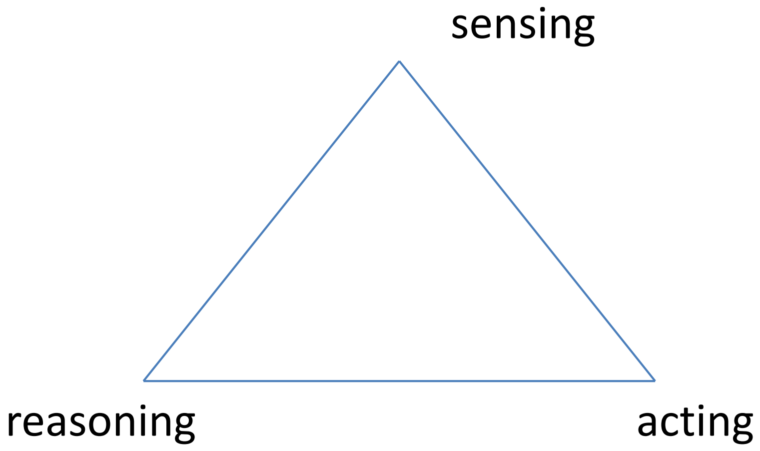
Perception for Robots

I.
BASICS
Thanks to Peter Corke for the use of some slides

What is a robot

- For the purposes of this class a robot is a goal oriented machine that can sense, reason and act.



BASIC QUESTIONS

- Where am I
- Where are you
- What are you
- How do I get there
- How to achieve a task

Where am I

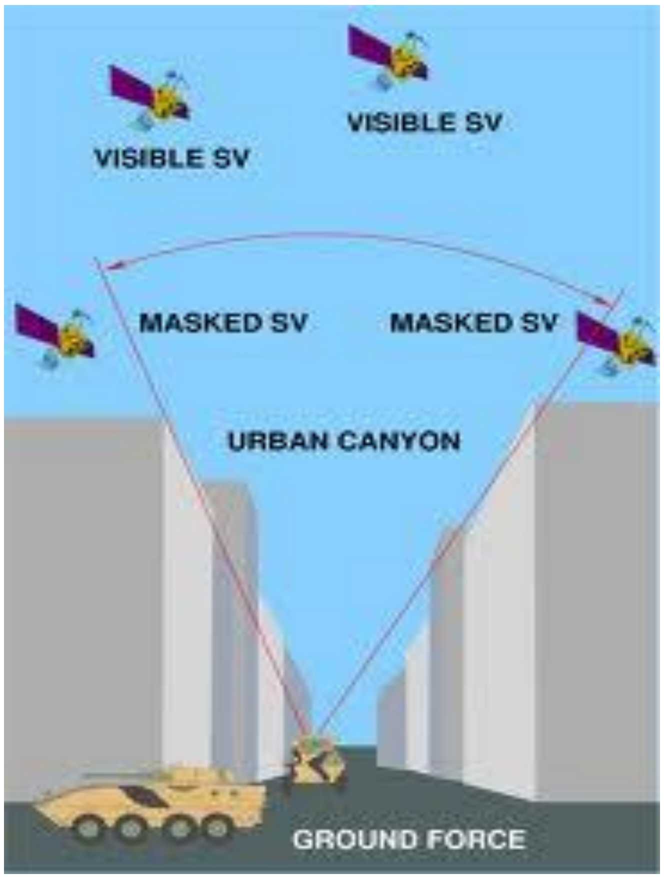
- Why not use GPS?

GPS is not perfect and has severe limitations is environments where robots are needed:

--- cities, mines, industrial sites, underwater,
deep forest.

It only tells where I am

Urban Canyon Problem



Industrial sites, mines

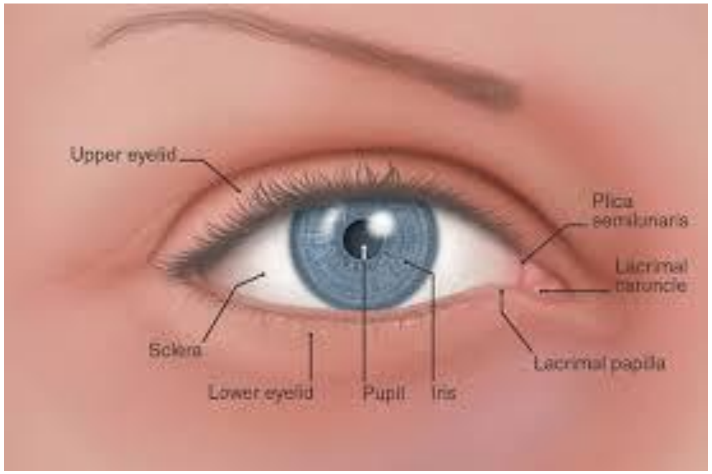
Underwater, deep forest

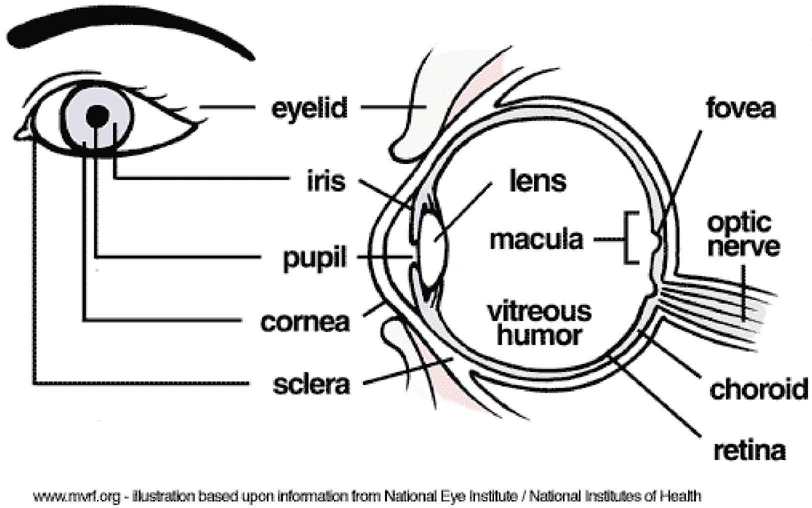
Humans and animals have a number of senses

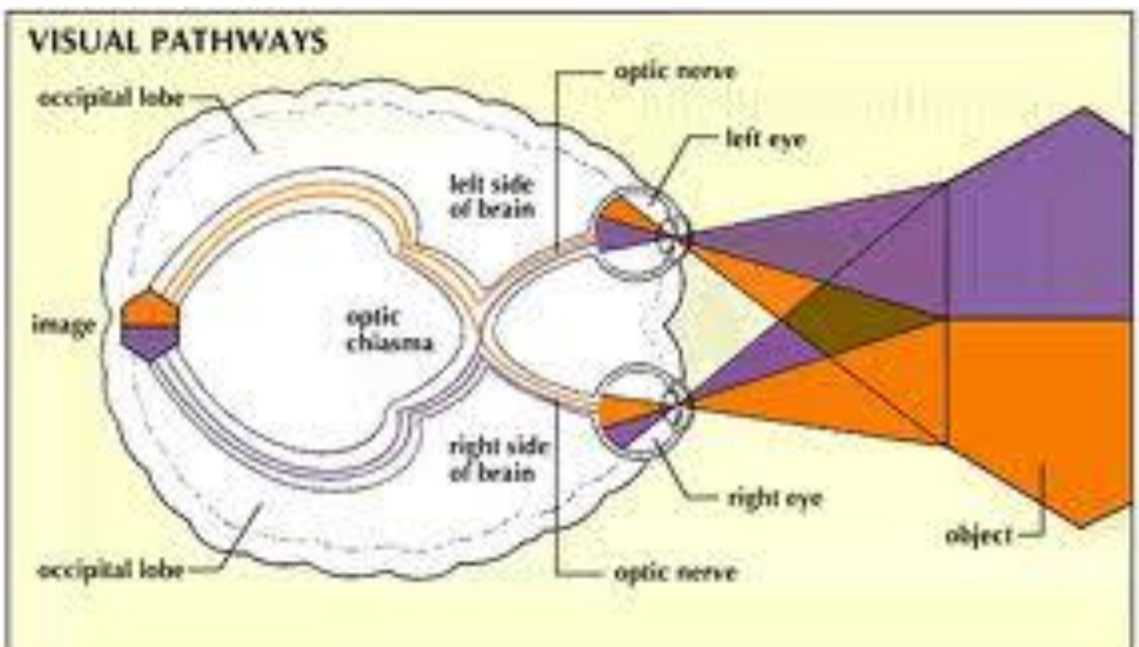
- Sight
- Hearing
- Touch
- Smell
- Taste
- Balance

- Echolocation: bats, electric fields: sharks, compass: birds

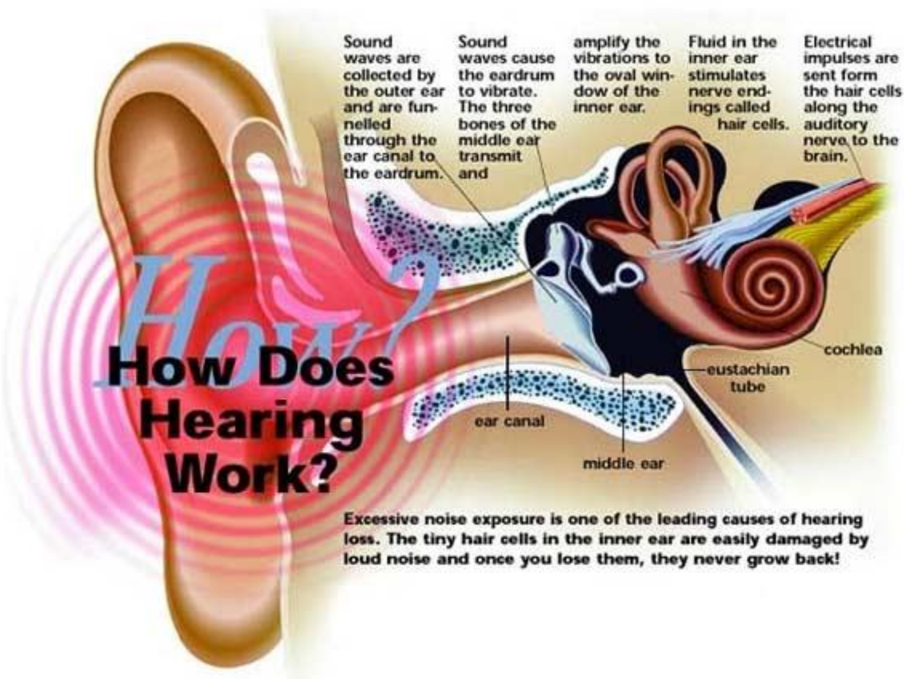
Vision


Upper eyelid
Plica semilunaris
Lacrimal caruncle
Sclera
Lower eyelid
Pupil
Iris
Lacrimal papilla

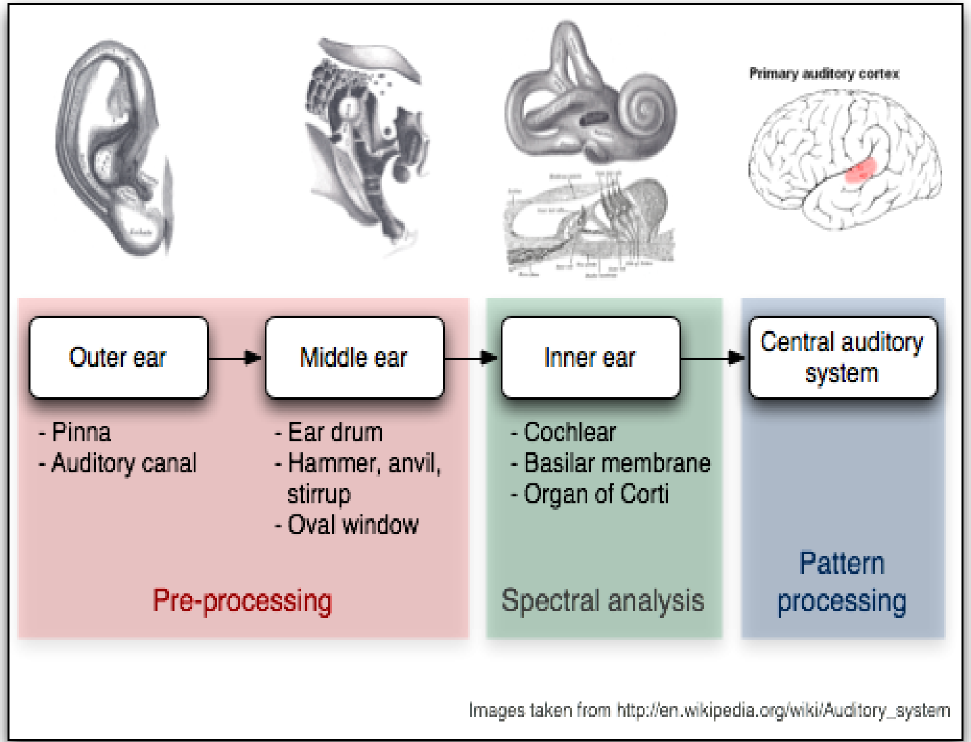

eyelid
iris
pupil
cornea
sclera
lens
macula
vitreous humor
fovea
optic nerve
choroid
retina
www.mvrf.org - illustration based upon information from National Eye Institute / National Institutes of Health


VISUAL PATHWAYS
occipital lobe
optic nerve
left eye
left side of brain
image
optic chiasma
right side of brain
right eye
object
occipital lobe
optic nerve

Hearing

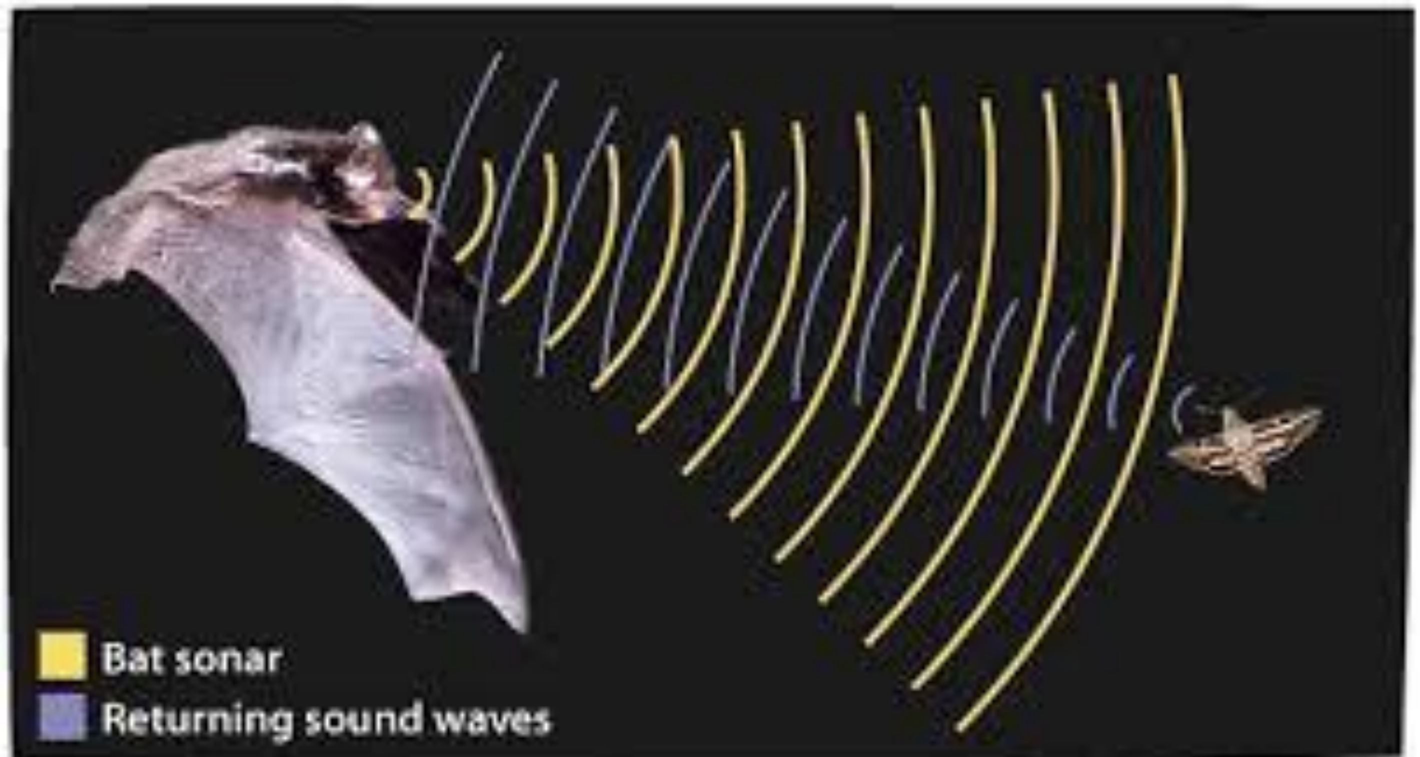


Sound waves are collected by the outer ear and are funnelled through the ear canal to the eardrum.
Sound waves cause the eardrum to vibrate. The three bones of the middle ear transmit and
amplify the vibrations to the oval window of the inner ear.
Fluid in the inner ear stimulates nerve endings called hair cells.
Electrical impulses are sent form the hair cells along the auditory nerve to the brain.
How?
How Does Hearing Work?
cochlea
eustachian tube
ear canal
middle ear
Excessive noise exposure is one of the leading causes of hearing loss. The tiny hair cells in the inner ear are easily damaged by loud noise and once you lose them, they never grow back!

