INVITED PAPER

Distributed Fusion of Heterogeneous Remote Sensing and Social Media Data: A Review and New Developments

Distributed computing strategies in remote sensing techniques and applications that use various data sources are comprehensively reviewed. A new distributed fusion framework that can accelerate the fusion of heterogeneous remote sensing and social media data is proposed by decomposing large data sets into small ones and processing them in parallel.

By JUN LI, *Fellow IEEE*, ZHENJIE LIU, *Student Member IEEE*, XINYA LEI, AND LIZHE WANG, *Fellow IEEE*

ABSTRACT | Despite the wide availability of remote sensing big data from numerous different Earth Observation (EO) instruments, the limitations in the spatial and temporal resolution of such EO sensors (as well as atmospheric opacity and other kinds of interferers) have led to many situations in which using only remote sensing data cannot fully meet the requirements of applications in which a (near) real-time response is needed. Examples of these applications include floods, earthquakes, and other kinds of natural disasters, such as typhoons. To address this issue, social media data have gradually been adopted to fill possible gaps in the analysis when remote sensing data are lacking or incomplete. In this case, the fusion of heterogeneous big data streams from multiple data sources introduces significant demands from a computational viewpoint. In order to meet these challenges, distributed computing is increasingly viewed as a feasible solution to parallelize the analysis of massive data coming from different sources (e.g., remote sensing and social media data). In this article, we provide an overview of available and new distributed strategies to address the computational challenges brought by massive heterogeneous data processing and fusion for real-time environmental monitoring and decision-making. The 2013 Boulder (Colorado) flood event is taken as a case study to evaluate several new distributed data fusion frameworks. Experimental results demonstrate that the proposed distributed frameworks are suitable in terms of response time and computational requirements for fusing large-volume heterogeneous data sources.

KEYWORDS | Data fusion; parallel and distributed computing; remote sensing; social media.

I. INTRODUCTION

Due to the limitations in the spatial and temporal resolutions of Earth Observation (EO) instruments, there is a need to integrate remote sensing data with other sources for EO applications, especially in the area of emergency response [1]. With the rapid development and availability of social media data, it has gradually become a complement to fill possible gaps in remotely sensed data [2]. However, methods for fusing heterogeneous remote sensing and social media data (e.g., domain

Manuscript received January 22, 2021; revised April 6, 2021 and May 7, 2021; accepted May 8, 2021. Date of publication May 21, 2021; date of current version July 19, 2021. This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA19090104, in part by the National Natural Science Foundation of China under Grant 42030111 and Grant 61771496, and in part by the National Key Research and Development Program of China under Grant 2017YFB0502900. *(Corresponding author: Lizhe Wang.)*

Jun Li and **Zhenjie Liu** are with the Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China.

Xinya Lei and **Lizhe Wang** are with the School of Computer Science, China University of Geosciences (Wuhan), Wuhan 430074, China (e-mail: lizhe.wang@gmail.com).

Digital Object Identifier 10.1109/JPROC.2021.3079176

adaptation techniques [3]) are computationally expensive, leading to the fact that the associated time constraints limit their applicability (especially, in the context of emergency response applications [4], [5]). Distributed computing, which has already been adopted in many remote sensing big data applications [6]–[8], is increasingly viewed as a feasible strategy to address the computational challenges brought by massive multisource data processing and fusion techniques [9]–[12].

In this work, we specifically focus on distributed computing techniques for the fusion of remote sensing data and social media data. The motivation of our work lies in two main aspects. On the one hand, this article gives a comprehensive overview of massive heterogeneous data fusion and its distributed computing implementations. On the other hand, we explore the feasibility of developing distributed parallel methods to address the computationally challenging problem of fusing heterogeneous data sources based on domain adaptation, providing an example of the distributed fusion of remote sensing and social media data during floods. The most relevant research objectives and contributions of our work can be summarized as follows.

1) We provide an overview of techniques for parallel and distributed fusion of heterogeneous big data from multiple data sources. Although there are some reviews on parallel and distributed computing for remote sensing data processing [7], [13], a specific review focused on the distributed fusion of remote sensing data and other sources data has not been published before in the literature and represents a completely novel contribution.
2) Based on our review of parallel and distributed techniques for fusing heterogeneous big data in remote sensing, new distributed fusion frameworks are proposed. The proposed frameworks can accelerate the fusion of heterogeneous remote sensing and social media data by decomposing large datasets into small ones and processing them simultaneously.
3) Last but not least, we present a case study of the 2013 floods in Boulder, Colorado, to explore the performance and advantages of the domain adaptation-based distributed fusion frameworks in a real scenario.

The remainder of this article is organized as follows. We first review existing techniques and applications to fuse remote sensing data with other data sources in Section II. Section III provides an overview of parallel and distributed fusion of remote sensing and other data sources. Section IV presents a case study to demonstrate the practical advantages of the fusion of heterogeneous remote sensing and social media data. Section V concludes this article with some remarks.

II. DATA FUSION BASED ON REMOTE SENSING AND OTHER SOURCES DATA

In this section, we provide an overview of available techniques and applications for fusing remote sensing and other sources of data. First, we describe available research using multisource remote sensing data. Then, we review some works that combine remote sensing and other kinds of auxiliary data. Finally, we focus on specific techniques and applications based on the fusion of remote sensing and social media data and also summarize the computational limitations inherent to the fusion of big heterogeneous data streams.

A. Multisource Remote Sensing Data

There have been many techniques based on multisource remote sensing data fusion for EO applications. This is mainly due to the wide availability of a diversity of remote sensing datasets from numerous sensors (multimodality). In the following, we structure the most relevant works according to their application to different areas of EO.

1) Land-Use and Land-Cover Classification: The fusion of multisource remote sensing data has been extensively used to improve land-use and land-cover (LULC) classification. In [14], a new framework based on a Bayesian formulation to fuse Landsat TM images and ERS-1 SAR images is proposed for land-use classification. Compared to conventional classifiers (based on a single remote sensing data source), the proposed framework could significantly improve the classification accuracy. The study in [15] integrated multisource remote sensing data into a homogeneous time series of land-cover maps by equalizing the levels of thematic content and spatial details. In [16], an optimized data mining classification approach was used to fuse SPOT-6, RADARSAT-2, and derived datasets for land-cover classification. Compared with two nonparametric classifiers [e.g., support vector machine (SVM) and random forest (RF)], the proposed approach produced classification results with higher accuracy. The research presented in [17] first fused RADARSAT-2 images with Thaichote (THEOS) and Landsat 8 OLI images through a principal component analysis (PCA) technique. Then, a new technique combining genetic algorithms (GAs) and SVM was proposed to improve land-use classification, which outperformed the traditional grid search approach. The work in [18] explored knowledge transfer between multiple remote sensing images for land-use classification. The focus was on the statistical alignment of the image of interest and another image with available ground-truth information. In [19], an orthogonal total variation component analysis (OTVCA)-based feature fusion method was adopted to fuse spectral, spatial, and elevation information extracted from hyperspectral images and light detection and ranging (LiDAR) measurements. The results showed that OTVCA-fusion could produce more accurate classification maps while using fewer features compared with RF and SVM classifiers. Similar studies for the fusion of hyperspectral images and LiDAR data were also given in [20]–[22].

2) Urban Monitoring: Urban environmental changes can be monitored and evaluated more accurately through multisource remote sensing data, which is of great significance to the sustainability of cities [23]. The work in [24] developed a comprehensive evaluation index to assess the urban environmental change in China from 2000 to 2012, considering three remote sensing indicators (e.g., PM2.5 concentration, land surface temperature, and vegetation cover) based on multisource remote sensing data. Another similar study exploring the interannual variations and trends in the urban environment of 17 megacities in Eurasia was the one in [25]. The research in [26] combined vegetation information from MODIS-NDVI data with urban socioeconomic information from the DMSP-OLS data to generate annual human settlement index (HSI) images. A linear least-squares model was fit between the mean MODIS-LST and HSIs to analyze the spatial and temporal changes of urban heat islands (UHIs). In [27], initial impervious surface area (ISA) data were generated based on the integration of the Visible Infrared Imaging Radiometer Suite Day/Night Band (VIIRS DNB) and Moderate Resolution Imaging Spectroradiometer (MODIS) data through a thresholding approach. Then, the normalized difference vegetation index (NDVI) and the normalized difference water index (NDWI) derived from Landsat images were used to remove vegetation and water areas from the initial ISA data, in order to construct the ISA mapping at 30-m spatial resolution in China.

3) Vegetation Mapping: Vegetation mapping is an EO area where the fusion of multisource remote sensing data has been widely applied. The work in [28] investigated different numerical combinations of integration techniques to fuse ERS-1 SAR geocoded images with Landsat TM data for vegetation cover monitoring and assessment. In [29], Landsat TM images were combined with ERS-1 SAR images through geometrical coregistration and resampling. Multilayer perceptron neural networks were used to identify eight different types of forests based on the integrated dataset. The research in [30] proposed a novel method for aboveground biomass (AGB) estimation of forest based on the structural analysis of mixed pixels and the RF model. Specifically, a correction factor estimated from MODIS data was used to create a model that scales from fine-resolution data (SPOT 5) to coarse-resolution data (MODIS). In [31], a wide range of vegetation and textural indices extracted from optical (Landsat and MODIS) and radar (ALOS-1 PALSAR and Sentinel-1) remote sensing data were imported into an RF regression model for AGB mapping of forests.

