CMT107 Visual Computing

Stereo Vision

Jing Wu

School of Computer Science and Informatics
Cardiff University

Overview

- Stereo Vision
  - Multi-view geometry problems
- Triangulation
- Epipolar Geometry
  - The epipolar constraint
  - Essential matrix and fundamental matrix
  - Eight-point algorithm

Acknowledgement
The majority of the slides in this section are from Svetlana Lazebnik at University of Illinois at Urbana-Champaign

Stereo Vision

- Geometric problem formulation: given several images of the same object or scene, compute a representation of its 3D shape.

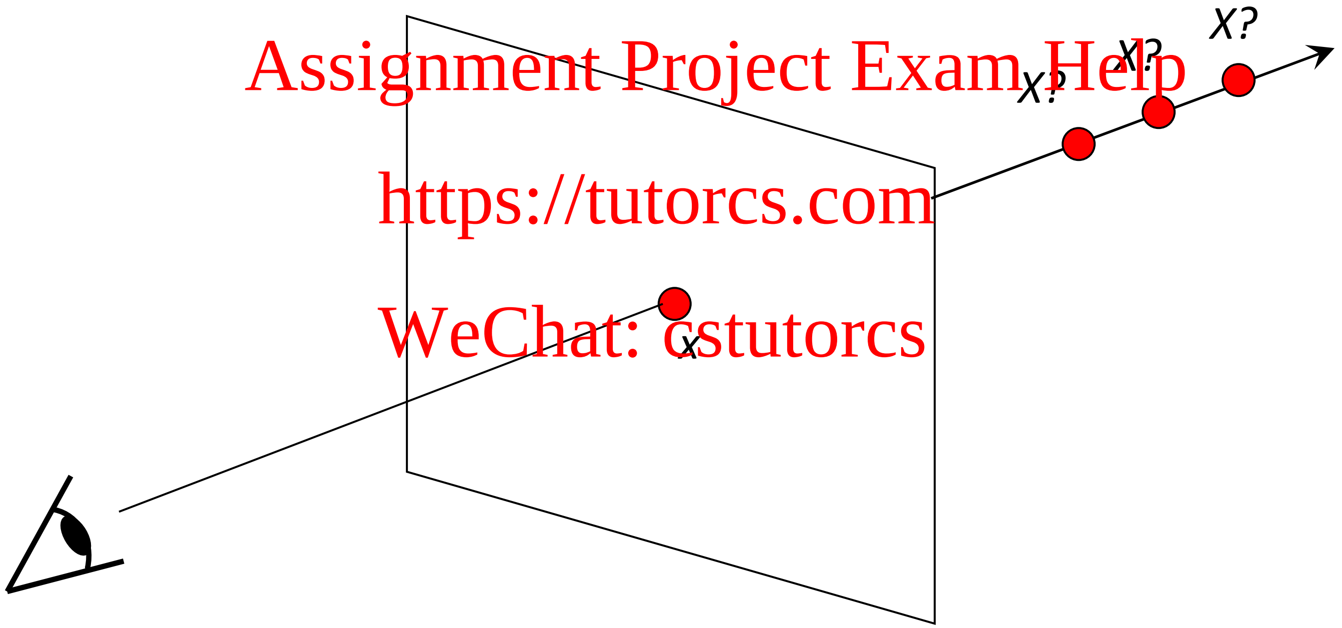


Goal: Recovery of 3D Structure

- Geometric problem formulation: given several images of the same object or scene, compute a representation of its 3D shape.
- "Images of the same object or scene"
  - Arbitrary number of images (from two to thousands)
  - Arbitrary camera positions (camera network or video sequence)
  - Calibration may be initially unknown
- "Representation of 3D shape"
  - Depth maps
  - Meshes
  - Point clouds
  - Patch clouds
  - Volumetric models
  - Layered models

