Robotics, Vision, & Mechatronics for Manufacturing.

Pose Estimation

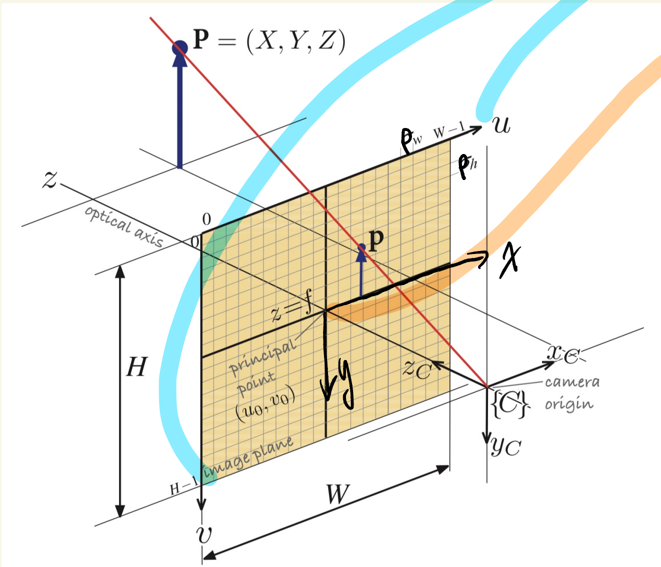
(Chap 11 in Peter Corke)

Xu Chen

Sp 2021

Recap:

Pixels & Discrete Image Plane



pixel axis: index u & v are nonnegative

principal point: geometric center (u_0, v_0)

⇒ coordinates in pixel domain:

$$u = \frac{x}{p_w} + u_0 \qquad v = \frac{y}{p_h} + v_0$$

p_w: pixel width $\qquad$ p_h: pixel height

$$\Rightarrow \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{1}{p_w} & 0 & u_0 \\ 0 & \frac{1}{p_h} & v_0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous form ⇒

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{p_w} & 0 & u_0 \\ 0 & \frac{1}{p_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \overbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}^{{}^C\tilde{P}}$$

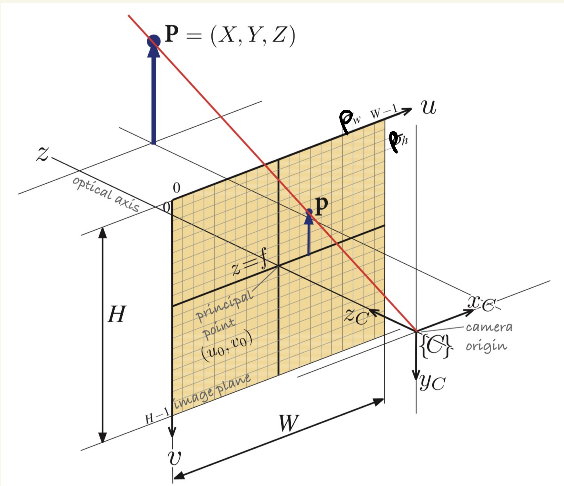
Introduce again homogeneous notation:

$$u = \frac{\tilde{u}}{\tilde{w}}, \quad v = \frac{\tilde{v}}{\tilde{w}} \qquad \tilde{w} = Z$$

normalizing ⟹

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} \triangleq Z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{p_w} & 0 & u_0 \\ 0 & \frac{1}{p_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \overbrace{Z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}} = K \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} {}^C\tilde{P}$$

Camera Matrix Revisited (the full camera matrix from global coordinate to pixel domain)



We have:

- pixel domain:

$$\tilde{p} = \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{\rho_w} & 0 & u_0 \\ 0 & \frac{1}{\rho_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{{}^{C}\tilde{p}}$$

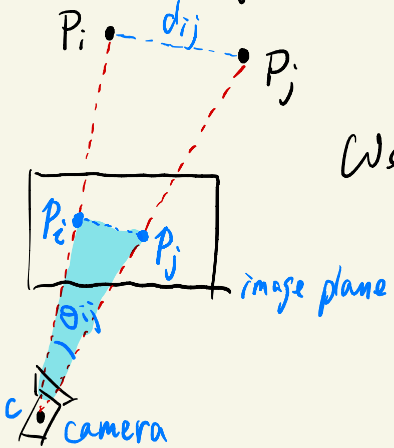
$$= \underbrace{\begin{bmatrix} \frac{f}{\rho_w} & 0 & u_0 \\ 0 & \frac{f}{\rho_h} & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\tilde{K}} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

