Intro to 3D + Camera Calibration

EECS 442 – Jeong Joon Park

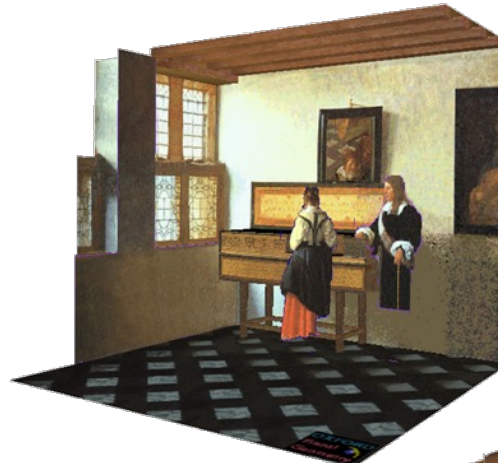
Winter 2024, University of Michigan

Administrivia

- Project proposals due today
- HW5 due next week

Our goal: Recovery of 3D structure

J. Vermeer, *Music Lesson*, 1662

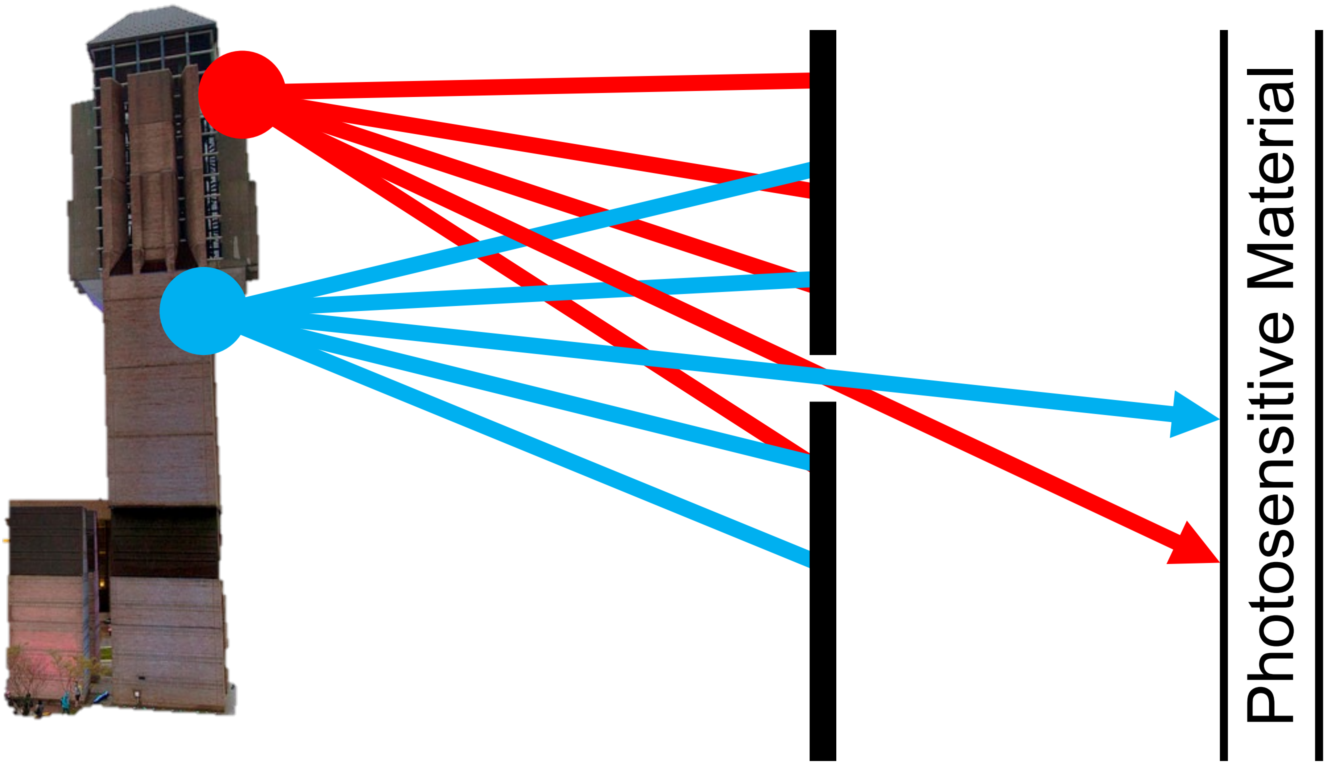


A. Criminisi, M. Kemp, and A. Zisserman, Bringing Pictorial Space to Life: computer techniques for the analysis of paintings, *Proc. Computers and the History of Art*, 2002

Next few classes

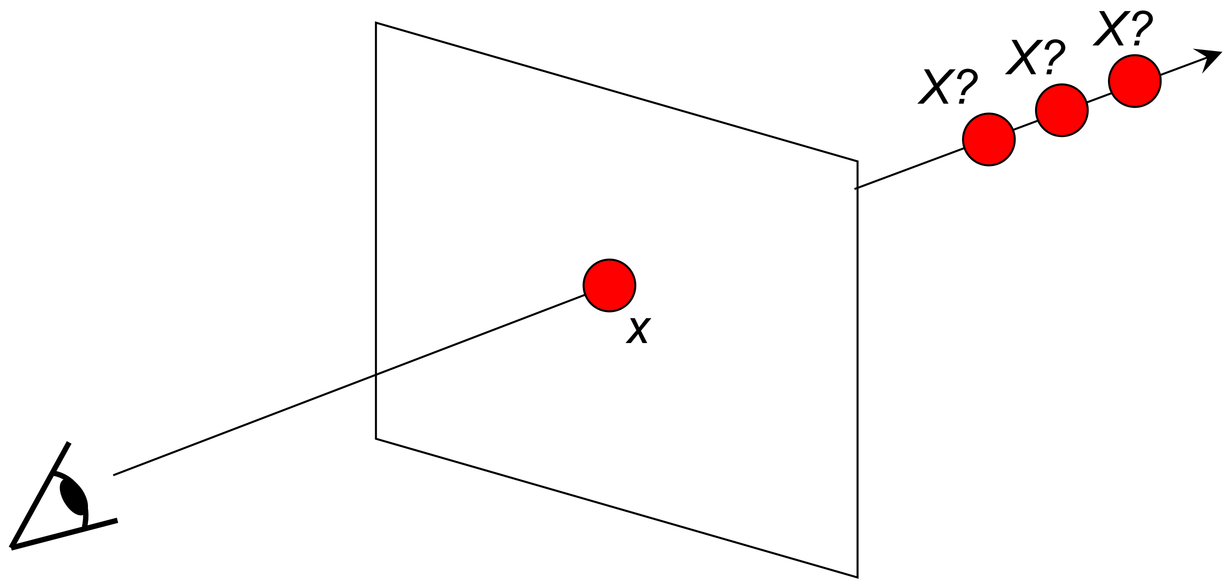
- First: some intuitions and maths about 3D vision.
- But first, a brief review

Let's Take a Picture!



Slide inspired by S. Seitz; image from Michigan Engineering

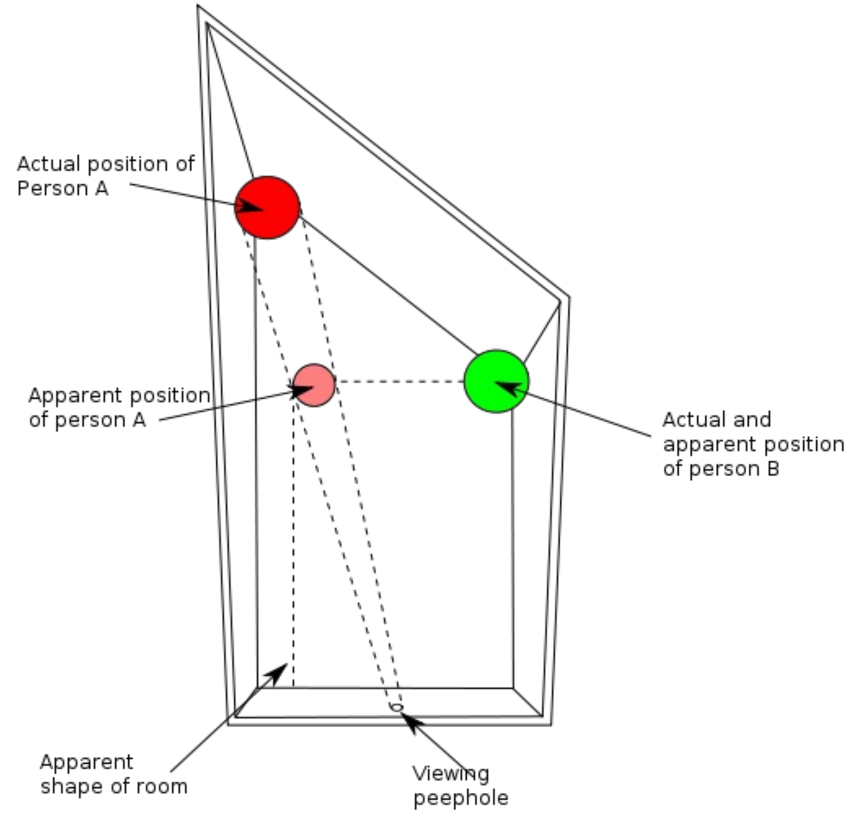
Single-view Ambiguity



- Given a *calibrated camera* and an image, we only know the ray corresponding to each pixel.
- Nowhere near enough constraints for X

Diagram credit: S. Lazebnik

Single-view Ambiguity



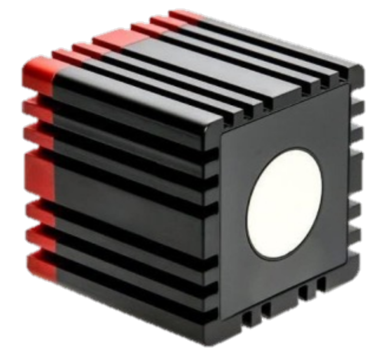
http://en.wikipedia.org/wiki/Ames_room

Slide Credit: J. Hays

Single-view Ambiguity

Diagram credit: J. Hays

Resolving Single-view Ambiguity

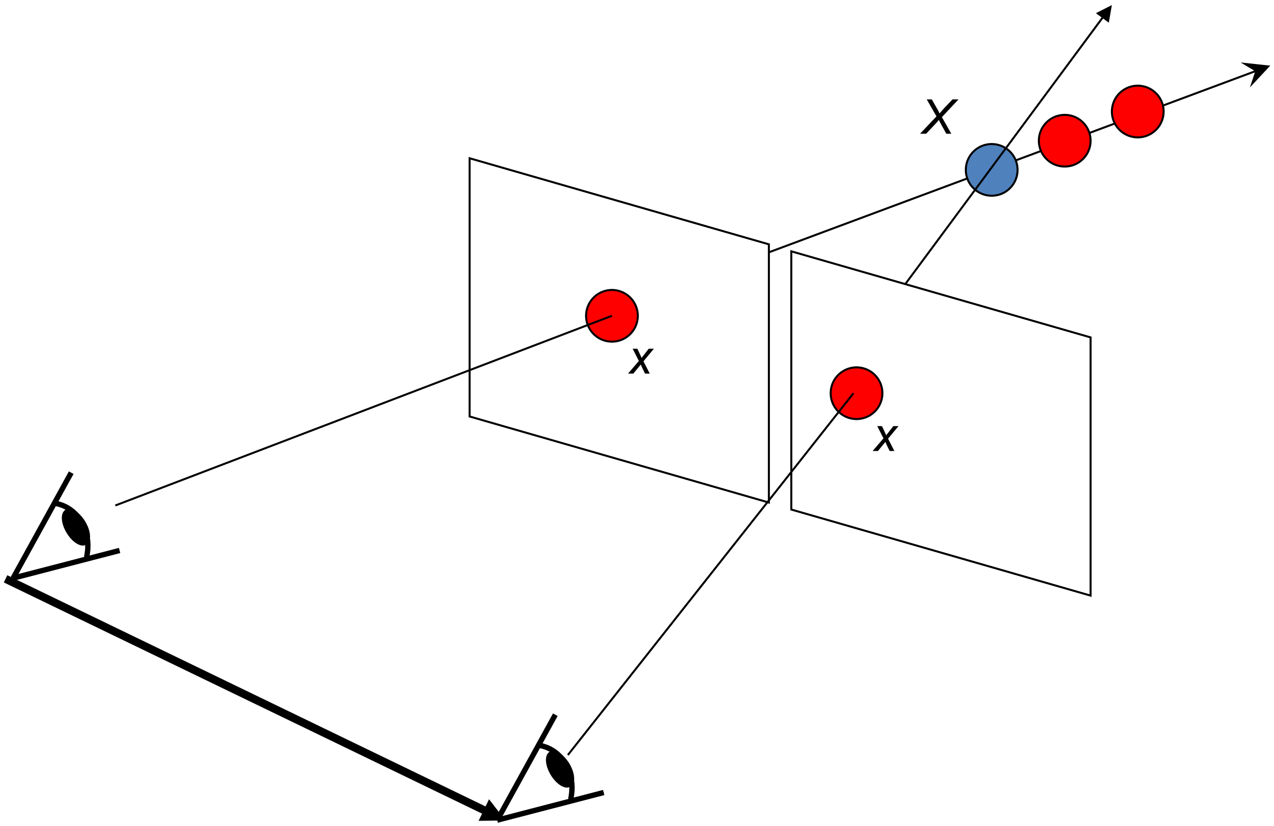


- Shoot light (lasers etc.) out of your eyes!
- Con: not so biologically plausible, dangerous?

