Lecture 22:
Structure from Motion (cont.)
+
Photometric Stereo

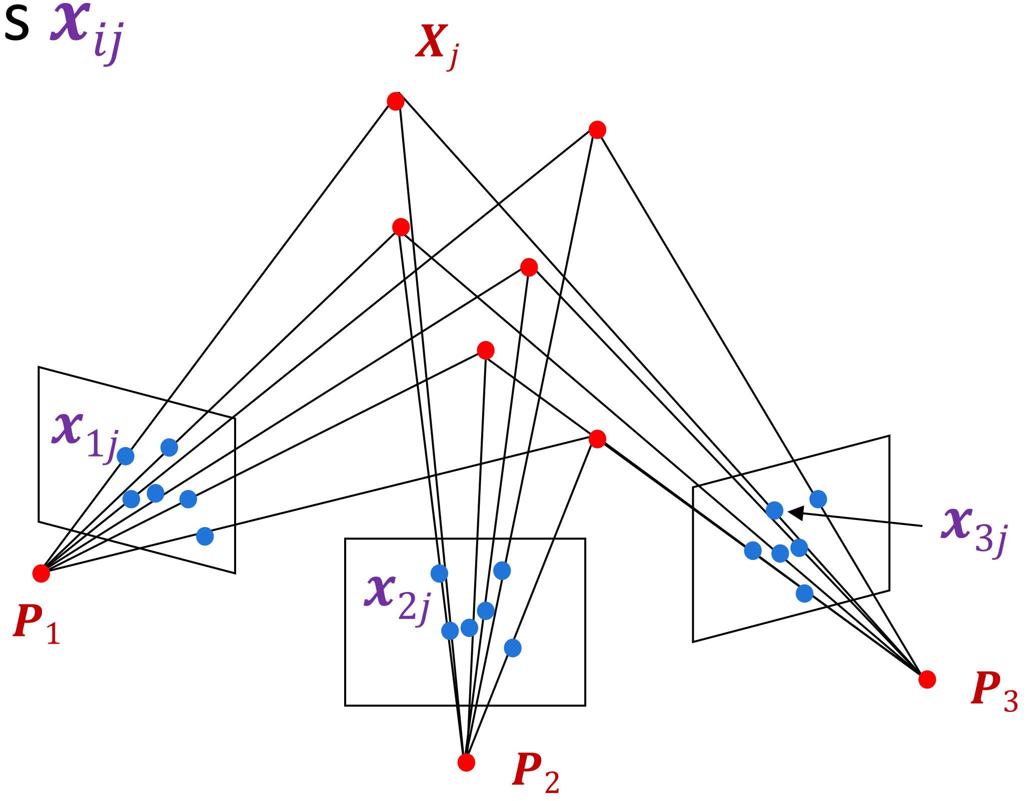
COMP 590/776: Computer Vision

Instructor: Soumyadip (Roni) Sengupta

Structure from Motion (cont.)

Projective structure from motion

- **Given**: m images of n fixed 3D points such that (ignoring visibility):
  - $\boldsymbol{x}_{ij} \cong \boldsymbol{P}_i \boldsymbol{X}_j, \qquad i = 1, \dots, m, \quad j = 1, \dots, n$
- **Problem**: estimate m projection matrices $\boldsymbol{P}_i$ and n 3D points $\boldsymbol{X}_j$ from the mn correspondences $\boldsymbol{x}_{ij}$

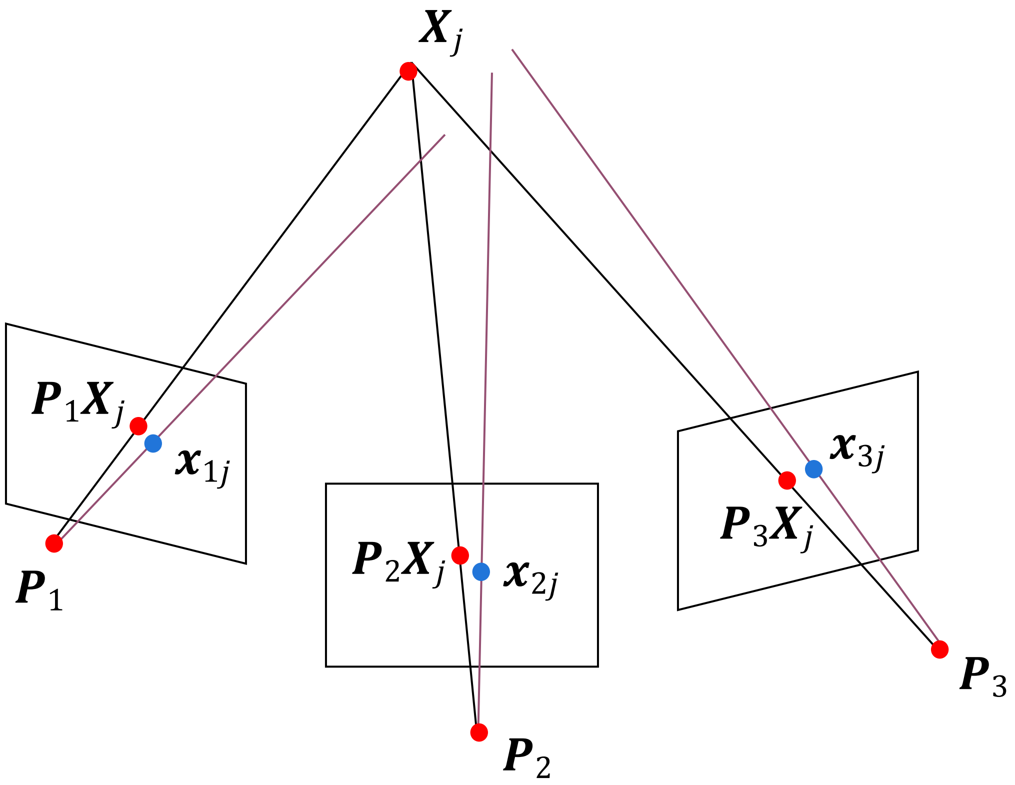


Bundle adjustment

- Non-linear method for refining structure and motion
- Minimize reprojection error (with lots of bells and whistles):

- $\sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} d\left(\boldsymbol{x}_{ij} - \text{proj}\left(\boldsymbol{P}_i \boldsymbol{X}_j\right)\right)^2$

visibility flag:
is point j visible in view i?



Bundle Adjustment is highly non-convex and requires good initialization – hence we require algebraic techniques to solve this (Factorization)

B. Triggs et al. [Bundle adjustment – A modern synthesis](). International Workshop on Vision Algorithms, 1999

Projective structure from motion

- **Given**: m images of n fixed 3D points such that (ignoring visibility):
  - $\boldsymbol{x}_{ij} \cong \boldsymbol{P}_i \boldsymbol{X}_j, \qquad i = 1, \ldots, m, \quad j = 1, \ldots, n$
- **Problem**: estimate m projection matrices $\boldsymbol{P}_i$ and n 3D points $\boldsymbol{X}_j$ from the mn correspondences $\boldsymbol{x}_{ij}$
- With no calibration info, cameras and points can only be recovered up to a 4×4 projective transformation $\boldsymbol{Q}$:
  - $\boldsymbol{X} \rightarrow \boldsymbol{QX}, \boldsymbol{P} \rightarrow \boldsymbol{PQ}^{-1}$
- We can solve for structure and motion when $2mn \geq 11m + 3n - 15$
- For two cameras, at least 7 points are needed
- Why is this hard to solve?
  - Factorization is hard as perspective projection is only upto a scale and we also need to search for a scale.

Affine structure from motion

- **Given**: m images of n fixed 3D points such that
  - $\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{X}_j + \boldsymbol{t}_i, \quad i = 1, \dots, m, \ j = 1, \dots, n$
  - $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \boldsymbol{AX} + \boldsymbol{t}$

Not in homogenous coordinate.

- **Problem**: use the mn correspondences $\boldsymbol{x}_{ij}$ to estimate m projection matrices $\boldsymbol{A}_i$ and translation vectors $\boldsymbol{t}_i$, and n points $\boldsymbol{X}_j$
- The reconstruction is defined up to an arbitrary *affine* transformation $\boldsymbol{Q}$ (12 degrees of freedom):

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{t} \\ 0^T & 1 \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{A} & \boldsymbol{t} \\ 0^T & 1 \end{bmatrix} \boldsymbol{Q}^{-1}, \qquad \begin{pmatrix} \boldsymbol{X}_j \\ 1 \end{pmatrix} \rightarrow \boldsymbol{Q} \begin{pmatrix} \boldsymbol{X}_j \\ 1 \end{pmatrix}$$

- How many knowns and unknowns for m images and n points?
  - $2mn$ knowns and $8m + 3n$ unknowns
  - To be able to solve this problem, we must have $2mn \geq 8m + 3n - 12$ (affine ambiguity takes away 12 dof)
  - E.g., for two views, we need four point correspondences

Affine structure from motion

$$\hat{\boldsymbol{x}}_{ij} = \boldsymbol{x}_{ij} - \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{x}_{ik} \qquad \hat{\boldsymbol{X}}_j = \boldsymbol{X}_j - \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{X}_k$$

Normalize 2D and 3D points

- Let's create a $2m \times n$ data (measurement) matrix:

- $$\boldsymbol{D} = \begin{bmatrix} \hat{\boldsymbol{x}}_{11} & \hat{\boldsymbol{x}}_{12} & \cdots & \hat{\boldsymbol{x}}_{1n} \\ \hat{\boldsymbol{x}}_{21} & \hat{\boldsymbol{x}}_{22} & \cdots & \hat{\boldsymbol{x}}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{x}}_{m1} & \hat{\boldsymbol{x}}_{m2} & \cdots & \hat{\boldsymbol{x}}_{mn} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \vdots \\ \boldsymbol{A}_m \end{bmatrix}}_{\boldsymbol{M}} \underbrace{\begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_n \end{bmatrix}}_{\boldsymbol{S}}$$

cameras ($2m \times 3$)

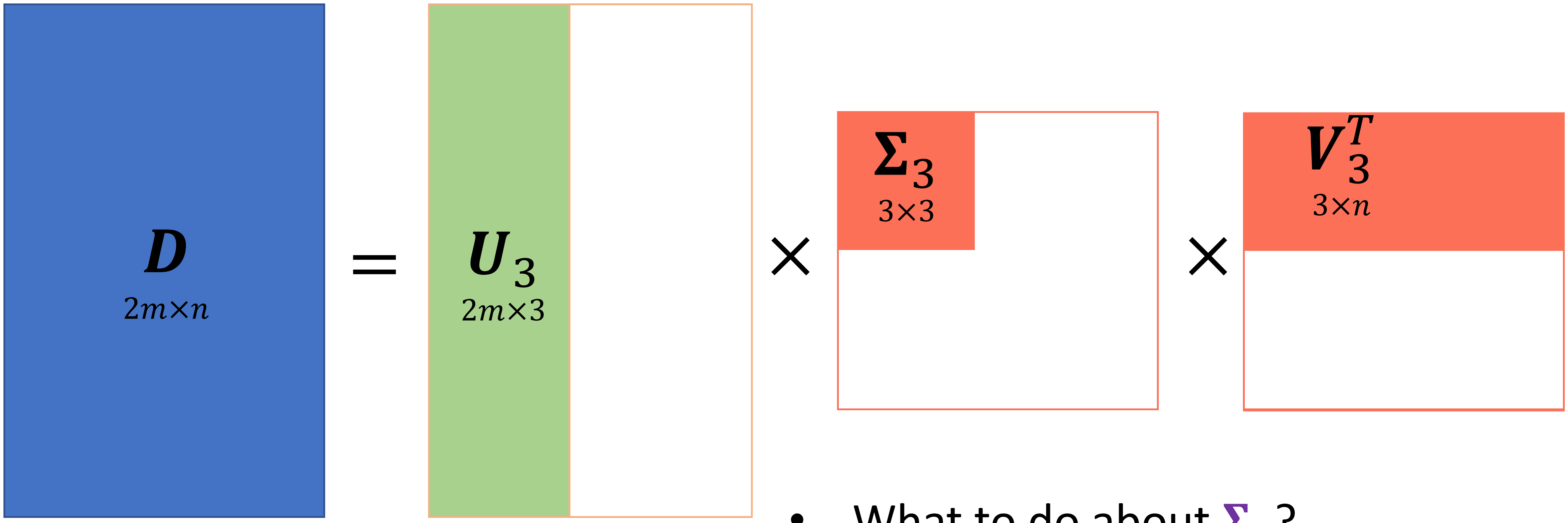
points ($3 \times n$)

- D is at most rank 3.

C. Tomasi and T. Kanade. [Shape and motion from image streams under orthography: A factorization method](). *IJCV*, 9(2):137-154, November 1992.

Factorizing the measurement matrix

- Keep top 3 singular values:
- This is the closest approximation of $\boldsymbol{D}$ with a rank-3 matrix in terms of Frobenius norm



- What to do about $\boldsymbol{\Sigma}_3$?
- One solution: $\boldsymbol{M} = \boldsymbol{U}_3\boldsymbol{\Sigma}_3^{\frac{1}{2}}, \boldsymbol{S} = \boldsymbol{\Sigma}_3^{\frac{1}{2}}\boldsymbol{V}_3^T$

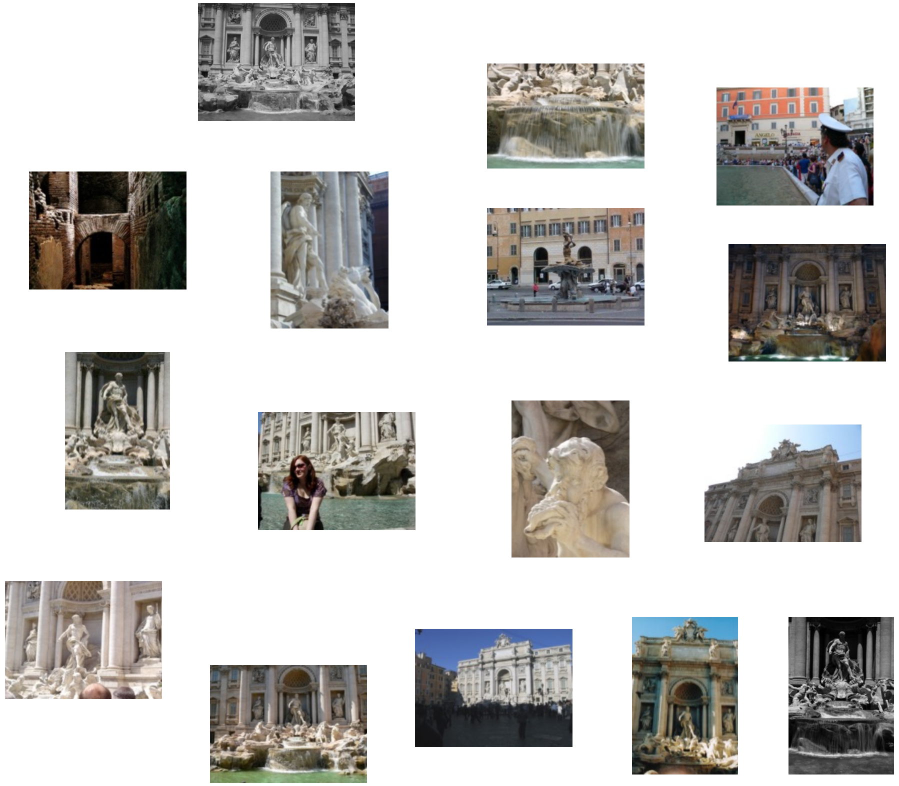
Further Reading: [Factorization methods for structure from motion]

Global Structure from Motion

SfM for large scale scenes

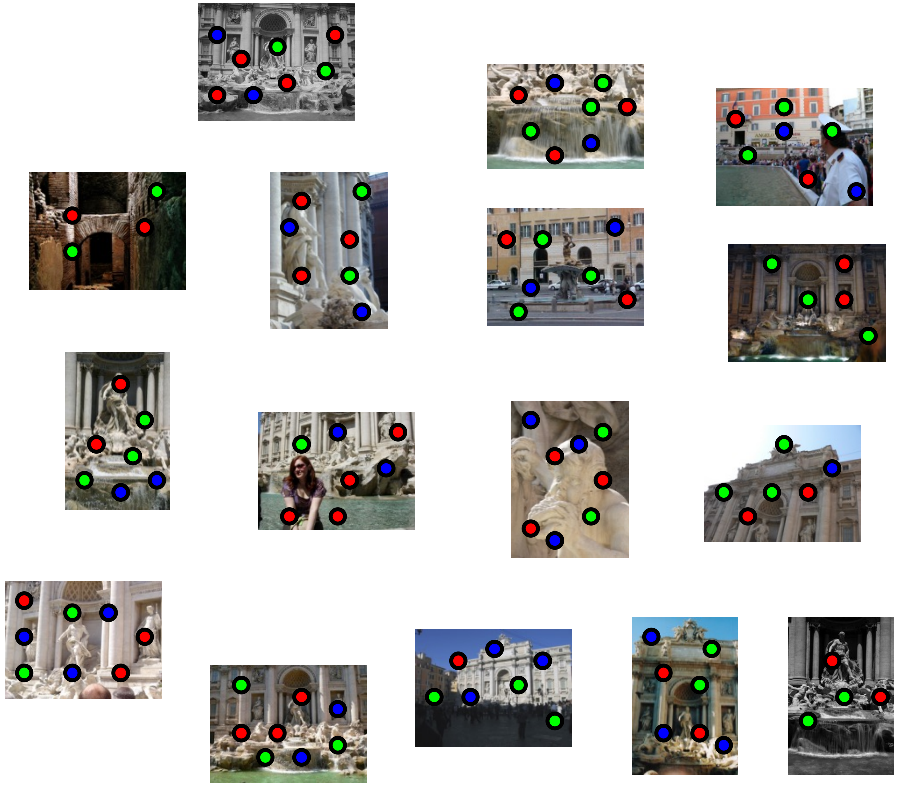
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



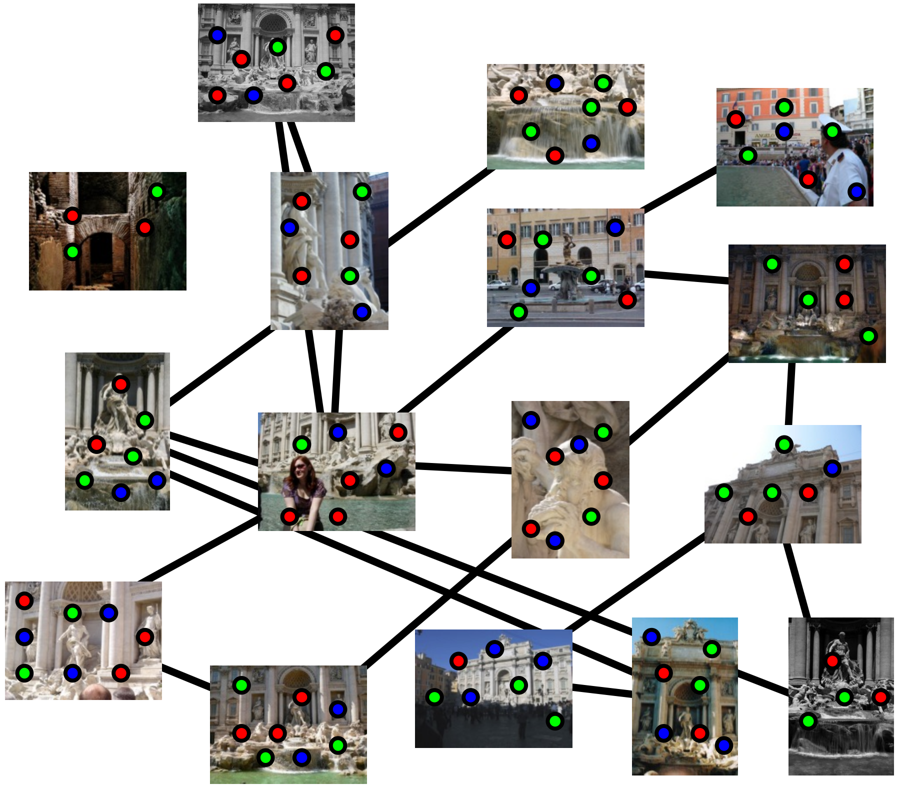
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



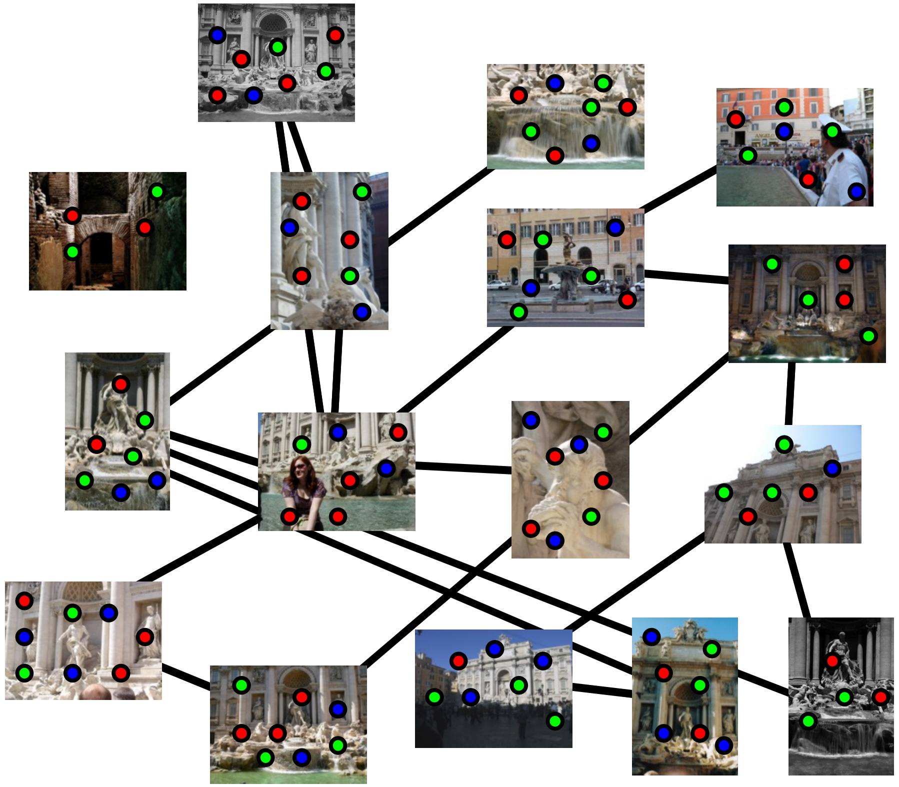
Feature matching

Match features between each pair of images



Feature matching

Refine matching using RANSAC to estimate fundamental matrix between each pair



Correspondence estimation

- Link up pairwise matches to form connected components of matches across several images

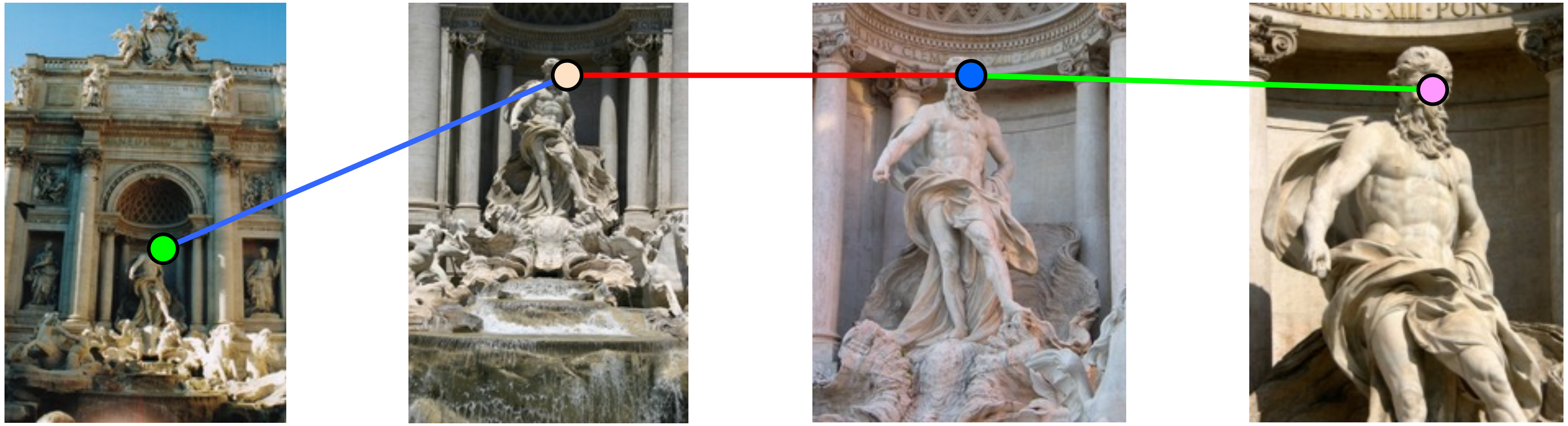


Image 1 Image 2 Image 3 Image 4

Global SfM

- Given N images, there are ${}^{N}C_2$ pairs. Many of these pairs will have no overlaps in views and/or Fundamental/Essential matrix between them can not be reliably estimated using RANSAC.
  - Consider we have N0 (N0 < ${}^{N}C_2$) pairs of images with fundamental matrix estimated

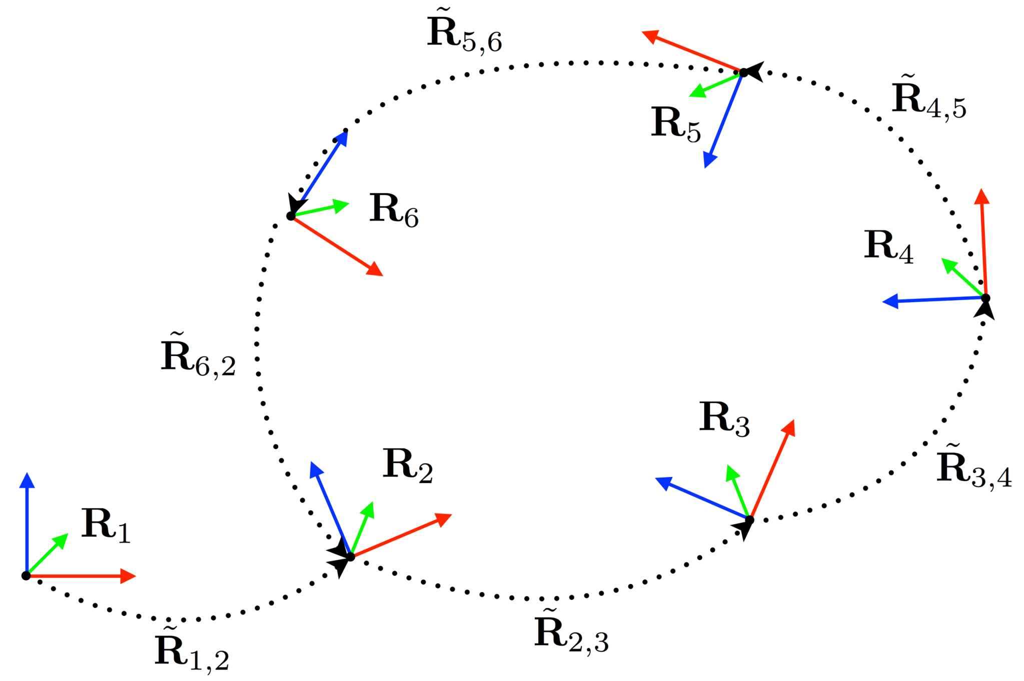
- For each N0 pairs of images decompose essential matrix into relative rotation and translation between two cameras: R_{ij} and t_{ij}.

