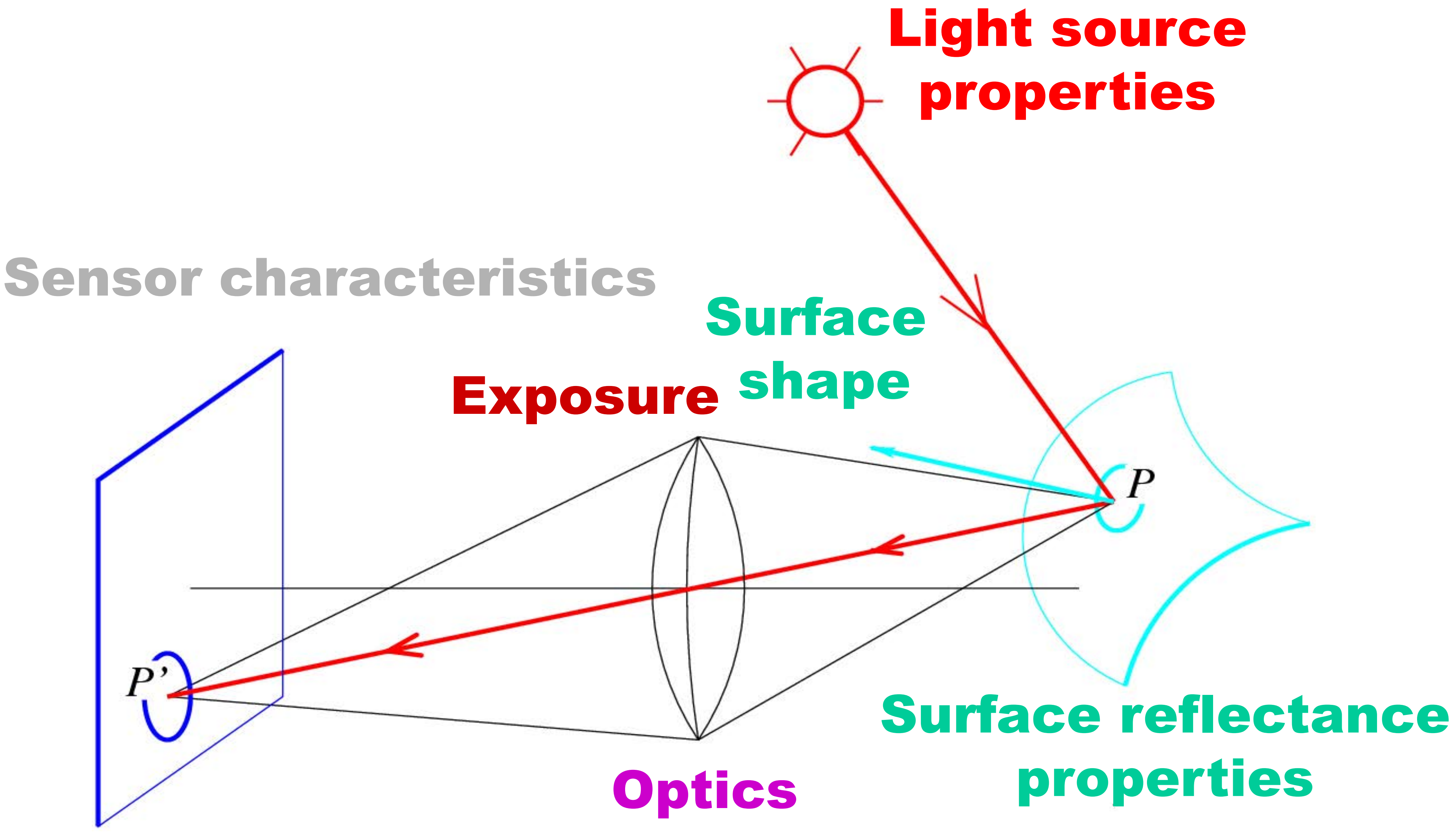
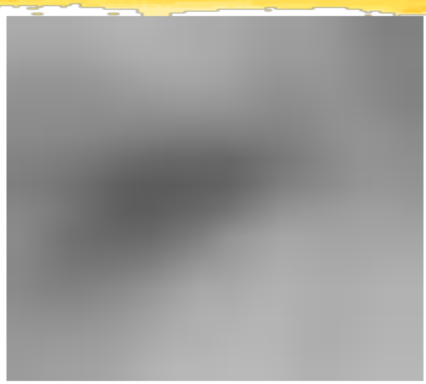
IMAGE FORMATION



ANALOG IMAGES

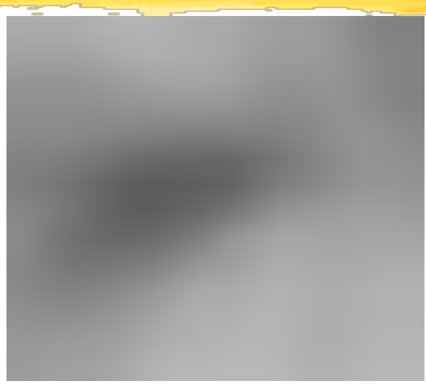


An image can be understood as a 2D light intensity function $f(x,y)$ where:

- x and y are spatial coordinates
- $f(x, y)$ is proportional to the brightness or gray value of the image at that point.

→ Cannot be stored as such on a digital computer.

DIGITAL IMAGES



136	134	161	159	163	168	171	173	173	171	166	159	157	155
152	145	136	130	151	149	151	154	158	161	163	163	159	151
145	149	149	145	140	133	145	143	145	145	145	146	148	148
148	143	141	145	145	145	141	136	136	135	135	136	135	133
131	131	129	129	133	136	140	142	142	138	130	128	126	120
115	111	108	106	106	110	120	130	137	142	144	141	129	123
117	109	098	094	094	094	100	110	125	136	141	147	147	145
136	124	116	105	096	096	100	107	116	131	141	147	150	152
152	152	137	124	113	108	105	108	117	129	139	150	157	159
159	157	157	159	135	121	120	120	121	127	136	147	158	163
165	165	163	163	163	166	136	131	135	138	140	145	154	163
166	168	170	168	166	168	170	173	145	143	147	148	152	159
168	173	173	175	173	171	170	173	177	178	151	151	153	156
161	170	176	177	177	179	176	174	174	176	177	179	155	157
161	162	168	176	180	180	180	182	180	175	175	178	180	180

A digitized image is one in which:

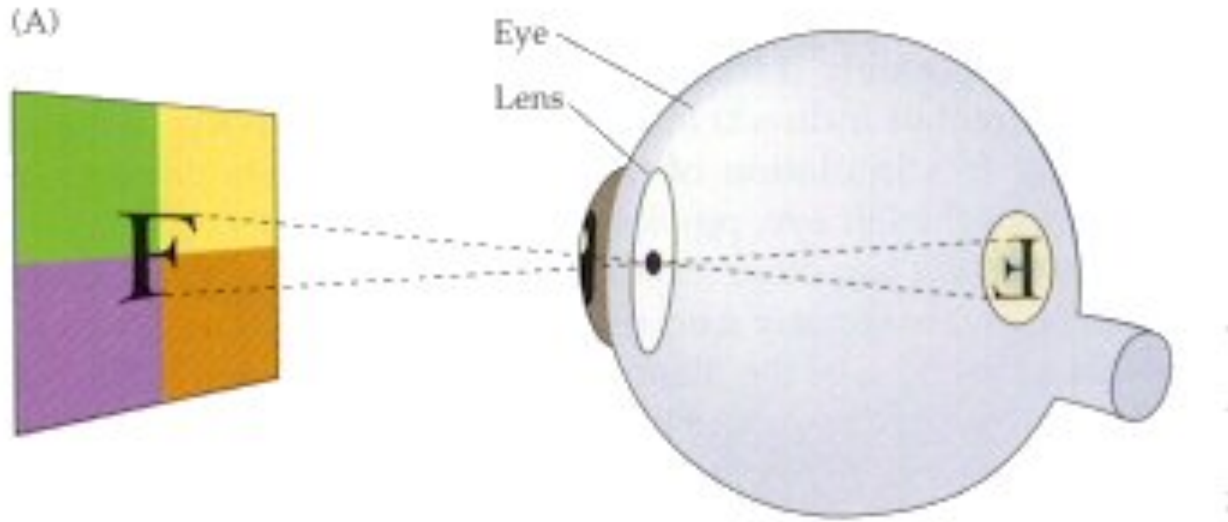
- Spatial and grayscale values have been made discrete.
- Intensities measured across a regularly spaced grid in x and y directions are sampled to
  - 8 bits (256 values) per point for black and white,
  - 3x8 bits per point for color images.

They are stored as a two dimensional arrays of gray-level values. The array elements are called pixels and identified by their x, y coordinates.

PIXELS

136	134	161	159	163	168	171	173	173	171	166	159	157	155
152	145	136	130	151	149	151	154	158	161	163	163	159	151
145	149	149	145	140	133	145	143	145	145	145	146	148	148
148	143	141	145	145	145	141	136	136	135	135	136	135	133
131	131	129	129	133	136	140	142	142	138	130	128	126	120
115	111	108	106	106	110	120	130	137	142	144	141	129	123
117	109	098	094	094	094	100	110	125	136	141	147	147	145
136	124	116	105	096	096	100	107	116	131	141	147	150	152
152	152	137	124	113	108	105	108	117	129	139	150	157	159
159	157	157	159	135	121	120	120	121	127	136	147	158	163
165	165	163	163	163	166	136	131	135	138	140	145	154	163
166	168	170	168	166	168	170	173	145	143	147	148	152	159
168	173	173	175	173	171	170	173	177	178	151	151	153	156
161	170	176	177	177	179	176	174	174	176	177	179	155	157
161	162	168	176	180	180	180	182	180	175	175	178	180	180

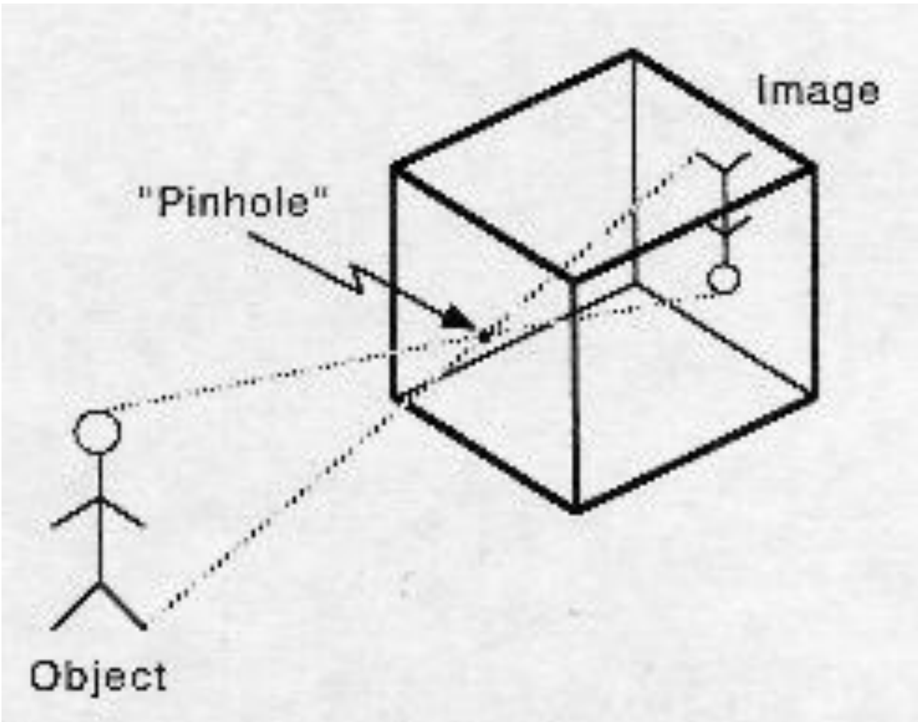
IMAGE FORMATION



Projection from surfaces to 2-D sensor.

- Where: Geometry
- How bright: Radiometry
- Stored how: Sensing

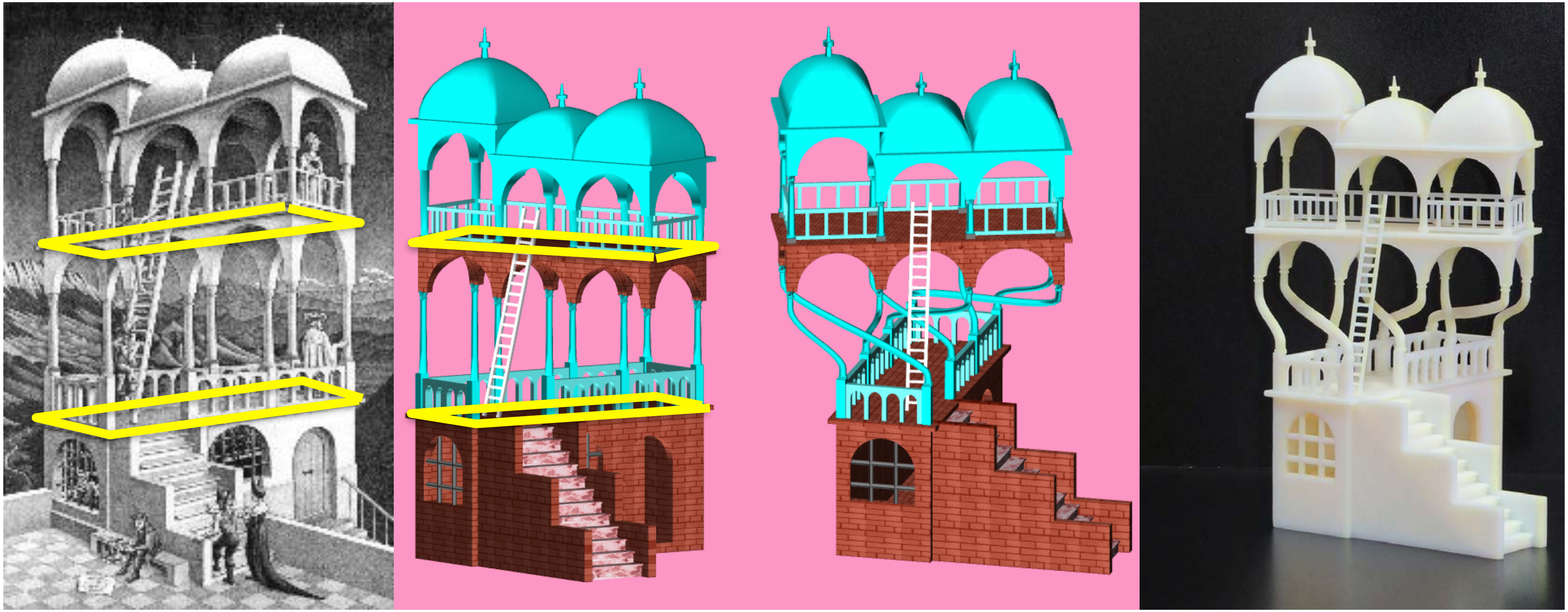
PINHOLE CAMERA MODEL



Idealized model of the perspective projection:

- All rays go through a hole and form a pencil of lines.
- The hole acts as a ray selector that allows an inverted image to form.

