Advances in Hyperspectral Image and Signal Processing

A comprehensive overview of the state of the art

LANDSCAPE—©ISTOCKPHOTO.COM/MAXIMKABB,
BACKGROUND ART—ALL-FREE-DOWNLOAD.COM

PEDRAM GHAMISI, NAOTO YOKOYA, JUN LI, WENZHI LIAO, SICONG LIU, JAVIER PLAZA, BEHNOOD RASTI, AND ANTONIO PLAZA

Recent advances in airborne and spaceborne hyperspectral imaging technology have provided end users with rich spectral, spatial, and temporal information. They have made a plethora of applications feasible for the analysis of large areas of the Earth's surface. However, a significant number of factors—such as the high dimensions and size of the hyperspectral data, the lack of training samples, mixed pixels, light-scattering mechanisms in the acquisition process, and different atmospheric and geometric distortions—make such data inherently nonlinear and complex, which poses major challenges for existing methodologies to effectively process and analyze the data sets. Hence, rigorous and innovative methodologies are required for hyperspectral image (HSI) and signal processing and have become a center of attention for researchers worldwide.

This article offers a comprehensive tutorial/overview focusing specifically on hyperspectral data analysis, which is categorized into seven broad topics: classification, spectral unmixing, dimensionality reduction (DR), resolution enhancement, HSI denoising and restoration, change detection (CD), and fast computing. For each topic, we provide a synopsis of the state-of-the-art approaches and numerical results for validating and evaluating different methodologies, followed by a discussion of future challenges and research directions.

Digital Object Identifier 10.1109/MGRS.2017.2762087
Date of publication: 27 December 2017

THE PROMISE OF HYPERSPECTRAL IMAGING

Remote sensing involves obtaining information from an object or a scene without any direct physical contact. This is possible because different objects uniquely reflect, absorb, and emit electromagnetic radiation based on their molecular composition and texture. If the radiation arriving at a sensor is measured at a detailed wavelength range, the consequent spectral signature, also known as a *spectrum*, can potentially be used to identify any given object of interest. To this end, the intent of hyperspectral imaging technology is to capture, from the immediate surface of the Earth, hundreds of spectral channels (i.e., to shape the spectra) that can precisely characterize the chemical composition of different materials.

Hyperspectral sensors sample mainly the reflective portion of the electromagnetic spectrum, ranging from the visible region (0.4–0.7 μm) to the short-wave infrared (SWIR) region (almost 2.4 μm) in hundreds of narrow contiguous spectral channels, each of which is 10-nm wide. There are, however, other types of hyperspectral sensors that are able to characterize the emissive properties of objects by collecting data in the range of the midwave and long-wave infrared region. (Hyperspectral imaging covers a broad range of imaging systems, such as medical hyperspectral imaging, atmospheric sounding, close-range hyperspectral imaging, and so on. Here, we focus solely on airborne or spaceborne remotely sensed HSIs with a spectral coverage ranging 0.4–2.5 μm.) Such detailed spectral sampling, making use of numerous small, commercial, high spatial and spectral instruments, has made HSIs a valuable source of information for a wide variety of applications, including precision agriculture (e.g., monitoring the development and health of crops), the food industry (e.g., characterizing product quality), environmental monitoring, mineralogy, defense and security-based applications (e.g., identification of man-made materials), chemical imaging, astronomy, ecological sciences, and many others.

A better understanding of HSIs can be gained from Figure 1. A three-dimensional (3-D) hyperspectral data cube consists of $n_1 \times n_2 \times d$ pixels, in which $n_1 \times n_2$ is the number of pixels in each spectral channel and d represents the number of spectral channels. An HSI can be characterized using one of the following more detailed definitions.

1) *Spectral perspective (or spectral dimension)*: From this perspective, a hyperspectral data cube is composed of $n_1 \times n_2$ pixels, where each pixel is a vector of d values. Each pixel corresponds to the reflected radiation of the specific region of the Earth and has multiple values in spectral bands. This detailed spectral information can be used to analyze different materials with precision. Figure 1(c) shows a spectral profile of one pixel, with multiple values for each band in the spectral dimension.
2) *Spatial perspective (or spatial dimension)*: In this context, a hyperspectral data cube consists of d gray-scale

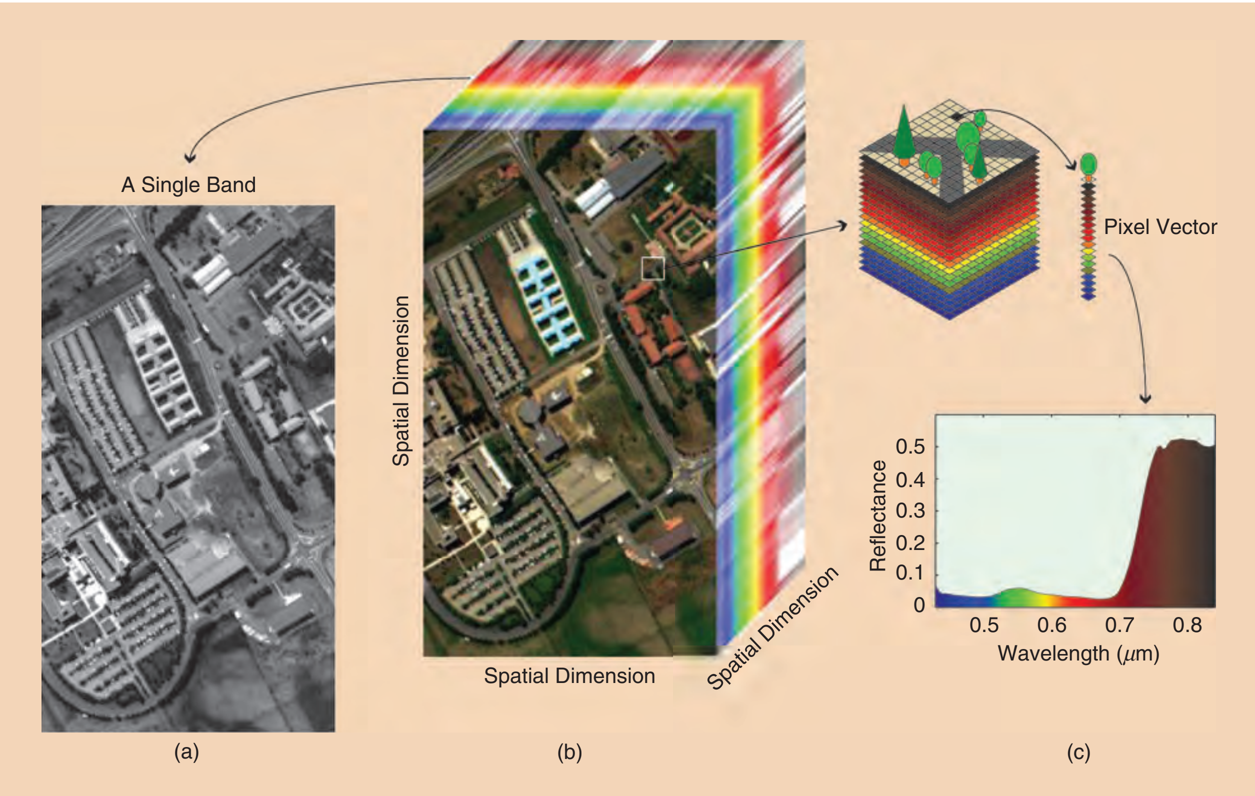

FIGURE 1. An example of a hyperspectral data cube: (a) a gray-scale image, (b) a hyperspectral data cube, and (c) a pixel vector and its corresponding spectral signature.

images, with a size of $n_1 \times n_2$. The values of all of the pixels in one spectral band shape a gray-scale image with two dimensions [as shown in Figure 1(a)], which are both spatial.

Although the greater dimensionality of HSIs compared with multispectral images improves data information content considerably, it does introduce new challenges to conventional image analysis techniques, which have been specifically designed for multispectral data. Furthermore, it is almost impossible for humans to visualize spaces of higher than three dimensions (e.g., red-green-blue images). A misunderstanding of high-dimensional spaces and conventional spaces sometimes leads to incorrect interpretations of HSIs and the inappropriate choice of the data processing technique. Bearing this in mind, in the next section, we provide an overview of a few common HSI challenges and their possible solutions.

MAIN CHALLENGES OF HYPERSPECTRAL IMAGE ANALYSIS AND POSSIBLE SOLUTIONS

Several factors make the analysis and processing of HSIs a challenging task. Figure 2 illustrates the main paths in HSI analysis that have been developed primarily to address these factors. In this section, we take a closer look at each of the applications shown in Figure 2. The common understanding of HSIs is that, because such data contain a rich amount of spectral information, the whole dimensionality needs to be used to define precise boundaries in the feature space for a specific application. The increasing spectral resolution of HSIs benefits precision applications (e.g., Earth observation, precision agriculture, and disease detection). However, it challenges conventional signal-processing techniques and, thus, hampers the abilities of HSIs in many real applications.

Taking classification as an example (because classification is one of the most popular applications for HSIs), we found in [1] that, when the number of training samples remains constant, after a few features, classification accuracy actually decreases as the number of features increases. Two solutions have been widely exploited to address this problem.

1) *Dimension (feature) reduction*: As mentioned in several studies, such as [2]–[4], a high-dimensional space is almost empty, and multivariate data can be represented in a lower-dimensional space, where the undesirable effects of high-dimensional geometric characteristics and the curse of dimensionality are reduced. This fact has led to a chain of research on dimension (feature) reduction, which will be detailed in the "Dimensionality Reduction" section.
2) *Robust classifiers*: The imbalance between the number of bands and available training samples has a dramatic influence on supervised classifiers. In this context, HSIs often demand a vast number of training samples to effectively estimate class parameters. To benefit from the rich spectral information of HSIs, one possible solution is based on using effective and efficient classification approaches that can handle high dimensionality, even if a limited number of training samples is available. In addition, along with the detailed spectral information provided by HSIs, it is possible to take advantage of available spatial information (in particular, for very-high-spatial-resolution HSIs) to further improve the eventual classification map. The "Classification" section elaborates on advances in HSI classification.

Spectral mixing (including both linear and nonlinear models) is another bottleneck for HSI analysis that occurs for a number of reasons, such as insufficient spatial resolution of the sensor and an intimate mixing effect. When mixing takes place, it is not possible to directly distinguish the materials available in the pixels from the corresponding measured spectral vectors. However, detailed spectral information provided by HSIs can be used to unmix hyperspectral pixels. The "Spectral Unmixing" section focuses on spectral unmixing to address these issues.

WHEN THE NUMBER OF TRAINING SAMPLES REMAINS CONSTANT, AFTER A FEW FEATURES, CLASSIFICATION ACCURACY ACTUALLY DECREASES AS THE NUMBER OF FEATURES INCREASES.

Spaceborne imaging spectrometers are usually designed to acquire HSIs with a moderate spatial resolution—e.g., a ground sampling distance (GSD) of 30 m—because of the inevitable tradeoffs among spatial resolution, spectral resolution, temporal resolution, and signal-to-noise ratio (SNR). Spatial resolution enhancement of HSIs is a technology essential to expanding the range of applications for spaceborne hyperspectral missions. In the "Resolution Enhancement" section, we discuss techniques for the resolution enhancement of HSIs.

The degradation mechanisms associated with the measurement process and atmospheric effects inject undesirable noise that substantially downgrades the quality of hyperspectral data. The HSI SNR is usually decreased during the imaging process, depending on different noise sources. In remote-sensing HSIs, highly corrupted bands must often be removed before any further processing. Alternatively, HSI restoration can recover those corrupted bands and also improve the HSI SNR, thereby improving the effectiveness of any further processing of the HSI. In this context, the "HSI Denoising and Image Restoration" section is dedicated to HSI denoising and image restoration techniques that address such effects.

Another emerging research domain in the hyperspectral community, CD is the process of identifying and examining spectral–temporal changes in signals. The detailed spectral sampling and representation in HSIs result in the potential identification of more subtle spectral variations, which are usually not easily detected in traditional multispectral images. Accordingly, land cover dynamic monitoring can be enhanced to a finer level. To this end, advanced CD techniques must be designed to address CD issues in multitemporal

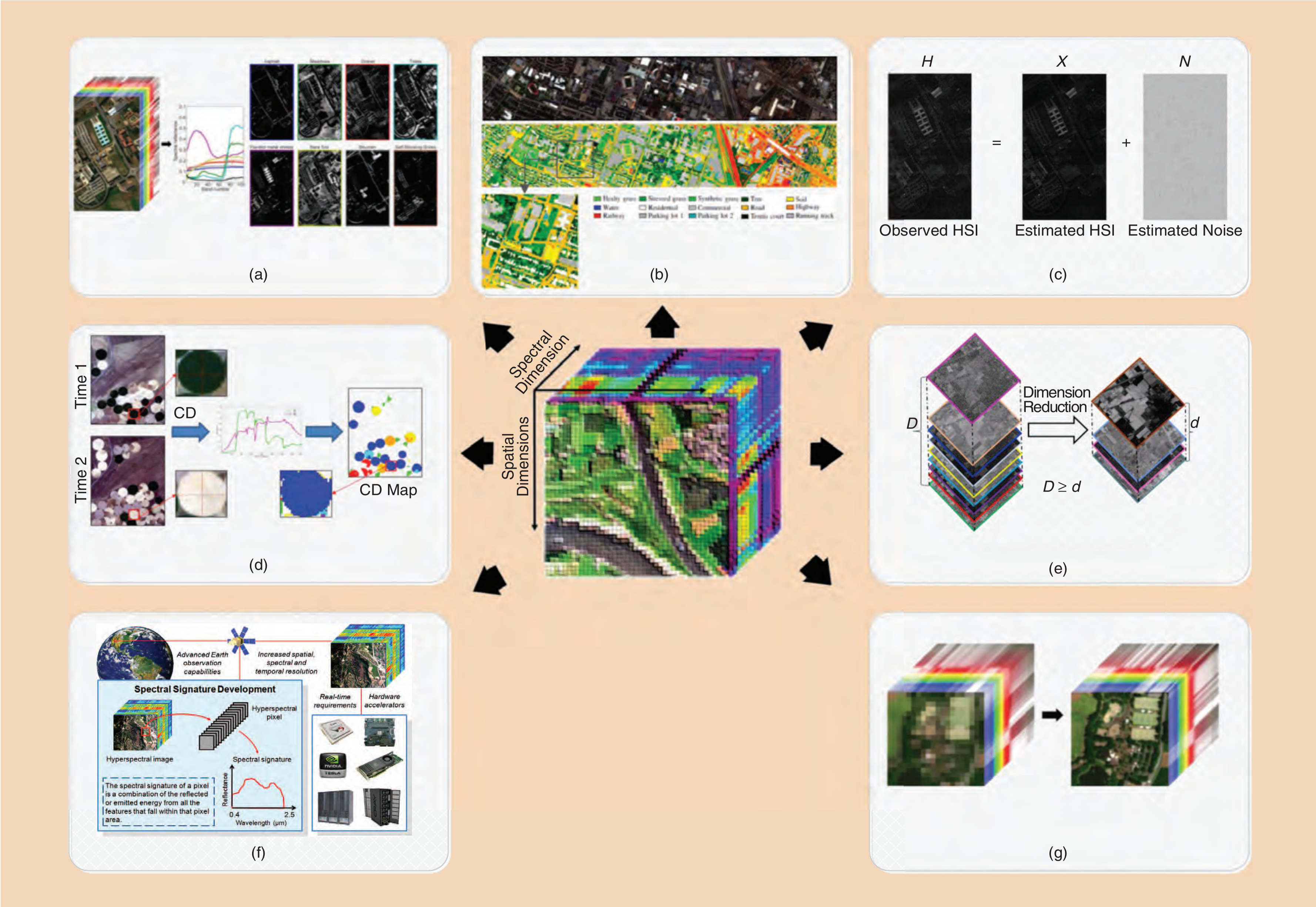

FIGURE 2. The different paths of HSI analysis: (a) spectral unmixing, (b) classification, (c) image restoration and denoising, (d) CD, (e) DR, (f) fast computing, and (g) resolution enhancement.

HSIs and, at the same time, overcome the challenges caused by the hyperspectral data set. We elaborate on different CD methods in the "Change Detection" section.

Another crucial aspect of HSI analysis to be precisely taken into account is that hyperspectral remote sensors are now in the era of massive automatic data collection resulting from the improved spatial, spectral, and temporal resolutions provided by several hyperspectral instruments. As a result, fast computing (detailed in the "Fast Computing" section) is critical to accelerating the efficient exploitation and analysis of HSIs.

MISSIONS AND STATISTICS

Several hyperspectral imaging instruments are currently available for the purpose of remote-sensing image and signal analysis, providing a large volume of images for various thematic applications. Airborne hyperspectral imaging sensors [e.g., the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Hyperspectral Digital Imagery Collection Experiment, Compact Airborne Spectrographic Imager (CASI), Airborne Prism Experiment, and HySpex] play a central role in acquiring data sources for the hyperspectral scientific community. The European Facility for Airborne Research (http://www.eufar.net) has established standards and protocols in the field of airborne hyperspectral remote sensing, allowing transnational access to national infrastructures. In recent years, unmanned-aerial-vehicle-based real-time hyperspectral imaging has become increasingly common in various applications, such as agricultural monitoring, while raising new challenges in image processing.

Table 1 presents the principal parameters of seven spaceborne imaging spectroscopy missions planned for the near future: the China commercial remote-sensing satellite system, DLR Earth sensing imaging spectrometer (DESIS) [5], environmental mapping and analysis program [6], hyperspectral imager suite (HISUI) [7], precursore iperspettrale della missione applicativa (PRISMA) [8], spaceborne hyperspectral applicative land and ocean mission (Shalom) [9], and hyperspectral infrared imager [10]. Many of those satellites are designed to have a GSD of 30 m, aiming at global coverage. Shalom's GSDs of 8 m and 10 m in the visible near-infrared (VNIR) and SWIR ranges, respectively, are driven by both operational and commercial needs. DESIS and HISUI will be mounted on the International Space Station. The launch of these satellites will further accelerate research on HSI processing and its applications.

Figure 3 shows statistics on articles related to HSIs and signal processing published in IEEE journals during 2009–2012 and 2013–2016. All articles were searched via IEEE *Xplore* using *hyperspectral* as the main keyword and then categorized into broad topics by analyzing keywords in the titles. The size of each pie in Figure 3 is proportional to the number of articles.

UNMANNED-AERIAL-VEHICLE-BASED REAL-TIME HYPERSPECTRAL IMAGING HAS BECOME INCREASINGLY COMMON IN VARIOUS APPLICATIONS.

The totals returned by this search are an indicator of the hyperspectral community's recent growth. Seven topics under investigation represented 61.5% of all of the articles published in 2013–2016. Classification was the most actively addressed topic in both periods, while spectral unmixing was the second most common. Classification- and unmixing-related studies accounted for 41.6% of the total. These top two topics were followed by DR and image restoration. Image restoration showed a high growth rate, indicating that the improvement in the quality of HSIs is significant in subsequent processing. Resolution enhancement received particular attention during 2013–2016, as demonstrated by the highest growth rate. Although the number of articles related to CD increased steadily, the overall number was still small, probably due to limited data sets.

CONTRIBUTION

This article introduces a detailed and organized overview of HSIs and signal processing, categorized into the seven different themes previously mentioned. In each section, we provide some numerical results, illustrations, a critical overview of the state of the art, current challenges, and possible future works. It is worth noting that the methodologies described

TABLE 1. THE PARAMETERS OF SEVEN SPACEBORNE IMAGING SPECTROSCOPY MISSIONS.

PARAMETER	CCRSS	DESIS	EnMAP	HISUI	PRISMA	Shalom	HyspIRI
Altitude (km)	30	400	653	400	615	600	626
GSD (m)	30	30	30	30	30	10	30
Bandwidth (nm)	5–20	3.3	5.25–12.5	10–12.5	≤ 12	10	≤ 10
Spectral coverage (μm)	0.4–2.5	0.4–1.0	0.42–2.45	0.44–2.5	0.4–2.5	0.4–2.5	0.38–2.5
Number of bands	328	180	228	185	237	241	210
Swath width (km)	30	30.7	30	20	30–60	10	45
Other sensor	Pan	—	—	—	Pan	Pan	TIR

TIR: thermal infrared; CCRSS: China commercial remote-sensing satellite system; EnMAP: environmental mapping and analysis program; HyspIRI: hyperspectral infrared imager.

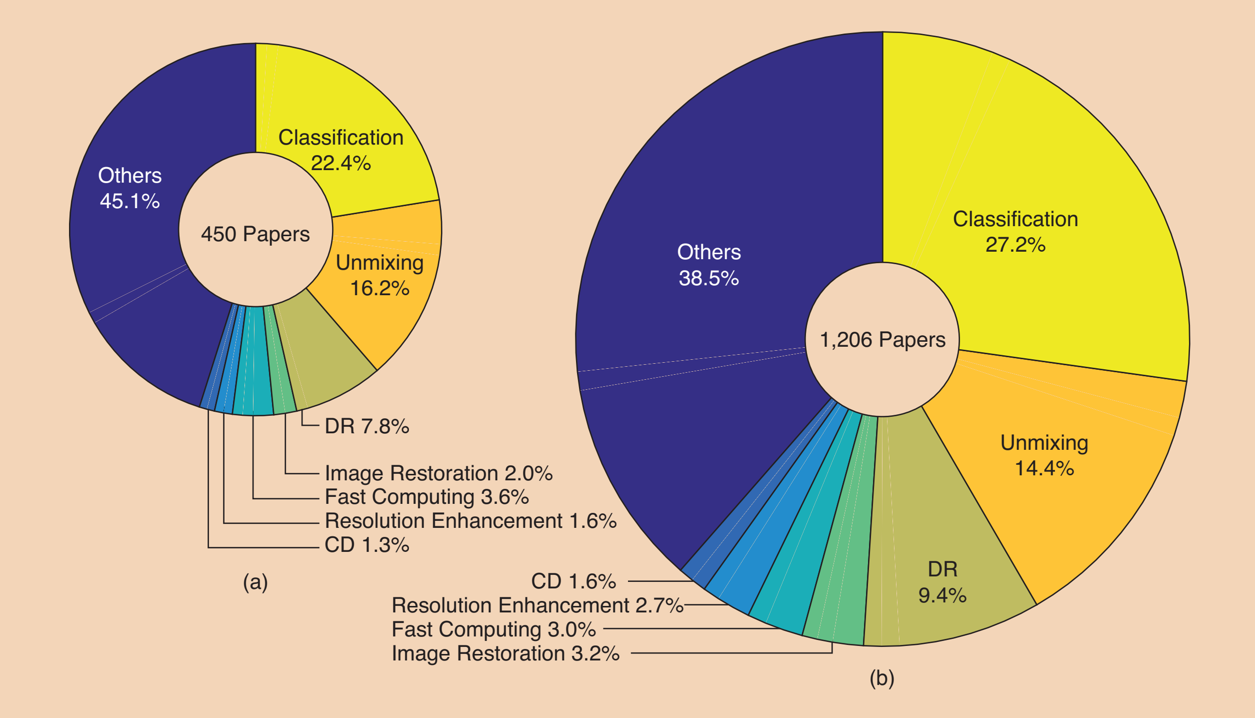

FIGURE 3. Some statistics on articles related to HSIs and signal processing published in IEEE journals during (a) 2009–2012 and (b) 2013–2016. The size of each pie chart is proportional to the number of articles.

or mentioned are rooted mainly in the signal and image processing, statistical inference, and machine-learning fields, with a particular emphasis on methodologies developed since 2013, after the publication of a previous survey article on hyperspectral remote-sensing image analysis [11].

