

# Intro to 3D + Camera Calibration

EECS 442 – Prof. David Fouhey

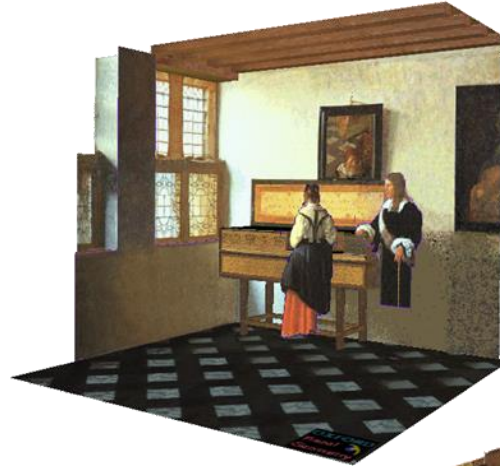
Winter 2019, University of Michigan

[http://web.eecs.umich.edu/~fouhey/teaching/EECS442\\_W19/](http://web.eecs.umich.edu/~fouhey/teaching/EECS442_W19/)

# Our goal: Recovery of 3D structure



J. Vermeer, *Music Lesson*, 1662

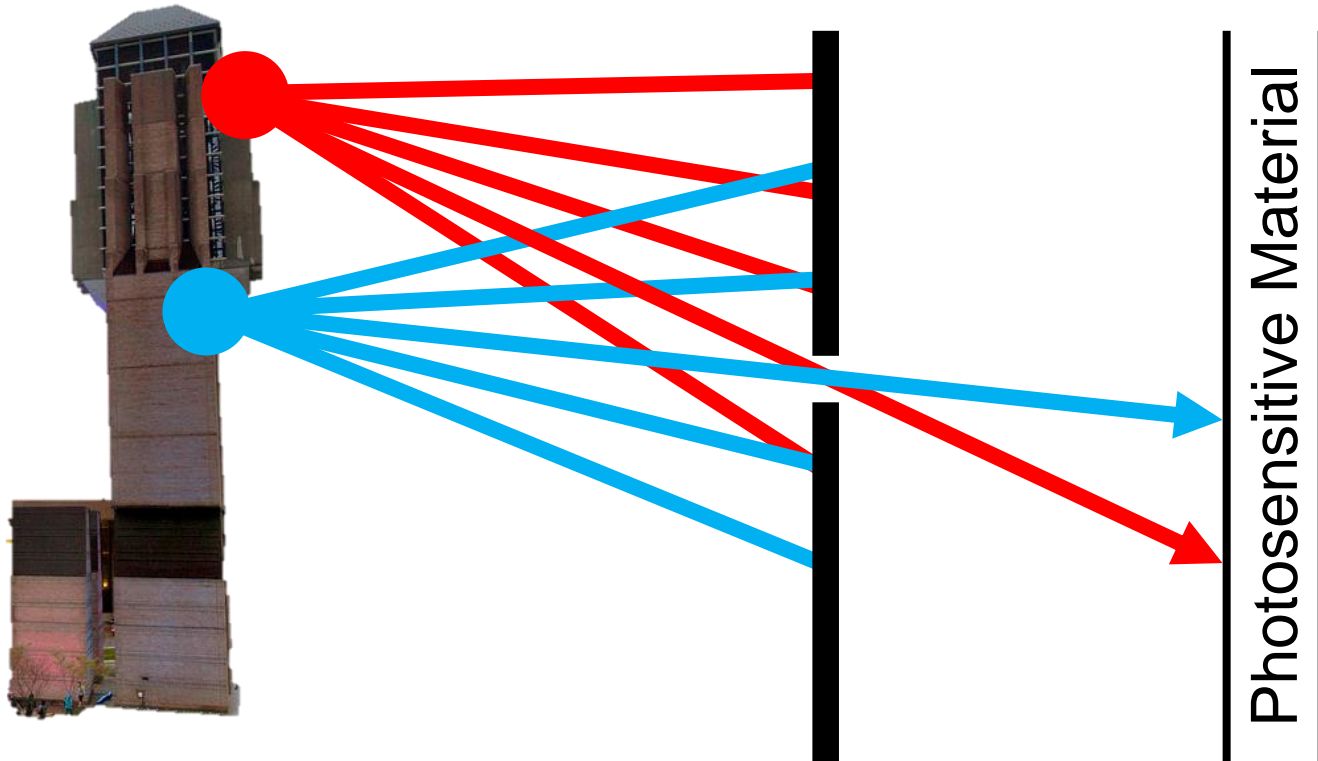


A. Criminisi, M. Kemp, and A. Zisserman, [Bringing Pictorial Space to Life: computer techniques for the analysis of paintings](#), *Proc. Computers and the History of Art*, 2002

# Next few classes

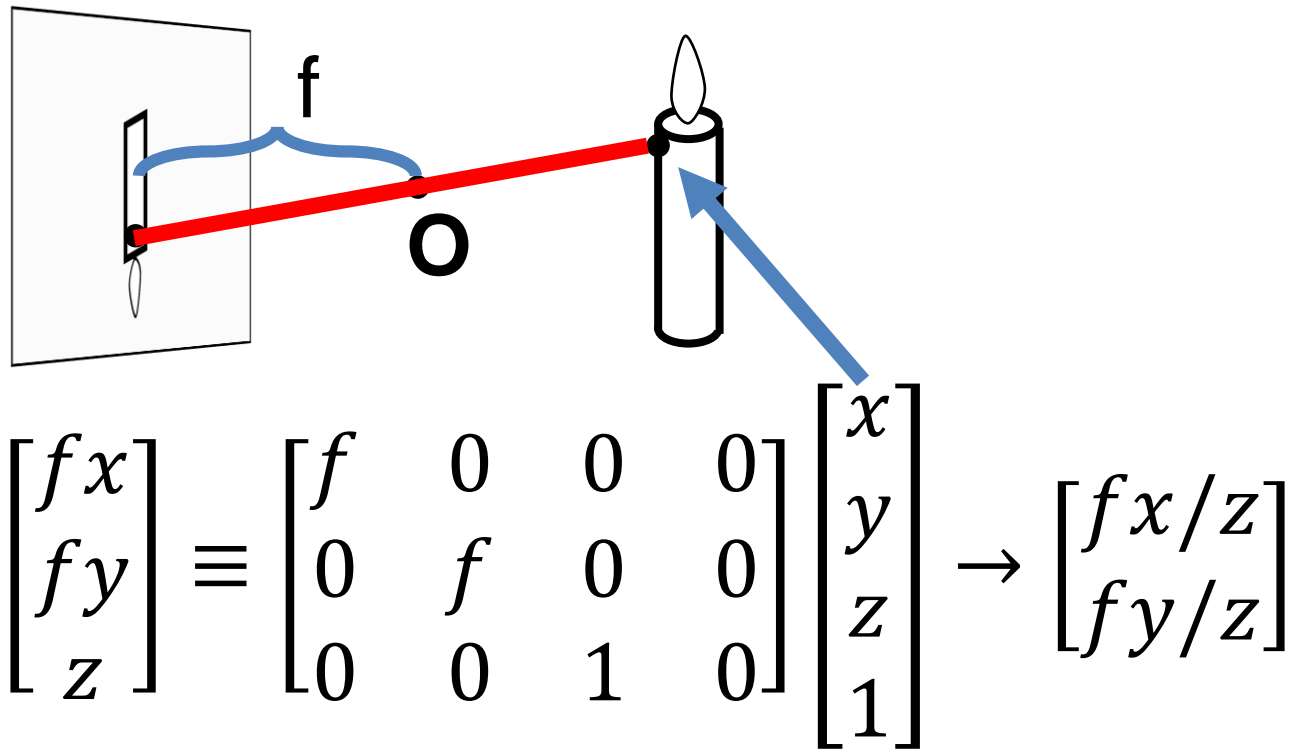
- First: some intuitions and examples from biological vision about 3D perception
- But first, a brief review

