# Hyperspectral Remote Sensing Data Analysis and Future Challenges

**JOSÉ M. BIOUCAS-DIAS,
ANTONIO PLAZA, GUSTAVO CAMPS-VALLS,
PAUL SCHEUNDERS, NASSER M. NASRABADI,
AND JOCELYN CHANUSSOT**

Abstract—Hyperspectral remote sensing technology has advanced significantly in the past two decades. Current sensors onboard airborne and spaceborne platforms cover large areas of the Earth surface with unprecedented spectral, spatial, and temporal resolutions. These characteristics enable a myriad of applications requiring fine identification of materials or estimation of physical parameters. Very often, these applications rely on sophisticated and complex data analysis methods. The sources of difficulties are, namely, the high dimensionality and size of the hyperspectral data, the spectral mixing (linear and nonlinear), and the degradation mechanisms associated to the measurement process such as noise and atmospheric effects. This paper presents a tutorial/overview cross section of some relevant hyperspectral data analysis methods and algorithms, organized in six main topics: data fusion, unmixing, classification, target detection, physical parameter retrieval, and fast computing. In all topics, we describe the state-of-the-art, provide illustrative examples, and point to future challenges and research directions.

© DYNAMIC GRAPHICS

I. INTRODUCTION

Hyperspectral remote sensing is concerned with the extraction of information from objects or scenes lying on the Earth surface, based on their radiance acquired by airborne or spaceborne sensors [1], [2]. Hyperspectral sensing, namely its imaging modality termed *hyperspectral imaging*, has been increasingly used in applications at lab scale (e.g., food safety, pharmaceutical process monitoring and quality control, biomedical, industrial, biometric, and forensic) using small, commercial, high spatial
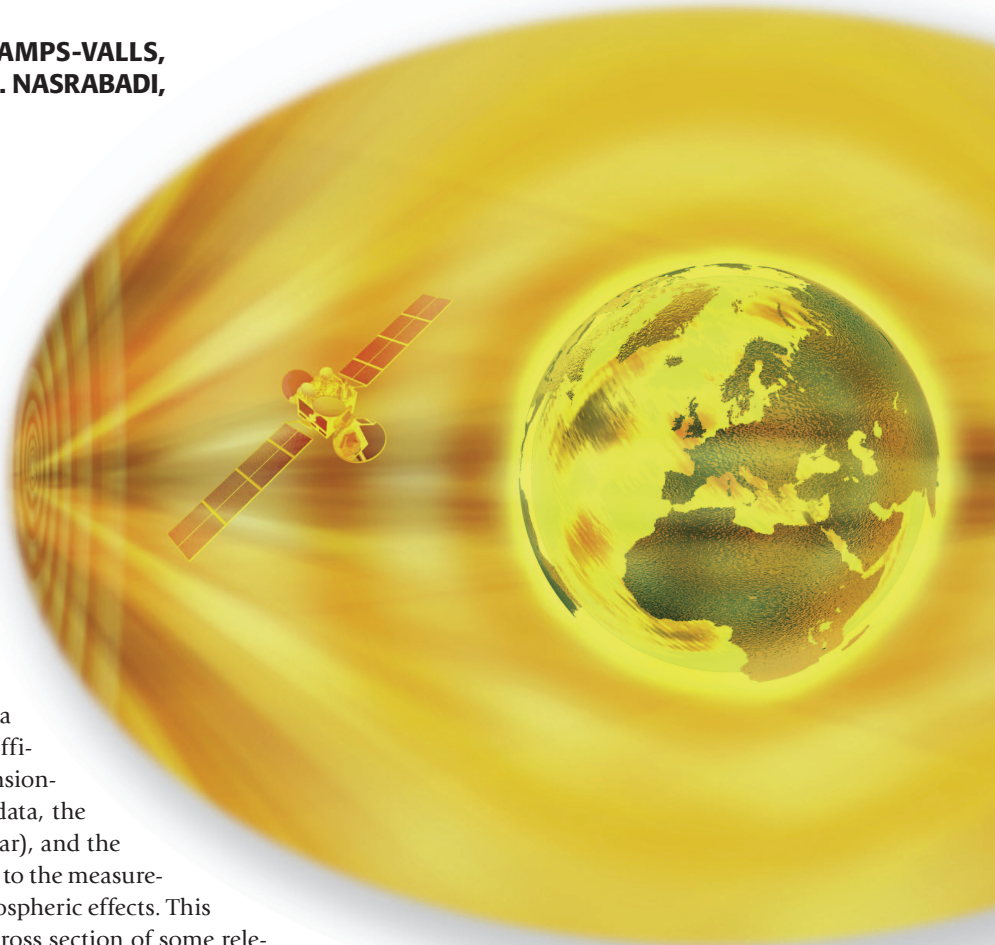
and spectral resolution instruments (see [3] and references therein).

Figure 1 gives a partial account of the relevance of hyperspectral applications, by comparing paper counts per year in the hyperspectral and radar areas. These results were obtained by searching the SCI-Expanded database of the ISI Web-Of-Science with the topics "(hyperspectral) and (remote sensing)," in the left hand side, and "(radar) and (remote sensing)," in the right hand side. We conclude that the number of items per year in 2011 is similar for the hyperspectral and radar areas, with a clear increasing trend in the former and a stabilization or decrease in the latter.

In hyperspectral imaging, also termed imaging spectroscopy [4], the sensor acquires a spectral vector with hundreds or thousands of elements from every pixel in a given scene. The result is the so-called hyperspectral image (HSI). It should be noted that HSIs are spectrally smooth and spatially piece-wise smooth: the values in neighboring locations and wavelengths are highly correlated. This can be observed by extremely non-diagonal covariance matrices and wide autocorrelation functions [1]. This piece-wise smoothness holds as well in the spatio-spectral direction. The characteristics are similar to those of natural photographic images and videos and, therefore, many tools that were developed for these data can be extended for HSI analysis.

An equivalent interpretation of an HSI is given by the acquisition of a stack of images representing the radiance in the respective band (wavelength interval). Due to this interpretation, the HSIs are also termed *hyperspectral data cubes*. These two points of view are illustrated in the top left hand side of Fig. 2, where the HSI has n_b spectral bands and $n_1 \times n_2$ pixels. The plots on the top right hand side show the spectra of pixels containing soil, vegetation, and water. Owing to the high spectral sampling, the spectral information is often highly correlated and thus lives in a low dimensional manifold. This is illustrated at the bottom of Fig. 2, where the spectral vectors of soil, vegetation, and water are represented as $\mathbb{R}^{n_b}$ dimensional points on a surface.

In terms of the geometrical properties of a remote sensing imaging system, the spatial resolution of a sensor is given by its field of view (FOV), and the obtained spectrum is the average of the material's reflectances within this FOV. The spectral resolution is determined by the bandwidth of the spectral bands. When spatially and spectrally sampling the information (we will assume that the sampling is performed at the sensors spatial and spectral resolution), a 3D "hypercube" $X \in \mathbb{R}^{n_1 \times n_2 \times n_b}$ is obtained, containing $n = n_1 \times n_2$ pixels and n_b bands (see Fig. 2). Different forms of representation can be used for HSIs:

- In the spectral representation, each pixel is defined in the spectral space $\mathbf{x} \in \mathbb{R}^{n_b}$. Since neighboring spectra correspond to similar materials, grouping in this spectral space is commonly applied to characterize materials. This can be done by clustering neighboring spectra, or by supervised classification (see section on Classification). Since the spectral correlation is high, the data are likely to reside on a very low-dimensional submanifold of the spectral space, and projection of the data on a subspace of dimension $d \ll n_b$, using, e.g., principal component analysis (PCA) [3], is commonly applied.

- In the spatial representation, each image band is a matrix $\mathbf{X}_i \in \mathbb{R}^{n_1 \times n_2}$. Because of the high spatial correlation, neighboring pixels are likely to belong to a similar material and spatial grouping (e.g., segmentation) is commonly applied.

- In the spatial-spectral representation spectral processing of a pixel is performed taking neighboring pixels into account, while spatial processing of an image band is performed by accounting for the other bands.

These representations have been actively exploited, namely, in dimensionality reduction, feature extraction, unmixing, classification, segmentation, and detection [1], [5]. Still related with the high dimensionality of the spectral information, the most recent trend is sparse and redundant modeling, which is currently reaching the areas of, e.g., restoration, unmixing, classification, segmentation, and detection (see, e.g., [6], [7] and references therein).

Since the output of a hyperspectral sensor provides raw digital number (DN) values and for quantification purposes, a conversion to apparent surface reflectance values is required before using advanced information extraction techniques such as those mentioned above [8]. The characteristics of the sensor itself are described by its transfer

José M. Bioucas-Dias is with the Instituto de Telecomunicações, Instituto Superior Técnico, Lisbon, 1049-1, Portugal (e-mail: bioucas@lx.it.pt). Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica de Cáceres, University of Extremadura, 10003 Cáceres, Spain (e-mail: aplaza@unex.es). Gustavo Camps-Valls is with the Image Processing Laboratory, Universitat de València, C/Catedrático Escardino, E-46980 Paterna (València), Spain (e-mail: gustavo.camps@uv.es). Paul Scheunders is with iMinds, Vision Lab, Department of Physics, University of Antwerp, 2610 Wilrijk, Belgium (e-mail: paul.scheunders@ua.ac.be). Nasser Nasrabadi is with the U.S. Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: nasser.m.nasrabadi.civ@mail.mil). Jocelyn Chanussot is with the GIPSA-Lab, Grenoble Institute of Technology, Grenoble, France (e-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr). This work was supported by the Portuguese Science and Technology Foundation, project PEst-OE/EEI/0008/2013 and by the Spanish Ministry of Economy and Competitiveness (MINECO) under projects TIN2012-38102-C03-01 and AYA2011-29334-C02-02.
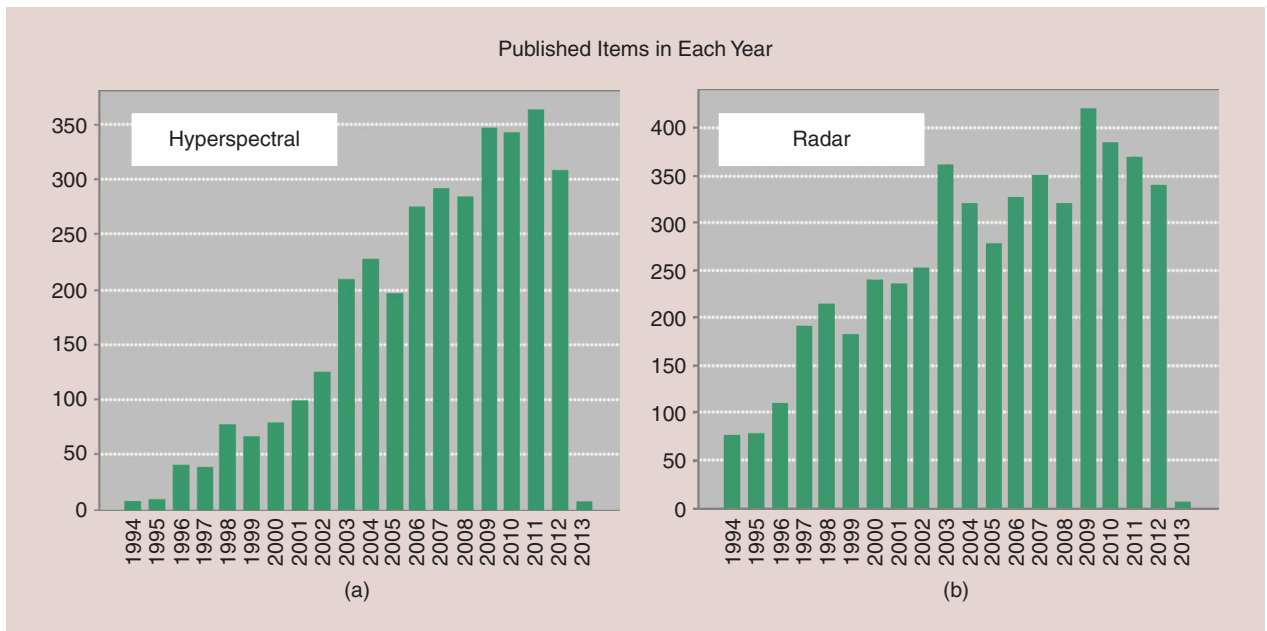
Published Items in Each Year

FIGURE 1. Paper counts per year in hyperspectral and radar topics obtained by searching the SCI-Expanded database of the ISI Web-Of-Science with the following topics: (a) hyperspectral and remote sensing; (b) radar and remote sensing. Search done on January 2013.
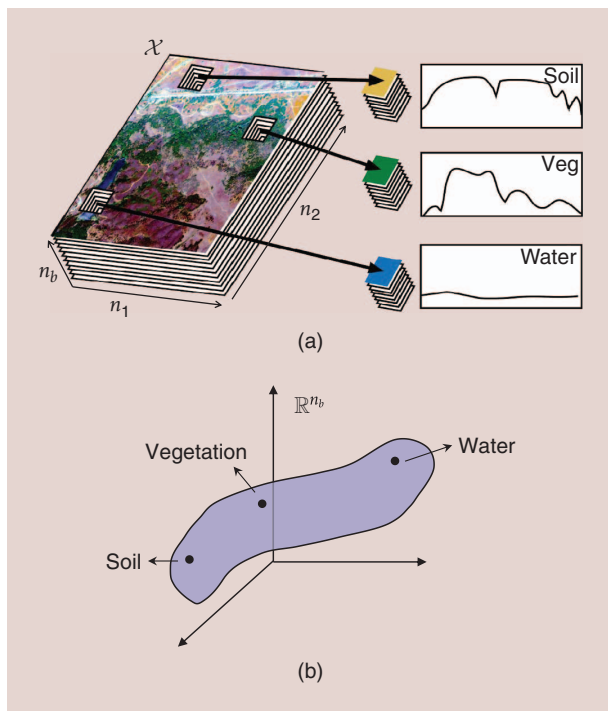

FIGURE 2. (a) Hyperspectral imaging concept. (b) Hyperspectral vectors represented in a low-dimensional manifold.

[9]. Finally, one has to account for the effects of illumination and viewing angle and the surfaces structural and optical properties, to lead to the surface reflectance values.

The interactions of the light with the atmosphere are extremely complex. The radiative transfer theory is often used to derive models for these interactions [10]. Originally developed for simulating TOA radiance for the preparation of future satellite missions, inversion of these models allows for atmospheric correction. The bidirectional reflectance distribution functions (BRDF) describe the reflected light on a surface as a function of the incoming and outgoing light directions. A first approximation for the BRDF, termed 'albedo', is the ratio between the reflectance and the sun's irradiance. BRDF's are usually accompanied by complex surface structure models such as leaf-to-canopy models. A schematic overview of the spectral characteristics of hyperspectral data is given in Fig. 3. We remark that, besides for the conversion to surface reflection, all these models are relevant as well for quantitative analysis (see Sections III and VI).

For illustrative purposes, Table 1 displays spatial and spectral parameters of eight hyperspectral instruments: two airborne (HYDICE[1] and AVIRIS[2]) and six spaceborne (HYPERION[3], EnMAP[4], PRISMA[5], CHRIS[6], HyspIRI[7] and IASI[8]). From this list, EnMAP, PRISMA and HyspIRI are

function. To account for this, first a radiometric calibration of the spectra is generally performed to obtain at-sensor or top-of-atmosphere (TOA) radiance values. As the reflected sunlight passes through the atmosphere, it gets partially absorbed and scattered. Since these effects have a huge influence on the spectral values, they need to be corrected to obtain the ground-leaving radiance or reflectance values

[1]http://rsd-www.nrl.navy.mil/hydice
[2]http://aviris.jpl.nasa.gov
[3]http://eo1.usgs.gov
[4]http://www.enmap.org
[5]http://www.asi.it/en/flash en/observing/prisma
[6]https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/proba
[7]http://hyspiri.jpl.nasa.gov
[8]http://smsc.cnes.fr/IASI

not yet operational. The spatial resolutions are higher for sensors carried by low altitude platforms and vice-versa. The spectral coverage of HYDICE, AVIRIS, HYPERION, EnMAP, PRISMA and HyspIRI corresponds to the visible, the near-infrared, and the shortwave infrared spectral bands, whereas CHRIS covers the visible bands and IASI covers the mid-infrared and the long-infrared bands. The number of bands is approximately 200 for HYDICE, AVIRIS, HYPERION, EnMAP, PRISMA and HyspIRI, with a spectral resolution of the order of 10 nm. The number of bands for CHRIS is 63, with spectral resolutions of 4 and 12 nm (depending on the region of the spectrum) and 8461 for IASI, with a resolution of $0.5\,\mathrm{cm}^{-1}$. In any case, the resolution is very high (offering a huge potential to discriminate materials) in the case of the first seven sensors, and to estimate physical parameters (temperature, moisture and trace gases across the atmospheric column), in the case of the IASI sensor. A summary of the characteristics of several hyperspectral imaging instruments currently in operation, under construction, and missions in a planning stage has been recently provided [11].

Several factors make the analysis of hyperspectral data an often complex and hard task calling for sophisticated methods and algorithms. Among these factors, we refer to spectral mixing (linear and nonlinear), and degradation mechanisms associated to the measurement process (e.g., noise and atmosphere). Another important issue is the extremely high dimensionality and size of the data, resulting from the improved spatial, spectral and temporal resolutions provided by hyperspectral instruments. This demands fast computing solutions that can accelerate the interpretation and efficient exploitation of hyperspectral data sets in various applications [12]. For example, it has been estimated by the NASA's Jet Propulsion Laboratory (JPL) that a volume of 4.5 TBytes of data will be daily produced by HyspIRI (1630 TBytes per year). Similar data volume ratios are expected for EnMAP and PRISMA. Unfortunately, this extraordinary amount of information jeopardizes the use of latest-generation hyperspectral instruments in real-time or near real-time applications, due to the prohibitive delays in the delivery of Earth Observation payload data to ground processing facilities [13]. In this respect, the European Space
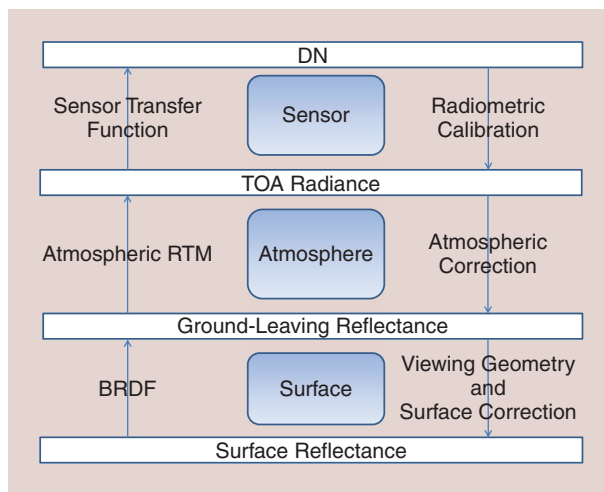

FIGURE 3. Spectral characterization of hyperspectral data.

Agency (ESA) already flagged up in 2011 that "*data rates and data volumes produced by payloads continue to increase, while the available downlink bandwidth to ground stations is comparatively stable*" [14]. In this context, the design of solutions aimed at taking advantage of the ever increasing dimensionality of remotely sensed hyperspectral images for near real-time applications has gained significant relevance and momentum during the last decade [15], [16].

This paper presents a tour over relevant and distinctive hyperspectral data analysis themes, organized in six main topics: data fusion, unmixing, classification, target detection, physical parameter retrieval, and fast computing. Most of the frameworks used in these topics are rooted on signal and image processing, statistical inference, and machine learning fields. In all topics, we describe the state-of-the-art and point to the most likely future challenges and research directions. Illustrative examples with real data are provided for some of the topics covered.

The remainder of the paper is organized as follows. Section II discusses processing techniques aimed at fusing spatial and spectral information from multiple observation and sources. Section III addresses linear and nonlinear hyperspectral mixing and unmixing. Section IV outlines some of the main techniques and challenges

TABLE 1. PARAMETERS OF EIGHT HYPERSPECTRAL INSTRUMENTS.