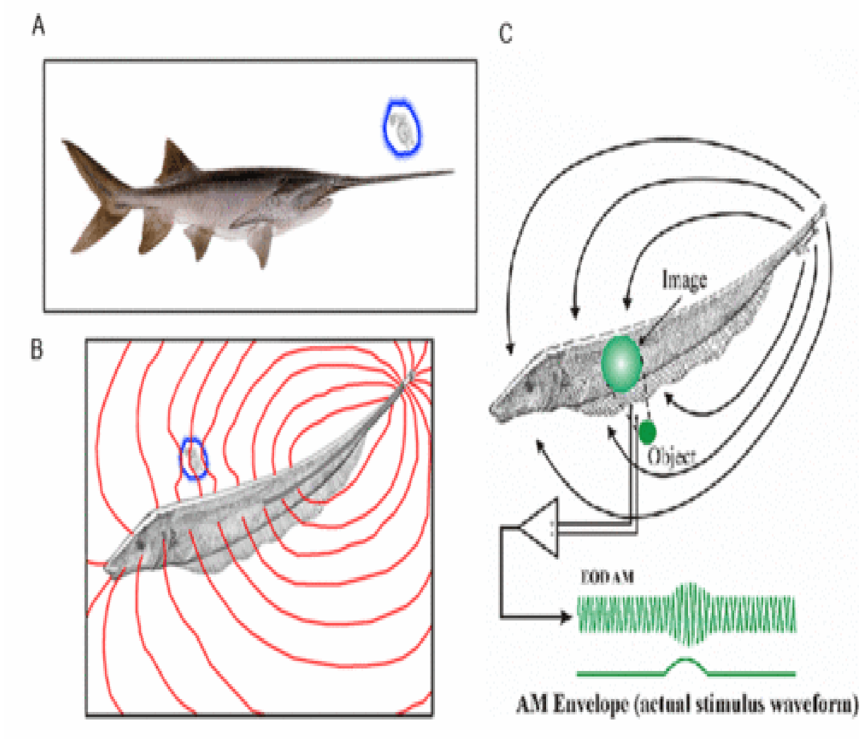


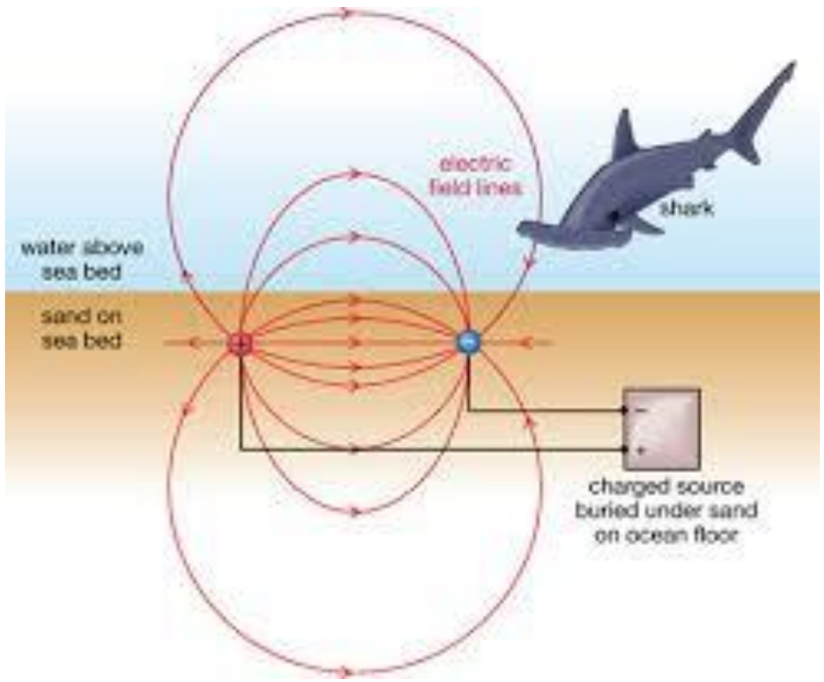
Primary auditory cortex
Outer ear
Middle ear
Inner ear
Central auditory system
- Pinna
- Auditory canal
- Ear drum
- Hammer, anvil, stirrup
- Oval window
- Cochlear
- Basilar membrane
- Organ of Corti
Pre-processing
Spectral analysis
Pattern processing
Images taken from http://en.wikipedia.org/wiki/Auditory_system

Echolocation of bats

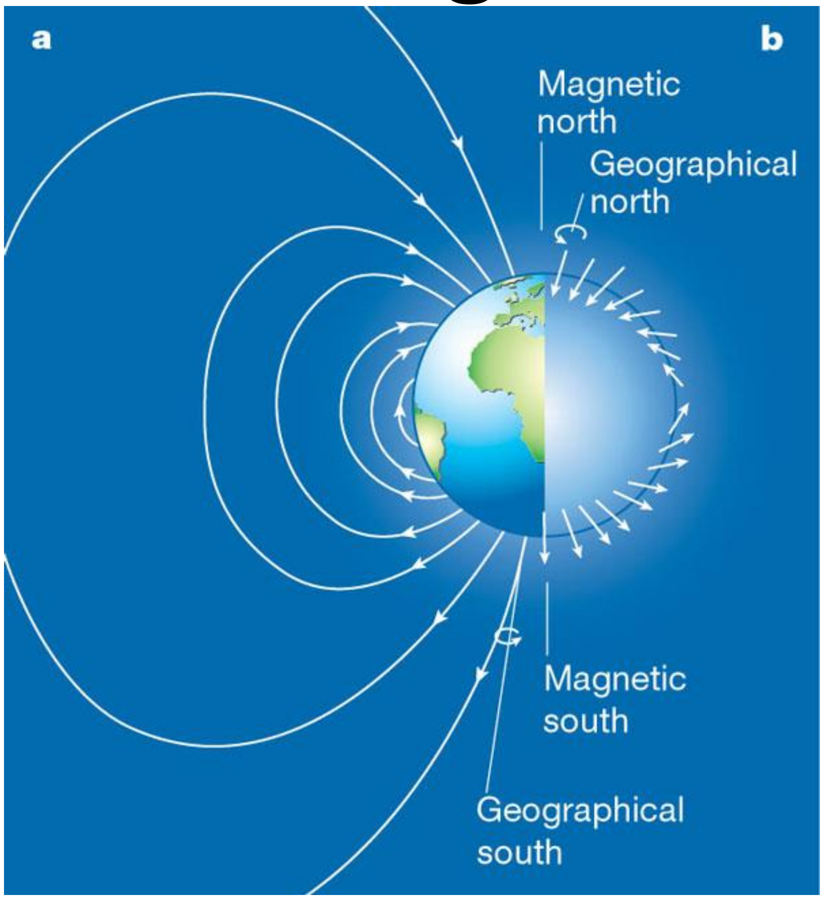


Electric field sensing





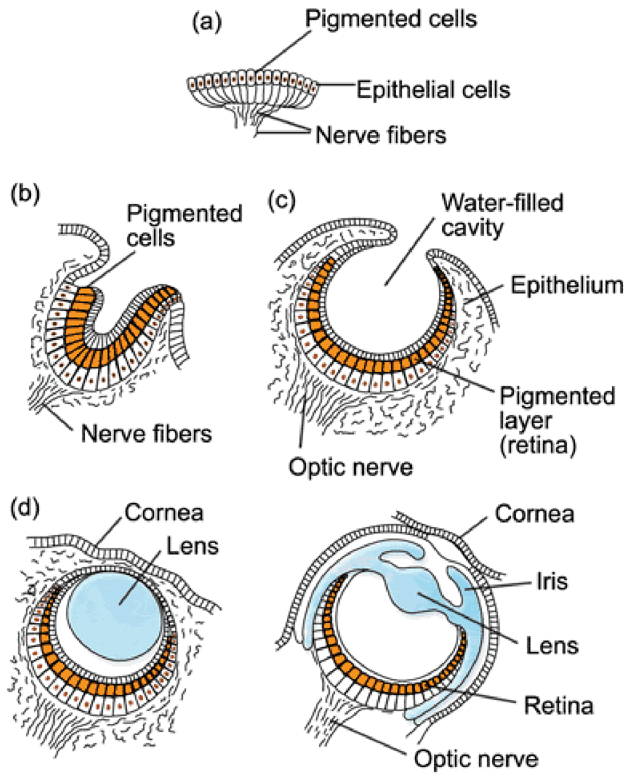
Magnetic field sensing

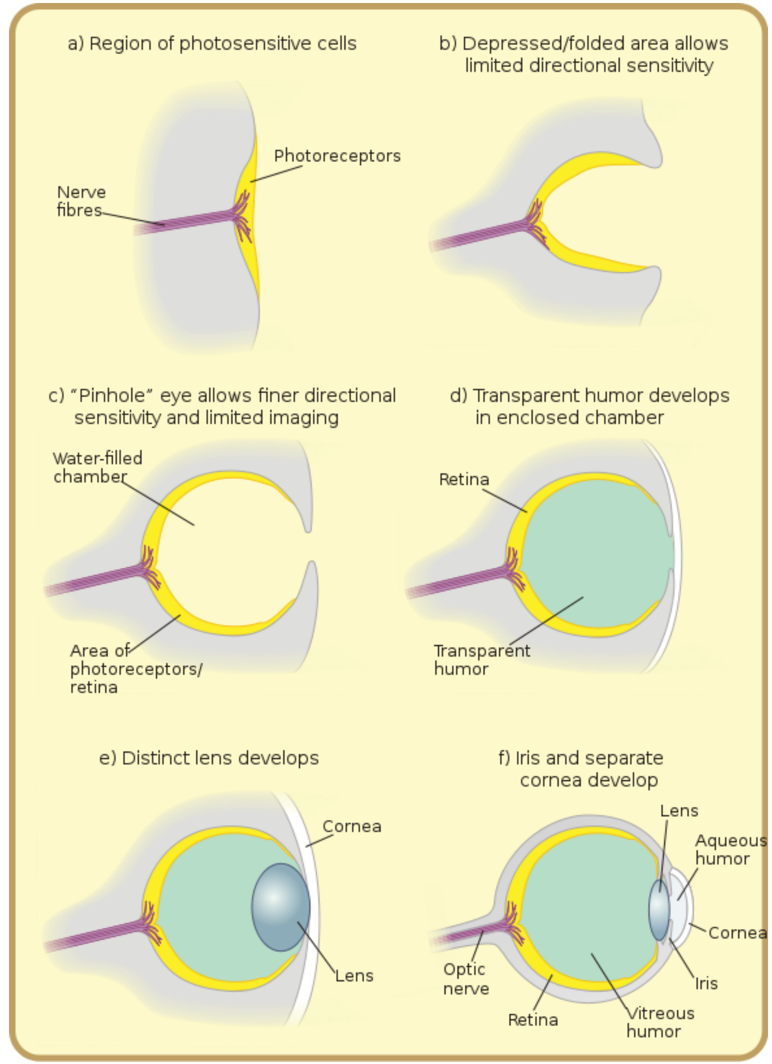


Vision: most powerful sense

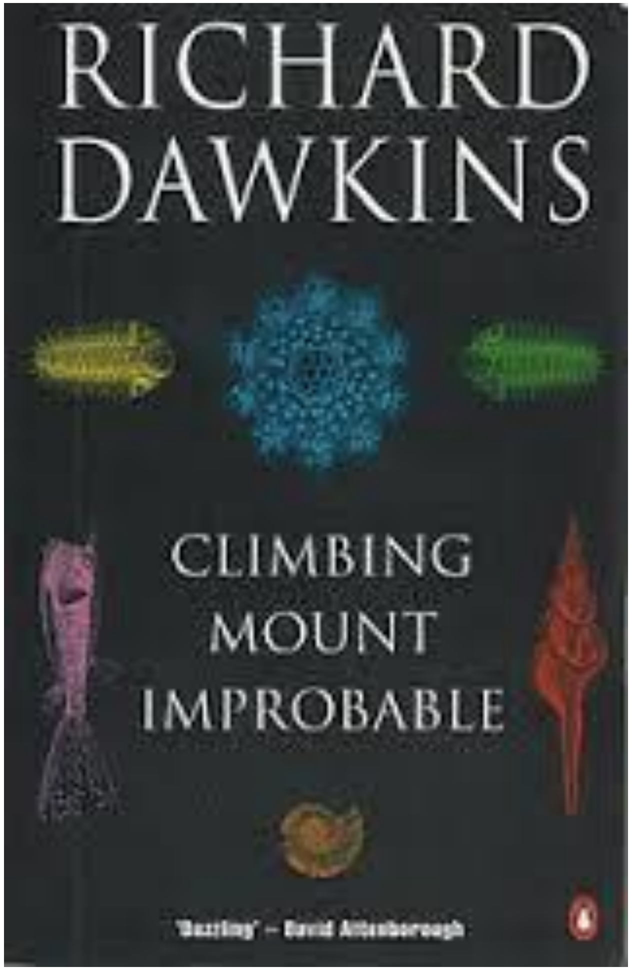
- Essential for survival: finding food, avoiding being food, finding mates
- Long range sensing: beyond our fingertip (vision is our way to touch the world)

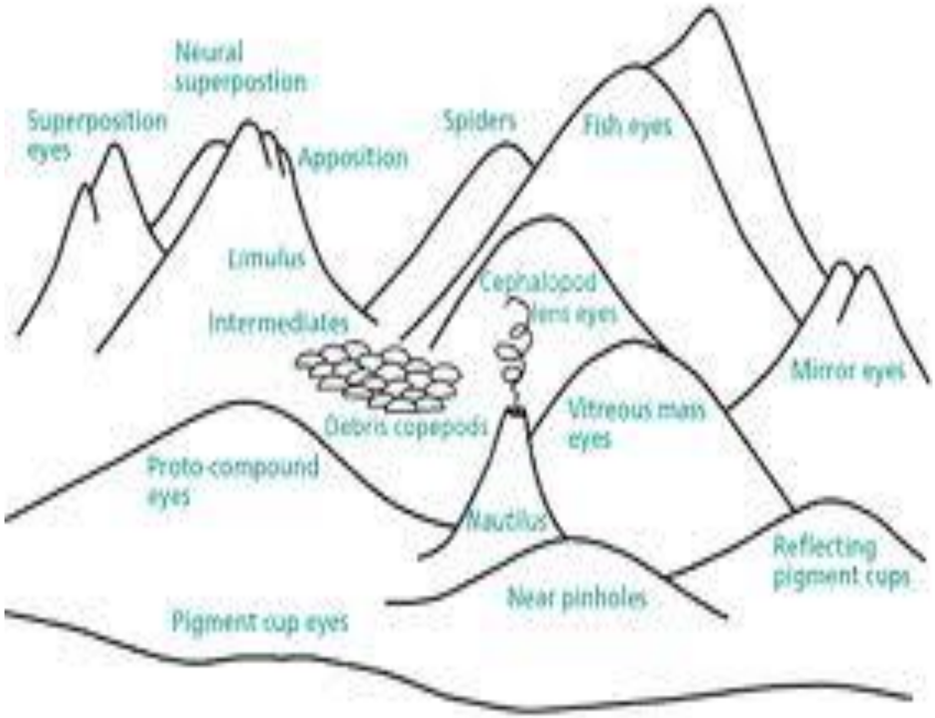
Evolution of the eye
½ billion years





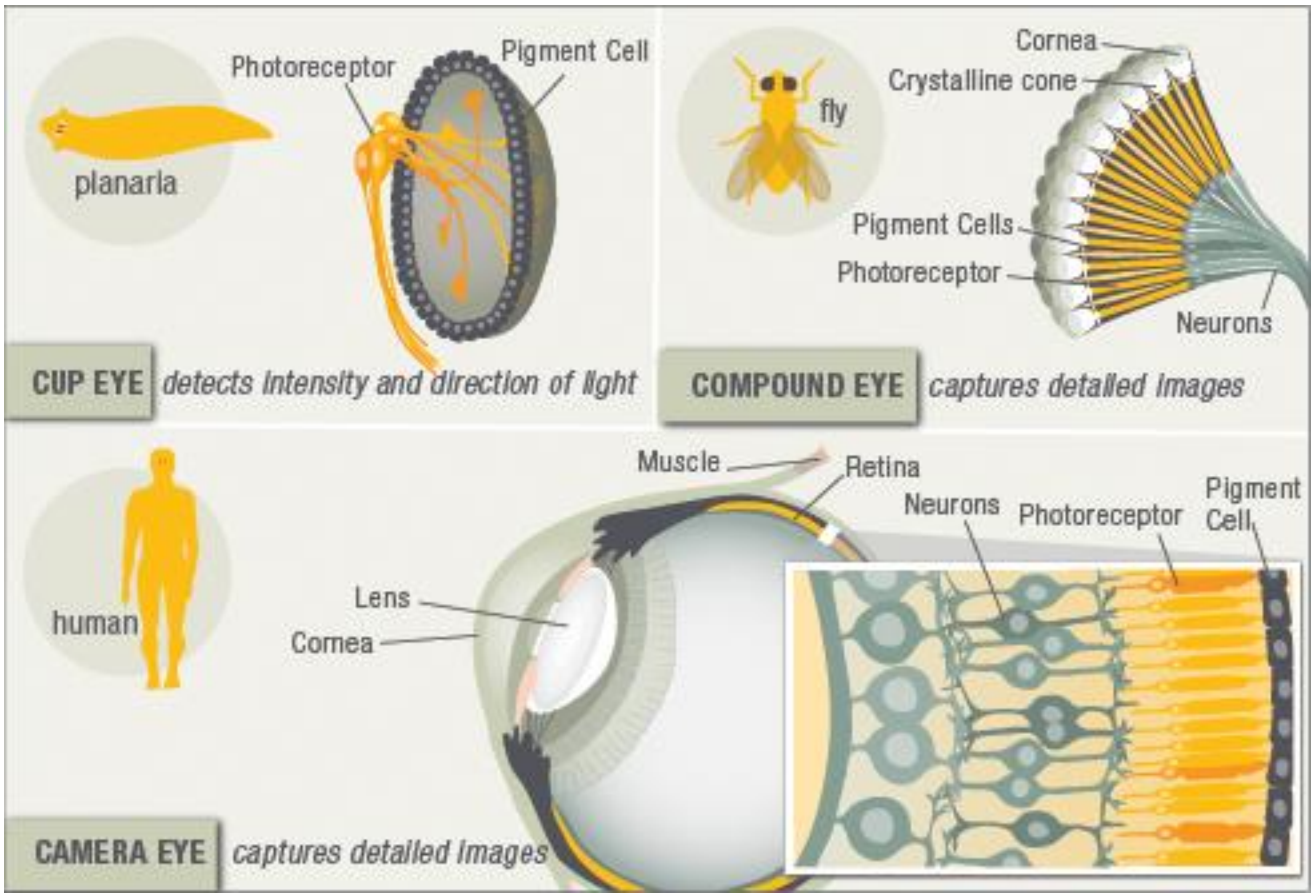
Climbing mount improbable
10 different designs



