Fundamentals of Computer Vision

Unit 1: Human Visual System. Marr's model

Jorge Bernal

Index

01

1. Human Visual System

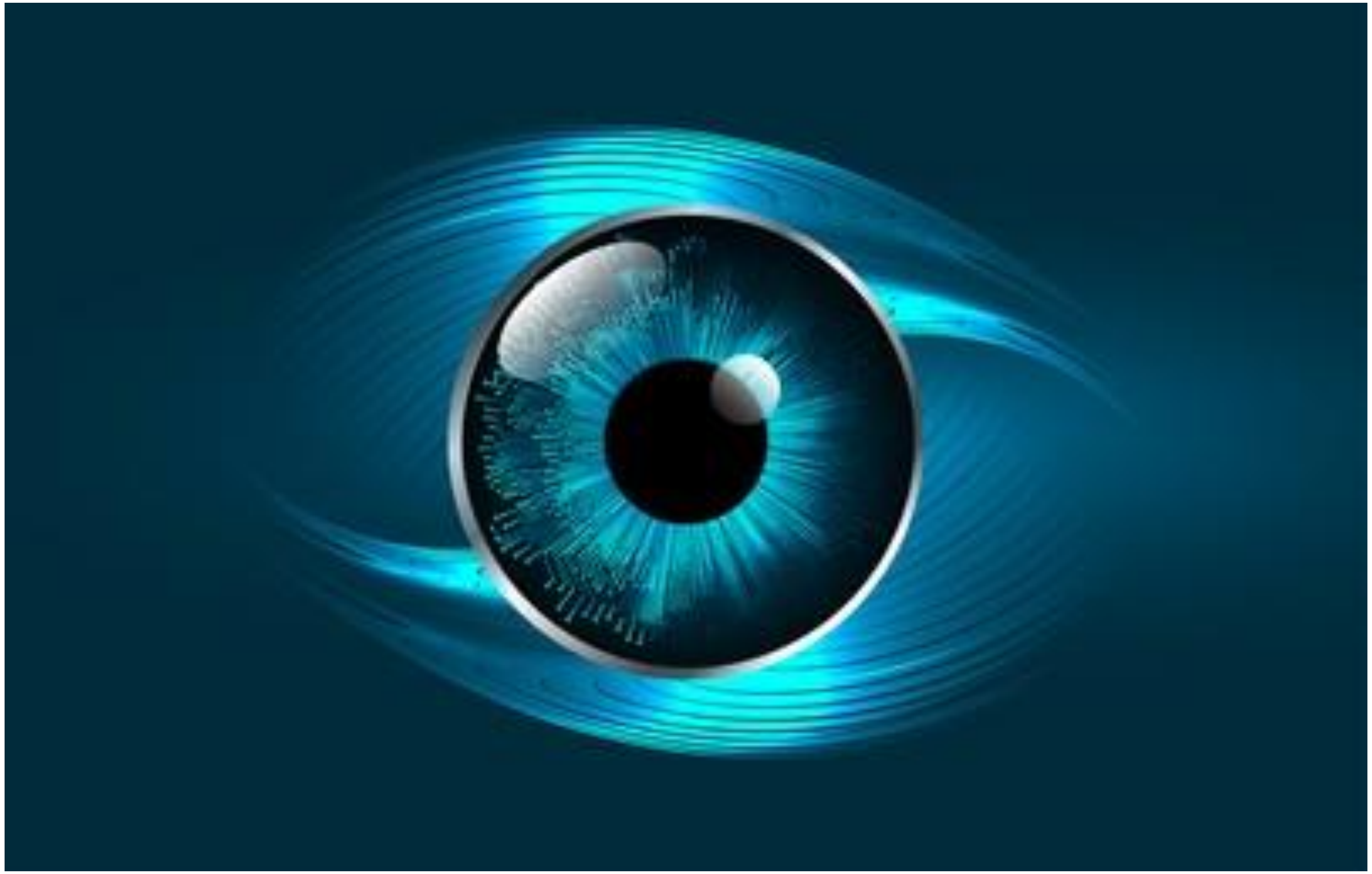
02

2. Marr’s model

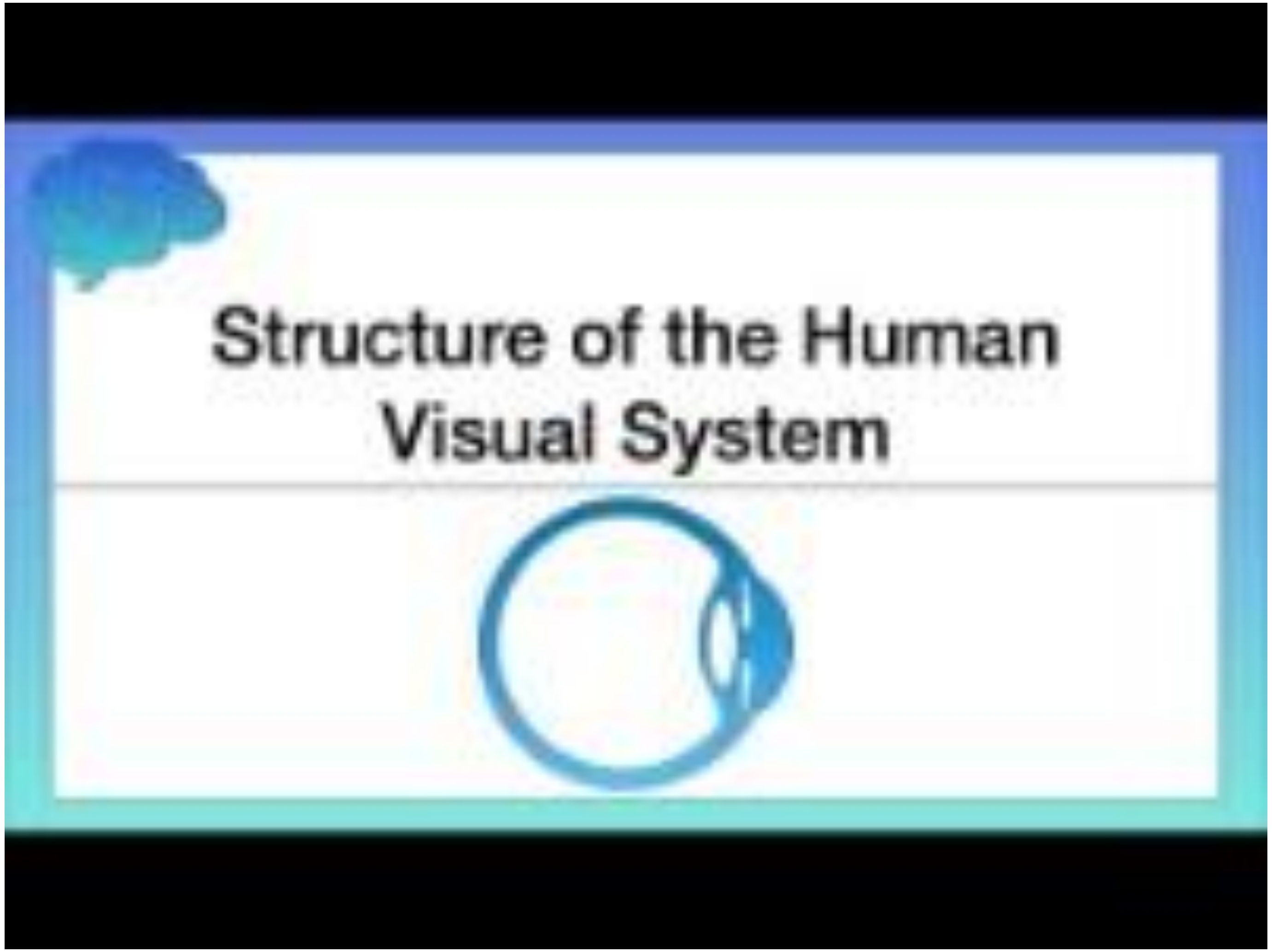
1

Human Visual System

Human Visual System



Human Visual System



Human Visual System

In many image processing applications, the objective is to help a human observer **perceive the visual information