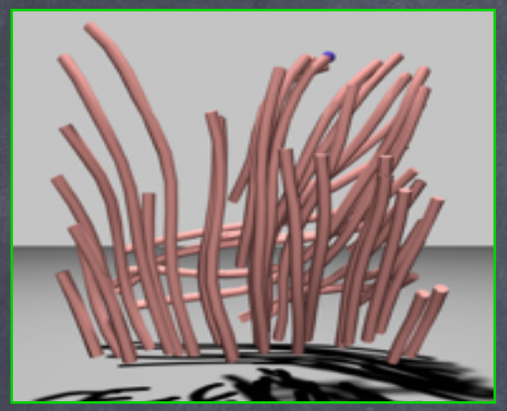
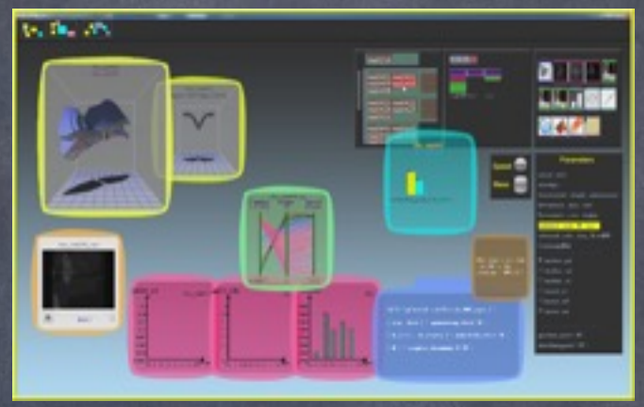
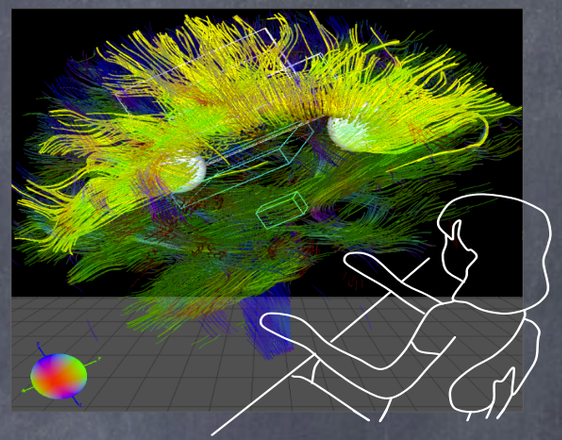
Accurate Visualization for Knowledge Discovery in Big-Data Science



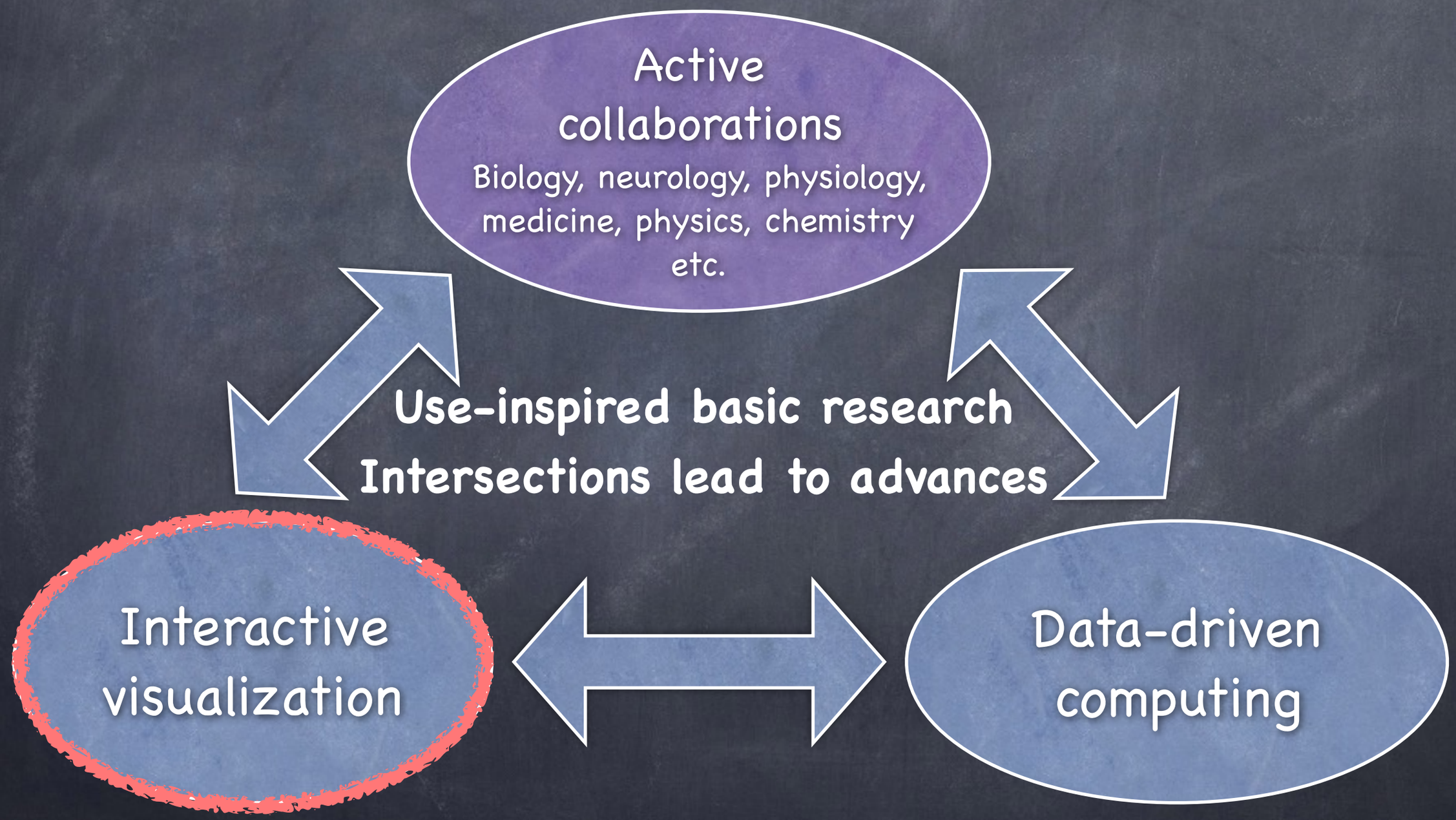




Jian Chen, PhD
http://www.csee.umbc.edu/~jichen
Interactive Visual Computing Laboratory (http://ivcl.umbc.edu)

We acknowledge support from:
NSF, "Supporting knowledge discovery through a 3D scientific visualization language."
NSF, "PathBubbles for dynamic visualization and integration of biological information."
NIST, "Understanding immersive metrology datasets: scientific and information visualization integration and hybrid input."
DoD, "An interactive visualization framework to support exploration and analysis of TBI/PTSD clinical data."

Research Program



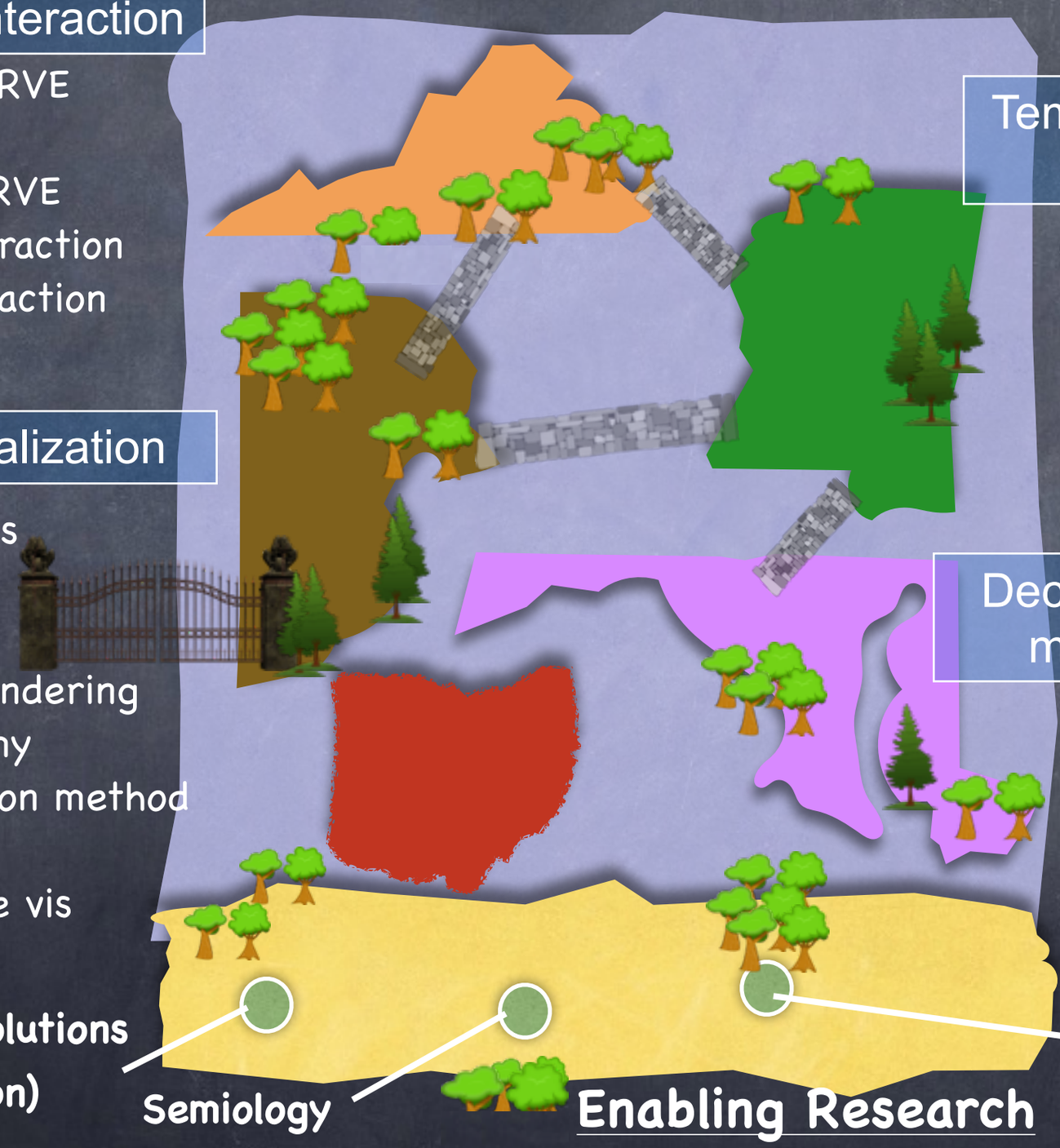
Chen's Garden of Use -Centered Visual Computing

Physical-spatial interaction

Resolution+size on IRVE
Stereo on DMRI
Text navigation in IRVE
Domain-specific interaction
Multi-platform interaction
Display ecology

Theory of Visualization

Language of DMRI vis
Coloring DMRI
Ranking encoding
Non-Photorealistic rendering
Visualization taxonomy
Visualization evaluation method
Visual feature space
Perceptually accurate vis



Temporal-dynamic workflow

Memory-driven
Schema-free exploration

Decision making from multifaceted data

Dimension reduction
Cost of context switching
Tool design

Chen's Garden of Human-Centered Visual Computing

Physical-spatial interaction

Resolution+size on IRVE
Stereo on DMRI
Text navigation in IRVE
Domain-specific interaction
Multi-platform interaction
Display ecology

Theory of Visualization

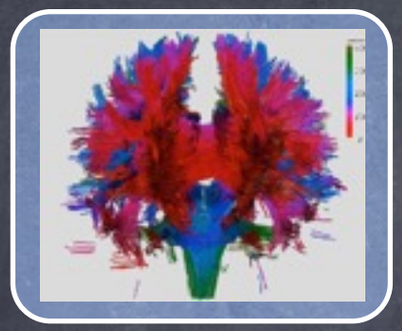
Language of DMRI vis
Coloring DMRI
Ranking encoding
Non-Photorealistic rendering
Visualization taxonomy
Visualization evaluation method
Visual feature space
Perceptually accurate vis

Temporal-dynamic workflow

Memory-driven
Schema-free exploration

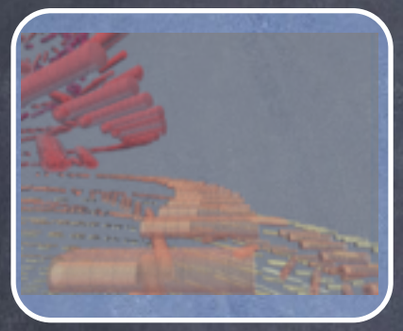
Decision making from multifaceted data

Dimension reduction
Cost of context switching
Tool design



1. A scientific visualization **language** for diffusion-tensor MRI visualization

Descriptive framework of seeing

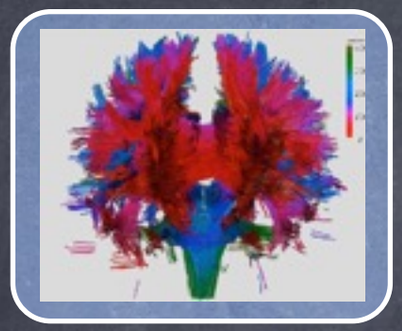


2. **Experiment**: design and evaluation of perceptually accurate visualization

Experiments

3. Workflow-driven design for time-varying bat flight **analysis**

Exploratory workflow



1. A scientific visualization **language** for diffusion-tensor MRI visualization

Descriptive framework of seeing

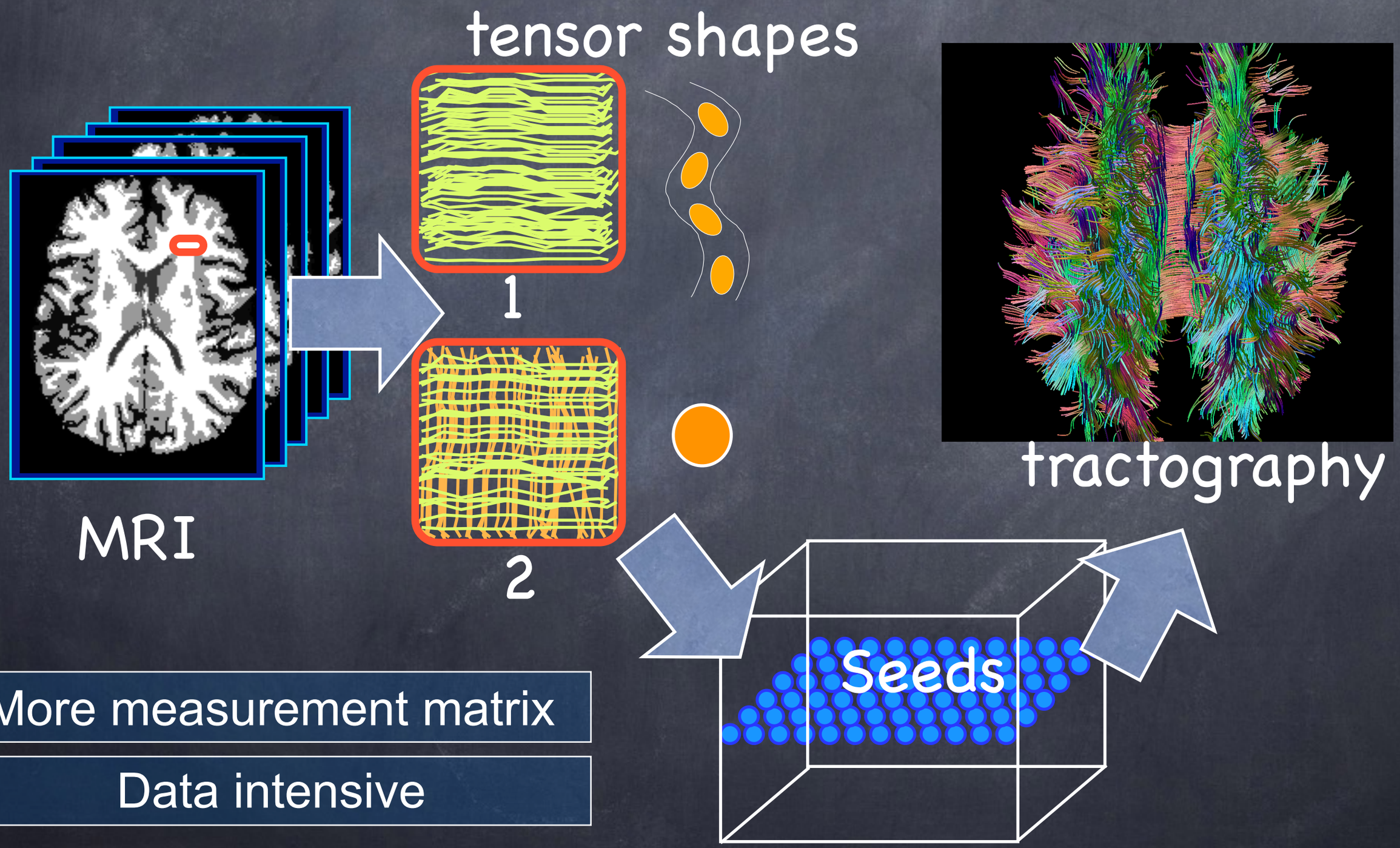
Collaborators:

Computer science: David H. Laidlaw (Brown)

Neurology: Alexander P. Auchus (UMMC)

- Research questions: how does visualization support **seeing**? and what do scientists **see** from mountains of data?

Diffusion-tensor MRI



More measurement matrix

Data intensive

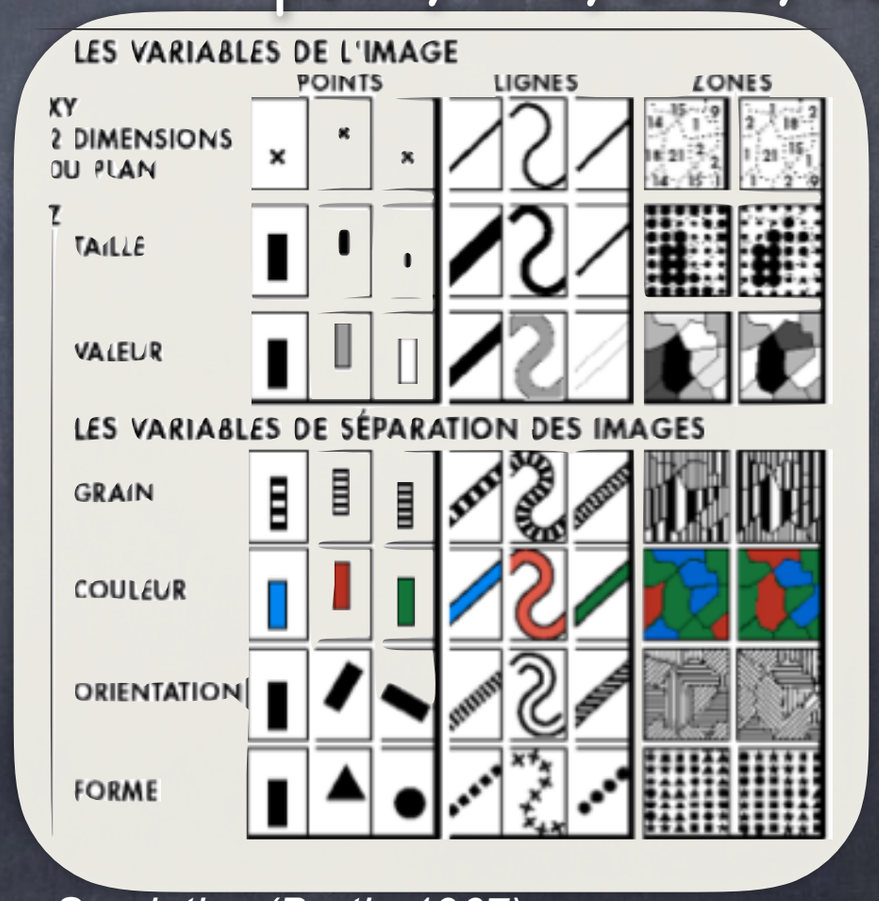
Extending Bertin's Sign Theory

data -> graphics signs

Applied to InfoVis by Mackinlay (Stanford, 1986, Tableau, 2012), Fry (MIT/Harvard, 2006), and Heer (Berkley 2007, UW, 2016).