# Let's Take a Picture!

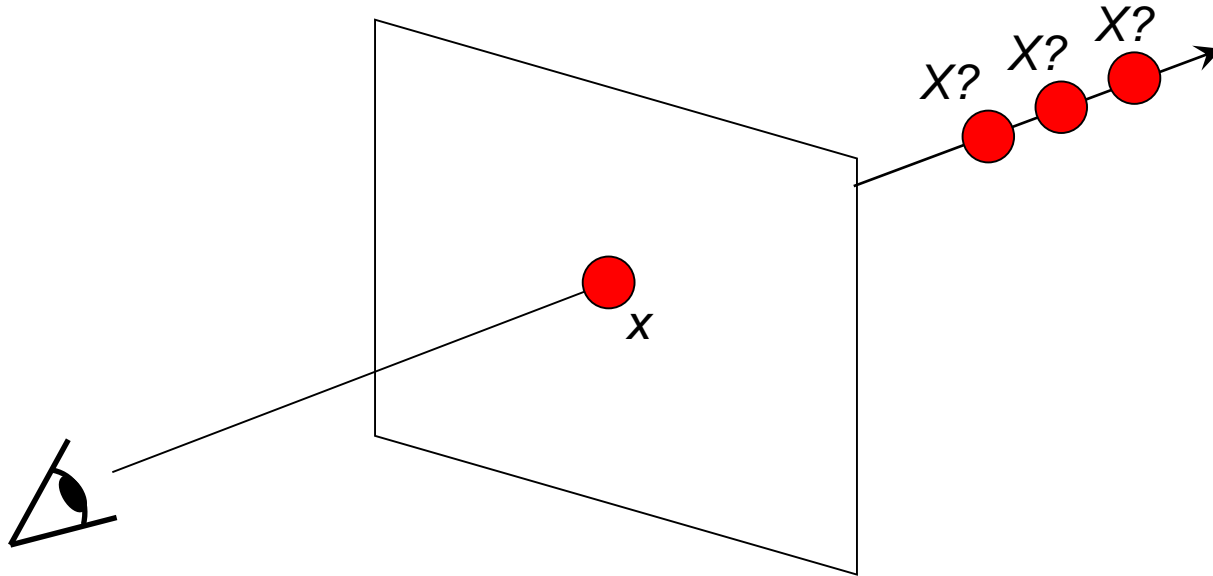


# Projection Matrix

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

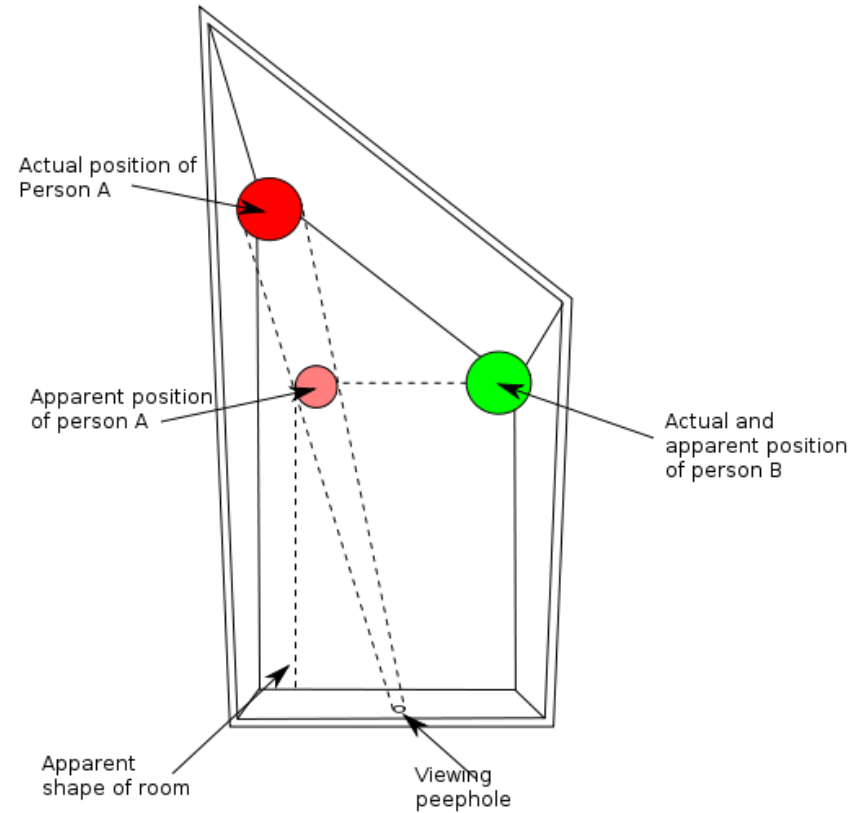


# Single-view Ambiguity



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

# Single-view Ambiguity



[http://en.wikipedia.org/wiki/Ames\\_room](http://en.wikipedia.org/wiki/Ames_room)



# Single-view Ambiguity



Diagram credit: J. Hays



# Single-view Ambiguity



[Rashad Alakbarov shadow sculptures](#)

# Resolving Single-view Ambiguity



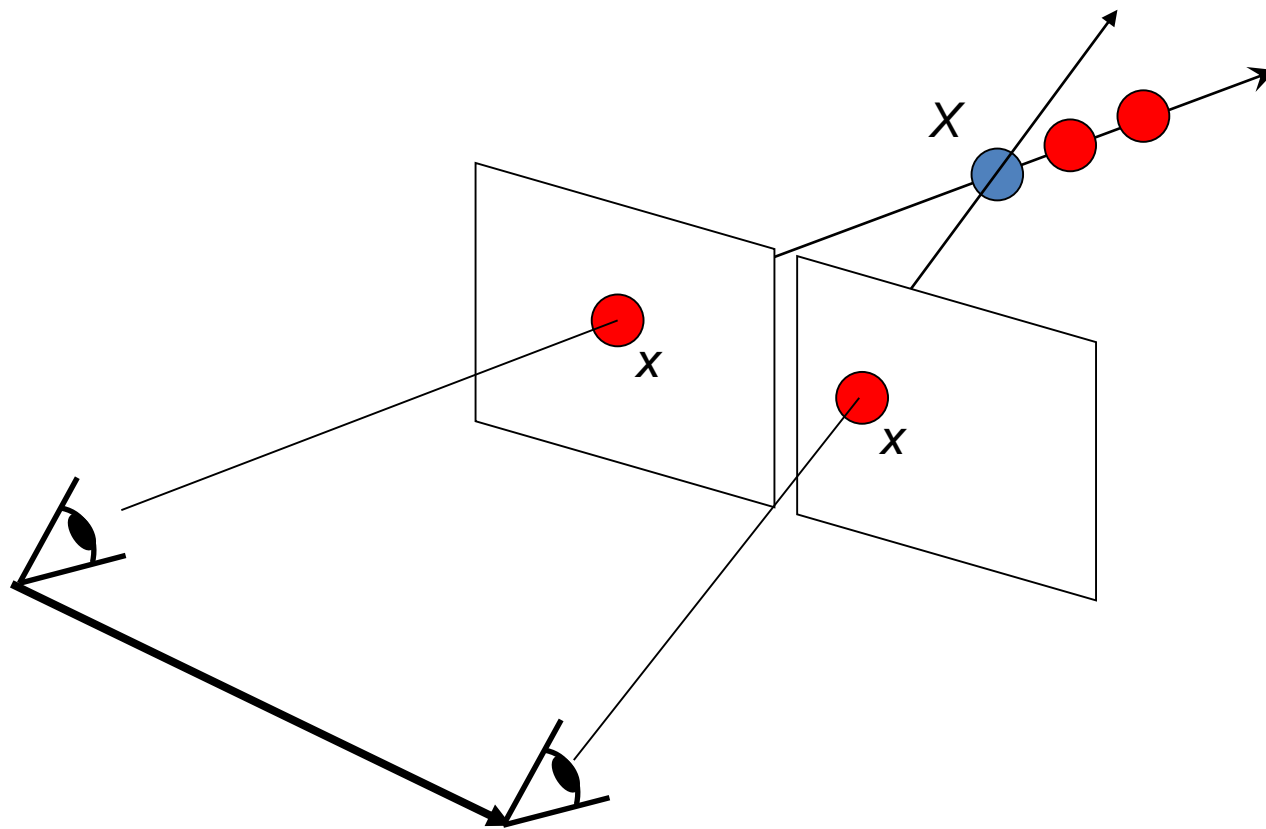
- Shoot light (lasers etc.) out of your eyes!
- Con: not so biologically plausible, dangerous?

# Resolving Single-view Ambiguity



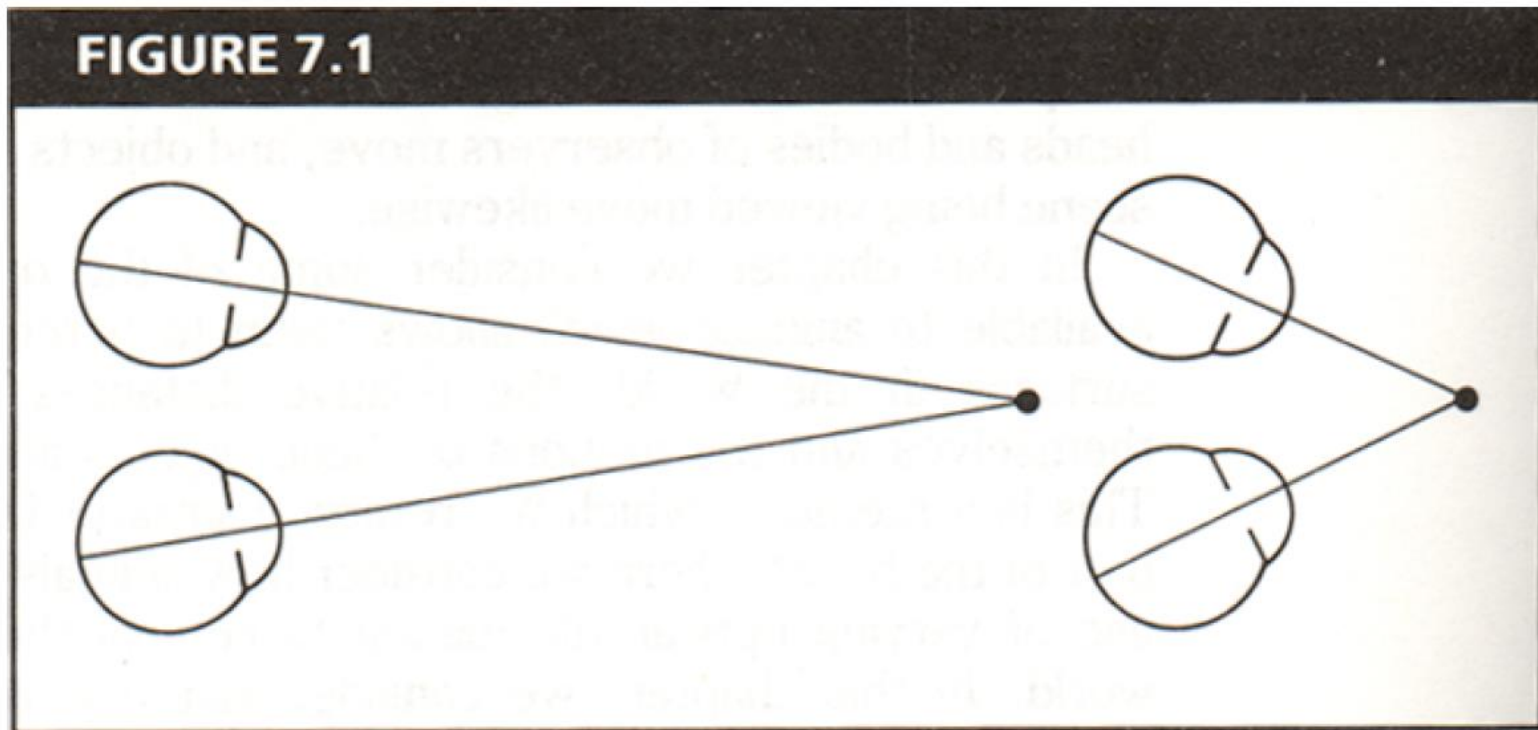
- Shoot light (lasers etc.) out of your eyes!
- Con: not so biologically plausible, dangerous?