ESCHER'S BELVEDERE

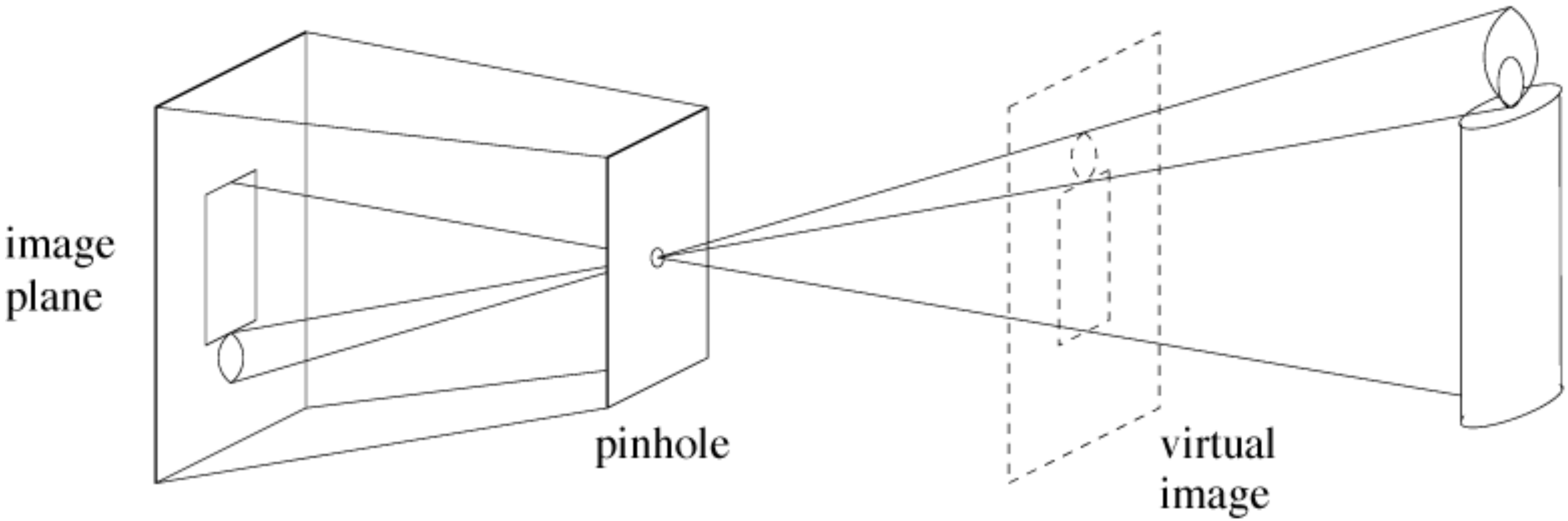


M. C. Escher | Impossible | Possible | Done

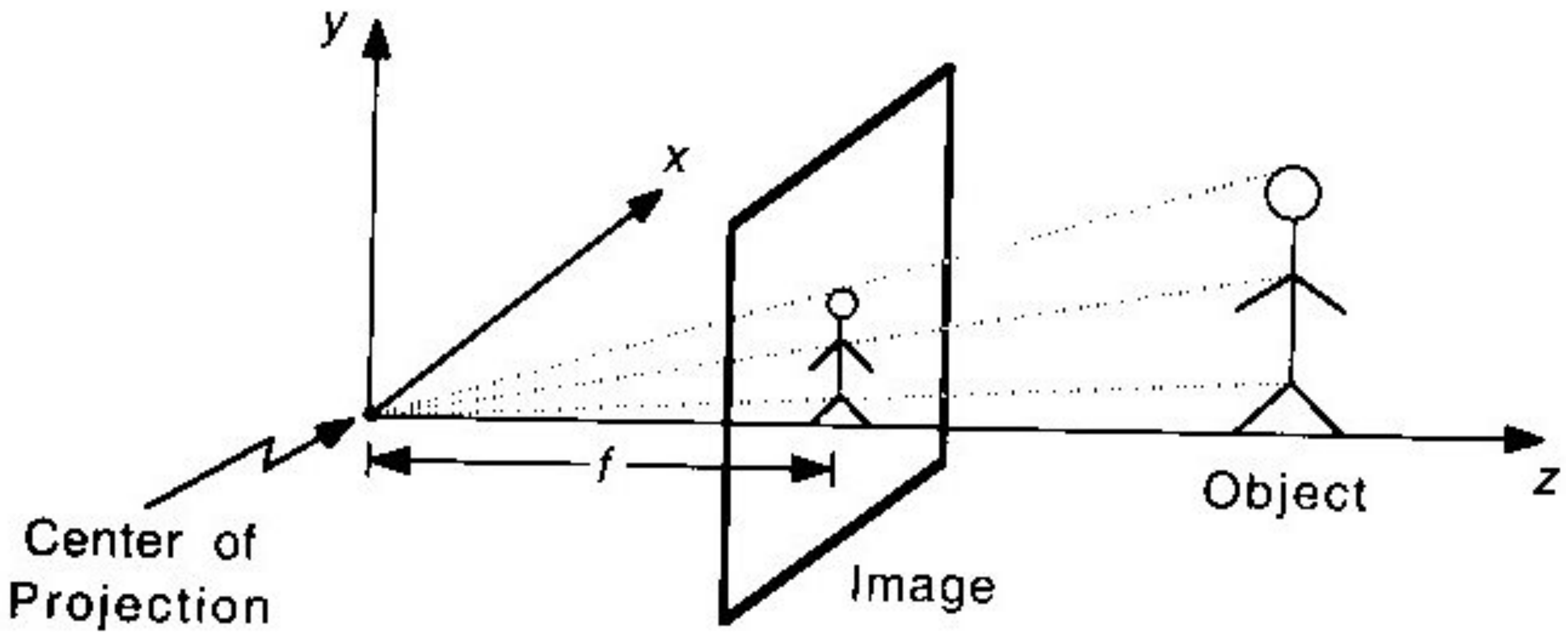
Gershon Helber
http://www.cs.technion.ac.il/~gershon/

MAGNET LIKE SLOPES

VIRTUAL IMAGE



CAMERA GEOMETRY



Pinhole geometry without image reversal

COORDINATE SYSTEMS

World, Camera, Image Coordinate Systems

World Coordinate System:

$$(X_w, Y_w, Z_w)$$

Camera Coordinate System:

$$(X_c, Y_c, Z_c)$$

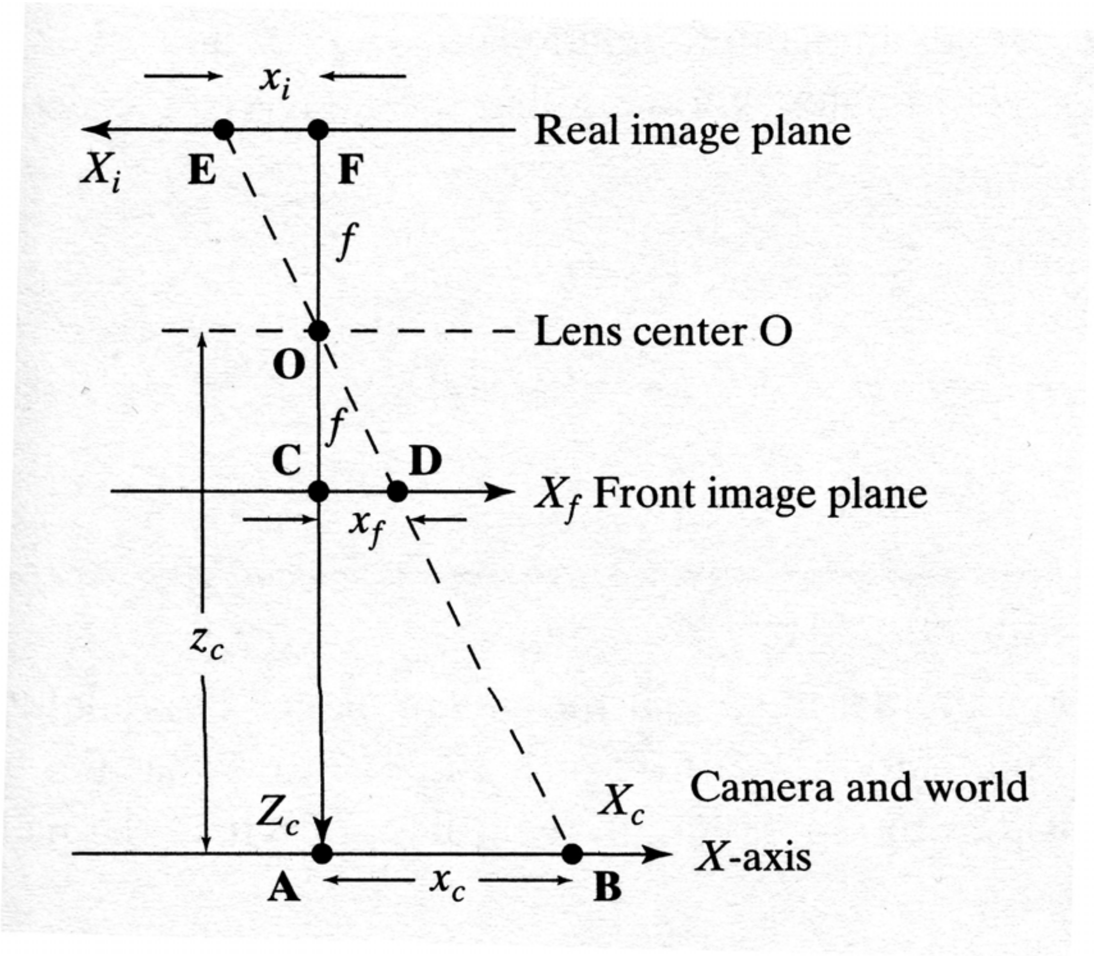
Image Coordinate System:

$$(X_i, Y_i, Z_i)$$

CAMERA COORDINATE SYSTEM

- The center of the projection coincides with the origin of the world.
- The camera axis (optical axis) is aligned with the world’s z-axis.
- To avoid image inversion, the image plane is in front of the center of projection.

