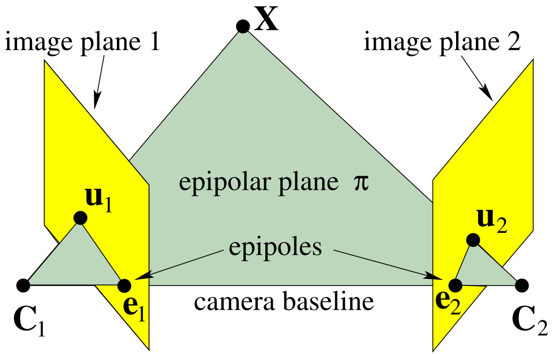
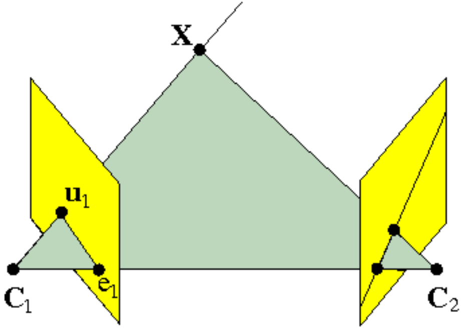
# Geometry of stereo vision



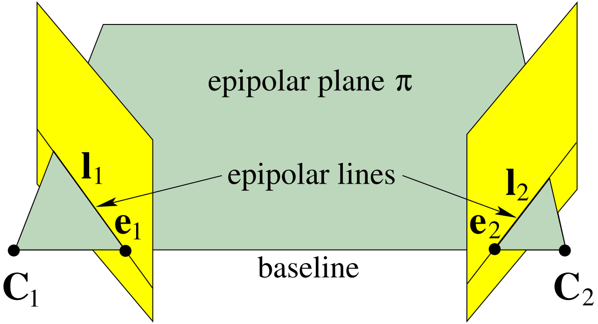

- **Baseline $\mathbf{C}_1\mathbf{C}_2$** connects two focal points.
- Baselines intersect image planes at epipoles.
- Two focal points and the spatial point **X** defines **epipolar plane**.

# Geometry of stereo vision: a video



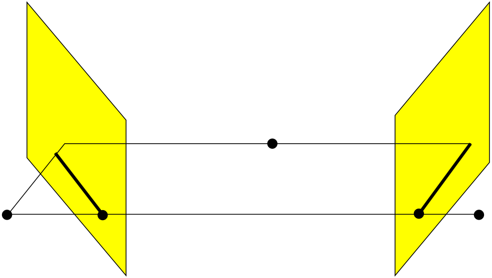

- Point **X** lies on line on ray back-projected using the point in the first image
- Point in the second image, corresponding to $\mathbf{u}_1$, lies on an **epipolar line**
  - $\rightarrow$ **epipolar constraint**
- Line $\mathbf{u}_1\mathbf{e}_1$ is the related epipolar line in the first image.

# Epipolar geometry




- Each plane, containing the baseline, is an epipolar plane
- Epipolar plane $\pi$ intersects the images at lines $\mathbf{l}_1$ and $\mathbf{l}_2$.
  - $\rightarrow$ Two epipolar lines correspond to each other.

# Epipolar geometry: video

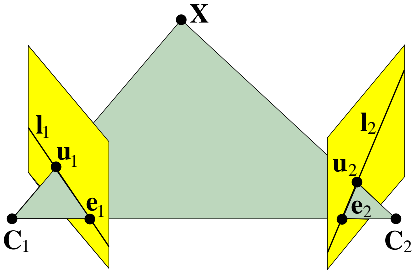


- Epipolar plane 'rotates' around the baseline.
- Each epipolar line contains epipole(s).

