



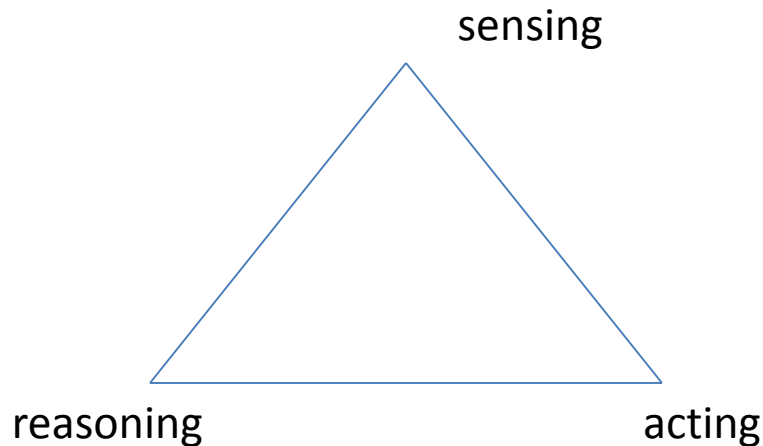
Perception for Robots

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

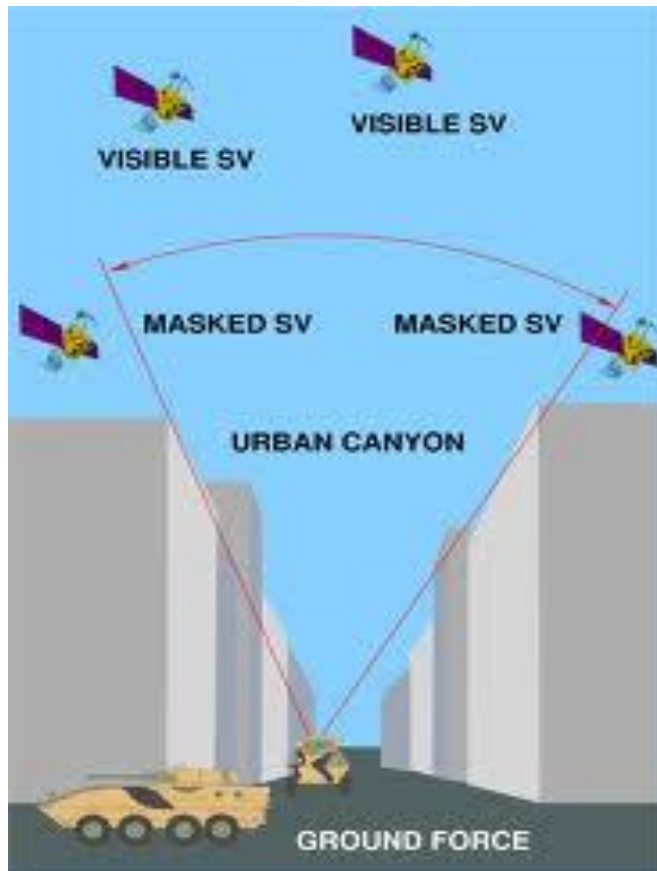
- Why not use GPS?

GPS is not perfect and has severe limitations in 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

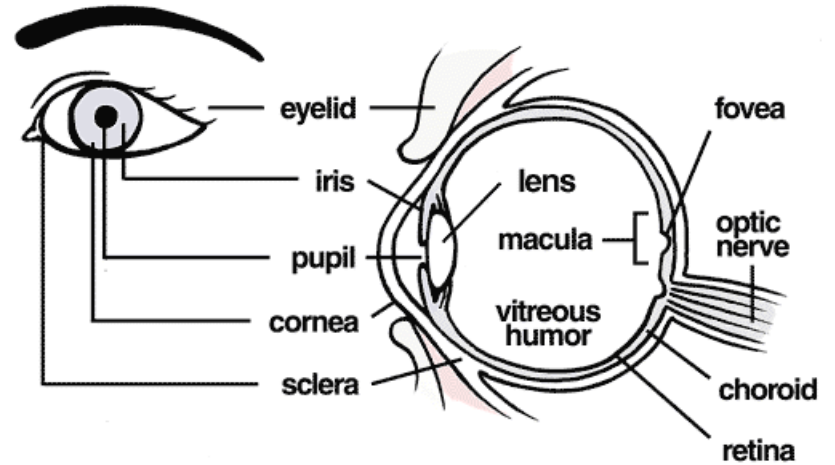
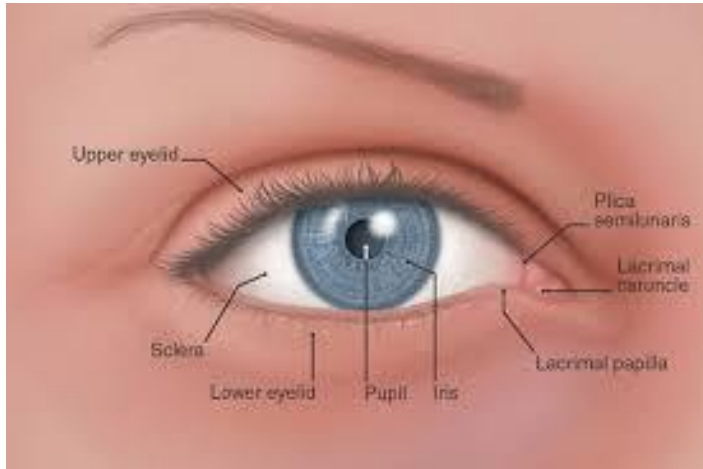


Humans and animals have a number of senses

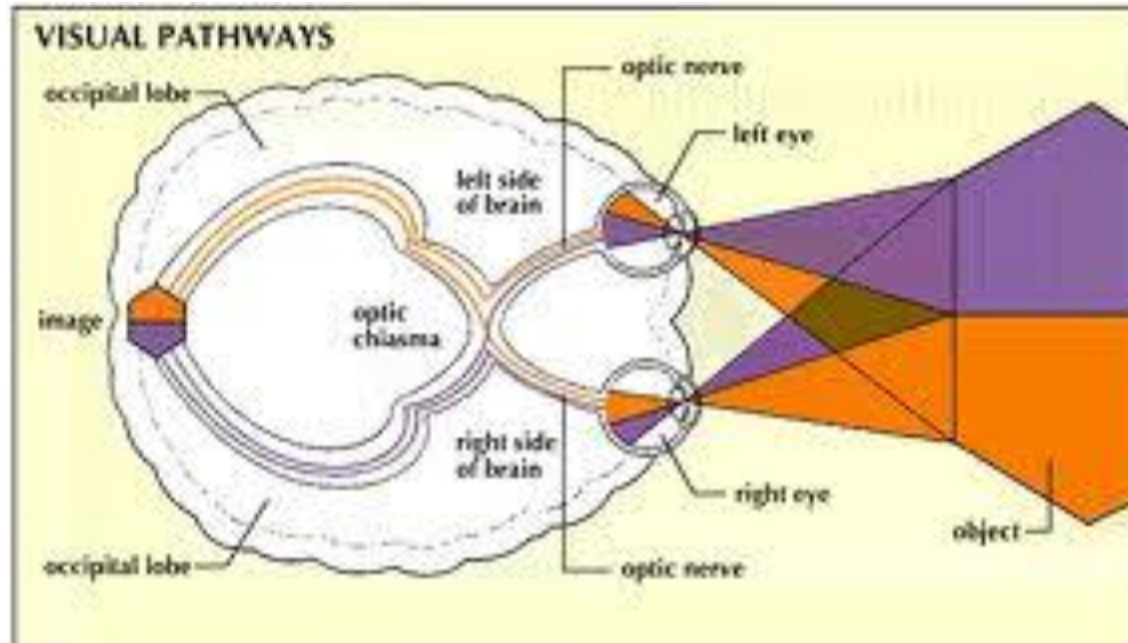
- Sight
- Hearing
- Touch
- Smell
- Taste
- Balance

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

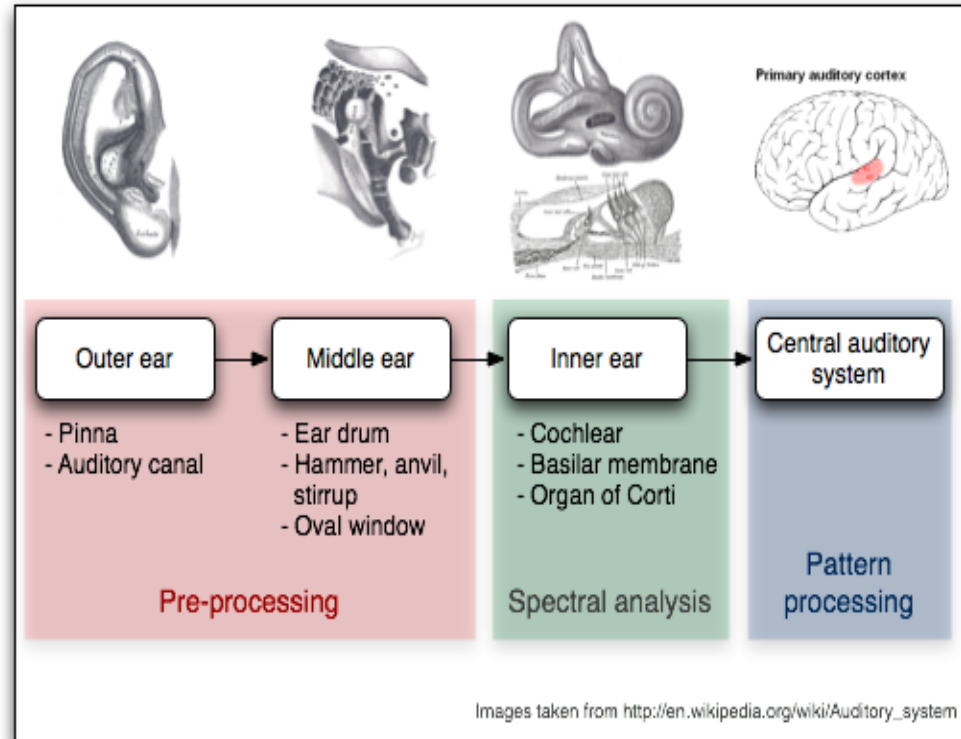
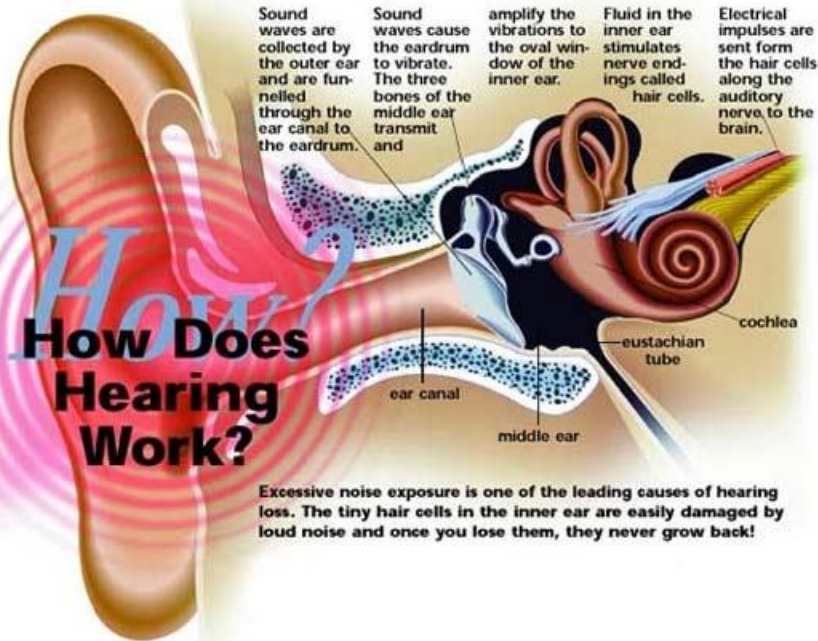
Vision



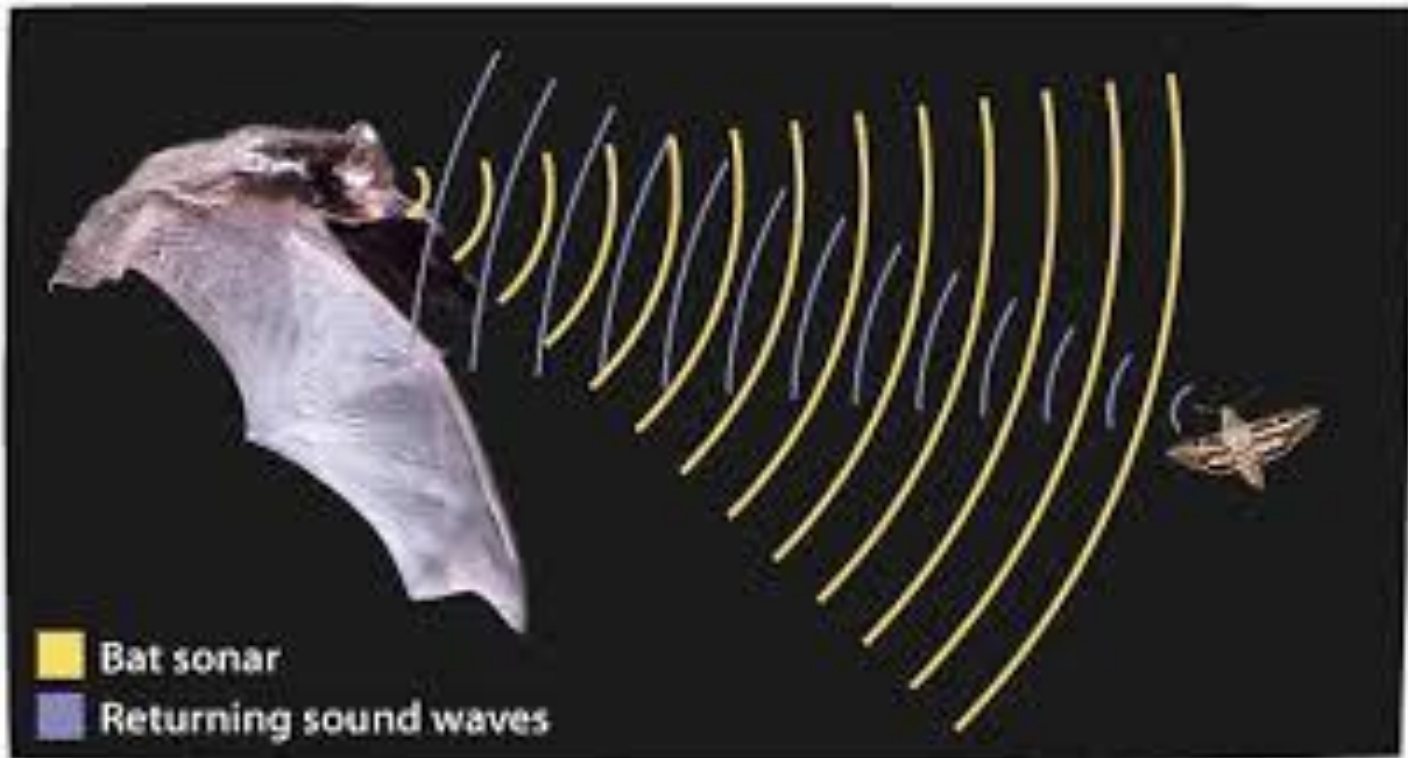
www.nvrf.org - illustration based upon information from National Eye Institute / National Institutes of Health



