At the end of last lecture, I sketched out how the visual system detects motion using cells in V1 that are oriented and also sensitive to changes in image intensity over time. Today I will elaborate on how this works.

I will begin by sketching out the general constraints of a computational model. The first challenge is to be clear what we mean 'image motion'. We think of motion in general as a change in position over time, and usually we are thinking of XYZ points in 3D space. When we speak about image motion, though, we are refering to position (x, y) in an image. The challenge is that, while we know the positions of points (pixels), we don't know what is moving and we need to define that (and infer it) somehow from image intensities.

Intensity Conservation and the Image Motion Constraint equation

The computational problem is to estimate a 2D image velocity vector (v_x, v_y) at each image position (x, y). This is the vector describing the local change in position over time as points (to be specified) move across the visual field. Let $(x(t), y(t))$ be the path of some point over t. Then the velocity of that point is the derivative $(v_x, v_y) = (\frac{dx(t)}{dt}, \frac{dy(t)}{dt})$. In order to estimate image velocity at a point and at some time, we need to relate the intensity information near that (x, y, t) to the motion path $(x(t), y(t))$. The intensity information we'll use here is the partial derivatives $\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t}$ at (x, y, t).

The key idea is to assume that the image intensity of a moving point doesn't change over time. Formally, we assume

$$\frac{d}{dt} I(x(t), y(t)) = 0.$$

This is called *intensity conservation*. Let's just pause here, and discuss this assumption.

Is the intensity conservation assumption always correct? In fact, no. When objects move in 3D space, they may go into and out of shadow or they can change orientation in space, and these changes affect the amount of light reaching each of their surfaces points. Hence each of the surface points (which traces some image path $(x(t), y(t))$) may undergo a changing image intensity over time. The issue here is whether these changes in intensity are typically large or small in relation to the changing position over time, and it turns out that the intensity changes tend to be small. So the assumption provides some leverage, in practice.

A different way to think of the intensity conservation assumption is that it *defines* what we mean by image motion. We're saying that image motion just *is* a path taken by points of some given intensity. Note that when points disappear (for example, one object behind another) the path of a point ends, and this event would need to be detected. As we will see below, there is another issue that arises as well.

A third comment about the intensity conservation assumption is that it is similar to the assumption we made when discussing how to estimate binocular disparity. With binocular disparity, we assumed that the left and right eye images $I_{left}(x, y)$ and $I_{right}(x, y)$ were the same except for local horizontal shifts by the disparity d which was the quantity that we wanted to estimate. Here with image motion, we assume that image positions of projected 3D scene points are moving over time and that the image intensity of each projected point stays the same over time. Here the quantity we want to estimate is the local velocity (v_x, v_y). The main difference between binocular disparity and motion is that with motion we can't just assume that the change in position is horizontal only.

Let's now examine the intensity conservation assumption more closely. Expanding the total

 ©Michael Langer

derivative gives

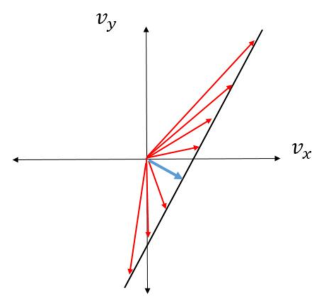
$$\frac{d}{dt}I(x(t),y(t)) = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t}$$

and since we are assuming that the intensity doesn't change over time, we get

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0.$$

This is known as the *motion constraint equation*.

Given a time varying image $I(x,y,t)$, one can compute the three partial derivatives. But can one estimate for (v_x, v_y) from these local derivatives at (x,y,t) alone? Unfortunately not, since the motion constraint equation only gives one linear constraint at each point and this equation has two unknowns, namely v_x and v_y. All we can say is that (v_x, v_y) lies on a particular line in the 2D space of (v_x, v_y). See figure below. The shortest such candidate velocity vector (shown in blue) is normal to the line, and hence it is called the *normal velocity*.



