Lecture 4: Introduction to Vision.

Reading Assignments:

Chapters 2 and 3 of textbook.

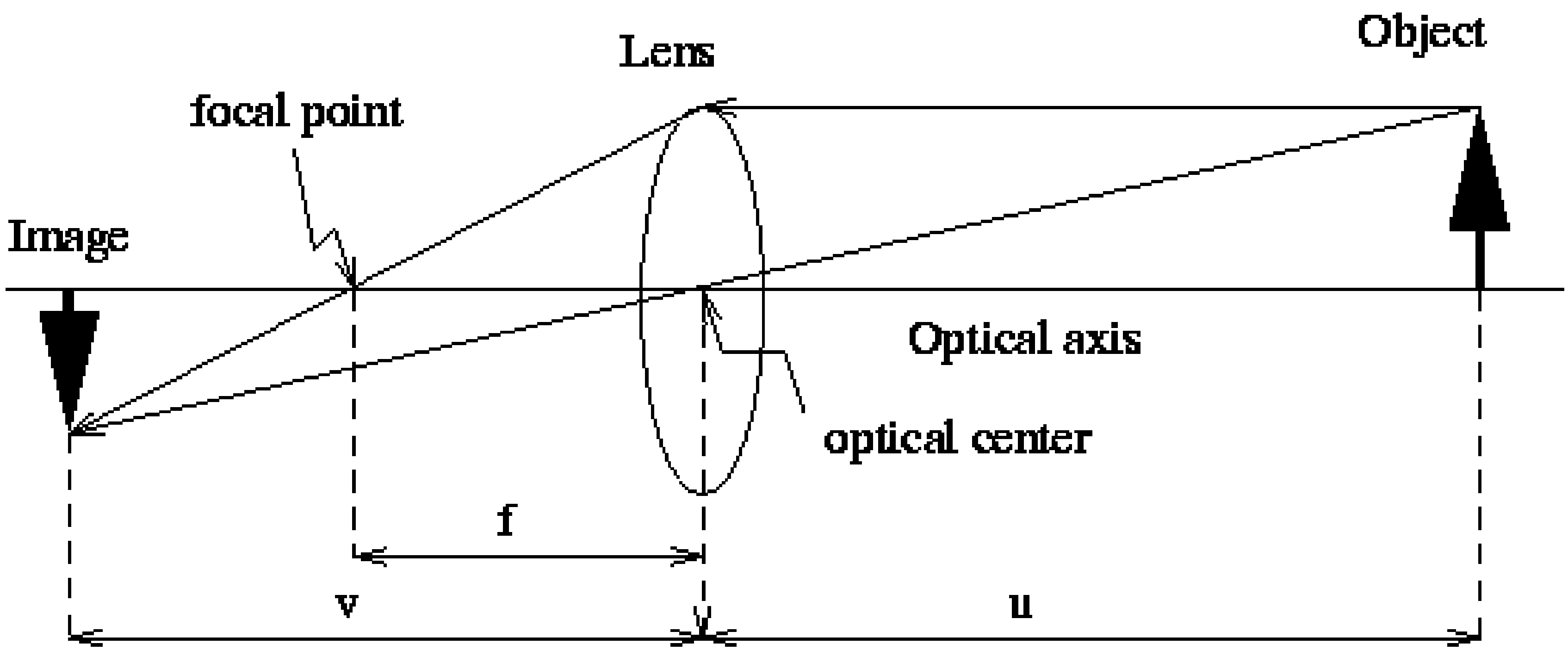
Today's lecture

- The challenges:
  - Optics and image formation
  - Sampling and image representation
  - Theoretical limits

- The biological approach:
  - Organization of the primate retina
  - Trading accuracy for coverage: moving eyes

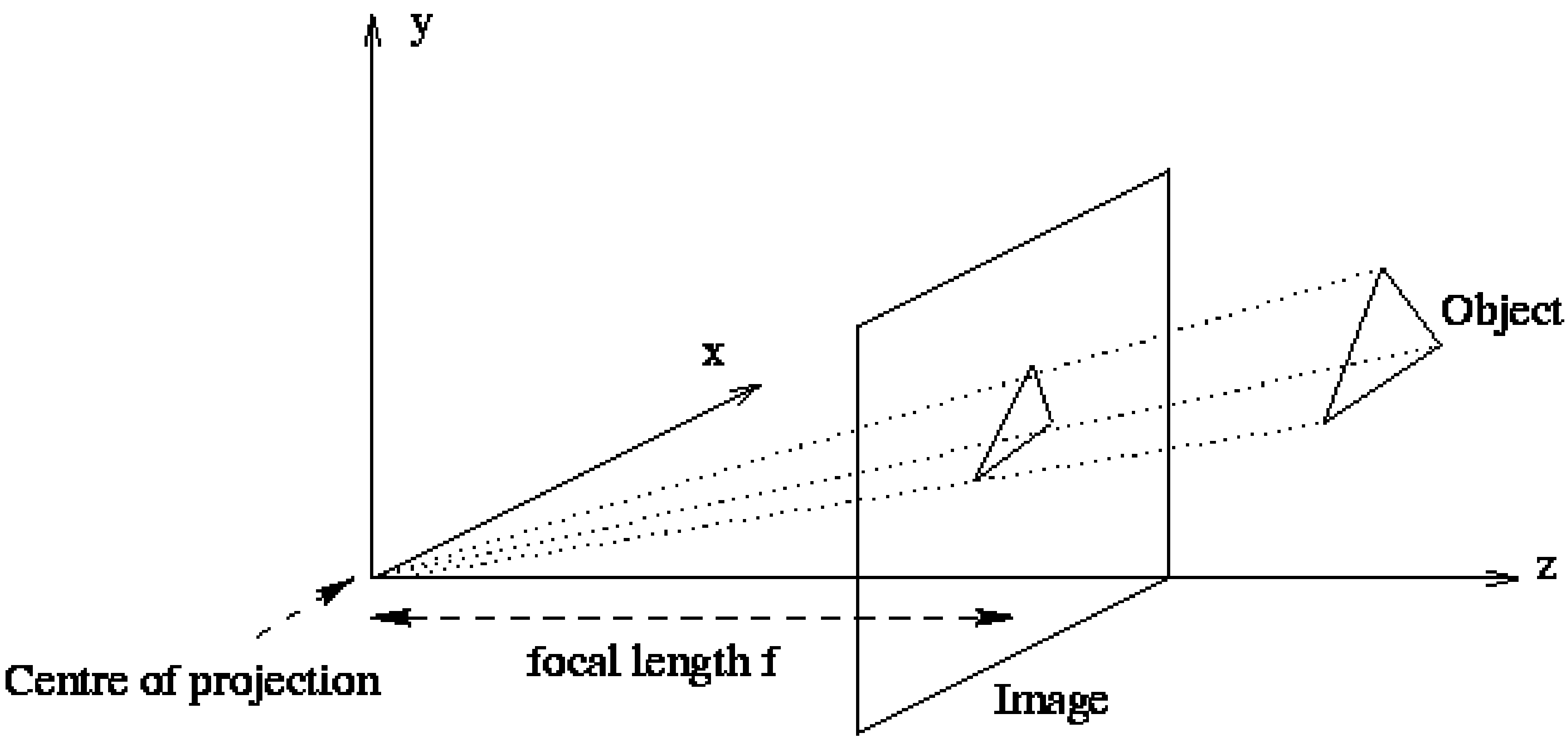
- The engineering approach:
  - Arrays of photosensitive sensors
  - On-board processing and VLSI sensors
  - Trading accuracy for coverage: multiple moving cameras

Projection



Object
Lens
focal point
Image
Optical axis
optical center
f
v
u

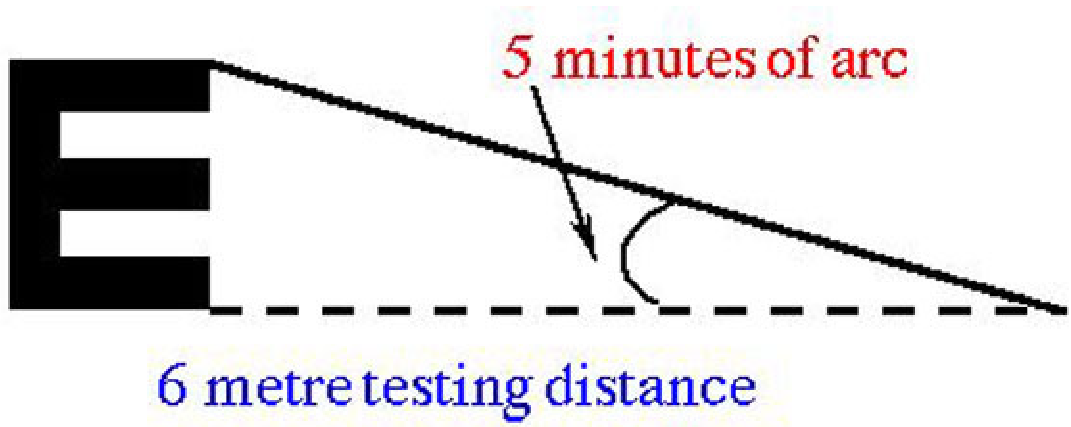
Projection



Convention: Visual Angle

Rather than reporting two numbers (size of object and distance to observer), we will combine both into a single number:

