Multisource and Multitemporal Data Fusion in Remote Sensing

Pedram Ghamisi, *Senior Member, IEEE,* Behnood Rasti, *Member, IEEE,* Naoto Yokoya, *Member, IEEE,*
Qunming Wang, Bernhard Höfle, Lorenzo Bruzzone, *Fellow, IEEE,* Francesca Bovolo, *Senior Member, IEEE,*
Mingmin Chi, *Senior Member, IEEE,* Katharina Anders, Richard Gloaguen,
Peter M. Atkinson, and Jón Atli Benediktsson, *Fellow, IEEE*

Abstract—The final version of the paper can be found in IEEE Geoscience and Remote Sensing Magazine.

The sharp and recent increase in the availability of data captured by different sensors combined with their considerably heterogeneous natures poses a serious challenge for the effective and efficient processing of remotely sensed data. Such an increase in remote sensing and ancillary datasets, however, opens up the possibility of utilizing multimodal datasets in a joint manner to further improve the performance of the processing approaches with respect to the application at hand. Multisource 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/backscattering information of the remotely sensed data is possible and helps to move from a representation of 2D/3D data to 4D data structures, where the time variable adds new information as well as challenges for the 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 in different paths according to each research community. This paper brings together the advances of multisource and multitemporal data fusion approaches with respect to different 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 paper 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, as well as big data and social media. In addition, the main challenges and possible future research for each section are outlined and discussed.

Index Terms—Fusion; Multisensor Fusion; Multitemporal Fusion; Downscaling; Pansharpening; Resolution Enhancement; Spatio-Temporal Fusion; Spatio-Spectral Fusion; Component Substitution; Multiresolution Analysis; Subspace Representation; Geostatistical Analysis; Low-Rank Models; Filtering; Composite Kernels; Deep Learning.

P. Ghamisi and R. Gloaguen are with the Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology (HIF), Exploration, D-09599 Freiberg, Germany (emails: p.ghamisi@gmail.com,r.gloaguen@hzdr.de).

B. Rasti is with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (email: behnood@hi.is).

N. Yokoya is with the RIKEN Center for Advanced Intelligence Project, RIKEN, 103-0027 Tokyo, Japan (e-mail: naoto.yokoya@riken.jp).

Q. Wang is with the College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China (email: wqm11111@126.com).

B. Höfle and K. Anders are with GIScience at the Institute of Geography, Heidelberg University, Germany (emails: hoefle@uni-heidelberg.de, katharina.anders@uni-heidelberg.de).

L. Bruzzone is with the department of Information Engineering and Computer Science, University of Trento, Trento, Italy (email: lorenzo.bruzzone@unitn.it).

F. Bovolo is with the Center for Information and Communication Technology, Fondazione Bruno Kessler, Trento, Italy (email: bovolo@fbk.eu).

M. Chi is with the school of Computer Science, Fudan University, China (email: mmchi@fudan.edu.cn).

P. M. Atkinson is with Lancaster Environment Centre, Lancaster University, Lancaster, U.K (email: pma@lancaster.ac.uk).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: benedikt@hi.is).

Manuscript received 2018.

I. INTRODUCTION

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 and planned for launch, and new platforms for proximate sensing such as unmanned aerial vehicles (UAVs) producing very fine spatial resolution data. While optical sensing capabilities have increased in quality and volume, the number of alternative modes of measurement has also grown including, most notably, airborne light detection and ranging (LiDAR) and terrestrial laser scanning (TLS), which produce point clouds representing elevation, as opposed to 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 the satellite mission Gravity Recovery And Climate Experiment (GRACE) or airborne electro-magnetic surveys are currently been also considered. In addition, there has been great interest in new sources of ancillary data, for example, from social media, crowd sourcing, scraping the internet and so on ([7]–[9]). These data have a very different modality 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 above platforms may vary greatly in multiple dimensions; for example, the types of properties sensed and the spatial and spectral resolutions of the data. This is true, even for sensors that are housed on the same platform (e.g., the many examples of multispectral and panchromatic sensors) or that are part of the same satellite configuration (e.g., the European Space Agency's (ESA's) series of Medium Resolution Imaging Spectrometer (MERIS) sensors). The rapid increase in the number and availability of data combined with their deeply heterogeneous natures 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 datasets pertaining to the problem and this creates a dilemma; how best to combine the datasets 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 above increase in the number and heterogeneity of data sources (presenting both challenge and opportunity) 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 exist a very wide range of approaches to data fusion (e.g., [11]–[13]). This paper seeks to review them by class of data modality (e.g., optical, SAR, 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 now 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 reviewed in the following sections. Due to the fact that the topic of multisensor data fusion is extremely broad and that specific aspects have been reviewed already 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.

We start by defining the space and properties of interest. In remote sensing, there have historically been considered to be four dimensions in which information is provided. These are: spatial, temporal, spectral, and radiometric; that is, 2D spatially, 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 definition of multiple wavebands (spectral dimension). LiDAR and TLS, in contrast to most optical and SAR sensors, measure a surface in 3D spatially. Recent developments in photo- and radargrammetry such as Structure from Motion (SfM) and InSAR, have increased the availability of 3D data. This expansion of the dimensionality of interest to 3D in space and 1D 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) measuring the visible and infrared parts of the EMS, and laser scanners measuring surface elevation in 3D. Only surface elevation is likely to be a primary interest, whereas reflectance and backscatter are likely to be only indirectly related to the property of interest.

Secondly, we 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, 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, 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, the sampling process implied by the image model above can be expanded and generalized to three key parameters (the sampling extent, the sampling scheme, and the sampling support), each of which has four further parameters (size, geometry, orientation, and position). The support is a key sampling parameter which 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 datasets with different spatial resolutions such as to create a unified dataset at the finest resolution [23]–[25]. Fig. 1(a) demonstrates schematically the multiscale nature (different spatial resolutions) of diverse datasets 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 the above measurement and sampling processes, which fundamentally limit the amount of information transferring from the scene to any one particular dataset. Indeed, in most cases of data fusion in remote sensing the different datasets 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, topographic surface). Understanding these measurement and sampling processes is, therefore, key to characterizing methods of data fusion since 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: (i) measurement to provide data, as described above, (ii) characterization of the data through model fitting, (iii) prediction of unobserved data given (ii), and (iv) forecasting [26]. (i), (ii), and (iii) are defined in space or space-time, while (iv) extends through time beyond the range of the current data.
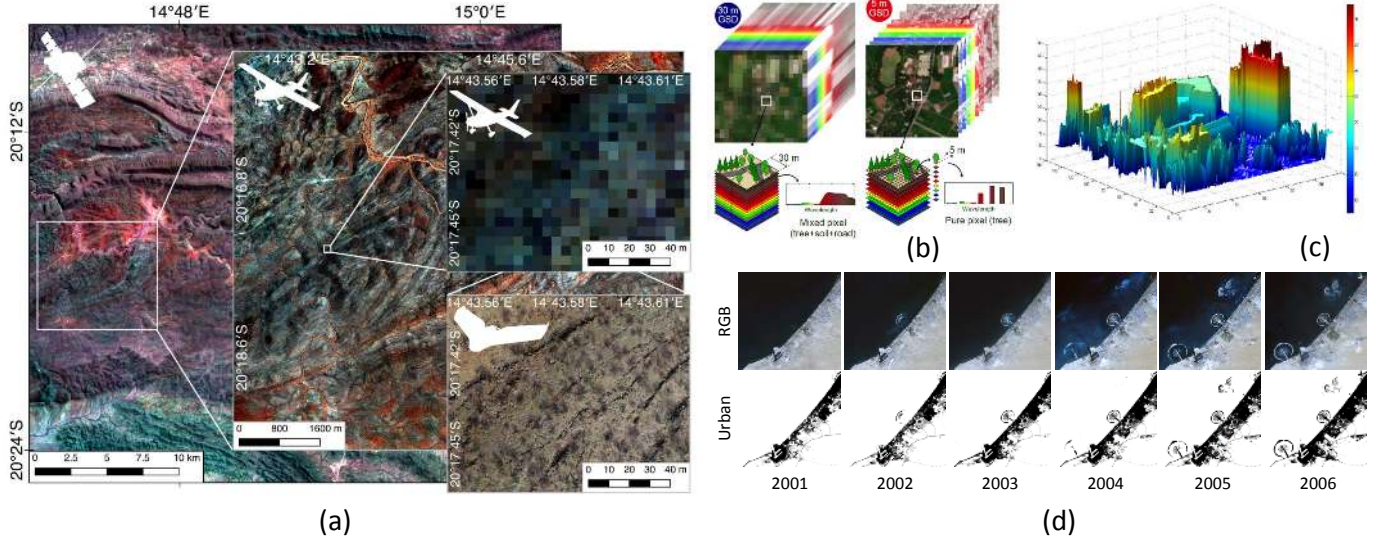
Fig. 1: (a) The multiscale nature of diverse datasets captured by multisensor data (spaceborne, airborne, and UAV sensors) in Nambia [14]; (b) The trade-off between spectral and spatial resolutions; (c) Elevation information obtained by LiDAR sensors from the University of Houston; (d) Time-series data analysis for assessing the dynamic of changes using RGB and urban images captured from 2001 to 2006 in Dubai.

Prediction (iii) can be of the measured property $\mathbf{x}$ (e.g., reflectance or topographic elevation, through interpolation) or it can be of a property of interest $\mathbf{y}$ to which the measured $\mathbf{x}$ data are related (e.g., land cover or vegetation biomass, through classification or regression-type approaches). Similarly, data fusion can be undertaken on $\mathbf{x}$ or it can be applied to predict $\mathbf{y}$ from $\mathbf{x}$. Generally, therefore, data fusion is applied either between (ii) and (iii) (e.g., fusion of $\mathbf{x}$ based on the model in (ii)), as part of prediction (e.g., fusion to predict $\mathbf{y}$) or after prediction of certain variables (e.g., ensemble unification). In this paper, the focus is on data fusion to predict $\mathbf{x}$.

Data fusion is made possible because each dataset 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 processes. Therefore, crucially, one should expect some level of coherence between the real world (the source) and the multiple datasets (the observations), as well as between the datasets themselves, and this is the basis of most data fusion methods. This concept of coherence is central to data fusion [27].

Attempts to fuse datasets 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 comprised of functional objects (e.g., residential houses, roads) that have expectations around their sizes and shapes, and such expectations can aid in defining objective functions (i.e., in optimization solutions) [29]. These sources of prior information (on real world structure) constrain the space of possible fusion solutions beyond the data themselves.

Many key application domains stand to benefit from data fusion processing. For example, there exists a very large number

of applications where an increase in spatial resolution would add utility, which is the center of focus in Section II of this paper. These include land cover classification, urban-rural definition, target identification, geological mapping, and so on (e.g., [30]). A large focus of attention currently is on the specific problem that arises from the trade-off in remote sensing between spatial resolution and temporal frequency; in particular the fusion of coarse-spatial-fine-temporal-resolution with fine-spatial-coarse-temporal-resolution space-time datasets such as to provide frequent data with fine spatial resolution [31]–[34], which will be detailed in Section II and V of this paper. Land cover classification is one of the most vibrant fields of research in the remote sensing community [35], [36], which attempts to differentiate between several land cover classes available in the scene, can substantially benefit from data fusion. Another example is the trade-off between spatial resolution and spectral resolution (Fig. 1(b)) to produce fine-spectral-spatial resolution images, which plays an important role for land cover classification and geological mapping. As can be seen in Fig. 1(b), both fine spectral and spatial resolutions are required to provide detailed spectral information and avoid the "mixed-pixel" phenomenon at the same time. Further information about this topic can be found in Section II. Elevation information provided by LiDAR and TLS (see Fig. 1(c)) can be used in addition to optical data to further increase classification and mapping accuracy, in particular for classes of objects, which are made up of the same materials (e.g., grassland, shrubs, and trees). Therefore, Sections III and IV of this paper are dedicated to the topic of elevation data fusion and their integration with passive data. Furthermore, new sources of ancillary data obtained from social media, crowd sourcing, 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 as well as

hazard monitoring and identification. This young, yet active, field of research is the focus of Section VI.

Many applications can benefit from fused fine-resolution, time-series datasets, particularly those that involve seasonal or rapid changes, which will be elaborated in Section V. Fig. 1(d) shows the dynamic of changes for an area in Dubai from 2001 to 2006 using time-series of RGB and urban images. For example, monitoring of vegetation phenology (the seasonal growing pattern of plants) is crucial to monitoring deforestation [37] and crop yield forecasting, which mitigates against food insecurity globally, natural hazards (e.g. earthquakes, landslides) or illegal activities such as pollutions (e.g. oil spills, chemical leakages). However, such information is provided globally only at very coarse resolution, meaning that local smallholder farmers cannot benefit from such knowledge. Data fusion can be used to provide frequent data needed for phenology monitoring, but 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 argument for frequent updates at fine resolution is obvious. While these application domains provide compelling arguments for data fusion, there exist many challenges including: (i) the data volumes produced at coarse resolution via sensors such as MODIS and MERIS are already vast, meaning that fusion of datasets most likely needs to be undertaken on a case-by-case basis as an on-demand service and (ii) rapid change events require ultra-fast 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 above, including the sources of the datasets to be fused. In the following sections, we review data fusion approaches in remote sensing according to the data sources to be fused only, but the further considerations introduced above are relevant in each section.

