Pan-tilt Camera Model and Target Tracking

Kyung Rak Jang

November 22, 2021

1 Abstract

This report presents an automatic target tracking system with a single pan-tilt camera. By using pinhole camera model, the model can simulate its field-of-view (FOV) on the camera position and orientation, as well as the target position in the region of interests (ROI). To keep the target in the center area of the camera's FOV, a simple PID controller is designed. After the movement of the target over time, the target is still in the center area of the camera's FOV. Therefore, the mathematical model and simulation introduced in this report demonstrate that the system can perform well in a workspace.

2 Introduction

Pan-tilt cameras (PT cameras) are used in specific spaces such as sporting events, video conferences, and broadcasting studios. PT cameras are also an essential part of surveillance systems because of its capability of remote control. PT cameras are characterized by a dynamic and bounded field-of-view (FOV)[1] that obtains the target measurements within limited range of angles. By adjusting the camera pan and tilt angles, the camera can detect the target in the camera's FOV. It can be also used to track the target automatically by keeping the target in the center of the camera's FOV.

This report demonstrates the ability to capture the moving target in the FOV and track the target within the bounded FOV by controlling the camera pan and tilt angles. A fixed-PT camera is designed as a surveillance camera. Pan angle of the camera initially rotates at given angular velocity, allowing the camera to find the target in the region of interests (ROI). Once the camera obtains target measurements, the motion of the camera FOV is automatically adjusted to keep the moving target in the center area of the camera's FOV. To control the pan and tilt angles, a PID controller is designed to continuously calculate error values and find desired angles.

Using numerical simulations in Python, the results in Section 5 show that the mathematical model of PT camera designed in this report finds and tracks the moving target efficiently. The camera simulation captures measurements' dependencies on camera state and target characteristics. Furthermore, the PID controller presented in Section 4 is effective at minimizing the error between a measured angle and a desired angle, allowing the camera to follow the moving target.

3 Mathematical Model of Pan-tilt Camera

As illustrated in Figure 1, it is assumed that the PT camera is operating in a workspace that is closed and bounded, denoted by $W \subset \mathbb{R}^3$. The workspace is populated with one (point) target defined in inertial XY-coordinate frame where, $X_T = [x_T, y_T]^T \in W$, and one PT camera where, $X_P = [x_P, y_P, z_P]^T \in W$. The camera is assumed to be fixed with respect to a camera-fixed frame $A \subset W$, and characterized by a dynamic and bounded FOV, $S(t) \subset W$. The camera lens is symmetric about an optical axis and FOV of the workspace can be imaged onto a 2D image plane which is located at the distance λ between the image plane and the position of the camera. The geometry of the camera FOV as well as the transformations of target coordinates between the pixel coordinates and an inertial frame can be obtained using the pinhole camera model [2].

The camera pan angle ψ and tilt angle ϕ can be represented by the Euler angle rotations, and the Euler rotation matrices can be defined as,

$$H_\psi = \begin{bmatrix} cos\psi & -sin\psi & 0 \\ sin\psi & cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}, H_\phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\phi & -sin\phi \\ 0 & sin\phi & cos\phi \end{bmatrix} \tag{1}$$

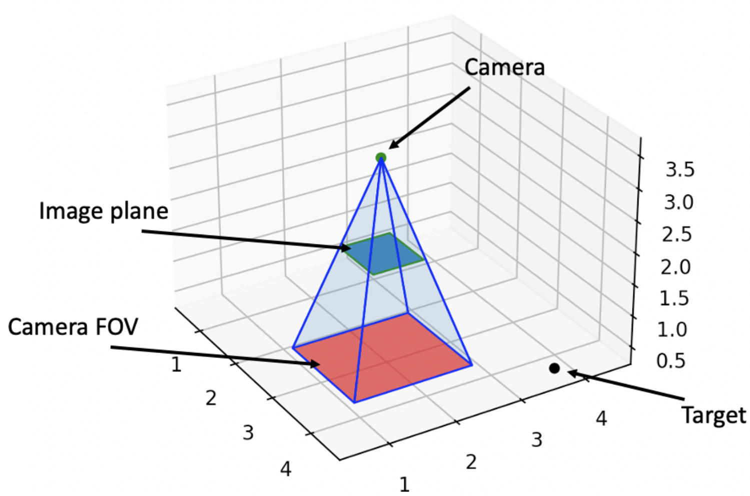

Figure 1: PT camera and ground target in the workspace W

Then, the target position X_T in the inertial frame can be transformed to the camera-fixed frame using the equation,

