

# Multisource and Multitemporal Data Fusion in Remote Sensing

*A comprehensive review of the state of the art*

©ISTOCKPHOTO.COM/PORBYTOV

The recent, sharp increase in the availability of data captured by different sensors, combined with their considerable heterogeneity, poses a serious challenge for the effective and efficient processing of remotely sensed data. Such an increase in remote sensing and ancillary data sets, however, opens up the possibility of utilizing multimodal data sets in a joint manner to further improve the performance of the processing approaches with respect to applications at hand. Multi-

source data fusion has, therefore, received enormous attention from researchers worldwide for a wide variety of applications.

Moreover, thanks to the revisit capability of several spaceborne sensors, the integration of the temporal information with the spatial and/or spectral/backscatter information of the remotely sensed data is possible, helping to move from a two-dimensional (2D) or three-dimensional (3D) data representation to four-dimensional

**PEDRAM GHAMISI, BEHNOOD RASTI, NAOTO YOKOYA, QUNMING WANG,  
BERNHARD HÖFLE, LORENZO BRUZZONE, FRANCESCA BOVOLO,  
MINGMIN CHI, KATHARINA ANDERS, RICHARD GLOAGUEN,  
PETER M. ATKINSON, AND JÓN ATLI BENEDIKTSSON**

Digital Object Identifier 10.1109/MGRS.2018.2890023  
Date of publication: 19 March 2019

(4D) data structures, where the time variable adds new information—and challenges—for information extraction algorithms. There are a huge number of research works dedicated to multisource and multitemporal data fusion, but the methods for the fusion of different modalities have expanded according to the different paths taken by each research community.

This article brings together the advances of multisource and multitemporal data fusion approaches with respect to the various research communities and provides a thorough and discipline-specific starting point for researchers at different levels (i.e., students, researchers, and senior researchers) willing to conduct novel investigations on this challenging topic by supplying sufficient detail and references. More specifically, this work provides a bird's-eye view of many important contributions specifically dedicated to the topics of pansharpening and resolution enhancement, point cloud data fusion, hyperspectral and lidar data fusion, multitemporal data fusion, and big data and social media. In addition, the main challenges and possible future research in each area are outlined and discussed.

### A DELUGE OF DATA

The number of data produced by sensing devices has increased exponentially in the last few decades, creating the big data phenomenon and leading to the creation of the new field of data science, including the popularization of machine-learning and deep-learning algorithms to deal with such data [1]–[3]. In the field of remote sensing, the number of platforms for producing remotely sensed data has similarly increased, with an ever-growing number of satellites in orbit or planned for launch and new platforms for proximate sensing, such as unmanned aerial vehicles (UAVs), producing extremely fine spatial resolution data. While optical-sensing capabilities have increased in quality and volume, the number of alternative modes of measurement has also grown, most notably including airborne lidar and terrestrial laser scanning (TLS), which produce point clouds representing elevation, as opposed to generating images [4].

The number of synthetic aperture radar (SAR) sensors, which measure radar backscatter, and satellite and airborne hyperspectral sensors, which extend optical-sensing capabilities by measuring in a larger number of wavebands, has also increased greatly [5], [6]. Airborne and spaceborne geophysical measurements, such as those from the Gravity Recovery and Climate Experiment satellite mission and airborne electromagnetic surveys, are currently also being considered. In addition, there has been great interest in new sources of ancillary data, from social media, crowdsourcing, and scraping the Internet [7]–[9], for example. These data have very different modalities compared to remote sensing data but may be related to the subject of interest and, therefore, may add useful information relevant to specific problems.

The remote sensors onboard the previously mentioned platforms may vary greatly in multiple dimensions, as in the types of properties sensed and the spatial and spectral

resolution of the data. This is true even for sensors housed on the same platform (e.g., the many kinds of multispectral and panchromatic sensors) or that are part of the same satellite configuration [e.g., the European Space Agency's series of Medium Resolution Imaging Spectrometer (MERIS) sensors]. The rapid expansion in the number and availability of data from widely differing sources creates serious challenges for their effective and efficient processing [10]. For a particular remote sensing application, there are likely to be multiple remote sensing and ancillary data sets pertaining to the problem; this creates a dilemma as to how best to combine the data sets for maximum utility. It is for this reason that multisource data fusion, in the context of remote sensing, has received so much attention in recent years [10]–[13].

Fortunately, the increase in the number and heterogeneity of data sources (presenting both challenges and opportunities) has been paralleled by increases in computing power; by efforts to make data more open, available, and interoperable; and by advances in methods for data fusion, which are reviewed here [15]. There exists a wide range of approaches to data fusion; see, for example, [11]–[13]. This article seeks to review them by class of data modality (e.g., optical, SAR, or laser scanning), because methods for these modalities have developed somewhat differently, according to each research community. Given this diversity, it is challenging to synthesize multisource data fusion approaches into a single framework, and that is not the goal here. Nevertheless, a general framework for measurement and sampling processes (i.e., forward processes) is described briefly to provide greater illumination of the various data fusion approaches (i.e., commonly inverse processes or with elements of inverse processing) that are overviewed in the following sections. Because the topic of multisensor data fusion is extremely broad and because specific aspects have already been reviewed, we have to restrict what is covered in the manuscript and, therefore, do not address a few topics, such as the fusion of SAR and optical data.

MANY APPLICATIONS CAN  
BENEFIT FROM FUSED FINE-  
RESOLUTION TIME-SERIES  
DATA SETS, PARTICULARLY  
THOSE THAT INVOLVE  
SEASONAL OR RAPID  
CHANGES.

### A GENERAL FRAMEWORK FOR DATA FUSION

We start by defining the space and properties of interest. Historically in remote sensing, four dimensions have been considered as providing information: spatial, temporal, spectral, and radiometric [that is, 2D spatially, one-dimensional (1D) temporally, and 1D spectrally, with *radiometric* referring to numerical precision]. The electromagnetic spectrum (EMS) exists as a continuum and, thus, lends itself to high-dimensional feature space exploration through the definition of multiple wavebands (the spectral dimension). In contrast to most optical and SAR sensors, lidar and TLS

measure a surface in three spatial dimensions. Recent developments in photo- and radargrammetry, such as Structure From Motion and interferometric SAR, have increased the availability of 3D data. This expansion of the dimensionality of interest to three dimensions in space and one dimension in time makes image and data fusion additionally challenging [4]. The properties measured in each case vary, with SAR measuring backscatter, optical sensors (including hyperspectral ones) measuring the visible and infrared parts of the EMS, and laser scanners measuring 3D surface elevation. Only surface elevation is likely to be a primary interest, whereas reflectance and backscatter are probably only indirectly related to the property of interest.

We next define measurement processes. A common physical model in remote sensing is one of four component models: scene model, atmosphere model, sensor model, and image model [16]–[21]. The scene model defines the subject of interest (e.g., land cover or topographic surface), while the atmosphere model is a transform of the EMS from surface to sensor. The sensor model represents a measurement process [e.g., involving a signal-to-noise ratio (SNR) or the point spread function], and the image model is a sampling process (e.g., to create the data as an image of pixels on a regular grid).

Third, we note that the sampling process implied by the image model can be expanded and generalized to three key

parameters (the sampling extent, sampling scheme, and sampling support), each of which has four further parameters (size, geometry, orientation, and position). The support is a key sampling parameter that defines the space on which each observation is made; it is most directly related to the point spread function in remote sensing and is

represented as an image pixel [22]. The combination and arrangement of pixels as an image defines the spatial resolution of the image. Fusion approaches are often concerned with the combination of two or more data sets with different spatial resolutions, such as to create a unified data set at the finest resolution [23]–[25]. Figure 1(a) demonstrates schematically the multiscale nature (different spatial resolutions) of diverse data sets captured by spaceborne, airborne, and UAV sensors. In principle, there is a relation between spatial resolution and scene coverage, i.e., data with a coarser spatial resolution (spaceborne data) have a larger scene coverage, while data with a finer spatial resolution have a limited coverage (UAV data).

All data fusion methods attempt to overcome these measurement and sampling processes, which fundamentally limit the amount of information transferring from the scene to any one particular data set. Indeed, in most cases of data fusion in remote sensing, the different data sets to be fused derive in different ways from the same scene model, at least

as defined in a specific space–time dimension and with specific measurable properties (e.g., land cover objects and topographic surface). Understanding these measurement and sampling processes is, therefore, key to characterizing methods of data fusion because each operates on different parts of the sequence from scene model to data. For example, it is equally possible to perform the data fusion process in the scene space (e.g., via some data-generating model, such as a geometric model) as in the data space (the more common approach) [21].

Finally, we define the statistical model framework as including

- 1) measurement to provide data, as described previously
- 2) characterization of the data through model fitting
- 3) prediction of unobserved data, given 2)
- 4) forecasting [26].

Items 1, 2, and 3 are defined in space or space–time, while item 4 extends through time beyond the range of the current data. Prediction—item 3—can be of the measured property  $x$  (e.g., reflectance or topographic elevation, through interpolation), or it can be of a property of interest  $y$  to which the measured  $x$  data are related (e.g., land cover or vegetation biomass through classification or regression-type approaches). Similarly, data fusion can be undertaken on  $x$ , or it can be applied to predict  $y$  from  $x$ . Generally, therefore, data fusion is applied either between items 2 and 3 (e.g., fusion of  $x$  based on the model in item 2), as part of prediction (e.g., fusion to predict  $y$ ), or after the prediction of certain variables (e.g., ensemble unification). In this article, the focus is on data fusion to predict  $x$ .

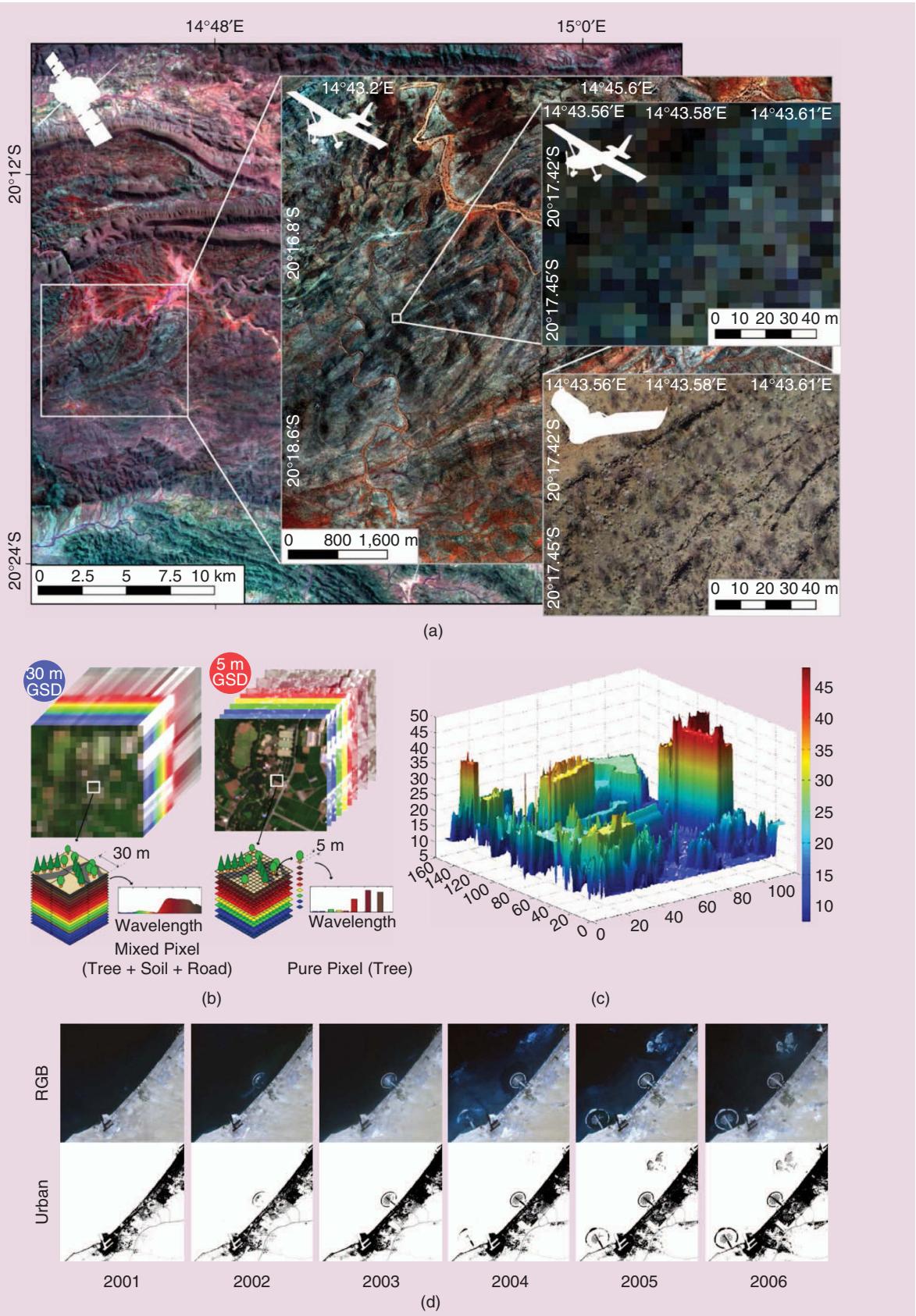
Data fusion is made possible because each data set to be fused represents a different view of the same real world defined in space and time (generalized by the scene model), with each view having its own measurable properties, measurement processes, and sampling procedures. Therefore, crucially, one should expect some level of coherence between the real world (the source) and the multiple data sets (the observations), as well as among the data sets themselves, and this is the basis of most data fusion methods. This concept of coherence is central to data fusion [27].

Attempts to fuse data sets are potentially aided by knowledge of the structure of the real world. The real world is spatially correlated, at least at some scale [28], and this phenomenon has been used in many algorithms (e.g., geostatistical models [27]). Moreover, the real world is often composed of functional objects (e.g., residential houses and roads) for which we have expectations in terms of their size and shape; such expectations can aid in defining objective functions (i.e., in optimization solutions) [29]. These sources of prior information (on real-world structures) constrain the space of possible fusion solutions beyond the data themselves.

## KEY APPLICATION DOMAINS

Many key application domains stand to benefit from data fusion processing. For example, there exists a large number of applications where an increase in spatial resolution would add utility (the focus in the section that follows). These

**ENORMOUS EFFORTS HAVE BEEN MADE TO DEVELOP DATA FUSION TECHNIQUES FOR RECONSTRUCTING SYNTHETIC DATA THAT HAVE THE ADVANTAGES OF DIFFERENT SENSORS.**



**FIGURE 1.** (a) The multiscale nature of diverse data sets captured by multisensor data (spaceborne, airborne, and UAV sensors) in Namibia [14]. (b) The tradeoff between spectral and spatial resolutions. (c) Elevation information obtained by lidar sensors from the University of Houston. (d) A time-series data analysis for assessing the dynamic of changes using red-green-blue (RGB) and urban images captured from 2001 to 2006 in Dubai, United Arab Emirates.

include land cover classification, urban–rural definition, target identification, and geological mapping (e.g., [30]). A large area of current attention is the specific problem that arises from the tradeoff in remote sensing between spatial resolution and temporal frequency—in particular, the fusion of coarse-spatial/fine-temporal-resolution space-time data sets with fine-spatial/coarse-temporal-resolution ones, so as to provide frequent data with fine spatial resolution [31]–[34]. This will be detailed in the “PanSharpening and Resolution Enhancement” and “Multitemporal Data Fusion” sections. Land cover classification—one of the most vibrant fields of remote sensing research [35], [36], attempting to differentiate between several land cover classes available in a scene—can substantially benefit from data fusion.

Another example is the tradeoff between spatial resolution and spectral resolution [Figure 1(b)] to produce fine spectral–spatial resolution images, which play an important role in land cover classification and geological mapping. As can be seen in Figure 1(b), both fine spectral and spatial resolutions are required to provide detailed spectral information and at the same time avoid the mixed-pixel phenomenon. Further information about this topic can be found in the next section. Elevation information provided by lidar and TLS [see Figure 1(c)] can be used in addition to optical data to further increase classification and mapping accuracy, in particular

for classes of objects that are made up of the same materials (e.g., grassland, shrubs, and trees). Therefore, the sections “Point Cloud Data Fusion” and “Hyperspectral and Lidar” are dedicated to the topic of elevation data fusion and their integration with passive data. Furthermore, new streams of ancillary data obtained from social media, crowdsourcing, and scraping the Internet can be used as additional sources

of information, together with airborne and spaceborne data, for smart-city and smart-environment applications and for hazard monitoring and identification. This young yet active field of research is the focus of the “Big Data and Social Media” section.

Many applications can benefit from fused fine-resolution time-series data sets, particularly those that involve seasonal or rapid changes, as discussed in the “Multitemporal Data Fusion” section. Figure 1(d) shows the dynamic of the changes for an area in Dubai, United Arab Emirates, from 2001 to 2006 using time series of RGB and urban images. For example, keeping track of vegetation phenology (the seasonal growing pattern of plants) is crucial to deforestation monitoring [37] and crop yield forecasting, which militate against global food insecurity, natural hazards (e.g., earthquakes and landslides), and illegal pollution activities (e.g., oil spills and chemical leaks). However, such

information is provided globally only at a very coarse resolution, meaning that local smallholder farmers cannot benefit from the knowledge. Data fusion can be used to provide the frequent data needed for phenology monitoring—and at a fine spatial resolution that is relevant to local farmers [38].

Similar arguments can be applied to deforestation, where frequent, fine-resolution data may aid in speeding up the timing of government interventions [37], [39]. The case for fused data is arguably even greater for rapid-change events, for example, forest fires and floods. In these circumstances, the need for frequent updates at a fine resolution is obvious. While these application domains provide compelling arguments for data fusion, there exist many challenges, including 1) the data volumes produced at a coarse resolution via such sensors as the Moderate Resolution Imaging Spectroradiometer (MODIS) and MERIS are already vast, meaning that data-set fusion most likely needs to be undertaken on a case-by-case basis as an on-demand service, and 2) rapid-change events require ultrafast processing, meaning that speed may outweigh accuracy in such cases [40].

In summary, data fusion approaches in remote sensing vary greatly, depending on the many considerations described previously, including the sources of the data sets to be fused. In the following sections, we review data fusion approaches in remote sensing in terms only of the data sources to be fused; however, the further considerations introduced previously are relevant to each section.

## PANSHARPENING AND RESOLUTION ENHANCEMENT

Optical Earth observation satellites have tradeoffs in spatial, spectral, and temporal resolutions. Enormous efforts have been made to develop data fusion techniques for reconstructing synthetic data that have the advantages of different sensors. Depending on which pair of resolutions has a tradeoff, these technologies can be divided into two categories: 1) spatirospectral fusion to merge fine spatial and fine spectral resolutions [see Figure 2(a)] and 2) spatiotemporal fusion to blend fine spatial and fine temporal resolutions [see Figure 2(b)]. This section provides overviews of these technologies, including recent advances.

## SPATIOSPECTRAL FUSION

Satellite sensors like WorldView and Landsat ETM+ can observe the Earth’s surface at different spatial resolutions in different wavelengths. For example, the spatial resolution of the eight-band WorldView multispectral image is 2 m, but the single-band panchromatic (PAN) image has a spatial resolution of 0.5 m. Spatirospectral fusion is a technique to fuse fine spatial resolution images (e.g., a 0.5-m WorldView PAN image) with coarse spatial resolution images (e.g., a 2-m WorldView multispectral image) to create fine spatial resolution images for all bands.

Spatiospectral fusion is also termed *pansharpening* when the available fine spatial resolution image is a single PAN image. When multiple fine spatial resolution bands are

FOR REMOTE SENSING-BASED GLOBAL MONITORING, THERE ALWAYS EXISTS A TRADEOFF BETWEEN SPATIAL RESOLUTION AND TEMPORAL VISIT FREQUENCY.

available, spatirospectral fusion is referred to as *multiband image fusion*, where two optical images with a tradeoff between spatial and spectral resolutions are fused to reconstruct fine spatial and fine spectral resolution imagery. Multiband image fusion tasks include multiresolution image fusion of single-satellite multispectral data (e.g., MODIS and Sentinel-2) and hyperspectral and multispectral data fusion [41].

