Image Projection

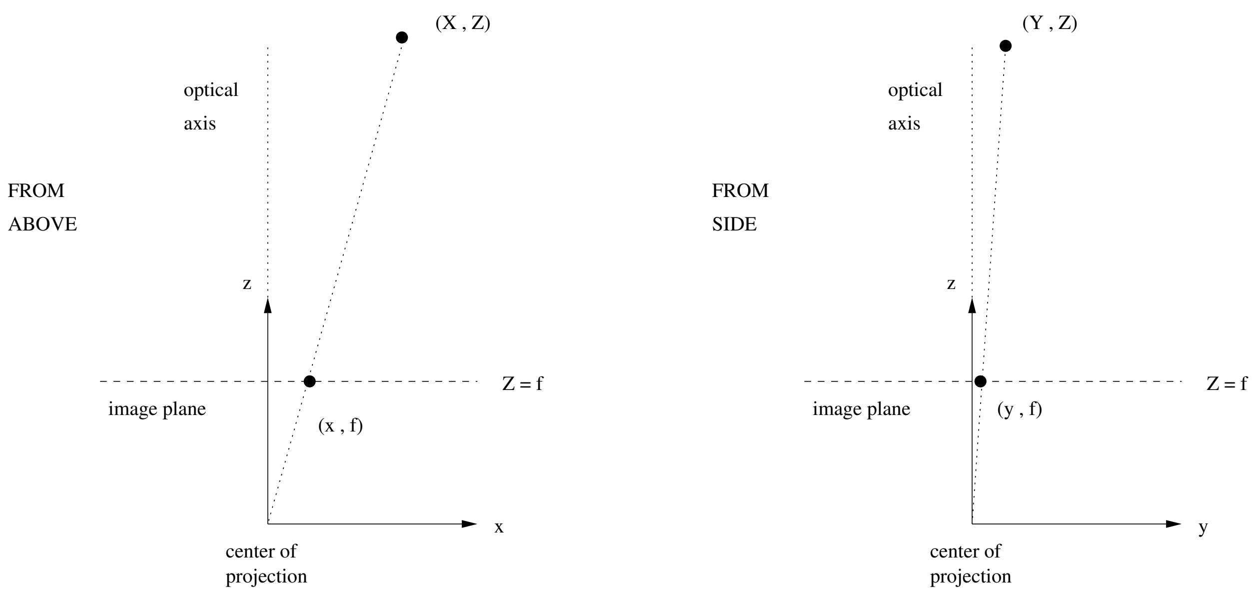
We begin by reviewing the basic geometry of image formation, namely how do points in 3D project to a 2D image? Consider one viewer i.e. a camera. Define the *camera coordinate system* (X, Y, Z) for the 3D points in the scene as follows. Let the position of the viewer be the origin $(0, 0, 0)$. Let the camera "point" in direction $(0, 0, 1)$, that is, in the direction of the Z axis. The Z axis is called the *optical axis*. The X axis is chosen to point to the right in the image plane, and the Y axis point upwards in the image plane. This a known as a left-handed coordinate system.

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*. Note that the image plane is placed in front of the center of projection, so we should not confuse it with the sensor plane. We'll have more to say later about the relationship between the projection plane and sensor plane.

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 scene plane

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

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

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

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

and so

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

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

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. Let the camera be height h above this *ground plane*. If the camera is pointing in a direction parallel to the ground plane, then the Y axis is in the gravity direction, and the equation of the ground plane is

$$Y = -h$$

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

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

Any fixed value of y is a horizontal line (row) in the image. Scene points that project to that line have constant depth (independent of x). The larger the depth, the nearer the y value is to 0. In the limit as $Z \to \infty$, we have $y \to 0$. Thus, $y = 0$ defines the horizon in this case.

Image motion seen by moving camera

Let's next consider what happens when the camera moves. The motion can be a translation or a rotation (or both). Motion causes the viewer to see the scene from different 3D positions and in different directions. The result is that scene points project to different positions in the image.

If we think of the depth $Z(x, y)$, then each point in the scene that is visible will change position in the image and so we will get a *vector field* of motion across the image, called the *motion field.* We will next examine the motion field for 3D translation motion of the camera. Next lecture we will examine rotational motion.

Translating the camera

Suppose the camera translates with 3D velocity (T_x, T_y, T_z). Forward camera motion is 3D velocity $(0, 0, 1)$. Rightward camera motion is 3D velocity $(1, 0, 0)$. Upward camera motion is 3D velocity $(0, 1, 0)$. When the camera translates, the position of any visible point varies over time. In the camera's coordinate system, the position of the point moves in the 3D direction and speed opposite to the camera. If the camera coordinates of a 3D scene point at time $t = 0$ are (X_0, Y_0, Z_0), then at time t the scene point will be at $(X_0 - T_x t,\ Y_0 - T_y\ t,\ Z_0 - T_z\ t)$ in camera coordinates.

