Last lecture we discussed V1 orientation selective cells and modelled them with Gabor functions. These cells only received input from one image i.e. they were *monocular*. They also did not have any temporal properties, as we were considering static images only. Today we will examine models of cells that are sensitive to both eye's images (binocular disparity tuned cells) and cells that are sensitive to motion. We will have more to say about both of these kinds of cells later in the course.

Estimating binocular disparity

Recall that the left and right eye images are sent to the LGN. The signals are not combined there, however. Rather, the left LGN carries the signals from the right visual field and the right LGN carries signals from the left visual field. However, the left and right eye's signals for each field are fed into different LGN layers and then relayed separately to V1. There is much known about how the two eye images are combined in V1. Our aim here is not to understand these circuit level details. Instead, let's try to understand the computational problem these cells are solving. We'll the consider a solution to his computational problem that can be expressed in terms of the Gabor functions that we discussed last class.

Consider a visual direction (x_0, y_0) in each eye's coordinate system. Think of this as two parallel rays reaching each eye. Suppose that the left and right images near this direction have similar intensities, except for a horizontal shift (see analglyph slide). This shift is the binocular disparity, which we discussed earlier in the course. This shift might vary over the image, because the depths of visible points will vary and the shift depends on depth.

Near (x_0, y_0), the visual system could attempt to estimate the shift to be the value d that minimizes

$$\sum_{(x,y)\in\mathcal{N}(x_0,y_0)} (I_{left}(x+d, y) - I_{right}(x, y))^2$$

where the sum is over (x, y) coordinates in a neighborhood of (x_0, y_0). The idea of this computation is that if you shift the left image by the correct disparity d, then the shifted left image should correspond pixel-by-pixel to the right image – at least in the local patch where the disparity is roughly constant. In that case, the above sum of squared differences should be 0 for the correct d. For other values of d, sometimes the left image will be brighter at a pixel than the right image and sometimes it will be darker, so the intensity difference at that pixel will be non-zero. We square the intensity differences because we only case how much it is different from 0, not whether it is positive or negative.[1] The idea for estimating disparity d near (x_0, y_0) for a particular left-right image pair is to choose the d value that minimizes this sum of squared differences.

Technical Note: Recall that $d = x_l - x_r$ and so when the left eye image is shifted to the right, we have positive disparity. Also note when $d > 0$, the term $I_{left}(x + d, y)$ corresponds to a *leftward* shift. The idea here is that, to find the correct match d which minimizes the above sum, we would need to do a leftward shift of the left image to cancel the positive disparity between the left and right images i.e. to properly register the left and right images so that their point-to-point difference would be 0.

While the above computational model works well (and is the basis for many computer vision methods for binocular stereo), the model is not "biologically plausible" since it ignores the processing

[1] We could have alteratively taken the absolute value, and indeed some computational models do that.

that is done in the brain, namely cells with local receptive fields. In particular, the first cells in the visual pathway that analyze both eye images are in V1, and these cells are orientation tuned. We model them using Gabor functions. So let's consider a model based on such cells. We restrict ourselves to vertical oriented cells. (In Assignment 2, you will explore why.)

Up to now we have considered monocular complex cells in V1 which were constructed from simple cells. We now consider binocular complex cells which are constructed from simple cells, namely Gabor cells for the left and right eyes. Using a similar idea as the computer vision method above, for a given (x_0, y_0), we could consider the following expressions which estimates the similarity in the responses of left eye cells at $(x_0 + d, y_0)$ and right eye cells at (x_0, y_0):

$$(< cosGabor(x - x_0 - d, y - y_0), I_{left}(x, y) > - < cosGabor(x - x_0, y - y_0), I_{right}(x, y))^2$$

$$+ (< sinGabor(x - x_0 - d, y - y_0), I_{left}(x, y) > - < sinGabor(x - x_0, y - y_0), I_{right}(x, y))^2$$

Here the d shift is for the sine and cosine Gabor for the left eye, that is, the Gabors for the left eye are centered at $(x_0 + d, y_0)$ and the Gabors for the right eye are centered at (x_0, y_0).

The idea is that if we place a cosine Gabor template at $(x_0 + d, y_0)$ in the left image and at (x_0, y_0) in the right image and if d is the true disparity in the images – then the linear responses of the left and right eye cosine Gabors will have the same value, so if we subtract one from the other then we get 0. Similarly, the linear responses of the sine Gabors will have the same value, so if we subtract one from the other we get 0. The shift d that minimizes the sum of squared differences in the above expression would the best estimate of the disparity.

