FEATURE EXTRACTION AND CONTENT BASED IMAGE RETRIEVAL ON EO ARCHIVES

Luis Veci[(1)], Jun Lu[(1)], Norman Fomferra[(2)], Ralf Quast[(2)], Kerstin Stelzer[(2)], Francesca Bovolo[(3)], Lorenzo Bruzzone[(3)], Begüm Demir[(3)], Michele Iapaolo[(4)], Pier Giorgio Marchetti[(4)]

[(1)]Array Systems Computing Inc., [(2)]Brockmann Consult GmbH, [(3)]University of Trento, [(4)]ESA ESRIN

1. INTRODUCTION

The aim of the Primitive Feature Extraction (PFA) project[1] is to demonstrate that feature extraction and analysis of EO data makes it possible to exploit the large volumes of various types of satellite data more efficiently, and through this to foster the scientific analysis of mission wide data sets. The project addresses emerging methods and tools for data product features and information extraction, in view of possible implementations for enriching data description and easing the use of archived data[1]. For this purpose, a number of EO data exploitation scenarios are being implemented to test the possibility of systematically extracting EO product features and demonstrate the usefulness of the performed feature extraction in the data selection process, both for the scientist and for the users of EO products. The project has selected to implement feature extractors for forest mapping, urban mapping, algal blooms and flood mapping. The entire EO archives from optical and SAR sensors will be processed using the ESA toolboxes BEAM and NEST[2] on the Calvalus Hadoop cluster[3]. Products are searched using content based image retrieval (CBIR) on single images and time series images[4].

2. APPROACH

Primitive, application-relevant features is extracted from large quantities of EO data and collected in a database. This database of primitive features is then exploited by performing analyses on the extracted features, and thus it serves as an auxiliary dataset complementary to the EO data archives it was generated from. This Auxiliary Enriched Database is complementary to the EO data archives and their catalogues and will be used to provide services that can exploit the existing EO data archives in a thematic manner using semantic queries.

Extracted features will be employed in three main scenarios: i) content based image retrieval; ii) content based time-series retrieval; and iii) unsupervised classification with kernel methods.

The first and second scenarios are devoted to fast and effective content based image retrieval on single images and image time-series, respectively. Query data can be an EO image, parameter values (such as a threshold values), temporal-trend of a time series or a step-change in bi-temporal images. Once the query is fixed, efficient approaches are required to retrieve from large EO data archives images (or time series) that match the query. Active-learning-based methods [5] are considered in the context of relevance feedback [5, 6] for both scenarios in order to efficiently exploit interaction with the user. The classification stage of the content based retrieval is based on a Support Vector Machine (SVM) classifier, which is nonparametric (and thus suitable for any kind of data) and widely recognized as effective [7].

The third scenario aims to search for the similar images to the query image without using any prior information on the archive. Here, the kernel k-means [8] method is used for unsupervised classification.

[1] THe PFA project is funded by European Space Agency (ESA)

3. APPLICATIONS

The feature extraction component of the project focuses on encoding relevant primitive features into an auxiliary database from the following applications:

From optical data

- Burned area mapping
- Algal bloom detection
- Intertidal flat selection

From SAR data

- Deforestation mapping
- Urban change mapping
- Flood mapping

4. SOFTWARE

The BEAM and NEST Toolboxes will be used for both client user interfaces as well as processors on the server system. The feature extraction software is developed using the ESA BEAM toolbox for optical data and ESA NEST for SAR data. The frontends for the content based image retrieval will be implemented as extensions to BEAM VISAT and NEST DAT. All software will be open-source and users of BEAM and NEST will be invited to install and use the new extension modules.

Bulk processing is done on the Calvalus EO Data Processing System that has been developed by Brockmann Consult within an LET-SME study under contract to ESA. The goal of Calvalus is the development of an efficient bulk-processing and analysis system for large amounts of EO satellite data. Calvalus is powered by Apache Hadoop, a framework that allows for the distributed processing of large data sets across clusters of computers using the simple map-reduce programming model and a robust distributed file system. The system will process the EO archive, apply pre-processing and feature extractors using BEAM and NEST, and generate the Auxiliary Enriched Database.

5. RESULTS

A prototype feature extraction system is in place and the CBIR based on Support Vectors Machines is currently being benchmarked for algal bloom detection based on MERIS data and urban area detection on ASAR data. An examples of the retrieved image patches is shown in figures 1 and 2.

Figure 1 MERIS Algal Bloom Search Results

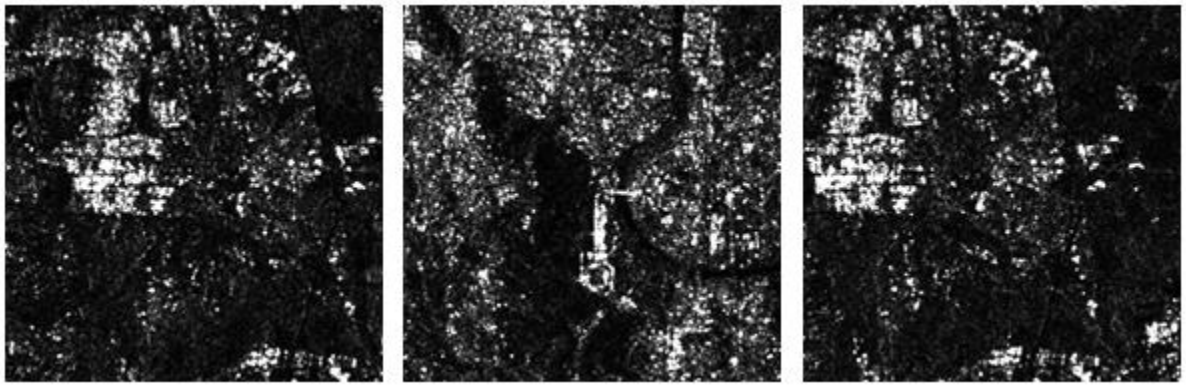

Figure 2 ASAR Urban Area Detection Search Results for High Urban

Initial results show searching by features using CBIR can produce relevant images. Furthermore, the use of active learning can reduce the number of training iterations on the SVM.

Further validation and applications will be available by mid-2014.

6. REFERENCES

[1] M. Datcu, S. D'Elia, R. L. King, and L. Bruzzone, "Introduction to the special section on image information mining for earth observation data," IEEE Trans. Geosci. Rem. Sen., vol. 45, no. 4, pp. 795–798, 2007.

[2] https://earth.esa.int/web/guest/software-tools

[3] http://www.brockmann-consult.de/calvalus/

[4] L. Bruzzone, M. Marconcini, U. Wegmuller, A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," IEEE Trans. Geosci. Rem. Sens., vol. 25, no. 13, 2004, pp. 1491-1500.

[5] B. Demir, L. Bruzzone, "An Effective Active Learning Method for Interactive Content-Based Retrieval in Remote Sensing Images", International Conference on Geoscience and Remote Sensing Symposium, Melbourne, Australia, 2013..

[6] M. Ferecatu, N. Boujemaa, "Interactive Remote-Sensing Image Retrieval Using Active Relevance Feedback", IEEE Trans. on Geosci. and Rem. Sensing, vol. 45, no. 4, pp. 818-826, 2007.

[7] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 43, no. 6, pp. 1351–1362, 2005.

[8] I.R. Zhang, A.I. Rudnicky, "A Large scale clustering scheme for kernel k-means," IEEE Int. Conf. on Pattern Recog., pp. 289-292, 2002.