ECE5554 – Computer Vision
Lecture 11a – Shape from Shading

Creed Jones, PhD

Course Update

- HW5 is due TONIGHT at 11:59 PM

- Final Exam
  - THIS Thursday, August 11
    - Sized to be a "one-hour" exam, but you will have two full hours (if you start on time!)
  - Final exam will cover all material, not just what the HW assignments involved
    - Mostly T/F and multiple choice, a few longer answer questions
    - May be some math, NO CODING
  - Open book and notes, no help from others
    - If you resort to help from online resources, you may be led astray

- SPOT Surveys

Today's Objectives

Shape via "passive" methods

- Shape from Shading
  - Lambertian reflection
  - Equations for calculating surface normal
  - Depth from normal
  - Photometric stereo
- Stereo – Focus – Texture

Shape via "active" methods

- Laser profiling – Projection – Structured light - LIDAR

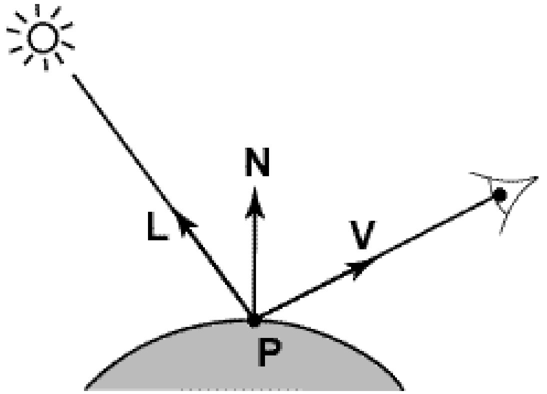
- Shape from Motion

First, a few terms related to light and its reflection

- Specular – “related to or having the properties of a mirror”
  - Light that is reflected at the complementary angle of the angle of incidence, with respect to the normal to the surface
- Diffuse – “spread over a large area; not concentrated”
  - Light that is reflected in directions other than the complementary angle
- Albedo – “the proportion of incident light that is reflected by a surface”
  - Fresh snow: albedo=0.95; asphalt: albedo=0.08
- Lambertian – “a surface that appears uniformly bright in all directions”
  - Examples: a ping-pong ball, a flat white wall, most white fabrics…

SHAPE FROM SHADING

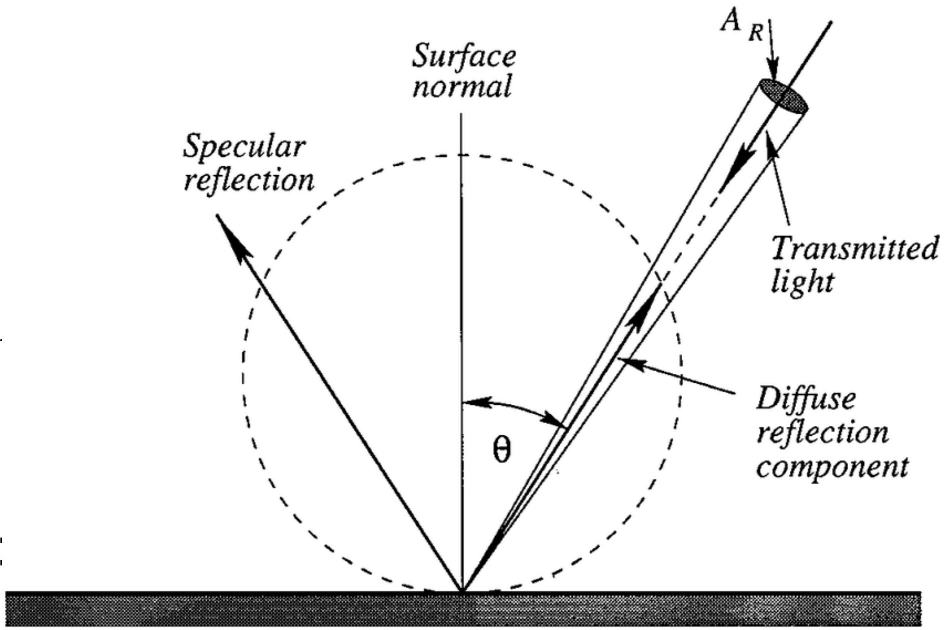
In diffuse reflection (not from shiny surfaces) the intensity of light reflected is governed by Lambert's law



$$R_e = k_d \mathbf{N} \cdot \mathbf{L} R_i$$

image intensity of **P**

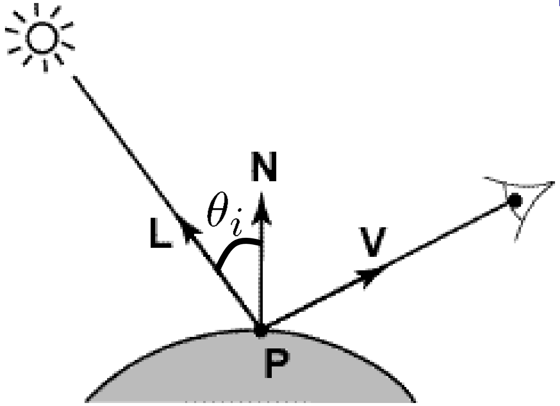
$$I = k_d \boldsymbol{N} \cdot \boldsymbol{L} = k_d \|\boldsymbol{N}\| \|\boldsymbol{L}\| \cos\theta$$



Simplifying assumptions

- $I = R_e$: camera response function f is the identity function:
  - can always achieve this in practice by solving for f and applying f^{-1} to each pixel in the image
- $R_i = 1$: light source intensity is 1
  - can achieve this by dividing each pixel in the image by R_i

To infer shape from shading, we can look at differences in measured brightness and deduce the angle of the normal vector



Lambertian reflection says that

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

Suppose $k_d = 1$:

$$I = \boldsymbol{N} \cdot \boldsymbol{L}$$

$$I \propto \cos \theta_i$$

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
- Its not easy to get it to work well in practice
  - plus, how many real objects have constant albedo?

Shape from Shading – start with the equation defining the surface (assume we are looking down on the scene)

- Assume surface is given by the height function of two dimensions $Z(x,y)$

- Let $p = \frac{\partial Z}{\partial x},\quad q = \frac{\partial Z}{\partial y}$ (note, these are surface gradients, not intensity gradients)

- So, the surface normal is $\mathbf{n} = \frac{1}{\sqrt{1+p^2+q^2}}\begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix}$; call the illumination vector $\boldsymbol{l} = \begin{pmatrix} l_x \\ l_y \\ l_z \end{pmatrix}$

- The image intensity at any pixel location is then $I = \frac{L\rho}{\sqrt{1+p^2+q^2}}\begin{pmatrix} l_x & l_y & l_z \end{pmatrix}\begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix}$

- Discretize: end up with one equation per pixel
- But this is *p* equations in 2*p* unknowns…

Shape from Shading – the governing equations

- $I(u,v) = \frac{L\rho}{\sqrt{1+p^2+q^2}} \begin{pmatrix} l_x & l_y & l_z \end{pmatrix} \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix}$, where $p = \frac{\partial Z}{\partial x}, \quad q = \frac{\partial Z}{\partial y}$
- ρ is a constant related to albedo (surface reflectance)

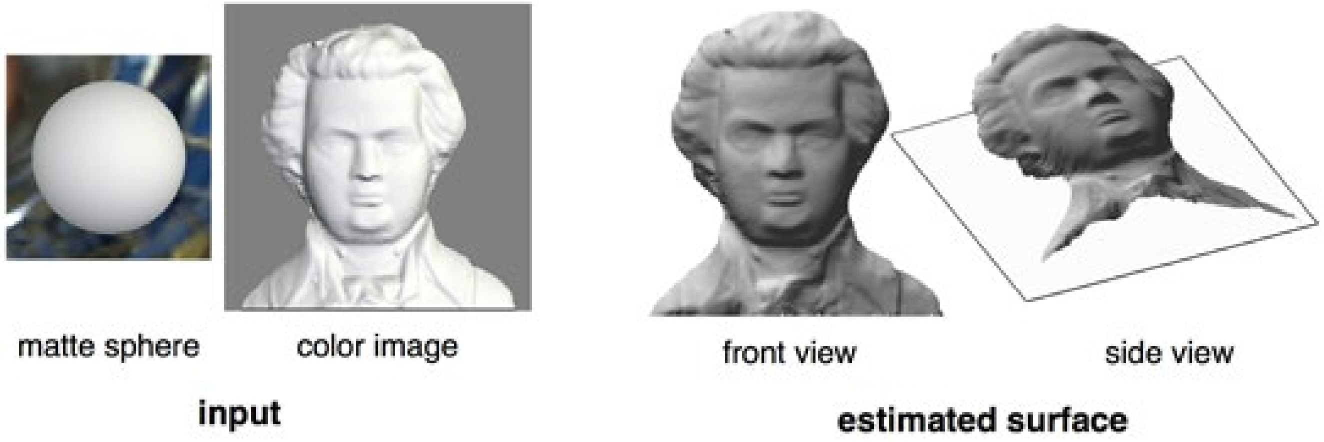
- Apply the integrability constraint (smoothness): $\frac{\partial^2 Z}{\partial x \partial y} = \frac{\partial^2 Z}{\partial y \partial x} \quad \Rightarrow \quad \frac{\partial p}{\partial y} = \frac{\partial q}{\partial x}$

- Wind up with system of 2*p* (nonlinear) differential equations
- <u>No solution in presence of noise or depth discontinuities</u>

Using these equations, 3D shape can be calculated from the shading observed on an object, by making some assumptions about the optical scenario

We make three assumptions:

1. The light source is far away, and close to a point source
2. Lambertian reflectance (light is reflected diffusely, with intensity proportional to the cosine of the angle between the surface normal and the illumination)
3. Uniform albedo – the surface doesn't vary in "brightness"



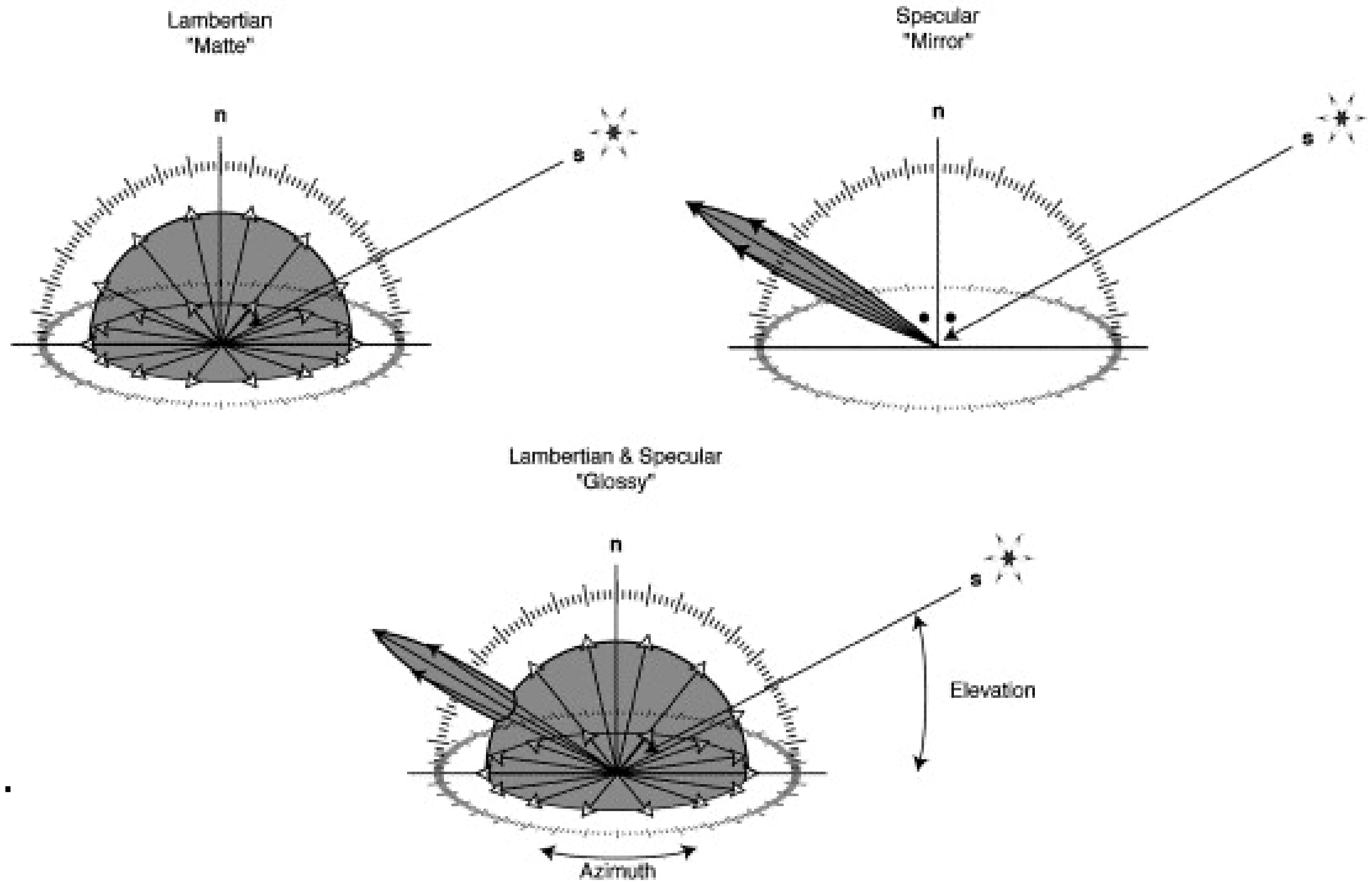
B. K. P. Horn, 1989. Obtaining shape from shading information. In B. K. P. Horn and M. J. Brooks, eds., Shape from Shading. MIT Press.