DATA SETS

Throughout this article, three benchmark hyperspectral data sets are referenced: Reflective Optics System Imaging Spectrometer (ROSIS)-03 Pavia University, CASI Houston University, and Hyperion Umatilla County.

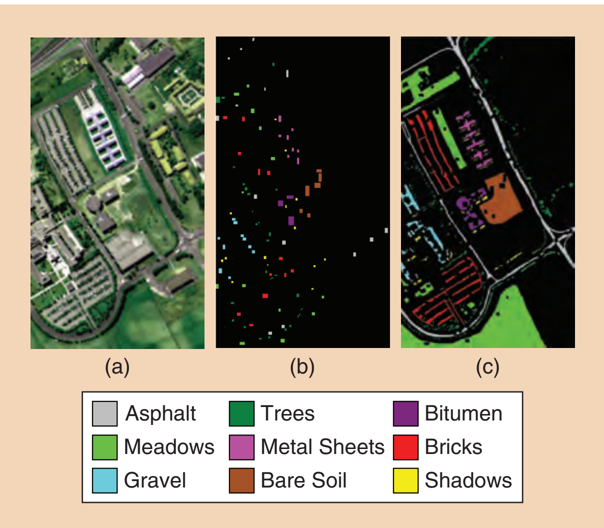

FIGURE 4. ROSIS-03 Pavia University: (a) a false color composite, (b) training samples, and (c) test samples.

PAVIA UNIVERSITY

This data set was captured over the University of Pavia, Italy, by the ROSIS-03 airborne instrument. The flight over the city of Pavia, Italy, was operated by the Deutsches Zentrum für Luft- und Raumfahrt (the German Aerospace Agency) within the context of the HySens project, managed and sponsored by the European Union. The ROSIS-03 sensor has 115 data channels with a spectral coverage ranging from 0.43 to 0.86 μm. Twelve channels have been removed because of the existence of noise, and the remaining 103 spectral channels processed. The data have been corrected atmospherically, but not geometrically. The spatial resolution is 1.3 m per pixel. The data set covers the Engineering School at the University of Pavia and consists of different classes, including trees, asphalt, bitumen, gravel, metal sheet, shadow, bricks, meadow, and soil. The subset data set investigated in this review article comprises 640 $\times$ 340 pixels. Figure 4 presents a false color image of ROSIS-03 Pavia University data and the corresponding training and test samples that have already been separated.

UNIVERSITY OF HOUSTON

This data set was captured by the CASI imager over the University of Houston campus and the neighboring urban area in June 2012. The size of the data is 349 $\times$ 1,905 pixels, with a spatial resolution of 2.5 m. This data set is composed of 144 spectral bands ranging from 0.38 to 1.05 μm. These data consist of

15 classes, including grass healthy, grass stressed, grass synthetic, tree, soil, water, residential, commercial, road, highway, railway, parking lot 1, parking lot 2, tennis court, and running track. Parking lot 1 includes parking garages at the ground level and also in elevated areas, while parking lot 2 corresponds to parked vehicles. Figure 5 shows a three-band false color image and its corresponding already-separated training and test samples.

UMATILLA COUNTY

A pair of real bitemporal Hyperion HSIs acquired on 1 May 2004 ($\mathbf{X}_1$) and 8 May 2007 ($\mathbf{X}_2$) were used to test several selected state-of-the-art CD approaches. This scene covers irrigated agricultural land in Umatilla County, Oregon. The images under consideration have a size of 180×225 pixels. The original Hyperion images contain 242 spectral bands, ranging from 0.35 to 2.58 μm, i.e., VNIR, and SWIR, with a spectral resolution of 0.01 μm and a spatial resolution of 30 m. Preprocessing operations, such as the removal of the uncalibrated and noisiest bands, bad stripes repair, atmospheric correction, and coregistration, have been carried out. Finally, 159 preprocessed bands (i.e., 8–57, 82–119, 131–164, 182–184, and 187–220) out of the original 242 bands were used in the CD experiment. Changes occurring in this scenario include the land cover class transitions between crops, bare soil, subtle variations in soil moisture, and water content of vegetation. More detailed descriptions of this data set can be found in [12]. Figure 6(a) and (b) shows the false color composite of $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. The false color composite of three spectral change vector (SCV) channels is shown in Figure 6(c); possible different changed pixels are illustrated in different colors, whereas the unchanged pixels are in gray. The multiclass change reference map is created based on careful image interpretation, as shown in Figure 6(d). Note that the possible subtle subpixel-level changes (e.g., the one associated with the road surrounding the irrigated agricultural land [12]) are not considered in this article, so that the quantitative comparison with other pixel-level-based approaches could be conducted fairly. Thus, six pixel-level changes were considered, as shown in Figure 6(d).

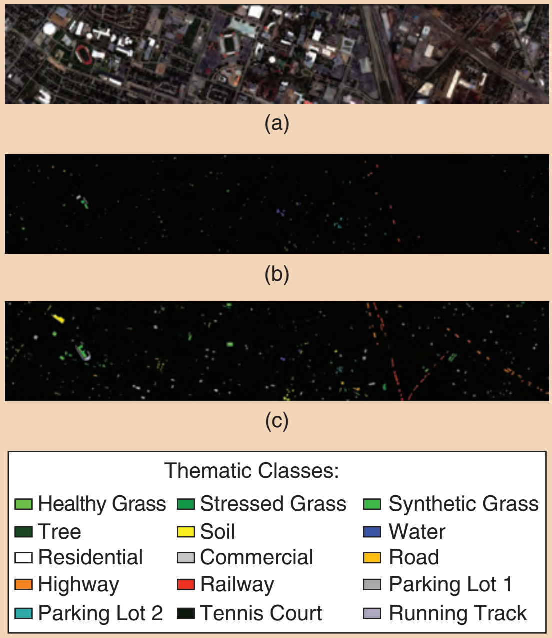

FIGURE 5. CASI Houston: (a) a false color composite (red: band 70, green: band 50, blue: band 20), (b) training samples, and (c) test samples.

DIMENSIONALITY REDUCTION

The increasing spectral resolution of hyperspectral data benefits precision pattern recognition, but it challenges both the memory capacity of ordinary personal computers and conventional signal-processing techniques. For an HSI with a spatial dimension of 600×400 pixels at 16 b-per-band-per-pixel, the data volume becomes 240 MB for 500 spectral bands. The data volume can be linearly increased

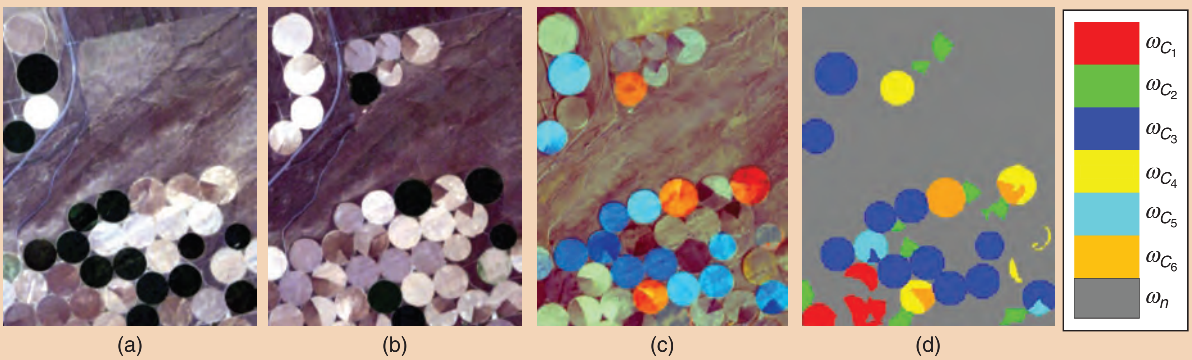

FIGURE 6. Umatilla County: (a) a false color composite (red: 650.67 nm, green: 548.92 nm, blue: 447.17 nm) of the bitemporal EO-1 Hyperion images acquired over an irrigated agricultural area in Umatilla County, Oregon, in 2004 ($\mathbf{X}_1$) and (b) in 2007 ($\mathbf{X}_2$); (c) a composite of three SCV channels (red: 823.65 nm, green: 721.90 nm, blue: 620.15 nm); and (d) a multiclass change reference map, in which six changes are in different colors, whereas the unchanged pixels are in gray.

when time-series hyperspectral data are acquired to monitor environmental changes. The complexities of storing and processing the data will easily exceed the memory capacity of ordinary personal computers. Moreover, as previously discussed, when the ratio between the spectral bands and the number of training samples is high, high-dimensional hyperspectral data suffer from the well-known issue of the curse of dimensionality.

DR, with the goal of identifying and eliminating statistical redundancies of hyperspectral data while keeping as much spectral information as possible, is widely used in hyperspectral data processing. Relatively few bands can represent most of the information in HSIs [13], making DR very useful for storage, transmission, classification, spectral unmixing, target detection [14], and visualization of remote-sensing data [13], [15]. Recent work demonstrates the benefits of using DR when extracting relevant information from HSIs for CD [16], forest management [17], and urban planning [18]. The applications of DR are of interest well beyond hyperspectral data, i.e., for various applications in signal processing and computer vision [19], and wherever interpretation and analysis of high-dimensional data are of interest.

MINIMUM NOISE-FRACTION TRANSFORMATION OBTAINS THE REDUCED FEATURES ACCORDING TO THE IMAGE QUALITY MEASURED BY THE SNR.

Hyperspectral DR consists of both feature selection and feature extraction [13]. Feature selection tries to select a minimal subset of D features $S = \{S_1, S_2, \ldots, S_D\}$ from the original feature set $F = \{F_1, F_2, \ldots, F_d\}$ based on an adopted selection criterion, where $D \le d$ and $S \subseteq F$, while aiming to achieve improved performances for a specific application (e.g., classification, target detection, and so forth). The objective of feature extraction is to find a transformation function $f: \mathbb{R}^d \to \mathbb{R}^D$ that can transform the high-dimensional data point $\{\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^N$ to $\mathbf{z}_i = f(\mathbf{x}_i)$, where $\{\mathbf{z}_i \in \mathbb{R}^D\}_{i=1}^N$ and $D \le d$, such that most information of the high-dimensional data is kept in a much lower-dimensional subspace. The term f can be a linear or nonlinear transformation. Unlike feature selection, feature extraction compresses the high-dimensional original data to generate a small number of new features, where each band often contributes to determining f, as shown in Figure 7. DR methods can be categorized into unsupervised, supervised, and semisupervised approaches, depending on whether the class label information is being used.

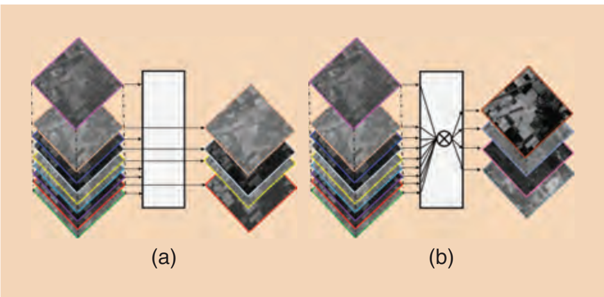

FIGURE 7. Hyperspectral DR: (a) feature selection and (b) feature extraction.

UNSUPERVISED DIMENSIONALITY REDUCTION

Unsupervised DR methods deal with cases where no labeled samples are available and aim to find another representation of the data in the lower-dimensional space by satisfying some given criterion. A variety of unsupervised DR methods have been introduced in the literature. The objective of these methods is not to optimize the accuracy for a given classification task, because they do not consider class-specific information provided by labeled samples. For example, principal component analysis (PCA) [20] reduces dimensionality by capturing the maximum variance in the data. Independent component analysis (ICA) [21] finds the project matrix by maximizing the statistical independence. Minimum noise-fraction (MNF) transformation [22] obtains the reduced features according to the image quality measured by the SNR, and local linear feature extraction (LLFE) [23]–[25] methods seek a projection direction in which neighborhood relationships are preserved in the feature spaces. The nonlinear versions of these methods, such as kernel methods (e.g., kernel PCA, kernel ICA, and kernel MNF [26]) and local methods (e.g., locally linear embedding [27], Laplacian eigenmap, and local tangent space alignment [19]) have been widely used to detect higher-order statistical redundancies. In the same manner, conventional unsupervised feature selection methods for DR select a subset of features from the original data according to a specific criterion, such as linear prediction error [28], entropy [29], or mutual information (by minimizing dependency) [30].

Recently, fusion-based methods and manifold-learning methods have been widely explored for HSI unsupervised DR. Graph-based fusion methods couple data fusion and DR in a unified framework for classification [31], [32]. Borhani and Ghassemian presented a kernel-based method to incorporate spectral and spatial information simultaneously for DR and classification of hyperspectral data [33], while Zhang et al. represented multiple features in a low-dimensional feature space where the complementary information of each feature was exploited by comanifold learning and cograph regularization [34]. In the approaches of [35], manifold learning was exploited for feature extraction and salient band selection of HSIs. In [36], orthogonal total variation component analysis (OTVCA) was proposed, where a nonconvex cost function was optimized to find the best representation for HSIs in a low-dimensional feature space while controlling the spatial smoothness of the features by using a total variation (TV) regularization. The TV penalty promotes piecewise smoothness (homogeneous spatial regions) on the extracted features and thus helps to extract spatial (local neighborhood) information that is very useful for classification. It was shown that OTVCA is highly robust to noise, because it exploits a penalized least-squares minimization framework.

SUPERVISED DIMENSIONALITY REDUCTION

Supervised methods rely on the existence of labeled samples to infer class separability. Several widely used supervised DR methods for HSIs are linear discriminant analysis (LDA) [37], nonparametric weighted feature extraction (NWFE) [38], band selection based on Jeffries–Matsushita (J–M) distance [39], and mutual information [40]. Many extensions of these methods have been proposed in past decades, including modified Fisher's LDA [41], regularized LDA [42], modified NWFE using spatial and spectral information [43], kernel NWFE [44], extended J–M to multiclass cases [40], J–M distance for spatially invariant features [45], minimal-redundancy/maximal-relevance based on mutual information [46], and normalized mutual information [47].

Recent supervised DR methods for hyperspectral data exploit the local neighborhood properties of data. Li et al. [48] employed local Fisher's LDA [49] to reduce the dimensionality of the data while preserving the corresponding multimodal structure. In [50], local neighborhood information was exploited in both the spectral and spatial domains to find a discriminative projection for DR of hyperspectral data. Cao et al. [51] proposed a supervised band selection, by introducing the local spatial smoothness of the HSI into the wrapper method. Dong et al. [52] presented an ensemble-discriminative local-metric-learning method for DR, where local spatial information was incorporated into distance metric learning to learn a subspace, keeping the samples from the same class closer while pushing those from different classes farther away.

Sparse graph embedding (SGE) explores the sparsity structure of the data for hyperspectral DR. Ly et al. [53] proposed block sparse-graph-based discriminant analysis, which learns a block sparse graph for a supervised DR. Xue et al. [54] proposed a spatially and spectrally regularized local discriminant embedding method for DR, where spatial information was integrated into the sparse graph learning process. In [55], a discriminative sparse multimodal learning was developed for multiple-feature selection. However, the sparse coding used in SGE is helpful for learning under conditions where the coding is local [56], which means locality is more important than sparsity. Unfortunately, the converse is not true: sparsity does not always guarantee locality [56]. He et al. [57] proposed a weighted sparse graph to overcome the drawback of sparse coding in SGE, where both the locality and sparsity of the training pixels are integrated.

Other trends in supervised DR methods exploit various algorithms and learning techniques from soft computing, artificial intelligence, and machine learning. Genetic algorithms (GAs) [58], particle swarm optimization (PSO) [59], and the combination of GAs and PSO are used to optimize feature selection [60], [61]. Deep learning techniques, e.g., stacked autoencoders [62] and convolutional neural networks (CNNs) [63], are used for spectral-spatial feature extraction for HSI classification [64], [65].

SEMISUPERVISED DIMENSIONALITY REDUCTION

In real-world applications, labeled data are usually very limited, and labeling a large amount of data may sometimes require considerable human resources or expertise. On the other hand, unlabeled data are available in large quantities at very low cost. For this reason, semisupervised methods [66]–[68], which aim at improved classification by utilizing both unlabeled and limited labeled data, have gained popularity in the machine-learning community. Some of the representative semisupervised learning methods include cotraining [66], transductive support vector machines (SVMs) [67], and graph-based semisupervised learning methods [68].

SOME SEMISUPERVISED FEATURE EXTRACTION METHODS ADD A REGULARIZATION TERM TO PRESERVE CERTAIN POTENTIAL PROPERTIES OF THE DATA.

Some semisupervised feature extraction methods add a regularization term to preserve certain potential properties of the data. For example, semisupervised discriminant analysis (SDA) [69] adds a regularizer into the objective function of LDA. The resulting method makes use of a limited number of labeled samples to maximize class discrimination and employs both labeled and unlabeled samples to preserve the local properties of the data. The approach of [70] proposed a general semisupervised DR framework based on pairwise constraints and employed regularization with sparse representation (SR). A semisupervised pairwise band selection method [71] was proposed for HSIs, in which an individual band selection process was performed only on each pair of classes. Other semisupervised feature extraction methods combine supervised methods with unsupervised ones using a tradeoff parameter, such as semisupervised local Fisher's (SELF) discriminant analysis [72].

It may not be easy, however, to specify the optimal parameter values in these and similar semisupervised techniques, as mentioned in [70] and [72]. Liao et al. [73] proposed a semisupervised local discriminant (SELD) analysis to overcome this problem by combining unsupervised methods (LLFE [23]–[25]) and a supervised method (LDA [37]) in a novel framework without any free parameters. They found an optimal projection matrix that preserves the local neighborhood information inferred from unlabeled samples, while simultaneously maximizing the class discrimination of the data inferred from the labeled samples. The approach of [74] improved SELD [73] by better modeling the differences and similarities between samples. Specifically, this method built a semisupervised graph where labeled samples were connected according to their label information and unlabeled samples by their nearest-neighborhood information.

Graph embedding and manifold-based SR were combined in a semisupervised framework for hyperspectral DR [75], where the sparse coefficients were exploited to construct the graph. Semisupervised manifold alignment [76] and semisupervised transfer component analysis [77] were proposed to find a transformation matrix to project high-dimensional multimodal images into a lower-dimensional feature space, where the geometry of each modality could be preserved.

These semisupervised DR methods try to build a similar objective function, i.e., maximizing class discrimination while at the same time preserving the intrinsic geometric structure of the data. The optimal solutions are acquired by solving generalized eigenvalue problems in the same manner [78]. These methods can be further expanded to shape a manifold learning method by using the kernel trick, similar to the approaches in [79].

EXPERIMENTAL RESULTS

Table 2 and Figure 8 show the performances of some DR methods on the classification of the Pavia University HSIs. We compared the performance when using raw hyperspectral data and seven DR methods (including unsupervised, supervised, and semisupervised DR methods) with three popular classifiers, the parameter settings of which are the same as those in [74]. The training samples were randomly selected from the training set, with the sample size corresponding to different cases: 20, 40, and 80 samples per class, respectively. The results were averaged over ten runs on different numbers of extracted features from one to 30, and the averaged overall accuracy (OA) of the classification was recorded for each method.

The results confirmed that DR can improve classification performance on HSIs. As the size of the training sample increased, classification accuracy increased. Semisupervised DR methods (especially those in [73] and [74], designed for hyperspectral data) outperformed both unsupervised and supervised methods for the one-nearest-neighbor classifier. DR methods that exploit spatial smoothness produced better results, even for unsupervised methods; e.g., OTVCA [36] outperformed the other methods for both random forest (RF) and SVM classifiers in terms of classification accuracy.

TABLE 2. THE DIMENSIONALITY REDUCTION FOR THE CLASSIFICATION OF PAVIA UNIVERSITY DATA, OA% (OPTIMAL NUMBER OF REDUCED FEATURES).

			NUMBER OF LABELED SAMPLES		
DR METHODS		CLASSIFIER	20	40	80
Unsupervised	Raw	1NN	63.14	68.36	69.88
		SVM	67.29	70.15	71.51
		RF	68.72	71.06	73.39
	PCA	1NN	67.38 (11)	69.46 (12)	70.23 (11)
		SVM	70.21 (8)	74.77 (8)	76.81 (9)
		RF	73.46 (11)	73.88 (11)	76.13 (11)
	LPP [24]	1NN	66.97 (18)	69.43 (18)	70.17 (12)
		SVM	71.80 (9)	77.14 (12)	77.95 (18)
		RF	72.46 (18)	74.20 (19)	75.82 (19)
	OTVCA [36]	1NN	71.93 (19)	75.81 (20)	80.39 (23)
		SVM	92.74 (19)	96.32 (21)	97.52 (18)
		RF	**96.24 (23)**	**97.75 (21)**	**98.79 (22)**
Supervised	NWFE [38]	1NN	71.34 (9)	73.34 (9)	74.42 (12)
		SVM	72.29 (8)	76.62 (8)	77.18 (9)
		RF	75.84 (9)	77.03 (11)	78.74 (8)
Semisupervised	SDA [69]	1NN	52.67 (7)	62.64 (8)	70.76 (8)
		SVM	51.62 (8)	63.64 (9)	70.96 (10)
		RF	55.72 (9)	66.97 (9)	71.34 (9)
	SELF [72]	1NN	61.42 (18)	68.75 (18)	69.56 (16)
		SVM	63.93 (19)	75.85 (19)	82.74 (18)
		RF	67.98 (18)	75.97 (12)	80.56 (12)
	SELD [73]	1NN	77.27 (17)	79.03 (10)	81.95 (11)
		SVM	75.71 (11)	76.20 (10)	82.03 (11)
		RF	75.53 (8)	77.50 (9)	82.26 (9)
	SEGL [74]	1NN	79.40 (9)	81.22 (9)	83.57 (10)
		SVM	77.57 (8)	78.16 (8)	82.26 (8)
		RF	76.67 (8)	79.14 (8)	83.51 (9)

LPP: locality preserving projections; SEGL: semisupervised graph learning; 1NN: one nearest neighbor.

We can also see from Figure 8 that, as the number of features increases, the classification performance does not always increase—in fact, some decrease. To achieve optimal classification performance, the number of reduced features needs to be optimized.

