
PENNSYLVANIA
FUZ
slender
Figure 1: AprilTag 36h11 ID = 12

We have the following :

Pixel Coordinates: $x = [U, V]^T$

$$K = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix}$$

Camera rotation matrix, R_{wc} and Camera Translation Vector t_{wc}

① $x_h = \begin{bmatrix} U \\ V \\ 1 \end{bmatrix}$ → Conversion to homogeneous coordinates

② $x_c = K^{-1} x_h$ → Convert pixel coordinates from image frame to camera frame using camera intrinsic matrix

③ $X_c = \lambda x_c = \begin{bmatrix} \lambda x_c \\ \lambda y_c \\ \lambda \end{bmatrix}$ → express the point in camera frame as a 3D vector with unknown depth λ

Step 4: $Y_w = R_{wc}^{-1}(X_c - t_{wc})$ → Convert the point from camera frame to world frame using camera pose

Step 5: $\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = R_{wc}^{-1}(\lambda x_c - t_{wc})$ → Solve for unknown depth, λ

Assuming $Z_w = 0$, we get $0 = r_3^T(\lambda x_c - t_{wc})$ → Assume point lies on x-y plane in world frame.

$$\lambda = \frac{r_3^T t_{wc}}{r_3^T x_c}$$ where r_3^T is the third row of R_{wc}^{-1}

$X_w = r_1^T(\lambda x_c - t_{wc})$ $Y_w = r_2^T(\lambda x_c - t_{wc})$.

Final 3D point is given by $X_w = \begin{bmatrix} X_w \\ Y_w \\ 0 \end{bmatrix}$