Recent Advances on Spectral–Spatial Hyperspectral Image Classification: An Overview and New Guidelines

Lin He, *Member, IEEE*, Jun Li, *Senior Member, IEEE*, Chenying Liu, *Student Member, IEEE*, and Shutao Li, *Senior Member, IEEE*

***Abstract*— Imaging spectroscopy, also known as hyperspectral imaging, has been transformed in the last four decades from being a sparse research tool into a commodity product available to a broad user community. Specially, in the last 10 years, a large number of new techniques able to take into account the special properties of hyperspectral data have been introduced for hyperspectral data processing, where hyperspectral image classification, as one of the most active topics, has drawn massive attentions. Spectral–spatial hyperspectral image classification can achieve better classification performance than its pixel-wise counterpart, since the former utilizes not only the information of spectral signature but also that from spatial domain. In this paper, we provide a comprehensive overview on the methods belonging to the category of spectral–spatial classification in a relatively unified context. First, we develop a concept of spatial dependency system that involves pixel dependency and label dependency, with two main factors: neighborhood covering and neighborhood importance. In terms of the way that the neighborhood information is used, the spatial dependency systems can be classified into fixed, adaptive, and global systems, which can accommodate various kinds of existing spectral–spatial methods. Based on such, the categorizations of single-dependency, bilayer-dependency, and multiple-dependency systems are further introduced. Second, we categorize the performings of existing spectral–spatial methods into four paradigms according to the different fusion stages wherein spatial information takes effect, i.e., preprocessing-based, integrated, postprocessing-based, and hybrid classifications. Then, typical methodologies are outlined. Finally, several representative spectral–spatial classification methods are applied on real-world hyperspectral data in our experiments.**

***Index Terms*— Hyperspectral image, neighborhood information, performing paradigm, spatial dependency system, spectral–spatial classification.**

Manuscript received May 11, 2017; revised September 4, 2017; accepted October 4, 2017. Date of publication November 8, 2017; date of current version February 27, 2018. This work was supported in part by the National Natural Science Foundations of China under Grant 61771496 and Grant 61571195, and in part by the Guangdong Provincial Natural Science Foundation under Grant 2016A030313254, Grant 2016A030313516, and Grant 2017A030313382. *(Corresponding authors: Lin He; Jun Li.)*

L. He is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: helin@scut.edu.cn).

J. Li and C. Liu are with the Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: lijun48@mail.sysu.edu.cn).

S. Li is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: shutao_li@hnu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2017.2765364

I. Context and Background

HYPERSPECTRAL images capture the spectral behavior of every pixel within observed scenes at hundreds of continuous and narrow bands, which offers great potentials for the subsequent information extracting [1]–[4]. As one of the major tasks of hyperspectral image processing, classification has drawn broad attentions and brought about a wide variety of methods since it arose, which aims at assigning a pixel (or a spectrum) to one of a certain set of classes [2], [4]–[7].

In the literature, many of the methods have concentrated on exploring the role of the spectral signatures of hyperspectral data in classification, employing exclusively the spectrum of a pixel to determine its class belonging. For such pixel-wise approaches with advantages of relative conceptual simplicity and implement ease, there are, however, two main dilemmas: 1) the relatively small training set versus the high-dimensional spectra and 2) spectral variabilities. The first issue, which has been often discussed from the point of view of the famous Hughes phenomenon [8], [9], actually brings about problems in two aspects. First, the small number of labeled samples likely leads to the singularity of sample covariance matrix, which results in ill-posed problems for some classification methods. Second, high-dimensional spectra result in many free parameters in a parametric model to be estimated, which is prone to the problem of overfitting and thus reduces the classifiers' generalization ability. With respect to the spectral variability, which is brought by many factors such as incident illumination, atmospheric effects, unwanted shade and shadow, natural spectrum variation, and instrument noises [2], [4], [10], there raise two main difficulties to hinder classification. On the one hand, high intraclass spectrum variability poses the identification of a given class very difficult. On the other hand, low interclass spectral variability makes the discriminating of different classes hard. All such difficulties make hyperspectral classification challenging and lead to unsatisfactory classification performances with pixel-wise methods.

As hyperspectral images are originally 3-D and pictorial, spatial dependency, which is complementary to spectral behavior, is naturally another useful information source in addition to spectra. The introduction of spatial dependency hence offers the possibility to boost the pixel-wise classification. Early attempt of incorporating spatial information into hyperspectral classification can trace back over a decade and some successful studies have shown its ability to enhance the classification

performance [5]–[7], [11]–[16]. Since then, spectral–spatial classification has witnessed a great surge of interest.

In this paper, we focus our work on the spectral–spatial classification methodologies for hyperspectral images, which is dedicated to providing a relatively general and comprehensive overview on the existing methods. The motivation of our work lies on two aspects. On the one hand, we expect this paper to clarify the mechanism behind existing spectral–spatial classification methods. There are many spectral–spatial methods, which are built from different kinds of viewpoints. Our work aims to analyze them in unified contexts and unveil some relevant mechanisms and principles underlying. On the other hand, our work intends to provide some guidelines for the design of new spectral–spatial approaches. Specifically, the research objectives and contributions can be identified as follows.

1) A new concept called spatial dependency system is developed to specify the spatial considerations functioning in spectral–spatial classification. In such a system, pixel dependency and label dependency reflect the entities carrying the spatial dependency, whereas neighborhood covering and neighborhood importance are ways to embed spatial information into the classification procedure. Such a system, which characterizes the spatial factors involved in spectral–spatial processing, can help to gain insight into how spatial information takes effect and to distinguish different kinds of methods. Although there already exist review publications on spectral–spatial classification, to the best of our knowledge, a similar concept of spatial dependency system has not occurred.
2) Based on the manner how the spatial neighborhood is determined, we organize spatial dependency systems into three major categories: fixed, adaptive, and global systems. Such a category organization of fixed, adaptive, and global spatial dependency systems has never been developed in existing reviews.
3) Based on the number of entities and the number of spatial dependencies, we further introduce the categorization of single-dependency, bilayer-dependency, and multiple-dependency systems. This categorization of single-dependency systems, bilayer-dependency systems, and multiple-dependency systems has never proposed in existing reviews.
4) In terms of the stage where spatial information takes effects, we group spectral–spatial methods into three essential paradigms, including preprocessing-based, integrated, and postprocessing-based classification approaches, and a more complex one, hybrid classification, with a comprehensive analysis. The entire set of these four paradigms, which can accommodate almost all existing spectral–spatial classification methods, has never arisen as a whole, although part of these threads have been mentioned in previous review publications, for example, postprocessing thread in [4] (Section IV-C).
5) Typical methodologies in the context of spectral–spatial classification are outlined, including structural filtering, morphological profile (MP), random field, sparse representation-based classification (SRC), and object (segmentation)-based classification.
6) Experimental results from the state-of-the-art methods on four real hyperspectral data are demonstrated.

The remainder of this paper is organized as follows. Section II introduces the spatial dependency systems. In Section III, the general performing threads for spectral–spatial classification and the relevant principles are reviewed. Section IV presents several typical methodologies. Section V demonstrates experiment-based evaluations of various spectral-only and spatial–spectral methods on real-world data sets. Finally, a summary is provided in Section VI, wherein we present some important issues, considerations, and suggestions for future spectral–spatial classification design.

II. Spatial Dependency Systems

Before we present the spatial dependency system, we first introduce the concept of spatial dependency. Herein, we refer *spatial dependency* to the correlation among pixels spatially related, according to the *Tobler's first law of geography*, in which the similarity between two objects on the same geography surface has an inverse relationship with their distance [17], defying the assumption that a hyperspectral image is made up of a collection of independent pixels who do not interact with one another. The spatially related pixels are then called *neighboring* pixels, and all those neighboring pixels are in the same *neighborhood*. Neighborhood resides in one of cores in the concept of the spatial dependency system.

A. Definition of Spatial Dependency System

In the context of spectral–spatial hyperspectral classification, two identities, pixel and the associated label, can probably carry spatial dependency information. In consequence, we associate a spatial dependency system with *pixel dependency* and *label dependency*.

1) *Pixel dependency* refers to the correlation of the neighboring pixels (refer to the pixels' spectral signatures or features), based on the fact that, in the real world, neighboring pixels in a hyperspectral image usually follow certain spatial structures or patterns. These spatial structures can occur in various forms.
2) *Label dependency* refers to the correlation of the labels of neighboring pixels, which generally arise in the manner of label consistency or label smoothness.

Accordingly, a spectral–spatial method actually refers to that where at least one of the above two spatial dependencies is utilized to determine the pixels' class-belongings, assuming that neighboring pixels or/and their labels would hold a certain kind of structure based on the introduction of the spatial information. The actions of pixel dependency and label dependency involve two factors, including *neighborhood covering* and *neighborhood importance*.

1) *Neighborhood covering* refers to the set of neighboring pixels, which stands for the spatial covering scope of the neighborhood.

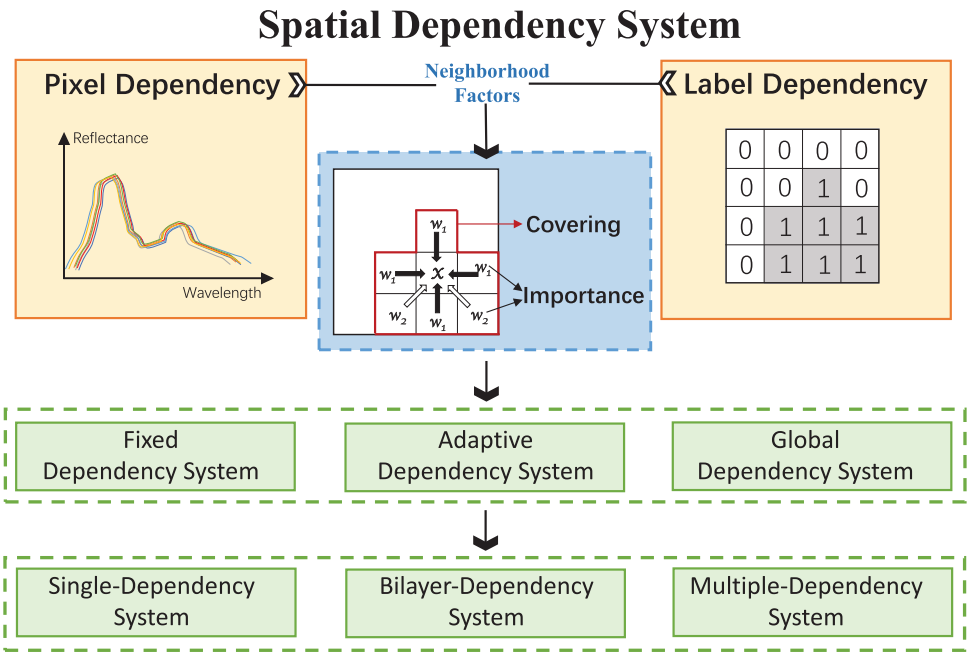


Fig. 1. Spatial dependency system in spectral–spatial classification.

2) *Neighborhood importance* denotes the degree of the contribution of every neighboring pixel in the neighborhood of a given pixel to determining the class belonging of that pixel.

To accomplish a spectral–spatial classification procedure, both neighborhood covering and neighborhood importance always have to be implicitly or explicitly set and utilized.

Under the present setup, the spectral–spatial classification will be analyzed in detail. *More formally, in this paper, a spatial dependency system is defined as an aggregate of pixel/label entity(ies) carrying spatial information and the connected neighborhood that has properties of covering and importance.* Fig. 1 illustrates the proposed spatial dependency systems. As shown in Fig. 1, the spatial information is nested in the system through the neighborhood covering and the neighborhood importance under label dependency or/and pixel dependency.

B. Fixed, Adaptive, and Global Dependency Systems

According to the design that how the neighborhood information is exploited, we can categorize the dependency systems into three kinds: fixed, adaptive, and global dependency systems.

1) Fixed Dependency Systems: In a fixed dependency system, for any pixel in the image, its neighboring pixels and their connected importance are fixed. There are many existing works following the fixed dependency system. For instance, predetermined masks or filters were used to extract spectral–spatial co-occurrence, wavelet, and Gabor features, respectively [18]–[28]. Camps-Valls *et al.* [16] constructed composite kernels to extract spatial moments in the kernel space. Similar developments based on composite kernel were presented in [29]–[32]. In [19] and [33]–[35], a fixed neighborhood is introduced into SRC, while in [36], the strategy of a sequence of fixed neighborhoods with different sizes for that procedure is employed. In [37] and [38], fixed neighborhoods are used to directly modify the objective of standard support vector machine (SVM). Guo *et al.* [39] developed a tensor SVM where fixed neighborhoods of pixels are used as input tensors.

Another representative collection of fixed dependency instantiations is related to MP-based methods [6], [13]–[15], [40], [41], which utilized mathematical morphology operators built on preset structural elements to obtain extended morphological MPs (EMAPs) that are used as features for classification purposes; lately, [42]–[50] introduced additional operations such as thresholding to develop more sophisticated morphological attribute profiles (APs).

