

Introduction to Neural Networks

U. Minn. Psy 5038

Spring, 1998

Representation of visual information

Primary visual cortex: anatomy, physiology, and functions

Outline

- Overview of anatomy from eye to cortex

 - chiasm

 - lateral geniculate nucleus

 - primary visual cortex (V1, striate, 17)

 - anatomical organization

 - simple, complex, end-stopped

 - model of simple cells

- Functions of primary visual cortex

 - binocular vision and stereopsis

 - motion

 - spatial frequency filtering

- Spatial frequency filtering

 - psychophysical evidence

 - neurophysiological evidence

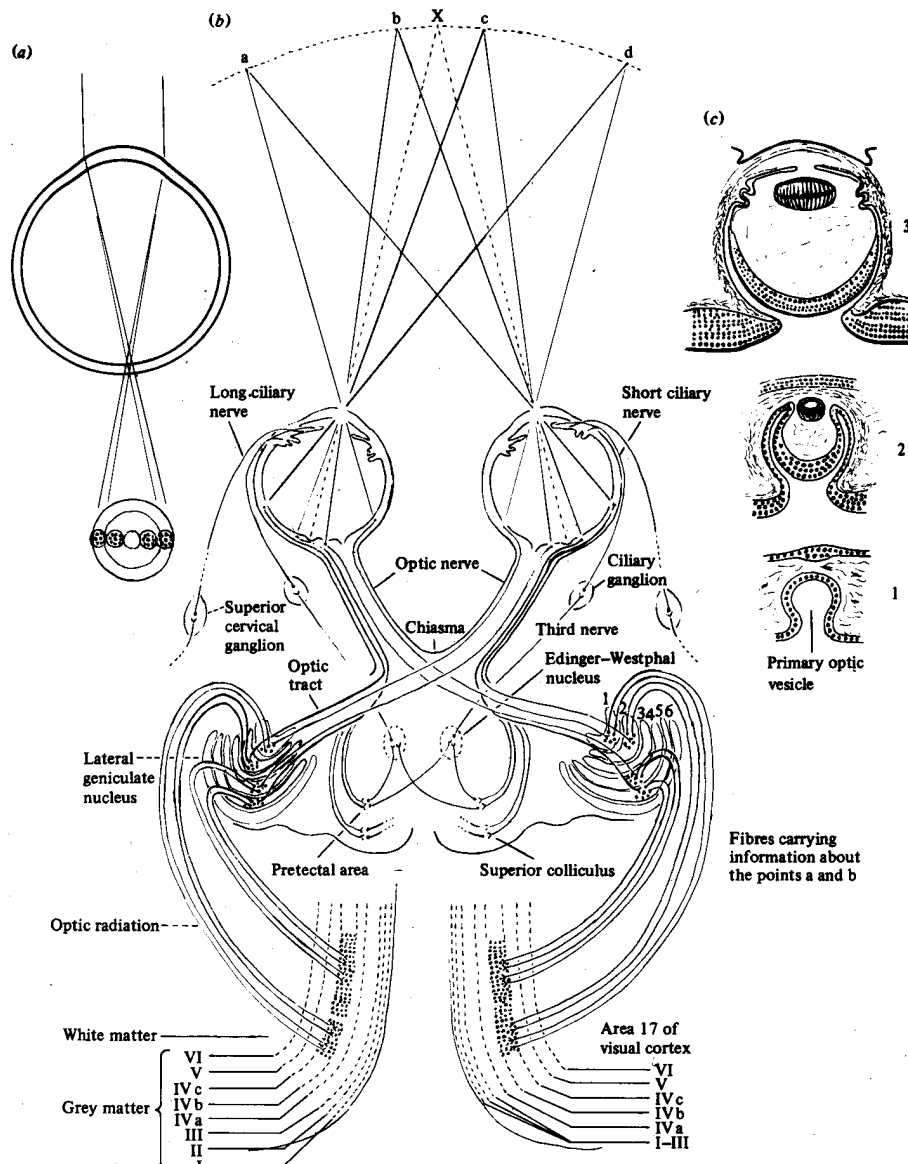
 - Why spatial filtering?

