

robotics hw4 cv

余承諺 r12922135

2023 / 11 / 29

Extend camera matrix to homogeneous form 4×4 and denote as CameraMat. Objects position in base coordination, gripper coordination and image coordination are denoted as obj_base, obj_gripper and obj_img respectively. Under the pinhole camera, we can write the projection relation from base coordination to image coordination as.

$$\begin{bmatrix} \text{obj_img_w} \times \text{obj_img_d} \\ \text{obj_img_h} \times \text{obj_img_d} \\ \text{obj_img_d} \\ 1 \end{bmatrix} = \text{CameraMat} \times T_{\text{gripper to camera}} \times T_{\text{base to gripper}} \begin{bmatrix} \text{obj_base_x} \\ \text{obj_base_y} \\ \text{obj_base_z} \\ 1 \end{bmatrix}$$

We have already known that

$$\text{obj_img_d} \times \sqrt{\text{obj_area}} = \text{obj_alpha} \quad \text{obj_img_d} = \text{obj_alpha} \times \text{obj_area}^{-0.5}$$

obj_alpha only depends on object. Then

$$\begin{aligned} & \begin{bmatrix} \text{obj_img_w} \times \text{obj_alpha} \times \text{obj_area}^{-0.5} \\ \text{obj_img_h} \times \text{obj_alpha} \times \text{obj_area}^{-0.5} \\ \text{obj_alpha} \times \text{obj_area}^{-0.5} \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \text{obj_alpha} & 0 & 0 & 0 \\ 0 & \text{obj_alpha} & 0 & 0 \\ 0 & 0 & \text{obj_alpha} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \text{obj_img_w} \times \text{obj_area}^{-0.5} \\ \text{obj_img_h} \times \text{obj_area}^{-0.5} \\ \text{obj_area}^{-0.5} \\ 1 \end{bmatrix} \\ &= \text{CameraMat} \times T_{\text{gripper to camera}} \times T_{\text{base to gripper}} \begin{bmatrix} \text{obj_base_x} \\ \text{obj_base_y} \\ \text{obj_base_z} \\ 1 \end{bmatrix} \end{aligned}$$

Merge obj_alpha, CameraMat and $T_{\text{gripper to camera}}$ to obj_M. Then

$$T_{\text{base to gripper}} \begin{bmatrix} \text{obj_base_x} \\ \text{obj_base_y} \\ \text{obj_base_z} \\ 1 \end{bmatrix} = \text{obj_M} \begin{bmatrix} \text{obj_img_w} \times \text{obj_area}^{-0.5} \\ \text{obj_img_h} \times \text{obj_area}^{-0.5} \\ \text{obj_area}^{-0.5} \\ 1 \end{bmatrix}$$

Regressing to acquire obj_M. The regression process needs N cases ($N \geq 4$). Each case contains: 1. a captured image, 2. the transformation from base coordination to gripper coordination when the image is captured, 3. object locations in base coordination.

When inferring, $T_{\text{base to gripper}}$ and obj_M show be known. obj_img and bj_area also can be detect from image. Then

$$\begin{bmatrix} \text{obj_base_x} \\ \text{obj_base_y} \\ \text{obj_base_z} \\ 1 \end{bmatrix} = T_{\text{base to gripper}}^{-1} \text{obj_M} \begin{bmatrix} \text{obj_img_w} \times \text{obj_area}^{-0.5} \\ \text{obj_img_h} \times \text{obj_area}^{-0.5} \\ \text{obj_area}^{-0.5} \\ 1 \end{bmatrix}$$