4) Disaster Monitoring: The integration of multisource remote sensing data can effectively improve disaster monitoring and assessment [32]. The work in [33] detected the spatial distribution of surface rupture zones after earthquakes by integrating multisource remote sensing images collected using different sensors (Landsat, SPOT, and ASTER). In [34], change detection and pansharpening were employed to monitor landslides at a regional scale using multitemporal SPOT images and IKONOS images. Another work focused on landslide analysis was presented in [35], in which geographic information system (GIS)-based statistical models [including analytical hierarchy process (AHP), weighted linear combination (WLC), and spatial multicriteria evaluation (SMCE) models] were used to identify the relationship between landslide locations and landslide-related factors extracted from SAR data and also from SPOT 5 and WorldView-1 images. Due to bad weather during a flood event, SAR images have been widely used for flood monitoring. For instance, in [36], a ranking matrix in 3-D multiplication mode was adopted to assess flood hazards using SAR and optical remote sensing images. In this method, flood-affected frequency and flood depth (obtained from the SAR images) were used as hydrologic parameters. Elevation heights, land cover classification, geomorphic division, and drainage network data obtained from optical remote sensing instruments were used through a GIS-based approach. The work presented in [37] monitored grassland snow disasters based on the combination of MODIS data and passive microwave AMSR-E data. Specifically, the snow map obtained by AMSR-E was resampled to the same spatial resolution as the snow map generated by MODIS in order to synthesize the missing pixel values of the MODIS area occluded by clouds.

5) Other Applications: There has also been a significant amount of research based on multisource remote sensing data for other applications. In [38], phenological features and backscattering features generated from Sentinel-1 images were combined with spectral features extracted from Sentinel-2 and Landsat 8 OLI images to construct multisource feature sets for crop mapping. The work presented in [39] described that freely available multisource data (e.g., Landsat TM, ETM+ imagery, JERS-1 active radar L-band imagery, and elevation data) contributed to a tree bagging classification procedure for wetland mapping. The research in [40] applied a genetic programming approach to fuse MODIS and Landsat TM/ETM+ data for water quality assessment. In [41], the SVM regression algorithm and artificial neural networks (ANNs) were employed to estimate soil salinity based on the model parameters extracted from multisource remote sensors (e.g., Sentinel-1A, Landsat 8 OLI, and MODIS). In [42], daily snow data from MODIS and snow classification results based on Landsat MSS and TM/ETM+ images were used along with supervised classification methods to monitor the changes in snow and glacier coverage. The work in [43] fused high-resolution SAR and optical images for the reconstruction of urban topography, using a hypothesis based on the greatest SAR-optical similarity. The work in [44] proved that multi-image fusion was able to improve cloud detection performance. In [45], a novel pansharpening method named SparseFI was developed by exploring the sparse representation of multispectral image patches with high and low spatial resolutions. Compared with conventional pansharpening methods, the proposed method was more robust and exhibited fewer spectral distortions. The work presented in [46] proposed a semisupervised manifold alignment method to align multisource remote sensing images, which was suitable for images with different spectral or spatial resolutions.

As shown in the aforementioned works, the fusion of multisource remote sensing data can provide important complementarities in terms of spectral, spatial, and temporal resolutions. In addition, using additional EO data sources often generates an increase in data volume, speed, and variety. The combination of remote sensing and other sources of data provides unprecedented opportunities for advanced EO and monitoring.

B. Remote Sensing and Auxiliary Data

There has also been a large amount of research based on the combination of remote sensing and auxiliary data. Auxiliary data mainly include EO data collected from stations, sensors, model simulations, and so on. Generally, auxiliary data have been used as input to improve the performance of remote sensing strategies. In the following, we summarize the most important works according to their different EO applications.

1) Ecological Monitoring: The combination of remote sensing and auxiliary data enables large-scale ecosystem monitoring while maintaining accuracy at the local scale [47]. In [48], the regression tree model was used to create ecosystem component predictions combining extensive ground measurements, remote sensing images from high-resolution satellites, and Landsat 8. In [49], multiple linear regression (MLR) analyses were used to estimate natural grassland biomass based on field and remote sensing data. The study presented in [50] generated a continuous 33-year 1-km net primary productivity (NPP) time series with the Carnegie–Ames–Stanford Approach (CASA) model by fusing multisource remote sensing data and station data. The research presented in [47] compared surface and regression methods for upscaling field-sampled aboveground carbon data using fine spatial resolution remote sensing data as auxiliary data. The results showed that the optimized integration of field data and multiscale remote sensing data could obtain a detailed mapping of aboveground carbon in heterogeneous landscapes. The work in [51] combined a large number of ground observations, MODIS data, and forest coverage/acquisition/loss to develop a novel AGB mapping method with a spatial resolution of 1 km.

2) Air Monitoring: Combining remote sensing and data from other sources has also been used to monitor air quality more quickly and accurately. The study in [52] developed a country-scale geographically weighted regression (GWR) model, which used fused satellite AOD as the main predictor to estimate China's daily PM2.5 concentration. The results show that the performance of the proposed model could be greatly improved by combining meteorological and land-use data. The research in [53] mapped hyperlocal air temperatures by integrating Sentinel, Landsat, and LiDAR data with crowd-sourced air temperatures data from private weather stations through RF regression models. In [54], a five-layer structured deep belief network (DBN) was employed to capture the complex and nonlinear relationship between remote sensing data and ground observation data for air temperature mapping. The work in [55] studied the concentration, physical properties, and chemical composition of aerosols by exploring the synergies between field measurement data and remote sensing images.

3) Disaster Monitoring: Various auxiliary sources of data have been gradually combined with multisatellite sensor data and applied to monitor different natural disasters. The work in [56] adopted a GIS-based arithmetic overlay approach to integrate multisource data (e.g., remote sensing images, topographic maps, and field data) for landslide susceptibility mapping. The results showed that the generated sensitivity maps exhibited close agreement with existing field instability conditions. In a related fashion, a landslide location map was generated by integrating satellite data, aerial photographs, and field observations, such as static and dynamic factors of the universal soil loss equation (USLE) index model. The study in [57] explored the complementary nature of remote sensing and ground observatory data through temporal analysis and revealed the changes in surface, air, atmosphere, and meteorological parameters after the Wenchuan Earthquake in 2018. In [58], a supervised classification technique was used to estimate the flooded extent based on RADARSAT remote sensing data, GIS data, and ground data. In [59], meteorological data and remote sensing data were used to derive drought prediction models, such as autoregressive integrated moving average (ARIMA), RF, and SVM. The results showed that integrated multisource data can help monitor and predict droughts with great accuracy.

4) Other Applications: There are other applications based on the integration of remote sensing and auxiliary data sources. The paper presented in [60] proposed a fuzzy approach to explore the inherent ambiguity of remote sensing data and ground data for the classification of suburban land cover. In [61], the layers of the total arable land and sown area of 17 major crops from the county-level agricultural census are overlapped with the land cover map derived from Landsat TM images, thus generating a distribution map of rice agriculture (with a resolution of 0.5°) in mainland China. In [62], a nonlinear ANN was developed for the classification of understorey vegetation (e.g., bamboo). In this method, a limited set of ground data were used for training, and widely available Landsat TM data were used as the input. The work in [63] analyzed the regression Kriging method that combines SPOT images and ground measurement data for soil salinity mapping. Compared with the purely regressive approaches, the regression Kriging method could monitor the soil salinity in arid areas more accurately using multisources data.

From the aforementioned works, we can conclude that the Earth can now be observed and modeled with unprecedented spectral ranges and spatiotemporal scales due to the wide availability of large EO datasets from numerous sensors and auxiliary data sources. However, in real applications, multisource remote sensing data and general *in situ* observations might be unavailable at the most urgent times and locations.

C. Remote Sensing and Social Media Data

With the rapid development and availability of social media data, its fusion with remote sensing data has recently attracted widespread attention. Meanwhile, publicly available social media data can complement and fill the gaps in remote sensing data. There has been a significant increase in research based on remote sensing and social media data. In the following, we describe the most relevant techniques categorized in terms of their EO applications.

1) Disaster Monitoring: The fusion of remote sensing and social media data can provide fast and effective monitoring of natural disasters, which is of great significance for disaster prevention and mitigation. The study in [64] imported geotagged photographs from social media, optical remote sensing, and high-resolution terrain data into a developed Bayesian statistical model in order to estimate the probability of flood inundation. The work in [65] first applied a kernel density smoothing operation to each layer of nonauthoritative data, including Twitter data, geolocated photographs, and online news. Then, a weighted sum overlay approach was used to integrate these layers with an RGB composite photograph, SAR imagery, digital elevation models (DEMs), and water height data to provide an estimation of flood extent. A similar study (based on general kernel density estimation) was the one in [66], in which a flood disaster map generation method using remote sensing and volunteered geographic data (e.g., Google News, videos, and photographs) was proposed. The results showed that even a small amount of voluntary ground data could significantly improve flood mapping precision. In related fashion, in [67], a novel method was developed to fuse NDWI extracted from remote sensing data with real-time volunteered geographic information (VGI) based on a Gaussian kernel function.

Detecting reliable information from massive social media data for emergency response is a very challenging task. In this context, text mining and natural language processing (NLP) techniques have been adopted. The study given in [68] grouped the detection methods of crisis-related Twitter messages into three categories, based on characteristics, crowdsourcing, and machine learning techniques, respectively. In [69], a deep learning method, e.g., convolutional neural network (CNN), was employed to filter disaster-related tweets. Based on filtering tweets, the work in [70] reconstructed the spatial and temporal distribution characteristics of natural disasters by combining clustering methods. The research in [71] presented a scalable system to enrich satellite imagery information in emergency situations (such as wildfires, earthquakes, or floods) by crawling and analyzing multimedia contextual content from social media. In [72], a system was proposed, which could automatically collect social media data related to natural disasters and automatically link it to remote sensing data through the text analysis in local languages. In order to improve the existing keyword-based NLP methods or machine learning algorithm methods that only rely on text, an enhanced text mining framework is proposed in [73]. The proposed framework combined location information from social media and remote sensing datasets to detect disaster-related social media posts. In addition to text mining, spatial mining has also been used. For instance, the work in [1] presented a cloud-based framework that integrates multisource data (e.g., social media, remote sensing, Wikipedia, and the Web) and analyzes the spatiotemporal patterns of each disaster type through spatial data mining and text mining techniques. The proposed framework could support rapid disaster analysis of historical and future events and provide flexibility in terms of computing and storage.