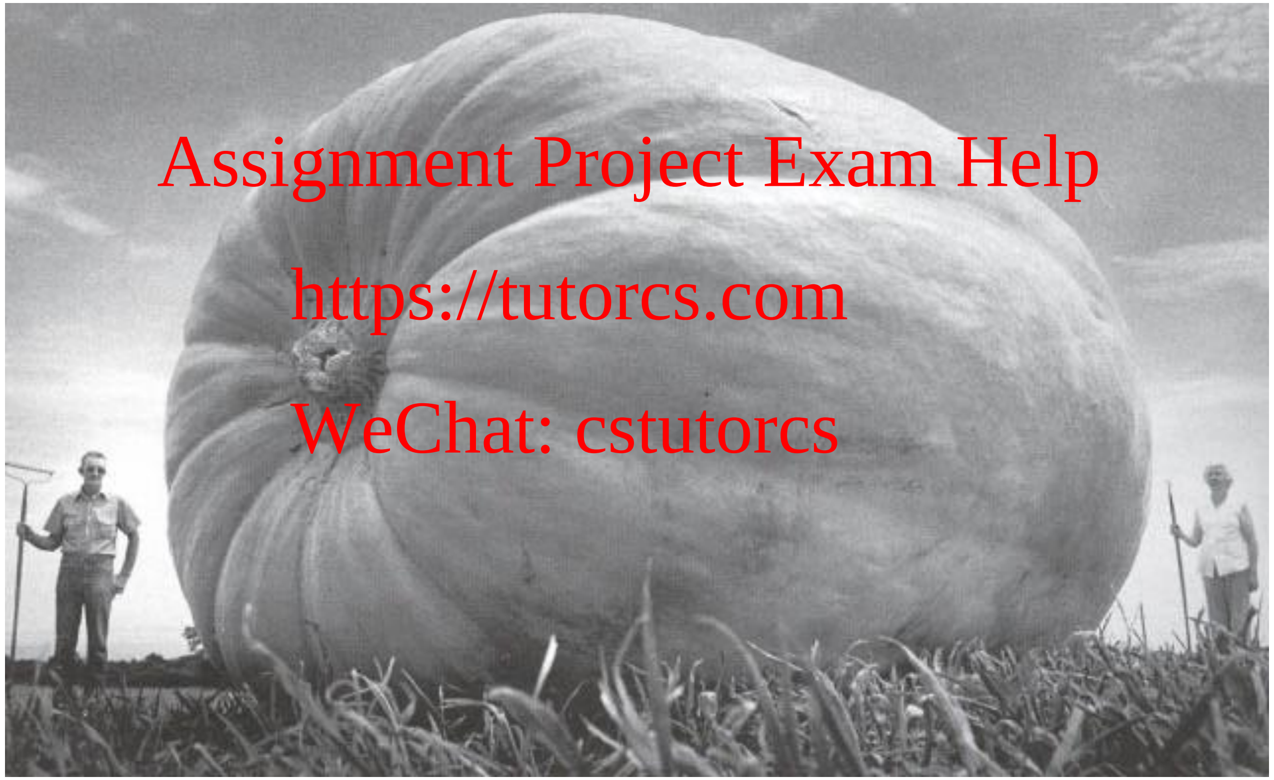
Goal: Recovery of 3D Structure

- Recovery of structure from one image is inherently ambiguous



Goal: Recovery of 3D Structure

- Recovery of structure from one image is inherently ambiguous

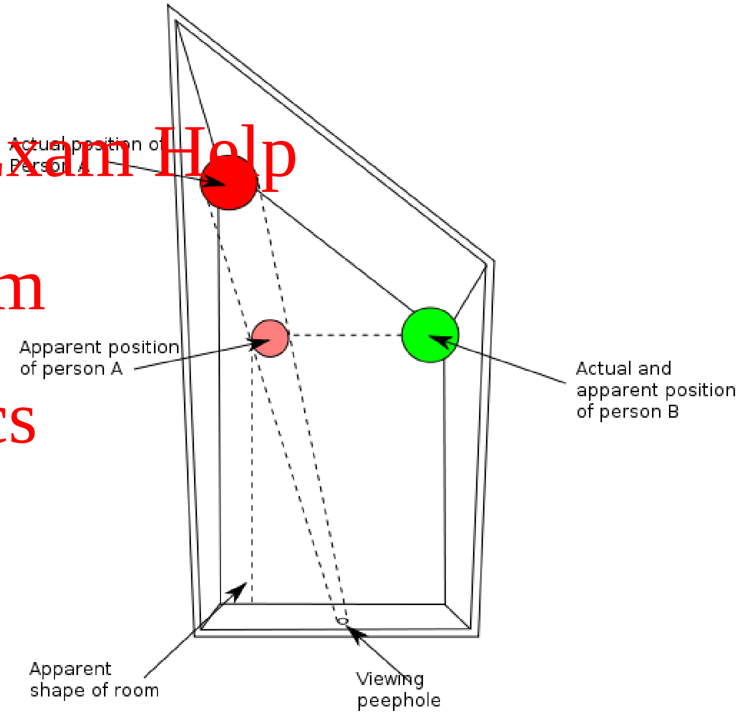


Goal: Recovery of 3D Structure

- Recovery of structure from one image is inherently ambiguous

Goal: Recovery of 3D Structure

- Recovery of structure from one image is inherently ambiguous



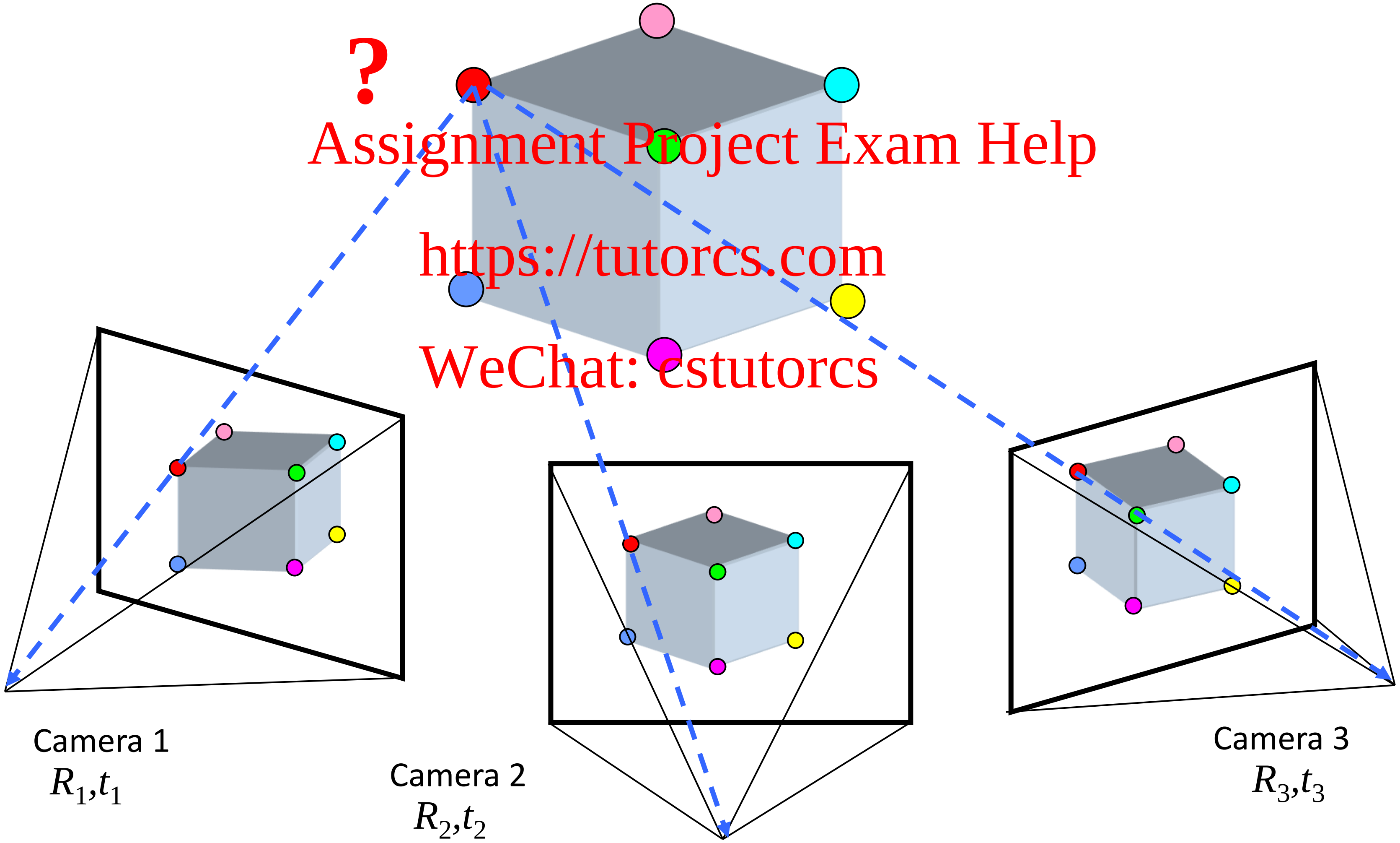
http://en.wikipedia.org/wiki/Ames_room

Goal: Recovery of 3D Structure

- We will need multi-view geometry

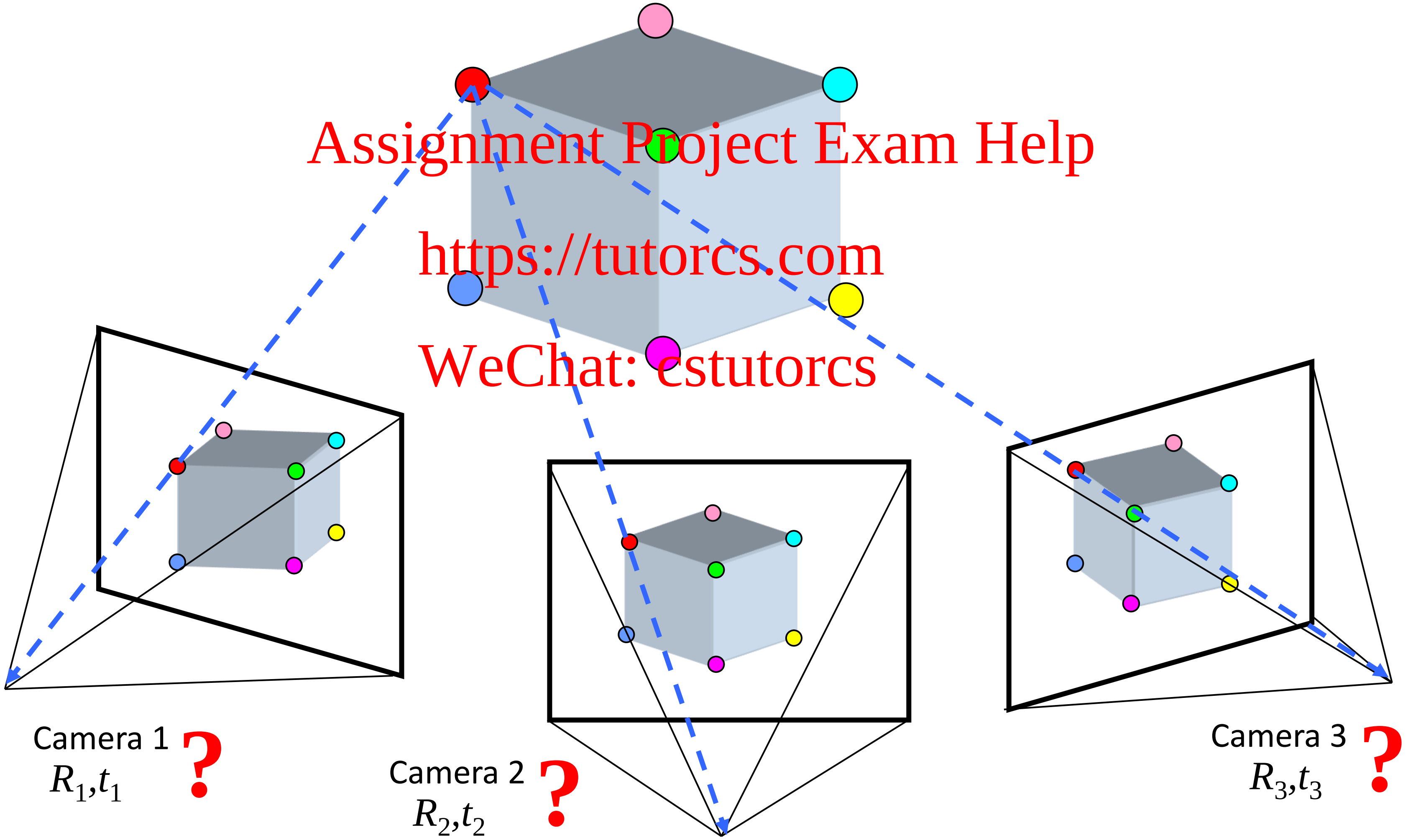
Multiview Geometry Problems

- Structure: Given projections of the same 3D point in two or more images, compute the 3D coordinates of that point



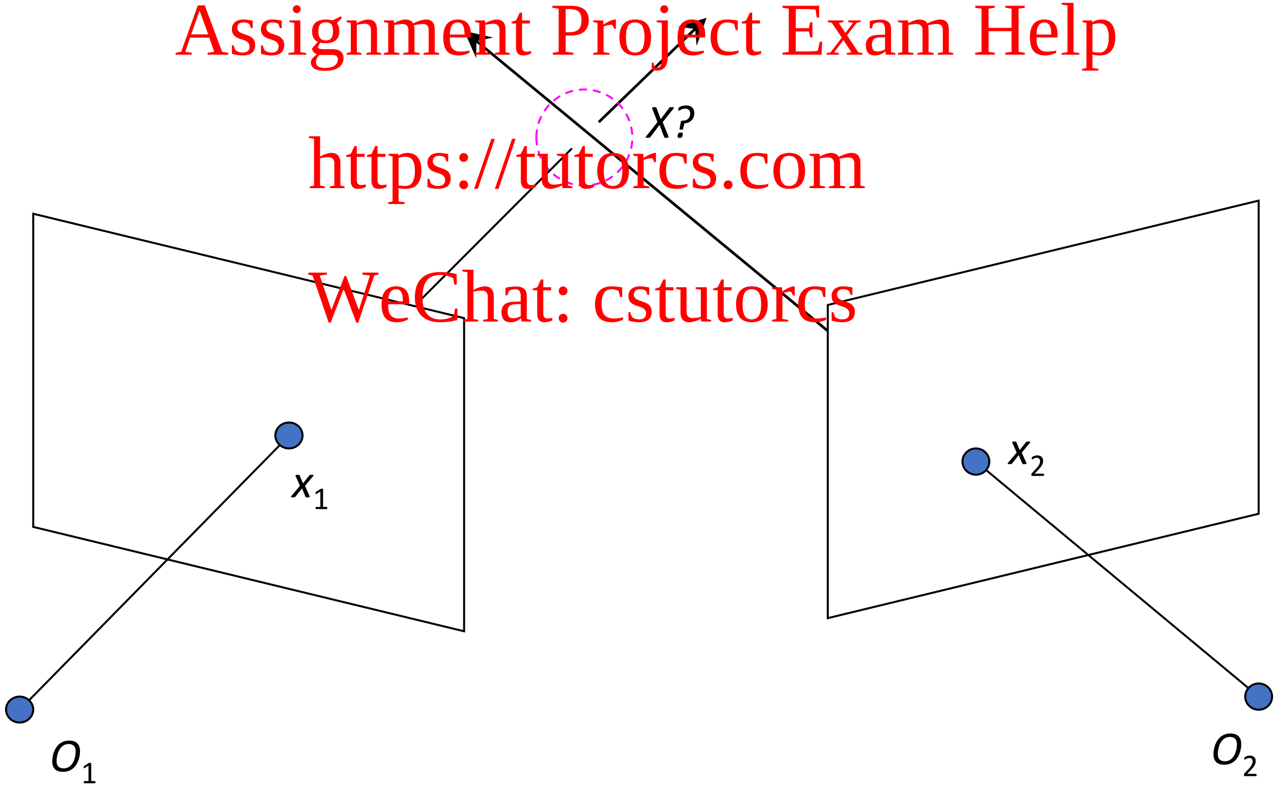
Multiview Geometry Problems

- Motion: Given a set of corresponding points in two or more images, compute the camera parameters