Resolving Single-view Ambiguity

- Shoot light (lasers etc.) out of your eyes!
- Con: not so biologically plausible, dangerous?

Resolving Single-view Ambiguity

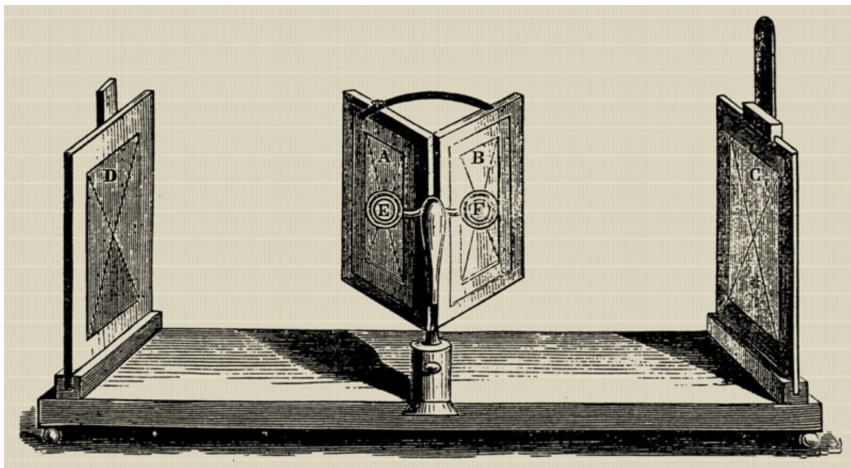


- Stereo: given 2 calibrated cameras in different views and correspondences, can solve for X

Original diagram credit: S. Lazebnik

Stereo photography and stereo viewers

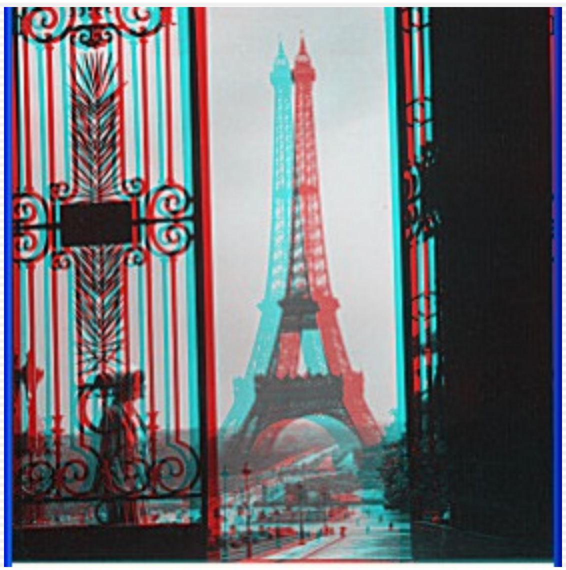
Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.



Invented by Sir Charles Wheatstone, 1838

Image from fisher-price.com

Slide credit: J. Hays



http://www.johnsonshawmuseum.org

Slide credit: J. Hays

http://www.well.com/~jimg/stereo/stereo_list.html

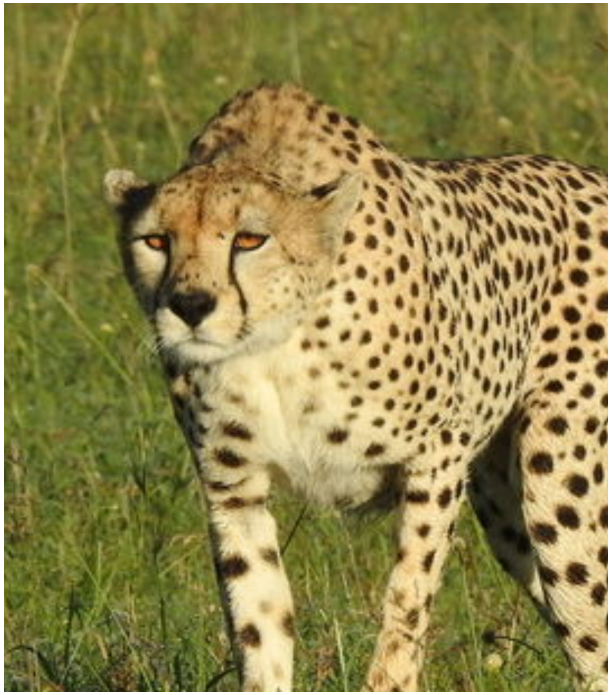
Slide credit: J. Hays

http://www.well.com/~jimg/stereo/stereo_list.html

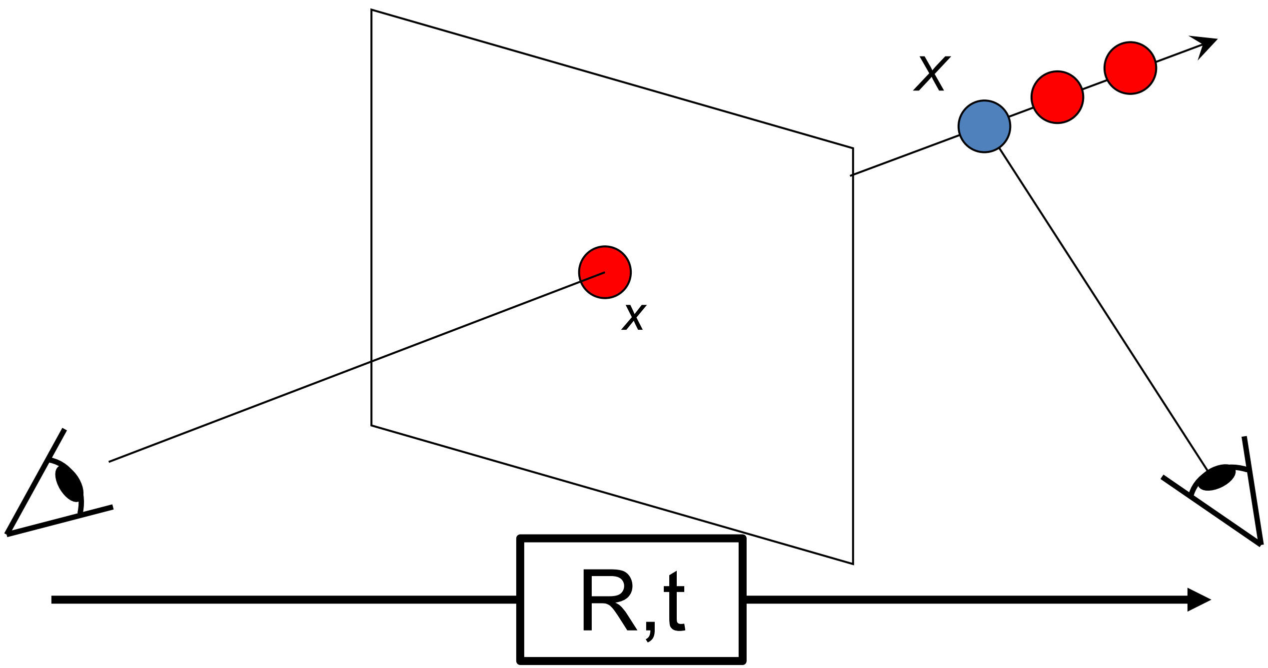
Slide credit: J. Hays

Yeah, yeah, but…

Not all animals see stereo:
Prey animals (large field of view to spot predators)
Stereoblind people



Resolving Single-view Ambiguity



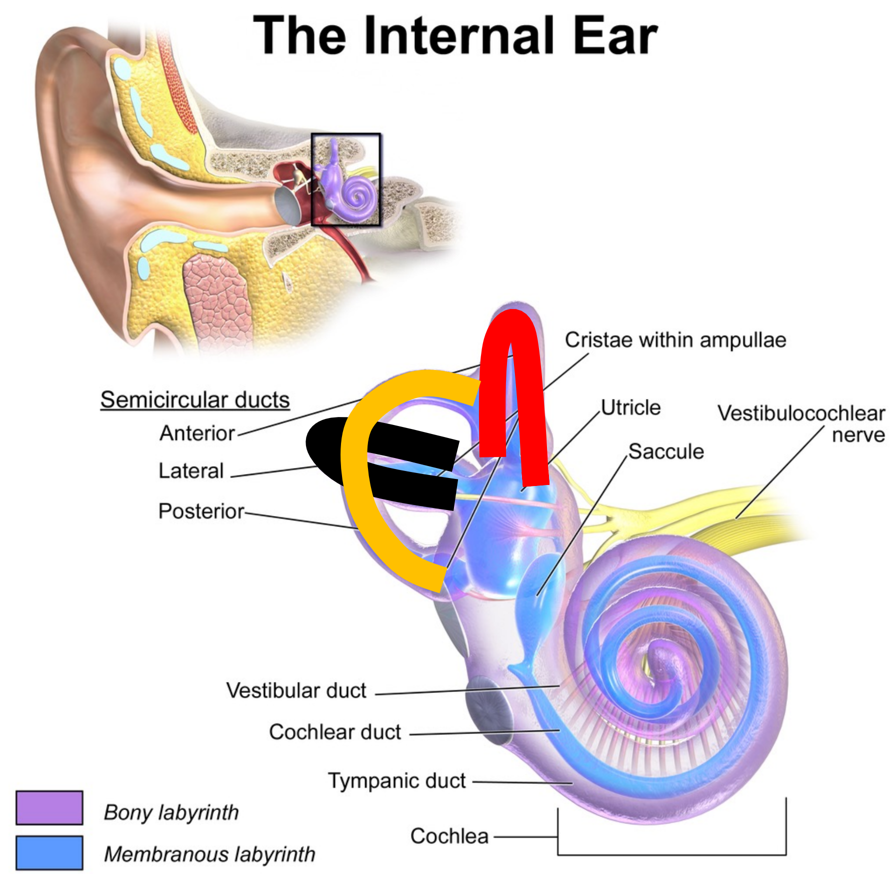
- One option: move, find correspondence.
- If you know how you moved and have a calibrated camera, can solve for X

Original diagram credit: S. Lazebnik

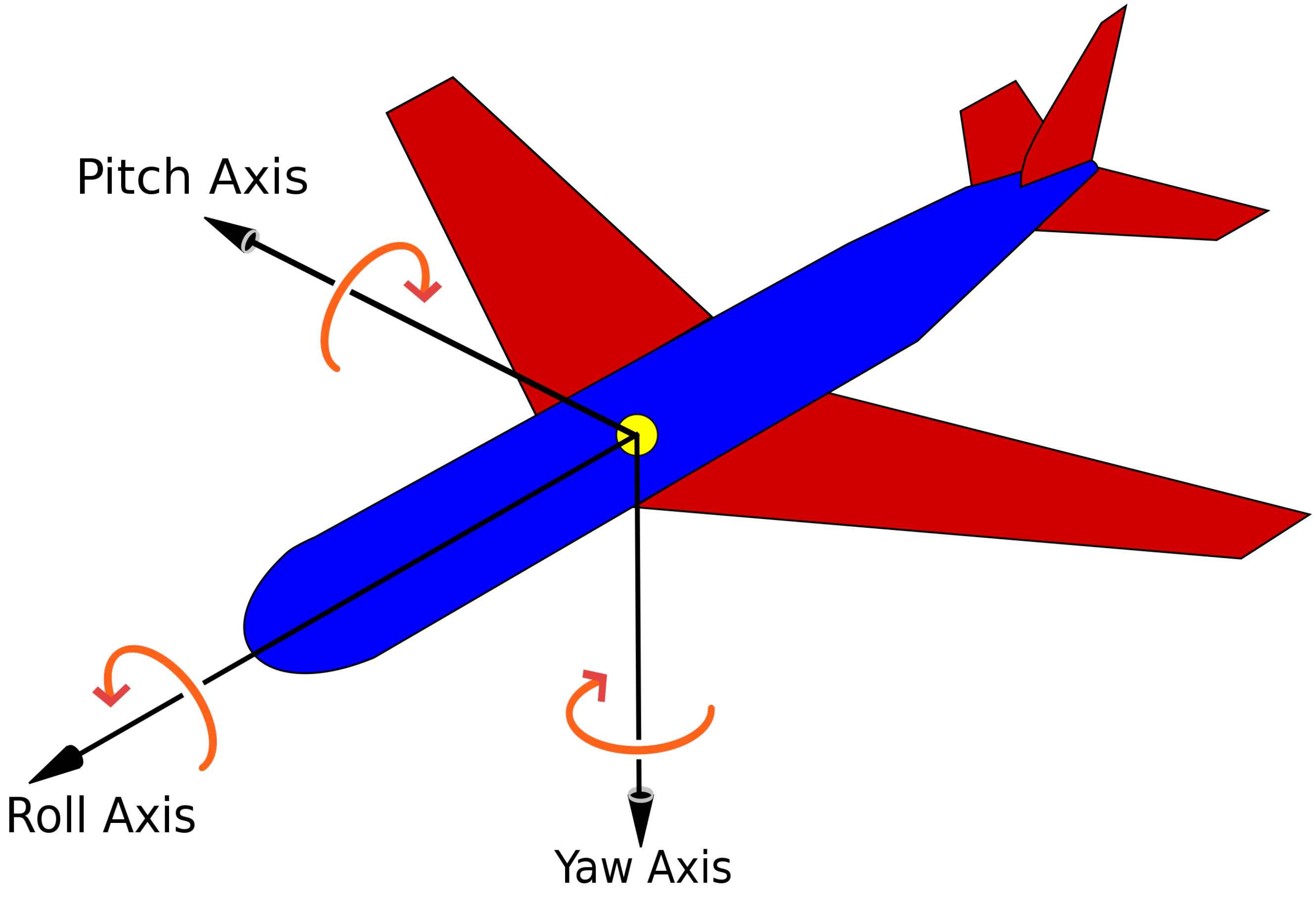
Knowing R,t

- How do you know how far you moved?
- Can solve via vision
- Can solve via ears
- **Why does your inner ear have 3 ducts?**

