

Intro to 3D + Camera Calibration

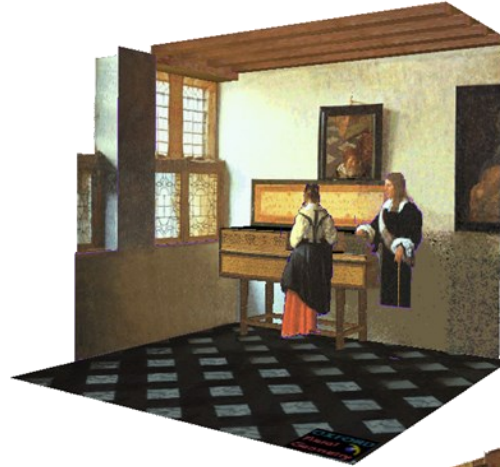
EECS 442 – David Fouhey and Justin Johnson
Winter 2021, University of Michigan

<https://web.eecs.umich.edu/~justincj/teaching/eecs442/WI2021/>

Our goal: Recovery of 3D structure



J. Vermeer, *Music Lesson*, 1662

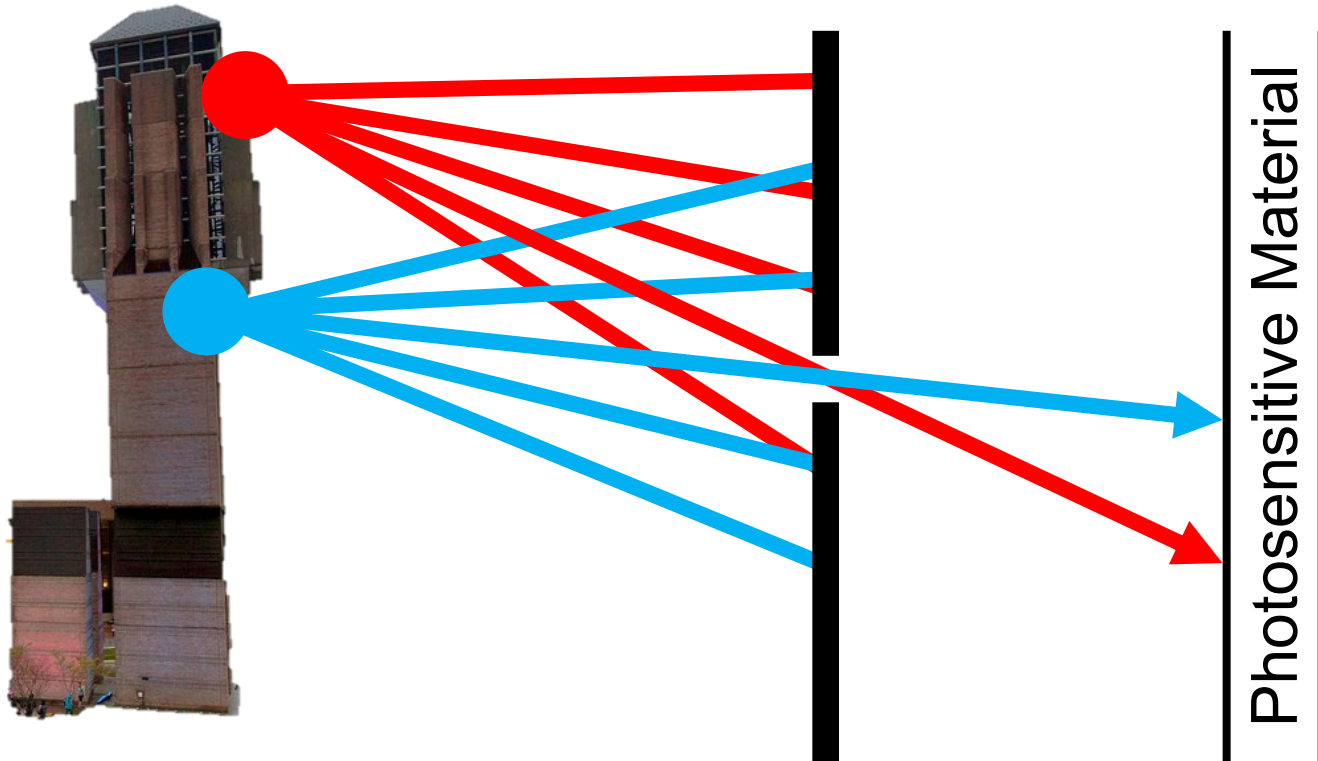


A. Criminisi, M. Kemp, and A. Zisserman, [Bringing Pictorial Space to Life: computer techniques for the analysis of paintings](#), *Proc. Computers and the History of Art*, 2002

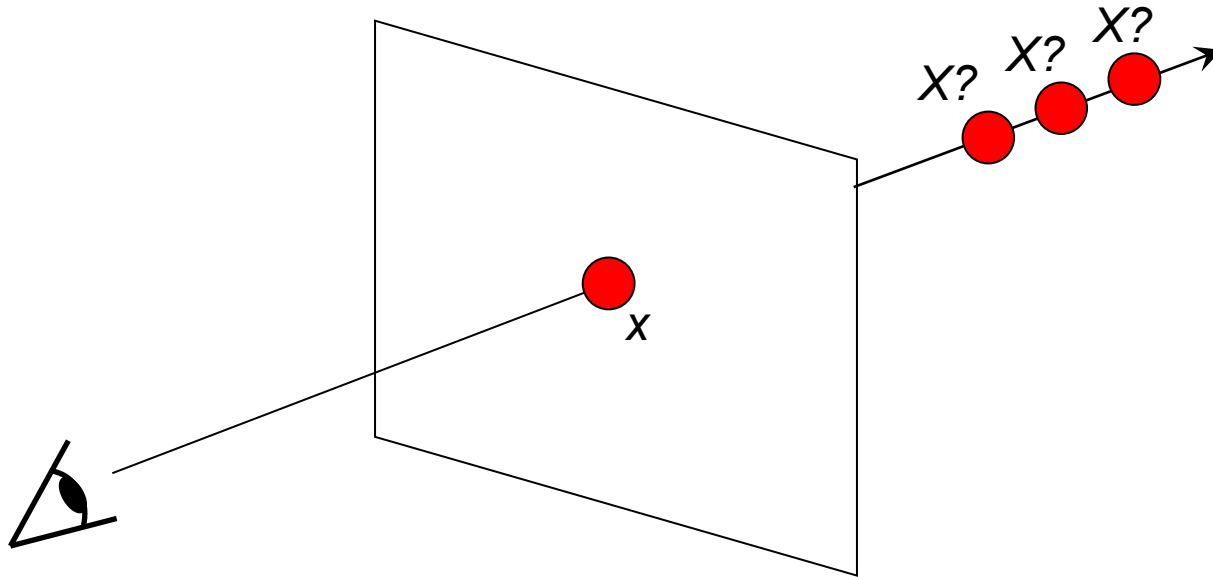
Next few classes

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

Let's Take a Picture!

