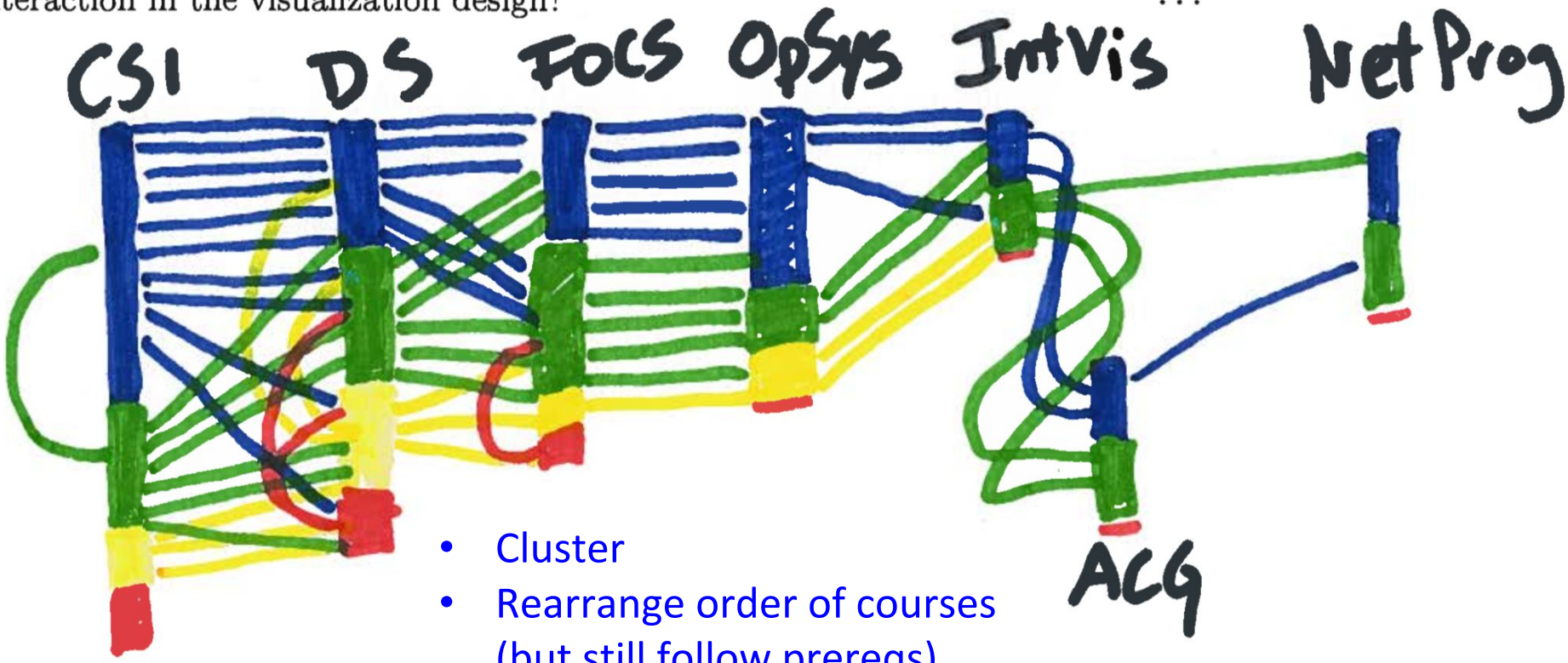
Handling Massive or Incomplete Information

Today

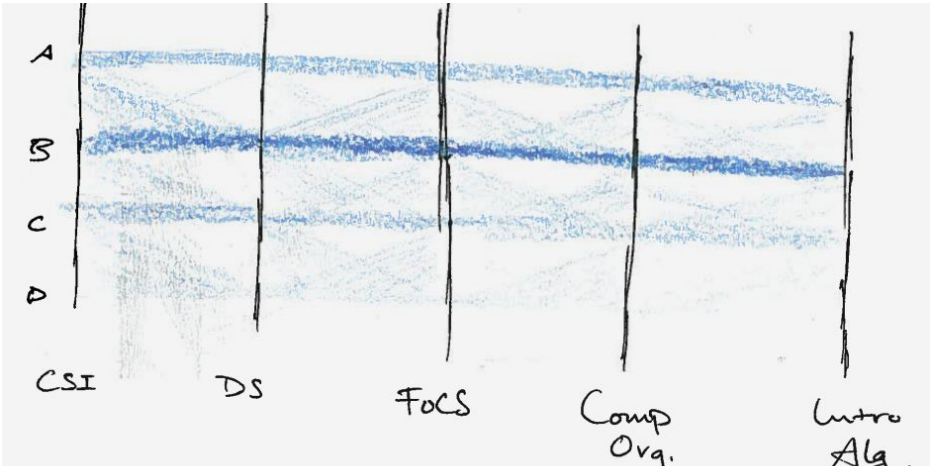
- Discuss Worksheet from Last Week
- Informed Consent & Data Anonymization
- Tangentially Related Graphics/Vision Topics
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
- Readings for Tuesday
- How to brainstorm/foster radical ideas?

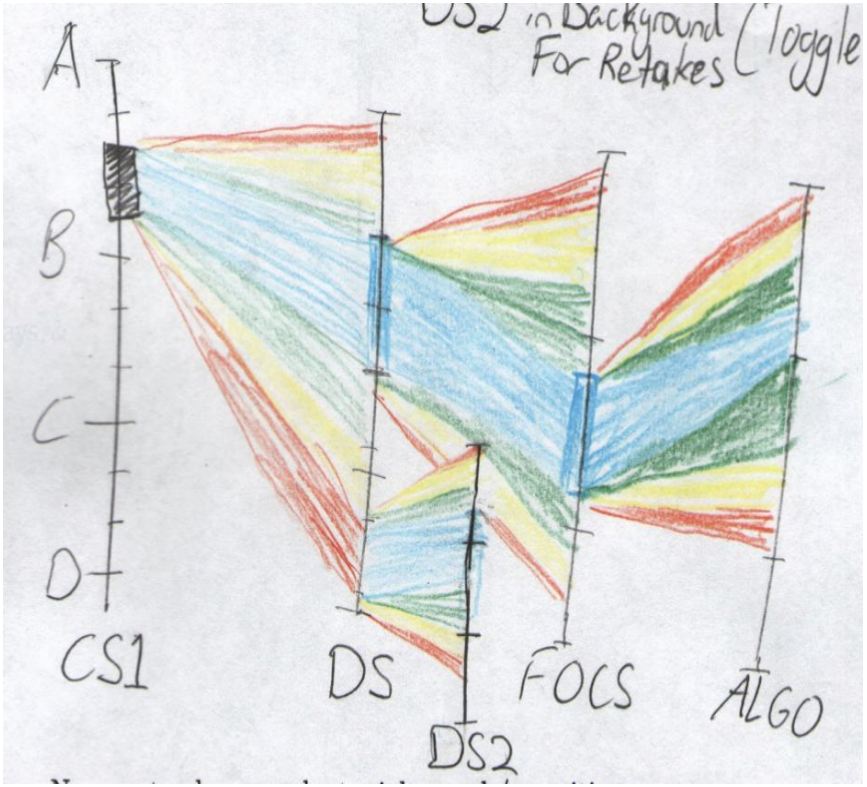
Say you have the final semester grades for 10 years of RPI classes. Sketch an anonymized visualization that explores the correlation in performance/final grade between courses. Let's assume that all offerings of each course are equivalent (no variation between terms/instructors). This visualization can advise students whether immediately repeating a course to improve their understanding of the material and course grade will positively impact their success at RPI in future terms. How would you leverage interaction in the visualization design?

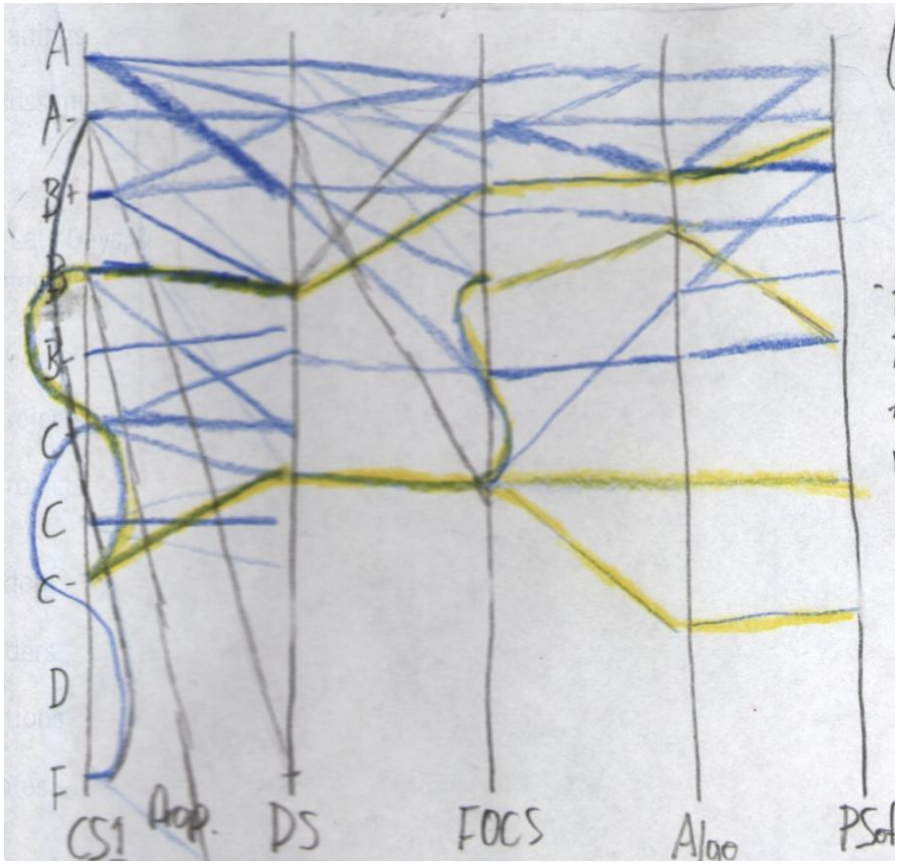
```
F15 John Doe CS1 A
F15 Bob Smith DS A-
...
S16 John Doe DS B-
S16 Bob Smith FOCS A
...
F16 John Doe FOCS B+
...
```

