EE 584
MACHINE VISION

Introduction

Relation with other areas
Image Formation & Sensing
Projections
Brightness
Lenses
Image Sensing

METU EE 584 Lecture Notes by A.Aydin ALATAN © 2012

Introduction

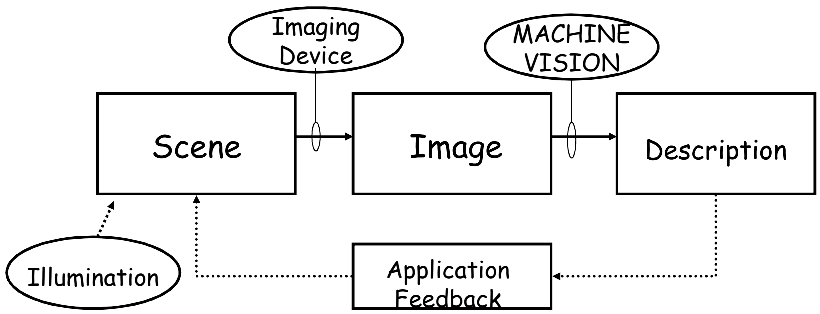
- Vision is the most powerful sense
- Vision is the most complicated sense

- The purpose of a general *machine/computer/robot vision* system is to produce a symbolic description of what is being imaged.

2

Machine Vision System

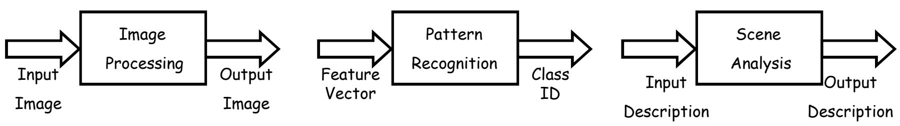
- A typical machine vision system :



Machine vision should be based on complete understanding of image formation

Relation to other fields :

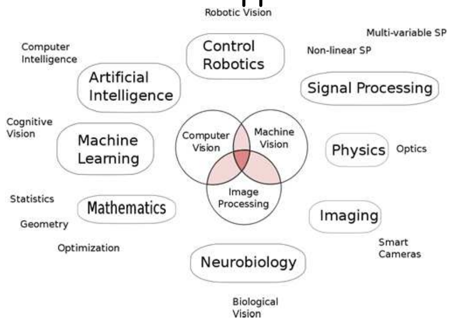
- 4 important related fields :
  - Imaging (science on creating all kinds of images)
  - Image processing
  - Pattern recognition
  - Scene analysis



None of them provides a solution to the problem "developing symbolic descriptions from images".

Relation to other fields :

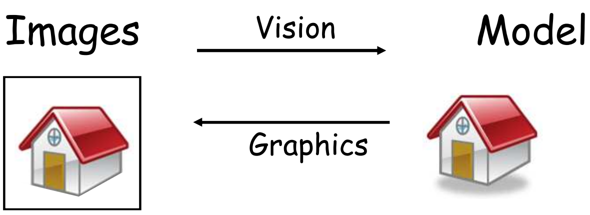
- Machine vision vs Computer vision:
  - Two terms can be used interchangeably
  - Machine vision → more constraints on the environment and focus on (industrial) applications
  - Computer vision → more generic in terms of content and applications



This course is about fundamentals of *vision* research

Image from http://en.wikipedia.org/wiki/Computer_vision

Vision and Graphics



Inverse problems: analysis and synthesis.

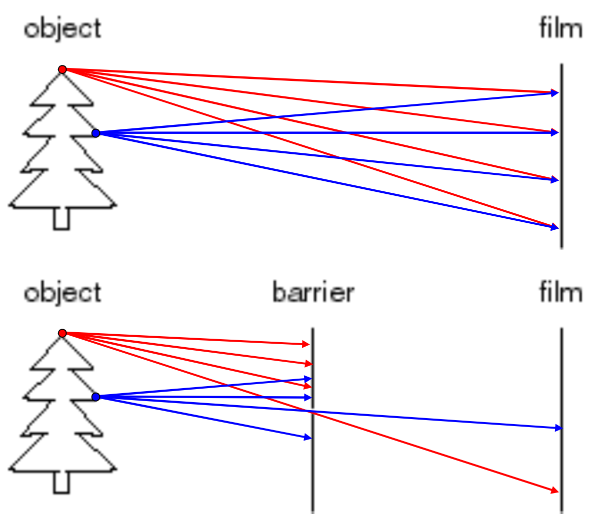
slide from Computer Vision Lecture Notes Trevor Darrell

Image Formation

- Projection of 3-D world onto 2-D image plane

- Two crucial questions :
  - What determines the *position* of a 3-D object point on the 2-D image plane?
  - What determines the *brightness* of a 3-D object point on the 2-D image plane?