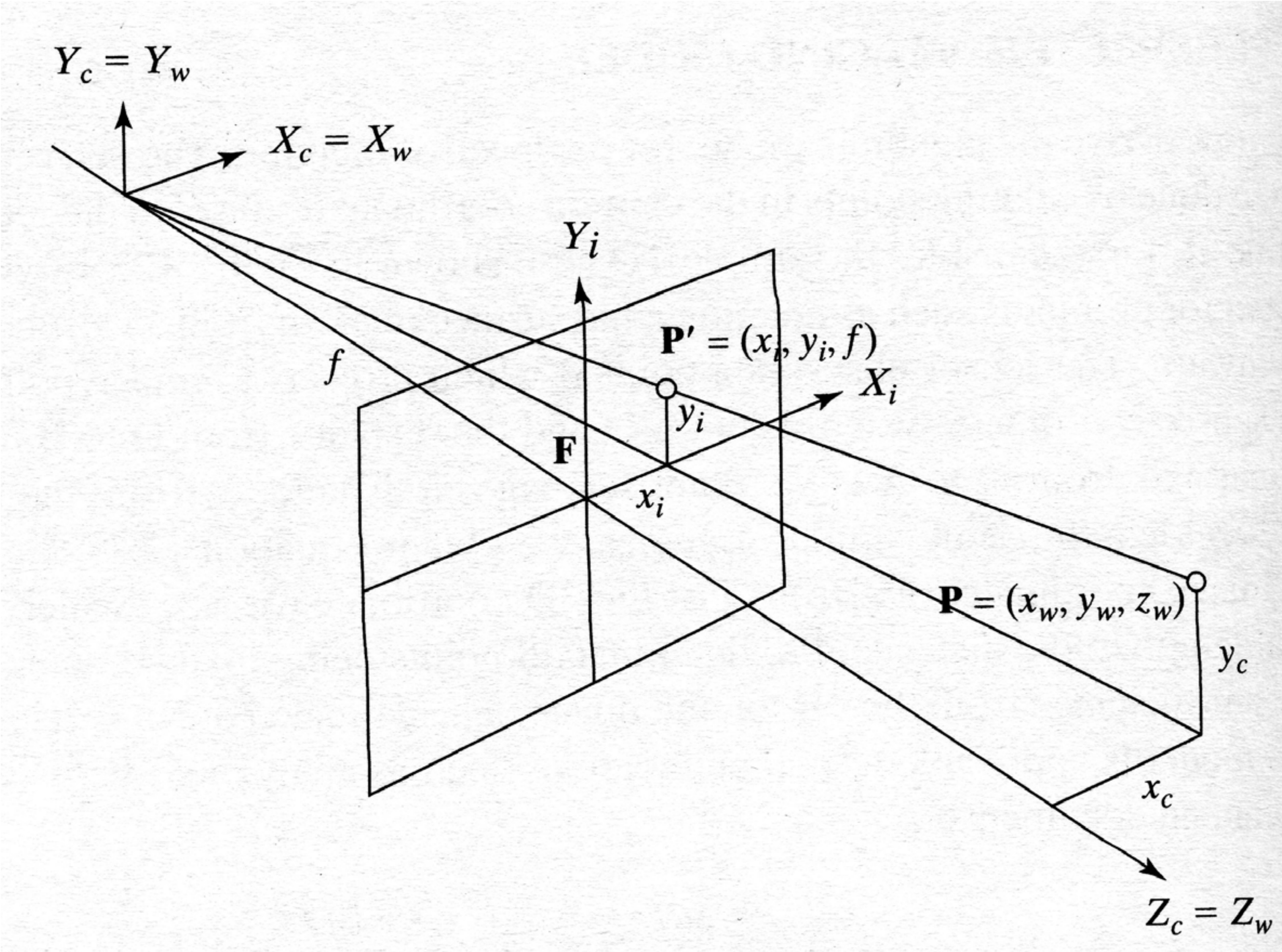
1D IMAGE



$$X_f / f = X_c / Z_c$$

$$X_f = f \frac{X_c}{Z_c}$$

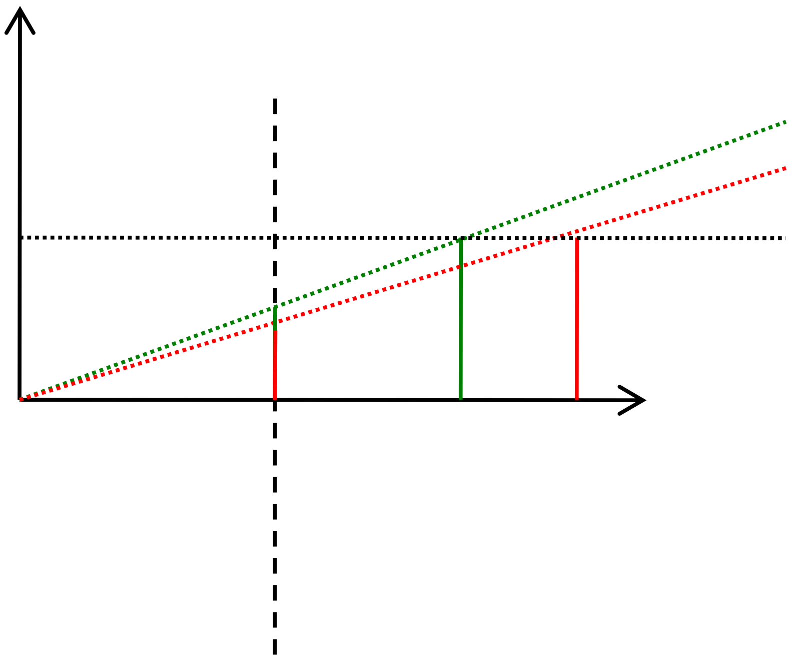
2D IMAGE

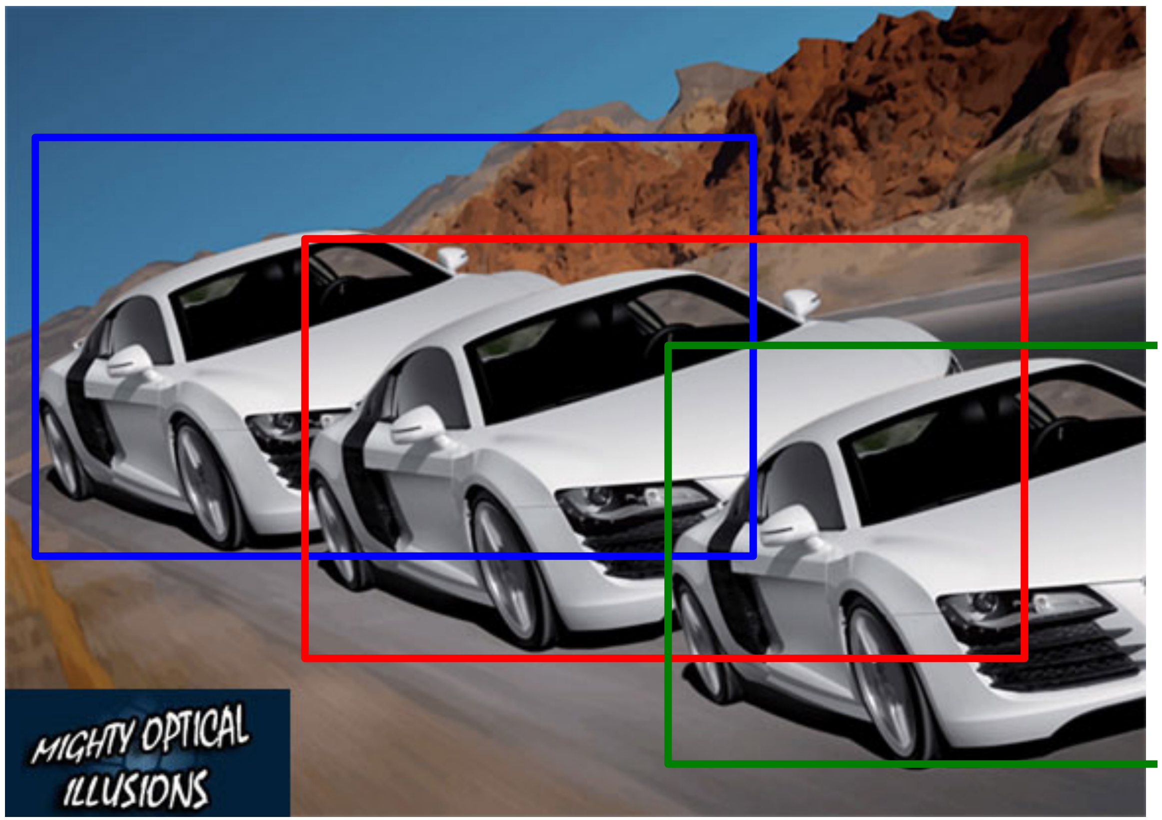


$$X_i = f \frac{X_c}{Z_c}$$

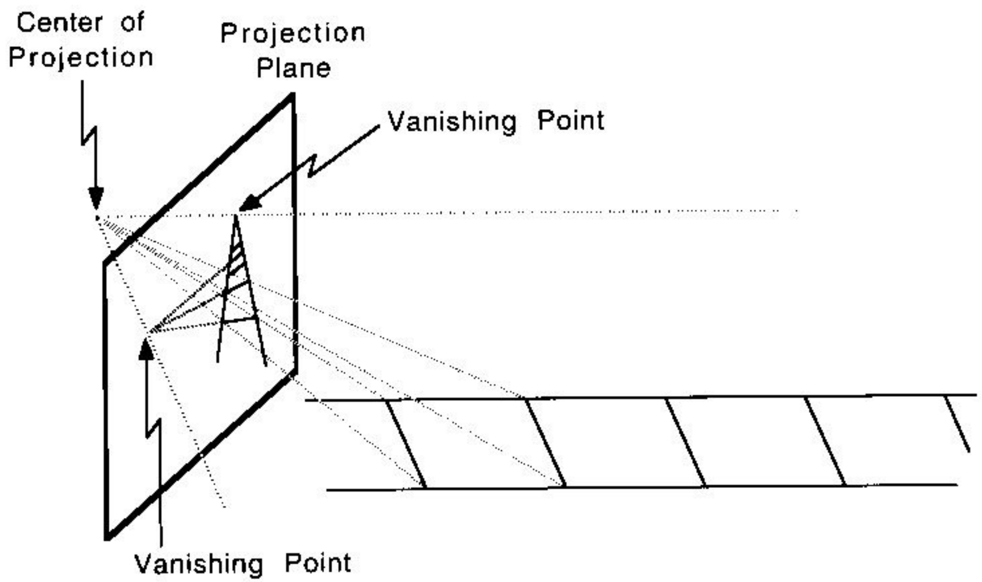
$$Y_i = f \frac{Y_c}{Z_c}$$

DISTANT OBJECTS APPEAR SMALLER





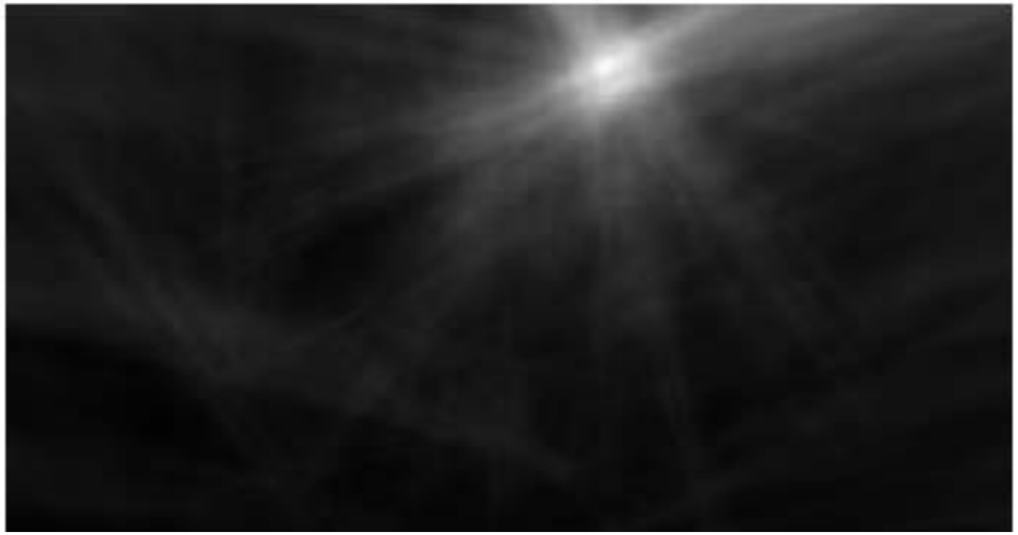
PARALLEL LINES MEET

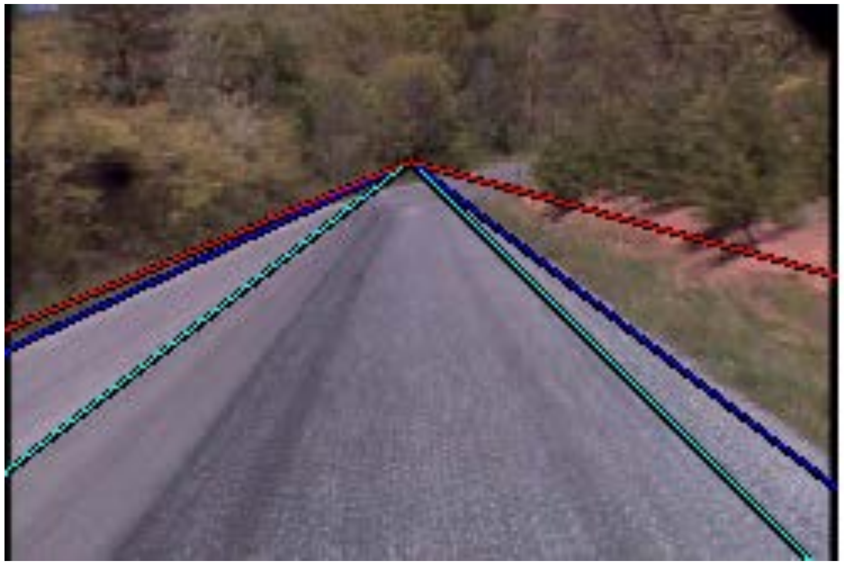


VANISHING POINTS

- The projections of parallel lines all meet at one point, called the vanishing point.
- As focal length and distance to camera increase, the image remains the same size but perspective effects diminish.

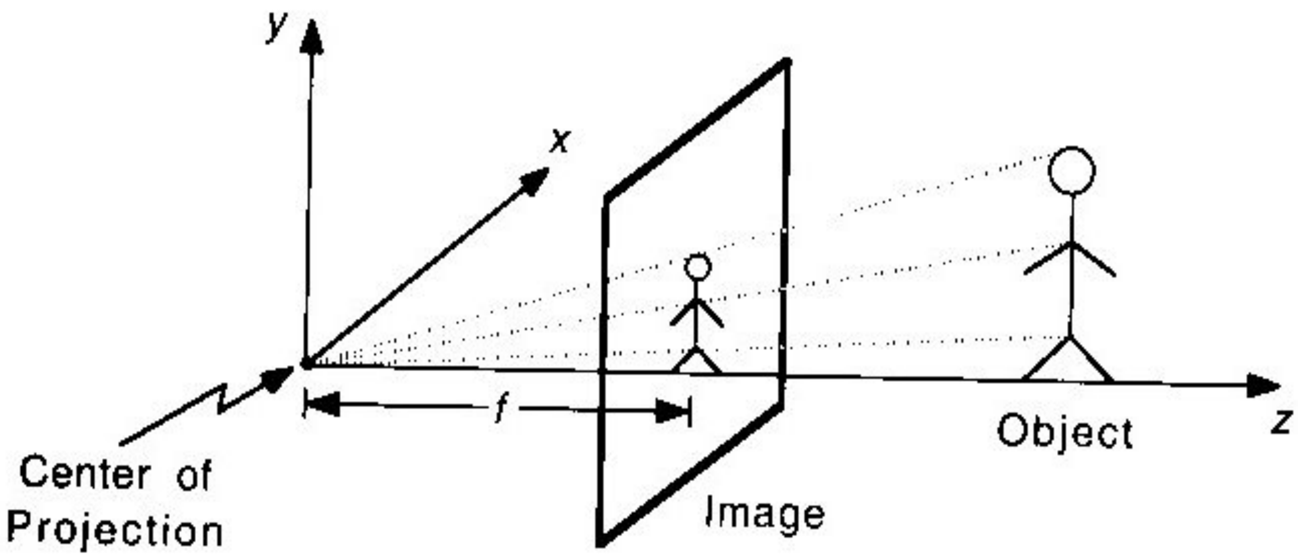
ROAD FOLLOWING





Rasmussen et al. , BMVC'04

PROJECTION IS NON LINEAR



$$u = X_i = f \frac{X_c}{Z_c}$$

$$v = Y_i = f \frac{Y_c}{Z_c}$$

→Reformulate it as a linear operation.