Besides the above methods with fixed pixel dependency, there are a few methods with fixed label dependency. Typical group in this aspect includes Markov random field (MRF)-based methods such as [51]–[57] and conditional random field (CRF) [58], [59], where fixed label masks were used to apply label smooth constraint. In addition, [60] used labels within given neighborhoods to extract relation features combining spatial information.

2) Adaptive Dependency Systems: In this category, the spatial dependency system used by a classification method contains neighborhood covering or neighborhood importance with adaptive behavior across the hyperspectral image. Such a capacity tends to make spatial dependency adapt the specific characteristic of an image and is explicitly data dependent.

Most classification methods under these kinds of systems adopted a certain image sensitive factor to tune the importance of pixels in a neighborhood, such as [61]–[64]. Especially, [65]–[67] wrapped edge sensitive factors to devise

TABLE I
SPATIAL DEPENDENCY SYSTEMS IN EXISTING WORKS

Related works	Pixel Dependency	Label Dependency
Fixed Dependency	Plaza *et al.* [13], [14], Benediktsson *et al.* [15], Camps-Valls *et al.* [16], Tsai *et al.* [18], He *et al.* [19], Qian *et al.* [20], Ye *et al.* [21], Velasco-Forero *et al.* [22], Bau *et al.* [23], He *et al.* [24], Shen *et al.* [25], Tang *et al.* [26], Rajadell *et al.* [27], Zhong *et al.* [28], Liu *et al.* [29], Gurram *et al.* [30], Zhou *et al.* [31], Li *et al.* [32], Chen *et al.* [33], [34], Soltani-Farani *et al.* [35], Fang *et al.* [36], Bruzzone *et al.* [37], Li *et al.* [38], Fauvel *et al.* [40], Gu *et al.* [41], Mura *et al.* [42], Ghamisi *et al.* [43], Xia *et al.* [44], Falco *et al.* [45], Song *et al.* [46], Liao *et al.* [47], Demir *et al.* [48], Li *et al.* [50], Guo *et al.* [39], He *et al.* [96]	Li *et al.* [51], Xia *et al.* [52], Bai *et al.* [54], Khodadadzadeh *et al.* [53], Li *et al.* [55], Wang *et al.* [56], Tarabalka *et al.* [57], Zhong *et al.* [58], [59], Guccione *et al.* [60]
Adaptive Dependency	Sun *et al.* [61], Kang *et al.* [62], Zhao *et al.* [63], Zhong *et al.* [64], Kang *et al.* [65], Xia *et al.* [66], Lunga *et al.* [67], Zhao *et al.* [68], Chen *et al.* [69], Kang *et al.* [70], Peng *et al.* [73], Zhang *et al.* [74], Fang *et al.* [75], Roscher *et al.* [76], Li *et al.* [77], Lu *et al.* [78], Valero *et al.* [82], Lu *et al.* [79], Veganzones *et al.* [83], Xue *et al.* [84], Phillips *et al.* [97], Huang *et al.* [98], Ji *et al.* [71], Liu *et al.* [72]	Zhong [64], Tarabalka *et al.* [88], [90], Ghamisi *et al.* [89], Zhong *et al.* [91], Zhang *et al.* [93], [92], Golipour *et al.* [99], Zhang *et al.* [100],Li *et al.* [85], Yuan *et al.* [86], Li *et al.* [87]
Global Dependency	Zhong *et al.* [58], [59], [94], Fang *et al.* [101], Li *et al.* [95]	Zhong *et al.* [94]

edge-preserving classifications. References [68] and [69] used convolutional neural networks (CNNs) under the feedback of classification output to adjust the neighborhood contributions to spectral–spatial feature. References [70]–[72] encoded the spatially local pixel differences into the Laplacian matrix of the image to make neighborhood adaptively work.

Some methods adopted dependency systems with adaptive neighborhood covering. A typical group of methods are superpixel/object/segmentation-based ones [73]–[81] where pixels within objects were collaboratively utilized for classification. Superpixel, object, and segmentation arose in some literatures as notions relevant to object-oriented methods, which are referred to as homogeneous subregions of the entire image normally but are different from the small scale to the large scale. As other examples involving pixels within adaptive covering, [82] and [83] used sequential operations of region merging and pruning on binary partition trees to adaptively mine the spatial relationships of pixels; [84] used shape-adaptive neighborhood achieved by an anisotropic local polynomial approximation (LPA) intersection of confidence intervals (ICI) strategy to detect spatial pixel dependency and applied it to spatial graph embedding; [85], [86] utilized the correlations between the central pixel and pixels in a fixed spatial window to adaptively set neighborhood covering and construct corresponding joint sparse/collaborative representations, while [87] used the central pixels as thresholds to generate adaptive neighborhood covering and accordingly get local binary patterns.

There are also methods operating adaptive spatial dependency in label level. Some typical examples are object-based methods such as [88]–[91], which used the votes of labels within an object yielded by segmentation to determine the class-belonging of the object, and MRF-based methods such as [92] and [93], which utilized the pixel variation sensitive indices to adjust the *a priori* distribution of image labels.

3) Global Dependency Systems: In this group, the utilization of spatial dependency is explicitly based on the pixels or labels of the whole image without the assumption of independence, while in fact such a global dependency usually contains implicit local relationships.

CRF-based methods can naturally accommodate global pixel dependency because its formulation involves the whole image while not imposes independence constraints [58], [59], [94]. In [58], [59], and [94], unary and higher order potentials of a CRF demand the join of the whole data. Another strategy for global pixel dependency is to use maximum a posterior model marginal (MAPM) where labels conditioning on the whole data are concerned. In [95], a Bayesian loopy propagation was used to infer MAPM.

In addition to global pixel dependency cases, there are the examples of global label dependency such as [94]. In this paper, besides involving the whole data, potentials in higher order than unary potentials and pairwise potentials are introduced to deal with high-order label dependency, which in principle can be extended up to the labels of all pixels.

Existing spectral–spatial classification methods built on fixed, adaptive, and global dependency systems are summarized in Table I.

C. Single-, Bilayer-, and Multiple-Dependency Systems

It should be noted that, in some cases, more than one spatial dependency is used to in-deep characterize the spatial information. Therefore, according to the number of the dependencies and the number of the entities carrying dependencies utilized, we can organize the dependency system strategies into the following groups.

1) Single-Dependency Systems: For the case of a single-dependency system, only one dependency is used in classification. Obviously, most spatial dependency systems employed

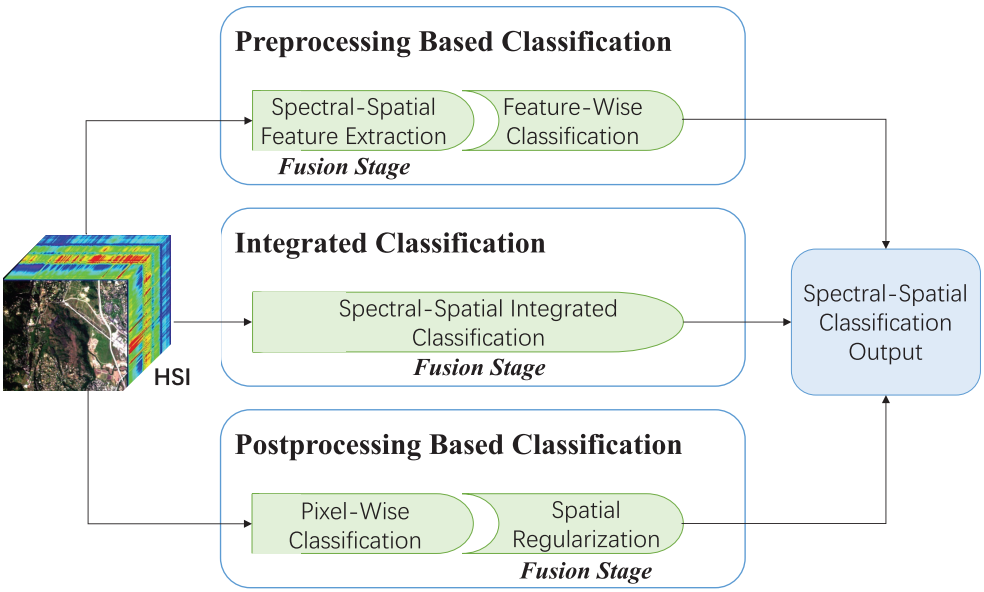


Fig. 2. Paradigms of preprocessing-based, integrated, and postprocessing-based spectral–spatial classifications.

in existing classification methods belong to this group because of its relative simplicity.

2) Bilayer-Dependency Systems: As mentioned, the pixel and label are two possible entities carrying spatial dependencies. In addition to the methods that involve the dependency either in pixel level or in label level, there are those that utilize both pixel dependency and label dependency simultaneously. Both MRF and CRF provide potential frameworks to bind pixel dependency and label dependency together. In [58], [59], and [94], the unary potential of CRF was built with global pixel relationship, while interaction potentials were modeled on label dependency as well as pixel interplaying. References [64] and [100] developed CRF variants where contextual information refined from the observed pixels was used to adjust the label smooth regularization term. In [92], [93], and [99], the traditional MRF formulating pairwise label relations were tuned by the contextual information built on neighboring pixels. Another strategy using both dependencies is based on superpixel, such as [74], where the unary potential of an MRF was constructed with the class-belonging probability of a superpixel obtained by averaging pixel-wise class belongings, and the pairwise potential was from the Potts model on labels of superpixels.

3) Multiple-Dependency Systems: In this group, more than one dependency is required. But such dependencies are in a single level, pixel level, or label level, which is different from bilayer dependency wherein different dependencies occur in both pixel and label levels simultaneously. Typical members include [74] and [78] where two spatial structures were used, one associated with spatial average operation to get the class belongings of adaptive areas, while the other involving the edge sensitive detection to adjust the label smoothness. Other examples include [85], which incorporated in-neighborhood correlations into joint sparsity model supposing in-neighborhood pixels share the same subspace.

III. Paradigms of Spectral–Spatial Classification

Under the definition of spatial dependency systems, in a spectral–spatial method, the spatial information can be plugged in via different paradigms. In terms of the stage where the spatial dependency takes effect or where the spectral information and the spatial information are fused, the paradigms of spatial–spectral classification methods can be approximately divided into three categories: preprocessing-based classification, postprocessing-based classification, and integrated classification (as shown in Fig. 2). For each paradigm, there is a spectral–spatial fusion stage. As suggested in Fig. 2, the fusion stage is eventually moving toward the tail of the whole classification procedure from preprocessing-based classification to postprocessing-based classification. It is noteworthy that apart from such three basic paradigms, there exists hybrid classification that is the mixture of elements from more than one basic paradigm and will also be discussed later.

A. Preprocessing-Based Classification

In preprocessing-based methods, the spatial information is pre-extracted in forms of spatial features, which is then pushed into the subsequent classifier. The classification process is typically divided into two phases: 1) spectral–spatial feature extraction that aims at extracting the representative features contained in the original hyperspectral image and 2) feature-wise classification performed using standard classifiers such as SVM and multinomial logistic regression (MLR). Obviously, the feature extraction step, which incorporates the spatial information, is the key to the performance of these techniques [102].

For instance, extended MPs constructed by mathematical morphology operations with structural elements in various sizes [6], [13]–[15], [40] and morphological APs obtained by more sophisticated morphological operations [42]–[45] are

the most remarkable ones for their excellent spatial features. He *et al.* [19] employed a spatial translation-invariant spline wavelet transform to extract spectral–spatial features that were then serialized with a linear programming (LP) SRC. References [20]–[22] also introduced wavelet features; specifically, [22] used a wavelet soft-shrinkage denoising strategy. References [23]–[28] used high-dimensional Gaussian enveloped harmonic to extract Gabor features. Specially, [24] designed a low-rank Gabor filtering that is not only computational efficient but also tailored for the special characteristics of hyperspectral images (HSIs). References [18] and [87] used modified co-occurrence matrix and local binary pattern to obtain texture features, respectively. In [96] and [103]–[105], empirical mode decomposition (EMD) or singular spectrum analysis was used for feature extraction. Reference [60] devised features by involving a spatial-dependent homogeneity index that was fed in a classifier that was collaborated with another pixel-wise classifier to form a semisupervised learning. There are also spectral–spatial preprocessings realized in kernel space, usually in the manner of the composite kernel, including those mentioned in [16] and [29]–[32] with fixed spatial dependency and those in [73] and [101] with adaptive spatial dependency.

B. Integrated Classification

In the integrated methods, the spectral and spatial information is utilized simultaneously to form an integrated classifier. That means the exploitation of spectral and spatial information is not explicitly separated in the classification process.

