Unit 1: Image Formation

1. Geometry
2. Optics
3. Photometry
4. Sensor

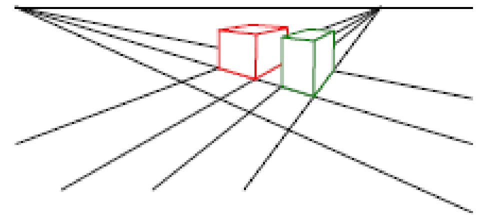
Readings

- Szeliski 2.1-2.3 & 6.3.5

Physical parameters of image formation

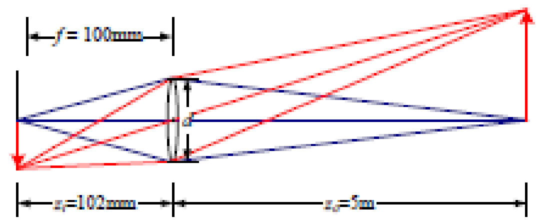
Geometric

- Type of projection
- Camera pose



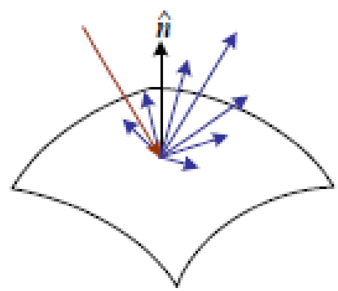
Optical

- Sensor's lens type
- focal length, field of view, aperture



Photometric

- Type, direction, intensity of light reaching sensor
- Surfaces' reflectance properties



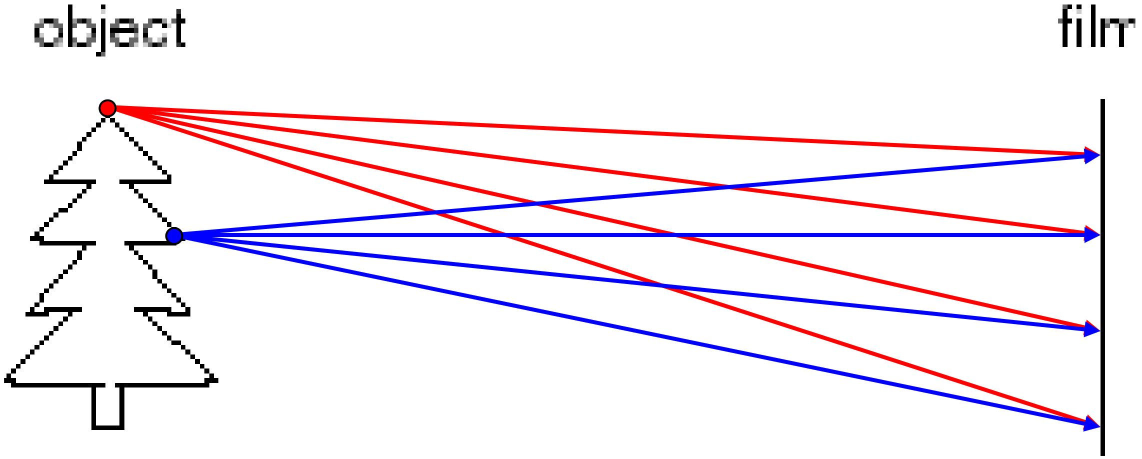
Sensor

- sampling, etc.

G	R	G	R
B	G	B	G
G	R	G	R
B	G	B	G

Bayer color filter array

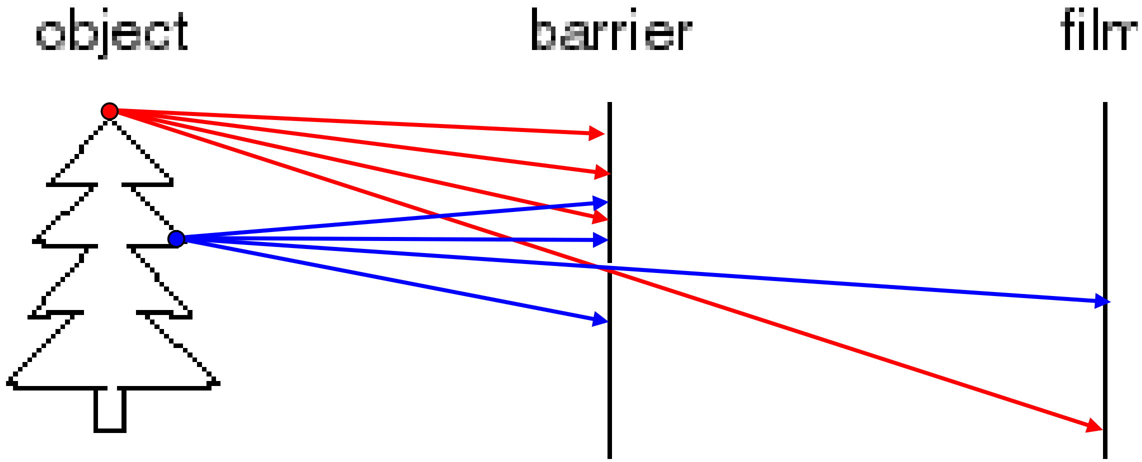
Image formation



Let's design a camera

- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?

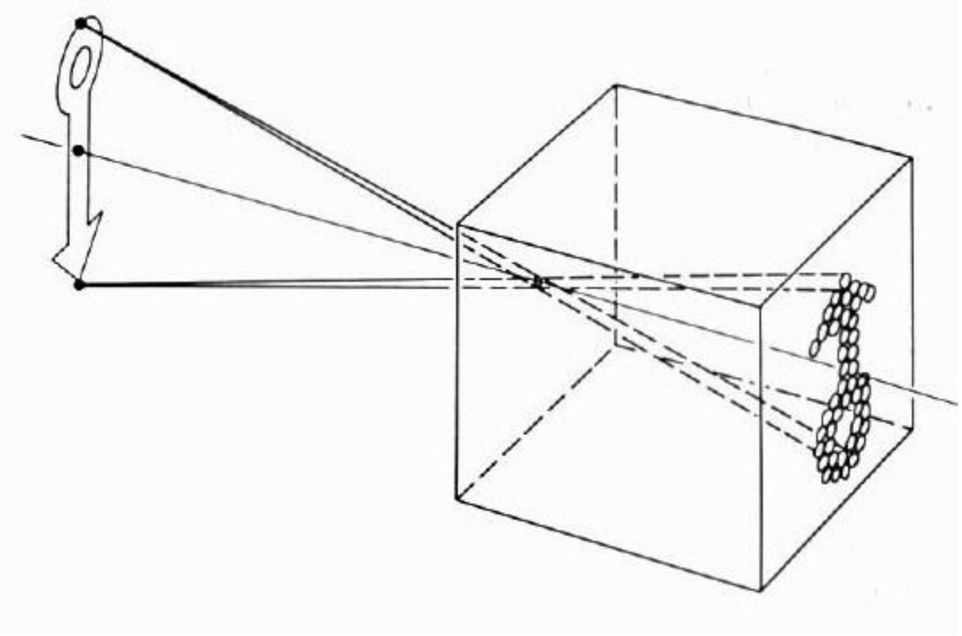
Pinhole camera



Add a barrier to block off most of the rays

- This reduces blurring
- The opening known as the **aperture**
- How does this transform the image?

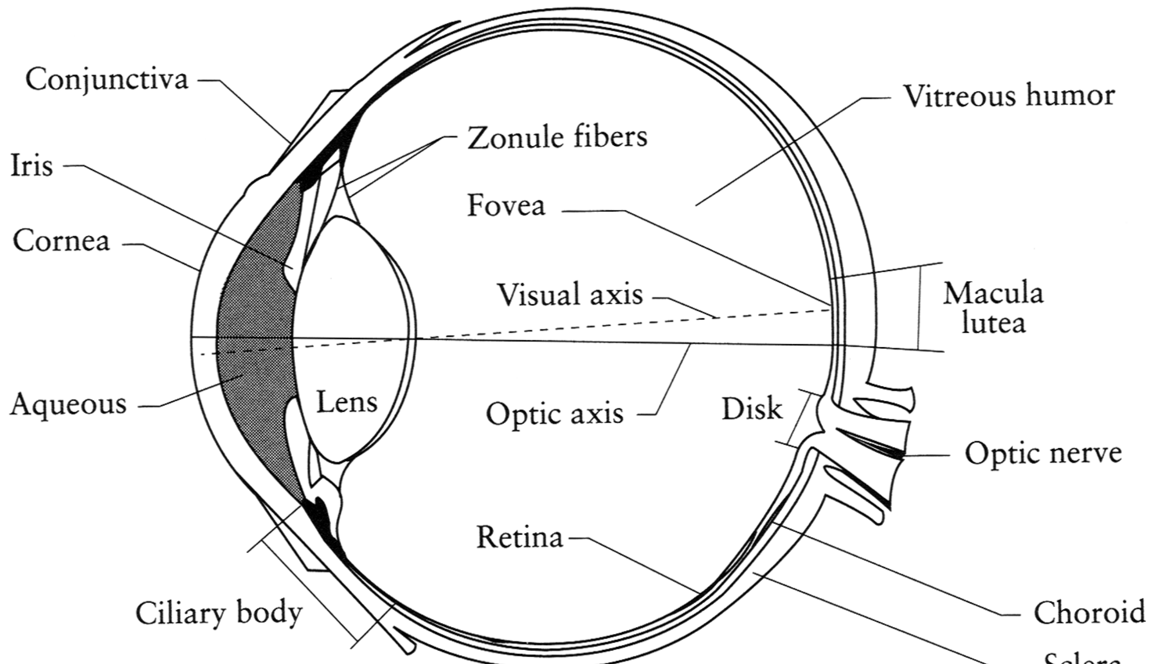
Camera Obscura



The first camera

- Known to Aristotle
- How does the aperture size affect the image?

