Big Data Visual Analysis

Chris Johnson
Scientific Computing and Imaging Institute
University of Utah
History of Computer Graphics in Utah

1. David Evans / Ivan Sutherland
   - Founded CS Dept at the U of U in 1963
   - Ivan Sutherland - Turing award
   - Founded Evans & Sutherland Company

2. John Warnock
   - Worked at Evans & Sutherland
   - Founded Adobe
   - Hidden Line Removal Algorithm
   - Helped invent Postscript at Adobe

3. Tom Stockham
   - Known for work in Signal Processing
   - Helped to invent the CD Player

4. Ed Catmull
   - Worked at Lucas Film
   - Co-Founded Pixar
   - President of Disney Animation Studios
   - Chair of CoE External Advisory Board

5. Alan Kay
   - Personal Computer
   - Turing Award Winner
   - Object-Oriented Languages

6. Jim Kajiya
   - VP Research at Microsoft

7. Jim Clark
   - Founded SCI, Netscape, Honda
   - Work in Geometry Pipelines

8. Nolan Bushnell
   - Invented Pong
   - Founded Atari

9. Allen Ashton
   - Word Perfect
   - My CFO Founder

10. Bui Tuong Phong
    - Invented Phong Reflection and Shading Models

11. Henri Gouraud
    - Invented Gouraud Shading Model

12. Ed Catmull
    - Worked at Lucas Film
    - Co-Founded Pixar
    - President of Disney Animation Studios
    - Chair of CoE External Advisory Board

13. Alan Kay
    - Personal Computer
    - Turing Award Winner
    - Object-Oriented Languages

14. Jim Kajiya
    - VP Research at Microsoft

15. Jim Clark
    - Founded SCI, Netscape, Honda
    - Work in Geometry Pipelines

16. Nolan Bushnell
    - Invented Pong
    - Founded Atari

17. Allen Ashton
    - Word Perfect
    - My CFO Founder

18. Bui Tuong Phong
    - Invented Phong Reflection and Shading Models

www.sci.utah.edu
Centers We Direct

NIH/NIGMS Center for Integrative Biomedical Computing

Utah Center for Neuroimage Analysis

UTAH Center for Computational Earth Sciences

Center for Extreme Data Management, Analysis, and Visualization (CEDMAV)

NVIDIA CUDA Center of Excellence at the Scientific Computing and Imaging Institute

Alliance for Computationally Guided Design of Energy Efficient Electronic Materials
National Centers We are Affiliated With

SDAV
Scalable Data Management, Analysis and Visualization

NIH NAMIC
NATIONAL INSTITUTES OF HEALTH

IAMCS
Institute for Applied Mathematics and Computational Science

CDC Decision-Support for Infectious Disease Epidemiology

EXACT
Center for Exascale Simulation of Combustion in Turbulence
Every two days we create as much data as we did from the beginning of mankind until 2003!
How Much is an Exabyte?

How many trees does it take to print out an Exabyte?

1 Exabyte = 1000 Petabytes = could hold approximately 500,000,000,000,000 pages of standard printed text

It takes one tree to produce **94,200** pages of a book

Thus it will take **530,785,562,327** trees to store an Exabyte of data

In 2005, there were **400,246,300,201** trees on Earth

We can store **.75** Exabytes of data using all the trees on the entire planet.

Sources: http://www.whatsabyte.com/ and http://wiki.answers.com
Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

Dan Ariely
Brain Information Bandwidth

- sight
- touch
- hearing/smell
- taste

Amount we're actually aware of (0.7%)
New Visual Analysis Techniques
James Watson and Francis Crick - 1953
Nobel Prize - 1962
X-ray diffraction data from Maurice Wilkins and Rosalind Franklin
Mario R. Capecchi, Ph.D., distinguished professor of human genetics and biology at the University of Utah’s Eccles Institute of Human Genetics has won the 2007 Nobel Prize in Physiology or Medicine.
Statistics of Shape, Connectivity, and Function