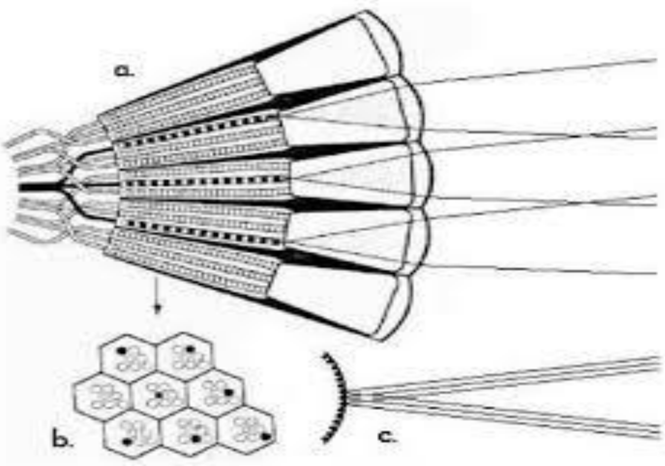


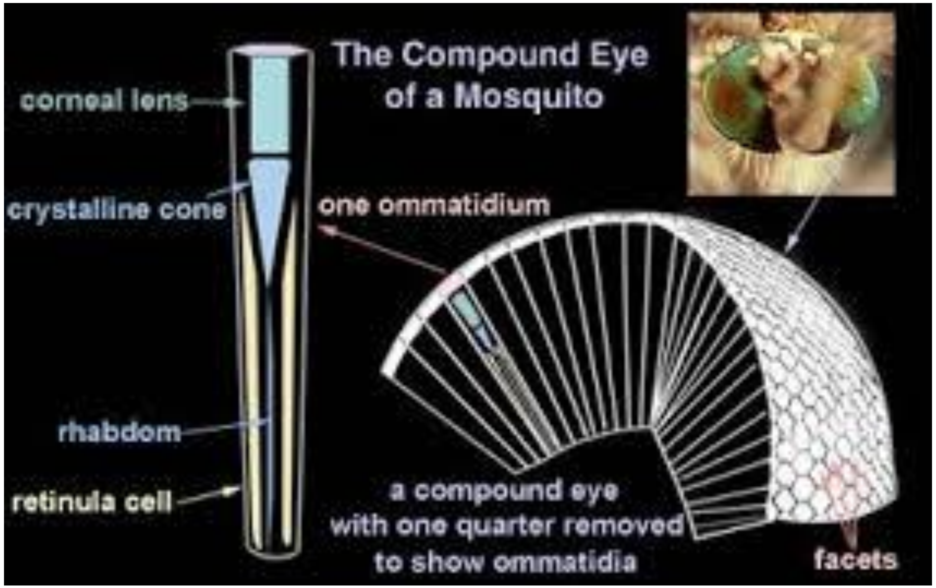
A plethora of eyes

Complex Eyes

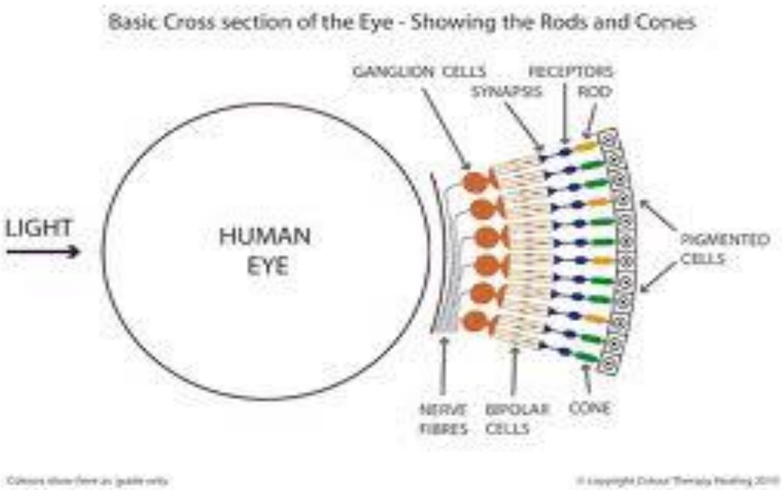


Compound eyes

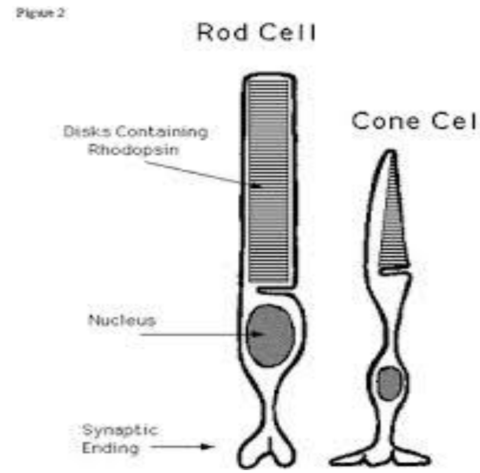




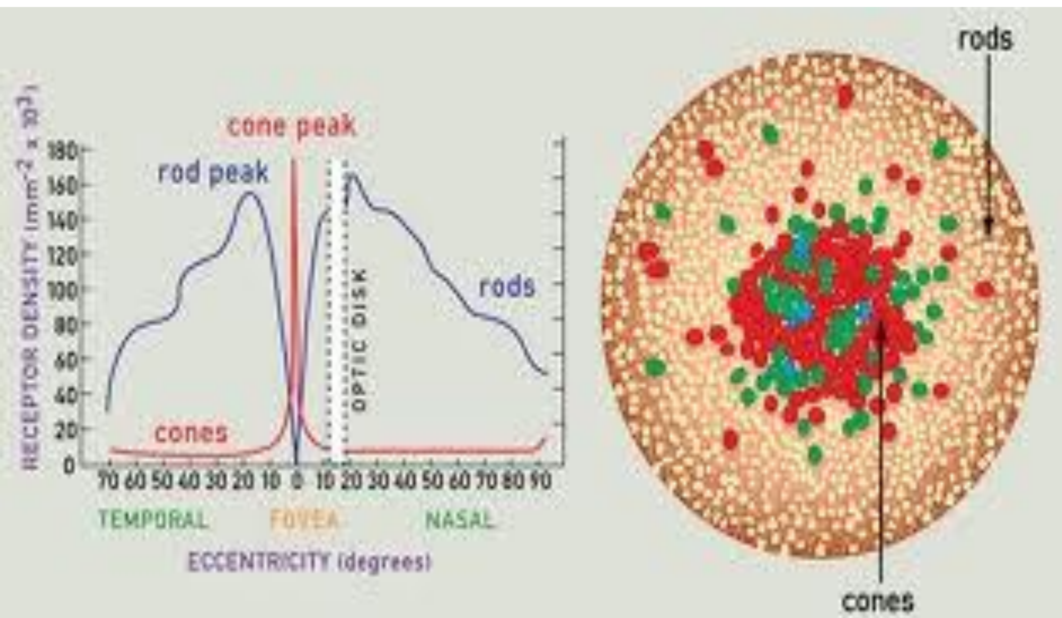
Human Eyes



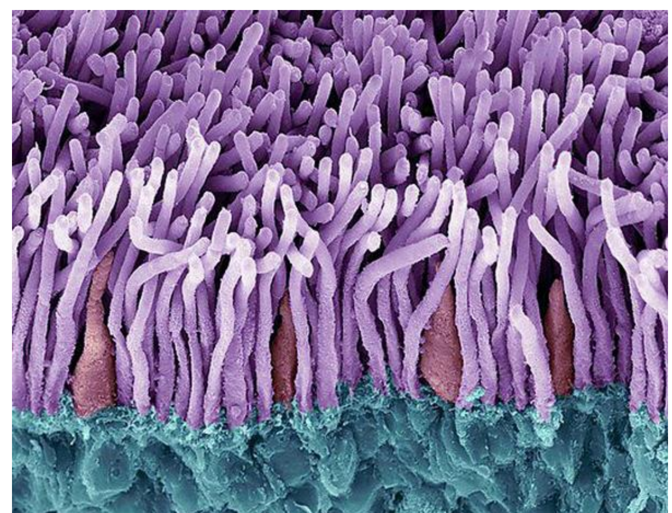
Basic Cross section of the Eye - Showing the Rods and Cones
GANGLION CELLS
RECEPTORS
SYNAPSIS
ROD
LIGHT
HUMAN EYE
PIGMENTED CELLS
NERVE FIBRES
BIPOLAR CELLS
CONE



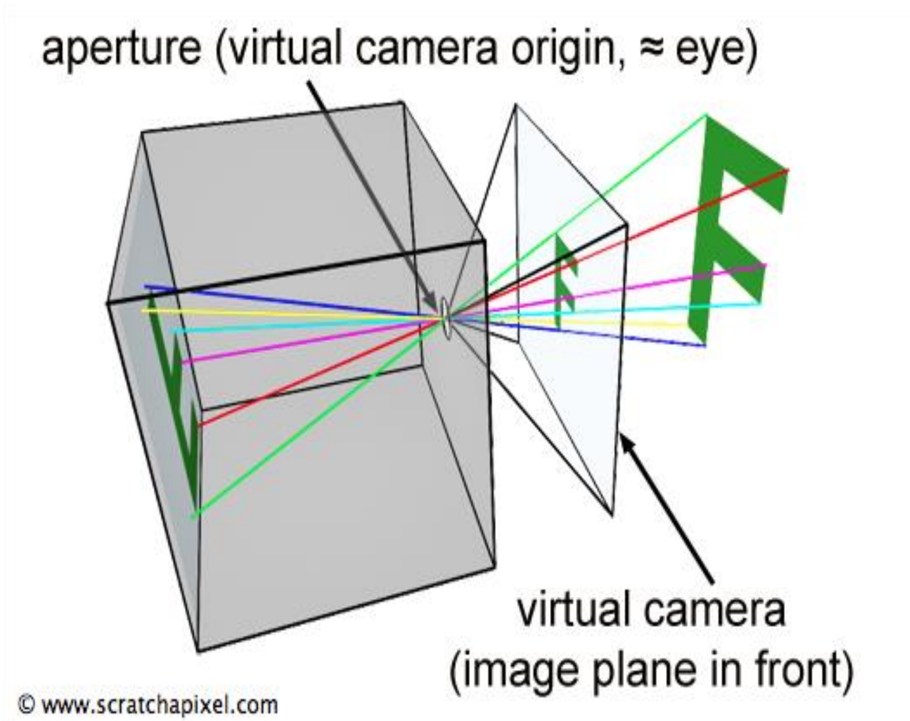
Rod Cell
Cone Cell
Disks Containing Rhodopsin
Nucleus
Synaptic Ending

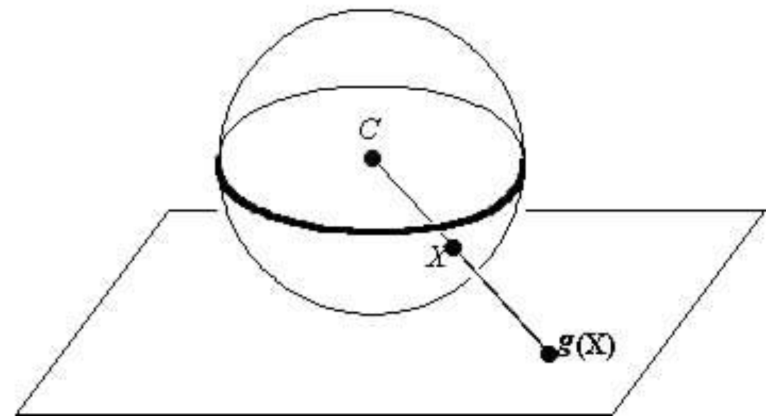


cone peak
rod peak
RECEPTOR DENSITY (mm⁻² x 10³)
180
160
140
120
100
80
60
40
20
0
rods
cones
OPTIC DISK
70 60 50 40 30 20 10 0 10 20 30 40 50 60 70 80 90
TEMPORAL
FOVEA
NASAL
ECCENTRICITY (degrees)
rods
cones



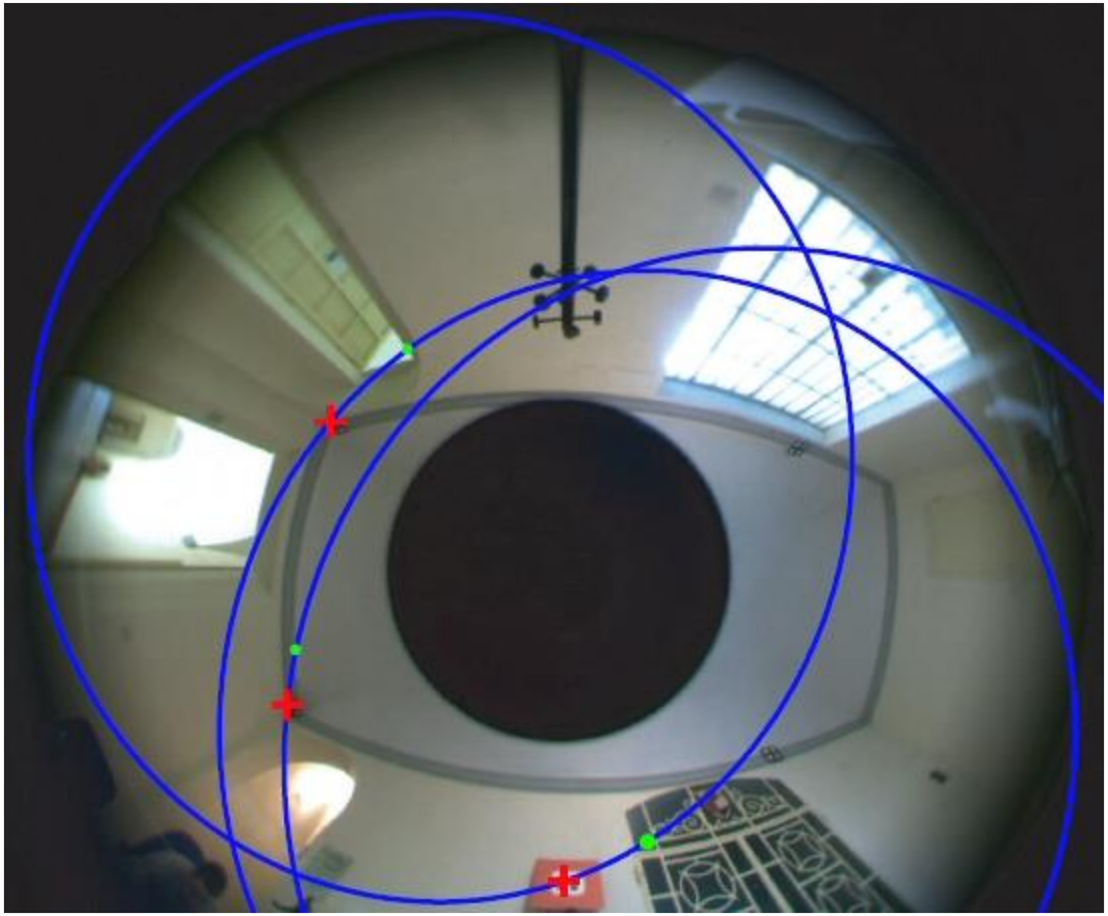
Two kinds of eyes at the top: Camera type or planar Spherical





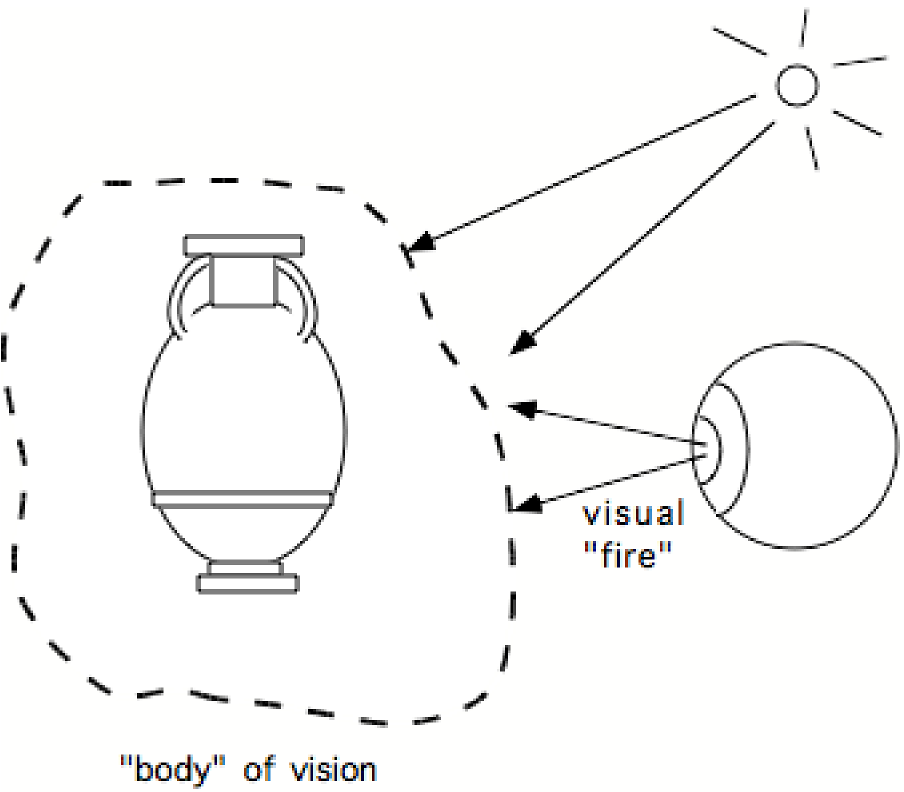
Many cameras in the market

Catadioptric – panoramic images

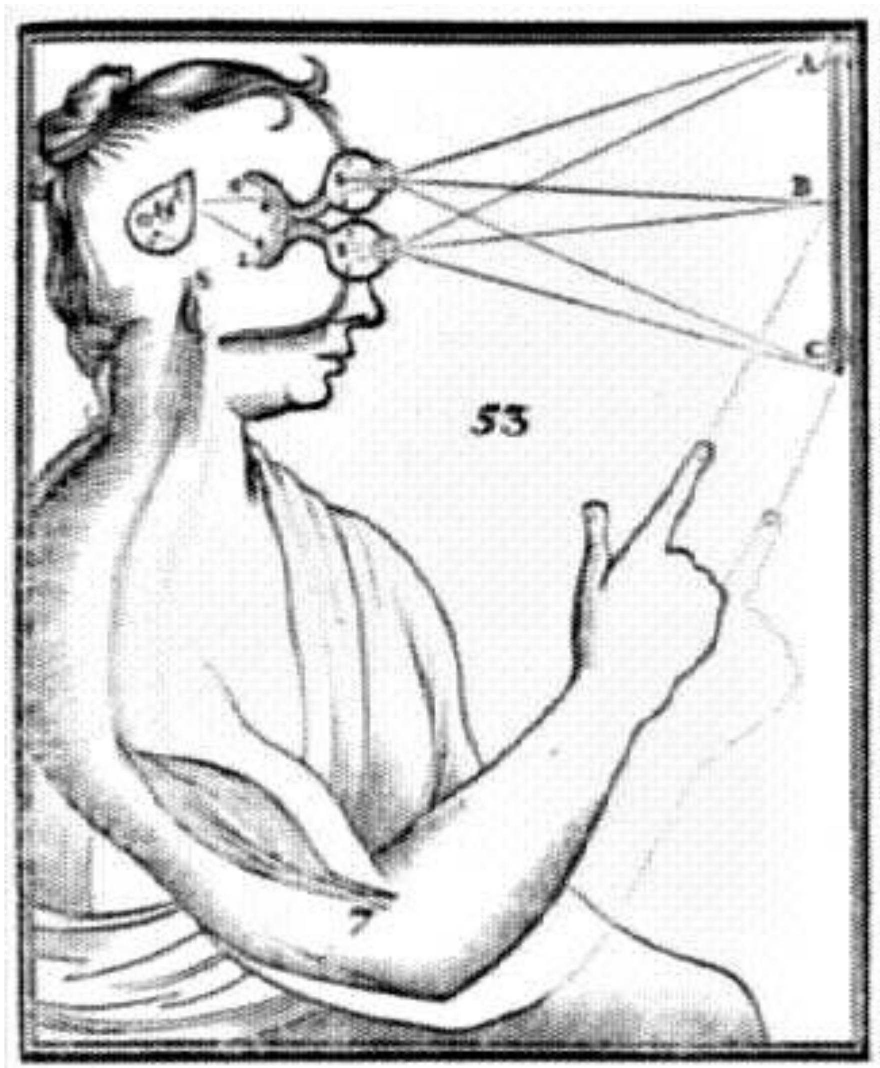


How does Vision work?