The eye



The human eye is a camera

- **Iris** - colored annulus with radial muscles
- **Pupil** - the hole (aperture) whose size is controlled by the iris
- What's the "film"?
  - photoreceptor cells (rods and cones) in the **retina**

Digital camera

A digital camera replaces film with a sensor array

- Each cell in the array is a **C**harge **C**oupled **D**evice
  - light-sensitive diode that converts photons to electrons
  - other variants exist: CMOS is becoming more popular
  - http://electronics.howstuffworks.com/digital-camera.htm

Issues with digital cameras

Noise
- big difference between consumer vs. SLR-style cameras
- low light is where you most notice noise

Compression
- creates artifacts except in uncompressed formats (tiff, raw)

Color
- color fringing artifacts from Bayer patterns

Blooming
- charge overflowing into neighboring pixels

In-camera processing
- oversharpening can produce halos

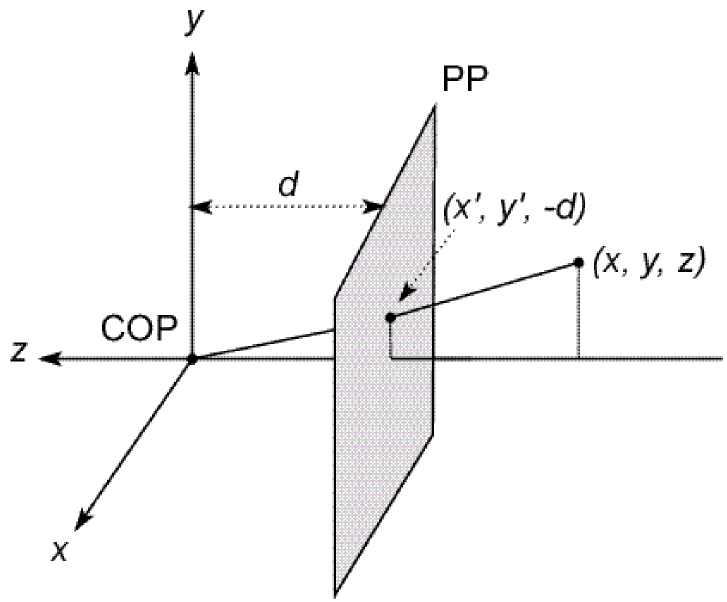
Interlaced vs. progressive scan video
- even/odd rows from different exposures

Are more megapixels better?
- requires higher quality lens
- noise issues

Stabilization
- compensate for camera shake (mechanical vs. electronic)

More info online, e.g.,
- http://electronics.howstuffworks.com/digital-camera.htm
- http://www.dpreview.com/

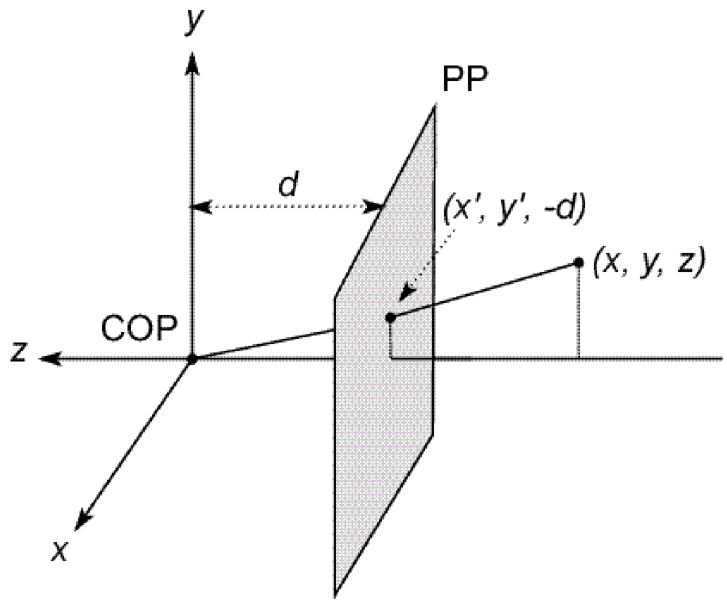
Geometric projection



The coordinate system

- We will use the pin-hole model as an approximation
- Put the optical center (**C**enter **O**f **P**rojection) at the origin
- Put the image plane (**P**rojection **P**lane) *in front* of the COP
  - Why?
- The camera looks down the *negative* z axis
  - we need this if we want right-handed-coordinates

Modeling projection



Projection equations

- Compute intersection with PP of ray from (x,y,z) to COP
- Derived using similar triangles (on board)

$$(x, y, z) \rightarrow (-d\frac{x}{z},\ -d\frac{y}{z},\ -d)$$

- We get the projection by throwing out the last coordinate:

$$(x, y, z) \rightarrow (-d\frac{x}{z},\ -d\frac{y}{z})$$

Homogeneous coordinates

Is this a linear transformation?

- no—division by z is nonlinear

Trick: add one more coordinate:

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous image coordinates

$$(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

homogeneous scene coordinates

Converting *from* homogeneous coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

Perspective Projection

Projection is a matrix multiply using homogeneous coordinates:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ -z/d \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \quad -d\frac{y}{z})$$

divide by third coordinate

This is known as **perspective projection**

- The matrix is the **projection matrix**
- Can also formulate as a 4x4 (today's reading does this)

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ -z/d \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \quad -d\frac{y}{z})$$

divide by fourth coordinate

Perspective Projection

How does scaling the projection matrix change the transformation?

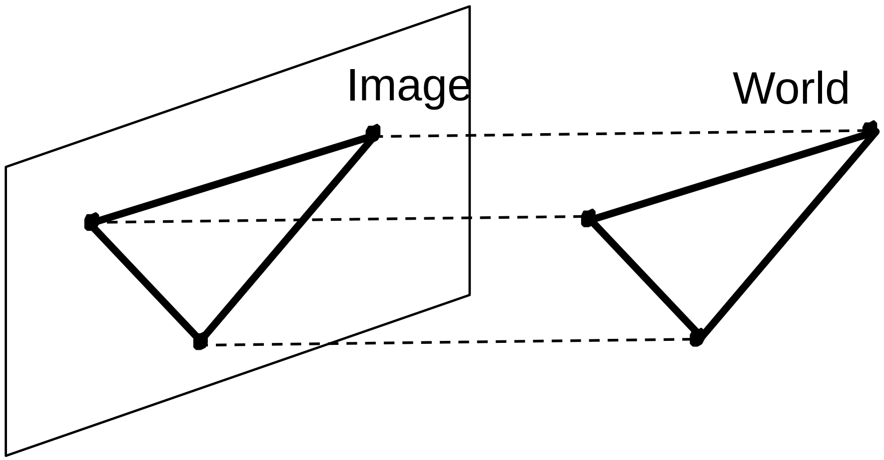
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ -z/d \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \ -d\frac{y}{z})$$

$$\begin{bmatrix} -d & 0 & 0 & 0 \\ 0 & -d & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} -dx \\ -dy \\ z \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \ -d\frac{y}{z})$$

Orthographic projection

Special case of perspective projection

- Distance from the COP to the PP is infinite

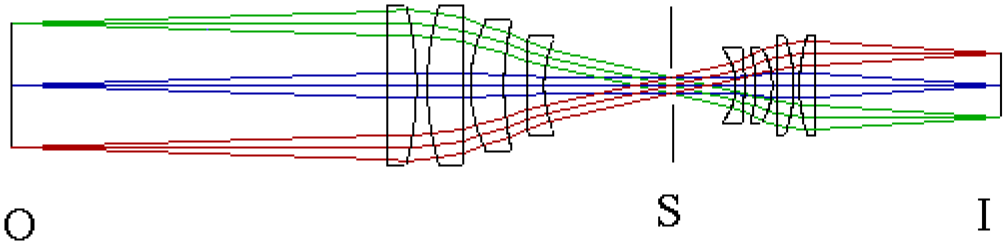


- Good approximation for telephoto optics
- Also called "parallel projection": (x, y, z) → (x, y)
- What's the projection matrix?

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \Rightarrow (x, y)$$

Orthographic ("telecentric") lenses

Navitar telecentric zoom lens



http://www.lhup.edu/~dsimanek/3d/telecent.htm

Variants of orthographic projection

Scaled orthographic

- Also called "weak perspective"

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/d \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1/d \end{bmatrix} \Rightarrow (dx, dy)$$

Affine projection

- Also called "paraperspective"

$$\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

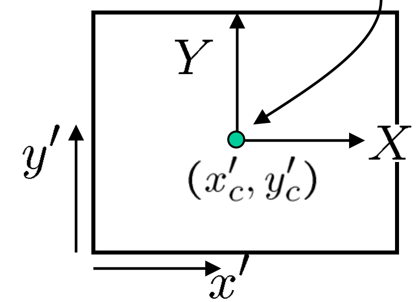
Camera parameters

A camera is described by several parameters

- Translation T of the optical center from the origin of world coords
- Rotation R of the image plane
- focal length f, principle point (c_x, c_y), pixel size (s_x, s_y)
- blue parameters are called "extrinsics," red are "intrinsics"

Projection equation

$$\mathbf{x} = \begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathbf{\Pi X}$$



- The projection matrix models the cumulative effect of all parameters
- Useful to decompose into a series of operations

identity matrix

$$\mathbf{\Pi} = \underbrace{\begin{bmatrix} fs_x & 0 & c_x \\ 0 & fs_y & c_y \\ 0 & 0 & 1 \end{bmatrix}}_{\text{intrinsics}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{projection}} \underbrace{\begin{bmatrix} \mathbf{R}_{3x3} & \mathbf{0}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix}}_{\text{rotation}} \underbrace{\begin{bmatrix} \mathbf{I}_{3x3} & \mathbf{T}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix}}_{\text{translation}}$$

- The definitions of these parameters are **not** completely standardized
  - especially intrinsics—varies from one book to another

3D-to-2D Camera Projection

Camera calibration matrix K is a 3X3 upper-triangular matrix

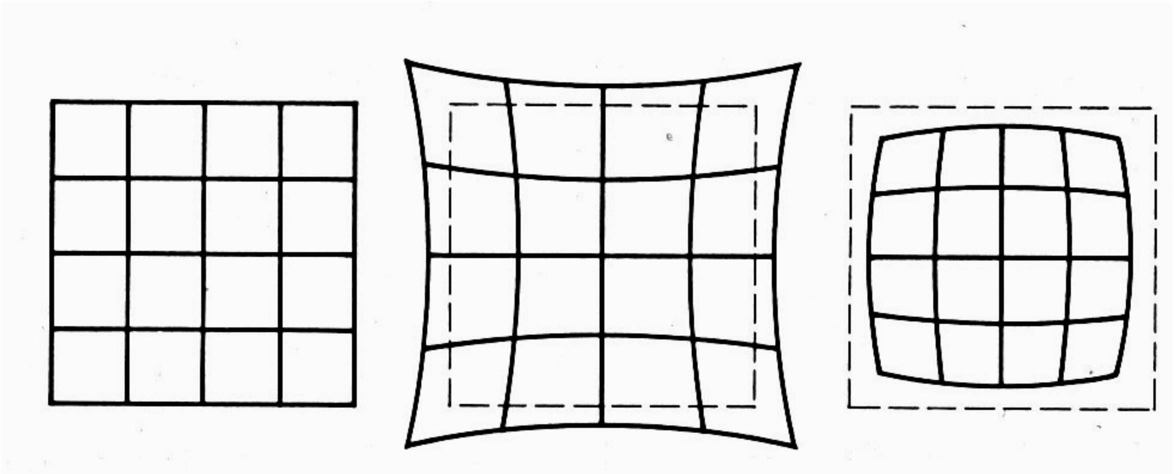
$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

The camera projection matrix P (3X4) which maps the 3D point coordinate (in world coordinate) to the corresponding 2D image coordinate is given by

$$\boldsymbol{P} = \boldsymbol{K}[\boldsymbol{R}|\boldsymbol{t}]$$

R & t: the camera extrinsic parameters

Lens Distortion



No distortion Pin cushion Barrel

Radial distortion of the image

- Caused by imperfect lenses
- Deviations are most noticeable for rays that pass through the edge of the lens
- The radial distortion model says that coordinates in the observed images are displaced away (barrel) or towards (pincushion) the image center by an amount proportional to their radial distance.

