CS294-6 (Fall 2004) Recognizing People, Objects and Actions Lecture: January 27, 2004

Human Visual System

Lecturer: Jitendra Malik *Scribe: Ryan White*

(Slide: layout of the brain) Facts about the brain:

- 10 billion cells
- 4 lobes
- 1.5 square feet

Vision processing starts in the striate cortex in the occipital lobe.

Of the 4 lobes, the parietal, temporal and occipital lobes are involved in vision. The fontal lobe tends to be higher level thought.

There are two main pathways for vision:

- The VENTRAL pathway: used for object recognition. Precedes from occipital to temporal lobe.
- The DORSAL pathway: vision by function. Precedes from the occipital to pariental lobe.

There are roughly 10^4 connections per neuron in the brain (remember that there are roughly 10^{10} neurons). Each neuron has inputs, some processing and then information travels through axon.

Important terms:

- AXON - the long part of the neuron that conducts the signal
- SYNAPSE - the method of communication between neurons
- ACTION POTENTIAL - electrical signal transmission, sent on axon
- GRADED POTENTIAL - change in potential of neuron caused by absorption of neurotransmitters

Signal transmission on the axon: resting potential is approximately -70 mV. Graded potential drives up voltage, absorption of sodium ions from surrounding cause a self-sustaining transmission of 'spike'. Voltage of axon reaches +40mV. This spike in voltage moves down the axon. The duration of the spike tends to be on the order of 1 millisecond. The 'spike rate' indicates something about activity of the neuron.

SYNAPSE: When the spike travelling down the axon reaches a junction, it causes the release of NEUROTRANSMITTERS, which are absorbed by the next neuron in the chain. Neurotransmitters are summed up over all of the inputs (in the form of graded potentials). Sum of many graded potentials causes self sustained chain reaction. Neurotransmitters can have a positive (excitatory) or negative (inhibitory) influence on the graded potential. example excitatory: -70mV to -50mV; example inhibitory: -70 mV to -80 mV. Different neurotransmitters for inhibitory and excitatory connections.

The brain is a relatively slow speed system: since spikes are on the order of a millisecond, the max rate is on the order of 300 Hz. Note that these aren't binary devices. In the following spike train, information is transmitted in the rate of spiking.

The traditional view is that spikes per second is the only important factor in transmitting information. However, new thoughts suggest that the timing between the spikes is important, since human response to stimuli is so quick.

VISUAL SYSTEM

- 2 eyes
- OPTIC NERVE: 1.5 million fibers per eye (each fiber is the axon from a neuron)
- 125 million rods
- 6 million cones

Note the approximately 100:1 compression from the number of receptors to the number of fibers in the optic nerve. The transmission rate through the optic nerve is quick - the fastest axons in the body tranmit at roughly 60 meters per second.

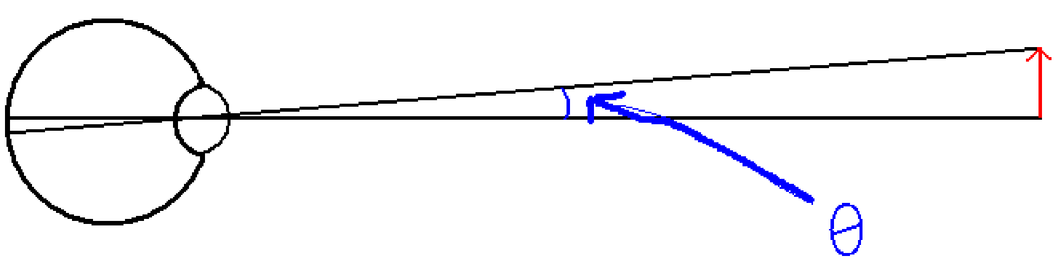
BLIND SPOT: where the optic nerve leaves the eye.

Lateral Geniculate Nucleus (LGN)

- transmission repeater station for the optical information
- partitions the fibers based on scene location
- send information pertaining to left portion of scene to right portion of brain and visa-versa.

Trnamission path: eye to LGN is a one way transmission. However, there are connections from the LGN to the cortex AND from the cortex to the LGN. Understanding of exact role of LGN still in progress.

Sizing objects in terms of the eye: compute the angle shown in the picture below.



Rough 'rule of thumb': the thumb at arms length is approximately 2 degrees

Monocular visual field is 160^o (w) x 135^o (h) Binocular visual field is 200^o (w) x 135^o (h)

Note that the visual field is much larger than most cameras. However, processing is not uniform across the visual field: 25% of cortex is devoted to the central five degrees of the field of view.

The eye can deal with a large amount of variation in illumination: apprixímately 14 orders of magnitude. At the low end of the scale, a rod can respond to a single photon. Typically it takes at least 6 photons for the brain to distinguish the signal from noise.

There are 3 kinds of cones for color vision.

