CS-565 Computer Vision

Nazar Khan

Department of Computer Science
University of the Punjab

19. Stereo Reconstruction

Stereo Reconstruction

- So far, we have only investigated the projective geometry of the monocular case (single pinhole camera).
- In this lecture, we study the binocular case (two pinhole cameras).
- Disparity between the two views allows reconstruction of the $3D$ scene.
- Known as *stereo reconstruction*.
- To do this, we will study stereo geometry (also called epipolar geometry).

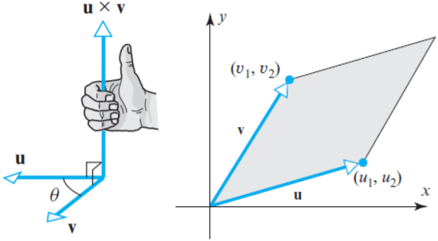
Cross Product

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}}_{[\mathbf{u}]_\times} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

- Only defined for 3-dimensional spaces such as $\mathbb{R}^3$ and $\mathbb{P}^2$.
- Matrix $[\mathbf{u}]_\times$ has two linearly independent rows.
  - *Proof*: u_1 row1 $+ u_2$ row2 $+ u_3$ row3 $= \mathbf{0}^T \implies$ any row can be written as a linear combination of the other two rows.
- $\mathbf{u} \times \mathbf{v}$ is another 3-dimensional vector orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

Cross Product

- $\|\mathbf{u} \times \mathbf{v}\|$ represents the area of the parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$.



- If $\mathbf{u}$ and $\mathbf{v}$ point in the same direction, then no parallelogram will be formed.
- Therefore $\|\mathbf{u} \times \mathbf{v}\|$ will be 0.
- The only vector with norm 0 is the $\mathbf{0}$ vector.
- Therefore, $\mathbf{u} \times \mathbf{v} = \mathbf{0}$ when $\mathbf{u}$ and $\mathbf{v}$ point in the same direction.

Lines and Points in Homogeneous Coordinates

- For a point $\mathbf{x} = (x, y)^T$ in Euclidean coordinates, the point $\tilde{\mathbf{x}} = (x, y, 1)^T$ is its counterpart in homogeneous coordinates.
- Let $\mathbf{m}_1 = (x_1, y_1)^T$ and $\mathbf{m}_2 = (x_2, y_2)^T$ be two different points (i.e. $\mathbf{m}_1 \neq \mathbf{m}_2$) in Euclidean coordinates.
- A $2D$ line can be represented as $\ell_1 = (a_1, b_1, c_1)^T$. It consists of all points (x, y) that satisfy $a_1x + b_1y + c_1 = 0$.
- In homogenous coordinates

$$\tilde{\mathbf{m}}_1 = \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad \tilde{\mathbf{m}}_2 = \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \quad \ell_1 = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix}$$

Lines and Points in Homogeneous Coordinates

1. If $\mathbf{m}_1$ lies on line ℓ_1 then $\tilde{\mathbf{m}}_1^T \ell_1 = 0$.

$$\begin{aligned} \mathbf{m}_1 \text{ lies on line } \ell_1 &\implies a_1 x_1 + b_1 y_1 + c_1 = 0 \\ &\implies \begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} = 0 \\ &\implies \tilde{\mathbf{m}}_1^T \ell_1 = 0 \end{aligned}$$

2. If $\mathbf{m}_1$ is the intersection of ℓ_1 and ℓ_2 then $\ell_1 \times \ell_2 \equiv \tilde{\mathbf{m}}_1$.
   Proof:
   Since $\mathbf{m}_1$ lies on both ℓ_1 and ℓ_2, $\tilde{\mathbf{m}}_1^T \ell_1 = 0$ and $\tilde{\mathbf{m}}_1^T \ell_2 = 0$.
   So vector $\tilde{\mathbf{m}}_1$ is orthogonal to both vector ℓ_1 and vector ℓ_2.
   We know that the vector that is orthogonal to both ℓ_1 and ℓ_2 is their cross-product $\ell_1 \times \ell_2$. Hence, $\tilde{\mathbf{m}}_1 \equiv \ell_1 \times \ell_2$.

Lines and Points in Homogeneous Coordinates

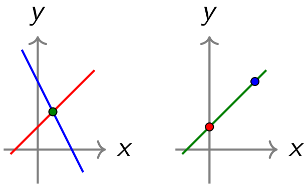
3. If line ℓ_1 connects points $\mathbf{m}_1$ and $\mathbf{m}_2$ then $\tilde{\mathbf{m}}_1 \times \tilde{\mathbf{m}}_2 \equiv \ell_1$.

