Lecture: Hands-on AI Based 3D Vision
Assignment 02

Due: May 27th, 12:00am

1 Epipolar Geometry Conventions

In epipolar geometry, there are two cameras C_0 and C_1, viewing the same scene from different viewpoints. C_0's coordinate system is assumed to be aligned with the world coordinate system.
In the lecture we used the following conventions for deriving the epipolar constraint:

- $\mathbf{R}$ is the rotation matrix that maps vectors from camera 1's frame to camera 0's frame.
- $\mathbf{t}$ is the position of camera 1 expressed in world frame (= camera ='s frame)

It follows, that the projection matrices of the cameras are

$$\mathbf{P}_0 = \mathbf{K}_0 \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \qquad \qquad \mathbf{P}_1 = \mathbf{K}_1 \begin{bmatrix} \mathbf{R}^T & -\mathbf{R}^T\mathbf{t} \end{bmatrix}$$

In this scenario, the epipolar constraint can be written as

$$\mathbf{u}_0^T(\mathbf{K}_0^{-T}[\mathbf{t}]_\times \mathbf{R}\mathbf{K}_1^{-1})\mathbf{u}_1 = 0$$

or in terms of normalized coordinates

$$\mathbf{x}_0^T \underbrace{([\mathbf{t}]_\times \mathbf{R})}_{\mathbf{E}} \mathbf{x}_1 = 0$$

In other sources (such as Hartley-Zisserman), you will find different conventions. Here $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$ denote camera 1's extrinsics, i.e.

$$\hat{\mathbf{P}}_1 = \mathbf{K}_1 \begin{bmatrix} \hat{\mathbf{R}} & \hat{\mathbf{t}} \end{bmatrix}$$

It follows that $\hat{\mathbf{R}} = \mathbf{R}^T$ and $\hat{\mathbf{t}} = -\mathbf{R}^T\mathbf{t}$. They then derive the epipolar constraint as

$$\mathbf{x}_1^T \underbrace{([\hat{\mathbf{t}}]_\times \hat{\mathbf{R}})}_{\hat{\mathbf{E}}} \mathbf{x}_0 = 0.$$

Note the swapping of $\mathbf{x}_0$ and $\mathbf{x}_1$!

(a) Show that the two conventions are equivalent, i.e. $\mathbf{E} = \hat{\mathbf{E}}^T$.
Hint: You may use the fact that $[\mathbf{R}\mathbf{v}]_\times = \mathbf{R}[\mathbf{v}]_\times \mathbf{R}^T$ *for rotation matrices* $\mathbf{R}$.

Hartley and Zisserman derive the following theorem (cf. Result 9.19)

Theorem 1 *For a given essential matrix* $\hat{\mathbf{E}} = \hat{\mathbf{U}}\,\mathrm{diag}(1,1,0)\,\hat{\mathbf{V}}^T$*, the extrinsics of camera* C_1 *are any combination of:*

- $\hat{\mathbf{t}} = \pm\hat{\mathbf{u}}_3$
- $\hat{\mathbf{R}} = \hat{\mathbf{U}}\mathbf{W}\hat{\mathbf{V}}^T$ *or* $\hat{\mathbf{U}}\mathbf{W}^T\hat{\mathbf{V}}^T$

where

$$\mathbf{W} = \begin{bmatrix} & -1 & \\ 1 & & \\ & & 1 \end{bmatrix}$$

(b) Show that from this we can derive the theorem for our convention, as shown in the lecture:

Theorem 2 *For a given essential matrix* $\mathbf{E} = \mathbf{U}\,\mathrm{diag}(1,1,0)\,\mathbf{V}^T$*, the extrinsics of camera* C_1 *are any combination of:*

- $\mathbf{t} = \pm\mathbf{u}_3$
- $\mathbf{R} = \mathbf{U}\mathbf{W}\mathbf{V}^T$ *or* $\mathbf{U}\mathbf{W}^T\mathbf{V}^T$