in an image**. Therefore, it is important to understand the human visual system.

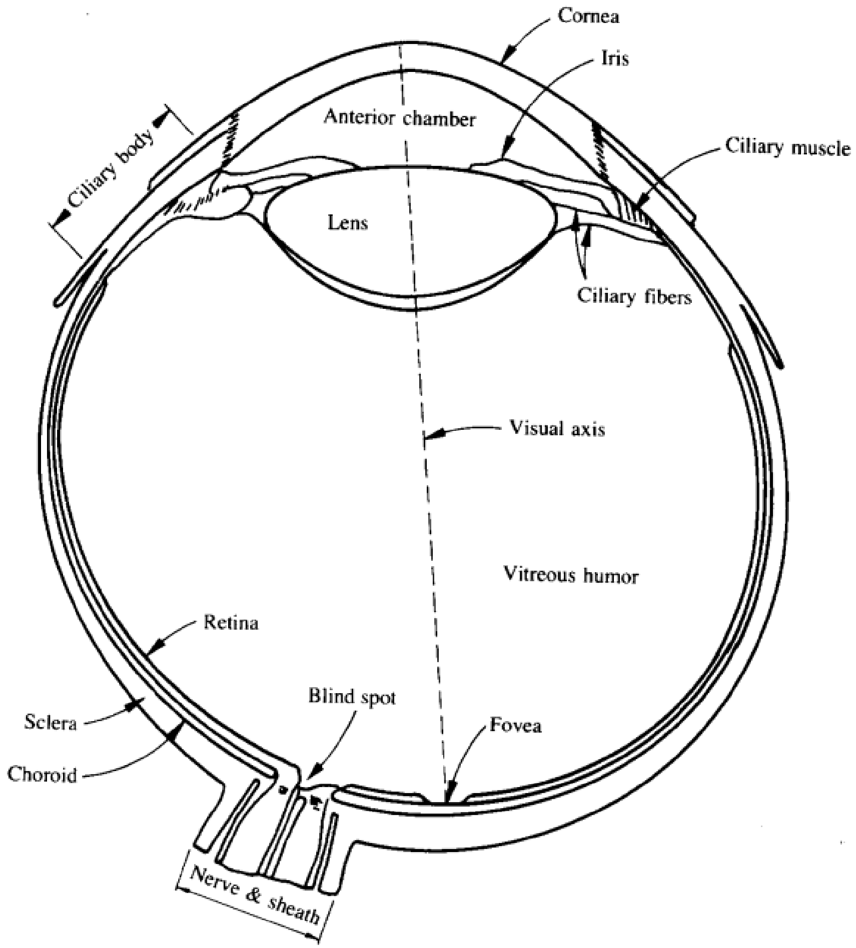
Human Visual System

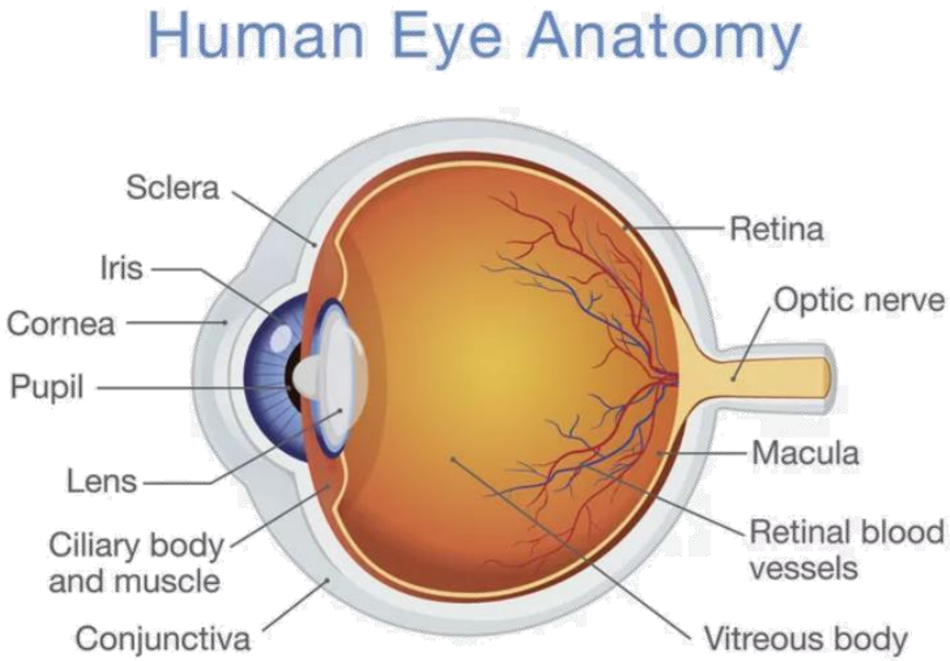
The human visual system consists mainly of the eye (image sensor or camera), optic nerve (transmission path), and brain (image information processing unit or computer).

It is one of the most sophisticated image processing and analysis systems.