The fusion of heterogeneous remote sensing and social media data generally suffers from the fact that the probability density functions may differ in different data sources [74]. Therefore, some research works explore domain adaptation techniques to align the representations of heterogeneous remote sensing and social media data. For instance, the work in [3] proposed an optimal transport (OT) strategy for both semisupervised and unsupervised cases, which could perform the alignment of the representations in the source and target domains. In [75], a robust theoretical framework was proposed to solve the inherent limitations of sensors and realize flood density estimation in near real time by using the OT algorithm. The framework was able to develop feasible flood maps in areas with the great impact of environmental hazards, such as hurricanes or severe weather. A more specific application-oriented study was given in [76], in which a new model called geographic OT (GOT) was developed. Compared with OT, the GOT model is able to simultaneously align representations and geolocations by considering two new forms of remote sensing features in flood events.

2) Disaster Assessment: Timely disaster assessment is essential for the coordination of disaster relief operations, the prioritizing of resource allocation, and the establishment of evacuation and supply routes [77]. The work in [78] overlaid three layers, including flooding areas extracted from Landsat 8 OLI images, a flooding and damage surface generated from tweets and geolocated photographs, and road networks for transportation infrastructure assessment during floods. The results showed that, when remote sensing data were missing or unavailable, social media data could provide effective information for evaluation. In [79], the Kriging interpolation was used to fuse crowdsourced, remote sensing, and social media data during and after Hurricane Sandy for transportation infrastructure assessment. It was proven that nonauthoritative data could be used to fill the gaps in remote sensing data. A similar study for the assessment of transportation infrastructure was also presented in [80]. The study in [81] first analyzed the spatial relationship between the brightness change of the satellite nightlight images and the density of power-related tweets. On this basis, the assessment of power outage information at a street-level resolution during Hurricane Sandy was realized.

3) Urban Land-Use Classification: Social media data can provide socioeconomic and demographic characteristics of urban land, thereby contributing to accurate remote sensing land-use mapping in complex urban systems [82]. In [83], a novel scene classification framework to identify dominant urban land-use type was proposed. In this framework, a land-use word dictionary was constructed by fusing natural–physical features from high spatial resolution images and socioeconomic semantic features from social media data. In related fashion, the study in [84] proposed a per-field classification approach to automatically map fine-grained urban land use. Specifically, the physical and socioeconomic information extracted from GF-2 imagery, points of interest (POIs), and geotagged Weibo posts was fused in an optimized RF model. The research in [85] developed a Crowd4RS system to generate labels of land use for remote sensing data. The proposed system combined the semantics of location-based social media photographs and big data analysis techniques (active learning, deep learning, clustering algorithms, and so on).

4) Urban Functional Zone Classification: Since human activities have a significant impact on urban morphology, social media data that can record information about human activities have been fused with remote sensing data for the division of urban functional zones [86]. For instance, in [87], a hierarchal clustering method was used to integrate landscape metrics from SPOT-5 images and human activity metrics from mobile phone positioning data for urban functional zones identification. The work in [88] presented a new method to identify the main center and subcenters of a polycentric city using nighttime light (NTL) imagery, social media data, cluster analysis, and GWR. In [89], the light gradient boosting machine (LightGBM) was used to fuse dual-modal data of high-resolution remote sensing images and user behavior data for urban functional zone classification. The study in [90] proposed a novel end-to-end deep learning-based remote and social sensing data fusion model. The proposed model mainly solved the asynchrony of the two data sources through enforcing cross-modal feature consistency (CMFC) and cross-modal triplet (CMT) constraints.

5) Other Applications: There have also been relevant works focused on the fusion of remote sensing and social media data in other application domains. For instance, the work presented in [91] proposed a new population mapping method by integrating the International Space Station (ISS) photography NTL data, POI data, and social media data through an RF model. The paper in [92] explored the potential of geotagged tweets to improve the quality of NTL images using an upsampling strategy for more accurate estimation of socioeconomic factors (e.g., personal incomes). In [93], a population-weighted metric was adopted to estimate dynamic population exposure to PM2.5, combining satellite-derived dynamic changes in PM2.5 concentrations and population distribution from social media data.

As it can be concluded from the aforementioned works, the fusion of heterogeneous remote sensing and social media data exhibits huge potential for different applications, especially for problems in which (near) real-time response is needed. Generally, the fusion of big heterogeneous data streams coming from various sensors is computationally very expensive. Therefore, such fusion demands effective parallel and distributed computing implementations, especially in the context of time-critical applications.

III. PARALLEL AND DISTRIBUTED COMPUTING IMPLEMENTATIONS OF TECHNIQUES COMBINING REMOTE SENSING AND OTHER DATA SOURCES

Currently, there are few research efforts devoted to the fusion of remote sensing and social media data using high-performance computing technologies. However, there have been several techniques and applications exploiting remote sensing and other sources of data independently using parallel and distributed computing. It has been proven that distributed computing is a feasible solution for the fusion of remote sensing big data. In a similar fashion, distributed computing may also be exploited to address the computational challenges brought by fusing remote sensing and social media data. In the following, we provide a literature review of existing techniques and applications. First, we focus on the most relevant efforts exploiting multisource remote sensing data. Then, we introduce some of the most important related works that are based on fusing remote sensing and other sources of information (e.g., traffic-site, climate, and social media data).

A. Multisource Remote Sensing Data

Fusing multisource remote sensing data can provide significant improvements in different applications. However, since the spatial, spectral, and temporal resolutions of remotely sensed data are continuously increasing, their size is becoming extremely large. Therefore, it generally takes a significant amount of time to fuse remote sensing data collected by different sensors. In turn, time-critical applications (such as damage assessment) require immediate responses [94]. To ensure high accuracy without compromising execution time, many researchers have been devoted to utilizing high-performance computing architectures to improve the computational performance of remote sensing techniques. Generally, there are two types of high-performance computing methods involved: parallel and distributed techniques. On the one hand, parallel computing (in a single computer) can be summarized into two categories: 1) computing based on multicore central processing units (CPUs) and 2) computing based on CPU processors plus accelerators, such as graphical processing units (GPUs). The former is called homogeneous parallel computing, while the latter is called heterogeneous parallel computing. Parallel computing on multicore CPUs is normally implemented via programming languages and libraries, such as message passing interface (MPI)[1] or OpenMP.[2] For instance, Yang *et al.* [95] used MPI to implement a parallel pansharpening algorithm for SPOT-5 multispectral and panchromatic images. The results demonstrated that the parallel algorithm, implemented on a computer with two CPU cores, was much faster than the serial version. In addition, the fusion results of the different versions had no obvious differences. The compute unified device architecture (CUDA)[3] is another programming model for parallel computing based on CPU + GPU hybrid heterogeneous architectures. This model has been used to accelerate many data-intensive algorithms, such as the intensity hue saturation (IHS) or the Y (perceived luminance), I, Q (color/luminance information)—YIQ—transform-based multisource remote sensing data fusion algorithms [96], [97]. Finally, distributed computing mainly includes cluster and cloud computing. Compared to parallel computing on a single computer, distributed computing needs additional computing resources but also provides more computing power and memory. In the following, we give a short survey of these architectures and related applications.

1) Cluster Computing: We can categorize cluster computing systems into two kinds: homogeneous and heterogeneous. A cluster system consists of at least one computing node. A typical homogeneous cluster is the CPU cluster system. Each computing node of a CPU cluster system has independent memory and may include multiple CPU processors. All the CPU processors can execute in parallel if the computing task of each CPU processor is independent of others. For instance, fusion algorithms using multisource remote sensing images can be independently executed for each pixel of a target remote sensing image. Therefore, it is suitable to perform parallel data fusion in CPU cluster systems. Experimental results in [98]–[100] showed that CPU-based cluster computing could significantly reduce the elapsed running time of fusion algorithms when the size of the datasets was large. However, if the size of the datasets was too small, the parallel performance generally decreases [101]. Unlike homogeneous cluster systems, heterogeneous clusters are equipped with different kinds of computing processors (e.g., CPUs and GPUs). MPI + CUDA is a typical distributed programming framework for GPU cluster systems. In [102], this framework was used to address the time-series quantitative retrieval problem, which could be implemented using multilevel parallelism including fine-grained parallelism (using the GPU) and coarse-grained parallelism (across CPUs). A similar study based on MPI + CUDA was presented in [103]. The results showed that the combination of MPI and CUDA in the discussed implementation could significantly speed up the oil detection process and provide a rapid response.

2) Cloud Computing: Compared to cluster computing systems, cloud computing systems provide computing resources in an easy-to-use and cheap way while managing and storing huge amounts of data in distributed, fault-tolerant environments. Therefore, these systems [e.g., Apache Spark and Google Earth Engine (GEE)] are suitable for parallel processing of massive volumes of multisource remote sensing data. For instance, the work in [104] implemented a novel remote sensing data flow (RESFlow) on Apache Spark for advancing machine learning to compute with 21 028 TB of remotely sensed imagery. The experimental results demonstrated that cloud computing could improve the efficiency of the algorithms that were based on a large amount of multisource remote sensing data. Moreover, the framework in [105] incorporated a task scheduling strategy to further exploit the parallelism of multispectral pansharpening algorithms. The algorithm was executed on the Apache Spark platform. In addition, the work in [106] proposed a cloud-based automatic remote sensing production system. This program supported storing and processing massive multisource remote sensing data, obtaining various remote sensing products. In addition to processing multisource remote sensing data, cloud computing has also been widely used in many other applications. For example, in [107], a framework using cloud computing was proposed, which could provide effective drought monitoring, evaluation, and prediction. Then, Mahdianpari *et al.* [108] produced a high-resolution 10-m wetland inventory map of Canada using multiyear, multisource (Sentinel-1 and Sentinel-2) EO data on the GEE. Also, based on the GEE, the study in [109] analyzed the influence of different physical surface properties using Landsat imagery collected between 1985 and 2018. In addition, the work in [110] developed a new cloud computing infrastructure for environmental monitoring using interagency EO data. Another related work based on a Hadoop cloud computing platform for drought monitoring was presented in [111]. The authors first proposed an abstract data format to achieve the unified description of remote sensing data. The data abstraction was intended to discretize multidimensional remote sensing data for simplified distributed storage and computation.