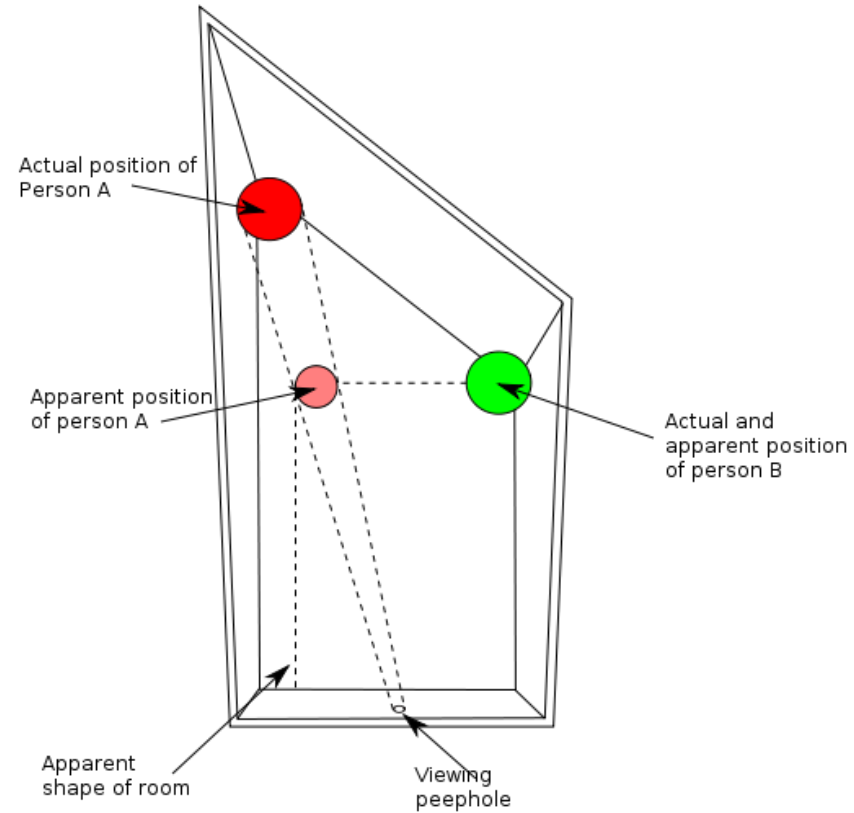


Single-view Ambiguity



- 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

Single-view Ambiguity



Diagram credit: J. Hays

Resolving Single-view Ambiguity



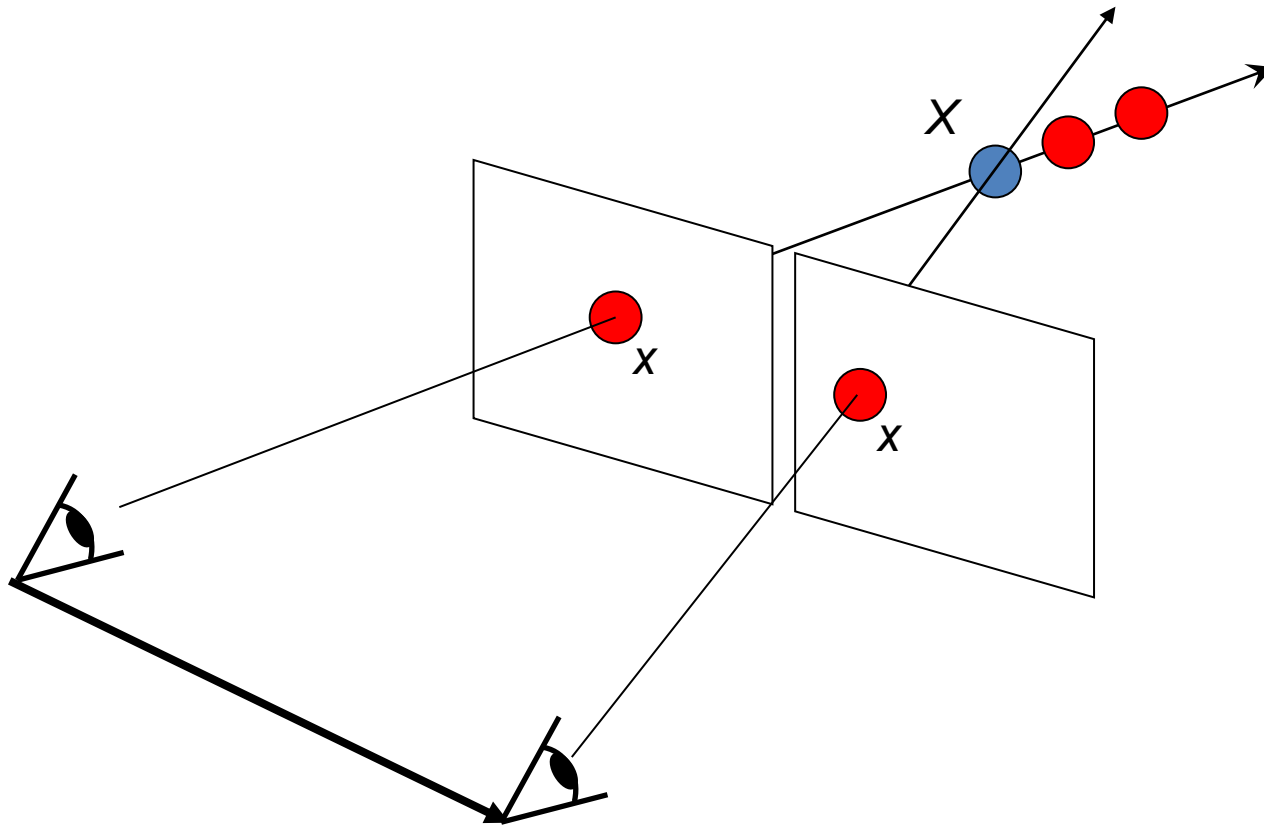
- Shoot light (lasers etc.) out of your eyes!
- Con: not so biologically plausible, dangerous?

Resolving Single-view Ambiguity



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

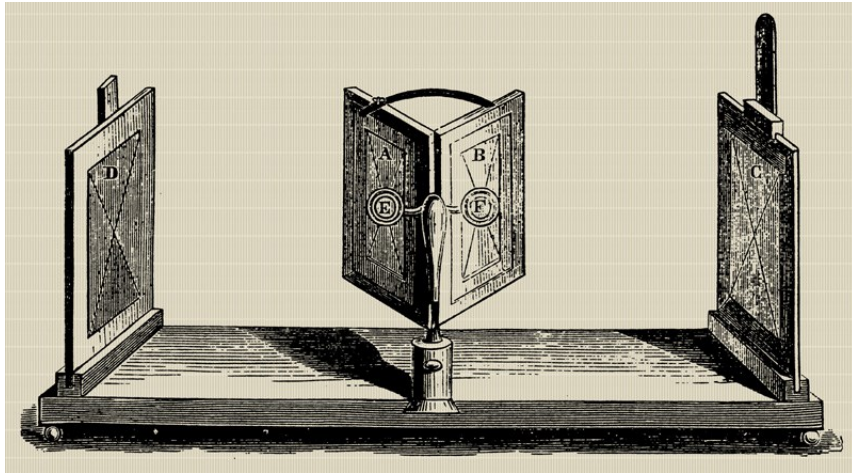
Resolving Single-view Ambiguity



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

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

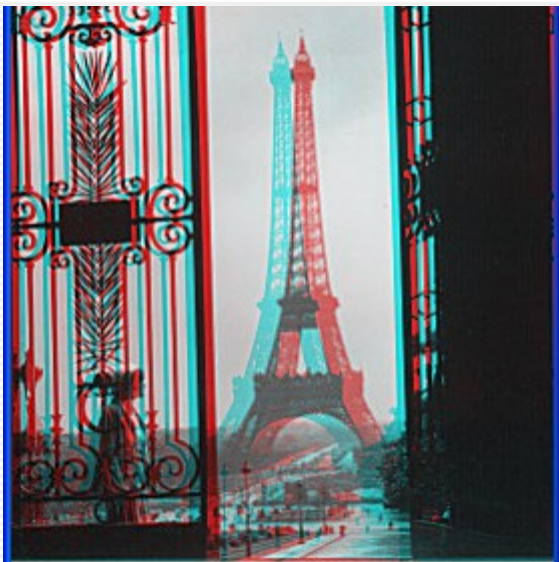


Slide credit: J. Hays



Image from fisher-price.com





© Copyright 2001 Johnson-Shaw Stereoscopic Museum

<http://www.johnsonshawmuseum.org>

Slide credit: J. Hays



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

Slide credit: J. Hays



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

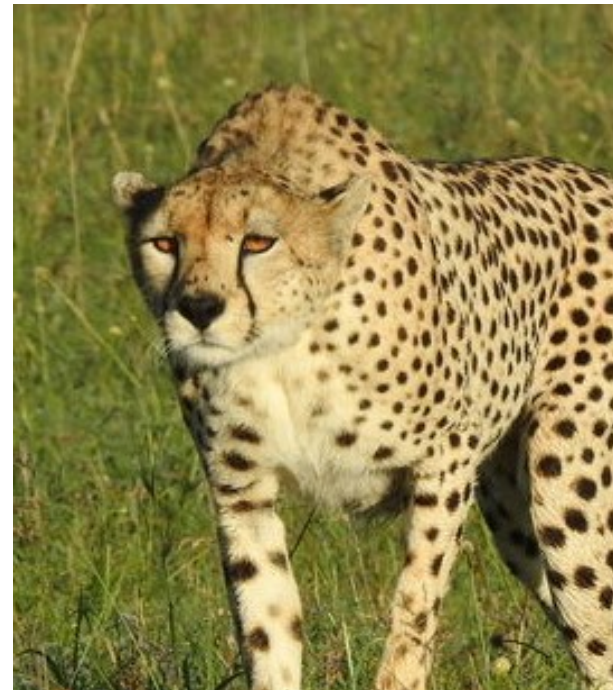


Yeah, yeah, but...

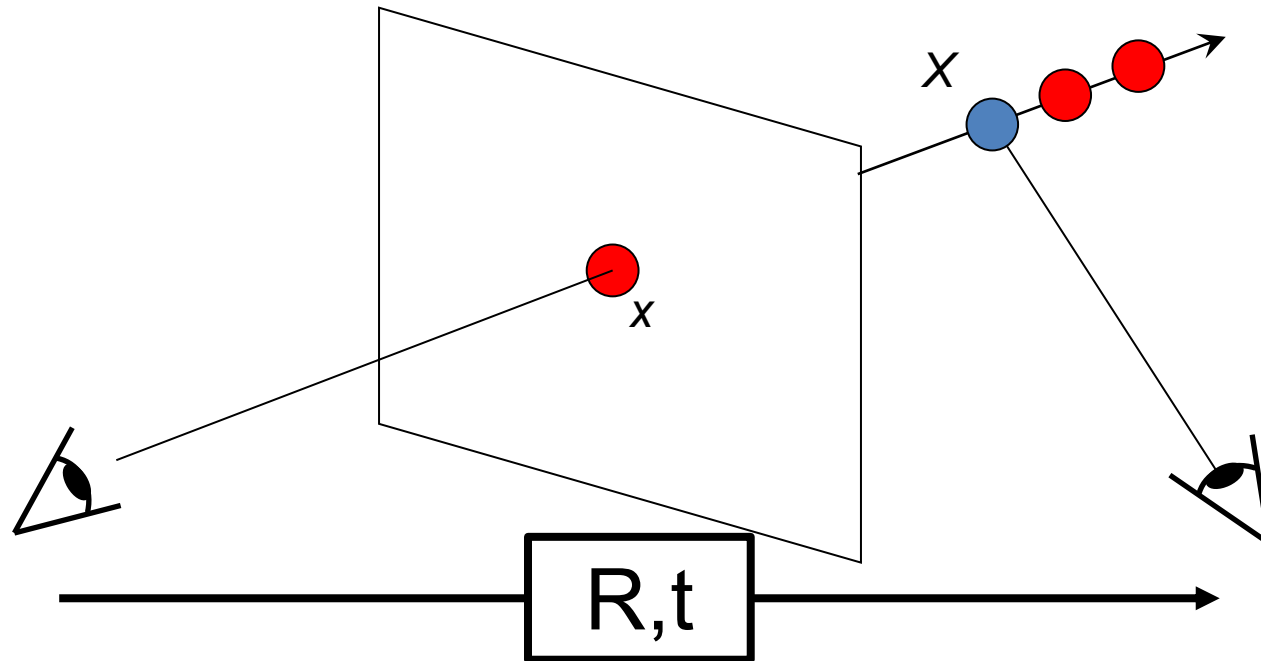
Not all animals see stereo:

Prey animals (large field of view to spot predators)