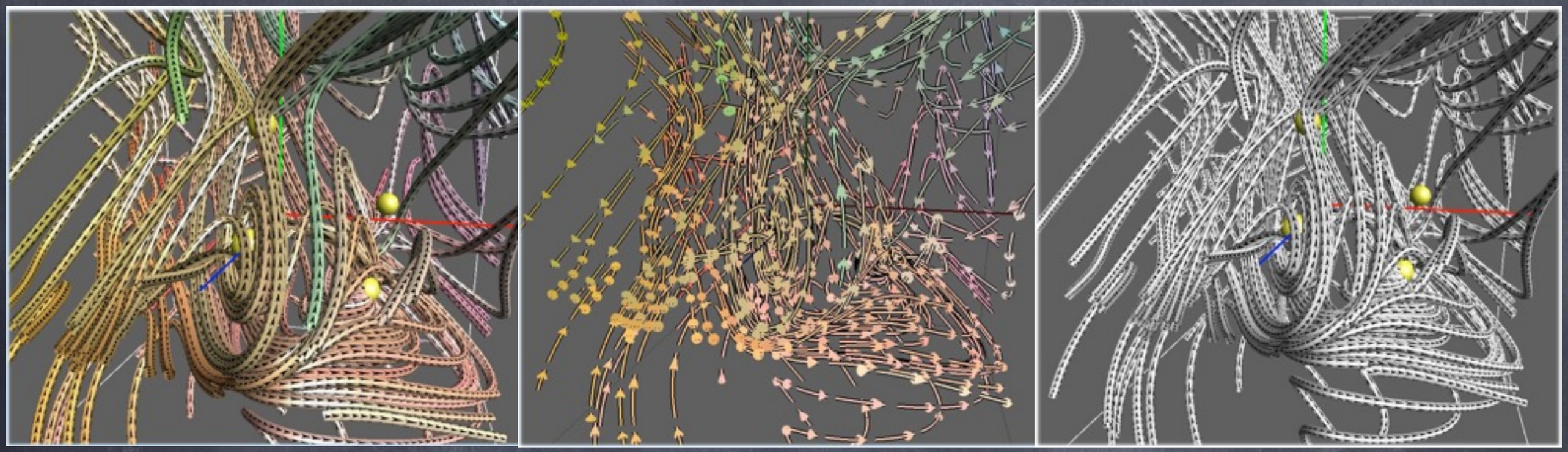
- Angular variables
  - shadow
  - depth
- Retinal variables
  - position
  - size
  - value
  - texture
  - color
  - orientation
  - shape

point, line, area, volume



Semiotics (Bertin 1967)

J. Chen,, On the semiological analysis of diffusion tensor field visualizations, IEEE VIS 2016, IEEE TVCG in submission.



flow direction -> color
flow speed -> texture
size

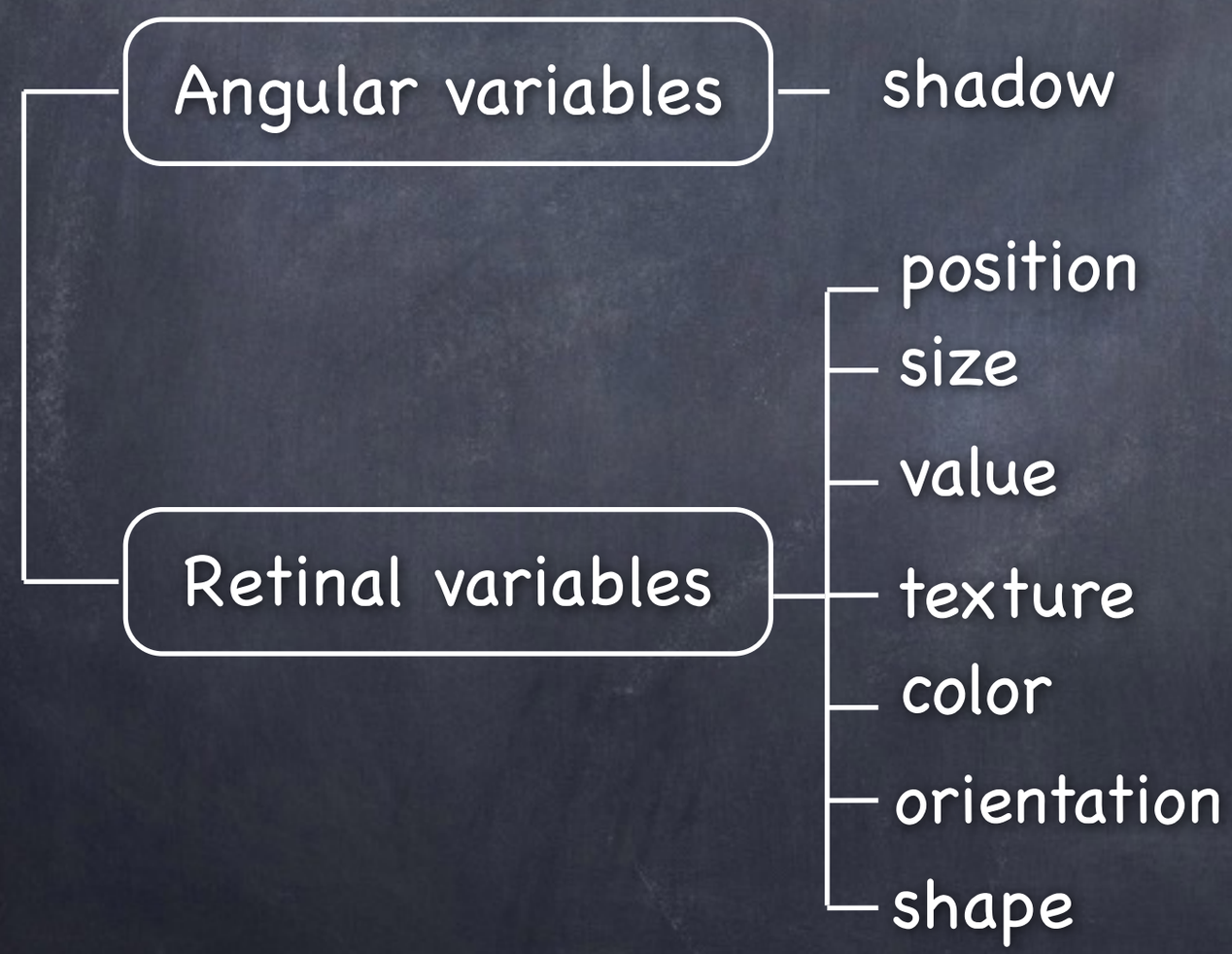
flow direction -> color
flow speed -> shape

flow speed -> texture
size

A.S. Forsberg, J. Chen, and D. H. Laidlaw, Comprising 3D vector field visualization methods: a user study, 15(6): 1219-1226, IEEE Transactions on Visualization and Computer Graphics (TVCG), 2009.

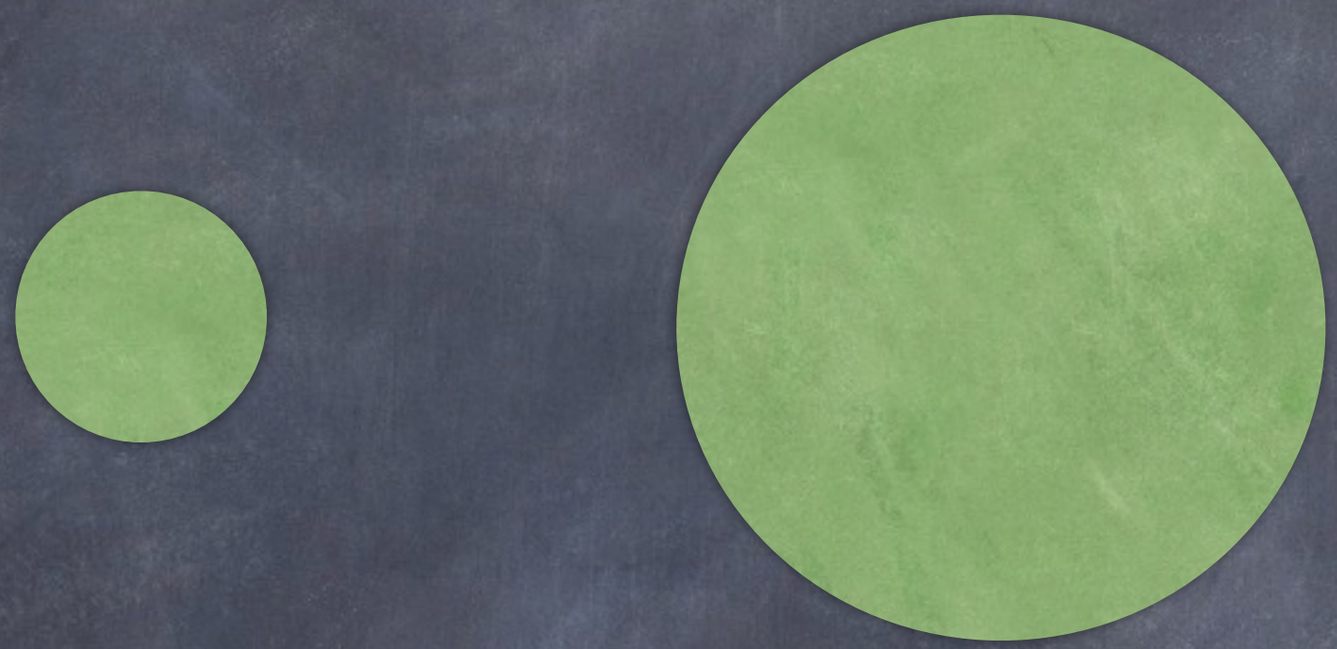
How to Study These Dimensions?

- Design space must inform design (visualization technique and problem solving environment)

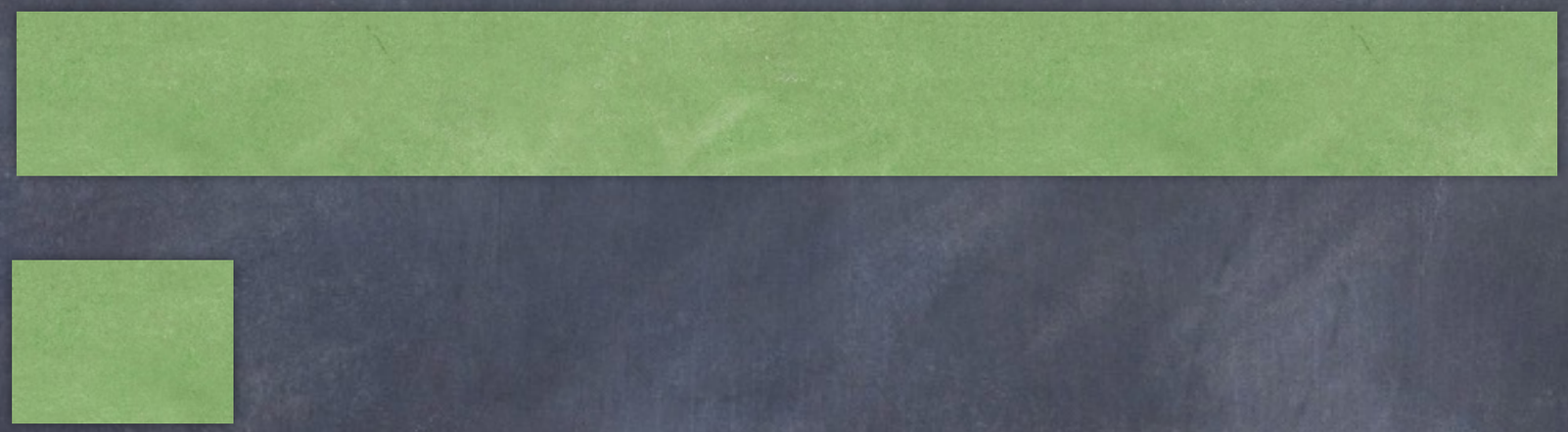


Which dimensions are most important?

Are these the right level of representation in a problem solving environment?

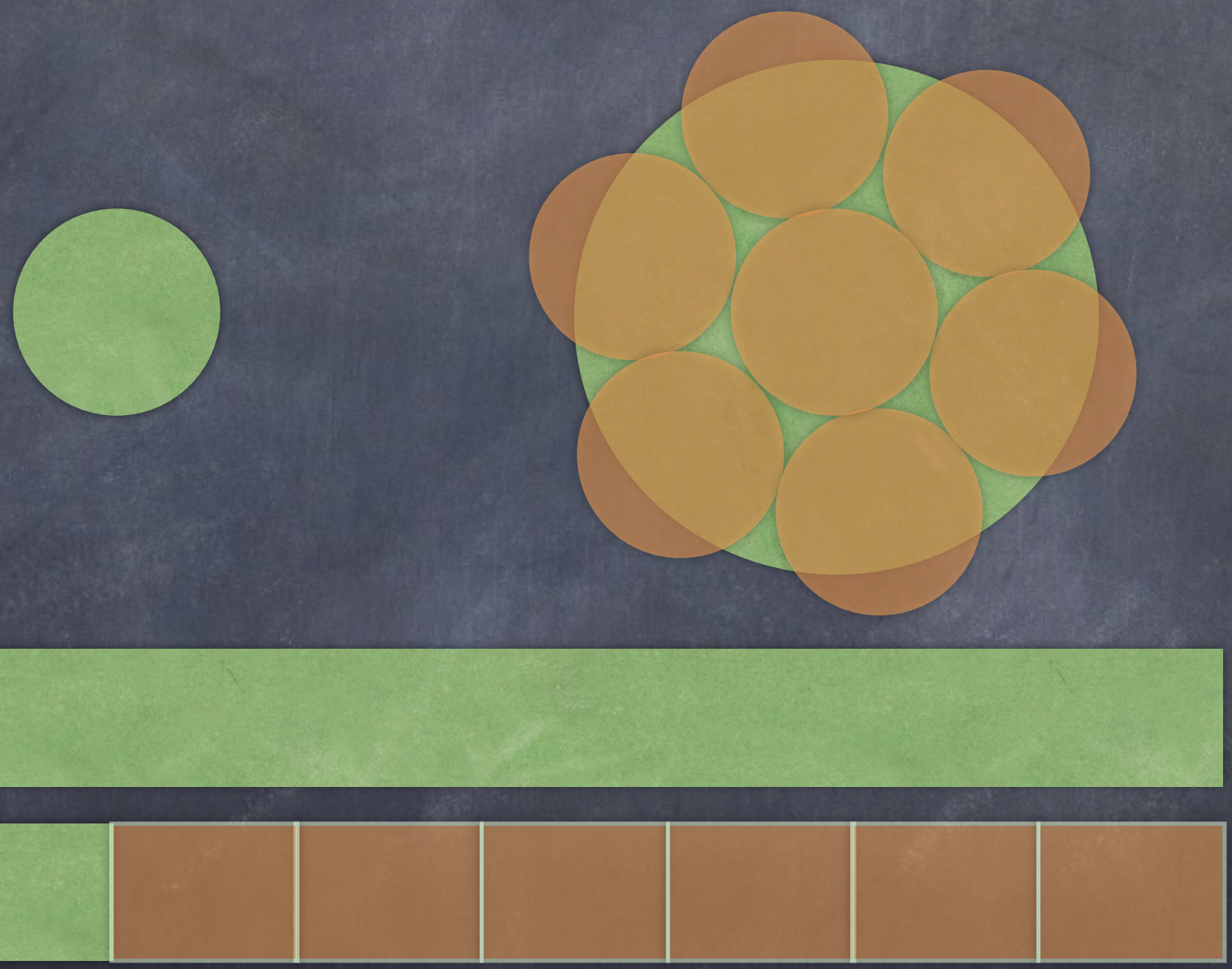


Compare **area** of circles



Compare **length** of bars

Human Perception is Inaccurate!

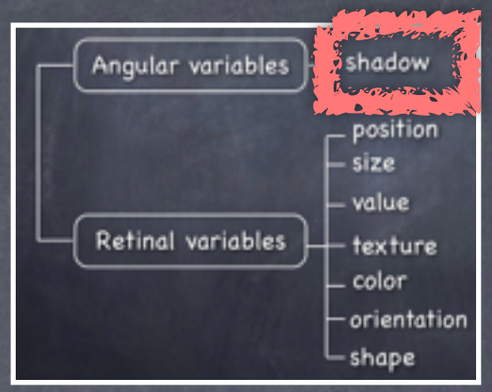


Length is more effective than area to represent quantities.

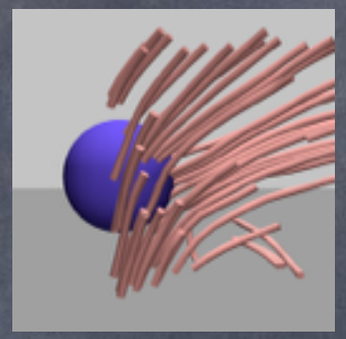
Our Approach

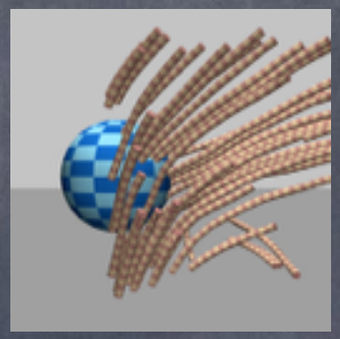
- Strongly hypothesis-driven experimentation
- End-to-end, breadth-first reciprocal research strategy
  - Deliver computational benefits immediately (build human-in-loop quality control for all algorithms to ensure trustworthy data features for down-stream research)
  - Psychophysics informs research on better visual features
- Corpus collection & data-driven research

