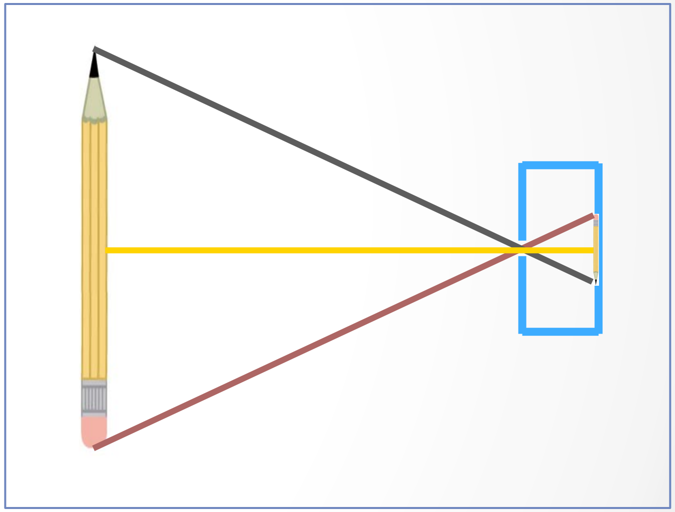
CSE578: Computer Vision

Spring 2017

01-Geometry: Imaging and Camera Model

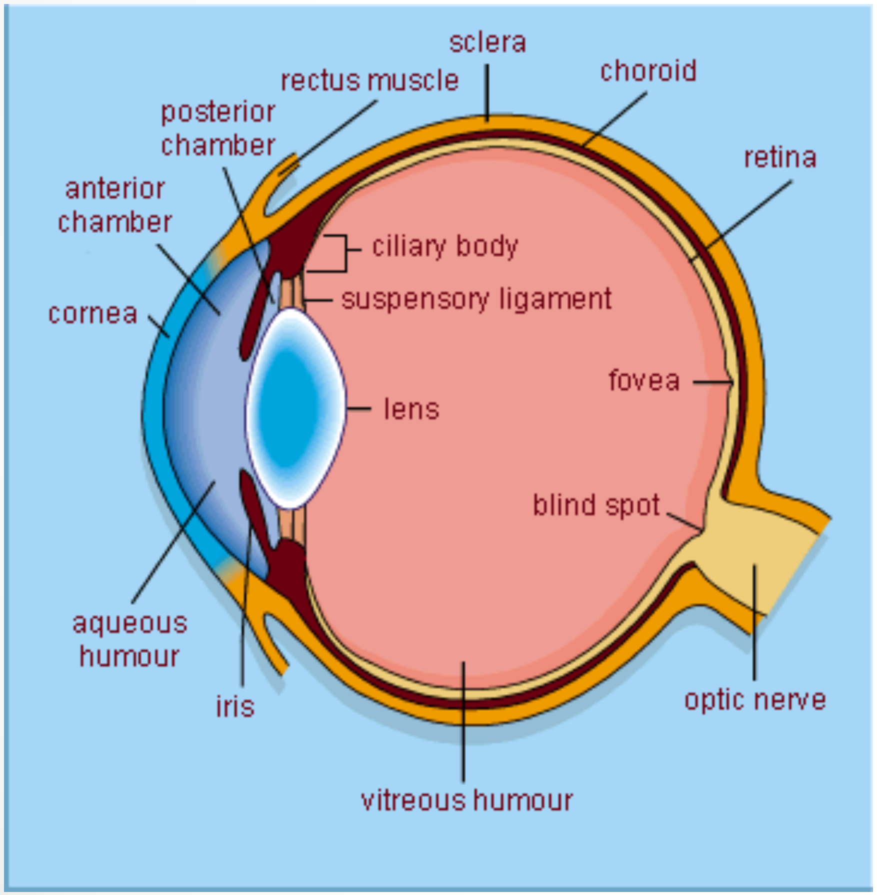


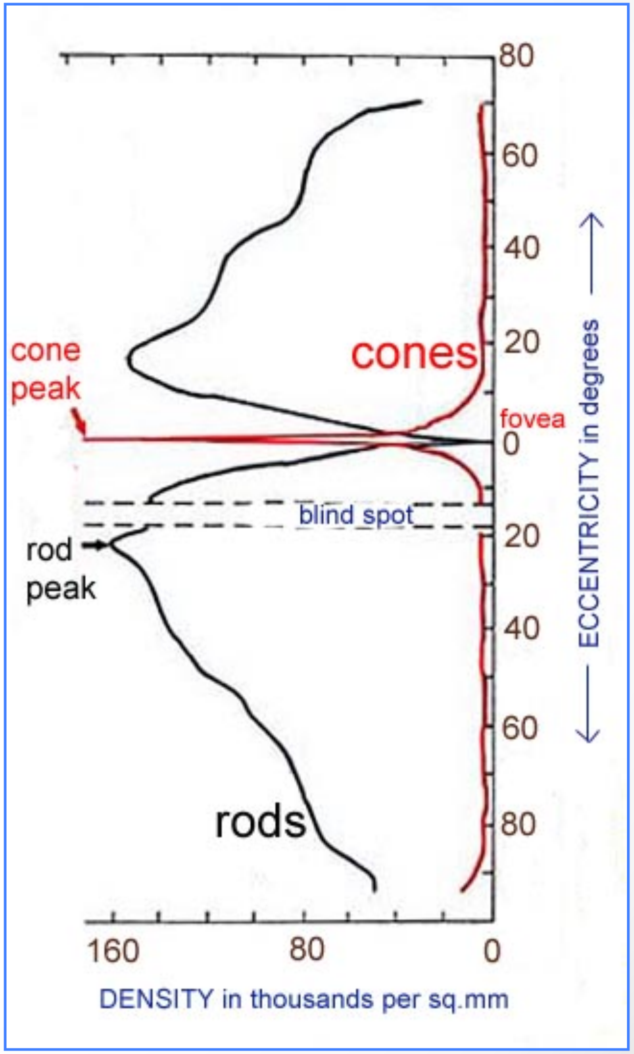
Anoop M. Namboodiri

Center for Visual Information Technology

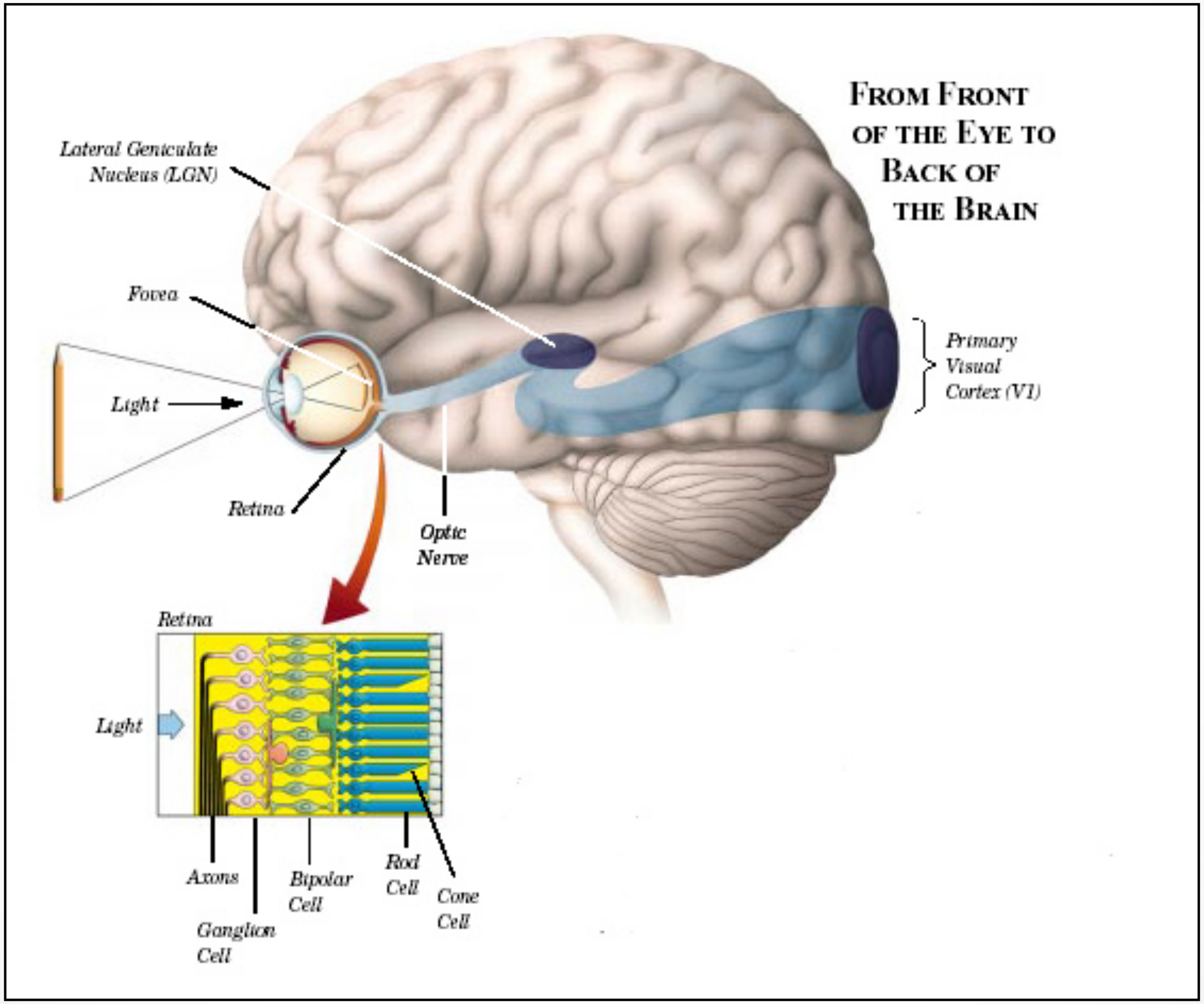
IIIT Hyderabad, INDIA

Human Eye


sclera