Stereoblind people



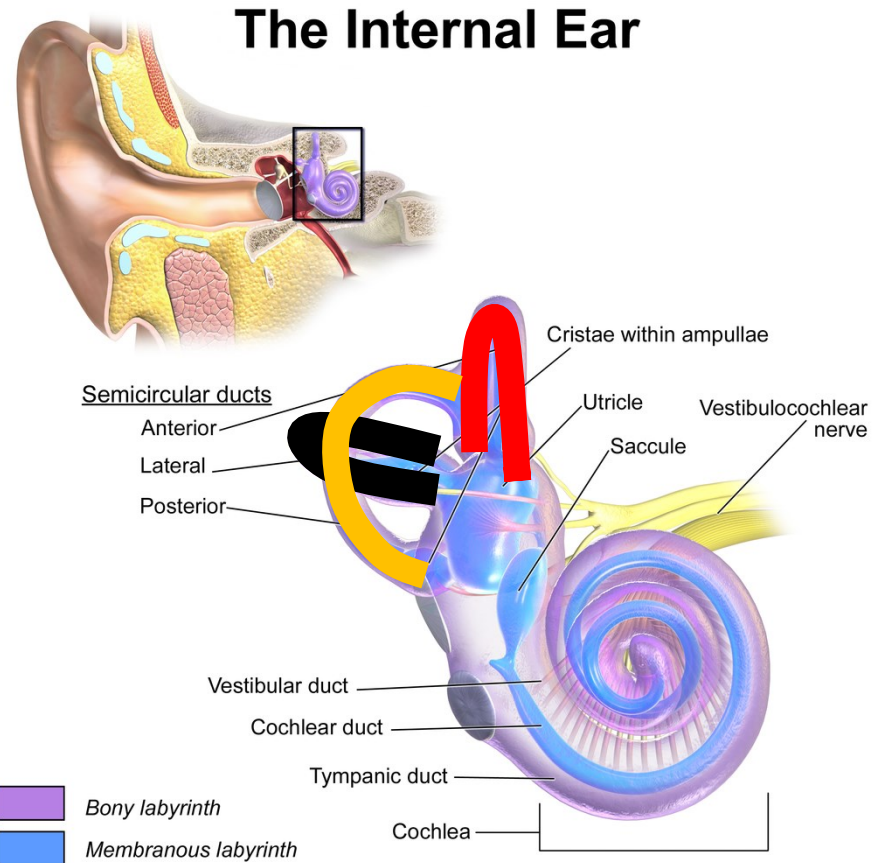
Resolving Single-view Ambiguity



- 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 (efference copy)

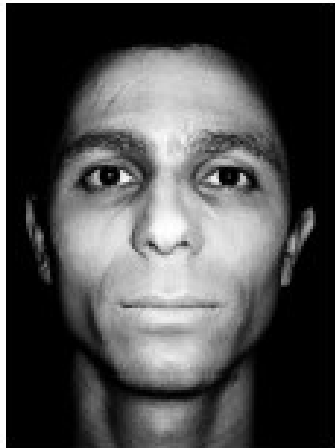


Yeah, yeah, but...

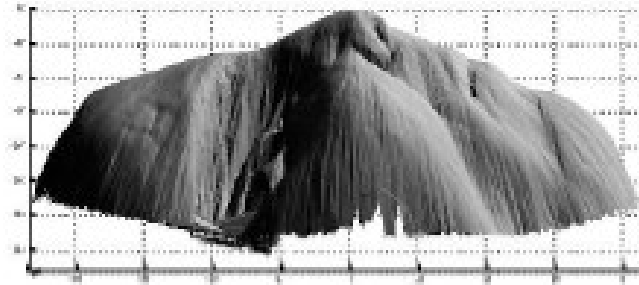
You haven't been here before, yet you probably have a fairly good understanding of this scene.



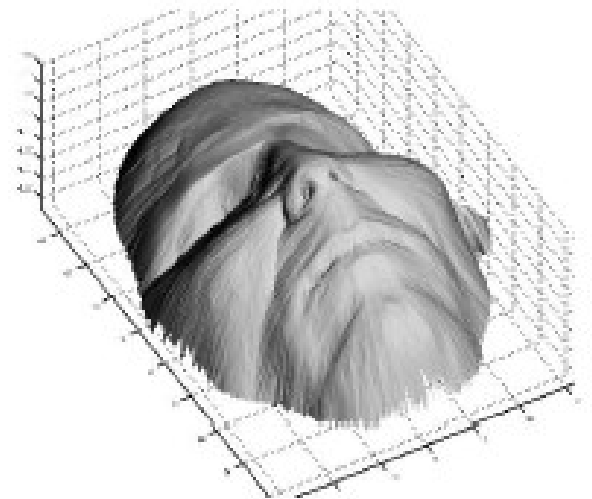
Pictorial Cues – Shading



a)



b)



c)

Pictorial Cues – Perspective effects



NATIONALGEOGRAPHIC.COM

© 2003 National Geographic Society. All rights reserved.

Pictorial Cues – Familiar Objects

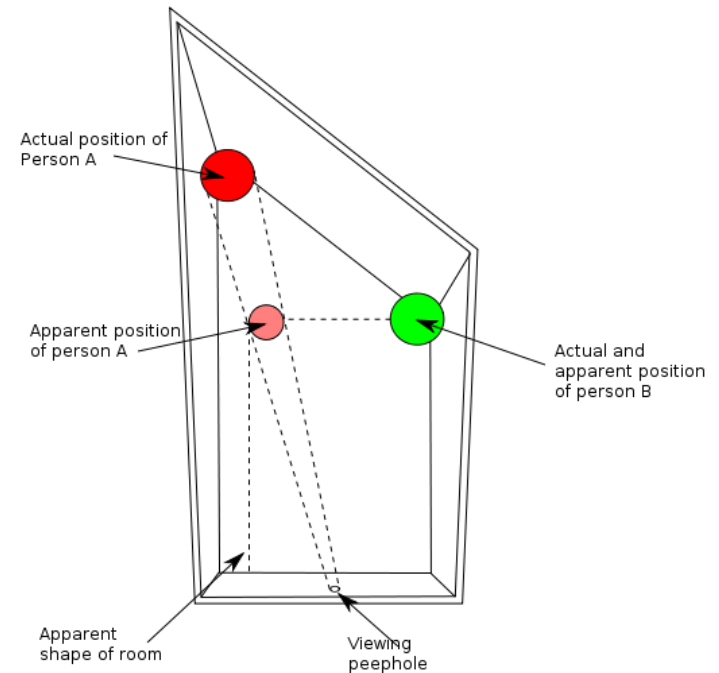


Reality of 3D Perception

- 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

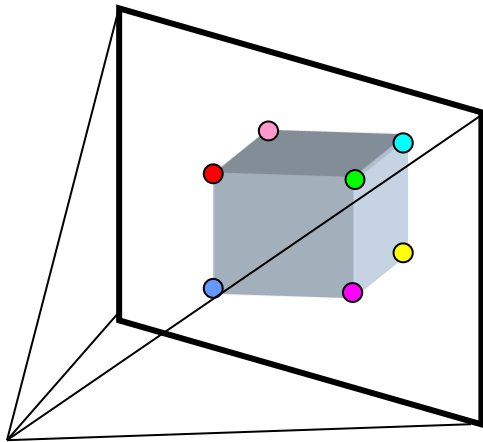
How are Cues Combined?

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

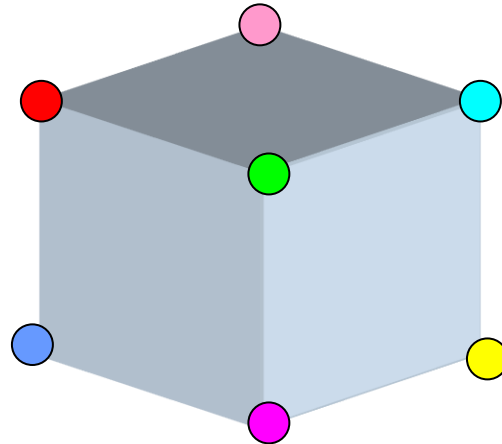


More Formally

Multi-view geometry problems



Camera 1
K ?

