3D Coordinates and Epipolar Geometry

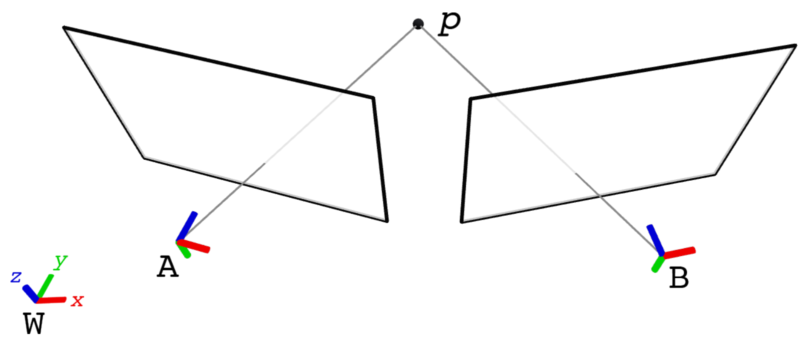
Discussion #8

Written by:
Brent*, Konpat*,
Chung Min*

This worksheet reviews 3D transformations and the basics of epipolar geometry.

1 Rigid Transforms

Let's consider a point p and three coordinate frames: the world W, camera A, and camera B:



Notation. We'll denote the coordinates of p in the world frame $p_{\mathrm{W}} \in \mathbb{R}^3$. The extrinsics[1] of camera A and B can be written as T_{WA} and T_{WB} respectively, where

$$\begin{bmatrix} p_{\mathrm{W}} \\ 1 \end{bmatrix} = T_{\mathrm{WA}} \begin{bmatrix} p_{\mathrm{A}} \\ 1 \end{bmatrix}.$$

(a) The same point p can be written as either p_{W}, p_{A}, and p_{B}. In words, what is the difference between these three vectors?

They are the same point represented relative to three different coordinate systems: one relative to the world origin, one relative to the camera A origin, and one relative to the camera B origin.

(b) Express p_{W} in terms of p_{A} and T_{WA}.

$p_W = T_{\mathrm{WA}} p_A$

[1] Following the "world-from-camera" / "camera-to-world" convention.

(c) Express p_{B} in terms of p_{W} and T_{WB}.

$p_B = T_{\mathrm{WB}}^{-1} p_W$

(d) How would you compute a transform that converts coordinates relative to camera B from camera A, in terms of T_{WA} and T_{WB}?

$T_{\mathrm{BA}} = T_{\mathrm{WB}}^{-1} T_{\mathrm{WA}}$

(e) Consider the case where camera A has intrinsics $K \in \mathbb{R}^{3\times 3}$. Given p_{W}, K, and T_{WA}, what are the image-space coordinates (u, v) of p viewed from camera A? You can leave your solution in homogeneous coordinates.

We first transform the point into the camera frame, drop the homogeneous term, then project:

$$K \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} (T_{\mathrm{WA}}^{-1} p_{\mathrm{W}})$$

(f) Let $T \in \mathrm{SE}(3)$ be composed of rotation $R \in SO(3)$ and $t \in \mathbb{R}^3$. We know that rotation matrices are orthonormal; $R^{-1} = R^\top$. Use this fact to derive an expression for the inverse T^{-1} without using matrix inversion.

Consider the case where T^{-1} has rotation term R' and translation term t':

$$\begin{bmatrix} R' & t' \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p \\ 1 \end{bmatrix} = \begin{bmatrix} p \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} R'(Rp + t) + t' \\ 1 \end{bmatrix} = \begin{bmatrix} p \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} R'Rp + R't + t' \\ 1 \end{bmatrix} = \begin{bmatrix} p \\ 1 \end{bmatrix}$$

