Computer Vision

CS308

Feng Zheng
SUSTech CS Vision Intelligence and Perception

Week 2

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Content

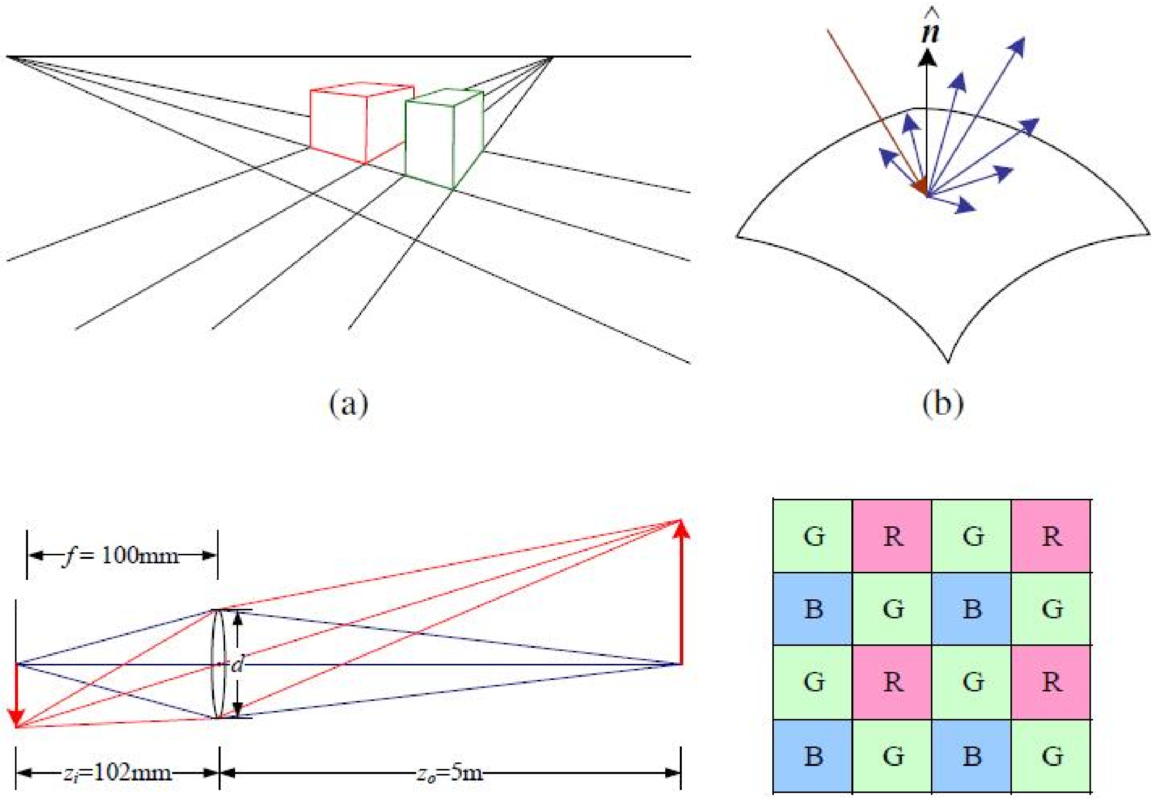
- Geometric primitives and transformations
- Projections
- Photometric image formation
- The digital camera

Image Formation

3D geometric primitives to 2D geometric primitives

Components of the Image Formation Process

- Image formation process: 3D (real-world) to 2D (matrix)
  - ➢ (a) Perspective projection
  - ➢ (b) Light scattering when hitting a surface
  - ➢ (c) Lens optics
  - ➢ (d) Bayer color filter array



(a) (b) (c) (d)

Geometric primitives and transformations

Geometric Primitives

- 2D points

$$\boldsymbol{x} = (x, y) \in \mathcal{R}^2 \qquad \boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

  - Homogeneous coordinates

$$\tilde{\boldsymbol{x}} = (\tilde{x}, \tilde{y}, \tilde{w}) \in \mathcal{P}^2$$

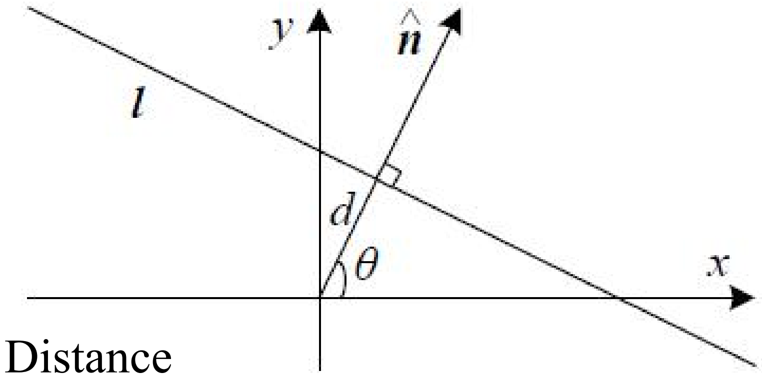
  - Augmented vector

$$\bar{\boldsymbol{x}} = (x, y, 1)$$

  - Relationship

$$\tilde{\boldsymbol{x}} = (\tilde{x}, \tilde{y}, \tilde{w}) = \tilde{w}(x, y, 1) = \tilde{w}\bar{\boldsymbol{x}},$$

Geometric Primitives



- 2D lines

$$\bar{x} \cdot \tilde{l} = \underline{ax + by + c = 0}$$
$$\tilde{l} = (a, b, c)$$

  - Polar coordinates

Direction Distance

$$l = (\hat{n}_x, \hat{n}_y, d) = (\hat{n}, d)$$

✓ The direction (normal vector) is a function of a rotation angle

$$\hat{n} = (\hat{n}_x, \hat{n}_y) = (\cos\theta, \sin\theta)$$

- Advantageous

  - Intersection of two lines
  - Line joining two points

Cross product operation

$$\tilde{x} = \tilde{l}_1 \times \tilde{l}_2 \qquad \tilde{l} = \tilde{x}_1 \times \tilde{x}_2$$

Geometric Primitives

- 3D points

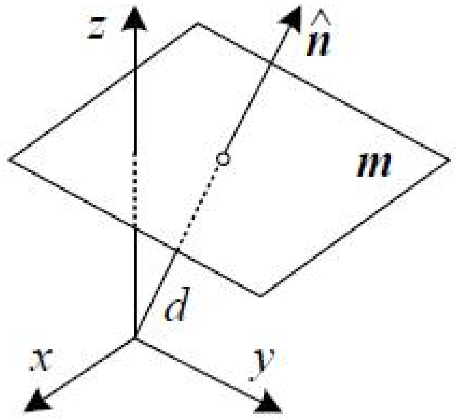
$$x = (x, y, z) \in \mathcal{R}^3 \qquad \tilde{x} = (\tilde{x}, \tilde{y}, \tilde{z}, \tilde{w}) \in \mathcal{P}^3$$

$$\bar{x} = (x, y, z, 1) \qquad \tilde{x} = \tilde{w}\bar{x}$$

- 3D planes

$$\bar{x} \cdot \tilde{m} = \underline{ax + by + cz + d = 0}$$

$$m = (\hat{n}_x, \hat{n}_y, \hat{n}_z, d) = (\hat{n}, d)$$

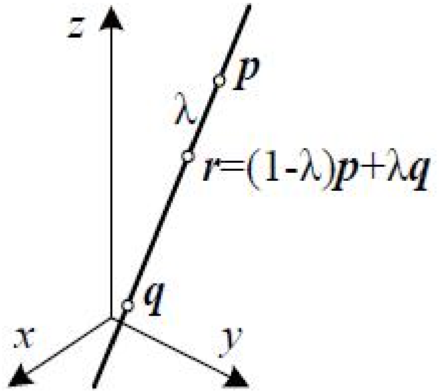


  - ➢ The direction (normal vector) is a function of two rotation angles

$$\hat{n} = (\cos\theta\cos\phi, \sin\theta\cos\phi, \sin\phi)$$

- 3D lines

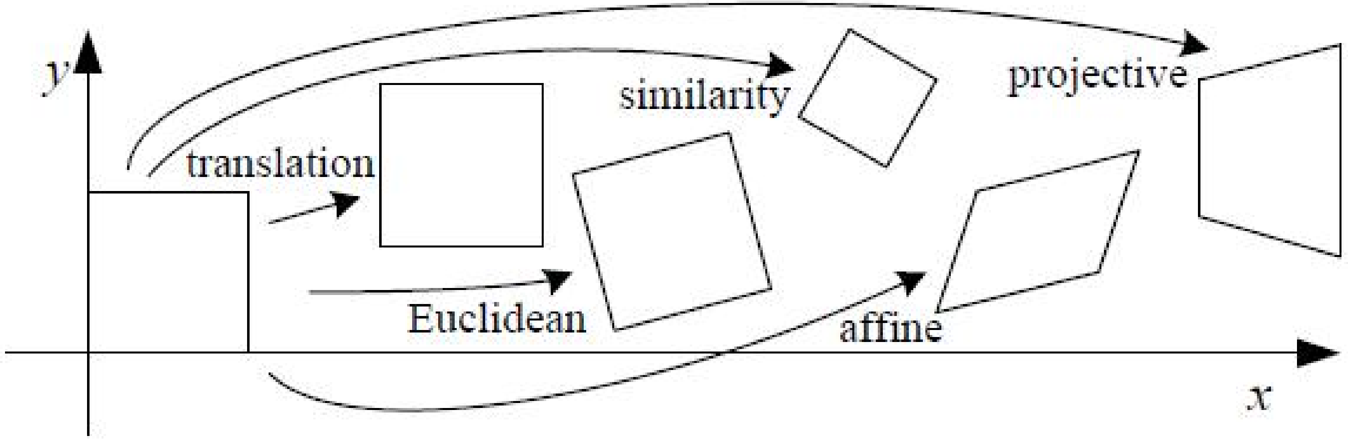
$$r = (1 - \lambda)p + \lambda q$$



Transformations

- 2D transformations
  - Translation $x' = x + t = \begin{bmatrix} I & t \end{bmatrix} \bar{x}$ $\qquad \bar{x}' = \begin{bmatrix} I & t \\ 0^T & 1 \end{bmatrix} \bar{x}$
  - Rotation + translation $x' = Rx + t = \boxed{\begin{bmatrix} R & t \end{bmatrix} \bar{x}}$