Now let's project the 3D point into the image plane. How does the image position of this point in the image vary with time? The image coordinate of the point is a function of t, namely,

$$(x(t), y(t)) \ = \ \left(\frac{X_0 - T_x\ t}{Z_0 - T_z\ t}, \frac{Y_0 - T_y\ t}{Z_0 - T_z\ t} \right) f.$$

Taking the derivative with respect to t at $t = 0$ yields an *image velocity vector* (v_x, v_y) :

$$(v_x, v_y) = \frac{d}{dt}(x(t), y(t)) \mid_{t=0} \ = \ \frac{f}{{Z_0}^2}(-\ T_x\ Z_0\ +\ T_z\ X_0, -\ T_y\ Z_0\ +\ T_z\ Y_0). \tag{4}$$

We will sometimes speak of the *motion field* (v_x, v_y) or *image velocity vector field* to be the 2D vector function, defined in the image plane. As we will see next, for camera translation, the velocity field depends on image position (x, y) and on the depth Z_0.

Lateral translation

Consider the case that $T_z = 0$. This means the camera is moving in a direction perpendicular to the optical axis. One often refers to this as *lateral motion*. It could be left/right motion, or up/down motion, or some combination of the two. Plugging $T_z = 0$ into the above equation yields:

$$(v_x, v_y) = \frac{f}{Z_0}(-T_x, \ \ -T_y) \ \ .$$

Note that the direction of the image velocity is the same for all points, and the magnitude (speed) depends on inverse depth.

A specific example is the case $T_x \neq 0$, but $T_y = T_z = 0$. The motion field corresponds to the camera pointing out the side window of a vehicle (car, train) as it moves forward. If we restrict the scene to be a single ground plane $Y = h$, then from the ground plane equation from earlier we have the image velocity is

$$(v_x, v_y) = -\frac{T_x}{h}(y, 0).$$

This produces a *shear field*, where the x-velocity is 0 at $y = 0$ (the horizon) and increases linearly with y. You have seen this shear motion pattern many times in your life when looking out the side window of the car or train.

Forward translation

Next take the case of forward translation ($T_x = T_y = 0$ but $T_z > 0$). In this case Eq. (4) reduces to

$$(v_x, v_y) \ = \ \frac{T_z}{Z_0}\ (x\ ,\ y) \tag{5}$$

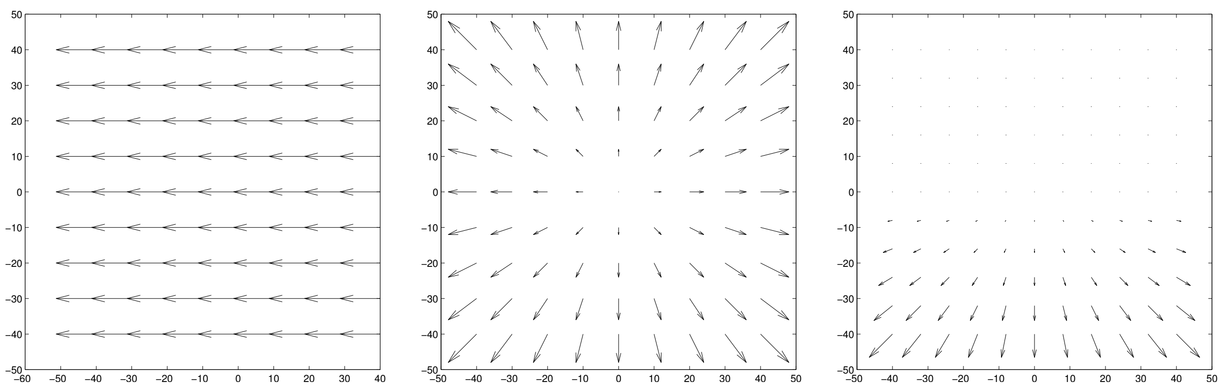
Verify that this field points away from the origin $(x, y) = (0, 0)$, and that the image speed (the length of the velocity vector) is

- proportional to the image distance from the origin i.e. $|(x, y)|$,
- inversely proportional to the depth Z
- proportional to the forward speed of the camera T_z.

In the case of the ground plane example from earlier ($Y = -h$, $y = \frac{hf}{Z}$), we get

$$(v_x, v_y) = \frac{yT_z}{hf} (x , y) = \frac{T_z}{hf} (xy , y^2)$$

See figure below and to the right. Note that this is the motion field that you see when you are walking or driving forward and the ground is an infinite flat plane.