[1] https://computing.llnl.gov/tutorials/mpi/
[2] https://www.openmp.org/
[3] https://developer.nvidia.com/CUDA-zone

Table 1 Advantages and Limitations of Existing Parallel and Distributed Computing Technologies

	Advantages	Limitations
MPI	Process-level parallelism within a single computer or across cluster computing nodes; Capable of tackling large scale data-intensive problems	Processes shared results through communication functions; Speedup limited by communication cost problems
OpenMP	Thread-level fine-grained parallelism within a single computer; The results of computation between threads can be shared; The modification of code is limited	Cannot resolve large-scale problems due to memory insufficient of a single computer
CUDA	Fine-grained parallelism within a GPU; Capable of accelerating deep learning models	Computing with multiple conditional branches is slow; Data transfers between GPU and CPU are costly; Limited memory of GPU
Google Earth Engine	Capable of processing peta-byte scale data in parallel	Internet required; Less security Limited size of uploaded user data

Then, they resolved the complexity of remote sensing algorithms using MapReduce in a Hadoop distributed environment. Moreover, the research in [112] provided a detailed description of a web platform that offered an integrated framework for disaster monitoring using cloud computing.

B. Multisource Remote Sensing Data and Auxiliary Data

There has been a significant amount of research based on exploiting multisource remote sensing data and auxiliary data using parallel and distributed computing. For instance, in [9], a parallel Markov model was presented for land-use prediction based on MapReduce, a distributed parallel framework. The experimental datasets included Landsat remote sensing images, traffic-site data, road networks, and location-address data. In [113], global positioning system (GPS) and position and orientation (POS) data were used to preprocess images collected by UAVs. Then, in order to shorten the time to find the relative and global orientation, the Levenberg–Marquard algorithm was parallelized on multicore CPUs. In order to reduce the computation cost of object-based image analysis (OBIA) for quick earthquake damage assessment, a distributed OBIA approach was used in [114]. After splitting the images (including Quickbird, DEM, LiDAR, and thematic maps) of a large study area into small subimages, two or more computers could be used in parallel for analyzing the subimages. A CPU + GPU heterogeneous parallel technique was used to implement deep learning algorithms for applications based on multisource data. The paper in [115] presented a 3-D virtual urban scene reconstruction method based on CNNs by combining maps, satellite optical images, and digital terrain models (DTMs). The CNN model was trained on a GPU. In [10], MODIS NDVI data, ground data, and climate data were used for investigating drought-vulnerable regions in North Korea. The climate data were downloaded from the Climate Engine, a web-based parallel cloud computing platform for processing climate and remote sensing datasets in real time [116]. The research in [11] used Landsat satellite and climate data to quantify the effectiveness of riparian restoration. The climate data were also downloaded from Climate Engine. A similar study for cloud-based remote sensing applications was proposed in [1]. This work pursues historical disaster analysis based on multisource data (e.g., social media, remote sensing, Wikipedia, and the Worldwide Web). The work in [12] mapped major land cover dynamics in Beijing with Landsat images and NVDI using GEE. The GEE platform provides many different source datasets and supports high-performance computing. Therefore, it can help researchers to address issues related to natural disasters, such as monitoring, prediction, evaluation, and prevention. For instance, the work in [117] presented an algorithm that exploited all available Sentinel-1 SAR images (in combination with historical Landsat and other auxiliary data sources hosted on the GEE) to rapidly map surface inundation during flood events. The authors also assessed their algorithm using three recent flood events with coincident VHR imagery and operational flood maps.

The advantages and limitations of existing parallel and distributed technologies for the fusion of remote sensing and other data sources are summarized in Table 1. Similarly, the fusion of remote sensing and social media data using high-performance computing capabilities can be conducted by combining suitable parallel and distributing technologies.

IV. CASE STUDY: HETEROGENEOUS DATA FUSION USING DISTRIBUTED COMPUTING

As a follow-up to Section III, we discuss a case study related to a flood event in Boulder (2013), which is addressed via distributed fusion of heterogeneous data. Specifically, the goal of this section is to discuss (with an example) how to fuse heterogeneous remote sensing and social media data by combining domain adaptation (i.e., GOT [76]) and distributed computing techniques (i.e., MPI [95] and OpenMP [118]). In the following, we first introduce the dataset used in our case study, and then, the considered frameworks for the distributed fusion of heterogeneous data are illustrated in detail.

A. Data Collection

In this study, the considered distributed fusion frameworks for heterogeneous data fusion are evaluated using the 2013 flood event in Boulder. During this disastrous event, the local rainfall exceeded 17 in, and nearly one year worth of rainfall was received in just eight days in

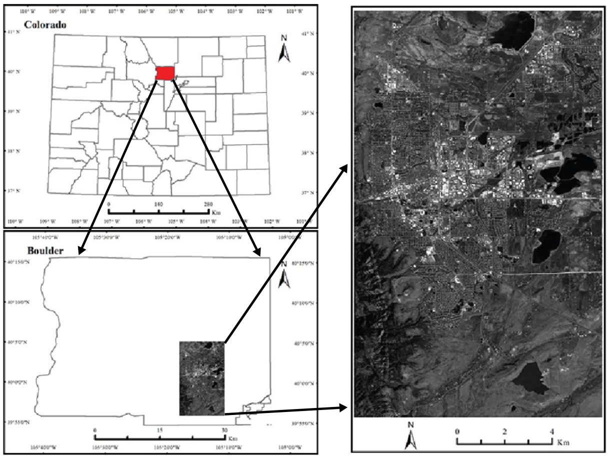


Fig. 1. ***Study area and its location in Boulder and Colorado.***

the Boulder area [119]. In total, the widespread flooding led by continuous rainfall killed three people, evacuated more than 1600 people, and caused significant infrastructure damage [120]. The study area (with a total area of 180 km^2) and its location are shown in Fig. 1.

1) Remote Sensing Data: The remote sensing data used in our study comprise two multispectral Landsat 8 operational land imager (OLI) images, which were completely cloud-free in the city of Boulder. Specifically, the images obtained on May 12 and September 17, 2013, can provide high-quality optical data before and after the flood event, respectively. These images are publicly available from the United States Geological Survey (USGS) website[4] with a fine spatial resolution of 30 m. In addition, the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm was adopted for atmospheric correction of the Landsat 8 OLI images [121]. In this context, remote sensing features (capable of characterizing floods) can be calculated for subsequent data fusion.

2) Social Media Data: Twitter is one of the largest social networking sites, providing microblogging services. Users can send and receive short messages through the Twitter website of mobile phones or computers. Due to the extensive coverage and real-time nature of these messages (called "tweets"), information can be quickly disseminated by both official government agencies and the public during emergencies and disasters [122]. The hashtags prefixed with the sign # are usually used to search and filter massive information from tweets [120]. In this study, the tweets containing hashtags of #boulderflood, #flood, or #coflood during September 11–18, 2013, were filtered. A total of 2254 geotagged tweets were harvested using Twitter APIs, extending from $105^\circ 18' 2''$ to $105^\circ 10' 40''$ W and $39^\circ 55' 54''$ to $40^\circ 5' 8''$ N [see Fig. 2(a)]. It is worth noting that, since geotagged tweets are generally sent from relatively safe places rather than directly on the flood sites, they tend to suffer a problem of spatial bias. Such biased geotagged tweets are used as the source distribution for data fusion purposes.

3) Historical Flood Data and Ground Truth: The special flood hazard area (SFHA) shown on the flood insurance rate maps (FIRMs) is defined as the area that has a 1% chance of being inundated in any given year. Generally, the SFHA is considered as the base flood or 100-year flood, which can well represent historical flood areas. The related data of SFHA in the city of Boulder can be downloaded from the website of the Federal Emergency Management Agency.[5] Fig. 2(a) illustrates the corresponding SFHA for Boulder.

In the weeks and months following the flood event, the spatial distribution of the floods was identified by combining hand-held GPS devices and photographs from communities. Then, an urban flood extent (UFE) map [see Fig. 2(b)] was constructed to describe the accurate inundated regions in the city of Boulder, which is publicly available on the website of Boulder city.[6] The UFE can be used as the ground truth to evaluate the data fusion.

B. Heterogeneous Data Fusion

Multisource data fusion can provide more accurate and reliable information, exhibiting significant advantages and potential for emergency response (e.g., flood monitoring) [120]. As mentioned earlier, the geotagged tweets that indicate floods are usually spatially biased. In this case, the flood-related applications will no doubt be affected

[4]http://www.earthexplorer.usgs.gov

[5]http://www.fema.gov
[6]http://www.bouldercolorado.gov

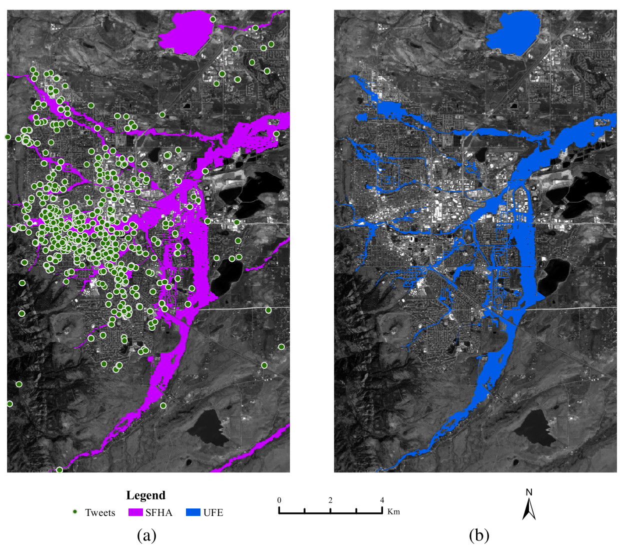


Fig. 2. ***Spatial distribution of the geotagged tweets, history flood data, and UFE for the 2013 Boulder flood event [76]. (a) Tweets and special flood hazard area. (b) Urban flood extent.***

when using such geotagged tweets directly. Therefore, the task of data fusion needs to relocate the biased geotagged tweets to the flooded areas (i.e., the UFE), in the combination of the prior flood information (i.e., the SFHA) and remote sensing features (capable of characterizing floods). It should be pointed out that domain adaptation techniques have been proven as effective means for representation alignment of multisource data in Section II-C, including the OT method [3], [75]. Compared with the OT, the GOT method in [76] achieved more remarkable results in the context of heterogeneous remote sensing and social media data fusion. In addition, the GOT method is able to simultaneously align representations and geolocations of multimodal data. Therefore, we adopt GOT for heterogeneous data fusion in this study.

