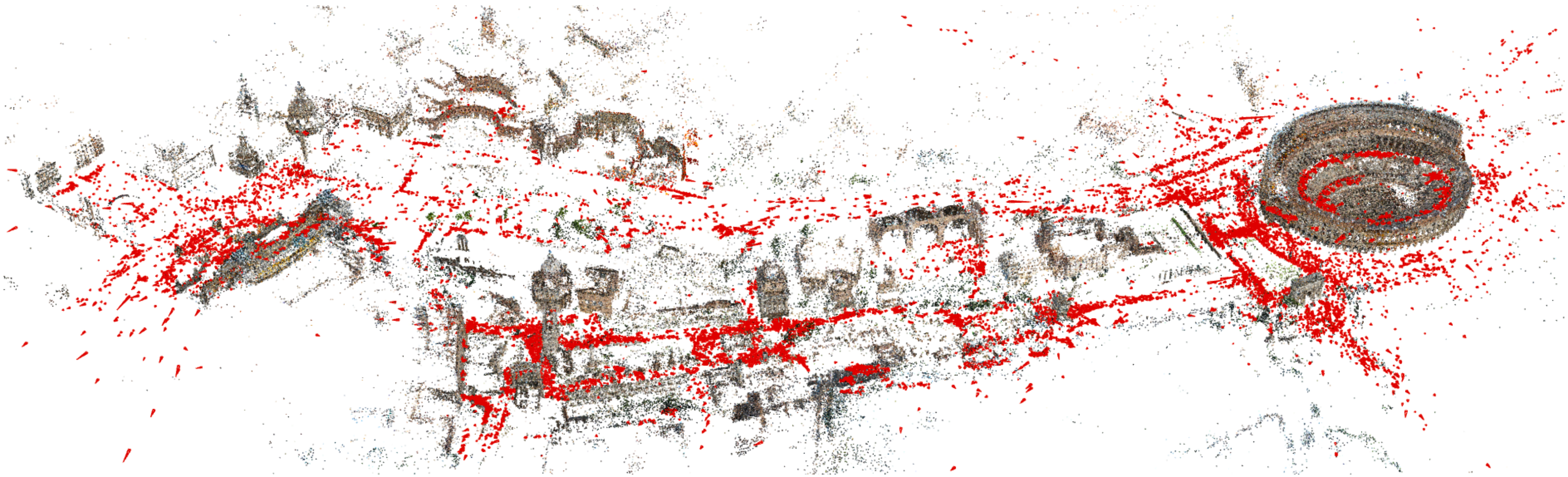
3D Vision: Epipolar Geometry



A lot of slides borrowed from Noah Snavely + Shree Nayar's YT series: First principals of Computer Vision

CS194: Intro to Computer Vision and Comp. Photo
Angjoo Kanazawa, UC Berkeley, Fall 2021

More cool things with 3D

3D photo

AR

Recap: Camera Model

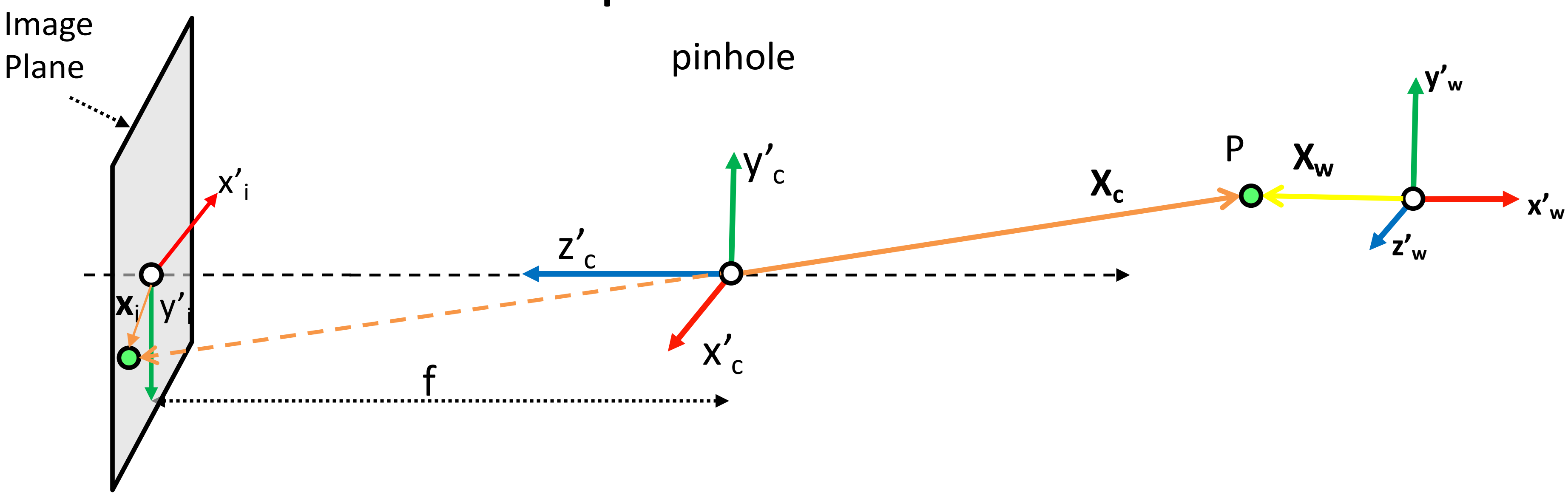


Image Coordinates

Camera Coordinates

World Coordinates

$$\mathbf{x}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

Perspective Projection

$$\mathbf{X}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}$$

Coordinate Transformation

$$\mathbf{X}_w = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}$$

$$\begin{bmatrix} f_x & 0 & o_x & 0 \\ 0 & f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} R_{3\times 3} & \mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$$

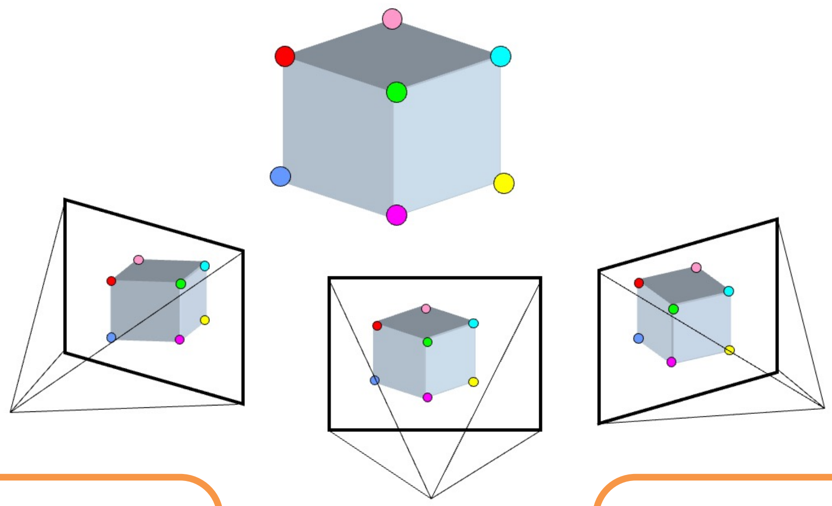
Intrinsics

Extrinsics

Slide inspired by Shree Nayar

Big picture: 3 key components in 3D

3D Points
(Structure)

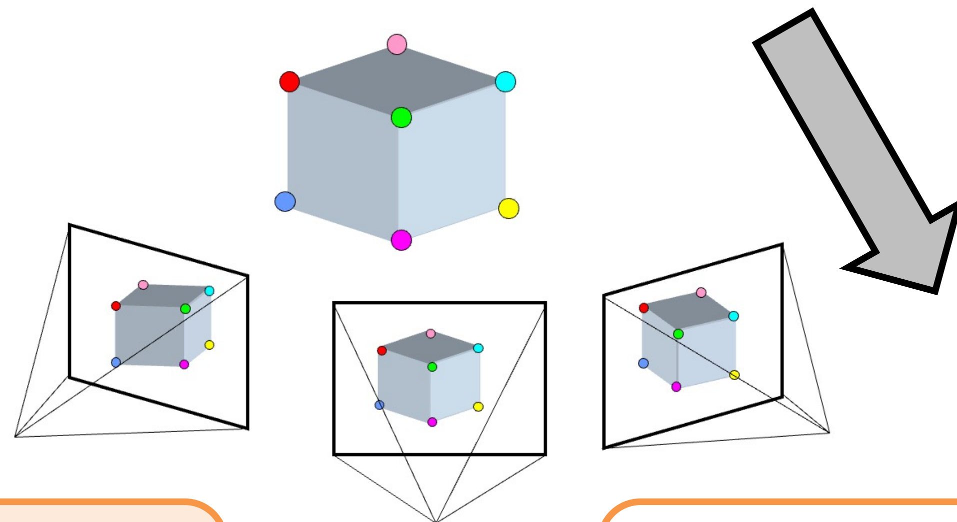


Correspondences

Camera
(Motion)

Big picture: 3 key components in 3D

3D Points
(Structure)

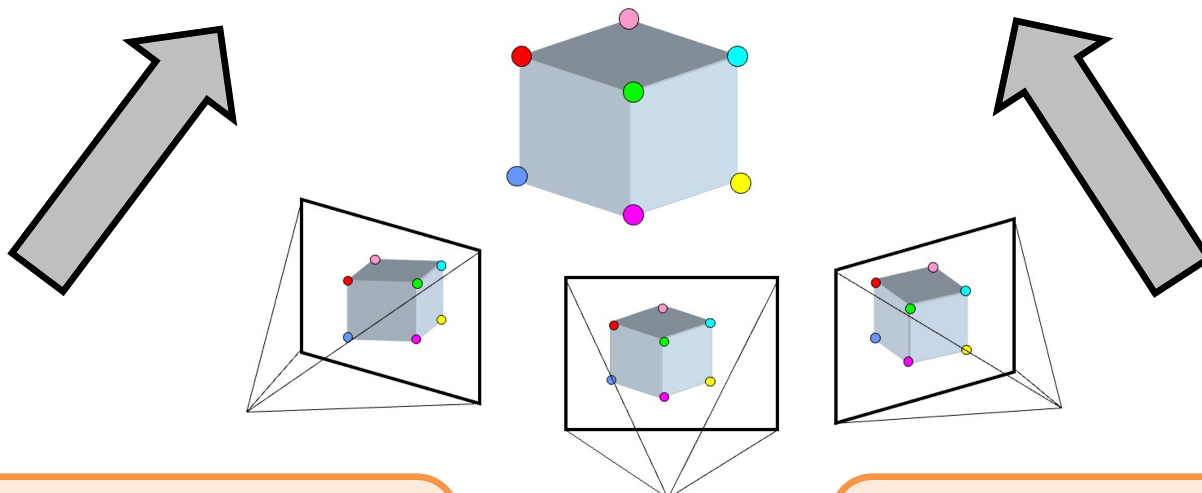


Correspondences

Camera
(Motion)

Big picture: 3 key components in 3D

3D Points
(Structure)

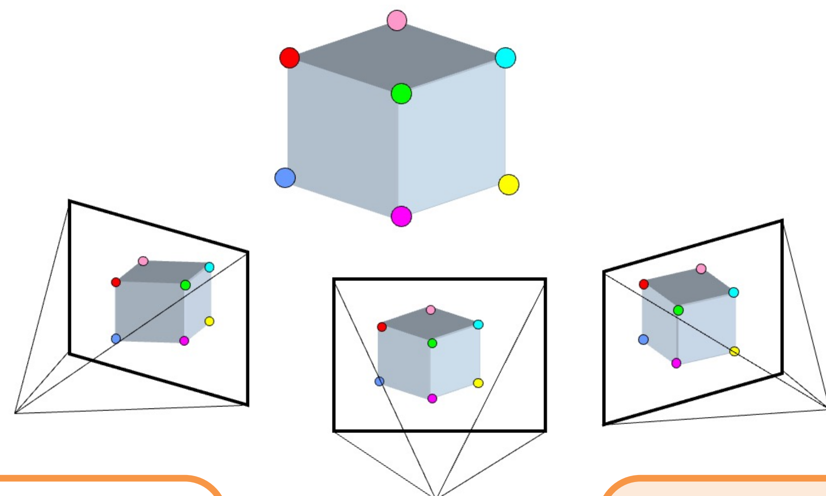


Correspondences

Camera
(Motion)

Big picture: 3 key components in 3D

3D Points
(Structure)

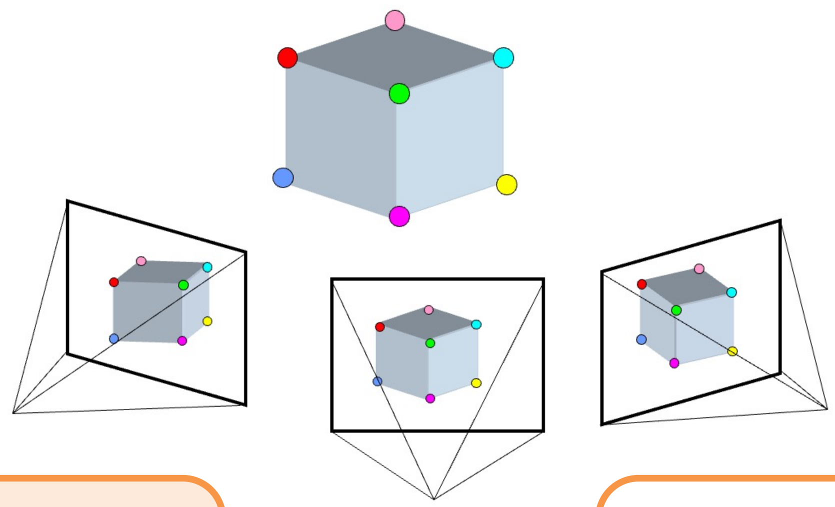


Correspondences

Camera
(Motion)

Big picture: 3 key components in 3D

3D Points
(Structure)



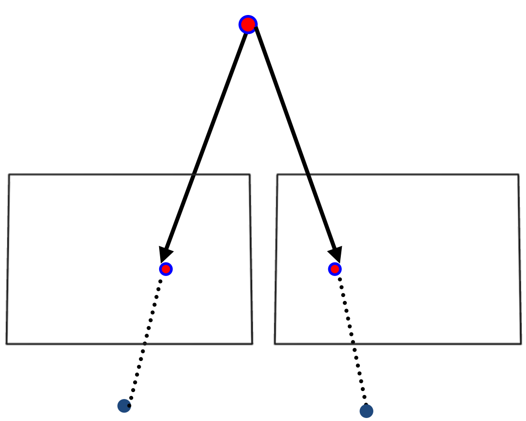
Correspondences

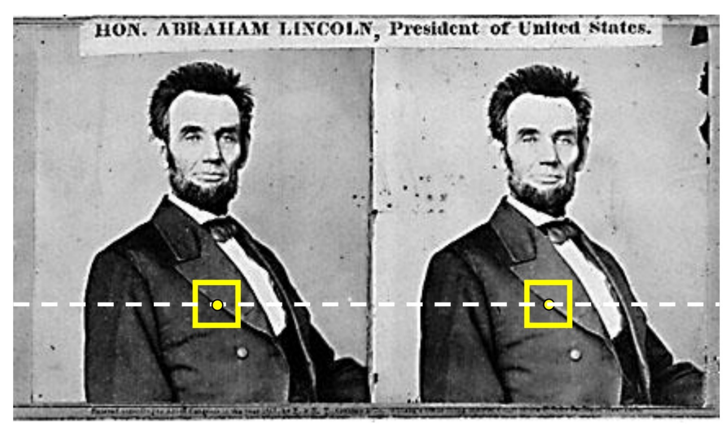
Camera
(Motion)

