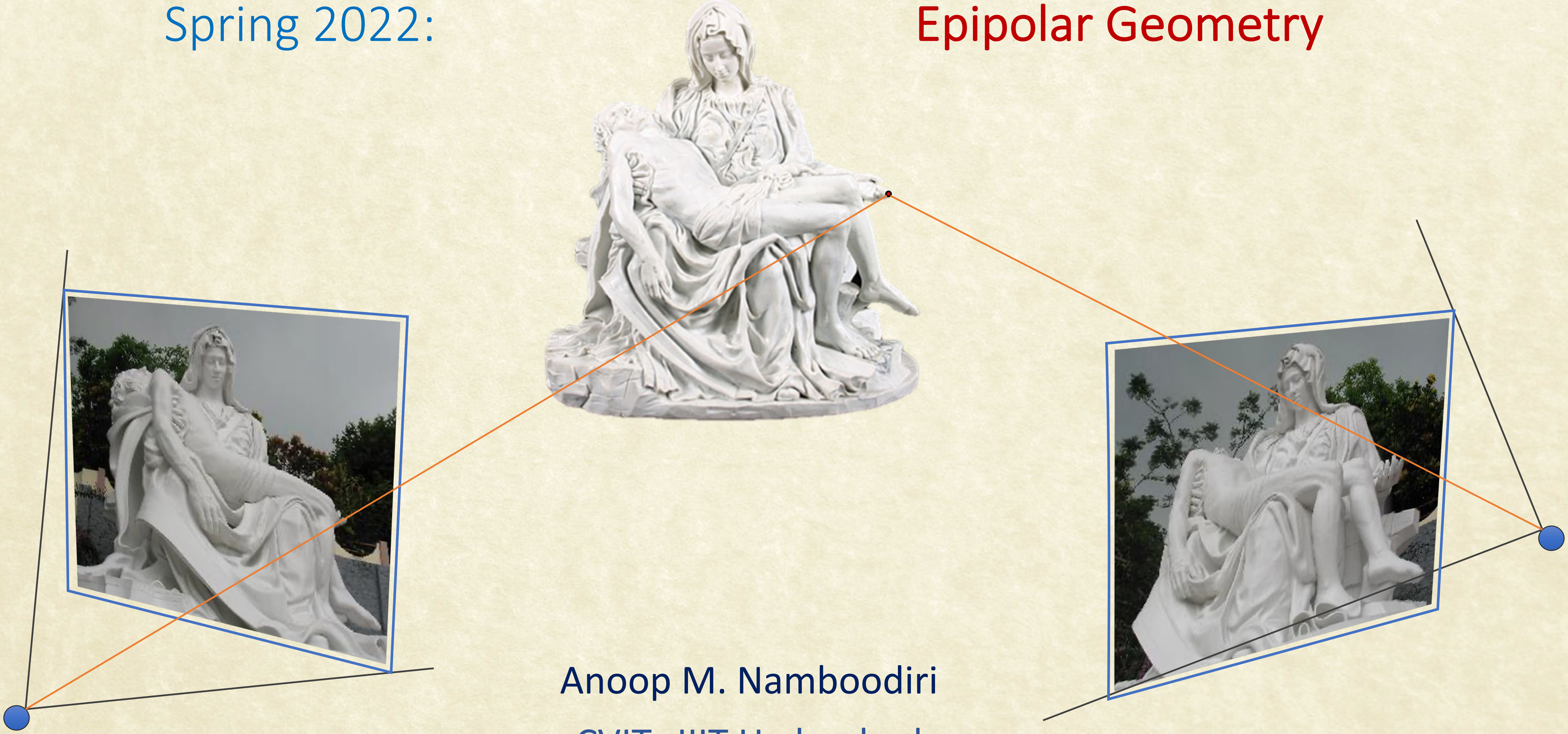
CS7.505: Computer Vision

Spring 2022: Epipolar Geometry



Anoop M. Namboodiri

CVIT, IIIT Hyderabad

Cross Product: A Recap

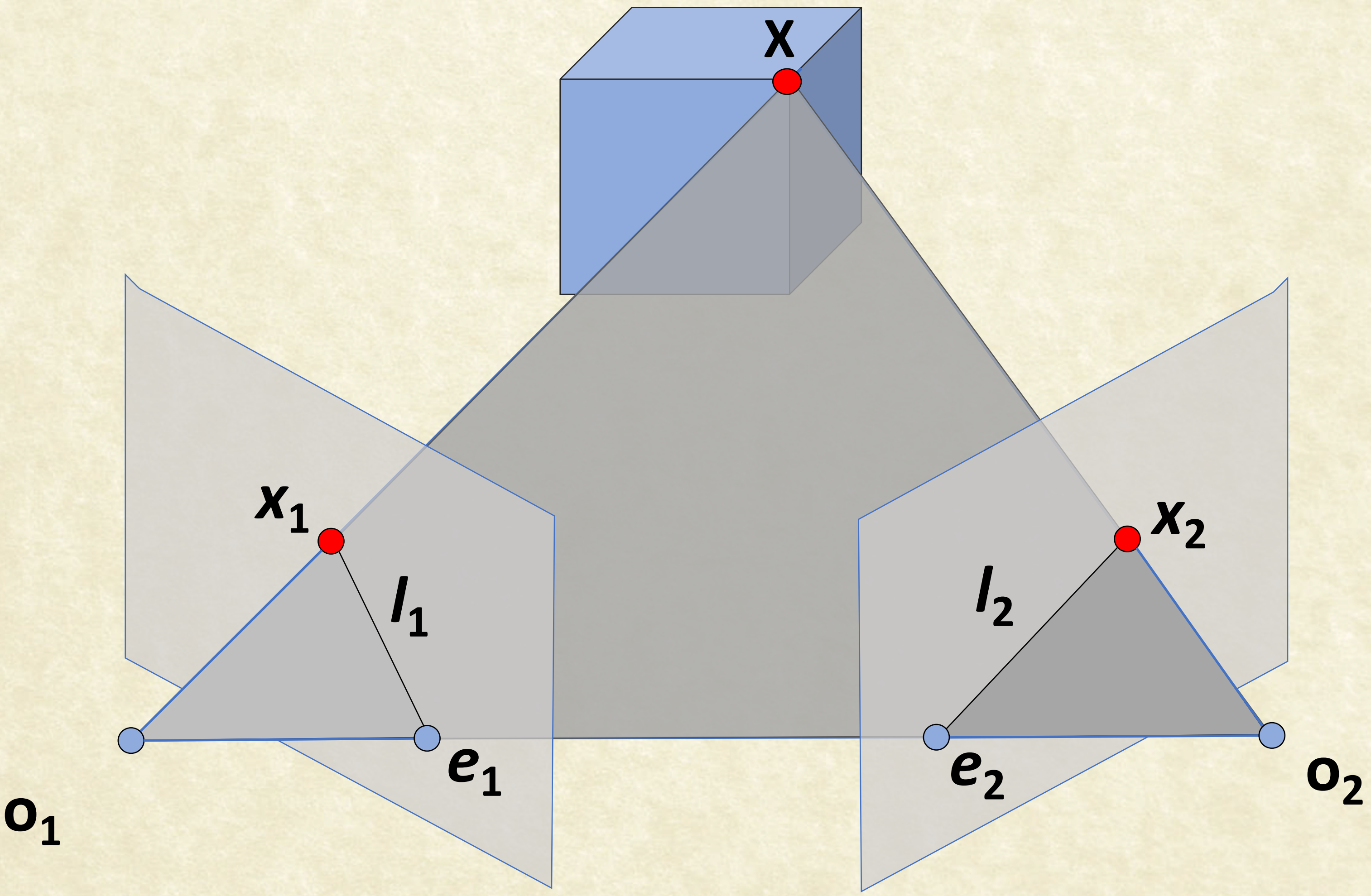
- Consider $\mathbf{A} = \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix}$; and Let $\widehat{\mathbf{A}} = \begin{bmatrix} 0 & -A_z & A_y \\ A_z & 0 & -A_x \\ -A_y & A_x & 0 \end{bmatrix}$

