CSCI 4550/6550 Interactive Visualization

https://www.cs.rpi.edu/~cutler/classes/visualization/S24/

Lecture 13: Handling Massive or Incomplete Information

Today

- Final Project Brainstorming Ideas Post
- Worksheet: Anonymized Course Grade Visualization
- Discussion about Anonymization & Opt-In/Opt-Out Consent
- Readings for Today:
  - “QSplat: A Multiresolution Point Rendering System for Large Meshes”
  - “LabelMe: online image annotation and applications”
  - “Dual Photography”
- Tangentially Related Graphics/Vision Topics
- Readings for Friday

Brainstorming Final Project Ideas

- Two Different Final Project Ideas
  - Who (audience), why (research question), & what (the visualization)
  - One technical challenge: What makes it difficult?
    What is a potential "risk" for completion? For example:
    - acquiring the data,
    - working with very large data,
    - implementing a new visualization design,
    - implementing a novel interaction scheme, or
    - revising the visualization design to validate your hypothesis.
  - Do you already have a partner?
- Make Submitty forum post by Thursday 2/29
- Read & *reply* to 3 other students by Monday 3/11 (after break)

Today

- Final Project Brainstorming Ideas Post
- Worksheet: Anonymized Course Grade Visualization
- Discussion about Anonymization & Opt-In/Opt-Out Consent
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
  - "Dual Photography"
- Tangentially Related Graphics/Vision Topics
- Readings for Friday

You are given the final semester grades for 10 years of RPI CSCI classes. Sketch an anonymized visualization exploring the *correlation in performance/final grade between specific courses*. E.g., perhaps if you get an 'B+' in Data Structures, you have a 42% chance of getting a 'B+' in Principles of Software. This visualization can help students decide whether retaking a course to improve their grade will positively impact their success at RPI in future terms and their graduation GPA. How would you leverage color and interaction in the visualization design?

```
F21 John Doe CS1 A
F21 Bob Smith DS A-
...
S22 John Doe DS B-
S22 Bob Smith FOCS A
...
F22 John Doe FOCS B+
...
```

- If I got a 'B' in DS, should I repeat it right away?
  Or should I wait until senior year and get an "easy A" in DS?
- If I got a 'C-' in FOCS, will I also struggle in IntroAlgorithms?
  Should I repeat FOCS or just drop the major right now?
- I've got a bad case of senioritis…
  - Which CSCI Options are the easiest?
  - Have the highest percentage of students getting an A or A-?

Piazza

- Question Answering Forum, Instructor Announcements, & Course Materials
- Opt-In for Access to Employers & Career Opportunities…

Classes taken on Piazza

CSCI 424/524: Computer Architecture
CSCI 435 (COLL 400): CSCI 435/535: Software Engin…
CSCI 301: CS 301 Software Development
CSCI 301: CS 301 Software Development
CSCI 241: Data Structures

Should they opt-in, the Piazza Network enables students and employers to connect. Students can contact employers directly, interact with content that employers have created for them, and also search for students and alumni with similar (or different) backgrounds and experiences to explore career opportunities and career guidance. The Piazza Network also enables employers to search for students that might be good fits for their diverse opportunities. Employers can search based on information that students fill out directly in their Piazza profile (graduation year, major, awards, job experience, etc) and information students choose to share from Piazza Q&A. Opted-in students can always control what specific information is and is not shared with employers. No class content is ever shared with employers — just metadata on the class (academic term, class number/title, role [TA vs student]).

Today

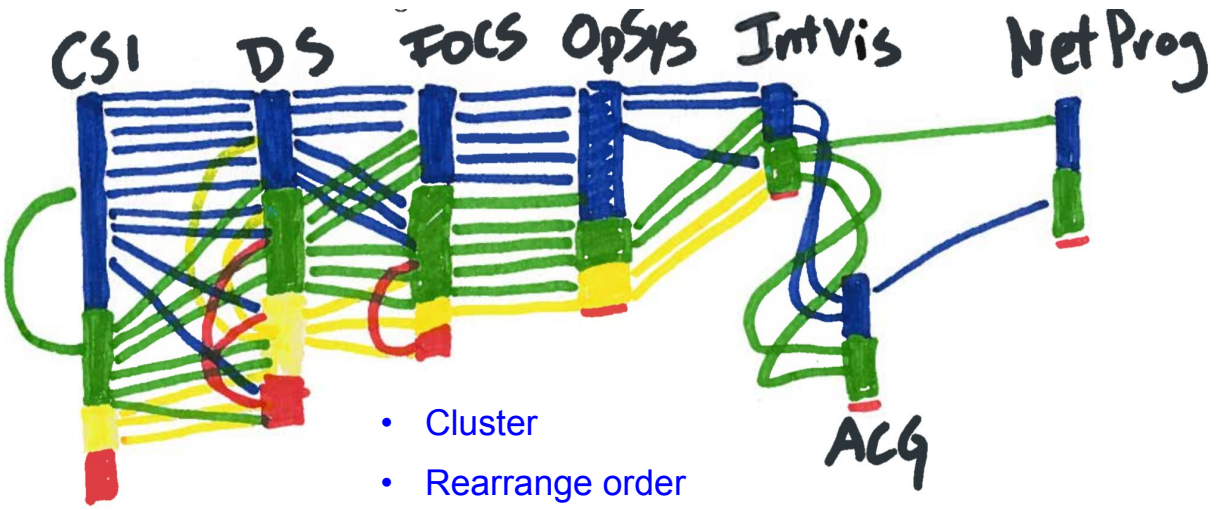
- Final Project Brainstorming Ideas Post
- Worksheet: Anonymized Course Grade Visualization
- Discussion about Anonymization & Opt-In/Opt-Out Consent
- Readings for Today:
  - “QSplat: A Multiresolution Point Rendering System for Large Meshes”
  - “LabelMe: online image annotation and applications”
  - “Dual Photography”
- Tangentially Related Graphics/Vision Topics
- Readings for Friday

