Earth Observation Big Data: New Paradigms

Mihai Datcu

(presented by Gottfried Schwarz)

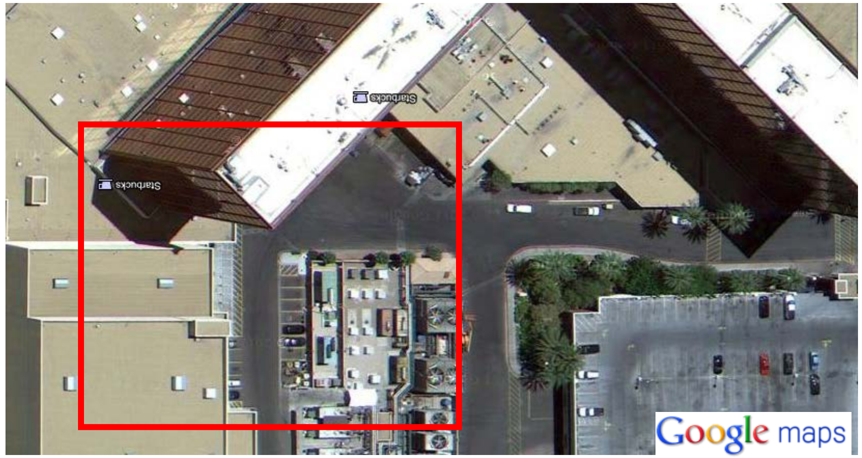
DLR

Multispectral Sensors: Sentinel-2

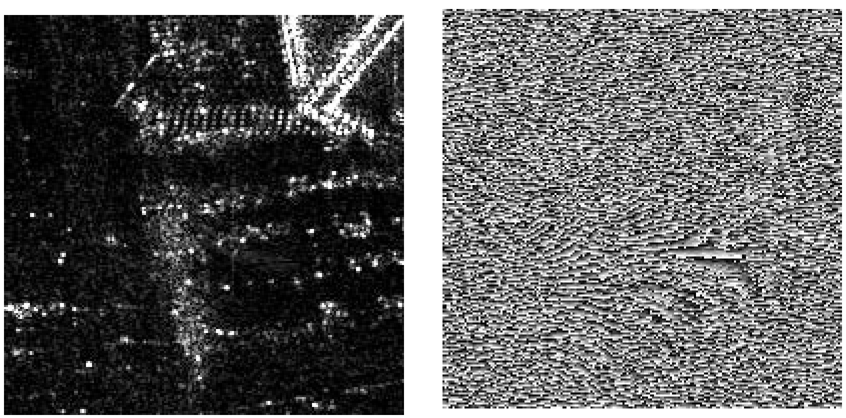
RGB Image

Infrared Image

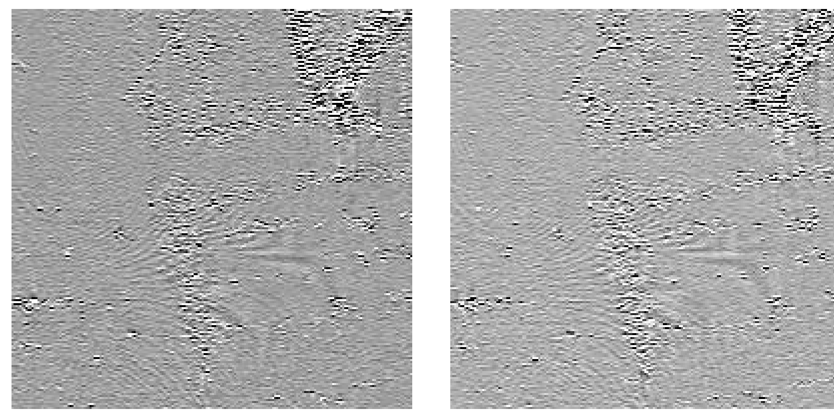
Synthetic Aperture Radar (SAR) vs. Optical Images