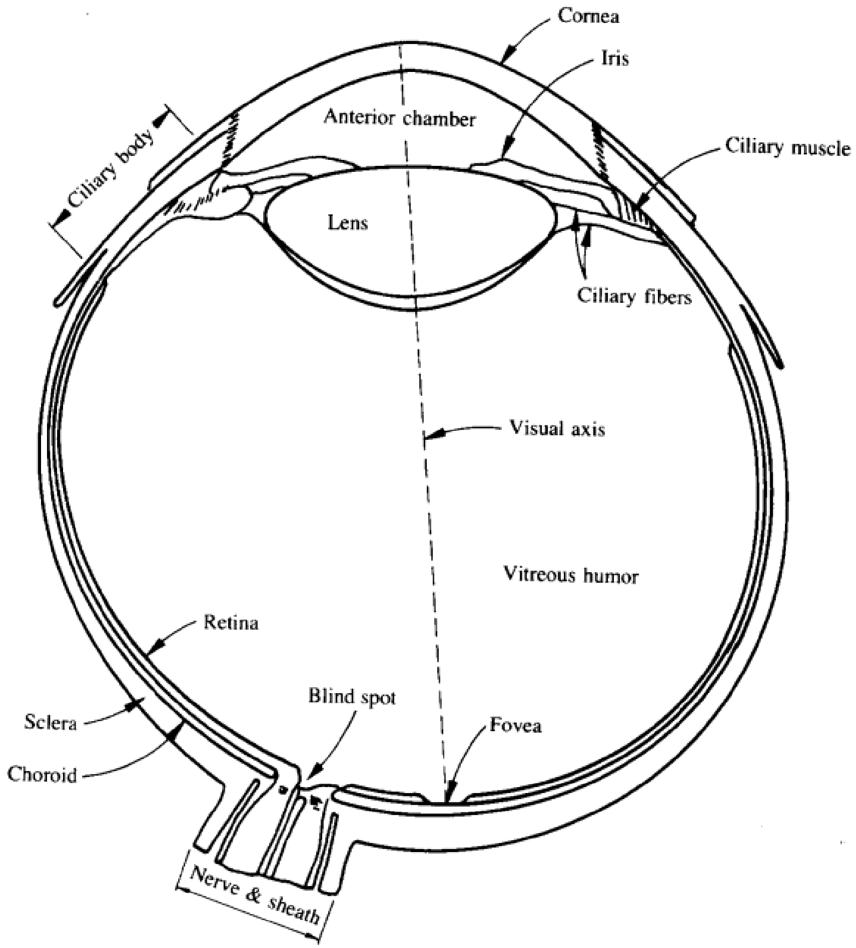
Its understanding would also help in the design of efficient, accurate and effective computer/machine vision systems.

Human Visual System





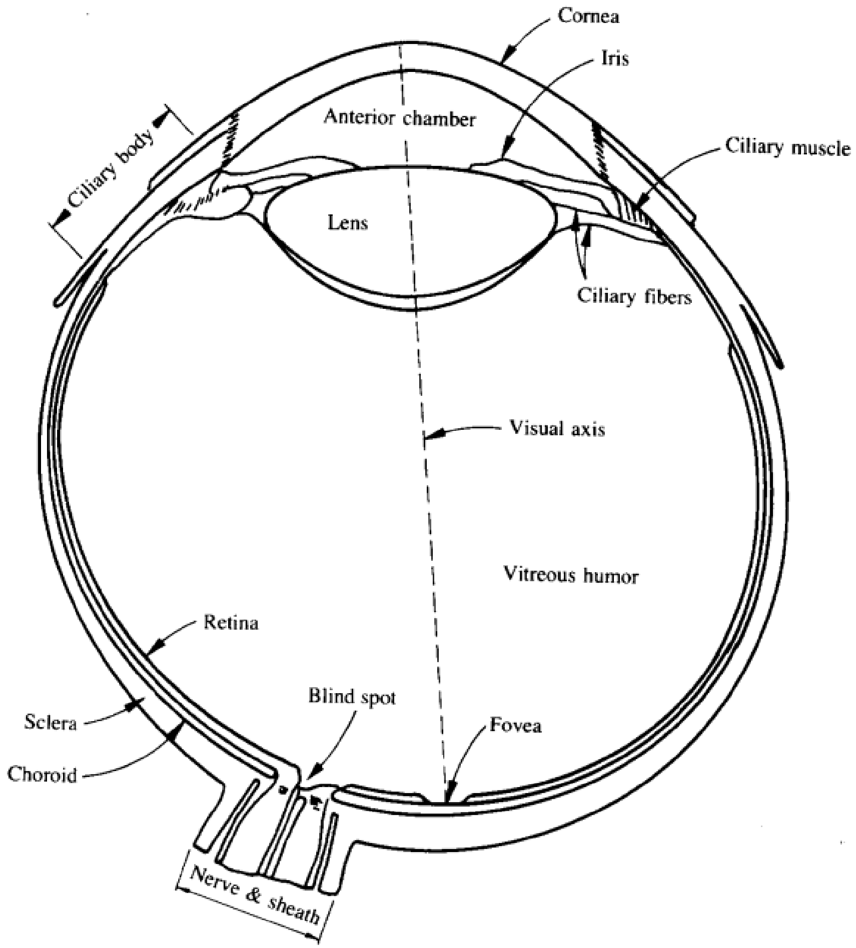
Human Visual System



Nearly spherical with a diameter of 20 mm (approx.).

- Cornea: Outer tough transparent membrane, covers anterior
- Sclera: Outer tough opaque membrane, covers rest of the optic globe.
- Choroid: Contains blood vessels, provides nutrition.
- Iris: Anterior portion of choroid, pigmented, gives color to the eye.

Human Visual System



- Pupil: Central opening of the Iris, controls the amount of light entering the eye (diameter varies from 2-8 mm).
- Lens: Made of concentric layers of fibrous cells, contains 60-70% water.
- Retina: Innermost layer, “screen” on which image is formed by the lens when properly focussed, contains photoreceptors (cells sensitive to light)

Human Visual System

Two types of photoreceptors: rods and cones (light sensors).