HOMOGENEOUS COORDINATES

Homogeneous representation of 2D point:

$$\mathbf{x} = (x_1, x_2, x_3) \text{ represents } (x_1/x_3, x_2/x_3)$$

Homogeneous representation of 3D point:

$$\mathbf{X} = (x_1, x_2, x_3, x_4) \text{ represents } (x_1/x_4, x_2/x_4, x_3/x_4)$$

→ **Projections become linear transformations.**

SIMPLE PROJECTION MATRIX

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \text{ with } X_{\mathrm{i}} = \frac{x}{z} = f\,\frac{X_c}{Z_c} \text{ and } Y_i = \frac{y}{z} = f\,\frac{Y_c}{Z_c}$$

$$= \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

INTRINSIC AND EXTRINSIC PARAMETERS

- Camera may not be at the origin, looking down the z-axis
  - → Extrinsic parameters
- One unit in camera coordinates may not be the same as one unit in world coordinates
  - → Intrinsic parameters

$$\begin{bmatrix} \mathrm{x} \\ \mathrm{y} \\ \mathrm{z} \end{bmatrix} = \begin{bmatrix} \text{Matrix of} \\ \text{intrinsic parameters} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \text{Matrix of} \\ \text{extrinsic parameters} \end{bmatrix} \begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{bmatrix}$$

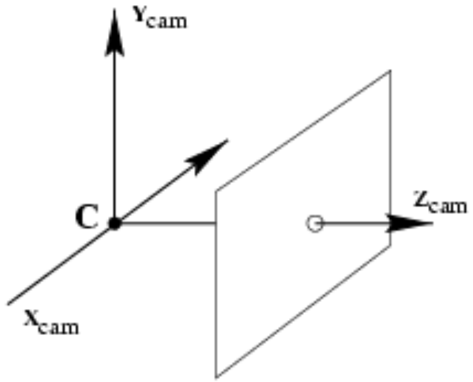
LINEAR CAMERA MODEL

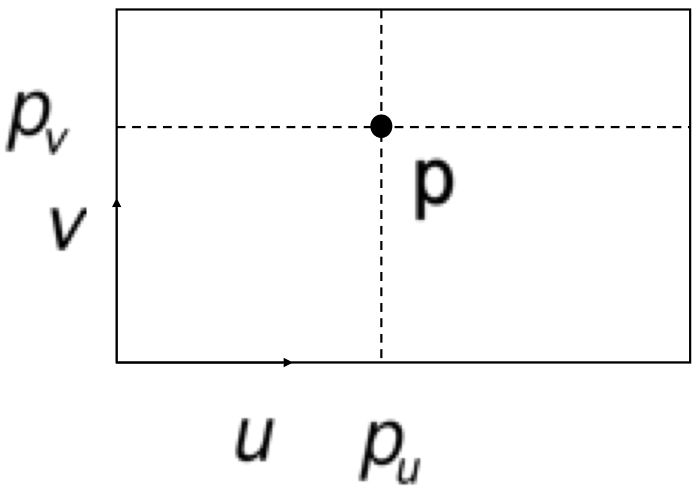
$$\begin{bmatrix} \mathrm{x} \\ \mathrm{y} \\ \mathrm{z} \end{bmatrix} = \begin{bmatrix} \text{Matrix of} \\ \text{intrinsic parameters} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \text{Matrix of} \\ \text{extrinsic parameters} \end{bmatrix} \begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{bmatrix}$$

$$= \mathbf{K} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{Rt} \begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{bmatrix}$$

where $\mathbf{K}$ is a 3×3 matrix and $\mathbf{Rt}$ a 4×4 matrix.

PRINCIPAL POINT





$$u = X_i + p_u = fX/Z + p_u$$
$$v = Y_i + p_v = fY/Z + p_v$$

$$\mathbf{K} = \begin{bmatrix} f & 0 & p_u \\ 0 & f & p_v \\ 0 & 0 & 1 \end{bmatrix}$$

INHOMOGENEOUS SCALING

$$
\begin{aligned}
u &= \alpha_u X_i + p_u = \alpha_u X/Z + p_u \\
v &= \alpha_v Y_i + p_v = \alpha_v Y/Z + p_v
\end{aligned}
$$

$$
\mathbf{K} = \begin{bmatrix} \alpha_u & 0 & p_u \\ 0 & \alpha_v & p_v \\ 0 & 0 & 1 \end{bmatrix}
$$

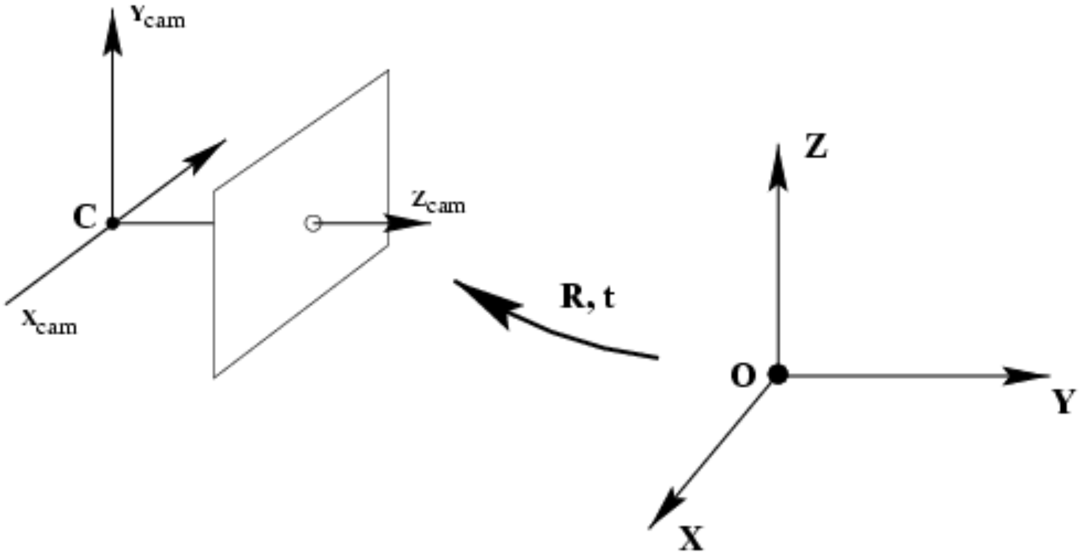
The pixels are not necessarily square.

AXIS SKEW

$$\mathbf{K} = \begin{bmatrix} \alpha_u & \mathrm{s} & p_u \\ 0 & \alpha_v & p_v \\ 0 & 0 & 1 \end{bmatrix}$$

s encodes the non-orthogonality of the u and v directions. Very close to zero in modern cameras.

ROTATION / TRANSLATION



$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{R}\left(\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} - \tilde{\mathbf{C}}\right) \text{ with } \mathbf{R}^t\mathbf{R} = \mathbf{I}$$

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \text{ with } \mathbf{T} = -\mathbf{R}\tilde{\mathbf{C}}$$

→Rotations and translations also expressed in terms of matrix multiplications in projective space.

FULL PROJECTION MATRIX

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a_u & s & p_u \\ 0 & a_v & p_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \text{ with } \mathbf{T} = -\mathbf{R}\tilde{\mathbf{C}} \text{ and } \mathbf{R}^t\mathbf{R} = \mathbf{I}$$

$$= \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\mathbf{x} = \mathbf{K}[\mathbf{R} \mid -\mathbf{R}\tilde{\mathbf{C}}]\mathbf{X} = \mathbf{KR}[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X} = [\mathbf{M} \mid \mathbf{m}]\mathbf{X} = \mathbf{PX}$$

Hartley, Chap 6.

CAMERA CALIBRATION

Internal Parameters:

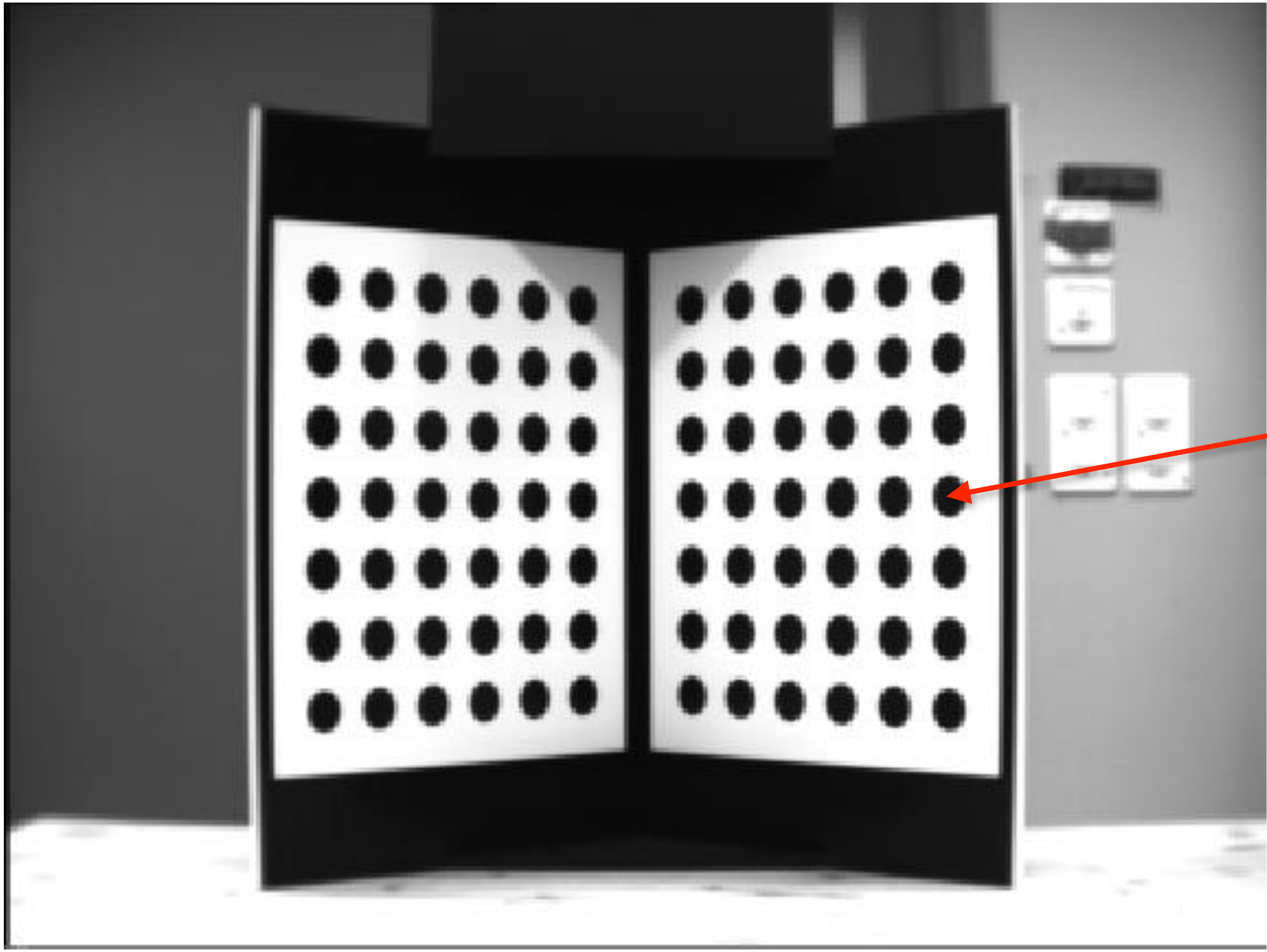
- Horizontal and vertical scaling (2)
- Principal points (2)
- Skew of the axis (1)

External Parameters:

- Rotations (3)
- Translations (3)