With our assumptions, the orientation of the surface can be found from the intensity of the image

In more complex cases, iterative approximations can be found for:

- specular surfaces
- more complex illumination
- variations in object brightness

- The math is significant…



Estimating Illumination and Albedo

- Need to know surface reflectance and Illumination brightness and direction
- In general, can’t compute from single image
- Certain assumptions permit estimating these
  - Assume uniform distribution of normals, look at distribution of intensities in image
  - Insert known reference object into image
  - Slightly specular object: estimate lighting from specular highlights, then discard pixels in highlights

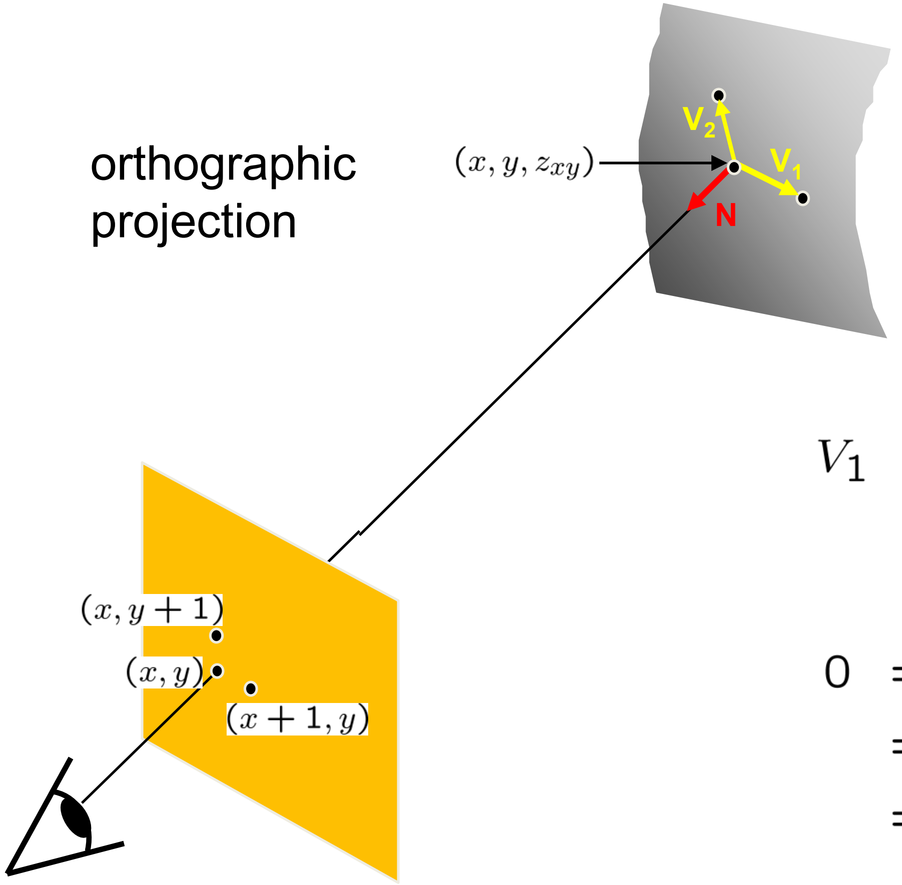
To determine the shape of an object, it's only necessary to be able to calculate the normal vector to its surface at any point

If the object's reflectance is uniform, since the intensity reflected is related to the angle of observation

$$I(x,y) = \frac{1}{\sqrt{\frac{\partial z}{\partial x}^2 + \frac{\partial z}{\partial y}^2}} \begin{bmatrix} -\frac{\partial z}{\partial x} \\ -\frac{\partial z}{\partial y} \end{bmatrix}$$

For non-uniform reflectances (but still known), the calculation is similar

Depth from normal

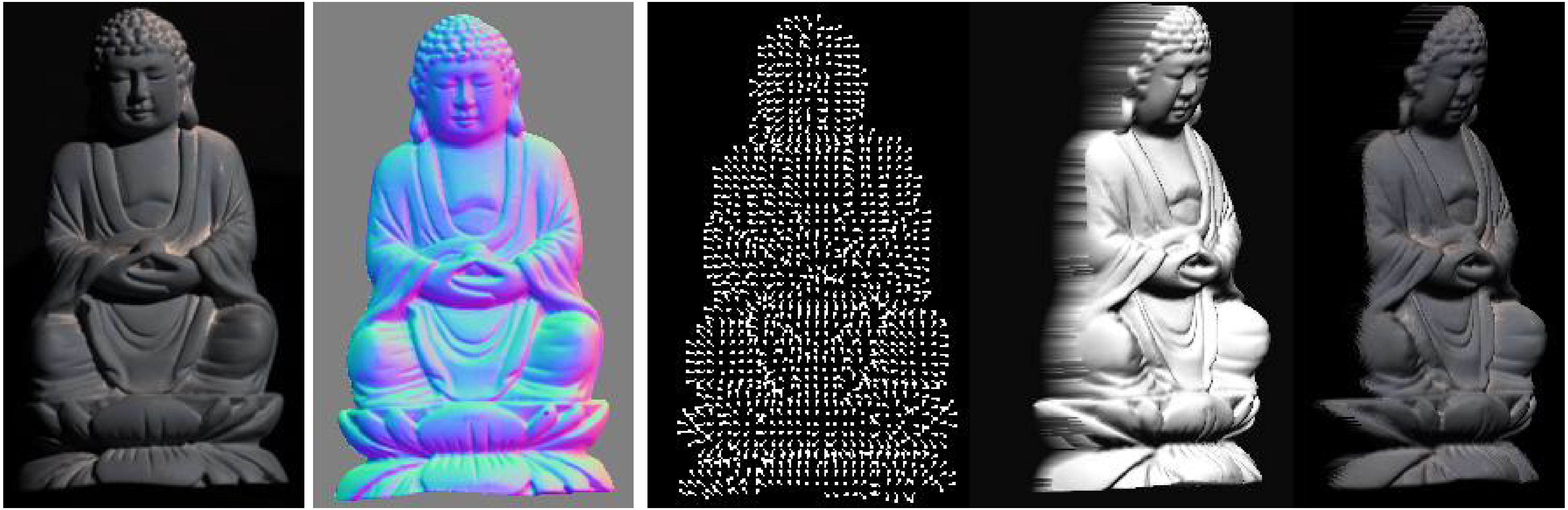


$$
\begin{aligned}
V_1 &= (x+1, y, z_{x+1,y}) - (x, y, z_{xy}) \\
&= (1, 0, z_{x+1,y} - z_{xy})
\end{aligned}
$$

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

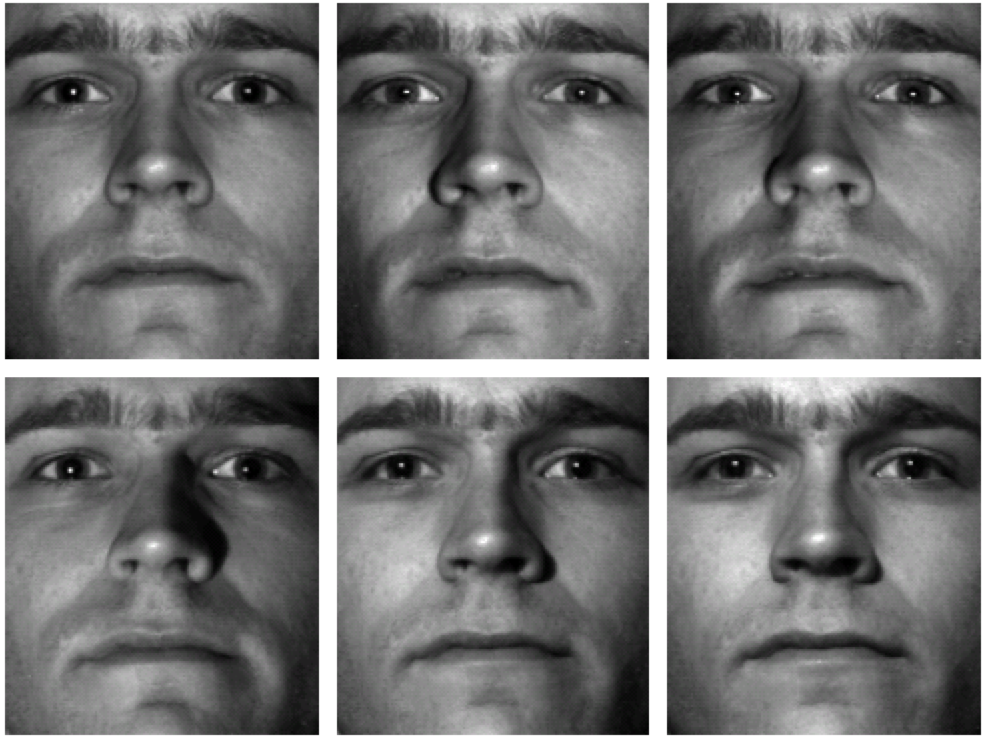
- Get a similar equation for $\mathbf{V_2}$
  - Each normal gives us two linear constraints on z
  - compute z values by solving a matrix equation

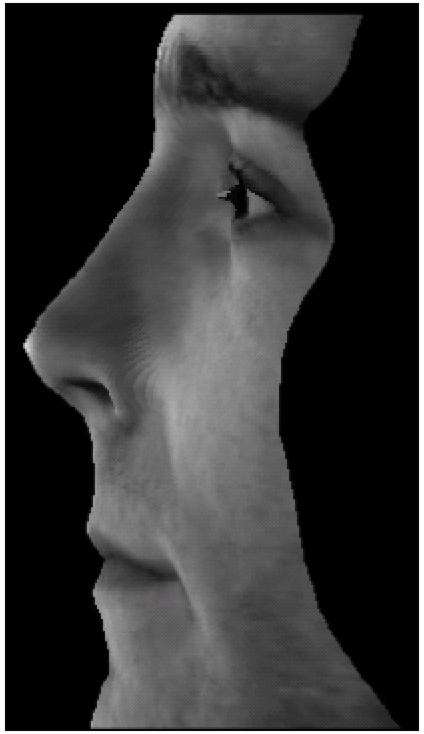
Results...



Input (1 of 12) | Normals | Normals | Shaded rendering | Textured rendering

Results of shape from shading on a human face





from Athos Georghiades

http://cvc.yale.edu/people/Athos.html

Limitations to Shape from Shading

- Big problems
  - doesn't work for shiny things, semi-translucent things
  - shadows, inter-reflections, multiple light sources
- 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

Variational Shape from Shading finds the surface profile that best fits the intensities observed

- Approach: energy minimization
- Given observed E(x,y), find shape Z(x,y) that minimizes energy

$$\mathrm{E} = \int \left((E(x,y) - L\rho\mathbf{l} \cdot \mathbf{n}(x,y))^2 + \lambda\left({p_x}^2 + {p_y}^2 + {q_x}^2 + {q_y}^2\right) \right) \, dxdy$$

- Regularization: minimize combination of disparity w. data, surface curvature

Variational Shape from Shading

- Solve by techniques from calculus of variations
- Use Euler-Lagrange equations to get a PDE, solve numerically
  - "greedy" methods tend not to work well

$$\frac{\partial \mathrm{E}}{\partial \mathrm{p}} - \frac{d}{dx}\frac{\partial \mathrm{E}}{\partial \mathrm{p_x}} - \frac{d}{dy}\frac{\partial \mathrm{E}}{\partial \mathrm{p_y}} = 0$$

$$\frac{\partial \mathrm{E}}{\partial \mathrm{q}} - \frac{d}{dx}\frac{\partial \mathrm{E}}{\partial \mathrm{q_x}} - \frac{d}{dy}\frac{\partial \mathrm{E}}{\partial \mathrm{q_y}} = 0$$

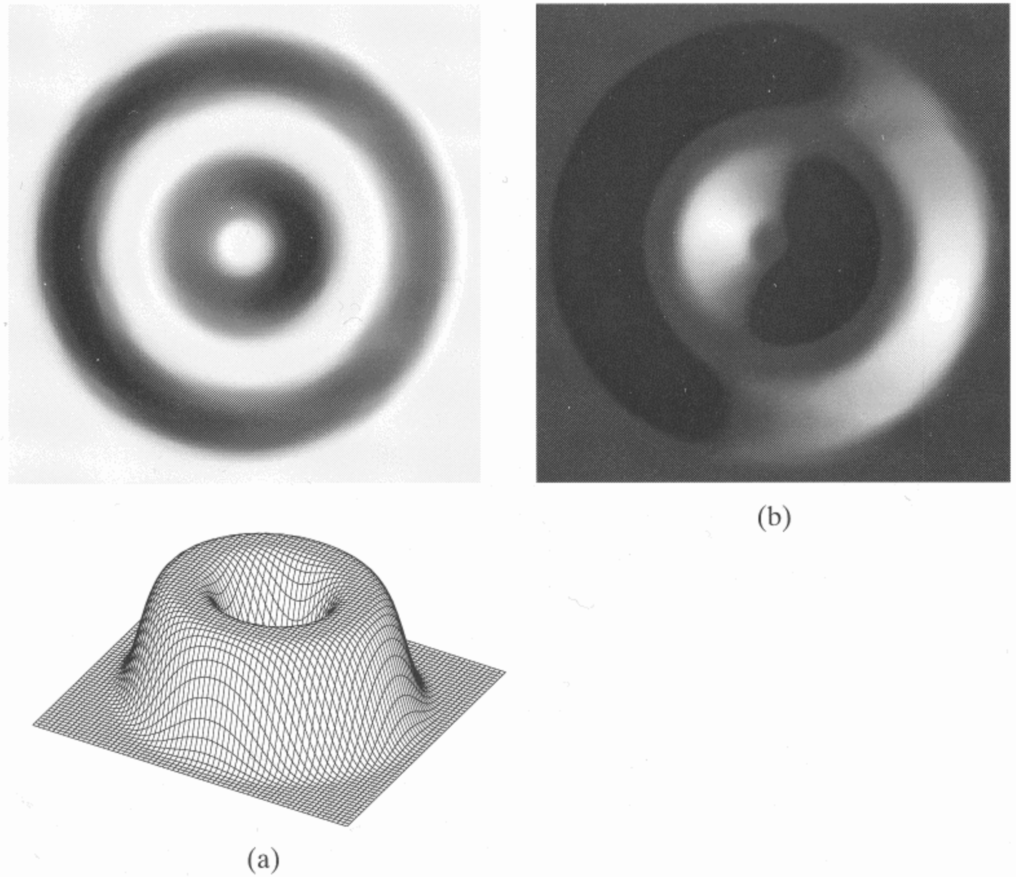


Figure 9.2 Two images of the same Lambertian surface seen from above but illuminated from different directions and 3-D rendering of the surface. Practically all the points in the top left image receive direct illumination ($\mathbf{i} = [0.20, 0, 0.98]^\top$); some regions of the top right image are in the dark due to self-shadowing effects ($\mathbf{i} = [0.94, 0.31, 0.16]^\top$).

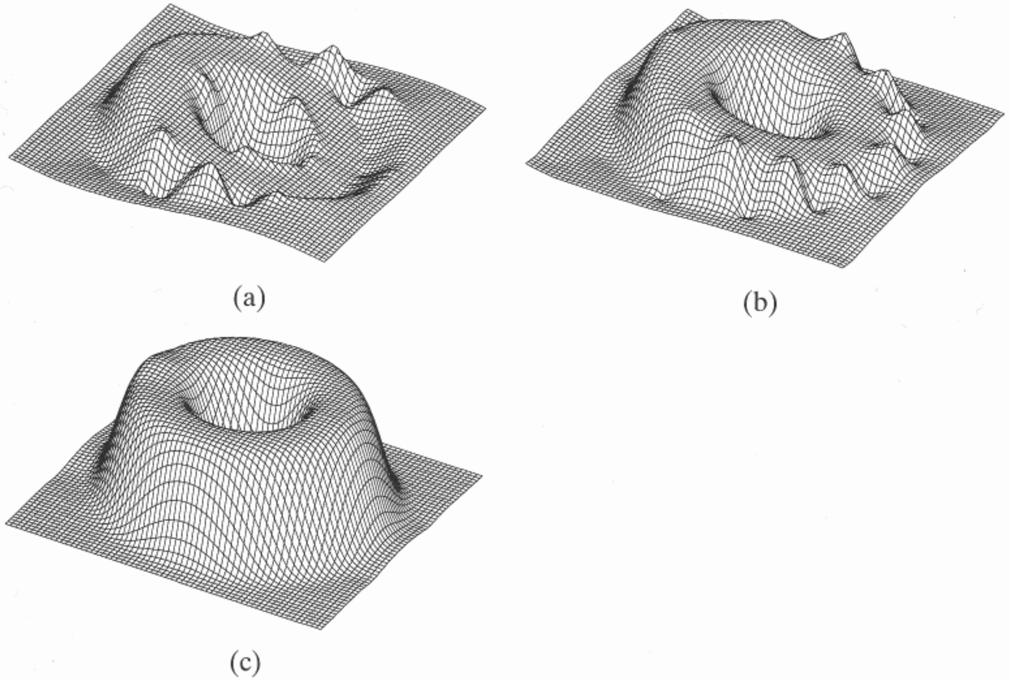


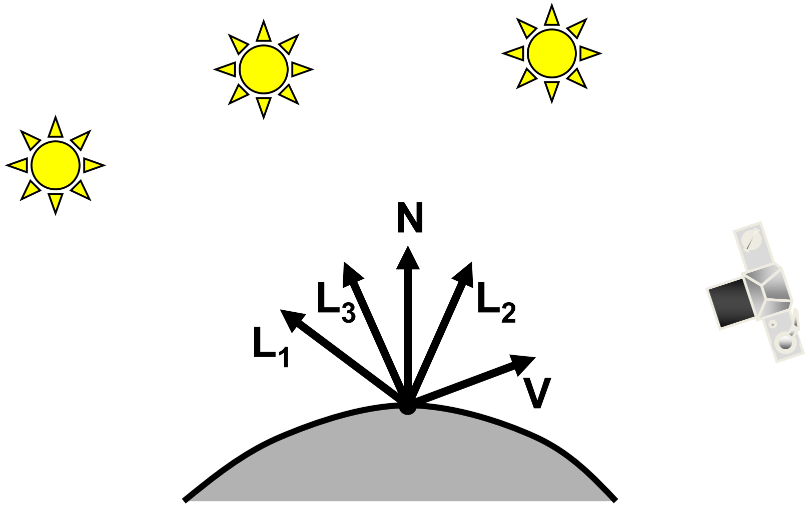
Figure 9.4 Reconstructions of the surface in Figure 9.2 after 100 (a), 1000 (b) and 2000 (c) iterations. The initial surface was a plane of constant height. The asymmetry of the first two reconstruction is due to the illuminant direction.