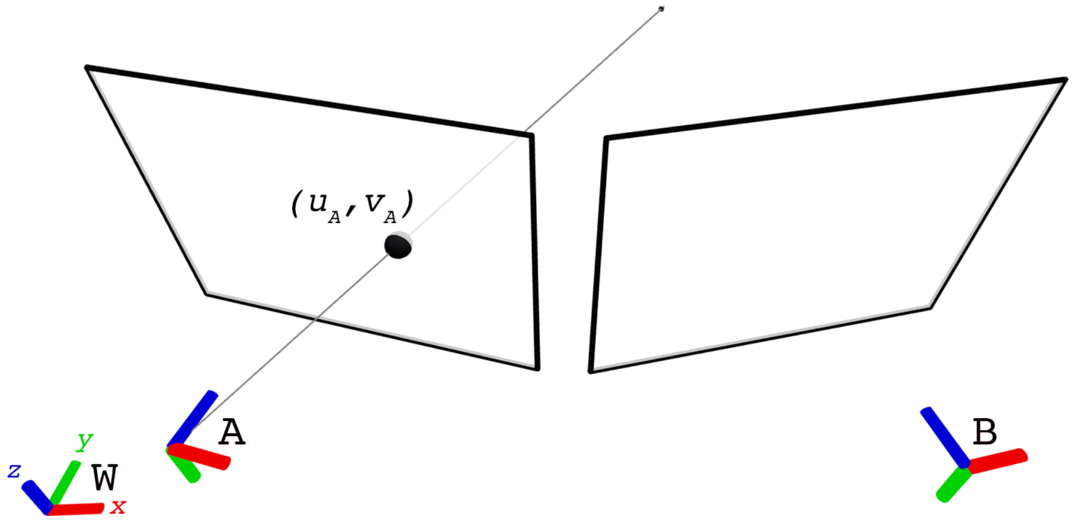
Therefore: $R' = R^\top$, $t' = -R't$:

$$T^{-1} = \begin{bmatrix} R^\top & -R^\top t \\ 0 & 1 \end{bmatrix}$$

2 Rays and Epipolar Geometry

3D computer vision algorithms typically begin with 2D observations, like coordinates extracted from corner and feature detectors.

We once again have cameras A and B, located in the world frame W. Consider the case where we observe a feature in the image from camera A, located at 2D coordinates (u_A, v_A):



The pixel at (u_A, v_A) corresponds to the projection of an unknown 3D point in the scene, p.

(a) Annotate the figure above with 3 possible locations for p.

(b) Annotate the image plane of camera B with the projection of your points from part (a). What pattern do these projected 2D points follow?

They should all lie on the same line. The epipolar line!

(c) Given (u_A, v_A) and intrinsics $K \in \mathbb{R}^{3\times 3}$, write an expression for all possible values of $p_A \in \mathbb{R}^3$.

From perspective projection, we know

$$\begin{aligned}\begin{bmatrix}u_A & v_A & 1\end{bmatrix}^\top &\sim Kp_A \\ p_A &\sim K^{-1}\begin{bmatrix}u_A & v_A & 1\end{bmatrix}^\top,\end{aligned}$$

therefore

$$p_A = \lambda K^{-1}\begin{bmatrix}u_A & v_A & 1\end{bmatrix}^\top$$

for any depth value $\lambda \geq 0$.

(d) Given (u_A, v_A), $K \in \mathbb{R}^{3\times3}$, T_WA, and T_WB, what are all possible coordinates for $p_\text{B} \in \mathbb{R}^3$?

We premultiply the answer from part (c) with T_BA

$$\begin{bmatrix} p_\text{B} \\ 1 \end{bmatrix} \sim T_\text{WB}^{-1} T_\text{WA} \begin{bmatrix} \lambda K^{-1} \begin{bmatrix} u_\text{A} & v_\text{A} & 1 \end{bmatrix}^\top \\ 1 \end{bmatrix}$$

for any depth value λ.

(e) Given the same inputs as part (d), how would you find all possible values of the point projected onto the image of camera B, (u_B, v_B)? Describe an approach in words.

You can project a pair of points from part (d) into the image plane of camera B, connect them with a line, and then write out an expression that contains points along that line.