Recap

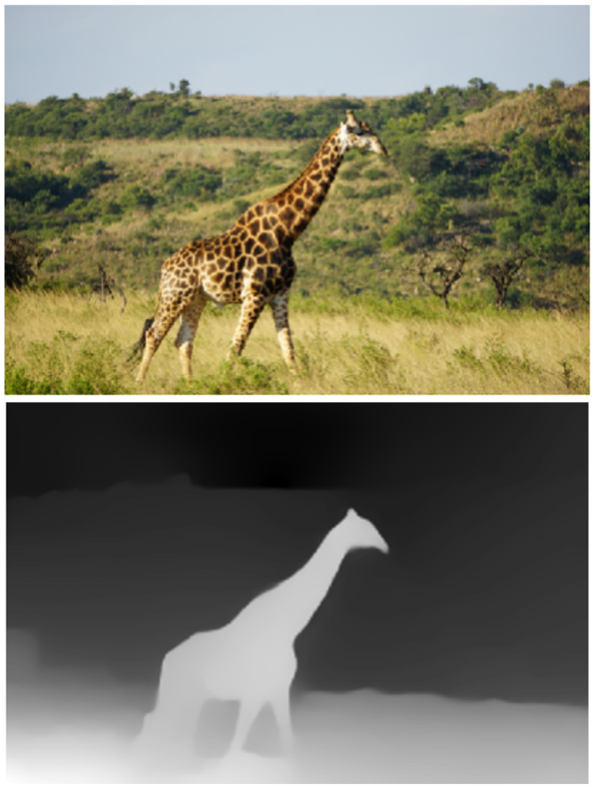
We covered:

- How to estimate the camera parameters
  - "Calibration"
  - Solve for intrinsics & extrinsics
- With a simple stereo, correspondences lie on horizontal lines
- depth is inversely proportional to disparity
  (how much the pixel moves)



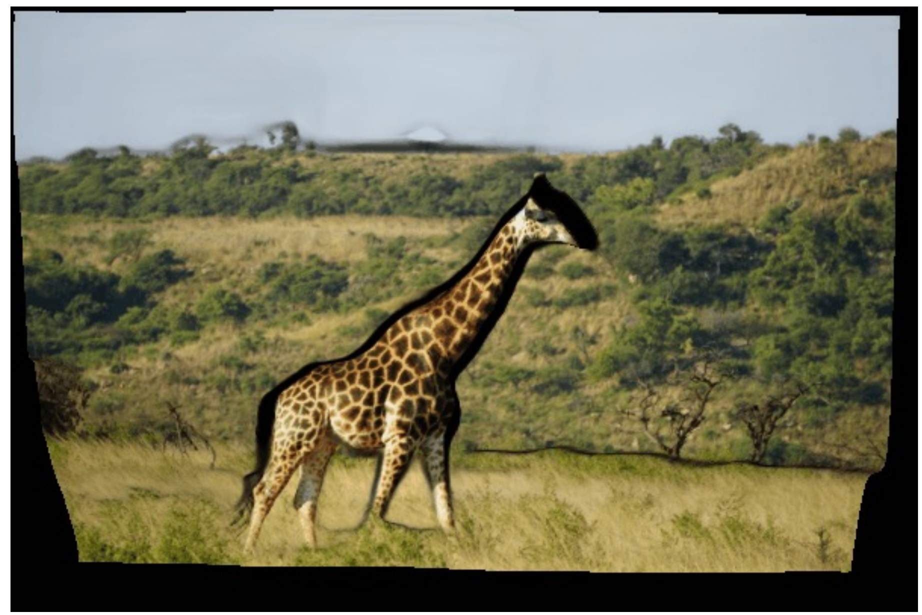


What Depth Map provides



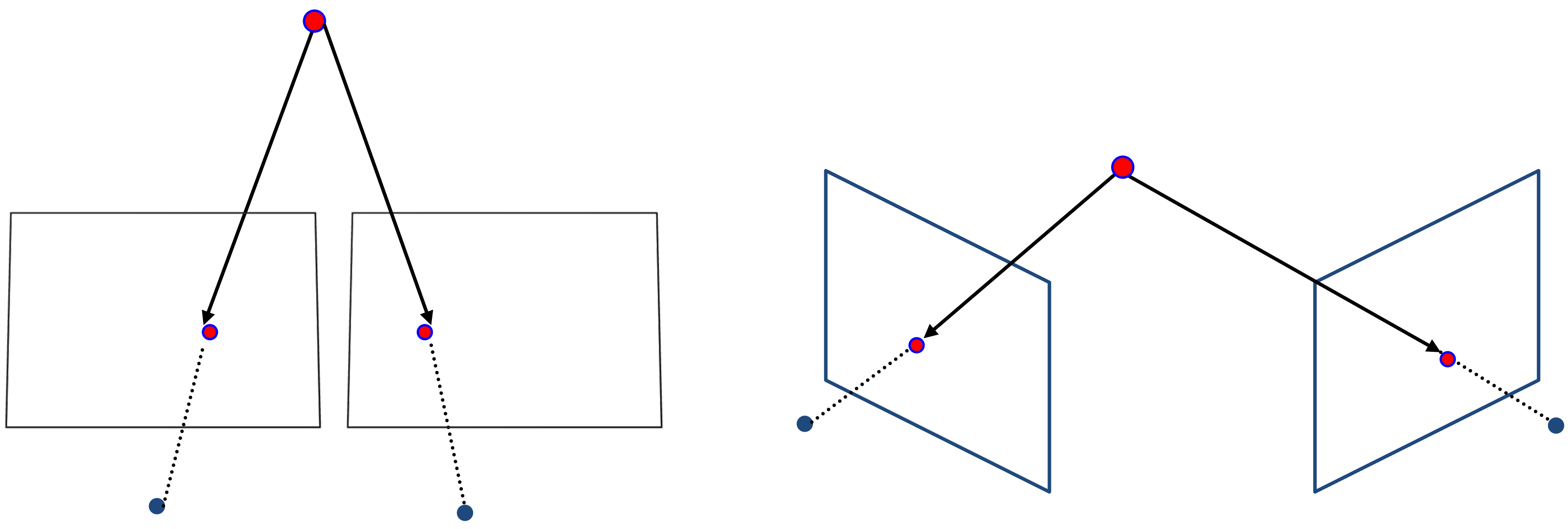
Monocular Depth Prediction [Ranftl et al. PAMI'20]

warping the pixel based on its depth as you change the views



Next: General case

- The two cameras need not have parallel optical axes.
- Assume cameras are calibrated

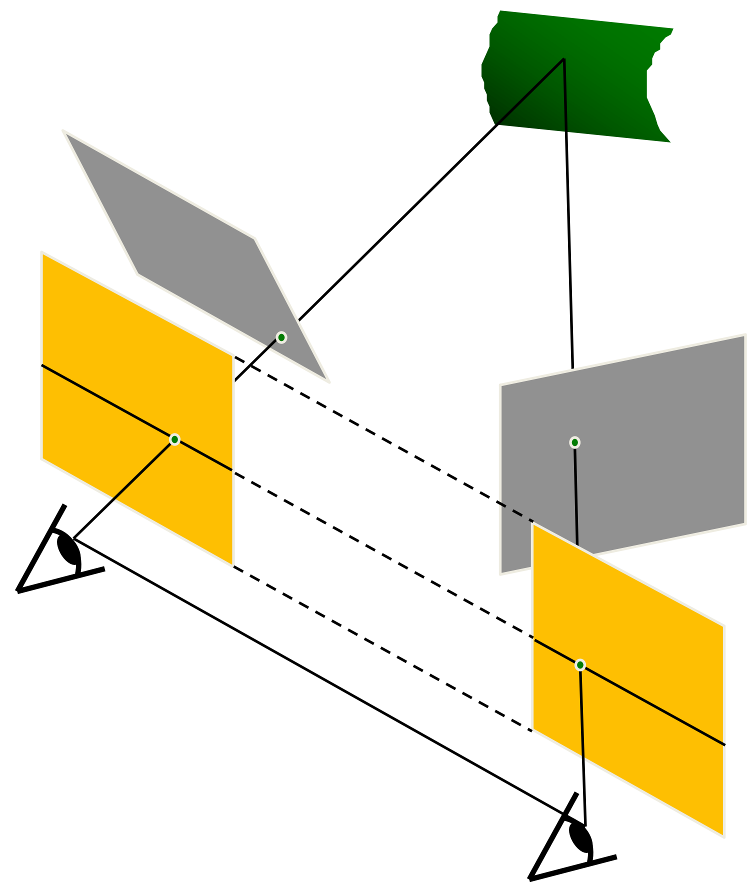


Same hammer:

Find the correspondences, then solve for structure

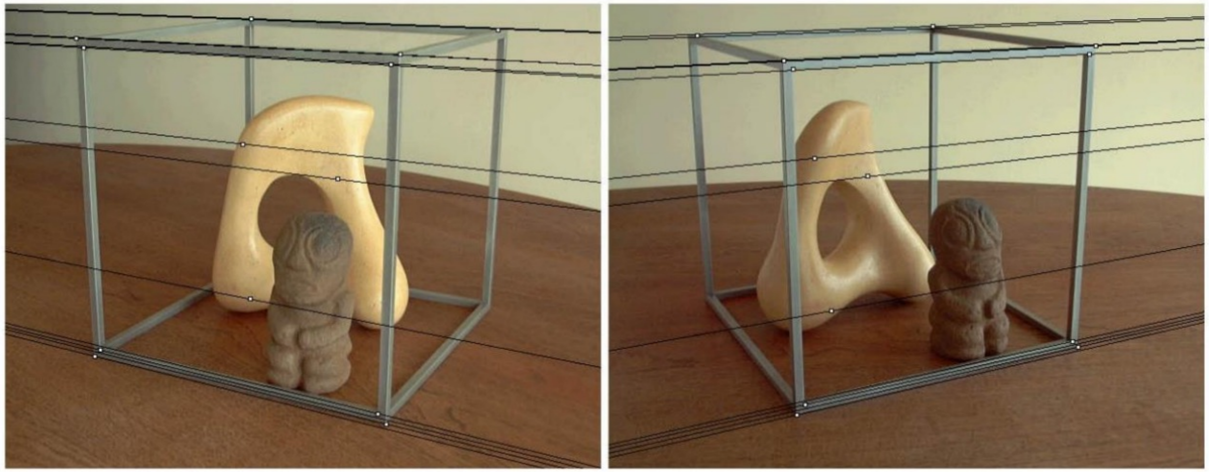
Option 1: Rectify via homography

- reproject image planes onto a common plane
  - plane parallel to the line between optical centers
- pixel motion is horizontal after this transformation
- Two homographies, one for each input image reprojection
  - C. Loop and Z. Zhang. [Computing Rectifying Homographies for Stereo Vision](). CVPR 1999.



Slide courtesy of Noah Snavely

Option 1: Rectify via homography



Original stereo pair

After rectification

Then find correspondences on the horizontal scan line

Slide courtesy of Noah Snavely

General case, known camera, find depth: Option 2

1. Find correspondences
2. Triangulate

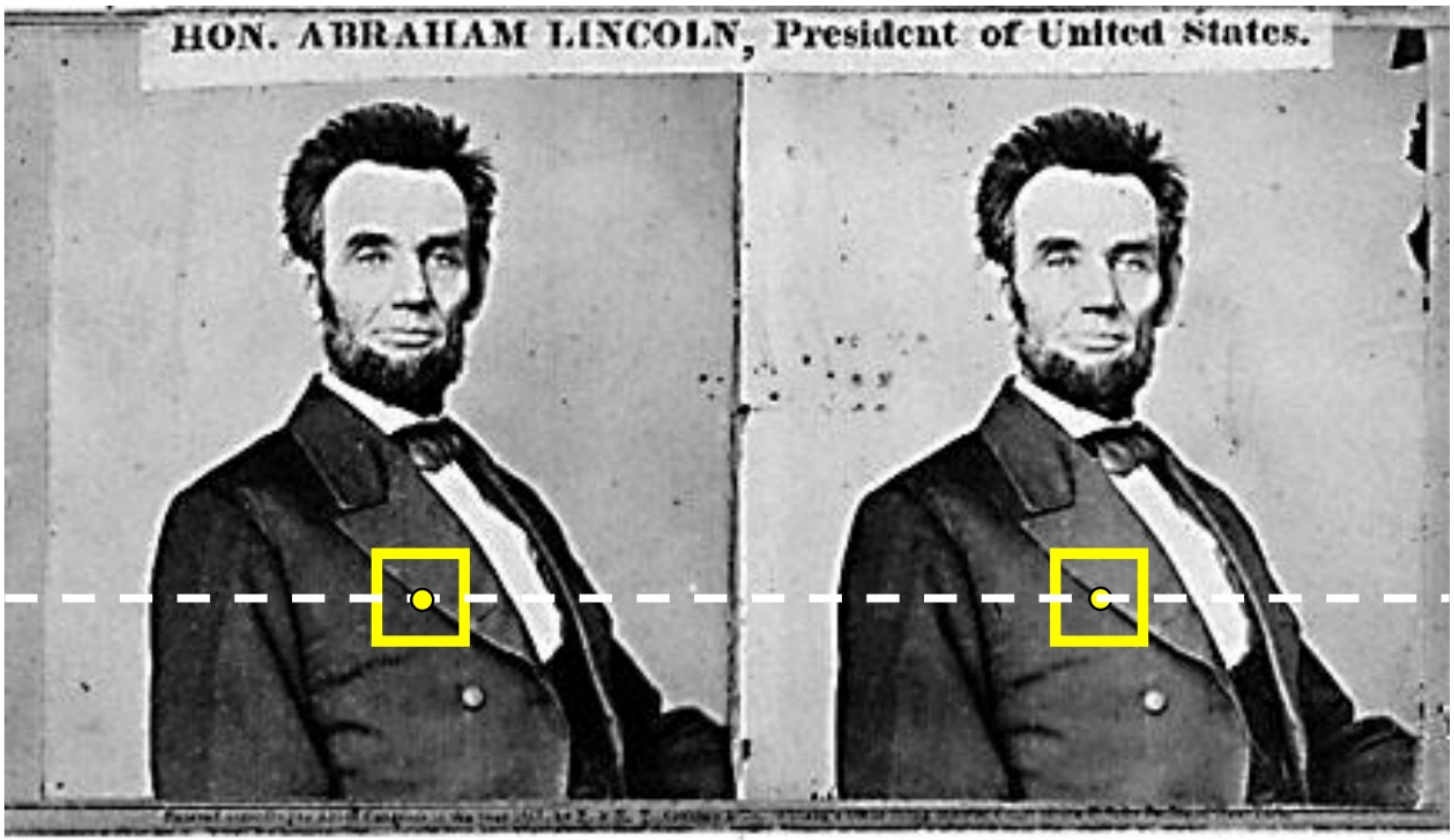
General case, known camera, find depth: Option 2

1. **Find correspondences**
2. Triangulate

Can we restrict the search space again to 1D?

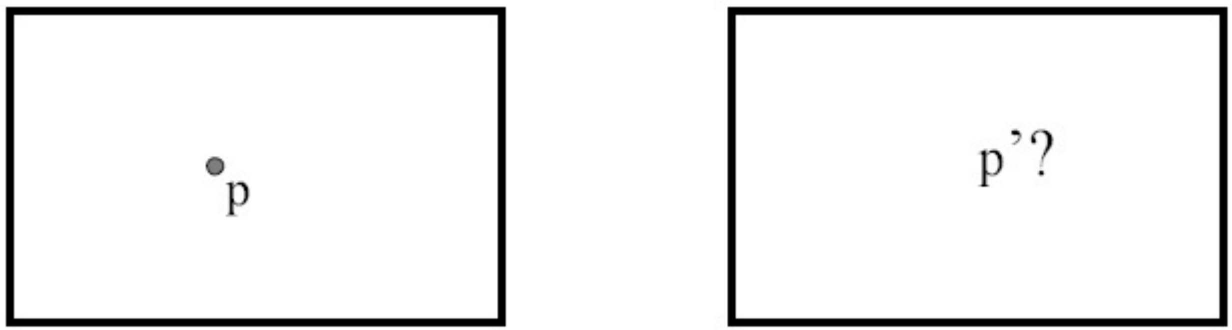
What is the relationship between the camera + the corresponding points?

Where do epipolar lines come from?



