Ethics & Privacy

Plan for Today

- “Pop Worksheet”
- Readings for this Week
  - "Adaptive Privacy-Preserving Visualization Using Parallel Coordinates”
  - "Agile Ethics for Massified Research and Visualization”
- Informed Consent & Data Anonymization
- Readings for Tuesday after break
- Tangentially Related Graphics/Vision Topics

"Adaptive Privacy-Preserving Visualization Using Parallel Coordinates", Dasgupta & Kosara, TVCG 2011.

- Different purposes for "Visualization":
  - Convey information or art? Is misleading necessarily bad when visualizing for art purposes? Context matters.
  - Find the appropriate balance between art (eye catching & memorable) and scientific accurate & honest data presentation.
- Some visualization techniques are already quite lossy (e.g., pie chart) or just confusing (e.g., parallel coordinates?) and obfuscation may not be necessary
  - Making false connections
  - Limit visualization to 500 pixels to thwart attacks(!?)
  - Rounding pixel coordinates
- Impractical in the real world?
- So much effort to make a visualization *less useful.*
  (Why make the visualization at all? Who is the target audience of this visualization?)
- Need to investigate reliability of data before reporting it
- K-anonymity & l-diversity
- (Trusted) server does clustering & only sends clusters to (untrusted) client
- Writing
  - Nice to read about the process and ideas for techniques that ultimately didn't work
  - Would have been better written/easier to read if they had stated upfront what they wanted to accomplish
  - Diagrams & charts not well explained

- Could be further secured:
  - Don't place unclustered data on the public facing machine
  - Don't ever re-cluster the data (prevent clustering attacks)
- Client-server model: Computation & networking costs?
- How well does this work for small datasets?
- Hadn't considered visualization as a means of data breach, reassuring that people are researching the problem.
- How much accuracy in analysis do we lose?
  - Obsuring & blurring data (opposite of our usual focus on clarity & accuracy!)
- What about more sophisticated attacks? …?
  - Maybe this isn't 100% secure, but it's important to do something!
- Would like a user study comparing their Privacy-Preserving Parallel Coordinates to original full data Parallel Coordinates.
- Quasi-Identifiers (is an anonymous form really anonymous?)
- Privacy policies that declare how they can and will share your personal data…

"Agile Ethics for Massified Research and Visualization", Neuhaus & Webmoor, Information, Communication & Society 2012.

- Privacy, confidentiality, anonymity, and informed consent
  - Minimize risk to participants
  - Observational study does not require consent
  - (current) IRB process does not/cannot work at massive scale
  - A single person can more easily do what used to take a team of researchers much more time.
- Agile ethics: too high level to be enforceable?
- Even though individuals made this data available, it is researchers responsibility to not put them in danger
- Not showing individuals, only general trends
- If the research team's information is similarly public/vulnerable, its ok?
- Geometry for twitter
- Public-private greyscale
- "Now that you've read this… Just be more careful, ok?"
- Writing
  - Confusing paper organization
  - Low resolution images
  - Unconventional acronyms
  - Footnotes at end of paper (prefer at end of each page)

- Jargon-y
- Unnecessarily lengthy?
- Websurfing is dangerous
- Twitter is scary (amount of personal data available surprising)
  - Users can (now?) disable location tracking
  - Are Twitter "protected" accounts new?
  - Read the fine print before sharing your data!
- TimeRose visualization is new to me
- Scattered topics
- Fitness tracking applications offer to post your morning run on facebook are dangerous
- Proposed solutions are too idealistic?

Interesting Tidbits

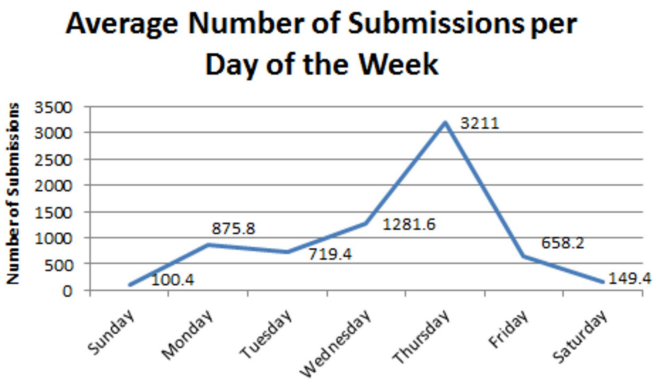
- Internet is an ocean of data
- Research results poured back into ocean of data
- Surveillance: Shopping malls are private spaces, but made to feel like public spaces
- File/log planned data collections in advance (pre-planning required, data more precious)
- Researchers should make themselves equally public