rectus muscle
choroid
posterior chamber
retina
anterior chamber
ciliary body
suspensory ligament
cornea
fovea
lens
blind spot
aqueous humour
iris
optic nerve
vitreous humour

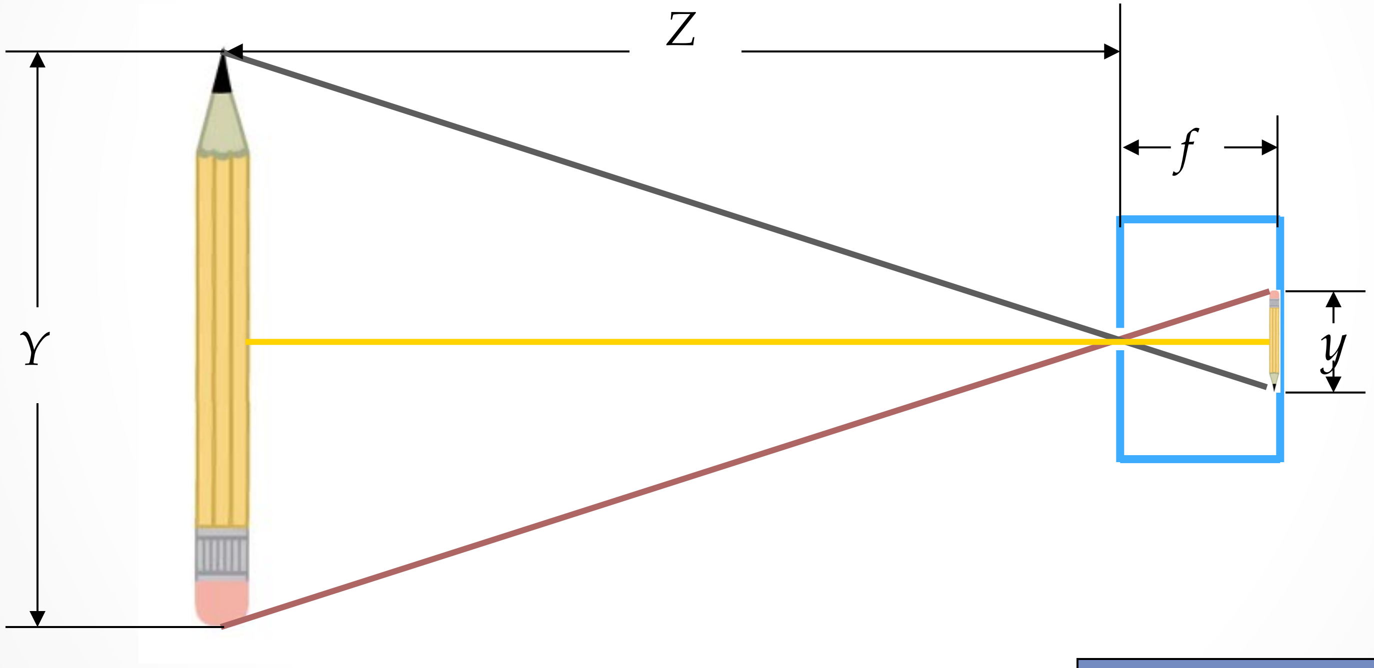

80
60
40
20
0
20
40
60
80
cone peak
cones
fovea
blind spot
rod peak
rods
ECCENTRICITY in degrees
160
80
0
DENSITY in thousands per sq.mm