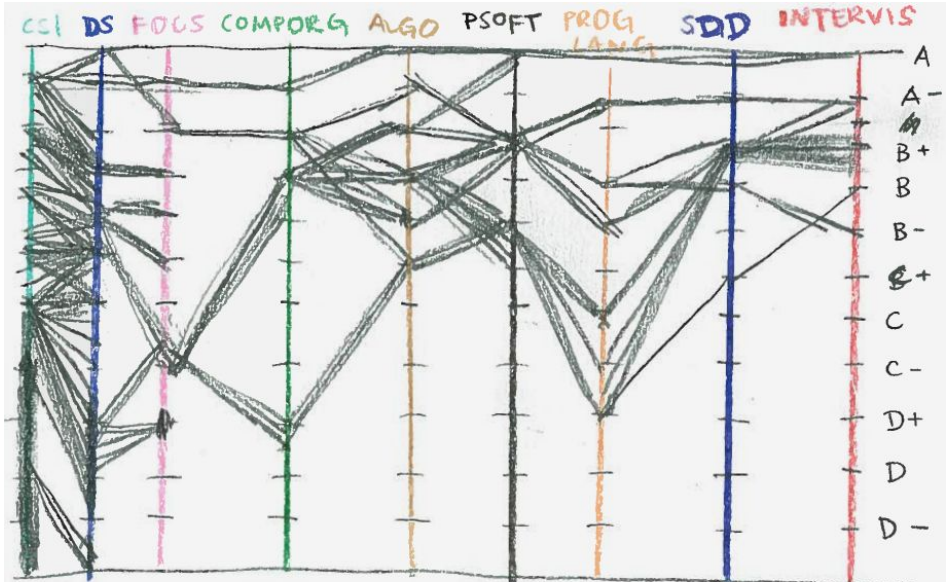


- Cluster
- Rearrange order of courses (but still follow prereqs)
- Brush/Filter by students who have taken certain courses, achieved certain grades, retaken specific course


A
B
C
D
CSI
DS
FoCS
Comp Org.
Intro Alg


DS2 in Background For Retakes (Toggle
A
B
C
D
CS1
DS
DS2
FOCS
ALGO


A
A-
B+
B
B-
C+
C
C-
D
F
CS1
Alop.
DS
FOCS
Algo
PSof


CS1
DS
FOCS
COMPORG
ALGO
PSOFT
PROG LANG
SDD
INTERVIS
A
A-
B+
B
B-
C+
C
C-
D+
D
D-

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

```
Jennifer Brown F14 CS1
Jennifer Brown F15 CompOrg
Jennifer Brown S18 DatabaseSystems
...
```

Can the job search/recruiting company reverse engineer students' GPA and grades without their permission? Can 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 (at least with high probability) the identity of some students
  - Small # of students who taken a specific unusual set of courses
  - Risk increases when applied to other majors/dual majors that have fewer total students with that selection of courses
- Applying repeated filtering, the viewer (or a computer crunching the extracted results) may find patterns. E.g.,
  - Students with weak GPAs select specific electives that have a reputation for being less difficult and/or have more grade inflation (GPA boosters)
  - Often when screening applications for a job or graduate school or papers submitted to for publication, the reviewer is looking for a reason to reject
- Assignment of grades in later courses, and hiring decisions should be made on current performance and current abilities *NOT* just on historical performance

Today

- Discuss Worksheet from Last Week
- Informed Consent & Data Anonymization
- Tangentially Related Graphics/Vision Topics
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
- Readings for Tuesday
- How to brainstorm/foster radical ideas?

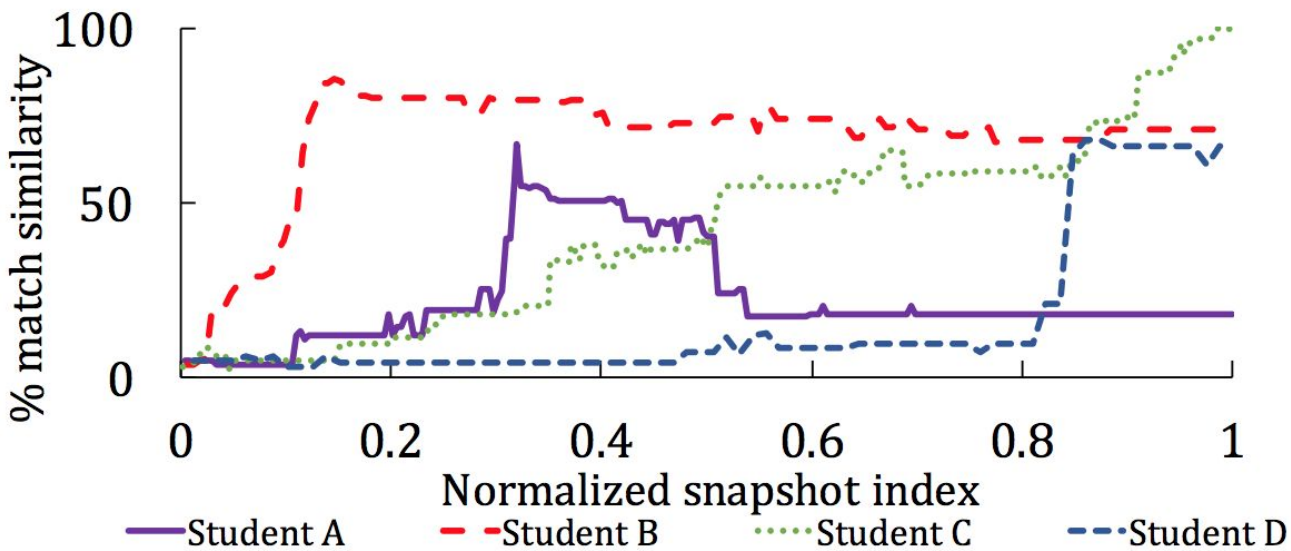
FERPA - The Family Educational Rights and Privacy Act

- 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

- Submitty stores all of your submissions
  - It's your choice when & what to submit

- What if we asked you to install a plugin for your IDE that:
  - Captured your files after every save? every keystroke?
  - Watched what other programs were used simultaneously?
  - Saved your physical location & who you were with
  - Spied through your camera/microphone?

This is creepy, we have no intention of doing this!



"TMOSS: Using Intermediate Assignment Work to Understand Excessive Collaboration in Large Classes", Yan et al, SIGCSE 2018

Privacy & Visualization

- Most visualization computation *assumes* unrestricted access to data
- How do we do this computation with partial information?
- How do you design hardware/software system to ensure data security?

- Who would potentially benefit from access to this data? (Why is this a grey area?)
  - 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?

Risks to users/participants?