visual angle



e.g., the moon: about 0.5deg visual angle

your thumb nail at arm's length: about 1.5deg visual angle

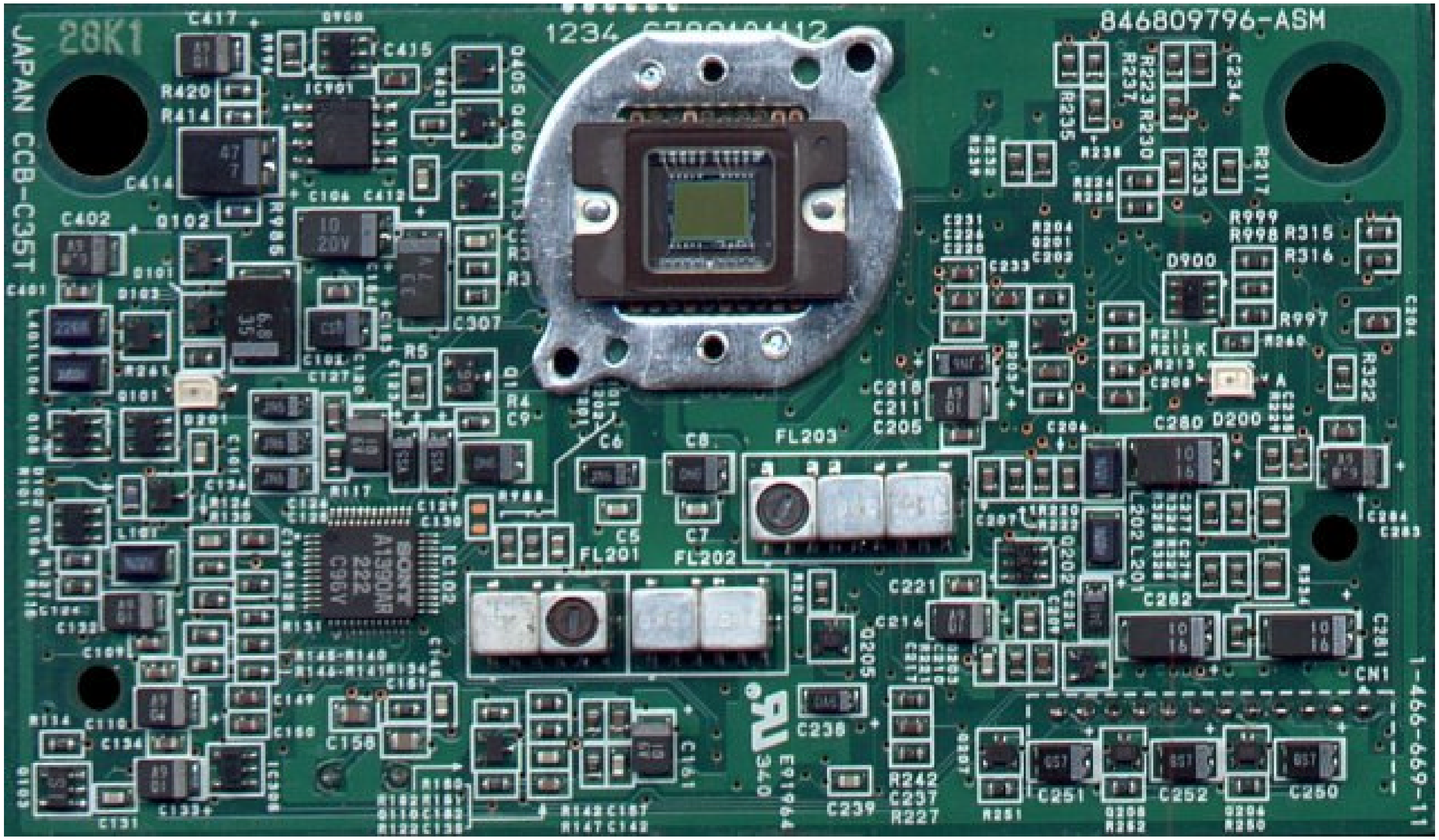
1deg visual angle: 0.3mm on retina

Charge-Coupled Devices

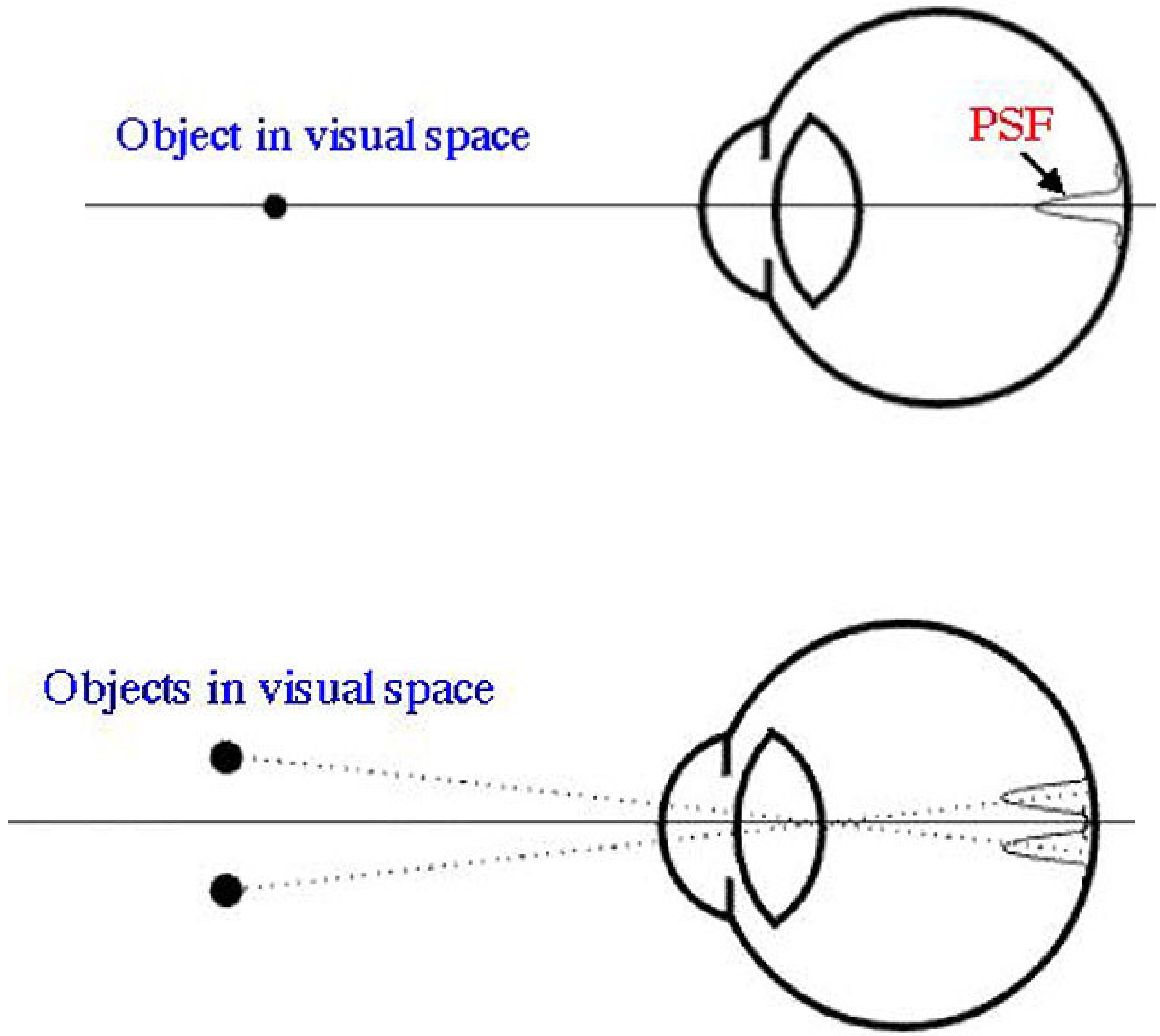
Uniform array of sensors

Very little on-board processing

Very inexpensive



Optics limitations: acuity



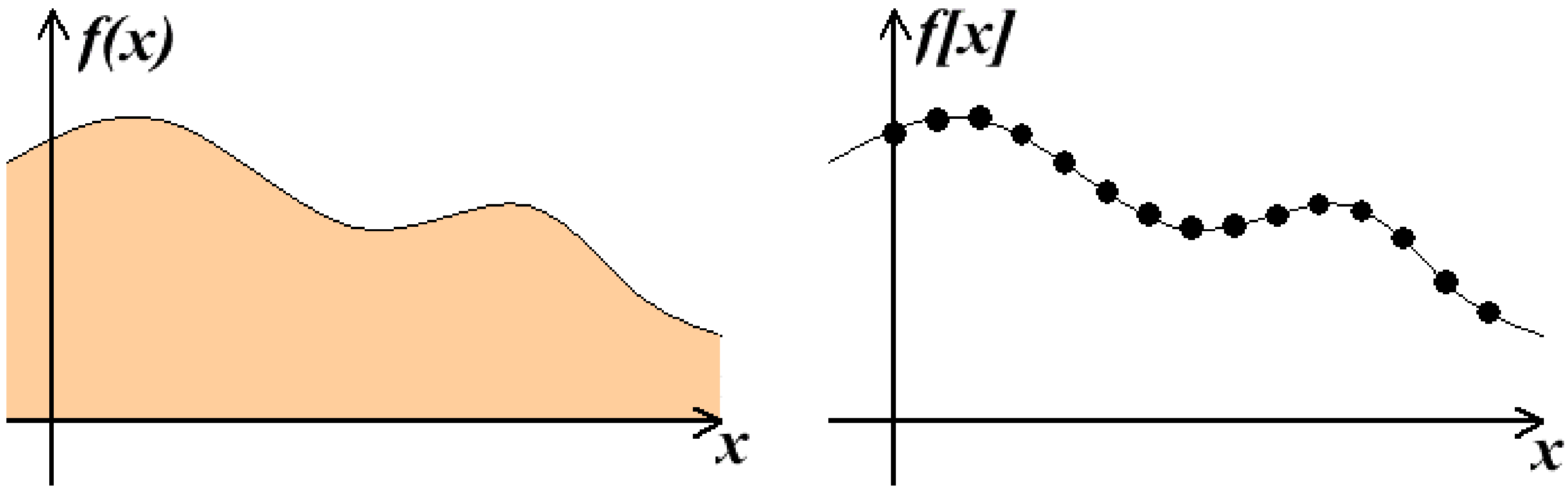
Sampling

We think of most things in the real world as *continuous*,
yet, everything in a computer is discrete

The process of mapping a continuous *function* to a discrete one is called *sampling*

The process of mapping a continuous *variable* to a discrete one is called *quantization*

When we represent or render an image using a computer we must both sample and quantize



Sampling

Sampling Grid

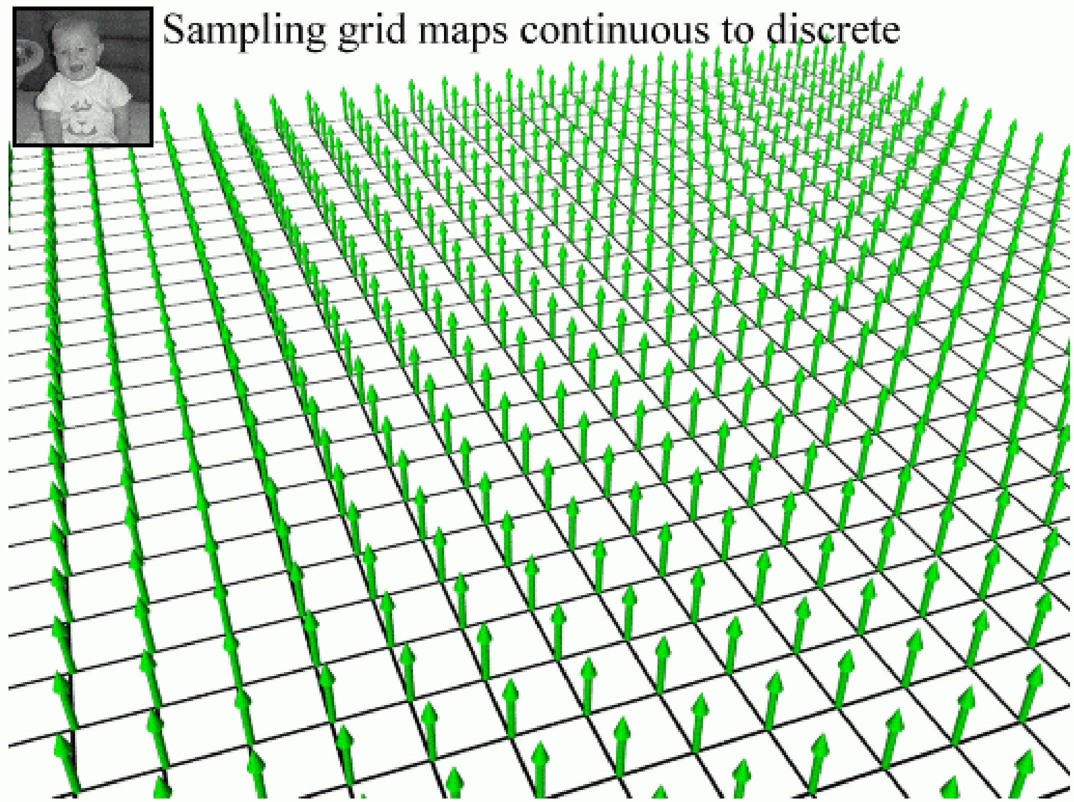
The most common (but not the only) way to generate the table values necessary to represent our function is to multiply the function by a *sampling grid*. A sampling grid is composed of periodically spaced Kronecker delta functions.

The definiton of the 2-D Kronecker delta is:

$$\delta(x,y) = \begin{cases} 1, & (x,y) = (0,0) \\ 0, & \text{otherwise} \end{cases}$$

And a 2-D sampling grid:

$$\sum_{j=0}^{h-1} \sum_{i=0}^{w-1} \delta(u-i, v-j)$$

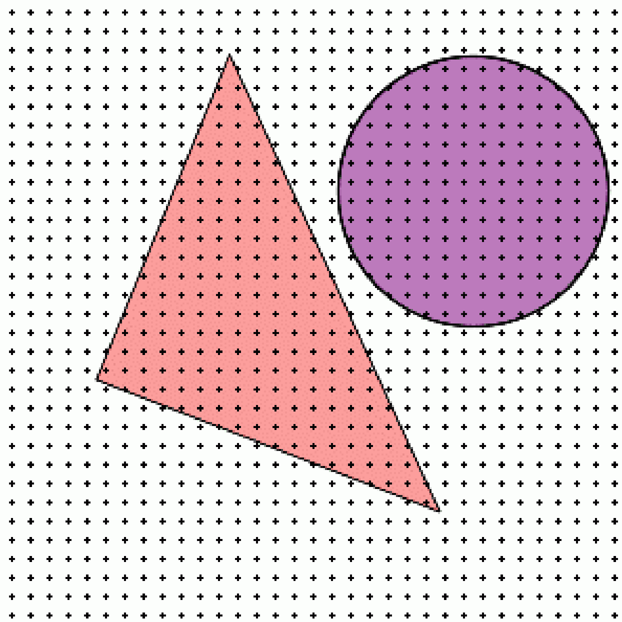


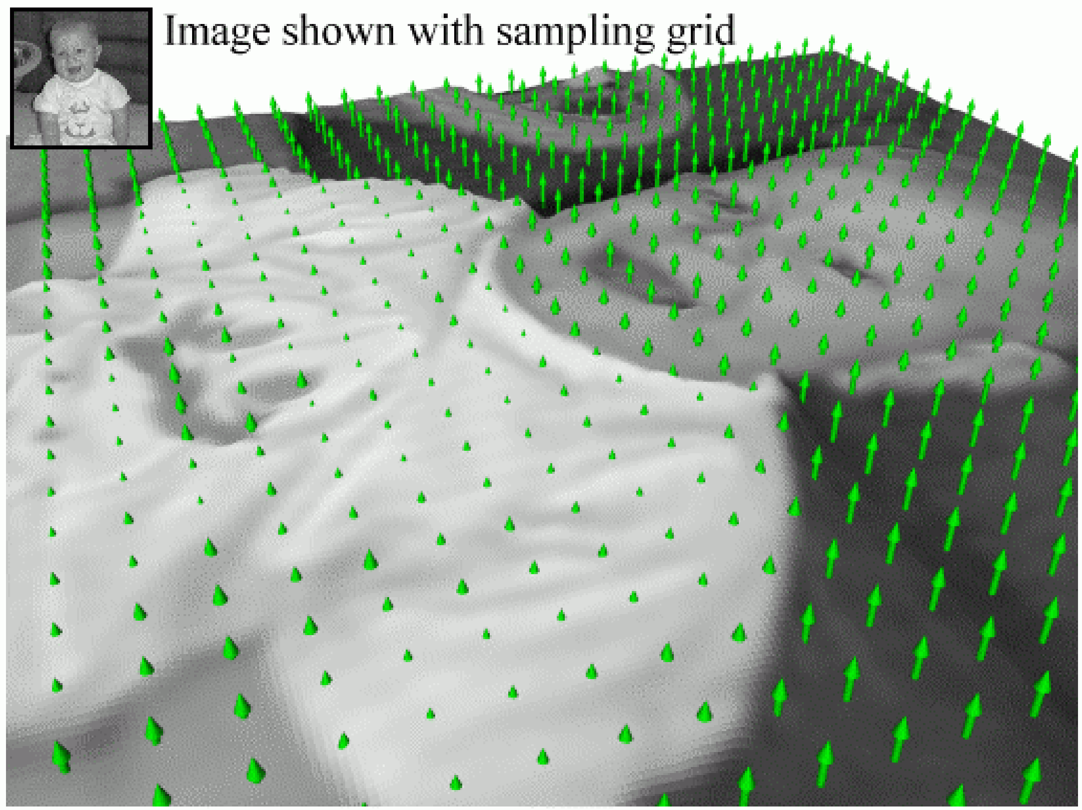
Sampling

Sampling an Image

When a continuous image is multiplied by a sampling grid a discrete set of points are generated. These points are called samples. These samples are pixels. We store them in memory as arrays of numbers representing the intensity of the underlying function.

The same analysis can be applied to geometric objects:





The Big Question

How densely must we sample an image in order to capture its essence?

Since our sampling grid is *periodic* we can appeal to Fourier analysis for an answer. Fourier analysis states that all periodic signals can be represented as a summation of sinusiodal waves. Thus every image function that we understand as a height field in the *spatial domain,* has a dual representaion in the *frequency domain.*

We can transform signals from one domain to the other using the Fourier transform.

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, e^{-i2\pi(ux+vy)} dxdy$$

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)\, e^{i2\pi(ux+vy)} dudv$$

Convolution & Fourier Transforms

In order to simplify our analysis we will consider 1–D signals for the moment. It will be straightforward to extend the result to 2–D.

Some operations that are diffcult to compute in the spatial domain are simplified when the function is tranformed to its dual representation in the frequency domain. One such function is *convolution*.