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

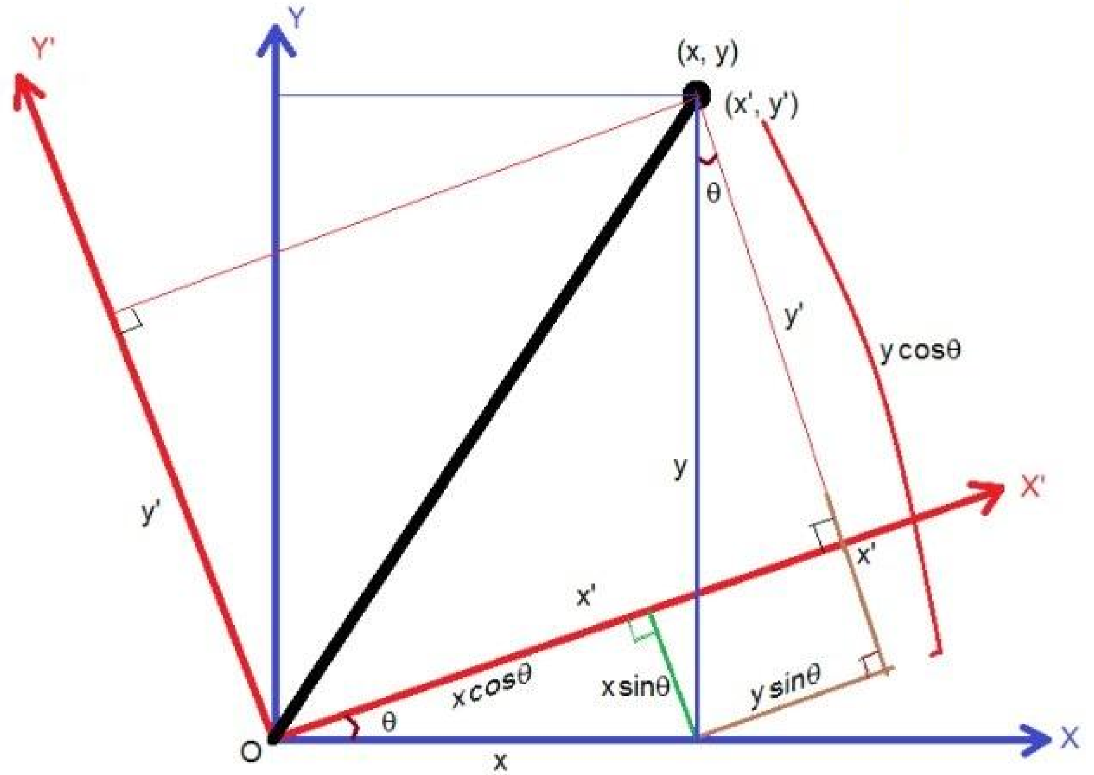


Transformations

- Rotation matrix
  - After the rectangular coordinate system is rotated by a certain angle
  - The relationship between the new and the old coordinate systems

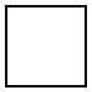
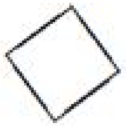
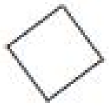
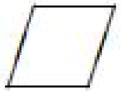
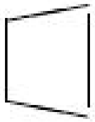
$$x' = x\cos\theta + y\sin\theta$$

$$y' = y\cos\theta - x\sin\theta$$



Transformations (2D-2D)

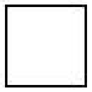
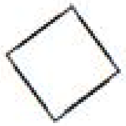
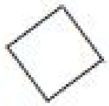
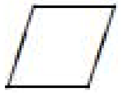
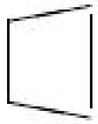
- Hierarchy of 2D coordinate transformations

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\begin{bmatrix} I \mid t \end{bmatrix}_{2\times 3}$	2	orientation	
rigid (Euclidean)	$\begin{bmatrix} R \mid t \end{bmatrix}_{2\times 3}$	3	lengths	
similarity	$\begin{bmatrix} sR \mid t \end{bmatrix}_{2\times 3}$	4	angles	
affine	$\begin{bmatrix} A \end{bmatrix}_{2\times 3}$	6	parallelism	
projective	$\begin{bmatrix} \tilde{H} \end{bmatrix}_{3\times 3}$	8	straight lines	

投影变换

Transformations (3D-3D)

- Hierarchy of 3D coordinate transformations

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\begin{bmatrix} I \mid t \end{bmatrix}_{3\times 4}$	3	orientation	
rigid (Euclidean)	$\begin{bmatrix} R \mid t \end{bmatrix}_{3\times 4}$	6	lengths	
similarity	$\begin{bmatrix} sR \mid t \end{bmatrix}_{3\times 4}$	7	angles	
affine	$\begin{bmatrix} A \end{bmatrix}_{3\times 4}$	12	parallelism	
projective	$\begin{bmatrix} \tilde{H} \end{bmatrix}_{4\times 4}$	15	straight lines	

投影变换

Transformations (3D-2D)

- 3D to 2D projections (what information you want to preserved)
  - ➢ Specify how 3D primitives are projected onto the image plane
  - ➢ Use a linear 3D to 2D projection matrix
- Orthography
  - ➢ Orthographic projection

$$x = [I_{2\times2}|0]\, p$$

$$\underset{\text{2D}}{\tilde{x}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \underset{\text{3D}}{\tilde{p}}$$

  - ➢ Scaled orthography
    - ✓First project the world points onto a local fronto-parallel image plane
    - ✓Then scale this image using regular perspective projection

$$x = [sI_{2\times2}|0]\, p$$

Transformations (3D-2D)

- Perspective
  - The most commonly used projection
  - Points projected onto the image plane by dividing them by their z component

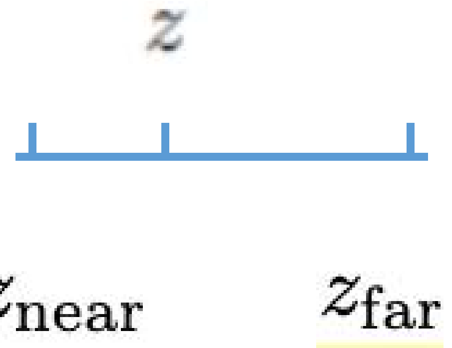
inhomogeneous $$\bar{x} = \mathcal{P}_z(p) = \begin{bmatrix} x/z \\ y/z \\ 1 \end{bmatrix}$$

  - A two-step projection
    - ✓ First project 3D points into normalized device coordinates in the range
    - ✓ Then rescale these coordinates to **integer** pixel coordinates

the near and far z clipping planes

$$z_{\text{range}} = z_{\text{far}} - z_{\text{near}}$$

$$\tilde{x} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -z_{\text{far}}/z_{\text{range}} & z_{\text{near}}z_{\text{far}}/z_{\text{range}} \\ 0 & 0 & 1 & 0 \end{bmatrix} \tilde{p}$$



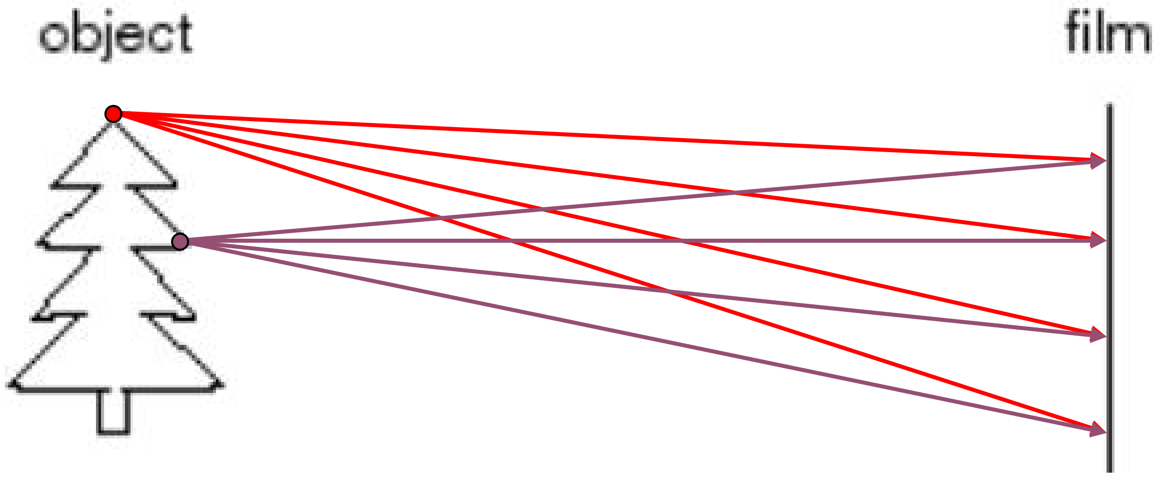
Projections

The Geometry of Image Formation

- Mapping between image and world coordinates
  - Pinhole camera model
  - Projective geometry
    - ✓ Vanishing points and lines
  - Projection matrix