Human Visual System



http://www.hhmi.org/senses

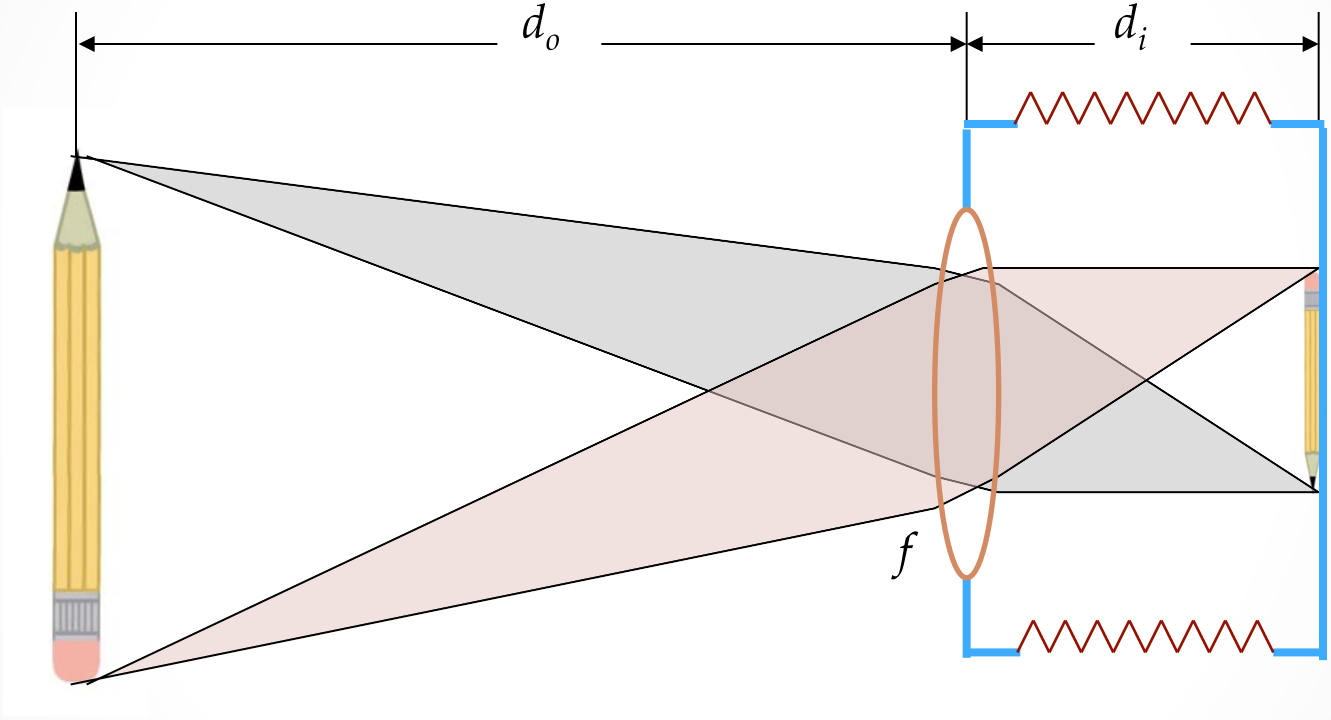
The Pinhole Camera



$y = f \,.\, Y / Z$

Pinhole Camera in practice

Camera with Lens

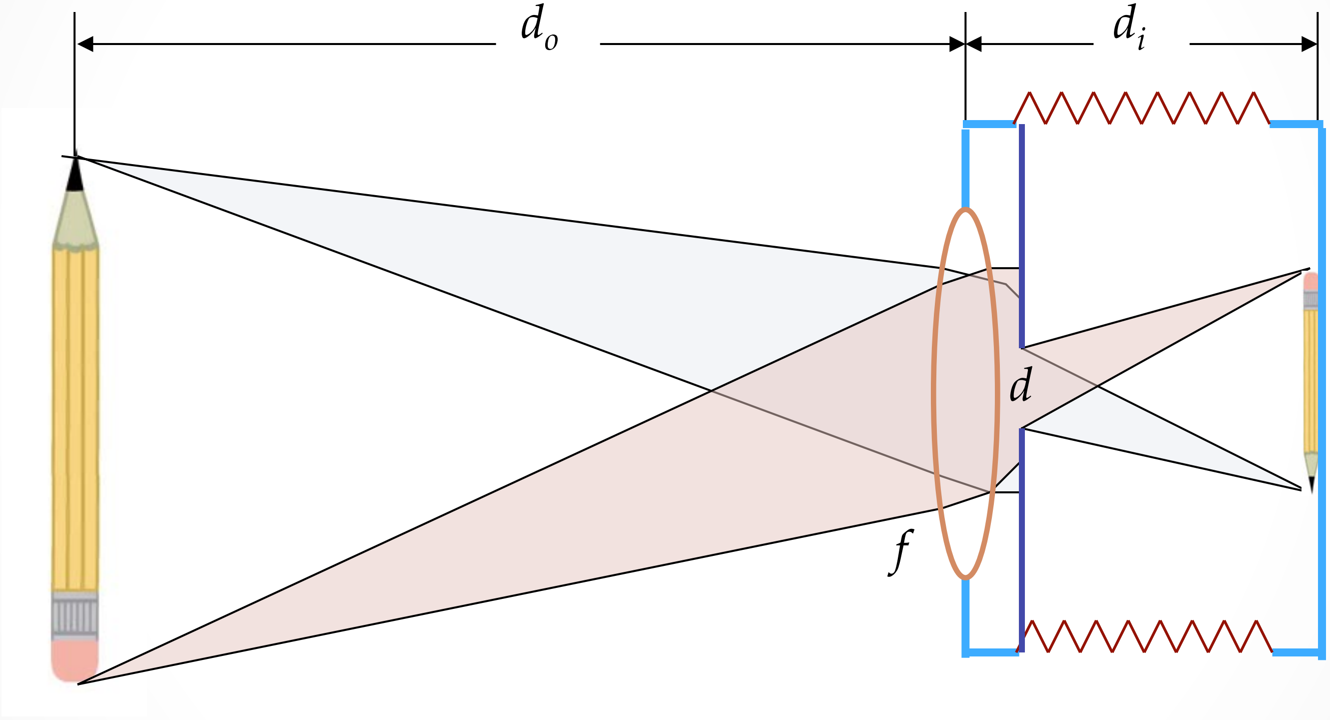


Thin lens equation $1/f = 1/d_o + 1/d_i$

$d_i = f \,.\, d_o \,/(d_o - f)$

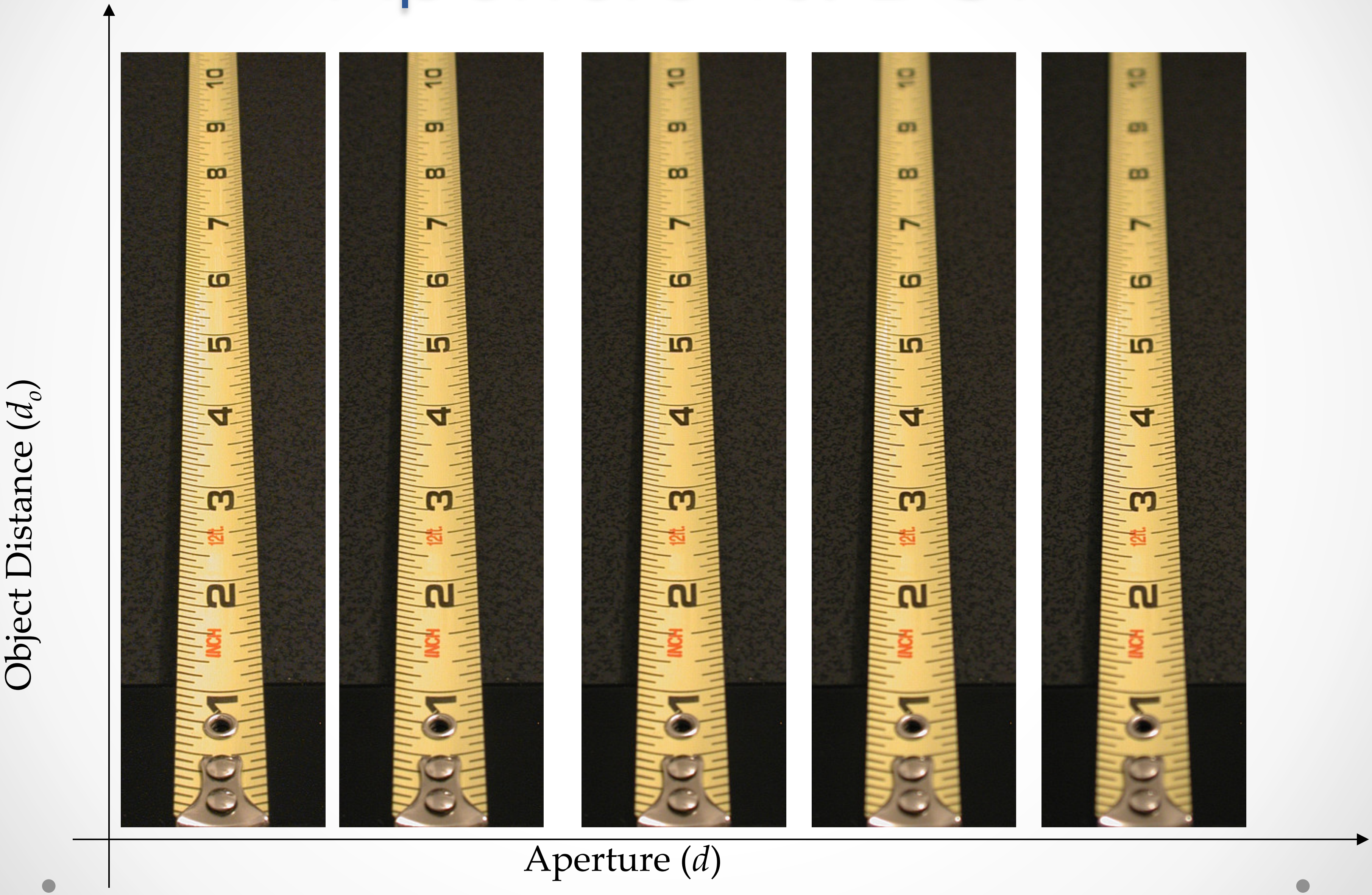
Focus and DOF

Aperture

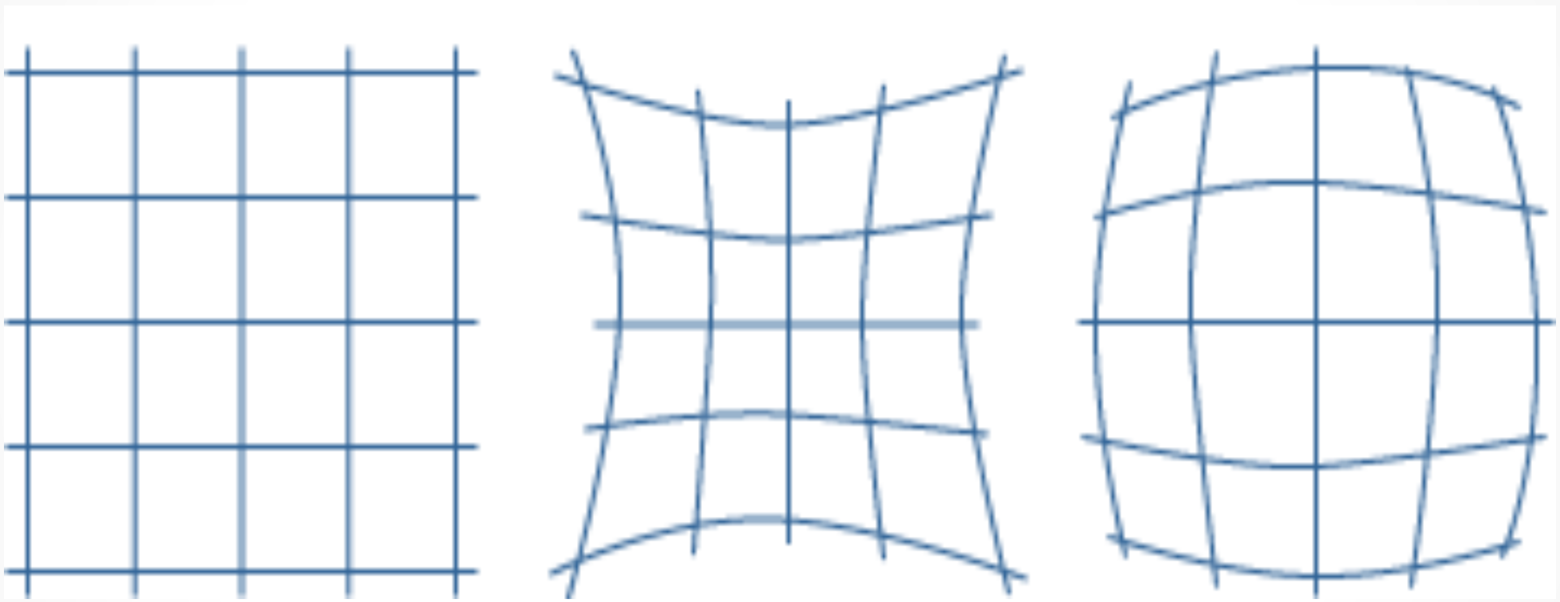


Focal Ratio = f / d

Aperture vs. DOF



Geometric Distortions



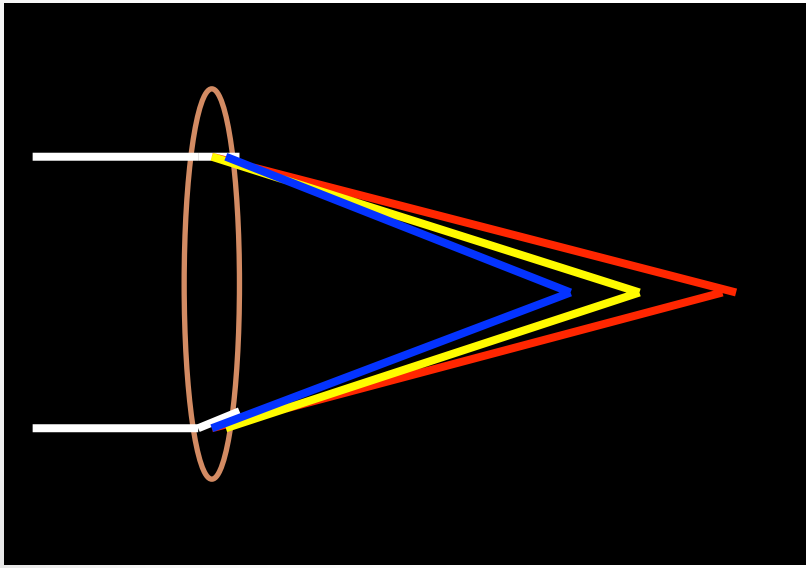
original pincushion barrel

Geometric Distortions

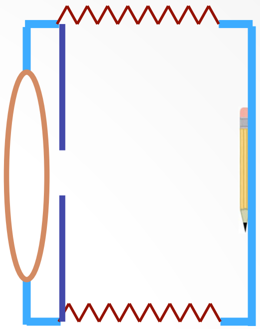
Lens Flare

Chromatic Aberration

