38. Wissenschaftlich-Technische Jahrestagung der DGPF und PFGK18 Tagung in München – Publikationen der DGPF, Band 27, 2018

A Service Architecture for Processing Big Earth Data in the Cloud with Geospatial Analytics and Machine Learning

WOLFGANG GLATZ & THOMAS BAHR[1]

Abstract: The Geospatial Services Framework (GSF) brings together data, geospatial analytics, and computing power in the cloud to enable the deployment of applications. The service architecture given by GSF is used by web clients to access and analyze on-demand remotely sensed data as well as for the automated, permanent processing of big geospatial data. GSF can be integrated in any public or internal server environments.
GSF is based on the concept of service engines and their workers. Harris provides the ready-to-use ENVI/IDL/SARscape/Machine Learning service engines. Available ENVI analytics include feature extraction, object identification, change detection, and classification. A specific machine learning algorithm for spectral-based land cover mapping is the Softmax Regression classifier. Harris machine learning contains deep learning capabilities, which focus on object recognition within scenes. They are designed for the unique characteristics of space- and airborne imagery of multiple modalities, and point cloud data sets.

1 Introduction

A continually increasing, massive amount of geospatial data, i.e. Big Earth Data, from different sources (commercial satellite constellations and small satellites, drones) and modalities (optical: Pan, RGB, MSI, HSI; SAR; LiDAR), enforces the automation of data processing. New tools and technologies are needed for hosting and managing distributed data processing in a high-performance computing environment within an enterprise or in the cloud.
The Geospatial Services Framework (GSF) brings together data, geospatial analytics, and computing power in the cloud to enable the deployment of applications, which solve problems at scale across industries.

2 Geospatial Services Framework (GSF)

GSF is based on the concept of service engines and their workers (HARRIS GEOSPATIAL SOLUTIONS 2017). Each worker uses multiple CPUs for parallel processing, the workers are processes executed in a dynamically scalable cluster of machines. Harris provides the ready-to-use ENVI/IDL/SARscape/Machine Learning service engines (BAHR & OKUBO 2013). In addition, customers may implement their own engines. The service architecture given by GSF is used by web clients to access and analyze on-demand remotely sensed data as well as for the automated, permanent processing of big geospatial data (Fig. 1).

[1] Harris Geospatial Solutions GmbH, Talhofstraße 32a, D-82205 Gilching,
E-Mail: [Wolfgang.Glatz, Thomas.Bahr]@harris.com

