1 Theory

1.1 Pinhole camera model

Without loss of generality, the pinhole is always located at $(0, 0, 0)$. Denote the position of an object by $\vec{r} = (x, y, z)$, its velocity by $\vec{v} = (v_x, v_y, v_z)$, its acceleration by $\vec{a} = (a_x, a_y, a_z)$. The screen is located at a vector $\vec{h_s}$ from the origin, its screen coordinates basic vectors are $\hat{h_0}$, $\vec{h_1}$ and $\vec{h_2}$. Note that

$$\vec{h_0} \perp \vec{h_1} \perp \vec{h_2} \tag{1}$$

$$\begin{pmatrix} \hat{h_0} \\ \hat{h_1} \\ \hat{h_2} \end{pmatrix} = T \begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} \tag{2}$$

where T is an orthonormal matrix.

Ideally, $\hat{h_0} // \vec{h_s}$. But there might be a small angle between them. Let's assume $\vec{h_s} = h_s \hat{h_0}$. The screen plane is given by

$$h_s \hat{h_0} + a\hat{h_1} + b\hat{h_2}, \quad a, b \in R \tag{3}$$

Then the projection of $\vec{r}$ on the screen, $\vec{r'} = (a, b)$, is given by

$$-k\vec{r} = h_s \hat{h_0} + a\vec{h_1} + b\vec{h_2} \tag{4}$$

$$\left(\vec{r}, \hat{h_1}, \hat{h_2}\right) \begin{pmatrix} k \\ a \\ b \end{pmatrix} = -h_s \hat{h_0} \tag{5}$$

or

$$\left(\hat{h_1}, \hat{h_2}, \vec{r}\right) \begin{pmatrix} a \\ b \\ k \end{pmatrix} = -h_s \hat{h_0}$$

Define

$$C^{-1} = \left(\hat{h_1}, \hat{h_2}, \vec{r}\right)^{-1} \tag{6}$$

as the inverse camera matrix. The screen coordinates of the object (a, b), as well as its distance k from the pinhole, is given by

$$\underline{\begin{pmatrix} a \\ b \\ k \end{pmatrix} = -C^{-1} h_s \hat{h_0}} \tag{7}$$

1.1.1 The characteristics of the screen coordinates

The world coordinates is denoted by $(\hat{x}, \hat{y}, \hat{z})$, where $\hat{z}$ is the upward vertical direction and $(\hat{x}, \hat{y})$ spans the ground. In typical vision, $\vec{h_0}$ is close to $\vec{h_s}$. And $\vec{h_2}$ is close to vertical, the angle between them being θ. Then

$$\hat{h_2}.\hat{z} = \cos\theta \approx 1 \tag{8}$$

$$T = (T_{ij}), 1 \leq i, j \leq 3$$

where $T_{31}, T_{32} \ll T_{33}$.

1.1.2 The solution to (4)

In

$$k\vec{r} + h_s\hat{h_0} + a\vec{h_1} + b\vec{h_2} = 0$$

Multiplying the equation by $\hat{h_i}$,$i = 0, 1, 2$:

$$k\left(\vec{r}.\hat{h_0}\right) = -h_s$$

$$k\left(\vec{r}.\hat{h_1}\right) + a = 0$$

$$k\left(\vec{r}.\hat{h_2}\right) + b = 0$$

Therefore

$$\underline{k = -\frac{h_s}{\vec{r}.\hat{h_0}}}, \quad a = -k\left(\vec{r}.\hat{h_1}\right) = \underline{h_s\frac{\vec{r}.\hat{h_1}}{\vec{r}.\hat{h_0}}}, \quad \underline{b = h_s\frac{\vec{r}.\hat{h_2}}{\vec{r}.\hat{h_0}}} \tag{9}$$

Note that if $\vec{r} \perp \hat{h_0}$,this object is out of the FOV and not on the screen.

1.2 Motion model

In a short time Δt, all three frameworks could move and rotate.