Correcting radial distortion

from [Helmut Dersch]

Radial Distortion Model

Let (x_c, y_c) be the pixel coordinates obtained after perspective division but before scaling by focal length *f* and shifting by the optical center (c_x, c_y), i.e.,

$$
\begin{aligned}
x_c &= \frac{\mathbf{r}_x \cdot \mathbf{p} + t_x}{\mathbf{r}_z \cdot \mathbf{p} + t_z} \\
y_c &= \frac{\mathbf{r}_y \cdot \mathbf{p} + t_y}{\mathbf{r}_z \cdot \mathbf{p} + t_z}
\end{aligned}
$$

The simplest radial distortion models use low-order polynomials, e.g.

$$
\begin{aligned}
\hat{x}_c &= x_c(1 + \kappa_1 r_c^2 + \kappa_2 r_c^4) \\
\hat{y}_c &= y_c(1 + \kappa_1 r_c^2 + \kappa_2 r_c^4)
\end{aligned}
\qquad r_c^2 = x_c^2 + y_c^2
$$

The final pixel coordinates can be computed using

$$
\begin{aligned}
x_s &= f x'_c + c_x \\
y_s &= f y'_c + c_y.
\end{aligned}
$$

Modeling distortion

Project $(\widehat{x}, \widehat{y}, \widehat{z})$ to "normalized" image coordinates

$$x'_n = \widehat{x}/\widehat{z}$$
$$y'_n = \widehat{y}/\widehat{z}$$

Apply radial distortion

$$r^2 = {x'_n}^2 + {y'_n}^2$$
$$x'_d = x'_n(1 + \kappa_1 r^2 + \kappa_2 r^4)$$
$$y'_d = y'_n(1 + \kappa_1 r^2 + \kappa_2 r^4)$$

Apply focal length translate image center

$$x' = f x'_d + x_c$$
$$y' = f y'_d + y_c$$

To model lens distortion

- Use above projection operation instead of standard projection matrix multiplication

360 degree field of view…

Basic approach

- Take a photo of a parabolic mirror with an orthographic lens (Nayar)
- Or buy a lens from a variety of omnicam manufacturers…
  - See http://www.cis.upenn.edu/~kostas/omni.html

Pinhole size / aperture

How does the size of the aperture affect the image we'd get?

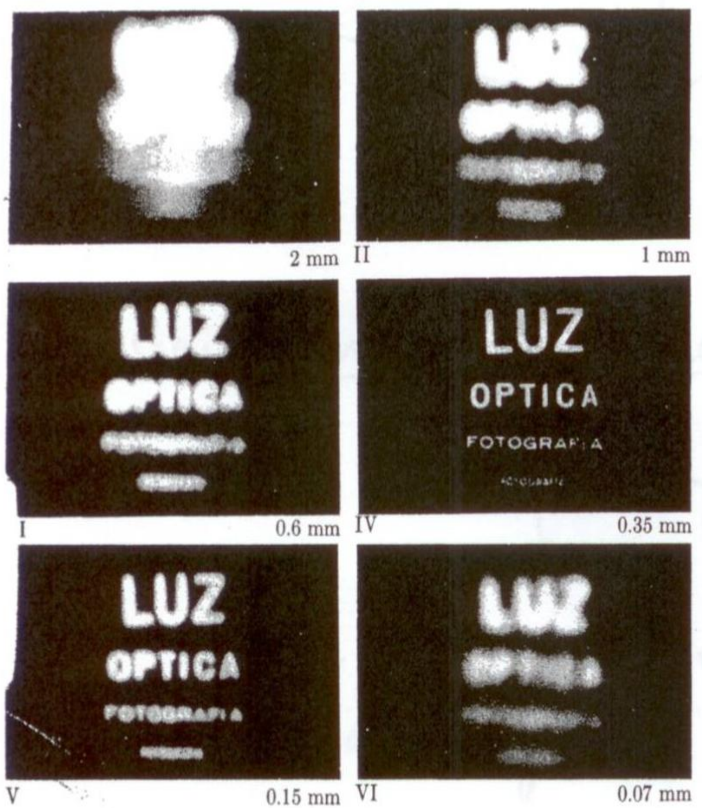
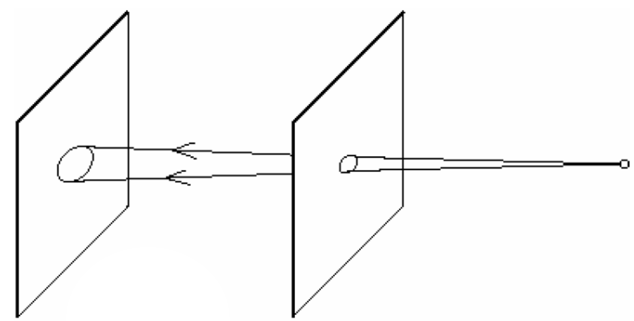


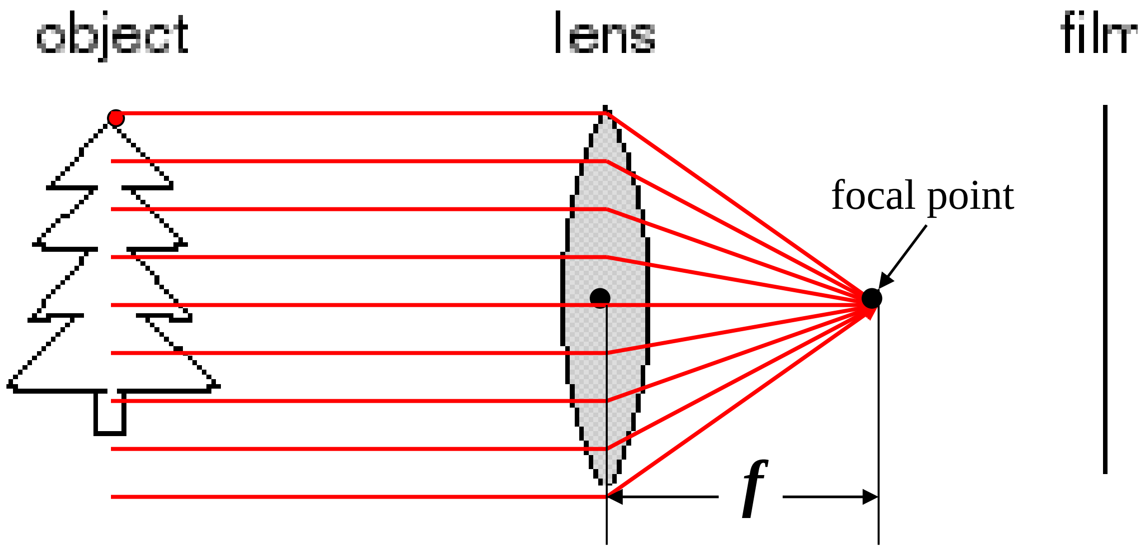
Fig. 5.96 The pinhole camera. Note the variation in image clarity as the hole diameter decreases. [Photos courtesy Dr. N. Joel, UNESCO.]

Larger

Smaller



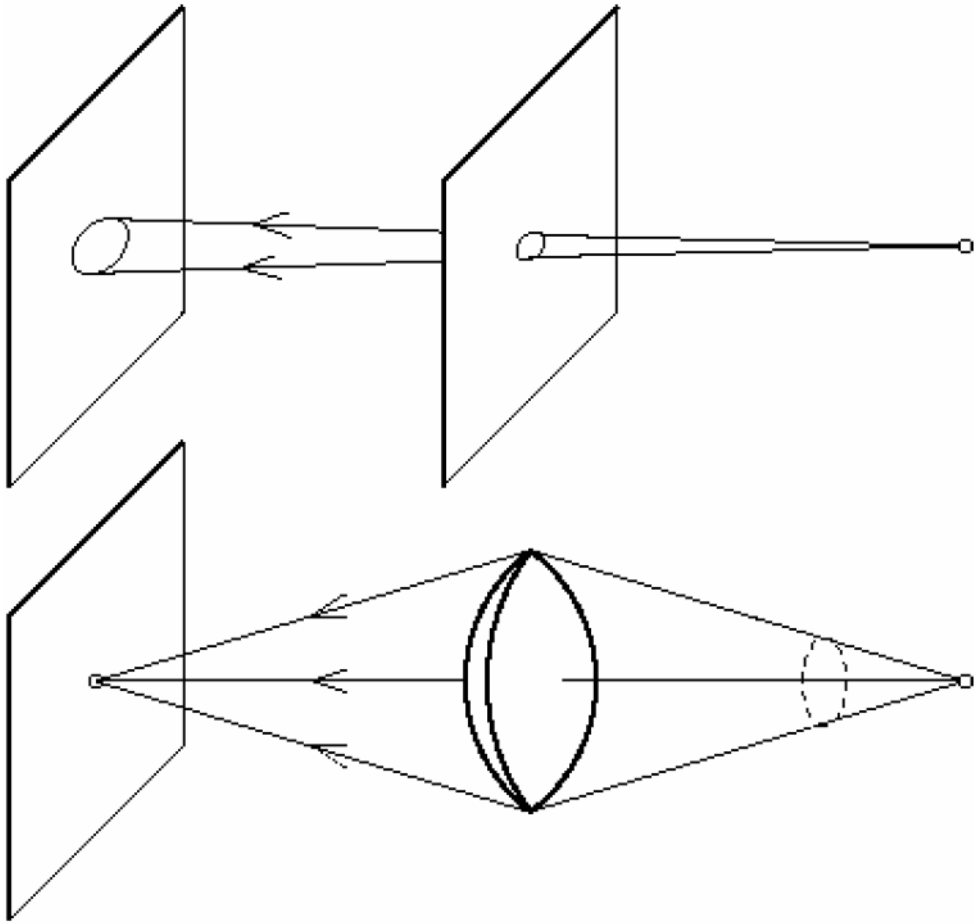
Adding a lens



- A lens focuses light onto the film
  - Rays passing through the center are not deviated
  - All parallel rays converge to one point on a plane located at the *focal length f*