As typical examples of such a group, Bruzzone and Persello [37] and Li *et al.* [38] used in-neighborhood contextual information to alter the objective and constraints of the traditional pixel-wise SVM classifier. Reference [39] also recasts the pixel-wise SVM, but with the tensor strategy, which can preserve the original data structure. Chen *et al.* [33], Fang *et al.* [36], and Li *et al.* [85] constructed a training sample-based collaborative dictionary and subsequently introduced contextual regression signals (pixels) to build a joint SRC/collaborative representation-based classification (CRC). References [82] and [83] used sequential binary partition tree-based region merging and pruning to segment hyperspectral images. Reference [95] introduced loopy belief propagation (LBP) strategy to infer MAPM, which directly determined the output of spectral–spatial classification. Zhong and Wang [58] employed a Gibbs distribution to represent the joint *a posteriori* probability to build a CRF for hyperspectral image classification. It is noteworthy that, since the *a posteriori* of pixel labels is directly formulated with stochastic distribution without intermediate processes in the original CRF, we organize the CRF-based methods into integrated classification categorization. References [68] and [69] used CNN strategy to fulfill the classification where feature layer and classification layer were trained jointly and the training was usually not disposed separately.

C. Postprocessing-Based Classification

In postprocessing-based methods, a pixel-wise classification is usually conducted in the first step, and then the spatial dependency is used to regularize the preobtained classification results.

In [51], an MLR was used to obtain the pixel-wise *posteriori* classification output first; then, an MRF regularizer connected to the *a priori* probability of the labels was serialized, which was solved by the α-expansion or the LBP strategies, to yield the final spectral–spatial *posteriori* classification. Similar developments based on postprocessing with the MRF were presented in [55], [57], [92], [93], and [106]. References [55] and [57] alternatively used probabilistic SVM and a Gaussian mixture classifier to model the spectral information, respectively. References [92] and [93] introduced a homogeneity index and a total variation regularization to adaptively adjust the weight of spatial contribution, respectively. It is noteworthy that [74] and [78] also carried out similar MRF-based postprocessing, but their work was in the superpixel level. In [88]–[91], pixel-wise classification output was obtained first, and then the strategy of majority voting of labels within homogenous subregions was used to achieve the final spectral–spatial classification output. Kang *et al.* [65] utilized SVM to obtain preliminary class-belonging probability, and then used guiding images that reflect the homogeneity disruption to guide an edge-preserving classification. Reference [72] used a kernel CRC to get pixel-wise class-belonging probabilities; the class-belonging probabilities were then estimated by a regression regularized by a term involving an adaptive weighted graph.

While the aforementioned categorized only some representative methods, a wider variety of existing spectral–spatial classification methods in terms of preprocessing-based, integrated, and postprocessing-based paradigms are presented in Table II.

D. Hybrid Classification

Based on the three essential paradigms mentioned above, there derives hybrid classification. In the hybrid classification case, a classifier comes from the coupling of factors of more than one basic paradigm. Reference [74] first used a postprocessing on pixel-wise class-belonging output to get superpixel class belonging and then introduced a postregularization on superpixel class belonging to obtain final classification output. The method embedded two scales of postprocessings, and thus two scales of spatial consistencies were concerned: one in pixel level and the other in superpixel level. Similar examples include [78]. References [64] and [100] used the pre-extracted features to tune the CRFs on HSIs, while [92], [93], and [99] utilized such features to adjust the postregularization terms of MRFs. Such methods are obviously examples that combine a preprocessing-based paradigm with another integrated/postprocessing-based paradigm.

E. Analysis of Spectral–Spatial Classification Paradigms

With different spectral–spatial classification paradigms adopted, there are different mechanisms for them to take effect. We explore the related principles as follows.

In preprocessing-based classification, the original observation space or feature space is transformed to another feature

TABLE II
TYPICAL METHODS OF PREPROCESSING-BASED, INTEGRATED, AND POSTPROCESSING-BASED SPECTRAL–SPATIAL CLASSIFICATIONS

	Typical Methods
Preprocessing Based Classification	Plaza *et al.* [13], [14], Benediktsson *et al.* [15], Camps-Valls *et al.* [16], Tsai *et al.* [18], He *et al.* [19], Qian *et al.* [20], Ye *et al.* [21], Velasco-Forero *et al.* [22], Bau *et al.* [23], He *et al.* [24], Shen *et al.* [25], Tang *et al.* [26], Rajadell *et al.* [27], Zhong *et al.* [28], Liu *et al.* [29],Gurram *et al.* [30], Zhou *et al.* [31],Li *et al.* [32], Fauvel *et al.* [6], [40], Gu *et al.* [41], Mura *et al.* [42], [44], Ghamisi *et al.* [43], Falco *et al.* [45], Song *et al.* [46], Liao *et al.* [47], Demir *et al.* [48], [103], Li *et al.* [50], Guccione *et al.* [60], Kang *et al.* [62], Xia *et al.* [66], Lunga [67], Zhao *et al.* [68], Xue *et al.* [84], Li *et al.* [87], He *et al.* [96], Phillips *et al.* [97], Huang *et al.* [98], Fang *et al.* [101], He *et al.* [104], Zabalza *et al.* [105]
Integrated Classification	Peng *et al.* [73], Chen *et al.* [33], [34], Soltani-Farani *et al.* [35], Fang *et al.* [36], Bruzzone *et al.* [37], Li *et al.* [38], Guo *et al.* [39], Zhong *et al.* [58], [59], Sun *et al.* [61], Zhao *et al.* [63], Zhong *et al.* [64], Chen *et al.* [69], Ji *et al.* [71], Fang *et al.* [75], Roscher *et al.* [76], Li *et al.* [77], [85], Lu *et al.* [79], Valero *et al.* [82], Veganzones *et al.* [83], Yuan *et al.* [86], Zhong *et al.* [94], Li *et al.* [95], Zhang *et al.* [100]
Postprocessing Based Classification	Li *et al.* [51], Xia *et al.* [52], Khodadadzadeh *et al.* [53], Bai *et al.* [54], Wang *et al.* [56], Li *et al.* [55], Tarabalka *et al.* [57], [88], Kang *et al.* [65], [70], Liu *et al.* [72], Zhang *et al.* [74], Lu *et al.* [78], Ghamisi *et al.* [89], Tarabalka *et al.* [90], Zhong *et al.* [91], Sun *et al.* [92], Zhang *et al.* [93], Golipour *et al.* [99], Huang *et al.* [106]

space thatt is spanned by the spectral–spatial features. Suppose that the feature dimensionalities of such two spaces remain unchanged. From the statistical and probabilistic perspective, that more information is used in the spectral–spatial feature space leads to less uncertainty in that domain, implying the likelihood function of a parametric model becomes sharper [107], [108]. This will yield influences in two aspects. On the one hand, the average curvatures of class-conditional data distributions increase [107]–[109], implying that it is expected to achieve a higher estimation accuracy of the parameters for a classifier, which is often quantitatively measured by Cramer–Rao lower bound [109], [110]. On the other hand, the overlaps of various class-conditional data distributions tend to reduce, thus decreasing the risk of error decisions [111], [112]. All the two points play positive roles in enhancing the discriminability of a classifier.

In integrated classification, the optimization objective and constraints connected with classification are directly modified in the same feature space by introducing spatial dependency, avoiding the usage of any intermediate processes. Normally, the optimization range and the feasible region that are determined by the objective and the constraints become more valid and restricted, making it easier to get a better estimate of the parameters for a classifier. Thus, a better classification result is received.

In postprocessing classification, a preliminary pixel-wise classification is adjusted with a postregularization to yield final spectral–spatial classification results. In the Bayesian framework, the postregularization term can be viewed as the term carrying a certain kind of prior information about the image that contains the spatial dependency. Through an appropriate postregularization, useful information in the image level is fused to yield image-wise *a posteriori* distributions, which sharpen the pixel-wise *a posteriori* distributions achieved by preliminary pixel-wise classification. Such sharper *a posteriori* distributions offer the potential to achieve better classification results.

In hybrid classification, several spatial dependencies are incorporated into different classification phases. This spatial information complementarily collaborates to implement the classification.

IV. TYPICAL METHODOLOGIES

In this section, we present several typical methodologies of spectral–spatial classification for hyperspectral images.

A. Structural Filtering

Structural filtering-based feature extraction is one of the most studied approaches for spatial preprocessing of hyperspectral image. In most cases, the contextual features are obtained via structural filtering. That is, given a fixed (or an adaptive) structural element, the hyperspectral data are then filtered such that we can obtain the spatial features. There are many researches going on this direction. We will group those approaches in the following lists.

1) One of the most simple but powerful ways is to extract the spatial information from a given region based on moment criteria, for instance, the mean or standard deviation (STD) of the neighborhood pixels in a window [16]. The inclusion of moments or cumulants is widely studied in the area of composite kernel learning [31], [32] and multiple kernel learning [29], where the features are pre-extracted and then used to build the kernel. Adaptations of the neighboring moments or cumulants are widely discussed for the classification of hyperspectral images in order to extract the spatial homogeneity, which at the same time remain the details [98], [113], [114]. For example, bilateral filter (edge-preserving filter) effectively removes noises and small details while preserves large-scale structures automatically [65], [66].
2) For structural filtering, given a fixed window, another research direction is to perform local harmonic analysis. The related works include spatial translation-invariant spectral–spatial wavelet features developed in [26]; spectral–spatial Gabor features used in [25]–[27]; and EMD-based features devised in [105].

3) A trend, in spatial filtering, is to use adaptive structures to extract features [97], such as adaptive multidimensional Wiener filtering [115] and adaptive region-based filtering [73]. Recently, superpixel-based filtering is used to characterize the adaptive spatial information [78], [101].

B. Morphological Profile

EMAPs, as the new generation of MPs, are an extension of APs via morphological operators using different kinds of structure elements, which are, therefore, special cases of structure filtering, obtained using different types of attributes and stacked together. The filtering operation implemented in EMAPs is based on the evaluation of how a given attribute $\mathcal{A}$ is computed for every connected components of a grayscale image f for a given reference value λ. For a connected component of the image, C_i, if the attribute meets a predefined condition [e.g., $\mathcal{A}(C_i) > \lambda$], then the region is kept unaltered; otherwise, it is set to the grayscale value of the adjacent region with a closer value, thereby merging C_i to a surrounding connected component. When the region is merged to the adjacent region of a lower (or greater) gray level, the operation performed is a thinning (or thickening). Given an ordered sequence of thresholds $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, an AP is obtained by applying a sequence of attribute thinning and attribute thickening operations as follows:

$$\mathrm{AP}(f) := \{\phi_n(f), \ldots, \phi_1(f), f, \gamma_1(f), \ldots, \gamma_n(f)\} \quad (1)$$

where ϕ_i and γ_i, respectively, denote the thickening and thinning transformations. Problem (1) focuses on a single feature (or spectral band) of panchromatic data. For hyperspectral images, we can perform attribute filtering on the full original data. However, on the one hand, hyperspectral data are very high dimensional, with a large number of spectral bands. This means that constructing the EAP on the original spectral bands would lead to very significant computational complexity. On the other hand, it is generally observed that the hyperspectral data live in a lower dimensional subspace than the original spectral space. As a result, we can perform dimensionality reduction using techniques such as principal component analysis (PCA) [116], and then perform attribute filtering on the first few principal components (PCs), as suggested in [117], in order to reduce computational complexity. For multispectral data, since there are only a few spectral bands available, we can perform attribute filtering on the full original spectral data. In this way, the EAP is obtained by generating an AP on each of the first q PCs (or any other features retained after applying feature selection on the multispectral/hyperspectral image) thus building a stacked vector using the AP on each feature. This leads to the following definition of the EAP for the pixel $\mathbf{x}_i$:

$$\mathrm{EAP} := \{\mathrm{AP}(f_1), \mathrm{AP}(f_2), \ldots, \mathrm{AP}(f_q)\} \quad (2)$$

where q is the number of retained features. From the EAP definition in (2), the consideration of multiple attributes leads to the concept of EMAP that combines the EAPs by concatenating them in a single vector of features and improves the capability to extract the spatial characteristics of the structures in the scene, where attribute filtering can be efficiently computed by applying a max-tree algorithm [118].