# Overview

1. Image-based 3D reconstruction
2. Geometry of stereo vision
   - Epipolar geometry
   - Essential and fundamental matrices
   - Estimation of the fundamental matrix
3. Standard stereo and rectification
   - Triangulation for standard stereo
   - Retification of stereo images
4. 3D reconstruction from stereo images
   - Triangulation and metric reconstruction
   - Projective reconstruction
   - Planar Motion
5. Summary

# Calibrated cameras: essential matrix 1/2




- Calibration matrix **K** is known, rotation **R** and translation **t** between coordinate systems are unknown.
- Lines $\mathbf{C}_1\mathbf{u}_1$, $\mathbf{C}_2\mathbf{u}_2$, $\mathbf{C}_1\mathbf{C}_2$ lay within the same plane:

$$\mathbf{C}_2\mathbf{u}_2 \cdot [\mathbf{C}_1\mathbf{C}_2 \times \mathbf{C}_1\mathbf{u}_1] = 0$$

# Calibrated cameras: essential matrix 2/2

- In the second camera system, the following equation holds if homogeneous coordinates are used:

$$\mathbf{u}_2 \cdot [\mathbf{t} \times \mathbf{R}\mathbf{u}_1] = 0$$

- Using the **essential matrix E** (Longuet-Higgins, 1981):

$$\mathbf{u}_2^T \mathbf{E} \mathbf{u}_1 = 0, \qquad (1)$$

where essential matrix is defined as

$$\mathbf{E} \doteq [\mathbf{t}]_\times R \qquad (2)$$

  - $[\mathbf{a}]_\times$ is the cross-product matrix:

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

# Properties of an essential matrix

- The equation $\mathbf{u}_2^T \mathbf{E} \mathbf{u}_1 = 0$ is valid if the 2D coorinates are normalized by $\mathbf{K}$.
  - Normalized camera matrix: $\mathbf{P} \longrightarrow \mathbf{K}^{-1}\mathbf{P} = [\mathbf{R}| - \mathbf{t}]$

  $\rightarrow$ Normalized coordinates: $\mathbf{u} \longrightarrow \mathbf{K}^{-1}\mathbf{u}$
- Matrix $\mathbf{E} = [\mathbf{t}]_\times R$ has 5 degree of freedom (DoF).
  - $3(\mathbf{R}) + 3(\mathbf{t}) - 1(\lambda)$
  - $\lambda$: (scalar unambigity)
- Rank of essential matrix is 2.
  - $\mathbf{E}$ has two equal, non-zero singular value.
- Matrix $\mathbf{E}$ can be decomposed to translation and rotation by SVD.
  - translation is up to an unknown scale
  - sign of $\mathbf{t}$ is also ambiguous

# Uncalibrated case: fundamental matrix

- Longuet-Higgins formula in case of **uncalibrated** cameras

$$\mathbf{u}_2^T F \mathbf{u}_1 = 0, \tag{3}$$

where the **fundamental matrix** is defined as

$$\mathbf{F} \doteq \mathbf{K}_2^{-T} \mathbf{E} \mathbf{K}_1^{-1} \tag{4}$$

  - $\mathbf{u}_1$ and $\mathbf{u}_2$ are unnormalized coordinates.
- Matrix **F** has 7 DoF.
- Rank of **F** is 2
  - Epipolar lines intersect each other in the same points
  - $\det \mathbf{F} = 0 \longrightarrow$ **F** cannot be inverted, it is non-singular.
- Epipolar lines: $\mathbf{l}_1 = \mathbf{F}^T \mathbf{u}_2$, $\mathbf{l}_2 = \mathbf{F} \mathbf{u}_1$
- Epipoles: $\mathbf{F}\mathbf{e}_1 = \mathbf{0}$, $\mathbf{F}^T \mathbf{e}_2 = \mathbf{0}^T$