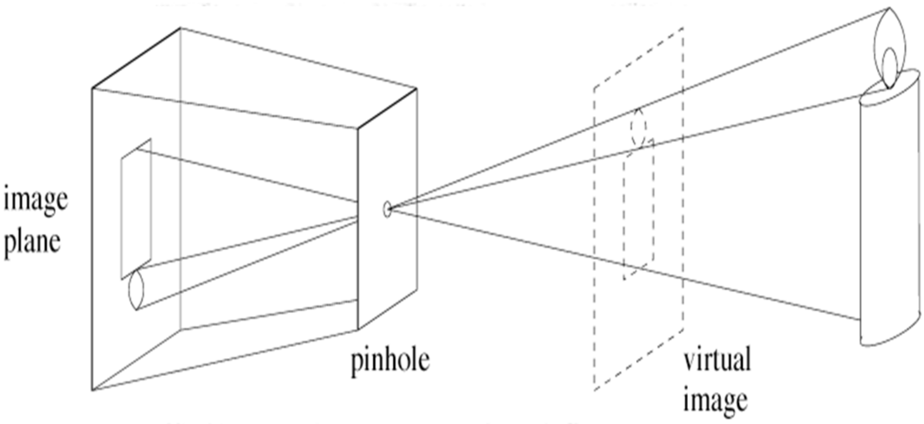
Image Formation



slide from Computer Vision Lecture Notes Trevor Darrell

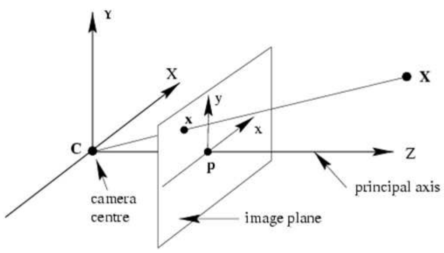
Camera Models

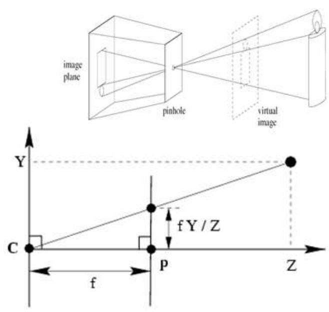
- Pinhole camera model



Camera Models

- Pinhole camera model

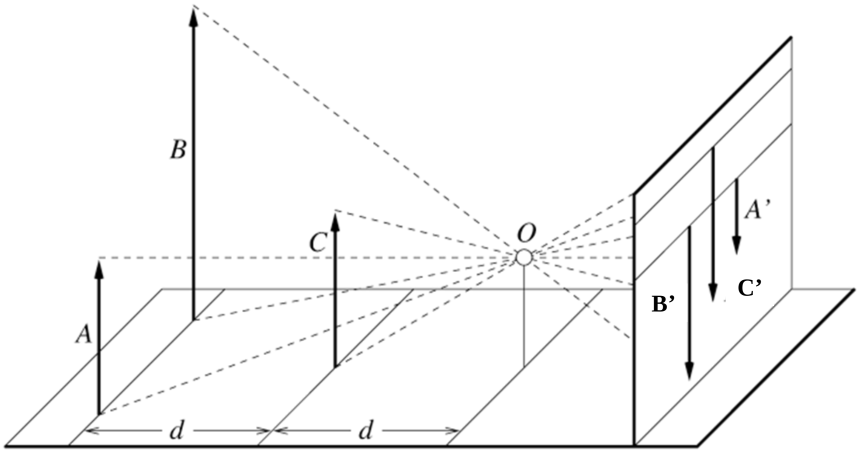




$$x = f\frac{X}{Z},\ y = f\frac{Y}{Z}$$

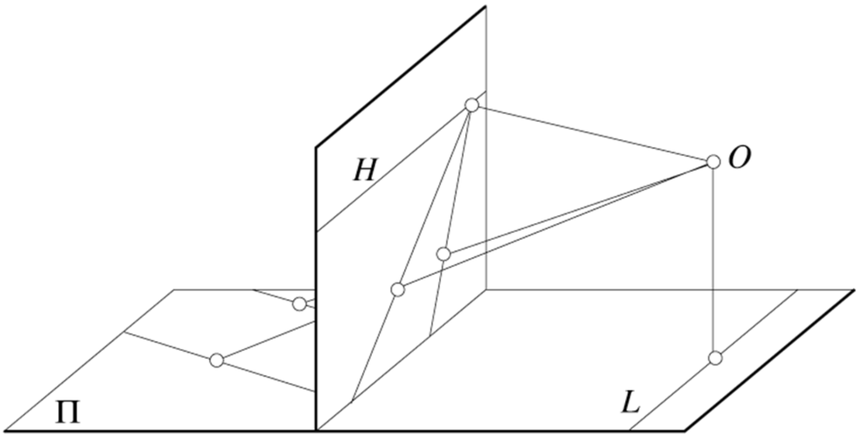
Camera Models

- In pin-hole camera, distant objects are observed smaller



Camera Models

- In pin-hole camera, parallel lines meet



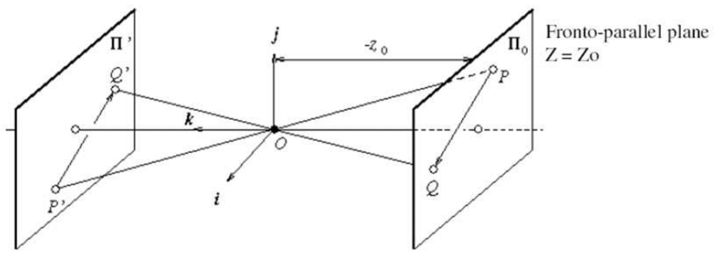
Perspective Projection

$$x = f\frac{X}{Z}, \; y = f\frac{Y}{Z}$$

- Pin-hole camera model is called *perspective projection*
- It is also possible to make approximations to perspective projection
  - Affine : Scene points are planar
  - Weak-perspective : Scene is approximated by a plane and assumed to be far away from camera