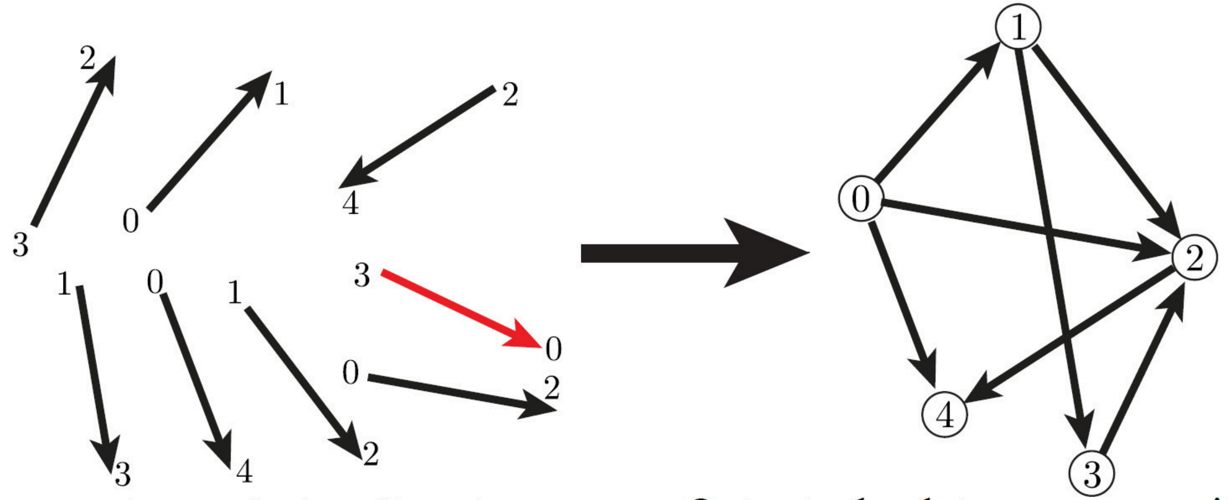
- Can we solve for global (world coordinate) rotation and translation of the cameras, given pairwise measurements, i.e.
  - Given R_{ij} and t_{ij} for N0 pairs, find R_k & T_k for N cameras.

- Once we have the cameras we can better initialize the Bundle Adjustment problem.

Rotation & Translation Averaging

Given R_{ij} and t_{ij} for N0 pairs, find R_k & T_k for N cameras



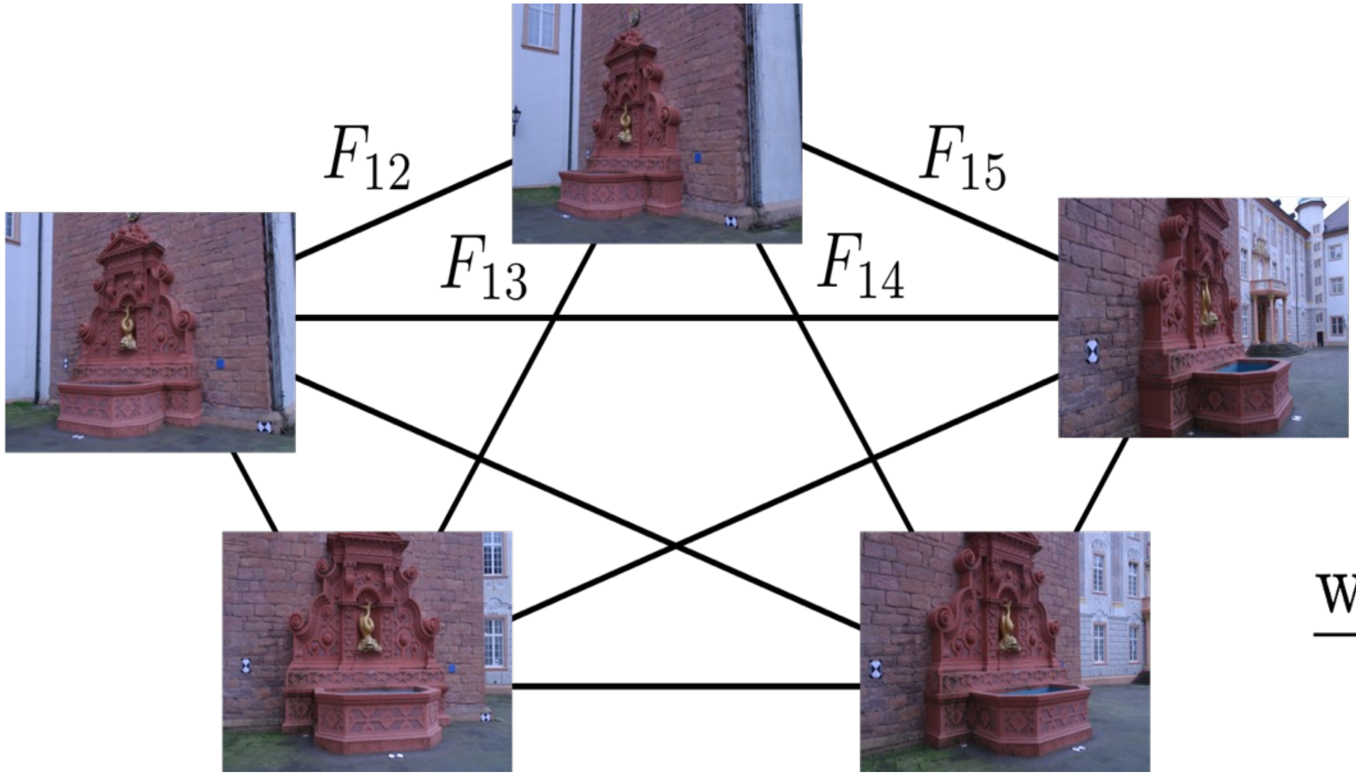


Input unit translation directions

Output: absolute camera positions

Credits: Venu Madhab Govindu

Camera Pose estimation as matrix completion over Fundamental matrices



$$F = \begin{bmatrix} 0 & F_{12} & F_{13} & F_{14} & F_{15} \\ F_{21} & 0 & F_{23} & F_{24} & F_{25} \\ F_{31} & F_{32} & 0 & F_{34} & F_{35} \\ F_{41} & F_{42} & F_{43} & 0 & F_{45} \\ F_{51} & F_{52} & F_{53} & F_{54} & 0 \end{bmatrix}$$

$$\underline{\text{with } F = A + A^T \text{ and } rank(A) = 3.}$$

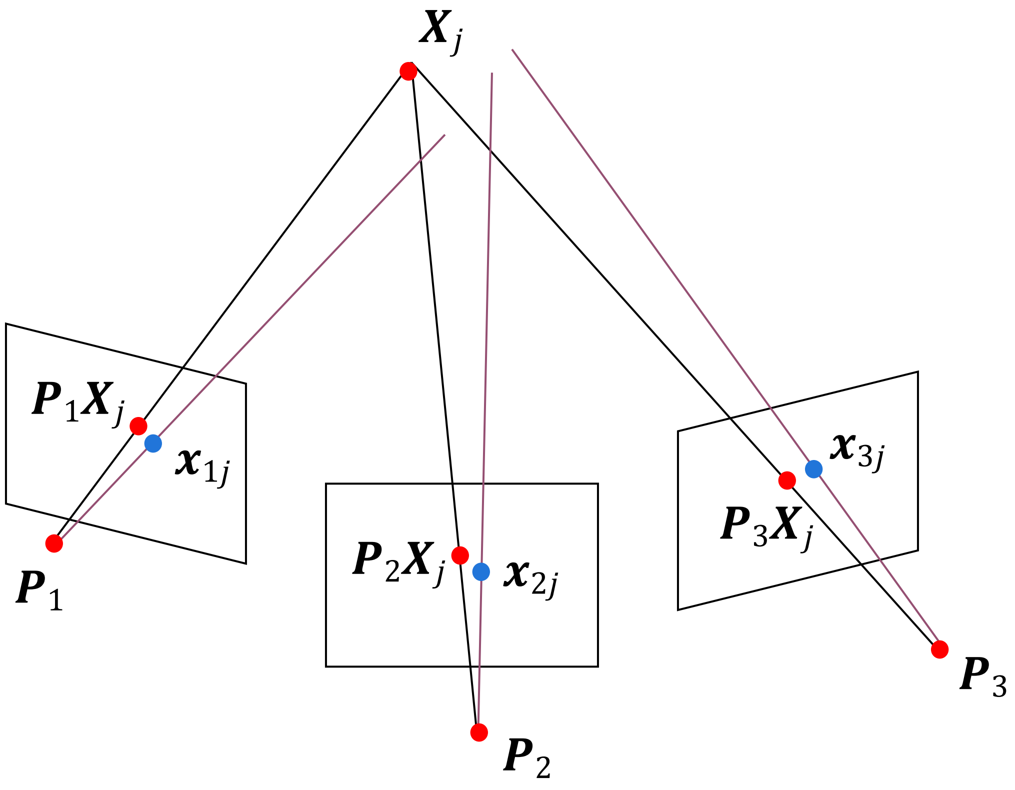
- Proves a low-rank property of all the cameras capturing different images of a scene.
- Solves a low-rank camera pose recovery algorithm from Structure from Motion.

"A New Rank Constraint on Multi-view Fundamental Matrices", Sengupta et. al. CVPR 2017.

Bundle adjustment

- Non-linear method for refining structure and motion
- Minimize reprojection error (with lots of bells and whistles):
-
- $\sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} d\left(\boldsymbol{x}_{ij} - \mathrm{proj}(\boldsymbol{P}_i \boldsymbol{X}_j)\right)^2$

visibility flag: is point j visible in view i?

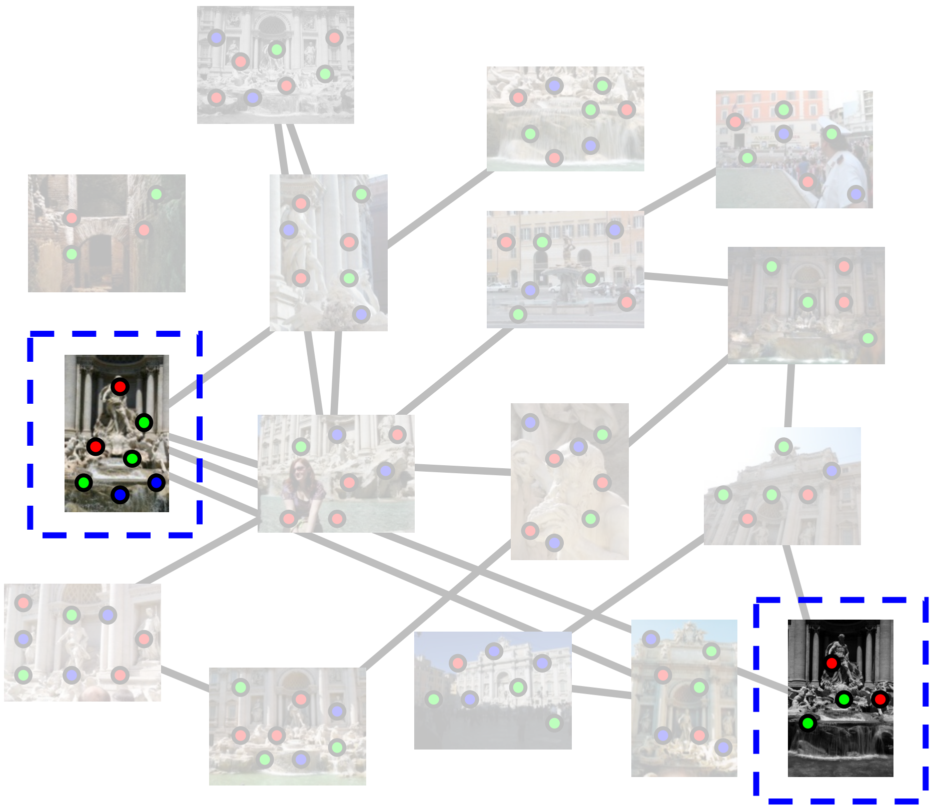


- **Initialize P_i's by solving global SfM**
  - **Rotation Averaging**
  - **Translation Averaging**

B. Triggs et al. [Bundle adjustment – A modern synthesis](). International Workshop on Vision Algorithms, 1999

Incremental SfM

Can handle large scale scene, more than Global SfM



- Automatically select an initial pair of images

1. Picking the initial pair

- We want a pair with many matches, but which has as large a baseline as possible

lots of matches
small baseline

large baseline
very few matches

large baseline
lots of matches

Incremental SFM

- Pick a pair of images with lots of inliers (and preferably, good EXIF data)
  - Initialize intrinsic parameters (focal length, principal point) from EXIF
  - Estimate extrinsic parameters ($\boldsymbol{R}$ and $\boldsymbol{t}$) using five-point algorithm (similar to 8-pt algorithm but for essential matrix)
  - Use triangulation to initialize model points
- While remaining images exist
  - Find an image with many feature matches with images in the model
  - Run RANSAC on feature matches to register new image to 3D model points
  - Triangulate new points
  - Perform bundle adjustment to re-optimize everything
  - Optionally, align with GPS from EXIF data or ground control points

Next Best View Problem

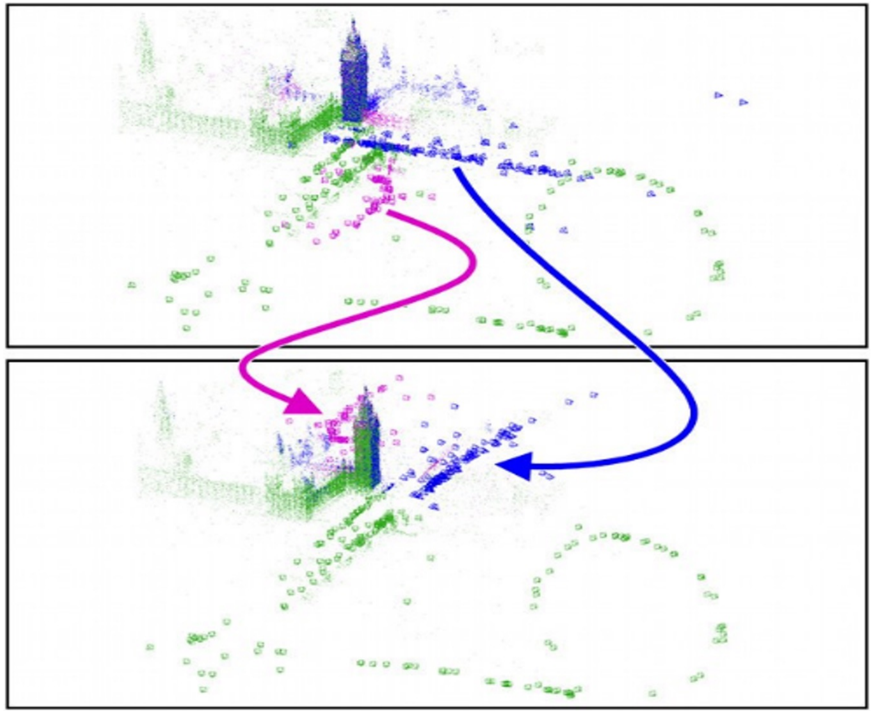
- Choice of next view impacts reconstruction quality
    - almost identical view => high uncertainty in triangulation
    - very different view => low overlap and high camera uncertainty
    - single bad choice may impact the whole reconstruction

- Popular next best view methods:
    - choose view with seeing the most triangulated points
    - minimize reconstruction uncertainty
        - depends on number of observations
        - distribution in the image

Challenges: The devil is in the details

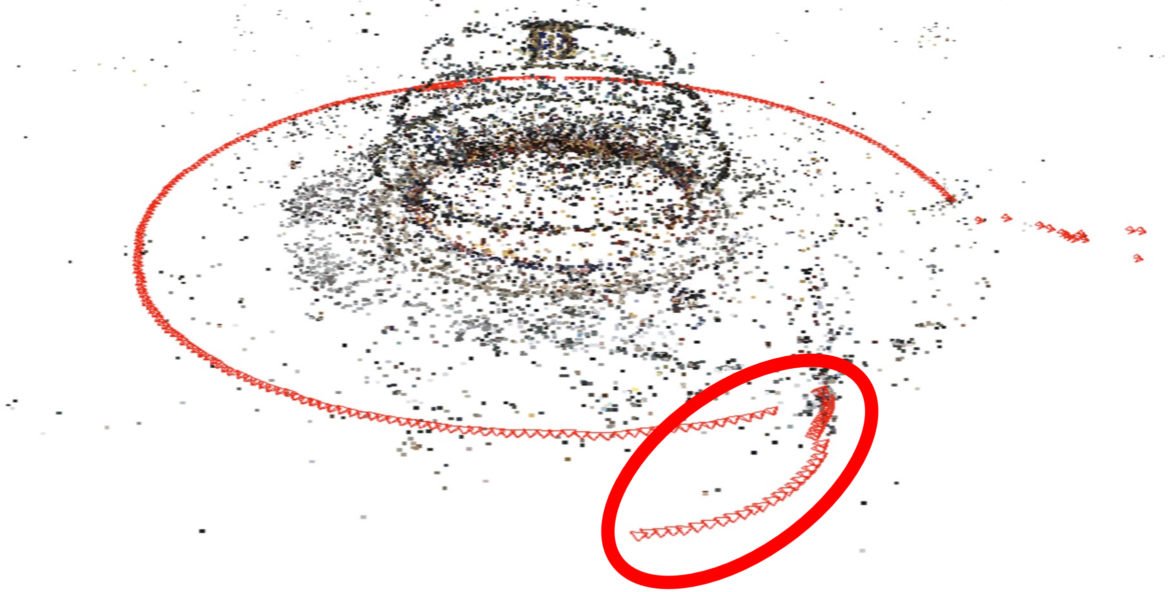
- Handling degenerate configurations (e.g., homographies)
- Filtering out incorrect matches
- Dealing with repetitions and symmetries

Repetitive structures cause catastrophic failures



https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Repetitive structures cause catastrophic failures



R. Kataria et al. [Improving Structure from Motion with Reliable Resectioning](). 3DV 2020

Challenges: The devil is in the details

- Handling degenerate configurations (e.g., homographies)
- Filtering out incorrect matches
- Dealing with repetitions and symmetries
- Reducing error accumulation and closing loops

Loop Detection/Closure

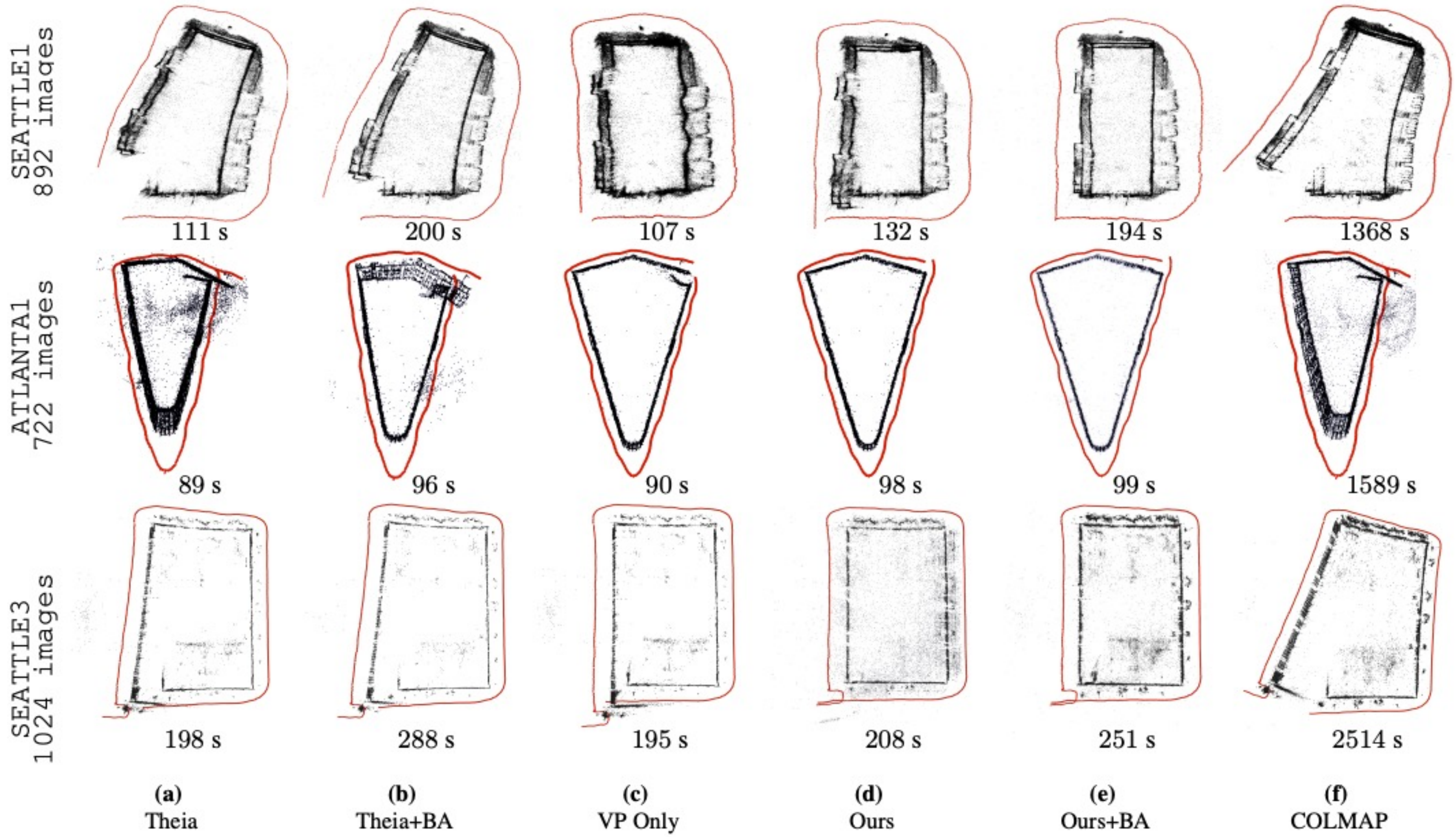
- Problem:
  - Structure from motion is an incremental process
  - Drift accumulates

- Mitigation:
  - Retrieval of long range connections

Reducing error accumulation and closing loops

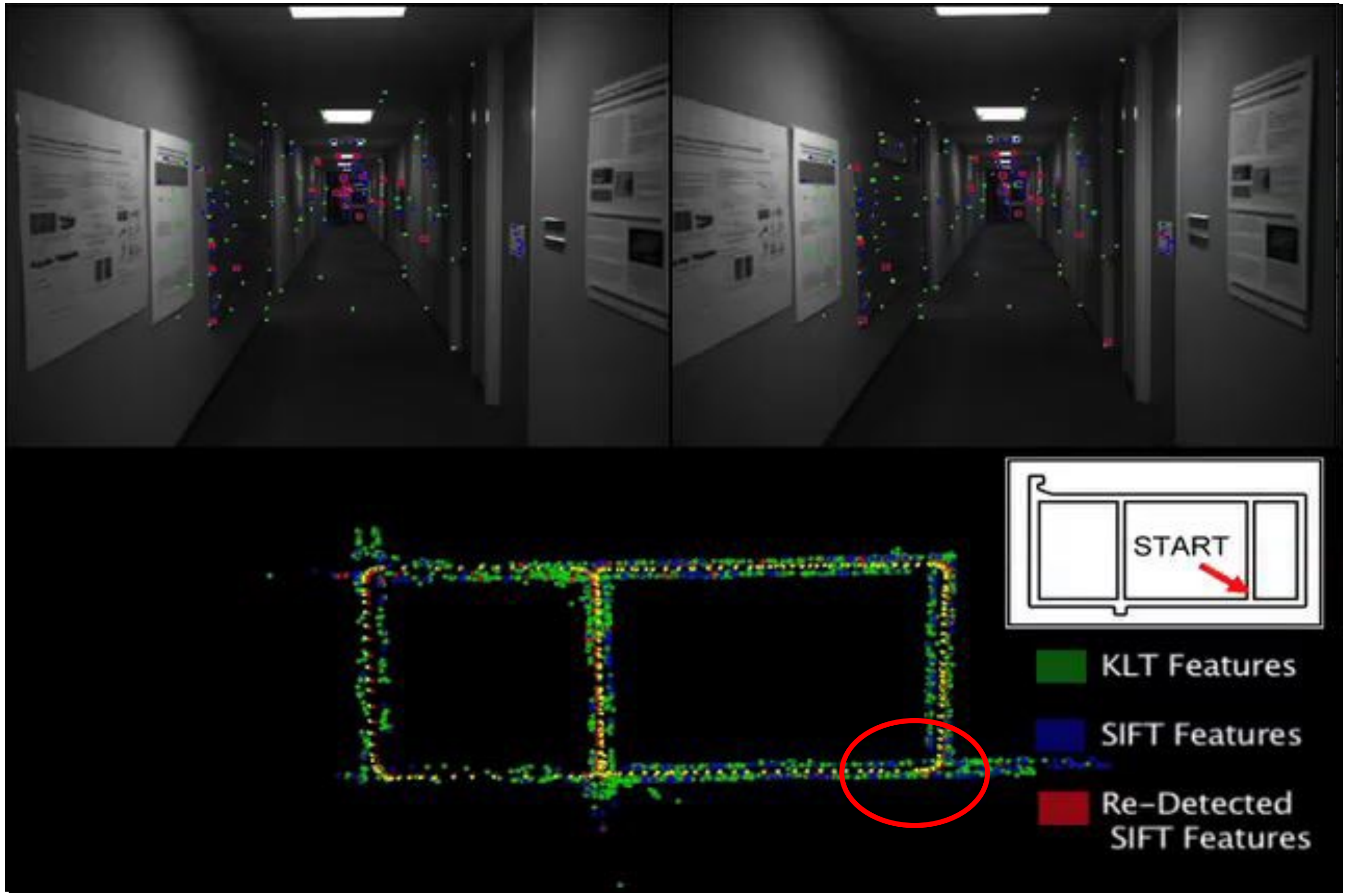
A. Holynski et al. [Reducing Drift in Structure From Motion Using Extended Features](). arXiv 2020

