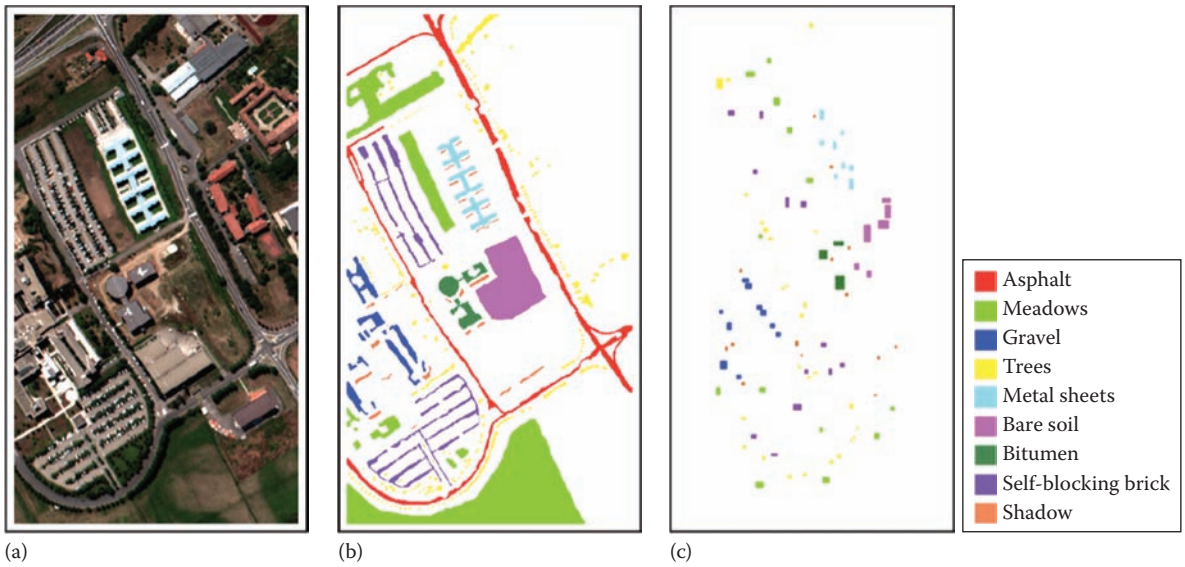


FIg u r e 12.9 The ROSIS Pavia University scene used in our experiments. The scene was collected by the ROSIS instrument in the framework of the HySens campaign. It comprises 103 spectral bands between 0.4 and 0.9 μm and was collected over an urban area at the University of Pavia, Italy. (a) ROSIS Pavia hyperspectral image. (b) Ground-truth classes. (c) Fixed training set.

resolution of 1.3 m per pixel. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). Figure 12.9a shows a false color composite of the image, while Figure 12.9b shows nine reference classes of interest, which comprise urban features, as well as soil and vegetation features. Finally, Figure 12.9c shows a fixed training set available for the scene, which comprises 3,921 training samples (42,776 samples are available for testing). This scene has been widely used in the hyperspectral imaging community to evaluate the performance of processing algorithms [6]. It represents a case study that integrates a challenging urban classification problem, with a data set comprising high spatial and spectral resolution, and a highly reliable ground truth, with a well-established training set. All these factors have made the scene a standard and an excellent test bed for evaluation of hyperspectral image classification algorithms, particularly those integrating the spatial and the spectral information.

Table 12.1 illustrates the classification results obtained by different supervised classifiers for the ROSIS University of Pavia scene in Figure 12.9a, using the same training data in Figure 12.9c to train the classifiers and a mutually exclusive set of labeled samples in Figure 12.9b to test the classifiers. As shown by Table 12.1, the SVM classifier obtained comparatively superior performance in terms of the overall classification accuracy when compared with discriminant classifiers such as LDA, QDA or LogDA.

TABLe 12.1 Classification Results Obtained for the ROSIS Pavia University Scene by the SVM Classifier as Compared with Several Discriminant Classifiers

Class	Training	Testing	LDA	QDA	LogDA	SVM
Asphalt (6,631)	548	6,083	69.45	67.75	70.89	83.71
Meadows (18,649)	532	18,117	81.92	75.73	76.72	70.25
Gravel (2,099)	265	1,834	39.11	59.79	55.31	70.32
Trees (3,064)	231	2,833	95.07	96.64	96.38	97.81
Metal Sheets (1,345)	375	970	99.41	99.93	100	99.41
Bare Soil (5,029)	540	4,489	46.59	73.49	75.06	92.25
Bitumen (1,330)	392	938	63.31	93.53	83.98	81.58
Self-Blocking Bricks (3,682)	524	3,158	88.29	89.52	87.91	92.59
Shadow (947)	514	433	99.79	99.26	99.79	96.62
Overall accuracy	—	—	77.95	77.95	78.41	80.99

LDA stands for linear discriminant analysis. QDA stands for quadratic discriminant analysis. LogDA stands for logarithmic discriminant analysis. SVM stands for support vector machine. The numbers in the parentheses are the total number of available samples.

