1 World, Camera, Image, and Pixel Coordinates.

In class we showed that the perspective projection of a scene point $\vec{X}_c$ is:

$$\vec{x} = \frac{f}{X_{c,3}} \vec{X}_c.$$  \hspace{1cm} (1)

Here $\vec{X}_c = (X_{c,1}, X_{c,2}, X_{c,3})^T$ are camera-centered coordinates, with the camera’s nodal point at the origin, the optical axis is on the $X_{c,3}$-axis, the image plane is perpendicular to the optical axis, and the optical axis pierces the image plane at the point $\vec{x} = (0, 0, f)^T$. Also, notice the image point $\vec{x} = (x_1, x_2, f)^T$ is a 3-vector in this camera centered frame.

Typically image coordinates are written in terms of pixels. A particular pixel can be denoted by $(q_1, q_2)$ with $q_1$ the column and $q_2$ the row in the sampled image. Here we take $(q_1, q_2) = (1, 1)$ to be the top left corner of the image. It is again convenient to express these pixel coordinates in terms of the 3-vector $\vec{q} = (q_1, q_2, 1)^T$. The transformation from the camera centered image coordinates, $\vec{x}$, to pixel coordinates, $\vec{q}$, can then be expressed as

$$\vec{q} = M_{in} \vec{x}$$ \hspace{1cm} (2)

where $M_{in}$ is the $3 \times 3$ matrix

$$M_{in} = \begin{pmatrix} 1/s_1 & 0 & o_1/f \\ 0 & 1/s_2 & o_2/f \\ 0 & 0 & 1/f \end{pmatrix}. \hspace{1cm} (3)$$

Here $(o_1, o_2)$ are the pixel coordinates for the point the optical axis pierces the image plane (i.e. the ‘center’ of the image), and $s_1, s_2$ specify the pixel spacing in the $x$ and $y$ directions. The constants $o_i, s_i$ for $i = 1, 2$, and $f$ are called the intrinsic camera parameters.

Let $\vec{X}_w = (X_{w,1}, X_{w,2}, X_{w,3})^T$ denote a world-based coordinate frame. In general, the coordinate transformation from $\vec{X}_w$ to the camera centered coordinates, $\vec{X}_c$, takes the form

$$\vec{X}_c = M_{ex} \begin{pmatrix} \vec{X}_w \\ 1 \end{pmatrix}, \hspace{1cm} (4)$$

where $M_{ex}$ is the $3 \times 4$ matrix

$$M_{ex} = \begin{pmatrix} R & -R\vec{d} \end{pmatrix}. \hspace{1cm} (5)$$

Here $R$ is the $3 \times 3$ rotation matrix specifying the relative rotation between the camera and world coordinate frames, and $\vec{d} = (d_1, d_2, d_3)^T$ specifies the location of the camera’s nodal point in world coordinates. The 3D position and orientation of the camera, as specified by $R$ and $\vec{d}$, are called the extrinsic camera parameters.
Combining the mappings $M_{int}$ and $M_{ext}$ we find from the above analysis that a world point $\tilde{X}_w$ is imaged to pixel $\tilde{q} = (q_1, q_2, 1)^T$ when

$$\beta \tilde{q} = M_{int} M_{ext} \begin{pmatrix} \tilde{X}_w \\ 1 \end{pmatrix}, \quad \beta = \frac{X_c}{f}. \quad (6)$$

Here $\beta$ must be positive in order for the scene point at $\tilde{X}_w$ to be in front of the camera’s nodal point.