Convolution describes how a system with impulse response, *h(x)*, reacts to a signal, *f(x)*.

$$f(x) * h(x) = \int_{-\infty}^{\infty} f(\lambda) h(x-\lambda) d\lambda$$

This integral evaluation is equivalent to multiplication in the frequency domain

$$f(x) * h(x) \to F(u) H(u)$$

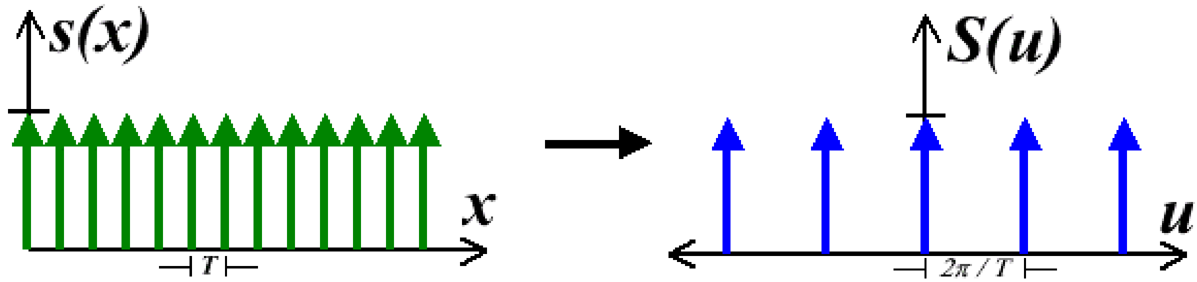
The converse is also true

$$F(u) * H(u) \to f(x) h(x)$$

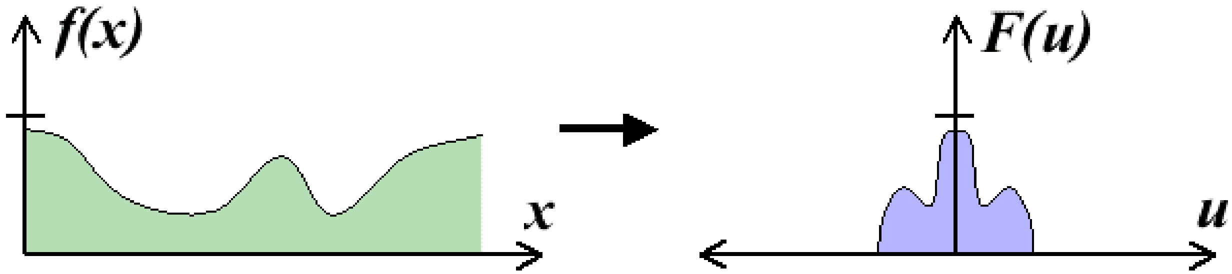
Sampling in the frequency domain

Image sampling was defined as multiplying a periodic series of delta functions by the continuous image. This is the same as convolution in the frequency domain.

Consider the sampling grid:

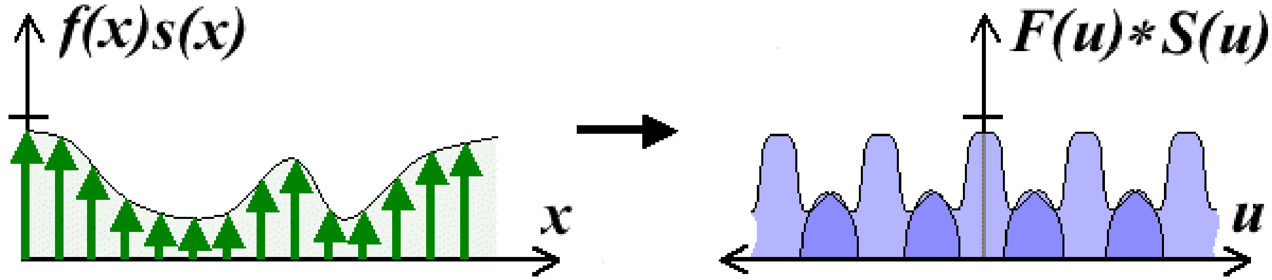


And the function being sampled

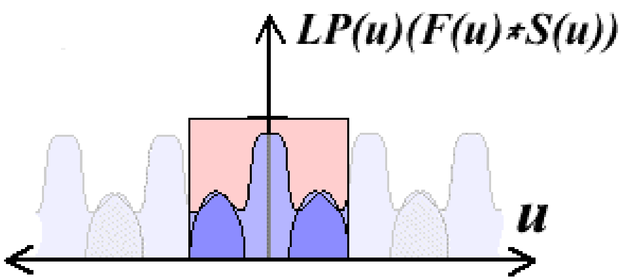


Reconstruction

This amounts to accumulating copies of the function's spectrum centered at the delta functions of the sampling grid.



Remember the goal of a sampled represention is to faithfully represent the underlying function. Ideally we would apply a *low–pass filter* to our sampled representation to *reconstruct* our original function. We will call this processing a *reconstruction filter*.

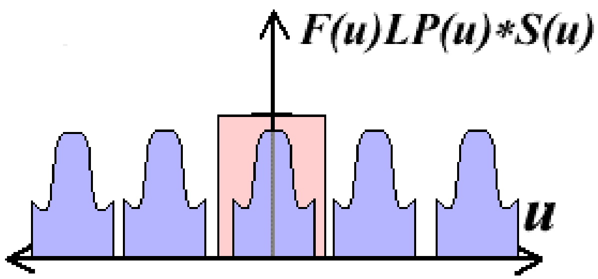


In this example we mixed together copies of our function (as seen in the darker overlap regions). In this case subsequent processing does not allow us to separate out a representative copy of our function.

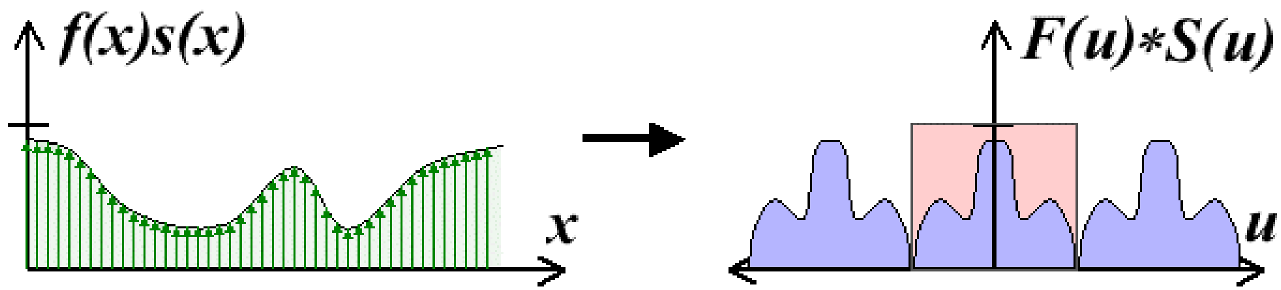
Aliasing

This mixing of spectrums is called *aliasing*. It has the effect of introducing high–frequencies into our original function. There are two ways of dealing with aliasing.

The first is to low pass filter our signal before we sample it.



The second is to increase the sampling frequency



Sampling Theorem

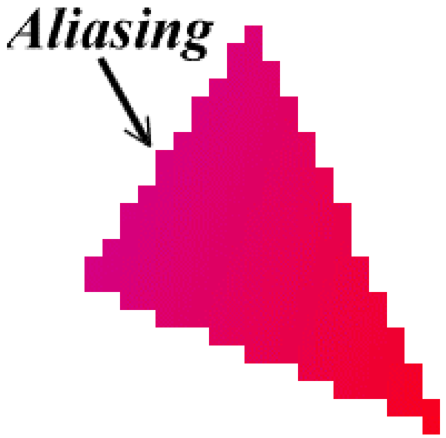
In order to have any hope of accurately reconstructing a function from a periodically sampled version of it, two conditions must be satisfied:

1. The function must be bandlimited.
2. The sampling frequency, f_s, must be at least twice the maximum frequency, f_{max}, of the function.

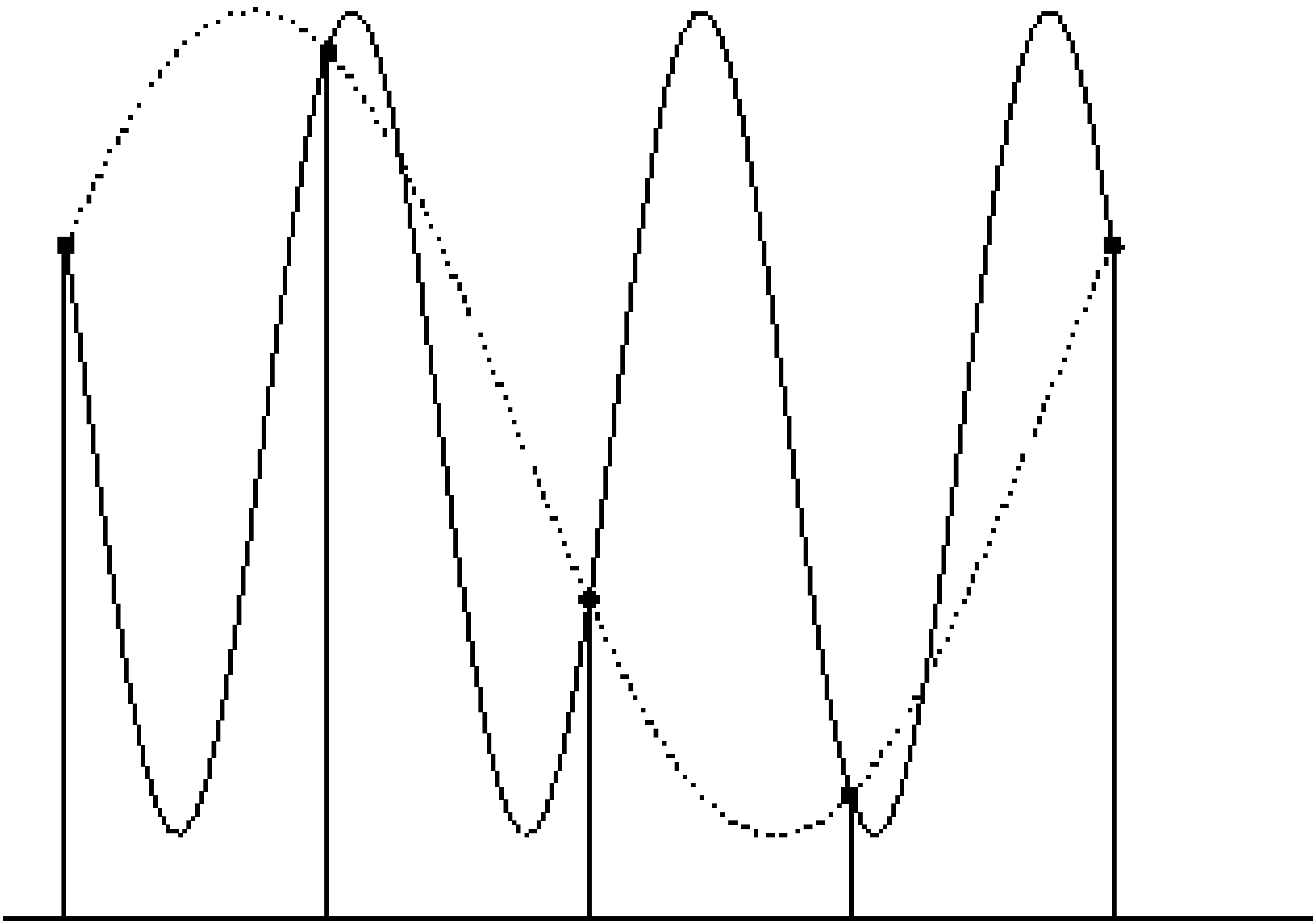
Satisfying these conditions will eliminate aliasing.

In practice:

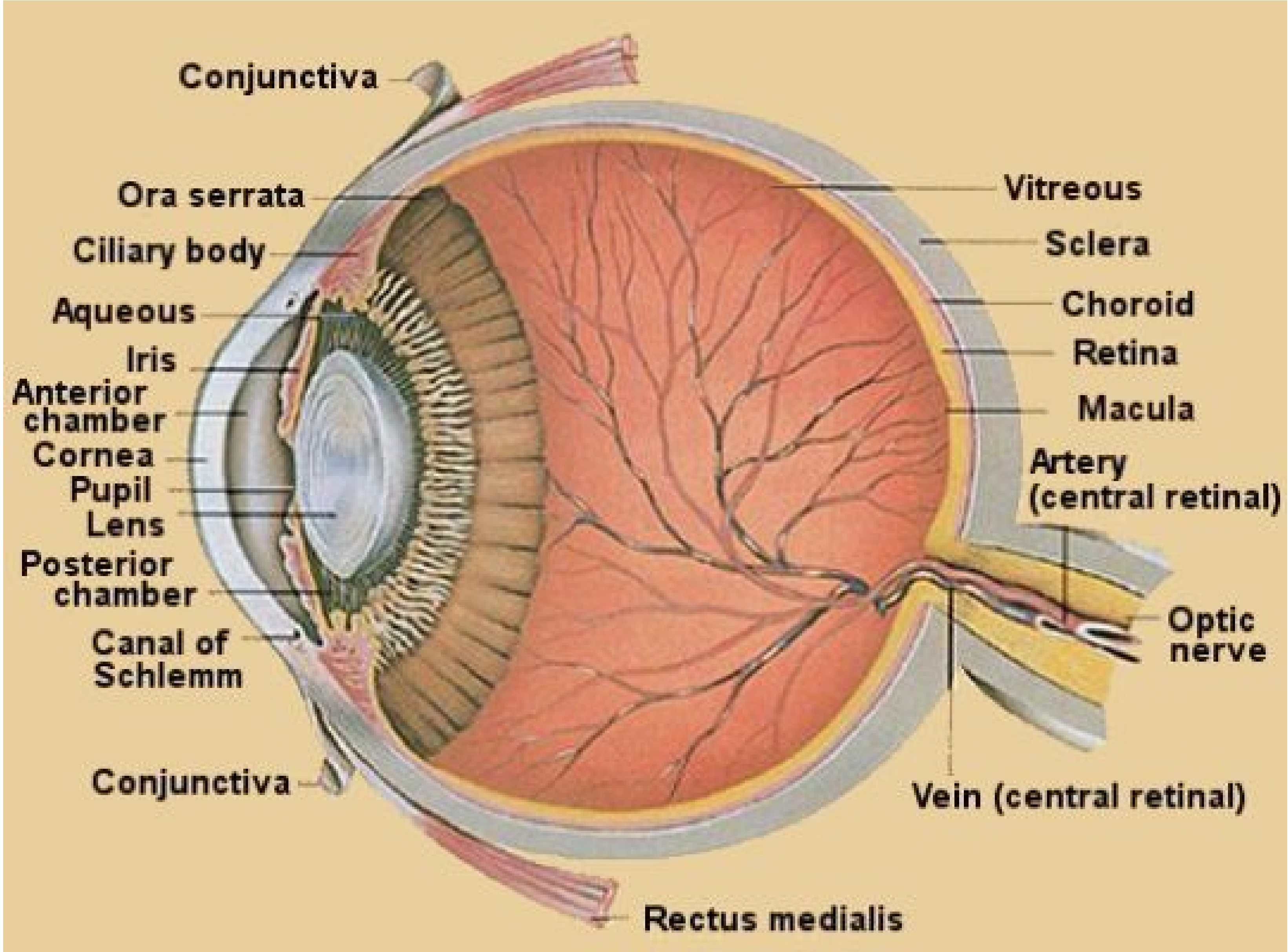
- "Jaggies" are aliasing
- Both of the techniques discussed are used
    1. Super-sampling (more samples than pixels)
    2. Low-pass prefiltering (averaging of super-samples)