Image Formation

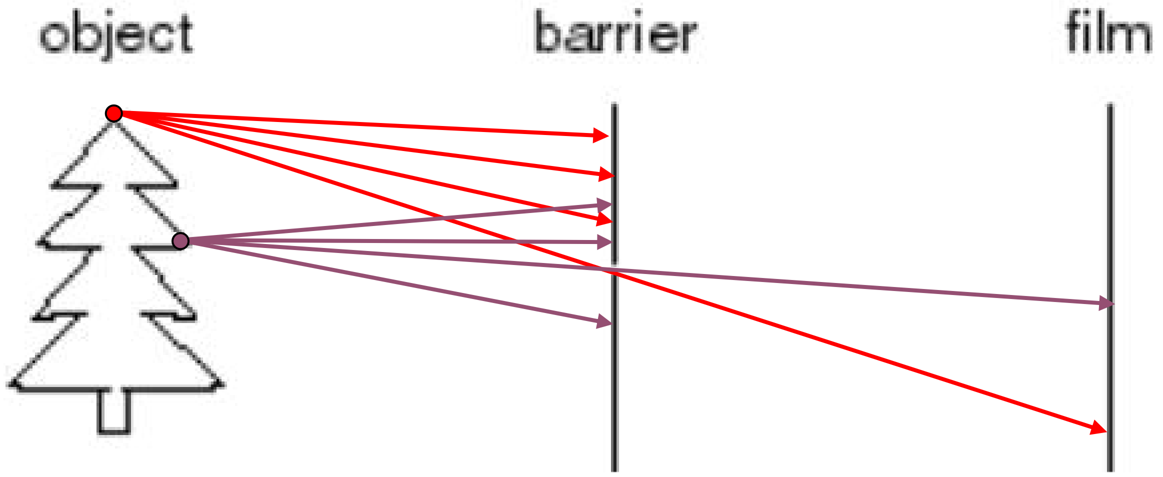
- Let's design a camera
  - Idea 1: put a piece of film in front of an object
  - Do we get a reasonable image?



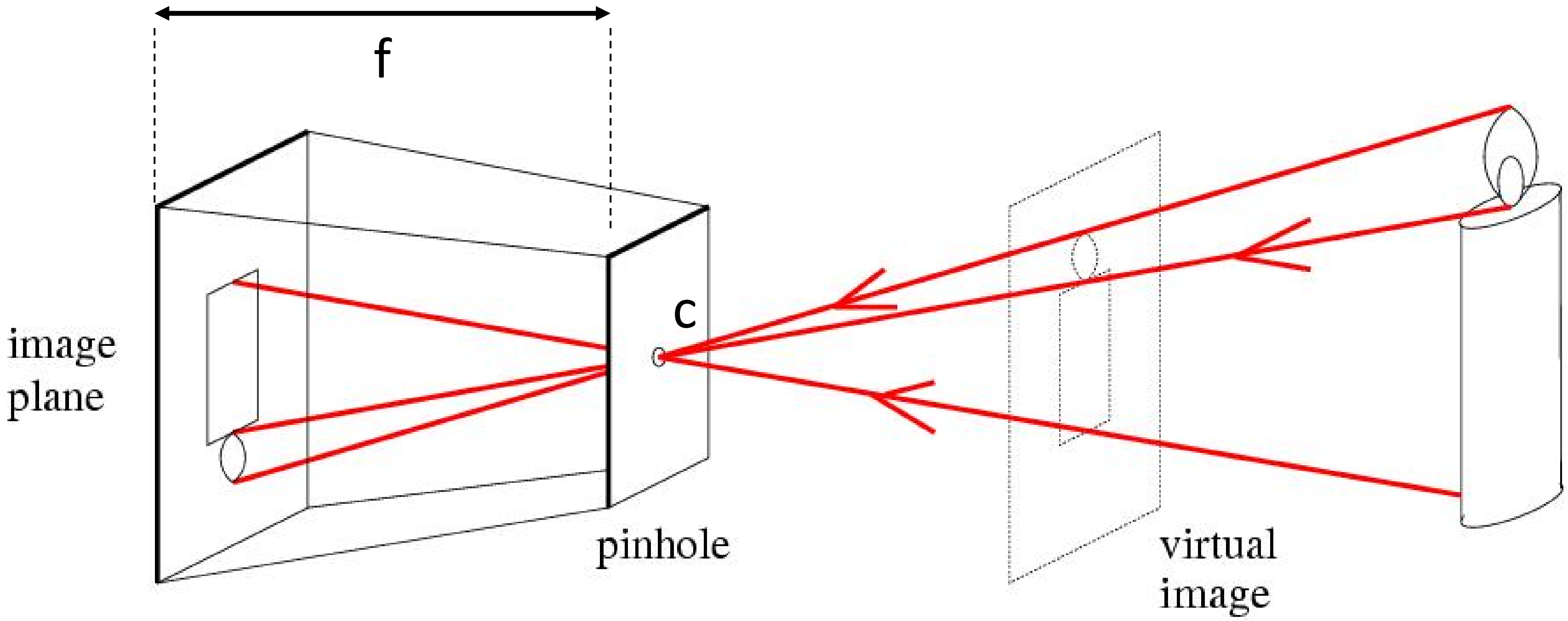
Slide source: Seitz

Pinhole Camera

- Idea 2: add a barrier to block off most of the rays
  - This reduces blurring
  - The opening known as the aperture



Pinhole Camera



f = focal length
c = center of the camera

Camera Obscura: the Pre-Camera

- Known during classical period in China and Greece (e.g. Mo-Ti, China, 470BC to 390BC)

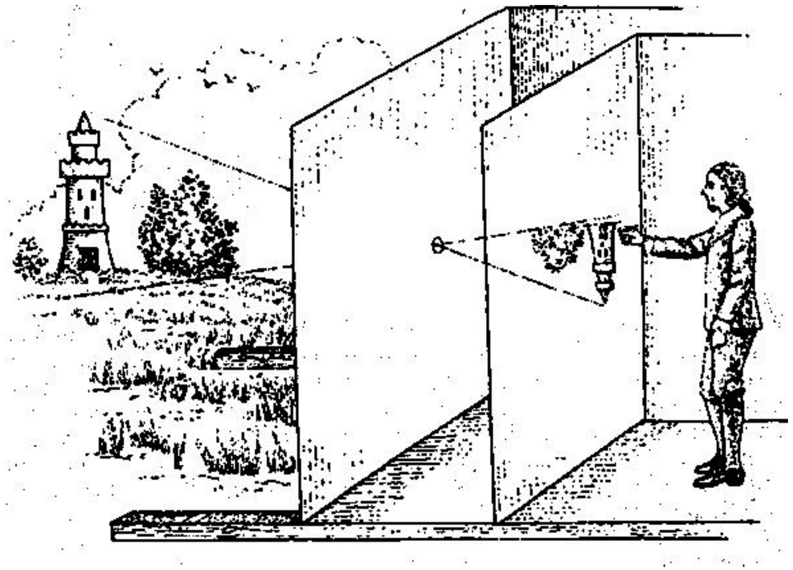


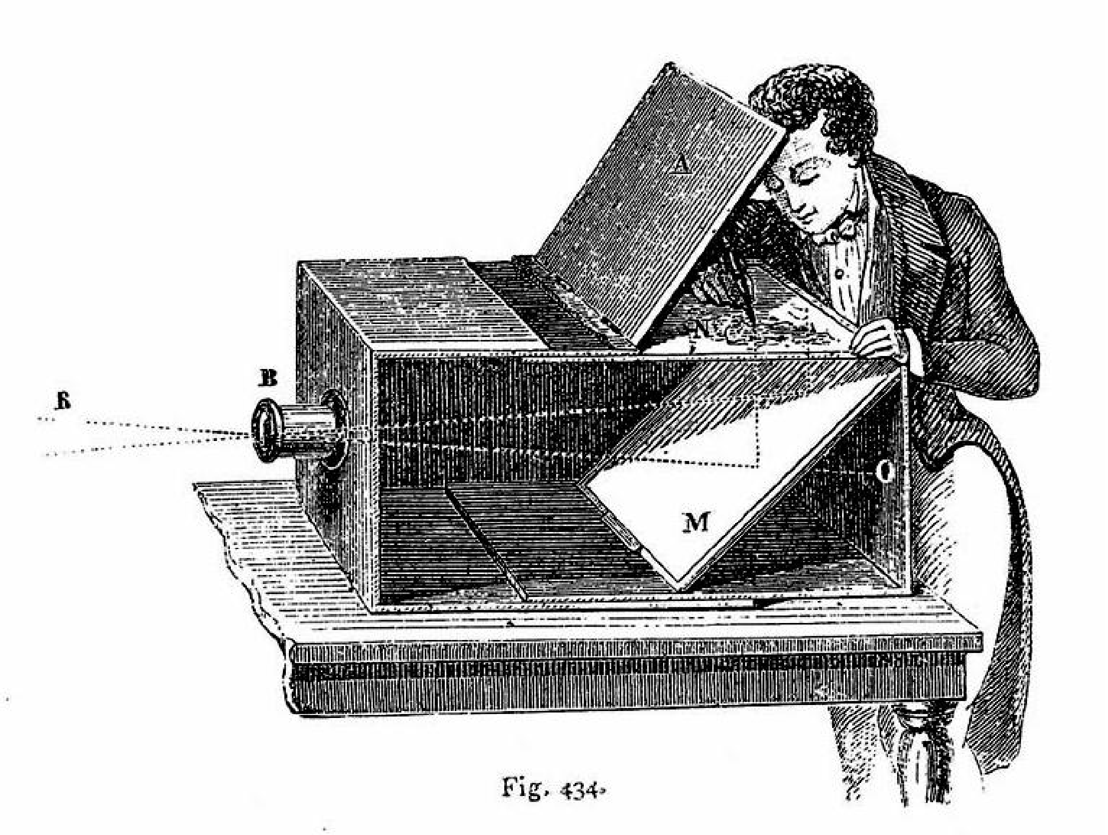
Illustration of Camera Obscura

Freestanding camera obscura at UNC Chapel Hill

Photo by Seth Ilys

“景到，在午有端，与景长。说在端。”

Camera Obscura used for Tracing



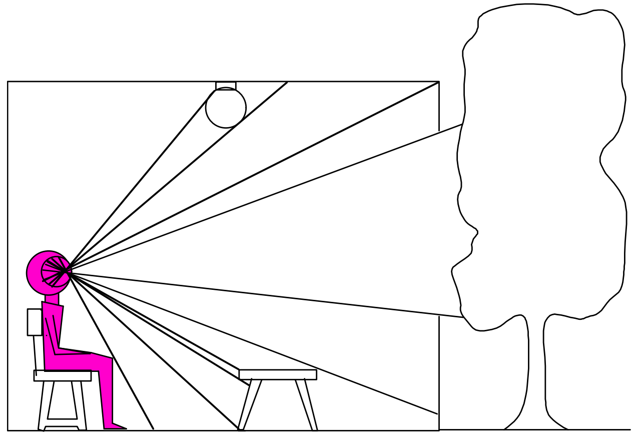
Lens Based Camera Obscura, 1568

Camera and World Geometry

- Questions:
  - How tall is this woman?
  - How high is the camera?
  - What is the camera rotation?
  - What is the focal length of the camera?
  - Which ball is closer?