C. Random Field

Let $\mathcal{S} \equiv \{1, \ldots, n\}$ denote a set of integers indexing the n pixels of a hyperspectral image; let $\mathbf{x} \equiv (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{d \times n}$ denote such hyperspectral image made up of d-D feature vectors; and let $\mathbf{y} \equiv (y_1, \ldots, y_n)$ denote an image of labels. In random fields, we build the *a posteriori* $p(\mathbf{y}|\mathbf{x})$ of the class labels $\mathbf{y}$ given the features $\mathbf{x}$, which is the engine for the class label inference. Following a discriminative approach, we model the distribution $p(\mathbf{y}|\mathbf{x})$ directly, instead of the joint distribution $p(\mathbf{y}, \mathbf{x})$, which quite often implies simplistic assumptions about the data generation mechanism. Furthermore, because the discriminative models are less complex than the corresponding generative counterparts, learning the former models often yields better results than learning the latter ones, namely, when the training data are limited. Therefore, the *a posteriori*[1] can be obtained using the following model:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\boldsymbol{\omega}, \mathbf{x})} \exp\left(\sum_{i \in \mathcal{S}} \log p(y_i|\mathbf{x}_i, \boldsymbol{\omega}) + \mu \sum_{(i,j) \in \mathcal{C}} \delta(y_i - y_j)\right) \quad (3)$$

where $\boldsymbol{\omega}$ is the learning parameter and $Z(\boldsymbol{\omega}, \mathbf{x})$ is the normalizing factor, also known as the partition function, which is

$$Z(\boldsymbol{\omega}, \mathbf{x}) = \sum_{\mathbf{y}} \exp\left(\sum_{i \in \mathcal{S}} \log p(y_i|\mathbf{x}_i, \boldsymbol{\omega}) + \mu \sum_{(i,j) \in \mathcal{C}} \delta(y_i - y_j)\right) \quad (4)$$

where $p(y_i|\mathbf{x}_i, \boldsymbol{\omega})$ denotes the class probability given by the learning parameter $\boldsymbol{\omega}$, μ is a parameter controlling the degree of smoothness on the image of labels, $\delta(y)$ is the unit impulse function,[2] and $\mathcal{C}$ is a set of cliques.[3] In the discriminative model (3), the term $p(y_i|\mathbf{x}_i, \boldsymbol{\omega})$ is itself a discriminative classifier, which gives the probability of label y_i given the feature vector $\mathbf{x}_i$, and the pairwise interaction term $\mu\,\delta(y_i - y_j)$ encodes spatial contextual information by attaching higher probability to equal neighboring labels than the other way around. Therefore, the term $\mu\,\delta(y_i - y_j)$ promotes piecewise smooth labelings, where μ controls the degree of smoothness.

It is notable that the *a posteriori* (3) is a particular case of a *discriminative random field* (DRF) [119] with *association potentials* given by $\log p(y_i|\mathbf{x}_i, \boldsymbol{\omega})$ and *interaction potentials* given by $\mu\,\delta(y_i - y_j)$. The DRFs, based on the concepts of *CRFs* [120], are, in a sense, a generalization of the *MRFs*.

Random fields have been widely studied for hyperspectral classification. On the one hand, MRF is favored due to the fact that the class parameters are usually estimated independently

[1] To keep the notation simple, we use $p(\cdot)$ to denote both continuous densities and discrete distributions of random variables. The meaning should be clear from the context.

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

[3] A clique is a set of labels that are neighbors of each other.

from the field parameters [51], [53], [55], [92], [121], [122], which can greatly simplify the computational complexity. On the other hand, CRFs, as a probabilistic framework for labeling and segmenting structured data, achieved very good performance in the exploration of spatial information [58], [59], [64], [94], [123]–[126]. CRFs define a conditional probability distribution, instead of a joint distribution, over the label sequences given by the observed hyperspectral data. The advantages of CRFs in comparison with MRFs are twofold. The first one is that the conditional nature of CRFs leads to the relaxation of the independence assumptions. The other one is that CRFs avoid the label bias problem, which is a weakness of MRF models based on directed graphical models. Besides, DRFs, having drawn broad attention in recent years [95], offer several advantages, namely: 1) the relaxation of conditional independence of the observed data; 2) the exploitation of probabilistic discriminative models instead of the generative MRFs; and 3) the simultaneous estimation of all DRF parameters from the training data.

D. Sparse Representation-Based Classification

In sparse representation models, the idea is that all the test samples can be represented by a (sparse) linear combination of atoms from an overcomplete training dictionary. Let us assume that we have a training dictionary, denoted by $\mathbf{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathbb{R}^{n \times l}$, with n samples of l dimensions, comprising a total of c distinct classes and that the dictionary is organized as $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_c]$, where $\mathbf{A}_k = \{\mathbf{x}_{k_1}, \ldots, \mathbf{x}_{k_{n_k}}\}$ (i.e., $\mathbf{A}_k$ holds the samples of class k in its columns, n_k is the number of samples in $\mathbf{A}_k$, and $\sum_{k=1}^{c} n_k = n$). A typical scenario in supervised classification of remote sensing data is that, normally, we have a (limited) set of labeled training samples for each class and then we use part of this information to train a classifier that is then tested with the remaining labeled samples. Let $\mathbf{x}_i$ be a test sample that can be appropriately represented by a linear combination of the atoms (training samples) in the dictionary $\mathbf{A}$ as follows:

$$\begin{aligned}\mathbf{x}_i &\approx \mathbf{x}_1\alpha_1 + \mathbf{x}_2\alpha_2 + \cdots + \mathbf{x}_n\alpha_n \\ &= [\mathbf{x}_1\mathbf{x}_2 \ldots \mathbf{x}_n][\alpha_1\alpha_2 \ldots \alpha_n]^T = \mathbf{A}\boldsymbol{\alpha} + \epsilon \end{aligned} \tag{5}$$

where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_c^T]^T$ is an n-D sparse vector (i.e., most elements of $\boldsymbol{\alpha}$ are zero), $\boldsymbol{\alpha}_i$ is the vector of regression coefficients associated with class i, and ϵ is the representation error. As in [33] and [127], the central assumption in our approach is that $\mathbf{x}_i$ is well approximated by $\mathbf{A}_i\boldsymbol{\alpha}_i$, i.e., $\boldsymbol{\alpha}_j = 0$, for $j \neq i$. Under this condition, the classification of $\mathbf{x}_i$ amounts to detect the support of $\boldsymbol{\alpha}$.

In order to obtain a sparse representation for a test sample $\mathbf{x}_i$ (which hereinafter is assumed to be obtained using an EMAP representation), we need to obtain a sparse vector $\boldsymbol{\alpha}$ satisfying $\mathbf{x}_i = \mathbf{A}\boldsymbol{\alpha} + \epsilon$. The sparse vector $\boldsymbol{\alpha}$ can be estimated by solving the following optimization problem:

$$\hat{\boldsymbol{\alpha}} = \arg\min \ \|\boldsymbol{\alpha}\|_0 \quad \text{s.t. } \mathbf{x}_i = \mathbf{A}\boldsymbol{\alpha} \tag{6}$$

where $\|\boldsymbol{\alpha}\|_0$ denotes the ℓ_0-norm that counts the nonzero components in the coefficient vector. Due to the presence of noises and possible modeling errors, the optimization (6) is often replaced by

$$\hat{\boldsymbol{\alpha}} = \arg\min \ \|\boldsymbol{\alpha}\|_0 \quad \text{s.t. } \|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2 \leq \delta \tag{7}$$

where δ is an error tolerance. The aforementioned problem is nondeterministic and NP-hard, and thus it is very difficult to solve. Recently, greedy algorithms such as basis pursuit (BP) [128] and orthogonal matching pursuit [129] have been proposed to tackle this problem. BP replaces the ℓ_0-norm with the ℓ_1-norm. Hence, the sparse vector $\boldsymbol{\alpha}$ in (7) can be obtained using ℓ_1-norm as

$$\hat{\boldsymbol{\alpha}} = \arg\min \ \|\boldsymbol{\alpha}\|_1 \quad \text{s.t. } \|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2 \leq \delta \tag{8}$$

where $\|\boldsymbol{\alpha}\|_1 = \sum_i |\alpha_i|$, for $i = 1, \ldots, n$. It should be noted that, contrary to problem (7), problem (8) is convex and can be solved using LP solvers. In fact, problem (8) is equivalent to the following unconstrained optimization problem:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_1 \tag{9}$$

where the parameter τ is a Lagrange multiplier that balances the tradeoff between the reconstruction error and the sparse solution: $\tau \to 0$ when $\epsilon \to 0$. Based on the conjecture that hyperspectral feature vectors are very likely to lie in the convex cone spanned by the atoms of the respective class, the coefficients generally hold the nonnegativity constraint $\boldsymbol{\alpha} \geq 0$. Consequently, we solve the constrained $\ell_2 - \ell_1$ optimization problem

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_1, \quad \boldsymbol{\alpha} \geq \mathbf{0}. \tag{10}$$

Problem (9) is the general form of sparse representation-based optimization for hyperspectral image classification. In order to incorporate the spatial information, there are three directions.

1) A general way to incorporate spatial information is to include a spatial weight to control the coefficients, which has flexible structure, is easy to implement, and can be precomputed

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \tau\|\mathbf{W}\boldsymbol{\alpha}\|_1, \quad \boldsymbol{\alpha} \geq \mathbf{0} \tag{11}$$

where $\mathbf{W}$ is the spatial weight matrix, which is generally built via the neighboring pixels [33], using the histogram [130], nonlocal information, neighboring pixels along with outliers [131], neighboring patch [86], and neighboring filtering [132]. More recently, superpixel-based spatial information [75] and shape adaptive neighborhood are also studied.

2) Another research trend is to include some spatial regularizer to incorporate spatial information

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{x}_i - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \tau\|\mathbf{W}\boldsymbol{\alpha}\|_1 + \eta\mathcal{S}(\boldsymbol{\alpha}), \quad \boldsymbol{\alpha} \geq \mathbf{0} \tag{12}$$

where $\mathcal{S}(\cdot)$ is a function incorporating the spatial information and η is the weight parameter controlling the sparsity of the spatial regularizer. There are many works focusing on building $\mathcal{S}(\cdot)$, such as total variation [133],

a centralized quadratic constraint [134], semilocal spatial graph regularization [135], and low-rank regularizer [136].

3) There are works considering spatial weight and regularizer at the same time to better exploit the spatial information [137], [138].

Finally, the class labels can be inferred based on the following criterion, which aims at introducing robustness with respect to the representation error ϵ and to the interclass correlation:

$$\widehat{\text{class}}(\mathbf{x}_i) = \arg \min_{j \in \{1,\ldots,c\}} \|\mathbf{x}_i - \mathbf{A}_j \boldsymbol{\alpha}_j\|_2.$$

E. Segmentation-Based Approaches

Some of hyperspectral image classification methods exploit segmentation, often as postprocessing, i.e., after a spectral-based classification has been conducted. One of the first classifiers with such a spatial postprocessing developed in the hyperspectral imaging literature was the well-known extraction and classification of homogeneous objects [9]. Another one is the strategy adopted in [90], which combines the output of a pixel-wise SVM classifier with the morphological watershed transformation [139] in order to provide a more spatially homogeneous classification. A similar strategy occurs in [140], in which the output of the SVM classifier is combined with the segmentation result provided by the recursive hierarchical segmentation algorithm.[4] These strategies can lead to much improved classification results with regard to spectral-based classification, as the segmentation-based postprocessing can lead to a better delineation of object borders as well as a refinement of spatial features that includes outlier removal and refinement of the classification results. A detailed overview of other techniques in this field is available in [6].

F. Deep Learning

Based on neural networks, deep learning, a kind of more complex hierarchical architecture simulating human brains, begins to apply in hyperspectral image classification. Compared with traditional shallow modules, it is constructed with multilayer structure, usually deeper than three layers [141]. The deep learning-based spectral–spatial classification framework can be decomposed into three major stages: 1) data input; 2) deep network constructing; and 3) classification [142]. Obviously, the core of a deep learning-based classification is the constructed deep networks, which often feature specific deep learning methods and are utilized to extract spatial information. Some deep learning models have been proven effective in handling hyperspectral classification tasks. These deep models stack the first two layers of a single shallow network, which is generally composed of three layers, layer by layer to extract more abstract features, such as the deep belief network (DBN) [143], the stacked autoencoder (SAE) [144], and the CNN [145]. More detailedly, DBNs are constructed on the basis of the layer-wise restricted Boltzmann machine (RBM). To adapt the high-dimensional characteristics of hyperspectral data, the learned features exported from a single RBM would be input to another RBM for more representative features. Analogously to DBNs, SAEs are layer-by-layer systems in which the input and hidden layers of the autoencoder work repeatedly. Besides, the deep CNNs for spectral–spatial classification consist of repetitive convolution and pooling layers, after which every input data vector can be transformed into a feature vector containing abundant spatial information.

V. EXPERIMENTS

This section is designed to illustrate the remarkable impacts of spatial information on hyperspectral image classification through a series of experiments. To evaluate and compare various spectral–spatial hyperspectral analysis algorithms, four benchmark images were used. The first and the fourth images were collected by the airborne visible infrared imaging spectrometer (AVIRIS) over Northwestern Indian and Salinas Valley, America, respectively; the third by the reflective optics imaging spectrometer system (ROSIS) over the urban area of the University of Pavia, Italy; and the rest by NASA EO-1 satellite over the Okavango Delta, Botswana. Apart from the characteristics of the data, the performance of a practical classification is strongly influenced by some other factors, such as the choice of classifier and parameter tuning. Therefore, before discussing the obtained results, we first introduce the experimental settings involved in our experiments.