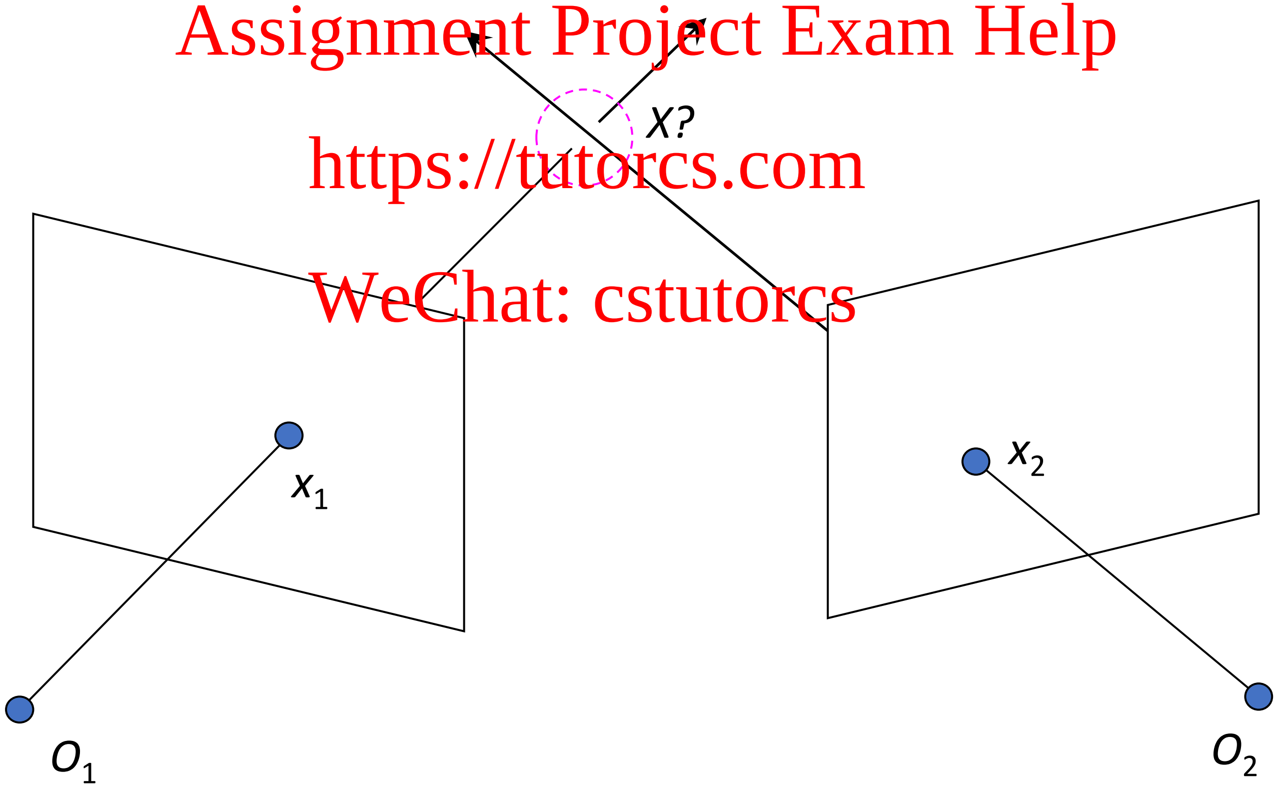
Triangulation

- Given projects of a 3D point in two or more images (with known camera matrices), find the coordinates of the point



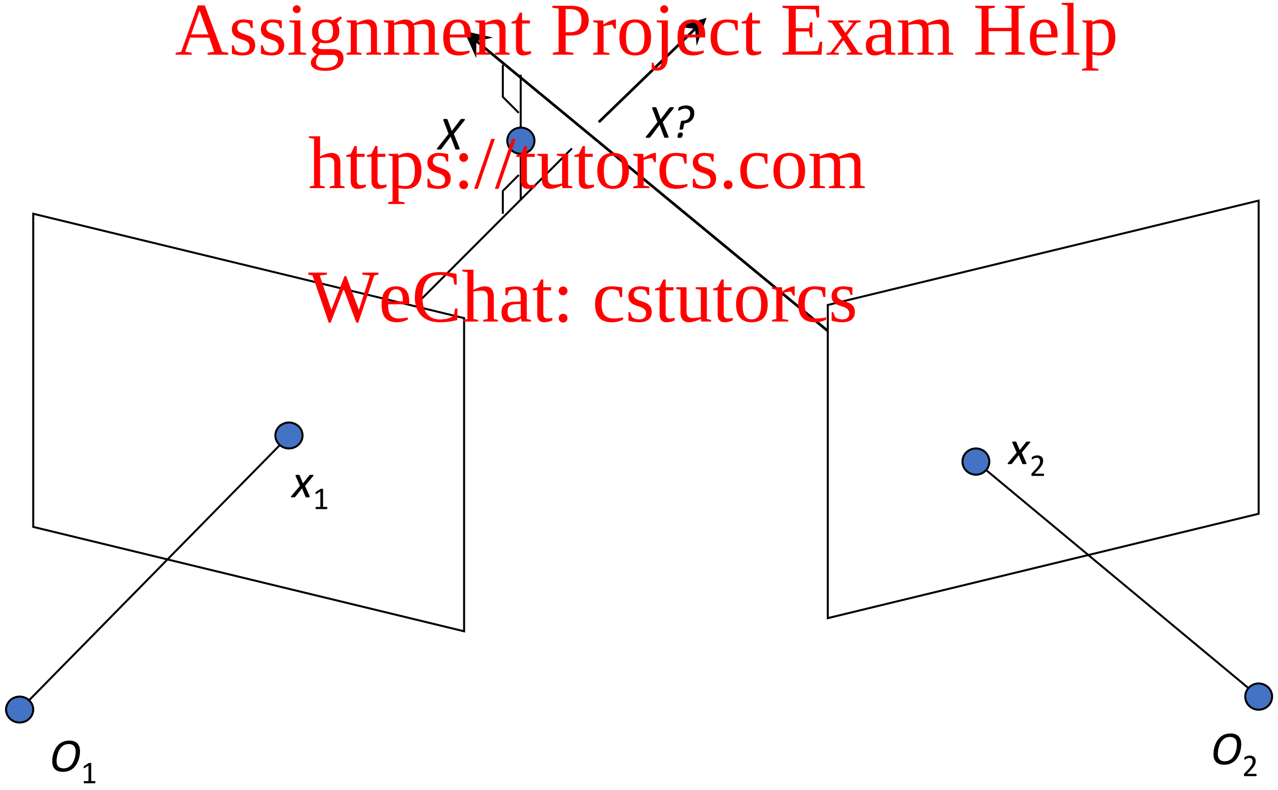
Triangulation

- We want to intersect the two visual rays corresponding to x_1 and x_2, but because of noise and numerical errors, they don't meet exactly



Triangulation: Geometric Approach

- Find shortest segment connecting the two viewing rays and let X be the midpoint of that segment.



Triangulation: Linear Approach

$$\lambda_1 \mathrm{x}_1 = \mathrm{P}_1 \mathrm{X} \qquad \mathrm{x}_1 \times \mathrm{P}_1 \mathrm{X} = 0 \qquad [\mathrm{x}_{1\times}]\mathrm{P}_1 \mathrm{X} = 0$$

$$\lambda_2 \mathrm{x}_2 = \mathrm{P}_2 \mathrm{X} \qquad \mathrm{x}_2 \times \mathrm{P}_2 \mathrm{X} = 0 \qquad [\mathrm{x}_{2\times}]\mathrm{P}_2 \mathrm{X} = 0$$

Two independent equations each in terms of three unknown entries of X

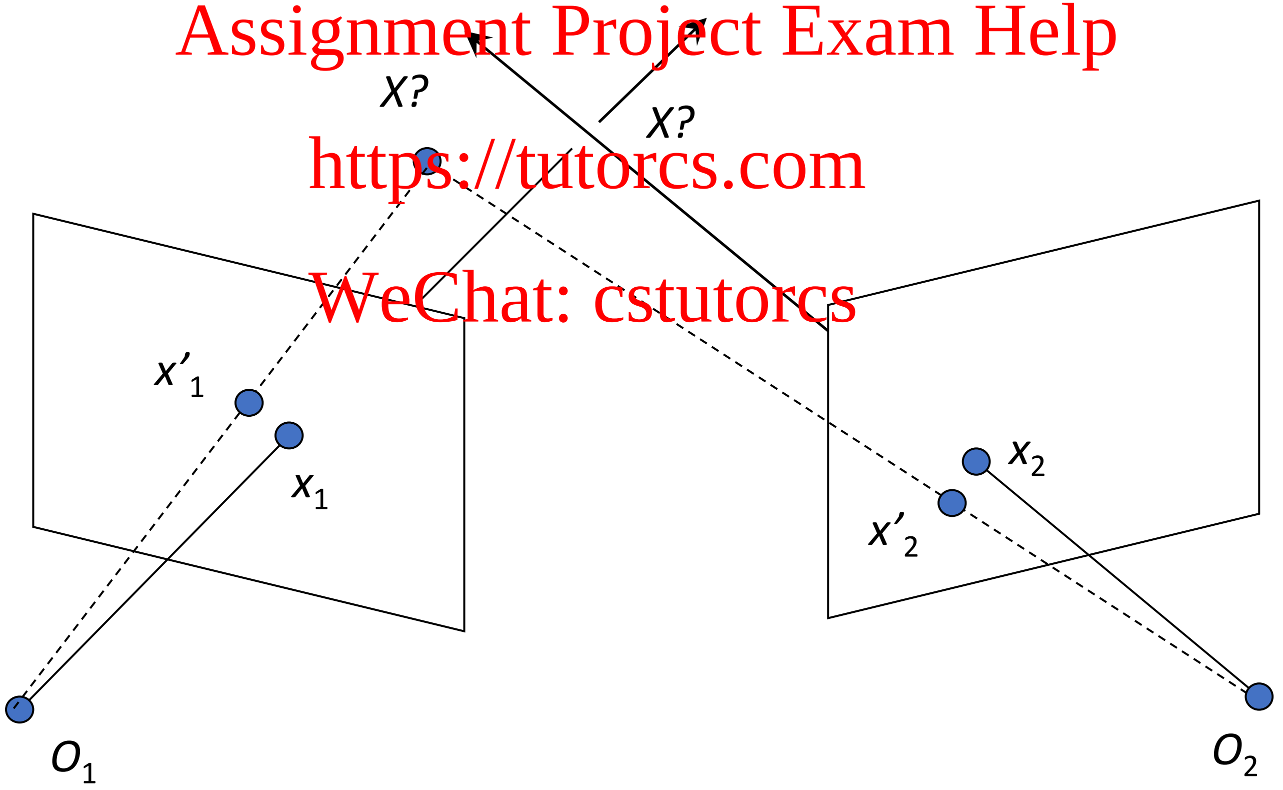
Cross product as matrix multiplication:

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

Triangulation: Non-Linear Approach

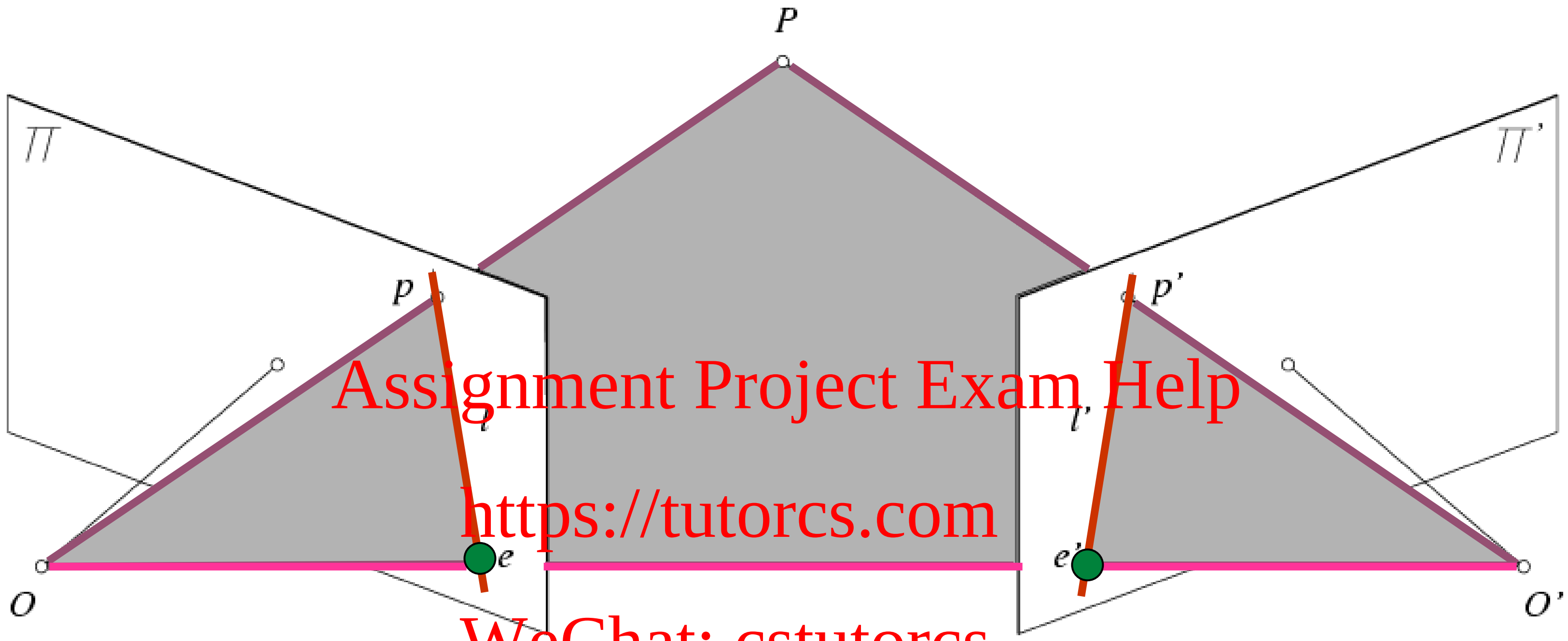
- Find X that minimizes

$$d^2(x_1, P_1 X) + d^2(x_2, P_2 X)$$



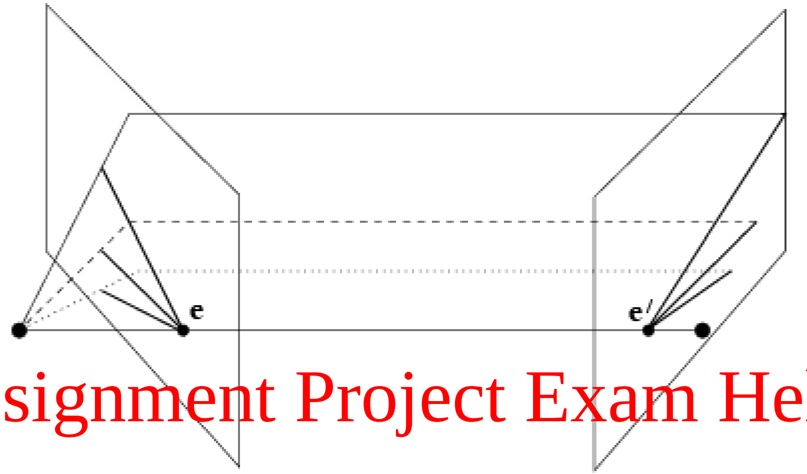
Two-view Geometry

Epipolar Geometry

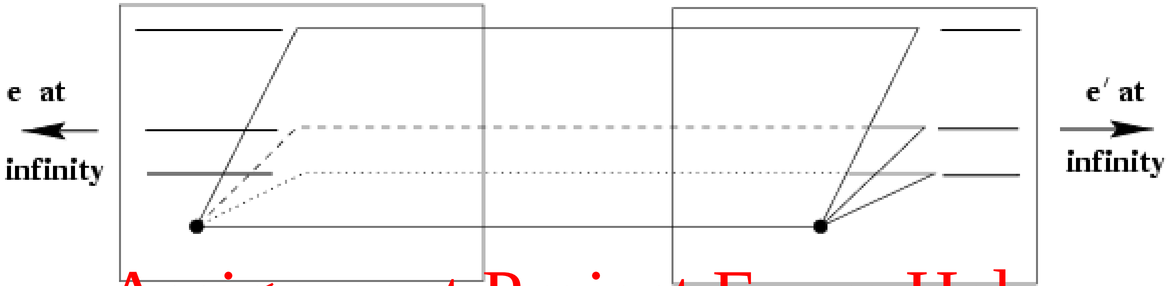


- Baseline – line connecting the two camera centres
- Epipolar Plane – plane containing baseline (1D family)
- Epipoles: intersections of baseline with image planes; projections of the other camera centres
- Epipolar Lines: intersections of epipolar plane with image planes (always come in corresponding pairs)