Informed Consent / Oversharing on Social Media

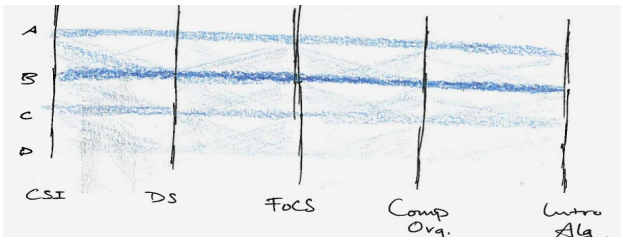
- Do you carefully read every document you sign?
  Every “agree to terms” button you click?
- Data can be taken out of context, used in ways other than intended
- *Previously:* Required a team of researchers to gather data
  Now: Easy for one person. No team discussion of ethical concerns.
- Opt-In / Opt-Out: It’s your choice to use the service! *Is this true?*
- Do we just accept that privacy doesn’t / can’t exist for anyone?
  In the future will we regret what has been shared?
- How do we educate young internet users on (lack of) online privacy?
  On how easy it is for algorithms to connect the dots even without usernames or obvious identifiers?

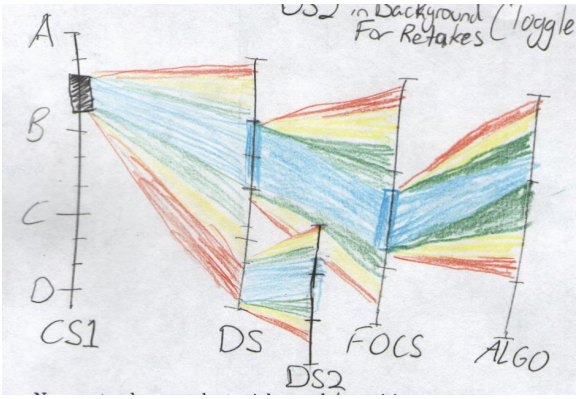
Risks to Users / Participants?

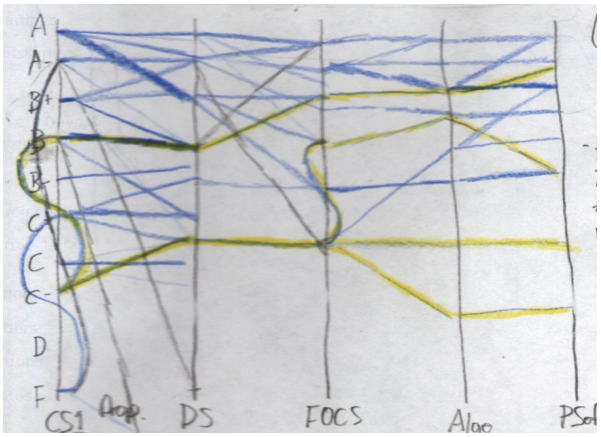
- Quasi-identifiers & Doxing/doxxing (document tracing):
  "Internet-based practice of researching and publishing personally identifiable information about an individual. The methods employed in pursuit of this information range from searching publicly available databases and social media websites like Facebook, to hacking, and social engineering. It is closely related to cyber-vigilantism, hacktivism and cyber-bullying." (definition from Wikipedia)
- If you're not interesting (now or ever in the future),
  Can you assume you have privacy?

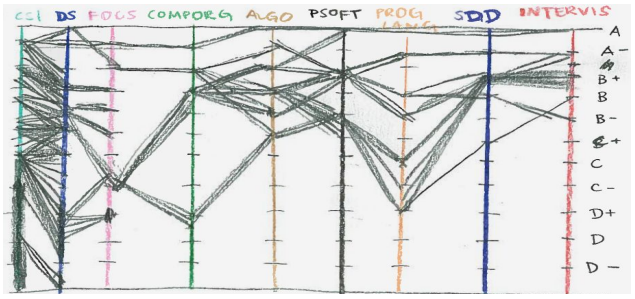


- Cluster
- Rearrange order
  of courses (but still follow prereqs)
- Brush/Filter by students who have
  taken certain courses, achieved
  certain grades, retaken specific course









Now pretend you work at a job search/recruiting company and only have access to the above anonymized visualization and publicly-available data of students from Piazza (don't worry, it doesn't include your final grade). *Note: Most students unknowingly "opt-in" to sharing this data.*