[Trucco & Verri]

Active Shape from Shading

- Idea: several (user-controlled) light sources
- More data
  - Allows determining surface normal directly
  - Allows spatially-varying reflectance
  - Redundant measurements: discard shadows and specular highlights
- Often called "photometric stereo"

Photometric stereo



$$I_1 = k_d \mathbf{N} \cdot \mathbf{L_1}$$
$$I_2 = k_d \mathbf{N} \cdot \mathbf{L_2}$$
$$I_3 = k_d \mathbf{N} \cdot \mathbf{L_3}$$

Can write this as a matrix equation:

$$\left[\begin{array}{ccc} I_1 & I_2 & I_3 \end{array} \right] = k_d \mathbf{N}^T \left[\begin{array}{ccc} \mathbf{L_1} & \mathbf{L_2} & \mathbf{L_3} \end{array} \right]$$

Solving the equations

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

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

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

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

More than three lights

- Get better results by using more lights

$$\left[\begin{array}{ccc} I_1 & \ldots & I_n \end{array}\right] = k_d \mathbf{N}^T \left[\begin{array}{ccc} \mathbf{L_1} & \ldots & \mathbf{L_n} \end{array}\right]$$

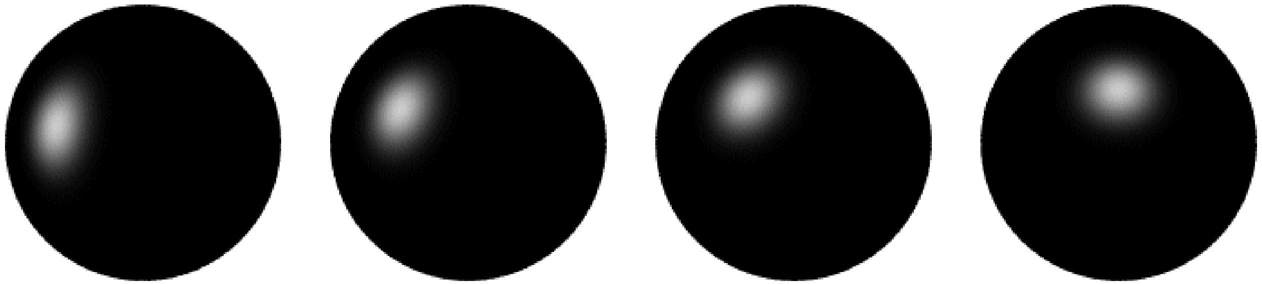
Least squares solution:

$$\begin{aligned} \mathbf{I} &= \mathbf{GL} \\ \mathbf{IL^T} &= \mathbf{GLL^T} \\ \mathbf{G} &= (\mathbf{IL^T})(\mathbf{LL^T})^{-1} \end{aligned}$$

Solve for N, k_d as before

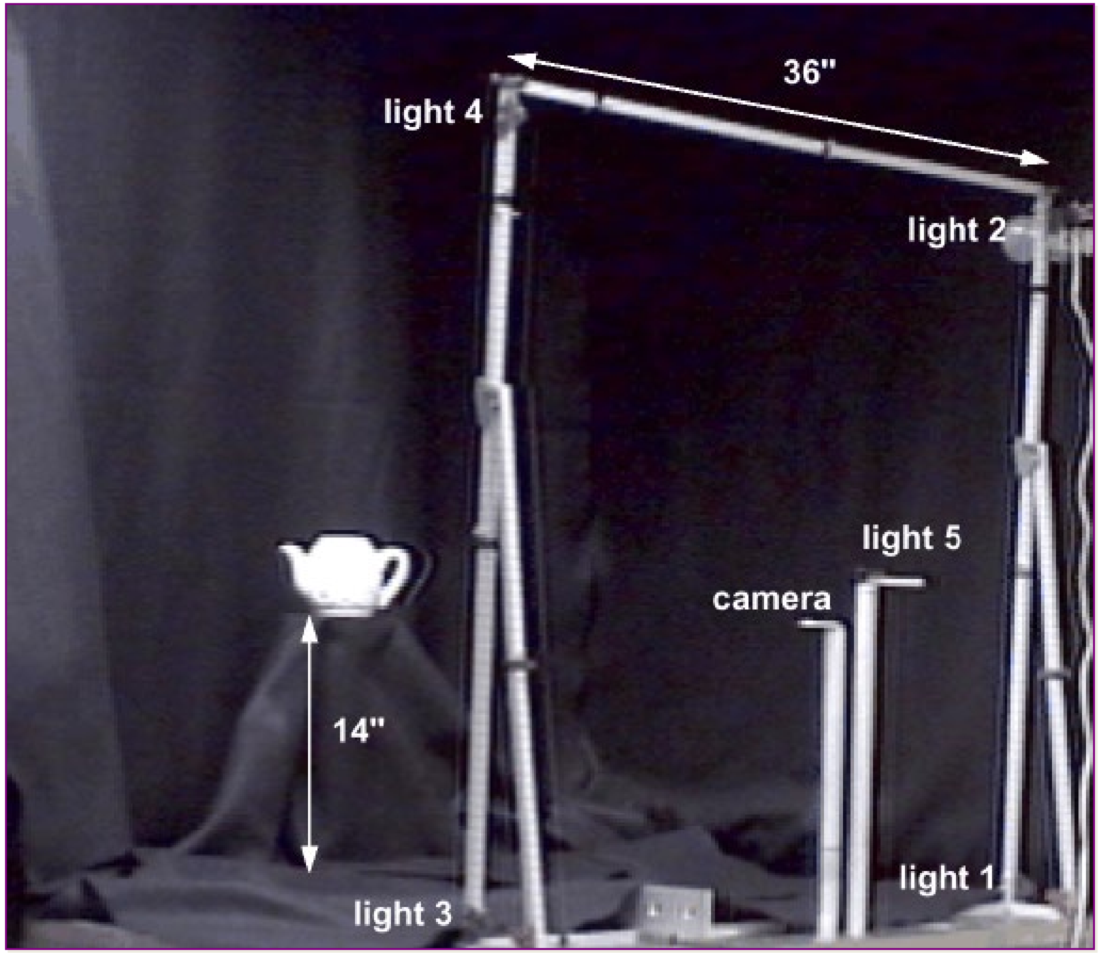
Computing light source directions

- Trick: place a chrome sphere in the scene



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

Photometric Stereo Setup



[Rushmeier et al., 1997]

Photometric Stereo Math

- For each point p, can write

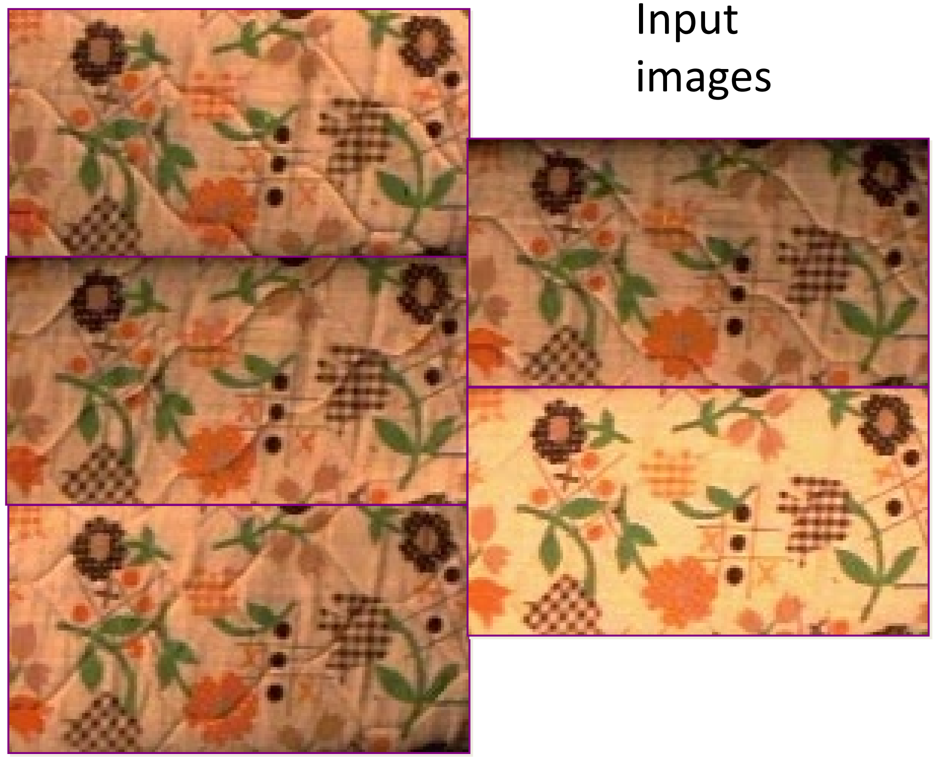
$$\rho_p \begin{bmatrix} l_{1,x} & l_{1,y} & l_{1,z} \\ l_{2,x} & l_{2,y} & l_{2,z} \\ l_{3,x} & l_{3,y} & l_{3,z} \end{bmatrix} \begin{bmatrix} n_{p,x} \\ n_{p,y} \\ n_{p,z} \end{bmatrix} = \alpha \begin{bmatrix} E_{p,1} \\ E_{p,2} \\ E_{p,3} \end{bmatrix}$$

- Constant α incorporates light source brightness, camera sensitivity, etc.

Photometric Stereo Math

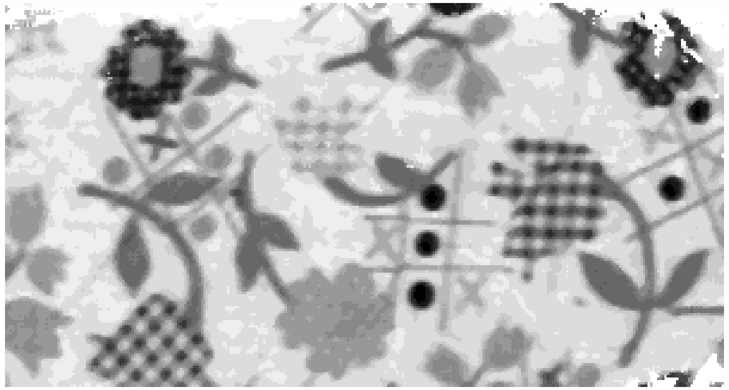
- Solving above equation gives (ρ/α) n
- n must be unit-length $\Rightarrow$ uniquely determined
- Determine ρ up to global constant
- With more than 3 light sources:
  - Discard highest and lowest measurements
  - If still more, solve by least squares

Photometric Stereo Results



Input images

Recovered normals

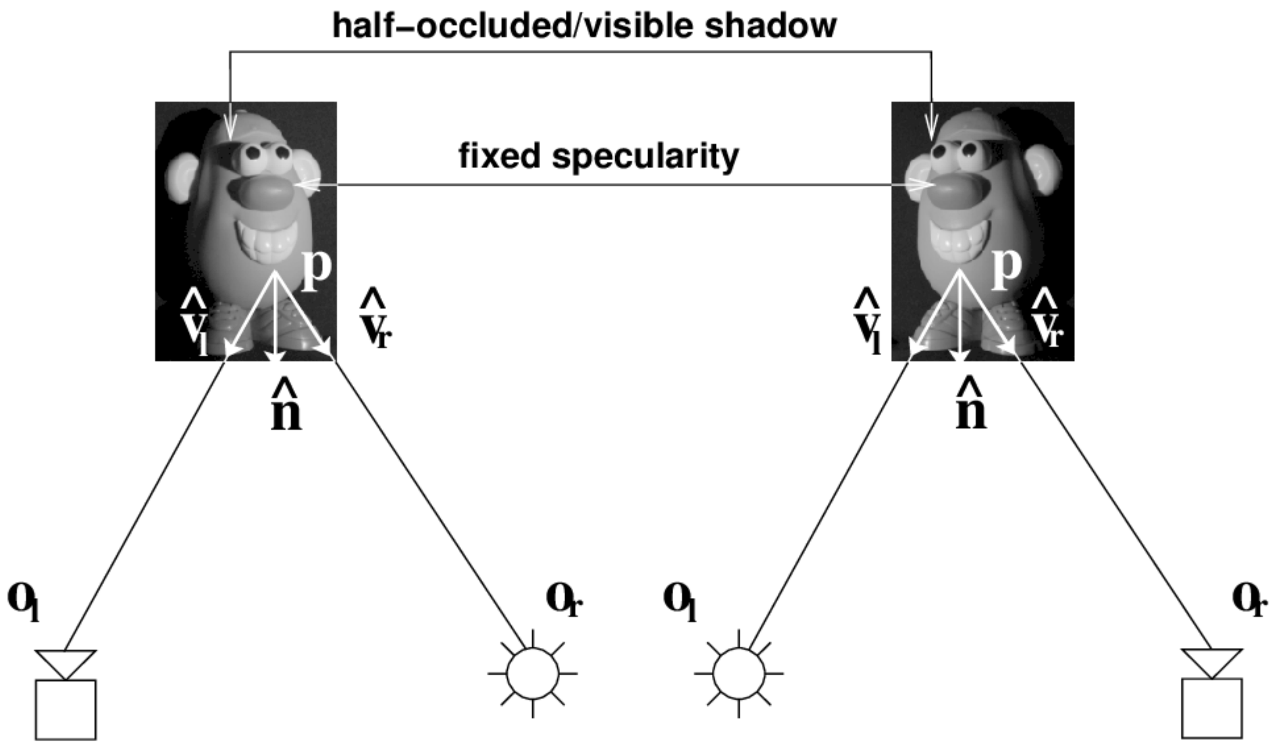


Recovered intensity

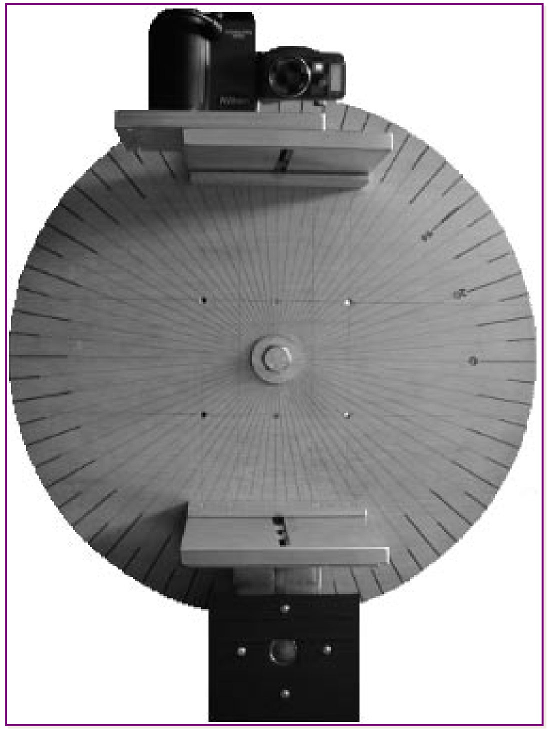
[Rushmeier et al., 1997]