- $\mathbf{A} \times \mathbf{B} = \begin{vmatrix} i & j & k \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} = \begin{bmatrix} A_yB_z - B_yA_z \\ B_xA_z - A_xB_z \\ A_xB_y - B_xA_y \end{bmatrix}$

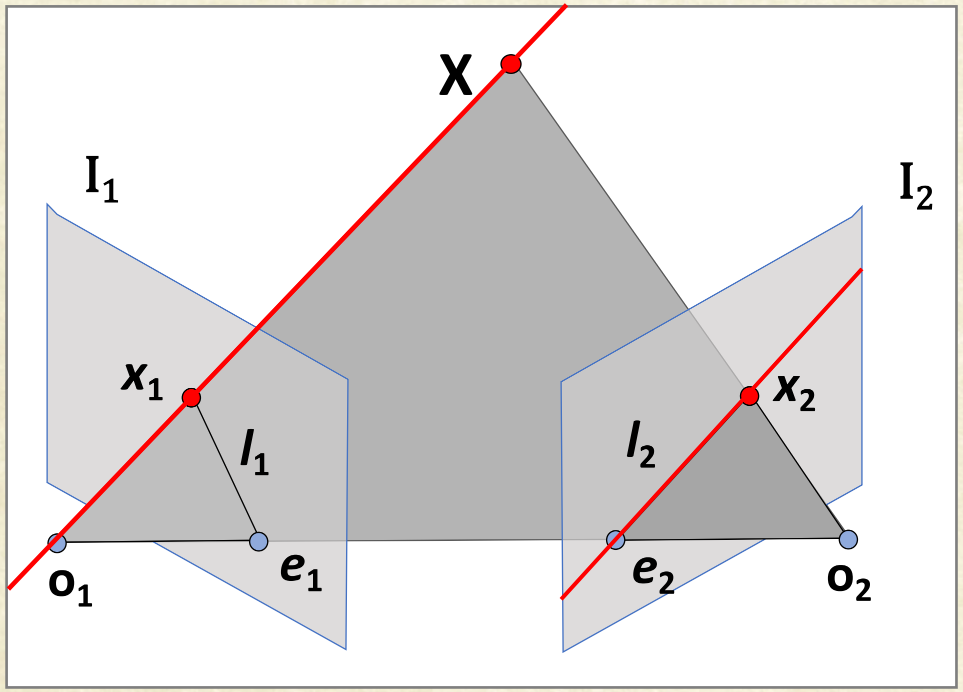
- $\widehat{\mathbf{A}}\mathbf{B} = \begin{bmatrix} 0 & -A_z & A_y \\ A_z & 0 & -A_x \\ -A_y & A_x & 0 \end{bmatrix} \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = \begin{bmatrix} A_yB_z - B_yA_z \\ B_xA_z - A_xB_z \\ A_xB_y - B_xA_y \end{bmatrix}$

- Note: The cross product, AxB or ÂB is a vector perpendicular to both A and B

Case 3: Generic World and Cameras



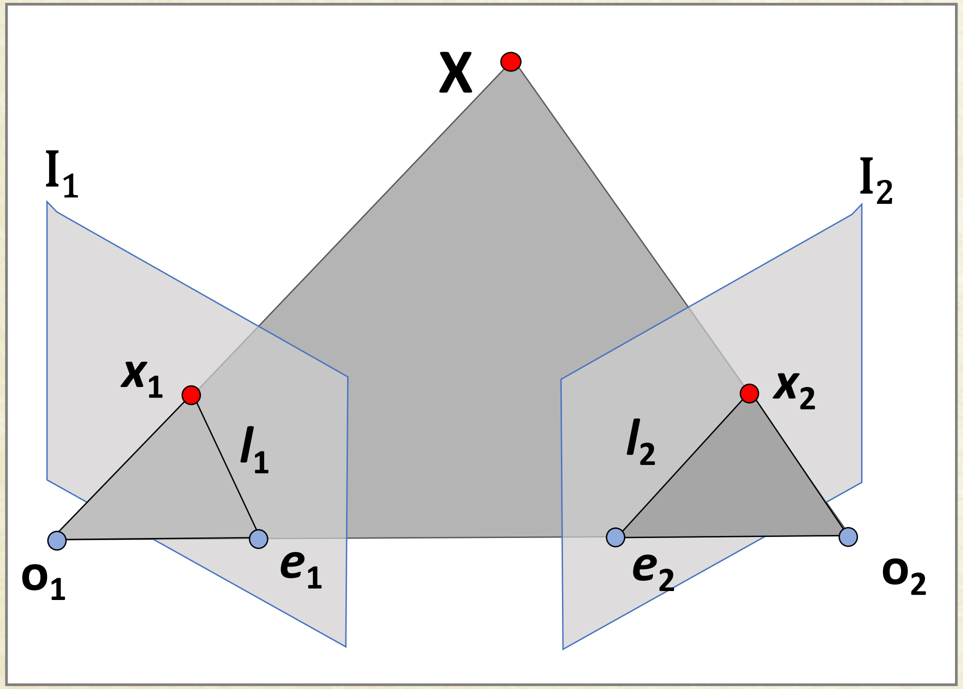
Epipolar Geometry



- All world points that map to $\boldsymbol{x}_1$ in $\mathbf{I}_1$ (pre-image of $\boldsymbol{x}_1$) map to a line $\boldsymbol{l}_2$ in $\mathbf{I}_2$, called an epipolar line. (so is $\boldsymbol{l}_1$)
- The image of $\boldsymbol{o}_1$ in $\mathbf{I}_2$ ($\boldsymbol{e}_2$) is an epipole. (so is $\boldsymbol{e}_1$)

- The plane containing these is called the epipolar plane.
- These result in a set of constraints, which are referred to as the *epipolar constraints* and the resulting geometry is called the *epipolar geometry.*

Epipolar Constraint: Essential Matrix



- Consider $\mathbf{X}$ in camera 1's coordinates:
  - $\lambda_1 \boldsymbol{x}_1 = \mathbf{X}$
- Now, viewing it in camera 2's coordinates:
- $\lambda_2 \boldsymbol{x}_2 = \mathbf{RX} + \mathbf{T}$

$$= \mathbf{R}(\lambda_1 \boldsymbol{x}_1) + \mathbf{T}$$

Pre-multiplying by $\widehat{\mathbf{T}}$, and then by $\boldsymbol{x}_2{}^T$, we get