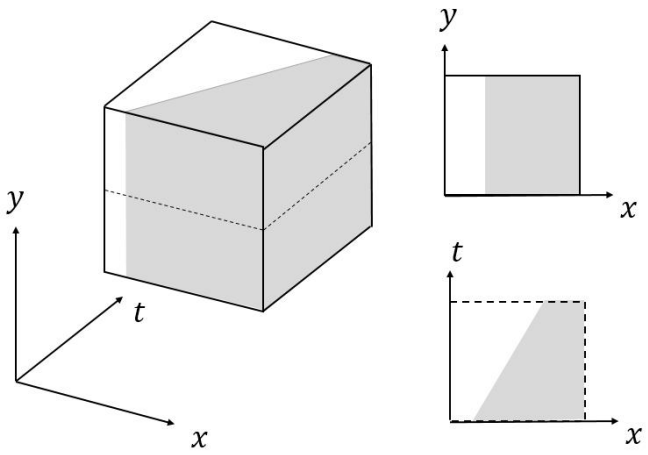
Please see the figures in the slides in this lecture. We will return to these ideas again a few lectures from now, and in Assignment 2. For now, we turn to time varying properties of cells in V1 and consider image motion.

Time varying images

Up to now we have said very little about how images vary over time. But of course they often do. Let's think of an image as a function of x, y and t, namely $I(x, y, t)$. In particular, the variations are often caused by objects moving or by the eye moving (changing the direction of view, or moving through space when the animal moves or sways).

Let's consider very simple image changes over time that are due to objects moving. An example is a vertical intensity edge drifting to the right. The figure below shows a small space-time cube through which the edge passes, and it shows an XY slice and an XT slice through the cube. (See lecture slides.) This edge drifts to the right with speed v_x so v_x is the slope of the edge in the XT slice, where slope is measured $\frac{dx}{dt}$, not $\frac{dt}{dx}$.

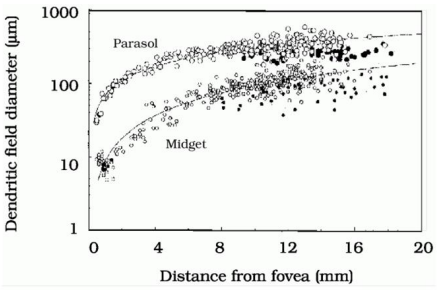
See the lecture slides for two other examples of $I(x, y, t)$. One is just a moving bar instead of a moving edge. The second is more interesting and shows a real video of a person walking from left to right. An XT slice reveals the motion pattern of the person's legs.



Retinal ganglion cells and time-varying images

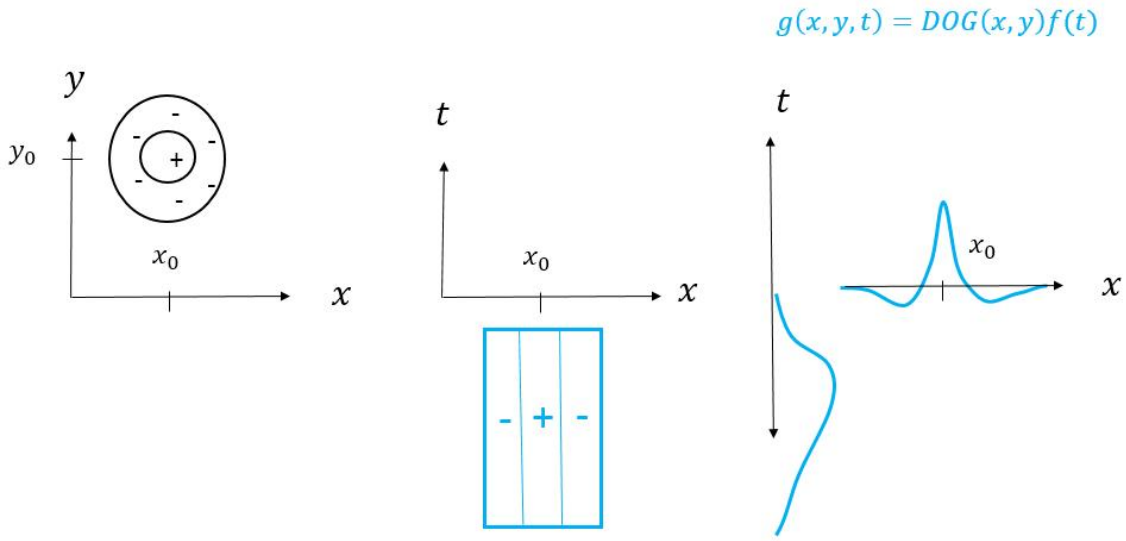
To model how the visual system estimates image motion, we need use model components that build on cells that respond to time varying images. Let's begin in the retina. Photoreceptors measure light intensity continuously over time (unlike digital video cameras which take discrete samples). A photoreceptor does not respond instantaneously, however. Rather there is a delay in the response. See illustration in the slides.

Retinal ganglion cells also have a temporal dependent response: they do not respond instantaneously, but rather their response at any time t depends on the image intensities in their receptive fields over some time period prior to t. It turns out there are there are two classes of retinal ganglion cells. These two classes differ in several ways, including their temporal responses.



