Lecture 2

August 26, 2017

The word camera comes from *camera obscura*, literally meaning *dark chamber*. The camera obscura is a dark chamber with a small hole in one wall. On the wall opposite the hole, we see an inverted image of the world outside. Today such a camera would be called a pinhole camera. A pinhole camera works because if the hole is infinitesimally small, for every point in the 3D world, only a single ray of light from that point will enter the box or the room, and will land on a precise point on the screen opposite, leading to a sharp image.

A pinhole camera unfortunately produces an inverted image. However we can consider a virtual screen in front of the pinhole. On this virtual screen, an erect image will be formed.

1 Geometry

Assume that the camera is at $\mathbf{O} = (0, 0, 0)$ and is looking along the Z axis. Suppose that the virtual screen is the plane $Z = f$. Consider a point $\mathbf{P} = (P_x, P_y, P_z)$ in the 3D world. Where does this point project?

This point projects where the ray OP intersects the virtual screen. Points on this ray can be written as $\mathbf{O} + t(\mathbf{P} - \mathbf{O}) = (tP_x, tP_y, tP_z)$. We can find the point at which OP intersects the $Z = f$ plane by finding the value of t for which the z coordinate equals f:

$$\begin{aligned} tP_z &= f \\ \Rightarrow t &= \tfrac{f}{P_z} \end{aligned} \tag{1}$$

The point of intersection is thus $(f\frac{P_x}{P_z}, f\frac{P_y}{P_z}, f)$, and so the coordinates of this point in the image are $\mathbf{p} = (f\frac{P_x}{P_z}, f\frac{P_y}{P_z})$. This transformation of points in the 3D world into points in the image plane is called *perspective projection*. Often, we will simply choose our units so that $f = 1$.

2 Effects of perspective projection

There are several interesting effects of perspective projection:

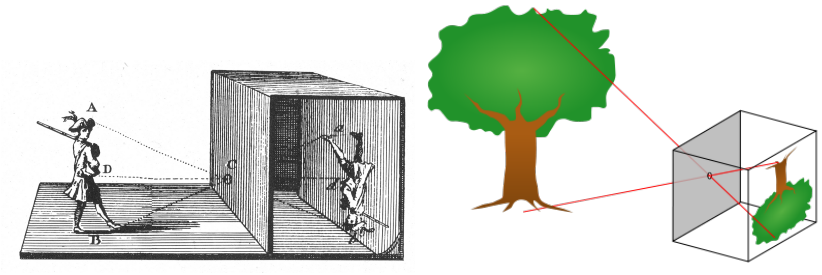

Figure 1: Pinhole camera

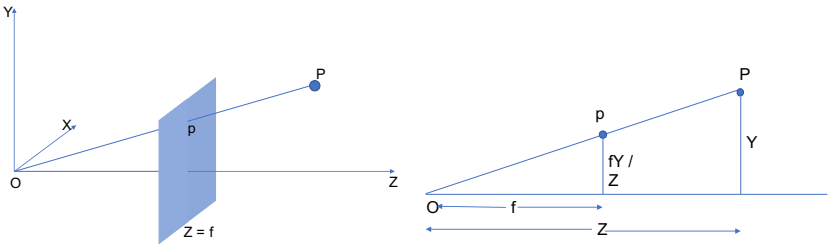

Figure 2: Geometry of a pinhole camera

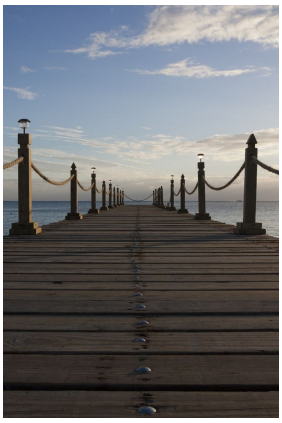

Figure 3: Perspective.

Parallel lines in the world converge in the image Why does this happen? Consider two parallel lines in the 3D world. A line in the 3D world can be represented by a point $\mathbf{A}$ and a direction $\mathbf{D}$. Points on this line then take the form $\mathbf{A}(t) = \mathbf{A} + t\mathbf{D}$. A line parallel to this line will have the same direction $\mathbf{D}$ but will pass through a different point $\mathbf{B}$. Points on this line are of the form $\mathbf{B}(t) = \mathbf{B} + t\mathbf{D}$.

$\mathbf{A}(t)$ projects to $\mathbf{a}(t) = (\frac{A_x+tD_x}{A_z+tD_z}, \frac{A_y+tD_y}{A_z+tD_z}) = (\frac{\frac{A_x}{t}+D_x}{\frac{A_z}{t}+D_z}, \frac{\frac{A_y}{t}+D_y}{\frac{A_z}{t}+D_z})$. The last equality was obtained by dividing both numerator and denominator by t. As $t \to \infty$, $\mathbf{a}(t) \to (\frac{D_x}{D_z}, \frac{D_y}{D_z})$. Similarly, as $t \to \infty$, $\mathbf{b}(t) \to (\frac{D_x}{D_z}, \frac{D_y}{D_z})$. Thus, far away points for both lines converge to the same image point. Thus the image of parallel lines in the world converge. Note that this only happens if $D_z \neq 0$.