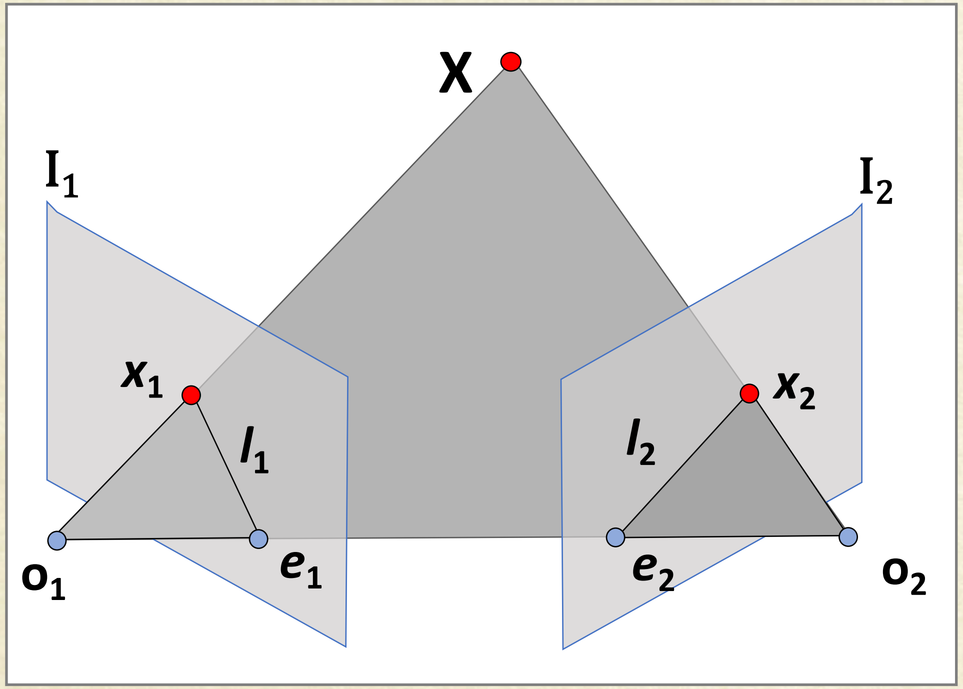
$$\widehat{\mathbf{T}}\lambda_2 \boldsymbol{x}_2 = \widehat{\mathbf{T}}\mathbf{R}\lambda_1 \boldsymbol{x}_1 + \cancel{\widehat{\mathbf{T}}\mathbf{T}}$$

$$\lambda_2 \cancel{\boldsymbol{x}_2{}^T \widehat{\mathbf{T}} \boldsymbol{x}_2} = \lambda_1 \boldsymbol{x}_2{}^T \widehat{\mathbf{T}}\mathbf{R}\boldsymbol{x}_1 + 0$$

$$\boldsymbol{x}_2{}^T \widehat{\mathbf{T}}\mathbf{R}\boldsymbol{x}_1 = 0$$

$$\boldsymbol{x}_2{}^T \mathbf{E}\boldsymbol{x}_1 = 0 \quad \text{or} \quad \boldsymbol{x}_1{}^T \mathbf{E}\boldsymbol{x}_2 = 0$$

Epipolar Constraint: Fundamental Matrix



$$\mathbf{x}_1 = \mathbf{K}_1\mathbf{X}$$

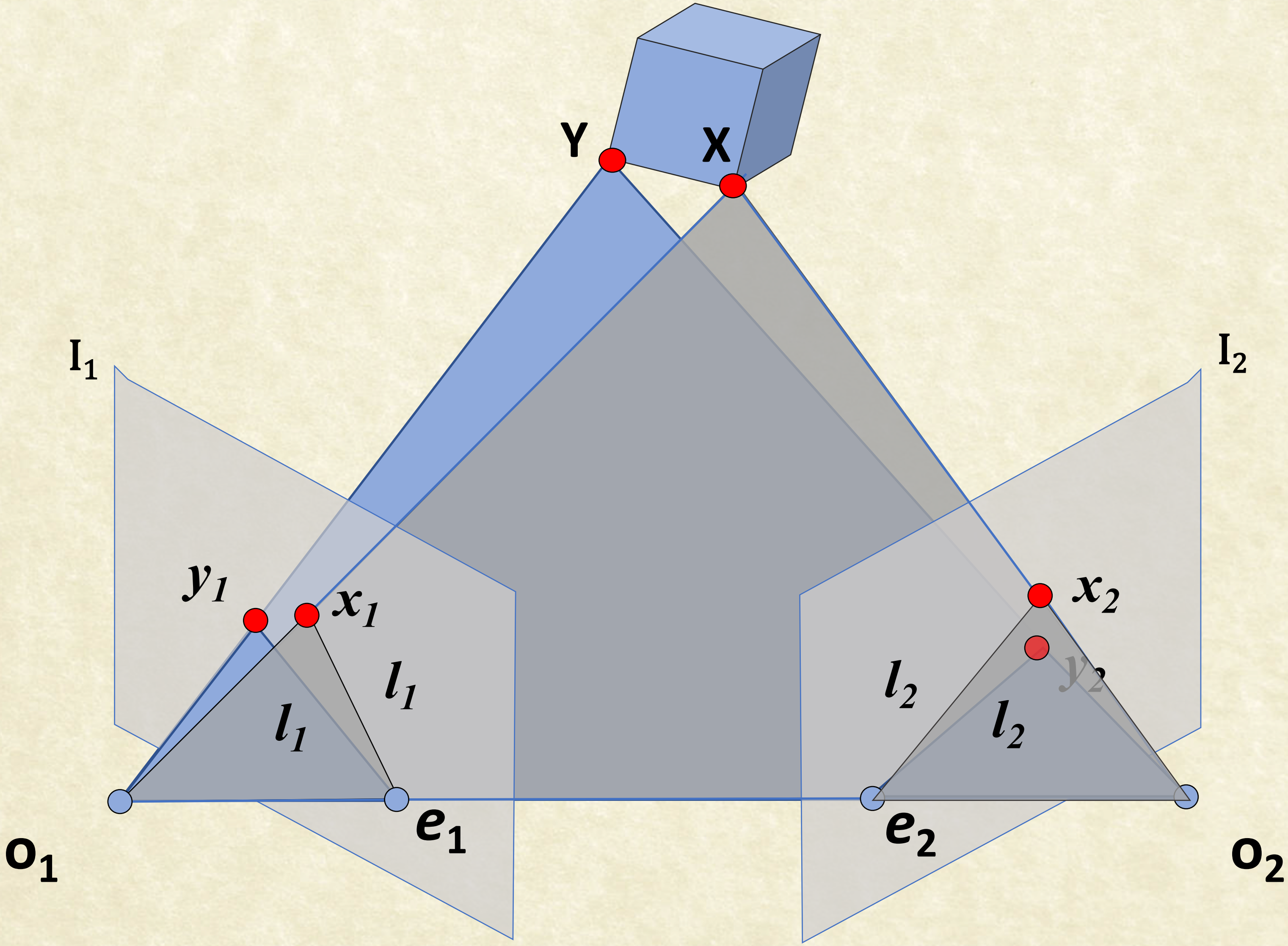
$$\mathbf{x}_2 = \mathbf{K}_2\mathbf{X}$$

$$\mathbf{x}_2^T\mathbf{K}_2^{-T}\hat{\mathbf{T}}\mathbf{R}\mathbf{K}_1^{-1}\mathbf{x}_1 = 0$$