Reducing error accumulation and closing loops

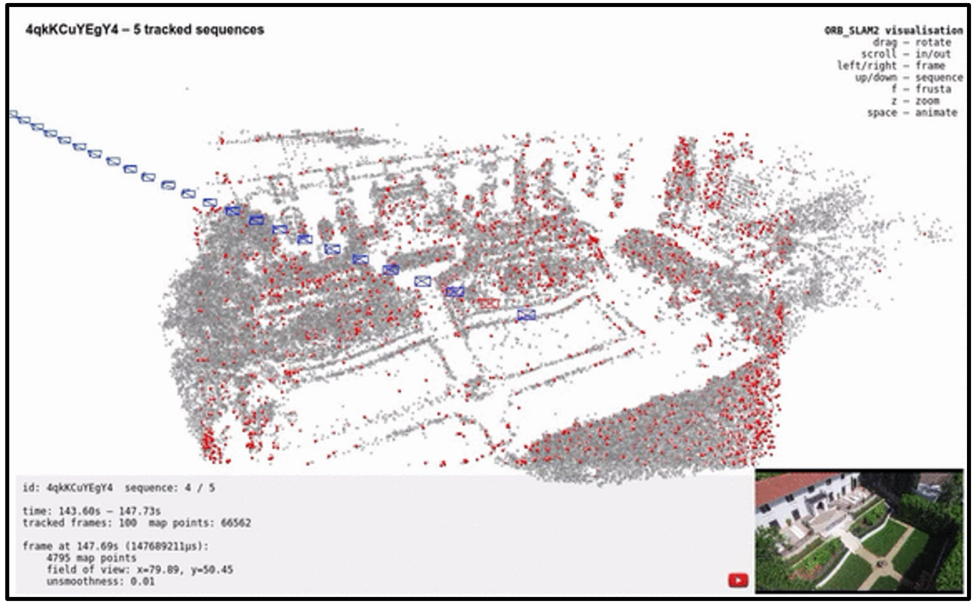


A. Holynski et al. [Reducing Drift in Structure From Motion Using Extended Features](). arXiv 2020

Loop Closure

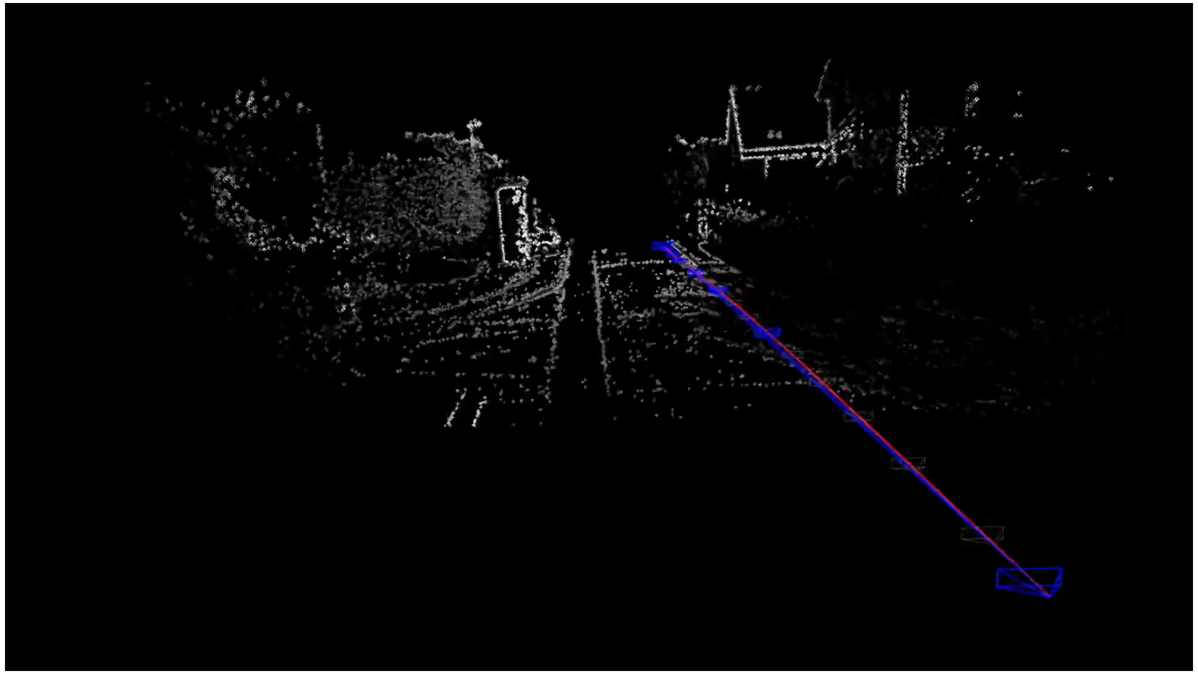


Can also compute camera poses from video (often called Visual SLAM)



Visual Simultaneous Localization and Mapping (V-SLAM)

- Main differences with SfM:
  - Continuous visual input from sensor(s) over time
  - Gives rise to problems such as loop closure
  - Often the goal is to be online / real-time



Video from Daniel Cremer's Lab

SFM software

- Bundler
- OpenSfM
- OpenMVG
- VisualSFM
- COLMAP (Structure-from-motion revisited, JL Schonberger, JM Frahm, CVPR 2016, from UNC!)

- See also Wikipedia's list of toolboxes

SfM applications

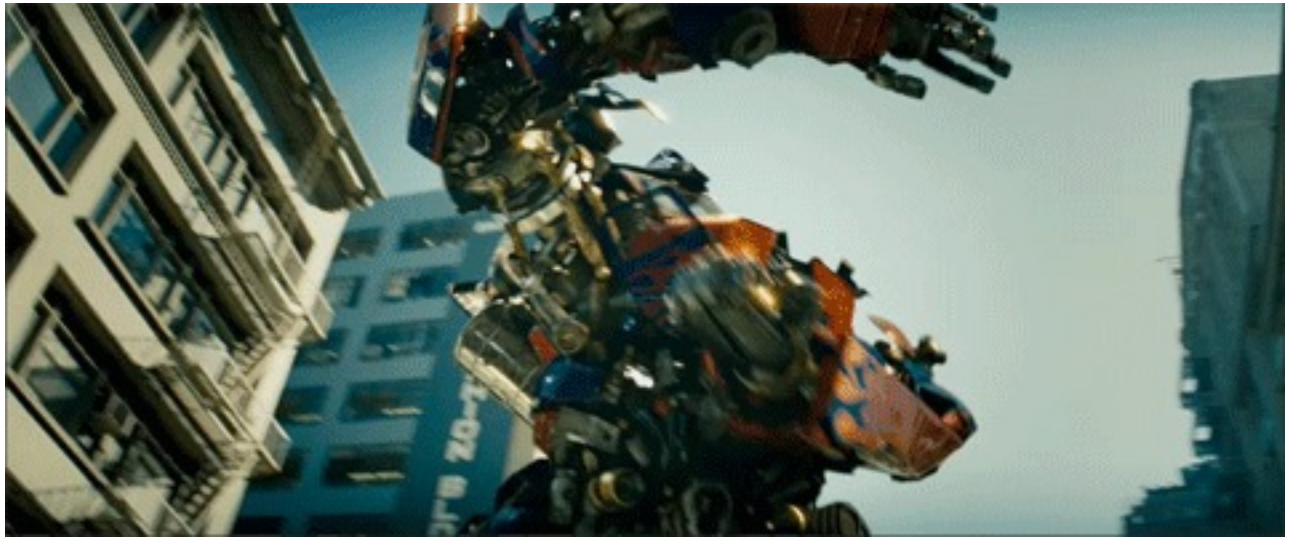
- 3D modeling
- Surveying
- Robot navigation and mapmaking
- Virtual and augmented reality
- Visual effects ("Match moving")
  - https://www.youtube.com/watch?v=RdYWp70P_kY

Applications: Match Moving

Or Motion tracking, solving for camera trajectory

Integral for visual effects (VFX)

Why?



Applications: Visual Reality & Augmented Reality

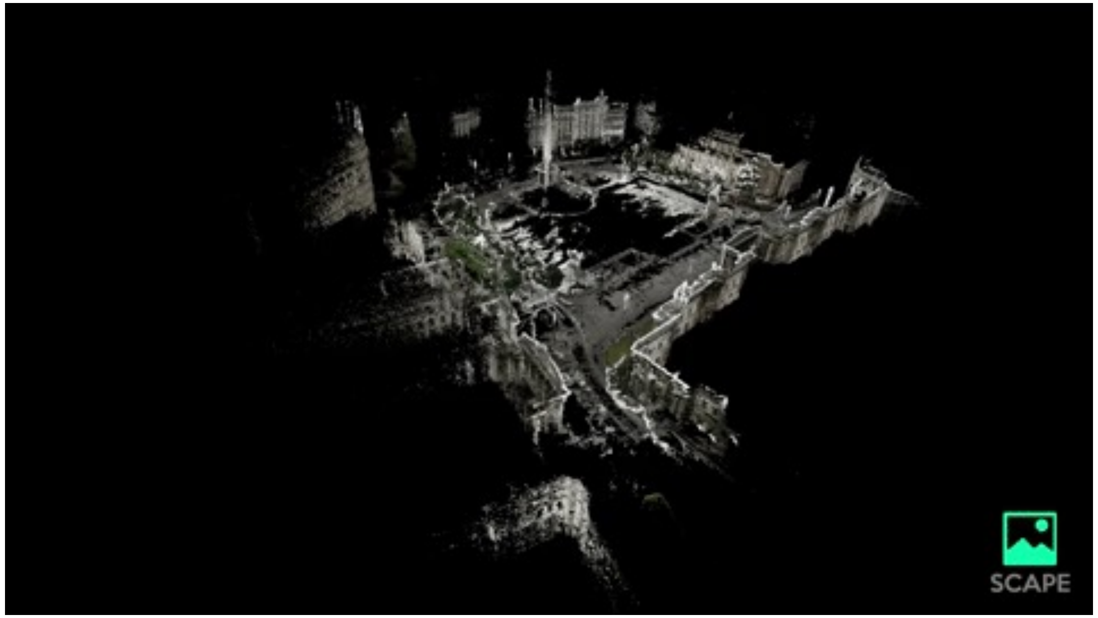
Oculus

https://www.youtube.com/watch?v=KOG7yTz1iTA

Hololens

https://www.youtube.com/watch?v=FMtvrTGnP04

Applications: Visual Reality & Augmented Reality



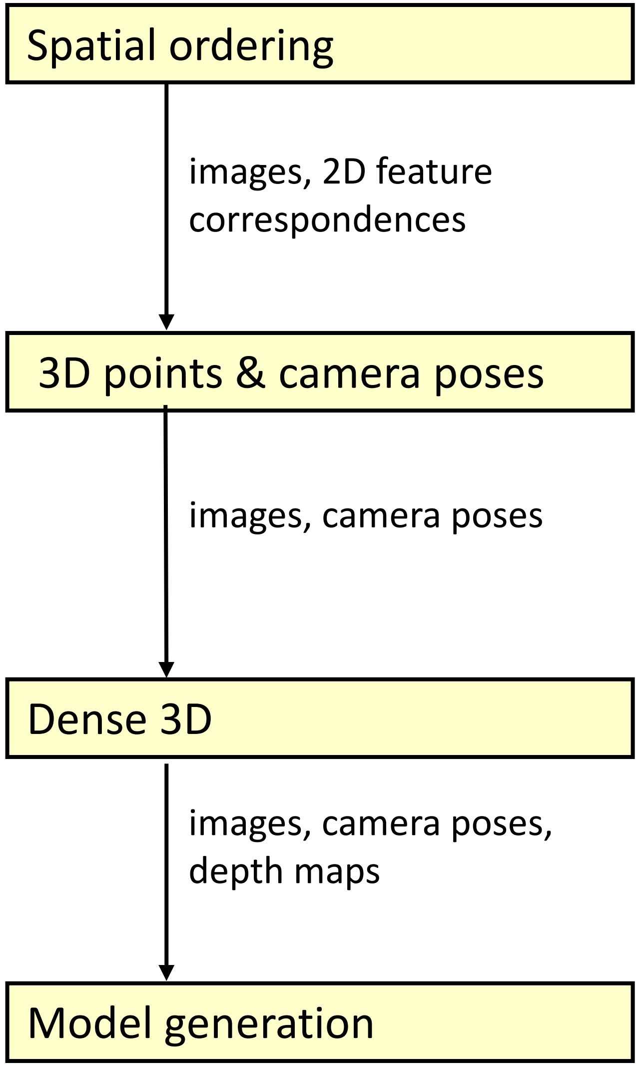
Scape: Building the 'AR Cloud': Part Three —3D Maps, the Digital Scaffolding of the 21st Century

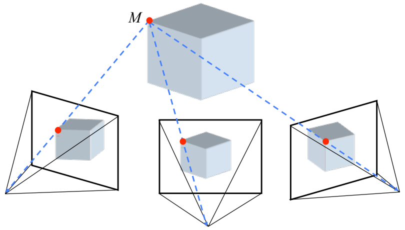
https://medium.com/scape-technologies/building-the-ar-cloud-part-three-3d-maps-the-digital-scaffolding-of-the-21st-century-465fa55782dd

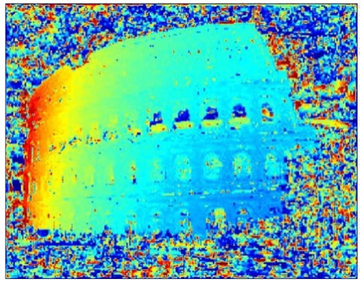
Application: AR walking directions

https://www.theverge.com/2019/8/8/20776247/google-maps-live-view-ar-walking-directions-ios-android-feature

3D model from video

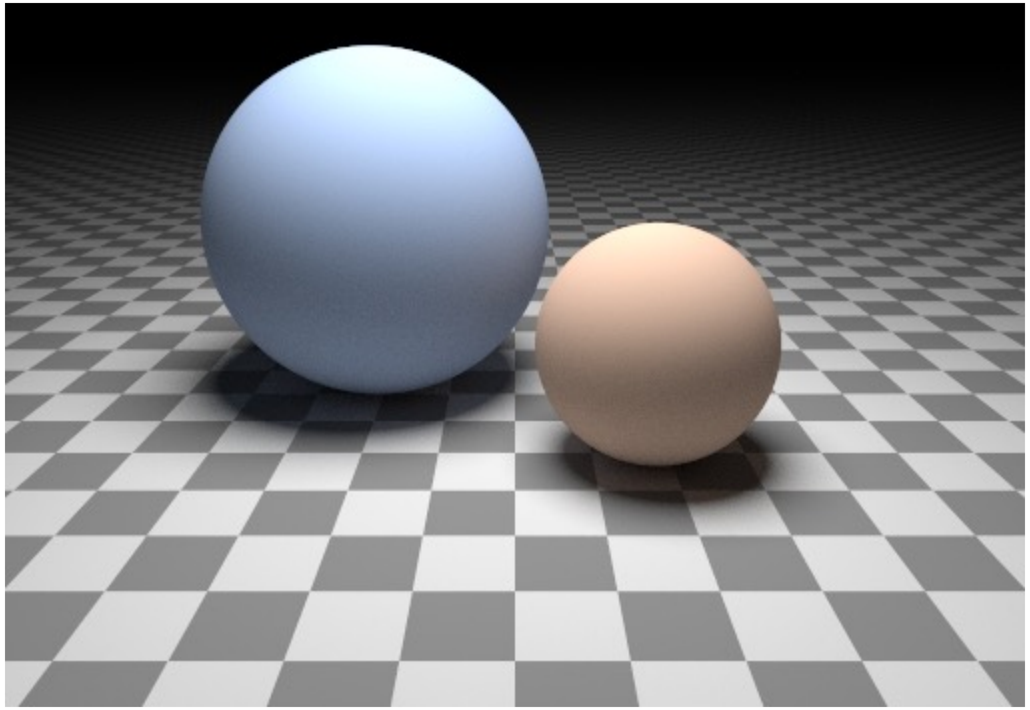






Photometric Stereo

Can we determine shape from lighting?



- Are these spheres?
  - Or just flat discs painted with varying color (albedo)?
  - There is ambiguity between *shading* and *reflectance*
  - But still, as humans we can understand the shapes of these objects

What we know: Stereo

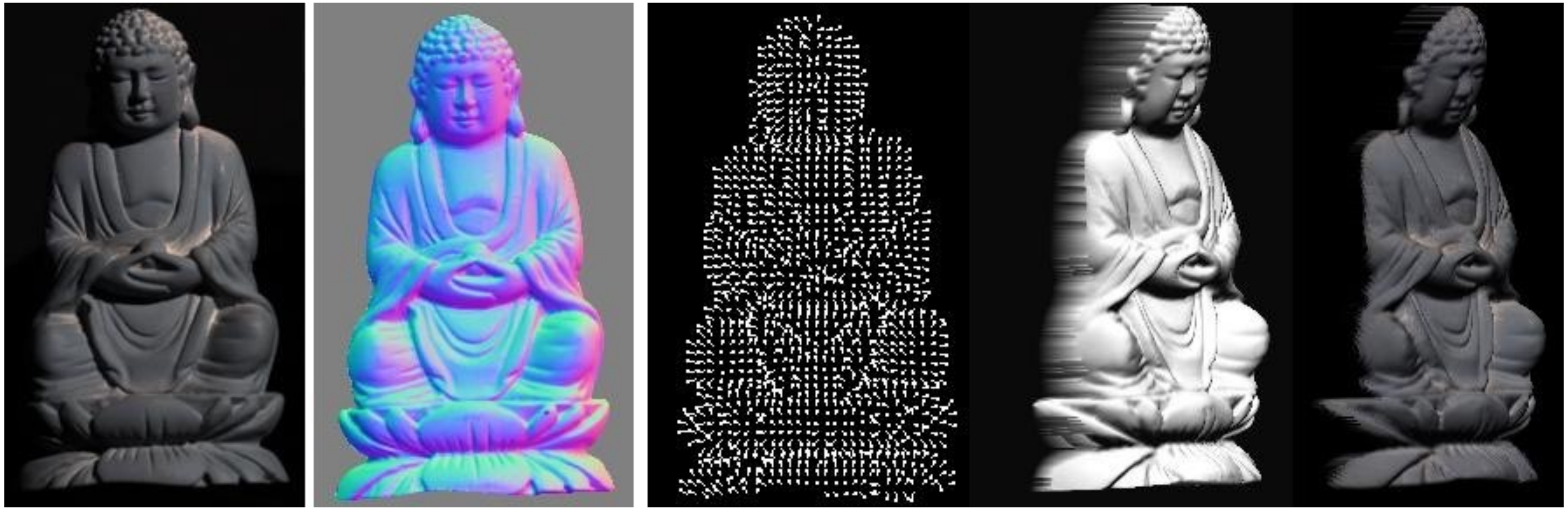
Key Idea: use camera motion to compute shape

Next: Photometric Stereo

Key Idea: use pixel brightness to understand shape

Photometric Stereo

What results can you get?



Input (1 of 12) | Normals (RGB colormap) | Normals (vectors) | Shaded 3D rendering | Textured 3D rendering

Today’s class

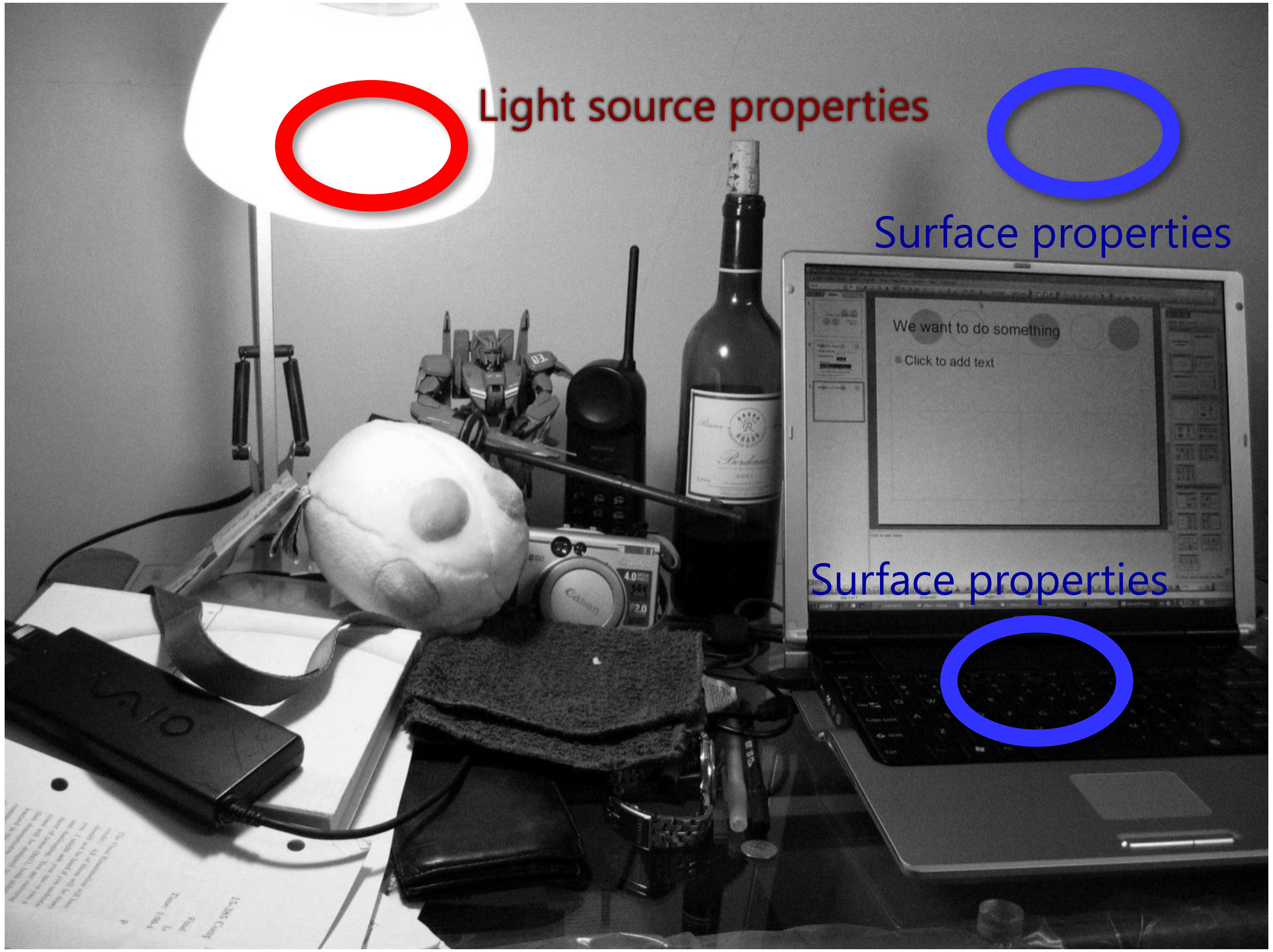
- Measuring Light (recap)
- Image formation with shape, reflectance, and illumination
- Shape from Shading
- Photometric Stereo
- Uncalibrated Photometric Stereo
- Generalized Bas-Relief Ambiguity
- Photometric Stereo in ‘deep learning era’.

Today’s class

- Measuring Light (recap)
- Image formation with shape, reflectance, and illumination
- Shape from Shading
- Photometric Stereo
- Uncalibrated Photometric Stereo
- Generalized Bas-Relief Ambiguity
- Photometric Stereo in ‘deep learning era’.

Radiometry

- What determines the brightness of a pixel?



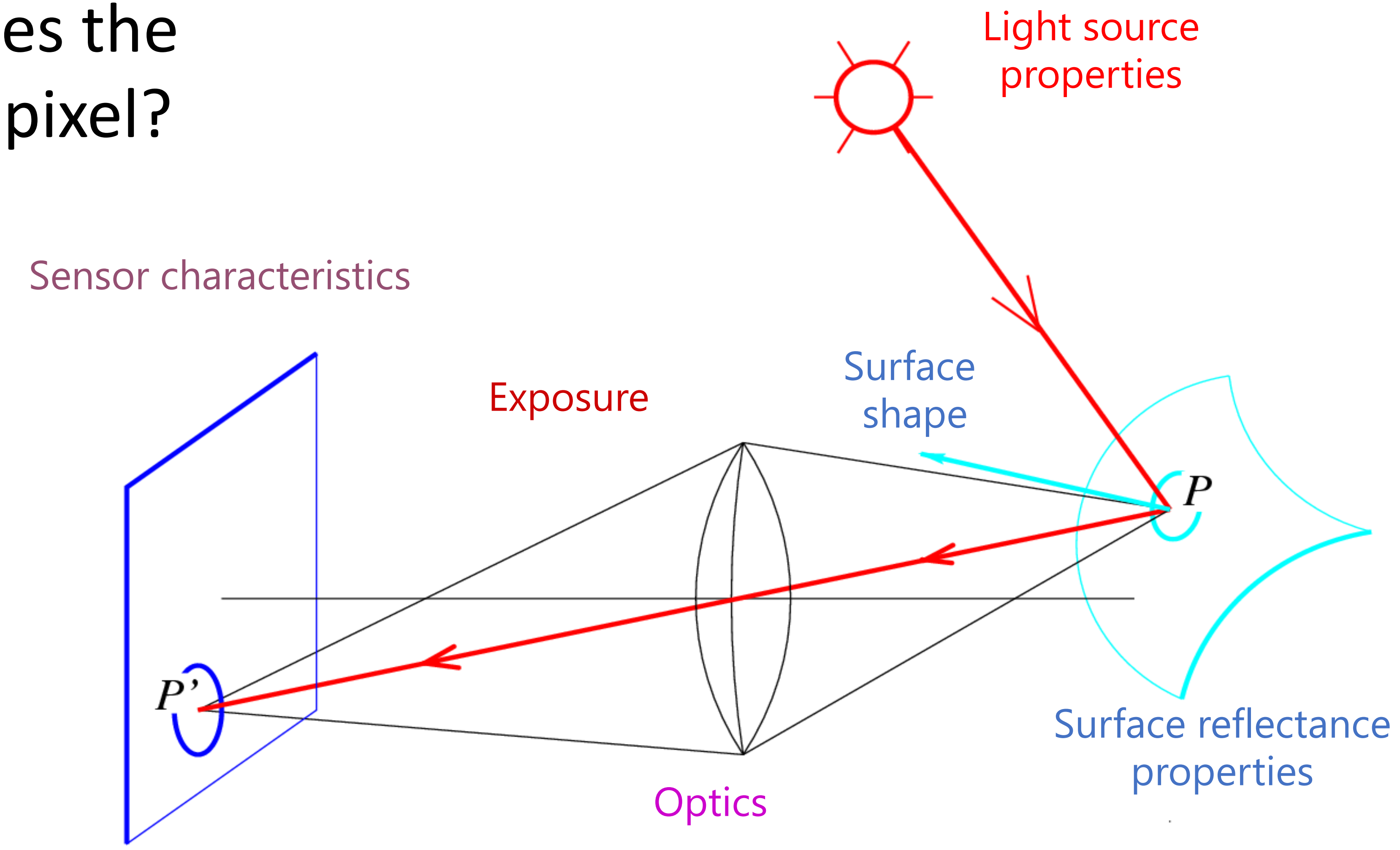
Radiometry

- What determines the brightness of a pixel?