  - Orthographic : Scene is approximated to be planar and far away from camera and camera distance does not change

Affine Projection

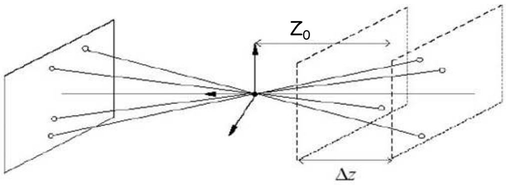
- If all scene points are on a plane



$$x = -f\frac{X}{Z_0}, \; y = -f\frac{Y}{Z_0} \quad \Rightarrow m \equiv \frac{f}{Z_0} \quad \Rightarrow x = -mX, \; y = -mY$$

Weak-perspective Projection

- Now, <u>assume</u> all scene points are on a plane

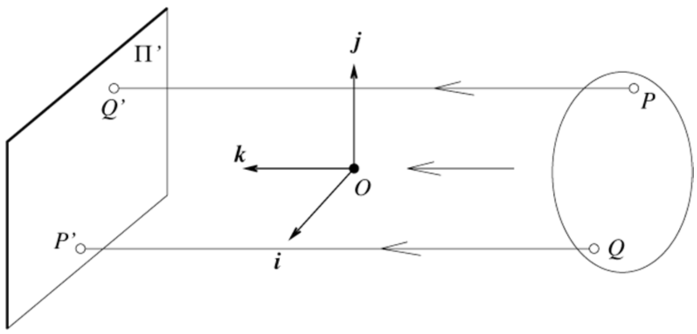


$$x \cong -f\frac{X}{Z_0}\,,\; y \cong -f\frac{Y}{Z_0} \quad \Rightarrow x \cong -mX\,,\; y \cong -mY$$

- This assumption can only be justified, if *scene depth range* is small compared to average distance from camera
  - Δz is small wrt Z_0

Orthographic Projection

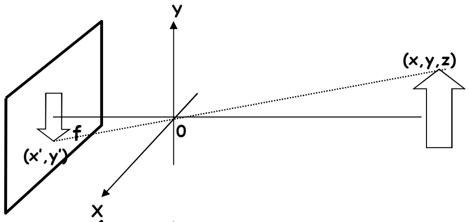
- If scene depth range is small wrt average depth <u>and</u> camera distance remain at a <u>constant distance</u> (i.e. Z_0 is contant)
  - Choose m=-1 ➔ x=X , y=Y