The remainder of this review is divided into the following sections. First, we review pansharpening and resolution enhancement approaches in Section II. Then, we will move on by discussing point cloud data fusion in Section III. Section IV is devoted to hyperspectral and LiDAR data fusion. Section V presents an overview of multitemporal data fusion. Major recent advances in big data and social media fusion are presented in Section IV. Finally, Section VII draws conclusions.

II. PANSHARPENING AND RESOLUTION ENHANCEMENT

Optical Earth observation satellites have trade-offs 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) spatio-spectral fusion to merge fine-spatial and fine-spectral resolutions [see Fig. 2(a)]; (2) spatio-temporal fusion to blend fine-spatial and fine-temporal resolutions [see Fig. 2(b)]. This
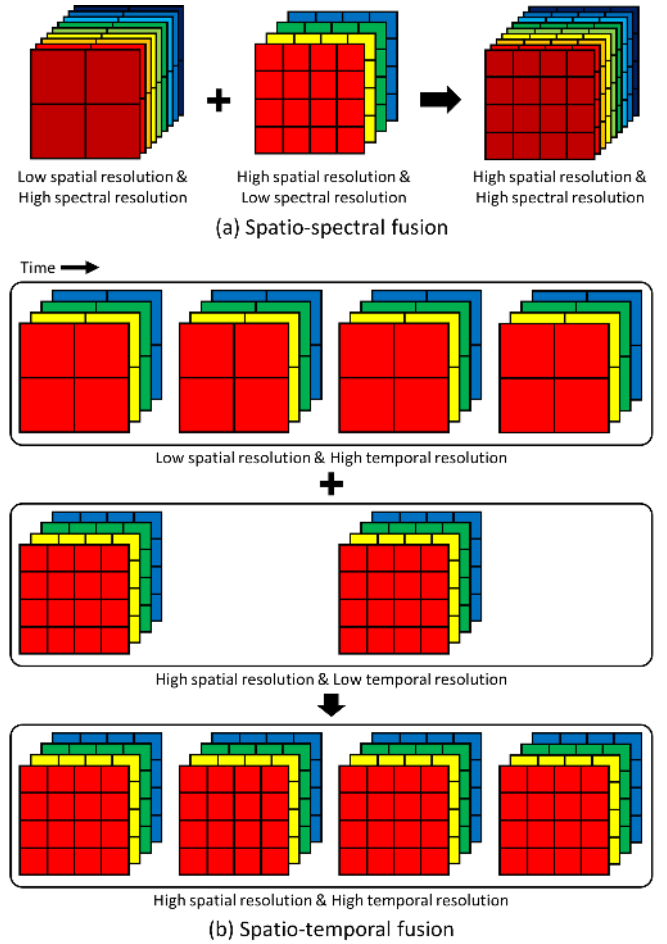

Fig. 2: Schematic illustrations of (a) spatio-spectral fusion and (b) spatio-temporal fusion.

section provides overviews of these technologies with recent advances.

A. Spatio-spectral fusion

Satellite sensors such as 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. Spatio-spectral fusion is a technique to fuse the fine spatial resolution images (e.g., 0.5 m WorldView PAN image) with coarse spatial resolution images (e.g., 2 m WorldView multispectral image) to create fine spatial resolution images for all bands. Spatio-spectral fusion is also termed pan-sharpening when the available fine spatial resolution image is a single PAN image. When multiple fine spatial resolution bands are available, spatio-spectral fusion is referred to as multiband image fusion, where two optical images with a trade-off 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].
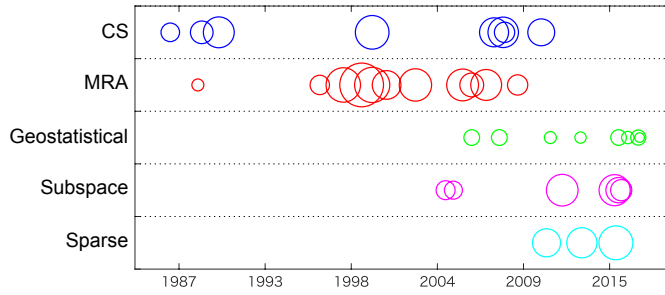
Fig. 3: The history of the representative literature of five approaches in spatio-spectral fusion. The size of each circle is proportional to the annual average number of citations. For each category, from left to right, circles correspond to [42]–[50] for *CS*, [51]–[57] for *MRA*, [58]–[61], [27], [62], [31], [63] for *Geostatistical*, [64]–[69] for *Subspace*, and [70]–[72] for *Sparse*.

Over the past decades, spatio-spectral fusion has motivated considerable research in the remote sensing community. Most spatio-spectral fusion techniques can be categorized into at least one of five approaches: 1) component substitution (CS), 2) multiresolution analysis (MRA), 3) geostatistical analysis, 4) subspace representation, and 5) sparse representation. Fig. 3 shows the history of representative 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 January 20, 2018), which indicates the impact of each approach in the field. The main concept and characteristics of each category are described below.

1) Component Substitution: CS-based pan-sharpening 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 (IHS) [44], principal component analysis (PCA) [43], and Gram-Schmidt [46] transformations. Next, the component that is supposed to contain the spatial information of the multispectral image is substituted by the PAN image after adjusting the intensity range of the PAN image to that of the component using histogram matching. Finally, the inverse transformation is performed on the modified data to obtain the sharpened image.

Aiazzi et al. (2007) proposed the general CS-based pan-sharpening 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-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 performances of traditional CS-based techniques were greatly increased, mitigating spectral distortion.

CS-based fusion techniques have been used widely owing to the following advantages: i) high fidelity of spatial details in the output, ii) low computational complexity, and iii)

robustness against misregistration. On the other hand, the CS methods suffer from global spectral distortions when the overlap of spectral response functions (SRFs) between the two sensors is limited.

2) Multiresolution Analysis: As shown in Fig. 3, great effort has been devoted to the study of MRA-based pan-sharpening algorithms particularly between 2000 and 2010 and they have been used widely as benchmark methods for more than ten years. The main concept of MRA-based pan-sharpening methods 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 pan-sharpening 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 techniques are based 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. (2015) proposed a general framework called hypersharpening that extends MRA-based pan-sharpening 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 its 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 techniques.

3) Geostatistical Analysis: Geostatistical solutions provide another family of approaches for spatio-spectral fusion. This type of approach can preserve the spectral properties of the original coarse images. That is, when the downscaled prediction is upscaled 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 downscaled the multispectral bands to a spatial resolution finer than any input images, including the PAN image. DSCK is a one-step method, and it 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 250 m Moderate Resolution Imaging Spectroradiometer (MODIS) bands 1-2 with 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 in Sales et al. [61], however, KED suffers from expensive computational cost, as it computes

TABLE I: Quantitative assessment of five representative pan-sharpening methods for the Hong Kong WorldView-2 dataset

Category	Method	PSNR	SAM	ERGAS	$Q2^n$
—	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

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 modelling and then the second step downscales the coarse residuals from the regression process with ATPK. The downscaled 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), which 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. ATPRK was also employed for fusion of the Sentinel-2 Multispectral Imager (MSI) images acquired from the recently launched Sentinel-2A satellite. Specifically, the six 20 m bands were downscaled to 10 m spatial resolution by fusing them with the four observed 10 m bands [31].

4) Subspace Representation: As indicated in Fig. 3, research on subspace-based fusion techniques has become very 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 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 due to iterative optimization. Recent efforts for speeding up the fusion procedure [69] are key to the applicability of this family of techniques for large-sized images obtained by operational satellites (e.g., Sentinel-2). Another drawback of the subspace-based methods is that they can introduce unnatural artifacts in the spectral domain due to imperfect subspace representations.

5) Sparse Representation: In recent years, spatio-spectral fusion approaches based on patch-wise sparse representation have been developed along with the theoretical development of compressed sensing and sparse signal recovery. Pan-sharpening based on sparse representation can be regarded as a special case of learning-based super-resolution, where correspondence between coarse- and fine-spatial-resolution patches are learned from a database (or a dictionary). Li et al. (2011) proposed the first sparse-representation-based pan-sharpening 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 setting of pan-sharpening, where there is only one pair of PAN and multispectral images is 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 that are not visible in the fine spatial resolution image at exactly the same location by reconstructing each patch of the output as a linear combination of non-local 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 pan-sharpening algorithms, namely, GSA [48], SFIM [54], MTF-GLP-HPM [76], ATPRK [27], and J-SparseFI-HM [77] using WorldView-2 data taken over Hong Kong. The original dataset consists of 0.5 m GSD PAN and 2 m GSD 8 multispectral bands. To assess the quality of pan-sharpened 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 signal-to-noise ratio (PSNR), spectral angle mapper (SAM), *erreur relative globale adimensionnelle de synthèse* (ERGAS) [78], and $Q2^n$ [79], which are all well-established quality measures in pan-sharpening. 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 both visually and quantitatively in Fig. 4 and Table I, respectively. The quality measures in Table I 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 Fig. 4, we can
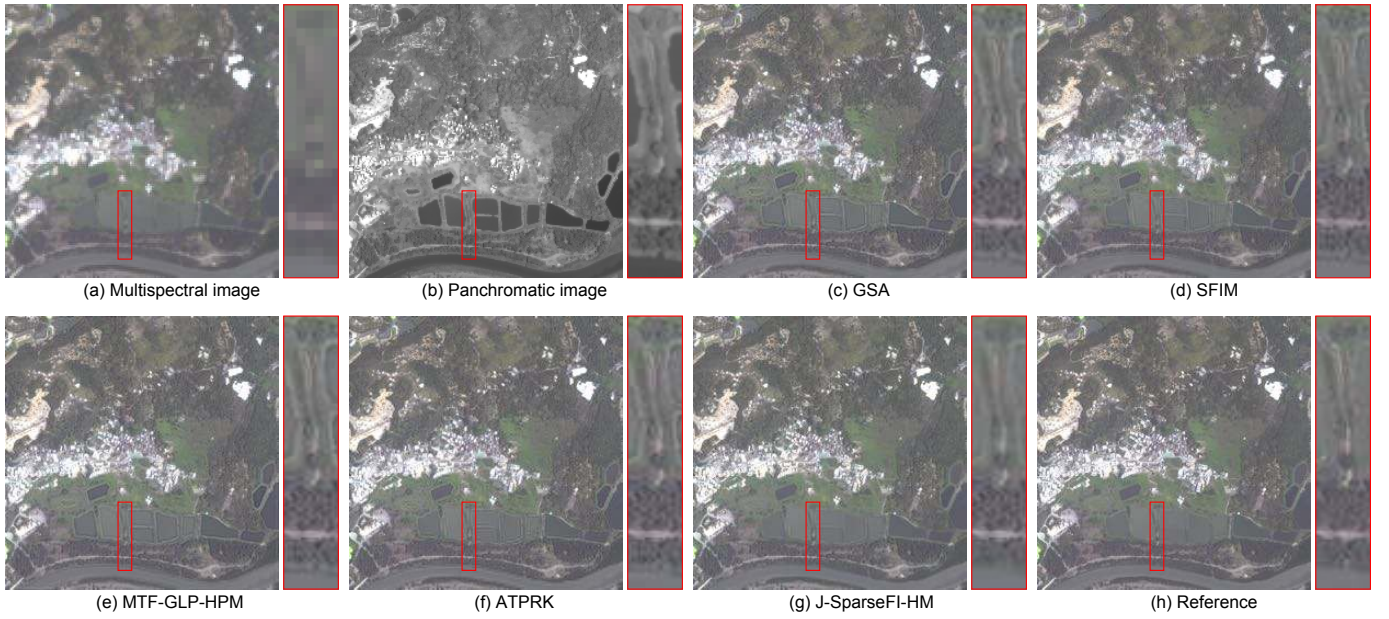
Fig. 4: The Hong Kong WorldView-2 dataset (bands 4, 3, and 2 as RGB). (a) 8 m coarse multispectral image, (b) 2 m PAN image (c) GSA, (d) SFIM, (e) MTF-GLP-HPM, (f) ATPRK, (g) J-SparseFI-HM, and (h) 2 m reference image.

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 non-local sparse representation capability.

B. Spatio-temporal fusion

For remote sensing-based global monitoring, there always exists a trade-off 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 m to 1000 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 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 or ecological phenomena. The recent development of remotely piloted aircraft systems (RPAS) or drones will provide a huge amount of multisource data with very high spatial and temporal resolutions.

Spatio-temporal fusion is a technique to blend fine spatial resolution, but coarse temporal resolution (e.g., Landsat) data and fine temporal resolution, but coarse spatial resolution data to create fine spatio-temporal resolution (e.g., MODIS) data [80]–[82]. Its implementation is performed based on the availability of at least one coarse-fine spatial resolution image pair (e.g., 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 spatio-temporal 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 spatio-temporal fusion methods. It is 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 were proposed, including sparse representation [86], [87], extreme learning machine [88], artical neural network [89], and deep learning [90]. These methods learn the relationship between the available coarse-fine image pairs, which 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 land-use database [95]. This type of methods is performed based on the strong assumption that there is no land-cover/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 classification of an available Landsat image [93], 30 m Landsat-like time-series were produced from 300 m Medium Resolution Imaging Spectrometer (MERIS) time-series to monitor vegetation seasonal dynamics. To maintain the similarity between the predicted endmembers and the pre-defined 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 spatio-temporal reflectance fusion model (U-STFM) using two coarse-fine image pairs. In addition, the image-pair-based and spatial unmixing-based methods can also be combined [32], [99], [100].