@robertwestonbreshears
https://www.instagram.com/p/BtgX55ZBhU-/

Radiometry

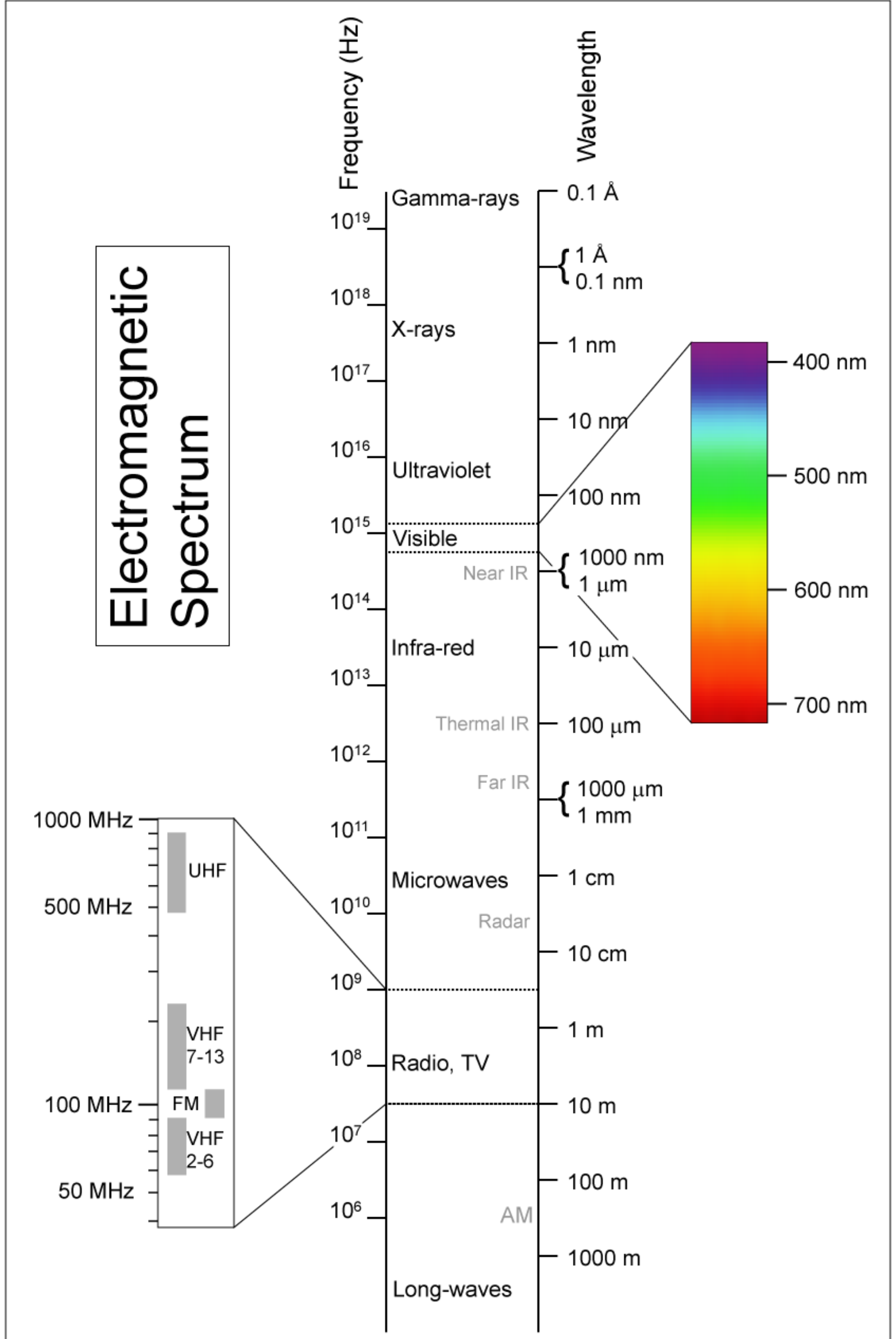
- What determines the brightness of a pixel?



Slide by L. Fei-Fei

Visible light

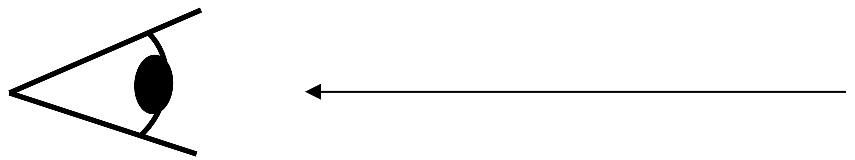
We "see" electromagnetic radiation in a range of wavelengths



What is light?

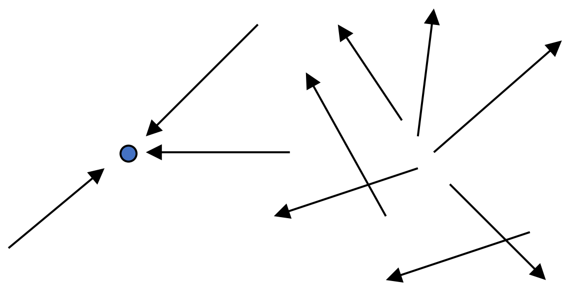
Electromagnetic radiation (EMR) moving along rays in space

- R(λ) is EMR, measured in units of power (watts)
  - λ is wavelength



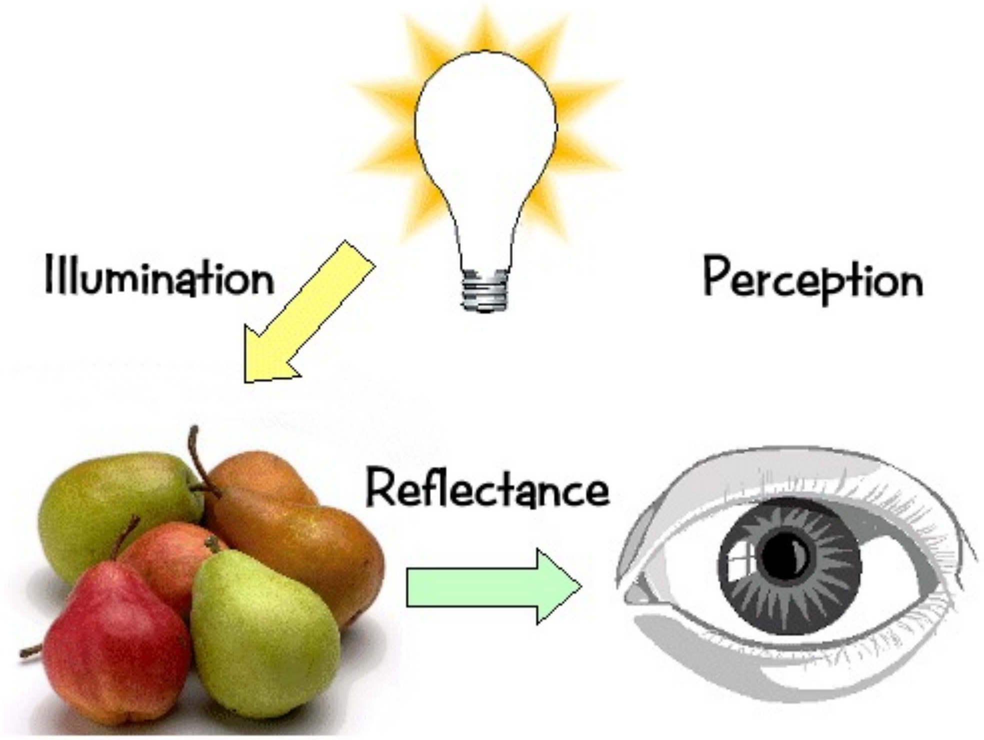
Light field

- We can describe all of the light in the scene by specifying the radiation (or **"radiance"** along all light rays) arriving at every point in space and from every direction



The ***plenoptic function*** describes all of this light: $R(X, Y, Z, \theta, \phi, \lambda, t)$

Light transport



Light sources

- Basic types
  - point source
  - directional source
    - a point source that is infinitely far away
  - area source
    - a union of point sources

- More generally
  - a light field can describe *any* distribution of light sources
  - Environment map

- What happens when light hits an object?

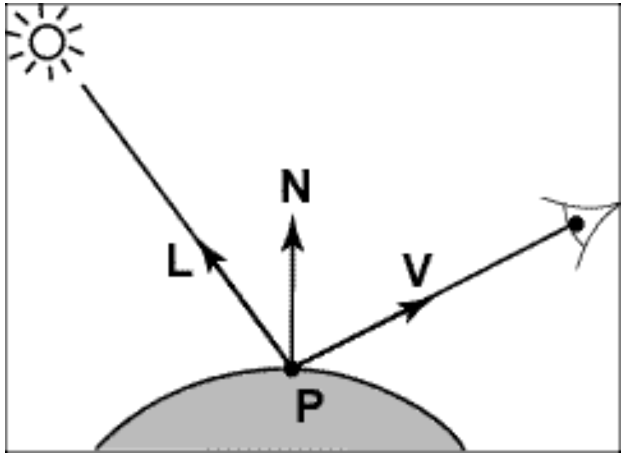
Today’s class

- Measuring Light (recap)
- Image formation with shape, reflectance, and illumination
- Shape from Shading
- Photometric Stereo
- Uncalibrated Photometric Stereo
- Generalized Bas-Relief Ambiguity
- Photometric Stereo in ‘deep learning era’.

Modeling Image Formation

We need to reason about:

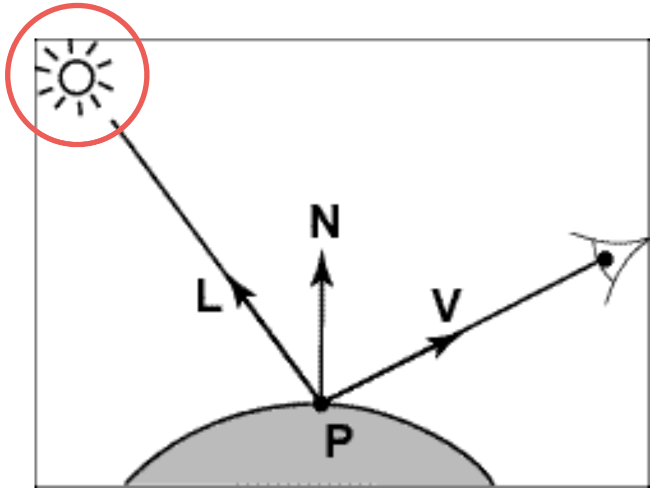
- How light interacts with the scene
- How a pixel value is related to light energy in the world



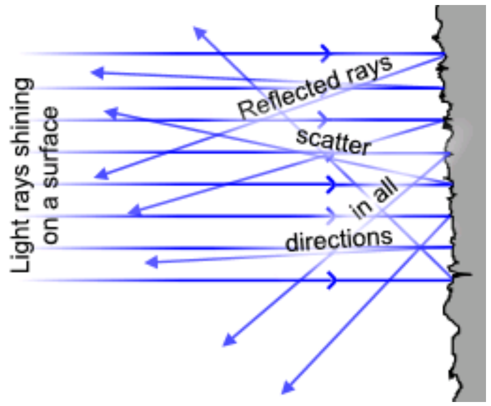
Track a “ray” of light all the way from light source to the sensor

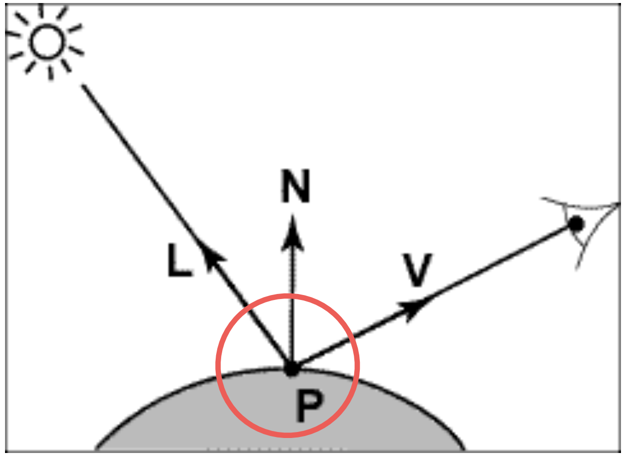
Directional Lighting

- Key property: all rays are parallel
- Equivalent to an infinitely distant point source



Lambertian Reflectance



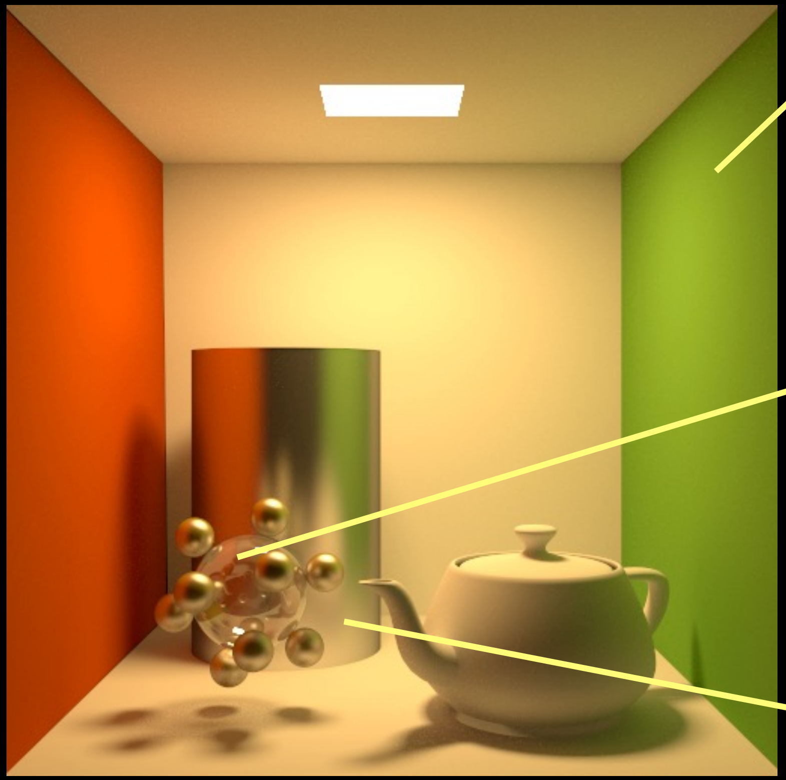


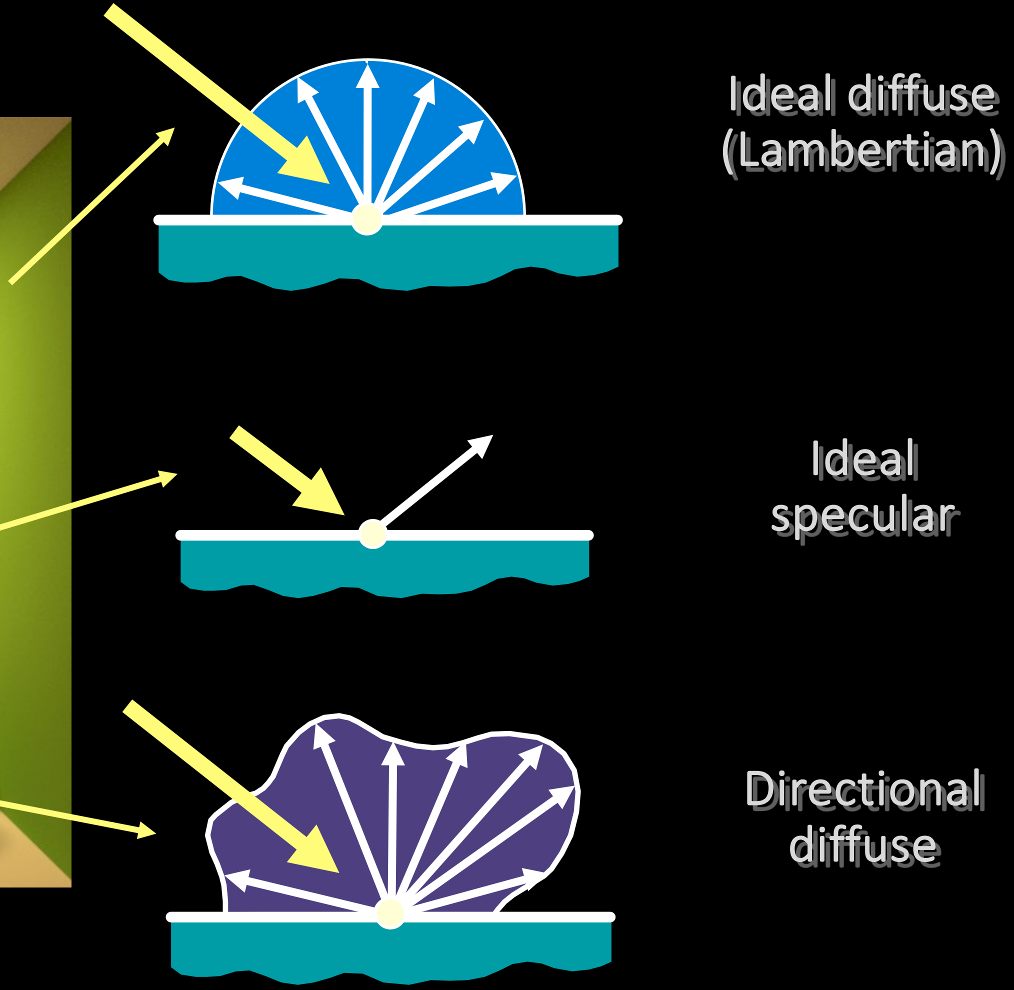
$$I = N \cdot L$$

Image intensity = Surface normal • Light direction

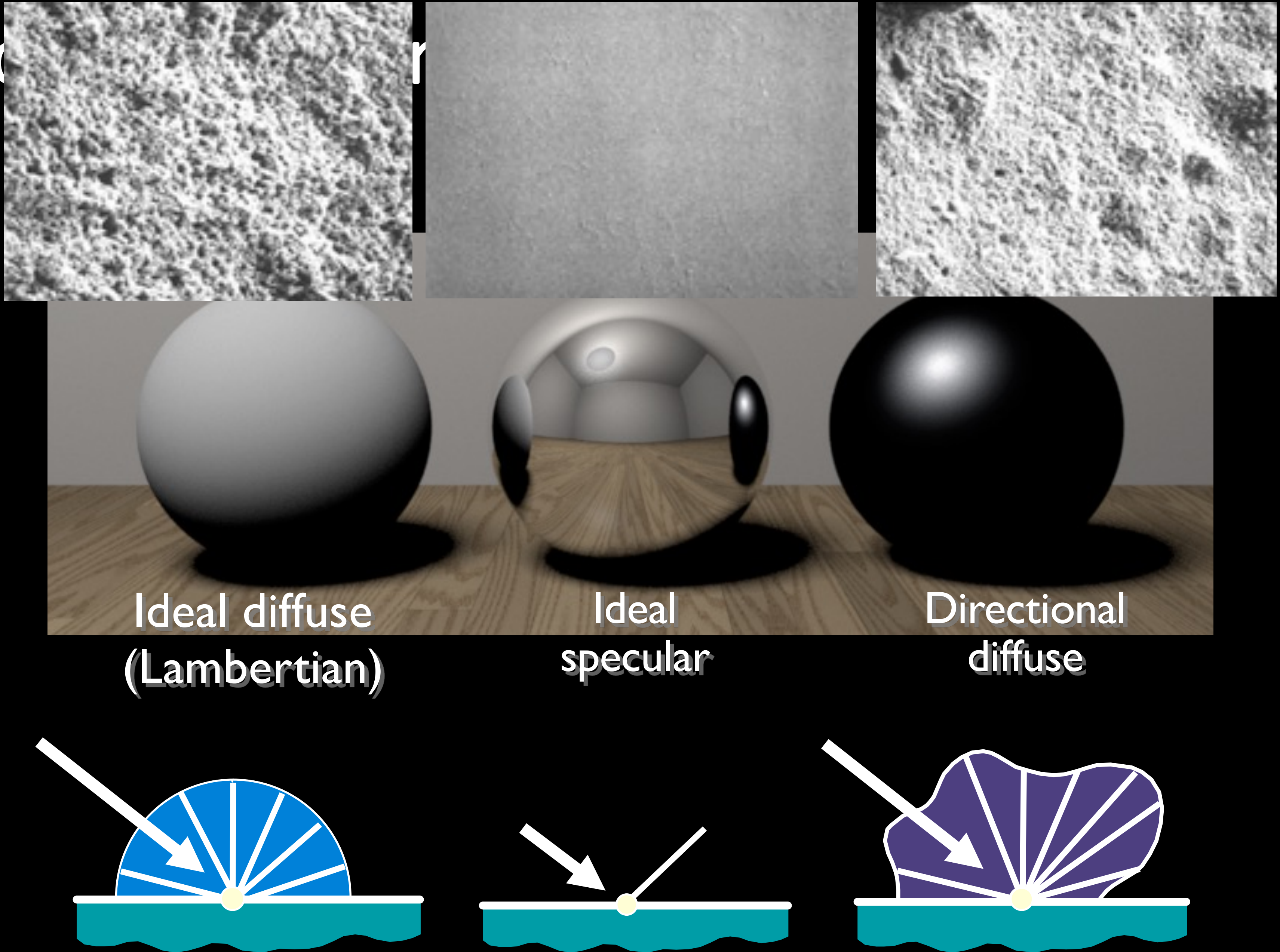
Image intensity $\propto$ cos(angle between N and L)

Materials - Three Forms



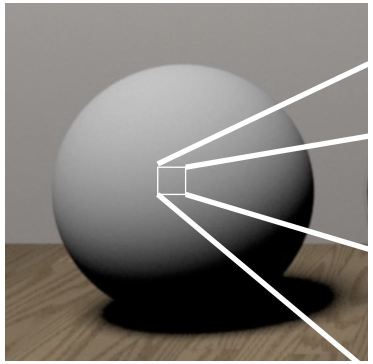


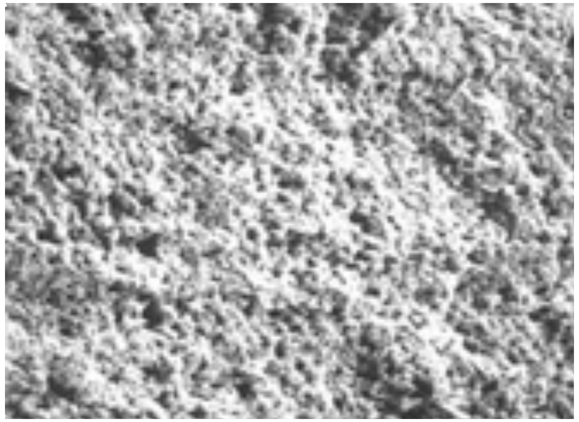
Refle


Ideal diffuse
(Lambertian)
Ideal
specular
Directional
diffuse

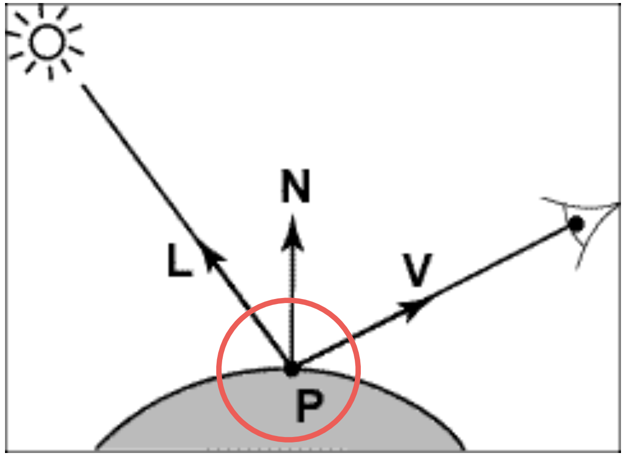
Ideal Diffuse Reflection

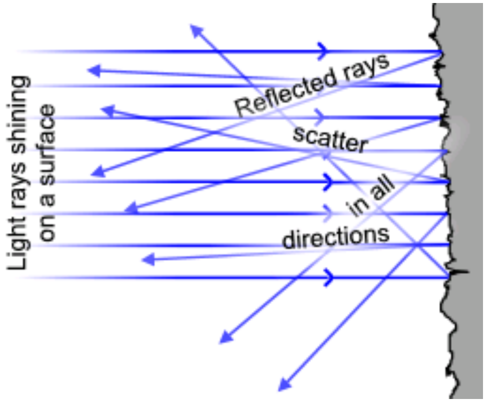
- Characteristic of multiple scattering materials
- An idealization but reasonable for matte surfaces





