THE UNIVERSITY OF MELBOURNE

Multi-view depth

Semester 2, 2021

Kris Ehinger

https://www.youtube.com/watch?v=0Pj-jzy6ESE

Outline

- Multi-view problem
- Camera calibration
- Epipolar geometry
  - Basics
  - Math

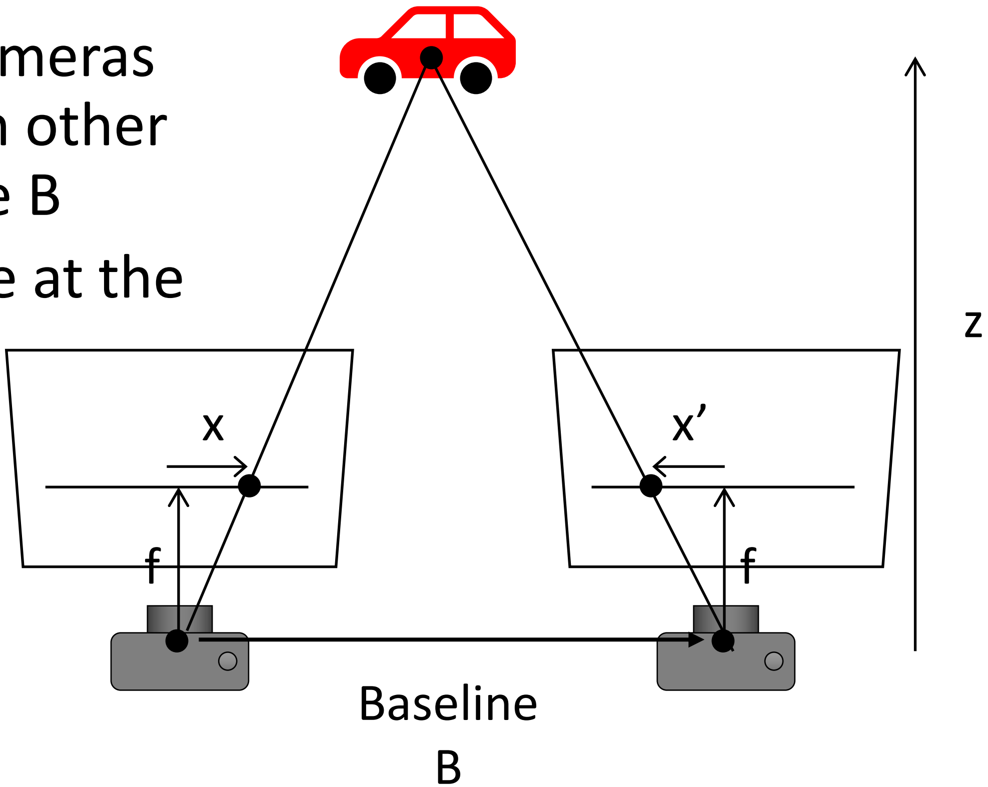
Learning outcomes

- Explain the unknowns that must be solved in a multi-view depth problem
- Explain two-view (epipolar geometry) and how it constrains the solution to this problem
- Define a “calibrated” camera and explain what is represented in a camera matrix

Multi-view problem

Standard stereo set-up

- Assume:
  - Image planes of cameras are parallel to each other and to the baseline B
  - Camera centres are at the same height
  - Focal lengths f are the same
- Goal: find z



Depth from multiple views

- What if you don't know the change in camera position between the views?
- What if the camera parameters (focal length, etc.) are unknown?

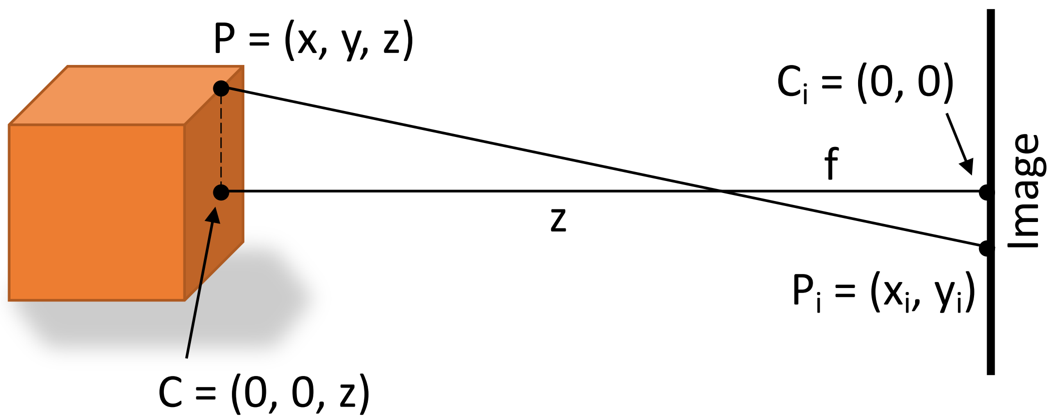
Multi-view problem

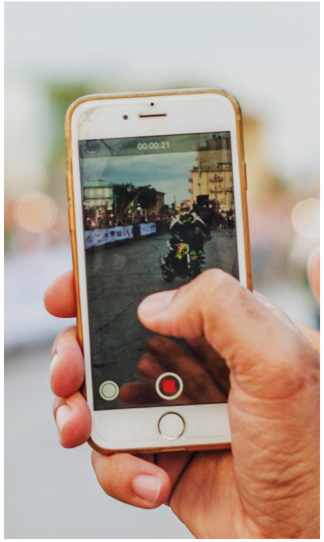
- Solve for:
  - Camera motion – what is the transform (camera translation + rotation) that relates the two views?
  - Camera parameters (e.g., focal length), if not known
  - Scene geometry – given corresponding image points (x, x’) in the two views, what is the position of the point X in 3D space?

Camera calibration

Camera calibration

- We know how to compute 3D (x,y,z) from image plane (x,y) when imaging parameters are known
- Usually, these parameters are unknown



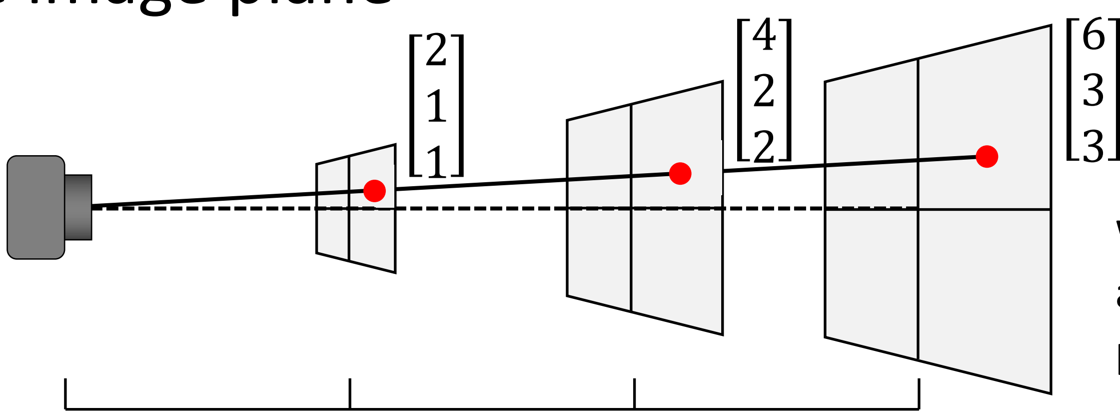


Camera parameters

- **Intrinsic parameters**: camera parameters related to image formation (focal length, optical centre, lens distortion)
- **Extrinsic parameters**: camera pose (location and orientation) relative to the world
- Camera calibration is a process to find the intrinsic parameters
- Usually, these parameters are learned from image data with unknown extrinsic parameters

Homogeneous coordinates

- When converting between world and image points, it is often convenient to use **homogeneous** (or **projective**) coordinates
- Image points are represented with 3 values (x_i, y_i, z_i)
- The third value can be thought of as the distance to the image plane



We generally assume the image plane is at $z_i=1$

Projection model

- The pinhole projection model can be represented as a matrix in homogenous coordinates:

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$x_i = (f_x x) + (c_x z)$$
$$y_i = (f_y y) + (c_y z)$$
$$z_i = z$$

Camera matrix (K)