TerraSAR-X Single Look Complex Image



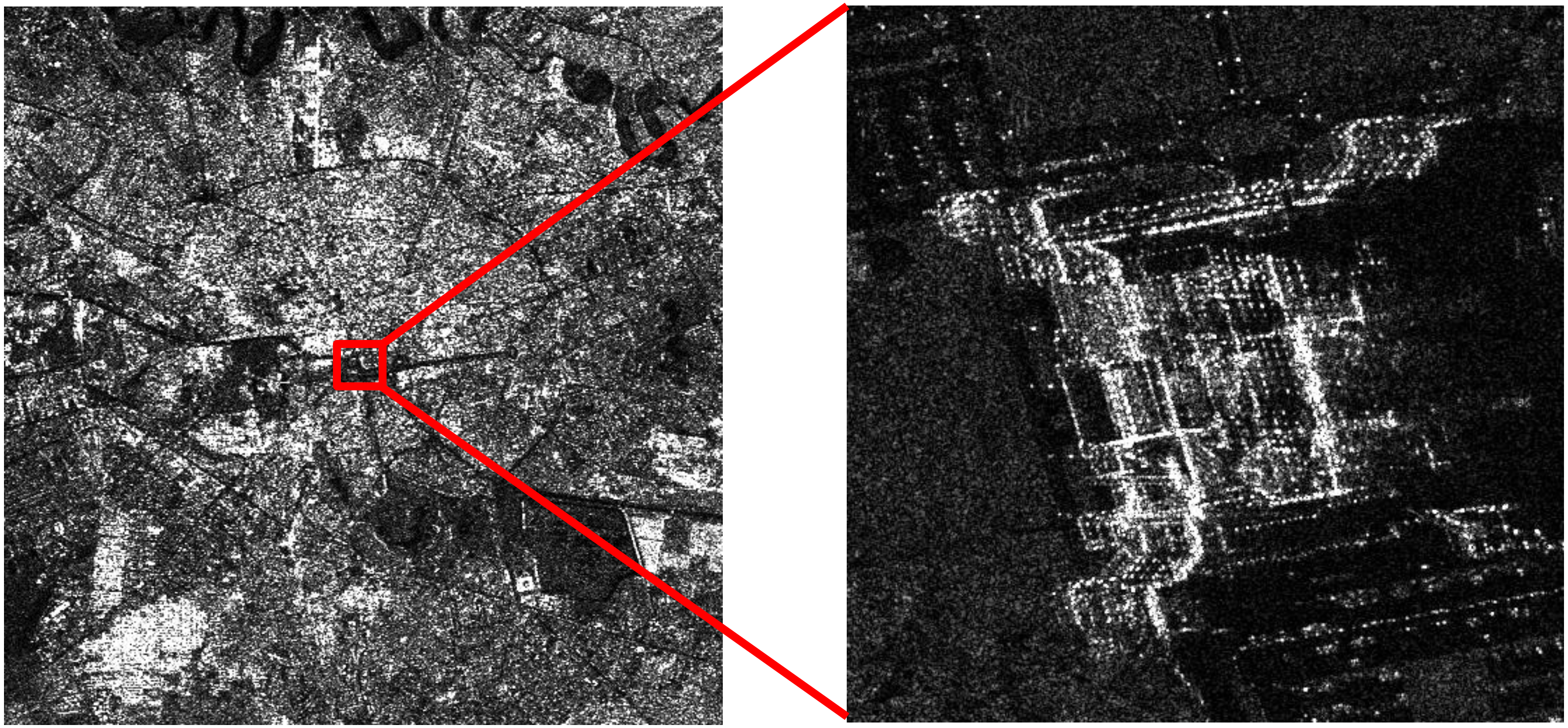
Amplitude and Phase



Real and Imaginary Components