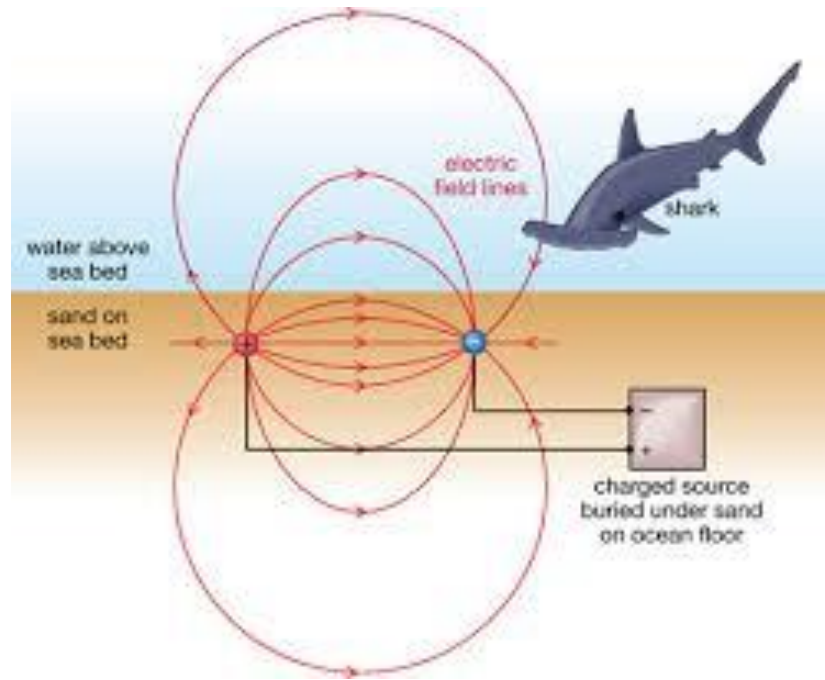
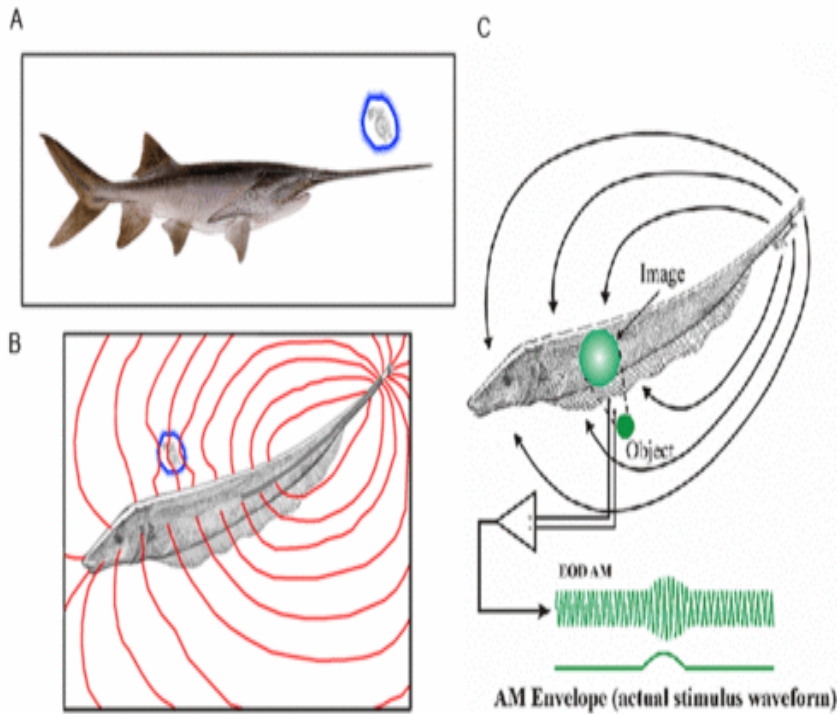
Hearing



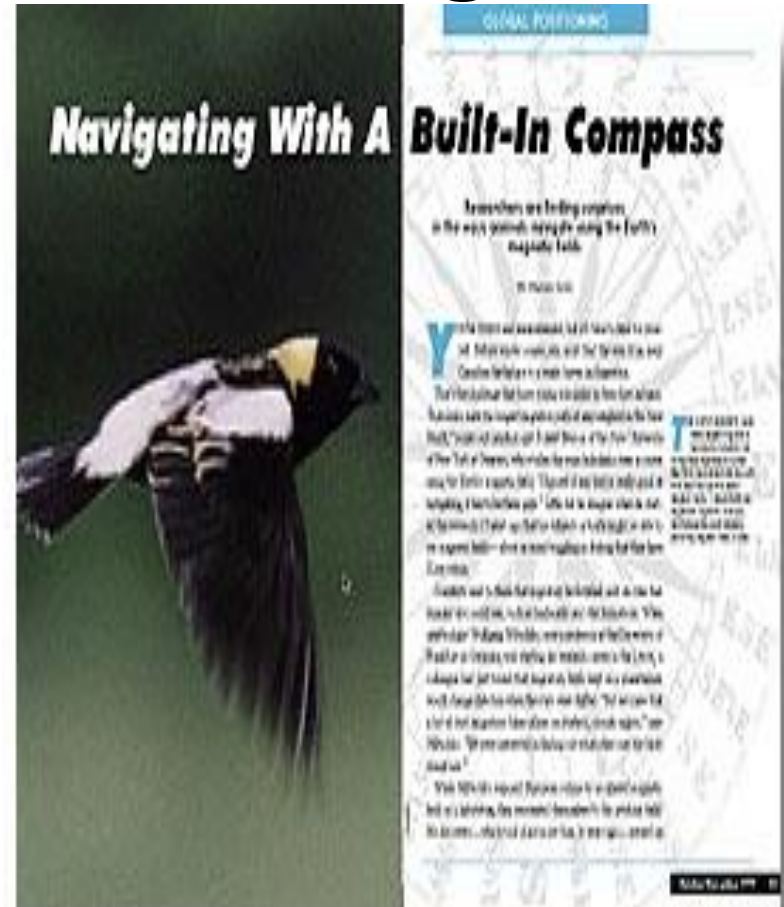
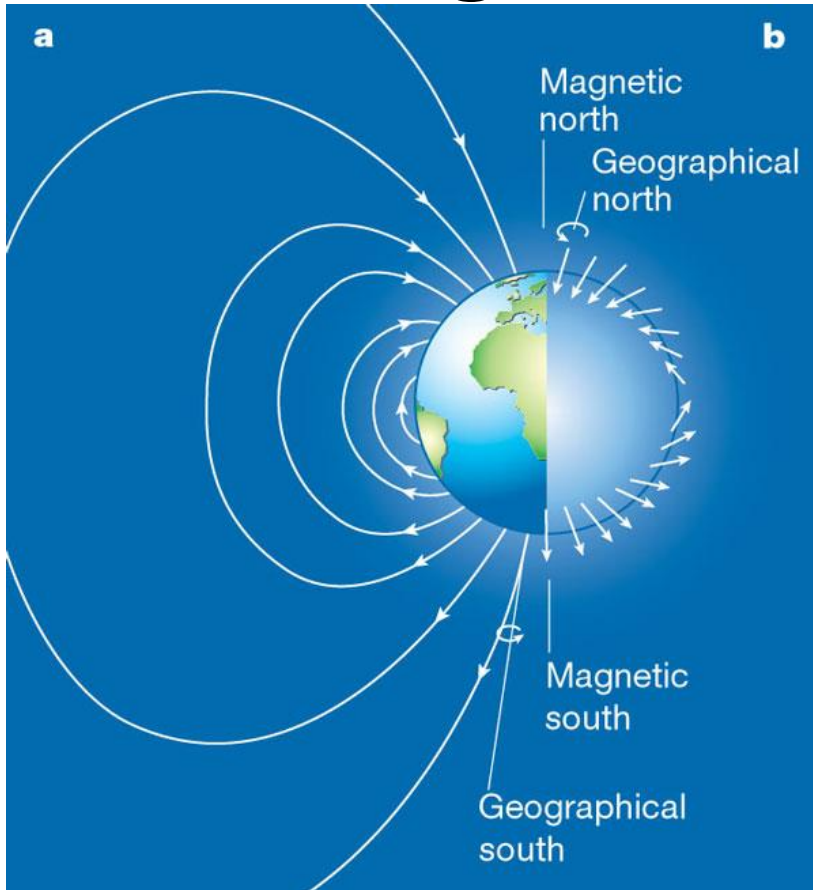
Echolocation of bats



Electric field sensing



Magnetic field sensing



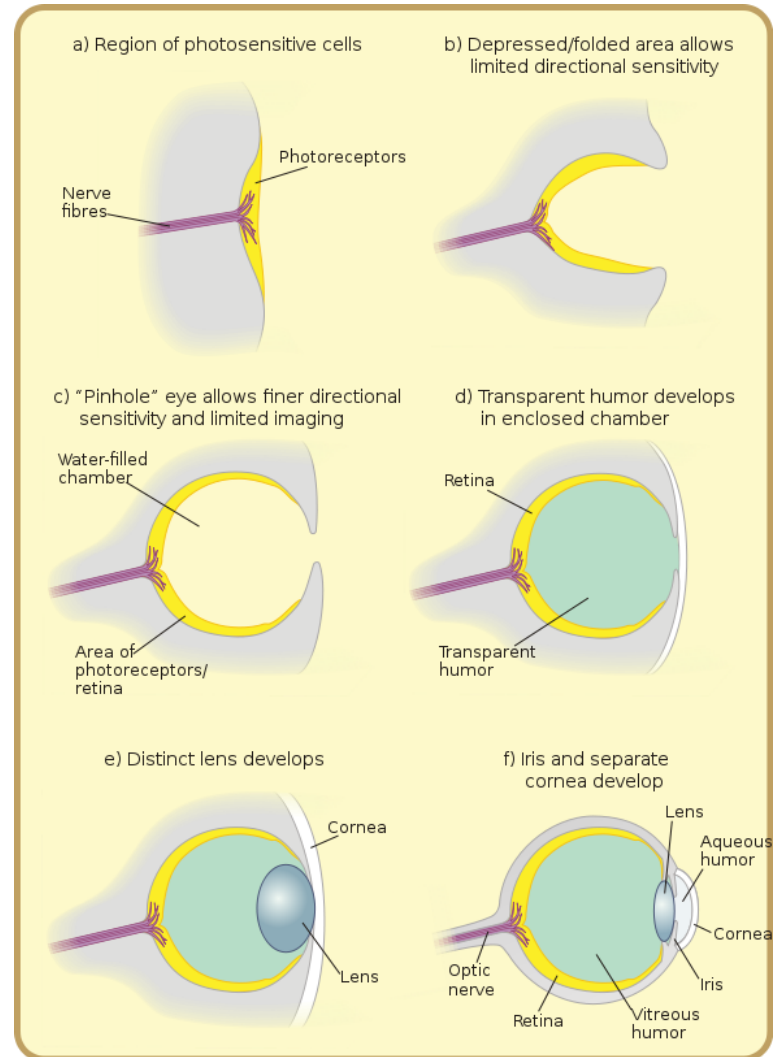
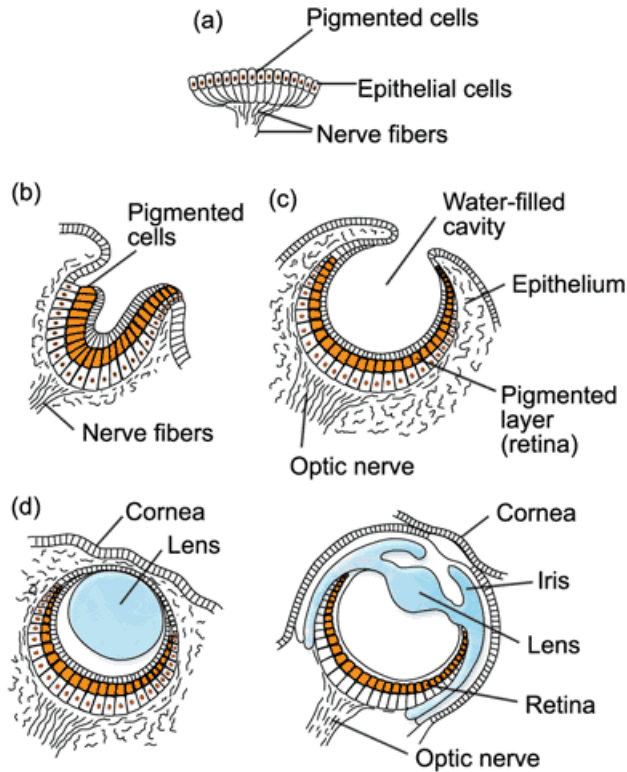
Vision: most powerful sense

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



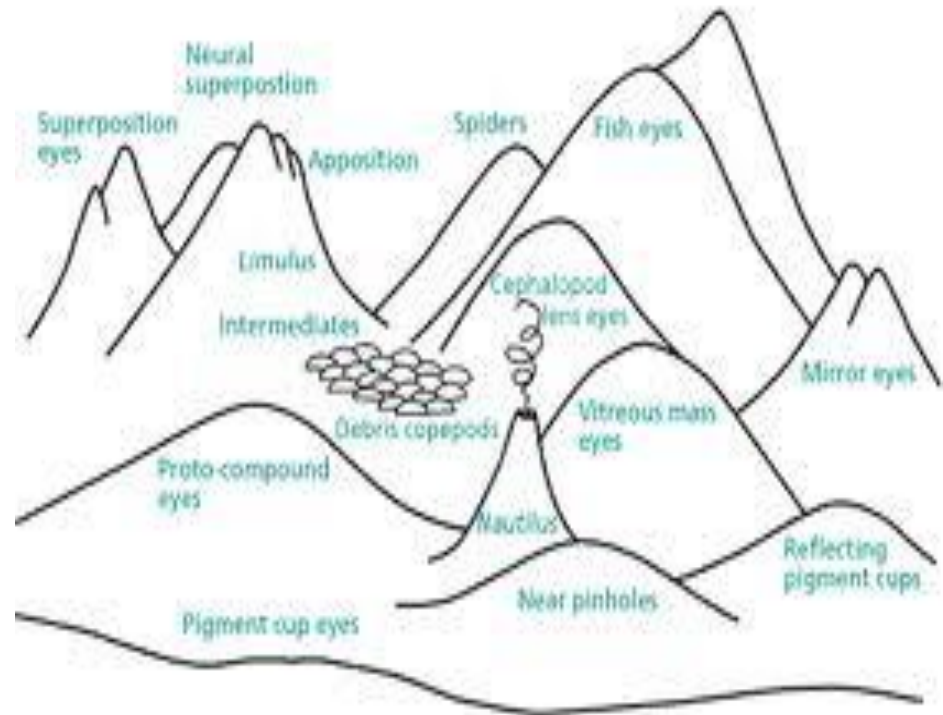
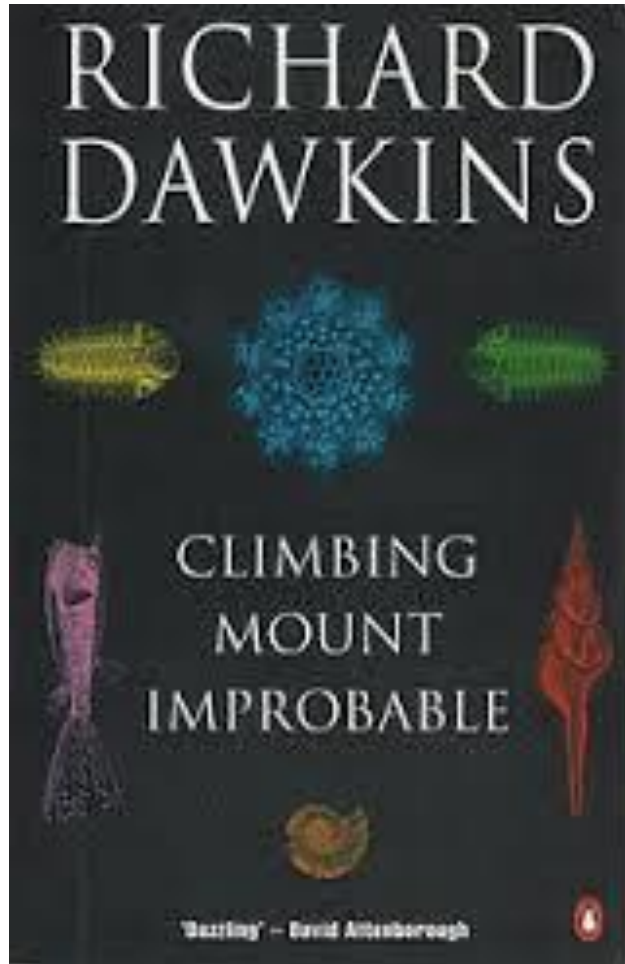
Evolution of the eye