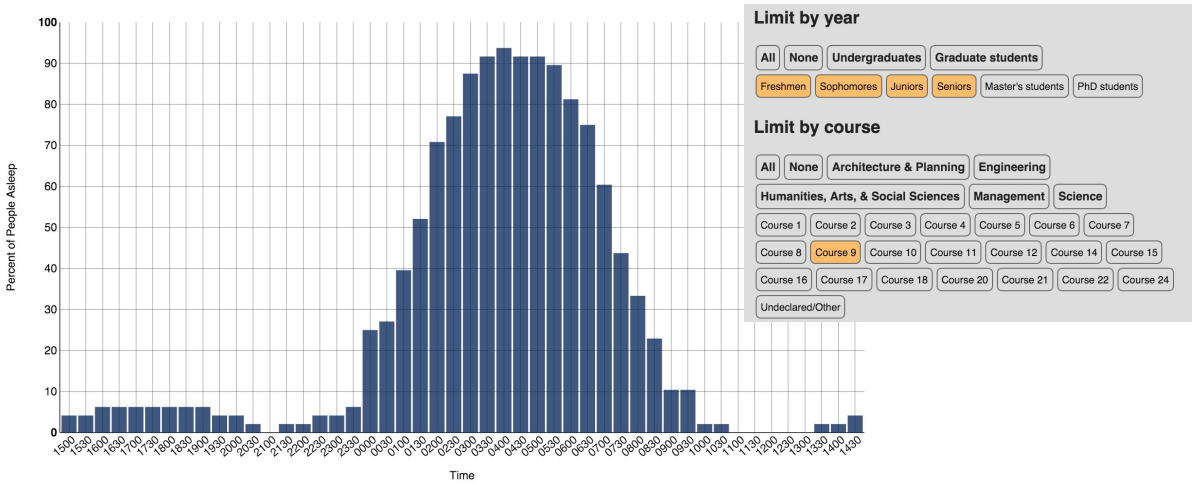
```
Jennifer Brown F20 CS1
Jennifer Brown F21 CompOrg
Jennifer Brown S23 DatabaseSystems
...
```

Is it possible for the job search/recruiting company reverse engineer students' GPA and grades without their permission? Could they automatically discard without human review the applications from students with any blemish on their record? How? Give a specific example. Is this a problem?

- May reveal (with high probability) the identity of some students
  - Small # of students who taken a specific unusual set of courses
  - Risk increases for majors/dual majors with fewer total students
- By repeatedly applying different variants of filters, the viewer (or algorithm) may find patterns. E.g.,
  - Students with weak GPAs select electives with reputation for being less difficult / more grade inflation (GPA boosters)
  - Job application process may only be looking for a reason to reject
- Grades in later courses and hiring decisions should be made on current performance and current abilities *NOT* just on historical performance

When are MIT students asleep?

- Leon Lin and Aaron Scheinberg, MIT Tech



http://tech.mit.edu/V132/N59/pressure/sleepinghours/index.htm

Privacy & Visualization

- Most visualization computation *assumes* unrestricted access to data
- Who would potentially benefit from access to this data?
  - Scientific discovery
  - Improve healthcare
  - Improve education
- What data has privacy concerns?
  - Corporate secrets
  - Health records
  - Academic records
  - Personal finances
  - Personal location

What if instructors for your current courses had access to your full RPI transcript?

FERPA - The Family Educational Rights and Privacy Act

- Students/parents can inspect & review information in their educational records
- Students/parents can request a correction to their record.
- Schools may disclose, without consent, "directory" information
  - @RPI: name, address, photographs, phone #, e-mail, date/location of birth, major field of study, academic load, participation in officially recognized activities and sports, weight and height of members of athletic teams, dates of attendance, degrees,
    honors and awards received, class year in school, and most recent previous educational institution attended
- *However, schools must allow students/parents to opt out of directory information disclosure*
- Students/parents must be regularly informed about their rights

Health Insurance Portability and Accountability Act (HIPAA)

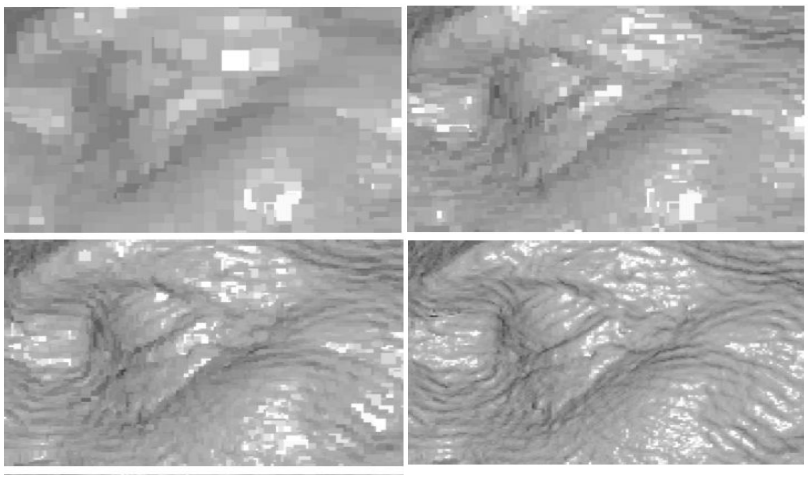
- Long Title: “An Act To amend the Internal Revenue Code of 1986 to improve portability and continuity of health insurance coverage in the group and individual markets, to combat waste, fraud, and abuse in health insurance and health care delivery, to promote the use of medical savings accounts, to improve access to long-term care services and coverage, to simplify the administration of health insurance, and for other purposes.”
- Unintended negative outcomes
  - Reduced retrospective chart-based research (responses dropped from 96% to 34% in one study on heart-attack follow up surveys)
  - Legalistic details on privacy preservation techniques has made informed consent forms even longer and less user-friendly
  - Stiff penalties for violations, lead doctors to withhold information (even sometimes from people who have rights to see it!)
  - Expensive to implement
  - Requires training healthcare providers

Today

- Final Project Brainstorming Ideas Post
- Worksheet: Anonymized Course Grade Visualization
- Discussion about Anonymization & Opt-In/Opt-Out Consent
- Readings for Today:
  - “QSplat: A Multiresolution Point Rendering System for Large Meshes”
  - “LabelMe: online image annotation and applications”
  - “Dual Photography”
- Tangentially Related Graphics/Vision Topics
- Readings for Friday