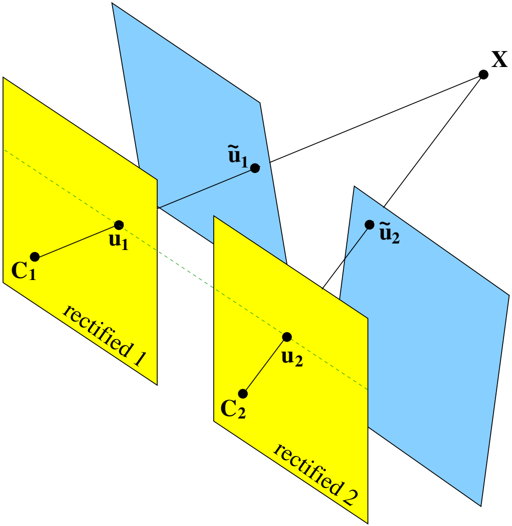
# Goals of rectification

- Input of rectification: non-standard stereo image pair
- Goal of rectification: make stereo matching more accurate
  - After rectification, corresponding pixels are located in the same row
  - $\rightarrow$ standard stereo, 1D search
- Rectification based on epipolar geometry
  - Images are transformed based on epipolar geometry
  - $\rightarrow$ after transformation, corresponding epipolar lines are placed on the same rows
  - $\rightarrow$ epipoles are in the infinity
- For rectification, only the fundamental matrix has to be known
  - $\rightarrow$ Fundamental matrix represents epipolar geometry

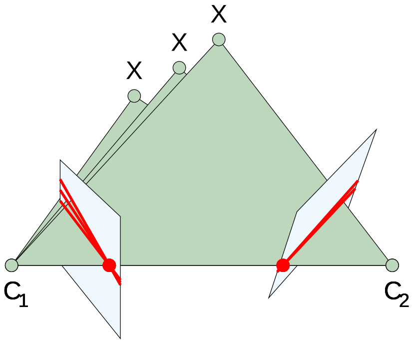
# Rectification methods

- Only the general principles are discussed here.
  - Rectification is a complex method.
  - Rectification **is not required**, it has both advantages and disadvantages.
- Rectification can be carried out by homographies.
- It has ambiguity: there are infinite number of rectification transformations for the same image pair.
- The aim is to find a 2D projective transformation that
  - fulfills the requirement for rectification and
  - distorts minimally the images.
- Knowledge of camera intrinsic parameters helps the rectification.

# Geometry of rectification

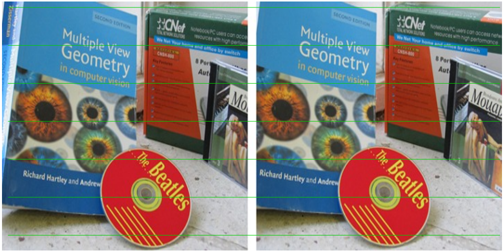


# Rectification: a video video

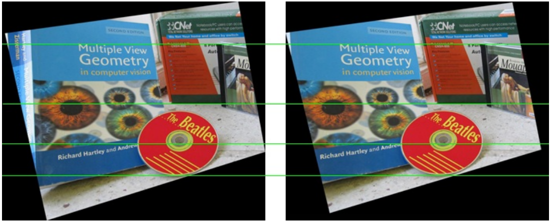



Epipoles transformed to infinity

# Rectification: an example




before




after

# Benefits of rectifications

- Modify the inage in order to get a standard stereo,
  - $\rightarrow$ then algorithms for standard stereo can be applyied.
- The properties of epipolar geometry can be visualized by rectifying the images.
- For practical purposes, the rectification has to be very accurate
  - otherwise there will be a shift between corresponding rows.
  - $\rightarrow$ feature matching more challenging, 1D cannot be run.

# Weak points of rectification

- Distortion under rectification hardly depends on baseline width.
- For wide-baseline stereo:
  - Rectification significantly destorts the image.
  - $\rightarrow$ Pixel-based method can be applied for feature matching
  - $\rightarrow$ Correspondence-based methods often fail.
- Size and shape of rectified images differ from original ones.
  - $\rightarrow$ Feature matching is more challenging.

$\rightarrow$ Many experts do not agree that rectification is necessary.

  - Epipolar lines can be followed if fundamental matrix is given.
  - Matching can be carried out in original frames.
  - $\rightarrow$ Then noise is not distorted by rectifying transformation.