→ **11 free parameters.**

CALIBRATION GRID



$x_i = PX_i$

PARAMETERS OF PROJECTION MATRIX

For all i, $\mathbf{x}_i = \mathbf{P}\mathbf{X}_i$

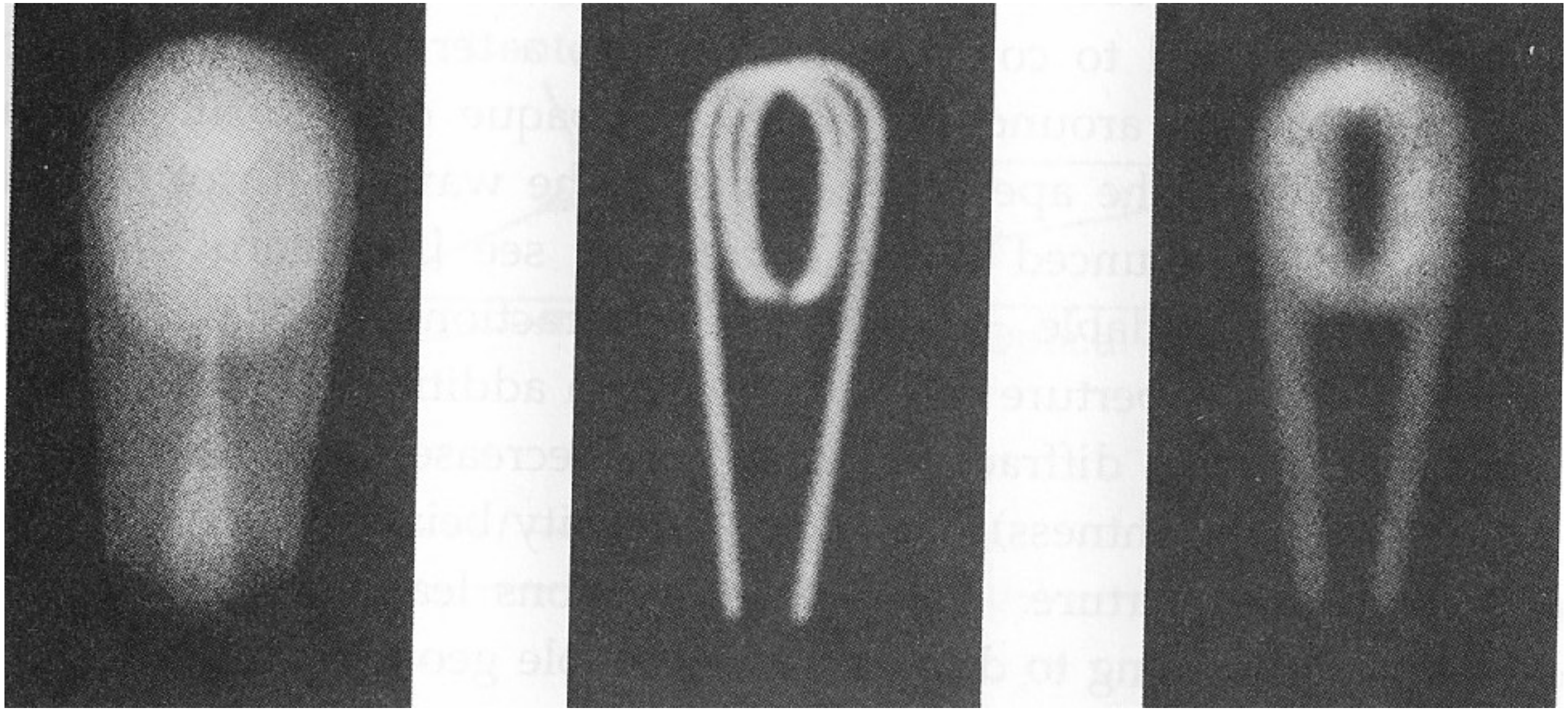
- Number of measurements required:
  - 11 degrees of freedom.
  - 2 constraints per correspondence.
- Direct linear transform:
  - Minimal solution for 6 correspondences
  - Over-constrained solutions by imposing

$$\|P\| = 1 \text{ or } P_{34} = 1$$

- Non linear optimization.

Hartley, Chap 7.

LIMITATIONS

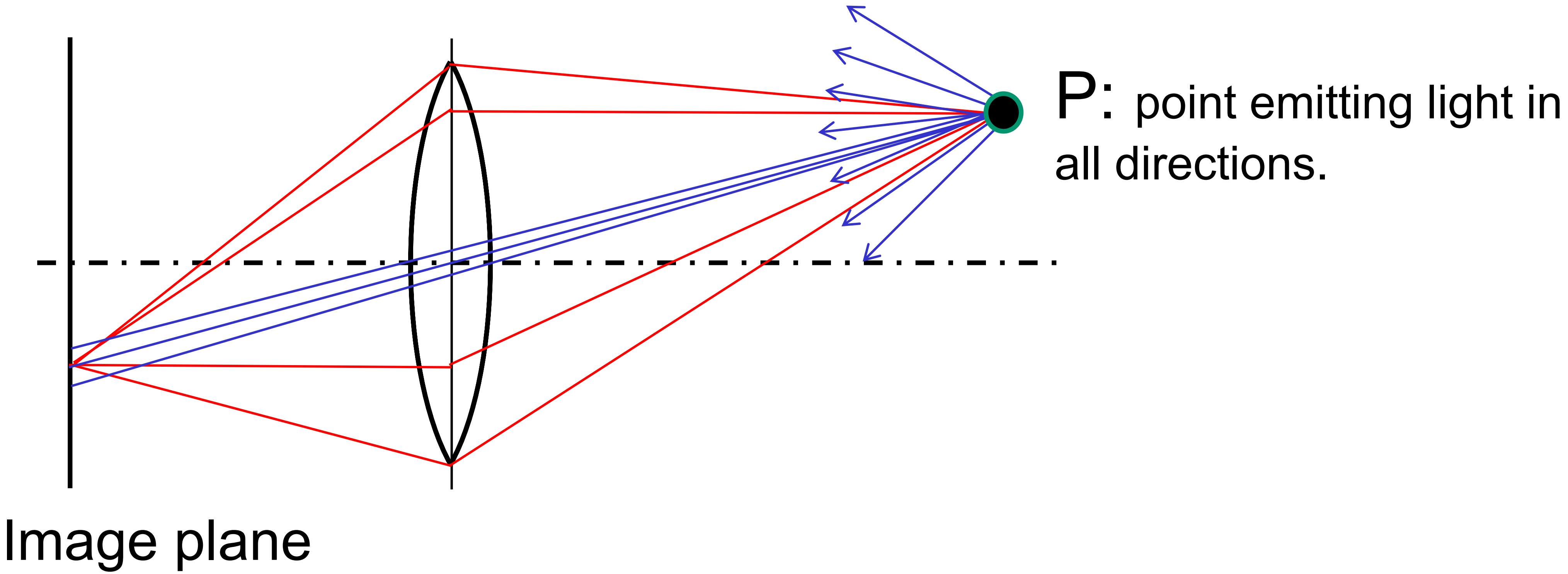


Idealization because the hole cannot be infinitely small

- Image would be infinitely dim
- Diffraction effects

→ Use of Lenses.

IMAGING WITH A LENS

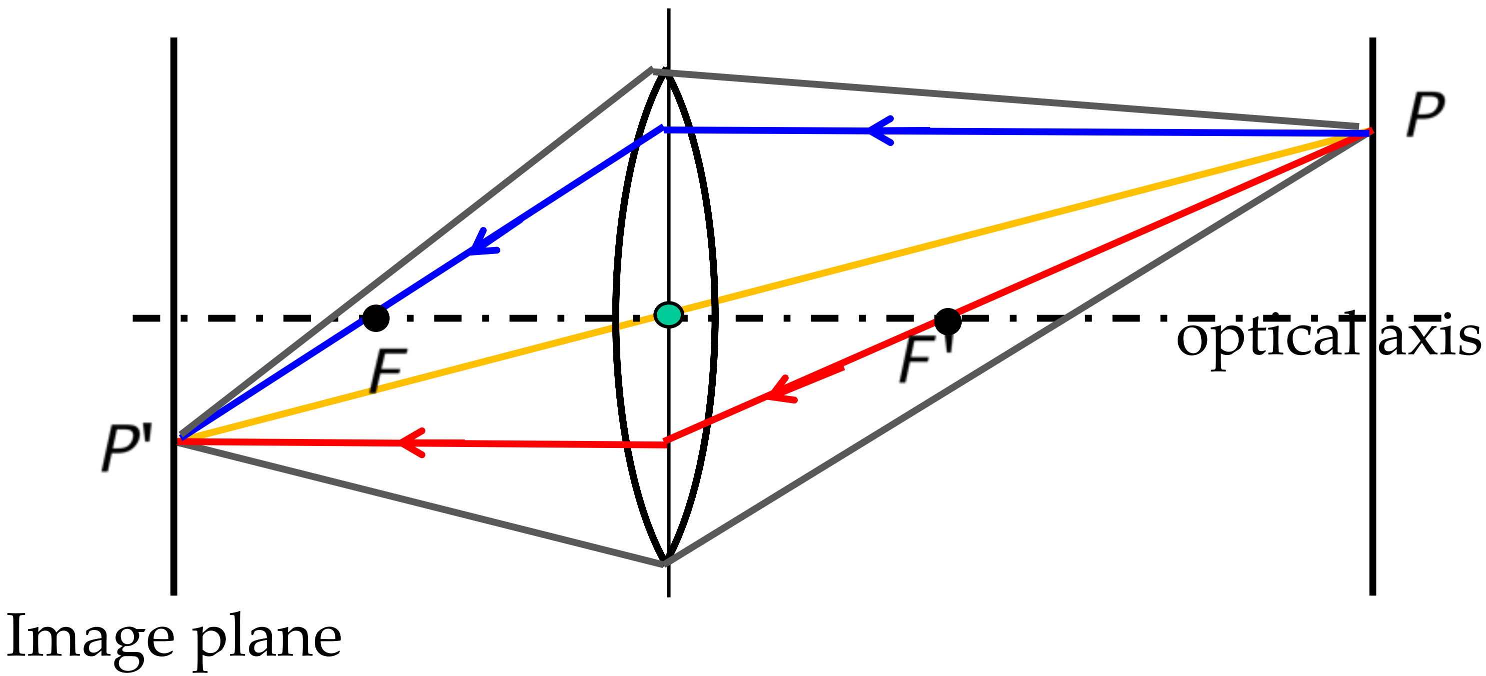


An ideal lens realizes the same projection as a pinhole but **gathers much more light**!

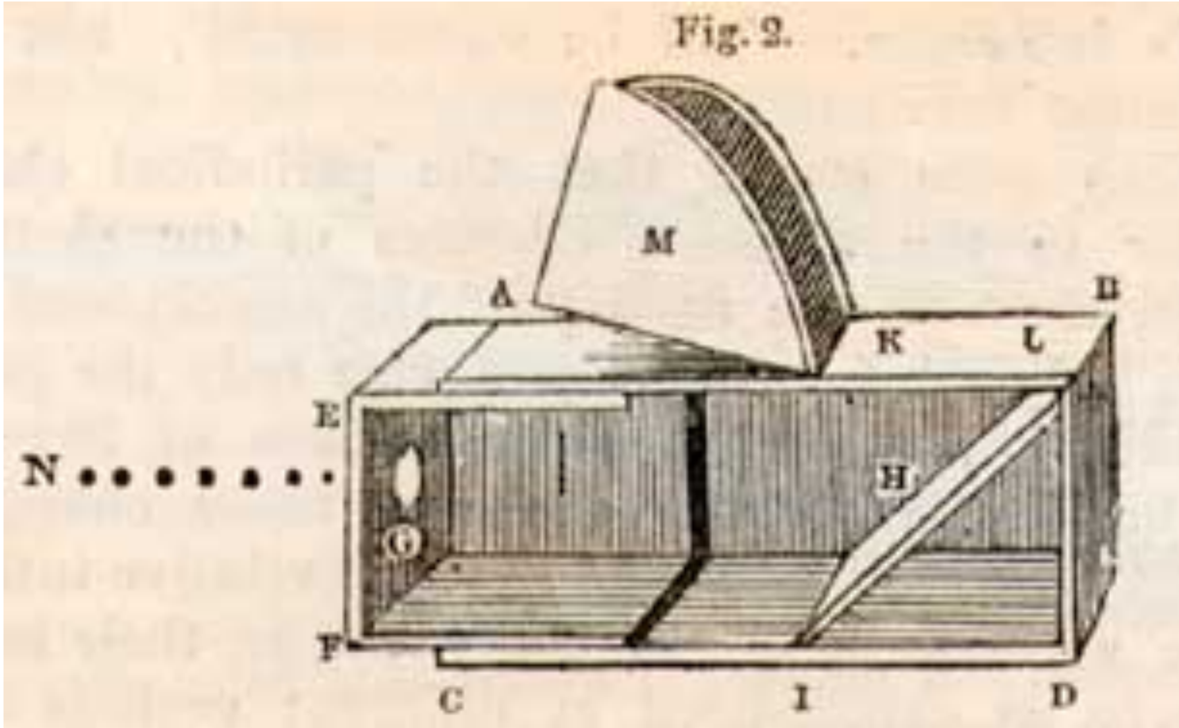
THIN LENS PROPERTIES

- An incident ray which passes through the center of the lens will in effect continue in the same direction that it had when it entered the lens.
- Any incident ray traveling parallel to the optical axis, will refract and travel through the focal point on the opposite side of the lens.
- Any incident ray traveling through the focal point on the way to the lens will be refracted and travel parallel to the principal axis.

- All rays emanating from P and entering the lens will converge at P’

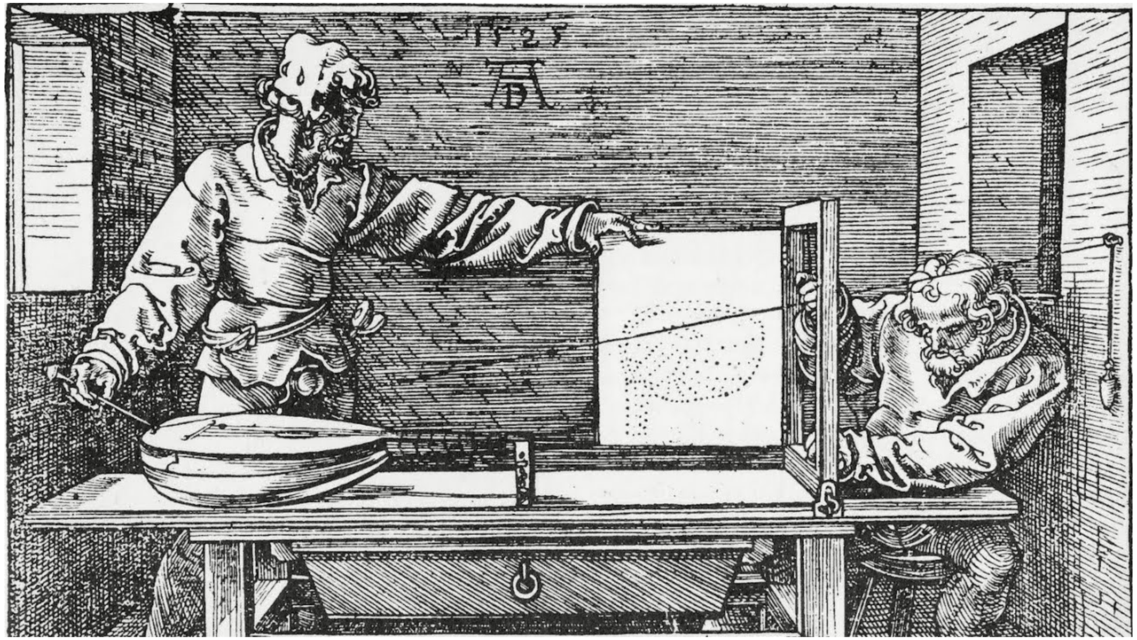


CAMERA OBSCURA



- Used by painters since the Renaissance to produce perspective projections.
- Direct ancestors to the first film cameras.

