Comp/Phys/APSc 715

Evaluation of Visualization
Vector Visualization Redesign

Example Videos

- Vis 2008, Wang: vis-1013_final_video.mp4
  - Focus + context display in 3D
- Vis 2008, Wangchao: idtvdv.avi
  - Importance-driven rendering
- Vis 2008, Zhou: 2008 Vis. Visibility Based Mesh Analysis.submission.mov
  - Importance-driven rendering from CAD model

Administrative

- There will be more data and more questions for all of the projects than was in the HW.
  - To determine, meet with scientist
- Let me know your project preferences
  - Total of 100 points to allocate to all 4
  - More points means more preference
  - Trade-offs to being both client and on team…
  - Email me by tomorrow (Friday)

Evaluation

- "…we often design and evaluate methods by presenting results informally to potential users." [Kosara et al 2003]
  - We will be doing this in this course
  - We'll also add a more formal task but only for a single person doing one task: see instructions
  - Note that even this will be a nontrivial effort – start planning for it now

Potential Types of Evaluation

- Re-use existing designs (art, cartography)
- Hire an expert visual designer to leapfrog into known "best-practice" space
- Videotaping one or more users working with the system
- User Studies: evaluating performance

Why Conduct User Studies?

Why Conduct User Studies?

- Offer scientifically sound method to measure a visualization's performance
  - Accuracy and speed
- Provide insight into why a technique is effective
  - By varying conditions and parameters to see effect
- Determine if theoretical principles derived from psychophysics apply to visualization design
  - Taking the study up one level of complexity

Types of Studies

- Perceptual studies
  - Very simple tasks and stimuli
    - "Which types of texture enhance surface perception, and which camouflage it?"
    - "What is the best color map to display ratio scalar fields with high spatial frequency data?"
- Usability studies
  - User performs a (perhaps complex) task

What to collect?

- Careful statistical data about performance
  - time and error measures
- Close observation of user behavior
  - when did they get frustrated?
  - when did they make errors?
- Free-form comments from the users

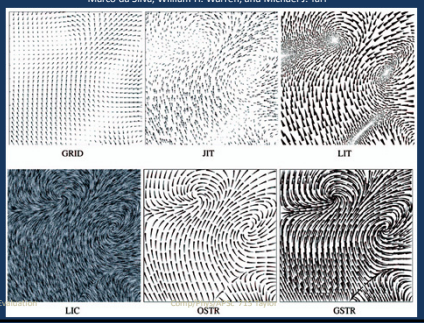
Doing Experiments on People is Serious Business

- Requires serious commitment of time and effort
  - Planning the experiment (seek help from psych!)
  - Evaluating the results (seek help from stat during plan!)
  - Iterating 2-3 times (uncompelling results)
- Requires approval of Institutional Review Board on campus
  - Seeks to preserve respect for and rights of subjects
  - Seeks to prevent new occurrences of egregious past acts of misconduct
- Kosara, et al, report that it is usually worth the effort

Laidlaw Vector Field Study

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 11, NO. 1, JANUARY/FEBRUARY 2005

David H. Laidlaw, Robert M. Kirby, Cullen D. Jackson, J. Scott Davidson, Timothy S. Miller, Marco da Silva, William H. Warren, and Michael J. Tarr



Laidlaw Vector Field Study

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 11, NO. 1, JANUARY/FEBRUARY 2005

David H. Laidlaw, Robert M. Kirby, Cullen D. Jackson, J. Scott Davidson, Timothy S. Miller, Marco da Silva, William H. Warren, and Michael J. Tarr

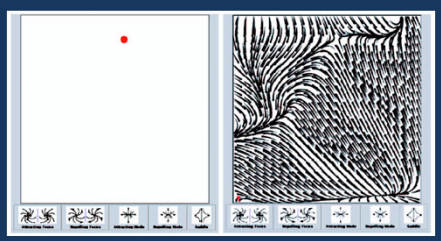
- Question 1: Where are the critical points?

Laidlaw Vector Field Study

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 11, NO. 1, JANUARY/FEBRUARY 2005

David H. Laidlaw, Robert M. Kirby, Cullen D. Jackson, J. Scott Davidson, Timothy S. Miller, Marco da Silva, William H. Warren, and Michael J. Tarr

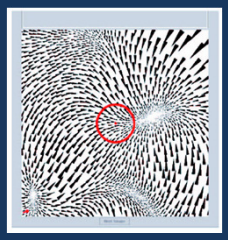
- Question 2: What type of critical point?