TABLe 1 2.2 Classification Results Obtained for the ROSIS Pavia University Scene by the SVM Classifier as Compared with a Composite SVM Obtained Using the Summation Kernel

Class	Training	Testing	SVM	Composite SVM
Asphalt (6,631)	548	6,083	83.71	79.85
Meadows (18,649)	532	18,117	70.25	84.76
Gravel (2,099)	265	1,834	70.32	81.87
Trees (3,064)	231	2,833	97.81	96.36
Metal Sheets (1,345)	375	970	99.41	99.37
Bare Soil (5,029)	540	4,489	92.25	93.55
Bitumen (1,330)	392	938	81.58	90.21
Self-Blocking Bricks (3,682)	524	3,158	92.59	92.81
Shadow (947)	514	433	96.62	95.35
Overall accuracy	—	—	80.99	87.18

Source: Camps-Valls, L. et al., *IEEE Geosci. Remote Sens. Lett.*, 3, 93, 2006.
The numbers in the parentheses are the total number of available samples.

In a second experiment, we compared the standard SVM classifier with the composite kernel strategy as defi ed in Ref. [33], which combines spatial and spectral information at the kernel level. After carefully evaluating all possible types of composite kernels, the summation kernel provided the best performance in our experiments as reported in Table 12.2. Th s table suggests the importance of using spatial and spectral information in the analysis of hyperspectral data.

Although the integration of spatial and spectral information carried out by the composite kernel in Table 12.2 is performed at the classification stage, the spatial information can also be included prior to classification. For illustrative purposes, Table 12.3 compares the classification results obtained by the SVM applied on the original hyperspectral image to those obtained using a combination of the morphological EMP for feature extraction followed by SVM for classification (EMP/SVM).

As shown in Table 12.3, the EMP/SVM provides good classification results for the ROSIS University of Pavia scene, which represent a good improvement over the results obtained using the original hyperspectral image as input to the classifier. These results confirm the importance of using spatial and spectral information for classification purposes, as it was already found in the experimental results reported in Table 12.2.

In order to illustrate other approaches that use spatial information as postprocessing, Tables 12.4 and 12.5, respectively, compare the classification results obtained by the traditional SVM with those found using the strategy adopted in Ref. [35], which combines the output of a pixel-wise SVM classifier with the morphological watershed, and in Ref. [36], in which the output of the SVM classifier is combined with the segmentation result provided by the RHSEG algorithm. As shown in Tables 12.4 and 12.5, these strategies lead to improved classification with regard to the traditional SVM. In fact, Tables 12.2 through 12.5 illustrate different aspects concerning the integration of spatial and spectral information. The results reported in Table 12.2 are obtained by integrating spatial and spectral information at the classification stage. On the other hand, the results reported in Table 12.3 are obtained by using spatial information

TABLE 12.3 Classification Results Obtained for the ROSIS Pavia University Scene by the SVM Classifier as Compared with Those Obtained Using a Combination of the Morphological EMP for Feature Extraction Followed by SVM for Classification (EMP/SVM)

Class	Training	Testing	SVM	EMP/SVM
Asphalt (6,631)	548	6,083	83.71	95.36
Meadows (18,649)	532	18,117	70.25	80.33
Gravel (2,099)	265	1,834	70.32	87.61
Trees (3,064)	231	2,833	97.81	98.37
Metal Sheets (1,345)	375	970	99.41	99.48
Bare Soil (5,029)	540	4,489	92.25	63.72
Bitumen (1,330)	392	938	81.58	98.87
Self-Blocking Bricks (3,682)	524	3,158	92.59	95.41
Shadow (947)	514	433	96.62	97.68
Overall accuracy	—	—	80.99	85.22

The numbers in the parentheses are the total number of available samples.