Reading for Today

- “QSplat: A Multiresolution Point Rendering System for Large Meshes”, Rusinkiewicz & Levoy, SIGGRAPH 2000



Design Criteria & Guaranteed Interactivity

- Low powered computer
  - portable
- Interactive visualization
  - E.g., find holes in data → suggest new scan points
- 127 million points
- Pixel size
  - Screen Space criteria
  - Visibility culling
- File layout & streaming data, pre-fetching

- Good discussion between memory & runtime
- Wanted to jump in and start coding! Explained in sufficient detail to reproduce.
  - Nice comparison of different ways to render splats
- Well-written
  - Why was related work at the end of the paper? (Where is the "right" place?)
  - Discussed the pros & cons of each decision they had to make
- Data structure description was lacking
- Arbitrary constants
- Relatively low frame rate? (5-10 fps)
- Non-traditional rendering algorithm
- Big data + interactivity + rendering
- Question about storing normals/colors

PointShop3D

- Why deal with triangles & connectivity when the triangles are smaller than a pixel?
- “Pointshop 3D: An Interactive System for Point-Based Surface Editing” Zwicker, Pauly, Knoll, Gross, SIGGRAPH 2002

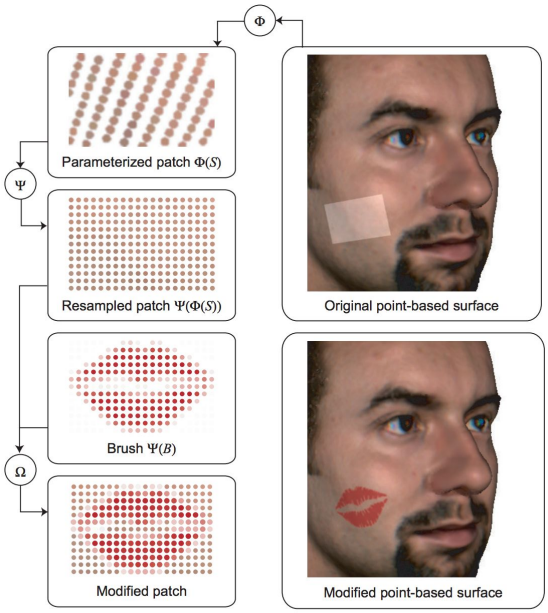


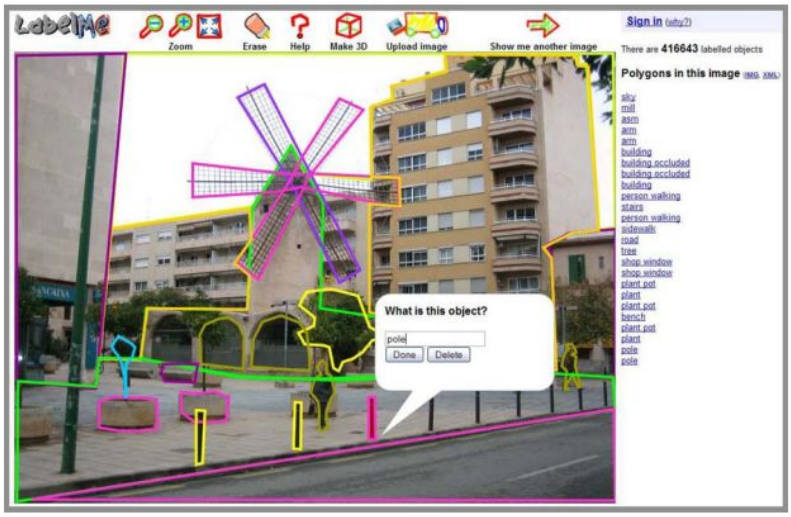
Figure 2: Overview of the operator framework for point-based surface editing.

Today

- Final Project Brainstorming Ideas Post
- Worksheet: Anonymized Course Grade Visualization
- Discussion about Anonymization & Opt-In/Opt-Out Consent
- Readings for Today:
  - “QSplat: A Multiresolution Point Rendering System for Large Meshes”
  - “LabelMe: online image annotation and applications”
  - “Dual Photography”
- Tangentially Related Graphics/Vision Topics
- Readings for Friday

Reading for Today

- "LabelMe: Online image annotation and applications" Torralba, Russell, & Yuen, IEEE 2010



- object recognition under a variety of conditions, object class recognition rather than object instance recognition, not just canonical pose, learning about objects embedded in a scene, not just caption or tag
- number of labels, number of objects with each label still growing over time, who are the labelers?
- visualize closest neighbors to a specific image in database
- visualization of images organized by similarity, smooth transition between some types of scenes
- what types of images are present in the database? lower performance on indoor scenes (fewer examples, more variability of visual appearance)
- automatically recovered spatial relationships between objects (standing on, supported by, supported by, part of)
- Current/Future work: extrapolate scene knowledge to unlabeled portion of image, infer 3D! & extend to video

- "Crowd-sourced" data labeling
  - free & public (lots of data, but will need to deal with some junk data, if it is intentional sabotage it may skew results)
  - Data collection is "less glamorous" than designing new algorithms
  - Limited to user's understanding of the object, and their energy/laziness to completely and accurately outline & label
  - How do they check for bad data?
  - Google's game for helping with image search discontinued because of spam/junk descriptions
- Good choice of saturated colors for contrast with the natural images that are being labeled
- Images should be higher resolution?
- Construct new scenes by pulling parts from other images
  - How easy is it for a 3rd party to download & use this data? Format seems complicated.
- When you collect this much data, something (can't necessarily predict what) will be discovered. But is it and the quantity of other conclusions worth the investment?
  - Conclusions somewhat unreliable
- Security camera footage