   Proof:

   Since ℓ_1 contains both $\mathbf{m}_1$ and $\mathbf{m}_2$, $\tilde{\mathbf{m}}_1^T \ell_1 = 0$ and $\tilde{\mathbf{m}}_2^T \ell_1 = 0$. So vector ℓ_1 is orthogonal to both vector $\tilde{\mathbf{m}}_1$ and vector $\tilde{\mathbf{m}}_2$.

   We know that the vector that is orthogonal to both $\tilde{\mathbf{m}}_1$ and $\tilde{\mathbf{m}}_2$ is their cross-product $\tilde{\mathbf{m}}_1 \times \tilde{\mathbf{m}}_2$. Hence, $\ell_1 \equiv \tilde{\mathbf{m}}_1 \times \tilde{\mathbf{m}}_2$.

- Exercises:
  1. Find the point of intersection of lines $y = x + 1$ and $y = -2x + 3$.
  2. Find the line passing through the points $(0, 1)$ and $(2, 3)$.



Single View

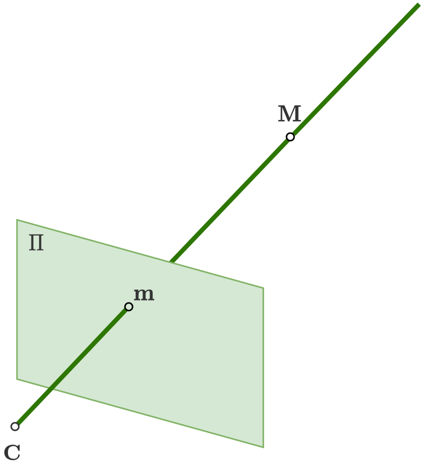


Figure: In a calibrated camera, any pixel **m** can be back-projected to form a ray $K^{-1}\tilde{\mathbf{m}}$ in the camera coordinate system. Author: N. Khan (2021)

Two Views

Triangulation

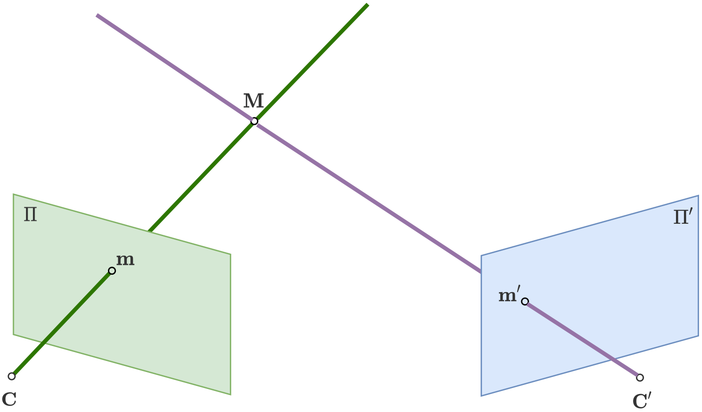


Figure: Back-projected rays from two cameras intersect at the $3D$ location of the world point $\mathbf{M}$. Given corresponding points $\mathbf{m}$ and $\mathbf{m}'$, recovering $\mathbf{M}$ in this manner is known as *triangulation*. Author: N. Khan (2021)

Epipoles

Figure: The image of one camera center in the other camera is called the *epipole*. Given calibrated cameras P and P', epipoles of the stereo setup can be computed as, $\mathbf{e} \equiv P\tilde{\mathbf{C}}'$ and $\mathbf{e}' \equiv P'\tilde{\mathbf{C}}$. Author: N. Khan (2021)

Epipolar Line

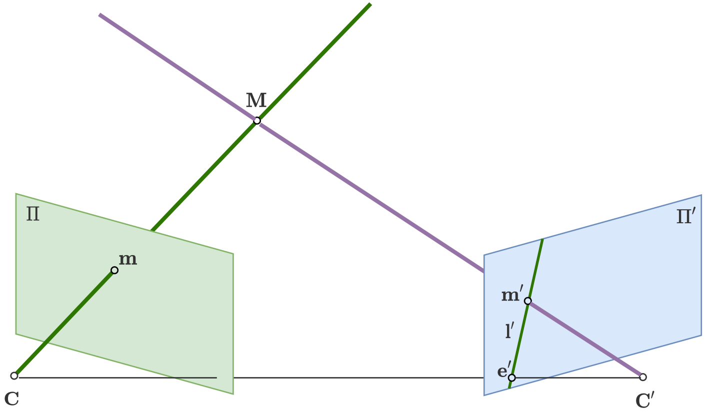


Figure: Given **m**, world point **M** can potentially lie anywhere along the back-projected ray. Therefore, in the second view, corresponding point **m**′ can lie anywhere along the image of the back-projected ray. Author: N. Khan (2021)

