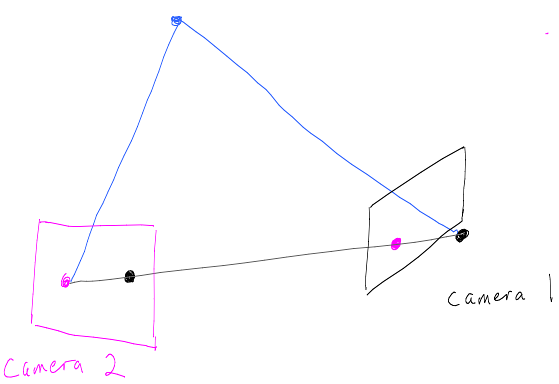
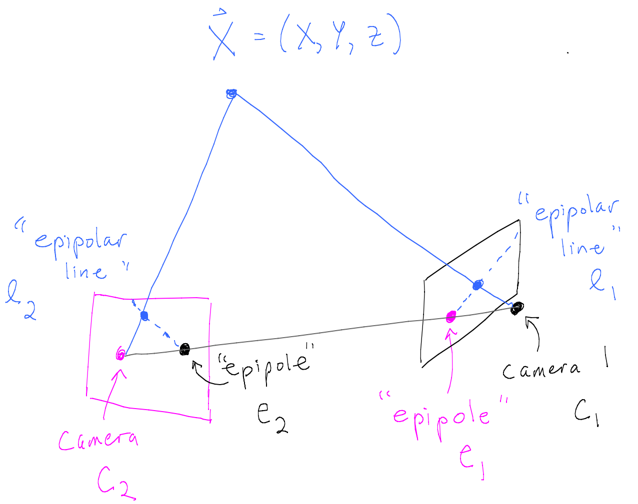
lecture 22

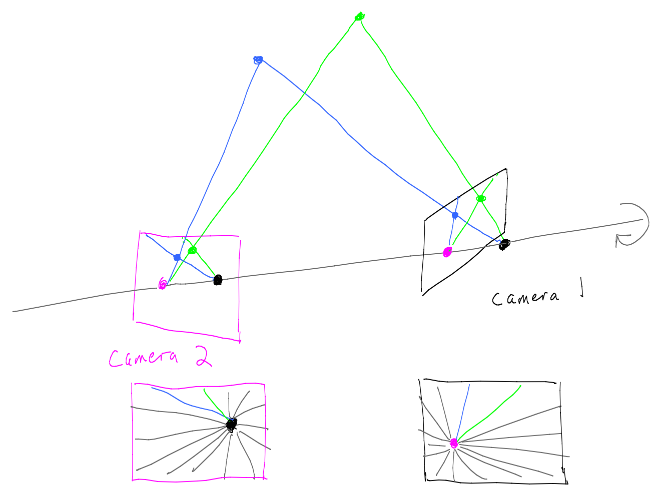
Binocular stereo
without camera calibration
(Epipolar Geometry)



Today's problem

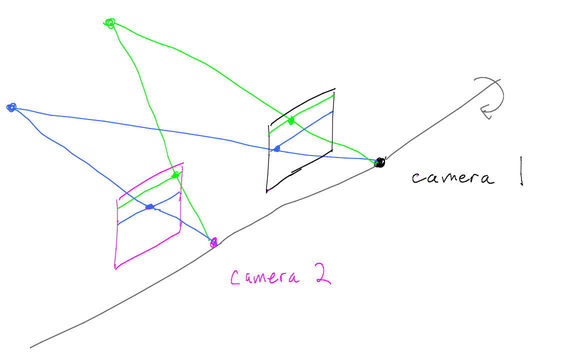
Stereo Geometry – what if the cameras are <u>not</u> calibrated? (We are given two images. What can we do?)