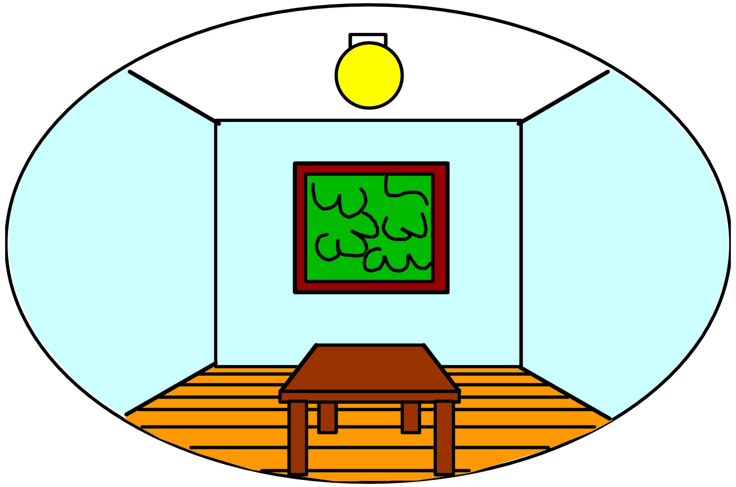
Dimensionality Reduction Machine (3D to 2D)

3D world



Point of observation

2D image

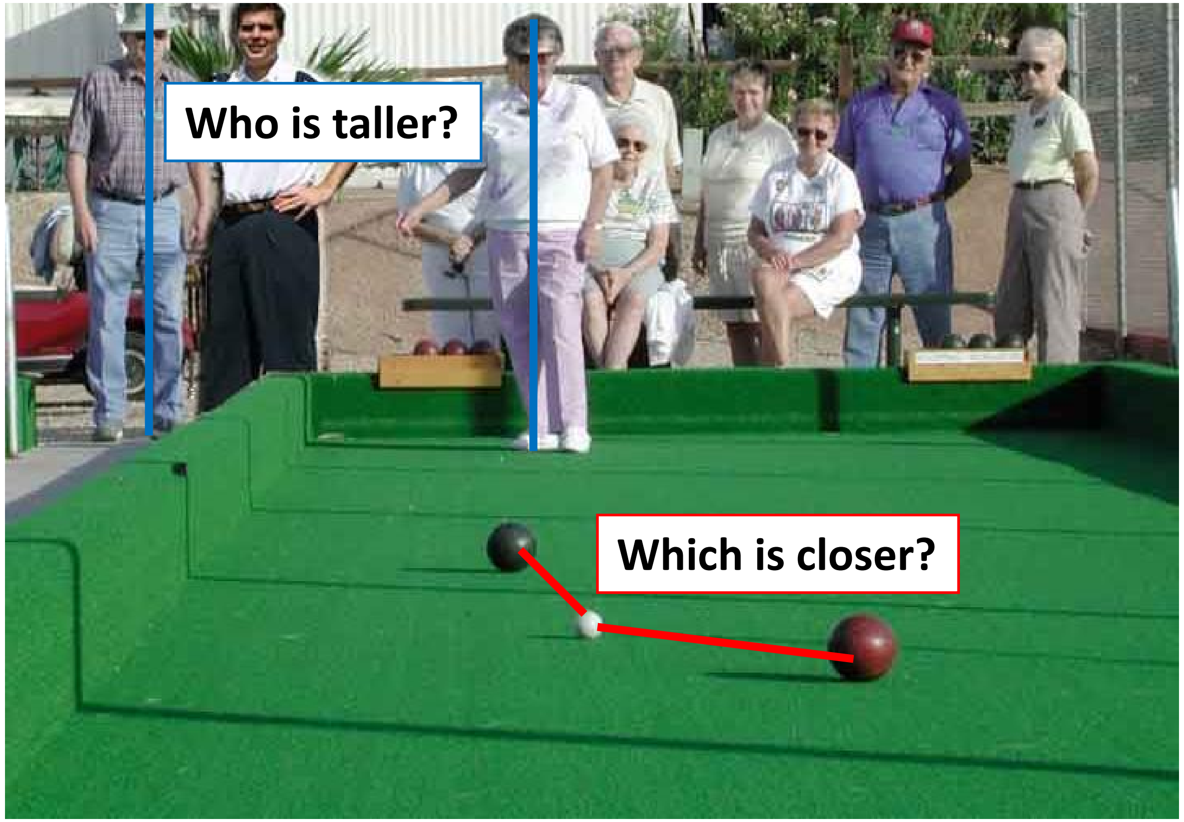


Projection Can Be Tricky…

From an another view, it is totally different

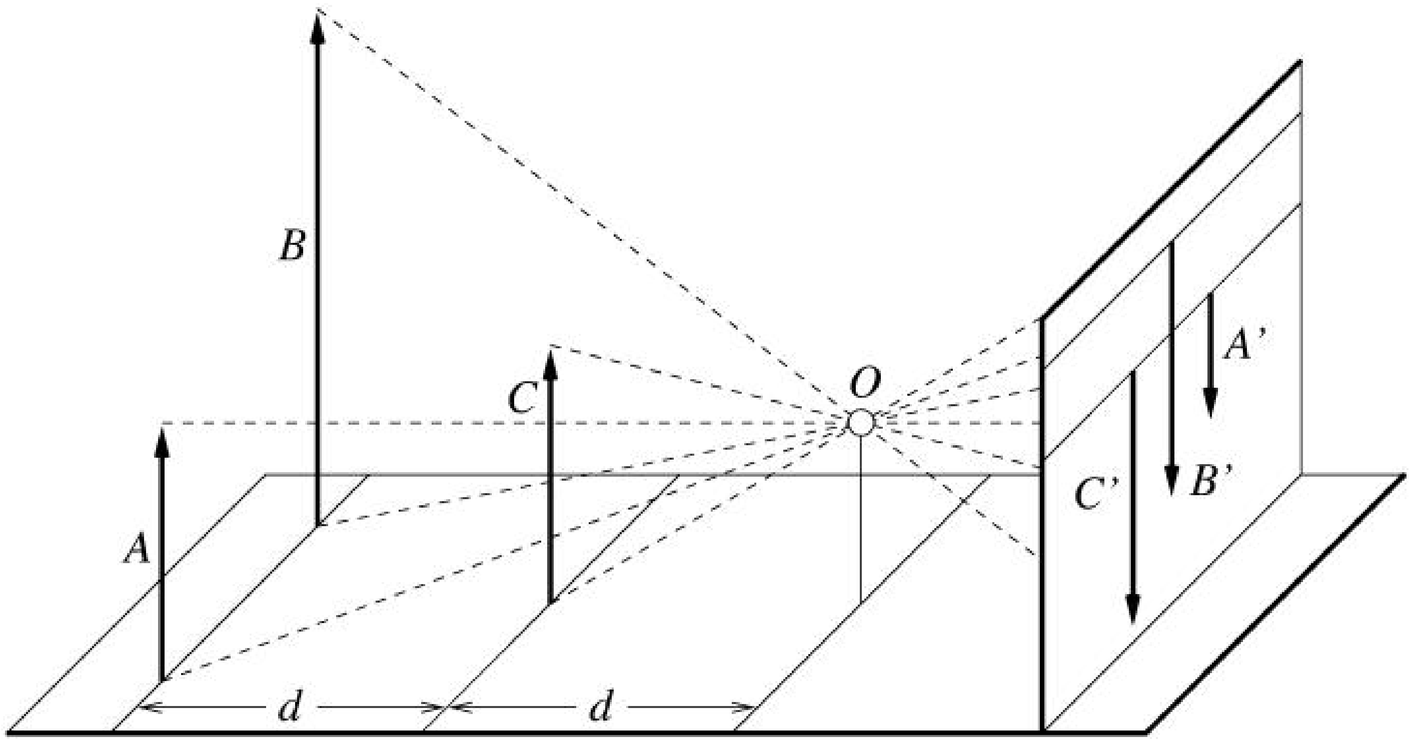
Projective Geometry

- What is lost?
