Perspective Projection in Homogeneous Coordinates

Carlo Tomasi

August 24, 2017

If standard Cartesian coordinates are used, a rigid transformation takes the form[1]

$$\mathbf{X}' = R(\mathbf{X} - \mathbf{t})$$

and the equations of perspective projection are of the following form:

$$x_1 = f\frac{X_1}{X_3} \quad \text{and} \quad x_2 = f\frac{X_2}{X_3} \ .$$

When describing the geometry of images taken from different viewpoints, one typically transforms the world coordinates of a point $\mathbf{X}$ through a rigid transformation to obtain the coordinates $\mathbf{X}'$ of that point in the camera reference frame D. The world reference frame is often attached to the first camera, and is therefore called C. The new point is then projected to the image plane to obtain camera coordinates[2] $\mathbf{y}' = [y_1', y_2']^T$, and this point in turn is converted to image coordinates $\boldsymbol{\eta}' = [\eta_1', \eta_2']^T$ through another affine transformation

$$\boldsymbol{\eta}' = \begin{bmatrix} s_1' & 0 \\ 0 & s_2' \end{bmatrix} \mathbf{y}' + \mathbf{c}_0'$$

as we saw in a previous note. In the equation above, $s_1', s_2', \mathbf{c}_0'$ are the internal parameters of the camera D.

Combining these transformations can get messy, and forces one to spell out formulas for individual coordinates:

$$\eta_1' = s_1' f' \frac{\mathbf{i}^T(\mathbf{X} - \mathbf{t})}{\mathbf{k}^T(\mathbf{X} - \mathbf{t})} + c_{01}' \quad \text{and} \quad \eta_2' = s_2' f' \frac{\mathbf{j}^T(\mathbf{X} - \mathbf{t})}{\mathbf{k}^T(\mathbf{X} - \mathbf{t})} + c_{02}' \ .$$

In this expression, $\mathbf{i}^T$, $\mathbf{j}^T$, $\mathbf{k}^T$ are the three rows of R.

This notational complication derives from the summation in affine transformations and from the division in the projection equations. Notation becomes simpler if one uses *homogeneous coordinates*, in which an additional dimension is added to every vector. It turns out that homogeneous coordinates also help accommodate points at infinity seamlessly.

Homogeneous Coordinates

In two dimensions, a point is determined by three homogeneous coordinates:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

[1] We are now starting to talk about multiple cameras, so we reserve different letters of the alphabet to quantities that relate to different cameras. Because of this, we now use subscripts to denote dimensions: x_1, x_2, x_3 instead of x, y, z.

[2] Observe that $\mathbf{X}$, $\mathbf{X}'$ are coordinates of the *same* point in two different reference frames, while $\mathbf{x}$ and $\mathbf{y}'$ are coordinates of two *different* points in two different reference frames.

which correspond to the point

$$e(\mathbf{x}) = \begin{bmatrix} \frac{x_1}{x_3} \\ \frac{x_2}{x_3} \end{bmatrix}$$

in Euclidean coordinates. Homogeneous coordinates are not unique: the same point $e(\mathbf{x})$ is represented by any vector of the form

$$\begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{bmatrix}$$

where α is a nonzero constant, because α cancels when fractions are taken to compute the Euclidean vector $e(\mathbf{x})$.

For the correspondence above to be well defined, x_3 must be nonzero. The definition of homogeneous coordinates weakens this requirement by asking only that **homogeneous coordinates not be zero simultaneously**. So $[0, 0, 0]^T$ is not a valid vector of homogeneous coordinates, but $\mathbf{x}_0 = [a, b, 0]^T$ is, as long as a and b are not both zero. Since the transformation $e(\mathbf{x})$ is then undefined, the point $\mathbf{x}_0$ above does not represent a Euclidean point, that is, a point that is defined in Euclidean geometry. To understand the significance of this point, consider the vector of homogeneous coordinates

$$\mathbf{x}_c = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

with nonzero c. Then,

$$e(\mathbf{x}_c) = \begin{bmatrix} \frac{a}{c} \\ \frac{b}{c} \end{bmatrix} .$$

As c varies, the point with Euclidean coordinates $e(\mathbf{x}_c)$—or homogeneous coordinates $\mathbf{x}_c$—moves along the line from the origin through $e(\mathbf{x}_1) = [a, b]^T$. So changing the last homogeneous coordinate *scales* the point. Because of this, that coordinate is called the *scaling factor*. If c tends to zero, the point moves further and further from the origin and in either direction, depending on the sign of c. One can therefore identify $\mathbf{x}_0$ as the *point at infinity* on that line. Euclidean coordinates have no way to give that point a name, but homogeneous coordinates do.

