GEOProcessing12 Conference – Valencia, Spain

Challenges in Handling Large Data Volumes for (Geo-)Processing

GEOProcessing12 Panel Discussion

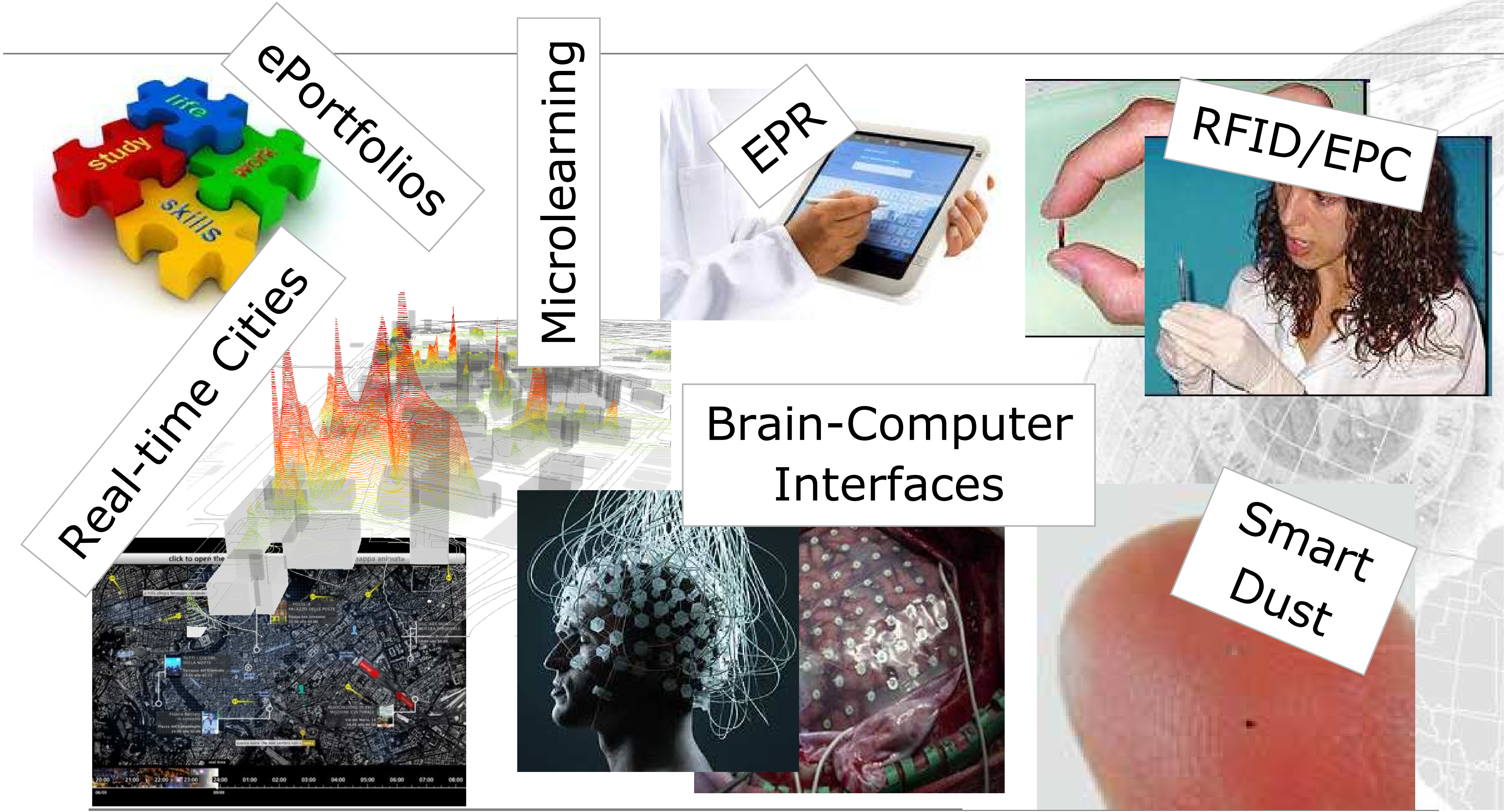
Prof. Dr. Bernd Resch

31 January 2012

Panelists

- Claus-Peter Rückemann
  Leibniz Universität Hannover / WWU / HLRN, Germany
- David Pheanis
  Arizona State University-Tempe, USA
- Yerach Doytsher, Technion
  Israel Institute of Technology - Haifa, Israel