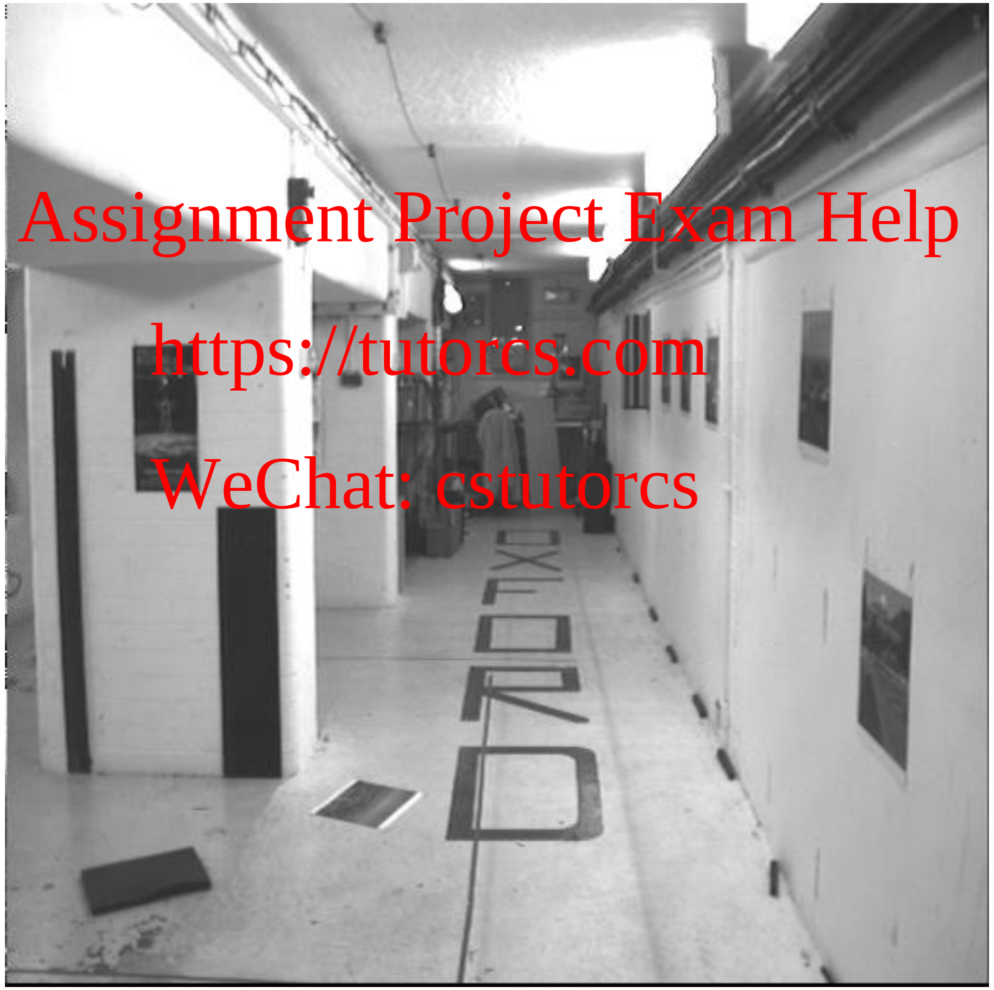
Example: Converging Cameras



Example: Motion Parallel to Image Plane

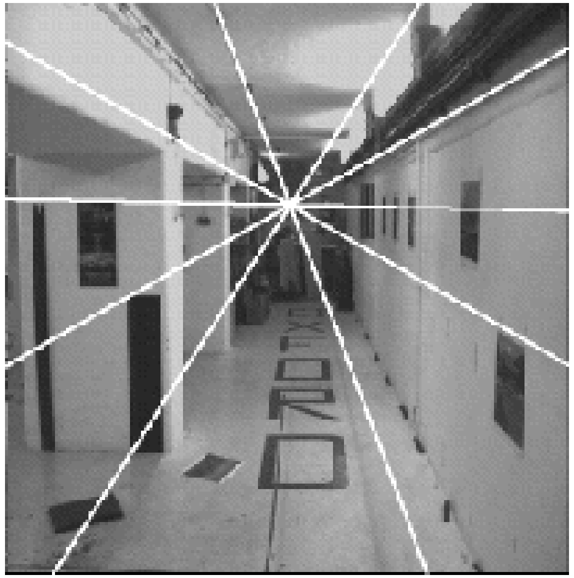


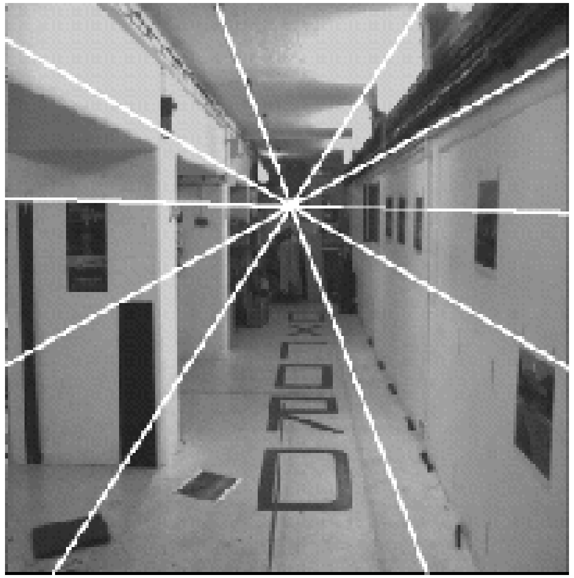
Example: Motion Perpendicular to Image Plane

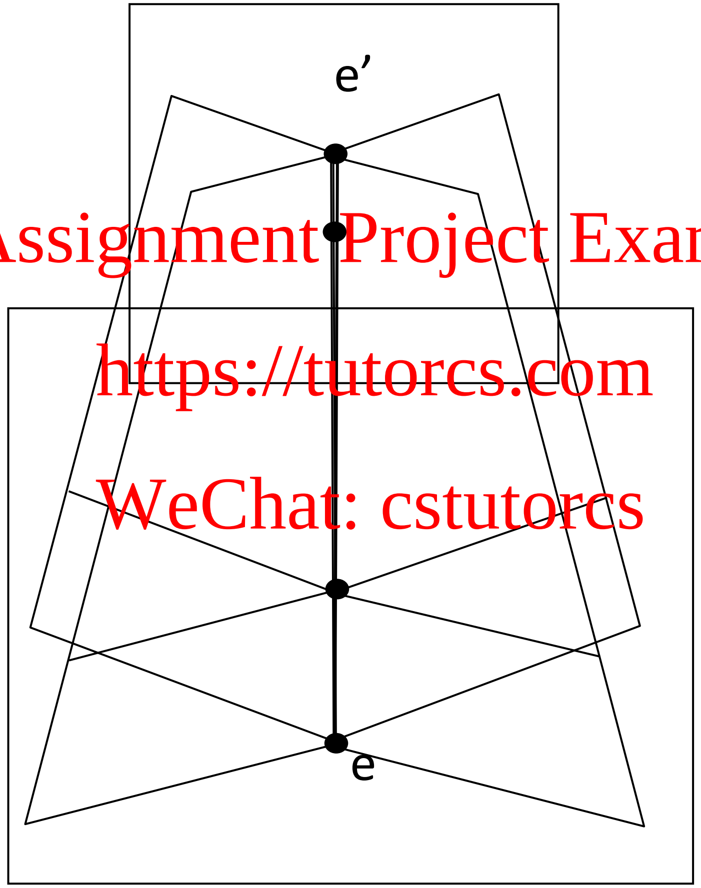


Example: Motion Perpendicular to Image Plane

Example: Motion Perpendicular to Image Plane

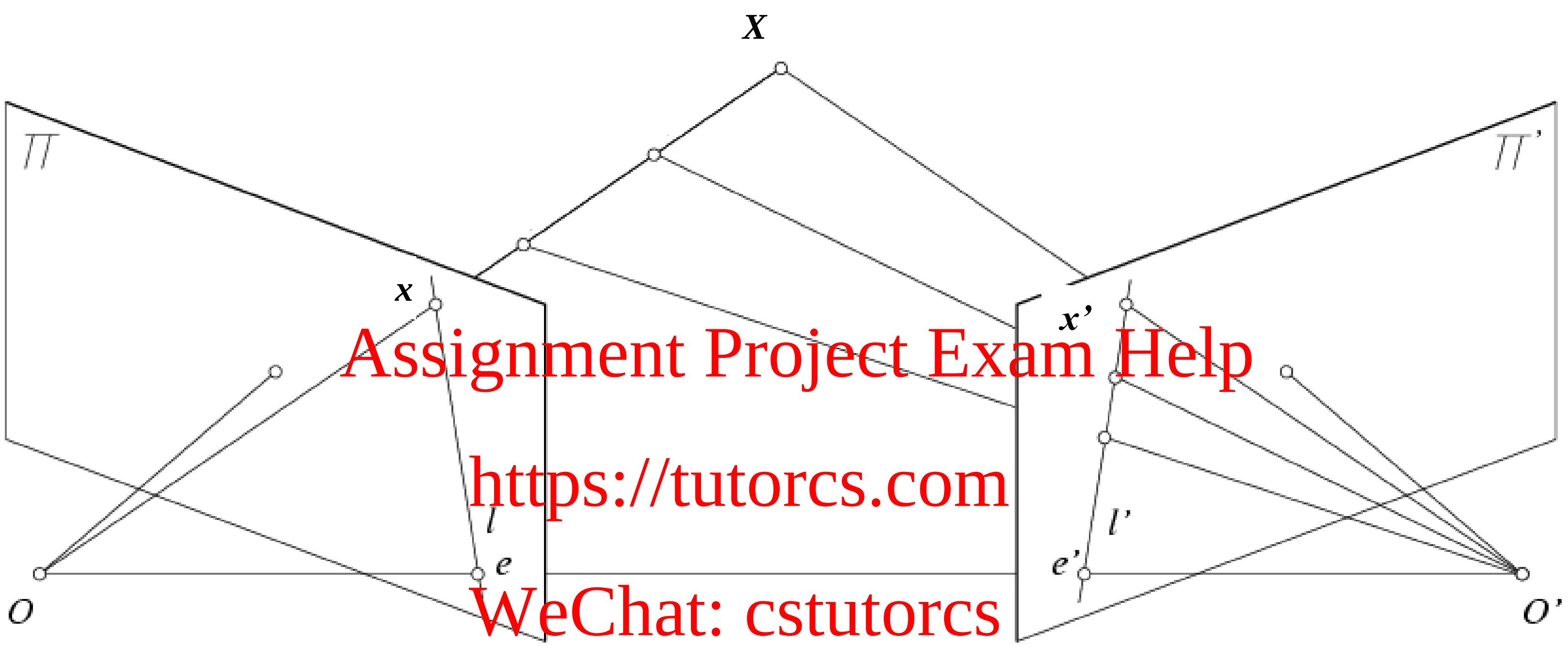






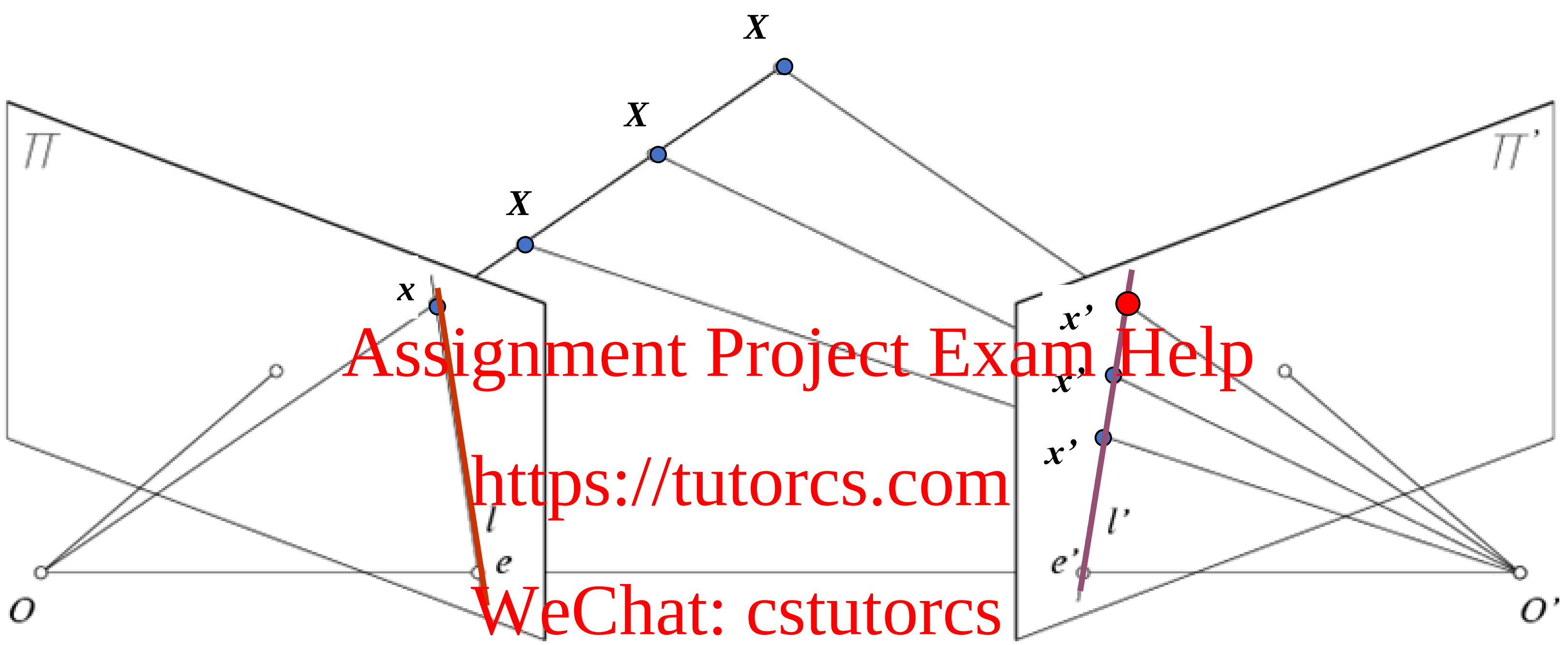
- Epipoles have same coordinates in both images.
- Points move along lines radiating from e: “focus of expansion”

Epipolar Constraint



- If we observe a point *x* in one image, where can the corresponding point *x'* be in the other image?

