Image motion seen by moving camera

Let's next consider what happens when the viewer moves. The motion can be either a translation or a rotation (or both). Motion causes the viewer to see the scene from different 3D positions and in different directions, and the result is that scene points project to different positions in the image. Later in the course we will see that if a vision system can measure the changes in image positions over time, then it is possible to compute the 3D positions of the points in the scene.

Translating the camera

We begin by looking at camera translation. Suppose the camera translates with 3D velocity (T_x, T_y, T_z). For example, 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 point at time $t = 0$ are (X_0, Y_0, Z_0), then at time t the 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) :

$$(v_x, v_y) = \frac{d}{dt}(x(t), y(t)) \mid_{t=0} \;=\; \frac{f}{{Z_0}^2}(-\, T_x\, Z_0 \;+\; T_z\, X_0, -\, T_y\, Z_0 \;+\; T_z\, Y_0). \tag{1}$$

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:

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

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 the (passenger!) seat of the car, as the car drives forward. If we restrict the scene to be a single ground plane $Y = h$, you can observe from the ground plane equation from last class, 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 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. (1) reduces to

$$(v_x, v_y) \;=\; \frac{T_z}{Z_0}\,(x\;,\;y) \tag{2}$$

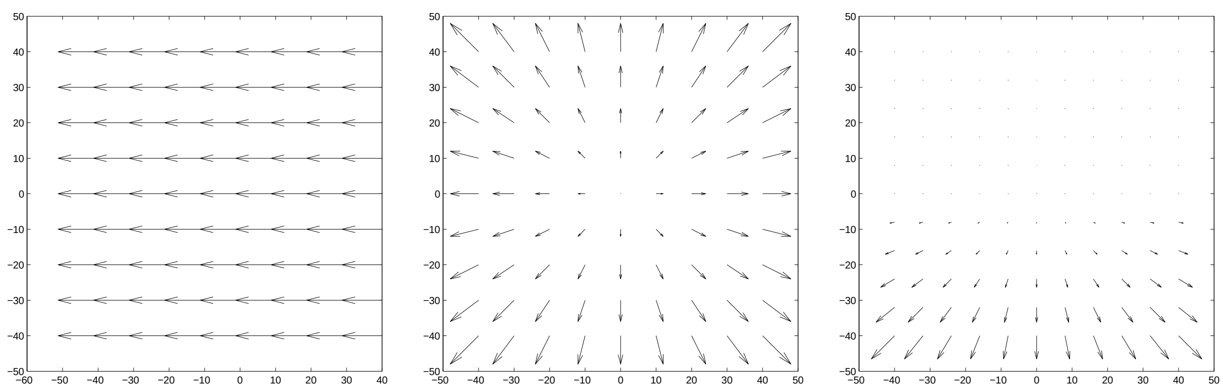
Verify that this field points away from the origin $(x, y) = (0, 0)$, and that the image speed (the length of the velocity vector) is

- 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 a ground plane ($Y = -h$, $y \;=\; \frac{hf}{Z}$), verify that 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.



$\mathbf{T} = (1, 0, 0)$ 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. (1), 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. Verify that we can rewrite Eq. (1) as:

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

There are a number of important properties of this general translation field. As we saw earlier, 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*. Notice that the point $(x, y) = (f\frac{T_x}{T_z},\ f\frac{T_y}{T_z})$, is on the projection plane $Z = f$ at the 3D point $(x, y, Z) = (f\frac{T_x}{T_z},\ f\frac{T_y}{T_z}, f)$ which is parallel to (T_x, T_y, T_z), namely to the direction of translation (heading).

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 orthogonal coordinate system. These surface boundaries also typically produce visible "edges" in images. There are typically many other lines in the scene that are parallel to each of these axes as well. For example, furniture such as desks and shelves often consider of rectangular surfaces and are often placed so that its 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 when we discussed image 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.

Note that 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.) This image line can be parameterized by t:

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

It is not immediately 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.

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

$$(x_v, y_v) \ = \ f(\frac{T_X}{T_Z}, \frac{T_Y}{T_Z}) \tag{3}$$

which is the *vanishing point* of the line. If $T_Z = 0$, then the 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 $t \rightarrow \infty$, 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 and don't intersect. Thus, if we have a set of parallel lines whose direction is perpendicular to the Z axis, then the vanishing point is a point at infinity in the direction of the lines.

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 (which might be finite, or at infinity). Thus, any set of 3D parallel lines have a common vanishing point.

Also notice that the above derivations are similar to the analysis of translational camera motion which we saw earlier in the lecture. 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. Recall that when you translate the camera, all points move away from the direction of heading. *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 typically contain three vanishing points. (It can contain more, for example, if there parallel lines in directions other than the scene's XYZ axes.)

Suppose that the scene's X, Z axes are north and west, and Y is the gravity direction. If the camera is pointing in the scene's X (or Z) direction and the camera's y axis is parallel to gravity, then the image is said to be a *one point perspective*, which means that there is one *finite* vanishing point, namely at the optical axis. The other two vanishing points are at infinity. A *two point perspective* arises, for example, when the camera's Y axis is parallel to the line of gravity, but the camera's X and Z axis differ from the scene's X and Z axes. In this case, the scenes X and Z axes both produce finite vanishing points. Since the gravity direction Y produces a vanishing point at infinity, there are only two finite vanishing points. We call this a two-point perspective because there are two finite vanishing points. Finally, if none of the three camera axes are parallel to the scene's XYZ axes, then you have a *three point perspective*. There are three finite vanishing points. See the slides for examples.