

$C$  is the full camera matrix:

$$C = A \times P \times E \quad (1)$$

$A$  is the affine matrix:

$$A = \begin{bmatrix} 1 & 0 & a_x \\ 0 & -1 & a_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

Where  $a_x$  and  $a_y$  equal 0.5, which shifts the screen-space projection so that the camera target is in the middle of the screen (screen space coordinates (0.5, 0.5)).

$P$  is the projection matrix:

$$P = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3)$$

Where  $f$  is the width of the focal plane. Since the field of view was used as an input parameter for the geometric construction of the camera, the following equation was used to calculate  $f$ :

$$f = \frac{1}{\tan FOV} \quad (4)$$

Where  $FOV$  is the field of view in radians.

$E$  is the extrinsic matrix:

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos p & \sin p & 0 \\ 0 & \sin p & -\cos p & c_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

Where  $p$  is the pitch from vertical in radians, and  $c_z$  is the camera Z-value in its own coordinate frame, which is the distance from the target point.

The geometric parameters for the final homography were:

$$FOV = 25.3^\circ, c_z = 36.3, p = 32.72^\circ$$

(Angles were converted from degrees to radians in software.)