Over the past decades, spatirospectral fusion has motivated considerable research in the remote sensing community. Most spatirospectral fusion techniques can be categorized according to at least one of five approaches:

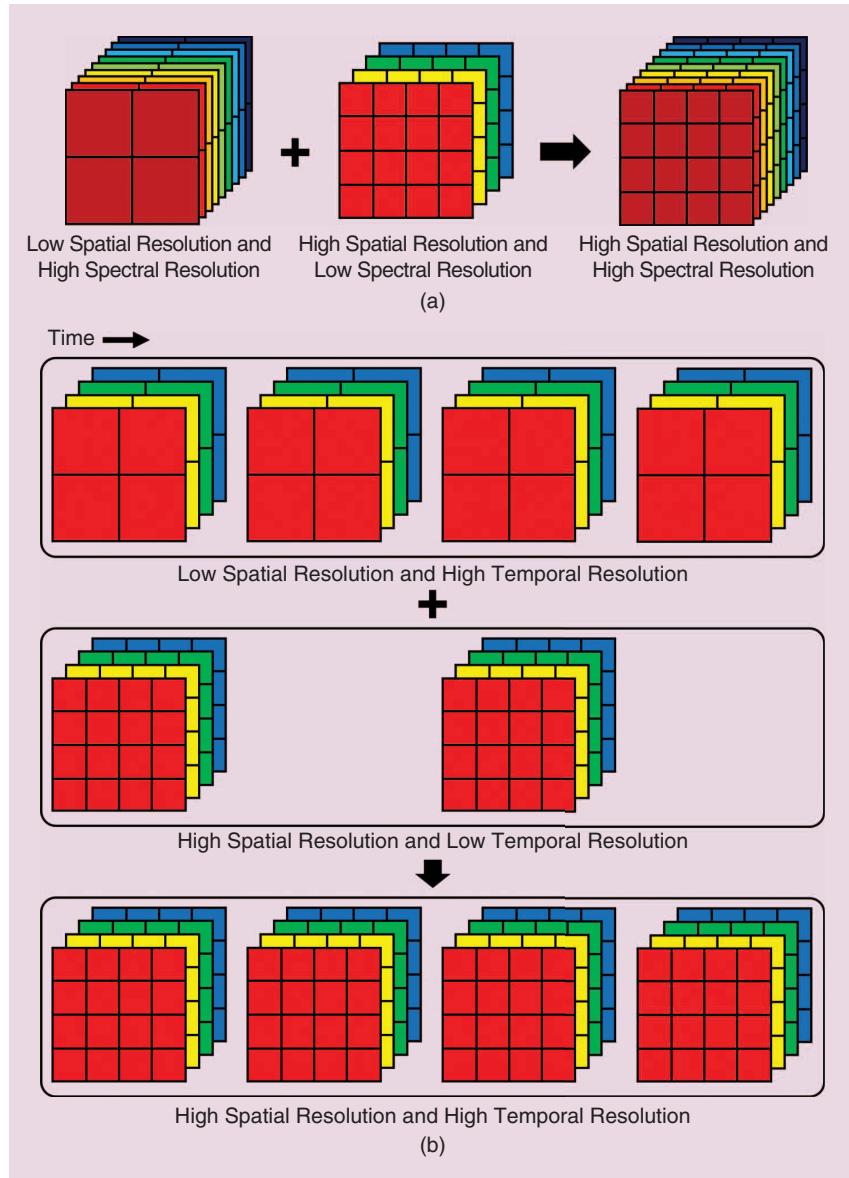
- ▷ component substitution (CS)
- ▷ multiresolution analysis (MRA)
- ▷ geostatistical analysis
- ▷ subspace representation
- ▷ sparse representation.

Figure 3 shows a history of representative works in the literature, with different colors (or rows) representing different categories of techniques. The size of each circle is proportional to the annual average number of citations (obtained by Google Scholar on 20 January 2018), which indicates the impact of each approach in the field. The main concept and the characteristics of each category are described in the following.

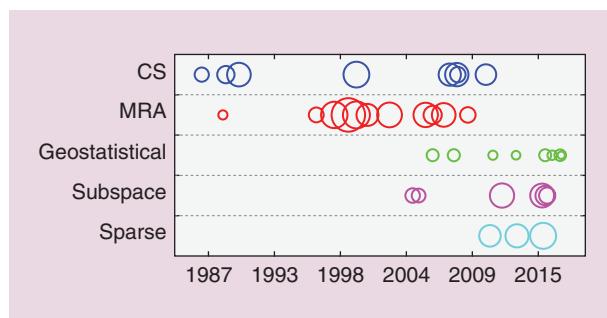
### COMPONENT SUBSTITUTION

CS-based pansharpening methods spectrally transform the multispectral data into another feature space to separate spatial and spectral information into different components. Typical transformation techniques include intensity/hue saturation [44], principal component analysis (PCA) [43], and Gram–Schmidt [46] transformations. Next, the component that is supposed to contain the multispectral image's spatial information is replaced by the PAN image after adjusting the intensity range of that image to the intensity range of the component, using histogram matching. Finally, the inverse transformation is performed on the modified data to obtain the sharpened image.

Aiazzi et al. in 2007 proposed the general CS-based pansharpening framework, where various methods based on different transformation techniques can be explained in a unified way [48]. In this framework, each multispectral band is sharpened by injecting spatial details obtained as the difference between the PAN image and a coarse spatial



**FIGURE 2.** Schematic illustrations of (a) spatirospectral fusion and (b) spatiotemporal fusion.



**FIGURE 3.** A history of the representative literature for five approaches in spatirospectral fusion. The size of each circle is proportional to the annual average number of citations. For each category, from left to right, the circles correspond to [42]–[50] for CS; [51]–[57] for MRA; [27], [31], and [58]–[63] for geostatistical; [64]–[69] for subspace; and [70]–[72] for sparse.

resolution synthetic component multiplied by a band-wise modulation coefficient. By creating the synthetic component based on linear regression between the PAN image and the multispectral image, the performance of traditional CS-based techniques was greatly increased, mitigating spectral distortion.

CS-based fusion techniques have been used widely, owing to the following advantages: 1) high fidelity of spatial details in the output, 2) low computational complexity, and 3) robustness against misregistration. On the other hand, CS methods suffer from global spectral distortions when the overlap of spectral response functions (SRFs) between the two sensors is limited.

### MULTIRESOLUTION ANALYSIS

As shown in Figure 3, great effort has been devoted to the study of MRA-based pansharpening algorithms, particularly between 2000 and 2010, and they have been used widely

as benchmark methods for more than 10 years. The main concept of MRA-based pansharpening is to extract spatial details (or high-frequency components) from the PAN image and inject the details, multiplied by gain coefficients, into the multispectral data. MRA-based pansharpening techniques can be characterized by 1) the algorithm used

for obtaining spatial details (e.g., spatial filtering or multiscale transform) and 2) the definition of the gain coefficients. Representative MRA-based fusion methods are founded on box filtering [54], Gaussian filtering [56], bilateral filtering [73], wavelet transform [53], [55], and curvelet transform [57]. The gain coefficients can be computed either locally or globally.

Selva et al. in 2015 proposed a general framework called *hypersharpener* that extends MRA-based pansharpening methods to multiband image fusion by creating a fine-spatial-resolution synthetic image for each coarse-spatial-resolution band as a linear combination of fine spatial resolution bands based on linear regression [74].

The main advantage of the MRA-based fusion techniques is their spectral consistency. In other words, if the fused image is degraded in the spatial domain, a degraded image is spectrally consistent with the input coarse spatial and fine spectral resolution image. The main shortcoming is that its computational complexity is greater than that of CS-based methods.

### GEOSTATISTICAL ANALYSIS

Geostatistical solutions provide another family of approaches for spatiotemporal fusion. This type of procedure can preserve the spectral properties of the original coarse

images. That is, when the downsampled prediction is up-scaled to the original coarse spatial resolution, the result is identical to the original one (i.e., perfect coherence). Pardo-Iguzquiza et al. [58] developed a downscaling cokriging (DSCK) method to fuse the Landsat ETM+ multispectral images with the PAN image. DSCK treats each multispectral image as the primary variable and the PAN image as the secondary variable. DSCK was extended with a spatially adaptive filtering scheme [60], in which the cokriging weights are determined on a pixel basis, rather than being fixed in the original DSCK. Atkinson et al. [59] extended DSCK to downscale the multispectral bands to a spatial resolution finer than any input images, including the PAN image. DSCK is a one-step method and involves auto-semivariogram and cross semivariogram modeling for each coarse band [61].

Sales et al. [61] developed a kriging with external drift (KED) method to fuse the 250-m-resolution MODIS bands 1 and 2 with the 500-m bands 3–7. KED requires only auto-semivariogram modeling for the observed coarse band and simplifies the semivariogram modeling procedure, which makes it easier to implement than DSCK. As admitted by Sales et al. [61], however, KED suffers from high computational cost, as it computes kriging weights locally for each fine pixel. The computing time increases linearly with the number of fine pixels to be predicted.

Wang et al. [27] proposed an area-to-point regression kriging (ATPRK) method to downscale MODIS images. ATPRK includes two steps: regression-based overall trend estimation and area-to-point kriging (ATPK)-based residual downscaling. The first step constructs the relationship between the fine and coarse spatial resolution bands by regression modeling; then, the second step downscales the coarse residuals from the regression process with ATPK. The downsampled residuals are finally added back to the regression predictions to produce fused images.

ATPRK requires only auto-semivariogram modeling and is much easier to automate and more user friendly than DSCK. Compared to KED, ATPRK calculates the kriging weights only once and is a much faster method. ATPRK was extended with an adaptive scheme (called AATPRK) that fits a regression model using a local scheme where the regression coefficients change across the image [62]. For fast fusion of hyperspectral and multispectral images, ATPRK was extended with an approximate version [63]. The approximate version greatly expedites ATPRK and also has a very similar performance in fusion. In addition, ATPRK was employed for the fusion of the Sentinel-2 Multispectral Imager (MSI) images acquired from the recently launched Sentinel-2A satellite. Specifically, the six 20-m bands were downsampled to a 10-m spatial resolution by fusing them with the four observed 10-m bands [31].

### SUBSPACE REPRESENTATION

As indicated in Figure 3, research on subspace-based fusion techniques has become quite popular recently. Most

of these techniques have been developed for multiband image fusion. The subspace-based methods solve the fusion problem via the analysis of the intrinsic spectral characteristics of the observed scene using a subspace spanned by a set of basis vectors (e.g., a principal component basis and the spectral signatures of endmembers). The problem is formulated as the estimation of the basis at a fine spectral resolution and the corresponding subspace coefficients at a fine spatial resolution. This category of techniques includes various methods based on Bayesian probability [68], matrix factorization [66], and spectral unmixing [75]. The interpretation of the fusion process is straightforward in the case of unmixing-based methods: endmembers and their fine spatial resolution fractional abundances are estimated from the input images; the output is reconstructed by multiplying the endmember matrix and the abundance matrix.

A recent comparative review on multiband image fusion in [41] demonstrated that unmixing-based methods are capable of achieving accurate reconstruction results even when the SRF overlap between the two sensors is limited. Many subspace-based algorithms are computationally expensive compared to CS- and MRA-based methods because of iterative optimization. Recent efforts for speeding up the fusion procedure [69] are key to the applicability of this family of techniques for large images obtained by operational satellites (e.g., the Sentinel-2 constellation). Another drawback of the subspace-based methods is that they can introduce unnatural artifacts in the spectral domain because of imperfect subspace representations.

### SPARSE REPRESENTATION

In recent years, spatiotemporal fusion approaches based on patch-wise sparse representation have been developed, along with the theoretical creation of compressed sensing and sparse signal recovery. Pansharpening based on sparse representation can be regarded as a special case of learning-based superresolution, where the correspondence between coarse and fine spatial resolution patches is learned from a database (or a dictionary). Li et al. in 2011 proposed the first sparse representation-based pansharpening method that exploits various external fine spatial resolution multispectral images as a database [70]. By considering the PAN image as a source for constructing a dictionary, it is possible to deal with the general problem of pansharpening, where there is only one PAN/multispectral image pair available [71]. Sparse representations have been introduced into the subspace-based fusion scheme to regularize fine spatial resolution subspace coefficients based on Bayesian probability [72].

It is noteworthy that sparse representation-based techniques are capable of sharpening spatial details in visible in the fine spatial resolution image at exactly the same location by reconstructing each patch of the output as a linear combination of nonlocal patches of the fine spatial resolution image. The critical drawback is its extremely high

computational complexity, sometimes requiring supercomputers to process fusion tasks in an acceptable time.

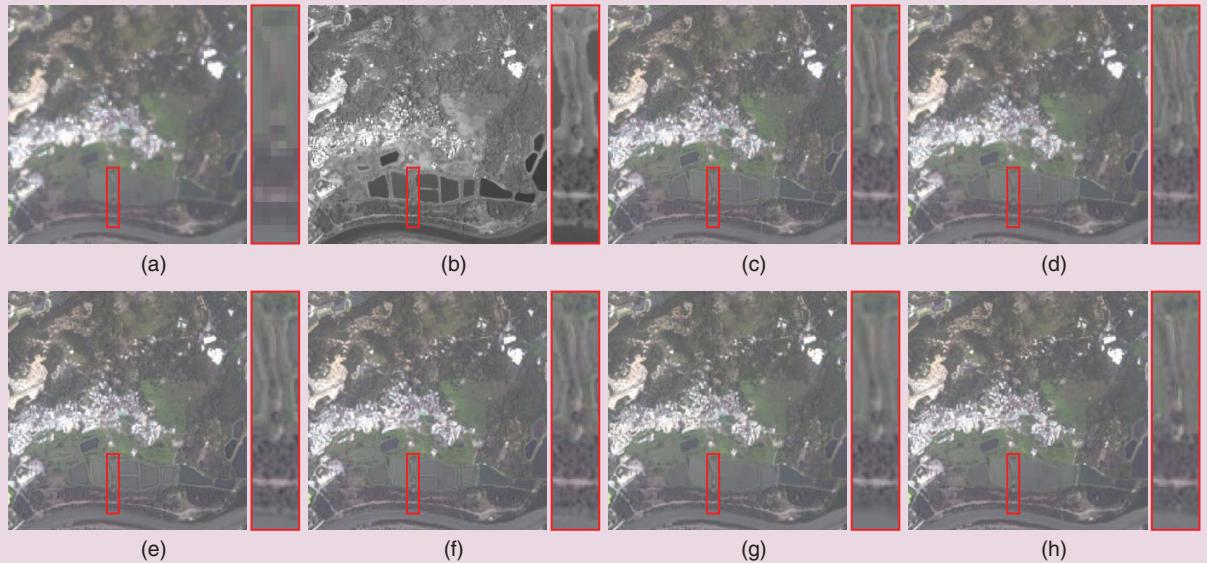
We compare five representative pansharpening algorithms, namely, Gram–Schmidt adaptive (GSA) [48], smoothing filtered-based intensity modulation (SFIM) [54], modulation-transfer-function-based generalized Laplacian pyramid with high-pass modulation (MTF-GLP-HPM) [76], ATPRK [27], and jointly sparse fusion of hyperspectral and multispectral imagery (J-SparseFI-HM) [77] using WorldView-2 data taken over Hong Kong. The original data set consists of one 0.5-m ground sample distance (GSD) PAN and eight 2-m GSD multispectral bands. To assess the quality of the pansharpened images, we adopt Wald's protocol, which degrades the original PAN and multispectral images to 2-m and 8-m GSDs, respectively, with the original multispectral bands being the reference. For quantitative evaluation, we use peak SNR (PSNR), spectral angle mapper (SAM), *erreur relative globale adimensionnelle de synthèse* (ERGAS) [78], and Q2n [79], all of which are well-established quality measures in pansharpening. PSNR quantifies the spatial reconstruction quality of each band, and the SAM index measures the spectral information preservation at each pixel. We use the average PSNR and SAM values. ERGAS and Q2n are global reconstruction indices.

The experimental results are compared visually and quantitatively in Figure 4 and Table 1, respectively. The quality measures in Table 1 are consistent with the literature: GSA, SFIM, and MTF-GLP-HPM provide the competitive baselines, ATPRK clearly outperforms the baselines, and J-SparseFI-HM achieves further increases in accuracy. In Figure 4, we can observe different characteristics of the investigated methods. For instance, GSA, SFIM, MTF-GLP-HPM, and ATPRK show sharper edges but also artifacts along object boundaries (e.g., between water and vegetation) where brightness is reversed between the PAN image and each band. J-SparseFI-HM deals with such situations and produces visually natural results, owing to its nonlocal sparse-representation capability.

### SPATIOTEMPORAL FUSION

For remote sensing-based global monitoring, there always exists a tradeoff between spatial resolution and temporal revisit frequency (i.e., temporal resolution). For example, the MODIS satellite can provide data on a daily basis, but the spatial resolution (250–1,000 m) is often too coarse to provide explicit land cover information, as such information may exist at a finer spatial scale than the sensor resolution. The Landsat sensor can acquire images at a much finer

**IN FUTURE RESEARCH, IT WILL BE OF GREAT INTEREST TO DEVELOP MORE ACCURATE METHODS TO ACCOUNT FOR LAND COVER CHANGES AND THE INCONSISTENCY BETWEEN COARSE AND FINE SPATIAL RESOLUTION TIME SERIES.**



**FIGURE 4.** The Hong Kong WorldView-2 data set (bands 4, 3, and 2 as RGB): (a) an 8-m coarse multispectral image, (b) a 2-m PAN image, (c) GSA, (d) SFIM, (e) MTF-GLP-HPM, (f) ATPRK, (g) J-SparseFI-HM, and (h) a 2-m reference image.

**TABLE 1. A QUANTITATIVE ASSESSMENT OF FIVE REPRESENTATIVE PANSHARPENING METHODS FOR THE HONG KONG WORLDVIEW-2 DATA SET.**

CATEGORY	METHOD	PSNR	SAM	ERGAS	Q2 <sup>n</sup>
—	Ideal	inf	0	0	1
CS	GSA	36.9624	1.9638	1.2816	0.86163
MRA	SFIM	36.4975	1.8866	1.2857	0.86619
MRA	MTF-GLP-HPM	36.9298	1.8765	1.258	0.85945
Geostatistical	ATPRK	37.9239	1.7875	1.1446	0.88082
Sparse	J-SparseFI-HM	37.6304	1.6782	1.0806	0.88814

spatial resolution of 30 m but has a limited revisit capability of 16 days. Fine spatial and temporal resolution data are crucial for timely monitoring of highly dynamic environmental, agricultural, and ecological phenomena. The recent development of remotely piloted aircraft systems, or drones, will provide a considerable amount of multisource data with very high spatial and temporal resolution.

Spatiotemporal fusion is a technique to blend fine spatial resolution data with coarse temporal resolution (e.g., Landsat) data and fine temporal resolution data with coarse spatial resolution data to create fine spatiotemporal resolution data (e.g., MODIS) [80]–[82]. Its implementation is performed based on the availability of at least one coarse/fine spatial resolution image pair (e.g., a MODIS–Landsat image pair acquired on the same day) or one fine spatial resolution land cover map that is temporally close to the prediction day. Over the past decade, several spatiotemporal fusion methods have been developed, and they can generally be categorized into image pair-based and spatial unmixing-based methods.

The spatial and temporal adaptive reflectance fusion model (STARFM) [83] is one of the earliest and most widely used spatiotemporal fusion methods. A typical image pair-based method, it assumes that the temporal changes of all land cover classes within a coarse pixel are consistent, which is more suitable for homogeneous landscapes dominated by pure coarse pixels. To enhance STARFM for heterogeneous landscapes, an enhanced STARFM (ESTARFM) method was developed [84]. ESTARFM requires two coarse/fine image pairs to estimate the temporal change rate of each class separately and assumes the change rates to be stable during the relevant period [85]. Moreover, some machine learning-based methods have been proposed, including sparse representation [86], [87], extreme learning machine [88], artificial neural networks [89], and deep learning [90]. These methods learn the relationship between the available coarse/fine image pairs, and this is used to guide the prediction of fine images from coarse images on other days.

Spatial unmixing-based methods can be performed using only one fine spatial resolution land cover map. The thematic map can be produced by interpretation of the available fine spatial resolution data [91]–[93] or from other sources, such as an aerial image [94] or a land use database [95]. These techniques are performed based on the strong assumption that there is no land cover or land use change during the period of interest. Using a fine spatial resolution land use database (LGN5) [95] or a 30-m thematic map obtained by the classification of an available Landsat image [93], 30-m Landsat-like time series were produced from 300-m MERIS time series to monitor vegetation seasonal dynamics.

To maintain the similarity between the predicted end-members and the predefined endmembers extracted from the coarse data, Amors-López et al. [91], [92] proposed

to include a new regularization term to the cost function of the spatial unmixing. Wu et al. [96] and Gevaert et al. [97] extended spatial unmixing to cases with one coarse/fine image pair available. The method estimates changes in class endmember spectra from the time of the image pair to prediction, before adding them to the known fine spatial resolution image. Furthermore, Huang and Zhang [98] developed an unmixing-based spatiotemporal reflectance fusion model using two coarse/fine image pairs. In addition, the image pair-based and spatial unmixing-based methods can be combined [32], [99], [100].