- Ancient Greeks: Extramission Theory



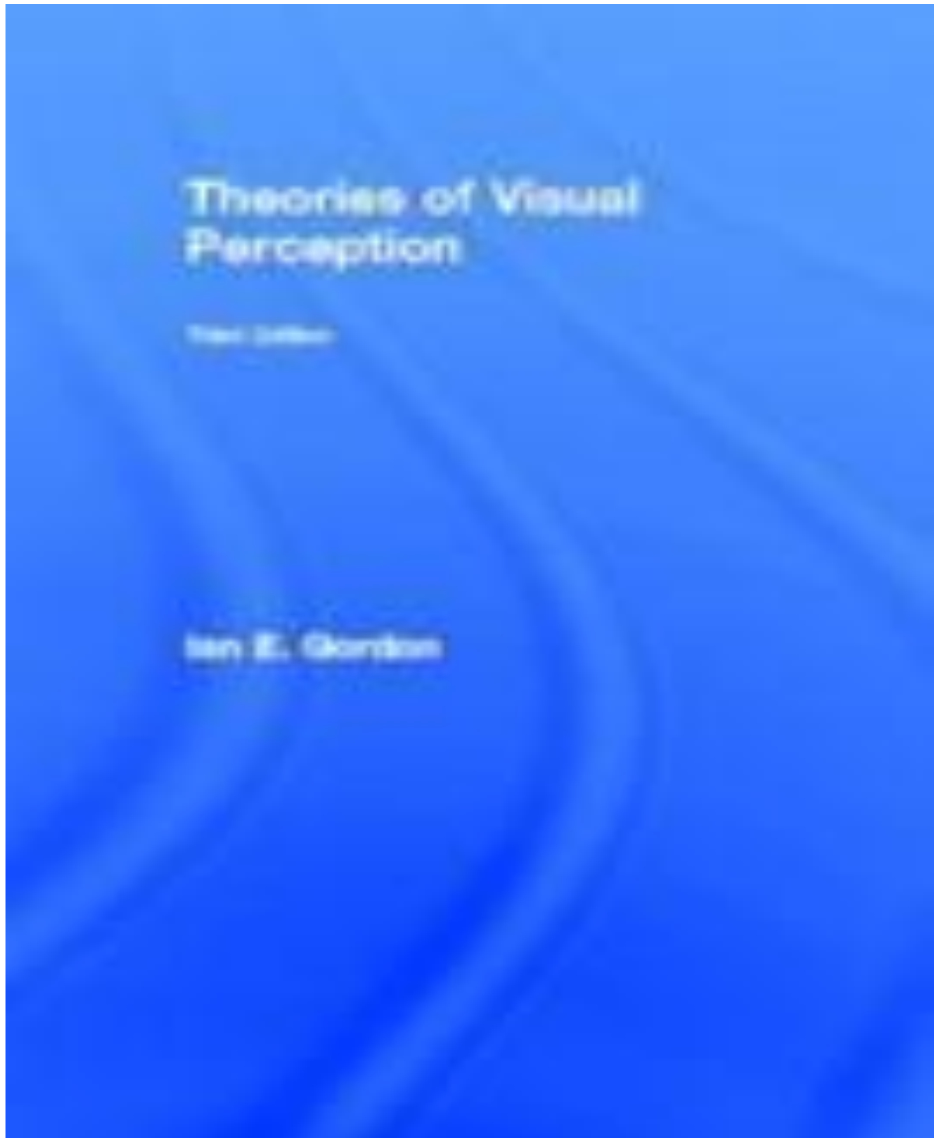
Descartes got it right



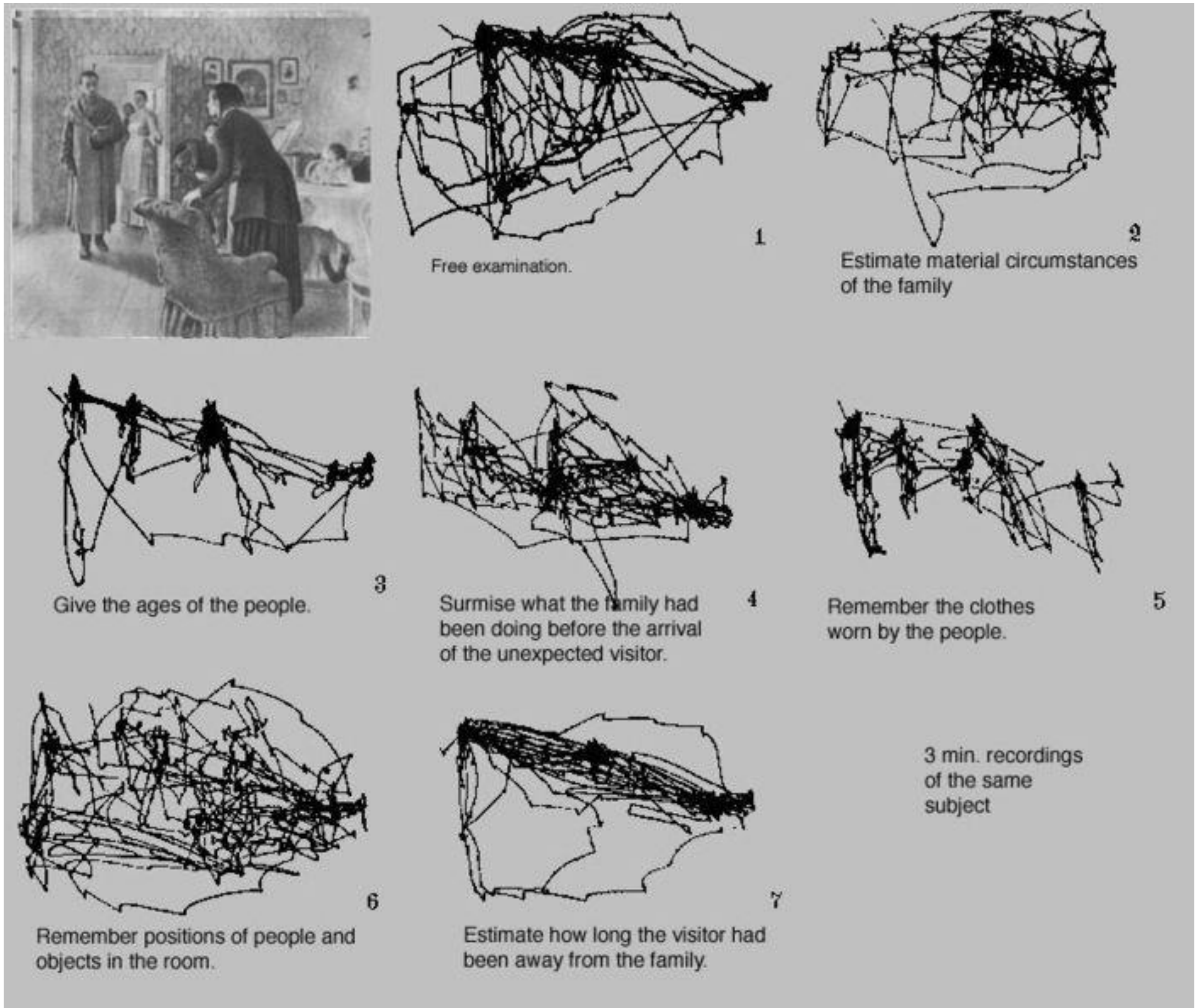
Many theories over the centuries

- The Gestaltists
- Von Helmholz: Unconscious inference
- David Marr: A reconstruction process that tells us where is what.