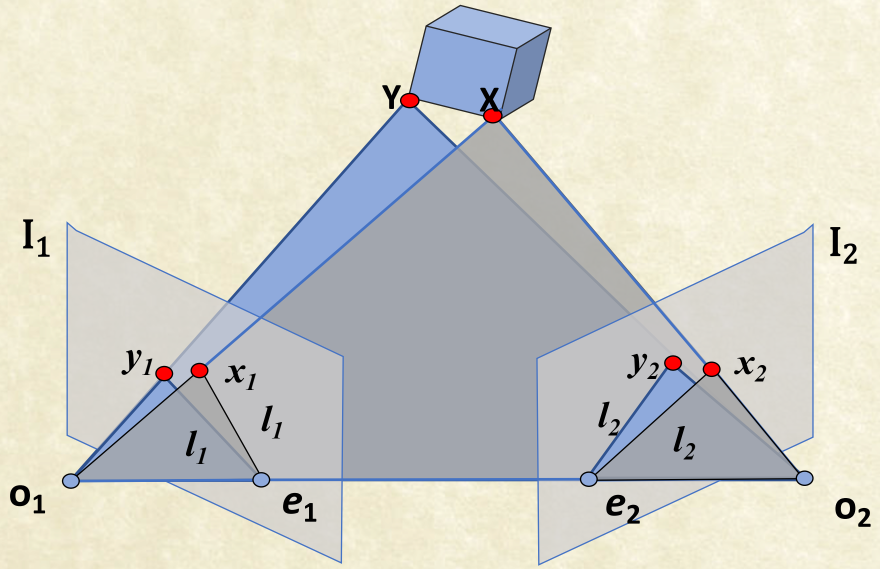
$$\mathbf{x}_1^T F\mathbf{x}_2 = 0$$

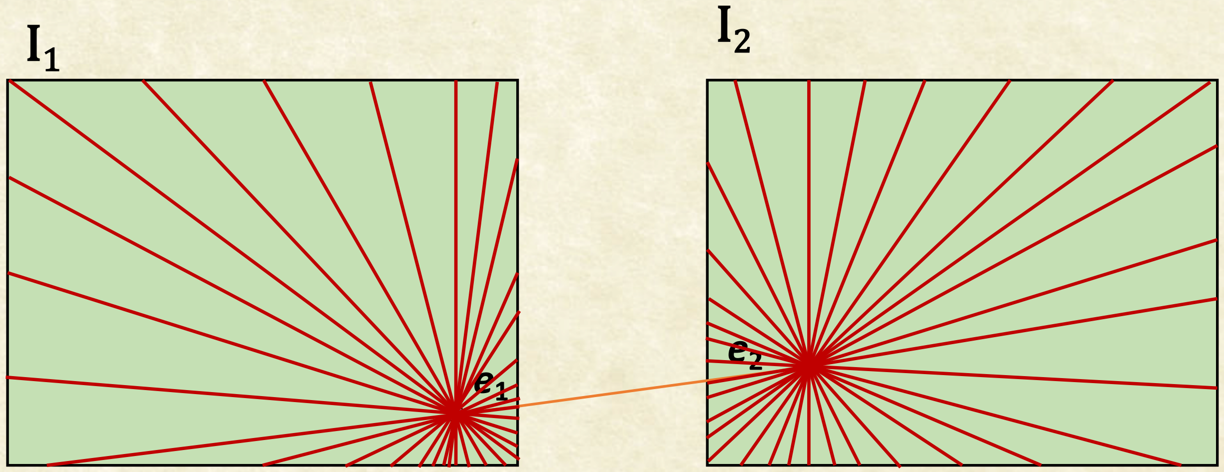
Both Essential and Fundamental matrices are 3x3 and are independent of the world point.