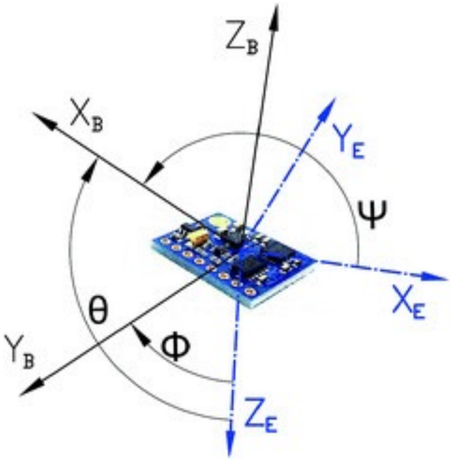
1. Nodding Yes
2. Nodding No
3. Tilting side to side

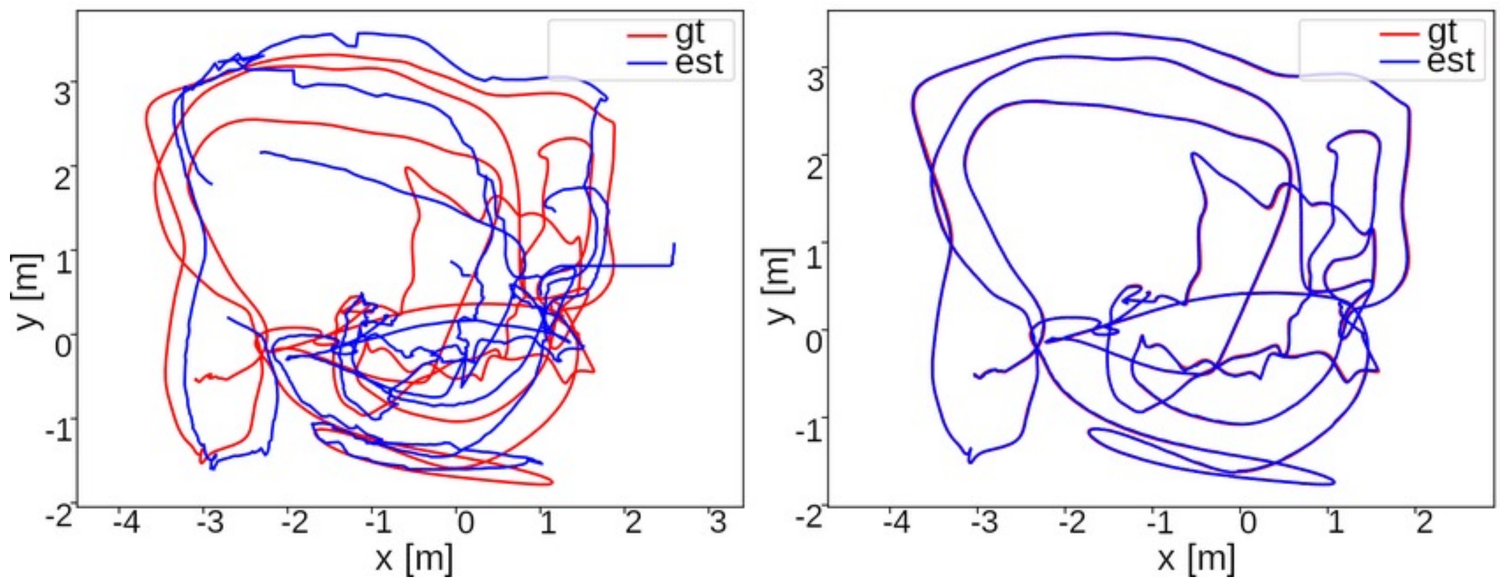


Knowing R,t



Inertial measurement unit



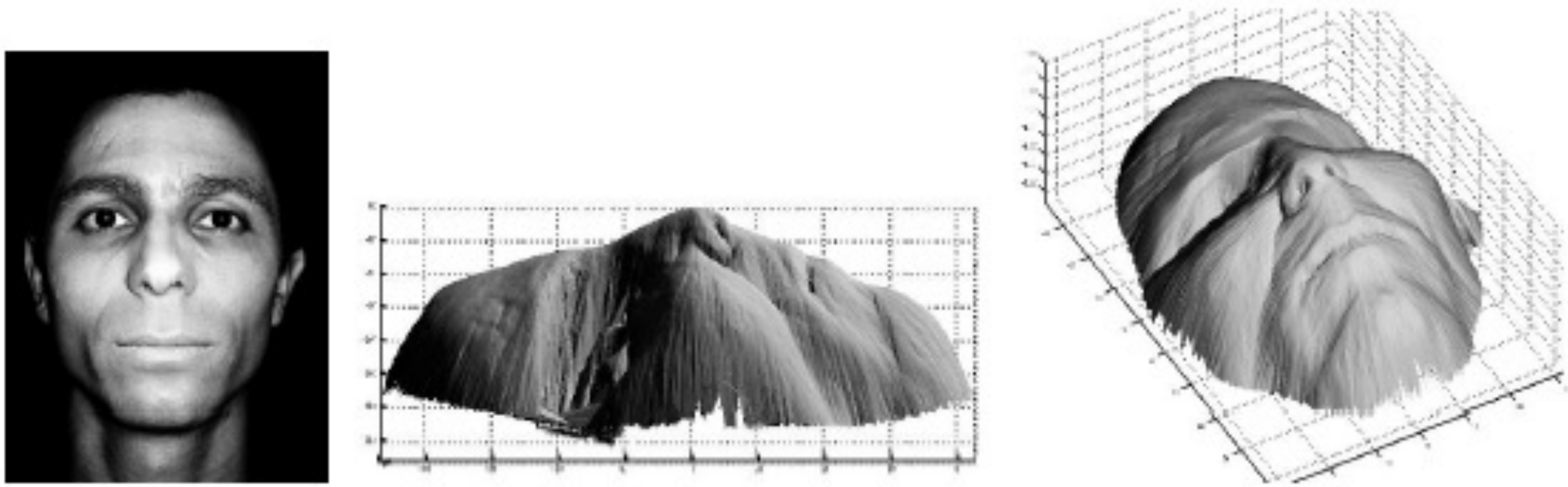


Credit: Merzlyakov et al.

Yeah, yeah, but…

You haven't been here before, yet you probably have a fairly good understanding of this scene.

Pictorial Cues – Shading



a) b) c)

[Figure from Prados & Faugeras 2006]

Pictorial Cues – Perspective effects

Image credit: S. Seitz

Pictorial Cues – Familiar Objects

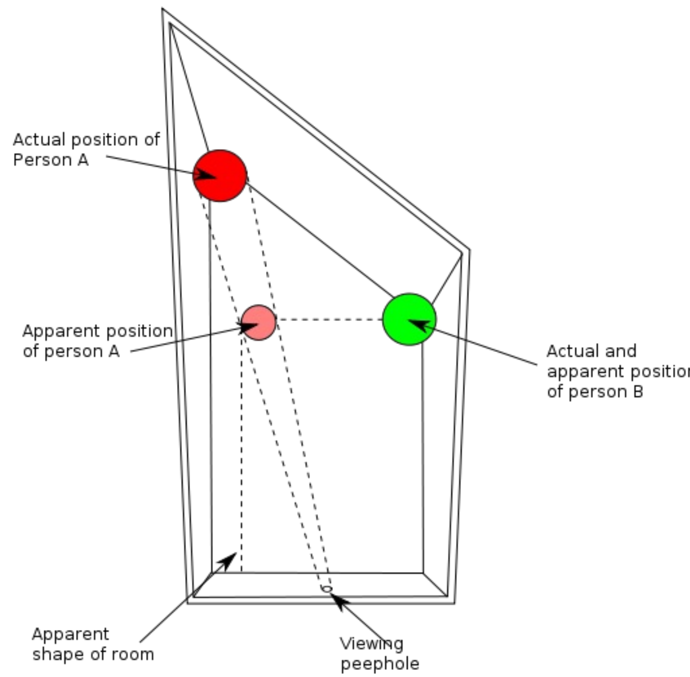
Reality of 3D Perception

- 3D perception is absurdly complex and involves integration of many cues:
  - Learned cues for 3D
  - Stereo between eyes
  - Stereo via motion
  - Integration of known motion signals to muscles (efferent copy), acceleration sensed via ears
  - Past experience of touching objects
- All connect: learned cues from 3D probably come from stereo/motion cues in large part

Really fantastic article on cues for 3D from Cutting and Vishton, 1995: https://pmvish.people.wm.edu/cutting%26vishton1995.pdf

How are Cues Combined?

Ames illusion persists (in a weaker form) even if you have stereo vision –guessing the texture is rectilinear is usually incredibly reliable

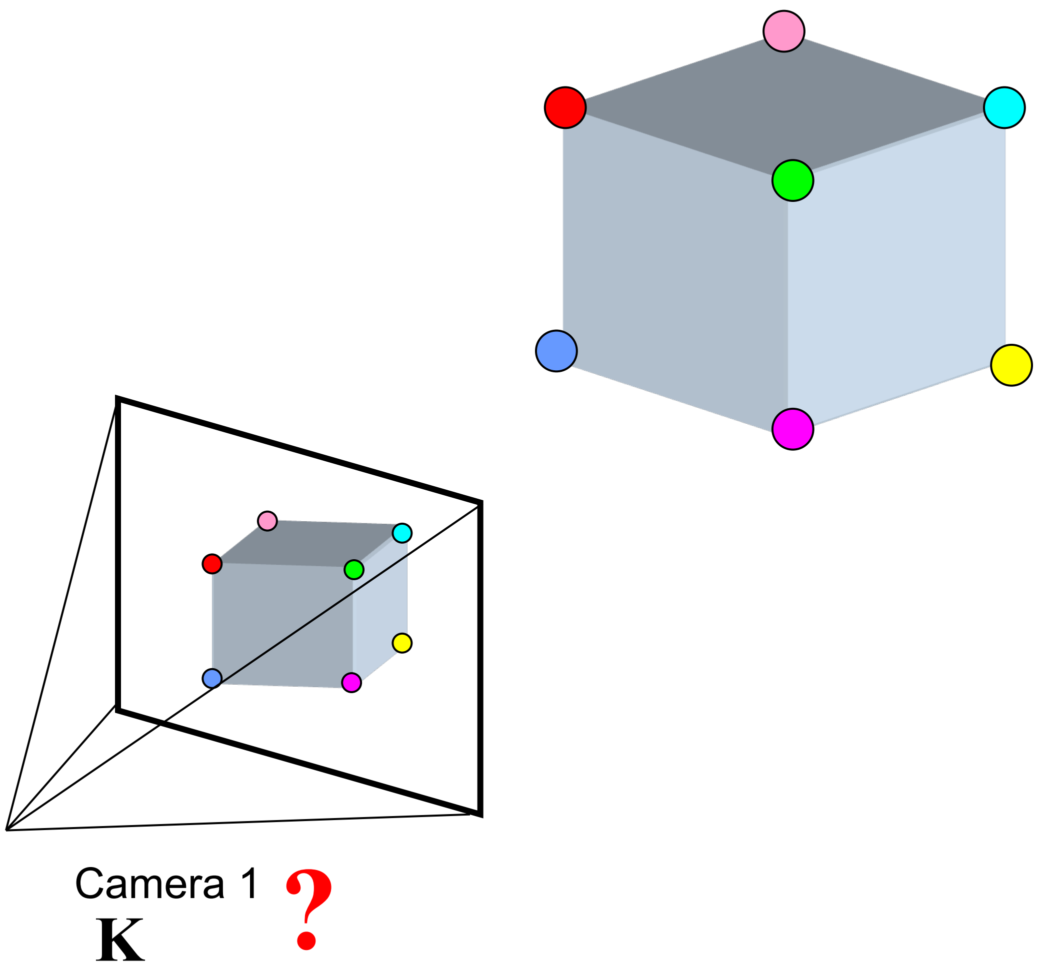


Gehringer and Engel, Journal of Experimental Psychology: Human Perception and Performance, 1986

More Formally

- Defining fundamental 3D vision problems.