Large Data Volumes in Research



ePortfolios

Microlearning

EPR

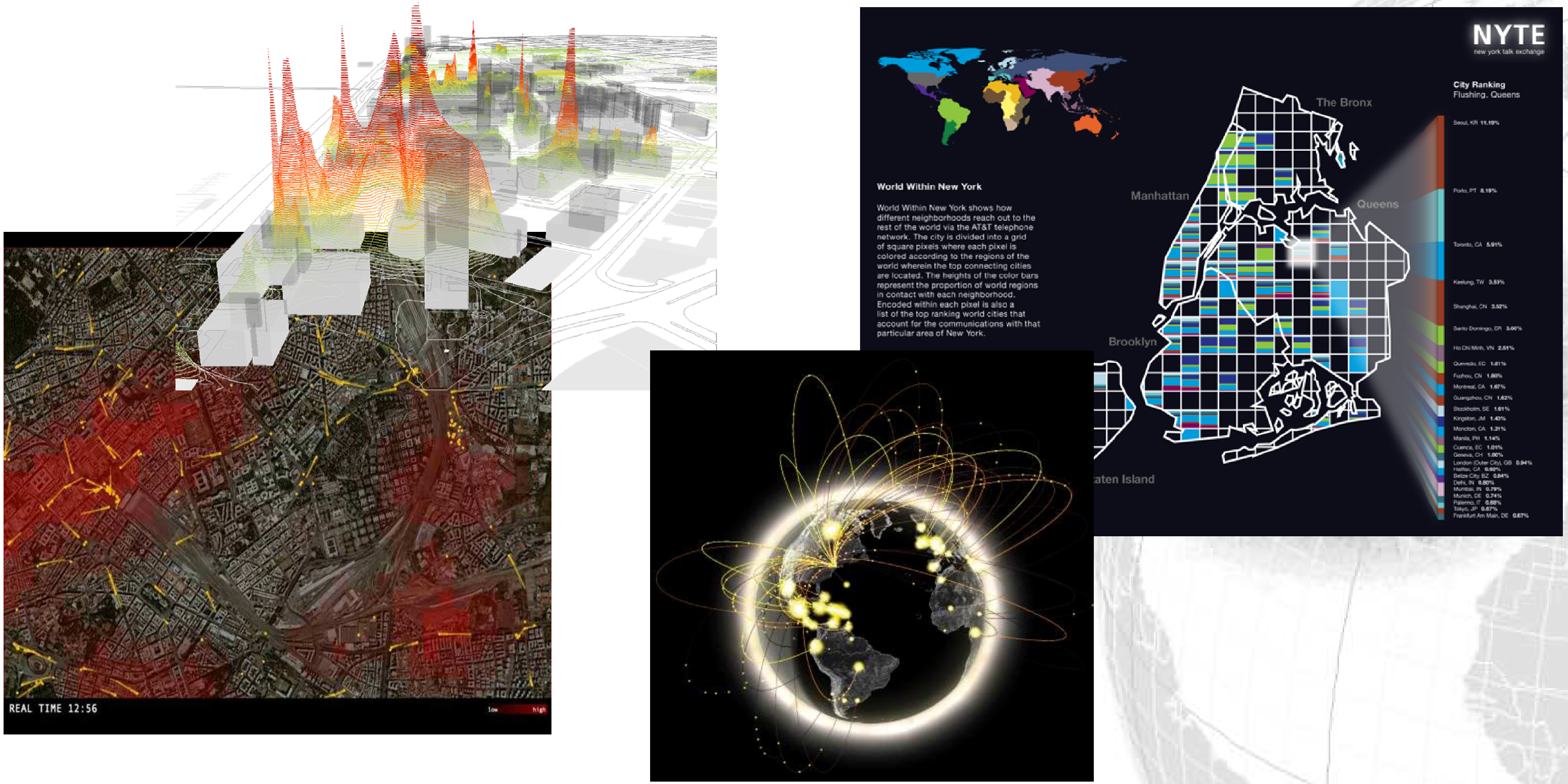
RFID/EPC

Real-time Cities

Brain-Computer Interfaces

Smart Dust

Large Data Volumes in GI Research



World Within New York

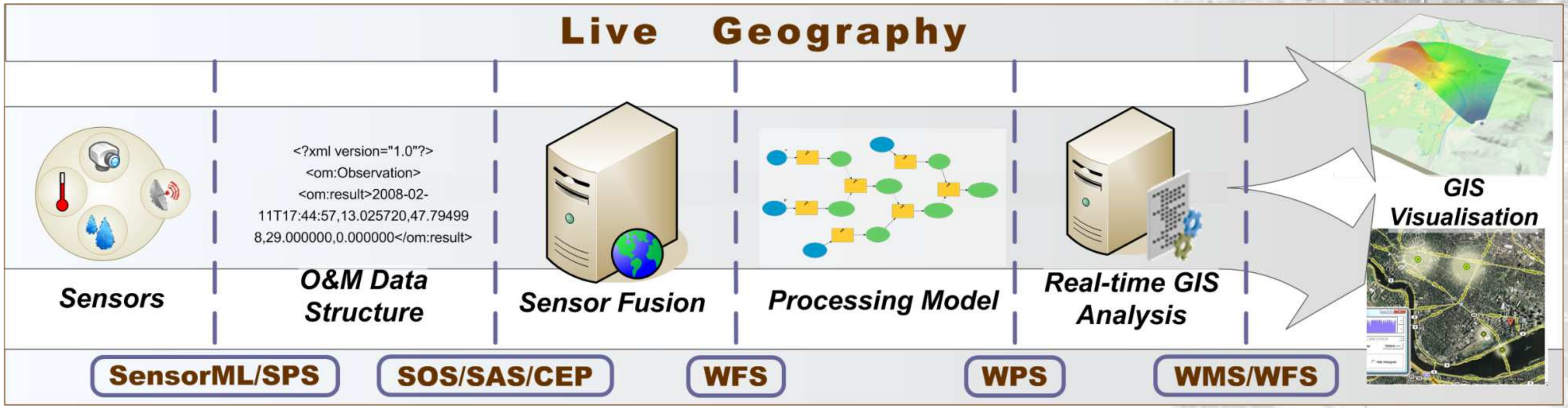
World Within New York shows how different neighborhoods reach out to the rest of the world via the AT&T telephone network. The city is divided into a grid of square pixels where each pixel is colored according to the regions of the world wherein the top connecting cities are located. The heights of the color bars represent the proportion of world regions in contact with each neighborhood. Encoded within each pixel is also a list of the top ranking world cities that account for the communications with that particular area of New York.

