Projective Geometry of Image Formation

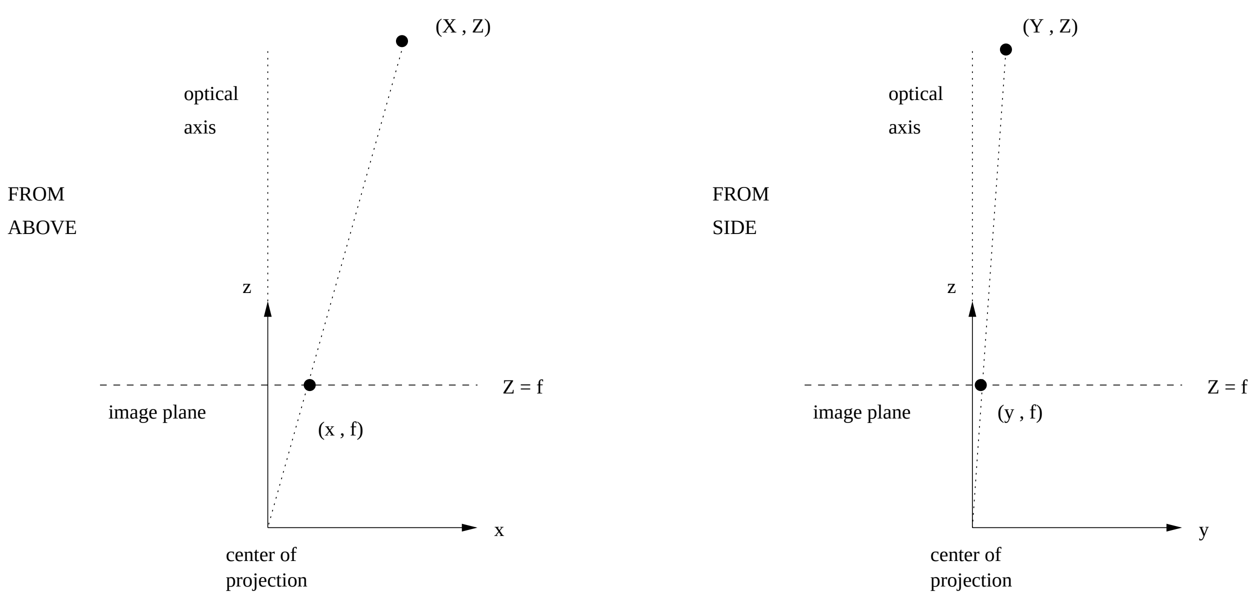
We begin by reviewing some of the basic geometry of image formation. How do points in 3D project to a 2D image? Consider one viewer i.e. a camera. Define the camera coordinate system (X, Y, Z) for the 3D points in the scene as follows. Let the position of the viewer be the origin $(0, 0, 0)$. Let the camera point in direction $(0, 0, 1)$, that is, in the direction of the Z axis. The Z axis is called the *optical axis*.

Next consider a 3D scene point (X_0, Y_0, Z_0) in this camera coordinate system. Consider the line from the origin through this point. We intersect this line with the plane $Z = f$. The plane $Z = f$ is called the *image plane* or *projection plane*. The point of intersection is the *image position*. The origin is the *center of projection*.

Using similar triangles, you can see that the image coordinates of the projected point are

$$(x, y) \;=\; (\frac{X_0}{Z_0}\, f, \frac{Y_0}{Z_0}\, f). \tag{1}$$

Of course, real cameras have the image plane behind the center of projection, so real images are upside down and backwards. We will return to this issue a few lectures from now.



General plane

Suppose the scene contains a plane, written in the camera's coordinate system as:

$$aX + bY + cZ = d. \tag{2}$$

[ASIDE: in class, I simplified the equation by assuming that $d = 1$. In these notes and in the posted slides, I will not make this assumption. This allows me the flexibility for the plane to have $d = 0$.]

Next, multiply both sides of the equation by f/Z, we get

$$\frac{afX}{Z} + \frac{bfY}{Z} + cf = \frac{fd}{Z}$$

and so

$$ax + by + cf = \frac{fd}{Z}$$

Notice that this 3D is still the equation of a plane, but now it is an equation of plane in the 3D space defined by $(x, y, \frac{1}{Z})$.

If we let $Z \to \infty$, we get the line

$$ax + by + cf = 0.$$

This line is sometimes called the *line at infinity*. In more familiar terms, it is called the *horizon*.

Example: Ground plane and the horizon

Consider a specific example. Suppose the only visible surface is the ground, which we approximate as a plane. Suppose the camera is a height h above this *ground plane* and is pointing in the Z direction, and so

$$Y = -h$$

where $h > 0$. From Eq. (1), we have

$$y \; = \; -\frac{hf}{Z}. \tag{3}$$

Any fixed value of y defines a horizontal line in the image, and scene points that project to that line have the same depth (independent of x). Similarly, points of a fixed depth $Z = Z_0$ all project to the same y value. In particular, the larger the depth, the nearer the y value is to 0. In the limit as $Z \to \infty$, we have $y \to 0$, which defines the horizon. Notice that the horizon passes through the center of the image. This property depends on the particular scene configuration we are assuming here.

