Course: MAE 4810/5810 Robot Perception
Instructor: Silvia Ferrari

Pinhole Camera Model and Field-of-view (FOV)

A controllable pan-tilt (PT) camera can be used to monitor a two-dimensional (2D) workspace over a time interval $(T_0, T_f]$, $\mathcal{W} \subset \mathbb{R}^2$, where,

$$\mathbf{x}_T = [x_T \quad y_T]^T \in \mathcal{W} \tag{1}$$

denotes the position of one (point) target defined in inertial XY-coordinate frame. As illustrated in Fig. 1, the PT camera is characterized by a dynamic and bounded field-of-view (FOV), $\mathcal{S}(t) \subset \mathcal{W}$, and thus obtains target measurements at a time t only if $\mathbf{x}_T(t) \in \mathcal{S}(t)$. By controlling the camera pan and tilt angles, the motion and shape of the camera FOV can be automatically adjusted so as to obtain the most informative target measurements.

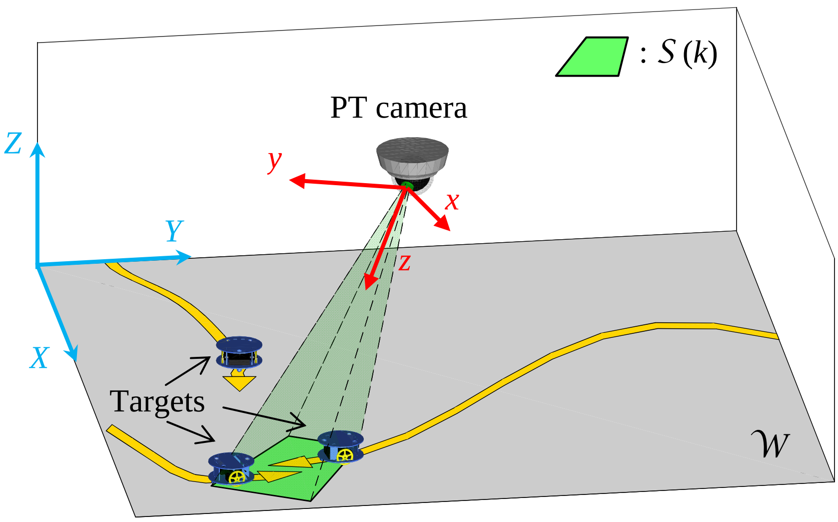

Figure 1: Example of pan-tilt (PT) camera observing three mobile ground targets in the workspace $\mathcal{W}$.

Based on the so-called *pinhole camera model*, the camera lens is symmetric about an optical axis, and images in $\mathcal{W}$ are projected onto a 2D virtual image plane perpendicular to the optical axis and located at a distance λ from the pinhole, as shown in Fig. 2. The distance λ between the virtual image plane and the pinhole is also known as focal length. A camera-fixed frame of reference (Fig. 2) is defined by placing the origin at the pinhole, aligning the z-axis with the optical axis and the x-axis parallel to the XY-plane, and obtaining the y-axis by the right-hand rule.

Then, the camera pan and tilt angles can be represented by the Euler angles known as yaw (ψ) and roll (ϕ) angles, respectively, and illustrated in Fig. 3. As explained in class, the yaw and roll Euler angles are defined by

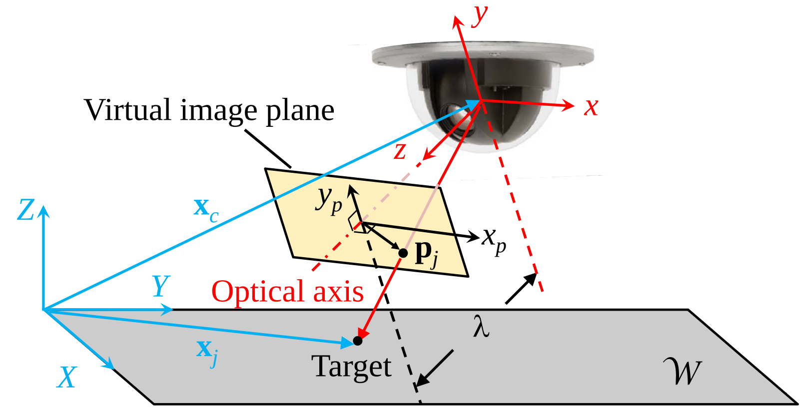

Figure 2: Pinhole camera model.

