Image-based Spatial Processing

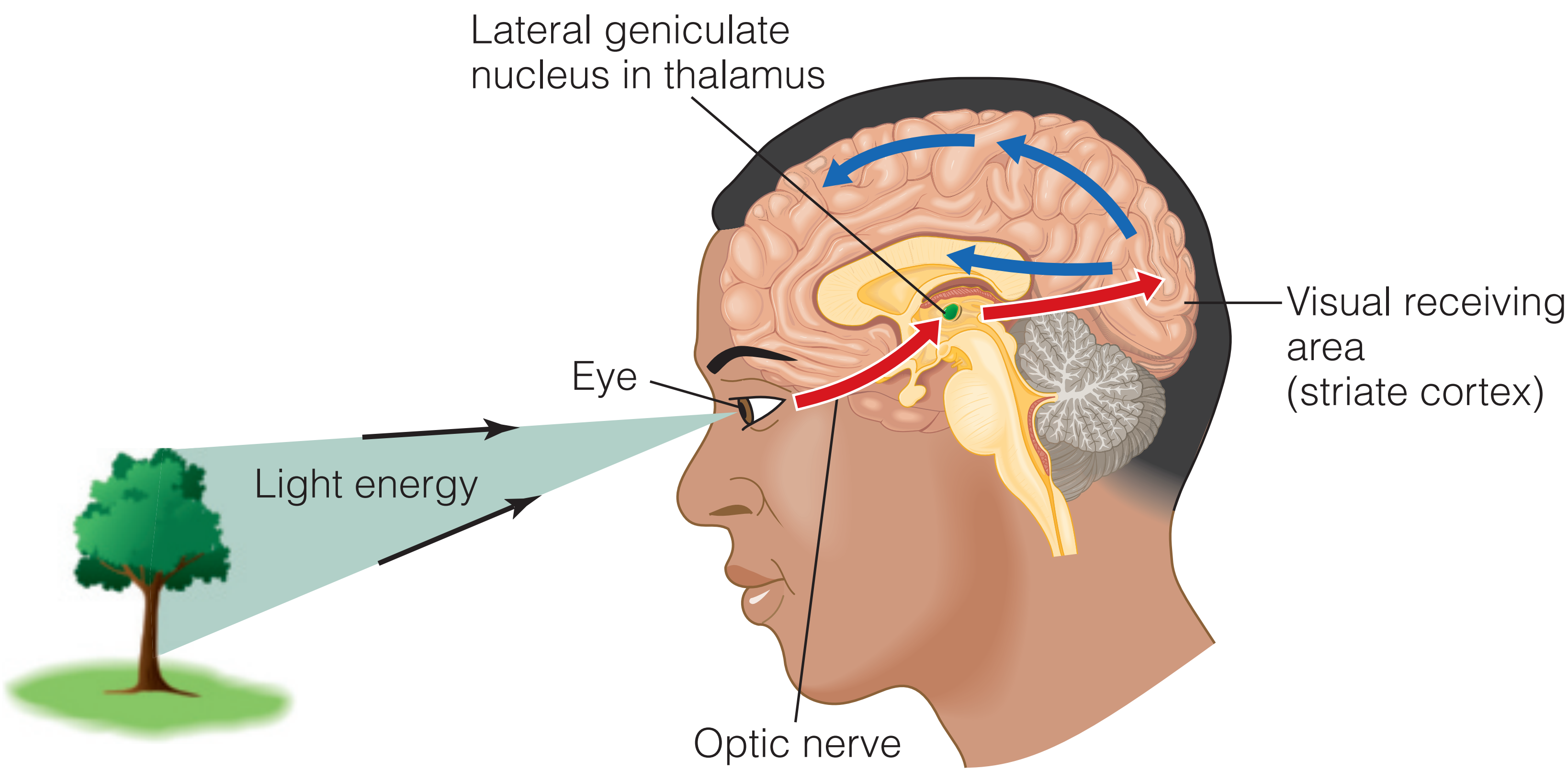
Wu Xihong

Peking University

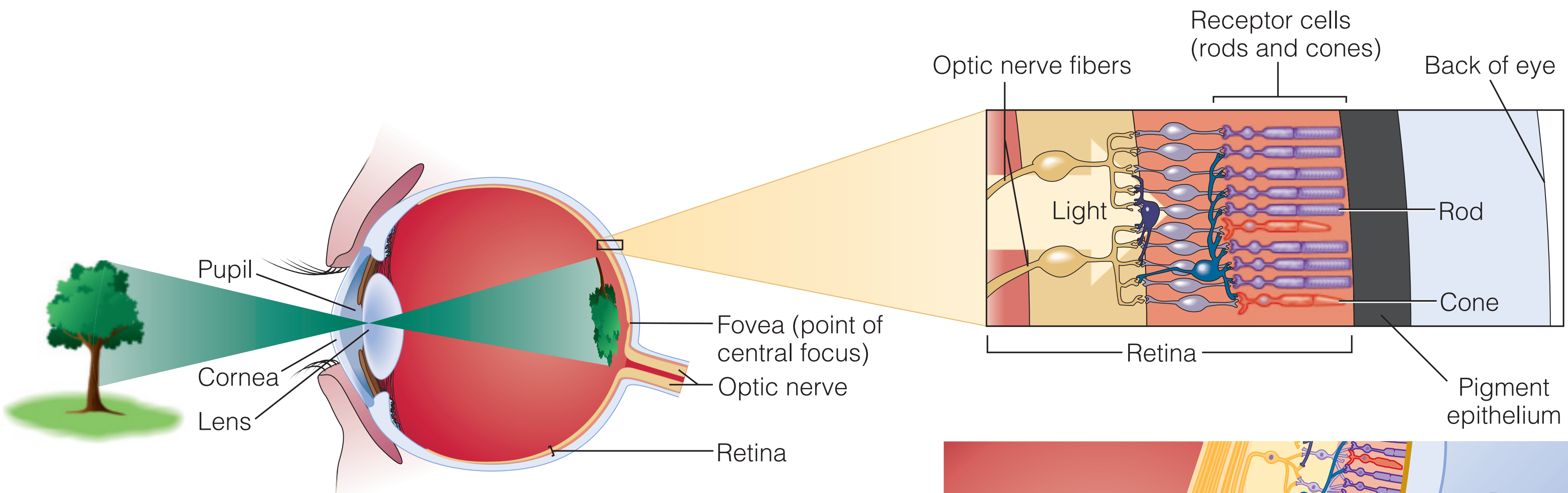
Introduction

- Image-based spatial processing
  - Those visual processes that primarily concern computing spatial features of the 2D retinal image rather than features of external surfaces and objects.
- Three principal approaches:
  - Physiological
  - Psychophysical
  - Computational

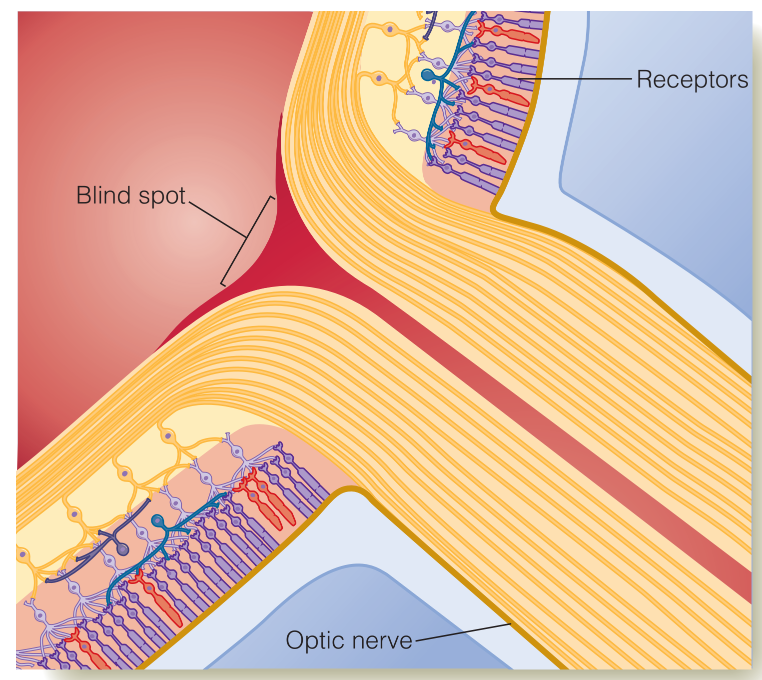
1. Physiological Mechanisms



Eye system

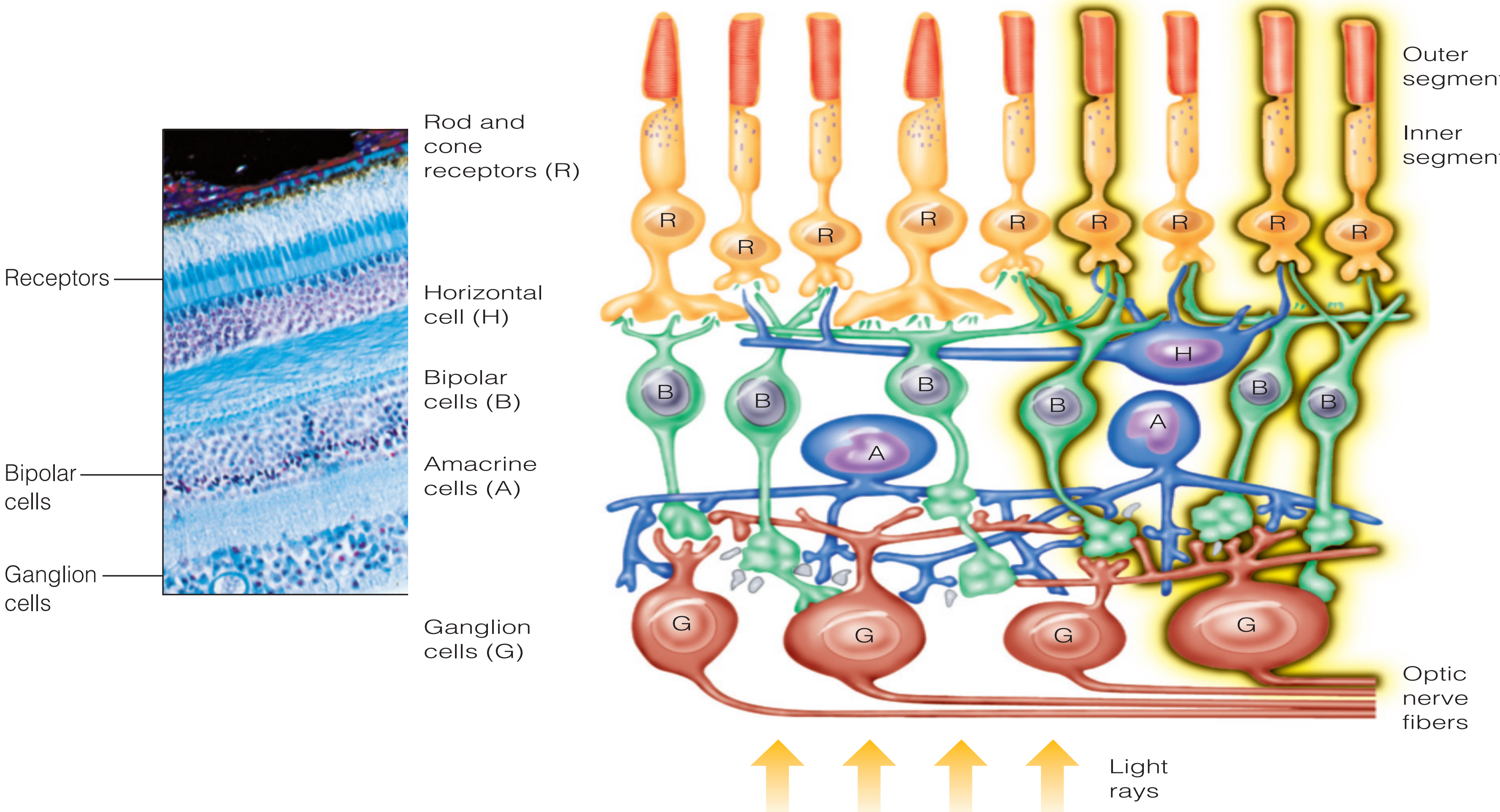


Pupil
Cornea
Lens
Fovea (point of central focus)
Optic nerve
Retina
Optic nerve fibers
Receptor cells (rods and cones)
Back of eye
Light
Rod
Cone
Retina
Pigment epithelium