 - edge detection

 - cortical basis set and economic representations

Pathways from eye-to-cortex

Schematic view



The primate retina has about 10^7 cones that send visual signals to the optic nerve via about 10^6 ganglion cells. The optic nerves from the two eyes meet at the optic chiasm where about half of the fibers cross over and the other half remain on the same side of the underside of the brain. Before synapsing in the lateral geniculate nucleus, about 20% of these fibers that now comprise the optic tract branch off to the superior colliculus—a structure involved with eye movements. The rest

of the optic tract fibers synapse on cells in the lateral geniculate nucleus. Cells in the lateral geniculate nucleus send their axons in a bundle called the optic radiation to layer IV (one of six layers) of primary visual cortex.

Functions of the Chiasm and Lateral geniculate nucleus (LGN)

The optic chiasm routes neuronal information so that information from corresponding points on the left and right eyes can come together at cortex for binocular vision, and in particular stereo vision. Typically animals with frontal vision have nearly complete cross-over, and animals with lateral eyes (e.g. fish) have little or no cross-over.

The nervous system has gone to considerable length to bring information from the two eyes together early on. Although there seems to be little if any interaction between neurons in the LGN, the arrangement of the optic chiasm is the first step towards the eventual construction of a topographic cortical map.

In fact, there is a general principle that becomes even more apparent when one looks at maps that pervade cortical organization:

Neural computations often require close physical connectivity between neurons

Anderson has a discussion of topographic and tonotopic maps in cortex. Later on we will see some of the consequences of self-organizing principles that serve to minimize wiring length when we study **Kohonen networks**.

The neurons of lateral geniculate nucleus do more high-pass filtering, and the cells are characterized by fairly symmetrical center-surround organization like the ganglion cells. They show even less response to uniform illumination than ganglion cells. Despite the fact that neurons from the two eyes exist within the same nucleus, no binocular neurons are found in LGN. We have to wait until cortex to see binocular neurons. Although the LGN is often considered a relay station, feedback from cortex suggests possible role of attention mechanisms (see Crick, 1984 for a speculative neural network theory of LGN and reticular function).

Sillito et al. (1994) have found Feature-lined synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, 369, N. 9, 479-482.

Although we will bypass a treatment of the superior colliculus, its primary role is in the control of eye movements-- a highly non-trivial problem requiring coordination of head and eye movements in the context of a constantly changing environment.

Primary Visual Cortex

Primary visual cortex: Large scale organization

■ Topographic map

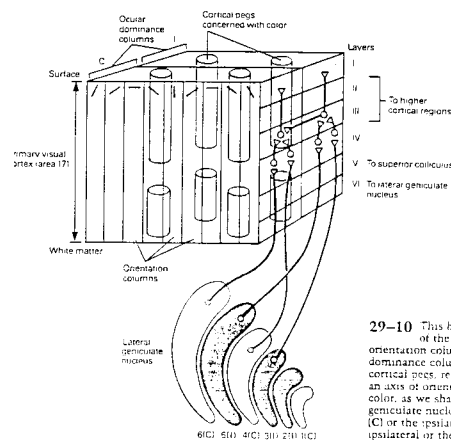
The striate cortex is laid out as non-linear topographic map with 80% of cortical area devoted to about 20% of visual field, reflecting the higher acuity of foveal vision. Because of the cross-over at the optic chiasm, the left visual field (right retina) maps to right hemisphere.

Axons from the optic radiation synapse on layer IV neurons of the primary visual cortex (also known as area 17, striate cortex or V1). Cortex is anatomically structured in layers, numbered from I (superficial) to VI (deep).