Following Liu *et al.* [76], the main function of GOT for data fusion can be expressed as

$$T\left(x_i^s\right) = \underset{x_j^t \in \widehat{P}_t}{\arg\min}\, c\left(x_i^s, x_j^t\right) \tag{1}$$

where x_i^s denotes a source element in the source distribution P_s and x_j^t denotes a target element in the target distribution P_t. $\widehat{P}_t$ denotes the empirical distribution of P_t when P_t is only accessible through discrete samples. $c(x_i^s, x_j^t)$ represents the cost to move a probability mass from x_i^s to x_j^t. Following Liu *et al.* [76], the transport cost $c(x_i^s, x_j^t)$ also considers distance and two kinds of remote sensing features (i.e., NDVI [123] and NDWI [124]) in this study. Specifically, the cost function considers the squared Euclidean distance between x_i^s and x_j^t, aimed at transporting the source to the target within a relatively small distance. As for the cost of remote sensing features, two new forms of features are derived based on NDVI and NDWI: NDVI difference and NDWI difference (considering the times before and after a flood event). Both NDVI difference and NDWI difference used in the transport cost $c(x_i^s, x_j^t)$ are bounded by extreme inundated scenarios, in order to transport sources to flooded areas. In this context, the spatially biased geotagged tweets can be relocated to the floods within a relatively small distance by combining the transport cost of distance and the aforementioned remote sensing features.

C. Distributed Frameworks

As mentioned before, distributed computing techniques have been widely applied to the fusion of remote sensing and other data sources. However, there are few research studies on the distributed fusion of remote sensing and social media data. Here, we introduce our newly proposed distributed fusion frameworks for heterogeneous data based on the aforementioned GOT. The main function of GOT [i.e., (1)] can be divided into three main parts: 1) computing array c (for $i = 1, \ldots, n_s$ and $j = 1, \ldots, n_t$); 2) computing array T (for $i = 1, \ldots, n_s$); and 3) other computations. After the profiling results of the GOT algorithm in Table 2, we find out that the time-consuming parts are the calculations of arrays c and T. To calculate array c, we need to use the function dist() in the library of OT to calculate the Euclidean distance matrix between the source samples and the target samples. It is obvious that computing c with different x_i^s's is an independent operation. Therefore, we can split the array x^s into smaller ones and then compute c in parallel. In a similar fashion, we can also compute T in parallel. According to the comparisons of existing parallel computing technologies for heterogeneous data fusion in Table 1, we employ an MPI + OpenMP hybrid programming model to implement this procedure. This model combines the advantages of MPI and OpenMP to implement two levels of parallelism (i.e., fine-grained parallelism within computing nodes and coarse-grained parallelism across nodes) for the computation of c and T.

Table 2 Profiling Results of the GOT Algorithm

Name	Occupation
Computing array c	81.16%
Computing array T	10.59%
Others	8.25%

MPI is a communication protocol for distributed computing based on distributed memory, which can be binded to Python language by using the `mpi4py` package [125]. MPI employs built-in functions to implement communications between processes. `Gatherv()` of MPI.Comm class is a communication function for gathering results, which can gather different lengths of the array to the root process from each process. OpenMP is a thread-level parallel programming model based on shared memory. However, OpenMP does not provide the corresponding version of Python. We use the thread class of the threading[7] package to implement thread-level parallelism, such as OpenMP. Thus, we denote this distributed computing method as "MPI-Thread." The pseudocode of our parallel implementation for the computation of array c and T is shown in Algorithm 1. First, the array $(x^s)^{n_s}$ is decomposed into K blocks, each with a size of (n_s/K). If n_s is not divisible by K, the rest will be computed by the last process. Then, the computation of array T and c is parallelized by K processes, each of which is executed by a single computing node. Meanwhile, the array $(x^s)^{n_s/K}$ (assigned to a process) is further divided into k blocks, while T and c are concurrently computed by k threads, each of which is executed by a single CPU core of a computing node. Finally, we gather the results in the root process by using `Gatherv()`. In addition, we also use MPI and threads to parallelize the GOT algorithm. Especially, when using the threads method to parallelize the GOT algorithm, we divide the array assigned to a thread into small ones and then calculate them serially on a single thread (mainly because the memory of one thread is limited).

[7] https://docs.python.org/3/library/threading.html

Algorithm 1 Pseudocode of the Parallel Algorithm for Computing Arrays c and T

Input: Size of the source distribution: n_s,
Array: $x^s[0{:}\ n_s]$, Number of processes: K, Number of threads: k
Output: Array: $c[0{:}\ n_s,\ 0{:}\ n_t]$, $T[0{:}\ n_s,\ 2]$
$\text{b}_{\text{size}} \leftarrow n_s // K$
$\text{b}_{\text{id}} \leftarrow$ process rank ▷ Obtain by MPI.Comm.rank()
$\text{t}_{\text{id}} \leftarrow$ thread id
$\text{t}_{\text{size}} \leftarrow b_{size} // k$
Parallel for $\text{b}_{\text{id}} < K$-1 **do**
  Parallel for $\text{t}_{\text{id}} < k$-1 **do**
    compute $c^{t_{size} \times n_t}$
    local_dis $\xleftarrow{append}$ argmin($c^{t_{size} \times n_t}$)
  end Parallel for
end Parallel for
if $\text{b}_{\text{id}} == K-1$ **then**
  Parallel for $\text{t}_{\text{id}} < k$-1 **do**
    compute $c^{t_{size} \times n_t}$
    local_dis $\xleftarrow{append}$ argmin($c^{t_{size} \times n_t}$)
  end Parallel for
end if
MPI.Comm.gatherv(local_dis, $T[0{:}\ n_s,\ 2]$)

Table 3 Evaluation of the GOT Algorithm Implemented Using Different Distributed Computing Strategies

Method	Times(s)
None	15.19
MPI	4.88
Thread	13.65
MPI-Thread	3.47

D. Experimental Results

Here, we explore the performance of the proposed frameworks for distributed heterogeneous data fusion. The experiments have been run on a CPU cluster called TianHe-2,[8] which is a distributed cluster system equipped with many computing nodes. Each computing node is equipped with multiple CPU cores. Multiple processes can be run on multiple nodes simultaneously. Multiple threads can run on multiple CPU cores at once. We first perform a quantitative evaluation of the GOT combining with different distributed computing strategies for the 2013 Boulder flood event. The datasets used here include 2254 tweets and two multispectral Landsat 8 OLI images with an area of 180 km^2. Table 3 shows the obtained results in terms of time consumption. Regarding the running time (in seconds), the heterogeneous data fusion framework requires 15 s. After parallelizing the computation of c and T, the proposed distributed data fusion frameworks based on MPI and MPI-Thread are much faster (only 4.88 and 3.47 s, respectively). Notice that the amount of tweets involved in this case study is only 2254. To reduce the time of opening and closing multiple processes, the MPI and MPI-Thread methods developed in this study use five processes. Correspondingly, the Thread and MPI-Thread methods both use five threads.

[8]http://en.nscc-gz.cn/

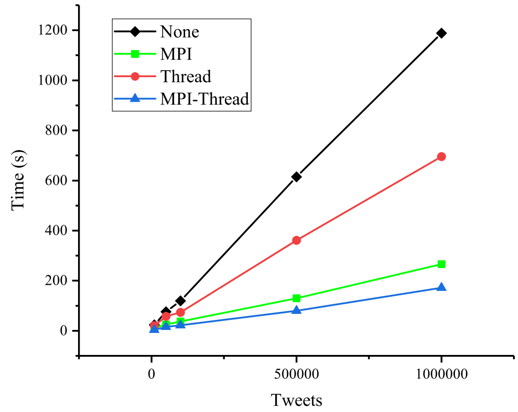


Fig. 3. ***Evaluation of the proposed distributed fusion frameworks after simulating different numbers of tweets.***

Furthermore, we perform additional experiments considering a large amount of social media data. Specifically, we artificially expanded the number of tweets for the 2013 Boulder flood to 10 000, 50 000, 100 000, 500 000, and 1 000 000, respectively, to simulate and test the performance of our distributed fusion frameworks. The results of the simulated experiments are shown in Fig. 3. Concerning the time consumption (in seconds), several conclusions can be obtained. First, the difference between the adopted data fusion framework (GOT) and the proposed distributed data fusion frameworks is relatively small when the number of tweets is less than 100 000. This is reasonable due to the inefficiency of distributed computing when considering small volumes of data [101]. In addition, with the increase in the number of tweets, the performance of our distributed frameworks increases significantly. Specifically, the proposed distributed data fusion framework using MPI-Thread exhibits the best computing performance. The running time of the MPI-Thread implementation is only 171 s when simulating 1 000 000 tweets, which is one-seventh of the total time taken by GOT. It should be pointed out that, with the expansion of remote sensing and social media data volumes, our proposed distributed data fusion frameworks can significantly scale in terms of computational efficiency.