Pose Estimation

$$|\overrightarrow{P_iP_j}|^2 = |\overrightarrow{cp_i}|^2 + |\overrightarrow{cp_j}|^2 - 2|\overrightarrow{cp_i}||\overrightarrow{cp_j}|\cos\theta_{ij}$$

Goal: based on 2D images of objects, determine 3D pose.

Basic Geometry:



P_i, P_j: reference points on the object

p_i, p_j: projection onto image plane

We know:

$$\underbrace{|\overrightarrow{P_iP_j}|^2}_{} = \underbrace{|\overrightarrow{CP_i}|^2}_{\text{unknown } ①} + \underbrace{|\overrightarrow{CP_j}|^2}_{\text{unknown } ②} - 2\underbrace{|\overrightarrow{CP_i}|}_{①}\underbrace{|\overrightarrow{CP_j}|}_{②}\underbrace{\cos\theta_{ij}}_{}$$

can be computed from image plane data in triangle

$$\breve{p} = \breve{K}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

can be obtained from perspective projection

$$|\overrightarrow{P_iP_j}|^2 = \left\| \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} - \begin{bmatrix} X_j \\ Y_j \\ Z_j \end{bmatrix} \right\|^2 = \left\| \breve{K}^{-1}\breve{p}_i - \breve{K}^{-1}\breve{p}_j \right\|^2$$

$$= \left(\breve{K}^{-1}(\breve{p}_i - \breve{p}_j)\right)^T \breve{K}^{-1}(\breve{p}_i - \breve{p}_j)$$

$$= \underline{(\breve{p}_i - \breve{p}_j)}^T \underline{(\breve{K}^{-1})^T \breve{K}^{-1}} \underline{(\breve{p}_i - \breve{p}_j)}$$

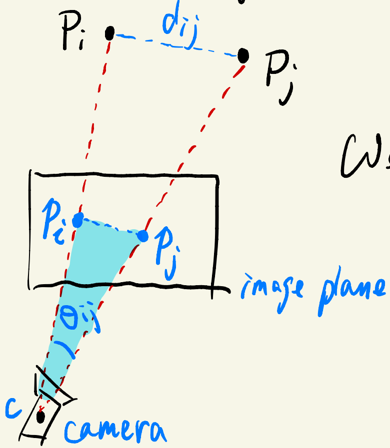
from image plane

camera intrinsic parameters

Pose Estimation

Goal: based on 2D images of objects, determine 3D pose.

Basic Geometry:



P_i, P_j: reference points on the object

p_i, p_j: projection onto image plane

We know:

$$\underbrace{|\overrightarrow{P_iP_j}|^2}_{\substack{\text{computed}\\ d_{ij}}} = \underbrace{|\overrightarrow{CP_i}|^2}_{\substack{\text{unknown } ①\\ \triangleq x_i}} + \underbrace{|\overrightarrow{CP_j}|^2}_{\substack{\text{unknown } ②\\ \triangleq x_j}} - 2\underbrace{|\overrightarrow{CP_i}|}_{①}\underbrace{|\overrightarrow{CP_j}|}_{②}\underbrace{\cos\theta_{ij}}_{\text{computed}}$$

for brevity, define $f_{ij}(x_i, x_j) = x_i^2 + x_j^2 - 2x_ix_j\cos\theta_{ij} - d_{ij}$

Take more reference points: n point $\Rightarrow C_n^2 = \frac{n(n-1)}{2}$ equations

$$\begin{cases} f_{12}(x_1, x_2) = 0 \\ f_{13}(x_1, x_3) = 0 \\ f_{23}(x_2, x_3) = 0 \\ \vdots \end{cases}$$

solution concept: each unknown x_i corresponds to one added point

turns out the equation set can be neatly solved.

References: Long Quan & Zhong-Dan Lan, "Linear N-point Camera Pose Estimation", 1999

& Gao et. al., "Complete Solution Classification for the Perspective 3-point Problem, 2003