INVITED PAPER

Big Data for Remote Sensing: Challenges and Opportunities

By Mingmin Chi, *Member IEEE*, Antonio Plaza, *Fellow IEEE*, Jón Atli Benediktsson, *Fellow IEEE*, Zhongyi Sun, Jinsheng Shen, and Yangyong Zhu

ABSTRACT | Every day a large number of Earth observation (EO) spaceborne and airborne sensors from many different countries provide a massive amount of remotely sensed data. Those data are used for different applications, such as natural hazard monitoring, global climate change, urban planning, etc. The applications are data driven and mostly interdisciplinary. Based on this it can truly be stated that we are now living in the age of big remote sensing data. Furthermore, these data are becoming an economic asset and a new important resource in many applications. In this paper, we specifically analyze the challenges and opportunities that big data bring in the context of remote sensing applications. Our focus is to analyze what exactly does big data mean in remote sensing applications and how can big data provide added value in this context. Furthermore, this paper describes the most challenging issues in managing, processing, and efficient exploitation of big data for remote sensing problems. In order to illustrate the aforementioned aspects, two case studies discussing the use of big data in remote sensing are demonstrated. In the first test case, big data are used to automatically detect marine oil spills using a large archive of remote sensing data. In the second test case, content-based information retrieval is performed using high-performance computing (HPC) to extract information from a large database of remote sensing images, collected after the terrorist attack to the World Trade Center in New York City. Both cases are used to illustrate the significant challenges and opportunities brought by the use of big data in remote sensing applications.

KEYWORDS | Big data; big data challenges; big data life cycle; big data opportunities; high-performance computing (HPC); remote sensing

I. INTRODUCTION

As moving data generators, human beings create data everyday. We are all connected by sharing data from social networks, intelligent devices, etc. Remote sensing devices have been widely used to observe our planet from various perspectives and to make our lives easier. It is not exaggerated to say that the whole Earth has now been made digital. Therefore, the digitized Earth plus the moving data generators are the main actors for big data in remote sensing, which can be used to make governments more efficient (e.g., improving services like police, healthcare and transportation) and also for business, i.e., to improve decision making, manufacturing, product innovation, consumer experience and service, etc.

As reported by IBM, 2.5 quintillion bytes of data are now generated every day. In other words, "90% of the data in the world today has been created in the last two years alone."[1] We are truly living in the big data age, and now government leaders, enterprises, and nonprofit organizations are quickly realizing that it is very important to

[1]"What is big data?" in http://www-01.ibm.com/software/data/bigdata/.

Manuscript received August 10, 2015; revised November 30, 2015; accepted November 30, 2015. This work was supported in part by the Natural Science Foundation of China under Contract 71331005, and in part by the Open Foundation of Second Institute of Oceanography (SOA) under Contract SOED1509.

M. Chi is with the School of Computer Science, Shanghai Key Laboratory of Data Science, Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China, and also with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography (SOA), Hangzhou 310012, China (e-mail: mmchi@fudan.edu.cn).

A. Plaza is with the Department of Technology of Computers and Communications, Escuela Politécnica de Cáceres, University of Extremadura, E-10003 Cáceres, Spain (e-mail: aplaza@unex.es).

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

Z. Sun, J. Shen, and Y. Zhu are with the School of Computer Science, Shanghai Key Laboratory of Data Science, Fudan University, Shanghai 200433, China (e-mail: 14210240051@fudan.edu.cn; 14210240050@fudan.edu.cn; yyzhu@fudan.edu.cn).

Digital Object Identifier: 10.1109/JPROC.2016.2598228

collect big data in different contexts. However, there still exists a common problem related to how we can gain insights into big data. This problem is a conundrum: On one hand, a wealth of big data can bring us big opportunities. On the other hand, we still do not know how to harness such big amount of data with tremendous complexity, diversity, and heterogeneity, yet with high potential values. This makes the data very difficult to process and analyze in a reasonable time.[2]

Big data can be mainly characterized by three features: volume, variety, and velocity, defined as three "V" dimensions by Meta Group (now Gartner) in 2001 [1]. It is worth noting that "value" is an important quality of big data, but it is not a defining characteristic. Big remote sensing data can be described by its own dimensions (referred hereinafter as 3Vs).

1) The archived data are characterized by their increasing volume, from terabytes (TB = 1024 GB) to petabytes (PB = 1024 TB), and even to exabytes (EB = 1024 PB). For instance, a huge amount of remote sensing data are now freely available from the NASA Open Government Initiative.[3] Only one of NASA archives, the Earth Science Data and Information System (ESDIS), holds 7.5 PB of data with nearly 7000 unique data sets and 1.5 million users in 2013 [2]. This volume only contains in-domain remote sensing data.
2) In terms of variety, we can see now that big remote sensing data consist of multisource (laser, radar, optical, etc.), multitemporal (collected on different dates), and multiresolution (different spatial resolution) remote sensing data, as well as data from different disciplines depending on several application domains [3].
3) The velocity of big data in remote sensing involves not only generation of data at a rapid growing rate, but also efficiency of data processing and analysis. In other words, the data should be analyzed in a (nearly) real or a reasonable time to achieve a given task, e.g., seconds can save hundreds of thousands of lives in an earthquake.

Although the 3Vs can describe big data, we consider that it is not necessary for big data in remote sensing to satisfy all the three V dimensions. For instance, any one of volume and velocity, volume and variety, or variety and velocity can already define a big data problem. Except for the common challenges of big data characterized by the 3Vs, there are other challenges for the remote sensing applications, such as extensibility to integrating multiple disparate management systems for different satellites for a remote sensing data center [4]. Of particular importance is the value of the data, an important quality hidden in the big data. Data processing methods can be utilized to discover such value, and then the value of big data can be realized in a real remote sensing application.

Therefore, to better understand big data, three perspectives should be unified, i.e., owning data, data applications, and data methods. In the paper, a trinity framework is proposed to better understand big data in the context of remote sensing applications. All the facets of such trinity share common challenges and different perspectives have individual challenges of its own.

In this work, these common and individual challenges are discussed in the context of remote sensing applications. In spite of such big challenges, the potentials of big remote sensing data are presented in detail. These potentials have been applied to deal with different real-world problems, such as archaeology [5], crop assessment and yield forecasting [6], [7], food security [8], human health [9], [10], land development and use [11], urban planning, management, and sustainability [12]–[15], forest monitoring [16], war and conflict studies [17], and several others. To illustrate the effectiveness of big remote sensing data, two case studies discussing the use of big data in remote sensing are demonstrated in this paper. In the first test case, social media data together with remote sensing images are identified to consist of big remote sensing data for automatical marine oil spill detection and then a new data methodology is adopted to deal with labeling challenges. In the second test case, content-based information retrieval is performed using high-performance computing (HPC) to extract information from a large database of remote sensing images, collected after the terrorist attack on the World Trade Center in New York City on September 11, 2001.