Computational Statistics in Nonlinear Spaces

Anatomical shape averaging and variability

Diffusion Tensor Image Analysis Autism project

Combined PET + MRI analysis Alzheimer’s disease project
Volume Rendering

enamel / background

dentin / background

dentin / enamel

dentin / pulp

1D: not possible
2D: specificity not as good
Volume Visualization

Scientific Computing and Imaging Institute, University of Utah
NIH Visible Male

Scientific Computing and Imaging Institute, University of Utah
Visible Human - High Resolution

Scientific Computing and Imaging Institute, University of Utah
The Need for High Resolution Visualization

“...the data show for the first time how detailed transport and chemistry effects can influence the mixing of reactive scalars. It may be advantageous to incorporate these effects within molecular mixing models. It is worth noting that at present it is impossible to obtain this type of information any other way than by using the type of highly resolved simulation performed here.”

Jacqueline Chen, Sandia National Laboratories
Topological Analysis of Massive Combustion Simulations

- Non-premixed DNS combustion (J. Chen, SNL): Analysis of the time evolution of extinction and reignition regions for the design of better fuels
ImageVis3D - Mobile

Sharing
Dataset: vishuman-ct.uvf
TransferFunction: vishuman-ct.1dt

The server is running on port 4000. Connect ImageVis3D Mobile to this system now.

Network Setup  Start  Close

Startpoint: 4000
The server is running on port 4000. Run the iPhone app now.
Add To List  Remove From List

bbit3.0m / bbit3.0t.png
bbit3.0m / bbit3.0t.png

ImageVis3D Mobile
Neuromodulation Research in Butson Lab

Transcranial Magnetic Stimulation for Depression

Deep Brain Stimulation for Traumatic Brain Injury: Primate, Human & Mouse

Cortical Stimulation for Depression

Deep Brain Stimulation (DBS) for Movement Disorders
Deep Brain Stimulation
ImageVis3D Mobile DBS App

Deep Brain Stimulation

DBP: Chris Butson

C. Butson, G. Tamm, S. Jain, T. Fogal and J. Krüger
Visualization of 10D Combustion Simulation of Jet CO/H2-Air Flames

10 dimensional data set describing the heat release wrt. to various chemical species in a combustion simulation.

- Local Extinction
- Pure Oxidizer
- Pure Fuel
Analysis of Combustion Simulations

Combustion Simulation of Jet CO/H$_2$-Air Flames

**Input:** Composition of 10 chemical species

**Output:** Temperature
Interactive Streamline Exploration and Manipulation using Deformation

Xin Tong\textsuperscript{1}, John Edwards\textsuperscript{2}, Chun-Ming Chen\textsuperscript{1}, Han-Wei Shen\textsuperscript{1}, Chris R. Johnson\textsuperscript{2}, Pak Chung Wong\textsuperscript{3}

\textsuperscript{1}The Ohio State University
\textsuperscript{2}Scientific Computing and Imaging Institute, University of Utah
\textsuperscript{3}Pacific Northwest National Laboratory
One billion polygons to billions of pixels
Welcome to the first gigapixel, multi-view rendering of
The Digital Michelangelo Project's David
Gigapixel David - iTunes App

This app is designed for both iPhone and iPad.

No Ratings
Rated 17+

GigapixelDavid
Valerio Pascucci

Welcome to the first gigapixel, multi-view rendering of The Digital Michelangelo Project's David.

The David model consists of 833 million triangles from a laser-scan of the original statue created by Professor Marc Levoy and members of The Digital Michelangelo Project at Stanford University. The original data is made of 8 million polygons, each about 20 mm in size, smaller than the thicknesses of 2 dime.

Each of the four views presented in this Gigapixel David app consists of 3 gigapixel stand frames (2928 x 704 x 18 pixels) rendered using the Manta Interactive Ray Tracer by researchers of the Scientific Computing Institute (SCI) at the University of Utah. In all, each frame took 30 hours to render using 64 cores each 2.56 GHz of the 81-institute 294 core Sun Ultra 1000 with 2.88GHz of RAM and 2.87GHz Intel Xeon 37542 saves. More information on Manta can be found at: http://researchwiki.sci.utah.edu/manta/index.php/Manta_Page

The final rendering is stored in using VISUS raw data format enabling efficient, streaming pipelines that process the information while in movement. This technology enables real-time management of large datasets on a variety of systems ranging from desktops and laptop computers to portable devices such as iPhones/iPad. VISUS has been deployed in a variety of large data applications such as the monitoring of large scientific simulations and the editing of massive images and panoramas.