The point of convergence of parallel points in an image is called a *vanishing point.*

Parallel planes in the world converge in the image A plane in 3D is represented by a normal $\mathbf{N} = (N_x, N_y, N_z)$ and a scalar d, and has an equation of the form:

$$N_x X + N_y Y + N_z Z = D \tag{2}$$

To see what happens to this plane under projection, let's divide by Z:

$$N_x \frac{X}{Z} + N_y \frac{Y}{Z} + N_z = \frac{D}{Z} \tag{3}$$

As $Z \to \infty$, this equation becomes:

$$N_x \frac{X}{Z} + N_y \frac{Y}{Z} + N_z = 0 \tag{4}$$

Since $(\frac{X}{Z}, \frac{Y}{Z})$ is the image of point (X, Y, Z), this equation represents a line in the image. Parallel lines will have the same normal $\mathbf{N}$, but might have different scalars D_1 and D_2. However, this limiting line does not depend on D, so all parallel planes will converge to this line in the image as Z approaches infinity. This line is a *vanishing line.* Note that this happens only if $N_x \neq 0$ or $N_y \neq 0$, since otherwise, the plane equation is $Z = d$.

Farther objects are smaller Why does this happen? Imagine a real world object of size H. Let's say that the bottom of this object is at location (X, Y, Z). Then, the top of this object is at $(X, Y+H, Z)$. The bottom projects to $(\frac{X}{Z}, \frac{Y}{Z})$, while the top projects to $(\frac{X}{Z}, \frac{Y+H}{Z})$. The *apparent height* is then $\frac{Y+H}{Z} - \frac{Y}{Z} = \frac{H}{Z}$, which decreases as distance from camera, i.e, Z increases.

Farther objects on the ground plane appear higher This effect is peculiar to our vantage point. We usually take photographs from some height above the ground plane. Suppose we are taking photographs from a height H_c above the ground plane. Then the equation of the ground plane is $Y = -H_c$. A point on the ground plane will be $(X, -H_c, Z)$. This point will project to $(\frac{X}{Z}, \frac{-H_c}{Z})$.

As Z increases, the y-coordinate of this image point becomes less negative, that is the point goes higher. In the limit, this y-coordinate will approach 0. Thus, points infinitely far away on the ground plane, lie on the line $y = 0$ in the image. This line is the *horizon*.

3 Projective plane and homogenous coordinates

An interesting fact about perspective projection is that all points on a ray through the origin project onto the same point in the image. For example if $\mathbf{P} = (P_x, P_y, P_z)$ is a point in 3D, then as we saw, points on the ray OP have the form $\lambda P = (\lambda P_x, \lambda P_y, \lambda P_z)$. All these points project to the same image point, because $(\frac{\lambda P_x}{\lambda P_z}, \frac{\lambda P_y}{\lambda P_z}) = (\frac{P_x}{P_z}, \frac{P_y}{P_z})$. Thus perspective projection creates an equivalence between all points that lie on a ray.

Consider now the space of all *rays* through the camera center $\mathbf{O}$. We will represent a ray by any point $\mathbf{P}$ on the ray, so each ray will have 3 coordinates, but we will stipulate that $\mathbf{P}$ and $\lambda\mathbf{P}$ are the same ray. Thus even though there are 3 coordinates, there are only 2 degrees of freedom. To distinguish rays from points, we will use a separate notation for them: $\vec{\mathbf{p}}$. Thus, we say that $\vec{\mathbf{p}} \sim \lambda\vec{\mathbf{p}}$, where $\sim$ denotes equivalence. This coordinate system where multiplying all coordinates by the same scalar doesn't change anything, is called homogenous coordinates.

This space of rays is the projective plane $\mathbb{P}^2$. This space more or less has a one-to-one mapping with coordinates on the image, which is $\mathbb{R}^2$. Any image point $\mathbf{p} = (x, y)$ corresponds to the ray $\vec{\mathbf{p}} = (x, y, 1)$. Conversely, for every ray $\lambda\vec{\mathbf{q}}$, there exists a corresponding point on the image plane, obtained by picking λ so that the z coordinate is 1. In other words, the ray $\vec{\mathbf{q}} = (x, y, z)$ corresponds to the point on the image plane $\mathbf{q} = (\frac{x}{z}, \frac{y}{z})$.