Multi-view geometry problems



Calibration:
Figure out intrinsics of camera (K).

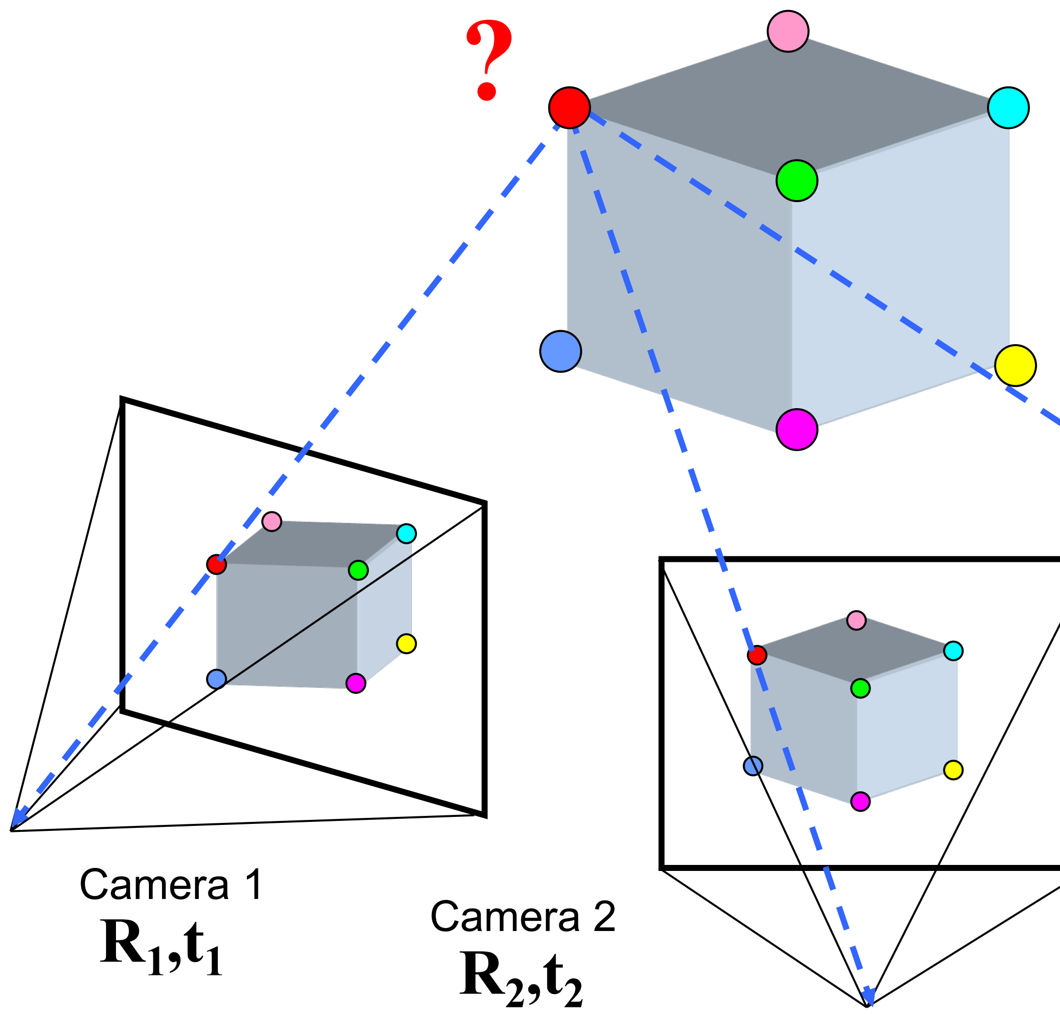
We need camera intrinsics / K in order to figure out where the rays are

Other cases assume known Ks

Slide credit: Noah Snavely

Multi-view geometry problems

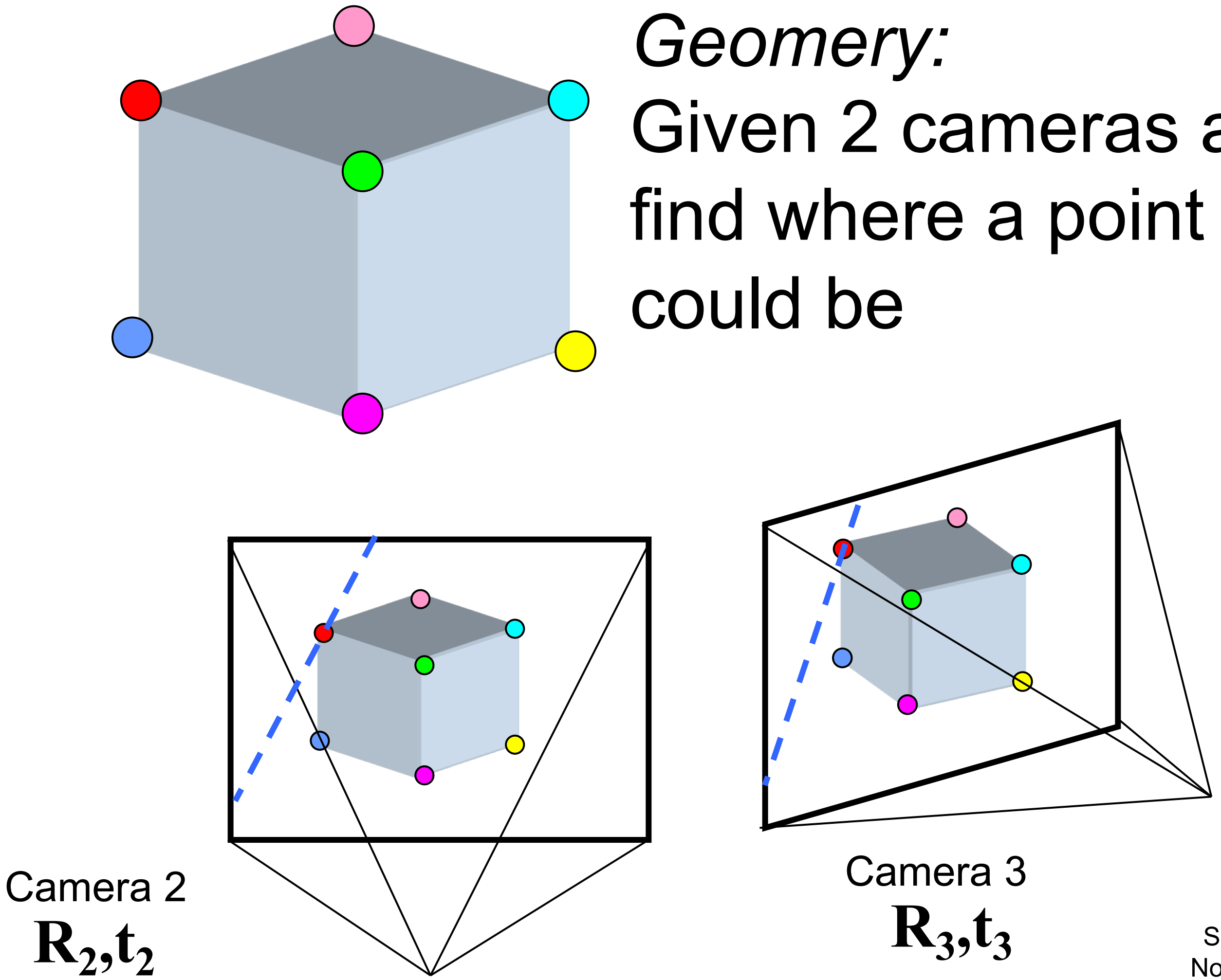
Recovering structure: Given cameras and correspondences, find 3D.



Slide credit: Noah Snavely

Multi-view geometry problems

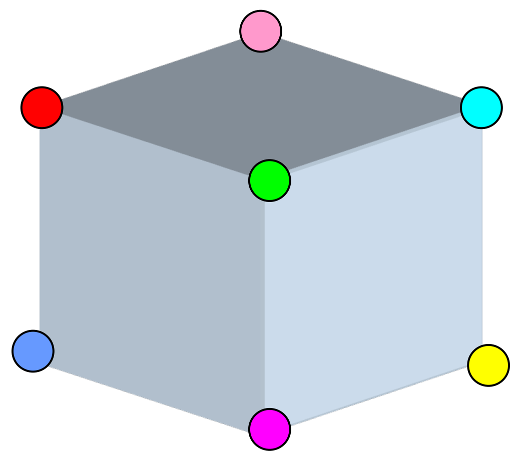
Stereo/Epipolar Geomery:
Given 2 cameras and find where a point could be

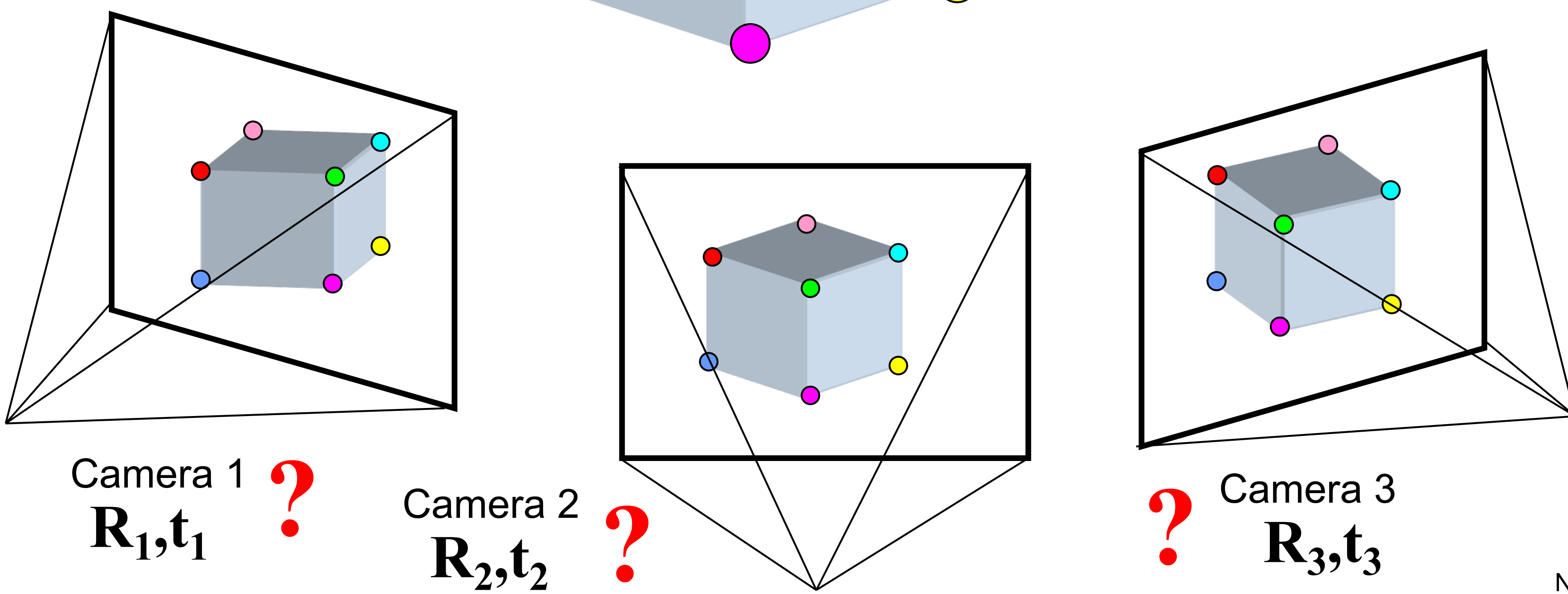


Slide credit: Noah Snavely

Multi-view geometry problems

Motion:
Figure out R, t for a set of cameras given correspondences





Slide credit: Noah Snavely

Outline

- (Today) Calibration:
  - Getting intrinsic matrix/K
- Stereo/Epipolar geometry:
  - 2 pictures → depthmap
- Structure from motion (SfM):
  - 2+ pictures → cameras, pointcloud
- Dense 3D reconstruction (NeRF):
  - 2+ pictures & cameras → dense 3D scene

Typical Perspective Model

principal point (image coords of camera origin on retina)
Just moves camera origin

focal length

rotation translation

$$\boldsymbol{p} \equiv \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{3x3} & \boldsymbol{t}_{3x1} \end{bmatrix} \boldsymbol{X}_{4x1}$$