Aliasing



Eye Anatomy



Conjunctiva
Ora serrata
Ciliary body
Aqueous
Iris
Anterior chamber
Cornea
Pupil
Lens
Posterior chamber
Canal of Schlemm
Conjunctiva
Vitreous
Sclera
Choroid
Retina
Macula
Artery (central retinal)
Optic nerve
Vein (central retinal)
Rectus medialis

Visual Pathways

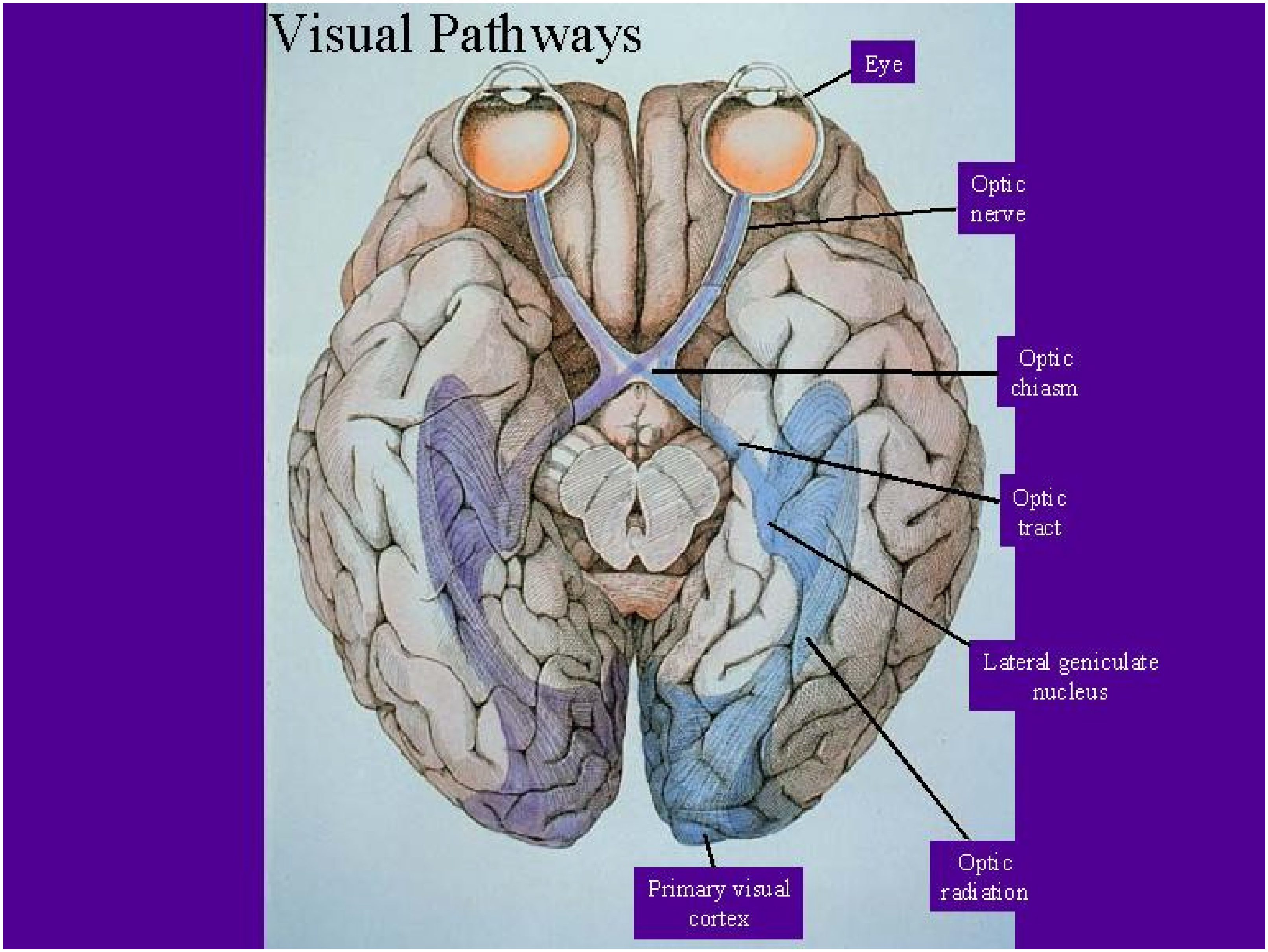
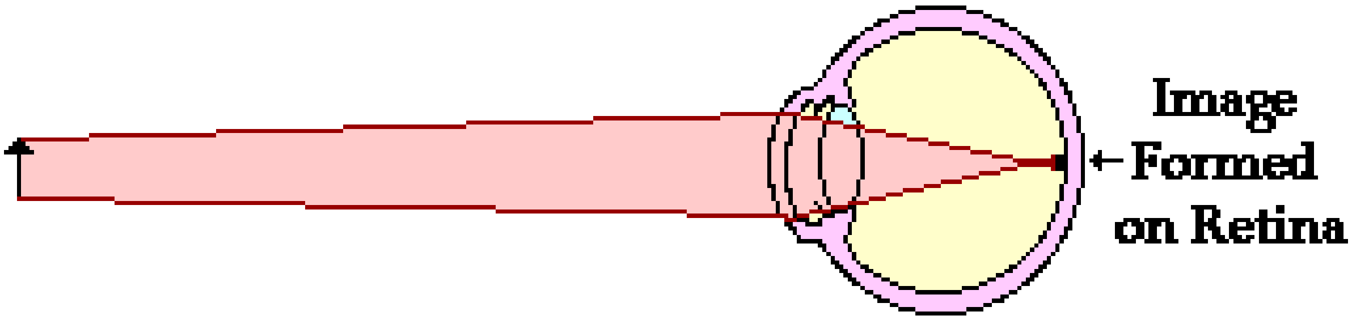
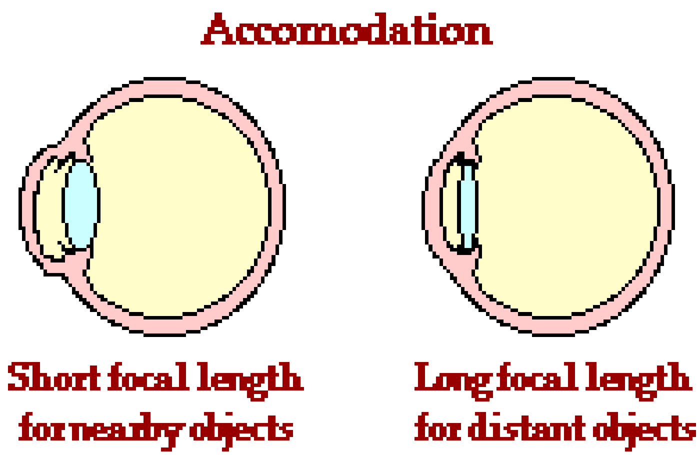


Image Formation



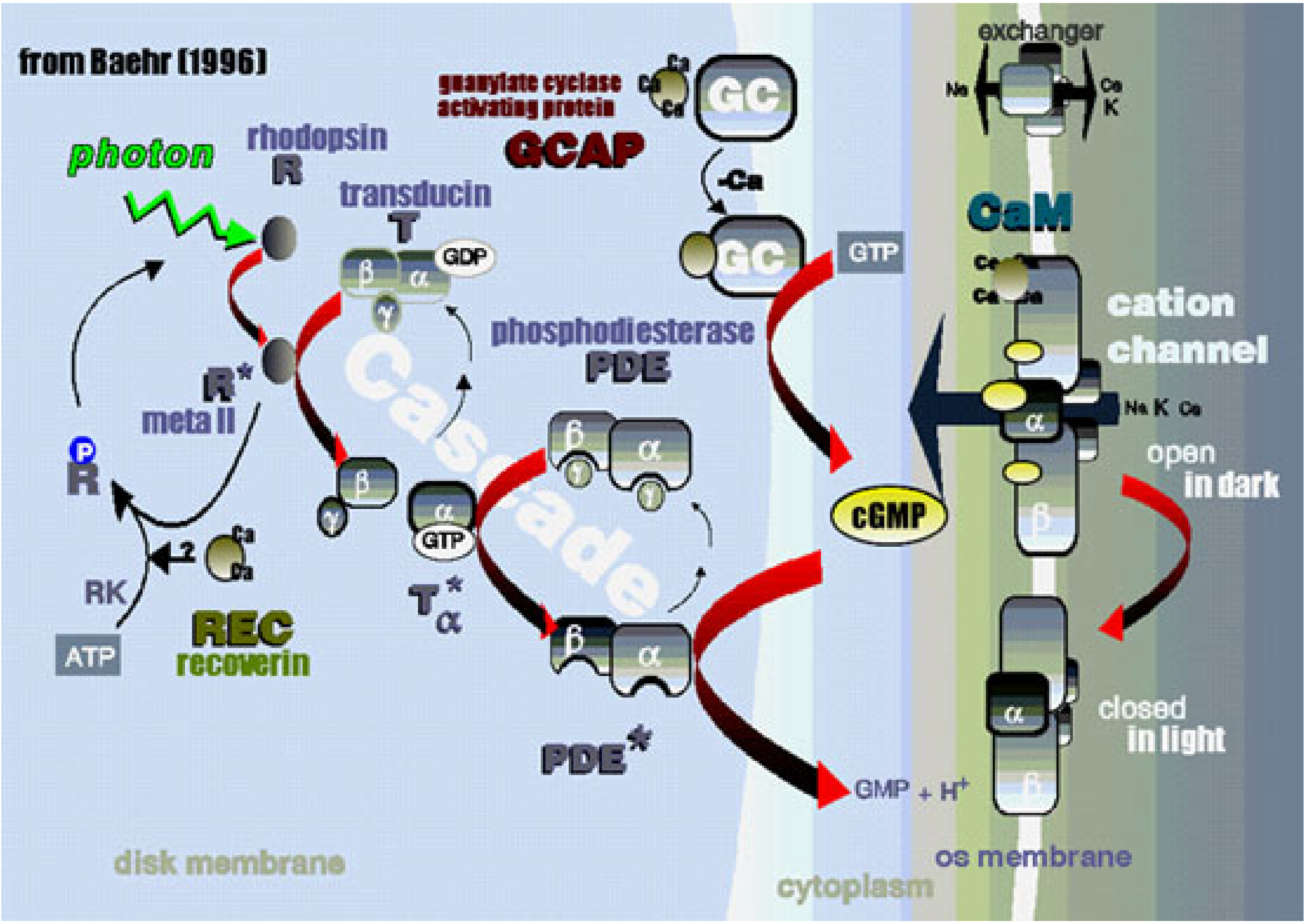
The cornea and lens serve to refract light and focus an image of the object upon the retinal surface.



Accomodation: ciliary muscles can adjust shape of lens, yielding an effect equivalent to an autofocus.

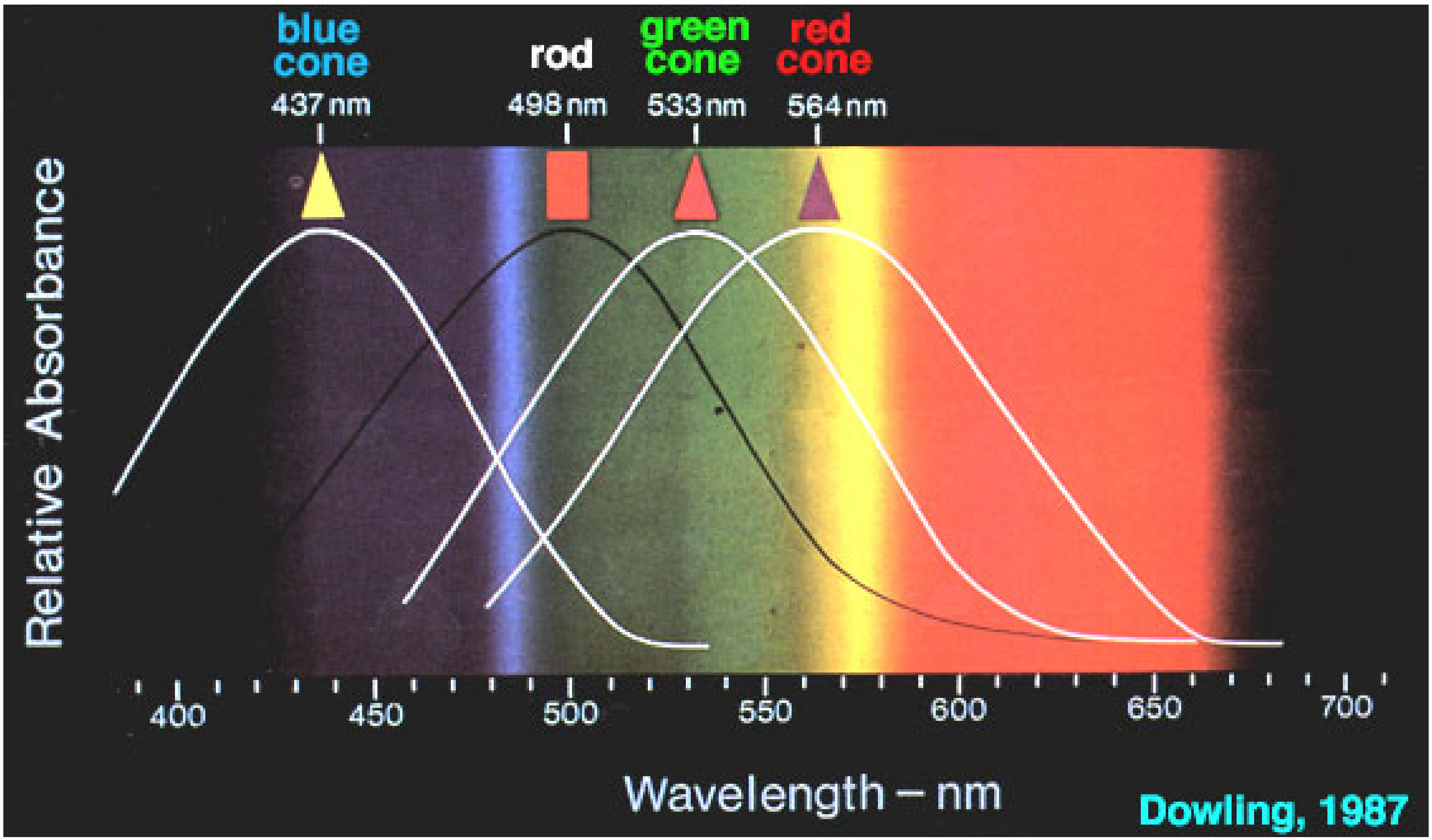
Phototransduction Cascade

Net effect: light (photons) is transformed into electrical (ionic) current.