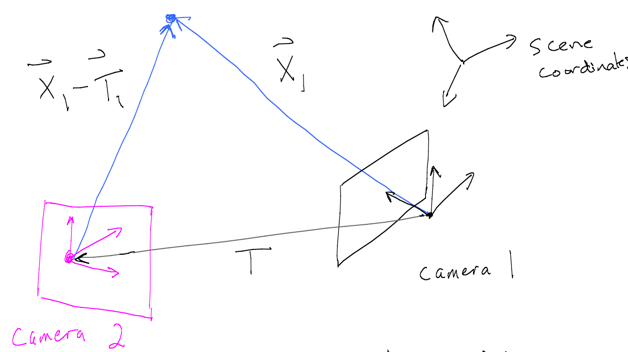




Example

Special Case



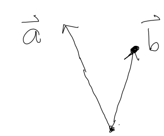


$$(\vec{X}_1 - \vec{T}_1) \cdot (\vec{T}_1 \times \vec{X}_1) = 0$$

black letters = written in Camera 1's coordinate system

Cross Product (recall lecture 3)

Let $\vec{a}$ and $\vec{b}$ be any vectors



$$\vec{a} \times \vec{b} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}$$

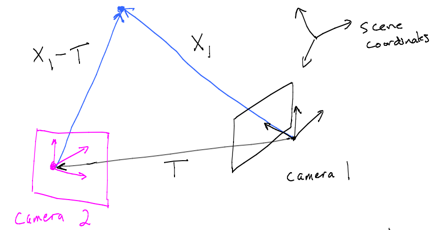
$$\vec{a} \times \vec{b} = (a_y b_z - a_z b_y,\ b_x a_z - a_x b_z,\ a_x b_y - a_y b_x)^T$$

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

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

Note:

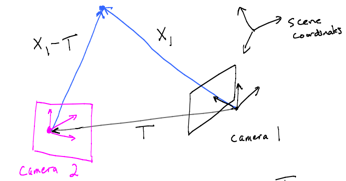
- $[\vec{a}]$ is anti-symmetric
- its null vector is $\vec{a}$
- $[\vec{a}]$ is of rank 2



$$(\vec{X}_1 - \vec{T}_1) \cdot (\vec{T}_1 \times \vec{X}_1) = 0$$

$$(R_1 R_2^T X_2) \cdot [T_1]_\times X_1 = 0$$

$$X_2^T R_2 R_1^T [T_1]_\times X_1 = 0$$

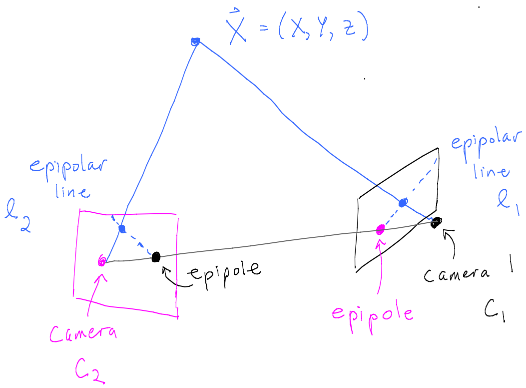


$$X_2^T \underbrace{R_2 R_1^T [T_1]_\times} X_1 = 0$$

called the "essential matrix" $E_{3\times 3}$

(the vector perpendicular to plane spanned by T & X_1 is perpendicular to $T - X_1$)

Epipolar Geometry - let's do the algebra!



Epipolar Lines

Choose X_1.

$$\vec{X}_2^T \underbrace{E \vec{X}_1}_{\vec{\ell}_2} = 0$$

i.e. $\vec{X}_2^T \cdot \vec{\ell}_2 = 0$

$\implies$ X_1's epipolar line in camera 2's projection plane is

$$(x_2, y_2, f_2) \cdot \vec{\ell}_2 = 0$$

Epipolar Lines

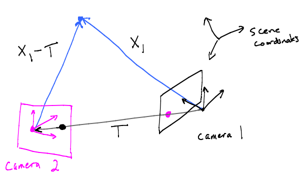
Choose X_2.