Spatiotemporal fusion is essentially an ill-posed problem involving inevitable uncertainty, especially for predicting abrupt changes and heterogeneous landscapes. To this end, Wang et al. [101] proposed to incorporate the freely available 250-m MODIS images into spatiotemporal fusion. Compared to the original 500-m MODIS data, the 250-m data can provide more information for abrupt changes and heterogeneous landscapes and, thus, can increase the accuracy of spatiotemporal fusion predictions.

Blending MODIS and Landsat has been the most common spatiotemporal fusion problem over the past decade. Sentinel-2 and Sentinel-3 are two recently launched satellite constellations for global monitoring. The Sentinel-2 MSI and Sentinel-3 Ocean and Land Color Instrument (OLCI) sensors have quite different spatial and temporal resolutions (Sentinel-2 MSI sensor: 10/20/60 m, 10 days, albeit five days with two sensors, conditional upon clear skies; Sentinel-3 OLCI sensor: 300 m, <1.4 days with two sensors). Wang et al. [34] proposed a new method, called *Fit-FC*, for spatiotemporal fusion of Sentinel-2 and Sentinel-3 images to create nearly daily Sentinel-2 images. *Fit-FC* is a three-step method consisting of regression model fitting, spatial filtering, and residual compensation. The *Fit-FC* method can be implemented using only one image pair and is particularly relevant for cases involving strong temporal changes.

### CHALLENGES AND TRENDS OF DOWNSCALING

The major remaining issue in the field of spatiotemporal fusion is how to conduct fair comparisons. Many researchers use their own simulated data sets, and the source code is rarely released. To fairly evaluate the performance of each algorithm, it is necessary to develop benchmark data sets that can be accessible to everyone and include various scenes. Additionally, it is always desirable to release the source code of each method to enable reproducible research. In several review papers, researchers have attempted to evaluate many methods with common data sets and disclose their source code, which is an excellent contribution to the community. However, the diversity of the studied scenes may not be enough to evaluate the generalization ability, and those data sets are not freely available because of the restrictive data policy of the original sources. Regarding source code, many research groups never release their code, while always outperforming state-of-the-art algorithms in their papers. It is an urgent issue for the community to arrange benchmark

data sets on a platform like the IEEE Geoscience and Remote Sensing Society (GRSS) Data and Algorithm Standard Evaluation website [102] so that everyone can fairly compete on algorithm performance.

With respect to spatiotemporal fusion, the main challenges lie in the reconstruction of land cover changes and eliminating the differences between coarse and fine spatial resolution time series. Because of the large difference in spatial resolution between coarse and fine spatial resolution time series (e.g., a ratio of 16 for MODIS–Landsat), the prediction of

land cover changes (especially abrupt changes) from coarse images always involves great uncertainty. Most existing methods are performed based on the strong assumption of no land cover change (classical STARFM and ESTARFM) and the spatial unmixing-based method. Furthermore, due to differences in sensor characteristics, atmospheric conditions, and acquisition geometry, the available coarse and fine spatial resolution data (e.g., MODIS and Landsat data) are not always perfectly consistent. The uncertainty is directly propagated to the spatiotemporal fusion process. In future research, it will be of great interest to develop more accurate methods to account for land cover changes and the inconsistency between coarse and fine spatial resolution time series.

**MULTISOURCE FUSION  
INCLUDING POINT CLOUDS  
IS ALREADY USED IN A WIDE  
VARIETY OF APPLICATION  
FIELDS AND REVEALS  
SEVERAL TRENDS.**

### POINT CLOUD DATA FUSION

Georeferenced point clouds have gained importance in recent years due to a multitude of developments in technology and research that increased their availability (e.g., hardware to capture 3D point clouds) and usability in applications (e.g., algorithms and methods to generate point clouds and analyze them) [103]. Research and development with point cloud data are driven by several disciplines (e.g., photogrammetry, computer science, geodesy, geoinformatics, and geography), scientific communities (e.g., lidar, computer vision, and robotics), and industry [104]. Spatial and temporal scales to utilize point clouds range from episodic countrywide and large-scale topographic mapping to near-real-time usage in autonomous driving applications. Sensors and methods to derive point clouds include predominantly lidar and photogrammetry, respectively [105]. A further very recent data source of point clouds in research is tomographic SAR [106]. Also, low-cost depth cameras are increasingly used [107]. Lidar, also referred to as *laser scanning*, is the only widely used method that records 3D points directly as an active remote and also is a close-range sensing technique [108].

### POINT CLOUD DATA MODEL

Although we draw a very broad picture, the common denominator is the point cloud data model, which is the

initial data model shared by all multisource fusion methods that include point clouds. Otepka et al. [109] defined the georeferenced point cloud data model as a set of points  $P_i$ ,  $i = 1, \dots, n$  in 3D Cartesian space that is related to a geospatial reference system (e.g., the Universal Transverse Mercator).  $P_i$  has at least three coordinates  $(x_i, y_i, z_i)^T \in \mathbb{R}^3$  for its position, and it can have additional point features, also referred to as attributes  $a_{j,i}$ , with  $j = 1, \dots, m_i$  as the number of point features of point  $i$ . A point feature  $a_j$  could be the color of a spectral band, lidar, or SAR backscatter value; a classification or segmentation ID; a local surface normal vector component (e.g.,  $n_x, n_y, n_z$ ); and so forth.

Figure 5 visualizes a point cloud with further point features stored in additional columns of a table with the 3D coordinates. Such point features can originate from the measurement process (e.g., lidar intensity [110]), or they can be derived by data postprocessing (e.g., segmentation) and fusion with other data sources. Interested readers can refer to [109] and [111] for a more detailed description of lidar point cloud features. A point in a point cloud  $P_i$  is a vector  $(x_i, y_i, z_i, a_{1,i}, \dots, a_{m_i,i})^T$ , of dimension  $3 + m_i$ , with the 3D coordinates as the first three dimensions (see Figure 5).

Generally, the point cloud model supports a variable number of point features  $m_i$  and leaves the 3D spatial distribution of  $(x_i, y_i, z_i)^T$  up to the point cloud generation process. The main challenges of the point cloud model for fusion with other data sources is the unstructured 3D spatial nature of  $P$  and that often no fixed spatial scale and accuracy exist across the data set. Local neighborhood

information must be derived explicitly, which is computationally intensive, and the definition of *neighborhood* depends on the application and respective processing task [109], [112].

## CONCEPTS OF POINT CLOUD FUSION

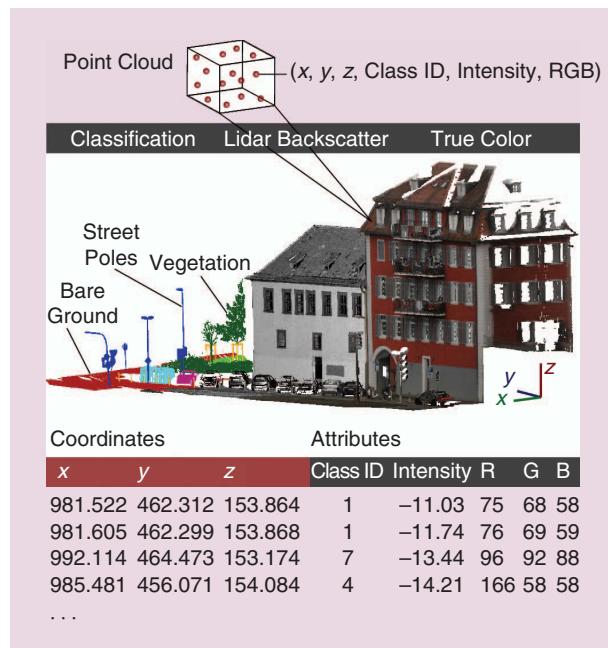
The main objectives of point cloud data fusion are to make use of the 3D geometric, spatial-structural, and lidar backscatter information inherent in point clouds and combine it with spectral data sources or other geoinformation layers, such as geographic information system (GIS) data. Zhang and Lin [104] gave a broad overview of applications involving the fusion of optical imagery and lidar point clouds. Looking more specifically at the methodology of fusion, three main methodological concepts can be distinguished in the literature with respect to the target model of multi-source point cloud fusion. The target data model of data fusion also determines which methods and software (e.g., image or point cloud processing) are primarily applied to classify the data sets. Based on the target data model (or product) we separate the following strategies (see Figure 6).

- 1) *Point cloud level*: Enrich the initial point cloud  $P$  with new point features.
- 2) *Image/voxel level*: Derive new image layers representing 3D point cloud information.
- 3) *Feature level*: Fuse point cloud information on the segment/object level.

## POINT CLOUD LEVEL: PIXEL TO POINT AND POINT TO POINT

Texturing point clouds with image data is a standard procedure with calibrated multisensor lidar systems for which the transformation from image to point cloud is well known from laboratory calibration, such as lidar systems with integrated multispectral cameras. For point clouds from photogrammetry—Structure from Motion and dense image matching—the spectral information is already given for each 3D point reconstructed from multiple 2D images [105]. Thus, the resulting point cloud  $P_i$  contains the respective pixel values from the images (e.g., R, G, and B) as point features and can be used for classification and object detection.

The labels of classified hyperspectral data can be transferred to the corresponding 3D points from lidar using precise coregistration. With this approach, Buckley et al. [113] related the spectra from close-range hyperspectral imaging pixels to terrestrial lidar point clouds to classify inaccessible geological outcrop surfaces. This enables improved visual inspection, but no joint 3D geometric and hyperspectral classification is conducted. A joint classification is presented by Vo et al. [114] in a paper for the 3D competition of the 2015 IEEE GRSS Data Fusion Contest [103]. They focused on lidar point clouds and RGB images and developed an end-to-end point cloud processing workflow. The authors made use of the colored point cloud and applied a supervised single-point classification (decision tree) to derive the target classes of ground, building,



**FIGURE 5.** A point cloud data model with the additional point features classification (ID per object class), intensity (lidar backscatter information), and true color (RGB values). Each point vector of the point cloud is stored in a table with its 3D coordinate and additional columns per attribute contained in the point cloud.

unassigned. This step was followed by the region-growing segmentation of the classified ground points to delineate roads. The point features of  $P$  were height, image intensity (RGB and hue-saturation value), laser intensity, height variation, surface roughness, and normal vector. RGB and laser-intensity data particularly supported the exclusion of grass areas, and joint classification increased the accuracy of a lidar-only solution by 2.3%. Generally, the majority of published approaches for multisource point cloud classification that resulted in a classified point cloud worked in the image domain and then transferred the classification results back to the point cloud [115]. This allows the use of fast and established image processing but limits the methods to single-point classification because the 3D point neighborhood information is not available in the classification procedure in the image domain as it is, for example, in point cloud segmentation.

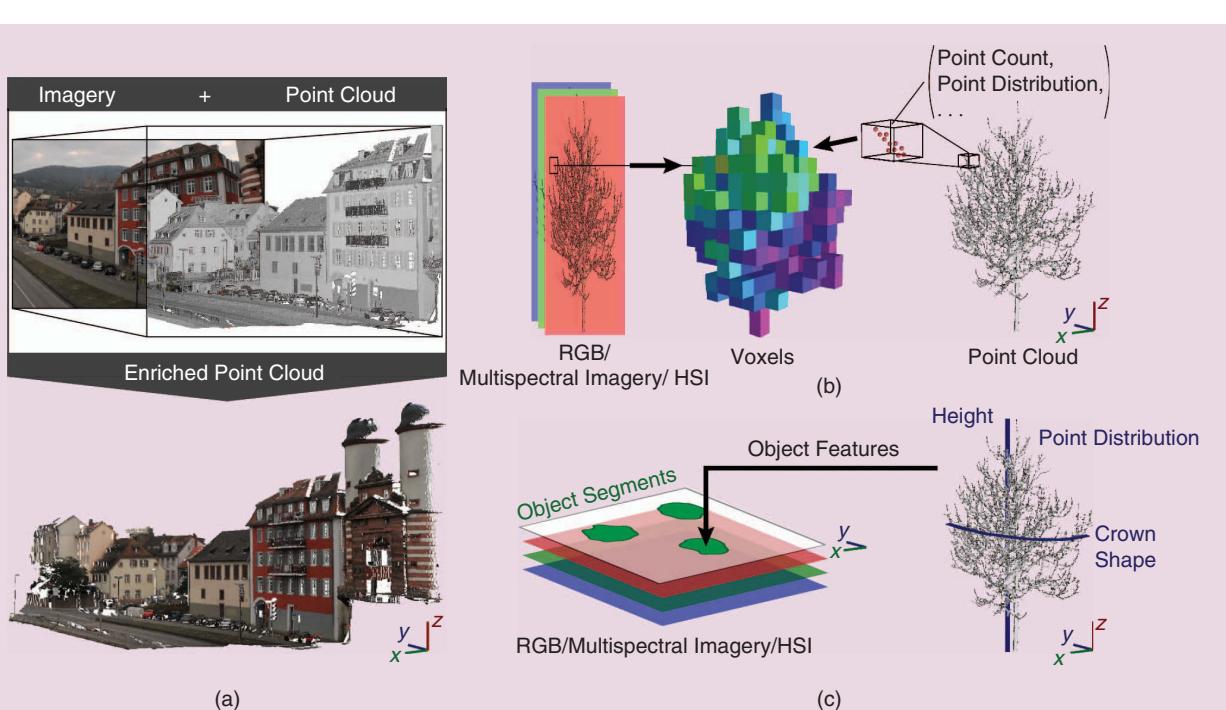
Point cloud-to-point cloud data fusion is known as *point cloud coregistration* or *alignment*. Coregistration of point clouds from the same sensor (e.g., within one lidar scanning campaign) is a standard preprocessing step in surveying with lidar from ground-based and airborne platforms [108]. Data fusion can be performed by different algorithms, such as point based (e.g., iterative closest point), keypoint based (e.g., scale invariant feature transform), surface based (e.g., local planes), or any combination of these [116]. This fusion principle is generally valuable in the merging of point clouds

from different sensor types, which have varied accuracies, spatial coverages, and spatial scales, and captured at different time stamps.

An image-based 2D registration for merging airborne and multiple terrestrial lidar point clouds was used by Paris et al. [117] to assess tree crown structures. They used the respective canopy height models for the registration, which was finally applied to the point cloud data sets to derive a fused point cloud.

A combination of data sets from different methods (e.g., lidar and photogrammetry) and platforms can lead to more accurate results compared to the use of an individual source. This was concluded in [118], where data sets from two different methods (lidar and photogrammetry) and three different platforms (a ground-based platform, small UAV platform, and manned-aircraft platform) were explored. The authors merged point clouds from UAV lidar, airborne manned lidar, and UAV photogrammetry spatially to a single point cloud to estimate the accuracy of bare earth elevation and the height of grasses and shrubs.

**FUTURE RESEARCH ON MULTISOURCE FUSION WITH POINT CLOUDS WILL NEED TO ADDRESS THE COMBINATION OF POINT CLOUDS FROM DIFFERENT SOURCES AND WITH STRONGLY HETEROGENEOUS CHARACTERISTICS.**



**FIGURE 6.** Strategies of point cloud data fusion on (a) the point cloud level, which visualizes the enrichment of the initial point cloud colored by lidar intensity using RGB information from imagery, with the RGB-colored point cloud as product; (b) the image/voxel level, which depicts a voxel model where each voxel contains information from a set of RGB and hyperspectral image layers as well as 3D point cloud features within each voxel; and (c) the feature/object level, which shows the assignment of features derived from the 3D point cloud to object segments created from raster image data.

## IMAGE/VOXEL LEVEL: POINT-TO-PIXEL/VOXEL

This concept transforms point cloud information into 2D images or voxels that can be analyzed by image processing approaches. In general, a multitude of images can be derived from the rich point clouds that come from point cloud geometry, lidar backscatter, and full-waveform lidar data directly. Those image bands usually represent elevation, geometric features (e.g., vertical distribution of points within a pixel), and lidar intensity-derived features. Ghamisi and Höfle [119] outlined several features that can be derived from lidar point clouds to encapsulate the 3D information into image bands for image classification, such as the laser echo ratio, variance of point elevation, plane fitting residuals, and echo intensity. The

fusion approach of lidar and hyperspectral imagery (HSI) in the classification of an urban scene is presented in the “Hyperspectral and Lidar” section. The experiment compares the classification results to the accuracies of the individual use of HSI.

Boulch et al. [115] used a pixel-based convolutional neural network (CNN)

to perform semantic labeling of point clouds based on RGB and geometric information (e.g., depth composite image). Every 3D point is labeled by assigning the derived pixel-wise label predictions to the single 3D points via back projection. The study could apply to both terrestrial lidar and photogrammetric point clouds.

A fusion of UAV-borne lidar, multispectral, and hyperspectral data was presented by Sankey et al. [120] for forest vegetation classification. They also used terrestrial lidar as a reference data set. The HSI was preclassified with the mixture-tuned matched filtering subpixel classification technique. The multisource fusion of UAV lidar and hyperspectral data (12-cm GSD) was performed via a decision tree classification approach. The fusion-based result achieved a higher accuracy for most target classes and also a greater overall accuracy (with an increase from 76% using only HSI to 88% for HSI and lidar data inputs). The largest increase, by adding lidar, was given for vegetation classes that separate well in height.

The combination of lidar's geometric quality with spectral information was used by Gerke and Xiao [121] to detect buildings, trees, vegetated ground, and sealed ground. They developed a method to fuse airborne lidar and multispectral imagery in two main consecutive steps: 1) point cloud segmentation (region growing) and classification (mean shift) using 3D lidar and spectral information [normalized difference vegetation index (NDVI)/saturation] and 2) supervised (random trees) or unsupervised voxel classification by a Markov random field framework using graph-cuts for energy optimization. The voxels contain features derived from 3D geometry and from the spectral image as well as the results

from the initial segmentation step. The results showed that spectral information supported the separation of vegetation from nonvegetation, but shadow areas still caused problems. Point cloud segmentation is sensitive to the color information also used in this process, which sometimes led to planes being missed out.

Airborne hyperspectral imagery was combined with full-waveform lidar data by Wang and Glennie [122] to classify nine target land cover classes (e.g., trees, bare ground, water, and asphalt road). The main goal was to generate synthetic vertical lidar waveforms by converting the raw lidar waveforms into a voxel model ( $1.2\text{ m} \times 1.2\text{ m} \times 0.15\text{ m}$ ). The voxels were then used to derive several raster features from the vertical distribution of backscatter intensity along the vertical voxels corresponding to one image pixel as well as metrics such as the height of the last return, penetration depth, and maximum lidar amplitude. In addition to these raster features, the authors derived principal components from the original 72 HSI bands and stacked them with the lidar features for classification. The fusion of lidar waveform data and HSI could increase the overall accuracy, using a support vector machine (SVM) classification, to 92.61% compared with 71.30% using only lidar and 85.82% using only HSI data.

## FEATURE/OBJECT LEVEL

This concept is based on the previously discussed concepts in terms of data model, which is used to derive objects followed by a classification step. Image or point cloud segmentation and combined pixel- and object-based approaches can be applied [123] to derive the entities for classification.

With airborne lidar images and full-waveform point cloud data, only one data source but two different data models for object-based urban tree classification were used by Höfle et al. [124]. The authors introduced a method to produce segments based on lidar point cloud-derived images [e.g., normalized digital surface model (nDSM) and echo ratio images]. The output segments were enriched by a multitude of geometric and full-waveform features computed directly in the 3D point clouds of each segment (e.g., the mean echo width). In particular, the geometric 3D point cloud features (e.g., the echo ratio) played an important role for vegetation classification because they encapsulated the 3D structure of the vegetation.

Alonso et al. [125] also worked at the single-tree/crown-object level and added HSI to the airborne lidar data set to map urban tree species. They applied canonical variates in a linear discriminant analysis classifier to assign tree-species labels to the segments, derived from the lidar canopy height model. Their lidar point cloud-derived structural variables included, for example, the median height of the returns in crowns and the average intensity below median height. Saarinen et al. [126] went one step further and fused UAV-borne lidar and HSI for mapping biodiversity indicators in boreal forests. After tree crown delineation by watershed segmentation, they derived

MULTISOURCE FUSION  
INCLUDING POINT CLOUDS  
IS ALREADY USED IN A WIDE  
VARIETY OF APPLICATION  
FIELDS AND REVEALS  
SEVERAL TRENDS.

point cloud-based segment features (e.g., height percentiles and average height) and also spectral segment features (e.g., mean and median spectra). Using nearest-neighbor estimation, it was possible to determine the variables of diameter at breast height, tree height, health status, and tree species for each crown segment. In the second step, the biodiversity indicators—structural complexity, extent of deciduous trees, and number of dead trees—were derived using single tree variables as input.