Spatio-temporal 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 spatio-temporal fusion. Compared to the original 500 m MODIS data, the 250 m data can provide more information for the abrupt changes and heterogeneous landscapes than, and thus, can increase the accuracy of spatio-temporal fusion predictions.

Blending MODIS and Landsat has been the most common spatio-temporal fusion problem over the past decade. Recently, Sentinel-2 and Sentinel-3 are two newly launched satellites for global monitoring. The Sentinel-2 MSI and Sentinel-3 Ocean and Land Colour Instrument (OLCI) sensors have very different spatial and temporal resolutions (Sentinel-2 MSI sensor 10 m, 20 m and 60 m, 10 days, albeit 5 days with 2 sensors, conditional upon clear skies; Sentinel-3 OLCI sensor 300 m, <1.4 days with 2 sensors). Wang et al. [34] proposed a new method, called Fit-FC, for spatio-temporal 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 (RM fitting), spatial filtering (SF) and residual compensation (RC). The Fit-FC method can be implemented using only one image pair and is particularly relevant for cases involving strong temporal changes.

C. Challenges and trends of downscaling

The major remaining issue in the field of spatio-spectral fusion is how to conduct fair comparisons. Many researchers use their own simulated datasets, and the source code is rarely released. To fairly evaluate the performance of each algorithm, it is necessary to develop benchmark datasets that can be accessible for everyone and include various scenes. Also, it is always desirable to release the source code of each method for enabling reproducible research. In several review papers, researchers have attempted to evaluate many methods with common datasets and to 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 generalization ability, and also those datasets are not freely available due to a restricted data policy of the original sources. Regarding the source code, there are still many research groups who never release their source code, while always outperforming state-of-the-art algorithms in their papers. It is an urgent issue of the community to arrange benchmark

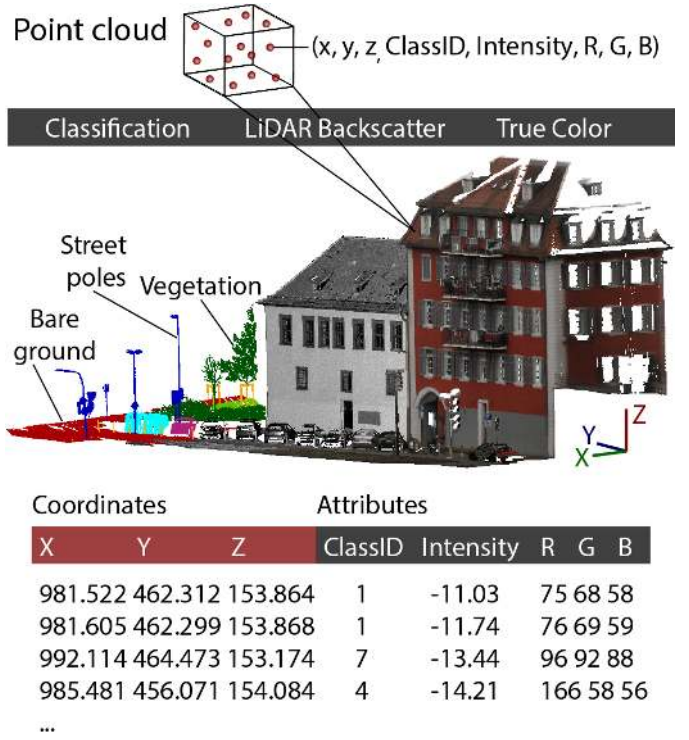| Coordinates			Attributes				
X	Y	Z	ClassID	Intensity	R	G	B
981.522	462.312	153.864	1	-11.03	75	68	58
981.605	462.299	153.868	1	-11.74	76	69	59
992.114	464.473	153.174	7	-13.44	96	92	88
985.481	456.071	154.084	4	-14.21	166	58	56
...							

Fig. 5: 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.

datasets on a platform like the GRSS Data and Algorithm Standard Evaluation (DASE) website [102] so that everyone can fairly compete for the performance of the algorithm.

With respect to spatio-temporal fusion, the main challenges lie in the reconstruction of land cover changes and eliminating the differences between coarse and fine spatial resolution time-series. Due to the large difference in the 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 for abrupt changes) from coarse images always involves great uncertainty. Most of the existing methods are performed based on the strong assumption of no land cover change, such as the classical STARFM, ESTARFM, and the spatial unmixing-based method. Furthermore, due to the differences in characteristics of sensors, atmospheric condition, and acquisition geometry, the available coarse and fine spatial resolution data (e.g., MODIS and Landsat data) are always not perfectly consistent. The uncertainty is directly propagated to the spatio-temporal fusion process. In future research, it will be of great interest to develop more accurate methods to account for the land cover changes and inconsistency between coarse and fine spatial resolution time-series.

III. 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 is driven from 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 country-wide, large-scale topographic mapping to near real-time usage in autonomous driving applications. Sensors and methods, respectively, to derive point clouds include predominantly LiDAR and photogrammetry [105]. A further very recent data source of point clouds in research is tomographic SAR [106]. Also low-cost depth cameras are used increasingly [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 close-range, sensing technique [108].

A. Point cloud data model

Although the above-mentioned aspects draw a very broad picture, the common denominator is the point cloud data model, which is the initial data model shared by all multi-source fusion methods that include point clouds. Otepka et al. [109] defined the georeferenced point cloud data model as a set of points, $\mathbf{P}_i$, $i = 1, ..., n$, in three-dimensional Cartesian space that is related to a geospatial reference system (e.g., UTM). $\mathbf{P}_i$ 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 $j = 1, ..., m_i$ 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, ID of classification or segmentation, local surface normal vector component (e.g., n_x, n_y, n_z), and so forth. Fig. 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 post-processing (e.g., segmentation) and fusion with other data sources. Please refer to [109] and [111] for a more detailed description of LiDAR point cloud features. A point in a point cloud, $\mathbf{P}_i$, is a vector, $(x_i, y_i, z_i, a_{1,i}, ..., a_{m_i,i})^T$, of dimension $3 + m_i$ with the 3D coordinates as the first three dimensions (see Fig. 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 three-dimensional spatial nature of $\mathbf{P}$ and that often no fixed spatial scale and accuracy exist across the dataset. 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].

B. Concepts of point cloud fusion

The main objectives of point cloud data fusion are to make use of the three-dimensional geometric, spatial-structural and LiDAR backscatter information inherent in point clouds and combine it with spectral data sources or other geoinformation layers, such as 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 datasets. Based on the target data model ("product") we separate the following strategies (see Fig. 6):

1) Point cloud level: Enrich the initial point cloud $\mathbf{P}$ with new point features.
2) Image/Voxel level: Derive new image layers representing 3D point cloud information.
3) Feature level: Fusion of point cloud information on the segment/object level.

1) Point cloud level - Pixel to point and point to point: Texturing point clouds with image data is a standard procedure for calibrated multi-sensor LiDAR systems for which the transformation from image to point cloud is well-known from lab 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 $\mathbf{P}_i$ contains the respective pixel values from the images (e.g., R, G, B) as point features and can be used for classification and object detection.

The labels of classified hyperspectral data can be transfered to the corresponding 3D points from LiDAR using precise co-registration. 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 of 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 ground, building, and unassigned. This step was followed by the region growing segmentation of the classified ground points to delineate roads. The point features of $\mathbf{P}$ were height, image intensity (RGB and HSV), 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 of multi-source point cloud classification, which resulted in a classified point cloud, worked in the image domain and then transfered back the classification results 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
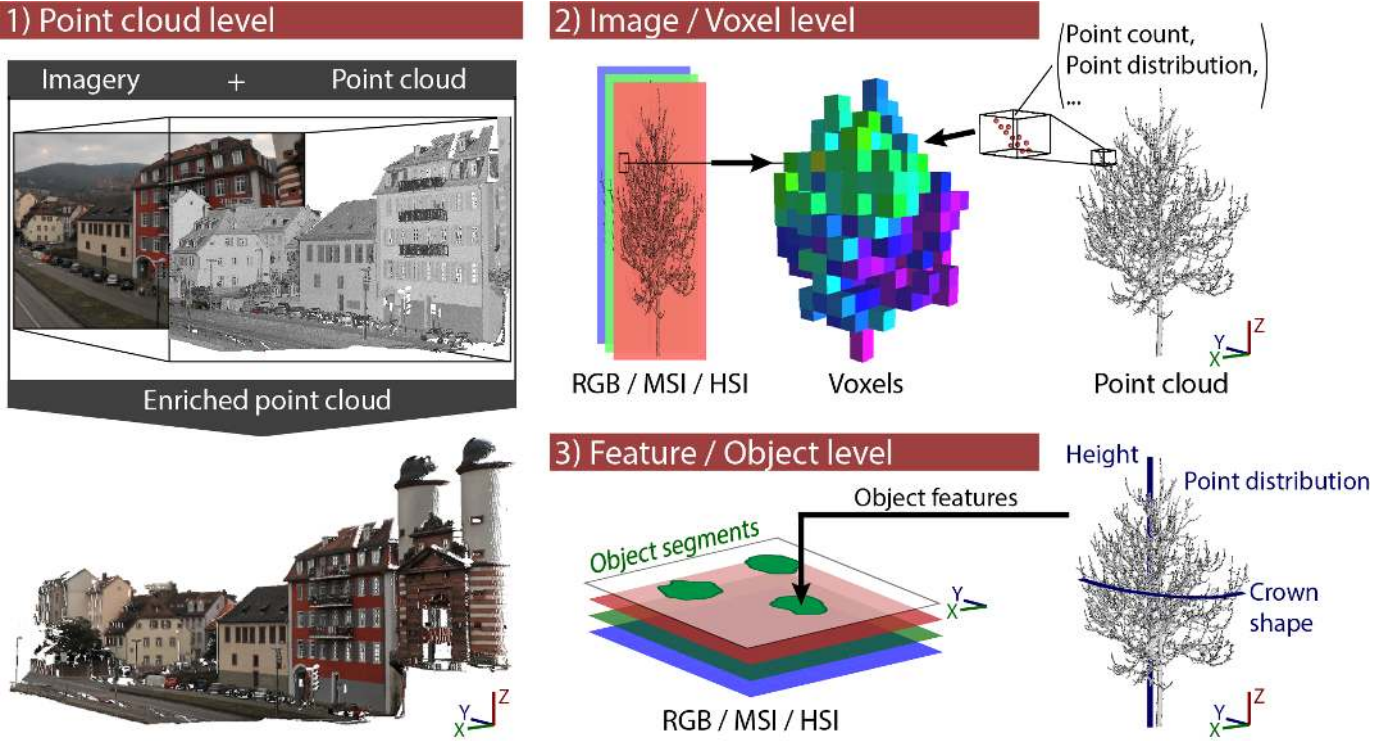
Fig. 6: Strategies of point cloud data fusion on (1) point cloud level, (2) image/voxel level, and (3) feature/object level. 1) Visualizes the enrichment of the initial point cloud colored by LiDAR intensity with RGB information from imagery with the RGB-colored point cloud as product. 2) Depicts a voxel 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. 3) Shows the assignment of features derived from the 3D point cloud to object segments created from raster image data.

3D point neighborhood information is not available in the classification procedure in the image domain, such as it is, for example, in point cloud segmentation.

Point cloud-to-point cloud data fusion is known as point cloud (co-)registration or alignment. Co-registration of point clouds from the same sensor (e.g., within one LiDAR scanning campaign) is a standard pre-processing 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 (ICP)], keypoint-based [e.g., Scale Invariant Feature Transform (SIFT)] or surface-based (e.g., local planes) or any combination [116]. This fusion principle is generally valuable if point clouds from different sensor types are merged, which have different accuracies, spatial coverages, and spatial scales as well as being captured at different timestamps.

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 datasets to derive a fused point cloud.

A combination of datasets from different methods (e.g., LiDAR and photogrammetry) and platforms can lead to more accurate results compared to the individual use of a source. This was concluded in [118], where datasets from two different methods (LiDAR and photogrammetry) and three different

platforms [a ground-based platform, a small unmanned aerial systems (UAS)-based platform, and a manned aircraft-based platform] were explored. They merged point clouds from UAS LiDAR, airborne manned LiDAR, and UAS photogrammetry spatially to a single point cloud to estimate the accuracy of bare earth elevation, heights of grasses, and shrubs.

2) 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 rich point clouds that derive from point cloud geometry, (LiDAR) backscatter, and also 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 laser echo ratio, variance of point elevation, plane fitting residuals, and echo intensity. The fusion approach of LiDAR and HSI and classification of an urban scene is presented in Section IV. The experiment compares classification results to accuracies of the individual use of HSI.

A pixel-based convolutional neural network (CNN) was used to perform semantic labeling of point clouds by Boulch et al. [115] 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 it 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. Furthermore, they used terrestrial LiDAR as reference dataset. The HSI was pre-classified with the mixture-tuned matched filtering subpixel classification technique. The multi-source fusion of UAV LiDAR and hyperspective data (12 cm GSD) was performed via a decision tree classification approach. The fusion-based result achieved higher user's accuracy for most target classes and also overall accuracy with an increase from 76% with 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 the geometric quality of LiDAR and 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 with two main consecutive steps: 1) Point cloud segmentation (region growing) and classification (mean shift) using 3D LiDAR and spectral information (NDVI/Saturation), 2) supervised (Random Trees) or unsupervised classification - by a Markov random field framework using graph-cuts for energy optimization - of voxels. 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 non-vegetation, but shadow areas still caused problems. Point cloud segmentation is sensitive to the color information that was 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, asphalt road, etc.). The main goal was to generate synthetic vertical LiDAR waveforms by converting the raw LiDAR waveforms into a voxel model (size of $1.2 \, \text{m} \times 1.2 \, \text{m} \times 0.15 \, \text{m}$). 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, and also metrics such as the height of the last return, penetration depth, and maximum LiDAR amplitude. In addition to these raster features, they 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 to 71.30% using only LiDAR and 85.82% using only HSI data.

