Computer Vision II: Multiple-view Geometry

Exercise 08

25th June, 2025

Robust Odometry Estimation for RGB-D Cameras

- Given two images I_1 and I_2, and the depth maps (Z_1 and Z_2)
  - The goal is to find a transformation which maps $I_2 \rightarrow I_1$
- Towards that, we have $I_1(x) = I_2\big(\tau(\xi, x)\big)$,
  - where, τ is a warping function and $\xi \in se(3)$ is the lie algebra defining the camera motion and x is a 2D pixel location **in the first image, $\boldsymbol{I_1}$**

Warping Function

- Step 1. We first unproject the 2D location x (from I_1 as)
  - $p = \pi^{-1}(x, Z_1(x)) = Z_1(x)\left(\frac{u+c_x}{f_x}, \frac{v+c_y}{f_y}, 1\right)^T$,
  - where, $K = \begin{pmatrix} f_x & 0 & -c_x \\ 0 & f_y & -c_y \\ 0 & 0 & 1 \end{pmatrix}$ is the intrinsic matrix.
  - Question: Is p now in world coordinate system? Of which camera?

Warping Function

- Step 2. The transformation function $\mathrm{T}(g, p)$ transforms the point p to I_2's coordinate system
  - $\mathrm{T}(g, p) = Rp + t$, where, $g(\xi) = \exp(\hat{\xi}) = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$ is the transformation matrix formed using $\xi \in se(3)$
  - Question: what is the dimensionality of ξ?
- Step 3. Next, we project $\mathrm{T}(g, p)$ to the image space of I_2 as $\pi(\mathrm{T}(g, p)) = \left(\frac{f_x x}{z} - c_x, \frac{f_y y}{z} - c_y\right)$
- Therefore, $\tau(\xi, x) = \pi\big(T(g(\xi), p)\big) = \pi\big(\, T(\, g(\xi),\ \pi^{-1}(x, z_1(x))\,)\,\big)$

Residual

- Our goal is to find the parameter ξ, therefore, we optimize to reduce the residual between the two images, $r_i(\xi) = I_2\big(\tau(\xi, x_i)\big) - I_1(x_i)$, for the i^{th} pixel.

- We assume that all the pixels are i.i.d. and the likelihood of the ξ over the entire image is $p(r \mid \xi) = \Pi_i(p(r_i|\xi)$

Maximum A Posteriori (MAP)

- Using Bayes' rule, $p(\xi|r) = \frac{p(r|\xi)p(\xi)}{p(r)}$
- $\xi_{MAP} = \underset{\xi}{argmax}\ p(\xi|r)$
$$= \underset{\xi}{argmax}\ p(r|\xi)p(\xi) = \underset{\xi}{argmin}\ -\Sigma_i \log p(r_i|\xi) - \log p(\xi)$$
- Minimizing the term, we obtain the optimal ξ, the transformation between the two images