Considering multiple sensors, hyperspectral and lidar data were fused in an approach proposed by Man et al. [123] for urban land use classification (15 classes) with a combined pixel and feature-level method. Lidar point cloud information was encapsulated in image layers. Furthermore, the authors aimed at assessing the contribution of lidar intensity and height information, particularly for the classification of shadow areas. Their methodology included pixel-based features, such as the nDSM and intensity image from lidar, along with the inverse minimum noise fraction rotation bands, NDVI, and texture features (GLCM) of HSI data. The derived features were input to a supervised, pixel-based classification (SVM and maximum-likelihood classifiers). Additionally, an edge-based segmentation algorithm was used to derive segments using lidar nDSM, intensity, and NDVI images; this was followed by a rule-based classification of the derived objects. The classification outputs of the pixel- and object-based methods were merged by GIS raster calculation. The combination of HSI and lidar increased overall accuracy by 6.8% (to 88.5%) compared with HSI classification alone. The joint pixel- and object-based method increased overall accuracy by 7.1%, to 94.7%.

Liu et al. [127] used HSI and airborne lidar data as complementary data sources for crown structure and physiological tree information to map 15 different urban tree species. First, crowns were segmented by watershed segmentation of the canopy height model. Second, lidar and hyperspectral features were extracted for the crown segments to be used in the subsequent segment-based random forest (RF) classification. The 22 lidar-derived crown structural features per segment included, for example, crown shape, laser return intensity, and laser point distribution. The researchers concluded that the combination of lidar and HSI increased single-source classification up to 8.9% in terms of overall accuracy.

A complex fusion strategy for lidar point cloud and HSI image data in a two-stage neural network classification was developed by Rand et al. [128]. First, spectral segmentation of the HSI data was performed by a stochastic expectation–maximization algorithm and spatial segmentation of the lidar point cloud with a combined mean-shift and dispersion-based approach. Second, the resulting segments from lidar and HSI data were input to a supervised cascaded neural network to derive the final object class labels. The final fusion classification map was produced in 3D using the elevation values from the lidar point cloud. The team's approach resulted in a large increase in overall classification

accuracy by multisource fusion (HSI and lidar) to 98.5%, compared with 74.5% overall accuracy with HSI input only.

## CHALLENGES AND TRENDS IN POINT CLOUD FUSION

Generally, we can see a large gain in the importance of point clouds. Multisource fusion including point clouds is already used in a wide variety of application fields [104] and reveals several trends

- There is increasing use of machine-learning methods, including point clouds or point cloud derivatives.
- The majority of current approaches transform and encapsulate 3D point cloud information into 2D images or voxels and perform fusion and analysis on images or objects. The derived classification labels are transferred back to points afterward.
- The implementation of fusion (or joint use) for spectral and 3D point cloud information from single-source photogrammetry (Structure from Motion and dense image matching) exists now. The link between point clouds and images is already given via several methodologies.
- The fusion of geometric and backscatter point cloud information from lidar exhibits increases in terms of classification accuracy.

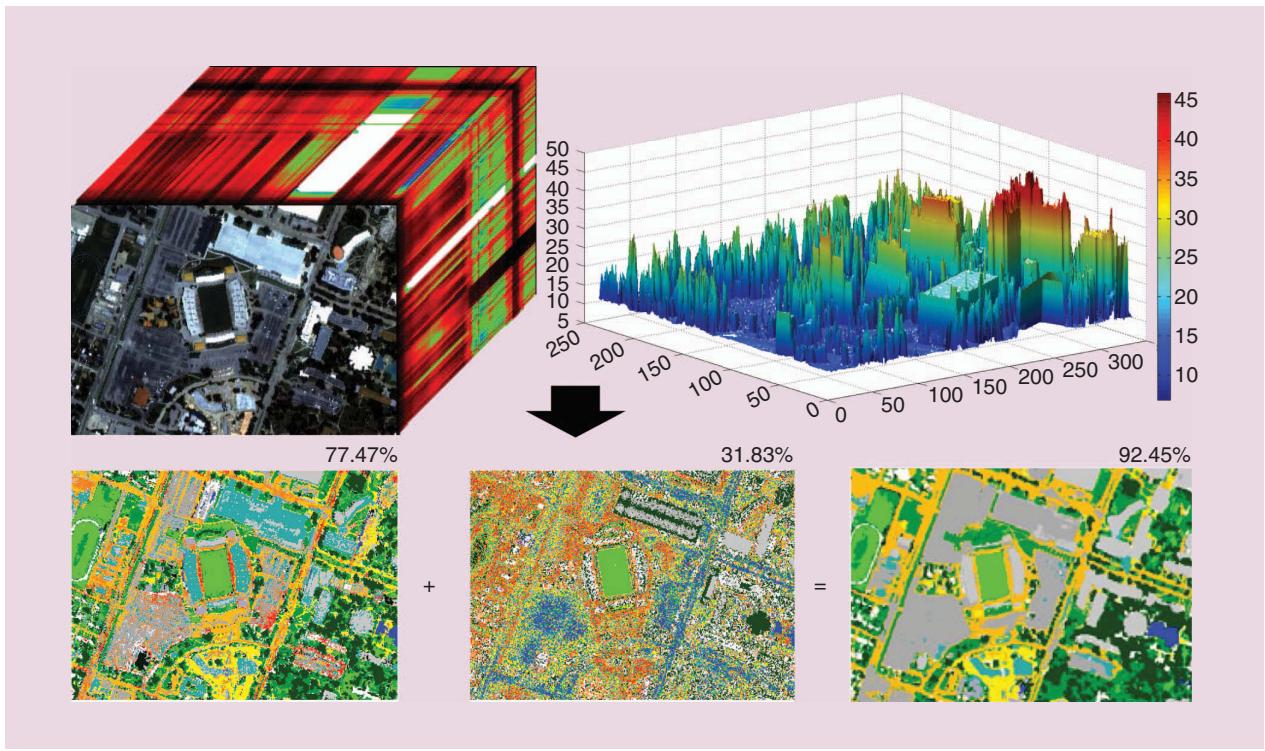
Future research on multisource fusion with point clouds will need to address the combination of point clouds from different sources and with strongly heterogeneous characteristics (e.g., point density and 3D accuracy). So far, mainly one source of point clouds is used in the fusion process, e.g., the joint use of HSI and lidar point clouds. Multispectral [129] and even hyperspectral lidar data [130] offer new possibilities for the fusion of point clouds as well as of point clouds with multispectral/hyperspectral data. The availability of 3D point cloud time series [110] will also enable investigation of how temporal aspects need to be addressed in fusion and classification approaches.

The number of contributions on HSI and lidar rasterized data fusion in the remote sensing community is quickly growing because of the complementary nature of such multisensor data. Therefore, the next section is specifically dedicated to the fusion of HSI and lidar-derived features and offers a review of such fusion schemes.

## HYPERSPECTRAL AND LIDAR

The efficacy of lidar (characterized as an active remote-sensing technique) for the classification of complex areas (e.g., where many classes are located close to each other) is limited by the lack of spectral information. On the other

**THE USE OF MULTITEMPORAL INFORMATION IS CRUCIAL FOR MANY IMPORTANT APPLICATIONS (FROM THE ANALYSIS OF SLOW AND SMOOTHLY EVOLVING PHENOMENA TO STEEP AND ABRUPT CHANGES).**



**FIGURE 7.** A depiction of HSI/lidar fusion. These images were generated based on studies in [138], where the overall classification accuracy of HSI (77.47%) and lidar (31.83%) was significantly increased to 92.45% using a feature fusion approach.

hand, hyperspectral sensors (characterized as passive remote sensing techniques) provide rich and continuous spectral information by sampling the reflective portion of the EMS, ranging from the visible region ( $0.4\text{--}0.7\ \mu\text{m}$ ) to the short-wave infrared region (almost  $2.4\ \mu\text{m}$ ) in hundreds of narrow contiguous spectral channels (often 10 nm wide).

Such detailed spectral information has made HSI a valuable source of data for complex scene classification. Detailed and systematic reviews on hyperspectral data classification for characterizing complex scenes have been published in [35] and [131]. However, hyperspectral images do not contain any information about the elevation and size of different materials, which imposes an inevitable constraint on classifying objects made up of similar materials (e.g., grassland, shrubs, and trees). The aforementioned limitations and capabilities of each sensor, as discussed in “The Deluge of Data” section, have provided the main motivation for fusing HSI and lidar.

The joint use of lidar and HSI has already been investigated for diverse applications, such as rigorous illumination correction [132] and quantifying riparian habitat structure [133]. However, the main application of this multisensor fusion technique is dedicated to scene classification, which is also the preeminent focus of this section.

Several studies, such as [134] and [135], investigated the differentiation of diverse species of trees in complex forested areas, while several other approaches dealt with complex urban area classification, e.g., [136]. Coregistered lidar

and HSI data were introduced in [137]. Figure 7 (generated based on some studies in [138]) demonstrates schematically that the fusion of HSI and lidar can considerably increase the classification accuracy over that of each source alone.

Now, we briefly discuss the fusion of lidar and HSI in terms of a few key approaches categorized in to four subsections: filtering strategies, low-rank models, composite kernels, and deep learning-based fusion approaches. Corresponding to each section, some numerical classification results obtained from the Compact Airborne Spectrographic Imager (CASI) Houston University data (details follow) are reported in Table 2. To obtain a better numerical evaluation, the classification accuracies of the individual use of HSI obtained by RF ( $\text{RF}_{\text{HSI}}$ ), SVM ( $\text{SVM}_{\text{HSI}}$ ), and CNN ( $\text{CNN}_{\text{HSI}}$ ) are also listed in Table 2.

## HOUSTON UNIVERSITY

The Houston University data for this section are composed of a lidar-derived DSM and an HSI, both captured over the University of Houston campus and the neighboring urban area. This data set was initially made publicly available for the 2013 GRSS Data Fusion Contest. The HSI and lidar data were captured on 23 June 2012 and 22 June 2012, respectively. The size of the data set is  $(1.2\text{ m} \times 1.2\text{ m} \times 0.15\text{ m})$ . pixels, with a GSD of 2.5 m. The HSI consists of 144 spectral bands ranging from  $0.38$  to  $1.05\mu\text{m}$ . Figure 8 illustrates the investigated data and the corresponding training and test samples. The number of training and test samples for different classes is detailed in Table 2.

## FILTERING

Filtering approaches have been used intensively in the literature to effectively extract contextual and spatial features by attenuating redundant spatial details (based on a criterion) and preserving the geometrical characteristics of the other regions. Among those techniques, one can refer to 1) morphological profiles (MPs) [144], which can be produced by the sequential implementation of opening and closing operators through reconstruction by considering a structuring element of increasing size; 2) attribute profiles (APs) [145], which can obtain a multilevel characterization of the input image by considering the repeated implementation of morphological attribute filters; and 3) extinction profiles (EPs) [146], which can obtain a multilevel characterization of the input image by considering the repeated implementation of a morphological extinction filter.

These approaches have been investigated frequently for the fusion of lidar and HSI because they are fast and conceptually simple and able to provide accurate classification results. For instance, in [147] and [148], the spatial features of HSI and lidar were extracted using APs. Then, they were concatenated and fed to a classifier, leading to quick, precise results in terms of classification accuracy. In [142], EPs were used to automatically extract the spatial and elevation features of HSI and lidar data. The extracted features were stacked and then classified using an RF classifier. (The results obtained by that approach can be found in Table 2 as  $\text{EP}_{\text{HSI+lidar}}$ .)

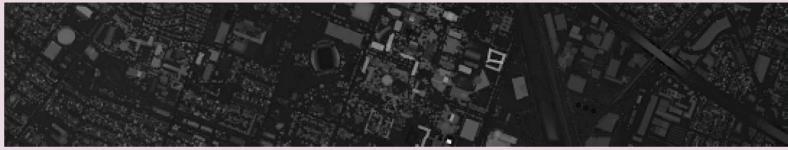
Filtering approaches like MPs, APs, and EPs suffer from two shortcomings: the curse of dimensionality and intensive processing time for the subsequent classification steps because they usually increase the number of dimensions by stacking spectral, spatial, and elevation features extracted from HSI and lidar, while the number of training samples remains the same. To address these deficiencies, composite kernel- and low-rank-based methods, which will be discussed in the following subsections, have been suggested in the literature to effectively fuse HSI and lidar.

## LOW-RANK MODELS

To avoid the curse of dimensionality and also increase the efficiency of the analysis compared to filtering approaches, low-rank models were

**TABLE 2. THE CLASSIFICATION ACCURACY VALUES FOR THE HOUSTON DATA SET ACHIEVED BY DIFFERENT STATE-OF-THE-ART APPROACHES.**

CLASS NAME	SPECTRAL						MULTISENSOR FUSION					
	RF <sub>HSI</sub> (%)	SVM <sub>HSI</sub> (%)	CNN <sub>HSI</sub> (%)	EP <sub>HSI+lidar</sub> (%)	GBFF (%)	FFCK (%)	MIR <sub>sub</sub> (%)	MIR <sub>MI-KSRC</sub> (%)	CNNGBFF (%)	SLRCA (%)	OTVCA (%)	
Grass, healthy	83.38	83.48	82.24	78.06	82.53	81.39	82.91	98.36	78.73	81.58	80.63	
Grass, stressed	98.4	96.43	98.31	84.96	98.68	99.91	81.48	98.59	94.92	99.44	99.62	
Grass, synthetic	98.02	99.8	70.69	100	100	100	100	100	100	98.61	100	
Tree	97.54	98.77	94.98	95.45	98.96	97.92	95.83	98.04	99.34	96.12	96.02	
Soil	96.4	98.11	97.25	98.76	100	100	99.05	93.15	99.62	99.72	99.43	
Water	97.2	95.1	79.02	95.80	95.1	95.8	91.61	100	95.8	98.6	95.8	
Residential	196.1/0.72	82.09	89.09	86.19	73.41	90.95	78.54	87.59	91.11	87.87	90.39	
Commercial	191.1/0.53	40.65	45.87	65.81	85.28	90.98	86.61	84.14	92.51	95.25	95.73	
Road	193.1/0.59	69.78	82.53	72.11	93.95	90.46	87.72	91.78	86.87	89.71	98.21	
Highway	191.1/0.36	57.63	83.2	55.21	67.08	60.91	68.82	86.2	94.66	81.18	63.42	
Railway	181.1/0.54	76.09	83.87	85.01	90.89	94.46	90.23	98.58	90.56	86.34	90.7	
Parking lot 1	192.1/1.041	49.38	70.99	60.23	88.56	99.14	98.08	92.32	90.74	92.7	91.07	
Parking lot 2	184/285	61.4	70.53	75.09	76.14	65.26	80.35	76.84	89.92	87.02	76.49	
Tennis court	181/247	99.6	100	83	100	100	100	99.6	98.58	99.19	100	
Running track	187/473	97.67	97.46	52.64	99.78	99.15	100	98.73	98.14	89.64	99.15	
Average accuracy (%)	-	80.34	86.34	77.19	88.54	91.24	91.02	90.65	NA	91.82	91.95	
Overall accuracy (%)	-	77.47	84.69	78.35	86.98	91.28	89.93	91.11	92.45	91.75	91.3	
Kappa coefficient K	-	0.7563	0.834	76.46	0.8592	0.903	0.891	0.8985	NA	0.9033	0.9056	
(no unit)											0.9181	



(a)



(b)



(c)



(d)

Thematic Classes:								
Grass, Healthy	Grass, Stressed	Grass, Synthetic	Tree					
Soil	Water	Residential	Commercial					
Road	Highway	Railway	Parking Lot 1					
Parking Lot 2	Tennis Court	Running Track						

**FIGURE 8.** (a) A lidar-derived rasterized DSM; (b) a color composite illustration of the CSSI Houston HSI, using bands 64, 43, and 22 as R, G, and B, respectively; (c) training samples; and (d) test samples.

investigated in [136], [138], [142], and [143], the main assumption of which was that the extracted features from HSI and lidar can be represented into a space of a lower dimension. All of those approaches followed a general framework, as demonstrated in Figure 9. This structure is composed of the following building blocks:

- ▷ DR<sub>1</sub> generates base images to build up an MP, AP, or EP.
- ▷ Filtering investigates the MP, AP, or EP to extract spatial features (e.g., EP/AP/MP<sub>HSI</sub>) and elevation features (e.g., EP/AP/MP<sub>lidar</sub>) from HSI and lidar, respectively.
- ▷ DR<sub>2</sub> is used to produce exactly the same number of spectral, spatial, and elevation features to put the same weight on each category. The other advantages of DR<sub>2</sub> are that it can reduce the executable computational cost and noise throughout the feature space. In [136], [138], [142], and [143], kernel PCA has been used for DR<sub>2</sub>.
- ▷ Finally, the outputs of the latter building block are fused and fed to a classification method. We next discuss [136], [138], [142], and [143] in more detail.

In [136], the spectral (HSI), spatial (MP<sub>HSI</sub>), and elevation features (MP<sub>lidar</sub>) were used (as the filtering step). A graph-based feature fusion (GBFF) technique was utilized (as the feature fusion step). Then, an SVM classifier was used to classify the fused features. (The results can be found in Table 2 as GBFF).

In [142], the spectral (HSI), spatial (EP<sub>HSI</sub>), and elevation features (EP<sub>lidar</sub>) were concatenated and fed to the GBFF and classified by a 2D CNN. These results can be found in Table 2 as CNNGBFF.

In [138], the following low-rank model was suggested to fuse HSI, EP<sub>HSI</sub>, and EP<sub>lidar</sub>:

$$\mathbf{F} = \mathbf{AV}^T + \mathbf{N}, \quad (1)$$

where  $\mathbf{F} = [\mathbf{f}_{(i)}]$  is an  $n \times p$  matrix that contains the vectorized features in its columns,  $\mathbf{V}$  is a  $p \times r$  unknown matrix containing the subspace basis,  $\mathbf{A} = [\mathbf{a}_{(i)}]$  is an  $n \times r$  matrix that contains the  $r$  unknown fused features in its columns, and  $\mathbf{N} = [\mathbf{n}_{(i)}]$  is the model error and noise. Note that  $r$  is the number of fused features. Also, hyperspectral bands and hyperspectral and lidar features are concatenated in matrix  $\mathbf{F}$  ( $\mathbf{F} = [\mathbf{EP}_{\text{HSI}}, \mathbf{HSI}, \mathbf{EP}_{\text{lidar}}]$ ).

In model (1), matrices  $\mathbf{A}$  and  $\mathbf{V}$  are both unknown. Therefore, they both need to be estimated. In [138], orthogonal total variation component analysis (OTVCA) [149] was suggested to solve this problem (as the feature fusion step shown in Figure 9). OTVCA is given by

$$\operatorname{argmin}_{\mathbf{A}, \mathbf{V}} \frac{1}{2} \|\mathbf{F} - \mathbf{AV}^T\|_F^2 + \lambda \sum_{i=1}^r \operatorname{TV}(\mathbf{a}_{(i)}) \text{ s.t. } \mathbf{V}^T \mathbf{V} = \mathbf{I}_r, \quad (2)$$

where the TV penalty is applied spatially on the fused features. TV preserves the spatial structure of the features while promoting piece-wise smoothness on the fused features. As a result, the final classification map contains homogeneous regions. The OTVCA fusion results can be found in Table 2 as OTVCA.

In [143], the extracted features were defined using the sparse and low-rank model given in [150],

$$\mathbf{F} = \mathbf{DWV}^T + \mathbf{N}, \quad (3)$$

where  $\mathbf{D}$  is an  $n \times n$  matrix that contains 2D wavelet basis, and  $\mathbf{W} = [\mathbf{w}_{(i)}]$  is an  $n \times r$  matrix containing the unknown

2D wavelet coefficients for the  $i$ th fused component. In [143], the sparse and low-rank component analysis [150], [151] was used to estimate  $\mathbf{W}$  and  $\mathbf{V}$ , given by

$$\operatorname{argmin}_{\mathbf{W}, \mathbf{V}} \frac{1}{2} \|\mathbf{F} - \mathbf{DWV}^T\|_F^2 + \lambda \sum_{i=1}^r \|\mathbf{w}_{(i)}\|_1 \text{ s.t. } \mathbf{V}^T \mathbf{V} = \mathbf{I}_r. \quad (4)$$

Note that the estimated fused features are given by  $\hat{\mathbf{F}}_{\text{fused}} = \mathbf{D}\hat{\mathbf{W}}$ . The fused features are expected to be sparse in the 2D wavelet basis. Therefore, in [150], to enforce the sparsity, an  $\ell_1$  penalty on the wavelet coefficients  $\mathbf{W}_r$  was used. As a result, promoting sparsity on the fused feature improves the SNR and final classification accuracies. The results for this approach can be found in Table 2 as SLRCA.