Plan for Today

- “Pop Worksheet”
- Readings for this Week
  - "Adaptive Privacy-Preserving Visualization Using Parallel Coordinates”
  - "Agile Ethics for Massified Research and Visualization”
- Informed Consent & Data Anonymization
- Readings for Tuesday after break
- Tangentially Related Graphics/Vision Topics

Informed Consent

- 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: a single person can do it alone - lost informal peer consultation of ethics concerns



- Research Questions of Submitty Data:
  - Does an individual student's grade rise over time (repeated submissions)?
  - Do students who start submitting earlier in the week have a higher final grade for that homework?
  - Do students who submit more often get higher grades?

What is sufficient to anonymize Submitty data? What are the sensitive attributes?

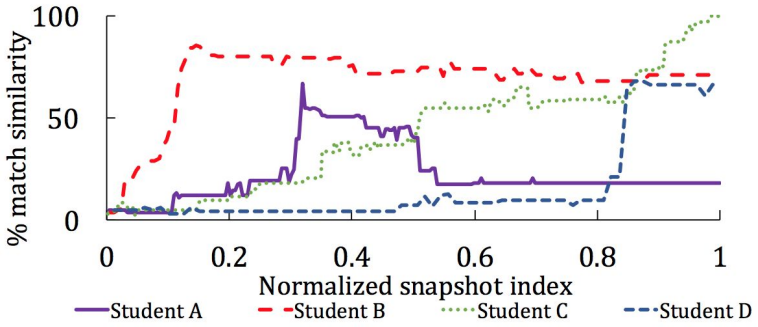
- grades, plagiarism, sleep patterns, who took the class, time/# of attempts needed to complete assignment
- Remove explicit identifiers (username, RIN)
- Small datasets cause problems
- Quasi identifiers
- Sanitize data on the fly, constraining the interaction (don't allow inspection of complete historical details of single student -- even if the name has been removed)
- Assume data holder is trusted and aware of data sensitivity *(and appropriately concerned)*

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

Risks to users/participants?

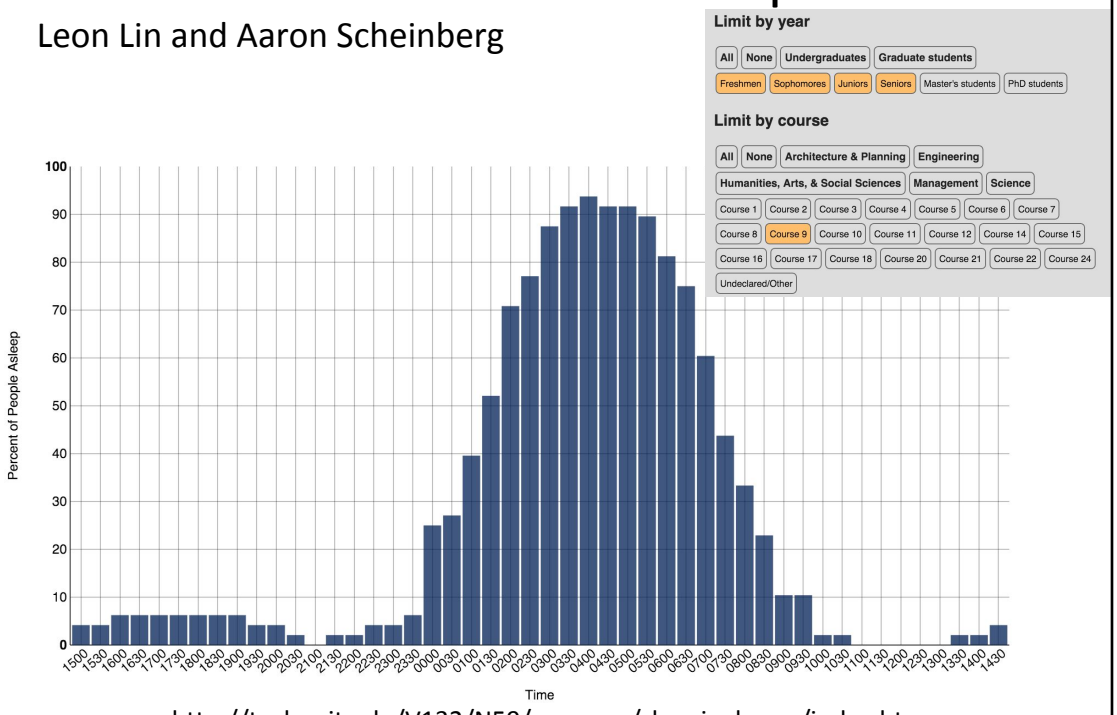
- Quasi-identifiers & Doxing/doxxing (document tracing): "Internet-based practice of researching and publishing personally identifiable information about an individual. The methods employed in pursuit of this information range from searching publicly available databases and social media websites like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