PARAMETER	HYDICE	AVIRIS	HYPERION	EnMAP	PRISMA	CHRIS	HyspIRI	IASI
Altitude (km)	1.6	20	705	653	614	556	626	817
Spatial resolution (m)	0.75	20	30	30	5–30	36	60	V: 1–2 km H: 25 km
Spectral resolution (nm)	7–14	10	10	6.5–10	10	1.3–12	4–12	0.5 cm^{-1}
Coverage (μm)	0.4–2.5	0.4–2.5	0.4–2.5	0.4–2.5	0.4–2.5	0.4–1.0	0.38–2.5 and 7.5–12	3.62–15.5 (645–2760 cm^{-1})
Number of bands	210	224	220	228	238	63	217	8461
Data cube size (sample × lines × bands)	200 × 320 × 210	512 × 614 × 224	660 × 256 × 220	1000 × 1000 × 228	400 × 880 × 238	748 × 748 × 63	620 × 512 × 210	765 × 120 × 8461

in hyperspectral image classification. Section V addresses hyperspectral target detection. Section VI reviews the main problems and methods in model inversion and estimation of physical parameters, and finally Section VII outlines several strategies to accelerate the hyperspectral image computations using different hardware architectures.

II. DATA FUSION

In this section, we will discuss hyperspectral processing techniques (image in-image out), that fuse spatial and spectral information from one or multiple hyperspectral observations, or a combination of hyperspectral images and other image sources. We will refer to this processing as data fusion. In Fig. 4, a schematic overview of the different strategies is given.

A. RESTORATION

Signal processing techniques can be applied to restore or improve the signal-to-noise ratio (SNR) and/or the spatial resolution. In the case of gray-scale images, many denoising and deconvolution techniques were developed to restore SNR and spatial resolution. It is clear that a band-by-band treatment of the restoration problem in HSIs would not benefit from the high spectral redundancy. The traditional image restoration techniques are extended to account for this spectral redundancy. In this way, hyperspectral image denoising techniques were recently developed by, e.g., employing spatial-spectral information [17] or employing tensor decompositions and multilinear algebra [18]. In [19], restoration of hyperspectral images was proposed based on anisotropic diffusion filtering. Remark that all the above mentioned methods preserve the original spatial and spectral sampling and thus do not improve the spatial resolution.

B. SPECTRAL DATA FUSION

Here, we discuss the fusion of spectral bands of an HSI, in this way removing high spectral redundancy. Since the high number of bands causes dimensionality problems, a dimensionality reduction of the hyperspectral vectors can highly facilitate the analysis afterwards. The goal is to obtain an image of reduced number of bands while trying to preserve the most useful spectral information as possible. The simplest way is to select a few of the available bands, but it is clear that better performance can be obtained when bands are fused together. Traditionally, methods based on PCA are applied that decorrelate bands. In many occasions, the dimensionality reduction is applied for an improved classification afterwards. This topic is treated in Section IV (IV.A.1 and IV.A.2).

A specific application of spectral data fusion is the visualization of HSIs. A user may need to visualize hyperspectral image data for exploration purposes, e.g., for generating ground reference data. However, an HSI contains far more image bands than can be displayed on a standard tristimulus display. By fusion of the spectral bands, an image of limited number of bands can be generated, e.g., a panchromatic image or an RGB image; how to fuse preserving as much information as possible is an issue. In [20], hyperspectral images are linearly projected onto color matching basis functions specifically designed as RGB primaries of a standard tristimulus display. A spatio-spectral approach allows to retain spatial details as well, and often generally generates high-contrast images. Spatio-spectral methods that were developed use e.g. wavelet transforms to fuse multiresolution information of the image bands [21], Markov Random Fields that model the spatial relationship between neighboring pixels [22] or constrained optimization to enforce spatial smoothness [23]. In Fig. 5, four different color visualizations of an AVIRIS hyperspectral image of 224 spectral bands are shown, obtained by PCA and the methods of [20]–[22] respectively.

C. SPATIAL DATA FUSION (MULTI-FRAME SUPERRESOLUTION)

The term (geometric) superresolution (SR) refers to the enhancement of the spatial resolution of imaging sensors by inferring information at the subpixel level. Subpixel image information is for instance available as subpixel shifts of multiple low-resolution observations (multiframe SR). In practice, the images are subsampled by dividing each pixel into $m \times m$ subpixels and interpolating the pixel gray levels. Then, corresponding areas between the multiple observations
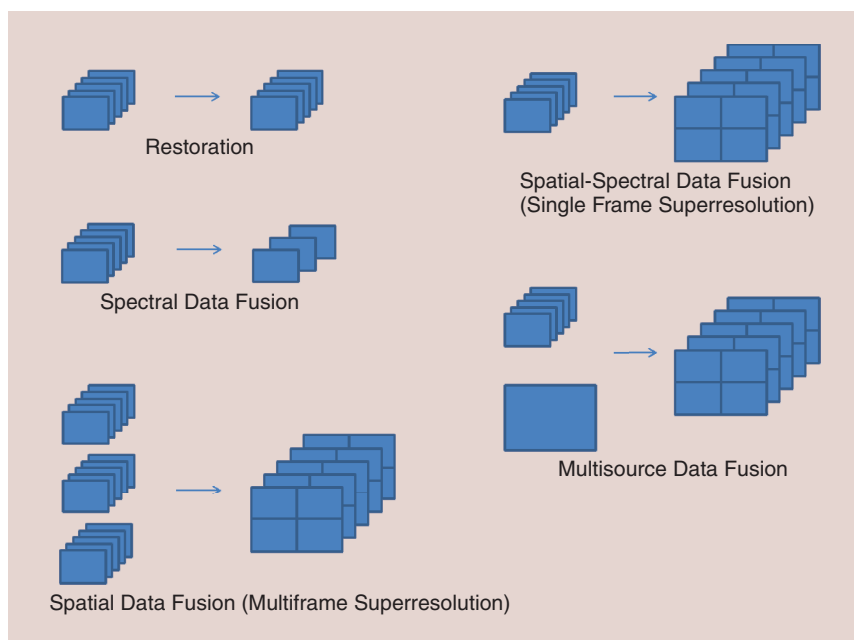

FIGURE 4. A schematic overview of the five different hyperspectral data fusion methodologies.

Restoration

Spectral Data Fusion

Spatial Data Fusion (Multiframe Superresolution)

Spatial-Spectral Data Fusion (Single Frame Superresolution)
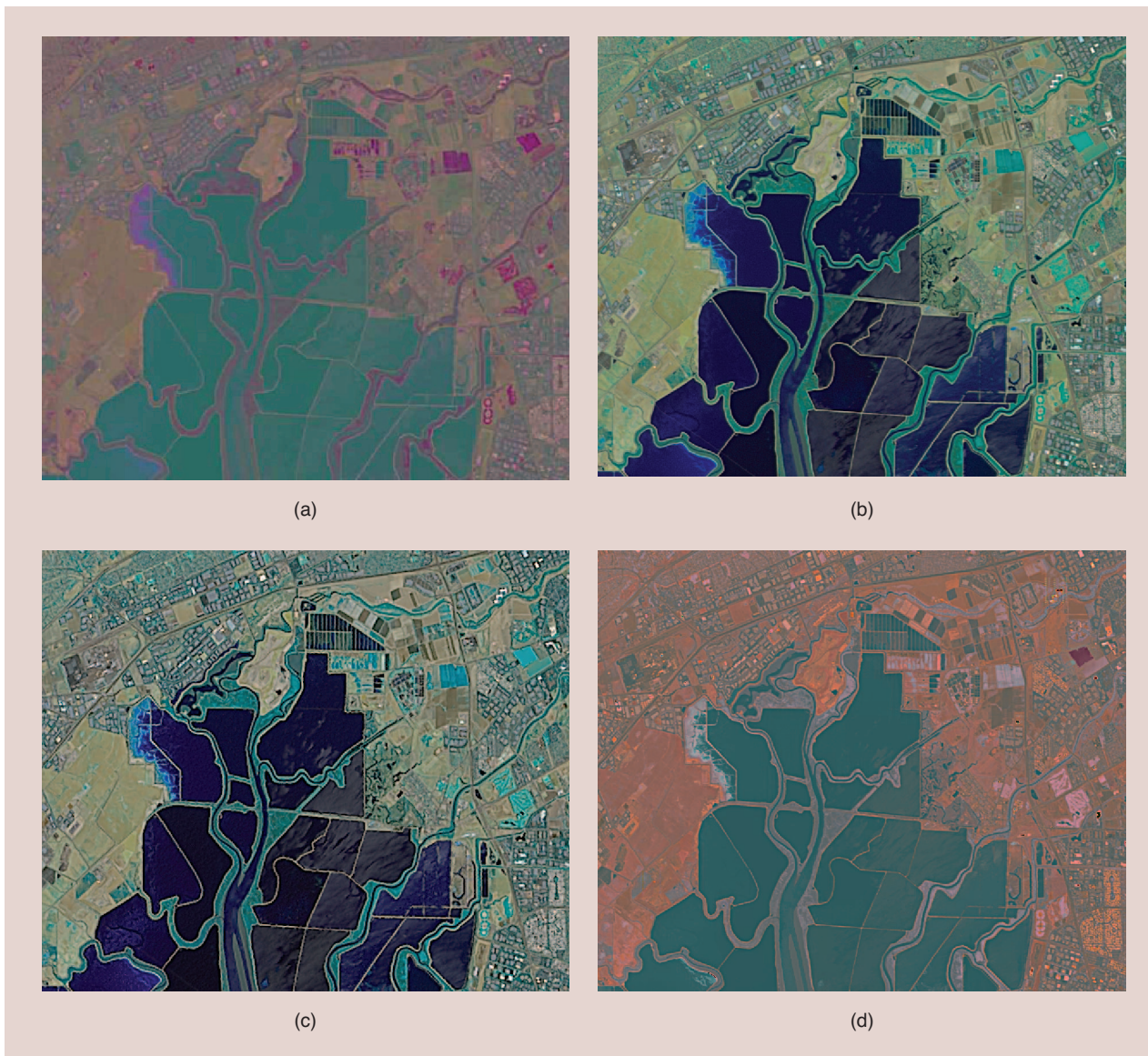
Multisource Data Fusion

FIGURE 5. Visualization of hyperspectral AVIRIS image, from (a) to (d) using PCA and the methods of [20]–[22], respectively.

are detected, geometrically registered and combined to generate one image of high spatial resolution.

When applying SR on hyperspectral images, each image band can be processed separately, but it is clear that a joint processing of all bands is superior [24]. Multiple low-resolution hyperspectral observations from the same scene are obtained for instance by overlapping flight lines, multiangle data [25], or multiframe instances in time [26].

D. SPATIAL-SPECTRAL DATA FUSION SUPERRESOLUTION

Another way of performing superresolution is by fusion of different parts of a single image (single frame SR), which in the case of hyperspectral data amounts to spatial-spectral fusion. In [27], the interband spatial subpixel shifts that are intrinsically present in a hyperspectral data cube are applied for obtaining a SR image.

Since a low-resolution hyperspectral pixel contains a spectral mixture of different materials, superresolution can be accomplished by a spatial localization of the materials fractions at subpixel level. How to obtain the fractions of the present materials will be explained in the section on VI devoted to spectral unmixing. The fractions can be obtained as well by using probabilistic classifiers that assign classification probabilities for each of the materials classes to the pixel. Then the pixel is subsampled. Subpixel mapping or super-resolution mapping refers to techniques that try to spatially organize the fractional spectra of the different materials within a pixel [28]. This mapping can then further be used to simulate a subsampled hyperspectral image [29], [30].

E. MULTISOURCE DATA FUSION

A third way of performing superresolution of an HSI is by the use of other available image sources of high spatial

resolution, acquired by other sensors [e.g. mounted on unmanned aerial vehicles (UAV)]. Several strategies are possible. First of all, many methods originally designed for pansharpening, i.e. fusion of multispectral images with a high spatial resolution panchromatic image [31] are easily transferable to HSIs. A large majority of these methods rely in one way or another on the injection of high spatial information of the panchromatic image into the hyperspectral image bands. Another approach is to assume a joint statistical model between the two image sources and apply Bayesian estimation techniques for enhancing the spatial resolution of the HSI [32], [33]. Alternatively, similar subpixel mapping strategies as in the spatial-spectral fusion can be applied, in which the high-resolution image can deliver the required materials spectral information [34], [35]. Alternatively, the local correlation with the high spatial resolution image can be employed [36].

F. CHALLENGES

All described methods for enhancement of the spatial resolution will generate images at a higher spatial sampling that show higher contrast and finer details, but this does not necessarily guarantee an improvement of the actual spatial resolution [37]. In particular for HSIs, the described methods will be very useful for exploratory analysis and visualization purposes. However, a quantitative analysis requires a high reliability of the obtained spectra. Further research needs to be conducted on validation methodologies of these fusion methods [38].

Moreover, with technological progress, the spatial resolution of sensors improves largely. Also the employment of UAV's leads to very high spatial resolution data. While most spatial resolution enhancement methods aim at a resolution improvement of a factor of 2-3, in practice, resolution differences on the order of a factor of 10 need to be bridged.

Finally, some but certainly not all of the methods mentioned explicitly make use of transfer function information of the employed sensors. When available, it is expected that this information improves the performance of data fusion methods.

III. HYPERSPECTRAL UNMIXING

The signal recorded by a hyperspectral sensor at a given band and from a given pixel, letting alone the effects of the atmosphere, is a mixture of the "light" scattered by the constituent substances located in the respective pixel coverage. Fig. 6 illustrates three types of mixtures owing to low spatial resolution of the sensor (a), presence of intimate mixtures (b), and multiple light scattering in a two-layer media (c). As a result, when mixing occurs, it is not anymore possible to determine what materials are present in the pixels directly from the respective measured spectral vectors. This is to say that the key feature of the spectral sensors, which is its ability to discriminate materials based on the their spectral responses, is compromised. This section addresses spectral mixing modeling, provide insights on the spectral unmixing inverse problems, and point to algorithms to solve them.

With the objective of recovering the ability to discriminate materials, an impressive amount of research work has been devoted to hyperspectral unmixing (HU) (see, e.g., [3], [39], and references therein). HU is, however, a hard inverse problem. The difficulties begin with its formulation. Put in simple terms, given a measured spectral vector $\mathbf{y} \in \mathbb{R}^{n_b}$, HU aims at explaining $\mathbf{y}$ in terms of the spectral properties of the materials present the respective pixel an of its distribution. An useful treatment of this problem cannot be given without a formal model, $\mathbf{y} = \mathbf{f}(\theta)$, where $\mathbf{f}(\cdot)$ is the so-called forward operator, linking the measurements $\mathbf{y}$ to the scene parameters θ. In conclusion, the precise meaning of HU depends on the meaning of parameter vector θ.
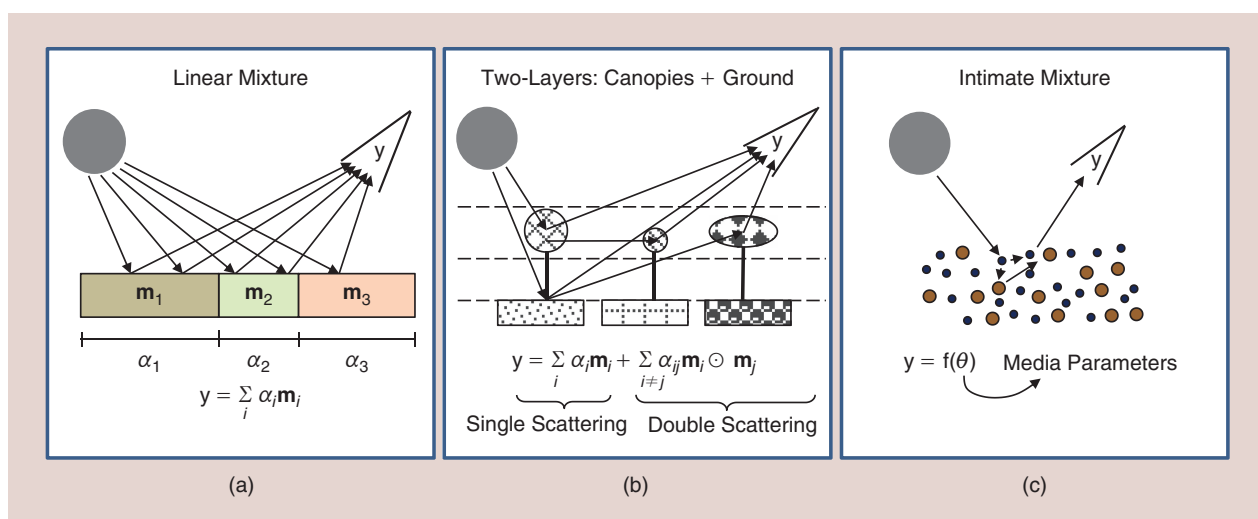

FIGURE 6. Schematic view of three types of spectral mixing. (a) Linear mixing in a checkerboard type surface. (b) Nonlinear (linear plus bilinear) mixing in a two-layer media. (c) Nonlinear mixing in an intimate (particulate) media.

RTT is a mathematical model for the transfer of energy as photons interacts with the materials in the scene, and thus to the derive forward operators necessary to solve HU problems. The core of the RTT is a differential equation describing radiance read by the sensor. It can be derived via the conservation of energy and the knowledge of the phase function, which represents the probability of light with a given propagation direction be scattered into a specified angle solid around a given scattering direction.

In general, the forward operator $\mathbf{f}(\boldsymbol{\theta})$ is not invertible, unless we have partial knowledge of vector $\boldsymbol{\theta}$, which usually depends on scene parameters often very hard to obtain. Three notable exceptions to these scenario, schematized in Fig. 6, are the linear model, the bilinear model, and the Hapke model [3], [40]. These are three approximations for the analytical solution to the RTT suitable to unsupervised applications, i.e., when no prior knowledge exits about the materials and its distributions.

The linear mixing holds true when the mixing scale is macroscopic and the incident light interacts with just one material, as is the case in checkerboard type scenes [40] schematized in Fig. 6. The light from the materials, although almost completely separated, is linearly mixed within the measuring instrument, owing to insufficient spatial resolution. Formally, the measured spectral vector $\mathbf{y} := [\gamma_1, \ldots, \gamma_{n_b}]^T$, holding the radiances at bands $i = 1, \ldots, n_b$, is expressed as

$$\mathbf{y} = \sum_{i=1}^{p} \alpha_i \mathbf{m}_i, \qquad (1)$$

where $\mathbf{m}_i \in \mathbb{R}^{n_b}$, for $i = 1, \ldots, p$, is the spectral signature of the ith material, termed *endmember*, and α_i is the percentage that the ith material occupies inside the pixel, termed *fractional abundance* or *simply* abundance. Inspired in the linear mixing model (LMM), the HU problem is very often defined as the unsupervised estimation of the endmembers and of the respective fractional abundances.

The LMM has been widely used in the past decade to address HU problems. The reason is threefold: a) despite its simplicity, LMM is an acceptable approximation for the light scattering in many real scenarios; b) under suitable conditions of the data set, LMM yields well-posed inverse problems; c) under the LMM, HU is interpretable as a blind source separation (BSS) problem or a nonnegative matrix factorization problem, which have been vastly researched in many signal processing areas. In section III-A, we address in more details relevant aspects of HU under the LMM.

In spite of the LMM attractiveness, researchers are beginning to expand more aggressively into the nonlinear mixing field to cope with the LMM limitations. In section III-F, we address in more detail relevant aspects of the nonlinear HU.

Unmixing via sparse regression (SR) is still another direction recently introduced to circumvent part of the limitations of the blind linear HU. This line of attack formulates HU as a semiblind approach in which the endmember identification is replaced with a SR over a library of spectral signatures, usually overcomplete, obtained in laboratory. Details of this approach are given in section III-G.