Theories influenced by the zeitgeist

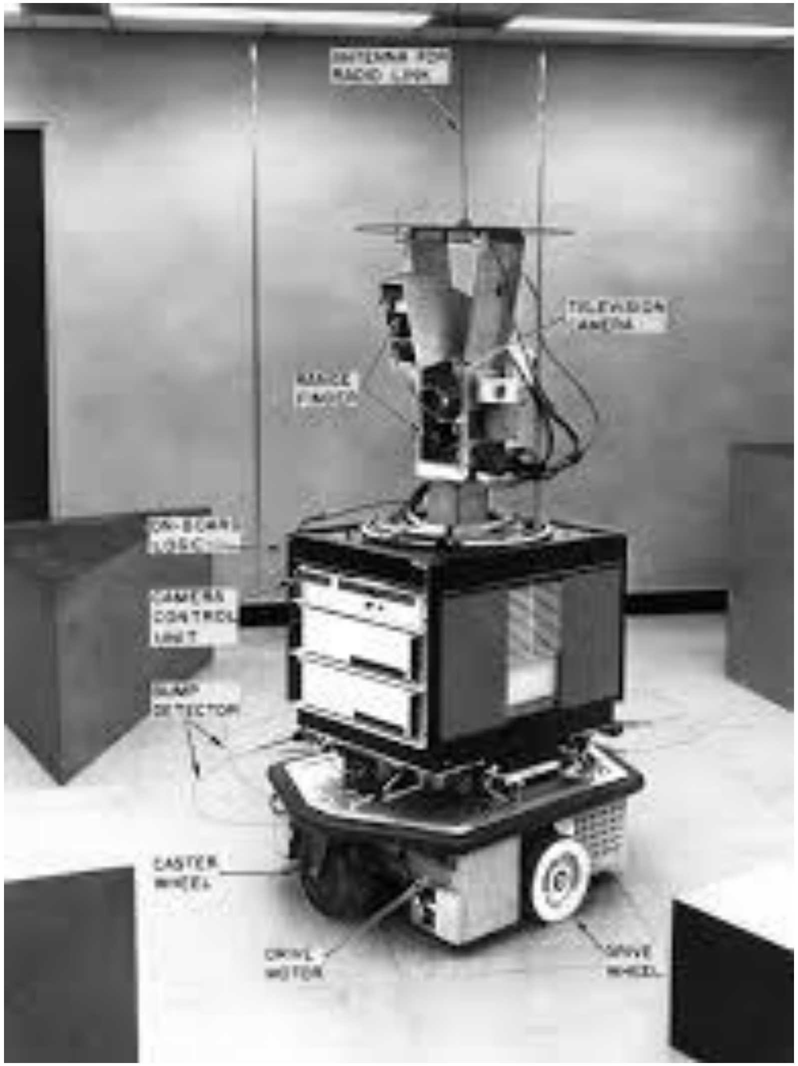


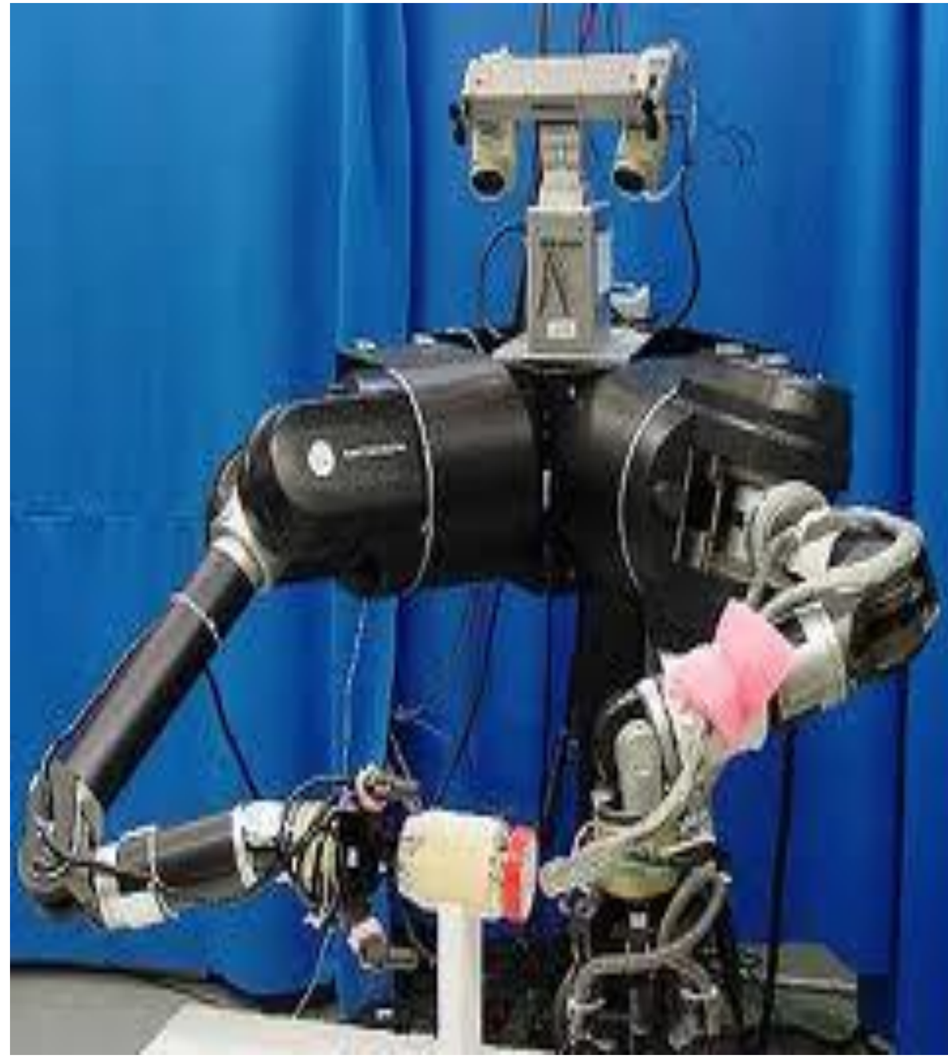
Animal perception is active



Measuring eye movements

Robots with Vision



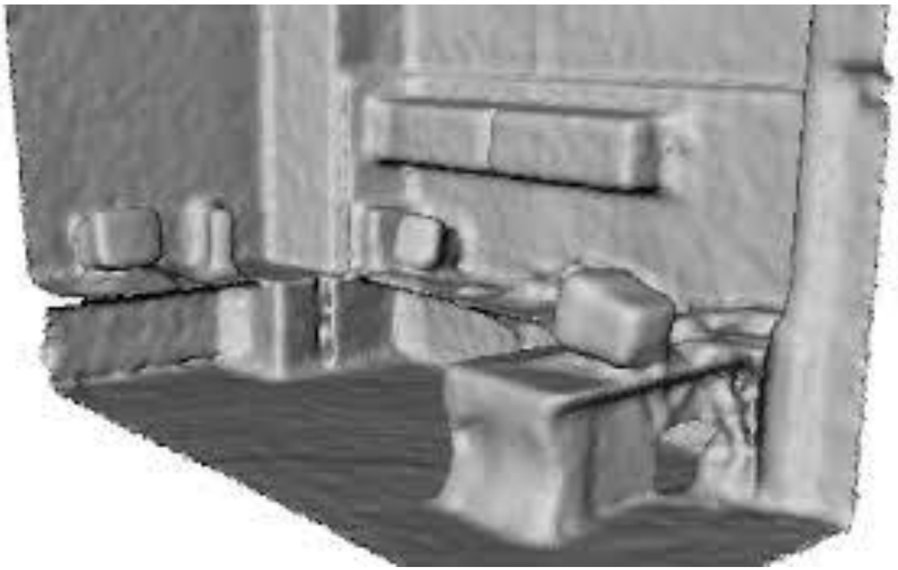


PR2 Humanoid

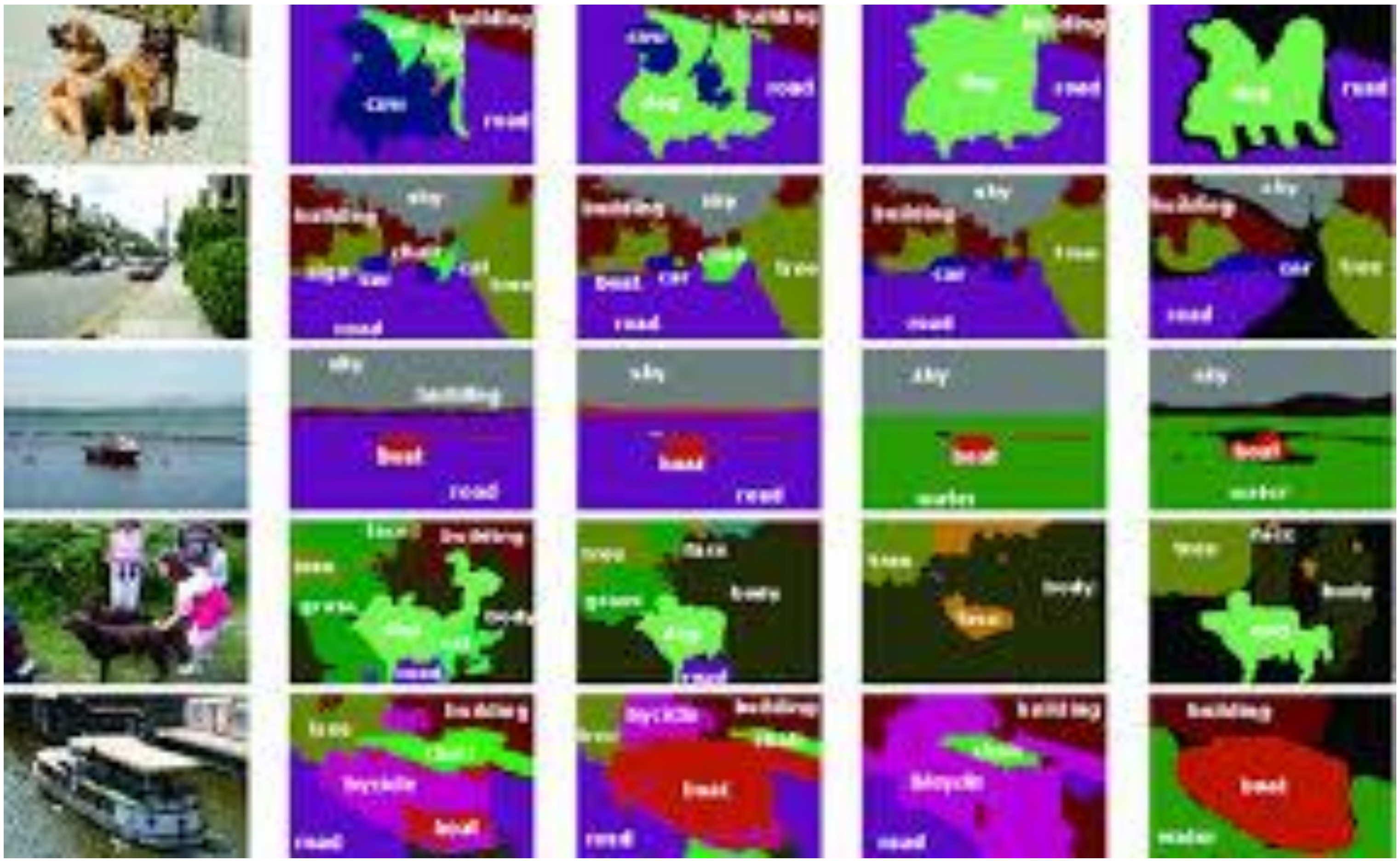
Perception for Robots
3 major problems

- Reconstruction
- Reorganization
- Recognition

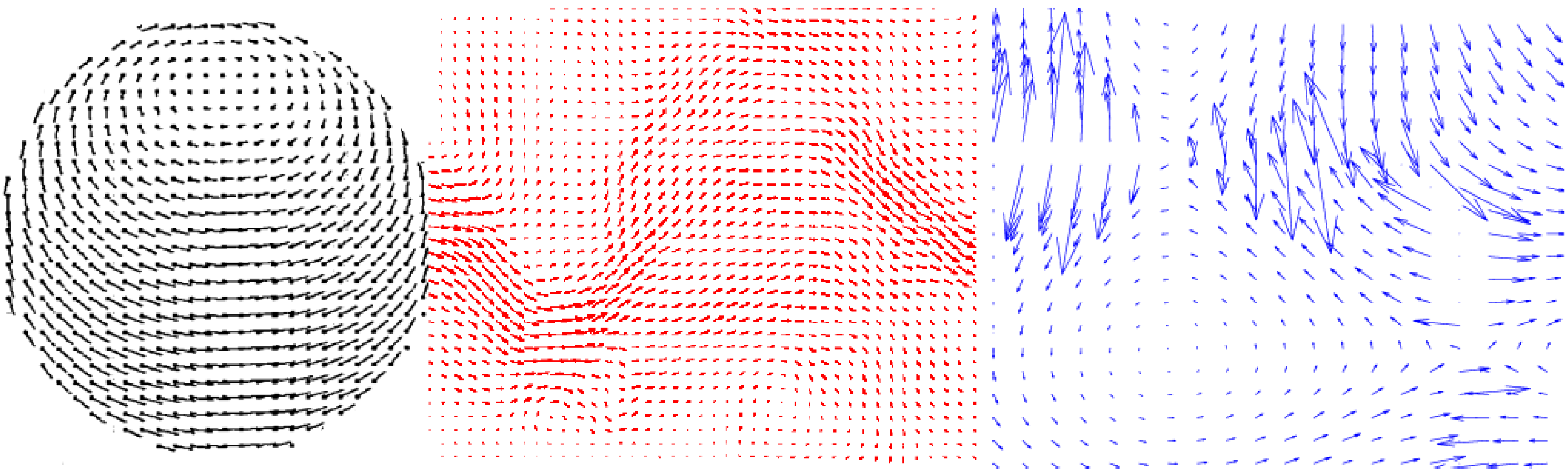
Reconstruction



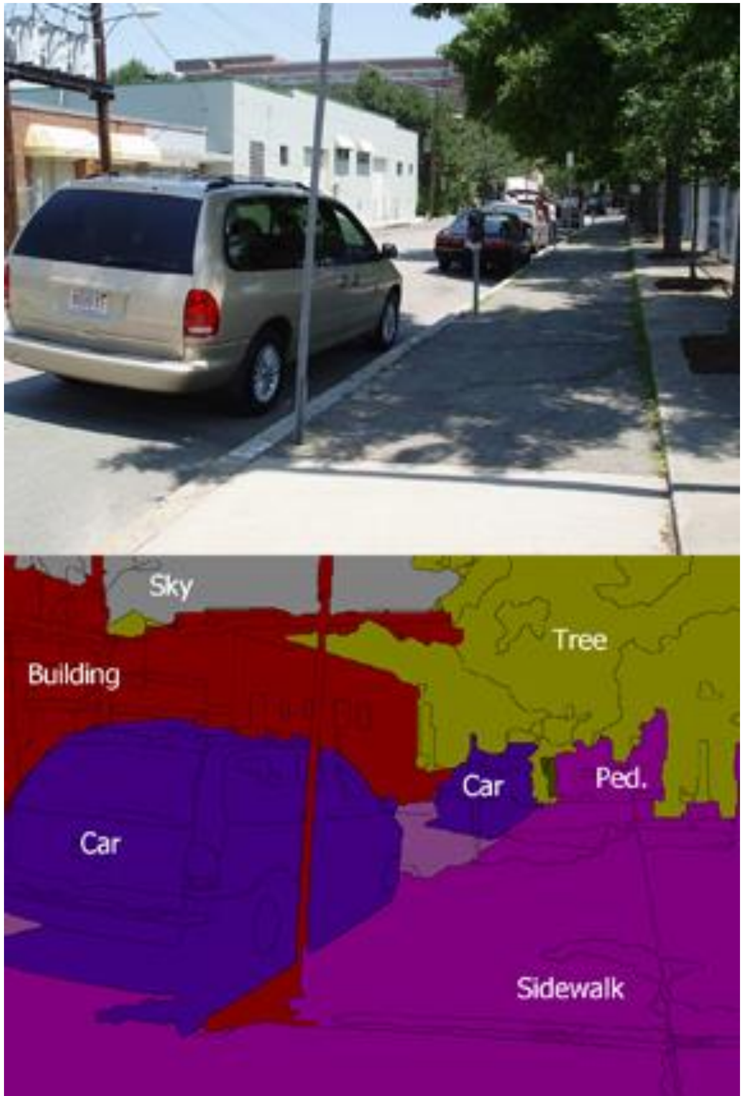
Reorganization: segmentation

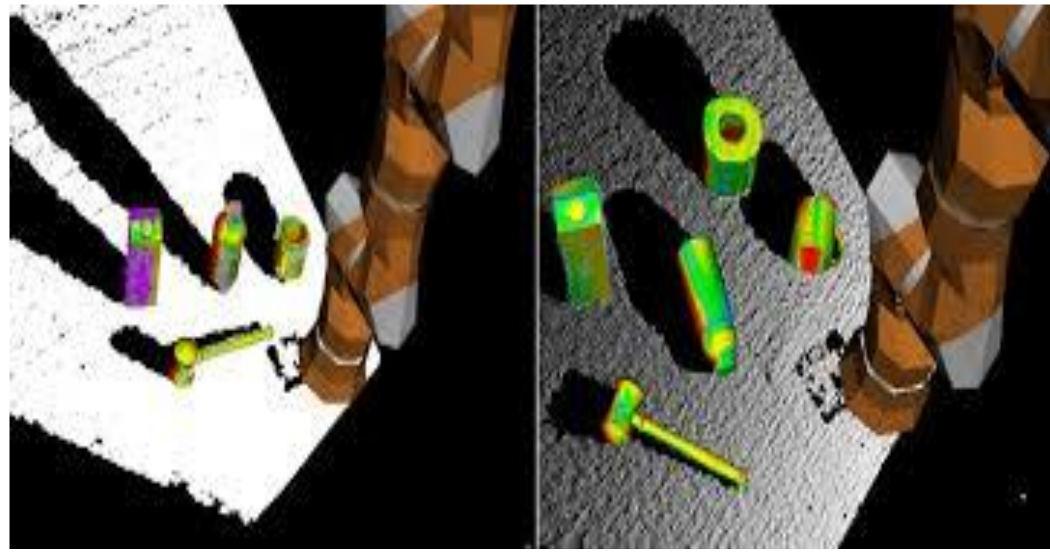


Reorganization: flow



Recognition

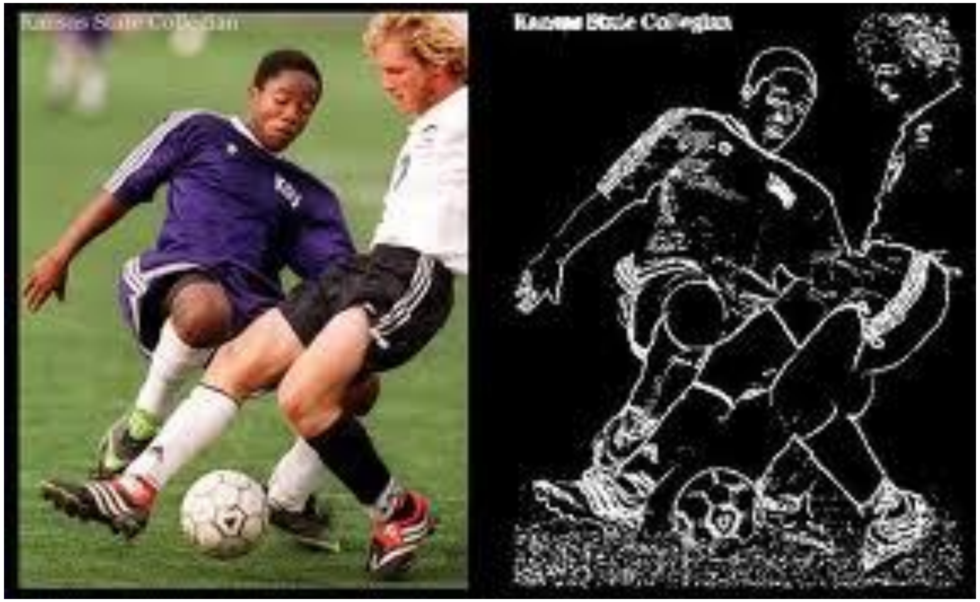




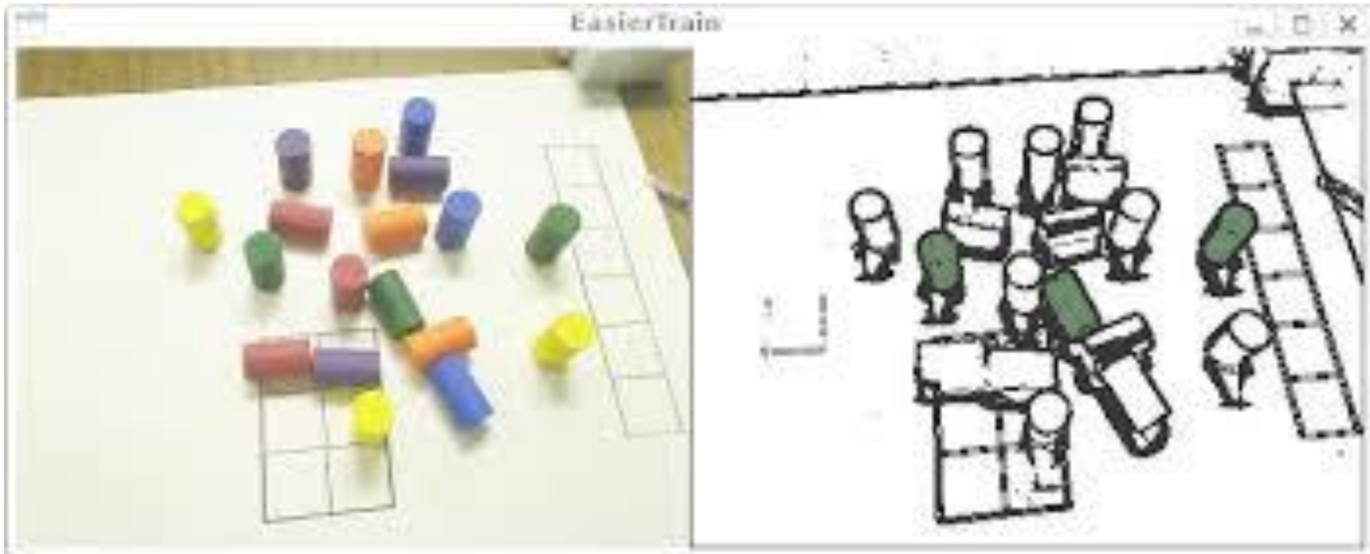
Images and Videos Contain

- Lines (contours, edges)
- Intensity and Color
- Texture
- movement

Lines



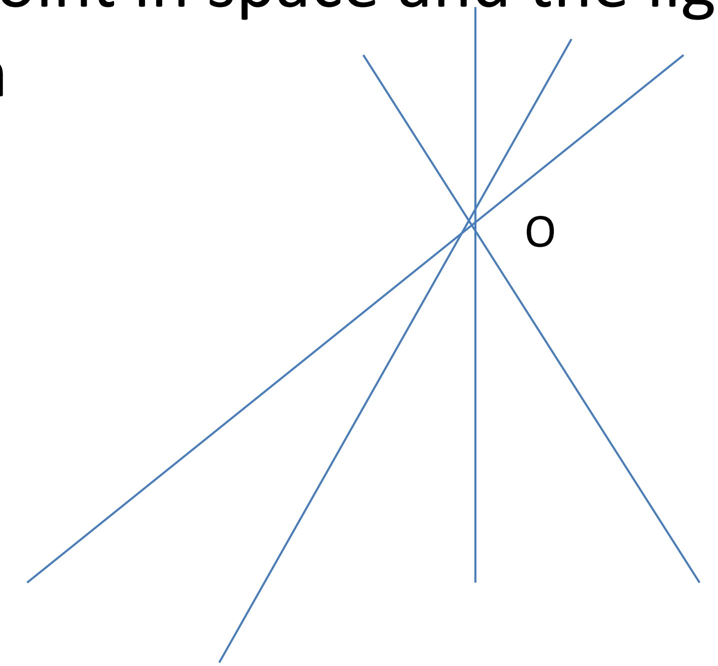
Color, Texture



Motion

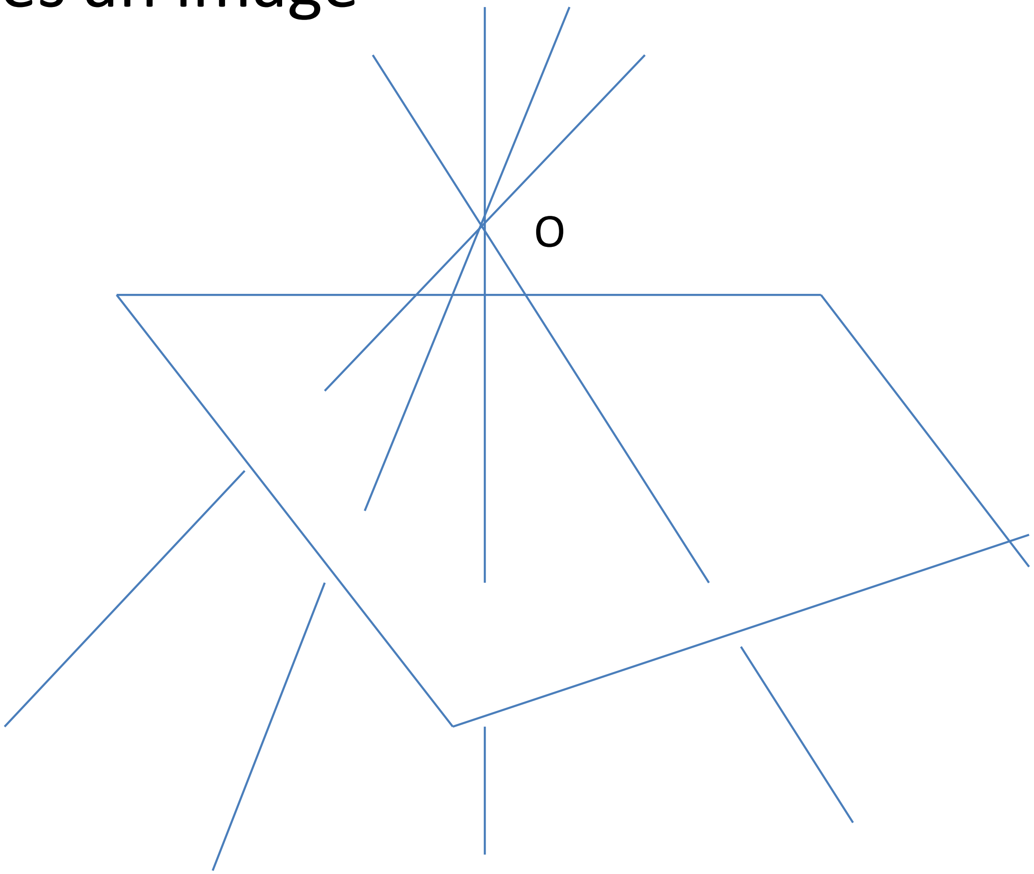