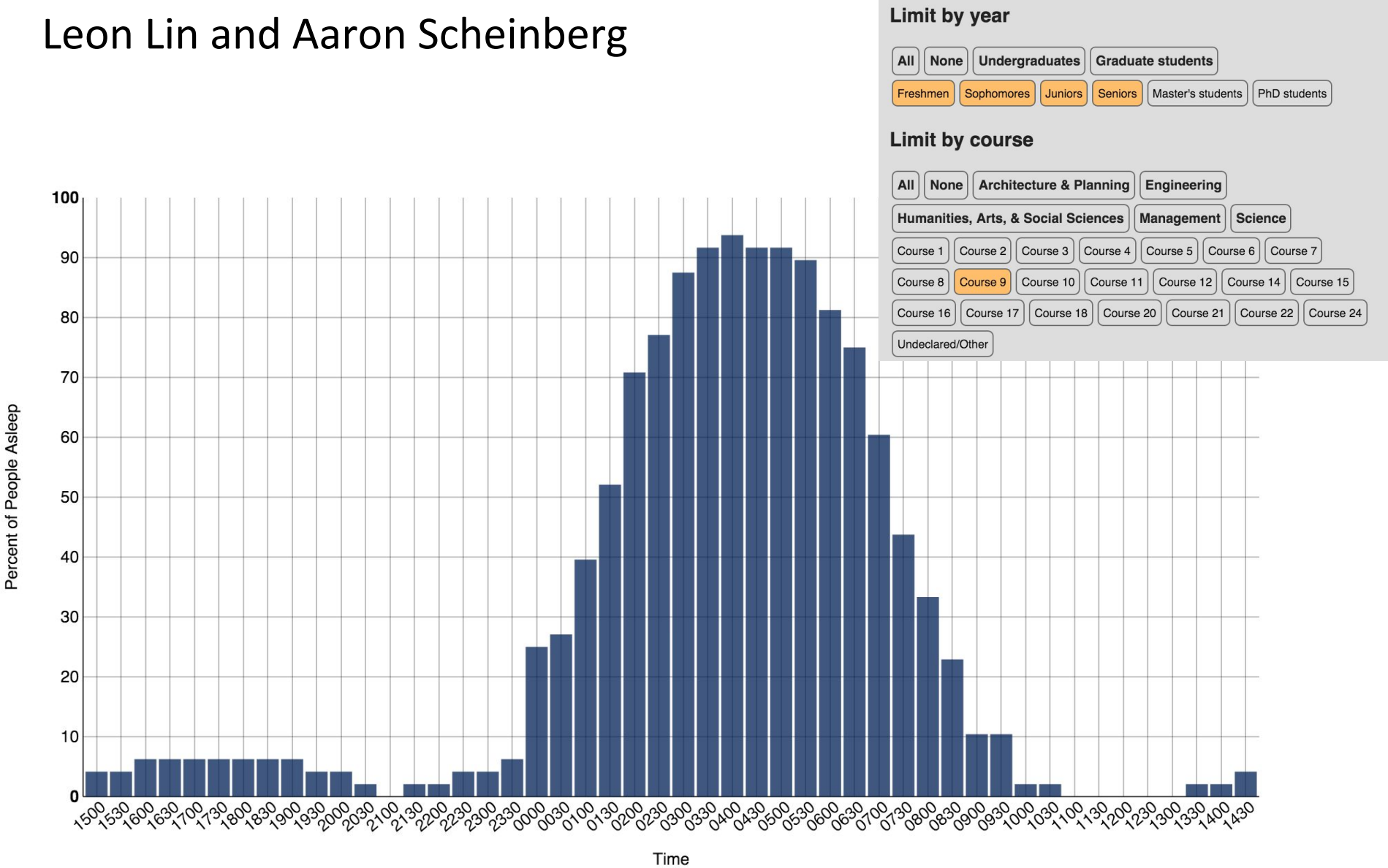
- Quasi-identifiers & Doxing/doxxing (document tracing): "Internet-based practice of researching and publishing personally identifiable information about an individual. The methods employed in pursuit of this information range from searching publicly available databases and social media websites like 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), you probably have privacy?

Facebook generation of oversharers?

- It’s your choice to share or not use the service at all (Is this true?)
- Generation that believes privacy doesn’t/can’t exist for anyone. Is there now or will there be regret for what has been shared?
- Do we have an obligation to educate young internet users on (lack of) online privacy? On how easy it is to connect the dots even without usernames or obvious identifiers?

When are MIT students asleep?

Leon Lin and Aaron Scheinberg



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

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 followup 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

- Discuss Worksheet from Last Week
- Informed Consent & Data Anonymization
- Tangentially Related Graphics/Vision Topics
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
- Readings for Tuesday
- How to brainstorm/foster radical ideas?

Synthetic aperture confocal imaging

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

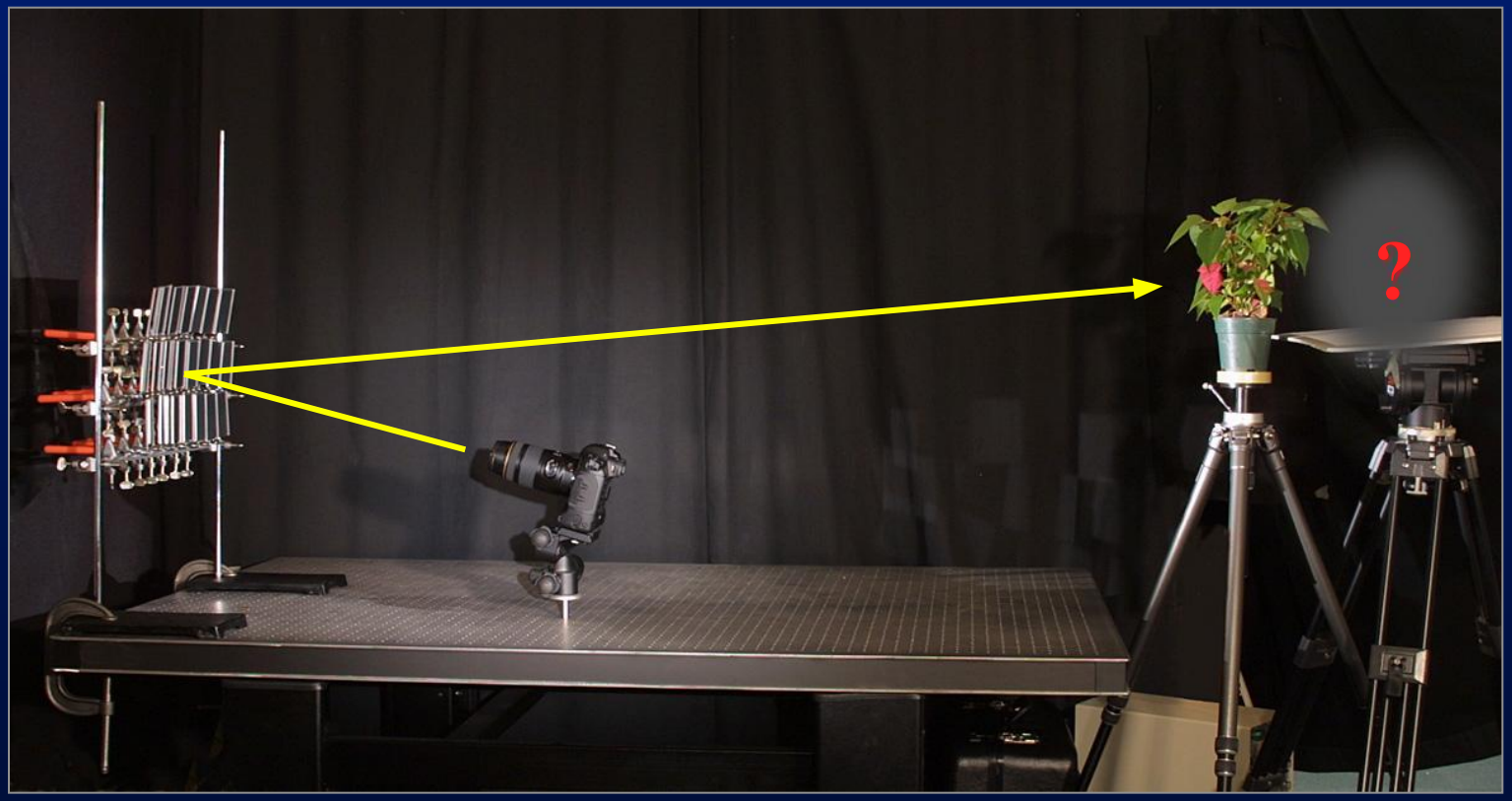
Stanford Multi-Camera Array

[Wilburn 2002]

- 640 × 480 pixels × 30fps × 128 cameras

- synchronized timing
- continuous video streaming
- flexible physical arrangement

Synthetic aperture photography using an array of mirrors



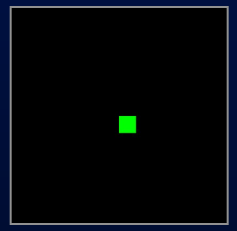
- 11-megapixel camera
- 22 planar mirrors

Confocal imaging in scattering media

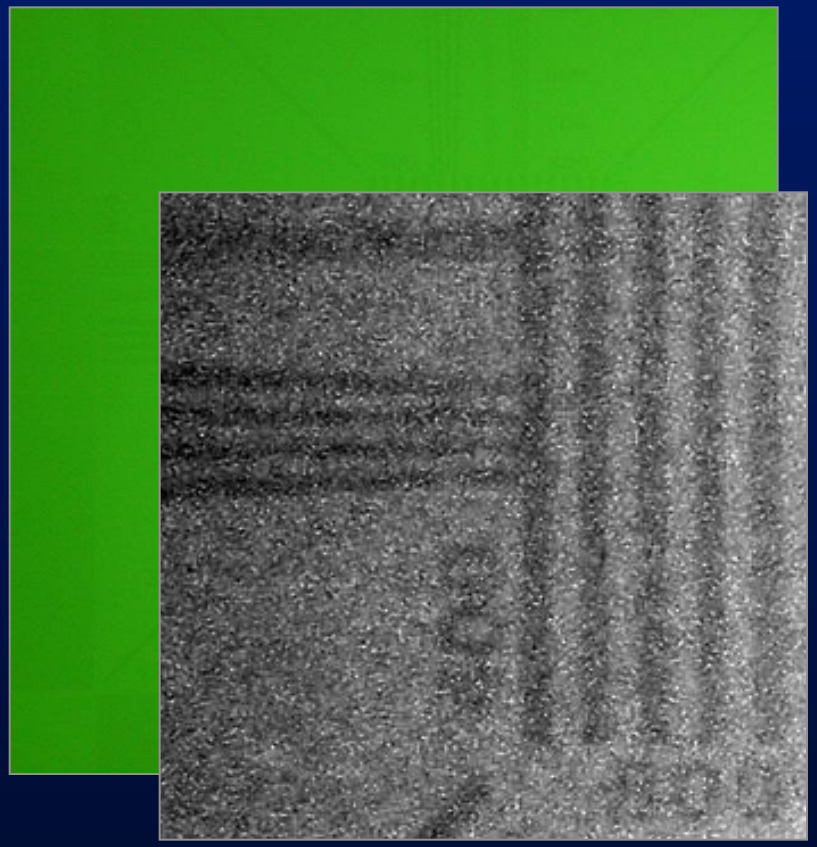
- small tank
  - too short for attenuation
  - lit by internal reflections