In our experiments, two kinds of classifiers are utilized to assign a label to each pixel. The first one is the superior MLR via variable splitting and augmented Lagrangian algorithm that is stated in detail in [146] including the corresponding parameter settings. Another one is the probabilistic SVM algorithm that improves the original SVM by combining all pairwise comparisons to estimate the class probabilities [147]. The SVM classifier employed in our experiments is provided by the well-known library for support vector machines library.[5] Here, 10-fold cross validation was used for optimizing the parameters involved in SVM. With respect to the embedded spatial dependencies, we used EMAPs [42] and Gabor filtering [24] as feature extraction tools in the preprocessing stage. The EMAP was built using the STD and the area attributes. The Gabor filtering was constructed by sin and cos harmonics in the spectral and spatial domains, respectively. It is noteworthy that we implemented PCA for dimensionality reduction after the data went through Gabor filters, and the number of remaining PCs was determined by cross validation. Additionally, we modeled the contextual information by means of the MRF multilevel logistic prior method to argue for postprocessing strategy. In comparison, we chose the label-dependent LBP method [95] as our tested integrated strategy. Thus, for each classifier, there were totally, excluding LBP, six spectral–spatial classification strategies, i.e., original spectral features only (abbreviated as RAW), spectral and EMAP features (abbreviated as EMAP), spectral and Gabor features

[4] http://opensource.gsfc.nasa.gov/projects/HSEG/

[5] Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/

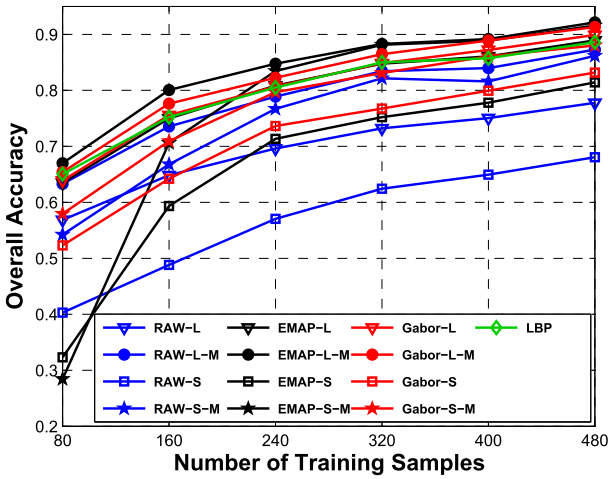


Fig. 3. OA curves (OAs versus the number of training samples) of the AVIRIS Indian Pines data set.

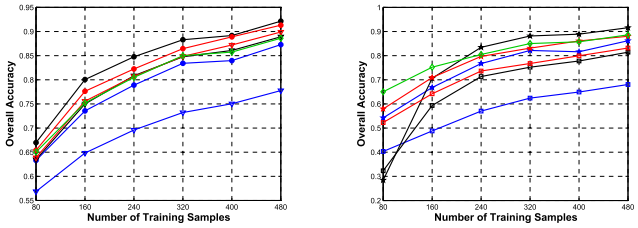


Fig. 4. OA curves (OAs versus the number of training samples) of the AVIRIS Indian Pines data set, obtained by MLR (L) and SVM (S), respectively, along with LBP. (a) OAs obtained by L and LBP. (b) OAs obtained by S and LBP.

(abbreviated as Gabor), RAW-MRF, EMAP-MRF, and Gabor-MRF. The sizes of training sets were set in the range from 5 to 30 per class with a step of five for the AVIRIS Indian Pines data set and the ROSIS Pavia University data set, while in the range from 3 to 18 per class with a step of three for the rest two data sets allowing for the quality of different data sets. If the number of samples in a certain class was less than the desired, we picked out half of them. Taking the statistical significance into account, we generated 10 different labeled sets from each data set for training purposes through independent Monte Carlo runs, and the associated averages are reported.

A. AVIRIS Indian Pines Data Set

The Indian Pines data set is composed of 220 spectral channels with a size of 145 × 145 samples, and its available ground-truth data contain 10 366 pixels with 16 classes [see Fig. 7(f)]. Fig. 3 presents all of our obtained overall accuracy (OA) results as functions of the number of training samples, where we used L, S, and M to represent MLR, SVM, and MRF, respectively (the subsequent curve Figs. 4–6, 8–11 will follow the same line specifications as Fig. 3). Obviously, compared with the OAs obtained based only on spectral features, those obtained with spatial information involved get varying degrees of improvements. However, the EMAP-S and EMAP-S-M perform badly at the beginning. It is partly due

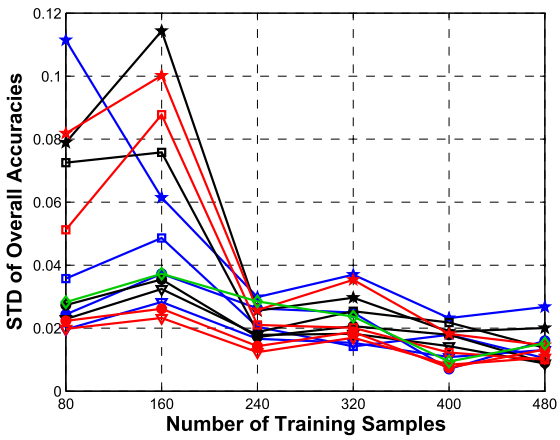


Fig. 5. STD curves (STDs of OAs versus the number of training samples) of the AVIRIS Indian Pines data set.

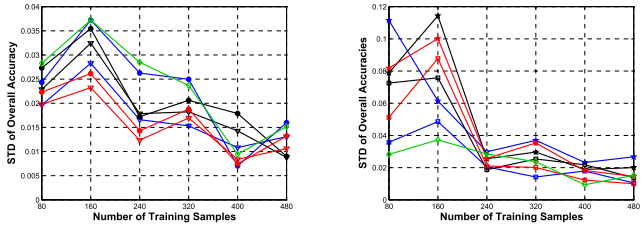
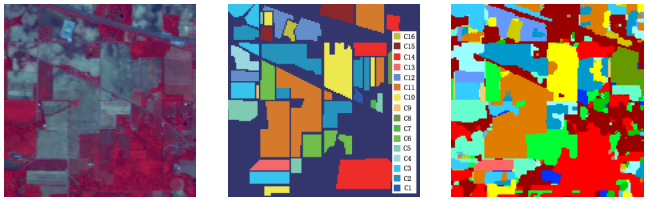


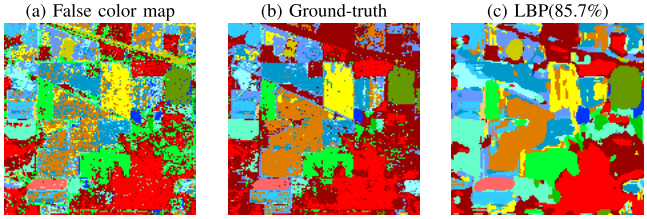
Fig. 6. STD curves (STDs of OAs versus the number of training samples) of the AVIRIS Indian Pines data set, obtained by MLR (L) and SVM (S), respectively, along with LBP. (a) STDs obtained by L and LBP. (b) STDs obtained by S and LBP.

to the SVM's poor capacity to deal with high-dimensional data with a small training set, whereas this hurdle is crossed with the increase in training samples. In this sense, MLR is more powerful. For more details, Fig. 4(a) and (b) separately lists the OA curves obtained by MLR (L) and SVM (S), along with LBP. As shown in Fig. 4(a) and (b), feature extraction tools implemented ahead of classification make a significant improvement in OAs. Furthermore, the advantages are amplified when MRF is subsequently included, which implies the accumulative effects of spatial dependencies utilized in different stages. In addition, EMAP benefits more from MRF than Gabor, in which case the OA curves of EMAP are below those of Gabor while EMAP-M above Gabor-M. Besides, LBP also makes a contribution to improving classification performances by means of global dependency.

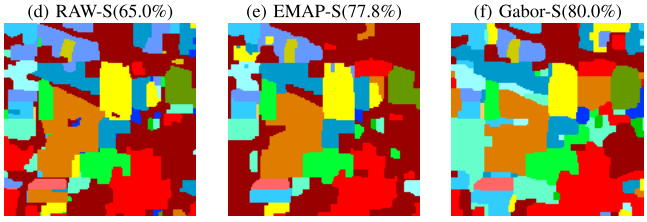
On the other hand, Fig. 5 shows the STDs of OAs obtained in 10 Monte Carlo runs as functions of the number of training samples, where MLR performs more steadily than SVM. Fig. 6 reveals the separate STD curves obtained with MLR and SVM. In both cases, the instability increases when MRF is contained in image processing, though it contributes to better classification results. Additionally, Gabor plays a more stabilizing role in spectral–spatial classification than EMAP, especially for MLR, whereas, as aforementioned, their promotion in OAs is similar. With regard to LBP, better performances are displayed just like the conclusions drawn from the analysis of OA curves.



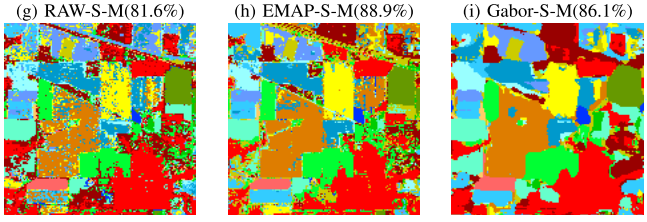
(a) False color map (b) Ground-truth (c) LBP(85.7%)



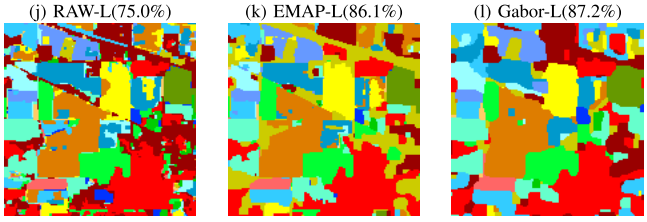
(d) RAW-S(65.0%) (e) EMAP-S(77.8%) (f) Gabor-S(80.0%)



(g) RAW-S-M(81.6%) (h) EMAP-S-M(88.9%) (i) Gabor-S-M(86.1%)



(j) RAW-L(75.0%) (k) EMAP-L(86.1%) (l) Gabor-L(87.2%)



(m) RAW-L-M(84.0%) (n) EMAP-L-M(89.2%) (o) Gabor-L-M(88.9%)

Fig. 7. AVIRIS Indian Pines scene and the classification maps obtained with 400 training samples and 9966 testing samples, along with the obtained OAs. The classes marked by C1–C16 in (f) represent Alfalfa, Corn-no till, Corn-min till, Corn, Grass/Pasture, Grass/Trees, Grass/Pasturemowed, Hay-windrowed, Oats, Soybeans-no till, Soybeansmin till, Soybeans-clean till, Wheat, Woods, Buildings-Grass-Tree-Drives, and Stone-steel-towers, respectively. (a) False color map. (b) Ground truth. (c) LBP (85.7%). (d) RAW-S (65.0%). (e) EMAP-S (77.8%). (f) Gabor-S (80.0%). (g) RAW-S-M (81.6%). (h) EMAP-S-M (88.9%). (i) Gabor-S-M (86.1%). (j) RAW-L (75.0%). (k) EMAP-L (86.1%). (l) Gabor-L (87.2%). (m) RAW-L-M (84.0%). (n) EMAP-L-M (89.2%). (o) Gabor-L-M (88.9%).

Finally, classification maps obtained with 25 training samples per class are listed in Fig. 7, along with corresponding OAs. Clearly, spatial information is beneficial for enhancing visual effects by accurately assigning more labels. First, pixel dependencies added in preprocessing stage contribute to smoother classification maps to some degree, especially with the Gabor filtering. Besides, the label dependencies introduced by MRF help improve the classification performance further. It should be noted that EMAP benefits more from MRF than Gabor. It is partly due to the fact that Gabor filtering is a powerful enough feature extraction tool. As for the integrated method, LBP also makes a significant difference by means of global dependency.

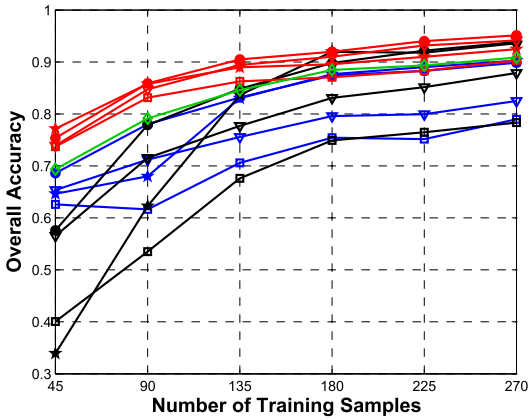


Fig. 8. OA curves (OAs versus the number of training samples) of the ROSIS Pavia University data set.