A theoretical model of an eye

- Pick a point in space and the light rays passing through



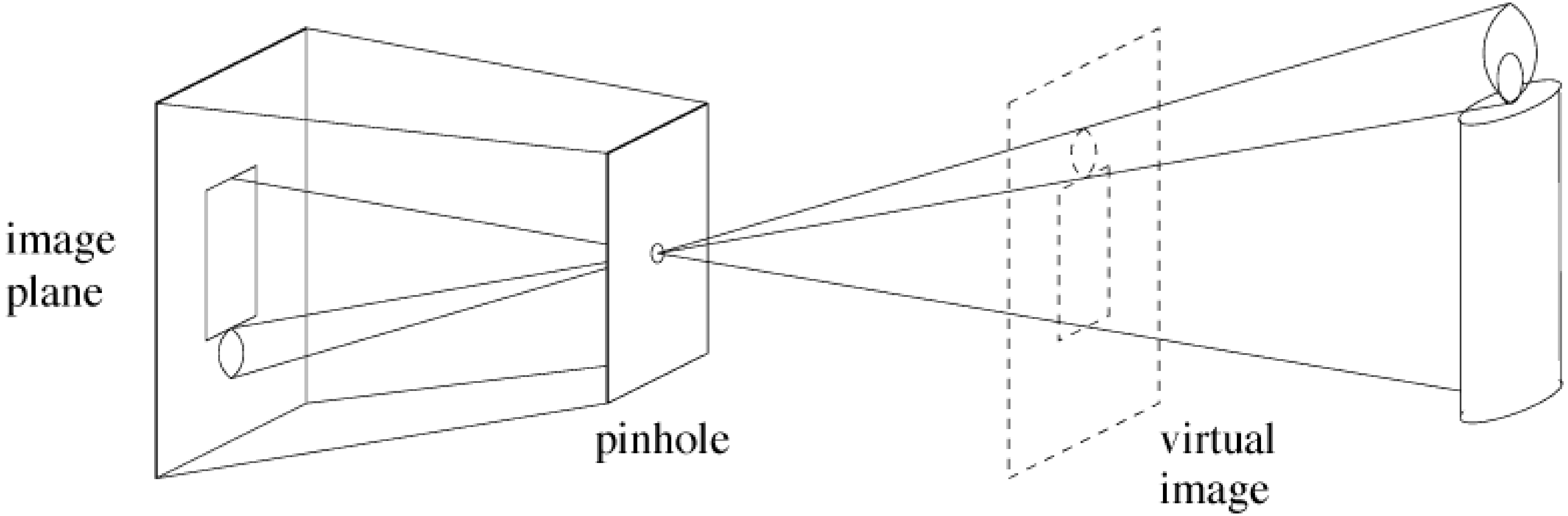
Then cut the rays with a plane

- This gives an image



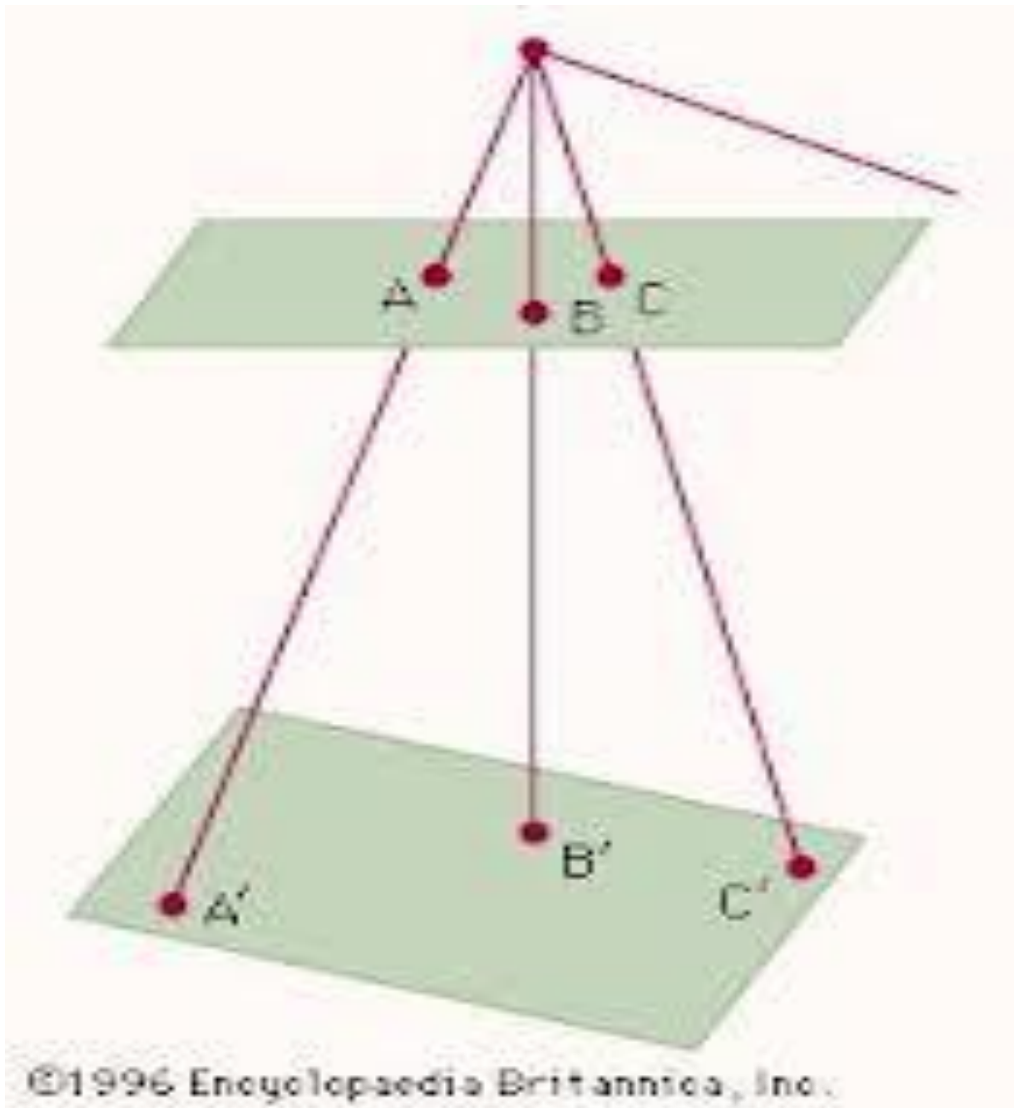
Pinhole cameras

- Abstract camera model - box with a small hole in it
- Pinhole cameras work in practice

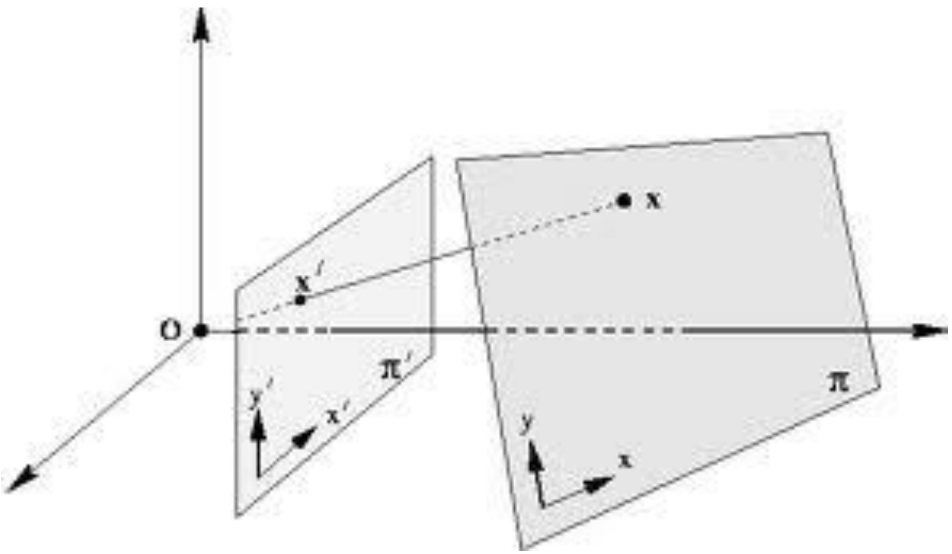


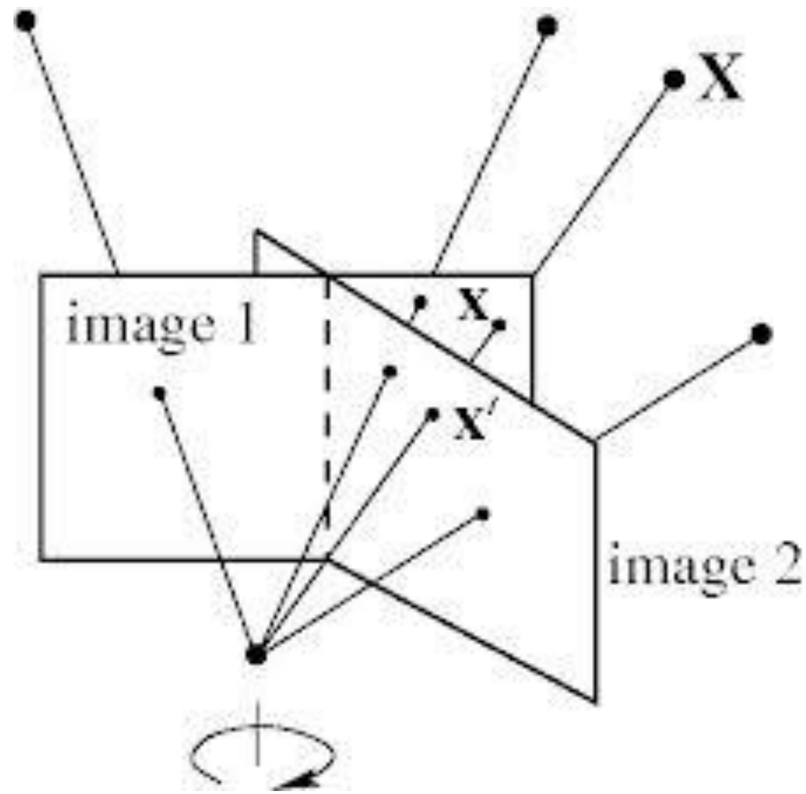
(Forsyth & Ponce)

If we change the plane, we get an new image

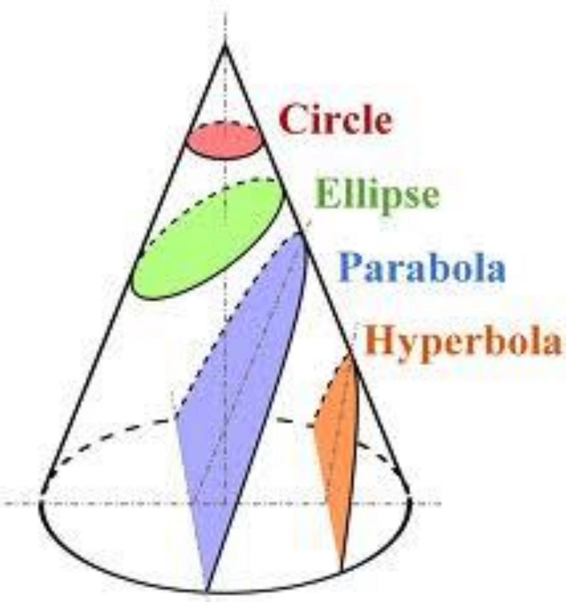


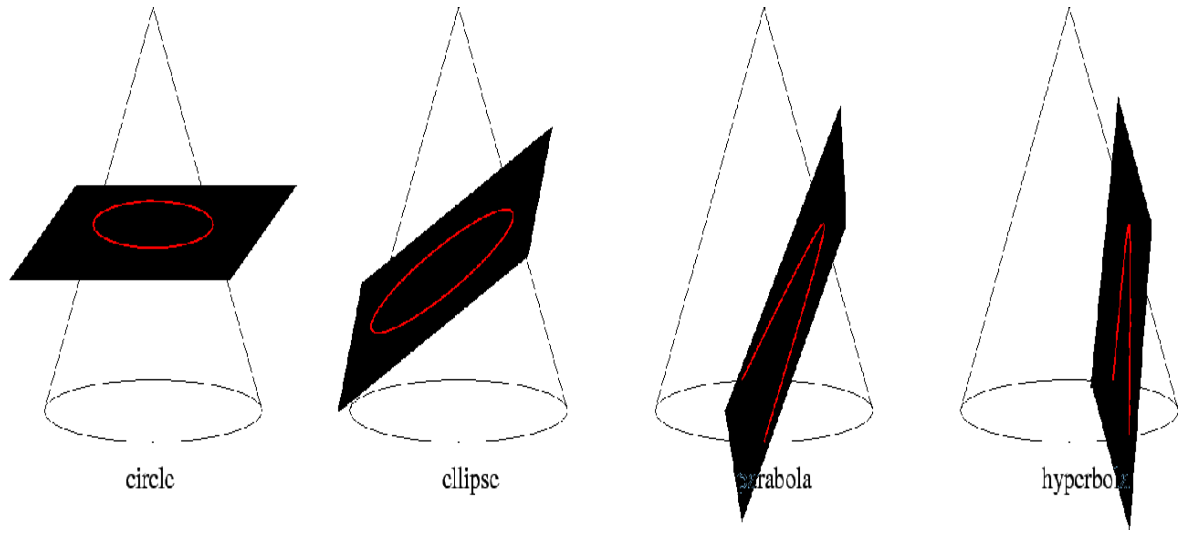
How are these images related? (what remains invariant?)





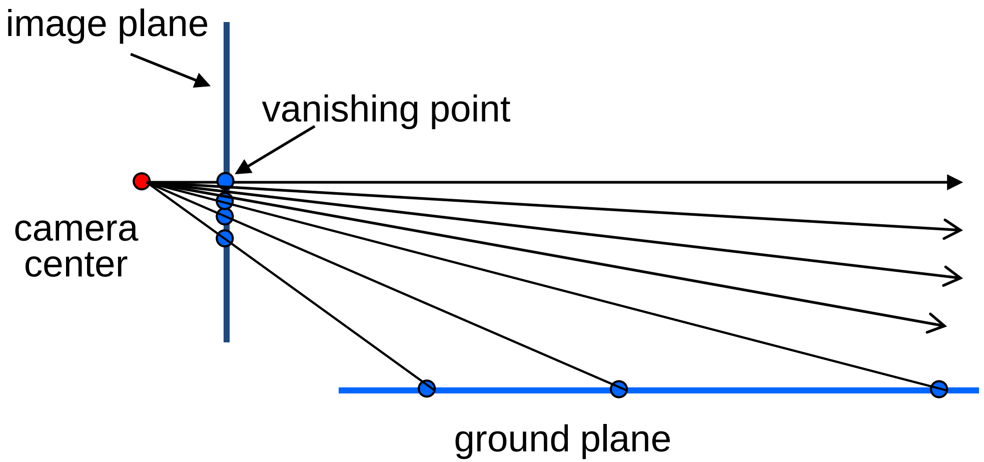
Conics





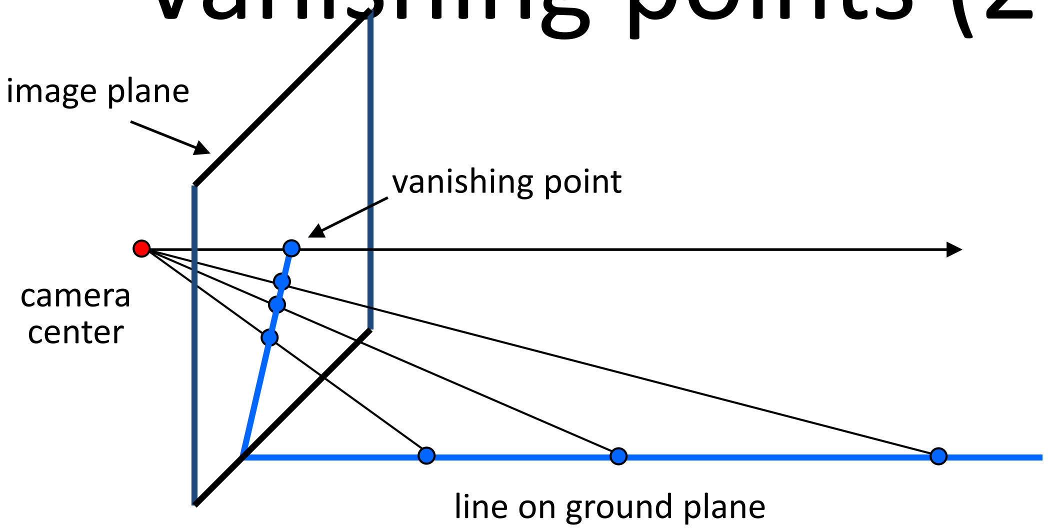
Projection of circle

Vanishing points

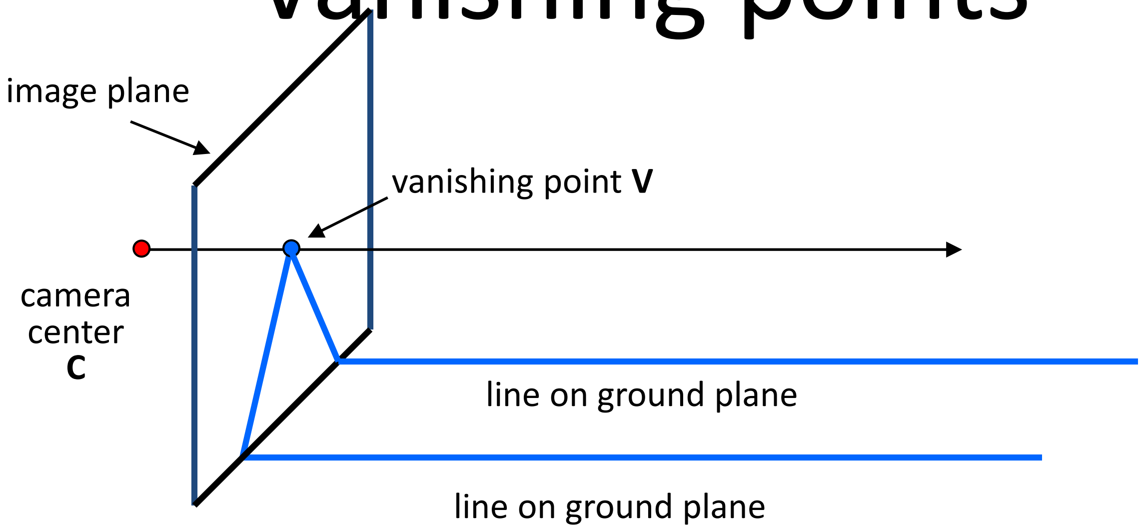


- Vanishing point
  - projection of a point at infinity

Vanishing points (2D)