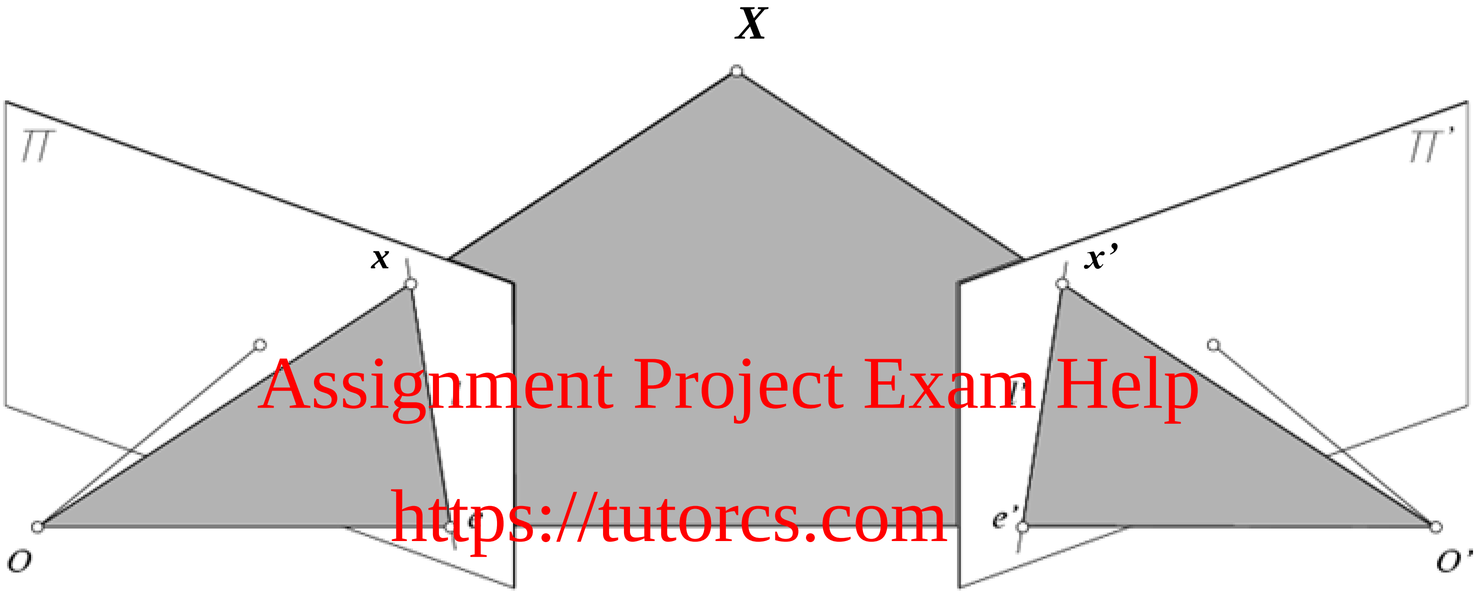
Epipolar Constraint



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

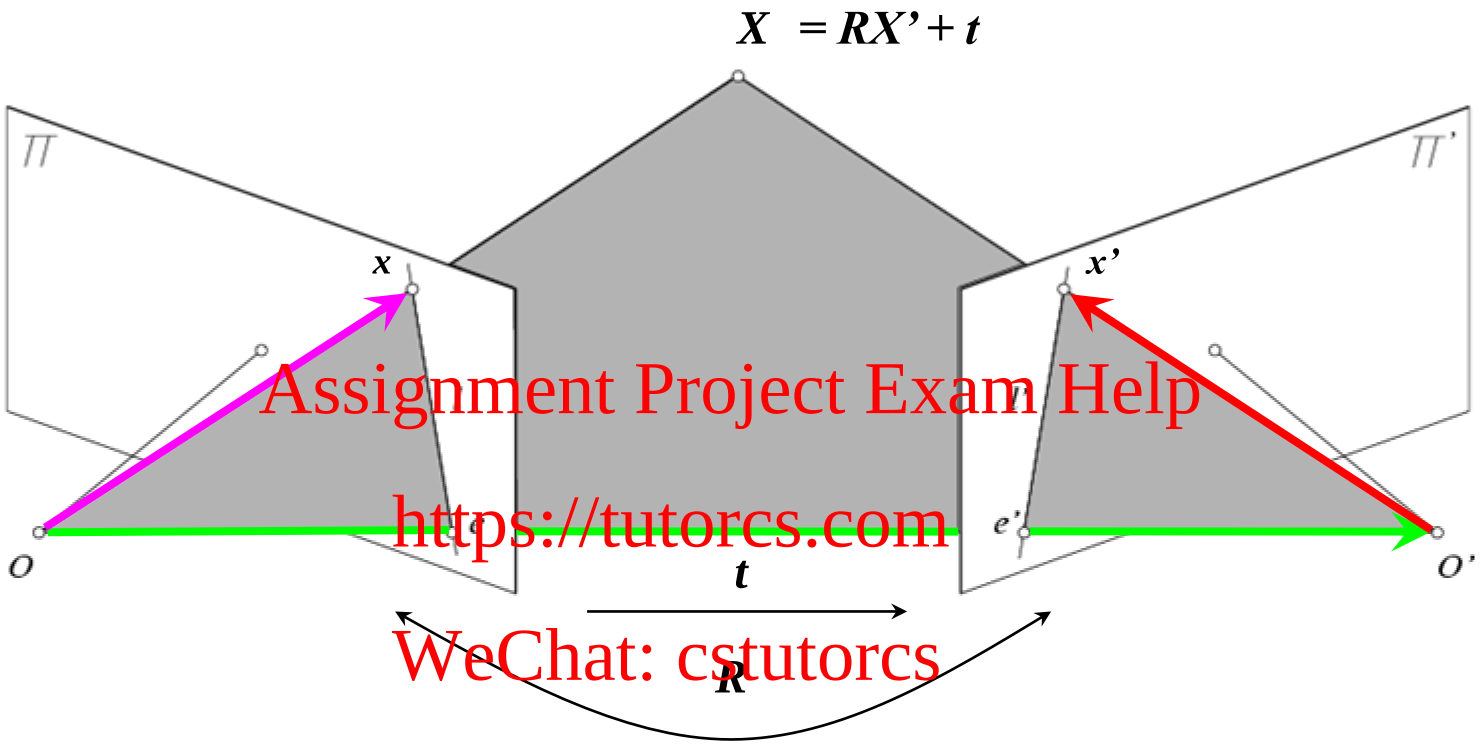
Epipolar Constraint Example

Epipolar Constraint: Calibrated Case



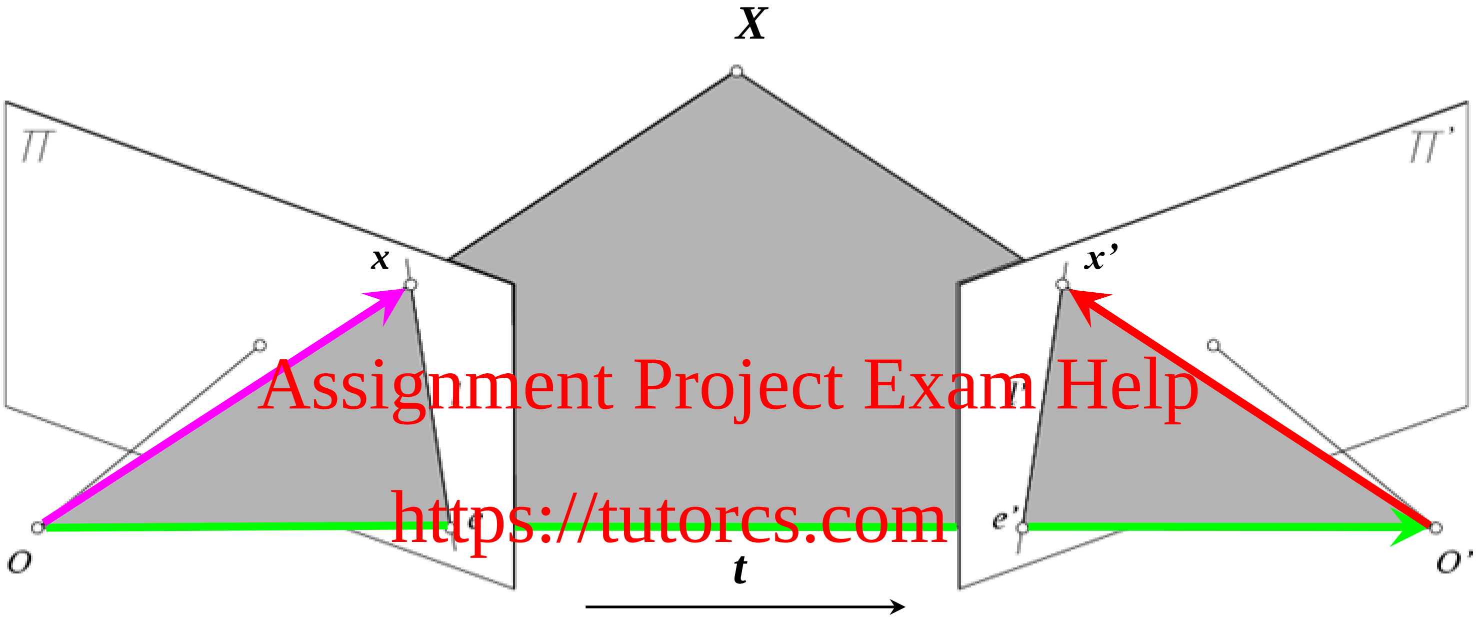
- Assume that the intrinsic and extrinsic parameters of the cameras are known
- We can multiply the projection matrix of each camera (and the image points) by the inverse of the calibration matrix to get *normalised* image coordinates.
- We can also set the global coordinate system to the coordinate system of the first camera. Then the projection matrix of the first camera is **[I | 0]**.