# Resolving Single-view Ambiguity



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

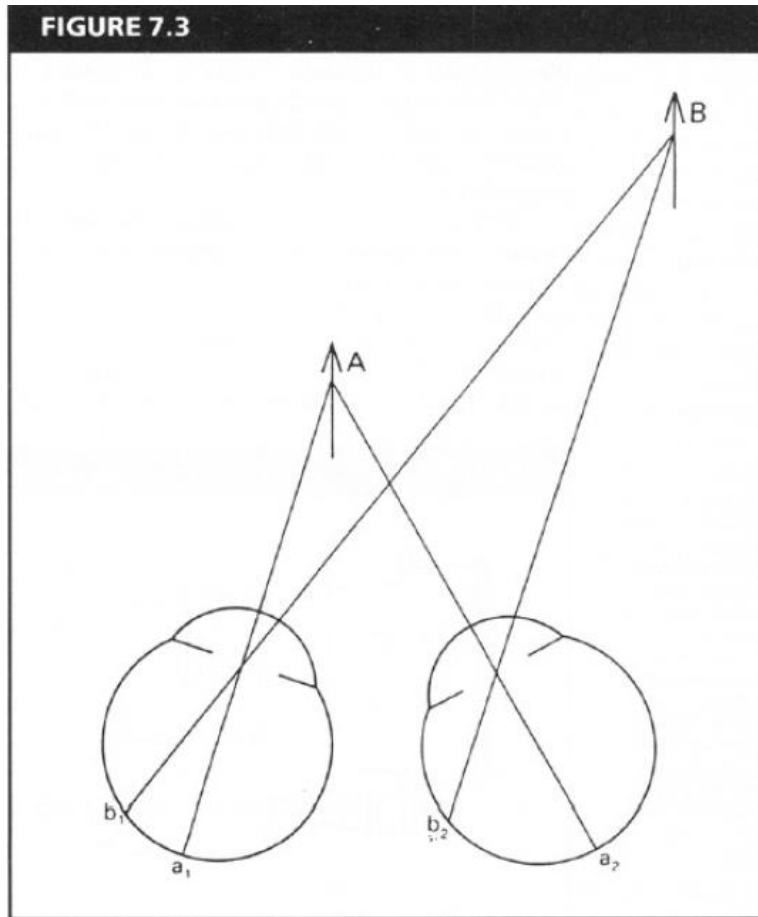
# Human stereopsis: disparity



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



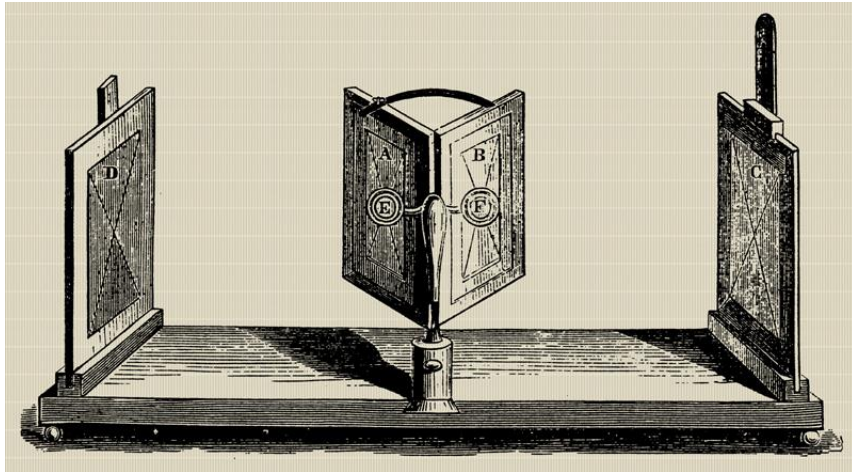
**Disparity** occurs when eyes fixate on one object; others appear at different visual angles

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



# 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



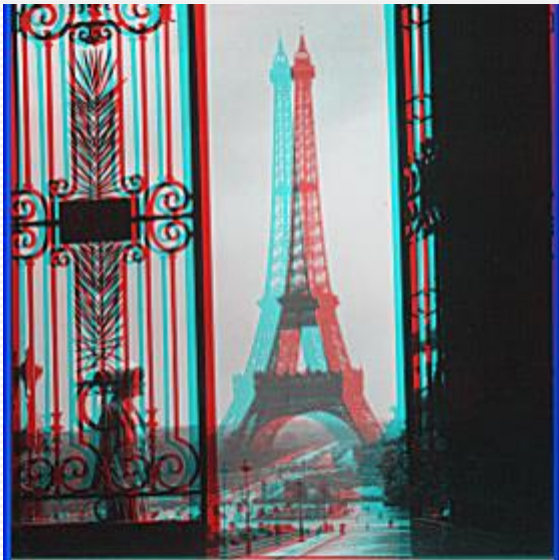
Slide credit: J. Hays



Image from fisher-price.com







© Copyright 2001 Johnson-Shaw Stereoscopic Museum

<http://www.johnsonshawmuseum.org>

Slide credit: J. Hays



**Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923**

Slide credit: J. Hays





[http://www.well.com/~jimmg/stereo/stereo\\_list.html](http://www.well.com/~jimmg/stereo/stereo_list.html)

Slide credit: J. Hays





[http://www.well.com/~jimmg/stereo/stereo\\_list.html](http://www.well.com/~jimmg/stereo/stereo_list.html)

Slide credit: J. Hays

# Autostereograms



Exploit disparity as depth cue using single image.

(Single image random dot stereogram, Single image stereogram)

# Autostereograms



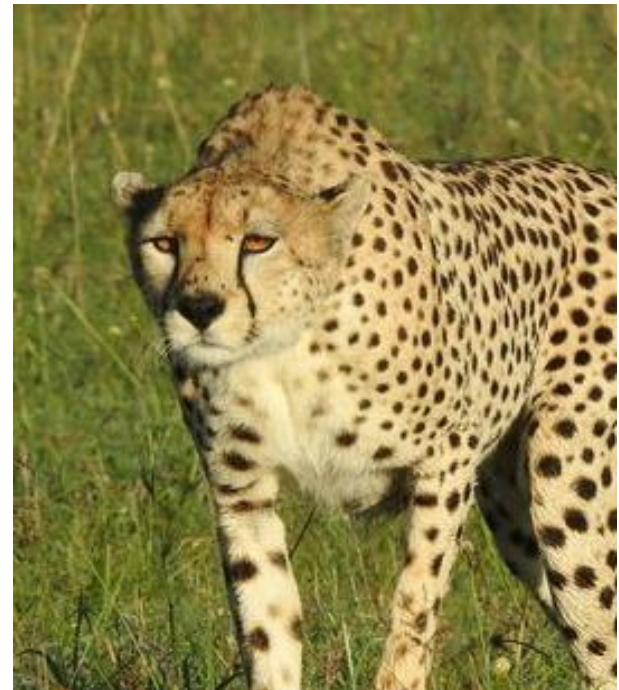


# Yeah, yeah, but...

Not all animals see stereo:

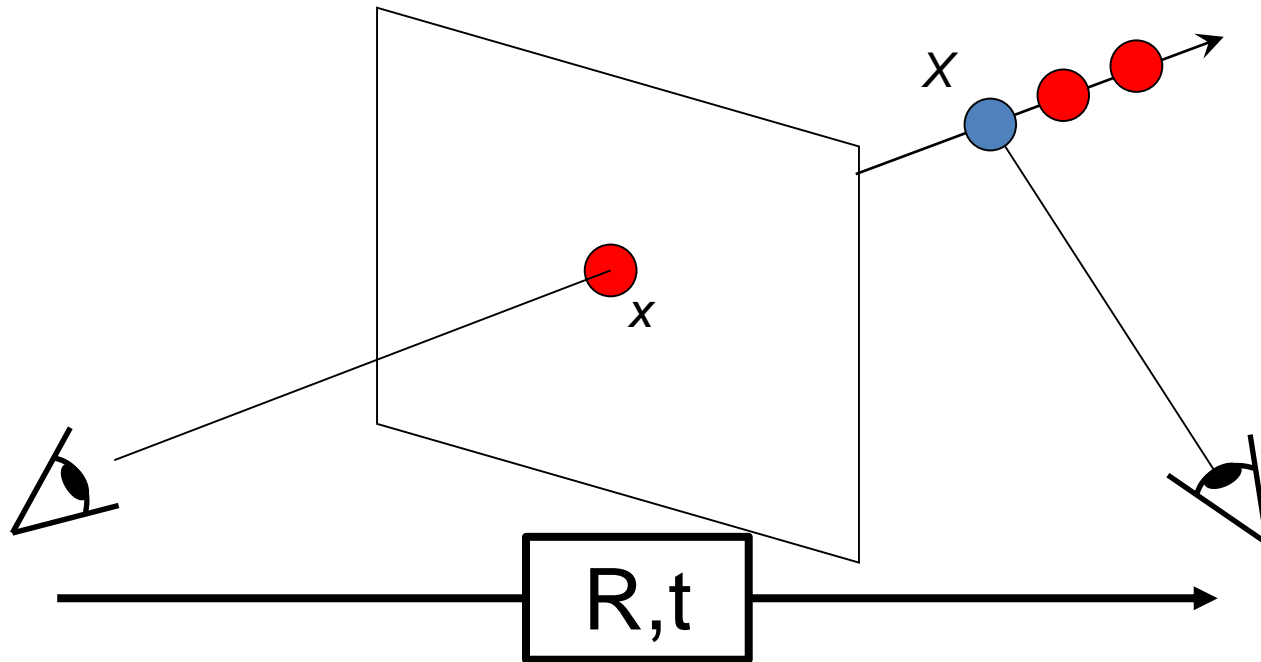
Prey animals (large field of view to spot predators)

Stereoblind people





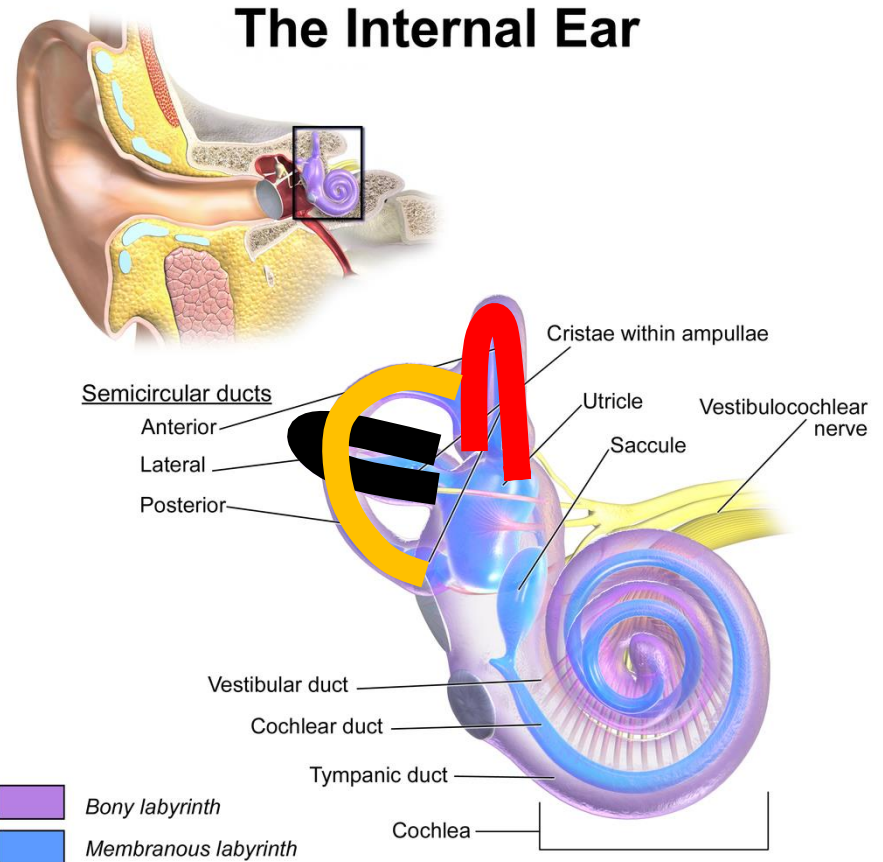
# Resolving Single-view Ambiguity



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

# 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?**
- Can solve via signals sent to muscles



# Yeah, yeah, but...

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



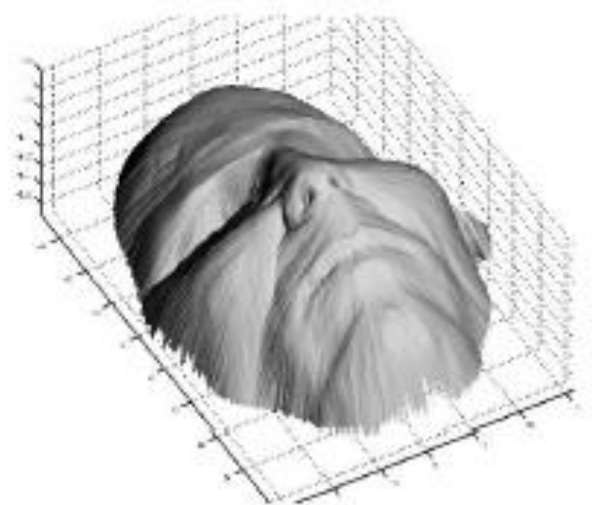
# Pictorial Cues – Shading



a)