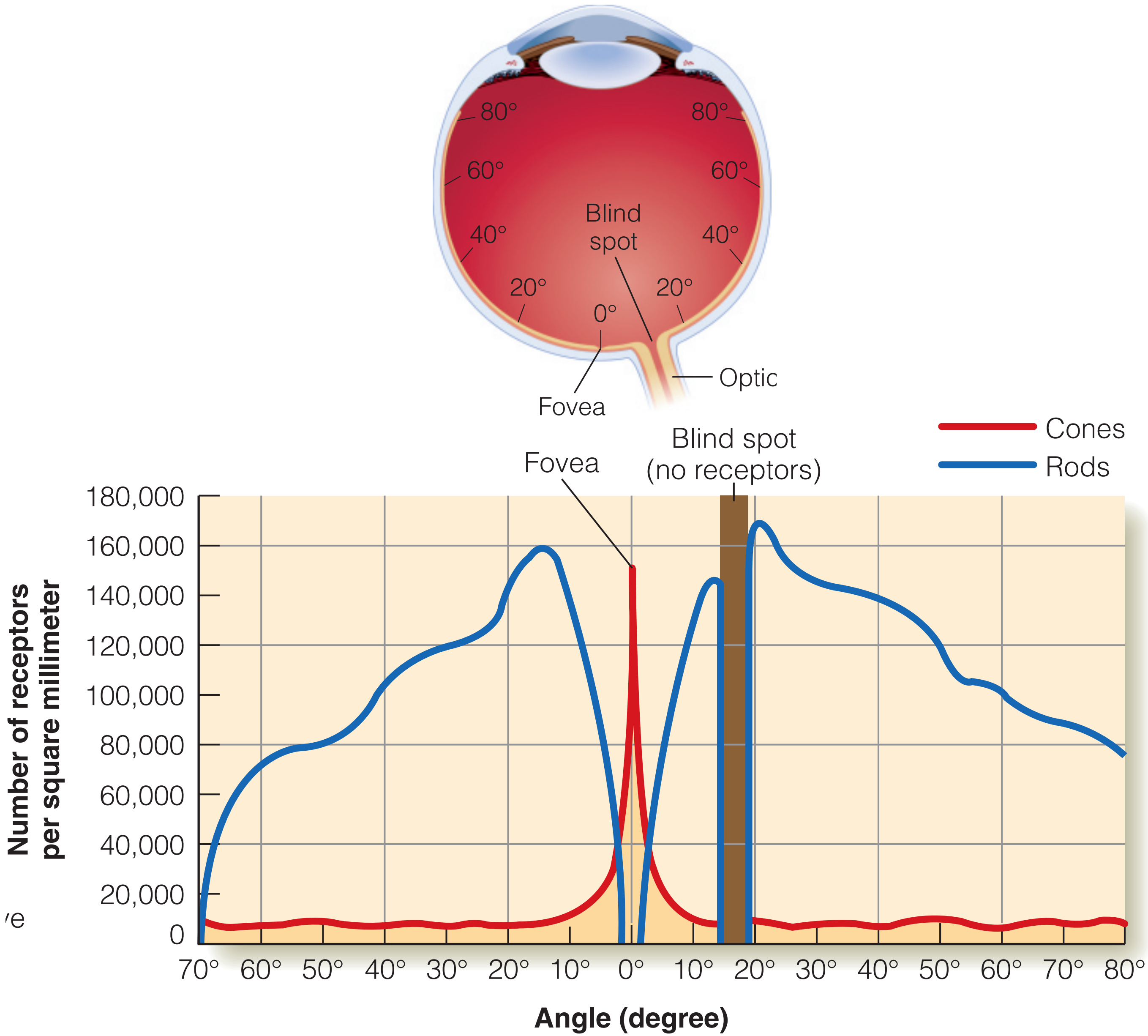


Receptors
Blind spot
Optic nerve

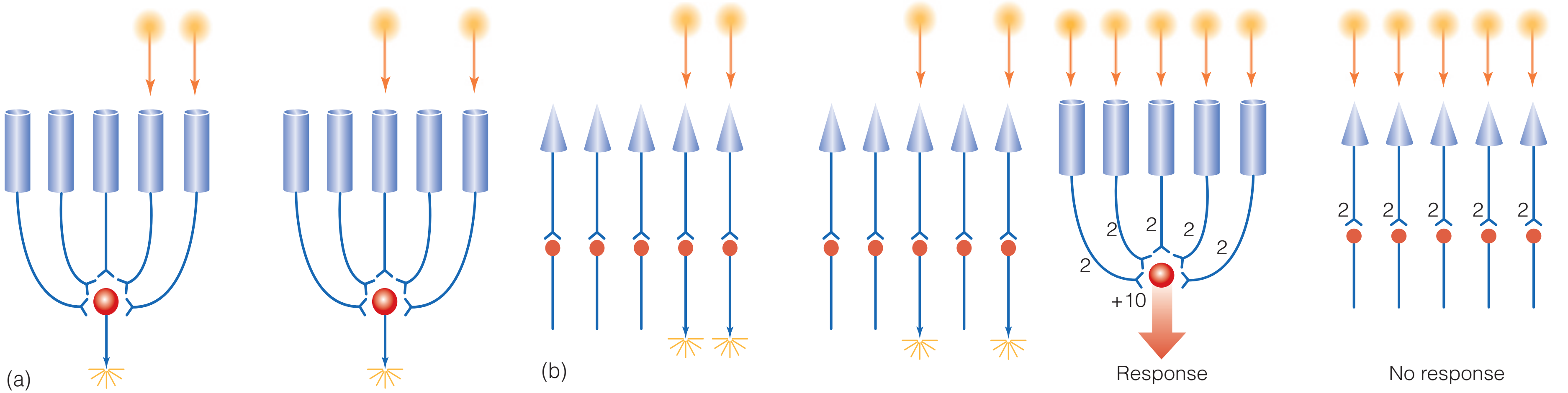
1.1 The Human retina

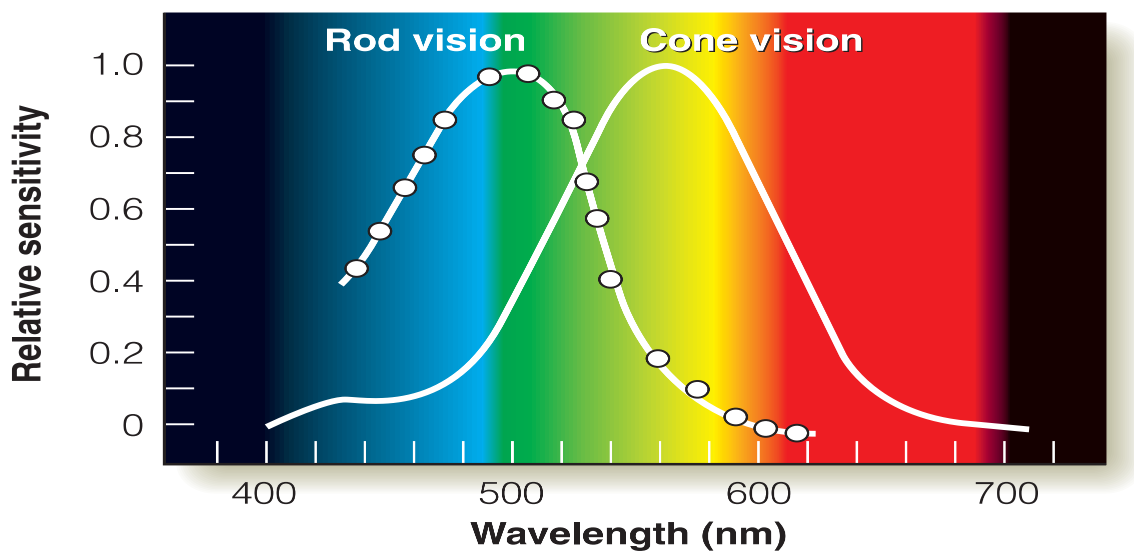


Rods and cones



Sensitivity and resolution

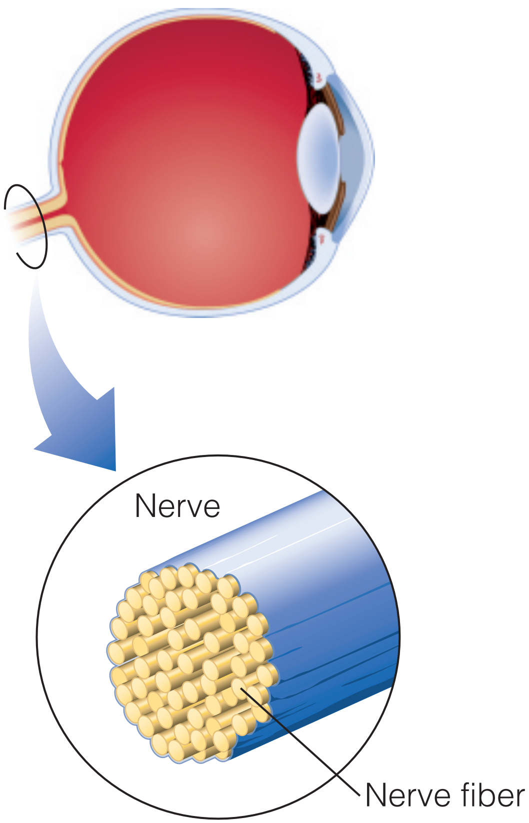


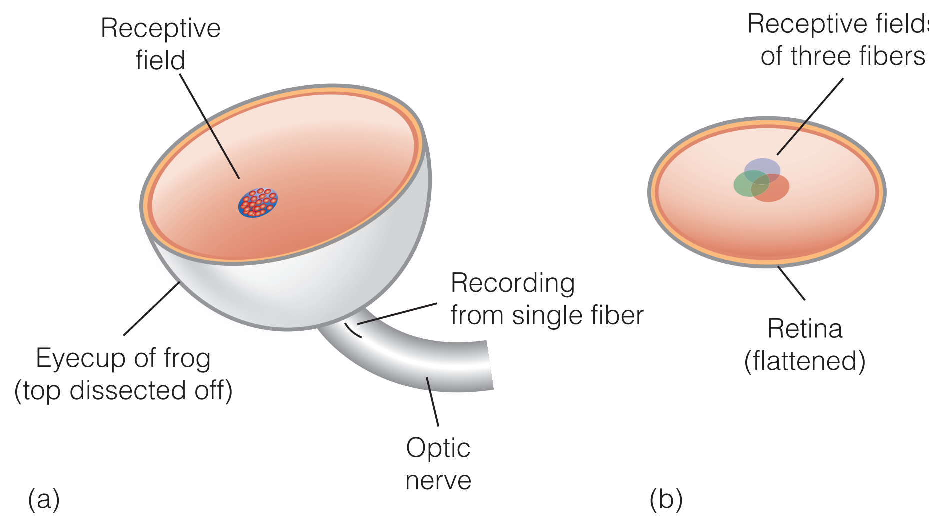


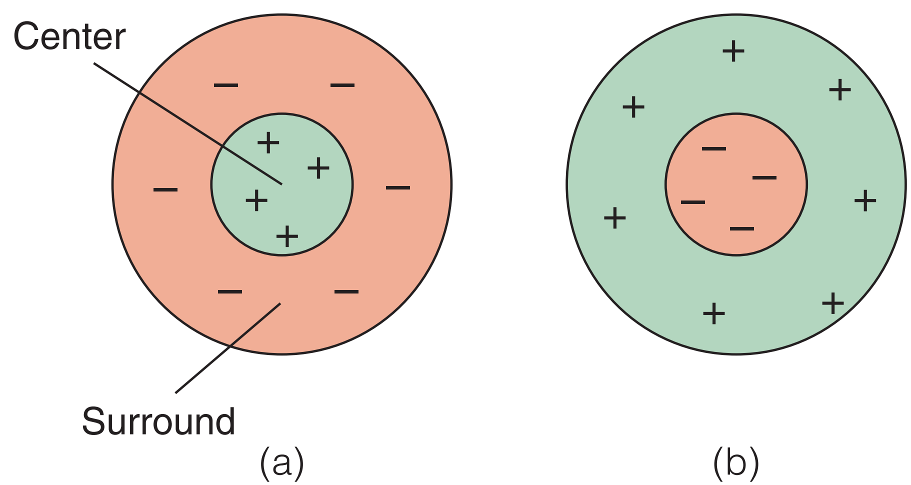
Bipolar cells and horizontal cells

- Receptors and bipolar cells respond by producing graded potentials, more slowly.
- Graded potentials must be recorded from inside the cell intracellular recording.
- Bipolar cells have circularly symmetric receptive fields with antagonistic relations between center and surround. ?

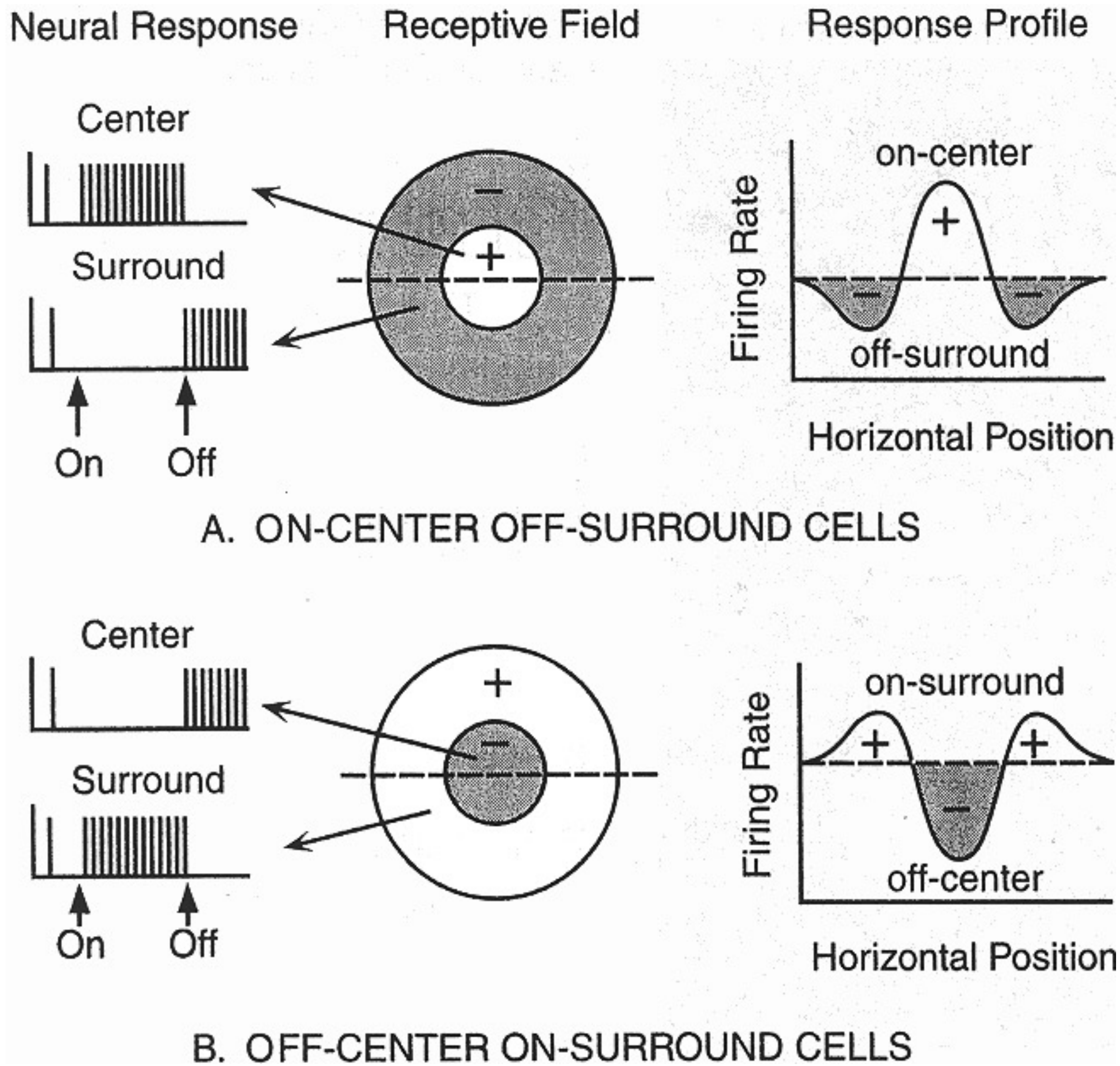
Single fiber‘s recoding and Receptive field

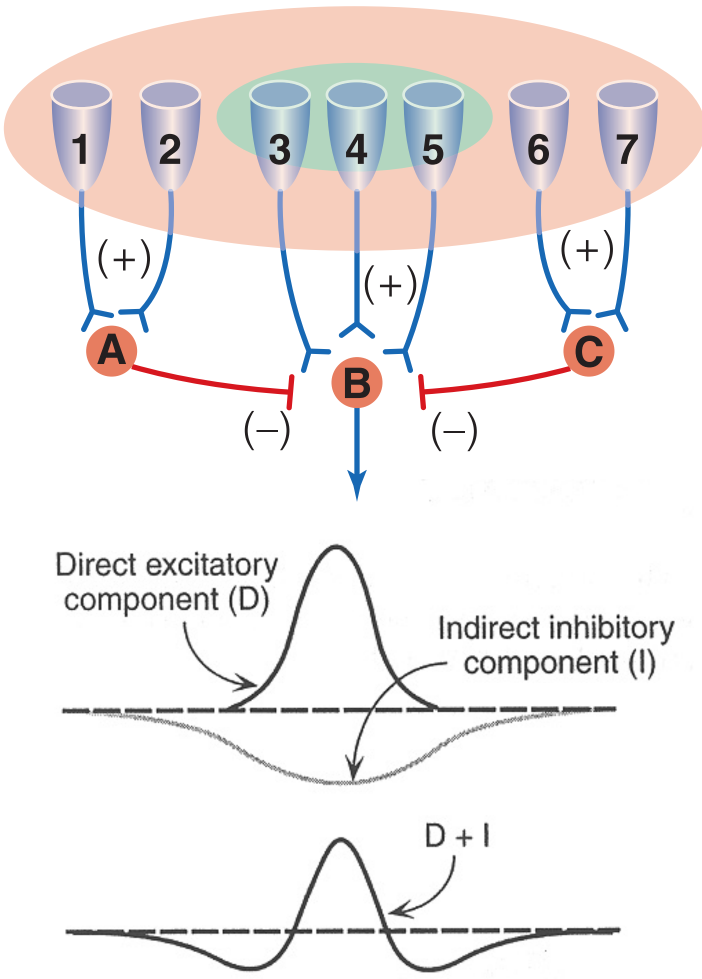






On-center cells and off-center cells



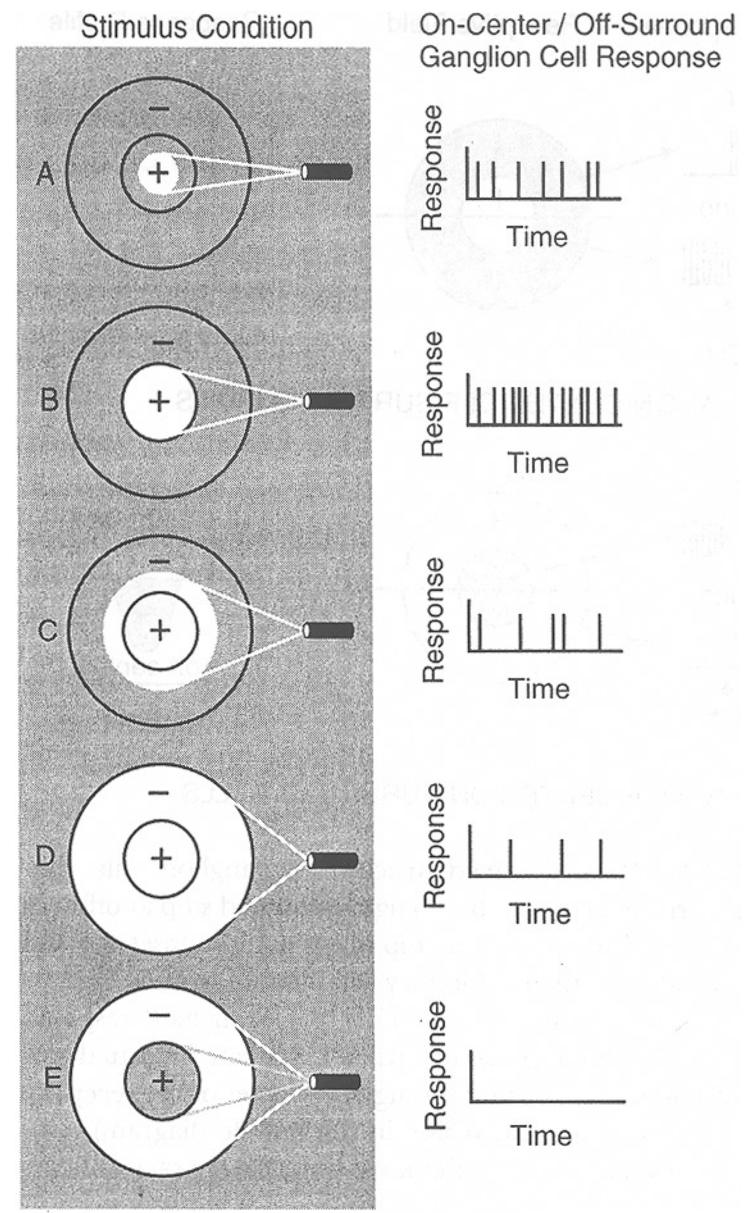


B. RECEPTIVE FIELD PROFILES

- The direct pathway from receptors to a given bipolar cell can be either excitatory or inhibitory, but the indirect pathway is always the opposite.

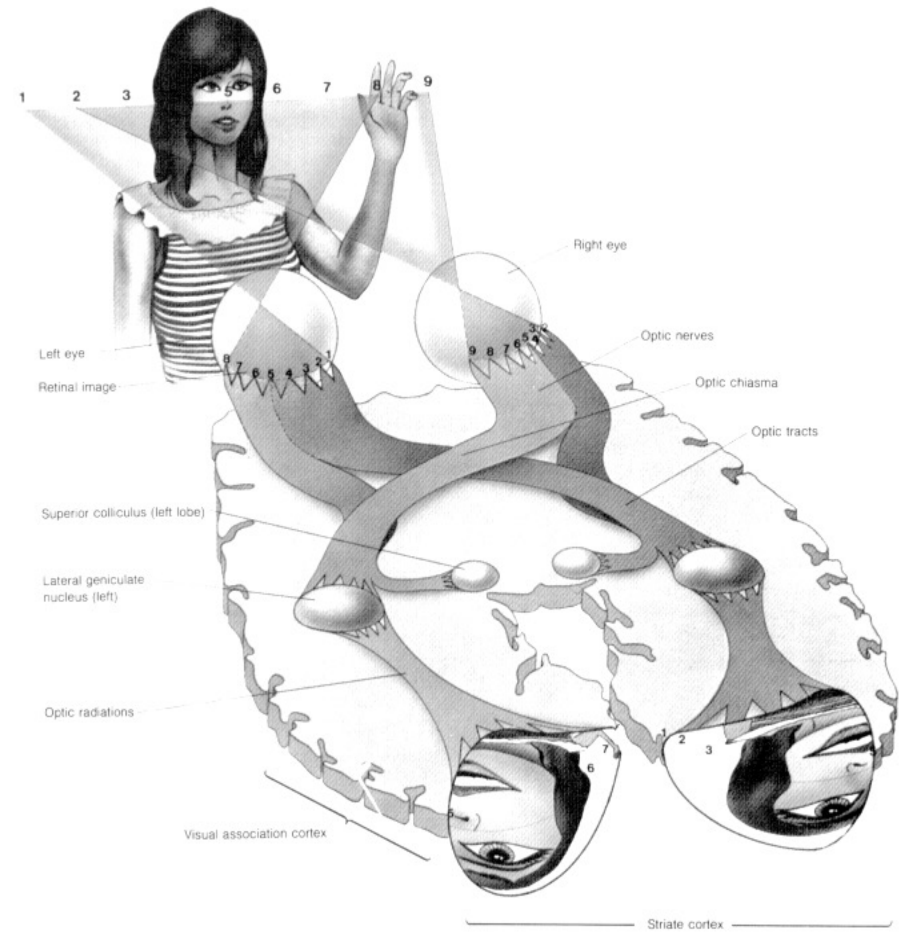
Ganglion cells (节细胞)

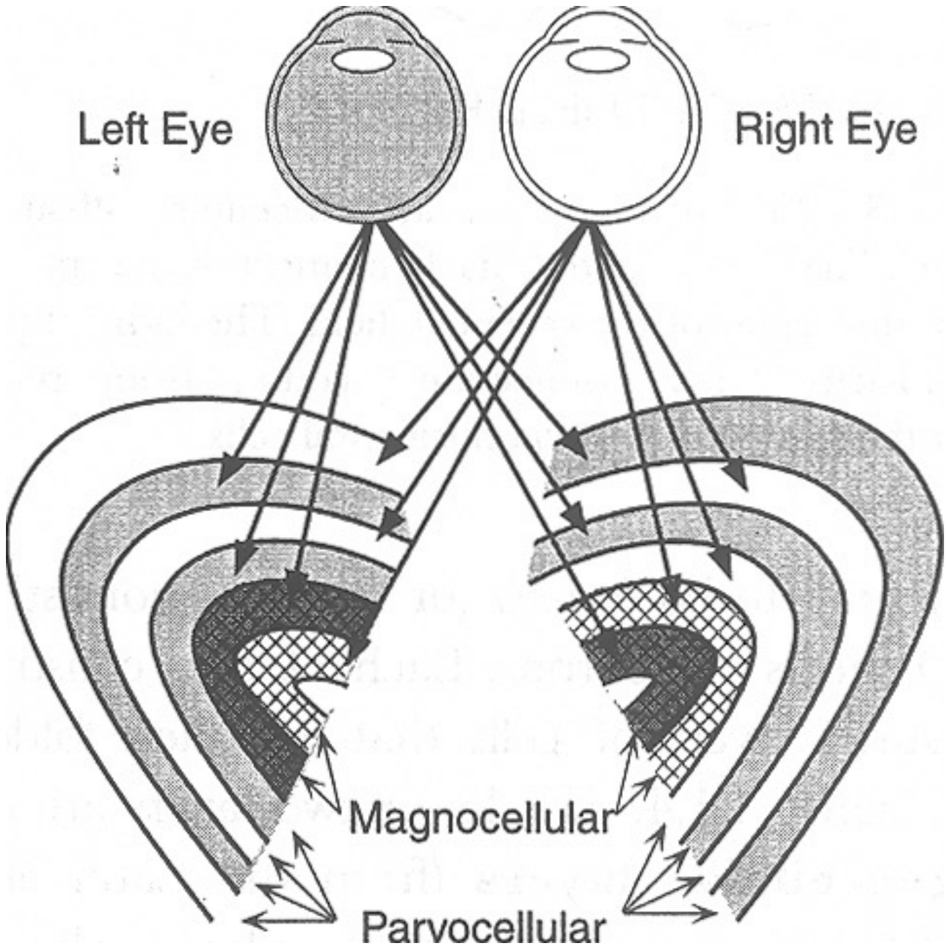
- Ganglion cells are the first cells in the visual system to produce spike discharges.
- Microelectrode extracellular recordings
- The firing rate was highest for a spot of light of a particular size at a particular position on the retina.
- If the size of the spot was either increased or decreased, the firing rate diminished.