3) Feature/Object level: This concept is based on the previous 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]. They introduced a method to produce segments based on LiDAR point cloud-derived images [e.g., normalized DSM (nDSM) and echo ratio images]. The output segments were enriched by a multitude of geometric and full-waveform features that were computed directly in the 3D point clouds of each segment (e.g., mean echo width). In particular, the geometric 3D point cloud features (e.g., echo ratio) played an important role for vegetation classification because they encapsulated the 3D structure of vegetation. Alonzo et al. [125] also worked at the single tree/crown object level and added HSI to the airborne LiDAR dataset to map urban tree species. They applied canonical variates in a linear discriminant analysis classifier to assign tree species labels to the segments, which were derived from the LiDAR canopy height model. Their LiDAR point cloud-derived structural variables included, for example, median height of returns in crown, average intensity below median height, and so forth. 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 point cloud-based segment features (e.g., height percentiles and average height) and also spectral segment features (e.g., mean and median spectra). By using nearest-neighbor estimation, the variables of diameter at breast height, tree height, health status, and tree species were determined for each crown segment. In the second step, the biodiversity indicators - structural complexity, amount of deciduous, and 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, they 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, and the inverse minimum noise fraction rotation (MNF) 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 based on LiDAR nDSM, intensity and NDVI images, which 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 to HSI classification alone. The joint pixel and object-based method increased the overall accuracy by 7.1% to 94.7%.

HSI and airborne LiDAR data were used as complementary data sources for crown structure and physiological tree information by Liu et al. [127] 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 for the subsequent segment-based random forest classification.

The 22 LiDAR-derived crown structural features per segment included, for example, crown shape, laser return intensity, laser point distribution, etc.. They concluded that the combination of LiDAR and HSI increased the 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 by using the elevation values from the LiDAR point cloud. Their approach resulted in a large increase in overall classification accuracy by multi-source fusion (HSI and LiDAR) to 98.5%, compared to 74.5% overall accuracy with HSI input only.

C. Challenges and trends of point cloud fusion

Generally, we can see a large gain in the importance of point clouds. Multi-source fusion including point clouds is already used in a huge variety of fields of applications (see [104]) and reveals several trends:

- The 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. Derived classification labels are transfered back to points afterwards.
- The fusion (or joint use) of spectral and 3D point cloud information from single-source photogrammetry (structure-from-motion and dense image matching). 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 multi-source 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 MSI/HSI 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 fast-growing due to the complementary nature of such multi-sensor data. Therefore, Section IV is specifically dedicated to the fusion of HSI and LiDAR-derived features to provide readers with an effective review of such fusion schemes.

IV. HYPERSPECTRAL AND LiDAR

The efficacy of LiDAR, which is 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 hand, hyperspectral sensors, which are characterized as passive remote sensing techniques, provide rich and continuous spectral information by sampling the reflective portion of the electromagnetic spectrum, ranging from the visible region (0.4-0.7 μm) to the short-wave infrared region (almost 2.4 μm) in hundreds of narrow contiguous spectral channels (often 10 nm wide). Such detailed spectral information has made HSIs 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], [131]. However, HSIs do not contain any information about the elevation and size of different materials, which imposes an inevitable constraint to classify objects that are made up of similar materials (e.g., grassland, shrubs, and trees). The aforementioned limitations and capabilities of each sensor, as discussed earlier in the introduction part, 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 multi-sensor fusion technique is dedicated to scene classification, which is also the pre-eminent focus of this section.

Several studies such as [134], [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]). Co-registered LiDAR and HSI data were introduced in [137]. Fig. 7 demonstrates schematically that the fusion of HSI and LiDAR can increase the classification accuracy above that of each individual source considerably (i.e., this figure was generated based on some studies in [138]).

Below, we discuss briefly a few key approaches for the fusion of LiDAR and HSI, which are categorized in four subsections: Filtering approaches, low-rank models, composite kernels, and deep learning-based fusion approaches. Corresponding to each section, some numerical classification results obtained from the CASI Houston University data (details below) are reported in Table II. To obtain a better numerical evaluation, the classification accuracies of the individual use of HSI obtained by random forest (**RF**$_{\text{HSI}}$), support vector machine (**SVM**$_{\text{HSI}}$), and convolutional neural network (**CNN**$_{\text{HSI}}$) are also listed in Table II.

A. Houston University

The Houston University data for this section are composed of a LiDAR-derived digital surface model (DSM) and an HSI both captured over the University of Houston campus and the neighboring urban area. This dataset was initially made publicly available for the 2013 GRSS data fusion contest. The HSI and LiDAR data were captured on June 23, 2012 and June 22, 2012, respectively. The size of the dataset is 349 × 1905

TABLE II: Houston - The classification accuracy values achieved by different state-of-the-art approaches. The indexes, average accuracy (AA) and overall accuracy (OA), are reported in percentages while the kappa coefficient (K) is of no unit.

| Class name | Train./Test | Spectral | | | Multisensor fusion | | | | | | | |
		RF_{HSI}	SVM_{HSI}	CNN_{HSI}	$EP_{HSI+LiDAR}$	GBFF[136]	FFCK[139]	MLR*sub*[140]	ALWMJ-KSRC[141]	CNNGBFF[142]	SLRCA[143]	OTVCA[138]
Grass Healthy	198/1053	83.38	83.48	82.24	78.06	82.53	81.39	82.91	98.36	78.73	81.58	80.63
Grass Stressed	190/1064	98.40	96.43	98.31	84.96	98.68	99.91	81.48	98.59	94.92	99.44	99.62
Grass Synthetis	192/505	98.02	99.80	70.69	100.00	100	100	100	100	100	98.61	100.00
Tree	188/1056	97.54	98.77	94.98	95.45	98.96	97.92	95.83	98.04	99.34	96.12	96.02
Soil	186/1056	96.40	98.11	97.25	98.76	100	100	99.05	93.15	99.62	99.72	99.43
Water	182/143	97.20	95.10	79.02	95.80	95.10	95.80	91.61	100	95.8	98.60	95.8
Residential	196/1072	82.09	89.09	86.19	73.41	90.95	78.54	87.59	91.11	87.87	90.39	86.01
Commercial	191/1053	40.65	45.87	65.81	85.28	90.98	86.61	84.14	92.51	95.25	95.73	93.54
Road	193/1059	69.78	82.53	72.11	93.95	90.46	87.72	91.78	86.87	89.71	98.21	97.07
Highway	191/1036	57.63	83.20	55.21	67.08	60.91	68.82	86.20	94.66	81.18	63.42	68.53
Railway	181/1054	76.09	83.87	85.01	90.89	94.46	90.23	98.58	90.56	86.34	90.70	98.86
Parking Lot 1	192/1041	49.38	70.99	60.23	88.56	99.14	98.08	92.32	90.74	92.7	91.07	100.00
Parking Lot 2	184/285	61.40	70.53	75.09	76.14	65.26	80.35	76.84	89.92	87.02	76.49	74.74
Tennis Court	181/247	99.60	100.00	83.00	100.00	100	100	99.60	98.58	99.19	100.00	100.00
Running Track	187/473	97.67	97.46	52.64	99.78	99.15	100	98.73	98.14	89.64	99.15	100.00
AA	–	80.34	86.34	77.19	88.54	91.24	91.02	90.65	NA	91.82	91.95	92.45
OA	–	77.47	84.69	78.35	86.98	91.28	89.93	91.11	92.45	91.75	91.3	92.68
K	–	0.7563	0.8340	76.46	0.8592	0.903	0.8910	0.8985	NA	0.9033	0.9056	0.9181
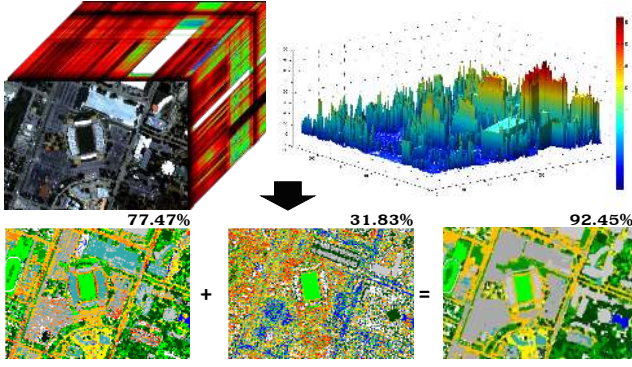
Fig. 7: HSI and LiDAR fusion. This figure was generated based on some studies in [138] where the overall classification accuracy of HSI (77.47%) and LiDAR (31.83%) is significantly increased to 92.45% using a feature fusion approach.
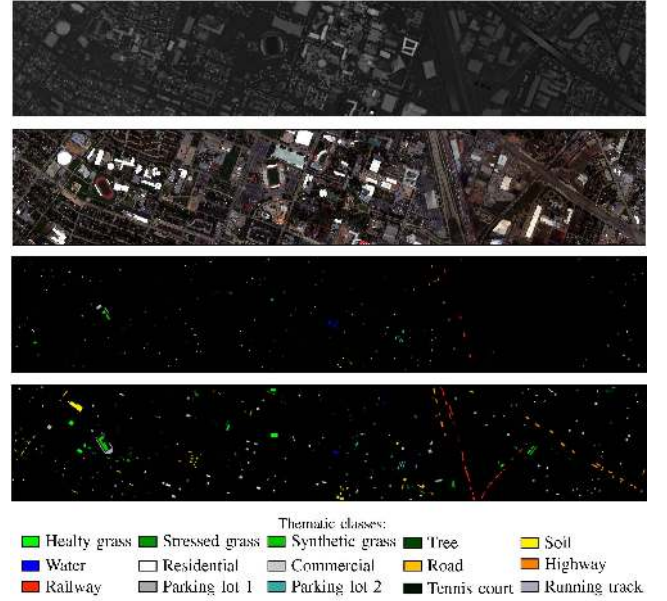

Fig. 8: Houston - From top to bottom: LiDAR-derived rasterized DSM, a color composite illustration of the CSI Houston HSI using bands 64, 43, and 22 as R, G, and B, respectively; Training samples; Test samples; and legend of different classes.

pixels with a ground sampling distance of 2.5 m. The HSI consists of 144 spectral bands ranging 0.38-1.05μm. Fig. 8 illustrates the investigated data and the corresponding training and test samples. The number of training and test samples for different classes are detailed in Table II.

B. 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 approaches, one can refer to morphological profiles (MPs [144], i.e., which can be produced by the sequential implementation of opening and closing operators by reconstruction by considering a structuring element of increasing size), attribute profiles (APs [145], i.e., which can obtain a multilevel characterization of the input image by considering the repeated implementation of morphological attribute filters), and extinction profiles (EPs [146], i.e., 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 since they are fast and conceptually simple and able to provide accurate classification results. For instance, in [147], [148], the spatial features of HSI and LiDAR were extracted using APs. Then, they were concatenated and fed to a classifier leading to precise results in terms of classification accuracy in a fast manner. 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 a random forest (RF) classifier (i.e., the results obtained by that approach can be found in Table II as **EP**$_{HSI+LiDAR}$).

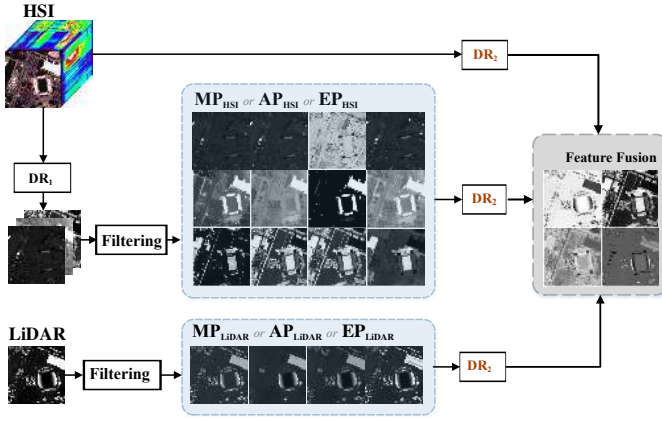Filtering approaches such as MPs, APs, and EPs suffer from

Fig. 9: Low-rank models. The use of DR$_2$ is optional. However, the studies investigated in [136], [142], [138] and [143] recommend the consideration of this extra extra step in order to provide more accurate classification maps.

two shortcomings: The curse of dimensionality and intensive processing time for the subsequent classification steps since 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 this shortcoming, composite kernel- and low rank-based approaches, which will be discussed in the following subsections, have been suggested in the literature to effectively fuse HSI and LiDAR.

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

1) DR$_1$ generates base images to build up MP, AP, or EP.
2) Filtering investigates MP, AP, or EP to extract spatial features (e.g., **EP/AP/MP**$_\text{HSI}$) and elevation features (e.g., **EP/AP/MP**$_\text{LiDAR}$) from HSI and LiDAR, respectively.
3) DR$_2$ 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$_2$ are that it can reduce the executable computational cost as well as noise throughout the feature space. In [136], [138], [142], [143], kernel PCA (KPCA) has been use for DR$_2$.
4) Finally, the outputs of (3) are fused and fed to a classification method. Below, we discuss [136], [142], [138] and [143] in more detail:

In [136], the spectral (**HSI**), spatial (**MP**$_\text{HSI}$), and elevation features (**MP**$_\text{LiDAR}$) were used (as the filtering step). A graph-based feature fusion (GBFF) technique was utilized (as the feature fusion step). Finally, an SVM classifier was used to classify the fused features (results can be found in Table II as **GBFF**).