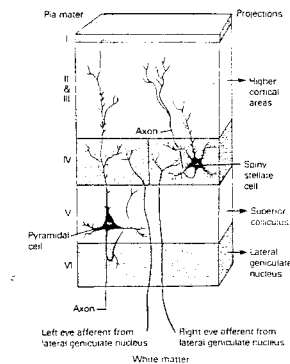
■ The hypercolumn of visual cortex

The cells of the primary cortex are organized into columns running roughly perpendicular to the surface in which cells tend to have the same orientation preference and degree of binocularity. A "hypercolumn" is a group of columns spanning all orientations and both eyes. In the cortex, we see for the first time binocular cells.

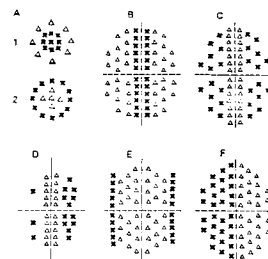
Embedded in the cortical hypercolumns are cytochrome oxidase blobs in which are found opponent color cells that seem to lack strong orientation selectivity (Livingstone, M. S., & Hubel, D. H., 1984; Livingstone, M. S., & Hubel, D. H., 1987).



29-10 This basic cortical module (hypercolumn) in area 17 of the visual cortex contains a complete set of orientation columns representing 360° and a set of ocular dominance columns. Each hypercolumn also contains several cortical 'pegs', regions of cortex in which the cells do not have an axis of orientation. The cells in the pegs are concerned with color, as we shall see in Chapter 30. Each layer of the lateral geniculate nucleus receives input from either the contralateral (C) or the ipsilateral (I) eye and projects in turn to the ipsilateral or the contralateral ocular dominance columns.



28-14 The afferent and efferent connections of the primary visual cortex are made in specific layers of cortex.



29-7 Comparison of the receptive fields of neurons in the retina and in the lateral geniculate nucleus with those of simple cortical cells in area 17. A. Cells of the retina and lateral geniculate fall into two classes: on-center (1) and off-center (2). B-F. Neurons of the primary visual cortex also fall into two major classes: simple and complex. Each of these classes, moreover, has several subclasses. This is illustrated here for simple cells. Despite this variety, however, all simple cells are characterized by three features: (1) specific retinal position, (2) their discrete excitatory (x) and inhibitory (i) zones, and (3) specific axis of orientation. For simplicity, only receptive fields with a vertical axis of orientation from 12 to 6 o'clock are shown in this figure; each has a rectilinear configuration. In fact, each region of the retina is represented in area 17, not only for this but for all axes of orientation—vertical, horizontal, and various obliques. (Adapted from Hubel and Wiesel, 1962.)

Primary visual cortex: Neuron properties

Apart from the neurons the LGN fibers synapse on, and in contrast with receptive field characteristics of earlier neurons, many cortical cells are:

- orientation selective
- binocular
- spatial frequency selective, with narrow tuning and
- motion selective
- spatial phase selective

Some of the major contributions to our understanding of visual cortex was due to the research of Hubel, D. H., & Wiesel, T. N. (see 1968 reference). Hubel and Wiesel won the Nobel prize for their work.

■ Simple cells

There are two main types of cells. The *simple* cells are roughly linear except for rectification, are spatially and temporally band-pass, and show spatial phase sensitivity. A first approximation model for simple cell response firing rate (in impulses/sec) is:

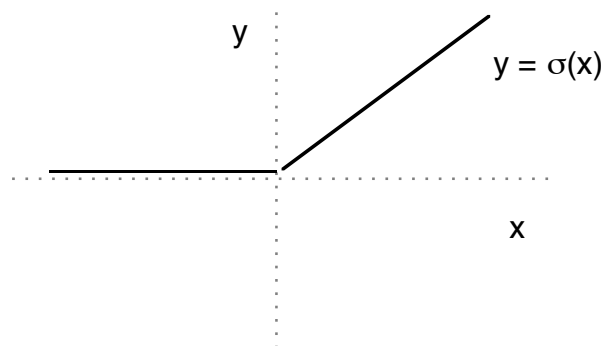
$$R = \sigma \left[\sum_{i,j} W_{ij} L_{ij} \right]$$

You can see that this has the same form as the generic neuron model, except that the inputs are the physical stimulus values.