Experiments in a large water tank

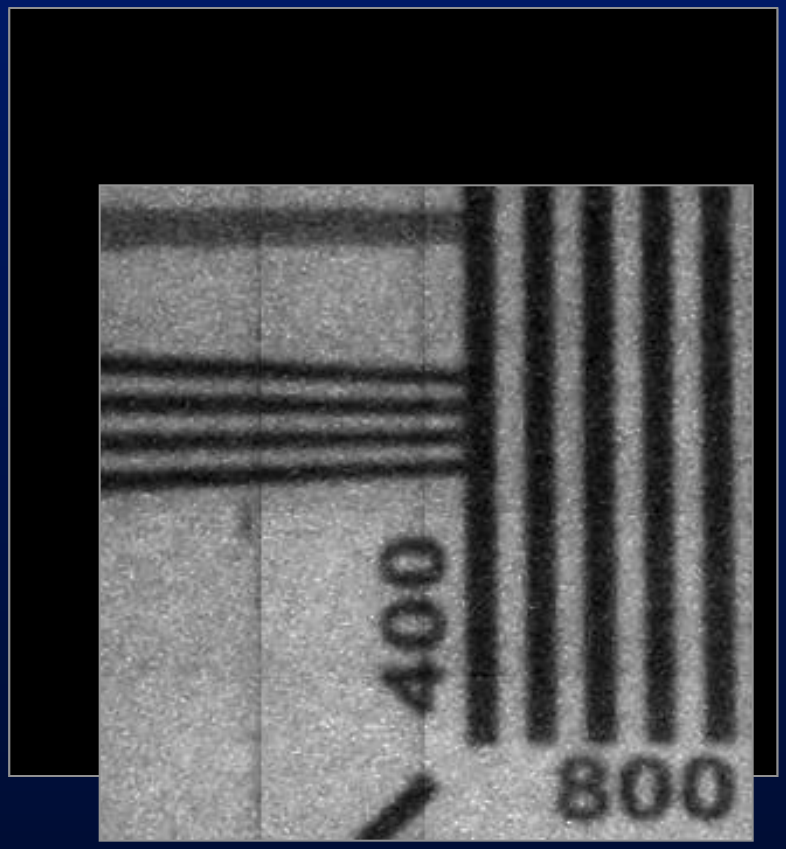
- stray light limits performance
- one projector suffices if no occluders



Seeing through turbid water

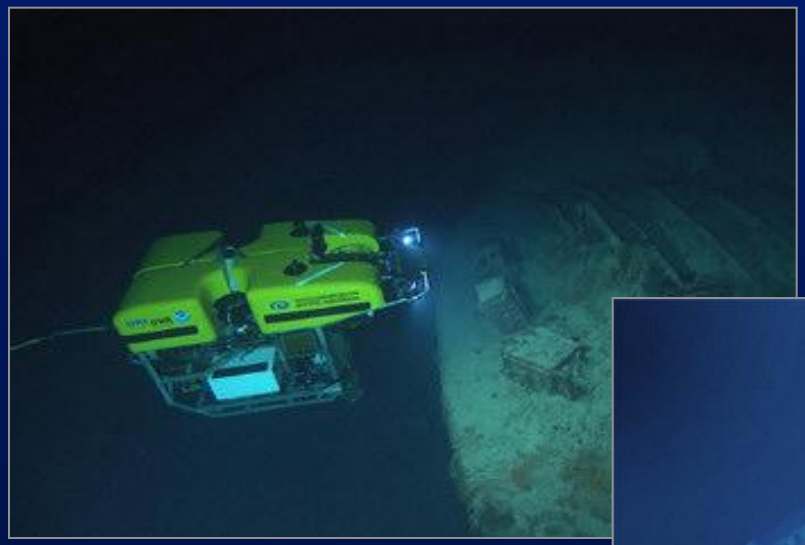


floodlit



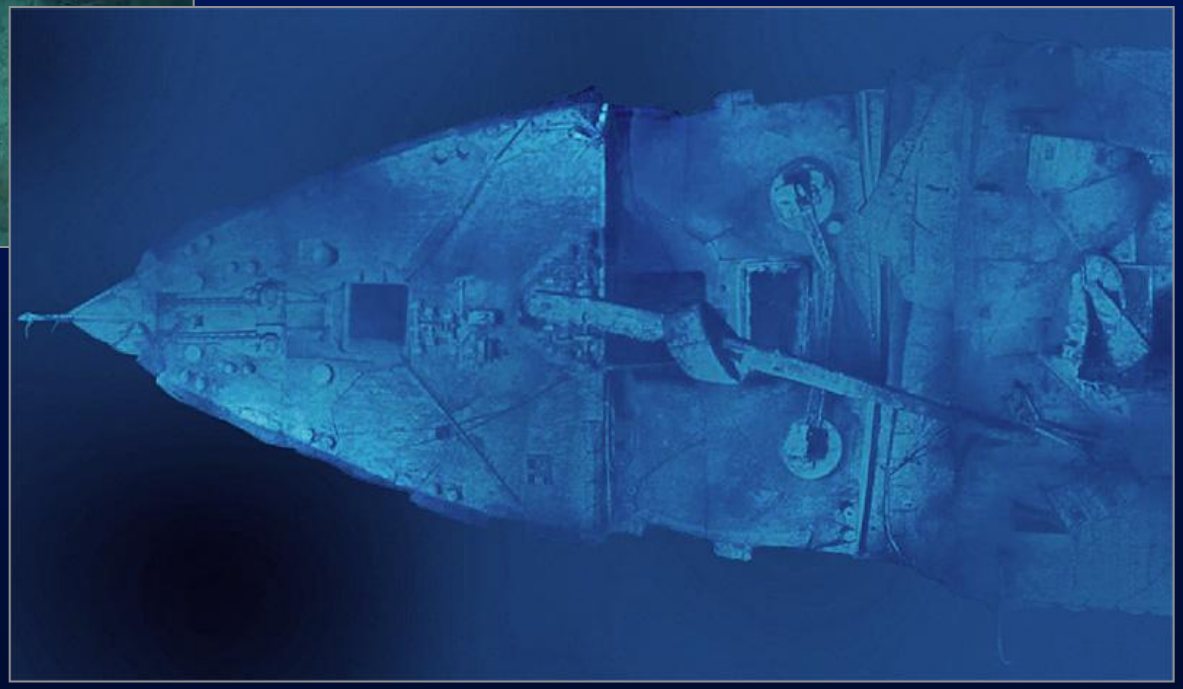
scanned tile

Application to underwater exploration



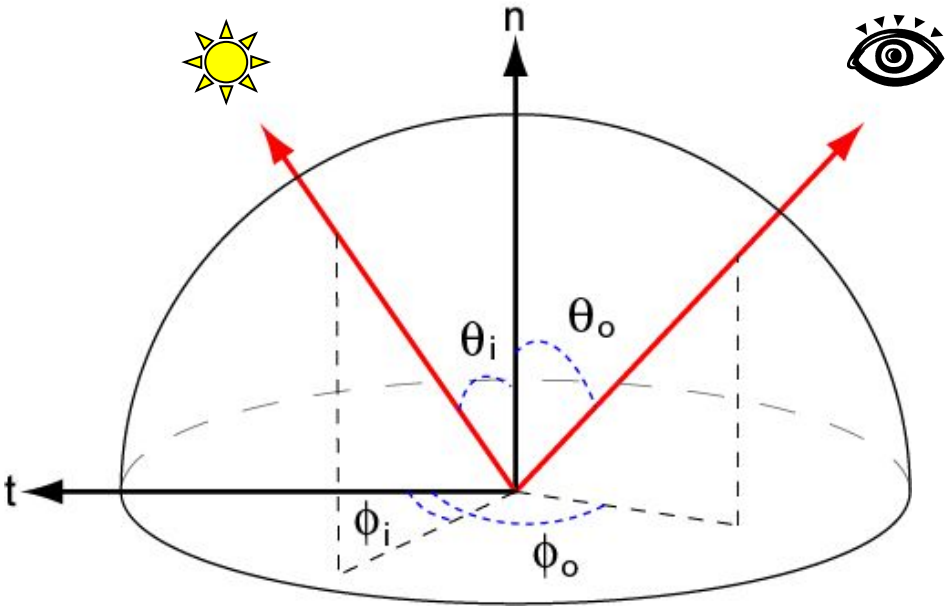
[Ballard/IFE 2004]

[Ballard/IFE 2004]