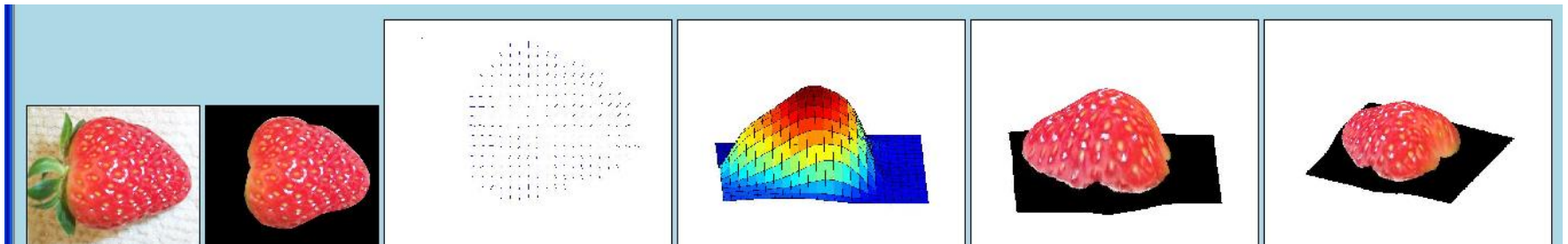
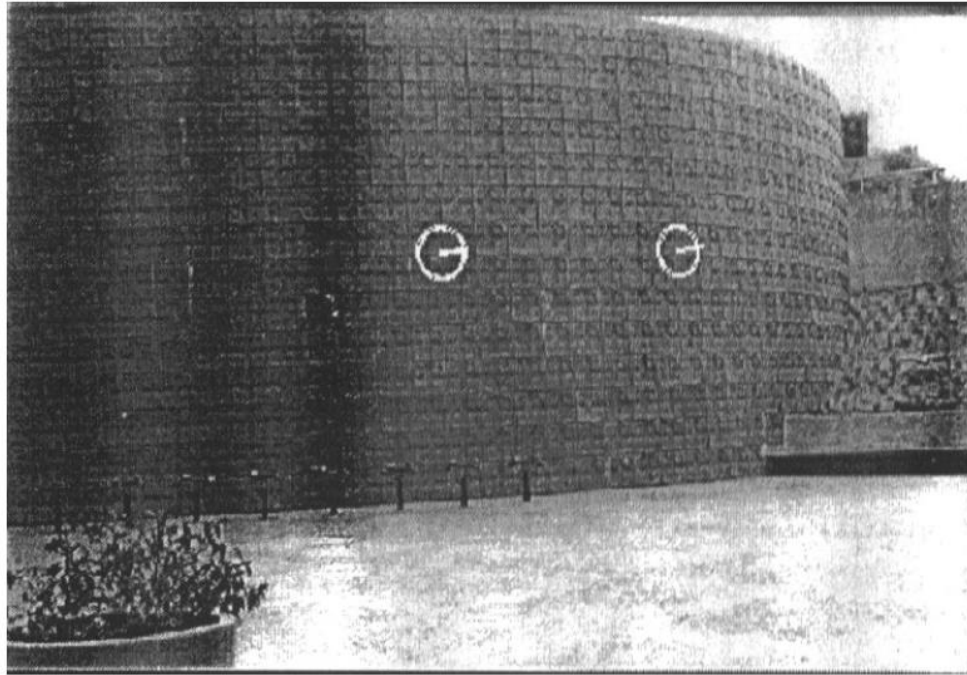


b)



c)

# Pictorial Cues – Texture



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



# Pictorial Cues – Perspective effects



Image credit: S. Seitz

# Pictorial Cues – Familiar Objects



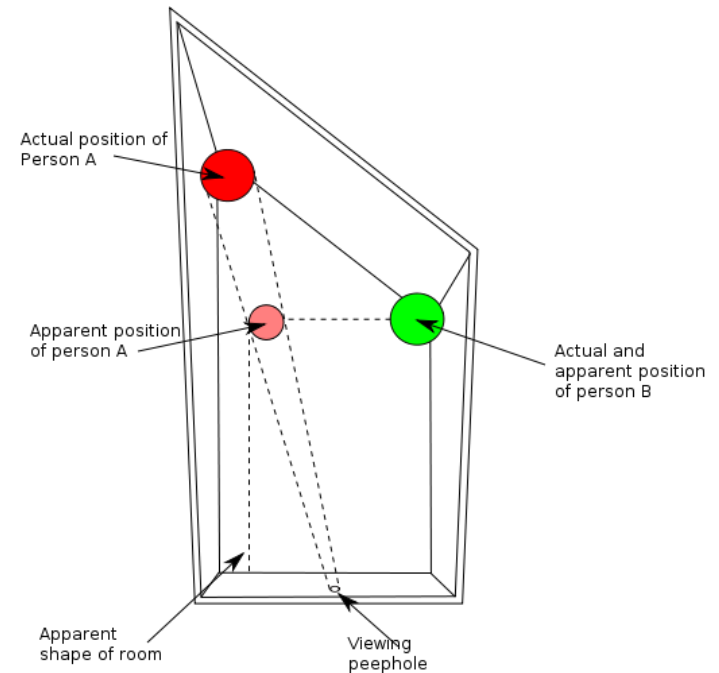


# Reality of 3D Perception

- 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 in part from stereo/motion cues

# How are Cues Combined?

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

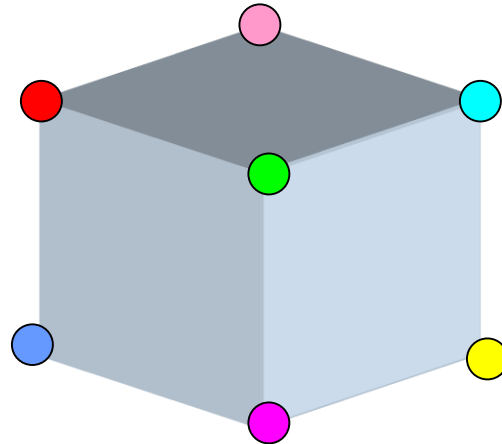
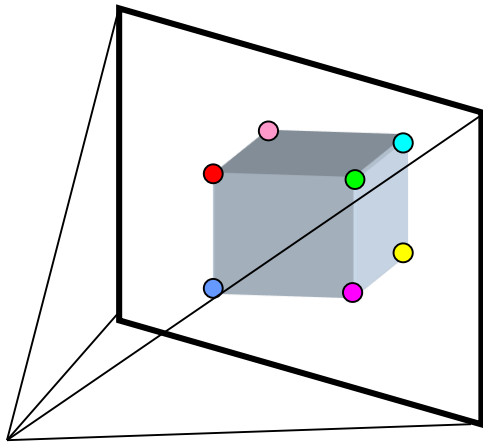


More Formally

# Multi-view geometry problems

*Calibration:*

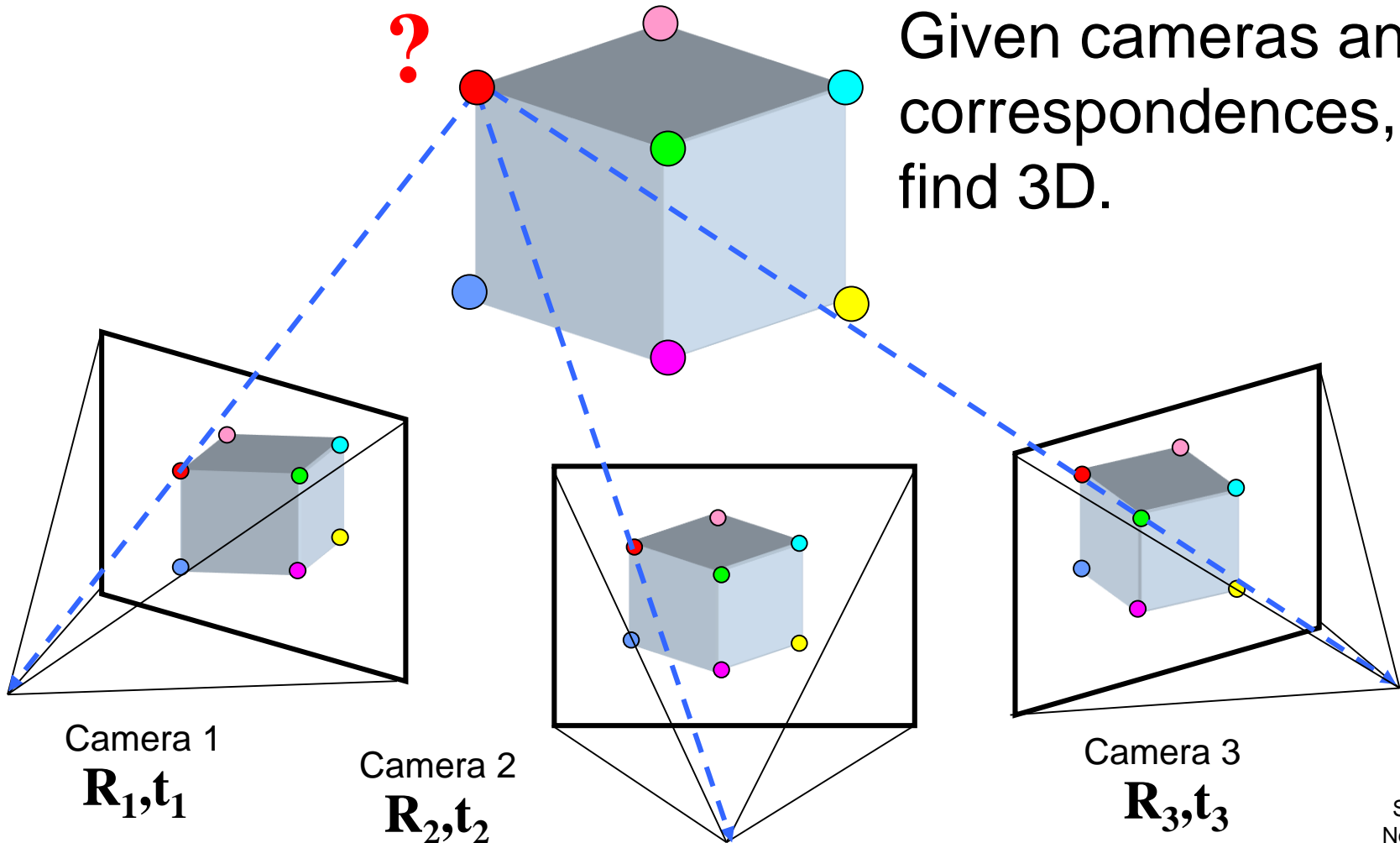
We need camera intrinsics /  $K$  in order to figure out where the rays are



Camera 1  
 $K$  ?