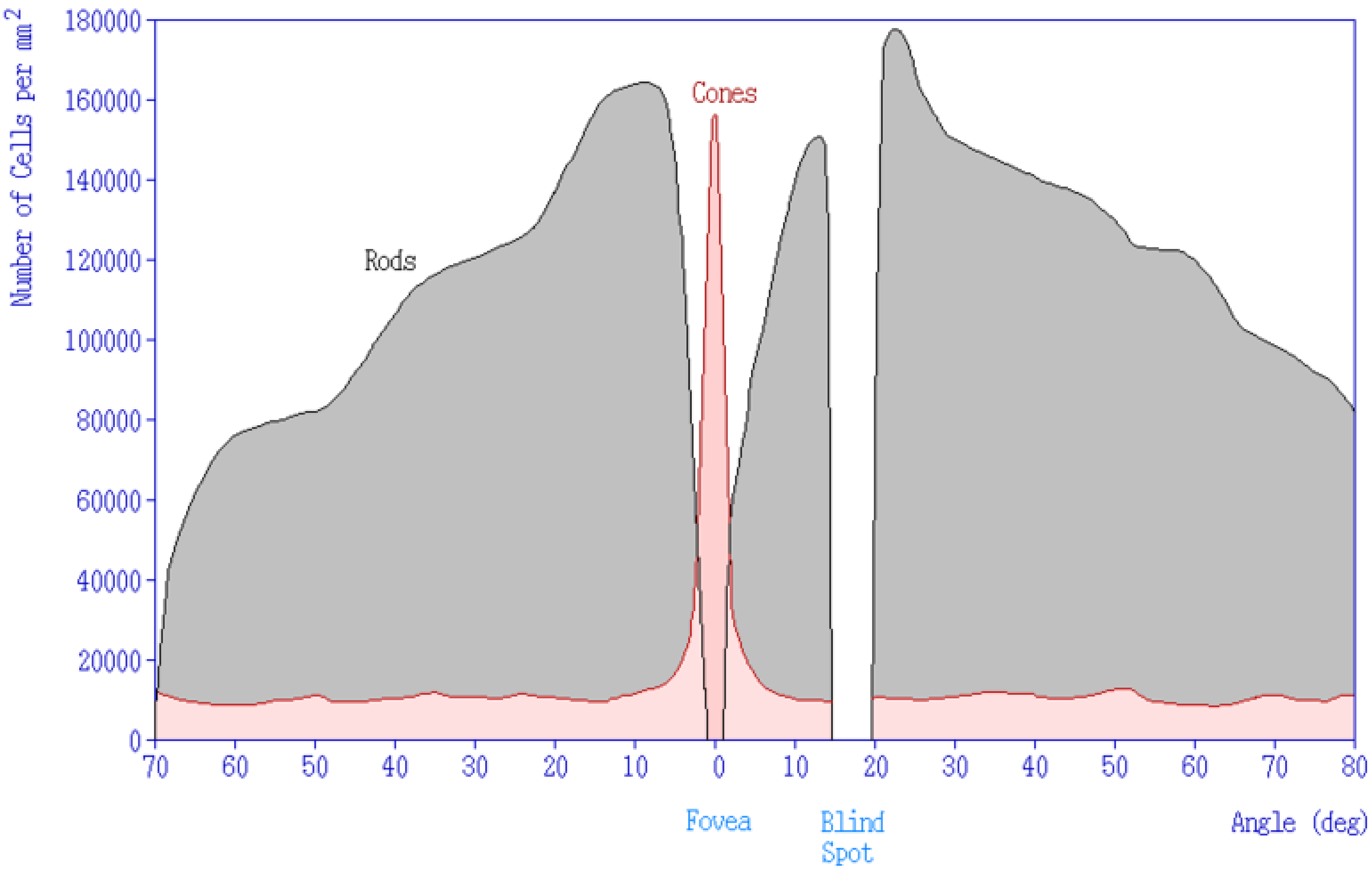
- Cones: 6-7 million, located in central portion of retina (fovea), responsible for photopic vision (bright-light vision) and color perception, can resolve fine details.
- Rods: 75-150 million, distributed over the entire retina, responsible for scotopic vision (dim-light vision), not color sensitive, gives general overall picture (not details).

Human Visual System

- Fovea: Circular indentation in center of retina, about 1.5mm diameter, dense with cones.

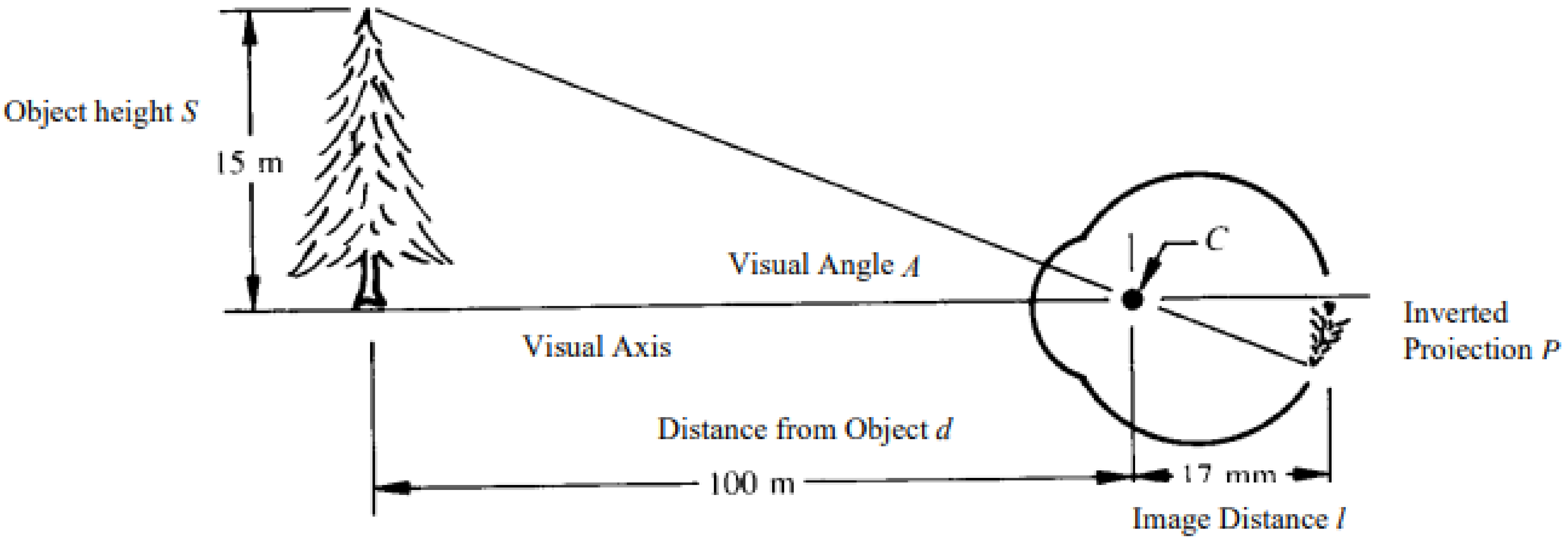
  Photoreceptors **around fovea responsible for spatial vision** (still images).
  Photoreceptors **around the periphery responsible for detecting motion**.
  Blind spot: Point on retina where optic nerve emerges, devoid of photoreceptors.