½ billion years



Climbing mount improbable

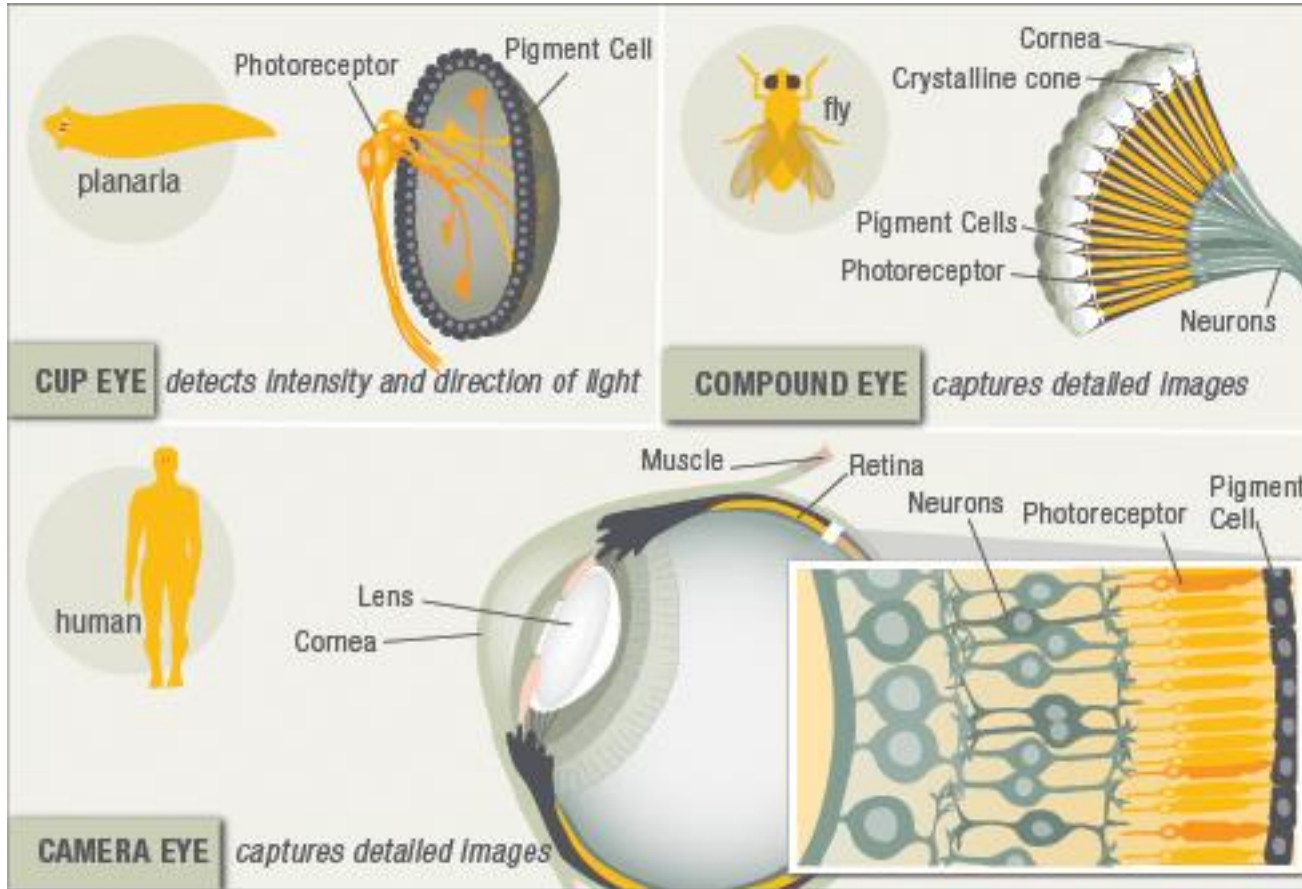
10 different designs



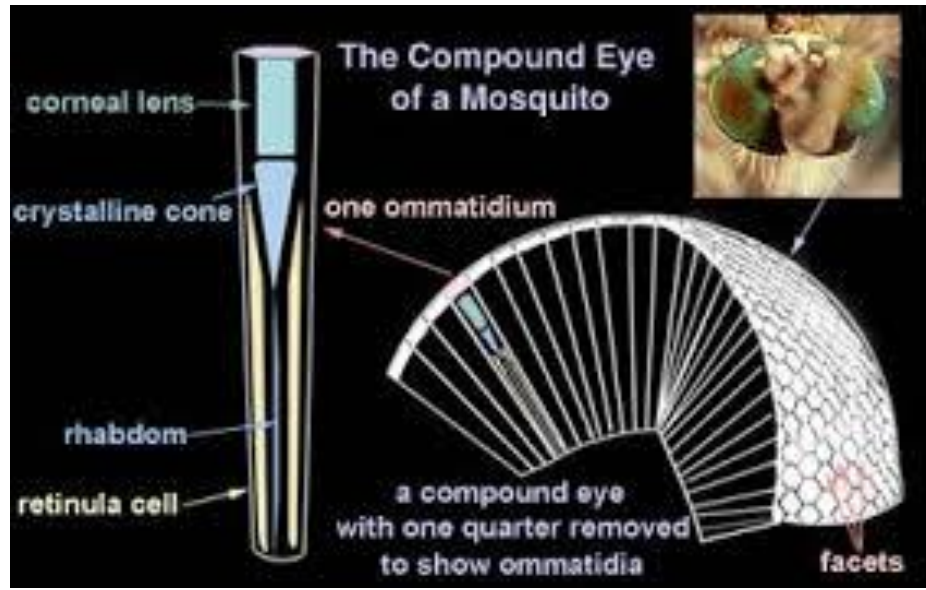
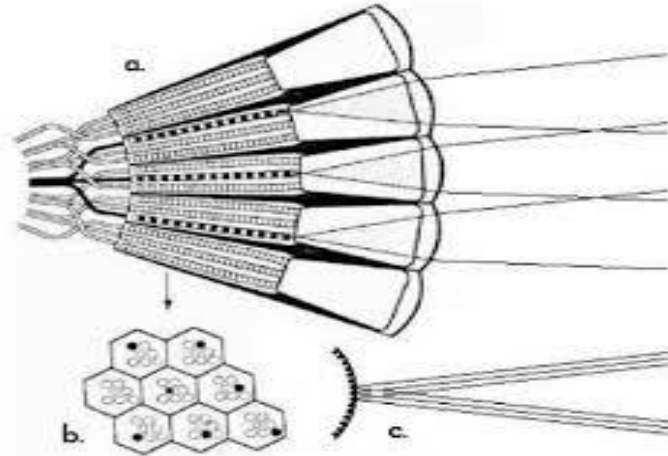
A plethora of eyes



Complex Eyes

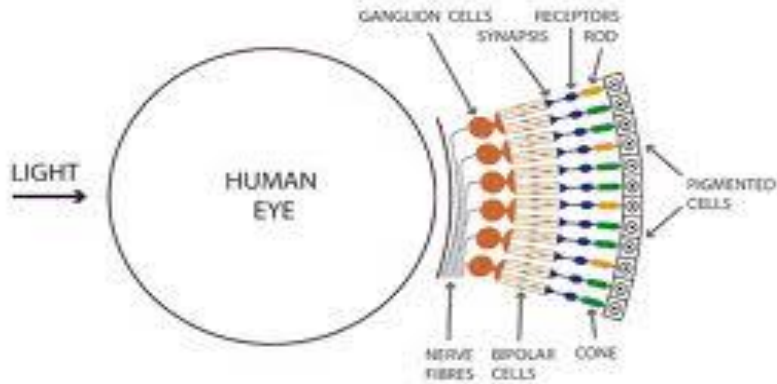


Compound eyes



Human Eyes

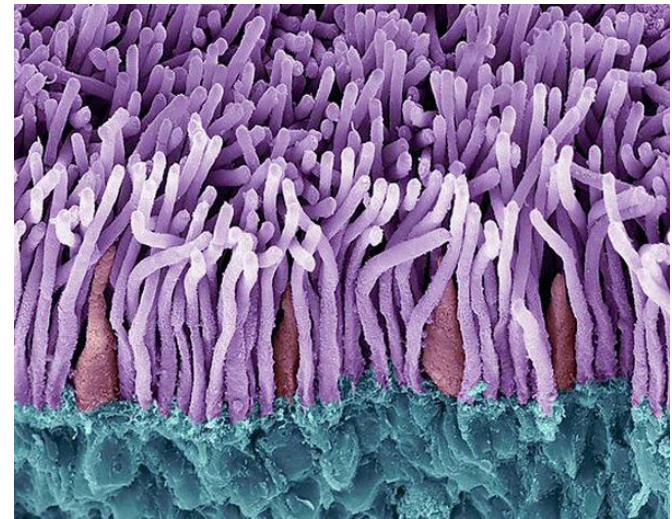
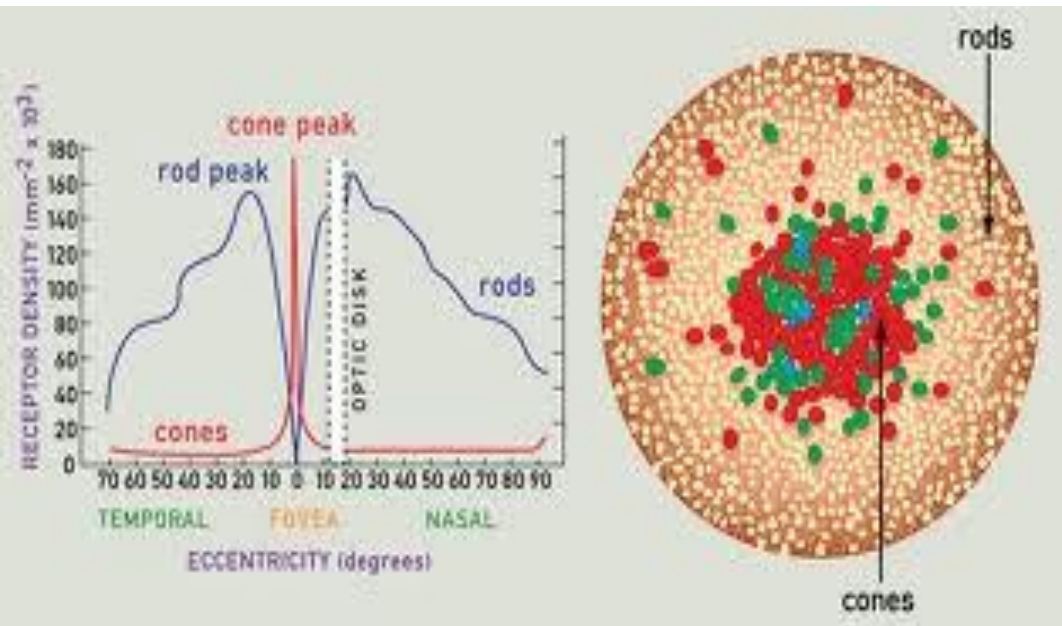
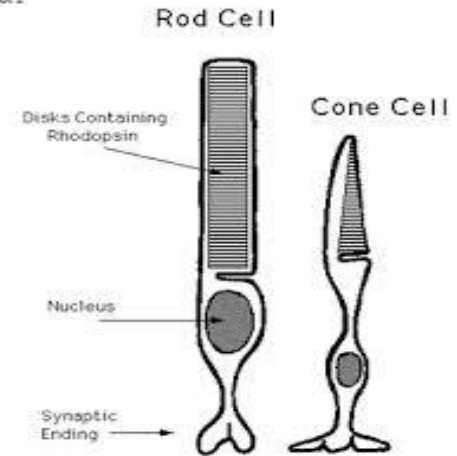
Basic Cross section of the Eye - Showing the Rods and Cones



Colors show from left to right

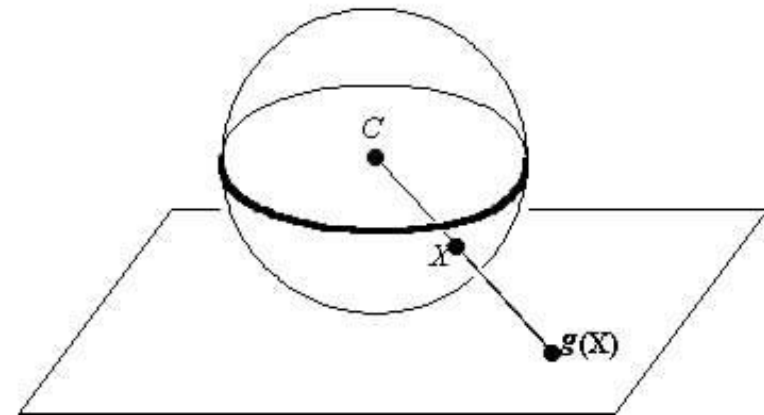
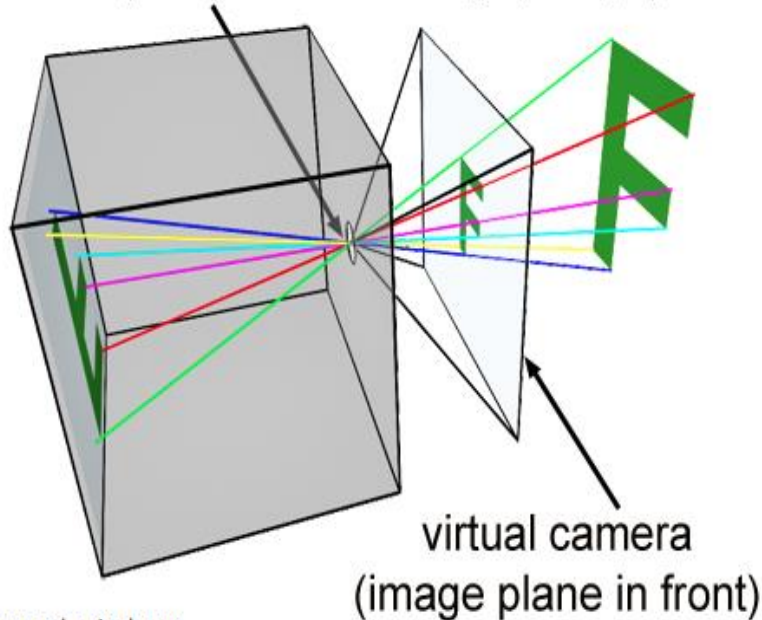
© Copyright Debra Therapy Training 2010

Figure 2



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

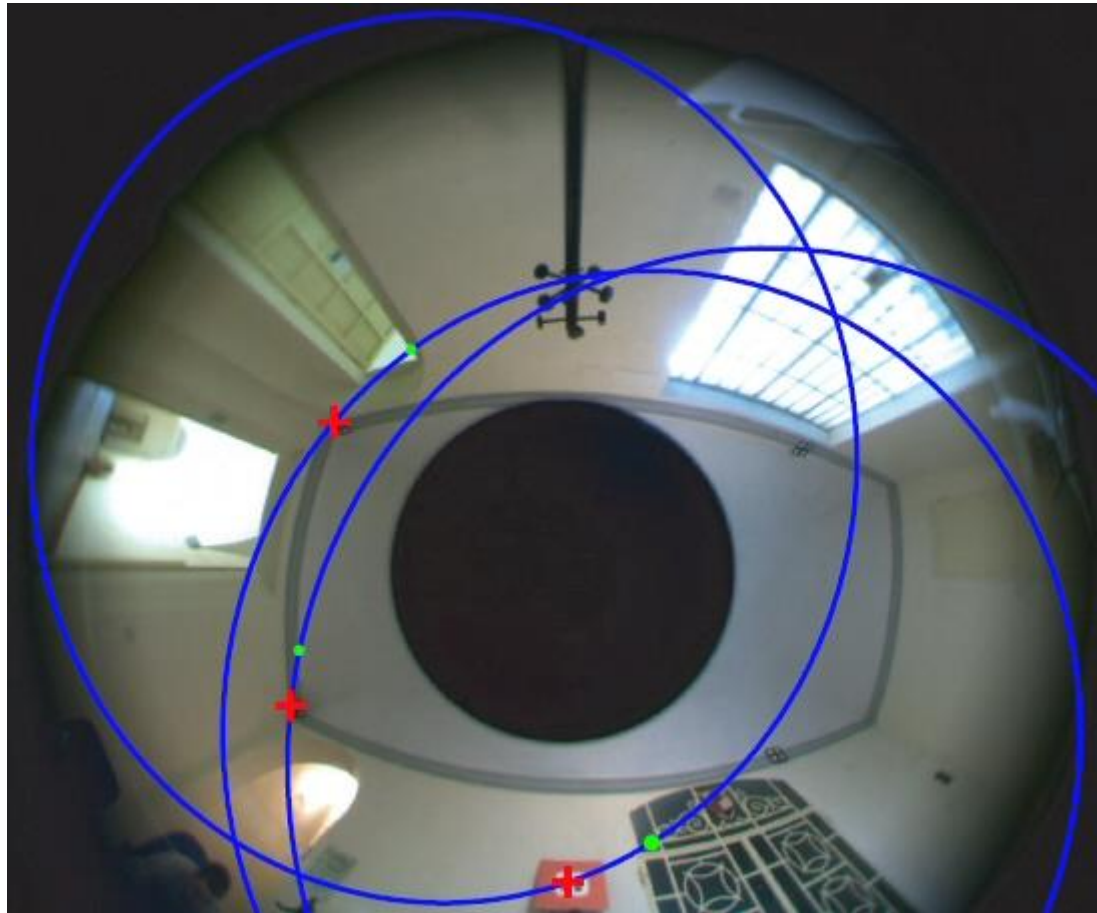
aperture (virtual camera origin, \approx eye)