Normal lenses diffract different wavelengths to different degree

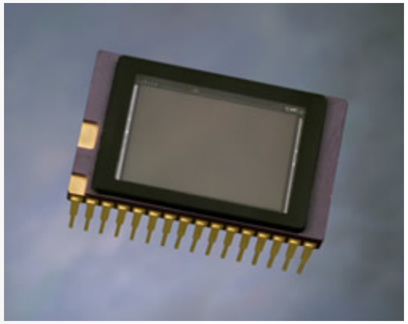


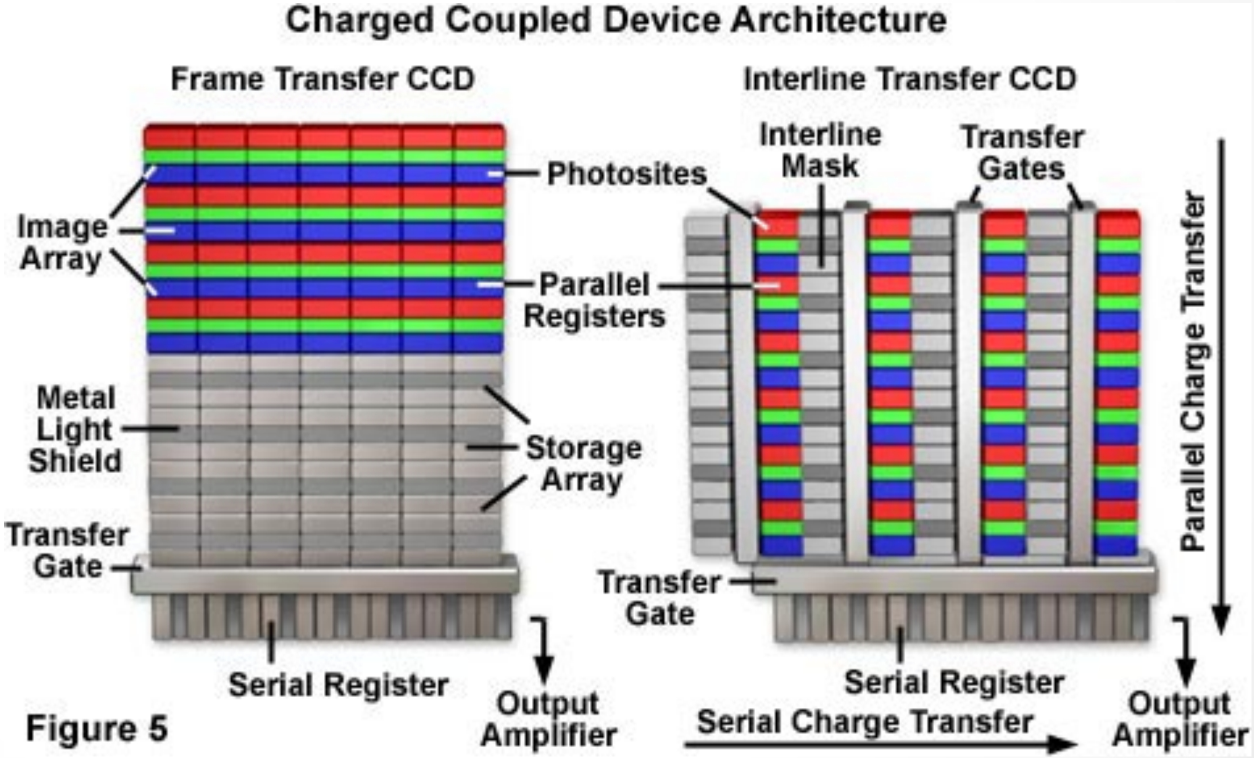
Sensor Element



- Film
  - Digital image created by scanning the film
- Vacuum Tube
  - An electron beam is used to scan the image formed by the lens on a photoconductive layer
- CCD and CMOS
  - An array of photosensitive elements is used at the image plane

CCD and CMOS cameras





http://micro.magnet.fsu.edu/primer/

Sampling

- Grayscale Image

  A grayscale image is a two-dimensional function ***f(x,y)*** of real values on the image plane.

- Sampling