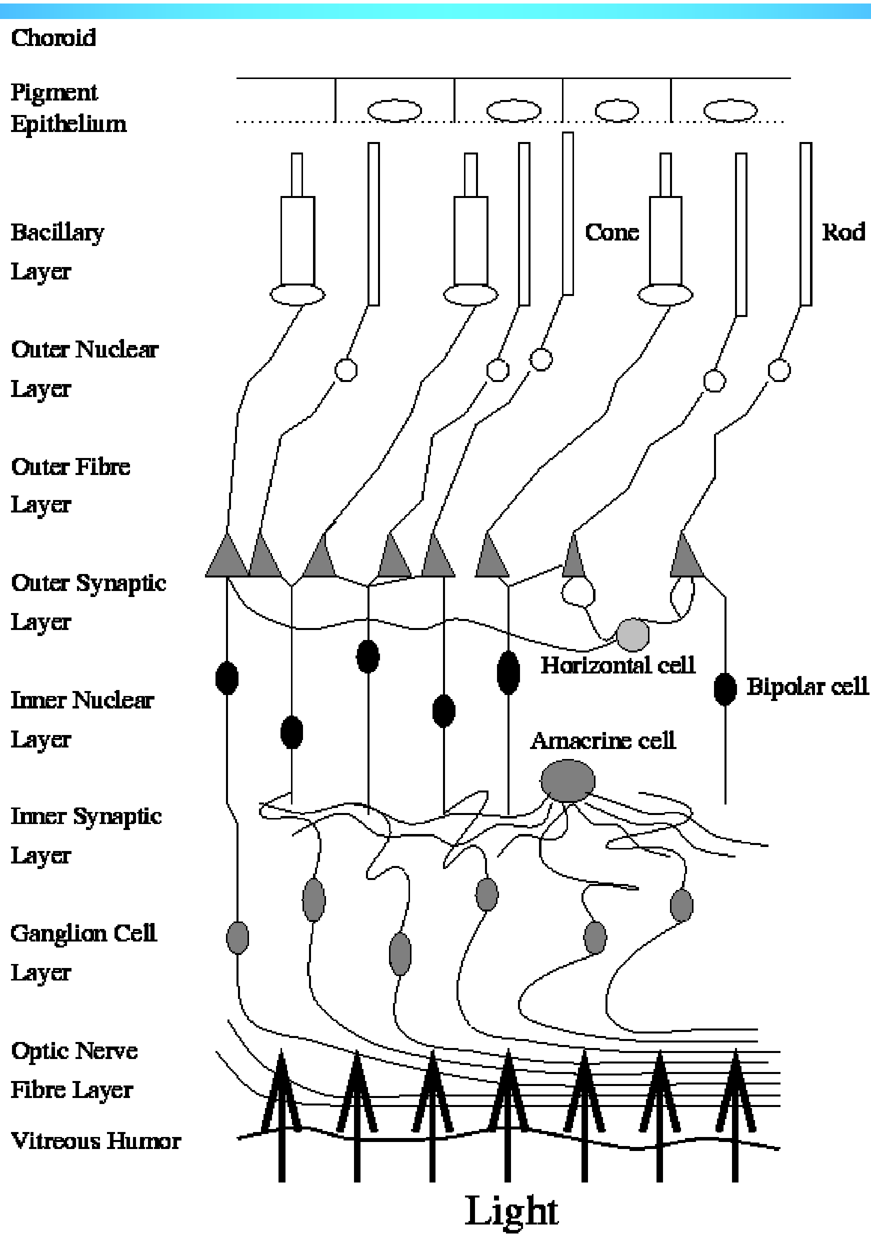


Rods and Cones

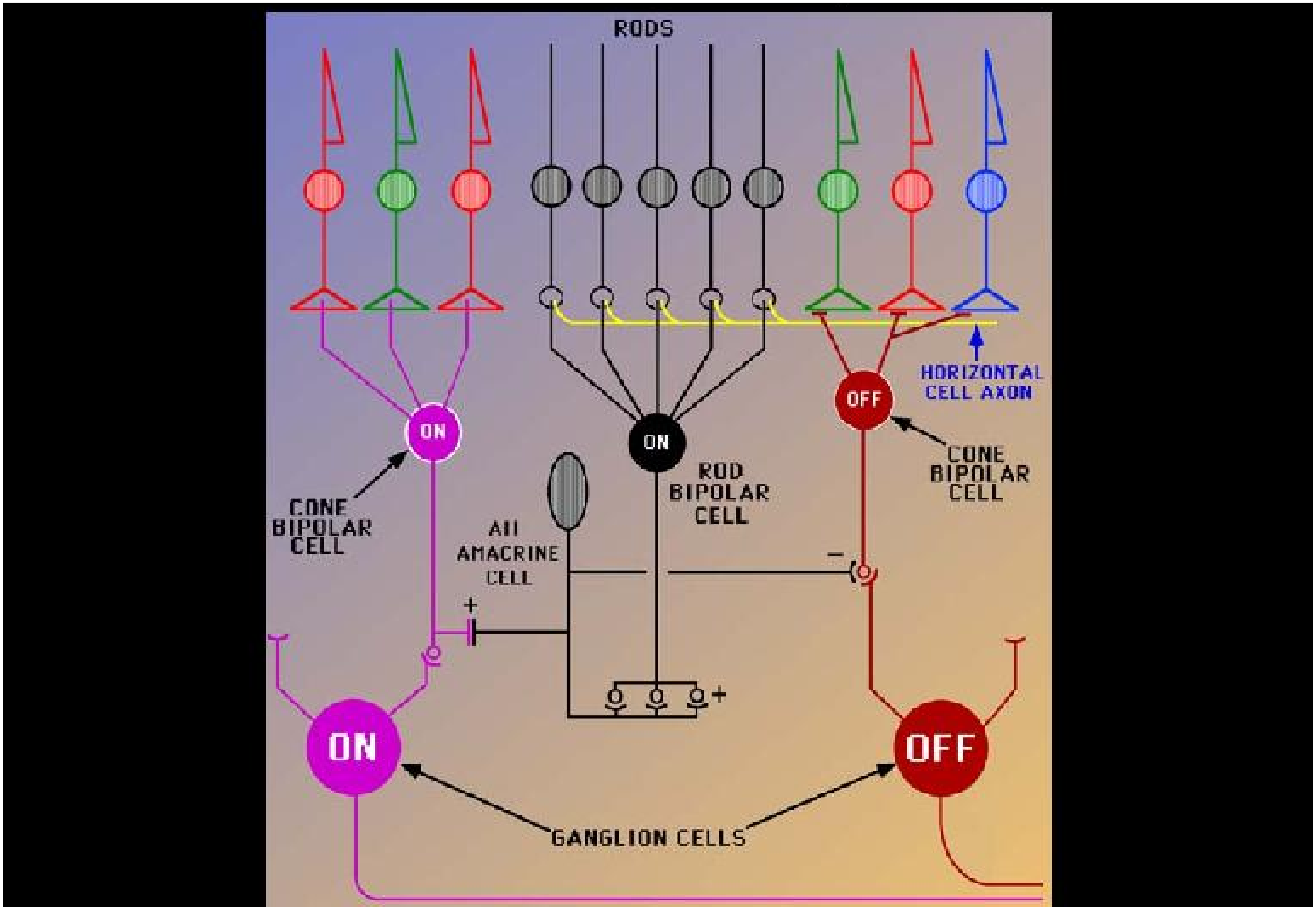
Roughly speaking: 3 types of cones, sensitive to red, green and blue.