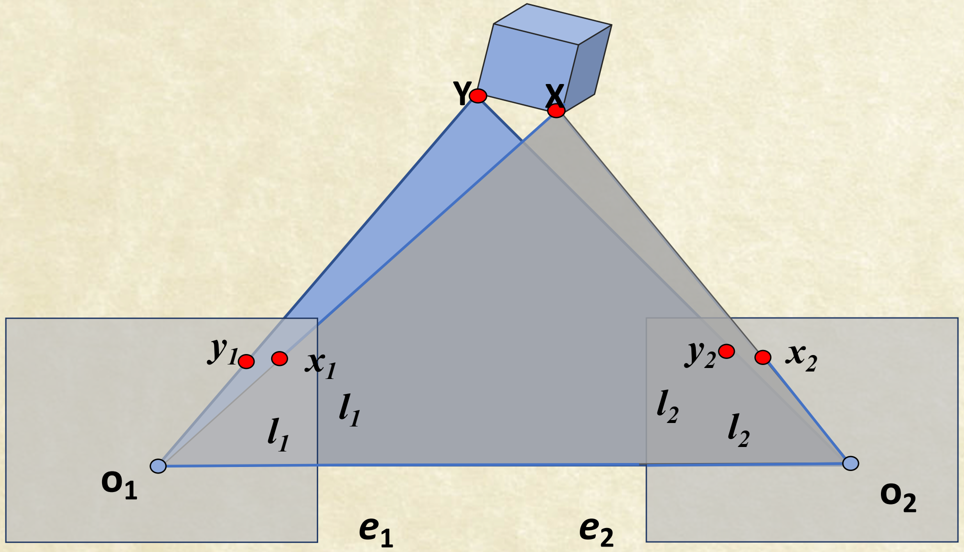
Epipolar Lines

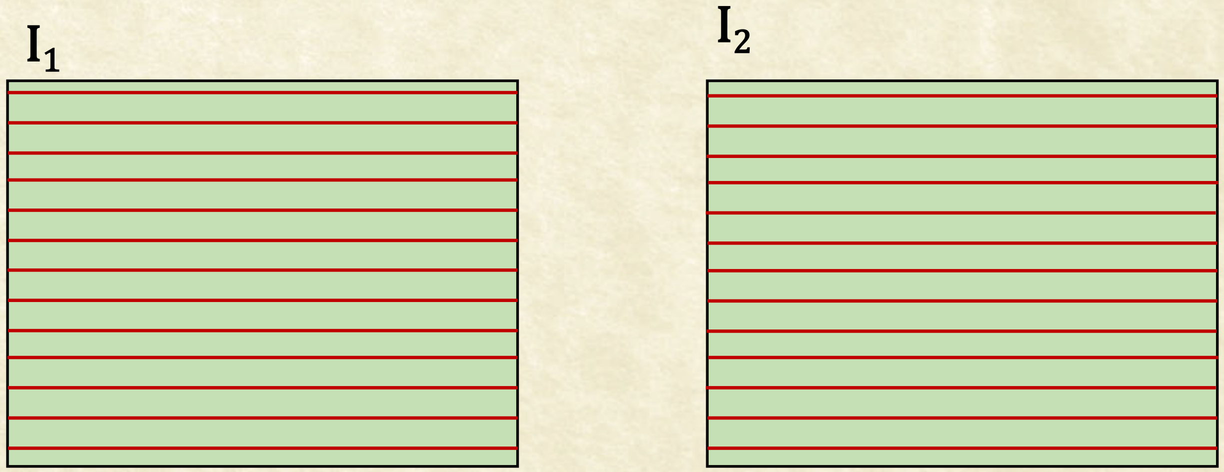


Epipolar Lines





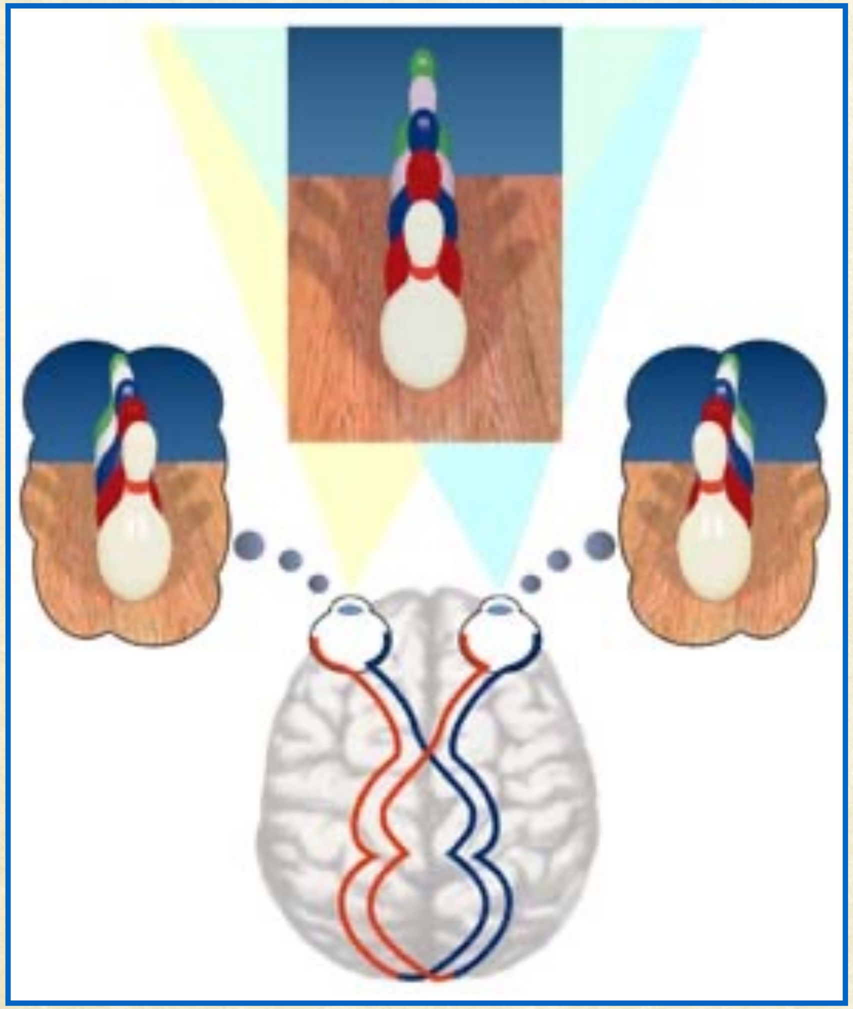




Summary

1. Two views of a planar world are related by a homography. (H: non-singular 3x3 matrix)
2. Two views or an arbitrary world from the same camera position are related by a homography.
3. Two views of an arbitrary world are constrained by epipolar geometry
   - Strong Calib.: Essential Matrix: $\boldsymbol{E}$; $x_1{}^T E x_2 = 0$
   - Weak Calib.: Fundamental Matrix: $\boldsymbol{F}$; $x_1{}^T F x_2 = 0$
4. All epipolar lines pass through the epipole in the image
   - The epipolar lines become parallel when the two image planes are coplanar

Stereo



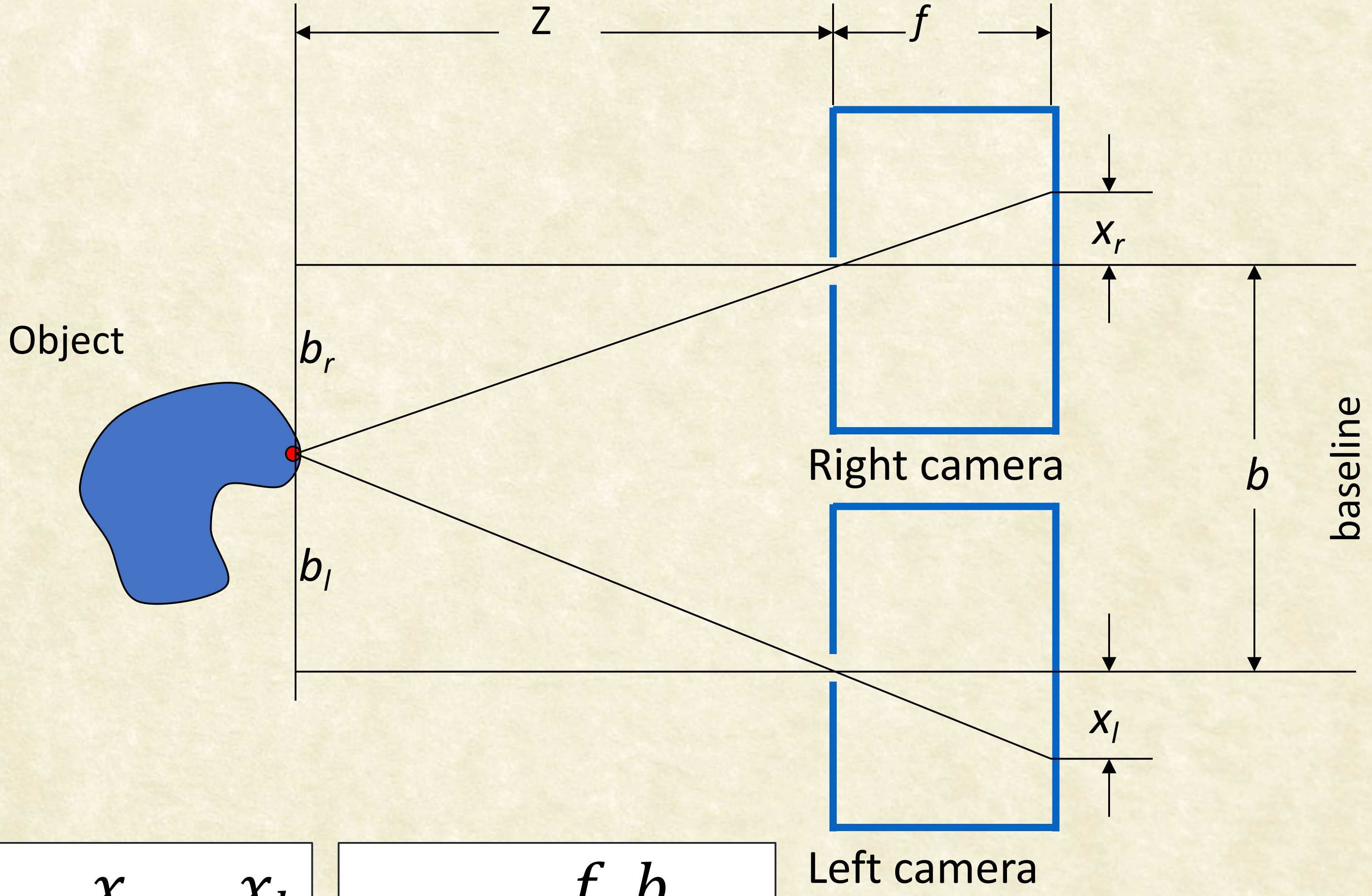
http://www.vision3d.com/stereo.html

1. We see a slightly different image of the world through the two eyes.

2. The shift in image is proportional to the the distance to the object.

3. Roadside trees are left behind, while the farther mountains follow you.

Stereo



$$\frac{b_r}{Z} + \frac{b_l}{Z} = \frac{x_r}{f} + \frac{x_l}{f}$$

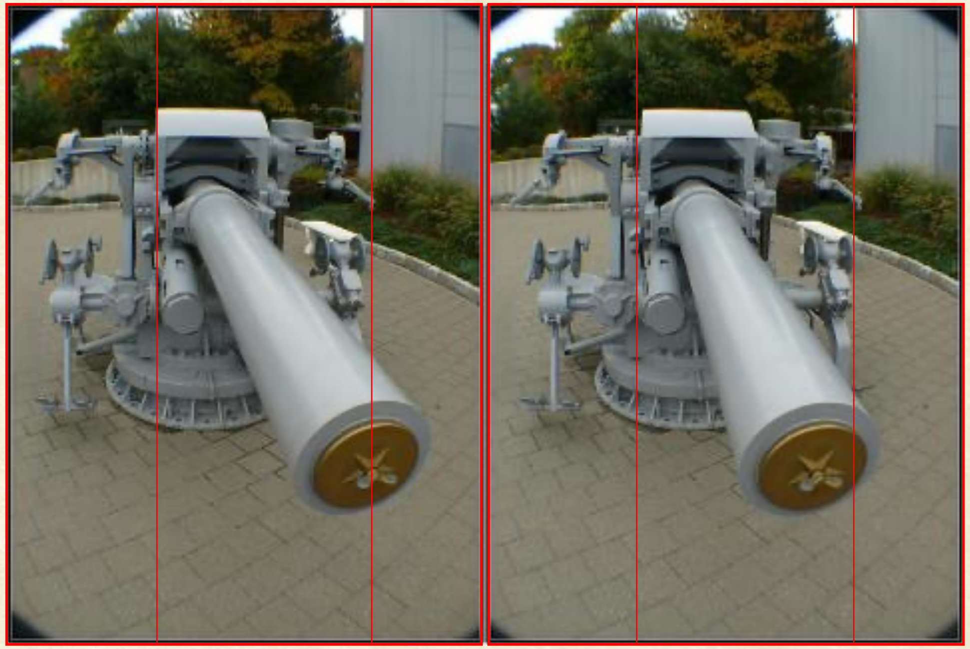
$$\mathrm{Z} = \frac{f \,.\, b}{(x_r + x_l)}$$

Stereo Geometry

- Farther the point, smaller the disparity and vice versa

- A large baseline can give more reliable estimates of depth. But, matching may become harder

- Basic step: Identify common points in the two camera views
  - How do we find the images of a single world point in two views?
  - Do they have the same intensity/color?
  - What else can be same/similar?