CHALLENGES FOR DIMENSIONALITY REDUCTION

Recent advances in sensor technologies and processing techniques strongly support the use of hyperspectral data. Moreover, global Earth observation missions (e.g., AVIRIS from NASA, the PROBA series from the European Space Agency, and the Gaofen series from China) make such data increasingly accessible. Furthermore, at lower altitudes, airplanes and unmanned aerial vehicles can deliver extremely high-resolution hyperspectral data from targeted locations. In addition, image processing techniques allow us to extract multiple-level features from these big hyperspectral data.

Two main challenges remain in hyperspectral DR: 1) mining complementary features (while reducing the dimension and redundancy) from multiple levels of big hyperspectral data and 2) coupling DR and applications in a unified framework, ensuring that optimal features for applications are obtained. Most state-of-the-art research has separated DR and applications into two different steps. For example, morphological operators were employed in [80] to extract low-level features (such as the size and shape of objects) from remote-sensing images. In [81] and [82], middle-level attribute features were extracted from HSIs for land cover mapping. High-level features, such as object-based [83] and so-called deep learning features [84], have been used for CD and classification.

State-of-the-art DR methods typically deal with either lower-level or higher-level features, but not with a combination of both. The features extracted at each level have their own characteristics: high-level features are usually more powerful but less robust, while low-level ones are less informative but more robust. On the other hand, classification is taken as one of the most popular applications to validate DR performances. Hyperspectral classification typically consists of two steps: 1) DR (via either feature extraction or feature selection) and 2) a training procedure for designing the classifier. However, it is difficult to ensure that the best features from the first step will optimize the classification performance of the following one.

CLASSIFICATION

HSI classification is a fast-growing and highly active field of research in the hyperspectral community. A classification algorithm is used to distinguish between different land covers by assigning unknown pixel vectors to one of the classes (or clusters). The individual classes are commonly differentiated based on the similarity to a certain class or by defining decision boundaries constructed in the feature space. The initial set of features for classification usually encompasses spectral channels [4].

With reference to Figure 1, two types of classification approaches can be broadly defined: spectral classifiers and spectral-spatial classifiers [4], where the former consider the HSIs to be a list of spectral measurements with no spatial organization, while the latter classify the input data by taking into account the spatial dependencies of adjacent pixels.

SPECTRAL CLASSIFIERS

Based on the availability of training samples (also referred to as *learning with a teacher*) for the training stage, classification approaches can be grouped into three categories: supervised, unsupervised (also known as *clustering*), and semisupervised approaches.

Supervised approaches classify input data using training samples. These samples are usually collected in one of two ways: 1) by manually labeling a small number of pixels in an image or 2) based on some field measurements. In contrast, unsupervised classification does not consider training samples. The supervised approach classifies input data based only on an arbitrary number of initial cluster centers that may be either user-specified or selected quite arbitrarily. During processing, each pixel is associated with one of the cluster centers, usually in an iterative way, based on a similarity criterion [85], [86]. In semisupervised approaches [87], the training stage is based not only on labeled training samples but also on unlabeled samples.

Because the consideration of training samples leads to higher classification accuracies than in situations where there is no class-specific information, supervised

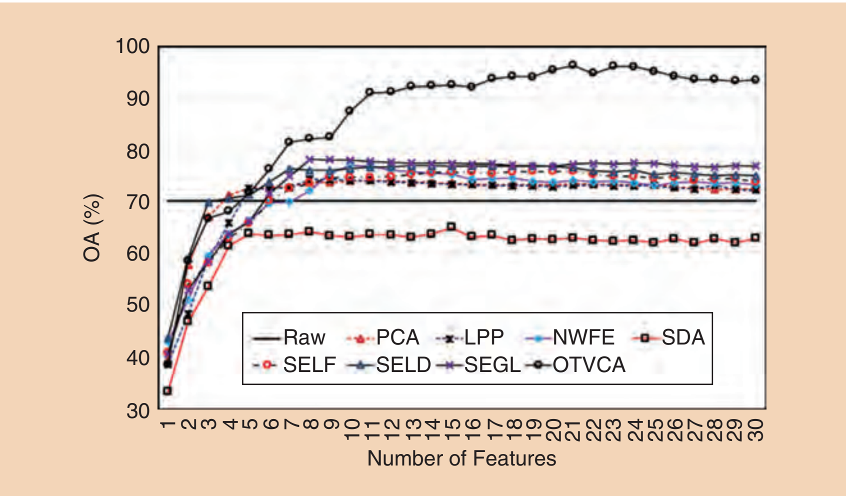

FIGURE 8. The classification performance on the Pavia University data as the number of features increases. Forty labeled training samples per class were randomly selected from the training set with an SVM classifier.

approaches have gained more attention in the hyperspectral community than unsupervised ones. However, the curse of dimensionality is a bottleneck for supervised classification techniques. In theory, a large number of training samples is required to define precise class boundaries in the feature space. This problem intensifies when the number of bands (features) increases. However, in practice, there are not enough training samples to train supervised classifiers, because collecting such samples is time consuming and/or costly. Therefore, classification approaches developed for HSIs need to be able to handle high-dimensional data with only a limited number of training samples.

CLASSIFICATION APPROACHES DEVELOPED FOR HSIS NEED TO BE ABLE TO HANDLE HIGH-DIMENSIONAL DATA WITH ONLY A LIMITED NUMBER OF TRAINING SAMPLES.

The most widely used supervised spectral classifiers have been studied precisely and compared in [88]. Table 3 demonstrates classification accuracies—i.e., OA, average accuracy (AA), and kappa coefficient—obtained on the University of Houston data set by a number of widely used supervised spectral classifiers in the hyperspectral community, including SVM [89], RF [90], rotation forest (RoF) [91], canonical correlation forest (CCF) [92], [93], back-propagation (BP) neural network [94], extreme learning machines (ELMs) [95], kernel ELMs (KELMs) [96], one-dimensional (1-D) deep CNNs [84], and multiple linear regression (MLR) [97]. For the algorithm setup, see [88] and [93].

Currently, a major contribution in the hyperspectral community is based on the use of deep learning for HSI classification. HSIs are highly influenced by various atmospheric scattering conditions, complicated light-scattering mechanisms, interclass similarity, and intraclass variability, which make the hyperspectral imaging procedure inherently nonlinear [65]. Compared to the so-called shallow models, deep learning approaches are expected to potentially extract high-level, hierarchical, and abstract features that are, by nature, more robust when handling the nonlinearities of the input hyperspectral data. Although the use of deep learning in the hyperspectral community is in its early days, some contributions in the community have focused on the use of deep learning for HSI classification. A stacked autoencoder and an autoencoder with sparse constraint were proposed for HSI classification [98], [99], where hierarchical features were extracted from the input data.

Another deep model, the deep belief network, was proposed for the classification of hyperspectral data through the learning of spectral-based features [100]. The critical comparison conducted in [88], specifically on supervised spectral classifiers, offered tantalizing hints about the logical selection of an appropriate classifier based on the application at hand. One of the main conclusions was that there is no classifier that can consistently provide the best performance in terms of classification accuracy when different data sets or different sets of training and test samples are considered. Instead, in addition to the resulting classification accuracies, the consideration of an appropriate classifier should be based on the complexity of the analysis scenario (e.g., the availability of training samples, processing requirements, tuning parameters, algorithm speed, and so forth) and on the considered application domain.

SPECTRAL-SPATIAL CLASSIFIERS

Neighboring pixels in HSIs are highly related or correlated because remote sensors acquire a significant amount of energy from adjacent pixels and homogeneous structures in the image scene are generally larger than the size of a pixel. This is especially evident for images of high spatial resolution. Spatial and contextual data can provide useful information about the shape of different structures. In addition, such information reduces the labeling uncertainty that exists when only spectral information is taken into account and helps to address the salt-and-pepper appearance of the resulting classification map. In general, spectral-spatial classification techniques are composed of three main stages:

1) extracting spectral information (i.e., based on spectral classifiers discussed in the "Spectral Classifiers" section)
2) extracting spatial information (to be discussed later in this section)
3) combining the spectral information extracted during the first stage and the spatial information extracted during the second.

To extract spatial information, two common strategies are available: the crisp neighborhood system and the adaptive neighborhood system. While the former considers spatial and contextual information in a predefined neighborhood system, the latter is more flexible and not confined to a given neighborhood system. In the following two

TABLE 3. SUPERVISED SPECTRAL CLASSIFIERS: CLASSIFICATION ACCURACIES OBTAINED FROM THE UNIVERSITY OF HOUSTON DATA.

CLASS	SVM	RF	ROF	CCF	BP	ELM	KELM	1-D CNN	MLR
OA	80.1	72.9	79.1	83.3	80.9	79.5	80.6	78.2	80.6
AA	83.0	76.9	82.0	85.7	83.1	82.4	82.9	81.2	83.0
Kappa	0.786	0.709	0.775	0.820	0.793	0.778	0.790	0.784	0.790

sections, each neighborhood system is briefly explained. It should be noted that these methods have been elaborated in detail in [4]. Table 4 demonstrates several classification accuracies obtained on the Pavia University data set by different spectral-spatial classification approaches we will now briefly discuss.

CRISP NEIGHBORHOOD SYSTEMS

Markov random fields (MRFs) are a family of probabilistic models that can be described as a two-dimensional (2-D) stochastic process over discrete pixel lattices. MRFs have been widely used to integrate spatial context into image classification problems. In this family of approaches, it is assumed that, for a predefined neighborhood of a given pixel, there is a high probability that its closest neighbors belong to the same object. In [101], a classification framework was introduced by integrating an SVM and MRF. The developed contextual generalization of an SVMs was achieved by analytically relating the Markovian minimum-energy criterion to the application of an SVM. In [102], Ghamisi et al. proposed a spectral-spatial classification approach based on a generalization of the MRF called *hidden MRF* (*HMRF*). In that work, spectral and spatial information was extracted using SVM and HMRF, respectively. Finally, the spectral and spatial information was combined via majority voting within each object. Xia et al. [103] integrated MRFs with the RoF classifier to further improve classification accuracy.

Another way of considering spatial information using the crisp neighborhood system is based on 2-D or 3-D deep CNNs [84]. CNNs consider local connections to deal with spatial dependencies using sharing weights, which can significantly reduce the number of network parameters (compared to its 1-D fully connected version) and extract spatial and contextual information using a predefined crisp neighborhood system [84]. In [104], an unsupervised approach was introduced to learn feature extraction frameworks from unlabeled hyperspectral imagery. This method extracts generalizable features by training on sufficiently large quantities of unlabeled data that are distinct from the target data set. The trained network is then able to extract features from smaller, labeled target data sets and address the curse of dimensionality.

In [65], a self-improving CNN (SICNN)-based approach was proposed for classifying hyperspectral data. This approach solves the curse of dimensionality and the lack of available training samples by iteratively selecting the most informative bands suitable for the designed network. Table 4 demonstrates the classification accuracies obtained by SICNN [65] and 2-D CNN [84] on the Pavia University data set. As can be seen, in all cases, the use of crisp-neighborhood-system-based spectral-spatial classification can improve the classification accuracy of spectral classifiers (e.g., RF and SVM). However, considering a set of crisp neighbors has some disadvantages:

TO ADDRESS THE SHORTCOMINGS OF CRISP NEIGHBORHOOD SYSTEMS, AN ADAPTIVE NEIGHBORHOOD SYSTEM CAN BE CONSIDERED.

1) The crisp neighborhood system may not contain enough samples, which downgrades the effectiveness of the classifier (particularly when the input data set is of high resolution and the neighboring pixels are highly correlated).
2) A larger neighborhood system may lead to intractable computational problems. Unfortunately, the closest fixed neighborhoods do not always accurately reflect information about spatial structures. For instance, they provoke assimilation of regions containing only a few pixels with their larger neighboring structures and do not provide accurate spatial information at the border of regions.
3) In general, the use of a crisp neighborhood system leads to acceptable results for big regions in the scene. Otherwise, it can make small structures in the scene disappear, merging them with larger surrounding objects.
4) They may cause oversmoothing on the border of different classes. This problem, however, has been addressed in [102] using a gradient step.

ADAPTIVE NEIGHBORHOOD SYSTEMS

To address the shortcomings of crisp neighborhood systems, an adaptive neighborhood system can be considered. One approach is to take advantage of different segmentation methods. Image segmentation is the process of partitioning a digital photo into multiple nonoverlapping regions or objects. In image segmentation, a label is assigned to each pixel in the image such that pixels with the same label share certain visual characteristics [105]. These objects provide more information than individual pixels,

TABLE 4. SUPERVISED SPECTRAL–SPATIAL CLASSIFIERS: CLASSIFICATION ACCURACIES OBTAINED FROM THE PAVIA UNIVERSITY HYPERSPECTRAL DATA.

CLASS	RF	SVM	2-D CNN	SICNN	FODPSO	MSF	EMP	EMAP	APDAFE	DBAPDA	RF-EMEP	ROF-EMEP	GCK	MFL
OA	71.3	78.8	78.8	83.4	88.1	91.1	77.7	90.7	97.0	98.0	96.1	96.3	98.0	97.5
AA	82.2	87.0	79.7	83.0	92.0	94.8	82.5	91.4	96.7	98.1	96.6	97.9	97.4	97.1
Kappa	0.648	0.735	0.734	0.778	0.848	0.880	0.710	0.877	0.960	0.974	0.949	0.952	0.974	0.967

MSF: minimum spanning forest; APDAFE: attribute profiles with discriminant analysis feature extraction; DBAPDA: decision boundary feature extraction and attribute profiles and discriminant analysis feature extraction.

because the interpretation of images based on objects is more meaningful than interpretations based on individual pixels.

For spectral-spatial classification of HSIs using segmentation approaches, there are usually two methods to consider: 1) segmentation and classification maps can be integrated using majority voting within each object by assigning the whole object to the most frequent classification label within that particular object [106] (such majority voting is described in [107]) and 2) segments can be considered to be input vectors for supervised classification [108]. In [109], however, a reverse view was employed, where markers for spatial regions were automatically obtained from classification results and then used as seeds for region-growing in the segmentation step. The classification accuracy of this segmentation method with an extra step (where the classification map is refined using the results of a pixelwise classification and a majority voting within the spatially connected regions) is shown in Table 4 as the minimum spanning forest.

EPS SOLVE THE MAIN ISSUE OF CONVENTIONAL APS, THE INITIALIZATION OF THE THRESHOLD VALUES.

One common way to segment an image is based on histogram thresholding. A commonly used exhaustive search for optimal thresholds in terms of between-class distances is based on the Otsu criterion [110]. The approach is easy to implement, but it has the disadvantage of being computationally expensive. An exhaustive search for an n-level segmentation (i.e., $n-1$ optimal thresholds) involves evaluations of the fitness of $n(L-n+1)^{n-1}$ combinations of thresholds, where L shows the number of intensity values. Therefore, this method is not suitable from a computational cost perspective. The task of determining $n-1$ optimal thresholds for n-level image thresholding could be formulated as a multidimensional optimization problem. In [106], a thresholding-based segmentation method was proposed, where an evolutionary-based optimization technique, called *fractional order Darwinian PSO* (*FODPSO*), sought to find the best set of thresholds with the highest between-class distance. The classification accuracy obtained by this segmentation method is given in Table 4. This method is very fast, even for large data sets, because it works on the image histogram instead of the image space.

Morphological profiles (MPs) are another set of approaches based on adaptive neighborhood pixels. MPs comprise a number of features constructed by applying a set of openings and closings by reconstruction with a structuring element (SE) of increasing size [111]. The result of the basic extension of the MP, extended MP (EMP), that is applicable to HSIs is shown in Table 4. Although MPs are a powerful approach for extracting spatial information, the concept suffers from some limitations:

- The SE shape is fixed, which imposes a constraint on model spatial structures within a scene.
- SEs are unable to characterize information about the gray-level characteristics of the regions, such as spectral homogeneity, contrast, and so on.

To address these MP shortcomings, [112] introduced a morphological attribute profile (AP) that provides a multilevel characterization of an image by using the sequential application of morphological attribute filters. A comprehensive survey on the use of APs for HSI classification can be found in [113] and [4]. The classification accuracy obtained by the extension of APs on HSIs, known as *extended multi-AP* (*EMAP*), is given in Table 4. There are two main difficulties of using the AP, however: not knowing 1) which attributes lead to a better discrimination ability for different classes and 2) which threshold values should be considered to initialize each AP. To solve these issues, several articles, such as [114]–[116], have tried to introduce automatic techniques for the use of APs. In [116] and [115], automatic spectral-spatial classification methods were proposed based on the use of EMAP and supervised/unsupervised feature extraction approaches. The classification accuracy of the APs with discriminant analysis feature extraction [116] and decision boundary feature extraction and APs and discriminant analysis feature extraction [115] are shown in Table 4.

MPs and APs produce extremely redundant features. To address this issue, a sparse classification using both spectral and spatial information was investigated in [117]. In [118], the performance of different feature extraction approaches, including linear, nonlinear, and manifold approaches, was investigated to generate base images for constructing EMAPs.

To further improve the conceptual capability of the AP and the corresponding classification accuracies, Ghamisi et al. proposed extinction profiles (EPs) in 2016 [119] by considering a set of connected idempotent filters and extinction filters. In contrast to the AP, the EP preserves the height of the extrema [119] and, as a result, shows better capability than the AP in terms of simplification for recognition. This advantage leads to higher classification accuracy for EPs than for APs. In addition, the EPs' parameters can be set automatically, independent of the kind of the attribute being used (e.g., area, volume, and so on). In other words, EPs solve the main issue of conventional APs, the initialization of the threshold values [119]. In [120], the concept of EPs was generalized to extract spatial and contextual information from HSIs, known as *extended multi-EP* (*EMEP*). The classification accuracy of EMEP using RF (RF-EMEP) and RoF (RoF-EMEP) is presented in Table 4.

COMPOSITE KERNELS

The main problem associated with the concept of spectral-spatial feature extraction approaches is that they usually increase the number of features, while the number of training samples remains the same. This can lead to the curse of dimensionality and high executable processing time. This problem has partially been addressed by combining different kernels for spectral and spatial information (i.e., composite kernels) [121] in the SVM classification process. However, classification using composite kernels and SVMs demands convex combination of kernels and a time-consuming optimization process. Therefore, the approach has been modified to deal with convex combinations of kernels through generalized composite kernels (GCK) [122], the results of which are shown in Table 4, and multiple-kernel learning [123]. In [124], a classification framework was introduced that combines multiple features with the linear and nonlinear class boundaries present in the data without requiring any regularization parameters to control the weights of the considered features (the results are shown as MFL in Table 4).

SEMISUPERVISED AND ACTIVE LEARNING

As previously discussed, the number of training samples is usually limited because the collection of such samples is either expensive or time consuming. In such situations, the limited number of training samples available may not be representative of the statistical distribution of the data, which can downgrade the quality of the classification map obtained by supervised classifiers. To partially address this issue, active learning, which aims to find the most informative training set, has gained popularity in the hyperspectral community.

Active learning starts an iterative process with a small and suboptimal initial training set and then selects a few additional samples from a large quantity of unlabeled samples. Active learning considers the result of the current model, ranking the unlabeled samples according to a criterion that allows selection of the most informative samples to improve the model, thus minimizing the number of training samples while preserving discrimination capabilities as much as possible [125]. For a complete survey on the use of active learning for remote-sensing image analysis, see [126] and [127].

Active learning and semisupervised learning share a similar conceptual background as both types of learning try to address the issue of limited labeled samples. In this manner, both approaches start with a small set of labeled samples and a large set of unlabeled data. Active learning usually requires a labor-intensive labeling process, while semisupervised learning, although avoiding manual labeling by assigning pseudolabels to unlabeled data, may introduce incorrect pseudolabels and consequently downgrade classification performance [128]. Although active learning and semisupervised learning follow different work flows, they both aim to make the most of unlabeled data while reducing manual labeling efforts [125]. Therefore, it is common to use both of these strategies to make the most of these two paradigms for HSI classification. In [128], active learning and semisupervised learning were collaboratively integrated to form an approach called *collaborative active and semisupervised learning* that improves pseudolabeling accuracy and thus facilitates semisupervised learning. This method was based on spectral information. In [125], active learning and hierarchical segmentation were combined for spectral-spatial classification of HSIs.

SPARSE REPRESENTATION CLASSIFICATION

SR classification (SRC)-based approaches with dictionary-based generative models [129], [130] have received considerable attention in the hyperspectral community. In this context, an input signal is represented by a sparse linear combination of samples (atoms) from a dictionary [129], where the training data are generally used as the dictionary. The main advantages of such approaches are that SRC avoids the heavy training procedure usually conducted by a supervised classifier and that the classification is performed directly on the dictionary. Classification can be improved by incorporating contextual information from the neighboring pixels into the classifier. This can be performed indirectly by exploiting the spatial correlation through a structured sparsity imposed earlier in the optimization process. If an adequate number of training samples is available, discriminative as well as compact class dictionaries can also be developed to improve classification performance [131].

THE RICH SPECTRAL RESOLUTION AVAILABLE IN HYPERSPECTRAL DATA CUBES CAN BE USED TO UNMIX HYPERSPECTRAL PIXELS.

CHALLENGES IN CLASSIFICATION

The main challenges for HSI classification are not particularly related to methodology. They are, rather, related to the lack of appropriate benchmark data sets and the corresponding training and test samples. As can be seen in Figure 3, most published contributions in the hyperspectral community are dedicated to HSI classification. The approaches are often capable of producing very accurate classification maps on the widely used Indian Pines and Pavia data sets, which makes real comparison of the approaches almost impossible. In other words, the existing data sets have already been saturated in terms of classification accuracies. Therefore, our community is in urgent need of more complex data sets to share (e.g., highly nonlinear data sets with greater area coverage that are composed of many classes). In addition, a standard set of training and test samples should be defined for each particular data set, to make the proposed approaches fully comparable with each other.

SPECTRAL UNMIXING

Spectral unmixing has been an alluring exploitation goal since the early days of HSI processing [132]. Mixed pixels are common in remotely sensed HSIs because of the imaging spectrometer's insufficient spatial resolution or due to intimate mixing effects. However, the rich spectral resolution available in hyperspectral data cubes can be used to unmix hyperspectral pixels. In fact, mixed pixels can also be obtained with high-spatial-resolution data because of intimate mixtures. This means that increasing the spatial resolution often does not solve the problem.

In other words, the mixture problem can be approached in a macroscopic fashion, which means that only a few macroscopic components and their associated abundances should be derived. However, intimate mixtures happen at microscopic scales, thus complicating the analysis with nonlinear mixing effects [133]. In addition to spectral mixing effects, there are many other interfering factors that can significantly affect the analysis of remotely sensed hyperspectral data. For instance, atmospheric interferers are a potential source of errors in spectral unmixing. Multiple scattering effects can also lead to model inaccuracies.

THE PIXEL PURITY INDEX IS PERHAPS THE MOST POPULAR ENDMEMBER EXTRACTION ALGORITHM BECAUSE OF ITS AVAILABILITY IN SOFTWARE PACKAGES.