Lateral Geniculate Nucleus

- LGN cells have center/ surround receptive fields but somewhat larger and with a stronger inhibitory surround.
- The LGN is a 3-D structure and receives input from both eyes.
- Each cell in LGN is monocular.





Each LGN is constructed of six distinct layers of cells.

- Magno and Parvo Cells
  - P cells receive input just from cones, and M cells receive input from both rods and cones.

	Magnocellular cells	**Parvocellular cells**
Contrast sensitivity	High	Low
Color sensitivity	Low	High
Spatial resolution	Low	High
Temporal resolution	Fast	Slow
Receptive field size	Large	Small
	motion and depth	color and shape

1.2 Striate Cortex

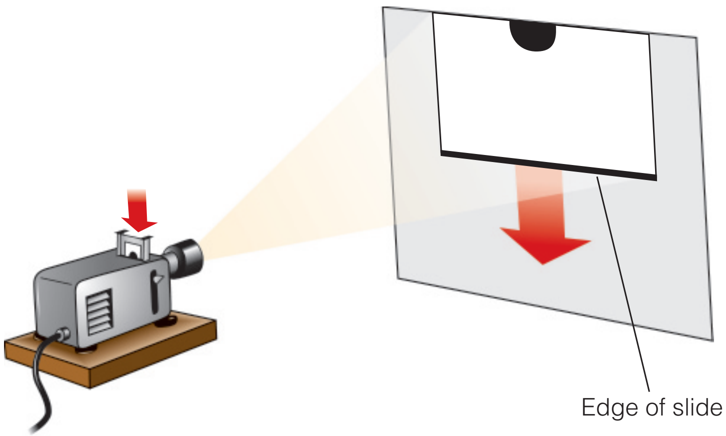
- Striate cortex (纹皮层)
  - The single largest cortical area in primates

Retinal ganglion cells	LGN cells	Striate cortex
1 million	1.5 million	200 million

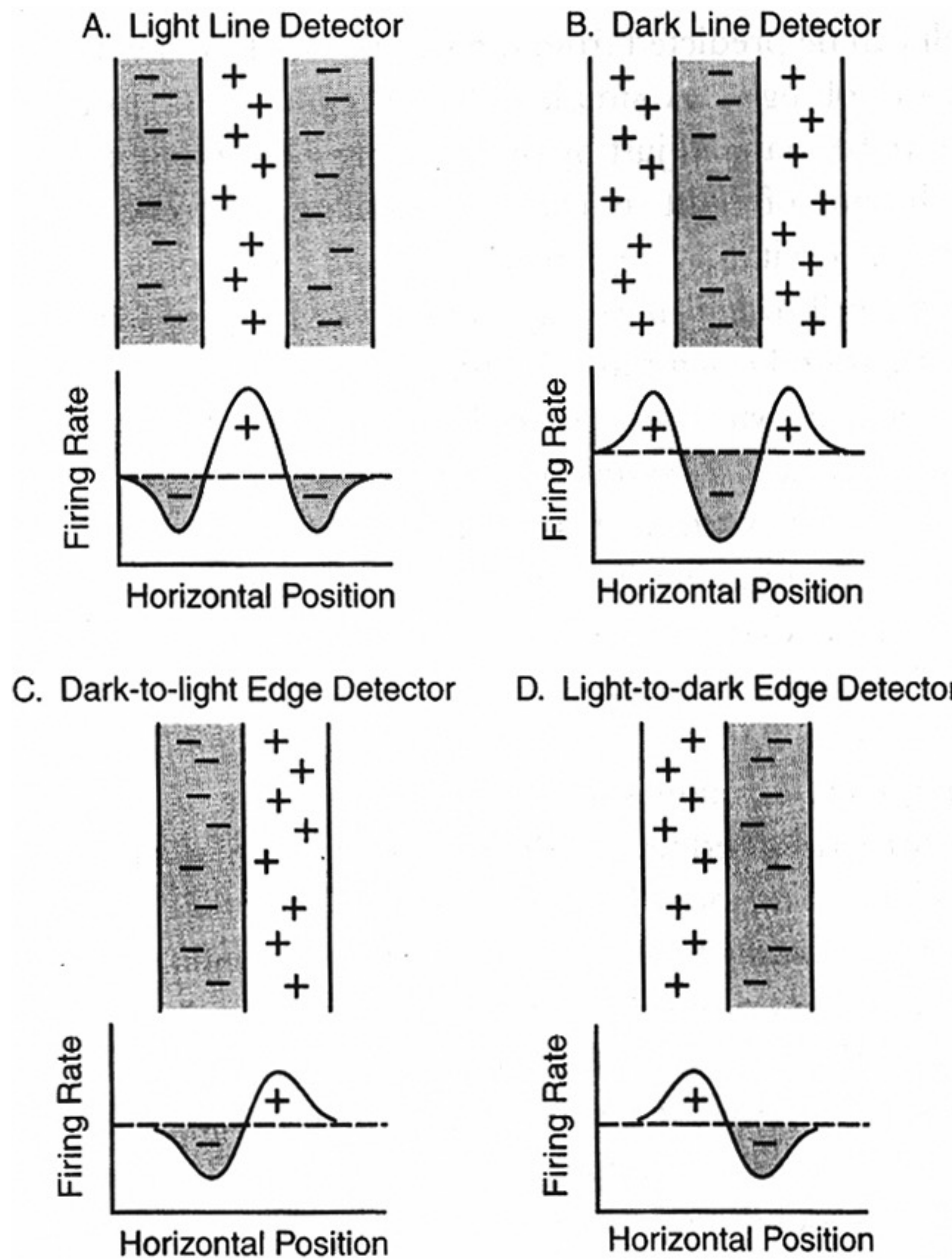
- Hubel and Wiesel's Discovery
  - First to successfully apply the receptive field mapping techniques to striate cortex.
  - Others had tried before them but were unable to find stimuli that made the cells fire.

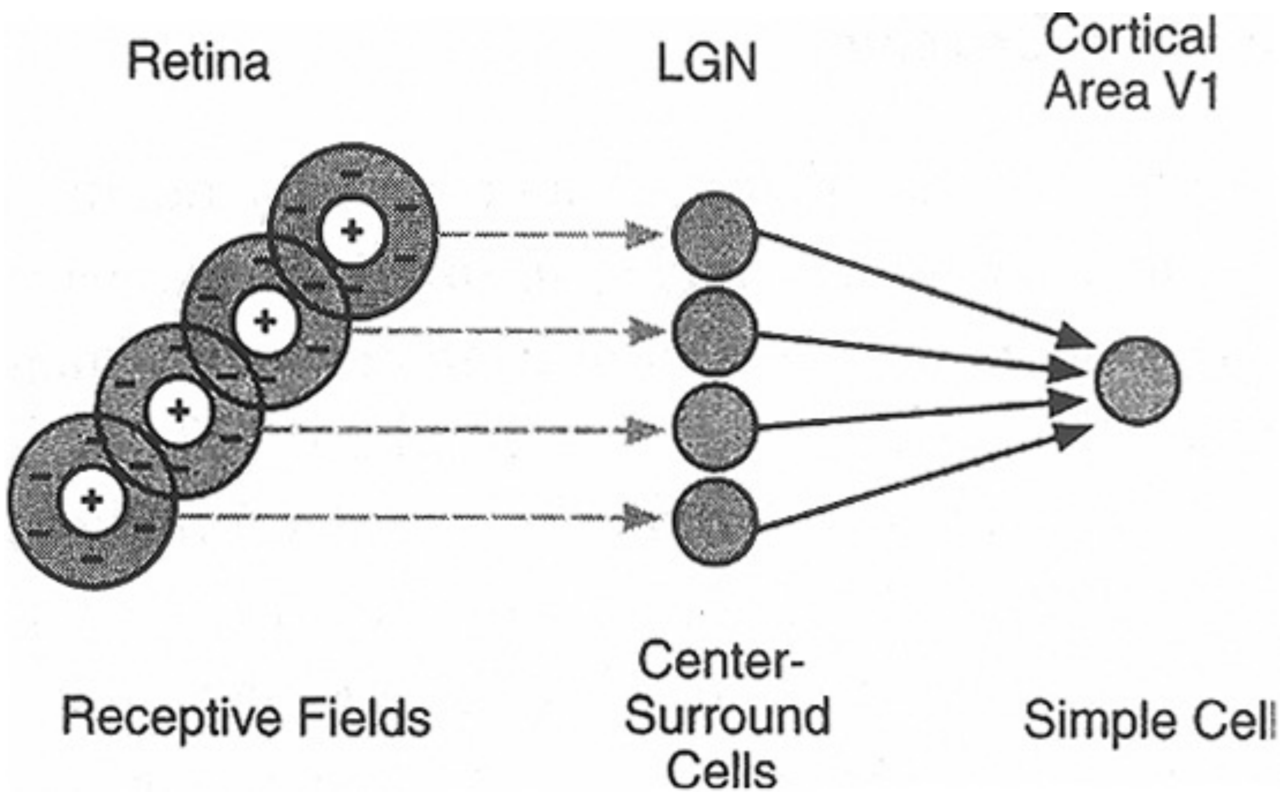
Hubel and Wiesel's experiment

- They realized that the cell's response had nothing to do with the spot.
- It was actually caused by the shadow of the slide's edge moving in a particular direction across the retina in a particular position.
- The orientation and direction of movement of this edge were indeed the crucial variables.
- Simple cells, complex cells, and hypercomplex cells



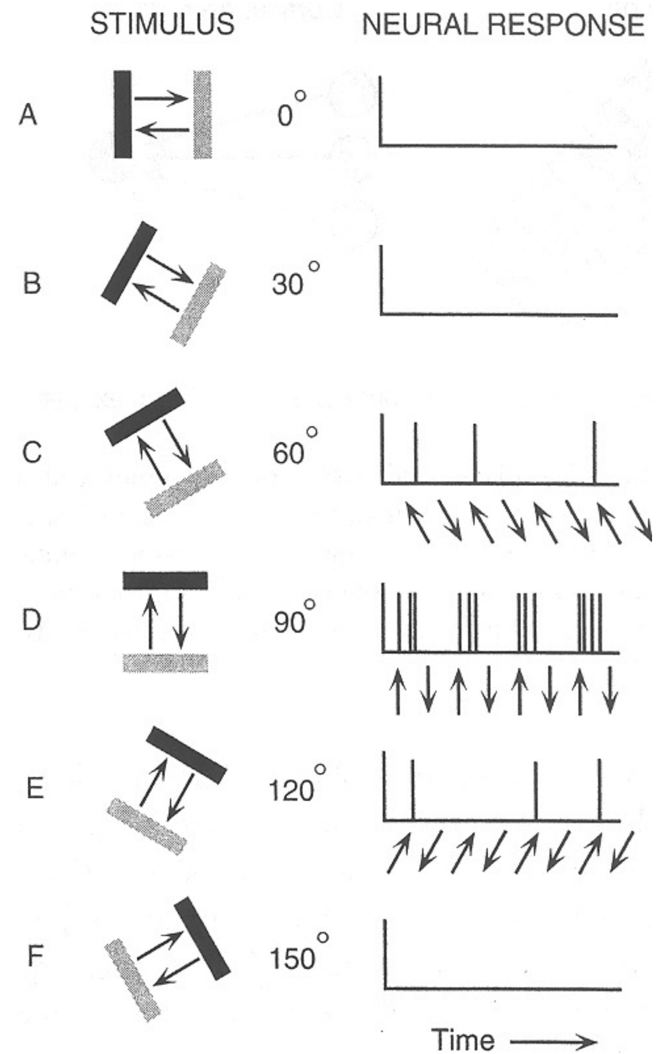
- Simple cells
  - The response to a larger, more complex pattern of stimulation can be roughly predicted by summing its responses to the set of small spots of light that compose it.
  - Several subtypes



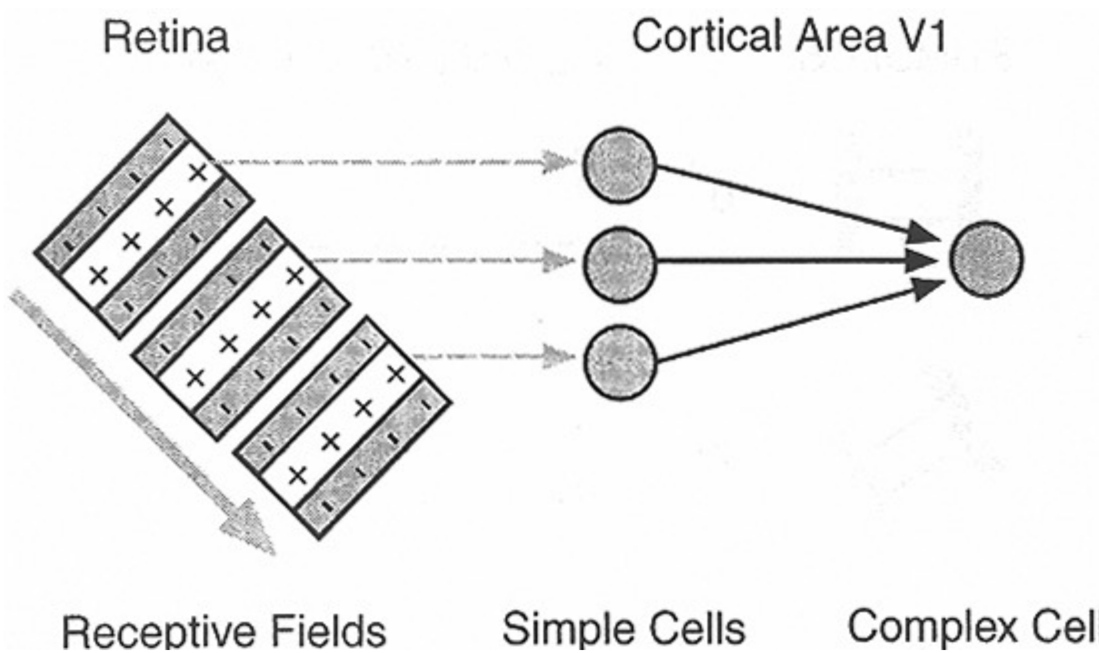


Possible neural wiring of cortical simple cells from LGNS

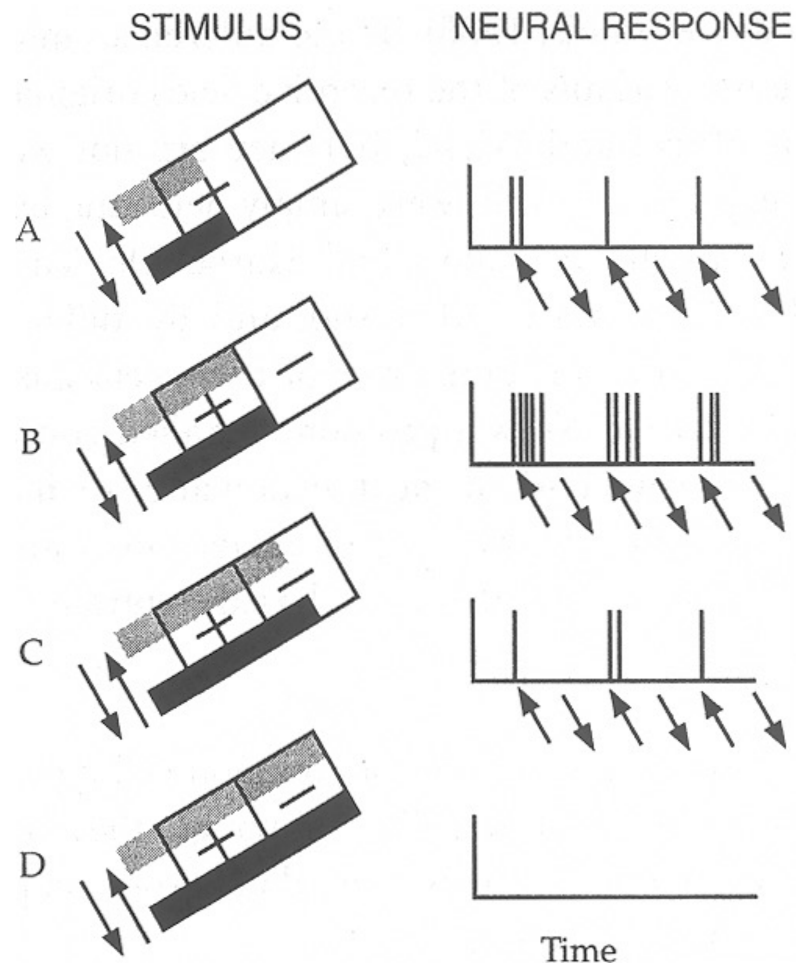
- Complex cells
  - The most common cell type in striate cortex (75%)
  - Differ from simple cells in several important respects
    - Nonlinearity: cannot be mapped by their response to individual spots in each retinal location.
    - Motion sensitivity: to be highly responsive to moving lines or edges. The motion sensitivity is specific to a particular direction of movement.
    - Position insensitivity:
    - Spatial extension: larger receptive fields



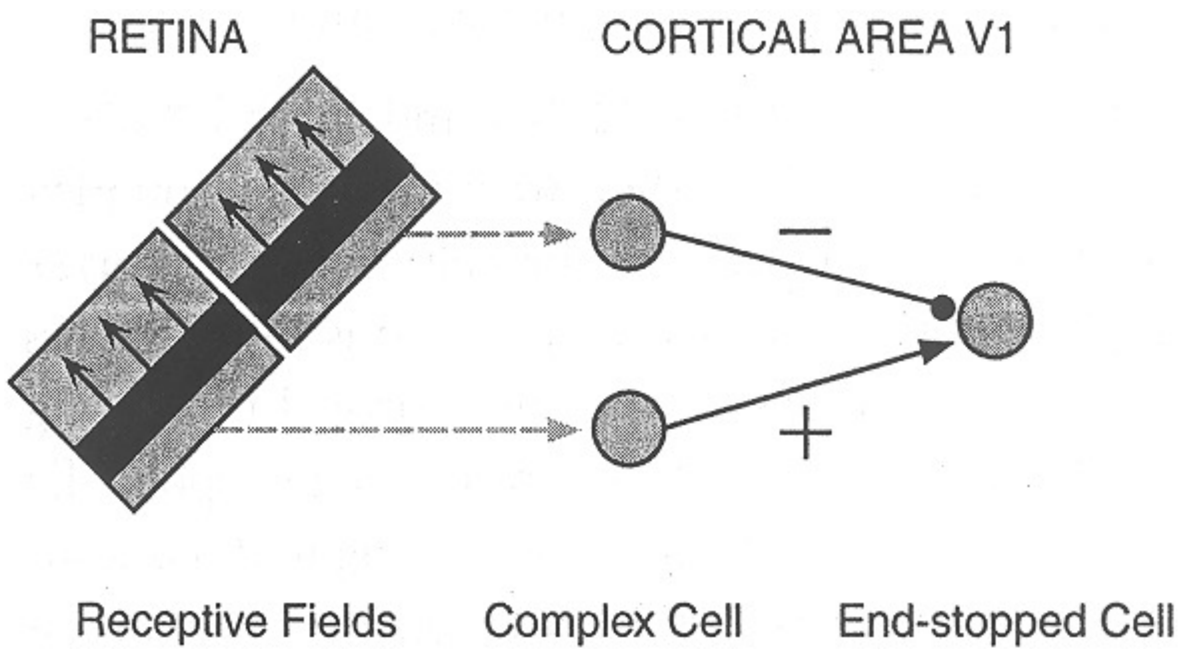
- Hulel and Wiesel proposed that complex cells were constructed by integrating the responses of many simple cells.