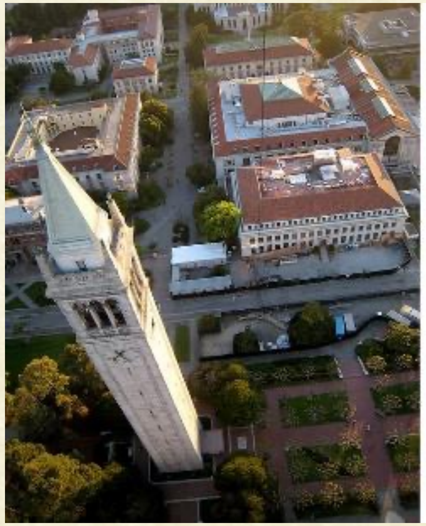
Identifying Common Points

- Find a world point in 2 or more views
- Appearance is the only clue to identify them
- Individual pixel colors are similar very often. However, the match is too noisy.
- Match a (small) neighborhood of colors from one image to a similar neighborhood in others
- Will work if local surface is fronto-parallel and images have similar magnification
- Foreshortening can happen when viewing an oblique surface
- Many ambiguities. We need a lot of help!

Some Examples






Matching Patches

- Compare $m \times m$ patches from two views; Form vectors v and v' of length m^2
- Matching scores between patches:
  - Sum of Absolute Difference (SAD): $\|v - v'\|_1$
  - Sum of squared difference (SSD): $\|v - v'\|_2$
  - Normalized correlation: $\frac{\bar{v}'^T \bar{v}}{\sqrt{\bar{v}^T \bar{v}}\sqrt{\bar{v}'^T \bar{v}'}}$. Range: $[-1, 1]$.

    where $\bar{v}$ and $\bar{v}'$ are vectors with respective patch-mean colour subtracted. Invariant to affine changes in intensity/colour.
- Census Transform
- Birchfield-Tomasi Match (fractional pixel match by interpolation)

Census Transform

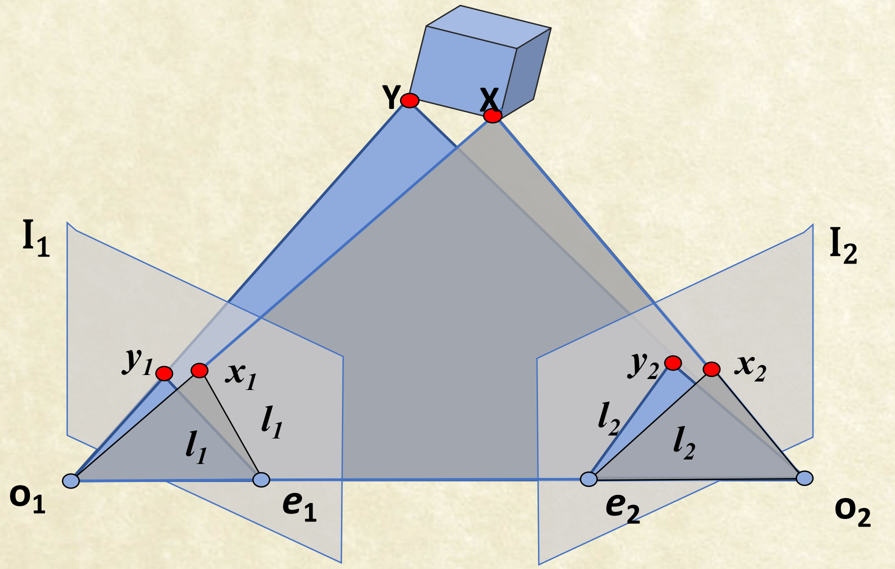
$$\begin{array}{|c|c|c|}\hline 124 & 74 & 32 \\ \hline 124 & 64 & 18 \\ \hline 157 & 116 & 84 \\ \hline \end{array} \longrightarrow \begin{array}{|c|c|c|}\hline 1 & 1 & 0 \\ \hline 1 & x & 0 \\ \hline 1 & 1 & 1 \\ \hline \end{array} \longrightarrow 11010111$$

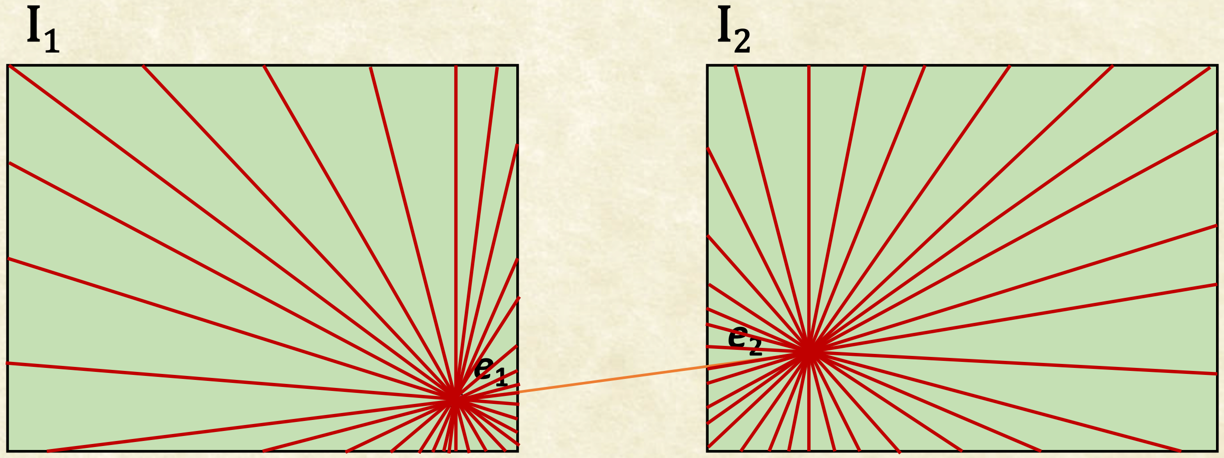
Constraints on Matching

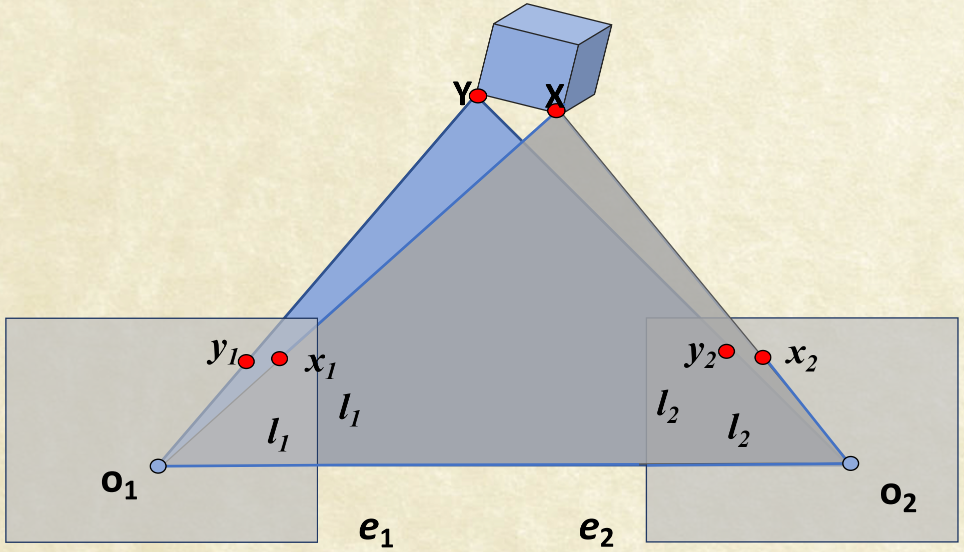
- **Epipolar:** Match lies on the epipolar line of the pixel
- **Colour Constancy:** The appearance/colour does not change from one view to another
- **Uniqueness:** A point on left image can match with only one point on the right and vice versa
- **Ordering or Monotonicity:** If point A is to the left of B in view 1, it will to the left of B in view 2 also. (Violated if great difference in depth exists)
- **Continuity:** Disparity values vary smoothly (violated at occlusion boundaries)