Projections : Summary

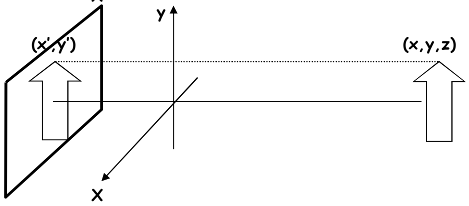
Perspective :

$$x' = f\frac{x}{z}, y' = f\frac{y}{z}$$



Orthographic :

$$x' = x, y' = y$$

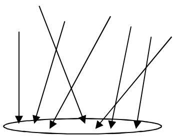


Perspective projection is a more realistic projection for (pin-hole) camera recordings

If *depth range* is small compared to average distance from the camera, orthographic is also a good approximation

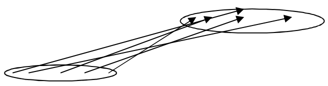
Brightness :

Two different brightness concepts :



- **Image brightness** : *irradiance*

  Light power per unit area <u>falling</u> on a (image) surface



- **Scene brightness** : *radiance*

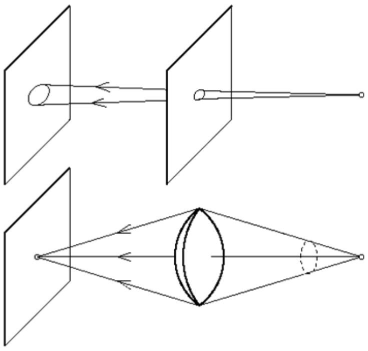
  Light power per unit area <u>emitted</u> into a solid angle from a (object) surface

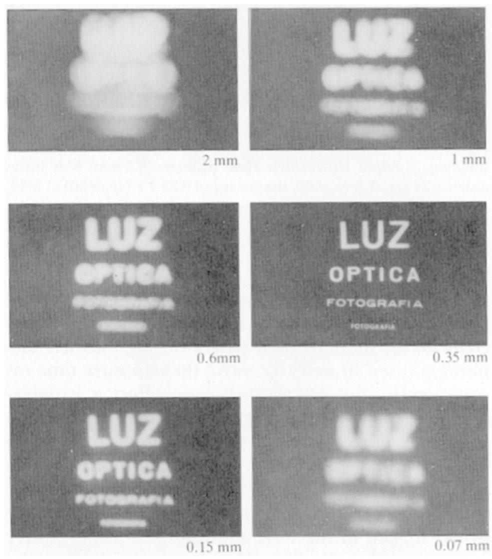
Image and scene brightness are proportional to each other

Pinhole camera needs non-zero diameter for enough light

Lenses :

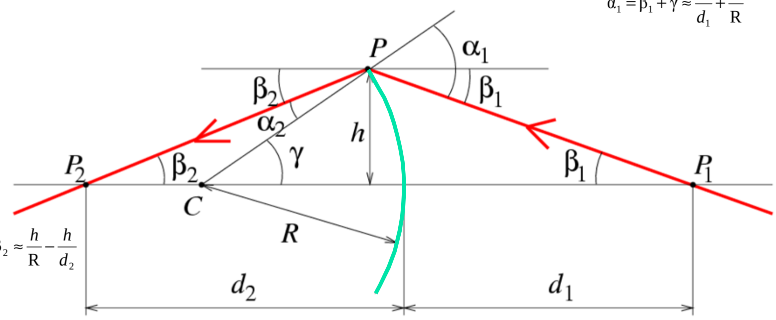
A pin-hole camera needs light





Lenses : Paraxial (or first-order) optics

$$\alpha_1 = \beta_1 + \gamma \approx \frac{h}{d_1} + \frac{h}{R}$$



$$\alpha_2 = \gamma - \beta_2 \approx \frac{h}{R} - \frac{h}{d_2}$$

Snell's law:

$$n_1 \sin \alpha_1 = n_2 \sin \alpha_2$$

Small angles:

$$n_1 \alpha_1 \approx n_2 \alpha_2$$

$$n_1\left(\frac{h}{d_1} + \frac{h}{R}\right) = n_2\left(\frac{h}{R} - \frac{h}{d_2}\right)$$

$$\frac{n_1}{d_1} + \frac{n_2}{d_2} = \frac{n_2 - n_1}{R}$$

Note that the relation is independent of β_1 and β_2 → all rays pass from *P1*, also pass *P2*.