V. CONCLUSION

In this article, we have presented a comprehensive review of the state of the art in remote sensing techniques and applications that use various data sources. First, we have focused on research efforts that are based on fusing remote sensing and other sources of data for EO applications. Next, we give an overview of parallel and distributed fusion implementations of techniques using multisource remote sensing data and other sources of data. In order to present a realistic case study of distributed fusion of heterogeneous remote sensing and other sources of data, a real flood event is discussed in detail. According to our experiments, the proposed distributed data fusion frameworks (exploiting both remote sensing and social media data) exhibit high processing efficiency and the potential for rapid computation in real scenarios. In future research, large-volume multisource remote sensing data should be considered in the distributed data fusion frameworks for various EO applications. For instance, the GEE platform has shown advanced cloud computing and storage capabilities [126], which provides unprecedented opportunities to process remotely sensed big data combined with other data sources. ■

REFERENCES

[1] Q. Huang, G. Cervone, and G. Zhang, "A cloud-enabled automatic disaster analysis system of multi-sourced data streams: An example synthesizing social media, remote sensing and Wikipedia data," *Comput., Environ. Urban Syst.*, vol. 66, pp. 23–37, Nov. 2017.

[2] J. Li *et al.*, "Social media: New perspectives to improve remote sensing for emergency response," *Proc. IEEE*, vol. 105, no. 10, pp. 1900–1912, Oct. 2017.

[3] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1853–1865, Sep. 2017.

[4] M. K. Pektürk and M. Ünal, "A review on real-time big data analysis in remote sensing applications," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, May 2017, pp. 1–4.

[5] C. Wickramaarachchi, A. Kumbhare, M. Frincu, C. Chelmis, and V. K. Prasanna, "Real-time analytics for fast evolving social graphs," in *Proc. 15th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput.*, May 2015, pp. 829–834.

[6] C. A. Lee, S. D. Gasster, A. Plaza, C.-I. Chang, and B. Huang, "Recent developments in high performance computing for remote sensing: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp. 508–527, Sep. 2011.

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

[8] L. Wang, Y. Ma, J. Yan, V. Chang, and A. Y. Zomaya, "PipsCloud: High performance cloud computing for remote sensing big data management and processing," *Future Gener. Comput. Syst.*, vol. 78, pp. 353–368, Jan. 2018.

[9] J. Kang, L. Fang, S. Li, and X. Wang, "Parallel cellular automata Markov model for land use change prediction over MapReduce framework," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 10, p. 454, Oct. 2019.

[10] J. Yu, J. Lim, and K.-S. Lee, "Investigation of drought-vulnerable regions in North Korea using remote sensing and cloud computing climate data," *Environ. Monitor. Assessment*, vol. 190, no. 3, p. 126, Mar. 2018.

[11] M. B. Hausner *et al.*, "Assessing the effectiveness of riparian restoration projects using Landsat and precipitation data from the cloud-computing application ClimateEngine.org," *Ecol. Eng.*, vol. 120, pp. 432–440, Sep. 2018.

[12] H. Huang *et al.*, "Mapping major land cover dynamics in Beijing using all Landsat images in Google earth engine," *Remote Sens. Environ.*, vol. 202, pp. 166–176, Dec. 2017.

[13] R. Wang, "Big data of urban flooding: Dance with social media, citizen science, and artificial intelligence," in *Proc. EGU Gen. Assem. Conf. Abstr.*, 2018, p. 404.

[14] A. H. S. Solberg, A. K. Jain, and T. Taxt, "Multisource classification of remotely sensed data: Fusion of Landsat TM and SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 768–778, Jul. 1994.

[15] C. C. Petit and E. F. Lambin, "Integration of multi-source remote sensing data for land cover change detection," *Int. J. Geograph. Inf. Sci.*, vol. 15, no. 8, pp. 785–803, Dec. 2001.

[16] M. B. A. Gibril, M. O. Idrees, H. Z. M. Shafri, and K. Yao, "Integrative image segmentation optimization and machine learning approach for high quality land-use and land-cover mapping using multisource remote sensing data," *J. Appl. Remote Sens.*, vol. 12, no. 1, 2018, Art. no. 016036.

[17] C. Sukawattanavijit, J. Chen, and H. Zhang, "GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 284–288, Mar. 2017.

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

[19] B. Rasti, P. Ghamisi, and R. Gloaguen, "Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3997–4007, Jul. 2017.

[20] M. Zhang, P. Ghamisi, and W. Li, "Classification of hyperspectral and LiDAR data using extinction profiles with feature fusion," *Remote Sens. Lett.*, vol. 8, no. 10, pp. 957–966, Oct. 2017.

[21] P. Ghamisi, B. Höfle, and X. X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, Jun. 2017.

[22] P. Ghamisi, J. A. Benediktsson, and S. Phinn, "Land-cover classification using both hyperspectral and LiDAR data," *Int. J. Image Data Fusion*, vol. 6, no. 3, pp. 189–215, Jul. 2015.

[23] P. C. Sutton, "An empirical environmental sustainability index derived solely from nighttime satellite imagery and ecosystem service valuation," *Population Environ.*, vol. 24, no. 4, pp. 293–311, 2003.

[24] C. He, B. Gao, Q. Huang, Q. Ma, and Y. Dou, "Environmental degradation in the urban areas of China: Evidence from multi-source remote sensing data," *Remote Sens. Environ.*, vol. 193, pp. 65–75, May 2017.

[25] L. Lu, Q. Weng, H. Guo, S. Feng, and Q. Li, "Assessment of urban environmental change using multi-source remote sensing time series (2000–2016): A comparative analysis in selected megacities in Eurasia," *Sci. Total Environ.*, vol. 684, pp. 567–577, Sep. 2019.

[26] Y. Zhang and J. Cheng, "Spatio-temporal analysis of urban heat island using multisource remote sensing data: A case study in Hangzhou, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3317–3326, Sep. 2019.

[27] G. Li, L. Li, D. Lu, W. Guo, and W. Kuang, "Mapping impervious surface distribution in China using multi-source remotely sensed data," *GISci. Remote Sens.*, vol. 57, no. 4, pp. 543–552, May 2020.

[28] Y. Smara, A. Belhadj-Aissa, B. Sansal, J. Lichtenegger, and A. Bouzenoune, "Multisource ERS-1 and optical data for vegetal cover assessment and monitoring in a semi-arid region of Algeria," *Int. J. Remote Sens.*, vol. 19, no. 18, pp. 3551–3568, Dec. 1998.

[29] G. G. Wilkinson, S. Folving, I. Kanellopoulos, N. McCormick, K. Fullerton, and J. Mégier, "Forest mapping from multi-source satellite data using neural network classifiers—An experiment in Portugal," *Remote Sens. Rev.*, vol. 12, nos. 1–2, pp. 83–106, Jan. 1995.

[30] X. Wang and H. Jiao, "Spatial scaling of forest aboveground biomass using multi-source remote sensing data," *IEEE Access*, vol. 8, pp. 178870–178885, 2020.

[31] C. Bourgoin *et al.*, "The potential of multisource remote sensing for mapping the biomass of a degraded Amazonian forest," *Forests*, vol. 9, no. 6, p. 303, May 2018.

[32] M. Xu *et al.*, "Change detection of the Tangjiashan barrier lake based on multi-source remote sensing data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 4, Jul. 2009, p. IV-303.

[33] B. Fu and A. Lin, "Spatial distribution of the surface rupture zone associated with the 2001 Ms 8.1 Central Kunlun earthquake, northern Tibet, revealed by satellite remote sensing data," *Int. J. Remote Sens.*, vol. 24, no. 10, pp. 2191–2198, Jan. 2003.

[34] J. Nichol and M. S. Wong, "Satellite remote sensing for detailed landslide inventories using change detection and image fusion," *Int. J. Remote Sens.*, vol. 26, no. 9, pp. 1913–1926, May 2005.

[35] H. Shahabi and M. Hashim, "Landslide susceptibility mapping using GIS-based statistical models and remote sensing data in tropical environment," *Sci. Rep.*, vol. 5, no. 1, pp. 1–15, Sep. 2015.

[36] A. M. Dewan, M. M. Islam, T. Kumamoto, and M. Nishigaki, "Evaluating flood hazard for land-use planning in greater Dhaka of Bangladesh using remote sensing and GIS techniques," *Water Resour. Manage.*, vol. 21, no. 9, pp. 1601–1612, Aug. 2007.

[37] X. Yang, Y. Cao, B. Xu, and X. Zhu, "The fusion and application of multi-source remote sensing data in real time monitoring of grassland snow of China," *Geograph. Res.*, vol. 28, no. 6, pp. 1704–1712, 2009.

[38] X. Liu *et al.*, "Large-scale crop mapping from multisource remote sensing images in Google earth engine," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 414–427, 2020.

[39] J.-R.-B. Bwangoy, M. C. Hansen, D. P. Roy, G. D. Grandi, and C. O. Justice, "Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 73–86, Jan. 2010.

[40] C. Doña *et al.*, "Integrated satellite data fusion and mining for monitoring lake water quality status of the Albufera de Valencia in Spain," *J. Environ. Manage.*, vol. 151, pp. 416–426, Mar. 2015.

[41] H. Jiang, Y. Rusuli, T. Amuti, and Q. He, "Quantitative assessment of soil salinity using multi-source remote sensing data based on the support vector machine and artificial neural network," *Int. J. Remote Sens.*, vol. 40, no. 1, pp. 284–306, Jan. 2019.

[42] J. She, Y. Zhang, X. Li, and Y. Chen, "Changes in snow and glacier cover in an arid watershed of the western Kunlun Mountains using multisource remote-sensing data," *Int. J. Remote Sens.*, vol. 35, no. 1, pp. 234–252, Jan. 2014.

[43] M. Schmitt and X. X. Zhu, "On the challenges in stereogrammetric fusion of SAR and optical imagery for urban areas," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 41, no. B7, pp. 719–722, 2016.