  The process of converting a continuous image $f(x,y)$ to a discrete representation $f_s[m,n]$ by recording the function values at points of intersection of a 2D grid. Each sampling point is called a pixel (picture element). The process is equivalent to multiplying the image function with an ideal 2D impulse train $\delta(x,y)$. X_0 and Y_0 are called sampling distances or intervals.

$$
\begin{aligned}
f_s[m,n] &= f(x,y).\sum_{m=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty}\delta(x-mX_0,y-nY_0) \\
&= \sum_{m=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty} f(mX_0,nY_0)\delta(x-mX_0,y-nY_0)
\end{aligned}
$$

Sampling

Image function

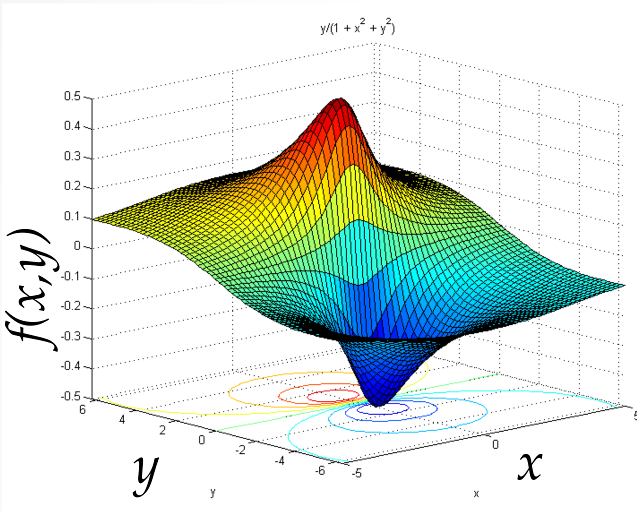
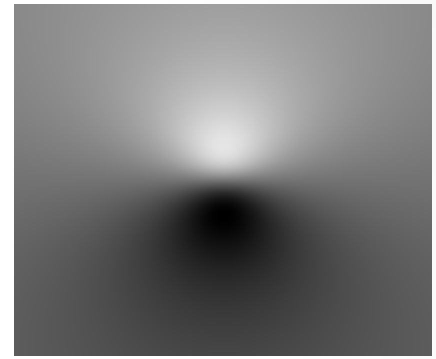