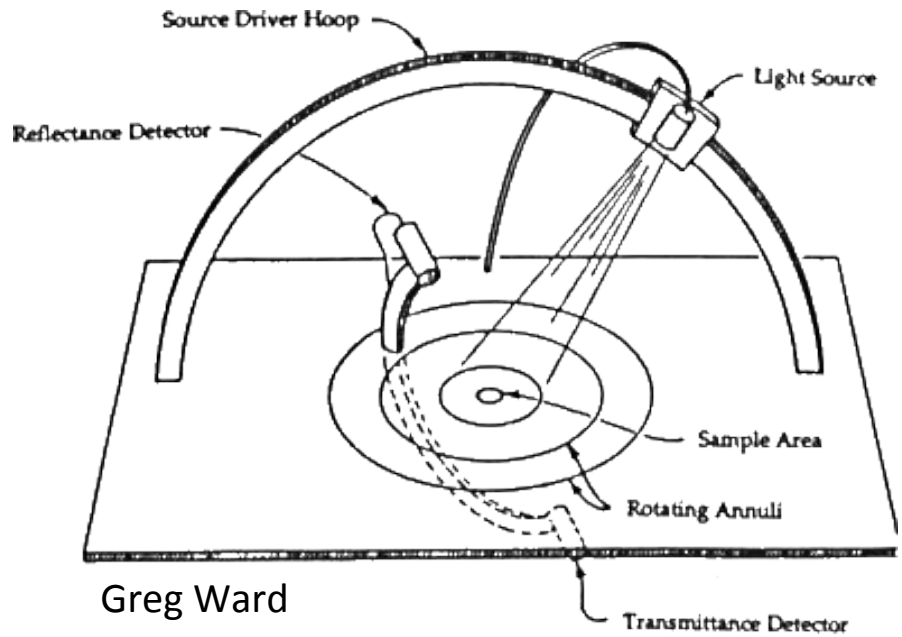


BRDF: Bidirectional Reflectance Distribution Function

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



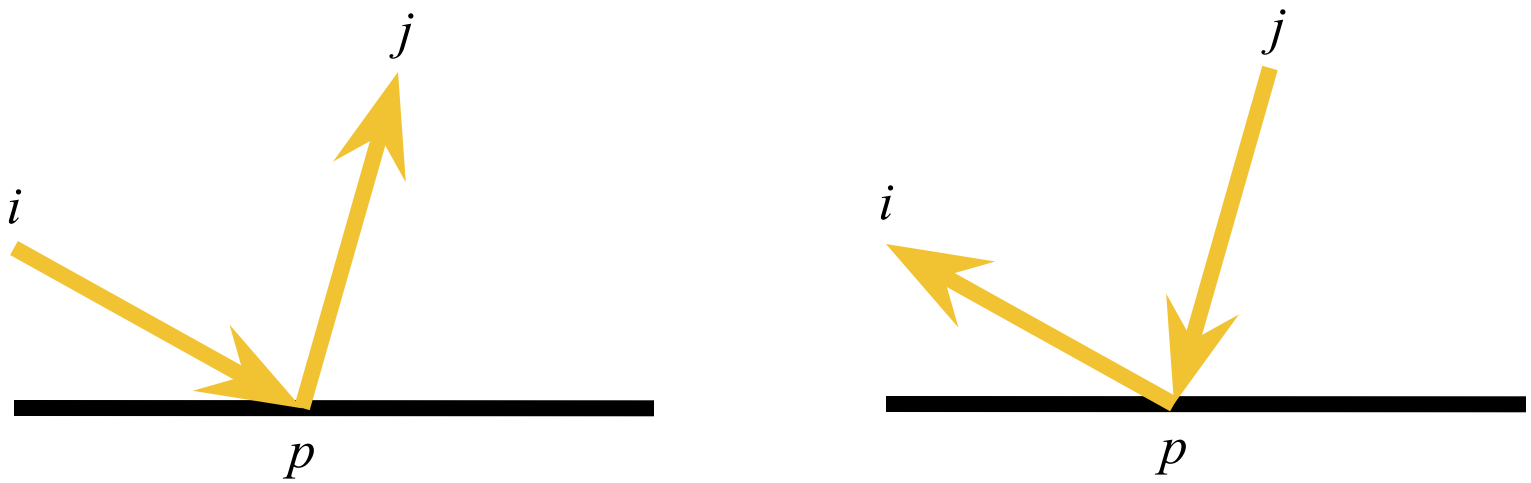
Complete material data capture: Gonioreflectometer