Lambertian Reflectance



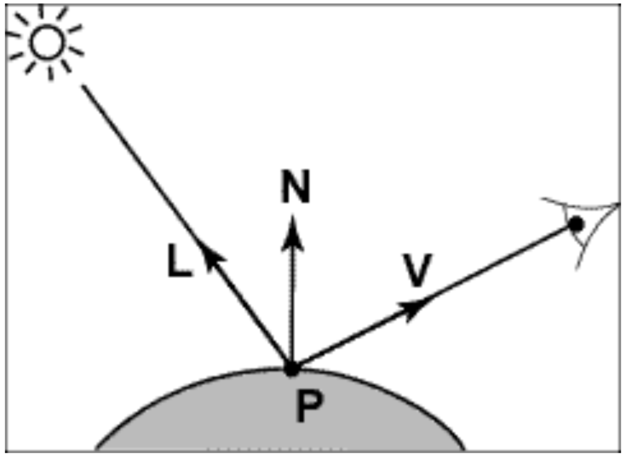


$$I = N \cdot L$$

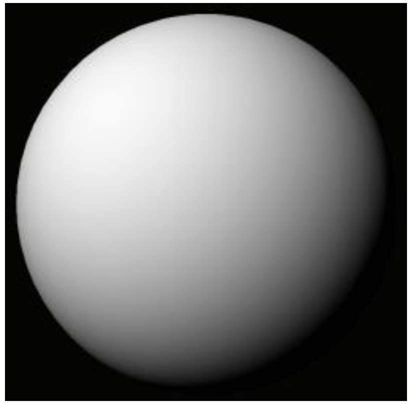
1. Reflected energy is proportional to cosine of angle between L and N **(incoming)**

2. Measured intensity is viewpoint-independent **(outgoing)**

Final Lambertian image formation model



$$I = k_d \mathbf{N} \cdot \mathbf{L}$$

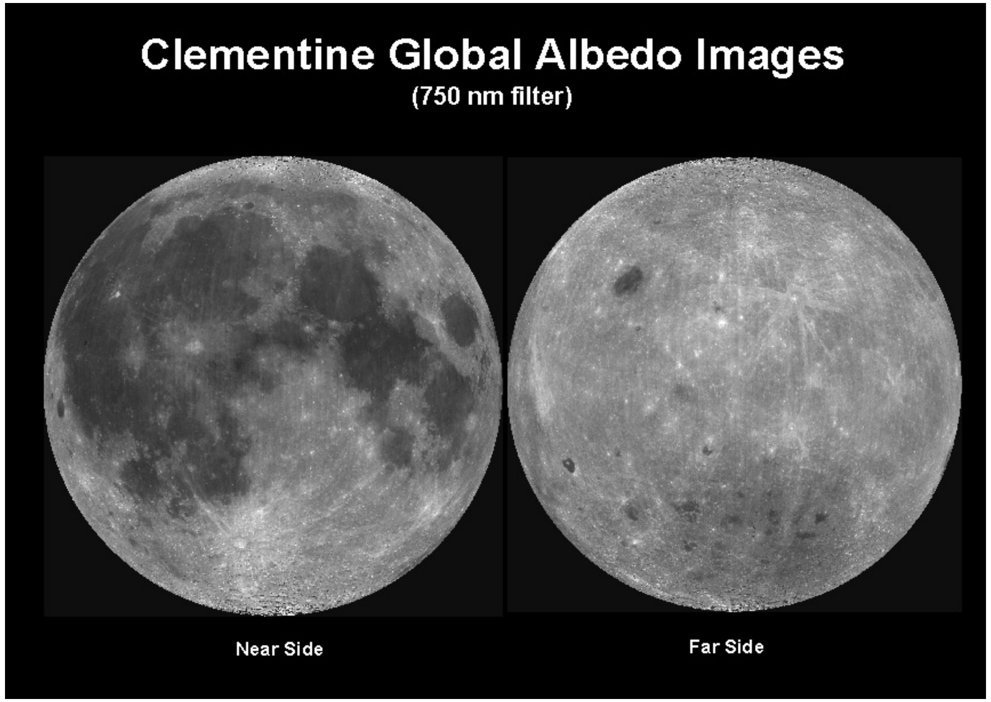


1. Diffuse **albedo**: what fraction of incoming light is reflected?
   - Introduce scale factor k_d
2. Light intensity: how much light is arriving?
   - Compensate with camera exposure (global scale factor)
3. Camera response function
   - Assume pixel value is linearly proportional to incoming energy (perform radiometric calibration if not)

Albedo

Sample albedos

Surface	Typical albedo
Fresh asphalt	0.04[4]
Open ocean	0.06[5]
Worn asphalt	0.12[4]
Conifer forest (Summer)	0.08,[6] 0.09 to 0.15[7]
Deciduous trees	0.15 to 0.18[7]
Bare soil	0.17[8]
Green grass	0.25[8]
Desert sand	0.40[9]
New concrete	0.55[8]
Ocean ice	0.5–0.7[8]
Fresh snow	0.80–0.90[8]



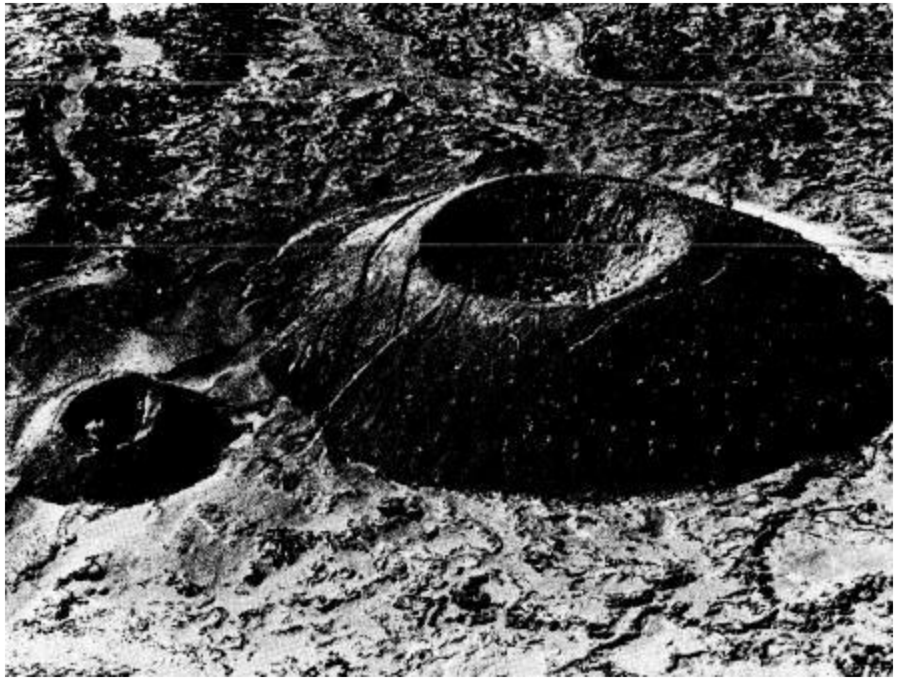
Objects can have varying albedo and albedo varies with wavelength

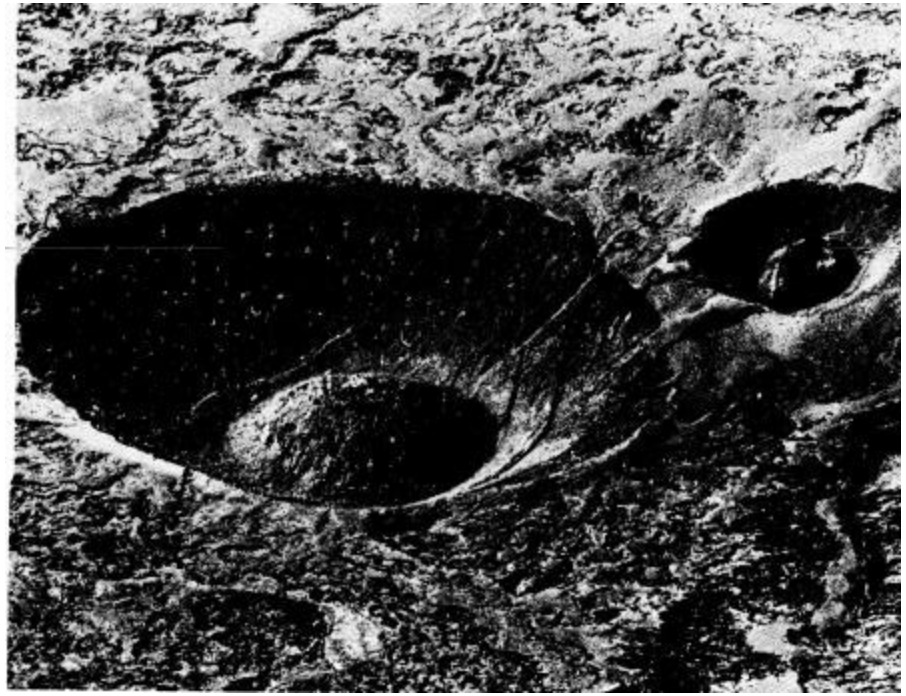
Source: https://en.wikipedia.org/wiki/Albedo

Today’s class

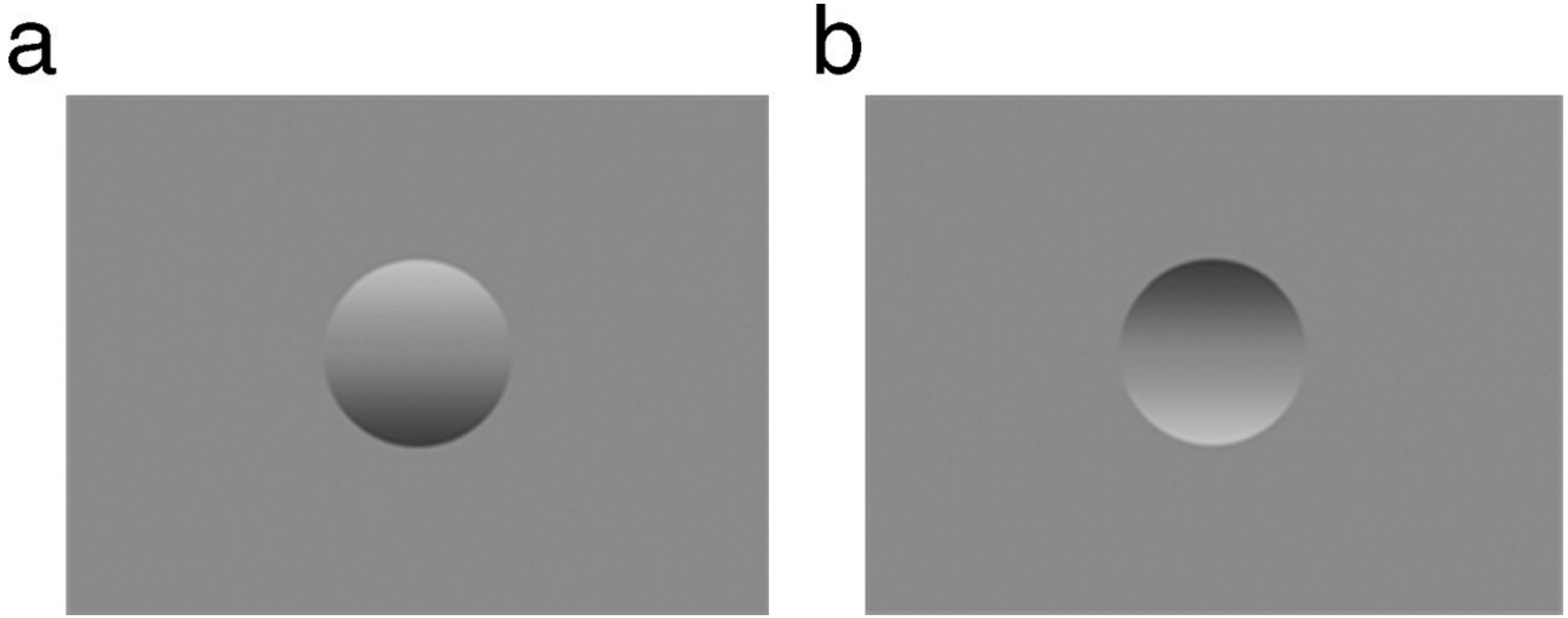
- Measuring Light (recap)
- Image formation with shape, reflectance, and illumination
- Shape from Shading
- Photometric Stereo
- Uncalibrated Photometric Stereo
- Generalized Bas-Relief Ambiguity
- Photometric Stereo in ‘deep learning era’.

Human Perception





Examples of the classic bump/dent stimuli used to test lighting assumptions when judging shape from shading, with shading orientations (a) 0° and (b) 180° from the vertical.



Thomas R et al. J Vis 2010;10:6

Human Perception

- Our brain often perceives shape from shading.

- Mostly, it makes many assumptions to do so.

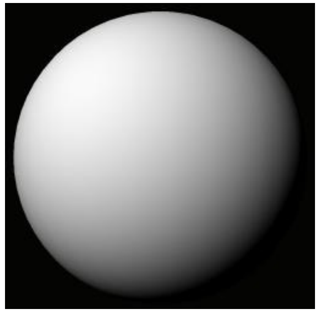
- For example:

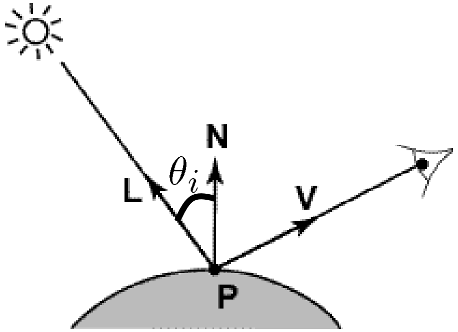
    Light is coming from above (sun).

    Biased by occluding contours.

by V. Ramachandran

A Single Image: Shape from shading





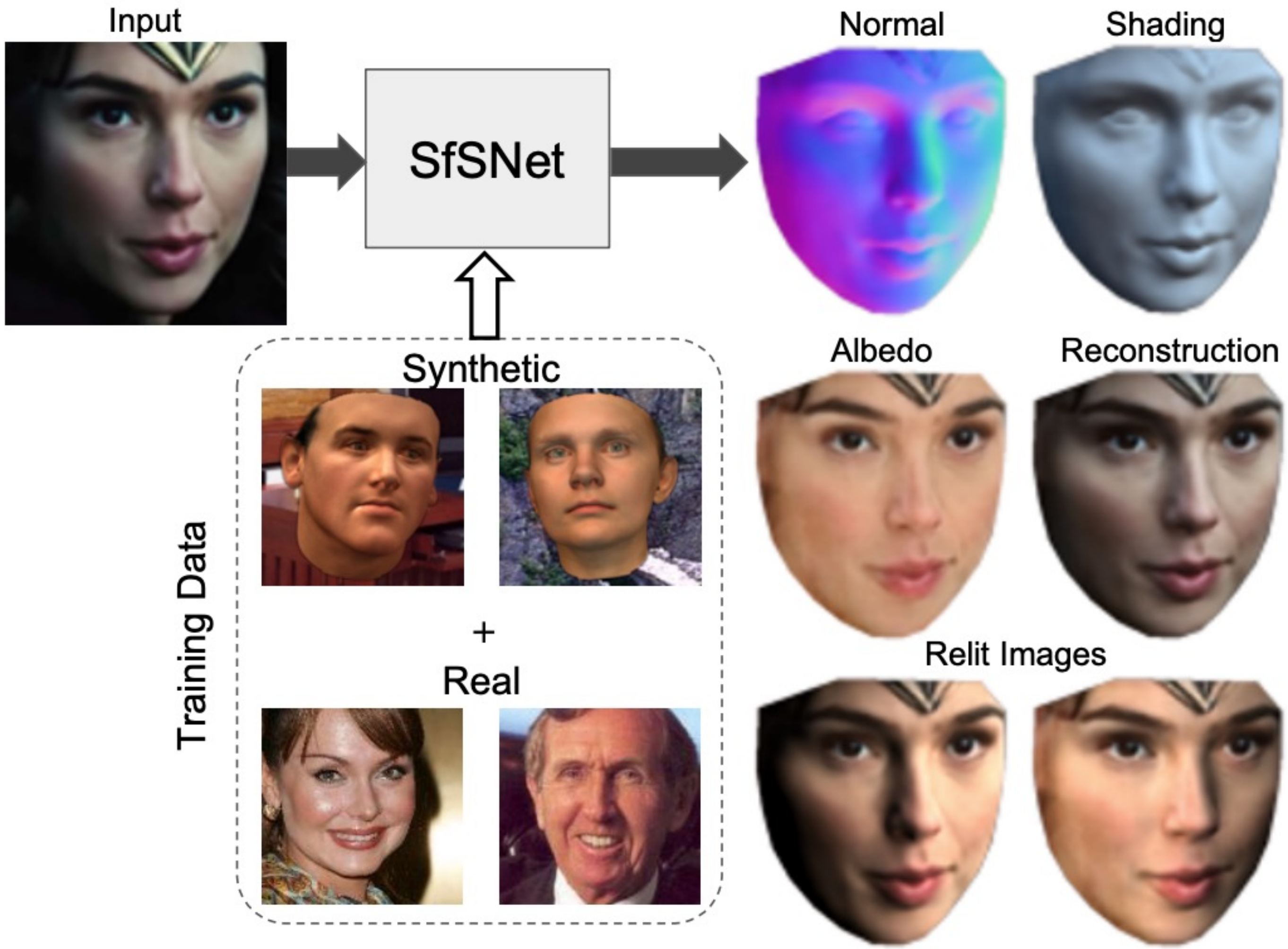
Suppose (for now) $k_d = 1$

$$\begin{aligned} I &= k_d \mathbf{N} \cdot \mathbf{L} \\ &= \mathbf{N} \cdot \mathbf{L} \\ &= cos\ \theta_i \end{aligned}$$

You can directly measure angle between normal and light source

- Not quite enough information to compute surface shape
- But can be if you add some additional info, for example
  - assume a few of the normals are known (e.g., along silhouette)
  - constraints on neighboring normals—"integrability"
  - smoothness
- Hard to get it to work well in practice
  - plus, how many real objects have constant albedo?
  - But, deep learning can help

Deep Learning for Shape from Shading

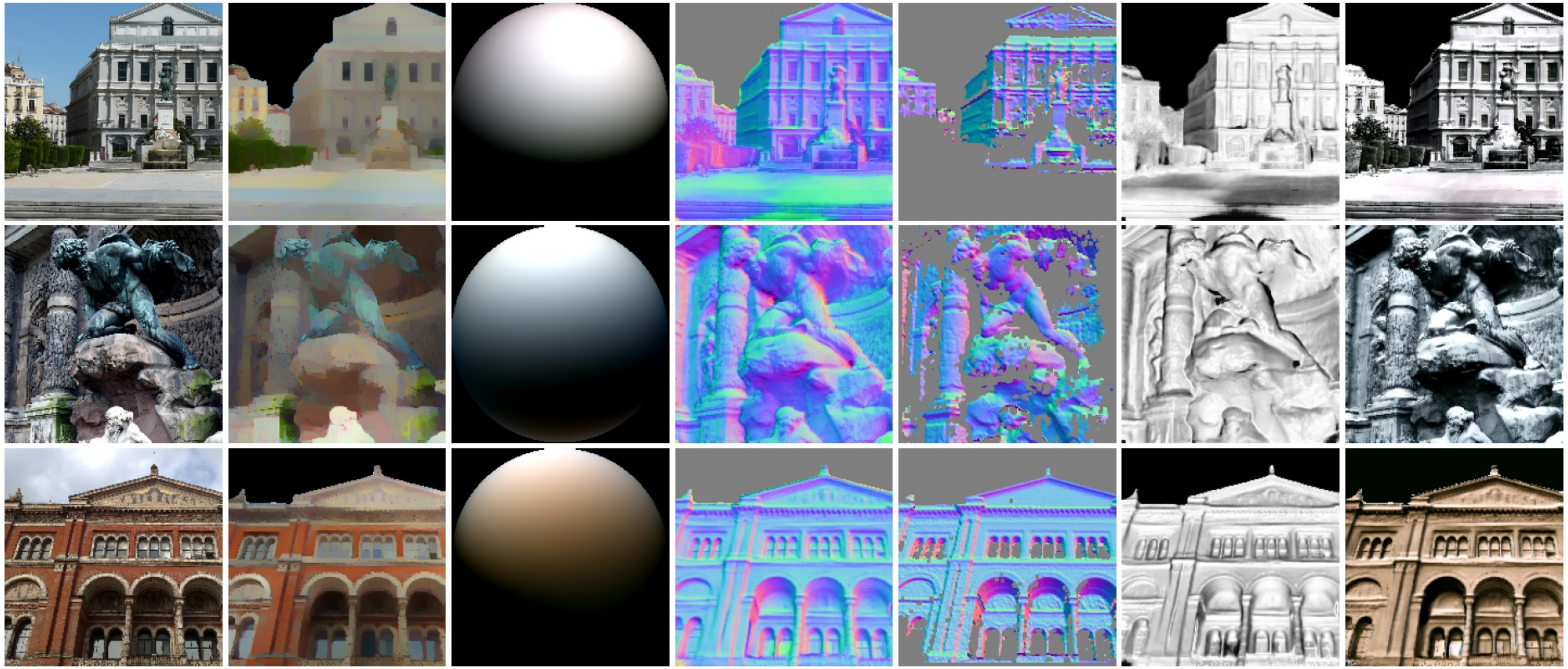


"SfSNet: Learning Shape, Reflectance and Illuminance of Faces in the Wild",
Sengupta, Kanazawa, Castillo, Jacobs, CVPR 2018.

InverseRenderNet: Learning single image inverse rendering

CVPR 2019

Ye Yu and William A. P. Smith
Department of Computer Science, University of York, UK
{yy1571,william.smith}@york.ac.uk



Input | Diffuse albedo | Illumination | NM prediction | NM from MVS | Frontal shading | Shading

Figure 1: From a single image (col. 1), we estimate albedo and normal maps and illumination (col. 2-4); comparison multi-view stereo result from several hundred images (col. 5); re-rendering of our shape with frontal/estimated lighting (col. 6-7).

Application: Detecting composite photos

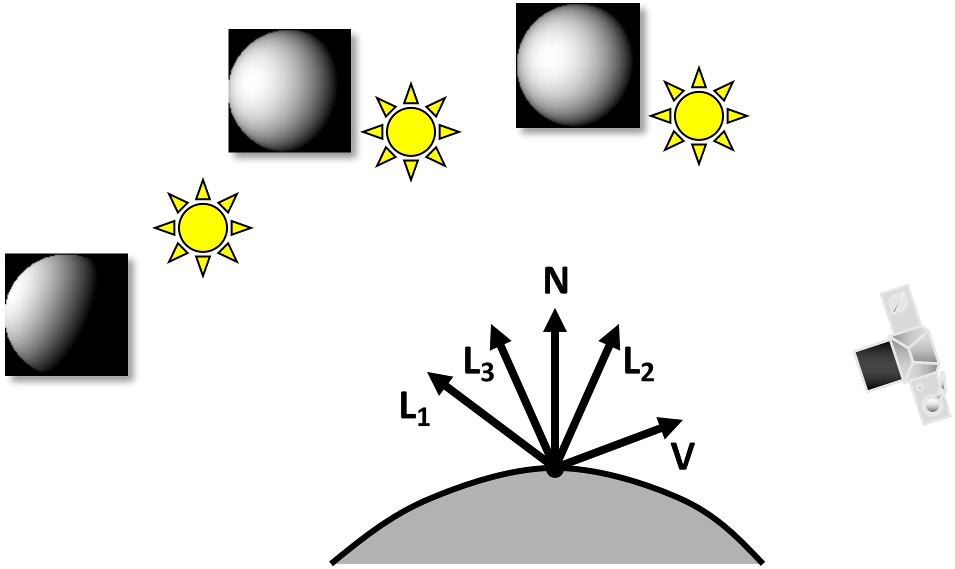
Fake photo

Real photo

Today’s class

- Measuring Light (recap)
- Image formation with shape, reflectance, and illumination
- Shape from Shading
- Photometric Stereo
- Uncalibrated Photometric Stereo
- Generalized Bas-Relief Ambiguity
- Photometric Stereo in ‘deep learning era’.