Many cameras in the market



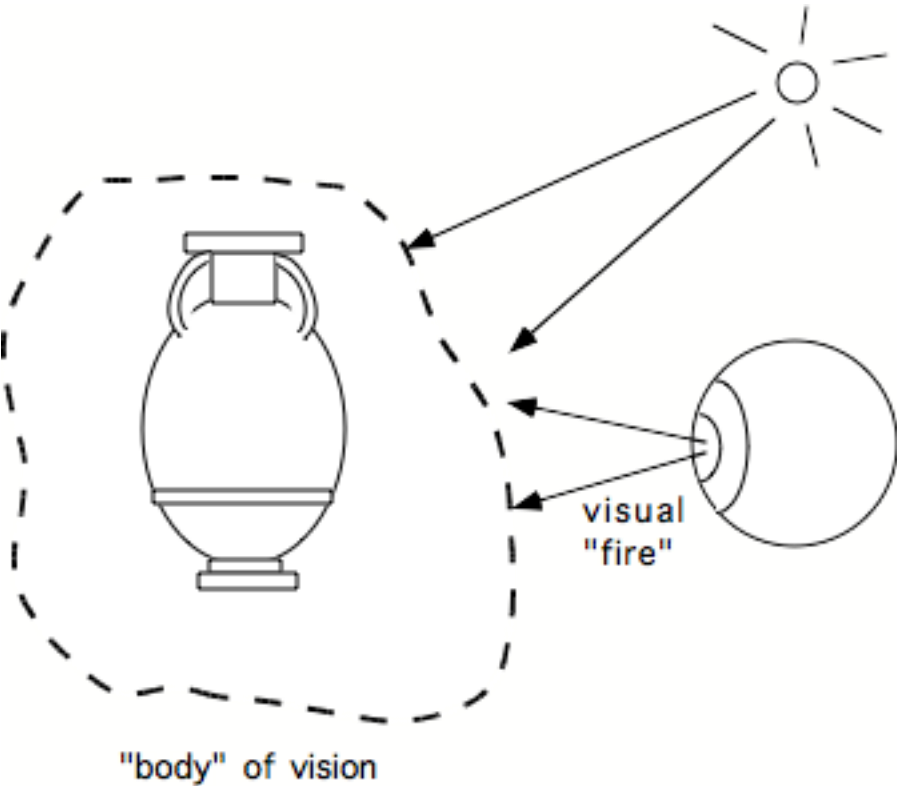
Catadioptric – panoramic images



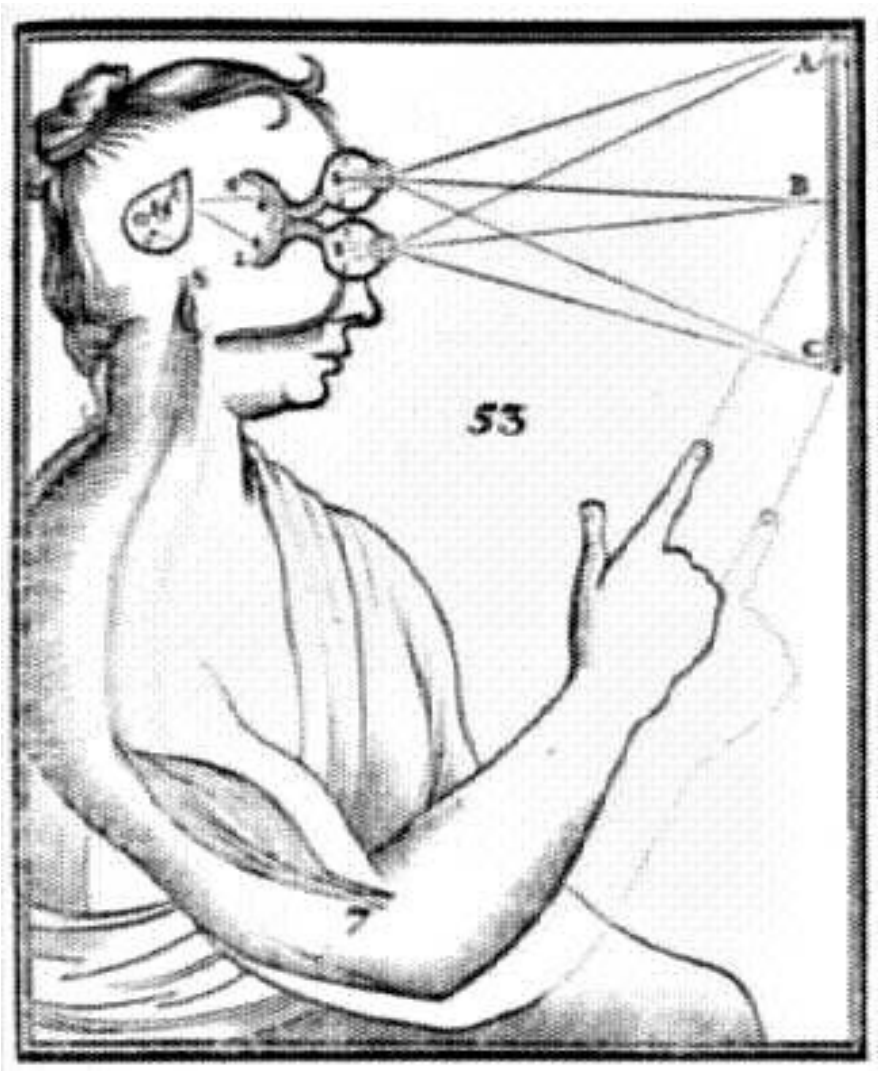


How does Vision work?

- Ancient Greeks: Extramission Theory



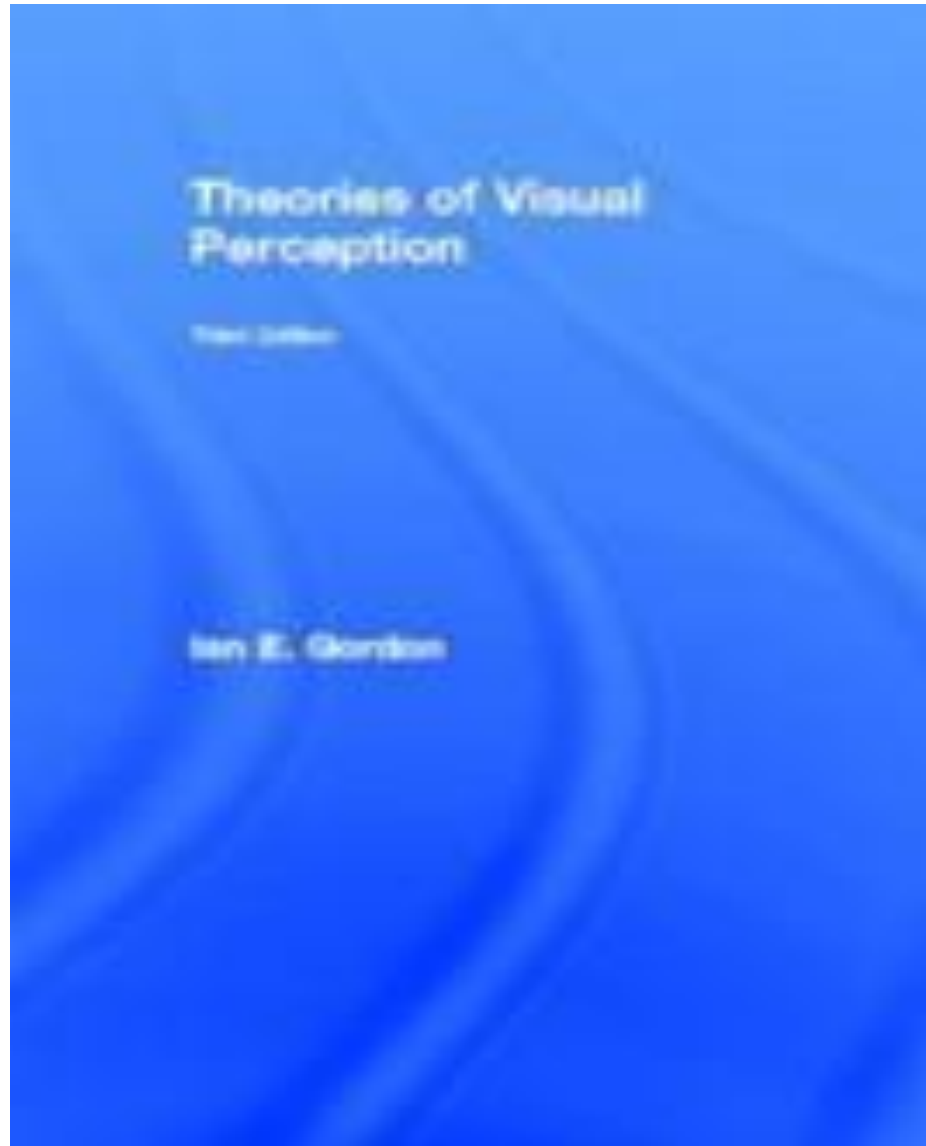
Descartes got it right




Many theories over the centuries

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

Theories influenced by the zeitgeist



Animal perception is active



Free examination. 1

Estimate material circumstances of the family 2

Give the ages of the people. 3

Surmise what the family had been doing before the arrival of the unexpected visitor. 4

Remember the clothes worn by the people. 5

Remember positions of people and objects in the room. 6

Estimate how long the visitor had been away from the family. 7

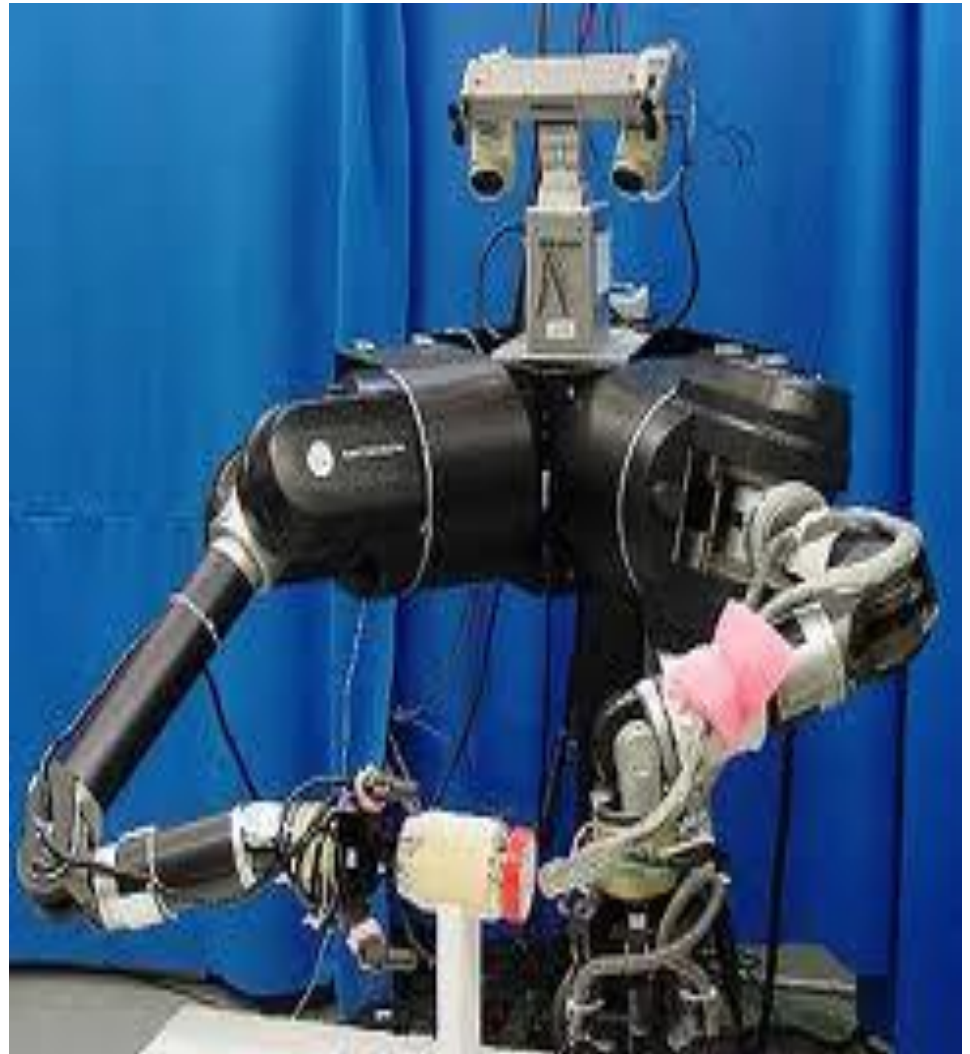
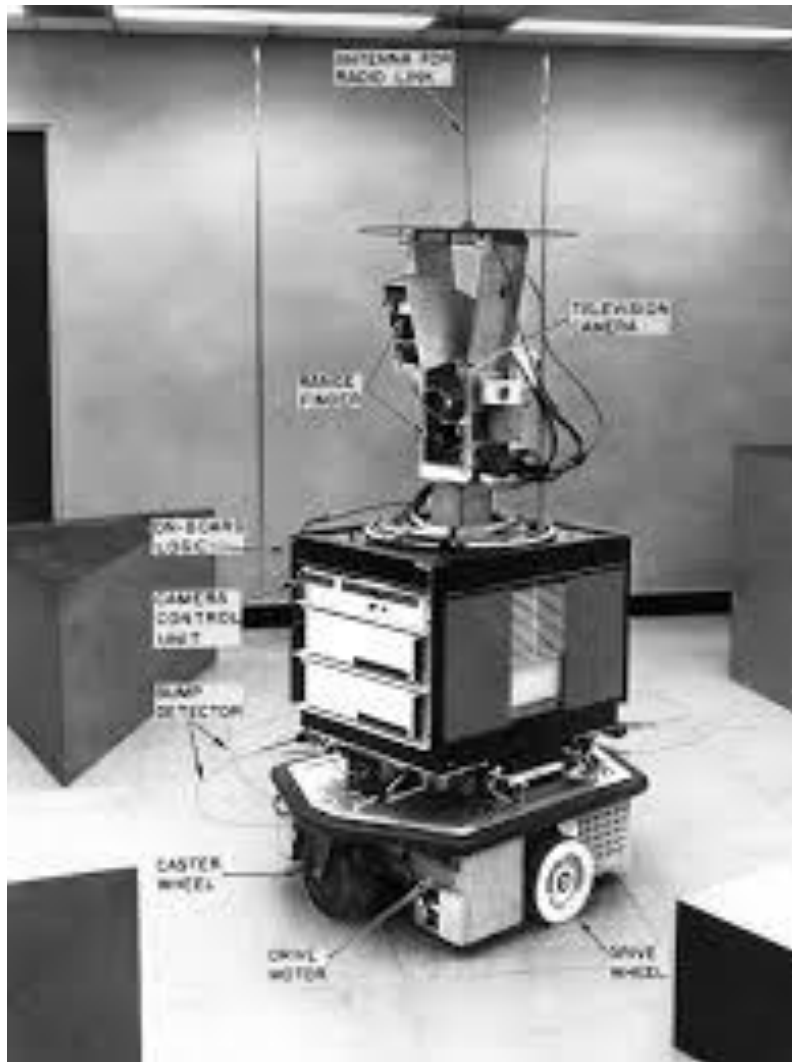
3 min. recordings of the same subject

The image displays a sequence of seven numbered diagrams (1-7) representing the active perception of a scene. Diagram 1 is a dense, chaotic scribble. Diagram 2 is a more structured scribble with some discernible shapes. Diagram 3 is a horizontal scribble with several distinct points. Diagram 4 is a horizontal scribble with a central cluster. Diagram 5 is a horizontal scribble with several distinct points. Diagram 6 is a dense, chaotic scribble. Diagram 7 is a horizontal scribble with a central cluster. The text '3 min. recordings of the same subject' is located to the right of diagram 7.