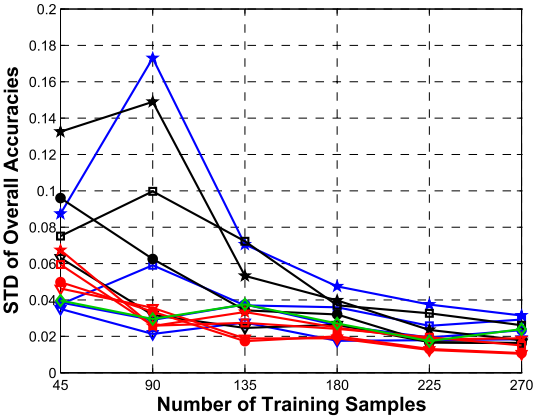
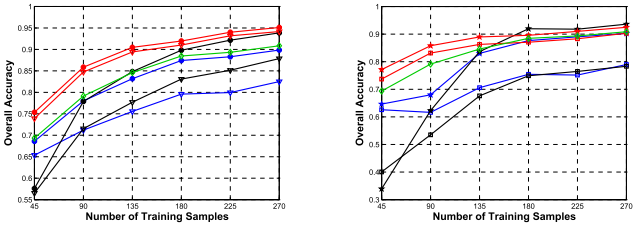


Fig. 9. STD curves (STDs of OAs versus the number of training samples) of the ROSIS Pavia University data set.



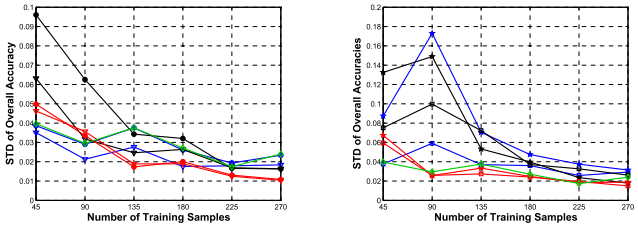
(a) (b)

Fig. 10. OA curves (OAs versus the number of training samples) of the ROSIS Pavia University data set, obtained by MLR (L) and SVM (S), respectively, along with LBP. (a) OAs obtained by L and LBP. (b) OAs obtained by S and LBP.

B. ROSIS Pavia University Data Set

The Pavia University data set is composed of 610 lines by 340 samples with 103 spectral bands. There are nine classes with 42 776 labeled samples available in the ground-truth image.

Similarly, we show the superiority of spectral–spatial classification by OAs, STDs, and classification maps on ROSIS Pavia University data set. As suggested in Figs. 8 and 9, Gabor stands out with higher OAs and lower STDs. However, in Fig. 10, the promotion of Gabor-MRF is attenuated while



(a) (b)

Fig. 11. STD curves (STDs of OAs versus the number of training samples) of the ROSIS Pavia University data set, obtained by MLR (L) and SVM (S), respectively, along with LBP. (a) STDs obtained by L and LBP. (b) STDs obtained by S and LBP.

MRF also boosts EMAP to great extent. This is in respect that Gabor is powerful in feature extraction, particularly for the data of high spatial resolution such as the Pavia University scene. The similar advantages reside in Fig. 11, wherein Gabor performs more steadily than EMAP. Moreover, the defect of MRF in stability remains in Fig. 11. In relation to the classification maps listed in Fig. 12, the analogous tendency of improvement is explicit when we compare the maps based on both spectral and spatial features with those only relying on spectral features. Gabor filtering, extracting features effectively before classification, outperforms EMAP to great degree, while EMAP gets more improvement with the assistance of MRF. In this case, LBP as a global spectral–spatial strategy also obtains a better result than the spectral classification methods.

C. EO-1 Botswana Data Set

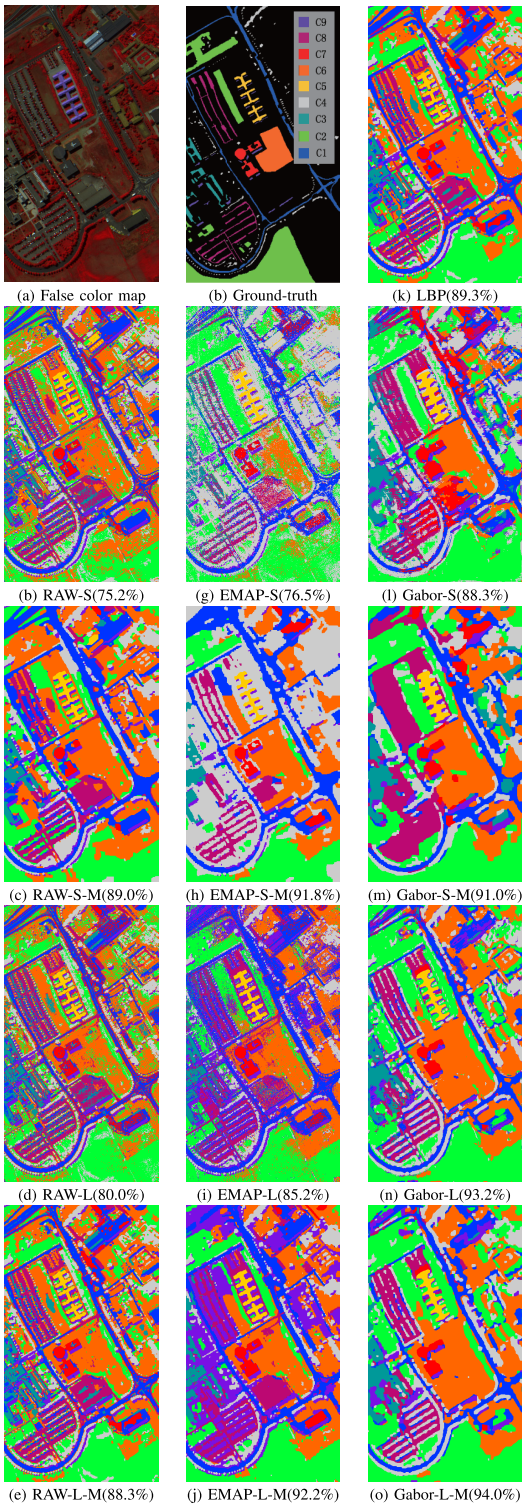
The third one is a subset from the EO-1 Botswana scene originally composed of 1476 lines and 256 columns with 145 spectral bands. In practice, we chose the top-middle part with a size of 460 lines and 256 columns for evaluation purposes. Finally, there are 1902 labeled samples with nine identified classes in the subset's ground-truth map [see Fig. 13(b)]. For simplicity, we list only the classification maps along with the corresponding OAs obtained by 18 training samples per class for the last two experiments.

As shown in Fig. 13, the preprocessing methods make a significant difference in improving the smoothness of classification maps. Particularly, Gabor filtering outperforms EMAP for all considered classifiers. Moreover, the postprocessing tool, MRF, makes a contribution to better visual effects in all cases, especially for EMAP-L and EMAP-L-M. Ultimately, BP as an integrated strategy can also obtain a good result as expected. Due to the lack of testing samples, the OA results are reported for reference only.

D. AVIRIS Salinas Data Set

The final data set, the Salinas scene, has a spatial coverage of 512×217 along with 224 spectral bands. The ground-truth map contains 54 129 labeled samples belonging to 16 classes [see Fig. 14(b)].

Fig. 14 shows the classification maps obtained with 288 training samples. Similar to the results of the other



(a) False color map (b) Ground-truth (k) LBP(89.3%)

(b) RAW-S(75.2%) (g) EMAP-S(76.5%) (l) Gabor-S(88.3%)

(c) RAW-S-M(89.0%) (h) EMAP-S-M(91.8%) (m) Gabor-S-M(91.0%)

(d) RAW-L(80.0%) (i) EMAP-L(85.2%) (n) Gabor-L(93.2%)

(e) RAW-L-M(88.3%) (j) EMAP-L-M(92.2%) (o) Gabor-L-M(94.0%)

Fig. 12. ROSIS Pavia University scene and the classification maps obtained with 225 training samples and 42 551 testing samples, along with the obtained OAs. The classes marked by C1–C9 in (f) represent Asphalt, Meadows, Gravel, Trees, Painted metal sheets, Bare Soil, Bitumen, Self-Blocking Bricks, and Shadows, respectively. (a) False color map. (b) RAW-S (75.2%). (c) RAW-S-M (89.0%). (d) RAW-L (80.0%). (e) RAW-L-M (88.3%). (f) Ground truth. (g) EMAP-S (76.5%). (h) EMAP-S-M (91.8%). (i) EMAP-L (85.2%). (j) EMAP-L-M (92.2%). (k) LBP (89.3%). (l) Gabor-S (88.3%). (m) Gabor-S-M (91.0%). (n) Gabor-L (93.2%). (o) Gabor-L-M (94.0%).

three experiments, these maps again prove the effectiveness of spatial dependencies. On the one hand, the improvement is evident with the pixel dependency involved, which resides in

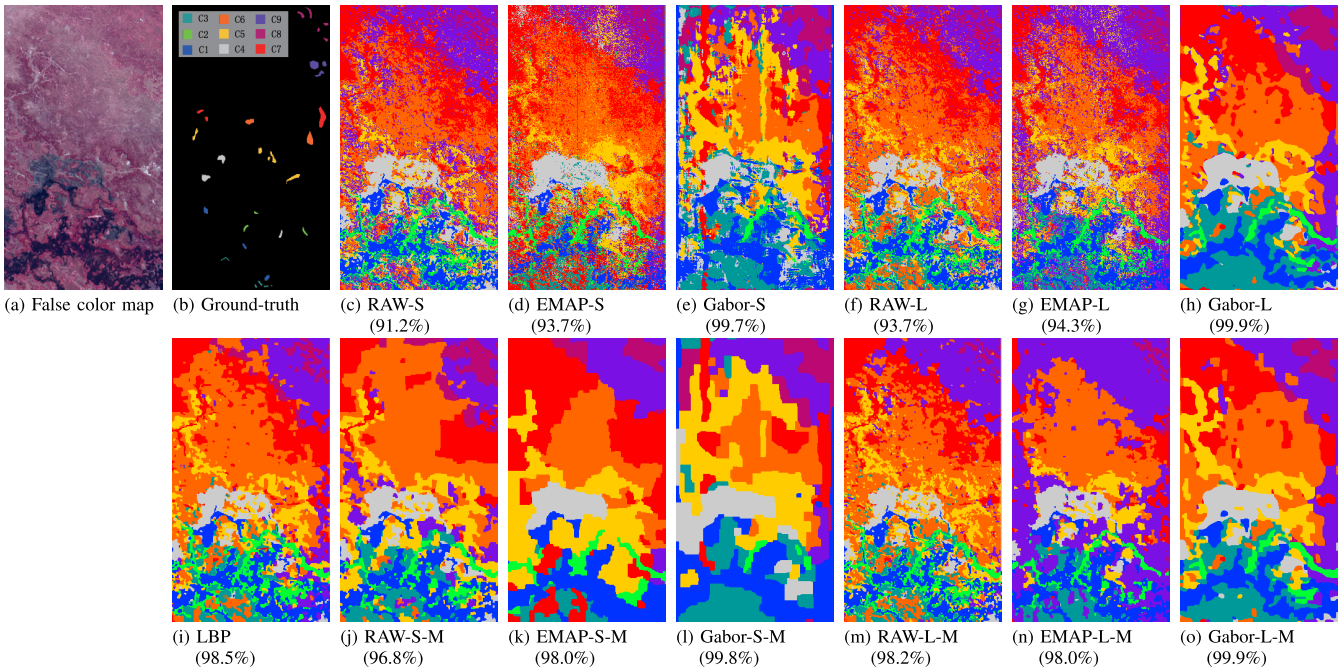


Fig. 13. Classification maps of the subset of the EO-1 Botswanan scene obtained with 162 training samples and 1740 testing samples, along with the obtained OAs. The identified classes marked by C1–C9 in (b) represent the land-cover types in seasonal swamps, occasional swamps, and drier woodlands located in the distal portion of the Delta. (a) False color map. (b) Ground truth. (c) RAW-S (91.2%). (d) EMAP-S (93.7%). (e) Gabor-S (99.7%). (f) RAW-L (93.7%). (g) EMAP-L (94.3%). (h) Gabor-L (99.9%). (i) LBP (98.5%). (j) RAW-S-M (96.8%). (k) EMAP-S-M (98.0%). (l) Gabor-S-M (99.8%). (m) RAW-L-M (98.2%). (n) EMAP-L-M (98.0%). (o) Gabor-L-M (99.9%).

Gabor filtering and EMAP. On the other hand, better results are obtained when the label dependency is stacked by MRF after classifications.

VI. Conclusion

Spectral–spatial hyperspectral image classification, which aims to collaboratively use spectral information and spatial information contained in data to improve hyperspectral classification, has been a multidisciplinary field attracting a lot of researchers over the recent years. We believe that in this stage, a systematic review and summary on existing spectral–spatial classification methods can promote the further development of the research area. Motivated by such, our work focuses on the following points: 1) introduced the concept of spatial dependency system, and accordingly unveiled the categorizations of fixed, adaptive, and global dependency systems and subsequently the groups of single-, bilayer-, and multiple-dependency systems; 2) presented spectral–spatial classification paradigms, including preprocessing-based, integrated, postprocessing-based, and hybrid classifications; and 3) investigated several typical spectral–spatial classification methodologies. Additionally, experimental comparisons were also carried out. Our work is expected to play roles in two aspects. On the one hand, it can help clarify mechanisms underlying the achievement of spectral–spatial classification; on the other hand, it implies guidelines that should be played by when designing a new method.