Where W_{ij} are the receptive field weights, and L_{ij} the image intensity values at spatial location (i,j) . An example would be a Gabor function (see discussion and figures below):

$$W_{ij} = e^{\frac{-i^2-j^2}{2\sigma^2}} \cos(2\pi(f_x i + f_y j) + \phi)$$

The half-wave rectification operation, σ , sets negative values to zero, and is linear for positive values:



And as we saw at the beginning of the course, a better model is obtained by replacing the straight sloping line with one that saturates at high values. This model is steady state. To include time domain dependencies requires the introduction of a band-pass temporal tuning characteristics.

■ Complex cells

The second major class of neurons is that of *complex* cells. Like simple cells, complex cells are spatially and temporally band-pass, show orientation and motion direction selectivity, but are insensitive to the phase of a stimulus such as a sine-wave grating. Rather than half-wave rectification, they show full-wave rectification. A model for complex cells would resemble the sum of the outputs of several subunits positioned at several nearby spatial locations. Each subunit would resemble simple cell with a linear spatial filter followed by a threshold non-linearity.

One way of obtaining the phase insensitivity would be to use subunits with cosine and sine phase receptive fields. We see below how a neural network can be built that can be used to detect edges--it combines simple cell outputs into outputs similar to those of complex cells.

The motion selectivity could be built in with appropriate inhibitory connections between subunits. Full-wave rectification could be built with subunit pairs that have excitatory and inhibitory receptive fields centers. Both simple and complex cells show contrast normalization--a feature not included in the above simple model. For a discussion of models of simple and complex cells see Heeger (1991).

■ End-stopped cells

A third class of cells are the *end-stopped* (or "hyper-complex") cells that have an optimal orientation for a bar or edge stimulus, but fire most actively if the bar or edge terminates within the receptive field, rather than extending beyond it. It has been suggested that these cells act as "curvature" detectors. (Dobbins, A., Zucker, S. W., & Cynader, M. S., 1987).

These cells are also thought to be important for detecting occluding surfaces and the perception of illusory contours.

Functions of Primary Cortex

■ Stereo

As mentioned earlier, primary cortex brings together information from the two eyes in single neurons. This information is important for coordinated eye movements and stereo vision. Although V1 cells are predominantly binocular, it was at first thought that disparity selectivity did not arise until V2 (Hubel and Wiesel, 1970). However, there is now evidence for disparity selective cells in V1 and V2 (Poggio, G., F., & Poggio, T., 1984). Disparity selectivity is a trivial task for single bar stimuli, and it wasn't until relatively recently that neurons were found that effectively solve the problem of false matching (Poggio and Talbot, 1981).

One possible algorithm for stereo vision is discussed in: Poggio, T. (1984). Vision by Man and Machine. Scientific American, 250, 106-115.

This algorithm is related to **Hopfield networks** that we will study later in this course.

Stereo vision has received a lot of attention in both computer and biological vision over the last 15 years. Later we will look at a neural network model of stereopsis.

■ Motion

The directional selectivity of cells in striate cortex provide a form of early motion detection, akin to that described for invertebrate and rabbit peripheral vision. This detection is only local and thus ambiguous. Cortical cells suffer from the "aperture problem", and further computation is required to disambiguate object motion. Cortical cells are also selective for speed (Orban et al., 1983). Both the motion selectivity and binocularity suggest a general hypothesis for cortical function: it links information likely to have a single environmental cause for subsequent extra-striate processing.

Anderson discusses a neural network solution to the aperture problem in Chapter 10.

■ Spatio-temporal filtering

Beginning with the psychophysical results of Campbell and Robson (1968), and continuing with studies of the spatial and temporal frequency selectivity of simple and complex cells, there has emerged a picture of how images may be processed in the visual cortex. Let us look at spatial frequency in detail with a view to understanding its computational function in vision.