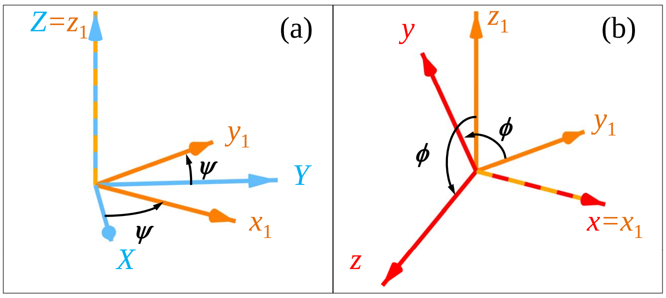

Figure 3: Sequence of yaw and roll Euler angle rotations comprised of the camera pan angle ψ, and tilt angle ϕ, respectively.

two successive right-hand rotations: a rotation by an angle ψ about the Z-axis, leading to the intermediate frame (x_1, y_1, z_1) (Fig. 3.a), followed by a rotation by an angle ϕ about the x_1-axis, leading to the camera-fixed frame (x, y, z) (Fig. 3.b). Then, any vector $\mathbf{x}$ in inertial frame can be resolved into the camera-fixed frame by the following transformation,

$$\mathbf{x}_c = \mathbf{R}_\phi \mathbf{R}_\psi \ \mathbf{x} \tag{2}$$

defined in terms of the Euler rotation matrices

$$\mathbf{R}_\phi \triangleq \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix}, \ \mathbf{R}_\psi \triangleq \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Consider now a 2D frame of reference (x_p, y_p) embedded in the virtual image plane, such that its origin lies at the intersection between the virtual image plane and the optical axis, the x_p-axis is chosen parallel to XY-plane, and the y_p-axis is orthogonal to the x_p-axis and the optical axis, as shown in Fig. 2. Then, from the pinhole camera model, it can be shown that the projection of any target state $\mathbf{x}_T$ onto the virtual image plane with respect to the (x_p, y_p)-frame is,

$$\mathbf{p}_T = \lambda \,[q_x/q_z \quad q_y/q_z]^T \triangleq [p_x \quad p_y]^T \tag{3}$$

where the position of the target with respect to camera-fixed frame is,

$$\mathbf{q}_T = \mathbf{R}_\phi^T \mathbf{R}_\psi^T \left([\mathbf{x}_T^T \; 0]^T - \mathbf{x}_c\right) \triangleq [q_x \quad q_y \quad q_z]^T \tag{4}$$

and $\mathbf{x}_c$ is the pinhole position with respect to the inertial frame.

For a moving target, the projection of the velocity,

$$\mathbf{v}_T = \frac{d\mathbf{x}_T}{dt} \triangleq \dot{\mathbf{x}}_T \tag{5}$$

onto the moving camera virtual image plane is found by taking the time derivative of both sides of (3), as follows,

$$\dot{\mathbf{p}}_T = [\dot{p}_x \quad \dot{p}_y]^T = \mathbf{H} \begin{bmatrix} \mathbf{R}_\phi^T \mathbf{R}_\psi^T & \mathbf{0} \\ \mathbf{0} & -\mathbf{R}_\phi^T \end{bmatrix} [\dot{\mathbf{x}}_T^T \quad 0 \quad \dot{\phi} \quad 0 \quad \dot{\psi}]^T \tag{6}$$

where,

$$\mathbf{H} \triangleq \begin{bmatrix} -\frac{\lambda}{q_z} & 0 & \frac{p_x}{q_z} & \frac{p_x p_y}{\lambda} & -\frac{\lambda^2 + p_x^2}{\lambda} & p_y \\ 0 & -\frac{\lambda}{q_z} & \frac{p_y}{q_z} & \frac{\lambda^2 + p_x^2}{\lambda} & -\frac{p_x p_y}{\lambda} & -p_x \end{bmatrix} \tag{7}$$

is the image Jacobian matrix.