In [142], the spectral (**HSI**), spatial (**EP**$_\text{HSI}$), and elevation features (**EP**$_\text{LiDAR}$) were concatenated and fed to the GBFF and classified by a 2D CNN. These results can be found in Table II as **CNNGBFF**.

In [138], the following low-rank model was suggested to fuse **HSI**, **EP**$_\text{HSI}$, and **EP**$_\text{LiDAR}$:

$$\mathbf{F} = \mathbf{A}\mathbf{V}^T + \mathbf{N}, \tag{1}$$

where $\mathbf{F} = \big[\mathbf{f}_{(i)}\big]$ is an $n \times p$ matrix which contains the vectorized features in its columns, $\mathbf{V}$ is a $p \times r$ unknown matrix containing the subspace basis, $\mathbf{A} = \big[\mathbf{a}_{(i)}\big]$ is a $n \times r$ matrix which contains the r unknown fused features in its columns, and $\mathbf{N} = \big[\mathbf{n}_{(i)}\big]$ 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 Fig. 9). OTVCA is given by

$$\arg\min_{\mathbf{A},\mathbf{V}} \ \frac{1}{2} \left\| \mathbf{F} - \mathbf{A}\mathbf{V}^T \right\|_F^2 + \lambda \sum_{i=1}^{r} \text{TV}(\mathbf{a}_{(i)}) \text{ s.t. } \mathbf{V}^T\mathbf{V} = \mathbf{I}_r, \tag{2}$$

where the total variation penalty (TV) is applied spatially on the fused features. TV preserves the spatial structure of the features while promotes 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 II as **OTVCA**.

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

$$\mathbf{F} = \mathbf{D}\mathbf{W}\mathbf{V}^T + \mathbf{N}, \tag{3}$$

where $\mathbf{D}$ is an $n \times n$ matrix which contains two-dimensional wavelet basis, and $\mathbf{W} = \big[\mathbf{w}_{(i)}\big]$ is an $n \times r$ matrix containing the unknown 2D wavelet coefficients for the ith fused component. In [143], the sparse and low-rank component analysis [150], [151] was used to estimate $\mathbf{W}$ and $\mathbf{V}$ given by

$$\arg\min_{\mathbf{W},\mathbf{V}} \ \frac{1}{2} \left\| \mathbf{F} - \mathbf{D}\mathbf{W}\mathbf{V}^T \right\|_F^2 + \lambda \sum_{i=1}^{r} \left\| \mathbf{w}_{(i)} \right\|_1 \text{ s.t. } \mathbf{V}^T\mathbf{V} = \mathbf{I}_r, \tag{4}$$

Note that the estimated fused features are given by $\hat{\mathbf{F}}_{fused} = \mathbf{D}\hat{\mathbf{W}}$. The fused features are expected to be sparse in the 2D wavelet basis. Therefore, in [150], to enforce the sparsity an ℓ_1 penalty on the wavelet coefficients $\mathbf{W}_r$ was used. As a result, promoting sparsity on the fused feature improves the SNR and the final classification accuracies. Results for this approach can be found in Table II as **SLRCA**.

D. Composite kernel

Composite kernel-based fusion approaches partially overcome the shortcomings of the filtering approaches by designing several kernels to handle spectral, spatial, and elevation features in feature space [152].

In [139], spectral, spatial (e.g., $\mathbf{EP}_{HSI}$), and elevation (e.g., $\mathbf{EP}_{LiDAR}$) information were fused using a local-region filter (LRF) and composite kernels. Results for this approach can be found in Table II as **FFCK**. The main shortcoming of this approach was that its obtained classification accuracy was dramatically influenced by the μ parameter which represents the amount of trade-off 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 approach was based on APs and multiple feature learning using the subspace multinomial logistic regression (MLRsub) classifier. The result of this approach is shown as **MLRsub** in Table II.

A joint sparse representation classification approach was proposed in [141] for multisource data fusion where the multisource data were weighted to have better sparse representation. The core idea of this approach was based on sparse representation classification. Then, the minimum distance (in the sense of ℓ_2 norm) between each sample and its sparse representation using subdictionaries (containing only training samples for one class) were used to allocate the class labels. However, in [141] the regularization term was weighted according to the data sources. Moreover, the method was also translated into the kernel space using kernel tricks. Results for the composite kernel version can be found in Table II as **ALWMJ-KSRC**.

E. 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 further magnified when 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 to the nonlinearities of the input data.

Deep learning is a fast-growing topic in the remote sensing community whose trace can also be found in the research area of LiDAR and HSI data fusion. 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 II as **CNNGBFF**. To extract spatial and elevation features in a more effective manner than in [142], two distinct CNN streams (as shown in Fig. 10) were employed in [153]. The heterogeneous features obtained by the previous CNNs were then classified by a fully connected deep neural network. In [154], a three-stream CNN with multisensor composite kernel was utilized to fuse spectral, spatial, and elevation features.

F. Trends of hyperspectral and LiDAR fusion

The following trends for the advancements of hyperspectral and LiDAR fusion need to be further investigated in the future:

- Due to the increased availability of large-scale DSMs and hyperspectral data, the further development of fast, accurate, and automatic classification/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 of areas with 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 further increase the quality of classification outputs.
- [155] took the first step in the remote sensing community to simulate DSM from single optical imagery. This work opens a new path in front of researchers to further modify this approach and design more sophisticated network architectures to produce more accurate elevation information from single optical images.

As stated above, the classification/fusion of large-scale data (e.g., big data) is a vitally important research line which will be further discussed in the next section.

V. Multitemporal Data Fusion

The use of multitemporal information is crucial for many important applications (from the analysis of slow and smooth evolving phenomena [156], [157] to steep and abrupt changes [158]–[160]). Fig. 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 earlier in the introduction part, 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 (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 information of the images opens the possibility to change the perspective also 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 as well as challenges for the information extraction algorithms.

Analyzing the literature, the most widely addressed applications of multitemporal data are the analysis/classification of image time-series and change detection [158], [162]. Nonetheless, there are many emerging topics based on the joint
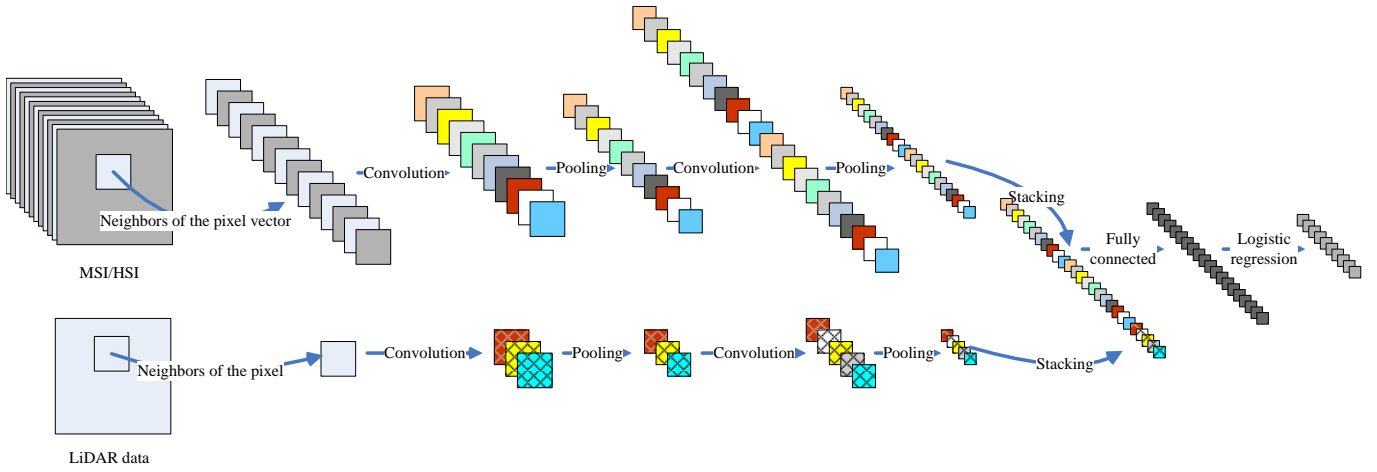
Fig. 10: Deep learning models [153].

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 with coarse/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, it is possible to study the phenology of the vegetation in specific local areas or to analyze the trends of vegetation in single fields of agricultural regions for precision farming applications [157], [166]. This can be done also by fusing the acquisitions of different satellites in a single multisensor time-series. Given the complexity and the 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 view point of the availability of labeled training data.

A. 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 data analysis. The goal of the classification can be to generate: i) a land-cover map associated with the most recent image (acquisition) of a time-series (or of a pair of acquisitions) (Fig. 12 (a)) [171]; ii) a land cover map for each item of the time-series, thus, producing a set of multitemporal land-cover maps (Fig. 12 (b)) that also implicitly models the land-cover transitions [172], [173]; iii) an annual/sesonal land-cover map with classes that represent the behavior of the temporal signature of each pixel/region in the images in a year/season

(Fig. 12 (c)) [174], [175]. 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 of obtaining sub-optimal results.

The use of the temporal information in classification dates to the early 1980s. First approaches used a stacked vector representation of the multitemporal data as input to the classifiers resulting in the so-called supervised direct multidate classification [159]. The main idea of such approaches 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 only the land covers of the most recent image. However, this is theoretically affordable under the assumption that both: i) there are no changes in the land covers between the considered image acquisition dates; and ii) 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 multidate direct classification approaches is to classify the land covers 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 to 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 combination of classes, including those associated to the changes. This is seldom possible in real application scenarios.

Many methodologies have been developed to address the above-mentioned issues of multidate direct classification. Swain [176] introduced in a pioneering paper a very interesting approach to the classification of multitemporal data based on the cascade classification of pairs of 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 to decouple
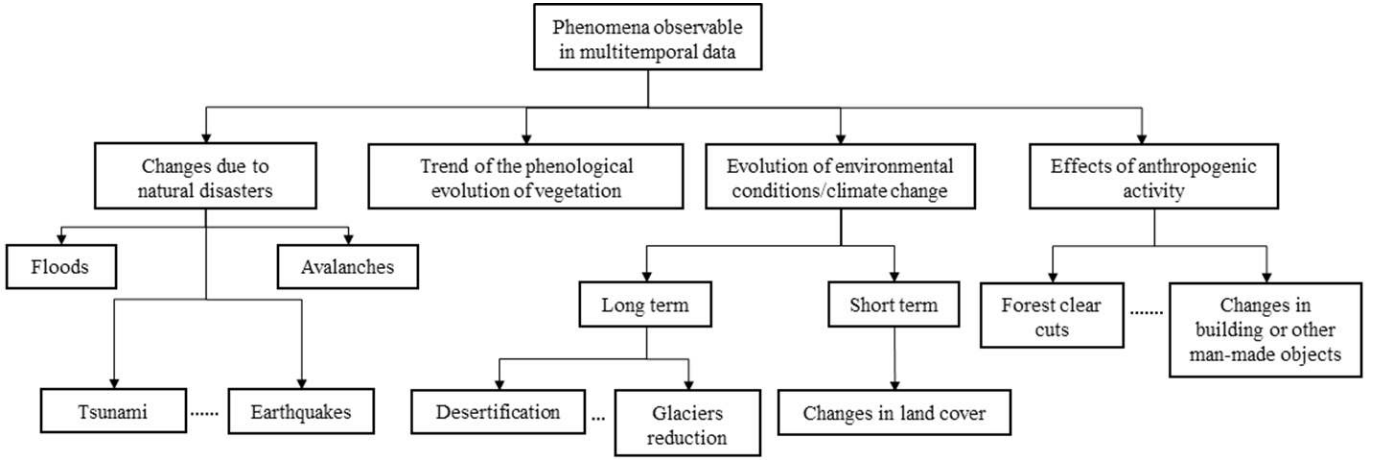
Fig. 11: Example of taxonomy of phenomena observable and detectable by using temporal information in remote sensing data analysis.
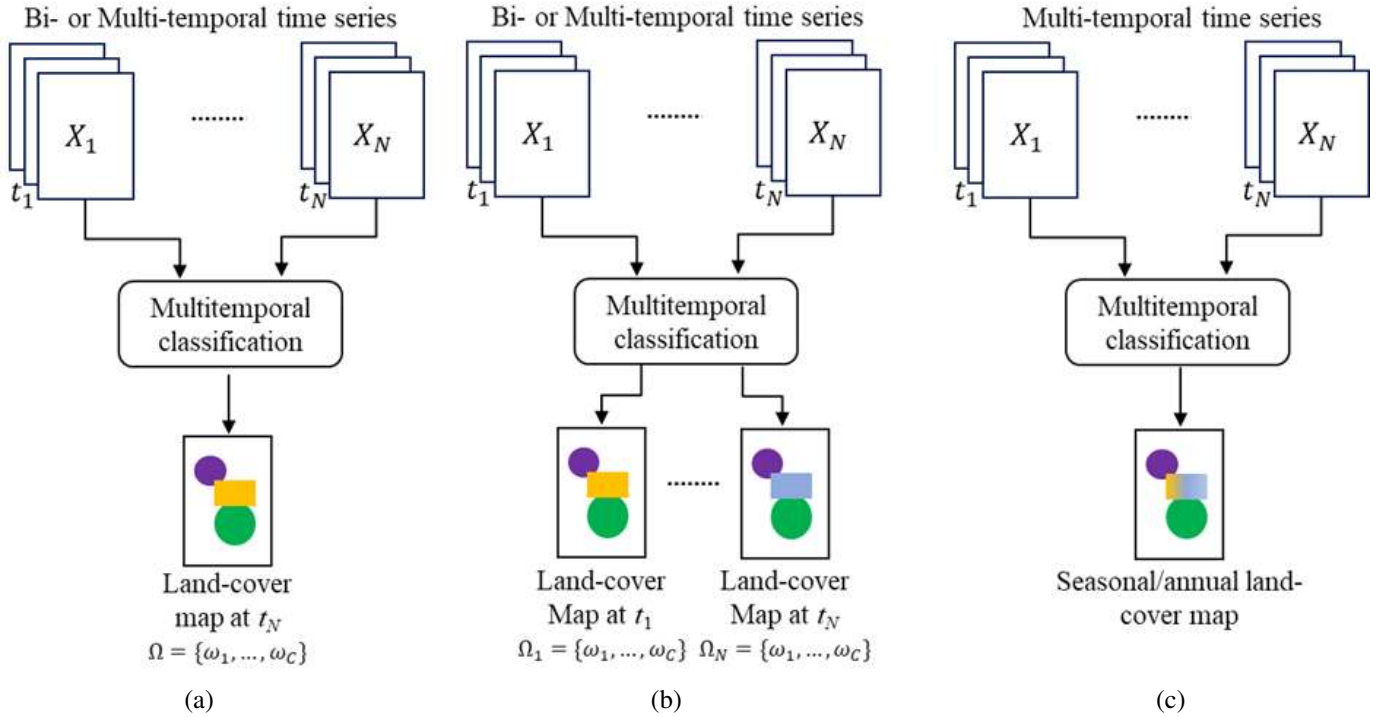