[44] S. Workman, M. U. Rafique, H. Blanton, C. Greenwell, and N. Jacobs, "Single image cloud detection via multi-image fusion," 2020, *arXiv:2007.15144*. [Online]. Available: http://arxiv.org/abs/2007.15144

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

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

[47] P. J. Leitão *et al.*, "From sample to pixel: Multi-scale remote sensing data for upscaling aboveground carbon data in heterogeneous landscapes," *Ecosphere*, vol. 9, no. 8, Aug. 2018, Art. no. e02298.

[48] M. Rigge *et al.*, "Quantifying western U.S. Rangelands as fractional components with multi-resolution remote sensing and *in situ* data," *Remote Sens.*, vol. 12, no. 3, p. 412, Jan. 2020.

[49] M. G. Filho, T. M. Kuplich, and F. L. F. D. Quadros, "Estimating natural grassland biomass by vegetation indices using Sentinel 2 remote sensing data," *Int. J. Remote Sens.*, vol. 41, no. 8, pp. 2861–2876, Apr. 2020.

[50] X. Guan *et al.*, "A 33-year NPP monitoring study in southwest China by the fusion of multi-source remote sensing and station data," *Remote Sens.*, vol. 9, no. 10, p. 1082, Oct. 2017.

[51] R. Zhang *et al.*, "Estimating aboveground biomass in subtropical forests of China by integrating multisource remote sensing and ground data," *Remote Sens. Environ.*, vol. 232, Oct. 2019, Art. no. 111341.

[52] Z. Ma, X. Hu, L. Huang, J. Bi, and Y. Liu, "Estimating ground-level $PM_{2.5}$ in China using satellite remote sensing," *Environ. Sci. Technol.*, vol. 48, no. 13, pp. 7436–7444, 2014.

[53] Z. S. Venter, O. Brousse, I. Esau, and F. Meier, "Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data," *Remote Sens. Environ.*, vol. 242, Jun. 2020, Art. no. 111791.

[54] H. Shen, Y. Jiang, T. Li, Q. Cheng, C. Zeng, and L. Zhang, "Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data," *Remote Sens. Environ.*, vol. 240, Apr. 2020, Art. no. 111692.

[55] S. Vratolis *et al.*, "Comparison and complementary use of *in situ* and remote sensing aerosol measurements in the Athens Metropolitan Area," *Atmos. Environ.*, vol. 228, May 2020, Art. no. 117439.

[56] S. Sarkar and D. P. Kanungo, "An integrated approach for landslide susceptibility mapping using remote sensing and GIS," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 617–625, May 2004.

[57] R. P. Singh, W. Mehdi, R. Gautam, J. S. Kumar, J. Zlotnicki, and M. Kafatos, "Precursory signals using satellite and ground data associated with the Wenchuan earthquake of 12 May 2008," *Int. J. Remote Sens.*, vol. 31, no. 13, pp. 3341–3354, Jul. 2010.

[58] R. Hoque, D. Nakayama, H. Matsuyama, and J. Matsumoto, "Flood monitoring, mapping and assessing capabilities using RADARSAT remote sensing, GIS and ground data for Bangladesh," *Natural Hazards*, vol. 57, no. 2, pp. 525–548, May 2011.

[59] J. Li, S. Zhang, L. Huang, T. Zhang, and P. Feng, "Drought prediction models driven by meteorological and remote sensing data in Guanzhong Area, China," *Hydrol. Res.*, vol. 51, no. 5, pp. 942–958, Oct. 2020.

[60] J. Zhang and G. M. Foody, "A fuzzy classification of sub-urban land cover from remotely sensed imagery," *Int. J. Remote Sens.*, vol. 19, no. 14, pp. 2721–2738, Jan. 1998.

[61] S. Frolking *et al.*, "Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in China," *Global Biogeochem. Cycles*, vol. 16, no. 4, pp. 1–38, 2002.

[62] M. Linderman *et al.*, "Using artificial neural networks to map the spatial distribution of understorey bamboo from remote sensing data," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1685–1700, May 2004.

[63] A. E. K. Douaoui, H. Nicolas, and C. Walter, "Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data," *Geoderma*, vol. 134, nos. 1–2, pp. 217–230, Sep. 2006.

[64] J. F. Rosser, D. G. Leibovici, and M. J. Jackson, "Rapid flood inundation mapping using social media, remote sensing and topographic data," *Natural Hazards*, vol. 87, no. 1, pp. 103–120, May 2017.

[65] E. Schnebele, G. Cervone, S. Kumar, and N. Waters, "Real time estimation of the calgary floods using limited remote sensing data," *Water*, vol. 6, no. 2, pp. 381–398, Feb. 2014.

[66] E. Schnebele and G. Cervone, "Improving remote sensing flood assessment using volunteered geographical data," *Natural Hazards Earth Syst. Sci.*, vol. 13, no. 3, pp. 669–677, Mar. 2013.

[67] X. Huang, C. Wang, and Z. Li, "A near real-time flood-mapping approach by integrating social media and post-event satellite imagery," *Ann. GIS*, vol. 24, no. 2, pp. 113–123, Apr. 2018.

[68] A. Kruspe, J. Kersten, and F. Klan, "Review article: Detection of informative tweets in crisis events," *Nature Hazards Earth Syst. Sci. Discuss.*, 2020. [Online]. Available: https://nhess.copernicus.org/preprints/nhess-2020-214/, doi: 10.5194/nhess-2020-214.

[69] J. Kersten, A. Kruspe, M. Wiegmann, and F. Klan, "Robust filtering of crisis-related tweets," in *Proc. 16th Int. Conf. Inf. Syst. Crisis Response Manage. (ISCRAM)*, 2019, pp. 1–11.

[70] J. Kersten and F. Klan, "What happens where during disasters? A workflow for the multifaceted characterization of crisis events based on Twitter data," *J. Contingencies Crisis Manage.*, vol. 28, no. 3, pp. 262–280, Sep. 2020.

[71] B. Bischke, D. Borth, C. Schulze, and A. Dengel, "Contextual enrichment of remote-sensed events with social media streams," in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 1077–1081.

[72] K. Ahmad, K. Pogorelov, M. Riegler, N. Conci, and P. Halvorsen, "Social media and satellites," *Multimedia Tools Appl.*, vol. 78, no. 3, pp. 2837–2875, Feb. 2019.

[73] C. J. Scheele and Q. Huang, "An enhanced text-mining framework for extracting disaster relevant data through social media and remote sensing data fusion," in *Proc. Amer. Geophys. Union Fall Meeting*, 2016, p. 1844.

[74] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[75] H. Wang, E. Skau, H. Krim, and G. Cervone, "Fusing heterogeneous data: A case for remote sensing and social media," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 6956–6968, Dec. 2018.

[76] Z. Liu, Q. Qiu, J. Li, L. Wang, and A. Plaza, "Geographic optimal transport for heterogeneous data: Fusing remote sensing and social media," *IEEE Trans. Geosci. Remote Sens.*, early access, Oct. 28, 2020, doi: 10.1109/TGRS.2020.3031337.

[77] K. E. Joyce, S. E. Belliss, S. V. Samsonov, S. J. McNeill, and P. J. Glassey, "A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters," *Prog. Phys. Geogr.*, vol. 33, no. 2, pp. 183–207, Apr. 2009.

[78] G. Cervone, E. Schnebele, N. Waters, M. Moccaldi, and R. Sicignano, "Using social media and satellite data for damage assessment in urban areas during emergencies," in *Seeing Cities Through Big Data*. Cham, Switzerland: Springer, 2017, pp. 443–457. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-40902-3_24

[79] C. E. Oxendine, E. Schnebele, G. Cervone, and N. Waters, "Fusing non-authoritative data to improve situational awareness in emergencies," in *Proc. 11th Int. Conf. Inf. Syst. Crisis Response Manage. (ISCRAM)*, 2014, pp. 762–766.

[80] E. Schnebele, C. Oxendine, G. Cervone, C. M. Ferreira, and N. Waters, "Using non-authoritative sources during emergencies in urban areas," in *Computational Approaches for Urban Environments*. Cham, Switzerland: Springer, 2015, pp. 337–361. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-11469-9_14

[81] C. Hultquist, M. Simpson, G. Cervone, and Q. Huang, "Using nightlight remote sensing imagery and Twitter data to study power outages," in *Proc. 1st ACM SIGSPATIAL Int. Workshop Use GIS Emergency Manage.*, Nov. 2015, pp. 1–6.

[82] T. Hu, J. Yang, X. Li, and P. Gong, "Mapping urban land use by using Landsat images and open social data," *Remote Sens.*, vol. 8, no. 2, p. 151, Feb. 2016.

[83] X. Liu *et al.*, "Classifying urban land use by integrating remote sensing and social media data," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 8, pp. 1675–1696, Aug. 2017.

[84] Y. Zhang, Q. Li, H. Huang, W. Wu, X. Du, and H. Wang, "The combined use of remote sensing and social sensing data in fine-grained urban land use mapping: A case study in Beijing, China," *Remote Sens.*, vol. 9, no. 9, p. 865, Aug. 2017.

[85] M. Chi, Z. Sun, Y. Qin, J. Shen, and J. A. Benediktsson, "A novel methodology to label urban remote sensing images based on location-based social media photos," *Proc. IEEE*, vol. 105, no. 10, pp. 1926–1936, Oct. 2017.

[86] W. Tu *et al.*, "Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 12, pp. 2331–2358, Dec. 2017.

[87] W. Tu *et al.*, "Portraying urban functional zones by coupling remote sensing imagery and human sensing data," *Remote Sens.*, vol. 10, no. 1, p. 141, Jan. 2018.

[88] J. Cai, B. Huang, and Y. Song, "Using multi-source geospatial big data to identify the structure of polycentric cities," *Remote Sens. Environ.*, vol. 202, pp. 210–221, Dec. 2017.

[89] C. Chen, J. Yan, L. Wang, D. Liang, and W. Zhang, "Classification of urban functional areas from remote sensing images and time-series user behavior data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1207–1221, 2021.

[90] R. Cao *et al.*, "Deep learning-based remote and social sensing data fusion for urban region function recognition," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 82–97, May 2020.