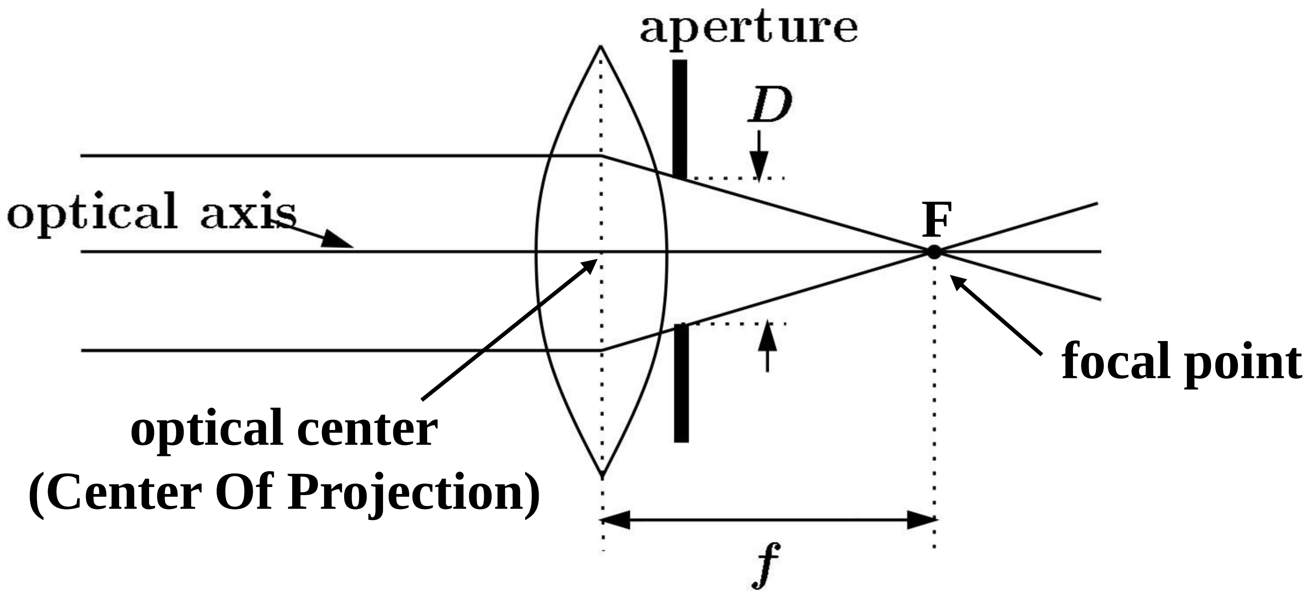
Slide by Steve Seitz

Pinhole vs. lens



K. Grauman

Cameras with lenses

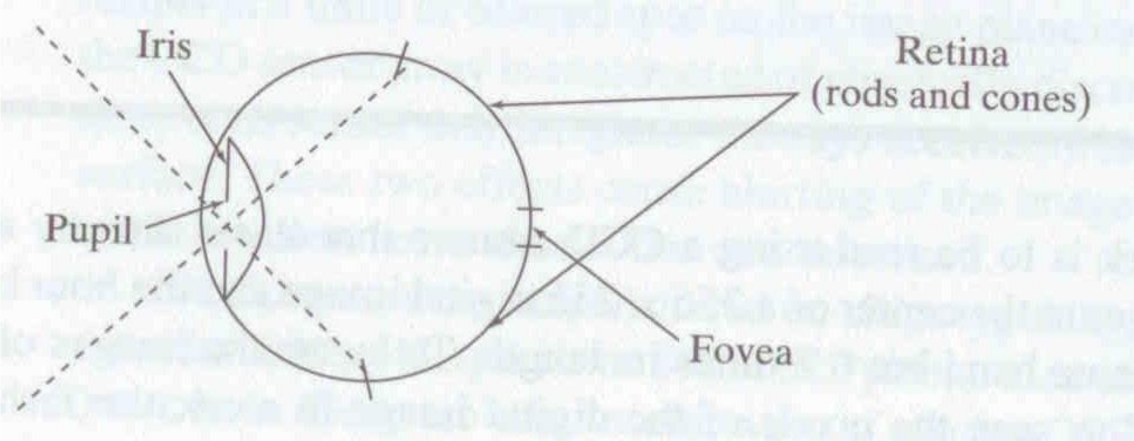


- A lens focuses parallel rays onto a single focal point
- Gather more light, while keeping focus; make pinhole perspective projection practical

K. Grauman

Human eye

Rough analogy with human visual system:



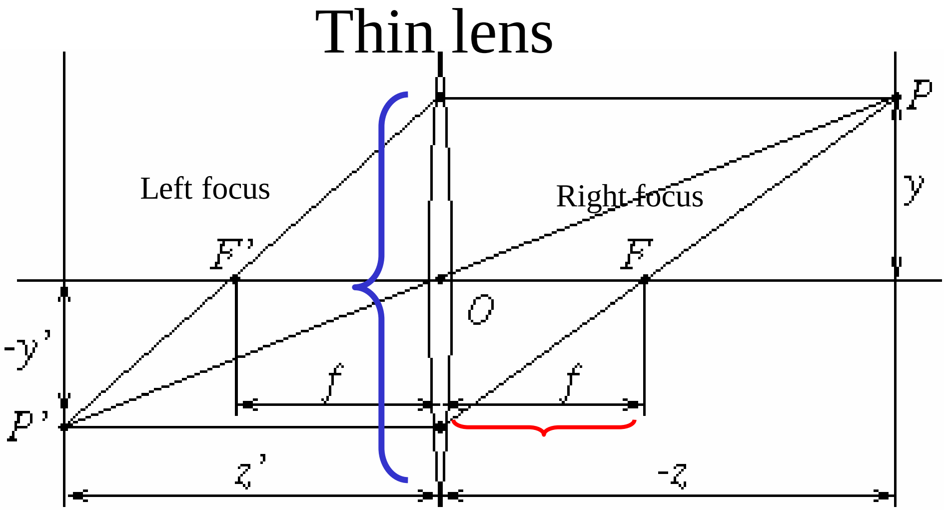
Pupil/Iris – control amount of light passing through lens

Retina - contains sensor cells, where image is formed

Fovea – highest concentration of cones

Fig from Shapiro and Stockman

Thin lens



Lens diameter *d*

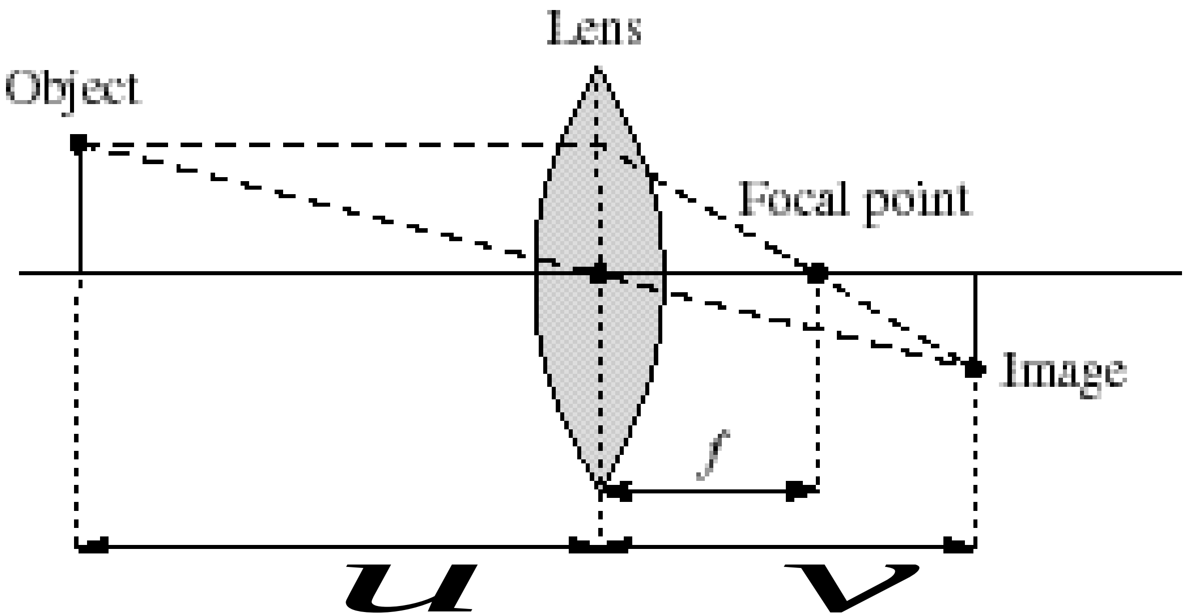
Focal length *f*

Rays entering parallel on one side go through focus on other, and vice versa.

In ideal case – all rays from P imaged at P'.

K. Grauman

Thin lens equation



$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$$

- Any object point satisfying this equation is in focus

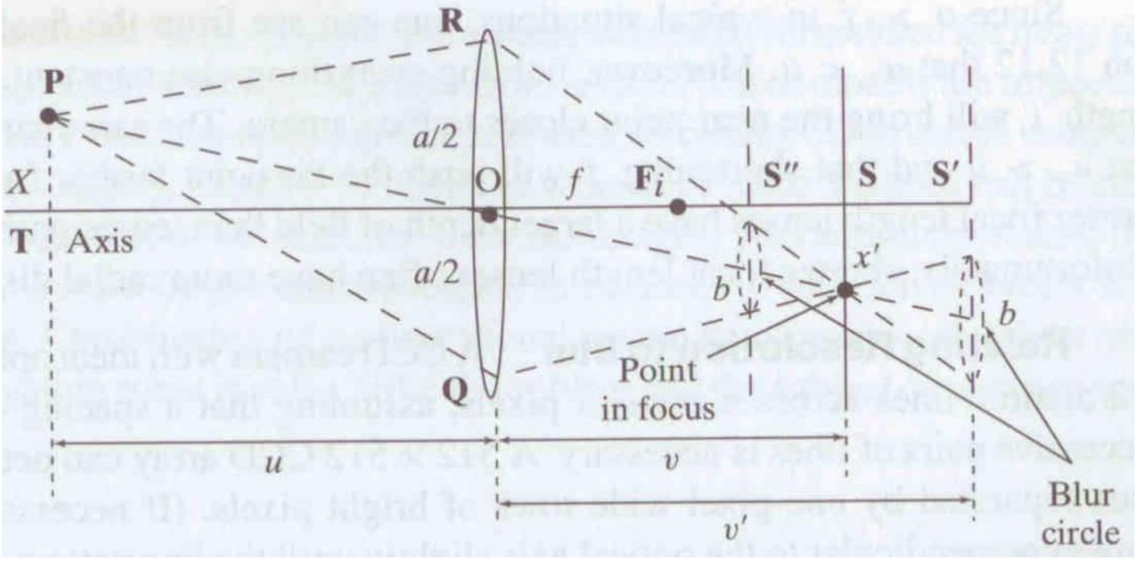
K. Grauman

Focus and depth of field

Image credit: cambridgeincolour.com

Focus and depth of field

Depth of field: distance between image planes where blur is tolerable



← "circles of confusion" →

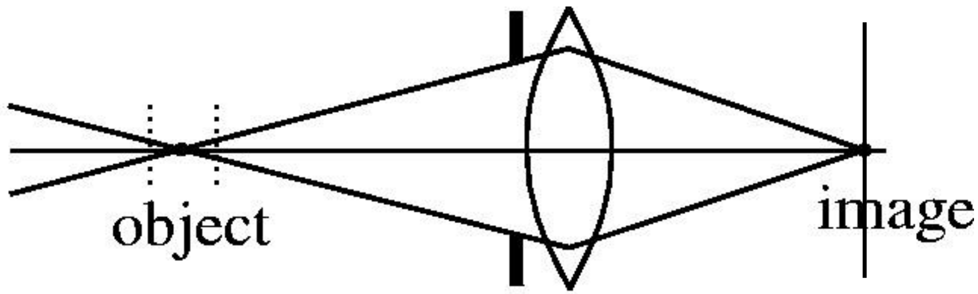
Thin lens: scene points at distinct depths come in focus at different image planes.