Processing layers in retina

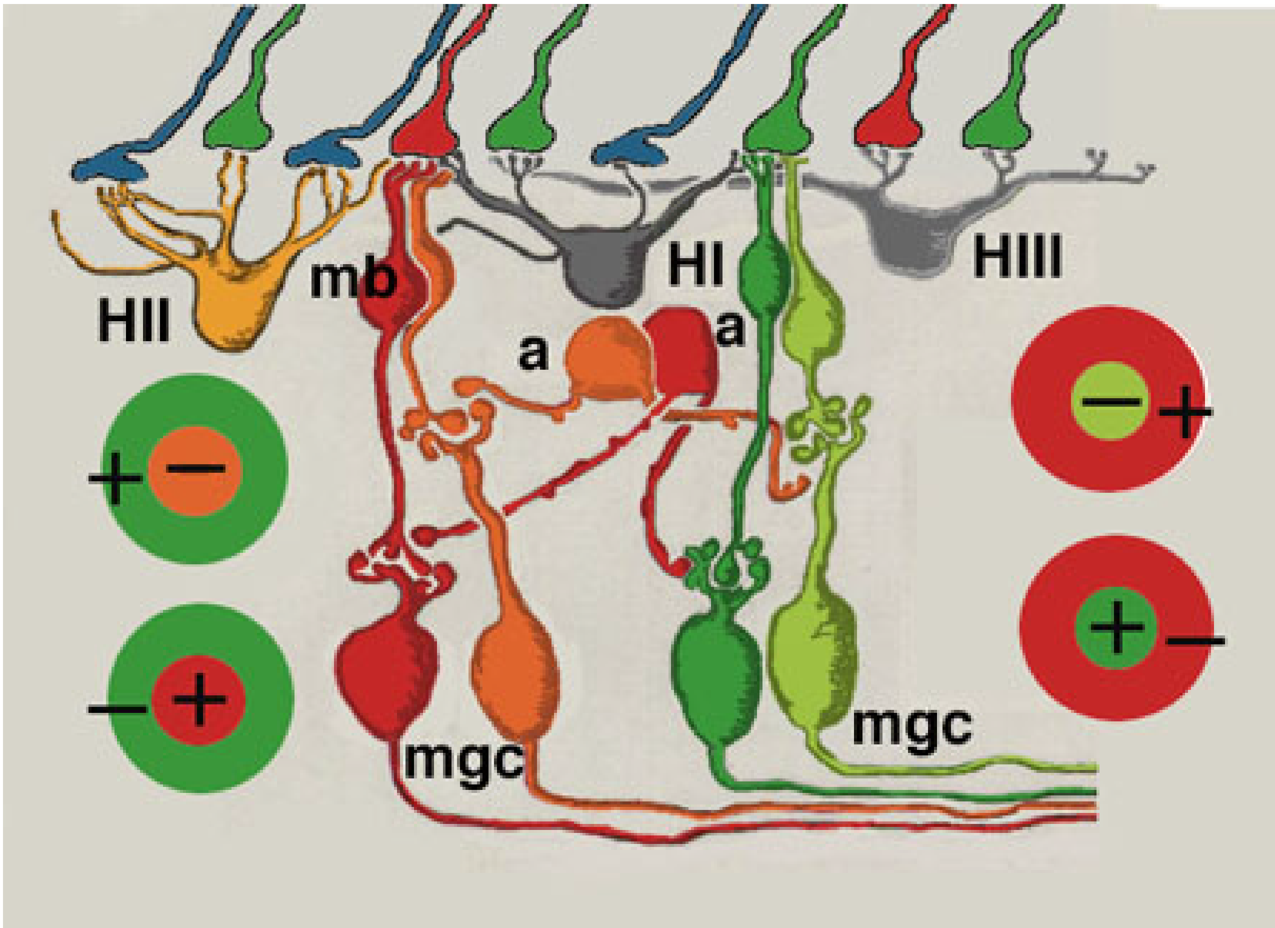


Retinal Processing

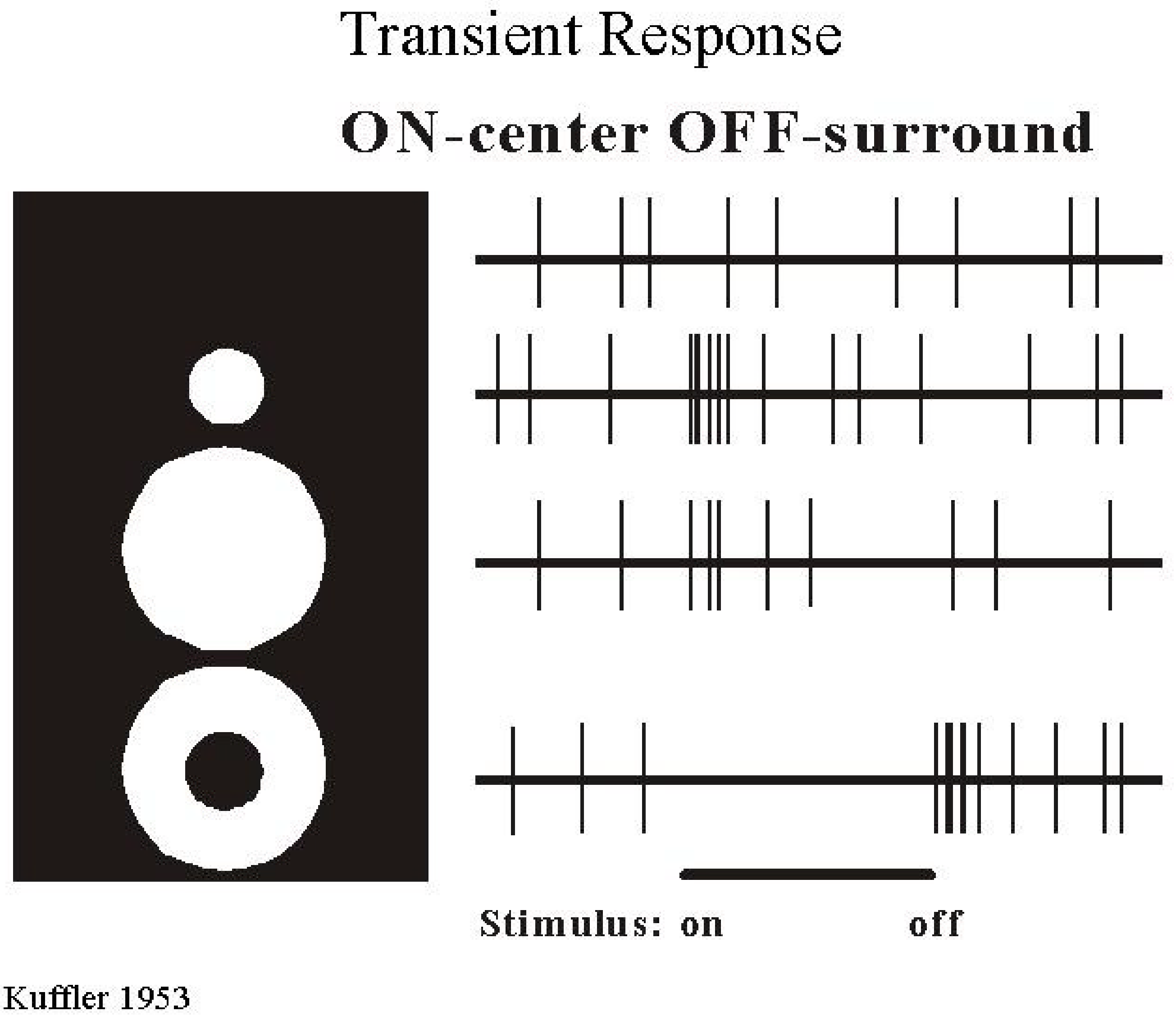


Center-Surround

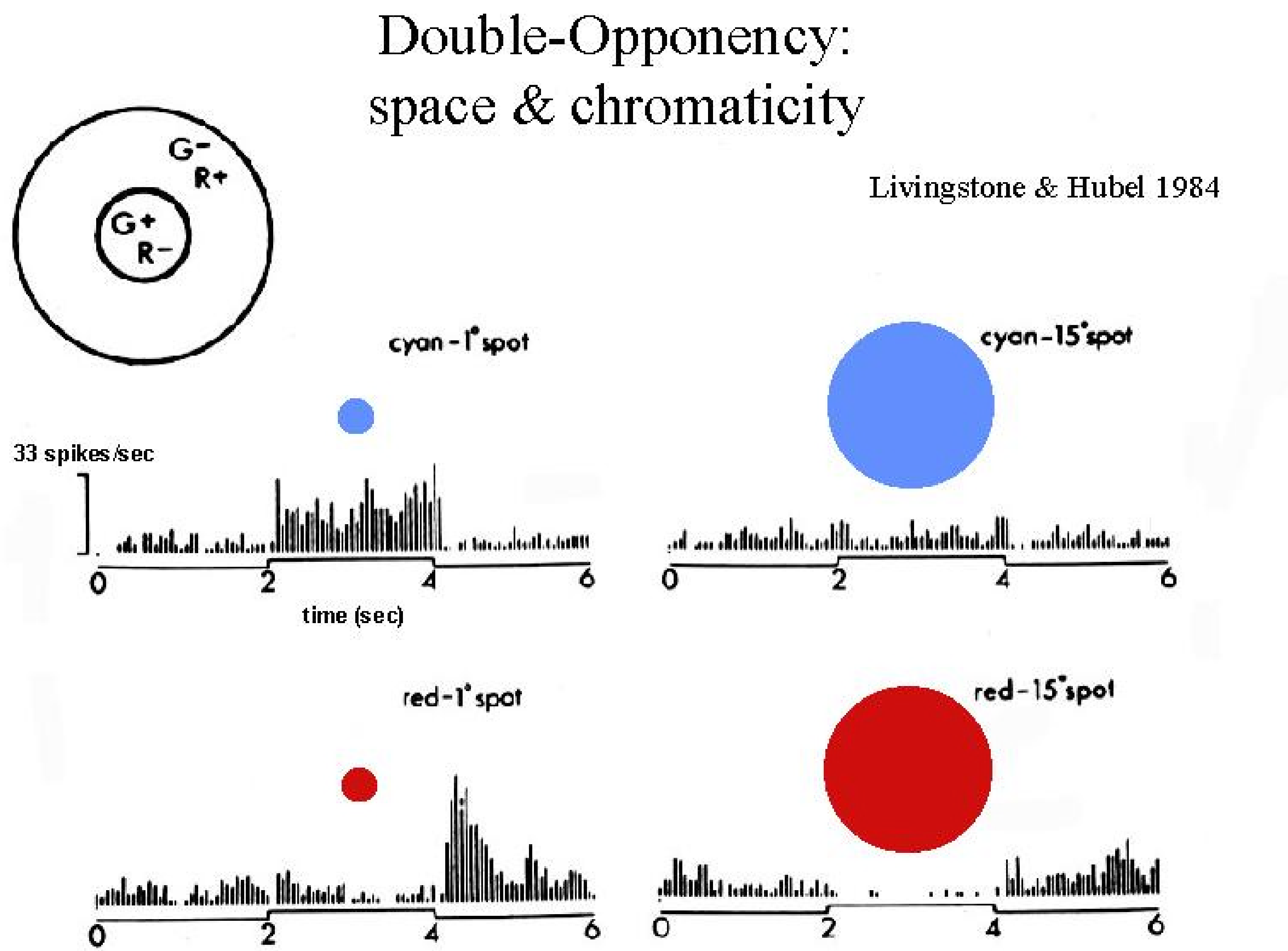
Center-surround organization: neurons with receptive field at given location receive inhibition from neurons with receptive fields at neighboring locations (via inhibitory interneurons).



Early Processing in Retina

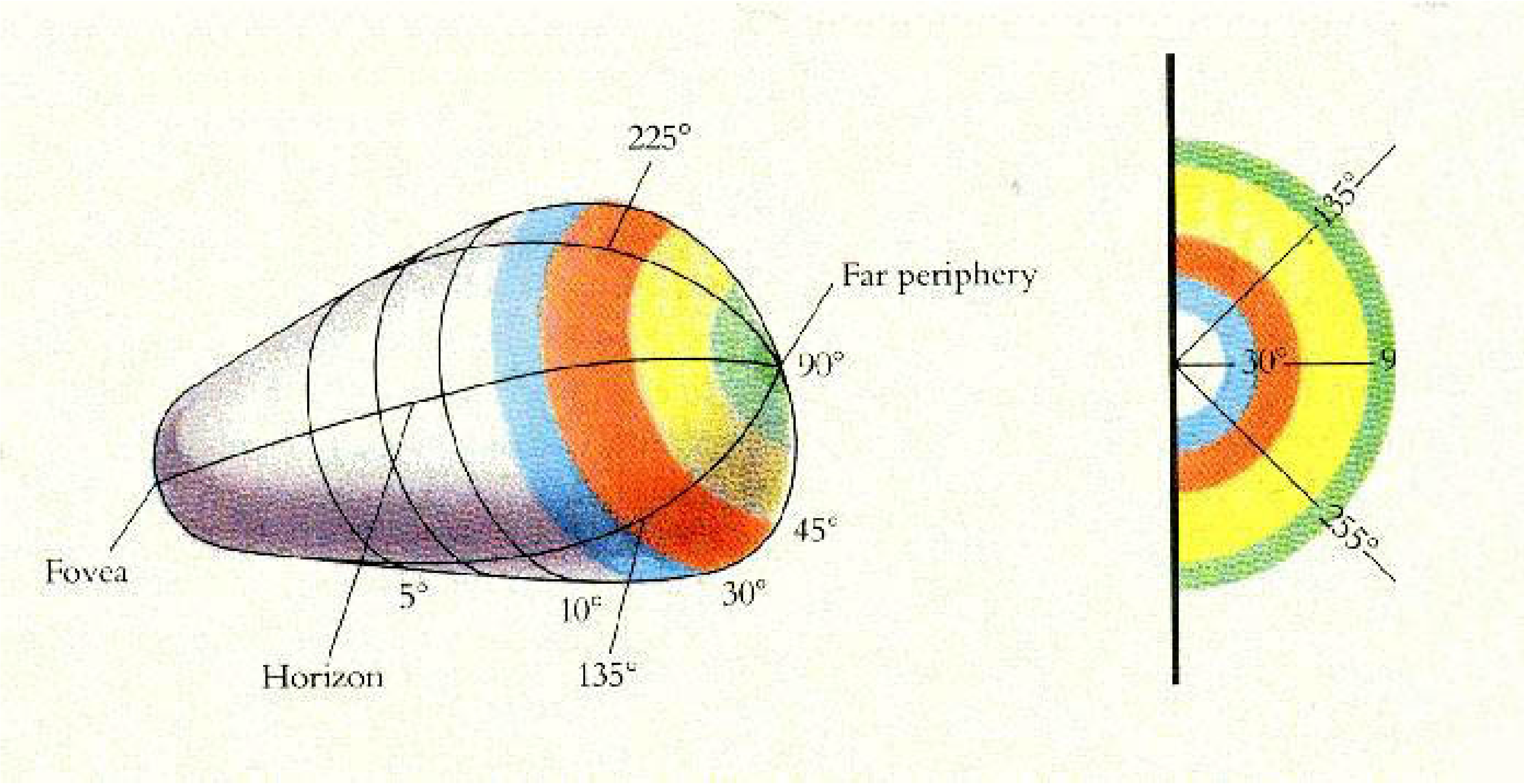


Color Processing

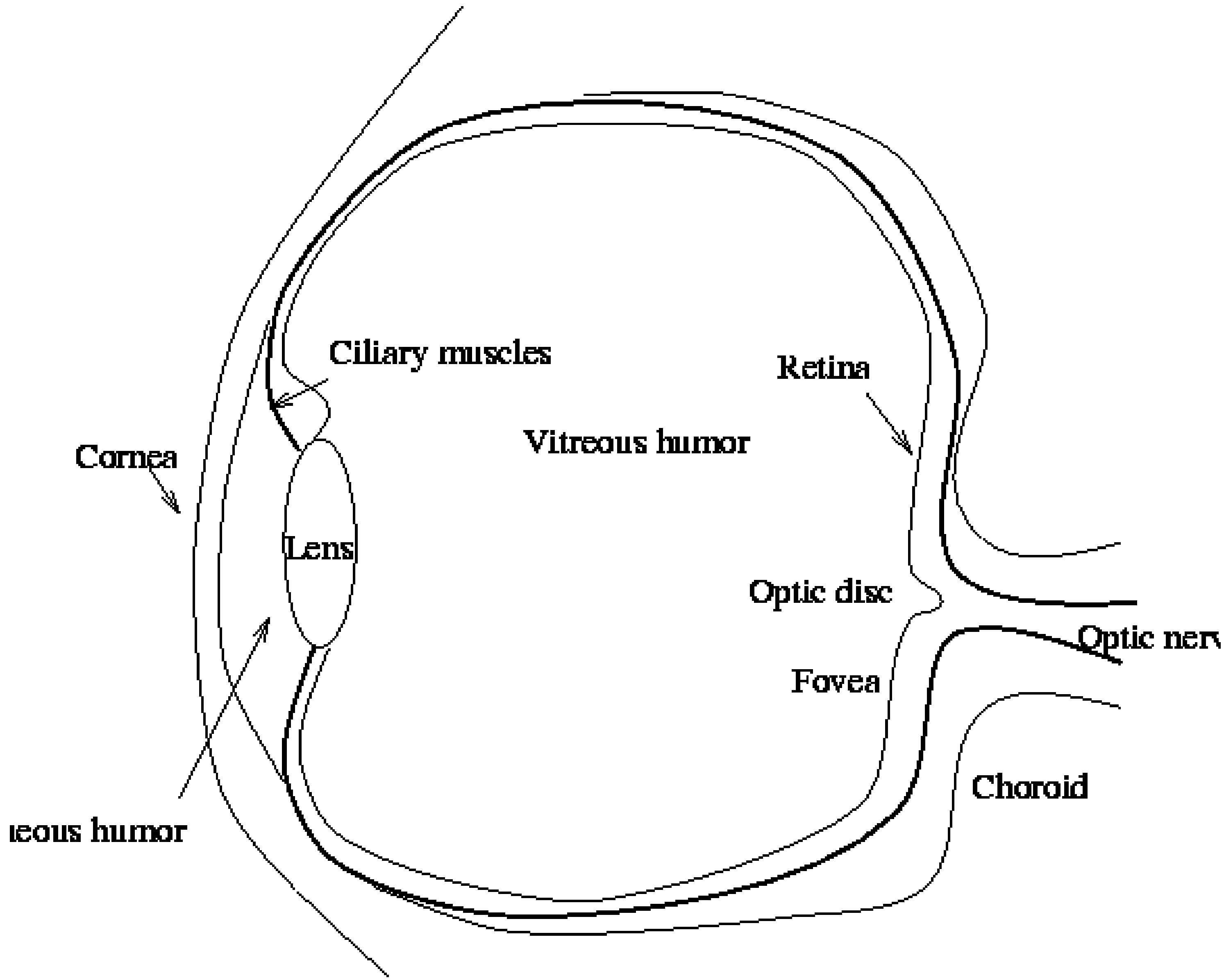


Over-representation of the Fovea

Fovea: central region of the retina (1-2deg diameter); has much higher density of receptors, and benefits from detailed cortical representation.