  - ➢ Length



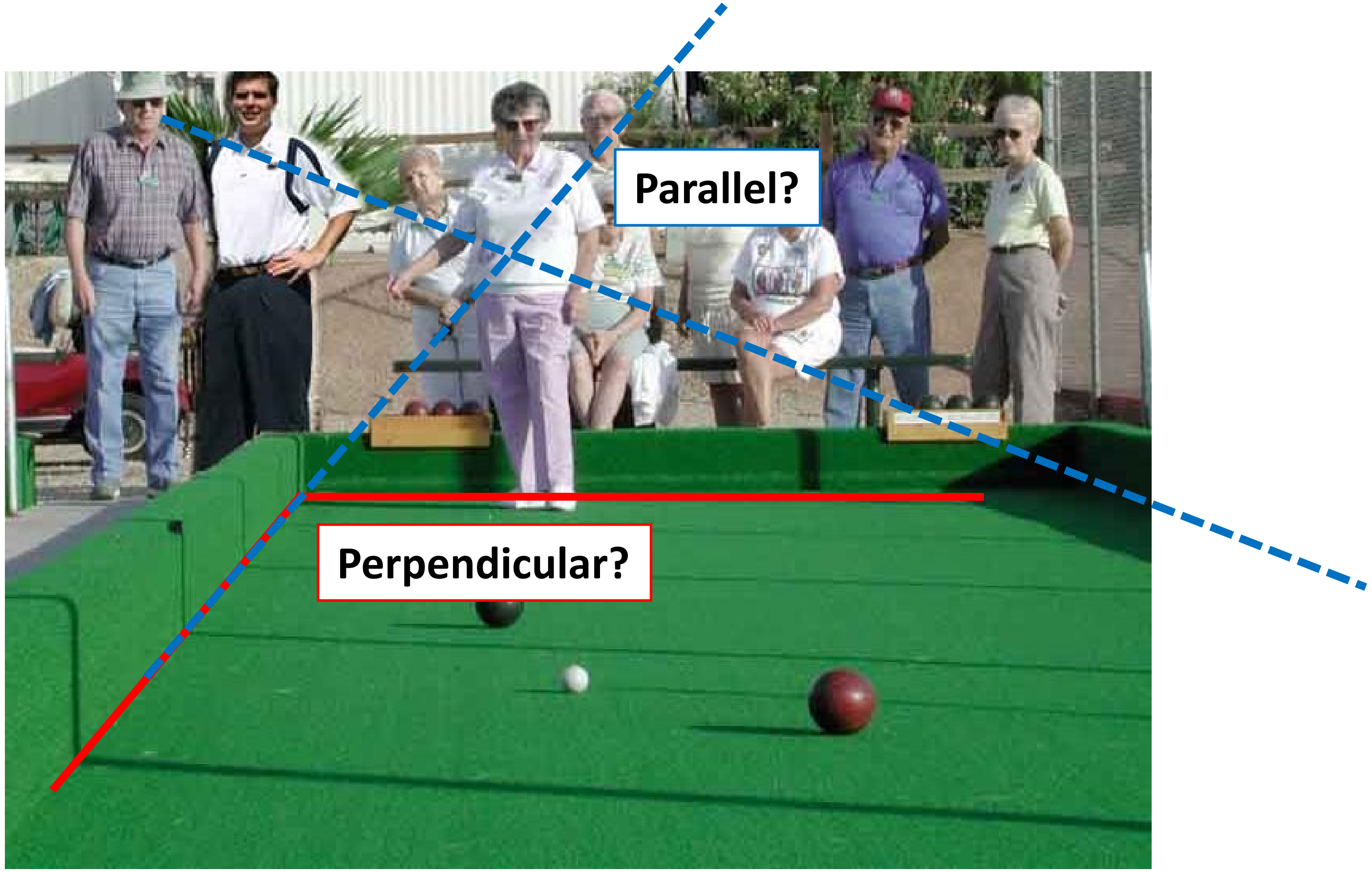
Projective Geometry

- What is lost?
  - Length and area are not preserved



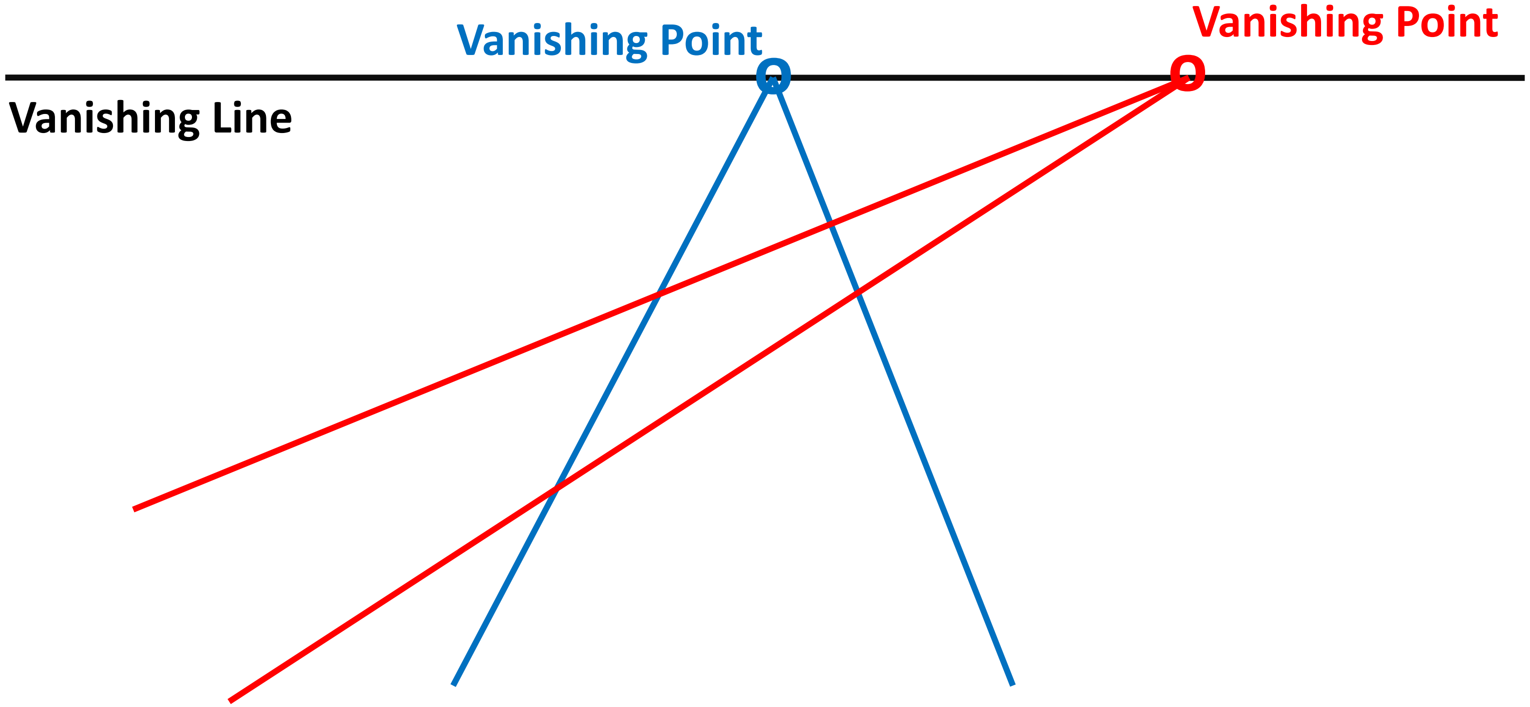
Projective Geometry

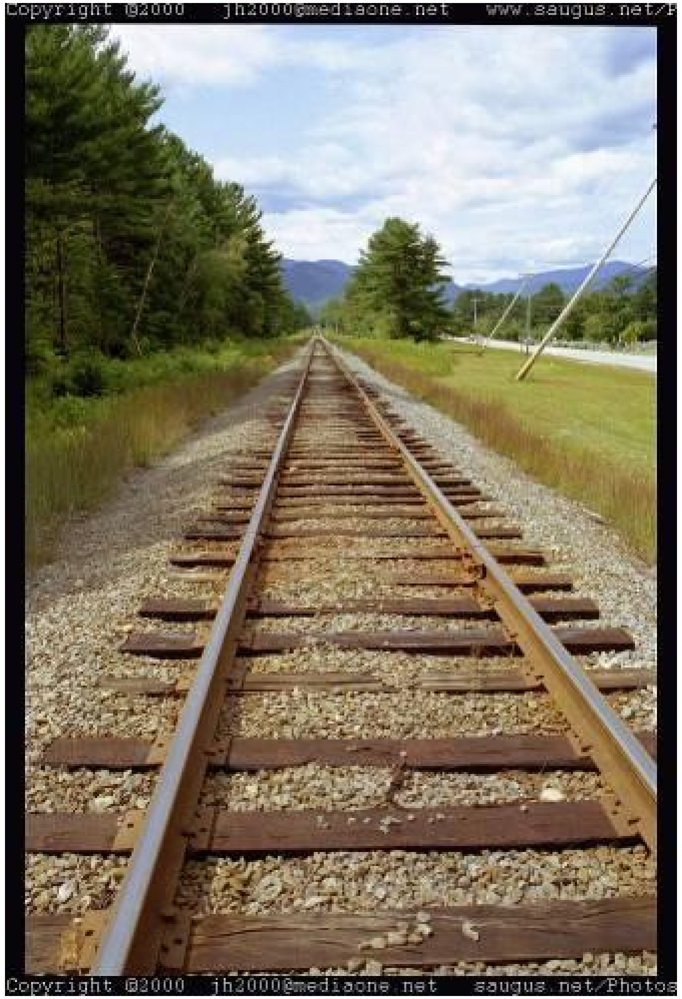
- What is lost?
  - Length
  - Angles
- What is preserved?
  - Straight lines are still straight