TABLe 12.4 Classification Results Obtained for the ROSIS Pavia University Scene by the SVM Classifier as Compared with Those Obtained Using the Strategy Adopted in [35], Which Combines the Output of a Pixel-Wise SVM Classifier with the Morphological Watershed

Class	Training	Testing	SVM	SVM + Watershed
Asphalt (6,631)	548	6,083	83.71	94.28
Meadows (18,649)	532	18,117	70.25	76.41
Gravel (2,099)	265	1,834	70.32	69.89
Trees (3,064)	231	2,833	97.81	98.30
Metal Sheets (1,345)	375	970	99.41	99.78
Bare Soil (5,029)	540	4,489	92.25	97.51
Bitumen (1,330)	392	938	81.58	97.14
Self-Blocking Bricks (3,682)	524	3,158	92.59	98.29
Shadow (947)	514	433	96.62	97.57
Overall accuracy	—	—	80.99	86.64

The numbers in the parentheses are the total number of available samples.

TABLe 12.5 Classification Results Obtained for the ROSIS Pavia University Scene by the SVM Classifier as Compared with Those Obtained Using the Strategy Adopted in [36], in Which the Output of the SVM Classifier Is Combined with the Segmentation Result Provided by the Recursive Hierarchical Segmentation (RHSEG) Algorithm Developed by James C. Tilton at NASA's Goddard Space Flight Center

Class	Training	Testing	SVM	SVM + RHSEG
Asphalt (6,631)	548	6,083	83.71	94.77
Meadows (18,649)	532	18,117	70.25	89.32
Gravel (2,099)	265	1,834	70.32	96.14
Trees (3,064)	231	2,833	97.81	98.08
Metal Sheets (1,345)	375	970	99.41	99.82
Bare Soil (5,029)	540	4,489	92.25	99.76
Bitumen (1,330)	392	938	81.58	100
Self-Blocking Bricks (3,682)	524	3,158	92.59	99.29
Shadow (947)	514	433	96.62	96.48
Overall accuracy	—	—	80.99	93.85

The numbers in the parentheses are the total number of available samples.

at a preprocessing step prior to classification. Finally, the results reported in Tables 12.4 and 12.5 correspond to cases in which spatial information is included at a postprocessing step after conducting spectral-based classification. As a result, the comparison reported in this section illustrates different scenarios in which spatial and spectral information are used in complementary fashion but following different strategies, that is, spatial information is included at different stages of the classification process (preprocessing, postprocessing, or kernel level).

After evaluating the importance of including spatial and spectral information, we now discuss the possibility to perform a better modeling of the hyperspectral data by working on a subspace. This is due to the fact that the dimensionality of the hyperspectral data is very high, and often, the data live in a subspace. Hence, if the proper subspace is identified prior to classification, adequate results can be obtained. In order to illustrate this concept, Table 12.6 shows the results obtained after comparing the SVM classifier with a subspace-based classifier such as the MLRsub [23], followed by an MRF-based spatial regularizer.

The idea of applying subspace projection methods relies on the basic assumption that the samples within each class can approximately lie in a lower-dimensional subspace. In the experiments reported in Table 12.6, it can also be seen that spatial information (included as an MRF-based postprocessing) can be greatly beneficial in order to improve classification performance.

Finally, it is worth noting that the results discussed in the previous text are all based on supervised classifiers that assume the sufficient availability of labeled training samples. In case that no sufficient labeled samples are available, semisupervised learning techniques can be used to generate additional unlabeled samples (from the initial set of labeled samples) that can complement the available labeled samples. The unlabeled samples can also be used to enhance subspace-based classifiers in case dimensionality issues are found to be relevant in the considered case

TABLE 12.6 Classification Results Obtained for the ROSIS Pavia University Scene by the SVM Classifier as Compared with a Subspace-Based Classifier Followed by Spatial Postprocessing (MLRsubMRF)

Class	Training	Testing	SVM	MLRsubMRF
Asphalt (6,631)	548	6,083	83.71	93.83
Meadows (18,649)	232	18,117	70.25	94.80
Gravel (2,099)	265	1,834	70.32	71.13
Trees (3,064)	231	2,833	97.81	92.17
Metal Sheets (1,345)	375	970	99.41	100
Bare Soil (5,029)	540	4,489	92.25	98.43
Bitumen (1,330)	392	938	81.58	99.32
Self-Blocking Bricks (3,682)	524	3,158	92.59	95.19
Shadow (947)	514	433	96.62	96.20
Overall accuracy	—	—	80.99	94.10

The numbers in the parentheses are the total number of available samples.

