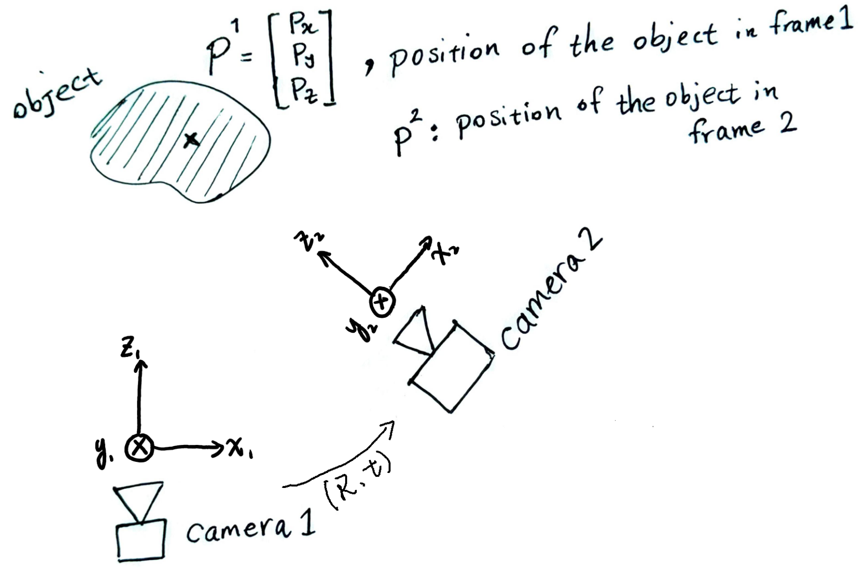
object
$p^1 = \begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix}$, position of the object in frame1
p^2 : position of the object in frame 2
z_2
x_2
y_2
camera 2
z_1
y_1
x_1
(R, t)
camera 1

Figure 1: The problem setup for the object's position estimation using two monocular cameras.

Problem Setup

We are given a setup, shown in Figure 1, to estimate the 3D position of an object's center. The object is stationary, and there are also two fixed monocular cameras in the room. A previously developed object detection algorithm can provide the object's center in the 2D image coordinates. However, the depth is not observable by a single RGB image. Therefore, our goal is to combine measurements from two cameras to estimate the object's 3D position in the frame of camera 1. The relative pose (orientation and translation) between the two cameras is accurately known and given.

One of your colleagues (who is good at math and modeling!) has already formulated the problem. But he has very poor knowledge of state estimation, which is why he reached out to you for the rest of the solution. Luckily, you are taking the Mobile Robotics class, and you can help him out this time, so he doesn't lose his job! So far, we have the following.

Measurement Models

From an elementary computer vision knowledge, we know that a monocular pinhole camera model takes the following form

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \cdot \frac{1}{p_z} \cdot \begin{bmatrix} p_x \\ p_y \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix} = K_f \pi(p) + c,$$

where $K_f = \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix}$ is called intrinsic camera matrix, $\begin{bmatrix} c_x \\ c_y \end{bmatrix}$ is the coordinates of the optical center of the image, and $\pi(\cdot)$ is the projection function (it takes a point and divides its coordinates by z, i.e., the last coordinate).

We denote the object's center in frame 1 by ${}^1p$ and ${}^2p$ as the object's center in frame 2. Given the rotation matrix, R, and the translation vector, t of frame 2 with respect to frame 1, we have the following relationship for the object's center between the two frames.

$${}^1p = R \cdot {}^2p + t,$$

and

$$^2p = R^\top \cdot {}^1p - R^\top t.$$

Combining everything so far, we have the following measurement models for camera 1

$$z = \begin{bmatrix} u \\ v \end{bmatrix} = {}^1K_f \pi({}^1p) + {}^1c := h_1({}^1p),$$

and camera 2

$$z = \begin{bmatrix} u \\ v \end{bmatrix} = {}^2K_f \pi({}^2p) + {}^2c = {}^2K_f \pi(R^\top \cdot {}^1p - R^\top t) + {}^2c := h_2({}^1p).$$

We also assume each measurement model is corrupted by a zero-mean white Gaussian noise.

$$\begin{aligned} z_1 &= h_1({}^1p) + v_1, \quad v_1 \sim \mathcal{N}(0, \Sigma_{v_1}), \\ z_2 &= h_2({}^1p) + v_2, \quad v_2 \sim \mathcal{N}(0, \Sigma_{v_2}). \end{aligned}$$

we may also stack two synchronized observations to form a stacked observation model as follows.

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}_{4\times1} = \begin{bmatrix} h_1({}^1p) \\ h_2({}^1p) \end{bmatrix}_{4\times1} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_{4\times1} := h({}^1p) + v,$$

where now $v \sim \mathcal{N}\left(0_{4\times1}, \text{blkdiag}(\Sigma_{v_1}, \Sigma_{v_2})\right)$, and blkdiag$(\cdot)$ forms a block diagonal matrix.

Motion Model

Furthermore, camera 1 is attached to a structure that can vibrate. The vibration is not so severe that we assume the camera moves, but to account for inaccuracies caused by the fixture vibration we use a discrete-time random walk process as its motion model.

$${}^1p_{k+1} = {}^1p_k + w, \quad w \sim \mathcal{N}(0, \Sigma_w).$$

Jacobians

The Jacobians are also given using the chain rule and the fact that if $q = Rp + t$, then $\frac{\partial q}{\partial p} = R$ (you may use a symbolic math software to compute the Jacobians as well). In the following we use ${}^1p = \begin{bmatrix} p_x & p_y & p_z \end{bmatrix}^\top$ and ${}^2p = \begin{bmatrix} q_x & q_y & q_z \end{bmatrix}^\top$.

$$H_1 = \frac{\partial h_1}{\partial\, {}^1p} = K_f \frac{\partial \pi}{\partial\, {}^1p} = K_f \begin{bmatrix} \frac{1}{p_z} & 0 & -\frac{p_x}{p_z^2} \\ 0 & \frac{1}{p_z} & -\frac{p_y}{p_z^2} \end{bmatrix},$$

$$H_2 = \frac{\partial h_2}{\partial\, {}^1p} = K_f \frac{\partial \pi}{\partial\, {}^1p} = K_f \frac{\partial \pi}{\partial\, {}^2p} \cdot \frac{\partial\, {}^2p}{\partial\, {}^1p} = K_f \begin{bmatrix} \frac{1}{q_z} & 0 & -\frac{q_x}{q_z^2} \\ 0 & \frac{1}{q_z} & -\frac{q_y}{q_z^2} \end{bmatrix} R^\top,$$

and the stacked measurement Jacobian can also be constrcuted as

$$H_{4\times3} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}.$$