Measuring eye movements



Robots with Vision



PR2 Humanoid

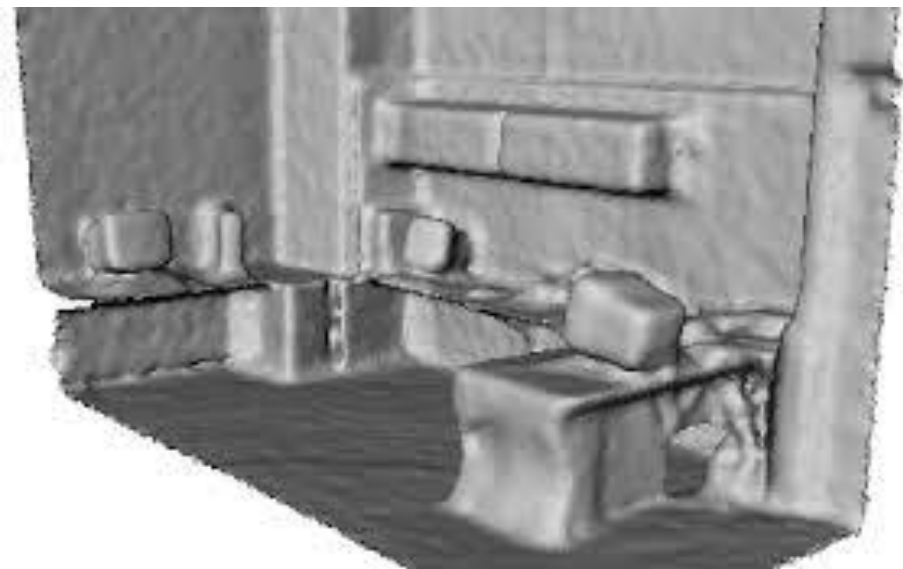


Perception for Robots

3 major problems

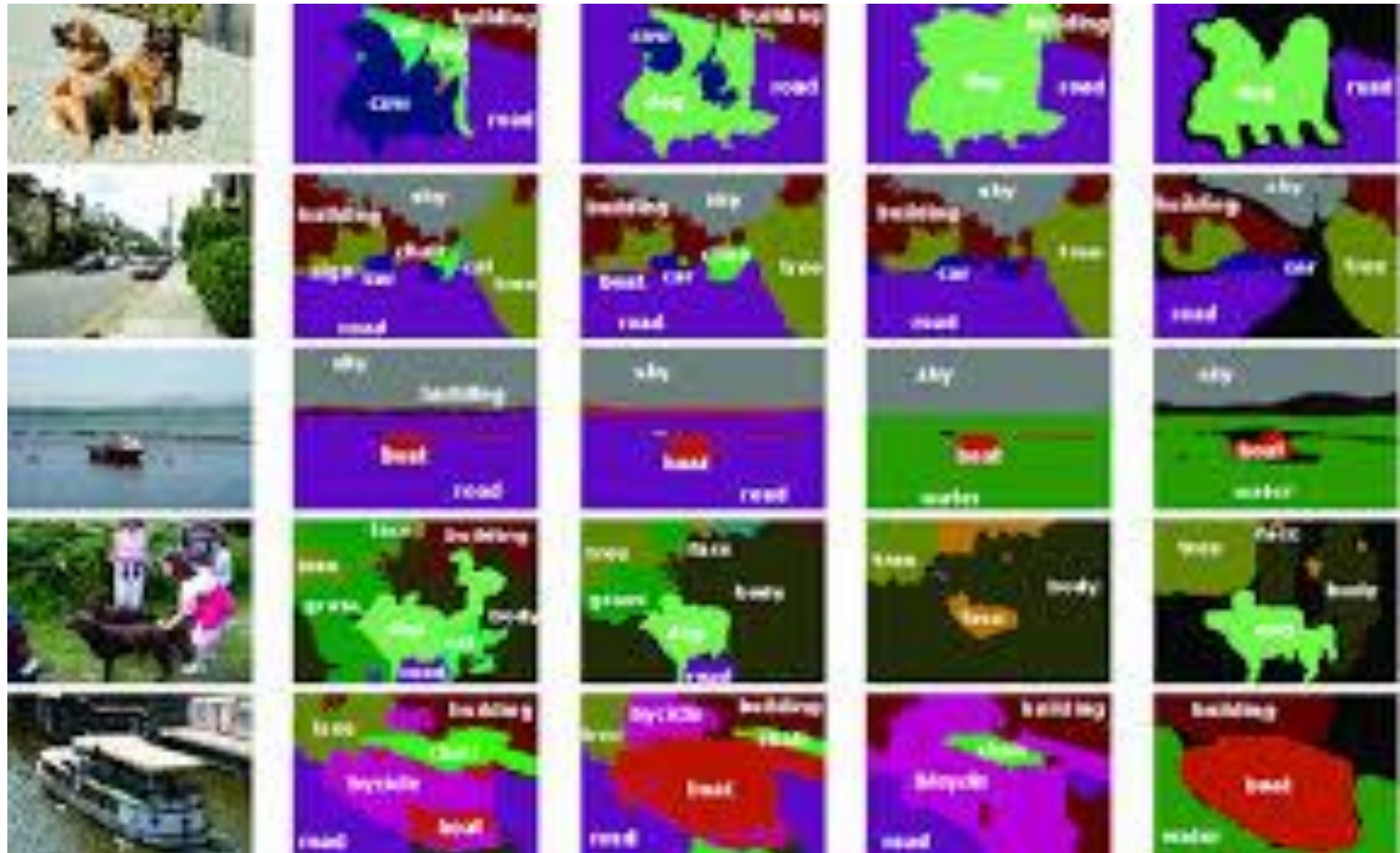
- Reconstruction
- Reorganization
- Recognition

Reconstruction

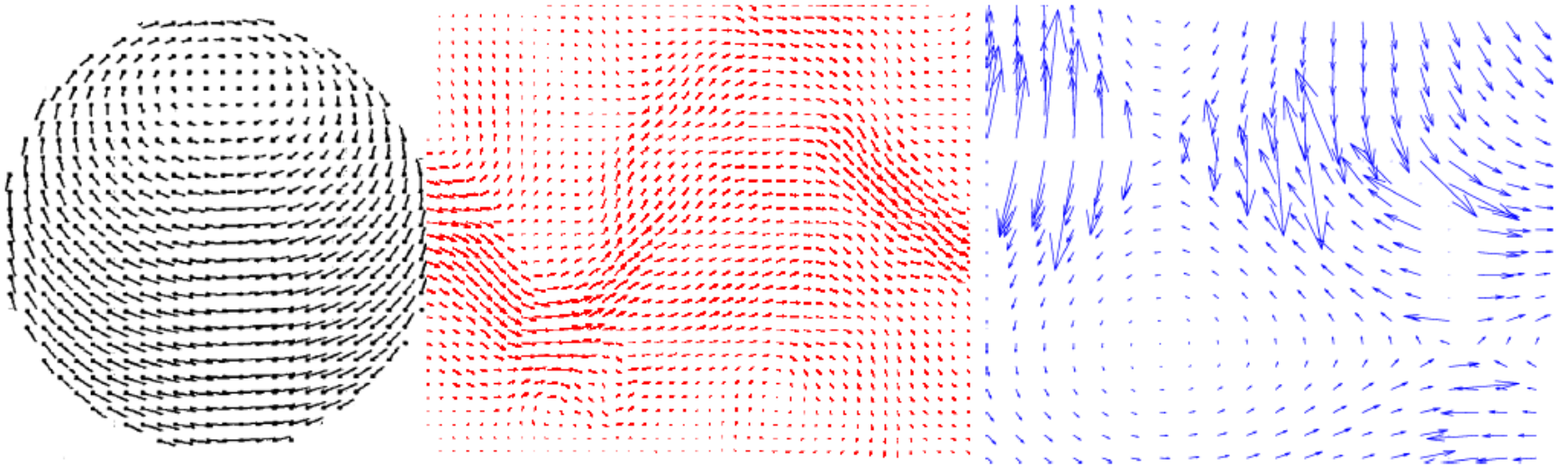




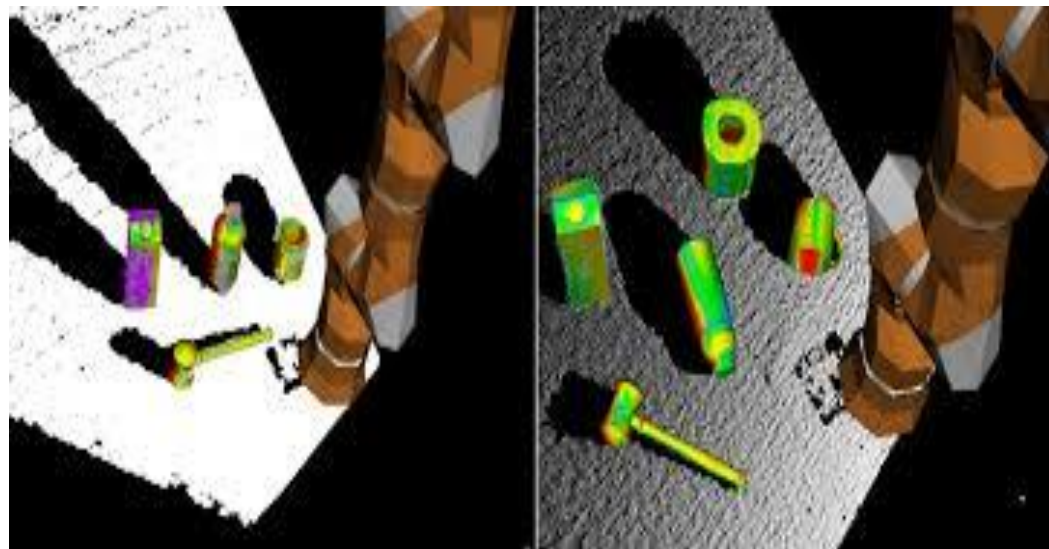
Reorganization: segmentation



Reorganization: flow



Recognition



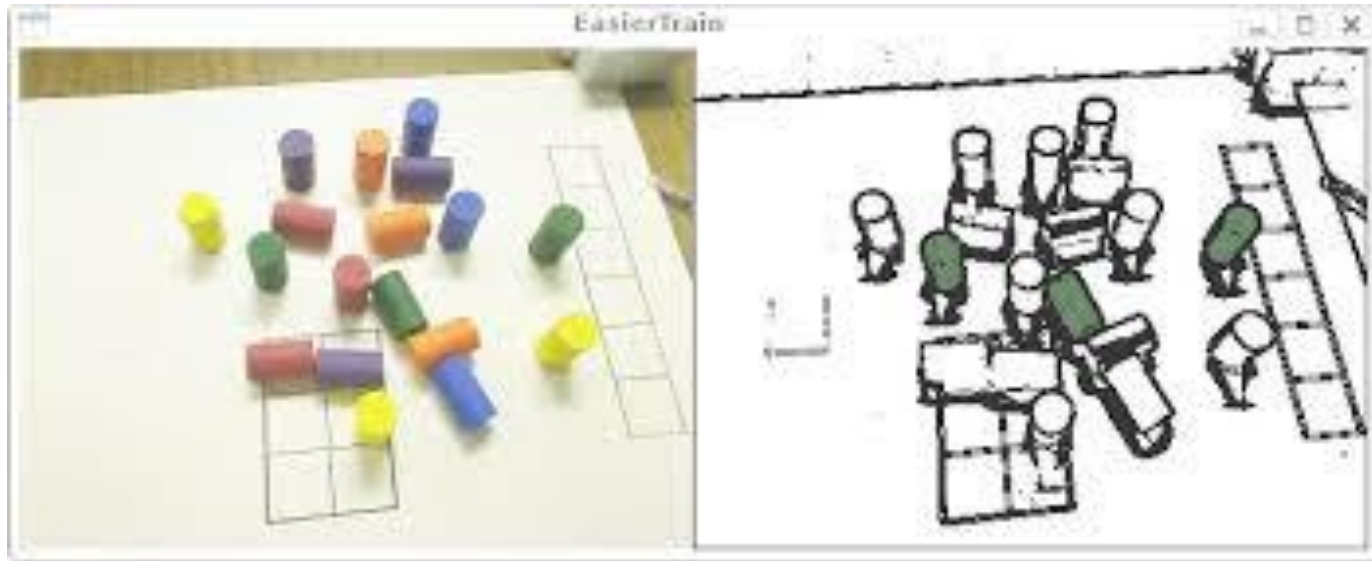
Images and Videos Contain

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

Lines



Color, Texture



Motion



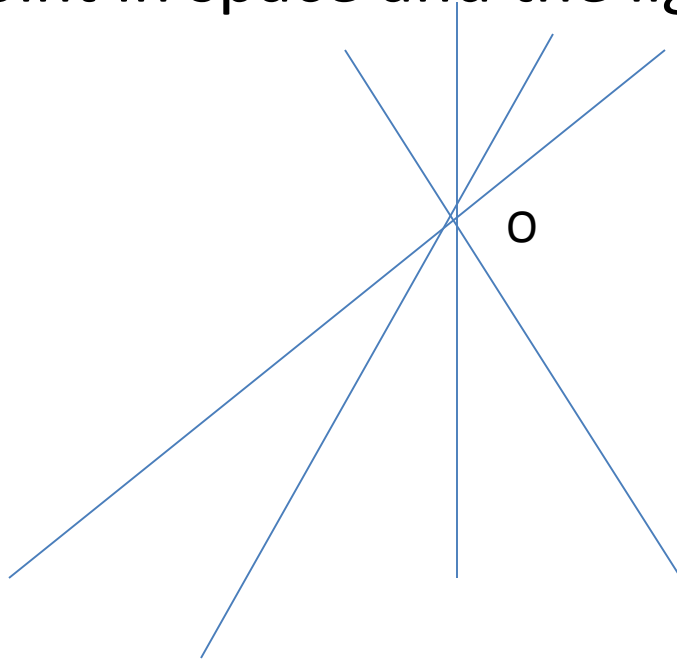
Contents of the Class

- **Image Processing:** Images, Light and Color, Filtering, Noise, Convolution, Edge detection, contour finding, texture analysis, segmentation and grouping.
- **3D Geometry:** Stereo, Multiple View Geometry, Epipolar Geometry, Projective Geometry
- **Motion:** Optical Flow, Egomotion, Motion Segmentation, Tracking
- **Navigation:** Map making, SLAM