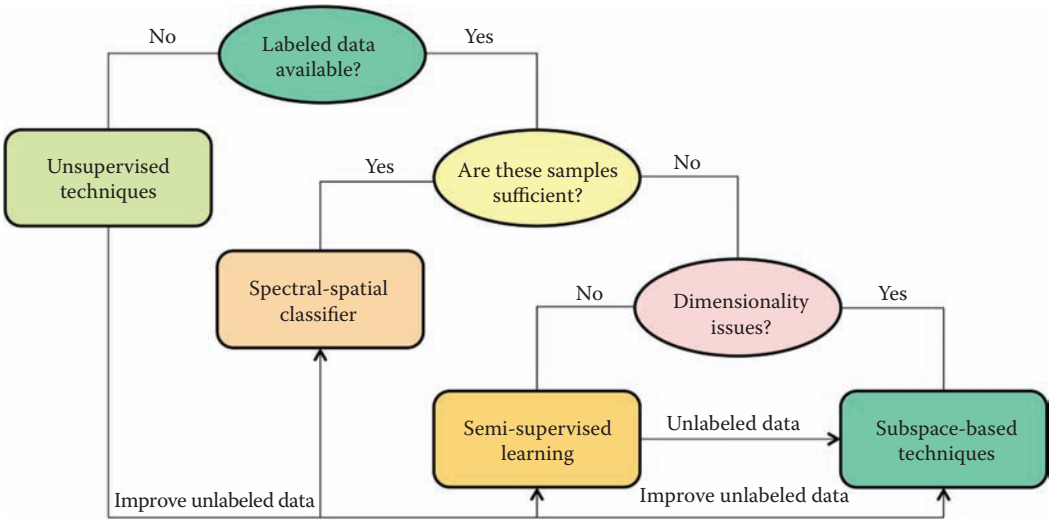


FIgure 12.10 Summary of contributions in hyperspectral image classification discussed in this chapter.

study. If sufficient labeled samples are available, then the use of a spectral–spatial classifier is generally recommended as spatial information can provide a very important complement to the spectral information. Finally, in case that labeled samples are not available at all, unsupervised techniques need to be used for classification purposes. For instance, a relevant unsupervised method successfully applied to hyperspectral image data is Tilton's RHSEG algorithm.* The different analysis scenarios for classification discussed in this chapter are summarized in Figure 12.10.

12.4 Conclusions and Future Directions

In this chapter, we have provided an overview of recent advances in techniques and methods for hyperspectral image classification. The array of techniques particularly illustrated in this chapter comprises supervised and semisupervised approaches, techniques able to exploit both spatial and spectral information, and techniques able to take advantage of a proper subspace representation of the hyperspectral data before conducting the classification in spatial–spectral terms. These approaches represent a subjective selection of the wide range of techniques currently adopted for hyperspectral image classification [3], which include other techniques and strategies that have not been covered in detail in this chapter for space considerations. Of particular importance is the recent role of sparse classification methods [2,20], which are gaining significant popularity and will likely play an important role in this research area in upcoming years.

Despite the wide arrange of techniques and strategies for hyperspectral data interpretation currently available, some unresolved issues still remain. For instance, the geometry of hyperspectral data is quite complex and dominated by nonlinear structures. This issue has undoubtedly an impact in the outcome of the classification techniques discussed in this section. In order to mitigate this, manifold learning has been proposed [48]. An important property of manifold learning is that it can model

* http://opensource.gsfc.nasa.gov/projects/HSEG/.

and characterize the complex nonlinear structure of the data prior to classification [49]. Another remaining issue is the very high computational complexity of some of the techniques available for classification of hyperspectral data [50]. In other words, there is a clear need to develop efficient classification techniques that can deal with the very large dimensionality and complexity of hyperspectral data.

Acknowledgments

The authors thank Prof. Paolo Gamba, the University of Pavia, and the HySens project for providing the ROSIS data used in the experiments.

