imec

An imec.icon research project | project results

ICT

BAHAMAS

A toolset for the processing, compression and analysis of big data in life and materials science applications

Big data analysis is becoming increasingly important. Large sets of data are being generated by the very latest and most advanced tools, but solutions are lacking that allow the scientists to efficiently cope with the overwhelming amount of information.

The BAHAMAS research project aims to offer life scientists and materials scientists an innovative platform for the storage, processing and analysis of big data. Four distinct use cases have been identified, focusing on the processing of:

- **3D electron microscopy data, supporting the analysis of sub-cellular organelles;**
- **hyperspectral data, to study the impact of environmental changes on plant growth;**
- **spectral and electron microscopy data, to develop a new framework for classifying potential gun-shot residue particles;**
- **high-throughput video, supporting the analysis of a group of fruit flies to study their (social) behavior.**

"Within the BAHAMAS project, four use cases were selected representing the fields of molecular biology, biotechnology, neuroscience and forensics," says Jan Lemeire, Research Lead of the BAHAMAS project. "Although the requirements are different for each individual case, they have one common challenge: the analysis of large sets of data. BAHAMAS wants to offer a generic solution that supports the end users in collecting, processing, analyzing and compressing (for future queries and data-mining) the data. By optimizing the value chain in a generic way, the project's outcome will be broadly applicable."

THE OUTCOMES

1. A novel generic microscope image analysis chain for processing 3D electron microscopy volumes

3D-scanning electron microscopy (3D-SEM) is used to investigate 3D ultra-structures (organelles) of biological tissues at nanometer resolution. One dataset contains almost 100GB of data. So far, by lack of automated tools, the segmentation (or the identification of the organelles in an image) had to be done manually which is extremely time-consuming. Jan Lemeire: "Within the project, we developed an automated image analysis chain for the processing and storage of large 3D-EM volumes. A plugin was developed for extracting and analyzing the images from the 3D-SEM (Fiji) tool. Fast GPU-based image segmentation algorithms were developed, reducing the cell segmentation time from weeks to only a couple of days. Data compression and storage is enabled by extending the JPEG codec for spectral data."

2. A novel processing framework for high-throughput hyperspectral data

In a second use case, hyperspectral imaging is used to study the impact of environmental changes (such as drought) on plant growth and physiology. The camera setup, part of an innovative 'Phenovision' system, acquires 3D data cubes per individual plant, and is able to generate 150TB of data per year. Since no automated tools are available for processing the data, the potential of the tool could not be fully exploited. The project partners developed a fully automated framework able to handle 30TB per year. Processing of the data per plant, including 3D reconstruction and modelling, can now be done within half a minute timeframe.

3. A new framework for classifying millions of potential gunshot residue particles

In the context of forensic analysis of gunshot residues (GSR), scanning electron microscopy and x-ray spectroscopy data is routinely generated from criminal case samples. Project Lead Bart Nys: "The data are used to investigate (1) whether any gunshot residues are present on the sample and (2) which type of ammunition has been used. Typically, 10.000 particles per sample are being characterized, and a subset of 100 particles is selected for manual verification. The forensic community is however challenged with a new evolution: novel ammunition has been put on the market that no longer contains (heavy) metals – on which previous classification was based. Within this project, a novel approach has therefore been followed, relying on the analysis of millions of (non-metal) particles. A framework was developed allowing to automatically classify the particles within a few days, and without human intervention, and compare them with known GSR samples stored in a central database."

4. A very fast automated fruit flies behavior analysis framework

A final challenge is in the use of high-throughput video data to study social behavior of fruit flies. Recording the flies with a high-speed camera for several hours generates gigabits of data, from which features are to be extracted – such as the social interactions between the flies. Within the project, an automated framework has been developed that allows real-time tracking of flies to detect different types of behavior (e.g. through training by human annotation). Algorithms have been developed that allow to represent and analyze the data.

NEXT STEPS

The developed tools and algorithms have substantially improved the efficiency and throughput of big data analysis, and extensions are possible beyond the specific use cases. Bart Nys: "The developments will e.g. facilitate the integration of 3D-SEM imaging as a standard tool in life sciences research. And, although no direct commercialization is envisaged, the availability of e.g. the automated phenotyping platform will result in a competitive advantage for the partner involved. For the use case on forensics, the proposed scientific and technological challenge is identical for all forensic laboratories worldwide. The GSR database can be made available to other forensic labs."

AGENTSCHAP INNOVEREN & ONDERNEMEN
Vlaanderen is ondernemen

The BAHAMAS project was co-funded by imec (iMinds), with project support from Agentschap Innoveren & Ondernemen.

FACTS

NAME	BAHAMAS
OBJECTIVE	Developing a toolkit for the collection, processing, compression and analysis of high-throughput data in the life sciences and materials sciences
TECHNOLOGIES USED	principal component analysis, neural networks, JPEG2000 codec, GPU processing, smart caching
TYPE	imec.icon project
DURATION	01/01/2015 – 31/12/2016
PROJECT LEAD	Bart Nys, NICC
RESEARCH LEAD	Jan Lemeire, imec - ETRO - VUB
BUDGET	2,077,707 euro
PROJECT PARTNERS	SMO, DCILabs, Pathomation, NICC, VIB
IMEC RESEARCH GROUPS	IPI - UGent, Vision Lab - University of Antwerp, ETRO - VUB

WHAT IS AN

IMEC.ICON PROJECT?

The imec.icon research program equals demand-driven, cooperative research. The driving force behind imec.icon projects are multidisciplinary teams of imec researchers, industry partners and / or social-profit organizations. Together, they lay the foundation of digital solutions which find their way into the product portfolios of the participating partners.

BAHAMAS project partners:

NICC INCC dcilabs VIB PATHOMATION smo

AMERICAS
raffaella.borzi@imec.be
T +1 408 386 8357

CHINA
timo.dong@imec-cn.cn
+86 13564515130

EUROPE & ISRAEL
michel.windal@imec.be
+32 478 96 67 29

JAPAN
isao.kawata@imec.be
T +81 90 9367 8463

TAIWAN & SE-ASIA
mavis.ho@imec.be
T +886 989 837 678

VIETNAM, BRAZIL, RUSSIA, MID EAST, INDIA
max.mirgoli@imec.be
T +1 415 480 4519