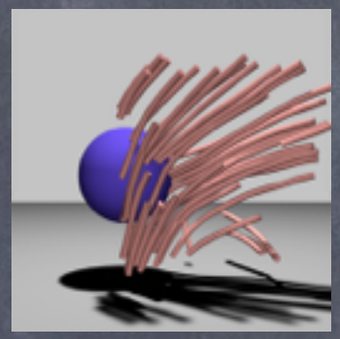
Illumination Model Comparison

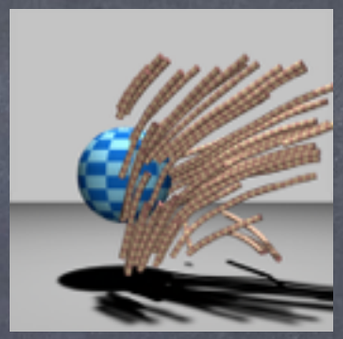


1. Illumination models
2. Texture
3. Motion

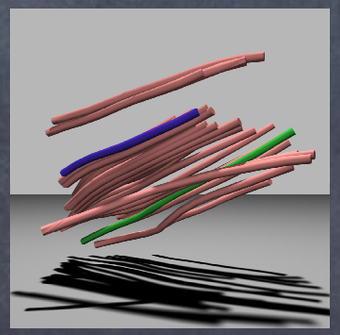

OpenGL

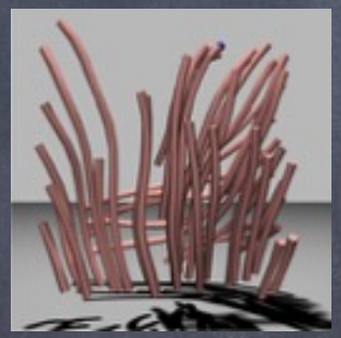

OpenGL+Texture

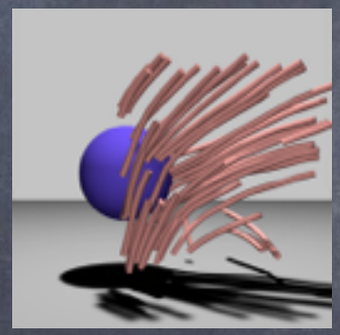

GI


GI + Texture

Three tasks


Depth Judgment


Visual Tracing

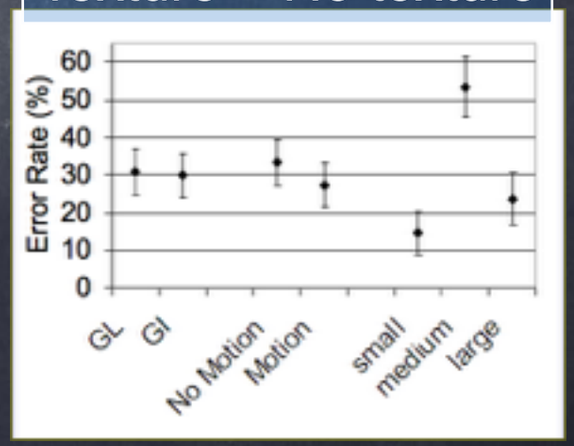

Contact Judgment

Hypotheses

GI > OpenGL

Motion > No motion

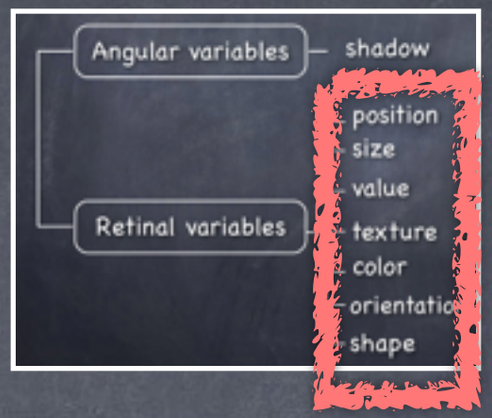
Texture > No texture

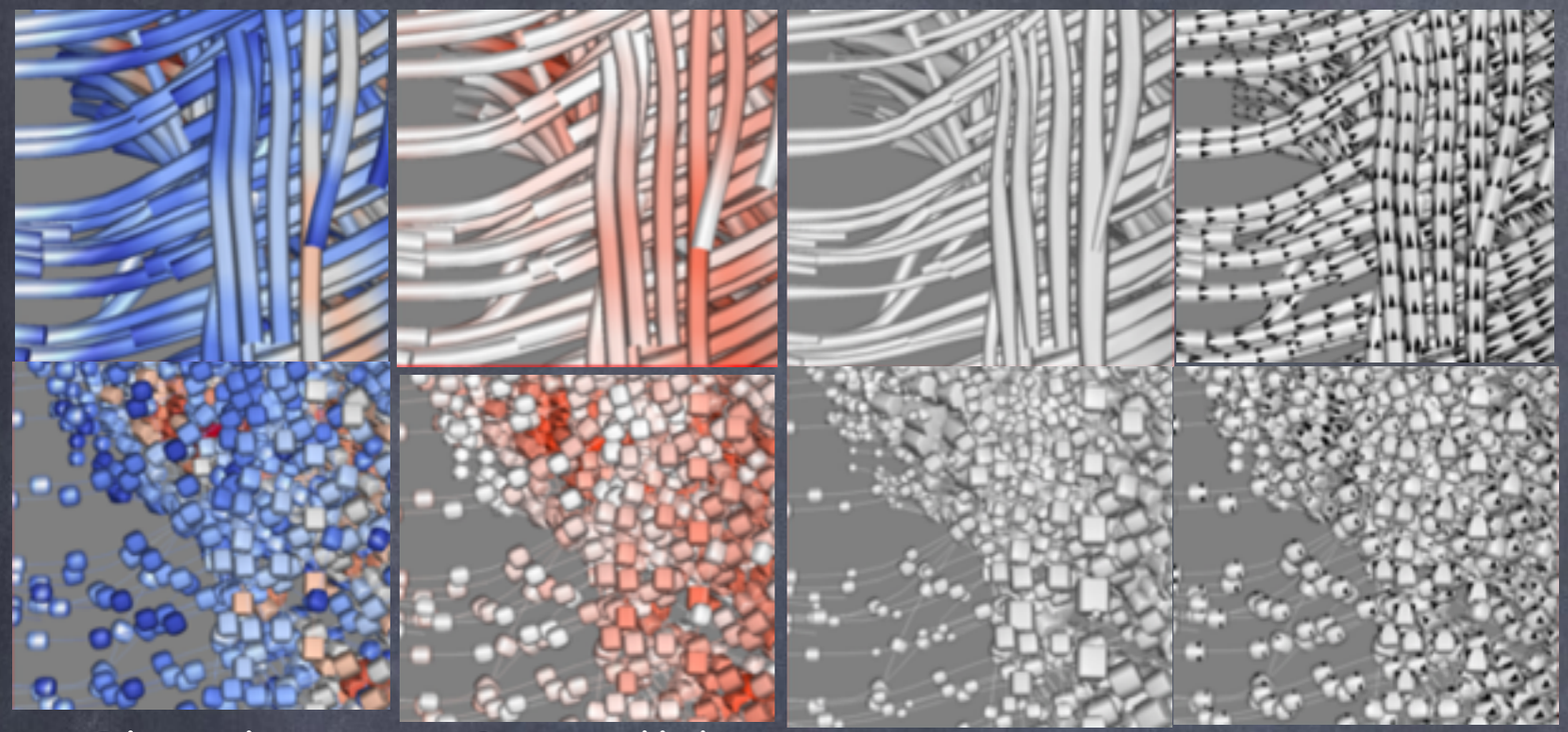


- Major results: (1) More realism -> higher subjective ratings; (2) Realism benefits local tumor contact detection only.

D. Penney, J. Chen, and D.H. Laidlaw, Effects of illumination, texture, and motion on task performance in 3D tensor-field streamtube visualizations, IEEE Pacific visualization 2012.

Ranking Visual Variable Encoding





Streamlines

Superquadrics

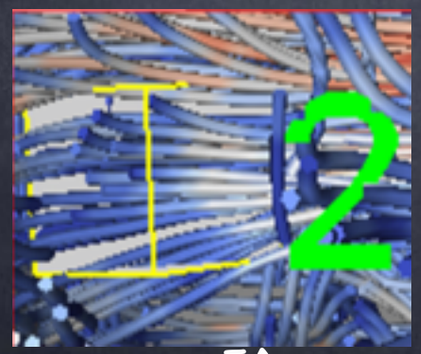
Diverging color

Sequential color

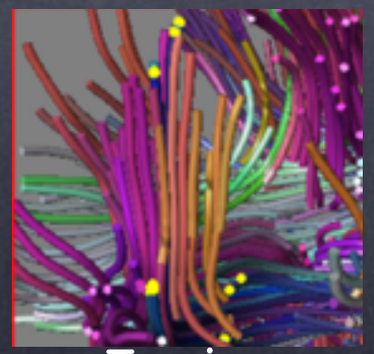
Size

Texture

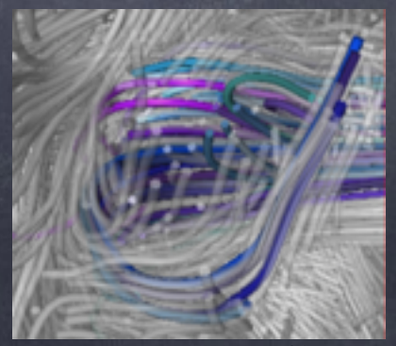
Three tasks



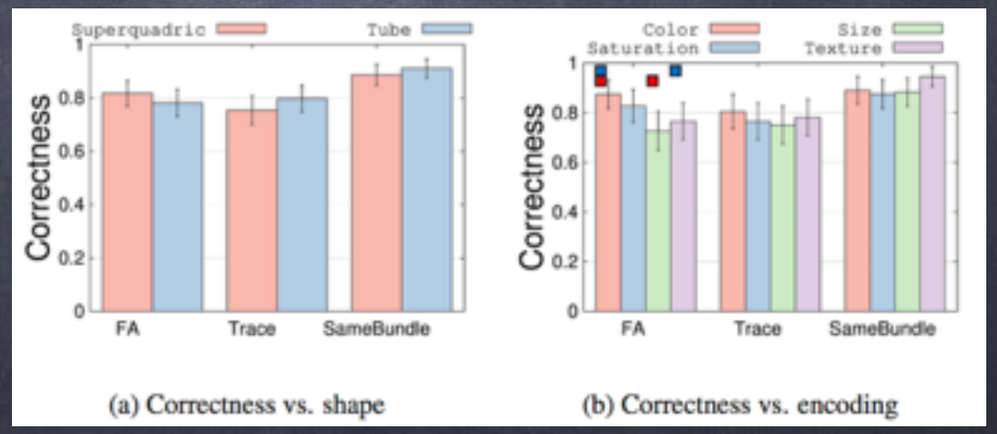
FA



Tracing

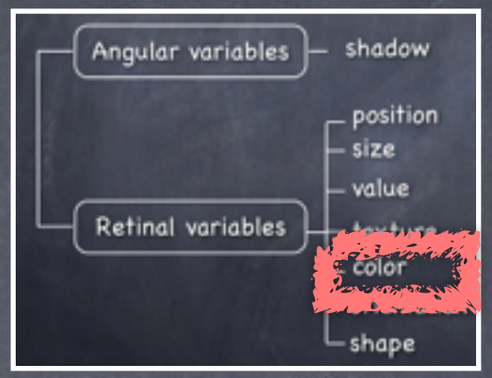


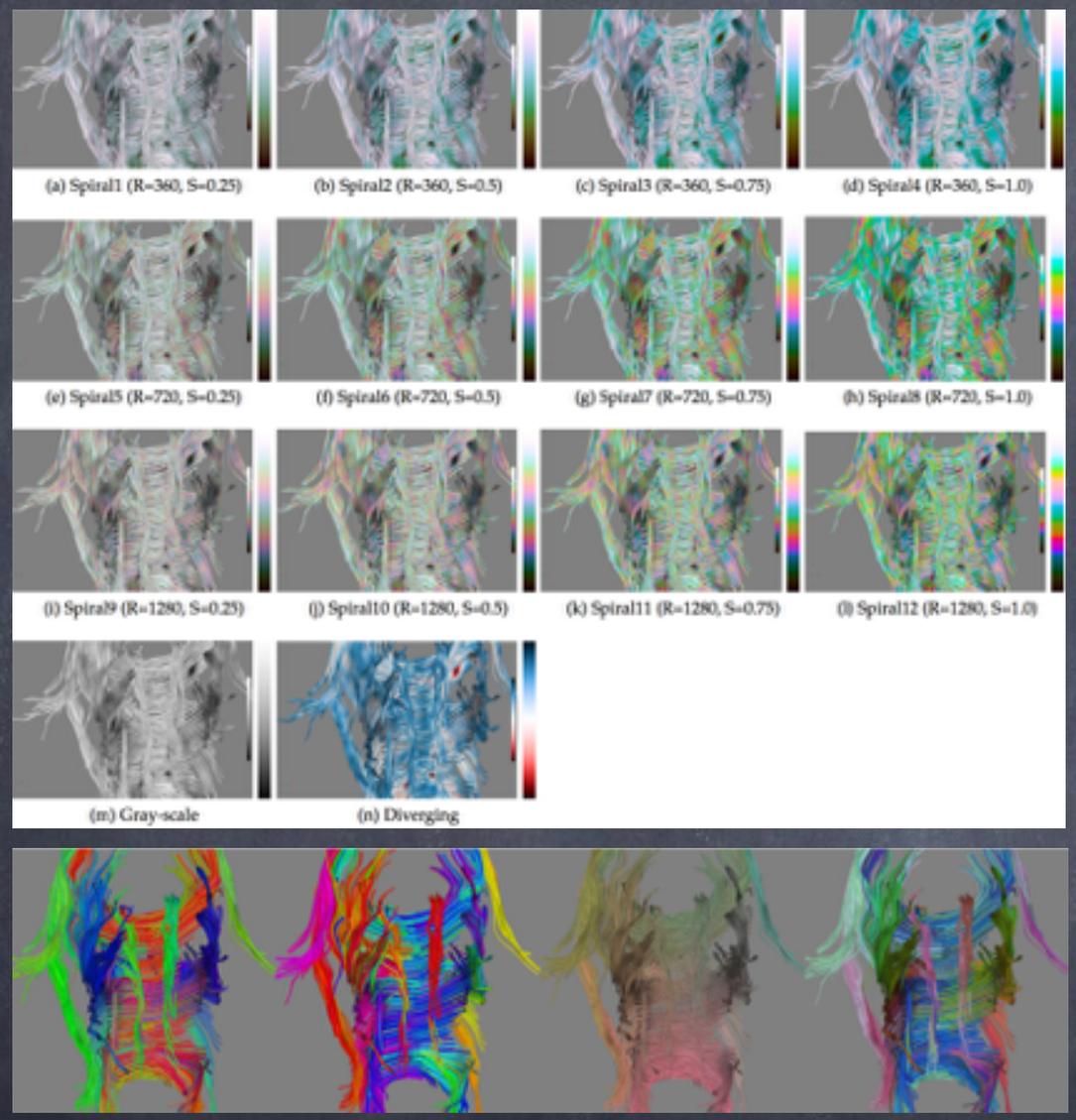
SameBundle

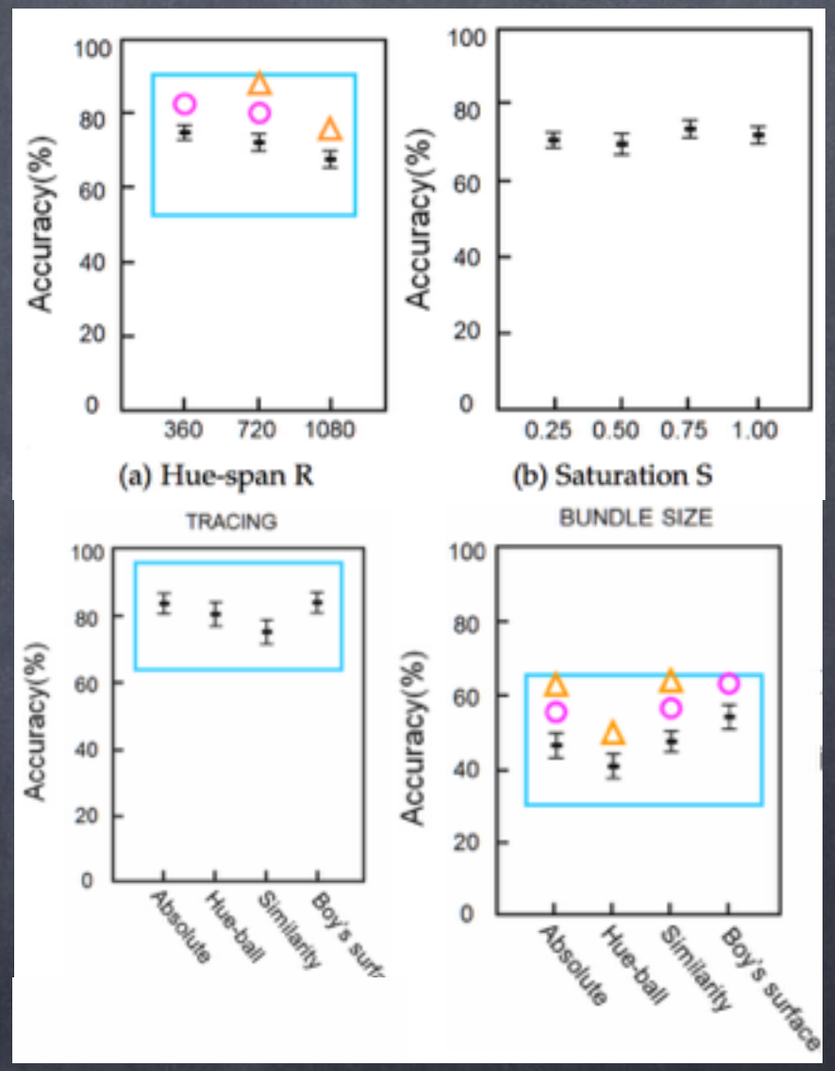


G. Zhang, K. Wu, H. Zhao, A.P. Auchus, and J. Chen, "Rank vis encoding: effectiveness of extended Bertin's retinal variables in diffusion tensor MRI tractography visualizations," IEEE Transactions on Visualization and Computer Graphics, 2017 (under review).

Effects of Coloring Schemes



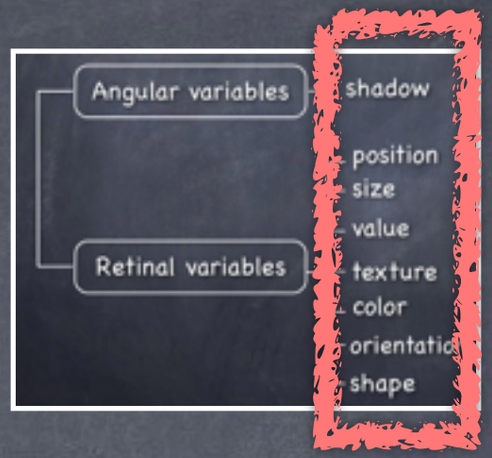


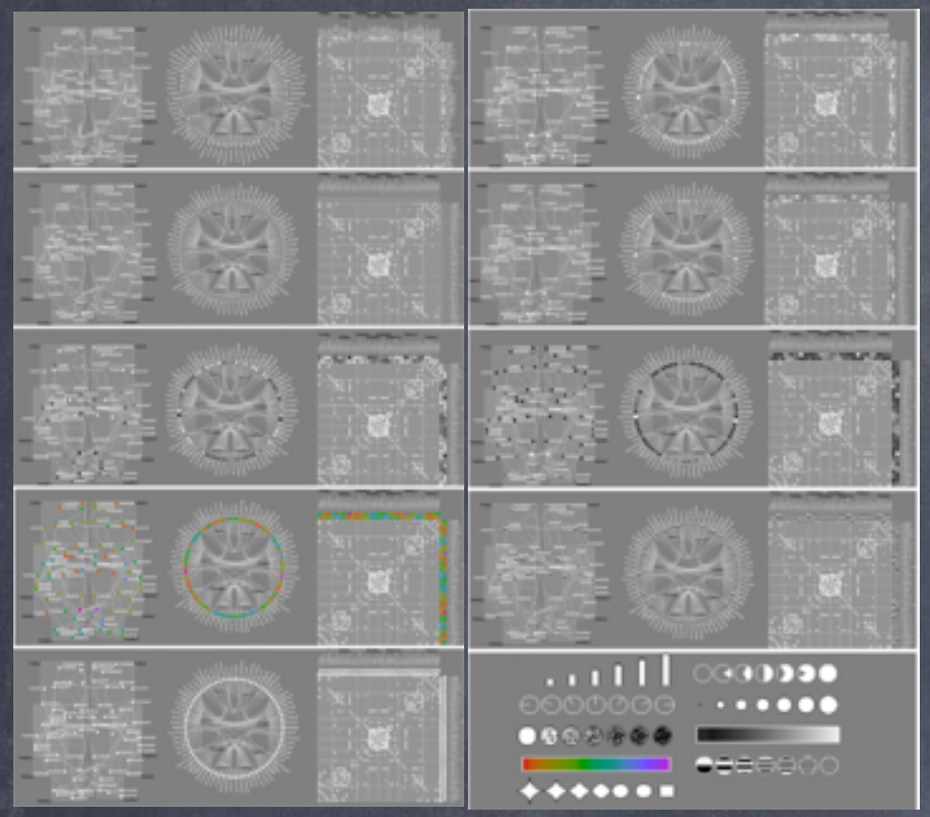


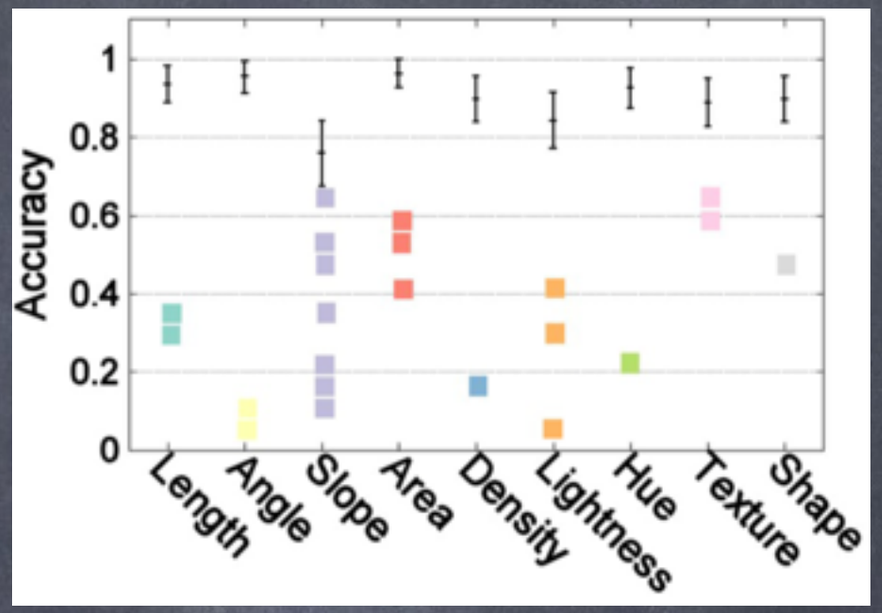
Three tasks FA Tracing BundleSize

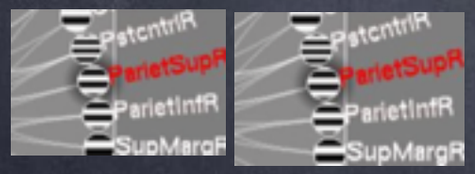
K. Wu, H. Zhao, A.P. Auchus, S. Correia, S.R. Gomez, D. Laidlaw, and J. Chen, "Effects of coloring schemes on 3D streamtube visualization of brain DMRI tractography", IEEE Transactions on Visualization and Computer Graphics, 2017 (under revision).