- Hypercomplex cells
  - Have more selective receptive fields than complex cells.
  - Extending a line or edge beyond a certain length causes them to fire less vigorously than they do to a shorter line or edge.(end-stopped cells)

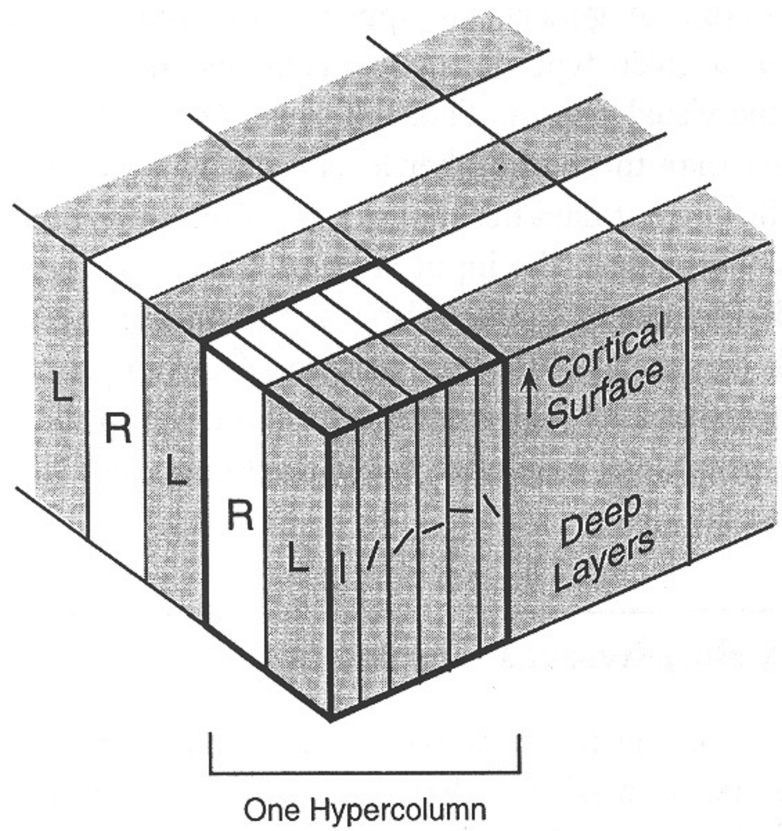


- Recent studies:
  - The degree of “end-stopping” is a continuum rather than an all-or-none phenomenon.
  - It may arise via lateral inhibition.



Striate Architecture

- The retinotopic map
  - The areas adjacent to each other on the retina are also adjacent in the cortex.
- Columnar structure
  - Run perpendicularly through all six cortical layers.
  - There is a regular progression along one dimension that represents the tuning cells for orientation.
  - Another functional dimension of each hypercolumn perpendicular to the orientation dimension that represents a regular progression of size-scale values.

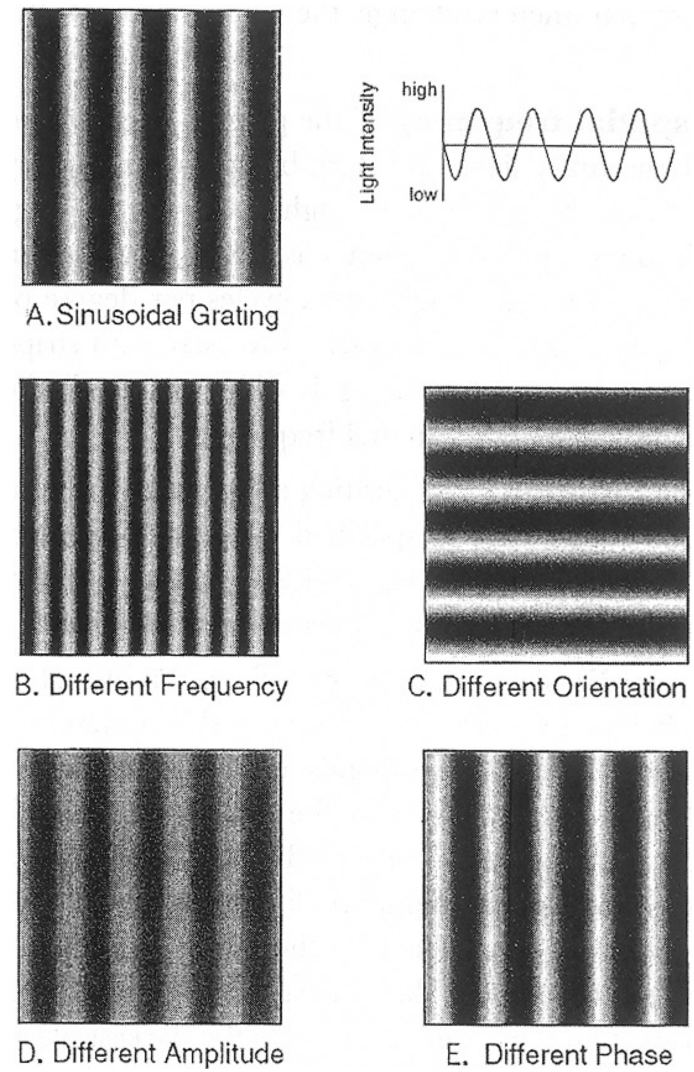


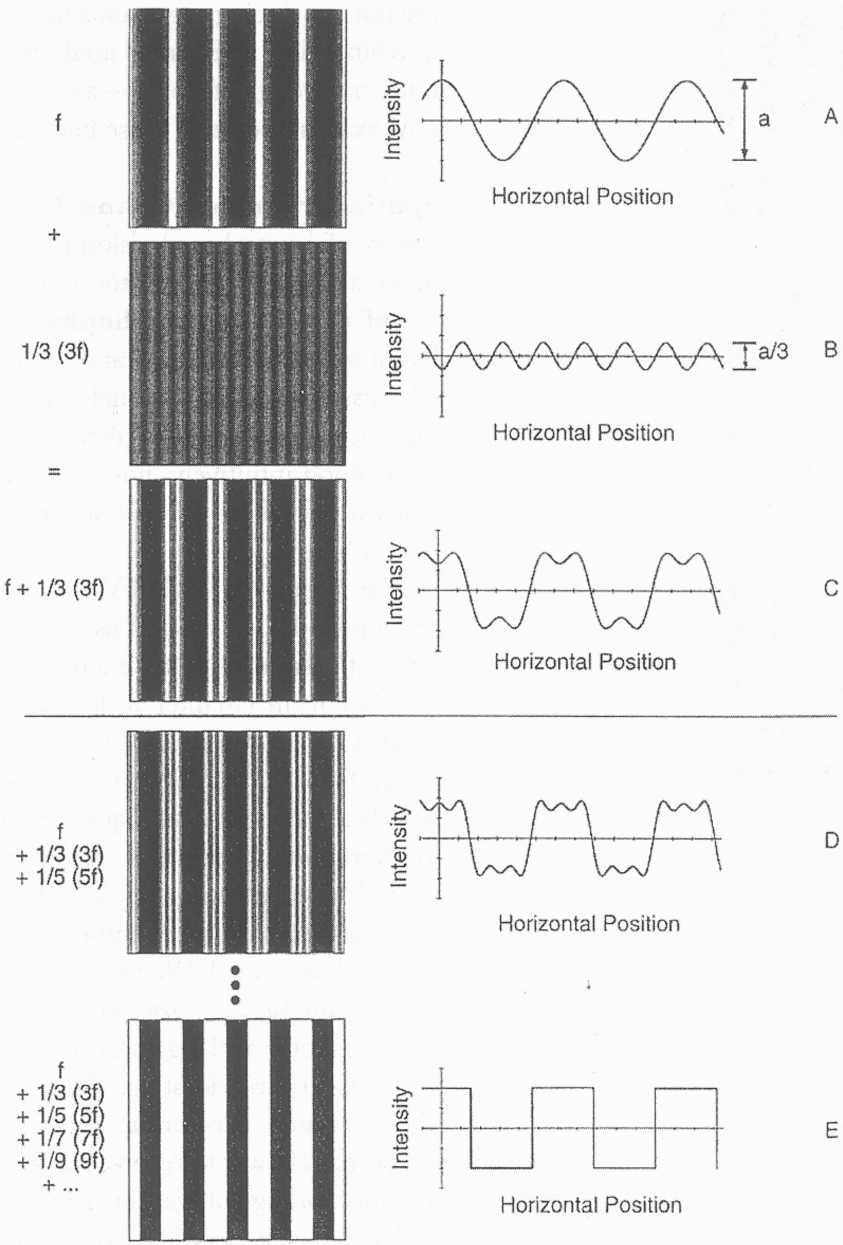
2. Psychophysical Mechanisms

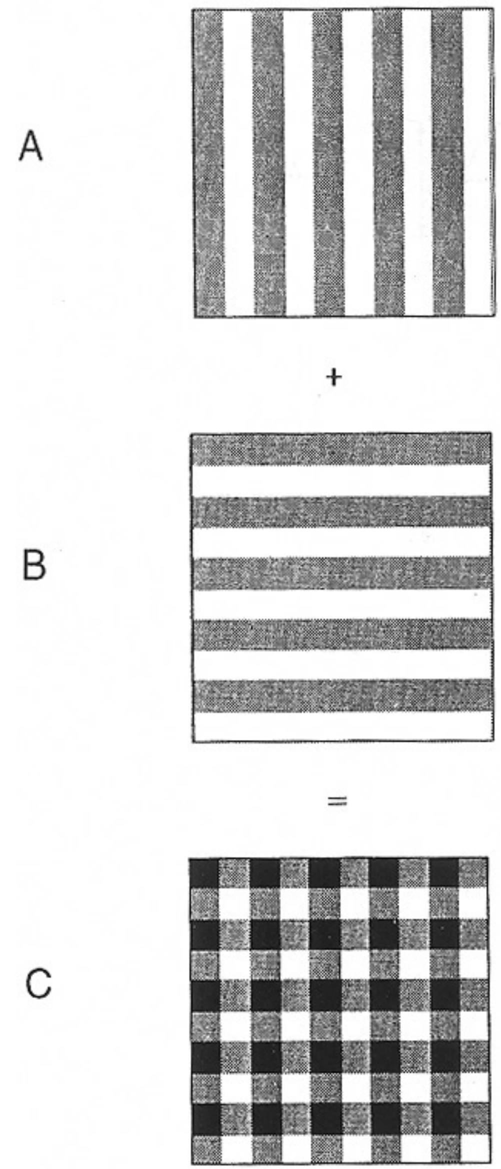
- Psychophysics is the study of quantitative relations between people's conscious experiences and properties of the physical world using behavioral methods.
- They measure people's performance in specific perceptual tasks and try to infer something about underlying mechanisms from behavioral measurements.
- These mechanisms must ultimately be implemented physiologically.

2.1 Spatial Frequency Theory

- The primitives of spatial frequency theory are sinusoidal gratings.
  - 2-D patterns whose luminance varies according to a sine wave over one spatial dimension
  - constant over the perpendicular dimension.



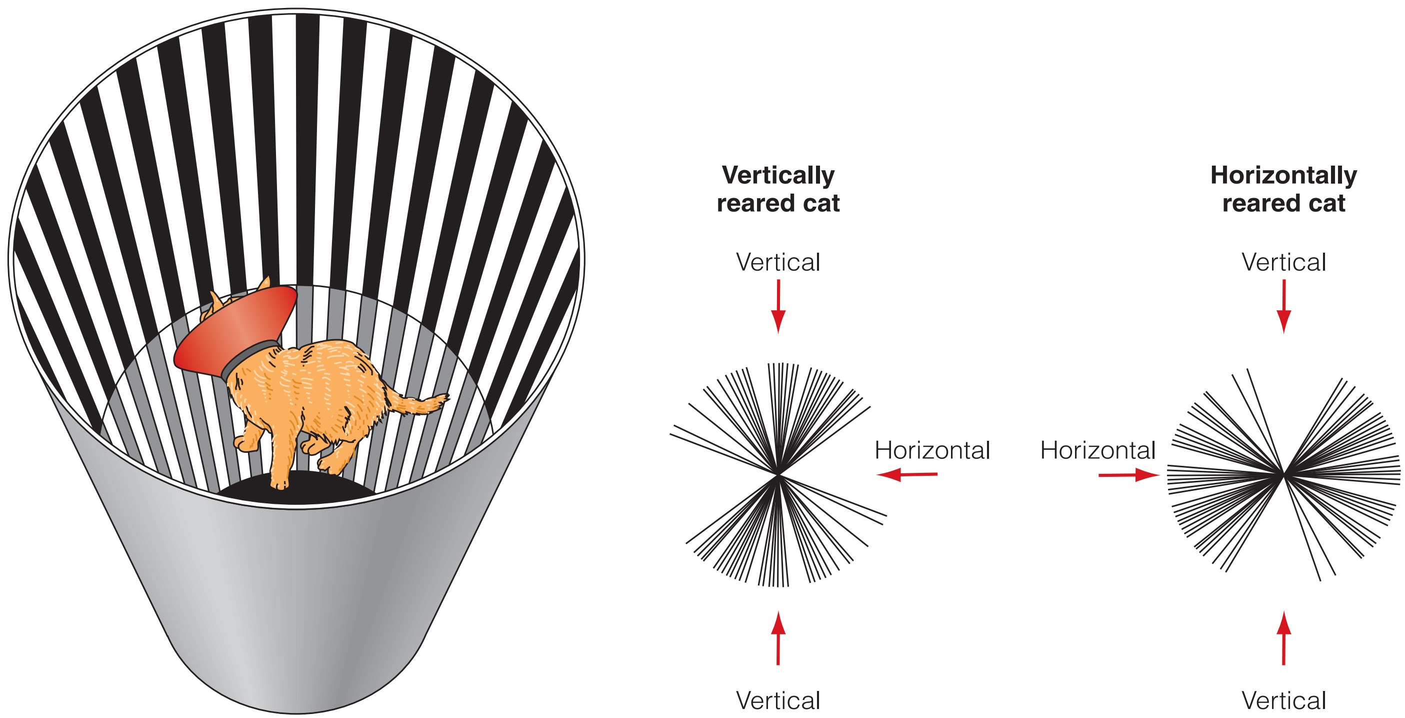

f
+
1/3 (3f)
=
f + 1/3 (3f)
f
+ 1/3 (3f)
+ 1/5 (5f)
f
+ 1/3 (3f)
+ 1/5 (5f)
+ 1/7 (7f)
+ 1/9 (9f)
+ ...
Intensity
Horizontal Position
a
a/3
A
B
C
D
E


A
+
B
=
C

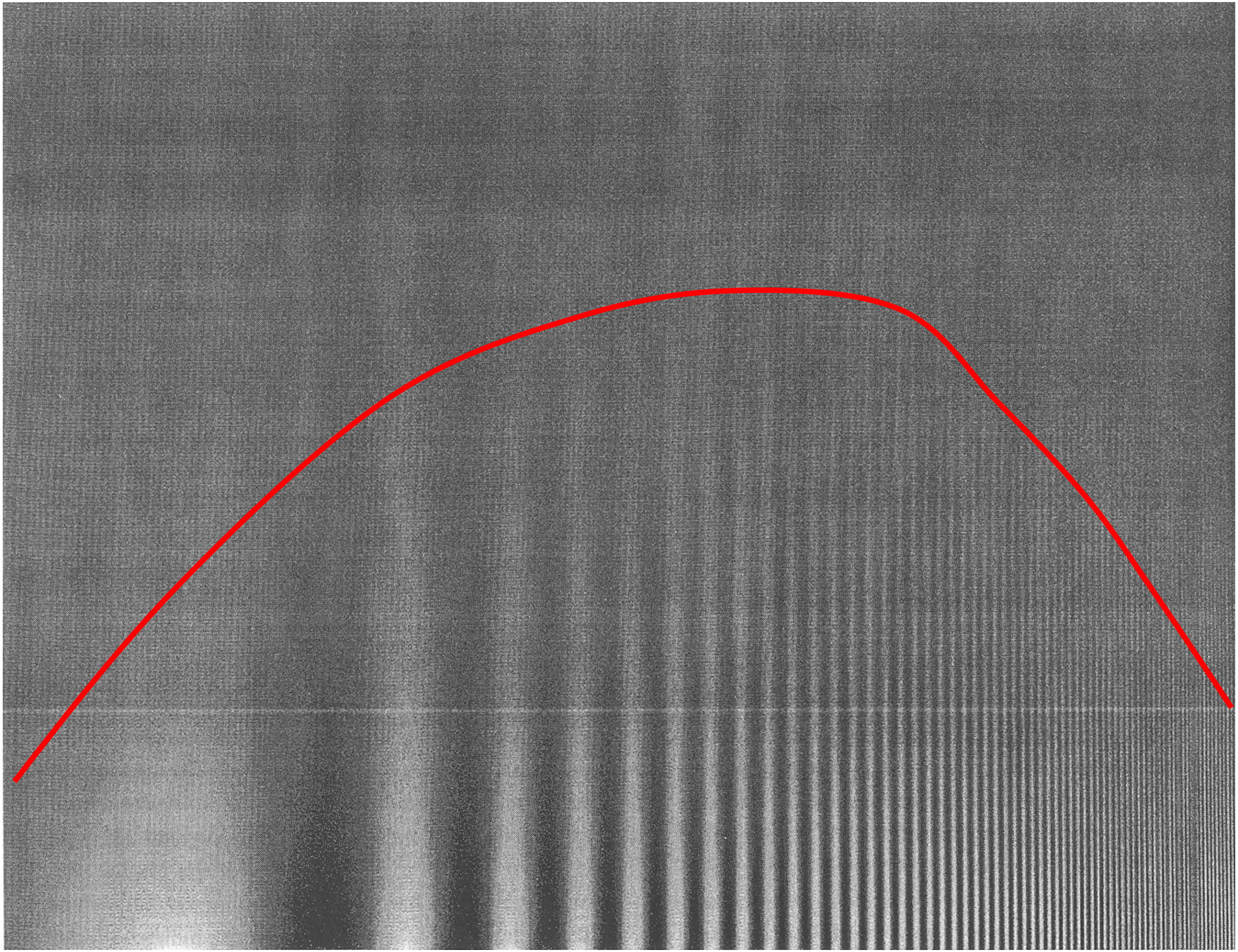
- Fourier analysis is not limited to these simple, regularly repeating patterns.
- We can demonstrate what kind of spatial information is carried by different ranges of spatial frequencies.

- Spatial Frequency Channels
  - Asserts that the visual system can be understood as consisting of many overlapping channels that are selectively tuned to different ranges of spatial frequencies and orientations, just like the channels in TV set.
  - Each channel is defined by the spatial frequency and orientation of the gratings to which it is maximally sensitive.
  - There is now a great deal of evidence to support this view.