Within the eye, there are 5 layers of processing organized as sheets:

- 1st sheet: photoreceptors - rods and cones that receive the incident light
- 2nd sheet: horizontal cells - inhibitory input to bipolar cells, responsible for combining information from surrounding photoreceptors
- 3rd sheet: bipolar cells - take exhitatory inputs from photoreceptors and inhibitory input from bipolar cells
- 4th sheet: amacrine cells
- 5th sheet: ganglion cells - the axons from the ganglion cells comprise the optic nerve

Normally the photoreceptors are in the back and the other sheets are towards the front - technically in front of the photoreceptors. However, in the fovea, these other cells are pushed aside.

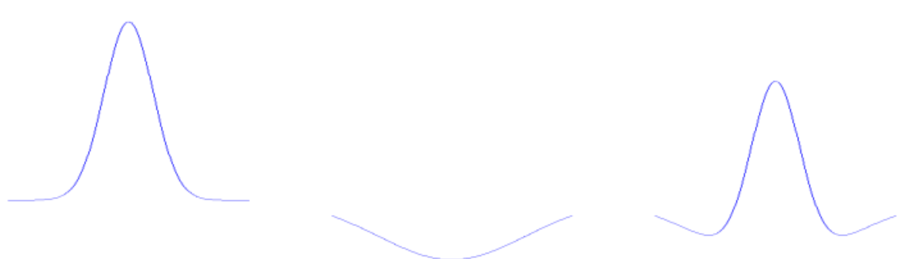
One can model the effect of the processing in the retina as the difference of gaussians. We will use the following definition of a gaussian:

$$G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \tag{1}$$

Now, the difference of gaussians is:

$$R(x) = G_{\sigma_1}(x) - G_{\sigma_2}(x) \tag{2}$$

This can be seen graphically as (from left to right $G_{\sigma_1}(x)$, $-G_{\sigma_2}(x)$, $R(x)$):



Convolution: $I_2 = I_1 * R$, where I_2 are the bipolar cell responses and I_1 is the original image. Note that convolution is the following integral:

$$\int\int I_1(x-u, y-v)R(u,v)dudv \tag{3}$$

The eye does not keep the absolute luminance level. It only keeps relative light values. In the fovea, there is almost no compression, but in the periphery, compression is greater than 100:1.

The difference of gaussians is called "ON center OFF surround"

The neurons are not capable of using negative numbers. So, to implement negative numbers, there is twice as much hardware - a positive number signal and a negative number signal.

RECEPTIVE FIELD - What part of the visual field is the stimulus in order to activate this neuron to respond. Also, the receptive field includes the type, size and shape of stimulus to cause maximal response. Can also think of the receptive field as a convolution kernel.

(slide: x-axis is frequency, y-axis is contrast) Contrast Sensitivity Function: the human visual system has varying sensistivity at different levels of spatial frequency. The 'best case resolution' is 60 cycles per degree ($\lambda = 1$ min)

(slide: receptor density vs eccentricity) Most of object recognition is done in the fovea. The fovea has very few rods. Eccentricity is the angle from the optical axis

Inside the cortex there are oriented bar filters. Some respond to light bars, others respond to dark bars. Some respond to edges. Variables include orientation, position, polarity and size. There are approximately 8 orientations.

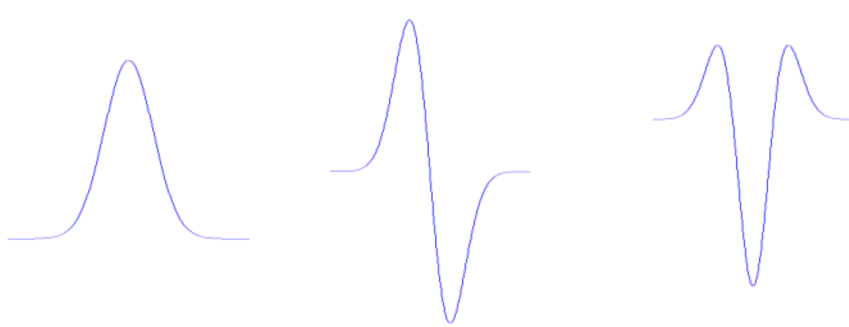
- Kuffler: 1950s - ON center OFF surround
- Hubel and Wiesel: 1962 - oriented cells. different bars respond to different orientations

The cortex has 10^3 factor more neurons than the OPTIC NERVE.

Models for the oriented bar filters:

- elongated directional gaussian derivatives
- 2nd derivative and hilbert transform
- l_1 normalized for scale invariance
- 6 orientations, 3 scales
- zero mean

From left to right: $G_\sigma(x)$,$G'_\sigma(x)$, $G''_\sigma(x)$:



To get a sample filter, can combine: $G_{\sigma_1}(x)G_{\sigma_2}(y)$

To get a rotated filter: $G_{\sigma_1}(u)G_{\sigma_2}(v)$, where:

$$u = x\cos(\theta) - y\sin(\theta) \tag{4}$$
$$v = x\sin(\theta) + y\cos(\theta) \tag{5}$$

Typically, $\frac{\sigma_2}{\sigma_1} \approx 3$

COMPLEX CELLS: respond to bars, but are locally position insensitive. Orientation Energy - know that bar is there, but position blur:

$$OE = (I * f_{odd})^2 + (I * f_{even})^2 \tag{6}$$