Large Data Volumes in GI Research

- Flexible and portable monitoring infrastructure

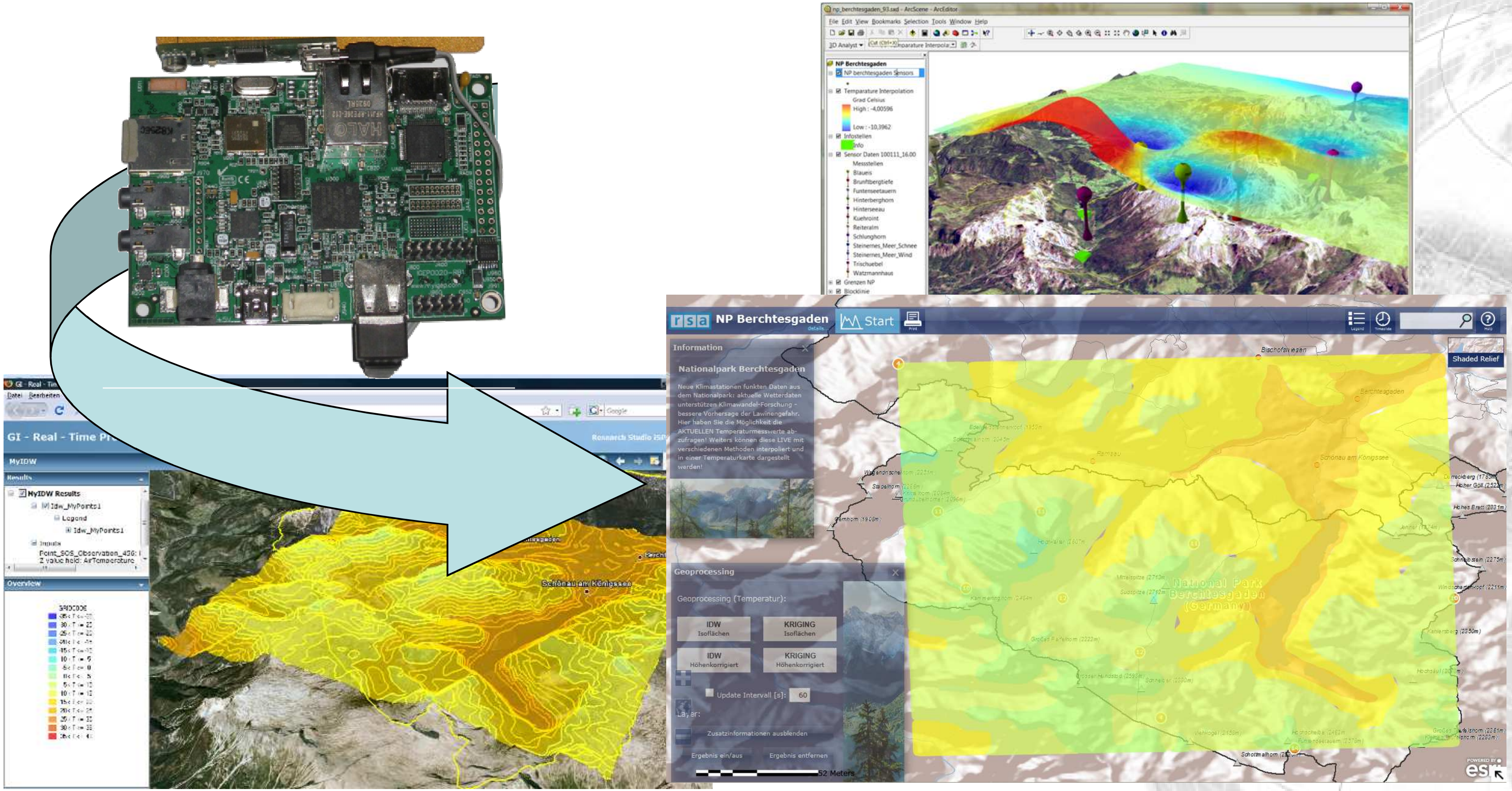
→ Standardisation enables a wide variety of monitoring applications

→ Sensor systems: not only view and analyse the world, but influence it in real-time



```
<?xml version="1.0"?>
<om:Observation>
<om:result>2008-02-
11T17:44:57,13.025720,47.79499
8,29.000000,0.000000</om:result>
```

Large Data Volumes in GI Research



Neue Klimastationen funkten Daten aus dem Nationalpark: aktuelle Wetterdaten unterstützen Klimawandel-Forschung - bessere Vorhersage der Lawinengefahr. Hier haben Sie die Möglichkeit die AKTUELLEN Temperaturmesswerte abzufragen! Weiters können diese LIVE mit verschiedenen Methoden interpoliert und in einer Temperaturkarte dargestellt werden!

Potential Challenges

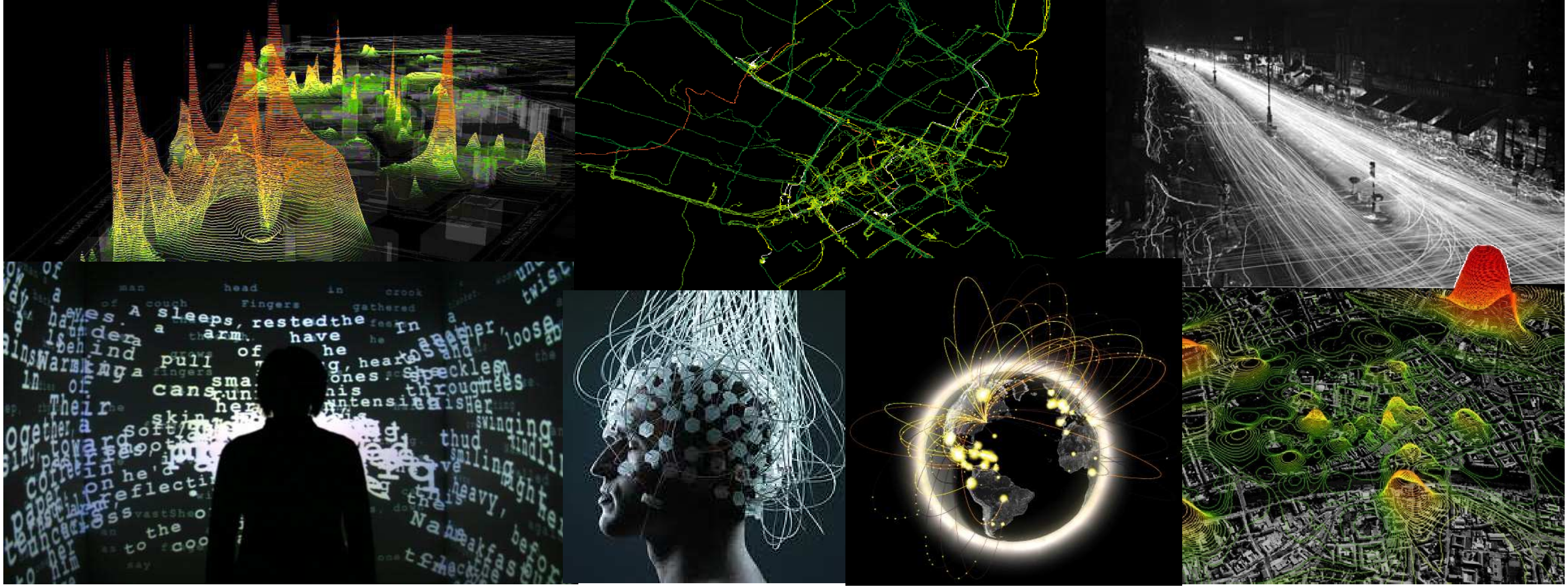
- **Real-time** capabilities in complex data processing
- **Distributed** sensor/data fusion, on-the-fly integration
- **Event detection** (CEP, ESP)
- **Integration** of VGI, UGC
  and People as Sensors
- **Visualisation**, 2D/3D
- **Interoperability**, Semantics
- …