### COMPOSITE KERNEL

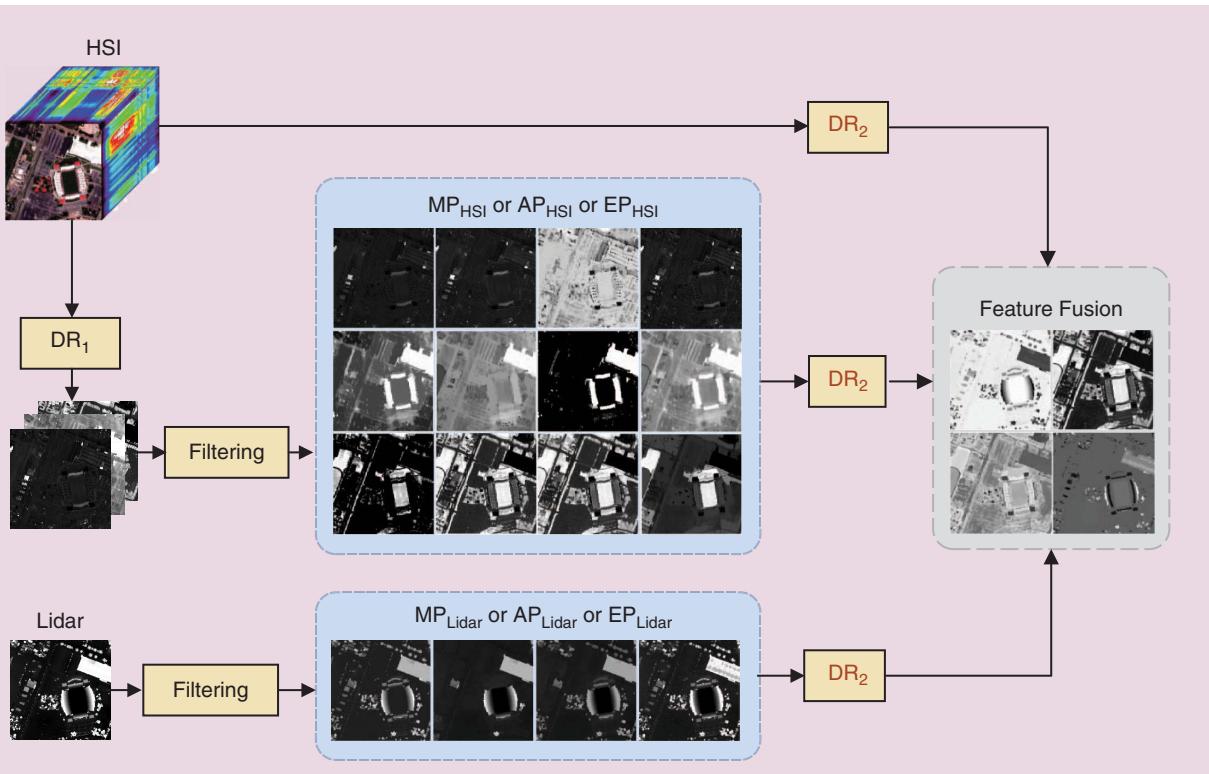
Composite kernel-based fusion techniques partially overcome the shortcomings of filtering methods by designing several kernels to handle spectral, spatial, and elevation features in feature space [152]. In [139], spectral, spatial (e.g., EP<sub>HSI</sub>), and elevation (e.g., EP<sub>Lidar</sub>) information were fused using a local-region filter and composite kernels. The results for this scheme can be found in Table 2 as FFCK. The main shortcoming here was that the obtained classification accuracy was dramatically influenced by the  $\mu$  parameter, which represents the amount of tradeoff between the spectral and spatial elevation kernels. To solve this issue, in [140], a fusion approach was introduced

capable of exhibiting substantial flexibility to integrate different feature sets without requiring any regularization parameters. That procedure was based on APs and multiple-feature learning using the subspace multinomial logistic regression (MLR<sub>sub</sub>) classifier. The results are shown as MLR<sub>sub</sub> in Table 2.

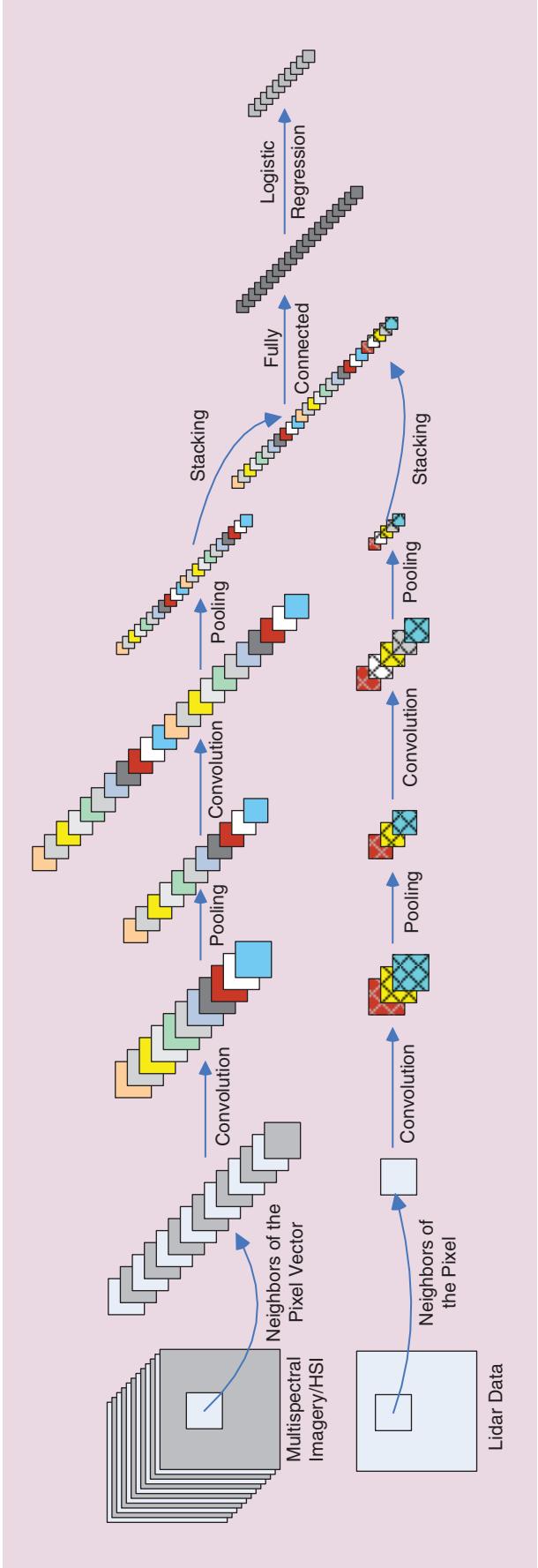
A joint sparse representation classification technique was proposed in [141] for multisource data fusion where the multisource data were weighted to have better sparse representation. The core idea of this scheme was based on sparse representation classification. Then, the minimum distance (in the sense of the  $\ell_2$  norm) between each sample and its sparse representation using subdictionaries (containing only training samples for one class) was used to allocate the class labels. However, in [141] the regularization term was weighted according to the data sources. Moreover, the method was translated into the kernel space using kernel tricks. The results for the composite kernel version can be found in Table 2 as ALWMJ-KSRC.

### DEEP LEARNING

Hyperspectral imaging often exhibits a nonlinear relation between the captured spectral information and the corresponding material. This nonlinear relation is the result of several factors, such as undesired scattering from other objects in the acquisition process, different atmospheric and geometric distortions, and intraclass variability of similar objects. This nonlinear characteristic is magnified when



**FIGURE 9.** Low-rank models. The use of DR2 is optional. However, the studies investigated in [136], [138], [142], and [143] recommend the consideration of this extra step to provide more accurate classification maps.



we deal with multisensor data. On the other hand, deep architectures are inherently able to extract high-level, hierarchical, and abstract features, which are usually invariant compared to the nonlinearities of the input data.

Deep learning is a fast-growing area in the remote sensing community, and its trace can be found in the research area of lidar and HSI data fusion as well. For instance, in [142], a classification method was developed to fuse spectral (HSI), spatial ( $EP_{HSI}$ ), and elevation features ( $EP_{lidar}$ ) using a 2D CNN and GBFF. The results for this approach can be found in Table 2 as CNNGBFF. To extract spatial and elevation features in a more effective manner than in [142], the authors of [153] employed two distinct CNN streams (as shown in Figure 10). The heterogeneous features obtained by the previous CNNs were then classified using a fully connected deep neural network. In [154], a three-stream CNN with a multisensor composite kernel was utilized to fuse spectral, spatial, and elevation features.

#### TRENDS OF HYPERSPECTRAL AND LIDAR FUSION

The following trends in the advancement of hyperspectral and lidar fusion need additional investigation:

- ▷ Because of the increased availability of large-scale DSMs and hyperspectral data, the further development of fast, accurate, and automatic classification and fusion techniques for the challenging task of transferable and large-area land cover mapping is of great interest.
- ▷ Investigation of the advanced machine-learning approaches (e.g., deep learning, domain adaptation, and transfer learning) for developing transferable classification/fusion schemes in areas with a limited number of training samples is in demand in our community.
- ▷ The development of sparse, low-rank, and subspace fusion approaches is another interesting line of research to address the high dimensionality of the heterogeneous features extracted from HSI and lidar to increase the quality of classification outputs.
- ▷ The researchers in [155] took the first step within the remote sensing community to simulate DSM from single optical imagery. This work opens a new path for researchers to increasingly modify this approach and design more sophisticated network architectures to produce more accurate elevation information from single optical images.

As stated previously, the classification and fusion of large-scale data (e.g., big data) are vitally important research lines and will be further discussed in the next section.

#### MULTITEMPORAL DATA FUSION

The use of multitemporal information is crucial for many important applications (from the analysis of slow and smoothly evolving phenomena [156], [157] to steep and abrupt changes [158]–[160]). Figure 11 shows a taxonomy of temporal phenomena that can be observed and detected by including the time variable in the analysis of remote sensing data.

As discussed in the introductory section, the recent availability of satellite constellations like Sentinel-1 and Sentinel-2 [161], which are characterized by the acquisition of fine-resolution images (up to 10 m) with a very short revisit time (a few days, depending on the latitude), is making the time dimension of satellite remote sensing images one of the most important sources of information to be exploited for the extraction of semantic content from a scene [158]. The time variable can expand the dimensionality of interest from 3D to 4D in space and time and can be exploited working with pairs of images, short time series, or long time series of either multispectral passive or SAR active images [158], [162]. Moreover, it is also possible to fuse together time series of multisensor images in a proper multitemporal sensor-fusion framework [163]–[165]. Fusion of temporal information with spatial and/or spectral/backscattering data from the images opens the possibility of also changing the perspective from the viewpoint of methodologies for data analysis. We can move from a representation of 3D cubes with multispectral images to 4D data structures, where the time variable adds new information (and also challenges) for information-extraction algorithms.

Analyzing the literature, the most widely addressed applications of multitemporal data are the analysis and classification of image time series and change detection [158], [162]. Nonetheless, there are many emerging topics based on the joint exploitation of the spectral, spatial, and temporal information of long and dense time series of fine spatial resolution images [166]–[169]. These topics were investigated widely in the past using coarse and medium spatial resolution images (i.e., at the scale of MODIS or MERIS/ENVISAT). However, with the availability of relatively dense time series of fine resolution images (e.g., Sentinel-1, Sentinel-2, and Landsat-8), it is now possible to develop studies at a dramatically increased resolution. For example, we can study the phenology of the vegetation in

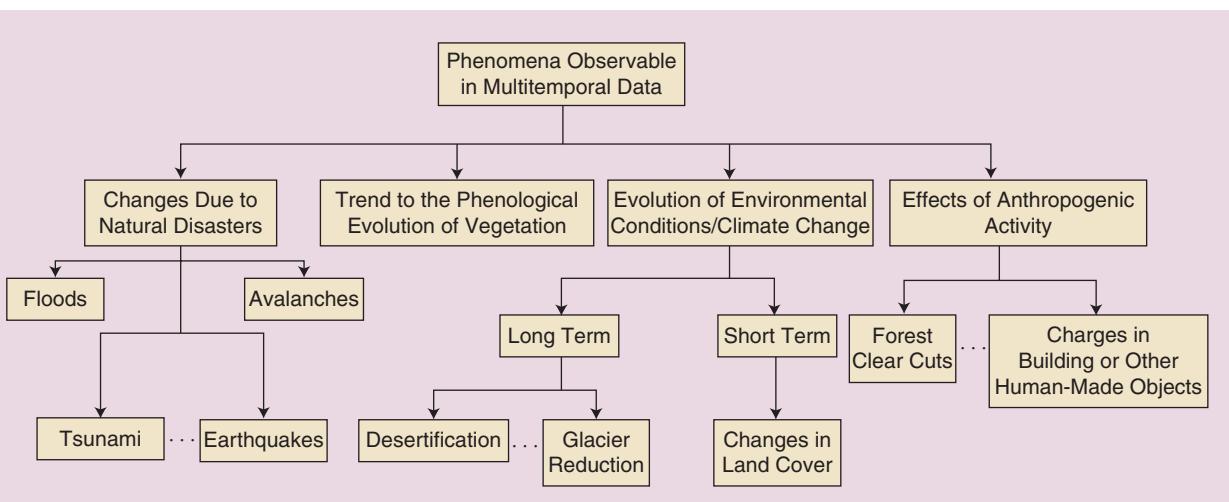
specific local areas or analyze the trends of vegetation in single fields of agricultural regions for precision farming applications [157], [166]. This can also be achieved by fusing the acquisitions of different satellites in a single multisensor time series.

Given the complexity and extent of the topic, in the following we analyze temporal information in relation to the classification problem, which is one of the most challenging lines of research and widely studied in the past [157], [167], [170]. First, we analyze the definition of the different classification problems with multitemporal data and briefly recall the methods presented in the literature for the solution of these problems. Then, we discuss the challenges related to multitemporal classification, especially from the viewpoint of the availability of labeled training data.

## MULTITEMPORAL INFORMATION IN CLASSIFICATION

Let us assume the availability of a set of multitemporal images (a time series with many images or at least a pair of images) acquired of the same geographical area at different times. The classification of these multitemporal data can be defined in different ways depending on the objective of the data analysis. The goal of the classification can be to generate

- ▶ a land cover map associated with the most recent image (acquisition) of a time series (or of a pair of acquisitions) [Figure 12(a)] [171]
- ▶ a land cover map for each item of the time series, thus producing a set of multitemporal land cover maps [Figure 12(b)] that also implicitly models the land cover transitions [172], [173]
- ▶ an annual or seasonal land cover map with classes that represent the behavior of the temporal signature of each pixel or region in the images in a year or season [Figure 12(c)] [174], [175].



**FIGURE 11.** An example of a taxonomy of the phenomena observable and detectable using temporal information in remote sensing data analysis.

These three definitions should result in significantly different classification approaches based on different assumptions. Unfortunately, in many cases in the literature and in the definition of application-oriented systems, the problems are not properly identified and modeled, with the implication that suboptimal results are obtained.

The use of temporal information in classification dates to the early 1980s. The first schemes used a stacked vector representation of the multitemporal data as input to the classifiers, resulting in so-called supervised direct multitemplate classification [159].

The main idea of such techniques is to characterize pixels by stacking the feature vectors of the images acquired at two (or more) times. Then the classification is carried out by training the classifiers to produce a map describing the land cover of only the most recent image. This is theoretically affordable under the assumption that both 1) there are no changes in the land cover between the considered image acquisition dates and 2) it is possible to properly model the complexity of the data distributions with respect to the classification methodology. The latter becomes critical when statistical Bayesian approaches are used.

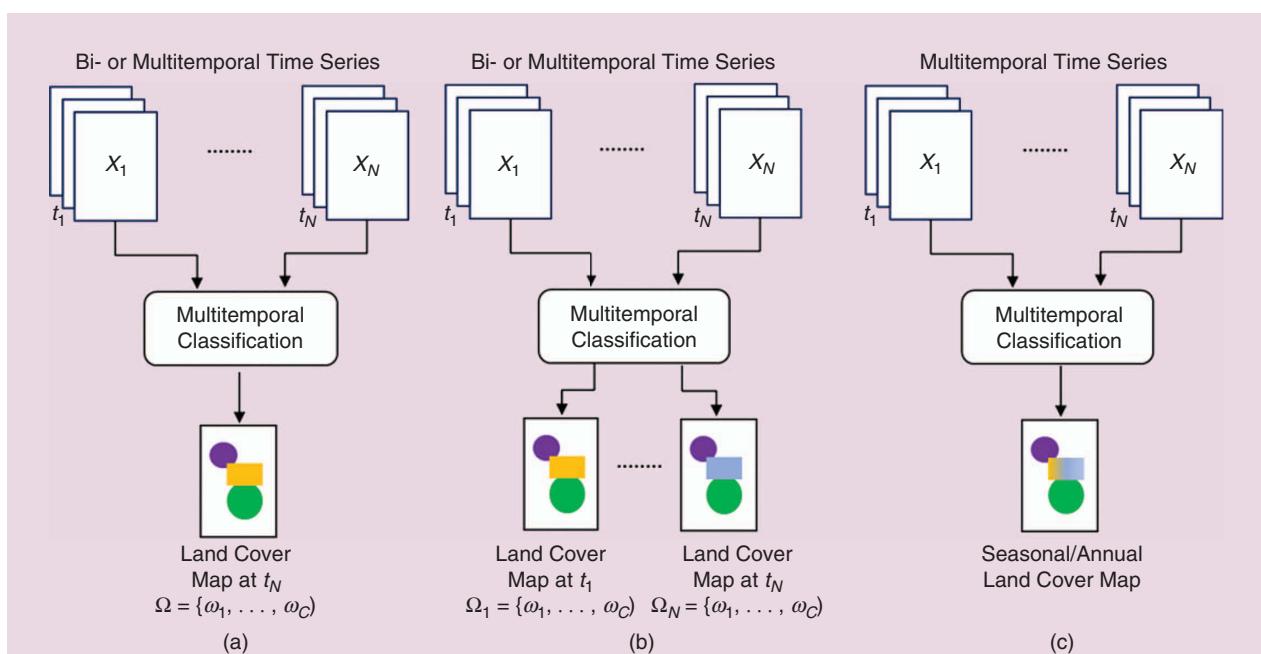
Another possible way of using multitemplate direct classification methods is to classify the land cover of each item of the time series, thus producing a land cover map for each available acquisition time. This allows one to explicitly identify land cover transitions and remove the assumption that there are no changes between the considered dates. However, a proper modeling of the change information requires the availability of labeled training data that can adequately represent the statistics of all possible class combinations, including those associated with the changes. This is seldom possible in real application scenarios.

Many methodologies have been developed to address these multitemplate direct classification issues. In a pioneering paper, Swain [176] introduced a very interesting approach to the classification of multitemporal data based on the cascade classification of paired images. He modeled the classification problem from a Bayesian perspective, introducing the temporal correlation in the classification process of images acquired at different times for linking class probabilities estimated on single images. In this way, it is possible in the multitemporal classification problem to decouple the modeling of the class distributions at each single date from the estimation of the temporal correlation between images.

Bruzzone et al. [171] developed and generalized this framework to the case of multitemporal and multisensor data, introducing a procedure to compound classification based on neural network classifiers' ability to properly merge in a Bayesian decision framework the distribution-free

## THE INTEGRATION OF MASSIVE GLOBAL-VIEW REMOTE SENSING DATA AND LOCAL, LIVE, LOCATION-BASED SOCIAL MEDIA DATA CAN OFFER NEW OPPORTUNITIES FOR SMART-CITY AND SMART-ENVIRONMENT APPLICATIONS.

under the assumption that both 1) there are no changes in the land cover between the considered image acquisition dates and 2) it is possible to properly model the complexity of the data distributions with respect to the classification methodology. The latter becomes critical when statistical Bayesian approaches are used.



**FIGURE 12.** A block scheme for achieving different goals in multitemporal classification: (a) a land cover map associated with the most recent image of a time series, (b) a land cover map for each item of the time series, and (c) an annual/seasonal land cover map with classes that represent the behavior of the temporal signature of each pixel/region in the time series.

estimations of the class parameters derived from both multispectral and SAR multitemporal images. This kind of fusion has been studied widely over the past two decades and developed in the context of different classification methodologies, including several neural models (e.g., multilayer perceptron neural networks and radial basis function networks), kernel methods (e.g., SVMs [177], [178]), and multiple classifier systems [175], [179], [180] (e.g., based on the fusion of neural and statistical classification algorithms). Also, the joint exploitation of the spatiotemporal information has been investigated, including Markov random fields in the modeling of the spatiotemporal context of multitemporal data.

Currently the challenge (still poorly addressed) is to exploit deep-learning architectures (e.g., CNNs) in the classification of multitemporal data [181], [182]. These architectures are intrinsically able to capture the spatiotemporal patterns in the definition of the classification model and, thus, increase the accuracy of the land cover/land cover transition maps. However, there is still a significant challenge in terms of defining theoretically sound and computationally affordable deep-learning architectures able to properly process multitemporal images. Indeed, the use of the 4D data structure sharply increases the complexity associated with deep-learning architectures and requires a number of training data that currently are far from being available.