Coordinates in the image

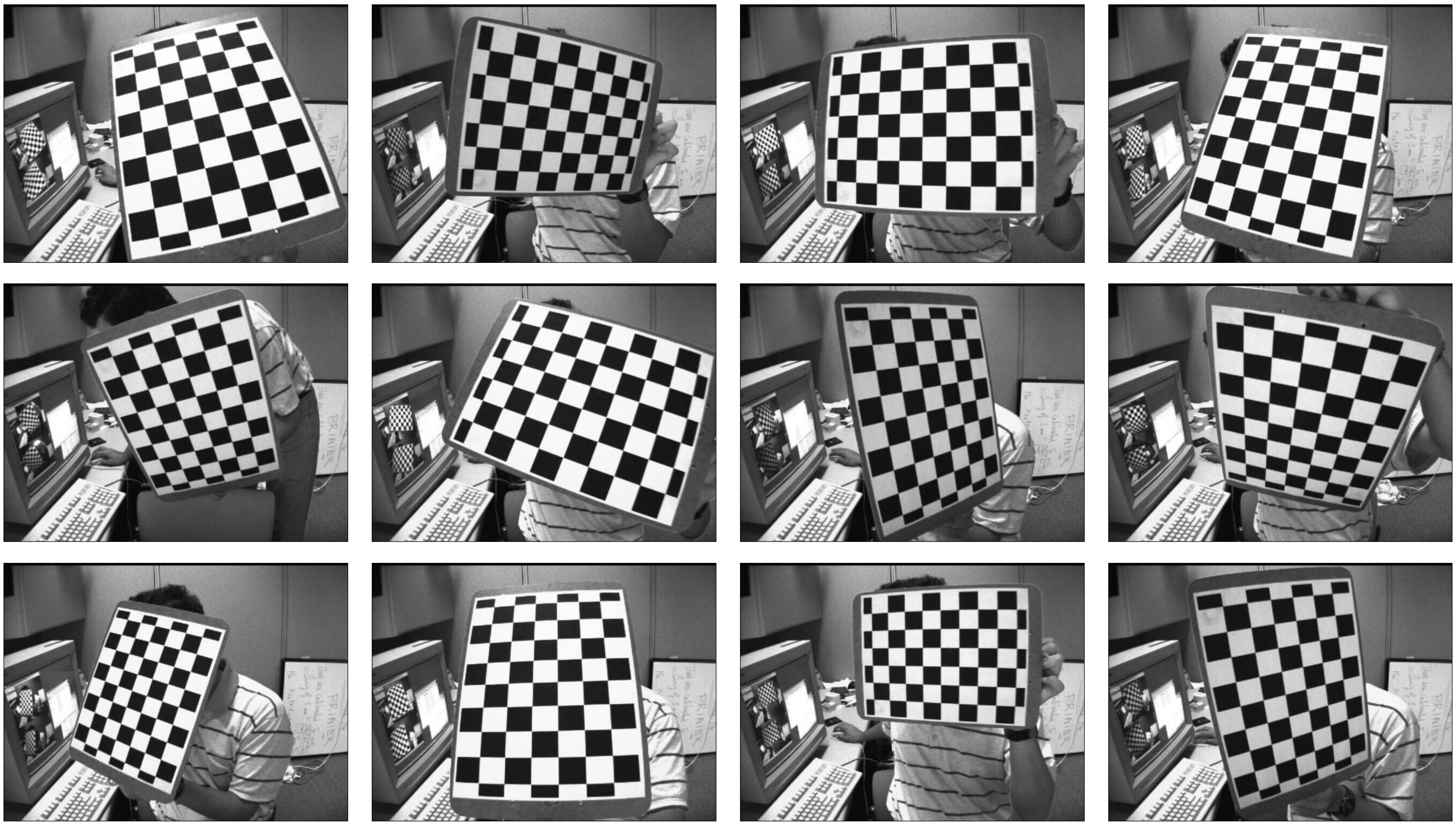
Coordinates in the world

(f_x, f_y) = focal length $\qquad (c_x, c_y)$ = optical centre

Camera calibration method

- Camera calibration requires a calibration target, a planar surface with a known pattern that is easily detected/tracked by feature detection methods
  - Common choices: checkerboard, squares, circles
- Take multiple photos (or a video) of the calibration target in many different poses
- Solve for intrinsic and extrinsic parameters

Calibration target



Projection model

- Relationship between points in the world and points in the image:

$$
\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & | & T \\ 000 & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}
$$

Camera matrix (K)

Rotation + translation

Coordinates in the image

Coordinates in the world

$(f_x, f_y) =$ focal length $\qquad (c_x, c_y) =$ optical centre

Projection model

- Relationship between points in the world and points in the image:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \left[\begin{array}{c|c} R & T \\ 000 & 1 \end{array}\right] \begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix}$$

For simplicity, assume the calibration target is aligned with the plane z=0

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix}$$

Camera calibration algorithm

- Given multiple images, can solve for H, and camera matrix using a system of linear equations
- Note that this model assumes no lens distortion
- Given best fit for H, estimate distortion parameters (different formulas for different distortion models)
- Iterate to refine parameters

Camera calibration result

- Output of calibration process is an estimate of camera intrinsic parameters (camera matrix, lens distortion parameters)
- Allows for accurate mapping between image coordinates and world coordinates

Alternative methods

- Calibration using planar surfaces in the world
  - Advantage: no need for a special calibration target
  - Disadvantage: more difficult to detect/track keypoints, may introduce errors
- Look up camera parameters from manufacturer specifications
  - Advantage: no computation!
  - Disadvantage: only for cameras with fixed focal length

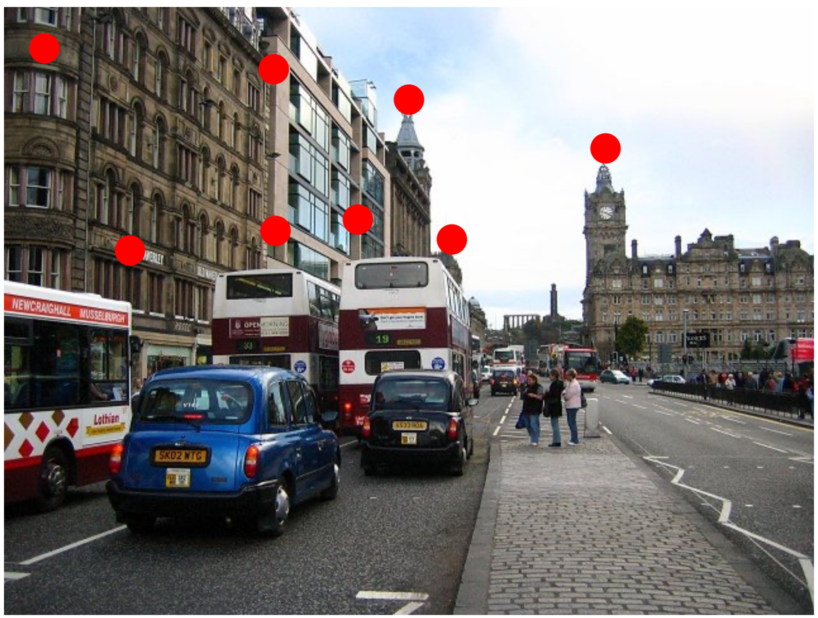
Summary

- Camera calibration is used to recover a camera's intrinsic parameters, expressed as a camera matrix
- Calibration is required for applications that involve accurate mapping between world and image points (e.g., augmented reality)

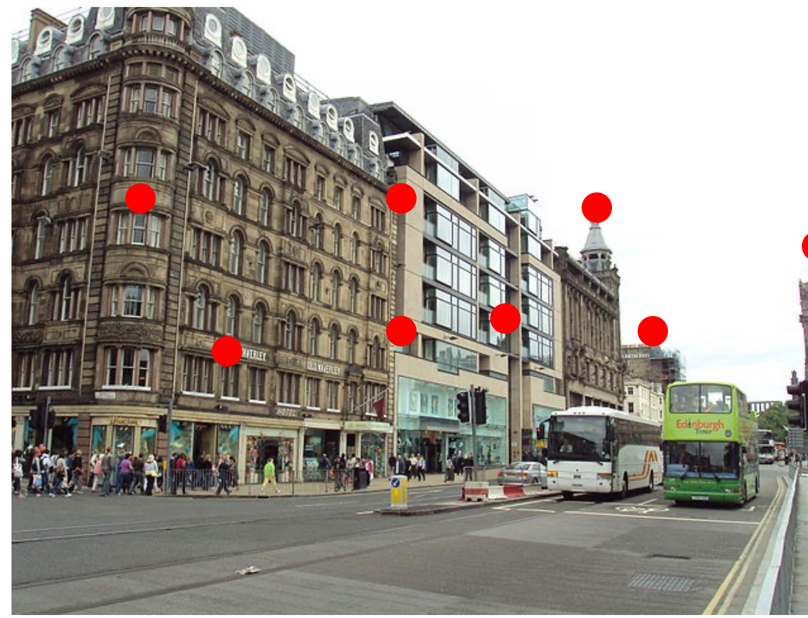
Epipolar geometry - basics

Two-view problem

- What is the camera transform (translation + rotation) that relates these two views?



