Special Issue on Big Data From Space

Mihai Datcu, *Fellow, IEEE*, Jacqueline Le Moigne, Sveinung Loekken, Pierre Soille, and Gui-Song Xia, *Senior Member, IEEE*

The deluge of data from space sensors counting hundreds of Terabytes per day needs to be converted into meaningful information, largely impacting science, technology, and overall the socio-economic-environmental triangle. Multispectral and microwave Earth Observation (EO) sensors are unceasingly streaming millions of samples per second, which must be interpreted to mine physical parameters to understand Earth spatio-temporal patterns and phenomena. Space telescopes covering from Gamma to radio waves spectra collect billions of measurements. Position, navigation, telemetry, and space data relays are other examples; the list is however much longer. The recent multiplication of open access initiatives to Big Data from Space is giving momentum to the field by widening substantially the spectrum of scientific communities and users, as well as awareness among the public, while offering new benefits at all levels from individual citizens to the whole society.

Big Data from Space is a very broad and highly interdisciplinary area. It encompasses the entire chain from sensors to scientists or application results. High complexity sensor data require in depth knowledge of the physical detection principles. New sensing paradigms such as computational imaging and sensor networks involve on-board intelligence as well as methods of compressive sensing and data compression. Ultra fast satellite optical communication relays enable Big Data to be received on Earth. Data access, catalogs, data delivery, and data quality, become very specific challenges. Storage, data indexing, data analytics and mining, knowledge discovery, visualization, the overall content extraction are very deep and particularly specialized topics of data science. Benchmark, test and validation procedure, and metrics of information quality have to be specialized as well. And last but not least, the data analytics learned lessons are used for the design of next generation of space sensors and systems.

We are very indebted to the academia, research, and industry communities having contributed with their results. We would like to acknowledge the scientific committee of the *Big Data from Space* BiDS'17 conference co-organised by the European Space Agency (ESA), the Joint Research Center of the European Commission (JRC), and the European Union Satellite Centre (SatCen) for their sustained efforts to achieve this special issue.

Following a detailed and rigorous review process, 14 articles have been selected out of 48 submissions for this special issue.

The first article, *Optical Compressive Imaging Technologies for Space Big Data*, considers the Big Data problematic at its source. It presents solutions for very particular space data acquisition constraints such as limited power, memory, computation, or data transmission resources and propose novel sensors concepts based on Compressive Sensing methods for space science, planetary exploration, or Earth Observation.

The subsequent two articles are concerned with telemetry data. On one hand, *Multiple Testing for Outlier Detection in Space Telemetries* is about on-board satellite measurements, like sensor temperature, voltage, or relative position of equipment during the space mission life. Telemetries accumulate terabytes of data, which have to be analysed to detect anomalies. The introduced techniques are detecting automatically anomalous events and are self-adaptive to the data and to the nature of anomalies. On the other hand, *Innovative Approach for Permanent Multipurpose Module Data Processing and Analytics* elaborate on the data analysis to provide upload cargo to increase the storage volume at the International Space Station (ISS) helping the utilization, maintenance and the crew logistic. Advanced machine learning methods are used to predict missing sensor measurements, as cabin fan assembly, pressure, heaters, or light assembly.

Two further articles fall in the domain of astronomical observations. The first is titled *Convolutional Neural Networks for Spectroscopic Redshift Estimation on Euclid Data*. The Euclid satellite will carry photometric and near-infrared sensors measuring the global parameters of the Universe to an exceptional precision. The objective is to achieve a better understanding of the nature of Dark Energy. The article presents a deep learning paradigm for the estimation of the redishift a fundamental parameter for the measurement of galaxy distances. The second, *ASTROIDE: A Unified Astronomical Big Data Processing Engine over Spark*, presents a distributed data server specialized for astronomical data, where astronomers can explore and manage large, highly complex data sets. The system is applied to the GAIA data which is producing a 3-D map of our Galaxy.

- *Mihai Datcu is with the German Aerospace Center DLR, 51147 Cologne, Germany and also with the University Politehnica Bucharest UPB, 060042 București, Romania. E-mail: mihai.datcu@dlr.de.*
- *Jacqueline Le Moigne is with NASA Earth Science Technology Office, Washington, DC 20546. E-mail: jacqueline.lemoigne@nasa.gov.*
- *Sveinung Loekken is with the European Space Agency ESA/ESRIN, 00044 Frascati, RM, Italy. E-mail: sveinung.loekken@esa.int.*
- *Pierre Soille is with the European Commission Joint Research Centre JRC, Ispra, Italy. E-mail: pierre.soille@jec.europa.eu.*
- *Gui-Song Xia is with the Wuhan University, Wuhan, Hubei 430072, China. E-mail: guising.xia@whu.edu.cn.*

Digital Object Identifier no. 10.1109/TBDATA.2020.3015536

The next three articles are approaching the thematic of optical satellite based Earth observation. *Beyond the Patchwise Classification: Spectral-Spatial Fully Convolutional Networks for Hyperspectral Image Classification*, elaborates a spectral-spatial fully convolutional network for hyperspectral observations. The classification is performed by an end-to-end, pixel-to-pixel paradigm, without the need of image partition in patches. *Exploiting Deep Features for Remote Sensing Image Retrieval: A Systematic Investigation*, proposes answers for the three main elements of the very high-resolution satellite image retrieval: the feature extraction, the similarity metrics and the relevance feedback. The article is a guide for the research and practice in the areas of content-based satellite image retrieval. *Mining Deep Semantic Representations for Scene Classification of High-Resolution Remote Sensing Imagery* presents a method based on the probabilistic latent semantic analysis (pLSA) applied to image features extracted by a deep convolutional neural network (CNN). The method minimizes the redundancy of the extracted features and is supported by multi-scale spatial information analysis to increase accuracy.