Image Segmentation from Approximate Outline

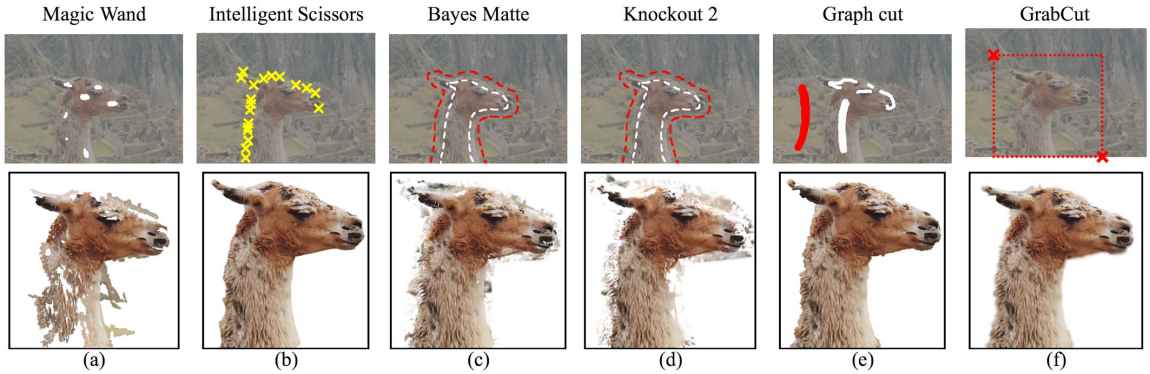
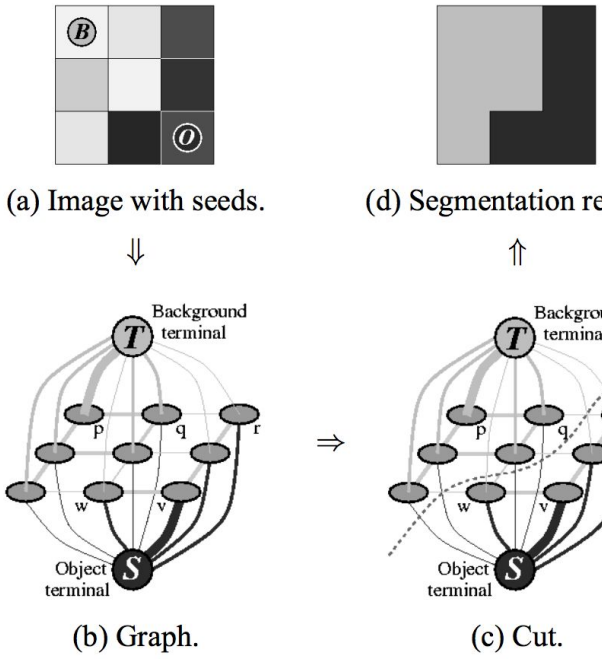
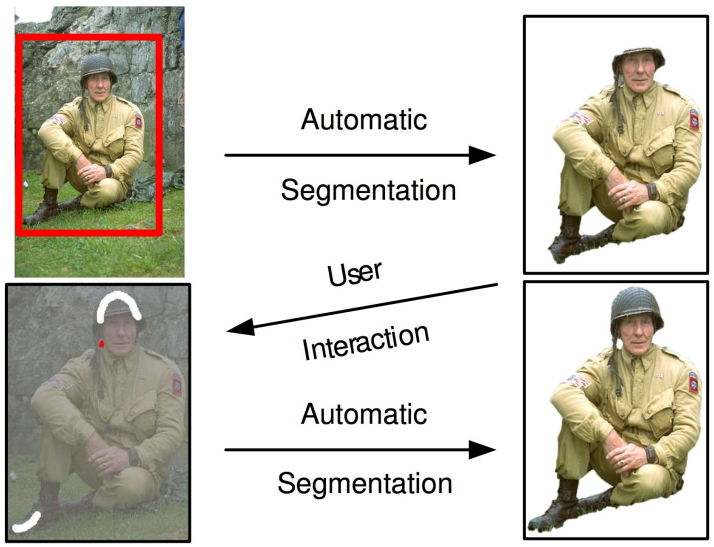


Figure 2: **Comparison of some matting and segmentation tools.** The top row shows the user interaction required to complete the segmentation or matting process: white brush/lasso (foreground), red brush/lasso (background), yellow crosses (boundary). The bottom row illustrates the resulting segmentation. GrabCut appears to outperform the other approaches both in terms of the simplicity of user input and the quality of results. Original images on the top row are displayed with reduced intensity to facilitate overlay; see fig. 1. for original. Note that our implementation of Graph Cut [Boykov and Jolly 2001] uses colour mixture models instead of grey value histograms.

"GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts",
Rother, Kolmogorov, Blake, SIGGRAPH 2004



"Graph Cuts and Efficient N-D Image Segmentation" Boykov & Funka-Lea, IJCV 2006



"GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts", Rother, Kolmogorov, Blake, SIGGRAPH 2004

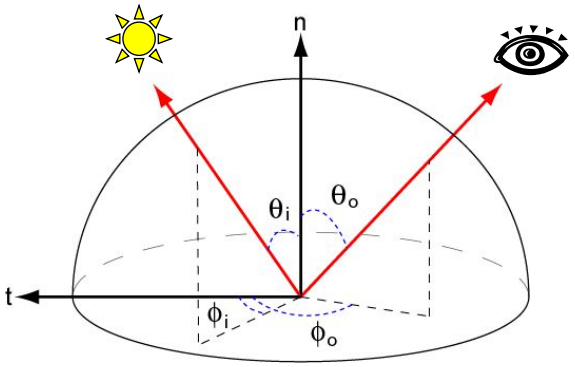
Figure 5: **User editing.** After the initial user interaction and segmentation (top row), further user edits (fig. 3) are necessary. Marking roughly with a foreground brush (white) and a background brush (red) is sufficient to obtain the desired result (bottom row).