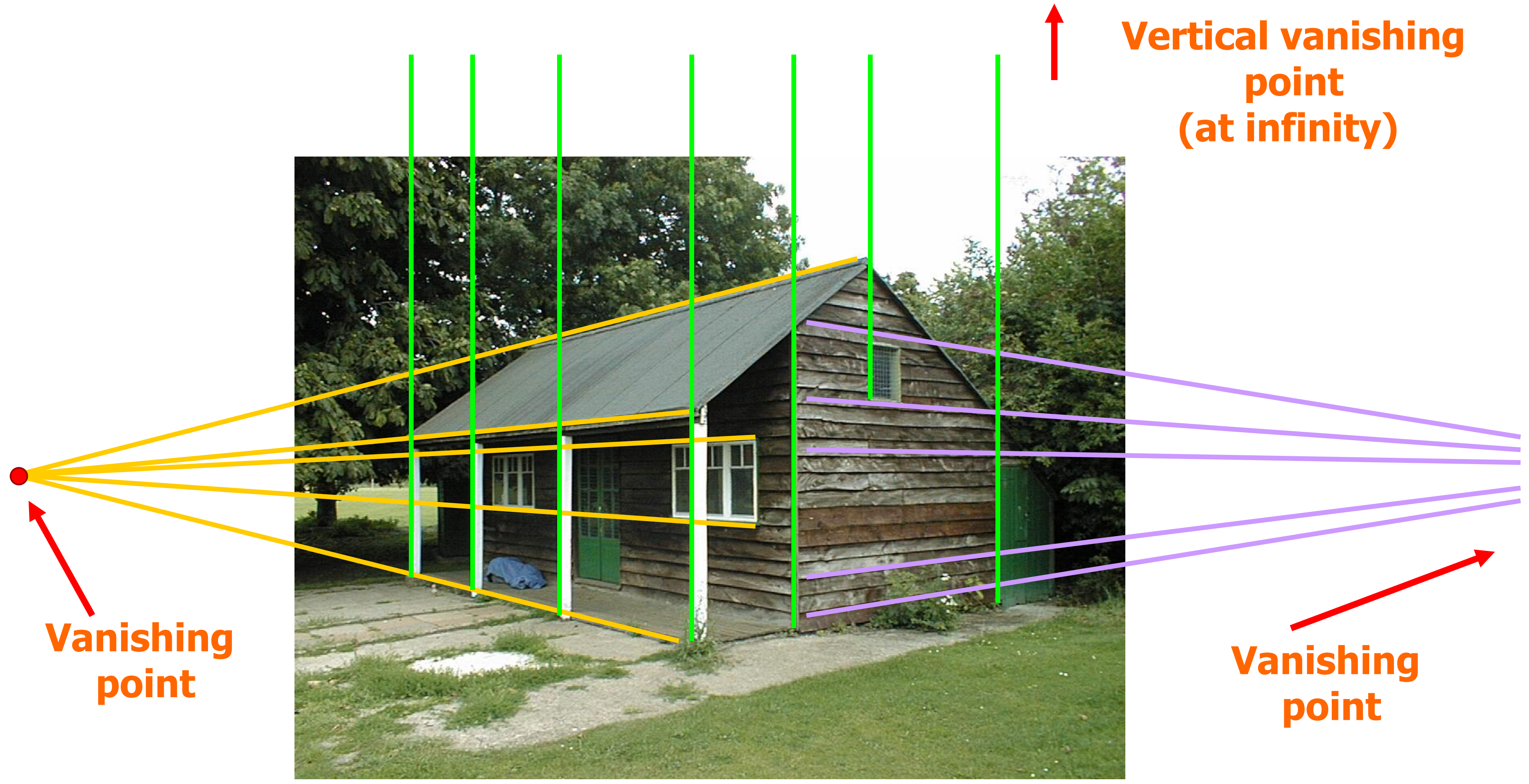
Projective Geometry

- Vanishing points and lines
  - Parallel lines in the world intersect in the image at a “vanishing point”

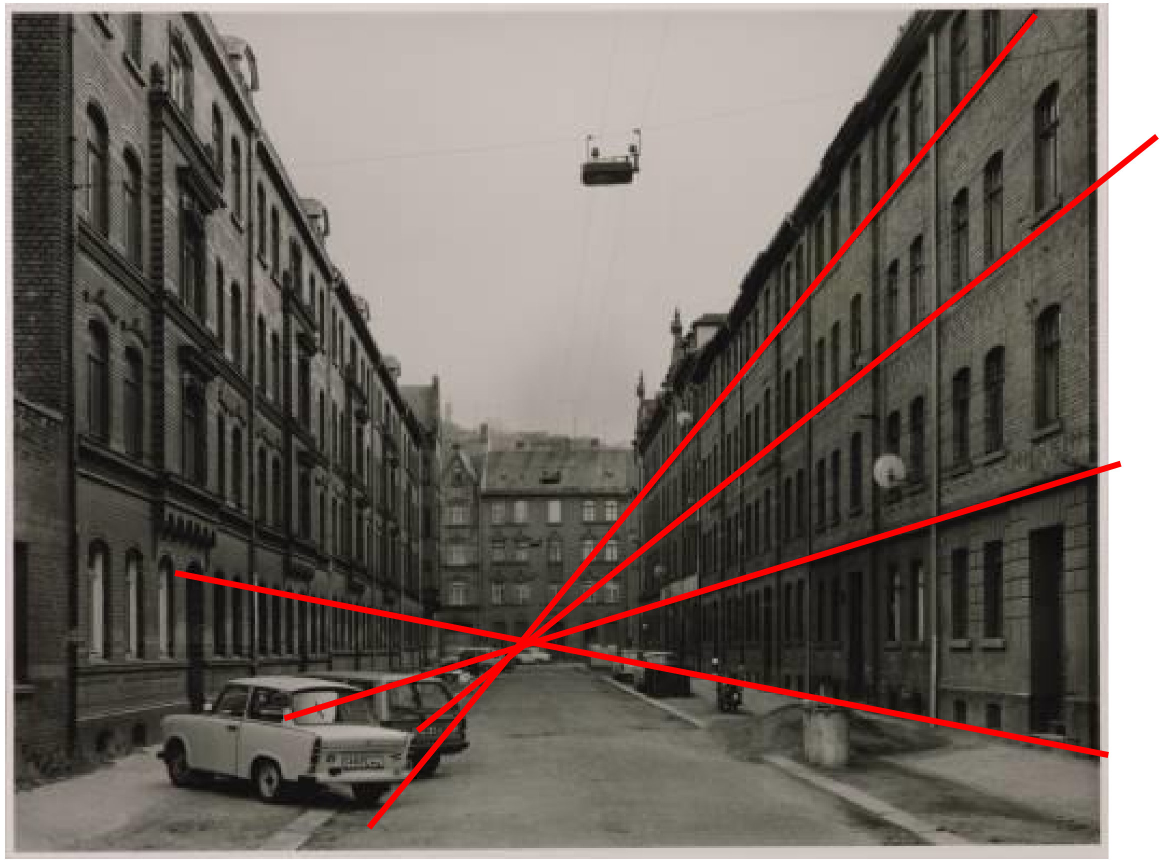




Projective Geometry



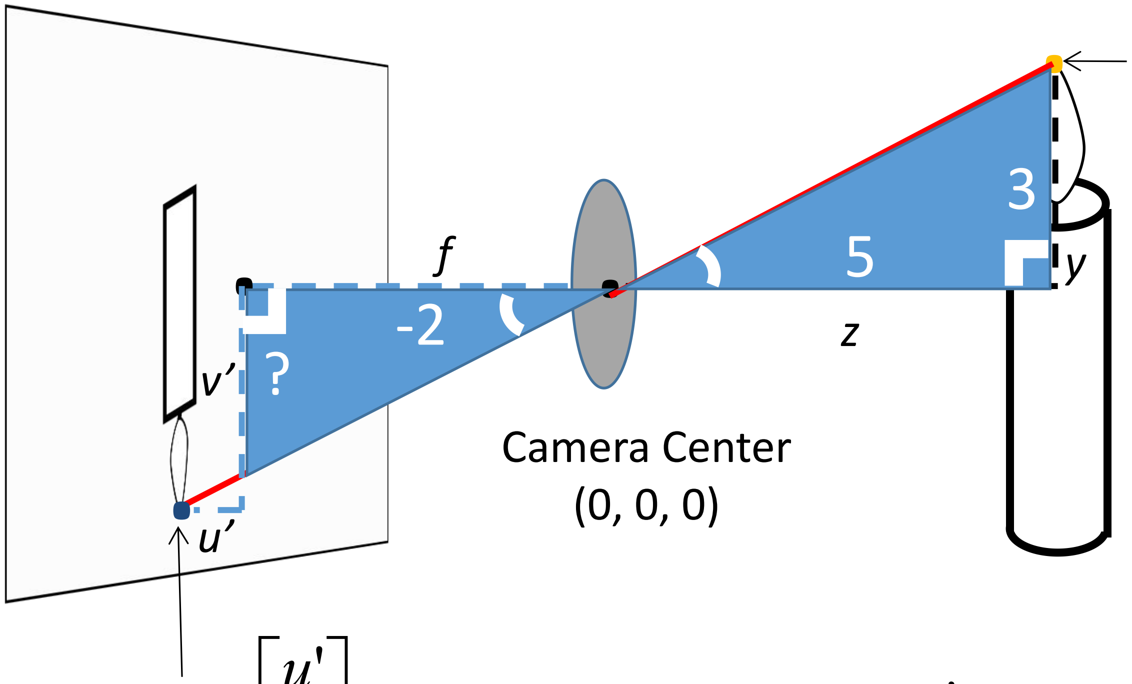
Projective Geometry



Questions: Why vertical parallel lines haven't have a finite vanishing point?

Projection

- World coordinates → image coordinates



$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$x = \begin{bmatrix} u' \\ v' \end{bmatrix}$$

If X = 2, Y = 3,
Z = 5, and f = 2
What are U and V?