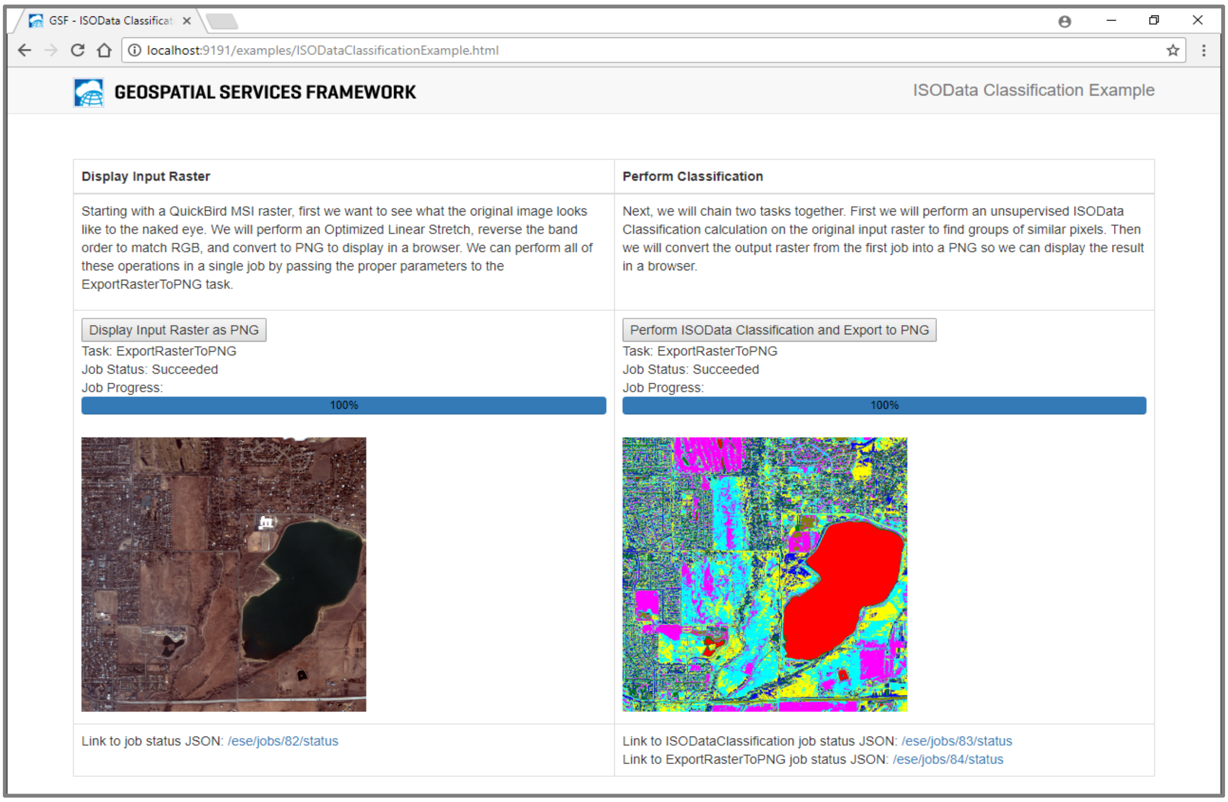

Fig. 1: GSF example: Web client showing the result of a land use classification processed in the cloud

GSF can be integrated in any public or internal server environments, such as Amazon Web Services (AWS), Microsoft Azure, or the Google Cloud Platform (Fig. 2).

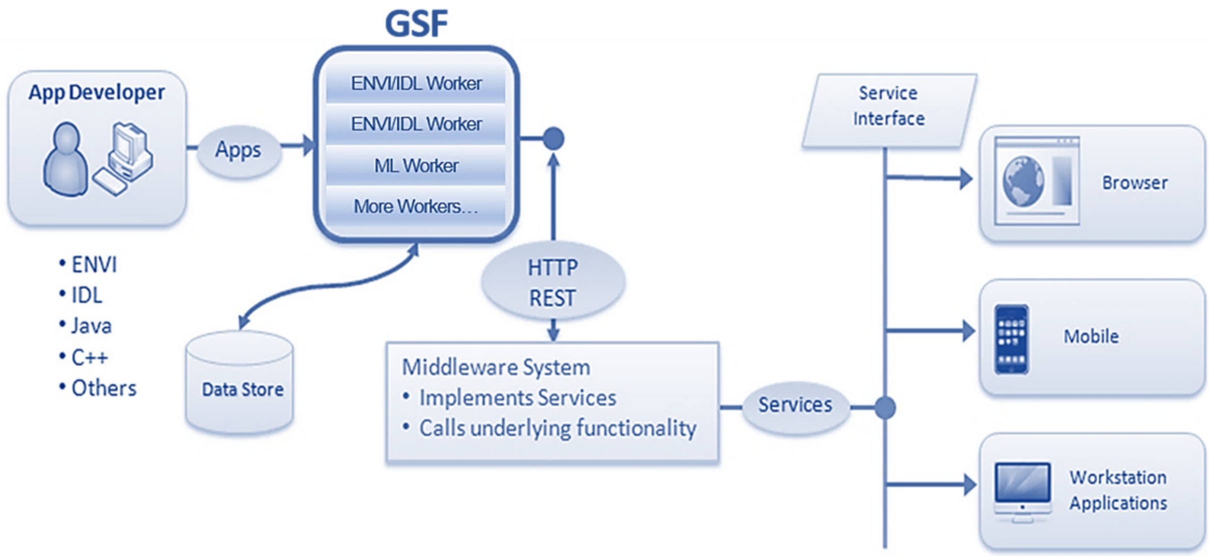

Fig. 2: GSF: A scalable framework for geospatial web services, using highly parallelized clusters of ENVI/IDL workers, Machine Learning (ML) workers, and other workers (processors)

Developers may use GSF to easily publish custom algorithms for the hosted service engines. These processing workflows can then be shared across the enterprise or cloud. For analysis of remotely sensed data, developers can resort to the full width of ENVI software analytics.
One of the essential features of GSF is the adaptation to varying utilization. On demand, the operator of the service architecture can add additional parallel workers (scalability) and additional geospatial workflows, if the amount of data and the number of service requests from any clients increases (Fig. 3).