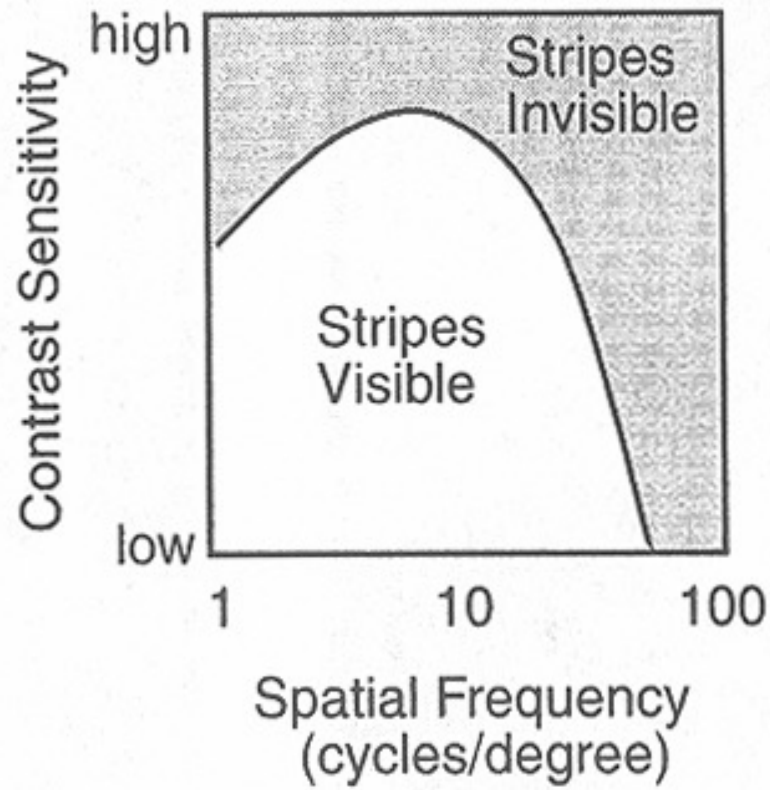
Striped Tube Experiment



(a) (b)

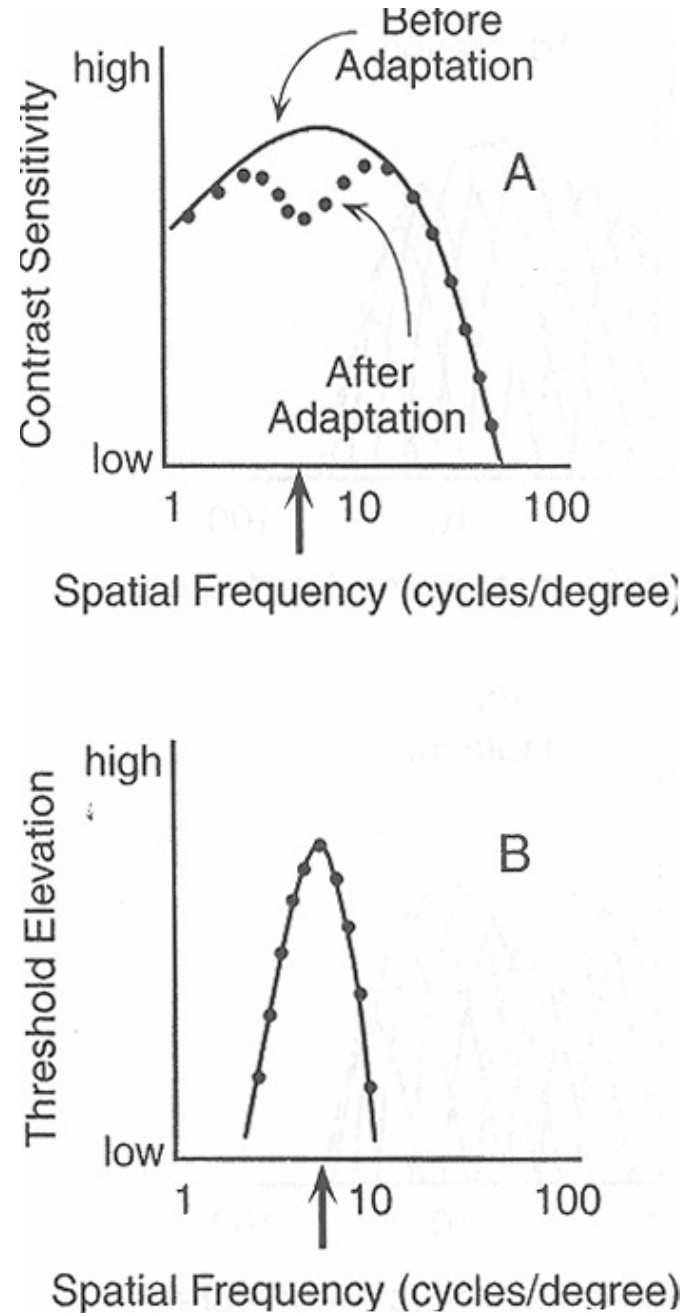


- Contrast Sensitivity Functions
  - Method: finding the lowest contrast at which the observer can just barely detect the difference between a sinusoidal grating and a uniform gray field.
  - Most sensitive at 4-5 cycles per degree of visual angle.
  - To determine the effects of adapting an observer to a particular spatial frequency grating by measuring their sensitivity to such gratings both before and after adaptation.

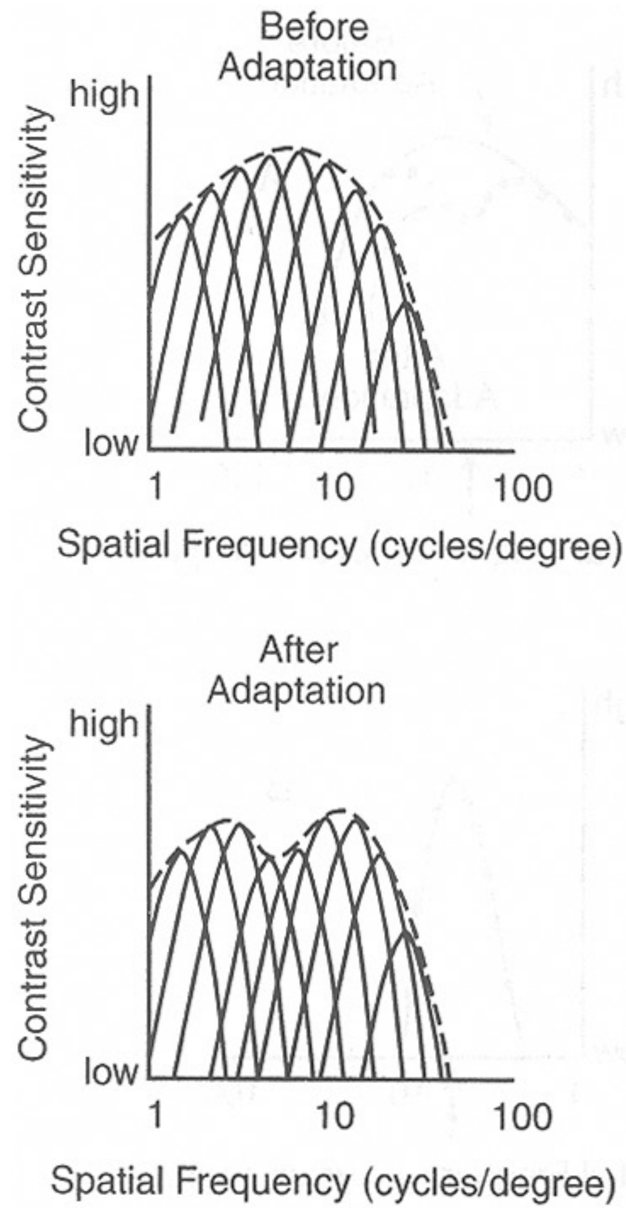


B. SENSITIVITY

- Selective Adaptation of Channels
  - The extended exposure to the grating caused the subject's visual system to adapt, that is to become less sensitive after the prolonged viewing experience, but only near the particular spatial frequency and orientation of the adapting grating.



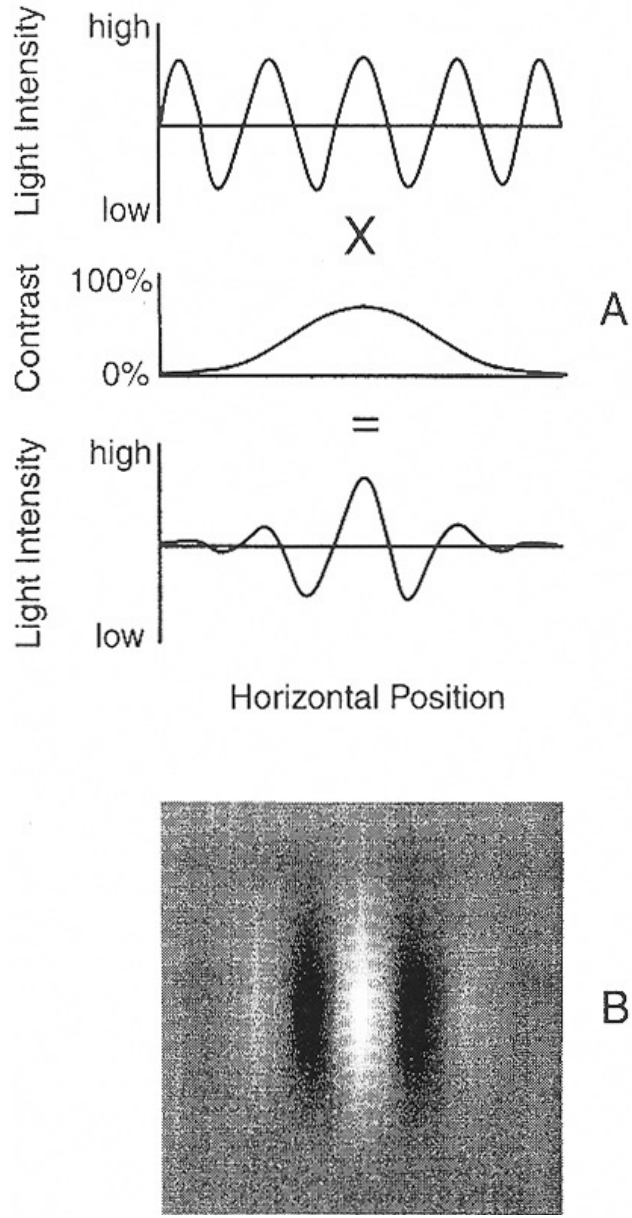
- Explanation
  - The theory states that the broad-band CSF that was originally measured actually represents the combined contribution of many overlapping narrow-band channels, each of which is sensitive to a different range of spatial frequencies.
  - After an extended period, the channel get tired and respond less vigorously.



2.2 Physiology of Spatial Frequency Channels

- Psychophysical channels are information processing constructs at Marr's algorithmic level of description rather than at his implementation level.
- They must be implemented somewhere in the visual nervous system.
- How? And where?

- The second theory about the function of the cells in striate cortex: these cells may be performing a **local spatial frequency analysis**.
- Gabor function: a local, piecewise, spatial frequency analysis can be accomplished through many small patches of sinusoidal gratings that “fade out” with distance from the center of the receptive field.



3. Computational Approaches

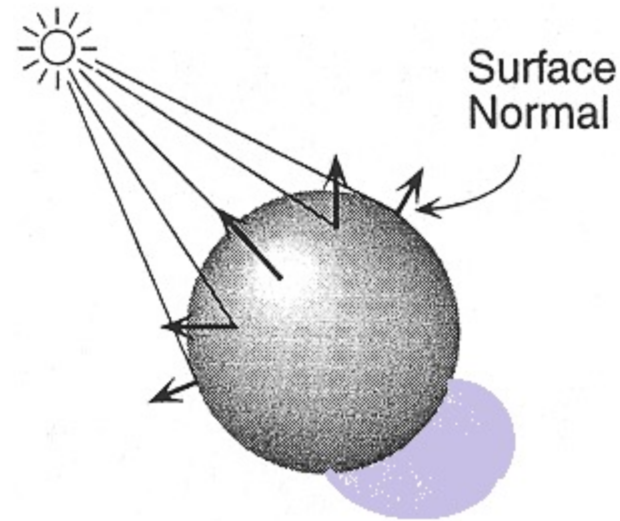
- The problem: how to detect naturalistic image features such as edges and lines in gray-scale images.
- Vision group at MIT
  - Line and edge detector theory

3.1 Marr’s Primal Sketches (要素图/简图)

- The raw primal sketch
  - The transition from the fully analog, gray-scale image to a largely symbolic representation of image-based features.
  - Edge, line, blob, termination
  - Parameters: position, size, orientation, contrast, direction of motion, speed of motion, and color contrast
- The full primal sketch
  - Perceptual organization at the image level
    - To link short line segments and edges together into longer ones
    - To group similar elements together
    - To characterize the texture of areas

Edge Definition

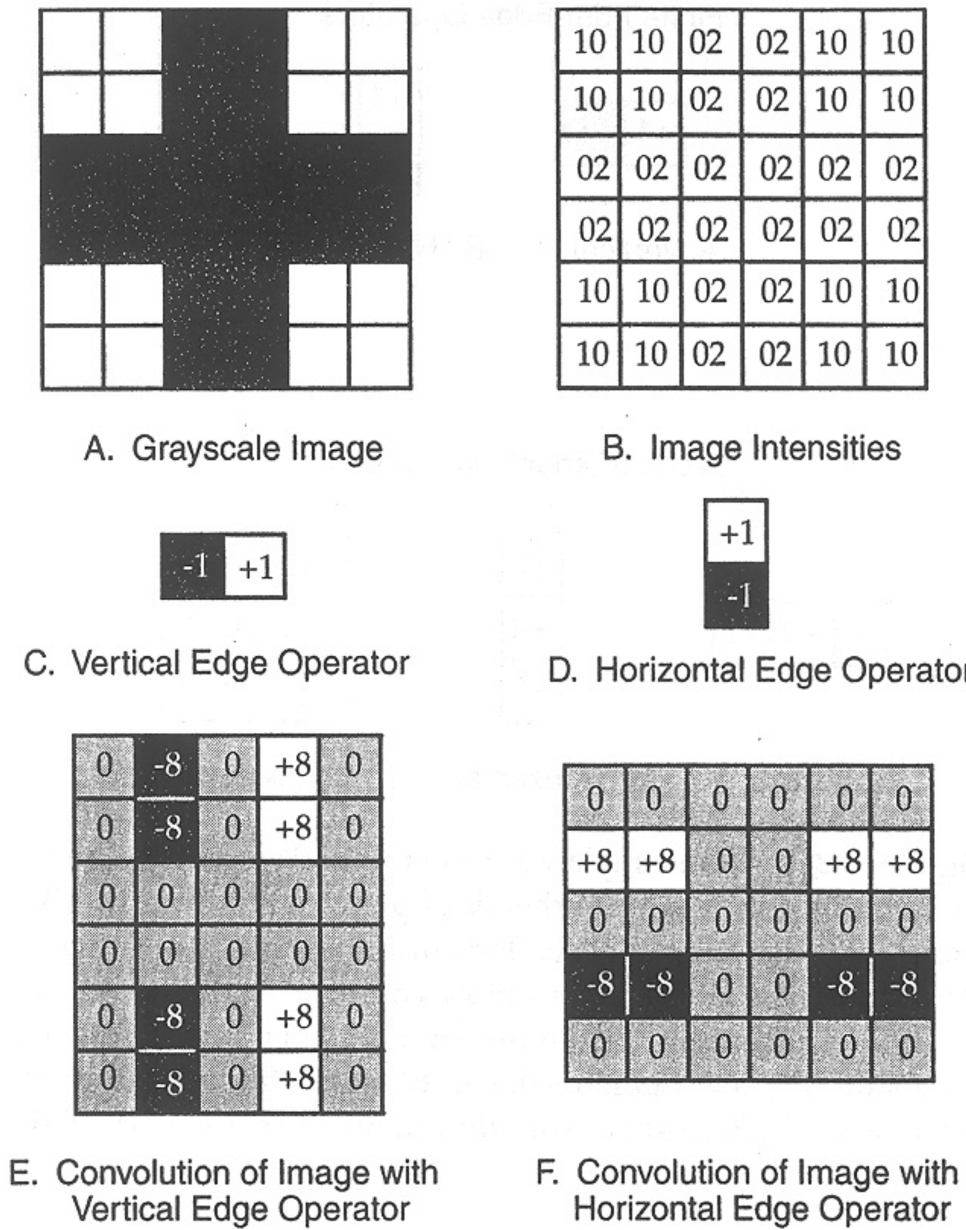
- The changes in luminance that take place along a uniformly colored smooth surface tend to be **gradual**.



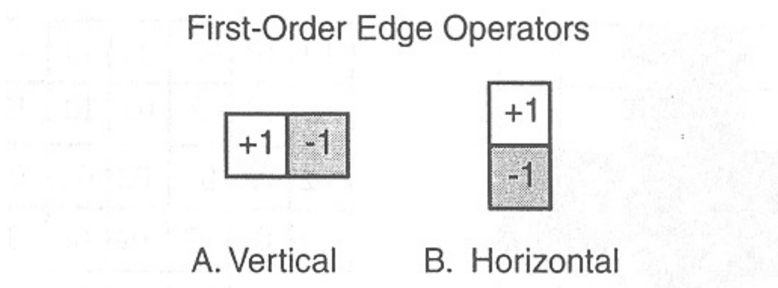
- The changes that take place across a transition from one surface to another are generally much more **abrupt**, and form luminance edges.
  - A change in the reflectance of the surface (**material**)
  - A change in the amount of light falling on it (**shadow**)
  - A change in surface orientation relative to the light source (**angle**)

Edge Operators and Convolution

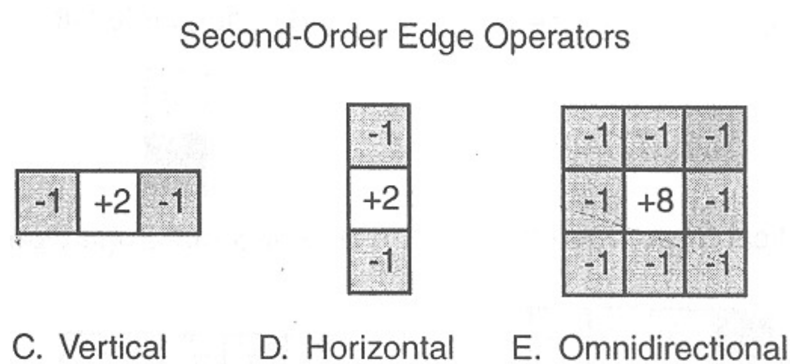
- Edges are detected by performing the **convolution** of an edge operator with an image.
- Large positive or negative values indicate the presence of dark-to-light or light-to-dark luminance edges.



- First-order differential operators
  - Take the simple difference between adjacent pixels.



- Second-order differential operators
  - Compute the difference between adjacent first-order operators.