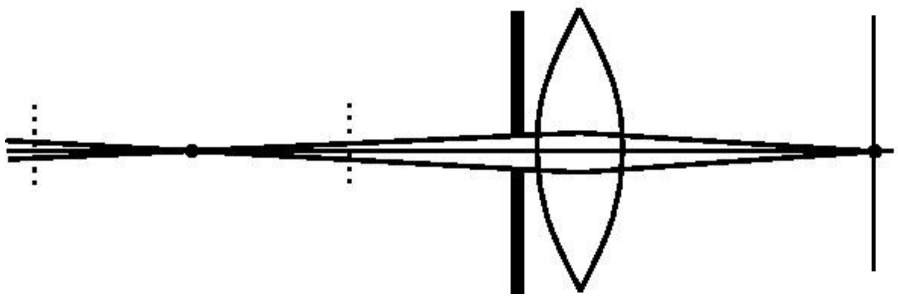
(Real camera lens systems have greater depth of field.)

Shapiro and Stockman

Focus and depth of field

How does the aperture affect the depth of field?



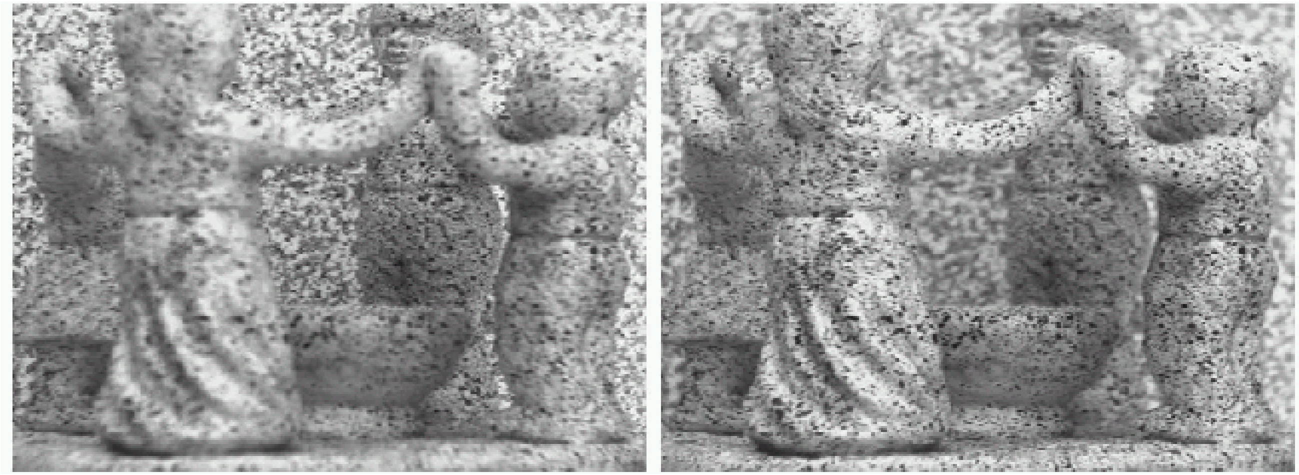


- A smaller aperture increases the range in which the object is approximately in focus

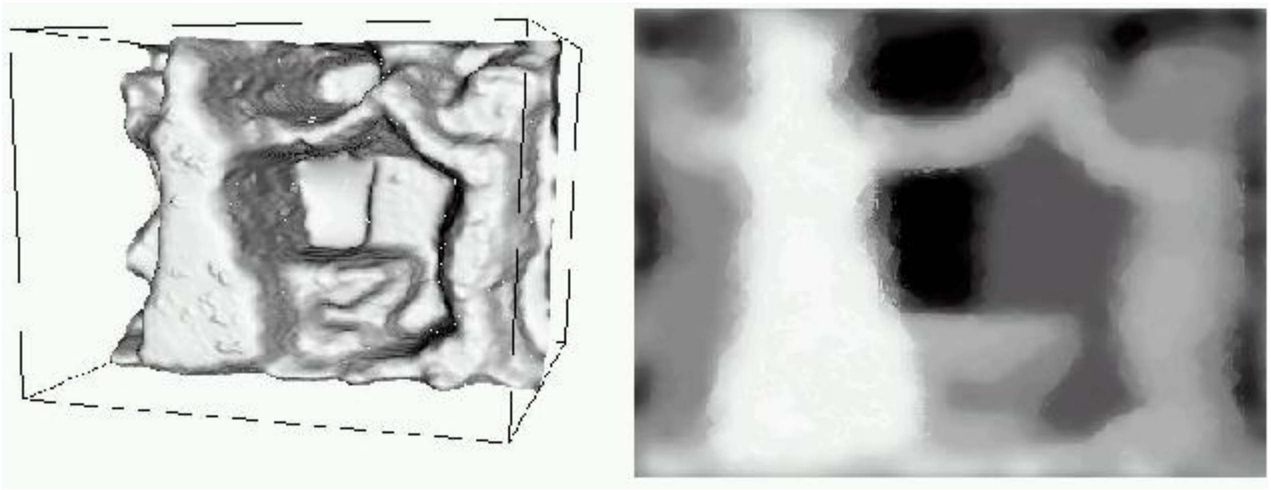
Flower images from Wikipedia http://en.wikipedia.org/wiki/Depth_of_field

Slide from S. Seitz

Depth from focus



Images from same point of view, different camera parameters



3d shape / depth estimates

[figs from H. Jin and P. Favaro, 2002]