Photometric stereo



$$\begin{aligned} I_1 &= k_d \mathbf{N} \cdot \mathbf{L_1} \\ I_2 &= k_d \mathbf{N} \cdot \mathbf{L_2} \\ I_3 &= k_d \mathbf{N} \cdot \mathbf{L_3} \end{aligned}$$

Can write this as a matrix equation:

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = k_d \begin{bmatrix} \mathbf{L_1}^T \\ \mathbf{L_2}^T \\ \mathbf{L_3}^T \end{bmatrix} \mathbf{N}$$

Solving the equations

$$\underbrace{\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}}_{\substack{\mathbf{I} \\ 3 \times 1}} = \underbrace{\begin{bmatrix} \mathbf{L_1}^T \\ \mathbf{L_2}^T \\ \mathbf{L_3}^T \end{bmatrix}}_{\substack{\mathbf{L} \\ 3 \times 3}} \underbrace{k_d \mathbf{N}}_{\substack{\mathbf{G} \\ 3 \times 1}}$$

$$\mathbf{G} = \mathbf{L}^{-1}\mathbf{I}$$

$$k_d = \|\mathbf{G}\|$$

$$\mathbf{N} = \frac{1}{k_d}\mathbf{G}$$

Solve one such linear system **per pixel** to solve for that pixel's surface normal

More than three lights

Can get better results by using more than 3 lights

$$\begin{array}{ccc} \begin{bmatrix} I_1 \\ \vdots \\ I_n \end{bmatrix} = & \begin{bmatrix} \mathbf{L_1} \\ \vdots \\ \mathbf{L_n} \end{bmatrix} & k_d\mathbf{N} \\ \text{nx3} & \text{nx3} & \text{3x1} \end{array}$$

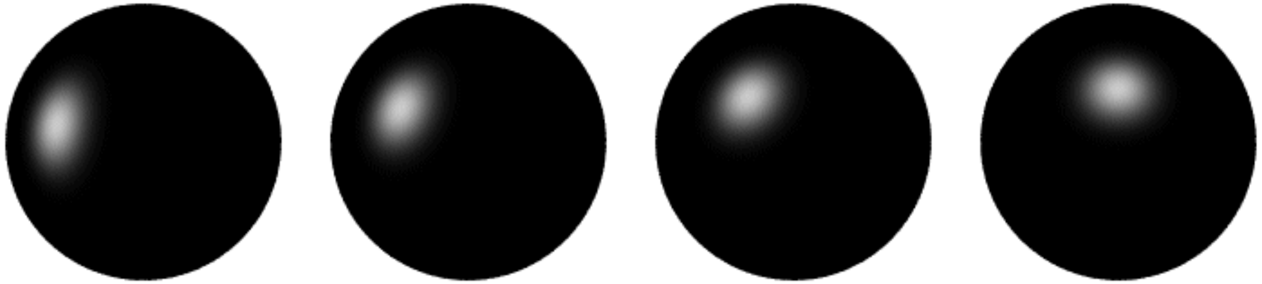
Least squares solution:

$$\begin{aligned} \mathbf{I} &= \mathbf{LG} \\ \mathbf{L^T I} &= \mathbf{L^T L G} \\ \mathbf{G} &= (\mathbf{L^T L})^{-1}(\mathbf{L^T I}) \end{aligned}$$

Solve for N, k_d as before

Calibrating Lighting Directions

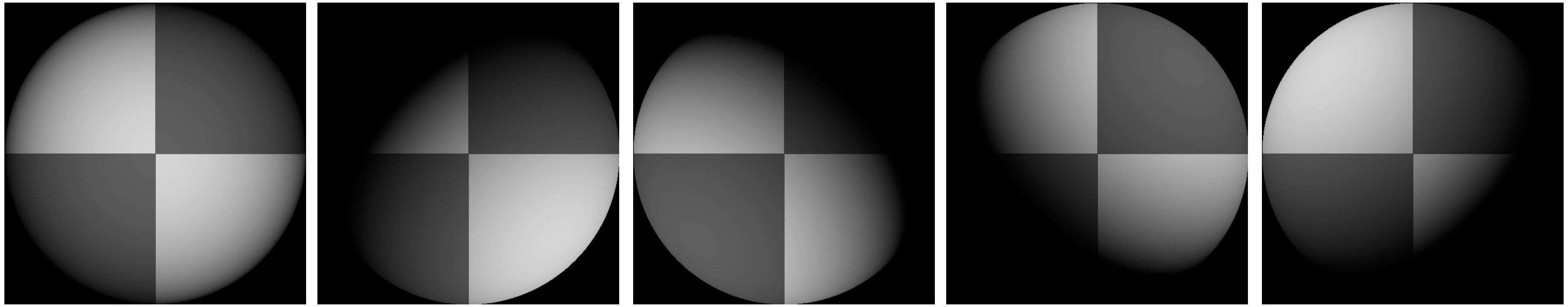
Trick: place a chrome sphere in the scene



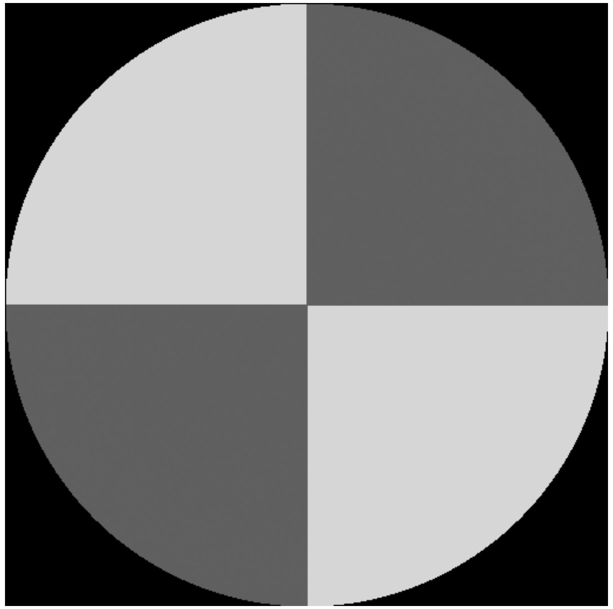
- the location of the highlight tells you where the light source is

Example

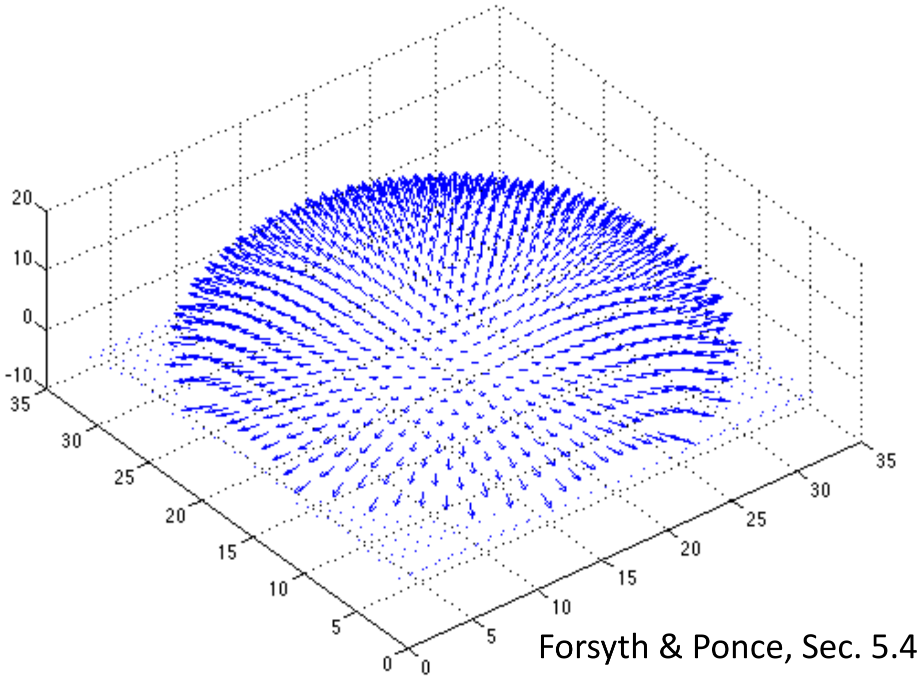
Input views



Recovered albedo



Recovered normal field



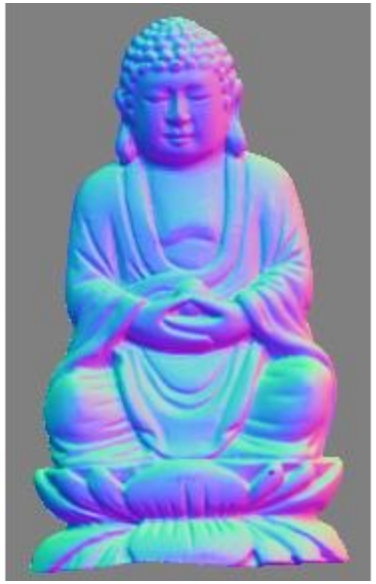
Forsyth & Ponce, Sec. 5.4

Depth from normals

- Solving the linear system per-pixel gives us an estimated surface normal for each pixel

- How can we compute depth from normals?
  - Normals are like the “derivative” of the true depth

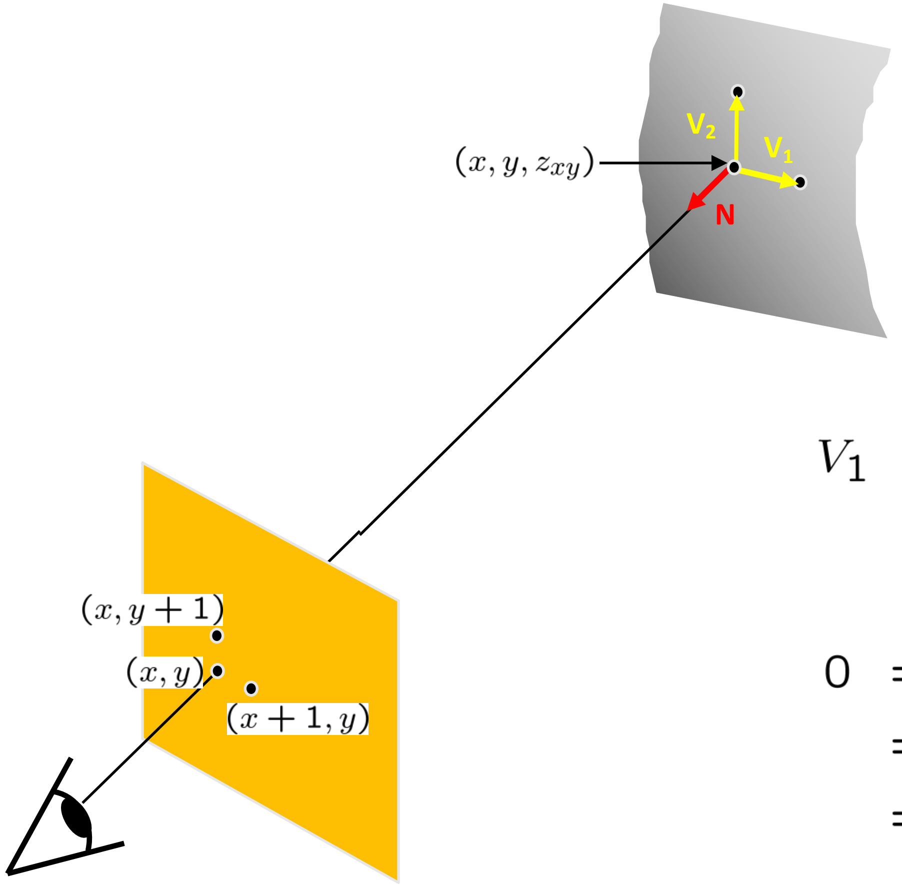
Input photo



Estimated normals

Estimated normals (needle diagram)

Depth from normals



$$\begin{aligned} V_1 &= (x+1, y, z_{x+1,y}) - (x, y, z_{xy}) \\ &= (1, 0, z_{x+1,y} - z_{xy}) \end{aligned}$$

$$\begin{aligned} 0 &= N \cdot V_1 \\ &= (n_x, n_y, n_z) \cdot (1, 0, z_{x+1,y} - z_{xy}) \\ &= n_x + n_z(z_{x+1,y} - z_{xy}) \end{aligned}$$

Get a similar equation for $\mathbf{V_2}$

- Each normal gives us two linear constraints on z
- compute z values by solving a matrix equation

Normal Integration

$$\nabla z = [p, q]^\top$$

where: → Linear Partial Differential Equations

$$\begin{cases} p = -\dfrac{n_1}{n_3} \\ q = -\dfrac{n_2}{n_3} \end{cases}$$

Integrability Constraint:

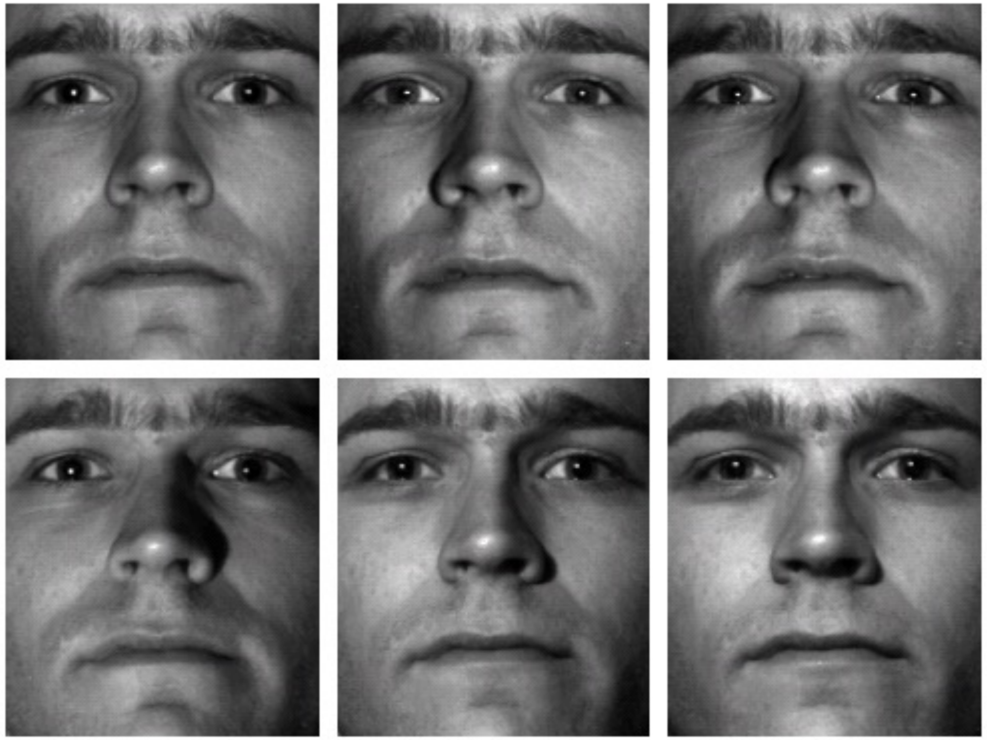
$$\partial_v p = \partial_u q$$

The order of taking 2nd order partial derivative with u & v (or x& y) shouldn't matter!

$$z(u, v) = z(u_0, v_0) + \int_{(r,s)=(u_0,v_0)}^{(u,v)} [p(r, s)\,\mathrm{d}r + q(r, s)\,\mathrm{d}s]$$

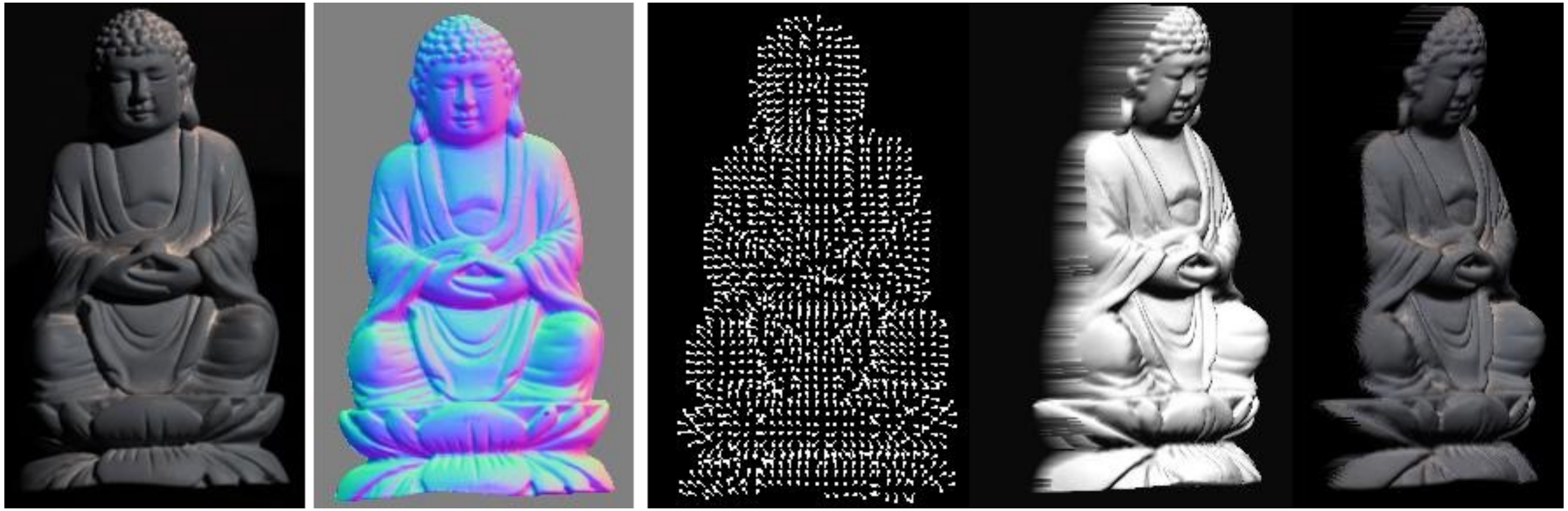
Read [Normal Integration: A Survey]() (if interested)

Results



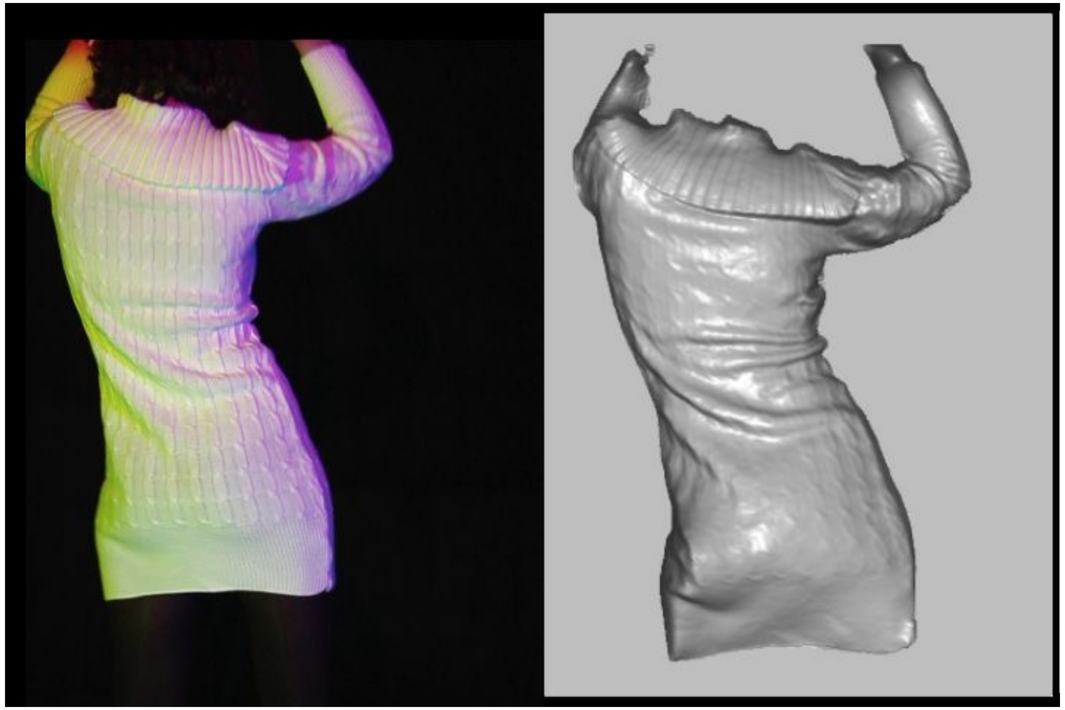
from Athos Georghiades

Results



Extension

- Photometric Stereo from Colored Lighting



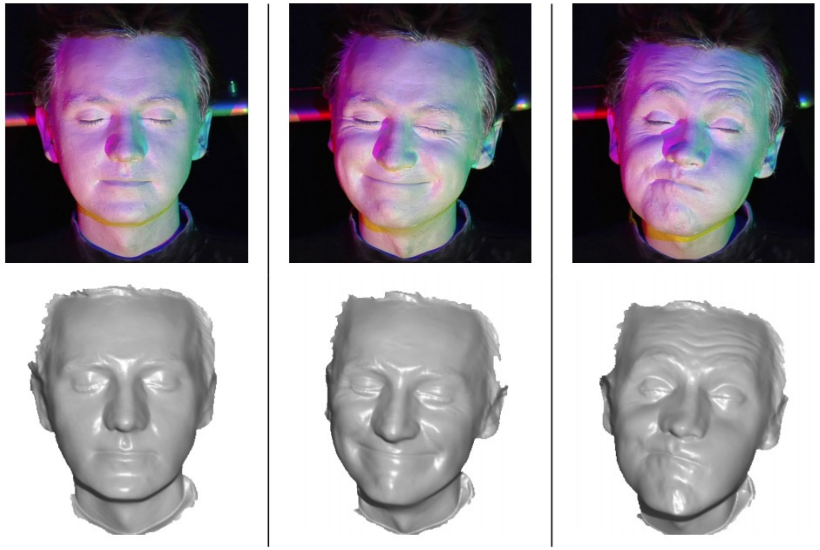


Fig. 2. **Applying the original algorithm to a face with white makeup.** Top: example input frames from video of an actor smiling and grimacing. Bottom: the resulting integrated surfaces.

Video Normals from Colored Lights
Gabriel J. Brostow, Carlos Hernández, George Vogiatzis, Björn Stenger, Roberto Cipolla
IEEE TPAMI, Vol. 33, No. 10, pages 2104-2114, October 2011.

Today’s class

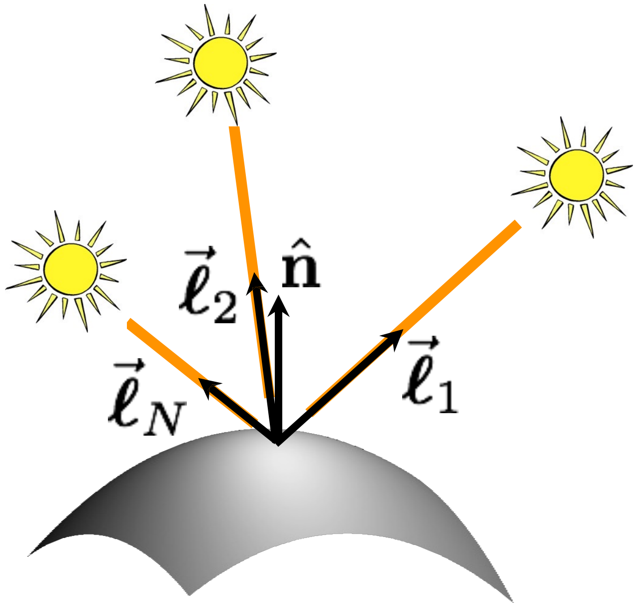
- Measuring Light (recap)
- Image formation with shape, reflectance, and illumination
- Shape from Shading
- Photometric Stereo
- Uncalibrated Photometric Stereo
- Generalized Bas-Relief Ambiguity
- Photometric Stereo in ‘deep learning era’.

What if the light directions are unknown?

a = albedo.

Previously k_d was used for albedo.

$$\boxed{\begin{aligned} I_1 &= a\hat{\boldsymbol{n}}^\top \vec{\boldsymbol{\ell}}_1 \\ I_2 &= a\hat{\boldsymbol{n}}^\top \vec{\boldsymbol{\ell}}_2 \\ &\vdots \\ I_N &= a\hat{\boldsymbol{n}}^\top \vec{\boldsymbol{\ell}}_N \end{aligned}}$$



define "pseudo-normal" $\vec{\boldsymbol{b}} \triangleq a\hat{\boldsymbol{n}}$

solve linear system for pseudo-normal

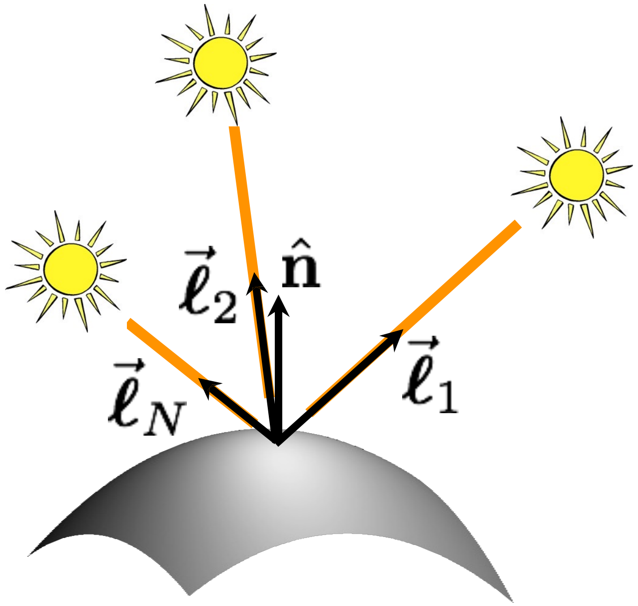
$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} \vec{\boldsymbol{\ell}}_1^\top \\ \vec{\boldsymbol{\ell}}_2^\top \\ \vdots \\ \vec{\boldsymbol{\ell}}_N^\top \end{bmatrix}_{N\times 3} \begin{bmatrix} \vec{\boldsymbol{b}} \end{bmatrix}_{3\times 1}$$

What if the light directions are unknown?

a = albedo.

Previously k_d was used for albedo.

$$\begin{aligned} I_1 &= a\hat{\boldsymbol{n}}^\top \vec{\boldsymbol{\ell}}_1 \\ I_2 &= a\hat{\boldsymbol{n}}^\top \vec{\boldsymbol{\ell}}_2 \\ &\vdots \\ I_N &= a\hat{\boldsymbol{n}}^\top \vec{\boldsymbol{\ell}}_N \end{aligned}$$



define “pseudo-normal” $\vec{\boldsymbol{b}} \triangleq a\hat{\boldsymbol{n}}$

solve linear system for pseudo-normal at each image pixel

$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix}_{N\times M} = \begin{bmatrix} \vec{\boldsymbol{\ell}}_1^\top \\ \vec{\boldsymbol{\ell}}_2^\top \\ \vdots \\ \vec{\boldsymbol{\ell}}_N^\top \end{bmatrix}_{N\times 3} \begin{bmatrix} B \end{bmatrix}_{3\times M}$$

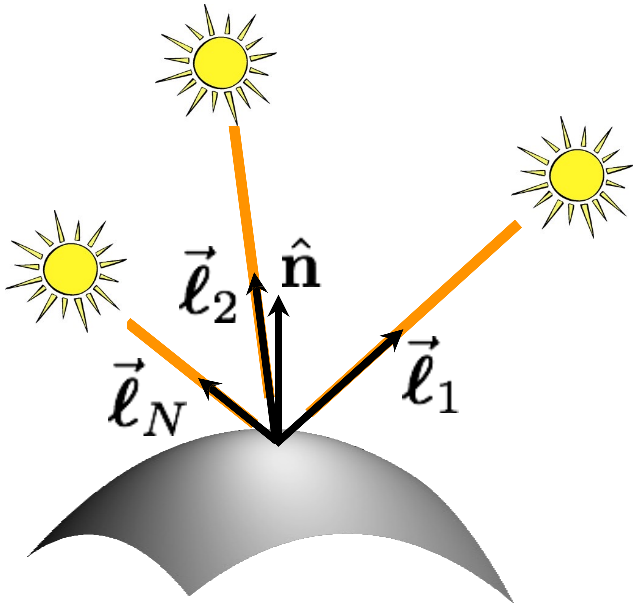
M: number of pixels

What if the light directions are unknown?

a = albedo.