Effects of Positioning and Encoding

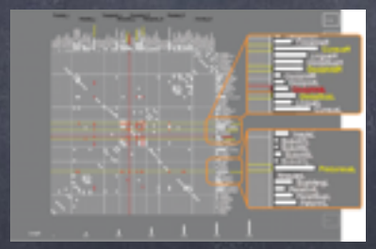




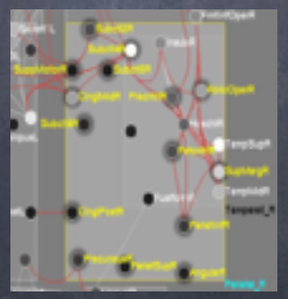




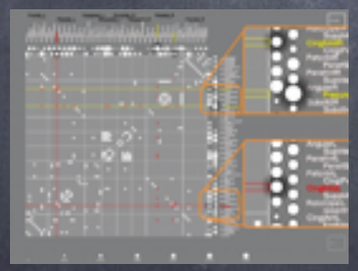
Change detection



Lobe: highest centrality



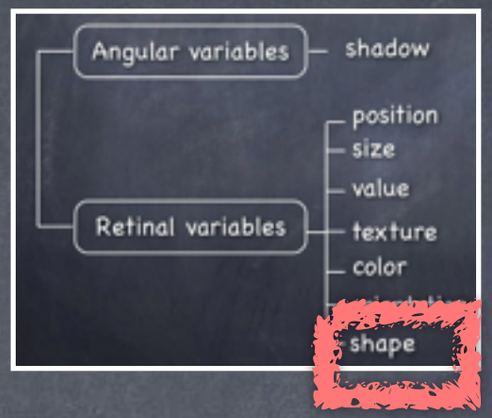
Hemisphere centrality

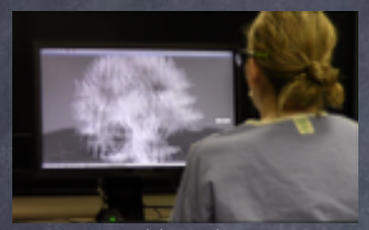


Neighbor: highest centrality

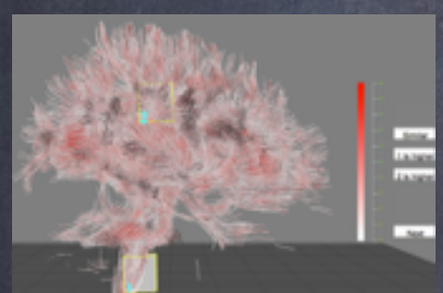
G. Zhang, A.P. Auchus, and J. Chen, "Overlaying quantitative measurement on brain networks: an evaluation of three positioning and nine visual marker techniques", IEEE Transactions on Visualization and Computer Graphics, 2017 (under revision).

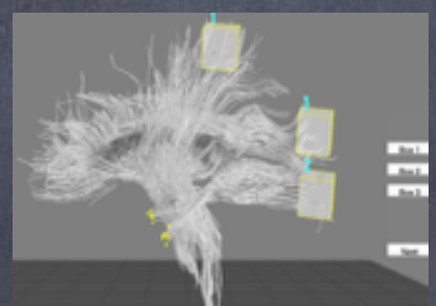
Effects of Stereo and Screen Size

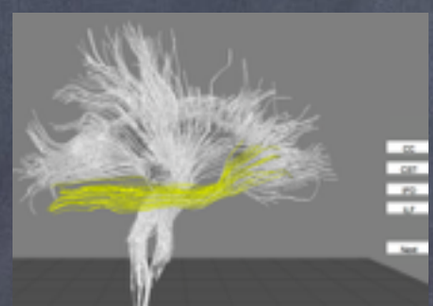


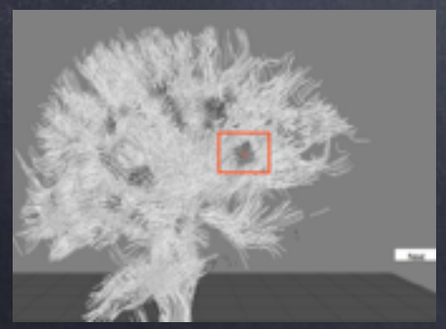

Small stereo

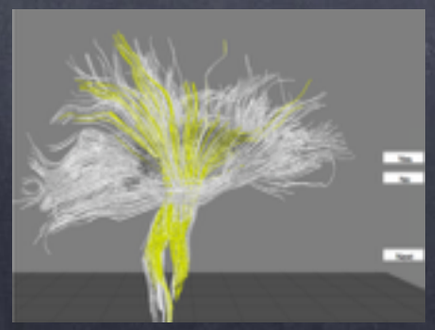

Large stereo

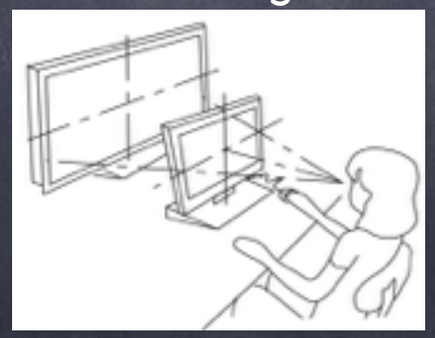

Average FA

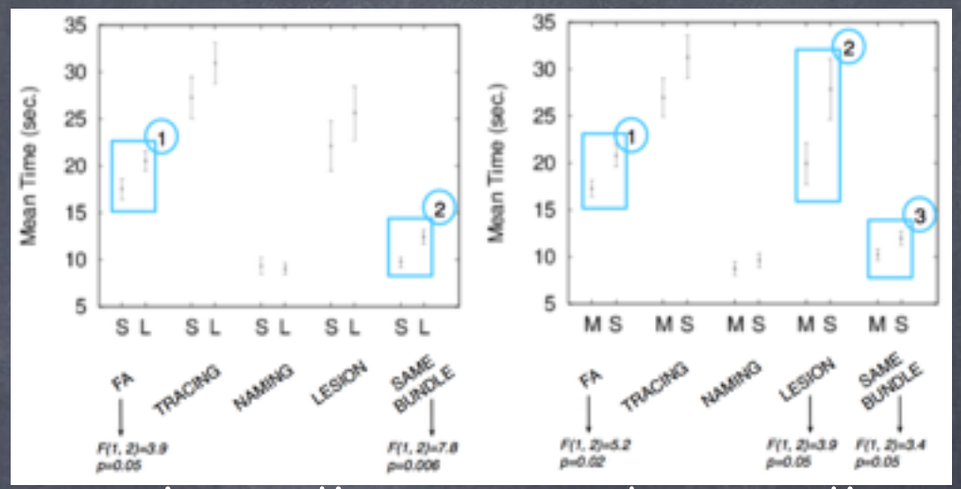

Tracing


Naming


Lesion

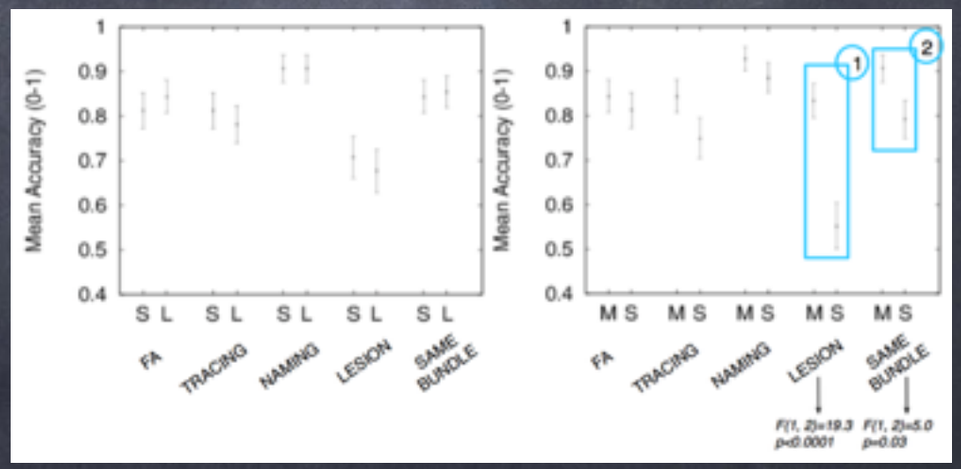

Same bundle


Display setup



size vs. time

stereo vs. time



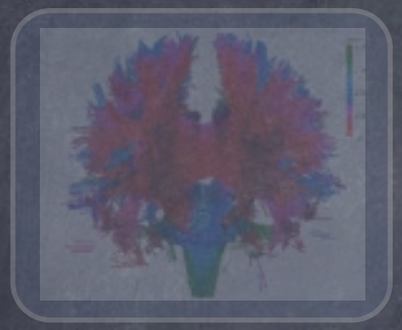
size vs. accuracy

stereo vs. accuracy

J. Chen, H. Cai, AP Auchus, and D. Laidlaw, "Effects of stereo and screen size on the legibility of three-dimensional streamtube visualization", IEEE Transactions on Visualization and Computer Graphics, 18(12): 2130-2139, 2012.

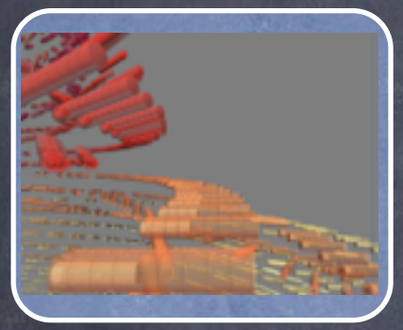
Contributions

- Significant first step
  - understanding effectiveness of data mapping
- Many design guidelines for spatial and non-spatial brain network analysis



1. A scientific visualization **language** for diffusion-tensor MRI visualization

Descriptive framework of seeing



2. **Experiment**: design and evaluation of perceptually accurate visualization

Experiments

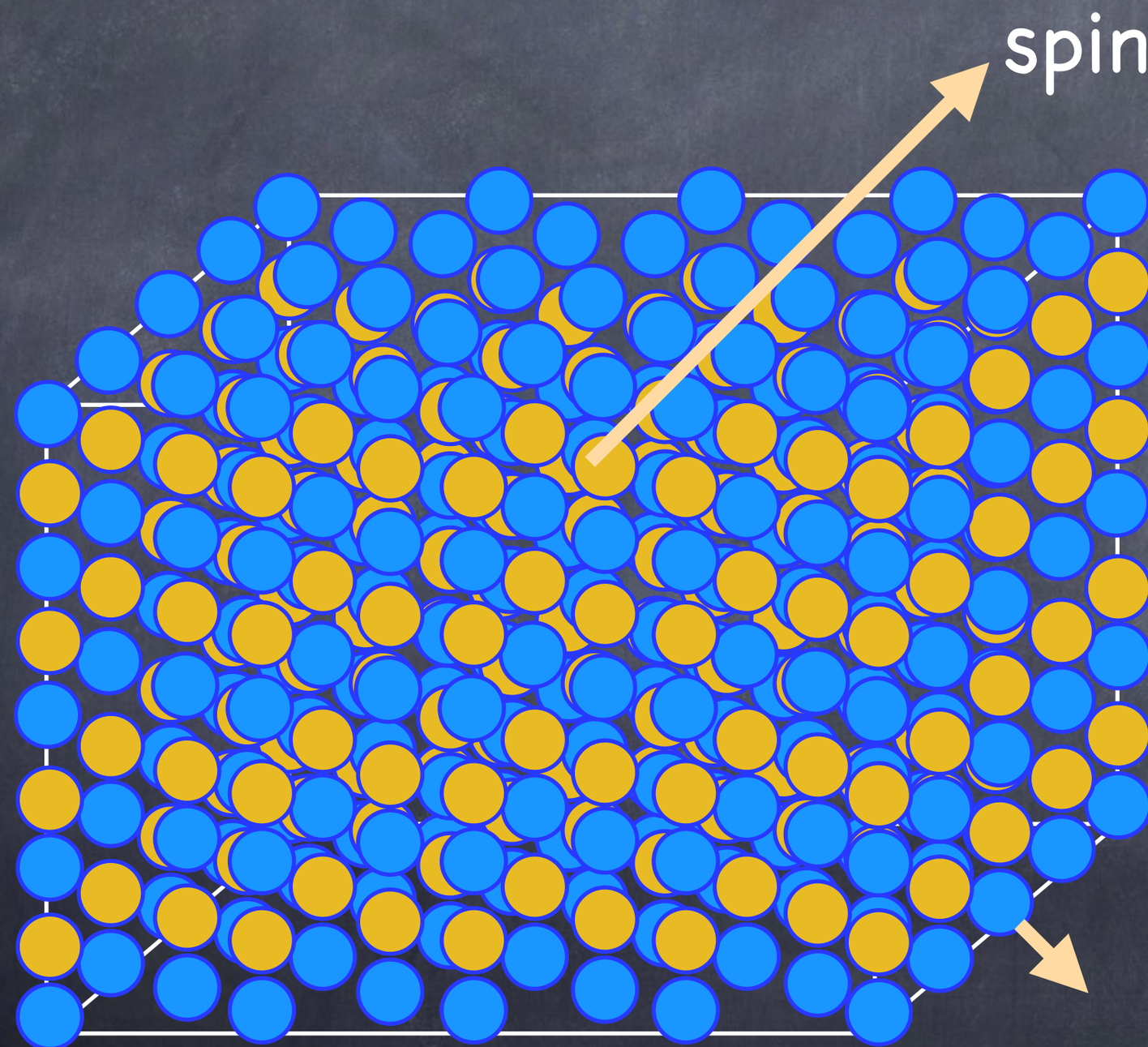
Collaborators:

Computer science: Judith Terrill (NIST)

Quantum physicist: Bryant Garnett (NIST)

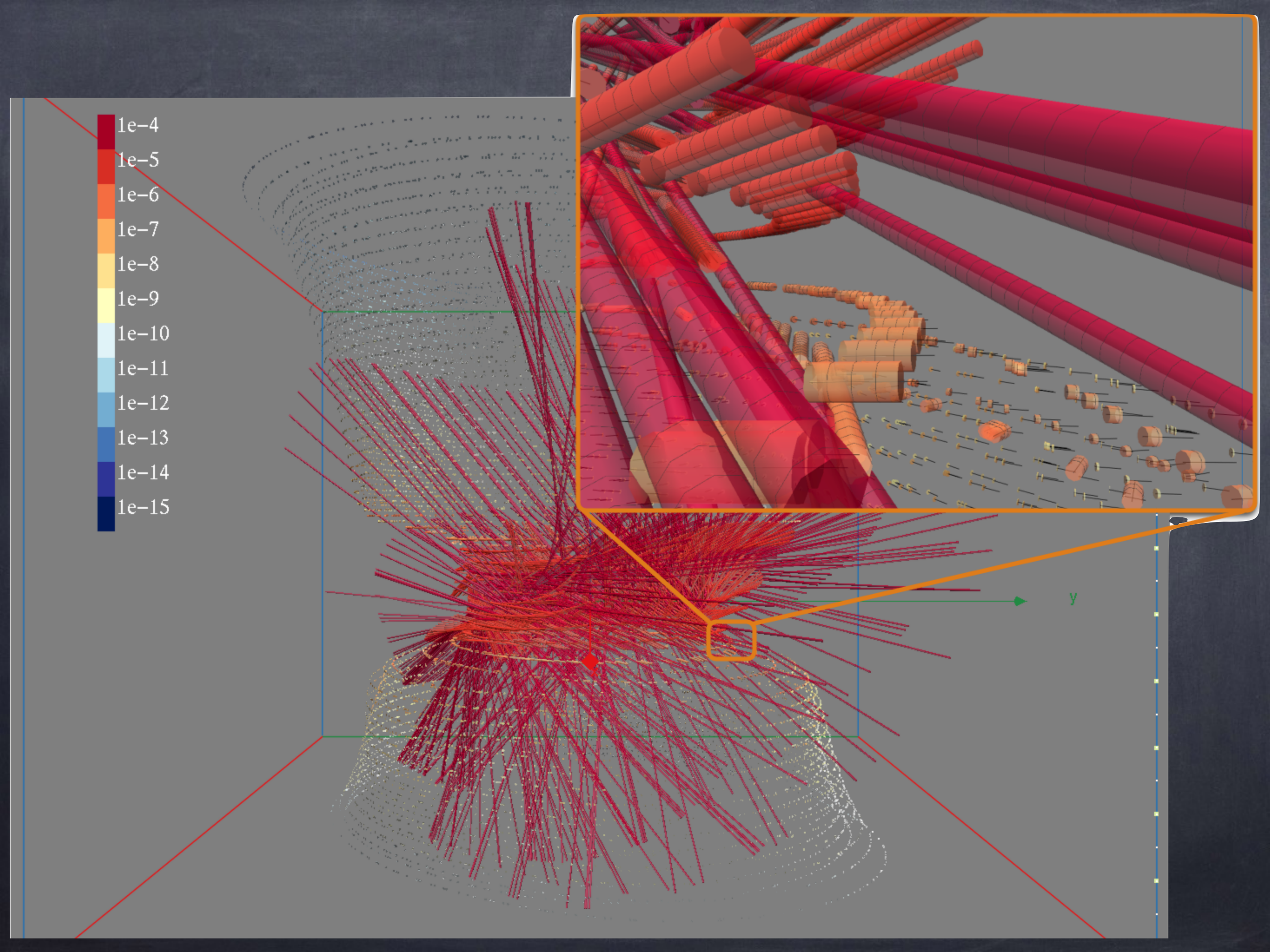
- Research questions: how does visualization support **seeing structures and large-magnitude range data** from highly dense spin vectors?

Quantum Simulators

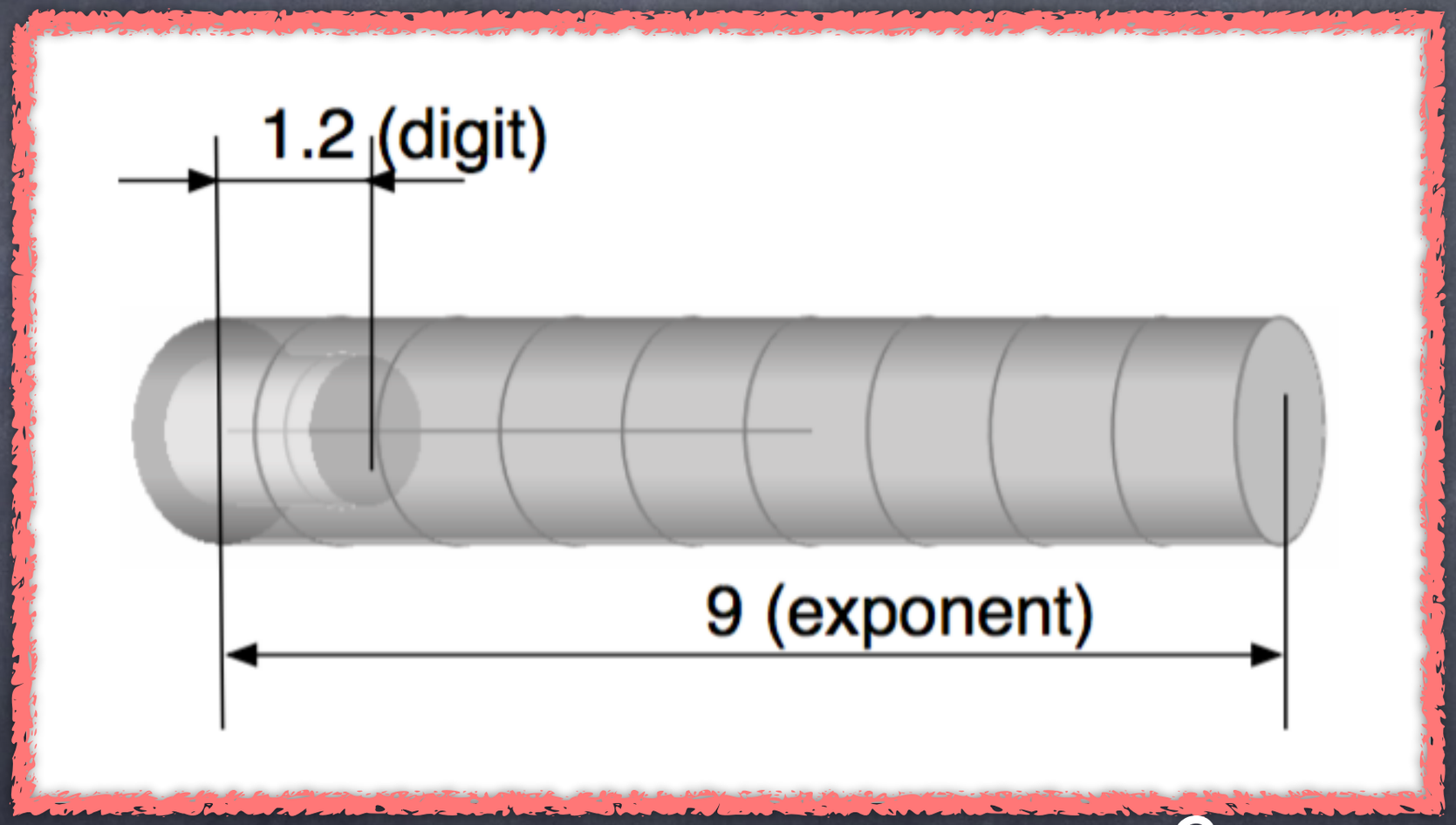


Quantify nanoparticle behaviors

$[10^{-15}, 10^{-4}]$


1e−4
1e−5
1e−6
1e−7
1e−8
1e−9
1e−10
1e−11
1e−12
1e−13
1e−14
1e−15
y

SplitVectors



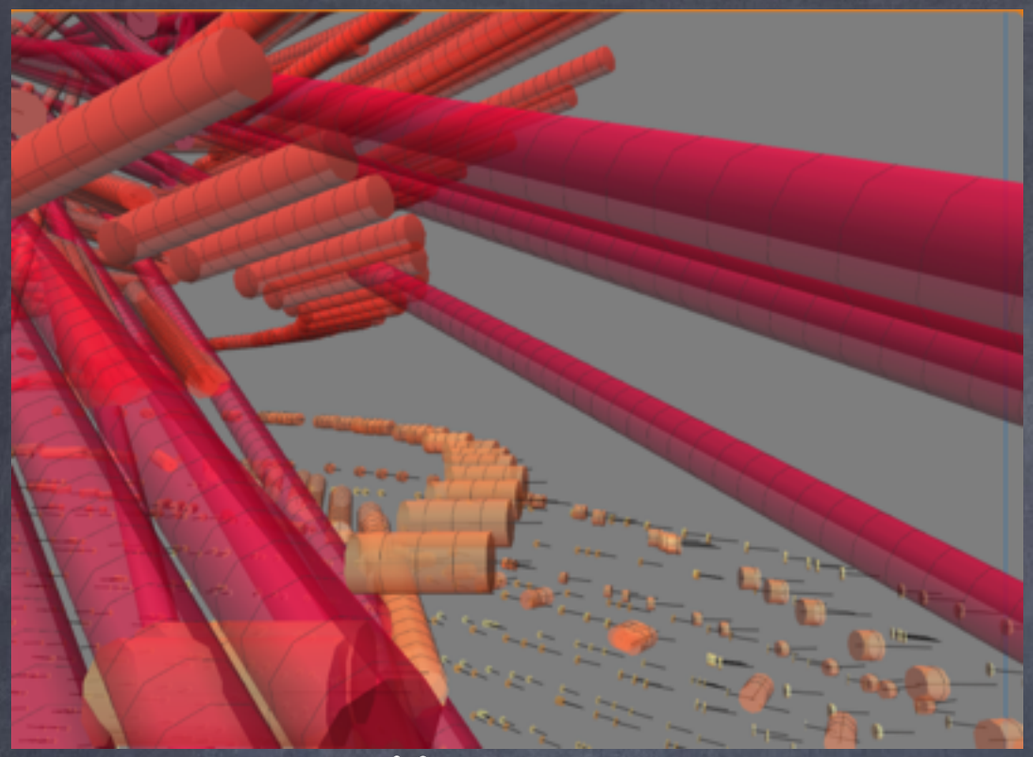
$$1200{,}000{,}000 = 1.2 \times 10^{9}$$