Spatial frequency filtering

■ Basis set for representing visual information

Psychophysics and physiology

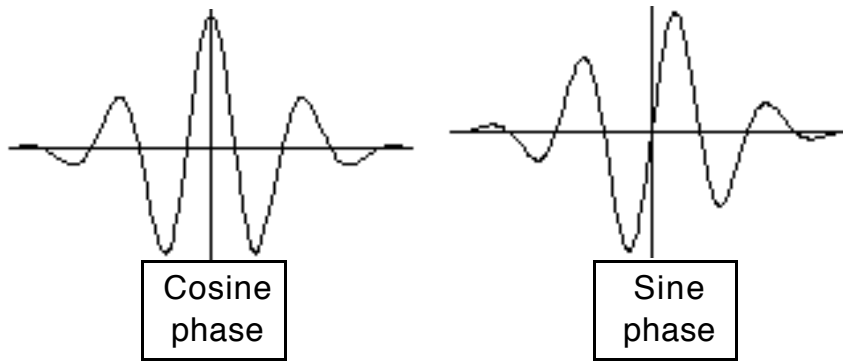
The results of masking, adaptation, and other psychophysical studies of spatial and orientation frequency selectivity in human vision are surprisingly consistent.

A cortical basis set for images

Both the psychophysical and neurophysiological data could be accounted for, in part, by assuming the visual system performed a quasi-Fourier analysis of the image. One possible model assumes that the visual system computes the coefficients (or spectrum) of an image with respect to the following basis set, called a Gabor set (Daugman, 1988):

$$\left\{ k e^{-\frac{(x^2 + y^2)}{2\sigma^2}} \cos(2\pi(f_x x + f_y y)), k e^{-\frac{(x^2 + y^2)}{2\sigma^2}} \sin(2\pi(f_x x + f_y y)) \right\}$$

The spectrum coefficients are represented by the firing rates of cells whose receptive field weights are represented by the above basis functions. In actuality, because as we saw earlier, simple cells behave more like linear filters followed by half-wave rectification, there should be two cells for each coefficient-- "on" and "off" cells). One difference between this basis set, and the Fourier basis set (i.e. the optical eigenfunctions) is that this set has a local spatial restriction because of the Gaussian envelope. A second difference, which has major implications for computation, is that the basis functions are, in general, not orthogonal. Graphs of these functions typically look like wave-packets:



Here we've plotted a one-dimensional slice through a sine, and cosine Gabor function. In two dimensions (with the standard deviation , and the x and y spatial frequencies equal to 1), we can visualize the receptive field weights as follows.

■ Visualizing the Gabor functions:

```
cgabor[x_,y_, fx_, fy_,s_] :=  
Exp[-(x^2 + y^2)/s^2] Cos[2 Pi(fx x + fy y)];  
sgabor[x_,y_, fx_, fy_, s_] :=  
Exp[-(x^2 + y^2)/s^2] Sin[2 Pi(fx x + fy y)];
```

■ Relation of horizontal and vertical frequencies to orientation and oriented frequency

$$\text{orientation} : \quad \theta = \tan^{-1} \left(\frac{f_y}{f_x} \right)$$

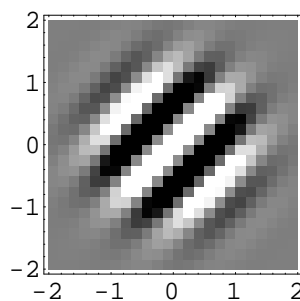
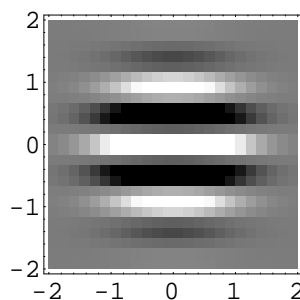
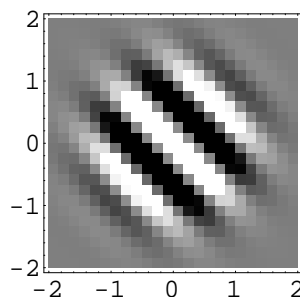
$$\text{frequency} : \quad f = \sqrt{f_x^2 + f_y^2}$$