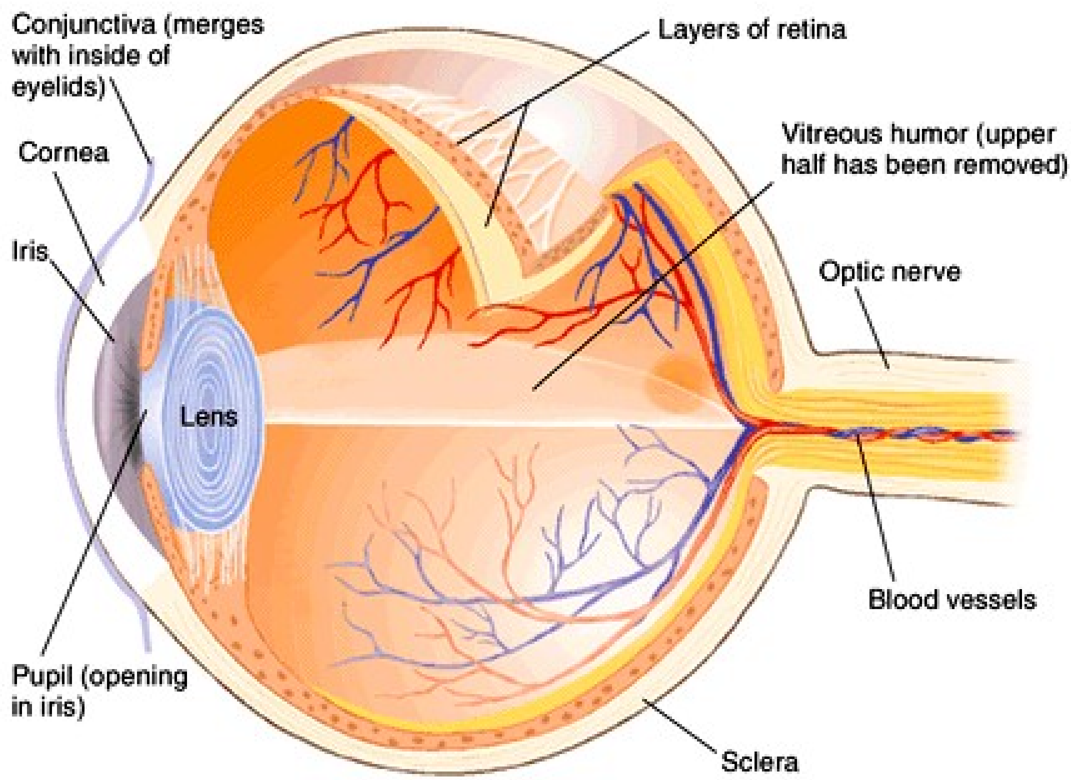


Fovea and Optic Nerve



Blind Spot

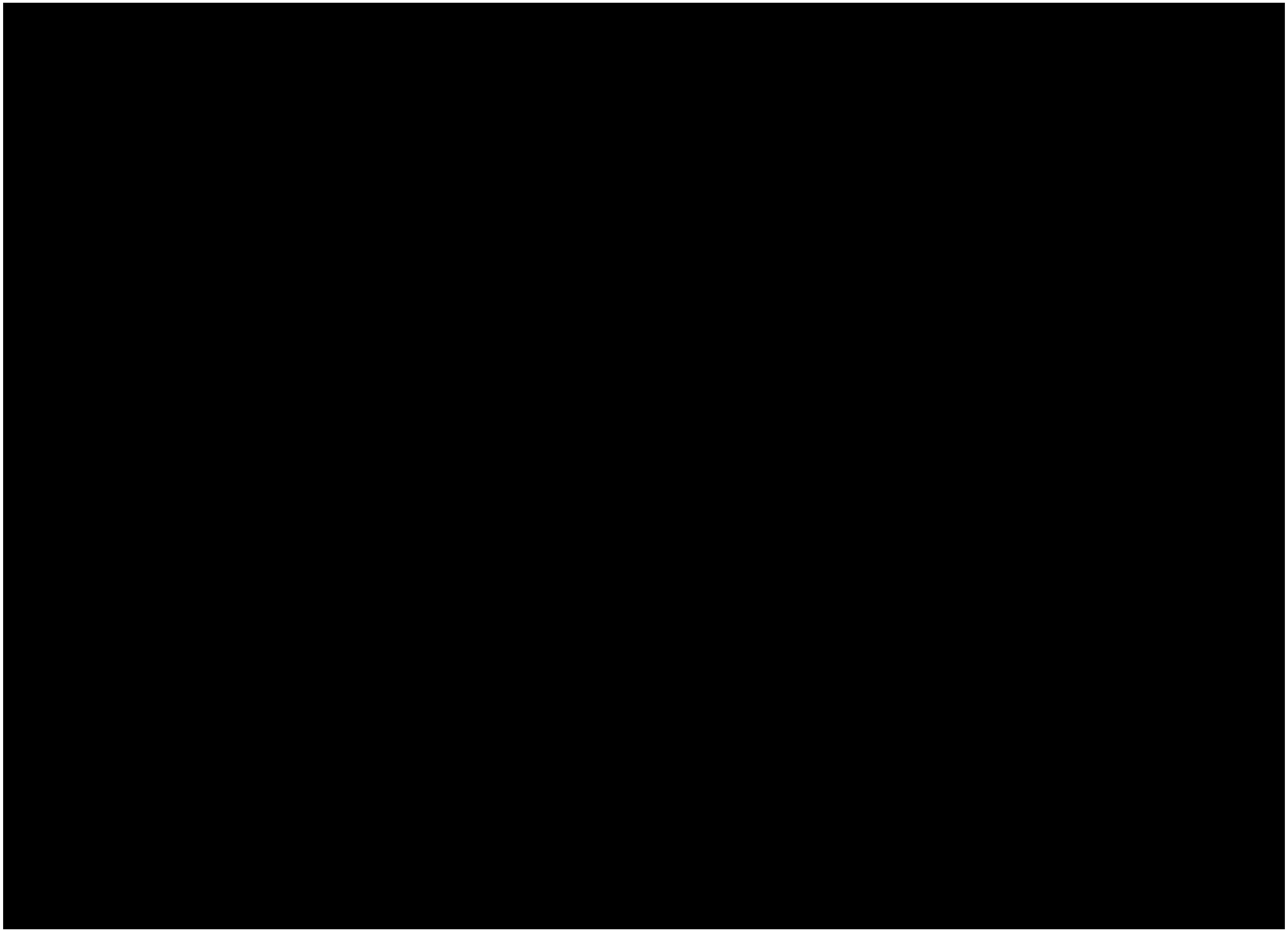
- **The Human Eye**



Retinal Sampling

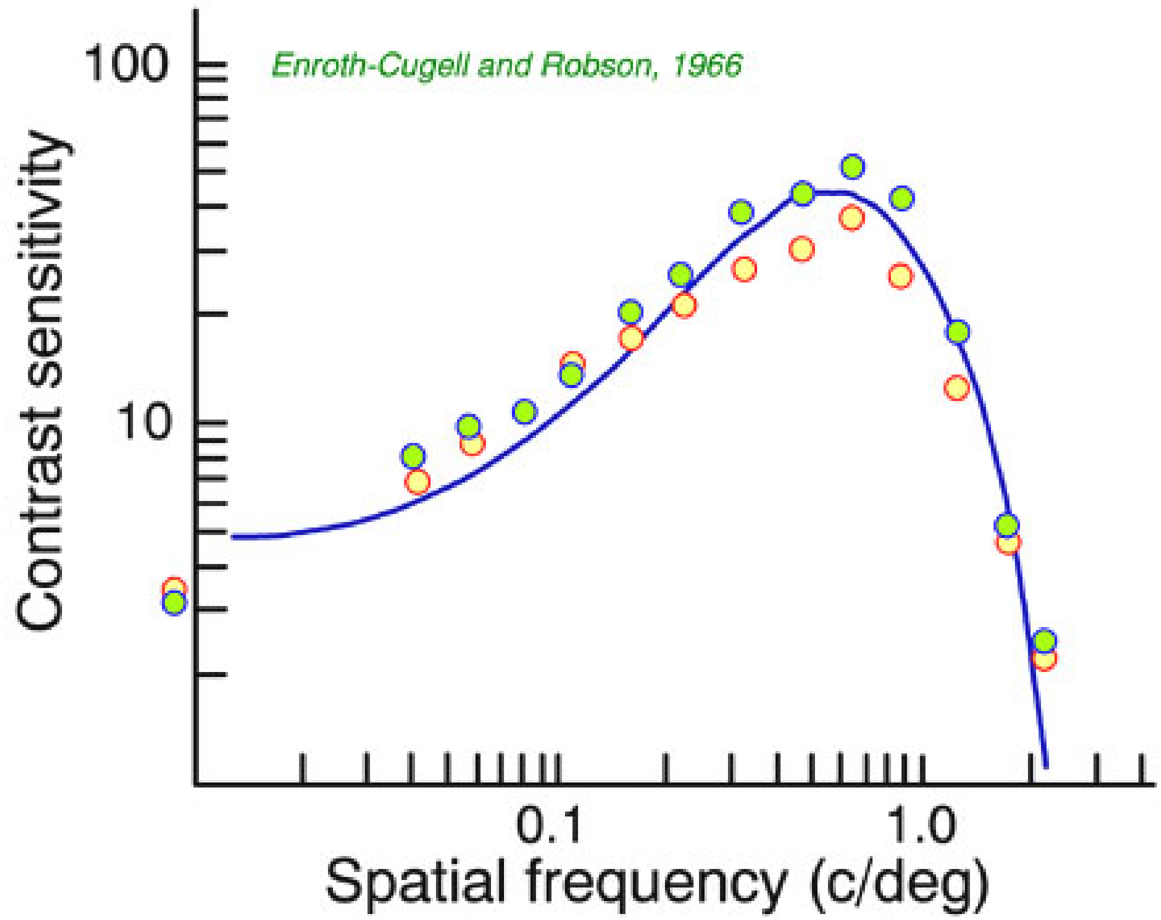
Retinal Sampling

Seeing the world through a retina

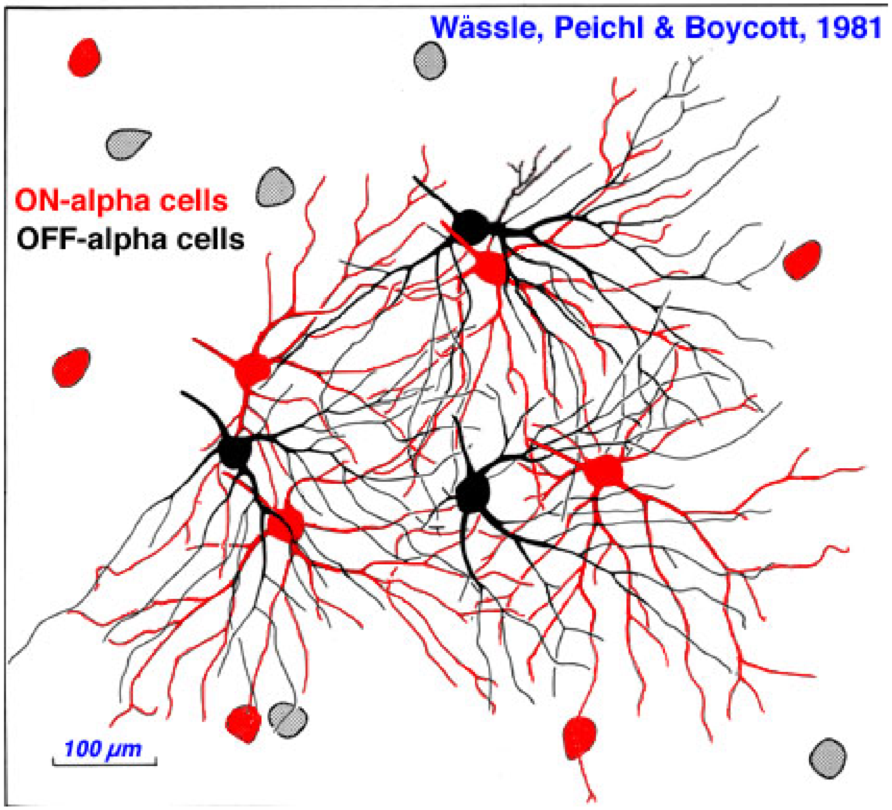


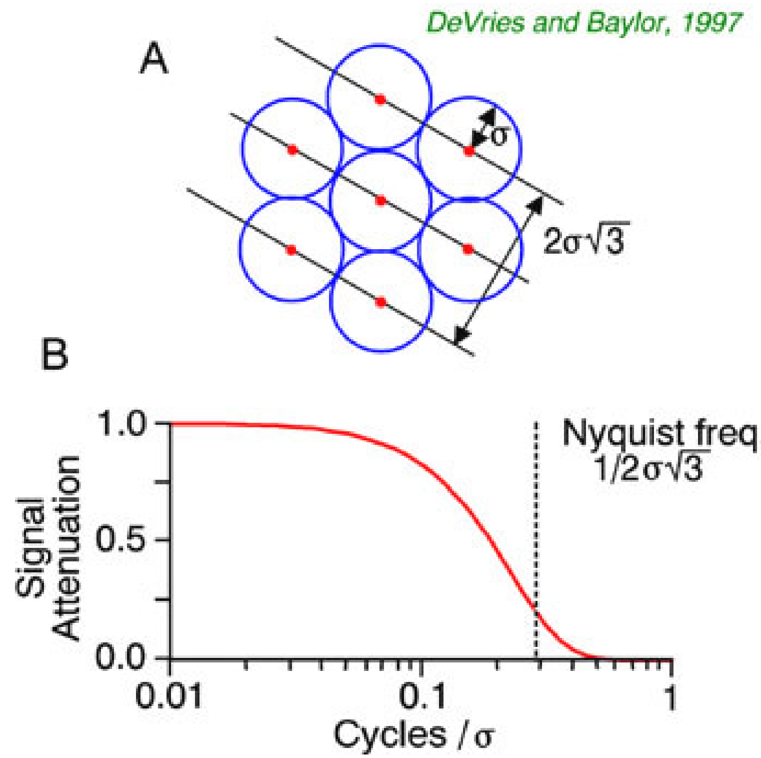
Sampling & Optics

Because of blurring by the optics, we cannot see infinitely small objects…



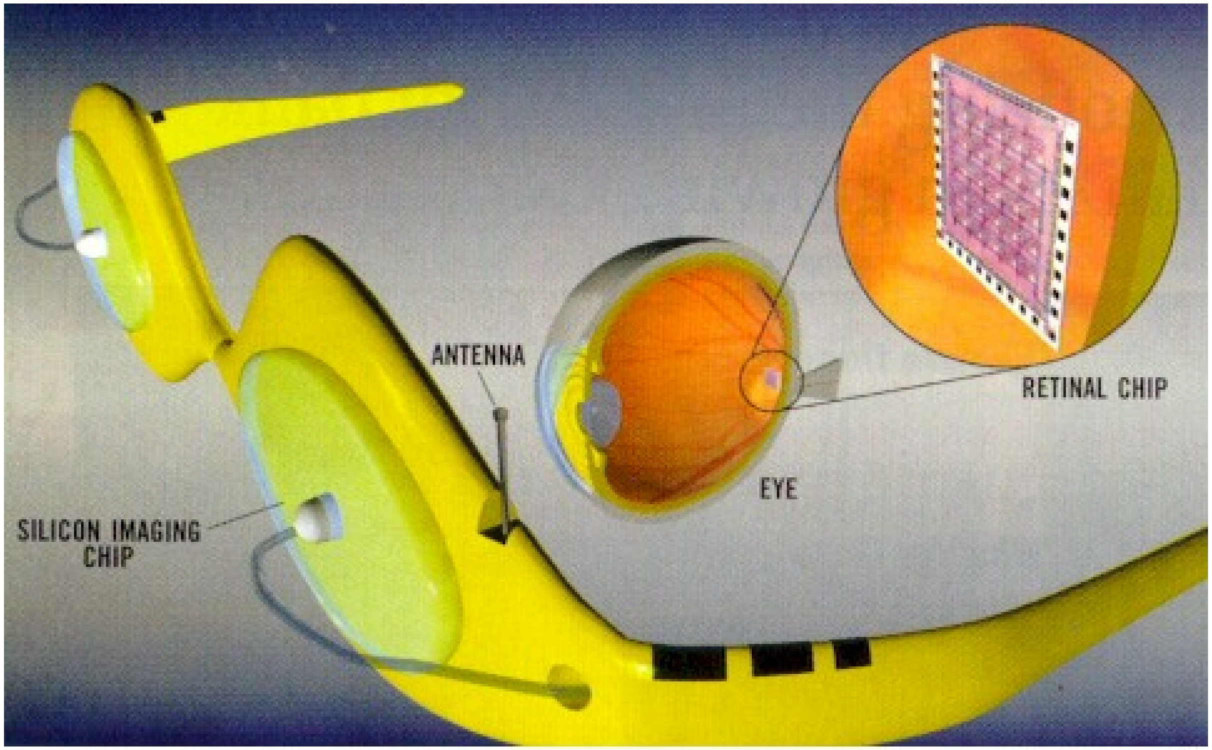
Sampling & optics





The sampling grid optimally corresponds to the amount of blurring due to the optics!

Retinal Implants



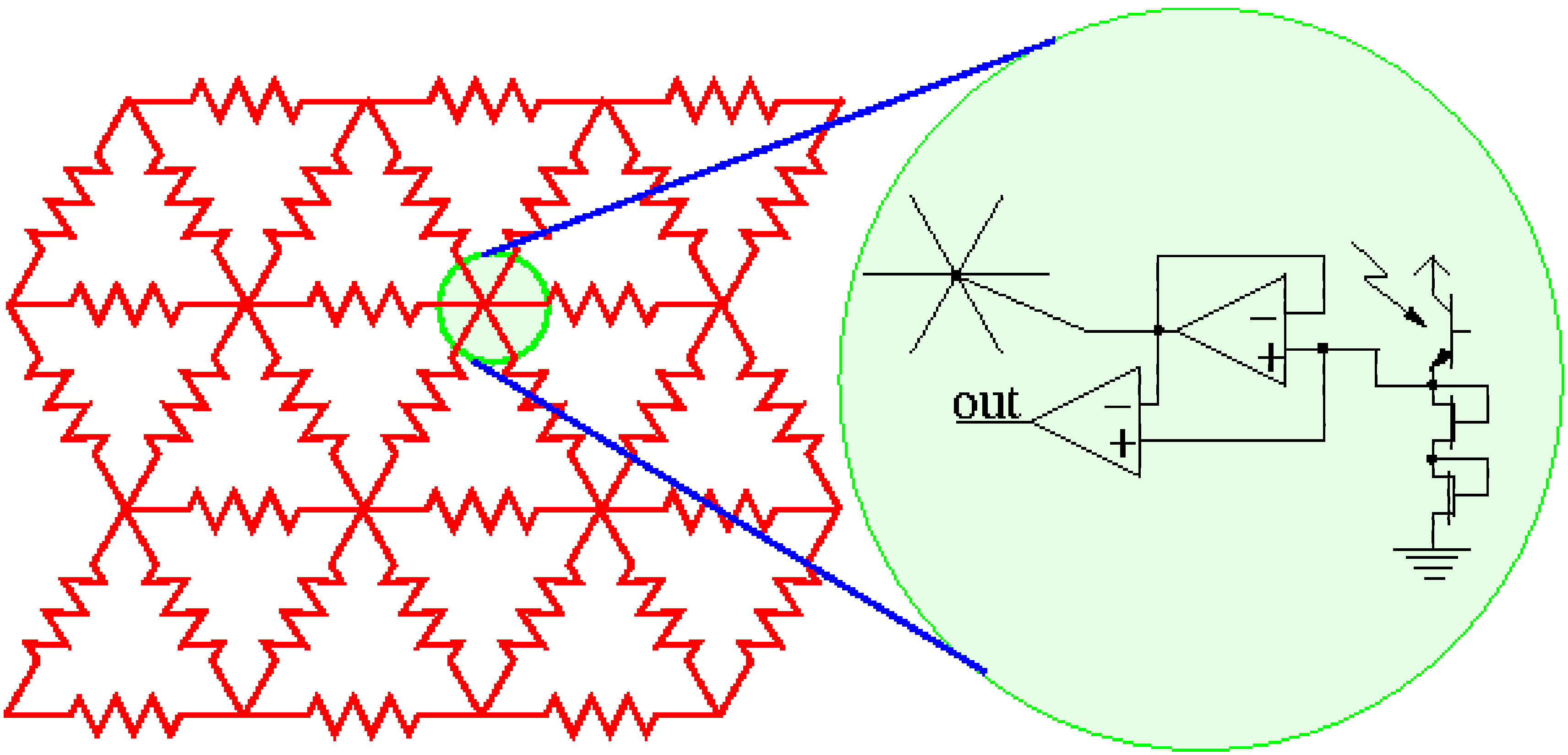
Retinal Prosthesis Project

Johns Hopkins University
North Carolina State University