DLR

Sensor Resolution



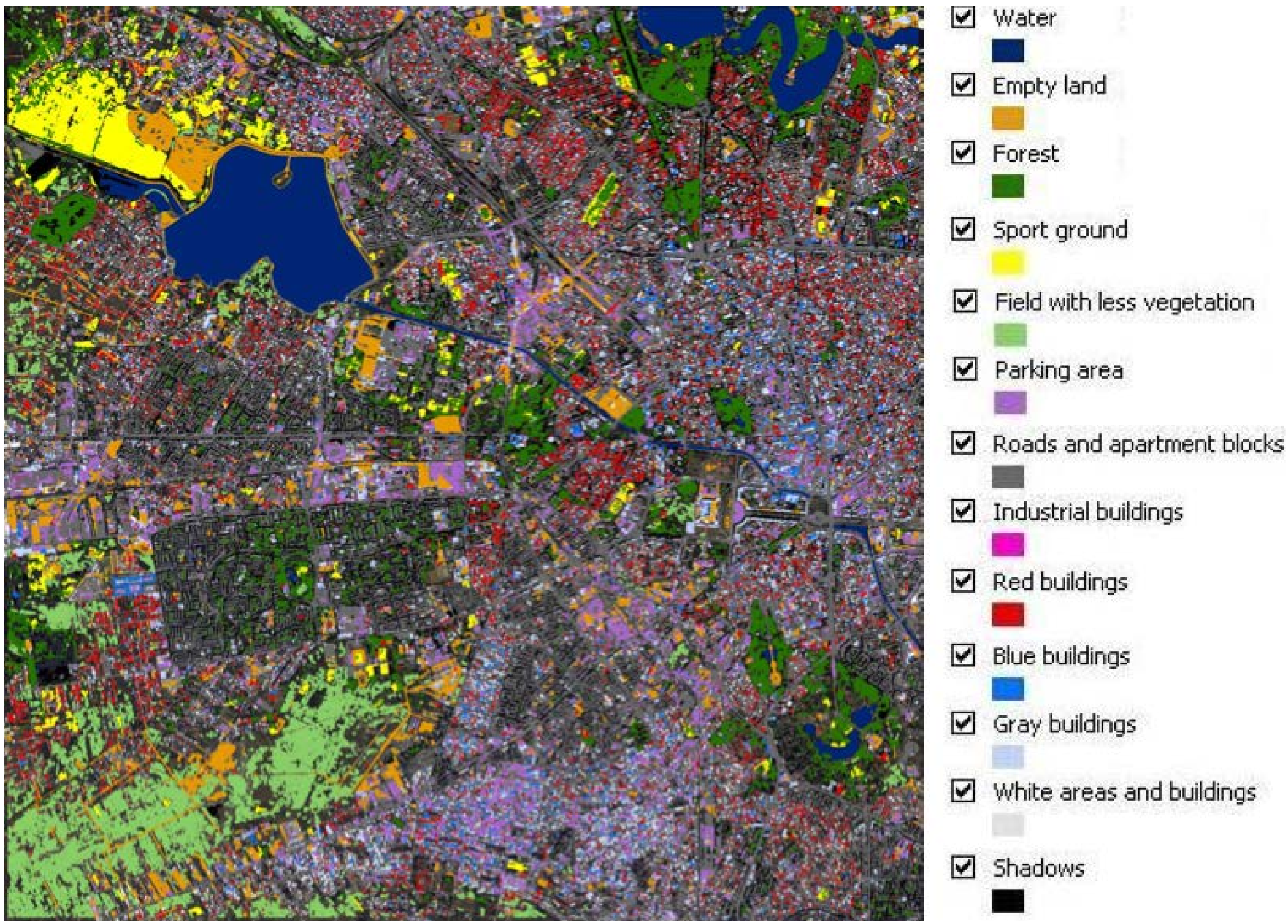
Mission: ERS1
Type: SAR_IMS_1P
Acquisition: 24-JUL-1992
512x512 pixels