Satellite image time series is addressed by the next article titled *A Time Series Mining Approach for Agriculture Area Detection*. The time series enable long temporal observations at large spatial scale. The present article proposes a framework based on an Active Learning approach for time series classification using less than 1 percent of the data as training data set. The method is highly impactful since the complexity of the satellite image time series is huge.

Three articles are then focusing on the methods for Big Synthetic Aperture Radar (SAR) Data. SAR is one of the most preeminent Earth observation technologies, it is day–night and cloud cover independent, providing extremely valuable information on the land cover, terrain height or millimeter precision deformations. However, SAR data analysis remains one of the major Big Data challenge due to its volume and high complexity. The first article on this topic, *Mosaicking Copernicus Sentinel-1 Data at Global Scale*, presents approaches for distribution discretization together with a false color composition and image rendering and mosaicking for Sentinel-1 dual polarization SAR products. The methods ensure fast automated processing, incremental adjustment and information extraction at global scale. The second, *National Scale Surface Deformation Time Series Generation through Advanced DInSAR Processing of Sentinel-1 Data within a Cloud Computing Environment*, presents an automatic Differential Interferometric SAR pipeline implemented within the Amazon Web Services (AWS) Cloud Computing. The method provides measurements of surface deformation time series and the velocity maps with millimeter to centimeter precision at very large geographical scales. The third, *Cloud Approach to Automated Crop Classification Using Sentinel-1 Imagery*, introduces a machine-learning paradigm for crop classification implemented in a cloud based workflow to face the Big Data challenges.

The special issue is concluded with two articles presenting interdisciplinary results in actual applications, which require physical parameters estimation of the observed scenes. The first, *Exploring Spring Onset at Continental Scales: Mapping Phenoregions and Correlating Temperature and Satellite-based Phenometrics*, analyses long mutitemporal observations at high-spatial resolution such as temperature-based phenological indices and land surface phenological metrics derived from satellite images to assess the "green-wave" of spring vegetation. The second, *Deep Neural Network-based Impacts Analysis of Multimodal Factors on Heat Demand Prediction*, is presenting a neural network approach for heat demand estimation based on the direct solar irradiance and wind speed.

We hope this special issue will bring the challenges of Big Data from space to the attention of the scientific community, thus new solutions and applications will be stimulated.

We also would like to take this opportunity to thank all the authors for making this special issue possible.

Acknowledgments

The Guest Editors would like to thank the IEEE Computer Society for its support in publishing this special issue; the reviewers of the papers, who played the most important role in the final selection and quality of the contributions published in this special issue; and, finally, of the IEEE staff for their valuable assistance in all phases of the preparation of this special issue.

Mihai Datcu (Fellow, IEEE) received the MS and PhD degrees in electronics and telecommunications from the University Politechnica Bucharest UPB, Romania, in 1978 and 1986, respectively. In 1999 he received the title Habilitation à diriger des recherches, in computer science from University Louis Pasteur, Strasbourg, France. Currently he is senior scientist and data intelligence and knowledge discovery research group leader with the Remote Sensing Technology Institute (IMF) of the German Aerospace Center (DLR), Oberpfaffenhofen, and a professor with the Department of Applied Electronics and Information Engineering, Faculty of Electronics, Telecommunications and Information Technology, UPB. From 1992 to 2002 he had a longer Invited professor assignment with the Swiss Federal Institute of Technology, ETH Zurich. From 2005 to 2013 he has been professor holder of the DLR-CNES Chair at ParisTech, Paris Institute of Technology, Telecom Paris. His interests are in data science, machine learning and artificial intelligence, and computational imaging for space applications. He is involved in Big Data from Space European, ESA, NASA, and national research programs and projects. He is a member of the ESA Big Data from Space Working Group. He received the Best Paper Award in 2006, IEEE Geoscience and Remote Sensing Society Prize, in 2008, the National Order of Merit with the rank of Knight, for outstanding international research results, awarded by the President of Romania, and in 1987 the Romanian Academy Prize Traian Vuia for the development of SAADI image analysis system and activity in image processing. He is holder of a 2017 Blaise Pascal International Chair of Excellence in Data Science at CEDRIC, CNAM.

