In the first half of the lecture, I'll define a general computational problem of estimating motion in an image, and how to solve it. This abstract formulation of the problem is similar to classical computer vision methods for computing local image motion, and the ideas of these models have been used in many human and non-human vision experiments, to understand how the biological motion estimation systems works.

In the second half of the lecture I will sketch a computational model for motion processing in the brain which is in terms of the XYT receptive fields of the V1 cells which we discussed last lecture.

Image motion constraint equation

The computational problem of *local* image motion estation is to estimate the local image velocity (v_x, v_y), which is the vector describing the local change in position over time as points move across the visual field. We would like to make such an estimate at each image position (x, y) and time t, but we can only do so if there is intensity information present that indicates the motion(s). 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 each point. Under certain conditions, this turns out to be enough for a basic formulation of the problem. Note that we do need the intensity to be changing locally over position, since if intensity is constant across a local patch then we cannot say that anything in the patch is moving since all visual directions in the patch look the same.

Suppose that the points in a small local patch have image velocity (v_x, v_y), and I'll say what that means below for points to "move". For now, let's not worry about the units, whether the space units are pixels or photoreceptors, mm on the retina, or visual angle and or whether the time units are seconds, or some frames in a video. With velocity (v_x, v_y), each point by a distance $(v_x \Delta t, v_y \Delta t)$ in a time interval Δt. If the image intensities $I(x, y, t)$ are smooth enough that we can compute local derivatives, and write a Taylor series expansion of the intensities near (x, y, t) as

$$I(x + v_x \Delta t, y + v_y \Delta t, t + \Delta t) \ = \ I(x,\ y,\ t) + \frac{\partial I}{\partial x} v_x \Delta t + \frac{\partial I}{\partial y} v_y \Delta t + \frac{\partial I}{\partial t} \Delta t + H.O.T.$$

where H.O.T. stands for "higher order terms", namely higher than first order derivatives. The partial derivatives are evaluated at (x, y, t).

We now make a key assumption about the motion, namely that the image intensity of a moving point doesn't change over time – this is sometimes called *intensity conservation.* Thus, when a point moves from (x, y) to $(x + v_x \Delta t, y + v_y \Delta t)$ from time t to time $t + \Delta t$, respectively, we have

$$I(x + v_x \Delta t, y + v_y \Delta t, t + \Delta t) = I(x,\ y,\ t).$$

This lets us cancel these two terms in the Taylor series above. If we further ignore the higher order terms, then we have:

$$\frac{\partial I}{\partial x} v_x \Delta t + \frac{\partial I}{\partial y} v_y \Delta t + \frac{\partial I}{\partial t} \Delta t \ = \ 0$$

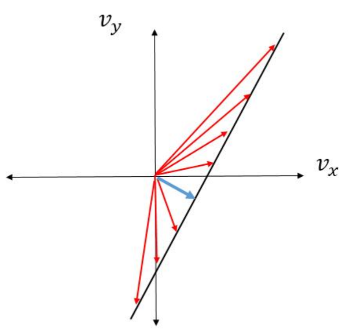
or

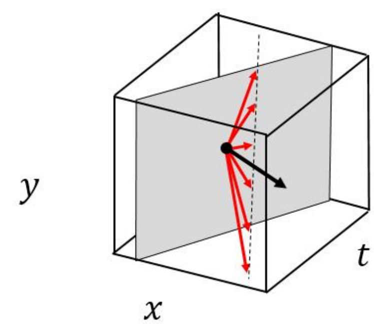
$$\frac{\partial I}{\partial x} v_x + \frac{\partial I}{\partial y} v_y + \frac{\partial I}{\partial t} = 0. \tag{1}$$

The latter is called the *motion constraint equation.* It expresses the relationship between the spatial and temporal derivatives of the image in terms of the image velocity (v_x, v_y). In particular, it expresses a relationship between what we want to estimate – (v_x, v_y) – and the image quantities that we can directly measure, namely partial derivatives of intensity.

The intensity conservation assumption is similar to the assumption we made when discussing how to estimate binocular disparity in lecture 5. 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).