■ Make a list of the orientations, and center frequencies of the basis set

```
vtheta = Table[i1 Pi/4, {i1,4}];  
vf = {.25, 1, 4, 8};
```


- Plot the elementary basis functions with the width, s , proportional to the reciprocal of spatial frequency. This maintains a constant bandwidth in octaves.

```
Table[DensityPlot[
  cgabor[x,y,vf[[i]] Cos[ vtheta[[j]] ],
    vf[[i]] Sin[ vtheta[[j]] ],
    1/vf[[i]] ], {x,-2,2}, {y,-2,2},
  PlotPoints->24, Mesh->False],
  {i, 2, 3}, {j, 1, 3}] // Short
```



\$Aborted

We've discretized the above continuous specification of the basis set. This leaves several free parameters. Most models of detection and masking get by with about no more than 6 spatial frequencies, about 12 orientations (specified by the ratio of horizontal and vertical spatial frequencies), and two phases (cosine and sine) at each retinal location. A subset of neurons representing a particular spatial frequency bandwidth makes up a spatial frequency channel. (Although there is neurophysiological evidence for pairs of V1 neurons having receptive fields with 90 deg phase shifted relative to each other,

there is evidence against absolute phase--i.e. there is not a predominance of edge or bar type receptive fields. See Field and Tolhurst). One parameter still left unspecified is the standard deviation or spread of the Gaussian envelope. If large, this basis set approaches that of regular and global Fourier analysis. The psychophysical data suggest that the standard deviation be such that the Gaussian envelope is about one cycle (at the $1/e$ point) of the sine wave. One cycle corresponds to about 1.5 octaves spatial frequency bandwidth.

Why a local Gabor-function representation of visual information?

Why would the visual system have such a representation? We have two types of explanations. One is that encoding over multiple spatial scales is important for subsequent processing that may involve edge detection, or stereoscopic matching, and so forth. Analogous pyramid schemes have been developed for computer vision. (See Adelson, E. H., Simoncelli, E., & Hingorani, R., 1987). The second explanation is in terms of economical encoding which we pick up on briefly below. (An interesting historical note is that many of early attempts to understand visual cortical receptive fields in terms of filters localized in space and spatial frequency were forerunners of modern wavelet theory.)

Edge Detection by Neural Networks

Because of the orientation selectivity of cortical cells, they have been sometimes interpreted as edge detectors. It is easy to see how a sine-phase Gabor function filter (1 cycle wide) would respond vigorously to an edge oriented with its receptive field. This type of receptive field behaves as a 2D smoothing operator followed by a first order directional derivative. If one took the outputs of two such cells, one vertical and one horizontal, the sum of the squares of their outputs would approximate the squared magnitude of the gradient of the smoothed image:

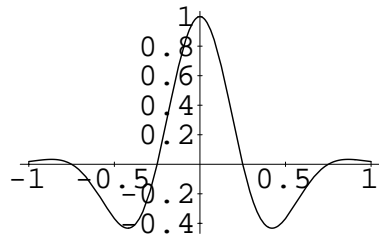
$$|\nabla G(x, y) * L(x, y)|^2$$

Morrone and Burr edge detector: combining sine and cosine phase filters

■ Define the filters

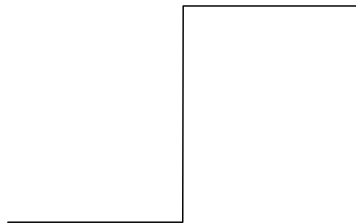
```
cosinefilter[x_,sigma_,f_] :=
    Exp[-(x/sigma)^2] Cos[2 Pi f x]
sinefilter[x_,sigma_,f_] :=
    Exp[-(x/sigma)^2] Sin[2 Pi f x]
```

```
Plot[cosinefilter[x,.5,1],{x,-1,1}]
```