Field of view

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

28 mm lens, 65.5° × 46.4°

50 mm lens, 39.6° × 27.0°

70 mm lens, 28.9° × 19.5°

210 mm lens, 9.8° × 6.5°

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

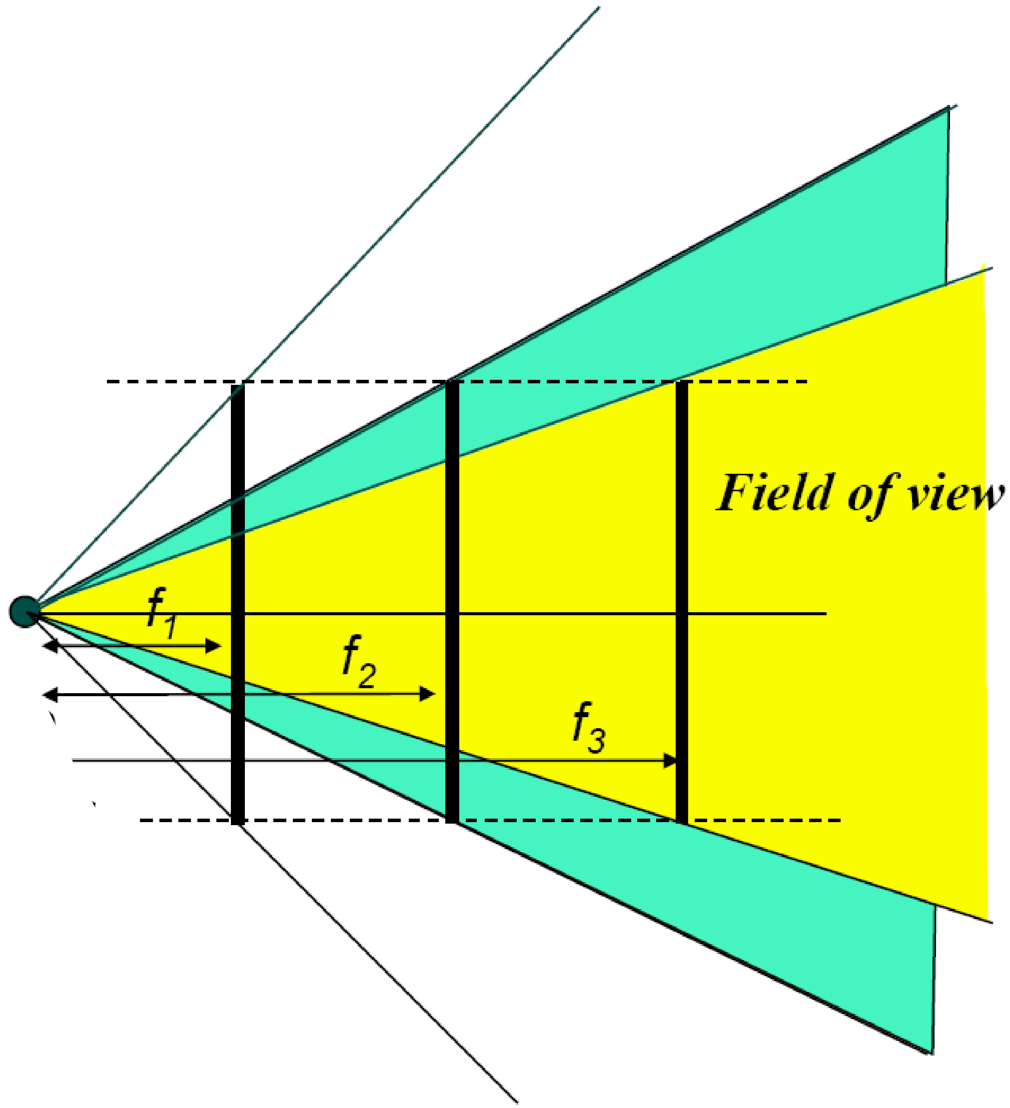
Field of view depends on focal length

As ***f*** gets smaller, image becomes more *wide angle*

- more world points project onto the finite image plane

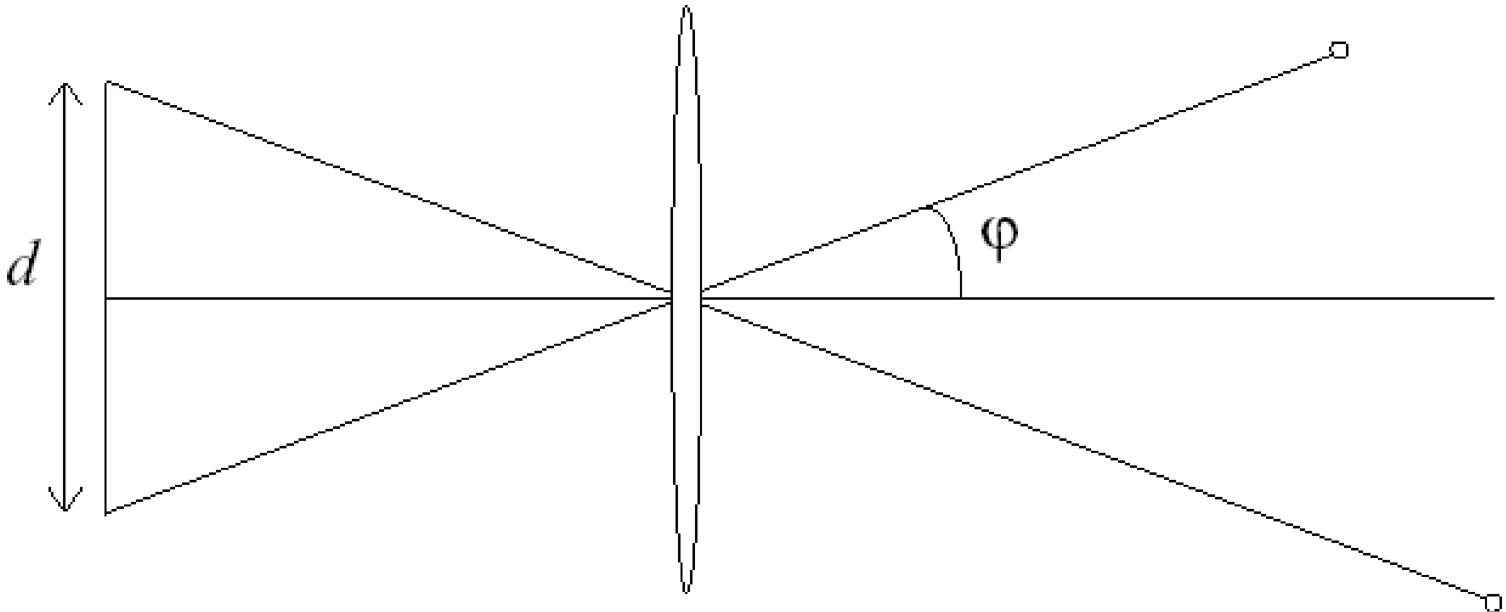
As ***f*** gets larger, image becomes more *telescopic*

- smaller part of the world projects onto the finite image plane



from R. Duraiswami

Field of view depends on focal length



Size of field of view governed by size of the camera retina:

$$\varphi = \tan^{-1}\left(\frac{d}{2f}\right)$$

Smaller FOV = larger Focal Length

Slide by A. Efros

Vignetting

http://www.ptgui.com/examples/vigntutorial.html

http://www.tlucretius.net/Photo/eHolga.html

Vignetting

"natural":

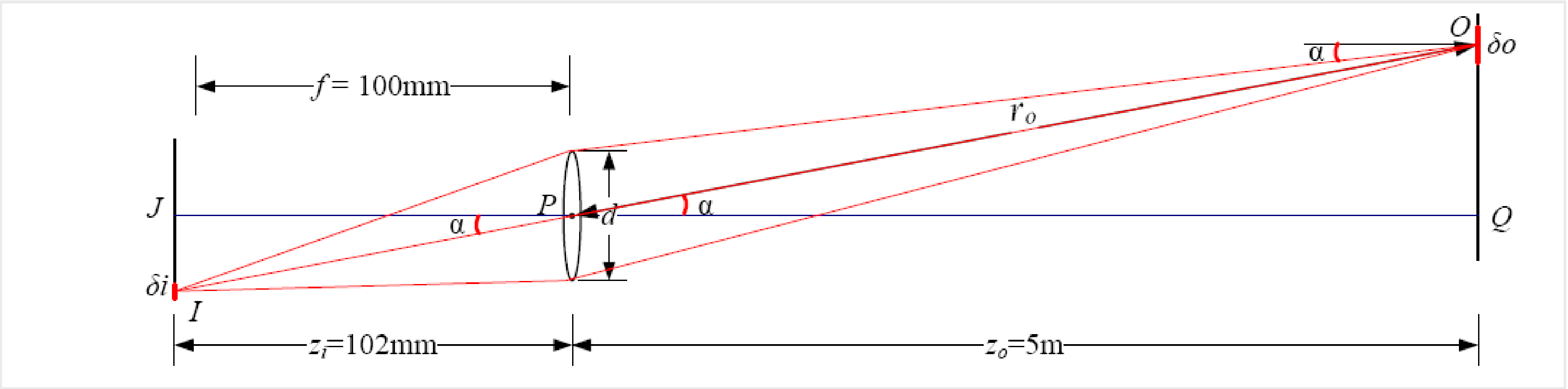


Figure 2.23: *The amount of light hitting a pixel of surface area δi depends on the square of the ratio of the aperture diameter d to the focal length f, as well as the fourth power of the off-axis angle α cosine, $\cos^4 \alpha$.*

Fundamental radiometric relation between the scene radiance L and the light (irradiance) E reaching the pixel sensor:

$$E = L \frac{\pi}{4} \left(\frac{d}{f} \right)^2 \cos^4 \alpha,$$

Physical parameters of image formation

Geometric

- Type of projection
- Camera pose

Optical

- Sensor’s lens type
- focal length, field of view, aperture

Photometric

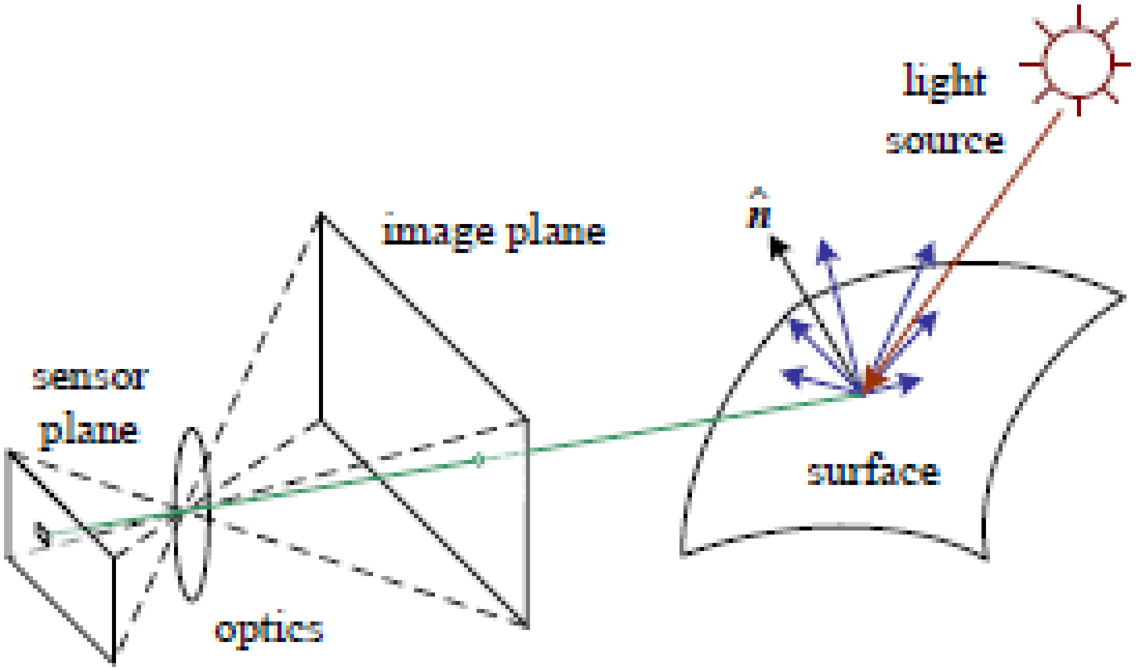
- Type, direction, intensity of light reaching sensor
- Surfaces’ reflectance properties

Sensor

- sampling, etc.

Photometric Image Formation

A simplified model of photometric image formation. Light is emitted by one or more light sources and is then reflected from an object’s surface. A portion of this light is directed towards the camera. This simplified model ignores multiple reflections, which often occur in real-world scenes.



BRDF (Bidirectional Reflectance Distribution Function)

$$f_r(\theta_i, \phi_i, \theta_r, \phi_r; \lambda)$$

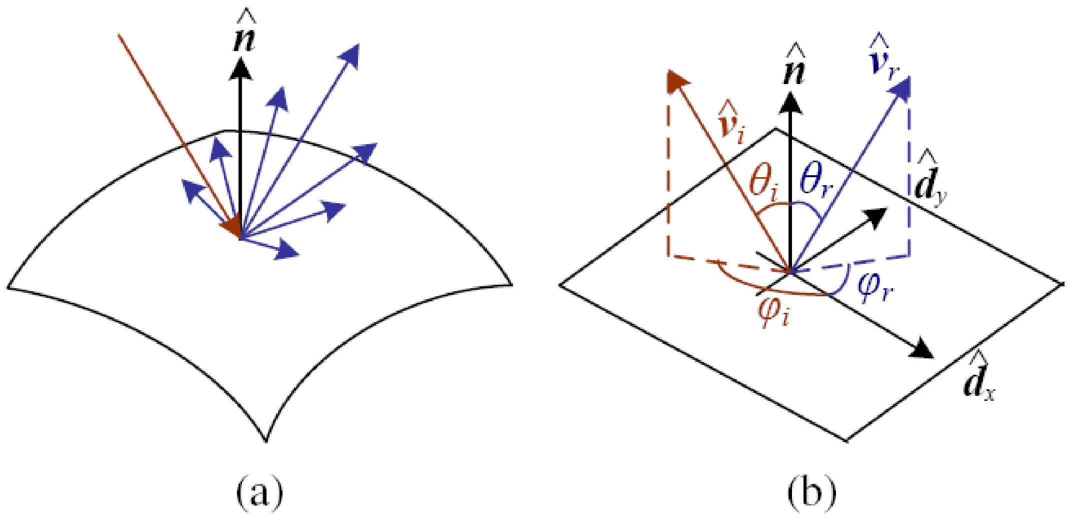


Figure 2.15: *(a) Light scattering when hitting a surface. (b) The bidirectional reflectance distribution function (BRDF) $f(\theta_i, \phi_i, \theta_r, \phi_r)$ is parameterized by the angles the incident $\hat{\boldsymbol{v}}_i$ and reflected $\hat{\boldsymbol{v}}_r$ light ray directions make with the local surface coordinate frame $(\hat{\boldsymbol{d}}_x, \hat{\boldsymbol{d}}_y, \hat{\boldsymbol{n}})$.*

For an isotropic material, we can simplify the BRDF to

$$f_r(\theta_i, \theta_r, |\phi_r - \phi_i|; \lambda) \quad \text{or} \quad f_r(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_r, \hat{\boldsymbol{n}}; \lambda),$$

Diffuse / Lambertian

Figure 2.16: *This close-up of a statue shows both diffuse (smooth shading) and specular (shiny highlight) reflection, as well as the darkening in the grooves and creases due to reduced light visibility and interreflections. (Photo courtesy of Alyosha Efros.)*

While light is scattered uniformly in all directions, i.e., the BRDF is constant,