In linear spectral unmixing, the macroscopically pure components are assumed to be homogeneously distributed in separate patches within the field of view. In nonlinear spectral unmixing, the microscopically pure components are intimately mixed. A challenge is how to derive the nonlinear function, because nonlinear spectral unmixing requires detailed a priori knowledge about the materials. Responding to this limitation, a vast majority of techniques have focused on linear spectral unmixing, where the goal is to find a set of macroscopically pure spectral components (called *endmembers*) that can be used to unmix all the other pixels in the data. Unmixing thus amounts to finding the fractional coverage (abundance) of each endmember in each pixel of the scene, which can be approached as a geometrical problem [134]. In the following section, we focus on the most relevant parts of the linear spectral unmixing chain. We also summarize the main efforts in nonlinear spectral unmixing.

ESTIMATION OF THE NUMBER OF ENDMEMBERS

Determining the number of pure spectral endmembers in HSIs is a challenging problem. One of the most commonly used approaches to this problem is the virtual dimensionality (VD) method [135], which follows the pigeon-hole principle. If we represent a signal source as a pigeon and a spectral band as a hole, we can use a spectral band to accommodate one source. Thus, if a signal source is present in our remotely sensed hyperspectral data set, we should be able to detect this particular source in the relevant spectral band. This can be accomplished by calculating the eigenvalues of both the data-correlation and covariance matrices. A source is present if their difference is positive.

Another popular approach is hyperspectral signal identification with minimum error (HySime) [136]. The idea of HySime is to find the first k eigenvectors that contain the most data information, i.e., to find k such that the mean-square error (MSE) between the original data and their projection onto the eigenvector subspace is minimized. Subspace k is ranked in terms of data variance, but noise variance is not unitary in different directions, and the contribution from signals may be smaller than from noise. HySime addresses this issue by using subspace projection techniques, thus contributing an additional feature with regard to VD: the modeling of noise before the estimation.

The eigenvalue likelihood maximization (ELM) method [137], in turn, implements a modification of the VD concept based on the following observations: 1) the eigenvalues corresponding to the noise are identical in the covariance and the correlation matrices, and 2) the eigenvalues corresponding to the signal (the endmembers) are larger in the correlation matrix than in the covariance matrix. The ELM takes advantage of this fact and provides a fully automatic method that does not need an input parameter (as does VD) or estimation of the noise (as does HySime).

Finally, the normal compositional model (NCM) [138] addresses the possibility that, in real images, there may not be any pure pixels. To address this issue, NCM assumes that the HSI pixels are linear combinations of an unknown number of random endmembers (the opposite of the deterministic approach). This model provides more flexibility with respect to the observed pixels and the endmembers, which are allowed to be a greater distance from the observed pixels.

ENDMEMBER EXTRACTION

The identification of endmembers is a challenging problem, for which many different strategies have been proposed [134]. To categorize algorithms, we consider three different scenarios.

1) The data contain at least one pure pixel per endmember, i.e., there is at least one spectral vector in each vertex of the data simplex (pure pixel assumption).
2) The data do not contain pure pixels but contain enough spectral vectors on each facet. In this case, we may fit a minimum volume simplex to the data.
3) The data are highly mixed, with no spectral vectors near the facets. In this case, minimum volume algorithms fail, and we need to resort to a statistical framework. We also consider algorithms that include spatial information in addition to spectral information for this purpose.

PURE PIXEL ASSUMPTION

Pure pixel methods assume a classic spectral unmixing chain with three stages: DR, endmember selection, and abundance estimation. Here, the endmembers are directly derived from the original hyperspectral scene. The pixel purity index (PPI) [139] is perhaps the most popular endmember extraction algorithm because of its availability in software packages. PPI has many parameters involved and is not an iterative algorithm. Manual intervention is required to select a final set of endmembers, which makes it unattractive for automatization purposes.

An alternative is the N-FINDR [140], which assumes the presence of pure pixels in the original hyperspectral scene and further maximizes the volume that can be formed with pixel vectors in the data cube. Orthogonal subspace projection (OSP) [141], in turn, uses the concept of orthogonal projections. Vertex component analysis (VCA) [142] iteratively projects data on a direction orthogonal to the subspace spanned by the endmembers, which have already been determined. In this regard, the algorithm is similar to OSP; the main difference is that VCA applies a noise characterization process to reduce the sensitivity to noise. This is accomplished by using singular value decomposition (SVD) to obtain the projection that best represents the data in the maximum-power sense.

Another important concept in this category is the endmember bundle, explored by algorithms such as multiple endmember spectral mixture analysis (MESMA) [143]. Although the shape of an endmember is fairly consistent, its amplitude generally varies because of illumination conditions, spectral variability, topographic modulation, and other circumstances. MESMA addresses this issue using endmember bundles, which incorporate variability by representing each endmember by a set or bundle of spectra, each of which could reasonably be the reflectance of an instance of the endmember.

MINIMUM VOLUME ALGORITHMS

If the data do not contain any pure signatures, we can fit a simplex of minimum volume in cases where we have enough spectral vectors on the facets. This idea is the opposite of the concept of maximum volume adopted by N-FINDR; here, the goal is to find the simplex with the minimum volume that encloses the data. From an optimization point of view, these algorithms are formulated by including a data term that minimizes the reconstruction error and a volume term that promotes mixing matrices of minimum volume. This is the case for the iterative constrained endmember (ICE) [144] and minimum volume constrained nonnegative matrix factorization approaches [145], the main differences of which are related to the way they define the data volume term.

The sparsity-promoting ICE approach [146] is an extension of the ICE algorithm, which incorporates sparsity-promoting priors aimed at finding the number of endmembers. The minimum volume estimation (MVES) algorithm [147] integrates concepts of convex analysis and volume minimization to provide a solution similar to that of previously mentioned algorithms but using cyclic minimization with linear programming. Again, the assumption is that the enclosing simplex with minimum volume should coincide with the true endmember simplex (MVES uses hard positivity constraints).

The minimum volume simplex analysis algorithm [148] follows a similar strategy but allows violations of the positivity constraint. This is because, due to the presence of noise or perturbations, spectral vectors may lie outside the true simplex, and this may introduce errors in the characterization. The minimum-volume enclosing algorithm for endmember identification and abundance estimation [149] also exploits this concept, allowing a certain number of outliers when estimating the minimum volume that encompasses the HSI.

HIGHLY MIXED DATA

When the spectral mixtures are highly mixed, geometrical-based methods yield poor results because there are not enough spectral vectors in the simplex facets. Statistical methods are a powerful alternative that usually comes with a price: higher computational complexity than with geometrical methods. Because, in most cases, the number of substances and their reflectances are not known, the problem can be approached as a blind-source separation problem, with some statistical unmixing approaches proposing variations on the ICA [150]. However, ICA applicability is compromised by the statistical dependence existing among abundances. This has been addressed (among other strategies presented in the recent literature) by the dependent component analysis algorithm [151]. Bayesian approaches have also been used because they can model statistical variability and impose priors to constrain solutions to physically meaningful ranges.

ONCE THE ENDMEMBER SIGNATURES HAVE BEEN DERIVED, DIFFERENT STRATEGIES CAN BE USED TO ESTIMATE THEIR FRACTIONAL ABUNDANCES.

INCLUSION OF SPATIAL INFORMATION

Most available algorithms for endmember identification do not consider information about spatial-contextual information. In certain scenarios, it is important to include the spatial information in the analysis. Automatic morphological endmember extraction [152] uses extended morphological transformations to integrate spatial and spectral information. However, spatial-spectral endmember extraction [153] uses a different approach. First, it processes the image using a local search window and applies SVD to determine a set of eigenvectors that describe most of the spectral variance in the window. Then, it performs a projection of all of the image data onto eigenvectors to determine candidate endmember pixels. Finally, it uses spatial constraints to combine and average spectrally similar candidate endmember pixels (preserving similar but distinct endmembers that occupy unique image regions).

To avoid modifying spectral-based algorithms for endmember extraction, spatial information can also be included as a preprocessing module, such as the spatial preprocessing algorithm [154]. A region-based approach [155] has also been developed to adaptively include spatial information. Finally, a spatial-spectral preprocessing approach [156] has been developed to derive a spatial homogeneity index that uses Gaussian filtering and is thus relatively insensitive to the noise present in the hyperspectral data.

ABUNDANCE ESTIMATION

Once the endmember signatures have been derived, different strategies can be used to estimate their fractional abundances [157]. The idea is to find the abundances that minimize the reconstruction error obtained after approximating the original hyperspectral scene using a linear mixture model. However, this is generally an unconstrained solution that does not satisfy the abundance nonnegativity constraints (ANCs) and the abundance sum-to-one constraints (ASCs). Whether or not abundance constraints should be imposed depends on the practical application. It has been argued that, if the linear mixture model is accurate, the two constraints should be satisfied automatically. In any event, the ANC is more important than the ASC. Due to noise and spectral variability, reinforcing the ASC may be prone to introducing additional estimation error.

SPECTRAL UNMIXING ALGORITHMS WITH THE PURE PIXEL ASSUMPTION REQUIRE THE PRESENCE OF PURE PIXELS IN THE SCENE FOR ENDMEMBER EXTRACTION.

When endmembers are unknown, endmember signatures should be extracted or estimated first. Some endmember extraction algorithms can provide abundance estimates simultaneously (e.g., algorithms without the pure pixel assumption) [134]. Another group of abundance estimation approaches based on blind-source separation, which does not require endmember signatures to be known a priori, has been developed [134]. Widely used matrix-factorization-based blind-source separation methods include ICA and nonnegative matrix factorization (NMF), which have been mostly used in the context of unsupervised (soft) classification. If the linear assumption does not hold, then nonlinear unmixing techniques should be used. In addition, as mentioned in the previous section, if spectral variability or endmember variability is being considered, the mixture model must be modified accordingly, which is traditionally accomplished by generating modified/extended linear mixture models [158]–[160].

SPARSE UNMIXING

Spectral unmixing algorithms with the pure pixel assumption require the presence of pure pixels in the scene for endmember extraction. Due to spatial resolution and mixing phenomena, this assumption cannot always be guaranteed. Spectral unmixing algorithms without the pure pixel assumption generate endmember signatures that often do not relate to real physical signatures. A possible solution is to use ground spectral libraries to perform spectral unmixing, but libraries are very large, and hence the problem becomes sparse and difficult to solve. Another problem is the difference between the ground library and the image data. To address these issues, sparse unmixing [161] expresses pixel vectors as linear combinations of a few pure spectral signatures obtained from a potentially very large spectral library of ground materials. An advantage is that it sidesteps the endmember extraction step (including the estimation of the number of endmembers).

To incorporate spatial information into the spectral unmixing formulation, a TV regularizer has been developed to enforce spatial homogeneity by including this term in the original objective function [162]. It produces spatially smooth abundance fractions that improve sparse unmixing performance, even in very high-noise conditions. Furthermore, because it is generally observed that spectral libraries are organized in the form of groups with different variations of the same component (e.g., different mineral alterations), exploiting the inherent group structure present in spectral libraries can improve the results of sparse unmixing by selectively enforcing groups. For this purpose, a group-based formulation of sparse unmixing has been introduced [163].

A further development has also been recently introduced based on the concept of collaborativity, which promotes solutions with a minimum number of active endmembers (the number of endmembers in a scene is generally low). This allows the number of endmembers participating in the final solution to be minimized [164], while also partially circumventing the need to estimate the number of endmembers in the scene [165].

NONLINEAR UNMIXING

Apart from the linear spectral unmixing algorithms, several alternatives dealing with the mixture problem can be found in the literature [166]–[168]. The bilinear mixing model (BMM) considers secondary illumination sources. This model represents a simplification of reality, as it only considers bilinear interactions (objects that can be illuminated by light reflected first by another object) [133]. BMMs can be generalized to deal with multiple endmembers, by trying to model bilinear interactions as new endmembers. The generalized bilinear model (GBM) provides more flexible solutions that can avoid some overfitting problems associated with bilinear models, but it still assumes that there are no self-bilinear interactions [133].

The BMM and GBM have inspired several posterior methodologies: some use polynomial functions to model the nonlinearities provided by layer partitions and scattering properties in multilayered scenarios [169], while others use different approaches to solve the proposed model [170]–[173]. In [174], the authors proposed a modification of the BMM in which the transformation of the spectral information is based on a two-degree polynomial, thus defining the polynomial postnonlinear model (PPNM), a main advantage of which is that it is able to deal with self-interactions; the PPNM has been recently extended to deal with all polynomial degrees [175]. In [169], a p-linear polynomial model is used to characterize nonlinearities in combination with a supervised artificial neural network (ANN) and polytope decomposition scheme. When the interactions are considered to exist at the photon level, nonlinear unmixing methods try to model the optical characteristics of the intimate mixture from a theoretical analysis of the reflectance behavior attending the specific geomorphical, chemical, and physical properties of the observed data. Several models have been proposed in this context, but the Hapke model [176] is still the most widely adopted approach.

In addition, supervised techniques based on the use of methods such as kernel-based ones and ANNs have been proposed to perform unmixing on a microscopic scale [177]–[186]. The main advantage of these techniques is that they can create a preliminary model of every nonlinear behavior. However, the need for reliable ground truth information about the training samples represents a major shortcoming.

CHALLENGES IN SPECTRAL UNMIXING

Despite the availability of several consolidated techniques for linear spectral unmixing, as well as a suite of incipient techniques for nonlinear spectral unmixing, an important challenge remains related to the nonlinearity of mixing phenomena in HSIs. The inherently nonlinear nature of the process and the dependence on the observed objects create the need to incorporate detailed information about the observed objects to properly model the multiple scattering phenomena occurring in the nonlinear case. The estimation of participating endmembers in the mixture also remains a challenge, despite the availability of some techniques that can provide reasonable approximations.

Another important hurdle for the automatic execution of sparse unmixing algorithms is the heterogeneity in available spectral libraries, although important efforts have been made toward the development of open-source libraries of spectral materials that can alleviate the need to estimate the number and the signatures of spectral endmembers in advance. In this regard, the availability of open libraries such as SPECCHIO (http://www.specchio.ch) provides an important first step toward the general use of open spectral libraries for spectral unmixing purposes.

RESOLUTION ENHANCEMENT

HSI resolution enhancement has attracted increasing attention in recent years, as shown in the statistical analysis of the trend described in the introduction. Resolution enhancement techniques for HSIs can be broadly categorized into four classes, as shown in Figure 9:

1) hyperspectral superresolution (multi-image/single-image)
2) subpixel mapping (or superresolution mapping)
3) hyperspectral pan-sharpening
4) hyperspectral–multispectral (HS–MS) data fusion.

The first class, hyperspectral superresolution, is an extension to HSIs of ordinary superresolution in computer vision. A high-resolution HSI is reconstructed from multiple low-resolution HSIs (or a single HSI) acquired by a single sensor. The second, subpixel mapping, is a resolution enhancement technique processed at a classification level using a single HSI as input. The remaining two classes, hyperspectral pan sharpening and HS–MS fusion, are multisensor superresolution techniques, in which an HSI is fused with an auxiliary higher-resolution data source (panchromatic or multispectral images) taken over the same area at the same time (or a similar period) to create a high-resolution HSI. The following sections provide an overview of recent advances in these four classes of techniques.

HYPERSPECTRAL SUPERRESOLUTION

A variety of superresolution image reconstruction techniques have been intensively pursued over the past three decades in computer vision. Superresolution techniques can be roughly divided into two types [187]: 1) classical multi-image superresolution, which obtains a high-resolution image (or sequence) from multiple low-resolution images for the same scene with different subpixel shifts, and 2) learning-based superresolution, which learns correspondences between

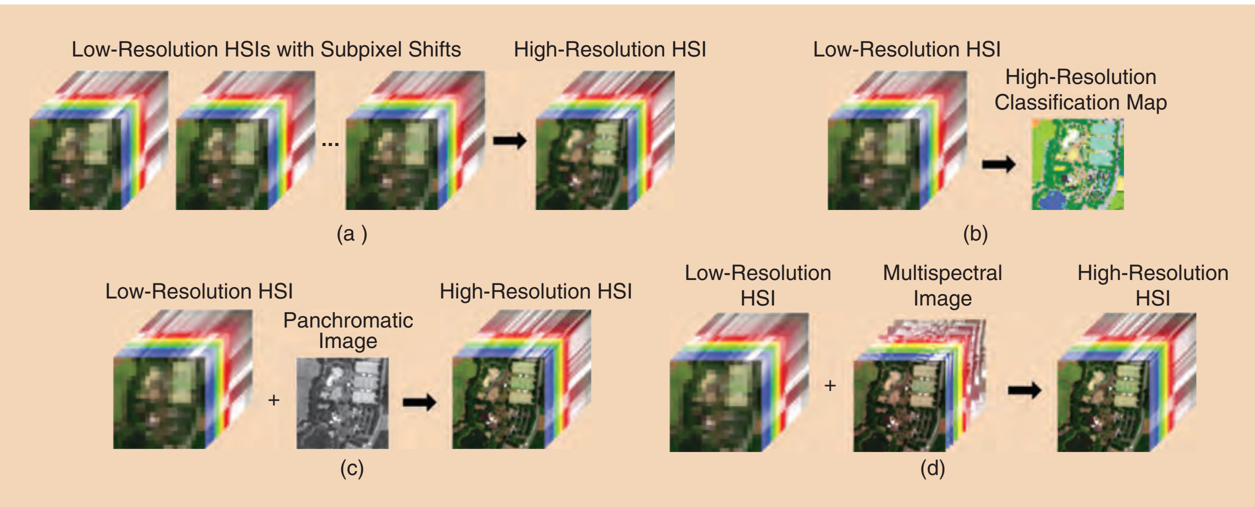

FIGURE 9. The four classes of resolution enhancement techniques for HSIs: (a) multi-image hyperspectral superresolution, (b) subpixel mapping, (c) hyperspectral pan-sharpening, and (d) HS–MS fusion.

low- and high-resolution image patches from an external training database.

Multi-image superresolution techniques have been extended to HSIs. Depending on the type of subpixel shifts in the low-resolution image, these techniques can be divided into two approaches: the first uses multiple HSIs acquired by the same sensor over the same scene, similar to classical multi-image superresolution; the second takes advantage of band-to-band misregistration (so-called keystone) in a single HSI. Akgum et al. first applied the multi-image superresolution technique to multiple HSIs based solely on simulations [188].

SUBPIXEL MAPPING TECHNIQUES HAVE BEEN ACTIVELY STUDIED USING MULTISPECTRAL IMAGES AND EXTENDED TO HSIS BY EXPLOITING RICH SPECTRAL INFORMATION.

In the remote-sensing community, multi-image hyperspectral superresolution has been mainly studied using multi-angular HSIs obtained by the Compact High-Resolution Imaging Spectrometer (CHRIS). In [189], conventional superresolution techniques were applied to a set of multi-angular CHRIS images. Here, Chan et al. developed a new superresolution technique for multi-angular HSIs based on a thin-plate spline nonrigid transform model. This approach was intended to improve the image registration procedure and demonstrated its effectiveness with experiments using the CHRIS images [190]. The impact of multi-angular superresolution on classification and unmixing applications was further investigated in [191] and [192].

Qian and Chen developed a technique for the second approach that enhances the spatial resolution of HSIs by exploiting the keystone characteristics [193]. Keystone is band-to-band misregistration in the cross-track direction caused by optical aberrations and misalignments in pushbroom systems. Different band images, including different subpixel miregistrations, can be used as input images in the multi-image superresolution framework. The advantage of this method is that it requires only a single HSI, but the limitation is that the spatial resolution can be enhanced only in the cross-track direction.

Zhao et al. proposed SR-based algorithms for learning-based hyperspectral superresolution [194], [195]. The high-resolution version of a given low-resolution patch is recovered by solving the sparse linear inverse problem with spectral regularization based on spectral unmixing, in which the patch dictionary is learned from a set of high-resolution panchromatic images or HSIs. Patel et al. developed a learning-based superresolution method for HSIs using an adaptive wavelet designed from training HSIs [196]. These learning-based superresolution techniques do not require multiple images over the same scene but instead require an external training database with target resolution. For the upcoming hyperspectral satellites, it is realistic to use images obtained by operational multispectral satellites (e.g., Sentinel-2) for the training database.

SUBPIXEL MAPPING

Subpixel mapping is a technique for enhancing the spatial resolution of spectral images by dividing a mixed pixel into subpixels and assigning land cover classes to these subpixels [197]. Subpixel mapping techniques have been actively studied using multispectral images and extended to HSIs by exploiting rich spectral information. Subpixel mapping is made up of two steps: 1) estimating fractional abundances of classes (or endmembers) at a coarse resolution by soft classification or spectral unmixing and 2) determining the subpixel location of each class within a pixel, assuming spatial dependence. Many recent advances in subpixel mapping of HSIs aim at improving the accuracy of inverse problems involving these two steps.

Tong et al. proposed a method that exploits not only attraction but also repulsion between subpixels to better retrieve spatial dependence [198]. Zhang et al. integrated learning-based superresolution into subpixel mapping, requiring an external training set [199]. The estimation of unknown spatial details of classes at a high resolution from a single low-resolution image is a typical ill-posed inverse problem. Different techniques for spatial regularization that have been recently studied to mitigate the ill-posedness have assumed spatial prior models, such as Laplacian, total variation, bilateral total variation, and nonlocal total variation [200], [201].

Obviously, the accuracy of the abundance maps obtained in the first step greatly affects the final performance of subpixel mapping. To address this issue, Tong et al. proposed a GA that can correct possible errors in the initial estimation of abundances using a mutation operator [202]. Xu et al. introduced a method that improves the accuracy of spectral unmixing in subpixel mapping by taking endmember variability into consideration, based on the assumption that different representative spectra for each endmember are available [201].

HYPERSPECTRAL PAN-SHARPENING

Pan sharpening is a technique that enhances the spatial resolution of multispectral imagery by fusing this imagery with a higher-resolution panchromatic image. Hyperspectral pan-sharpening is an extension to HSIs of conventional pan-sharpening and is also a special case of HS–MS fusion. Naturally, there are two main approaches: 1) extensions of pan-sharpening methods and 2) subspace-based methods (originally developed for HS–MS fusion).

Hyperspectral pan-sharpening is motivated by spaceborne imaging spectroscopy missions that mount both hyperspectral and panchromatic imaging sensors, such as EO-1/Hyperion-advanced land imager (ALI) [203], PRISMA [8], and Shalom [9]. The main advantage of HSIs on multispectral images is the rich spectral information, which enables the discrimination and identification of spectrally similar materials. In other words, any spectral distortions will lead to inaccurate analysis results in the subsequent data processing. Therefore, the challenge of hyperspectral pan-sharpening is to enhance the spatial resolution of hyperspectral data while minimizing spectral distortion.

Typical pan-sharpening techniques include the component substitution (CS) [204]–[206], multiresolution analysis (MRA) [207], [208], and SR [209], [210] algorithms. Aiazzi et al. [206] proposed a general CS-based pan-sharpening scheme in which, a multispectral image is sharpened through the addition of spatial details obtained when multiplying the difference between a panchromatic image and a synthetic intensity component by a band-wise modulation coefficient. Gram–Schmidt adaptive (GSA) is one of the benchmark CS-based pan-sharpening algorithms; here, the synthetic intensity component is computed via linear regression between a panchromatic image and lower-resolution bands.