A. LINEAR UNMIXING

Under the LMM (1), a given measured hyperspectral vector can be written as $\mathbf{y} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{w}$, where $\mathbf{M} := [\mathbf{m}_1, \ldots, \mathbf{m}_p]$ stands for the mixing matrix, $\boldsymbol{\alpha} := [\alpha_1, \ldots, \alpha_p]^T$ stands for the fractional abundance vector, and $\mathbf{w}$ accounts for additive perturbations due to, for example, model mismatches and additive noise. Because the components of $\boldsymbol{\alpha}$ represent fractions, then they satisfy the constraints $\alpha_i \geq 0$, for $i = 1, \ldots, p$ and $\sum_{i=1}^{p} \alpha_i = 1$, termed *abundance nonnegativity constraint* (ANC) and *abundance sum constraint* (ASC), respectively. Owing to signature variability, the ASC is seldom observed in real applications. Nevertheless, because the spectral vectors are non-negative, is it always possible to build rescaled versions thereof, belonging to an affine set [41], and thus satisfying the ASC (see [3] for details). We assume, therefore, that the ASC holds true.

> **SPECTRAL UNMIXING IS THE INVERSE PROBLEM OF DETERMINING THE SPECTRALLY PURE COMPONENTS (ENDMEMBERS) PRESENT IN MIXED PIXELS, AS WELL AS THEIR ABUNDANCE FRACTIONS.**

Before unmixing, the hyperspectral data set usually undergoes atmospheric calibration and dimension reduction. The atmospheric calibration step converts the measured radiance into reflectance, which is an intrinsic characteristic of the materials. However, the unmixing inverse problem can also be formulated in the radiance data, provided that the effects of atmosphere are pixel invariant. The dimension reduction step (see IV-A for additional details) identifies the subspace where the spectral vectors live and projects them onto this subspace. Given that the identified subspace is generally of much lower dimension than that of the spectral vectors, this projection yields considerable gains in algorithm performance and complexity, data storage, and noise reduction.

Suppose we are given a hyperspectral data set containing n spectral vectors of size n_b arranged in the matrix $\mathbf{Y} := [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \mathbb{R}^{n_b \times n}$. Defining the abundance fraction matrix $\mathbf{A} := [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n]$, where $\boldsymbol{\alpha}_i$ represents the fractional abundance vector of the ith pixel, then the linear HU inverse problem can be stated as

$$\min_{\mathbf{M}, \mathbf{A}} \| \mathbf{Y} - \mathbf{M}\mathbf{A} \|_F$$
$$\text{subject to: } \mathbf{A} \geq \mathbf{0}, \ \mathbf{1}_p^T \mathbf{A} = \mathbf{1}_n, \qquad (2)$$

where $\| \mathbf{X} \|_F := \sqrt{\text{trace}\{\mathbf{X}\mathbf{X}^T\}}$, $\mathbf{A} \geq \mathbf{0}$ is to be understood in the componentwise sense and $\mathbf{1}_p$ and $\mathbf{1}_p$ are column vectors with p and n ones, respectively. Note the inequality $\mathbf{A} \geq \mathbf{0}$, is the ANC and the equality $\mathbf{1}_p^T \mathbf{A} = \mathbf{1}_n$ is the ASC.
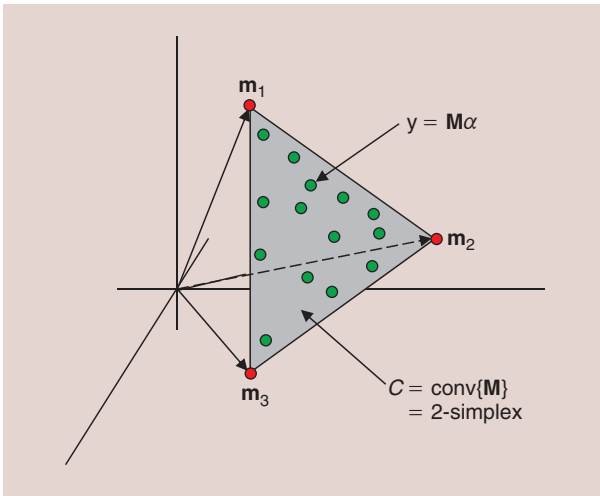
FIGURE 7. Illustration of the simplex set C for $p = 3$ (C is the convex hull of the columns of **M**, $C = \text{conv}\{\mathbf{M}\}$). Green circles represent spectral vectors. Red circles represent vertices of the simplex and correspond to the endmembers.

The optimization (2) is interpretable both as a linear blind source separation problem and as a matrix factorization problem. In the former case the independent component analysis (ICA) come to mind to separate sources (i.e., the fractional abundances). ICA have in fact been considered to solve spectral unmixing problems. Unfortunately, ICA is based on the assumption of mutually independent sources, which is not the case of hyperspectral data, since the sum of abundance fractions is constant, implying statistical dependence among them. This dependence compromises ICA applicability to hyperspectral data as shown in [42].

1) The convex geometry of linear unmixing: In order to shed light into the linear HU problem, we now give an interpretation of problem (2) based on convex geometry. The set

$$C := \left\{ \mathbf{y} = \mathbf{M}\boldsymbol{\alpha} : \sum_{j=1}^{p} \alpha_j = 1, \; \alpha_j \geq 0, \; j = 1, \ldots, p \right\}$$

i.e., the convex hull of the columns of **M**, is a $(p - 1)$-simplex in $\mathbb{R}^{n_b}$. Fig. 7 illustrates a 2-simplex C for a hypothetical mixing matrix **M** containing three endmembers. The points in green denote non-pure spectral vectors, whereas the points in red are pure spectral vectors, thus corresponding to the vertices of the simplex. Note that the inference of the mixing matrix **M** amounts to identify the vertices of the simplex C. This geometrical point of view has been exploited by many unmixing algorithms, which can be mainly classified either as pure pixel or non-pure pixel based.

B. PURE PIXEL BASED ALGORITHMS

In the pure pixel based algorithms it is assumed the presence in the data of at least one pure pixel per endmember, meaning that there is at least one spectral vector on each vertex of the data simplex. This geometric picture is illustrated in the left hand side of Fig. 8. The pure pixel assumption, though enabling the design of very efficient algorithms from the computational point of view, is a strong requisite that may not hold in many datasets. These class of algorithms have been the most often used in linear HU applications, perhaps because of their light computational burden and clear conceptual meaning.

Most of the pure pixel based algorithms exploit one of the following properties of the endmember signatures: a) the extremes of the projection of the spectral vectors onto any subspace correspond to endmembers; b) the volume defined by any set of p spectral vectors is maximum when those are endmembers. Representative algorithms of class a) are *pixel purity index* (PPI) [43], *vertex component analysis* (VCA) [44], simplex growing algorithm (SGA) [45] *successive volume maximization* (SVMAX) [46], and the *recursive algorithm for separable NMF* (RSSNMF) [47]; Representative algorithms of class b) are N-FINDR [48], *iterative error analysis* (IEA), [49], *sequential maximum angle convex cone* (SMACC), and *alternating volume maximization* (AVMAX) [46].

C. NON-PURE PIXEL BASED ALGORITHMS

Fig. 8, middle and right hand side, schematizes two data sets without pure pixels; the data set in the middle does not contain pure pixels but contains at least $p - 1$ spectral vectors on each facet. In this data set, the endmembers may be inferred by fitting a minimum volume (MV) simplex to the data; this rather simple and yet powerful idea, introduced by Craig in his seminal work [41], underlies several geometrical based unmixing algorithms.
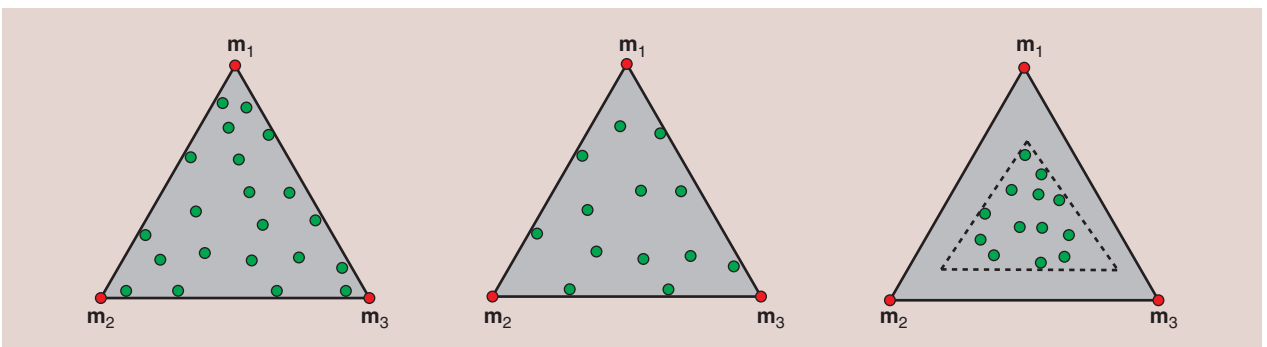

FIGURE 8. Illustration of the concept of simplex of minimum volume containing the data for three data sets.

From an optimization point of view, the MV based unmixing algorithms are formulated as

$$\min_{\mathbf{M},\mathbf{A}} \| \mathbf{Y} - \mathbf{MA} \|_F^2 + \lambda V(\mathbf{M})$$
$$\text{subject to: } \mathbf{A} \geq \mathbf{0}, \ \mathbf{1}_p^T \mathbf{A} = \mathbf{1}_n, \tag{3}$$

where $V(\mathbf{M})$ is a volume regularizer, promoting mixing matrices of "minimum volume" and $\lambda > 0$ is a regularization parameter setting the relative weight between the data term and the volume term. Most of the methods adopting the above formulation implement an a nonlinear block Gauss-Seidel iterative scheme minimizing successively with respect to $\mathbf{M}$ and to $\mathbf{A}$. This is the case of *iterative constrained endmembers* (ICE) algorithm [50] and of the *minimum volume transform-nonnegative matrix factorization* (MVC-NMF) [51], whose main differences are related with the way they define the regularizer $V(\mathbf{M})$. For variations of these ideas recently introduced, see [3]. The *sparsity-promoting ICE* (SPICE) [52] is an extension of the ICE algorithm that incorporates sparsity-promoting priors aiming at finding the number of endmembers.

Problem (3) is non-convex. Thus the solutions provided by greedy solvers are strongly dependent on the initialization. This handicap was circumvented in the *simplex identification via variable splitting and augmented Lagrangian* (SISAL) [53], the *minimum volume enclosing simplex* (MVES) [54] by reformulating (3) with respect to $\mathbf{M}^{-1}$ instead of $\mathbf{M}$.

D. STATISTICAL ALGORITHMS

The MV simplex shown in the right hand side example of Fig. 8 is smaller than the true one. This situation corresponds to a highly mixed data set where there are no spectral vectors near the facets. For these classes of problems, the MV algorithms fail and we usually resort to the statistical framework, formulating HU as a statistical inference problem, usually adopting the Bayesian paradigm.

The Bayesian approaches often have the following flavor (see, e.g., [55] [3] and references therein): The posterior distribution of the parameters of interest is computed from the linear observation model (1) within a hierarchical Bayesian model, where conjugate prior distributions are chosen for some unknown parameters to account for physical constraints. The hyperparameters involved in the definition of the parameter priors are then assigned non-informative priors. Due to the complexity in obtaining close-form expression for the posterior density, the parameters of interest, namely the mixing matrix and the fractional abundances, are, often, estimated from samples of the posterior density generated with Markov chain Monte Carlo (MCMC) techniques.

A clear illustration of the potential of the Bayesian approach to cope with highly mixed data sets is provided by the DECA [56] algorithm; it models the abundance fractions as mixtures of Dirichlet densities. A cyclic minimization algorithm is developed where: 1) the number of Dirichlet modes is inferred based on the minimum description length (MDL) principle; 2) a generalized expectation maximization (GEM) algorithm is derived to infer the model parameters.

Finally, we note that most of the matrix factorization methods referred to in sections III-B and III-C may be also be formulated as Bayesian inference problems, with the advantage of attaching meaning to the model parameters and providing a suitable framework to deal with them.

E. UNMIXING EXAMPLE

In this section, we illustrate part of the concepts presented before by unmixing the publicly available TERRAIN HSI[9] acquired by the HYDICE sensor [57] (see HYDICE parameters in Table 1). The low SNR bands due to water absorption were removed yielding a data set with 166 bands.

The TERRAIN HSI, shown in the top left column of Fig. 9, was calibrated to reflectance, has size 500×307, and is mainly composed of soil, trees, grass, a lake, and shadows, disposed on a flat surface. The signal subspace was identified with the HySime [58] algorithm and the original data projected onto this subspace. The identified subspace dimension was 20. We have, however, discarded those orthogonal directions corresponding to SNR $<$ 10 to avoid instability of the endmember identification (see [3] for more details). After this procedure, we ended up with a subspace of dimension 6.

The plots on the top right column of Fig. 9 show the identified endmember signatures with the VCA algorithm [44]. The corresponding pixels are referenced in the original image. They represent three types of soil, trees, grass, and a spectrum obtained in the lake, which we termed shade due to its low amplitude. The figure in the middle of the left column shows a scatterogram of the data set projected on the subspace defined by the first two subspace eigen directions determined by HySime. The endmembers identified by VCA and N-FINDR area also represented. The solution provided by the two algorithms are identical and, due to the high spatial resolution of the sensor, correspond to nearly pure pixels. Notice that there are endmembers placed in all the "extremes" of the scatterogram, which is coherent with the pure pixel hypothesis. The remaining parts of Fig. 9 shows the estimated abundance fractions for soil 1, trees, and grass.

F. NONLINEAR UNMIXING

A complete physics based approach to nonlinear HU would involve the inversion of the RTT, which is an extremely complex ill-posed problem, relying on scene parameters
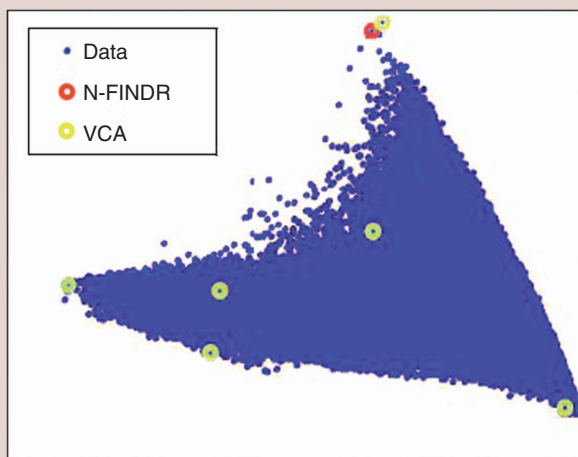
> **NONLINEAR UNMIXING IS A MORE COMPLEX PROBLEM THAN LINEAR UNMIXING. FULLY UNSUPERVISED NONLINEAR UNMIXING METHODS HAVE ONLY VERY RECENTLY BEEN EXPLORED.**

---

[9]Data set available at http://www.agc.army.mil/Missions/Hypercube.aspx.
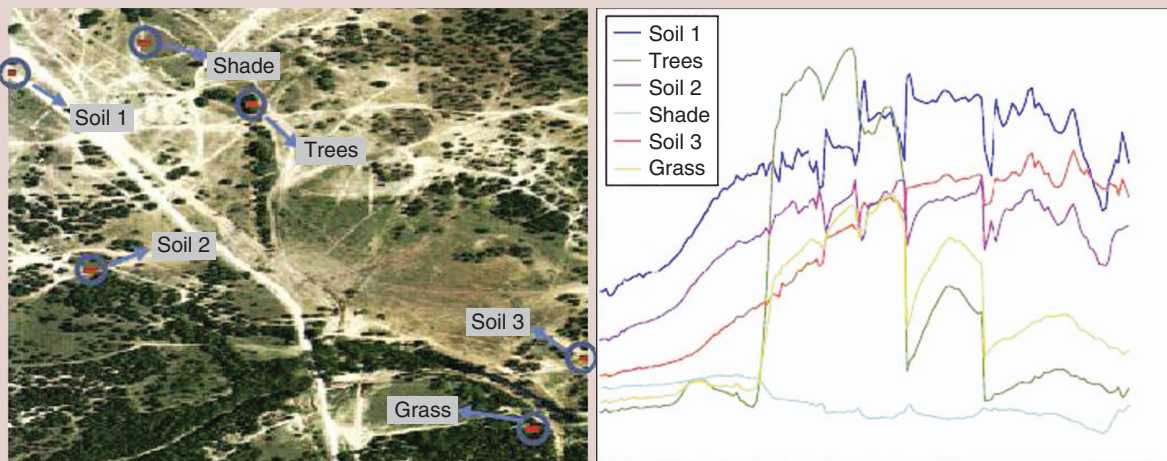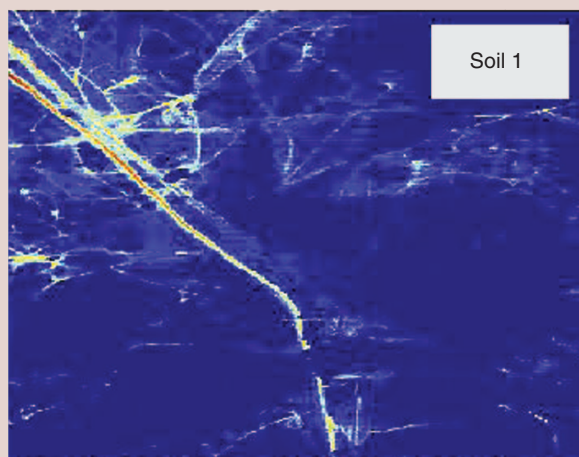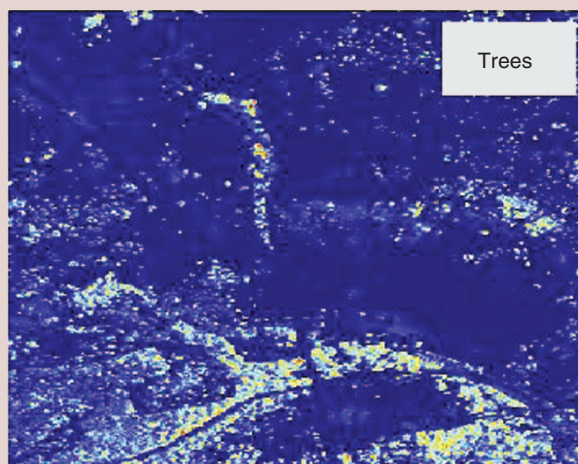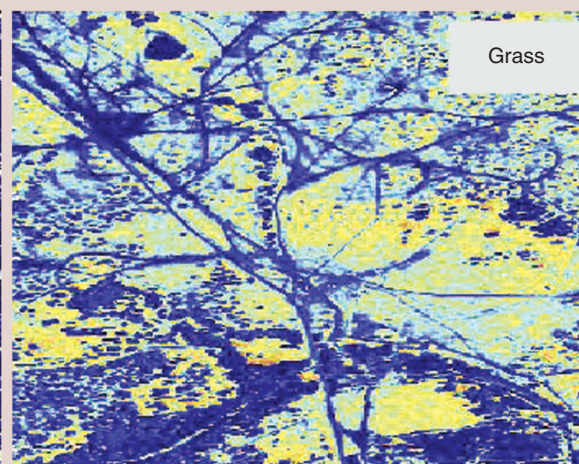
FIGURE 9. Unmixing results: (a) TERRAIN HSI; (b) identified endmembers; (c) data projection onto the subspace defined by the first two eigen directions; (d) soil abundance map; (e) tree abundance map; and (f) grass abundance map.

that are very hard or impossible to obtain. For this reason, the research on nonlinear HU is far more immature compared to linear HU. To avoid the complex physical models, usually simpler strategies are applied using data-driven but physics-inspired models, such as the bilinear and Hapke's models.

The bilinear model is valid when the scene can be partitioned in successive layers with similar scattering properties. Fig. 6, in the middle, schematizes a two-layer scene and shows the expression for the measured light. The sum on the left hand side accounts for the single scattering and is similar to the LMM; the sum on the right hand side accounts for the double scattering, where the vectors $\mathbf{m}_i \odot \mathbf{m}_j \in \mathbb{R}^{n_b}$ (symbol $\odot$ stands for elementwise multiplication) account for pairwise interactions.