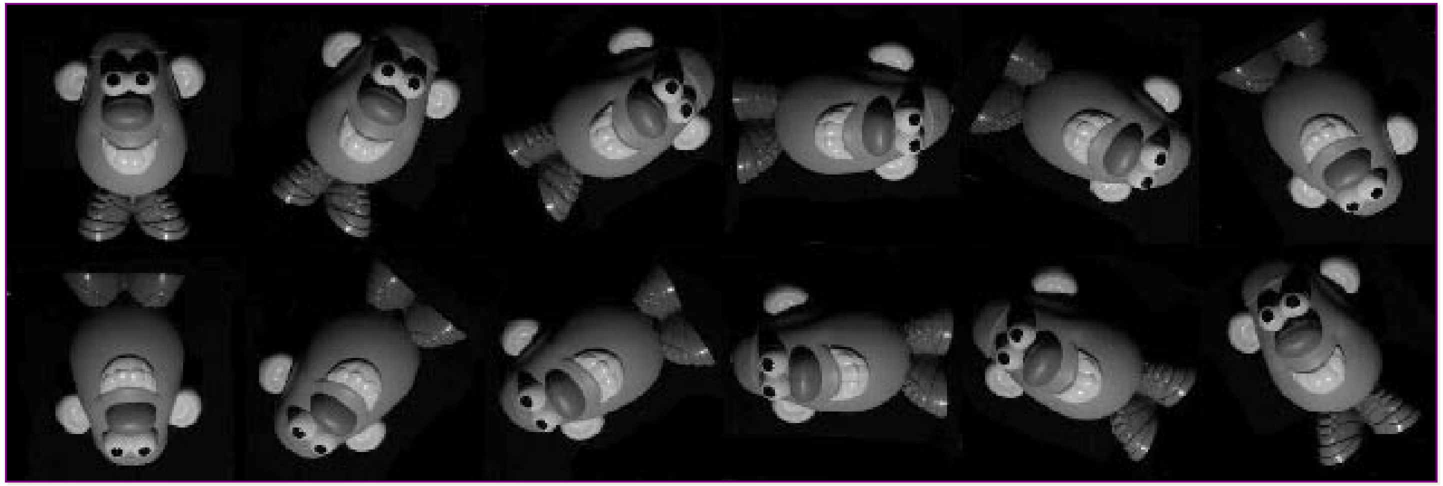
Helmholtz Stereopsis removes dependence on surface reflectance

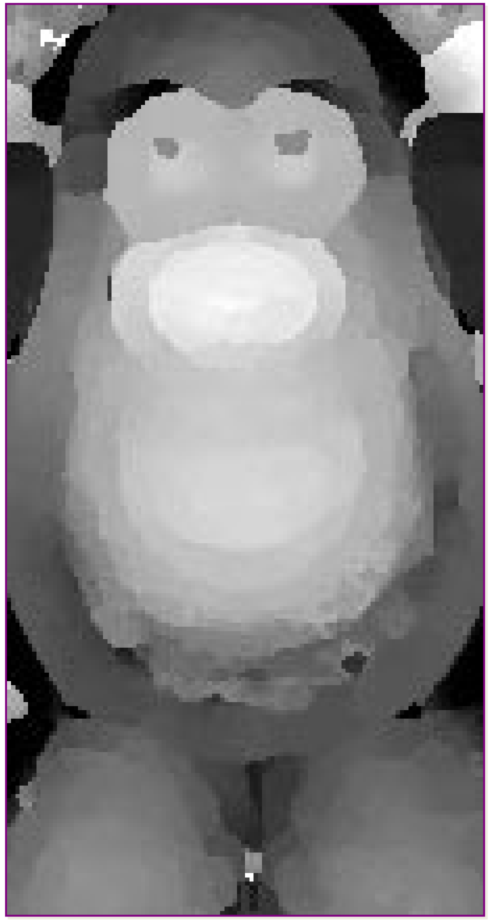
- Based on Helmholtz reciprocity: surface reflectance is the same under interchange of light, viewer
- So, take pairs of observations with viewer and light interchanged
- Ratio of the observations in a pair is independent of surface material!



Helmholtz Stereopsis using many images

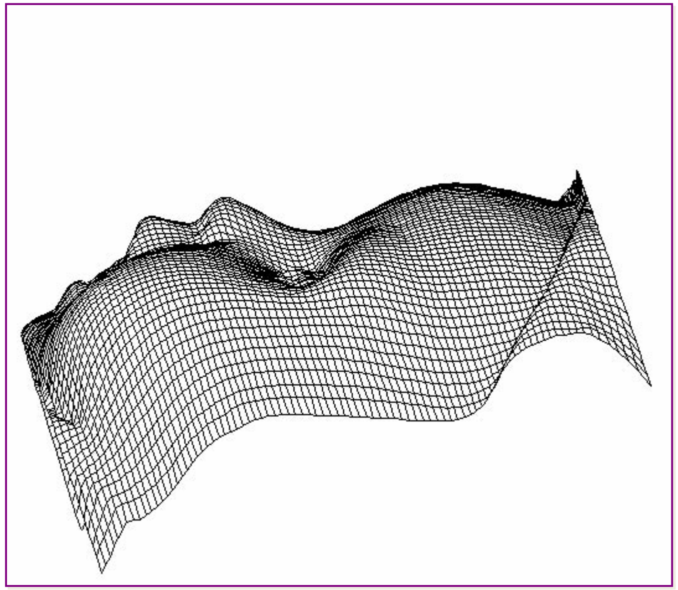


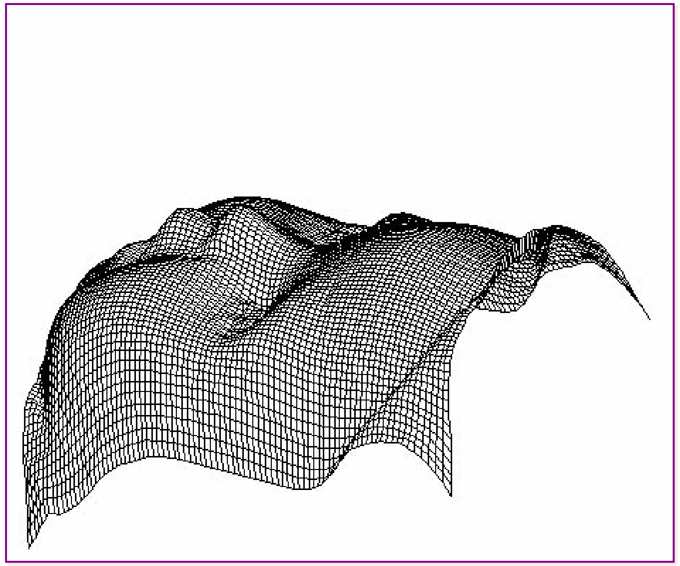




[Zickler, Belhumeur, & Kriegman]

Helmholtz Stereopsis





SHAPE VIA PASSIVE METHODS

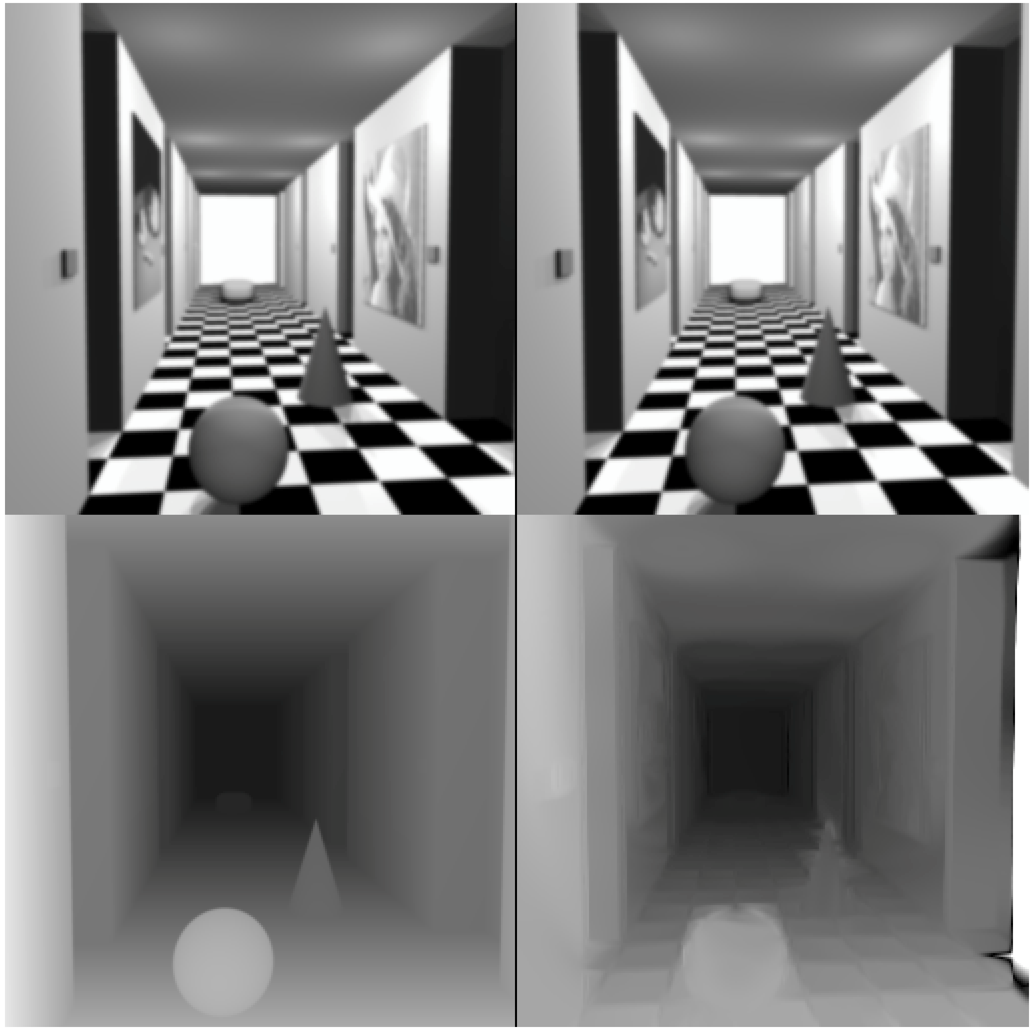
Passive sensing methods (one or more cameras observe the scene, without actively transmitting energy into the scene)

- Stereo (and MVS = multi-view stereo):
  Estimate depth with known inter-camera pose
- Structure from motion (SFM):
  Estimate 3D structure and camera pose
- Camera vergence
- Focus
- Shading
- Texture
- . . .

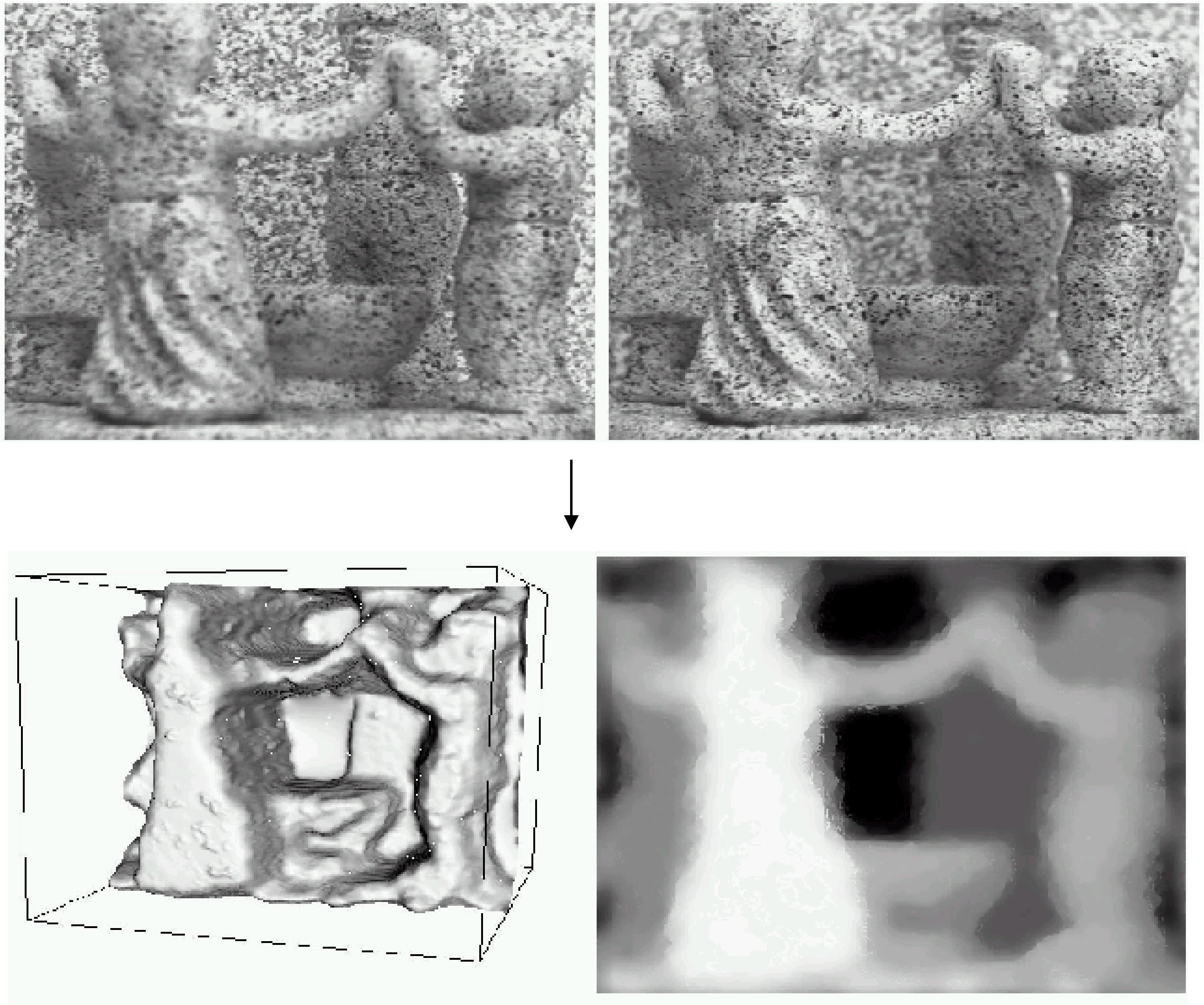
Humans use many different visual cues to estimate distances

As we have seen, stereo imaging can be used to find distance to objects in the scene by finding *disparity*

- The disparity map is calculated from two stereo images which were captured with known geometry
- The disparity is the change in position from one image to the next of each image pixel
- Done by finding key points in both images and interpolating
- Necessary to reject false correspondences
  - Epipolar geometry
  - Ordering
  - Smoothness