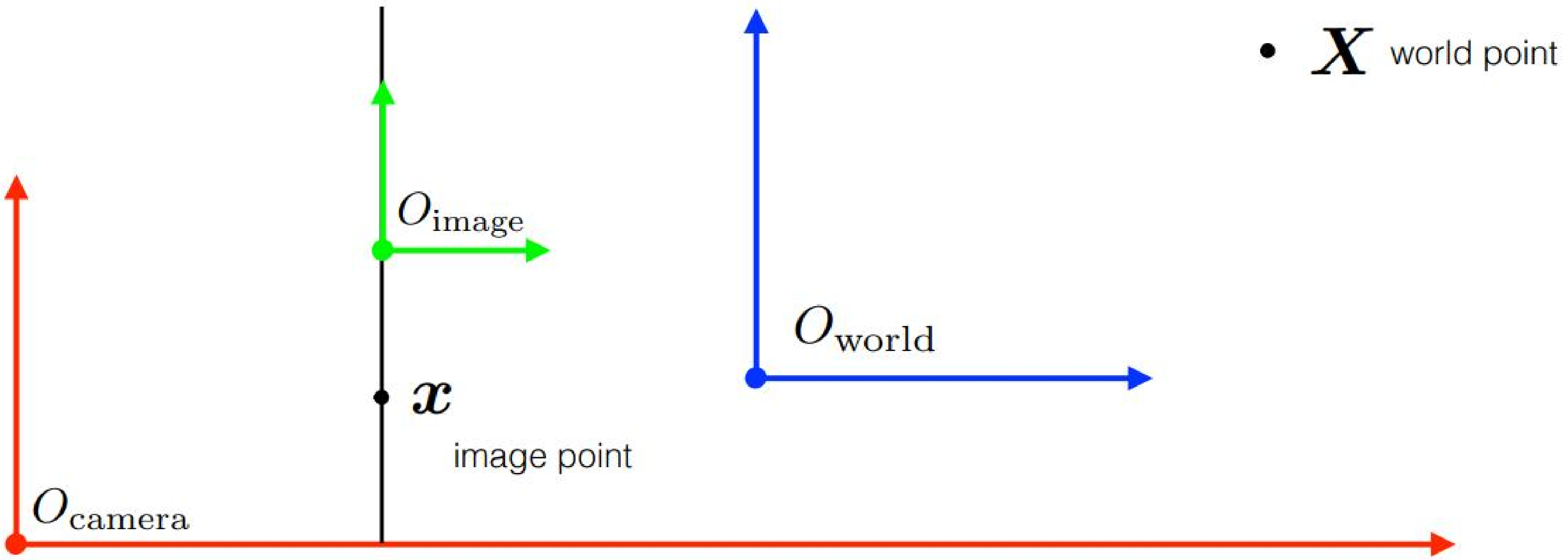
$$\frac{v'}{-f} = \frac{y}{z}$$

$$u' = -x * \frac{f}{z} \qquad u' = -2 * \frac{2}{5}$$

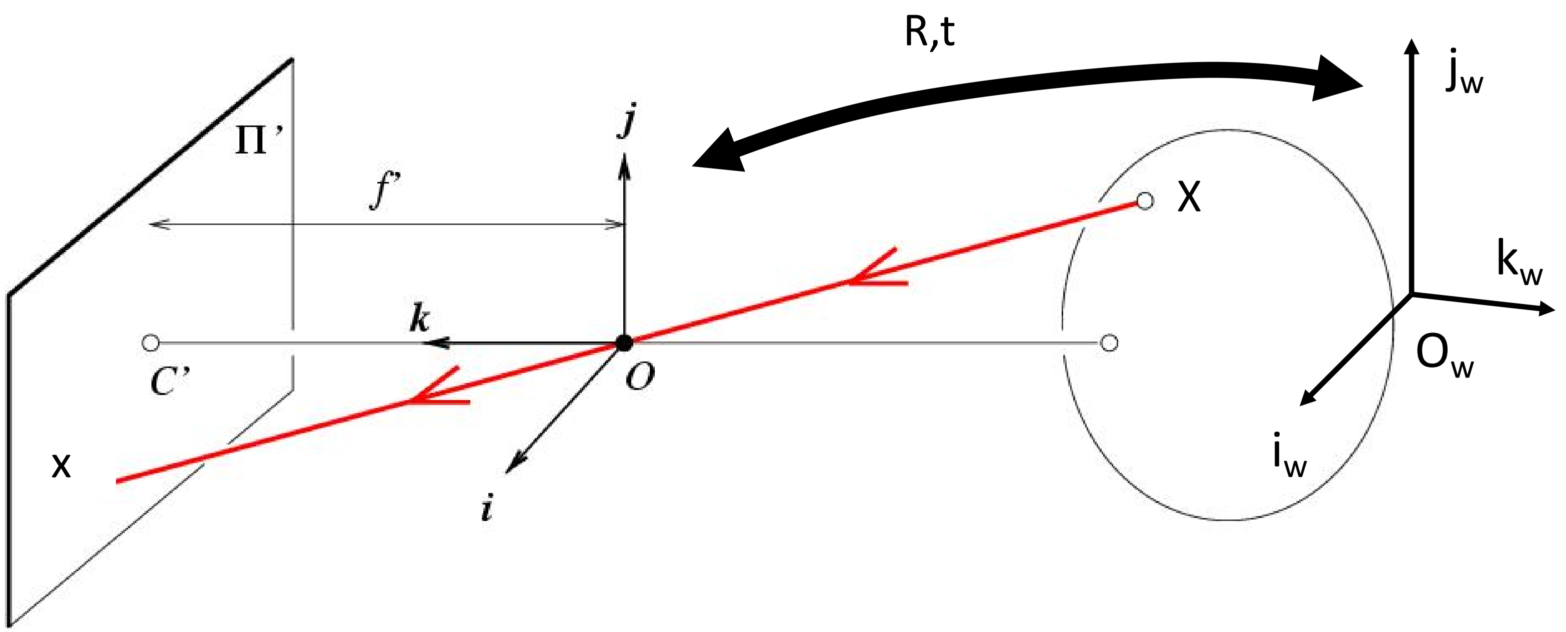
$$v' = -y * \frac{f}{z} \qquad v' = -3 * \frac{2}{5}$$

Three Different Coordinate Systems

- You need the know the transformations between them



Projection Matrix



$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\mathbf{X}$$

x: Image Coordinates: (u,v,1)
K: Intrinsic Matrix (3x3)
R: Rotation (3x3)
t: Translation (3x1)
X: World Coordinates: (X,Y,Z,1)

Projection Matrix

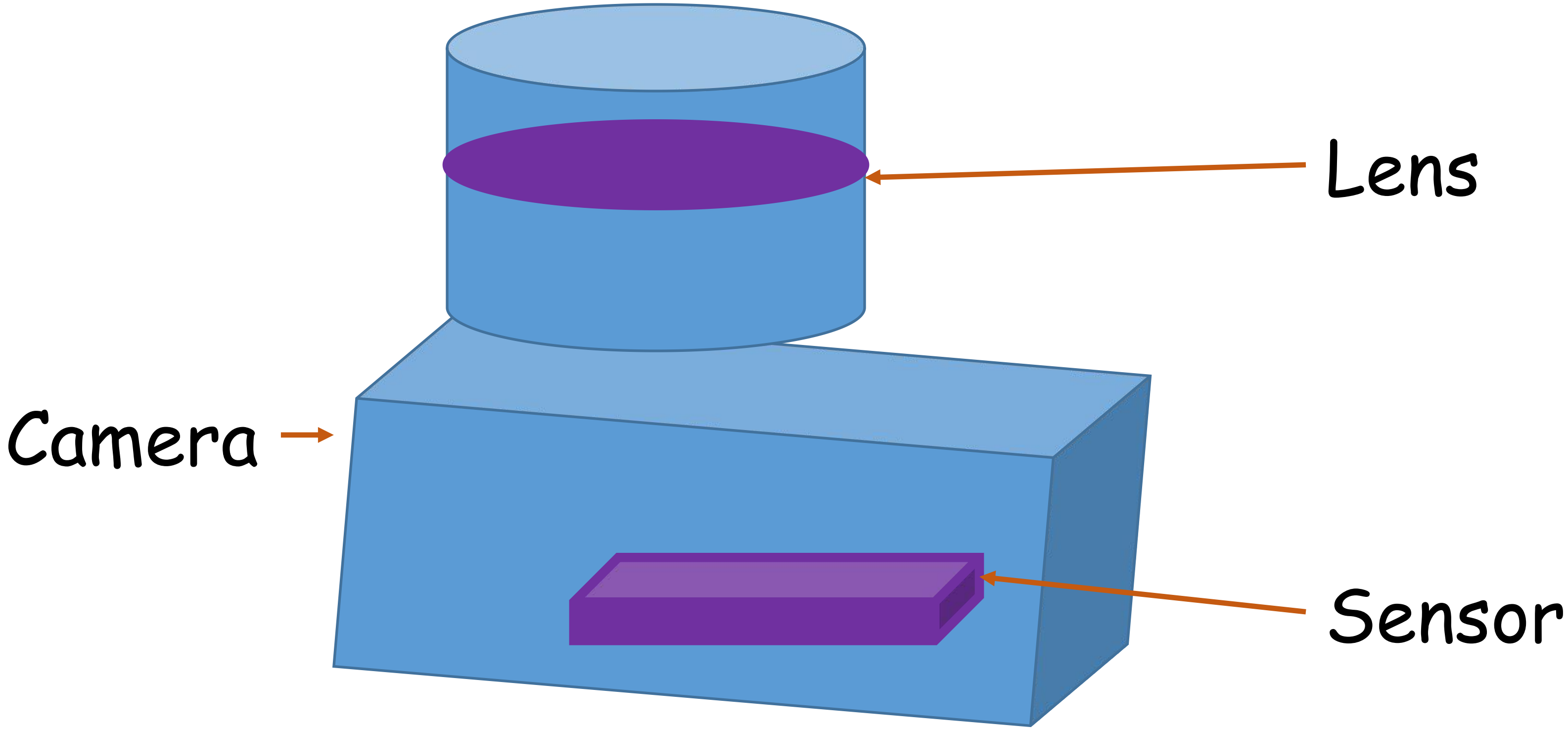
- Inserting photographed objects into images (SIGGRAPH 2007)