$$(q_T)_B = H_\phi^T H_\psi^T [(X_T)_I - (X_P)_I] = [q_x, q_y, q_z]^T \tag{2}$$

The target position in the virtual image plane can be obtained from

$$P_T = \lambda [q_x/q_z, q_y/q_z]^T = [p_x, p_y]^T \tag{3}$$

The camera FOV can be obtained by transforming the image sensor's physical bounds from the virtual image plane to the inertial frame. The bounds of a rectangular image plane where a target can be observed is,

$$(q_1)_V = \begin{bmatrix} a/2 \\ b/2 \end{bmatrix}, (q_2)_V = \begin{bmatrix} a/2 \\ -b/2 \end{bmatrix}, (q_3)_V = \begin{bmatrix} -a/2 \\ -b/2 \end{bmatrix}, (q_4)_V = \begin{bmatrix} -a/2 \\ b/2 \end{bmatrix}, \tag{4}$$

where a is the width and b is the height of the image sensor. Thus, the target can be detected from the camera sensor when the target position in the virtual image plane is within the bounds of a rectangular image plane, that is,

$$P_T \in [-a/2, a/2] \times [-b/2, b/2] \tag{5}$$

The geometry of the camera FOV can be obtained in terms of the tetragon's four vertices. The vertices relative to the camera-fixed frame is

$$(q_l)_B = [(q_l)_V^T, \lambda]^T, \qquad l = 1, ..., 4. \tag{6}$$

By performing the inverse transformation, the vertices of the sensor physical bounds in inertial frame can be obtained as follows,

$$(q_l)_I = H_\psi H_\phi (q_l)_B, \qquad l = 1, ..., 4. \tag{7}$$

Then, each of the vectors is dilated by a factor p_l,

$$(q_l')_I = p_l (q_l)_I, \qquad l = 1, ..., 4, \tag{8}$$

where,

$$p_1, 4 = \frac{-X_{p_z}}{\frac{b}{2} sin\phi + \lambda cos\phi} \qquad and \qquad p_2, 3 = \frac{-X_{p_z}}{-\frac{b}{2} sin\phi + \lambda cos\phi} \tag{9}$$

Finally, the FOV vertices in the inertial frame are given by

$$\zeta_l = X_p + q_l', \qquad l = 1, ..., 4. \tag{10}$$

In this report, the PT camera is designed that the pan and tilt angles are adjustable. Thus, the pan and tilt angles are constrained to the ranges $\psi \in [0, 2\pi]$ and $\phi \in [2\pi/3, \pi]$, and the pan and tilt angular velocities are constrained to the ranges $\dot{\psi} \in [-100°, 100°]$ and $\dot{\phi} \in [-100°, 100°]$.

4 Target Detection and Tracking

For simplicity, the camera measurements are assumed to be noise free in this report. The time interval is known and can be discretized and indexed by $k = 0, 1, 2, ..., k_f$. Thus, the position and velocity of the target can be fully observable by the sensor measurements.

Target detection by the camera can be determined from equation (5). It follows that the target lies in the camera FOV in workspace if and only if P_T is in the image plane. The figure 2 shows that the position of target is inside of the image plane, meaning that the camera captures the target. Once the target is detected, the distance between the target and the center of image plane can be measured.