Human Visual System



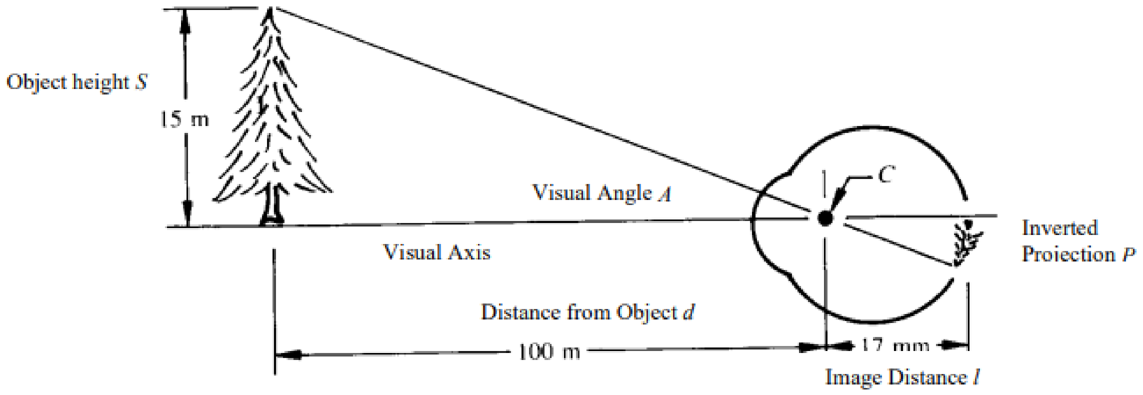
Human Visual System

Simple model for image formation:



Human Visual System

Simple model for image formation:

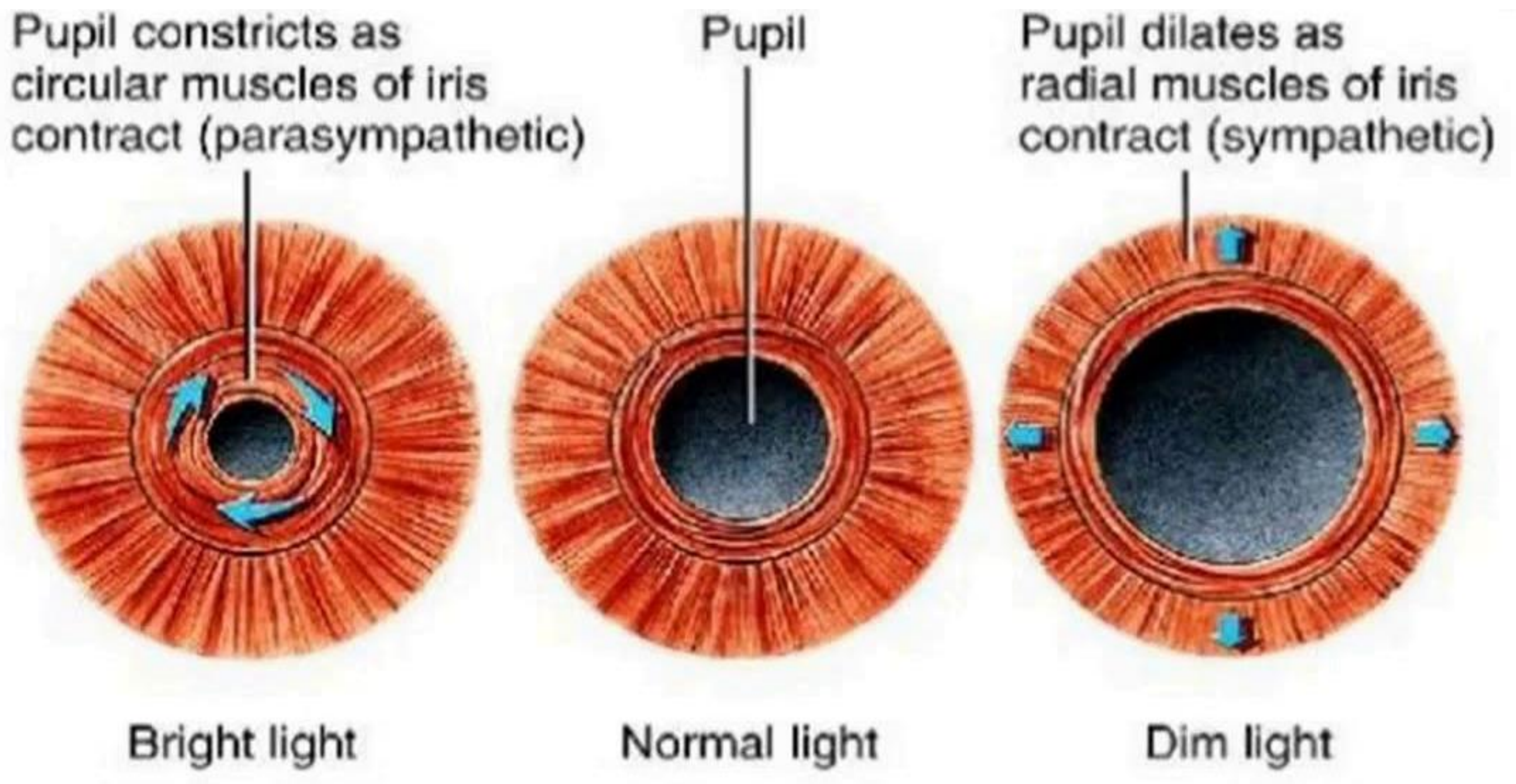


- Distance between center of lens and retina varies from 14 to 17 mm
- The farther the objetc, smaller the refractive power of the lens, the large the focal length

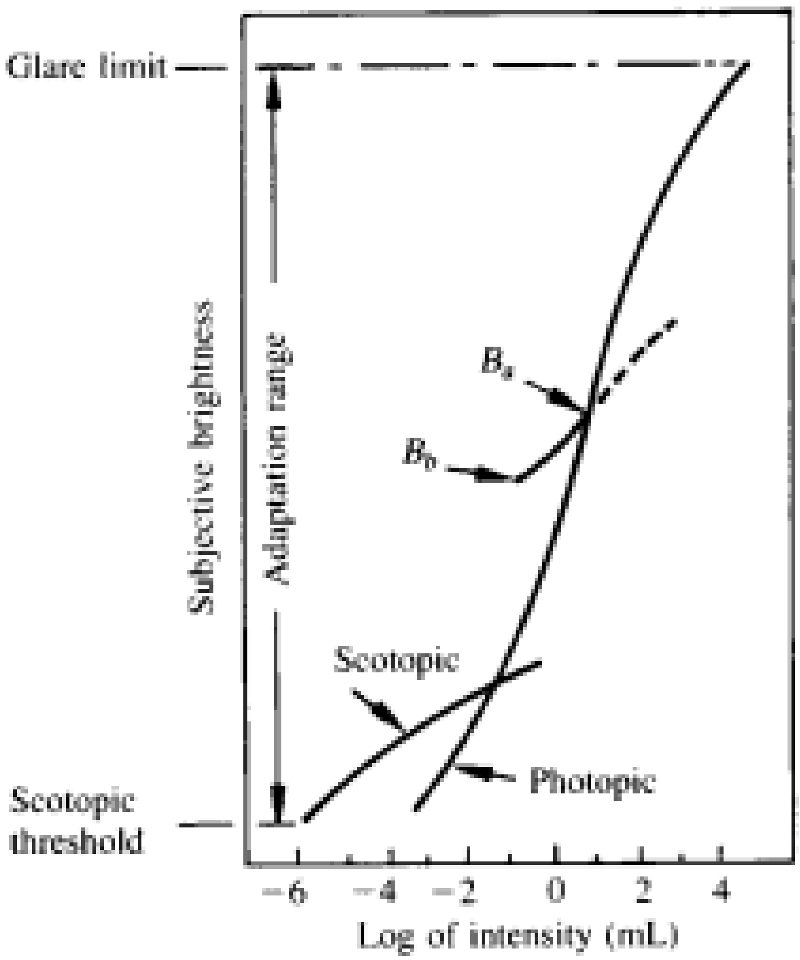
Human Visual System

- Human eye can adapt to an enormous range of light intensity levels, almost 10 orders of magnitude!
- Brightness perceived (subjective brightness) is a logarithmic function of light intensity.
- Eye cannot simultaneously operate over such a range of intensity levels.
- This is accomplished by changing the overall sensitivity→ Brightness adaptation.