[91] L. Wang, H. Fan, and Y. Wang, "Fine-resolution population mapping from International Space Station nighttime photography and multisource social sensing data based on similarity matching," *Remote Sens.*, vol. 11, no. 16, p. 1900, Aug. 2019.

[92] N. Zhao, W. Zhang, Y. Liu, E. L. Samson, Y. Chen, and G. Cao, "Improving nighttime light imagery with location-based social media data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2161–2172, Apr. 2019.

[93] Y. Song, B. Huang, Q. He, B. Chen, J. Wei, and R. Mahmood, "Dynamic assessment of $PM_{2.5}$ exposure and health risk using remote sensing and geo-spatial big data," *Environ. Pollut.*, vol. 253, pp. 288–296, Oct. 2019.

[94] U. Bhangale, S. Durbha, A. Potnis, and R. Shinde, "Rapid earthquake damage detection using deep learning from VHR remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 2654–2657.

[95] J. Yang, J. Zhang, and G. Huang, "A parallel computing paradigm for pan-sharpening algorithms of remotely sensed images on a multi-core computer," *Remote Sens.*, vol. 6, no. 7, pp. 6039–6063, Jun. 2014.

[96] J. Lu and B. Zhang, "An accelerated IHS transform fusion of remote sensing image data based on GPU," in *Proc. Int. Conf. Environ. Sci. Inf. Appl. Technol.*, vol. 1, 2009, pp. 492–496.

[97] J. Zhao and H. Zhou, "Design and optimization of remote sensing image fusion parallel algorithms based on CPU-GPU heterogeneous platforms," in *Proc. Int. Congr. Image Signal Process.*, vol. 3, 2011, pp. 1623–1627.

[98] R. Guerra, S. López, and R. Sarmiento, "A computationally efficient algorithm for fusing multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5712–5728, Oct. 2016.

[99] X. Xue, F. Xiang, H. Wang, and J. Peng, "A parallel fusion algorithm of remote sensing images based on wavelet transform," in *Proc. IEEE Int. Conf. High Perform. Comput. Commun., IEEE Int. Conf. Embedded Ubiquitous Comput.*, Nov. 2013, pp. 1113–1118.

[100] X. Xue, L. Guo, H. Wang, and F. Xiang, "A parallel fusion method of remote sensing image based on IHS transformation," in *Proc. Int. Congr. Image Signal Process.*, vol. 3, 2011, pp. 1600–1603.

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

[102] J. Liu, Y. Xue, K. Ren, J. Song, C. Windmill, and P. Merritt, "High-performance time-series quantitative retrieval from satellite images on a GPU cluster," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2810–2821, Aug. 2019.

[103] U. Bhangale, S. S. Durbha, R. L. King, N. H. Younan, and R. Vatsavai, "High performance GPU computing based approaches for oil spill detection from multi-temporal remote sensing data," *Remote Sens. Environ.*, vol. 202, pp. 28–44, Dec. 2017.

[104] D. Lunga, J. Gerrand, L. Yang, C. Layton, and R. Stewart, "Apache spark accelerated deep learning inference for large scale satellite image analytics," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 271–283, 2020.

[105] J. Sun *et al.*, "An efficient and scalable framework for processing remotely sensed big data in cloud computing environments," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4294–4308, Jul. 2019.

[106] J. Yan, Y. Ma, L. Wang, K.-K.-R. Choo, and W. Jie, "A cloud-based remote sensing data production system," *Future Gener. Comput. Syst.*, vol. 86, pp. 1154–1166, Sep. 2018.

[107] A. Kaur and S. K. Sood, "Deep learning based drought assessment and prediction framework," *Ecol. Informat.*, vol. 57, May 2020, Art. no. 101067.

[108] M. Mahdianpari *et al.*, "Big data for a big country: The first generation of Canadian wetland inventory map at a spatial resolution of 10-m using Sentinel-1 and Sentinel-2 data on the Google earth engine cloud computing platform," *Can. J. Remote Sens.*, vol. 46, no. 1, pp. 15–33, Jan. 2020.

[109] L. Nill, T. Ullmann, C. Kneisel, J. Sobiech-Wolf, and R. Baumhauer, "Assessing spatiotemporal variations of Landsat land surface temperature and multispectral indices in the Arctic Mackenzie Delta Region between 1985 and 2018," *Remote Sens.*, vol. 11, no. 19, p. 2329, Oct. 2019.

[110] W. Yu, J. Xie, and G. Li, "Cloud computing for environmental monitoring using multi-source earth observation data," in *Proc. Int. Conf. Agro-Geoinform.*, 2016, pp. 485–488.

[111] Q. Zou, G. Li, and W. Yu, "MapReduce functions to remote sensing distributed data processing-global vegetation drought monitoring as example," *Softw., Pract. Exper.*, vol. 48, no. 7, pp. 1352–1367, Jul. 2018.

[112] Q. Zou, G. Li, and W. Yu, "An integrated disaster rapid cloud service platform using remote sensing data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 5221–5224.

[113] R. Xiang, M. Sun, C. Jiang, L. Liu, H. Zheng, and X. Li, "A method of fast mosaic for massive UAV images," *Proc. SPIE*, vol. 9260, Nov. 2014, Art. no. 92603W.

[114] J. Sun and T. T. Vu, "Distributed and hierarchical object-based image analysis for damage assessment: A case study of 2008 Wenchuan earthquake, China," *Geomatics, Natural Hazards Risk*, vol. 7, no. 6, pp. 1962–1972, Nov. 2016.

[115] S. Li, Z. Zhu, H. Wang, and F. Xu, "3D virtual urban scene reconstruction from a single optical remote sensing image," *IEEE Access*, vol. 7, pp. 68305–68315, 2019.

[116] J. L. Huntington, K. C. Hegewisch, B. Daudert, C. G. Morton, and T. Erickson, "Climate engine: Cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding," *Bull. Amer. Meteorol. Soc.*, vol. 98, no. 11, pp. 2397–2409, 2017.

[117] B. DeVries, C. Huang, J. Armston, W. Huang, J. W. Jones, and M. W. Lang, "Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google earth engine," *Remote Sens. Environ.*, vol. 240, Apr. 2020, Art. no. 111664.

[118] H. Wang, J. Zhang, X. Liu, and X. Huang, "Parallel algorithm design for remote sensing image processing in the PC cluster environment," in *Proc. 18th Int. Conf. Geoinform.*, 2010, pp. 1–5.

[119] S. Dashti *et al.*, "Supporting disaster reconnaissance with social media data: A design-oriented case study of the 2013 Colorado floods," in *Proc. ISCRAM*, 2014, pp. 1–10.

[120] G. Cervone, E. Sava, Q. Huang, E. Schnebele, J. Harrison, and N. Waters, "Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study," *Int. J. Remote Sens.*, vol. 37, no. 1, pp. 100–124, Jan. 2016.

[121] T. Cooley *et al.*, "FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 3, Jun. 2002, pp. 1414–1418.

[122] N. Waters, "Social network analysis," in *Handbook of Regional Science*. Berlin, Germany: Springer-Verlag, 2021, pp. 725–740. [Online]. Available: https://link.springer.com/referenceworkentry/10.1007%2F978-3-662-60723-7_49

[123] Y. Ke, J. Im, J. Lee, H. Gong, and Y. Ryu, "Characteristics of Landsat 8 OLI-derived NDVI by comparison with multiple satellite sensors and *in-situ* observations," *Remote Sens. Environ.*, vol. 164, pp. 298–313, Jul. 2015.

[124] H. Xu, "A study on information extraction of water body with the modified normalized difference water index (MNDWI)," *J. Remote Sens.*, vol. 5, pp. 589–595, Sep. 2005.

[125] L. D. Dalcin, R. R. Paz, P. A. Kler, and A. Cosimo, "Parallel distributed computing using Python," *Adv. Water Resour.*, vol. 34, no. 9, pp. 1124–1139, 2011.

[126] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017.

ABOUT THE AUTHORS

Jun Li (Fellow, IEEE) received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004, the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

She is currently a Full Professor with Sun Yat-sen University, Guangzhou, China. Her main research interests comprise remotely sensed hyperspectral image analysis, signal processing, supervised/semisupervised learning, and active learning.

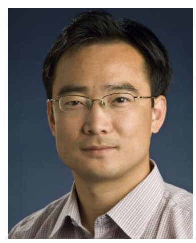
Prof. Li is also the Editor-in-Chief of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She has been a Guest Editor for several journals, including PROCEEDINGS OF THE IEEE and the *ISPRS Journal of Photogrammetry and Remote Sensing*.

Zhenjie Liu (Student Member, IEEE) received the B.S. and M.E. degrees from South China Agricultural University, Guangzhou, China, in 2016 and 2019, respectively. He is currently working toward the Ph.D. degree at Sun Yat-sen University, Guangzhou, China.

His research interests include the fusion of remote sensing and social media, and urban flood mapping.

Xinya Lei received the B.E. and M.E. degrees in information security from the China University of Geosciences (Wuhan), Wuhan, China, in 2017 and 2020, respectively, where she is currently working toward the Ph.D. degree.

Her research interests include parallel computing, remote sensing data processing, and knowledge graph.



Lizhe Wang (Fellow, IEEE) received the B.E. and M.E. degrees in electrical engineering and automation from Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the D.E. degree *(magna cum laude)* in applied computing science from the Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2007.

He is currently the "ChuTian" Chair Professor and the Dean of the School of Computer Science, China University of Geosciences (Wuhan), Wuhan, China. His research interests include remote sensing data processing, digital Earth, and big data computing.

Dr. Wang is a Fellow of the Institution of Engineering and Technology and the British Computer Society. He was a recipient of Distinguished Young Scholars of the National Natural Science Foundation of China, National Leading Talents of Science and Technology Innovation, and 100-Talents Program of the Chinese Academy of Sciences. He is an Associate Editor of *Remote Sensing*, the *International Journal of Digital Earth, ACM Computing Surveys*, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, and IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING.