April 15, 2015

Getting Started With High Performance Computing for Humanities, Arts, and Social Science

Extreme Science and Engineering Discovery Environment

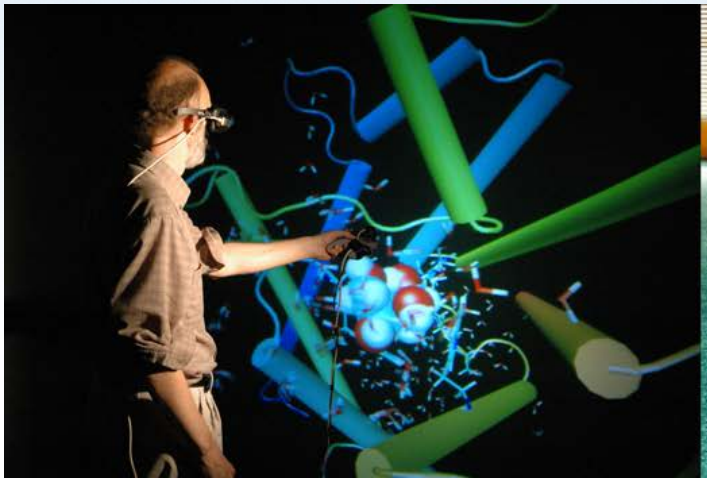
Alan B. Craig, PhD
acraig@ncsa.uiuc.edu

Who am I?

- HASS Specialist for XSEDE
- Senior Associate Director for I-CHASS
- Research Scientist at NCSA (28 years)

- I am a point of contact for you to help engage with appropriate experts

What are my interests outside of XSEDE?



Virtual Reality

Augmented Reality

Personal Fabrication

Visualization
Representation of Information
Human-Computer Interaction

The Order of this Presentation

- Today will be *extremely informal.* Let's have a dialog, not a one way presentation

- Concepts
- Application Areas
- Mechanics
- Application Examples throughout

First – Some Definitions

- HASS
- HPC
- Supercomputer
- Data
- Model
- XSEDE
- ECSS

Why would anyone in HASS want to use HPC?

- Humans are good at certain things
  - Reasoning
  - Interpretation
  - And more
- Computers are good at certain things
  - Repetitive Tasks
  - Identifying potential anomalies
  - Some reasoning, interpretation
  - Identifying relationships
  - And more

Why would anyone in HASS want to use HPC?

- It's an issue of *scale*
- It's an issue of *scale*
- It's an issue of *scale*

- *Scale of amount of data*
- *Scale of amount of computation*
- *Scale of amount of storage*
- *Scale of problem scope*

Scale Example

Scale Example

Scale Example

Scale Example

Scale Example

Time to get busy!!!

It Takes a Team

- Large projects often require interdisciplinary team
- Content Expertise
- Computational Expertise
- Visualization Expertise
- Project Management Expertise

What is Data?

- Data is information
- In our case it is information in electronic form
- Data can be:
  - Numeric
  - Text
  - Visual Image
  - Audio
  - Video
  - Etc. (any kind of signals)

Representation of Information

- Data can be manipulated in useful ways
- Computer can help with this
- We need to choose good representations for different purposes, and have the computer do the work
- We can transform from and to different representations

What Are Some Things People Do?

- Text Analysis
- Image Analysis
- Video Analysis
- Network Analysis
- GIS
- Simulation
- Visualization
- Display and Interaction

Text Analysis

- Statistics (word counts, co-occurrences, etc.)
- Entity recognition
- Clustering / categorization
- Etc.

- Genre identification
- Topic Modeling
- Sentiment Analysis
- Etc.

Text Analysis

- Find things in a collection of text
- What is in this collection of text?
- Machine Learning applications

Text Analysis

- Interesting tools:
  - Mallet
    - MAchine Learning for LanguagE Toolkit
    - Topic Modeling
    - Installing on Blacklight at PSC
    - http://mallet.cs.umass.edu
- We can discuss appropriate tools for your project. Different strengths and weaknesses
- Some scale better, different capabilities…

Image Analysis

- What is in this image?
- Find images that…
- OCR
- Who painted this?
- Who wrote this manuscript?
- Again, machine learning often used

Image Analysis – Authorship Example

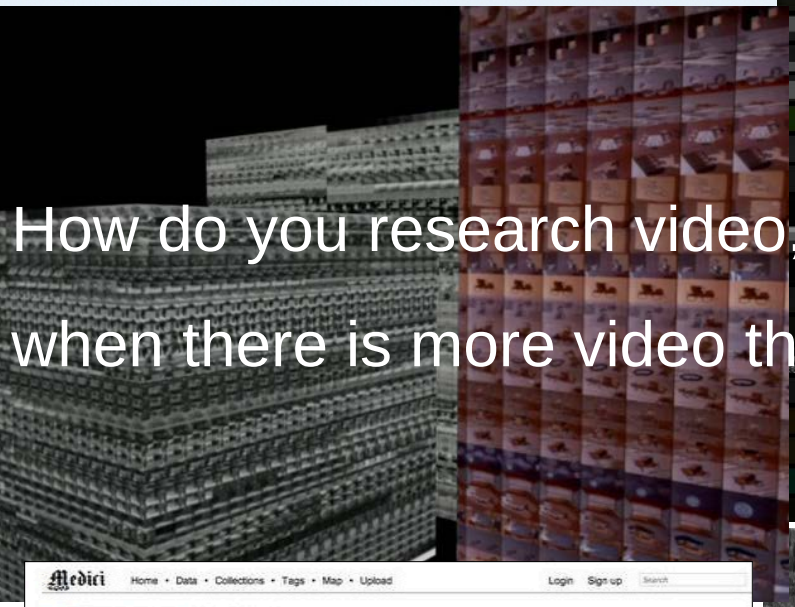
15th-century manuscripts, 17th and 18th-century maps, and 19th and 20th-century quilts

ISDA - DID

Video Analysis

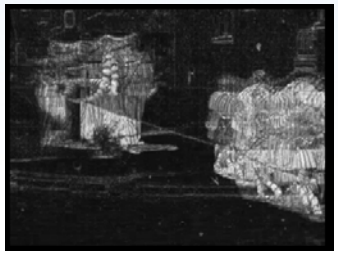
- Video is being created far faster than we can view it – EG – YouTube, Surveillance, etc.
- It is more than just the case of multiple images
- Scene Identification
- Contents – EG Phones
- Cinematographic Elements – Camera moves, lighting, etc.
- Visualization of movies

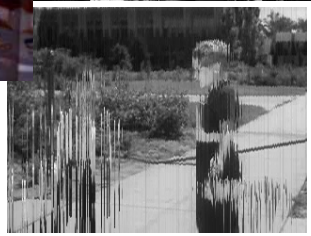
Large Scale Video Analytics: On-Demand, iterative inquiry for moving image research

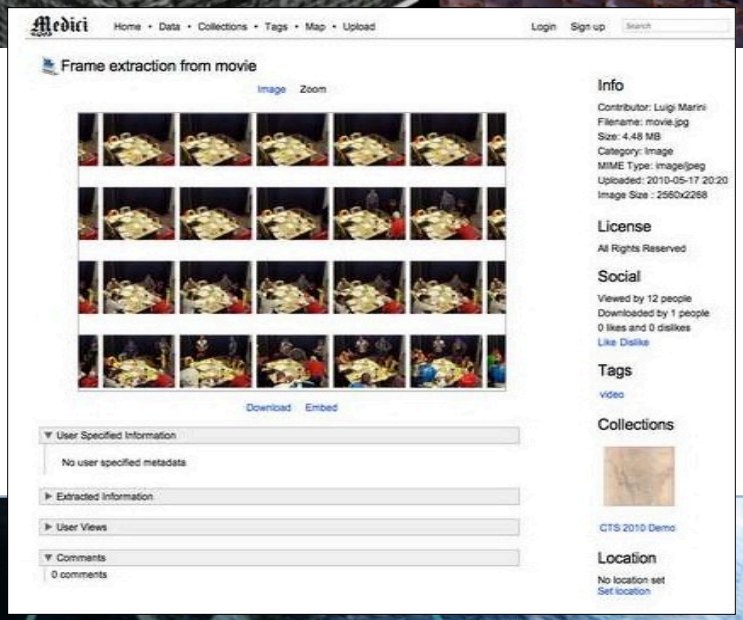


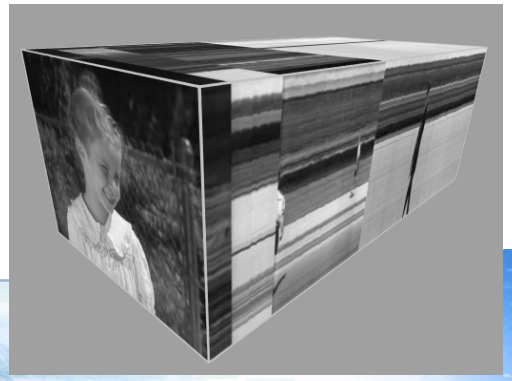
How do you research video,

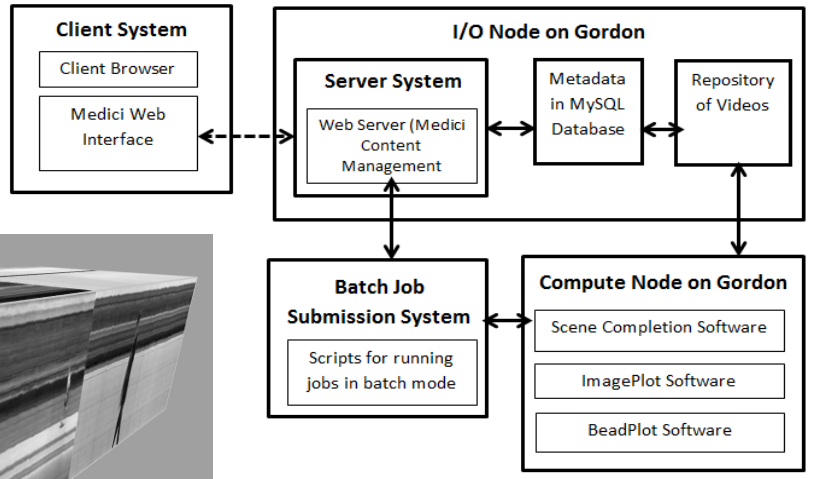
when there is more video than can ever be watched?











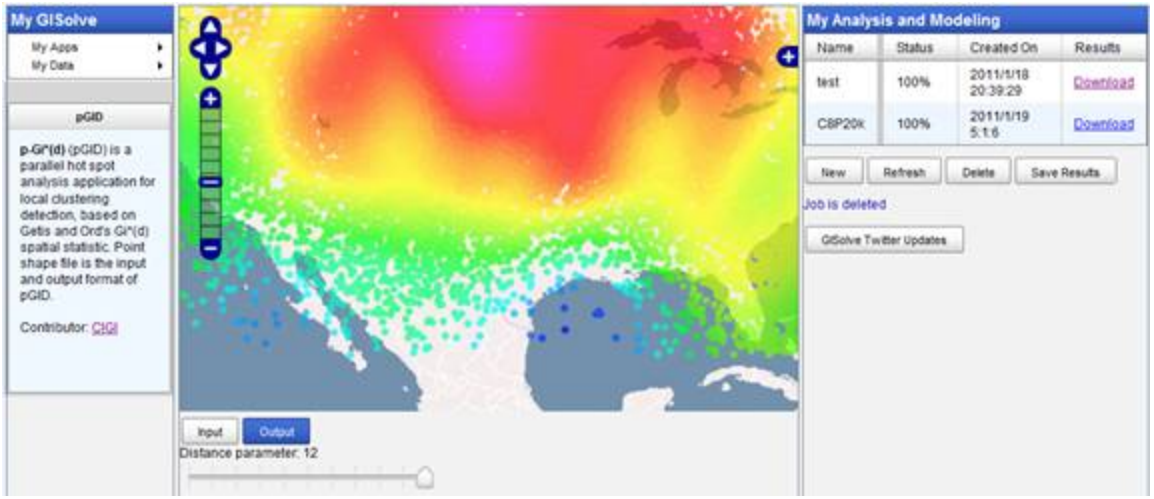
XSEDE

Network Analysis

- Looks at relationships
- Who is connected to who?
- Who is connected to where?
- Etc.
- Think about facebook… think about twitter… think about other social media….
- Think about world events and news items…

GIS

- Map based information
- Spatial Studies
- Often combined with others listed here

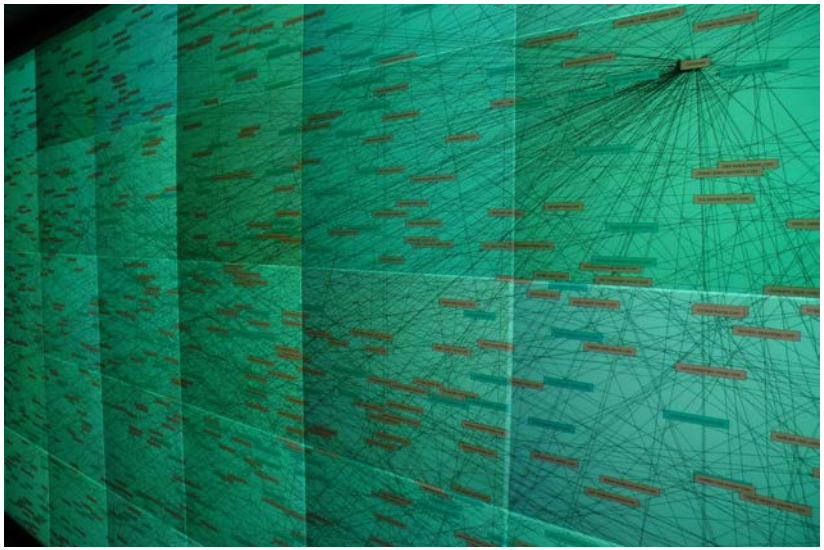


Simulation

- One of traditional uses of supercomputers for scientists
- Can simulate events to see how they might play out
- Think about propagation of ideas as diffusion…
- Historical counterfactuals
- Other ideas?

Visualization

- Represent data visually (and other senses)
- Can show relationships
- Can show the unseen

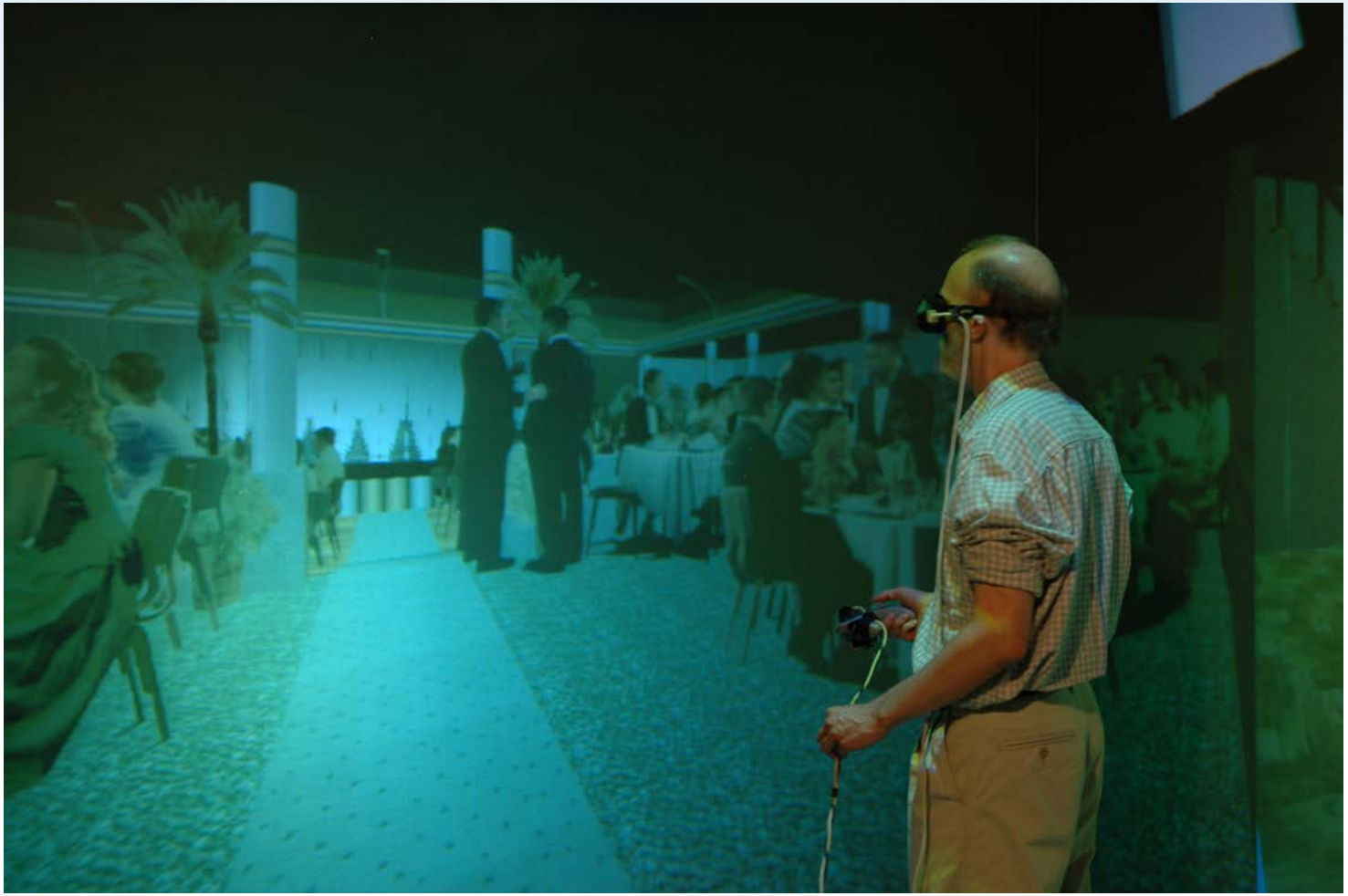


Display and Interaction

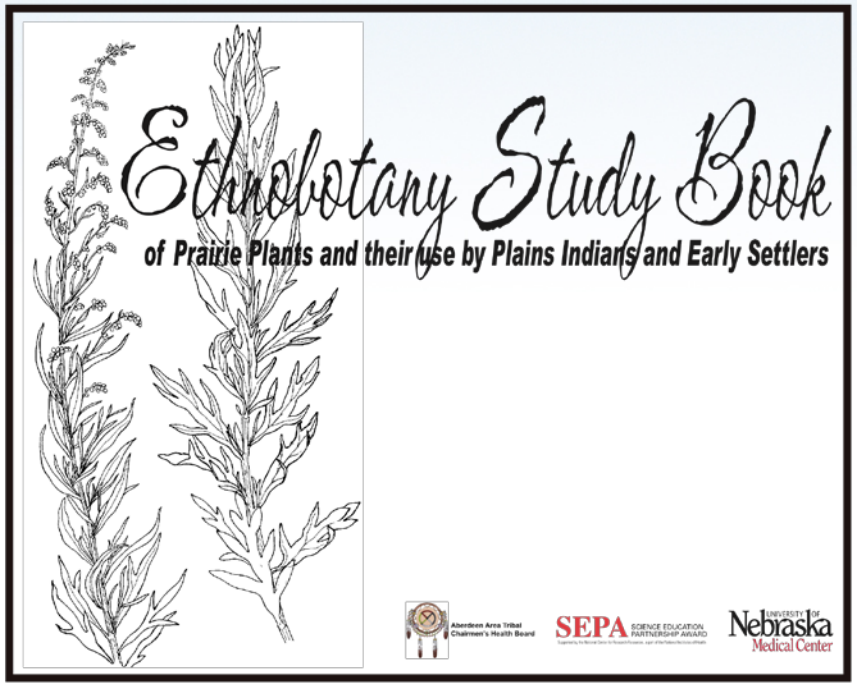
- Virtual Reality
  - Purely synthetic environment
  - Bodily engaged
- Augmented Reality
  - Real world combined with digital
  - Bodily engaged
- Note that Facebook just bought Oculus for $2B
- Microsoft bought $150M in VR and AR patents