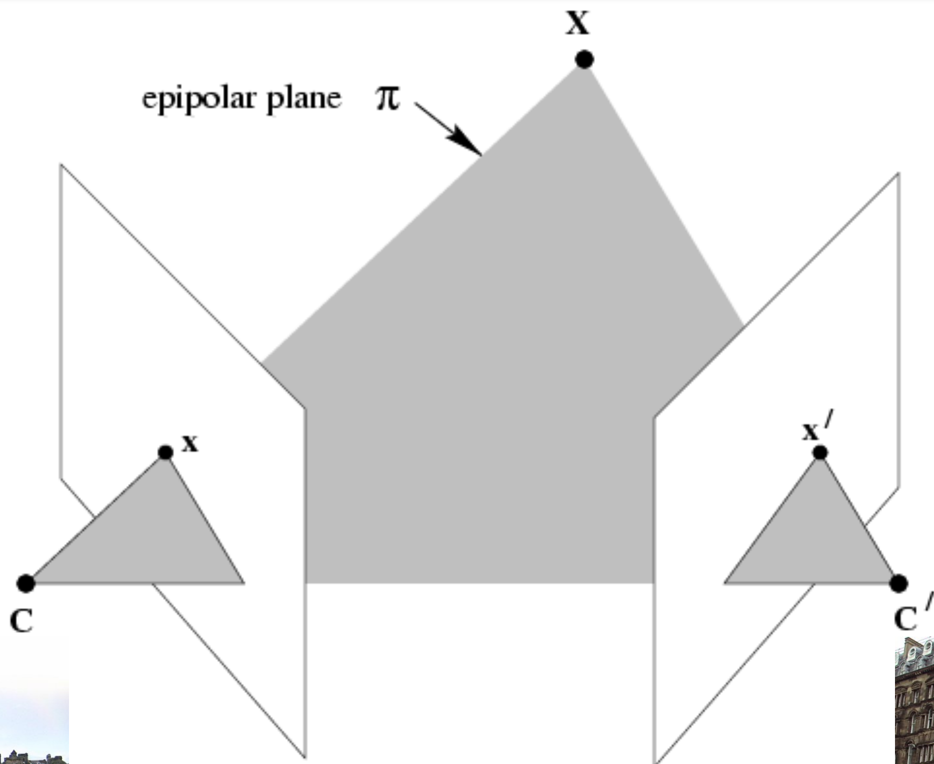
View from camera 1



View from camera 2

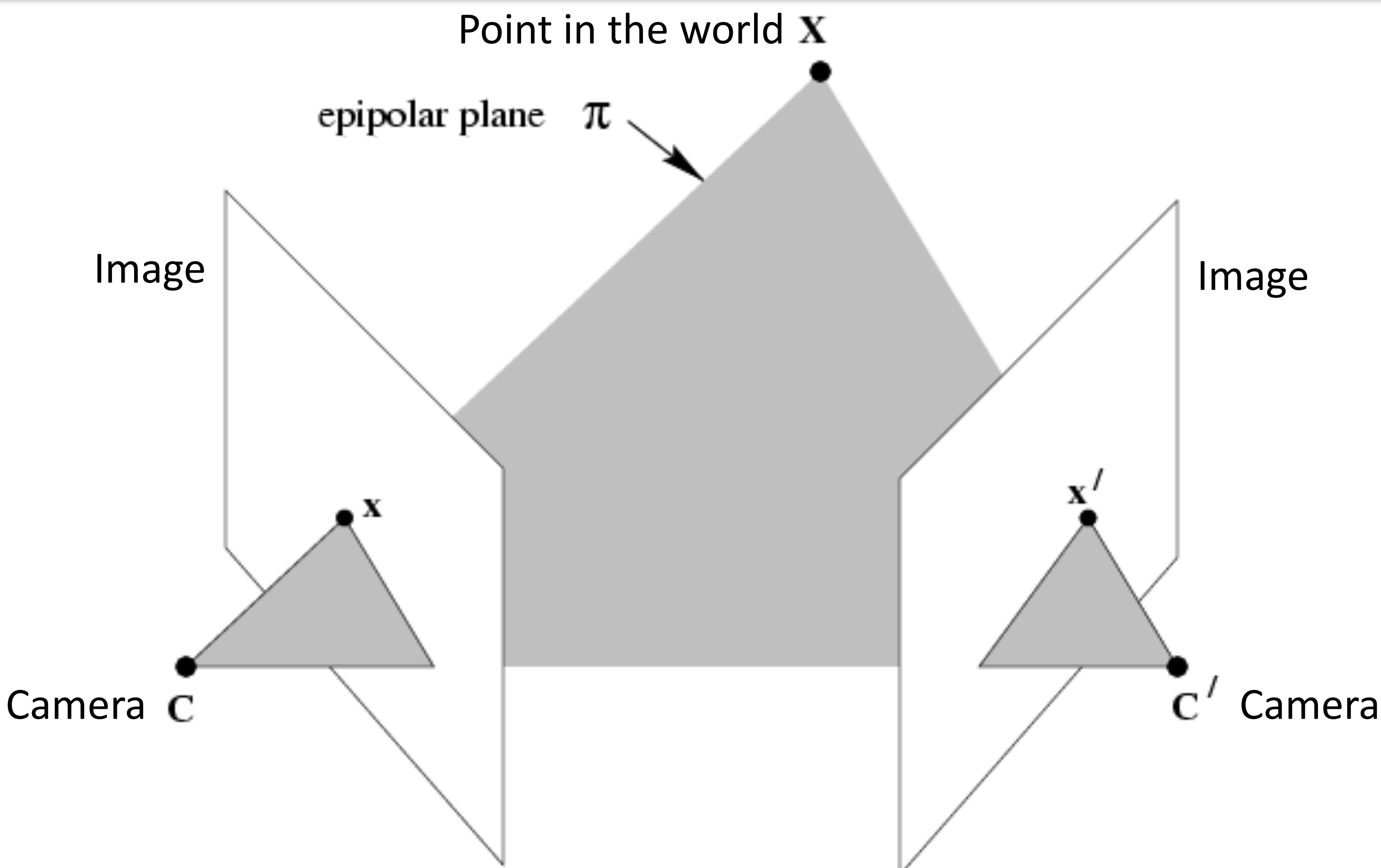
Estimating camera transform

What is the transformation matrix that relates these two views?



Camera translation+rotation

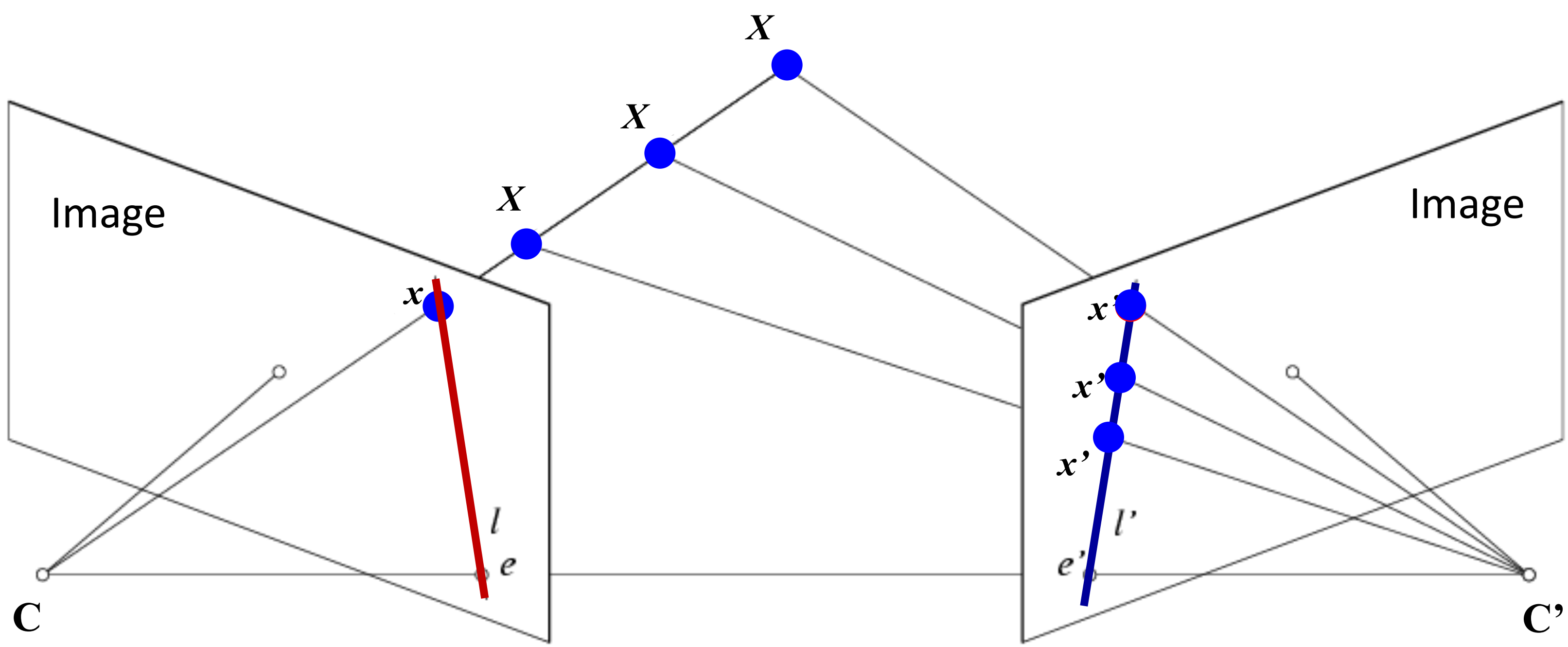
Key idea: Epipolar constraint



C, C’, x, x’ and X are coplanar

Figure: Hartley & Zisserman (2004)

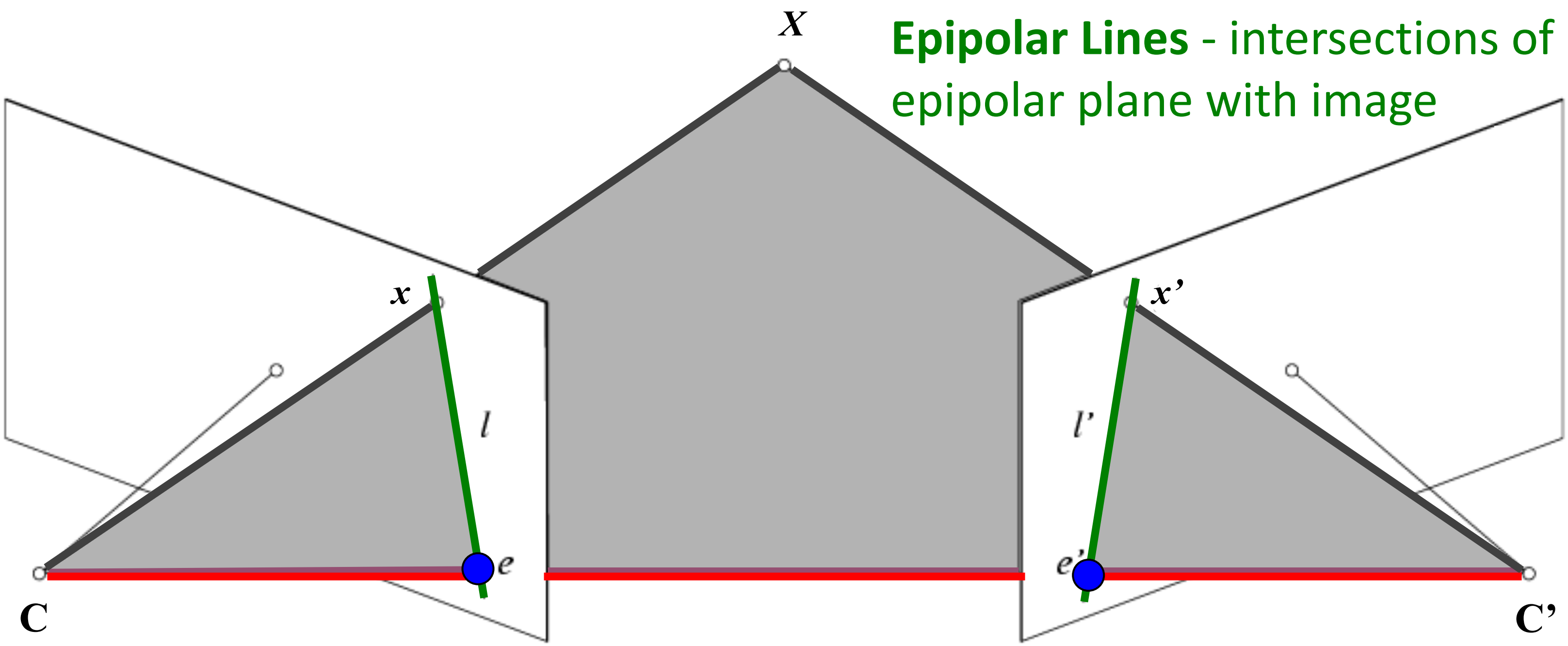
Key idea: Epipolar constraint



Potential matches for x must lie on the line l'.