In the MRA-based pan-sharpening scheme, spatial details of each multispectral band are obtained by use of MRA, which calculates the difference between a panchromatic image and its low-pass version multiplied by a gain factor. The gain factor can be computed either locally or globally. Representative MRA pan-sharpening algorithms include the smoothing filtered-based intensity modulation (SFIM) method [207], the additive wavelet luminance proportional method [211], and the generalized Laplacian pyramid (GLP) method [208].

The SR-based pan-sharpening approach can be regarded as a special case of learning-based superresolution, i.e., learning correspondences between low- and high-resolution image patches from a panchromatic image [210] or an external database, including multiple high-resolution multispectral images [212]. In [213], Alparone et al. demonstrated the fusion of Hyperion and ALI panchromatic images using CS- and MRA-based algorithms [213].

The subspace-based approach exploits the intrinsic spectral characteristics of the scene via a subspace spanned by a set of basis vectors, enhancing the spatial resolution of subspace coefficients. Because most subspace-based methods were developed for HS–MS fusion, we discuss them in the next section.

Eleven state-of-the-art algorithms from the CS-, MRA-, and subspace-based approaches were applied to hyperspectral pan-sharpening and compared in [214]. MRA- and subspace-based algorithms demonstrated relatively high-quality and stable results. However, it was evident that large spectral distortion was inevitable for all the algorithms under comparison, implying room for further technology development [214].

HYPERSPECTRAL AND MULTISPECTRAL DATA FUSION

HS–MS fusion is a technique that fuses an HSI with a higher-resolution multispectral image to create high-resolution hyperspectral data. Unlike hyperspectral pan-sharpening, this technique yields higher spectral quality owing to spectral information from the high-resolution data source. Enormous efforts have made to develop algorithms in the last decade. Quite recently, a comparative review of HS–MS fusion was reported in [215]. Most of the HS–MS fusion algorithms can be categorized into at least one of the following six classes of approaches: 1) CS, 2) MRA, 3) SR, 4) unmixing, 5) Bayesian, and 6) hybrid. The CS, MRA, and SR methods are extensions of pan-sharpening techniques, whereas the unmixing and Bayesian approaches fall into the same broader category of subspace-based techniques.

CS-based methods can be adapted to HS–MS fusion by splitting the fusion problem into several pan-sharpening subproblems and applying CS-based pan-sharpening to these subproblems. A key procedure for the CS-based approach is to divide the HSI into several groups and assign a high-resolution image selected or synthesized from the multispectral image to each group. Chen et al. proposed a framework that solves the HS–MS fusion problem by dividing the spectrum of hyperspectral data into several regions and fusing hyperspectral and multispectral band images in each region by conventional pan-sharpening techniques [216]. The hypersharpening proposed by Selva et al. is a framework that effectively adapts MRA-based pan-sharpening methods to HS–MS fusion by synthesizing a high-resolution image for each hyperspectral band as a linear combination of multispectral band images via linear regression [217]. An SR-based pan-sharpening technique was adapted to HS–MS fusion with spectral grouping and joint sparse recovery in [218].

Subspace-based methods have been actively developed for HS–MS fusion problems. The HS–MS fusion task is formulated as the inverse problem of estimating the subspace basis and coefficients of the high-resolution HSI from the two input images [219]. The resolution-enhanced HSI can be reconstructed as the product of the basis matrix and the high-resolution coefficient matrix.

In recent years, a perspective of spectral unmixing has been attracting considerable attention in the context of subspace-based HS–MS fusion, owing to its straightforward interpretation of the fusion process. The basis matrix is defined as a set of spectral signatures of intrinsic materials (endmembers), and the subspace coefficients correspond to the fractional abundances. Several unmixing-based methods have been proposed for HS–MS fusion with various optimization formulations [220]–[227]. In [215], it was noted that the high-resolution abundances estimation step has a significant influence on fusion performance. Outstanding and stable performance could be achieved by minimizing the unmixing reconstruction errors with respect to both the hyperspectral and multispectral images rather than only the multispectral image, particularly when the overlap of spectral response functions between hyperspectral and multispectral sensors was small. Such algorithms include coupled NMF (CNMF) [224], coupled spectral unmixing with a projected gradient method [228], and HySURE [229].

One important aspect of developing subspace-based methods has been determining how to mitigate the ill-posedness of inverse problems involving the estimation of high-resolution subspace coefficients. Regularization on subspace coefficients has been extensively explored. Sparsity-promoting regularization has been commonly adopted for subspace-based methods [222], [225], [226], assuming that there is a limited number of endmembers at each high-resolution pixel. HySURE solves a convex optimization problem with vector-total-variation-based regularization of the spatial distribution of subspace coefficients, leading to implicit denoising effects in the fusion results [229]. Wei et al. developed a Bayesian HS–MS fusion methodology in which different types of regularization terms on subspace coefficients can be designed flexibly, based on information from the prior distribution in the observed scene [230]. A Sylvester equation-based explicit solution was further integrated into the Bayesian methodology to speed it up, leading to fast fusion based on Sylvester equation (FUSE) [231]. Veganzones et al. [232] introduced local image processing into an unmixing-based approach, where well-posed inverse problems are solved for each small patch, assuming that the number of endmembers in each patch is smaller than the number of multispectral bands.

One interesting finding of the comparative study in [215] was that hypersharpening methods and unmixing-based methods (CNMF and HySURE), which are entirely different approaches, showed comparably high numerical performances, although the characteristics of the resolution-enhanced HSIs are different. This finding implies that hybrid methods combining different approaches can be expected to further improve fusion performance.

Table 5 shows the overall quantitative assessment results of 12 HS–MS fusion algorithms for eight simulated HS–MS data sets used in [215], including those based on the Pavia University and University of Houston data sets. The quantitative assessment of fusion performance was carried out based on a version of Wald's protocol presented in [232] and [215].

The reconstruction performances of algorithms has posed the greatest concern for researchers; however, their general versatility, computational costs, and impacts on applications are also important for users. Because each method has advantages and disadvantages, it is essential to choose a method based on the fusion and analysis scenarios. The impact of HS–MS fusion on applications has been recently investigated via classification and unmixing [233], [215]. Further research on real data is still necessary to verify the practicability of HS–MS fusion for future hyperspectral satellites.

CHALLENGES IN RESOLUTION ENHANCEMENT

The main challenges in the resolution enhancement of HSIs relate to practical issues. The development of multisensor superresolution algorithms has recently surged; however, very few publications in the literature discuss experiments on real data. Studies regarding the impact of HSI resolution enhancement on applications are still lacking. It is necessary to clarify benefits and address practical issues of resolution-enhancement technology to promote its operational application for future hyperspectral satellite missions. For instance, temporal mismatches included in the input images raise challenges for HSI resolution enhancement. Furthermore, a no-reference image-quality assessment for HSIs needs to be developed to provide reliability information regarding resolution-enhanced hyperspectral data products for end users.

In multisensor superresolution, the problem remains that large spectral distortions are inevitable when the mismatch between the two imaging sensors is large in either the spatial or spectral domain (e.g., a large GSD ratio or hyperspectral pan-sharpening). To address this issue, a possible future direction for performance improvement lies in developing algorithms that exploit a spectral library or spatial information of a high-resolution image.

TABLE 5. THE AVERAGE QUALITY MEASURES FOR EIGHT DATA SETS FROM [215].

CATEGORY	METHOD	PSNR	SAM	ERGAS	$Q2^n$
CS	GSA [206]	39.221	2.063	1.885	0.8796
MRA	SFIM-HS [207], [217]	41.074	1.707	1.789	0.8867
	GLP-HS [208], [217]	41.322	1.664	1.733	0.8984
Unmixing	CNMF [224]	42.092	1.607	**1.681**	0.9060
	ECCV '14 [225]	38.101	2.658	4.561	0.8431
	ICCV '15 [227]	40.470	1.672	1.902	0.8847
	HySURE [229]	**42.336**	**1.602**	1.766	**0.9089**
Bayesian	MAP-SMM [219]	40.008	1.822	1.997	0.8635
	FUSE [231]	40.568	1.881	1.908	0.8693
	FUSE-S [231]	41.177	1.803	1.804	0.8764
Ideal		∞	0	0	1

Bold numbers indicate the best performances. HS: hyperspectral; ECCV: European Conference on Computer Vision; ICCV: International Conference on Computer Vision; FUSE-S: sparse FUSE.

HSI DENOISING AND IMAGE RESTORATION

Image restoration generally refers to the reconstruction of the true image based on its corrupted version. The true image is unknown and, therefore, is estimated through the observed image, which has been degraded by different sources. Degradation sources depend on the imaging technology, system, environment, and other factors. When the observed signal is degraded by noise sources, the estimation task is usually called *denoising*. *Image restoration* generally refers to a broader set of methods that also include applications such as deconvolution, deblurring, and inpainting. However, in HSI analysis, the term *restoration* is often used to refer to denoising.

In hyperspectral imaging the received radiance at the sensor is degraded by different sources, such as atmospheric haze and instrumental noise. The atmospheric effects are usually compensated for by applying atmospheric corrections. Instrumental (sensor) noise includes thermal (Johnson) noise, quantization noise, and shot (photon) noise. Spectral bands are often corrupted to some degree. The presence of corrupted bands (also called *junk bands*) could degrade the efficiency of the image analysis technique, and, therefore, they are usually removed from the data before any further processing. The information lost by removing those bands can be substantial; hence, an alternative approach is to recover those bands and improve the SNR of the HSI. As a result, HSI restoration can be considered an important preprocessing step in HSI analysis.

An HSI can be modeled by

$$\mathbf{X} = \mathbf{H} + \mathbf{N}, \tag{1}$$

where $\mathbf{X}$ is an $n_{12} \times d$ matrix $(n_{12} = n_1 \times n_2)$ containing the vectorized observed image at band i in its ith columns, $\mathbf{H}$ is the true unknown signal to be estimated and is represented as an $n_{12} \times d$ matrix containing the unknown vectorized image at band i in its ith columns, and $\mathbf{N}$ is an $n_{12} \times d$ matrix containing the vectorized noise at band i in its ith columns. Note that all of the previously mentioned noises can be assumed to be additive noise. The restoration task is to estimate the original (unknown) signal $\mathbf{H}$. Penalized (regularized) least squares is one of the most popular and common minimization frameworks used for estimation in HSI restoration. Penalized least squares is usually composed of a fidelity term and a penalty term. The penalty term is often chosen based on the prior knowledge of the signal and might be a combination of penalties. Penalized least squares might also be solved subject to some constraints.

NOISE SOURCE ASSUMPTIONS IN HYPERSPECTRAL IMAGES

The presence of different noise sources in HSIs makes their modeling and restoration highly challenging. Therefore, HSI restoration often employs one or a mixture of the following approaches.

SIGNAL-INDEPENDENT NOISE

Thermal noise and quantization noise in HSIs are modeled by signal-independent Gaussian additive noise [234], [235]. Usually, noise is assumed to be uncorrelated spectrally, i.e., the noise covariance matrix is diagonal [236], [235]. The Gaussian assumption has been broadly used in hyperspectral analysis because it considerably simplifies the analysis. In addition, noise parameter (variance) estimation is simpler under this assumption.

***IMAGE RESTORATION* GENERALLY REFERS TO THE RECONSTRUCTION OF THE TRUE IMAGE BASED ON ITS CORRUPTED VERSION.**

SIGNAL-DEPENDENT NOISE

Shot (photon) noise in HSIs is modeled by a Poission distribution because the noise variance is dependent on the signal level. The noise parameter (variance) estimation is more challenging under this assumption compared to the signal-independent case because it varies with respect to the signal level.

SPARSE NOISE

Impulsive noise, missing pixels, missing lines, and other outliers often exist in the acquired HSI, usually due to the malfunctioning of the sensor. In this review, we categorize these as sparse noise because of their sparse characteristics. Sparsity techniques or sparse and low-rank decomposition techniques have been used to remove sparse noise from the signal. In [237], impulsive noise was removed using an ℓ_1-norm for both penalty and data fidelity terms in the minimization problem suggested.

STRIPING NOISE

Hyperspectral imaging systems might also induce artifacts in HSIs, usually referred to as *pattern noise*. For instance, in pushbroom imaging systems, the target is scanned line by line, and the image line is acquired in different wavelengths by an area-array detector (usually a charged coupled device). This line-by-line scanning causes an artifact called *striping noise*, which is often due to calibration error and the sensitivity variation of the detector [238]. Striping noise reduction (also referred to as *destriping* in the literature) for pushbroom scanning techniques has been widely studied in the remote-sensing literature [239], [240], including work on HSI remote sensing [238], [241]–[243].

EVOLUTION OF HYPERSPECTRAL IMAGE RESTORATION APPROACHES

HSI restoration has been developed considerably over the past few years. Conventional restoration methods based on 2-D modeling and convex optimization techniques were not effective enough for HSIs because of the lack of understanding of spectral information. The highly correlated spectral bands in HSIs have been found very useful for HSI restoration. This is the main reason for the success in using MLR as an estimation technique [244]. In [244], it is assumed that each band is a linear combination of the other bands; therefore, the *i*th band is estimated using least squares estimation. Note that this technique has been used for noise parameter estimation in several HSI restoration approaches [245], [246]. Many restoration approaches have been suggested in the literature to exploit the spectral information, and they can be categorized into three main groups.

APPROACHES THAT USE 3-D MODELS INSTEAD OF 2-D ONES

In [247], the discrete Fourier transform was used to decorrelate the signal in the spectral domain, and the 2-D discrete wavelet transform (DWT) was investigated to denoise the signal in the spatial domain. In [248], the HSI was treated as a 3-D data cube, and an HSI restoration technique was proposed based on sparse analysis regularization and undecimated wavelet transform. The advantages of 3-D wavelets (orthogonal and undecimated) over 2-D ones for HSI restoration were also discussed in [248] and [236].

APPROACHES THAT PROPOSE NEW PENALTIES FOR PENALIZED LEAST SQUARES THAT ALSO TAKE INTO ACCOUNT SPECTRAL INFORMATION

An algorithm given in [249] uses the 2-D DWT and a sparse restoration criterion based on penalized least squares having a group of ℓ_2 penalties on the wavelet coefficients. This method was improved on in [250] for HSI restoration by taking into account the spectral noise variance in the minimization problem and solving it using the alternating direction method of multipliers. Subsequently, because of the redundancy and high correlation in the spectral bands in HSIs, penalized least squares using a first-order spectral roughness penalty (FOSRP) was proposed for HSI restoration [251]. The new cost function was formulated in the wavelet domain to exploit the MRA property of wavelets. The Stein unbiased risk estimator (SURE) was used to automatically select the tuning parameters. It was shown that FOSRP outperforms sparsity penalties for HSI restoration. This method was improved upon using a combination of a spectral roughness penalty and a group lasso penalty [252].

Cubic TV (CTV), proposed in [253], exploits the gradient in the spectral axis to improve the restoration results compared to TV denoising band by band. In [254], an adaptive version of CTV was applied for preserving both texture and edges simultaneously. In [255], a spatial-spectral HSI denoising approach was developed. The spectral derivation was proposed to concentrate the noise in the low frequency. Then, noise was removed by applying the 2-D DWT on the spatial domain and the 1-D DWT on the spectral domain. A spatial-spectral prior for maximum a posteriori information was proposed in [256]. The prior was based on five derivatives, one along the spectral direction and the rest applied on the spatial domain for four neighborhood pixels.

APPROACHES THAT USE LOW-RANK MODELS

Because of the redundancy along the spectral direction, low-rank modeling has been widely used in HSI analyses and applications such as DR, feature extraction, unmixing, and compression. PCA was used for hyperspectral restoration in [257] to spectrally decorrelate the noise and signal. A low-rank representation method is a three-mode factor analysis called *Tucker3 decomposition* [258] used for HSI restoration [259]. The HSI is assumed to be a third-order tensor, and the best lower rank of the decomposition is chosen by minimizing a Frobenius norm. A similar idea was exploited for HSI restoration by applying more reduction spectrally [260]. A GA was developed to choose the rank of the Tucker3 decomposition [261]. This work was followed by [262], in which the kernel function (Gaussian radial basis) was applied for each spectral band, with the idea of efficiently using multilinear algebra.

Multidimensional Wiener filtering that exploits Tucker3 decomposition was used in [263], where the flattening of the HSI was achieved by estimating the main direction that gives the smallest rank. Parallel factor analysis is also a low-rank modeling used in [264] for HSI denoising. A new 3-D linear model was proposed for HSI in [265], where 2-D wavelets were used for spatial projection and spectral-singular vectors of the observed HSI were used for spectral projection. A convex optimization was used for the restoration task based on the 3-D linear model and ℓ_1 penalty. Additionally, SURE was used for regularization parameter selection.

Low-rank modeling has been used in synthesis and analysis penalized least squares [266], [250] and also in TV regularization [267], [250]. A 3-D low-rank model in the form of model (2) in the next section was proposed in [246], where 2-D wavelets were used as the spatial basis while the spectral basis was assumed to be an unknown low-rank orthogonal matrix. Therefore, an orthogonality constraint was added to the optimization problem for the simultaneous estimation of the two unknown matrices in the minimization problem, which led to a nonconvex optimization.

HYPERSPECTRAL IMAGE MODEL SELECTION

Further studies in HSI modeling and restoration [250] confirmed that capturing spectral redundancy by low-rank modeling is more appropriate than full-rank modeling for HSI modeling and restoration. In [250], a model selection criterion was given for a general model of the form

$$\mathbf{X} = \mathbf{A}\mathbf{W}_r\mathbf{M}_r^T + \mathbf{N}, \tag{2}$$

where $\mathbf{A}$ ($n_{12} \times n_{12}$ matrix) and $\mathbf{M}_r$ ($d \times r$ matrix, $r \le \min(n_{12}, d)$) could be 2-D and 1-D orthogonal (known) bases, respectively, and $\mathbf{W}_r$ ($n_{12} \times r$ matrix) contains the corresponding coefficients for the unknown hyperspectral data, $\mathbf{H}$. The term $\mathbf{W}_r$ is estimated using ℓ_1 penalized least squares, and the signal is restored by $\hat{\mathbf{H}} = \mathbf{A}\hat{\mathbf{W}}_r\mathbf{M}_r^T$. Ideally, the model that gives the lowest MSE is the best choice among the candidates. However, the MSE is uncomputable in practice because it depends on the true (uncorrupted) data. Therefore, SURE is suggested for use as an estimator of MSE for HSIs. The results have confirmed that low-rank models give lower MSE (estimated by SURE) compared with full-rank ones.

HYPERSPECTRAL IMAGE RESTORATION WITH MIXED-NOISE ASSUMPTION

A mixed-noise assumption has also been taken into consideration in HSI modeling and restoration. In a mixed-noise assumption, the HSI ($\mathbf{X}$) in model (1) is assumed to be corrupted by a mixture of the noise sources described in the "Noise Source Assumptions in Hyperspectral Images" section. A mixture of the signal-dependent noise ($\mathbf{N}_{\mathrm{SD}}$) and signal-independent noise ($\mathbf{N}_{\mathrm{SI}}$) models has been taken into account in [245], [268], and [269] as $\mathbf{N} = \mathbf{N}_{\mathrm{SI}} + \mathbf{N}_{\mathrm{SD}}$. Therefore, two parameters need to be estimated: the variances of $\mathbf{N}_{\mathrm{SI}}$ and $\mathbf{N}_{\mathrm{SD}}$, which have Gaussian distribution and Poission distribution, respectively. In [268], a 3-D (block-wise) nonlocal sparse restoration method was suggested for HSIs. The minimization problem, which uses a group lasso penalty and a dictionary consisting of a 3-D discrete cosine transform and a 3-D DWT, was solved by using the accelerated proximal gradient method. In [269], $\mathbf{N}_{\mathrm{SI}}$ and $\mathbf{N}_{\mathrm{SD}}$ were removed sequentially. Maximum likelihood was used to estimate the two parameters of the noise model where MLR was investigated for an earlier estimation of the noise. In [241], a subspace-based approach was given to restore HSI corrupted by striping noise and $\mathbf{N}_{\mathrm{ID}}$.

A widely used mixed-noise assumption for HSI restoration is $\mathbf{N} = \mathbf{N}_{\mathrm{SI}} + \mathbf{N}_{\mathrm{S}}$, where $\mathbf{N}_{\mathrm{S}}$ represents the sparse noise defined in the "Noise Source Assumptions in Hyperspectral Images" section. This mixture assumption was used in [270], where low-rank and sparse matrix decomposition was taken into account to restore HSIs. This method was improved on by augmenting the total variation penalty to the restoration criterion [271]. Also, the noise-adjusted iterative low-rank matrix approximation given in [272] approximates an HSI with a low-rank matrix, while taking into account the changes of the noise variance through the spectral bands. The authors of [273] proposed using a weighted Schatten p-norm as a nonconvex low-rank regularizer for low-rank and sparse decomposition of HSIs degraded by sparse and Gaussian noises.

EXPERIMENTAL RESULTS

In this section, we present some experimental results for HSI restoration. Figure 10 shows the evolution of HSI restoration techniques based on SNR. The experiments were applied on a portion ($128 \times 128 \times 96$) of the Pavia University data set; here, the variance of the added Gaussian noise varies along the spectral axis (σ_i^2), like a Gaussian shape centered at the middle band ($d/2$), as

$$\sigma_i^2 = \sigma^2 \frac{e^{-\frac{(i-d/2)^2}{2\eta^2}}}{\sum_{j=1}^{d} e^{-\frac{(j-d/2)^2}{2\eta^2}}}, \tag{3}$$

where the power of the noise is controlled by σ and η behaves like the standard deviation for the Gaussian bell curve [244]. To evaluate the restoration results for the simulated data set, the SNR in decibels is used as

$$\mathrm{SNR}_{\mathrm{out}} = 10 \quad \log_{10}\left(\|\mathbf{H}\|_F^2 / \|\mathbf{H} - \hat{\mathbf{H}}\|_F^2\right),$$

where $\|\,.\,\|_F$ is the Frobenius norm and the noise input level for the whole cube is defined as

$$\mathrm{SNR}_{\mathrm{in}} = 10 \quad \log_{10}(\|\mathbf{H}\|_F^2 / \|\mathbf{H} - \mathbf{X}\|_F^2).$$

In Figure 10, the results are shown when $\mathrm{SNR}_{\mathrm{in}}$ varies from 5 to 40 dB in increments of 5 dB. Note that the results shown are means-over-ten experiments (adding random Gaussian noise) and the error bars show the standard deviations. In this experiment, six restoration methods are compared based on SNR: ℓ_1 penalized least squares using 2-D wavelet modeling (2-D wavelet) [274] and 3-D wavelet modeling (3-D wavelet) [248]; penalized least squares using FOSRP (FOSRPDN) [251]; ℓ_1 penalized least squares using a wavelet-based low-rank model called sparse wavelet-based model and rank selection (SPAWMARS) [250]; low-rank matrix recovery (LRMR) [270]; and noise-adjusted iterative low-rank matrix approximation [272]. Note that 2-D wavelet, 3-D wavelet, FOSRPDN, and SPAWMARS all exploit SURE as a parameter selection technique and, therefore, are parameter-free methods. Also, note that SPAWMARS is a

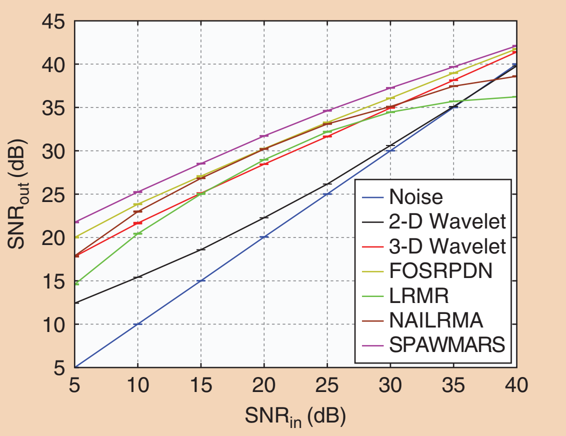

FIGURE 10. A comparison of the performance of different HSI restoration methods for different levels of input Gaussian noise in decibels. NAILRMA: noise-adjusted iterative low-rank matrix approximation.