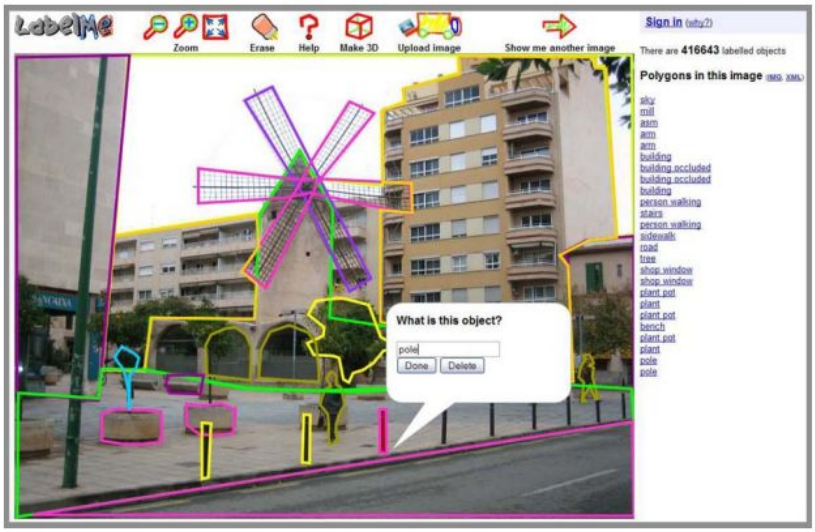
- Long Title: “An Act To amend the Internal Revenue Code of 1986 to improve portability and continuity of health insurance coverage in the group and individual markets, to combat waste, fraud, and abuse in health insurance and health care delivery, to promote the use of medical savings accounts, to improve 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

Plan for Today

- “Pop Worksheet”
- Readings for this Week
  - "Adaptive Privacy-Preserving Visualization Using Parallel Coordinates”
  - "Agile Ethics for Massified Research and Visualization”
- Informed Consent & Data Anonymization
- Readings for Tuesday after Break
- Tangentially Related Graphics/Vision Topics

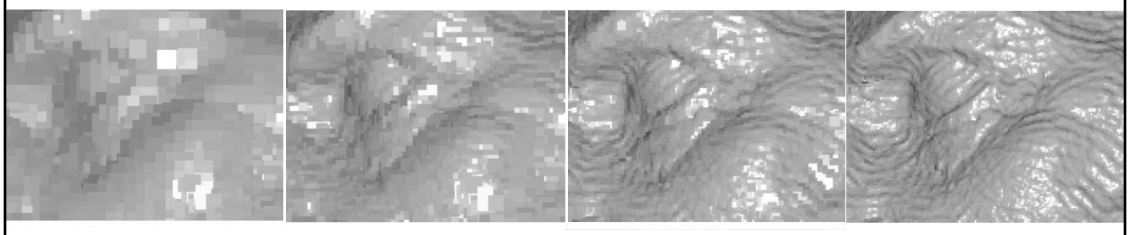
Reading for Tuesday: *(choose one)*

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



Reading for Tuesday: *(choose one)*

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



Plan for Today

- "Pop Worksheet"
- Readings for this Week
  - "Adaptive Privacy-Preserving Visualization Using Parallel Coordinates"
  - "Agile Ethics for Massified Research and Visualization"
- Informed Consent
- Data Anonymization
- Readings for Tuesday after Break
- Tangentially Related Graphics/Vision Topics