$\mathbf{T} = (1, 0, 0)$ lateral + 'Z = constant' | $\mathbf{T} = (0, 0, 1)$ forward, and Z = constant | $\mathbf{T} = (0, 0, 1)$ forward and ground plane

General translation

Returning to Eq. (4), let's take the case of a general non-lateral translation direction i.e. $T_z \neq 0$ and either T_x or T_y are not zero. In this case we can rewrite Eq. (4) as:

$$(v_x, v_y) = \frac{T_z}{Z_0} (x - f\frac{T_x}{T_z}, \ y - f\frac{T_y}{T_z})$$

Again, image speed is proportional to T_z and inversely proportional to depth Z. In addition, the image velocities point away from a particular image position,

$$(f\frac{T_x}{T_z}, \ f\frac{T_y}{T_z})$$

which is sometimes called the *direction of heading.* Note the magnitude of the velocity vector (the speed) depends linearly with the distance of image position (x, y) from the focus of expansion.

Vanishing points

One of the most interesting phenomena in perspective geometry is that parallel lines in the 3D world typically project to non-parallel lines in the image which intersect at a single point, called the *vanishing point*. Parallel lines in 3D are quite common in man made environments. The boundaries of floors, ceilings and doorways typically align with a natural XYZ coordinate system. These surface boundaries also typically produce edges in images. There are typically many other lines in the scene that are parallel to each of the XYZ axes as well. For example, furniture such as desks and shelves often consist of rectangular surfaces placed so that their natural coordinate system is parallel to the scene's coordinate system. Thus the images of such scenes often contain vanishing points.

Interestingly, to derive the expression for a vanishing point, we use a very similar argument to what we used above for the motion field under camera translation. Take a point (X_0, Y_0, Z_0) in space and a direction (T_X, T_Y, T_Z). This defines a line

$$(X_0, Y_0, Z_0) + t(T_X, T_Y, T_Z).$$

We consider the case that the camera center $(X, Y, Z) = (0, 0, 0)$ does not lie on the line. In this case, the camera center and the line together define a unique plane. The image projection of the line is the intersection of this plane with the image projection plane. (The intersection of the two planes is a line.) From Eq. 1, this image line is:

$$(x(t), y(t)) \;=\; (\,\frac{X_0 + T_X\ t}{Z_0 + T_Z\ t}, \frac{Y_0 + T_Y\ t}{Z_0 + T_Z\ t})\ f.$$

It is not at all obvious from the above expressions that $(x(t), y(t))$ define a line. The fact that it *does* define a line follows from the geometric argument above, namely that this set of points is the intersection of two 3D planes.

If $T_Z \neq 0$, then we can let $t \to \infty$ and we get

$$(x_v, y_v) \;=\; f(\,\frac{T_X}{T_Z}, \frac{T_Y}{T_Z}) \tag{6}$$

This is called the *vanishing point* of the line. If $T_Z = 0$, then our 3D line lies in a constant Z plane, and the image projection of the line is

$$(x(t), y(t)) \;=\; (\,\frac{X_0 + T_X\ t}{Z_0}, \frac{Y_0 + T_Y\ t}{Z_0})\ f.$$

As $t \to \infty$, we go to a point at infinity in direction (T_x, T_y) in the image plane. That is, the projected lines in the image are (in this case) parallel. So they don't intersect.

Notice that the vanishing point is only defined by the direction vector (T_X, T_Y, T_Z), not by the point (X_0, Y_0, Z_0). This means that we can vary the latter point however we like (not just along the line) and we will always get the same vanishing point. Thus, *any set of 3D parallel lines have a common vanishing point.*

Note the above derivations are similar to the analysis of translational camera motion which we saw earlier. Why? When you translate the camera, all points in the scene travel along straight lines relative to the camera position. From the camera's perspective, there is no difference between translating the camera and keeping the world fixed versus translating the world and keeping the camera fixed. *The direction of heading is thus mathematically equivalent to the vanishing point.*

Finally, you may have heard of vanishing points in the context of classical painting and drawing. In particular, you may have heard of 1,2, and 3 point perspective. What do these refer to? In a scene where there are many 3D lines/edges that are parallel to the scene's X, Y, Z axes, the image projection projection plane will contain up to three vanishing points, corresponding to the XYZ axes. (It can contain more, for example, if there parallel lines in directions other than the scene's XYZ axes.)

Consider a typical man-made scene that contains three orthogonal sets of lines. If the camera's Z axes are aligned with one of these sets of lines, then the other two scene axes will be parallel to the camera's projection plane. In this case, there is one vanishing point in the projection plane, namely at $(x, y) = (0, 0)$, and two vanishing points at infinity. We say such an image is a *one point perspective*, i.e. there is one *finite* vanishing point, namely at the optical axis. The other two vanishing points are at infinity. A *two point perspective* arises when just one set of the scene's parallel lines is parallel to the camera projection plane. In this case, the other two sets of parallel lines each define a vanishing point that is not at infinity. Finally, if none of the three camera axes are parallel to the scene's three sets of orthogonal parallel lines, then we have a *three point perspective.* There are three finite vanishing points. See the slides for examples.