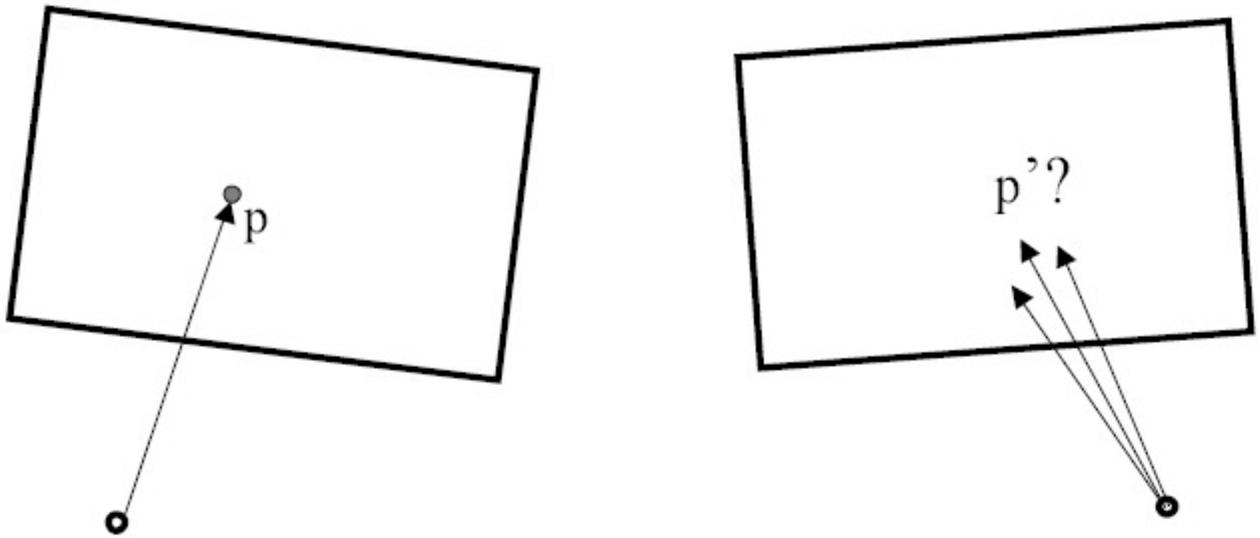
Slide courtesy of Noah Snavely

Stereo correspondence constraints



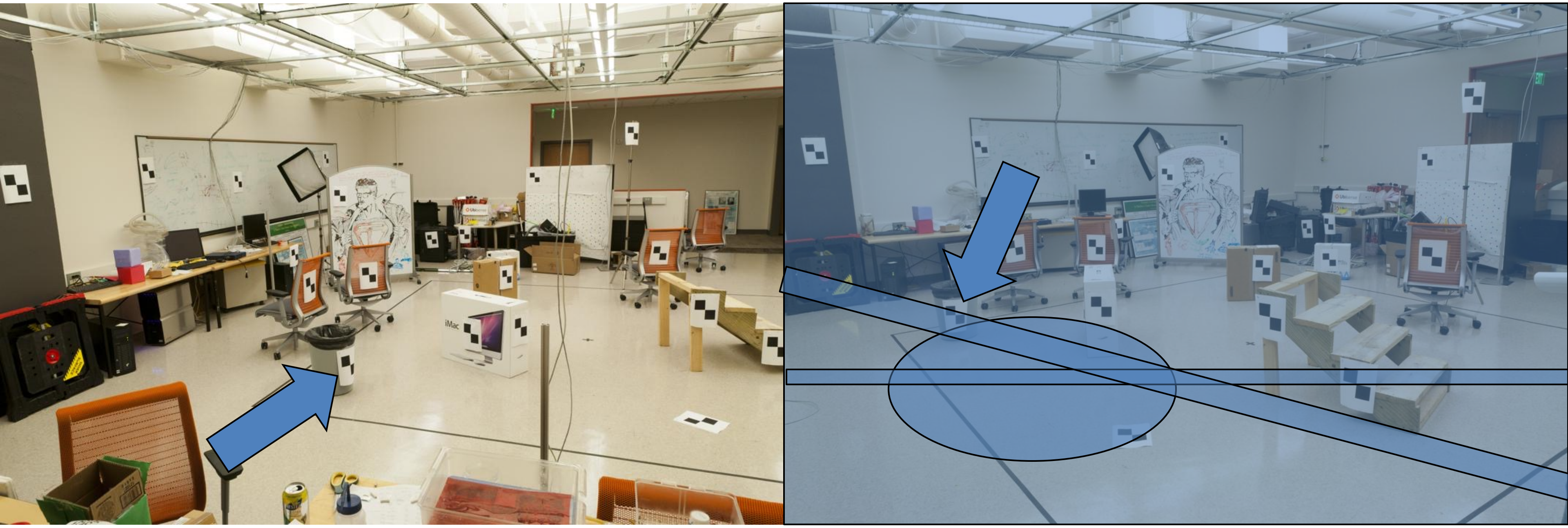
- Given p in left image, where can corresponding point p' be?

Stereo correspondence constraints



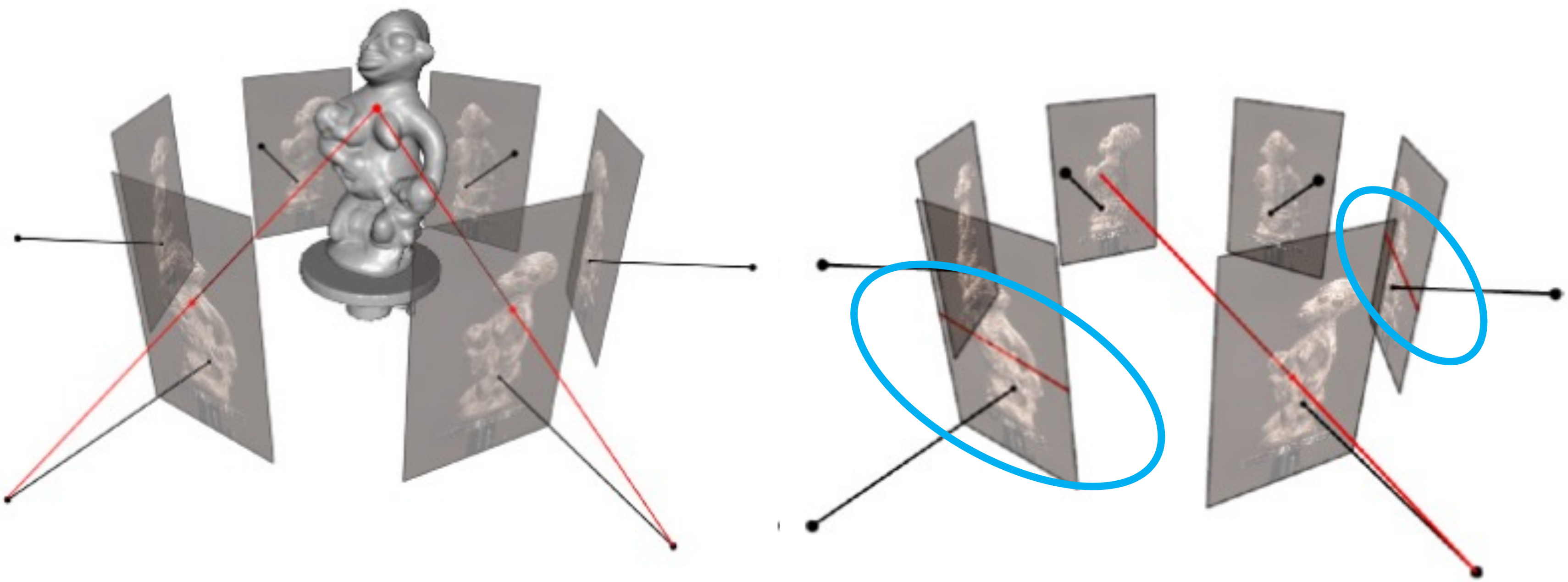
- Given p in left image, where can corresponding point p' be?

Where do we need to search?



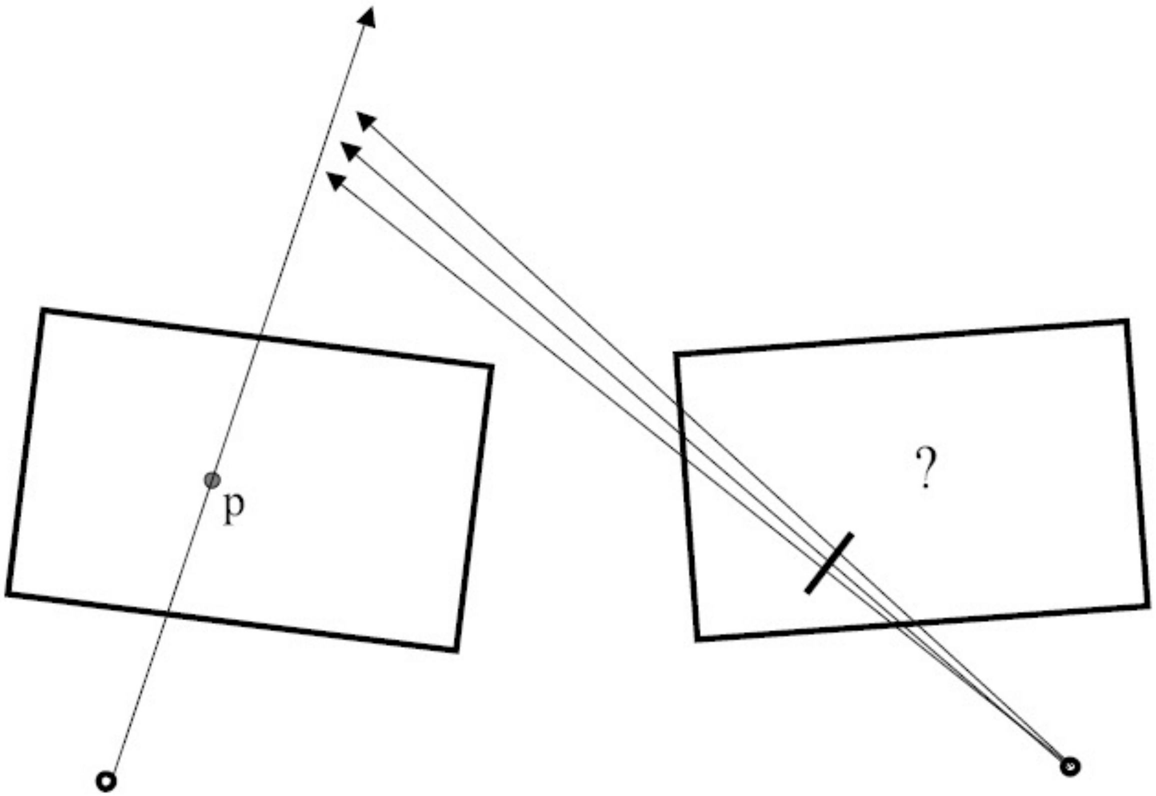
Slide by James Hays

Epipolar Geometry

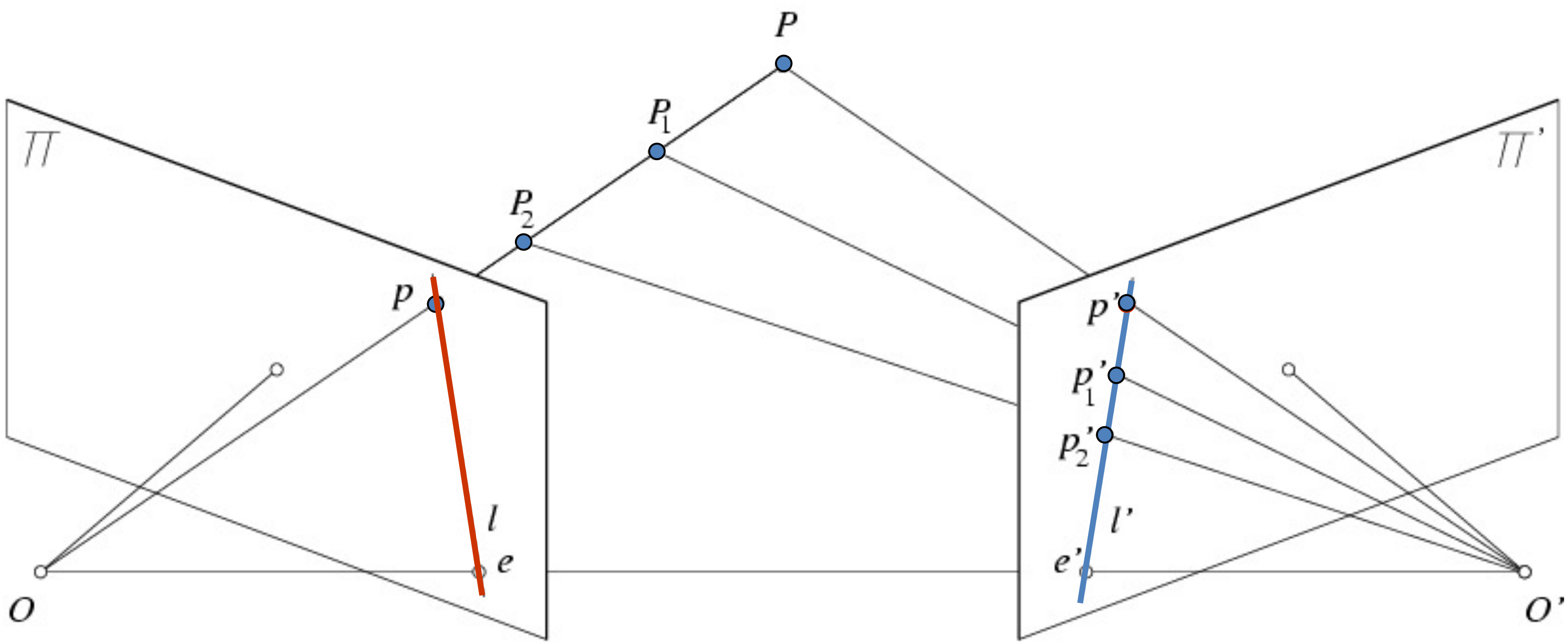


Figures by Carlos Hernandez

Stereo correspondence constraints



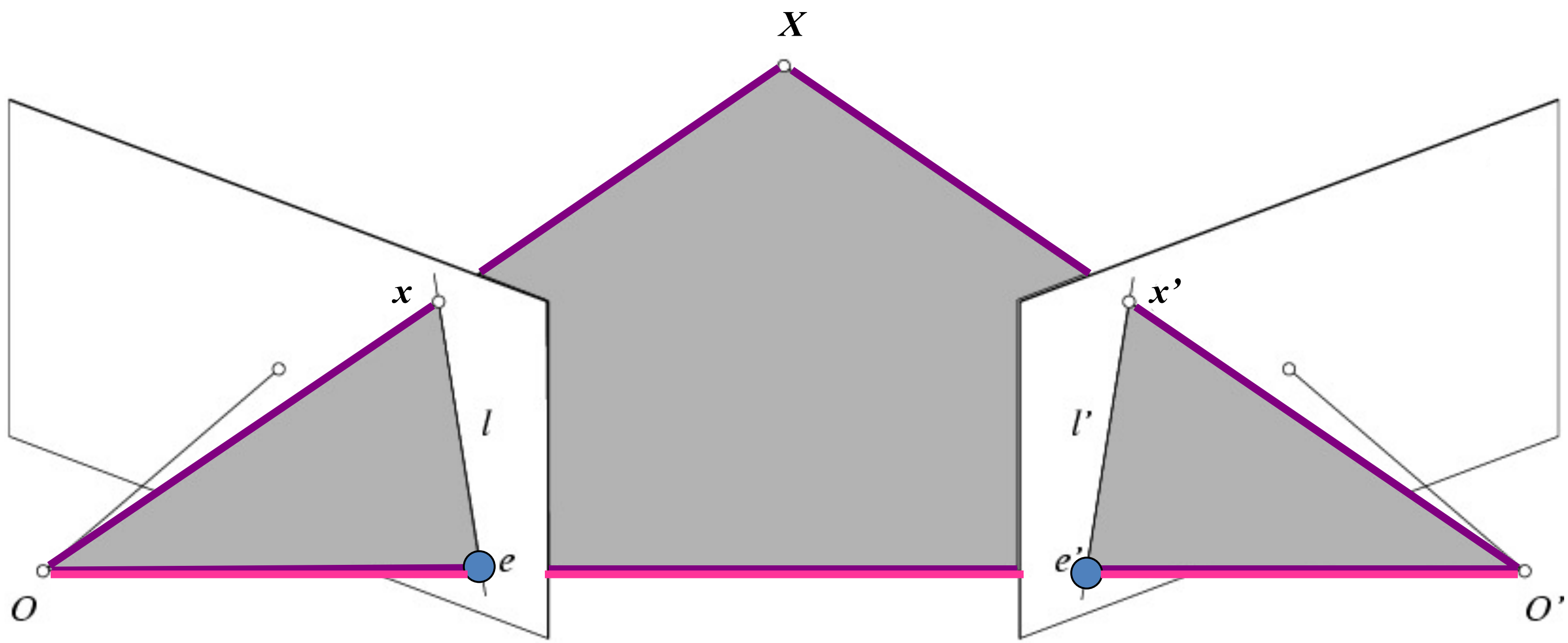
Epipolar constraint



- Potential matches for *p* have to lie on the corresponding epipolar line *l'*.
- Potential matches for *p'* have to lie on the corresponding epipolar line *l*.