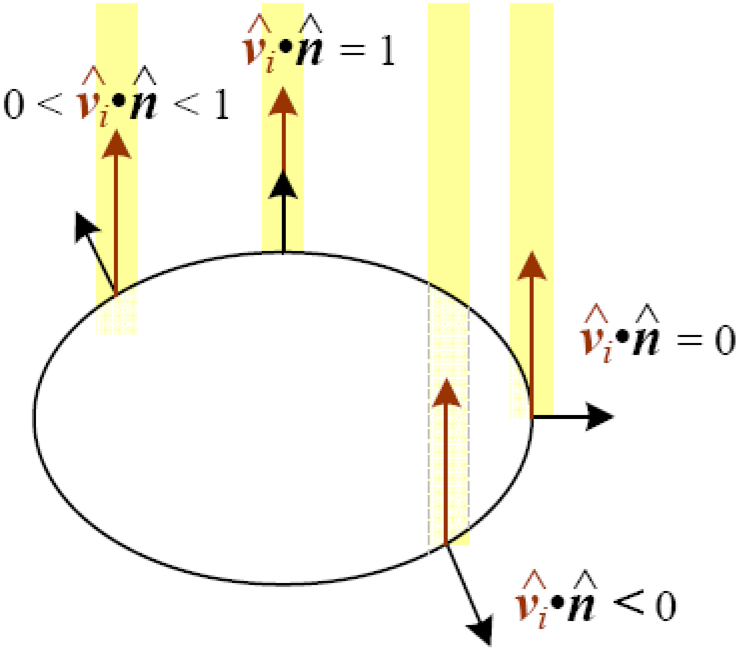
$$f_d(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_r, \hat{\boldsymbol{n}}; \lambda) = f_d(\lambda),$$

Shading equation for diffuse reflection :

$$L_d(\hat{\boldsymbol{v}}_r; \lambda) = \sum_i L_i(\lambda) f_d(\lambda) \cos^+ \theta_i = \sum_i L_i(\lambda) f_d(\lambda) [\hat{\boldsymbol{v}}_i \cdot \hat{\boldsymbol{n}}]^+,$$

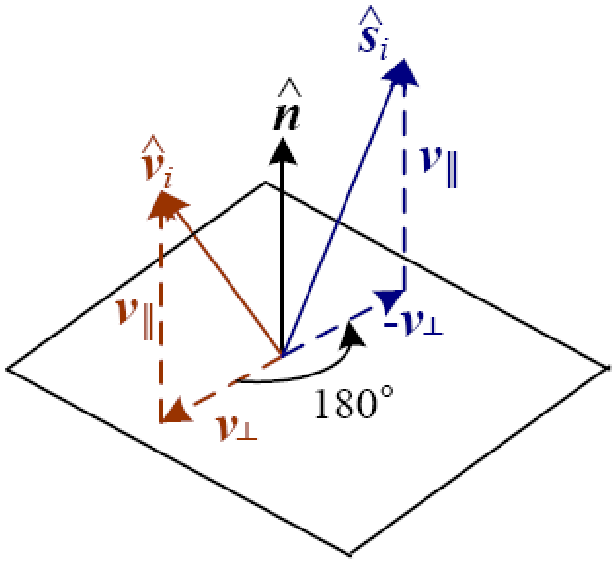
$$[\hat{\boldsymbol{v}}_i \cdot \hat{\boldsymbol{n}}]^+ = \max(0, \hat{\boldsymbol{v}}_i \cdot \hat{\boldsymbol{n}})$$

Foreshortening



The diminution of returned light caused by foreshortening *depends on* $\hat{\boldsymbol{v}}_i \cdot \hat{\boldsymbol{n}}$, *the cosine of the angle between the incident light direction* $\hat{\boldsymbol{v}}_i$ *and the surface normal* $\hat{\boldsymbol{n}}$.

Specular reflection



The amount of light reflected in a given direction $\hat{v}_r$ thus depends on the angle $\theta_s = \cos^{-1}(\hat{v}_r \cdot \hat{s}_i)$ between the view direction $\hat{v}_r$ and the specular direction $\hat{s}_i$. For example, the Phong (1975) model uses a power of the cosine of the angle,

$$f_s(\theta_s; \lambda) = k_s(\lambda) \cos^{k_e} \theta_s, \tag{2.90}$$

while the Torrance and Sparrow (1967) micro-facet model uses a Gaussian,

$$f_s(\theta_s; \lambda) = k_s(\lambda) \exp(-c_s^2 \theta_s^2). \tag{2.91}$$

Phong

Diffuse+specular+ambient:

$$f_d(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_r, \hat{\boldsymbol{n}}; \lambda) = f_d(\lambda),$$

$$f_s(\theta_s; \lambda) = k_s(\lambda) \cos^{k_e} \theta_s,$$

$$f_a(\lambda) = k_a(\lambda) L_a(\lambda).$$

$$L_r(\hat{\boldsymbol{v}}_r; \lambda) = k_a(\lambda) L_a(\lambda) + k_d(\lambda) \sum_i L_i(\lambda) [\hat{\boldsymbol{v}}_i \cdot \hat{\boldsymbol{n}}]^+ + k_s(\lambda) \sum_i L_i(\lambda) (\hat{\boldsymbol{v}}_r \cdot \hat{\boldsymbol{s}}_i)^{k_e}.$$

Physical parameters of image formation

Geometric

- Type of projection
- Camera pose

Optical

- Sensor’s lens type
- focal length, field of view, aperture

Photometric

- Type, direction, intensity of light reaching sensor
- Surfaces’ reflectance properties

Sensor

- sampling, etc.

Digital cameras

Film → sensor array

Often an array of charge coupled devices

Each CCD is light sensitive diode that converts photons (light energy) to electrons

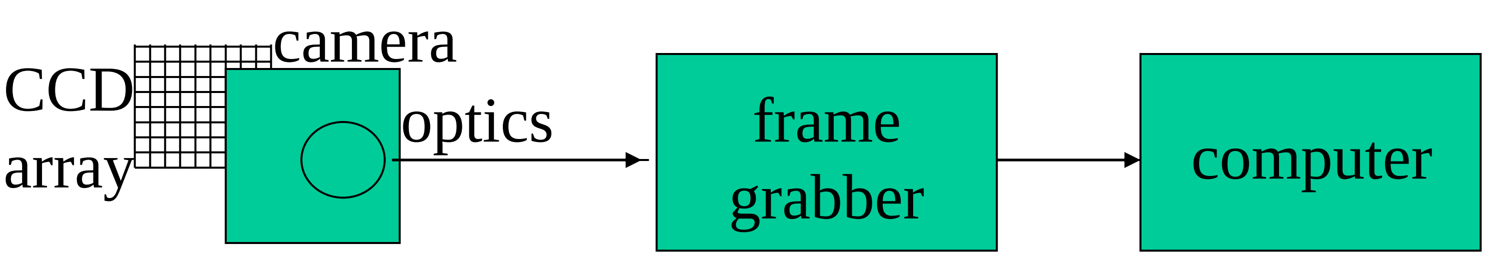
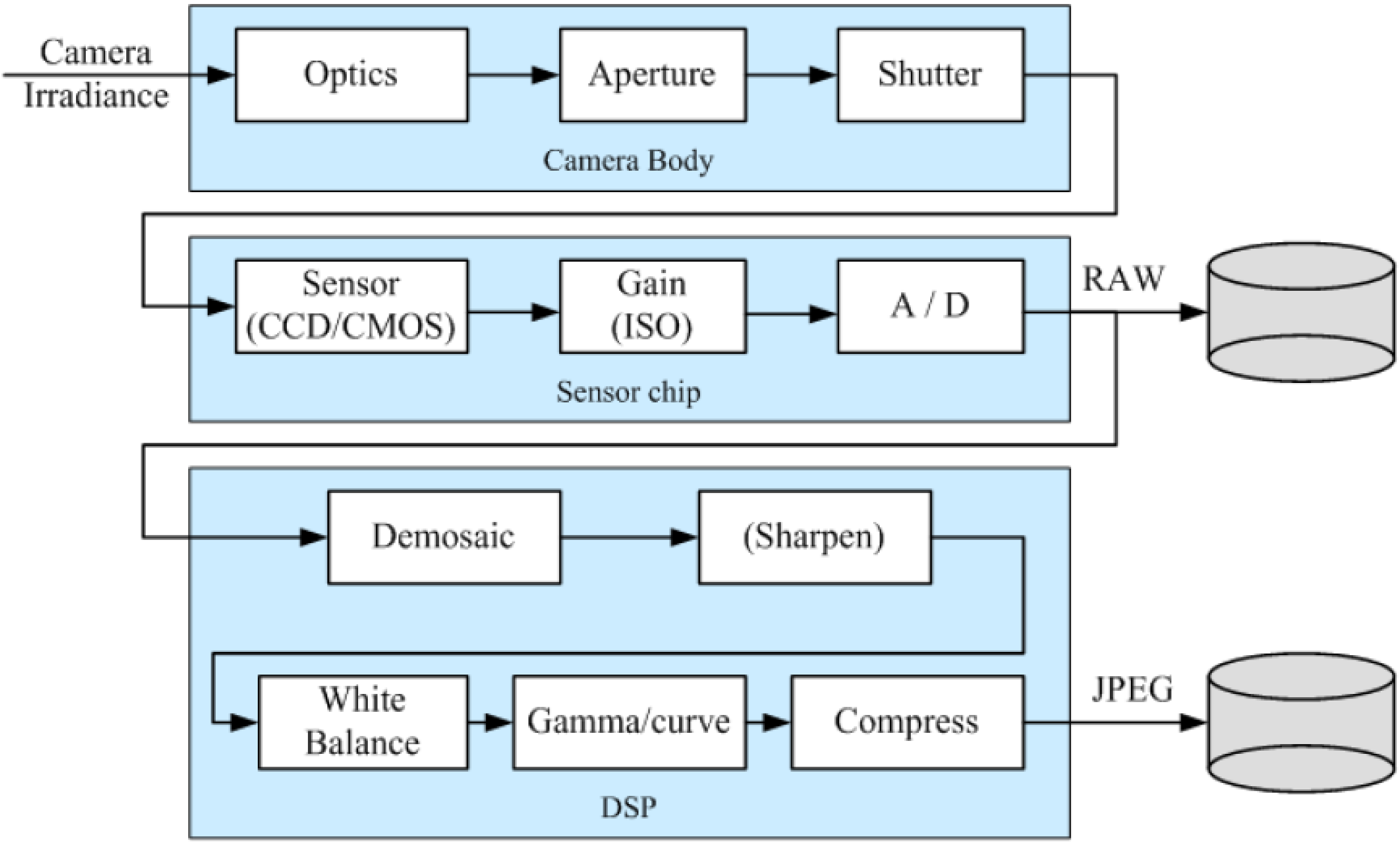
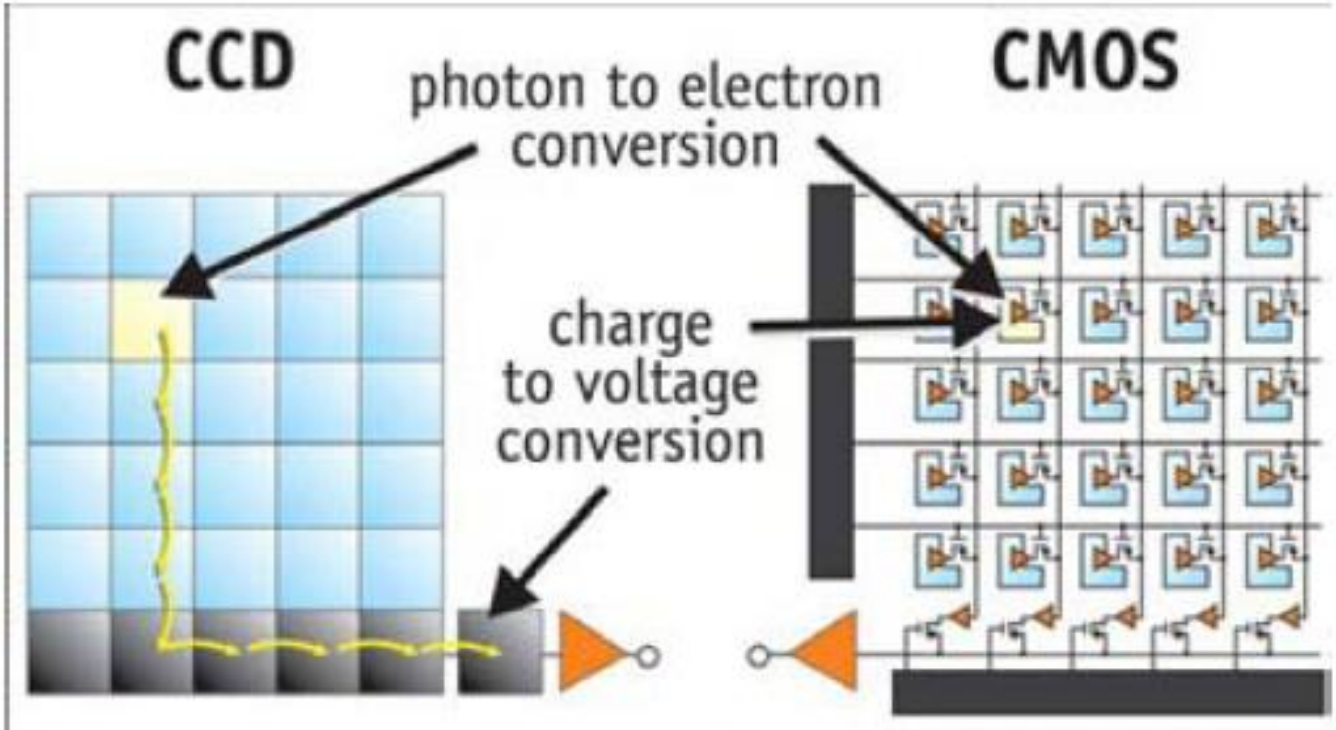