Human Visual System



Human Visual System



- At a given sensitivity, the eye can simultaneously discriminate only a small number of intensity levels.
- For a given condition, the sensitivity of the visual system is called the brightness adaptation level (ex. B_a).
- At this adaptation, the eye can perceive brightness in the range B_b (below which, everything is perceived as black) to B_a (above which, the eye adapts to a different sensitivity

Human Visual System

Brightness discrimination: The ability of the eye to discriminate between changes in brightness levels is called brightness discrimination.

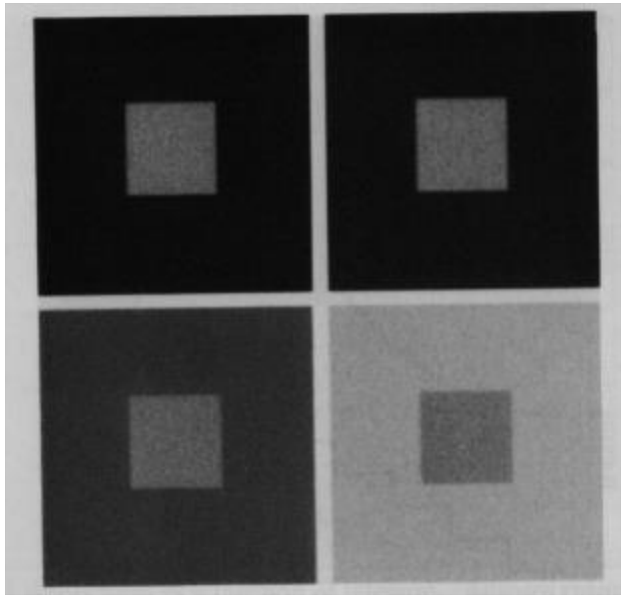
The increment of intensity ΔIc that is discriminable over a background intensity of I is measured.

- Weber ratio --- it is the ratio $\Delta Ic / I$. •
- Small value of Weber ratio: good brightness discrimination, a small percentage change in intensity is discriminable.
- Large value of Weber ratio: poor brightness discrimination, a large percentage change in intensity is required.
- At high intensities the brightness discrimination is good (small Weber ratio), than at low intensities.

Human Visual System

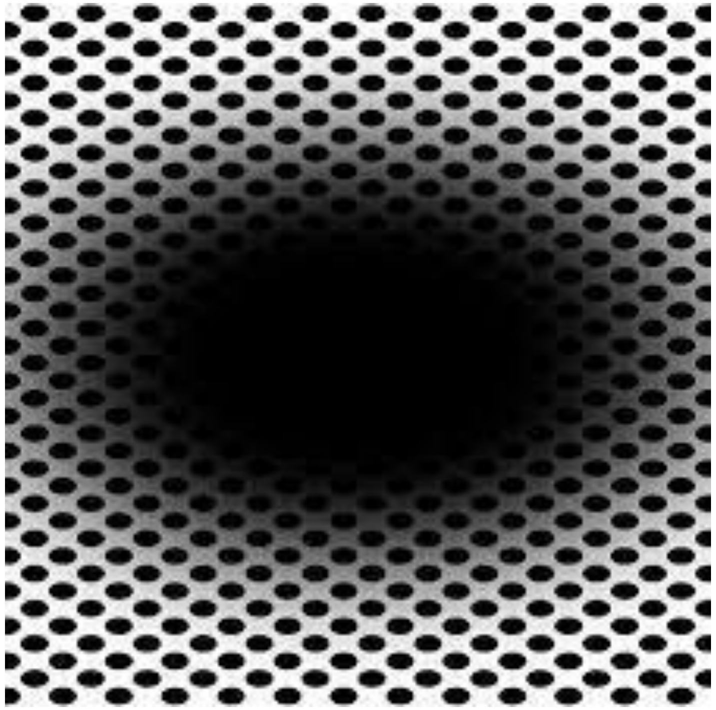
Important!! A region's perceived brightness is not a function of only its intensity, but it depends on the background intensity as well.

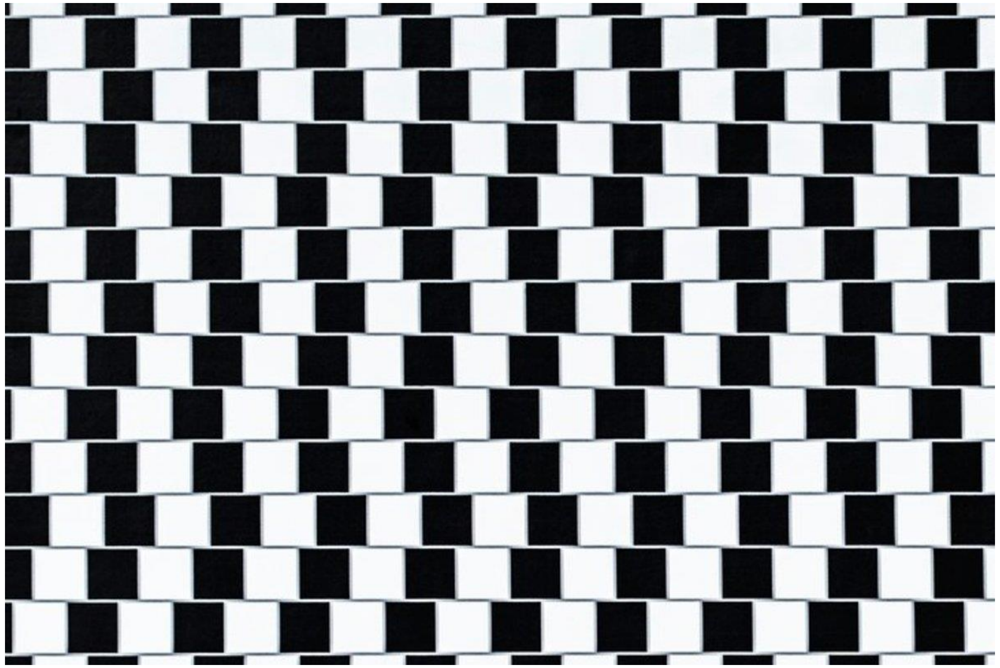
All the center squares in the figure below have exactly the same intensity. However, they appear to the human eye to become darker as the background becomes brighter.



