Computer Vision	Forsyth, Ponce
Chapter 1	Geometric Camera Models

1 Reading

The image plane for a camera is usually rectangular, but the image plane for a human (i.e. a retina) is more spherical.

1.1. Image Formation

1.1.1. Pinhole Perspective

Since a pinhole setup creates inverted images on the backplane of the dark chamber, we typically consider a virtual image in front of the pinhole (`Z(pinhole) - Z(backplane)` away), where the image will appear right side up.

The projections of two parallel lines lying within a plane will converge at a horizon line in the image. We can prove such properties geometrically. To do so, we define a coordinate system (i, j, k) with an origin O at the pinhole (with k pointing in the direction of the physical image plane). The **optical axis** is then defined as the line perpendicular to the image plane that passes through the pinhole, and the **image center** c is the intersection of the optical axis with the image plane. (The pinhole is d away from the image plane.)

$P = (X, Y, Z)$ is a scene point; $p = (x, y, d)$ is its projection in the image. P, the pinhole O, and p are collinear, so $\overrightarrow{Op} = \lambda\overrightarrow{OP}$ and $(x, y, d) = (\lambda X, \lambda Y, \lambda Z)$. In other words,

$$\lambda = \frac{x}{X} = \frac{y}{Y} = \frac{d}{Z}$$

meaning $x = dX/Z$ and $y = dY/Z$.

1.1.2. Weak Perspective

In **weak perspective** (aka *scaled orthography*), the entire scene exists (approximately) on a fronto-parallel plane Π_0 at $Z = Z_0$ which is sufficiently far from the camera. For each point in Π_0,

$$x = dX/Z_0 = -mX$$
$$y = dY/Z_0 = -mY$$

for $m = -d/Z_0$. Since Π_0 is in front of the camera (meaning Z_0 is negative), the magnification m is positive. It is called the "magnification" because for P and Q in Π_0 (with image projections p and q),

$$\begin{aligned}
\|\overrightarrow{pq}\| &= \|q - p\| \\
&= \sqrt{(q_i - p_i)^2 + (q_j - p_j)^2} \\
&= \sqrt{(-mQ_i + mP_i)^2 + (-mQ_j + mP_j)^2} \\
&= \sqrt{m^2(Q_i - P_i)^2 + m^2(Q_j - P_j)^2} \\
&= m\|Q - P\| \\
&= m\|\overrightarrow{PQ}\|
\end{aligned}$$

If the camera is always at a constant distance from the scene, we can avoid reversing the image by defining m to be -1. Then $x = X$ and $y = Y$. This is known as **orthographic projection**.

2 Exercises

References

[1] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Always learning. Pearson, 2012.