One way the two classes differ is the size of their receptive fields. The first class (called "midget") has receptive fields that are roughly factor of 10 smaller than the second class (called "parasol"). This difference in receptive field sizes is illustrated in the figure below, which shows samples of receptive field sizes measured by the diameter of dendrite tree (done using staining techniques). Notice that the sizes of both classes of cells increase from the center of the field of view into the periphery. Think of the σ of the DOG functions as increasing with eccentricity. This size changes are a big effect. Note the abscissa ("x axis") in the figure is on a linear scale whereas the ordinate ("y axis") is on a log scale.

Another way that the two types of cells differ is in their temporal properties. The response of any ganglion cell at any time t will depend on the image in some local spatial neighborhood and on some local time interval prior to t. Let's first discuss a basic model for the small (midget) retinal ganglion cells. Consider the XT slice for the cell shown below in the middle. Its temporal receptive field lies in the range $t < 0$ and this is meant to illustrate the receptive field weights for determining the response (firing rate) at time $t = 0$. The receptive field can be non-zero only for $t < 0$ since the cell's response cannot depend on something that hasn't happened yet.



For simplicity, I have given this cell a *separable* response function, namely a DOG in XY and a function $f(t)$ to describe the temporal dependence. So its space time receptive field is modelled as

$$DOG(x, y; x_0, y_0, \sigma)\ f(t)$$

The x_0, y_0 pair denote the center of the cell in XY space as usual, and $f(t)$ indicates how the cell's response at time t will depend *only* on the past. The spatial σ defines the spatial window of the receptive field. So its reponse at time $t = 0$ is modelled as

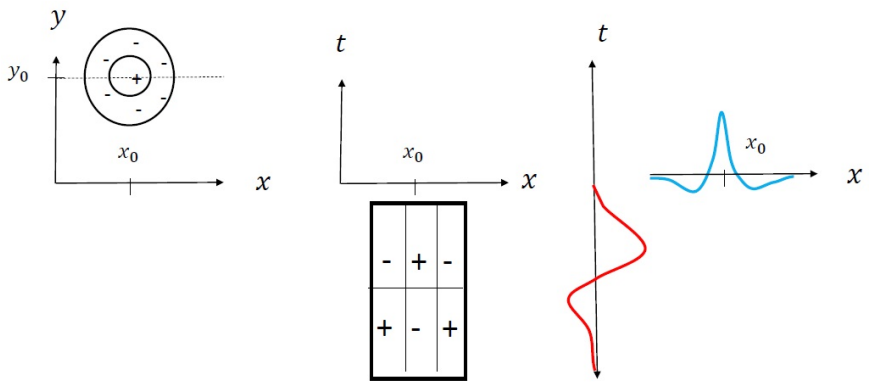
$$\sum_{x',y',t'} DOG(x_0 + x', y_0 + y', \sigma)\ f(t')\ I(x', y', t')$$

In the slides, I gave an example of two different stimuli. One is a single bright spot of light that lasts for a duration longer than the integration time of the cell (i.e the y length of the receptive field in the figure); the other was a single bright spot of light that lasts for only a brief duration. If the brief duration spot were proportionally brighter then the longer spot, then these two stimuli could give the same response at time $t = 0$ i.e. if the above summations in the two cases gave the same result. In this sense, we say that the model cell is not sensitive to temporal properties of the stimulus.

Next consider the larger retinal ganglion cells. These cells *are* sensitive to changes in the stimulus over time. Here the dependence on time $f(t)$ has both a positive and negative part. This cell would

not respond well to a *static* bright spot in its center, since the cell's negative and positive weights in the center would give opposite weights that cancel out over time. But notice that the cell would give a response to a *brief* spot of light in the recent past. It would also give a negative response (be inhibited) by a brief spot of light that occured slightly longer ago in the past. Thus, such a cell would be more suitable at coding temporal events such as spots of light appearing or disappearing at a pixel at some time.

ASIDE: I didn't go into details in the lecture, but you should also consider the spatial properties of the receptive field, namely that it is center-surround. This lateral inhibition effect serves to enhance spatial changes in the image, as you know. But now there is a temporal dependence that interacts with the spatial dependence. The surround can be either inhibitory or excitatory, depending on which time in the past we are talking about. In fact, spatiotemporal receptive field are a bit more complicated than this, namely they are not separable. Much is known about these cell properties, but this is not so important for our needs in this course – and so I will leave it here, with this basic example.



Directionally selective cells in V1

At the end of the lecture, I briefly described directionally selective cells in V1. These are cells that have oriented structure in XY, and that are sensitive to motion in a direction perpendicular to the orientation. I only gave a sketch for now, so *please see the slides*. I will return to these cells later and we'll say more about how motion is computed.