Potential matches for x' must lie on the line l.

Epipolar geometry: Notation



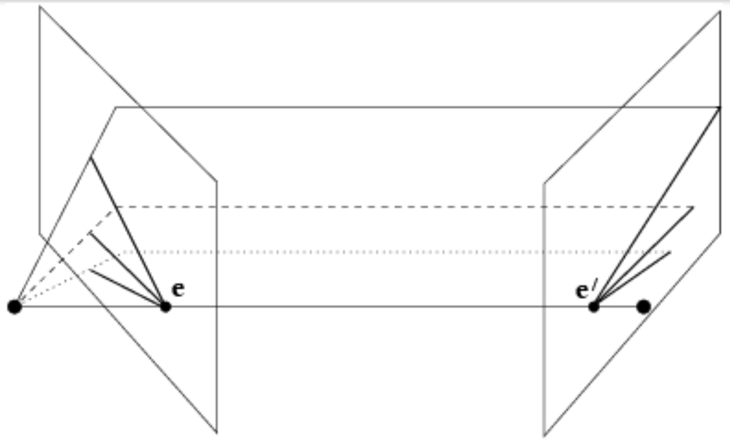
Epipolar Lines - intersections of epipolar plane with image

Baseline – line connecting the two camera centers

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

- **Epipolar Plane** – plane containing baseline

Example: Epipolar lines



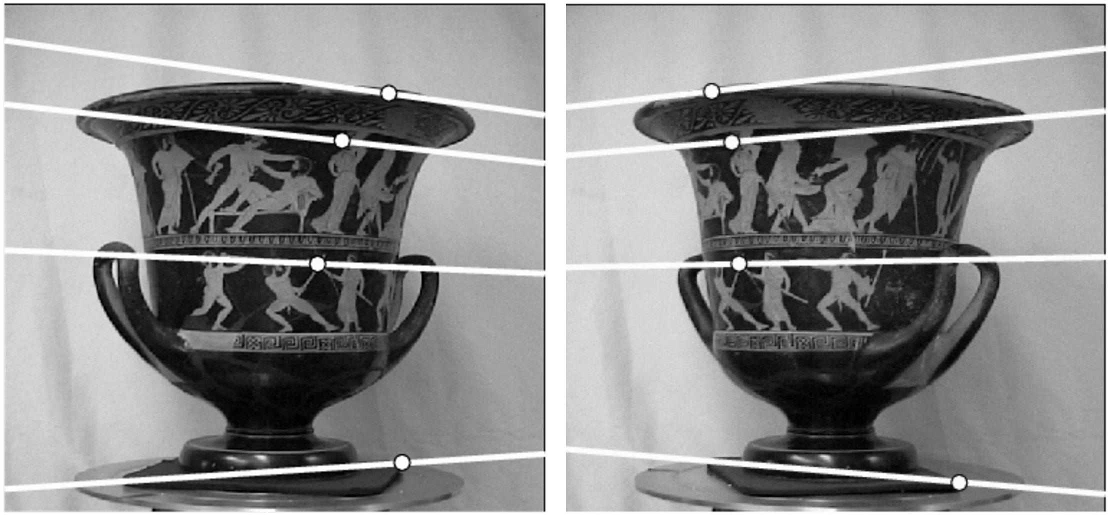


Figure: Hartley & Zisserman (2004)

Example: Horizontal motion

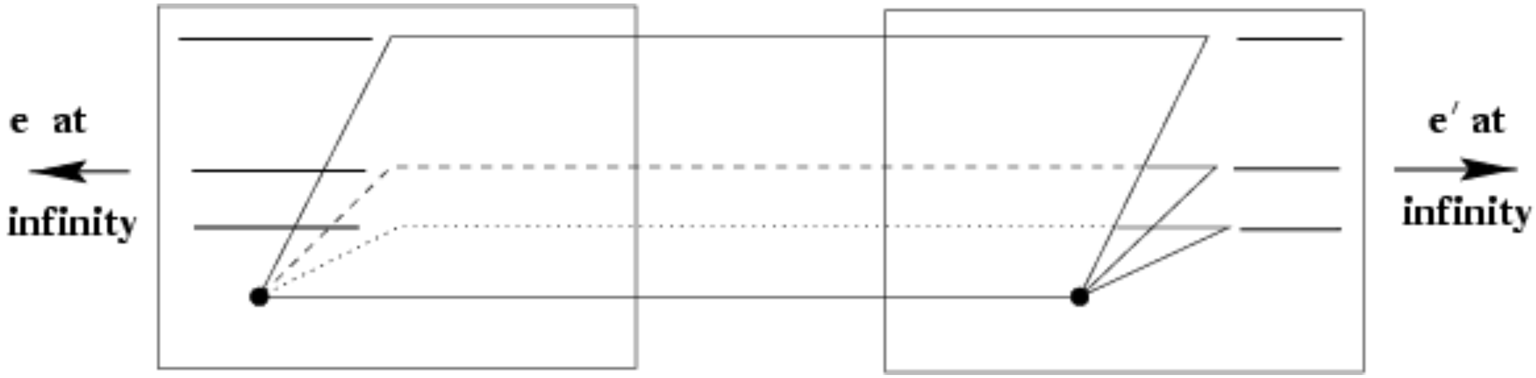


Figure: Hartley & Zisserman (2004)

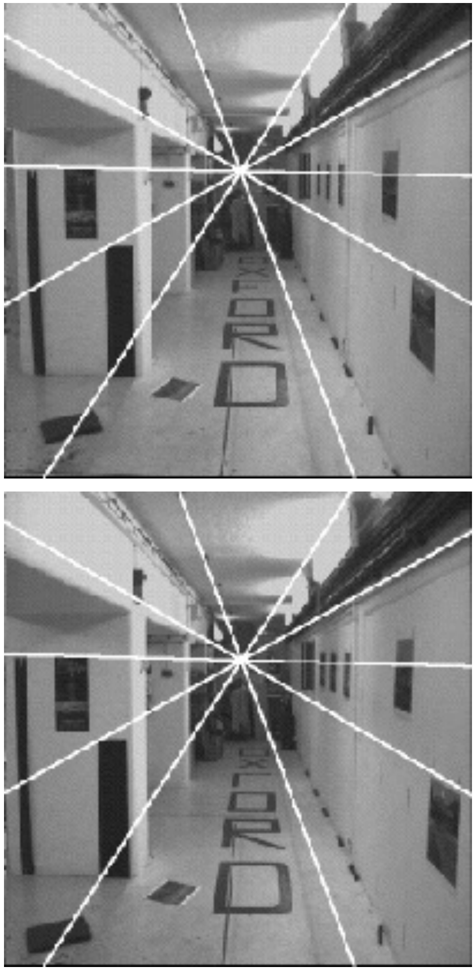
Example: Forward motion

 image: Hartley & Zisserman (2004)

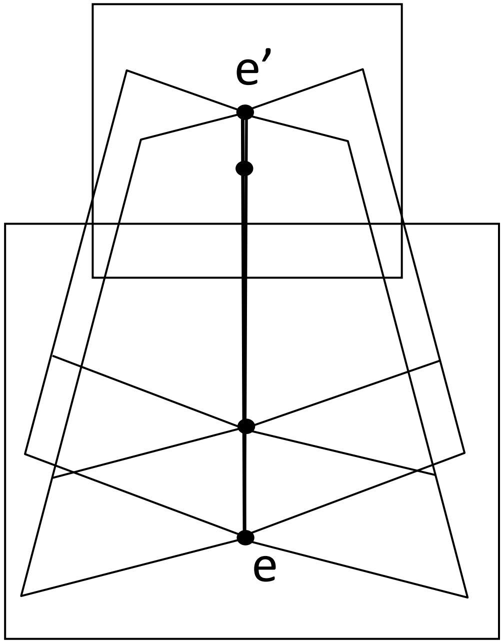
Example: Forward motion

 image: Hartley & Zisserman (2004)

Example: Forward motion



Epipole has same coordinates in both images
Points move along lines radiating from epipole e
(called the "focus of expansion")



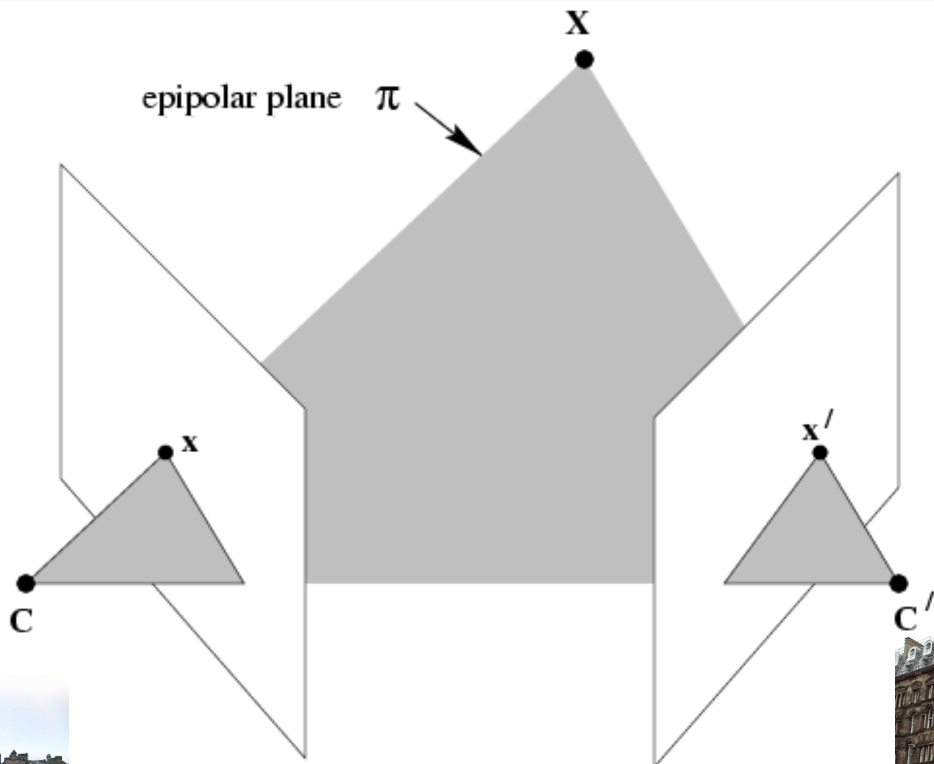
Summary

- Epipolar geometry describes relations between points in two views
- A point in one image lies along an **epipolar line** in the other image
- Epipolar lines in an image meet at a point called the **epipole**
- The epipole is the projection of one camera in the other image