$$\underbrace{\vec{X}_2^T E}_{\vec{\ell}_1} \vec{X}_1 = 0$$

i.e. $\vec{\ell}_1 \cdot \vec{X}_1 = 0$

$\implies$ X_2's epipolar line in camera 1's projection plane is

$$\vec{\ell}_1 \cdot (x_1, y_1, f_1) = 0$$



$$\vec{X}_2^T E \vec{X}_1 = 0$$

$$X_2^T R_2 R_1^{-1} [T_1]_\times X_1 = 0$$

Epipoles

$$[T_1]_\times T_1 = \vec{0} \implies X_2^T E T_1 = 0 \text{ for all } X_2$$

$$T_2^T R_2 R_1^{-1} [T_1]_\times = \vec{0} \implies T_2^T E X_1 = 0 \text{ for all } X_1$$

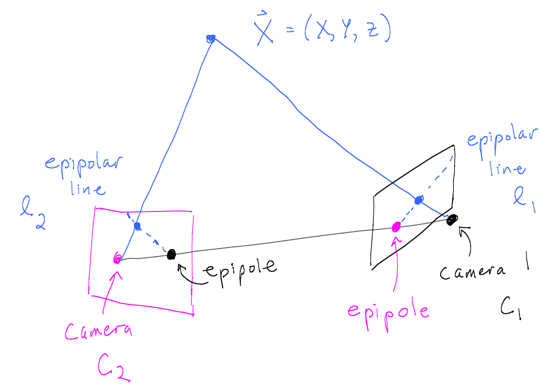
Fundamental Matrix

$$\vec{X}_2^T E \vec{X}_1 = 0$$

$$X_2^T K_2^T K_2^{-T} E K_1^{-1} K_1 \vec{X}_1 = 0$$

$$(x_2\ y_2\ 1) \underbrace{K_2^{-T} E K_1^{-1}}_{F} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = 0$$

Epipolar Geometry — let's do it in pixel coordinates



Epipolar Lines (in pixel units)

$$\vec{x}_2^T \underbrace{F \vec{x}_1}_{\vec{l}_2} = 0 \qquad \overbrace{\vec{x}_2^T F}^{\vec{l}_1}$$

$\vec{x}_2^T \vec{l}_2 = 0$ — $\vec{x}_1$'s epipolar line in image 2

$\vec{l}_1^T \vec{x}_1 = 0$ — $\vec{x}_2$'s epipolar line in image 1

Epipoles

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

$F \vec{e}_1 = \vec{0} \Rightarrow \vec{l}_1 \cdot \vec{e}_1 = 0$ for all $\vec{l}_2$

$\vec{e}_2^T F = \vec{0} \Rightarrow \vec{e}_2 \cdot l_2 = 0$ for all $\vec{l}_2$

i.e. epipoles are intersection of epipolar lines.

Remaining Questions

- Given 2 images and a set of corresponding $(x_1 y_1), (x_2 y_2)$ pairs, how can we estimate F?
- (More basic) Given two images, how can we automatically find a set of corresponding pairs?
- Given F, how can we find correspondences for all pixels?
- Given all correspondences, what are limits on 3D reconstruction?

Given 2 images, and a set of corresponding points $(x, y_1), (x_2 y_2)$, how can we estimate the fundamental matrix?

$$[x_2 \; y_2 \; 1] \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = 0$$

$$\underbrace{\begin{bmatrix} x_2x_1 & x_2y_1 & x_2 & y_2x_1 & y_2y_1 & y_2 & x_1 & y_1 & 1 \\ & & & & & & & & \end{bmatrix}}_{N \times 9} \begin{bmatrix} F_{11} \\ F_{12} \\ F_{13} \\ \vdots \\ F_{33} \end{bmatrix} = 0$$

$$A = \begin{bmatrix} x_2x_1 & x_2y_1 & x_2 & y_2x_1 & y_2y_1 & y_2 & x_1 & y_1 & 1 \\ & & & & \vdots & & & & \end{bmatrix}$$

need at least 8 corresponding pairs $(x_1 y_1)(x_2 \; y_2)$

Compute $A = U \Sigma V^T$ and construct F from last column of V.
i.e. goes with smallest σ_i.

ASIDE: For your information.....

In fact, F has 7 (not 8) degrees of freedom.

- homogeneous i.e. you can scale it
- $\det(F) = 0$

Remaining Questions

- Given 2 images and a set of corresponding $(x_1 y_1), (x_2 y_2)$ pairs, how can we estimate F?
- (More basic) Given two images, how can we automatically find a set of corresponding pairs?
- Given F, how can we find correspondences for all pixels?
- Given all correspondences, what are limits on 3D reconstruction?