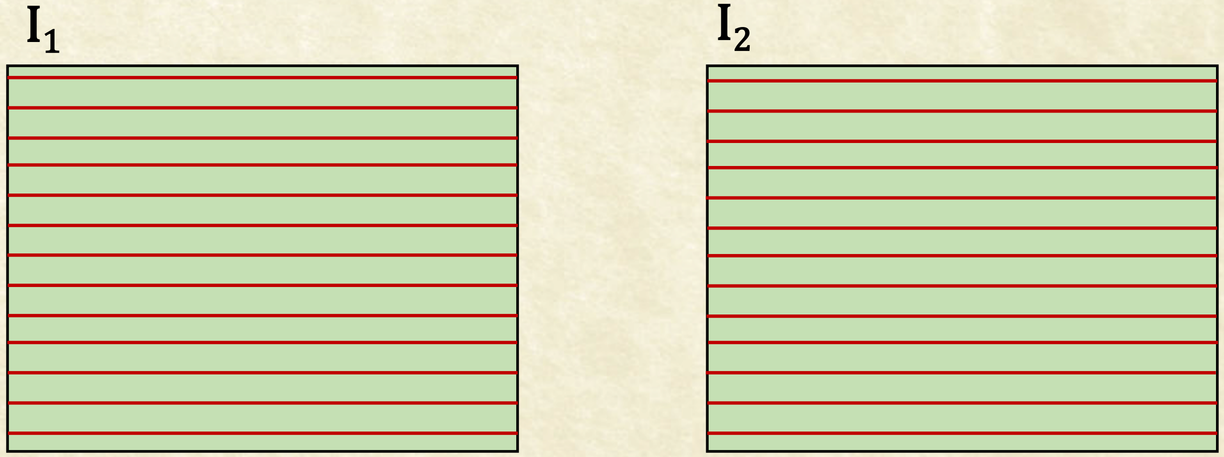
- **Sparse correspondence:** only for good feature points
- **Dense correspondence:** a match for every pixel

Epipolar Constraint









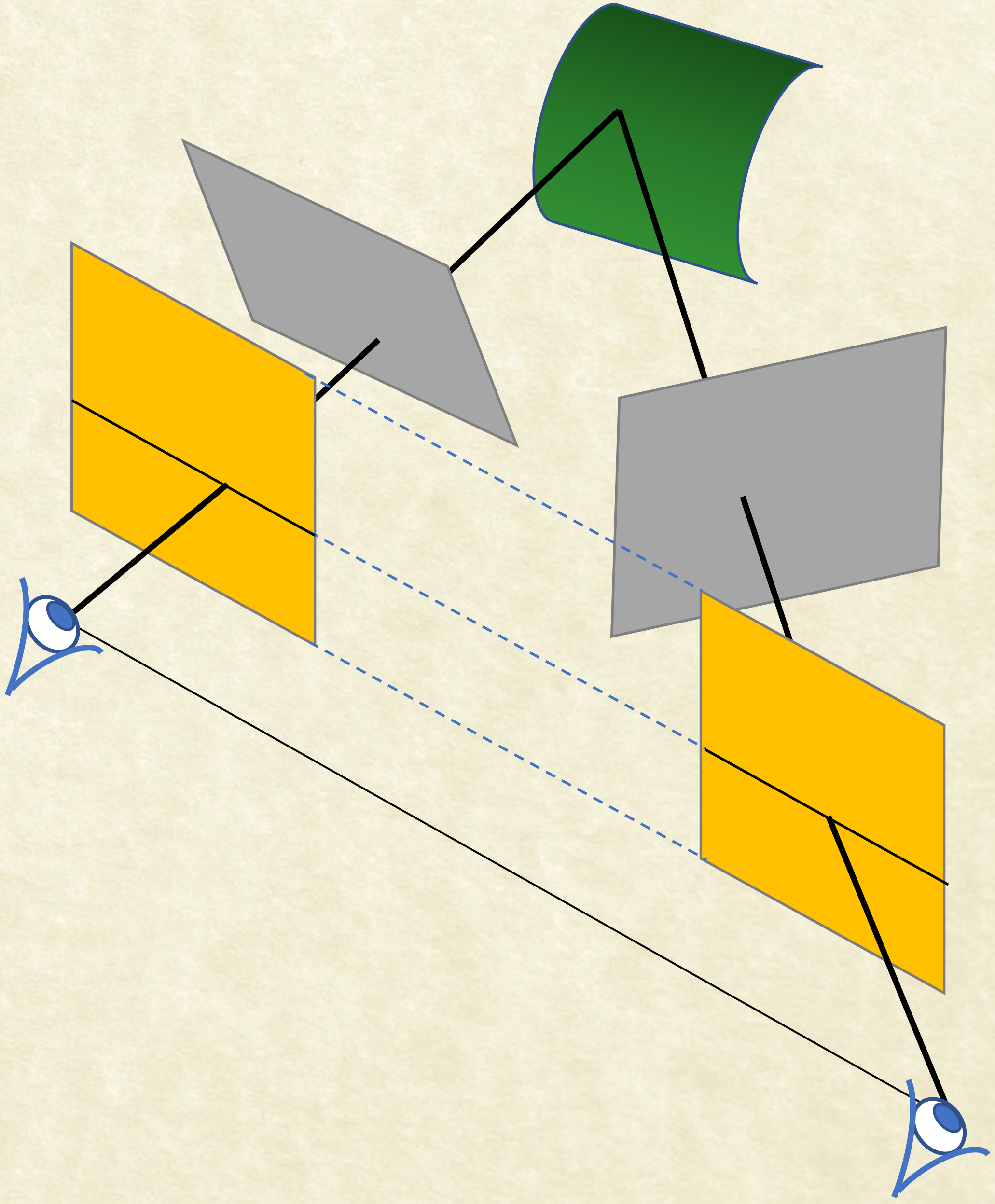
Epipolar: Reduced Search and Rectification

- Search is limited to a line if fundamental matrix is known (weakly calibrated)
- Simplest if left and right cameras have same image plane and pure X-translation between them.
  - Fundamental matrix has a simple form and Epipolar constraint reduces to $y' = y$.
  - Matches constrained to lie in the same scan line
- **Rectification:** A rotation of the camera (to make image planes parallel) and a change in K matrix (focal length, image center).
  - Can be represented using a homography H to align one image plane to the other
  - Or, homographies H_1, H_2 to align them to a third plane.