slide from Computer Vision Lecture Notes by Marc Pollefeys

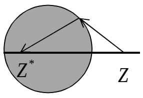
20

Lenses : Thin Lenses

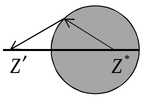
spherical lens surfaces;
thickness << radii; same refractive index on both sides

$$\boxed{\frac{n_1}{d_1}+\frac{n_2}{d_2}=\frac{n_2-n_1}{R}}$$

$$\frac{1}{Z}+\frac{n}{Z^*}=\frac{n-1}{R}$$

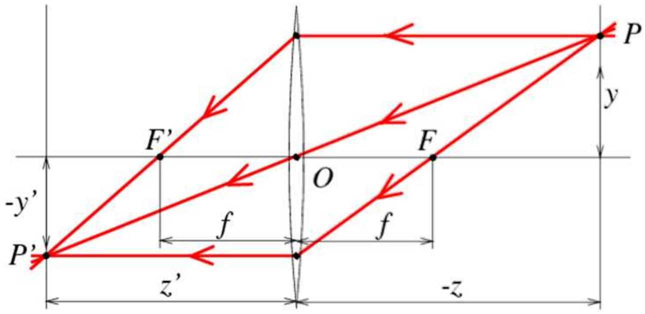


$$\frac{n}{Z^*}+\frac{1}{Z'}=\frac{1-n}{R}$$



$$\frac{n}{Z^*}=\frac{n-1}{R}-\frac{1}{Z} \qquad \frac{n}{Z^*}=\frac{1-n}{R}-\frac{1}{Z'}$$

$$\frac{n-1}{R}-\frac{1-n}{R}=\frac{1}{Z}-\frac{1}{Z'}$$

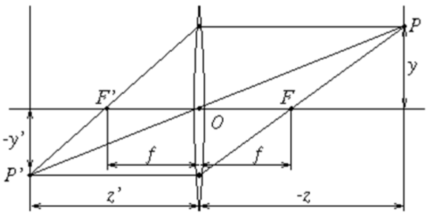


$$\boxed{\frac{1}{z'}-\frac{1}{z}=\frac{1}{f}} \quad \text{and} \quad \boxed{f=\frac{R}{2(n-1)}}$$

slide from Computer Vision Lecture Notes by Marc Pollefeys

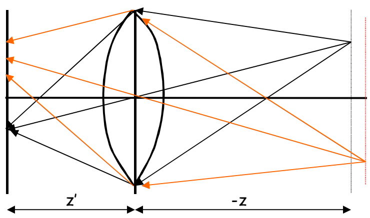
Lenses :

An ideal thin lens produces the same projection with a pinhole camera, plus some finite amount of light.



$$\frac{1}{z'}+\frac{1}{-z}=\frac{1}{f}$$

Once you focus for one distance z, points on other distances will be blurred.



Lenses : Deviations from the lens model

3 assumptions :

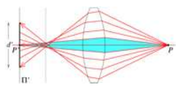
1. all rays from a point are focused onto 1 image point
2. all image points in a single plane
3. magnification is constant

deviations from this ideal are *aberrations*

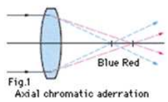
2 types of aberrations:

geometrical : small for paraxial rays

study through 3rd order optics $\sin(\theta) \approx \theta - \frac{\theta^3}{6}$

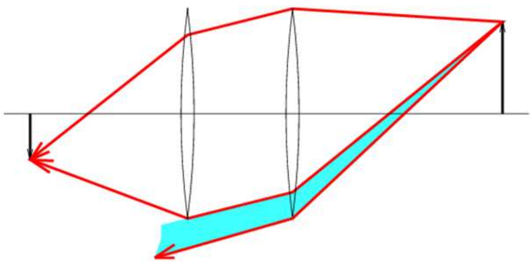


chromatic : refractive index function of wavelength



Lenses :