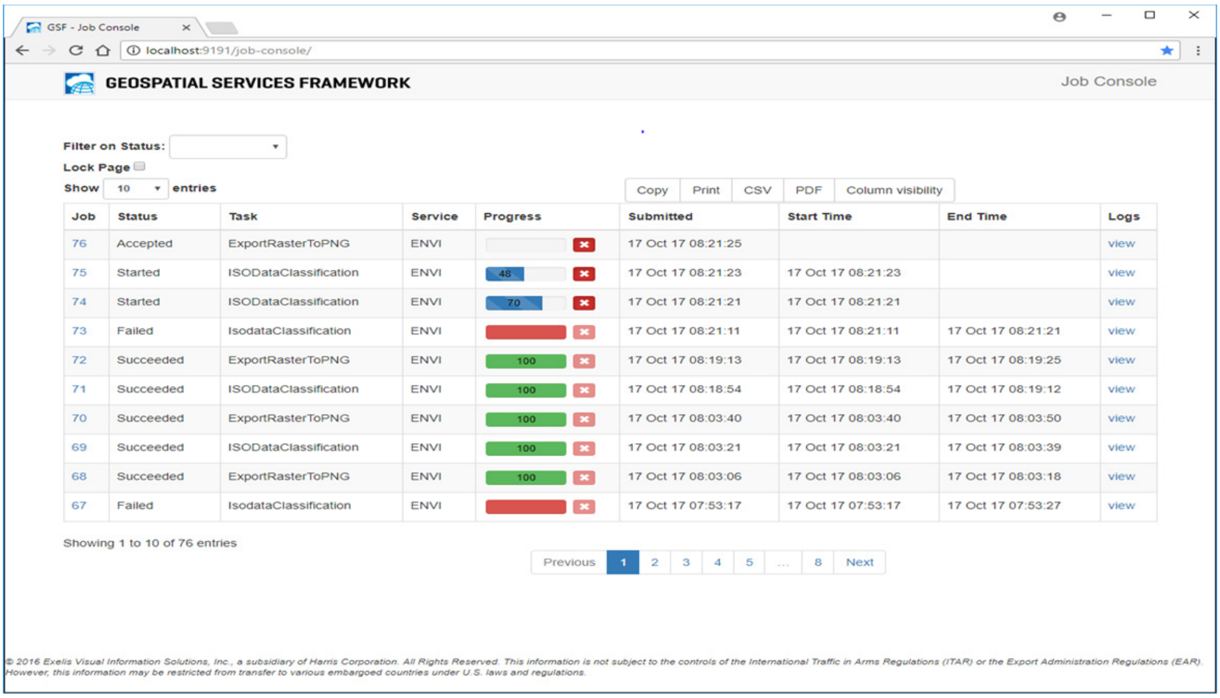

Fig. 3: GSF Job Console: State list of the geospatial analytics jobs executed by the GSF

3 ENVI Software Analytics

ENVI combines image processing and analysis technology to derive detailed information from all geospatial data, i.e. optical imagery, SAR, and LiDAR. Available analytics include feature extraction, classification, object identification, change detection, and more (HARRIS GEOSPATIAL SOLUTIONS 2018). A specific machine learning algorithm for spectral-based land cover mapping is the Softmax Regression classifier (WOLFE et al. 2017). It can be created and trained on a reference dataset using spectral and spatial information and then be applied to similar data multiple times (Fig. 4). Implemented in a classification framework, it provides a flexible approach to customize a classification process.
All described geoprocessing tools are capable of being fully integrated with ArcGIS® for Server from ESRI via a Python client library.