Epipolar Line

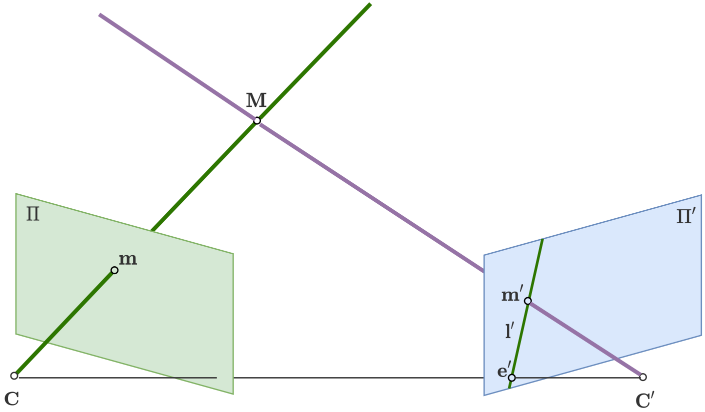


Figure: Image of the back-projected ray is called the *epipolar line* $\mathbf{l}'$. Since it must pass through $\mathbf{m}'$ and the epipole $\mathbf{e}'$, it can be computed as $\mathbf{l}' \equiv \mathbf{m}' \times \mathbf{e}'$. Author: N. Khan (2021)

Epipolar Geometry

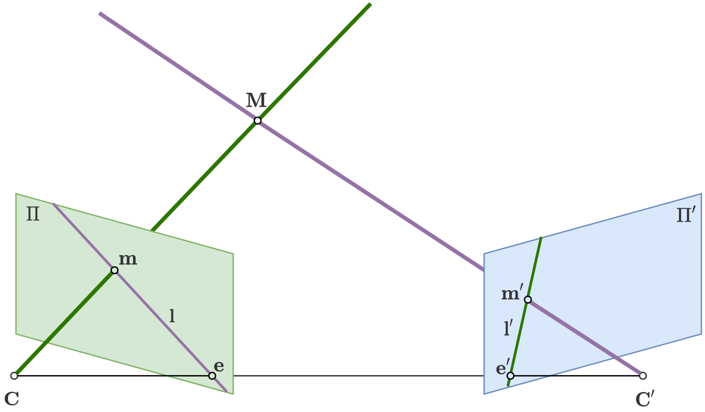


Figure: Line $\mathbf{l}'$ is the epipolar line in the second camera corresponding to pixel $\mathbf{m}$ in the first camera. Similarly, $\mathbf{l} \equiv \mathbf{m} \times \mathbf{e}$ is the epipolar line in the first camera corresponding to pixel $\mathbf{m}'$ in the second camera. *Each pixel in one view has a corresponding epipolar line in the other view.* Author: N. Khan (2021)

Epipolar Plane

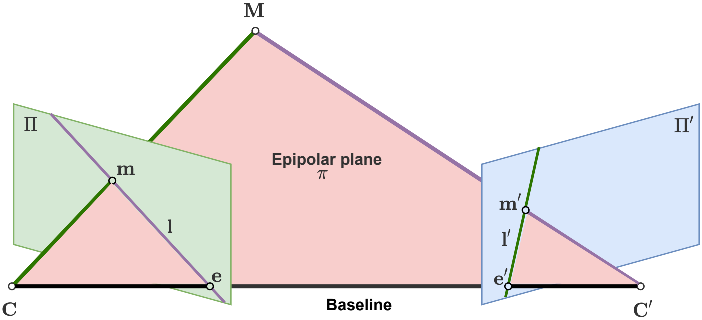


Figure: The world point **M** (alternatively, pixel **m**) and the two camera centers **C** and **C**′ define a plane in $3D$ called the *epipolar plane*. Notice that corresponding points $\mathbf{m}, \mathbf{m}'$, epipoles $\mathbf{e}, \mathbf{e}'$, and epipolar lines $\mathbf{l}, \mathbf{l}'$ also lie on the epipolar plane. Author: N. Khan (2021)

Epipolar Plane

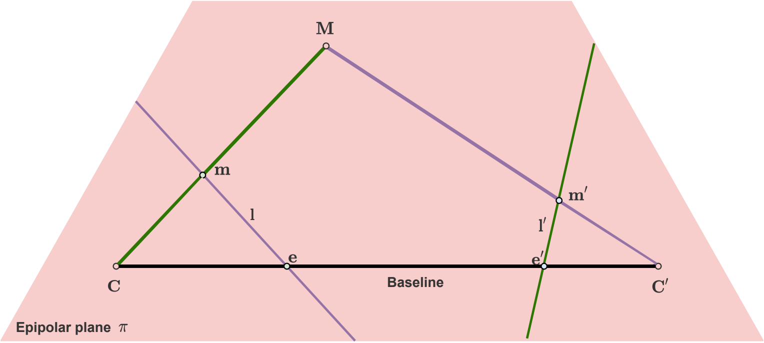
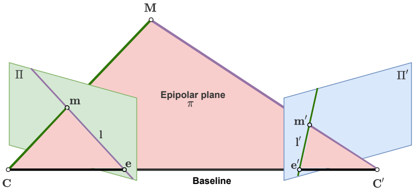


