The OpenVisus Framework for Extreme Data Management, Analysis and Visualization

Steve Petruzza, Aniketh Venkat, Nate Morrical, Giorgio Scorzelli, Valerio Pascucci, and Many Many Many Others
Massive Simulations and High Resolution Sensing Generate Big Data Challenges and Opportunities
OpenViSUS: A Data Intensive Collaborative Analytics and Visualization Platform

Agriculture

Geology

Desktop

Web & Mobile Clients

Server & Central Data, Database

Histopathology
OpenViSUS Software Platform

A set of interoperable components easily customized and deployed for different solutions as needed
PIDX demo

- Reference: https://github.com/sci-visus/PIDX/

- Optional: go through the code of the PIDX examples
- Run checkpoint-restart example to produce data in IDX format using PIDX
- Add dataset to the data portal
- Show data from the web viewer of the data portal
Demo: Interactive Remote Analysis and Visualization of 6TB Imaging Data

- EM datasets of resolution: 130Kx130Kx340

6.4 TB of raw data
High Resolution Seismic Models
Demo: large Scale Geology Data
A Science Cyberinfrastructure Requires Efficient Big Data Management and Processing

- **Advanced data storage techniques:**
  - Data re-organization.
  - Compression.

- **Advanced algorithmic techniques:**
  - Streaming.
  - Progressive multi-resolution.
  - Out of core computations.

- **Scalability across a wide range of running conditions:**
  - From laptop, to office desktop, to cluster of PC, to BG/L.
  - Memory, to disk, to remote data access.
A Scalable Solution for Acquisition and Processing High Resolution Data

Applications: Microscopy for Neuroscience

(1) Tissue Clearing (31 days)

Image data

(2) Data Source

(3) Preliminary Interactive Analytics

(4) Optimal Asynchronous Parallel Processing

(5) Interactive, Exploratory Assessment and Feedback

Large Scale Data Management and High Performance Computing Fully Integrated in a Brain Connectomics Workflow.
Demo Data Portal: convert and publish


Sample data (to unzip):
- Asteroid (500x500x500 float32):
  [https://drive.google.com/open?id=1wHdWynj9jnIf5NrrNZyOHbOMwKmltQ](https://drive.google.com/open?id=1wHdWynj9jnIf5NrrNZyOHbOMwKmltQ)
- Heart (stack of images):
  [https://drive.google.com/open?id=1NLWe2FVhUJoEaaA8JZ2w61H_dT77z1La](https://drive.google.com/open?id=1NLWe2FVhUJoEaaA8JZ2w61H_dT77z1La)

**Demo**

- Start/login your Docker installation
- docker pull visus/dataportal
- docker run –p 8080:80 visus/dataportal

 **Portal login:**
- User: admin
- Password: password

**Manage data**
- Asteroid: convert single file (browse, upload, select, insert information, convert)
- Heart: convert stack of images (browse, upload folder, select folder, convert)

**Add to server**

After 1-2 min data should be available for streaming on the viewer (Explore data)
We Characterize Algorithmic Classes Based on Effect in a Processing Network

1. Standard data access (bricks, slices, row-major, …)
2. Linear Streaming
3. Guided Streaming
4. Progressive Streaming
5. Adaptive Progressive Streaming

Cache oblivious raw data access

main memory local disk remote data

Data Layout (Cache Oblivious)

We characterize algorithmic classes based on effect in processing network.
The use of top-down and bottom-up processes have a strong impact on the data stream.

Progressive refinement: coarse representation immediately available

Challenge: minimize the quality differential

Benefit: pipeline of progressive modules

Decimation: full resolution data needed first
We Allow Distributed Computations at Different Stages of the Data Stream

- Progressive Image Differencing + Editable GPU filter.

Two data sources (11 GB each)

Progressive differencing + GPU edge detection
We are Developing Progressive Scheme for Content Based Image Processing

- Sample:

- Progressive Analysis:
Poisson Solver for Image Cloning in Massive Image Collections

- Color correction of 600+ images in real time
Poisson solver for composition of massive images

- Pasting a 300GB satellite image of a city in background world map merged in real time
Server can be wrapped in Apache plug-in
Client can be run in a web browser

One billion polygons to billions of pixels
Welcome to the first gigapixel, multi-view rendering of
The Digital Michelangelo Project's David

GIGAPIXELDAVID
High Performance Data Movements for Real-Time Monitoring of Large Scale Simulations

Scale simulation dumps to 130K cores with better performance than state of the art libraries while enabling real-time, remote visualization.