Human Visual System

Optical illusion: the eye fills in non-existent information or wrongly perceives geometrical properties of objects





Human Visual System

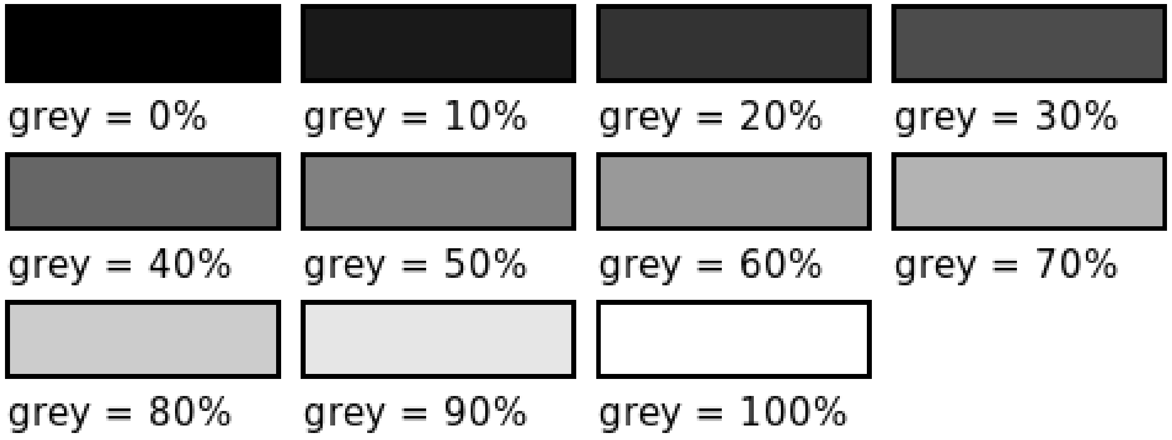
Light and EM Spectrum:

- Electromagnetic (EM) waves or radiation can be visualized as propagating sinusoidal waves with some wavelength λ or equivalently a frequency ν where $\lambda\nu = c$, c being the velocity of light.
- Equivalently, they can be considered as a stream of (massless) particles (or photons), each having an energy E proportional to its frequency ν; $E = h\nu$, where h is Planck's constant.
- EM spectrum ranges from high energy radiations like gamma rays and X-rays to low energy radiations like radio waves.
- Light is a form of EM radiation that can be sensed or detected by the human eye. It has a wavelength between 0.43 to 0.79 micron.

Human Visual System

Light and EM Spectrum:

- Different regions of the visible light spectrum corresponds to different colors.
- Light that is relatively balanced in all visible wavelengths appears white (i.e. is devoid of any color). This is usually referred to as achromatic or monochromatic light.
- The only attribute of such light is its intensity or amount. It is denoted by a gray value or gray level. White corresponds to the highest gray level and black to the lowest gray level.



Human Visual System

Attributes associated to describe a chromatic light source:

- **Radiance** is the total amount of energy (in unit time) that flows from the source and it is measure in Watt (W).
- **Luminance** is a measure of the amount of light energy that is received by an observer. It is measured in lumens (lm).
- **Brightness** is a subjective descriptor of light measure (as perceived by a human).

Human Visual System

The wavelength of EM radiation used depends on the imaging application.

In general, the wavelength of an EM wave required to “see” an object must be of the same size (or smaller) than that of the object.

Besides EM waves, other sources of energy such as sound waves (ultra sound imaging) and electron beams (electron microscopy) are also used in imaging

2

Marr’s Model

Marr’s Model

Vision is an information-processing task:
- Representation
- Processes (computations)

A multi-level explanation is required:
- Human being perspective
- Psychologist perspective
- Programming perspective

Marr’s Model

Background:

- During 1950s and 1960s, several important discoveries in many different fields (AI, neuroscience, psychophysics)

- A big gap until 1970, no neurophysiologist had recorded new and clear high-level correlates of perception

- The ones leading the field moved away from what they were doing
    - Hubel and Wiesel focused on anatomy
    - Barlow focused on psychophysics

Marr’s Model

Background:

- During 1950s and 1960s, several important discoveries in many different fields (AI, neuroscience, psychophysics)

- A big gap until 1970, no neurophysiologist had recorded new and clear high-level correlates of perception

- The ones leading the field moved away from what they were doing
  - Hubel and Wiesel focused on anatomy
  - Barlow focused on psychophysics

Marr’s Model

More problems:

- Gross developed hand detector that can tell whether they exist or not, but not why and how do they work

- Neurphysiology and psychophysics can describe the behaviour of cells or subjects but cannot explain such behaviour

Marr’s Model

Several questions to be answered:

- What is the visual area of the cortex actually doing?
- What are the problems related to what is being done that need explanation?
- At which level of description should such explanation go?

Marr’s Model

Marr’s core idea:

- An additional level of understanding should exist at which the character of the information-processing tasks carried out during perception are analyzed and understood in a way that is independent of the particular mechanisms and structures that implement them in our head.

Marr’s Model

Marr’s core idea:

- What was missing was **the analysis of the problem as an information-processing task.**
- This does not imply to neglect the understanding at other levels (neurons, computer programs) → it is a complement to understand the functions of all these neurons

Marr’s Model

Representation vs. description

- Representation: a formal system for making explicit certain entities or types of information
- Description: the result of using representation

Marr’s Model

Process

- The transformation, interaction, connection and/or manipulation of representations
- Two components:
  - Representation
  - algorithm

Marr’s Model

The three levels of Marr’s Theory

Computational theory:
- What is the goal of computation?
- Why should we use it?
- What is the logic of the strategy by which it can be carried out?

COMPUTATIONAL

Marr’s Model

The three levels of Marr’s Theory

Computational theory:

- What is the goal of computation?
- Why should we use it?
- What is the logic of the strategy by which it can be carried out?

COMPUTATIONAL

Representation and algorithm:

- How can this computational theory be implemented?
- How to represent inputs and outputs?
- Which algorithm can we use for the transformation?

COGNITIVE

Marr’s Model

The three levels of Marr’s Theory

Computational theory:
- What is the goal of computation?
- Why should we use it?
- What is the logic of the strategy by which it can be carried out?

COMPUTATIONAL

Representation and algorithm:
- How can this computational theory be implemented?
- How to represent inputs and outputs?
- Which algorithm can we use for the transformation?

COGNITIVE

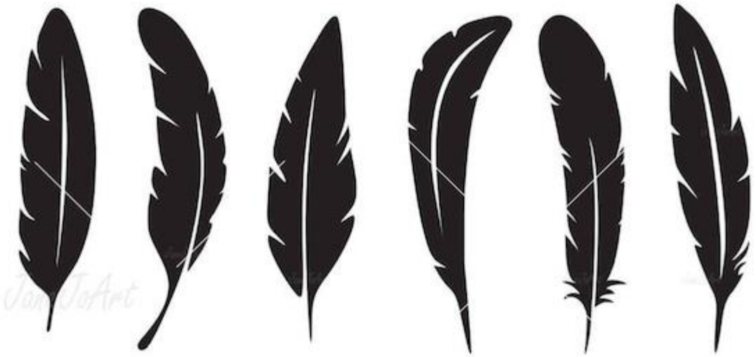
Hardware implementation:
- How can the representation and algorithm be realized physically?