The remainder of the paper is organized as follows. The next section discusses our understanding on big data in remote sensing from three different perspectives. According to our view on big data, Section III divides big data challenges into common challenges for all remote sensing applications and individual challenges in individual facets of the so-called trinity of big data. Then, the potentials of big remote sensing data are presented in Section IV. Section V presents the aforementioned case studies of big data in remote sensing. Finally, Section VI draws some conclusions of the work and discusses future developments.

II. UNDERSTANDING BIG DATA IN REMOTE SENSING

From a general perspective, we can understand big data as having different connotations regarding those who own the big data, those who can process and analyze the big data, and those who utilize the big data. Accordingly, different data methods may be exploited to tackle big data challenges in order to efficiently derive the value of

[2]The 462nd Session of the Xiangshan Science Conference, Beijing, China, May 29–31, 2013.

[3]http://www.nasa.gov/open/

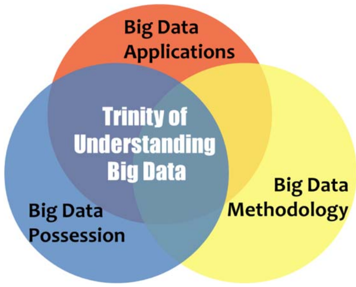


Fig. 1. ***Trinity for understanding big data, i.e., three facets of big data from different perspectives related to who owns big data, who has innovative big data methods and methodologies, and who needs big data applications.***

those data. In the following, a trinity (three in one) is discussed for the understanding of big data (with particular focus on remote sensing applications). Here, we identify three facets for understanding big data, i.e., owning data, data methods, and data applications, which contribute together to a single big data life cycle. The trinity concept of big data is illustrated in Fig. 1. There are common and different challenges in the individual facets of understanding big data, which are detailed next.

A. First Facet: Owning Data

This is an important aspect of big data based on which we can identify applications and utilize or design proper data methods to address a real problem (e.g., a remote sensing problem). The corresponding opportunities are based on the fact that more diverse data can be acquired by intelligent devices where most of human beings have access to the internet now to become both individual and moving data generators. Accordingly, data values can be derived from those complex, diverse, heterogeneous, and high-dimensional remote sensing data and other data from cyberspace. However, big challenges arise at each step when obtaining and organizing big remote sensing data. For instance, remote sensing data are acquired from satellites, airplanes, or other sensing devices while the other forms of data are retrieved from cyberspace. Remote sensing data are preprocessed by geometric and radiometric correction, georeferencing, noise removal, etc. [18], and the data from cyberspace should be cleaned to reduce errors and noise, in which data quality can be improved. Remote sensing data should be delivered from satellites to ground stations, and from ground stations to customers. Other related issues are data compression, data archiving, data retrieval, data rights and protection, etc. We emphasize that data are of no value until they are utilized for applications. The key difference between traditional data and big data is how to identify the right data sets and how to combine them to solve a challenging or novel problem.

B. Second Facet: Big Data Methodologies

A big data methodology should be designed to systematically address big data problems from different remote sensing domains. Such methodology is used to design new data methods for big remote sensing data preparation, data deployment, information extraction, data modeling, data fusion, data visualization, and data interpretation. These aspects are particularly crucial in remote sensing applications, in which preprocessing steps are as equally important as information extraction steps. However, data processing and analysis represent a multistep pipeline and data-driven methods could be significantly different from the viewpoint of specific applications and domains.

Due to the aforementioned heterogeneity and high dimensionality of big data in remote sensing, we also face important computational and statistical challenges related to processing scalability, noise accumulation, spurious correlation, incidental endogeneity, and measurement errors [19], [20]. These challenges require new computational and statistical techniques in order to tackle big data analysis and processing. The analysis and processing techniques are data driven and can benefit from theories and methods from the fields of statistics, machine learning, pattern recognition, artificial intelligence, data mining, etc. Domain knowledge is another crucial aspect that should be tightly linked to data analysis.

C. Third Facet: Big Data Applications

A main goal in big data applications is to identify the right data to solve the problems at hand, which are difficult to be addressed or mostly cannot be manipulated by traditional remote sensing data. Then, the next problem is how to collect, organize, and utilize these big data to deal with real remote sensing problems.

To identify the right data, we should be closely linked to the first facet of understanding big data. In other words, to harness big data firstly one should obtain data from the related data agents (or, in general, data industry or organization). In order to access the data, collaboration across domains or organization should be taken into account in an efficient manner. This is one of crucial challenges in remote sensing applications.

After obtaining the right data, such as remote sensing data, textual data and pictures from social networks, innovative data methodologies should be developed to discover, realize, and demonstrate the value of big data for remote sensing applications.

III. BIG DATA, BIG CHALLENGES

The challenges of big data in remote sensing involves not only dealing with high volumes of data [21]. In particular, challenges on data acquisition, storage, management, and analysis are also related to remote sensing problems

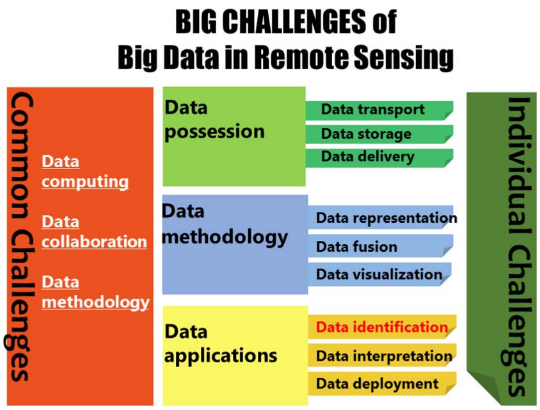


Fig. 2. ***A summary of the challenges introduced by big data.***

involving big data. In this section, we particularly analyze the challenges of big data in remote sensing which involve the different facets of understanding big data in the previous section.

From different perspectives of understanding big data, we are facing big challenges in leveraging the value that data have to offer. In the three facets, the same challenges are shared, such as data computing, data collaboration, and data methodologies for different applications; in the meantime, we are facing different challenges in the individual facets of understanding big data. Fig. 2 summarizes the common and different challenges, which are described in detail in subsequent sections.

A. Common Challenges

In the following, three common challenges, i.e., big data computing, big data collaboration, and big data methodologies, are listed according to the trinity of understanding big data in remote sensing.

1) Big Data Computing: A challenge in the design of high-performance systems for big data computing is to develop more heterogeneous systems able to integrate resources in different locations [22]. Although cloud computing systems have been shown to realize a high level of aggregate performance in remote sensing applications, there are still challenges remaining regarding the progressive incorporation of the concept of cloud computing to remote sensing studies [23]. The ultimate goal should be making distributed collections of data easy to access from different users. However, a remaining challenge is the energy consumption, which is still difficult to leverage in massively parallel platforms or even in onboard processing scenarios. Addressing these challenges will be important for the full incorporation of big data computing techniques to remote sensing applications. Literature for big data in remote sensing mainly focuses on the voluminous issue of big data computing and considers it as a data-intensive computing problem [24]. Usually, an HPC paradigm is exploited for (nearly) real-time big data processing [20], [23], [25].