TABLE 6. THE CPU PROCESSING TIME IN SECONDS FOR DIFFERENT RESTORATION APPROACHES USED IN THE EXPERIMENT.

2-D WAVELET	3-D WAVELET	FOSRPDN	LRMR	NAILRMA	SPAWMARS
1.16	1.72	1.09	25.70	4.79	**0.62**

fully automatic version of an SVD-based sparse regularization with reduction [236] with rank selection. The MATLAB codes for FOSRPDN and SPAWMARS are available online in [275] and [276], respectively.

The blue line in Figure 10 indicates the noise levels, and, therefore, the effectivity of the HSI restoration techniques can be compared with respect to the noise levels. It can be seen that 3-D wavelet restoration considerably improves on conventional 2-D wavelet techniques. Also, FOSRPDN, which is based on a 2-D wavelet and an FOSRP, outperforms 3-D wavelet restoration, which confirms the importance of the spectral correlations. Finally, SPAWMARS, which utilizes a wavelet-based low rank, outperforms the other techniques used in this experiment. Note that Wavelab Fast (a fast wavelet toolbox), which is provided in [277], was considered for the implementation of wavelet transforms. A Daubechies wavelet with two and ten coefficients for spectral and spatial bases, respectively, in five decomposition levels was used in all the experiments. The central processing unit (CPU) processing time in seconds for different restoration approaches used in the experiment confirms that SPAWMARS is the most efficient method (Table 6). All the experiments in this section were performed in MATLAB on a computer having an Intel Core i7-4710HQ CPU at 2.5 GHz, with 12 GB of memory, and a 64-b operating system.

CHALLENGES IN HYPERSPECTRAL IMAGE DENOISING AND IMAGE RESTORATION

HSI restoration and modeling face several future challenges.

1) The HSI model selection and noise parameter estimation need more attention. For instance, a model selection criterion that is not restricted to the Gaussian noise model would be very useful for HSI modeling and restoration. The main advantage of a model selection criterion is that it provides an instrument to compare the restoration techniques without using a simulated (noisy) HSI but rather using the observed HSI itself.
2) It is of interest to investigate the contributions of the various HSI restoration approaches, such as CD, unmixing, and resolution enhancement, as preprocessing steps for further HSI analysis.
3) In mixed-noise scenarios, investigating the dominant noise type within HSIs should be considered.
4) HSI restoration approaches, which are computationally efficient and parameter free (automatic), can simply be used as a preprocessing step in real-world applications. Fast computing techniques may be considered for the swift implementation of HSI restoration approaches in the future.

CHANGE DETECTION

In this section, we define a CD problem in a pair of bitemporal HSIs (those in the multitemporal domain can be addressed straightforwardly, pair by pair). Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be two HSIs acquired over the same geographical area at times t_1 and t_2, respectively. The hyperdimensional difference image $\mathbf{X}_D$, i.e., the SCVs, can be computed by subtracting the bitemporal images pixel by pixel, i.e.,

$$\mathbf{X}_D = \mathbf{X}_2 - \mathbf{X}_1. \tag{3}$$

Let $\Omega = \{\omega_n, \Omega_c\}$ be the set of all classes in $\mathbf{X}_D$, where ω_n is the no-change class and $\Omega_c = \{\omega_{C_1}, \omega_{C_2}, \dots, \omega_{C_K}\}$ be the set of K possible change classes. The considered binary CD problem can be formalized to separate the ω_n and Ω_c classes without distinguishing different classes in Ω_c; the objective of the multiclass CD task is to detect the changed pixels in Ω_c and identify their classes in $\{\omega_{C_1}, \omega_{C_2}, \dots, \omega_{C_K}\}$.

Continuous satellite observation resulted in the acquisition of a large number of multitemporal remote-sensing images. By analyzing these images, a better understanding of the changes and evolutions on Earth's surface can be gained. CD is a technique that enables the land cover changes occurring over a geographical area at different observation times to be identified [278]. In past decades, CD has played important roles in various multitemporal remote-sensing applications, such as urban sprawl analysis, disaster loss evaluation, and forest and environmental monitoring [279]–[281]. For optical remote-sensing images, CD based on multispectral images has been intensively investigated because of the availability of multispectral sensors onboard the last-generation Earth observation satellites. With the launch of new-generation Earth observation satellites carrying hyperspectral sensors, there are further opportunities to implement CD in multitemporal HSIs.

Unlike multispectral images, detailed spectral sampling in HSIs allows the potential detection of more spectral variations, especially those with subtle, spectrally insignificant changes [282]. Compared with the abrupt changes that present in coarse multispectral images, changes in HSIs are more sophisticated, implicit, and structurally complex. One emerging problem within this context is the detection of small and subtle changes compared to the large unchanging background in multitemporal

HSIs. This is related to the definition of an anomaly CD (ACD) problem, which is not easily addressed by classical CD techniques.

Traditional CD identifies large land cover changes characterized by significant spectral variations. An empirical definition of the change concept in multitemporal HSIs from the global and local spectrum discriminability can be found in [282]. For CD techniques in multispectral images, exhaustive investigations have been made in the past few decades. However, we could find relatively few works covering CD in HSIs. Recently, a book chapter [283] analyzed this challenging task and provided a literature review and comparison between CD in hyperspectral and multispectral images. Problems and challenges in the existing methods were also discussed.

This section reviews several recent CD techniques for HSIs, discussing and analyzing their properties. In addition, a quantitative analysis comparing the CD results obtained by some state-of-the-art CD techniques on the Umatilla County Hyperion data set is provided. A summary of several classical and recent techniques for multitemporal HSI CD techniques is provided in Table 7, along with some details such as their categories and principal characteristics. Note that this summary is not exhaustive.

ANOMALY CHANGE DETECTION

ACD in HSIs is intended to distinguish anomalous changes from the nonchanges and pervasive changes in multitemporal HSIs [284]. The main idea of ACD is to design a robust detector able to maximize the difference between the

TABLE 7. A SUMMARY OF SOME CLASSICAL AND RECENT TECHNIQUES FOR MULTITEMPORAL HSI CD (NOT EXHAUSTIVE).

APPLICATION PURPOSE	SUPERVISED/SEMISUPERVISED/ UNSUPERVISED	TECHNIQUES	MAIN CHARACTERISTICS
ACD	Unsupervised	Chronochrome [285], covariance-equalization [286]	An anomaly change estimation based on linear predictors.
	Unsupervised	Elliptically contoured distributions-based methods [287]	Anomalous change detectors based on elliptically contoured distributions.
	Unsupervised	SFA-based RX algorithm [284]	SFA is used to construct the change residual image.
	Unsupervised	CKRX [288]	Background pixels are clustered, and the cluster centers are used in ACD.
	Unsupervised	Eliminating external factors (e.g., image parallax errors [289], vegetation and illumination variations [290], diurnal and seasonal variations [291])	Enhancing the CD performance.
Binary CD	Unsupervised	MAD [294], IR-MAD [296], T-PCA [297], ICA [298]	Changes are represented and highlighted in transformed and reduced feature space.
	Supervised	Subspace-based approach [299]	Changes are identified as anomalous pixels according to the subspace distance.
	Semisupervised	Distance metric-based method [300]	The distance metric is learned in a Laplacian regularized metric framework.
	Unsupervised	Slight CD method [16]	Using block processing and locally linear embedding.
	Unsupervised/supervised	Unmixing-based methods [301]–[304]	Unmixing single-time or stacked multitemporal images to investigate the subpixel-level change.
Multiple CD	Supervised	PCC [305]	Independently classifying each single-time image; detailed land cover transitions (i.e., from–to information) can be obtained.
	Supervised	3-D spectral modeling approach [306]	A spatial/spectral/temporal joint modeling and unmixing of multitemporal data.
	Unsupervised	HSCVA [282]	Hierarchical clustering and detecting of spectral variations at different significant levels.
	Semisupervised	S^2CVA [307]	The adaptive and sequential projection of SCVs to discover and detect multiple changes.
	Unsupervised	MSU [12]	Spectral–temporal unmixing to identify the unique multitemporal endmembers for representing multiclass changes.
	Unsupervised/supervised	Band-selection-based method [309]	The most informative band subset of the different images is selected to enhance CD performance.

anomaly changes and the unchanged scene background but to minimize the possible class difference within the background. In this context, multivariate statistical techniques based on linear predictors, e.g., cross covariance (chronochrome) [285] and covariance-equalization [286], are typical ACD and binary CD detectors. Recently, algorithms were proposed to detect anomalous changes in HSIs by modeling the data with elliptically contoured distributions [287]. A new ACD method was proposed in [284] that constructed the change residual image based on slow feature analysis (SFA) and detected anomaly changes by using the Reed–Xiaoli (RX) algorithm. A cluster kernel RX (CKRX) algorithm was developed in [288] that clustered the background pixels, then used the cluster centers in the anomaly detection. Other investigations have also focused on various specific changes—e.g., eliminating image parallax errors [289], vegetation and illumination variation [290], and diurnal and seasonal variations [291].

FOR BINARY CD, THE AIM IS TO SEPARATE THE CHANGED AND UNCHANGED PIXELS IN THE IMAGES.

BINARY CHANGE DETECTION

For binary CD, the aim is to separate the changed and unchanged pixels in the images. In this case, classical techniques in multispectral imaging such as thresholding [292] or clustering [293] can still be considered, based on the computed magnitude of the SCVs. Transform-based methods constitute an important class. Such techniques represent the original data in a feature space where the significant change information is concentrated in a few transformed components. This not only reduces data dimensionality and noise but also focuses on the specific changes of interest in specific components.

The multivariate alteration detection (MAD) method originally proposed for multispectral images in [294] was applied to HSIs to detect vegetation change based on canonical correlation analysis [295]. It was extended to an iterative reweighted version (IR-MAD) [296] to better emphasize and detect changes. A temporal PCA (T-PCA) was proposed in [297] that exploits the temporal variances in the combined multitemporal HSIs. The no-change information is represented in the first principal component and the change information in the second, which is orthogonal to the first one.

Recently, in [298], ICA was applied, combined with the uniform feature design strategy, by following a hierarchical framework to identify specific vegetation changes. In [299], a subspace-based CD approach was designed using the undesired land cover class spectral signature as prior knowledge, where the subspace distance was computed to determine the anomalous pixels as change with respect to the background subspace. A semisupervised CD method was proposed in [300], where a Laplacian regularized metric framework was exploited to learn a distance metric for detecting change under noisy conditions. In [16], an unsupervised CD method was developed for slight change extraction in a multitemporal HSI sequence. The feature space was built using block processing and locally linear embedding, and then a final CD map was generated by clustering the change class and the no-change class. In [301] and [302], CD in HSIs using unmixing was investigated on a single-time image and stacked multitemporal images, respectively. Also, in [303] and [304], sparse unmixing and a decision-fusion-based spectral-mixture approach were exploited, respectively, to detect the subpixel-level change information.

MULTIPLE CHANGE DETECTION

The aim of multiple CD is more complex and challenging than binary CD. Changed pixels are detected while different classes of change are distinguished. If multitemporal ground reference samples are available, this task can be addressed according to a typical supervised postclassification comparison (PCC) method [305] by classifying independently two (or more) images at different times and then comparing the pixel class label to detect changes. The main advantage of PCC is that detailed land cover transitions are obtained (i.e., from–to information). However, the accuracy of the CD performance depends greatly on the accuracy of a single-time image classification result.

In [306], a new approach for modeling the temporal variations of the reflectance response as a function of time period and wavelength was developed. In this approach, a library of known endmembers that depends mainly on the 3-D surface reconstruction quality and similarity measure is used to perform a classification task. Changes are detected, and the approach provides better modeling of seasonal variations. However, in practical applications, comprehensive multitemporal ground reference samples are usually not available. Accordingly, unsupervised techniques that do not rely on reference samples are more valuable.

In [282], an unsupervised coarse-to-fine hierarchical SCV analysis (HSCVA) approach was proposed. It was designed to cluster and detect changes that include spectral variations at different significant levels according to their discriminable spectral behaviors. A semisupervised sequential SCV analysis (S^2CVA) technique for discovering, identifying, and discriminating multiple changes in HSIs was proposed in [307]. This method detects different kinds of changes according to the adaptive and sequential projection of SCVs at each level of the hierarchy. S^2CVA successfully extends the compressed change vector analysis method [308], which is suitable for dealing only with multispectral CD cases having few spectral channels. To identify the subpixel-level spectral changes so that a more accurate multiple CD result can be produced, an unsupervised multitemporal spectral unmixing (MSU) model was proposed in [12]. The MSU investigates in detail the spectral-temporal mixture properties in multitemporal HSIs, and the considered multiple CD problem is addressed by analyzing the abundances of different distinct multitemporal endmembers at a subpixel level.

In addition, both unsupervised and supervised band-selection-based CD approaches were designed to evaluate the potential detectability of the change detector in a reduced feature space [309]. Experimental results confirmed that the most informative band subset selected can help to improve CD performance over the use of the original full-dimensional HSIs. A multiclass CD experimental comparison was carried out on the Umatilla County Hyperion data set using the state-of-the-art CD techniques introduced above, including HSCVA [282], S^2CVA [307], and MSU [12]. Numeric experimental results are provided in Table 8, which includes the obtained OA, kappa coefficient, and omission and commission errors.

From Table 8, we can see that the MSU resulted in the best performance in this data set, with respect to the highest OA (95.37%) and kappa (0.8826) values and lowest number of total errors (3,114 pixels). This indicates the superior detection of spectral variations at a fine level in the MSU approach. The other two state-of-the-art methods also obtained a high level of OA and kappa values, which demonstrates the effectiveness of the considered methods. In practical applications, two pixel-level multiple CD techniques are sufficient to address the challenging multiple CD task, especially in an unsupervised/semisupervised fashion, without using ground reference samples.

TABLE 8. THE CD ACCURACIES AND ERRORS OBTAINED BY CONSIDERED STATE-OF-THE-ART TECHNIQUES ON THE UMATILLA COUNTY HYPERION DATA.

METHODS	HSCVA [282]	S^2CVA [307]	MSU [12]
OA (%)	95.26	95.36	95.37
Kappa	0.8800	0.8822	0.8826
Omission (pixels)	1,715	1,878	1,875
Commission (pixels)	1,918	1,676	1,239
Total errors (pixels)	3,633	3,554	3,114

CHALLENGES IN CHANGE DETECTION

For CD tasks in multitemporal HSIs, challenges come from the intrinsic properties of the hyperspectral data and the design of sophisticated CD techniques to handle the complexities in the CD process. From the hyperspectral data perspective, the high dimension inevitably leads to information redundancy and high computational cost. Moreover, changes are to be more implicitly represented and detected. In other words, from the spectral signature point of view, similar changes are more likely to overlap, especially subtle changes. Thus, the difficulty in discriminating their classes in high-dimensional feature space increases.

From the CD methodology perspective, advanced approaches need to be designed that 1) can more adaptively detect multiple complex changes (i.e., those having different spectral significance), 2) can be implemented more automatically, and 3) can be computationally effective. In particular, the development of unsupervised CD techniques for HSIs is more important for real applications. In fact, several subproblems should be considered within a CD process, such as the identification of the number of multiple changes, the separation of changed and unchanged pixels, and the discrimination of multiclass changes. Each of these subproblems deserves to be investigated in detail, independently or simultaneously, to generate more accurate CD results. In addition, the subpixel and superpixel implementation of CD techniques at different detection scales is expected to enhance CD performance. Learning from the limited prior change information transferred between multitemporal HSIs could be another interesting research direction.

FAST COMPUTING

The high-dimensional nature of hyperspectral data sets, together with the complexity of the processing algorithms, calls for advanced processing techniques to accelerate hyperspectral-related computations [310]. Traditionally, software for hyperspectral analysis has been written for serial computation, i.e., to be run on a single computer having a single CPU in which a problem is broken down into a discrete series of instructions. Instructions are then executed one after another so that only one may execute at any moment in time. In turn, parallel computing allows the simultaneous use of multiple computer resources. In other words, a problem is run on multiple CPUs by being broken down into discrete parts so that each part can be executed simultaneously. *Load balancing* refers to the practice of distributing work among tasks so that all tasks are kept busy all the time. This practice is considered to minimize task idle time.

PARALLEL COMPUTING ALLOWS THE SIMULTANEOUS USE OF MULTIPLE COMPUTER RESOURCES.

Taking advantage of these concepts, several techniques have been developed to accelerate hyperspectral imaging computation. In this section, we summarize some of the available strategies, which vary according to the adopted platform for fast computing and acceleration.

CLUSTER COMPUTERS

Perhaps the most widely used high-performance computing architecture for accelerating hyperspectral-related computations is cluster computing. A cluster is a collection of commodity computers interconnected by a computer network. To efficiently execute a parallel problem in a cluster, partitioning strategies can be used so that the original problem is broken down into subtasks allocated to the different computers.

In the case of hyperspectral imaging, two kinds of strategies have been used to partition the original hyperspectral data set into subsets for efficient cluster-based processing. In spectral-domain partitioning, a single pixel vector (spectral signature) may be stored in different processing units, and communications would be required for individual pixel-based calculations, such as spectral angle computations [311]. In spatial-domain partitioning, every pixel vector (spectral signature) is stored in the same processing unit [312]. These concepts have been explored not only in clusters [313], [314] but also in heterogeneous workstation networks [315].

RECONFIGURABILITY IS NO LONGER A PROMISE BUT A REALITY, AND IT IS FEASIBLE TO HAVE SEVERAL ALGORITHMS IMPLEMENTED ON THE SAME BOARD.

HARDWARE ACCELERATORS

In addition to clusters, other kinds of high-performance computing architectures have been used. Specifically, onboard processing of HSIs is required in certain contexts. This is particularly the case in time-critical applications [316], [317]. For this purpose, other solutions have been explored. In particular, reconfigurable computing provides higher performance (throughput and processing power) compared to other specialized hardware processors [318]. Reconfigurability is no longer a promise but a reality, and it is feasible to have several algorithms implemented on the same board and dynamically select one out of a pool of algorithms from a control station [319]. A field-programmable gate array (FPGA) is a chip in which there is a matrix of blank cells called *configurable logic blocks*. This device can be used to implement any circuit (provided there is a sufficient number of logic blocks). FPGAs have been widely used to accelerate hyperspectral imaging algorithms for onboard processing [320]–[322].

Another high-performance computing architecture that has provided excellent performance when accelerating hyperspectral imaging computations is the graphical processing unit (GPU) [323]. GPUs are now fully programmable using high-level languages such as NVIDIA CUDA (http://developer.nvidia.com). The GPU specializes in computer-intensive, massive-data-parallel computation (which exactly describes graphics rendering). Therefore, more transistors can be devoted to data processing rather than to data caching and flow control. The fast-growing video game industry exerts strong economic pressure for constant innovation. This has motivated the extended use of GPUs for accelerating many different hyperspectral imaging-related tasks [316], [324]–[335].

CLOUD COMPUTING

Recently, more sophisticated high-performance architectures have been used to process hyperspectral data. For instance, cloud computing platforms are increasingly employed for processing hyperspectral data in distributed architectures. In this sense, cloud computing offers advanced capabilities for service-oriented and high-performance computing [336]. Furthermore, using cloud computing to analyze large repositories of hyperspectral data can be considered a natural solution resulting from the evolution of techniques previously developed for other types of computing platforms [311]. In particular, using GPUs within distributed scenarios has been radically extended worldwide, thanks in part to the increasing development of deep learning-based frameworks (e.g., Apache Spark, Caffe, Theano, Torch, and TensorFlow), which also find application in the hyperspectral analysis community [64], [65], [84], [337], [338].

However, the recent literature still provides few examples of the use of cloud infrastructures to implement hyperspectral analysis techniques in general and to perform supervised classification of hyperspectral data in particular. This may be due to the lack of open repositories of HSIs available for public use, a situation that is expected to change in the near future, as large distributed repositories of hyperspectral data for open use become available to the scientific community.

CHALLENGES IN FAST COMPUTING

The most important challenge related to consolidating fast-computing techniques to analyze hyperspectral data (particularly in the context of real-time platforms) is still the high energy consumption required by high-performance computing architectures, which reduces their applicability in real scenarios for onboard operation. Currently, the power consumption required by devices such as GPUs is too high for their incorporation into satellite platforms. Another shortcoming is the fact that these platforms are often subject to radiation tolerance issues. Future developments in hardware instruments for onboard operation are required for efficient real-time processing of HSIs, particularly in the context of satellite missions.

CONCLUSIONS

The role of HSI analysis cannot be underestimated for a plethora of applications, especially those related to CD and scene classification. Without a doubt, the use of such valuable data has been well established in the remote-sensing community, and the precise investigation of such data is increasing significantly. To this end, the considerable number of airborne and spaceborne hyperspectral missions as well as the increasing number of scientific publications on this particular subject demonstrate that the area of HSI analysis is substantial, dynamic, and vibrant.

The field of hyperspectral imagery is extremely broad, and it is impossible to investigate it comprehensively in one literature review. This article focuses particularly on algorithmic approaches that have been developed, adapted, or proposed since 2013, covering a number of key research areas, such as DR, classification, spectral unmixing, resolution enhancement, image restoration, CD, and fast computing. It is certainly of significant interest to summarize other survey articles from the application point of view, where the usefulness of hyperspectral imagery can be demonstrated through different practical aspects such as mineralogy, environmental mapping and monitoring, geology, and so on.

In addition to the material presented in this article, hyperspectral data preprocessing plays a vital role in fostering application-oriented tasks. This important subject lies beyond the scope of this investigation, which mainly focuses on the development of algorithms for HSI analysis and processing. To this end, we see the need for some tutorials and survey articles with a primary focus on hyperspectral data preprocessing and preparation designed for atmospheric corrections, geometric and radiometric corrections, coregistration, and quality assessment.