Epipolar Constraint: Calibrated Case



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

Epipolar Constraint: Calibrated Case

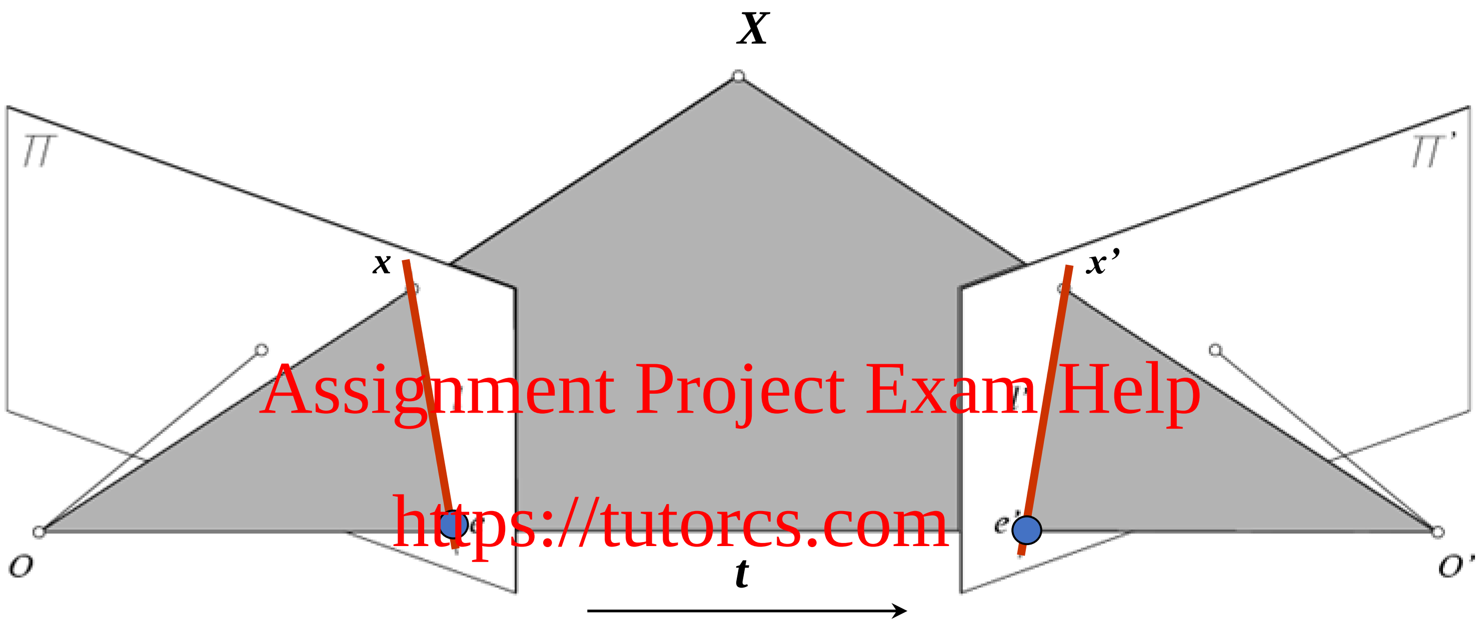


$$x \cdot [t \times (Rx')] = 0 \quad \Rightarrow \quad x^T E x' = 0 \quad \text{with} \quad E = [t_\times]R$$

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

Essential Matrix
(Longuet-Higgins, 1981)

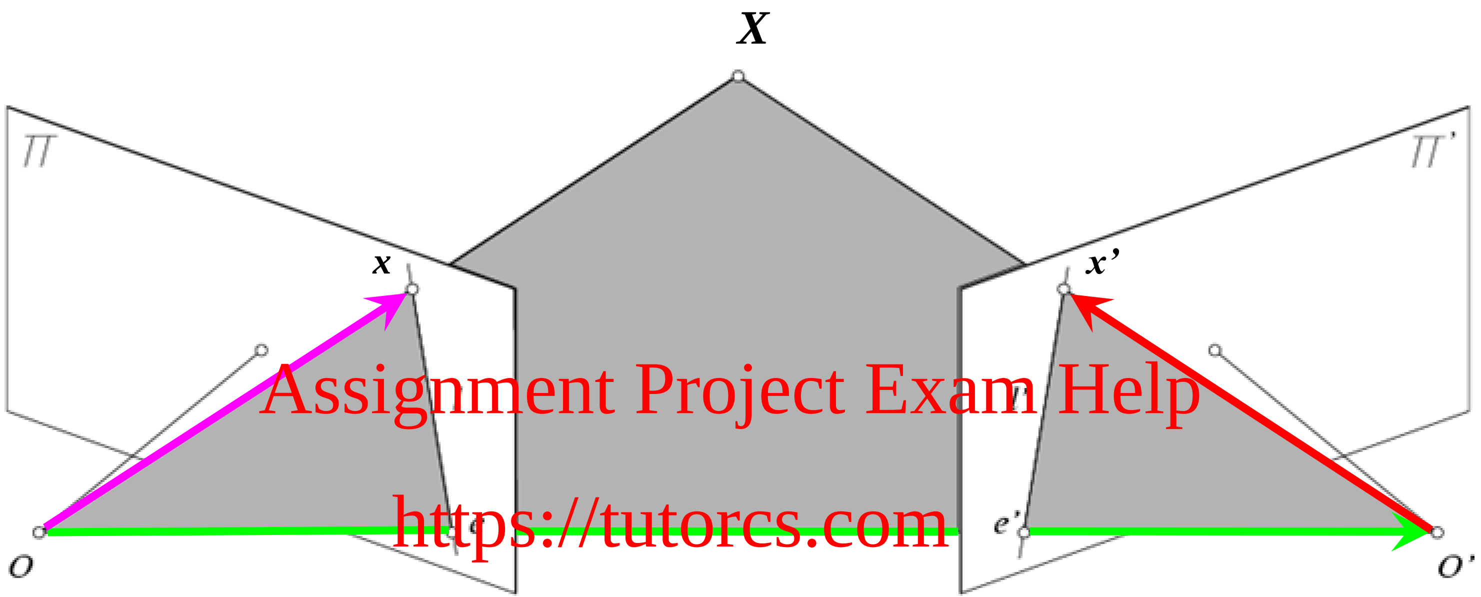
Epipolar Constraint: Calibrated Case



$$x \cdot [t \times (Rx')] = 0 \quad \Rightarrow \quad x^T E x' = 0 \quad \text{with} \quad E = [t_\times]R$$

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

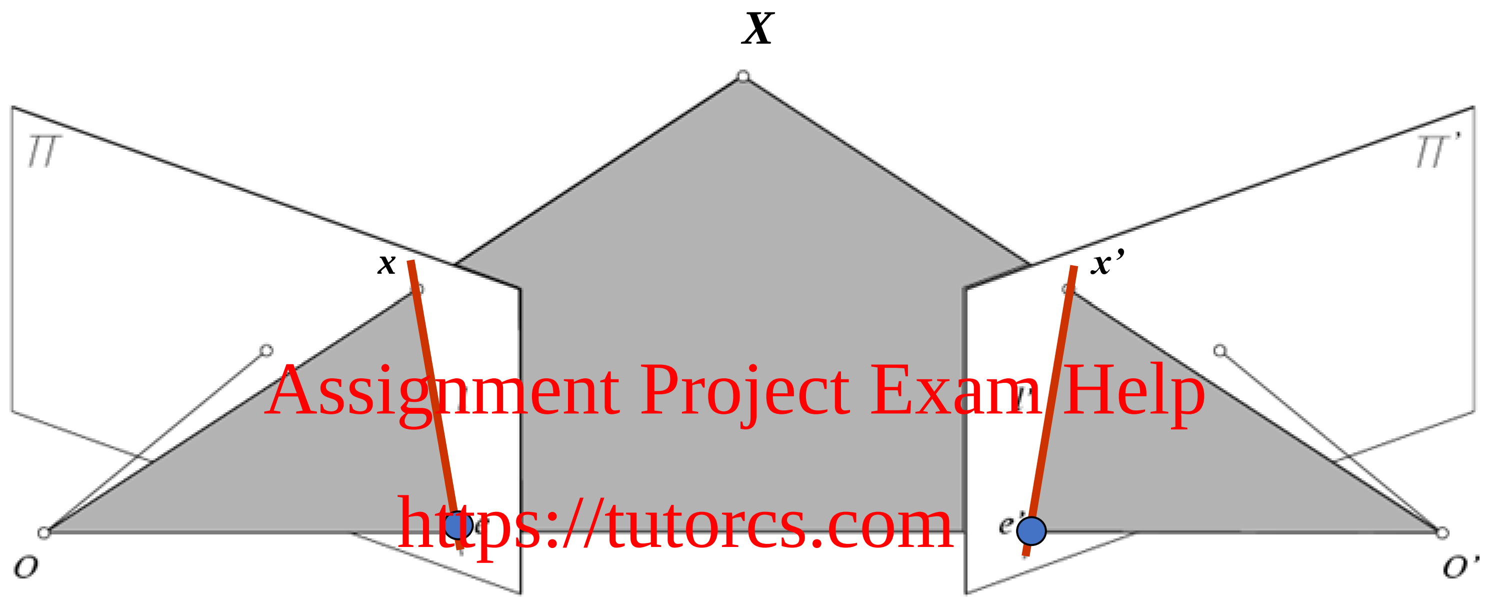
Epipolar Constraint: Uncalibrated Case



- The calibration matrices *K* and *K'* of the two cameras are unknown
- We can write the epipolar constraint in terms of *unknown* normalized coordinates

$$\hat{x}^T E \hat{x}' = 0 \qquad\qquad x = K\hat{x}, \quad x' = K'\hat{x}'$$

Epipolar Constraint: Uncalibrated Case



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

- Fx' is the epipolar line associated with x' ($l = Fx'$)
- $F^T x$ is the epipolar line associated with x ($l' = F^T x$)
- $Fe' = 0$ and $F^T e = 0$
- F is singular (rank two), and F has seven degrees of freedom