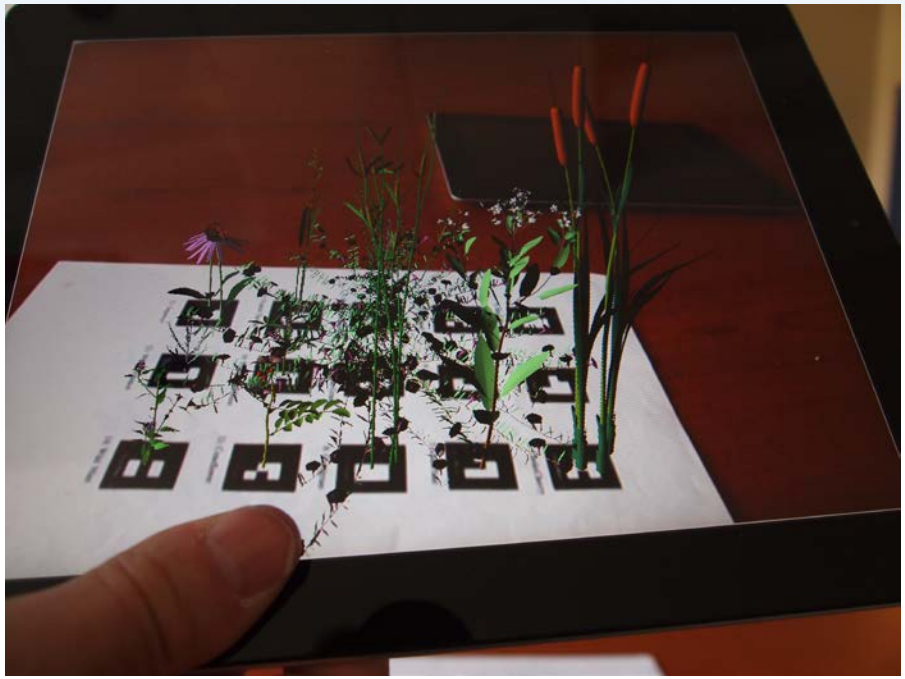
VR Example – Harlem in 1920s

VR Example – Harlem in 1920s



Augmented Reality Examples / Demos





Gateways

- Strategy for building communities, lowering barrier to entry
- GIS gateway already exists
- Building Video Analysis Gateway
- Building Text Analysis Gateway
- Building Network Analysis Gateway

Current state of HASS and HPC

- There are HASS projects using XSEDE successfully
- Many others are interested, but haven't taken the leap yet
- Some HASS problems don't fit the XSEDE mold…. We are changing the mold!
- XSEDE / HPC Communities can benefit from input from HASS Communities

Four categories of researchers

- Have code, have expertise
- 3rd party codes, may or may not be on HPC
- No code, but great idea
- No idea what to do but interested

- We can work with all of these folks

Language issues between HASS and HPC

- We speak different languages… words overlap
  - Model
  - Data
  - Simulation
  - Big
- We have different expectations
  - Black and white vs. grey answers
  - Researcher fit to the computer vs. the computer fit to the researcher

The workflow issue

- Batch vs. real time / on demand interactive

Technical hurdles in HASS computing

- Data / Digitization
- Application development

- There is help available!

Getting Started With XSEDE

- Contact me!
- Resources are free! (by proposal)
- Technical assistance available

April 15, 2015

Thanks for listening!

Alan B. Craig, Ph.D.

Associate Director, Human-Computer Interaction

Institute for Computing in Humanities, Arts, and Social Science

Research Scientist

National Center for Supercomputing Applications

University of Illinois

acraig@ncsa.uiuc.edu

XSEDE

Extreme Science and Engineering Discovery Environment