As indicated several times throughout this article, although the area of HSI analysis is well established, there are still many doors open for further investigation. We hope that our work here will raise new possibilities for researchers to further investigate the remaining issues by developing fast, accurate, and automatic methodologies suitable for applications at hand.

ACKNOWLEDGMENTS

We extend our sincere appreciation to Prof. Paolo Gamba of the University of Pavia, Italy, and Prof. D. Landgrebe of Purdue University, West Lafayette, Indiana, for providing the Pavia University and Indian Pines data sets, respectively. In addition, the authors thank the National Center for Airborne Laser Mapping at the University of Houston, Texas, for providing the CASI Houston data set and the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee for organizing the 2013 Data Fusion Contest.

AUTHOR INFORMATION

Pedram Ghamisi (p.ghamisi@gmail.com) is with the German Aerospace Center, Remote Sensing Technology Institute, and the Technische Universität München, Signal Processing in Earth Observation, Munich, Germany. He is a Member of the IEEE.

Naoto Yokoya (yokoya@sal.rcast.u-tokyo.ac.jp) is with the Department of Advanced Interdisciplinary Studies, the University of Tokyo, Japan. He is a Member of the IEEE.

Jun Li (lijun48@mail.sysu.edu.cn) is with the Guangdong Provincial Key Laboratory of Urbanization and Geosimulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou, China. He is a Senior Member of the IEEE.

Wenzhi Liao (wenzhi.liao@ugent.be) is with the Department of Telecommunications and Information Processing, Ghent University, Belgium. He is a Senior Member of the IEEE.

Sicong Liu (sicong.liu@tongji.edu.cn) is with the College of Surveying and Geoinformatics, Tongji University, Shanghai, China. He is a Member of the IEEE.

Javier Plaza is with the Department of Technology of Computers and Communications, University of Extremadura, Spain. He is a Senior Member of the IEEE.

Behnood Rasti (behnood@hi.is) is with the Keilir Institute of Technology, Reykjanesbaer, Iceland, and the Department of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland. He is a Member of the IEEE.

Antonio Plaza is with the Department of Technology of Computers and Communications, University of Extremadura, Spain. He is a Fellow of the IEEE.

REFERENCES

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

[2] N. Keshava, J. Kerekes, D. Manolakis, and G. Shaw, "An algorithm taxonomy for hyperspectral unmixing," in *Proc. SPIE AeroSense Conf. Algorithms for Multispectral and Hyperspectral Imagery VI*, vol. 4049, 2000, pp. 42–63.

[3] E. J. Wegman, "Hyperdimensional data analysis using parallel coordinates," *J. Amer. Statist. Assoc.*, vol. 85, no. 411, pp. 664–675, 1990.

[4] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Norwood, MA: Artech House, 2015.

[5] A. Eckardt, J. Horack, F. Lehmann, D. Krutz, J. Drescher, M. Whorton, and M. Soutullo, "DESIS (DLR Earth sensing imaging spectrometer for the ISS-MUSES platform)," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, July 2015, pp. 1457–1459.

[6] L. Guanter, H. Kaufmann, K. Segl, S. Förster, C. Rogaß, S. Chabrillat, T. Küster, A. Hollstein, G. Rossner, C. Chlebek, C. Straif, S. Fischer, S. Schrader, T. Storch, U. Heiden, A. Mueller, M. Bachmann, H. Mühle, R. Müller, M. Habermeyer, A. Ohndorf, J. Hill, H. Buddenbaum, P. Hostert, S. van der Linden, P. J. Leitão, A. Rabe, R. Doerffer, H. Krasemann, H. Xi, W. Mauser, T. Hank, M. Locherer, M. Rast, K. Staenz, and B. Sang, "The EnMAP spaceborne imaging spectroscopy mission for earth observation," *Remote Sens.*, vol. 7, no. 7, pp. 8830–8857, 2015.

[7] A. Iwasaki, N. Ohgi, J. Tanii, T. Kawashima, and H. Inada, "Hyperspectral imager suite (HISUI): Japanese hyper-multi spectral radiometer," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, July 2011, pp. 1025–1028.

[8] P. Stefano, P. Angelo, P. Simone, R. Filomena, S. Federico, S. Tiziana, A. Umberto, C. Vincenzo, N. Acito, D. Marco, M. Stefania, C. Giovanni, C. Raffaele, D. B. Roberto, L. Giovanni, and A. Cristina, "The PRISMA hyperspectral mission: Science activities and opportunities for agriculture and land monitoring," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, July 2013, pp. 4558–4561.

[9] T. Feingersh and E. B. Dor, "Shalom: A commercial hyperspectral space mission," in *Optical Payloads for Space Missions*, S.-E. Qian, Ed. Chichester, U.K.: Wiley, 2015, ch. 11, pp. 247–263.

[10] R. Green, G. Asner, S. Ungar, and R. Knox, "NASA mission to measure global plant physiology and functional types," in *Proc. Aerospace Conf.*, Mar. 2008, pp. 1–7.

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

[12] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Unsupervised multitemporal spectral unmixing for detecting multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 1–16, 2016.

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

[14] D. M. O. Kuybeda and M. Barzohar, "Rank estimation and redundancy reduction of high dimensional noisy signals with preservation of rare vectors," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5579–5592, 2007.

[15] J. Ren, J. Zabalza, S. Marshall, and J. Zheng, "Effective feature extraction and data reduction in remote sensing using hyperspectral imaging," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 149–154, June 2014.

[16] J. Yang and W. Sun, "Automatic analysis of the slight change image for unsupervised change detection," *J. Appl. Remote Sens.*, vol. 9, no. 1, pp. 1–24, 2015.

[17] B. Somers and G. P. Asnerb, "Multi-temporal hyperspectral mixture analysis and feature selection for invasive species mapping in rainforests," *Remote Sens. Environ.*, vol. 136, pp. 14–27, May 2013.

[18] L. Liu, N. C. Coops, N. W. Aven, and Y. Pang, "Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data," *Remote Sens. Environ.*, vol. 200, pp. 170–182, 2017.

[19] L. van der Maaten, E. Postma, and J. van den Herik, "Dimensionality reduction: A comparative review," Tilburg University, The Netherlands, Tech. Rep., TiCC-TR 2009-005, 2009.

[20] I. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.

[21] A. Hyvrinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[22] A. Green, M. Berman, P. Switzer, and M. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[23] X. F. He, D. Cai, S. C. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Computer Vision*, vol. 2, 2005, pp. 1208–1213.

[24] X. F. He and P. Niyogi, *Locality Preserving Projections*. Cambridge, MA: MIT Press, 2004.

[25] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4172–4186, Dec. 2007.

[26] A. Nielsen, "Kernel maximum autocorrelation factor and minimum noise fraction transformations," *IEEE Trans. Image Process.*, vol. no. 203, pp. 612–624, Mar. 2011.

[27] G. Chen and S.-E. Qian, "Dimensionality reduction of hyperspectral imagery using improved locally linear embedding," *J. Appl. Remote Sens.*, vol. 1, pp. 1–10, 2007.

[28] P. Mitra, C. A. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 301–312, 2002.

[29] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogramm. Eng. Remote Sens. J.*, vol. 70, no. 7, pp. 793–802, 2004.

[30] J. M. Sotoca, F. Pla, and J. S. Sanchez, "Band selection in multispectral images by minimization of dependent information," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 2, pp. 258–267, 2007.

[31] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS Data Fusion Contest," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, June 2014.

[32] W. Liao, M. D. Mura, J. Chanussot, and A. Pizurica, "Fusion of spectral and spatial information for classification of hyperspectral remote-sensed imagery by local graph," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 2, pp. 583–594, 2016.

[33] M. Borhani and H. Ghassemian, "Kernel multivariate spectralspatial analysis of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 2418–2426, 2015.

[34] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, 2017. doi: 10.1109/TCYB.2016.2605044.

[35] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, 2016.

[36] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6976–6985, Dec. 2016.

[37] C.-I. Chang and H. Ren, "An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1044–1063, 2000.

[38] B. Kuo and D. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.

[39] P. H. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*. New York: McGraw-Hill, 1978.

[40] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–550, July 1994.

[41] Q. Du, N. Younan, R. King, and V. Shah, "On the performance evaluation of pan-sharpening techniques," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 518–522, Oct. 2007.

[42] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant

analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, 2009.

[43] B. C. Kuo, C. W. Chang, C. C. Hung, and H. P. Wang, "A modified nonparametric weight feature extraction using spatial and spectral information," in *Proc. Int. Geoscience and Remote Sensing Symp.*, 2006, pp. 172–175.

[44] B. Kuo, C. Li, and J. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1139–1155, Apr. 2009.

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

[46] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.

[47] P. A. Estevez, T. Michel, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, 2009.

[48] W. Li, S. Prasad, J. E. Fowled, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, 2012.

[49] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, no. 5, pp. 1027–1061, 2007.

[50] 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, 2015.

[51] X. Cao, T. Xiong, and L. Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 329–333, 2016.

[52] Y. Dong, B. Du, L. Zhang, and L. Zhang, "Dimensionality reduction and classification of hyperspectral images using ensemble discriminative local metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2509–2524, May 2017.

[53] N. H. Ly, D. Qian, and J. E. Fowler, "Sparse graph-based discriminant analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3872–3884, 2014.

[54] 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. 1–20, 2015.

[55] Q. Zhang, Y. Tian, Y. Yang, and C. Pan, "Automatic spatial-spectral feature selection for hyperspectral image via discriminative sparse multimodal learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 261–279, 2015.

[56] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proc. Advances in Neural Information Processing Systems*, 2009, pp. 2223–2231.

[57] W. He, H. Zhang, L. Zhang, W. Philips, and W. Liao, "Weighted sparse graph based dimensionality reduction for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 686–690, 2016.

[58] D. Whitley, "A genetic algorithm tutorial," *Statist. Computing*, vol. 4, no. 2, pp. 65–85, 1994.

[59] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Networks*, vol. IV, 1995, pp. 1942–1948.

[60] M. Pedergnana, P. R. Marpu, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "A novel technique for optimal feature selection in attribute profiles based on genetic algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3514–3528, 2013.

[61] P. Ghamisi, M. S. Couceiro, and J. A. Benediktsson, "A novel feature selection approach based on FODPSO and SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2935–2947, 2015.

[62] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[63] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980.

[64] 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, 2016.

[65] P. Ghamisi, Y. Chen, and X. Zhu, "A self-improving convolution neural network for the classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1537–1541, 2016.

[66] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Computational Learning Theory*, 1998, pp. 92–100.

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

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

[69] D. Cai, X. F. He, and J. W. Han, "Semi-supervised learning literature survey," in *Proc. IEEE 11th Int. Conf. Computer Vision*, 2008, pp. 1–7.

[70] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, 2011.

[71] J. Bai, S. Xiang, L. Shi, and C. Pan, "Semisupervised pair-wise band selection for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 2798–2813, 2015.

[72] M. Sugiyama, T. Ide, S. Nakajima, and J. Sese, "Semi-supervised local Fisher discriminant analysis for dimensionality reduction," *Mach. Learn.*, vol. 78, pp. 35–61, 2010.

[73] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, 2013.

[74] R. Luo, W. Liao, X. Huang, Y. Pi, and W. Philips, "Feature extraction of hyperspectral images with semi-supervised graph

learning," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 9, pp. 4389–4399, 2016.

[75] F. Luo, H. Huang, Z. Ma, and J. Liu, "Semisupervised sparse manifold discriminative analysis for feature extraction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6197–6211, 2016.

[76] D. Tuia, M. Volpi, M. Trolliet, and G. Camps-Valls, "Semisupervised manifold alignment of multimodal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7708–7720, 2014.

[77] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semisupervised transfer component analysis for domain adaptation in remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3550–3564, 2015.

[78] K. Uto, Y. Kosugi, and G. Saito, "Semi-supervised hyperspectral subspace learning based on a generalized eigenvalue problem for regression and dimensionality reduction," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sensing*, vol. 7, no. 6, pp. 2583–2599, 2014.

[79] F. Morsier, M. Borgeaud, V. Gass, J. Thiran, and D. Tuia, "Kernel low rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3410–3420, 2016.

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

[81] 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, 2016.

[82] W. Liao, J. Chanussot, M. D. Mura, X. Huang, R. Bellens, S. Gautama, and W. Philips, "Promoting partial reconstruction for the morphological analysis of very high resolution urban remote sensing images," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 8–28, 2017.

[83] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, 2010.

[84] 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.

[85] P. Ghamisi, A.-R. Ali, M. Couceiro, and J. Benediktsson, "A novel evolutionary swarm fuzzy clustering approach for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 2447–2456, 2015.

[86] Y. Chen, S. Ma, X. Chen, and P. Ghamisi, "Hyperspectral data clustering based on density analysis ensemble," *Remote Sens. Lett.*, vol. 8, no. 2, pp. 194–203, 2017.

[87] Q. Jackson and D. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, 2002.

[88] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.

[89] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004.

[90] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Patt. Recog. Lett.*, vol. 27, no. 4, pp. 294–300, 2006.

[91] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 239–243, Jan. 2014.

[92] T. Rainforth and F. Wood. (2017, Aug.). Canonical correlation forests. arXiv. [Online]. Available: https://arxiv.org/abs/1507.05444.

[93] J. Xia, N. Yokoya, and A. Iwasaki, "Hyperspectral image classification with canonical correlation forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 421–431, Jan. 2017.

[94] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 540–552, 1990.

[95] M. Pal, "Extreme-learning-machine-based land cover classification," *Int. J. Remote Sens.*, vol. 30, no. 14, pp. 3835–3841, 2009.

[96] M. Pal, A. E. Maxwell, and T. A. Warner, "Kernel-based extreme learning machine for remote-sensing image classification," *Remote Sens. Lett.*, vol. 4, no. 9, pp. 853–862, 2013.

[97] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, 2010.

[98] 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. in Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, 2014.

[99] C. Tao, H. Pan, Y. Li, and Z. Zou, "Unsupervised spectralspatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2438–2442, 2015.

[100] 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. in Remote Sens.*, vol. 8, no. 6, pp. 2381–2292, 2015.

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

[102] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral–spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2565–2574, 2014.

[103] 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, 2015.

[104] R. Kemker and C. Kanan, "Self-taught feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2693–2705, May 2017.

[105] P. Ghamisi, M. S. Couceiro, J. A. Benediktsson, and N. M. F. Ferreira, "An efficient method for segmentation of

images based on fractional calculus and natural selection," *Expert Syst. Applicat.*, vol. 39, no. 16, pp. 12 407–12 417, 2012.

[106] P. Ghamisi, M. S. Couceiro, F. M. Martins, and J. A. Benediktsson, "Multilevel image segmentation approach for remote sensing images based on fractional-order Darwinian particle swarm optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2382–2394, 2014.

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

[108] P. Ghamisi, M. Couceiro, M. Fauvel, and J. A. Benediktsson, "Integration of segmentation techniques for classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 342–346, Jan. 2014.

[109] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, 2010.

[110] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, pp. 62–66, 1979.

[111] 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, 2005.

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

[113] P. Ghamisi, M. Dalla 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, 2015.

[114] M. Pedergnana, P. R. Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "A novel technique for optimal feature selection in attribute profiles based on genetic algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3514–3528, June 2013.

[115] P. Ghamisi, J. A. Benediktsson, G. Cavallaro, and A. Plaza, "Automatic framework for spectral-spatial classification based on supervised feature extraction and morphological attribute profiles," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2147–2160, 2014.

[116] 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. 5, pp. 5771–5782, 2014.

[117] B. Song, J. Li, M. Dalla Mura, P. Li, A. Plaza, J. M. Bioucas-Dias, J. A. Benediktsson, and J. Chanussot, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, 2014.

[118] X. Huang, X. Guan, J. A. Benediktsson, L. Zhang, J. Li, A. Plaza, and M. Dalla Mura, "Multiple morphological profiles from multicomponent base images for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4653–4669, Dec. 2014.

[119] P. Ghamisi, R. Souza, J. A. Beneiktsson, X. X. Zhu, L. Rittner, and R. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, 2016.

[120] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1641–1645, 2016.

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

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

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

[124] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.

[125] Z. Zhang, E. Pasolli, M. M. Crawford, and J. C. Tilton, "An active learning framework for hyperspectral image classification using hierarchical segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 2, pp. 640–654, Feb. 2016.

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

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

[128] L. Wan, K. Tang, M. Li, Y. Zhong, and A. K. Qin, "Collaborative active and semisupervised learning for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2384–2396, May 2015.

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

[130] Z. Lai, W. K. Wong, Y. Xu, C. Zhao, and M. Sun. (2014, Oct.). Sparse alignment for robust tensor learning. *IEEE Trans. Neural Netw. Learn. Syst.* [Online]. 25(10), pp. 1779–1792. Available: http://dx.doi.org/10.1109/TNNLS.2013.2295717

[131] A. Castrodad, Z. Xing, J. Greer, E. Bosch, L. Carin, and G. Sapiro, "Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4263–4281, Dec. 2011.

[132] A. Goetz, G. Vane, J. Solomon, and B. Rock, "Imaging spectrometry for Earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, June 1985.

[133] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, June 2014.

[134] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[135] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, 2004.

[136] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, 2008.

[137] B. Luo, J. Chanussot, S. Doute, and L. Zhang, "Empirical automatic estimation of the number of endmembers in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 24–28, Jan. 2013.

[138] O. Eches, N. Dobigeon, and J. Y. Tourneret, "Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical Bayesian algorithm," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 582–591, June 2010.

[139] J. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. JPL Airborne Earth Science Workshop*, 1995, pp. 23–26.

[140] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. Society of Photographic Instrumentation Engineers, Imaging Spectrometry V*, 2003, vol. 3753, pp. 266–277.

[141] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, 1994.

[142] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, 2005.

[143] B. Somers, S. Delalieux, J. Stuckens, W. W. Verstraeten, and P. Coppin, "A weighted linear spectral mixture analysis approach to address endmember variability in agricultural production systems," *Int. J. Remote Sens.*, vol. 30, no. 1, pp. 139–147, 2009.

[144] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: a statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, 2004.

[145] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, 2007.

[146] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, July 2007.

[147] T. H. Chan, C. Y. Chi, Y. M. Huang, and W. K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[148] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sept. 2015.

[149] E. M. T. Hendrix, I. Garcia, J. Plaza, G. Martin, and A. Plaza, "A new minimum-volume enclosing algorithm for endmember identification and abundance estimation in hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2744–2757, July 2012.

[150] C. Chen and X. Zhang, "Independent component analysis for remote sensing study," *in Proc. SPIE Image and Signal Processing Remote Sensing V*, vol. 3871, 1999, pp. 150–158.

[151] J. Nascimento and J. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, 2012.

[152] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2025–2041, 2002.

[153] D. M. Rogge, B. Rivard, J. Zhang, A. Sánchez, J. Harris, and J. Feng, "Integration of spatial–spectral information for the improved extraction of endmembers," *Remote Sens. Environ.*, vol. 110, no. 3, pp. 287–303, 2007.

[154] M. Zortea and A. Plaza, "Spatial preprocessing for endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, pp. 2679–2693, 2009.

[155] G. Martin and A. Plaza, "Region-based spatial preprocessing for endmember extraction and spectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 745–749, July 2011.

[156] G. Martin and A. Plaza, "Spatial-spectral preprocessing prior to endmember identification and unmixing of remotely sensed hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sensing*, vol. 5, no. 2, pp. 380–395, Apr. 2012.

[157] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA: Kluwer, 2003.

[158] A. Halimi, P. Honeine, and J. M. Bioucas-Dias, "Hyperspectral unmixing in presence of endmember variability, nonlinearity, or mismodeling effects," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4565–4579, 2016.

[159] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, 2016.

[160] L. Drumetz, M.-A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, 2016.

[161] M. D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, 2011.

[162] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing,"

IEEE Trans. Geosci. Remote Sens., vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
[163] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Music-CSR: Multiple signal classification collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4364–4382, 2013.
[164] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, 2014.
[165] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6076–6090, Oct. 2016.
[166] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, 2014.
[167] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with l1-norm spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2654–2665, 2014.
[168] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4810–4819, 2015.
[169] A. Marinoni, J. Plaza, A. Plaza, and P. Gamba, "Nonlinear hyperspectral unmixing using nonlinearity order estimation and polytope decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 2644–2654, June 2015.
[170] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.
[171] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2442–2453, May 2013.
[172] N. Yokoya, J. Chanussot, and A. Iwasaki, "Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1430–1437, 2014.
[173] Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Abundance estimation for bilinear mixture models via joint sparse and low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4404–4423, 2014.
[174] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a post-nonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017–3025, 2012.
[175] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Ttrans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 240–251, 2016.
[176] B. Hapke, "Bidirectional reflectance spectroscopy: 1. Theory," *J. Geophys. Res.*, vol. 86, no. B4, pp. 3039–3054, Apr. 1981.
[177] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation mode," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, 2013.
[178] L. Zhang, B. Wu, B. Huang, and P. Li, "Nonlinear estimation of subpixel proportion via kernel least square regression," *Int. J. Remote Sens.*, vol. 28, no. 18, pp. 4157–4172, 2007.
[179] M. Brown, H. G. Lewis, and S. R. Gunn, "Linear spectral mixture models and support vector machines for remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2346–2360, 2000.
[180] L. Wang and X. Jia, "Integration of soft and hard classifications using extended support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 6, no. 3, pp. 543–547, 2009.
[181] G. Foody, "Relating the land-cover composition of mixed pixels to artificial neural network classification output," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 5, pp. 491–499, 1996.
[182] G. Foody, R. Lucas, P. Curran, and M. Honzak, "Non-linear mixture modeling without end-members using an artificial neural network," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 937–953, 1997.
[183] A. Baraldi, E. Binaghi, P. Blonda, P. A. Brivio, and A. Rampini, "Comparison of the multilayer perceptron with neuro-fuzzy techniques in the estimation of cover class mixture in remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 5, pp. 994–1005, 2001.
[184] J. Plaza, A. Plaza, R. Perez, and P. Martinez, "On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images," *Pattern Recogn.*, vol. 42, no. 11, pp. 3032–3045, Nov. 2009.
[185] J. Plaza, A. Plaza, R. Perez, and P. Martinez, "Joint linear/nonlinear spectral unmixing of hyperspectral image data," in *Proc. Int. Geoscience and Remote Sensing Symp.*, Barcelona, Spain, July 2007, p.4037–4040.
[186] G. Licciardi and F. Del Frate, "Pixel unmixing in hyperspectral data by means of neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4163–4172, Nov. 2011.
[187] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. Int. Conf. Computer Vision*, 2009, pp. 349–356.
[188] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1860–1875, 2005.
[189] J. C.-W. Chan, J. Ma, and F. Canters, "A comparison of superresolution reconstruction methods for multi-angle CHRIS/Proba images," in *Proc. SPIE, Image and Signal Processing for Remote Sensing XIV*, 710904, vol. 7109, Oct. 1, 2008. doi: 10.1117/12.800256
[190] J. C.-W. Chan, J. Ma, P. Kempeneers, and F. Canters, "Superresolution enhancement of hyperspectral CHRIS/Proba images with a thin-plate spline nonrigid transform model," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2569–2579, 2010.
[191] J. C.-W. Chan, J. Ma, T. Van de Voorde, and F. Canters, "Preliminary results of superresolution-enhanced angular hyperspectral (CHRIS/Proba) images for land-cover classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1011–1015, 2011.
[192] L. Demarchi, J. C.-W. Chan, J. Ma, and F. Canters, "Mapping impervious surfaces from superresolution enhanced CHRIS/

Proba imagery using multiple endmember unmixing," *ISPRS J. Photogramm. Remote Sens.*, vol. 72, pp. 99–112, Aug. 2012.