2D Projection of X

3D point

Camera Calibration

$$\boldsymbol{p} \equiv \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{3x3} & \boldsymbol{t}_{3x1} \end{bmatrix} \boldsymbol{X}_{4x1}$$

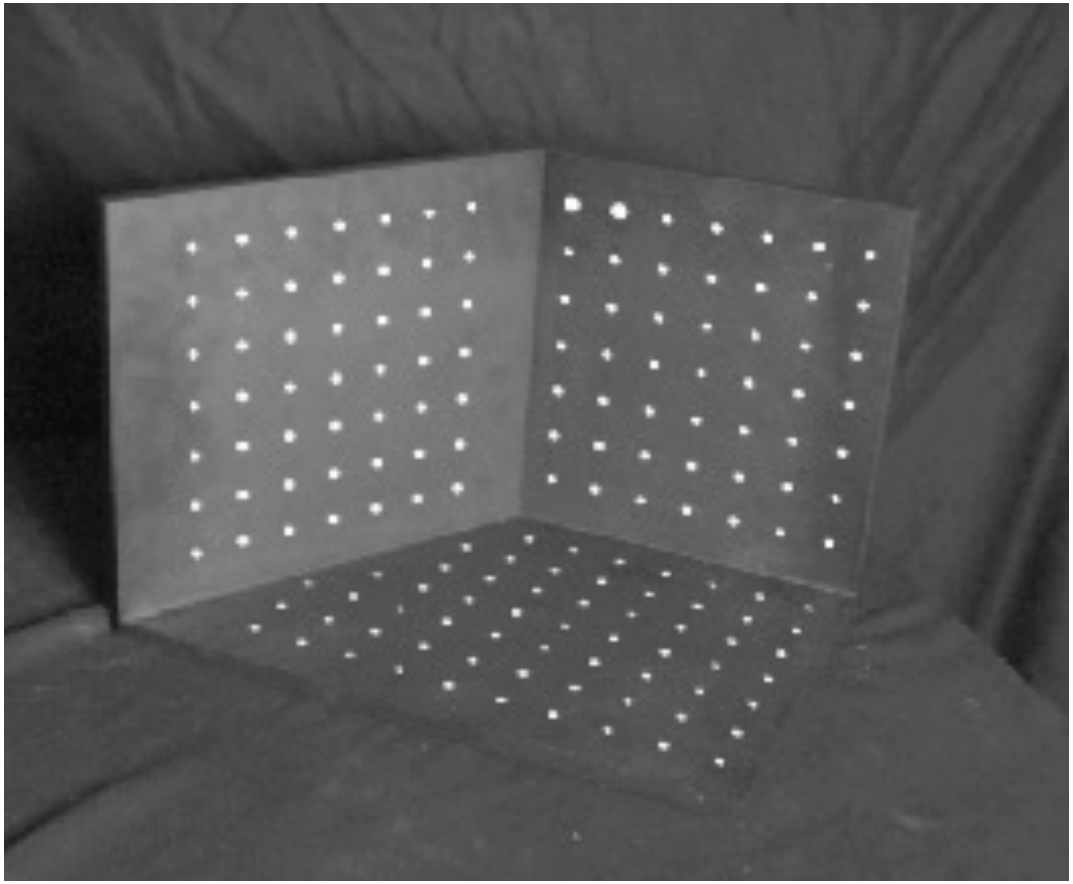
$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \equiv \boldsymbol{M}_{3x4} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Pairs of [X,Y,Z] and [u,v] $\rightarrow$ eqns to constrain **M**

How do I get [X,Y,Z], [u,v]?

Camera Calibration

A funny object with multiple planes.



Camera Calibration Targets

Using a tape measure

Known 2d image coords

880	214
43	203
270	197
886	347
745	302
943	128
476	590
419	214
317	335
783	521
235	427
665	429
655	362
427	333
412	415
746	351
434	415
525	234
716	308
602	187

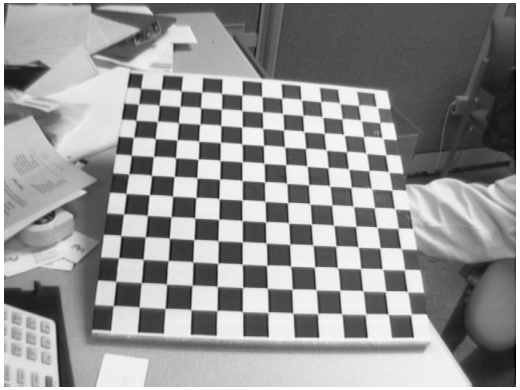
Known 3d locations

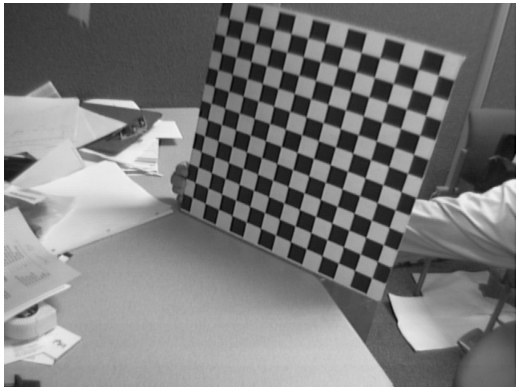
312.747	309.140	30.086
305.796	311.649	30.356
307.694	312.358	30.418
310.149	307.186	29.298
311.937	310.105	29.216
311.202	307.572	30.682
307.106	306.876	28.660
309.317	312.490	30.230
307.435	310.151	29.318
308.253	306.300	28.881
306.650	309.301	28.905
308.069	306.831	29.189
309.671	308.834	29.029
308.255	309.955	29.267
307.546	308.613	28.963
311.036	309.206	28.913
307.518	308.175	29.069
309.950	311.262	29.990
312.160	310.772	29.080
311.988	312.709	30.514

Image credit: J. Hays

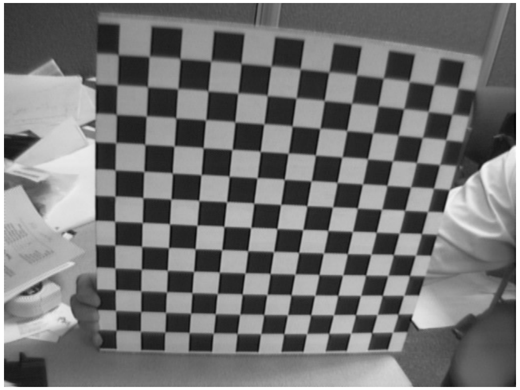
Camera Calibration Targets

A set of views of a plane (not covered today)



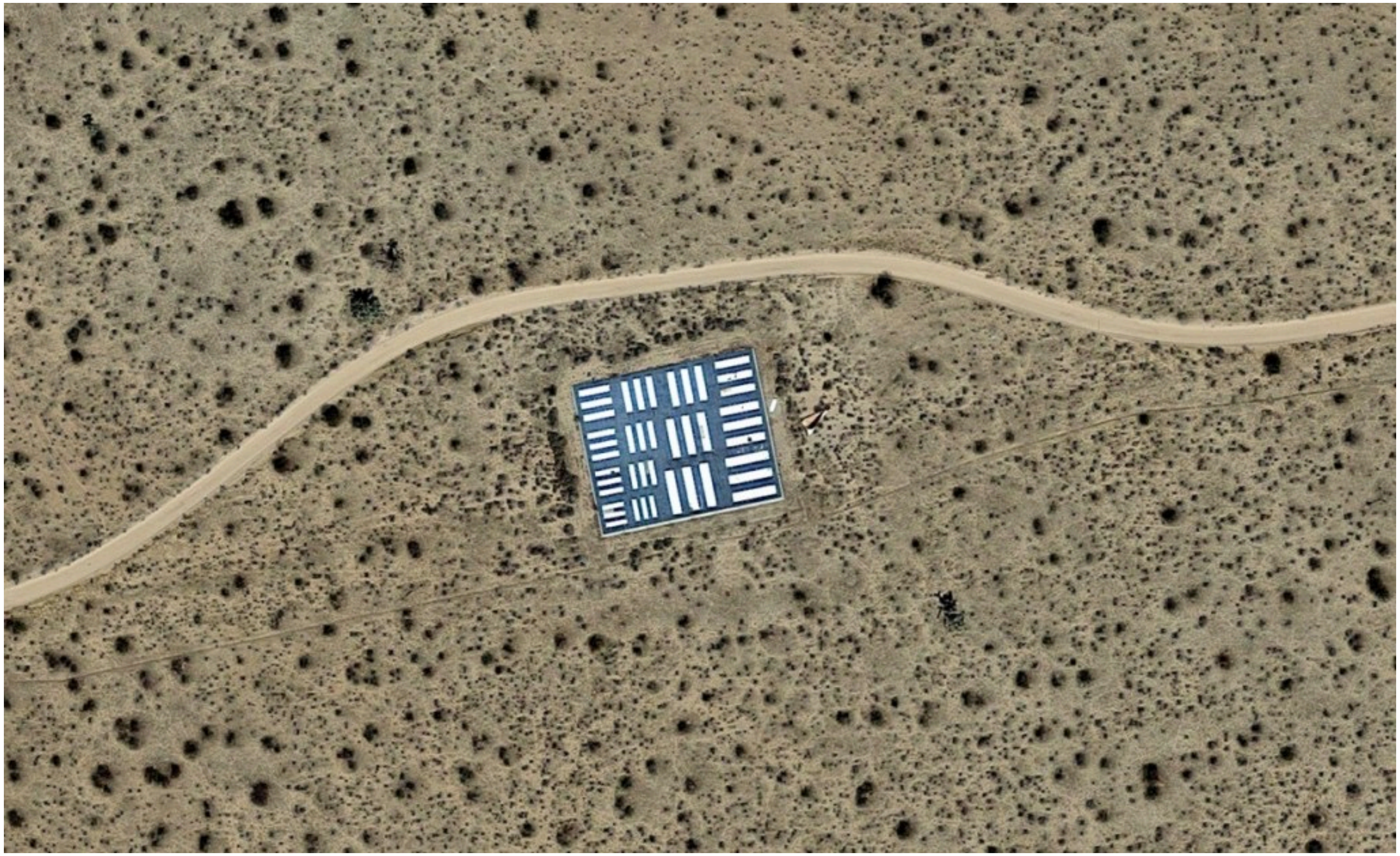


. . .



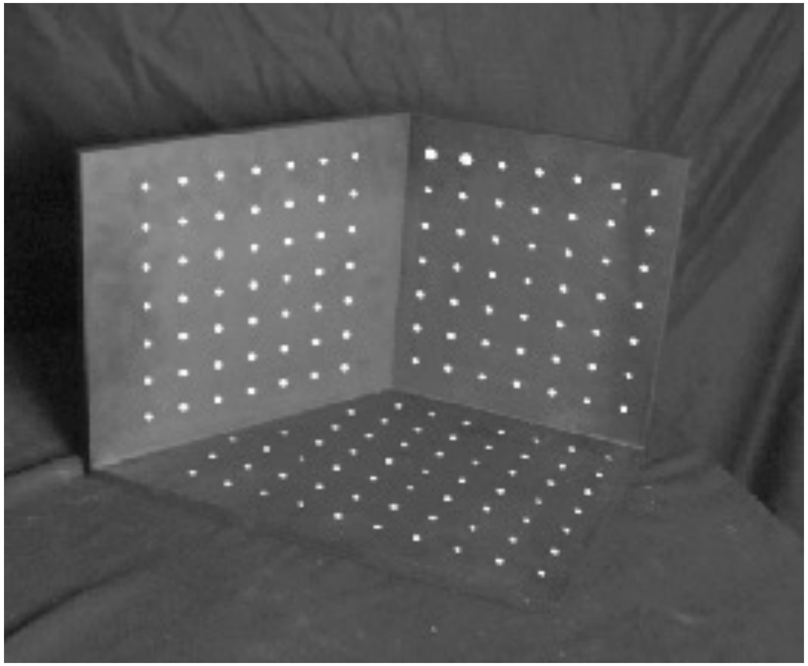
Camera Calibration Targets

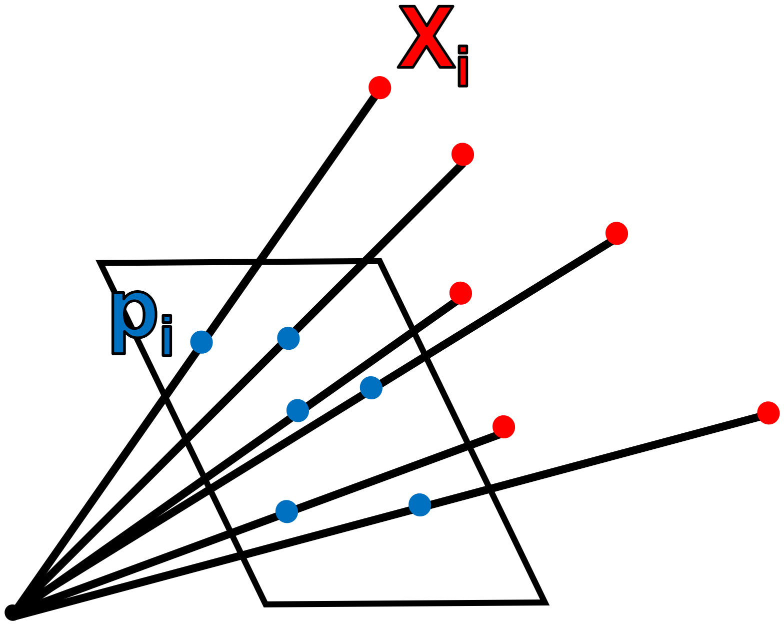
A single, huge plane. **What's this for?**



