robotics hw4 cv2

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

$$\text{obj\_img\_d} \times \sqrt{\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}}} = \text{alpha} \qquad \text{obj\_img\_d} = \text{alpha} \times (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5}$$

alpha only depends on object. Then

$$\begin{bmatrix} \text{obj\_img\_w} \times \text{alpha} \times (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ \text{obj\_img\_h} \times \text{alpha} \times (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ \text{alpha} \times (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} \text{alpha} & 0 & 0 & 0 \\ 0 & \text{alpha} & 0 & 0 \\ 0 & 0 & \text{alpha} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \text{obj\_img\_w} \times (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ \text{obj\_img\_h} \times (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ 1 \end{bmatrix}$$
$$= \text{CameraMat} \times T_{\text{gripper to camera}} \times T_{\text{base to gripper}} \begin{bmatrix} \text{obj\_base\_x} \\ \text{obj\_base\_y} \\ \text{obj\_base\_z} \\ 1 \end{bmatrix}$$

Merge alpha, CameraMat and $T_{\text{gripper to camera}}$ to M_camera_to_gripper. Then

$$T_{\text{base to gripper}} \begin{bmatrix} \text{obj\_base\_x} \\ \text{obj\_base\_y} \\ \text{obj\_base\_z} \\ 1 \end{bmatrix} = \text{M\_camera\_to\_gripper} \begin{bmatrix} \text{obj\_img\_w} \times (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ \text{obj\_img\_h} \times (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ (\tfrac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ 1 \end{bmatrix}$$

Regressing to acquire M_camera_to_gripper. 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 M_camera_to_gripper show be known. obj_img and obj_area also can be detect from image. Then

$$\begin{bmatrix} \text{obj\_base\_x} \\ \text{obj\_base\_y} \\ \text{obj\_base\_z} \\ 1 \end{bmatrix} = T^{-1}_{\text{base to gripper}} \text{M\_camera\_to\_gripper} \begin{bmatrix} \text{obj\_img\_w} \times (\frac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ \text{obj\_img\_h} \times (\frac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ (\frac{\text{obj\_area}}{\text{obj\_area\_factor}})^{-0.5} \\ 1 \end{bmatrix}$$