As regards future research lines in such a field, we believe the following might be promising or be worth special attentions.

1) *Application-Specific Spectral–Spatial Classification*: Many of existing spectral–spatial classification methods are over general, not tailored for a specific application or a specific task, meaning that they can be generally used for many scenarios while may neglect some special characteristics of the task. Such special characteristics usually imply certain *a priori* spatial dependency information that can be incorporated into the classification procedure to achieve better classification performance for the connected specific application.
2) *Superpixel Based Spectral–Spatial Classification*: Many traditional image processing techniques have been introduced into hyperspectral image analysis, one of which is the superpixel technique. Superpixel is based on the oversegmentation strategy to partition an image into homogenous subregions. Such homogenous subregions are usually spatially irregular units but perceptually consistent, i.e., all pixels in a superpixel are most likely uniform, meaning that superpixel methods are adaptive to real scenes. This implies an automatic determination of the neighborhood covering for a spatial dependency system in a data-dependent manner. Obviously, superpixel is associated with adaptive spatial dependency systems. Therefore, combining superpixel partitioning with conventional spectral–spatial classification approaches is expected to boost the performances, considering the variations of spatial characteristics of hyperspectral images. Although there already exist several publications on this point, such as [74], [75], and [79], we believe that

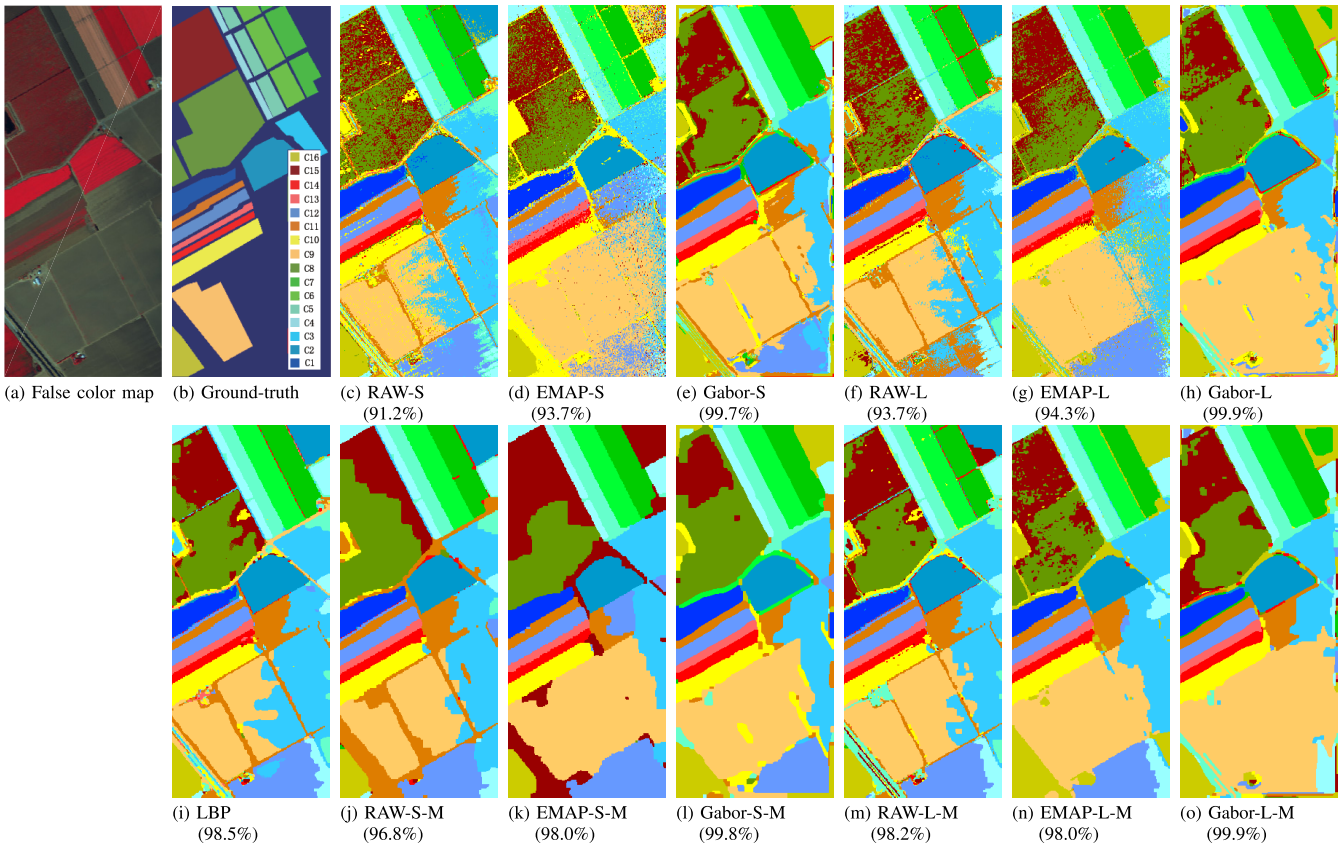


Fig. 14. Classification maps of the AVIRIS Salinas scene obtained with 288 training samples and 53 841 testing samples, along with the obtained OAs. The classes marked by C1–C9 in (b) represent Brocoli green weeds1, Brocoli green weeds2, Fallow, Fallow rough plow, Fallow smooth, Stubble, Celery, Grapes untrained, Soil vinyard develop, Corn senesced green weeds, Lettuce romaine 4wk, Lettuce romaine 5wk, Lettuce romaine 6wk, Lettuce romaine 7wk, Vinyard untrained, and Vinyard vertical trellis, respectively. (a) False color map. (b) Ground truth. (c) RAW-S (91.2%). (d) EMAP-S (93.7%). (e) Gabor-S (99.7%). (f) RAW-L (93.7%). (g) EMAP-L (94.3%). (h) Gabor-L (99.9%). (i) LBP (98.5%). (j) RAW-S-M (96.8%). (k) EMAP-S-M (98.0%). (l) Gabor-S-M (99.8%). (m) RAW-L-M (98.2%). (n) EMAP-L-M (98.0%). (o) Gabor-L-M (99.9%).

using superpixel or similar adaptive spatial dependency strategies for classification worths further explorations.

3) *Deep Learning Based Spectral–Spatial Classification*: Having achieved some breakthroughs in image classification [148]–[150] and started out to be applied on hyperspectral image processing [68], [69], [151], [152], deep learning is able to bind together multiple level features and the subsequent classifier in an end-to-end hierarchical network representation. Although introducing this technique into hyperspectral image classification offers the potential for improvements, there are still some problems that have to be considered, with the following two as our major concerns. First, large amounts of parameters, resulting from the manner in deep hierarchical networks, versus limited number of training samples, lead to the problem of obvious overfitting. Second, it is difficult to obtain the optimal or near optimal solution of a deep network in a feasible time by the current solvers, thus restricting the network's learning ability. Adjusting the network structure or bringing in additional information or constraints exploiting the spectral–spatial dependencies to the formulations might be helpful in relieving these problems, which is still in its early stage and deserves further efforts for the hyperspectral spectral–spatial classification.

REFERENCES

[1] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, Jun. 1985.

[2] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.

[3] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[4] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[5] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[6] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[7] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.

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

[9] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ, USA: Wiley, 2003, ch. 3.

[10] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.

[11] P.-F. Hsieh and D. Landgrebe, "Lowpass filter for increasing class separability," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 5. Jul. 1998, pp. 2691–2693

[12] F. Dell'Acqua, P. Gamba, A. Ferrari, J. A. Palmason, J. A. Benediktsson, and K. Arnason, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 322–326, Oct. 2004.

[13] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profiles," *Pattern Recognit.*, vol. 37, no. 6, pp. 1097–1116, 2004.

[14] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.

[15] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[16] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[17] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Econ. Geography*, vol. 46, pp. 234–240, Jun. 1970.

[18] F. Tsai and J.-S. Lai, "Feature extraction of hyperspectral image cubes using three-dimensional gray-level cooccurrence," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3504–3513, Jun. 2013.

[19] L. He, Y. Li, X. Li, and W. Wu, "Spectral–spatial classification of hyperspectral images via spatial translation-invariant wavelet-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2696–2712, May 2015.

[20] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[21] Z. Ye, S. Prasad, W. Li, J. E. Fowler, and M. He, "Classification based on 3-D DWT and decision fusion for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 173–177, Jan. 2014.

[22] S. Velasco-Forero and V. Manian, "Improving hyperspectral image classification using spatial preprocessing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 297–301, Apr. 2009.

[23] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral region classification using a three-dimensional Gabor filterbank," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3457–3464, Sep. 2010.

[24] L. He, J. Li, A. Plaza, and Y. Li, "Discriminative low-rank Gabor filtering for spectral–spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1381–1395, Mar. 2017.

[25] L. Shen and S. Jia, "Three-dimensional Gabor wavelets for pixel-based hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5039–5046, Dec. 2011.

[26] Y. Y. Tang, Y. Lu, and H. Yuan, "Hyperspectral image classification based on three-dimensional scattering wavelet transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2467–2480, May 2015.

[27] O. Rajadell, P. García-Sevilla, and F. Pla, "Spectral–spatial pixel characterization using Gabor filters for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 860–864, Jul. 2013.

[28] Z. Zhong *et al.*, "Discriminant tensor spectral–spatial feature extraction for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1028–1032, May 2015.

[29] K.-H. Liu, Y.-Y. Lin, and C.-S. Chen, "Linear spectral mixture analysis via multiple-kernel learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2254–2269, Apr. 2015.

[30] P. Gurram and H. Kwon, "Sparse kernel-based ensemble learning with fully optimized kernel parameters for hyperspectral classification problems," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 787–802, Feb. 2013.

[31] Y. Zhou, J. Peng, and C. L. P. Chen, "Extreme learning machine with composite kernels for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2351–2360, Jun. 2015.

[32] H. Li, Z. Ye, and G. Xiao, "Hyperspectral image classification using spectral-spatial composite kernels discriminant analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2341–2350, Jun. 2015.

[33] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

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

[35] A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 527–541, Jan. 2015.

[36] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.

[37] L. Bruzzone and C. Persello, "A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2142–2154, Jul. 2009.

[38] C.-H. Li, B.-C. Kuo, C.-T. Lin, and C.-S. Huang, "A spatial–contextual support vector machine for remotely sensed image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 784–799, Mar. 2012.

[39] X. Guo, X. Huang, L. Zhang, L. Zhang, A. Plaza, and J. A. Benediktsson, "Support tensor machines for classification of hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3248–3264, Jun. 2016.

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

[41] Y. Gu, T. Liu, X. Jia, J. A. Benediktsson, and J. Chanussot, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.

[42] M. D. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.

[43] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, Sep. 2014.

[44] J. Xia, M. D. Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, Sep. 2015.

[45] N. Falco, J. A. Benediktsson, and L. Bruzzone, "Spectral and spatial classification of hyperspectral images based on ICA and reduced morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6223–6240, Nov. 2015.

[46] B. Song *et al.*, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, Aug. 2014.

[47] W. Liao, M. D. Mura, J. Chanussot, R. Bellens, and W. Philips, "Morphological attribute profiles with partial reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1738–1756, Mar. 2016.

[48] B. Demir and L. Bruzzone, "Histogram-based attribute profiles for classification of very high resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2096–2107, Apr. 2016.

[49] P. Ghamisi, M. D. Mura, and J. A. Benediktsson, "A survey on spectral–spatial classification techniques based on attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, May 2015.

[50] J. Li *et al.*, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.

[51] 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 Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[52] J. Xia, J. Chanussot, P. Du, and X. He, "Spectral–spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2532–2546, May 2015.

[53] M. Khodadadzadeh, J. Li, A. Plaza, H. Ghassemian, J. M. Bioucas-Dias, and X. Li, "Spectral–spatial classification of hyperspectral data using local and global probabilities for mixed pixel characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6298–6314, Oct. 2014.

[54] J. Bai, S. Xiang, and C. Pan, "A graph-based classification method for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 803–817, Feb. 2013.

[55] W. Li, S. Prasad, and J. E. Fowler, "Hyperspectral image classification using Gaussian mixture models and Markov random fields," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 153–157, Jan. 2014.

[56] Z. Wang, N. M. Nasrabadi, and T. S. Huang, "Spatial–spectral classification of hyperspectral images using discriminative dictionary designed by learning vector quantization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4808–4822, Aug. 2014.

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

[58] P. Zhong and R. Wang, "Learning conditional random fields for classification of hyperspectral images," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1890–1907, Jul. 2010.

[59] P. Zhong and R. Wang, "Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1319–1334, Jul. 2014.