The Eight-point Algorithm

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

$$(u, v, 1)\begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0$$

$$(uu', uv', u, vu', vv', v, u', v', 1)\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = 0$$

$$\begin{pmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' \\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' \\ u_3u_3' & u_3v_3' & u_3 & v_3u_3' & v_3v_3' & v_3 & u_3' & v_3' \\ u_4u_4' & u_4v_4' & u_4 & v_4u_4' & v_4v_4' & v_4 & u_4' & v_4' \\ u_5u_5' & u_5v_5' & u_5 & v_5u_5' & v_5v_5' & v_5 & u_5' & v_5' \\ u_6u_6' & u_6v_6' & u_6 & v_6u_6' & v_6v_6' & v_6 & u_6' & v_6' \\ u_7u_7' & u_7v_7' & u_7 & v_7u_7' & v_7v_7' & v_7 & u_7' & v_7' \\ u_8u_8' & u_8v_8' & u_8 & v_8u_8' & v_8v_8' & v_8 & u_8' & v_8' \end{pmatrix}\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix} = -\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Minimize:

$$\sum_{i=1}^{N} (x_i^T F x_i')^2$$

under the constraint

$F_{33} = 1$

The Eight-point Algorithm

- Meaning of error $$\sum_{i=1}^{N}(x_i^T F x_i')^2$$

  Sum of Euclidean distances between points x_i and epipolar line Fx_i' (or points x_i' and epipolar line F^Tx_i) multiplied by a scale factor

- Non-linear approach: minimize

$$\sum_{i=1}^{N}\left[\mathrm{d}^2(x_i, F x_i') + \mathrm{d}^2(x_i', F^T x_i)\right]$$

Problem with Eight-point Algorithm

$$
\begin{pmatrix}
u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' \\
u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' \\
u_3u_3' & u_3v_3' & u_3 & v_3u_3' & v_3v_3' & v_3 & u_3' & v_3' \\
u_4u_4' & u_4v_4' & u_4 & v_4u_4' & v_4v_4' & v_4 & u_4' & v_4' \\
u_5u_5' & u_5v_5' & u_5 & v_5u_5' & v_5v_5' & v_5 & u_5' & v_5' \\
u_6u_6' & u_6v_6' & u_6 & v_6u_6' & v_6v_6' & v_6 & u_6' & v_6' \\
u_7u_7' & u_7v_7' & u_7 & v_7u_7' & v_7v_7' & v_7 & u_7' & v_7' \\
u_8u_8' & u_8v_8' & u_8 & v_8u_8' & v_8v_8' & v_8 & u_8' & v_8'
\end{pmatrix}
\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix}
= -
\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}
$$

Problem with Eight-point Algorithm

$$
\begin{bmatrix}
250906.36 & 183269.57 & 921.81 & 200931.10 & 146766.13 & 738.21 & 272.19 & 198.81 \\
2692.28 & 131633.03 & 176.27 & 6196.73 & 302975.59 & 405.71 & 15.27 & 746.79 \\
416374.23 & 871684.30 & 935.47 & 408110.89 & 854384.92 & 916.90 & 445.10 & 931.81 \\
191183.60 & 171759.40 & 410.27 & 416435.62 & 374125.90 & 893.65 & 465.99 & 418.65 \\
48988.86 & 30401.76 & 57.89 & 298604.57 & 185309.58 & 352.87 & 846.22 & 525.15 \\
164786.04 & 546559.67 & 813.17 & 1998.37 & 6628.15 & 9.86 & 202.65 & 672.14 \\
116407.01 & 2727.75 & 138.89 & 169941.27 & 3982.21 & 202.77 & 838.12 & 19.64 \\
135384.58 & 75411.13 & 198.72 & 411350.03 & 229127.78 & 603.79 & 681.28 & 379.48
\end{bmatrix}
\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix}
= -\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}
$$

- Poor numerical conditioning
- Can be fixed by rescaling the data

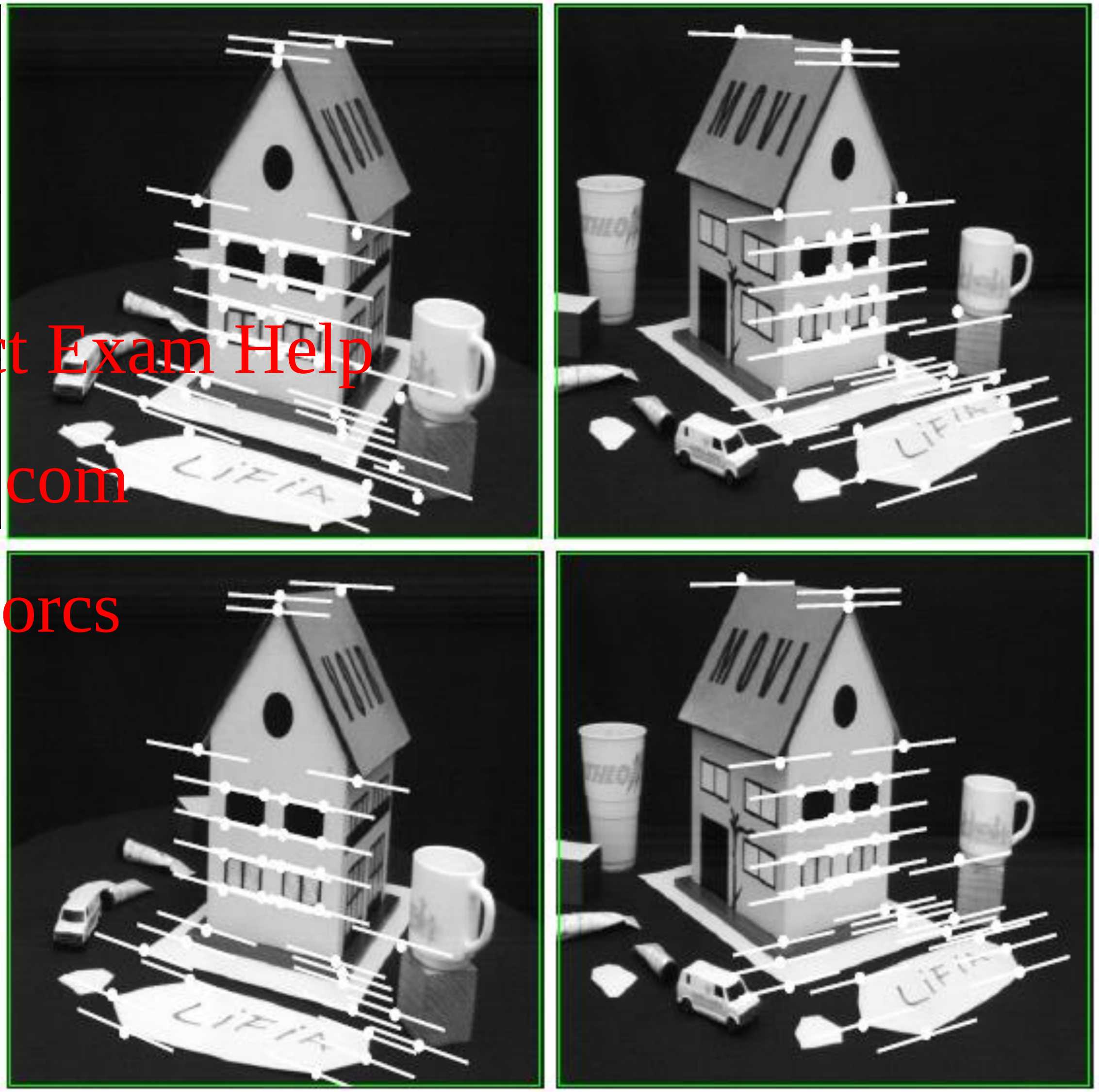
The Normalized Eight-point Algorithm

(Hartley, 1995)

- Centre the image data at the origin, and scale it so that the mean squared distance between the origin and the data points is 2 pixels
- Use the eight-point algorithm to compute *F* from the normalized points
- Enforce the rank 2 constraint (for example, take SVD of *F* and throw out the smallest singular value)
- Transform fundamental matrix back to the original units: if *T* and *T'* are the normalized transformations in the two images, then the fundamental matrix in the original coordinates is $T^T F T'$

Comparison of Estimation Algorithms

	Av. Dist. 1	Av. Dist. 2
8-point	2.33 pixels	2.18 pixels
Normalized 8-point	0.92 pixel	0.85 pixel
Nonlinear least squares	0.86 pixel	0.80 pixel



From Epipolar Geometry to Camera Calibration

- Estimating the fundamental matrix is known as "weak" calibration
- If we know the calibration matrices of the two cameras, we can estimate the essential matrix: $E = K^T F K'$
- The essential matrix gives us the relative rotation and translation between the cameras, or their extrinsic parameters

Summary

- What is the problem of stereo vision?
- What is baseline? What are epipole, epipolar line, and epipolar plane? How to determine epipolar lines?
- What is essential matrix? What is fundamental matrix?
- Describe eight-point algorithm