Break

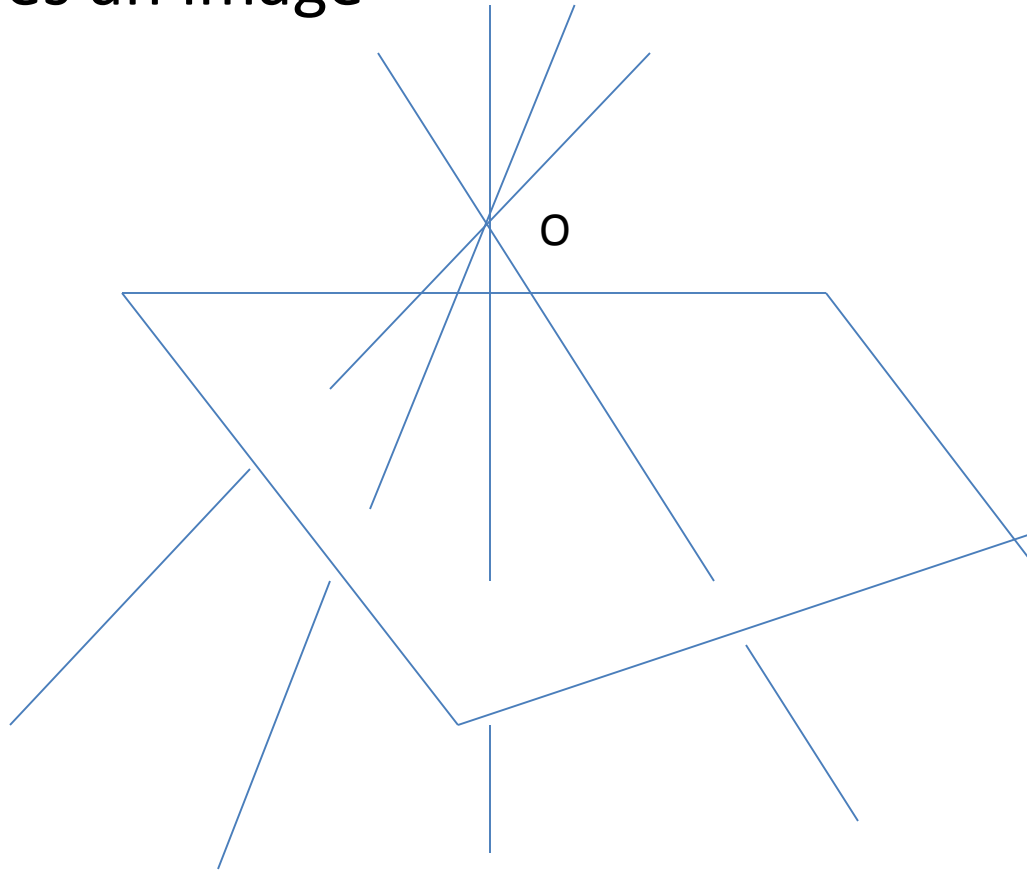
A theoretical model of an eye

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



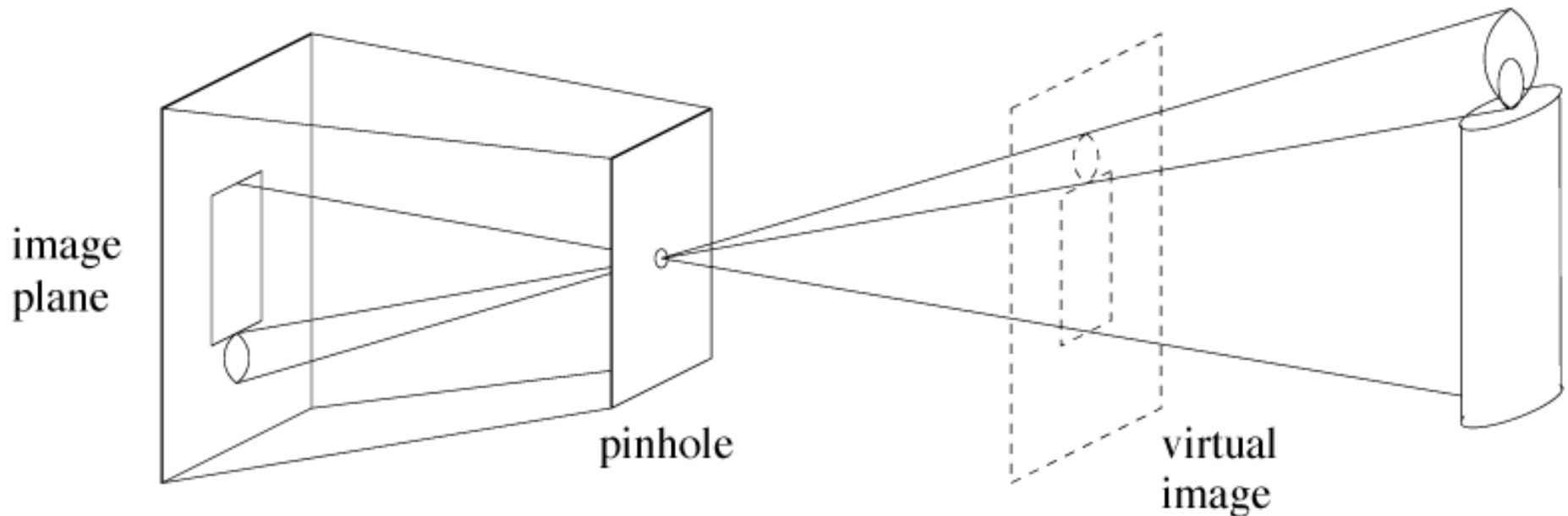
Then cut the rays with a plane

- This gives an image



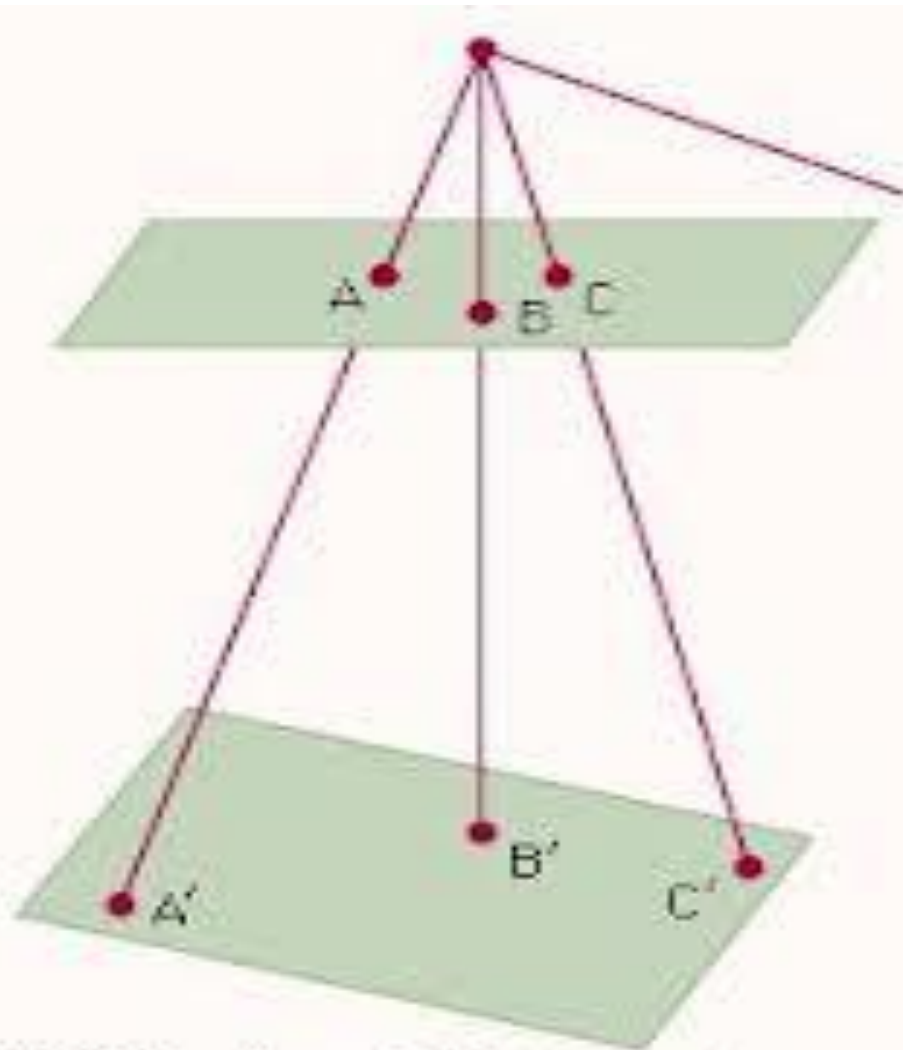
Pinhole cameras

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

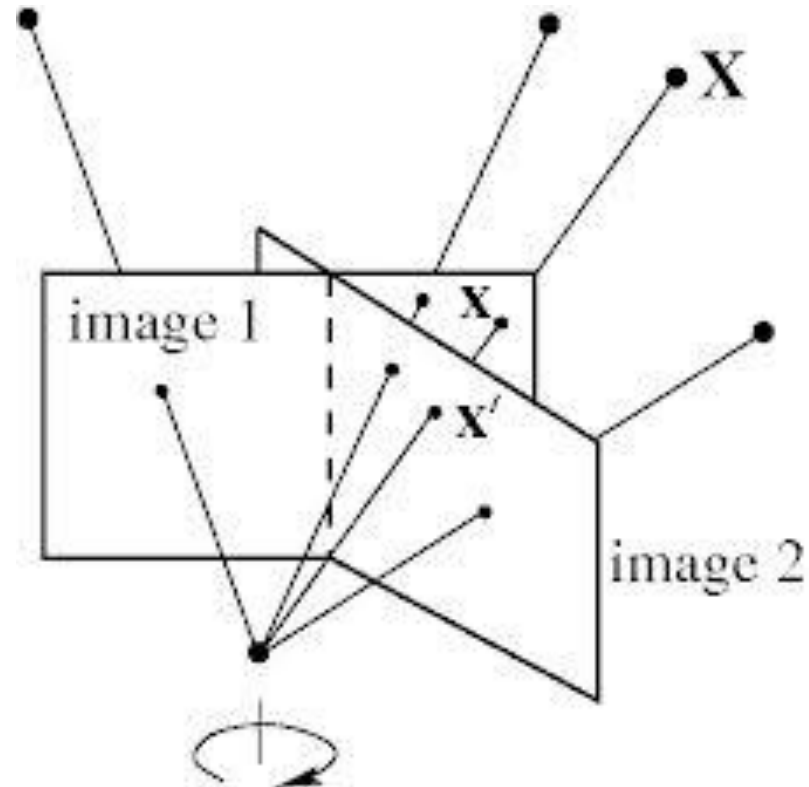
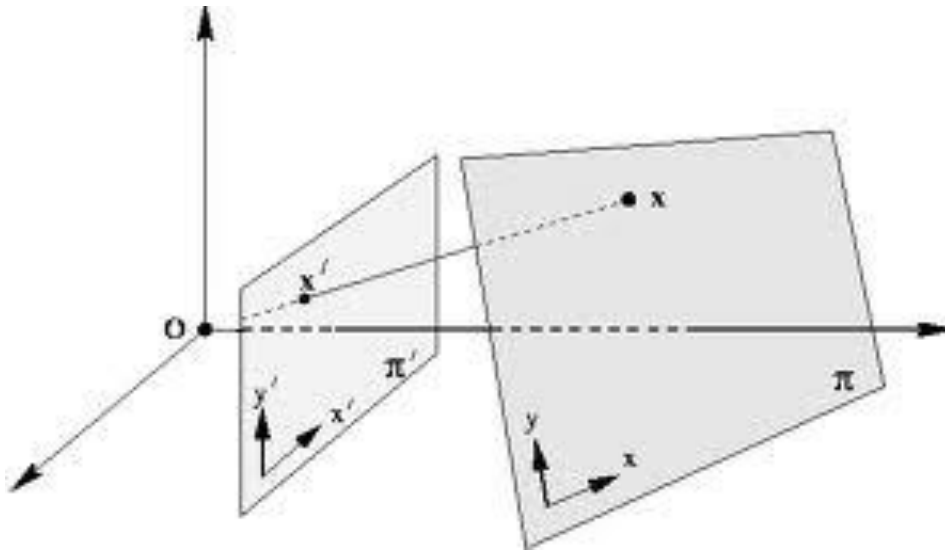


(Forsyth & Ponce)

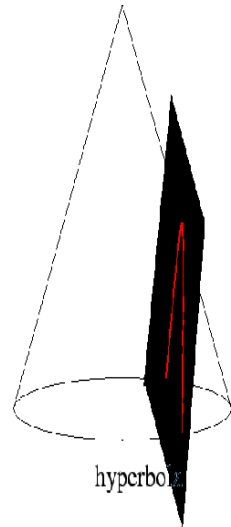
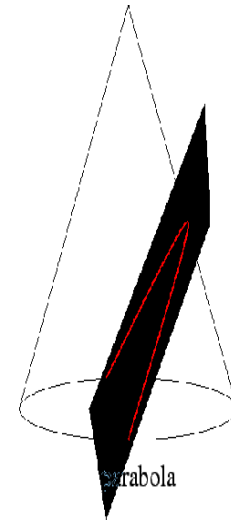
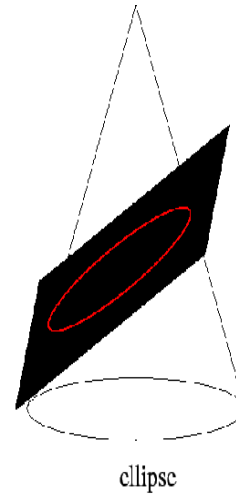
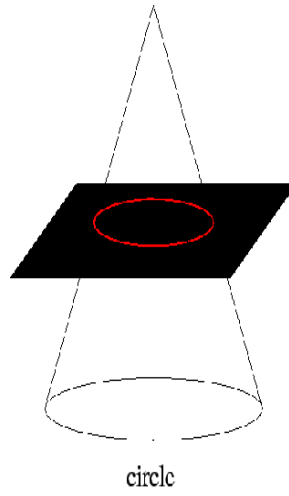
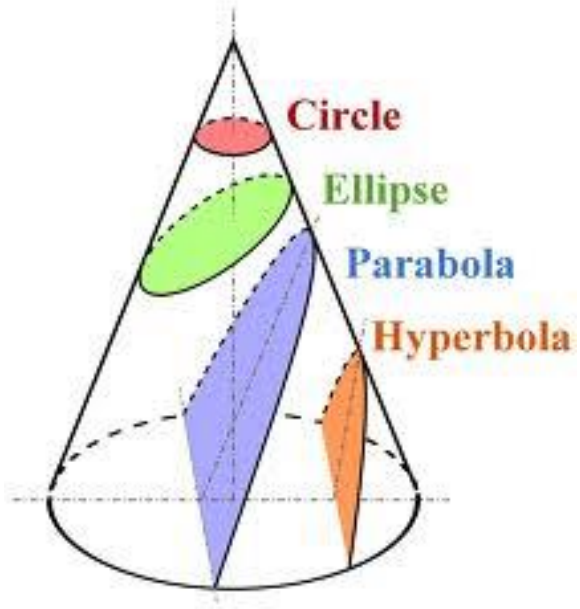
If we change the plane, we get an new image



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



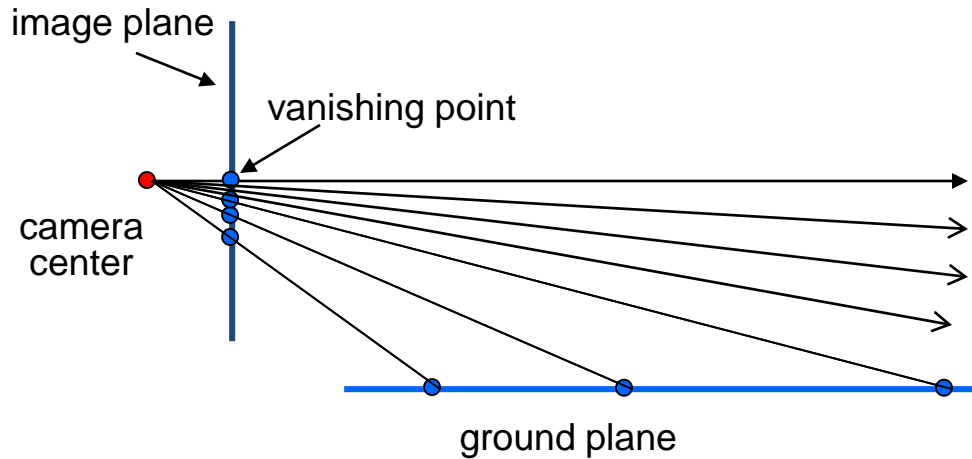
Conics



Projection of circle

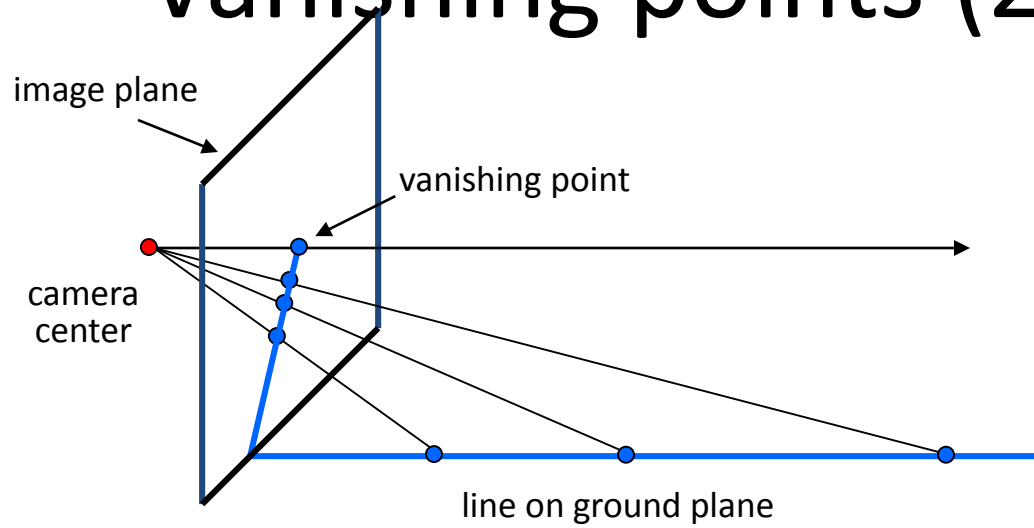


Vanishing points

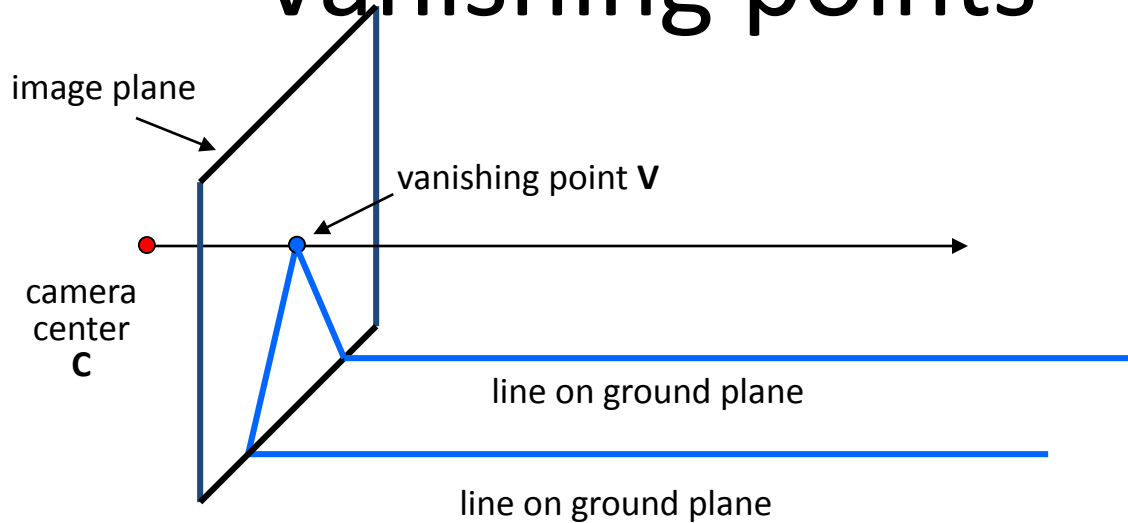


- Vanishing point
 - projection of a point at infinity

Vanishing points (2D)