Focus/defocus



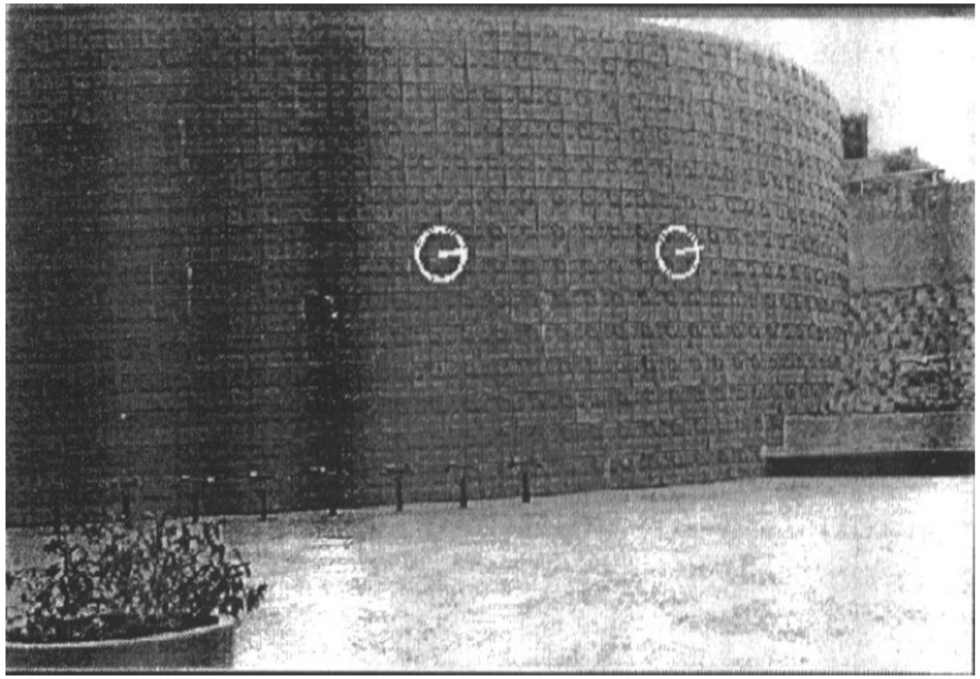
Images from same point of view, different camera parameters

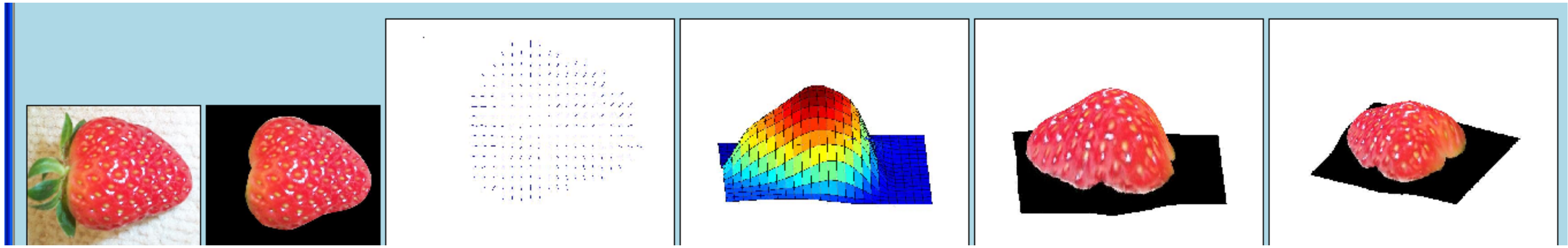
3d shape / depth estimates

[figs from H. Jin and P. Favaro, 2002]

Slide credit: Kristen Grauman

Texture





[From A.M. Loh. The recovery of 3-D structure using visual texture patterns. PhD thesis]

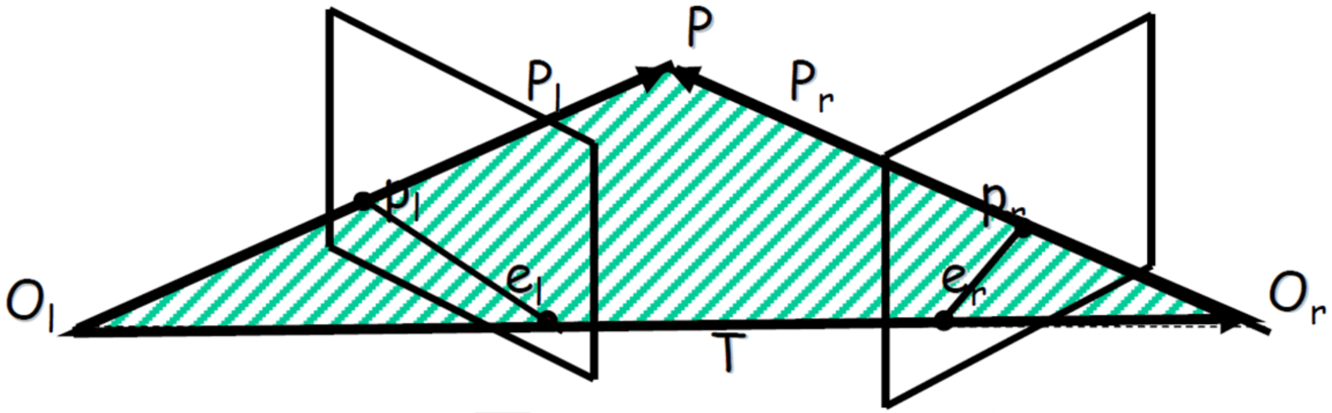
Slide credit: Kristen Grauman

Structure from motion (SFM)

Figures from L. Zhang

Slide credit: Kristen Grauman

Remember our approach to stereo vision; two cameras with different views of an object will have image point correspondences related by the Essential Matrix of the system



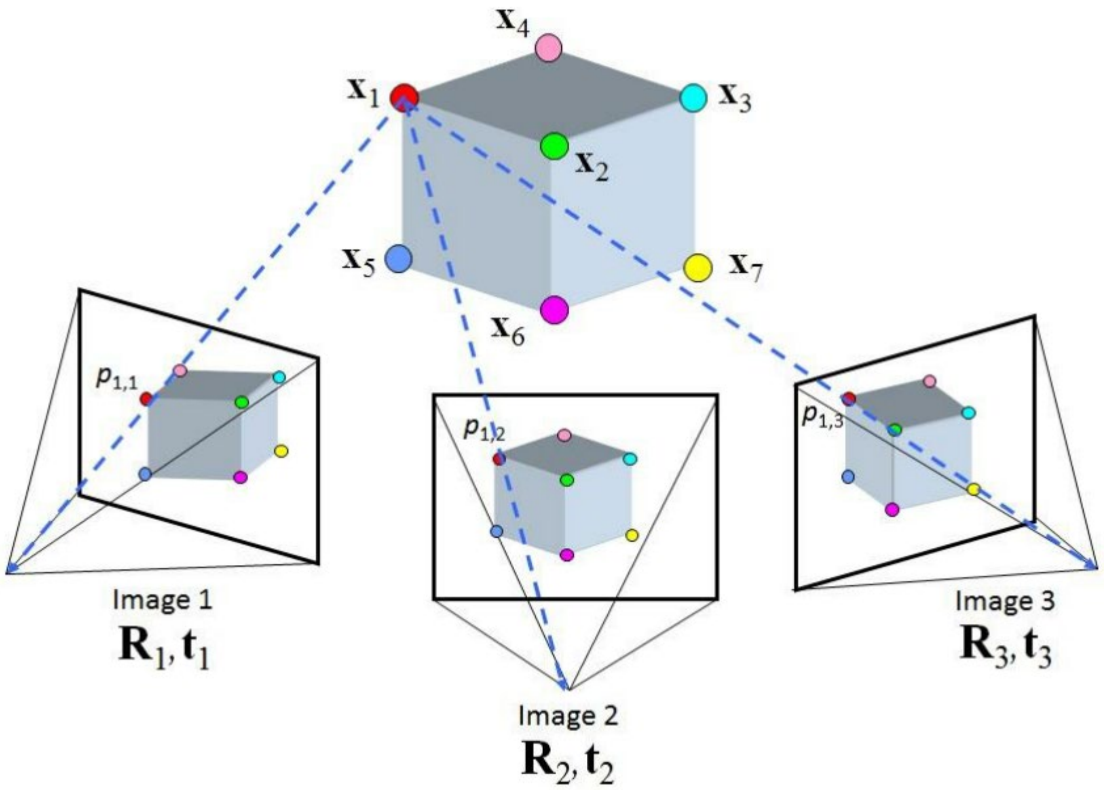
$$P_r^T R S P_l = 0$$

Essential Matrix:

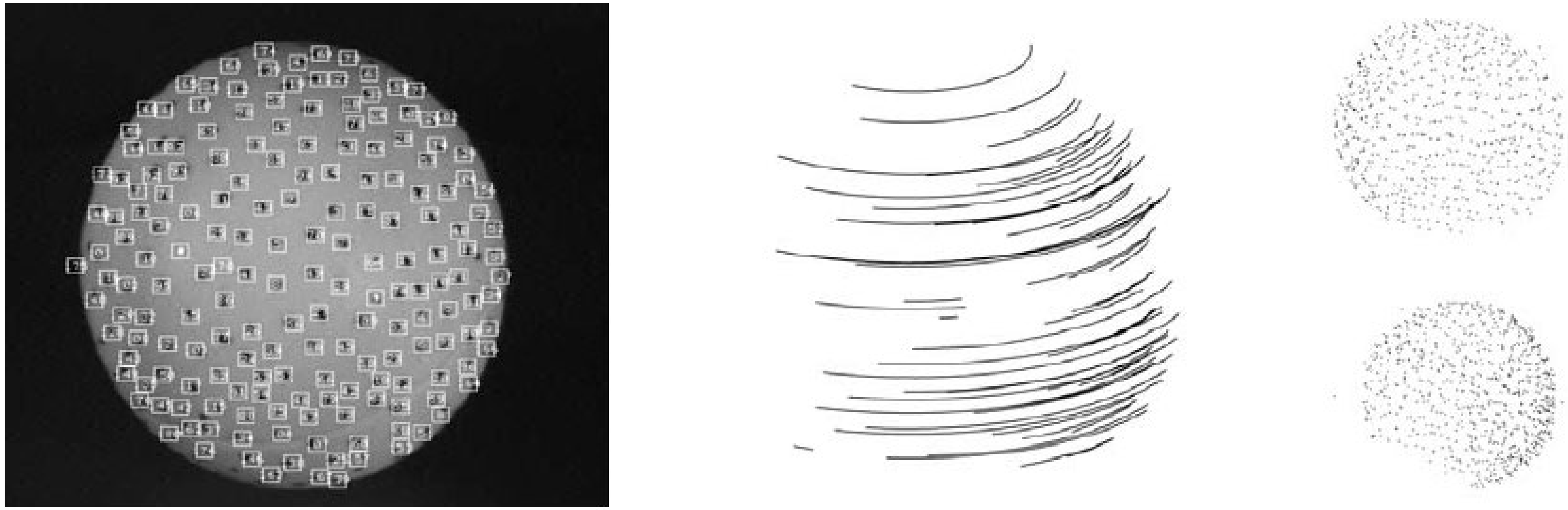
$$E = RS \qquad P_r^T E P_l = 0$$

First, consider a rotating object; note that two images taken from the same camera at different times can be considered as two cameras with different views

- Whether there are:
  - multiple cameras, or
  - a single camera revolving around a stationary object, or
  - a rotating object,
- we can obtain suitable images for structure reconstruction using stereo techniques
- If the motions are known!

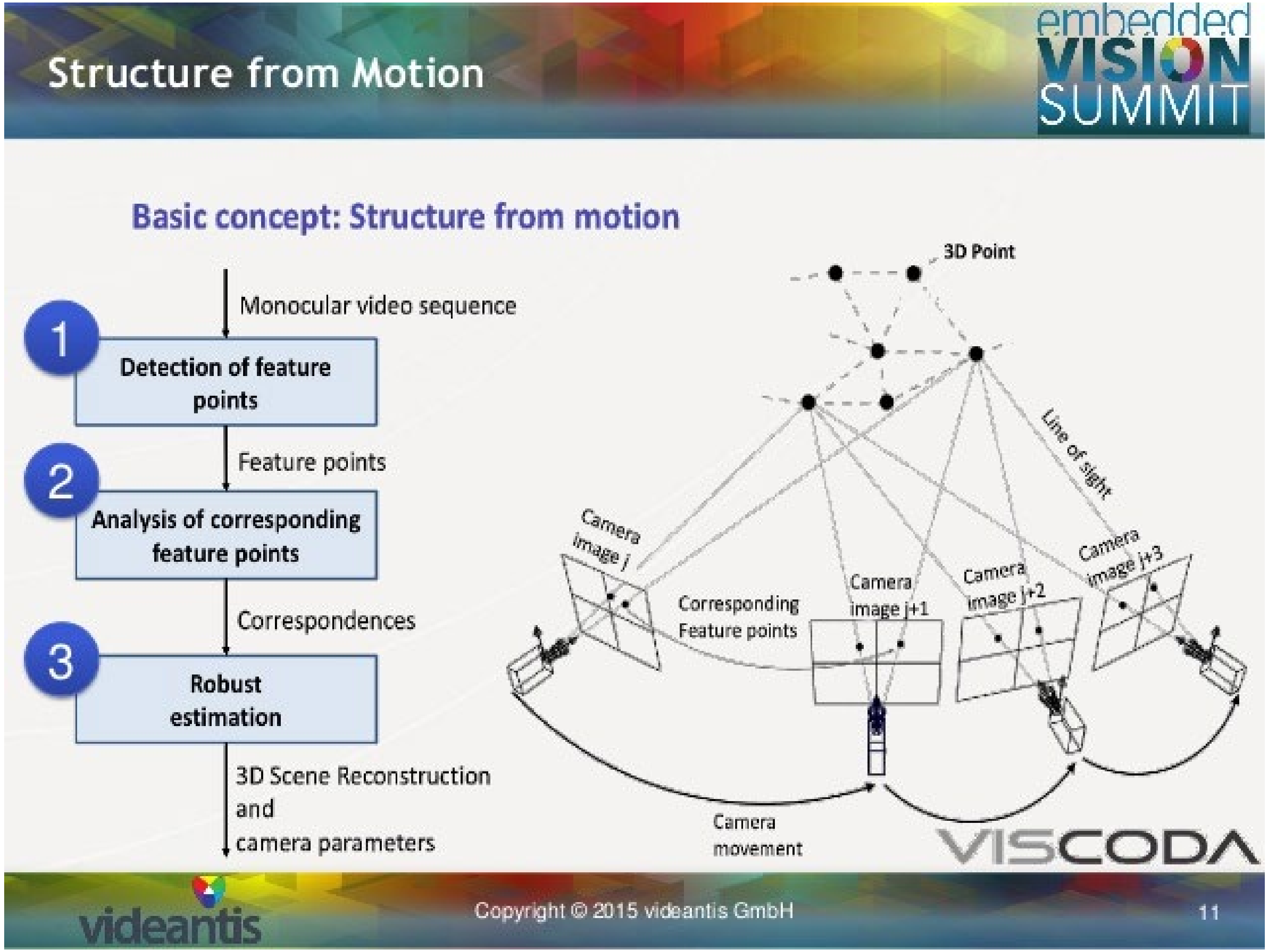


Shape from Motion (from Szeliski p. 316)



(a) (b) (c)

Extraction of more complex movement is possible



The key steps in more general shape from motion computation are:

- Take a set of images (video frames) containing the object or scene
- Extract key points from each: the average locations are $[\bar{\mathbf{X}}, \bar{\mathbf{Y}}, \bar{\mathbf{Z}}] = [0, 0, 0]$
- $\mathbf{M}_\mathrm{j}$ is the upper part of the projection matrix for image j (they differ because of motion) and $\mathbf{p}_\mathrm{i}$ is the i^{th} point $[\mathrm{X}, \mathrm{Y}, \mathrm{Z}]$
- The image points are given by $\tilde{\boldsymbol{x}}_{ji} = \boldsymbol{M}_j \boldsymbol{p}_i$
- Define a measurement matrix $\hat{\mathbf{X}} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_\mathrm{M} \end{bmatrix} [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_\mathrm{N}] = \hat{\mathbf{M}}\hat{\mathbf{S}}$
- $\hat{\mathbf{M}}$ is the measurement matrix and $\hat{\mathbf{S}}$ is the structure matrix
- Use singular value decomposition to find $\hat{\boldsymbol{X}} = \boldsymbol{U\Sigma V^T} = [\boldsymbol{UQ}][\boldsymbol{Q^{-1}\Sigma V^T}]$
- Q can be found by assuming some things about the motion (rotation only, displacement, etc.)