# Multi-view geometry problems

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

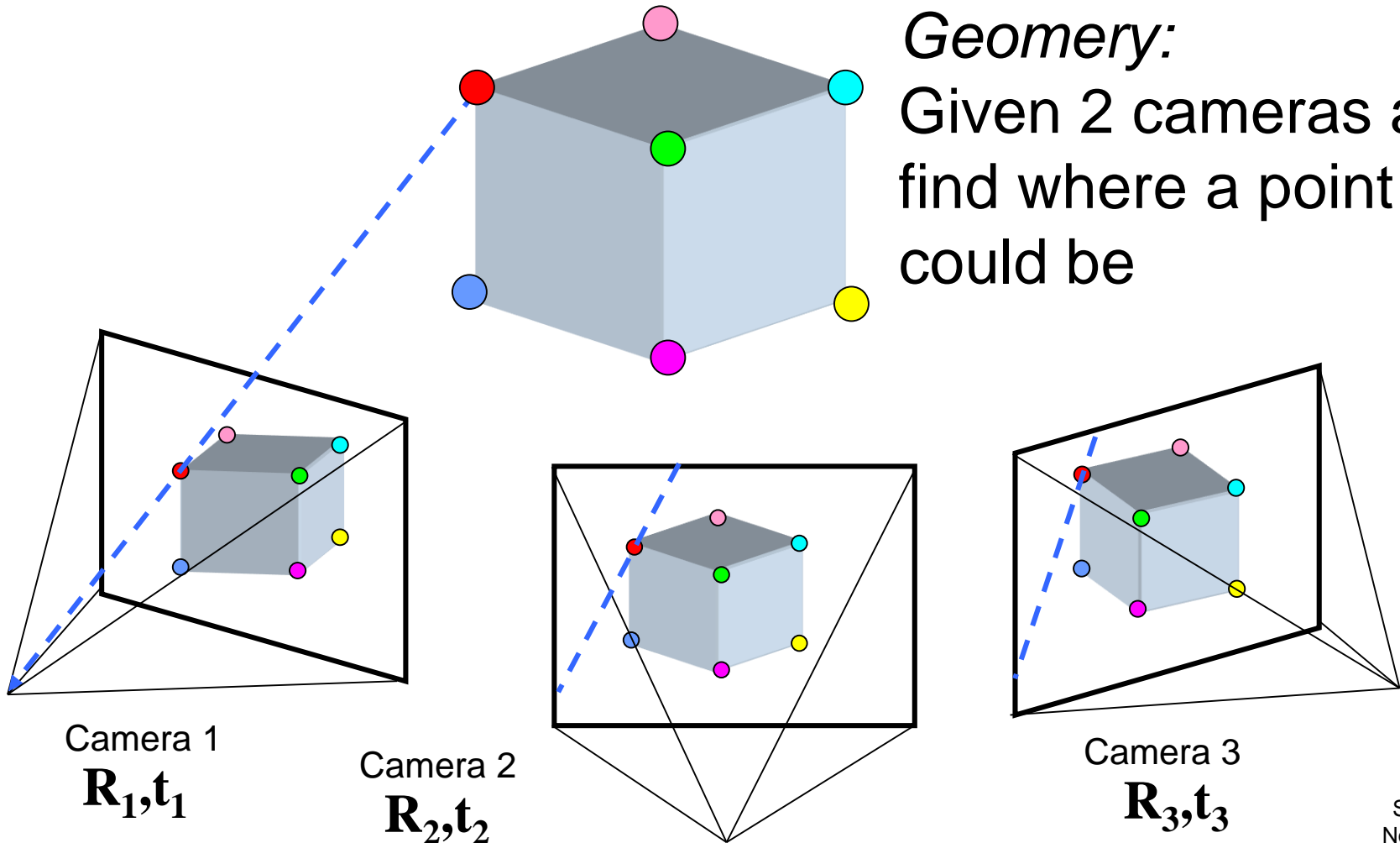




# Multi-view geometry problems

*Stereo/Epipolar  
Geometry:*

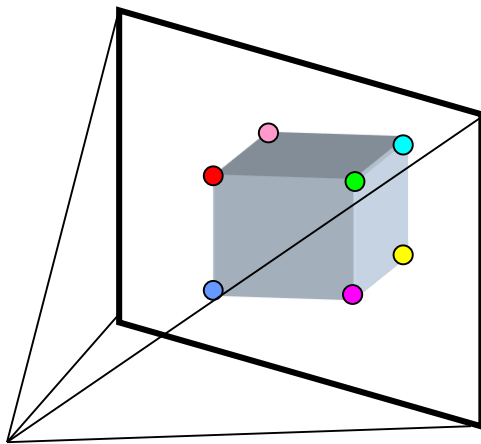
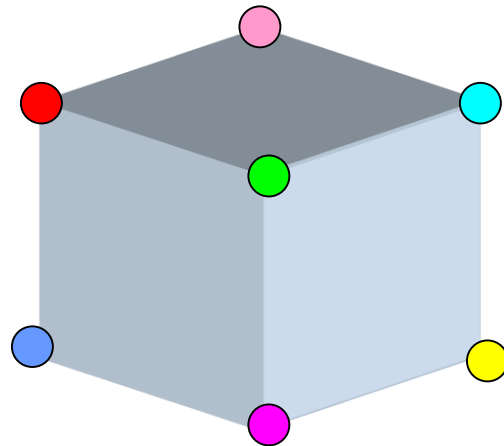
Given 2 cameras and  
find where a point  
could be



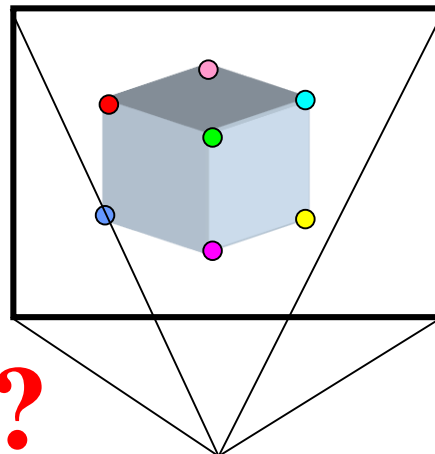
# Multi-view geometry problems

*Motion:*

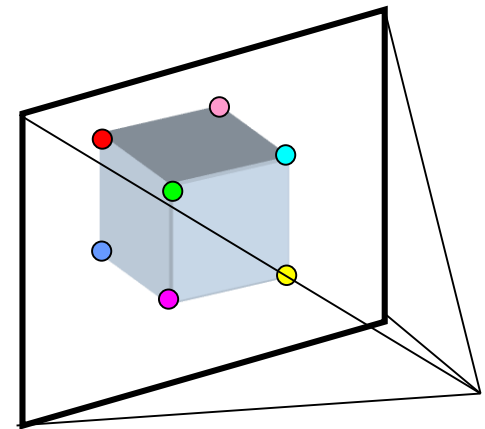
Figure out  $R, t$  for a set of cameras given correspondences



Camera 1  
 $R_1, t_1$  ?



Camera 2  
 $R_2, t_2$  ?



? Camera 3  
 $R_3, t_3$

# Outline

- (Today) Calibration:
  - Getting intrinsic matrix/K
- Single view geometry:
  - measurements with 1 image
- Stereo/Epipolar geometry:
  - 2 pictures → depthmap
- Structure from motion (SfM):
  - 2+ pictures → cameras, pointcloud

# Typical Perspective Model

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

focal length

$$\mathbf{p} \equiv \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

2D Projection of X

rotation translation

$$[\mathbf{R}_{3 \times 3} \quad \mathbf{t}_{3 \times 1}] \quad \mathbf{X}_{4 \times 1}$$

3D point

# Camera Calibration

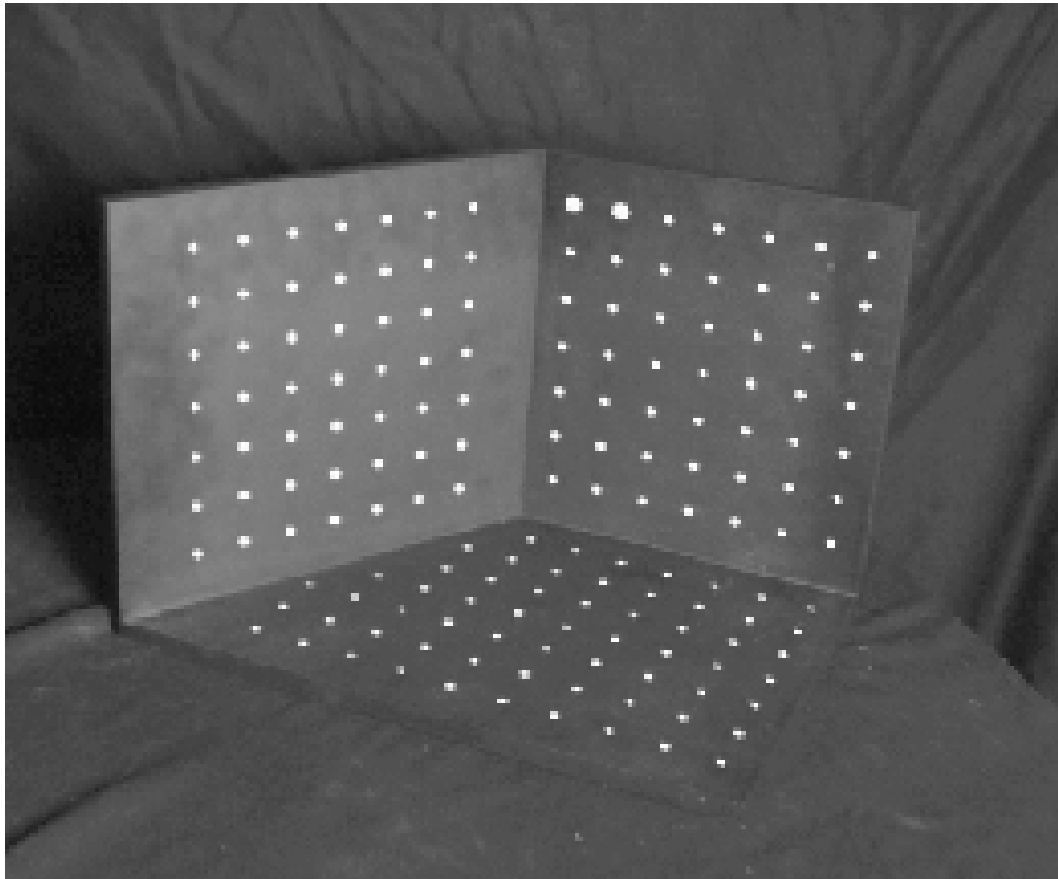
$$\mathbf{p} \equiv \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad [\mathbf{R}_{3 \times 3} \quad \mathbf{t}_{3 \times 1}] \quad \mathbf{X}_{4 \times 1}$$
$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \equiv \mathbf{M}_{3 \times 4} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