Source: M. Pollefeys

Epipolar geometry



- **Baseline** – line connecting the two camera centers
- **Epipolar Plane** – plane containing baseline (1D family)
- **Epipoles**

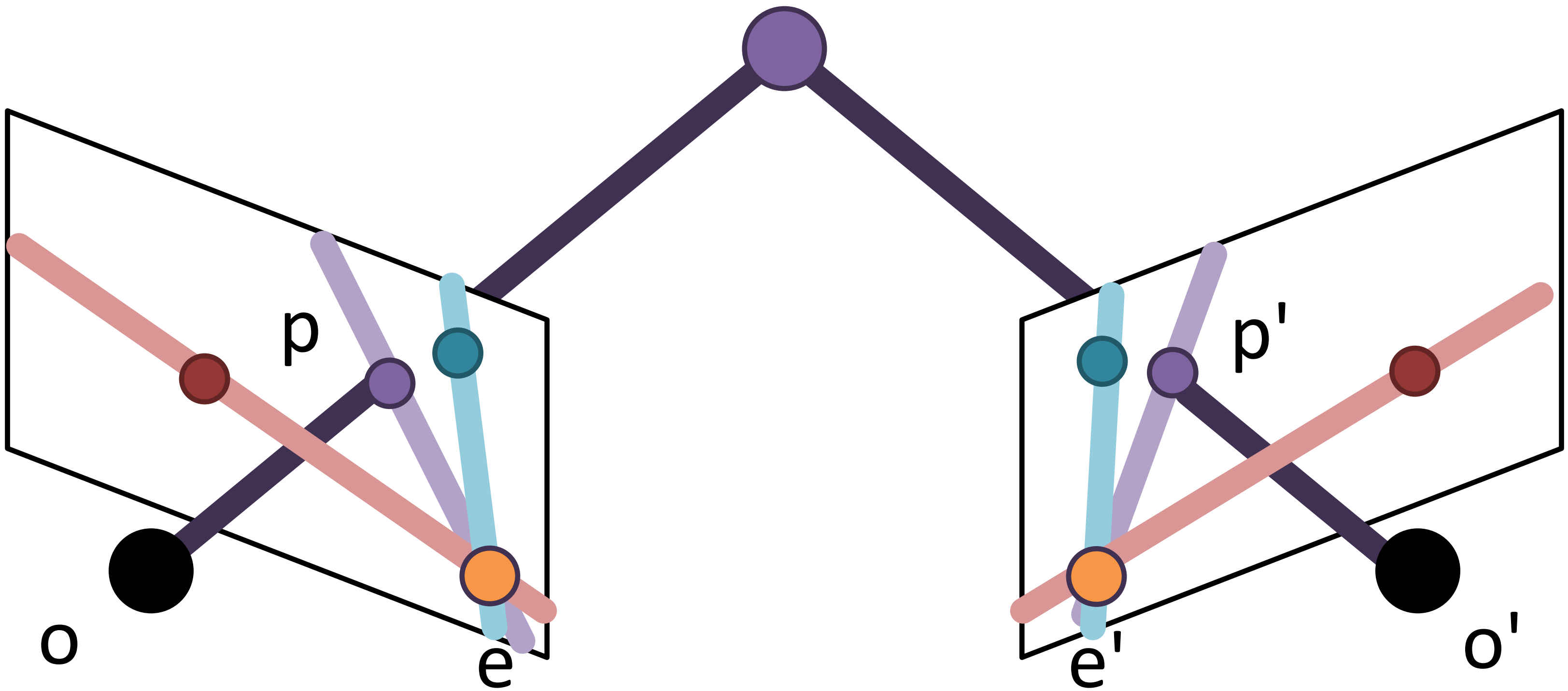
= intersections of baseline with image planes
= projections of the other camera center

Slide credit: David Fouhey

The Epipole

Photo by Frank Dellaert

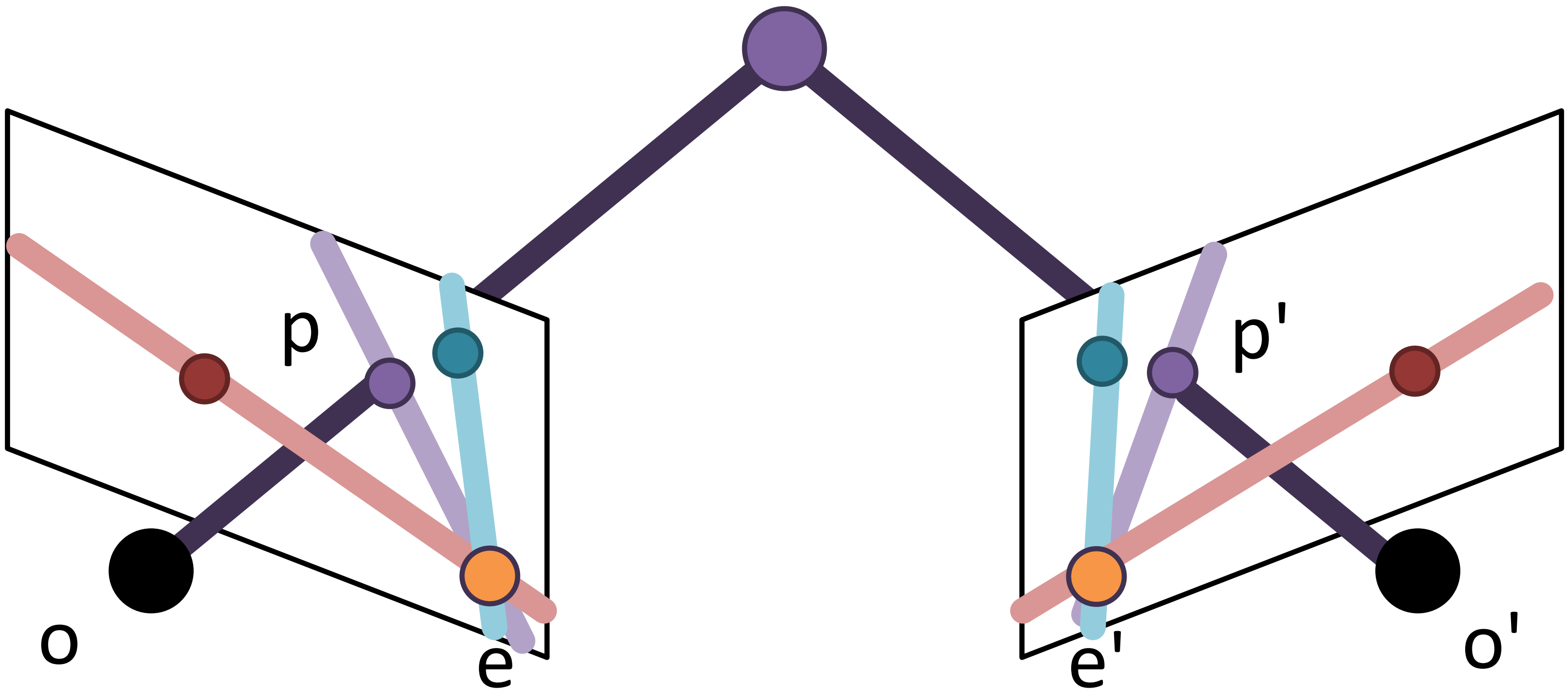
Example: Converging Cameras



Epipoles finite, maybe in image; epipolar lines converge

Slide credit: David Fouhey

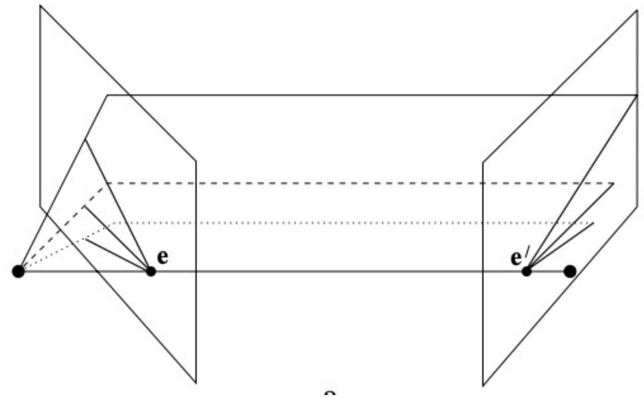
Example: Converging Cameras



Epipolar lines come in pairs:
given a point p, we can construct the epipolar line for p’.

Slide credit: David Fouhey

Example 1: Converging Cameras



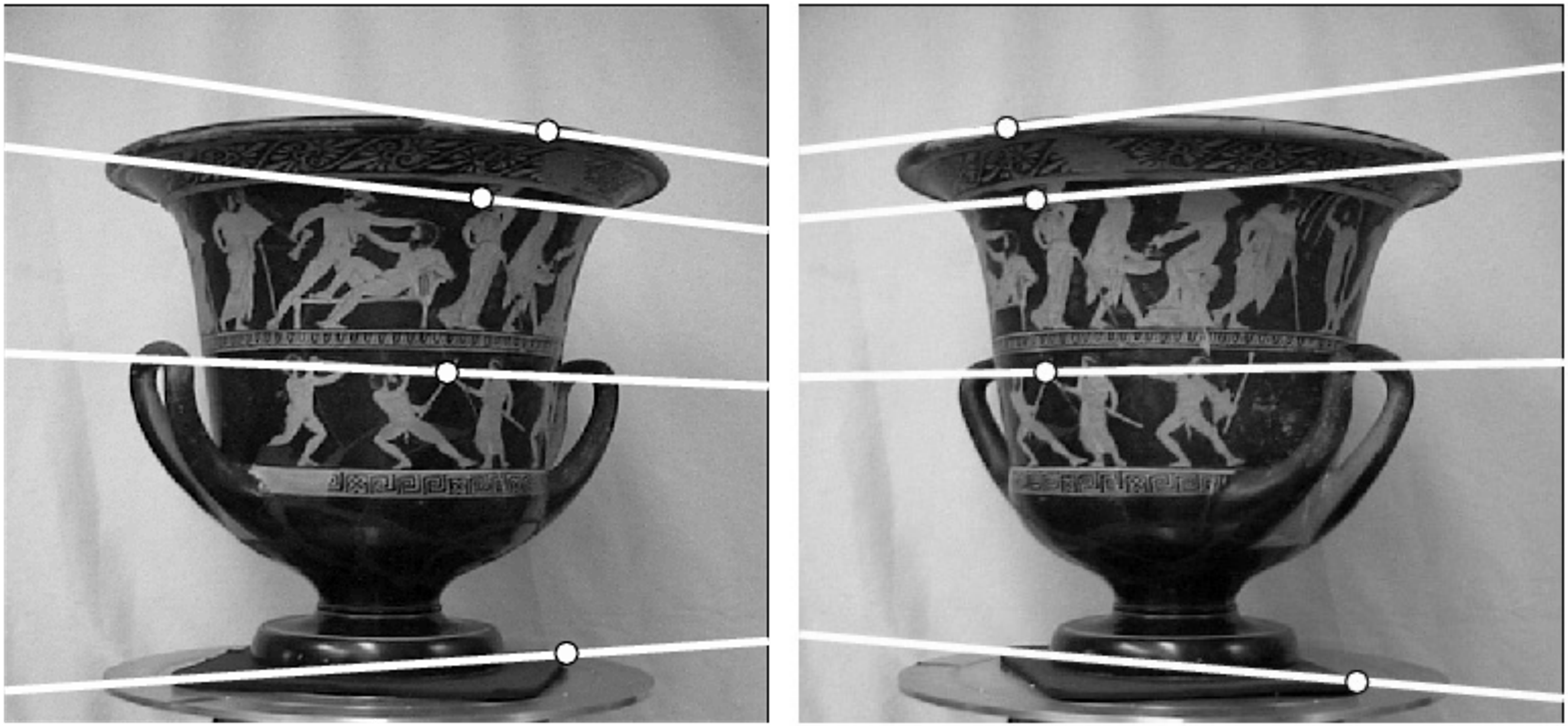
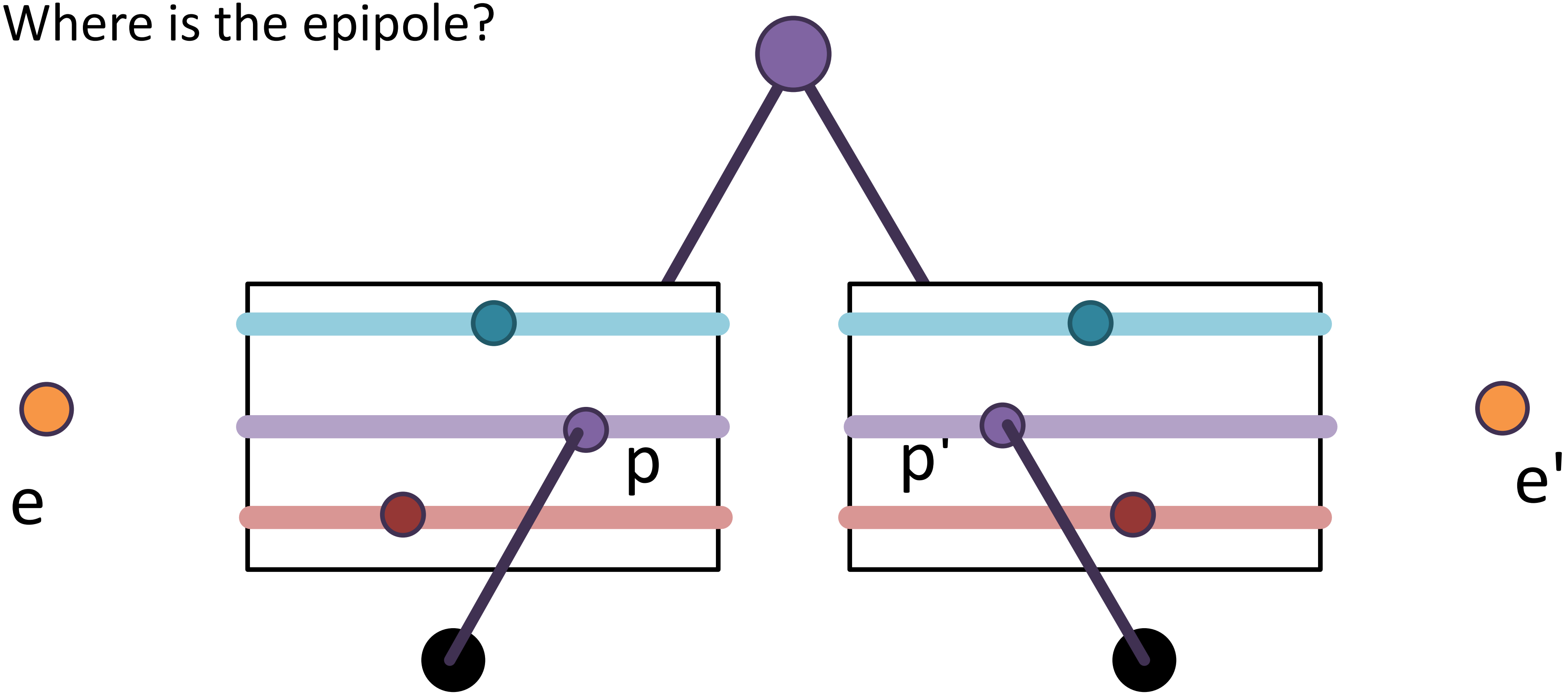


Image Credit: Hartley & Zisserman

Example: Parallel to Image Plane

Where is the epipole?