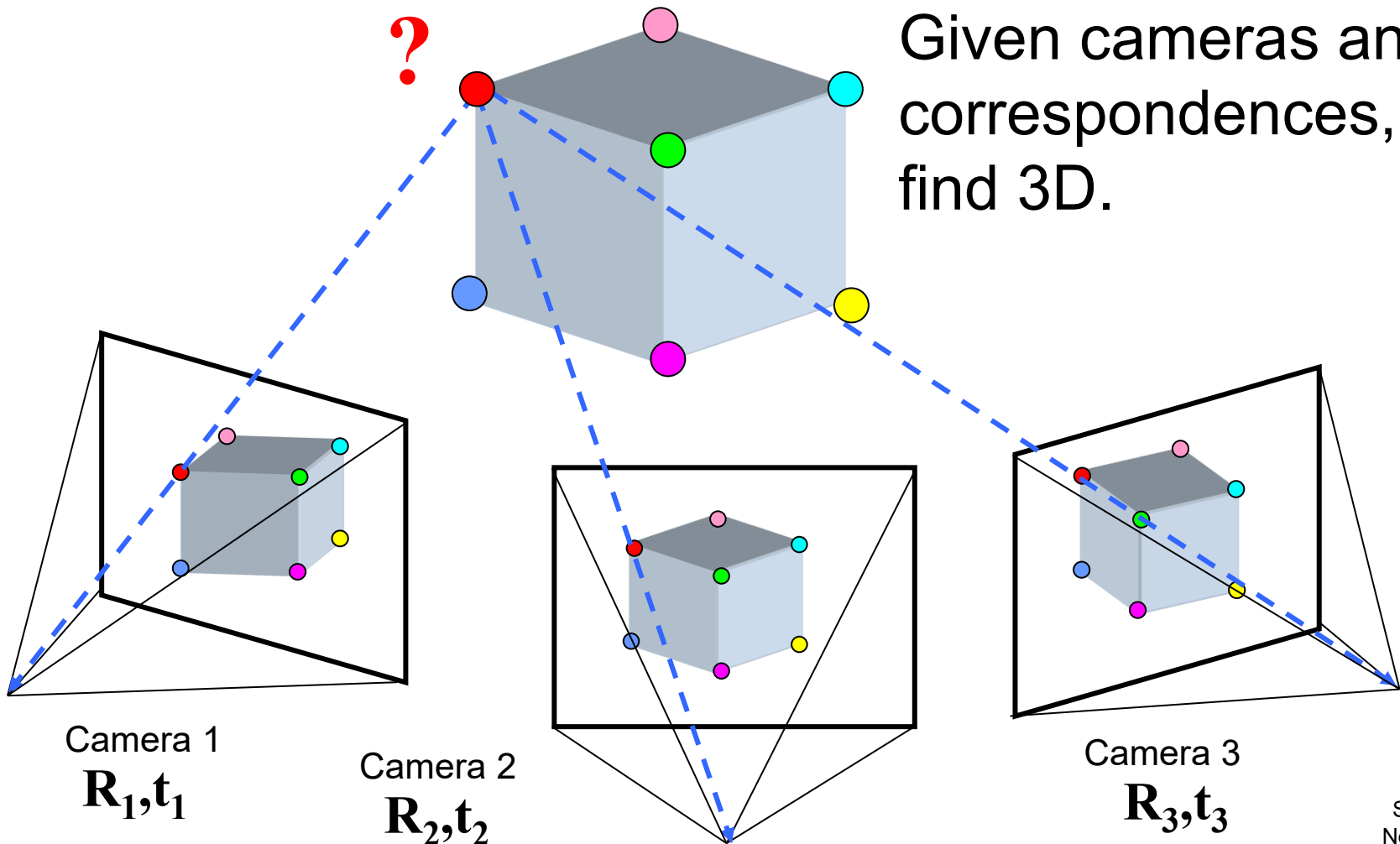


Calibration:
Figure out intrinsics
of camera (K).

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

Multi-view geometry problems

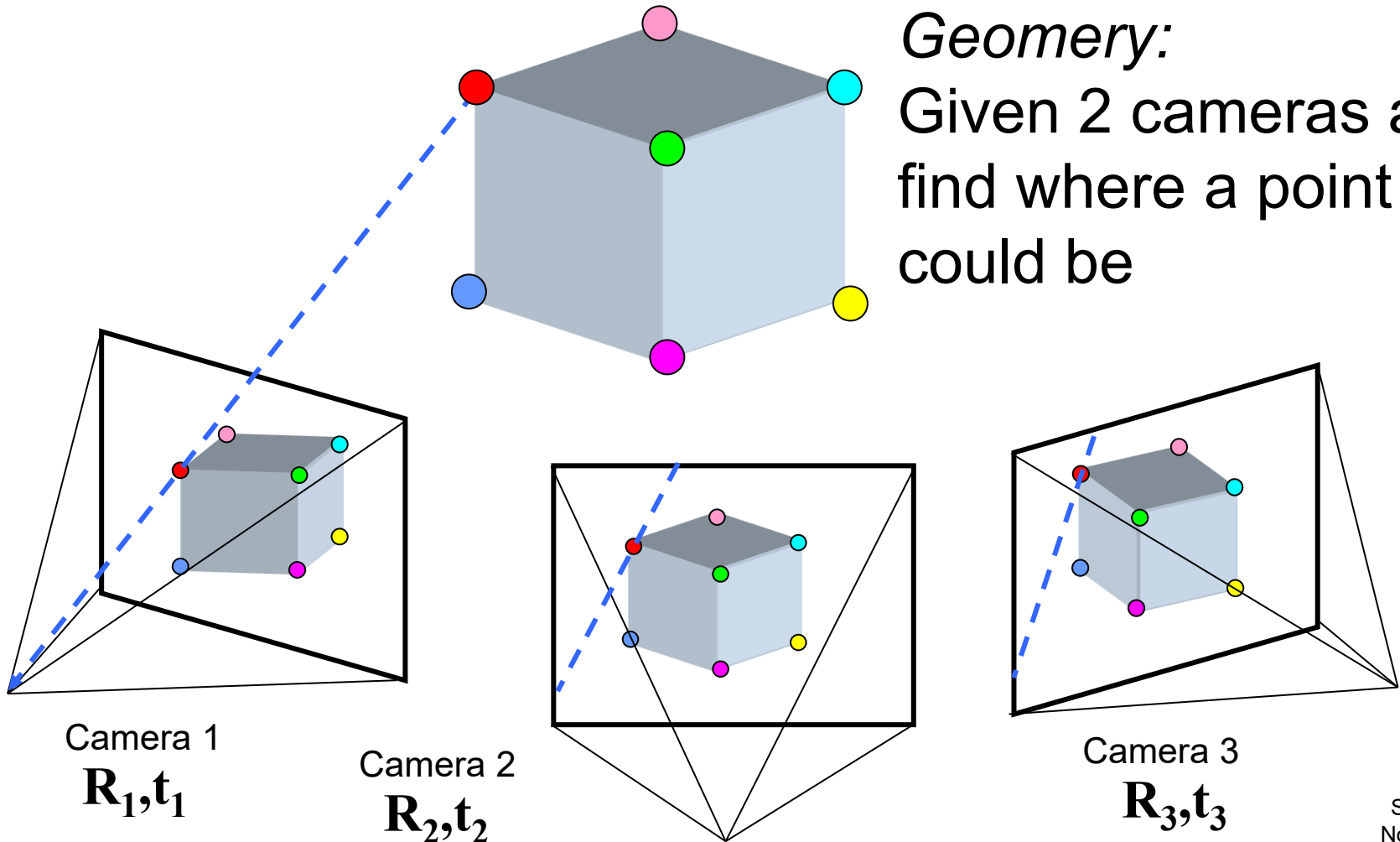
Recovering structure:
Given cameras and correspondences, find 3D.



Multi-view geometry problems

*Stereo/Epipolar
Geometry:*

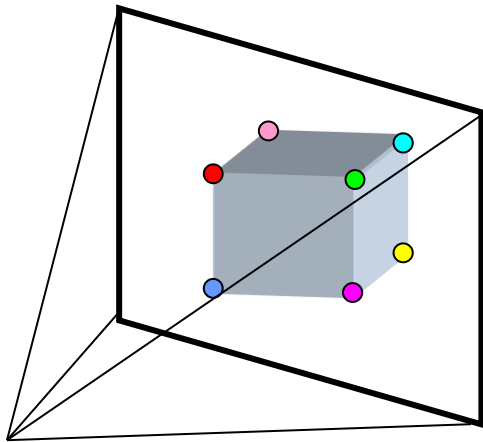
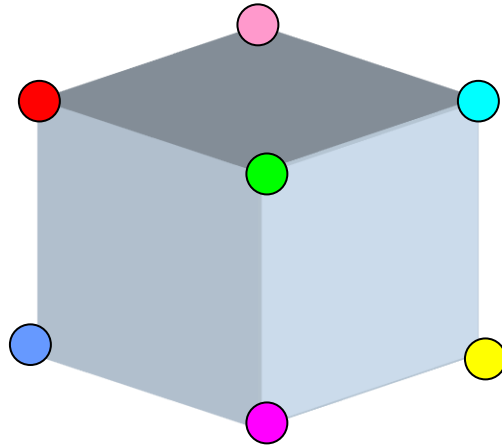
Given 2 cameras and
find where a point
could be



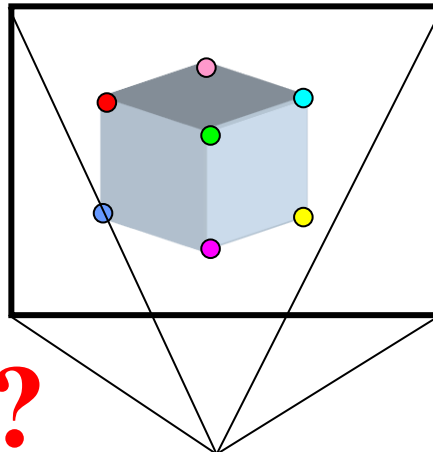
Multi-view geometry problems

Motion:

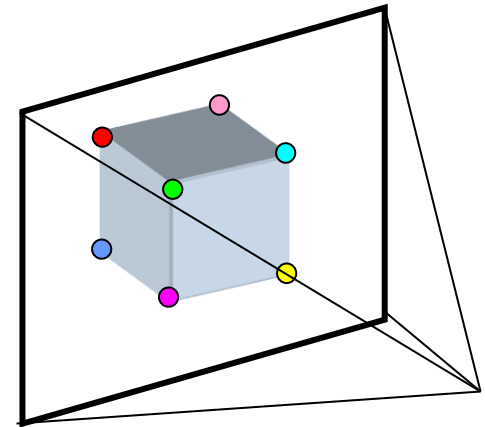
Figure out R, t for a set of cameras given correspondences



Camera 1
 R_1, t_1 ?



Camera 2
 R_2, t_2 ?