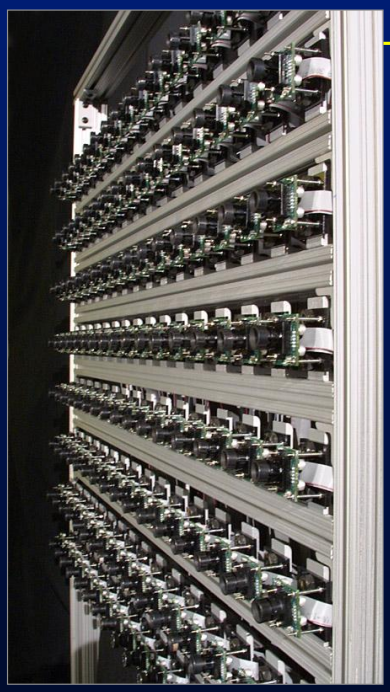
Synthetic aperture confocal imaging

Option for Reading for Tuesday?

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

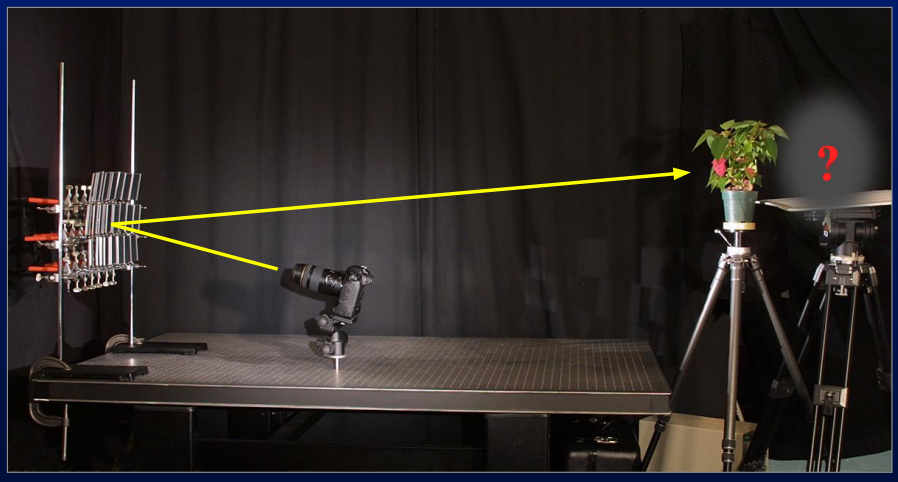
Stanford Multi-Camera Array

[Wilburn 2002]