References

1. A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, Imaging spectrometry for Earth remote sensing, *Science*, 228(4704), 1147–1153, 1985.
2. J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, Hyperspectral remote sensing data analysis and future challenges, *IEEE Geoscience and Remote Sensing Magazine*, 1(2), 6–36, 2013.
3. M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, Advances in spectral-spatial classifi ation of hyperspectral images, *Proceedings of the IEEE*, 101(3), 652–675, 2013.
4. J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, Hyperspectral unmixing overview: Geometrical, statistical and sparse regression-based approaches, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 354–379, 2012.
5. J. Chanussot, M. M. Crawford, and B.-C. Kuo, Foreword to the special issue on hyperspectral image and signal processing, *IEEE Transactions on Geoscience and Remote Sensing*, 48(11), 3871–3876, 2010.
6. A. Plaza, J. M. Bioucas-Dias, A. Simic, and W. J. Blackwell, Foreword to the special issue on hyperspectral image and signal processing, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 347–353, 2012.
7. N. Keshava and J. Mustard, Spectral unmixing, *IEEE Signal Processing Magazine*, 19(1), 44–57, 2002.
8. W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. Plaza, A. Ambikapathi, and C.-Y. Chi, A signal processing perspective on hyperspectral unmixing: Insights from remote sensing, *IEEE Signal Processing Magazine*, 31(1), 67–81, 2014.
9. A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot et al., Recent advances in techniques for hyperspectral image processing, *Remote Sensing of Environment*, 113(Suppl 1), 110–122, 2009.
10. D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: Wiley, 2003.
11. L. O. Jimenez, J. L. Rivera-Medina, E. Rodriguez-Diaz, E. Arzuaga-Cruz, and M. Ramirez-Velez, Integration of spatial and spectral information by means of unsupervised extraction and classifi ation for homogenous objects applied to multispectral and hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing*, 43(1), 844–851, 2005.
12. P. Gamba, F. Dell'Acqua, A. Ferrari, J. A. Palmason, J. A. Benediktsson, and J. Arnasson, Exploiting spectral and spatial information in hyperspectral urban data with high resolution, *IEEE Geoscience and Remote Sensing Letters*, 1(3), 322–326, 2004.
13. J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York: Springer-Verlag, 2006.
14. G. F. Hughes, On the mean accuracy of statistical pattern recognizers, *IEEE Transactions on Information Theory*, 14(1), 55–63, 1968.
15. G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*. Hoboken, NJ: Wiley, 2009.
16. L. Bruzzone and C. Persello, A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples, *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), 2142–2154, 2009.
17. J. Muñoz-Marí, D. Tuia, and G. Camps-Valls, Semisupervised classifi ation of remote sensing images with active queries, *IEEE Transactions on Geoscience and Remote Sensing*, 50(10), 3751–3763, 2012.
18. D. Tuia and G. Camps-Valls, Semisupervised remote sensing image classifi ation with cluster kernels, *IEEE Geoscience and Remote Sensing Letters*, 6(2), 224–228, 2009.
19. Y. Tarabalka, M. Fauvel, J. Chanussot, and J. Benediktsson, SVM- and MRF-based method for accurate classifi ation of hyperspectral images, *IEEE Geoscience and Remote Sensing Letters*, 7(4), 736–740, 2010.
20. B. Song, J. Li, M. Dalla Mura, P. Li, A. Plaza, J. M. Bioucas-Dias, J. A. Benediktsson, and J. Chanussot, Remotely sensed image classifi ation using sparse representations of morphological attribute profiles, *IEEE Transactions on Geoscience and Remote Sensing*, 52(8), 5122–5136, 2014.
21. J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, Generalized composite kernel framework for hyperspectral image classifi ation, *IEEE Transactions on Geoscience and Remote Sensing*, 51(9), 4816–4829, 2013.
22. J. Li, J. M. Bioucas-Dias, and A. Plaza, Spectral-spatial classifi ation of hyperspectral data using loopy belief propagation and active learning, *IEEE Transactions on Geoscience and Remote Sensing*, 51(2), 844–856, 2013.
23. J. Li, J. Bioucas-Dias, and A. Plaza, Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fi lds, *IEEE Transactions on Geoscience and Remote Sensing*, 50(3), 809–823, 2012.
24. J. Bioucas-Dias and J. Nascimento, Hyperspectral subspace identifi ation, *IEEE Transactions on Geoscience and Remote Sensing*, 46(8), 2435–2445, 2008.