Experimental Design

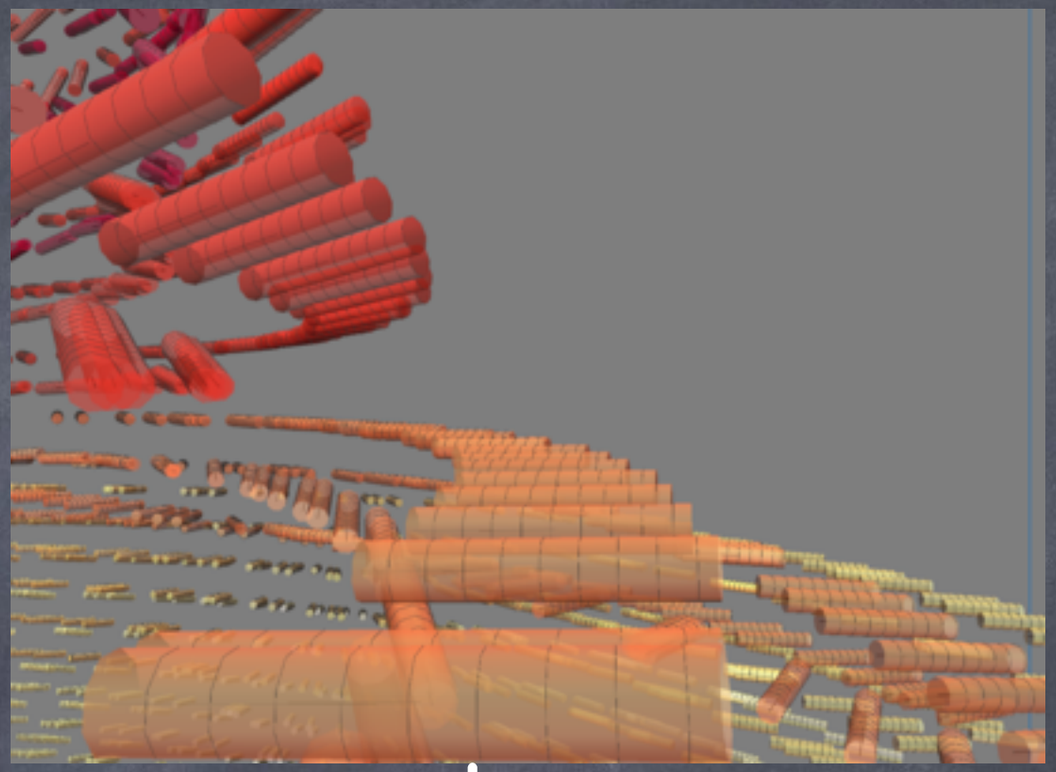
Selected hypothesis: SplitVectors leads to most perceptually accurate representations.

- Independent variables:
  - Encoding method and stereo
- Depend variables
  - Time, error, sensitivity analysis
- Participants: 20

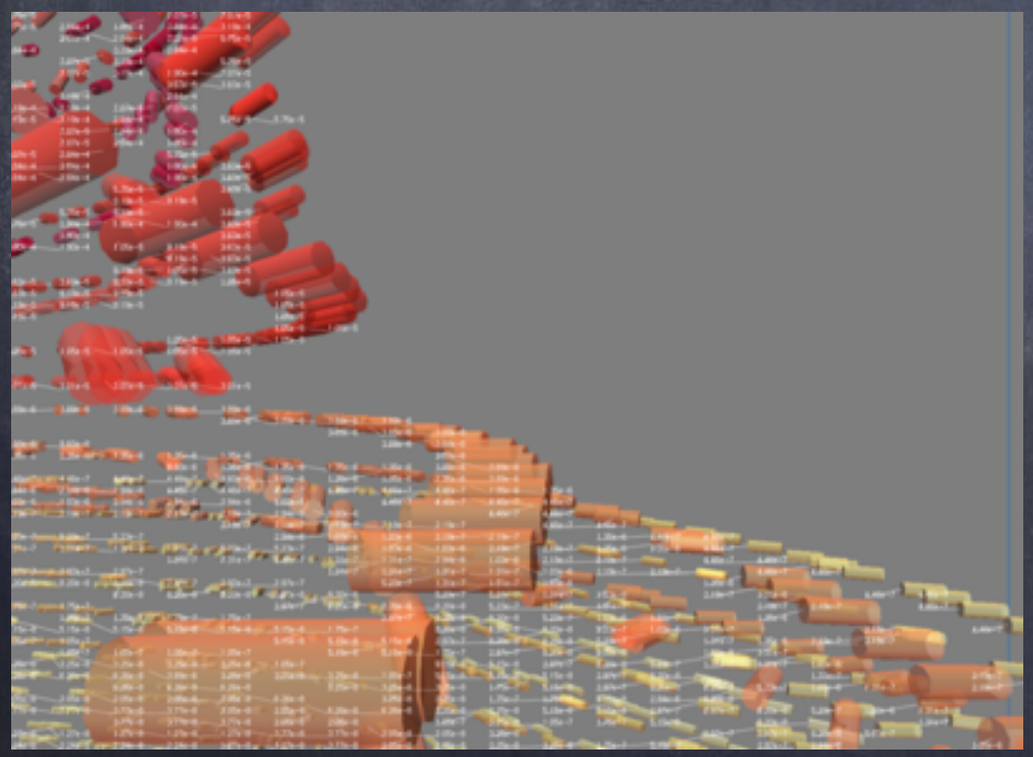
H. Zhao, G. Bryant, W. Griffin, J. Terrill, and J. Chen, "Validation of splitVectors encoding for quantitative visualization of large-magnitude-range vector fields," IEEE Transactions on Visualization and Computer Graphics, 2016.



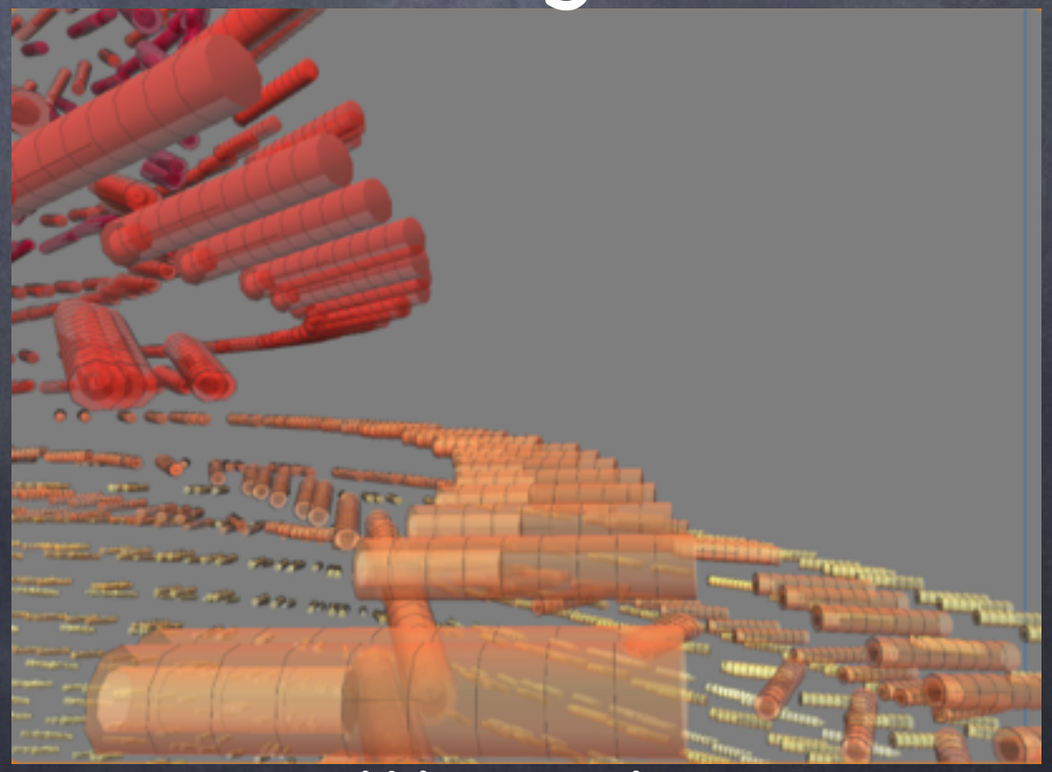
linear



log

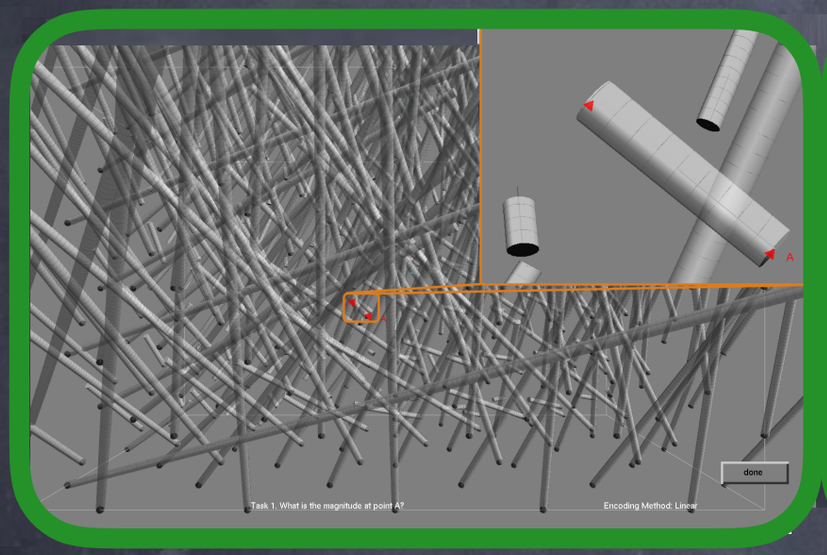


text

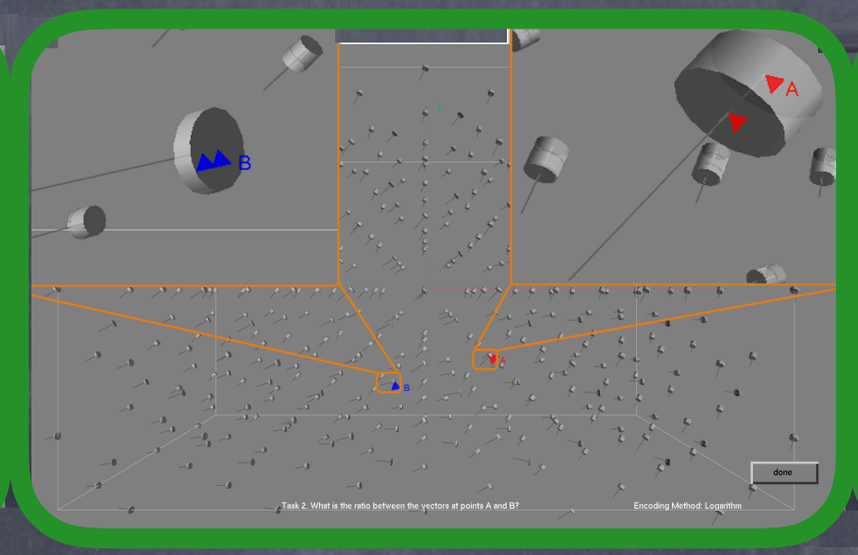


splitVectors

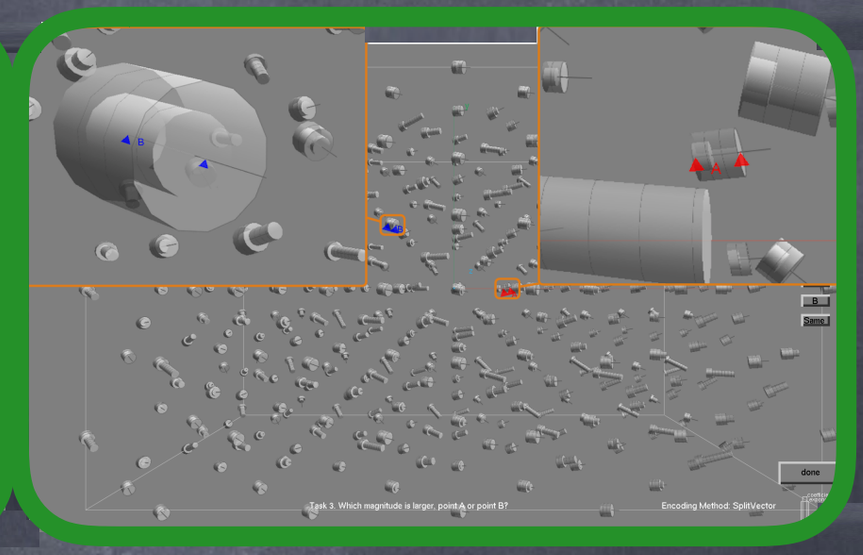
Three Tasks



Magnitude reading

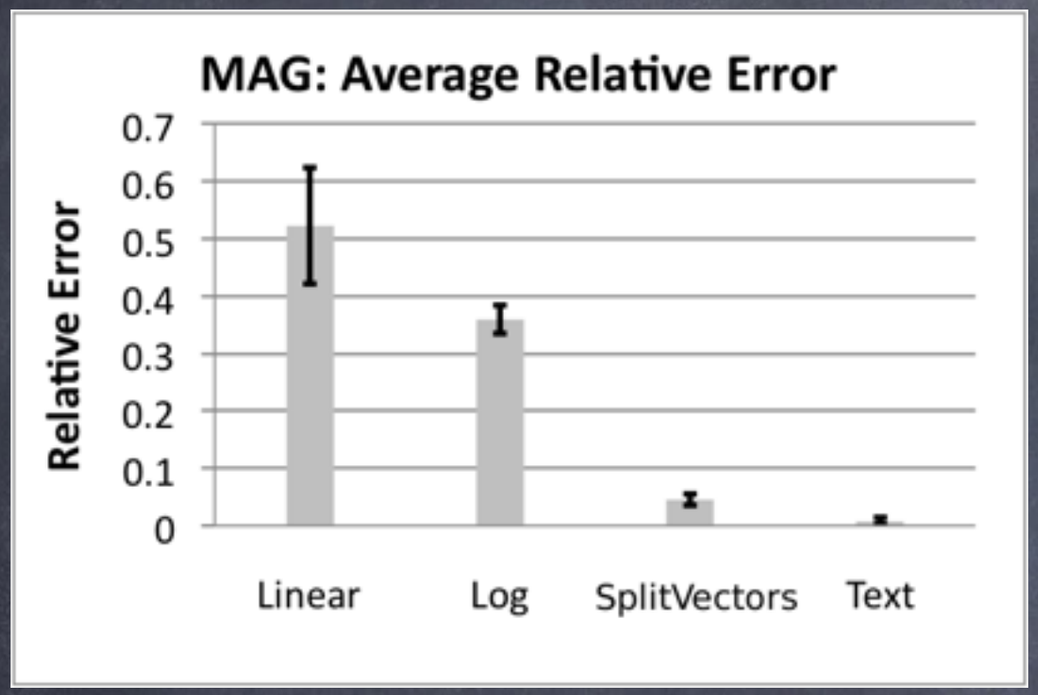


Ratio

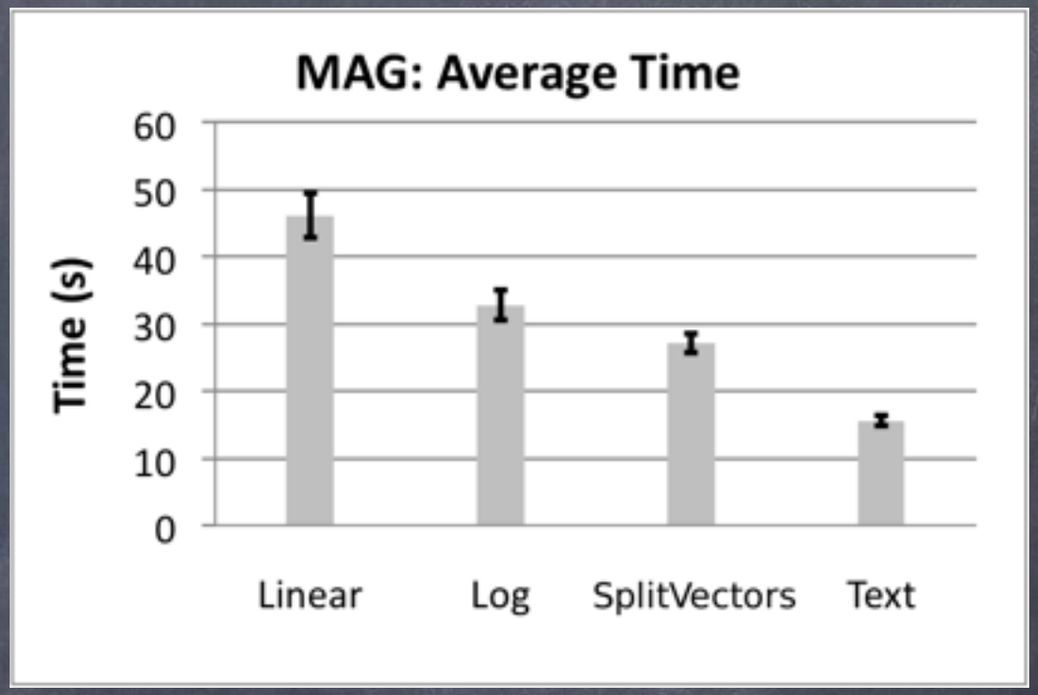


Comparison (which one is larger?)