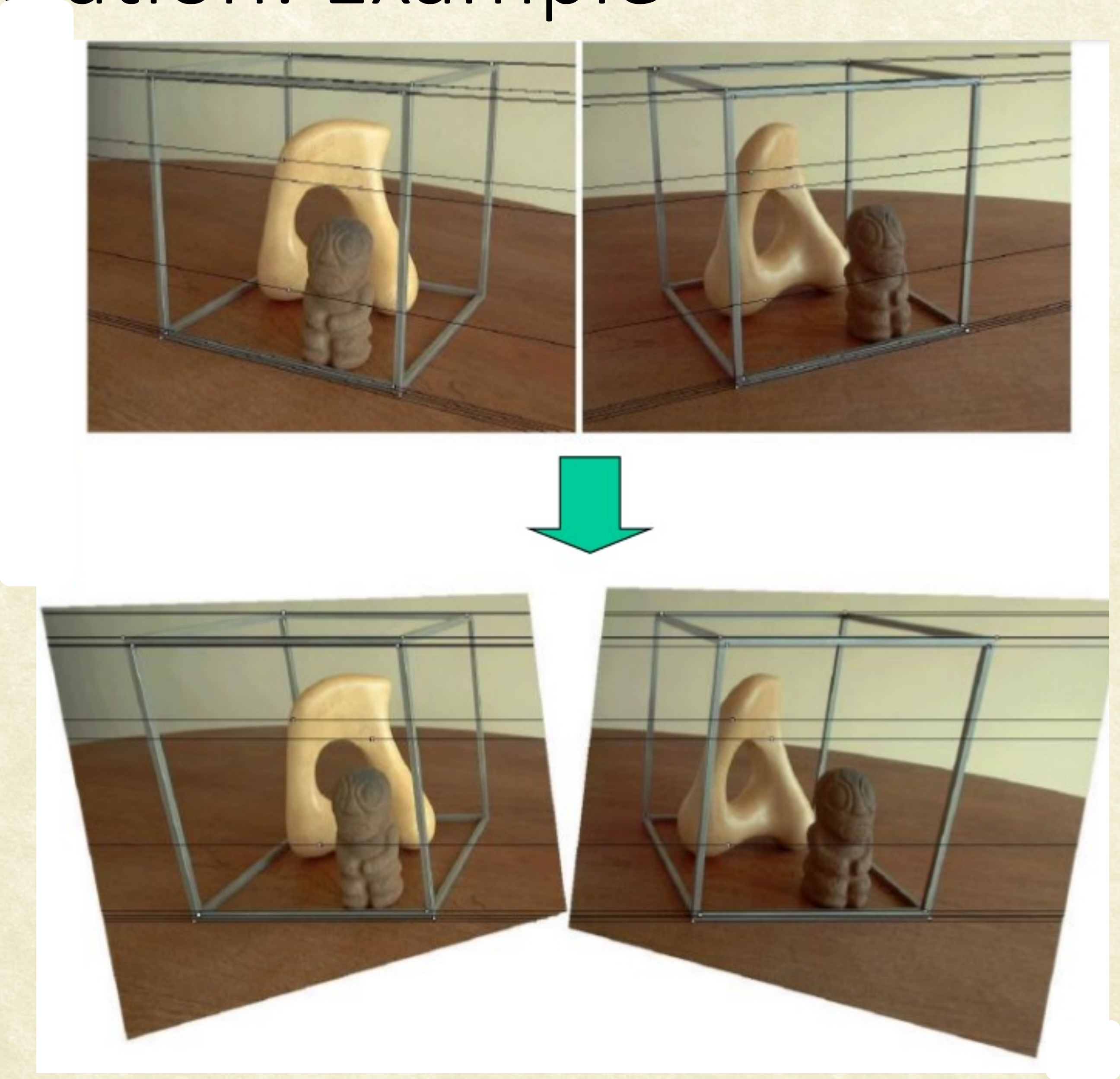
Stereo Rectification

- Reproject images onto a common plane parallel to the line between the optical centers
- Pixel motion is horizontal after this transformation

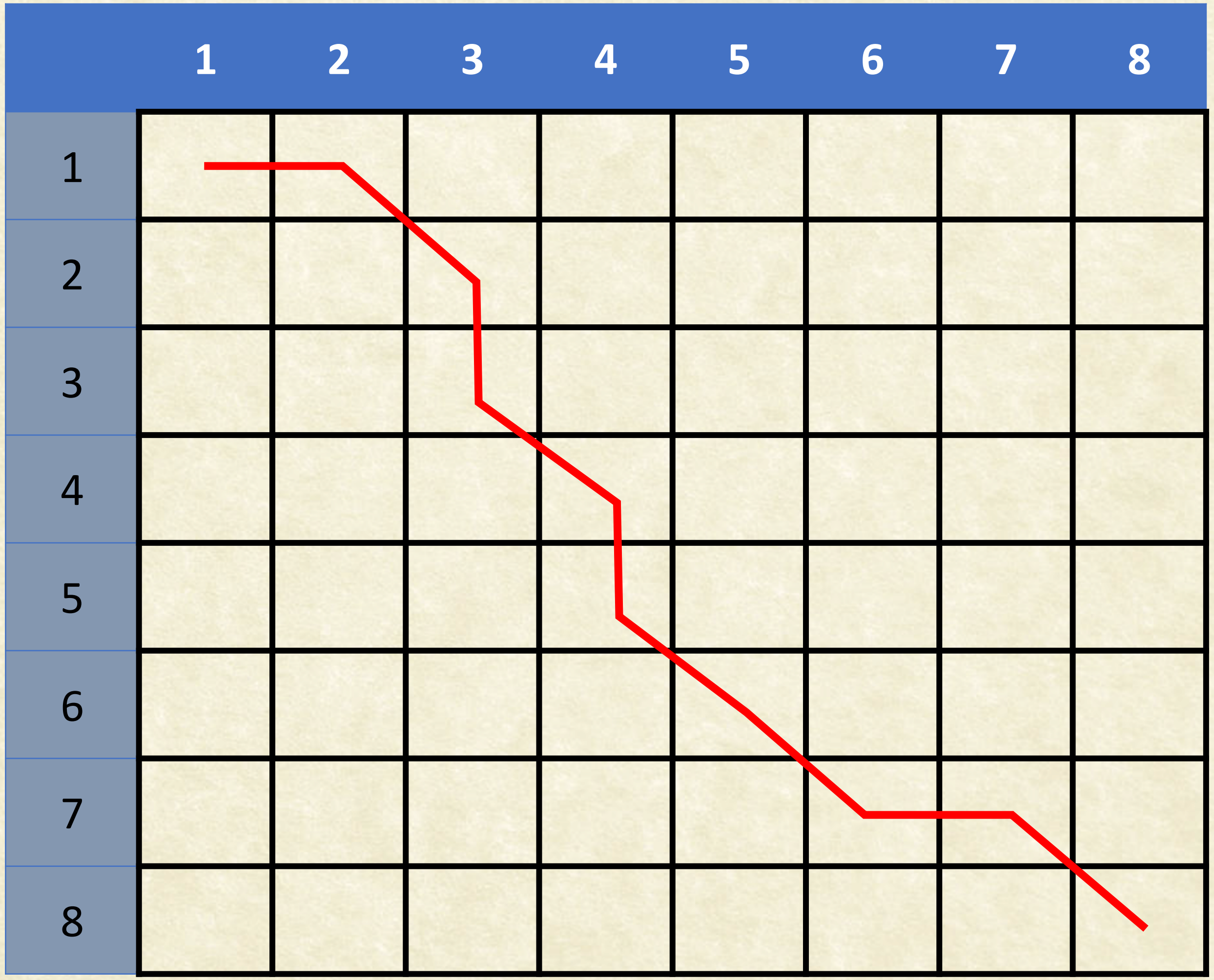
C. Loop and Z. Zhang "Computing Rectifying Homographies for Stereo Vision", IEEE Computer Vision and Pattern Recognition 1999



Rectification: Example



Dynamic Programming



Stereo vs. Optical Flow

- Calibrated vs. Uncalibrated cameras
- 1-D vs. 2-D Matching
- Occlusions are handled (marked) explicitly
- Disparities are quantized, allowing special optimization techniques (DP, Graph Cuts)
- Special constraints on matching
- Simultaneous vs. sequential image capture

Questions?