Fig. 4: Softmax Regression classifier: Merge of two classification images (bottom). The classifier was trained on one attribute image (top right) and then applied to a similar attribute image (top left)

4 Deep Learning Capabilities

Harris machine learning contains deep learning capabilities, which focus on object recognition within scenes.

Fig. 5: Training of the Deep Learning network for recognition of commercial airplanes with negative (left) and positive (right) training chips

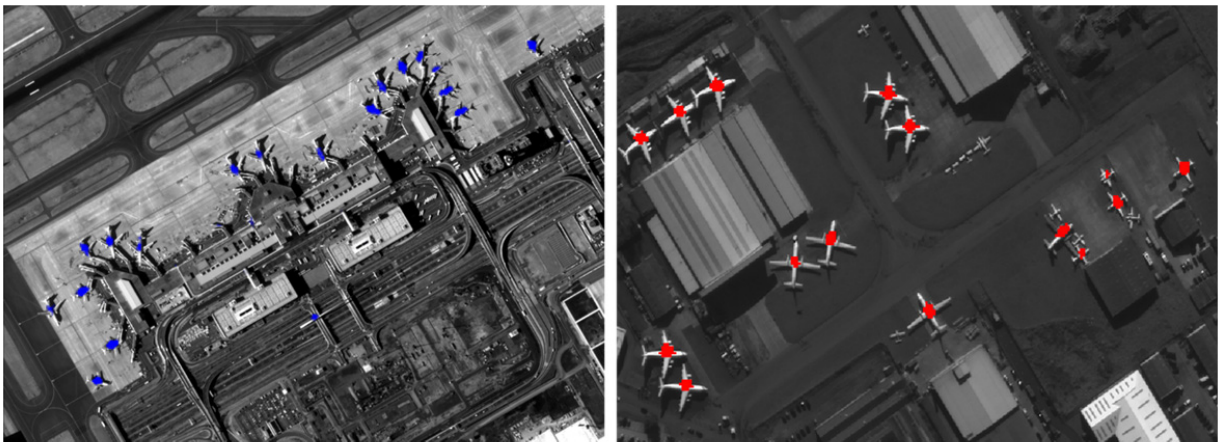

Fig. 6: Application of the trained network to different data for plane recognition: Ikonos (left), WorldView-3 (right)

They are designed for the unique characteristics of space- and airborne imagery of multiple modalities, as well as point cloud data sets. Successful sample applications, for instance on Pan imagery, included the detection of airplanes, storage tanks, cooling towers, athletic fields, paved roads, overpasses, and tollbooths (Fig. 5, 6).

5 Conclusion

The Geospatial Services Framework (GSF) brings together data, geospatial analytics, and computing power in the cloud for processing remotely sensed data with geospatial analytics and machine learning. In particular, the Harris modules for Deep Learning can be applied for automated object recognition processes and executed as GSF services.
Overall, GSF is a substantial contribution to operational, spatio-temporal analytics of Big Earth Data.

6 References

BAHR, T. & OKUBO, B., 2013b: A New Cloud-based Deployment of Image Analysis Functionality. GI_Forum 2013, Jekel, T., Car, A., Strobl, J. & Griesebner, G. (Eds.), Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach, ÖAW Verlag, Wien, 243-250.

HARRIS GEOSPATIAL SOLUTIONS (Ed.), 2017: Geospatial Services Framework Documentation. Geospatial Services Framework v2.1.

HARRIS GEOSPATIAL SOLUTIONS (Ed.), 2018: ENVI Documentation, ENVI v5.5.

STANFORD UNIVERSITY (Ed.), 2013: Unsupervised Feature Learning and Deep Learning. http://ufldl.stanford.edu/wiki/ index.php/Softmax_Regression (updated April 2013, accessed September 2016).

WOLFE, J., JIN, X., BAHR, T. & HOLZER, N., 2017: Application of Softmax Regression and its Validation for Spectral-Based Land Cover Mapping. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., **42**(1/W1), 455-459.