Laidlaw Vector Field Study

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 11, NO. 1, JANUARY/FEBRUARY 2005
David H. Laidlaw, Robert M. Kirby, Cullen D. Jackson, J. Scott Davidson, Timothy S. Miller, Marco da Silva, William H. Warren, and Michael J. Tarr

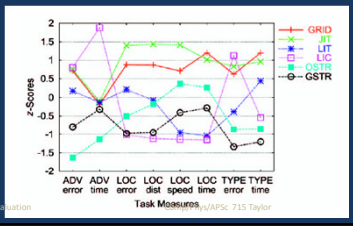
- Question 3: Where would the point go?



Laidlaw Vector Field Study

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 11, NO. 1, JANUARY/FEBRUARY 2005
David H. Laidlaw, Robert M. Kirby, Cullen D. Jackson, J. Scott Davidson, Timothy S. Miller, Marco da Silva, William H. Warren, and Michael J. Tarr

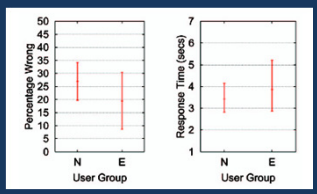
- Results: Which was the best?
  - It depends on the task
  - GSTR better than average on all metrics



Laidlaw Vector Field Study

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 11, NO. 1, JANUARY/FEBRUARY 2005
David H. Laidlaw, Robert M. Kirby, Cullen D. Jackson, J. Scott Davidson, Timothy S. Miller, Marco da Silva, William H. Warren, and Michael J. Tarr

- Of note: Experts and non-experts similar!
  → Brief training sufficient

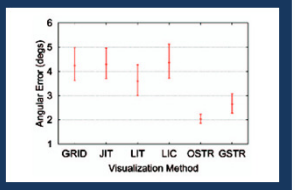


N= non-
E= expert

Laidlaw Vector Field Study

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 11, NO. 1, JANUARY/FEBRUARY 2005
David H. Laidlaw, Robert M. Kirby, Cullen D. Jackson, J. Scott Davidson, Timothy S. Miller, Marco da Silva, William H. Warren, and Michael J. Tarr

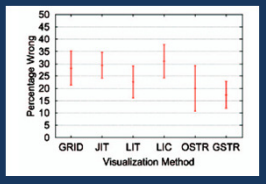
- Of note: Advection was always pretty good!
  - < ~5 degrees of error

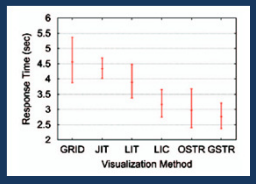


Laidlaw Vector Field Study

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 11, NO. 1, JANUARY/FEBRUARY 2005
David H. Laidlaw, Robert M. Kirby, Cullen D. Jackson, J. Scott Davidson, Timothy S. Miller, Marco da Silva, William H. Warren, and Michael J. Tarr

- Of note: Not significantly better, but faster
  - – for critical point type





Compared to Class Guesses…

Par	*Li*	*Gl1*	*2*	*3*	*4*	*5*	*Tex*	*Sfc*	*Clr*	
L	L+	Ok/4	5	3	2	1	Ok/6	X	X	Sources and sinks (Identify critical points)
L	L+	Ok/4	5	3	2	1	Ok/1	X	X	Sources and sinks (Locate critical points)
Ok	Tr	Ok/4	5	1	3	2	Ok/6	X	Ok	Fast/slow/still (Find zero-flow locations)
Ok	L	Ok					Ok	X	X	Center of rotation
Ok	+	Ok					Ok	X	X	Shape of flow
Ok	++	?					Ok	X	X	Where is flow laminar vs. turbulent?
++	T	L/4	4	3	1	2	Dye/4	L	X	Where would a pushed object end up?
T	T	L					T	L	X	Where does a concentration come from?
Ok	Ok	Ok					?	+	X	Where does stress cause strain?
T	T	Ok					T	T	++	Positive vs. negative field (scalar)?

Evaluation in this class

- Formal
  - Asks primary goal of the scientist
  - On a data set truth is known for (often synthetic)
  - Non-team-member who has not seen the data
- Informal client feedback
  - What new things did the client learn?
  - How is it better/worse than existing tools?
  - How do they like it?

3D DDS User Study

- How well does 3DS work?
  - At what?
  - Compared to what?
- More specific
  - At identifying relationships and extracting values.
  - Compared to other glyph-based technique.