[ASIDE: In the lecture, the material was presented in a different order. In particular, the following example was presented last.]

Suppose the ground plane is covered in square tiles. This could be a sidewalk, or tiles in a building. Suppose that the width and length of each tile is 1 unit, i.e. $\Delta X = 1$ and $\Delta Z = 1$. What is the width Δx and height Δy of a tile projected in the image?

[The following was modified to Sept. 16 in response to a clarification request.]

For the width, we take a sideways step on the surface $(\Delta X, \Delta Z) = (1, 0)$, i.e. the Z value doesn't change. Let the points marking the width have X values X_1, X_2, so $X_2 - X_1 = 1$ and so $\Delta x = x_2 - x_1$ satisfies

$$\Delta x = -\frac{f \Delta X}{Z} = -\frac{f}{Z}$$

For the height, we take a step on the surface $(\Delta X, \Delta Z) = (0, 1)$. Since $y = -\frac{fh}{Z}$, we can take the derivative dy/dZ and approximate the image height of the tile by:

$$\Delta y \; \approx \; \frac{fh}{Z^2} \Delta Z \; = \; \frac{fh}{Z^2}.$$

Note Δx and Δy both depend on $\frac{1}{Z}$, and in particular both decrease to 0 as $Z \to \infty$. Thus, the tiles appear smaller in the image as the points get further away, i.e. toward the horizon. (No surprise there.) Note also that Δx and Δy decrease at different rates as we go near the horizon, and in particular

$$\frac{\Delta y}{\Delta x} = \frac{h}{Z}.$$

Thus, the image of each tile also becomes more compressed (foreshortened) as $Z \to \infty$, i.e. towards the horizon $y = 0$. You can see both of these effects (size and foreshortening) in the image below.

Depth gradient of a plane, slant and tilt

One familiar perceptual property of a plane is how it slopes away from you, relative to the line of sight. *(See examples in the slides.)* Does it slope to the right, or to the left, or downward like a ceiling or upward like a floor ?

Slope depends on the first order properties of a surface, that is, the depth gradient

$$\nabla Z \equiv (\frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y}). \tag{4}$$

It is sometimes useful to describe the depth gradient in the following way. (We will come back to this later in the course.)

Define the angle τ by the direction of the vector ∇Z, namely

$$(\cos \tau, \sin \tau) = \frac{\nabla Z}{|\nabla Z|}. \tag{5}$$

This direction is well-defined provided that $|\nabla Z| \neq 0$.

The length $|\nabla Z|$ of the gradient vector is the slope of the surface in direction the depth gradient, namely in direction $(\cos\tau, \sin\tau)$. By inspection, the slope defines an angle σ, which is the angle between the plane and a constant-Z plane, and so

$$|\nabla Z| = \tan\sigma. \tag{6}$$

Traditionally, the angle σ is called the *slant* and angle τ is called the *tilt*.

An equivalent definition of the angles τ and σ is that they define spherical coordinates for the surface normal vector, as follows. For a plane, $aX + bY + cZ = d$, you should remember from your linear algebra that the surface normal is in direction (a, b, c). The angle σ can now be interpreted as the angle between the Z axis and the surface normal. Thus, from geometric reasoning,

$$\tan\sigma = \frac{\sqrt{a^2 + b^2}}{|c|}. \tag{7}$$

Note this is consistent with the definition of σ above. To see why, note from the equation of the plane that

$$\nabla Z = (-\frac{a}{c}, -\frac{b}{c}), \tag{8}$$

and

$$|\nabla Z| = \sqrt{(\frac{\partial Z}{\partial X})^2 + (\frac{\partial Z}{\partial Y})^2} = \sqrt{\frac{a^2 + b^2}{c^2}} \tag{9}$$

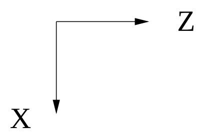
and so, from Eqs 7 and 9

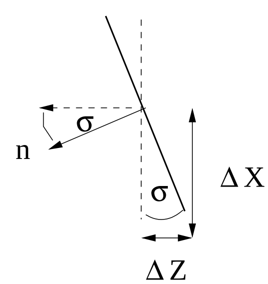
$$\tan\sigma = |\nabla Z|.$$

which is what we saw with Eq. (6).

Treating the $-Z$ direction as the north pole of a sphere, the slant angle σ may be thought of as the *latitute* angle (measured from the north pole) and the angle τ may be thought of as the longitude τ (measured from the X direction). Slant σ can go from 0 to $\pi/2$ or 90 degrees. Tilt τ can go from 0 to 2π or 360 degrees.

To illustrate the two ways of thinking about σ, here is a case where $\tau = 0$ and so the slope is entirely in the X direction.





[ASIDE: At the end of the lecture, I briefly discussed the difference between the formal definition of slant and tilt given above (which is based on a plane in the scene), and a perhaps more intuitive notion of surface slant which is based on the angle between the surface normal and the direction from a surface point to the viewer. Please see the slides to try to get the basic idea.]