Fig. 6, right hand side, illustrates an intimate mixture, meaning that the materials are in close proximity and the mixture occurs at a microscopic level. The Hapke approximation [40] for intimate mixtures models the reflectance as a nonlinear function of a convex combination of the individual endmember albedos. The coefficients of the linear combination are the relative geometric cross-sections of the components. When the endmember particle sizes and densities are similar, the coefficients are good approximations for the mass fractions of the different endmembers. However, in general, one needs information concerning the particle sizes of the components to relate the mass fractions and the relative geometric cross-sections [59].

Several strategies have successfully applied the bilinear model to treat the double scattering problem, such as Bayesian algorithms, where prior models are chosen to satisfy the positivity and sum-to-one constraints [60]. On the other hand, kernel-based methods can design flexible kernels to handle the problem of intimate mixtures. Linear kernels, radial-basis functions, polynomial, and physics-based kernels were proposed [61]. To cope with both scattering and intimate mixture problems simultaneously, machine learning technologies have been proposed, where training samples were used to train artificial neural networks for nonlinearities (see [62] and references therein). Polynomial functions can be applied as well to model the nonlinearities [63].

A disadvantage of the above methods is that they require the knowledge of the endmember signatures. Fully unsupervised nonlinear unmixing methods have only very recently been explored. One possibility is to work directly on the nonlinear data manifold on which it can be shown that the concepts of convex geometry still hold. The geometry of the data manifold is described by graph-based methods, and geodesic distances between spectra are approximated by shortest-path distances on this graph. If then endmember extraction and unmixing methods can be completely rewritten in terms of distance geometry, a complete nonlinear unmixing version is obtained [5]. Although this method is completely data-driven, geodesic distances can be calculated as well on manifolds induced by nonlinear models, such as the bilinear model [64].

Other methods utilize the LMM and Hapke's approximation to model macroscopic and intimate mixtures, respectively. The mixtures are estimated directly from the data without the need for a priori knowledge of the mixture types. In addition, the explicit modeling of both mixture types allows for direct estimation of the endmembers [65].

G. UNMIXING VIA SPARSE REGRESSION

HU via SR has recently been introduced with the objective of coping with data sets not fulfilling the geometrical or statistical assumptions on which the HU methodologies presented in the previous sections rely. In the SR formulation, it is assumed that the measured spectral vectors can be expressed as linear combinations of a small number of pure spectral signatures known in advance [6] (e.g., spectra collected on the ground by a field spectro-radiometer). Unmixing then amounts to finding the optimal subset of signatures in a (potentially very large) spectral library (*dictionary* in the SR jargon) that can best model each mixed pixel in the scene. In practice, this is a combinatorial problem, which calls for efficient linear SR techniques based on sparsity-inducing regularizers. Linear SR is an area of very active research with strong links to compressed sensing [66].

> **SPARSE REGRESSION TECHNIQUES ALLOW TACKLING THE UNMIXING PROBLEM USING (POTENTIALLY VERY LARGE) GROUND SPECTRAL LIBRARIES.**

Let us assume that we are given a spectral library $\mathbf{A} \in \mathbb{R}^{n_b \times m}$ containing m spectral samples. Usually, we have $m > n_b$ and, therefore, the linear problem in hand is underdetermined. Let $\mathbf{x} \in \mathbb{R}^n$ denote the fractional abundance vector with respect to the library $\mathbf{A}$. With these definitions in place, we can now write our SR problem as

$$\min_{\mathbf{x}} \| \mathbf{x} \|_0 \text{ subject to } \| \mathbf{Ax} - \mathbf{y} \|_2 \le \delta, \ \mathbf{x} \ge \mathbf{0}, \qquad (4)$$

where $\| \mathbf{x} \|_0$, termed the ℓ_0 norm, denotes the number of nonzero components of $\mathbf{x}$, and $\delta \ge 0$ is the error tolerance due to noise and modeling errors. Problem (4) is NP-hard [67] and therefore there is no hope in solving it in a straightforward way. Greedy algorithms such as the orthogonal matching pursuit (OMP) [68] and convex relaxations replacing the ℓ_0 norm with the ℓ_1 norm are alternative approaches to compute the sparsest solution [69].

Contrary to problem (4), there are efficient solvers to solve the convex approximations of it conceived to HU applications [70]. What is, perhaps, totally unexpected is that sparse fractional abundances vectors can be exactly reconstructed by the convex relaxations, provided that the columns of matrix $\mathbf{A}$ are incoherent in a given sense [69]. The applicability of sparse regression to HU was studied in detail in [6]. Two main conclusions were drawn:

a) spectral signatures tend to be highly correlated what imposes limits to the quality of the results provided by solving the convex relaxations of (4).
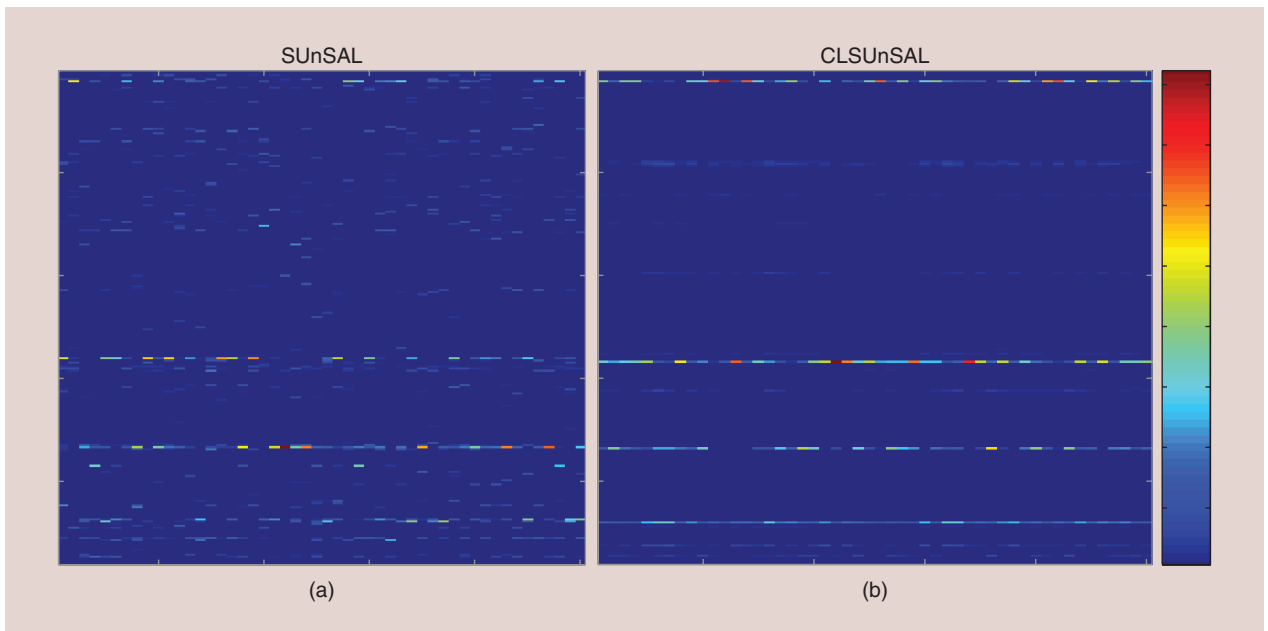
FIGURE 10. Sparse regression solutions **X** for 50 simulated spectral LMM generated with 4 endmembers and SNR = 30 dB and using a subset of the USGS library with 250 signatures. The horizontal axis represents pixels and the vertical axis represents endmembers. (a) Solution computed by the SUnSAL algorithm treating independently each pixel. (b) Solution computed by CLSUnSAL algorithm, enforcing collaborative sparsity.

b) the limitation imposed by the high correlation of the spectral signatures is mitigated by the high level of sparsity often observed in the hyperspectral mixtures.

The current research efforts to cope with the high correlation of the spectral signature are aligned with the recent advances in the area of structured sparsity [71]. This research direction exploits prior information about patterns of sparsity known to exist in specific applications. The types of structured sparsity exploited in HU are directly linked with two characteristics of hyperspectral data: i) the fractional abundance maps are piecewise smooth; and ii) the fractional abundance vectors for the different pixels in the HSI share the same support (i.e., the set of non-zero elements).

The structured sparsity referred to in i) is linked to the fact that, in piece-wise smooth maps, it is very likely that neighboring pixels have very close values. These ideas are exploited in [72] by including the total variation (TV) regularization term in the objective function (4), which promotes piecewise-smooth abundance maps.

The structured sparsity referred to in ii), termed *collaborative sparsity*, is promoted, for example, by the $\ell_{2,1}$ mixed norm $\|\mathbf{X}\|_{2,1} := \sum_{i=1}^{n} \|\mathbf{x}^i\|_2$, where $\mathbf{X} := [\mathbf{x}_1, ..., \mathbf{x}_n]$ holds the fractional abundance vectors in its columns and $\mathbf{x}^i$ is the *i*th row of $\mathbf{X}$ [73]. The work [74] introduces the collaborative sparsity approach in HU. Fig. 10 shows in false color the inference of $\mathbf{X} \in \mathbb{R}^{250 \times 50}$ using a subset of the USGS library splib06[10] with 250 signatures of size 220. The simulated ground-truth abundances contain

50 pixels and 4 endmembers randomly extracted from the library. The simulated measurement were contaminated with additive noise and SNR = 30 dB. The image on the left hand side corresponds to the ℓ_1 relaxation of (4) computed with the SUnSAL algorithm [70], thus treating each pixel independently. The image on the right hand side is the solution of the group sparse problem using the mixed $\ell_{2,1}$ norm and was computed with the CLSUnSAL algorithm introduced in [74]. The collaborative regularization yields a cleaner solution with many rows set to zero.

H. CHALLENGES

As a result of intense work in the last ten years, the research boundary in HU has advanced considerably. Many instances of HU are, however, hard inverse problems far from being solved in reasonable terms. The need for reliable unmixing results will continue to foster active research in HU, namely in areas of mixing models, accounting for the measurement process, and data representation or prior knowledge. In the area of mixing models, researchers are starting to derive and use nonlinear forward models, usually based on the RTT, which are application specific. In the area of data representation or prior knowledge, the objective is to approximate the low dimensional data manifolds using sparse and redundant representation models, possibly learned from the data. Array signal processing is yet another research direction merging ideas coming from sparse representation and multiple measurements, with a large potential for example in pruning the size of the hyperspectral libraries.

[10]http://speclab.cr.usgs.gov/spectral.lib06

IV. CLASSIFICATION

Hyperspectral image classification has been a very active area of research in recent years [75]. Given a set of observations (i.e., pixel vectors in a hyperspectral image), the goal of classification is to assign a unique label to each pixel vector so that it is well-defined by a given class.

The availability of hyperspectral data with high spatial resolution has been quite important for classification techniques, as their main assumption is that the spatial resolution of the data is high enough to assume that the data mostly contains pure pixels (i.e., pixels represented by a single predominant spectral signature). In the opposite scenario (i.e., the data mostly contains mixed pixels) it is preferable to use spectral unmixing techniques to perform the analysis.

In this section, we outline some of the main techniques and challenges in hyperspectral image classification. We focus mainly on supervised and semi-supervised classification, although techniques for unsupervised classification and/or clustering have also been used in the literature [3]. For instance, a relevant unsupervised method successfully applied to hyperspectral image data is Tilton's recursive hierarchical segmentation (RHSEG) algorithm[11]. Supervised classification has been more widely used [76], but it also faces challenges related with the high dimensionality of the data and the limited availability of training samples [75].

In order to address these issues, feature mining [3], subspace-based approaches [58] and semi-supervised learning techniques [1] have been developed. In feature mining and subspace approaches, the goal is to reduce the dimensionality of the input space in order to better exploit the (limited) training samples available. In semi-supervised learning, the idea is to exploit the information conveyed by additional (unlabeled) samples, which can complement the available labeled samples with a certain degree of confidence. In all cases, there is a clear need to integrate the spatial and spectral information to take advantage of the complementarities that both sources of information can provide [76]. An overview of these different aspects, which are crucial to hyperspectral image classification, is provided in the following subsections.

A. FEATURE MINING

Hyperspectral imaging is characterized by the high spectral resolution available, which allows capturing fine details of the spectral characteristics of materials in a wide range of applications. However, it has been demonstrated that the original spectral features contain high redundancy [3]. Specifically, there is a high correlation between adjacent bands and the number of the original spectral features may be too high for classification purposes [3], [75]. In addition, the original spectral features may not be the most effective ones to separate the objects of interest from others. These observations have fostered the use of feature mining techniques so that an effective set of features can be identified prior to classification. In this subsection we briefly outline some of the available approaches for feature mining from hyperspectral data sets.

1) Feature extraction: Several strategies have been used in the hyperspectral imaging literature to perform feature extraction prior to classification purposes. A distinguishing characteristic of feature extraction methods is that they exploit all available spectral measurements in order to extract relevant features. A widely used approach has been the generation of features in a new space, like those obtained from the PCA [3] or the minimum noise fraction (MNF) [77]. In these techniques, the hyperspectral data are projected onto a new space in which the first few components account for most of the total information in the data, and therefore only the first few features could be retained. The segmented PCA [78] reduces the computational load significantly for feature extraction, compared with the conventional PCA. Another spectral-based approach to generate new features has been the discrete wavelet transform (DWT), which allows for the separation of high and low frequency components separately. This allows a form of derivative analysis which has been also used to generate features prior to hyperspectral image classification [79]. Another popular strategy has been canonical analysis [3], which is focused on the extraction of features that maximize the ratio between the variance among classes and the average variance within the classes. However, this approach requires good estimates of the class covariance matrices, and therefore a generally large number of training samples (which may not be available in practice) are often required. An alternative strategy to deal with this problem has been to use semi-supervised feature extraction [80], in which only a few labeled samples and additional unlabeled samples are used. Other widely used methods have been nonparametric weighted feature extraction (NWFE) or decision boundary feature extraction (DBFE) [75]. Another strategy for feature extraction has been grouping of neighboring bands, using techniques such as the weighted sum or average of each group [81]. A free Matlab toolbox for linear and nonlinear feature extraction methods is simFEAT[12].

2) Feature selection: In feature selection, the idea is to select a set of spectral bands from the initial pool of bands available prior to classification. A particular characteristic of feature selection methods is that they tend to retain the spectral meaning (while reducing the number of bands). In unsupervised feature selection, the goal is to automatically find statistically important features. The advantage of unsupervised methods is that they do not need training data. Quite opposite, supervised feature selection is based on general/expert knowledge, and require labeled and (often) unlabeled training samples. Techniques in the latter category comprise methods based on class separability measures using standard distance metrics (e.g., Euclidean,

---

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

[12]http://www.uv.es/gcamps/code/simfeat.htm

Mutual information, Bhattacharyya), or more sophisticated class separability measures in feature space [3]. In this regard, methods have been proposed that implement an exhaustive search of optimal features, such as the progressive two-class decision classifier [82]. Other advanced feature selection strategies (e.g., using kernels) are described in [83].

B. SUPERVISED CLASSIFICATION

Several techniques have been used to perform supervised classification of hyperspectral data. For instance, in discriminant classifiers several types of discriminant functions can be applied: nearest neighbor, decision trees, linear functions, nonlinear functions, etc. In linear discriminant analysis (LDA) [84], a linear function is used in order to maximize the discriminatory power and separate the available classes effectively. However, such a linear function may not be the best choice and nonlinear strategies such as quadratic discriminant analysis (QDA) or logarithmic discriminant analysis (LogDA) have also been used. The main problem of these classic supervised classifiers, however, is their sensitivity to the Hughes effect.

In this context, kernel methods such as the support vector machine (SVM) have been widely used in order to deal effectively with the Hughes phenomenon [85], [86]. The SVM was first investigated as a binary classifier [87]. Given a training set mapped into an Hilbert space by some mapping, the SVM separates the data by an optimal hyperplane that maximizes the margin. If the data are not linearly separable, soft margin classification with slack variables can be used to allow mis-classification of difficult or noisy cases. However, the most widely used approach in hyperspectral classification is to combine soft margin classification with a kernel trick that allows separation of the classes in a higher dimensional space by means of a nonlinear transformation. In other words, the SVM used with a kernel function is a nonlinear classifier, where the nonlinear ability is included in the kernel and different kernels lead to different types of SVMs. The extension of SVM to the multi-class cases is usually done by combining several binary classifiers. Two classical procedures are the *one versus the rest* and the *one versus one* [87].

In the following, we illustrate the performance of SVMs (implemented using the Gaussian radial basis function kernel) by processing a widely used hyperspectral data set collected by the Reflective Optics Imaging Spectrographic System (ROSIS) optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Commission. The image size in pixels is 610×340, with very high spatial resolution of 1.3 meters per pixel. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). Fig. 11(a) shows a false color composite of
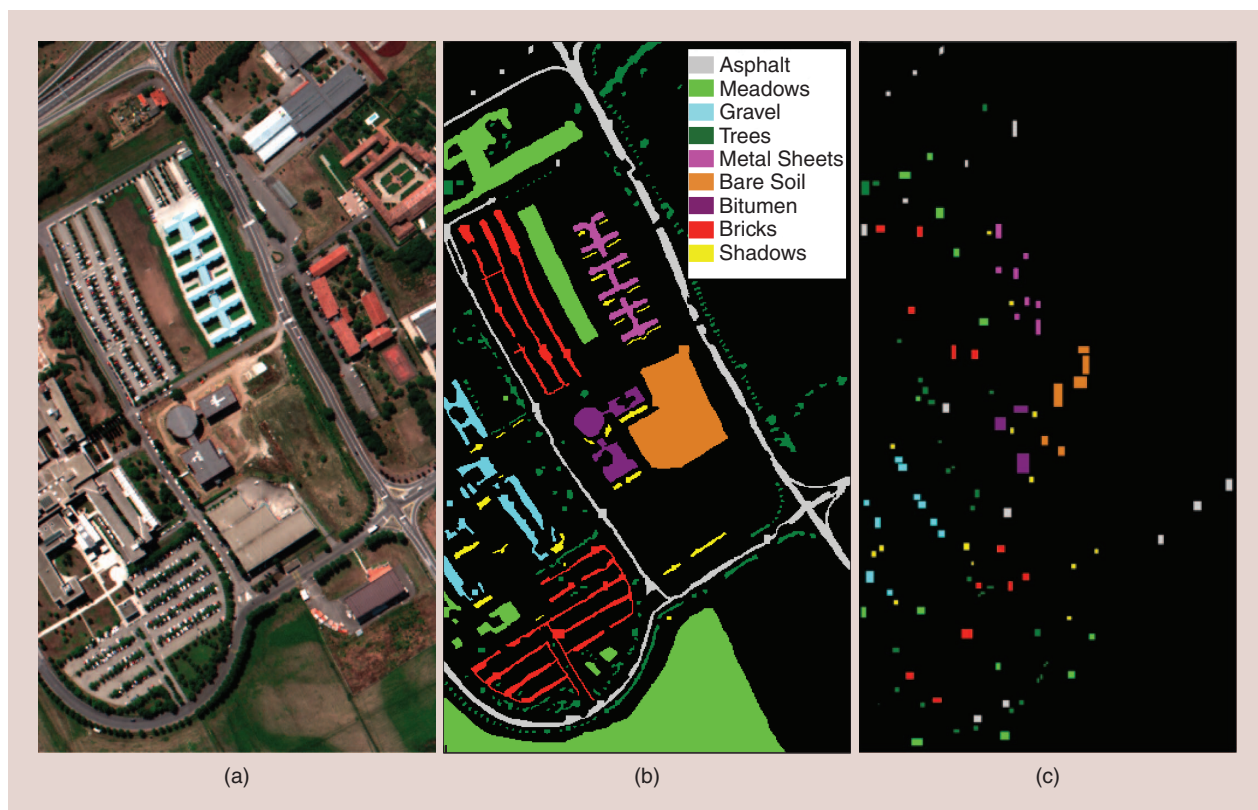

FIGURE 11. (a) False color composition of the ROSIS University of Pavia scene. (b) Reference map containing 9 mutually exclusive land-cover classes. (c) Training set used in experiments.

the image, while Fig. 11(b) shows nine reference classes of interest, which comprise urban features, as well as soil and vegetation features. Finally, Fig. 11(c) shows a fixed training set available for the scene which comprises 3921 training samples (42776 samples are available for testing).