A. Grayscale Image

10	10	02	02	10	10
10	10	02	02	10	10
02	02	02	02	02	02
02	02	02	02	02	02
10	10	02	02	10	10
10	10	02	02	10	10

B. Image Intensities

-1	+2	-1

C. Vertical

-1
+2
-1

D. Horizontal

Second-Order Edge Operators

-1	-1	-1
-1	+8	-1
-1	-1	-1

E. Omnidirectional

+8	–8	–8	+8
+8	–8	–8	+8
0	0	0	0
0	0	0	0
+8	–8	–8	+8
+8	–8	–8	+8

F. Convolution of Image with Vertical Edge Operator

+8	+8	0	0	+8	+8
–8	–8	0	0	–8	–8
–8	–8	0	0	–8	–8
+8	+8	0	0	+8	+8

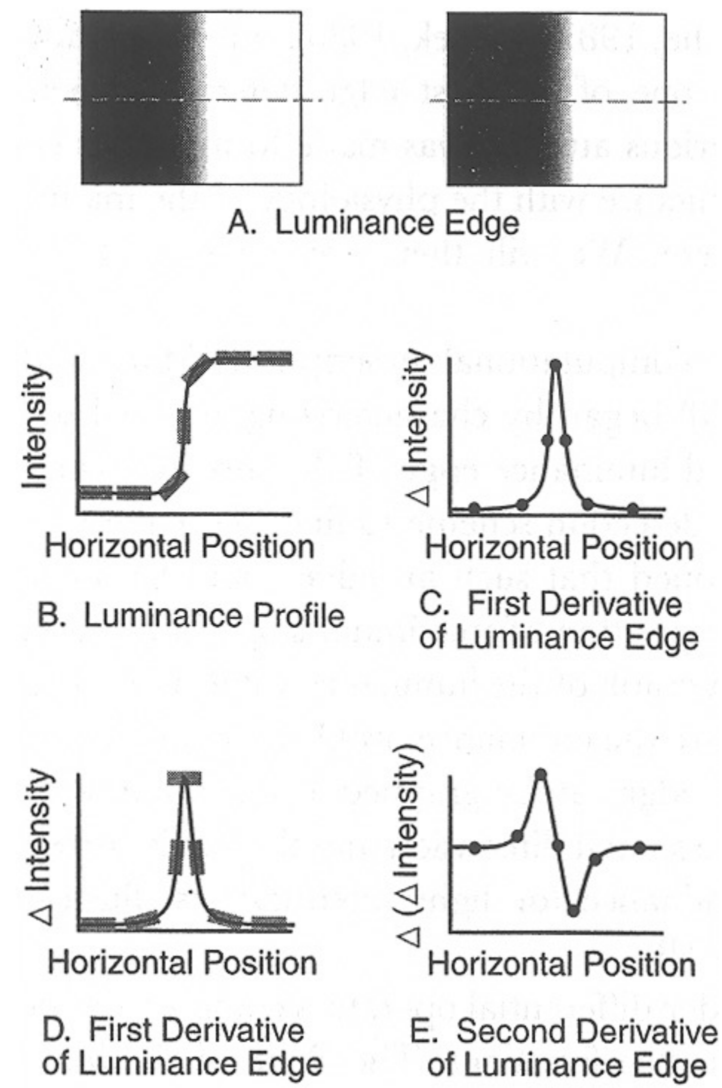
G. Convolution of Image with Horizontal Edge Operator

40	-16	-16	40
-16	–8	–8	-16
-16	–8	–8	-16
40	-16	-16	40

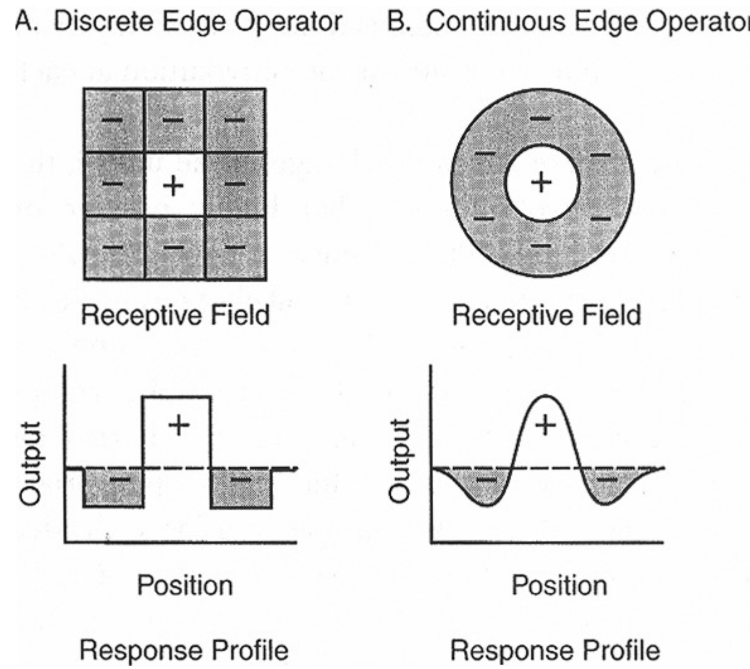
H. Convolution of Image with Omnidirectional Edge Operator

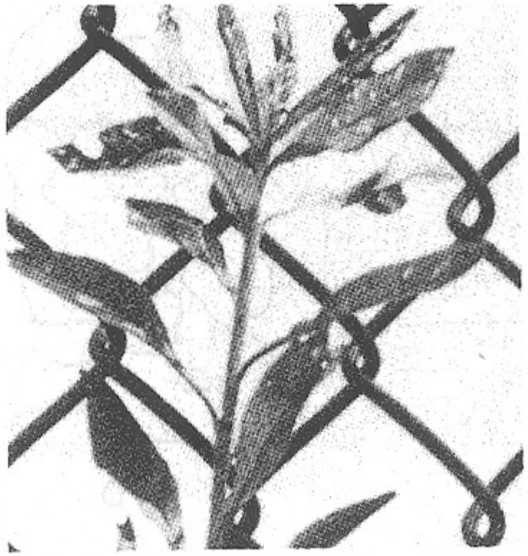
The Marr-Hildreth Zero-Crossing algorithm

- To detect edges by the first derivative
- To detect zero-crossings of the second derivative, the position of the luminance edge corresponds to the zero value in between a highly positive and highly negative value.
- Criticizing: First-order edge operators must be computed at every possible orientation, and this process was computationally wasteful.

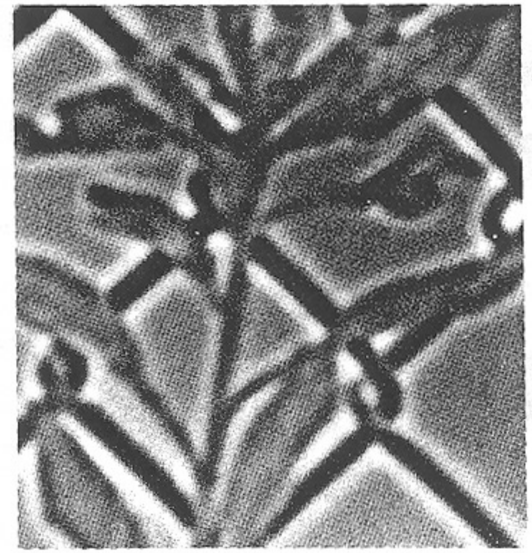


- The computational advantage of the zero-crossing of the second derivative
  - The second-order operators are symmetrical about their midpoint, the second derivative of a 2-D image can be computed in all orientations simultaneously by the single 2-D operator.
  - It is essentially the simultaneous combination of second-order operators at vertical, horizontal, left-diagonal, and right-diagonal orientations.





A

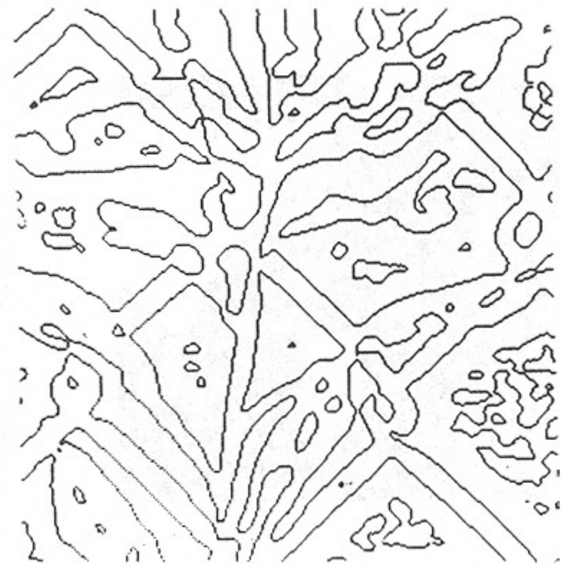


B

where zero is represented as neutral gray, positive values as lighter gray and negative values as darker gray

C

positive values as white and negative values as black

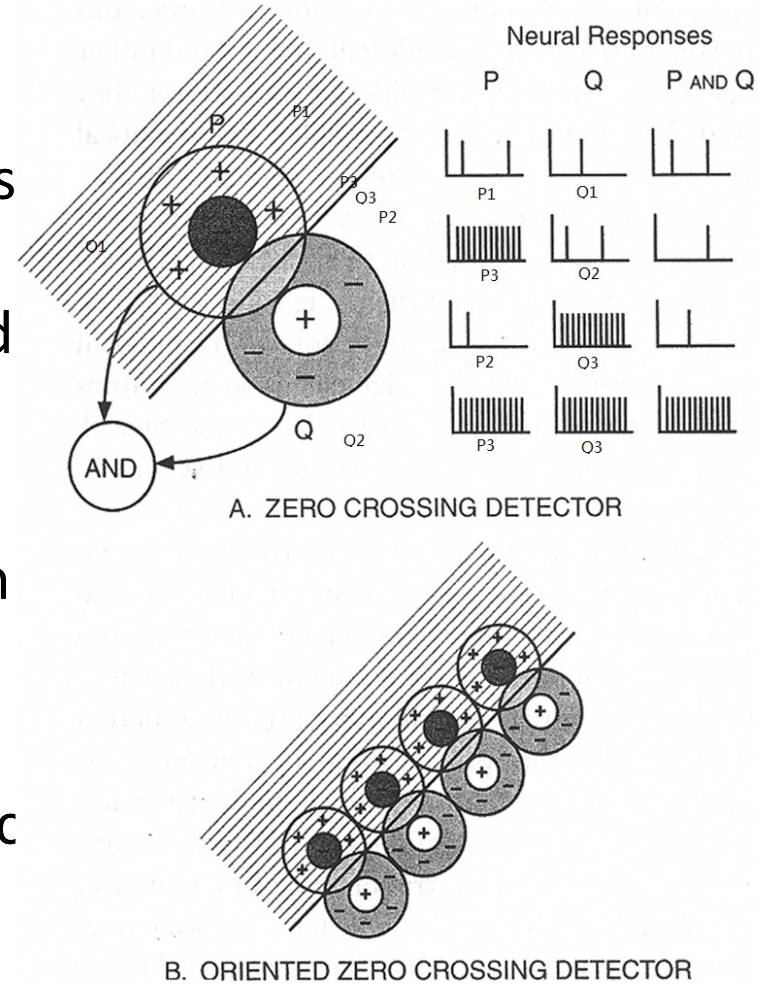


D

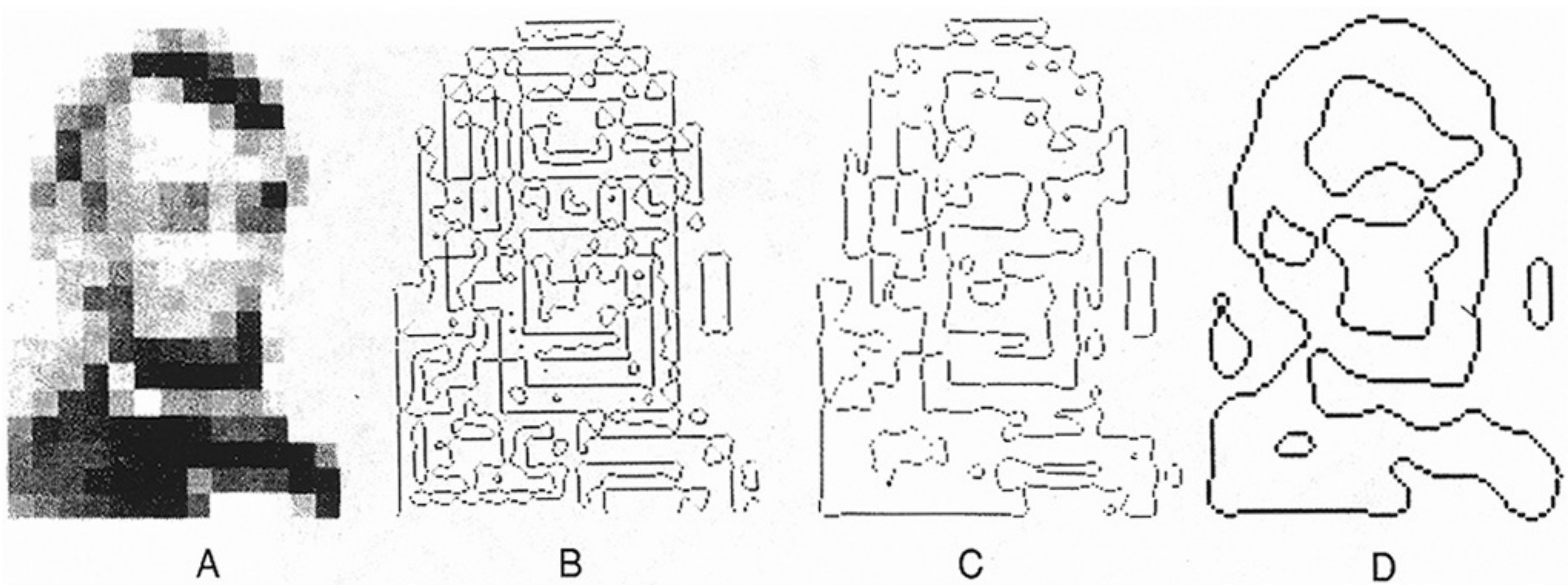
The zero-crossing in B

Neural Implementation

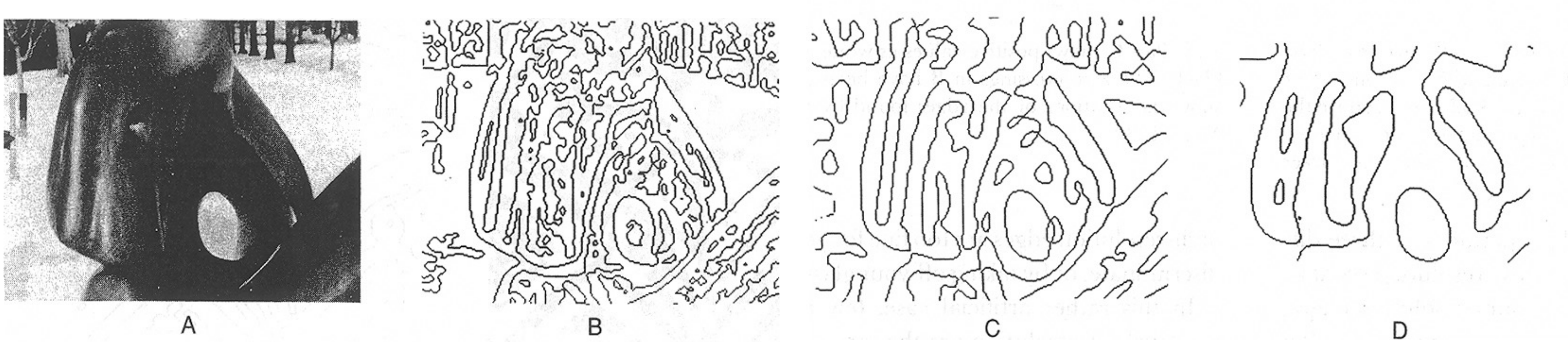
- The convolution of the image with the second-order edge operator was accomplished by ganglion and LGN cells whose receptive fields have a rotationally symmetric center/surround organization.
- Zero-crossings can be accomplished by constructing by a logical AND operation on adjacent on-center and off-center geniculate cells.
- Edge detectors sharply tuned to specific orientations can be constructed by combining the output of many aligned zero-crossing detectors.



- Second-order edge operators of three different sizes
  - Small ones to detect high-resolution edges.
  - Medium ones to detect medium-resolution edges
  - Large ones to detect low-resolution edges

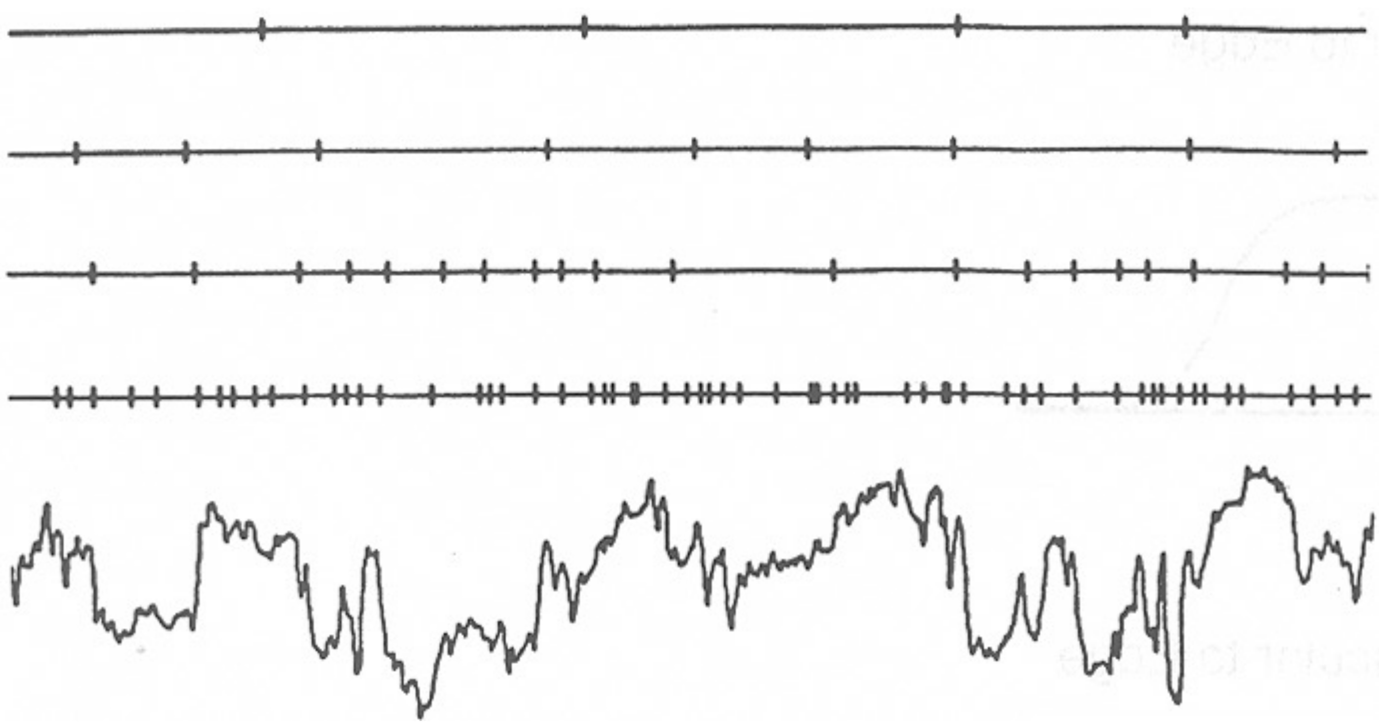


- In natural images, most real edges at low levels of resolution will also be found in the same general location at higher levels of resolution.
- Finding edges at multiple levels of resolution in the same location increases the likelihood that a real edge is present.