NEUROLOGIC

Marr’s Model

- An algorithm is likely to be understood better if **we understand the nature** of the computational problem to be solved rather than by examining the different lines of code or the hardware it is implemented on

- It is not useful to understand perception by studying only neurons



Marr’s Model

Representational framework for vision

Name	Purpose	Primitives
Image	Represents intensity	Intensity value at each point in the image
Primal sketch	Makes explicit important information about the 2D image, primarily **intensity change** and their geometrical distribution and organization	Zero-crossings Blobs Terminations and discontinuities Edges Boundaries

Marr’s Model

Representational framework for vision

Name	Purpose	Primitives
Image	Represents intensity	Intensity value at each point in the image
Primal sketch	Makes explicit important information about the 2D image, primarily **intensity change** and their geometrical distribution and organization	Zero-crossings Blobs Terminations and discontinuities Edges Boundaries
2’5D sketch	Make explicit the orientation and rough Depth of visible surfaces and the contours of discontinuities in these quantities in a **viewer-centered** coordinate frame.	Local surface orientation (needles primitives) Distance from view Discontinuities in Depth Discontinuities in Surface orientation

Marr’s Model

Representational framework for vision

Name	Purpose	Primitives
2’5D sketch	Make explicit the orientation and rough Depth of visible surfaces and the contours of discontinuities in these quantities in a **viewer-centered** coordinate frame.	Local surface orientation (needles primitives) Distance from view Discontinuities in Depth Discontinuities in Surface orientation
3D model	Describes shapes and their spatial organization in an object-centered coordinate frame. Modular hierarchical representation that includes volumetric and Surface primitives.	3D models arranged hierarchically, each one based on a spatial configuration of a few sticks or axes to which volumetric or Surface shape primitive are attached

Marr’s Model

- To achieve this, a sequence of representations are necessary
- Starting with images on retina, we gradually reach a more objective, physical properties about an object’s shape
- With this, we can go from 2D images to viewer-centered surfaces description to object-centered object description

Marr’s Model

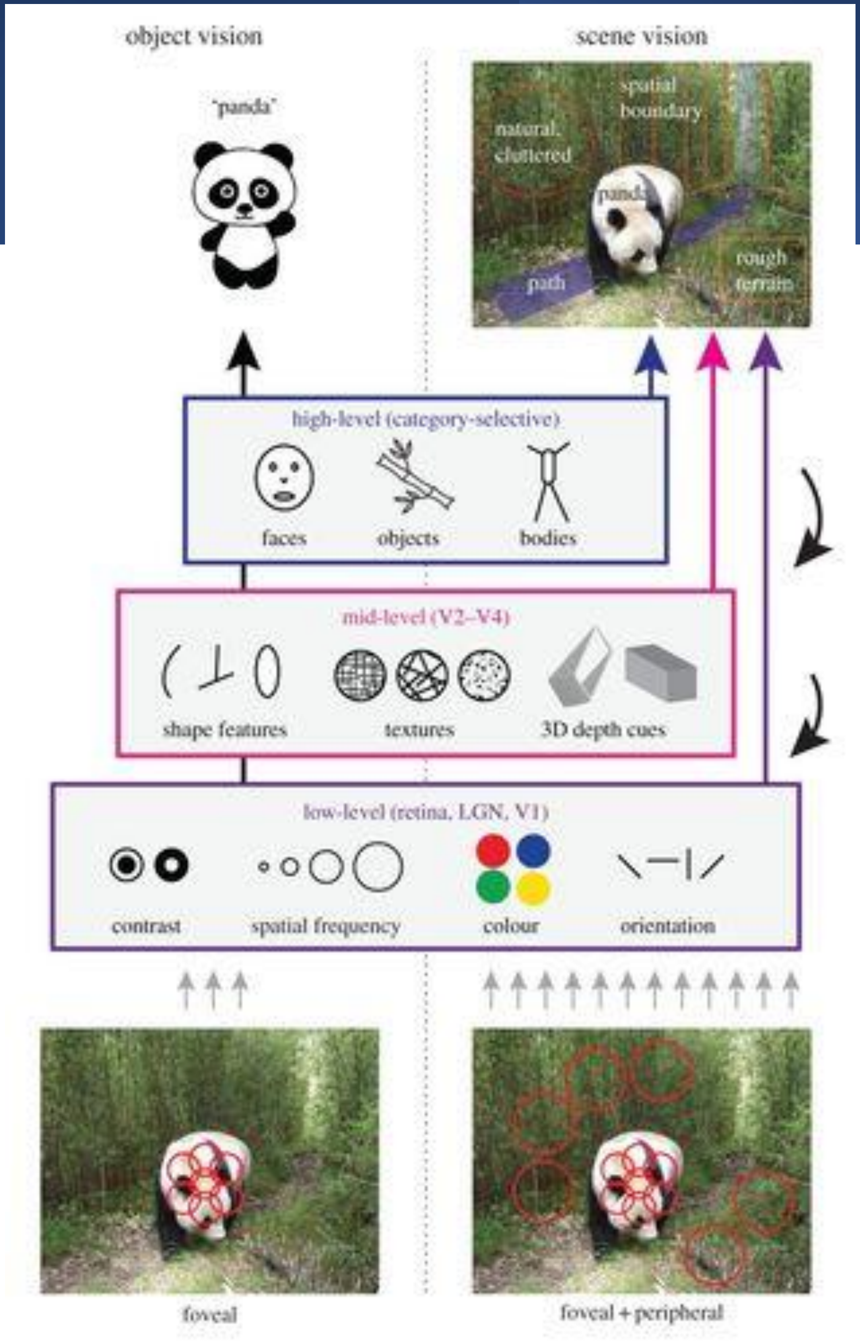
- How do we use these levels today and what do they had to do with vision?

1. Low-level vision: image preprocessing operations (edge and corner extraction, filtering, morphological operations)

2. Mid-level vision: integrate information from previous stage to organize it in objects and surfaces. Geometry and movement should be considered → object segmentation and tracking

3. High-level vision: image semantics, understand what is being shown → object recognition and scene understanding

Marr’s Model



object vision
scene vision
'panda'
spatial boundary
natural, cluttered
panda
rough terrain
path
high-level (category-selective)
faces
objects
bodies
mid-level (V2–V4)
shape features
textures
3D depth cues
low-level (retina, LGN, V1)
contrast
spatial frequency
colour
orientation
foveal
foveal + peripheral