Vanishing points

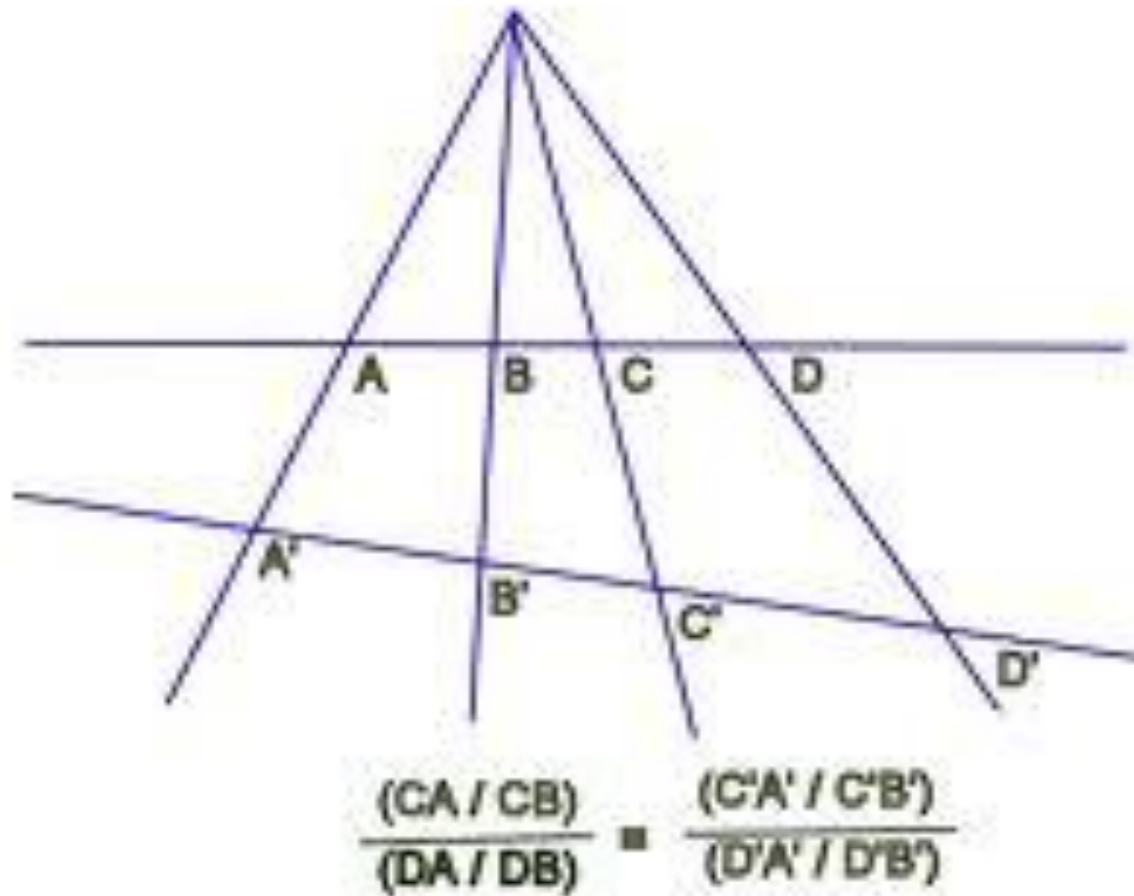


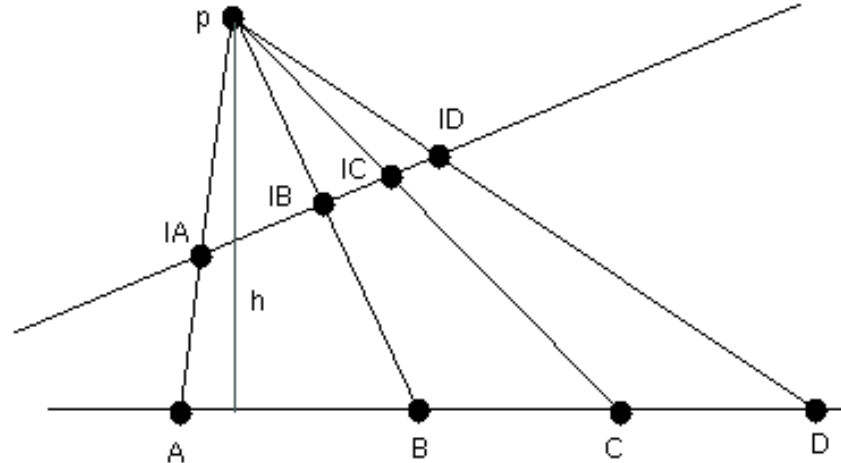
- Properties
 - Any two parallel lines have the same vanishing point v
 - The ray from C through v is parallel to the lines
 - An image may have more than one vanishing point

Parallelism (angles) not invariant



Cross ratio = only invariant





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 sine of the angle between them. Using this, we get:

$$\text{Area}(pAC) = h/2(AC) = 1/2 (pA)(pC) \sin(\angle ApC)$$

$$\text{Area}(pBC) = h/2(BC) = 1/2 (pB)(pC) \sin(\angle BpC)$$

$$\text{Area}(pAD) = h/2(AD) = 1/2 (pA)(pD) \sin(\angle ApD)$$

$$\text{Area}(pBD) = h/2(BD) = 1/2 (pB)(pD) \sin(\angle BpD)$$

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

$$= [(pA)(pC) \sin(\angle ApC)/(pB)(pC) \sin(\angle BpC)]$$

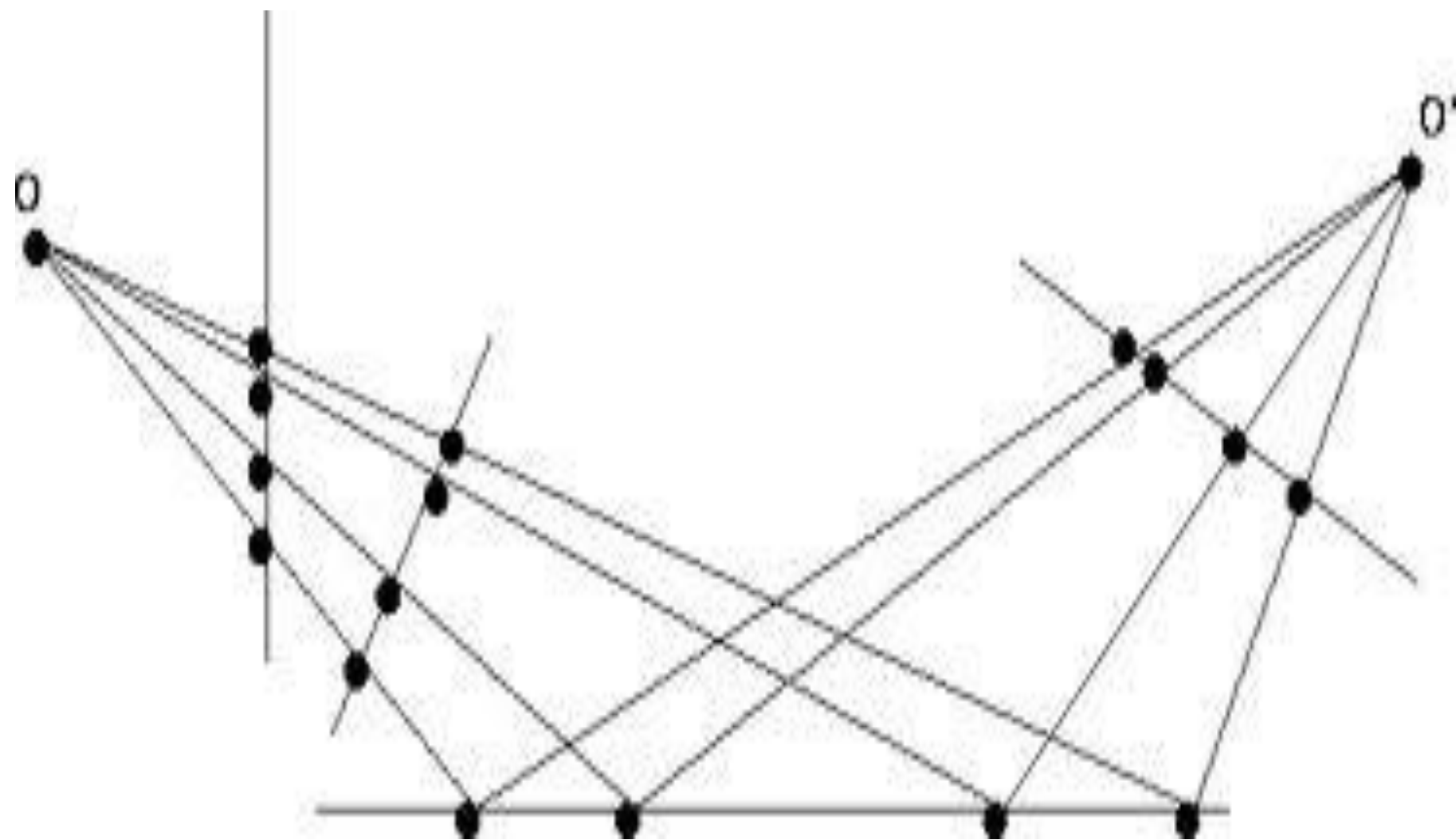
$$-----$$

$$[(pA)(pD) \sin(\angle ApD)/(pB)(pD) \sin(\angle BpD)]$$

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

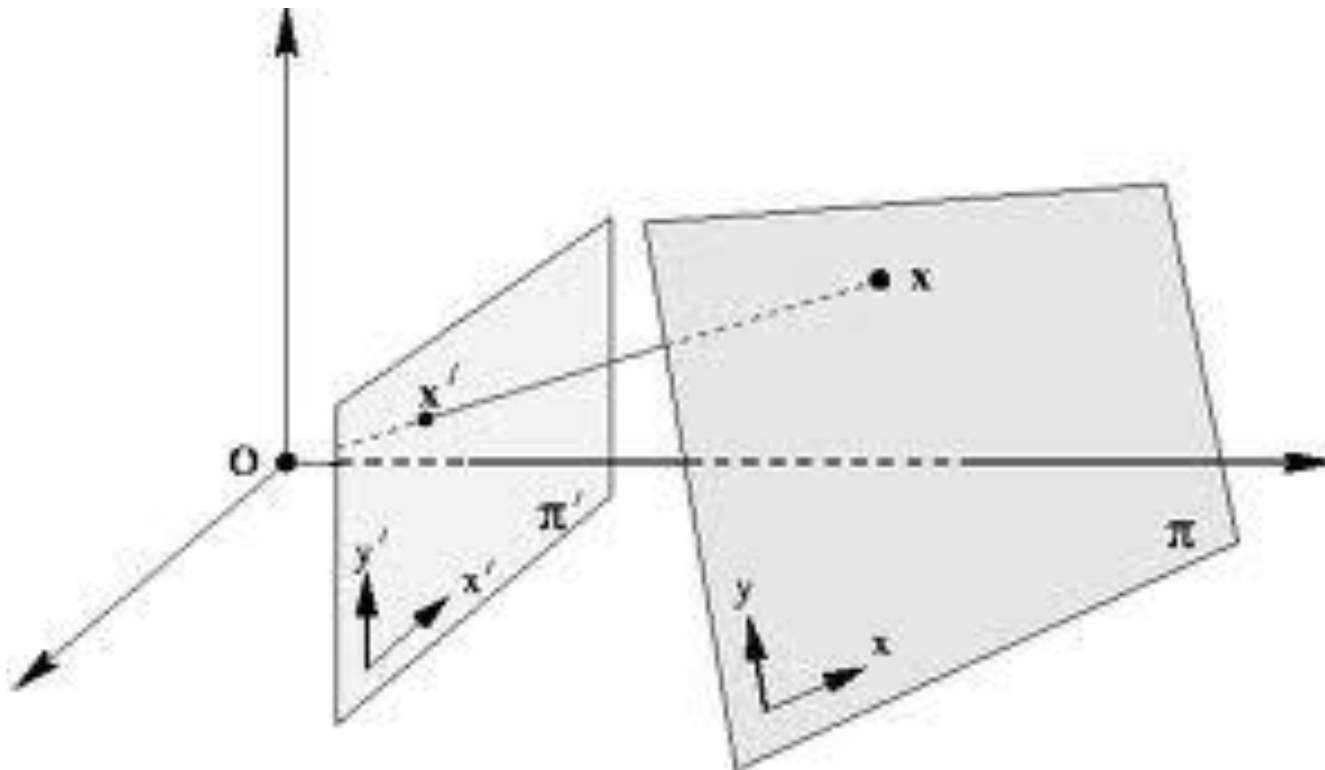
This last quantity is independent of the line we project to. Thus cross ratios are invariant under projection.