Today

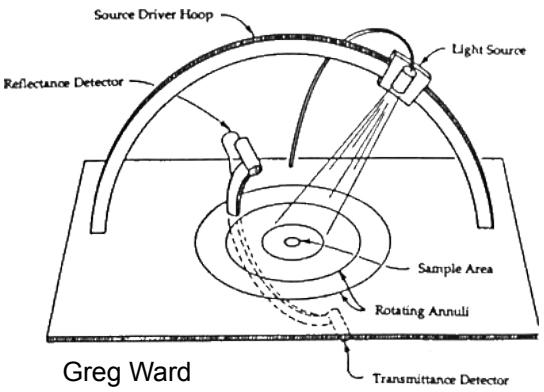
- Final Project Brainstorming Ideas Post
- Worksheet: Anonymized Course Grade Visualization
- Discussion about Anonymization & Opt-In/Opt-Out Consent
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
  - "Dual Photography"
- Tangentially Related Graphics/Vision Topics
- Readings for Friday

BRDF: Bidirectional Reflectance Distribution Function

- Ratio of light from one direction that gets reflected in another direction
- 4D function: incoming θ_i φ_i & outgoing θ_o φ_o



Complete material data capture: Gonioreflectometer

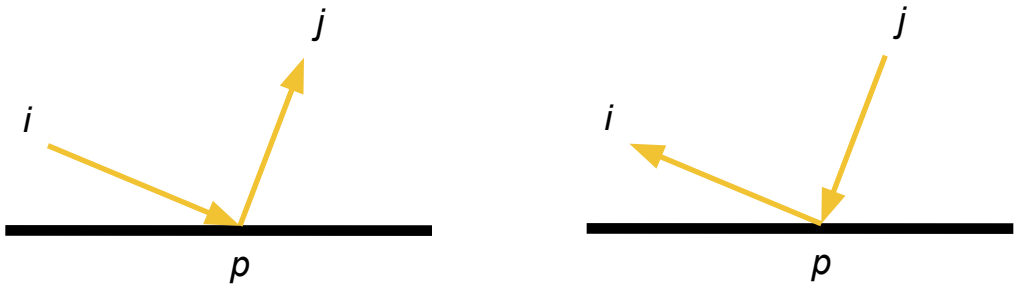


Greg Ward

Helmholtz Reciprocity

- BRDF is symmetric

  % of light reflected from direction *i* off surface point *p* to direction *j*
  is the same as the
  % of light reflected from direction *j* off surface point *p* to direction *i*

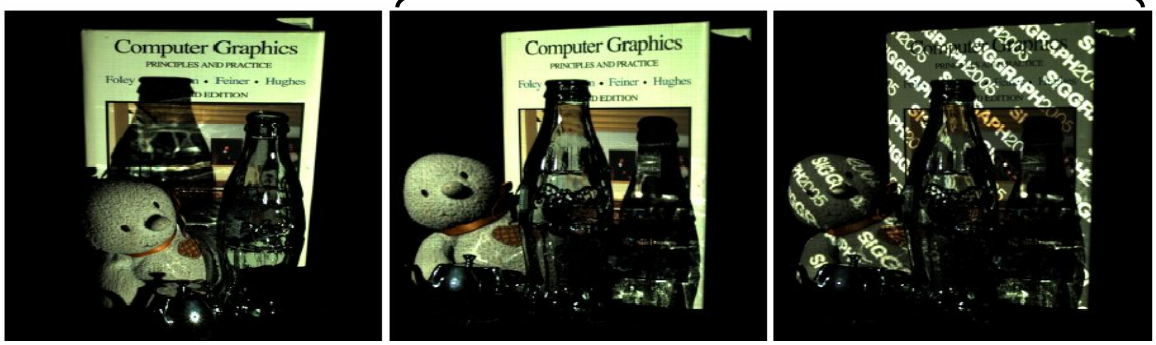


Helmholtz Reciprocity

"Dual Photography",
Sen, Chen, Garg, Marschner, Horowitz,
Levoy, & Lensch, *SIGGRAPH 2005*

real scene

synthetic renderings



Reading for Today

- "Dual Photography", Sen, Chen, Garg, Marschner, Horowitz, Levoy, & Lensch, SIGGRAPH 2005