Panelists

- Claus-Peter Rückemann
  Leibniz Universität Hannover / WWU / HLRN, Germany
- David Pheanis
  Arizona State University-Tempe, USA
- Yerach Doytsher, Technion
  Israel Institute of Technology - Haifa, Israel

Conclusions for (GI) Research

- Tight coupling with cloud computing developments
- Efficient methods for transmission and storage needed
- Comprehensive semantic models as a central enabler
- Strong focus on efficient algorithms and implementations
- Profiling (generalisation)
- Data sub-selection



Challenges in Handling Large Data Volumes for (Geo-) Processing

GEOProcessing12
Panel Discussion

IARIA

UNIVERSITÄT
OSNABRÜCK

IGF

31 January 2012

PANEL GEOProcessing

Handling Large Data Volume

Seamless Nationwide Topographic Multi-Databases Challenges

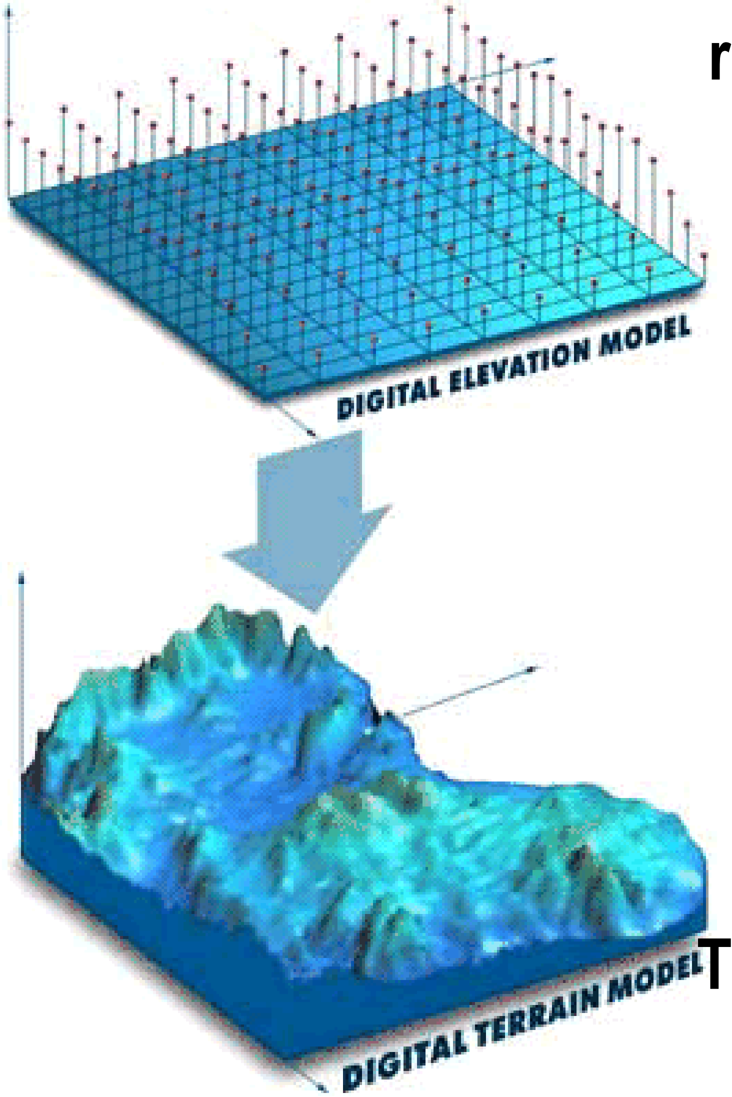
Yerach Doytsher

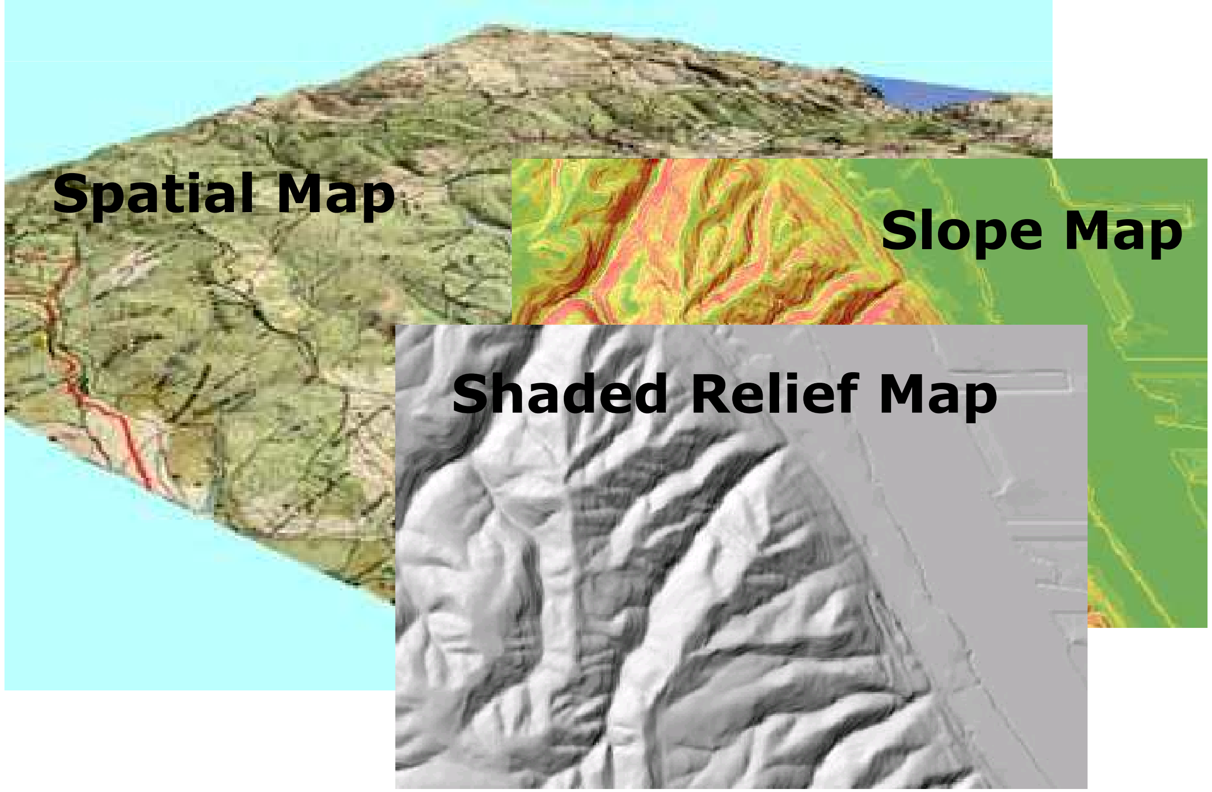
Mapping and Geo-Information Engineering

Technion, Israel

Introduction

DTM (Digital Terrain Model) - topographic computerized model representing the terrain relief (alterations in the topography).

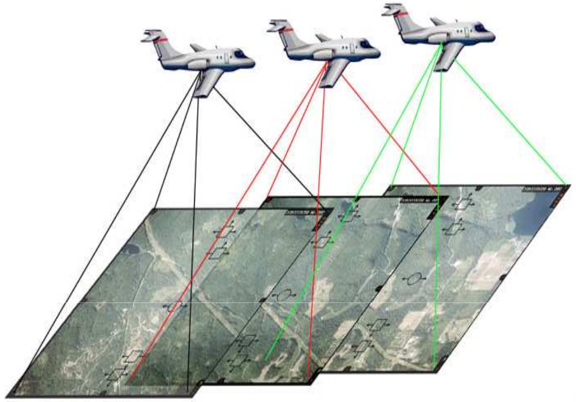