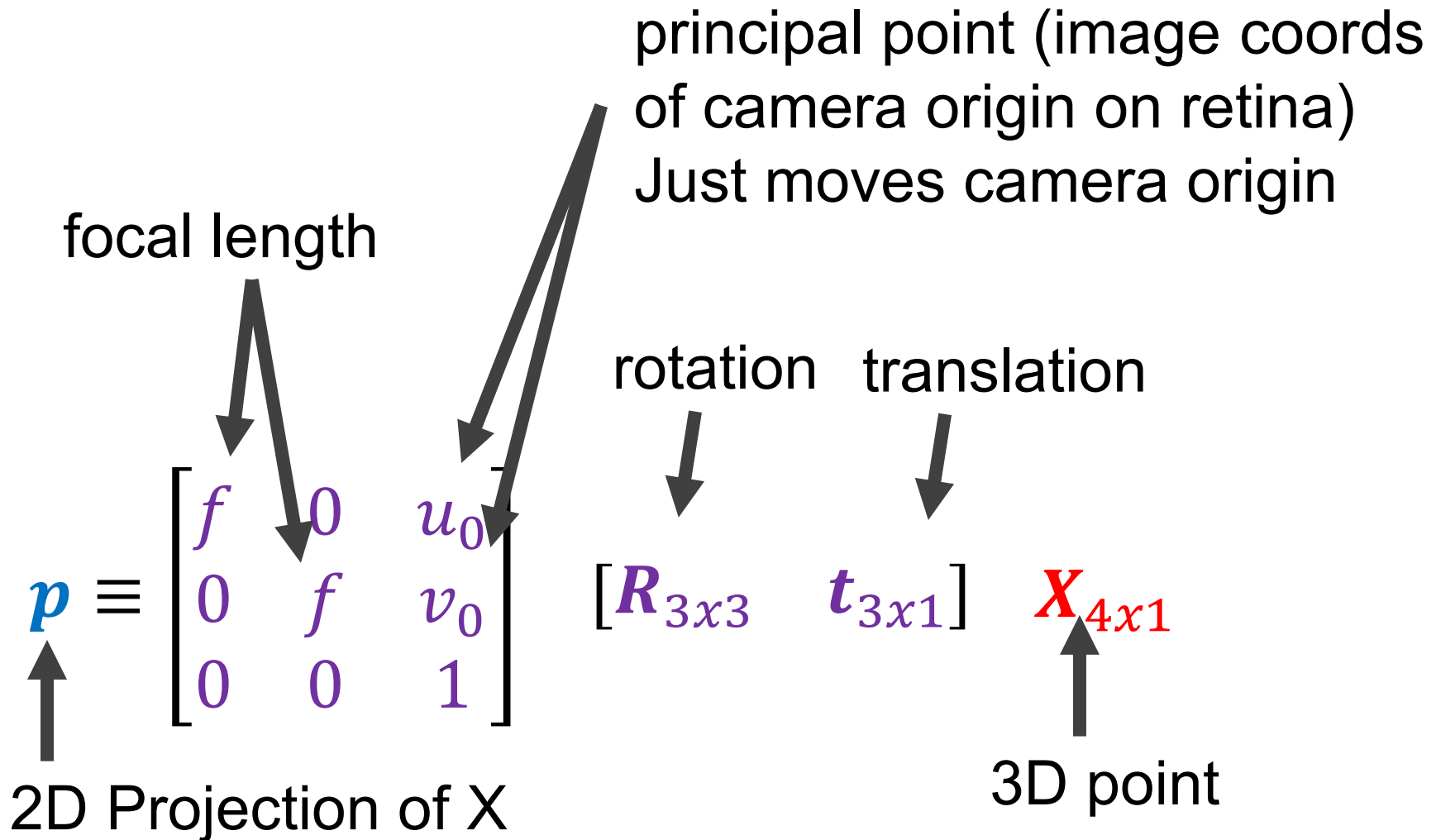


? Camera 3
 R_3, t_3

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



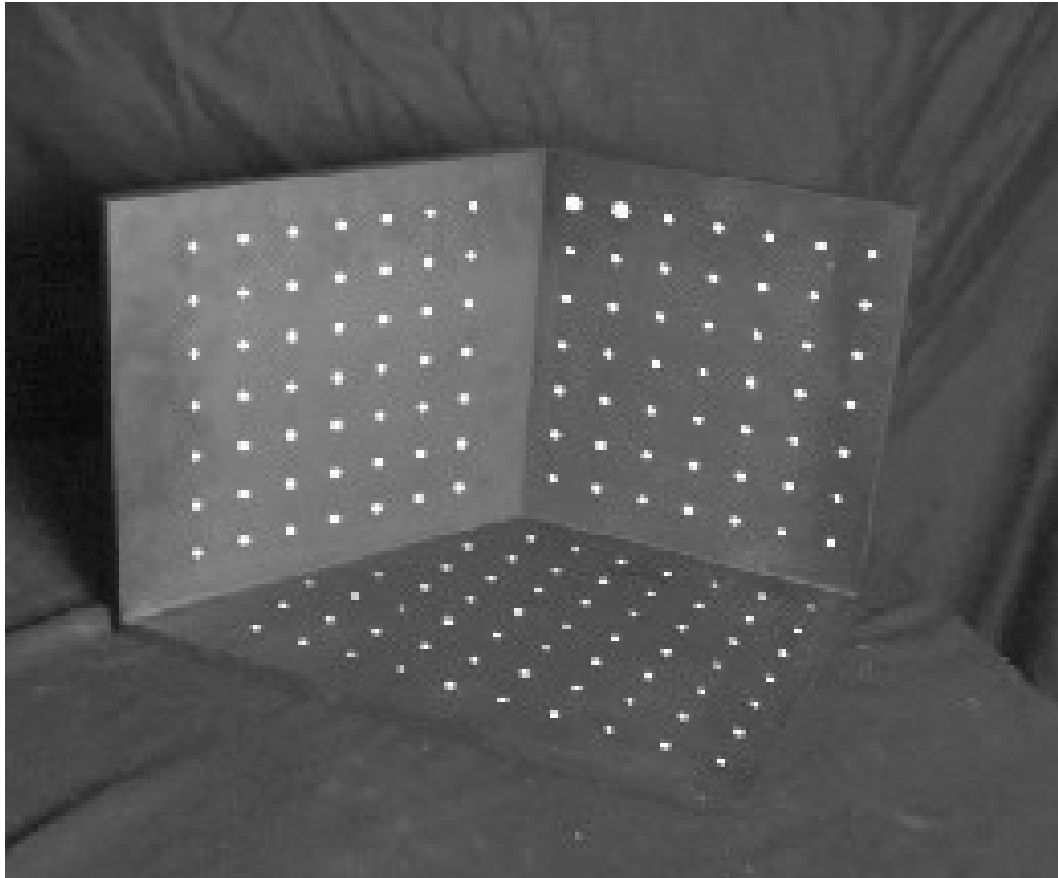
Camera Calibration

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

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

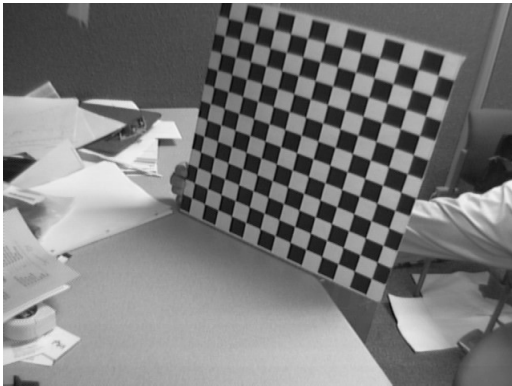
880	214
43	203
270	197
886	347
745	302
943	128
476	590
419	214
317	335
783	521
235	427
665	429
655	362
427	333
412	415
746	351
434	415
525	234
716	308
602	187

Known 3d
locations

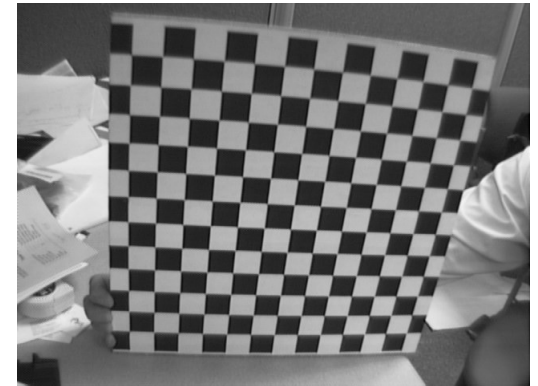
312.747	309.140	30.086
305.796	311.649	30.356
307.694	312.358	30.418
310.149	307.186	29.298
311.937	310.105	29.216
311.202	307.572	30.682
307.106	306.876	28.660
309.317	312.490	30.230
307.435	310.151	29.318
308.253	306.300	28.881
306.650	309.301	28.905
308.069	306.831	29.189
309.671	308.834	29.029
308.255	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

Camera Calibration Targets

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

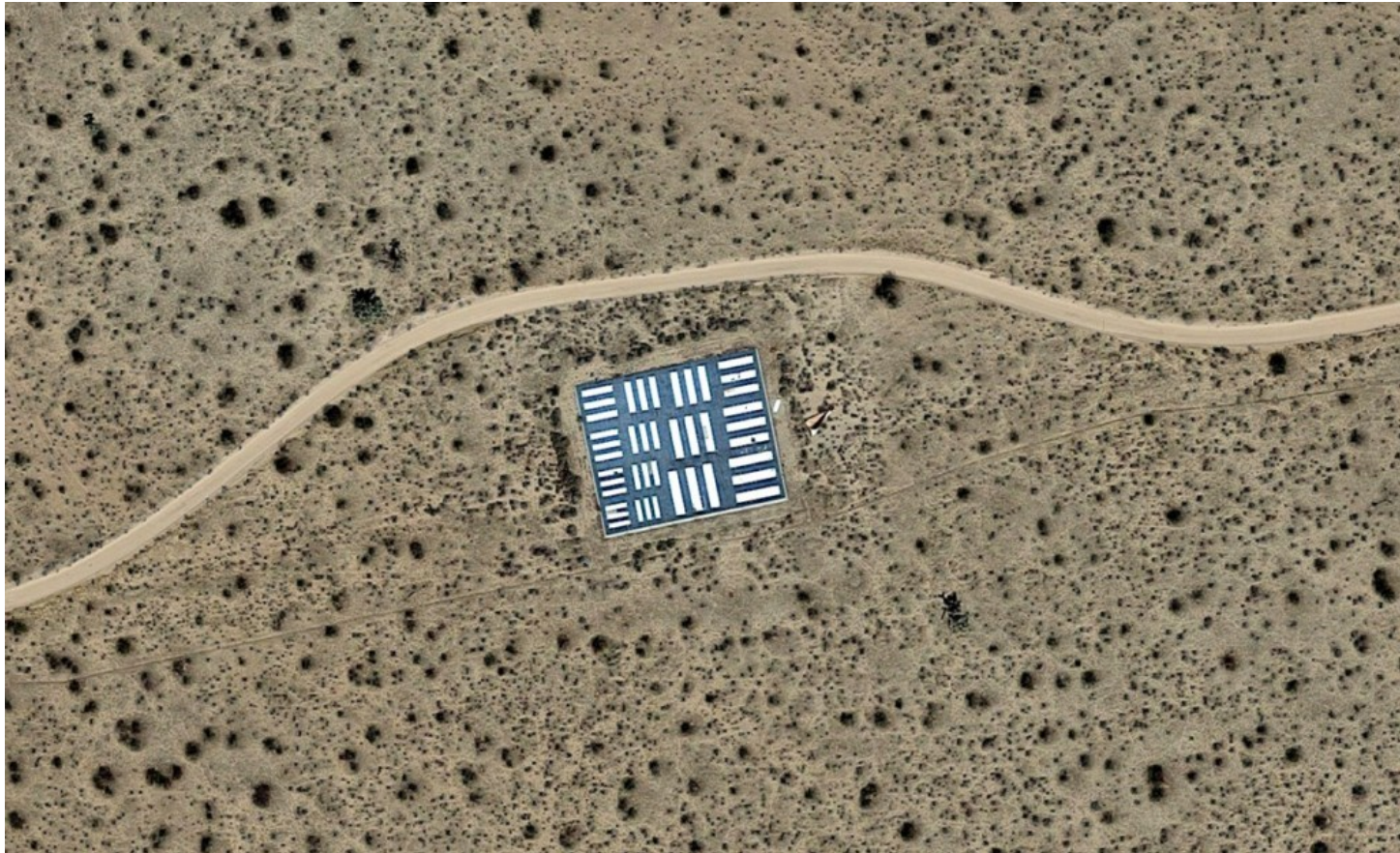


...