This is a fundamental distinction, and one can create a whole new geometry—called *projective geometry*—based on it. The Euclidean plane augmented with all the points at infinity is called the *projective plane* $\mathcal{P}^2$ and its points are called *projective points*. A projective point at infinity, $[a, b, 0]^T$ with $(a, b) \neq (0, 0)$, can be usefully used to represent the *direction* of the half-line through the origin and Euclidean point $[a, b]^T$. The set of all points at infinity on $\mathcal{P}^2$ is called the *line at infinity*.

Of course, there is also a *projective space* $\mathcal{P}^3$ of all the projective points with homogeneous coordinates

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \quad \text{with} \quad (X_1, X_2, X_3, X_4) \neq (0, 0, 0, 0) .$$

When $X_4 \neq 0$, this point corresponds to the Euclidean point

$$e(\mathbf{X}) = \begin{bmatrix} \frac{X_1}{X_4} \\ \frac{X_2}{X_4} \\ \frac{X_3}{X_4} \end{bmatrix} .$$

Similar considerations hold for $\mathcal{P}^3$ as do for $\mathcal{P}^2$, and the set of all points at infinity on $\mathcal{P}^3$ is called the *plane at infinity*.

Projection Equations in Homogeneous Coordinates

For us, the main advantage of using homogeneous coordinates is that both affine transformations and projections become linear. Specifically, let the *standard reference frame* for a camera C be a right-handed Cartesian frame with its origin at the center of projection of C, its positive X_3 axis pointing towards the scene along the optical axis of the lens, and its positive X_1 axis pointing to the right[3] along the rows of the camera sensor. As a consequence, the positive X_2 axis points downwards along the columns of the sensor.

Then, let $\mathbf{X}$ and $\mathbf{X}'$ denote the homogeneous coordinates of the same 3D point $\mathcal{P}$ in the standard reference frames of two cameras C and D. We attach the world reference system to C, so un-primed coordinates refer to C and are also world coordinates. Let

$$\mathbf{X}' \sim G\mathbf{X} \quad \text{where} \quad G \sim \begin{bmatrix} R & -R\mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{1}$$

be the rigid transformation between the two reference systems. In the expression for G above, the column vector $\mathbf{0}$ contains three zeros, and the 3×3 matrix R represents a rotation, so that

$$R^T R = R R^T = I_3 \quad \text{and} \quad \det(R) = 1 \, . \tag{2}$$

The rows of R are the unit vectors along the positive axes of D in the reference frame of C. The 3×1 vector $\mathbf{t}$ contains the coordinates of the center of projection of D in the reference frame of C, so that the vector of Euclidean coordinates of the center of projection of C in the reference frame of D is

$$\mathbf{s}' = -R\mathbf{t} \tag{3}$$

as was shown in a previous note. Then, the key relationships can be expressed as follows in homogeneous coordinates.[4]

$$\mathbf{X}' \sim G\mathbf{X} \quad \text{where} \quad \mathbf{X}, \mathbf{X}' \in \mathcal{P}^3 \quad \text{and} \quad G \sim \left[\begin{array}{c|c} R & -R\mathbf{t} \\ \hline \mathbf{0}_3^T & 1 \end{array}\right]$$

$$\mathbf{x} \sim \Pi\mathbf{X} \quad \text{and} \quad \mathbf{y}' \sim \Pi\mathbf{X}' \quad \text{where} \quad \mathbf{x}, \mathbf{y}' \in \mathcal{P}^2 \quad , \quad \Pi \sim \left[\begin{array}{c|c} I_3 & \mathbf{0} \end{array}\right]$$

$$\boldsymbol{\xi} \sim K_s K_f \mathbf{x} \quad \text{and} \quad \boldsymbol{\eta}' \sim K_s' K_f' \mathbf{y}' \quad \text{where} \quad \boldsymbol{\xi}, \boldsymbol{\eta}' \in \mathcal{P}^2 \quad , \quad K_s \sim \left[\begin{array}{c|c} S & \mathbf{c}_0 \\ \hline \mathbf{0}_2^T & 1 \end{array}\right] \quad , \quad K_s' \sim \left[\begin{array}{c|c} S' & \mathbf{c}_0' \\ \hline \mathbf{0}_2^T & 1 \end{array}\right] ,$$

$$K_f = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad K_f' = \begin{bmatrix} f' & 0 & 0 \\ 0 & f' & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

[3] When the camera is upside-up and viewed from behind it, as when looking through its viewfinder.

[4] With arguable inconsistency of notation, we use K and K' to denote possibly different camera calibration matrices, rather than using different letters, such as K and L. On the other hand, a camera calibration matrix is always expressed in its own reference system, so this causes no problems.