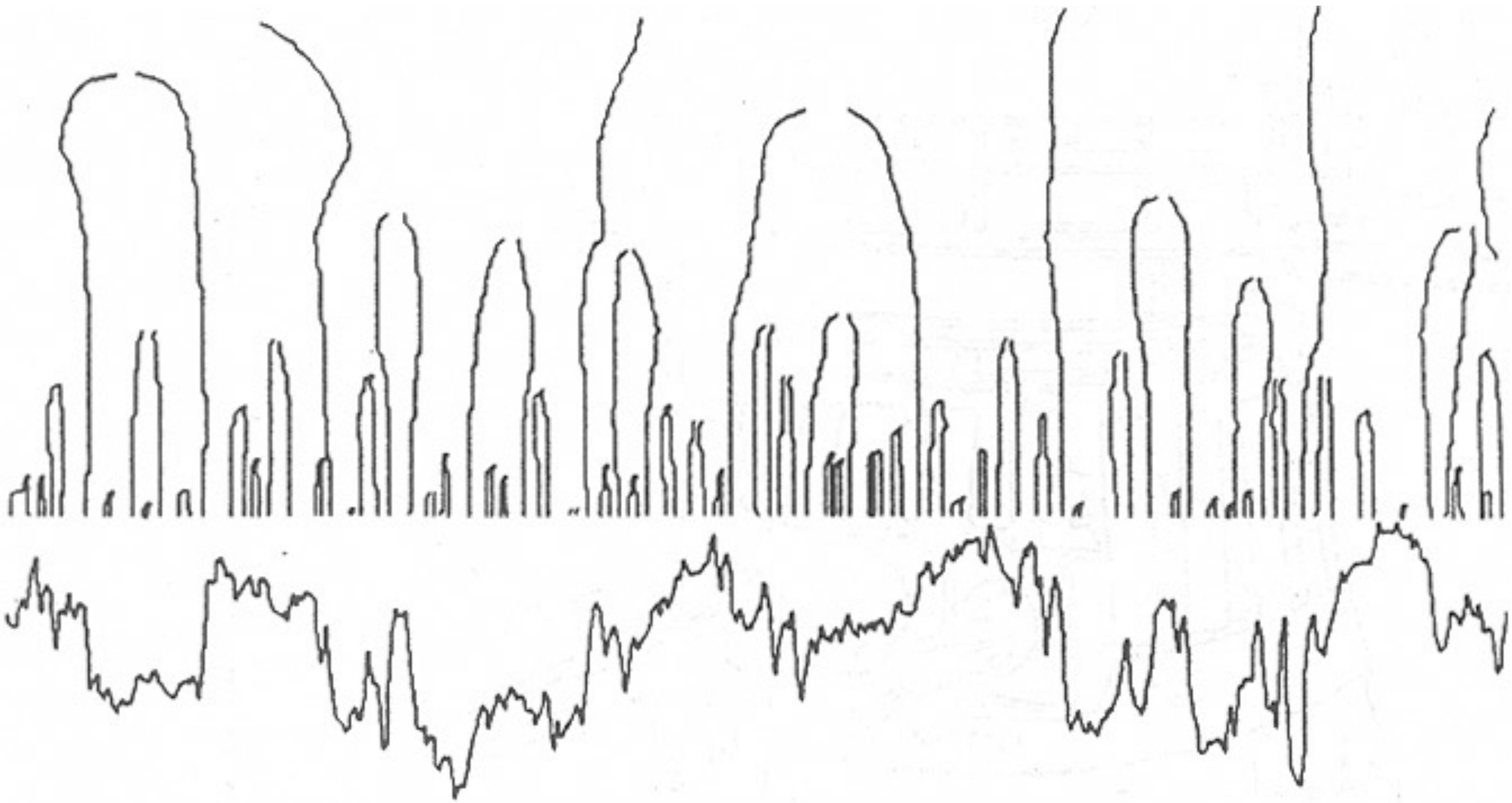


Witkin’s Scale Integration

- The visual system needs to determine how the different edges at different scales match up:
  - Which ones are due to the same edge?
  - And which one are not?

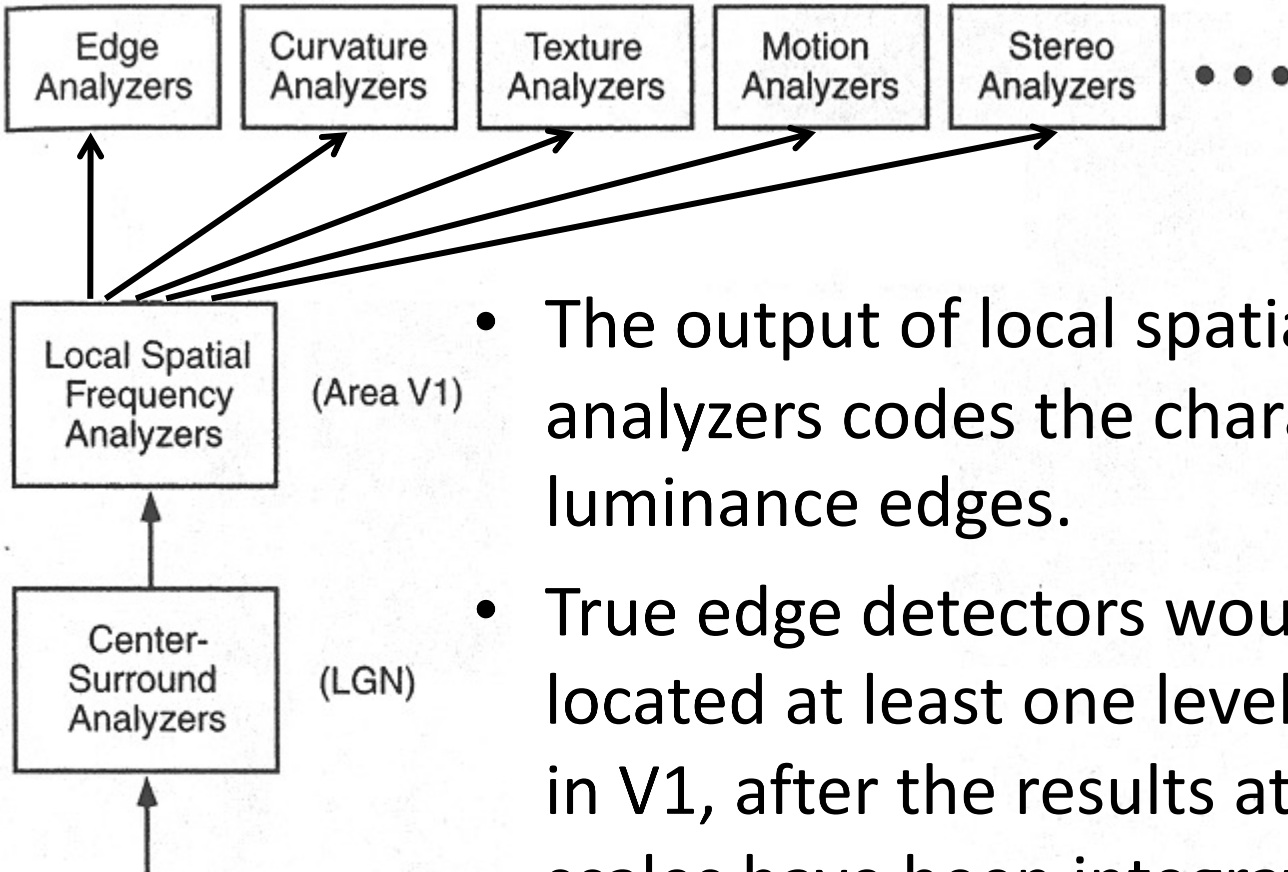


- The most obvious strategy is to base the correspondence on the location and orientation of the edges.
  - But the location and orientation of a given edge may differ from one scale to another.
- Scale space algorithm:
  - A continuous scale space that ranges from very small edge operators to very large ones.
  - A virtual continumm of scale sizes, edge information can be followed throughout this dimension to determine which edges go with which.
  - The receptive fields of cortical cells include a fairly dense representation of sizes in the spatial frequency dimension.



3.2 A Theoretical Synthesis

- Different views
  - The psychophysical view: the cortical cells describe images in a piecewise, local Gabor analysis.
  - These cells are actually the physiological implementation of edge detection mechanisms at different spatial scales.
  - Witkin’s scale space analysis theory
- How can we tell what the real function of these cells might be?

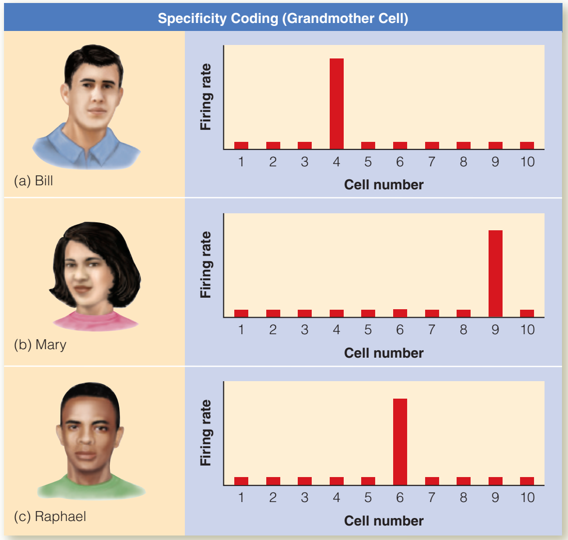


- The output of local spatial frequency analyzers codes the characteristic of luminance edges.
- True edge detectors would then be located at least one level beyond the cells in V1, after the results at different spatial scales have been integrated.
- All of the cortical processing of visual information in primates goes through area V1 and then fans out to a variety of other cortical areas.

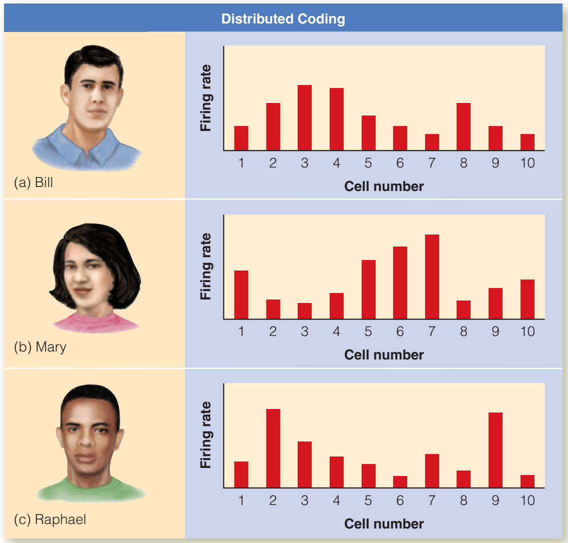
- The representation of image structure in V1 may well be a kind of unified image-based representation.
- V1 appears to represent the visual image in terms of the continuous output of analyzers at a variety of different positions, orientations, spatial scales, and phases.
- Whereas Marr suggested that the primal sketch was a symbolic representation constructed of bars, edges, blobs, and terminators. (wrong)

Sensory Code?

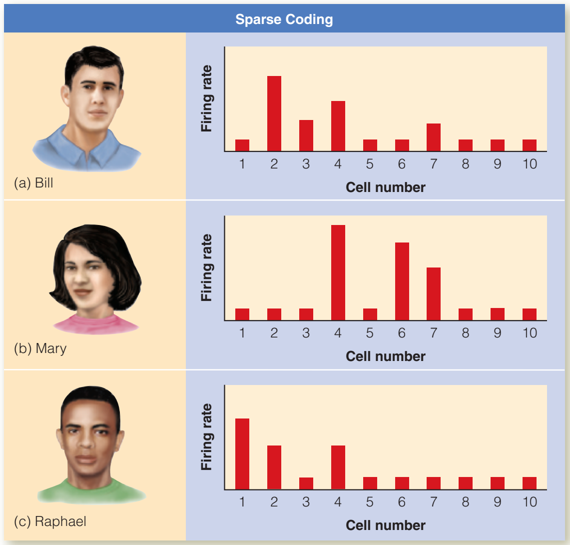
- Specificity Coding: Representation by the firing of a specialized neuron
- Distributed Coding: Representation by the firing of large groups of neurons
- Sparse Coding: Representation by the firing of a small number of neurons



Specificity Coding

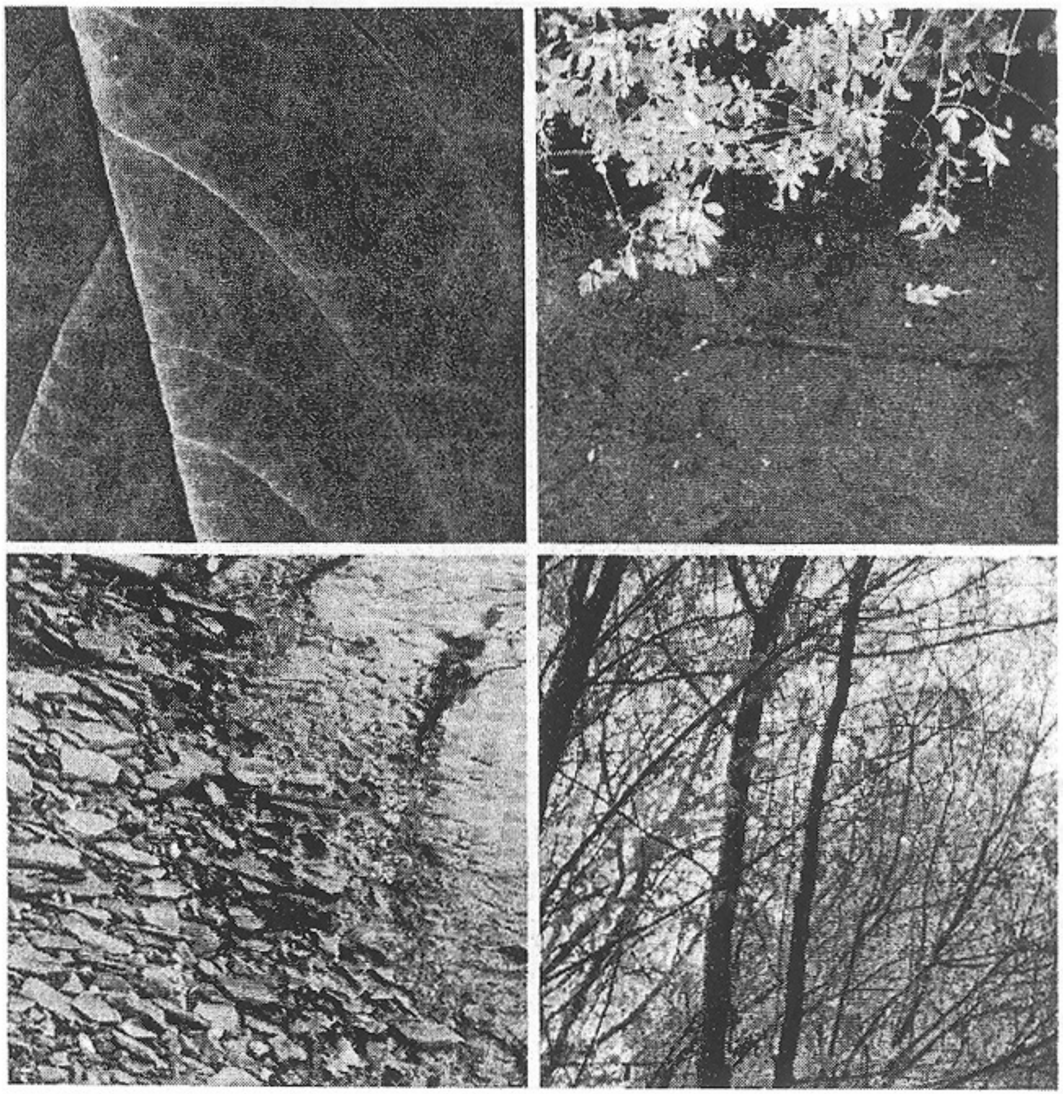


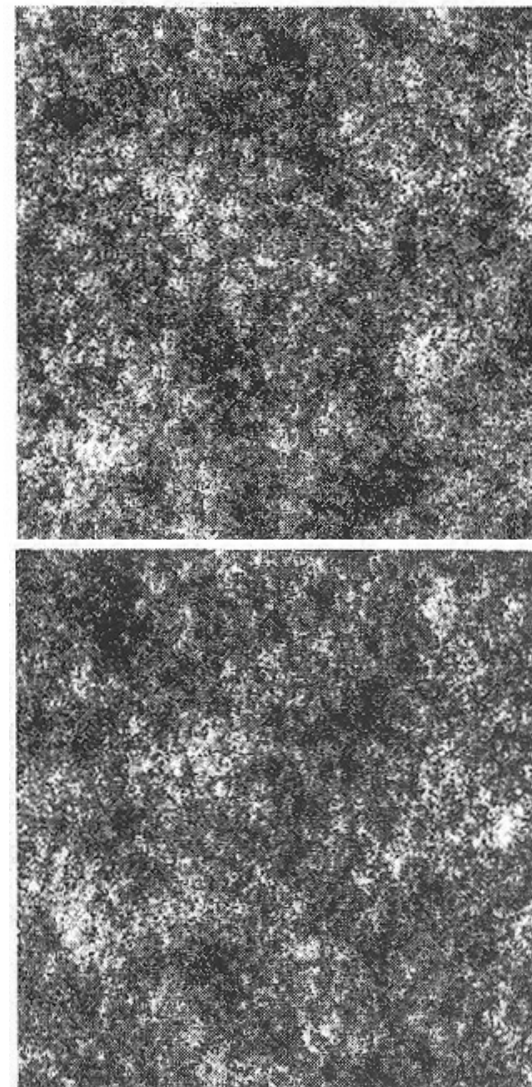
Distributed Coding.



Sparse Coding

Exploiting the Structure of Natural Images





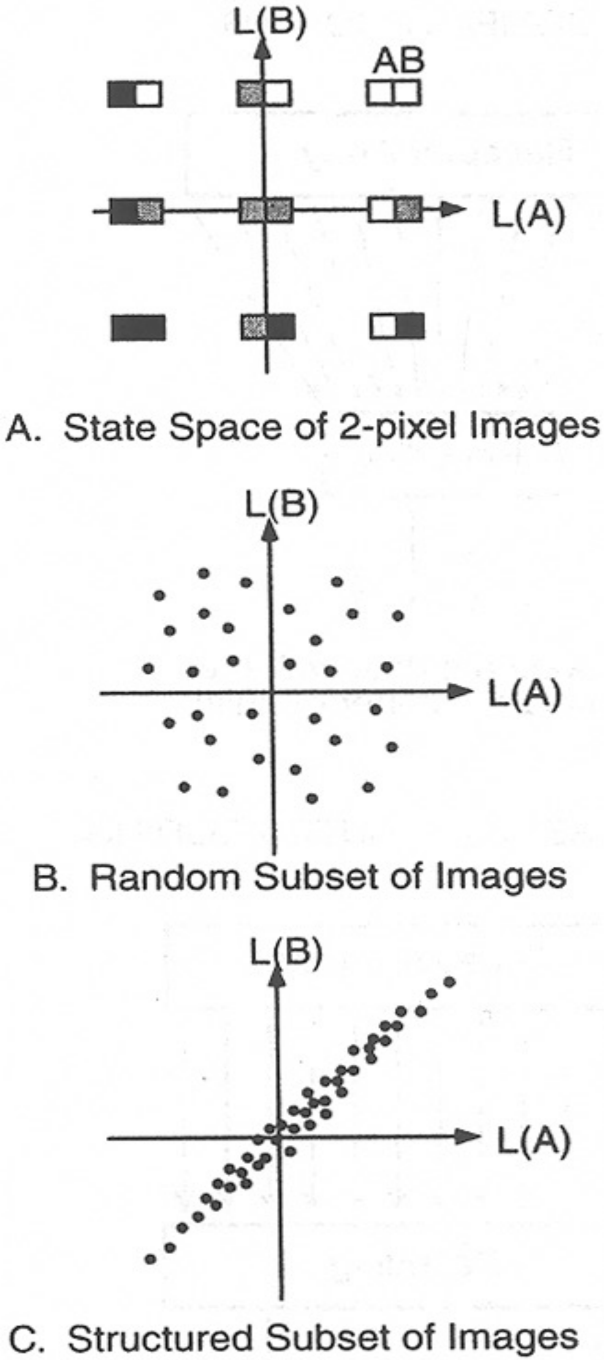
Natural images vs. unnatural images

Exploiting the Structure of Natural Images

- If the visual system does represent image-based information in this fashion, why might it do so?
- The most plausible explanation is that such a representation arises to exploit structure in natural images that are projected onto our retinae.
  - The term **“exploiting structure”** refers to eliminating redundancies so that the information from such natural images can be represented more efficiently.
  - The receptive field of V1 cells may be optimized for extracting the **statistical structure** of natural images according to particular computational principles.