### CHALLENGES IN MULTITEMPORAL CLASSIFICATION

The main challenges associated with the exploitation of the time information source in the classification of remote sensing data are related to the availability of adequate labeled samples for the definition of training sets suitable for the learning of supervised algorithms. The problem is to define statistically significant training sets able to represent the structured information content present in the data [172], [183]. This problem is much more critical than in the classification of single images, given the complexity associated with the possible combinations of land cover classes in the spatiotemporal domain. A proper modeling of the temporal information would require multitemporal ground reference samples (or reliably annotated multitemporal images) with samples that represent

- ▶ all of the multitemporal classes
- ▶ the interrelation among classes along the time series (e.g., land cover transitions or different kinds of changes), with a reliable statistic
- ▶ the high temporal and spatial variabilities in large scenes.

These constraints are very difficult to satisfy in real applications. For this reason, great attention has been and is still devoted to the use of methods that address the limitations of the real training set.

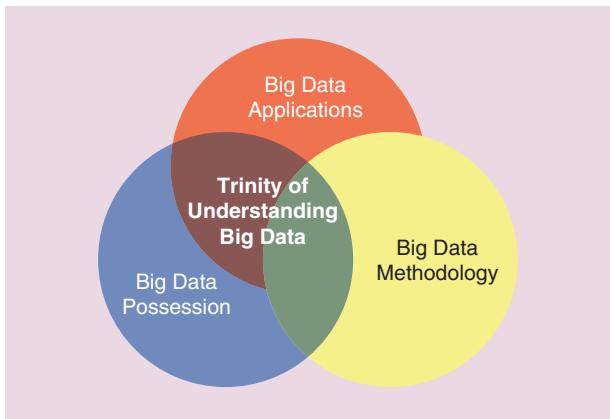
In this context, the scientific community's activities have been focused on semisupervised (also called *partially unsupervised*) classification methods [183]–[188]. These

techniques jointly exploit the available labeled training data and the distribution of the observed images to improve the modeling of the spatiotemporal properties of the analyzed time series. Early attempts to use these approaches in remote sensing are related to the use of the expectation-maximization algorithm in the context of land cover map updating with a maximum likelihood classifier [183]. This has been extended to use with the cascade and compound classification methods for the classification of bitemporal images [171], [184]. The approaches can integrate multispectral and SAR multitemporal data as well as multisensor images.

The problem of semisupervised learning with multitemporal data has been formulated in the more general theoretical problem of domain adaptation, for which different solutions can be found in the literature [189]. For example, the use of a semisupervised SVM has been widely investigated with different methodological implementations [177]. The use of active learning in the framework of compound classification for optimizing the definition of training data while minimizing the cost associated with their collection was proposed in [185]. The main idea was to collect ad hoc training samples in portions of the images where there is high multitemporal uncertainty on the class labels. Transfer learning approaches were proposed in [186]–[188], where change detection-based techniques were defined to propagate the labels of available data for a given image to the training sets of other images in the time series. The main observation at the basis of these techniques is that the available class labels can be propagated within the time series to all of the pixels that have not been changed between the considered acquisitions. In this way, unsupervised change detection can become a way to increase the amount of supervision that can be injected in the learning of a multitemporal classifier.

However, despite the large number of papers on semisupervised methods capable of capturing the information of 4D data structures, this is still a critical issue in the classification of multitemporal data and definitely an open issue in the framework of multitemporal classification with deep-learning architectures. In this last case, the challenges are related to decoupling, in the architecture of the network, the learning of the relation between the spatiotemporal patterns to achieve feasible requirements on the amount and the characteristics of training samples without significantly degrading the ability to extract the semantics of spatiotemporal patterns from the data. We expect that these crucial issues will be widely addressed in future years.

**THE CHALLENGES FOR  
FUSING REMOTE SENSING  
AND SOCIAL MEDIA DATA  
ARE SIMILAR TO THOSE  
IN A GENERAL BIG DATA  
PROBLEM, I.E., DATA  
REPRESENTATION, DATA  
ANALYSIS, AND DATA-  
ACCELERATED COMPUTING.**



**FIGURE 13.** A trinity for understanding big data, i.e., three facets of big data from different perspectives related to who owns the data, who has innovative big data methods and methodologies, and who needs big data applications [190]. This list can be further extended by adding big data visualization and big data accelerated computing.

### BIG DATA AND SOCIAL MEDIA

In the past decade, the topic of big data has become very important in many research areas, including remote sensing applications [190]. Every day, a massive number of remote sensing data is provided by a profusion of Earth observation spaceborne and airborne sensors from many different countries. In the near future, all-day, all-weather, and full-spectrum acquisition segment data sets will be provided by commercial satellites, such as the Jilin-1 constellation, which launched 10 fine-spatial-resolution satellites as of February 2018 and will have 60 satellites in orbit by 2020, with the ability to observe any arbitrary global point, with a 30-min revisit frequency [191]. Those video satellites, with an extremely fine temporal and spatial resolution, can effectively be exploited to monitor our location-based living environments as do couple-charged device cameras but on a much larger scale [192].

From a broader spatial perspective, new opportunities for humankind can be provided jointly by the remote sensing big data those satellites acquire along with social media data, providing local and live/real-time information to better monitor our living environment [192], especially in the applications of smart cities [193], [194], emergencies, environmental hazards [195], [196], and the like. On the other hand, based on unprecedented access to a huge number of remote sensing data, rapidly leading to a data-rich but knowledge-poor environment, new challenges can appear. Here, the semantic gap of remote

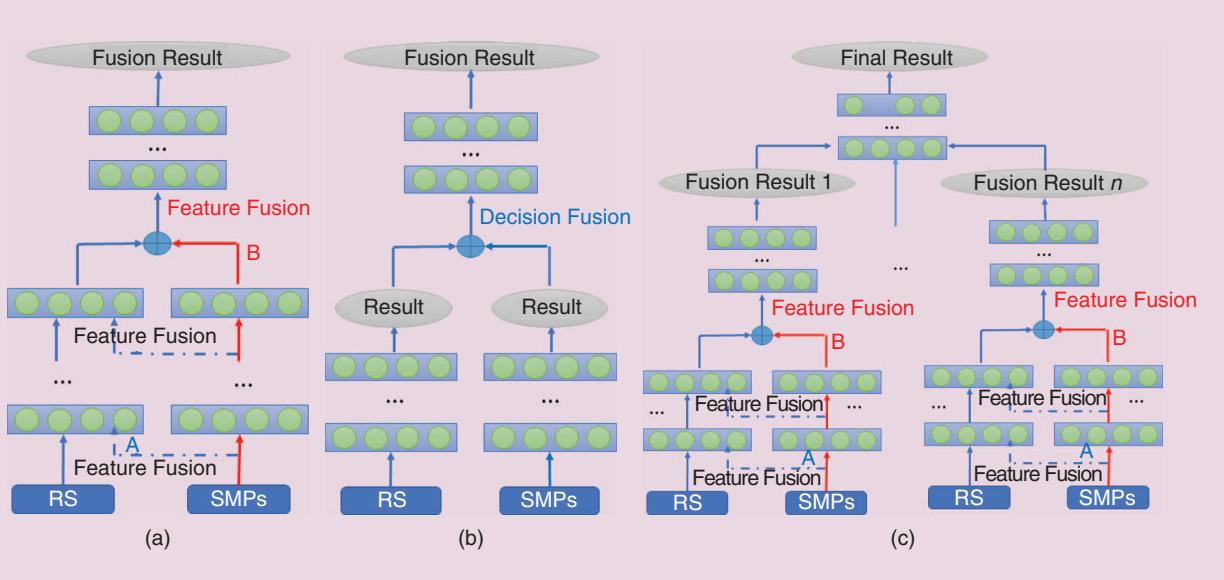
**REMOTE SENSING BIG DATA CAN PROVIDE RICH INFORMATION ON THE CONTENT AND LOCATION OF THE SCENES FOR AUGMENTED REALITY APPLICATIONS, 3D RECONSTRUCTION, INDOOR POSITIONING, AND SO ON.**

sensing data is usually caused by the lack of certain land cover or land use remote sensing categories on site. As an example, one can analyze the change in remote sensing images before and after floods. In this context, it is possible to roughly determine the damaged area using unsupervised learning algorithms, but it is difficult to assess the details (e.g., the damage to transportation infrastructure) [192].

Social media data, which provide one of the most important data sources from human activities, are usually composed of geolocated posts, tweets, photos, video, and audio with rich spatial information. With the fast development of computer technologies and Internet innovations, social media data are easily created by wearable and intelligent mobile devices equipped with GPS receivers. Those data can be disseminated quickly to social networks like Facebook, Twitter, YouTube, and Weibo and, in particular, to messaging apps like SnapChat, WhatsApp, and WeChat. Accordingly, the big data produced by the integration of massive global-view remote sensing data and local, live, location-based social media data can offer new opportunities for smart-city and smart-environment applications, with the ground reference collection through social sensing [197] and crowd sensing [198], in particular for hazards and disaster identification or tracking [192], [199]–[201].

To better analyze and utilize remote sensing big data with social media data, as in the definition of the connotations of big data in [190], big data can be expressed in the context of a trinity framework with three perspectives, i.e., owning data, data applications, and data methods. The big data trinity concept is illustrated in Figure 13. Each of the three perspectives has individual challenges but all share common concerns, discussed in detail in [190].

To derive the value of big data, combining remote sensing and social medial data, one of the most important challenges is determining how to process and analyze those data by novel methodologies. Because remote sensing data have significantly different properties than those of social media data, typical data fusion methods cannot be exploited directly for combining the two forms. Often, remote sensing data consist of multisource (laser, radar, optical, and so forth), multitemporal (collected on different dates), and multiresolution (different spatial resolution) data. Most remote sensing data are images. Social media data have a much wider array of formats, including images, videos, audio, and texts, with texts containing different types of textual information, such as geotagging, hashtags, posts, tweets, and Rich Site Summary. Nevertheless, the challenges for fusing remote sensing and social media data are similar to those in a general big data problem, i.e., data representation, data analysis, and data-accelerated computing [190]. However, advances in artificial intelligence (AI) techniques—in particular, deep neural networks—have merged data representation and data analysis into a unified AI model. In recent decades, high-performance computing has developed for



**FIGURE 14.** In the context of deep neural networks, an illustration of the integration of remote sensing (RS) images with SMPs: (a) feature-based fusion, (b) decision-based fusion, and (c) feature–decision-based fusion.

data-accelerated-based computing on big data platforms, such as Hadoop [202] and SPARK [203]. Notably, with advances in AI, graphics processing unit (GPU)-accelerated computing using a GPU and AI (especially, deep learning) chips have developed quickly in recent years for accelerating deep-learning computing in a heterogeneous platform by combining CPUs, GPUs, field-programmable gate arrays, and the like.

Social media data, such as photos with geotaggings, can be integrated with remote sensing data at the feature, decision, or feature–decision levels, in the context of deep neural networks. In feature-based data fusion, social media photos (SMPs) can be integrated into the same deep neural network to extract the features for further processing, as shown in Figure 14(a). Here, the feature extractor can have a deep architecture, and the features generated from SMPs can be integrated in each layer (or arbitrary layers) of the deep neural network. Nonetheless, features can be extracted individually from remote sensing images and SMPs by different deep neural networks, as shown in Figure 14(b). After that, the fused features can be sent to a deep neural network for further processing with feature convolutional layers, activation layers, pooling layers, and final classification to achieve a more reliable and accurate result.

In decision-based fusion, each deep neural network is designed to first extract the features of remote sensing images or SMPs; then, the classification result is generated by individual features. At the decision level, those results provided by the remote sensing and social media data are integrated into a unified deep neural network, as shown in Figure 14(b). In this case, social media data can have diverse types (images, texts, and so on), such that different types of social media data can build

**TABLE 3. OVERALL ACCURACIES COMPARED IN TERMS OF THE TRAINING DATA SETS LABELED BY VOLUNTEERS WITH AND WITHOUT SMPs.**

MODELS	OVERALL ACCURACY (%)
With SMPs	FCN
	CNN
	SVM
Without SMPs	78.91
	74.85
	62.40
Without SMPs	71.23
	65.72
	61.16

different deep neural networks for further decision fusion. By combining the two properties of feature-based and decision-based fusion strategies, feature–decision-based fusion can be easily derived based on deep neural networks. The challenge is how to design a unified deep neural network model to efficiently and accurately fuse heterogeneous data.

Except for directly modeling a deep-learning algorithm by integrating remote sensing and social media data, SMPs can be used to label remote scene images, especially for fine spatial resolution data. For instance, SMPs with the same positions as the fine resolution remote sensing data can be acquired to help volunteers with no professional knowledge effectively label remote sensing scene images. To validate the effectiveness of using SMPs for labeling, fine spatial resolution remote sensing images in Frankfurt, Germany, acquired by a Jilin-1 satellite are utilized for remote sensing image classification. The classification models are trained on training data sets labeled with and without SMPs. Table 3 shows the prediction results on the test data. Both the fully convolutional network (FCN)

model [204] and the CNN model are constructed based on the pretrained ImageNet VGG-16 network [205] with cross entropy loss. The SVM model with radial basis function kernels is adopted for a further comparison. Figure 15 illustrates several classification maps obtained by SVM, CNN, and FCN with and without using SMPs.

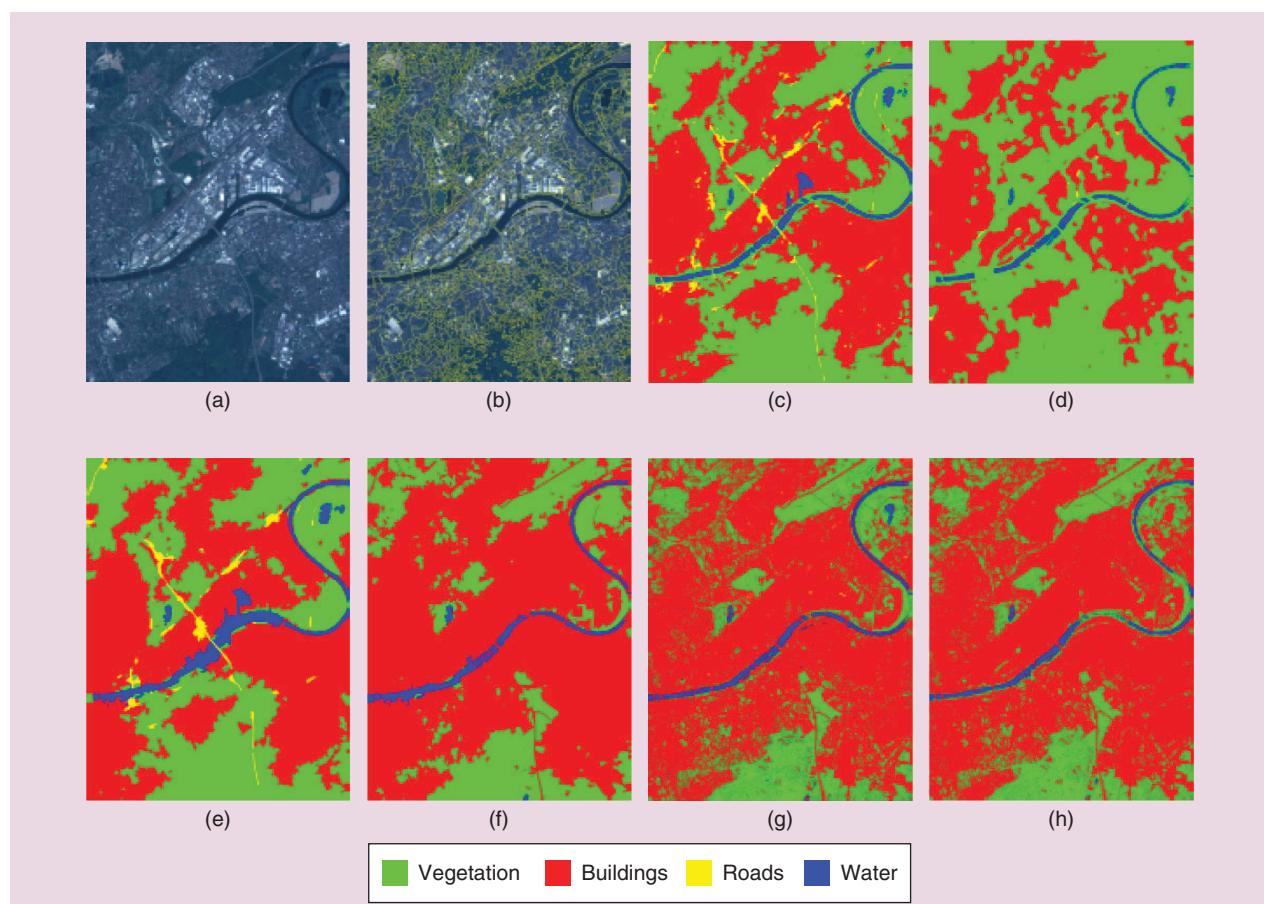
**THERE ARE STILL MANY DOORS OPEN FOR FURTHER INVESTIGATION, BOTH FROM THE THEORETICAL AND APPLICATION PERSPECTIVES, EVEN THOUGH THE FIELD OF REMOTE SENSING DATA FUSION IS MATURE.**

human-made disasters, such as factory explosions. In addition, remote sensing big data can provide rich information on the content and location of the scenes for augmented reality applications, 3D reconstruction, indoor positioning, and so on.

## CONCLUSIONS

The ever-growing availability of data captured by different sensors, coupled with advances in methodological approaches and computational tools, makes it desirable to fuse considerably heterogeneous complementary data sets to increase the efficacy and efficiency of the remotely sensed data processing approaches with respect to the problems at hand. The field of multisensor and multitemporal data fusion for remotely sensed imagery is enormously broad, which makes it challenging to treat it comprehensively in one literature review. This article focused particularly on advances in multisource and multitemporal data fusion approaches with respect to different research communities because the methods for fusing different modalities have expanded along different paths with respect to each research area. In this context, several vibrant fusion topics, including pansharpening and resolution enhancement, point cloud data fusion, hyperspectral and lidar data fusion, multitemporal data fusion, and big data and social media, were detailed, and their corresponding challenges and possible future research directions were outlined and discussed.

As demonstrated through the challenges and possible future research noted in each section, there are



**FIGURE 15.** Several classification maps obtained by SVM, CNN, and FCN with or without using SMPs: (a) a test image, (b) segmentation, (c) FCN + SMPs, (d) FCN, (e) CNN + SMPs, (f) CNN, (g) SVM + SMPs, and (h) SVM.

still many doors open for further investigation, both from the theoretical and application perspectives, even though the field of remote sensing data fusion is mature. We hope this review reveals new possibilities for readers to further explore the remaining issues in developing sophisticated fusion approaches suitable for the applications at hand.

## ACKNOWLEDGMENTS

This research was supported by the Helmholtz-Zentrum Dresden-Rossendorf High Potential Program (specifically supporting Pedram Ghamisi), the Natural Science Foundation of China under contract 71331005, and the State Key Research and Development Program of China under contract 2016YFE0100300. In addition, we would like to thank the National Center for Airborne Laser Mapping at the University of Houston for providing the CASI Houston data set and the IEEE Geoscience and Remote Sensing Society's Image Analysis and Data Fusion Technical Committee for organizing the 2013 Data Fusion Contest. Moreover, we wish to thank Dr. Claas Grohnfeldt for processing J-SparseFI-HM.

## AUTHOR INFORMATION

**Pedram Ghamisi** (p.ghamisi@gmail.com) received his B.Sc. degree in civil (survey) engineering from the Tehran South Campus, Azad University, Iran, in 2008; his M.Sc. (first-class honors) degree in remote sensing from the K.N. Toosi University of Technology, Tehran, Iran, in 2012; and his Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, in 2015. Since early 2018, he has worked as the head of the machine-learning group at Helmholtz-Zentrum Dresden-Rossendorf. He received the Best Researcher Award for M.Sc. students from the K.N. Toosi University of Technology during the 2010–2011 academic year. In 2017, he won the Data Fusion Contest organized by the Image Analysis and Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society. He was the winner of the 2017 IEEE Geoscience and Remote Sensing Letters Best Reviewer Prize. He received the prestigious Alexander von Humboldt Fellowship in 2015 and is an associate editor of *IEEE Geoscience and Remote Sensing Letters*. He is a Senior Member of the IEEE. For more info, please see <http://pedram-ghamisi.com/>.