If I can get pairs of  $[X, Y, Z]$  and  $[u, v]$   
→ equations to constrain  $\mathbf{M}$



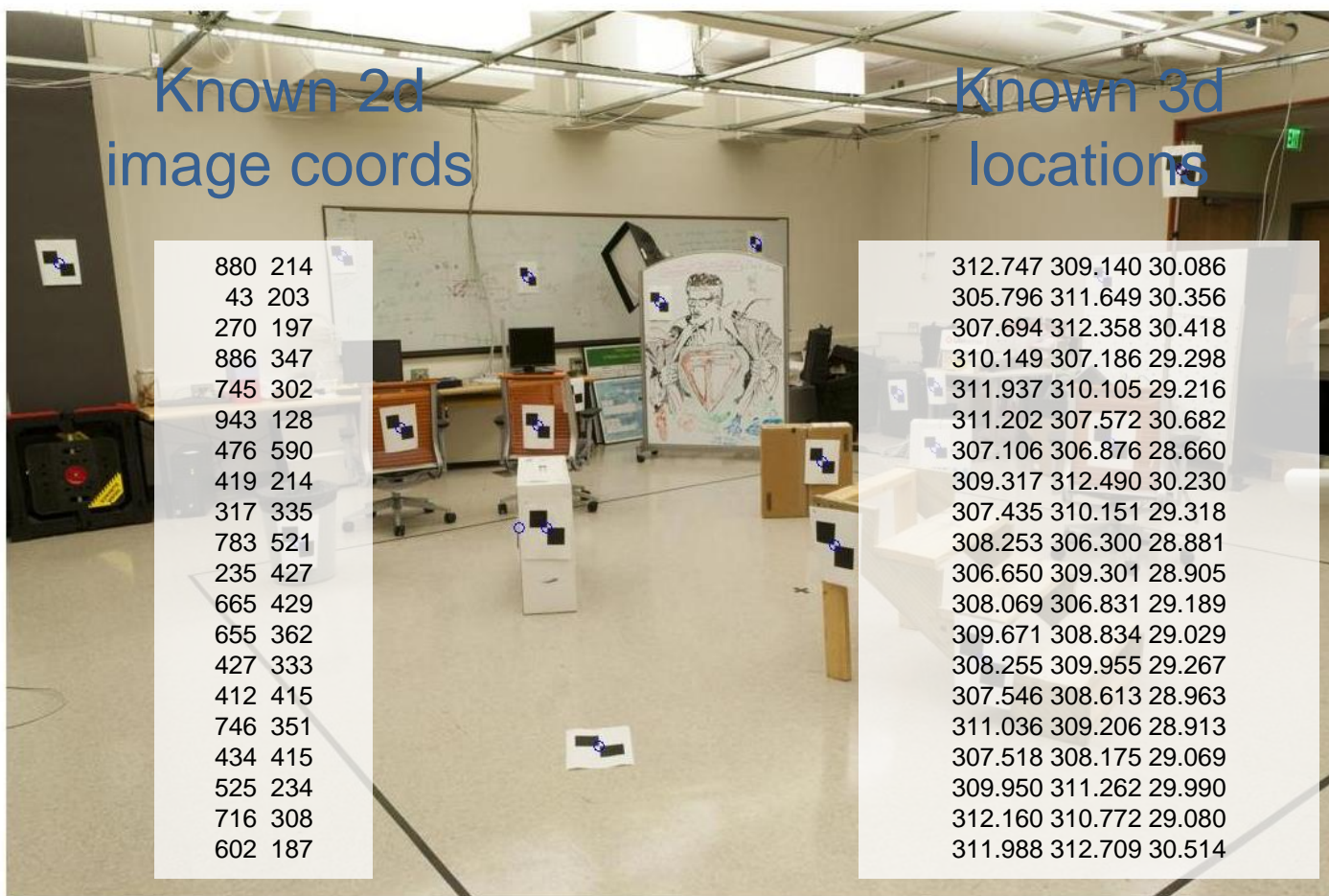
# Camera Calibration

A funny object with multiple planes.



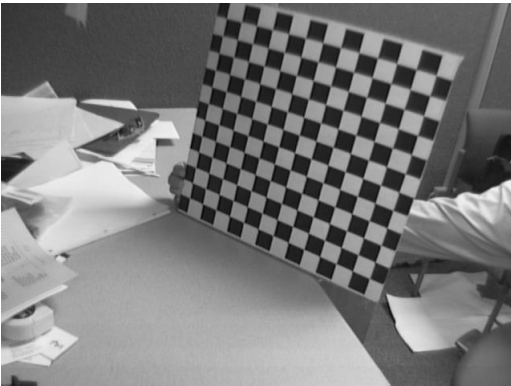
# Camera Calibration Targets

Someone used a tape measure

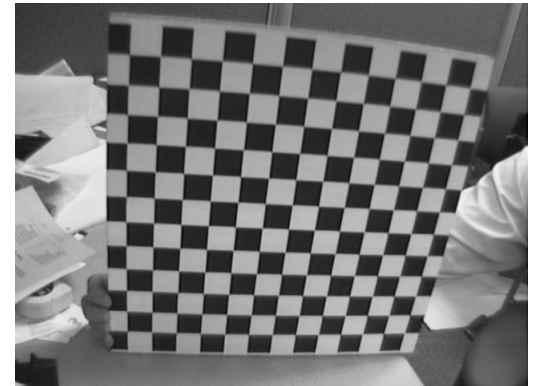


# Camera Calibration Targets

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

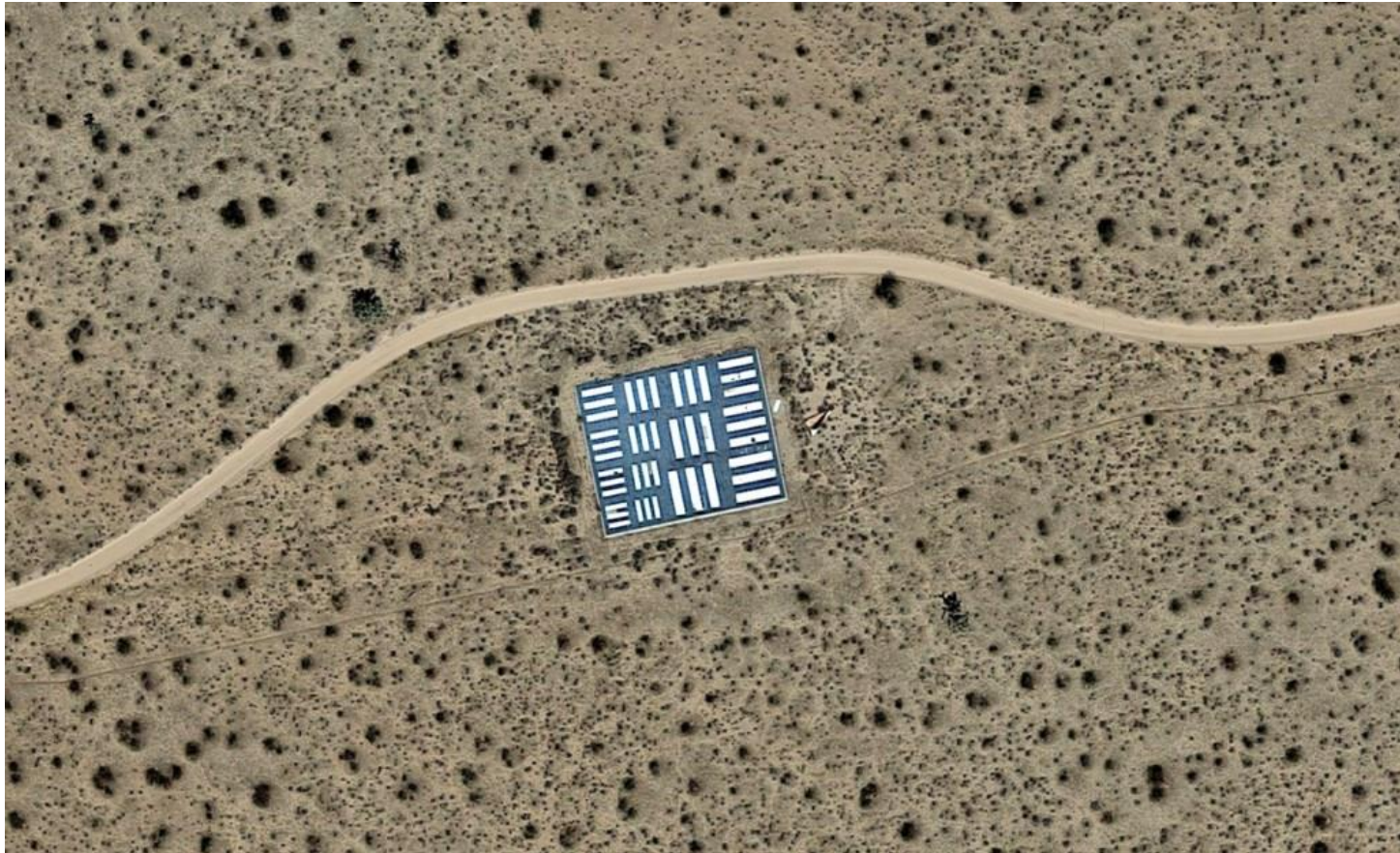


...