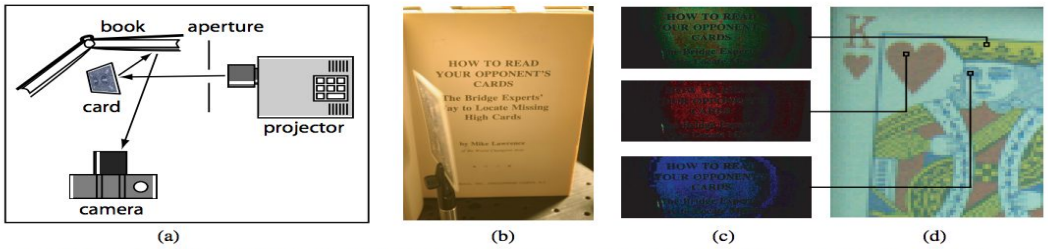
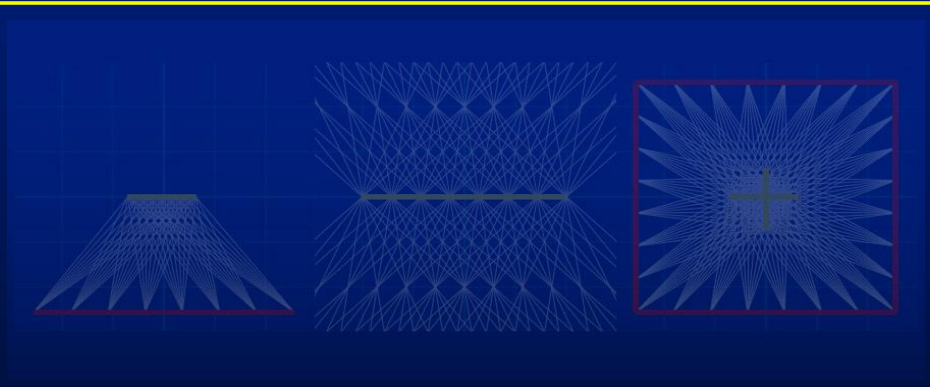


Figure 16: Dual photography with indirect light transport. (a) A projector illuminates the front of a playing card while the camera sees only the back of the card and the diffuse page of the book. An aperture in front of the projector limits the illumination only onto the card. The card was adjusted so that its specular lobe from the projector did not land on the book. Thus, the only light that reached the camera underwent a diffuse bounce at the card and another at the book. (b) Complete camera view under room lighting. The back of the card and the page of the book are visible. It seems impossible to determine the identity of the card from this point of view simply by varying the incident illumination. To acquire the transport matrix, a 3×3 white pixel was scanned by the projector and 5742 images were acquired to produce a dual image of resolution 66×87. (c) Sample images acquired when the projector scanned the indicated points on the card. The dark level has been subtracted and the images gamma-corrected to amplify the contrast. We see that the diffuse reflection changes depending on the color of the card at the point of illumination. After acquiring the T matrix in this manner, we can reconstruct the floodlit dual image (d). It shows the playing card from the perspective of the projector being indirectly lit by the camera. No contrast enhancement has been applied. Note that the resulting image has been automatically antialiased over the area of each projector pixel.

Today

- Final Project Brainstorming Ideas Post
- Worksheet: Anonymized Course Grade Visualization
- Discussion about Anonymization & Opt-In/Opt-Out Consent
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
  - "Dual Photography"
- Tangentially Related Graphics/Vision Topics
- Readings for Friday

Synthetic Aperture Confocal Imaging



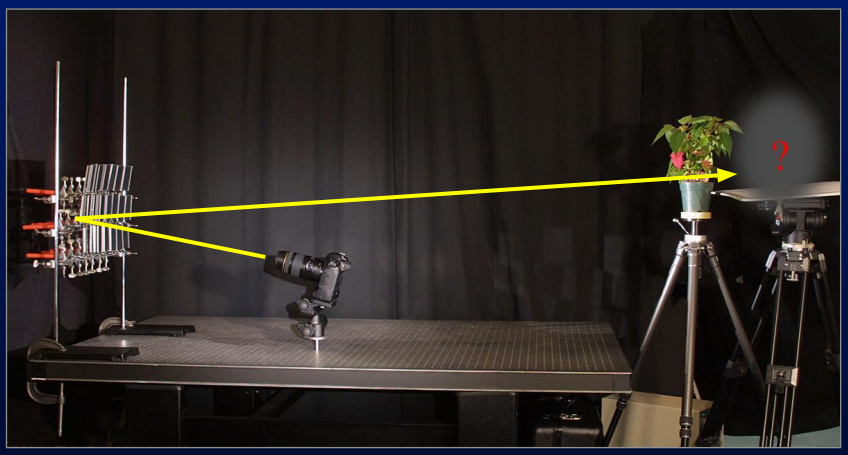
Marc Levoy
Billy Chen
Vaibhav Vaish
Mark Horowitz
Ian McDowall
Mark Bolas

Stanford Multi-Camera Array [Wilburn 2002]

- 640 × 480 pixels × 30fps × 128 cameras

- synchronized timing
- continuous video streaming
- flexible physical arrangement