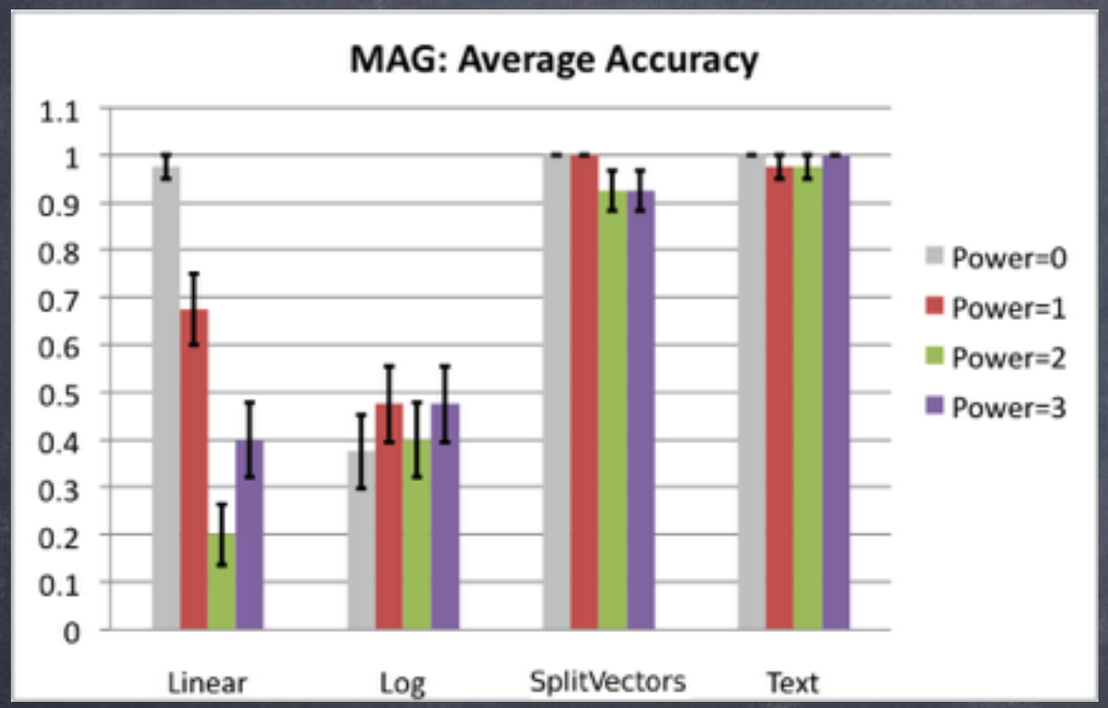
Major results

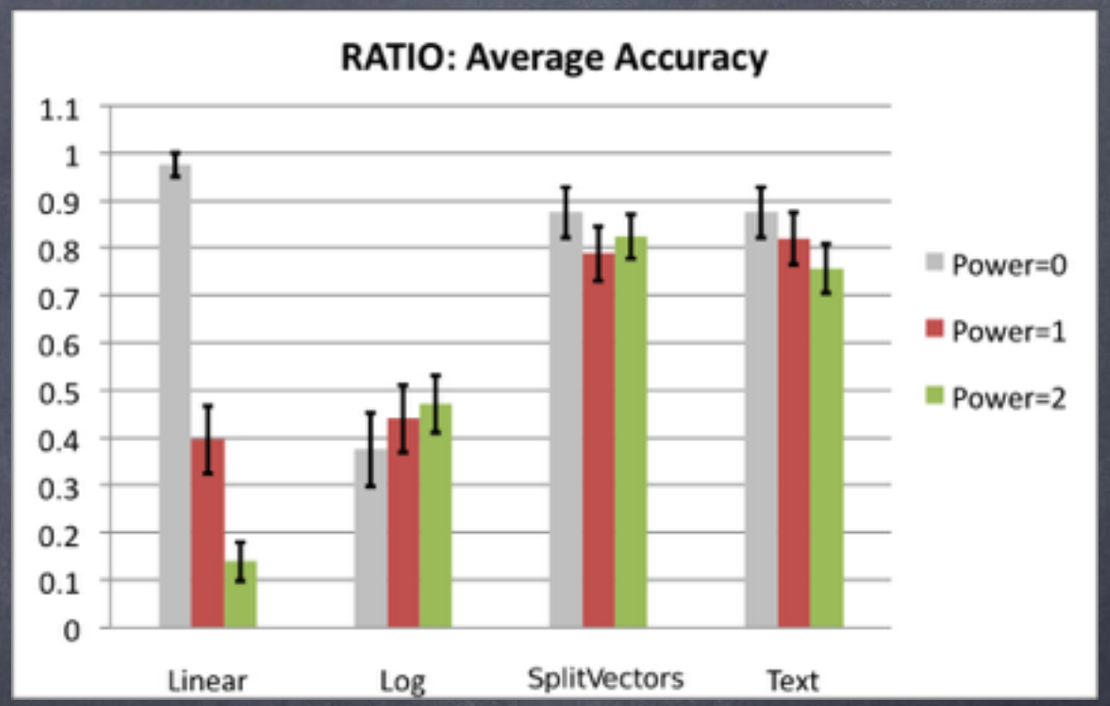


$F(3, 639)=22.5, p<0.0001$



$F(3, 639)=35, p<0.0001$



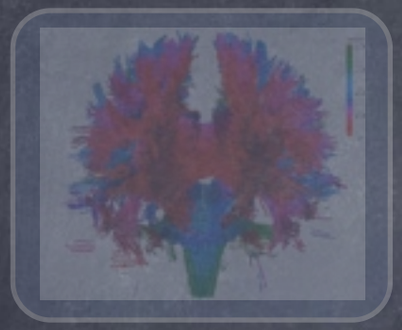


E.g., $92 = 9.2 \times 10^1$

Contributions

- A new perceptually accurate visualization
- New quantitative tasks to discriminate numeric values
- Results could have impact on other types of 3D large-range field visualizations

H. Zhao, G. Bryant, W. Griffin, J. Terrill, and J. Chen, "Validation of splitVectors encoding for quantitative visualization of large-magnitude-range vector fields," IEEE Transactions on Visualization and Computer Graphics, 2016.



1. A scientific visualization **language** for diffusion-tensor MRI visualization

2. **Experiment**: design and evaluation of perceptually accurate visualization

3. Workflow-driven design for time-varying bat flight **analysis**

Knowledge discovery

Goal: invisible visual interfaces for knowledge discovery

Collaborators:

Computer science: Andrew Bragdon (Microsoft Research), Andy van Dam, David H. Laidlaw

Biology: Sharon M. Swartz, Rhea von Busse

Problem Domain

- Kinematics
  - Complex wing bone interaction
  - Time-varying wing deformation
- Kinetics
  - Unmanned vehicle design

Recording @ 1000 fps
Playback @ 30 fps
~ 33x slow down

Video courtesy of Brown University

Conventional problem solving approach

Downstroke Upstroke

- Observations (bio)
- Matlab feature extraction (bio, cs, math)
- Visualization (cs)
- Hypothesis formation (bio, eng)
- Comparison (cs, bio)

Work in multiple environments

Extremely complex and dynamic process

Barriers to knowledge discovery

- Observations (bio)
- Matlab feature extraction (bio, cs, math)
- Visualization (cs)
- Hypothesis formation (bio, eng)
- Comparison (cs, bio)

Error-prone computing

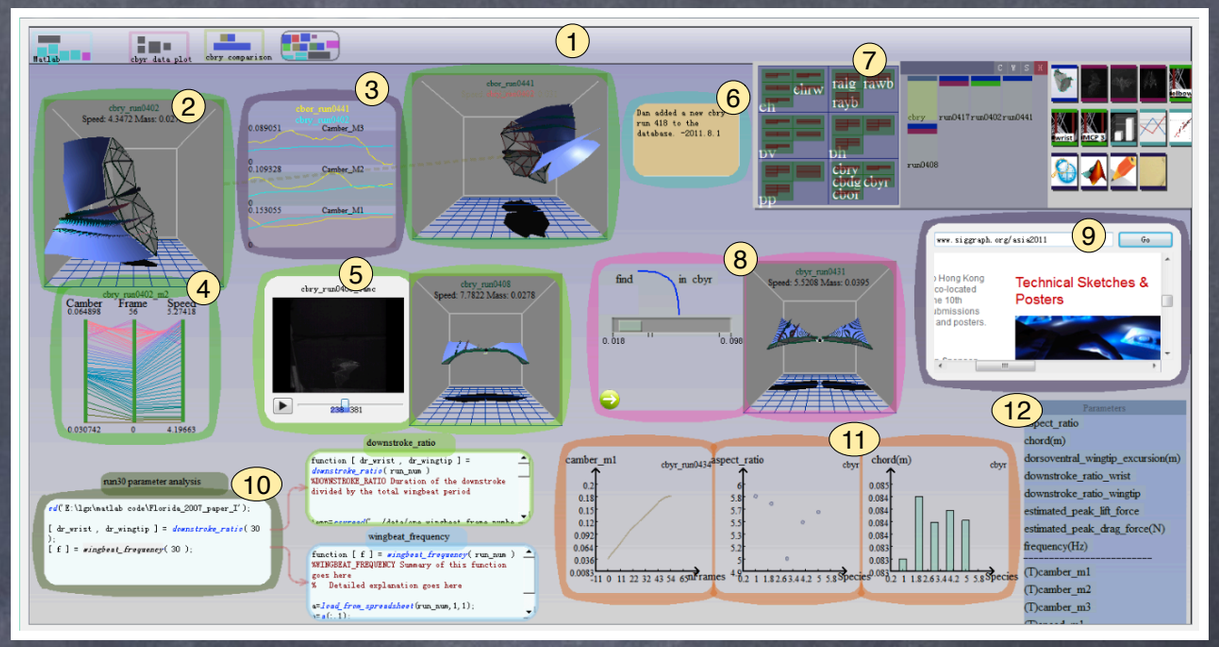
Inefficient collaborative social dynamics

Education

Difficulties in visualization

Our Solution: VisBubbles

- In the nutshell, it is a metaphorical interface
  - **Organizing theme**: units of visualization; interface theme matches task theme
  - **Immediacy**: A programming environment for data handling cross-linked to visualization
  - **Proximity**: interactive regrouping and rearrangement
  - An asynchronous collaborative environment



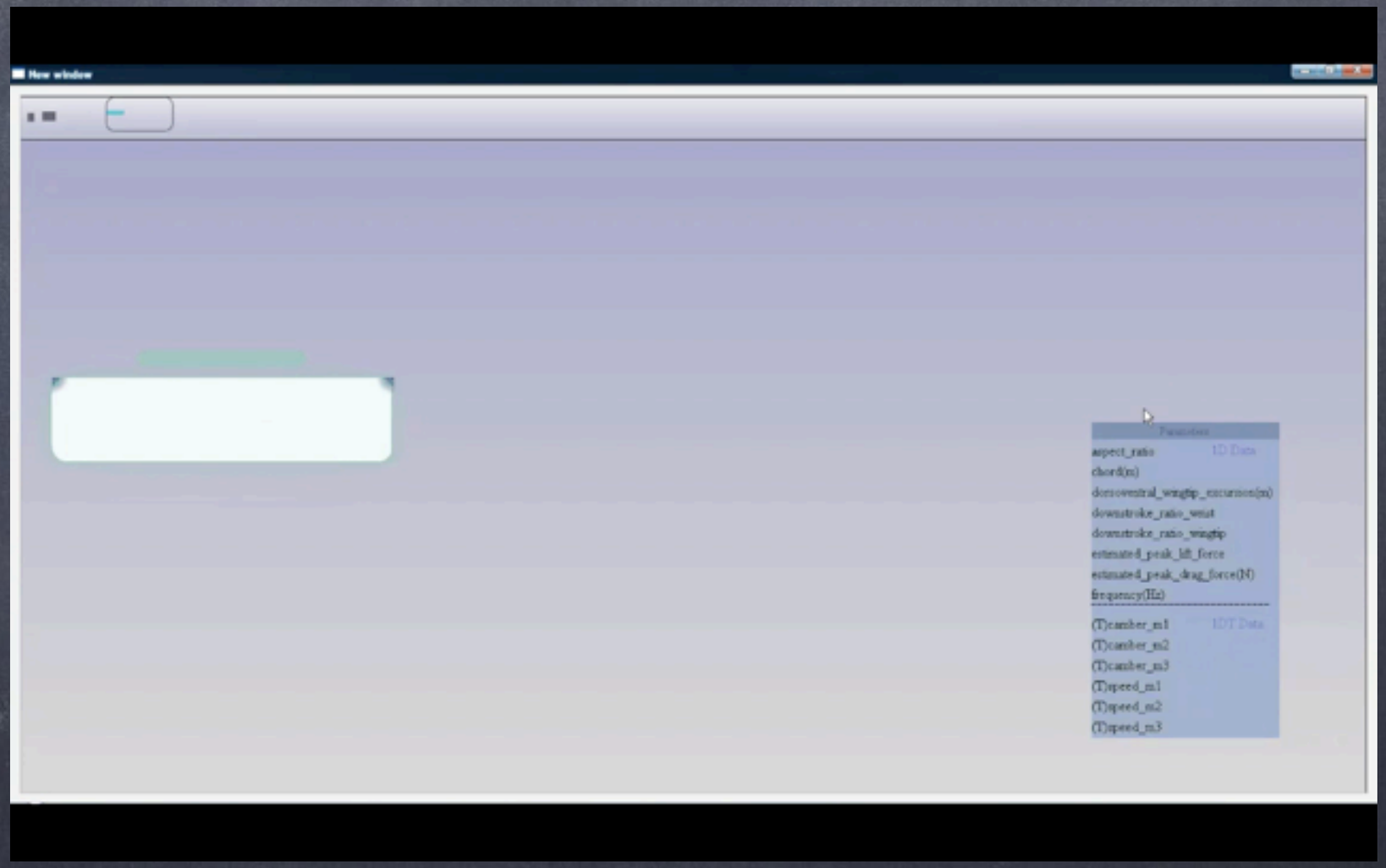
Error-prone computing

Difficulties in visualization

Inefficient collaborative social dynamics / education

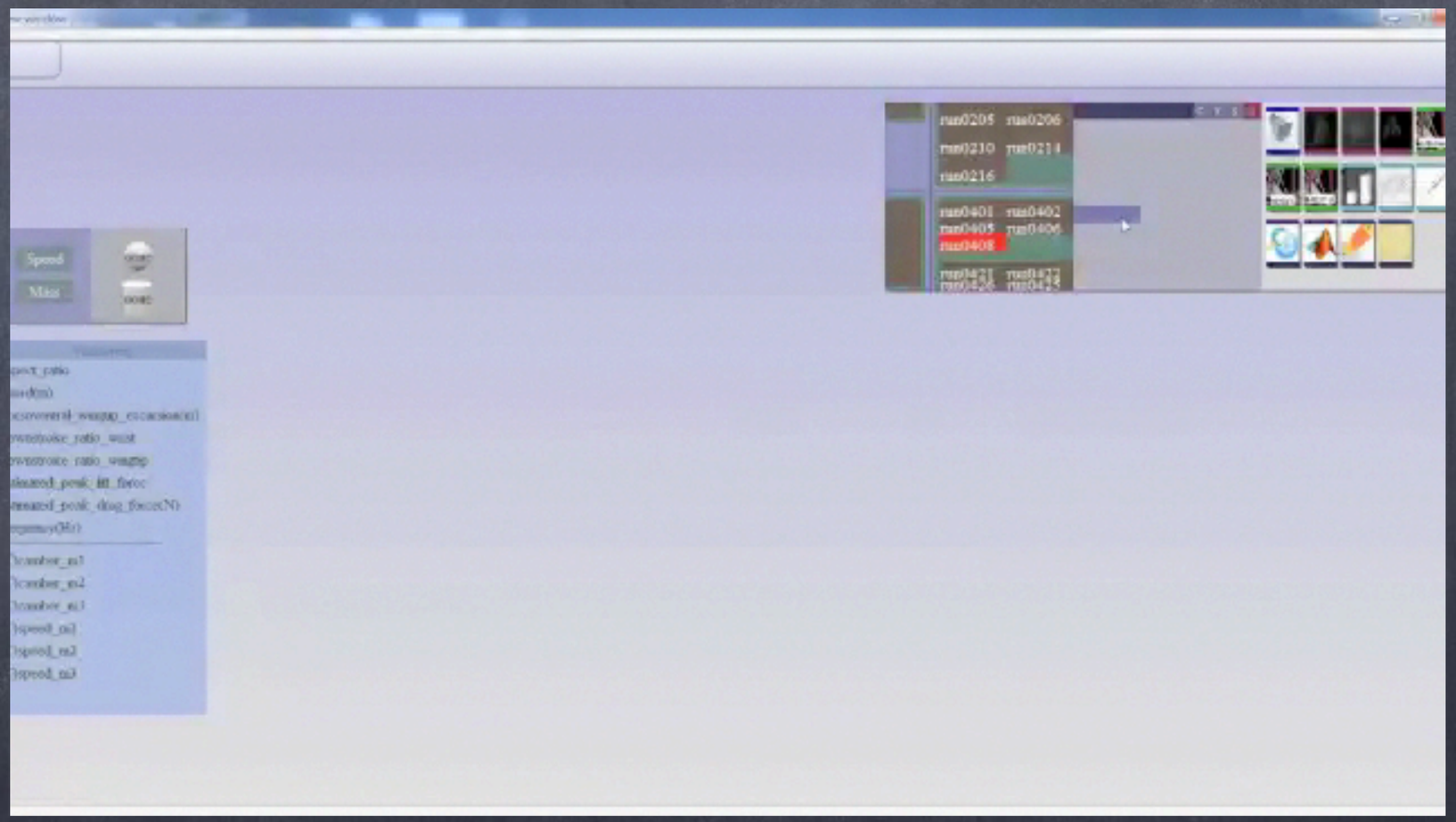
Code Navigation Efficiency

- **Organizing theme:** Concurrent display with automatic layout.
- **Task proximity:** Editing multiple fragment of functions.
- **Immediacy:** Inspecting execution results visually.



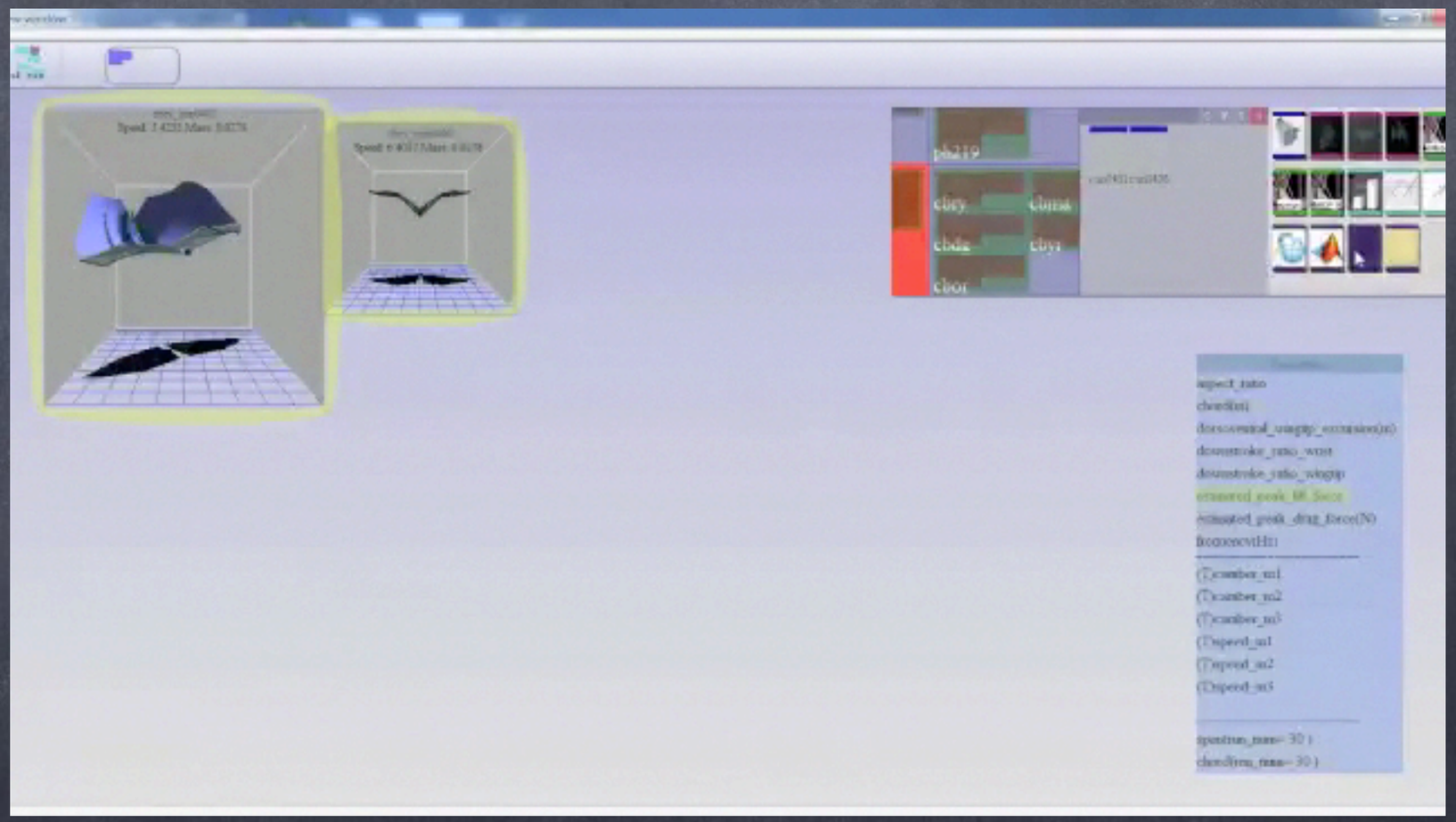
Visual Comparison Efficiency

- **Organizing theme**: units of visualization; interface theme matches task theme
- **Immediacy**: A programming environment for data handling cross-linked to visualization
- **Proximity**: interactive regrouping and rearrangement
- An asynchronous collaborative environment