Epipolar geometry - math

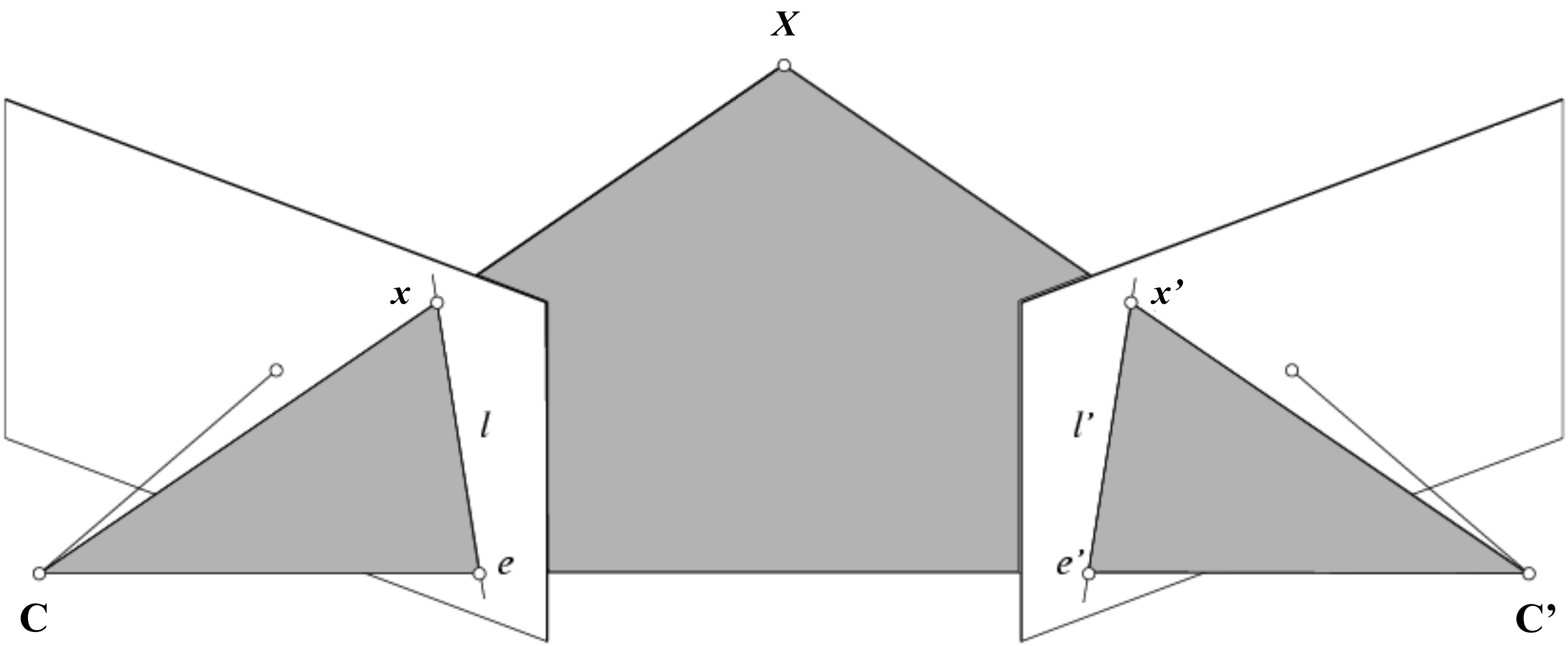
Two-view problem

What is the transformation matrix that relates these two views?



Camera
translation+rotation

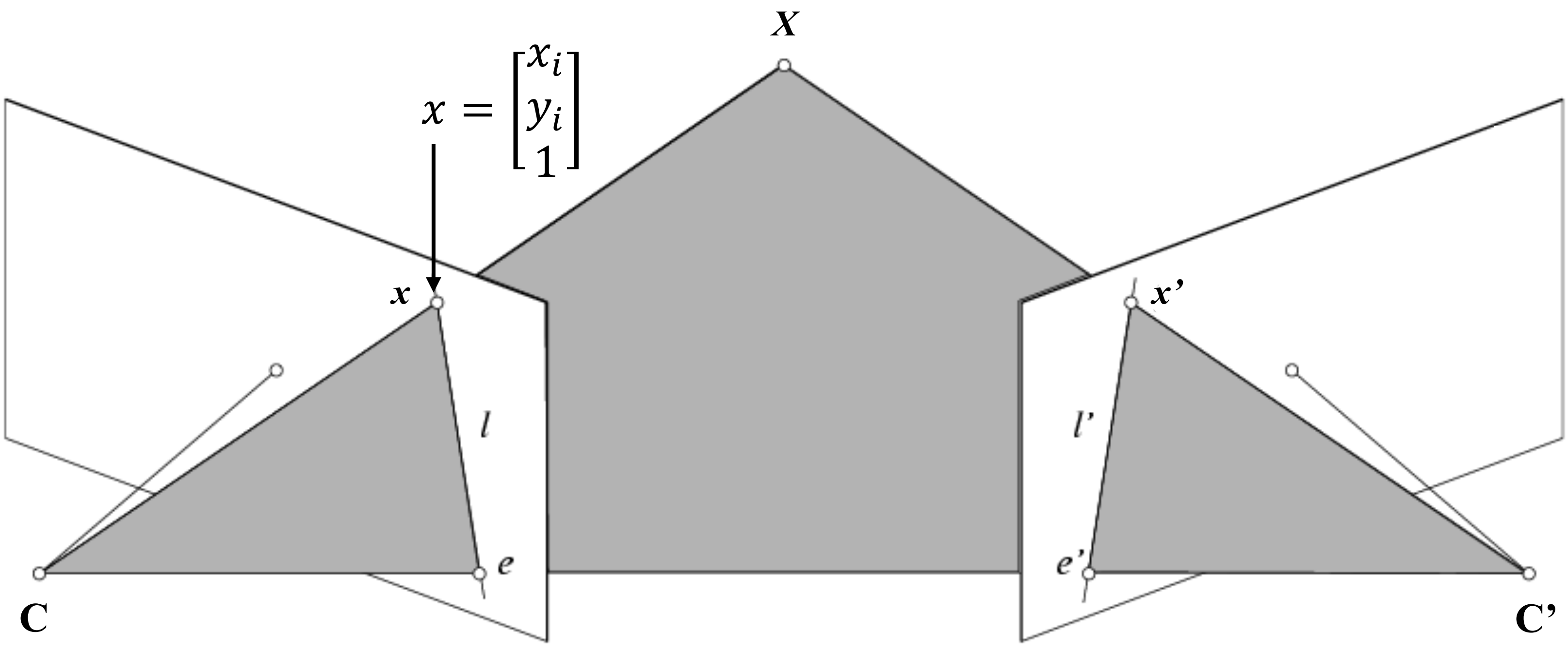
Epipolar constraint



If the camera matrix (K) is known, use it to convert image points x and x' to homogeneous coordinates

$$\hat{x} = \underbrace{\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}}_{K} X$$