**Behnoor Rasti** (behnoor@hi.is) received his B.Sc. and M.Sc. degrees in electrical and electronics engineering from the University of Guilan, Rasht, Iran, in 2006 and 2009, respectively, and his Ph.D. degree in electrical and computer engineering from the Department of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland, in 2014. From 2015 to 2016, he was a post-doctoral researcher with the University of Iceland. From 2016 to 2018, he was with the Keilir Institute of Technology, Reykjanesbaer, Iceland. He is currently with the Department of Electrical and Computer Engineering, University of Iceland. His research interests include signal

and image processing, hyperspectral image analysis, remote sensing data fusion, biomedical engineering, biomedical data analysis, control systems, and robotics. He is a Member of the IEEE.

**Naoto Yokoya** (naoto.yokoya@riken.jp) received his M.Sc. and Ph.D. degrees in aerospace engineering from the University of Tokyo in 2010 and 2013, respectively. From 2012 to 2013, he was a research fellow with the Japan Society for the Promotion of Science, Tokyo. From 2013 to 2017, he was an assistant professor with the University of Tokyo. From 2015 to 2017, he was an Alexander von Humboldt Research Fellow with the German Aerospace Center, Oberpfaffenhofen, and the Technical University of Munich, Germany. Since 2018, he has been the leader of the Geo-informatics Unit at the RIKEN Center for Advanced Intelligence Project, Tokyo. His research interests include image analysis and data fusion in remote sensing. In 2017, he won the Data Fusion Contest organized by the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society (GRSS). He is currently a cochair of the IEEE GRSS IADF TC and the secretary of the IEEE GRSS All Japan Joint Chapter. He is a Member of the IEEE.

**Qunming Wang** received his Ph.D. degree from Hong Kong Polytechnic University, China, in 2015. He is now a professor in the College of Surveying and Geo-Informatics at Tongji University, Shanghai, China. He was a lecturer (assistant professor) and senior research associate in the Lancaster Environment Centre at Lancaster University, United Kingdom, from 2017 to 2018 and 2015 to 2017, respectively. His three-year Ph.D. study was supported by the Hong Kong Ph.D. fellowship, and his Ph.D. thesis was awarded as the outstanding thesis in the faculty. His research interests focus on remote sensing, image processing, and geostatistics. He has published more than 40 peer-reviewed articles in international journals such as *Remote Sensing of Environment*, *IEEE Transactions on Geoscience and Remote Sensing* and *ISPRS Journal of Photogrammetry and Remote Sensing*. He serves as a reviewer for 30 international journals, including almost all international journals in remote sensing. He is now an associate editor for *Photogrammetric Engineering and Remote Sensing and Computers and Geosciences*. He is a Member of the IEEE.

**Bernhard Höfle** (hoefle@uni-heidelberg.de) received his Ph.D. degree with excellence in the field of physical geography and geoinformatics at the University of Innsbruck, Austria, in 2007. He was a junior professor of GIScience and 3D spatial data processing at the Institute of Geography, Heidelberg University, Germany, from 2011 to 2017. In 2017, he became professor of 3D spatial data processing at the Institute of Geography and head of the 3DGeo research group at the Interdisciplinary Center for Scientific Computing at Heidelberg University, Germany. Since 2014, he has been the cochair of the working group Geoinformatics of the German Society for Photogrammetry, Remote Sensing, and Geoinformation. His research interests include the

development of computational methods for 3D geospatial data processing using the full sensor data streams, such as lidar backscatter information, and the extraction of geoinformation from 3D/4D geodata acquired by different sensor systems.

**Lorenzo Bruzzone** received his M.S. degree in electronic engineering (summa cum laude) and his Ph.D. degree in telecommunications from the University of Genoa, Italy, in 1993 and 1998, respectively. He is currently a full professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, and digital communications. He is the founder and director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His current research interests are in the areas of remote sensing, radar and synthetic aperture radar, signal processing, machine learning, and pattern recognition. He was the recipient of many international and national honors and awards, including the recent IEEE Geoscience and Remote Sensing Society 2015 Outstanding Service Award and the IEEE International Geoscience and Remote Sensing Symposium Prize Paper Award for 2017 and 2018. He is the founder of *IEEE Geoscience and Remote Sensing Magazine*, for which he was the editor-in-chief between 2013 and 2017. He is a Fellow of the IEEE.

**Francesca Bovolo** received her B.S. and M.S. degrees in telecommunication engineering (summa cum laude) and her Ph.D. degree in communication and information technologies from the University of Trento, Italy, in 2001, 2003, and 2006, respectively, where she remained as a research fellow until June 2013. She is the founder and head of the Remote Sensing for Digital Earth unit at Fondazione Bruno Kessler, Trento, Italy, and a member of the Remote Sensing Laboratory in Trento. Her main research interests include multitemporal remote sensing image analysis; change detection in multispectral, hyperspectral, and synthetic aperture radar images and very-high resolution images; time series analysis; content-based time series retrieval; domain adaptation; and lidar and radar sounders. She was the publication chair for the 2015 IEEE International Geoscience and Remote Sensing Symposium. She is the cochair of the Society of Photographic Instrumentation Engineers International Conference on Signal and Image Processing for Remote Sensing. She is a Senior Member of the IEEE.

**Mingmin Chi** (mmchi@fudan.edu.cn) received her B.S. and M.S. degrees in electrical engineering from Changchun University of Science and Technology, China, and Xiamen University, China, in 1998 and 2002, respectively. She received her Ph.D. degree in computer science from the University of Trento, Italy, in 2006. Currently, she is an associate professor in the School of Computer Science, Fudan University, Shanghai, China, and a group leader at the Shanghai Key Laboratory of Data Science, China. She is a guest editor for the special issue on big data in remote sensing for *IEEE Journal of Selected Topics in Applied Earth*

*Observations and Remote Sensing* and for the special issue on analysis of big data in remote sensing for *Remote Sensing*. Her research interests include data science, big data, and machine learning with applications to astronomy, remote sensing, computer vision, and natural language processing. She is a Senior Member of the IEEE.

**Katharina Anders** (katharina.anders@uni-heidelberg.de) received her B.Sc. and M.Sc. degrees in geography at Heidelberg University, Germany, with a major in geoinformatics. She is currently pursuing her Ph.D. degree in the 3D Geospatial Data Processing Group at the Institute of Geography and the Interdisciplinary Center for Scientific Computing at Heidelberg University, Germany. Her research interests include geoinformation extraction from 3D point cloud data and methods for geomorphic change analysis using lidar time series.

**Richard Gloaguen** (r.gloaguen@hzdr.de) received his Ph.D. degree in marine geosciences "Communitatis Europae" from the University of Western Brittany, Brest, France, in collaboration with the Royal Holloway University of London and Göttingen University, Germany, in 2000. He was a post-doctoral research associate at the Royal Holloway University of London from 2000 to 2003. He led the Remote Sensing Group at the University Bergakademie Freiberg, Germany, from 2003 to 2013. Since 2013, he has been leading the Exploration Technology division at the Helmholtz-Institute Freiberg for Resource Technology. He is currently involved in unmanned aerial vehicle-based hyperspectral imaging, laser-induced fluorescence, and remote sensing-based tectonic geomorphology. His research interests include multisource multiscale remote sensing integration. He is a Member of the IEEE.

**Peter M. Atkinson** received his B.Sc. degree in geography from the University of Nottingham, United Kingdom, in 1986 and his Ph.D. degree from the University of Sheffield, United Kingdom, in 1990. He received his M.B.A. degree from the University of Southampton, United Kingdom, in 2012. He is currently a visiting professor at Queen's University Belfast, United Kingdom, the University of Southampton, and the Chinese Academy of Sciences, Beijing. In 2015, he was the Belle van Zuylen chair at Utrecht University, The Netherlands, and was the 2016 laureat of the Peter Burrough Award of the International Spatial Accuracy Research Organisation. His research interests include remote sensing, spatial data science, spatial (and space-time) statistics, and spatially explicit dynamic models applied to a range of environmental science and socioeconomic problems. He has published more than 270 peer-reviewed articles in international scientific journals and more than 50 refereed book chapters. He is an associate editor for *Computers and Geosciences* and sits on the editorial boards of several journals, including *Geographical Analysis*, *Spatial Statistics*, the *International Journal of Applied Earth Observation and Geoinformation*, and *Environmental Informatics*.

**Jón Atli Benediktsson** (benedikt@hi.is) received his Cand. Sci. degree in electrical engineering from the University of

Iceland, Reykjavik, in 1984 and his M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, Indiana, in 1987 and 1990, respectively. He is the president and rector of the University of Iceland, Reykjavik, where he is also a professor of electrical and computer engineering. He is a cofounder of the biomedical start-up company Oxymap ([www.oxymap.com](http://www.oxymap.com)). In 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society (GRSS) and the Outstanding Electrical and Computer Engineering Award from the School of Electrical and Computer Engineering, Purdue University, in 2016. He was a corecipient of the 2012 IEEE Transactions on Geoscience and Remote Sensing Paper Award, and, in 2013, he was a corecipient of the GRSS Highest Impact Paper Award. In 2013, he received the IEEE/Association of Chartered Engineers in Iceland (VFI) Electrical Engineer of the Year Award. He is a fellow of the International Society for Optics and Photonics and a Fellow of the IEEE.

## REFERENCES

- [1] M. Appel, F. Lahn, W. Buytaert, and E. Pebesma, "Open and scalable analytics of large earth observation datasets: From scenes to multidimensional arrays using SCIDB and GDAL," *ISPRS J. Photogrammetry Remote Sens.*, vol. 138, pp. 47–56, Apr. 2018.
- [2] N. Audebert, B. L. Saux, and S. Lefvre, "Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 140, pp. 20–32, June 2018.
- [3] J. E. Ball, D. T. Anderson, and C. S. Chan, "Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017. doi: 10.1117/1.JRS.11.042609.
- [4] J. Xia, N. Yokoya, and A. Iwasaki, "Fusion of hyperspectral and lidar data with a novel ensemble classifier," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 6, pp. 957–961, 2018.
- [5] M. Eslami and A. Mohammadzadeh, "Developing a spectral-based strategy for urban object detection from airborne hyperspectral TIR and visible data," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 5, pp. 1808–1816, 2016.
- [6] J. Zhu, J. Hu, S. Jia, X. Jia, and Q. Li, "Multiple 3-D feature fusion framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1873–1886, 2018.
- [7] L. Estes et al., "A platform for crowdsourcing the creation of representative, accurate landcover maps," *Environ. Modelling Software*, vol. 80, pp. 41–53, June 2016.
- [8] R. Foulser-Piggott, R. Spence, R. Eguchi, and A. King, "Using remote sensing for building damage assessment: Geocan study and validation for 2011 Christchurch earthquake," *Earthquake Spectra*, vol. 32, no. 1, pp. 611–631, 2016.
- [9] J. Li, J. A. Benediktsson, B. Zhang, T. Yang, and A. Plaza, "Spatial technology and social media in remote sensing: A survey," *Proc. IEEE*, vol. 105, no. 10, pp. 1855–1864, 2017.
- [10] N. Yokoya et al., "Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS Data Fusion Contest," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 11, no. 5, pp. 1363–1377, 2018.
- [11] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, Nov. 2016.
- [12] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, Jan. 2017.
- [13] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, July 2018.
- [14] R. Booyens, "Multiscale hyperspectral remote sensing of REE deposits in southern Africa," M.S. thesis, Univ. Witwatersrand, Johannesburg, South Africa, 2018.
- [15] S. Nieland, N. Moran, B. Kleinschmit, and M. Frster, "An ontological system for interoperable spatial generalisation in biodiversity monitoring," *Comput. Geosci.*, vol. 84, pp. 86–95, Nov. 2015.
- [16] J. P. Kerekes and D. A. Landgrebe, "Parameter trade-offs for imaging spectroscopy systems [Remote Sensing]," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 1, pp. 57–65, 1991.
- [17] G. Joseph, "How well do we understand earth observation electro-optical sensor parameters?" *ISPRS J. Photogrammetry Remote Sens.*, vol. 55, no. 1, pp. 9–12, 2000.
- [18] W. Zhao, M. Tamura, and H. Takahashi, "Atmospheric and spectral corrections for estimating surface albedo from satellite data using 6s code," *Remote Sens. Environ.*, vol. 76, no. 2, pp. 202–212, 2001.
- [19] M. Parente, J. T. Clark, A. J. Brown, and J. L. Bishop, "End-to-end simulation and analytical model of remotesensing systems: Application to crism," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 3877–3888, 2010.
- [20] X. He and X. Xu, "Physically based model for multispectral image simulation of earth observation sensors," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 10, no. 5, pp. 1897–1908, 2017.
- [21] S. Lefvre, D. Tuia, J. D. Wegner, T. Produt, and A. S. Nassar, "Toward seamless multiview scene analysis from satellite to street level," *Proc. IEEE*, vol. 105, no. 10, pp. 1884–1899, 2017.
- [22] Q. Wang and P. M. Atkinson, "The effect of the point spread function on sub-pixel mapping," *Remote Sens. Environ.*, vol. 193, pp. 127–137, May 2017.
- [23] P. M. Atkinson, "Downscaling in remote sensing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 22, pp. 106–114, June 2013.
- [24] A. Garzelli, "A review of image fusion algorithms based on the super-resolution paradigm," *Remote Sens.*, vol. 8, no. 10, p. 797, 2016.
- [25] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, 2018.
- [26] A. Z. Mangion, N. Cressie, and C. Shumack, "On statistical approaches to generate Level 3 products from satellite remote sensing retrievals," *Remote Sens.*, vol. 10, no. 1, 2018. doi: 10.3390/rs10010155.
- [27] Q. Wang, W. Shi, P. M. Atkinson, and Y. Zhao, "Downscaling modis images with area-to-point regression kriging," *Remote Sens. Environ.*, vol. 166, no. Supplement C, pp. 191–204, 2015.

- [28] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Econ. Geog.*, vol. 46, pp. 234–240, June 1970.
- [29] C. Zhang et al., "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 140, pp. 133–144, June 2018.
- [30] A. Shelestov, M. Lavreniuk, N. Kussul, A. Novikov, and S. Skakun, "Exploring Google Earth Engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping," *Front. Earth Sci.*, vol. 5, p. 17, Feb. 2017.
- [31] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson, "Fusion of Sentinel-2 images," *Remote Sens. Environ.*, vol. 187, no. Supplement C, pp. 241–252, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0034425716304023>
- [32] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.*, vol. 172, no. Supplement C, pp. 165–177, 2016.
- [33] J. Wei, L. Wang, P. Liu, X. Chen, W. Li, and A. Y. Zomaya, "Spatiotemporal fusion of modis and landsat-7 reflectance images via compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7126–7139, 2017.
- [34] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, no. Supplement C, pp. 31–42, 2018.
- [35] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, 2017.
- [36] P. Ghamisi et al., "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, 2017.
- [37] Y. Zhang, P. M. Atkinson, X. Li, F. Ling, Q. Wang, and Y. Du, "Learning-based spatial-temporal superresolution mapping of forest cover with modis images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 600–614, 2017.
- [38] A. O. Onojeghuo, G. A. Blackburn, Q. Wang, P. M. Atkinson, D. Kindred, and Y. Miao, "Mapping paddy rice fields by applying machine learning algorithms to multi-temporal Sentinel-1a and Landsat data," *Int. J. Remote Sens.*, vol. 39, no. 4, pp. 1042–1067, 2018.
- [39] Y. Feng, D. Lu, E. F. Moran, L. V. Dutra, M. F. Calvi, and M. A. F. de Oliveira, "Examining spatial distribution and dynamic change of urban land covers in the Brazilian Amazon using multitemporal multisensor high spatial resolution satellite imagery," *Remote Sens.*, vol. 9, no. 4, p. 381, 2017.
- [40] Z. Zhu, "Change detection using Landsat time series: A review of frequencies, preprocessing, algorithms, and applications," *ISPRS J. Photogrammetry Remote Sens.*, vol. 130, pp. 370–384, Aug. 2017.
- [41] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, 2017.
- [42] R. Welch, "Merging multiresolution SPOT HRV and Landsat TM data," *Photogrammetric Eng. Remote Sens.*, vol. 53, no. 3, pp. 301–303, 1987.
- [43] P. Kwarteng and A. Chavez, "Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis," *Photogrammetric Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, 1989.
- [44] W. Carper, T. M. Lillesand, and P. W. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogrammetric Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.
- [45] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation," *Photogrammetric Eng. Remote Sens.*, vol. 66, no. 1, pp. 49–61, 2000.
- [46] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6011875, Jan. 4, 2000.
- [47] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, 2002.
- [48] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS + Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, 2007.
- [49] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, 2008.
- [50] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, 2010.
- [51] D. A. Yocky, "Multiresolution wavelet decomposition image merger of Landsat Thematic Mapper and SPOT panchromatic data," *Photogrammetric Eng. Remote Sens.*, vol. 62, no. 9, pp. 1067–1074, 1996.
- [52] J. Zhou, D. Civco, and J. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.
- [53] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, 1999.
- [54] J. G. Liu, "Smoothing Filter-based Intensity Modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, 2000.
- [55] X. Otazu, M. González-Audicana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods: Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, 2005.
- [56] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and PAN imagery," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, 2006.

- [57] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion*, vol. 8, no. 2, pp. 143–156, 2007.
- [58] E. Pardo-Iguzquiza, M. Chica-Olmo, and P. M. Atkinson, "Downscaling cokriging for image sharpening," *Remote Sens. Environ.*, vol. 102, no. 1, pp. 86–98, 2006. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S003442570600068X>
- [59] P. M. Atkinson, E. Pardo-Iguzquiza, and M. Chica-Olmo, "Downscaling cokriging for super-resolution mapping of continua in remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 2, pp. 573–580, 2008.
- [60] E. Pardo-Iguzquiza, V. F. Rodriguez-Galiano, M. Chica-Olmo, and P. M. Atkinson, "Image fusion by spatially adaptive filtering using downscaling cokriging," *ISPRS J. Photogrammetry Remote Sens.*, vol. 66, no. 3, pp. 337–346, 2011.
- [61] M. H. R. Sales, C. M. Souza, and P. C. Kyriakidis, "Fusion of modis images using kriging with external drift," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2250–2259, 2013.
- [62] Q. Wang, W. Shi, and P. M. Atkinson, "Area-to-point regression kriging for pan-sharpening," *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, no. Supplement C, pp. 151–165, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271616000496>
- [63] Q. Wang, W. Shi, P. M. Atkinson, and Q. Wei, "Approximate area-to-point regression kriging for fast hyperspectral image sharpening," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 10, no. 1, pp. 286–295, 2017.
- [64] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, 2004.
- [65] M. T. Eismann and R. C. Hardie, "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 455–465, 2005.
- [66] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, 2012.
- [67] M. Simões, J. B. Dias, L. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, 2015.
- [68] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multiband images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, 2015.
- [69] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multiband images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, 2015.
- [70] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, 2011.
- [71] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, 2013.
- [72] Q. Wei, J. M. B. Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, 2015.
- [73] W. Sun, B. Chen, and D. Messinger, "Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images," *Opt. Eng.*, vol. 53, no. 1, 2014. doi: 10.1117/1.OE.53.1.013107.
- [74] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hyper-sharpening: A first approach on SIM-GA data," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 3008–3024, 2015.
- [75] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2015, pp. 3586–3594.
- [76] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, 2014.
- [77] C. Grohnfeldt, "Multi-sensor data fusion for multi- and hyperspectral resolution enhancement based on sparse representations," Ph.D. dissertation, Tech. Univ. Munich, Germany, 2017.
- [78] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. 3rd Conf. Fusion Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images*, 2000, pp. 99–103.
- [79] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, 2009.
- [80] F. Gao et al., "Fusing Landsat and MODIS data for vegetation monitoring," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 47–60, 2015.
- [81] H. K. Zhang, B. Huang, M. Zhang, K. Cao, and L. Yu, "A generalization of spatial and temporal fusion methods for remotely sensed surface parameters," *Int. J. Remote Sens.*, vol. 36, no. 17, pp. 4411–4445, 2015.
- [82] B. Chen, B. Huang, and B. Xu, "Comparison of spatiotemporal fusion models: A review," *Remote Sens.*, vol. 7, no. 2, pp. 1798–1835, 2015.
- [83] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, 2006.
- [84] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, 2010.
- [85] I. V. Emelyanova, T. R. McVicar, T. G. V. Niel, L. T. Li, and A. I. van Dijk, "Assessing the accuracy of blending Landsat-MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection," *Remote Sens. Environ.*, vol. 133, no. Supplement C, pp. 193–209, 2013.
- [86] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716, 2012.