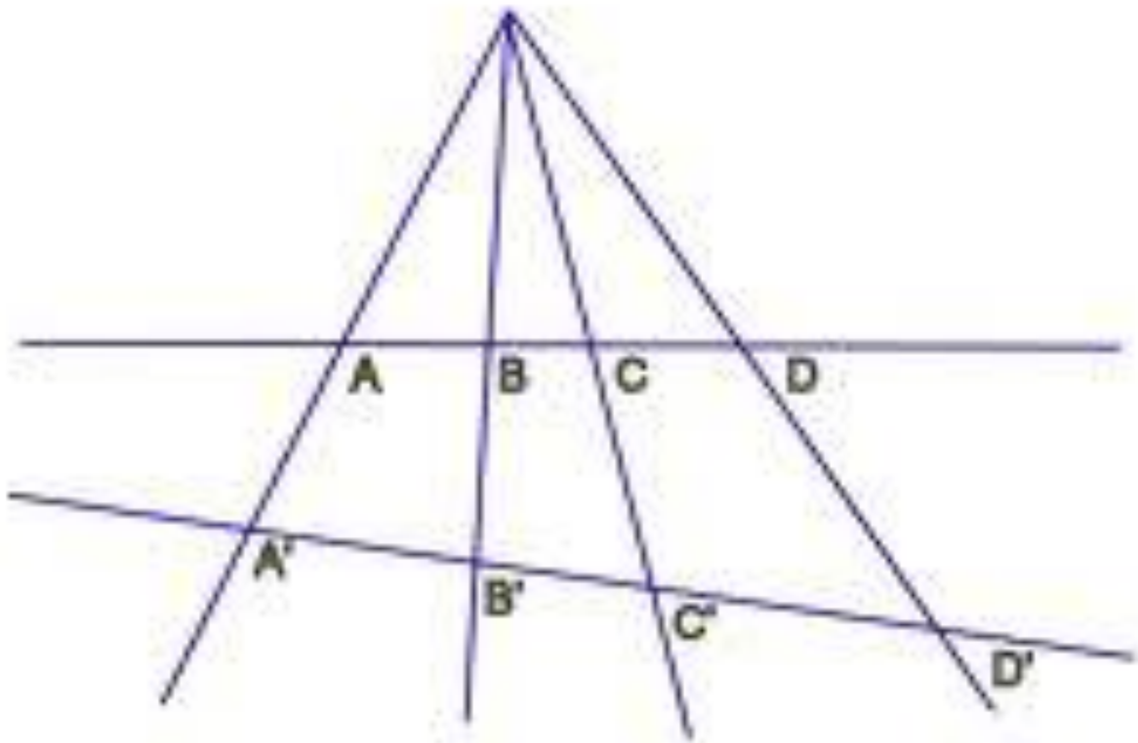
Vanishing points



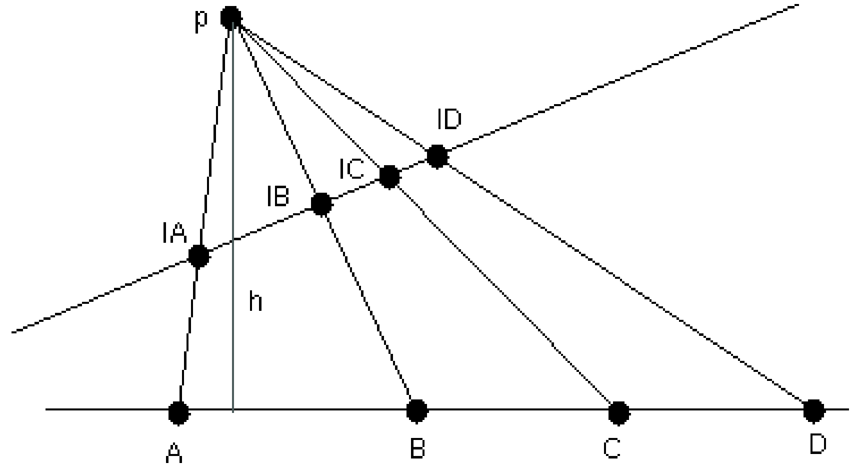
- Properties
  - Any two parallel lines have the same vanishing point **v**
  - The ray from **C** through **v** is parallel to the lines
  - An image may have more than one vanishing point

Parallelism (angles) not invariant

Cross ratio = only invariant



$$\frac{(CA\ /\ CB)}{(DA\ /\ DB)} = \frac{(C'A'\ /\ C'B')}{(D'A'\ /\ D'B')}$$



Remember that the area of a triangle is 1/2 the base times the height.
It is also the product of two sides times the side of the angle between them. Using this, we get:

Area(pAC) = h/2(AC) = 1/2 (pA)(pC) sin(ApC)
Area(pBC) = h/2(BC) = 1/2 (pB)(pC) sin(BpC)
Area(pAD) = h/2(AD) = 1/2 (pA)(pD) sin(ApD)
Area(pBD) = h/2(BD) = 1/2 (pB)(pD) sin(BpD)

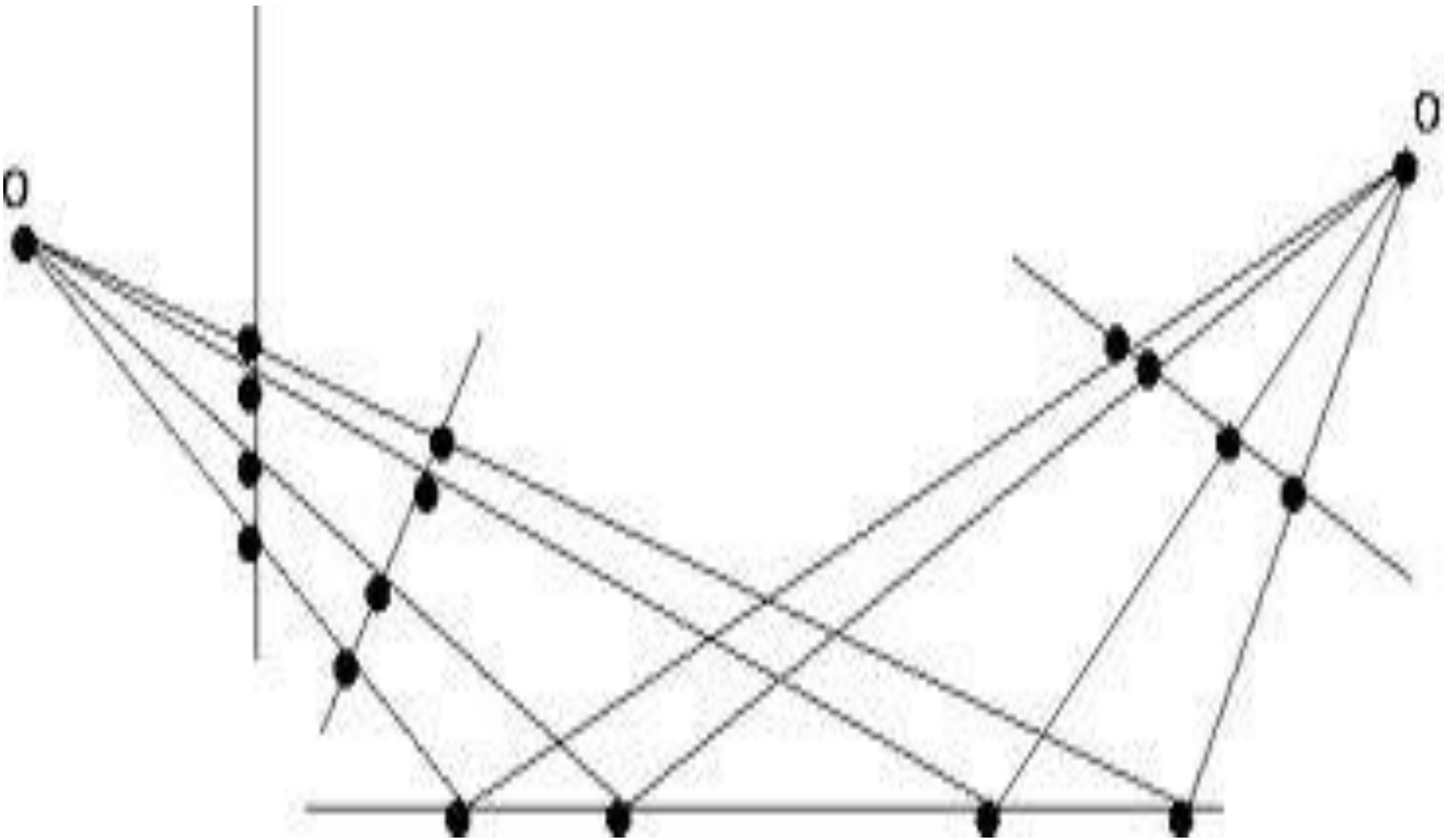
Thus the cross ratio of A,B,C,D = [(AC)/(BC)]/[(AD)/(BD)]

$$= \frac{[(pA)(pC)\sin(ApC)/(pB)(pC)\sin(BpC)]}{[(pA)(pD)\sin(ApD)/(pB)(pD)\sin(BpD)]}$$

$$= [\sin(ApC)/\sin(BpC)]/[\sin(ApD)/\sin(BpD)]$$

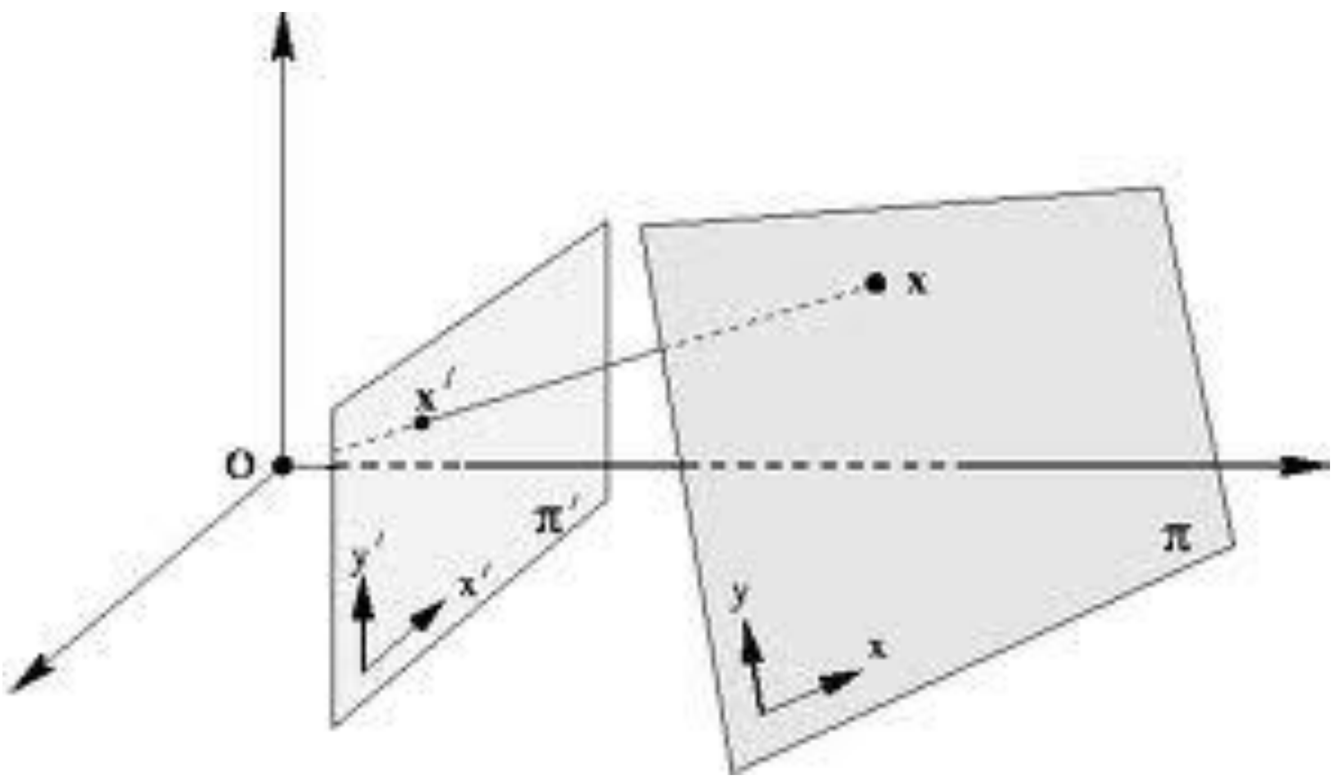
This last quantity is independent of the line we project to. Thus cross ratios are invariant under projection.

This discussion is based on Courant and Robbins, "What Is Mathematics."

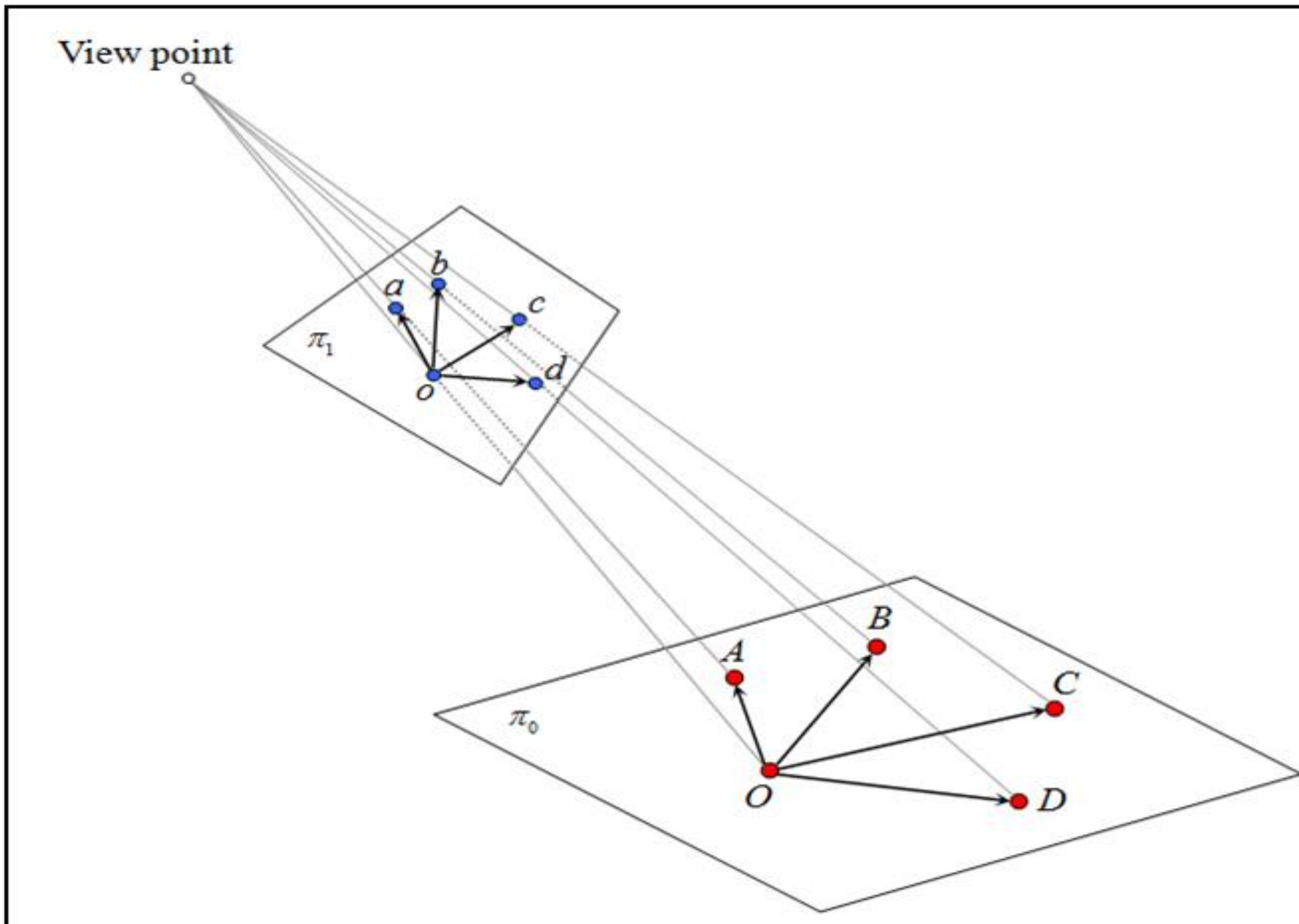

O
O'

Back to our question: how are the 2 images related to each other

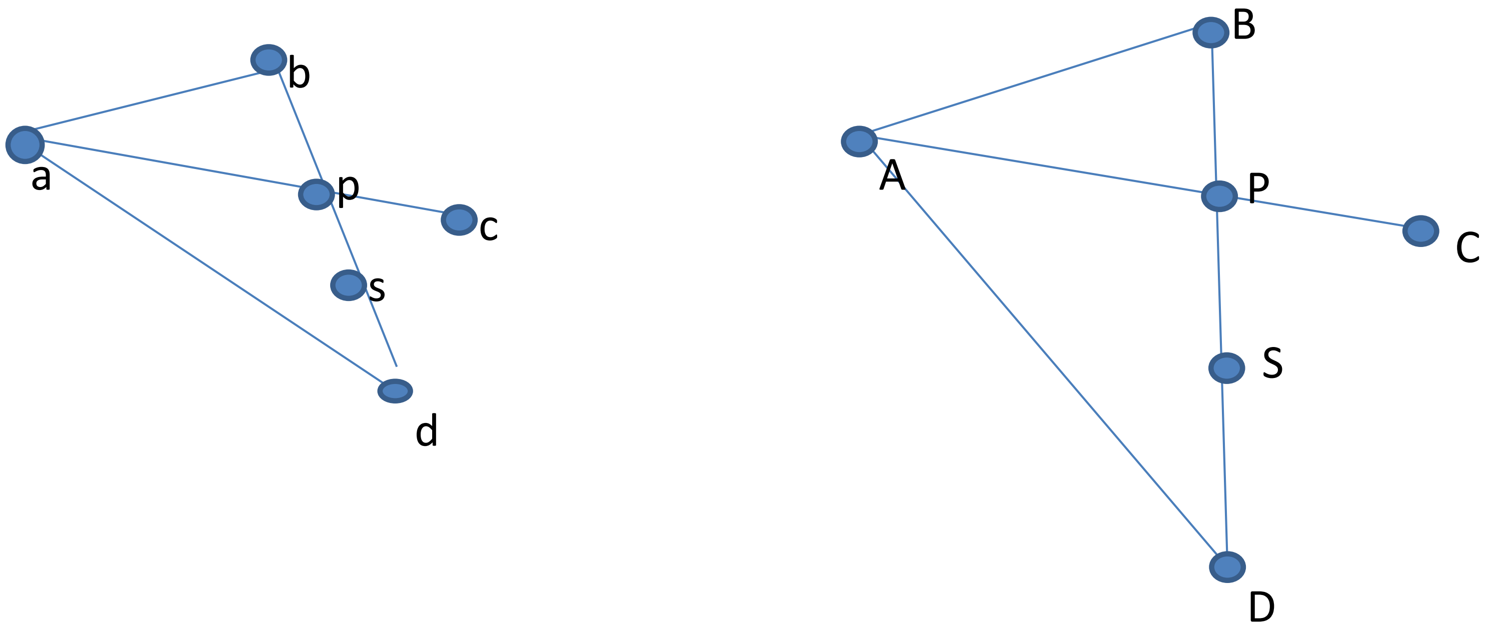
Can we find a map, a function mapping x' to x?