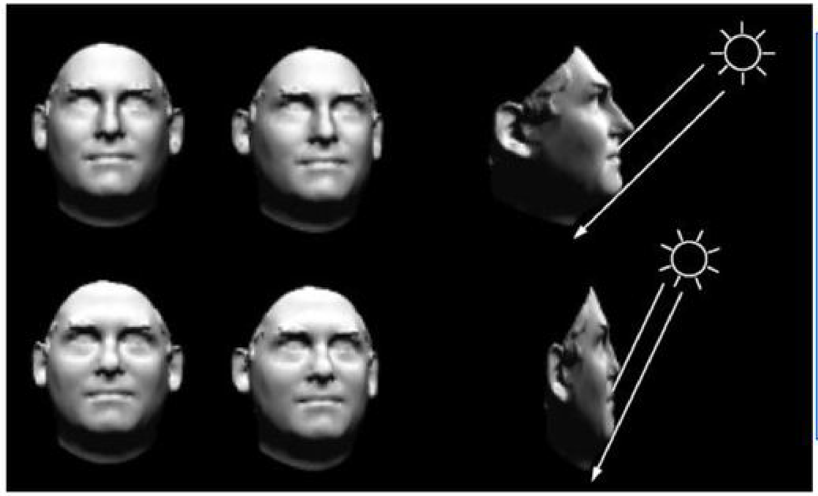
Even after shape from motion is calculated, there is still the possibility of the bas-relief ambiguity

Figure 1: Frontal and side views of a pair of marble bas-relief sculptures: Notice how the frontal views appear to have full 3-dimensional depth, while the side views reveal the flattening – the sculptures rise only 5 centimeters from the background plane. While subtle shading is apparent on the faces, the shadows on the women's pleats are the dominant perceptual cue in the body.

Belhumeur 2012 – The bas-relief ambiguity; https://ieeexplore.ieee.org/document/609461

Do ambiguities exist ?

Can two different objects produce the same illumination cone ? **YES** "Bas-relief" ambiguity

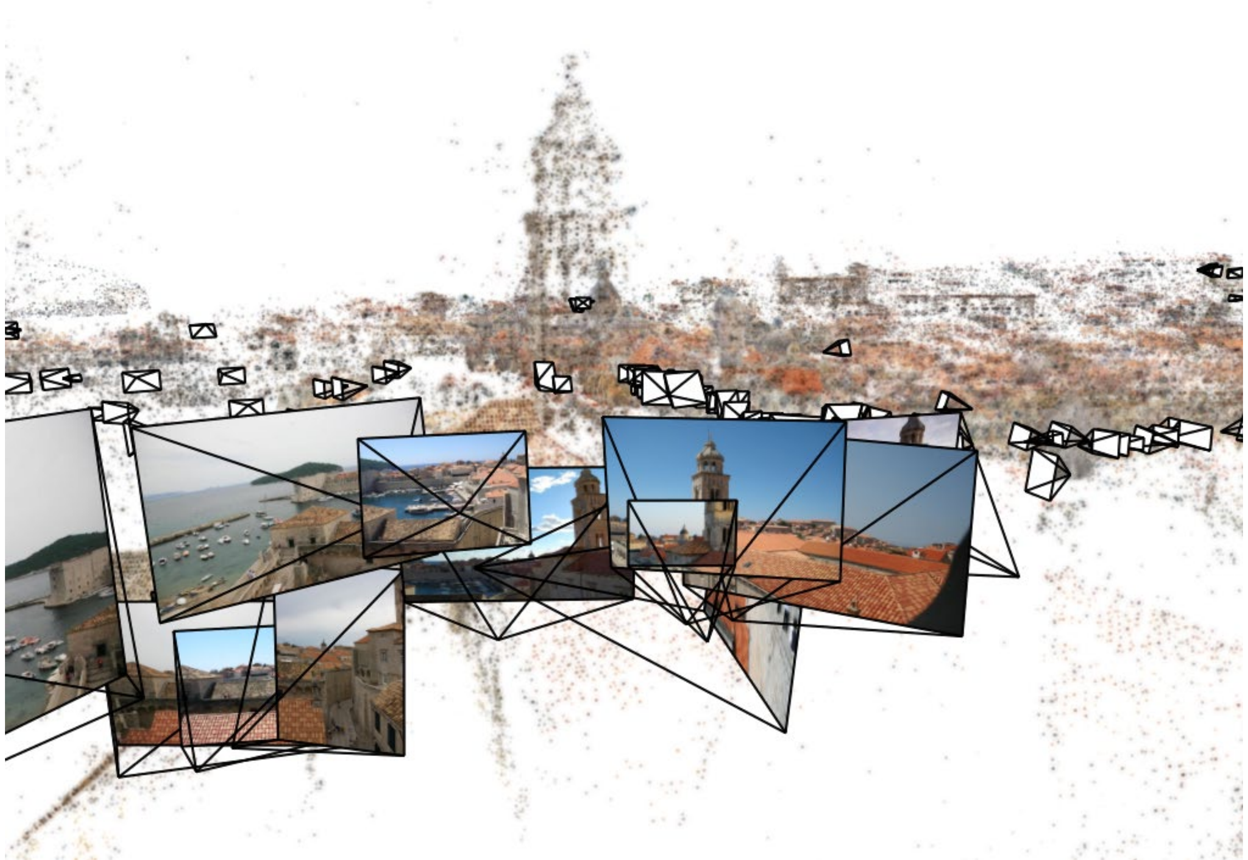


Convex object

- B span L
- Any $A \in GL(3)$, $B^{*}=BA$ span L
- $I=B^{*}S^{*}=(BA)(A^{-1}S)=BS$
  Same image B lighted with S and B^{*} lighted with S^{*}

When doing PCA the resulting basis is generally not normal*albedo

A collaborative group is working on a 3D model of the world (!) by using Structure from Motion on online photo collections

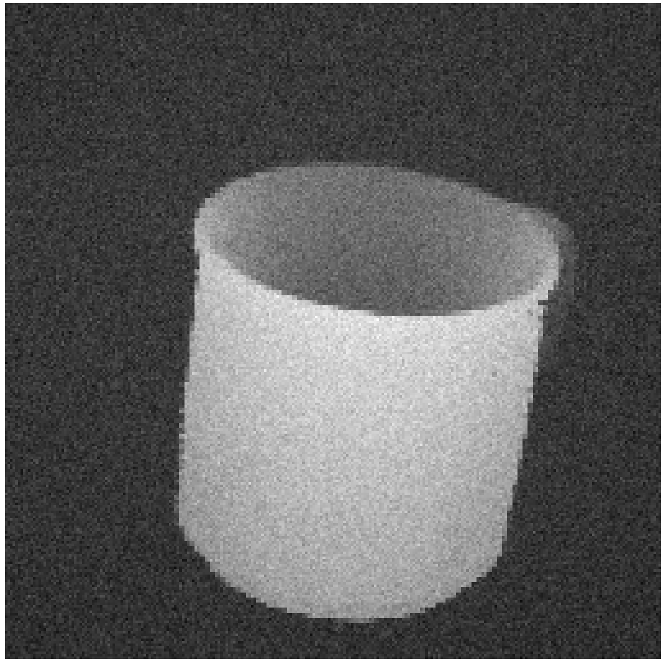


http://www.cs.cornell.edu/projects/bigsfm/

SHAPE VIA ACTIVE METHODS

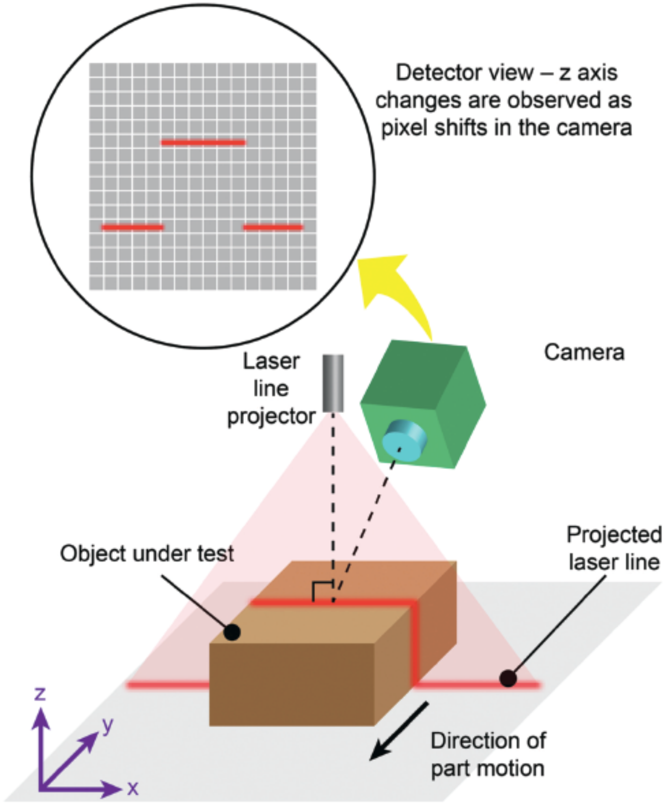
Active sensing methods (energy is transmitted into the scene; often “time of flight“ is measured to determine distance)

- Radar
- Lidar
- Ultrasound
- Tomographic imaging



- **There are also hybrid methods**
  - Structured light
  - Photometric stereo

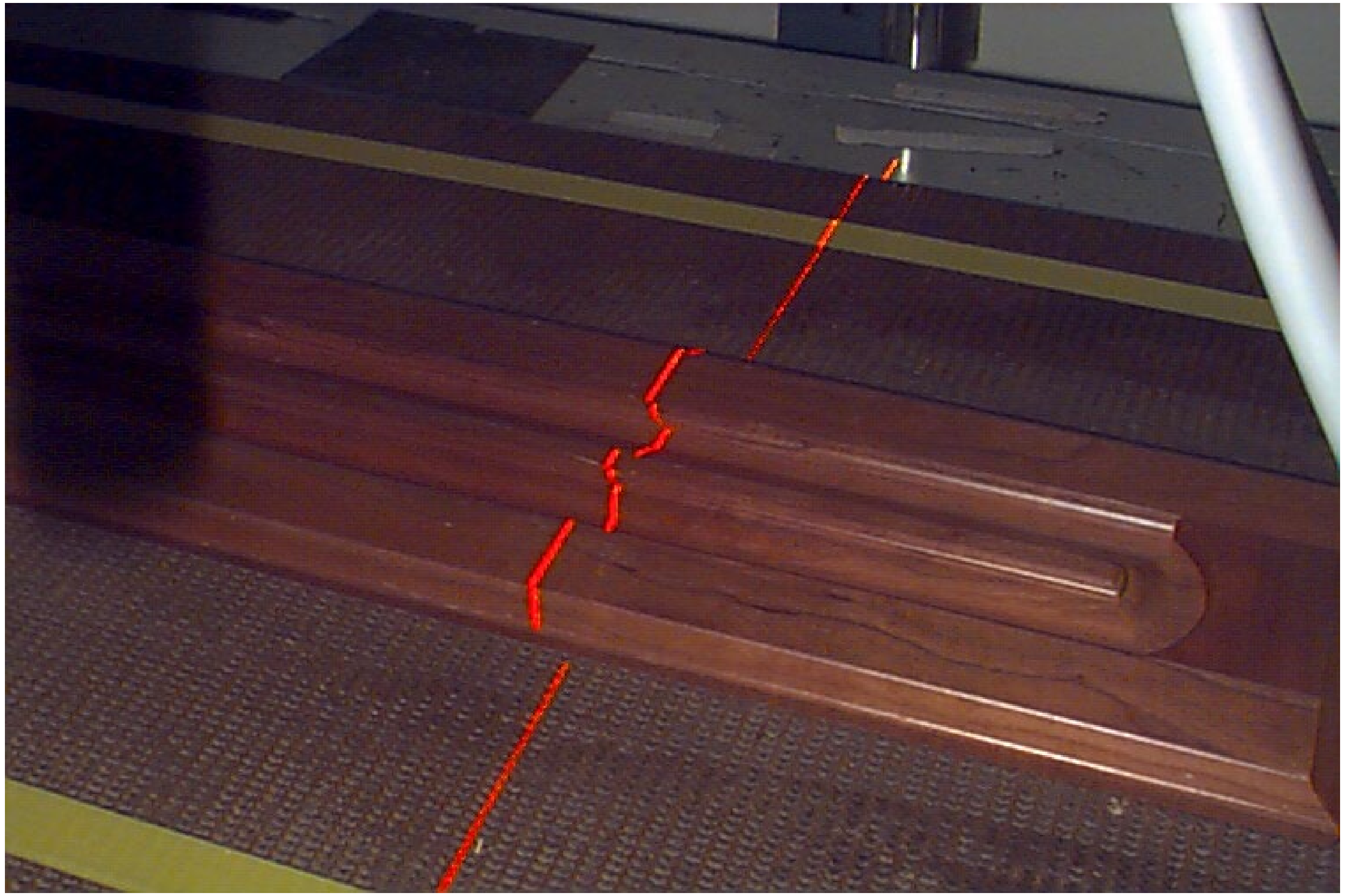
In well-controlled situations, simple structured light using triangulation can be used to infer object depth