Jacqueline Le Moigne received the BS and MS degree in mathematics, and the PhD degree in computer science from the University Pierre and Marie Curie (UPMC), Paris, France, now known as Sorbonne Université. She has been with NASA, since 1998, currently she is a manager of the Advanced Information Systems Technology (AIST) Program of the NASA Science Mission Directorate (SMD) Earth Science Technology Office (ESTO). The AIST Program focuses on advanced computer science and information systems technologies that reduce the risk, cost, size, and development time of future Earth science space-based and ground-based data and information systems; including the design of novel observation systems and of analytic center frameworks. Previously, she was an assistant chief for Technology in the Software Engineering Division at NASA Goddard, also working with NASA Space Technology Research Grants Program. Her PhD thesis dealt with biomedical imagery, specifically the classification of blood platelets. She has published more than 180 journal, conference publications and book chapter articles, including more than 25 journal papers; she co- authored an edited book on "Image Registration for Remote Sensing" and holds a Patent on this topic. She has been PI of several projects focused on Distributed Spacecraft Missions. She has been associate editor for the *IEEE Transactions on Geoscience and Remote Sensing* and for the *Journal Pattern Recognition*. She is a NASA Goddard senior fellow, and was a program evaluator for the Accreditation Board in Engineering and Technology and a NATO Science for Peace and Security Committee Panel member. She has been recipient of a NASA Exceptional Service Medal and of the Goddard Information Science and Technology Award.

Sveinung Loekken received the BSc degree from the University of Oslo, in 1991, and the MSc degree from Texas A&M, in 1993. He has been with the European Space Agency, since 1994, working on various aspects of technology research and development on the European Payload Data Ground Segment, with the primary objective of boosting the uptake and societal impact of satellite-based EO. In this role he conceived, set up and conducted the Thematic Exploitation Platforms initiative, which defined the ESA approach to a new data exploitation paradigm. In 2013 he received, together with ESA and NASA teams, a NASA Award for the successful collaborative effort with the MERIS/MODIS-SeaWiFS data exchange. In 2017 he co-founded the ESA Φ-lab, an innovation laboratory in the Earth Observation programme directorate working in partnership with research institutions and industry with the aim of accelerating the development and uptake of new technologies in Earth Observation. In his current role he focuses on making sense of the methodological revolution determined by Artificial Intelligence techniques meeting traditional earth science in the context of EO, and called and conducted the first AI for Earth Observation (AI4EO) workshop in ESRIN, in 2018. Sveinung co-chaired the Big Data From Space conference, in 2019 (BIDS '19) and will co-chair the same conference, in 2021. He is the guest editor for the 2020 "Machine and Deep Learning for Earth Observation Data Analysis" special issue of MDPI Remote Sensing.

Pierre Soille received the engineering degree from the University of Louvain, Louvain-la-Neuve, Belgium, in 1988, and the PhD degree in agronomical sciences from the same university in collaboration with the Centre for Mathematical Morphology of the Ecole des Mines de Paris, France, in 1992, and the Habilitation à Diriger des Recherches (HDR), in computer science, in 1997 from Montpellier II University, France. After completion of his PhD on mathematical morphology applied to digital elevation models and satellite image analysis, he pursued research on image analysis and mathematical morphology at the CSIRO Mathematical and Information Sciences in Sydney, Australia, the Centre for Mathematical Morphology of the Ecole des Mines de Paris, France, and the Pattern Recognition Department of the Fraunhofer-Institut IPK in Berlin, Germany. During the period 1995–1998 he was an assistant professor with the Ecole de Mines d'Alès and EERIE, Nîmes, France. In 1999, he was appointed as a senior researcher with the Biotechnology and Biological Sciences Research Council, United Kingdom. He has been with the Joint Research Centre (JRC) of the European Commission, Ispra, Italy, since December 1999. His research activities are dedicated to the automatic information retrieval from high to very high resolution satellite image data for the generation of new and enhanced global datasets. He is currently leading the JRC Big Data Analytics project and is responsible for the JRC Big Data Platform that is distilling multi-petabyte scale data sources into policy relevant insights and foresight. He has been programme co-chair of Big Data from Space conference series, organised by the European Space Agency, the European Commission's JRC and the European Union Satellite Centre, since its inception, in 2014. He has authored and co-authored numerous and highly cited image analysis, pattern recognition, and machine learning algorithms and methods published in peer-reviewed journal and conference papers as well as book chapters. He has also edited various books and is author of a reference monograph on morphological image analysis.

Gui-Song Xia (Senior Member, IEEE) received the PhD degree in image processing and computer vision from CNRS LTCI, Telecom ParisTech, Paris, France, in 2011. From 2011 to 2012, he has been a post-doctoral researcher with the Centre de Recherche en Mathematiques de la Decision (CEREMADE), CNRS, Paris-Dauphine University, Paris, for one and a half years. He is currently working as a full professor in computer vision and photogrammetry at Wuhan University. He has also been working as visiting scholar at DMA, Ecole Normale Superieure (ENS-Paris) for two months in 2018. His current research interests include mathematical modeling of images and videos, structure from motion, perceptual grouping, and remote sensing image understanding. He serves on the Editorial Boards of the journals *Pattern Recognition*, *Signal Processing: Image Communications*, and *EURASIP Journal on Image & Video Processing*.