2) Big Data Collaboration: The ownership of data in remote sensing problems is generally fragmented across data agents or industries [26]. Accordingly, data access and connectivity can be an obstacle. Legitimate concerns can be raised to achieve cross-sector collaboration which motivates data sharing, such as social text or social media. However, individuals often resist to sharing personal data due to security and privacy. This is contradictory to the idea of data personalization. In addition, numerous data firms regard big data as proprietary and thus do not obtain an incentive to share data. Concurrently, it is an important challenge for government institutions to share data unless all participants can achieve material benefits and incentives in data sharing that outweigh the risks [27]. For instance, even if NASA is now sharing a significant amount of remote sensing data under the open government initiative,[4] most high-quality, high-spatial-resolution images are still unavailable to the public. Therefore, it is necessary to find new ways of collaboration for improved big data access in remote sensing problems.

3) Big Data Methodologies: The problem of analyzing big data in remote sensing can be simply formalized as follows. Let $\mathbf{X}$ be an input data set and let $f(\mathbf{X})$ be a mapping function between an input $\mathbf{x} \in \mathbf{X}$ and the output $\mathbf{y}$. Then, a common data analysis task can be formulated as

$$\mathbf{y} = f(\mathbf{X})$$

where the corresponding processing can be carried out in the memory of a computer containing the data input.

However, big data analysis should generally adopt a mechanism to partition the data input into a distributed and/or parallel architecture, i.e., $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N\}$, which means splitting the bigger set $\mathbf{X}$ to N smaller data sets. The adopted data methods or algorithms, i.e., $f(\cdot)$, should be modified to satisfy the new computing environments. Although this is, in general, a simplification (as the smaller data sets may not be easy to process independently and involve some synchronization and/or communication in the associated processing task), an important challenge for this processing scheme is that not all exiting algorithms can be distributed or efficiently implemented in parallel form. Even if data processing methods can do so, it is challenging to collect the distributed data and to deliver those data to the right computing node. As a result, big data processing in general (and

[4] https://www.opengov.com

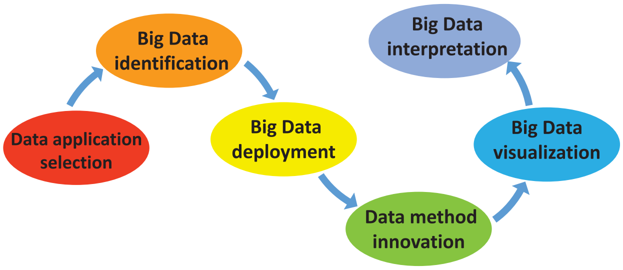


Fig. 3. ***Life cycle to address big data tasks in remote sensing applications.***

in remote sensing in particular) needs new computational and statistical paradigms with regards to standard data processing strategies.

B. Individual Challenges

In this section, a few crucial challenges are discussed in the context of designing a big remote sensing data life cycle (see Fig. 3). After understanding the business need (e.g., a remote sensing application involving big data), some important steps are to identify the right type of data across different disciplines, to deploy the big data, to utilize or design innovative data methods, and finally to visualize and interpret the obtained results. Here, data methods include data analysis, data modeling, data processing, etc.

1) Proper Data Identification: Big remote sensing data usually include in-domain data and out-domain data. In the past, those different formats of data have been seldom combined to fulfil remote sensing applications/tasks. Therefore, the data are of no value unless harnessed to accomplish a specific task. Accordingly, the key difference between traditional remote sensing data and big remote sensing data lies in how to select and combine different formats of data to address real-world problems previously deemed intractable. This is a key challenge of big data, i.e., how to identify and exploit the proper data to solve the problem at hand.

In remote sensing, we have many different kinds of data [28], including optical (e.g., multispectral and hyperspectral), radar [e.g., synthetic aperture radar (SAR)], or laser [e.g., light detection and ranging (LiDAR)] provided by airplane or satellite or ground sensors. Other kinds of data sources can also be integrated in remote sensing problems, i.e., internet textual data (e.g., news, web logs, etc.) can be used to help labeling data patterns provided by remote sensors [3], such as through active learning [29] or crowdsourcing techniques [30], which involve low or no cost. Also, image data taken by individuals from social networks can be taken into account for assisting in remote sensing data interpretation tasks. Other data formats such as census data, meteorological data, intelligent transportation data, high-fidelity geographical data, healthcare data, and so on, can be of significant help to solve a specific real-world problem, e.g., monitoring food security [31].

2) Challenges in Data Possession: After the data have been transmitted to the ground station, those data should be stored in a system. A data storage system usually consists of hardware and software components. In the former, the hardware infrastructure should be flexibly adapted to different application environments. In terms of software, the data storage system is usually equipped with various interfaces, data archives, and queries from web services for users' interactions. With the rapid growth of remote sensing data, traditional structured related database management systems (RDBMSs) cannot meet the requirements of managing big data in remote sensing. Accordingly, it is urgent to adopt or to design a novel data storage system which can meet the rapid growth of big remote sensing data in PB scale or larger. This general discussion on big data storage can be referred to [32].

Data delivery provides access to remote sensing data and metadata to users, both at main ground stations or networks of receiving ground station. Usually, this consists of graphical web portals that provide access to data and searching of metadata to users. Traditionally, users download data of interest from a central archive to their local computers for analysis. This cannot work in big data applications as the sharp growth of data sizes cannot allow the current system to deliver the data to users for local computing. In particular, if an emergency such as an earthquake occurred, a large amount of data should be received for data analysis in a very short time, as few seconds can save many lives by timely warnings. This is another big challenge for those owning big data, as extremely diverse and high-dimensional data should be delivered and analyzed in a short time interval due to the volume and velocity properties of big data. Therefore, a real-time big data analysis platform should be developed to deal with online remote sensing data together with offline data in local data center or from distributed data centers for a real-time application, such as weather forecast, hazard warning, etc.

3) Data Deployment: As discussed in Section III-B1, a critical challenge is to identify the proper data source to achieve a specific goal which is difficult to fulfill without big data. Another challenge of big data is how to deploy the data for real applications. In the phase of big data application, big data deployment encompasses data preparation, data management technologies, data methods and techniques, and so on. That is, how to obtain the data, how to store the data in the computing environment, and how to build models to get insight of big data should be carefully designed in the big data deployment step.

Due to the volume and velocity properties of big data, traditional methods cannot be used for deployment purposes. Accordingly, new technologies should be taken into account, such as distributed data management technologies, schema-less data models, active visualization techniques and so on, to gain meaningful insights on the big data.

4) Data Representation: Various sources of remote sensing data have different spectral and spatial resolutions and usually are acquired on different dates [33]. For instance, in optical data the spectral signature of every material is unique in a laboratory measure. However, spectral signatures of field data are changeable due to variation of materials, environmental effects, surface contaminants, adjacency effects by nearby objects, seasonal changes, and so on [34]. This can lead to the phenomenon that similar signatures could denote different objects, while different signatures might denote the same object. This phenomenon is similar to the "semantic gap" observed in computer vision, i.e., the divergence between the information coming from data and the knowledge interpreted by users [35]. In recent years, deep neural networks (DNNs) [36] have successfully addressed the classification problem in computer vision to fill the semantic gap via automatic feature extraction in a deep manner [37]. Although DNNs have been adopted for feature selection and classification tasks when analyzing remote sensing images, in most of the works conducted so far only spectral features or the transformed spectral features (e.g., principal components) are used as inputs to the DNNs to generate a "better" representation of remote sensing images [38]–[40]. In the published work, usually components from the original feature vectors correspond to the spectral reflectance of land-cover targets, which means that the feature components have clear physical meaning. In this regard, remote sensing images can be interpreted as structured data. Although classification accuracy and robustness can be slightly improved by incorporating unlabeled samples for feature learning, it is not clear how much the classification performance can benefit from a "better" feature representation due to lack of large amounts of training data and the limited number of layers that can be used in practice when implementing neural networks. For instance, the classification accuracy cannot be significantly improved with more than five layers as described in number of layers adopted in [40].

Besides, although various types of remote sensing data (acquired by different sensors, from different locations in different dates) are acquired and exploited to deal with a challenging application problem together with out-of-remote-sensing data, existing data methods cannot manipulate those data to retrieve the value of those data. Meanwhile, remote sensing data comprise different dimensions and spatial resolutions, such as spaceborne multispectral moderate resolution imaging spectroradiometer (MODIS) in 36 spectral bands with ground spatial 250 m (bands 1–2), 500 m (bands 3–7), and 1000 m (bands 8–36)[5] and airborne hyperspectral reflective optics system imaging spectrometer (ROSIS) with ground resolution less than 1 m in 115 bands)[6] Furthermore, the representations of the out-of-remote-sensing data could be unstructured (e.g., individual pictures), which are significantly different from those of optical or microwave remote sensing data. Therefore, different data representation becomes a big obstacle for the exploitation of big remote sensing data.

5) Data Fusion: Due to the data representation challenge discussed in Section III-B4, a follow-up challenge is how to integrate the data from various sources, where data features are significantly different (e.g., spectral signatures in optical remote sensing data, electromagnetic radiation in microwave data, structural features of texts, unstructured features of images by a digital camera, etc.).

Traditionally, data fusion can be carried out in terms of pixel-level fusion, feature-level fusion, and decision-level fusion [41]. However, big data in remote sensing usually comprise different scales and/or formats. As a result, traditional approaches cannot be utilized to integrate the information for data fusion. Therefore, new methods should be developed to tackle the fusion of big data in remote sensing. For instance, in urban applications, each pixel can be annotated by photos taken by individuals from a social network in the same location [3] by means of a crowdsourcing technique [30]. Measuring correlation between different sources of data also becomes an additional challenge by the aid of artificial intelligence, data mining, machine learning, or statistics.

6) Data Visualization and Interpretation: Visualization not only enables users/decision-makers to gain better insights into big data, but is also important to understand and analyze big data in remote sensing to bring out data details relevant for the current aims or objectives. Accordingly, visualization should be considered early, along with other upstream tasks shown in Fig. 3, such as data acquisition and preprocessing. This requires a novel visualization technique with prior interdisciplinary domain knowledge through closely collaborating with domain experts who have posed the task to address real problems.

In order to effectively use visualization, remote sensing big data should be aggregated from diverse sources in a huge volume, and imported to a model which allows decision making in minutes rather than weeks or months. This is a big challenge for PB level or larger volume of data inputs, for instance, in applications related

[5]http://modis.gsfc.nasa.gov/about/specifications.php

[6]http://messtec.dlr.de/en/technology/dlr-remote-sensing-technology-institute/hyperspectral-systems-airborne-rosis-hyspex/?sid=3b724ae1718878a22607b4d4b92da16754914a4adcdc3

with hazard monitoring. Therefore, visualization of big remote sensing data should deal with challenges of large data visualization as well as interactive exploration of data for an improved understanding. Note that data visualization continues throughout the life cycle of big data, but has individual challenges in different phases.

IV. BIG OPPORTUNITIES

Despite the aforementioned big challenges, the potential value of big remote sensing data is impressive. Actually, remote sensing techniques have been successfully used for different applications, such as agriculture applications (e.g., food security monitoring, pasture monitoring), oceanic applications (e.g., ship detection, oil spill detection), urban planning, urban monitoring, human settlements (both urban and rural), food security monitoring, water quality monitoring, energy assessment, population of disease, ecosystem assessment, global warming, global change, global forest resources assessment, ancient site discovery (archaeology), and so on.

Combined with human activities and data from social science, remote sensing techniques listed above have become much more powerful tools to significantly improve the effectiveness of production and operation for human welfare. In this way, big remote sensing data provides the capacity to accomplish targets which were hard or impossible to achieve in traditional ways. For instance, a hidden relic site can be found by high-resolution remote sensing data in a dense forest without modern infrastructure, which is an incredible barrier for field archaeologists to penetrate. A successful application is on Maya research in the Petén region of northern Guatemala [5].

In urban planning applications, ground measurements as well as spaceborne and airborne remote sensing images are integrated to result in better and timely urban planning, management, and sustainability [12], [14], [15]. In this context, remote sensing data can be acquired over a large area in a sequence in a very high resolution (i.e., less than 1 m/pixel) using advanced remote sensing techniques. Related projects include the 100 cities project for urban environmental characterization, monitoring, and government decision making,[7] and global urban footprint using very high spatial resolution of a total of 180 000 TerraSAR-X and TanDEM-X scenes for the worldwide mapping of settlements.[8] Combined with population census data, remote sensing data were integrated to understand land development, land use, and urban sprawl in Puerto Rico [11]. Together with socioeconomic variables, high-resolution satellite images have been used to analyze urban population growth which is closely related to economic growth [13].

[7]http://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-9628/16557_read-40454/

[8]http://cesa.asu.edu/urban-systems/100-cities-project/

Since the early 1990s, remote sensing data have been used for agricultural applications with regional phenological change and associated meteorological factors [6]. In precision agriculture, the role of remote sensing data becomes more and more important in terms of sustainable agriculture, including food security [8], or assessing crop condition and yield forecasting [7]. In addition, average yield gaps are large among nations for major cereal crops, maize, wheat and rice, etc. Usually, agricultural intensification could greatly reduce these yield gaps [42]. In this case, remote sensing has proved to be of great help for monitoring crops in a large area, such as mapping the bioenergy potential of maize crops [43] by incorporating the effects of climate and soils on yields [44].

In particular, food security is a key factor of intelligent agricultural systems and only remote sensing from Earth Observing satellites (e.g., Landsat, Resourcesat, MODIS) can provide consistent, repeated, and high-quality data for characterizing and mapping key cropland parameters for global cropland estimation and food security analysis in combination with national statistics, field-plot data, and secondary data [long (50–100 year) records of precipitation and temperature, soil types, and administrative boundaries] [31], [45]. Together with demographical and health survey data, many applications can benefit from the analysis of remote sensing data [9], [10], and further the relationship between human health and environmental changes can be accurately modeled [10]. In addition, remote sensing data analysis can be used to global insurance markets, such as crop damage and flood and fire risk assessment [46].

