

PROCESSING BIG REMOTE SENSING DATA FOR FAST FLOOD DETECTION IN A DISTRIBUTED COMPUTING ENVIRONMENT

A. Olasz^{a*}, D. Kristóf^a, B. Nguyen Thai^b, M. Belényesi^a, R. Giachetta^b

^a Dept. of Geodesy, Remote Sensing and Land Administration, Government Office of the Capital City Budapest,
5. Bosnyák sqr. Budapest, 1149 Hungary (olasz.angela, kristof.daniel, belenyesi.marta)@bfkh.gov.hu

^b Dept. of Cartography and Geoinformatics, Eötvös Loránd University (ELTE), 1/A Pázmány Péter sétány, Budapest, 1117 Hungary,
(ntb, groberto)@inf.elte.hu

Commission ICWG IV/II

KEY WORDS: Distributed Computing, Geospatial Big Data, Cloud Computing, Fast Flood detection, Big Earth Observation Data

ABSTRACT:

The Earth observation (EO) missions of the space agencies and space industry (ESA, NASA, national and commercial companies) are evolving as never before. These missions aim to develop and launch next-generation series of satellites and sensors and often provide huge amounts of data, even free of charge, to enable novel monitoring services. The wide geospatial sector is targeted to handle new challenges to store, process and visualize these geospatial data, reaching the level of Big Data by their volume, variety, velocity, along with the need of multi-source spatio-temporal geospatial data processing. Handling and analysis of remote sensing data has always been a cumbersome task due to the ever-increasing size and frequency of collected information. This paper presents the achievements of the IQmulus EU FP7 research and development project with respect to processing and analysis of geospatial big data in the context of flood and waterlogging detection.

1 INTRODUCTION

Geospatial Big Data has an established and well-known definition (Lee and Kang 2015, Li et al. 2016, Olasz A. and Nguyen Thai B and Kristóf D. 2016). Huge datasets are available but up to now there are only limited automatic procedures for processing; thus, as a huge amount of data remains unprocessed, a wealth of information is latent in many datasets. However, a number of distributed data processing solutions exist, but have primarily been focused on processing simple structured documents rather than complex geospatial data. This paper presents a multi-sensor remote sensing processing solution within the IQmulus Project to realize fast flood and waterlogging detection in a cloud environment.

2 THE IQMULUS PROJECT

The four-year IQmulus project (www.iqmulus.eu) targeted the optimized use of large, heterogeneous geospatial datasets (“Geospatial Big Data”) for better decision making through a high-volume fusion and analysis information management platform. An end-to-end involvement of users was ensured through the implementation of three concrete “test beds” (Maritime Spatial Planning & Land Applications for Rapid Response and Territorial Management) to show the benefits of the approach (Kristóf et al. 2014). In this study we are focusing on the Land Application for Rapid Response test bed, which includes the showcase of Detection and characterization of flood and waterlogging. The project was successfully concluded in October 2016; this paper documents the latest results in terms of processing services and system architecture.

3 IMPLEMENTATION DETAILS

According to IQmulus specification, services should run on distributed environment. The Apache Hadoop (<http://hadoop.apache.org/>) open source framework has been selected as a basis for architecture development. IQmulus developed functional and domain processing services in order to maximize the use of geospatial big data and provide support for analysing quickly changing environmental conditions. The platform (together with a web user interface) supports the processing of both vector data and remotely sensed images using a wide variety of algorithms. The development has been focusing on the usage of open source software.

4 RESULTS AND CONCLUSION

Our presentation will focus on the solutions on the developments and results related to preprocessing and classification of satellite images for the detection of flooded and waterlogged areas. Quick response in such a situation is critical, thus the processing time of the satellite image series in a given time is of major importance. The developed solution provides a higher level of automation via smarter algorithms supporting the process of multi-sensor remote sensing resources (SPOT, Landsat, and Sentinel 2). Our presentation will include description of the newly available operative processing workflow.

ACKNOWLEDGEMENTS

This research is co-funded by the project “IQmulus” (A High-volume Fusion and Analysis Platform for Geospatial Point Clouds, Coverages and Volumetric Data Sets) funded from the 7th Framework Programme of the European Commission, call identifier FP7-ICT-2011-8.

REFERENCES

Lee J.-G. and Kang M. 2015. Geospatial Big Data: Challenges and Opportunities, *Big Data Research* Vol. 2, no. 2, pp. 74–81.

Li S. , Dragicevic S., Anton F., M. Sester, S. Winter, A. Coltekin, C. Pettit, B. Jiang, J. Haworth, A. Stein, and Cheng T. 2015. Geospatial Big Data Handling Theory and Methods: A Review and Research Challenges, pp. 2-19.

Kristóf, D., Giachetta, R., Olasz, A., Nguyen Thai B. 2014. Big geospatial data processing and analysis developments in the IQmulus project, *Proceedings of the 2014 Conference on Big Data from Space* (BiDS'14), pp. 214-217

Olasz A., Nguyen Thai B., and Kristóf D., 2016. A new initiative for Tiling, Stitching and Processing Geospatial Big Data in Distributed Computing Environments, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. Volume III-4, pp. 111–118,