DURER 1471-1528



He clearly knew all about the perspective transform!

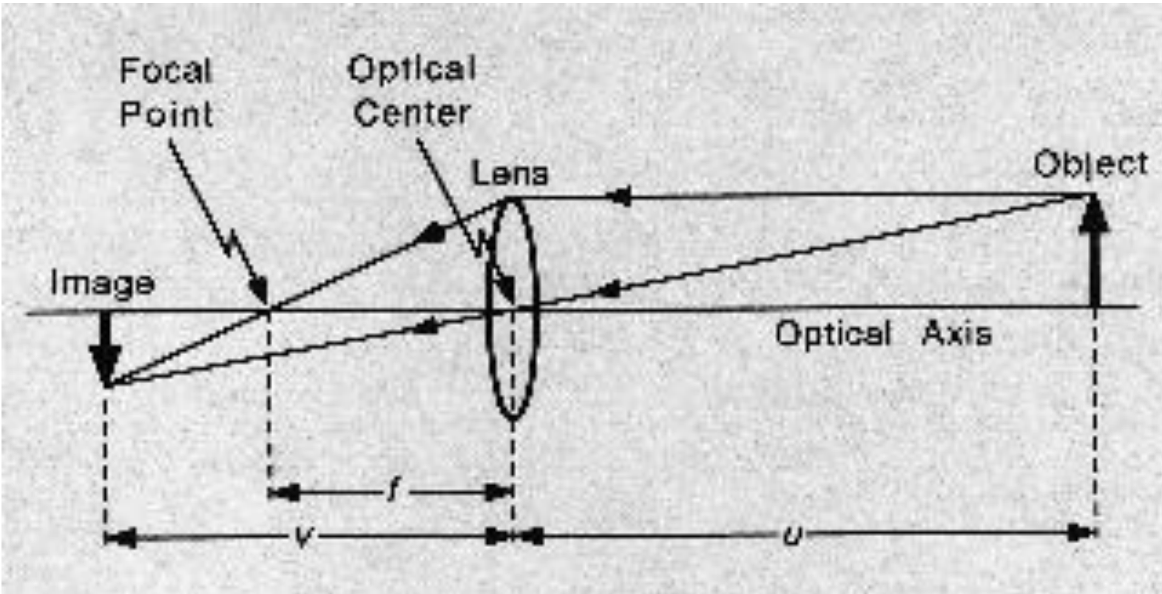
SHIFTING PERSPECTIVE

Buddha cutting his hair, 8th c.

China, 8th century:

- The focal point moves from one part of the image to the other.
- The characters are always seen at eye-level as the picture is unrolled.

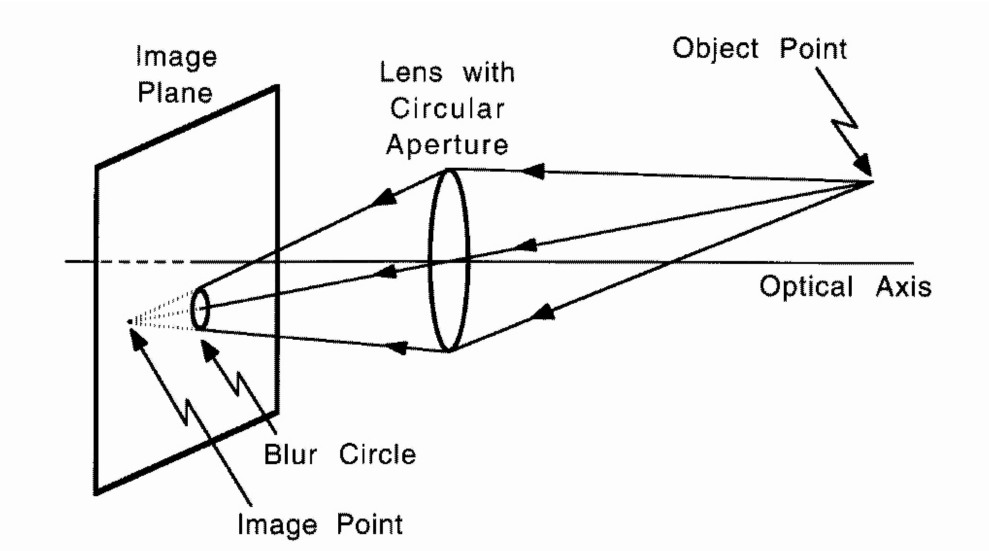
THIN LENS EQUATION



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

→ Lens with focal distance f equivalent to pinhole camera with similar focal distance but larger aperture.

DEPTH OF FIELD vs APERTURE



Large Aperture:

- Large blur circles
- Shallow depth of field

Small Aperture:

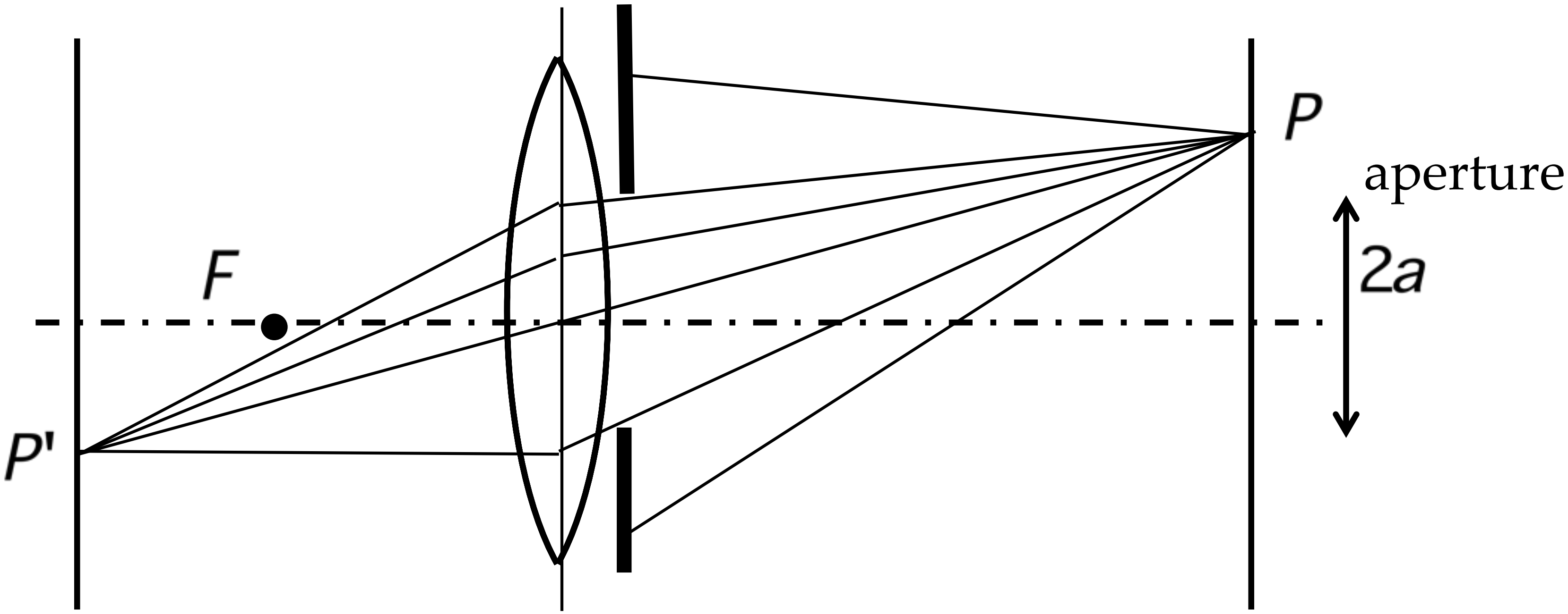
- Low intensity
- Long exposure time

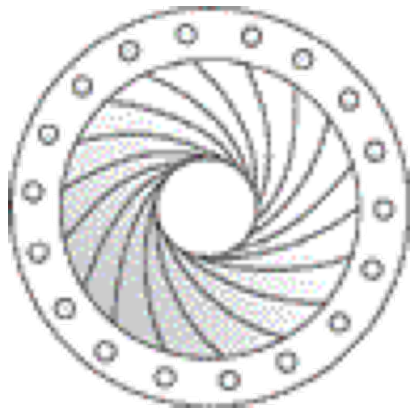
DEPTH OF FIELD

- Range of object distances (d-d’) over which the image is sufficiently well focused.
- Range for which blur circle is less than the resolution of the sensor.

Small focal length —> Large depth of field.

APERTURE

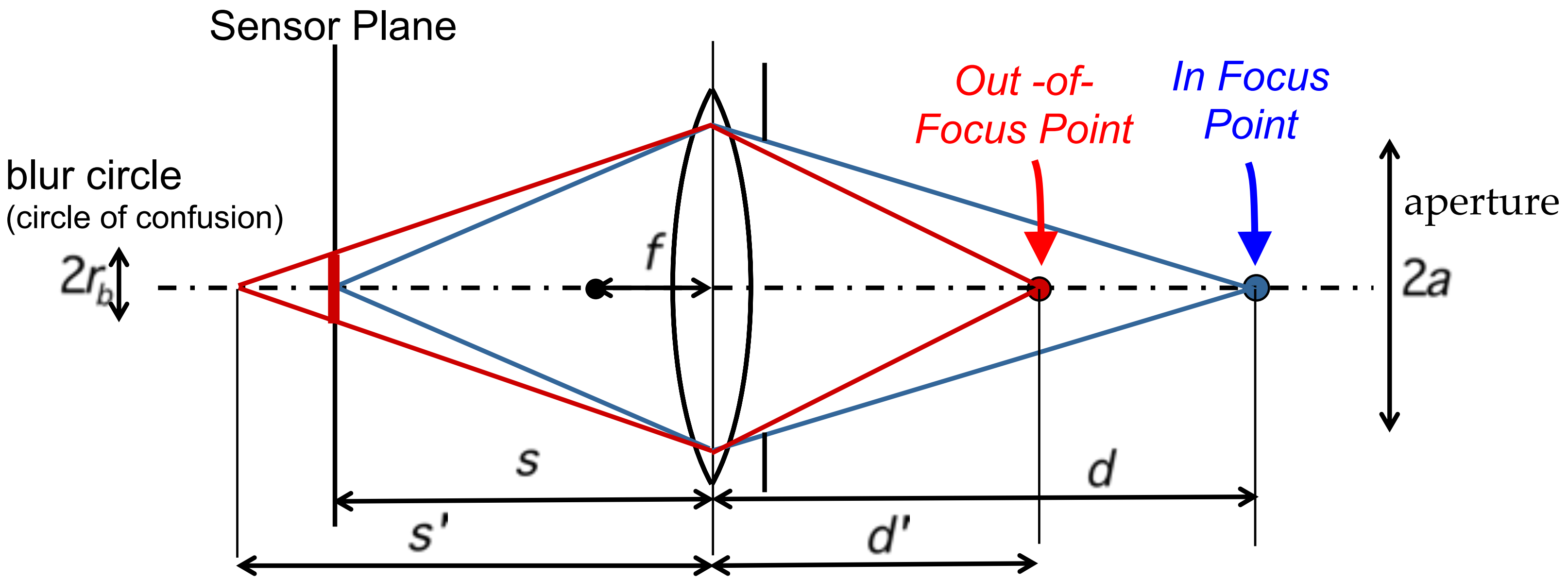




iris diaphragm

Diameter d of the lens that is exposed to light.

BLUR CIRCLE



- Simple geometry:

$$r_b = \frac{a}{s'}|s' - s|$$

- Thin lens equation:

$$\frac{1}{d} + \frac{1}{s} = \frac{1}{f} \Rightarrow s = \frac{df}{d-f}$$

$$\frac{1}{d'} + \frac{1}{s'} = \frac{1}{f} \Rightarrow s' = \frac{d'f}{d'-f}$$

$$\boxed{(s' - s) = \frac{f}{(d' - f)} \frac{f}{(d - f)}(d - d')}$$

CHANGING APERTURE

f/11 1/30sec

Small aperture, long exposure.