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

projector

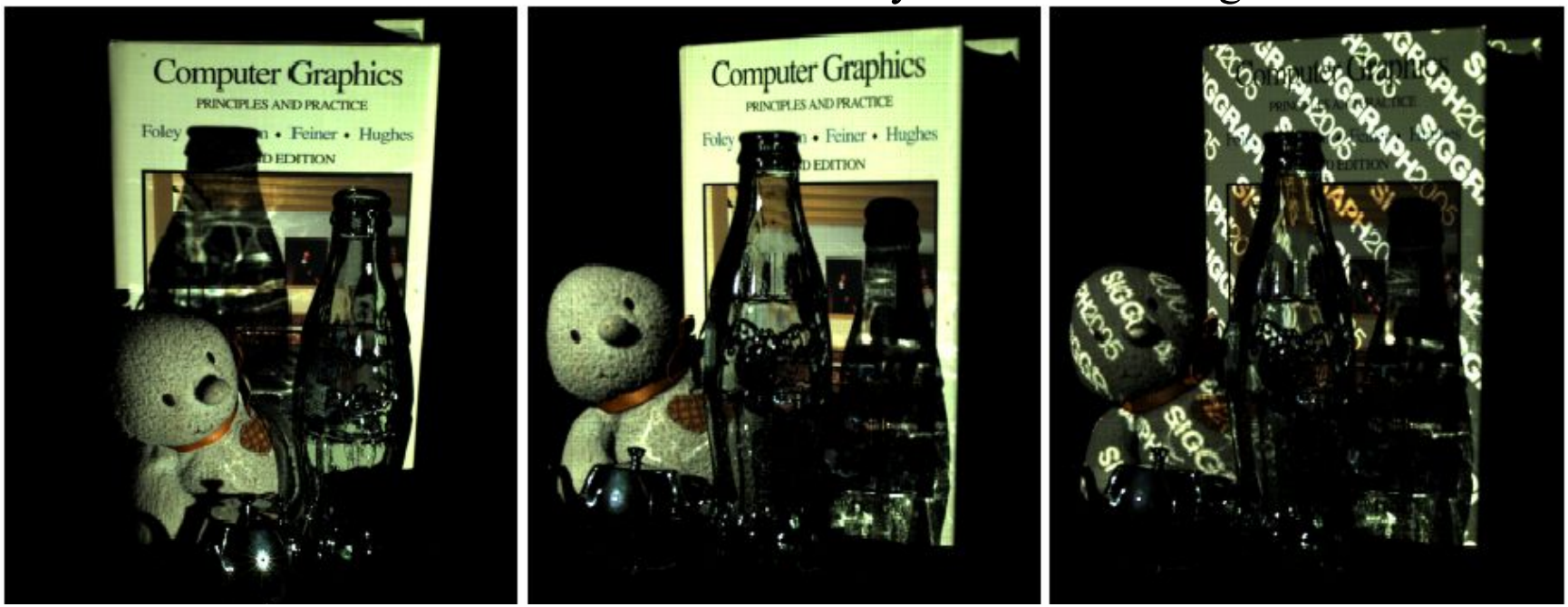
camera

camera

projector

real scene

synthetic renderings



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

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

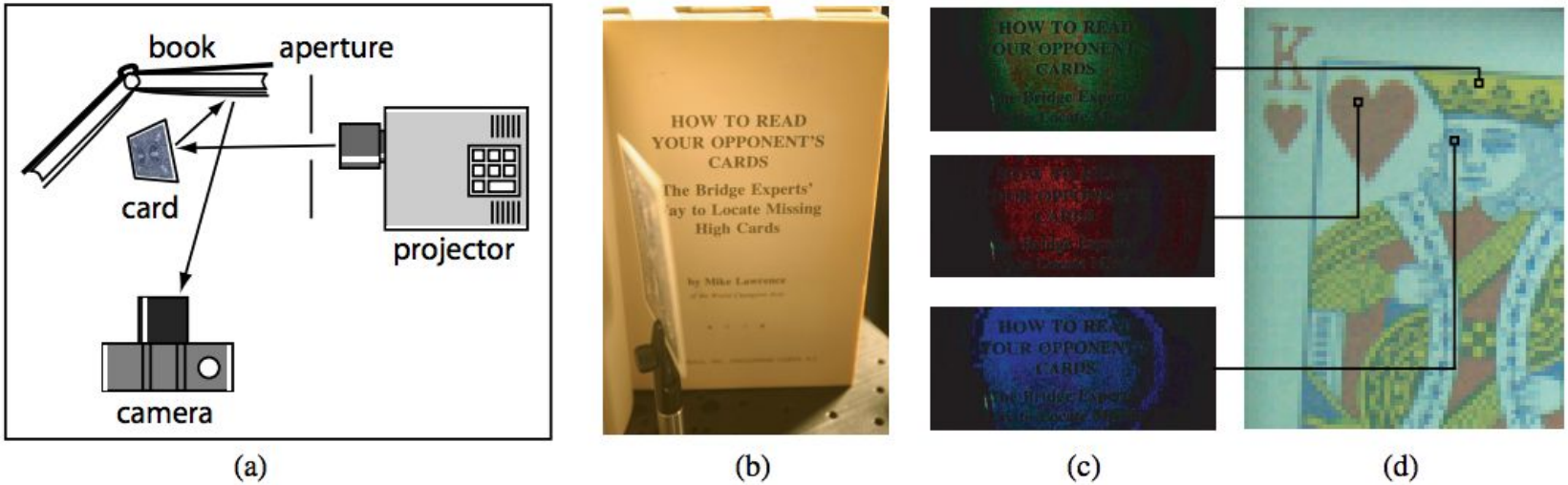


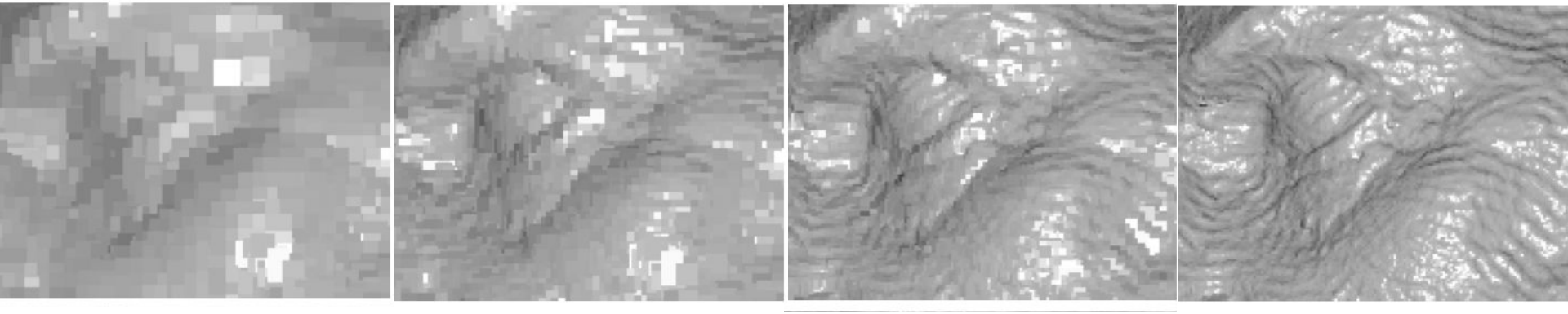
Figure 16: Dual photography with indirect light transport. (a) A projector illuminates the front of a playing card while the camera sees only the back of the card and the diffuse page of the book. An aperture in front of the projector limits the illumination only onto the card. The card was adjusted 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

- Discuss Worksheet from Last Week
- Informed Consent & Data Anonymization
- Tangentially Related Graphics/Vision Topics
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
- Readings for Tuesday
- How to brainstorm/foster radical ideas?

Readings 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

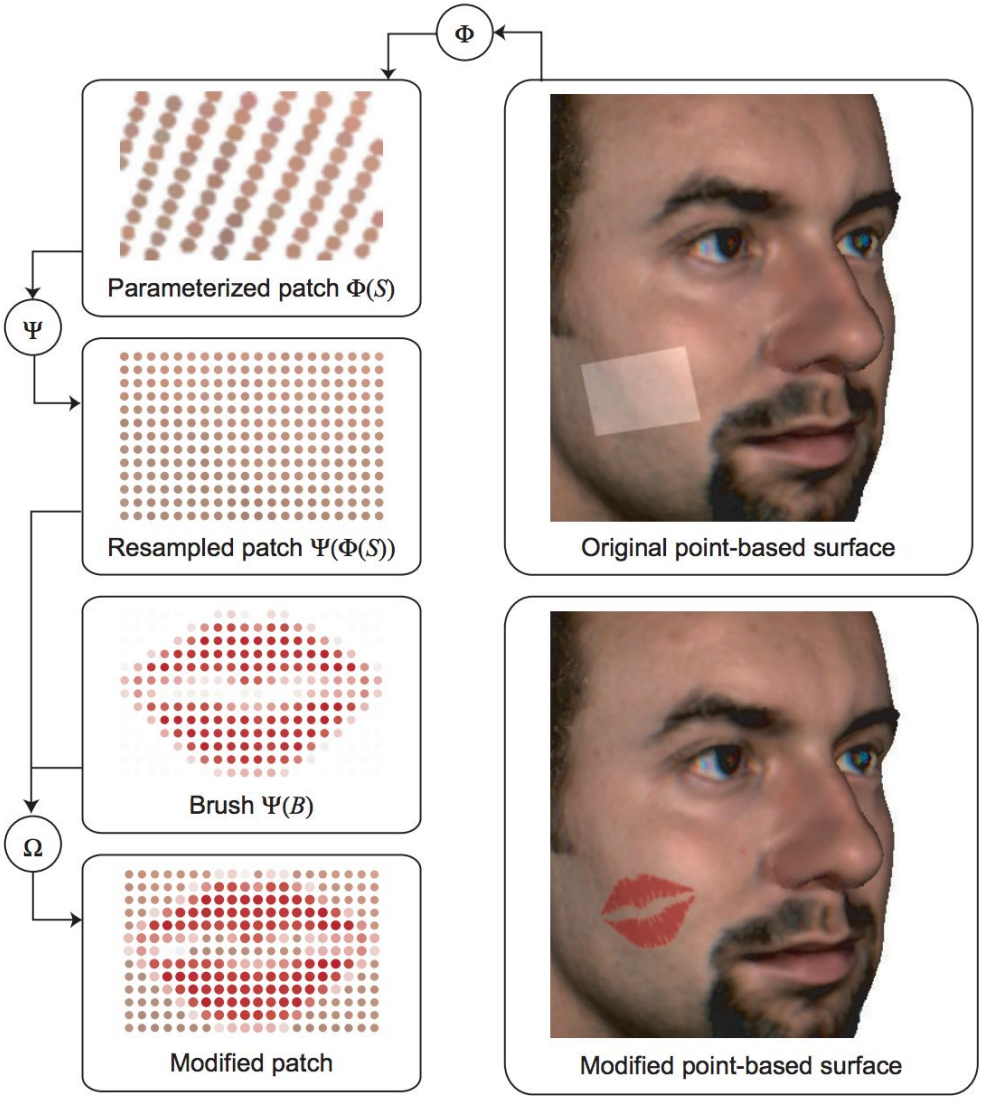


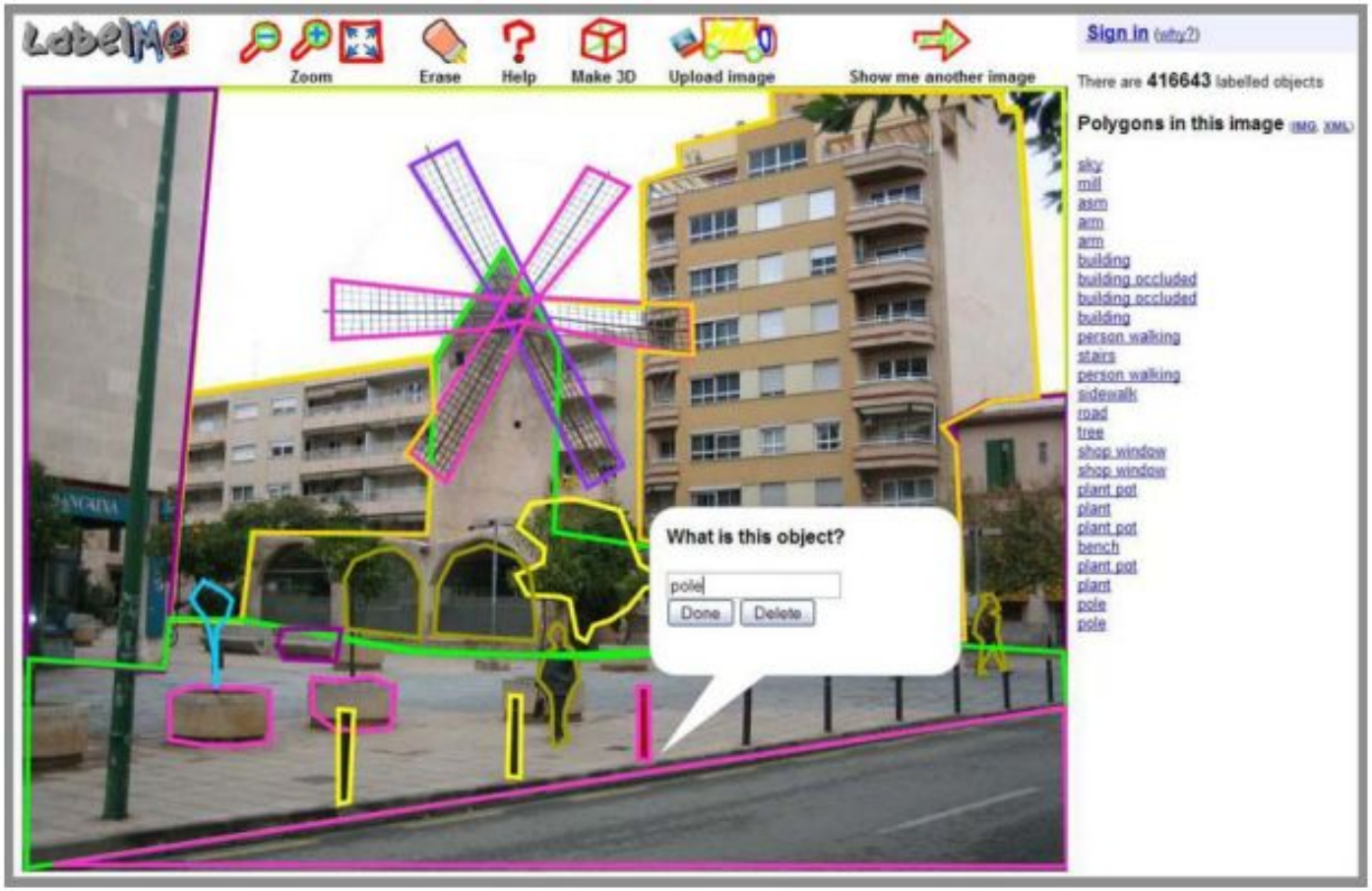
Figure 2: Overview of the operator framework for point-based surface editing.

Today

- Discuss Worksheet from Last Week
- Informed Consent & Data Anonymization
- Tangentially Related Graphics/Vision Topics
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
- Readings for Tuesday
- How to brainstorm/foster radical ideas?