These models serve as a continuous, constant, reliable and homogenous infrastructure.

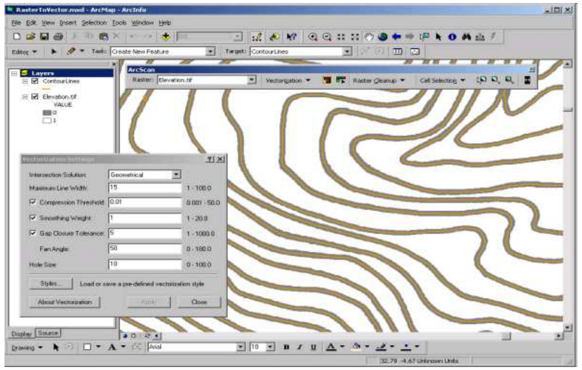
DTM sources

Photogrammetry



Field Surveying

Digitization/scanning



Vectorization Settings

Intersection Solution: Geometrical

Maximum Line Width: 15 (1 - 100.0)

☑ Compression Threshold: 0.01 (0.001 - 50.0)

☑ Smoothing Weight: 1 (1 - 20.0)

☑ Gap Closure Tolerance: 5 (1 - 1000.0)

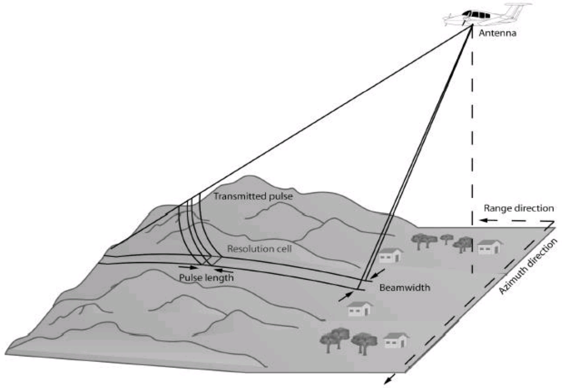
Fan Angle: 50 (0 - 180.0)

Hole Size: 10 (0 - 100.0)

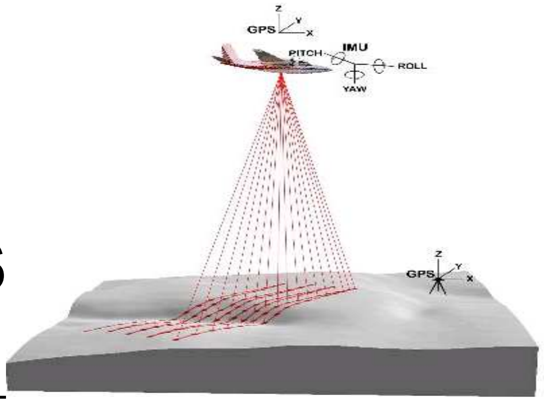
Styles... Load or save a pre-defined vectorization style

About Vectorization | Apply | Close

Radar based systems



ALS (LiDAR) Systems



DTMs: Accuracy of Different Sources

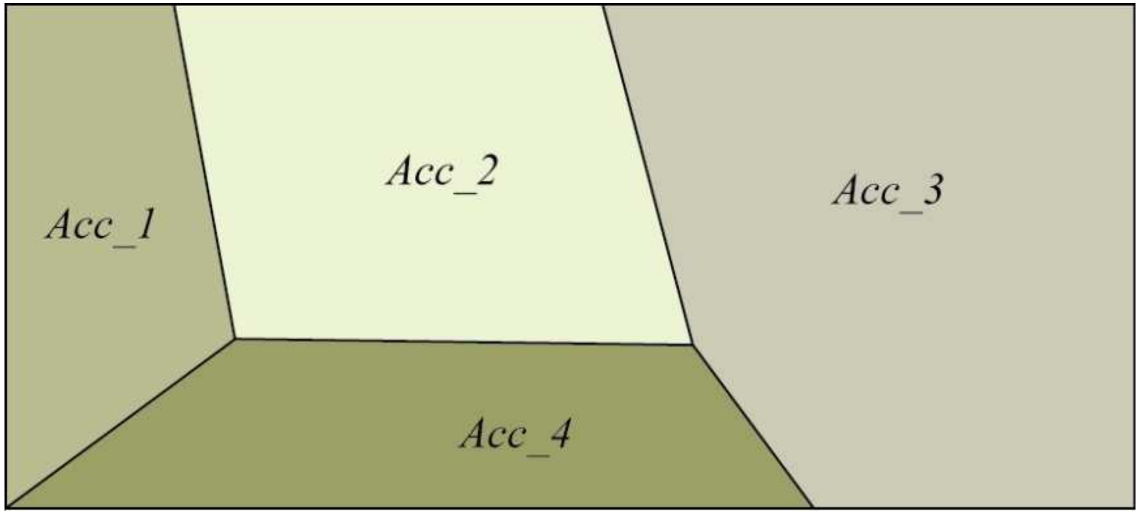
Technique/Technology	Vertical Accuracy (m)
Aerial photogrammetry	0.1 – 1
Satellite photogrammetry	1 – 10
Field surveying	0.01 – 0.1
Digitization	1/3 of contouring interval
Aerial radargrammetry	2 – 5
Satellite SAR inteferometery	5 – 20
LiDAR	0.1 – 0.2

Multi-Sources Challenges

Need - integration of DTMs is essential for obtaining computerized topographic infrastructure.