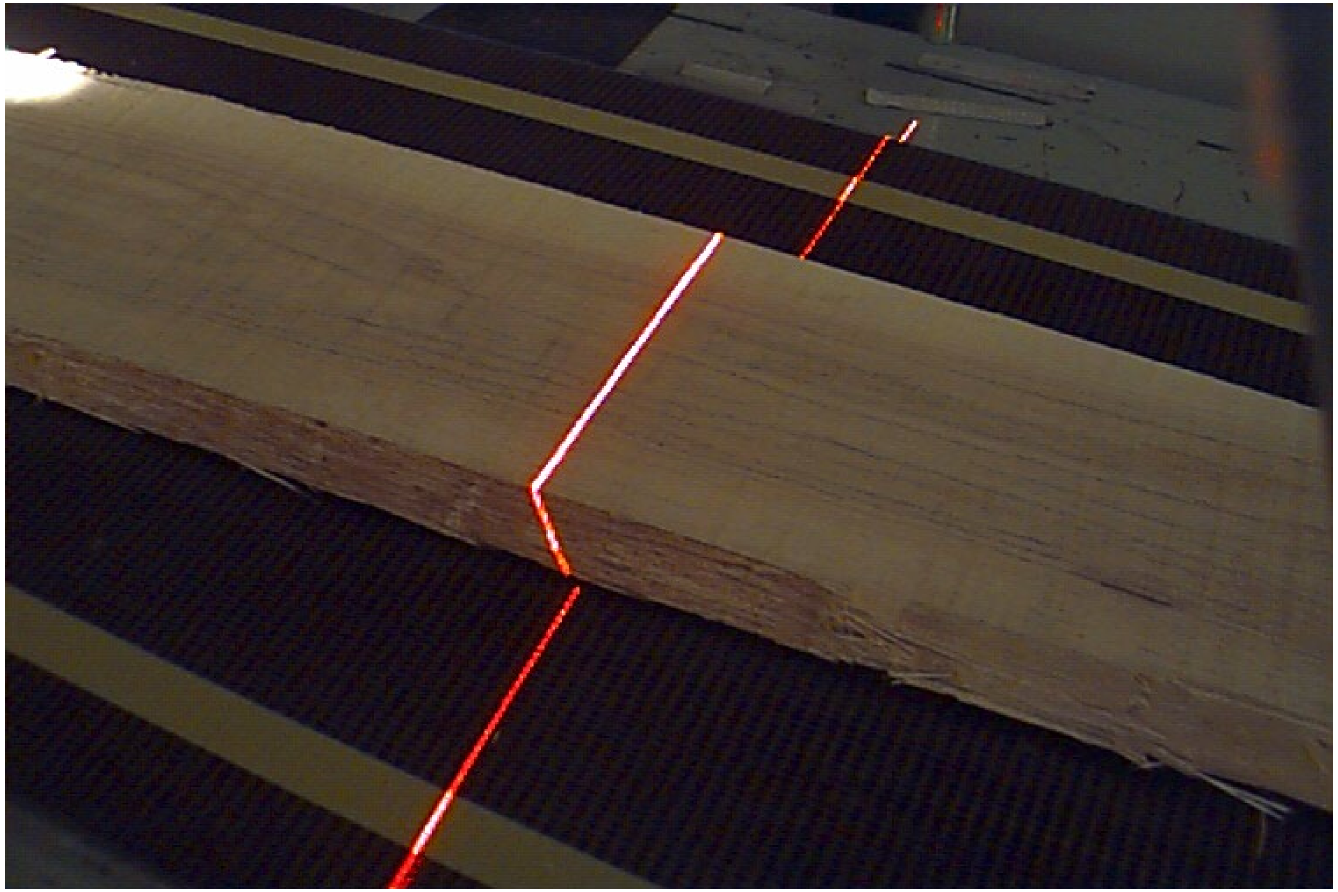
- Often, the camera is vertical and the light projector is at an angle (opposite to what is shown here)
- Deflection of the laser line is proportional to the z-axis height:

$$\Delta z = \Delta x \tan^{-1} \theta$$

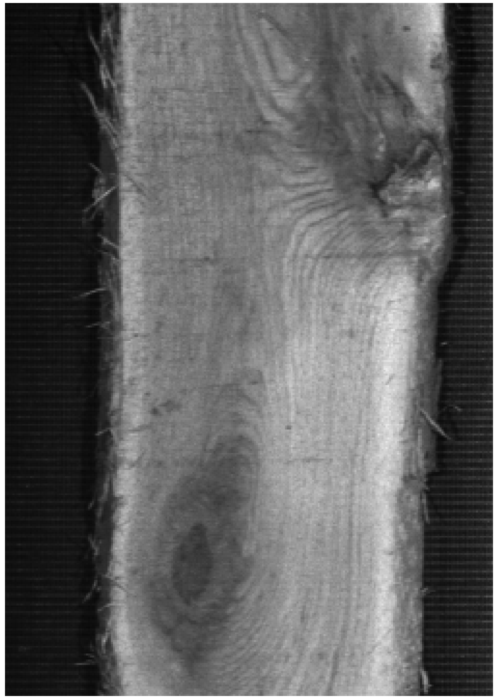
Laser profiling can determine complex shapes along the optic axis – such as these raised-panel doors



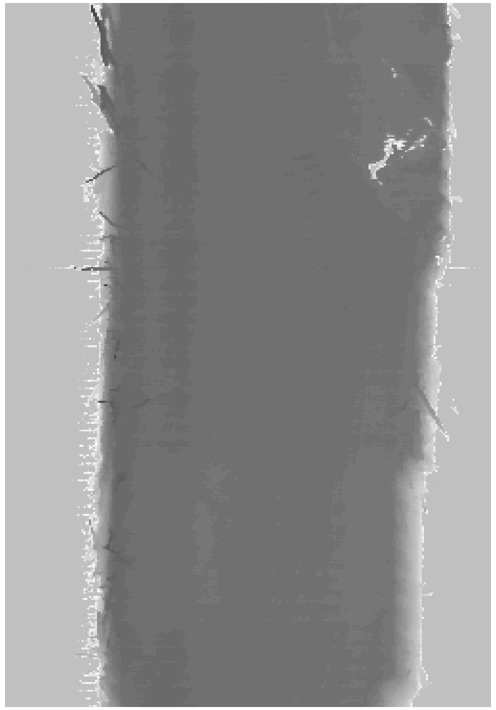
Lumber



VT Laboratory prototype

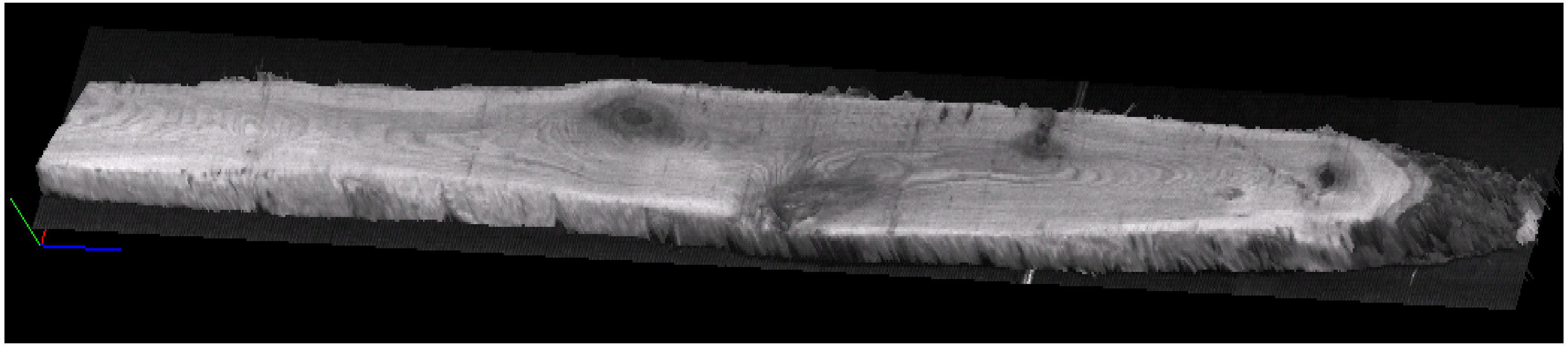


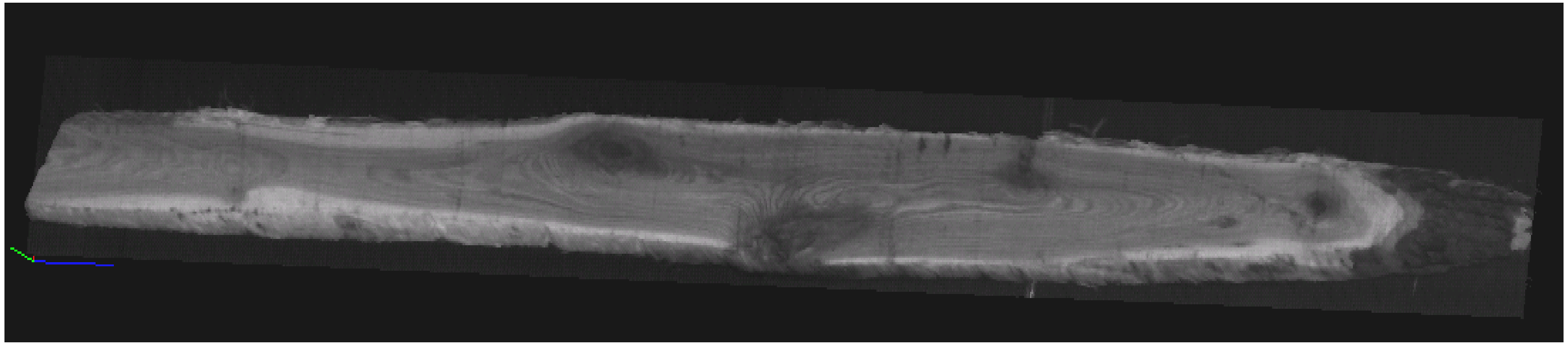
Gray-scale image



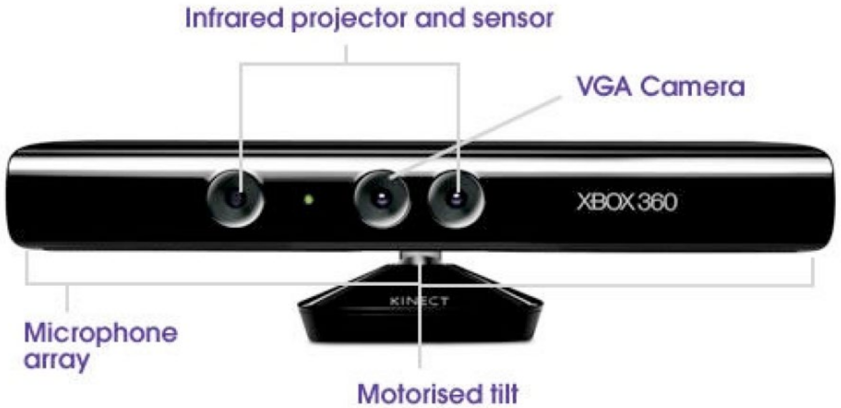
"Profile" image

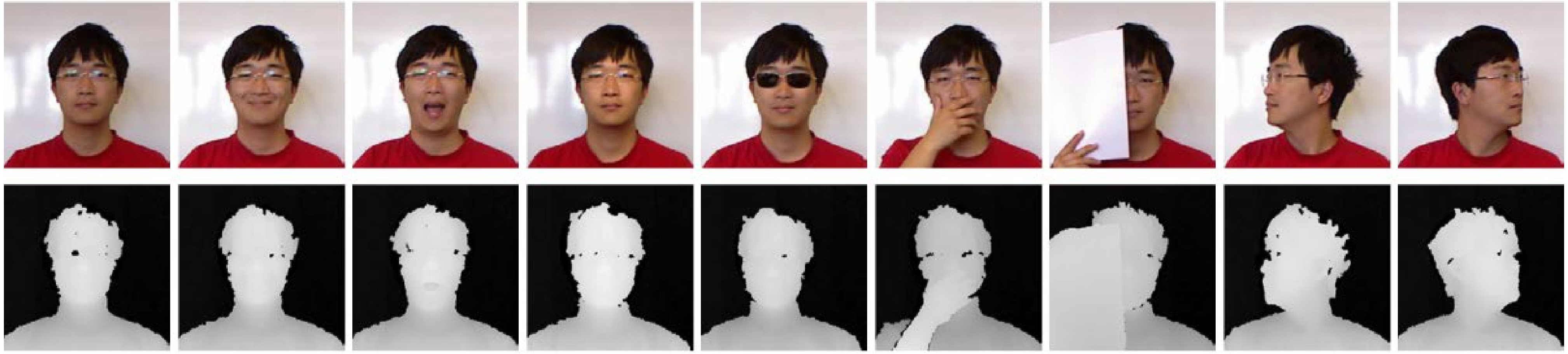
Example 3D plots





Some RGBD sensors project light in known patterns and use the image return to deduce the z-axis coordinate





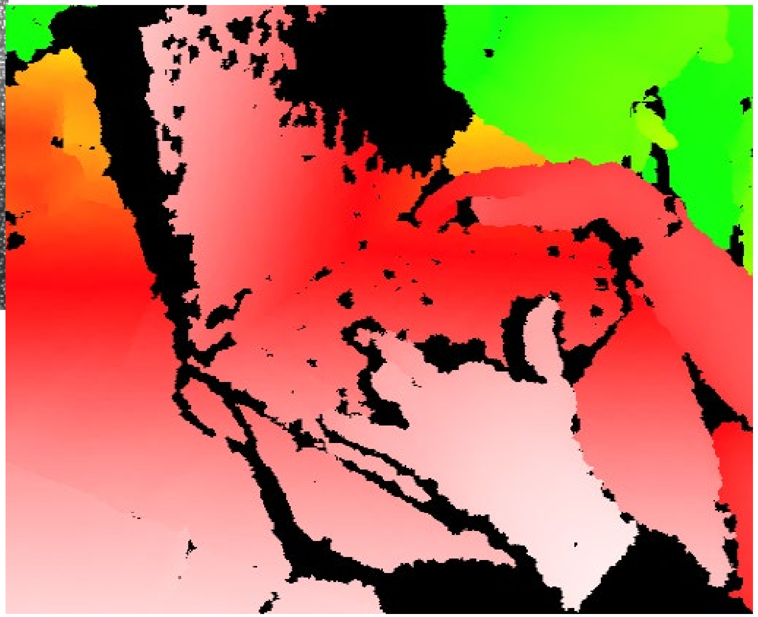
(a) (b) (c) (d) (e) (f) (g) (h) (i)

http://rgb-d.eurecom.fr/

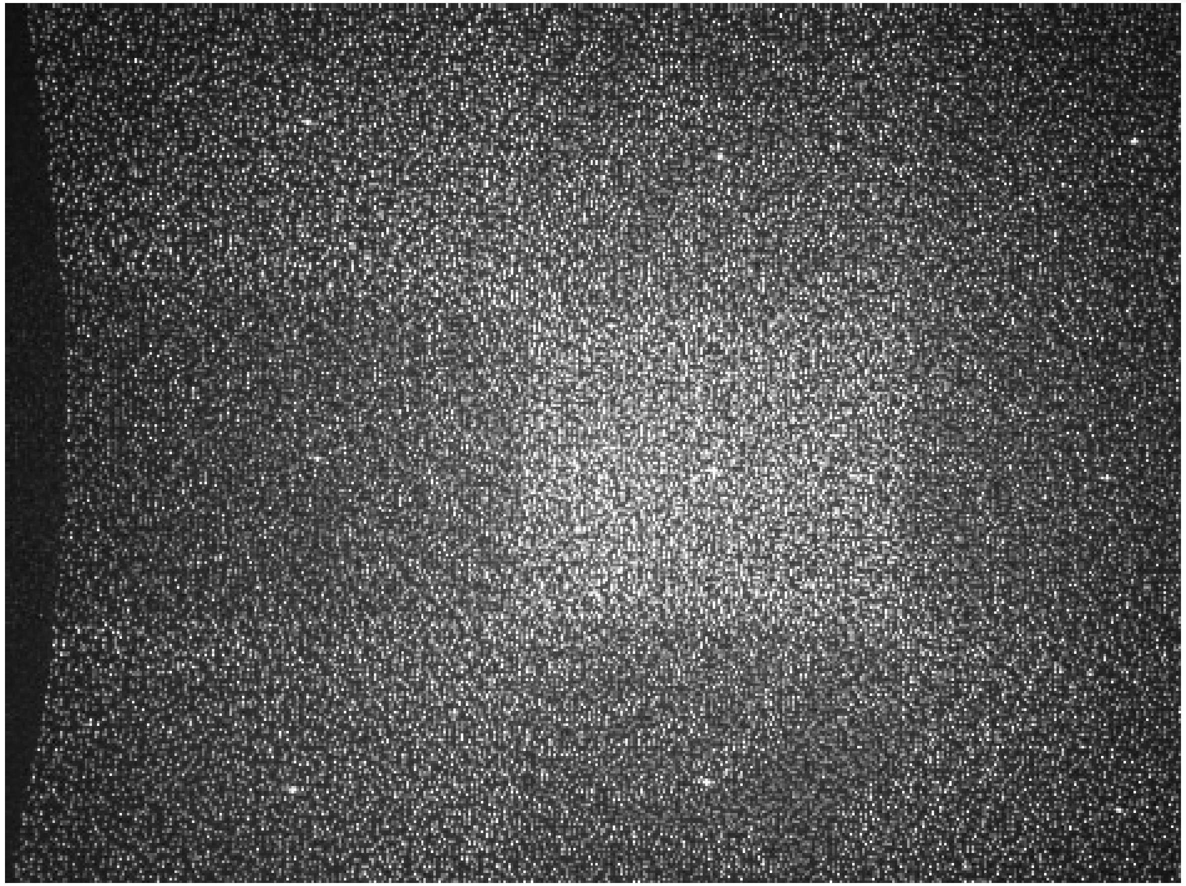
The Microsoft Kinect (V1) sensor projected a pattern of dots and from their position in the image, deduced an approximation to the range values