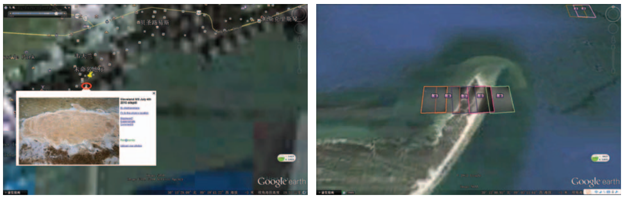
In summary, remote sensing data as well as other domain data provide great opportunities for applications in natural sciences, such as mapping tree density at a global scale [16], but also on social science, such as urban studies, demography, archeology, war and conflict studies, and so on [17].

V. CASE STUDIES

In this section, two case studies demonstrating the effectiveness of big data in remote sensing applications are described. In both cases, using a new data processing methodology and powerful computing architectures are essential. The problems addressed are automatic oil spill detection and content-based information retrieval from a large repository of multispectral and hyperspectral remote sensing data and related data from other domains, respectively.

A. Big Data for Oil Spill Detection

In traditional remote sensing classification applications, labeled samples are obtained according to ground surveys, image photointerpretation, or a combination of the aforementioned strategies [47]. *In situ* ground surveys can lead to a high accuracy of labeling but these



(a) (b)

Fig. 4. ***External data for oil spill labeling: (a) data provided by a governmental institute; and (b) images by social media.***

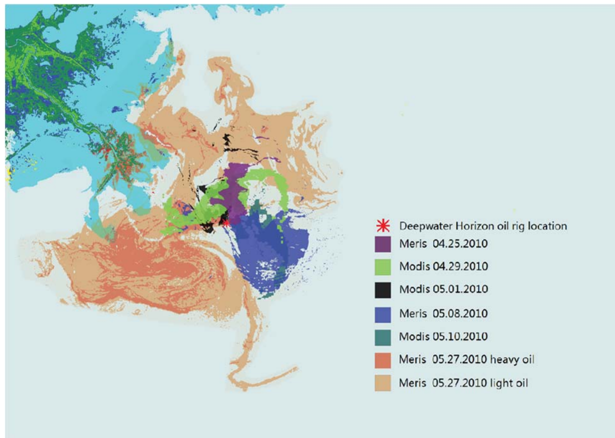


Fig. 5. ***Oil spill detection on multitemporal and multisource spaceborne remote sensing images using big data in remote sensing.***

techniques are costly and time consuming. Image photointerpretation is fast and cheap, but cannot guarantee a high labeling quality. Although hybrid solutions can take advantage of ground surveys and image photointerpretation in most remote sensing problems, it is still difficult to label marine oil spills using the hybrid solution in terms of remote sensing data provided by air/spaceborne instruments due to oil drift and diffusion. Therefore, the labeling of marine oil spills brings a great challenge to the oil spill detection task. In this case study, we first identify proper data consisting of big remote sensing data and then tackle the labeling challenge by a novel data methodology, i.e., by the integration of social media data with aid of crowdsourcing [3] and active learning techniques [48], [49].

Specifically, we selected the big oil spill event occurred in the Gulf of Mexico (USA) on 2010 as our study case as more social media data and other forms of data can be obtained for big data analysis. The optical remote sensing data used contain multitemporal and multisource images, i.e., data from the medium resolution imaging spectrometer (MERIS), operated by the European Space Agency (ESA), and the moderate resolution imaging spectrometer (MODIS), operated by NASA. Other forms of data in this context include social media data, i.e., the pictures from social media and textual description. For instance, pictures from Panoramio,[9] a geolocation-oriented photosharing site, can be easily obtained and are often geotagged, in the form of precise coordinates of the location from where these pictures have been taken, as well as textual tagging [see Fig. 4(a)]. Also, the airborne data in the polluted area can be used to label remote sensing images, such as the oil spills detected by airborne sensors from an official institute [see Fig. 4(b)]. Those different forms of big data can be used to improve the oil spill detection accuracy in this specific context.

It should be noted that it is time consuming for the labeling process to incorporate the idea of crowdsourcing and the external data cannot cover all pixels in the remote sensing images. Accordingly, it is important to intelligently select a reduced number of informative samples for labeling in order to guarantee the accuracy of the classification task. Here, the labeling process has been done through active learning in an iterative way [48], [49].

After removing data that are heavily corrupted by clouds, multispectral remote sensing images from different dates (i.e., multitemporal images) and images from different sensors (i.e., multisource images) were exploited to detect oil spills using machine learning algorithms. Here, we used popular classifiers such as the support vector machine (SVM) [50]–[52], backpropagation neural networks [53], [54], and the k-nearest neighbor classifier [28]. In our experiments, the SVM gave the best classification accuracies and was also most robust. The obtained classification map by SVM for the considered oil spill problem in the Gulf of Mexico is given in Fig. 5, which shows oil spills spreading around the deepwater oil rig location.

There are still many open problems for pattern labeling when combining remote sensing images and social media data. For instance, an efficient strategy should be developed in order to obtain most relevant external data for a specific task. In the mean time, those external data such as photos and textual information should be automatically associated with the corresponding samples.

B. Content-Based Image Retrieval From Hyperspectral Data Repositories

In this second case study, we address a specific case study of content-based image retrieval (CBIR) applied to remotely sensed hyperspectral data, which are characterized by its high dimensionality in the spectral domain [55]. The system, introduced in [56], is validated using a complex hyperspectral image database, and implemented on a Beowulf cluster at NASA's Goddard Space Flight Center. In this context, the main challenge of this case

[9] www.panoramio.com

Fig. 6. ***AVIRIS hyperspectral image collected over the World Trade Center (left) and detail of the area used as input query (right).***

study is to deal with the voluminous challenge of big remote sensing data, which in our experiment comprises a collection of 154 high-resolution hyperspectral data sets (more than 20 TB of data) gathered by NASA over the World Trade Center (WTC) area in New York City during the last two weeks of September 2001, just several days after the terrorist attacks that collapsed the two main towers and other buildings in the WTC complex. The spatial resolution of the data is 3.7 m/pixel, and the spectral resolution is 224 narrow spectral bands between 0.4 and 2.5 μm. Fig. 6 shows a false color composite of one of such images, with 614 $\times$ 512 pixels and 224 bands. The false color composition has been formed using the 1682-, 1107-, and 655-nm channels, displayed as red, green, and blue, respectively. Vegetated areas appear green in Fig. 6, while burned areas appear dark gray. Smoke coming from the WTC area appears bright blue due to high spectral reflectance in the 655-nm channel. The area used as input query in our experiment is shown in a red rectangle, and is centered at the region where the towers collapsed.

Using the search area in the rightmost part of Fig. 6 as input query, the proposed parallel CBIR system successfully retrieved all image instances containing the WTC complex across the database, with no false positive detections. For illustrative purposes, Fig. 7 shows the seven full image flightlines in the considered AVIRIS database that contain the searched area centered at the WTC complex.