Status - multi-source DTM:

- **Produced via various technologies and techniques;**
- **Influenced/affected by rapid data-updates.**

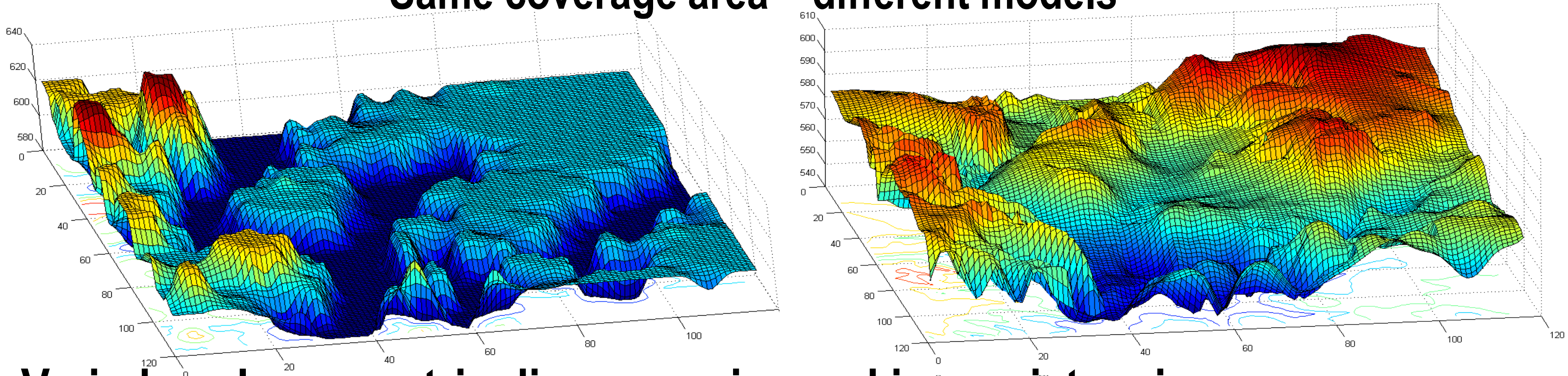


A typical accuracy zoning map

Problem Challenges

Different DTMs can vary and present different data-characterizations: structure, data-density, level-of-detail, accuracy, resolution, datum, …

Same coverage area – different models



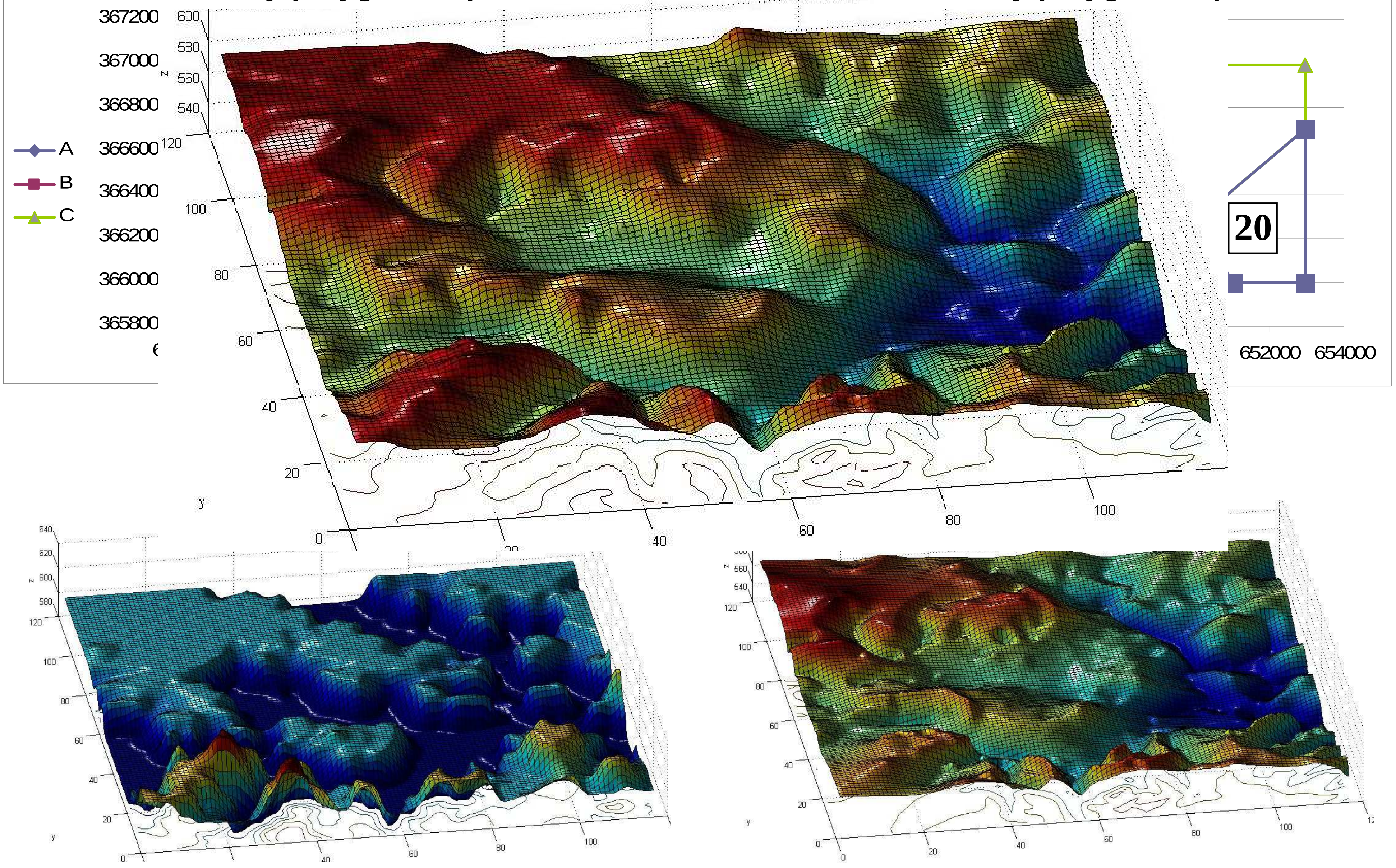
- **Varied-scale geometric discrepancies and inconsistencies;**
- **Global-systematic and local-random inaccuracies;**