Epipolar constraint



$$x = \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

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

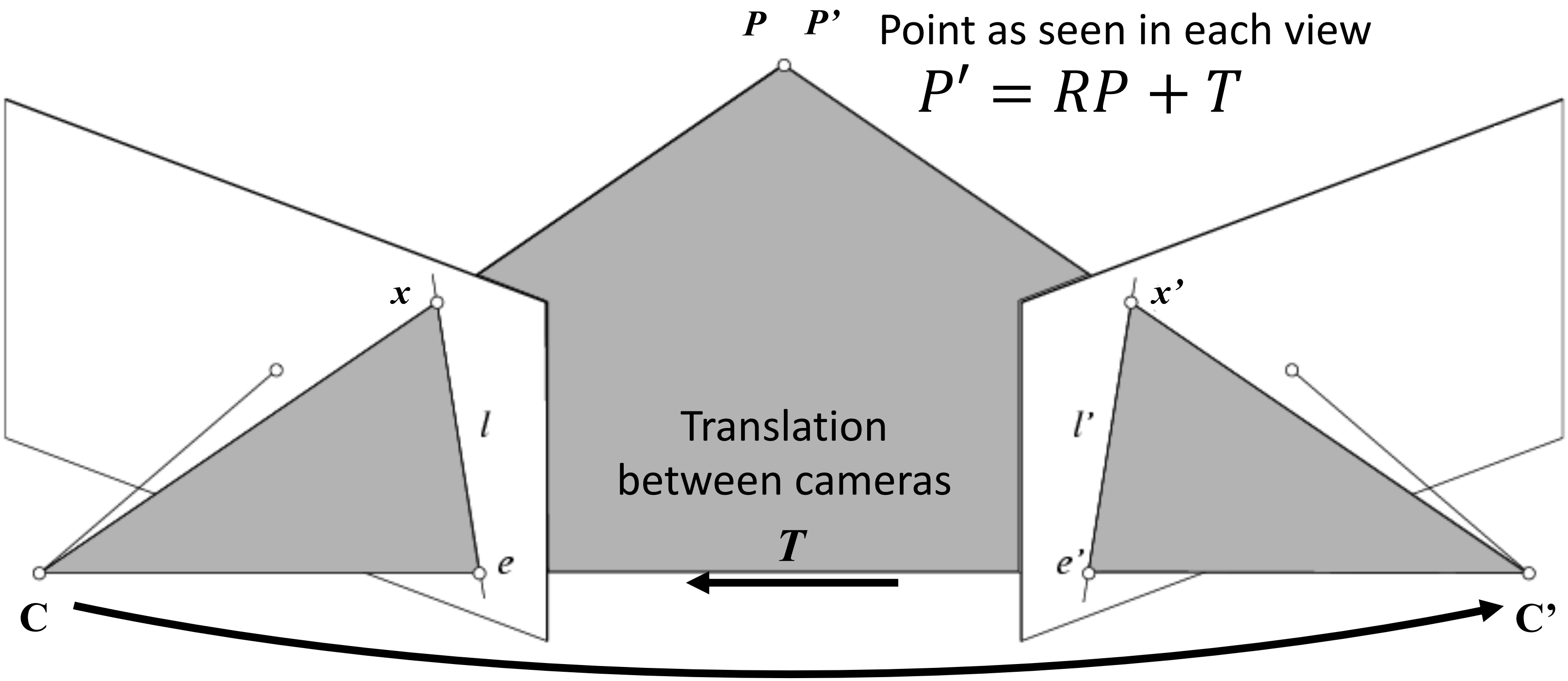
View ray in homogeneous coordinates

Camera C matrix

Point in image

Camera C' matrix

Epipolar constraint



Point as seen in each view

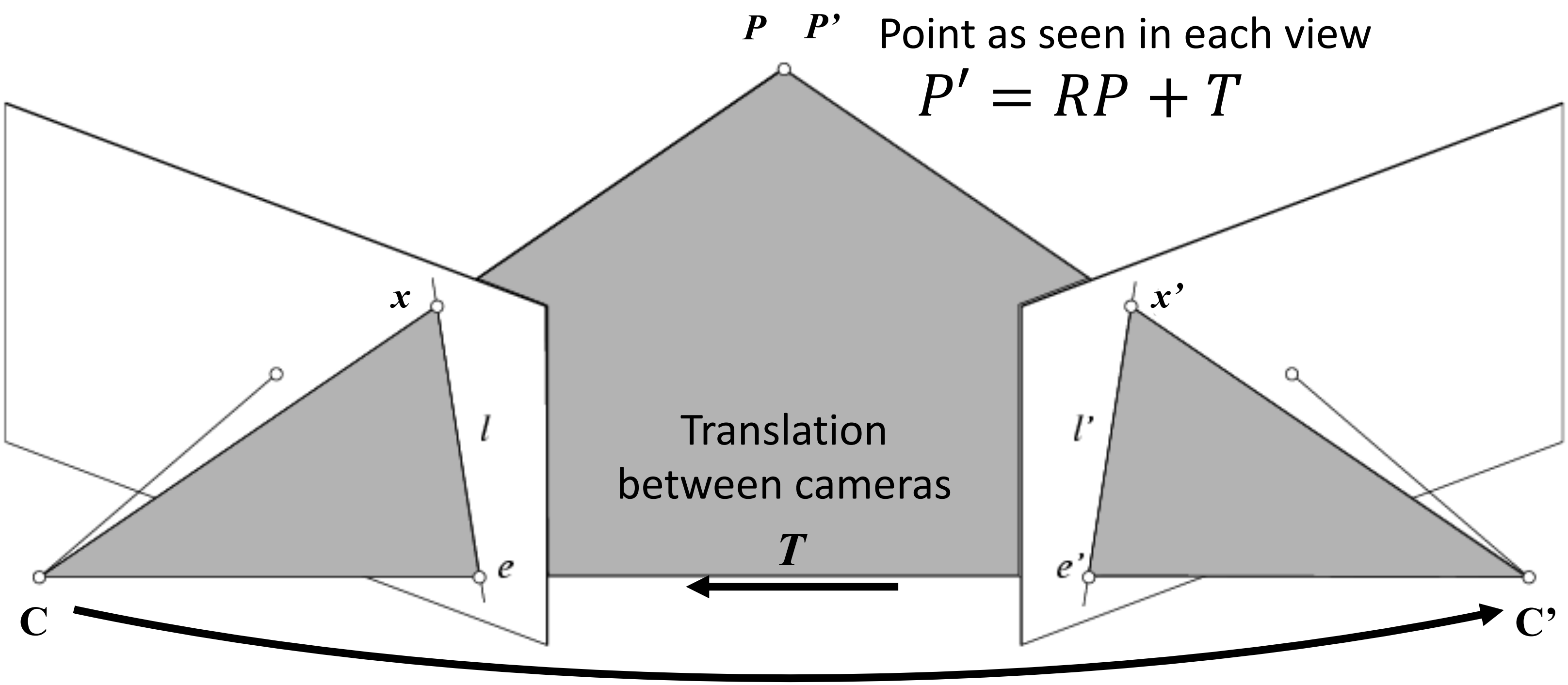
$P' = RP + T$

Translation between cameras

Rotation between cameras

Key constraint: vectors $\widehat{x'}, T, R\hat{x}$ and are all coplanar

Epipolar constraint



Point as seen in each view

$$P' = RP + T$$

Translation between cameras

Rotation between cameras

$$\widehat{x'} \cdot [T \times R\hat{x}] = 0$$

$$\widehat{x'}^T \left[[T]_x R \right] \hat{x} = 0$$

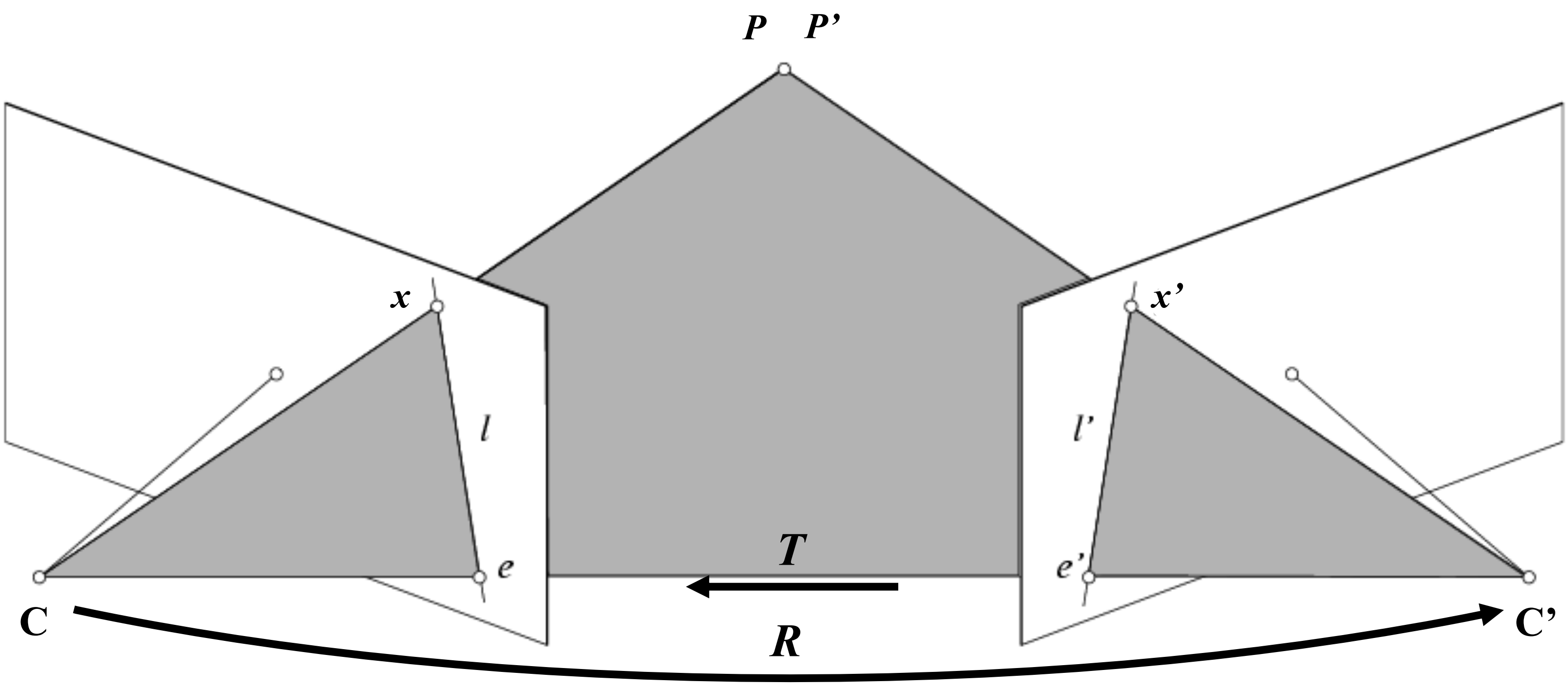
Cross product as matrix notation

$$\widehat{x'}^T E \hat{x} = 0$$

$$E = [T]_x R$$

E = Essential matrix

Properties of Essential matrix



$E\hat{x}$ is the epipolar line l',$E^T\hat{x}'$ is the epipolar line l

$Ee' = 0$ gives the epipole e' in the right image (projection of left camera's centre C)

$E^Te = 0$ gives the epipole e in the left image (projection of right camera's centre C')

E has 5 degrees of freedom (rotation (3) + translation (2) missing a scaling factor)

Fundamental matrix

- What if the camera parameters are unknown (uncalibrated cameras)?
- Can define a similar relationship using the unknown K and K':