Table 2 illustrates the classification results obtained by different supervised classifiers for the ROSIS University of Pavia scene in Fig. 11(a), using the same training data in Fig. 11(c) to train the classifiers and a mutually exclusive set of labeled samples in Fig. 11(b) to test the classifiers. As shown by Table 2, the SVM classifier obtained comparatively superior performance in terms of the overall classification accuracy (OA), average classification accuracy (AV) and kappa statistic [88] when compared with discriminant classifiers such as LDA, QDA or LogDA. In this experiment, the SVM was also slightly superior to the multinomial logistic regression (MLR) classifier [89], which has been recently explored in hyperspectral imaging as a technique able to model the posterior class distributions in a Bayesian framework, thus supplying (in addition to the boundaries between the classes) a degree of plausibility for such classes [90]. A subspace-based version of this classifier, called MLR*sub* [91], is also included in the comparison given in Table 2. The idea of applying subspace projection methods relies on the basic assumption that the samples within each class can approximately lie in a lower dimensional subspace. However, in the experiments reported in [91] for the MLR*sub* it was observed that spatial information needs to be included in this (and other classifiers) in order to improve classification performance. In the following subsection, we summarize some techniques for spatial-spectral classification.

C. SPATIAL-SPECTRAL CLASSIFICATION

Several efforts have been performed in the literature in order to integrate spatial-contextual information in spectral-based classifiers for hyperspectral data [76]. It is now commonly accepted that using the spatial and the spectral information simultaneously provides significant advantages in terms of improving the performance of classification techniques. Some of these approaches include spatial information prior to the classification, during the feature extraction stage. Mathematical morphology [92] has been particularly successful for this purpose. Morphology is a widely used approach for modeling the spatial characteristics of the objects in remotely sensed images. Advanced morphological techniques such as morphological profiles (MPs) [93] have been successfully used for feature

extraction prior to classification of hyperspectral data by extracting the first few principal components of the data using the PCA [3], and then building so-called extended morphological profiles (EMPs) on the first few components to extract relevant features for classification [94].

As shown by Table 3, the combination of EMP for feature extraction followed by SVM for classification (EMP/SVM) provides good classification results for the ROSIS University of Pavia scene. Recently, morphological attribute profiles (APs) [95] were introduced as an advanced mechanism to obtain a detailed multilevel characterization of a hyperspectral image created by the sequential application of morphological attribute filters that can be used (prior to classification) to model different kinds of the structural information. According to the type of the attributes considered in the morphological attribute transformation, different parametric features can be modeled. The use of different attributes leads to the concept of extended multi-attribute profiles (EMAPs) which have been also used successfully for hyperspectral image classification purposes [96].

Another strategy in the literature has been to exploit simultaneously the spatial and the spectral information. For instance, in order to incorporate the spatial context into kernel-based classifiers, a pixel entity can be redefined simultaneously both in the spectral domain (using its spectral content) and also in the spatial domain, by applying some feature extraction to its surrounding area which yields spatial (contextual) features, e.g., the mean or standard deviation per spectral band. These separated entities lead to two different kernel matrices, which can be easily computed. At this point, one can sum spectral and textural dedicated kernel matrices and introduce the cross-information between textural and spectral features in the formulation. This simple methodology yields a full family of new kernel methods for hyperspectral data classification, defined in [97] and implemented using the SVM classifier thus providing a composite kernel-based SVM (SVM-CK) illustrated in Table 3 (using the summation kernel). Recently, composite kernels have been generalized in [98], using the MLR

TABLE 2. ACCURACIES OBTAINED BY DIFFERENT SUPERVISED CLASSIFIERS FOR THE ROSIS UNIVERSITY OF PAVIA SCENE.

METRIC	LDA	QDA	LogDA	SVM	MLR	MLR*sub*
OA	77.95	77.95	78.41	80.99	80.11	67.08
AV	73.67	78.73	79.82	88.28	87.80	77.20
kappa	0.606	0.770	0.720	0.761	0.750	0.703

TABLE 3. OVERALL ACCURACIES OBTAINED BY DIFFERENT SUPERVISED SPATIAL-SPECTRAL CLASSIFIERS FOR THE ROSIS UNIVERSITY OF PAVIA SCENE.

METRIC	ECHO	LDA-MLL	QDA-MLL	LOGDA-MLL	SVM-CK	EMP/SVM	SVM-W	SVM-RHSEG	MLR-MLL	MLR*sub* MLL	MLR-GCK
OA	87.58	80.27	89.48	87.04	87.18	85.22	85.42	93.85	85.57	94.10	98.09
AV	92.16	78.05	91.91	83.32	90.47	90.76	91.31	97.07	92.54	93.45	97.76
kappa	0.839	0.739	0.864	0.872	0.871	0.808	0.813	0.918	0.818	0.922	0.974

classifier and EMAPs to define spatial context. The resulting generalized composite kernel-based MLR (MLR-GCK) can linearly combine multiple kernels without any restriction of convexity. This introduces a different approach with regards to the SVM-CK and multiple kernel learning methods, in which composite kernels need to be convex combinations of kernels. This approach provided the best classification result for the ROSIS University of Pavia scene in Table 3.

Another approach to jointly exploit spatial and spectral information is to use Markov random fields (MRFs) for the characterization of spatial information. MRFs exploit the continuity, in probability sense, of neighboring labels [99]. In this regard, several techniques have introduced an MRF-based multinomial logistic level (MLL) prior which encourages neighboring pixels to have the same label when performing probabilistic classification of hyperspectral data sets. As it can be seen in Table 2, such MLL prior can significantly improve the results provided by spectral-based classifiers such as the LDA, QDA, LogDA, MLR and, most notably, the MLR*sub* described in the previous subsection. At this point, it is worth noting that the combination of a subspace-based classifier such as the MLR*sub* with an MLL prior enforcing spatial homogeneity in the resulting segmentation provides one of the best classification results in Table 3.

Several other approaches include spatial information as a post-processing, *i.e.*, after a spectral-based classification has been conducted. One of the first classifiers with spatial post-processing developed in the hyperspectral imaging literature was the well-known ECHO (extraction and classification of homogeneous objects) [75]. Another one is the strategy adopted in [100], which combines the output of a pixel-wise SVM classifier with the morphological watershed (SVM-W) transformation [92] in order to provide a more spatially homogeneous classification. A similar strategy is adopted in [101], in which the output of the SVM classifier is combined with the segmentation result provided by the RHSEG segmentation algorithm (SVM-RHSEG). These strategies lead to much improved classification results with regards to spectral-based classification, as it can be observed in Table 3. A detailed overview of recent advances in spatial-spectral classification of hyperspectral data is available at [102].

Last but not least, an important recent development has been the use of sparse representation classifiers using dictionary-based generative models [103]. In this case, an input signal is represented by a sparse linear combination of training samples (atoms) from a dictionary [103]. The classification can be improved by incorporating the contextual information from the neighboring pixels into the classifier. This can be done indirectly by exploiting the spatial correlation through an structured sparsity prior imposed in the optimization process. As shown in [103], the performance of different sparsity-based classifiers are comparable to the state-of-the-art SVM-CK classifier. Given sufficient training data some researchers have also developed discriminative as well as compact class dictionaries to improve classification performance [104].

D. SEMI-SUPERVISED CLASSIFICATION

A relevant challenge for supervised classification techniques is the limited availability of labeled training samples, since their collection generally involves expensive ground campaigns [105]. While the collection of labeled samples is generally difficult, expensive and time-consuming, unlabeled samples can be generated in a much easier way. This observation has fostered the idea of adopting semi-supervised learning techniques in hyperspectral image classification. The main assumption of such techniques is that new (unlabeled) training samples can be obtained from a (limited) set of available labeled samples without significant effort/cost.

In contrast to supervised classification, semi-supervised algorithms generally assume that a limited number of labeled samples are available *a priori*, and then enlarge the training set using unlabeled samples, thus allowing these approaches to address ill-posed problems. However, in order for this strategy to work, several requirements need to be met. First and foremost, the new (unlabeled) samples should be obtained without significant cost/effort. Second, the number of unlabeled samples required for the semi-supervised classifier to perform properly should not be too high in order to avoid increasing computational complexity in the classification stage. In other words, as the number of unlabeled samples increases, it may be unbearable for the classifier to properly exploit all the available training samples due to computational issues. Further, if the unlabeled samples are not properly selected, these may confuse the classifier, thus introducing significant divergence or even reducing the classification accuracy obtained with the initial set of labeled samples. In order to address these issues, it is very important that the most highly informative unlabeled samples are identified in computationally efficient fashion, so that significant improvements in classification performance can be obtained without the need to use a very high number of unlabeled samples.

The area of semi-supervised learning for remote sensing data analysis has experienced a significant evolution in recent years. For instance, in [106] transductive SVMs (TSVMs) are used to gradually search a reliable separating hyperplane (in the kernel space) with a transductive process that incorporates both labeled and unlabeled samples in the training phase. In [107], a semi-supervised method is presented that exploits the wealth of unlabeled samples in the image, and naturally gives relative importance to the labeled ones through a graph-based methodology. In [108], kernels combining spectral-spatial information are constructed by applying spatial smoothing over the original hyperspectral data and then using composite kernels in graph-based classifiers. In [109], a semi-supervised SVM is presented that exploits the wealth of unlabeled samples for regularizing the training kernel representation locally by means of cluster kernels. In [90], a new semi-supervised approach is presented that exploits unlabeled training samples (selected by means of an active selection strategy based on the entropy of

the samples). Here, unlabeled samples are used to improve the estimation of the class distributions, and the obtained classification is refined by using a spatial multi-level logistic prior. In [110], a novel context-sensitive semi-supervised SVM is presented that exploits the contextual information of the pixels belonging to the neighborhood system of each training sample in the learning phase to improve the robustness to possible mislabeled training patterns.

In [111], two semi-supervised one-class (SVM-based) approaches are presented in which the information provided by unlabeled samples present in the scene is used to improve classification accuracy and alleviate the problem of free-parameter selection. The first approach models data marginal distribution with the graph Laplacian built with both labeled and unlabeled samples. The second approach is a modification of the SVM cost function that penalizes more the errors made when classifying samples of the target class. In [112] a new method combines labeled and unlabeled pixels to increase classification reliability and accuracy, thus addressing the sample selection bias problem. In [113], an SVM is trained with the linear combination of two kernels: a base kernel working only with labeled examples is deformed by a likelihood kernel encoding similarities between labeled and unlabeled examples, and then applied in the context of urban hyperspectral image classification. In [114], similar concepts to those addressed before are adopted using a neural network as the baseline classifier. In [115], a semi-automatic procedure to generate land cover maps from remote sensing images using active queries is presented and discussed.

At this point, it should be noted that active learning techniques have been mainly exploited in a supervised context, *i.e.* a given supervised classifier is trained with the most representative training samples selected after a (machine-human) interaction process in which the samples are actively selected according to some criteria based on the considered classifier, and then the labels of those samples are assigned by a trained expert in fully supervised fashion [90], [116]–[118]. In this supervised context, samples with high uncertainty are generally preferred as they are usually more informative. At the same time, since the samples are labeled by a human expert, high confidence can be expected in the class label assignments. As a result, classic (supervised) active learning generally focuses on samples with high confidence at the human level and high uncertainty at the machine level.

Recently, standard active learning methods have been adapted into a semi-supervised self-learning scenario [119]. The main idea is to obtain unlabeled samples (from a pool of samples) using machine-machine interaction instead of human supervision. The first (machine) level—similar to the human level in classic (supervised) active learning—is used to infer a set of candidate unlabeled samples with high confidence. In a second (machine) level—similar to the machine level for supervised active learning—the machine learning algorithm itself automatically selects the samples

with highest uncertainty from the obtained candidate set. This strategy relies on two main assumptions. The first assumption (global) is that training samples having the same spectral structure likely belonging to the same class. The second assumption (local) is that spatially neighboring pixels likely belong to the same class. As a result, this approach naturally integrates the spatial and the spectral information in the semi-supervised classification process.

E. CHALLENGES
There are several important challenges when performing hyperspectral image classification. Supervised classification faces challenges related with the unbalance between high dimensionality and limited availability of training samples, or the presence of mixed pixels in the data (which may compromise classification results for coarse spatial resolutions). Another relevant challenge is the need to integrate the spatial and spectral information to take advantage of the complementarities that both sources of information can provide. These challenges are quite important for future developments and solutions to some of them have been outlined in this section. Specifically, we have explored techniques such as supervised and semi-supervised techniques for hyperspectral image classification, strategies for integrating the spatial and the spectral information, or sparse classifiers that can bring solutions to the aforementioned problems. However, some issues still remain. For instance, the geometry of hyperspectral data is quite complex and dominated by nonlinear structures. This issue has undoubtedly an impact in the outcome of the classification techniques discussed in this section. In order to mitigate this, manifold learning has been proposed [120]. An important property of manifold learning is that it can model and characterize the complex nonlinear structure of the data prior to classification [121]. Another remaining issue is the very high computational complexity of some of these classifiers discussed in this section. In other words, there is a need to develop efficient classification techniques that can deal with the very large dimensionality and complexity of hyperspectral data. In Section VII we discuss fast computing solutions for hyperspectral imaging algorithms.

Last but not least, we emphasize that the techniques described in this section only represent a small sample (and somehow subjective selection) of the vast collection of approaches presented in recent years for hyperspectral image classification. For a more exhaustive summary of available techniques and future challenges in this area, we point interested readers to [1].

V. HYPERSPECTRAL TARGET DETECTION
Hyperspectral imagery has been used in reconnaissance and surveillance applications where targets of interest are detected and identified [7], [122]. The process of detecting and identifying a target in hyperspectral imagery can be considered as consisting of two stages. The first stage is an anomaly detector [123], [124] which identifies spectral

vectors that have significant spectral differences from their surrounding background pixels. Man-made anomalies can also be detected through change detectors [125], [126], which are used to identify changes within a scene that occur over time. The second stage is to identify whether or not the anomaly is a target or natural clutter. This stage can be achieved if the spectral signature of the target is known which can be obtained from a spectral library [122] or from a set of training data which could also be synthetically generated [127]. Almost all the classical target detection techniques in the literature [127]–[131] are based on a linear process that only exploits the first and second order statistics to identify anomalies or targets. Advanced nonlinear detection techniques based on statistical kernel learning theory [132] have also been developed in [133] that indirectly exploit the higher order statistics between the spectral bands through a kernel function [132].

A. ANOMALY DETECTION

Anomaly detectors, outlier detectors, or novelty detectors are pattern recognition or statistical schemes that are used to detect objects that stand out from their cluttered background. In spectral anomaly detection algorithms [123], [124], [134]–[136] pixels (materials) that have a significantly different spectral signature from their neighboring background clutter pixels are identified as spectral anomalies. In such algorithms, no prior knowledge of the target spectral signature is utilized or assumed. In [134], a spectral anomaly detection algorithm was developed for detecting targets of unknown spectral distribution against a background with unknown spectral covariance. This algorithm is now commonly referred to as the Reed-Xiaoli (RX) anomaly detector, has been successfully applied to many hyperspectral target detection applications [7], [124], [135], [136] and is considered as the benchmark anomaly detection algorithm for multispectral/hyperspectral data. The RX algorithm is a constant false alarm rate adaptive anomaly detector which is derived from the generalized likelihood ratio test (GLRT). Assuming a single pixel target $\mathbf{y}$ as the observation test vector, the results of RX-algorithm is given by

$$RX(\mathbf{y}) = (\mathbf{y} - \hat{\boldsymbol{\mu}}_b)^T \hat{\mathbf{C}}_b^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}_b), \tag{5}$$
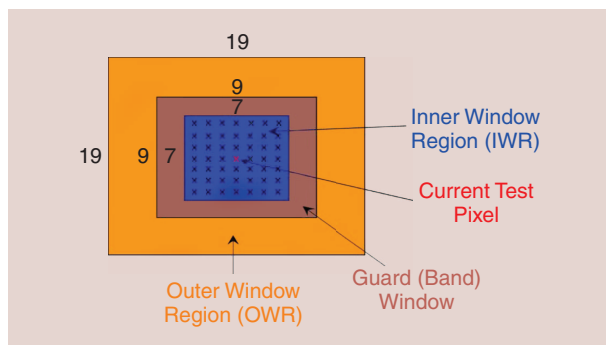

FIGURE 12. An example of a dual window.

where $\hat{\boldsymbol{\mu}}_b$ is the estimated background clutter sample mean and $\hat{\mathbf{C}}_b$ is the estimated background clutter covariance.

The background mean and covariance matrix can be estimated globally from the whole hyperspectral image or locally using a double concentric window approach [123]. To estimate $\hat{\mathbf{C}}_b$ globally the background pixels are usually modeled as a mixture of multivariate Gaussian distributions [137], linear subspace [138], [139], linear or stochastic mixture models [140] or by some clustering or segmentation techniques [141]–[143] which are used to segment the background into several clusters. On the other hand the local background covariance matrix can be estimated by using a sliding double concentric window, centered at each test pixel, which consists of a small inner window region (IWR) centered within a larger outer window region (OWR), as shown in Fig. 12. The local background mean vector and covariance matrix are then computed from the spectral pixels falling within the OWR. The size of the inner window is assumed to be the size of the typical target of interest in the image. A guard band surrounding the IWR is also sometimes used to prevent the target pixels from corrupting the calculation of the background OWR statistics. The whole background probability density function has also been modeled by a single class support vector machine in [144] and spectral pixels that fall outside this model are considered as anomalies. Anomaly detection techniques formulated as eliminating the whole or local background subspace from every pixel have also been investigated in [145].

Several variations of the RX detector that attempt to alleviate the limitation of RX have been proposed in the literature [135]–[137]. In [146], a modification to the RX algorithm called SubSpace RX (SSRX) was outlined that is based on the PCA of the background covariance matrix. In the SSRX algorithm, several high-variance background dimensions are deleted before applying the RX algorithm as these are assumed to capture non-normal background clutter variance. Another consideration in RX implementation is potential ill-conditioning of the local covariance matrix due to the high correlation, high dimensionality of the hyperspectral data and a limited background sample size. This ill-conditioning is typically addressed by a regularization procedure such as PCA-based regularization or adding a scaled identity matrix to the background covariance matrix [147].

B. SIGNATURE-BASED TARGET DETECTION

In some applications, we have some prior knowledge about the spectral characteristics of the desired targets. In these situations, the target spectral characteristics can be defined by a single target spectrum [148] or by a signal subspace [129]. The GLRT detector using a single targets spectrum is referred to as the spectral matched filter (SMF) and the maximum likelihood abundance estimate of the target in a test pixel $\mathbf{y}$ is given in [130] as

$$D_{\mathrm{SMF}}(\mathbf{y}) = \frac{\mathbf{s}^T \hat{\mathbf{C}}^{-1} \mathbf{y}}{\mathbf{s}^T \hat{\mathbf{C}}^{-1} \mathbf{s}}, \tag{6}$$

where **s** is the spectral signature of the target and $\hat{\mathbf{C}}$ represents the estimated covariance matrix for the centered observation data. The SMF model is based on the assumption that the background clutter noise has a Gaussian distribution $\mathcal{N}(\mathbf{0}, \hat{\mathbf{C}})$ and the target distribution is also a Gaussian $\mathcal{N}(a\mathbf{s}, \hat{\mathbf{C}})$ with the same covariance statistics but with a mean of $a\mathbf{s}$, where a is an scalar abundance value representing the target strength.