Fig. 12: Block scheme for achieving different goals in multitemporal classification: (a) a land-cover map associated with the most recent image of a time-series; (b) a land cover map for each item of the time-series; (c) an annual/seasonal land-cover map with classes that represent the behavior of the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

in the multitemporal classification problem the modeling of the class distributions at each single date with 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 an approach to compound classification based on neural networks classifiers being able to properly merge in a Bayesian decision framework the distribution free estimations of the class parameters derived from both multispectral and SAR multitemporal images. This kind of fusion has been studied widely in the past two decades, and developed in the context of different classification methodologies including several neural models (e.g., multilayer perceptron neural networks, radial basis function networks), kernel methods (e.g., support vector machines [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 spatio-temporal information has been investigated including Markov Random Fields in the modeling of the spatio-temporal context of multitemporal data.

Nowadays, the challenge, still poorly addressed, is to exploit deep learning architectures (e.g., convolutional neural networks) in the classification of multitemporal data [181], [182]. These architectures are intrinsically able to capture the spatio-temporal patterns in the definition of the classification model and, thus, to increase the accuracy of the land-cover/land-cover-transition maps. However, there is still a very significant challenge to define theoretically sound and computational 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 an amount of training data that currently is far from being available.

B. 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 spatio-temporal domain. A proper modeling of the temporal information would require multitemporal ground reference samples (or reliably annotated multitemporal images) with samples that represent: i) all the multitemporal classes; ii) the inter-relation between classes along the time-series (e.g., land-cover transitions or different kinds of changes) with a reliable statistic; and iii) the high temporal and spatial variabilities in large scenes. These constraints are very difficult to satisfy in real applications. For this reason, a large 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 activities have been focused on semi-supervised (also called partially unsupervised) classification methods [183]–[188]. These methods jointly exploit the available labeled training data and the distribution of the observed images for improving the modeling of the spatio-temporal 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 the use with the cascade and compound classification approaches to the classification of bi-temporal images [171], [184]. The approaches can integrate multispectral and SAR multitemporal data, as well as multisensor images. The problem of semi-supervised learning with multitemporal data has been then 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 semi-supervised 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 labels of the classes. Transfer learning approaches were proposed in [186]–[188], where change detection-based techniques were defined for propagating 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 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 amount of papers on methods capable to capture in a semi-supervised way 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 decouple in the architecture of the network the learning of the relation between the spatio-temporal patterns to achieve feasible requirements on the amount and the characteristics of training samples without degrading significantly the capability to extract the semantic of spatio-temporal pattern from the data. We expect that these crucial issues will be widely addressed in future years.

VI. BIG DATA AND SOCIAL MEDIA

In the recent decade, big data has become a very important topic in many research areas, e.g., remote sensing applications [190]. Every day a massive number of remote sensing data is provided by a large number of Earth observation (EO) space borne and airborne sensors from many different countries. In the near future, all-day, all-weather and full spectrum acquisition segment datasets will be provided by commercial satellites, such as the Jilin-1 constellation, which has launched 10 fine-spatial resolution satellites by February 2018 and will have 60 satellites in orbit by 2020 with a capability of observing any global arbitrary point with a 30 minute revisit frequency [191]. Those video satellites with a (very) fine temporal and spatial resolution can effectively be exploited to monitor our location-based living environments like CCD cameras but on a much larger scale [192]. From a broader spatial perspective, new opportunities for humankind can be provided by the big remote sensing data acquired by those satellites jointly 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], emergency and environmental hazards [195], [196], etc.. On the other hand, new challenges can appear from unprecedented access to a huge number of remote sensing data that are leading to a data-rich but knowledge-poor environment in a fast manner. Here, the semantic gap of remote sensing data is usually caused due to 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
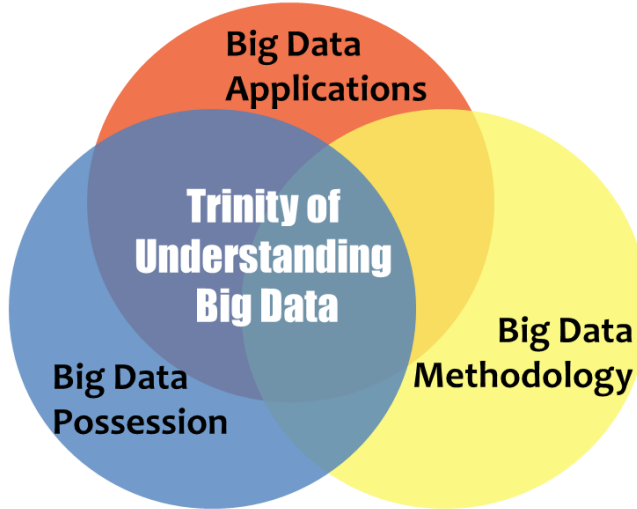
Fig. 13: Trinity for understanding big data, i.e., three facets of big data from different perspectives related to who owns big data, who has innovative big data methods and methodologies, and who needs big data applications [190]. This list can further be extended by big data visualization and big data accelerated computing.

images before and after floods. In this context, it is possible to roughly determine the damaged area by unsupervised learning algorithms, but it is difficult to assess the details (e.g., the damage to transportation infrastructure) [192].

Social media data provides one of the most important data sources from human activities and are usually comprised 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 Global Position System (GPS) receivers. Those data can be disseminated quickly to social networks like Facebook, Twitter, YouTube, Weibo and in particular, messaging apps like SnapChat, WhatsApp, and WeChat. Accordingly, the big data produced by the integration of massive global-view remote sensing data and local-but-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 big data in remote sensing with social media data, as in the definition of connotations of big data in [190], it can be expressed in the context of a trinity framework with three perspectives, i.e., owning data, data applications and data methods. The trinity concept of big data is illustrated in Fig. 13. Accordingly, different perspectives have individual challenges and all the facets of such a trinity share common challenges, which have been 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 how to process and analyze those data by novel methods or methodologies. Since remote sensing data have significantly different properties from those of social media data, typical data fusion methods cannot be exploited directly for combining remote sensing and social medial data. Often, remote sensing data consist of multi-source (laser, radar, optical, etc.), multi-temporal (collected on different dates), and multi-resolution (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, where texts contain different types of textual information, such as geotagging, hashtags, posts, tweets, RSS (Rich Site Summary), etc.. Nevertheless, the challenges for data fusion of 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 of artificial intelligence (AI) techniques in particular, deep neural networks, have merged data representation and data analysis to a unified AI model. In recent decades, high performance computing has developed for data accelerated-based computing on big data platforms, such as Hadoop [202] or SPARK [203]. In particular, with the fast development of artificial intelligence (AI), GPU-accelerated computing by using a graphics processing unit (GPU) and AI (in particular deep learning) chips have developed quickly in recent years for accelerating deep learning computing in a heterogeneous platform by combining CPU, GPU, FPGA, etc..

Social media data, such as photos with geotaggings, can be integrated with remote sensing data in feature or decision or feature-decision levels, respectively, in the context of deep neural networks. In the feature-based data fusion, social media photos (SMPs) can be integrated to the same deep neural network to extract the features for further processing as shown in Fig. 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 the SMPs by different deep neural networks as shown in Fig. 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 get a more reliable and more accurate result. In the decision-based fusion, each deep neural network is designed to firstly extract the features of remote sensing images or social media photos, and then the classification result is generated by individual features. In the decision level, those results provided by the remote sensing and social media data, respectively, are integrated to a unified deep neural network as shown in Fig. 14. In this case, social media data can have diverse types, such as images, texts and so on, such that different types of the social media data can build different deep neural networks for further decision fusion. By combining two properties of feature-based and decision-based fusion strategies, the feature-decision-based fusion can be easily derived based on deep neural networks. The challenge is how to design a unified DNN model to efficiently and accurately fuse the heterogeneous data.
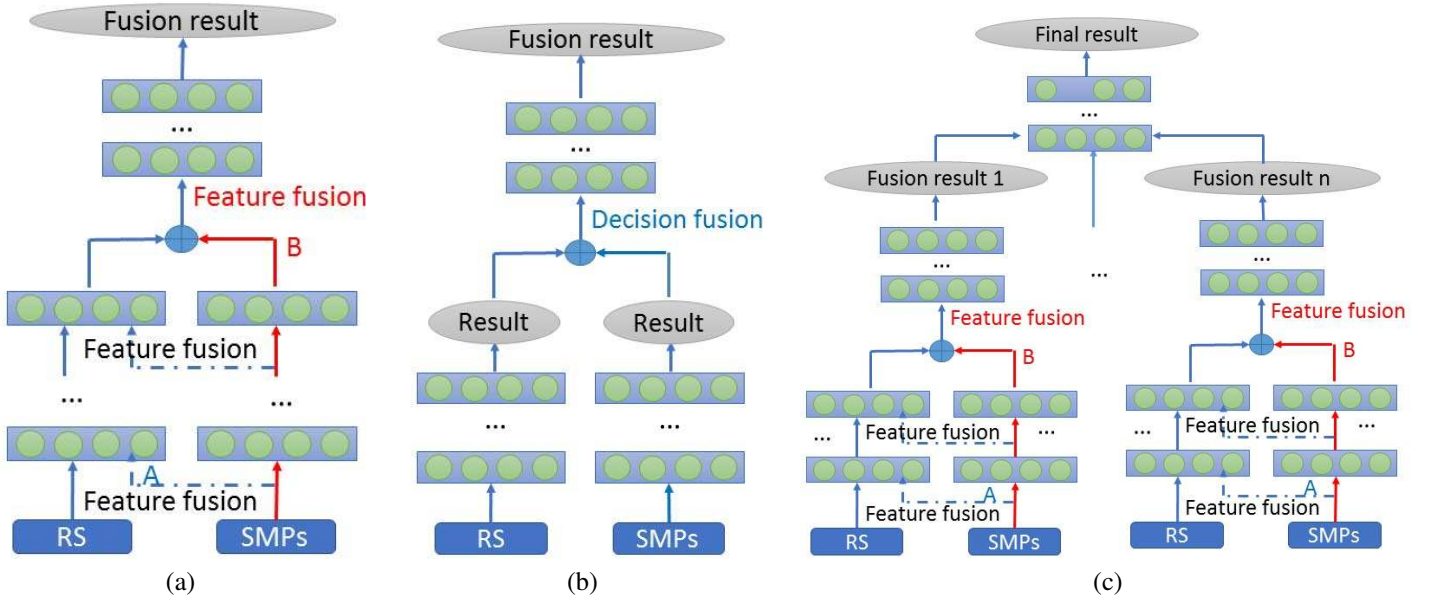
Fig. 14: On the context of deep neural networks, the integration of remote sensing (RS) images with social media photos (SMP): (a) Feature-based fusion, (b) Decision-based fusion, and (c) Feature-decision-based fusion, respectively.

TABLE III: The overall accuracies are compared in terms of the training datasets labeled by volunteers with social media photos (SMPs) and without SMPs, respectively.

Models		OA%
	FCN	78.91
With SMPs	CNN	74.85
	SVM	62.40
	FCN	71.23
Without SMPs	CNN	65.72
	SVM	61.16

Except for directly modeling a deep learning algorithm by the integration of remote sensing and social media data, social media photos can be utilized 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 without any professional knowledge to 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 the Jilin-1 satellite are utilized for remote sensing image classification. The classification models are trained respectively on the training datasets labeled with and without SMPs. Table III 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 pre-trained ImageNet VGG-16 network [205] with the cross-entropy loss. The SVM model with the RBF kernels is adopted for a further comparison. Fig. 15 illustrates Several classification maps obtained by SVM, CNN, and FCN with or without using SMPs.