Epipoles *infinitely* far away, epipolar lines parallel

Slide credit: David Fouhey

Example: Forward Motion

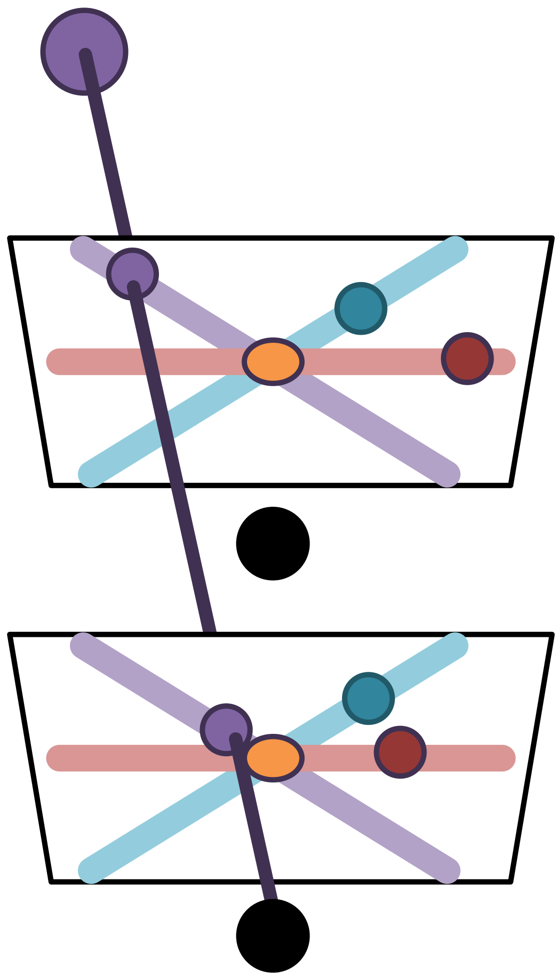
Image Credit: Hartley & Zisserman

Example: Forward Motion

Image Credit: Hartley & Zisserman

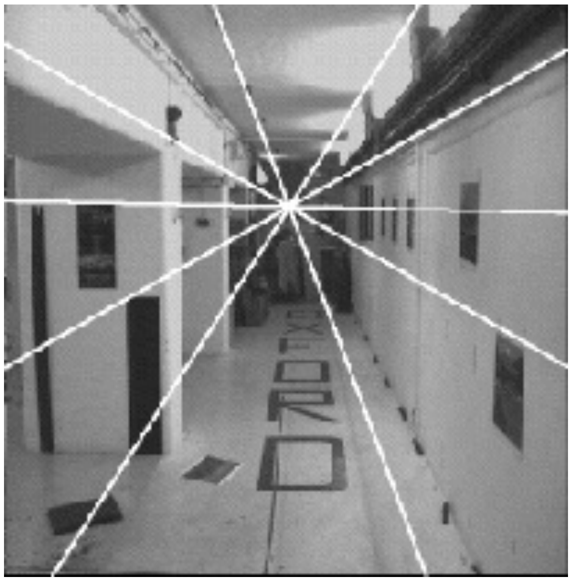
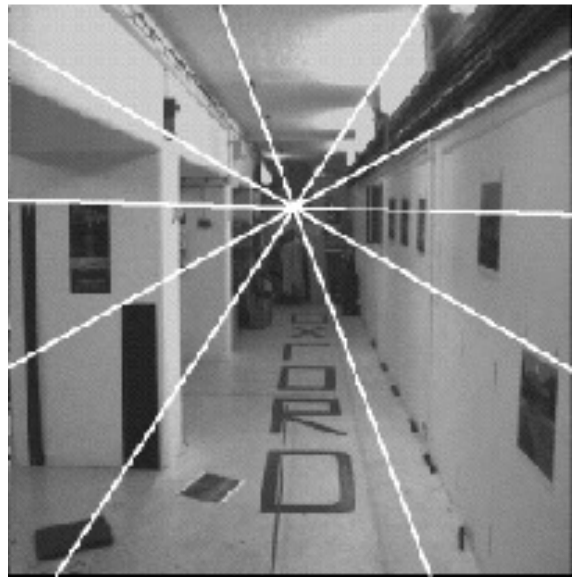
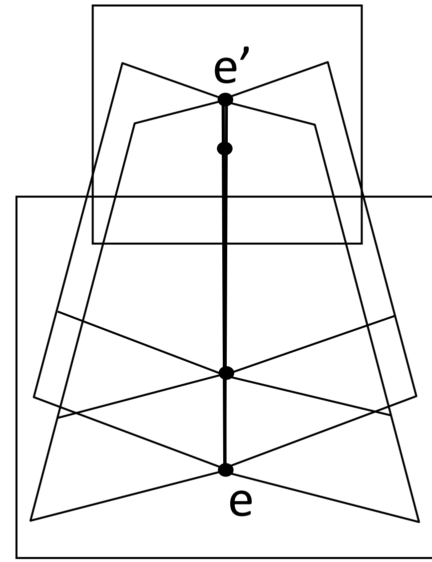
Example: Forward Motion

Epipole is focus of expansion / principal point of the camera.



Epipolar lines go out from principal point

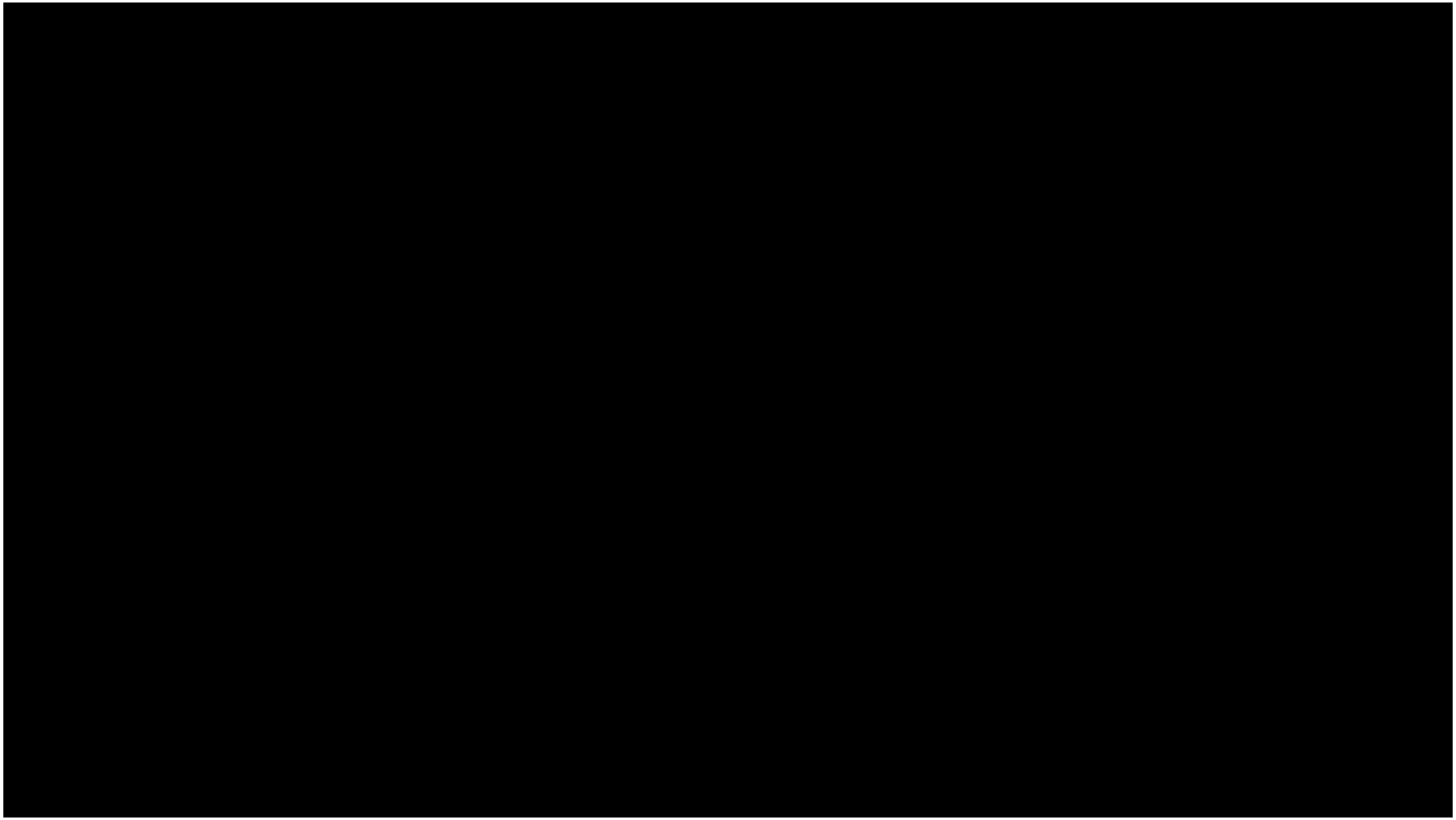
Example: forward motion

Epipole has same coordinates in both images.
Points move along lines radiating from e: “Focus of expansion”

Figure from Hartley & Zisserman

Motion perpendicular to image plane



http://vimeo.com/48425421

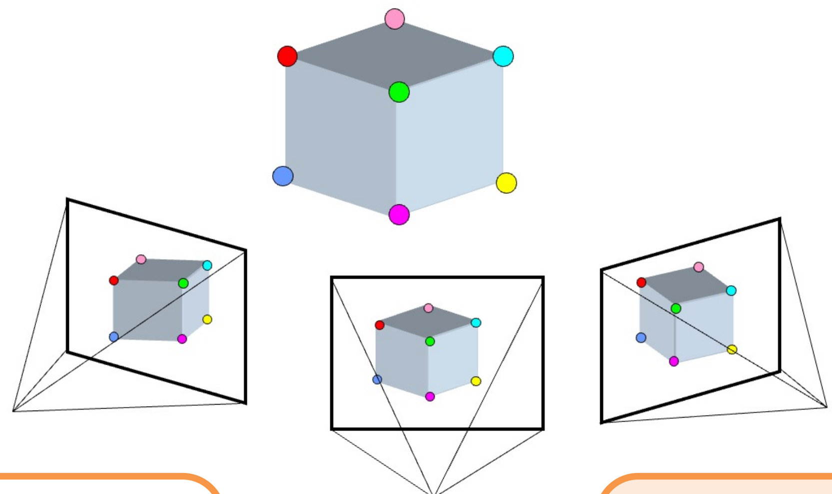
Slide credit: David Fouhey

Where were we?

- Why is this relevant?
- Assume camera is calibrated
- Goal: 3D reconstruction of corresponding points in the image

Big picture: 3 key components in 3D

3D Points
(Structure)

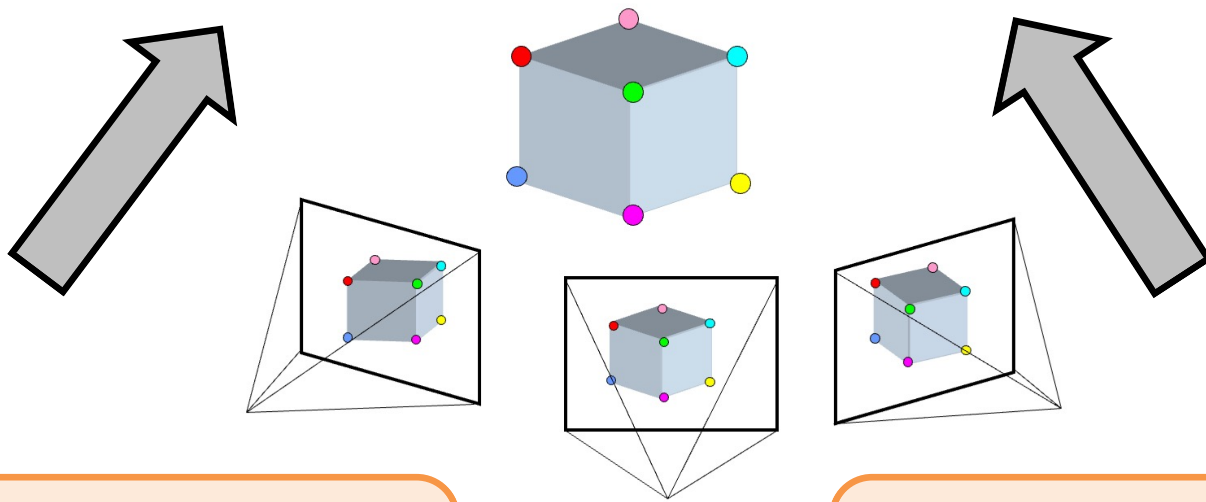


Correspondences

Camera
(Motion)

Big picture: 3 key components in 3D

3D Points
(Structure)



Correspondences

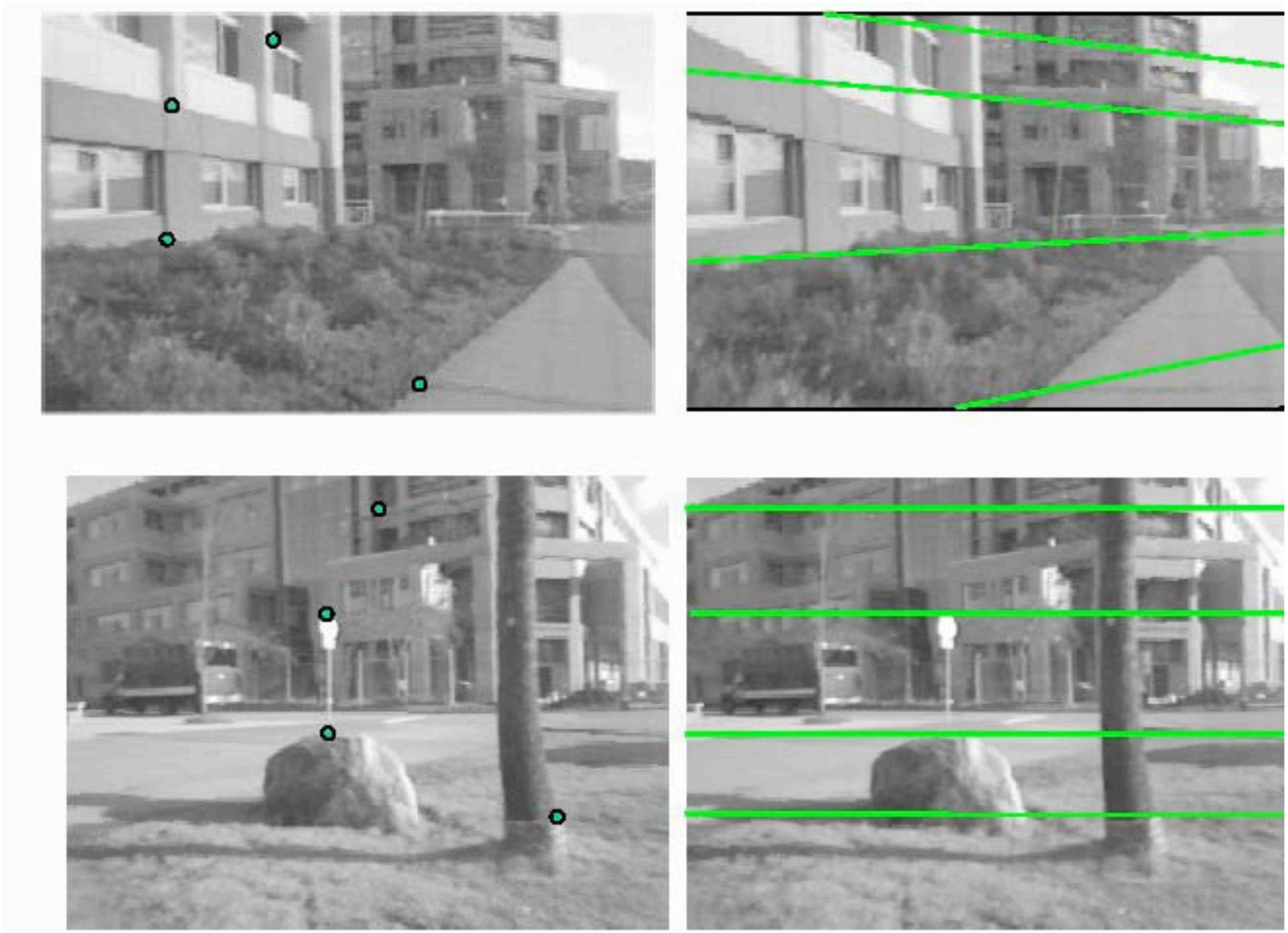
Camera
(Motion)

Epipolar Geometry

- If I want to do stereo, I want to find a corresponding pixel for each pixel in the image:
- Naïve search:
  - For each pixel, search every other pixel
- With epipolar geometry:
  - For each pixel, search along each line (1D search)!