Mission: TerraSAR-X
Type: SSC__HS_S
Acquisition : 11-OCT-2008
512x512 pixels

TerraSAR-X: Non-Visual Data

Information Content

WorldView 8 bands, 2 m: Spectral Classes



Spatial Categories

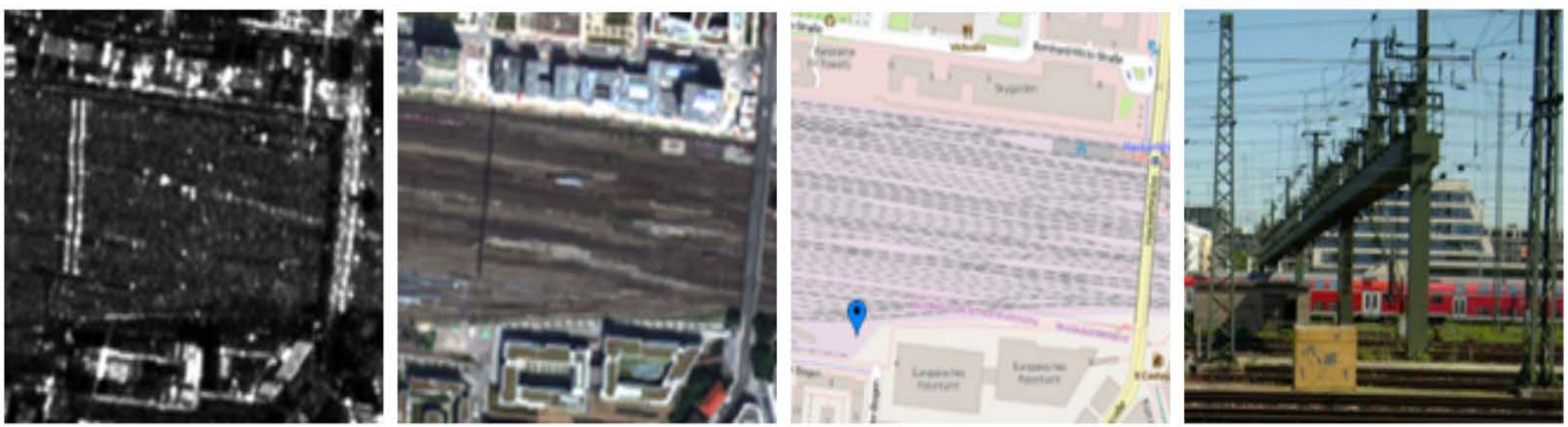
DLR

The Earth Observation Context

Earth observation data are always **used jointly** with information extracted from **other sources** such as Geographical Information Systems (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.

Note data volumes, their heterogeneity, unstructured nature, and their complexity.



(a) SAR (b) Multispectral (c) Map - OpenStreetMap (d) LUCAS

Our Particular Challenges

Challenge 1: Data Volume and Heterogeneity

Challenge 2: Big Data Earth Observation Analytics

Challenge 3: Big Data Earth Observation Data Mining

Challenge 4: Human-Machine Communication

Challenge 5: Information Platforms

Challenge 1: Data Volume and Heterogeneity

- Earth observation images: multisensory data, e.g., multispectral, SAR, altimeter data, etc.
- These are multidimensional signals, acquired by sensors or instruments
- Sensor data carry physical meaning, radiation levels, wavelengths, etc.
- They are measuring land, ocean, or atmospheric parameters
- Very high resolution Earth observation images observe detailed spatial structures and objects
- Satellite image time series observe evolution processes over long periods of time.

- An important particularity of Earth observation images is their "**instrument**" nature, i.e., they result from **sensing physical parameters, and they are often sensing outside of the visual spectrum.**

DLR

Challenge 1: Data Volume and Heterogeneity

- Earth observation product metadata are describing location, time of acquisition, instrument parameters, orbit information, product processing level, etc. They appear in text and numerical formats with a broad diversity of meaning, physical units, format representations, etc.

- We also use GIS and maps, on various themes (urban areas, vegetation, topography, etc.), in standardized digital formats.