Synthetic aperture photography using array of mirrors



- 11-megapixel camera
- 22 planar mirrors

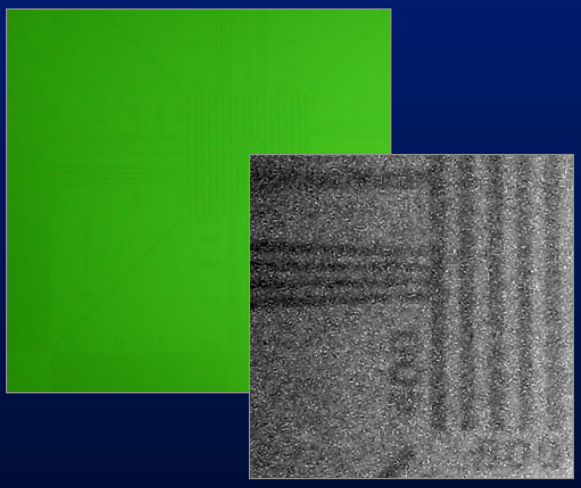
Confocal imaging in scattering media

- small tank
  - too short for attenuation
  - lit by internal reflections

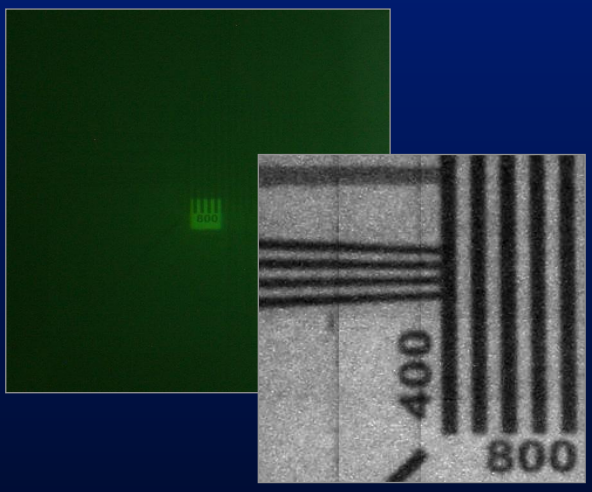
Experiments in a large water tank

- stray light limits performance
- one projector suffices if no occluders

Seeing through turbid water

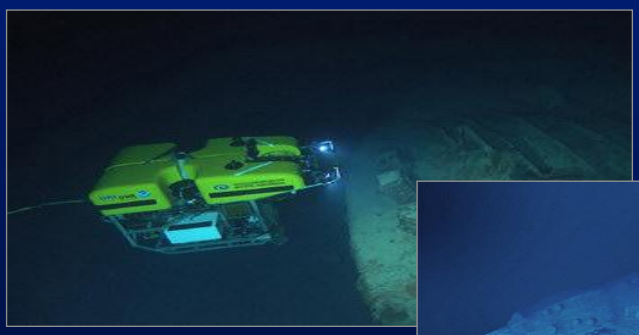


floodlit



scanned tile

Application to underwater exploration



[Ballard/IFE 2004]

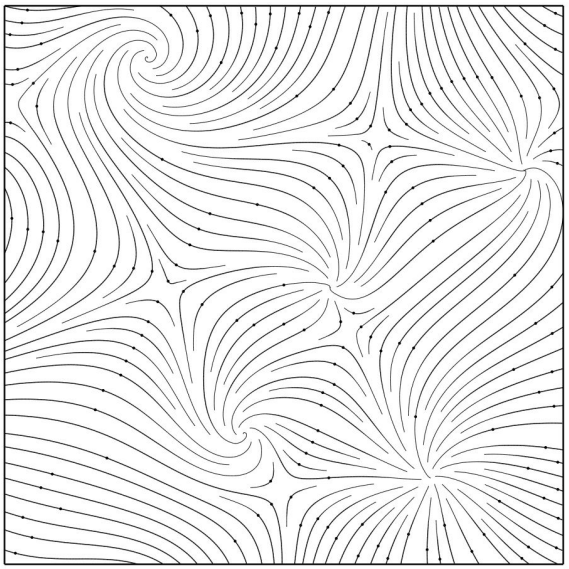
[Ballard/IFE 2004]

Today

- Final Project Brainstorming Ideas Post
- Worksheet: Anonymized Course Grade Visualization
- Discussion about Anonymization & Opt-In/Opt-Out Consent
- Readings for Today:
  - “QSplat: A Multiresolution Point Rendering System for Large Meshes”
  - “LabelMe: online image annotation and applications”
  - “Dual Photography”
- Tangentially Related Graphics/Vision Topics
- Readings for Friday

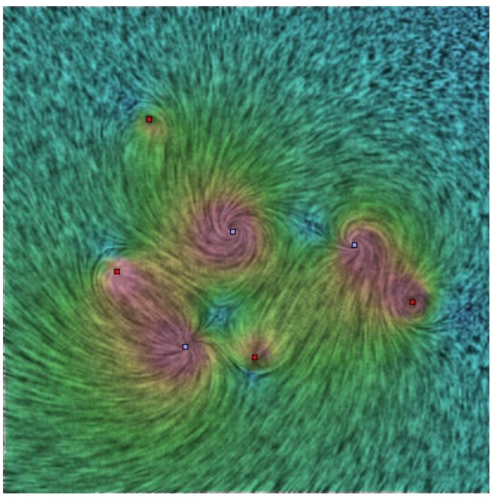
Reading for Friday *pick one*

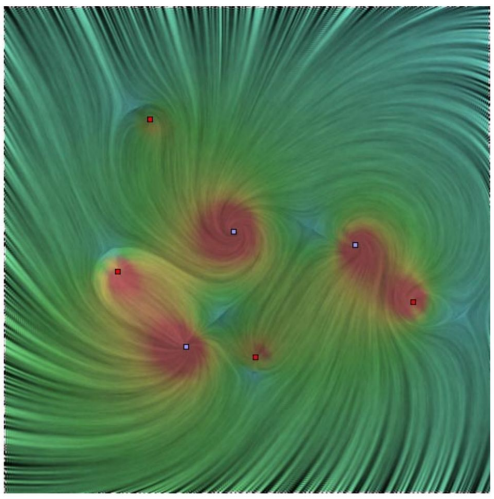
- "Farthest Point Seeding for Efficient Placement of Streamlines", Mebarki, Alliez, & Devillers, IEEE Visualization 2005



Reading for Friday *pick one*

- "Image Based Flow Visualization", Jarke J. van Wij, SIGGRAPH 2002.





Reading for Friday *pick one*

- “Strategy for Seeding 3D Streamlines”, Ye, Kao, & Pang, IEEE Visualization 2005