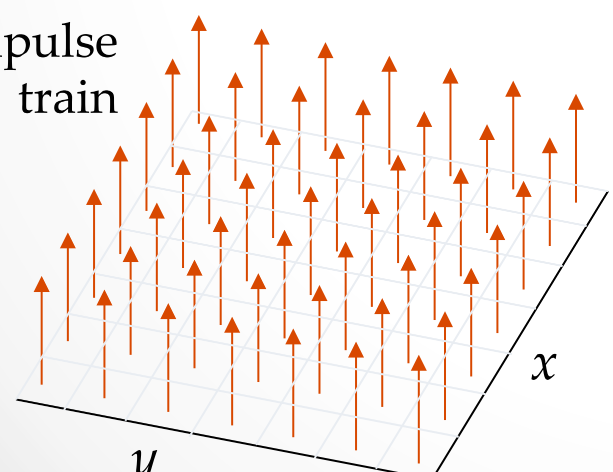
Image formed by the lens on the sensor

y

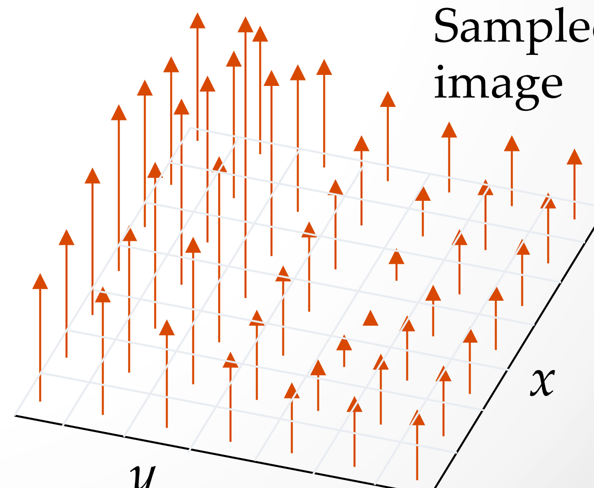


x

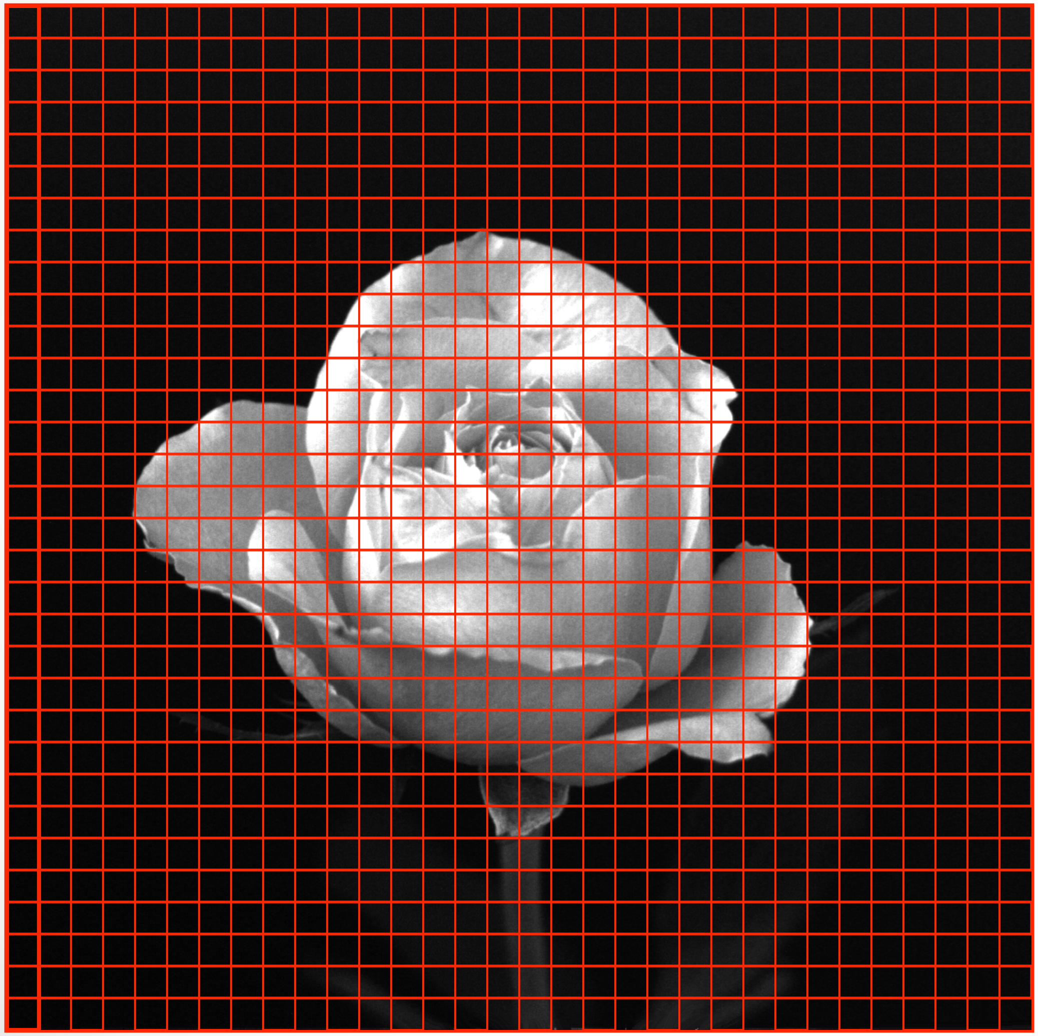
2D impulse train



Sampled image



Sampling an image



Resolution

- The number of samples in an image (number of sensor elements) is referred to as its resolution
- The resolution is typically represented as the product of number of samples in the horizontal and vertical directions in the image. e.g.: 32x32, 256x256, 640x480

Common Resolutions:

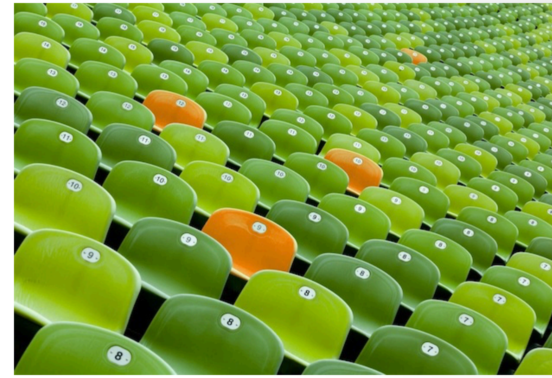
NTSC:	648 x 486
Typical Webcam:	1280 x 720
High-end Digital Camera:	8,650 x 5,800
Hubbles Telescope:	1,600 x 1,600

Camera Module: Objectives

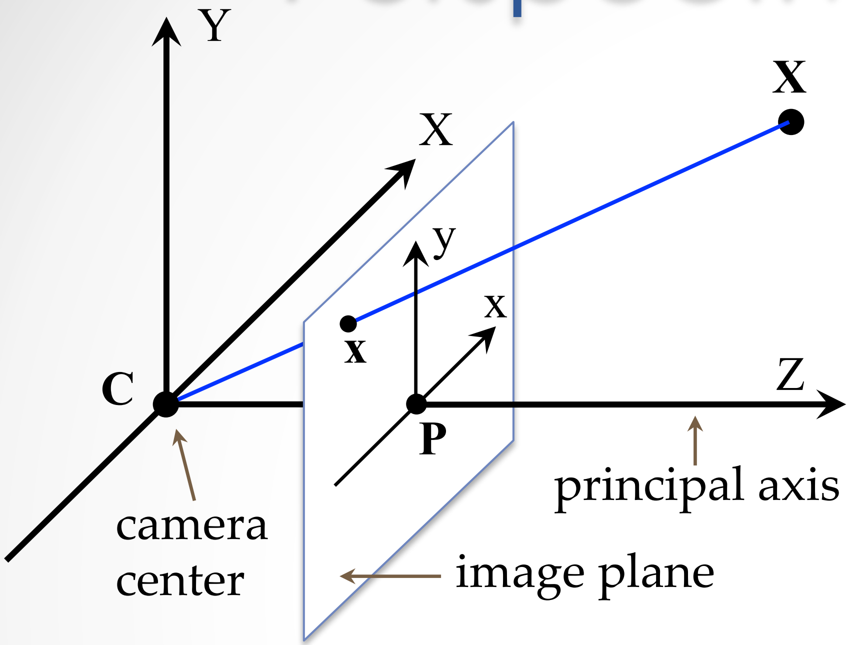
- Mathematically model what a camera does
  - Also understand what the model means
- Getting the model for a real-world camera
  - Estimation from real world measurements
- Special imaging configurations with simpler properties
  - Simpler relationships
- General theory on fitting linear models under noisy observations
  - Techniques that work across problems

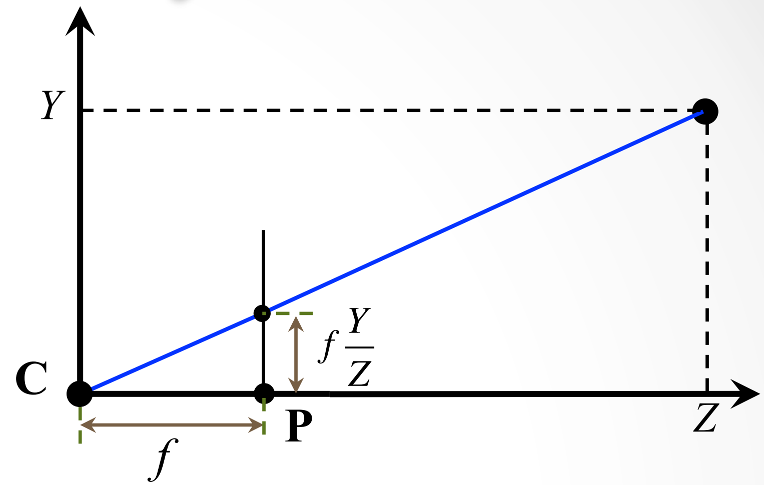
What does a Camera do?