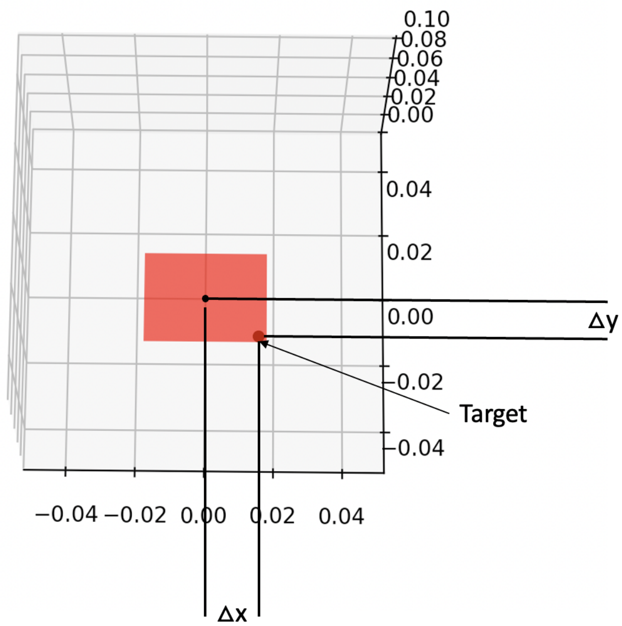

Figure 2: Position of target in the image plane

The distance Δx and Δy shown in Figure 2 can be used to calculate the error values between current angles and desired angles. The desired angle ψ and angle ϕ are determined by PT camera kinematic equations and simple control input [2] because the FOV shape changes based on the orientation of the camera. Let $s = [\psi, \phi, \dot{\psi}, \dot{\phi}]^T$ which is the dynamic state of the camera. For simplicity, it is assumed that the Δx changes based on the orientation of ψ angle and Δy changes based on the orientation of ϕ angle. Then, the camera control input with respect to error values can be designed with proportional gain and derivative gain because the position of the target changes over discrete time intervals.

$$u = \begin{bmatrix} u_1 = K_p e_1(k) + K_d \frac{e_1(k) - e_1(k-1)}{\Delta t} \\ u_2 = K_p e_2(k) + K_d \frac{e_2(k) - e_2(k-1)}{\Delta t} \end{bmatrix} \tag{11}$$

where $e_1 = \Delta x$ and $e_2 = \Delta y$. Then, the camera kinematic equation can be expressed in state-space form,

$$s(k+1) = As(k) + Bu(k) \tag{12}$$

where,

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ b_1 & 0 \\ 0 & b_2 \end{bmatrix} \tag{13}$$

and b_1 and b_2 are two constant motor parameters.

The calculated dynamic state of the camera $s(k+1)$ can be used to adjust the desired pan and tilt angles. It is significant to notice that the pan and tilt angles are not adjusted simultaneously because it is assumed that the error values e_1 and e_2 are controlled discretely.

5 Simulation Results

In this section, the effectiveness of the camera measurement and target tracking method presented in this report is demonstrated using simulation in Python. The workspace W in this simulation consists of one (point) target and one (point) camera. For simplicity, the camera is stationary at fixed position and the point target is moving linearly in the inertial XY-coordinate frame. The surveillance camera initially monitors the ROI by controlling its pan angle in the range $\psi \in [0, 2\pi]$. Once the point target is within boundary of the camera image plane, the camera keeps the target in the center area of the camera's FOV.

As shown in Figure 3, the camera captures the target and tracks simultaneously when there is a controller designed. Compared to the simulation (a) when there is no tracking control, it can be seen in simulation (b) that the position and orientation of the camera's FOV is changed when there is a control for target tracking.