Fundamental Theorem: If we know how 4 points map to each other in the two planes, then we know how all points map. (if a→A, b→B, c→C,d→D, then we can map any point)

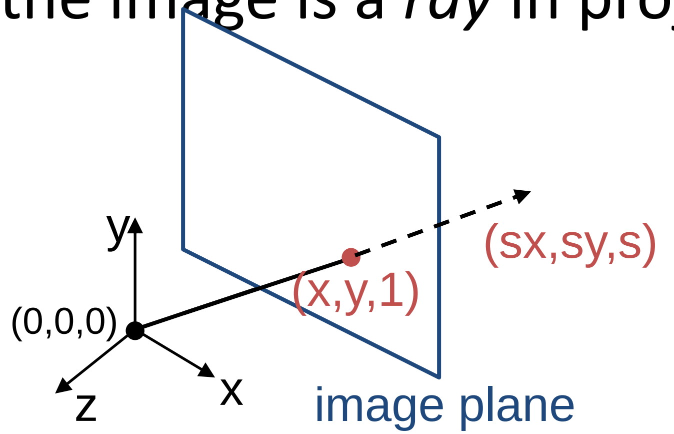


Proof



The projective plane

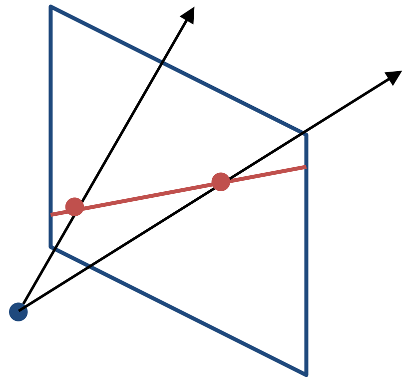
- Why do we need homogeneous coordinates?
  - represent points at infinity, homographies, perspective projection, multi-view relationships
- What is the geometric intuition?
  - a point in the image is a *ray* in projective space



- Each *point* (x,y) on the plane is represented by a *ray* (sx,sy,s)
  - all points on the ray are equivalent: (x, y, 1) ≅ (sx, sy, s)

Projective lines

- What does a line in the image correspond to in projective space?



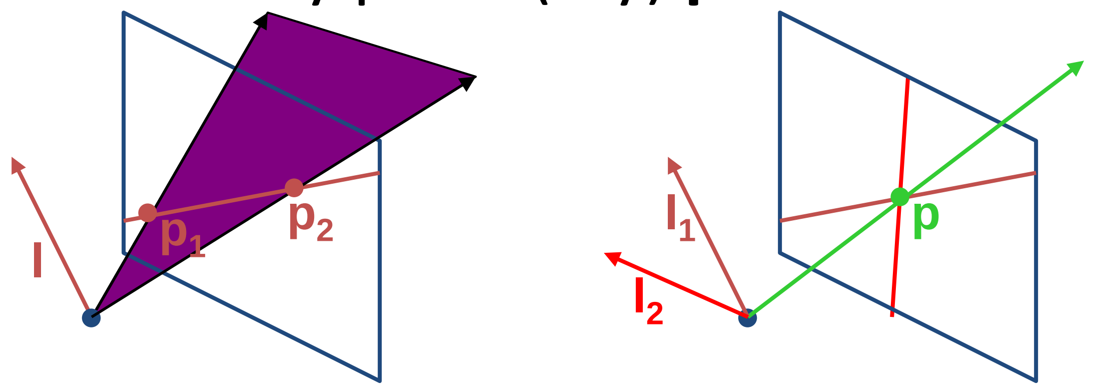
- A line is a *plane* of rays through origin
  - all rays (x,y,z) satisfying: ax + by + cz = 0

$$\text{in vector notation:} \qquad 0 = \underbrace{\begin{bmatrix} a & b & c \end{bmatrix}}_{\mathbf{l}} \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{\mathbf{p}}$$

- A line is also represented as a homogeneous 3-vector **l**

Point and line duality

- A line $\mathbf{l}$ is a homogeneous 3-vector
- It is $\perp$ to every point (ray) $\mathbf{p}$ on the line: $\mathbf{l}\ \mathbf{p}=0$



What is the line $\mathbf{l}$ spanned by rays $\mathbf{p_1}$ and $\mathbf{p_2}$?

- $\mathbf{l}$ is $\perp$ to $\mathbf{p_1}$ and $\mathbf{p_2}$ $\Rightarrow$ $\mathbf{l} = \mathbf{p_1} \times \mathbf{p_2}$
- $\mathbf{l}$ is the plane normal

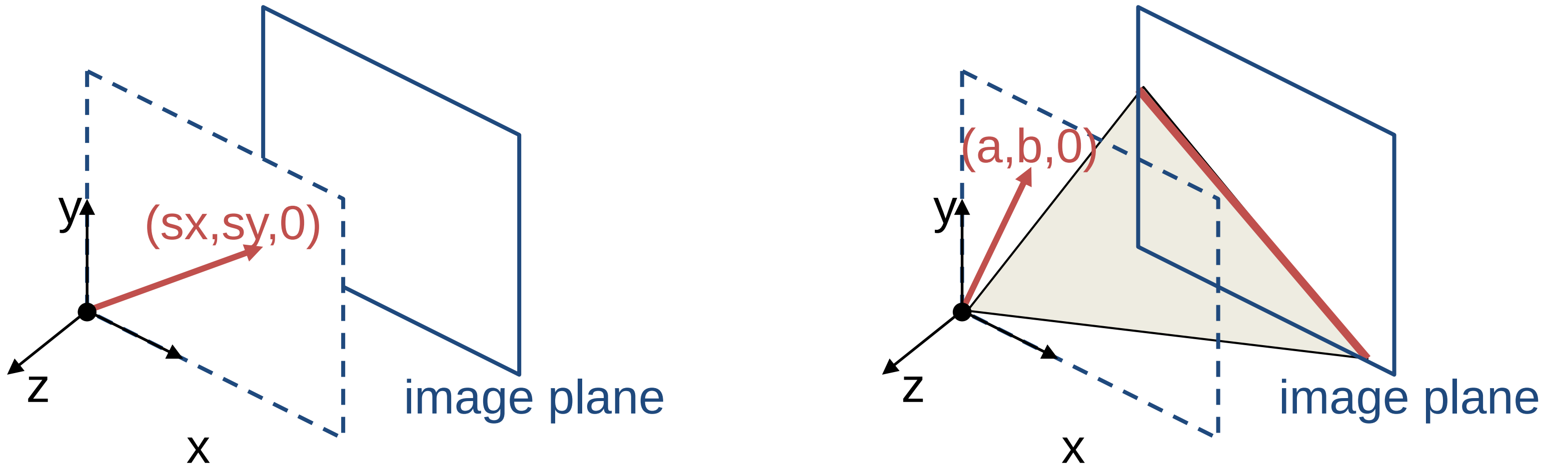
What is the intersection of two lines $\mathbf{l_1}$ and $\mathbf{l_2}$?

- $\mathbf{p}$ is $\perp$ to $\mathbf{l_1}$ and $\mathbf{l_2}$ $\Rightarrow$ $\mathbf{p} = \mathbf{l_1} \times \mathbf{l_2}$

Points and lines are *dual* in projective space

- given any formula, can switch the meanings of points and lines to get another formula

Ideal points and lines



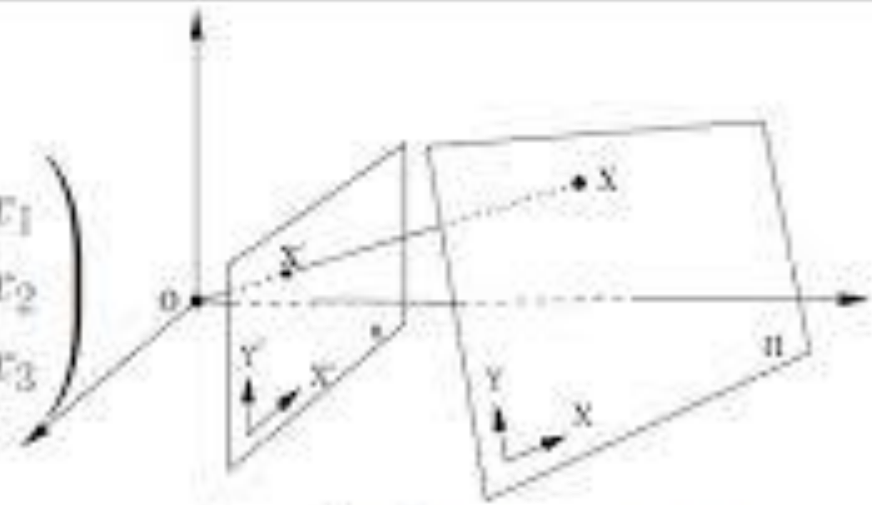
- Ideal point ("point at infinity")
  - $p \cong (x, y, 0)$ – parallel to image plane
  - It has infinite image coordinates

Ideal line

- $l \cong (a, b, 0)$ – parallel to image plane
- Corresponds to a line in the image (finite coordinates)

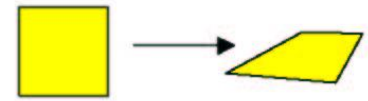
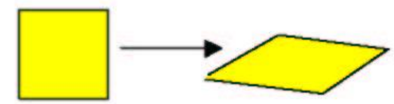
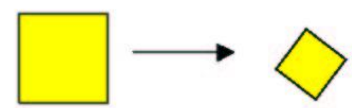
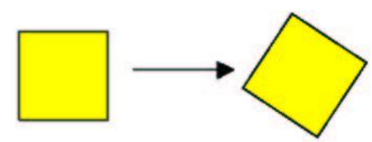
Fundamental Theorem
(homography or collineation)

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$



or $\mathrm{x}' = \mathrm{H}\mathrm{x}$, where H is a 3×3 non-singular homogeneous matrix.

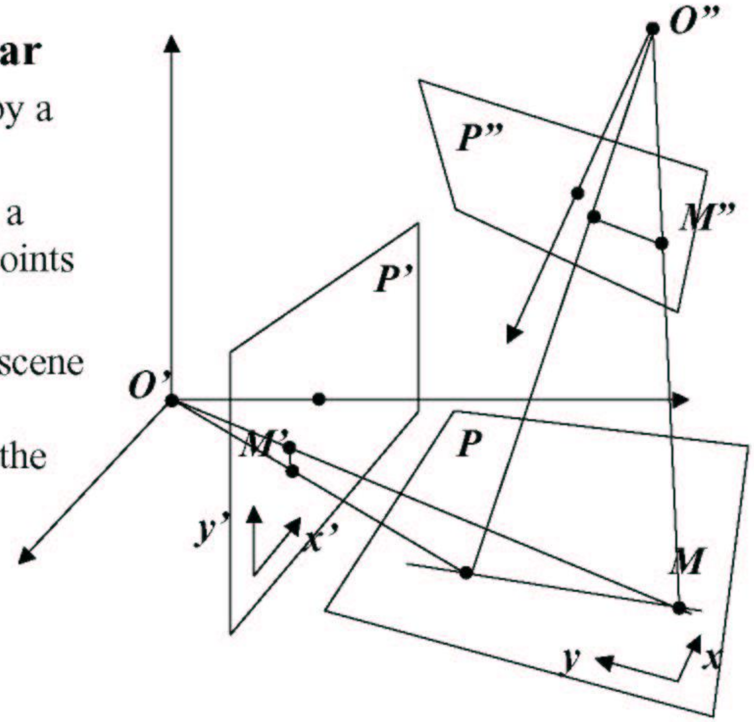
Special Projectivities

		Invariants	
Projectivity 8 dof	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$	Collinearity, Cross-ratios	
Affine transform 6 dof	$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_x \\ 0 & 0 & 1 \end{bmatrix}$	Parallelism, Ratios of areas, Length ratios	
Similarity 4 dof	$\begin{bmatrix} s\,r_{11} & s\,r_{12} & t_x \\ s\,r_{21} & s\,r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$	Angles, Length ratios	
Euclidean transform 3 dof	$\begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$	Angles, Lengths, Areas	

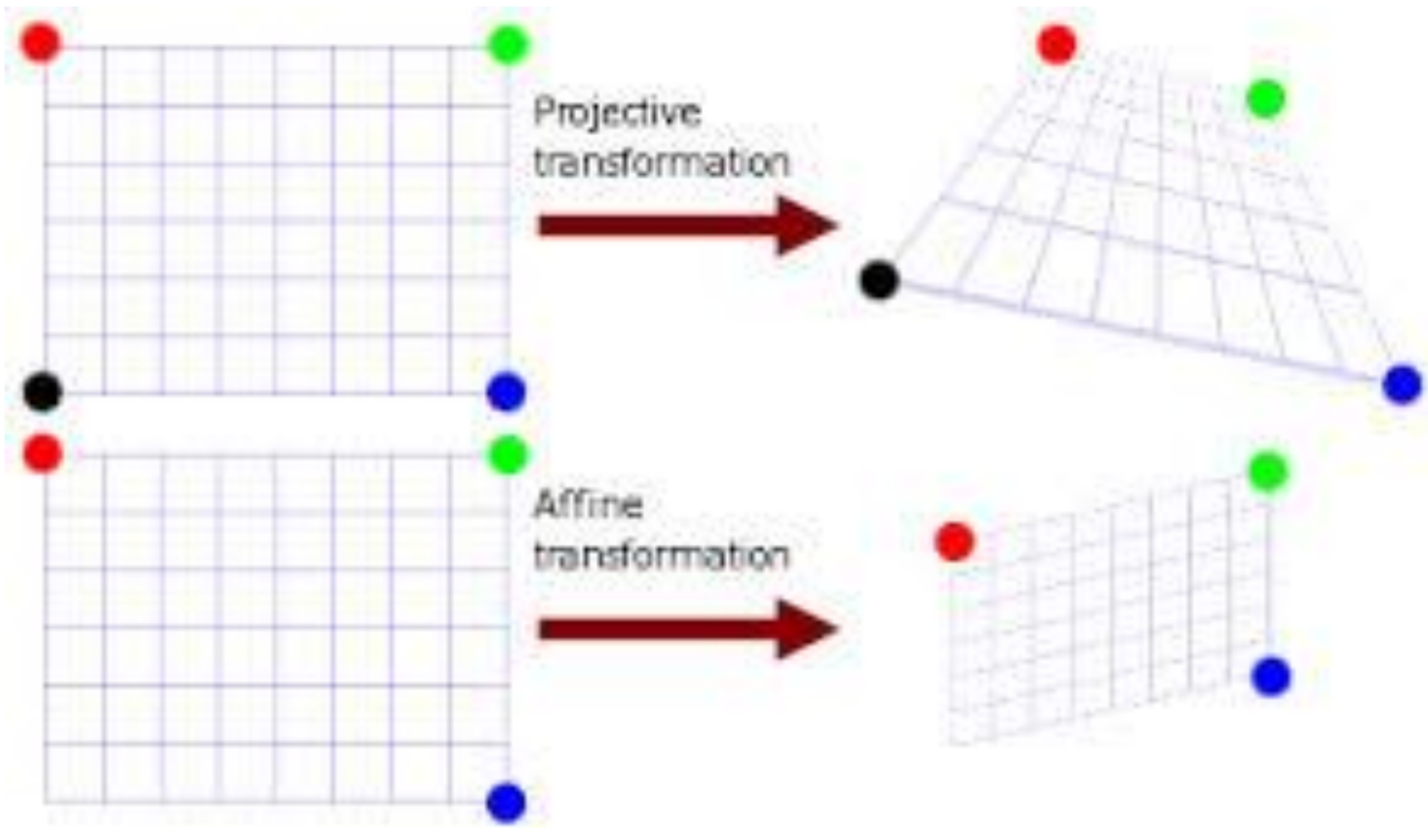
Examples of Projective Transformations

- Central projection maps **planar scene** points to image plane by a projectivity
  - True because all points on a scene line are mapped to points on its image line
- The image of the same planar scene from a second camera can be obtained from the image from the first camera by a projectivity
  - True because
    $\mathbf{x}'_i = \mathbf{H}'\,\mathbf{x}_i$, $\mathbf{x}''_i = \mathbf{H}''\,\mathbf{x}_i$

    so $\mathbf{x}''_i = \mathbf{H}''\,\mathbf{H}'^{-1}\,\mathbf{x}'_i$



Projective vs Affine



Rectification