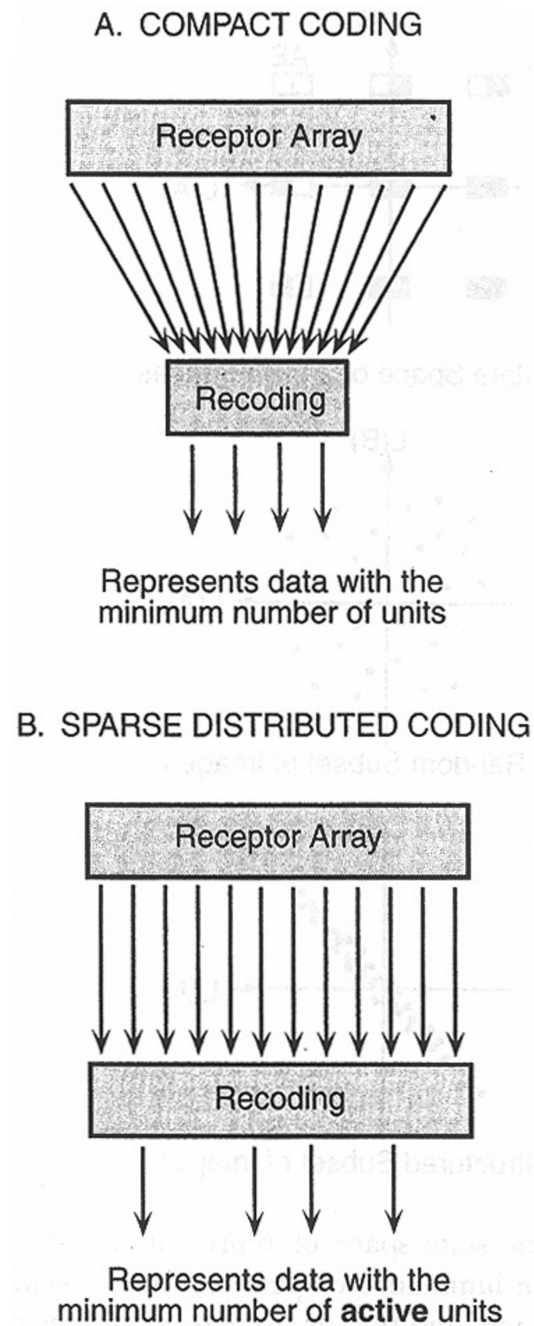
- The **state space of** a receptor array
  - Given an array composed of n receptors, each of which can represent any value within a range of luminances, every possible image that can be represented in that array corresponds to a single point in an n-dimensional space.
  - The entire state space represents the set of all possible images that the array can encode.
  - If the array consists of 1024x1024 receptors, each of which has 256 gray levels, then the state space would be able to represent $256^{1024X1024}$ distinct gray-scale images.
  - There are a million receptors and a nearly continuous range of luminance levels.

- What does such a state space have to do with the statistical structure of natural images?
  - Because the set of natural images is a small subset of all possible images, it will necessarily occupy only a very restricted set of locations in the state space.
  - The key question is how these natural images are distributed within the state space.
    - Are they more or less randomly located over the whole space, or
    - are they clumped together in systematic ways?

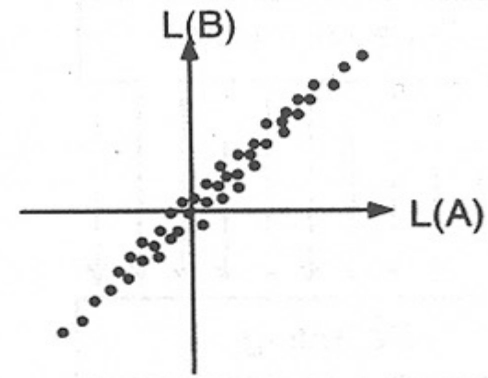


- Fig.A Consider the state space for two-pixel images consisting of 3 luminance values
- Fig.B No statistical structure for the visual system to exploit.
- Fig.C The visual system could take advantage of this structure to increase efficiency.

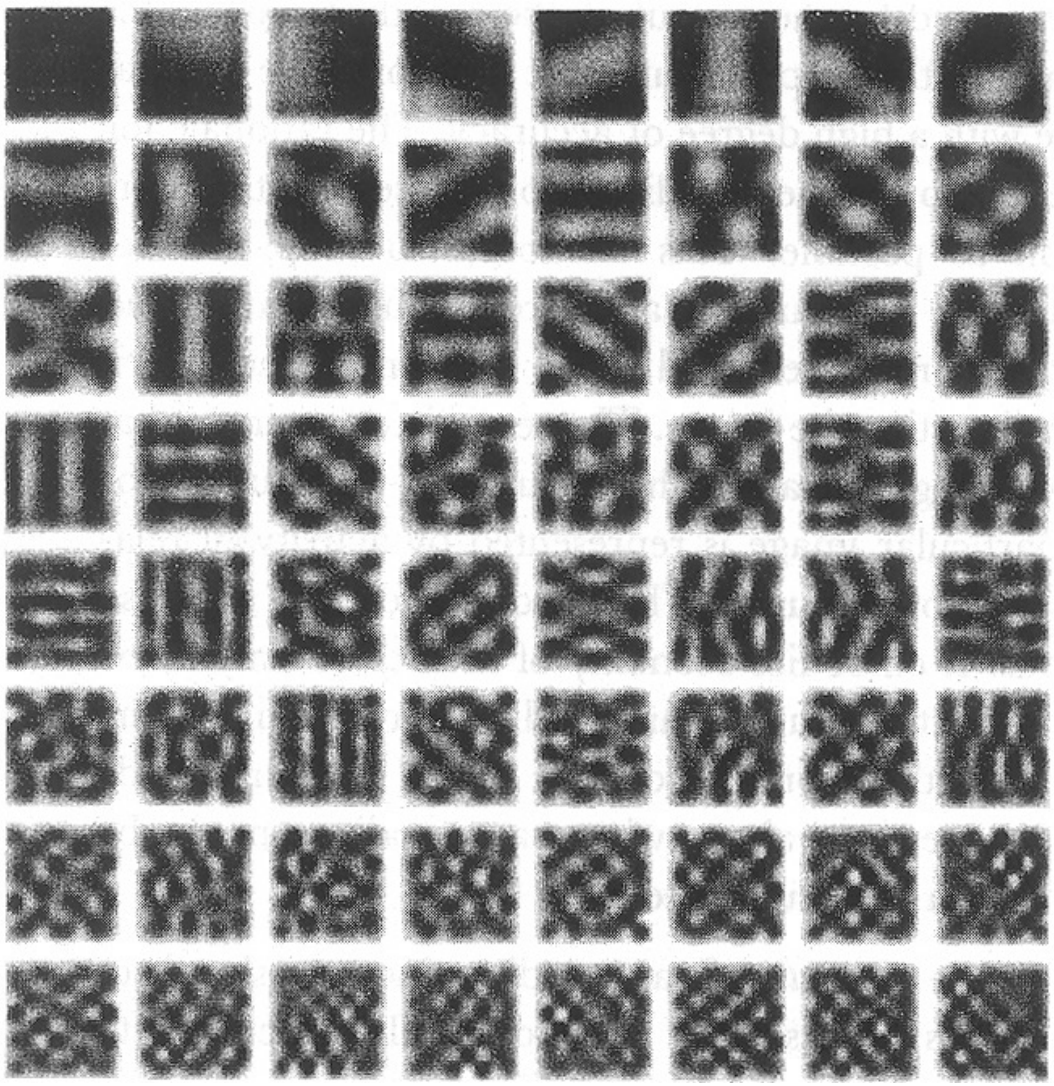
- How the visual system might exploit such structure?
- Two possibilities
  - **Compact coding**: the output of the receptor array should be recoded so that the number of units needed to represent the image is minimized.
  - **Sparse coding**: the receptor output should be recoded so that the number of active units is minimized.



- Compact Coding – Principle Components Analysis
  - A procedure that identifies **a reduced set of orthogonal basis functions that capture the maximum variance of the subset of points** in the state space.



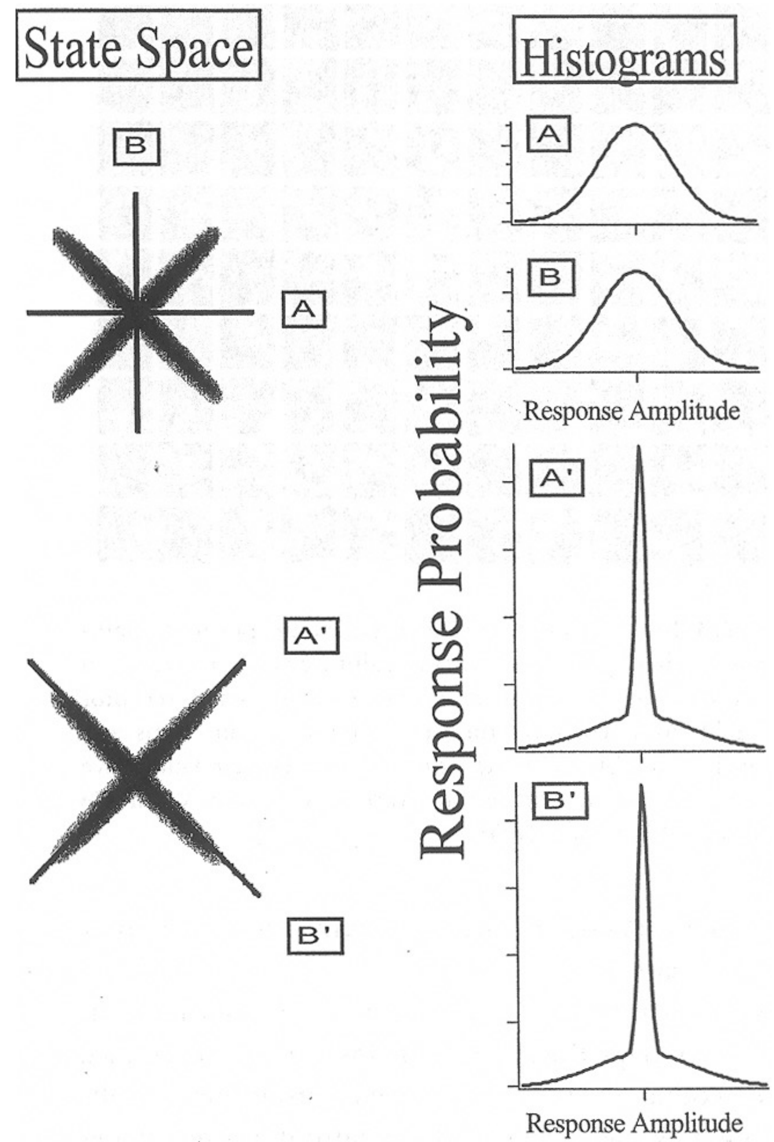
  - Representing the images in PCA provides a very efficient representation because the dimensionality is reduced by half without degrading the structure of the image much at all.
  - The method would not do a very good job of providing an accurate representation of random images.
  - Because natural images have such structure, theorists conjectured that the receptive field of V1 cells might be learned from extensive viewing of natural images using the compact coding principle.

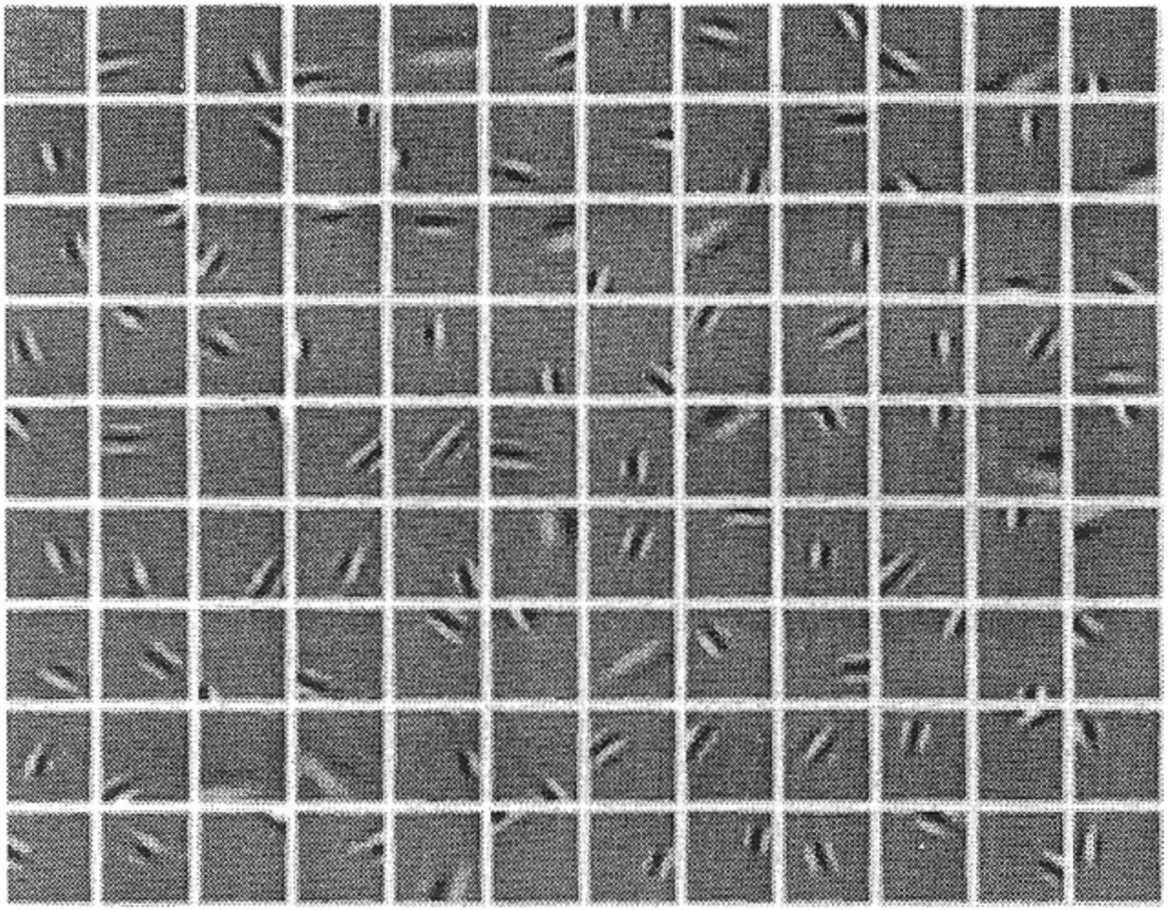


They do not correspond closely enough to actual receptive fields in V1 cells to provide a convincing explanation.

The set of receptive fields produced by PCA on small patches (8X8 pixels) of natural images

- Sparse coding- Independent Component Analysis
  - Efficient representation is achieved by minimizing the number of active units rather than the number of available units.
  - **The peakedness of a probability distribution** is characteristic of sparse codings.
  - It means that sparse recodings will have relatively few active units and relatively many inactive units.



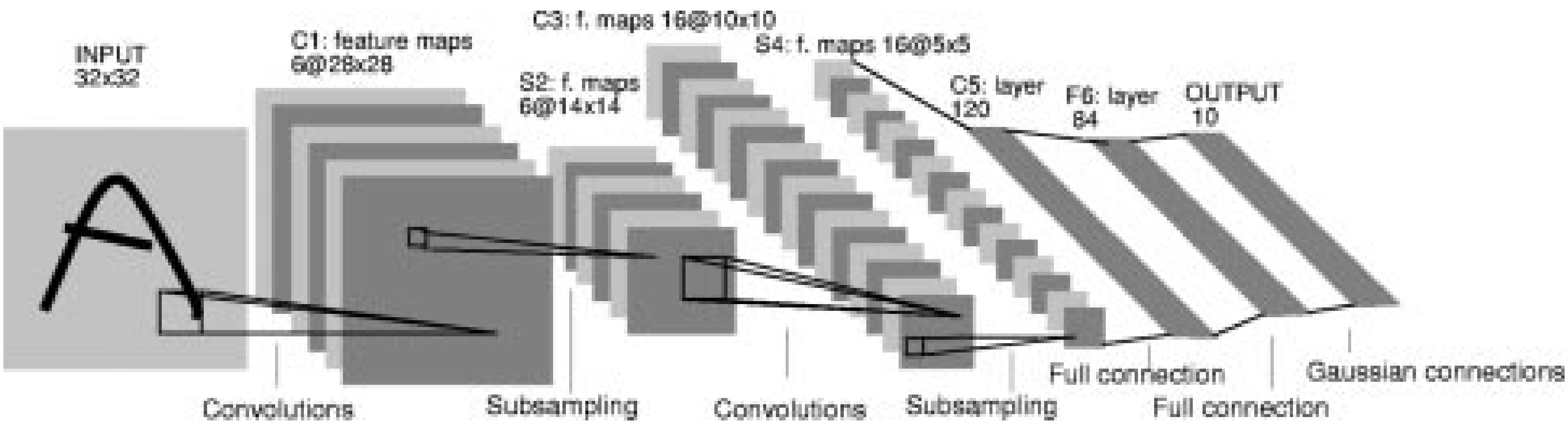


- The receptive fields show localized, oriented structure at a variety of different spatial scales.
- They are very much like the Gabor functions that characterize V1 simple cells.

Many natural image patches (16 X 16 pixels) were presented to a network of 192 recoding units

- Field (1994) gave three suggestions about why sparse coding might be desirable in the visual system.
  - **Signal-to-noise ratios**. It increase SNR because only a few cells will be active with many more inactive.
  - **Feature detection**. It should assist later recognition processes based on the detection of specific features because relatively few feature detectors will be active, and more active than those in dense coding.
  - **Storage and retrieval from associative memory**. Allow networks to store more memories and to retrieve them more effectively.

Convolutional Neural Networks



- Yann Lecun, et al. Gradient-based learning applied to document recognition, Proceedings of the IEEE, Vol.86,No.11,1998

Sensory Code?

1. Distributed Coding
2. Sparse Coding
3. Specificity Coding

4. Conclusion

- Physiological mechanisms
  - Receptive fields of Ganglion cells and Geniculate cells: on-center and off-center
  - The structure of Lateral Geniculate nucleus
  - The structure of Striate cortex
- Psychophysical mechanisms
  - Spatial frequency theory: Fourier analysis
- Computational Approach
  - Marr's Primal Sketches: Edge Detection
  - A Theoretical Synthesis: Compact coding vs. Sparse coding

Question?