Simulation

ViSUS IDX File Format

Storage Nodes

Real-time Data Analysis and Remote Visualization

End User

[SC12a] Efficient Data Restructuring and Aggregation for IO Acceleration in PIDX
PIDX High Write Performance on Mira

![Graph showing performance comparison between PIDX, IOR Shared File (MPI Collective IO), and IOR File per Process (POSIX IO) across different core counts.]  
- PIDX: 187GB/s

Throughput (GB/Sec) vs. Cores (1K to 768K)

- PIDX
- IOR Shared File (MPI Collective IO)
- IOR File per Process (POSIX I/O)
Server can be wrapped in Apache plug-in
Client can be run in a web browser
A Scalable Solution for Acquisition and Processing High Resolution Data

Applications: Microscopy for Neuroscience

(1) Tissue Clearing (31 days)

Large Scale Data Management and High Performance Computing Fully Integrated in a Brain Connectomics Workflow.

(2) Data Source

(3) Preliminary Interactive Analytics

(4) Optimal Asynchronous Parallel Processing

(5) Interactive, Exploratory Assessment and Feedback
Remote Monitoring of Data Quality During Acquisition
High Performance Data Movements for Real-Time Access to Large Scale Experimental Data

- Experiment run at Advance Photon Source at ANL
- Materials Scientists at University of Utah
Scalable Deployment: Real Time Exploration of 3.5 Petabytes of Weather/Climate Data

Workflow

- **Data creation**
- **Data Management**

- Processing
- Analysis
- Visualization

**Workflow**

---

**Simulation Servers**

**IDX Data Server**

**IDX cache**

**Climate Data Converter**

**Interaction**

**Collaboration**

---

**Distributed Resources**

- 3.5 PB of data store in NASA
- Primary ViSUS server in LLNL
- Secondary ViSUS server in Utah

---

**Clients connect remotely**

**Work without additional HPC resources**

---

http://atlantis.sci.utah.edu/visus/webviewer/
nature_2007_aer1_hourly

---

- 7km GEOS-5 “Nature Run” -> 1 dataset, 3.5 PB
- theoretically: openly accessible -> practically: precomputed pics
Streaming Analytics and Visualization

Live demonstration from ANL to SLC

Infrastructure that scales gracefully with available hardware resources

Cores available
Demo Interactive Data Analysis with Python

- Reference: https://github.com/sci-visus/OpenVisus
- Jupyter examples: https://github.com/sci-visus/OpenVisus/tree/master/Samples/jupyter

- Installation with pip:
  - python -m pip install --user numpy OpenVisus==1.3.3

Demo

- Use standalone viewer and python scripting for interactive analysis
- Use Jupyter notebook to fetch data from a server visualize, analyze and share with other tools (e.g., with the standalone viewer)
High Resolution Display Platforms for High Resolution Outcrop and Seismic Data
ViSUS PowerWall: Installed and Fully Operational in a Few Hours at KAUST
ViSUS PowerWall: RIKEN
We Allow Distributed Computations at Different Stages of the Data Stream

- Progressive Image Differencing + Editable GPU filter.

Two data sources (11 GB each)

Progressive differencing + GPU edge detection
https://visoar.org/mamografia/
ViSOAR: a Unified Solution for Distribution of Imaging Data in Medical and Geology

Example of Visualization and Annotation of outcrop and medical data

Healthcare: remote access and diagnostics for Doctor and Patients on commodity devices

Mohamed E. Salama, MD, Chief of Hematopathology, Professor of Pathology
Teaching Histopathology with a Great Group of Students!!!

For more information see: www.visus.org
Computed Tomography Exploration:
Volume Rendering
An OpenViSUS Pipeline for Dynamic Data-Intensive Agricultural Applications
ViSUS Gigapixel Progressive Ditching for Aerial Imagery

ViSOAR

Registration & Stitching

Analytics
ViSUS Gigapixel stitching for aerial imagery
Farmers want actionable information...

To achieve this goal and enable wide-spread adoption of aerial imagery use in crop management there is an urgent need for real-time image stitching on high-latency low-bandwidth networks in rural areas.
OpenVisus references

- Main website: https://www.visus.org/
- Documentation website: https://wiki.visus.org/
- Gitter chat: https://gitter.im/sci-visus/OpenViSUS

- PIDX code: https://github.com/sci-visus/PIDX/
- OpenViSUS code: https://github.com/sci-visus/OpenVisus
- Jupyter examples: https://github.com/sci-visus/OpenVisus/tree/master/Samples/jupyter