[60] P. Guccione, L. Mascolo, and A. Appice, "Iterative hyperspectral image classification using spectral–spatial relational features," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3615–3627, Jul. 2015.

[61] X. Sun, N. M. Nasrabadi, and T. D. Tran, "Task-driven dictionary learning for hyperspectral image classification with structured sparsity constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4457–4471, Aug. 2015.

[62] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, Jun. 2014.

[63] J. Zhao, Y. Zhong, and L. Zhang, "Detail-preserving smoothing classifier based on conditional random fields for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2440–2452, May 2015.

[64] Y. Zhong, X. Lin, and L. Zhang, "A support vector conditional random fields classifier with a Mahalanobis distance boundary constraint for high spatial resolution remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1314–1330, Apr. 2014.

[65] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.

[66] J. Xia, L. Bombrun, T. Adali, Y. Berthoumieu, and C. Germain, "Spectral–spatial classification of hyperspectral images using ICA and edge-preserving filter via an ensemble strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4971–4982, Aug. 2016.

[67] D. Lunga and O. Ersoy, "Multidimensional artificial field embedding with spatial sensitivity," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1518–1532, Feb. 2014.

[68] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[69] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[70] X. Kang, S. Li, L. Fang, M. Li, and J. A. Benediktsson, "Extended random walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.

[71] R. Ji, Y. Gao, R. Hong, Q. Liu, D. Tao, and X. Li, "Spectral-spatial constraint hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1811–1824, Mar. 2014.

[72] J. Liu, Z. Wu, J. Li, A. Plaza, and Y. Yuan, "Probabilistic-kernel collaborative representation for spatial–spectral hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2371–2384, Apr. 2016.

[73] J. Peng, Y. Zhou, and C. L. P. Chen, "Region-kernel-based support vector machines for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4810–4824, Sep. 2015.

[74] G. Zhang, X. Jia, and J. Hu, "Superpixel-based graphical model for remote sensing image mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5861–5871, Nov. 2015.

[75] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.

[76] R. Roscher and B. Waske, "Shapelet-based sparse representation for landcover classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1623–1634, Mar. 2016.

[77] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, Oct. 2015.

[78] Z. Lu and J. He, "Spectral–spatial hyperspectral image classification with adaptive mean filter and jump regression detection," *Electron. Lett.*, vol. 51, no. 21, pp. 1658–1660, Oct. 2015.

[79] T. Lu, S. Li, L. Fang, L. Bruzzone, and J. A. Benediktsson, "Set-to-set distance-based spectral–spatial classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7122–7134, Dec. 2016.

[80] X. Zhang, S. E. Chew, Z. Xu, and N. D. Cahill, "SLIC superpixels for efficient graph-based dimensionality reduction of hyperspectral imagery," *Proc. SPIE*, vol. 9472, p. 947209, May 2015.

[81] Y. Liu, G. Cao, Q. Sun, and M. Siegel, "Hyperspectral classification via deep networks and superpixel segmentation," *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3459–3482, Jul. 2015.

[82] S. Valero, P. Salembier, and J. Chanussot, "Hyperspectral image representation and processing with binary partition trees," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1430–1443, Apr. 2013.

[83] M. A. Veganzones, G. Tochon, M. Dalla-Mura, A. J. Plaza, and J. Chanussot, "Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3574–3589, Aug. 2014.

[84] Z. Xue, P. Du, J. Li, and H. Su, "Simultaneous sparse graph embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6114–6133, Nov. 2015.

[85] J. Li, H. Zhang, Y. Huang, and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707–3719, Jun. 2014.

[86] H. Yuan and Y. Y. Tang, "Sparse representation based on set-to-set distance for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2464–2472, Jun. 2015.

[87] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.

[88] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[89] P. Ghamisi, M. S. Couceiro, F. M. L. Martins, and J. A. Benediktsson, "Multilevel image segmentation based on fractional-order darwinian particle swarm optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2382–2394, May 2014.

[90] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, 2010.

[91] Y. Zhong, J. Zhao, and L. Zhang, "A hybrid object-oriented conditional random field classification framework for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7023–7037, Nov. 2014.

[92] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1490–1503, Mar. 2015.

[93] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive Markov random field approach for classification of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 973–977, Sep. 2011.

[94] P. Zhong and R. Wang, "Modeling and classifying hyperspectral imagery by CRFs with sparse higher order potentials," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 688–705, Feb. 2011.

[95] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.

[96] Z. He, Q. Wang, Y. Shen, and M. Sun, "Kernel sparse multitask learning for hyperspectral image classification with empirical mode decomposition and morphological wavelet-based features," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5150–5163, Aug. 2014.

[97] R. D. Phillips, C. E. Blinn, L. T. Watson, and R. H. Wynne, "An adaptive noise-filtering algorithm for aviris data with implications for classification accuracy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3168–3179, Sep. 2009.

[98] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[99] M. Golipour, H. Ghassemian, and F. Mirzapour, "Integrating hierarchical segmentation maps with MRF prior for classification of hyperspectral images in a Bayesian framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 805–816, Feb. 2016.

[100] G. Zhang and X. Jia, "Simplified conditional random fields with class boundary constraint for spectral-spatial based remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 856–860, Sep. 2012.

[101] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.

[102] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.

[103] B. Demir and S. Erturk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4071–4084, Nov. 2010.

[104] Z. He, J. Li, L. Liu, K. Liu, and L. Zhuo, "Fast three-dimensional empirical mode decomposition of hyperspectral images for class-oriented multitask learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6625–6643, Nov. 2016.

[105] J. Zabalza *et al.*, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4418–4433, Aug. 2015.

[106] X. Huang, Q. Lu, L. Zhang, and A. Plaza, "New postprocessing methods for remote sensing image classification: A systematic study," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7140–7159, Nov. 2014.

[107] L. He, Q. Pan, W. Di, and Y. Li, "Anomaly detection in hyperspectral imagery based on maximum entropy and nonparametric estimation," *Pattern Recognit. Lett.*, vol. 29, no. 9, pp. 1392–1403, 2008.

[108] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991, ch. 11.

[109] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1. Hoboken, NJ, USA: Prentice Hall, 1993, ch. 3.

[110] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA, USA: Addison-Wesley, 2006, ch. 6.

[111] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2001, ch. 2.

[112] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. San Diego, CA, USA: Academic, 1990, ch. 3.

[113] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.

[114] Y. Fang *et al.*, "Dimensionality reduction of hyperspectral images based on robust spatial information using locally linear embedding," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1712–1716, Oct. 2014.

[115] S. Bourennane, C. Fossati, and A. Cailly, "Improvement of classification for hyperspectral images based on tensor modeling," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 801–805, Oct. 2010.

[116] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. Berlin, Germany: Springer, 2006.

[117] P. R. Marpu, M. Pedergnana, M. D. Mura, S. Peeters, J. A. Benediktsson, and L. Bruzzone, "Classification of hyperspectral data using extended attribute profiles based on supervised and unsupervised feature extraction techniques," *Int. J. Image Data Fusion*, vol. 3, no. 3, pp. 269–298, 2012.

[118] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, Apr. 1998. [Online]. Available: http://dx.doi.org/10.1109/83.663500

[119] S. Kumar and M. Hebert, "Discriminative random fields," *Int. J. Comput. Vis.*, vol. 68, no. 2, pp. 179–201, 2006.

[120] J. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Mach. Learn. (ICML)*, 2001, pp. 282–289.

[121] S. Sun, P. Zhong, H. Xiao, and R. Wang, "An MRF model-based active learning framework for the spectral-spatial classification of hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1074–1088, Sep. 2015.

[122] J. S. Borges, J. M. Bioucas-Dias, and A. R. S. Marcal, "Bayesian hyperspectral image segmentation with discriminative class learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2151–2164, Jun. 2011.

[123] P. Zhong and R. Wang, "Learning sparse CRFs for feature selection and classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4186–4197, Dec. 2008.

[124] F. Li, L. Xu, P. Siva, A. Wong, and D. A. Clausi, "Hyperspectral image classification with limited labeled training samples using enhanced ensemble learning and conditional random fields," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2427–2438, Jun. 2015.

[125] J. Zhao, Y. Zhong, Y. Wu, L. Zhang, and H. Shu, "Sub-pixel mapping based on conditional random fields for hyperspectral remote sensing imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1049–1060, Sep. 2015.

[126] F. Yao, Y. Qian, Z. Hu, and J. Li, "A novel hyperspectral remote sensing images classification using Gaussian Processes with conditional random fields," in *Proc. Int. Conf. Intell. Syst. Knowl. Eng. (ISKE)*, Nov. 2010, pp. 197–202.

[127] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[128] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 41, no. 1, pp. 129–159, 2001.

[129] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[130] D. Ni and H. Ma, "Hyperspectral image classification via sparse code histogram," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1843–1847, Sep. 2015.

[131] C. Li, Y. Ma, X. Mei, C. Liu, and J. Ma, "Hyperspectral image classification with robust sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 641–645, May 2016.

[132] J. Liu, Z. Wu, Z. Wei, L. Xiao, and L. Sun, "Spatial–spectral kernel sparse representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2462–2471, Dec. 2013.

[133] P. Du, Z. Xue, J. Li, and A. Plaza, "Learning discriminative sparse representations for hyperspectral image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1089–1104, Sep. 2015.

[134] H. Yuan, Y. Y. Tang, Y. Lu, L. Yang, and H. Luo, "Hyperspectral image classification based on regularized sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2174–2182, Jun. 2014.

[135] J. Liu, Z. Wu, L. Sun, Z. Wei, and L. Xiao, "Hyperspectral image classification using kernel sparse representation and semilocal spatial graph regularization," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 8, pp. 1320–1324, Aug. 2014.

[136] S. Jia, X. Zhang, and Q. Li, "Spectral-spatial hyperspectral image classification using $\ell_{1/2}$ regularized low-rank representation and sparse representation-based graph cuts," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2473–2484, Jun. 2015.

[137] C. Chen, N. Chen, and J. Peng, "Nearest regularized joint sparse representation for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 424–428, Mar. 2016.

[138] X. Sun, Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Structured priors for sparse-representation-based hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1235–1239, Jul. 2014.

[139] P. Soille, *Morphological Image Analysis, Principles and Applications*, 2nd ed. Berlin, Germany: Springer-Verlag, 2003.

[140] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral–spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.
[141] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.
[142] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.
[143] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.
[144] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
[145] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016.
[146] J. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented Lagrangian tools," Inst. Superior Técnico, Lisbon, Portugal, Tech. Rep., 2009.
[147] H.-T. Lin, C.-J. Lin, and R. C. Weng, "A note on Platt's probabilistic outputs for support vector machines," *Mach. Learn.*, vol. 68, no. 3, pp. 267–276, 2007.
[148] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[149] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[150] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[151] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, 2016.
[152] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016.

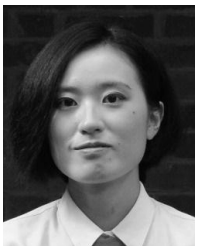
Lin He (S'05–M'12) received the B.S. degree from the Xi'an Institute of Technology, Xi'an, China, in 1995, the M.S. degree from Chongqing University, Chongqing, China, in 2003, and the Ph.D. degree from Northwestern Polytechnical University, Xi'an, in 2007.

Since 2007, he has been with the College of Automation Science and Engineering, South China University of Technology, Guangzhou, China, where he is currently an Associate Professor. His research interests include statistical pattern recognition, hyperspectral image processing, and biomedical signal processing.

Jun Li (M'12–SM'16) received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004, the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotcnica e de Computadores and the Instituto de Telecomunicações, IST, Universidade Técnica de Lisboa, in the framework of the European Doctorate for signal processing. She has also been actively involved in the hyperspectral imaging network, a Marie Curie Research Training Network involving 15 partners in 12 countries and intended to foster research, training, and cooperation on hyperspectral imaging at the European level. Since 2011, she has been a Post-Doctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain. Currently, she is a Professor with Sun Yat-Sen University, Guangzhou, China. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li is an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and has been a reviewer for several journals.



Chenying Liu (S'16) received the B.S. degree from Sun Yat-sen University, Guangzhou, China, in 2017, where she is currently pursuing the M.S. degree with the School of Geography and Planning.

Her research interests include hyperspectral image classification and machine learning.

Shutao Li (M'07–SM'15) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he was with the College of Electrical and Information Engineering, Hunan University. In 2001, he was also a Research Associate with the Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong. From 2002 to 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London, Surrey, U.K. In 2005, he was a Visiting Professor with the Department of Computer Science, The Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He is also a Chang-Jiang Scholar Professor appointed by the Ministry of Education of China. He has authored or co-authored over 180 refereed papers. His research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

Dr. Li was a recipient of two Second-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006. He is an Editorial Board Member of the *Information Fusion, Sensing, and Imaging*. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT. In 2013, he was granted the National Science Fund for Distinguished Young Scholars in China.