Slide credit: David Fouhey

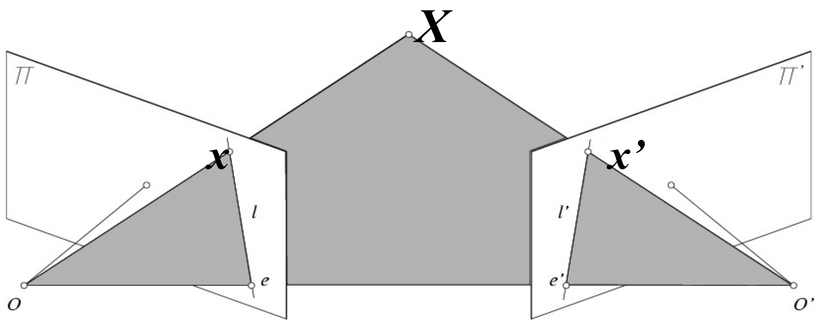
Epipolar constraint example



Slide Credit: S. Lazebnik

How do we compute the Epipolar line?

Step 0: Normalized image coordinates



$$x = K[R \;\; t]X$$

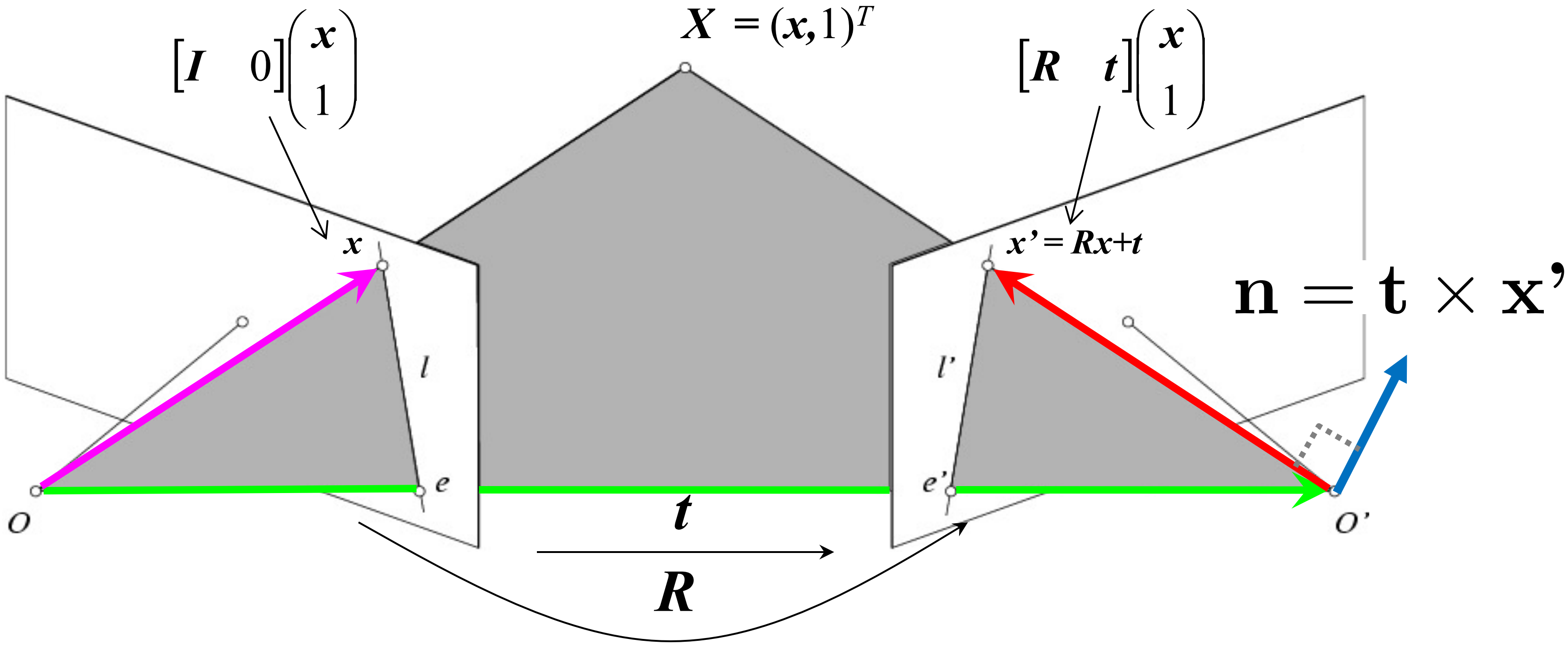
$$K^{-1}x = [R \;\; t]X$$

- Let's factor out the effect of K
- Since we know the intrinsics K, apply its inverse to x
- This is called the *normalized* image coordinates. It may be thought of as a set of points with K = Identity

$$\boldsymbol{x}_{\text{norm}} = \boldsymbol{K}^{-1}\boldsymbol{x}_{\text{pixel}} = [\boldsymbol{I} \;\; 0]\boldsymbol{X}, \qquad \boldsymbol{x}'_{\text{norm}} = \boldsymbol{K}'^{-1}\boldsymbol{x}'_{\text{pixel}} = [\boldsymbol{R} \;\; \boldsymbol{t}]\boldsymbol{X}$$

- Assume that the points are normalized from here on

Epipolar constraint: Calibrated case



The vectors x, t, and x' are coplanar

What can you say about their relationships, given $\mathbf{n} = \mathbf{t} \times \mathbf{x'}$?

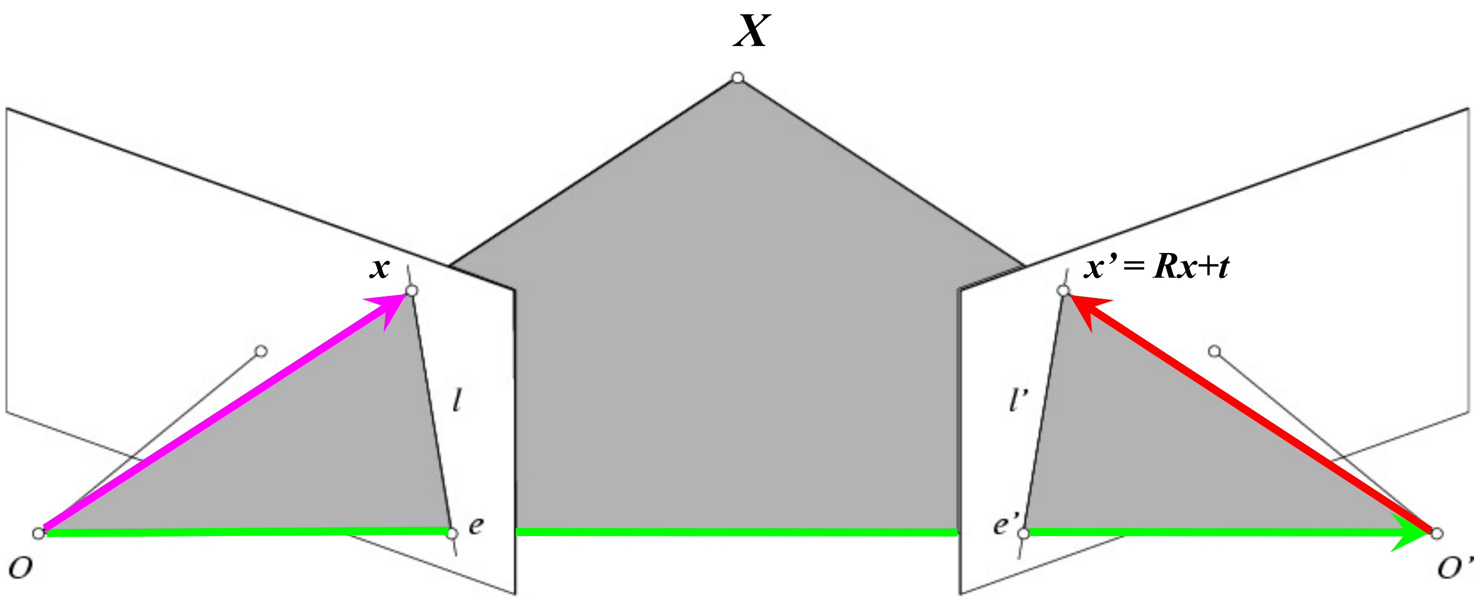
$$\boxed{\mathbf{x'} \cdot (\mathbf{t} \times \mathbf{x'}) = 0}$$

$$\mathbf{x'} \cdot (\mathbf{t} \times (R\mathbf{x} + \mathbf{t})) = 0$$

$$\mathbf{x'} \cdot (\mathbf{t} \times R\mathbf{x} + \underbrace{\cancel{\mathbf{t} \times \mathbf{t}}}_{0})) = 0$$

$$\boxed{\mathbf{x'} \cdot (\mathbf{t} \times R\mathbf{x}) = 0}$$

Epipolar constraint: Calibrated case

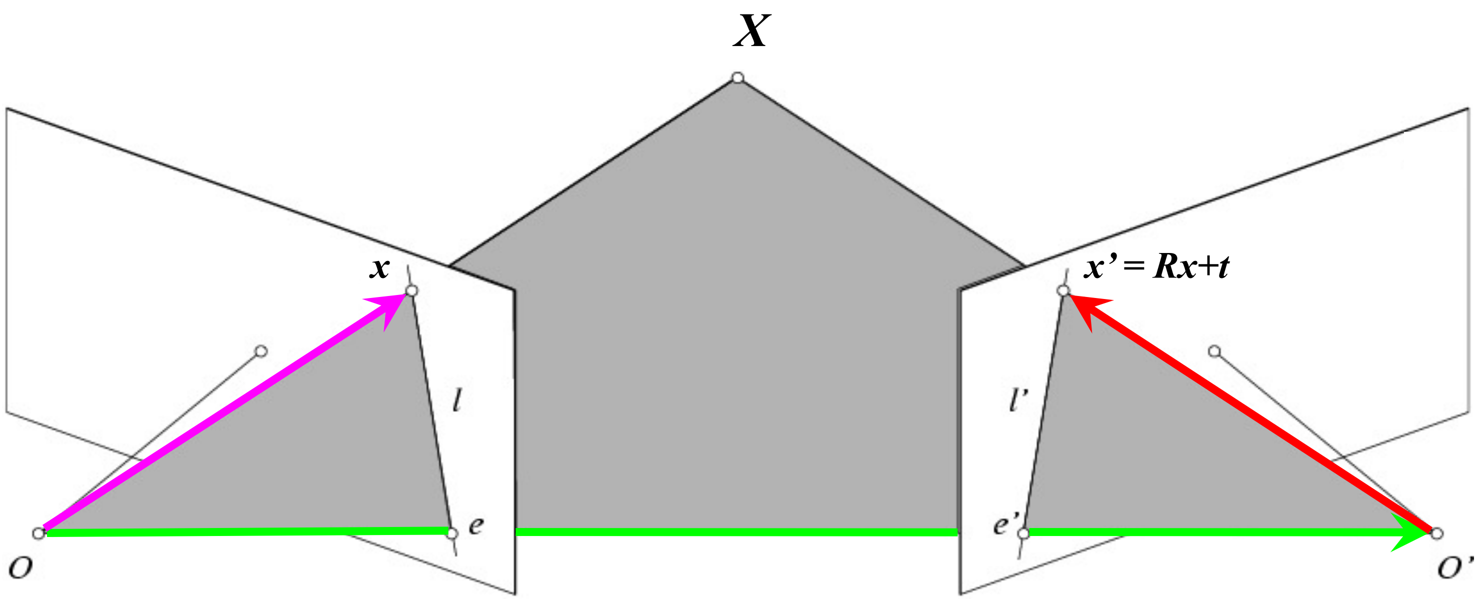


$$\boldsymbol{x}' \cdot [\boldsymbol{t} \times (\boldsymbol{R}\boldsymbol{x})] = 0 \quad \Rightarrow \quad \boldsymbol{x}'^T [\boldsymbol{t}_\times] \boldsymbol{R}\boldsymbol{x} = 0$$

Recall: $\mathbf{a} \times \mathbf{b} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = [\mathbf{a}_\times]\mathbf{b}$

The vectors ***Rx***, ***t***, and ***x'*** are coplanar

Epipolar constraint: Calibrated case



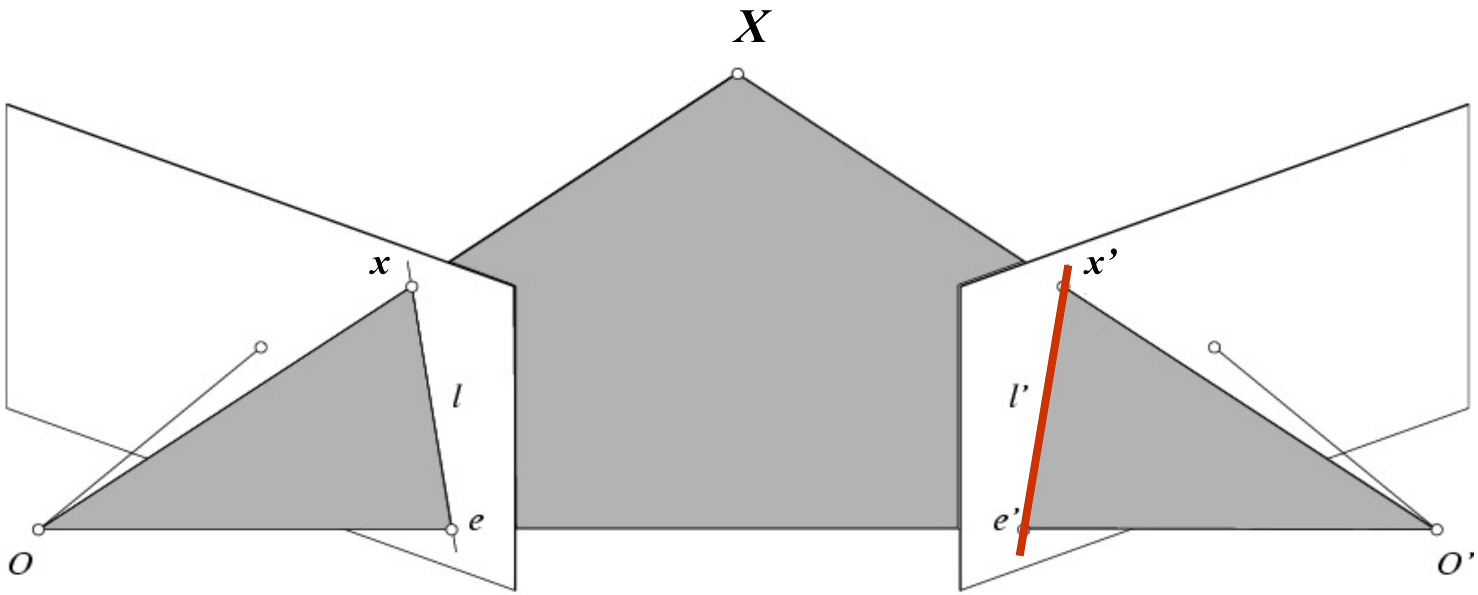
$$\boldsymbol{x}' \cdot [\boldsymbol{t} \times (\boldsymbol{R}\boldsymbol{x})] = 0 \quad \Rightarrow \quad \boldsymbol{x}'^{T} \underbrace{[\boldsymbol{t}_{\times}]\boldsymbol{R}}_{E} \boldsymbol{x} = 0 \quad \Rightarrow \quad \boldsymbol{x}'^{T} E \boldsymbol{x} = 0$$