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 Eq. (1) 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 to consider XYT space and write Eq. (1) as

$$(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t}) \cdot (v_x, v_y, 1) = 0.$$

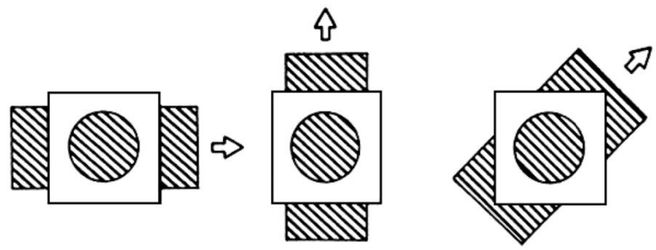
which says that the 3D vectors $(v_x, v_y, 1)$ in red are constrained to be perpendicular to the 3D intensity gradient vector.

Aperture problem

The ambiguity of the motion constraint equation is often called the *aperture problem*, since we can think of viewing the image through a small aperture in space-time 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.

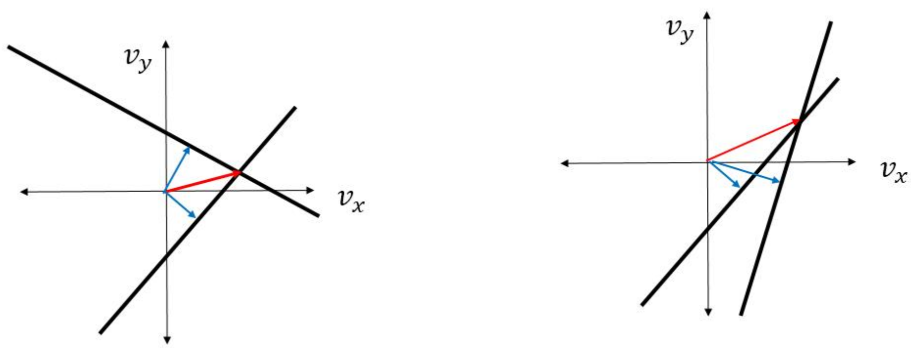
The aperture problem is more general than this, though. It applies anytime one has a moving 1D pattern. For example, the illustration below shows a set of oblique parallel stripes that are moving, either horizontally, vertically, or obliquely. Given only the motion in the aperture, one cannot say what 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 constraints of Eq. (1) from two nearby points whose spatial gradients $(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial x})$ differ. Since the two points would define two different lines in (v_x, v_y) space and 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.[1]



[1]There are many experiments in which humans *do* perceive image velocities using a "vector averaging" solution, rather than an IOC solution.

Gabor cells and the motion constraint equation

Recall the XYT Gabor cells that we discussed last class. These cells were defined by taking a sine wave with some spatial frequency (k_x, k_y) and some temporal frequency ω, and multiplying by a Gaussian window. Let's briefly review a few properties of the underlying sine wave. The sine way travels with velocity $(\frac{2\pi}{N}k_x, \frac{2\pi}{N}k_y, \frac{2\pi}{T}\omega)$ in XYT which is the vector perpendicular to constant values of the sine function. Note this is similar to the idea of the motion constraint equation, where we define "motion" to be a path of a point of some intensity. For the moving sine wave, we define the motion to the normal velocity only, namely the velocity normal to the set of points of constant intensity i.e constant value of sine.

The XYT Gabor cell response to a moving image $I(x, y, t)$ is defined in the usual way by taking the inner product of the Gabor function and the image function over XYT. The response at some time t for a Gabor cell located at position (x_0, y_0) will depend on the image in the recent past before t and in a spatial neighborhood of that point. We are modelling the space-time window by a Gaussian, but of course one needs to note that the response can only depend on the past and so a more accurate model would have a hard cutoff for the window. The same is true for space, in fact, as the Gaussian in theory has an infinite extent.

Such a Gabor will generally give its largest response when the image contains spatial variations in the direction $(\frac{2\pi}{N}k_x, \frac{2\pi}{N}k_y)$ and when the image component of the velocity in that direction matches the normal velocity of the Gabor's underlying sine wave. Note that the response depends on both of these factors.