[193] S. Qian and G. Chen, "Enhancing spatial resolution of hyperspectral imagery using sensor's intrinsic keystone distortion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 12, pp. 5033–5048, 2012.

[194] Y. Zhao, J. Yang, Q. Zhang, L. Song, Y. Cheng, and Q. Pan, "Hyperspectral imagery super-resolution by sparse representation and spectral regularization," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 87, 2011.

[195] Y. Zhao, J. Yang, and J. C.-W. Chan, "Hyperspectral imagery super-resolution by spatial–spectral joint nonlocal similarity," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 7, no. 6, pp. 2671–2679, 2014.

[196] R. C. Patel and M. V. Joshi, "Super-resolution of hyperspectral images: Use of optimum wavelet filter coefficients and sparsity regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1728–1736, 2015.

[197] P. M. Atkinson, "Mapping sub-pixel boundaries from remotely sensed images," *Innovations in GIS*, vol. 4, pp. 166–180, Apr. 1997.

[198] X. Tong, X. Zhang, J. Shan, H. Xie, and M. Liu, "Attraction-repulsion model-based subpixel mapping of multi-/hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2799–2814, 2013.

[199] Y. Zhang, Y. Du, F. Ling, S. Fang, and X. Li, "Example-based super-resolution land cover mapping using support vector regression," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 7, no. 4, pp. 1271–1283, 2014.

[200] Y. Zhong, Y. Wu, X. Xu, and L. Zhang, "An adaptive subpixel mapping method based on map model and class determination strategy for hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1411–1426, 2015.

[201] R. Feng, Y. Zhong, Y. Wu, D. He, X. Xu, and L. Zhang, "Nonlocal total variation subpixel mapping for hyperspectral remote sensing imagery," *Remote Sens.*, vol. 8, no. 3, p. 250, 2016.

[202] X. Tong, X. Xu, A. Plaza, H. Xie, H. Pan, W. Cao, and D. Lv, "A new genetic method for subpixel mapping using hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 9, pp. 4480–4491, 2016.

[203] J. Pearlman, S. Carman, C. Segal, P. Jarecke, P. Clancy, and W. Browne, "Overview of the Hyperion imaging spectrometer for the NASA EO-1 mission," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, vol. 7, 2001, pp. 3036–3038.

[204] W. Carper, T. M. Lillesand, and P. W. Kiefer, "The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, Apr. 1990.

[205] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6 011 875 A, 2000.

[206] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS + Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.

[207] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Jan. 2000.

[208] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

[209] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[210] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May 2013.

[211] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods: Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, 2005.

[212] B. Li, R. Yang, and H. Jiang, "Remote-sensing image compression using two-dimensional oriented wavelet transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 236–250, Jan. 2011.

[213] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, *Remote Sensing Image Fusion*. Boca Raton, FL: CRC, 2015.

[214] L. Loncan, L. B. Almeida, J. B. Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simões, J. Y. Tourneret, M. A. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sept. 2015.

[215] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, June 2017.

[216] Z. Chen, H. Pu, B. Wang, and G.-M. Jiang, "Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 8, pp. 1418–1422, Aug. 2014.

[217] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hyper-sharpening: A first approach on SIM-GA data," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 3008–3024, June 2015.

[218] C. Grohnfeldt, "Multi-sensor data fusion for multi- and hyperspectral resolution enhancement based on sparse representations," Ph.D. dissertation, Technical University of Munich, Germany, 2017.

[219] M. T. Eismann and R. C. Hardie, "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 455–465, Mar. 2005.

[220] O. Berné, A. Helens, P. Pilleri, and C. Joblin, "Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy," in *Proc. IEEE Workshop Hyperspectral Image and Signal Processing (WHISPERS)*, June 2010, pp. 1–4.

[221] N. Yokoya, T. Yairi, and A. Iwasaki, "Hyperspectral, multispectral, and panchromatic data fusion based on non-negative matrix factorization," *in Proc. IEEE Workshop Hyperspectral Image and Signal Processing (WHISPERS)*, June 2011, pp. 1–4.

[222] R. Kawakami, J. Wright, Y. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," *in Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2011, pp. 2329–2336.

[223] J. Bieniarz, D. Cerra, J. Avbelj, P. Reinartz, and R. Müller, "Hyperspectral image resolution enhancement based on specral unmixing and information fusion," in *Proc. ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Sept. 2011, pp. 33–37.

[224] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[225] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. European Conf. Computer Vision*, 2014, pp. 63–78.

[226] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu, "Spatial and spectral image fusion using sparse matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1693–1704, Mar. 2014.

[227] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Computer Vision*, Dec. 2015, pp. 3586–3594.

[228] C. Lanaras, E. Baltsavias, and K. Schindler, "Advances in hyperspectral and multispectral image fusion and spectral unmixing," in *Proc. ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-3/W3, 2015, pp. 451–458.

[229] M. Simões, J. B. Dias, L. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, June 2015.

[230] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multiband images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sept. 2015.

[231] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multiband images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.

[232] M. Veganzones, M. Simões, G. Licciardi, N. Yokoya, J. Bioucas-Dias, and J. Chanussot, "Hyperspectral superresolution of locally low rank images from complementary multisource data," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 274–288, Jan. 2016.

[233] N. Yokoya, J. C. W. Chan, and K. Segl, "Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated EnMAP and Sentinel-2 images," *Remote Sens.*, vol. 8, no. 3, p. 172, 2016.

[234] J. P. Kerekes and J. E. Baum, "Hyperspectral imaging system modeling," *Lincoln Laboratory*, vol. 14, no. 1, pp. 117–130, 2003.

[235] D. Landgrebe and E. Malaret, "Noise in remote-sensing systems: The effect on classification error," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-24, no. 2, pp. 294– 300, Mar. 1986.

[236] B. Rasti, J. R. Sveinsson, and M. O. Ulfarsson, "SURE based model selection for hyperspectral imaging," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, pp. 4636–4639, July 2014.

[237] H. K. Aggarwal and A. Majumdar. (2015). Exploiting spatiospectral correlation for impulse denoising in hyperspectral images. *J. Electron. Imaging.* [Online]. *24*(*1*). Available: http://dx.doi.org/10.1117/1.JEI.24.1.013027

[238] L. Gomez-Chova, L. Alonso, L. Guanter, G. Camps-Valls, J. Calpe, and J. Moreno, "Correction of systematic spatial noise in push-broom hyperspectral sensors: Application to CHRIS/PROBA images," *Appl. Optics*, vol. 47, no. 28, pp. f46–f60, 2008.

[239] M. di Bisceglie, R. Episcopo, C. Galdi, and S. L. Ullo, "Destriping modis data using overlapping field-of-view method," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 2, pp. 637–651, Feb. 2009.

[240] P. Rakwatin, W. Takeuchi, and Y. Yasuoka, "Stripe noise reduction in modis data by combining histogram matching with facet filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1844–1856, June 2007.

[241] N. Acito, M. Diani, and G. Corsini, "Subspace-based striping noise reduction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1325–1342, Apr. 2011.

[242] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, July 2013.

[243] P. Meza, J. E. Pezoa, and S. N. Torres, "Multidimensional striping noise compensation in hyperspectral imaging: Exploiting hypercubes' spatial, spectral, and temporal redundancy," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 9, pp. 4428–4441, Sept. 2016.

[244] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[245] N. Acito, M. Diani, and G. Corsini, "Signal-dependent noise modeling and model parameter estimation in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2957–2971, 2011.

[246] B. Rasti, J. Sveinsson, and M. Ulfarsson, "Wavelet-based sparse reduced-rank regression for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6688–6698, Oct. 2014.

[247] I. Atkinson, F. Kamalabadi, and D. Jones, "Wavelet-based hyperspectral image estimation," in *Proc.* IEEE *Int. Geoscience and Remote Sensing Symp.*, vol. 2, July 2003, pp. 743–745.

[248] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using 3D wavelets," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, pp. 1349–1352, 2012.

[249] A. Zelinski and V. Goyal, "Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, pp. 387–390, Aug 2006.

[250] B. Rasti, "Sparse hyperspectral image modeling and restoration," Ph.D. dissertation, Univ. Iceland, Reykjavik, Dec. 2014.

[251] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 7, no. 6, pp. 2458–2467, June 2014.

[252] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Wavelet based hyperspectral image restoration using spatial and spectral penalties," in *Proc. Society of Photographic Instrumentation Engineers*, vol. 8892, 2013, pp. 88 920I–88 920I–8. doi: 10.1117/12.2029257.

[253] H. Zhang, "Hyperspectral image denoising with cubic total variation model," *ISPRS Ann. Photogramm., Remote Sens. and Spatial Inform. Sci.*, vol. I-7, pp. 95–98, 2012. doi: 10.5194/isprsannals-I-7-95-2012.

[254] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, 2012.

[255] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.

[256] S. L. Chen, X. Y. Hu, and S. L. Peng, "Hyperspectral imagery denoising using a spatial-spectral domain mixing prior," *J. Comput. Sci. Technol.*, vol. 27, no. 4, pp. 851–861, 2012.

[257] G. Chen and S. E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Mar. 2011.

[258] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sept. 1966.

[259] L. D. Lathauwer, B. D. Moor, J. Vandewalle, "On the best Rank-1 and Rank-(R1,R2,…,Rn) approximation of higher-order tensors," *SIAM J. Matrix Anal. Applicat.*, vol. 21, no. 4, pp. 1324–1342, Mar. 2000.

[260] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.

[261] A. Karami, M. Yazdi, and A. Asli, "Best rank-r tensor selection using genetic algorithm for better noise reduction and compression of hyperspectral images," in *Proc. 5th Int. Conf. Digital Information Management (ICDIM)*, July 2010, pp. 169–173.

[262] A. Karami, M. Yazdi, and A. Zolghadre Asli, "Noise reduction of hyperspectral images using kernel non-negative tucker decomposition," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 487–493, June 2011.

[263] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, July 2008.

[264] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.

[265] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "A new linear model and sparse regularization," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, July 2013, pp. 457–460.

[266] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image restoration using wavelets," *Proc. Society of Photographic Instrumentation Engineers*, vol. 8892, p. 889. 2013. doi: 10.1117/12.2029240.

[267] B. Rasti, J. R. Sveinsson, and M. O. Ulfarsson, "Total variation based hyperspectral feature extraction," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, July 2014, pp. 4644–4647.

[268] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr. 2013.

[269] X. Liu, S. Bourennane, and C. Fossati, "Reduction of signal-dependent noise from hyperspectral images for target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5396–5411, Sept. 2014.

[270] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[271] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[272] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, June 2015.

[273] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, "Hyperspectral image restoration via iteratively regularized weighted Schatten p-norm minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4642–4659, Aug. 2016.

[274] D. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[275] B. Rasti. (2016, May). FORPDN_SURE. [Online]. Available: https://www.researchgate.net/publication/303445288_FORPDN_SURE

[276] B. Rasti. (2016, May). HySURE. [Online]. Available: https://www.researchgate.net/publication/303784304_HySURE

[277] B. Rasti. (2016, May). Wavelab_fast. [Online]. Available: https://www.researchgate.net/publication/303445667_Wavelab_fast

[278] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 609–630, 2013.

[279] P. Du, S. Liu, P. Gamba, and K. Tan, "Fusion of difference images for change detection over urban areas," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 5, no. 4, pp. 1076–1086, 2012.

[280] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Digital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, 2004.

[281] S. Liu, M. Chi, Y. Zou, A. Samat, J. A. Benediktsson, and A. Plaza, "Oil spill detection via multitemporal optical remote sensing images: a change detection perspective," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 324–328, 2017.

[282] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Hierarchical unsupervised change detection in multitemporal hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 244–260, 2015.

[283] L. Bruzzone, S. Liu, F. Bovolo, and P. Du, "Change detection in multitemporal hyperspectral images," in *Multitemporal Remote Sensing Remote Sensing and Digital Image Processing*, vol. 20, Y. Ban, Ed. Cham, Switzerland: Springer, 2016, pp. 63–88.

[284] C. Wu, L. Zhang, and B. Du, "Hyperspectral anomaly change detection with slow feature analysis," *Neurocomput.*, vol. 151, pp. 175–187, Mar. 2015.

[285] A. Schaum and A. Stocker, "Long-interval chronochrome target detection," in *Proc. Int. Symp. Spectral Sensing Research*, 1998, pp. 1760–1770.

[286] A. P. Schaum and A. Stocker, "Hyperspectral change detection and supervised matched filtering based on covariance equalization," in *Proc. SPIE 5425, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X, The Int. Society for Optical Engineering*, vol. 5425, 2004, pp. 77–90.

[287] J. Theiler, C. Scovel, B. Wohlberg, and B. R. Foy, "Elliptically contoured distributions for anomalous change detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 271–275, Apr. 2010.

[288] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, 2016.

[289] K. Vongsy, M. J. Mendenhall, M. T. Eismann, and G. L. Peterson, "Removing parallax-induced changes in hyperspectral change detection," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, 2012, pp. 2012–2015.

[290] J. Meola, M. T. Eismann, K. J. Barnard, and R. C. Hardie, "Analysis of hyperspectral change detection as affected by vegetation and illumination variations," in *Proc.* SPIE 6565, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII, 2007. doi: 10.1117/12.716324.

[291] M. T. Eismann, J. Meola, and R. C. Hardie, "Hyperspectral change detection in the presence of diurnal and seasonal variations," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 237–249, 2008.

[292] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, 2000.

[293] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, 2009.

[294] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, 1998.

[295] M. Frank and M. Canty, "Unsupervised change detection for hyperspectral images," in *Proc. 12th JPL Airborne Earth Science Workshop*, 2013, pp. 63–72.

[296] A. A. Nielsen, "The regularized iteratively reweighted mad method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, 2007.

[297] V. Ortizrivera, M. Vlezreyes, and B. Roysam, "Change detection in hyperspectral imagery using temporal principal components," in *SPIE: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*, 2006. doi: 10.1117/12.667961.

[298] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Unsupervised hierarchical spectral analysis for change detection in hyperspectral images," in *Proc. 4th Workshop Hyperspectral Image and Signal Processing (WHISPERS)*, 2012, pp. 1–4.

[299] C. Wu, B. Du, and L. Zhang, "A subspace-based change detection method for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 6, no. 2, pp. 815–830, 2013.

[300] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomput.*, vol. 148, pp. 363–375, Jan. 2015.

[301] Q. Du, L. Wasson, and R. King, "Unsupervised linear unmixing for change detection in multitemporal airborne hyperspectral imagery," in *Proc. 2005 Int. Workshop Analysis Multi-Temporal Remote Sensing Images*, 2005, pp. 136–140.

[302] A. Erturk and A. Plaza, "Informative change detection by unmixing for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1252–1256, 2015.

[303] A. Ertrk, M. D. Iordache, and A. Plaza, "Sparse unmixing-based change detection for multitemporal hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 2, pp. 708–719, 2015.

[304] P. Du, S. Liu, P. Liu, K. Tan, and L. Cheng, "Sub-pixel change detection for urban land-cover analysis via multi-temporal remote sensing images," *Geo-spatial Inf. Sci.*, vol. 17, no. 1, pp. 26–38, 2014.

[305] G. M. Foody, "Monitoring the magnitude of land-cover change around the southern limits of the Sahara," *Photogramm. Eng. Remote Sens.*, vol. 67, no. 7, pp. 841–848, 2001.

[306] S. Hemissi, I. R. Farah, K. S. Ettabaa, and B. Solaiman, "Multi-spectro-temporal analysis of hyperspectral imagery based on 3-D spectral modeling and multilinear algebra," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 199–216, 2013.

[307] S. Liu, L. Bruzzone, F. Bovolo, and M. Zanetti, "Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4363–4378, 2015.

[308] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2196–2212, 2012.

[309] S. Liu, Q. Du, and X. Tong, "Band selection for change detection from hyperspectral images," vol. 9, no. 10, p. 1008, 2017.

[310] A. Plaza, J. Plaza, A. Paz, and S. Sánchez, "Parallel hyperspectral image and signal processing," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 119–126, 2011.
[311] A. Plaza, Q. Du, Y.-L. Chang, and R. King, "High performance computing for hyperspectral remote sensing," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 4, no. 3, pp. 528–544, 2011.
[312] A. Plaza, D. Valencia, J. Plaza, and P. Martinez, "Commodity cluster-based parallel processing of hyperspectral imagery," *J. Parallel and Distributed Comput.*, vol. 66, pp. 345–358, 2006.
[313] S. Sánchez, A. Paz, G. Martin, and A. Plaza, "Parallel unmixing of remotely sensed hyperspectral images on commodity graphics processing units," *Concurrency and Computation: Practice and Experience*, vol. 23, no. 13, pp. 1538–1557, 2011.
[314] A. Plaza, D. Valencia, J. Plaza, and P. Martinez, "Commodity cluster-based parallel processing of hyperspectral imagery," *J. Parallel Distrib. Comput.*, vol. 66, no. 3, pp. 345–358, 2006.
[315] A. Plaza, J. Plaza, and A. Paz, "Parallel heterogeneous CBIR system for efficient hyperspectral image retrieval using spectral mixture analysis," *Concurrency and Computation: Practice and Experience*, vol. 22, no. 9, pp. 1138–1159, 2010.
[316] S. Sánchez, R. Ramalho, L. Sousa, and A. Plaza, "Real-time implementation of remotely sensed hyperspectral image unmixing on GPUs," *J. Real-Time Image Process.*, vol. 10, no. 3, pp. 469–483, Sept. 2015.
[317] C. Gonzalez, S. Sánchez, A. Paz, J. Resano, D. Mozos, and A. Plaza, "Use of FPGA or GPU-based architectures for remotely sensed hyperspectral image processing," *INTEGRATION: The VLSI J.*, vol. 46, no. 2, pp. 89–103, 2013.
[318] S. Lopez, T. Vladimirova, C. Gonzalez, J. Resano, D. Mozos, and A. Plaza, "The promise of reconfigurable computing for hyperspectral imaging onboard systems: A review and trends," *Proc. IEEE*, vol. 101, no. 3, pp. 698–722, 2013.
[319] L. Santos, L. Berrojo, J. Moreno, J. F. López, and R. Sarmiento, "Multispectral and hyperspectral lossless compressor for space applications (HyLoC): A low-complexity FPGA implementation of the CCSDS 123 standard," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 2, pp. 757–770, 2016.
[320] A. Plaza and C.-I. Chang, "Clusters versus FPGA for parallel processing of hyperspectral imagery," *Int. J. High Performance Comput. Applicat.*, vol. 22, no. 4, pp. 366–385, 2008.
[321] D. Keymeulen, N. Aranki, A. Bakhshi, H. Luong, C. Sarture, and D. Dolman, "Airborne demonstration of FPGA implementation of fast lossless hyperspectral data compression system," in *Proc. NASA/ESA Conf. Adaptive Hardware and Systems (AHS)*, 2014, pp. 278–284.
[322] C. Gonzalez, S. Lopez, D. Mozos, and R. Sarmiento, "FPGA implementation of the HySime algorithm for the determination of the number of endmembers in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 2870–2883, 2015.
[323] J. Nickolls and W. J. Dally, "The GPU computing era," *IEEE Micro*, vol. 30, no. 2, pp. 56–69, Mar. 2010.
[324] S. Bernabé, G. Botella, J. Navarro, C. Orueta, F. Igual, M. Prieto-Matias, and A. Plaza, "Performance-power evaluation of an opencl implementation of the simplex growing algorithm for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 304–308, 2017.
[325] S. Bernabe, S. Lopez, A. Plaza, and R. Sarmiento, "GPU implementation of an automatic target detection and classification algorithm for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 221–225, 2013.
[326] S. Sánchez, A. Paz, G. Martin, and A. Plaza, "Parallel unmixing of remotely sensed hyperspectral images on commodity graphics processing units," *Concurrency and Computation: Practice and Experience*, vol. 23, no. 13, pp. 1538–1557, 2011.
[327] X. Wu, B. Huang, A. Plaza, Y. Li, and C. Wu, "Real-time implementation of the pixel purity index algorithm for endmember identification on GPUs," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 955–959, 2014.
[328] S. Bernabé, P. R. Marpu, and A. Plaza, "Spectral-spatial classification of multispectral images using kernel feature space representation," *IEEE Geosci. Remote Sens. Lett.*, 2013.
[329] J. M. Nascimento, J. M. Bioucas-Dias, J. M. R. Alves, V. Silva, and A. Plaza, "Parallel hyperspectral unmixing on GPUs," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 666–670, 2014.
[330] Z. Wu, Q. Wang, A. Plaza, J. Li, J. Liu, and Z. Wei, "Parallel implementation of sparse representation classifiers for hyperspectral imagery on GPUs," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 2912–2925, 2015.
[331] J. Sevilla, S. Bernabe, and A. Plaza, "Unmixing-based content retrieval system for remotely sensed hyperspectral imagery on GPUs," *J. Supercomput.*, vol. 70, no. 2, pp. 588–599, 2014.
[332] J. Delgado, G. Martín, J. Plaza, L. I. Jiménez, and A. Plaza, "Fast spatial preprocessing for spectral unmixing of hyperspectral data on graphics processing units," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 2, pp. 952–961, 2016.
[333] E. Torti, G. Danese, F. Leporati, and A. Plaza, "A hybrid CPU–GPU real-time hyperspectral unmixing chain," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 2, pp. 945–951, 2016.
[334] L. Jiménez, S. Sánchez, G. Martín, J. Plaza, and A. Plaza, "Parallel implementation of spatial-spectral endmember extraction on graphics processing units," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 10, no. 4, pp. 1247–1255, 2017.
[335] L. Riha, J. L. Moigne, and T. El-Gazawhi, "Optimization of selected remote sensing algorithms for many-core architectures," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 12, pp. 5576–5587, 2016.
[336] C. Lee, S. Gasster, A. Plaza, C.-I. Chang, and B. Huang, "Recent developments in high performance computing for remote sensing: A review," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 4, no. 3, pp. 508–527, 2011.
[337] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, 2017.
[338] E. Aptoula, M. Ozdemir, and B. Yanikoglu, "Deep learning with attribute profiles for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1970–1974, 2016.

GRS