To investigate the parallel properties of the proposed CBIR system, we have evaluated its performance when implemented on NASA's Thunderhead Beowulf cluster, a system composed of 256 dual 2.4-GHz Intel Xeon nodes, each with 1 GB of memory and 80 GB of main memory, interconnected with 2-GHz optical fiber Myrinet. Using 256 processors on Thunderhead, the system was able to search the most similar scenes across the full database of 154 images (with precomputed metadata) in only 4 s, resulting in a total processing of approximately 10 s to catalog and fully describe a new entry in the database. This represents a significant improvement over the implementation of the same CBIR process on a single Thunderhead processor, which took over 1 h of computation for the same operation.

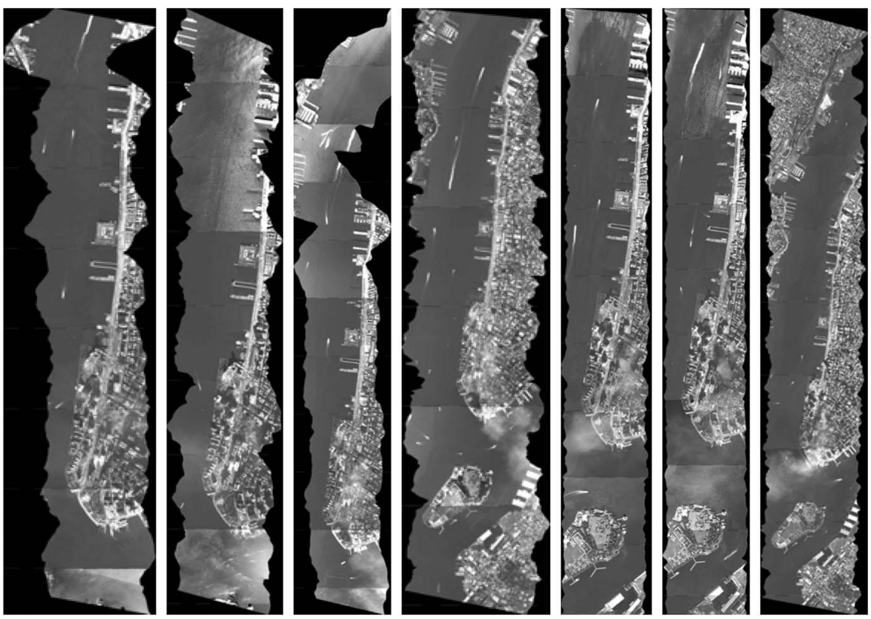


Fig. 7. ***Full flightlines collected by the AVIRIS sensor over the World Trade Center area which contain the search area in Fig. 6. Typically, each flightline contains five to seven hyperspectral images (each with 224 spectral bands).***

VI. CONCLUSION

In this paper, the connotations of big remote sensing data have been discussed. Big data in remote sensing can contain a variety of remotely sensed data from different spectral reflectance, different ground spatial resolutions, and different locations (such as optical, radar, microwave, etc.), as well as the data from other domains, such as archeology, demographics, economics (which refers to the "variety" of the three V properties of big data). As a result, the big remote sensing data have the same three V characteristics as big data in general [1] with the increasingly accumulated volume of remote sensing data from TB to PB and even to EB scale. With the voluminous data, on one hand, the tasks which are difficult to be attacked can be achieved in a reasonable time (which refers to the "velocity" of the three V and results in big opportunities); on the other hand, the big remote sensing data with any of 2Vs or 3Vs bring big challenges for those owning big data, analyzing big data and utilizing big data, respectively.

Then, a trinity framework for understanding big data in remote sensing has been proposed for those who own big data, those who can provide data methods, and those who need to exploit big data to solve real-world problems. In terms of the framework, common and individual challenges of big data have been discussed in the context of remote sensing applications. As a key common challenge, big remote sensing data should first be identified to cope with real remote sensing applications. Then, it is necessary to have the capability for highly efficient computations in order to deal with voluminous data. In addition, novel data methods or even completely new data methodologies should be developed to attack the complexities of big remote sensing data. Of course, there exist other common challenges when dealing with big remote sensing data, such as how to manipulate data quality from different perspectives by individual data providers and data recipients.

Except for the common challenges of big data in remote sensing, individual challenges should be taken into account from the three perspectives of the trinity of big data. For those who own big remote sensing data, three critical factors should be carefully designed, i.e., data transmission from air/satellite-borne sensing system to a ground station, then data storage to a system, and data delivery to users of interest. For those who exploit big data to remote sensing applications, the key challenges are the identification of the right data to achieve the given task, the deployment of the big data for later data processing and analysis, and the interpretation of the results provided by data methods. For those who are capable of developing novel data methods and/or data methodologies for remote sensing applications, data representation should first be managed due to the diversity of multisource and multitemporal remote sensing data and the data from other domains. Then, the data described in different attribute formats should be integrated to better analyze and process the big data. After that, the results provided by data analysis techniques need to be well visualized for improved data analysis and data interpretation.

Although the potential of big remote sensing data has already been anticipated, it is important to note that the data often come from heterogeneous sources and require significant computational efforts in terms of interpretation. Therefore, the big opportunity is to integrate remote sensing data together with other external data to transfer these potentials to reality. In this context, we can benefit from large-scale, consistent, repeated, and high-quality big remote sensing data in order to address applications related to monitoring food security, urbanization progress, population density, etc. This can be further used to address other relevant applications related with human health, environmental changes, or human activities in general.

In order to benefit from remote sensing in big data, proper data from different sources should be first identified to solve a specific application. Except for multitemporal, multiresolution, multiradiometric remote sensing data, how to identify related complementary out-of-domain data and how to obtain those data sets represent the biggest challenge for big remote sensing data application. Then, a novel data methodology should be carefully designed for data processing, data fusion, and so on. Although there are many applications combining remote sensing data and data coming from other domains, most of the works available are based on a sampling technique for estimation, even for the recent work to globally estimate tree population estimation based on 429775 ground-sourced measurements of tree density from every continent on Earth [16]. How can we use all the data available deserves further study for big remote sensing data task. Last but not the least, how to evaluate the performance and how to guarantee data quality are other interesting research lines to be further explored. ■

REFERENCES

[1] D. Laney, "3D data management: Controlling data volume, velocity and variety," Gartner, Tech. Rep., Feb. 2001. [Online]. Available: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

[2] H. Ramapriyan, J. Brennan, J. Walter, and J. Behnke, "Managing big data: NASA tackles complex data challenges," *Earth Imaging J.*, 2013. [Online]. Available: http://eijournal.com/print/articles/managing-big-data.

[3] M. Chi, J. Shen, Z. Sun, F. Chen, and J. A. Benediktsson, "Oil spill detection based on web crawling images," in *Proc. IEEE Int. Geosci. Remote Sensing Symp.* Québec, QC, Canada, Jul. 2014, pp. 1–4.

[4] J. Shao, D. Xu, C. Feng, and M. Chi, "Big data challenges in China Centre for Resources Satellite Data and Application," in *Proc. Hyperspectral Image Signal Process.:*

Evol. Remote Sens. (WHISPERS), Tokyo, Japan, Jun. 2015, pp. 1–41. [Online]. Available: http://www.ieee-whispers.com/images/techprograms/whispers_2015_technical_program.pdf.