- Form an image on the 2D image plane of the 3D world visible to it.
- Image is *behind* the lens; the scene is in front.
- 3D world is **projected** down to a 2D plane.
- Significant loss of information as one dimension is dropped.
- Mathematical depiction of this projection ...



Perspective Projection





- Cartesian image coordinates: $x = f\dfrac{X}{Z}, \quad y = f\dfrac{Y}{Z}$
- In matrix form (homogeneous):

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{PX}$$

Basic Camera Equation

A pinhole camera projects a 3D point $\mathbf{X_c}$ in camera coords to an image point $\mathbf{x}$ via the 3x4 camera matrix $\mathbf{P}$ as:

$$\mathbf{x} = \mathbf{PX_c} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{I} \mid \mathbf{0}]\mathbf{X_c} = \mathbf{K}[\mathbf{I} \mid \mathbf{0}]\mathbf{X_c},$$

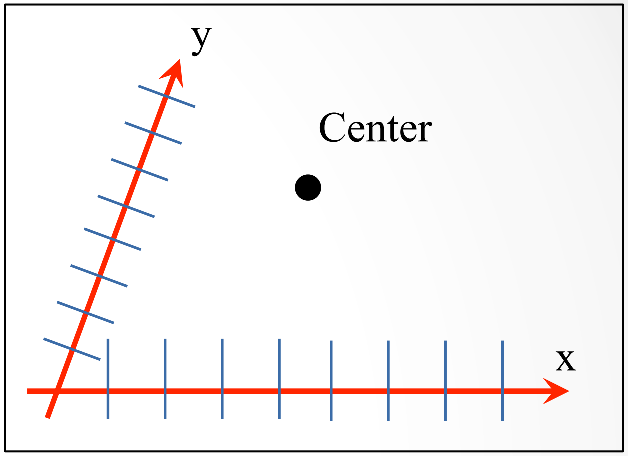
where $\mathbf{K}$ is the internal camera calibration matrix.

Note that:

- The camera is at origin
- Z is the Camera or Optical axis
- Principal Point: Center of the image
- Focal length in pixel units
- Orthogonal image axes with uniform scale

A General Camera

Image center at (x_0, y_0), Non-orthogonal axes with skew s, and different scales for axes with focal lengths, α_x and α_y.

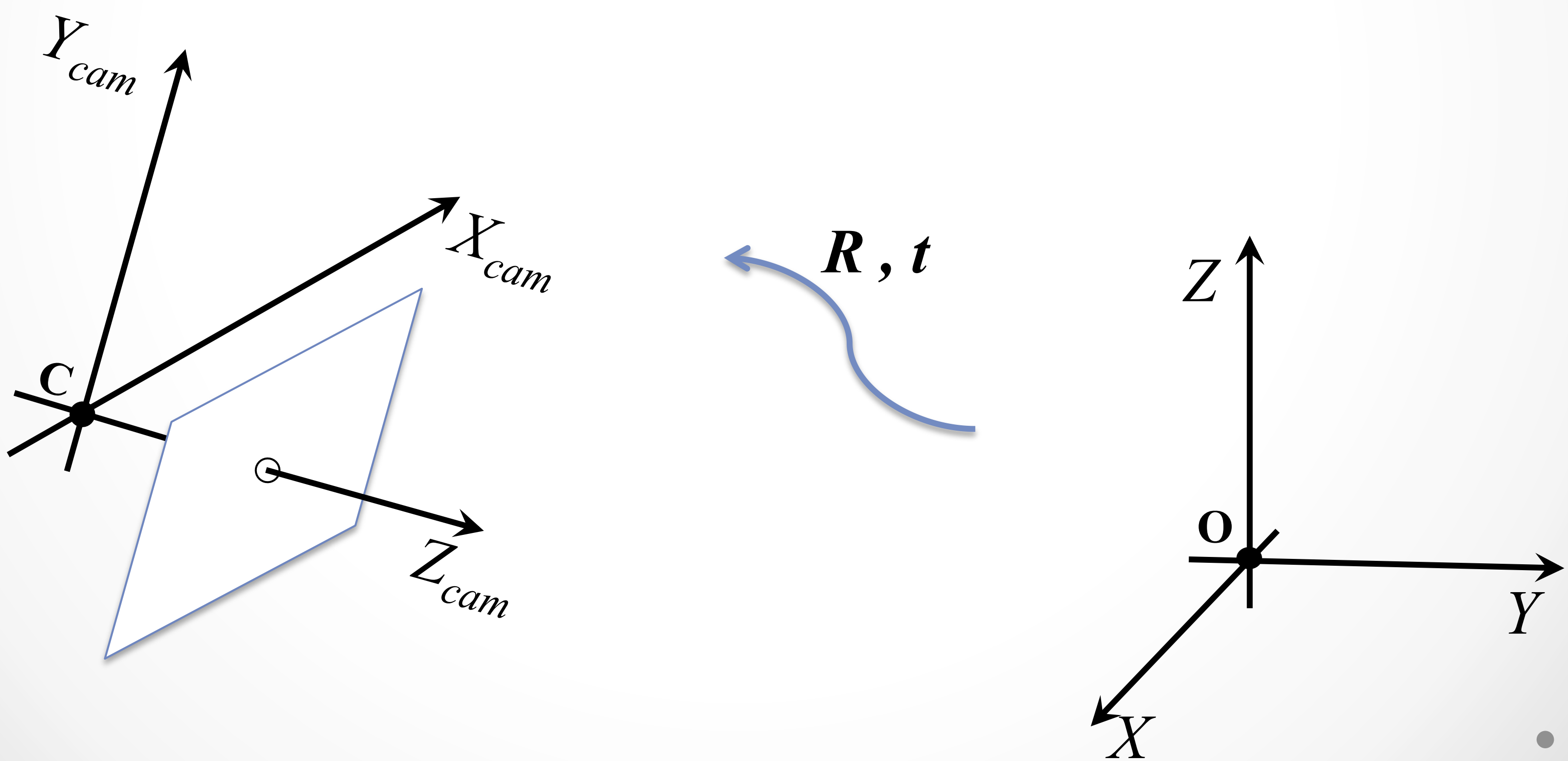


$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

K an upper diagonal matrix with 5 degrees of freedom.

Moving the Camera from Origin

- General Setting: Camera is not at origin and Z is not the optical axis.
- Camera is at a point C in world coordinates. The camera axes are also rotated by a matrix R.