Camera calibration

- Given *n* points with known 3D coordinates $\mathbf{X}_i$ and known image projections $\mathbf{p}_i$, estimate the camera parameters





Slide credit: S. Lazebnik

Camera Calibration: Linear Method

$$\boldsymbol{p_i} \equiv \boldsymbol{M X_i}$$

Remember homogenous coordinate:

$$\boldsymbol{p_i} \rightarrow [\boldsymbol{u_i}, \boldsymbol{v_i}] = [\frac{(\boldsymbol{M X_i})_1}{(\boldsymbol{M X_i})_3}, \frac{(\boldsymbol{M X_i})_2}{(\boldsymbol{M X_i})_3}]$$

$$= [\frac{\boldsymbol{X_i^T M_1}}{\boldsymbol{X_i^T M_3}}, \frac{\boldsymbol{X_i^T M_2}}{\boldsymbol{X_i^T M_3}}]$$

M1, M2, M3 : the three rows of the projection matrix M

Camera Calibration: Linear Method

$$\boldsymbol{p_i} \equiv \boldsymbol{M X_i}$$

Remember homogenous coordinate:

$$\boldsymbol{p_i} \rightarrow [\boldsymbol{u_i}, \boldsymbol{v_i}] = [\frac{\boldsymbol{X_i^T M_1}}{\boldsymbol{X_i^T M_3}}, \frac{\boldsymbol{X_i^T M_2}}{\boldsymbol{X_i^T M_3}}]$$

$$\boldsymbol{X_i^T M_1 - u_i\, X_i^T M_3 = 0}$$
$$\boldsymbol{X_i^T M_2 - v_i\, X_i^T M_3 = 0}$$

M1, M2, M3 : the three rows of the projection matrix M

Camera Calibration: Linear Method

$$\boldsymbol{p_i \equiv M X_i}$$

Remember homogenous coordinate:

$$\boldsymbol{X_i^T M_1 - u_i\, X_i^T M_3 = 0}$$
$$\boldsymbol{X_i^T M_2 - v_i\, X_i^T M_3 = 0}$$

Very similar equations to homographies!

M1, M2, M3 : the three rows of the projection matrix M

Camera Calibration: Linear Method

$$\boldsymbol{p_i} \equiv \boldsymbol{M X_i}$$

Remember homogenous coordinate:

$$\boldsymbol{X_i^T M_1 - u_i\, X_i^T M_3 = 0}$$

$$\boldsymbol{X_i^T M_2 - v_i\, X_i^T M_3 = 0}$$

How many equations per [u,v] + [X,Y,Z] pair?

2

If M is 3x4, how many degrees of freedom?

11

Minimum number of points required?

6

Camera Calibration: Linear Method

$$\begin{bmatrix} \mathbf{0}^T & \boldsymbol{X}_1^T & -v_1\boldsymbol{X}_1^T \\ \boldsymbol{X}_1^T & \mathbf{0}^T & -u_1\boldsymbol{X}_1^T \\ \cdots & \cdots & \cdots \\ \mathbf{0}^T & \boldsymbol{X}_n^T & -v_1\boldsymbol{X}_n^T \\ \boldsymbol{X}_n^T & \mathbf{0}^T & -u_n\boldsymbol{X}_n^T \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_1^T \\ \boldsymbol{M}_2^T \\ \boldsymbol{M}_3^T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

How do we solve problems of the form

$\arg\min \|\boldsymbol{An}\|_2^2, \|\boldsymbol{n}\|_2^2 = 1$ **?**

Eigenvector of **A^{T}A** with smallest eigenvalue

Derivation from L. Lazebnik; note we negate one of the equations from the cross product

In Practice

We want: $\boldsymbol{p} \equiv \boldsymbol{K}_{3x3}[\boldsymbol{R}_{3x3}, \boldsymbol{t}_{3x1}] \quad \boldsymbol{X}_{4x1}$

We get: $\boldsymbol{p} \equiv \boldsymbol{M}_{3x4}\boldsymbol{X}_{4x1}$

What's the difference between K[R,t] and M?

Unfactored. Often times we want K

In Practice

We want: $\boldsymbol{p} \equiv \boldsymbol{K}_{3x3}[\boldsymbol{R}_{3x3}, \boldsymbol{t}_{3x1}] \quad \boldsymbol{X}_{4x1}$

We get: $\boldsymbol{p} \equiv \boldsymbol{M}_{3x4}\boldsymbol{X}_{4x1}$

What's the difference between K[R,t] and M?

$$\boldsymbol{M} = [\boldsymbol{M}_{3x3}|\boldsymbol{t}_{3x1}]$$

$$\boldsymbol{M}_{3x3} = \boldsymbol{K}\boldsymbol{R}$$

Solution: RQ-decomposition on left-most 3x3 matrix
→ finite options of a upper triangular matrix * orthogonal matrix (rotation)

In Practice

If $\mathbf{p}_i = \mathbf{MX}_i$ is overconstrained, the objective function isn't actually the one you care about. Geometrically, it measures co-linearity.

Instead:

1) initialize parameters with linear model
2) Apply off-the-shelf non-linear optimizer (e.g., gradient descent) to:

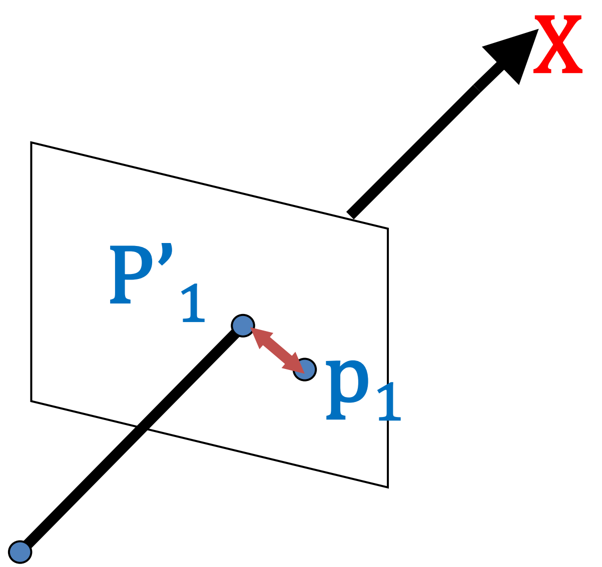
$$\sum \|\mathrm{proj}(\boldsymbol{MX_i}) - [u_i, v_i]^T\|_2^2$$

Another Advantage: can also add radial distortion and other constraints

In Practice

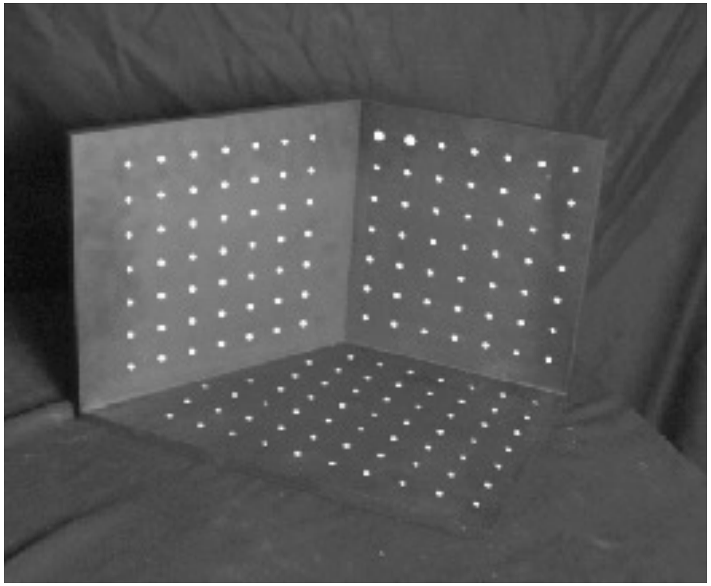
Direct reprojection loss:
2) Minimize Euclidean distance between predicted and ground-truth 2D points

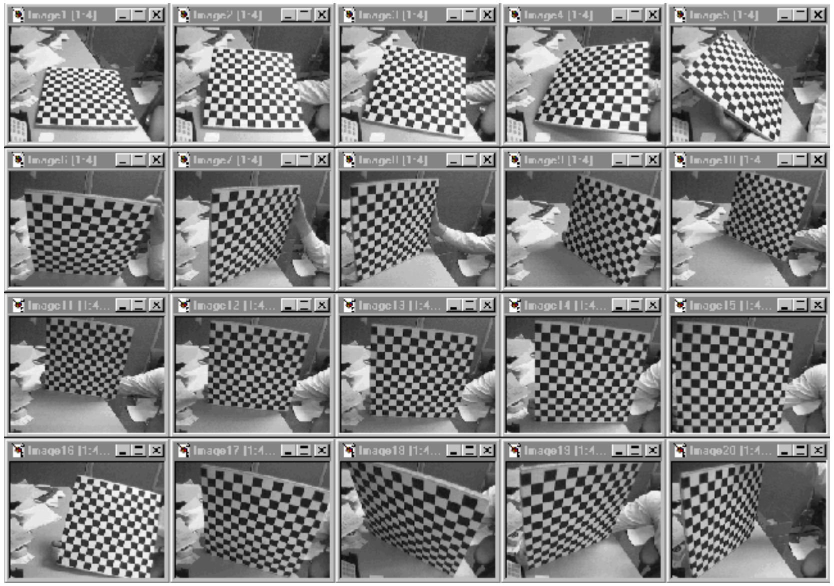
$$\sum \|\mathrm{proj}(\boldsymbol{M}\boldsymbol{X}_{\boldsymbol{i}}) - [u_i, v_i]^T\|_2^2$$



In Practice

Degenerate configurations (e.g., all points on one plane) an issue. Usually need multiplane targets.





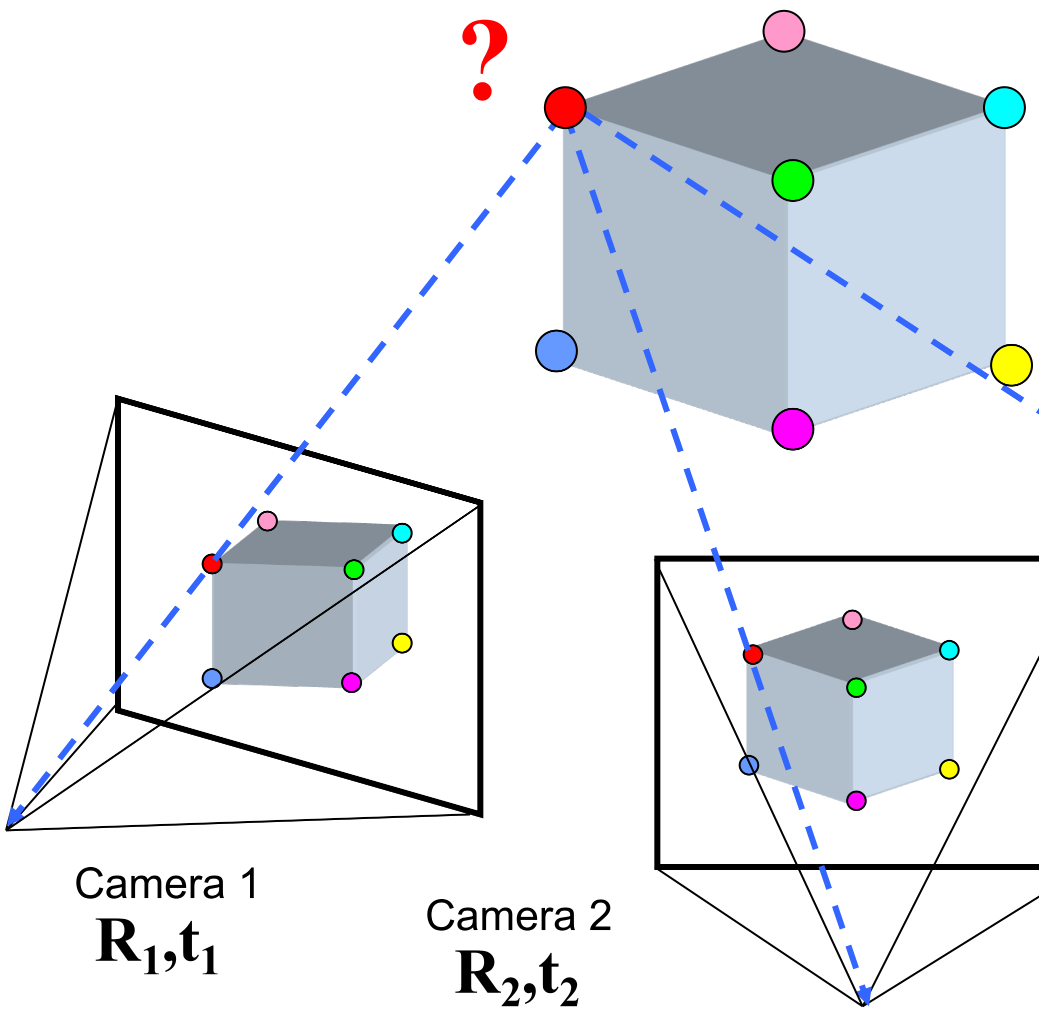
Supplementary read on multi-plane calibration: [Zhang et al.]()

What Does This Get You?