Recall: $\mathbf{a} \times \mathbf{b} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = [\mathbf{a}_{\times}]\mathbf{b}$

Essential Matrix
(Longuet-Higgins, 1981)

The vectors $\boldsymbol{x}$, $\boldsymbol{t}$, and $\boldsymbol{x}'$ are coplanar

Epipolar constraint: Calibrated case

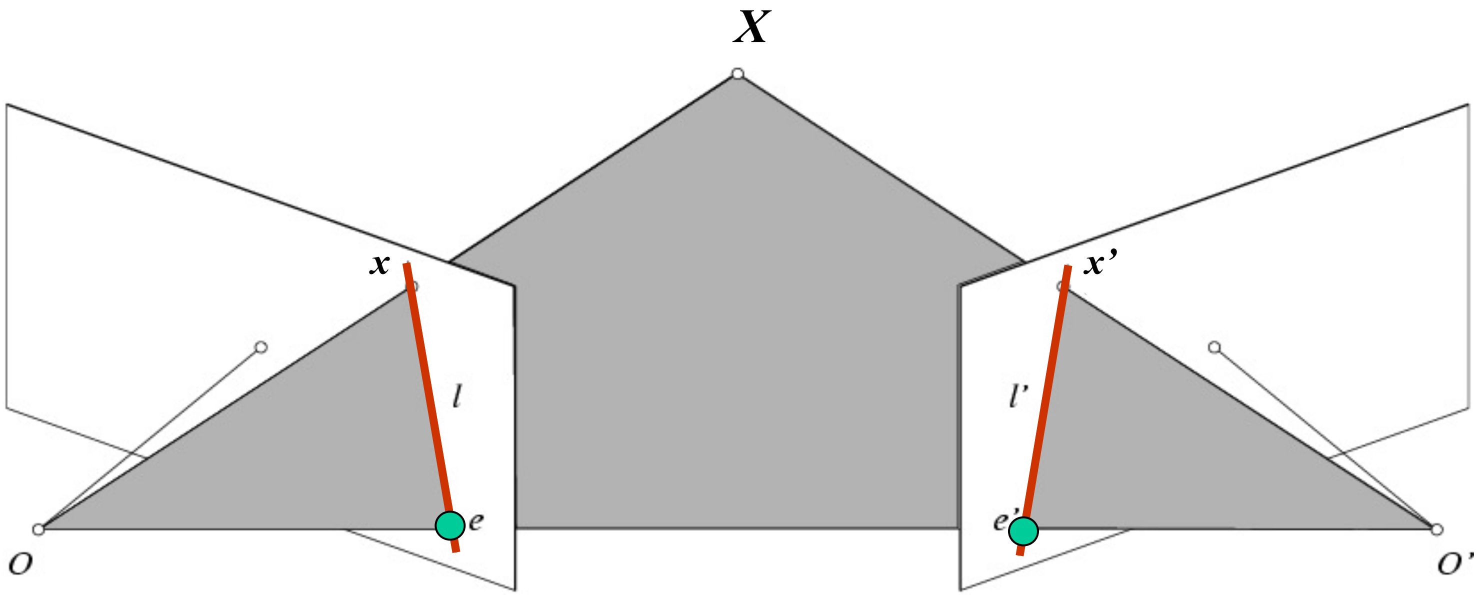


$$\boxed{\boldsymbol{x}'^{T} E \boldsymbol{x} = 0}$$

- ***E x*** is the epipolar line associated with ***x*** (***l'*** = ***E x***)
  - Recall: a line is given by $ax + by + c = 0$ or

$$\mathbf{l}^T \mathbf{x} = 0 \quad \text{where} \quad \mathbf{l} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Epipolar constraint: Calibrated case



$$\boxed{\boldsymbol{x}'^T E \boldsymbol{x} = 0}$$

- ***E x*** is the epipolar line associated with ***x*** (***l'*** = ***E x***)
- ***E***^{T}***x'*** is the epipolar line associated with ***x'*** (***l*** = ***E***^{T}***x'***)
- ***E*** **e** = 0 and ***E***^{T}**e'** = *0*
- ***E*** is singular (rank two)
- ***E*** has five degrees of freedom

Epipolar constraint: Uncalibrated case

- Recall that we normalized the coordinates

$$x = K^{-1}\hat{x} \quad x' = K'^{-1}\hat{x}'$$

$$\hat{x} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

where $\hat{x}$ is the image coordinates

- But in the *uncalibrated* case, ***K*** and ***K'*** are unknown!
- We can write the epipolar constraint in terms of *unknown* normalized coordinates:

$$x'^T E x = 0$$

$$(K'^{-1}\hat{x}')'^T E (K^{-1}\hat{x}) = 0$$

$$\hat{x}'^T \underbrace{K'^{-T} E K^{-1}}_{F} \hat{x} = 0$$

$$\hat{x}'^T F \hat{x} = 0$$

$$F = K'^{-T} E K^{-1}$$

Fundamental Matrix
(Faugeras and Luong, 1992)

Essential vs Fundamental matrix

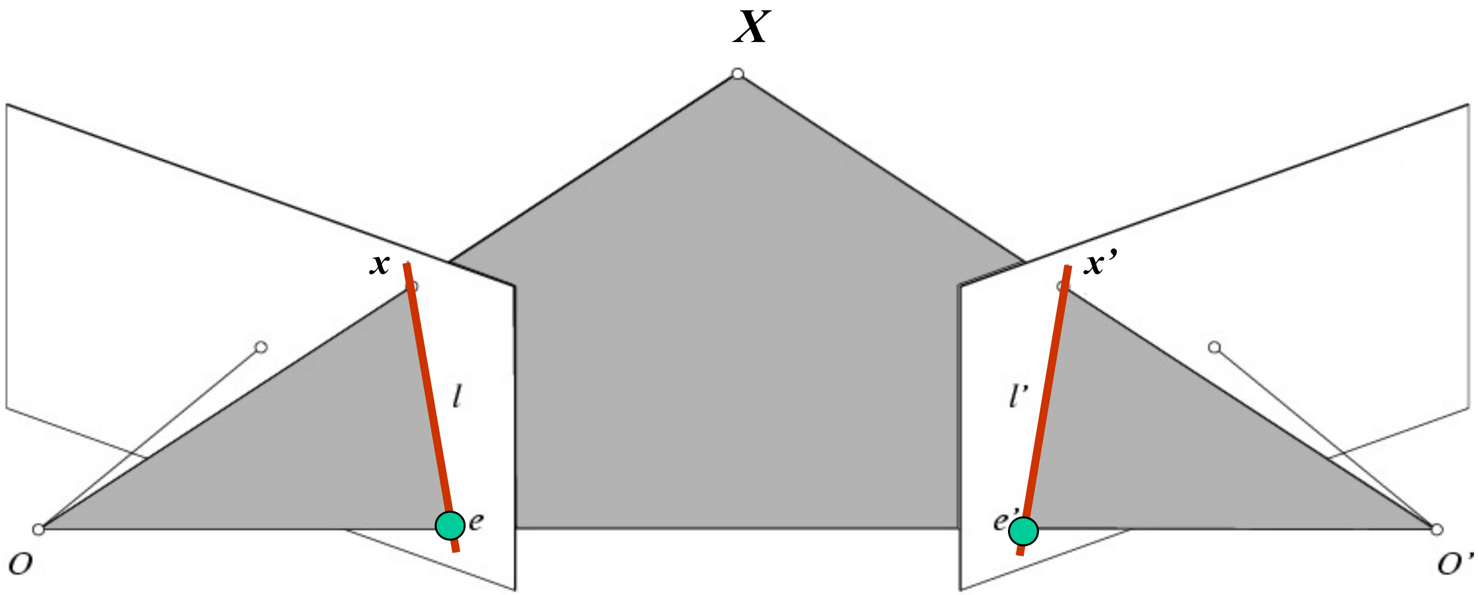
What is the difference??

$$x'^T E x = 0$$

$$\hat{x}'^T F \hat{x} = 0$$

$$F = K'^{-T} E K^{-1}$$

Epipolar constraint: Uncalibrated case



$x'^T E x = 0$ ⇒ $\hat{x}'^T F \hat{x} = 0$ with $F = K'^{-T} E K^{-1}$

- $\boldsymbol{F}\hat{\boldsymbol{x}}$ is the epipolar line associated with $\hat{\boldsymbol{x}}$ ($\boldsymbol{l}' = \boldsymbol{F}\hat{\boldsymbol{x}}$)
- $\boldsymbol{F}^T\hat{\boldsymbol{x}}'$ is the epipolar line associated with $\hat{\boldsymbol{x}}'$ ($\boldsymbol{l} = \boldsymbol{F}^T\hat{\boldsymbol{x}}'$)
- $\boldsymbol{F}\mathbf{e} = 0$ and $\boldsymbol{F}^T\mathbf{e}' = 0$
- ***F*** is singular (rank two)
- ***F*** has *seven* degrees of freedom

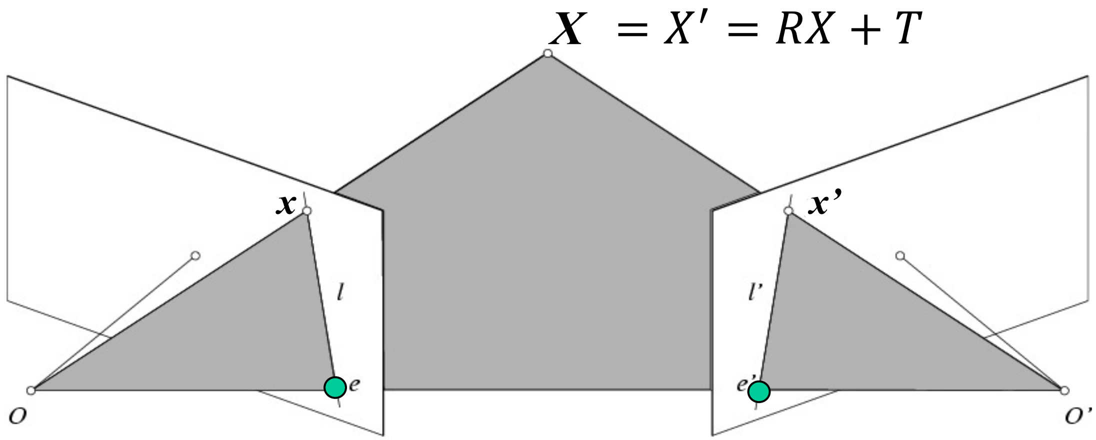
Where are we?

Recall we’re trying to get the 3D points of corresponding images, with calibrated cameras (known K and R, T)

1. Solve for correspondences using epipolar constraints from known camera (1D search)
2. Triangulate to get depth!

Finally: computing depth by triangulation

We know about the camera, K_1, K_2 and [R t]:



and that these are corresponding points: $x \leftrightarrow x'$

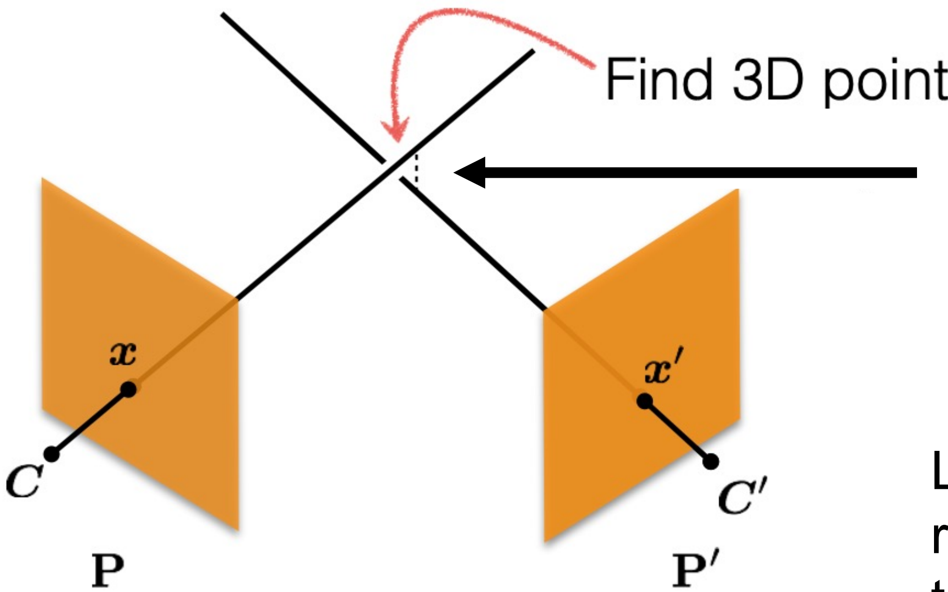
$$x = K\boxed{X} \qquad x' = K'X'$$
$$= K'(R\boxed{X} + T)$$

only unknowns!

How many unknowns
+ how many equations
do we have?

Solve by least squares

Triangulation Disclaimer: Noise



Ray's don't always intersect because of noise!!!

$$\mathbf{X} \text{ s.t.}$$

$$\mathbf{x} = \mathbf{P}\mathbf{X}, \;\; \mathbf{x}' = \mathbf{P}'\mathbf{X}$$

Least squares get you to a reasonable solution but it's not the actual geometric error (it's how far away the solution is from Ax = 0)

In practice with noise, you do non-linear least squares, or "bundle adjustment" (more than 2 image case, next lecture..)

Slide credit: Shubham Tulsiani

Summary: Two-view, known camera

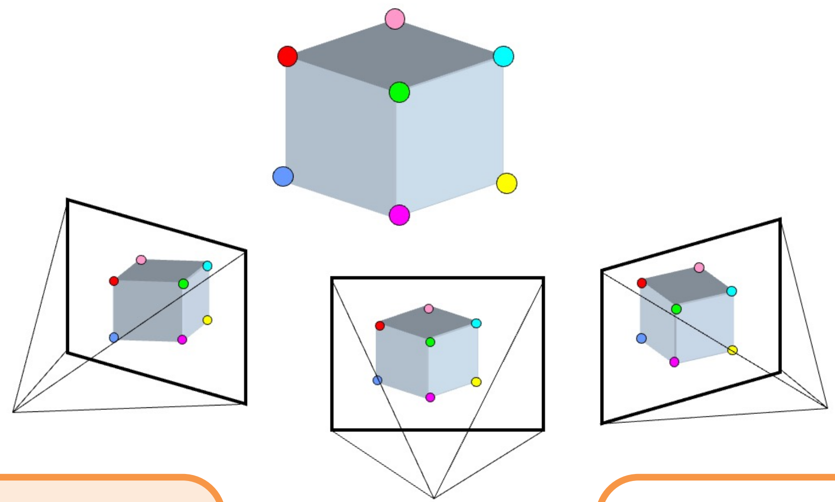
0. Calibrate the camera.

1. Find correspondences:
   - Reduce this to 1D search with Epipolar Geometry!

2. Get depth:
   - If simple stereo, disparity (difference of corresponding points) is inversely proportional to depth
   - In the general case, triangulate.