Except for the challenge of designing a novel data technology by fusing big remote sensing data combining remote sensing and social media data, how to exploit the two different types of data is another challenging problem. As future trends, big remote sensing data will be utilized for monitoring natural hazards as well as human-made disasters, such as factory explosion. In addition, big remote sensing data can provide rich information on the contents and locations of the scenes for Augmented Reality (AR) applications, 3D reconstruction, indoor positioning, and so on.

VII. CONCLUSIONS

The ever-growing increase in the availability of data captured by different sensors coupled with advances in methodological approaches and computational tools makes it desirable to fuse the considerably heterogeneous complementary datasets to increase the efficacy and efficiency of the remotely sensed data processing approaches with respect to the problem 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 focuses particularly on advances in multisource and multitemporal data fusion approaches with respect to different research communities since the methods for the fusion of different modalities have expanded along different paths with respect to each research community. In this context, several vibrant fusion topics, including pansharpening and resolution enhancement, point cloud data fusion, hyperspectral and LiDAR data fusion, multitemporal data fusion, as well as 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 of each section, although the field of remote sensing data fusion is mature, there are still many doors left open for further investigation, both from the theoretical and application

(a) Test Image (b) Segmentation (c) FCN+SMPs

(d) FCN (e) CNN+SMPs (f) CNN

(g) SVM+SMPs (h) SVM

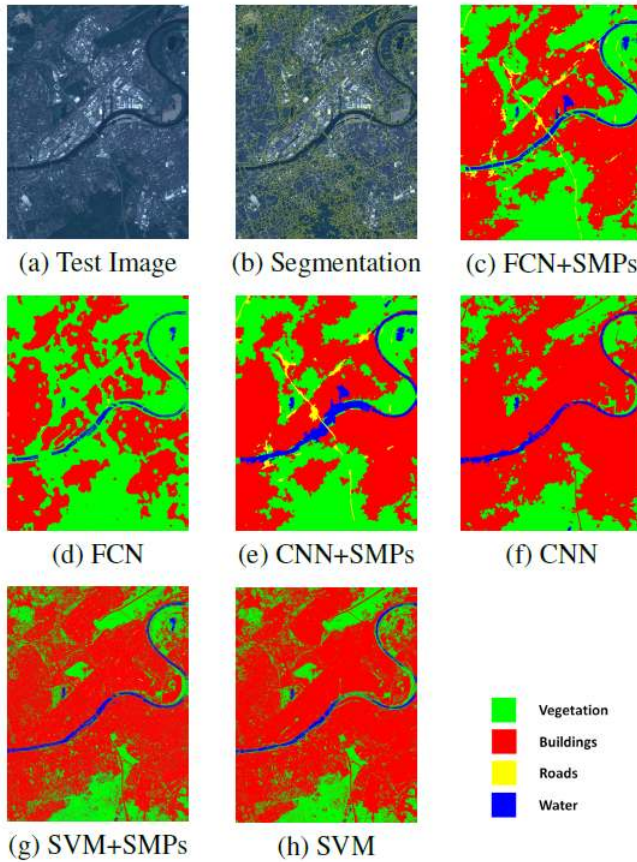■ Vegetation
■ Buildings
■ Roads
■ Water

Fig. 15: Several classification maps obtained by SVM, CNN, and FCN with or without using SMPs.

perspectives. We hope that this review opens up new possibilities for readers to further investigate the remaining challenges to developing sophisticated fusion approaches suitable for the applications at hand.

VIII. ACKNOWLEDGMENT

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 Journal of Photogrammetry and Remote Sensing*, vol. 138, pp. 47 – 56, 2018.

[2] N. Audebert, B. L. Saux, and S. Lefvre, "Beyond rgb: Very high resolution urban remote sensing with multimodal deep networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 20 – 32, 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," *Journal of Applied Remote Sensing*, vol. 11, pp. 042 609–1 – 042 609–54, 2017.

[4] J. Xia, N. Yokoya, and A. Iwasaki, "Fusion of hyperspectral and lidar data with a novel ensemble classifier," *IEEE Geos. Remote Sens. Let.*, vol. 15, no. 6, pp. 957–961, June 2018.

[5] M. Eslami and A. Mohammadzadeh, "Developing a spectral-based strategy for urban object detection from airborne hyperspectral tir and visible data," *IEEE Jour. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 9, no. 5, pp. 1808–1816, May 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. Geos. Remote Sens.*, vol. 56, no. 4, pp. 1873–1886, April 2018.

[7] L. Estes, D. McRitchie, J. Choi, S. Debats, T. Evans, W. Guthe, D. Luo, G. Ragazzo, R. Zempleni, and K. Caylor, "A platform for crowdsourcing the creation of representative, accurate landcover maps," *Environmental Modelling & Software*, vol. 80, pp. 41 – 53, 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, Oct 2017.

[10] N. Yokoya, P. Ghamisi, J. Xia, S. Sukhanov, R. Heremans, I. Tankoyeu, B. Bechtel, B. L. Saux, G. Moser, and D. Tuia, "Open data for global multimodal land use classification: Outcome of the 2017 ieee grss data fusion contest," *IEEE Jour. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 11, no. 5, pp. 1363–1377, May 2018.

[11] H. Ghassemian, "A review of remote sensing image fusion methods," *Information Fusion*, vol. 32, pp. 75 – 89, 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," *Information Fusion*, vol. 33, pp. 100 – 112, 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," *Information Fusion*, vol. 42, pp. 158 – 173, 2018.

[14] R. Booysen, R. Zimmermann, S. Lorenz, R. Gloaguen, P. A. M. Nex, L. Andreani, and R. Möckel, "Towards multiscale and multisource remote sensing mineral exploration using rpas. a case study in the lofdal carbonatite-hosted ree deposit, namibia." *Remote Sensing*, 2018.

[15] S. Nieland, N. Moran, B. Kleinschmit, and M. Frster, "An ontological system for interoperable spatial generalisation in biodiversity monitoring," *Computers & Geosciences*, vol. 84, pp. 86 – 95, 2015.

[16] J. P. Kerekes and D. A. Landgrebe, "Parameter trade-offs for imaging spectroscopy systems [remote sensing]," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 29, no. 1, pp. 57–65, Jan 1991.

[17] G. Joseph, "How well do we understand earth observation electro-optical sensor parameters?" *ISPRS Jour. Photo. 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. Env.*, 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 remote-sensing systems: Application to crism," *IEEE Trans. Geos. Remote Sens.*, vol. 48, no. 11, pp. 3877–3888, Nov 2010.

[20] X. He and X. Xu, "Physically based model for multispectral image simulation of earth observation sensors," *IEEE Jour. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 10, no. 5, pp. 1897–1908, May 2017.

[21] S. Lefvre, D. Tuia, J. D. Wegner, T. Produit, and A. S. Nassar, "Toward seamless multiview scene analysis from satellite to street level," *Proc. IEEE*, vol. 105, no. 10, pp. 1884–1899, Oct 2017.

[22] Q. Wang and P. M. Atkinson, "The effect of the point spread function on sub-pixel mapping," *Remote Sensing of Environment*, vol. 193, pp. 127 – 137, 2017.

[23] P. M. Atkinson, "Downscaling in remote sensing," *Int. Jour. App. Earth Obs. Geo.*, vol. 22, pp. 106 – 114, 2013, spatial Statistics for Mapping the Environment.

[24] A. Garzelli, "A review of image fusion algorithms based on the super-resolution paradigm," *Remote Sens.*, vol. 8, no. 10, 2016.

[25] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Proc.*, vol. 27, no. 7, pp. 3418–3431, July 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, p. 155, 2018.

[27] Q. Wang, W. Shi, P. M. Atkinson, and Y. Zhao, "Downscaling modis images with area-to-point regression kriging," *Remote Sensing of Environment*, vol. 166, no. Supplement C, pp. 191 – 204, 2015.

[28] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic Geog.*, vol. 46, pp. 234–240, 1970.

[29] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P. M. Atkinson, "A hybrid mlp-cnn classifier for very fine resolution remotely sensed image classification," *ISPRS Jour. Photo. Remote Sens.*, vol. 140, pp. 133 – 144, 2018, geospatial Computer Vision.

[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, 2017.

[31] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson, "Fusion of sentinel-2 images," *Remote Sensing of Environment*, 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 Sensing of Environment*, 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. Geos. Remote Sens.*, vol. 55, no. 12, pp. 7126–7139, Dec 2017.

[34] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily sentinel-2 images," *Remote Sensing of Environment*, 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 Geoscience and Remote Sensing Magazine*, vol. 5, no. 1, pp. 8–32, March 2017.

[36] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geos. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec 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. Geos. Remote Sens.*, vol. 55, no. 1, pp. 600–614, Jan 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. Jour. 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, 2017.

[40] Z. Zhu, "Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications," *ISPRS Jour. Photo. Remote Sens.*, vol. 130, pp. 370 – 384, 2017.

[41] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: a comparative review of the recent literature," vol. 5, no. 2, pp. 29–56, June 2017.

[42] R. Welch, "Merging multiresolution spot hrv and landsat tm data," *Photogrammetric Engineering & Remote Sensing*, 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," *Photogramm. Eng. Remote Sens*, vol. 55, 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," *Photo. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, Apr. 1990.

[45] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: the arsis concept and its implementation," *Photogrammetric Engineering and Remote Sensing*, 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," Jan. 4 2000, uS Patent 6,011,875.

[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 Transactions on geoscience and remote sensing*, 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," vol. 45, no. 10, pp. 3230–3239, Oct. 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 transactions on geoscience and remote sensing*, 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 Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 746–750, 2010.

[51] D. A. Yocky, "Multiresolution wavelet decomposition i me merger of landsat thematic mapper and spot panchromatic data," *Photogrammetric Engineering & Remote Sensing*, 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," *International journal of remote sensing*, 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 Transactions on Geoscience and Remote sensing*, 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. Jour. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Jan. 2000.

[55] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods," *IEEE Trans. Geos. 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," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

[57] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Information Fusion*, vol. 8, no. 2, pp. 143–156, 2007.

[58] E. Pardo-Igzquiza, M. Chica-Olmo, and P. M. Atkinson, "Downscaling cokriging for image sharpening," *Remote Sensing of Environment*, 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 Transactions on Geoscience and Remote Sensing*, vol. 46, no. 2, pp. 573–580, Feb 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 Journal of Photogrammetry and Remote Sensing*, 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 Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2250–2259, April 2013.

[62] Q. Wang, W. Shi, and P. M. Atkinson, "Area-to-point regression kriging for pan-sharpening," *ISPRS Journal of Photogrammetry and Remote Sensing*, 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 286–295, Jan 2017.

[64] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," vol. 13, no. 9, pp. 1174–1184, Sep. 2004.

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

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

[67] M. Simões, J. B. Dias, L. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," vol. 53, no. 6, pp. 3373–3388, Jun. 2015.

[68] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," vol. 9, no. 6, pp. 1117–1127, Sep. 2015.

[69] ——, "Fast fusion of multi-band images based on solving a Sylvester equation," vol. 24, no. 11, pp. 4109–4121, Nov. 2015.

[70] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," vol. 49, no. 2, pp. 738–746, Feb. 2011.

[71] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," vol. 51, no. 5, pp. 2827–2836, May 2013.

[72] Q. Wei, J. M. B. Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[73] W. Sun, B. Chen, and D. Messinger, "Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images," *Optical Engineering*, vol. 53, no. 1, p. 013107, 2014.

[74] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hypersharpening: A first approach on SIM-GA data," vol. 8, no. 6, pp. 3008–3024, Jun. 2015.

[75] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral superresolution by coupled spectral unmixing," in *Proc. IEEE ICCV*, Dec. 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 Geoscience and Remote Sensing Letters*, 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, Technical University of Munich, 2017.

[78] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. the Third Conference Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images*, Jan. 2000, pp. 99–103.

[79] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," vol. 6, no. 4, pp. 662–665, Oct. 2009.

[80] F. Gao, T. Hilker, X. Zhu, M. Anderson, J. Masek, P. Wang, and Y. Yang, "Fusing landsat and modis data for vegetation monitoring," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 47–60, Sept 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," *International Journal of Remote Sensing*, vol. 36, no. 17, pp. 4411–4445, 2015.

[82] B. Chen, B. Huang, and B. Xu, "Comparison of spatiotemporal fusion models: A review," *Remote Sensing*, 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 Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2207–2218, Aug 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 Sensing of Environment*, 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 landsatmodis surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection," *Remote Sensing of Environment*, vol. 133, no. Supplement C, pp. 193 – 209, 2013.

[86] B. Huang and H. Song, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3707–3716, Oct 2012.

[87] H. Song and B. Huang, "Spatiotemporal satellite image fusion through one-pair image learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 1883–1896, April 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 Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 2039–2043, Dec 2016.

[89] V. Moosavi, A. Talebi, M. H. Mokhtari, S. R. F. Shamsi, and Y. Niazi, "A wavelet-artificial intelligence fusion approach (waifa) for blending landsat and modis surface temperature," *Remote Sensing of Environment*, 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 Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1984–1988, Dec 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 Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 844–848, Sept 2011.

[92] J. Amors-Lpez, L. Gmez-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," *International Journal of Applied Earth Observation and Geoinformation*, 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 Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 453–457, July 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 Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4980–4991, Aug 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 Sensing of Environment*, vol. 113, no. 9, pp. 1874 – 1885, 2009.