Image sensing pipeline



Digital Sensors



CCDs move photogenerated charge from pixel to pixel and convert it to voltage at an output node. CMOS imagers convert charge to voltage inside each pixel.

Resolution

sensor: size of real world scene element a that images to a single pixel

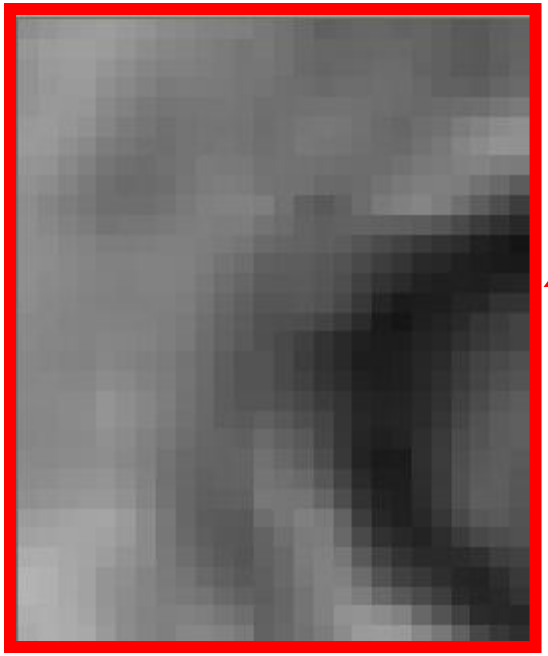
image: number of pixels

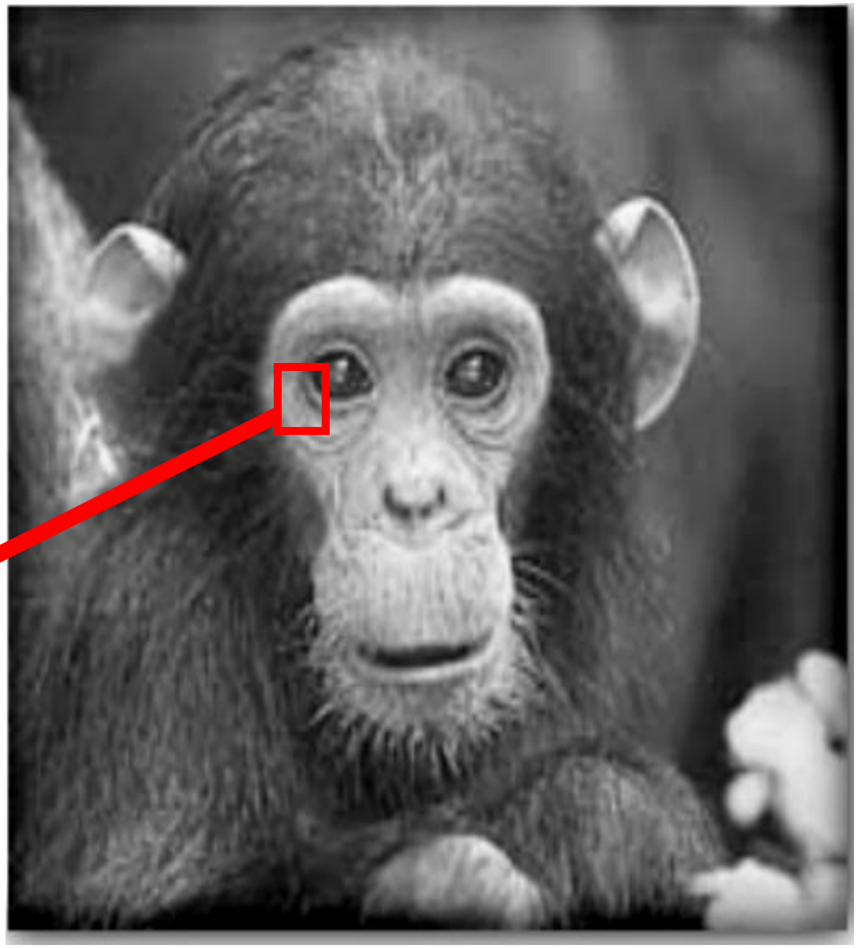
Influences what analysis is feasible, affects best representation choice.

[fig from Mori et al]

Digital images

Think of images as matrices taken from CCD array.

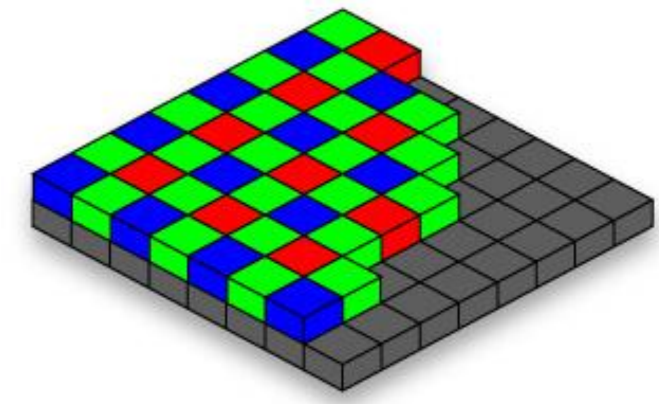


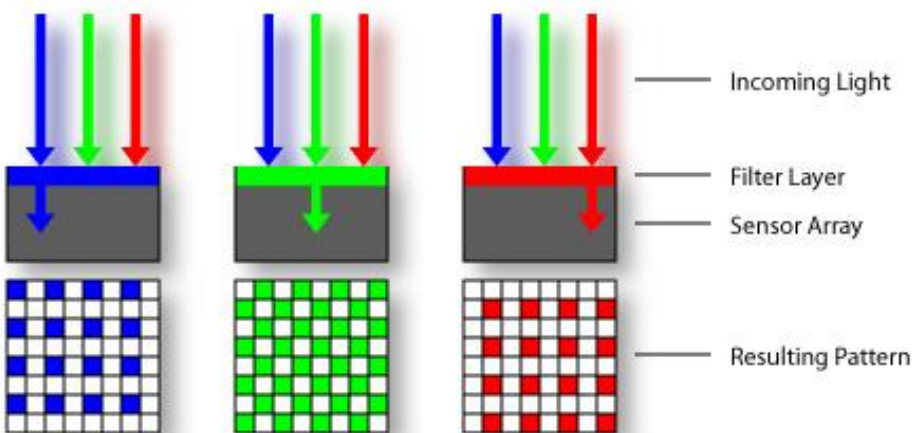


K. Grauman

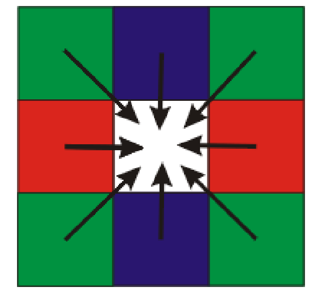
Color sensing in digital cameras

Bayer grid



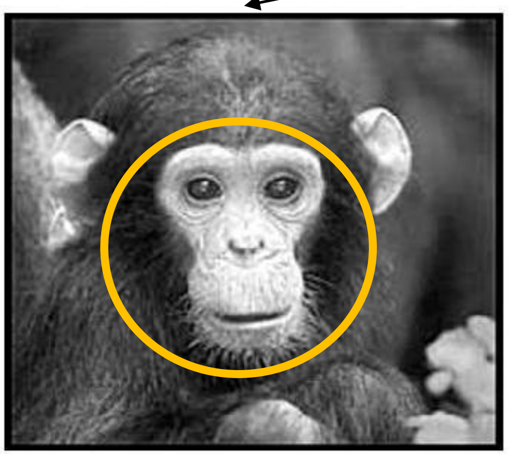


Estimate missing components from neighboring values (demosaicing)



Source: Steve Seitz

Color images,
RGB color space



R

G

B

Much more on color in next lecture…

Summary

- Geometric projection models
- Optical issues
- Photometric models
- Image sensing in digital camera