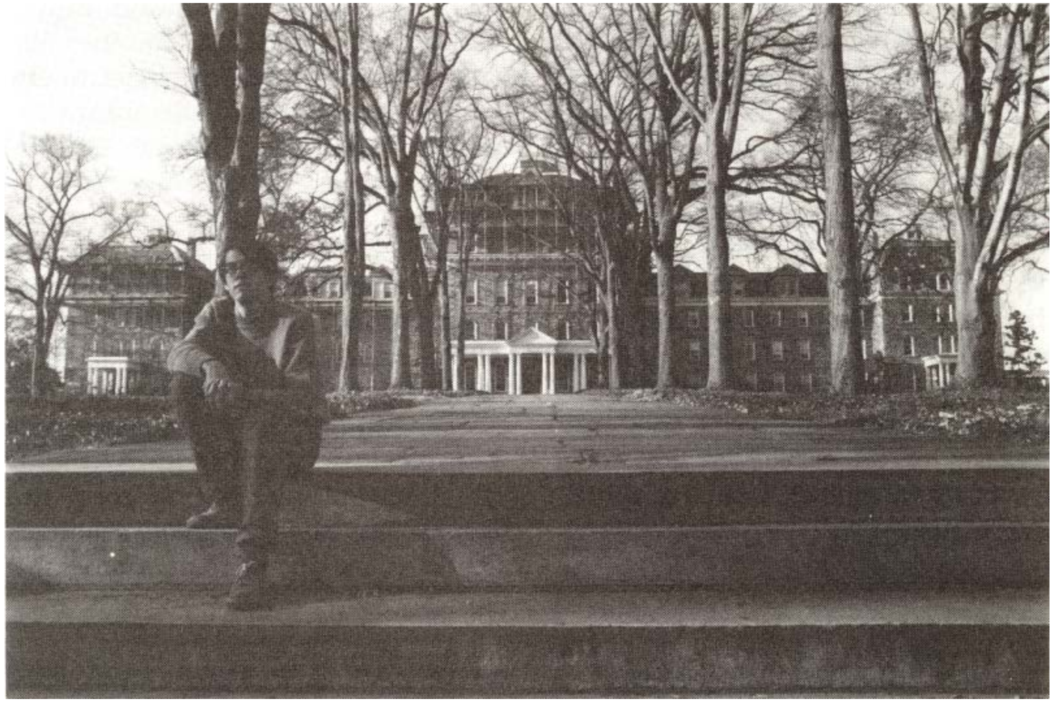
f/2.8 1/500sec

Large aperture, short exposure.

$$r_b = \frac{a}{s'} \left| \frac{f^2}{(d'-f)(d-f)} \right|$$

$$\text{Small } a \rightarrow \text{Small } r_b$$

CHANGING FOCAL LENGTH



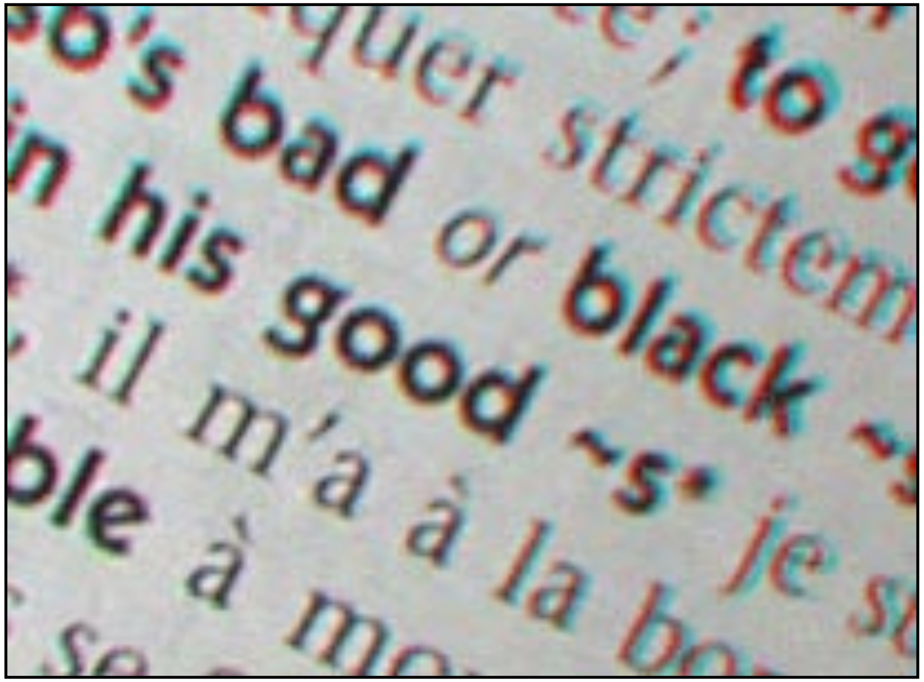
Wide field of view
(small f)

Narrow field of view
(large f)

$$r_b = \frac{a}{s'}\left|\frac{f^2}{(d'-f)(d-f)}\right|$$

$$\text{Small } f \rightarrow \text{Small } r_b$$

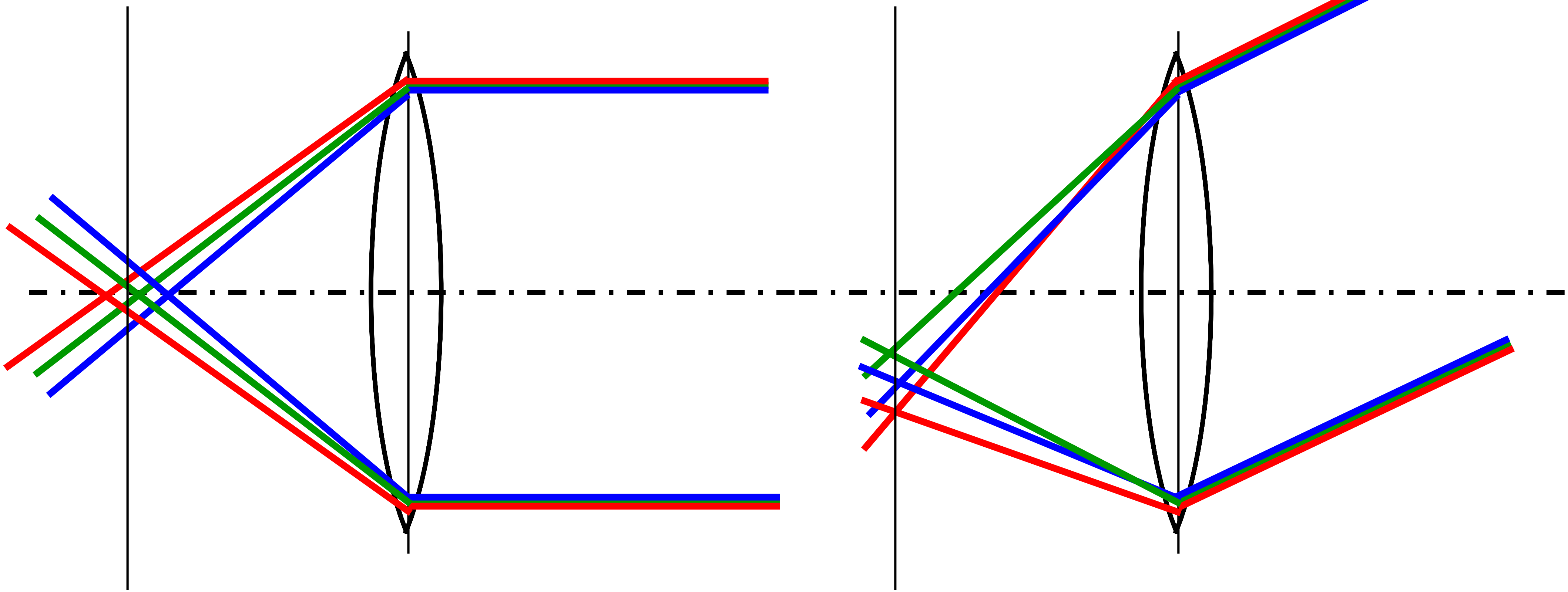
DISTORTIONS



The lens is not exactly a “thin lens:”

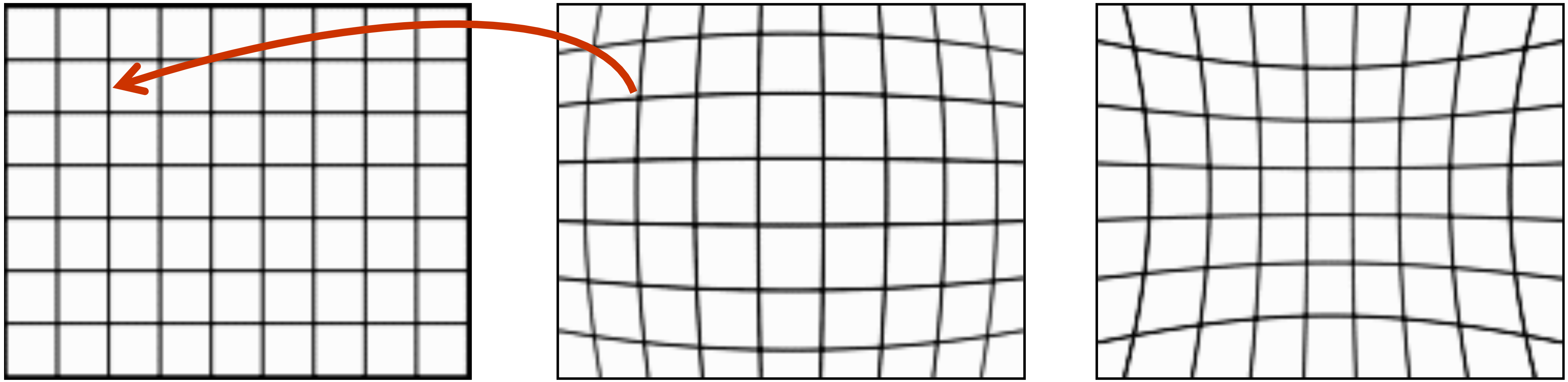
- Different wave lengths refracted differently
- Barrel Distortion

CHROMATIC ABERRATION



Different wavelengths are refracted differently.

RADIAL LENS DISTORTIONS

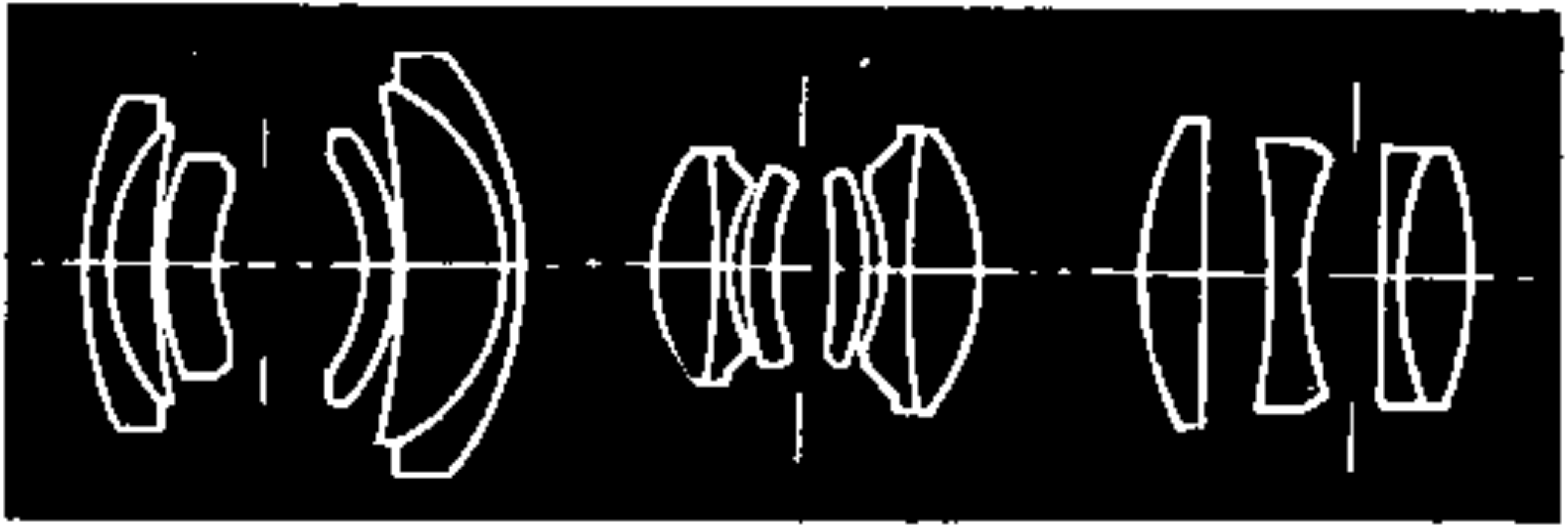


No Distortion | Barrel Distortion | Pincushion Distortion

Radial distance from Image Center:

$$r_u = r_d + k_1 r_d^3$$

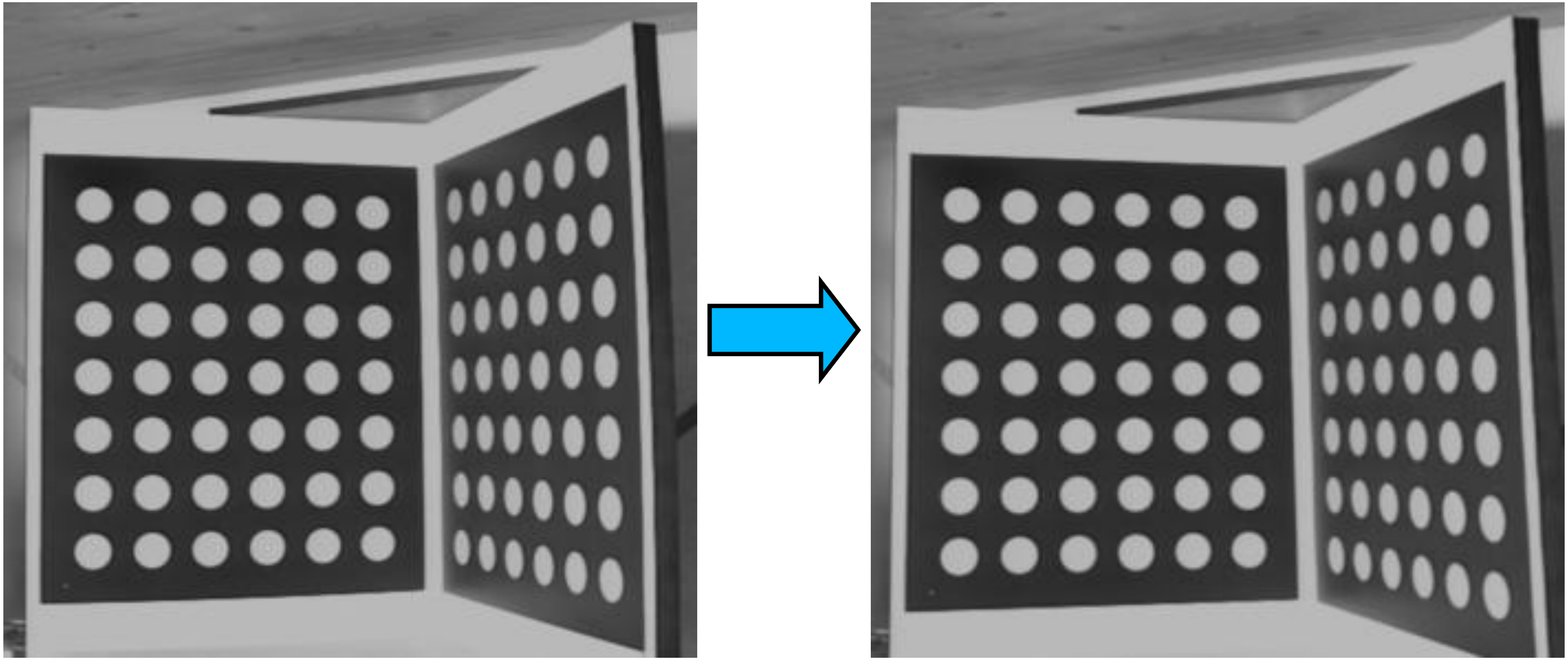
LENS SYSTEMS



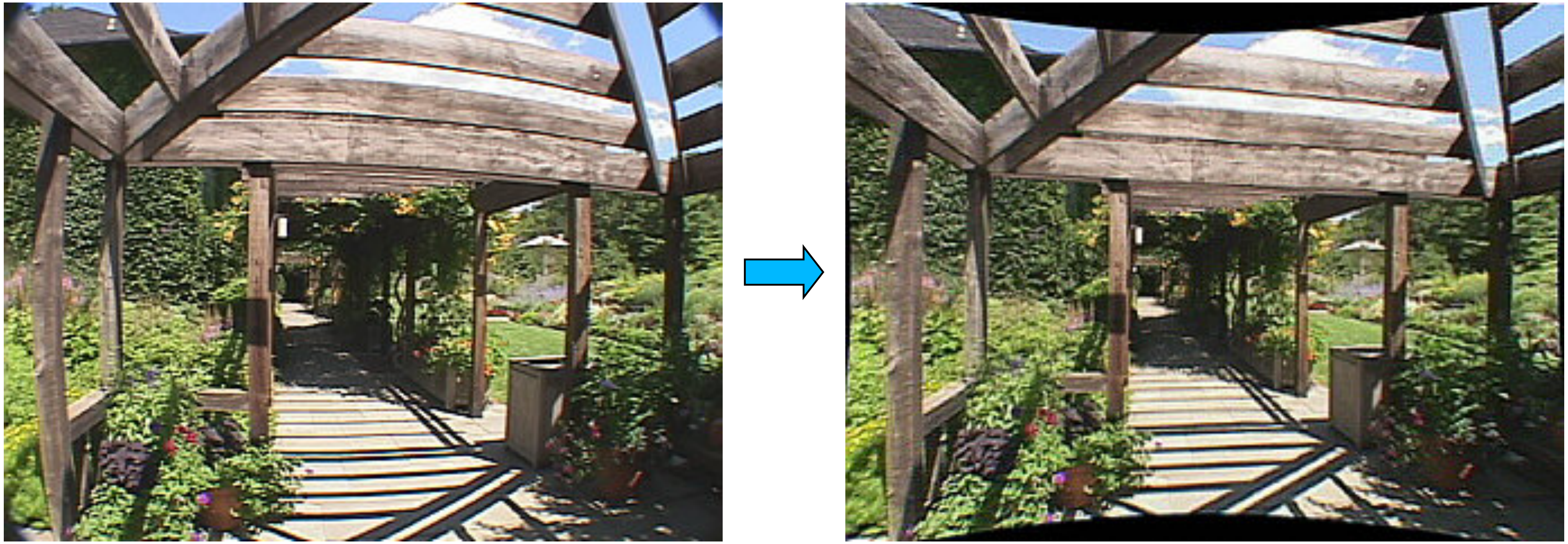
Aberrations can be minimized by aligning several lenses with well chosen

- Shapes,
- Refraction indices.

UNDISTORTING



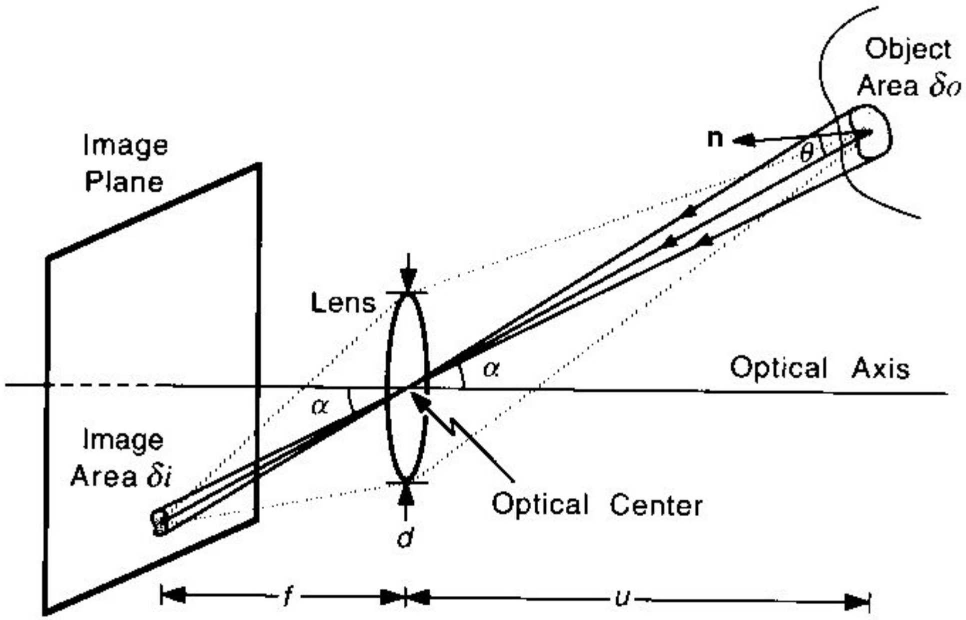
UNDISTORTING



Once the image is undistorted, the camera projection can be formulated as a projective transform.

→ The pinhole camera model applies.

RADIOMETRY



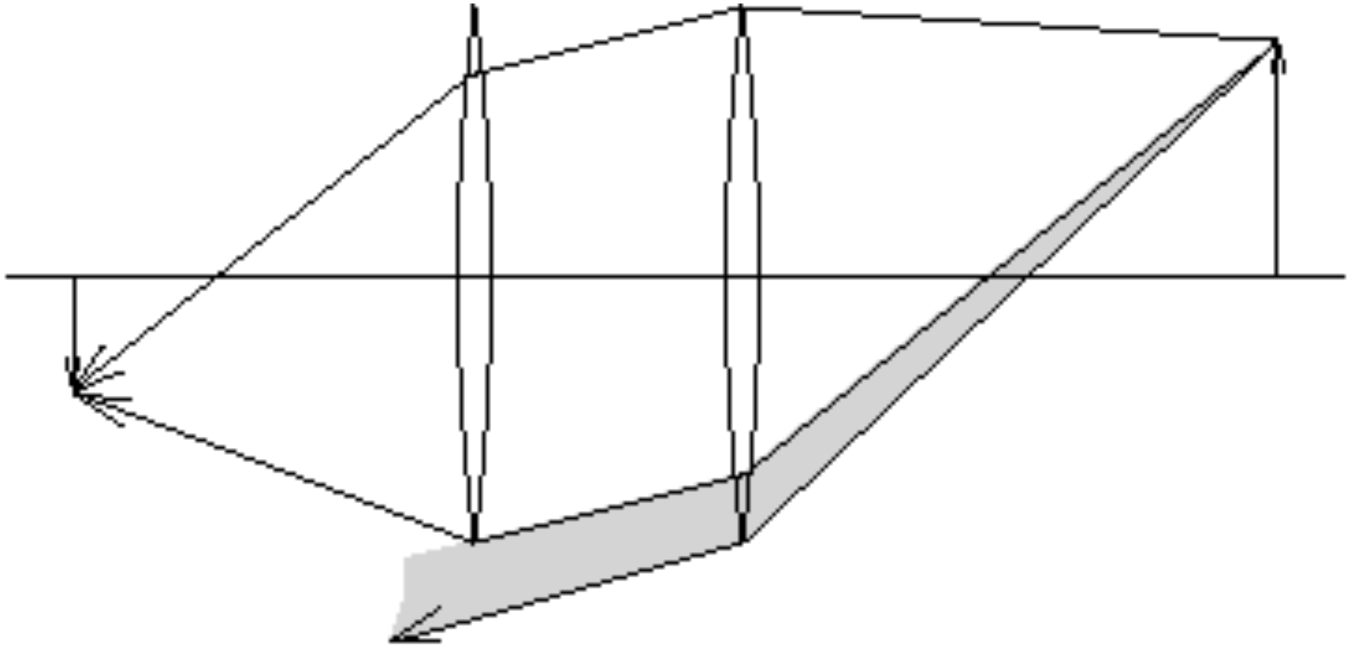
Scene Radiance: Amount of light radiation from a surface point (Watt / m2 / Steradian)

Image Irradiance: Amount of light incident at the image of the surface point. (Watt / m2)

Fundamental Radiometric Equation:

$$\mathrm{Irr} = \frac{\pi}{4}\left(\frac{d}{f}\right)^2 \cos^4(a)\mathrm{Rad}$$

VIGNETTING

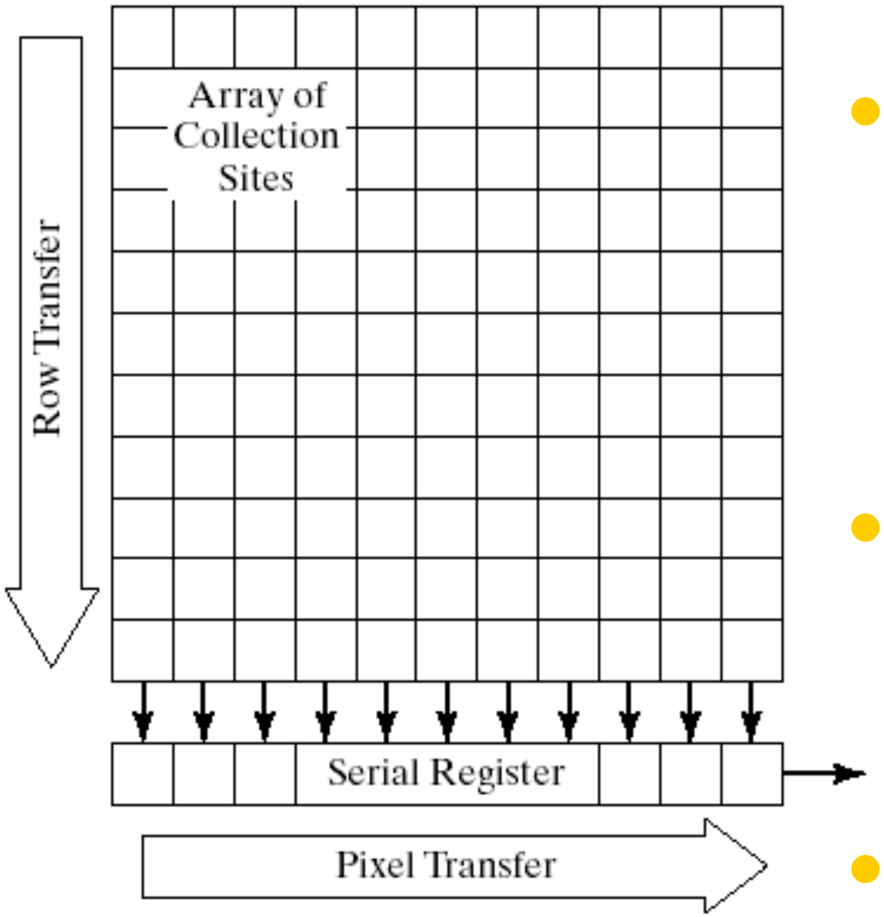


Images can get darker towards their edges because some of the light does not go through all the lenses.

DE VIGNETTING

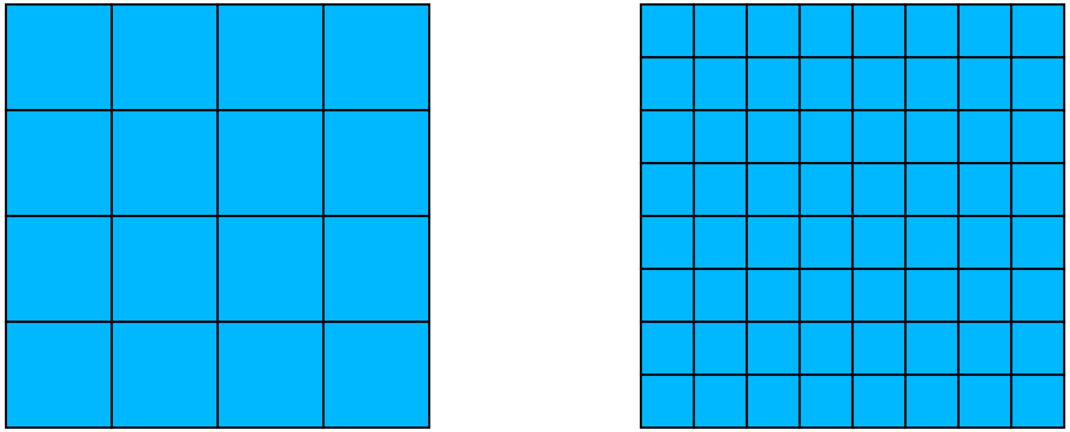
Y. Zheng, S. Lin, and S.B. Kang, CVPR'08

SENSOR ARRAY



- Photons free up electrons that are then captured by a potential well.
- Charges are transferred row by row wise to a register.
- Pixel values are read from the register.

SENSING



Conversion of the "optical image" into an "electrical image":

$$
\begin{aligned}
E(x,y) &= \int_{t_0}^{t_1}\int_0^{\Lambda} \mathrm{Irr}(x,y,t,\lambda)s(\lambda)dtd\lambda \\
I(m,n) &= \mathrm{Quantize}\left(\int_{x_0}^{x_1}\int_{y_0}^{y_1} E(x,y)dxdy\right)
\end{aligned}
$$

→ Quantization in

- Time
- Space

IN SHORT

- Camera geometry can be modeled in terms of the pinhole camera model, which is linear in projective space.
- Image radiance is roughly proportional to surface radiance and the two can be used interchangeably for our purposes.