The common "coordinate-based superimposition" is not reliable and the results are not acceptable; it should be replaced with a "feature-based" matching/fusion solution (hierarchical approach)

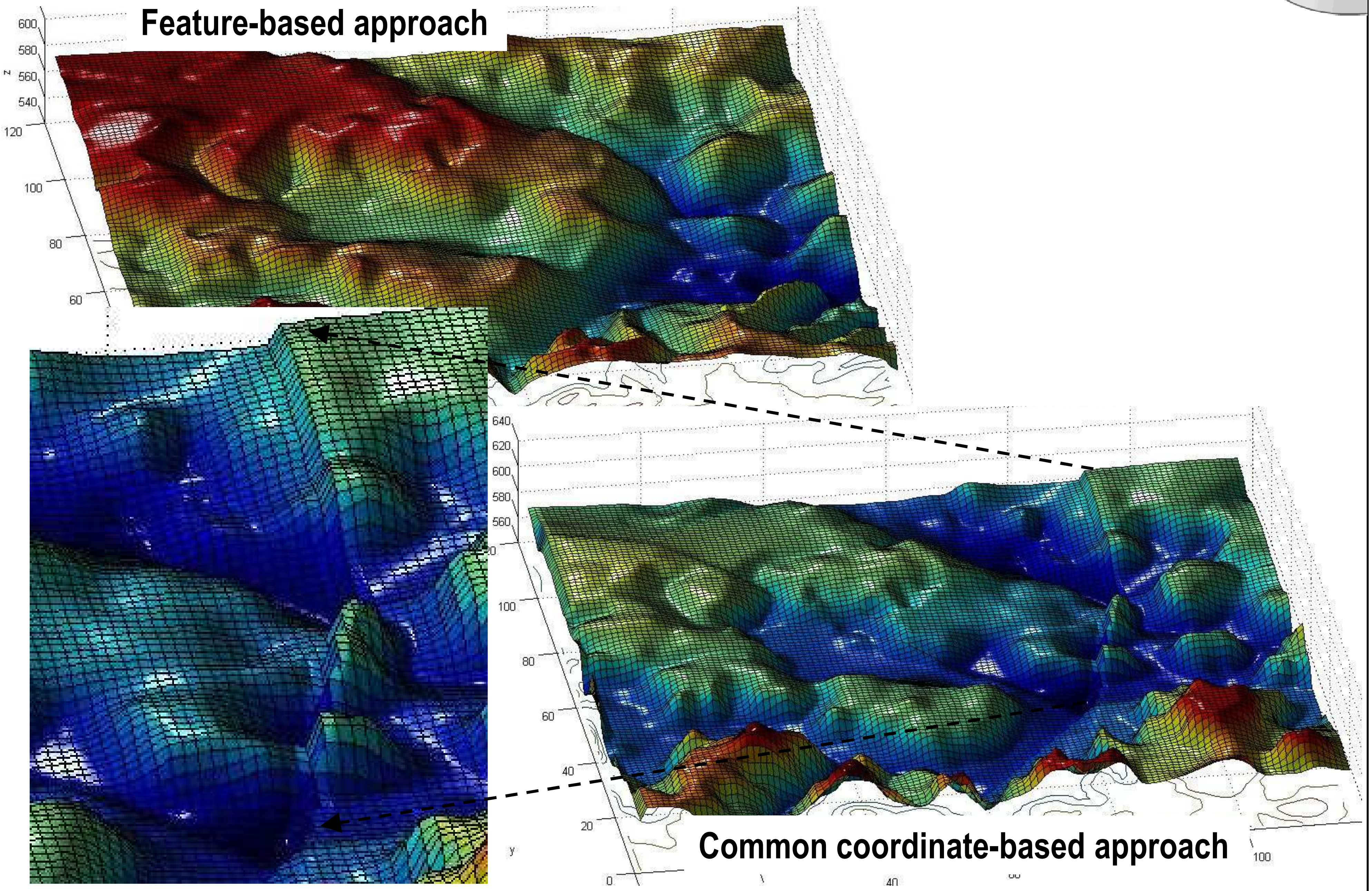
Results

Accuracy polygon map I

Accuracy polygon map II



Results



Feature-based approach

Common coordinate-based approach

International Panel GEOProcessing 2012
Challenges in Handling Large Data Volume for GEO Processing

High End Computing and Data Volume

International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012)
DigitalWorld 2012, January 31, 2012, Valencia, Spain

IARIA

Dr. rer. nat. Claus-Peter Rückemann[1,2,3]

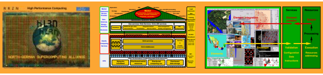
[1] Leibniz Universität Hannover, Hannover, Germany
[2] Westfälische Wilhelms-Universität Münster (WWU), Münster, Germany
[3] North-German Supercomputing Alliance (HLRN), Germany

ruckema(at)uni-muenster.de

Westfälische Wilhelms-Universität Münster

GEXI

ACT@MAP



Status on High End Computing and Data Volume: Supercomputing

Supercomputing, High Performance Computing

Challenges with users application scenarios, service providers, resource providers large volume data handling for geo-data and funding point of view:

Disciplines:

- Seismics and Seismology (SEG-Y), Geosciences, Natural Sciences . . .
- Geophysics, Global Environment, Climatology (NetCDF), Archaeology, . . .
- Visualization, streaming data . . .