Given projection $\mathbf{p_i}$ of unknown 3D point $\mathbf{X}$ in two or more images (with known cameras $\mathbf{M_i}$), find $\mathbf{X}$

Multi-view geometry problems

Recovering structure: Given cameras and correspondences, find 3D.

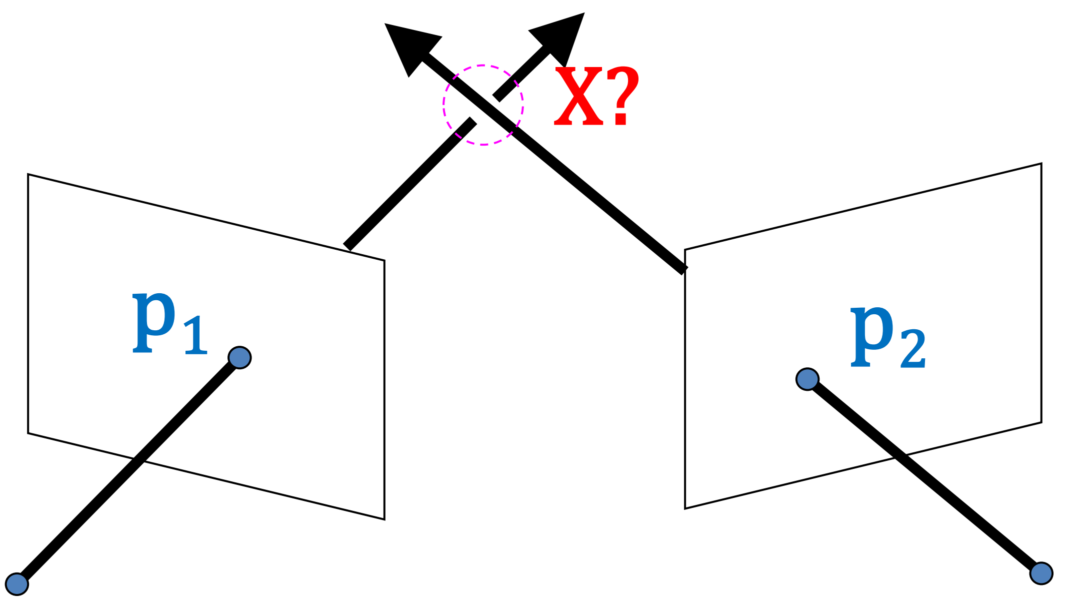


Slide credit: Noah Snavely

Triangulation

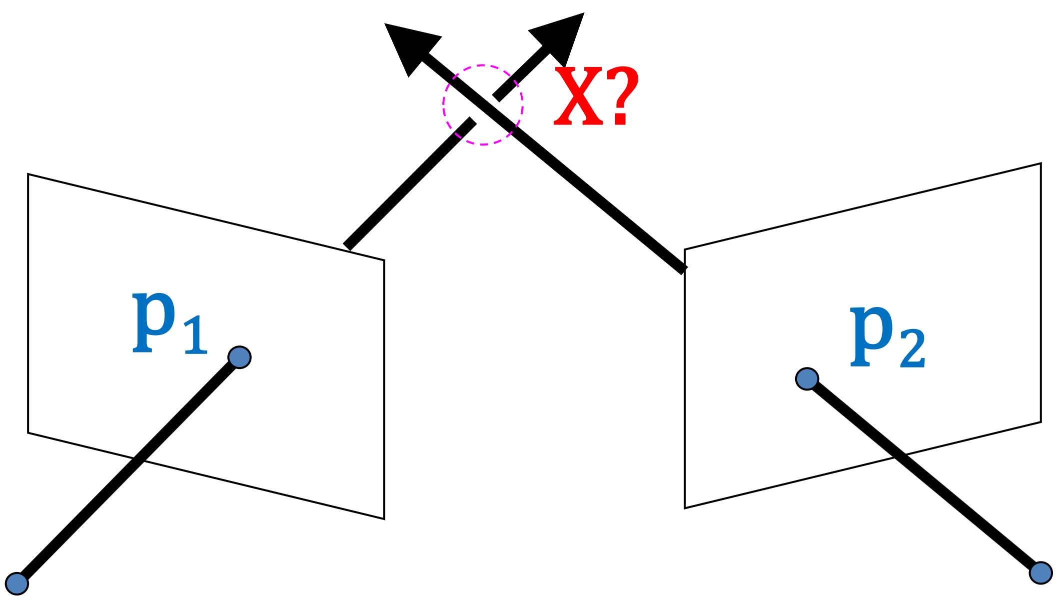
Given projection $\mathbf{p_i}$ of unknown 3D point $\mathbf{X}$ in two or more images (with known cameras $\mathbf{M_i}$), find $\mathbf{X}$

Why is the calibration here important?



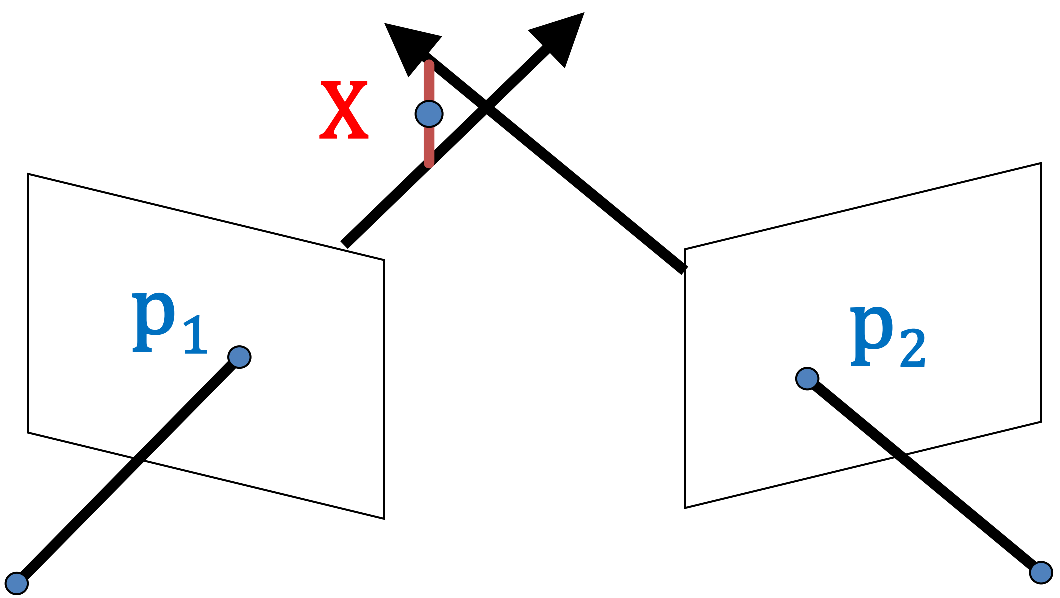
Triangulation

Rays in principle should intersect, but in practice usually don't exactly due to noise, numerical errors.



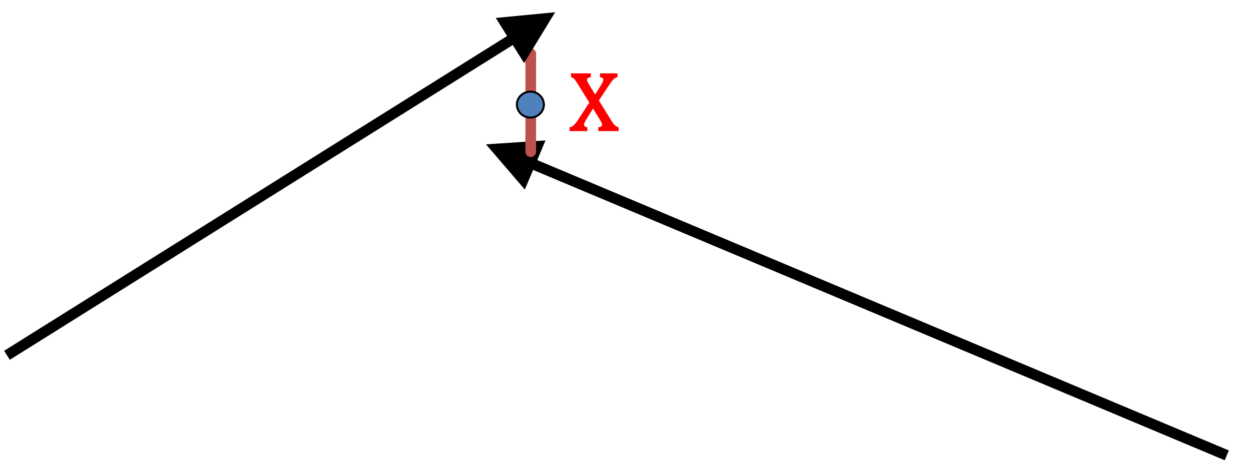
Triangulation – Geometry

Find shortest segment between viewing rays, set **X** to be the midpoint of the segment.



Triangulation – Geometry

Find shortest segment between viewing rays, set **X** to be the midpoint of the segment.



Triangulation – Linear Optimization

Solving for X given p's and M's
Another way of looking:
p and MX should be colinear/parallel

$$\boldsymbol{p_1} \equiv \boldsymbol{M_1 X} \quad \Rightarrow \quad \boldsymbol{p_1} \times \boldsymbol{M_1 X} = \boldsymbol{0} \quad \Rightarrow \quad [\boldsymbol{p_{1x}}]\boldsymbol{M_1 X} = \boldsymbol{0}$$

$$\boldsymbol{p_2} \equiv \boldsymbol{M_2 X} \quad \quad \boldsymbol{p_2} \times \boldsymbol{M_2 X} = \boldsymbol{0} \quad \quad [\boldsymbol{p_{2x}}]\boldsymbol{M_2 X} = \boldsymbol{0}$$

Cross Prod.
as matrix

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = [\boldsymbol{a}_x]\boldsymbol{b}$$

Triangulation – Linear Optimization

$$\boldsymbol{p_1} \equiv \boldsymbol{M_1 X} \quad \rightarrow \quad \boldsymbol{p_1} \times \boldsymbol{M_1 X} = \mathbf{0} \quad \rightarrow \quad [\boldsymbol{p_{1x}}]\boldsymbol{M_1 X} = \mathbf{0}$$

$$\boldsymbol{p_2} \equiv \boldsymbol{M_2 X} \quad \rightarrow \quad \boldsymbol{p_2} \times \boldsymbol{M_2 X} = \mathbf{0} \quad \rightarrow \quad [\boldsymbol{p_{2x}}]\boldsymbol{M_2 X} = \mathbf{0}$$

Cross Prod.
as matrix

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = [\boldsymbol{a}_x]\boldsymbol{b}$$

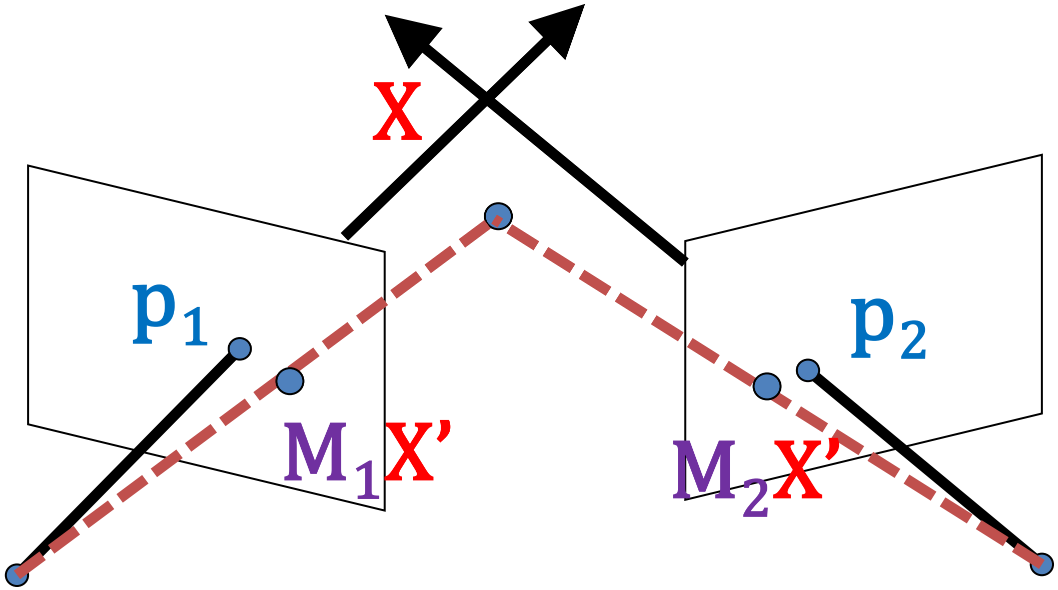
$$[\boldsymbol{p_{1x}}]\boldsymbol{M_1 X} = \mathbf{0} \quad \rightarrow \quad ([\boldsymbol{p_{1x}}]\boldsymbol{M_1})\boldsymbol{X} = \mathbf{0}$$

$$[\boldsymbol{p_{2x}}]\boldsymbol{M_2 X} = \mathbf{0} \quad \rightarrow \quad ([\boldsymbol{p_{2x}}]\boldsymbol{M_2})\boldsymbol{X} = \mathbf{0}$$

Two equations per camera (u,v)
for 3 unknowns in X (x,y,z)

Triangulation – Non-linear Optim.