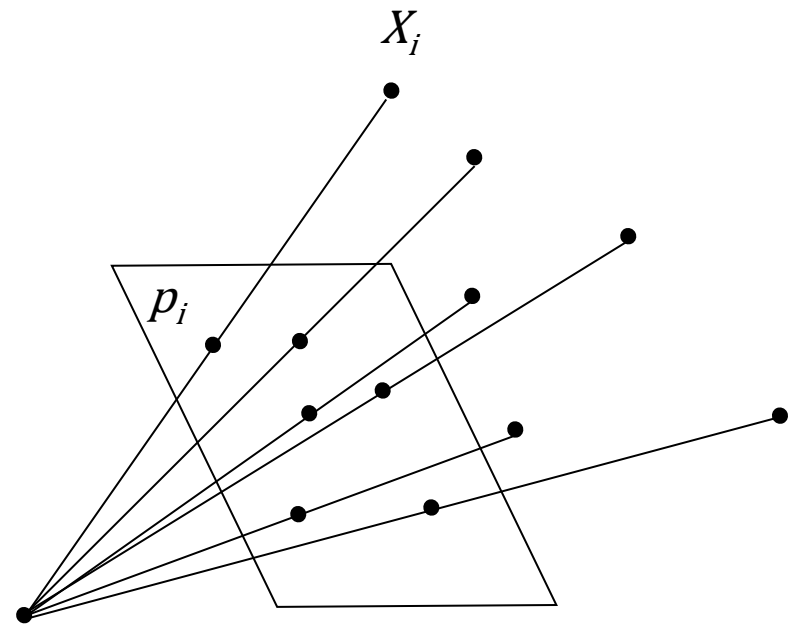
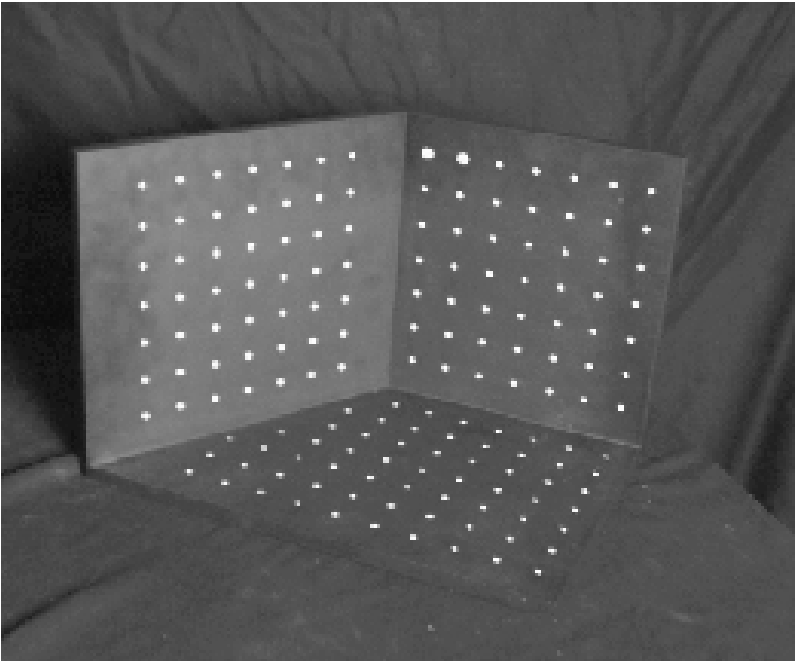
# Camera Calibration Targets

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





# Camera Calibration: Linear Method

$$\mathbf{p}_i \equiv \mathbf{M}\mathbf{X}_i$$

Remember (from geometry): this implies  $\mathbf{M}\mathbf{X}_i$   $\mathbf{p}_i$  are scaled copies of each other

$$\mathbf{p}_i = \lambda \mathbf{M}\mathbf{X}_i, \lambda \neq 0$$

Remember (from homography fitting): this implies their cross product is  $\mathbf{0}$

$$\mathbf{p}_i \times \mathbf{M}\mathbf{X}_i = \mathbf{0}$$

# Camera Calibration: Linear Method

$$\mathbf{p}_i \times \mathbf{M} \mathbf{X}_i = \mathbf{0}$$

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \times \begin{bmatrix} \mathbf{M}_1 \mathbf{X}_i \\ \mathbf{M}_2 \mathbf{X}_i \\ \mathbf{M}_3 \mathbf{X}_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Some tedious math occurs  
(see Homography derivation)

$$\begin{bmatrix} \mathbf{0}^T & -\mathbf{X}_i^T & y_i \mathbf{X}_i^T \\ \mathbf{X}_i^T & \mathbf{0}^T & -u_i \mathbf{X}_i^T \\ -v_i \mathbf{X}_i^T & u_i \mathbf{X}_i^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \mathbf{M}_1^T \\ \mathbf{M}_2^T \\ \mathbf{M}_3^T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

# Camera Calibration: Linear Method

$$\begin{bmatrix} \mathbf{0}^T & -X_i^T & v_i X_i^T \\ X_i^T & \mathbf{0}^T & -u_i X_i^T \\ -v_i X_i^T & u_i X_i^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} M_1^T \\ M_2^T \\ M_3^T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

**How many linearly independent equations?**

2

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

2

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

11

# Camera Calibration: Linear Method

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

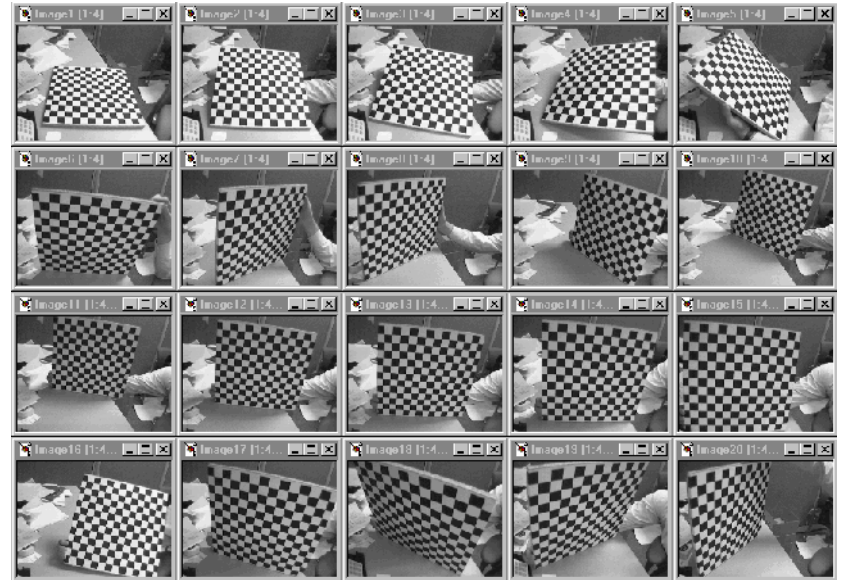
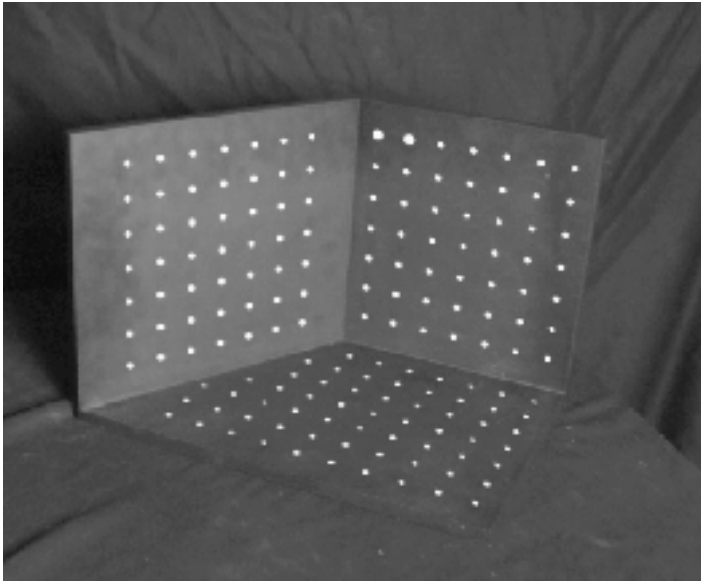
**How do we solve problems of the form**

$$\arg \min \|\mathbf{A}\mathbf{n}\|_2^2, \|\mathbf{n}\|_2^2 = 1 ?$$

Eigenvector of  $\mathbf{A}^T\mathbf{A}$  with smallest eigenvalue

# In Practice

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



# In Practice

I pulled a fast one.

We want:  $\mathbf{p} \equiv \mathbf{K}_{3 \times 3} [\mathbf{R}_{3 \times 3}, \mathbf{t}_{3 \times 1}] \mathbf{X}_{4 \times 1}$

We get:  $\mathbf{p} \equiv \mathbf{M}_{3 \times 4} \mathbf{X}_{4 \times 1}$

**What's the difference between  $\mathbf{K}[\mathbf{R}, \mathbf{t}]$  and  $\mathbf{M}$ ?**

Solution: QR-decomposition  $\rightarrow$  finite choices



# In Practice

If  $\mathbf{p}_i = \mathbf{M}\mathbf{x}_i$  is overconstrained, the objective function isn't actually the one you care about.