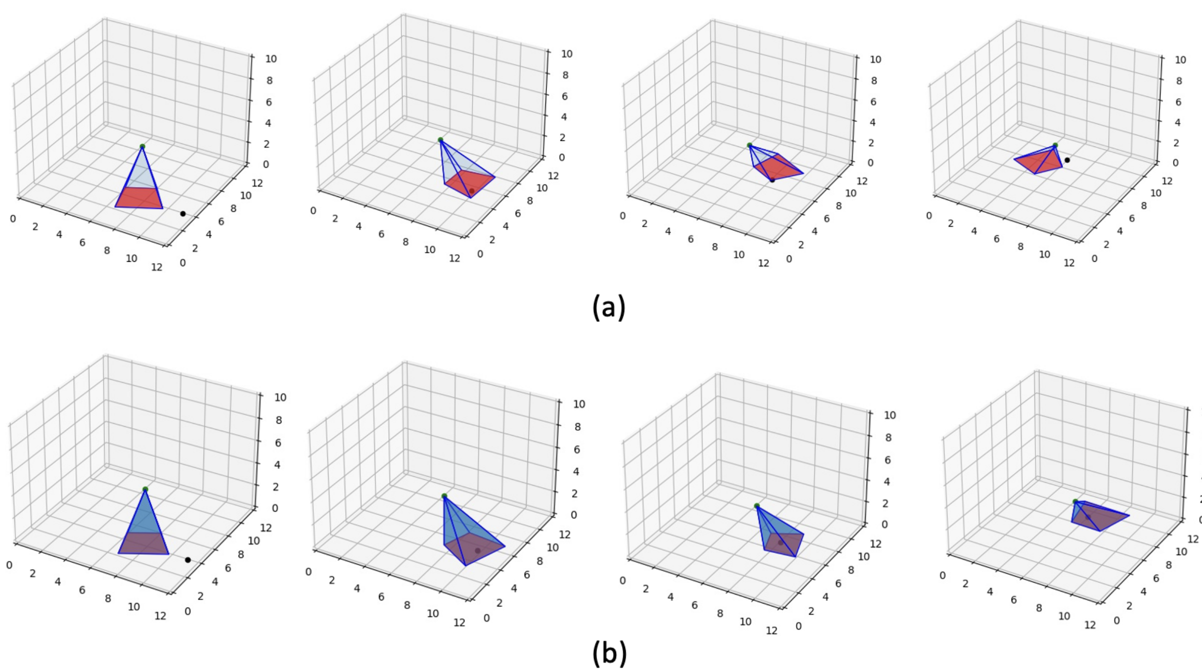

Figure 3: Camera simulation (a) when there is no target tracking, and (b) when there is a control for target tracking

From the Figure 4, it can be seen that the error values are close to zero when the camera keeps the target in the image plane. When the target is detected at first, the error values are relatively large but quickly converge to zero, meaning that the response of control system is fast enough. Therefore, the camera can successfully track the target without losing its controllability. Lastly, it can also be seen that there are small noises because of estimation errors in the system, but it does not significantly affect the performance of the camera to capture the target.

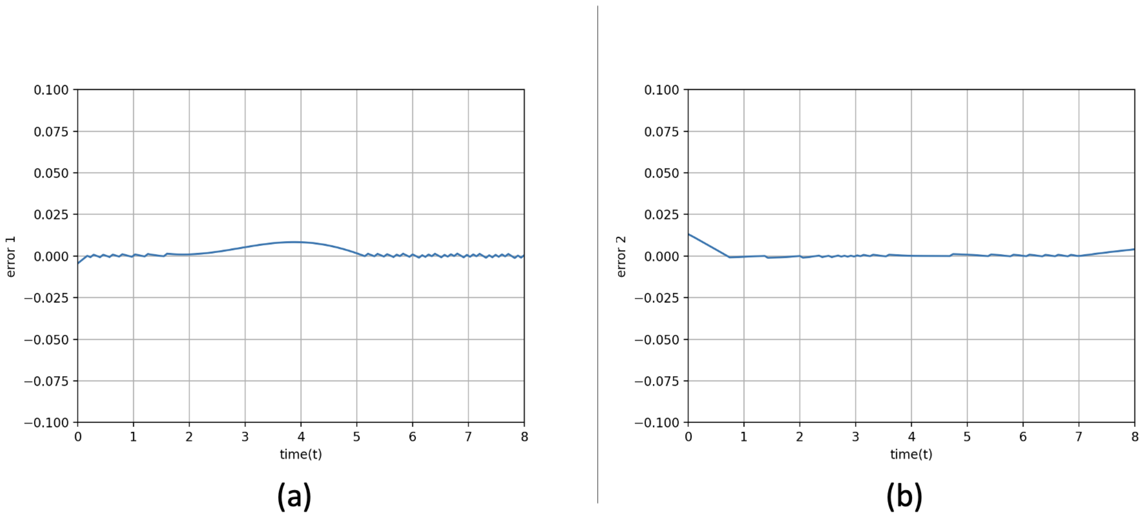

Figure 4: error values over time (a) $error_1 = \Delta x$, (b) $error_2 = \Delta y$

6 Conclusion

This report presents a method of moving target tracking for the pan-tilt camera, characterized by bounded FOV. Using the camera model, the linearly moving target in the workspace can be detected. Also, using the PID controller, the error values can be minimize effectively, and the modeled camera can followed the target maneuver by keeping the position in the center of camera's FOV.

According to the results, it concludes that the controller for target tracking experiences some noises when measuring the error values. Therefore, although the mathematical model of the camera and controller are demonstrated in the simulation result section, a more robust model can be presented in the future using Linear-Quadratic Regulator(LQR) to optimize the controller, as well as the estimation model such as Kalman filter in order to reduce noises.

References

[1] S. Ferrari and T. Wettergren, *Information-driven Planning and Control*. Cambridge, Massachusetts: Cyber physical systems series. MIT Press, 2021.

[2] S. Ferrari, "Pinhole camera model and field-of-view(fov)."