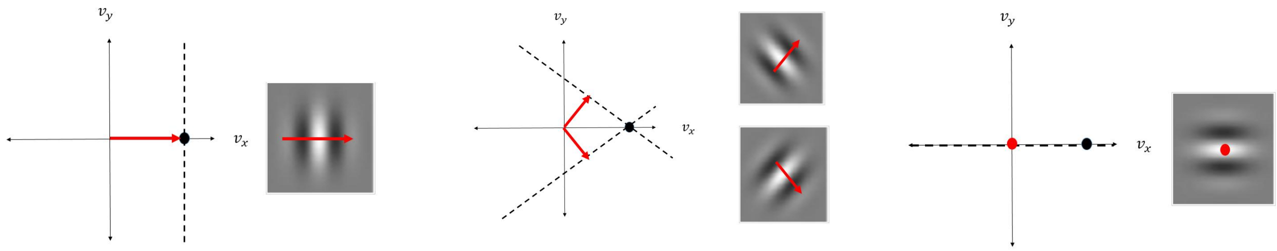
How can a set of XYT Gabors cells be used to estimate the image velocity (v_x, v_y) at a point, when each of the Gabor cells can only detect a normal component of velocity? The answer to this question is similar to intersection of constraint solution described above. Suppose we wanted to design a cell at a further layer of processing, such that this cell would respond best when the image velocity near (x_0, y_0) had some given value (v_x, v_y). Call this a *velocity tuned cell.*

When velocity (v_x, v_y) is present, which Gabor cells would give a good response? The answer is: those Gabor cells whose underlying sine wave has normal velocity $(\frac{2\pi}{N}k_x, \frac{2\pi}{N}k_y, \frac{2\pi}{T}\omega)$ satisfying the motion constraint equation:

$$(\frac{2\pi}{N}k_x, \frac{2\pi}{N}k_y, \frac{2\pi}{T}\omega) \cdot (v_x, v_y, 1) = 0.$$

This defines a family of Gabor functions, namely those whose 2D motion motion constraint line passes through (v_x, v_y). As an example, take a motion $(v_x, v_y) = (v_0, 0)$ at some speed v_0 in the x direction.

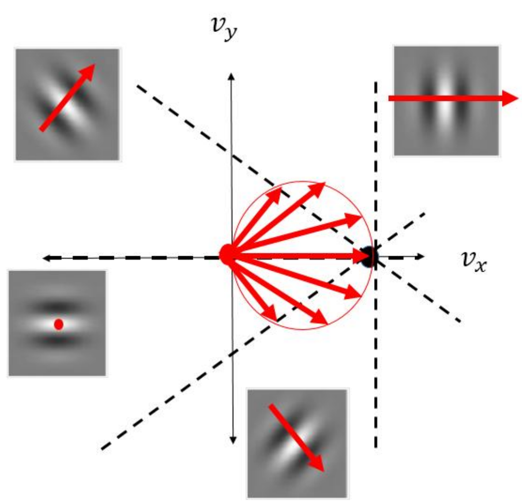
The figure above illustrates several Gabors whose motion constraint lines intersect that particular velocity vector $(v_0, 0)$. The Gabor on the left has a normal velocity equal to $(v_0, 0)$. The two Gabors in the middle panel have normal velocities that are 45 degrees away from $(v_0, 0)$ and have a smaller speed, namely $v_0/\sqrt{2}$. The Gabor on the right has an orientation that parallel to the x axis, and it is most sensitive to zero normal velocity. That is, it prefers no motion in its normal velocity direction. (See Exercises.)



Velocity tuned cells in area MT (middle temporal lobe)

There are cells in the visual system that are velocity tuned. These cells are found in the temporal lobe in an area known as MT (middle temporal).[2] There are direct connections from V1 to MT.

The basic model for an MT cell that is tuned for motion in some direction (v_x, v_y) is illustrated below. This cell has excitatory inputs from a set of XYT Gabor cells, namely from those V1 cells whose underlying Gabors have normal velocity motion constraint line passing through or close to (v_x, v_y).



I will not give further details because it would take too long. If you are curious, have a look at this paper: `http://www.cns.nyu.edu/pub/lcv/simoncelli96-reprint.pdf`
I put one of the figures in the slides and briefly discussed it during the lecture.

[2]MT also has normal velocity cells, but I won't discuss them here.