Previously k_d was used for albedo.

$$
\begin{aligned}
I_1 &= a\hat{\boldsymbol{n}}^\top \vec{\boldsymbol{\ell}}_1 \\
I_2 &= a\hat{\boldsymbol{n}}^\top \vec{\boldsymbol{\ell}}_2 \\
&\vdots \\
I_N &= a\hat{\boldsymbol{n}}^\top \vec{\boldsymbol{\ell}}_N
\end{aligned}
$$



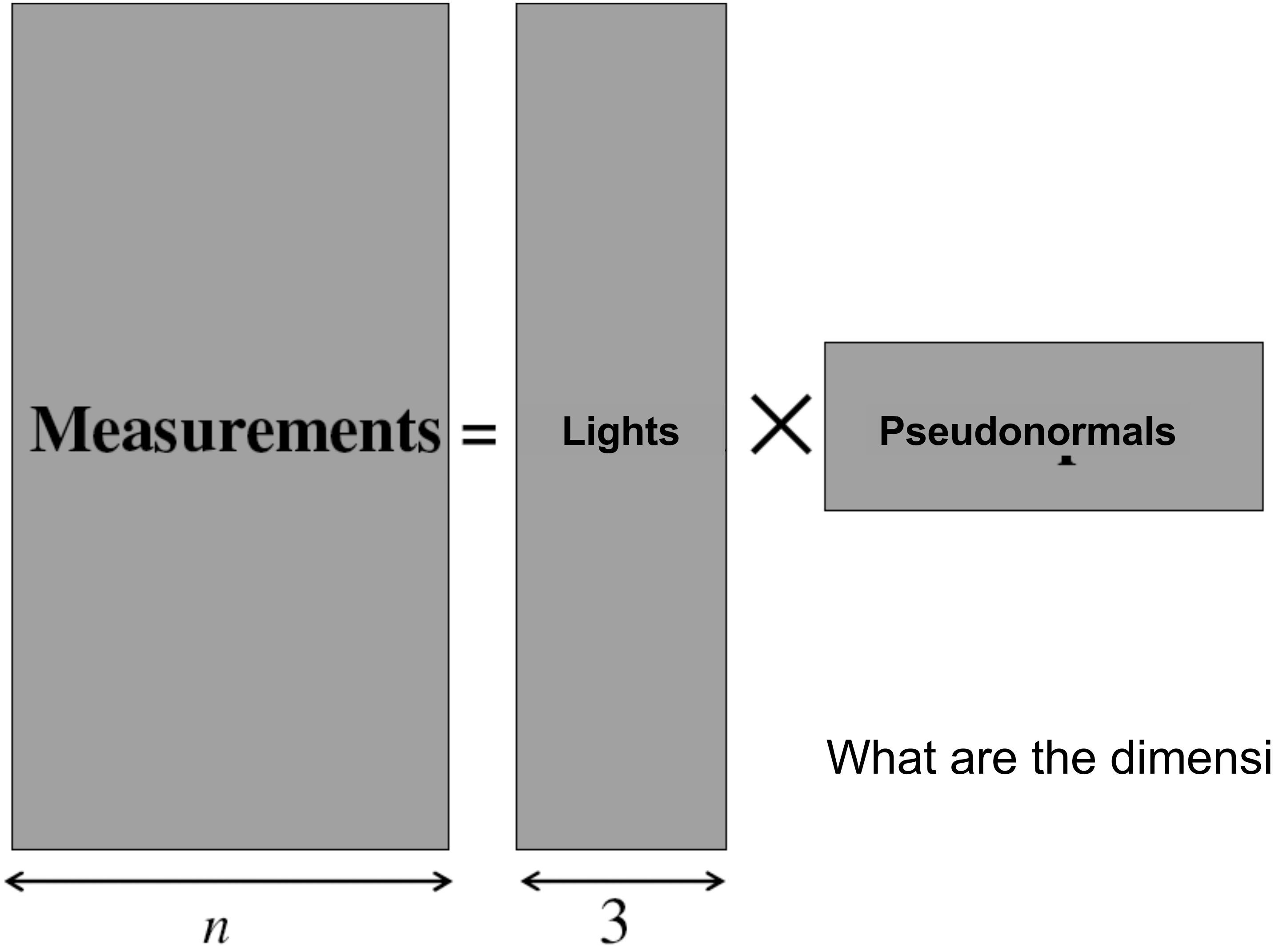
define "pseudo-normal" $\vec{\boldsymbol{b}} \triangleq a\hat{\boldsymbol{n}}$

solve linear system for pseudo-normal at each image pixel

$$
\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix}_{N\times M} = \begin{bmatrix} \vec{\ell}_1^\top \\ \vec{\ell}_2^\top \\ \vdots \\ \vec{\ell}_N^\top \end{bmatrix}_{N\times 3} \begin{bmatrix} B \end{bmatrix}_{3\times M}
$$

How do we solve this system without knowing light matrix L?

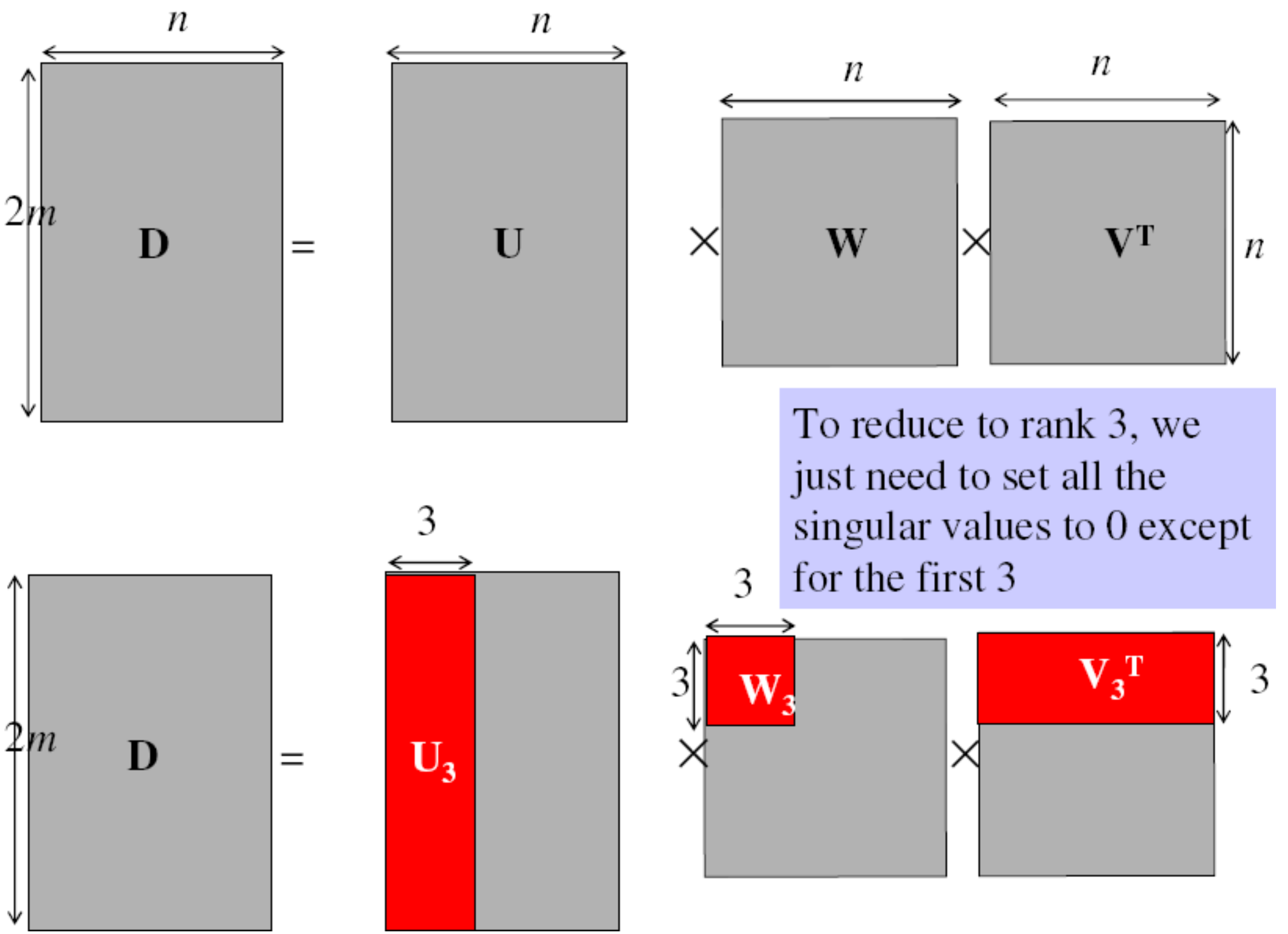
Factorizing the measurement matrix



What are the dimensions?

Factorizing the measurement matrix

- Singular value decomposition:



To reduce to rank 3, we just need to set all the singular values to 0 except for the first 3

This decomposition minimizes $|\mathbf{I}-\mathbf{LB}|^2$

Are the results unique?

We can insert any 3x3 matrix Q in the decomposition and get the same images:

$$\mathbf{I = L\,B = (L\,Q^{-1})\,(Q\,B)}$$

Can we use any assumptions to remove some of these 9 degrees of freedom?

Today’s class

- Measuring Light (recap)
- Image formation with shape, reflectance, and illumination
- Shape from Shading
- Photometric Stereo
- Uncalibrated Photometric Stereo
- Generalized Bas-Relief Ambiguity
- Photometric Stereo in ‘deep learning era’.

Generalized Bas-Relief ambiguity

We can insert any 3x3 matrix Q in the decomposition and get the same images:

$$\mathbf{I} = \mathbf{L}\,\mathbf{B} = (\mathbf{L}\,\mathbf{Q}^{-1})\,(\mathbf{Q}\,\mathbf{B})$$

Can we use any assumptions to remove some of these 9 degrees of freedom?

Generalized Bas-Relief ambiguity to rescue!

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{bmatrix}.$$

G has 3 degrees of freedom.

What does G mean?

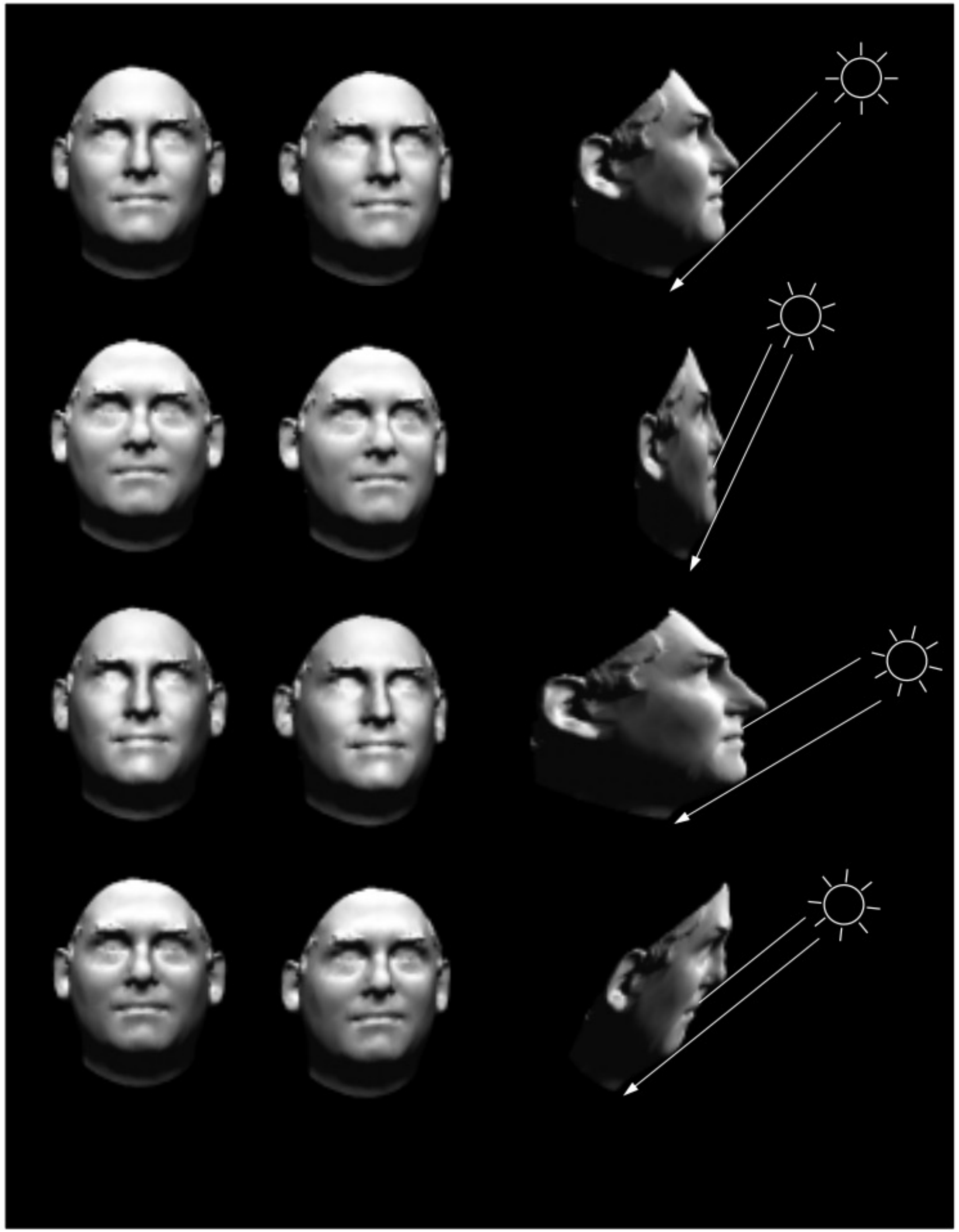
How do we obtain G? What constraints lead us to G?

Generalized Bas-Relief ambiguity

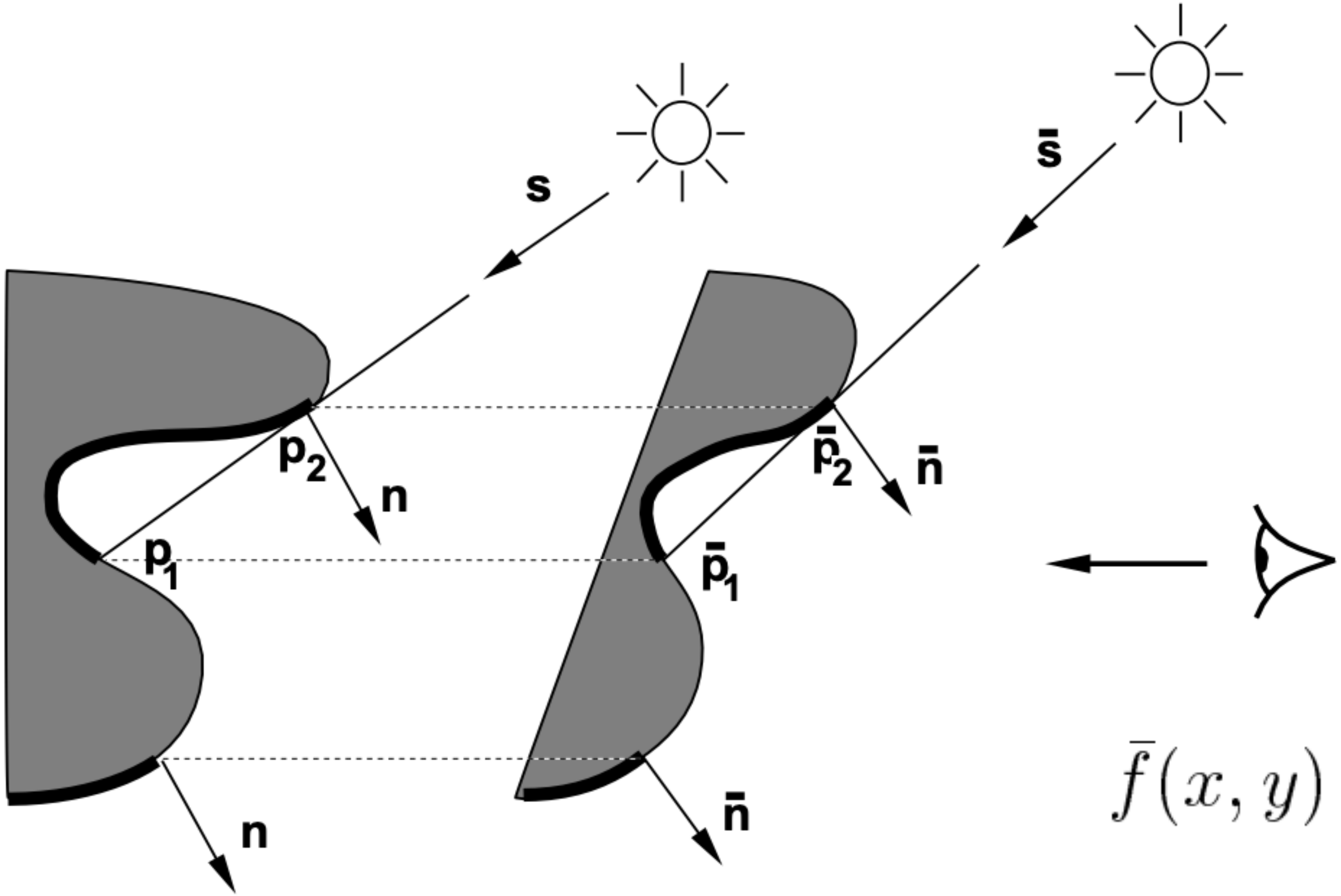
Artists have exploited GBR ambiguity in creating statues!

- On can flatten a surface and yet give an impression of full 3D to a viewer

"The Bas-Relief Ambiguity", Peter N. Belhumeur, David J. Kriegman, Alan L. Yuille, IJCV 99



Generalized Bas-Relief ambiguity



$$z = f(x, y)$$

$$\mathbf{n}(x, y) = \begin{bmatrix} -f_x \\ -f_y \\ 1 \end{bmatrix}$$

$$\bar{f}(x, y) = \lambda f(x, y) + \mu x + \nu y$$

Generalized Bas-Relief ambiguity

Note that if $\mathbf{p} = (x, y, f(x, y))$ and $\bar{\mathbf{p}} = (x, y, \bar{f}(x, y))$, then $\bar{\mathbf{p}} = G\mathbf{p}$ where

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{bmatrix}.$$

$$\bar{\mathbf{n}} = G^{-T}\mathbf{n}$$

$$G^{-1} = \frac{1}{\lambda} \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ -\mu & -\nu & 1 \end{bmatrix}.$$

Generalized Bas-Relief ambiguity

We can insert any 3x3 matrix Q in the decomposition and get the same images:

$$\mathbf{I} = \mathbf{L}\,\mathbf{B} = (\mathbf{L}\,\mathbf{Q}^{-1})\,(\mathbf{Q}\,\mathbf{B})$$

Can we use any assumptions to remove some of these 9 degrees of freedom?

Generalized Bas-Relief ambiguity to rescue!

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{bmatrix}.$$

G has 3 degrees of freedom.

G indicates integrable surface:
The order of taking 2^{nd} order partial derivative with u & v (or x& y) shouldn't matter!

Enforcing integrability

What does the integrability constraint correspond to?

- Differentiation order should not matter:

$$\frac{d}{dy}\frac{df(x,y)}{dx} = \frac{d}{dx}\frac{df(x,y)}{dy}$$

$$\mathbf{I} = \mathbf{L}\,\mathbf{B} = (\mathbf{L}\,\mathbf{Q}^{-1})\,(\mathbf{Q}\,\mathbf{B})$$

If B is integrable, then:

- $B' = G^{-T} \cdot B$ is also integrable for all G of the form ($\lambda \neq 0$)

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{bmatrix}$$

For now, ignore specular reflection

Slides from Photometric Methods for 3D Modeling, Matsushita, Wilburn, Ben-Ezra

And Refraction…

Slides from Photometric Methods for 3D Modeling, Matsushita, Wilburn, Ben-Ezra

And Interreflections…

Slides from Photometric Methods for 3D Modeling, Matsushita, Wilburn, Ben-Ezra

And Subsurface Scattering…

Slides from Photometric Methods for 3D Modeling, Matsushita, Wilburn, Ben-Ezra

What assumptions have we made for all this?

- Lambertian BRDF
- Directional lighting
- Distant Lighting
- Orthographic camera
- No interreflections or scattering

Limitations

Bigger problems

- doesn't work for shiny things, semi-translucent things
- shadows, inter-reflections

Smaller problems

- camera and lights have to be distant
- calibration requirements
  - measure light source directions, intensities
  - camera response function

Newer work addresses some of these issues

Some pointers for further reading:

- Zickler, Belhumeur, and Kriegman, "*Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction*." IJCV, Vol. 49 No. 2/3, pp 215-227.
- Hertzmann & Seitz, "*Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs*." IEEE Trans. PAMI 2005

Today's class

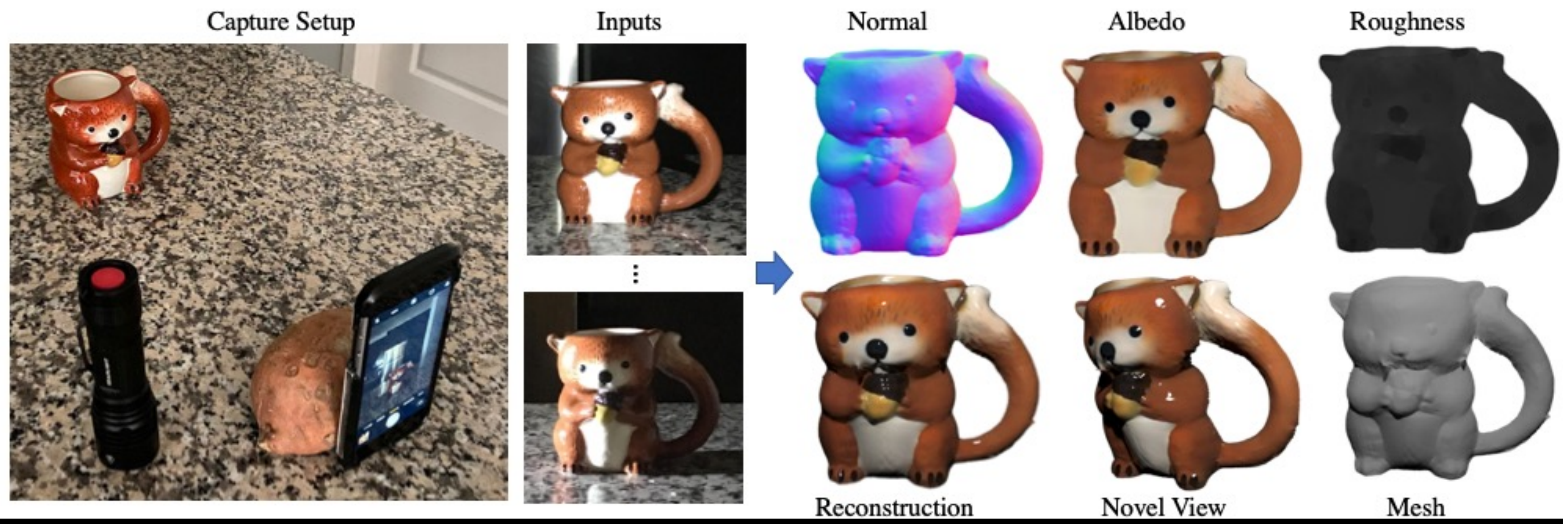
- Measuring Light (recap)
- Image formation with shape, reflectance, and illumination
- Shape from Shading
- Photometric Stereo
- Uncalibrated Photometric Stereo
- Generalized Bas-Relief Ambiguity
- Photometric Stereo in 'deep learning era'.

Photometric Stereo now … in Deep Learning era!

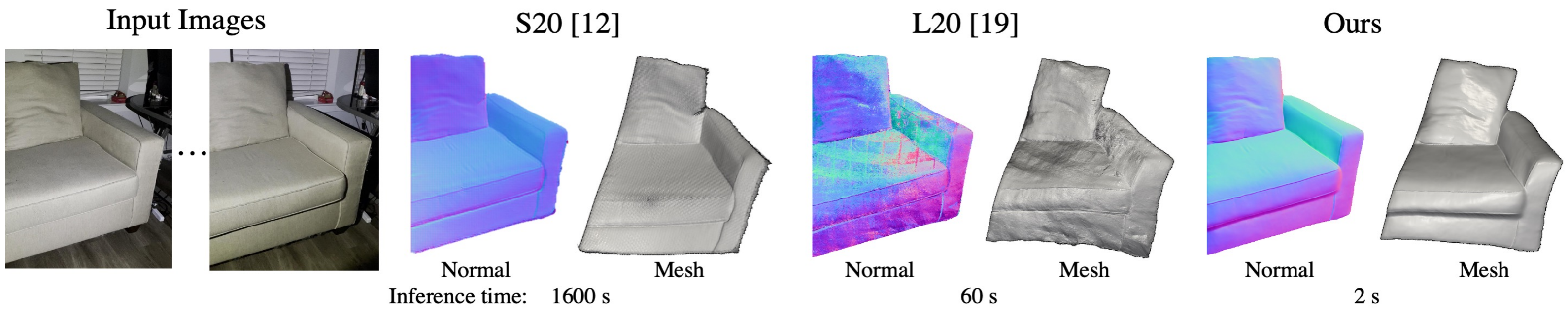
- Exploiting High-quality CG rendering for training data
- Designing deep neural network architectures
- Designing loss functions

- GBR ambiguity is still a problem! -> Flattened objects reconstructed.