- Other types of geo-information are geo-morphological models, models of evolution, textual descriptions, etc.

- We exploit in-situ information and photography, continuously growing with the diversity of sensors in larger and larger networks, measuring physical parameters such as air quality or water content, etc.

- Also note location information, multimedia location awareness, GPS, tagging, spatial context, social networks in the Internet or mobile communication information, with a great evolution in diversity and volume, and containing unexpected important information.

Challenge 1: Data Volume and Heterogeneity

- All these are autonomous sources with distributed and decentralized control.

- In this context, Big Data Earth observation seeks to explore complex and evolving terrestrial processes and their inter-relationships impacting environmental, and socio-economic developments.

- Therefore, Big Data Earth observation has another very specific dimension, namely its large and diverse area of applications, and users from a large diversity of disciplines.

Challenge 2: Big Data Earth Observation Analytics

- Today's techniques, methods, and tools for automated data analysis are insufficient for the analysis and information extraction from Earth observation 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.

Challenge 2: Big Data Earth Observation Analytics

Methods of **Computer Vision** and **Pattern Recognition** are needed for new tasks:

- **Detecting, localizing** and recognizing objects
- Recognition and extraction of **semantic descriptions** from sensor data images
- Extraction of **quantitative measures** of physically meaningful parameters from a scene
- **Co-registration** of multi-sensor multi-temporal data
- **Exploitation of the various imaging modes** to provide different types of information about various structures
- Recognition methods to **distinguish the huge variability of scene classes** and objects with very good precision

Challenge 3: Big Data Earth Observation Data Mining

Big Data needs more and more **machine or statistical learning** for "**discovery**" functions.

The discrepancy between the **data volume** explosion and our **analysis potential** is continuously growing; new solutions are required:

- Detection of **irrelevant** data

- New sensors based on **Compressive Sensing/Sampling**, recoding smaller data volumes but with pertinent content
- Data **compression**

- **Machine/statistical learning** algorithms for fast prediction
- **DNNs** for large scale prediction
- Content analysis to extract **higher-level analytics**
- Extraction and **formalization of knowledge** for data classification and understand-ing

DLR

Challenge 3: Big Data Earth Observation Data Mining

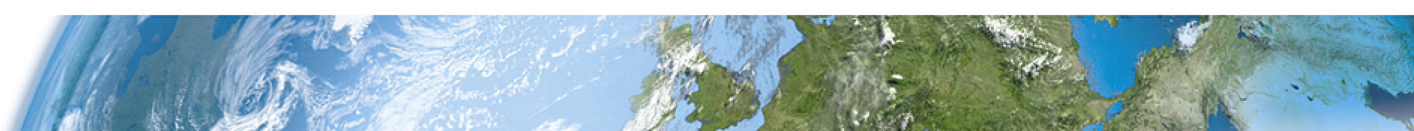
Advanced topics, **beyond** today's techniques and methods:

- Computational imaging
- Sensor networks
- Quantum sensors

Machine Learning and Analytic Systems **beyond** the physical **spatio-temporal** domain

- Quantum information theory
- Quantum signal processing
- Quantum machine learning

- Quantum computers



Challenge 4: Human-Machine Communication

Predictive, adaptive and natural user interfaces

Learning and anticipating the user **behavior,** and collaboration with the user

Understanding and learning the user **intentions and context**, establishing a **dialog**

Transformation of **non-visual sensor** data and information into easily understandable representations for humans.

Challenge 5: Information Platforms

Web-based interactive technologies and tools

Distributed architecture systems

We have to cope with demanding computational loads and requirements on the data volumes to be accessed, the complexity of the information to be extracted, analyzed and presented, adaptations to specific applications, and the speed of interactive operations.

Cloud computing should enable tasks not achievable with current resources.

New methods are further needed, since tools such as Hadoop or MapReduce have reached their limits.

Potential solutions are foreseen in virtual Earth observation centers, connected and communicating across clouds for enhanced potential to share hardware resources and data.