1. The object moves to $\vec{r} + \vec{v}\Delta t$,
2. The pinhole moves to $(0, 0, 0) + \vec{u}\Delta t$. We can safely assume $\vec{u}$ is very small, $\vec{u} \approx 0$.
3. The screen rotates from T to $T.\Delta T$, where ΔT is a "small" orthonormal matrix

then its projection on the screen is given by

$$\left(\hat{h_1}, \hat{h_2}, \vec{r} + (\vec{v} - \vec{u})\Delta t\right)\begin{pmatrix} a + \Delta a \\ b + \Delta b \\ k + \Delta k \end{pmatrix} = -h_s\hat{h_0}$$

According to Eq. (9), the new projection point is

$$a + \Delta a = h_s \frac{(\vec{r} + \Delta\vec{r}) . \left(\hat{h_1} + \Delta\hat{h_1}\right)}{(\vec{r} + \Delta\vec{r}) . \left(\hat{h_0} + \Delta\hat{h_0}\right)} \tag{10}$$

$$b + \Delta b = h_s \frac{(\vec{r} + \Delta\vec{r}) . \left(\hat{h_2} + \Delta\hat{h_2}\right)}{(\vec{r} + \Delta\vec{r}) . \left(\hat{h_0} + \Delta\hat{h_0}\right)} \tag{11}$$

Let $\Delta t \to 0$, the velocities of a, b are given by

$$\frac{da}{dt} = h_s \frac{\left(\vec{r}'.\hat{h_1} + \vec{r}.\hat{h}_1'\right)\left(\vec{r}.\hat{h_0}\right) - \left(\vec{r}.\hat{h_1}\right).\left(\vec{r}'.\hat{h_0} + \vec{r}.\hat{h}_0'\right)}{\left(\vec{r}.\hat{h_0}\right)^2} = h_s \frac{U}{\left(\vec{r}.\hat{h_0}\right)^2} \tag{12}$$

$$U = \vec{r}.\left[\left(\vec{r}'.\hat{h_1} + \vec{r}.\hat{h}_1'\right)\hat{h_0} - \left(\vec{r}'.\hat{h_0} + \vec{r}.\hat{h}_0'\right).\hat{h_1}\right]$$

If we simply let $\hat{h}_1' = \hat{h}_0' = 0$, i.e, the screen and the pinhole do not move, then

$$U = \vec{r}.\left[\left(\vec{r}'.\hat{h_1}\right)\hat{h_0} - \left(\vec{r}'.\hat{h_0}\right)\hat{h_1}\right]$$

$$U = \vec{r}.\left[\left(\vec{v}.\hat{h_1}\right)\hat{h_0} - \left(\vec{v}.\hat{h_0}\right)\hat{h_1}\right] \tag{13}$$

Recall the *vector triple product* identity, $\vec{a} \times \left(\vec{b} \times \vec{c}\right) = (\vec{a}.\vec{c})\,\vec{b} - \left(\vec{a}.\vec{b}\right)\vec{c}$, then (13) becomes

$$U = \vec{r}.\left(\vec{v} \times \left(\vec{h_0} \times \vec{h_1}\right)\right) = \vec{r}.\left(\vec{v} \times \vec{h_2}\right) \tag{14}$$

$$\boxed{\frac{da}{dt} = h_s \frac{\vec{r}.\left(\vec{v} \times \vec{h_2}\right)}{\left(\vec{r}.\hat{h_0}\right)^2}} \tag{15}$$

$$\boxed{\frac{db}{dt} = -h_s \frac{\vec{r}.\left(\vec{v} \times \vec{h_1}\right)}{\left(\vec{r}.\hat{h_0}\right)^2}} \tag{16}$$

Note that

$$\vec{A}.\left(\vec{B} \times \vec{C}\right) = \left(\vec{A}, \vec{B}, \vec{C}\right) = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix}$$

1.3 Network

1. Yolo: gives the bounding box of the ball in a frame, (a_1, b_1), (a_2, b_2), denoted by

$$B = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{pmatrix}$$

2. Physical network (PNN):

   (a) input: a series of bounding boxes in consecutive frames, B_1, B_2, ..., B_m

2 Appendix

2.1 Matrix op

Let A^{-1} be the inverse matrix of A. Consider the inverse matrix of $A + dA$, where $|dA| \ll |A|$.

$$A + dA = A(I + A^{-1}dA)$$

Let $dB = A^{-1}dA$, then

$$(I - dB)\, A^{-1}.A\,(I + dB) = (I - dB)\,(I + dB) = I - (dB)^2 \approx I$$

Therefore

$$(A + dA)^{-1} \approx \boxed{A^{-1} - A^{-1}\,(dA)\,A^{-1}}$$

2.2 Differential vectors

For a dot product of two vectors $u = \vec{x}.\vec{y}$, its derivative is

$$\frac{du}{dt} = \frac{d}{dt}\,(\vec{x}.\vec{y}) = \frac{d\vec{x}}{dt}.\vec{y} + \vec{x}.\frac{d\vec{y}}{dt}$$