Instead:

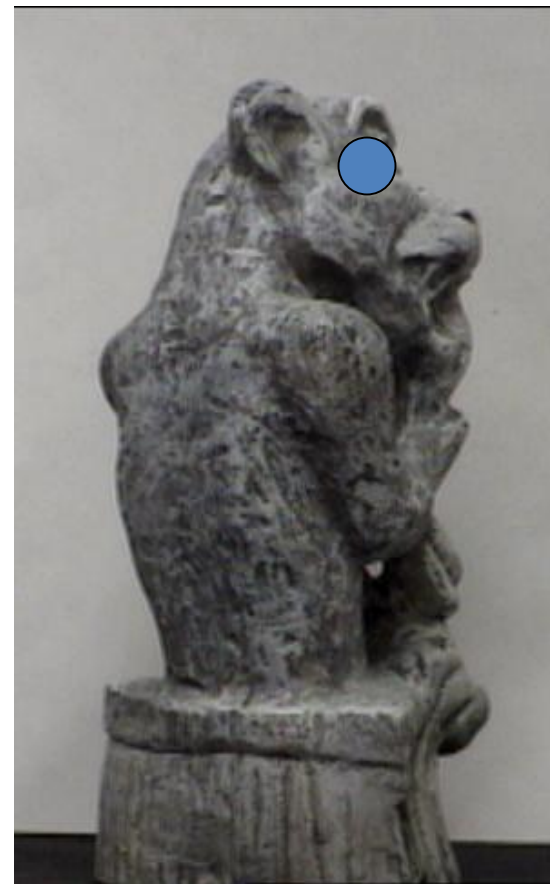
- 1) initialize parameters with linear model
- 2) Apply off-the-shelf non-linear optimizer to:

$$\sum \|\text{proj}(\mathbf{M}\mathbf{X}_i) - [u_i, v_i]^T\|_2^2$$

Advantage: can also add radial distortion, not optimize over known variables, add constraints

# What Does This Get You?

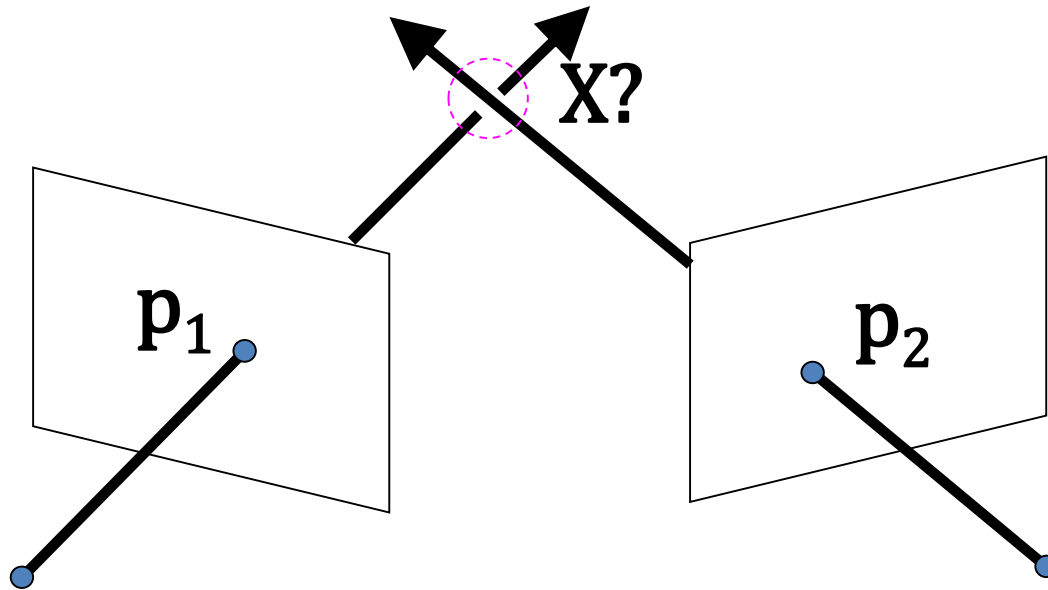
Given projection  $\mathbf{p}_i$  of unknown 3D point  $\mathbf{X}$  in two or more images (with known cameras  $\mathbf{M}_i$ ), find  $\mathbf{X}$



# Triangulation

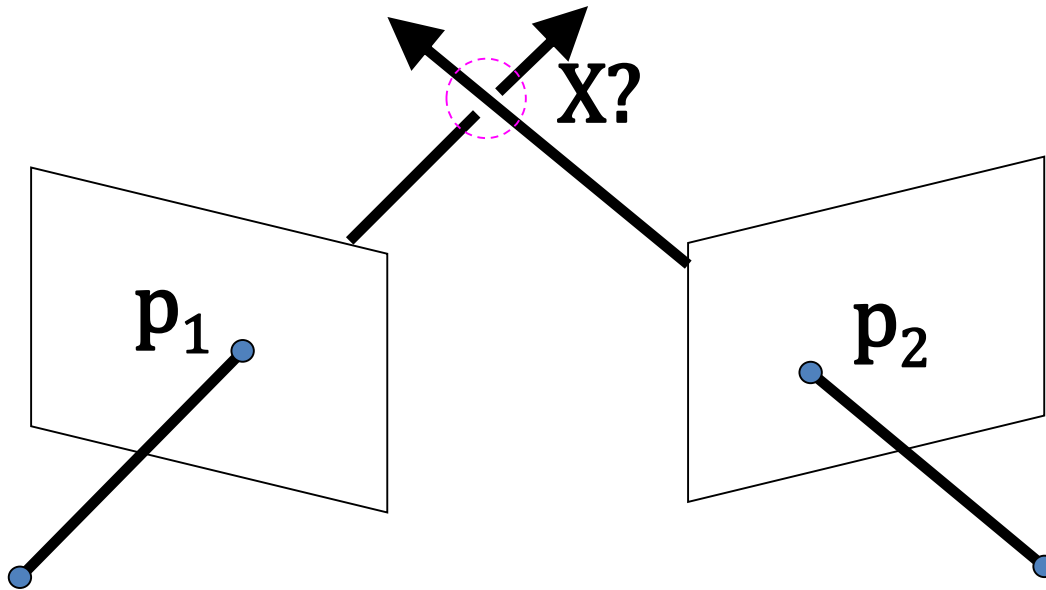
Given projection  $\mathbf{p}_i$  of unknown 3D point  $\mathbf{X}$  in two or more images (with known cameras  $\mathbf{M}_i$ ), find  $\mathbf{X}$

**Why is the calibration here important?**



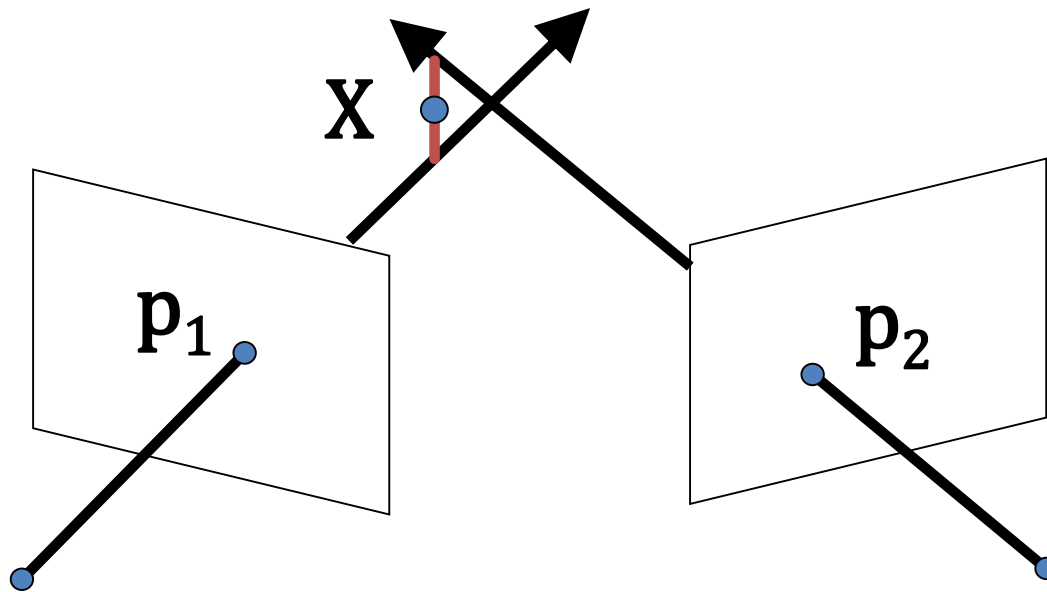
# Triangulation

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



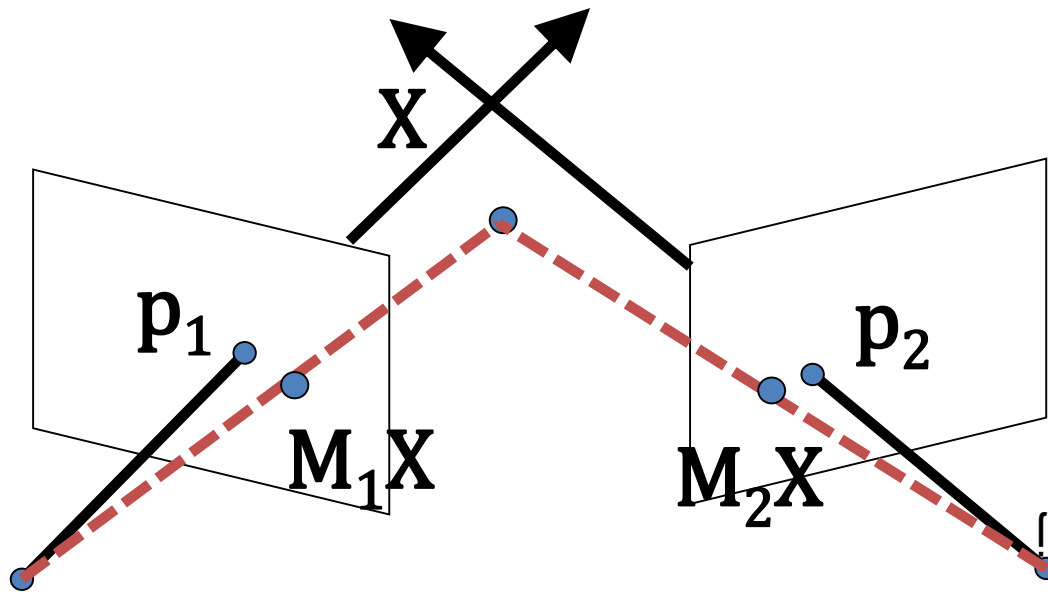
# Triangulation – Geometry

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



# Triangulation – Non-linear Optim.

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





# Triangulation – Linear Optimization

$$\begin{array}{l} \mathbf{p}_1 \equiv \mathbf{M}_1 \mathbf{X} \\ \mathbf{p}_2 \equiv \mathbf{M}_2 \mathbf{X} \end{array} \quad \rightarrow \quad \begin{array}{l} \mathbf{p}_1 \times \mathbf{M}_1 \mathbf{X} = \mathbf{0} \\ \mathbf{p}_2 \times \mathbf{M}_1 \mathbf{X} = \mathbf{0} \end{array} \quad \rightarrow \quad \begin{array}{l} [\mathbf{p}_{1x}] \mathbf{M}_1 \mathbf{X} = \mathbf{0} \\ [\mathbf{p}_{2x}] \mathbf{M}_2 \mathbf{X} = \mathbf{0} \end{array}$$

Cross Prod.  
as matrix

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

$$\begin{array}{l} [\mathbf{p}_{1x}] \mathbf{M}_1 \mathbf{X} = \mathbf{0} \\ [\mathbf{p}_{2x}] \mathbf{M}_2 \mathbf{X} = \mathbf{0} \end{array} \quad \rightarrow \quad \begin{array}{l} ([\mathbf{p}_{1x}] \mathbf{M}_1) \mathbf{X} = \mathbf{0} \\ ([\mathbf{p}_{2x}] \mathbf{M}_2) \mathbf{X} = \mathbf{0} \end{array} \quad \rightarrow \quad \begin{array}{l} \text{Two eqns per} \\ \text{camera for 3} \\ \text{unkn. in } \mathbf{X} \end{array}$$