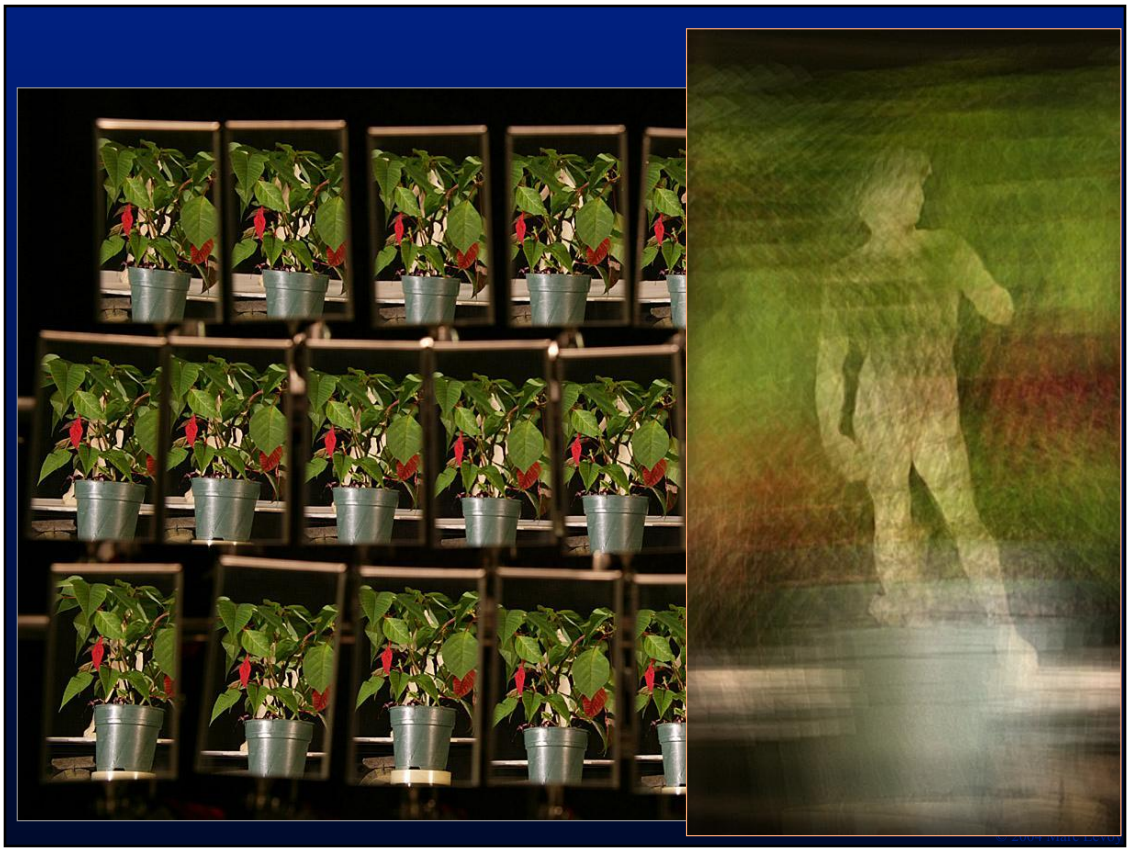
- 640 × 480 pixels × 30fps × 128 cameras
- synchronized timing
- continuous video streaming
- flexible physical arrangement

Synthetic aperture photography using an array of mirrors



- 11-megapixel camera
- 22 planar mirrors


© 2004 Marc Levoy

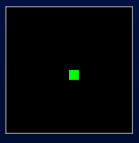


Confocal imaging in scattering media

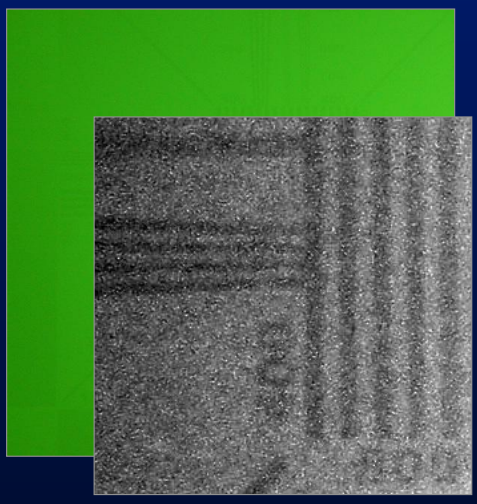
- small tank
  - too short for attenuation
  - lit by internal reflections

Experiments in a large water tank

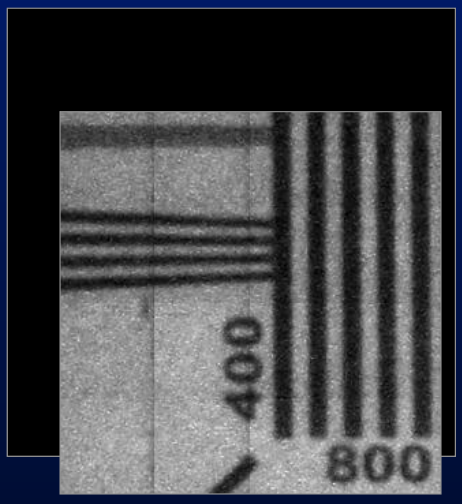
- stray light limits performance
- one projector suffices if no occluders



Seeing through turbid water



floodlit



scanned tile

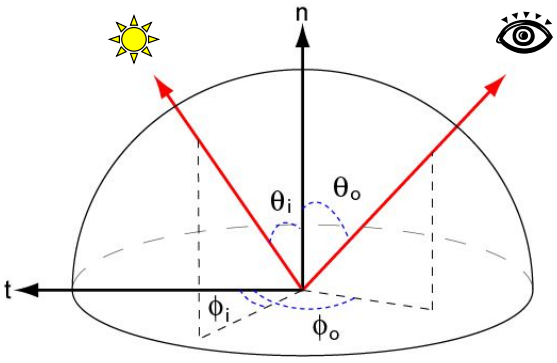
Application to underwater exploration

[Ballard/IFE 2004]