Services and Resources:

- Parallel IO on large TeraByte-datasets.
- Large numbers of datasets on parallel high end filesystems.
- Running parallel NetCDF on High End Resources.
- Data storage and archiving of large data.
- Transfer and broad band for geo-computation data.
- HW architecture aspects, bottlenecks (appliances, storage targets, IO servers).
- Data-staging and pre-staging for large geo data.

Development of Applications and Methods:

- MPI and efficiency for large datasets.
- Funding for handling of parallelisation and parallel IO.
- Future IO challenges and strategies . . .

Status on High End Computing and Data Volume: Distributed Computing

Distributed Computing

Disciplines:

- Interactive communication requirements (quantity and quality).
- Data transfer to/from distributed resources (interactive and batch).

Services and Resources:

- Communication: Insufficient data transfer rates for large data volumes.
- Storage: Insufficient backup capacities for large data volumes.

Development of Applications and Methods:

- . . . depending on funding, physical resources, consulting.
- . . . depending on reliability, high availability, security.

Vision

What has to be done for the next generation of high end systems

Next generation: The next generation of complex high end systems will not be just only the next sizeof-factor multiplied with all the items required or just another project. The next generation step can neither be reached by hardware nor by software alone.

Holistic view: For reaching a general state of the art there is no efficient solution available other than an holistic handling, integrating user applications, services, and resources.

Common understanding: An overall understanding of integrated use, operation, and provisioning regarding software and hardware is neccessary for appropriate funding.

Real life case studies: Real life large data case studies with keyplayer participants from academia and industry representing disciplines, services, and resources providers.

Challenges on Future Prerequisites

Prerequisites on development and technology

- Funding for integrated systems with academia and industry for user applications and developer support. This has been shown not to be possible with standard restricted project funding.
- Common availibility of data transfer to compute resources, international Broadband resources (TeraBytes per hour, per user group).
- Fast and massive I/O ($>$ 100 GB per second, per application; $>$ 1 GB per second, per core).
- Fast massive communication ($>$ 100000 requests per interactive application, per operation).
- Fast Archiving, Storage and Retrieval (PetaBytes per day, per user group).
- Reliable and secure data and resources access (homomorphic?).
- Data-Staging (hardware and software technology efficiently using the high end resources, availability of about 1 TeraByte per 5 Minutes).

Using GPS-Enabled Cell Phones for Traffic-Flow Data Collection

Bruce Beyeler and David C. Pheanis
Computer Science and Engineering
Ira A. Fulton School of Engineering
Arizona State University
Tempe, AZ 85287-8809
Bruce.Beyeler@ASU.edu or David.Pheanis@ASU.edu

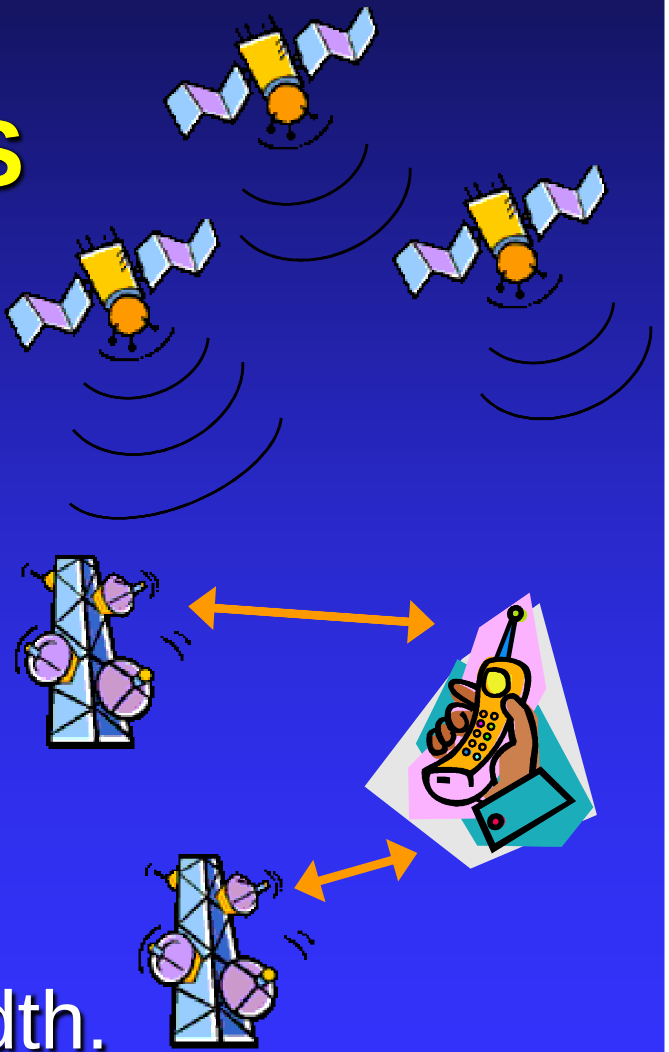
Context

- **Measuring Traffic Flow**
  - Using cell phones as sensors
  - Need to understand provided accuracy
- **Questions to be Answered**
  - How accurate are cell-phone reports?
  - How does accuracy vary?
  - Are cell-phone accuracy reports accurate?
- **Research Goal**
  - Optimize collection algorithms
  - Minimize required data collection.

Existing Research

- **Static Collectors**
  - Inductive loops, cameras, …
  - Vehicle counts, speeds, at fixed locations
  - Average results per road segment
- **Dynamic Collectors**
  - Cell phones, GPS probes, …
  - Report data based on time, distance, etc.
  - Provide individual reports continuously
- **Comparison to Static Collectors**
  - Cell-phone data transformed in order to compare against static data.

Cell-Phone GPS



- **Cell phones include GPS**
- **Network-Assisted GPS**
- **Assumptions**
  - Majority of future phones will be GPS-enabled
  - Accuracy will improve
  - Capability of running user programs
  - Available low-cost bandwidth.

Compare to Ground Truth

- **Ground-Truth Measurements**
  - Utilize known map data
  - Vehicle data
  - Operator inputs
- **Current Research Focus**
  - Accuracy of individual cell phones
  - Different conditions/positions
  - Goal: Determine the *actual* speed and position of a test vehicle at any given time.