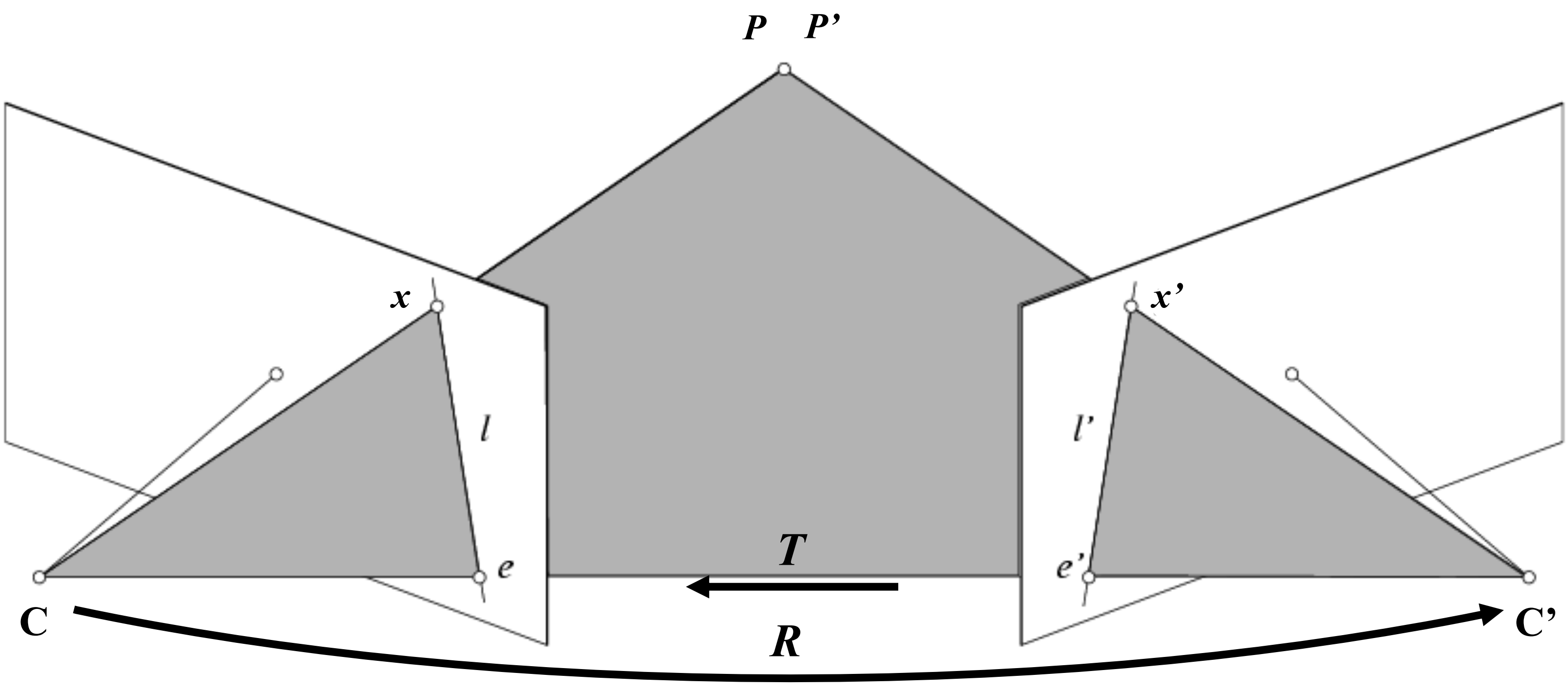
$$\widehat{x'}^T E \hat{x} = 0 \qquad\qquad x'^T F x = 0$$

$$\hat{x} = K^{-1} x \qquad\qquad F = K'^{-T} E K^{-1}$$

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

F = Fundamental matrix

Properties of Fundamental matrix

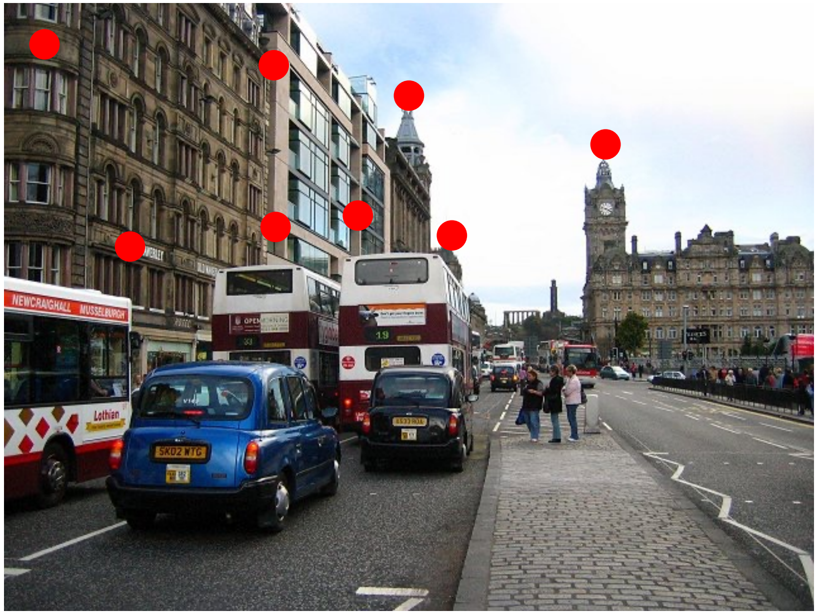


Fx is the epipolar line l',$F^T x'$ is the epipolar line l

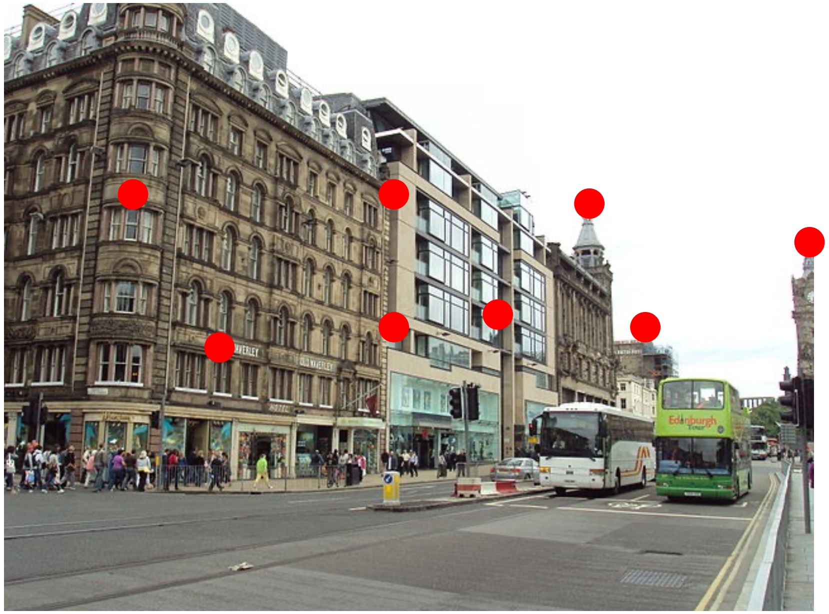
F has 7 degrees of freedom (3x3 matrix, but it has an unknown scaling factor and Det(F)=0, which removes 2 degrees of freedom)

Solving for F

- How to solve for F when K and K' are unknown?
- Match pairs of points across views and find matrix F that explains the correspondences



View from camera 1

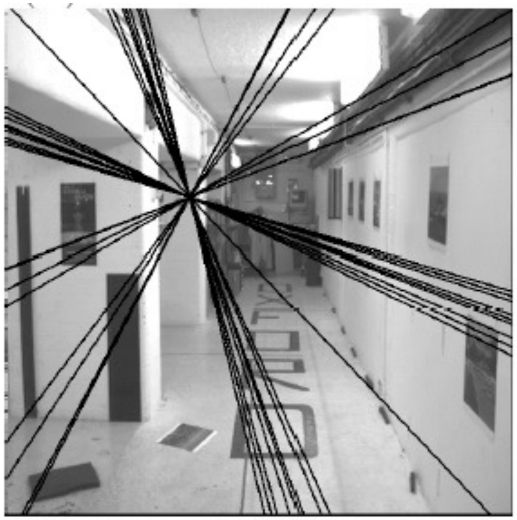


View from camera 2

Solving for F

- 8-point algorithm
  - Requires 8 matching points
  - Solve for F as a linear system of equations
  - Additional steps (SVD = singular value decomposition) to ensure that F has the correct form

Epipolar lines, using F from least squares solution to linear system



Epipolar lines, using final estimate of F

Figure: Hartley & Zisserman (2004)

The 8 point algorithm

Each correspondence between two views gives us:

$$\begin{pmatrix} q_1 & q_2 & 1 \end{pmatrix} \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{pmatrix} p_1 \\ p_2 \\ 1 \end{pmatrix} = 0$$

This equation can be expanded out to give:

$$p_1q_1F_{11} + p_2q_1F_{12} + q_1F_{13} + p_1q_2F_{21} + p_2q_2F_{22} + q_2F_{23} + p_1F_{31} + p_2F_{32} + F_{33} = 0$$

This is a single linear constraint on the values of F and can be rewritten in matrix form

The 8 point algorithm

$$\begin{pmatrix} p_1q_1 & p_2q_1 & q_1 & p_1q_2 & p_2q_2 & q_2 & p_1 & p_2 & 1 \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = 0$$

We get one pair of p and q for each correspondence

But the same F must apply to all correspondences

The 8 point algorithm

So with 8 correspondences we get a matrix like this (where each row has the p and q from different correspondences

$$\begin{pmatrix} p_1q_1 & p_2q_1 & q_1 & p_1q_2 & p_2q_2 & q_2 & p_1 & p_2 & 1 \\ p_1q_1 & p_2q_1 & q_1 & p_1q_2 & p_2q_2 & q_2 & p_1 & p_2 & 1 \\ p_1q_1 & p_2q_1 & q_1 & p_1q_2 & p_2q_2 & q_2 & p_1 & p_2 & 1 \\ p_1q_1 & p_2q_1 & q_1 & p_1q_2 & p_2q_2 & q_2 & p_1 & p_2 & 1 \\ p_1q_1 & p_2q_1 & q_1 & p_1q_2 & p_2q_2 & q_2 & p_1 & p_2 & 1 \\ p_1q_1 & p_2q_1 & q_1 & p_1q_2 & p_2q_2 & q_2 & p_1 & p_2 & 1 \\ p_1q_1 & p_2q_1 & q_1 & p_1q_2 & p_2q_2 & q_2 & p_1 & p_2 & 1 \\ p_1q_1 & p_2q_1 & q_1 & p_1q_2 & p_2q_2 & q_2 & p_1 & p_2 & 1 \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The vector of F values that we want is the null space of this matrix

Limitations

- System is only solved up to a scaling factor, need at least one known distance to solve for real-world positions
- Degenerate cases: can’t solve if the system has too few degrees of freedom
  - Points in the world are all coplanar
  - Camera translation = 0 (just rotation)

Example: Refining GPS locations

- Google Streetview GPS coordinates are not always exact – refine locations using epipolar geometry

Image 1

Where is camera 2 in image 1?