Because the geometry of the FOV, $\mathcal{S}(t)$, changes with respect to the camera state, it is easier to derive the analytical form of the geometric constraints by projecting $\mathbf{x}_T(t)$ and $\mathcal{S}(t)$ onto the virtual image plane, as illustrated in Fig. 2. From (3)-(4), the projection of $\mathbf{x}_T(t)$ onto the virtual image plane, $\mathbf{p}_T = [p_x \quad p_y]^T$, is obtained by the pinhole camera model. In addition, the projection of $\mathcal{S}(t)$ onto the virtual image plane is a rectangle with the same size as the image sensor. Let a and b denote the width and height of the image sensor, respectively. It follows that the target lies in the camera FOV in $\mathcal{W}$ if and only if $\mathbf{p}_T$ is in the image sensor, that is,

$$\mathbf{x}_T(t) \in \mathcal{S}(t) \Leftrightarrow \mathbf{p}_T \in [-a/2, \; a/2] \times [-b/2, \; b/2] \tag{8}$$

For simplicity, in this document, the camera measurements are assumed to be noise free. The sampling interval Δt is a known finite constant, such that the time interval $(T_0, T_f]$ can be discretized and indexed by $k = 0, 1, \ldots, k_f$. Then, the target position and velocity can be measured (are fully observable) provided the target is in the camera FOV, i.e.:

$$\mathbf{z}(k) = \begin{cases} \mathbf{h}[\mathbf{x}_T(k), \mathbf{v}_T(k)], & \text{if } \mathbf{x}_T(k) \in \mathcal{S}(k) \\ \varnothing, & \text{if } \mathbf{x}_T(k) \notin \mathcal{S}(k) \end{cases} \tag{9}$$

where $\varnothing$ is the empty set. The vector function $\mathbf{h}[\cdot]$ in the measurement equation (9) can be derived using the above projections and transformations, by recognizing that the camera measurement vector consists of the target position and velocity in the x_p, y_p-frame, or $\mathbf{z}(k) = [\mathbf{p}_T^T(k) \quad \dot{\mathbf{p}}_T^T(k)]^T$. Therefore, the functional elements of $\mathbf{h}[\cdot]$ are given by (3)-(6), completing the definition of the PT camera measurement equations.

The PT camera kinematic equations can be obtained from the kinematic constraints on the motor and pinhole or lens movements, which determine the motion constraints on the FOV, $\mathcal{S}$. The FOV shape changes based on the orientation of the camera with respect to $\mathcal{W}$. Let $\mathbf{s} = [\psi \quad \phi \quad \dot{\psi} \quad \dot{\phi}]^T$ denote the dynamic state of the camera, and $\mathbf{u} = [u_1 \quad u_2]^T$ denote the camera control input, where u_1 and u_2 are two voltage levels that, applied to the motors, can independently adjust the pan and tilt angles. Then, the camera kinematic equation can be expressed by a difference equation in state-space form,

$$\mathbf{s}(k+1) = \mathbf{A}\mathbf{s}(k) + \mathbf{B}\mathbf{u}(k) \tag{10}$$

where,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ b_1 & 0 \\ 0 & b_2 \end{bmatrix} \tag{11}$$

and b_1 and b_2 are two constant motor parameters (Table 1).

In addition to the kinematic constraint (10), the camera state and control must also obey inequality constraints that reflect physical bounds imposed by the instrumentation. The pan and tilt angles are constrained to the ranges $\psi \in [0,\ 2\pi)$ and $\phi \in [\pi/2,\ \pi]$, and the pan and tilt angular velocities are bounded by the constants $\dot{\psi}_{max}$ and $\dot{\phi}_{max}$ (Table 1), respectively. Then, by normalizing the camera input voltages such that their upper bounds are equal to one, the full set of camera inequality constraints can be expressed as,

$$\begin{cases} \mathbf{b}_1 \leq \mathbf{s} \leq \mathbf{b}_2 \\ |\mathbf{u}| \leq \mathbf{1}_2 \end{cases} \tag{12}$$

where $\mathbf{b}_1 = [0 \quad \pi/2 \quad -\dot{\psi}_{max} \quad -\dot{\phi}_{max}]^T$, $\mathbf{b}_2 = [2\pi \quad \pi \quad \dot{\psi}_{max} \quad \dot{\phi}_{max}]^T$, $\mathbf{1}_n$ denotes an $n \times 1$ vector of ones, and input scaling coefficients have been absorbed into the matrix $\mathbf{B}$ in (11).

Table 1: Parameters of PT Camera

Description	Variable	Value
Maximum pan angular velocity	$\dot{\psi}_m$	100°/s
Maximum tilt angular velocity	$\dot{\phi}_m$	100°/s
Motor coefficients	b_1, b_2	$100^\circ/(\mathrm{Vs}^2)$