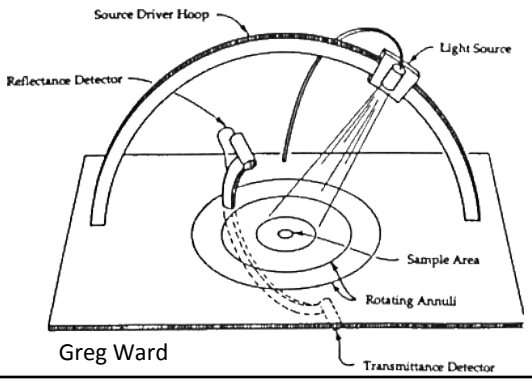
[Ballard/IFE 2004]

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

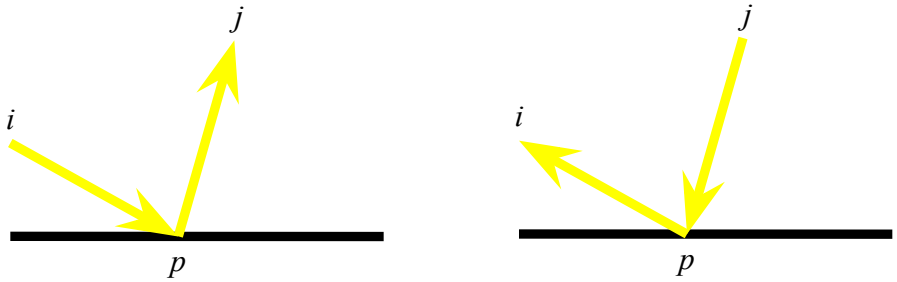


Complete material data capture: Gonioreflectometer

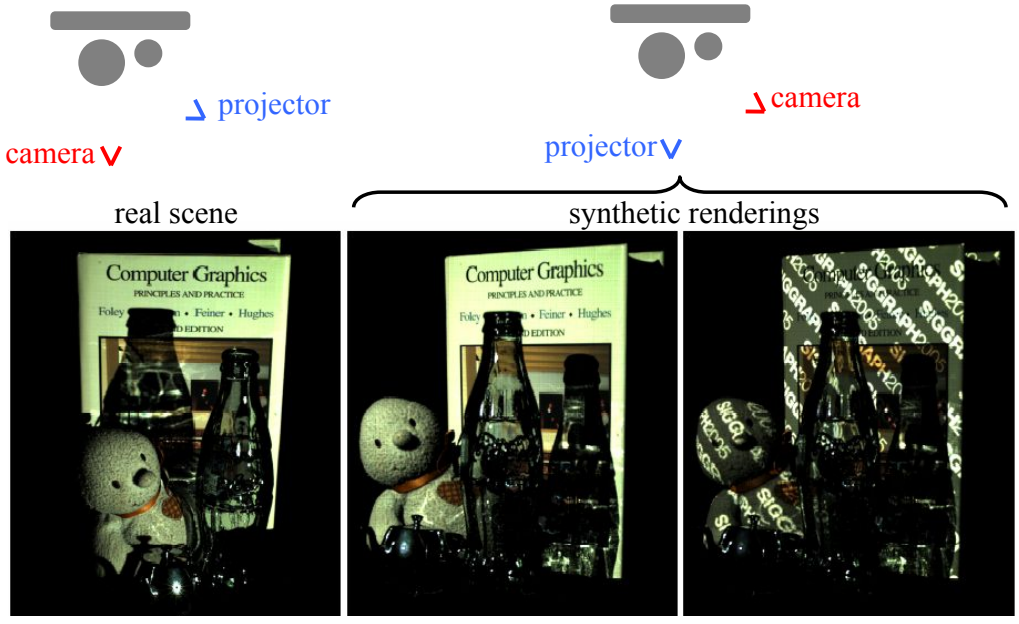


Helmholtz Reciprocity

- 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*

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

Option for Reading for Tuesday?

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.