Where do correspondences

$(x_1 \; y_1) \Leftrightarrow (x_2 \; y_2)$

Come from?

- detect SIFT keypoints in each image $\{(x_1, y_1)\}_i$ $\{(x_2 y_2)_j\}$
- match keypoints between images based on the similarity of their SIFT descriptors

RANSAC (similar to estimation of H)

- repeat
  - random sample 8 quadruples $(x_1 \; y_1 \; x_2 \; y_2)_i$ = 1 "trial"
  - fit a model for this trial
  - go through all other candidate matches $\tilde{x}_1 \tilde{y}_1 \tilde{x}_2 \tilde{y}_2$ and build the "consensus set" for this F
  - increment counter

  until (consensus set $> \tau_2$) or (counter == numTrials)
- refit F using largest consensus set and return

Because of noise in $(x_1,y_1)(x_2,y_2)$, F will not be of rank 2

Compute $F = U \Sigma V^T$

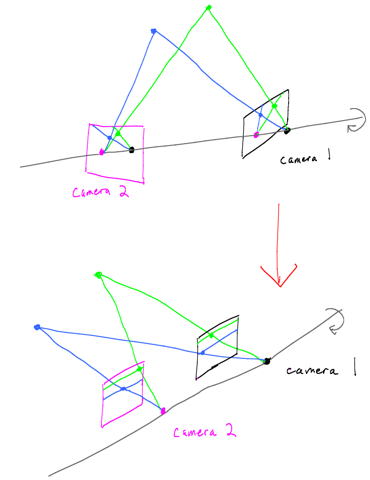
Set σ_3 to 0.

$$\hat{F} = U \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & 0 \end{bmatrix} V^T$$

Gives best rank 2 approximation to F

Remaining Questions

- Given 2 images and a set of corresponding $(x_1,y_1), (x_2,y_2)$ pairs, how can we estimate F?
- (More basic) Given two images, how can we automatically find a set of corresponding pairs?
- Given F, how can we find correspondences for all pixels?
- Given all correspondences, what are limits on 3D reconstruction?



Pixel image rectification can be done, given only F!

Find homographies H_1, H_2 that map epipoles e_1, e_2 to $(\pm 1, 0, 0)$.

Non-trivial

∴ Details omitted.

The task now is to find corresponding points for all pixels, not just keypoints.

We will do that next lecture.

However, if cameras are not calibrated (today), then we won't be able to estimate 3D positions (X, Y Z). Why not?

Projective Reconstruction Theorem

Suppose

$$\begin{pmatrix} w x_1 \\ w y_1 \\ w \end{pmatrix} = P_1 \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

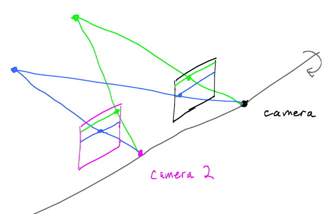
$$\begin{pmatrix} w x_2 \\ w y_2 \\ w \end{pmatrix} = P_2 \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Consider any invertible 4x4 matrix M.

$$\begin{pmatrix} w x_1 \\ w y_1 \\ w \end{pmatrix} = \overbrace{P_1 M}^{\text{new camera 1}} M^{-1} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} w x_2 \\ w y_2 \\ w \end{pmatrix} = \underbrace{P_2 M}_{\text{new camera 2}} \underbrace{M^{-1} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}}_{\text{new geometry}}$$

Therefore, given 2 images, we can at best reconstruct the scene and camera matrix up to a "projective ambiguity".



Assume

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 1 & 0 & 0 & -T \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$x_1 = \frac{X}{Z}$$

$$x_2 = \frac{X - T}{Z} \qquad x_1 - x_2 = \frac{T}{Z}$$

$\Rightarrow$ compute (X, Y, Z) up to a "projective reconstruction"

- Friday (review for final)
- Exercises 2 posted
- For fun, look up

Fundamental Matrix Song (youtube)