■ Define the input stimulus: an ideal edge

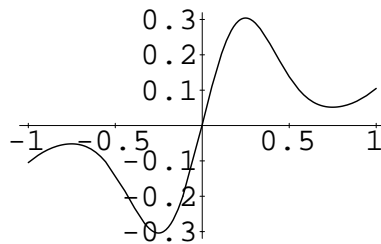
```
Plot[Sign[x],{x,-1,1},Axes->None]
```



■ Calculate the response of a bank of cosine filters to the edge

```
cr[x_] :=  
NIntegrate[cosinefilter[(x - x1),.5,1] Sign[x1],{x1,-1.5,1.5}]
```

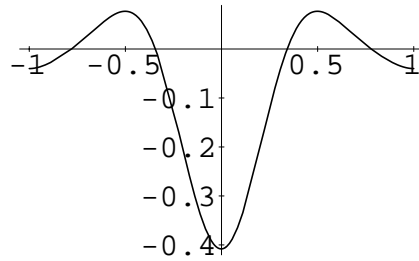
```
Plot[cr[x],{x,-1,1},PlotPoints->10];
```



■ Calculate the response of a bank of sine filters to the edge

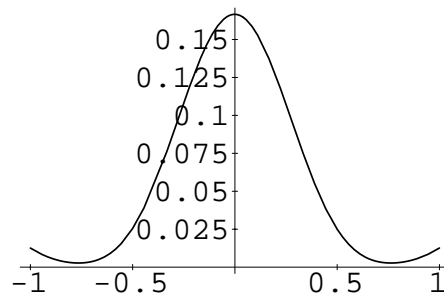
```
sr[x_] :=  
NIntegrate[sinefilter[(x - x1),.5,1] Sign[x1],{x1,-1.5,1.5}]
```

```
Plot[sr[x], {x, -1, 1}, PlotPoints -> 10];
```



■ Combine the outputs from the two banks by squaring and adding:

```
Plot[cr[x]^2 + sr[x]^2, {x, -1, 1}, PlotPoints -> 10]
```



The peak of the squared sum is at the location of the edge.

Note the basic structure of the neural network for edge detection:

1. A linear operations using weights from a cosine function, followed by a squaring operation

(how does this squaring relate to the sigmoidal non-linearities we have been using?)

2. A linear operation using weights from a sine function, followed by a squaring operation

3. A linear sum of the outputs

Perceived edges don't necessarily correspond to a sharp ideal edge, but can be blurry and noisy.

Morrone and Burr went on to show that one could do the same operation with different sizes of filters (i.e. different values of **sigma**), and each time the peak of the above operation for an ideal edge occurs at the edge transition. But even for blurry edges, the larger scale filters will still find a point in the transition region. Thus by adding up a whole set of neural outputs

over a range of scales, one could detect an edge. Another way of viewing this network is one that detects

phase coherence. Fourier theory shows that a step function can be built up of sine-waves of various frequencies whose zero crossings all line up with (say positive slope) at the edge transition.

Although one can build edge detectors from oriented filters, simple cells cannot uniquely signal the presence of an edge for several reasons. One is that their response is a function of many different parameters. A low contrast bar at an optimal orientation will produce the same response as a bar of higher contrast at a non-optimal orientation. There is a similar trade-off with other parameters such as spatial frequency and temporal frequency. In order to make explicit the location of an edge from the responses of a population of cells, one would have to compute something like the "center-of-mass" over the population, where response rate takes the place of mass. Another problem is that edge detection has to take into account a range of spatial scales. The cortical basis set does encompass a range of spatial scales, and in fact may be "self-similar" across these scales. See Koenderink (1990) for a recent theoretical discussion of "ideal" receptive field properties from the point of view of basis elements.