- [87] H. Song and B. Huang, "Spatiotemporal satellite image fusion through one-pair image learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1883–1896, 2013.
- [88] X. Liu, C. Deng, S. Wang, G. B. Huang, B. Zhao, and P. Lauren, "Fast and accurate spatiotemporal fusion based upon extreme learning machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 2039–2043, 2016.
- [89] V. Moosavi, A. Talebi, M. H. Mokhtari, S. R. F. Shamsi, and Y. Niazi, "A wavelet-artificial intelligence fusion approach (WAI-FA) for blending Landsat and MODIS surface temperature," *Remote Sens. Environ.*, vol. 169, no. Supplement C, pp. 243–254, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0034425715301036>
- [90] M. Das and S. K. Ghosh, "Deep-STEP: A deep learning approach for spatiotemporal prediction of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1984–1988, 2016.
- [91] J. Amoros-Lopez, L. Gomez-Chova, L. Alonso, L. Guanter, J. Moreno, and G. Camps-Valls, "Regularized multiresolution spatial unmixing for ENVISAT/MERIS and Landsat/TM image fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 844–848, 2011.
- [92] J. Amoros-Lopez, L. Gomez-Chova, L. Alonso, L. Guanter, R. Zurita-Milla, J. Moreno, and G. Camps-Valls, "Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 23, no. Supplement C, pp. 132–141, 2013.
- [93] R. Zurita-Milla, J. G. P. W. Clevers, and M. E. Schaepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453–457, 2008.
- [94] Y. T. Mustafa, V. A. Tolpekin, and A. Stein, "Improvement of spatio-temporal growth estimates in heterogeneous forests using Gaussian Bayesian networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4980–4991, 2014.
- [95] R. Zurita-Milla, G. Kaiser, J. Clevers, W. Schneider, and M. Schaepman, "Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics," *Remote Sens. Environ.*, vol. 113, no. 9, pp. 1874–1885, 2009.
- [96] M. Wu, Z. Niu, C. Wang, C. Wu, and L. Wang, "Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model," *J. Appl. Remote Sens.*, vol. 6, no. 1, 2012. doi: 10.1111/1.JRS.6.063507.
- [97] C. M. Gevaert and F. J. Garca-Haro, "A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion," *Remote Sens. Environ.*, vol. 156, no. Supplement C, pp. 34–44, 2015.
- [98] B. Huang and H. Zhang, "Spatio-temporal reflectance fusion via unmixing: Accounting for both phenological and land-cover changes," *Int. J. Remote Sens.*, vol. 35, no. 16, pp. 6213–6233, 2014.
- [99] Y. Xu, B. Huang, Y. Xu, K. Cao, C. Guo, and D. Meng, "Spatial and temporal image fusion via regularized spatial unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1362–1366, 2015.
- [100] D. Xie et al., "An improved STARFM with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions," *Sensors*, vol. 16, no. 2, p. 207, 2016.
- [101] Q. Wang, Y. Zhang, A. O. Onojehuo, X. Zhu, and P. M. Atkinson, "Enhancing spatio-temporal fusion of MODIS and Landsat data by incorporating 250 m MODIS data," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 10, no. 9, pp. 4116–4123, 2017.
- [102] IEEE Geoscience and Remote Sensing Society. GRSS Data and Algorithm Standard Evaluation website. [Online]. Available: <http://dase.grss-ieee.org/>
- [103] A. V. Vo et al., "Processing of extremely high resolution lidar and RGB data: Outcome of the 2015 IEEE GRSS Data Fusion Contest," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 12, pp. 5560–5575, 2016.
- [104] J. Zhang and X. Lin, "Advances in fusion of optical imagery and lidar point cloud applied to photogrammetry and remote sensing," *Int. J. Image Data Fusion*, vol. 8, no. 1, pp. 1–31, 2017.
- [105] F. Remondino, M. G. Spera, E. Nocerino, F. Menna, and F. Nex, "State of the art in high density image matching," *Photogrammetric Rec.*, vol. 29, no. 146, pp. 144–166, 2014.
- [106] X. X. Zhu and M. Shahzad, "Facade reconstruction using multiview spaceborne TomoSAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3541–3552, 2014.
- [107] M. Hämmerle and B. Höfle, "Mobile low-cost 3D camera maize crop height measurements under field conditions," *Precision Agric.*, vol. 19, no. 4, pp. 630–647, Oct. 2017.
- [108] J. Shan and C. Toth, *Topographic Laser Ranging and Scanning: Principles and Processing*. Boca Raton, FL: CRC Press, 2009.
- [109] J. Otepka, S. Ghuffar, C. Waldhauser, R. Hochreiter, and N. Pfeifer, "Georeferenced point clouds: A survey of features and point cloud management," *ISPRS Int. J. Geo-Inf.*, vol. 2, no. 4, pp. 1038–1065, 2013.
- [110] J. U. Eitel et al., "Beyond 3-D: The new spectrum of lidar applications for earth and ecological sciences," *Remote Sens. Environ.*, vol. 186, pp. 372–392, Dec. 2016.
- [111] K. Koenig and B. Höfle, "Full-waveform airborne laser scanning in vegetation studies: A review of point cloud and waveform features for tree species classification," *Forests*, vol. 7, no. 9, p. 198, 2016.
- [112] S. Filin and N. Pfeifer, "Segmentation of airborne laser scanning data using a slope adaptive neighborhood," *ISPRS J. Photogrammetry Remote Sens.*, vol. 60, no. 2, pp. 71–80, 2006.
- [113] S. J. Buckley, T. H. Kurz, J. A. Howell, and D. Schneider, "Terrestrial lidar and hyperspectral data fusion products for geological outcrop analysis," *Comput. Geosci.*, vol. 54, pp. 249–258, Apr. 2013.
- [114] A. V. Vo, L. Truong-Hong, and D. F. Laefer, "Aerial laser scanning and imagery data fusion for road detection in city scale," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.* 2015, pp. 4177–4180.
- [115] A. Boulch, B. L. Saux, and N. Audebert, "Unstructured point cloud semantic labeling using deep segmentation networks," in *Eurographics Workshop on 3D Object Retrieval*, I. Pratikakis, F. Dupont, and M. Ovsjanikov, Eds. Eurographics Association, 2017.
- [116] R. Hänsch, T. Weber, and O. Hellwich, "Comparison of 3D interest point detectors and descriptors for point cloud fusion,"

- ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. II-3, pp. 57–64, 2014. doi: 10.5194/isprsaannals-II-3-57-2014.
- [117] C. Paris, D. Kelbe, J. van Aardt, and L. Bruzzone, “A novel automatic method for the fusion of ALS and TLS lidar data for robust assessment of tree crown structure,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3679–3693, 2017.
- [118] T. L. Swetnam et al., “Considerations for achieving cross-platform point cloud data fusion across different dryland ecosystem structural states,” *Front. Plant Sci.*, vol. 8, p. 2144, Jan. 2018.
- [119] P. Ghamisi and B. Höfle, “Lidar data classification using extinction profiles and a composite kernel support vector machine,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 659–663, 2017.
- [120] T. Sankey, J. Donager, J. McVay, and J. B. Sankey, “UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA,” *Remote Sens. Environ.*, vol. 195, pp. 30–43, June 2017.
- [121] M. Gerke and J. Xiao, “Fusion of airborne laserscanning point clouds and images for supervised and unsupervised scene classification,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 87, pp. 78–92, Jan. 2014.
- [122] H. Wang and C. Glennie, “Fusion of waveform lidar data and hyperspectral imagery for land cover classification,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 108, pp. 1–11, Oct. 2015.
- [123] Q. Man, P. Dong, and H. Guo, “Pixel- and feature-level fusion of hyperspectral and lidar data for urban land-use classification,” *Int. J. Remote Sens.*, vol. 36, no. 6, pp. 1618–1644, 2015.
- [124] B. Höfle, M. Hollaus, and J. Hagenauer, “Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 67, pp. 134–147, Jan. 2012.
- [125] M. Alonso, B. Bookhagen, and D. A. Roberts, “Urban tree species mapping using hyperspectral and lidar data fusion,” *Remote Sens. Environ.*, vol. 148, pp. 70–83, May 2014.
- [126] N. Saarinen et al., “UAV-based photogrammetric point clouds and hyperspectral imaging for mapping biodiversity indicators in boreal forests,” *ISPRS Int. Archives Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-3/W3, pp. 171–175, Oct. 2017.
- [127] L. Liu, N. C. Coops, N. W. Aven, and Y. Pang, “Mapping urban tree species using integrated airborne hyperspectral and lidar remote sensing data,” *Remote Sens. Environ.*, vol. 200, pp. 170–182, Oct. 2017.
- [128] R. S. Rand, T. Khuon, and E. Truslow, “Distributed adaptive framework for multispectral/hyperspectral imagery and three-dimensional point cloud fusion,” *Opt. Eng.*, vol. 55, 2016. doi: 10.1117/1.OE.55.7.073101.
- [129] K. Karila, L. Matikainen, E. Puttonen, and J. Hyypä, “Feasibility of multispectral airborne laser scanning data for road mapping,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 294–298, 2017.
- [130] J. Vauhkonen et al., “Classification of spruce and pine trees using active hyperspectral lidar,” *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1138–1141, 2013.
- [131] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Norwood, MA: Artech House, 2015, ch. 5 and 6.
- [132] M. Brell, K. Segl, L. Guanter, and B. Bookhagen, “Hyperspectral and LiDAR intensity data fusion: A framework for the rigorous correction of illumination, anisotropic effects, and cross calibration,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2799–2810, 2017.
- [133] R. K. Hall et al., “Quantifying structural physical habitat attributes using LIDAR and hyperspectral imagery,” *Environ. Monitoring Assessment*, vol. 159, pp. 63–83, Dec. 2009.
- [134] T. Matsuki, N. Yokoya, and A. Iwasaki, “Hyperspectral tree species classification of Japanese complex mixed forest with the aid of lidar data,” *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 5, pp. 2177–2187, 2015.
- [135] M. Dalponte, L. Bruzzone, and D. Gianelle, “Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, 2008.
- [136] C. Debes et al., “Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest,” *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, 2014.
- [137] E. J. Lentilucci, “New SHARE 2010 HSI-LiDAR dataset: Recalibration, detection assessment and delivery,” in *Proc. SPIE Imaging Spectrometry XXI*, 2016. doi: 10.1117/12.2239044.
- [138] B. Rasti, P. Ghamisi, and R. Gloaguen, “Hyperspectral and lidar fusion using extinction profiles and total variation component analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3997–4007, 2017.
- [139] M. Zhang, P. Ghamisi, and W. Li, “Classification of hyperspectral and lidar data using extinction profiles with feature fusion,” *Remote Sens. Lett.*, vol. 8, no. 10, pp. 957–966, 2017.
- [140] M. Khodadadzadeh, J. Li, S. Prasad, and A. Plaza, “Fusion of hyperspectral and LiDAR remote sensing data using multiple feature learning,” *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 6, pp. 2971–2983, 2015.
- [141] Y. Zhang and S. Prasad, “Multisource geospatial data fusion via local joint sparse representation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3265–3276, 2016.
- [142] P. Ghamisi, B. Höfle, and X. X. Zhu, “Hyperspectral and lidar data fusion using extinction profiles and deep convolutional neural network,” *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, 2017.
- [143] B. Rasti, P. Ghamisi, J. Plaza, and A. Plaza, “Fusion of hyperspectral and LiDAR data using sparse and low-rank component analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6354–6365, 2017.
- [144] J. A. Benediktsson, J. A. Palmason, and J. R. Steinsson, “Classification of hyperspectral data from urban areas based on extended morphological profiles,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, 2005.
- [145] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, “Morphological attribute profiles for the analysis of very high resolution images,” *IEEE Trans. Geos. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, 2010.
- [146] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. Lotufo, “Extinction profiles for the classification of remote sensing data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, 2016.

- [147] M. Pedergnana, P. R. Marpu, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "A novel technique for optimal feature selection in attribute profiles based on genetic algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3514–3528, 2013.
- [148] P. Ghamisi, J. A. Benediktsson, and S. Phinn, "Land-cover classification using both hyperspectral and lidar data," *Int. J. Image Data Fusion*, vol. 6, no. 3, pp. 189–215, 2015.
- [149] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6976–6985, 2016.
- [150] B. Rasti, J. R. Sveinsson, and M. O. Ulfarsson, "Wavelet-based sparse reduced-rank regression for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6688–6698, 2014.
- [151] B. Rasti and K. S. Gudmundsson, "Sparse and low-rank feature extraction for the classification of target's tracking capability," *Proc. SPIE*, vol. 9970, 2016. doi: 10.1117/12.2240282. [Online]. Available: <http://dx.doi.org/10.1117/12.2240282>
- [152] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, 2006.
- [153] Y. Chen, C. Li, P. Ghamisi, X. Jia, and Y. Gu, "Deep fusion of remote sensing data for accurate classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1253–1257, 2017.
- [154] H. Li, P. Ghamisi, U. Soergel, and X. X. Zhu, "Hyperspectral and lidar fusion using deep three-stream convolutional neural networks," *Remote Sens.*, vol. 10, no. 10, p. 1649, 2018.
- [155] P. Ghamisi and N. Yokoya, "Img2dsm: Height simulation from single imagery using conditional generative adversarial net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 794–798, May 2018.
- [156] C. Kuenzer, S. Dech, and W. Wagner, Eds., *Remote Sensing Time Series: Revealing Land Surface Dynamics*. Cham, Switzerland: Springer-Verlag, 2015, pp. 1–24. [Online]. Available: <https://doi.org/10.1007/978-3-319-15967-6>
- [157] C. Gomez, J. C. White, and M. A. Wulder, "Optical remotely sensed time series data for land cover classification: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 116, pp. 55–72, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271616000769>
- [158] F. Bovolo and L. Bruzzone, "The time variable in data fusion: A change detection perspective," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 8–26, 2015.
- [159] A. Singh, "Digital change detection techniques using remotely sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1988.
- [160] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 609–630, 2013.
- [161] ESA Sentinel Online. [Online]. Available: <https://sentinel.esa.int/web/sentinel/home>
- [162] F. Bovolo, L. Bruzzone, and Y. T. S. Correa, *Multitemporal Analysis of Remotely Sensed Image Data*, vol. 2, Amsterdam, The Netherlands: Elsevier, 2017, pp. 156–185.
- [163] Y. T. Solano-Correa, F. Bovolo, and L. Bruzzone, "An approach for unsupervised change detection in multitemporal VHR images acquired by different multispectral sensors," *Remote Sens.*, vol. 10, no. 4, p. 533, 2018.
- [164] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, 2008.
- [165] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, 2010.
- [166] Y. T. Solano-Correa, F. Bovolo, L. Bruzzone, and D. Fernandez-Prieto, "Spatio-temporal evolution of crop fields in Sentinel-2 satellite image time series," in *Proc. 9th Int. Workshop Analysis Multitemporal Remote Sensing Images (MultiTemp)*, 2017, pp. 1–4.
- [167] J. Verbesselt, R. Hyndman, G. Newham, and D. Culvenor, "Detecting trend and seasonal changes in satellite image time series," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 106–115, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S003442570900265X>
- [168] J. Pekel, A. Cottam, N. Gorelick, and A. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, no. 7633, pp. 418–422, 2016.
- [169] A. Zanchetta, G. Bitelli, and A. Karnieli, "Monitoring desertification by remote sensing using the tasseled cap transform for long-term change detection," *Natural Hazards*, vol. 83, no. 1, pp. 223–237, 2016. [Online]. Available: <https://doi.org/10.1007/s11069-016-2342-9>
- [170] D. Roy, Y. Jin, P. Lewis, and C. Justice, "Prototyping a global algorithm for systematic fire-affected area mapping using modis time series data," *Remote Sens. Environ.*, vol. 97, no. 2, pp. 137–162, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0034425705001136>
- [171] L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1350–1359, 1999.
- [172] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, 1997.
- [173] Y. Guo, X. Jia, and D. Paull, "Effective sequential classifier training for SVM-based multitemporal remote sensing image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3036–3048, 2018.
- [174] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1321–1334, 2004.
- [175] L. Bruzzone, R. Cossu, and G. Vernazza, "Detection of land-cover transitions by combining multidate classifiers," *Pattern Recogn. Lett.*, vol. 25, no. 13, pp. 1491–1500, 2004. [Online]. Available: <http://dx.doi.org/10.1016/j.patrec.2004.06.002>

- [176] P. H. Swain, "Bayesian classification in a time-varying environment," *IEEE Trans. Syst., Man, Cybern.*, vol. 8, no. 12, pp. 879–883, 1978.
- [177] L. Bruzzone and M. Marconcini, "Toward the automatic updating of land-cover maps by a domain-adaptation SVM classifier and a circular validation strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1108–1122, 2009.
- [178] F. Bovolo, G. Camps-Valls, and L. Bruzzone, "A support vector domain method for change detection in multitemporal images," *Pattern Recog. Lett.*, vol. 31, no. 10, pp. 1148–1154, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865509001676>
- [179] L. Bruzzone and R. Cossu, "A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 1984–1996, 2002.
- [180] L. Bruzzone, R. Cossu, and G. Vernazza, "Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images," *Inf. Fusion*, vol. 3, no. 4, pp. 289–297, 2002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S156625350200091X>
- [181] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, 2017.
- [182] D. Ienco, R. Gaetano, C. Dupacquier, and P. Maurel, "Land cover classification via multitemporal spatial data by deep recurrent neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1685–1689, 2017.
- [183] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 456–460, 2001.
- [184] L. Bruzzone and D. F. Prieto, "A partially unsupervised cascade classifier for the analysis of multitemporal remote-sensing images," *Pattern Recog. Lett.*, vol. 23, no. 9, pp. 1063–1071, 2002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865502000533>
- [185] B. Demir, F. Bovolo, and L. Bruzzone, "Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1930–1941, 2012.
- [186] K. Bahirat, F. Bovolo, L. Bruzzone, and S. Chaudhuri, "A novel domain adaptation bayesian classifier for updating land-cover maps with class differences in source and target domains," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2810–2826, 2012.
- [187] B. Demir, F. Bovolo, and L. Bruzzone, "Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 300–312, 2013.
- [188] B. Demir, F. Bovolo, and L. Bruzzone, "Classification of time series of multispectral images with limited training data," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3219–3233, 2013.
- [189] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A DASVM classification technique and a circular validation strategy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 770–787, 2010.
- [190] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proc. IEEE*, vol. 104, no. 11, pp. 2207–2219, 2016.
- [191] Xinhua, "China to launch 60 Jilin-1 video satellites by 2020," China Daily, Jan. 27, 2018. [Online]. Available: <http://www.chinadaily.com.cn/a/201801/27/WS5a6c360ea3106e7dcc137247.html>
- [192] M. Chi, Z. Sun, Y. Qin, J. Shen, and J. A. Benediktsson, "A novel methodology to label urban remote sensing images based on location-based social media photos," *Proc. IEEE*, vol. 105, no. 10, pp. 1926–1936, 2017.
- [193] J. M. Shapiro, "Smart cities: Quality of life, productivity, and the growth effects of human capital," *Rev. Econ. Statist.*, vol. 88, no. 2, pp. 324–335, 2006.
- [194] P. Gamba, "Human settlements: A global challenge for EO data processing and interpretation," *Proc. IEEE*, vol. 101, no. 3, pp. 570–581, 2013.
- [195] 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, 2009.
- [196] S. B. Serpico, S. Dellepiane, G. Boni, G. Moser, E. Angiati, and R. Rudari, "Information extraction from remote sensing images for flood monitoring and damage evaluation," *Proc. IEEE*, vol. 100, no. 10, pp. 2946–2970, 2012.
- [197] C. Aggarwal and T. Abdelzaher, *Managing and Mining Sensor Data*. New York: Springer-Verlag, 2013, ch. 8, "Social Sensing," pp. 237–275.
- [198] R. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, 2011.
- [199] L. Barrington, S. Ghosh, M. Greene, and S. Harnoy, "Crowdsourcing earthquake damage assessment using remote sensing imagery," *Ann. Geophys.*, vol. 54, no. 6, pp. 680–687, 2011.
- [200] E. Schnebele, C. Oxendine, G. Cervone, C. M. Ferreira, and N. Waters, "Using non-authoritative sources during emergencies in urban areas," in *Geotechnologies and the Environment: Computational Approaches for Urban Environments*, vol. 13, M. Helbich, J. J. Arsanjani, and M. Leitner, Eds. Switzerland: Springer-Verlag, 2015, ch. 14, pp. 337–361.
- [201] G. Cervone, E. Sava, Q. Huang, E. Schnebele, J. Harrison, and N. Waters, "Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study," *Int. J. Remote Sens.*, vol. 37, no. 1, pp. 100–124, 2015.
- [202] T. White, *Hadoop: The Definitive Guide*, 1st ed. Sebastopol, CA: O'Reilly Media, Inc., 2009.
- [203] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets," in *Proc. 2nd USENIX Conf. Hot Topics Cloud Computing*. Berkeley, CA, 2010, pp. 1–7.
- [204] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation. 2014. [Online]. Available: [arxiv:abs/1411.4038](http://arxiv.org/abs/1411.4038)
- [205] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition. 2014. [Online]. Available: [arxiv:abs/1409.1556](http://arxiv.org/abs/1409.1556)