Vignetting : Brightness drop in image periphery for compound lenses



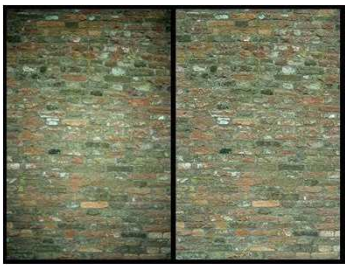
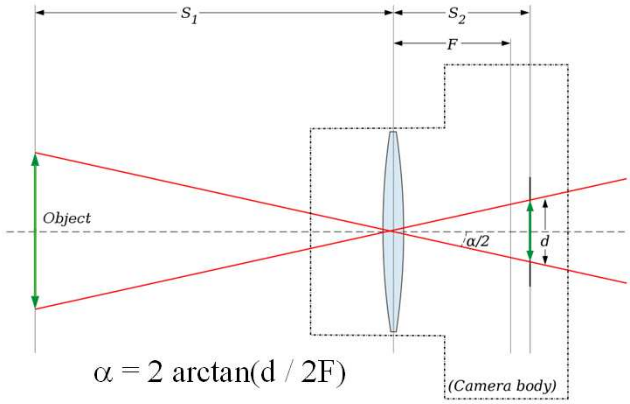


Figure from http://www.vanwalree.com/optics/vignetting.html

Camera Field of View

Angular measure of the portion of 3D space seen by the camera

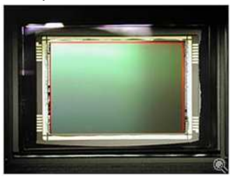


$$\alpha = 2 \arctan(d / 2F)$$

Image from http://en.wikipedia.org/wiki/Angle_of_view

Image Sensing :

Light photons striking a suitable (vacuum or semi-conductor device) surface generate electron-hole pairs which are measured to determine the irradiance.



Quantum efficiency : ratio of electron flux to incident photon flux & depends on energy (wavelength) of photon

Solid-state devices almost ideal for some wavelengths

Photographic films have poor quantum efficiency

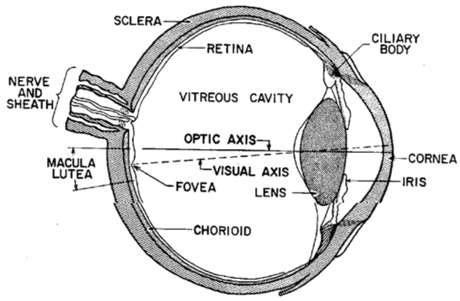
Quantization of Image

Electrons should be measured/averaged at some predefined regions on the image sensor -> Spatial quantization

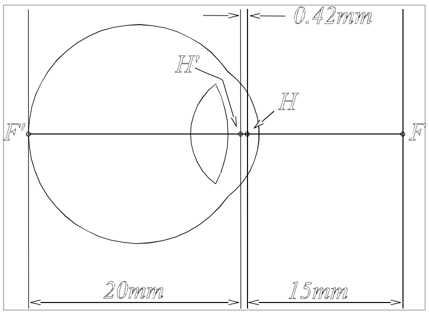
These regions can be square, rectangular or hexagonal

Each predefined region represents a *pixel* (picture element) location and the quantized values are *pixel* values (usually 0 to 255)

The Human Eye

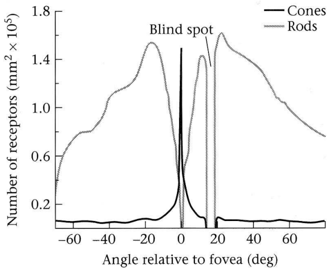


Reproduced by permission, the American Society of Photogrammetry and Remote Sensing. A.L. Nowicki, "Stereoscopy." Manual of Photogrammetry, Thompson, Radlinski, and Speert (eds.), third edition, 1966.



Helmoltz's Schematic Eye

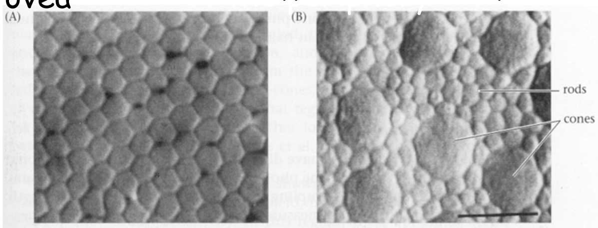
The distribution of rods and cones across the retina



Reprinted from Foundations of Vision, by B. Wandell, Sinauer Associates, Inc., (1995). © 1995 Sinauer Associates, Inc.

Cones in the fovea

3 types of cones that yield color perception



Reprinted from Foundations of Vision, by B. Wandell, Sinauer Associates, Inc., (1995). © 1995 Sinauer Associates, Inc.