Visual Comparison Efficiency

- **Organizing theme**: units of visualization; interface theme matches task theme
- **Immediacy**: A programming environment for data handling cross-linked to visualization
- **Proximity**: interactive regrouping and rearrangement
- An asynchronous collaborative environment



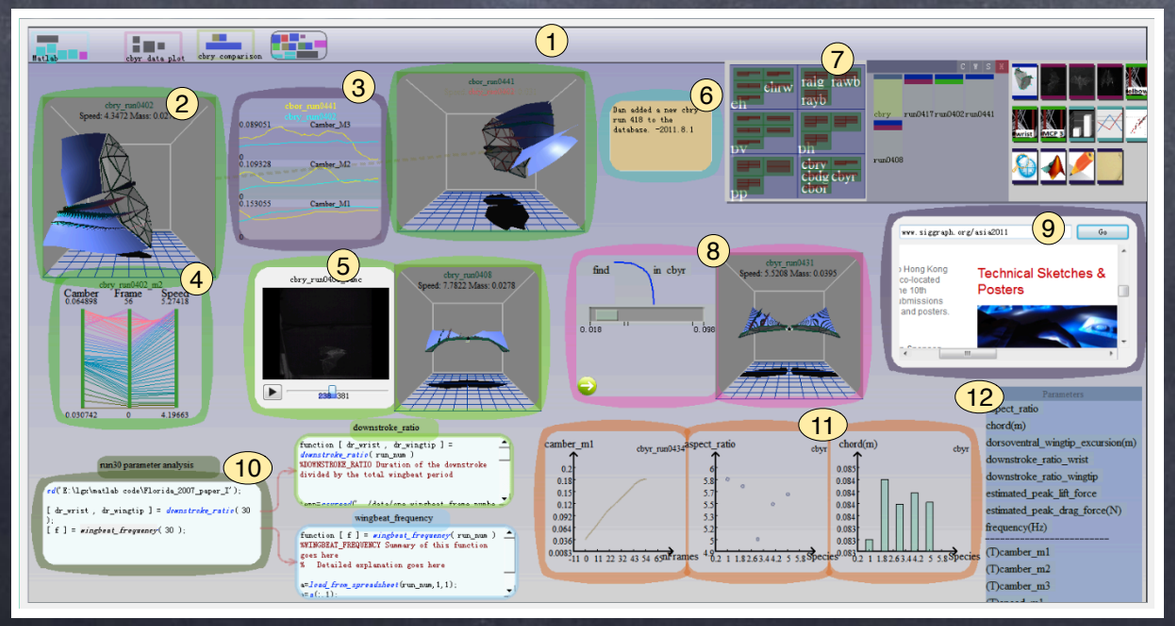
How to Support Creativity?

- memory sequencing (spatial locations, predicting next step): overview, small functions and composition

Reduce interruption

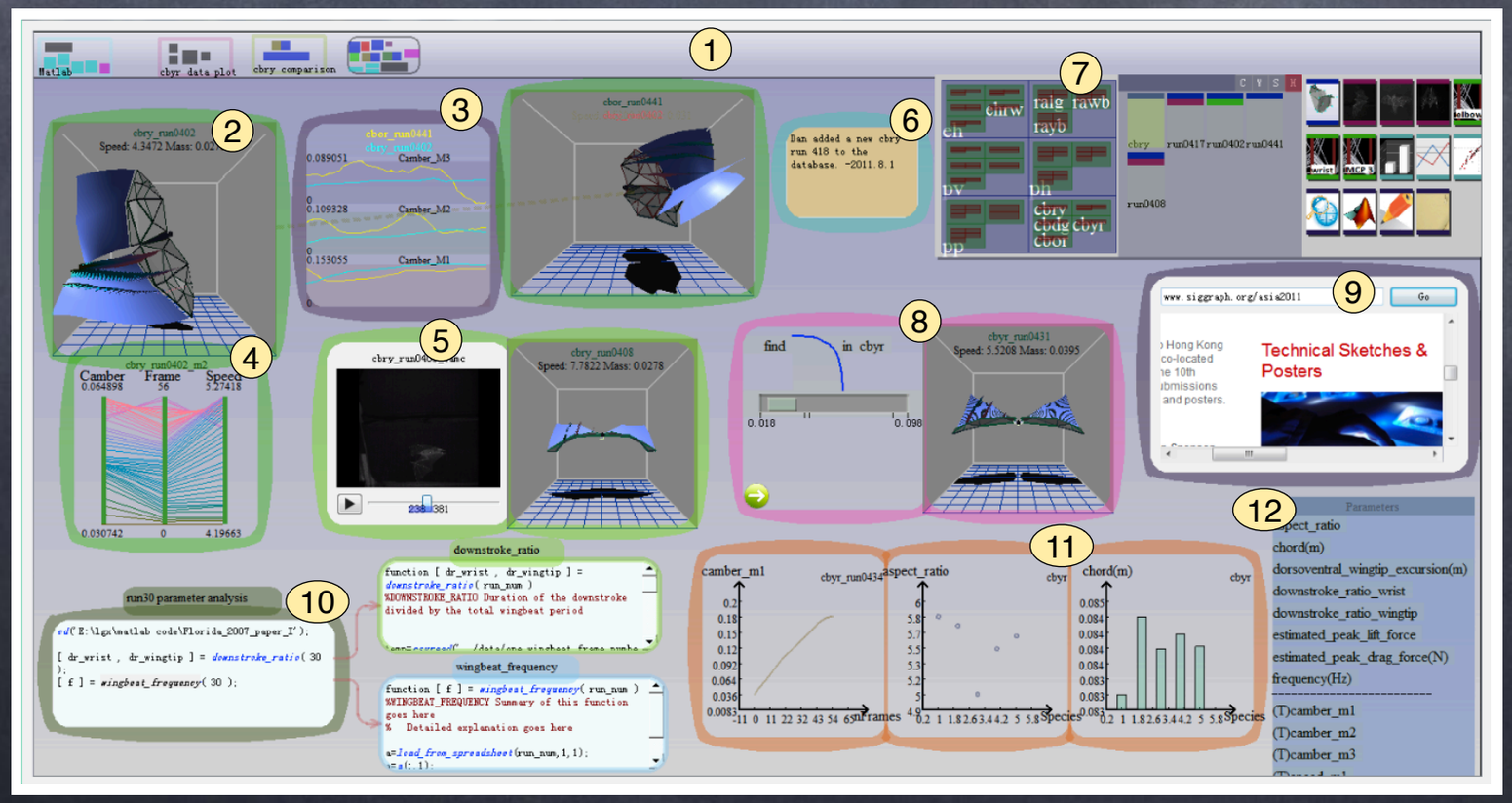
- forming schema (Barlett 32, mental structure representing knowledge): proximity->grouping->linking

Consistency



Forming Schema

- Bubbles metaphor (Bragdon 2010)
- User behavior -> interface action
  - grouping -> linking



Jian: Andy, look at the wonderful thing I did :-).
Andy: Jian, …… focus on science and tell me why it works…. and focus on interaction…

@Image courtesy of van Dam

Pak C. Wong, Hanwei Shen, and Chaomei Chen, Top ten interaction challenges in extreme-scale visual analytics, IEEE CG&A, 2012.

- The Renaissance of conventional wisdom: overview->detail.
- The human bottleneck
- Design web-based tools

Biological Pathway Exploration

- Proximity-driven comparison: a mirror metaphor

Contributions

- Novel metaphors (bubbles, mirrors) and query methods
- Progressive refinement environments
- Memory-proximity driven design
  - visual proximity; task proximity; user proximity

Our collaborators are using these features for hypothesis confirmation and generation.

Current work: Nuances in Network Visualizations

- Challenges: encoding quantitative values in networks
- Method:
  - Network entity grouping utilizing Gestalt principles (coloring)
  - Calibrated-columns for patter recognition and extreme value detection

Current work: Interactive Discourse Analysis

Deeper analysis

- Is there an accessible structure in space usage pattern within the knowledge discovery discourse?
- How might one exploit this?
- Answering these questions?
  - Is inherently multidisciplinary
  - Requires expansive effort and vision
  - Promising great rewards
- A key component is mental imagery in discourse.

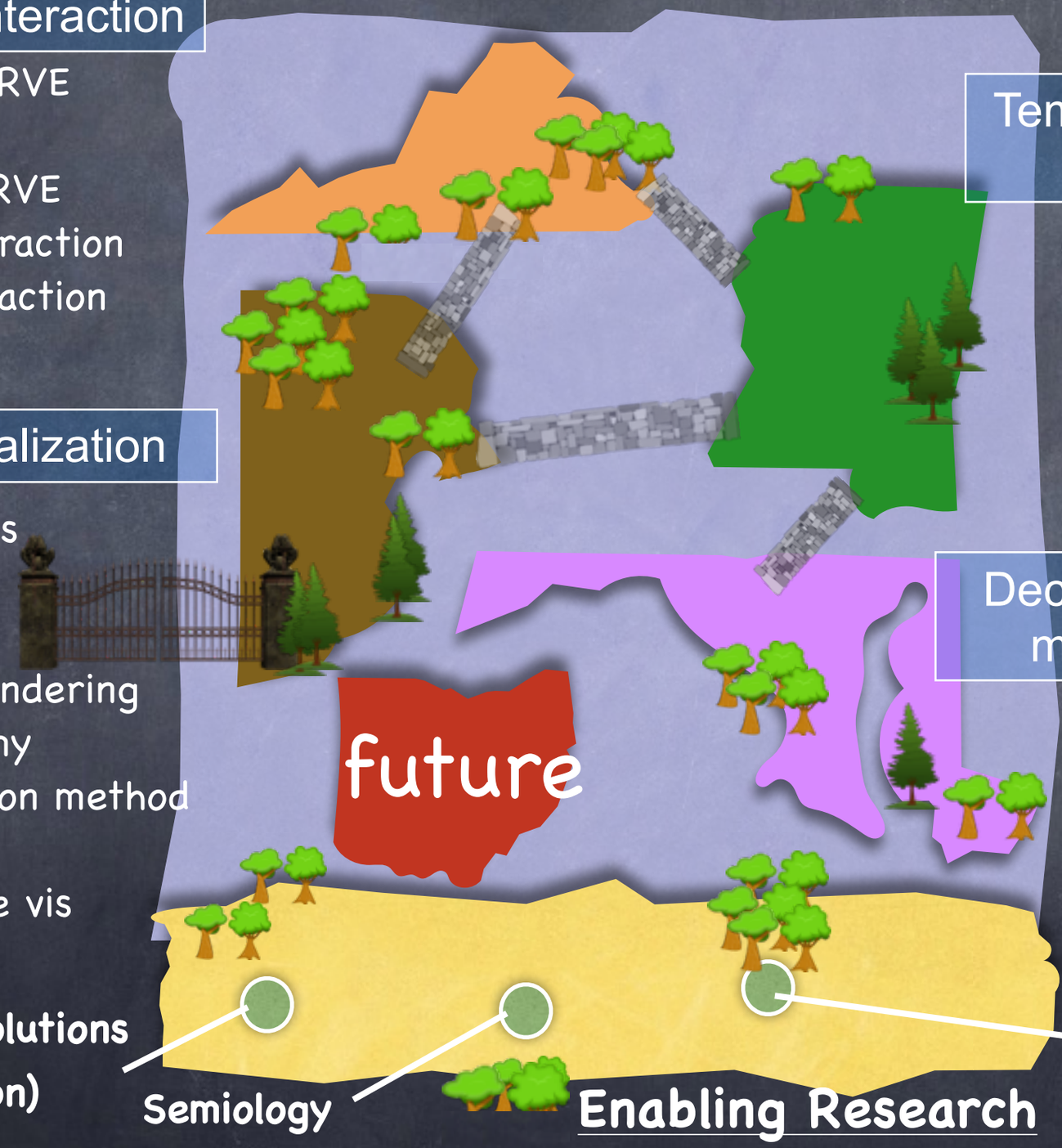
Chen's Garden of Human-Centered Visual Computing

Physical-spatial interaction

Resolution+size on IRVE
Stereo on DMRI
Text navigation in IRVE
Domain-specific interaction
Multi-platform interaction
Display ecology

Theory of Visualization

Language of DMRI vis
Coloring DMRI
Ranking encoding
Non-Photorealistic rendering
Visualization taxonomy
Visualization evaluation method
Visual feature space
Perceptually accurate vis



Temporal-dynamic workflow

Memory-driven
Schema-free exploration

Decision making from multifaceted data

Dimension reduction
Cost of context switching
Tool design

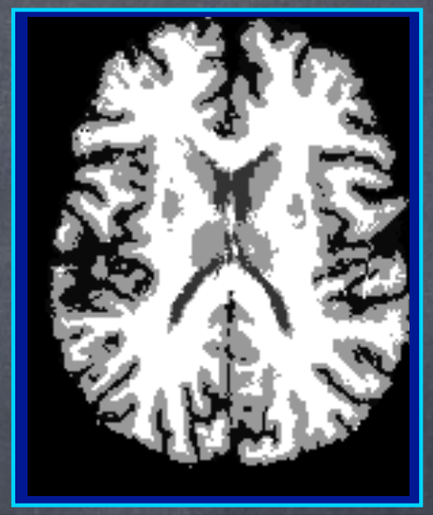
Trend: Perceptually Accurate (Medical) imaging

- Spatialization and dimension reduction
- Encoding accuracy
  - Hybrid search (Jeremy Wolfe)
  - Benefits of VR display
  - Benefits of coloring and other encoding
  - Interactive comparison (show all other patients; change lighting)

Do I Really See A Bone ?

Organizer
Raghu Machiraju, The Ohio State University

Panelists
Chris Johnson, University of Utah
Terry Yoo, National Library Of Medicine
Roger Crawfis, The Ohio State University
David Ebert, Purdue University
Don Stredney, Ohio Supercomputer Center



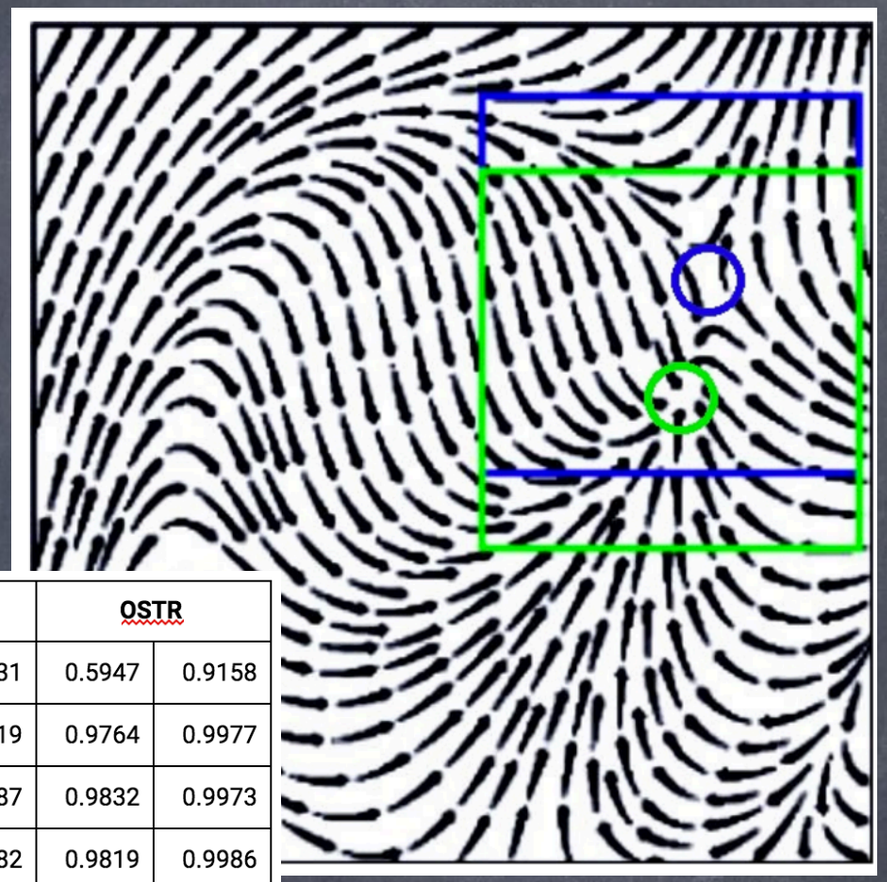
Department of Health and Human Services
Part 1. Overview Information

Participating Organization(s)
National Institutes of Health (NIH)

Components of Participating Organizations
National Cancer Institute (NCI)
National Institute of Biomedical Imaging and Bioengineering (NIBIB)

Funding Opportunity Title
Perception and Cognition Research to Inform Cancer Image Interpretation (R01)

Jeremy Wolfe (Harvard), Cancer imaging



32 Labels	Grid		Jitter		LIC		OSTR	
1st Epoch	0.3007	0.7133	0.3265	0.7649	0.3188	0.7831	0.5947	0.9158
10th Epoch	0.5938	0.9139	0.9008	0.9832	0.7133	0.9719	0.9764	0.9977
20th Epoch	0.9751	0.9937	0.9701	0.9882	0.8261	0.9887	0.9832	0.9973
20th Epoch	0.9728	0.9923	0.9715	0.9855	0.8424	0.9882	0.9819	0.9986
40th Epoch	0.9755	0.9932	0.9719	0.986	0.8406	0.9869	0.9846	0.9977
50th Epoch	0.976	0.9928	0.971	0.9878	0.8419	0.9864	0.9837	0.9982

What visual features can be detected by machines?