The mean and covariance matrix appearing in (6) are estimated from the data either locally, or globally under the assumption that the number of target pixels contaminating the estimation of the covariance matrix is insignificant. The spectral target signatures are usually obtained from the spectral library or from a target training data set. An essential pre-processing stage to implement a signature-based detector is to estimate and compensate the atmospheric effects on the data [149] in order to transform the known target spectrum and measurement data into a common domain where an algorithm such as the SMF can be applied. Another consideration in SMF implementation is potential ill-conditioning of the covariance matrix due to the high correlation, high dimensionality of the hyperspectral data and from a limited background sample size. Representing the inverse covariance matrix in terms of its eigenvector-eigenvalue decomposition, it becomes clear that the behavior of the inverse covariance matrix depends heavily on the small eigenvalues which could render it unstable. In order to reduce SMF sensitivity to statistical and numerical errors, eigenvectors corresponding to eigenvalues below an appropriate condition number or estimated sensor noise level are discarded [150], or a scaled identity matrix is added to the background clutter covariance before inverting, which is equivalent to including a regularization term (penalty term) in the design of the SMF [151].

In a situation when the target and background pixel characteristics are modeled by two linear subspaces with additive noise the resulting detector is referred to as the subspace matched detector (MSD). The GLRT for such a two subspace model is given in [129] as

$$D_{\text{MSD}}(\mathbf{y}) = \frac{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_\mathbf{B})\mathbf{y}}{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_{\text{TB}})\mathbf{y}}, \qquad (7)$$

where $\mathbf{P}_\mathbf{B}$ is the projection matrix associated with the background subspace $\langle \mathbf{B} \rangle$, and $\mathbf{P}_{\text{TB}}$ is the projection matrix associated with the target-and-background subspace $\langle \mathbf{TB} \rangle$. Usually, the eigenvectors corresponding to the significant eigenvalues of the target and background covariance matrices are used to generate the columns of **T** and **B**, respectively. To generate appropriate target and background linear subspaces researchers have used scene samples from the hyperspectral image itself or have used the MODTRAN software package to generate a large number of synthetic target and background spectral pixels [127] for a given environment in order to estimate the two environmentally invariant subspaces **T** and **B**.

Using some variations of the models used in SMF and MSD, researchers have also developed a number of different detectors, such as the adaptive subspace detector (ASD) [131] and orthogonal subspace projection (OSP) [128]. In ASD the target signature is represented by a linear subspace and the background statistics by a zero-mean Gaussian distribution whose covariance is estimated from the hyperspectral image data itself. In the case of OSP the target signature is represented by a single spectral vector s and the background is represented by the spectral signatures of the undesired background endmembers **B**. The OSP algorithm is based on maximizing the signal-to-noise ratio (SNR) in the subspace orthogonal to the background subspace which is given by $D_{OSP} = \mathbf{s}^T \mathbf{P}_{\dot{B}}^\perp \mathbf{y}$ where $\mathbf{P}_{\dot{B}}^\perp = \mathbf{I} - \mathbf{B}\mathbf{B}^\#$ is the background rejection operator and $(\cdot)^\#$ denotes the pseudo inverse matrix.

C. SPARSE REPRESENTATION TARGET DETECTION

In the sparse representation classifiers (SRC) a test sample is approximately represented by very few training samples from both target and background dictionaries, and the recovered sparse representation is directly used for classification [152]. Given the concatenated training samples from the target and background dictionaries as $\mathbf{A} = [\mathbf{A}_b \ \mathbf{A}_t]$, the sparse representation vector $\boldsymbol{\gamma}$ satisfying $\mathbf{A}\boldsymbol{\gamma} = \mathbf{y}$ can be obtained by solving the following optimization problem

$$\hat{\boldsymbol{\gamma}} = \arg\min \|\boldsymbol{\gamma}\|_0 \quad \text{subject to} \quad \mathbf{A}\boldsymbol{\gamma} = \mathbf{y}, \qquad (8)$$

where $\|\boldsymbol{\gamma}\|_0$, termed the ℓ_0-norm, is defined as the number of non-zero entries in the vector $\boldsymbol{\gamma}$ that also represents a sparsity prior or penalty in (8). The above problem of minimizing the ℓ_0-norm is NP-hard, but its approximate solution can be obtained by greedy algorithms [68]. The ℓ_0-norm can also be replaced by an ℓ_1-norm regularization prior term in (8) [69], where standard convex optimization algorithms [70] can be used. Once the sparse coefficient vector $\boldsymbol{\gamma}$ is obtained, the class of the test pixel **y** can be determined by comparing the residuals $r_b(\mathbf{y}) = \|\mathbf{y} - \mathbf{A}_b\hat{\boldsymbol{\alpha}}\|_2$ and $r_t(\mathbf{y}) = \|\mathbf{y} - \mathbf{A}_t\hat{\boldsymbol{\beta}}\|_2$, where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ represent the recovered sparse coefficient vectors corresponding to the background and target dictionaries, respectively.

In the above process, the sparsity-based target detector is applied to each pixel in the test region independently without considering the correlation (contextual) between neighboring pixels. To incorporate contextual information within the SRC algorithm other sparsity priors or penalties, such as joint sparsity (collaborative) or an l_2-norm smoothness constraint [153] can be considered. Fig. 13 shows the receiver operating characteristic (ROC) curves for several different target detection techniques on a typical hyperspectral image, the forest radiance I data collection (FR-I), obtained from a hyperspectral digital imagery collection experiment (HYDICE) sensor which consists of 210 bands across the whole spectral range from 0.4 to 2.5 μm which includes the visible and short-wave infrared bands. As seen from the ROC curves in Fig. 13 the SVM-CK and sparsity-based with smoothing classifiers outperform the classical
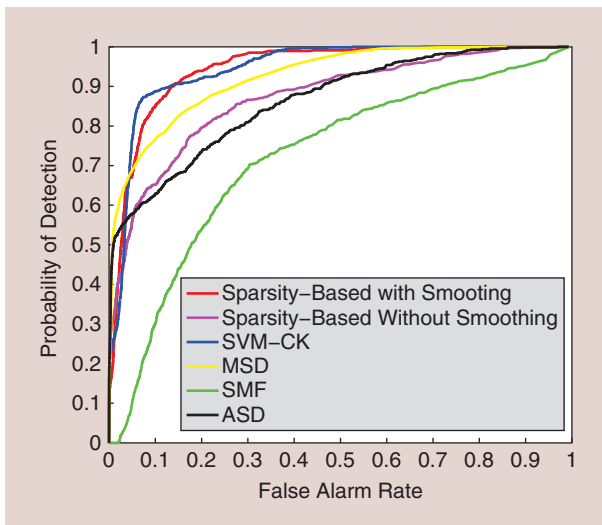
FIGURE 13. ROC curves for a typical hyperspectral image (FR-I) with several military targets reproduced from [152].

signature-based target detection techniques, the experimental detail and results for couple of other hyperspectral images can be found in [152].

D. NONLINEAR DETECTORS

Almost all the anomaly and target detectors are based on only exploiting first and second order statistics in order to identify anomalies or targets. Kernel machine learning theory [132] has emerged as a new nonlinear-based learning approach for extending the classical pattern recognition algorithms. The implicit exploitation of nonlinear features through kernels provides crucial information about a given dataset which, in general, the learning methods based on linear models cannot achieve. The RX anomaly detection algorithm, the statistical target detection algorithms and the sparsity-based target classifier have all been extended

FIGURE 14. Forward (solid lines) and inverse (dashed lines) problems in remote sensing. Figure adapted from [157].

to their nonlinear versions [133], [154] by using the ideas in kernel machine learning theory. Experimental results show that typically the kernel-based algorithms outperform their linear versions.

E. CHALLENGES

The major challenges in the classical anomaly and target detection techniques (RX, SMF, MSD, ASD, OSP) are still the need for developing new approaches for estimating the background/target covariance matrices or their corresponding subspaces given a limited training data. Further research is also needed in the classical techniques to incorporate the spatial-contextual information about the targets that are more than one pixel size. In the case of sparsity-based techniques more research is needed to develop the appropriate class sub-dictionaries as well as compact discriminative dictionaries. More advanced structured sparsity priors are to be incorporated and their performance evaluated. Currently most of the non-linear methods are based on kernel learning theory, other nonlinear approaches beside kernel-based methods need to be introduced.

VI. ESTIMATION OF LAND PHYSICAL PARAMETERS

This section reviews the main problems and methods in the field of model inversion and estimation of physical parameters. Our main goal in remote sensing is to monitor the Earth and its interaction with the atmosphere. The analysis can be done at local or global scales by looking at bio-geo-chemical cycles, atmospheric state and evolution, and vegetation dynamics [155], [156]. All these complex interactions can be studied through the definition of physical parameters representing different properties for land (e.g., surface temperature, biomass, leaf area coverage), water (e.g., yellow substance, ocean color, suspended matter) or the atmosphere (e.g., temperature and moisture profiles at different altitudes). The field of physical parameter estimation is an intermediate modeling step necessary to transform the measurements[13] into useful estimates [157].

The remote sensing modeling system is illustrated in Fig. 14. The *forward* (or direct) problem involves *radiative transfer models* (RTMs). These models summarize the physical processes involved in the energy transfer from canopies and atmosphere to measured radiance. They simulate the reflected or emitted radiation transmitted through the atmosphere for a given observation configuration (e.g., wavelength, view and illumination directions) and some auxiliary variables (e.g., vegetation and atmosphere characteristics). Solving the *inversion* problem implies the design of algorithms that, starting from the radiation acquired by the sensor, can give accurate estimates of the variables of interest, thus "inverting" the RTM. In the inversion process, a priori information of the variables of interest can also be
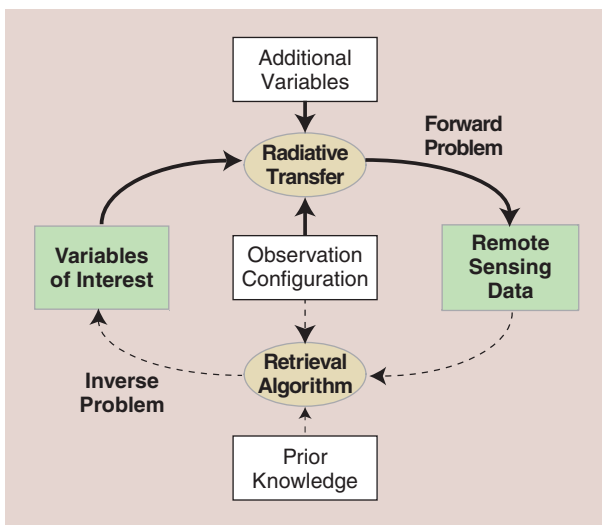
---

[13]The acquired data may consist of multispectral or hyperspectral images provided by satellite or airborne sensors, but can also integrate spectra acquired by *in situ* (field) radiometers, GIS data that help to integrate geographical information, or radiosonde measures for weather monitoring.

included to improve the performance, such as the type of surface, geolocation, or acquisition time.

Notationally, a *discrete forward model* can be expressed as

$$\mathbf{Y} = f(\mathbf{x}, \theta) + \mathbf{n},$$

where $\mathbf{Y}$ is a set of measurements (such as the expected radiance); $\mathbf{x}$ is the state vector that describes the system (e.g., parameters such as temperature or moisture); θ contains a set of controllable measurement conditions (e.g., different combinations of wavelength, viewing direction, time, Sun position, and polarization); $\mathbf{n}$ is an additive noise vector; and $f(\cdot)$ is a complex nonlinear continuous function that relates $\mathbf{x}$ with $\mathbf{Y}$. The *discrete inverse model* is then defined as

$$\hat{\mathbf{x}} = g(\mathbf{Y}, \omega),$$

where $g(\cdot)$ is a possibly nonlinear function, parameterized by weights ω, that approximates the measurement conditions, $\mathbf{x}$, using a set of observations as inputs, $\mathbf{Y}$.

A. TAXONOMY OF INVERSION METHODS

Model inversion methods can be roughly divided into statistical, physical or hybrid. In what follows, we review the main contributions in each family for land parameter retrieval. The reader is referred to [1] for an updated review on atmospheric and ocean parameter retrieval methodologies.

1) Statistical inversion: *Statistical* inversion can be done either with parametric or non-parametric models. Parametric models rely on physical knowledge of the problem and build explicit parameterized expressions that relate a few spectral channels with the bio-geo-physical parameter of interest. Different narrowband vegetation indices (VIs) have been proposed to study the vegetation status by estimating chlorophyll content and other leaf pigments, as well as vegetation density parameters like the leaf area index (LAI) and the fractional vegetation cover (FVC) [158]. The simple calculation of these indices has made possible deriving reasonable maps of vegetation properties in a quick and easy way. Furthermore, since the launch of imaging spectrometers into space crafts, these VIs have been applied at canopy level on ecosystems across the globe [159]. Nevertheless, the majority of the indices only use up to five bands thus under-exploiting the full potential information contained in hyperspectral images [160].

Alternatively, non-parametric models estimate the variable of interest using a set of input-output training data pairs, which come from concurrent measurements of the parameter and the corresponding reflectance/radiance observation. A terrestrial campaign is thus necessary at the same time the satellite overpasses the study area to measure the surface parameter. Several nonparametric approaches have been introduced for land parameter retrieval. In [161], biomass was estimated using common spectral band ratios, vegetation indices and linear/stepwise multiple regression models. Partial least squares (PLS) regression has been

used for mapping canopy nitrogen [162]. Nonlinear extension of PLS was introduced via kernels in [163] for chlorophyll concentration estimation. Recently, the support vector regression (SVR) has yielded good results in modeling oceanic chlorophyll [164]. In the recent years, Gaussian Process (GP) regression has shown very good properties to tackle the physical parameter estimation: GPs are simpler to train than other models, they show good numerical performance and stability and provide sensible confidence intervals for the predictions [165].

2) Physical inversion: Statistical approaches may lack transferability, generality, or robustness to new geographical areas. Physical inversion models can alleviate these shortcomings by coupling models from lower to higher levels (e.g., canopy level models built upon leaf models). Therefore, they provide a physically-based, interpretable and broad linkage between Earth observation data and biochemical or structural state variables [166]. Running RTMs in forward mode enables creating a database covering a wide range of situations and configurations. Forward RTM simulations allow for sensitivity studies of canopy parameters relative to diverse observation specifications, and allow us to better understand the observed signal. The use of RTMs to generate datasets is a common practice, and especially convenient because acquisition campaigns are very costly (in terms of time, money, and human resources), and usually limited in terms of parameter combinations. RTMs are also widely used in the preliminary phase of a new sensor design, which allows understanding both the limits and capabilities of the instrument for the retrieval tasks. There exist many RTMs implemented in software packages to deal with the forward modeling. For example, PROSPECT is an RTM accounting for leaf optical properties while SAIL constitutes a canopy bidirectional reflectance model. Their combination led to the PROSAIL model. Other RTMs are more specific to ocean or atmosphere applications.

Physical inversion models essentially try to reverse (invert) data generated with an RTM. The basic assumption for inverting RTMs is that the forward model, $f(\cdot)$, contains all the necessary information about the problem, so its inversion may lead to accurate parameter estimates. When a unique solution is not achieved, more a priori information is required to overcome the ill-posed problem. After generating the dataset, $\{\mathbf{Y}, \mathbf{x}\}$, the problem reduces to, given new spectra, assign the parameter of the 'closest' spectra. Several approaches and metrics have been considered to solve the problem: iterative optimization, look-up-tables (LUTs), simulated annealing and genetic algorithms, and Markov chain Monte Carlo (MCMC). See [156] and references therein.

3) Hybrid inversion: *Hybrid* inversion models combine the previous approaches: they exploit the input-output data generated by RTMs simulations, $\{\mathbf{Y}, \mathbf{x}\}$, and train statistical methods (typically neural networks) to invert the model, i.e., learn $g(\cdot)$. Nonparametric statistical inversion is computationally efficient and can replace more costly physical inversion learning $g(\cdot)$. Regression trees has been used for

example to estimate land surface variables like LAI, fraction of photo-synthetically active radiation (FAPAR), and chlorophyll content [167]. However, the vast majority of hybrid inversion methods consider the use of neural networks [168]–[170] for canopy parameter retrieval, mainly due to their capability to ingest huge databases. Neural network revealed less biased than standard LUT methods in LAI retrieval [171]. In [172], neural networks were trained on a reflectance database made of RTM simulations, and LAI, FAPAR and FCOVER were accurately retrieved as compared with ground measurements. Very recently, the combination of clustering and neural networks inverted simulated data with additive noise. Inclusion of multiangle images improved the LAI estimations. Lately, in [173], neural networks were successfully developed over RTMs to estimate LAI, FCOVER and FAPAR. Only very recently, kernel methods [174] have been used: In [175], the SVR was used to retrieve LAI by inverting PROSAIL.

RTMs have become important tools for the analysis of Earth observation data, providing meaningful links between radiometry and environmental applications. However these models are still often perceived as excessively complicated tools and are not always easily accessible. A recent effort that is worth mentioning is the automated radiative transfer models operator (ARTMO) Matlab toolbox. ARTMO provides all necessary tools for running and inverting a suite of plant RT models, both at the leaf and at the canopy level. The toolbox is freely available[14], and will soon integrate advanced statistical regression for inversion. Currently, the simpleR toolbox[15] provides easy Matlab code to develop parametric and nonparametric retrieval algorithms.

B. EXPERIMENTS

We here illustrate the performance of both empirical and statistical approaches to retrieve chlorophyll concentration from hyperspectral images. The data were obtained in the SPARC-2003 (SPectra bARrax Campaign) and SPARC-2004 campaigns in Barrax, La Mancha, Spain. The region consists of approximately 65% dry land and 35% irrigated land. The methodology applied to obtain the *in situ* leaf-level Chl_{ab} data consisted of measuring samples with a calibrated CCM-200 Chlorophyll Content Meter in the field. Concurrently, we used CHRIS images Mode 1 (62 spectral bands, 34 m spatial resolution at nadir). The images were preprocessed, geometrically and atmospherically corrected. A total of $n = 136$ datapoints in a 62-dimensional space and the measured chlorophyll concentration constitute the database.

Performances of a wide array of established vegetation indices, linear regression with all bands (LR), SVR, and GP [165] were tested. Models were run for a total of 50 random realizations of the training and test data. Averaged correlation coefficients are shown for the test set in Table 4. Nonparametric methods show the best results both in correlation and stability, with GP performing best of the tested methods.

The best GP model was used for prediction on the whole CHRIS image to generate a pixel-by-pixel map of *Chl* and its confidence map (see Fig. 15). The maps show clearly the irrigated crops (the circles in orange-red), the semi-natural areas (light blue) and the bare soil areas (dark blue). Gaussian Processes also provide confidence intervals for the predictions, Fig. 15(b), which may be helpful to identify anomalies. For example, the high confidences (western part of the image) were the fields sampled the most, while low confidence predictions (center of the image) correspond to areas particularly underrepresented in the training data, such as dry barley, harvested barley, and bright bare soils. This product may be used to set sensitivity margins of field instruments quite intuitively: areas are thresholded with error levels above 10% of the total *Chl* range (e.g. >6.5 $\mu g/cm^{-2}$), see Fig. 15(c).

C. CHALLENGES

We reviewed the very active field of physical parameter estimation from acquired images. We presented the main

TABLE 4. CORRELATION COEFFICIENT R RESULTS OF STANDARD NARROWBAND AND BROADBAND INDICES ALONG WITH RECENT NONPARAMETRIC NONLINEAR MODELS.