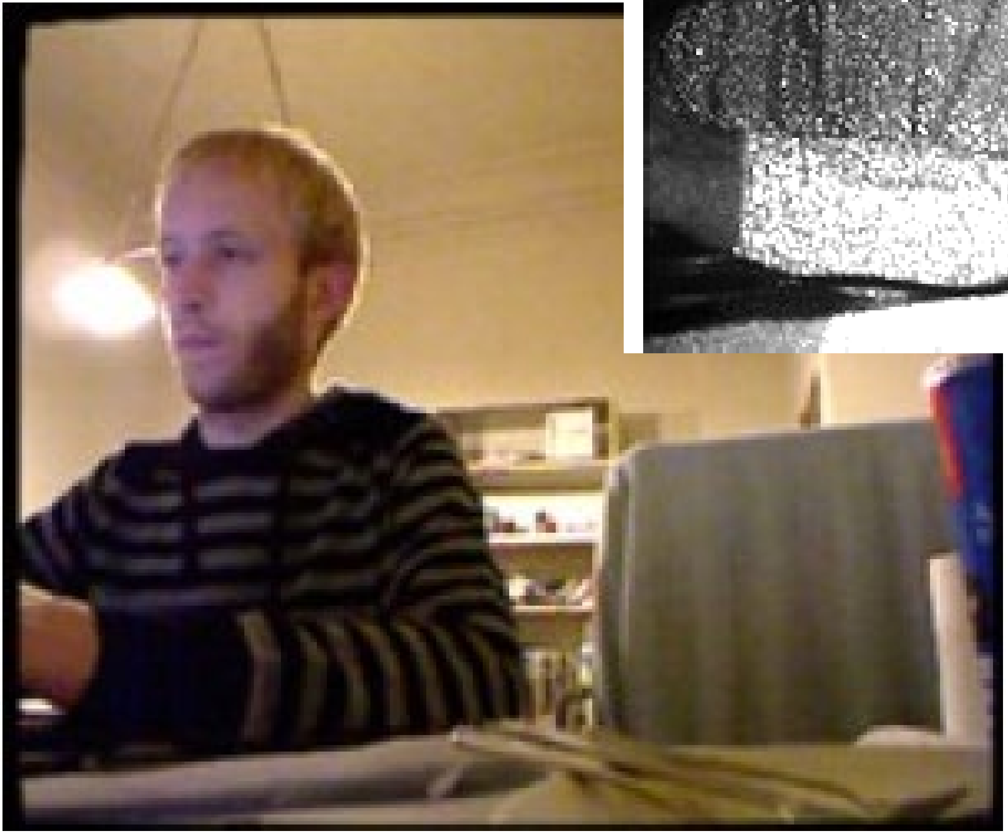
- **Triangulation is used to establish depth**

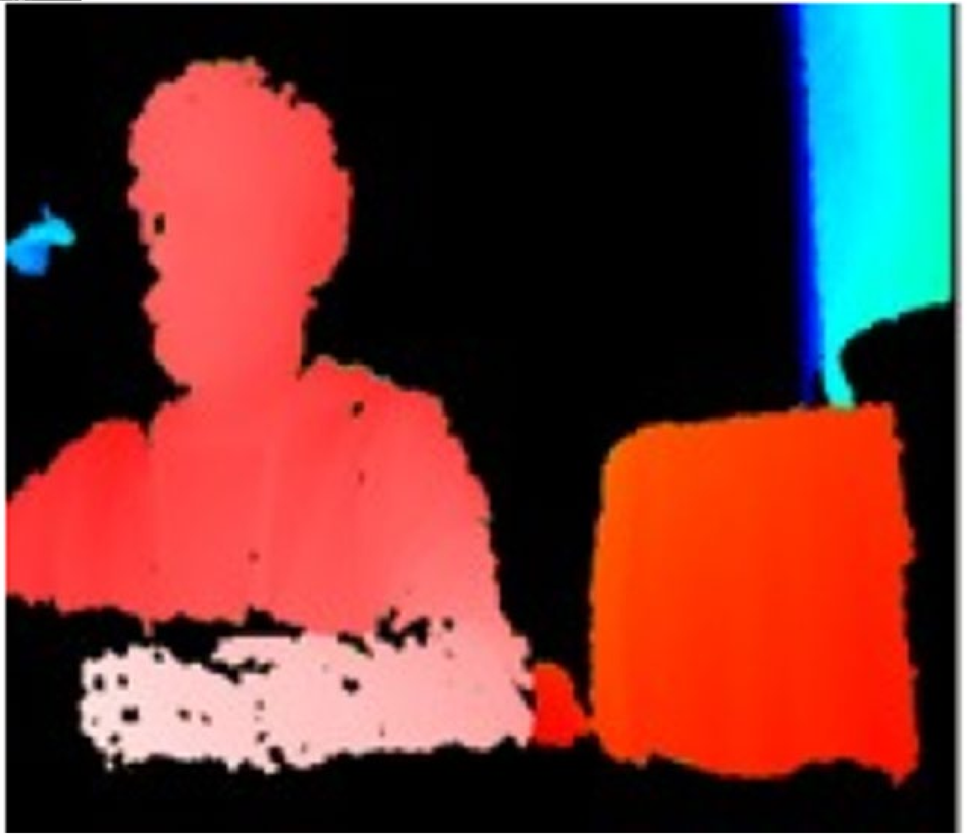
- **Depth sensing specs**
  512 x 424
  30 Hz
  FOV: 70 x 60
  One mode: .5–4.5 meters



- Transmit a known pattern into the scene
- Use infrared wavelengths so it is not visible to humans
- Use correlation methods to find matches
  (e.g., 9x9 windows, then interpolate for subpixel resolution)
- Run at 10 to 30 frames per second







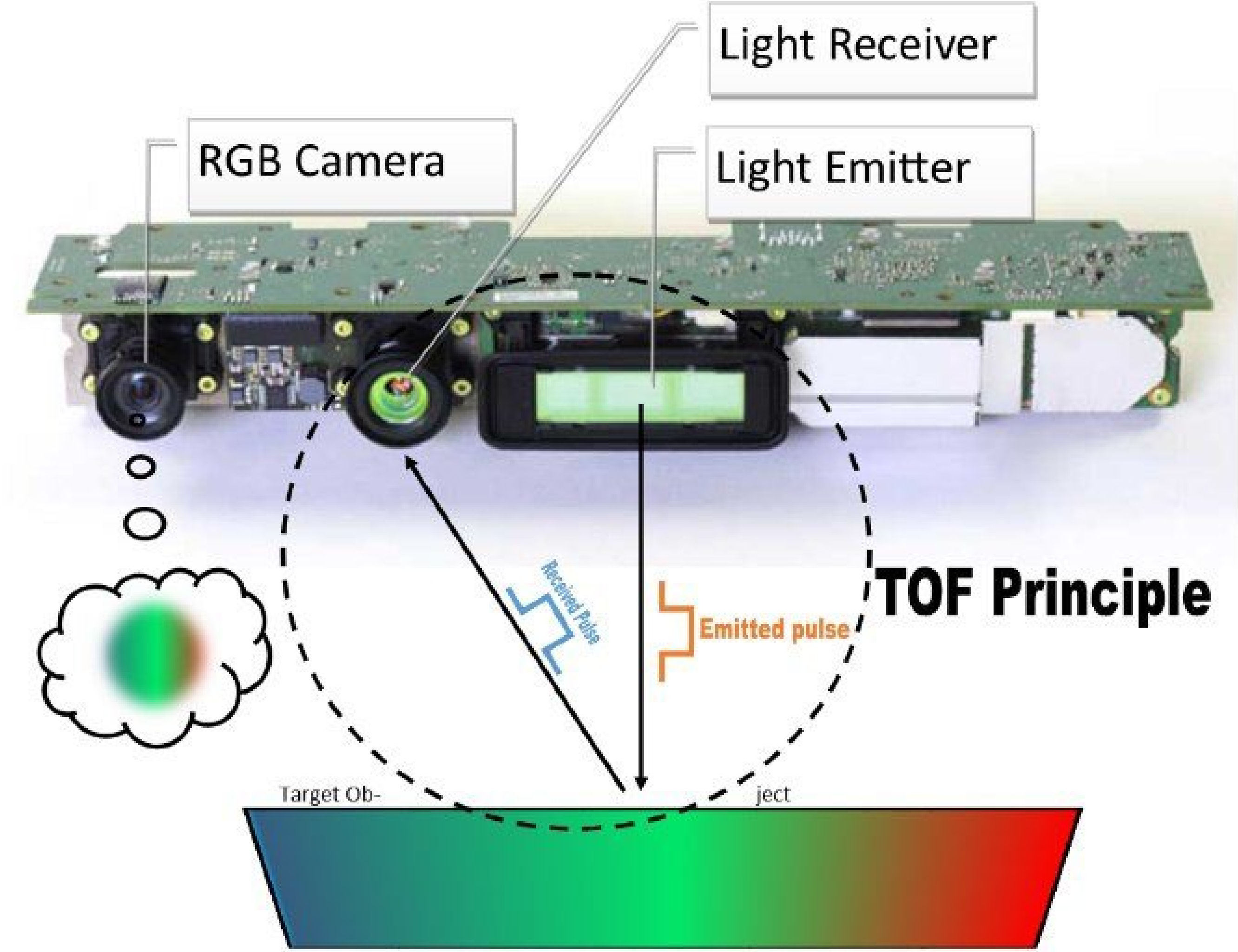
The Microsoft Kinect V2 sensor uses time-of-flight to determine depth

Feature	Kinect for Windows 1	Kinect for Windows 2
Color Camera	640 x 480 @30 fps	1920 x 1080 @30 fps
Depth Camera	320 x 240	512 x 424
Max Depth Distance	~4.5 M	~4.5 M
Min Depth Distance	40 cm in near mode	50 cm
Horizontal Field of View	57 degrees	70 degrees
Vertical Field of View	43 degrees	60 degrees
Tilt Motor	yes	no
Skeleton Joints Defined	20 joints	26 joints
Full Skeletons Tracked	2	6
USB Standard	2.0	3.0
Supported OS	Win 7, Win 8	Win 8
Price	$299	TBD

Kinect For Windows 2

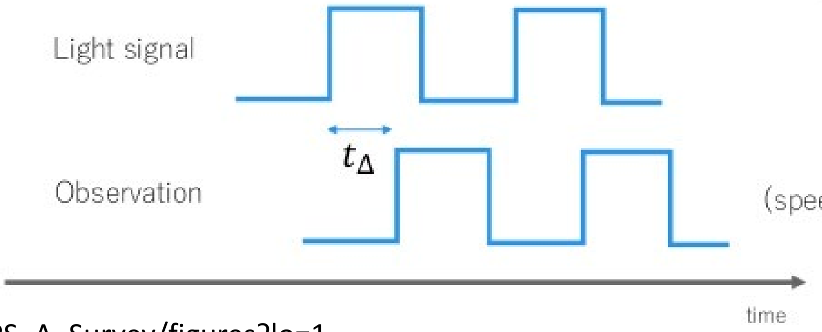
Image via http://blogs.msdn.com

zugara



Time-of-Flight (ToF) Camera

- Depth sensor based on time delay of light
- Kinect v2, Project Tango, etc.



$$d = \frac{ct_\Delta}{2}$$

(speed of light x time delay)

https://www.researchgate.net/publication/333151664_Techniques_of_Indoor_Positioning_Systems_IPS_A_Survey/figures?lo=1

The iPhone X FaceID system uses near infrared time-of-flight to acquire 3D data for face recognition

These are just a couple of examples of “time-of-flight” depth measurement technologies

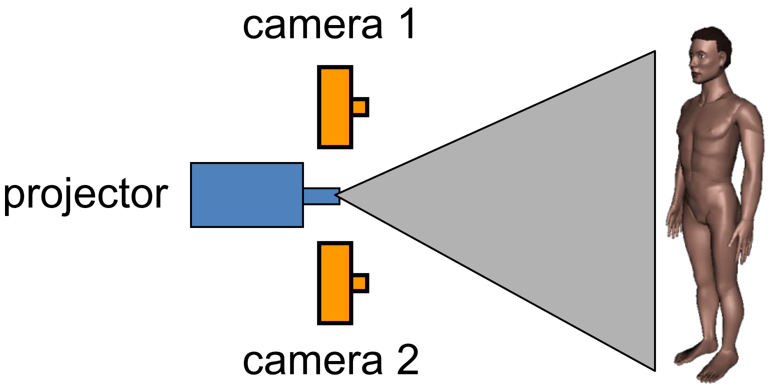
- Each has limitations in accuracy and working range
- Each also has limits on response times
- With each, the depth information must be registered to and fused with the image acquired

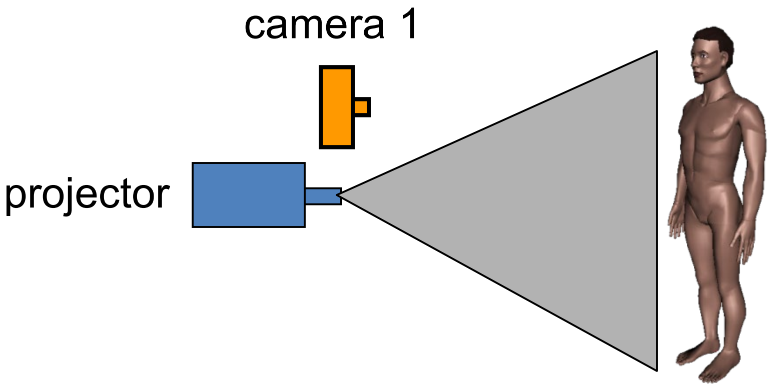
Technology	Accuracy	Coverage (m)
Ultra-wideband	cm-m	1-50
~~Bluetooth~~	~~20 m~~	~~Bluetooth MAX~~
~~RFID~~	~~dm-m~~	~~1-50~~
~~Wi-Fi~~	~~m~~	~~20-50~~
Ultrasound	cm	2-10
~~AGNSS~~	~~6-61.2 m~~	~~global~~
~~Magnetic Localization~~	~~mm-cm~~	~~1-20~~
~~INS~~	~~1 %~~	~~10-100~~
Visual Localization	0.1mm-dm	1-10
Infrared	cm-m	1-5

https://www.researchgate.net/publication/333151664_Techniques_of_Indoor_Positioning_Systems_IPS_A_Survey/figures?lo=1

Active stereo with structured light

Li Zhang's one-shot stereo





- Project known (structured) light patterns onto the object
- Greatly simplifies simplifies stereo matching

Today’s Objectives

Shape via “passive” methods

- Shape from Shading
  - Lambertian reflection
  - Equations for calculating surface normal
  - Depth from normal
  - Photometric stereo
- Stereo – Focus – Texture

Shape via “active” methods

- Laser profiling – Projection – Structured light - LIDAR

- Shape from Motion