[96] C. W. C. W. L. W. Mingquan Wu, Zheng Niu, "Use of modis and landsat time series data to generate high-resolution temporal synthetic landsat data using a spatial and temporal reflectance fusion model," *Journal of Applied Remote Sensing*, vol. 6, pp. 6 – 6 – 14, 2012.

[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 Sensing of Environment*, 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," *International Journal of Remote Sensing*, 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 Geoscience and Remote Sensing Letters*, vol. 12, no. 6, pp. 1362–1366, June 2015.

[100] D. Xie, J. Zhang, X. Zhu, Y. Pan, H. Liu, Z. Yuan, and Y. Yun, "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, 2016.

[101] Q. Wang, Y. Zhang, A. O. Onojeghuo, X. Zhu, and P. M. Atkinson, "Enhancing spatio-temporal fusion of modis and landsat data by incorporating 250 m modis data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 9, pp. 4116–4123, Sept 2017.

[102] GRSS Data and Algorithm Standard Evaluation website. [Online]. Available: http://dase.grss-ieee.org/

[103] A. V. Vo, L. Truong-Hong, D. F. Laefer, D. Tiede, S. d'Oleire Oltmanns, A. Baraldi, M. Shimoni, G. Moser, and D. Tuia, "Processing of extremely high resolution lidar and rgb data: Outcome of the 2015 ieee grss data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 12, pp. 5560–5575, Dec 2016.

[104] J. Zhang and X. Lin, "Advances in fusion of optical imagery and lidar point cloud applied to photogrammetry and remote sensing," *International Journal of Image and 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," *The Photogrammetric Record*, vol. 29, no. 146, pp. 144–166, 2014.

[106] X. X. Zhu and M. Shahzad, "Facade reconstruction using multiview spaceborne tomosar point clouds," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3541–3552, June 2014.

[107] M. Hämmerle and B. Höfle, "Mobile low-cost 3d camera maize crop height measurements under field conditions," *Precision Agriculture*, Oct 2017.

[108] J. Shan and C. Toth, "Topographic laser ranging and scanning : Principles and processing," *CRC Press/Taylor and Francis Group*, 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 International Journal of Geo-Information*, vol. 2, no. 4, 2013.

[110] J. U. Eitel, B. Hfle, L. A. Vierling, A. Abelln, G. P. Asner, J. S. Deems, C. L. Glennie, P. C. Joerg, A. L. LeWinter, T. S. Magney, G. Mandlburger, D. C. Morton, J. Mller, and K. T. Vierling, "Beyond 3-d: The new spectrum of lidar applications for earth and ecological sciences," *Remote Sensing of Environment*, vol. 186, pp. 372 – 392, 2016.

[111] K. Koenig and B. Höfle, "Full-waveform airborne laser scanning in vegetation studiesa review of point cloud and waveform features for tree species classification," *Forests*, vol. 7, no. 9, 2016.

[112] S. Filin and N. Pfeifer, "Segmentation of airborne laser scanning data using a slope adaptive neighborhood," *ISPRS Journal of Photogrammetry and Remote Sensing*, 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," *Computers & Geosciences*, vol. 54, pp. 249 – 258, 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 *IGARSS 2015*, 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. The Eurographics Association, 2017.

[116] R. Hänsch, T. Weber, and O. Hellwich, "Comparison of 3d interest point detectors and descriptors for point cloud fusion," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3, pp. 57–64, 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 Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3679–3693, July 2017.

[118] T. L. Swetnam, J. K. Gillan, T. T. Sankey, M. P. McClaran, M. H. Nichols, P. Heilman, and J. McVay, "Considerations for achieving cross-platform point cloud data fusion across different dryland ecosystem structural states," *Frontiers in Plant Science*, vol. 8, p. 2144, 2018.

[119] P. Ghamisi and B. Höfle, "Lidar data classification using extinction profiles and a composite kernel support vector machine," *IEEE Geoscience and Remote Sensing Letters*, 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 Sensing of Environment*, vol. 195, pp. 30 – 43, 2017.

[121] M. Gerke and J. Xiao, "Fusion of airborne laserscanning point clouds and images for supervised and unsupervised scene classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 78 – 92, 2014.

[122] H. Wang and C. Glennie, "Fusion of waveform lidar data and hyperspectral imagery for land cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 108, pp. 1 – 11, 2015.

[123] Q. Man, P. Dong, and H. Guo, "Pixel- and feature-level fusion of hyperspectral and lidar data for urban land-use classification," *International Journal of Remote Sensing*, 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 Jour. Photo. Remote Sens.*, vol. 67, pp. 134–147, 2012.

[125] M. Alonzo, B. Bookhagen, and D. A. Roberts, "Urban tree species mapping using hyperspectral and lidar data fusion," *Remote Sensing of Environment*, vol. 148, pp. 70 – 83, 2014.

[126] N. Saarinen, M. Vastaranta, R. Näsi, T. Rosnell, T. Hakala, E. Honkavaara, M. A. Wulder, V. Luoma, A. M. G. Tommaselli, N. N. Imai, E. A. W. Ribeiro, R. B. Guimarães, M. Holopainen, and J. Hyyppä, "Uav-based photogrammetric point clouds and hyperspectral imaging for mapping biodiversity indicators in boreal forests," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-3/W3, pp. 171–175, 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 Sensing of Environment*, vol. 200, pp. 170 – 182, 2017.

[128] E. T. Robert S Rand, Timothy S Khuon, "Distributed adaptive framework for multispectral/hyperspectral imagery and three-dimensional point cloud fusion," *Optical Engineering*, vol. 55, pp. 55 – 55 – 14, 2016.

[129] K. Karila, L. Matikainen, E. Puttonen, and J. Hyyppa, "Feasibility of multispectral airborne laser scanning data for road mapping," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 3, pp. 294–298, March 2017.

[130] J. Vauhkonen, T. Hakala, J. Suomalainen, S. Kaasalainen, O. Nevalainen, M. Vastaranta, M. Holopainen, and J. Hyyppa, "Classification of spruce and pine trees using active hyperspectral lidar," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1138–1141, 2013.

[131] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Chapters 5 and 6, Artech House Publishers, INC, Boston, USA, 2015.

[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 Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2799–2810, May 2017.

[133] R. K. Hall, R. L. Watkins, D. T. Heggem, K. B. Jones, P. R. Kaufmann, and S. B. Moore, "Quantifying structural physical habitat attributes using lidar and hyperspectral imagery," *Environmental Monitoring and Assessment*, vol. 159, pp. 63–83, 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 Jour. Sel. Topi. App. Earth Obs. Remote Sens.*, vol. 8, no. 5, pp. 2177–2187, May 2015.

[135] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416–1427, May 2008.

[136] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE Jour. Selec. Top. App. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, June 2014.

[137] E. J. Ientilucci, "New SHARE 2010 HSI-LiDAR dataset: re-calibration, detection assessment and delivery," in *Imaging Spectrometry XXI*, vol. 9976, Sep. 2016, p. 99760I.

[138] B. Rasti, P. Ghamisi, and R. Gloaguen, "Hyperspectral and lidar fusion using extinction profiles and total variation component analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3997–4007, July 2017.

[139] M. Zhang, P. Ghamisi, and W. Li, "Classification of hyperspectral and lidar data using extinction profiles with feature fusion," *Remote Sensing Letters*, 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 Jour. Sel. Top. App. Earth Obs. 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 Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3265–3276, June 2016.

[142] P. Ghamisi, B. Höfle, and X. X. Zhu, "Hyperspectral and lidar data fusion using extinction profiles and deep convolutional neural network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 3011–3024, June 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 Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–12, 2017.

[144] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, 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. Beneiktsson, X. X. Zhu, L. Rittner, and R. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, 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 Transactions on Geoscience and Remote Sensing*, 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," *International Journal of Image and 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 Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 6976–6985, Dec 2016.

[150] B. Rasti, J. R. Sveinsson, and M. O. Ulfarsson, "Wavelet-based sparse reduced-rank regression for hyperspectral image restoration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6688–6698, Oct 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, pp. 99 701U–99 701U–7, 2016. [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 classfication," *IEEE Geoscience and Remote Sensing Letters*, 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 Geoscience and Remote Sensing Letters*, vol. 14, no. 8, pp. 1253–1257, Aug 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, 2018.

[155] P. Ghamisi and N. Yokoya, "Img2dsm: Height simulation from single imagery using conditional generative adversarial net," *IEEE Geoscience and Remote Sensing Letters*, vol. PP, no. 99, pp. 1–5, 2018.

[156] C. Kuenzer, S. Dech, and W. Wagner, *Remote Sensing Time Series Revealing Land Surface Dynamics: Status Quo and the Pathway Ahead*. Cham: Springer International Publishing, 2015, pp. 1–24. [Online]. Available: https://doi.org/10.1007/978-3-319-15967-6_1

[157] C. Gomez, J. C. White, and M. A. Wulder, "Optical remotely sensed time series data for land cover classification: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, 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 Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 8–26, Sept 2015.

[159] A. Singh, "Digital change detection techniques using remotely sensed data," vol. 10, pp. 989–1003, 11 1988.

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

[161] [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*. Elsevier, 12 2017, vol. 2, 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 Sensing*, vol. 10, no. 4, 2018.

[164] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1428–1441, May 2008.

[165] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using vhr optical and sar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2403–2420, May 2010.

[166] Y. T. Solano-Correa, F. Bovolo, L. Bruzzone, and D. Fernndez-Prieto, "Spatio-temporal evolution of crop fields in sentinel-2 satellite image time series," in *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, June 2017, pp. 1–4.

[167] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor, "Detecting trend and seasonal changes in satellite image time series," *Remote Sensing of Environment*, 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, pp. 418–422, 2016.

[169] A. Zanchetta, G. Bitelli, and A. Karnieli, "Monitoring desertification by remote sensing using the tasselled cap transform for long-term change detection," *Natural Hazards*, vol. 83, no. 1, pp. 223–237, Oct 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 Sensing of Environment*, 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 Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1350–1359, May 1999.

[172] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 858–867, Jul 1997.

[173] Y. Guo, X. Jia, and D. Paull, "Effective sequential classifier training for svm-based multitemporal remote sensing image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3036–3048, June 2018.

[174] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 6, pp. 1321–1334, June 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, Oct. 2004. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2004.06.002

[176] "Bayesian classification in a time-varying environment," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 12, pp. 879–883, Dec 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 Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1108–1122, April 2009.

[178] F. Bovolo, G. Camps-Valls, and L. Bruzzone, "A support vector domain method for change detection in multitemporal images," *Pattern Recognition Letters*, vol. 31, no. 10, pp. 1148 – 1154, 2010, pattern Recognition in Remote Sensing. [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 Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 1984–1996, Sep 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," *Information 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 Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 778– 782, May 2017.

[182] D. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel, "Land cover classification via multitemporal spatial data by deep recurrent neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1685–1689, Oct 2017.

[183] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 456–460, Feb 2001.

[184] ——, "A partially unsupervised cascade classifier for the analysis of multitemporal remote-sensing images," *Pattern Recognition Letters*, 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 Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1930–1941, May 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 Transactions on Geoscience and Remote Sensing*, vol. 50, no. 7, pp. 2810–2826, July 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 Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 300–312, Jan 2013.

[188] ——, "Classification of time series of multispectral images with limited training data," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3219–3233, Aug 2013.

[189] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A dasvm classification technique and a circular validation strategy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 770–787, May 2010.

[190] M. Chi, P. A., B. J. A., Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proceedings of the IEEE*, vol. 104, no. 11, pp. 2207–2219, Nov. 2016.

[191] Xinhua, "China to launch 60 jilin-1 video satellites by 2020," http://www.chinadaily.com.cn/a/201801/27/WS5a6c360ea3106e7dcc137247.html, January 2018.

[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," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1926 – 1936, Oct. 2017.

[193] J. M. Shapiro, "Smart cities: Quality of life, productivity, and the growth effects of human capital," *Review of Economics and Statistics*, vol. 88, no. 2, pp. 324–335, 2006.

[194] P. Gamba, "Human settlements: A global challenge for EO data processing and interpretation," *Proceedings of the 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," *Proceedings of the 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," *Proceedings of the IEEE*, vol. 100, no. 10, pp. 2946–2970, 2012.

[197] C. Aggarwal and T. Abdelzaher, *Managing and Mining Sensor Data*. Springer, 2013, ch. Social Sensing, pp. 237–275.

[198] R. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, p. 32?C39, 2011.

[199] L. Barrington, S. Ghosh, M. Greene, and S. Harnoy, "Crowdsourcing earthquake damage assessment using remote sensing imagery," *Annals of geophysics*, vol. 54, no. 6, pp. 680–687, 2011.

[200] E. Schnebele, C. Oxendine, G. Cervone, C. M. Ferreira, and N. Waters, *Computational Approaches for Urban Environments*. Springer, 2015, ch. 14, Using Non-authoritative Sources During Emergencies in Urban Areas, 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," *International Journal of Remote Sensing*, vol. 37, no. 1, pp. 100–124, Dec. 2015.

[202] T. White, *Hadoop: The Definitive Guide*, 1st ed. 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 *Proceedings of the 2Nd USENIX Conference on Hot Topics in Cloud Computing*, ser. HotCloud'10. Berkeley, CA, USA: USENIX Association, 2010, pp. 10–10.

[204] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *CoRR*, vol. abs/1411.4038, 2014. [Online]. Available: http://arxiv.org/abs/1411.4038

[205] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.