Economical representations by neurons in primary cortex

We might expect something like Fourier analysis of the image to result in efficient coding because of the close relationship between Fourier rotations and Principal Components Analysis (e.g. Appendix A, Andrews, 1983). Fourier coefficients for natural images tend to be uncorrelated. Some work has been completed toward a functional explanation for the orientation and spatial frequency tuning properties of cortical receptive fields based on the statistics of natural images (Field, 1994), but the story is far from complete. Recently, Barlow has argued that a decorrelated representation of sensory information is important for efficient learning (Barlow, 1990).

There has been recent rapid progress in the relationship between self-organizing models of visual cortex, and efficient coding of image information. For more on this, see: Linsker, R. (1990) and Barlow, H. B., & Foldiak, P. (1989). Linsker's computational studies show, for example, that orientation tuning, and band-pass properties of simple cells can emerge as a consequence of maximum information transfer (in terms of variance) given the constraint that the inputs are already band-pass, and the receptive field connectivity is a priori limited.

Extra-striate cortex

Figure 10.9 in the Anderson book shows that cells in the visual cortex send their visual information to an incredibly complex, and yet structured collection of extra-striate areas. Any hypothesized function of striate cortex must eventually take into account what the information is to be used for.

Bibliography

- Adelson, E. H., Simoncelli, E., & Hingorani, R. (1987). Orthogonal Pyramid Transforms for Image Coding. Proc. SPIE - Visual Communication & Image Proc. II, Cambridge, MA.
- Barlow, H. B., & Foldiak, P. (1989). Adaptation and decorrelation in the cortex. In C. Miall, R. M. Durban, & G. J. Mitchison (Ed.), *The Computing Neuron* Addison-Wesley.
- Barlow, H. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research*, 30(11), 1561-1572.
- Campbell, F. W., & Robson, J. R. (1968). Application of Fourier Analysis to the Visibility of Gratings. 197, 551-566.
- Crick, F. (1984). Function of the Thalamic Reticular Complex: The Searchlight Hypothesis. 81, 4586-4590.
- Daugman, J. G. (1988). An information-theoretic view of analog representation in striate cortex. In *Computational Neuroscience* Cambridge, Massachusetts: M.I.T. Press.
- DeValois, R., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545-559).
- Dobbins, A., Zucker, S. W., & Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329(6138), 438-441.
- Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, 6, 559-601.
- Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual cortex. In M. & M. Landy A. (Ed.), *Computational Models of Visual Processing* (pp. 119-133). Cambridge, Massachusetts: M.I.T. Press.
- Hubel, D. H., & Wiesel, T. N. (1968,). Receptive Fields and Functional Architecture of Monkey Striate Cortex. *J. Physiol.*, London.pp. 215-243.
- Koenderink, J. J., & van Doorn, A. J. (1990). Receptive field families. *Biol. Cybern.*, 63, 291-297.
- Linsker, R. (1990). Perceptual neural organization: some approaches based on network models and information theory. *Annual Review of Neuroscience*, 13, 257-281.
- Livingstone, M. S., & Hubel, D. H. (1984). Anatomy and Physiology of a Color System in the Primate Visual Cortex. 4(1), 309-356;
- Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical Evidence for Separate Channels for the Perception of Form, Color, Movement and Depth. *The Journal of Neuroscience*, 7(11), 3416-3468).
- Poggio, G., F., & Poggio, T. ,1984. The Analysis of Stereopsis. *Annual Review of Neuroscience*, 7, 379-412).
- Poggio, T. (1984). Vision by Man and Machine. *Scientific American*, 250, 106-115.
- Poggio, G. F., & Talbot, W. H. (1981). Mechanisms of Static and Dynamic Stereopsis in Foveal Cortex of the Rhesus Monkey. 315, 469-492.
- Sillito, A. M., Jones, H. E., Gerstein, G. L., & West, D. C. (1994). Feature-lined synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, 369, N. 9, 479-482.