[5] T. L. Sever and D. E. Irwin, "Landscape archaeology: Remote-sensing investigation of the ancient Maya in the Peten rainforest of northern Guatemala," *Ancient Mesoamerica*, vol. 14, no. 1, pp. 113–122, Jul. 2003.

[6] M. D. Steven and J. A. Clark, Eds., *Applications of Remote Sensing in Agriculture*. London, U.K.: Butterworth-Heinemann, Jan. 1991.

[7] S. Liaghat and S. Balasundram, "A review: The role of remote sensing in precision agriculture," *Amer. J. Agricultural Biol. Sci.*, vol. 5, no. 1, pp. 50–55, 2010.

[8] I. Becker-Reshef *et al.*, "NASA contribution to the group on Earth observation (GEO) global agricultural monitoring system of systems," *Earth Observations*, vol. 21, pp. 24–29, 2009.

[9] "Contributions of land remote sensing for decisions about food security and human health," Committee on the Earth System Science for Decisions About Human Welfare: Contributions of Remote Sensing, Geographical Sciences Committee, National Research Council, Tech. Rep., 2007.

[10] M. E. Brown, K. Grace, G. Shively, K. B. Johnson, and M. Carroll, "Using satellite remote sensing and household survey data to assess human health and nutrition response to environmental change," *Population Environ.*, vol. 36, no. 1, pp. 48–72, Sep. 2014.

[11] S. Martinuzzia, W. A. Goulda, and O. M. R. Gonzaleza, "Land development, land use, and urban sprawl in puerto rico integrating remote sensing and population census data," *Landscape Urban Planning*, vol. 79, pp. 288–297, 2007.

[12] M. Netzband, W. Stefanov, and C. Redman, Eds., *Applied Remote Sensing for Urban Planning, Governance and Sustainability*. Berlin, Germany: Springer-Verlag, 2007.

[13] X. Deng, J. Huang, S. Rozelle, and E. Uchida, "Growth, population and industrialization, and urban land expansion of China," *J. Urban Econ.*, vol. 63, pp. 96–115, 2008.

[14] P. Gamba and M. Herold, Eds., *Global Mapping of Human Settlement: Experiences, Datasets, and Prospects*. Boca Raton, FL, USA: CRC Press, 2009.

[15] B. Bhatta, Ed., *Analysis of Urban Growth and Sprawl from Remote Sensing Data*. Berlin, Germany: Springer-Verlag, 2010.

[16] T. W. Crowther *et al.*, "Mapping tree density at a global scale," *Nature*, vol. 525, pp. 201–205, Sep. 2015.

[17] O. Hall, "Remote sensing in social science research," *Open Remote Sensing J.*, vol. 3, pp. 1–16, 2010.

[18] P. M. Mather and M. Koch, *Computer Processing of Remotely-Sensed Images: An Introduction*, 4th ed. New York, NY, USA: Wiley, Jan. 2011.

[19] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *Nat. Sci. Rev.*, vol. 1, no. 2, pp. 293–314, Jun. 2014.

[20] Y. Ma *et al.*, "Remote sensing big data computing: Challenges and opportunities," *Future Generat. Comput. Syst.*, vol. 51, pp. 47–60, 2015.

[21] Gartner, "Gartner says solving 'big data' challenge involves more than just managing volumes of data," Tech. Rep., Jun. 2011.

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

[23] A. Plaza and C.-I. Chang, *High Performance Computing in Remote Sensing*. Boca Raton, FL. USA: CRC Press, 2007.

[24] T. Hey, S. Tansley, and K. Tolle, Eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA, USA: Microsoft Research, 2009.

[25] M. M. U. Rathore *et al.*, "Real-time big data analytical architecture for remote sensing application," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 8, no. 10, pp. 4610–4621, 2015.

[26] M. A. Wulder and N. C. Coops, "Satellites: Make earth observations open access," *Nature*, vol. 513, pp. 30–31, Sep. 2014.

[27] M. Kovac, "E-health demystified: An e-government showcase," *Computer*, vol. 47, no. 10, pp.34–42, 2014.

[28] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York, NY, USA: Springer-Verlag, 2006.

[29] B. Settles, "Active Learning Literature Survey," Univ. Wisconsin Madison, Madison, WI, USA, Comput. Sci. Tech. Rep. 1648, Jan. 2010.

[30] J. Howe, "The rise of crowdsourcing," *Wired Mag.*, vol. 14, no. 6, 2006. [Online]. Available: http://www.wired.com/wired/archive/14.06/crowds.html

[31] P. Thenkabail, G. Lyon, H. Turral, and C. Biradar, Eds., *Remote Sensing of Global Croplands for Food Security*. Boca Raton, FL, USA/London, U.K.: CRC Press/Taylor & Francis, Jun. 2009.

[32] J. Hurwitz, A. Nugent, F. Halper, and M. Kaufman, *Big Data For Dummies*, 1st ed. New York, NY, USA: Wiley, 2013.

[33] L. Bruzzone, D. Prieto, and S. Serpico, "A neural-statistical approach to multitemporal andmultisource remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1350–1359, 1999.

[34] J. Bao, M. Chi, and J. Benediktsson, "Spectral derivative features for classification of hyperspectral remote sensing images: Experimental evaluation," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 6, no. 2, pp. 594–601, 2013.

[35] V. N. Gudivada and V. V. Raghavan, "Content based image retrieval systems," *Computer*, vol. 28, no. 9, pp. 18–22, 1995.

[36] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1106–1114, 2012.

[38] M. Chi, J. Bao, Y. Zou, and J. A. Benediktsson, "Deep neural networks for remote sensing image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Melbourne, VIC, Australia, Jul. 2013, pp. 1–172. [Online]. Available: http://www.igarss2013.org/IG13_FinalProgram_Web.pdf.

[39] A. Merentitis and C. Debes, "Automatic fusion and classification using random forests and features extracted with deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Milan, Italy, Jul. 2015, pp. 2943–2946.

[40] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Top. Appl. Earth Observat. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, 2015.

[41] E. Waltz and T. Waltz, *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed. Boca Raton, FL, USA: CRC Press, Nov. 2008, ch. 5, pp. 89–114.

[42] N. D. Mueller, *et al.*et al., "Closing yield gaps through nutrient and water management," *Nature*, vol. 490, pp. 254–257, 2012.

[43] T. Udelhoven *et al.*, "Retrieving the bioenergy potential from maize crops using hyperspectral remote sensing," *Remote Sens.*, vol. 5, no. 1, pp. 254–273, 2013.

[44] C. Atzberger, "Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs," *Remote Sens.*, vol. 5, pp. 949–981, 2013.

[45] P. Teluguntla *et al.*, "Global Cropland Area Database (GCAD) derived from remote sensing in support of food security in the Twenty-First Century: Current achievements and future possibilities," *Land Resources: Monitoring, Modelling, and Mapping, Remote Sensing Handbook*. Boca Raton, FL, USA: CRC Press, Jun. 2015, vol. II.