The VISUS Gigapixel David viewer is also currently available as a Windows web browser plugin (Firefox and Chrome) or as a standalone application for Windows, Mac OS X, or OpenOSSL, available via http://www.sci.utah.edu/visus.html
341 Sections
90nm thick sections
~32GB/Section
~1000 tiles/section
4096x4096 pixels/tile
2.18 nm/Pixel
16.5 TB after processing
... my work, which I've done for a long time, was not pursued in order to gain the praise I now enjoy, but chiefly from a craving after knowledge, which I notice resides in me more than in most other men. And therewithal, whenever I found out anything remarkable, I have thought it my duty to put down my discovery on paper, so that all ingenious people might be informed thereof.

Antony van Leeuwenhoek. Letter of June 12, 1716
Connectome
PROBLEM-DRIVEN VISUALIZATION RESEARCH for biological data

- target specific biological problems
- close collaboration with biologists
- rapid, iterative prototyping
- focus on genomic and molecular data
Genome-wide synteny through highly sensitive sequence alignment: Satsuma
M. Grabherr, et al.
Visualization of Biological Data

MizBee
Browser that enables analysis of comparative genomics data through visualization across multiple scales.
Uncertainty Visualization

When is the last time you’ve seen an error bar in a 3D visualization?
Uncertainty Visualization

Surfaces imply certainty
Uncertainty Visualization

Surfaces imply certainty
Uncertainty Visualization

Surfaces imply certainty
Topological Uncertainty

Figure 6: Increasing the uncertainty of a random vector field: a) certain topology of mean vector field; b) $\|T\|_F = 0.2$; c) $\|T\|_F = 5.0$.

ProbVis

**Visualization & Exploration for Distributions**

- Show differences between PDFs
- Summarize all data in a single view

Visualizing Uncertainty

Fuzzy Sensitivity Confidence

QuizLens: A Multi-lens approach for uncertainty exploration

- Global information important for qualitative evaluation & context
- Local information necessary for quantitative understanding
- Interchangeable lenses to explore various data characteristics
Ensemble-Vis: A Framework for the Statistical Visualization of Ensemble Data

### What is ensemble data?
Collection of data sets generated by computational simulations.

Used to simulate complex systems, mitigate uncertainty, unknowns in initial conditions, and parameter sensitivity.

These data sets are:
- Multidimensional
- Multivariate
- Multivalued

### Ensemble-Vis Workflow
- User-driven
- Component-based

### Spatial Overviews:
Mean and standard deviation encoded through colormaps and contours.

### Temporal Overviews:
Filmstrip and animation. Small multiples show every time step. User can select desired temporal location.

### TREND CHARTS:
Show minimum and maximum, innerquartile range.

### QUARTILE CHARTS:
Show minimum and maximum, innerquartile range.

### PLUME CHARTS:
Show every member and mean. Color coded based on model. Deselect members to hide. Drill-down to direct data display.

### QUERY CONTOURS:
- User-driven query
- Select subset based on conditions
- Returns % of satisfying members
- Displayed as nested contours

### SPAGHETTI CHARTS:
- Show variation across space
- User chosen contour value
- Isocontour for each desired member
- Highlights outliers and divergence

### Implementation
- **SREF Weather Explorer**
  - VTK filters, Qt Widgets
  - Relational database:
    - MySQL/ Netezza

- **VISUS**
  - Climate Data Analysis Tools (CDAT) integration
  - C++, python, FLTK
  - Out-of-core, streaming
Productivity Machines
Acknowledgments

NIH/NIGMS Center for Integrative Biomedical Computing

CEDMAV

Utah Center for Neuroimage Analysis

SDAV
Scalable Data Management, Analysis and Visualization

NIH NAMIC

IAMCS
Institute for Applied Mathematics and Computational Science

Center for Extreme Data Management, Analysis, and Visualization