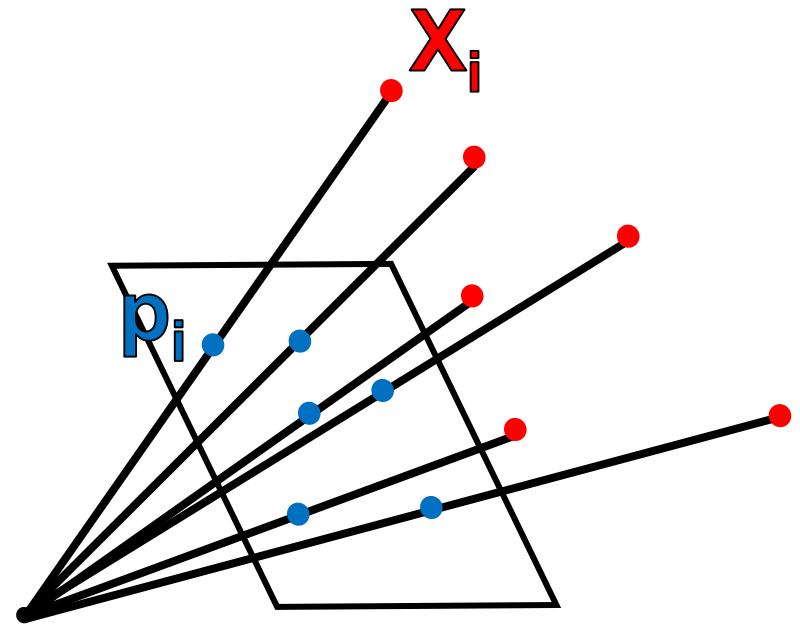
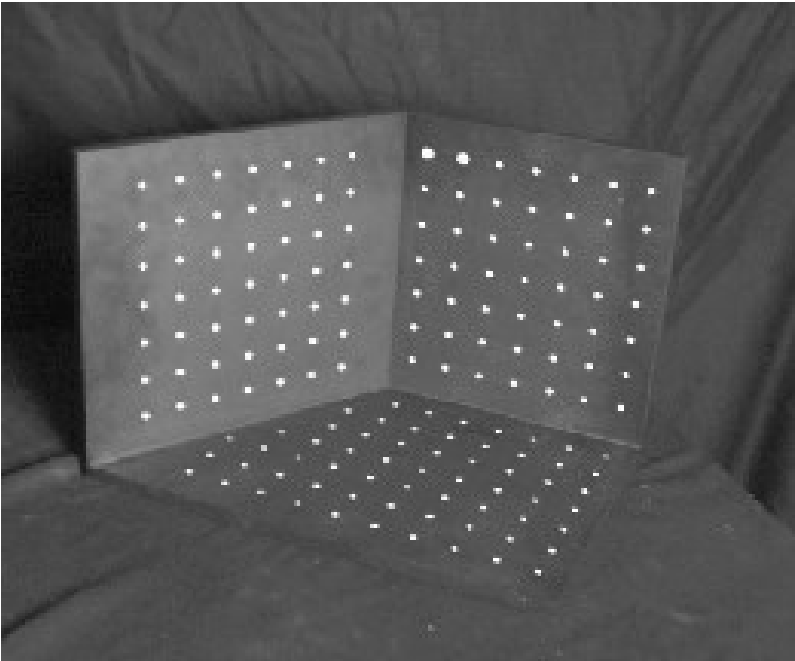
Camera Calibration Targets

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



Camera calibration

- Given n points with known 3D coordinates X_i and known image projections 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 proportional/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)

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

Camera Calibration: Linear Method

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

How many linearly independent equations?

2

How many equations per $[\mathbf{u}, \mathbf{v}] + [\mathbf{X}, \mathbf{Y}, \mathbf{Z}]$ pair?

2

If \mathbf{M} is 3×4 , how many degrees of freedom?

11

Camera Calibration: Linear Method

$$\begin{bmatrix} \mathbf{0}^T & \mathbf{X}_1^T & -v_1 \mathbf{X}_1^T \\ \mathbf{X}_1^T & \mathbf{0}^T & -u_1 \mathbf{X}_1^T \\ \dots & \dots & \dots \\ \mathbf{0}^T & \mathbf{X}_n^T & -v_n \mathbf{X}_n^T \\ \mathbf{X}_n^T & \mathbf{0}^T & -u_n \mathbf{X}_n^T \end{bmatrix} \begin{bmatrix} M_1^T \\ M_2^T \\ M_3^T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

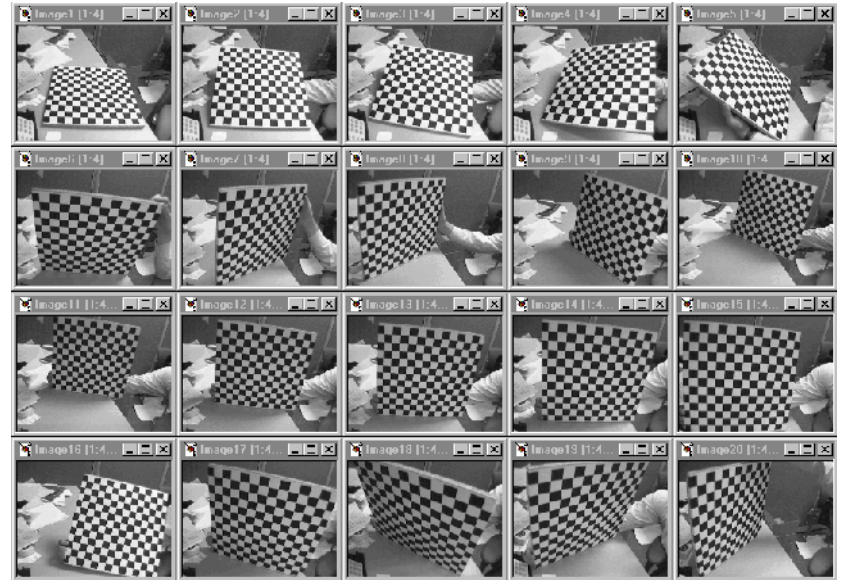
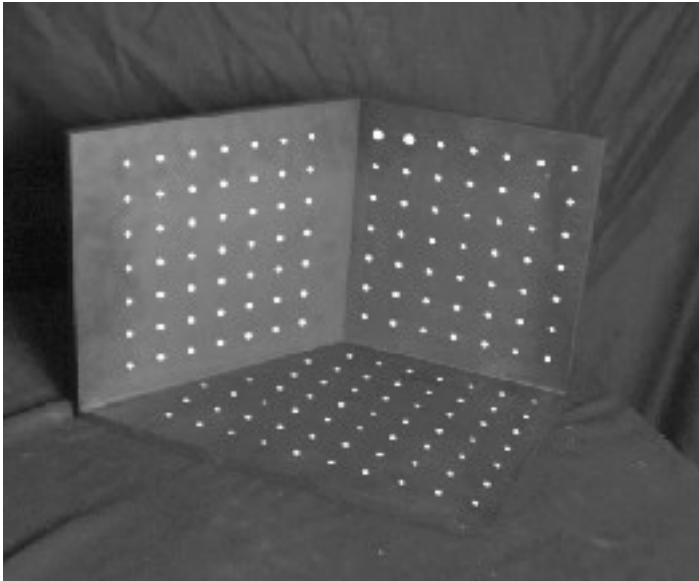
How do we solve problems of the form

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

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 on left-most 3x3 matrix
→ finite options of a upper triangular matrix * rotation

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:

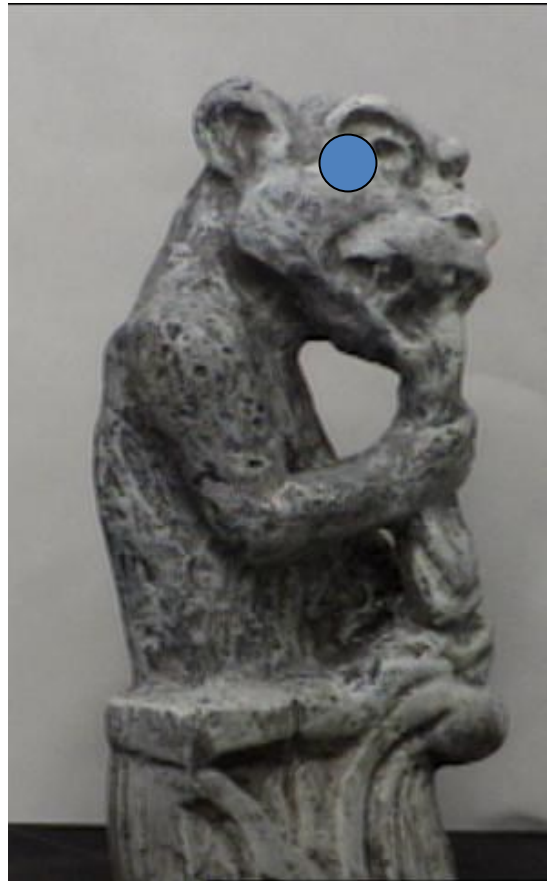
- 1) initialize parameters with linear model
- 2) Apply off-the-shelf non-linear optimizer to:

$$\sum \left\| \text{proj}(\mathbf{M}\mathbf{X}_i) - [\mathbf{u}_i, \mathbf{v}_i]^T \right\|_2^2$$

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

What Does This Get You?