Find X minimizing $d(\boldsymbol{p}_1, \boldsymbol{M}_1\boldsymbol{X'})^2 + d(\boldsymbol{p}_2, \boldsymbol{M}_2\boldsymbol{X'})^2$



Summarizing

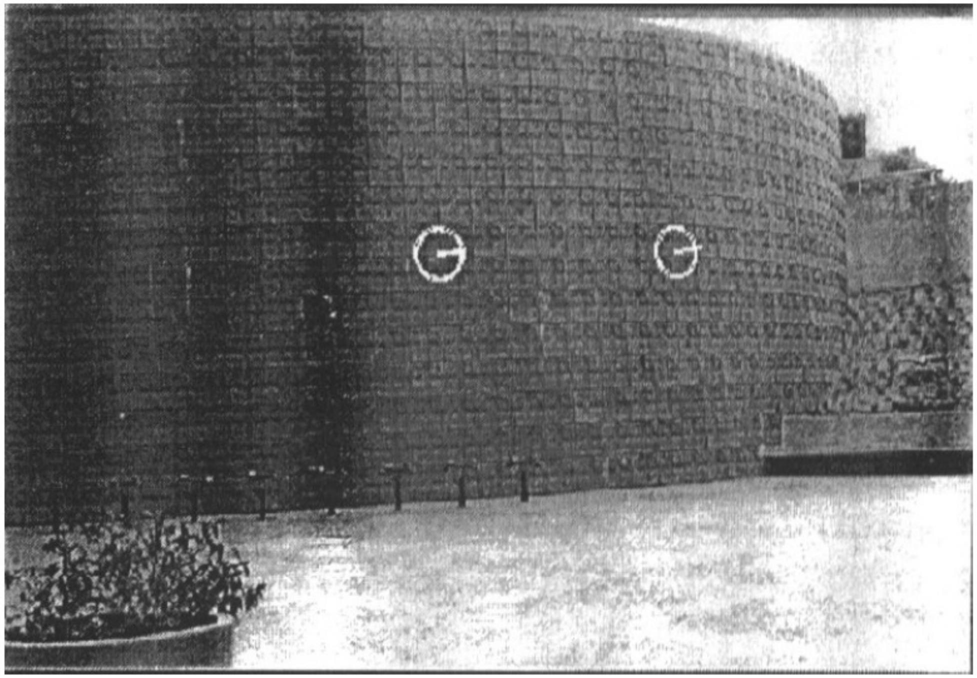
- 3D is complicated
- Given **p = MX**, you can derive equations that let you solve for **M** (calibration) or **X** (triangulation)
- Next time: two-view stereo

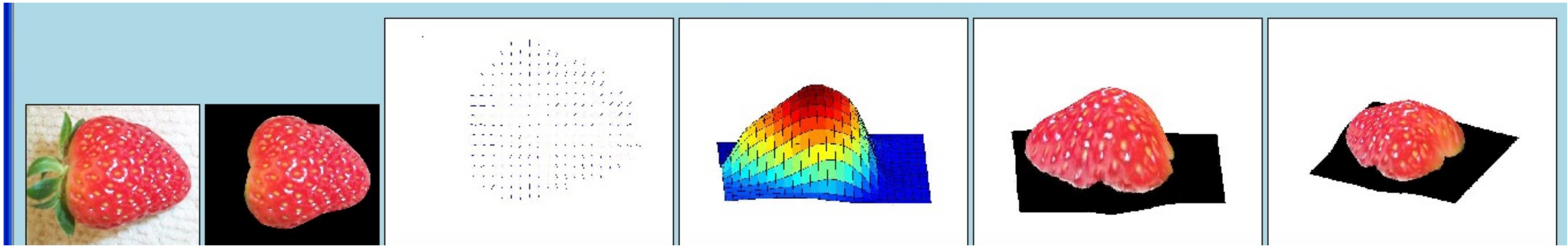
Bonus Fun

Single-view Ambiguity

Rashad Alakbarov shadow sculptures

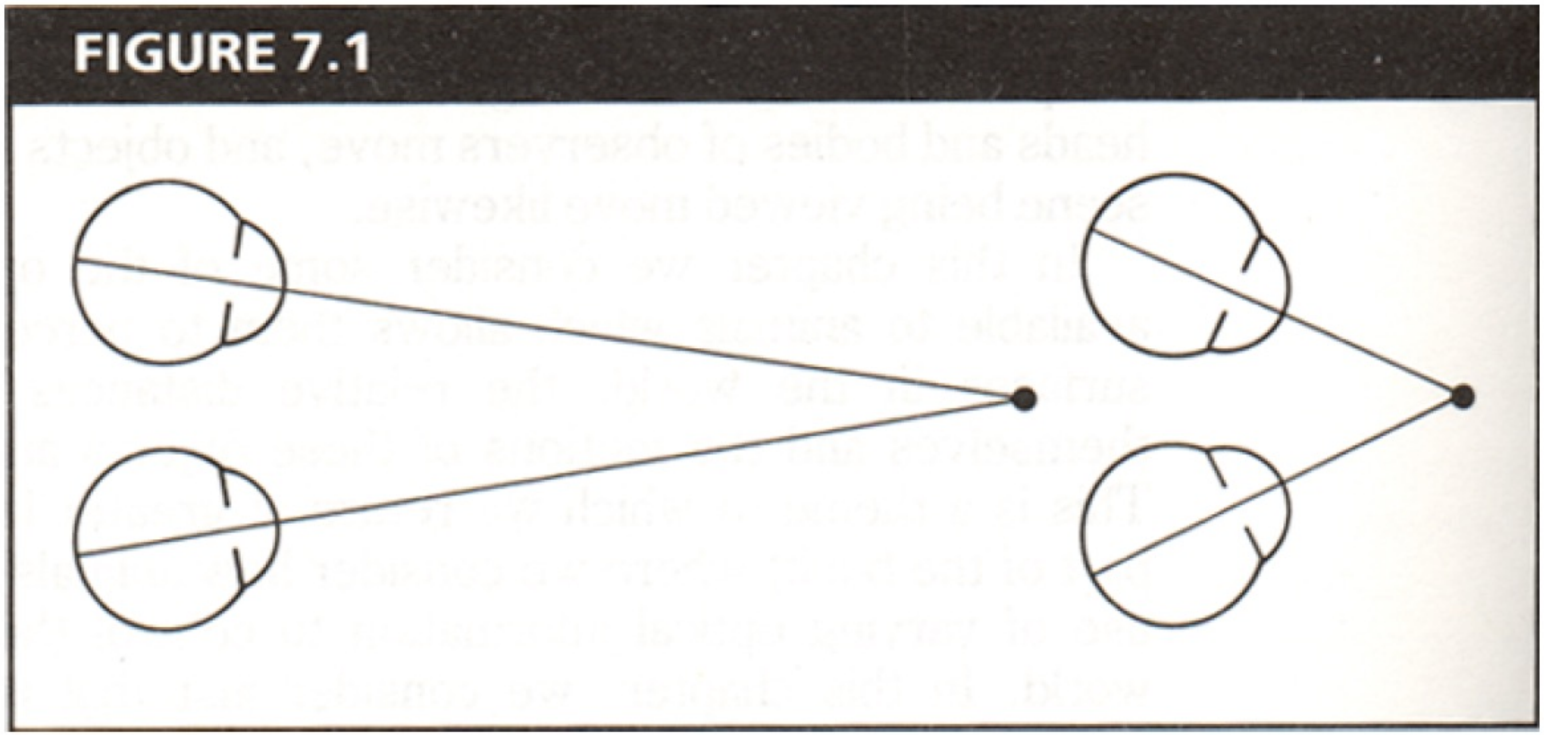
Pictorial Cues – Texture





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

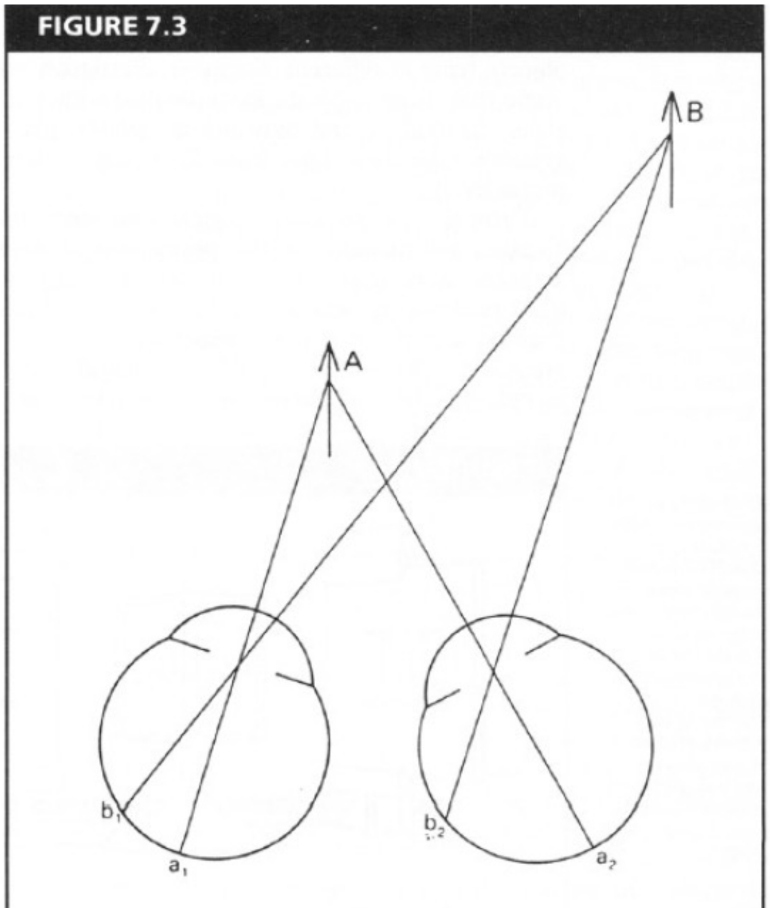
Human stereopsis: disparity



From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

Human eyes **fixate** on point in space – rotate so that corresponding images form in centers of fovea.

Human stereopsis: disparity



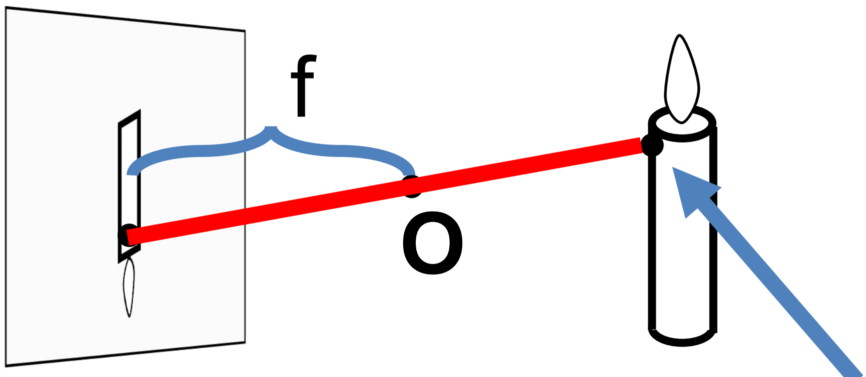
From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

Disparity occurs when eyes fixate on one object; others appear at different visual angles

Adapted from David Forsyth, UC Berkeley

Projection Matrix

Projection (fx/z, fy/z) is matrix multiplication



$$\begin{bmatrix} fx \\ fy \\ z \end{bmatrix} \equiv \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} fx/z \\ fy/z \end{bmatrix}$$

Slide inspired from L. Lazebnik