METHOD	FORMULATION	R
GI	R_{672}/R_{550}	0.52 (0.09)
GVI	$(R_{682}-R_{553})/(R_{682}+R_{553})$	0.66 (0.07)
MCARI2	$1.2[2.5(R_{800}-R_{670})-1.3(R_{800}-R_{550})]$	0.71 (0.12)
mNDVI	$(R_{800}-R_{680})/(R_{800}+R_{680}-2R_{445})$	0.77 (0.12)
mNDVI$_{705}$	$(R_{750}-R_{705})/(R_{750}+R_{705}-2R_{445})$	0.80 (0.07)
mSR$_{705}$	$(R_{750}-R_{445})/(R_{705}+R_{445})$	0.72 (0.07)
mTVI	$1.2[1.2(R_{800}-R_{550})-2.5(R_{670}-R_{550})])$	0.73 (0.07)
NDVI	$(R_{800}-R_{670})/(R_{800}+R_{670})$	0.77 (0.08)
NDVI2	$(R_{750}-R_{705})/(R_{750}+R_{705})$	0.81 (0.06)
NPCI	$(R_{680}-R_{430})/(R_{680}+R_{430})$	0.72 (0.08)
OSAVI	$1.16(R_{800}-R_{670})/(R_{800}+R_{670}+0.16)$	0.79 (0.09)
PRI	$(R_{531}-R_{570})/(R_{531}+R_{570})$	0.77 (0.07)
PRI2	$(R_{570}-R_{539})/(R_{570}+R_{539})$	0.76 (0.07)
PSRI	$(R_{680}-R_{500})/R_{750}$	0.79 (0.08)
RDVI	$(R_{800}-R_{670})/\sqrt{(R_{800}+R_{670})}$	0.76 (0.08)
SIPI	$(R_{800}-R_{445})/(R_{800}-R_{680})$	0.78 (0.08)
SPVI	$0.4[3.7(R_{800}-R_{670})-1.2(R_{530}-R_{670})]$	0.70 (0.08)
SR1	R_{750}/R_{700}	0.74 (0.07)
SR3	R_{750}/R_{550}	0.75 (0.07)
SR4	R_{672}/R_{550}	0.76 (0.10)
SRPI	R_{430}/R_{680}	0.76 (0.09)
TVI	$0.5[120R_{750}-R_{550})-200(R_{670}-R_{550})]$	0.70 (0.10)
VOG	R_{740}/R_{720}	0.76 (0.06)
NAOC	Area in [643,795]	0.79 (0.09)
LR	Least squares linear regression	0.88 (0.06)
SVR	RBF kernel	0.98 (0.03)
KRR	RBF kernel	0.98 (0.04)
GP [165]	Anisotropic RBF kernel	0.99 (0.02)

---

[14]https://sites.google.com/site/jochemverrelst/ARTMO
[15]http://www.uv.es/gcamps/code/simpleR.html

approaches in the literature and introduced the principles and standard terminology. The use and performance of the different existing approaches were illustrated in a real problem of vegetation monitoring which confirmed the excellent results obtained by nonparametric statistical approaches. The field encompasses both physics of land and atmosphere, optimization, and machine learning. The future poses some challenging problems: the community will be confronted to the availability of huge amount of training data coming from RTMs, the design of more sophisticated and realistic RTMs, the combination of both statistical and physically-based models, and the specification of models that can adapt to multitemporal domains.

FIGURE 15. Chlorophyll concentration estimated map (a), predictive standard deviation (b), and masked confidence map (c) generated with GP on the CHRIS 12-07-2003 nadir image.

VII. FAST COMPUTING

In this section, we outline several strategies to accelerate hyperspectral image computations using different kinds of high-performance computing (HPC) architectures. As mentioned before in this contribution, the improved spatial, spectral and temporal resolutions provided by hyperspectral instruments demand fast computing solutions that can a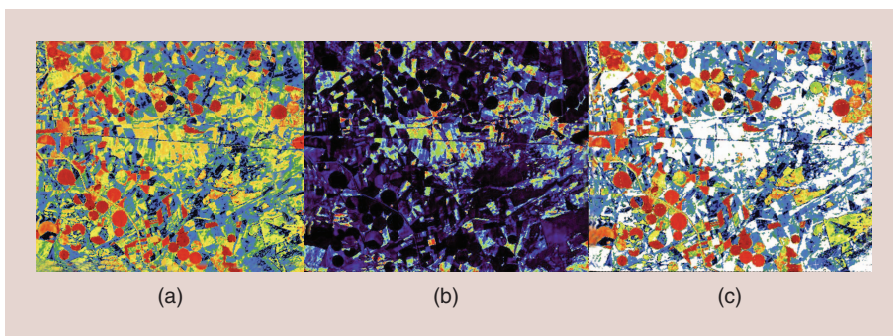ccelerate the efficient exploitation of hyperspectral data sets. Specifically, we describe techniques based on different types of hardware accelerators (see Fig. 16), such as clusters, distributed platforms and specialized devices such as commodity graphics processing units (GPUs) or field programmable gate arrays (FPGAs). The section concludes with a summary of the main challenges in the exploitation of HPC platforms in hyperspectral remote sensing missions.

A. CLUSTERS AND DISTRIBUTED PLATFORMS FOR HYPERSPECTRAL PROCESSING

Clusters were originally developed with the purpose of creating a cost-effective parallel computing system able to satisfy specific computational requirements in different applications. In remote sensing, the need for large amounts of computation was first identified for processing multispectral
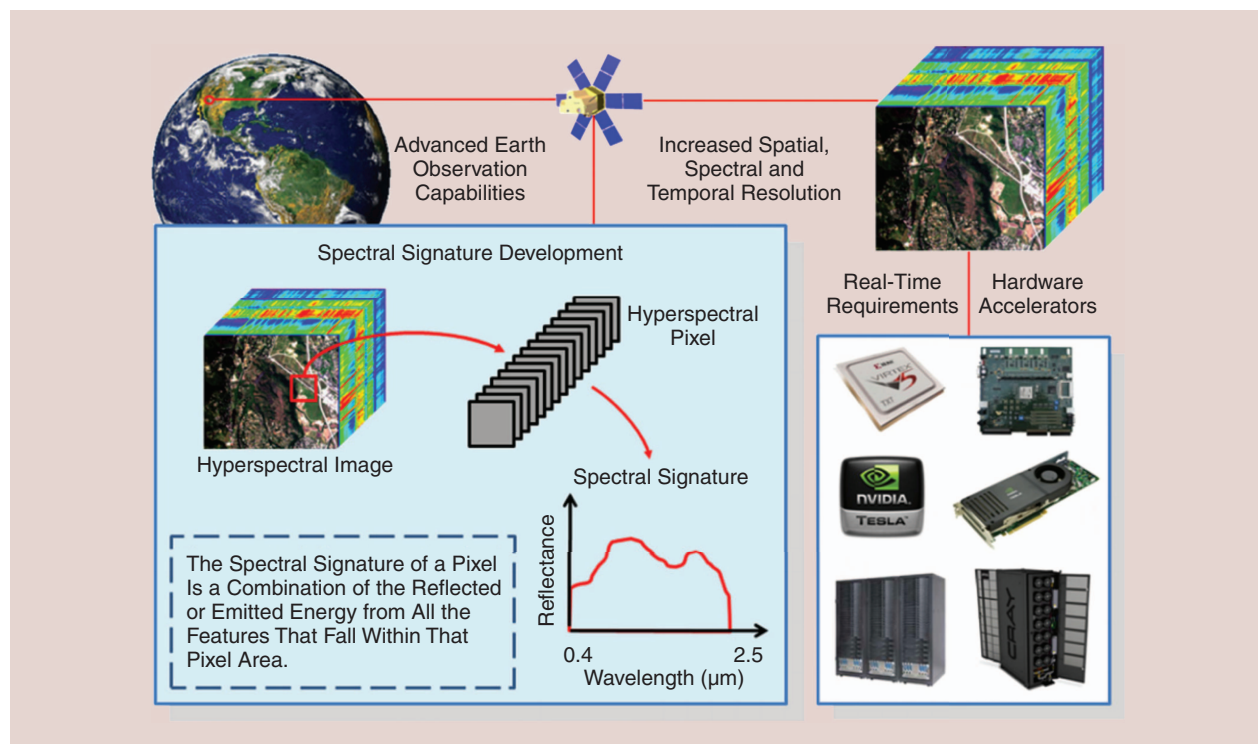

FIGURE 16. Different types of hardware accelerators commonly used to improve computational performance of hyperspectral imaging applications.

imagery with tens of bands. As sensor instruments incorporated hyperspectral capabilities, it was soon recognized that computer mainframes and mini-computers could not provide sufficient power for effectively processing this kind of data. It is worth noting that NASA and ESA are currently supporting massively parallel clusters for remote sensing applications including hyperspectral imaging, such as the Columbia supercomputer[16] at NASA Ames Research Center. Another example of massively parallel computing facility which has been exploited for hyperspectral imaging applications is located at the High Performance Computing Collaboratory (HPC^2) at Mississippi State University[17], which has several supercomputing facilities that have been used in hyperspectral imaging studies.

Homogeneous clusters have already offered access to greatly increased computational power at a low cost (commensurate with falling commercial PC costs) in a number of hyperspectral imaging applications, such as classification or spectral unmixing [15]. However, a recent trend in the design of HPC systems for data-intensive problems, such as those involved in hyperspectral image analysis, is to utilize highly heterogeneous computing resources [176]. In this regard, networks of heterogeneous workstations can realize a very high level of aggregate performance in hyperspectral imaging applications, and the pervasive availability of these resources resulted in the current notions of *grid* and, later, *cloud* computing, which are yet to be fully exploited in hyperspectral imaging problems [177].

Although hyperspectral processing algorithms generally map quite nicely to clusters or networks of CPUs, these systems are generally expensive and difficult to adapt to onboard remote sensing data processing scenarios, in which low-weight and low-power integrated components are essential to reduce mission payload and obtain analysis results in real-time, i.e., at the same time as the data is collected by the sensor. In this regard, the emergence of specialized hardware devices such as FPGAs [178] or GPUs [179] exhibit the potential to bridge the gap towards onboard and real-time analysis of remote sensing data.

B. GPUS FOR HYPERSPECTRAL PROCESSING

In recent years GPUs have evolved into highly parallel, multi-threaded, many-core coprocessors with tremendous computational power, consumption and memory bandwidth [179]. The combined features of general-purpose supercomputing, high parallelism, high memory bandwidth, low cost, compact size, and excellent programmability are now making GPU-based desktop computers an appealing alternative to a massively parallel systems made up of commodity CPUs. The exploding GPU capability has attracted more and more scientists and engineers to use it as a cost-effective high-performance computing platform, including scientists in hyperspectral processing areas. In addition, GPUs can also

significantly increase the computational power of cluster-based and distributed systems (e.g., the fastest supercomputers in the world are now clusters of GPUs[18]).

Several efforts exploiting GPU technology can already be found in the hyperspectral imaging literature [15], [16], [180]. For instance, only in the area of spectral unmixing there have been many developments already. A GPU-based implementation of an automated morphological endmember extraction (AMEE) algorithm for pure spectral signature (endmember) identification is described in [181]. In this case, speedups on the order of 15× were reported. A full spectral unmixing chain comprising the automatic estimation of the number of endmembers, the identification of the endmember signatures, and quantification of endmember fractional abundances has been reported in [182] with speedups superior to 50×. Additional efforts towards real-time and on-board hyperspectral target detection and classification [183], [184] using GPUs have also been recently available. It should be noted that, despite the increasing programmability of low-power GPUs such as those available in smartphones, radiation-tolerance and power consumption issues still prevent the full incorporation of GPUs to spaceborne Earth observation missions.

C. FPGAS FOR HYPERSPECTRAL PROCESSING

An FPGA [178] can be roughly defined as an array of interconnected logic blocks. One of the main advantages of these devices is that both the logic blocks and their interconnections can be (re)configured by their users as many times as needed in order to implement different combinational or sequential logic functions. This characteristic provides FPGAs with the advantages of both software and hardware systems in the sense that FPGAs exhibit more flexibility and shorter development times than application specific integrated circuits (ASICs) but, at the same time, are able to provide much more competent levels of performance, closer to those offered by GPUs (but with much lower power consumption). In fact, the power and energy efficiency of FPGAs has significantly improved during the last decade. FPGA vendors have achieved this goal improving the FPGA architectures, including optimized hardware modules, and taking advantage of the most recent silicon technology. For instance, manufacturing companies such as Xilinx[19] or Altera[20] have reported a 50% reduction in the power consumption when moving from their previous generation of FPGAs. This feature, together with the availability of more FPGAs with increased tolerance to ionizing radiation in space, have consolidated FPGAs as the current standard choice for on-board hyperspectral remote sensing. In the following, we outline several hyperspectral analysis techniques that have been recently implemented in FPGAs.

If we consider the area of spectral unmixing, implementation of endmember extraction algorithms using a Xilinx

---

Virtex-4 FPGA have been recently described in [185]. This FPGA model is similar to radiation-hardened FPGAs certified for space operation. The acceleration factor or speedup of this implementation, compared with a software description developed in C language and executed on a PC with AMD Athlon 2.6 GHz processor and 512 Mb of RAM, is 37x for the well-known AVIRIS Cuprite scene (16 endmembers)[21], 38x for a hyperspectral image collected also in the Cuprite mining district by EO-1 Hyperion (21 endmembers), and 37x for an AVIRIS image collected over the Jasper Ridge biological preserve in California (19 endmembers). The speedup factor is quite constant across all the images, even taking into account the differences in the number of endmembers. Similarly FPGA implementations of abundance estimation algorithms have also been described in [186]. This implementation was tested in the same FPGA used in [185], and achieved a speedup factor of 10x when processing the AVIRIS Cuprite scene and over 12× when it comes to the AVIRIS Jasper Ridge scene. Authors also reach the conclusion that, using FPGAs, the execution time scales linearly with the size of the image. FPGA implementations of other classic unmixing algorithms have also been discussed in chapter 2 of [180].

Other areas in which FPGA implementations have been particularly relevant is target detection and classification. In this context, [180] discusses several examples. Specifically, chapter 15 in [180] discusses the use of FPGAs in detection applications and provides specific application case studies. Chapter 16 in [180] describes FPGA implementations of techniques for hyperspectral target detection applications. Chapter 17 in [180] describes an on-board real-time processing technique for fast and accurate target detection and discrimination in hyperspectral data. Real-time implementations of several popular detection and classification algorithms for hyperspectral imagery have also been discussed in [187].

D. CHALLENGES

Despite the individual success of the different types of HPC architectures described in this section in different problems, a key aspect still missing is the integration of such systems in complementary fashion. Although the role of each type of accelerator depends heavily on the considered application domain, cluster-based systems seem particularly appropriate for efficient information extraction from very large data archives comprising data sets already transmitted to Earth, while the time-critical constraints introduced by many remote sensing applications call for on-board and, often, real-time processing developments which require the use of specialized hardware architectures such as GPUs and FPGAs. What is still missing is an infrastructure in which these computing resources are available on-demand, possibly from a distributed cloud resource that can support coupled HPC codes with strict processing deadlines. Clearly,

[21]http://aviris.jpl.nasa.gov/data/free data.html

such a grand challenge system could support a wider variety of hyperspectral imaging applications.

VIII. CONCLUDING REMARKS

Among the remote sensing modalities, the role of the hyperspectral technology in the detection and identification of materials, determination of physical parameters, and change detection cannot be overstated. A few signs of the crescent importance of the hyperspectral remote sensing technology are the increasing number of hyperspectral spaceborne and airborne sensors and applications supported on hyperspectral remote sensing data. Another sign is the growing number of scientific publications, which has reached a figure comparable to that of radar remote sensing, as documented in the Introduction.

> **THE ADOPTION OF NEW STATISTICAL METHODS IN HYPERSPECTRAL IMAGE ANALYSIS IS A FACT TODAY. NEW EXCITING ADVANCES IN THE THEORY AND APPLICATIONS ARE YET TO COME.**

Owing to several factors among which we refer to the high dimensionality and size of the hyperspectral data, the spectral mixing (linear and nonlinear), and the degradation mechanisms associated to the measurement process such as noise and atmospheric effects, the extraction of information from hyperspectral remote sense data relies on sophisticated and complex data analysis methods. In this paper, we presented a tour over a number of representative and attractive hyperspectral data analysis methods and algorithms, organized in six main topics: data fusion, unmixing, classification, target detection, physical parameter retrieval, and fast computing. In all topics, we described the state-of-the-art, provided illustrative examples, and pointed to future challenges and research directions.

As the reader has noted, the remote sensing data analysis chain is very broad. For this reason, we could not cover all the interesting and relevant aspects exhaustively. For example, we do not cover the important field of change detection and multitemporal classification, which are very active areas of research. Also note that the relevant field of image compression is missing, while the crucial steps of atmospheric and geometric corrections, co-registration, deconvolution, or image restoration and quality assessment have been treated only superficially.

As it can be concluded from the material presented in the paper, hyperspectral remote sensing data analysis is a multidisciplinary area using and adapting frontier concepts, frameworks, and algorithms from the field of signal and image processing, statistical inference, and machine learning. Note that the methods used in hyperspectral remote sensing data analysis are not always simple adaptations of well-known methods developed in the above fields. For example, hyperspectral unmixing has provided a unique

problem scenario to the development of new blind source separation solutions that are not interpretable as particularizations of known solutions. The same is true, in different degrees, in the remaining addressed topics.

As a conclusion, hyperspectral remote sensing data analysis is a mature field in the intersection of signal and image processing, statistical inference, and machine learning, contributing actively with frontier cross-disciplinary research activities. We hope that this paper be useful for researchers working in the field and foster curiosity in postgraduate students looking for a PhD theme and in researchers looking for a new research area.

REFERENCES

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

[2] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York; Berlin, Germany; Heidelberg, Germany: Springer-Verlag, 2006.

[3] 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. Select. Topics Appl. Earth Observ. Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.

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

[5] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *IEEE J. Select. Topics Signal Processing*, vol. 5, no. 3, pp. 534–542, 2011.

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

[7] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, June 2009.

[8] J. Keef and T. U. of Arizona. *Optical Sciences*, Hyper-spectral Sensor Calibration Extrapolated from Multi-spectral Measurements. The University of Arizona, 2008.

[9] W. Moses, "Atmospheric correction of hyperspectral data: Execution of and comparison between FLAASH and TAFKAA_6S," Cornell Univ., Ithaca, New York, Aug. 2005.

[10] W. Verhoef and H. Bach, "Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models," *Remote Sens. Environ.*, vol. 87, no. 1, pp. 23–41, 2003.

[11] K. Staenz, A. Mueller, A. Held, and U. Heiden, "Technical committees corner: International spaceborne imaging spectroscopy (ISIS) technical committee," *IEEE Geosci. Remote Sensing Newsletter*, no. 165, pp. 38–42, 2012.

[12] A. Plaza, J. Plaza, A. Paz, and S. Sánchez, "Parallel hyperspectral image and signal processing," *IEEE Signal Processing Mag.*, vol. 28, no. 3, pp. 119–126, 2011.

[13] S. López, T. Vladimirova, C. Gónzalez, J. Resano, D. Mozos, and A. Plaza, "The promise of reconfigurable computing for hyperspectral imaging on-board systems: Review and trends," *Proc. IEEE*, to be published.

[14] R. Trautner. (2011). ESA's roadmap for next generation payload data processors. presented at DASIA Conf. [Online]. *1*. Available: http://www.esa.int/TEC/OBDP/

[15] Special Issue on High Performance Computing for Hyperspectral Imaging, *Int. J. High Perform. Comput.*, vol. 4, no. 3, pp. 528–544, 2011.

[16] Special Issue on Architectures and Techniques for Real-Time Processing of Remotely Sensed Images, *J. Real-Time Image Processing*, vol. 4, no. 3, pp. 191–193, 2009.

[17] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 10 Pt 1, pp. 3660–3677, 2012.

[18] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 10 Pt 1, pp. 3717–3724, 2012.

[19] R. Mendez-Rial, M. Calvino-Cancela, and J. Martin-Herrero, "Accurate implementation of anisotropic diffusion in the hypercube," *IEEE Geosci. Remote Sensing Lett.*, vol. 7, no. 4, pp. 870–874, Oct. 2010.

[20] N. Jacobson and M. Gupta, "Design goals and solutions for display of hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 11, pp. 2684–2692, Nov. 2005.

[21] Z. Mahmood and P. Scheunders, "Enhanced visualization of hyperspectral images," *IEEE Geosci. Remote Sensing Lett.*, vol. 8, no. 5, pp. 869–873, 2011.

[22] M. Mignotte, "A bicriteria optimization approach based dimensionality reduction model for the color display of hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 2, pp. 501–513, 2012.

[23] K. Kotwal and S. Chaudhuri, "An optimization-based approach to fusion of hyperspectral images," (special issue), *IEEE J. Select. Topics Appl. Earth Observ. Remote Sensing*, vol. 5, no. 2, pp. 501–509, 2012.