Another way to express the same constraint is

$$(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}) \cdot (v_x, v_y) + \frac{\partial I}{\partial t} = 0.$$

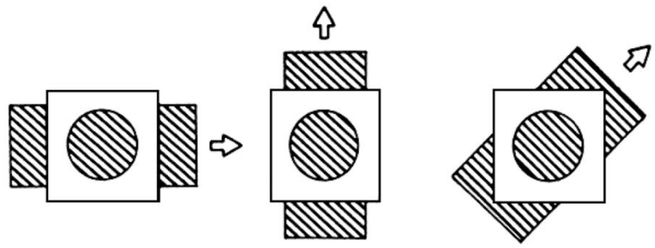
and so we now see that we only have a constraint on the component of the (v_x, v_y) vector in the direction of the spatial gradient of intensity. We will elaborate on this important point next.

Aperture problem

The ambiguity of the motion constraint equation is often called the *aperture problem*. We can think of viewing the image through a small aperture in space-time (XYT) such that only the first order partial derivatives can be computed. Note by "aperture" here, I'm not talking about a camera aperture like in lecture 2. Rather I'm just talking about a receptive field– a limited image window.

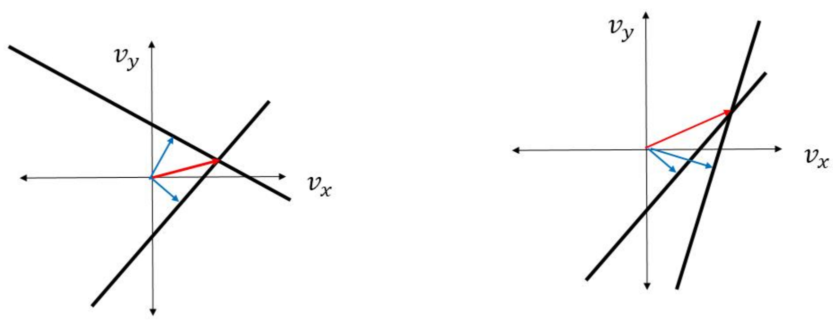
The aperture problem is more general than this, though. It applies anytime one has a moving 1D pattern. For example, the illustration below shows a set of oblique parallel stripes that are moving, either horizontally, vertically, or obliquely. Given only the motion in the aperture, one cannot say what the "true" velocity vector is. ASIDE: This problem is also related to the barber pole illusion:

`http://www.opticalillusion.net/optical-illusions/the-barber-pole-illusion/`



To avoid the aperture problem and estimate a unique velocity vector, one needs two or more such equations. The natural way to do so is to assume that the velocity vector (v_x, v_y) is constant over some local image region, and to combine the motion constraint equations from the two nearby points whose spatial gradients $(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})$ have different directions. (This requirement is not always met, as in the above example.) Since two points with different intensity gradient directions would define two intersecting lines in (v_x, v_y) space and since the true velocity must lie on both these lines, one can solve for the true velocity vector by computing the intersection of the two lines. This is called the *intersection of constraints* (IOC) solution.

As examples, see figure below. The red vector is the IOC solution and the blue vectors are the normal velocities. The one on the right is counterintuitive because both lines have a normal velocity that is downward to the right, but the true solution is upwards to the right. This is surprising because one might have expected that the true solution should be "between" the normal velocity motion vectors defined by the two given constraints. For example, one might expect the solution to be the average of the two normal velocities. See the slides for more detailed versions of these examples.



Motion sensitive cells in V1 and MT

Many orientation selective cells in V1 – both simple and complex – are also sensitive to motion direction. We can model the responses of these cells by combining responses to an image $I(x, y, t)$ at two different times t and $t + \Delta t$. One way to do this is analogous to what we considered with binocular stereo, namely we shift the receptive field between the two times.

See the slides for examples of how this is done, and how this related to normal velocities and motion constraint lines. You will explore this model in detail in Assignment 2.

 ©Michael Langer