In these expressions, $\mathbf{0}_k$ is a column vector of k zeros, I_3 is the 3×3 identity, and

$$S = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \quad , \quad S' = \begin{bmatrix} s'_1 & 0 \\ 0 & s'_2 \end{bmatrix} .$$

If image coordinates $\boldsymbol{\xi}, \boldsymbol{\eta}'$ are in pixels and world and image coordinates $\mathbf{X}, \mathbf{X}', \mathbf{x}, \mathbf{y}'$ are, say, in millimeters, then s_1, s_2, s'_1, s'_2 are in pixels per millimeter and f, f' are in millimeters.

The symbol '$\sim$' represents projective equality: two vectors $\mathbf{X}$, $\mathbf{Y}$ for which

$$\mathbf{X} \sim \mathbf{Y}$$

represent the same point in projective space, but are merely proportional, rather than equal, to each other as vectors:

$$\mathbf{X} \sim \mathbf{Y} \quad \Leftrightarrow \quad \mathbf{X} = \alpha \mathbf{Y} \quad \text{for some} \quad \alpha \neq 0 .$$

With these definitions, the relationships between world point $\mathcal{P}$ and its image points $\boldsymbol{\xi}$ and $\boldsymbol{\eta}'$ can be expressed concisely in matrix form as follows:

$$\boldsymbol{\xi} = P\mathbf{X} \quad \text{and} \quad \boldsymbol{\eta}' = P'\mathbf{X} \quad \text{where} \quad P = K_s K_f \Pi \quad \text{and} \quad P' = K'_s K'_f \Pi G .$$

The matrices P, P', Π, K, K' are all 3×4. The matrices P, P' are called *projection matrices*, the matrix Π is called the *canonical projection matrix*, and the matrices K_f, K'_f, K_s, K'_s are called *internal camera calibration matrices*.

The Canonical Camera

Perhaps the subtlest of the equations above concerns canonical projection, $\mathbf{x} \sim \Pi\mathbf{X}$, so here it is again, spelled out with all its coordinates:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} .$$

This equation eliminates the scaling constant X_4 altogether, consistently with the fact that all points in $\mathcal{P}^3$ on the line through the origin and the point with Euclidean coordinates $[X_1, X_2, X_3]^T$ project onto the same image point, because the origin is the center of projection. It then reinterprets $x_3 = X_3$ to be the scaling factor of $\mathbf{x} \in \mathcal{P}^2$. So all points with *Euclidean* coordinates

$$e(\mathbf{X}) = \begin{bmatrix} \frac{X_1}{X_4} \\ \frac{X_2}{X_4} \\ \frac{X_3}{X_4} \end{bmatrix} \in \mathbb{R}^3$$

as well as the projective point

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 0 \end{bmatrix}$$

project to image point

$$e(\mathbf{x}) = \begin{bmatrix} \frac{x_1}{x_3} \\ \frac{x_2}{x_3} \end{bmatrix} = \begin{bmatrix} \frac{X_1}{X_3} \\ \frac{X_2}{X_3} \end{bmatrix} .$$

A point in $\mathcal{P}^3$ with $X_3 = 0$ is a point on the (projective) plane through the center of projection and parallel to the image plane. Appropriately, it projects to point at infinity

$$\begin{bmatrix} X_1 \\ X_2 \\ 0 \end{bmatrix}$$

on the image plane $\mathcal{P}$. If one introduces a transformation from Euclidean to homogeneous coordinates

$$h(\mathbf{x}) = \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} ,$$

then the equation $\mathbf{x} \sim \Pi\mathbf{X}$ can also be written as follows for Euclidean points:

$$\mathbf{x} \sim h(e(\mathbf{X}))$$

(check this!), but not for points at infinity.

The coordinates $\mathbf{x}$ and $\mathbf{y}'$ are called *canonical image coordinates*, and the reference system in which they are measured is called the *canonical reference system*. The canonical projection is the projection matrix P that would be obtained when $K_s = K_f = I_3$, the 3×3 identity matrix. Because of this, the canonical coordinates $\mathbf{x}$ and $\mathbf{y}'$ can be viewed as the homogeneous coordinates of image points taken with a *canonical camera* that has a focal distance of 1, and for which image coordinates are measured relative to the principal point.

We point out a convenient coincidence regarding coordinates of image points in the canonical reference system. In this system, the third Euclidean coordinate of an image point is 1—the canonical focal distance. Because of this, the canonical coordinates $\mathbf{x}$, $\mathbf{y}'$ can be viewed in two different ways: They are either vectors of homogeneous coordinates for the two-dimensional image point, or vectors of Euclidean coordinates of the three-dimensional vectors from the center of projection to the image point. In other words, in the canonical reference frame the distinction between homogeneous coordinates and Euclidean coordinates is mainly in the eyes of the beholder.