We said that this was a more or less one-to-one correspondence. However, there are rays that do not correspond to points on the image plane. Consider a ray through the camera center $\mathbf{O}$ but *parallel* to the image plane. For such a ray, the z coordinate is 0: these rays are of the form $(x, y, 0)$. These rays *never intersect the image plane*, and in fact can be seen as image plane points *at infinity*. Thus, the projective plane $\mathbb{P}^2$ can be seen as the real plane $\mathbb{R}^2$ *embellished with points at infinity*.

To recap, we have two mappings:

1. A mapping from *inhomogenous coordinates* (x, y) to *homogenous coordinates* $(x, y, 1)$, and

2. A mapping from *homogenous coordinates* (x, y, z) to *inhomogenous coordinates* $(\frac{x}{z}, \frac{y}{z})$.

There is nothing special about the 2D plane. We can also have homogenous coordinates in 3D. Thus the 3D point represented by the inhomogenous coordinates (x, y, z) is represented in homogenous coordinates by $(x, y, z, 1)$. Similarly,

the 3D point represented by the homogenous coordinates (x, y, z, w) corresponds to the inhomogenous coordinates $(\frac{x}{w}, \frac{y}{w}, \frac{z}{w})$.

So how does perspective projection work in homogenous coordinates? We claim that perspective projection is actually *linear* in homogenous coordinates. In particular, we claim that perspective projection amounts to pre-multiplying by the 3×4 projective matrix $\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$. This matrix basically drops the last coordinate. To see this, let's calculate the projection of a point represented in homogenous coordinates by $\vec{\mathbf{Q}} = (x, y, z, w)$. This point in inhomogenous coordinates is $(\frac{x}{w}, \frac{y}{w}, \frac{z}{w})$. It projects to the point $(\frac{\frac{x}{w}}{\frac{z}{w}}, \frac{\frac{y}{w}}{\frac{z}{w}}) = (\frac{x}{z}, \frac{y}{z})$, which will be represented by $\vec{\mathbf{q}} = (\frac{x}{z}, \frac{y}{z}, 1)$ in homogenous coordinates. We can see that

$$\mathbf{P}\vec{\mathbf{Q}} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \sim \vec{\mathbf{q}} \tag{5}$$

Thus, perspective projection is *linear* in homogenous coordinates. Because of this, we will often choose to work in homogenous coordinates.

Another interesting property of homogenous coordinates is that translation is also represented as a matrix multiplication in homogenous coordinates. For example, multiplying the point $(x, y, z, 1)$ by the matrix $M = \begin{pmatrix} \mathbf{I} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix}$ produces the point $(x + t_x, y + t_y, z + t_z, 1)$, which represents the translated point.

Finally, any transformation represented as a matrix multiplication in non-homogenous coordinates is still a matrix multiplication in homogenous coordinates. To see this, observe that:

$$\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} \mathbf{P} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{AP} \\ 1 \end{pmatrix} \tag{6}$$

4 The full projection equation

Till now, we chose coordinate systems of our choice: we placed the origin at the camera, and the Z axis along the camera's viewing direction. We also assumed that the image plane used the same coordinate system. However, in general this is not true.

First, the world may be represented in a different coordinate system with a different origin. So we will first need to (a) rotate the coordinate axes so that the axes align with the camera coordinate system, and (b) translate the origin to the camera. Thus if $\vec{\mathbf{Q}}^w$ is a point in world homogenous coordinates, in the camera coordinates, this point is:

$$\vec{\mathbf{Q}} \sim \begin{pmatrix} \mathbf{I} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} \vec{\mathbf{Q}}^w \tag{7}$$

$$\Rightarrow \vec{\mathbf{Q}} \sim \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \vec{\mathbf{Q}}^w \tag{8}$$

where $\mathbf{R}$ is a rotation of the axes, and $\mathbf{t}$ is a translation.

We then project to the image by simply pre-multiplying by the projection matrix $\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$:

$$\vec{\mathbf{q}} \sim \mathbf{P} \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \vec{\mathbf{Q}}^w = [\mathbf{R}|\mathbf{t}]\vec{\mathbf{Q}}^w \tag{9}$$

Finally, we will need to convert this into the image coordinates in terms of pixels. We will represent this conversion using a simple 3×3 matrix $\mathbf{K}$:

$$\vec{\mathbf{q}} \sim \mathbf{K}[\mathbf{R}|\mathbf{t}]\vec{\mathbf{Q}}^w \tag{10}$$

Thus the full projection of the camera involves three sets of parameters: $\mathbf{t}$, representing the translation to the camera location, $\mathbf{R}$, representing the rotation of the axes, and $\mathbf{K}$ representing the conversion to the image coordinate system. $\mathbf{R}$ and $\mathbf{t}$ are called *extrinsic camera parameters*, and $\mathbf{K}$ contains the *intrinsic camera parameters*.