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

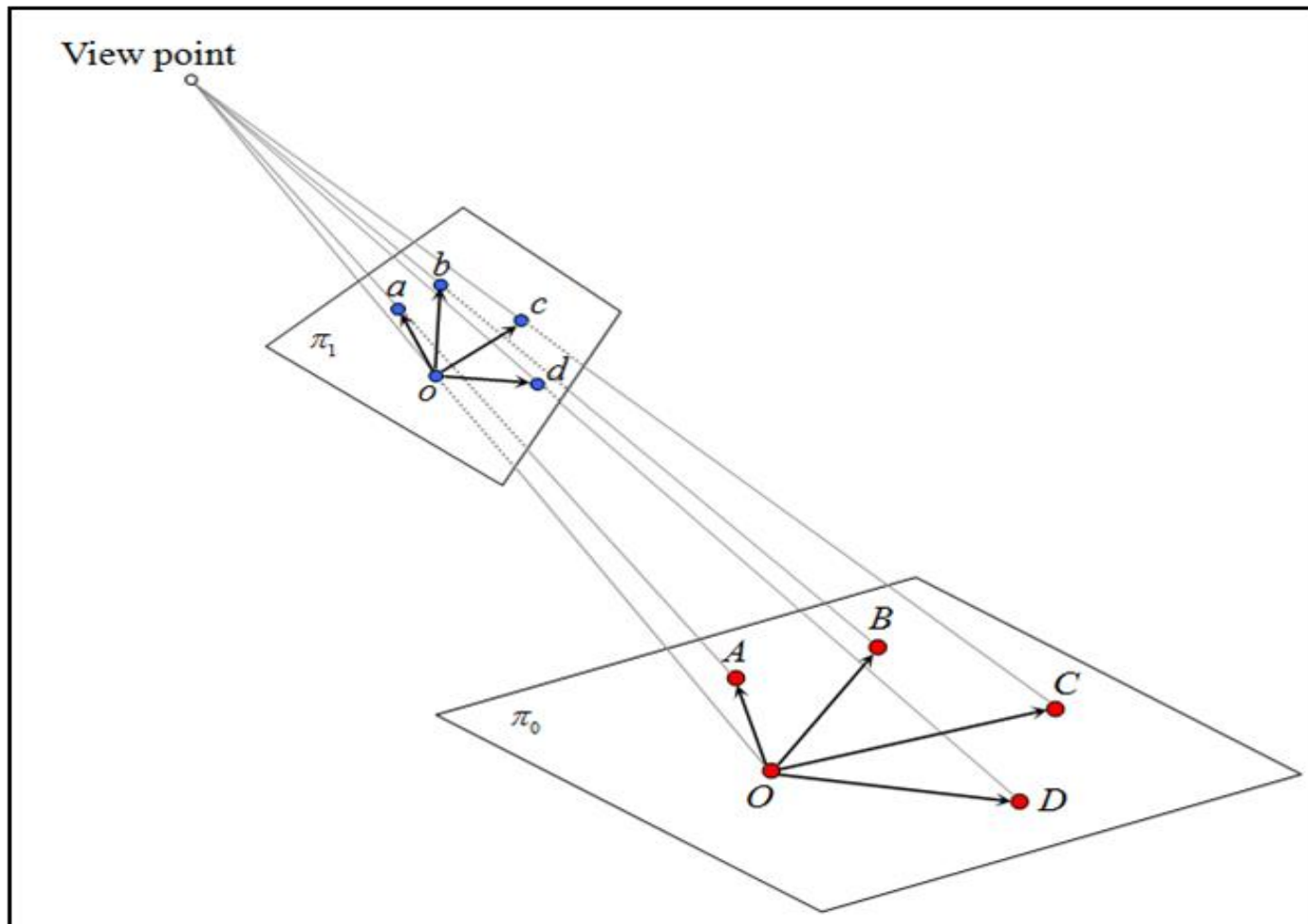


Back to our question: how are the 2 images related to each other

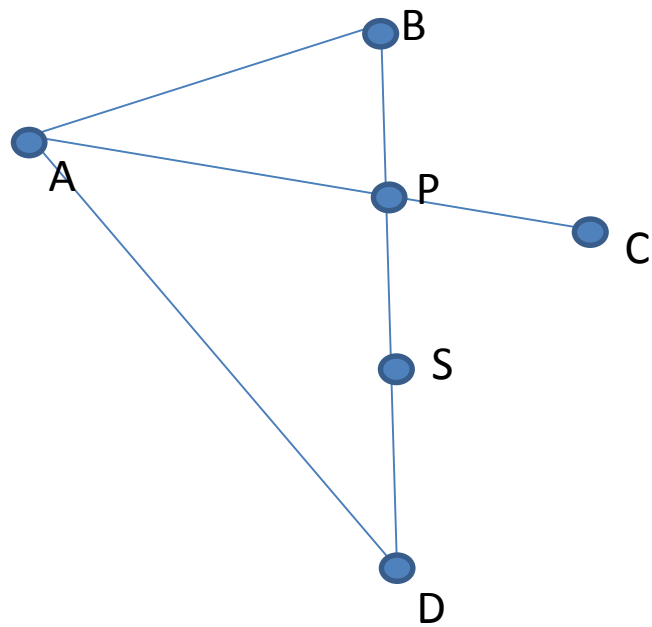
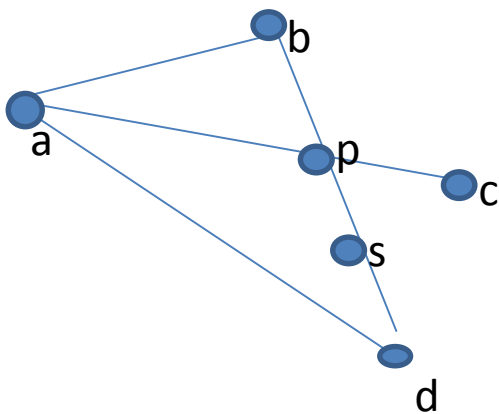
Can we find a map, a function mapping x' to x ?



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

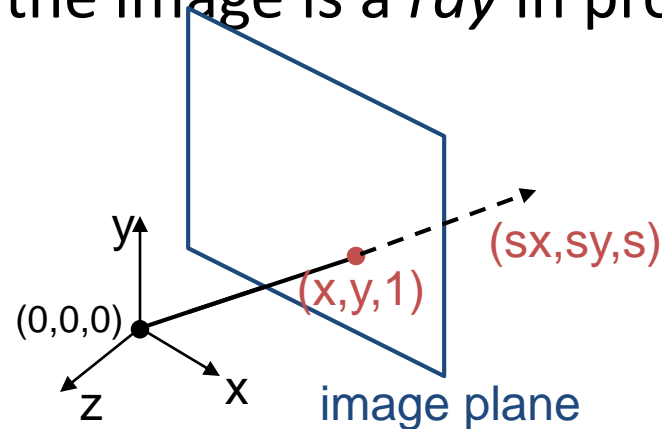


Proof



The projective plane

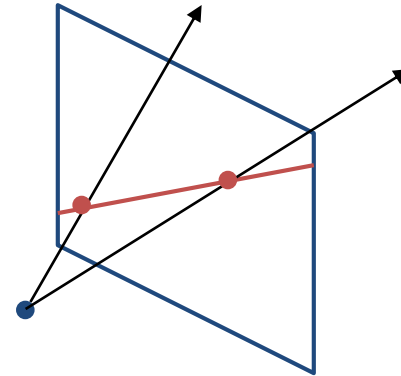
- Why do we need homogeneous coordinates?
 - represent points at infinity, homographies, perspective projection, multi-view relationships
- What is the geometric intuition?
 - a point in the image is a *ray* in projective space



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

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: $ax + by + cz = 0$

in vector notation :

$$0 = \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

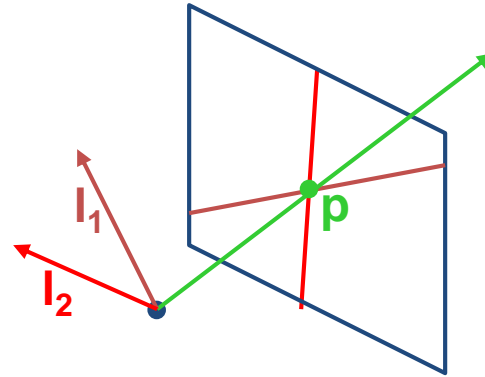
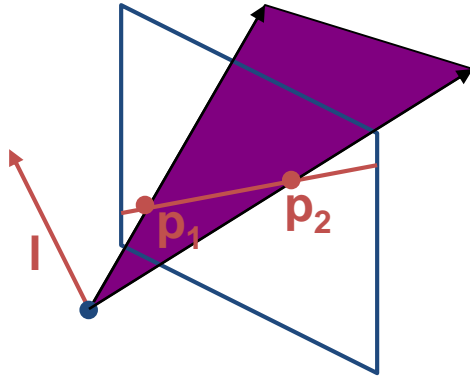
l **p**

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

Point and line duality

– A line l is a homogeneous 3-vector

– It is \perp to every point (ray) p on the line: $l \cdot p = 0$



What is the line l spanned by rays p_1 and p_2 ?

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

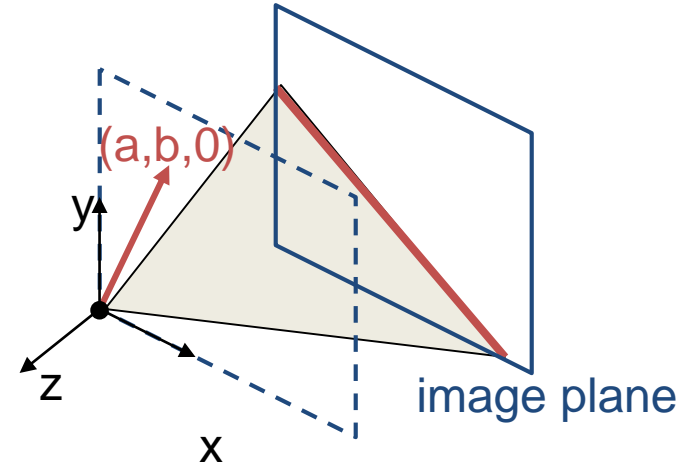
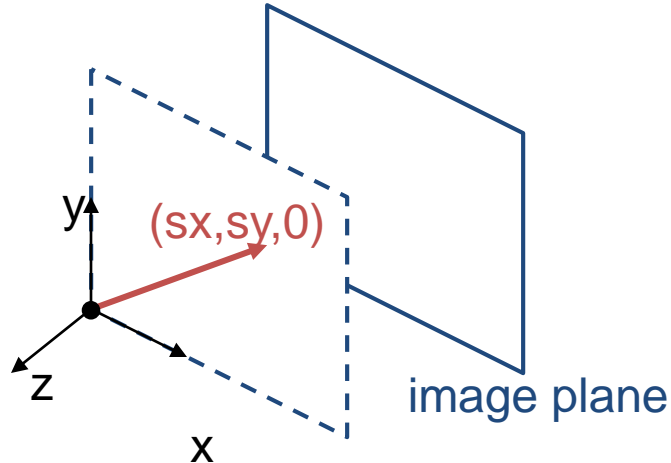
What is the intersection of two lines l_1 and l_2 ?

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

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



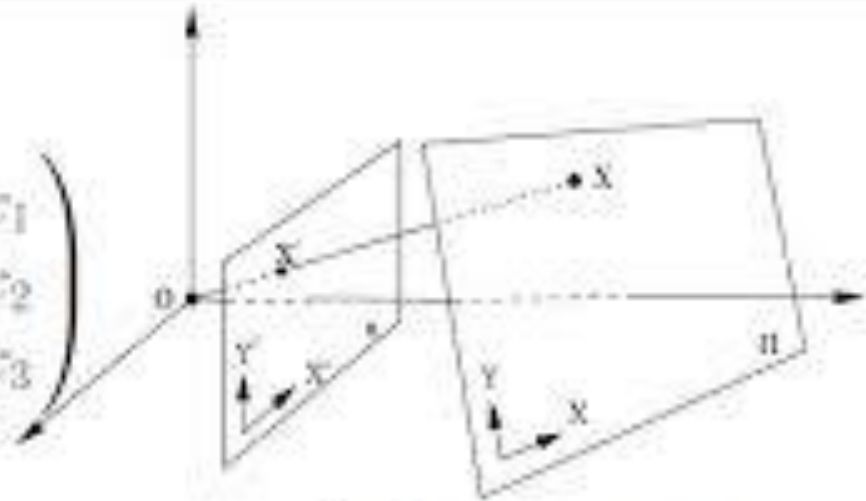
- Ideal point (“point at infinity”)
 - $p \cong (x, y, 0)$ – parallel to image plane
 - It has infinite image coordinates

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 $x' = Hx$, where H is a 3×3 non-singular homogeneous matrix.

Special Projectivities



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

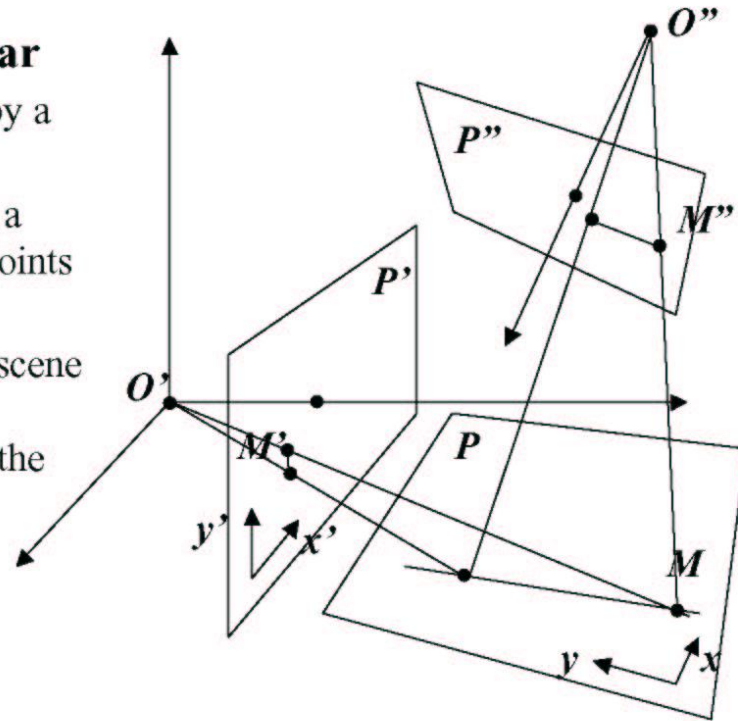
Examples of Projective Transformations



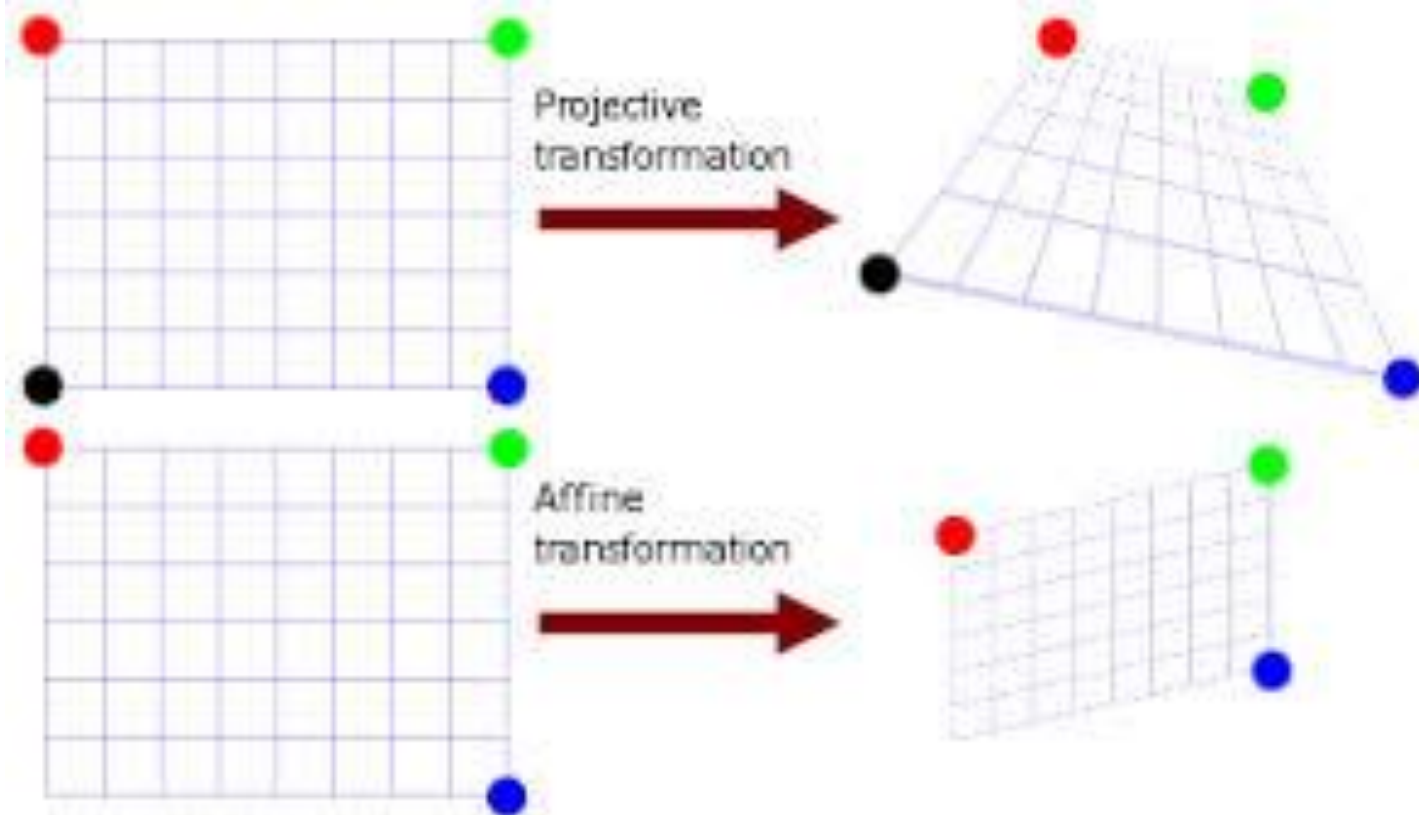
- Central projection maps **planar scene** points to image plane by a projectivity
 - True because all points on a scene line are mapped to points on its image line
- The image of the same planar scene from a second camera can be obtained from the image from the first camera by a projectivity
 - True because

$$\mathbf{x}'_i = \mathbf{H}' \mathbf{x}_i, \mathbf{x}''_i = \mathbf{H}'' \mathbf{x}_i$$

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



Projective vs Affine



Rectification