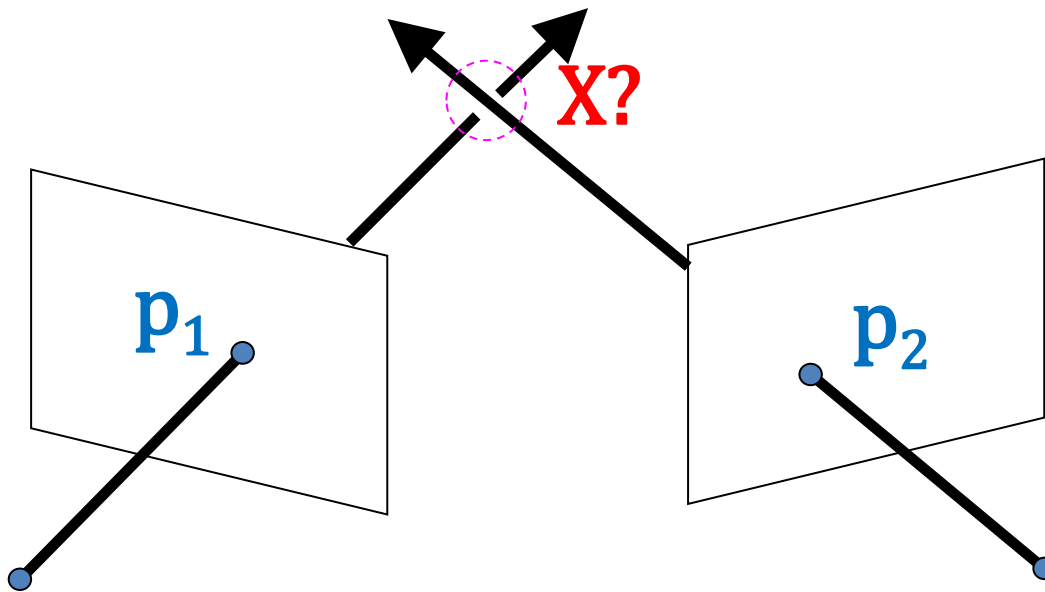
Given projection \mathbf{p}_i of unknown 3D point \mathbf{X} in two or more images (with known cameras \mathbf{M}_i), find \mathbf{X}



Triangulation

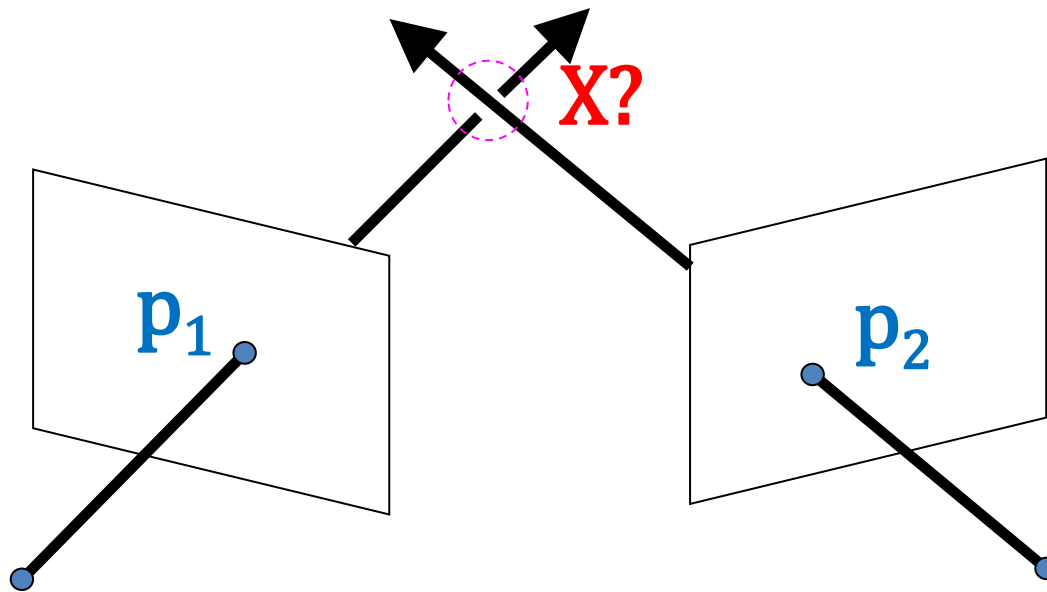
Given projection \mathbf{p}_i of unknown 3D point \mathbf{X} in two or more images (with known cameras \mathbf{M}_i), find \mathbf{X}

Why is the calibration here important?



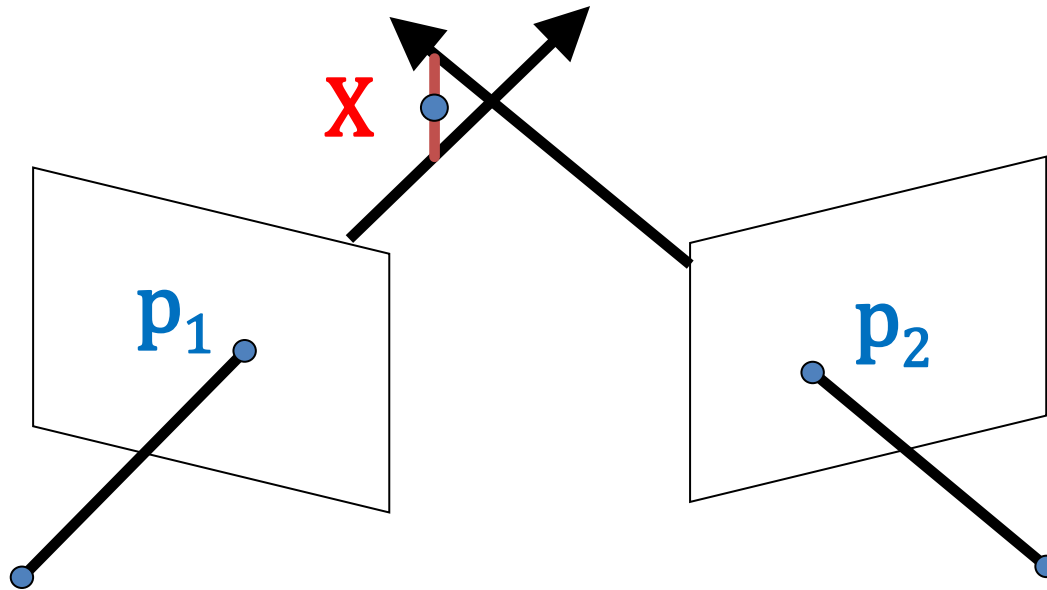
Triangulation

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



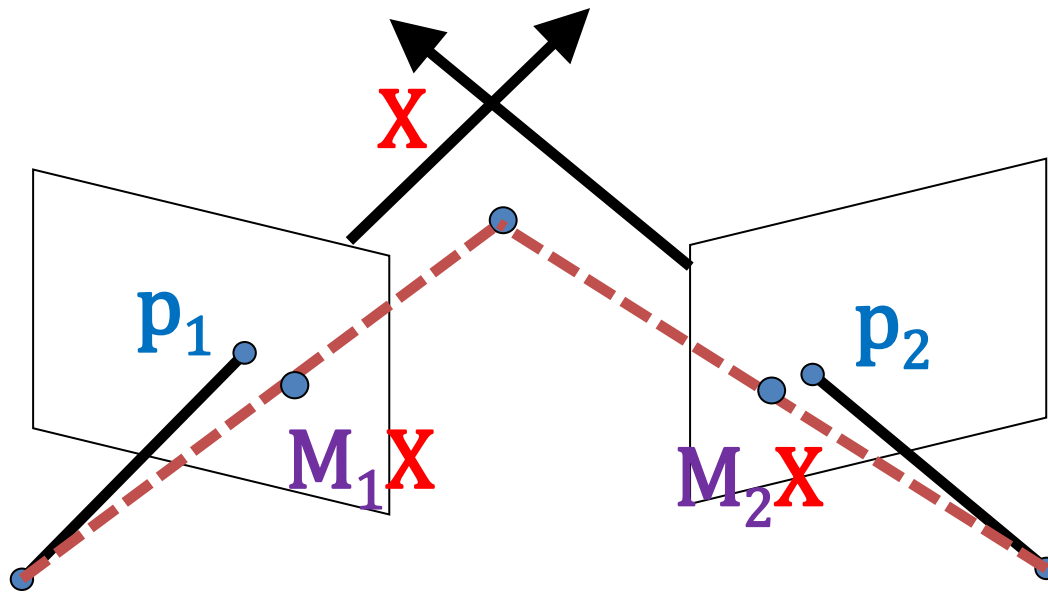
Triangulation – Geometry

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



Triangulation – Non-linear Optim.

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



Triangulation – Linear Optimization

$$\begin{array}{l} p_1 \equiv M_1 X \\ p_2 \equiv M_2 X \end{array} \Rightarrow \begin{array}{l} p_1 \times M_1 X = 0 \\ p_2 \times M_2 X = 0 \end{array} \Rightarrow \begin{array}{l} [p_{1x}] M_1 X = 0 \\ [p_{2x}] M_2 X = 0 \end{array}$$

Cross Prod.
as matrix

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

$$\begin{array}{l} [p_{1x}] M_1 X = 0 \\ [p_{2x}] M_2 X = 0 \end{array} \Rightarrow \begin{array}{l} ([p_{1x}] M_1) X = 0 \\ ([p_{2x}] M_2) X = 0 \end{array} \Rightarrow \begin{array}{l} \text{Two eqns per} \\ \text{camera for 3} \\ \text{unkn. in } X \end{array}$$

Summarizing

- 3D is complicated
- Given $\mathbf{p} = \mathbf{M}\mathbf{X}$, you can derive equations that let you solve for \mathbf{M} (calibration) or \mathbf{X} (triangulation)
- Next time: what can you do from a single image itself?