Figure: Any two points of a planar surface are related by an invertible 3×3 matrix, i.e., homography. We will denote the homography that maps $\mathbf{m}$ to $\mathbf{m}'$ by H_π. Therefore, $\mathbf{m}' \equiv H_\pi \mathbf{m}$. Author: N. Khan (2021)

Fundamental Matrix

- Gathering everything together

$$\begin{aligned} \mathbf{l}' &\equiv \mathbf{e}' \times \mathbf{m}' \\ &\equiv \mathbf{e}' \times H_\pi \mathbf{m} \\ &\equiv [\mathbf{e}']_\times H_\pi \mathbf{m} \\ &\equiv F\mathbf{m} \end{aligned}$$



- The 3×3 matrix F is called the *fundamental matrix*.
- It maps any pixel in the first camera to its corresponding epipolar line in the second camera.
- Note that $\text{rank}(F) = 2$.

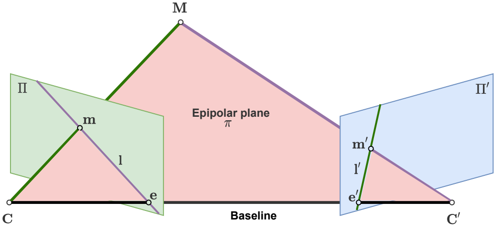
$$\text{rank}(F) = \min\left(\text{rank}([\mathbf{e}']_\times), \text{rank}(H_\pi)\right) = \min(2, 3) = 2$$

- F has 7 degrees of freedom (9 parameters -1 for scale -1 for rank)

Epipolar Constraint

- Since $\mathbf{m}'$ lies on the epipolar line $\mathbf{l}'$

$$\mathbf{m}'^T \mathbf{l}' = 0$$
$$\implies \mathbf{m}'^T F \mathbf{m} = 0$$



- This is known as the *epipolar constraint*.
- Corresponding points $\mathbf{m}$ and $\mathbf{m}'$ must satisfy the epipolar constraint.

Fundamental Matrix Estimation

- When cameras are not calibrated, F can be estimated via $N \geq 8$ correspondences $\mathbf{m}_i, \mathbf{m}'_i$.
- Epipolar constraint can be written as

$$0 = \mathbf{m}'^T F\mathbf{m} = \begin{bmatrix} x' & y' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\begin{aligned} &\quad xx'f_{11} + yx'f_{12} + x'f_{13} \\ &= +xy'f_{21} + yy'f_{22} + y'f_{23} \\ &\quad +xf_{31} + yf_{32} + f_{33} \\ &= \mathbf{s}^T\mathbf{f} \end{aligned}$$

where

$$\mathbf{s} = \begin{bmatrix} xx' & yx' & x' & xy' & yy' & y' & x & y & 1 \end{bmatrix}^T$$

$$\mathbf{f} = \begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{21} & f_{22} & f_{23} & f_{31} & f_{32} & f_{33} \end{bmatrix}^T$$

Fundamental Matrix Estimation

- Vector $\mathbf{f}$ can be estimated by minimizing the sum-squared-error

$$E(\mathbf{f}) = \sum_{i=1}^{N} \left(\mathbf{s}_i^T \mathbf{f}\right)^2 = \sum_{i=1}^{N} \left(\mathbf{s}_i^T \mathbf{f}\right)^T \left(\mathbf{s}_i^T \mathbf{f}\right) = \mathbf{f}^T \underbrace{\left(\sum_{i=1}^{N} \mathbf{s}_i \mathbf{s}_i^T\right)}_{A} \mathbf{f}$$

  subject to $\|\mathbf{f}\| = 1$ to avoid the trivial solution $\mathbf{f} = \mathbf{0}$.
- Constrained optimization problem

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \mathbf{f}^T A \mathbf{f} \qquad \text{s.t. } \|\mathbf{f}\| = 1$$

- $\mathbf{f}^*$ is the eigenvector of A corresponding to the *smallest* eigenvalue.
- $\mathbf{f}^*$ can be reshaped to yield the fundamental matrix F.

Enforcing Rank-2 Constraint

- Recall that rank(F) must be 2.
- If necessary, this can be enforced via SVD.
- SVD allows F to be decomposed as a product of three rank-3 matrices

$$F = UDV^T$$

where

$$D = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$$

with $\sigma_i = \sqrt{\lambda_i(F^T F)}$.

- Replacing σ_3 by 0 and multiplying the three matrices together yields the closest rank-2 approximation of F.

Stereo Setups
Parallel Camera Stereo Rig

Stereo Setups

Forward Translating Camera

Stereo Correspondence

Triangulation