25. I. Dopido, J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, Semi-supervised self-learning for hyperspectral image classifi ation, *IEEE Transactions on Geoscience and Remote Sensing*, 51(7), 4032–4044, 2013.
26. T. V. Bandos, L. Bruzzone, and G. Camps-Valls, Classifi ation of hyperspectral images with regularized linear discriminant analysis, *IEEE Transactions on Geoscience and Remote Sensing*, 47(3), 862–873, 2009.
27. F. Melgani and L. Bruzzone, Classifi ation of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, 42(8), 1778–1790, 2004.
28. G. Camps-Valls and L. Bruzzone, Kernel-based methods for hyperspectral image classifi ation, *IEEE Transactions on Geoscience and Remote Sensing*, 43, 1351–1362, 2005.
29. B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA: MIT Press, 2002.
30. P. Soille, *Morphological Image Analysis, Principles and Applications*, 2nd edn. Berlin, Germany: Springer-Verlag, 2003.
31. M. Pesaresi and J. Benediktsson, A new approach for the morphological segmentation of high-resolution satellite imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 39(2), 309–320, 2001.
32. J. Benediktsson, J. Palmason, and J. Sveinsson, Classifi ation of hyperspectral data from urban areas based on extended morphological profiles, *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 480–491, 2005.
33. G. Camps-Valls, L. Goméz-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, Composite kernels for hyperspectral image classifi ation, *IEEE Geoscience and Remote Sensing Letters*, 3, 93–97, 2006.
34. M. Khodadadzadeh, J. Li, A. Plaza, H. Ghassemian, J. Bioucas-Dias, and X. Li, Spectral–spatial classifi ation of hyperspectral data using local and global probabilities for mixed pixel characterization, *IEEE Transactions on Geoscience and Remote Sensing*, 52(10), 6298–6314, 2014.
35. Y. Tarabalka, J. Chanussot, and J. Benediktsson, Segmentation and classifi ation of hyperspectral images using watershed transformation, *Pattern Recognition*, 43, 2367–2379, 2010.
36. Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, Multiple spectral-spatial classifi ation approach for hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing*, 48(11), 4122–4132, 2011.
37. D. Böhning, Multinomial logistic regression algorithm, *Annals of the Institute of Statistics and Mathematics*, 44, 197–200, 1992.
38. F. Bovolo, L. Bruzzone, and L. Carlin, A novel technique for subpixel image classifi ation based on support vector machine, *IEEE Transactions on Image Processing*, 19, 2983–2999, 2010.
39. L. Bruzzone, M. Chi, and M. Marconcini, A novel transductive SVM for the semisupervised classifi ation of remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 11, 3363–3373, 2006.
40. G. Camps-Valls, T. Bandos, and D. Zhou, Semi-supervised graph-based hyperspectral image classifi ation, *IEEE Transactions on Geoscience and Remote Sensing*, 45, 3044–3054, 2007.
41. S. Velasco-Forero and V. Manian, Improving hyperspectral image classifi ation using spatial preprocessing, *IEEE Geoscience and Remote Sensing Letters*, 6, 297–301, 2009.
42. J. Li, J. Bioucas-Dias, and A. Plaza, Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning, *IEEE Transactions on Geoscience and Remote Sensing*, 48(11), 4085–4098, 2010.
43. J. Muñoz Marí, F. Bovolo, L. Gómez-Chova, L. Bruzzone, and G. Camp-Valls, Semisupervised one-class support vector machines for classifi ation of remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, 48(8), 3188–3197, 2010.
44. L. Gómez-Chova, G. Camps-Valls, L. Bruzzone, and J. Calpe-Maravilla, Mean MAP kernel methods for semisupervised cloud classifi ation, *IEEE Transactions on Geoscience and Remote Sensing*, 48(1), 207–220, 2010.
45. D. Tuia and G. Camps-Valls, Urban image classifi ation with semisupervised multiscale cluster kernels, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(1), 65–74, 2011.
46. F. Ratle, G. Camps-Valls, and J. Weston, Semisupervised neural networks for effici t hyperspectral image classifi ation, *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), 2271–2282, 2010.
47. G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jiménez, and J. Malo, *Remote Sensing Image Processing*. San Rafael, CA: Morgan and Claypool, 2011.
48. L. Ma, M. Crawford, and J. Tian, Local manifold learning based k-nearest-neighbor for hyperspectral image classifi ation, *IEEE Transactions on Geoscience and Remote Sensing*, 48(11), 4099–4109, 2010.
49. W. Kim and M. Crawford, Adaptive classifi ation for hyperspectral image data using manifold regularization kernel machines, *IEEE Transactions on Geoscience Remote Sensing*, 48(11), 4110–4121, 2010.
50. A. Plaza, J. Plaza, A. Paz, and S. Sanchez, Parallel hyperspectral image and signal processing, *IEEE Signal Processing Magazine*, 28(3), 119–126, 2011.