Trend: Human-computer collaboration Approach to Construction and Cancer Imaging

- Borrow ideas from computer vision community (imageNet, Vision Genome, Feifei et al.)
  - Vision models
  - Validation:
    - Visual abstraction
    - Visual features (visual GIST) not encoding features (size, color, texture)

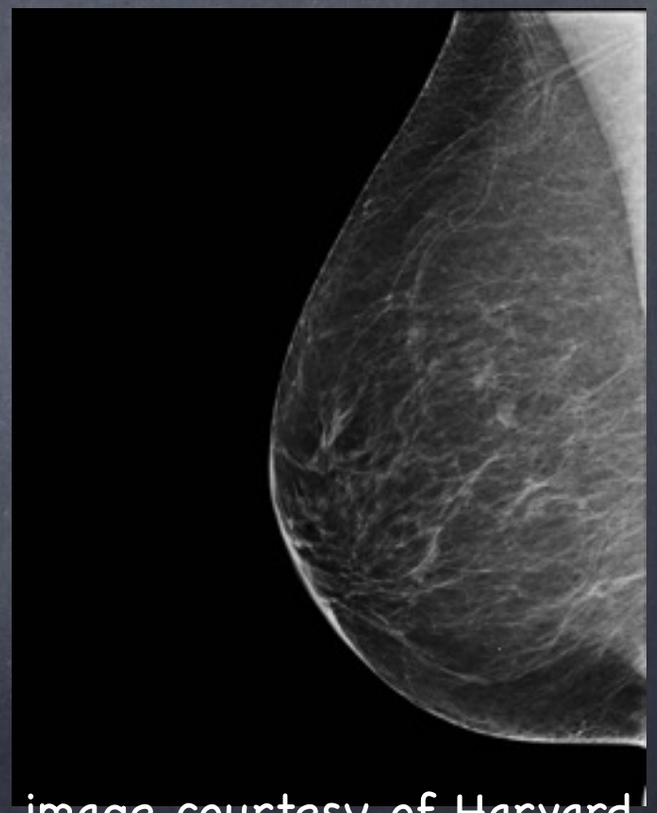


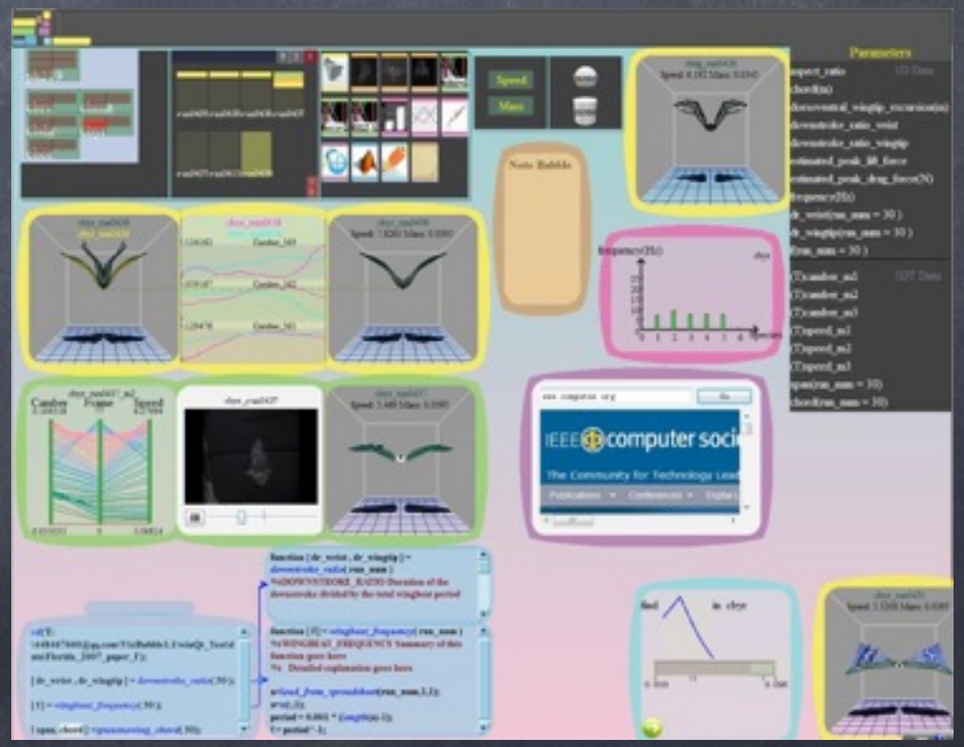
image courtesy of Harvard

Jeremy Wolfe (Harvard), Cancer imaging

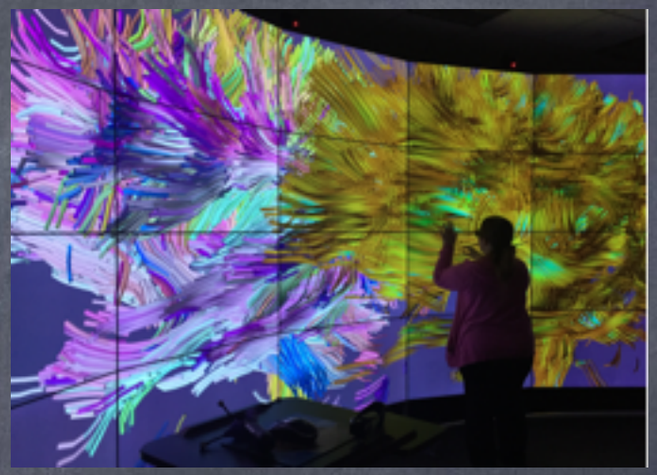
Trend: increased importance of design process (tools, provenance, and teaching)

- Visual ecology (storytelling, creativity, integrating infoVis + sciVis + data mining + learning)

  - Space -> what makes us smarter? (embodied interaction)
  - Mining / learning / recommend workflow
  - Database + perception
  - VisGoogle

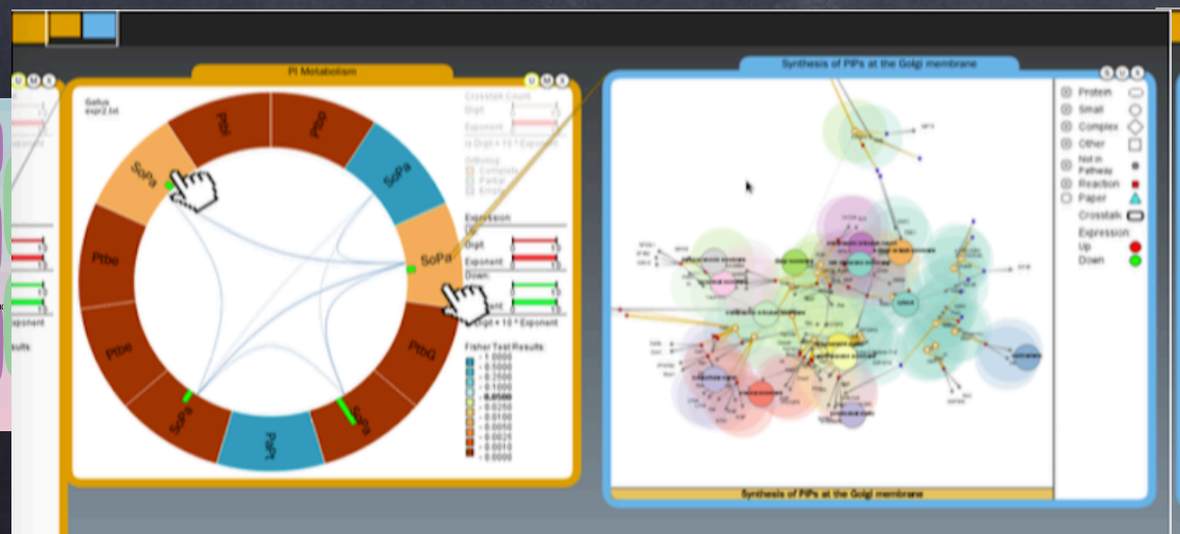


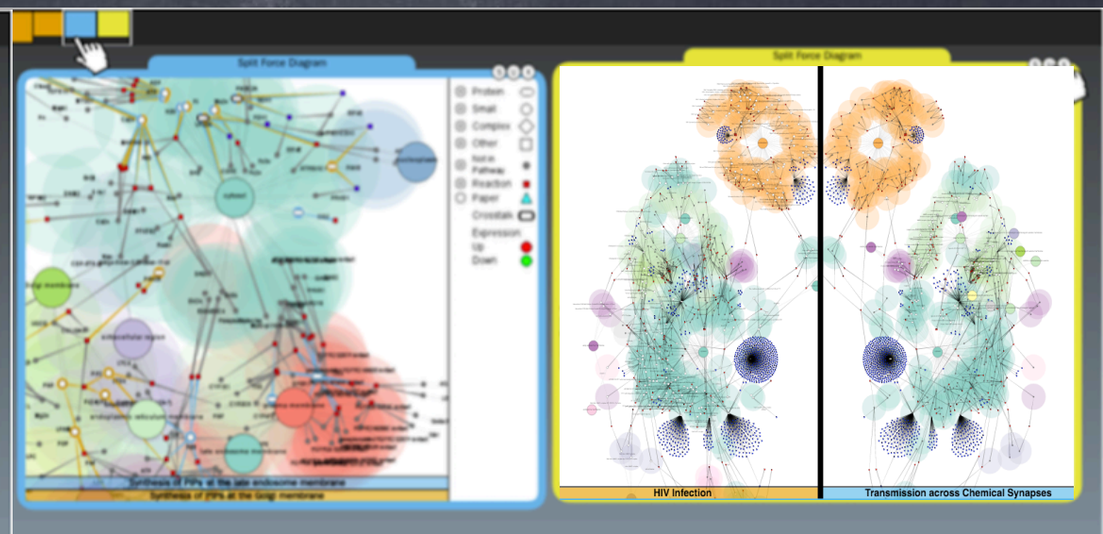
Trend: Immersive Analytics: Metaphors and Interactive Biological Pathway Analytics



- Comparative pathway visualization
  - Big space for big cognition
  - Big space for big algorithm
  - Big space for big collaboration
- Visual debugging

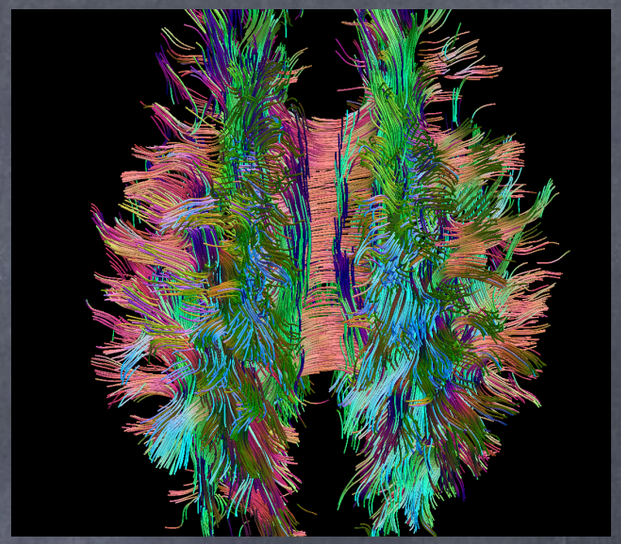
image courtesy of NOAA





Trend: understand uncertainty

- Combining, integrating image-based techniques and the geometry-based methods



- Error bars for measurement errors?

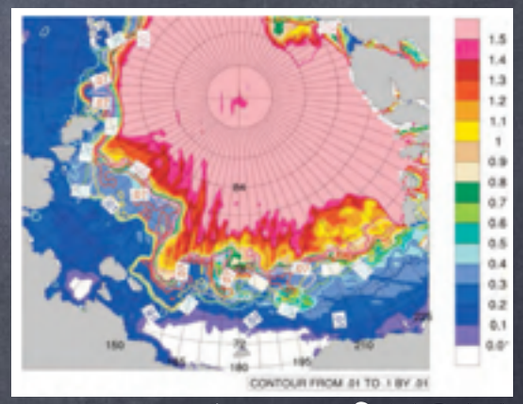
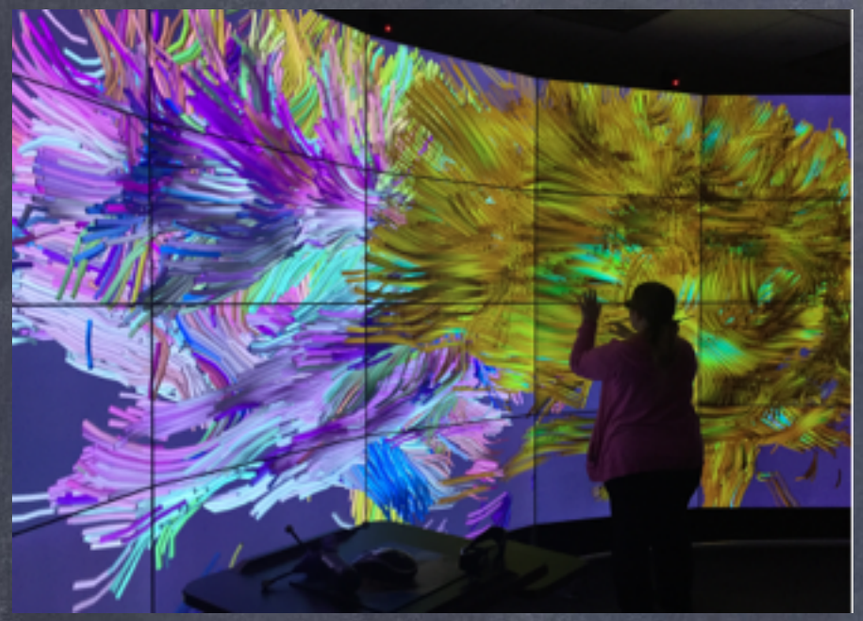
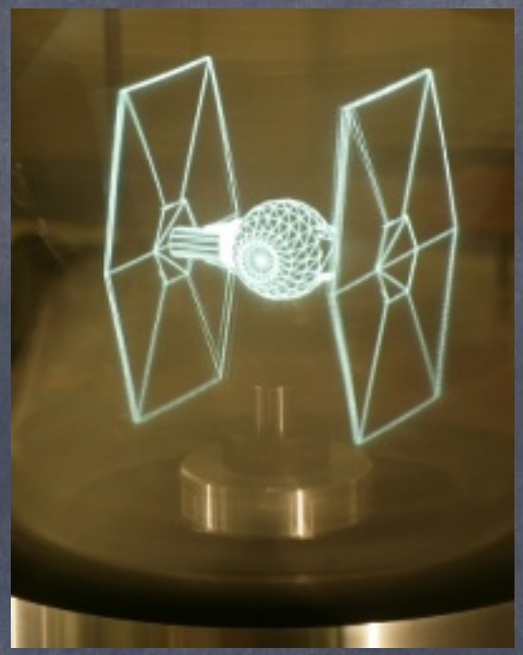


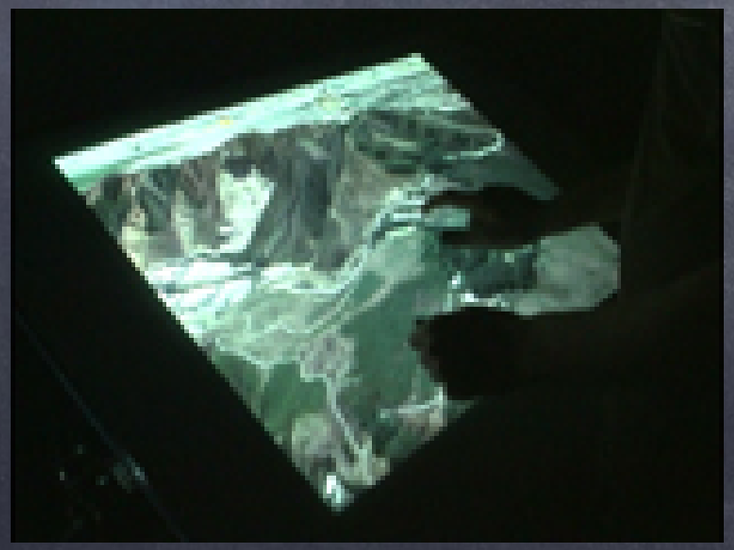
image courtesy of NOAA

- Accessibility:
  - Fast rendering algorithms on mobile devices;
  - Reducing the polygons and showing saliency regions;
  - Languages combining visualization and interactivity

Trend: Display, Exploration, Display Ecology







high-performance computing,
pervasive computing,
Interaction & visualization

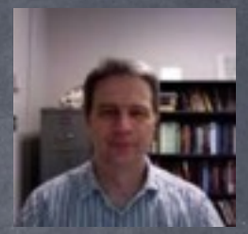

David H. Laidlaw (Brown)

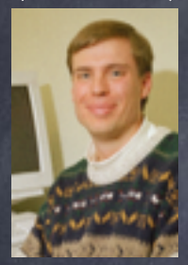

Andries van Dam (Brown)


Alexander P. Auchus (UMC)


Peter Kuchnov (UMB)


Elliot Hong (UMB)


Carl Schmidt (U. Delaware)


Doug Bowman (VT)


Francis Quek (Texas A&M)


Sharon M. Swartz (Brown)

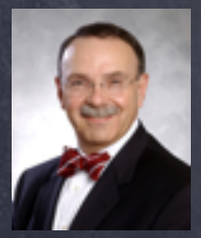

Magenta (Harvard)


Todd Hughes (Next Century)

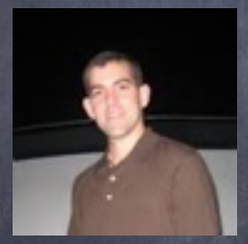

Penny Rheingans


Karl Steiner


Michael Summers


R. Bowen Loftin (U. Missouri)


Andrew Bragdon (Brown)

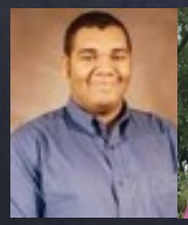

Jesus Caban (DoD)

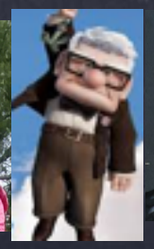
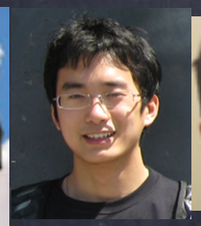

Garnett Bryant (NIST)

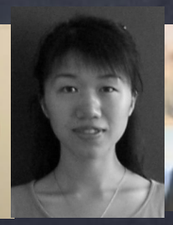
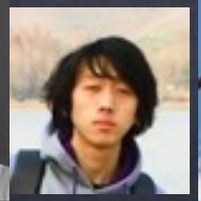
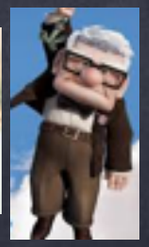

Judith Terrill (NIST)


Craig Saper





Michio Masutani

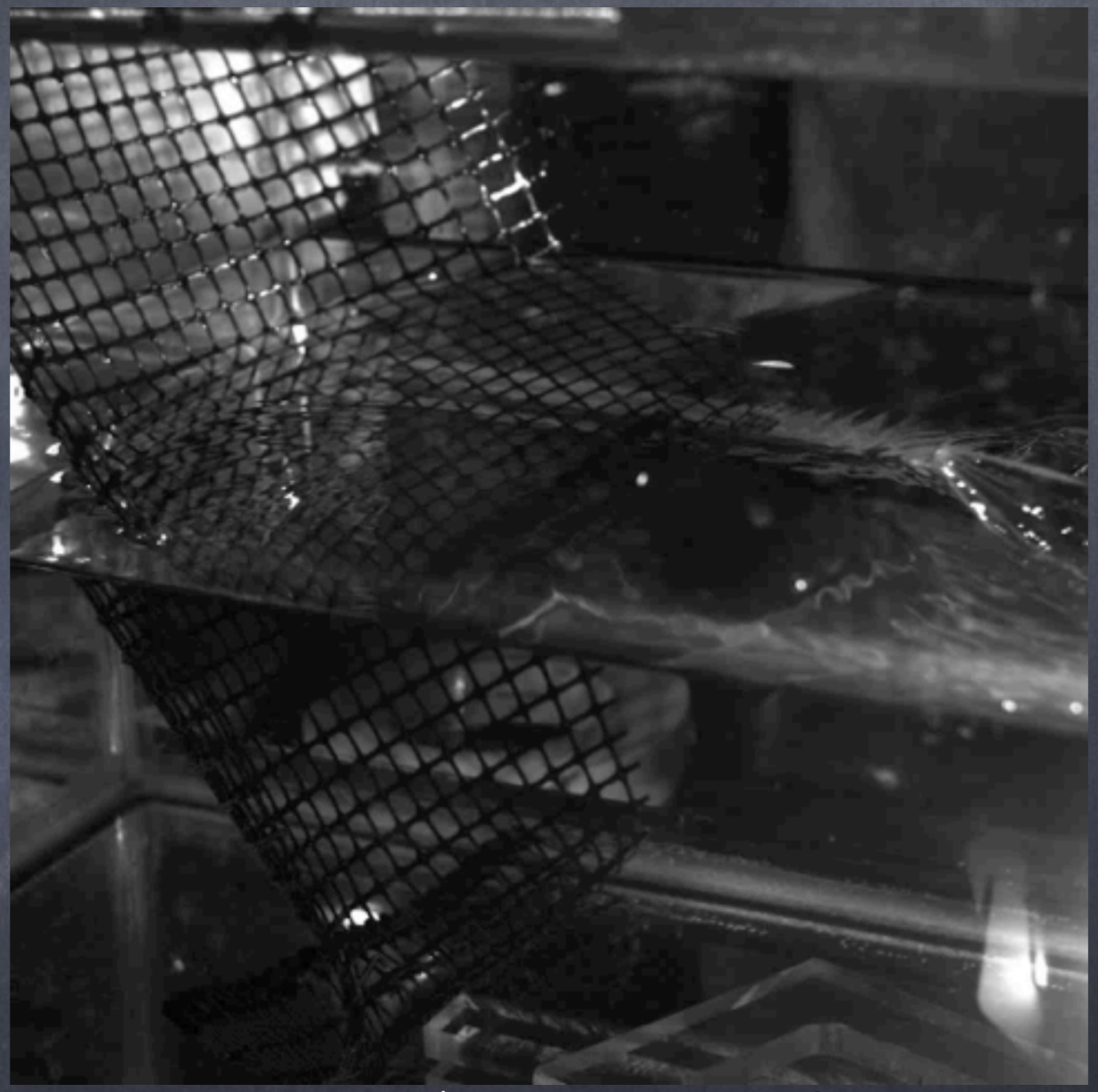
Dr. Keqin Wu

Acknowledgements

- NSF: Supporting knowledge discovery through a 3D scientific visualization language
- NIST: Understanding immersive metrology datasets: scientific and information visualization integration and hybrid input
- DoD: An interactive visualization framework to support exploration and analysis of TBI/PTSD clinical data
- NSF: PathBubbles for dynamic visualization and integration of biological information.
- Other grants: DHS: Sport security; NSF DUE (TUES)

Thank you!

http://www.csee.umbc.edu/~jichen
jianchen0@gmail.com



A swimming bat @ Brown (Video courtesy of the Swartz lab)

A current collaborative project: SwimmingBubbles