[46] J. de Leeuw *et al.*, "The potential and uptake of remote sensing in insurance: A review," *Remote Sens.*, vol. 6, no. 11, pp. 10 888–10 912, 2014.

[47] C. Chen, *Signal and Image Processing for Remote Sensing*, 2nd ed. Boca Raton, FL, USA: CRC Press, Feb. 2012.

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

[49] Z. Sun and M. Chi, "Superpixel-based active learning for the classification of hyperspectral images," in *Proc. Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*," Tokyo, Japan, Jun. 2015, pp. 1–41. [Online]. Available: http://www.ieee-whispers.com/images/techprograms/whispers_2015_technical_program.pdf.

[50] B. Waske, M. Chi, J. A. Benediktsson, S. van der Linden, and B. Koetz, "Algorithms and applications for land cover classification—A review," in *Geospatial Technology for Earth Observation*, D. Li, J. Shan, and J. Gong, Eds. New York, NY, USA: Springer-Verlag , 2009, pp. 203–233.

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

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

[53] D. M. Miller, E. J. Kaminsky, and S. Rana, "Neural network classification of remote-sensing data," *Comput. Geosci.*, vol. 21, no. 3, pp. 377–386, Apr. 1995.

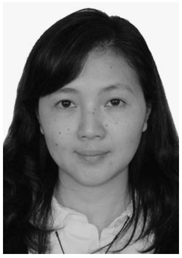
[54] P. M. Atkinson and A. R. L. Tatnall, "Introduction neural networks in remote

sensing," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 699–709, 1997.

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

[56] A. Plaza, J. Plaza, and A. Paz, "Parallel heterogeneous CBIR system for efficient hyperspectral image retrieval using spectral mixture analysis," *Concurrency Comput., Practice Exp.*, vol. 22, no. 9, pp. 1138–1159, 2010.

ABOUT THE AUTHORS



Mingmin Chi (Member, IEEE) received the B.S. degree in electrical engineering from Changchun University of Science and Technology, Changchun, China, in 1998, the M.S. degree in electrical engineering from Xiamen University, Xiamen, China, in 2002, and the Ph.D. degree in computer science from University of Trento, Trento, Italy, in 2006. She was also a student visitor from May 2005 to March 2006 at the Department of Schoelkopf, Max-Planck Institute for Biological Cybernetics, Tuebingen, Germany.

Currently, she is an Associate Professor at the School of Computer Science, Fudan University, Shanghai Key Laboratory of Data Science, Shanghai, China. Her research interests include data science, big data, and machine learning with applications to remote sensing and social media.

Prof. Chi has been an Associate Editor for the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing since 2013.

Antonio Plaza (Fellow, IEEE) received the Computer Engineering degree and the M.Sc. and Ph.D. degrees in computer engineering from the University of Extremadura, Badajoz, Spain, in 1998, 2000, and 2002, respectively.

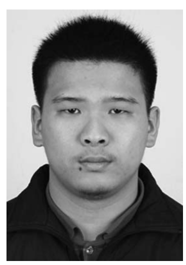
He is an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, where he is the Head of the Hyperspectral Computing Laboratory (HyperComp). He was the Coordinator of the Hyperspectral Imaging Network, a European project with total funding of 2.8 million € (2007–2011). He has authored more than 370 publications, including more than 100 JCR journal papers (60 in IEEE journals), 20 book chapters, and over 230 peer-reviewed conference proceeding papers (90 in IEEE conferences). He has guest edited seven special issues in JCR journals (three in IEEE journals).

Prof. Plaza was a Chair for the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing in 2011. He is an Associate Editor for the IEEE Geoscience and Remote Sensing Magazine, and was a member of the Editorial Board of the IEEE Geoscience and Remote Sensing Newsletter from 2011 to 2012, and a member of the Steering Committee of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing in 2012. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012, and has been the President of the Spanish Chapter of the IEEE GRSS since November 2012. He has been the Editor-in-Chief of the IEEE Transactions on Geoscience and Remote Sensing since January 2013. He was a recipient of the recognition of Best Reviewers of the IEEE Geoscience and Remote Sensing Letters in 2009 and a recipient of the recognition of Best Reviewers of the IEEE Transactions on Geoscience and Remote Sensing in 2010, a journal for which he has served as an Associate Editor from 2007 to 2012.

Jón Atli Benediktsson (Fellow, IEEE) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

On July 1, 2015, he became the Rector of the University of Iceland. From 2009 to 2015, he was the Pro Rector of Science and Academic Affairs and Professor of Electrical and Computer Engineering at the University of Iceland. He is a cofounder of the biomedical startup company Oxymap (www.oxymap.com). His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing. He has published extensively in those fields.

Prof. Benediktsson was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS AdCom since 2000. He was the Editor-in-Chief of the IEEE Transactions on Geoscience and Remote Sensing (TGRS) from 2003 to 2008 and has served as an Associate Editor of TGRS since 1999, the IEEE Geoscience and Remote Sensing Letters since 2003, and IEEE Access since 2013. He is on the Editorial Board of the Proceedings of the IEEE, the International Editorial Board of the *International Journal of Image and Data Fusion* and was the Chairman of the Steering Committee of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS) in 2007–2010. He is a Fellow of the International Society for Optics and Photonics (SPIE). He is a member of the 2014 IEEE Fellow Committee. He received the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award. In 2000, he was granted the IEEE Third Millennium Medal, and in 2004, he was a corecipient of the University of Iceland's Technology Innovation Award. In 2006, he received the yearly research award from the Engineering Research Institute of the University of Iceland, and in 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society. He was a corecipient of the 2012 IEEE Transactions on Geoscience and Remote Sensing Paper Award, and in 2013, he was a corecipient of the IEEE GRSS Highest Impact Paper Award. In 2013, he received the IEEE/VFI Electrical Engineer of the Year Award. In 2014, he was a corecipient of the *International Journal of Image and Data Fusion* Best Paper Award. He is a member of the Association of Chartered Engineers in Iceland (VFI), Societas Scinetiarum Islandica, and Tau Beta Pi.



Zhongyi Sun received the Diploma in computer science from Fudan University, Shanghai, China, in 2014, where he is currently a graduate student at the School of Computer Science.

His research focuses on remote sensing, machine learning, and social media.

Jinsheng Shen received the B.S. degree in computer science from Fudan University, Shanghai, China, in 2014, where he is currently a graduate student at the School of Computer Science.

His research focuses on remote sensing and big data.

Yangyong Zhureceived the Ph.D. degree in computer and software theory from Fudan University, Shanghai, China, in 1994.

He is a Professor of Computer Science in Fudan University. His research interests include database, data mining, and their applications in finance, economics, insurance, bioinformatics, and sociology. He has published more than 100 papers. His research has been supported in part by the National High-Tech Research and Development Plan (863) of China, the National Natural Science Foundation of China (NSFC), and the Development Fund of Shanghai Science and Technology Commission. He is the founder and the director of the Shanghai Key Laboratory of Data Science, Fudan University.