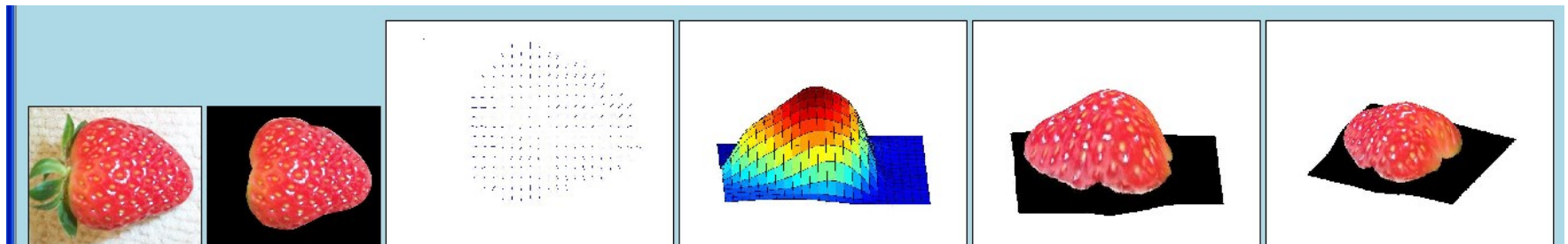
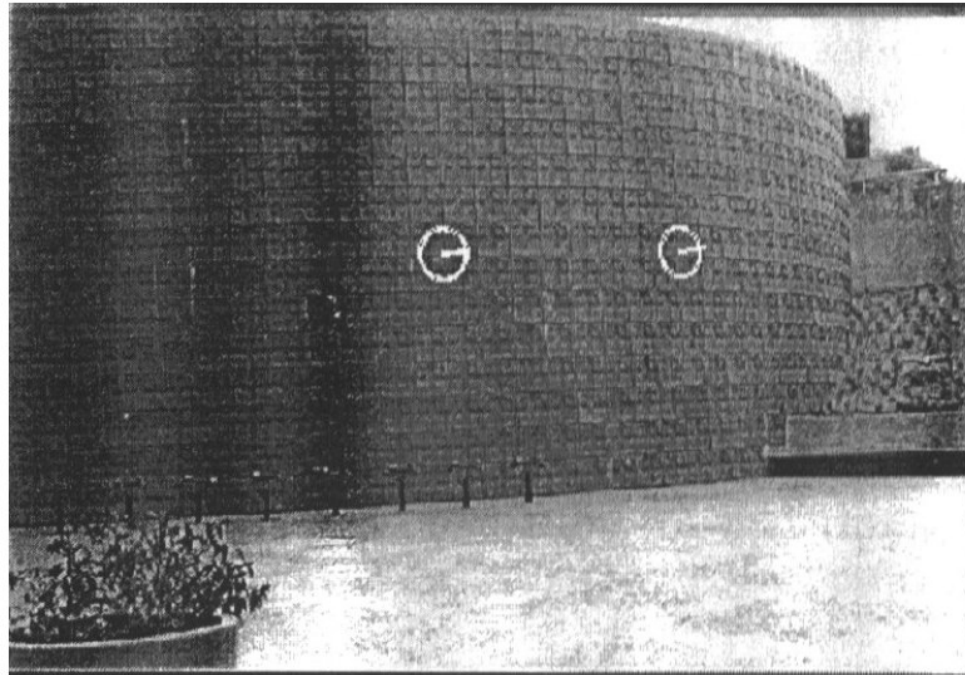
Bonus Fun

Single-view Ambiguity



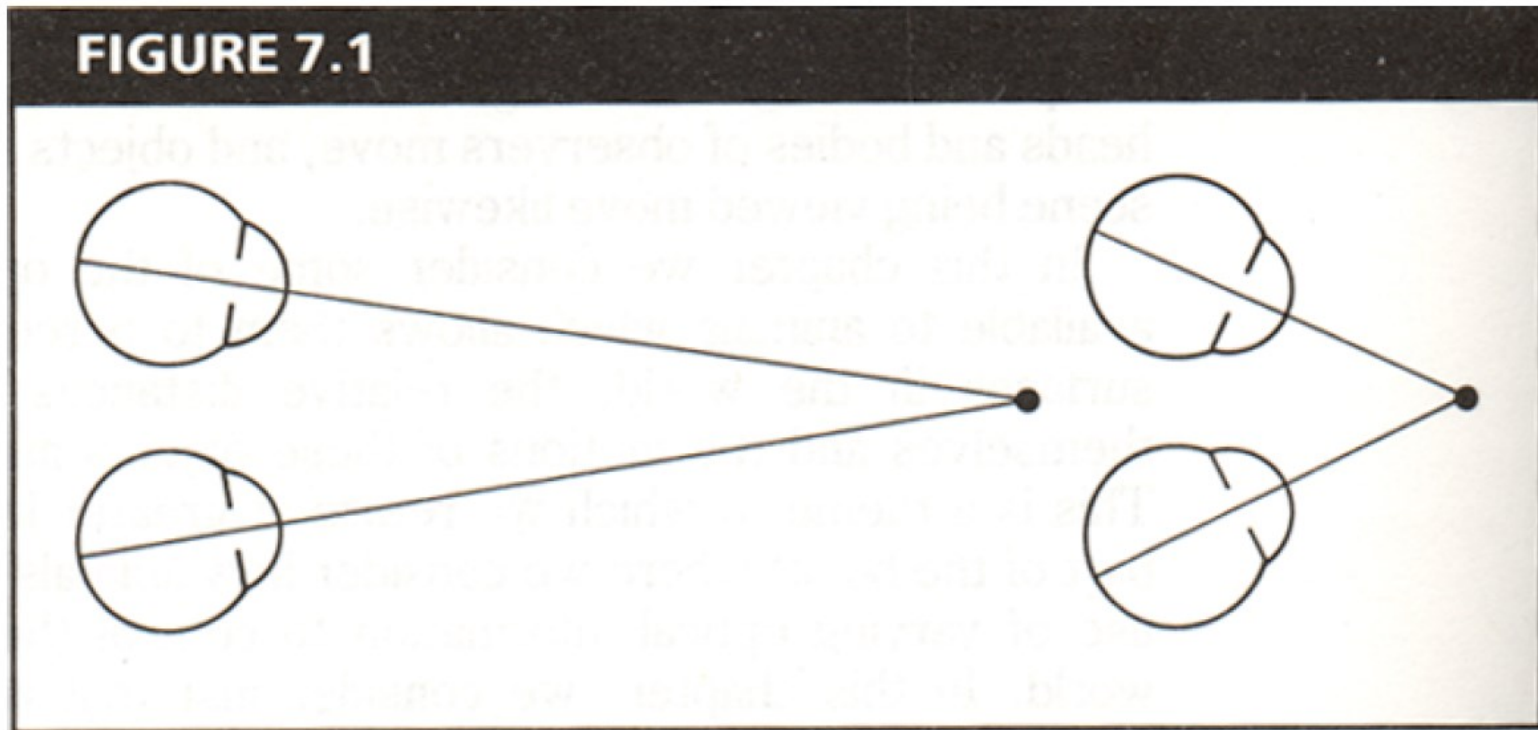
[Rashad Alakbarov shadow sculptures](#)

Pictorial Cues – Texture



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

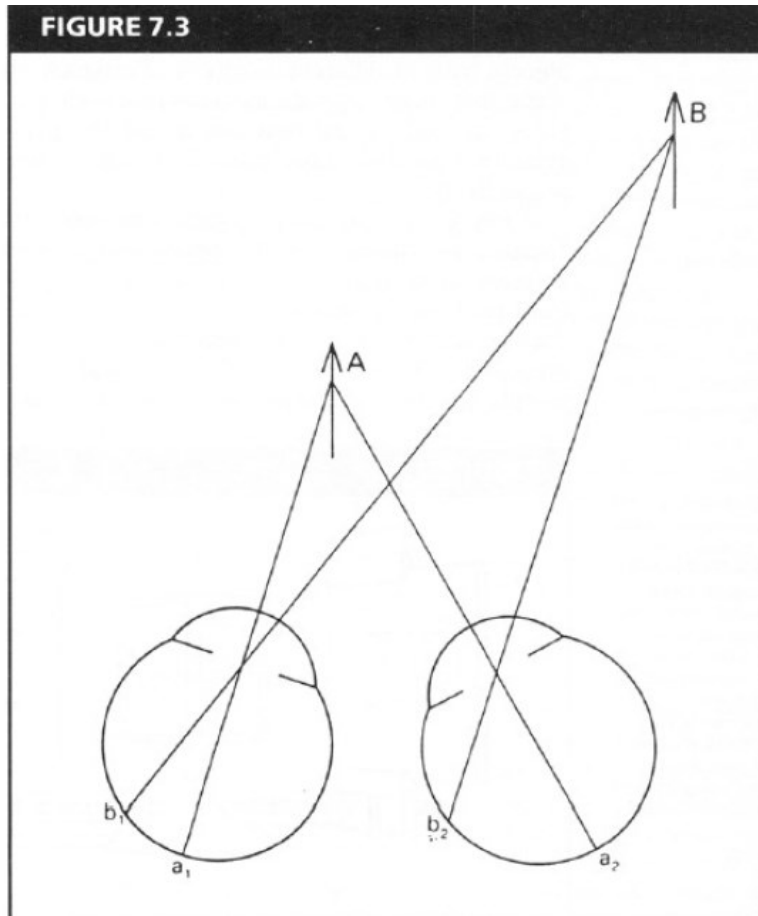
Human stereopsis: disparity



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



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

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

Projection Matrix

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