Feng D., Lee, Y., Kwock L., and Taylor, R., "Evaluation of Glyph-based Scalar Multivariate Volume Visualization Techniques," in *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization 2009*. ACM Press, New York, NY, pp. 61-68.

Relationships

- What kinds of relationships?
  - **Linear**
  - Overlap/Intersection
  - Multivariate
  - Etc.
- Data
  - Real? No. goal is to discover relationship
  - Fake? What kind?

Data

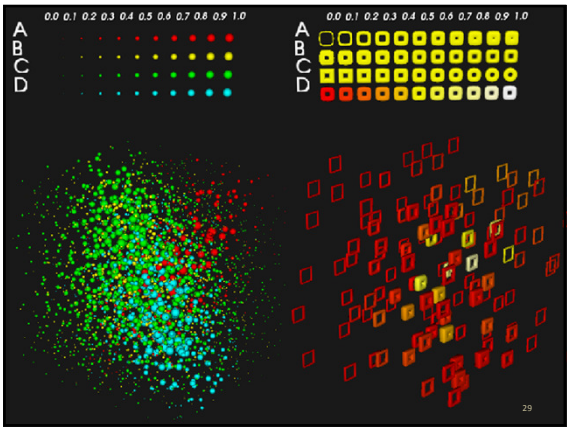
- Application driven
  - Controlled, but resembles original data
  - 3D randomly oriented Gaussian splats
- What resolution?
  - Again, application driven
  - 15x15x15

Compare to SQ Glyphs

- Superquadric glyphs
- Recently published for use in multivariate 3D scalar vis.
- 4 parameters
  - 2 roundnesses
  - thickness
  - color

Legend?

- 2D Legend?
- 3DS
  - It's a 3D glyph, pointless for size-varying
- SQ
  - 4D parameter space.
  - Can't show it all
  - Four examples: full range in 1 var, middle in others.

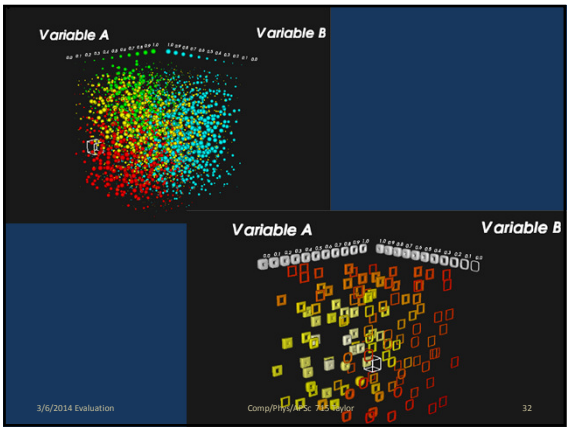


Other Controls

- Control skill level
  - No mouse, keyboard
  - Spacebar for camera rotate, keypad for value selection
- Control environment
  - Dark room
  - 3D stereo glasses, Eye-separation corrected

Value Extraction

- Easy in 2D, how to label a spot in 3D?
  - Dot, sphere, cube…
- Wireframe cube
  - What color?
  - White probably a bad choice in-band for color). Oops.
- Average value? Interpolated value?
  - Confused users…



Between Sub vs. In Sub

- Between subject
  - Compare absolute performance of different participants between conditions
  - Compare Sphere avg to SQ average
- In subject
  - Compare relative performance of different participants between conditions.
  - Average improvement

How to pick?

- Variability between subjects
  - First-person-shooter-playing students might be better?
  - Experts vs. non-experts
- Fewer controls → Variability → More subjects
  - Oh boy…
- Lead David to pick In-Subjects design

What to capture?

- Ideally: everything
  - System Interaction
    - Mouse events, keyboard events, etc
  - **Interviews**
  - **Timing**
  - **Performance**

Details details…

- How many subjects do we need?
  - Run a pilot, ask stats person
- How many trials should each participant do?
  - Run a pilot, ask stats person
- How much training to I need to do?
  - Run a pilot
- What age range to we sample?
- Do you offer compensation? How much?
- How much help to give?
- What do you do with outliers?
- How do I know this applies to my real data?
  - Uh…
- Help, I don't know statistics!
  - Me neither…

Vector Visualization Redesign

- Keller & Keller
- How does wind velocity correlate with temperature?
  - Magnitude
  - Direction
  - Critical Points