Readings for This Week:

- "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 Quick Approximate Outlining

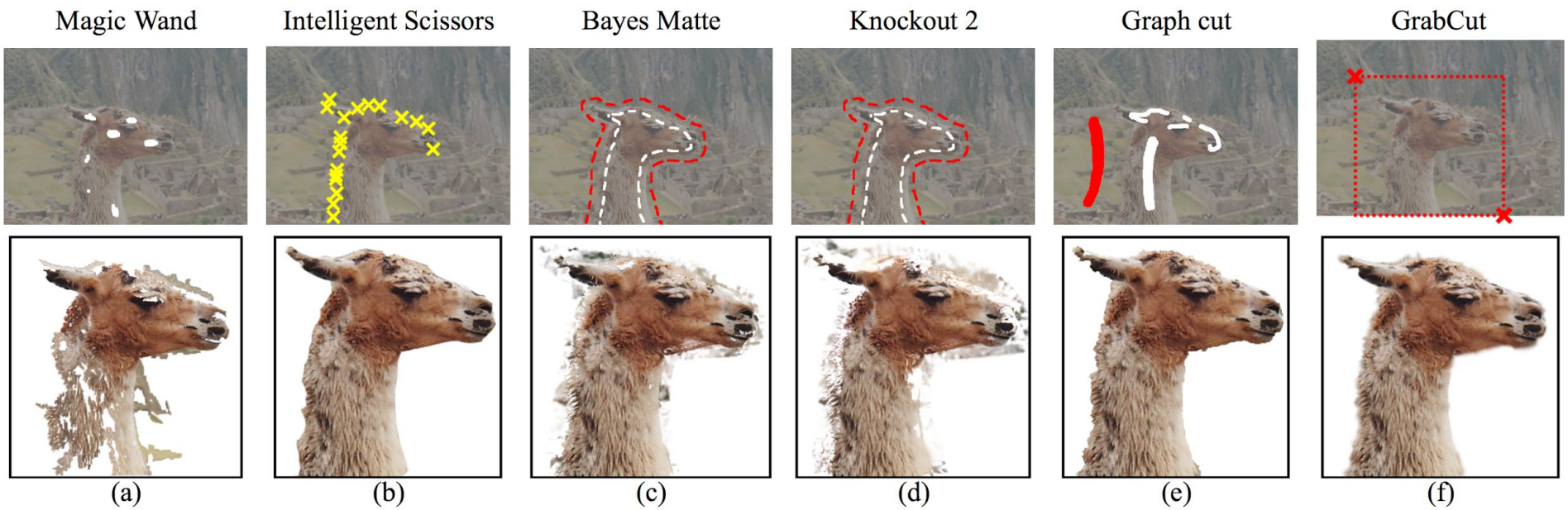
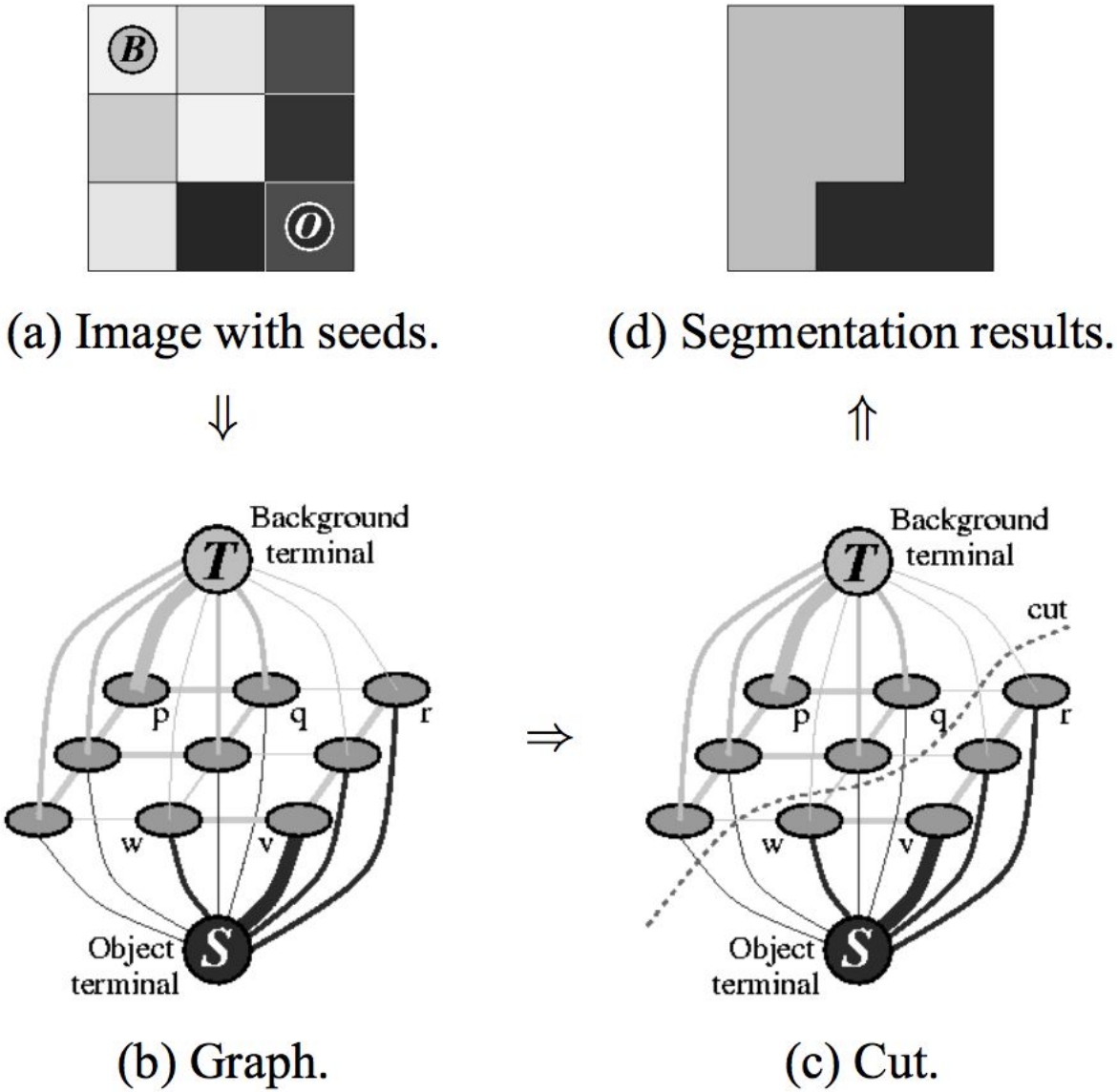


Figure 2: **Comparison of some matting and segmentation tools.** The top row shows the user interaction required to complete the segmentation or matting process: white brush/lasso (foreground), red brush/lasso (background), yellow crosses (boundary). The bottom row illustrates the resulting segmentation. 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

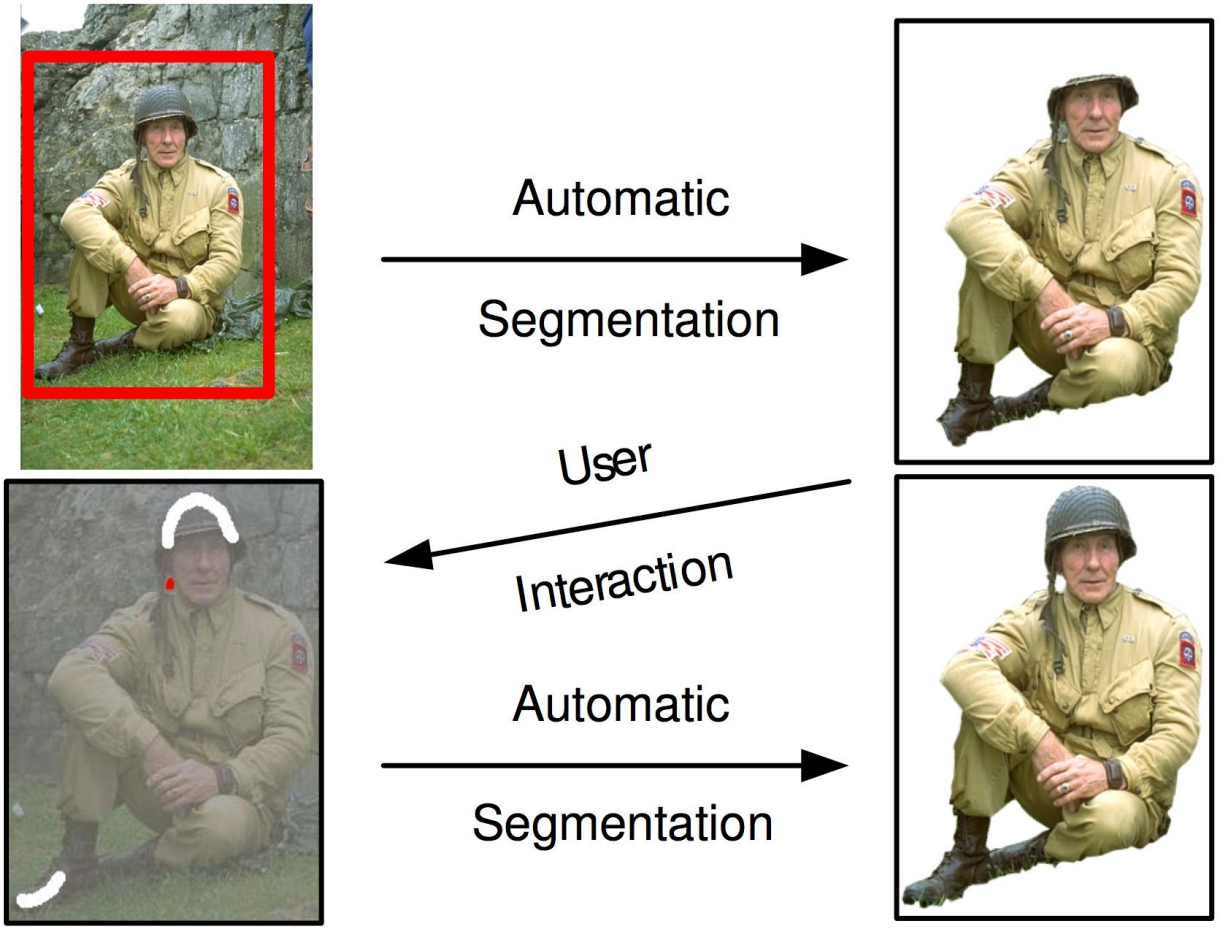


Figure 5: **User editing.** After the initial user interaction and segmentation (top row), further user edits (fig. 3) are necessary. Marking roughly with a foreground brush (white) and a background brush (red) is sufficient to obtain the desired result (bottom row).

