Big data science for Earth observation: large scale visual analytics and knowledge discovery

Mihai Datcu
Motivation

Big Earth Observation Data

ENVISAT

TerraSAR-X and TanDEM-X

Copernicus Sentinel 1 and 2

The DLR - DFD EO Digital Library
Multispectral sensors: Sentinel 2

RGB Image

IR Image
SAR vs. optical images

TerraSAR-X Single Look Complex Image

Amplitude and Phase

Real and Imaginary components
The EO context

Earth Observation data is always used jointly with information extracted from other sources such as GIS, in-situ observations, or maps.

The goal is the exploration of these data and the timely delivery of focused information and knowledge in a simple understandable format.

The data volumes, their heterogeneity, unstructured nature and their complexity have become a major Big EO Data challenge for all applications.
EO images: multisensory, e.g. MS, SAR, altimeter, etc.

These are multidimensional signals, acquired by sensors or instruments

Sensor data carry physical meaning, radiation level, wavelength, etc.

They are measuring land, ocean, or atmospheric parameters

The VHR EO images observe detailed spatial structures and objects

Satellite Image Time Series observe evolution processes over long period of time.

An important particularity of EO images should be considered, is their “instrument” nature, i.e. they are sensing physical parameters, and they are often sensing outside of the visual spectrum.
Big EO Data Analytics

• The today techniques, methods, and tools, for automated data analysis are insufficient for the analysis and information extraction from EO data sources.

• A new goal has become the gathering of the user’s interest, together with the transformation of the data into reduced information and knowledge items, and adaptation to direct and easy understanding.

• The capability of retrieving information interactively and the use of data-driven paradigms are now more than ever necessary due to the huge data volumes being involved.
Examples
Multisensor search engine
Query Results

Optical River-Rank by Similarity

Query
Post. Map

Optical Tree-Rank by Probability

Query
Post. Map

SAR Railway-Rank by Probability

Query
Post. Map

Optical Tree-Rank by Probability

Query
Post. Map
Cascaded active learning

- Two main components: Feature extraction and Learning
Semantic annotation

The location of the 300 TerraSAR-X scenes and the distribution of the scenes
From CLC to our semantic taxonomy

Legend - categories defined for Venice using CORINE Land Cover nomenclature:
- Marine waters – coastal lagoons
- Marine waters – sea and ocean
- Urban fabric
- Pastures
- Forest
- Heterogeneous agricultural areas
- Open spaces with little or no vegetation
- Industrial, commercial and transport units
- Open spaces with little or no vegetation
- Artificial, non-agricultural vegetated areas

• Using: CLC – 10 categories;
  our methodology – 17 categories
• In the case of CLC some categories are mixed together (e.g., the bridges are included in marine waters – coastal lagoons)
Semantic catalogues

- Bangkok (Thailand);
- Shenyang (China);
- Nazca Lines (Peru);
- Havana (Cuba);
- Venice (Italy);
- Vasteras (Sweden);
- Oran (Algeria);
- Bogota (Columbia).
1 HS TerraSAR-X Scene = up to 10,000 image patches (100 x 100 m)
Knowledge Discovery application example

The damages in the agriculture can be clearly seen by comparing the classification in pre disaster image (left figure) with the post disaster image (right figure).

TerraSAR-X scene **before** Tsunami
20.10.2010

TerraSAR-X scene **after** Tsunami
12.03.2011
Data Analytics: Tsunami effects assessment
Leading Edge: Big Data Analytics
Immersive Visual Information Mining for EO image archives

Navigation inside the EO image collections using the CAVE automatic virtual environment
Immersive Visual Information Mining for EO image archives
EOLIB – Earth Observation Image Librarian

- Big Data Mining in the TerraSAR-X Ground Segment System
Selected references