General Camera Equation

- Camera and world are related by:

$$\mathbf{X_c} = \begin{bmatrix} \mathbf{R} & \mathbf{-RC} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \mathbf{X_w}$$

- 2D projection $\mathbf{x}$ of a 3D point $\mathbf{X_w}$ given by:
  - $\mathbf{x = K[\ I\ |\ 0\ ]\ X_c = K\ [\ R\ |\ \text{-}RC\ ]\ X_w}$
- $\mathbf{x = PX_w}$; camera matrix $\mathbf{P = [KR\ |\ \text{-}KRC] = [M\ |\ p_4]}$
- Common K: General K:

$$\begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

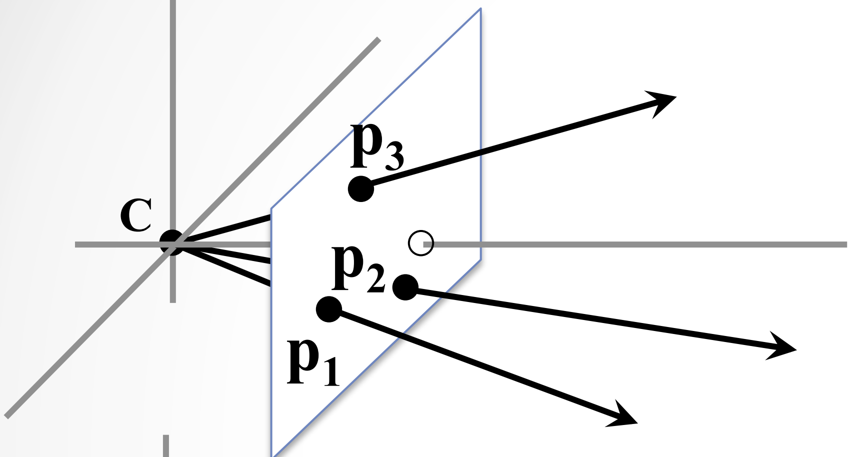
General Camera Equation

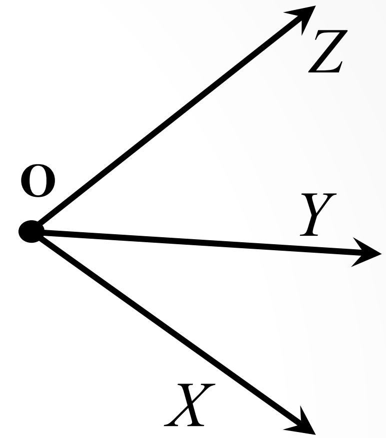
- General projection equation in world coordinates:

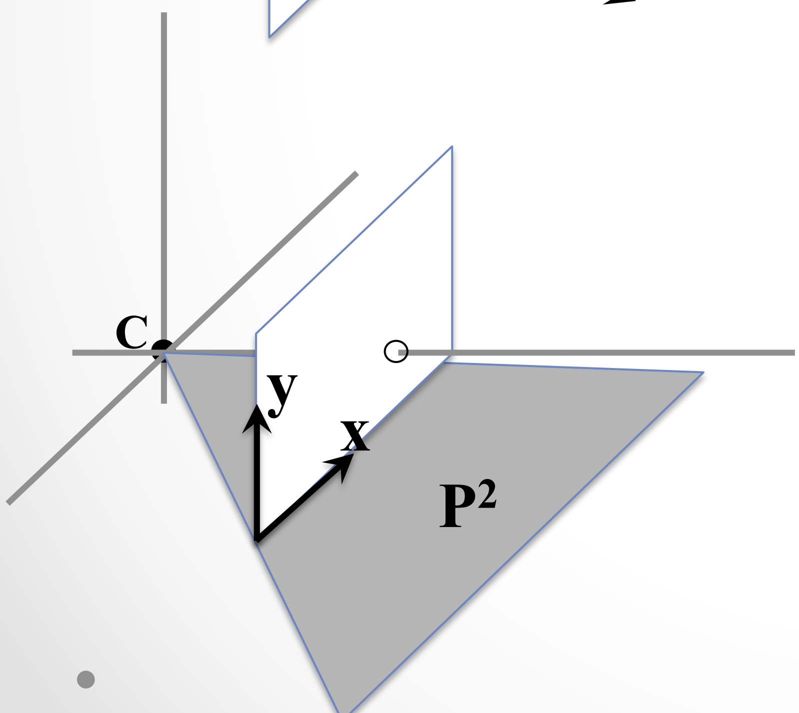
$$\mathbf{x} = \mathbf{K}\,[\mathbf{R} \mid -\mathbf{RC}]\,\mathbf{X}_w = [\mathbf{KR} \mid -\mathbf{KRC}]\,\mathbf{X}_w = [\mathbf{M} \mid \mathbf{p}_4]\,\mathbf{X}_w$$

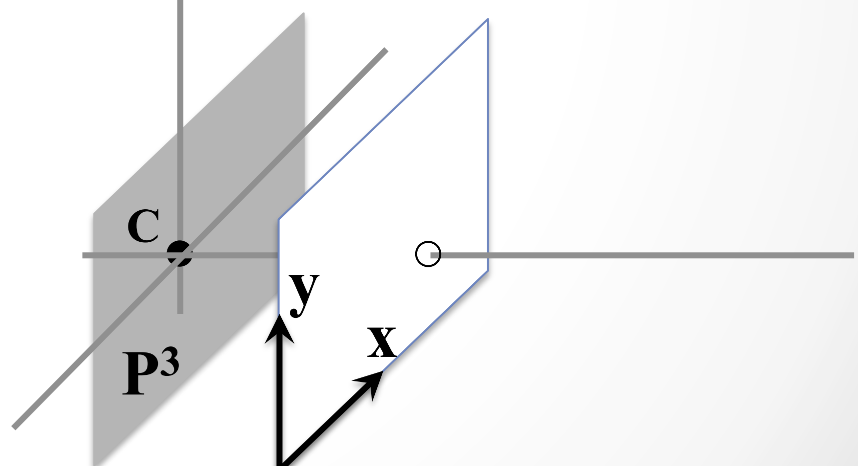
- 3x4 matrix **P** maps/projects World-C to Image-C
  - Left 3 × 3 submatrix is non-singluar for finite cameras
  - Orthographic projection: left submatrix is singular
- Any 3 × 4 matrix P with non-singular left submatrix represents a camera! It can be decomposed as:
  - A non-singular upper diagonal matrix K
  - An orthonormal matrix R and a vector C
    with the usual meanings!!

Axis Points and Planes









Camera Matrix Anatomy

$$\mathbf{P} = [\,\mathbf{p}_1\ \ \mathbf{p}_2\ \ \mathbf{p}_3\ \ \mathbf{p}_4\,] = [\,\mathbf{P}^1\ \ \mathbf{P}^2\ \ \mathbf{P}^3\,]^\mathrm{T}$$

- 4-vector $\mathbf{C}$ with $\mathbf{PC} = \mathbf{0}$ is the camera center.
  - Camera center is the only point with no projection or projects to the vector **0**, which is undefined in **P**^{2}.
- Columns $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$ are the images of vanishing points of the world X, Y and Z directions.
  - $\mathbf{p}_1 = \mathbf{P}\,[1\ 0\ 0\ 0]^\mathrm{T}$, the image of the ideal point in X direction and similarly the rest.
- $\mathbf{p}_4$ is the image of world origin: $\mathbf{p}_4 = \mathbf{P}\,[0\ 0\ 0\ 1]^\mathrm{T}$.

Prove the Following

- Row vector $\mathbf{P}^3$ is the principal plane.
- Row vectors $\mathbf{P}^1$ and $\mathbf{P}^2$ are axis planes for image Y and X axes.
- The principal point (or image center) is given by $x_0 = M\, m_3$, with m_3 the third row vector of matrix M.
- $\mathbf{det(M)m_3}$ gives the principal axis as a vector from the camera centre through the prinicipal point to the front of the camera.