Using lighting as a cue for 3D reconstruction (Photometric Stereo)



“Shape & Material Capture at Home”, Lichy, Wu, Sengupta, Jacobs, CVPR 2021



“Real-Time Light-Weight Near-Field Photometric Stereo”, Lichy, Sengupta, Jacobs, CVPR 2022

Captured Images: Right

Single iPhone Image with Built-In Flash

Image 1/1

Mesh

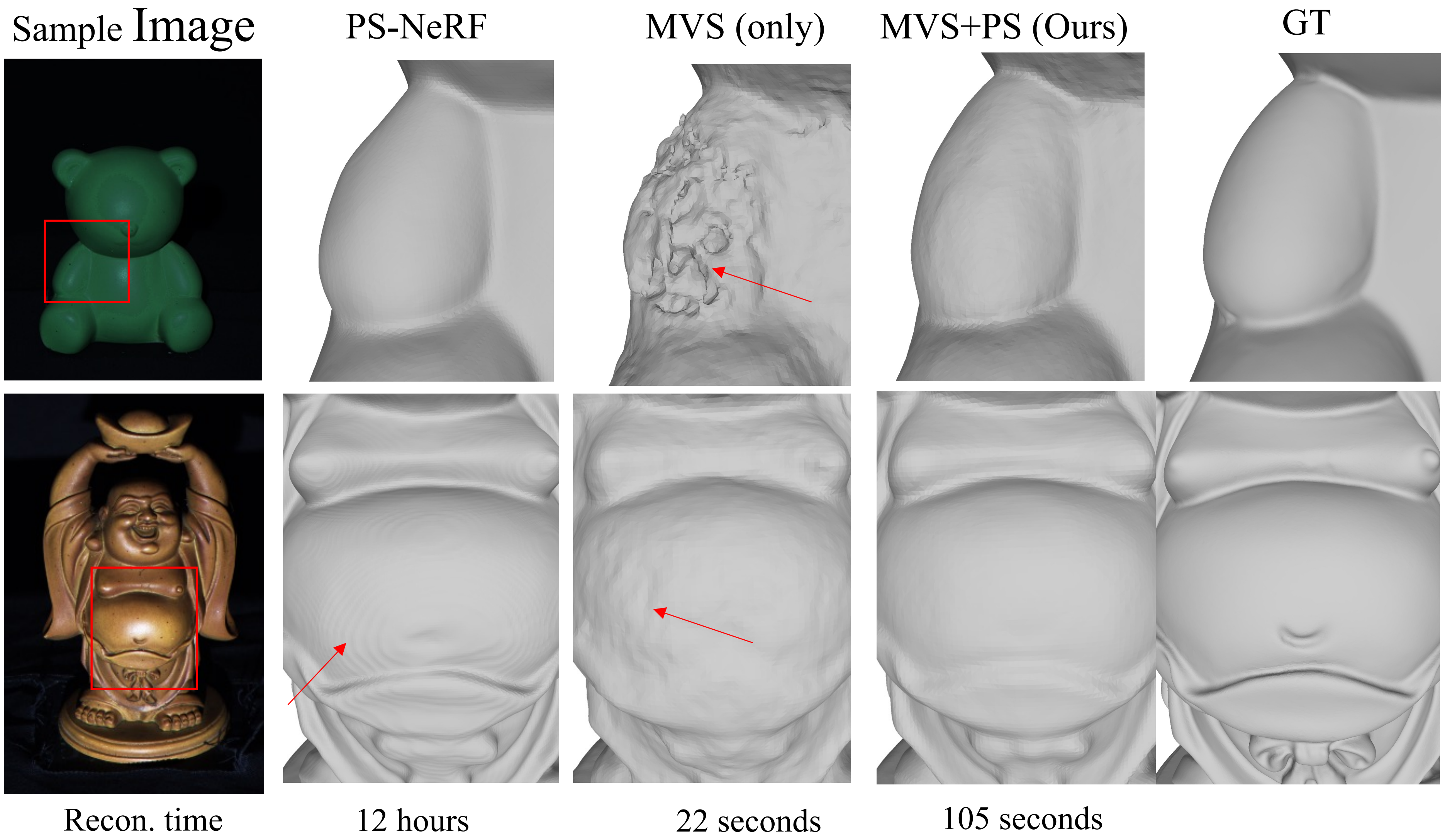
Photometric Stereo + SLAM for colon reconstruction in colonoscopy

SLAM only

Photometric Stereo +
SLAM (Ours)

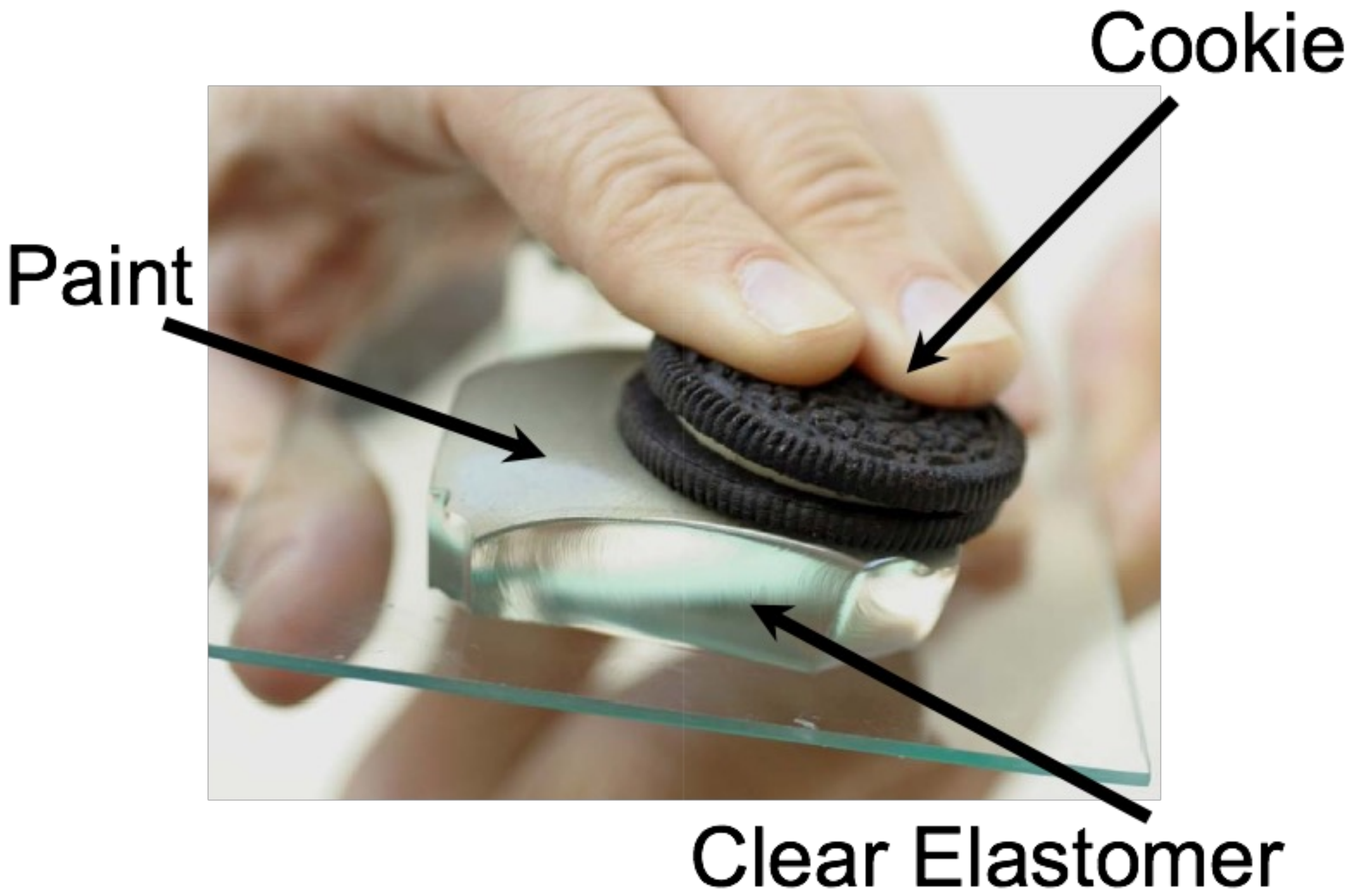
“A Surface-normal Based Neural Framework for Colonoscopy Reconstruction”, Sherry Wang, Yubo Zhang, Sarah McGill, Julian Rosenman, Jan-Michael Frahm, Soumyadip Sengupta, Steve Pizer, IPMI 2023.

Photometric Stereo + Multi-view Stereo for fast 3D reconstruction



"MVPSNet: Fast Generalizable Multi-view Photometric Stereo", Dongxu Zhao, Daniel Lichy, Pierre-Nicolas Perrin, Jan-Michael Frahm, Soumyadip Sengupta, in submission.

Johnson and Adelson, 2009

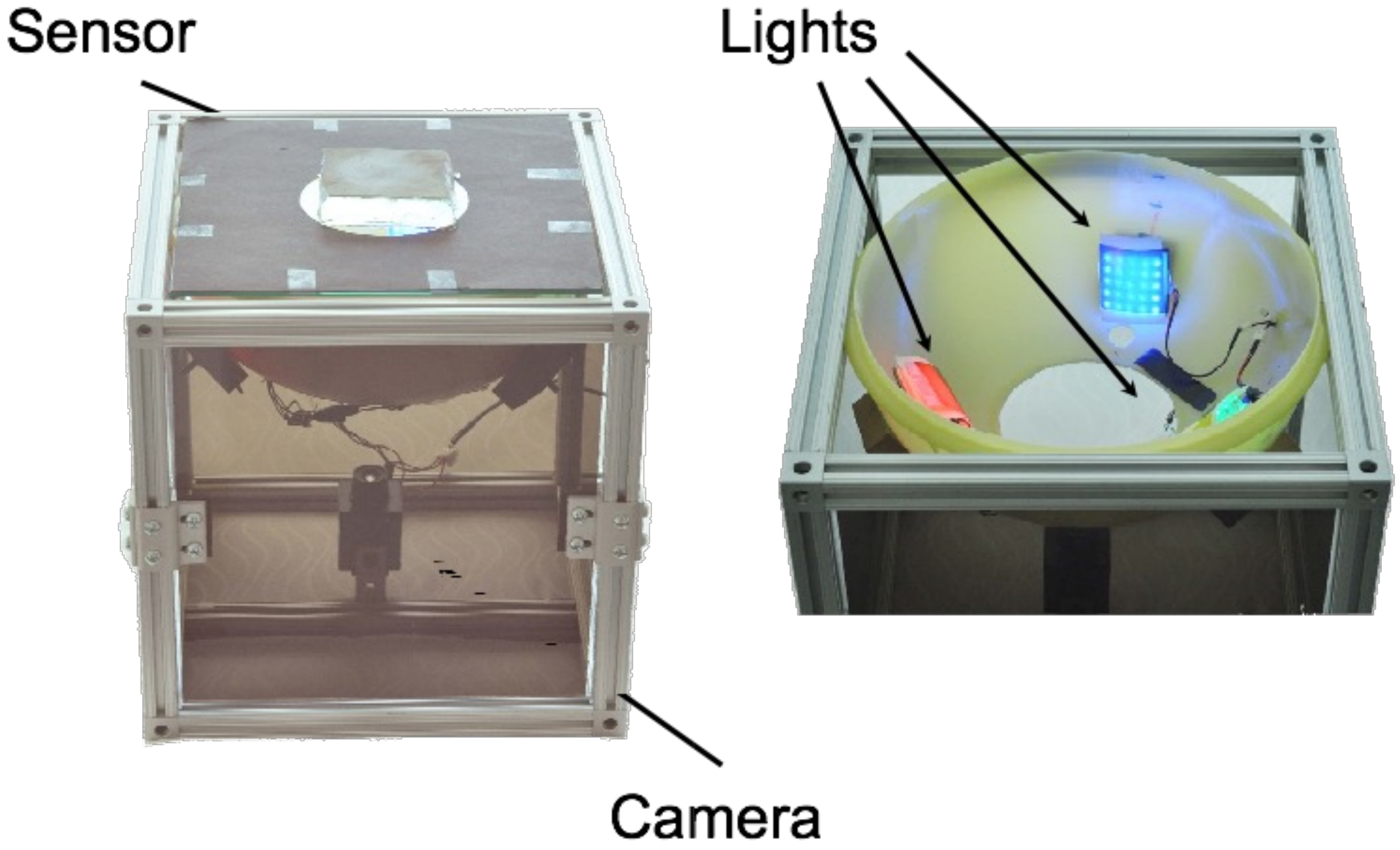


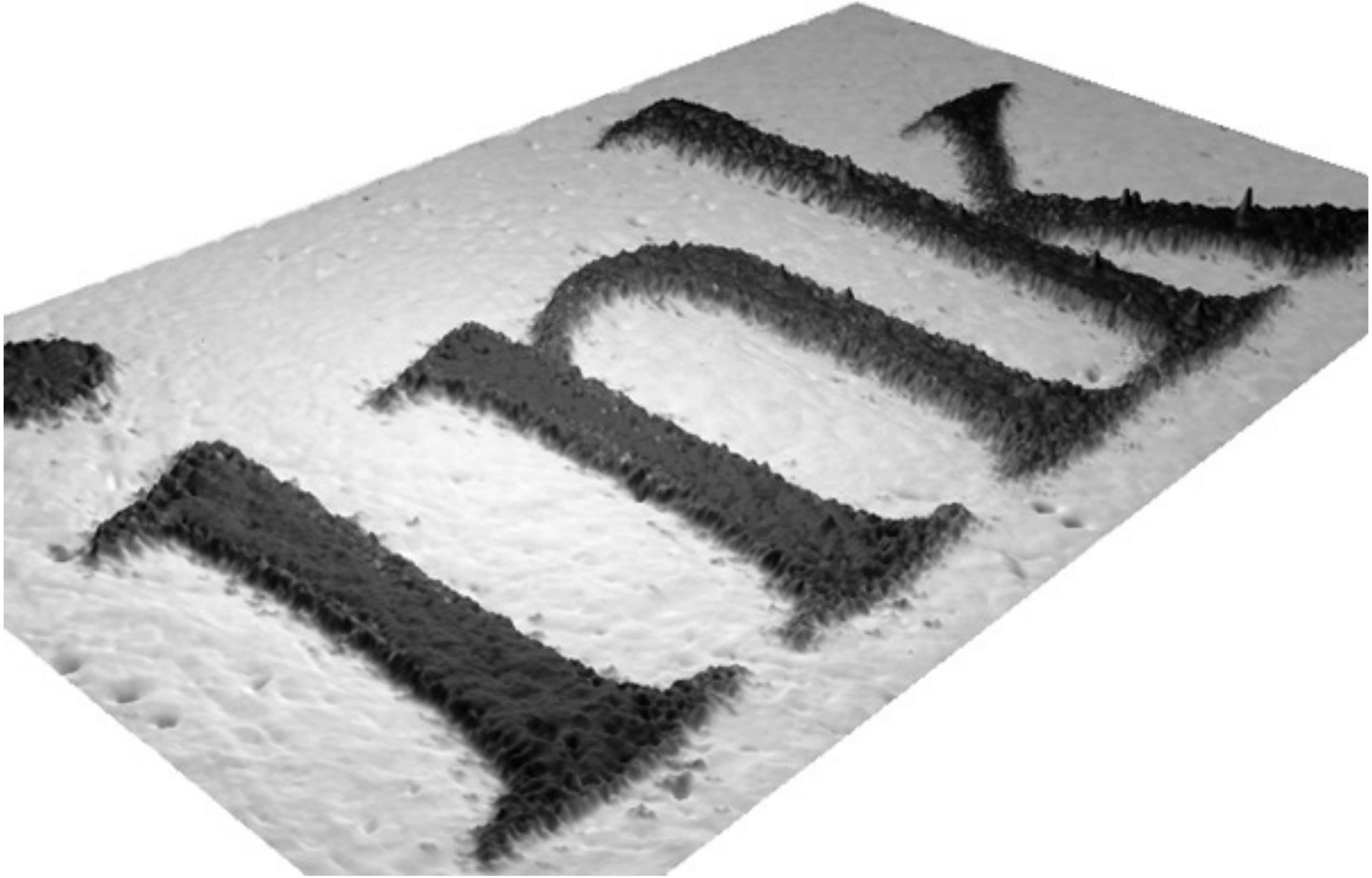
Johnson and Adelson, 2009

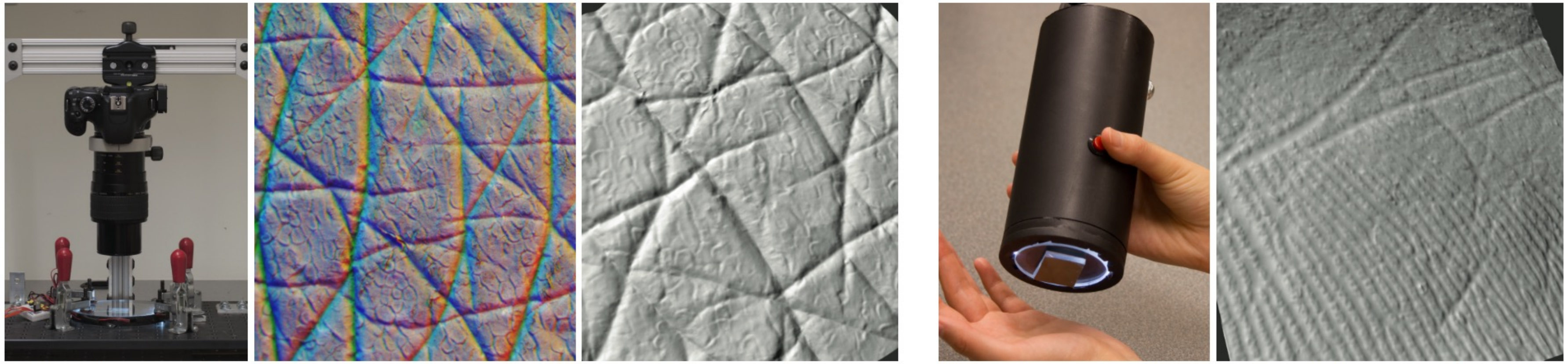

OREO


OREO

Lights, camera, action




ink



(a) bench configuration (b) captured (c) reconstruction (d) portable configuration (e) reconstruction

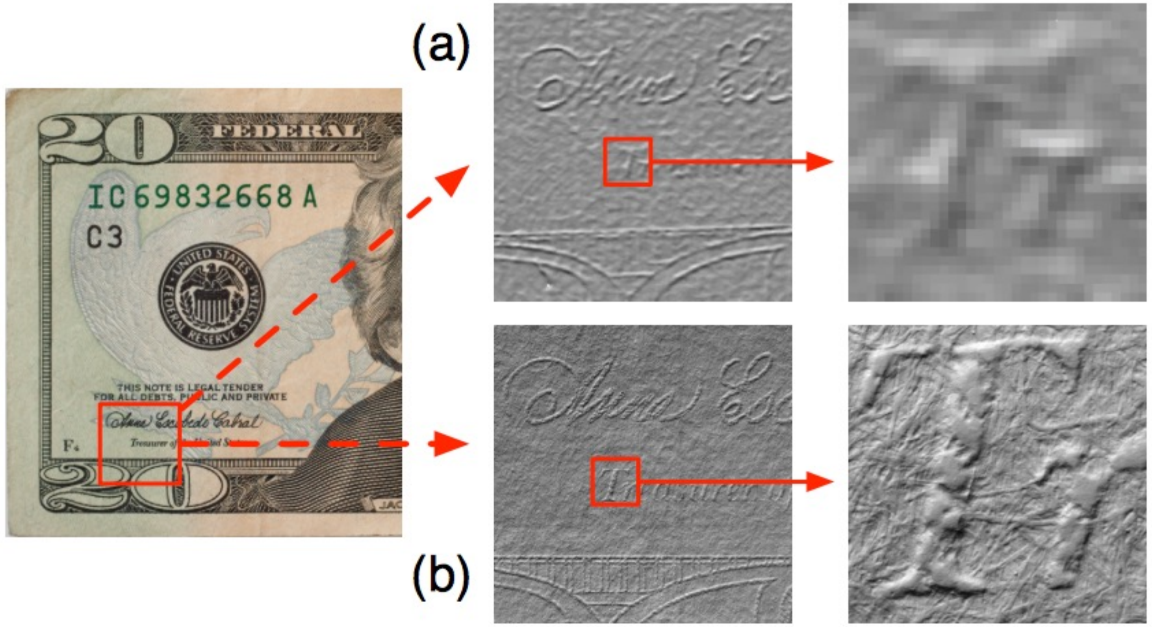


Figure 7: *Comparison with the high-resolution result from the original retrographic sensor. (a) Rendering of the high-resolution \$20 bill example from the original retrographic sensor with a close-up view. (b) Rendering of the captured geometry using our method.*

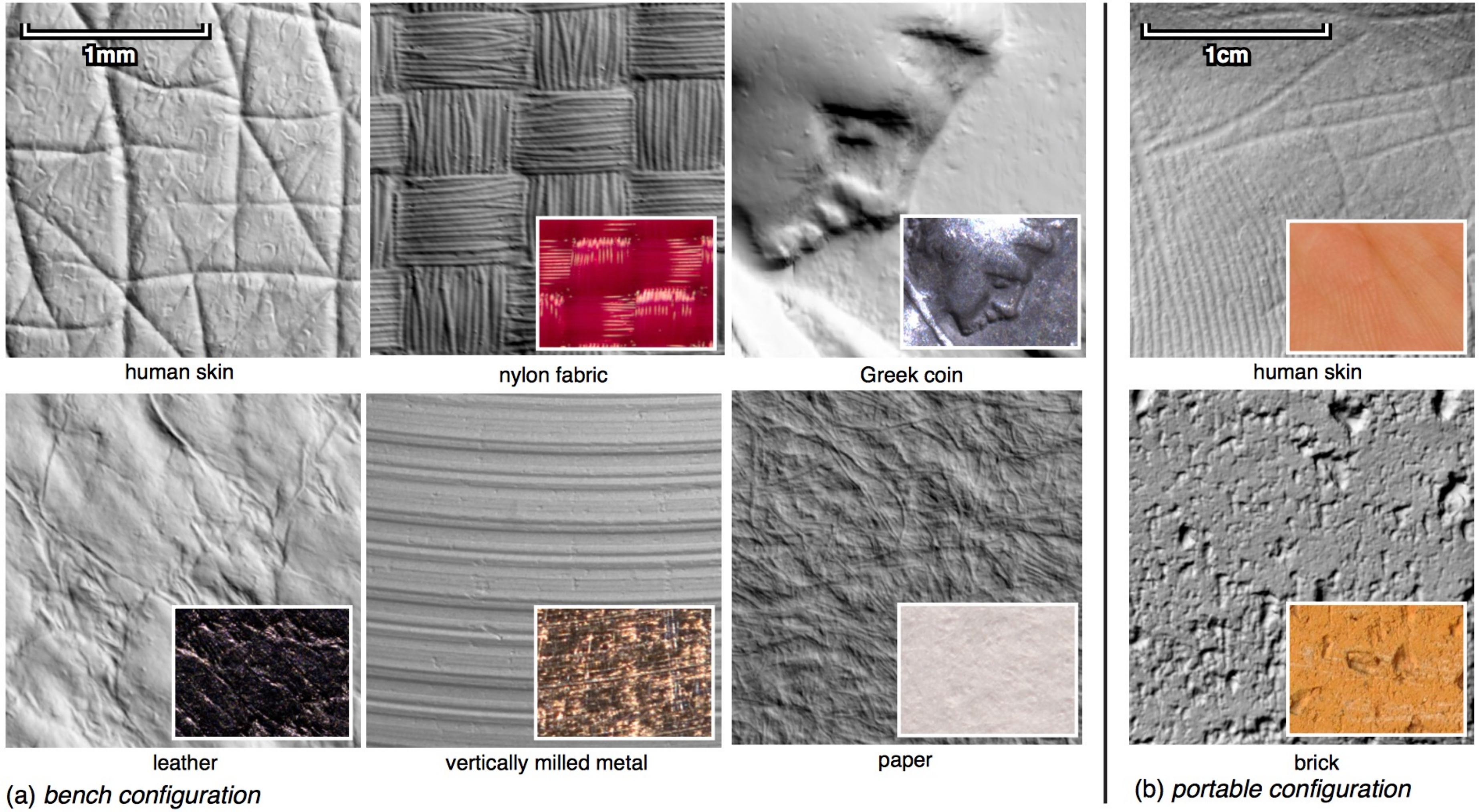
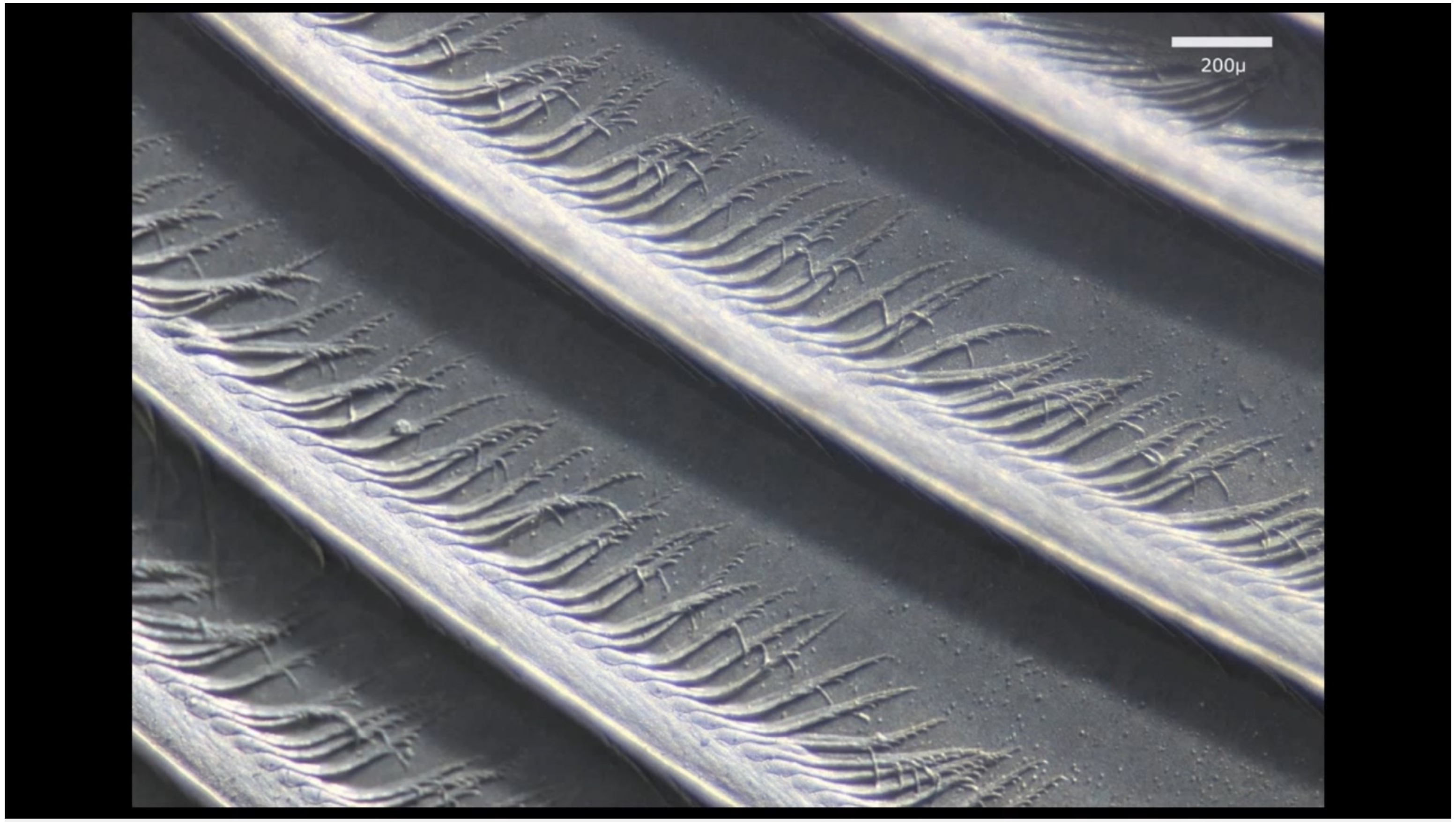


Figure 9: *Example geometry measured with the bench and portable configurations. Outer image: rendering under direct lighting. Inset: macro photograph of original sample. Scale shown in upper left. Color images are shown for context and are to similar, but not exact scale.*



Sensing Surfaces with GelSight

kimoatmit

Subscribe 64

138,850 views

https://www.youtube.com/watch?v=S7gXih4XS7A