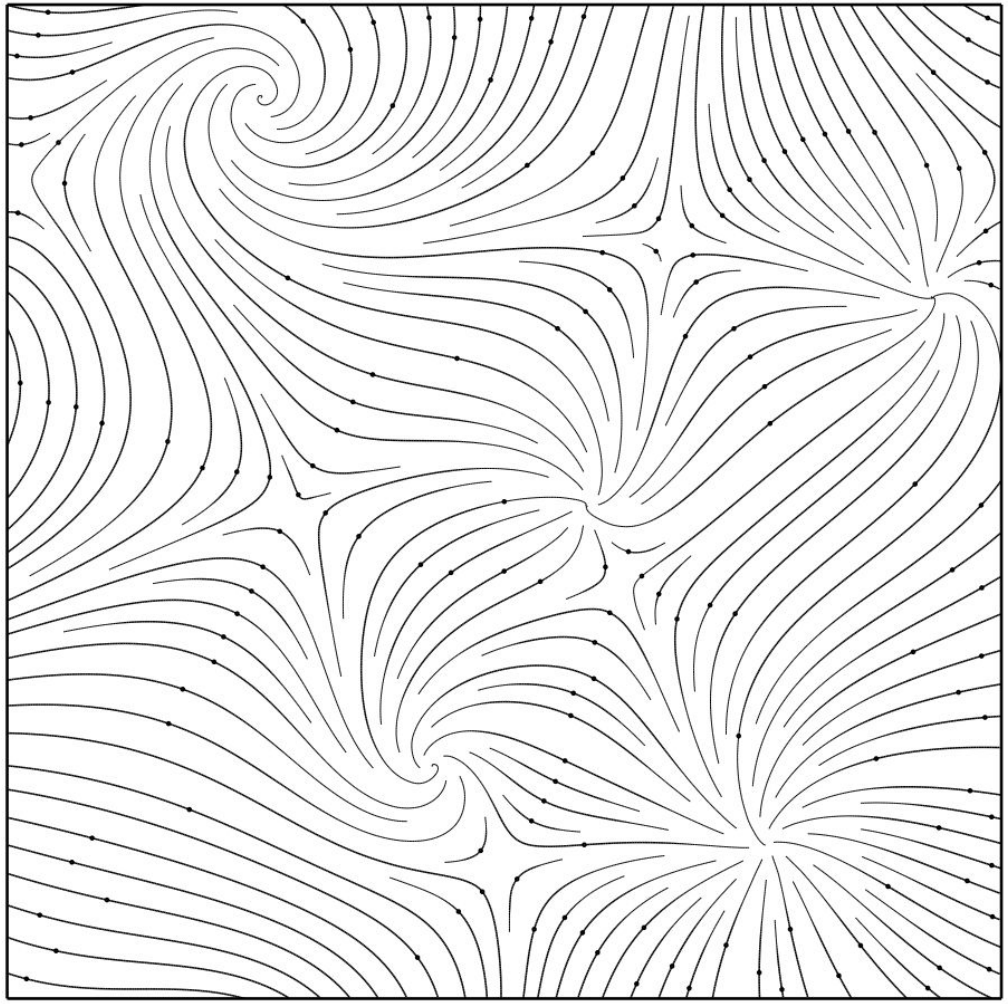
"GrabCut -Interactive Foreground Extraction using Iterated Graph Cuts", Rother, Kolmogorov, Blake, SIGGRAPH 2004

Today

- Discuss Worksheet from Last Week
- Informed Consent & Data Anonymization
- Tangentially Related Graphics/Vision Topics
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
- Readings for Tuesday
- How to brainstorm/foster radical ideas?

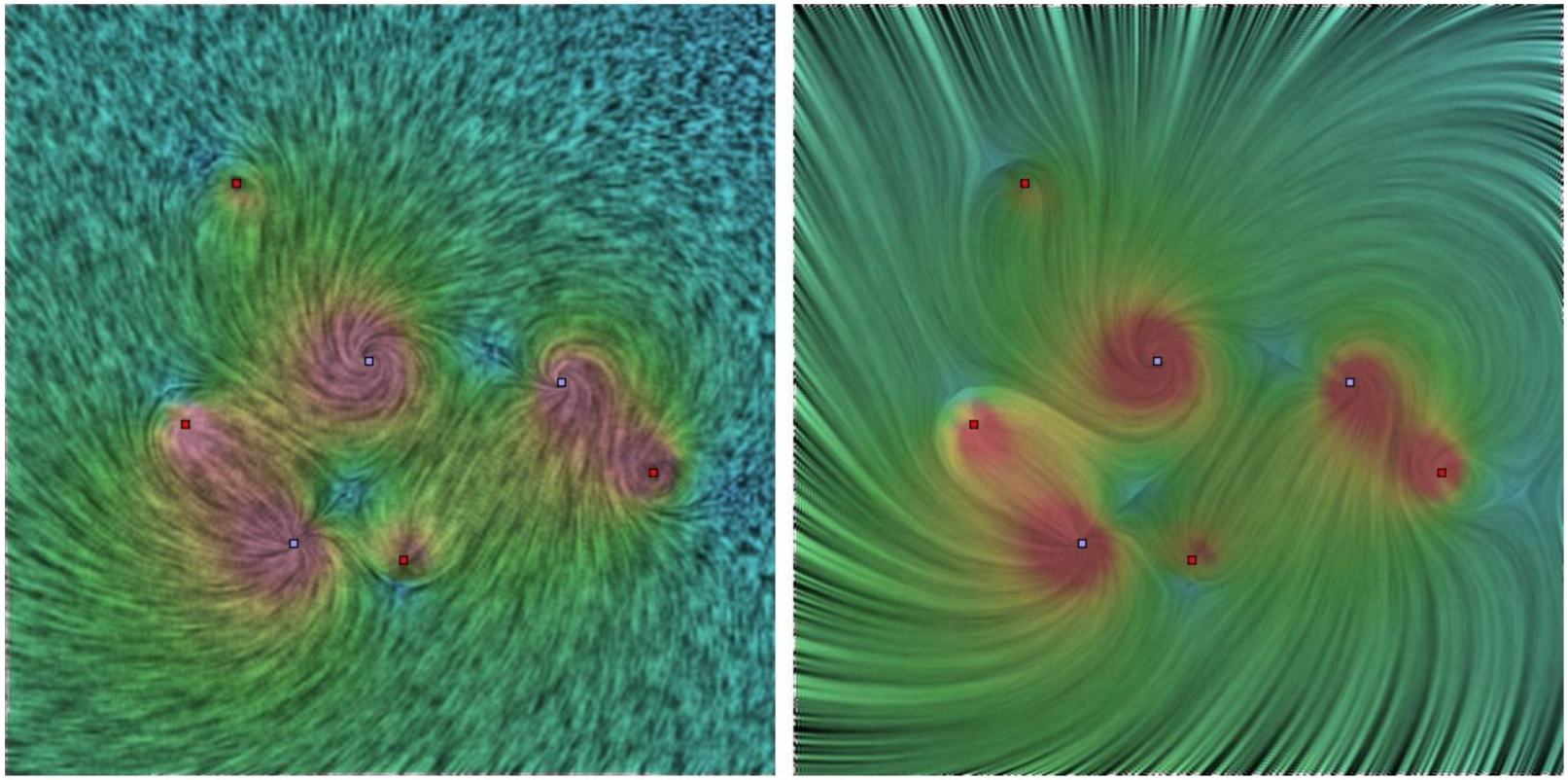
Readings for Friday: *(choose one)*

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



Readings for Friday: *(choose one)*

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



Today

- Discuss Worksheet from Last Week
- Informed Consent & Data Anonymization
- Tangentially Related Graphics/Vision Topics
- Readings for Today:
  - "QSplat: A Multiresolution Point Rendering System for Large Meshes"
  - "LabelMe: online image annotation and applications"
- Readings for Tuesday
- How to brainstorm/foster radical ideas?

Homework 7: Final Project Ideas

- Invent 2 different Final Project Ideas
  - “Who" (audience), "why" (research question), "what" (the finished visualization)
  - One technical challenge for the project. 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 Saturday 3/7 @ midnight
- Read & *reply* to 3 other students by Monday 3/16 (after break)