[24] T. Akgun, Y. Altunbasak, and R. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1860–1875, 2005.

[25] 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 Sensing*, vol. 48, no. 6, pp. 2569–2579, 2010.

[26] H. Zhang, L. Zhang, and H. Shen, "A super-resolution reconstruction algorithm for hyperspectral images," *Signal Processing*, vol. 92, no. 9, pp. 2082–2096, 2012.

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

[28] K. C. Mertens, B. De Baets, L. P. C. Verbeke, and R. R. De Wulf, "A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models," *Int. J. Remote Sens.*, vol. 27, no. 15, pp. 3293–3310, 2006.

[29] Y. Gu, Y. Zhang, and J. Zhang, "Integration of spatial-spectral information for resolution enhancement in hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 5, pp. 1347–1358, 2008.

[30] F. A. Mianji, Y. Gu, Y. Zhang, and J. Zhang, "Enhanced self-training superresolution mapping technique for hyperspectral imagery," *IEEE Geosci. Remote Sensing Letters*, vol. 8, no. 4, pp. 671–675, 2011.

[31] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sensing*, vol. 45, no. 10, pp. 3012–3021, 2007.

[32] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 47, no. 11, pp. 3834–3843, 2009.

[33] Y. Zhang, A. Duijster, and P. Scheunders, "A Bayesian restoration approach for hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 9, pp. 3453–3462, 2012.

[34] R. Zurita-Milla, J. G. P. W. Clevers, and M. E. Schdepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote Sensing Lett.*, vol. 5, no. 3, pp. 453–457, 2008.

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

[36] M. Elbakary and M. Alam, "Superresolution construction of multispectral imagery based on local enhancement," *IEEE Geosci. Remote Sensing Lett.*, vol. 5, no. 2, pp. 276–279, 2008.

[37] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 5, pp. 1301–1312, 2008.

[38] K. Kotwal and S. Chaudhuri, "A novel approach to quantitative evaluation of hyperspectral image fusion techniques," *Inform. Fusion*, vol. 14, no. 1, pp. 5–18, 2013.

[39] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 44–57, 2002.

[40] B. Hapke, "Bidirection reflectance spectroscopy. I. Theory," *J. Geophys. Res.*, vol. 86, pp. 3039–3054, 1981.

[41] M. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 542–552, 1994.

[42] J. Nascimento and J. Bioucas-Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 1, pp. 175–187, 2005.

[43] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Ann. JPL Airborne Geoscience Workshop*, 1993, vol. 1, pp. 11–14.

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

[45] C.-I. Chang, C.-C. Wu, W. Liu, and Y.-C. Ouyang, "A new growing method for simplex-based endmember extraction algorithm," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 10, pp. 2804–2819, 2006.

[46] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 11, 2011.

[47] N. Gillis and S. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," arXiv preprint arXiv:1208.1237, 2012.

[48] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end member determination in hyperspectral data," in *Proc. SPIE Image Spectrometry V*, 1999, vol. 3753, pp. 266–277.

[49] R. A. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, "Automatic end member extraction from hyperspectral data for mineral exploration," in *Proc. Canadian Symp. Remote Sensing*, 1999, pp. 21–24.

[50] 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 Sensing*, vol. 42, no. 10, pp. 2085–2095, 2004.

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

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

[53] J. Bioucas-Dias, "A variable splitting augmented lagragian approach to linear spectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Processing: Evolution Remote Sensing (WHISPERS)*, 2009, pp. 1–4.

[54] T. Chan, C. Chi, Y., Huang, and W. Ma, "Convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Processing*, vol. 57, no. 11, pp. 4418–4432, 2009.

[55] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Process.*, vol. 89, no. 12, pp. 2657–2669, Dec. 2009.

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

[57] R. Basedow, D. Carmer, and M. Anderson, "Hydice system: Implementation and performance," in *Proc. SPIE's 1995 Symp. OE/Aerospace Sensing and Dual Use Photonics*, pp. 258–267.

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

[59] J. Mustard and C. Pieters, "Quantitative abundance estimates from bidirectional reflectance measurements," *J. Geophys. Res.*, vol. 92, pp. E617–E626, 1987.

[60] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sensing*, no. 11, pp. 4153–4162, Nov. 2011.

[61] J. Broadwater, A. Banerjee, and P. Burlina, "Kernel methods for unmixing hyperspectral imagery," in *Optical Remote Sensing Advances in Signal Processing and Exploitation*, L. B. S. Prasad and E. J. Chanussot, Eds. Springer, 2011, pp. 247–269.

[62] A. Plaza, G. Martin, J. Plaza, M. Zortea, and S. Sánchez, "Recent developments in spectral unmixing and endmember extraction," in *Optical Remote Sensing*, S. Prasad, L. M. Bruce, and J. Chanussot, Eds. Berlin, Germany: Springer-Verlag, 2011, ch. 12, pp. 235–267.

[63] 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 Processing*, vol. 21, no. 6, pp. 3017–3025, 2012.

[64] R. Heylen and P. Scheunders, "Calculation of geodesic distances in nonlinear mixing models: Application to the generalized bilinear model," *IEEE Geosci. Remote Sensing Lett.*, vol. 9, no. 4, pp. 644–648, 2012.

[65] R. Close, "Endmember and proportion estimation using physics-based macroscopic and microscopic mixture models," Ph.D. dissertation, Univ. Florida, CA, Dec. 2011.

[66] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[67] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[68] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *1993 Conference Record 27th Asilomar Conf. Signals, Systems and Computers*, pp. 40–44.

[69] E. Candès and J. J. Romberg, "Sparsity and incoherence in compressive sampling," *IEEE Trans. Image Processing*, vol. 23, pp. 969–985, 2007.

[70] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Processing: Evolution Remote Sensing (WHISPERS)*, 2010, vol. 1, pp. 1–4.

[71] J. Mairal, R. Jenatton, G. Obozinski, and F. Bach, "Convex and network flow optimization for structured sparsity," *J. Mach. Learn. Res.*, vol. 12, pp. 2681–2720, 2011.

[72] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.

[73] J. Tropp, A. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximation. Part I. Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.

[74] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," submitted for publication.

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

[76] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, J. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, 2009.

[77] 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 Sensing*, vol. 26, pp. 65–74, 1988.

[78] X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote sensing image display and classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 538–542, 1999.

[79] L. Bruce and J. Li, "Wavelets for computationally efficient hyperspectral derivative analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 1540–1546, 2001.

[80] 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 Sensing*, vol. 51, pp. 184–198, 2013.

[81] S. Kumar, J. Ghosh, and M. M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 1368–1379, 2001.

[82] X. Jia and J. A. Richards, "Progressive two-class decision classifier for optimization class discriminations," *Remote Sens. Environ.*, vol. 63, pp. 289–297, 1998.

[83] G. Camps-Valls, J. Mooij, and B. Schölkopf, "Remote sensing feature selection by kernel dependence estimation," *IEEE Trans. Geosci. Remote Sensing*, vol. 7, pp. 587–591, 2010.

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

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

[86] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, pp. 1351–1362, 2005.

[87] B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA: MIT Press, 2002.

[88] G. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, 2004.

[89] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Statist. Math.*, vol. 44, pp. 197–200, 1992.

[90] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, pp. 4085–4098, 2010.

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

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

[93] M. Pesaresi and J. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001.

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

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

[96] M. Dalla Mura, J. Atli Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sensing*, vol. 31, no. 22, pp. 5975–5991, 2010.

[97] G. Camps-Valls, L. Goméz-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sensing Lett.*, vol. 3, pp. 93–97, 2006.

[98] 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 Sensing*, to be published.

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

[100] Y. Tarabalka, J. Chanussot, and J. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, pp. 2367–2379, 2010.

[101] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 11, pp. 4122–4132, 2011.

[102] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, to be published.

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

[104] 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 Sensing*, vol. 49, no. 11, pp. 4263–4281, Dec. 2011.

[105] F. Bovolo, L. Bruzzone, and L. Carlin, "A novel technique for subpixel image classification based on support vector machine," *IEEE Trans. Image Processing*, vol. 19, pp. 2983–2999, 2010.

[106] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for the semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 11, pp. 3363–3373, 2006.

[107] G. Camps-Valls, T. Bandos, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 45, pp. 3044–3054, Oct. 2007.

[108] S. Velasco-Forero and V. Manian, "Improving hyperspectral image classification using spatial preprocessing," *IEEE Geosci. Remote Sensing Lett.*, vol. 6, pp. 297–301, 2009.

[109] D. Tuia and G. Camps-Valls, "Semisupervised remote sensing image classification with cluster kernels," *IEEE Geosci. Remote Sensing Lett.*, vol. 6, no. 2, pp. 224–228, Apr. 2009.

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

[111] J. Muñoz Marí, F. Bovolo, L. Gómez-Chova, L. Bruzzone, and G. Camp-Valls, "Semisupervised one-class support vector machines for classification of remote sensing data," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 8, pp. 3188–3197, 2010.

[112] L. Gómez-Chova, G. Camps-Valls, L. Bruzzone, and J. Calpe-Maravilla, "Mean MAP kernel methods for semisupervised cloud classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 1, pp. 207–220, 2010.

[113] D. Tuia and G. Camps-Valls, "Urban image classification with semisupervised multiscale cluster kernels," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sensing*, vol. 4, no. 1, pp. 65–74, Mar. 2011.

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

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

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

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

[118] S. Patra and L. Bruzzone, "A batch-mode active learning technique based on multiple uncertainty for SVM classifier," *IEEE Geosci. Remote Sensing Lett.*, vol. 9, no. 3, pp. 497–501, May 2012.

[119] I. Dopido, J. Li, P. Marpu, A. Plaza, J. Bioucas-Dias, and J. A. Benediktsson, "Semi-supervised self-learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, to be published.

[120] L. Ma, M. Crawford, and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 48, no. 11, pp. 4099–4109, 2010.

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

[122] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[123] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerospace Elect. Systems Mag.*, vol. 25, no. 7, pp. 5–27, July 2010.

[124] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Mag.*, vol. 19, pp. 58–69, 2002.

[125] A. P. Schaum and A. D. Stocker, "Subclutter target detection using sequences of thermal infrared multispectral imagery," in *Proc. SPIE*, 1997, vol. 3071, pp. 12–22.

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

[127] G. Healey and D. Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 6, pp. 2706–2717, 1999.

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

[129] L. Scharf and B. Friedlander, "Matched subspace detectors," *Signal Process.*, vol. 42, no. 8, pp. 2146–2157, 1994.

[130] F. Robey, D. Fuhrmann, and E. Kelly, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, 1992.

[131] S. Kraut, L. Scharf, and L. McWhorter, "Adaptive subspace detectors," *Signal Process.*, vol. 49, no. 1, pp. 208–216, 2001.

[132] B. Schölkopf and A. J. Smola, *Learning with Kernels*. Cambridge, MA: MIT Press, 2002.

[133] H. Kwon and N. Nasrabadi, "A comparative analysis of kernel subspace target detectors for hyperspectral imagery," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 193–193, 2007.

[134] S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, 1990.

[135] C. Willis, "Comparison of anomaly detection methods for hyperspectral imagery," in *Proc. SPIE*, Orlando, FL, 2005, vol. 5988, pp. B1–B12.

[136] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ: Wiley, 2007.

[137] A. Stocker and A. Schaum, "Application of stochastic mixing models to the hyperspectral detection problems," in *Proc. SPIE*, Orlando, FL, 1997, vol. 3071, pp. 47–60.

[138] H. Kwon, S. Der, and N. Nasrabadi, "Adaptive anomaly detection using subspace separation for hyperspectral images," *Opt. Eng.*, vol. 42, no. 11, pp. 3342–3351, 2003.

[139] K. Ranney and M. Soumekh, "Hyperspectral anomaly detection within the signal subspace," *IEEE Trans. Geosci. Remote Sensing Lett.*, vol. 3, no. 3, pp. 312–316, 2006.

[140] D. Stein, "Stochastic compositional models applied to subpixel analysis of hyperspectral imagery," in *Proc. SPIE*, 2001, vol. 4480, pp. 49–56.

[141] E. Ashton, "Detection of subpixel anomalies in multispectral infrared imagery using an adaptive bayesian classifier," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 2, pp. 506–517, 1998.

[142] S. Schweizer and J. Moura, "Hperspectral imagery: Clutter adaptation in anomaly detection," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1855–1871, 2000.

[143] M. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 2, pp. 374–387, 2005.

[144] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 8, pp. 2282–2291, 2006.

[145] H. Goldberg and N. Nasrabadi, "A comparative study of linear and nonlinear anomaly detectors for hyperspectral imagery," in *Proc. SPIE*, Orlando, FL, 2007, vol. 6497.

[146] A. Schaum, "Joint subspace detection of hyperspectral targets," in *Proc. IEEE Aerospace Conf.*, 2004, vol. 3.

[147] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

[148] L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.

[149] D. Gu, A. Gillespie, A. Kahle, and F. Palluconi, "Autonomous atmospheric compensation (AAC) of high resolution hyperspectral thermal infrared remote-sensing imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 11, pp. 2557–2570, 2000.

[150] P. Villeneuve, H. Fry, J. Theiler, and W. Clodius, "Improved matched-filter detection techniques," in *Proc. SPIE*, Orlando, FL, 1999, vol. 3753.

[151] N. Nasrabadi, "Regularized spectral matched filter for target recognition in hyperspectral imagery," *IEEE Trans. Signal Processing Lett.*, vol. 15, pp. 317–320, 2008.

[152] Y. Chen, N. Nasrabadi, and T. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Select. Topics Signal Processing*, vol. 5, no. 3, pp. 629–640, 2011.

[153] Y. Chen, N. Nasrabadi, and T. Tran, "Simultaneous joint sparsity model for target detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing Lett.*, vol. 8, no. 4, pp. 676–680, 2011.

[154] C. Chen, N. Nasrabadi, and T. Tran, "Kernel sparse representation for hyperspectral target detection," in *Proc. SPIE*, Baltimore, MD, 2012, vol. 8390.

[155] C. Rodgers, *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Singapore: World Scientific, 2000.

[156] S. Liang, *Quantitative Remote Sensing of Land Surfaces*. Hoboken, NJ: Wiley, 2004.

[157] F. Baret and S. Buis, "Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems," in *Advances in Land Remote Sensing: System, Modeling, Inversion and Applications*. Germany: Springer-Verlag, 2008.

[158] J. Berni, P. Zarco-Tejada, L. Suárez, and E. Fereres, "Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle," *IEEE Trans. Geosci. Remote Sensing*, vol. 47, no. 3, pp. 722–738, 2009.

[159] J. Verrelst, M. Schaepman, B. Koetz, and M. Kneubhler, "Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2341–2353, 2008.

[160] M. Schaepman, S. Ustin, A. Plaza, T. Painter, J. Verrelst, and S. Liang, "Earth system science related imaging spectroscopy: An assessment," *Remote Sens. Environ.*, vol. 113, no. 1, pp. S123–S137, 2009.

[161] L. Sarker and J. Nichol, "Improved forest biomass estimates using ALOS AVNIR-2 texture indices," *Remote Sens. Environ.*, vol. 115, no. 4, pp. 968–977, 2011.

[162] P. Townsend, J. Foster, R. J. Chastain, and W. Currie, "Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains using Hyperion and AVIRIS," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 6, pp. 1347–1354, June 2003.

[163] J. Arenas-García and G. Camps-Valls, "Efficient kernel orthonormalized PLS for remote sensing applications," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 10, pp. 2872–2881, Oct. 2008.

[164] G. Camps-Valls, L. Bruzzone, J. L. Rojo-Álvarez, and F. Melgani, "Robust support vector regression for biophysical variable estimation from remotely sensed images," *IEEE Geosci. Remote Sensing Lett.*, vol. 3, no. 3, pp. 339–343, 2006.

[165] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, "Retrieval of canopy parameters using Gaussian processes techniques," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, 2011.

[166] S. Jacquemoud, C. Bacour, H. Poilvé, and J.-P. Frangi, "Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode," *Remote Sens. Environ.*, vol. 74, no. 3, pp. 471–481, 2000.

[167] M. Weiss and F. Baret, "Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data," *Remote Sens. Environ.*, vol. 70, pp. 293–306, 19991.

[168] F. Baret, J. Clevers, and M. Steven, "The robustness of canopy gap fraction estimates from red and near-infrared reflectance: A comparison of approaches," *Remote Sens. Environ.*, vol. 54, pp. 141–151, 1995.

[169] S. Gopal and C. Woodcock, "Remote sensing of forest change using artificial neural networks," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 398–404, 1996.

[170] F. Baret and T. Fourty, "Estimation of leaf water content and specific leaf weight from reflectance and transmittance measurements," *Agronomie*, vol. 17, nos. 9–10, pp. 455–464, 1997.

[171] H. Fang and S. Liang, "A hybrid inversion method for mapping leaf area index from MODIS data: Experiments and application to broadleaf and needleleaf canopies," *Remote Sens. Environ.*, vol. 94, no. 3, pp. 405–424, 2005.

[172] C. Bacour, F. Baret, D. Béal, M. Weiss, and K. Pavageau, "Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation," *Remote Sens. Environ.*, vol. 105, no. 4, pp. 313–325, 2006.

[173] A. Verger, F. Baret, and F. Camacho, "Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations," *Remote Sens. Environ.*, vol. 115, no. 2, pp. 415–426, 2011.

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

[175] S. Durbha, R. King, and N. Younan, "Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer," *Remote Sens. Environ.*, vol. 107, nos. 1–2, pp. 348–361, 2007.

[176] A. Plaza, D. Valencia, and J. Plaza, "An experimental comparison of parallel algorithms for hyperspectral analysis using homogeneous and heterogeneous networks of workstations," *Parallel Comput.*, vol. 34, no. 2, pp. 92–114, 2008.

[177] J. Brazile, R. A. Neville, K. Staenz, D. Schlaepfer, L. Sun, and K. I. Itten, "Cluster versus grid for operation generation of ATCOR's MODTRAN-based look up table," *Parallel Compt.*, vol. 34, pp. 32–46, 2008.

[178] S. Hauck, "The roles of FPGAs in reprogrammable systems," *Proc. IEEE*, vol. 86, no. 4, pp. 615–639, 1998.

[179] J. Nickolls and W. J. Dally, "The GPU computing era," *IEEE Micro*, vol. 30, pp. 56–69, 2010.

[180] A. Plaza and C.-I. Chang, *High Performance Computing in Remote Sensing*. New York: Taylor & Francis, 2007.

[181] J. Setoain, M. Prieto, C. Tenllado, A. Plaza, and F. Tirado, "Parallel morphological endmember extraction using commodity graphics hardware," *IEEE Geosci. Remote Sensing Lett.*, vol. 43, no. 3, pp. 441–445, 2007.

[182] S. Sánchez, A. Paz, G. Martin, and A. Plaza, "Parallel unmixing of remotely sensed hyperspectral images on commodity graphics processing units," *Concurrency Comput. Pract. Exp.*, vol. 23, no. 13, pp. 1538–1557, 2011.

[183] Y. Tarabalka, T. V. Haavardsholm, I. Kasen, and T. Skauli, "Real-time anomaly detection in hyperspectral images using multivariate normal mixture models and GPU processing," *J. Real-Time Image Process.*, vol. 4, pp. 1–14, 2009.

[184] 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 Sensing Lett.*, vol. 10, no. 2, pp. 221–225, 2013.

[185] C. González, D. Mozos, J. Resano, and A. Plaza, "FPGA implementation of the N-FINDR algorithm for remotely sensed hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 2, pp. 374–374, 2012.

[186] C. Gonzalez, J. Resano, A. Plaza, and D. Mozos, "FPGA implementation of abundance estimation for spectral unmixing of hyperspectral data using the image space reconstruction algorithm," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sensing*, vol. 5, no. 1, pp. 248–261, 2012.

[187] Q. Du and R. Nekovei, "Fast real-time onboard processing of hyperspectral imagery for detection and classification," *J. Real-Time Image Process.*, vol. 4, no. 3, pp. 273–286, 2009.

GRS