What if we don't know the camera?

3D Points
(Structure)



Correspondences → Camera
(Motion)

What if we don't know the camera?

Assume we know the correspondences:

$\hat{x}'$ and $\hat{x}$ in the image

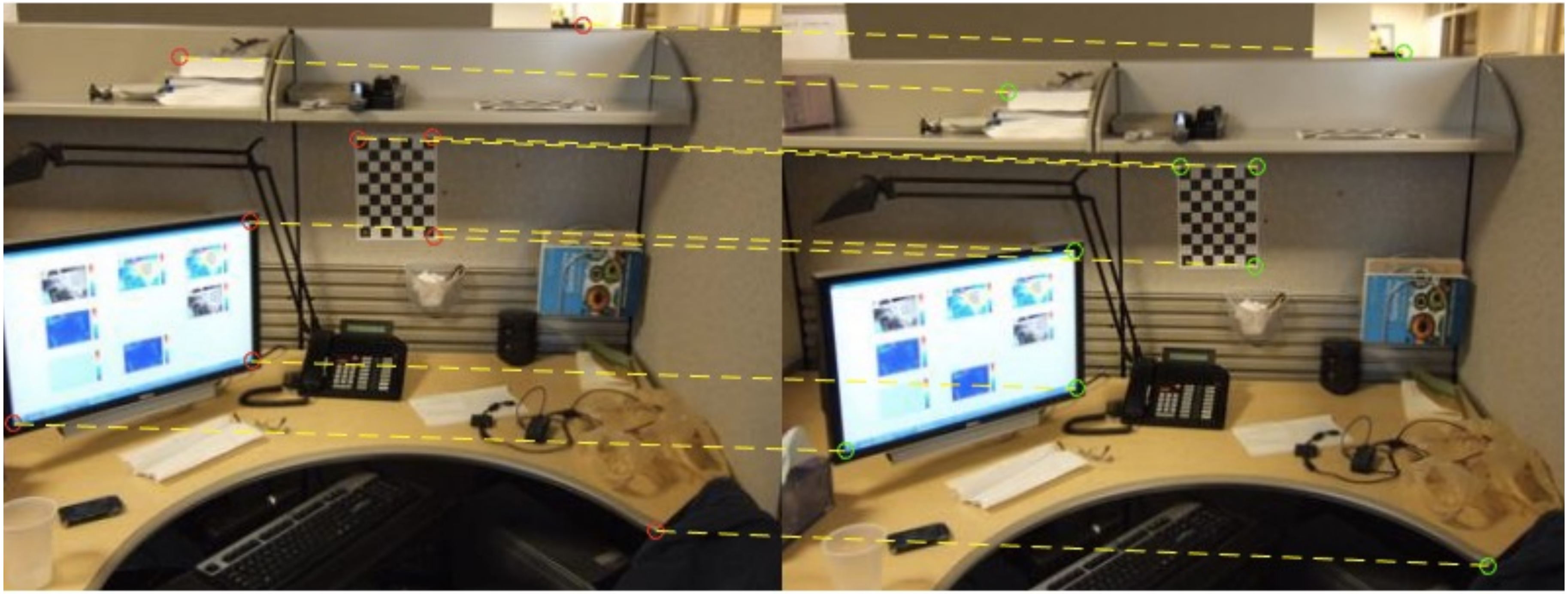
$$\hat{x}'^T F \hat{x} = 0$$

$$\hat{x} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} u' & v' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

How many correspondences do we need?

Estimating the fundamental matrix



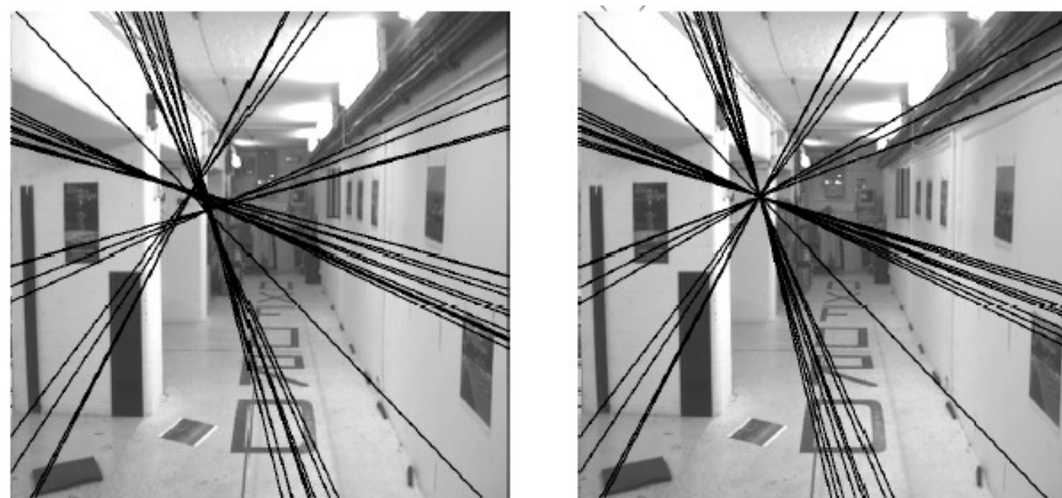
The eight-point algorithm

$$\boldsymbol{x} = (u, v, 1)^T, \quad \boldsymbol{x}' = (u', v', 1)$$

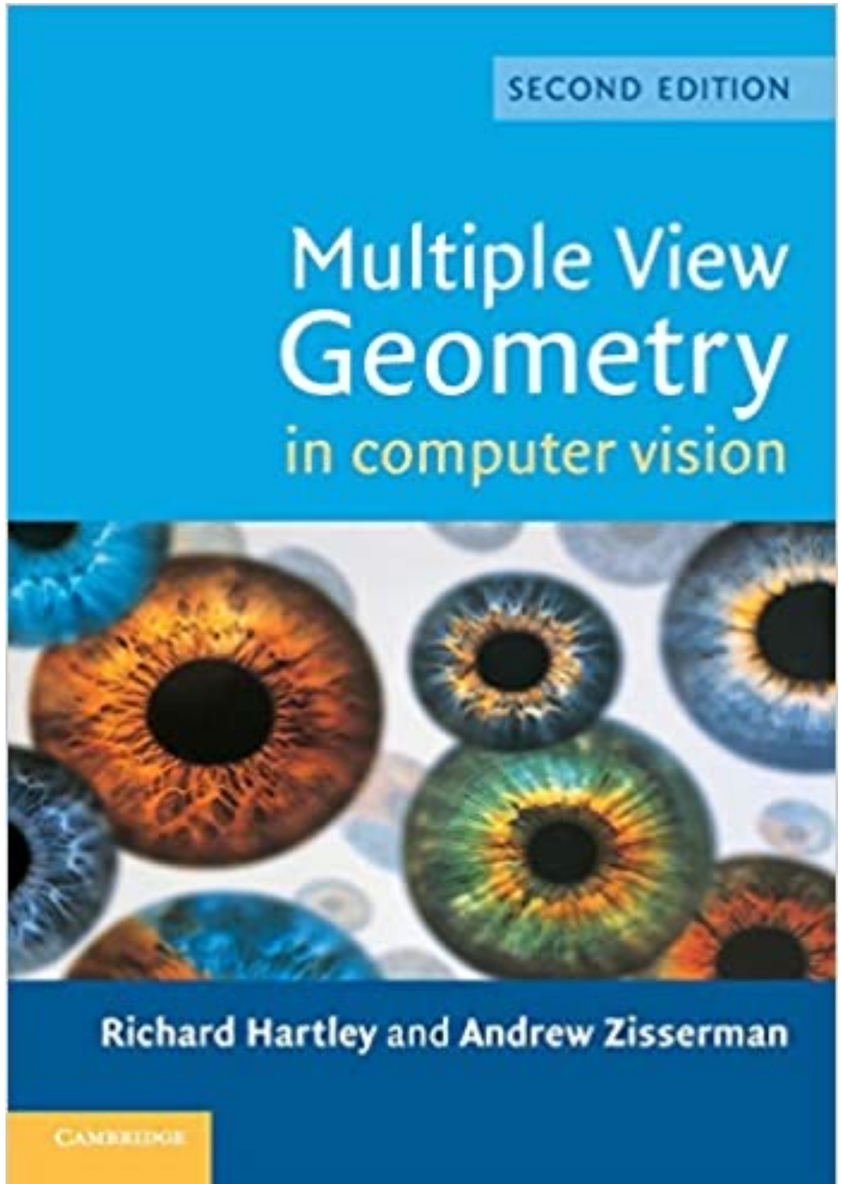
$$\begin{bmatrix} u' & v' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0 \quad \Rightarrow \quad \begin{bmatrix} u'u & u'v & u' & v'u & v'v & v' & u & v & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

Solve homogeneous linear system using eight or more matches

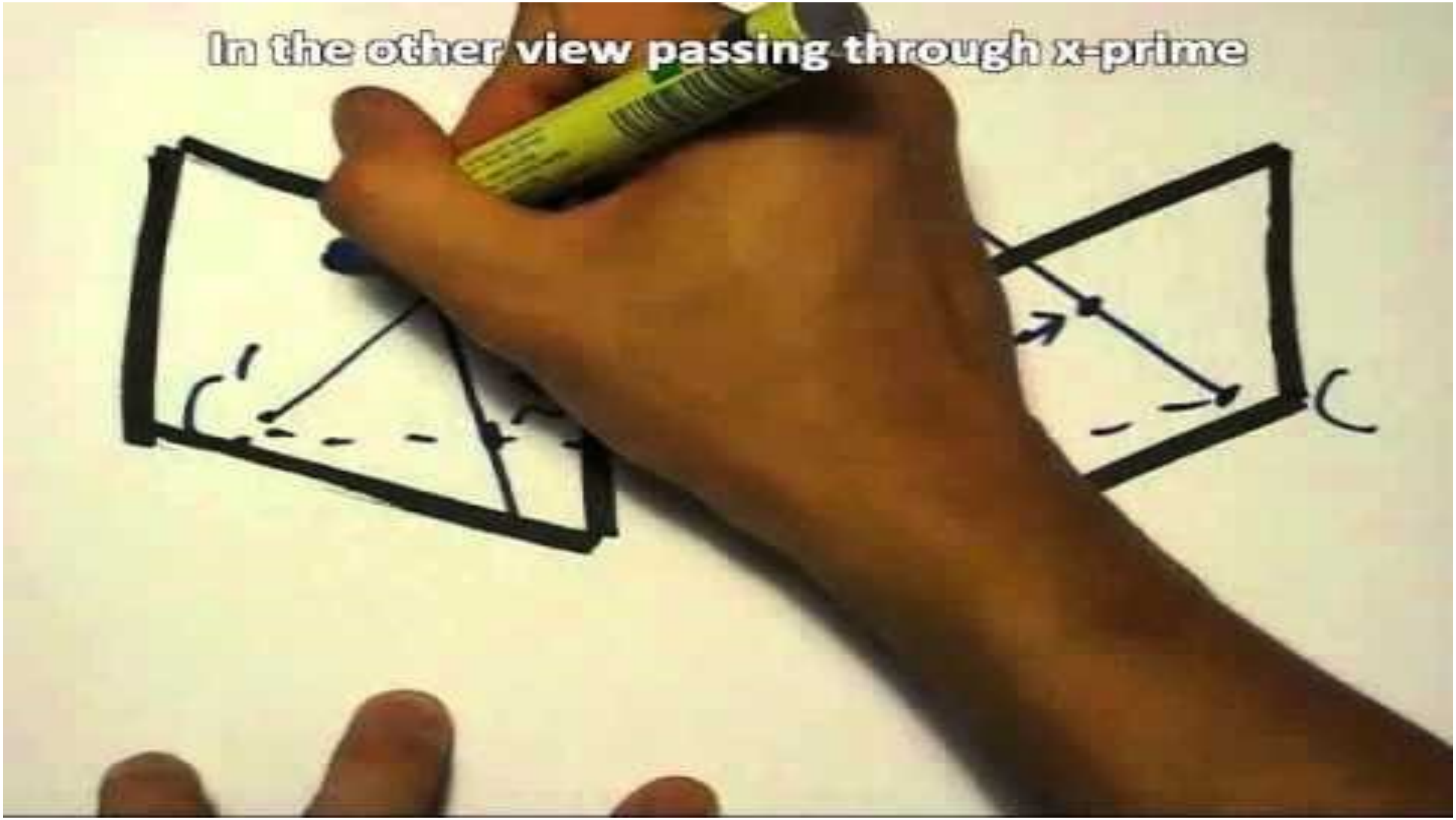
Enforce rank-2 constraint (take SVD of ***F*** and throw out the smallest singular value)



The Bible by Hartley & Zisserman



The Fundamental Matrix Song



http://danielwedge.com/fmatrix/
https://www.youtube.com/watch?time_continue=8&v=DgGV3l82NTk&feature=emb_title

What about more than two views?

The geometry of three views is described by a 3 x 3 x 3 tensor called the *trifocal tensor*

The geometry of four views is described by a 3 x 3 x 3 x 3 tensor called the *quadrifocal tensor*

After this it starts to get complicated…

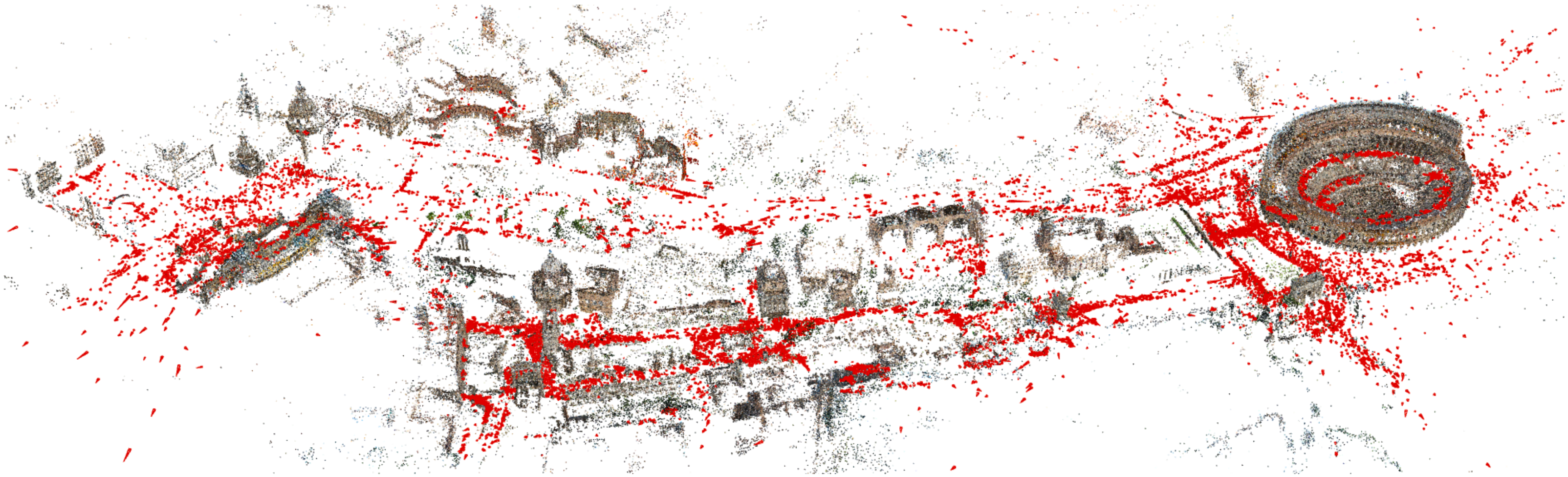
Next: Large-scale structure from motion

Dubrovnik, Croatia. 4,619 images (out of an initial 57,845).
Total reconstruction time: 23 hours
Number of cores: 352

Building Rome in a Day, Agarwal et al. ICCV 2009

Slide courtesy of Noah Snavely

Large-scale structure from motion



Result using COLMAP: Schönberger et al. CVPR '16