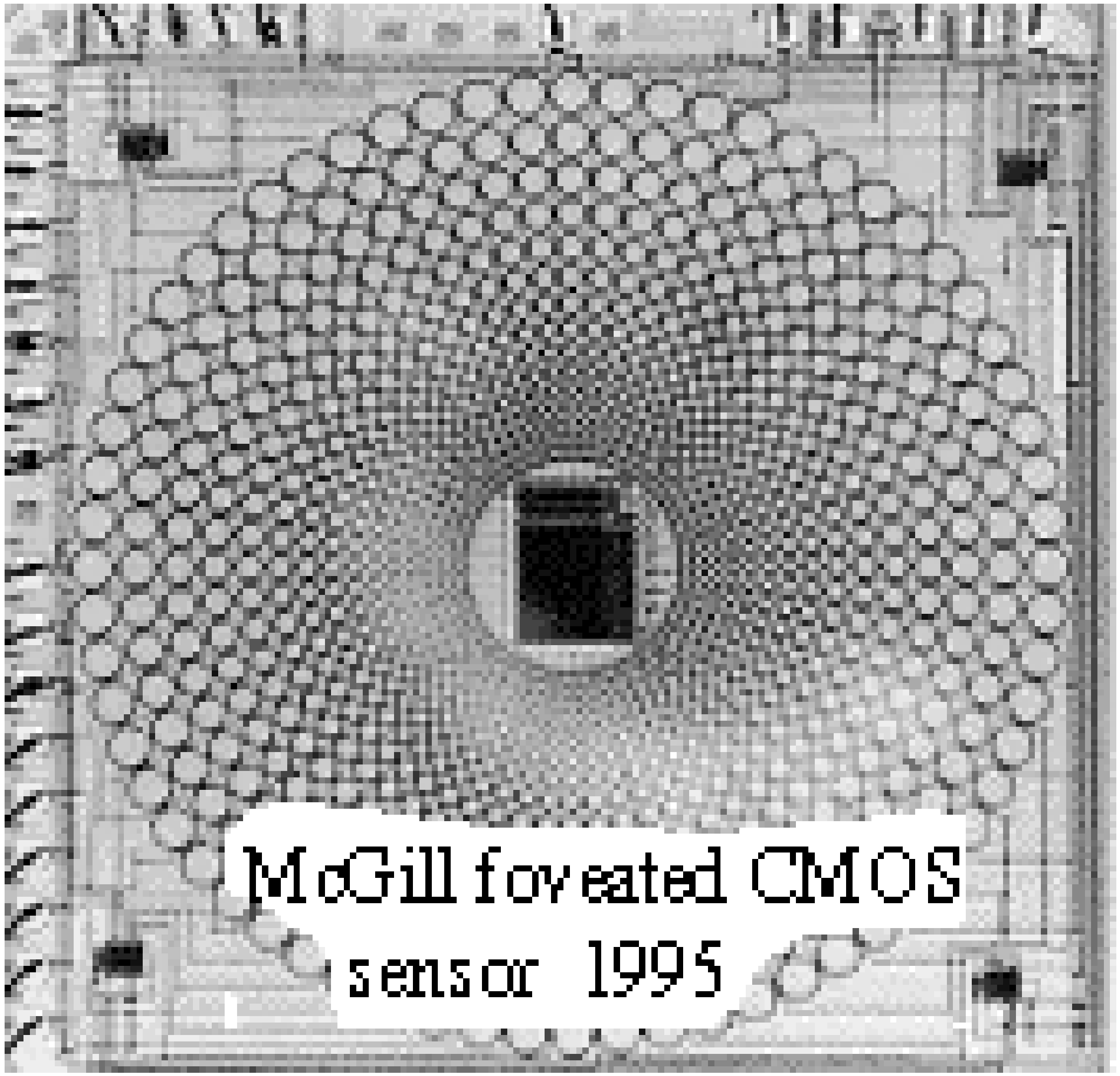
Problems: - long-term toxicity (neurons close to electrodes die)
- image quality so low that users hate it!

Silicon Retina

Smoothing network: allows system to adapt to various light levels.



VLSI sensor vith retinal organization



Combining cameras

From fixed cameras with overlapping field of view, we can reconstruct a virtual camera with any point of view: virtual PTZ (pan-tilt-zoom).

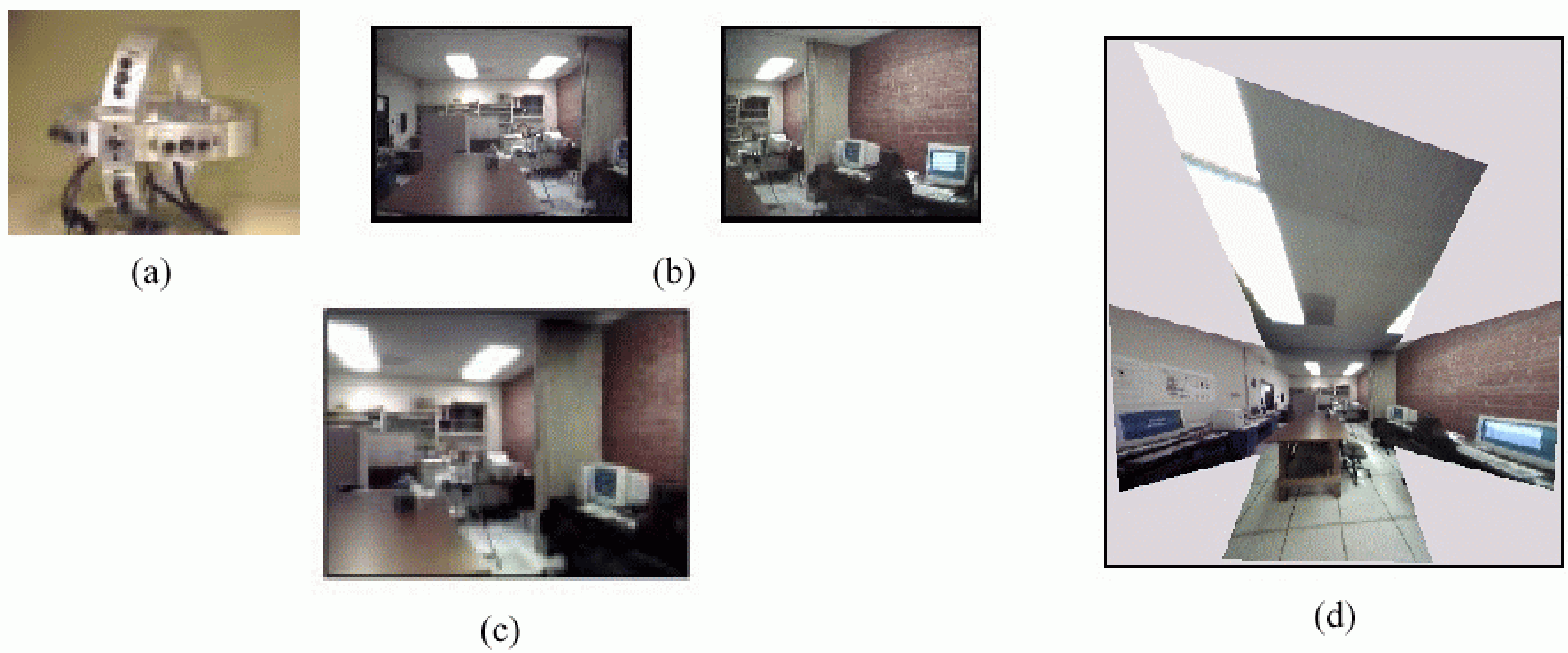


Figure 2: (a) The GlobeAll system, (b) Real images acquired by two cameras of the system, (c) represents a virtual camera generated from the two views. (d) shows a planar mosaic, depicting a real time view of the monitored scene.