Image 2

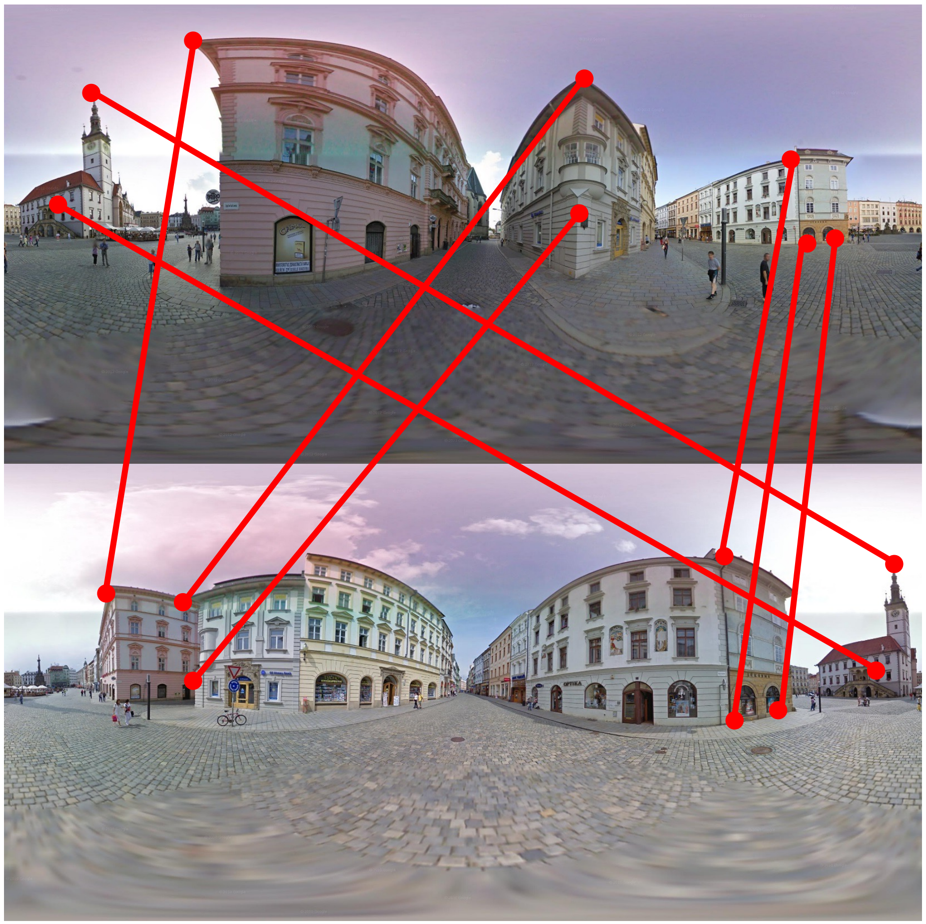
Where is camera 1 in image 2?

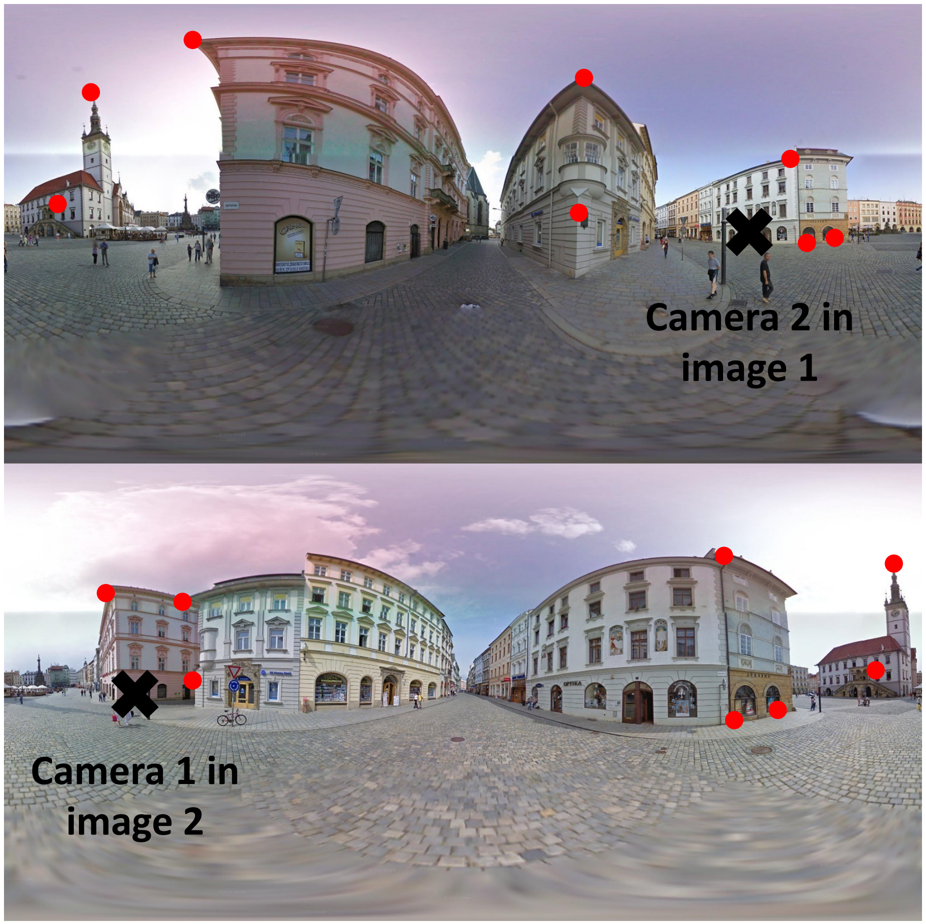
Images: Google

Example: Refining GPS location

- Algorithm:
  - Detect ASIFT (affine-invariant SIFT) keypoints in each image, find potential matches using ratio test
  - Use RANSAC to find the Fundamental matrix (F) that relates the two views, and the inlier matches that are explained by that transform
  - Compute Essential matrix (E) from F using known camera matrices
  - Decompose E into rotation and translation between cameras

Example: Refining GPS location



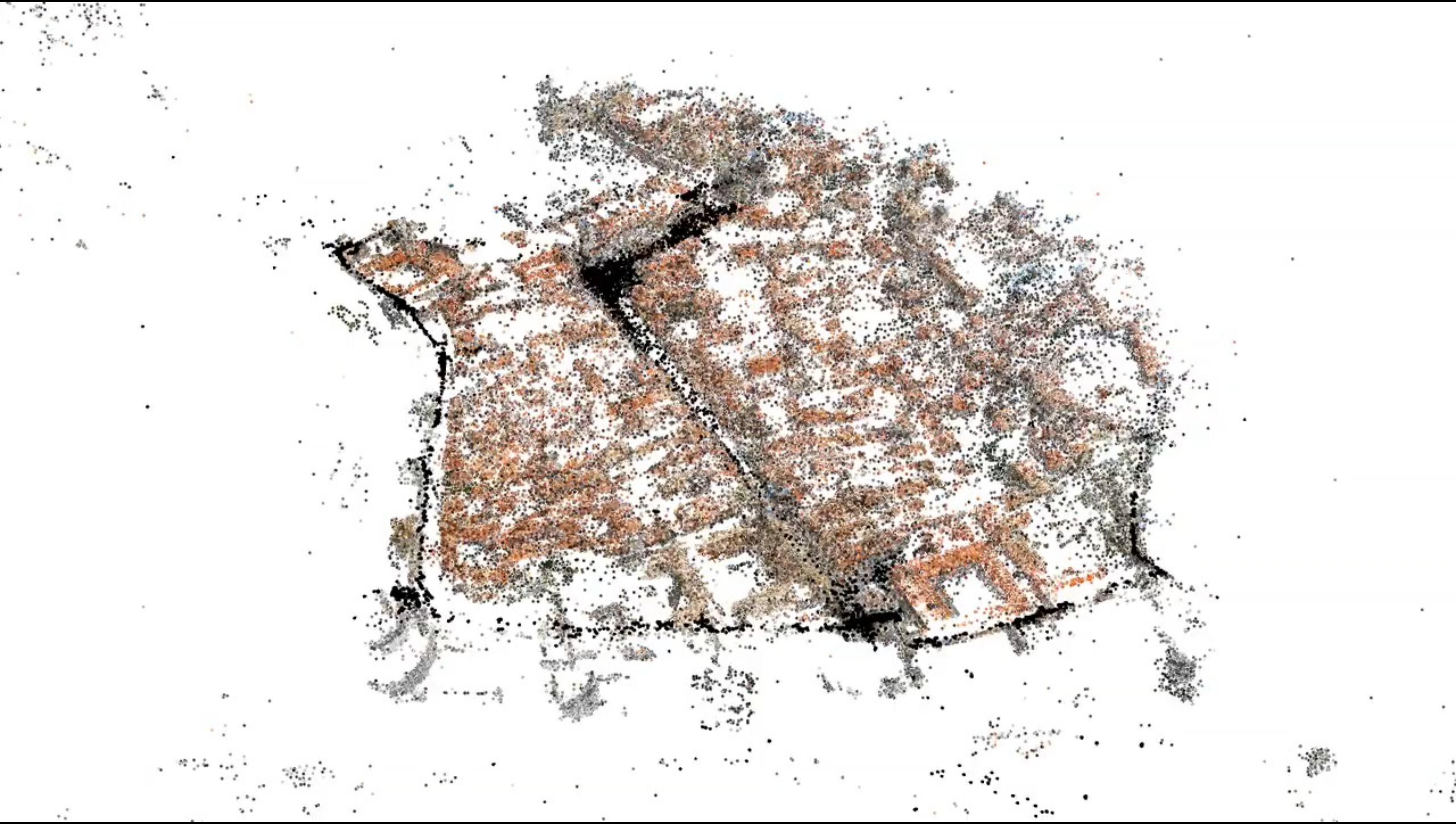


Summary

- Epipolar geometry describes how a point in 3D space is imaged through a pair of cameras
- Essential and Fundamental matrices map points in one image to a line (epipolar line) in the other image
- Typically, use feature detection to find matching points in the two views, then solve for Fundamental matrix (e.g., using RANSAC)

Beyond two-view geometry

- Better depth results can be obtained by combining more than two views:
  - Structure from motion
  - Simultaneous localisation and mapping (SLAM)



https://grail.cs.washington.edu/rome/