Original

Created

Camera Intrinsic

- Potential problems caused by the production process



Camera Intrinsic

- Pixel values indexed by integer pixel coordinates
- Starting at the upper-left corner of the image
  - ✓ First scale the pixel values by the pixel spacing
  - ✓ Then describe the orientation of the sensor array relative to the camera projection center

the sensor planes at location

a sensor homography

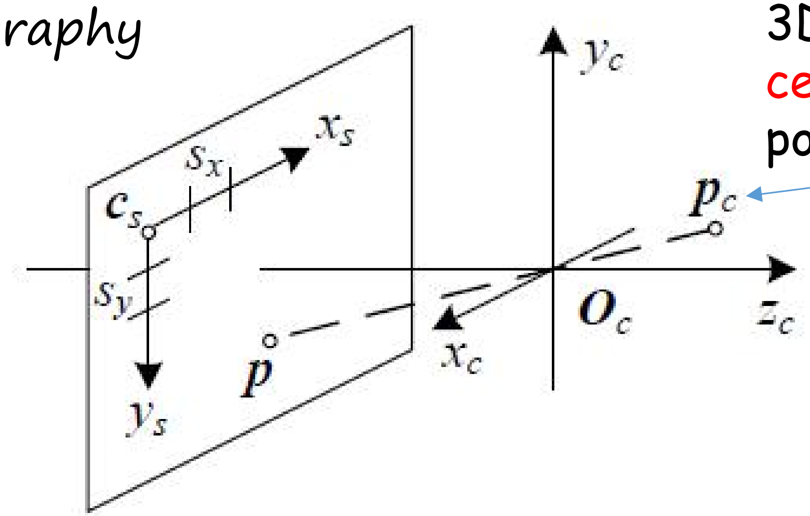
$$p = \left[\, R_s \mid c_s \,\right] \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} = M_s \bar{x}_s$$

3D rotation

origin

scale

integer pixel coordinates



3D camera-centered points

Camera Intrinsic

- The relationship between the 3D pixel center and the 3D camera-centered point is given by an unknown scaling s
  - ➢ The calibration matrix describes the camera intrinsics

$$p = s p_c \qquad \tilde{x}_s = s M_s^{-1} p_c = K p_c$$

the sensor planes at location

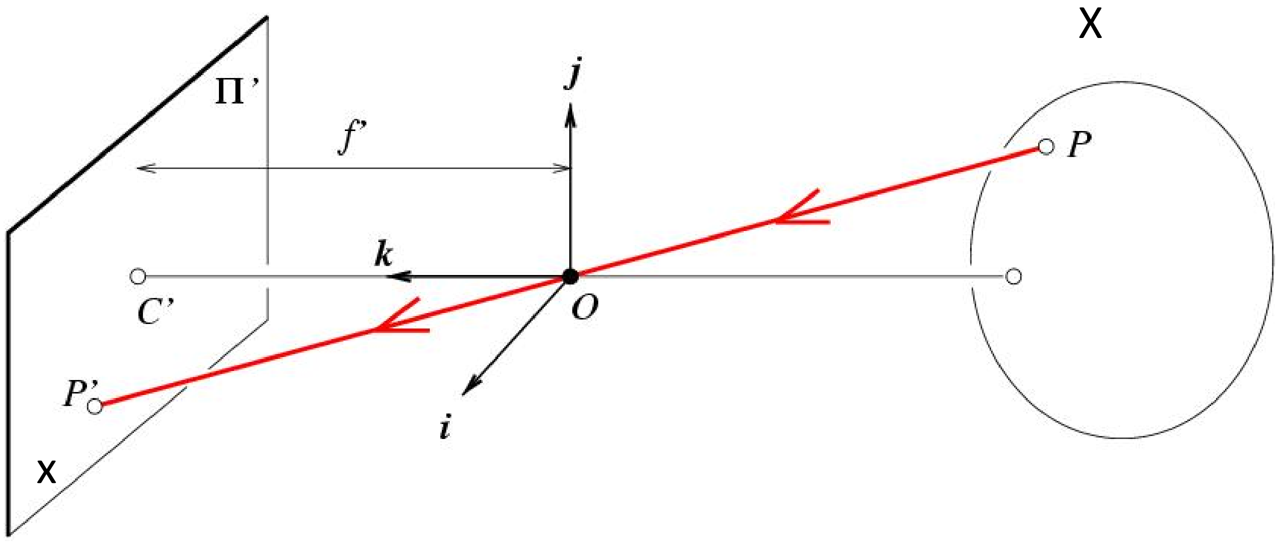
3D camera-centered points

pixel address

calibration matrix

Projection (Camera) matrix

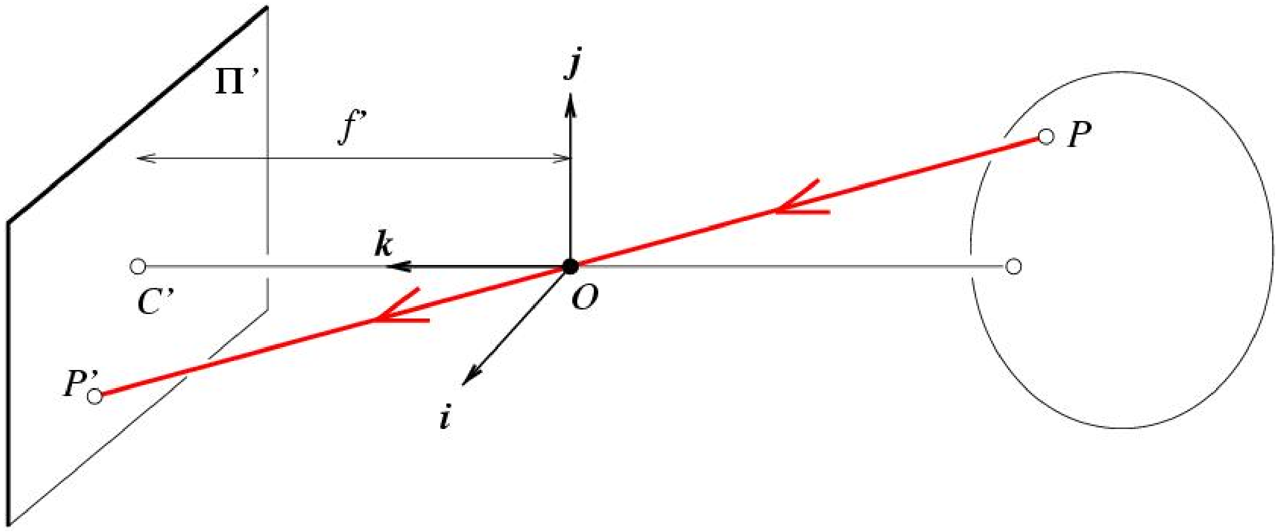
- Intrinsic Assumptions
  - Unit aspect ratio
  - Optical center at (0,0)
  - No skew
- Extrinsic Assumptions
  - No rotation
  - Camera at (0,0,0)



$$\mathbf{x} = \mathbf{K}\underbrace{\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}}_{\text{Perspective}}\mathbf{X} \Rightarrow w\begin{bmatrix}u\\ v\\ 1\end{bmatrix} = \begin{bmatrix}\overbrace{\begin{matrix}f & 0 & 0\\ 0 & f & 0\\ 0 & 0 & 1\end{matrix}}^{\mathrm{K}} & \begin{matrix}0\\ 0\\ 0\end{matrix}\end{bmatrix}\begin{bmatrix}x\\ y\\ z\\ 1\end{bmatrix}$$

Projection (Camera) matrix

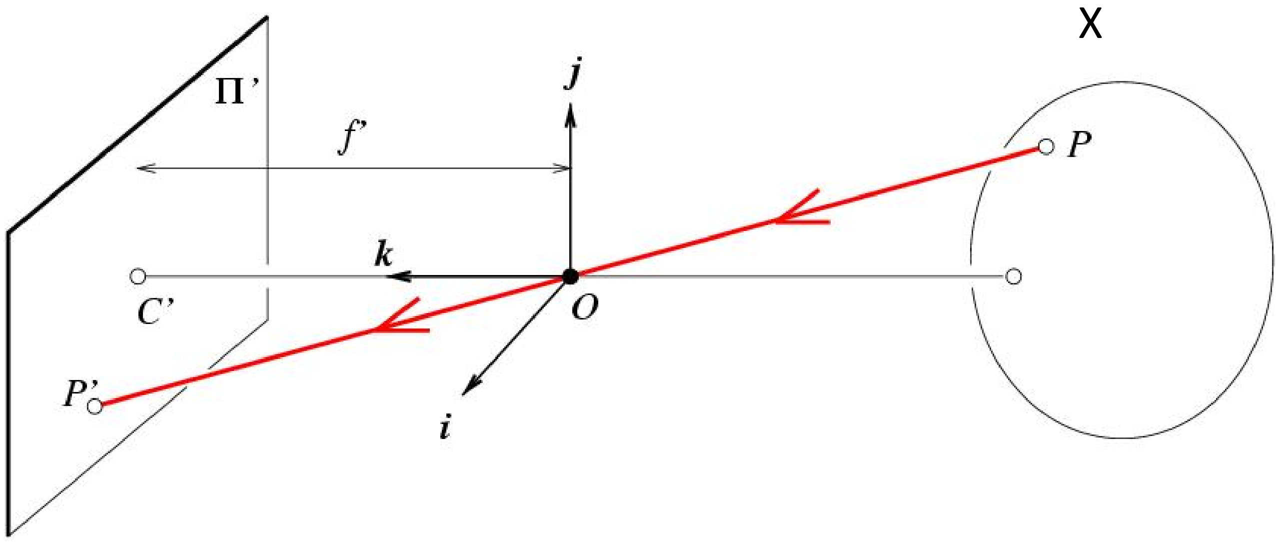
- Intrinsic Assumptions
  - Unit aspect ratio
  -
  - No skew
- Extrinsic Assumptions
  - No rotation
  - Camera at (0,0,0)



$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \Rightarrow w\begin{bmatrix}u\\v\\1\end{bmatrix} = \begin{bmatrix}f & 0 & u_0 & 0\\0 & f & v_0 & 0\\0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix}$$

Projection (Camera) matrix

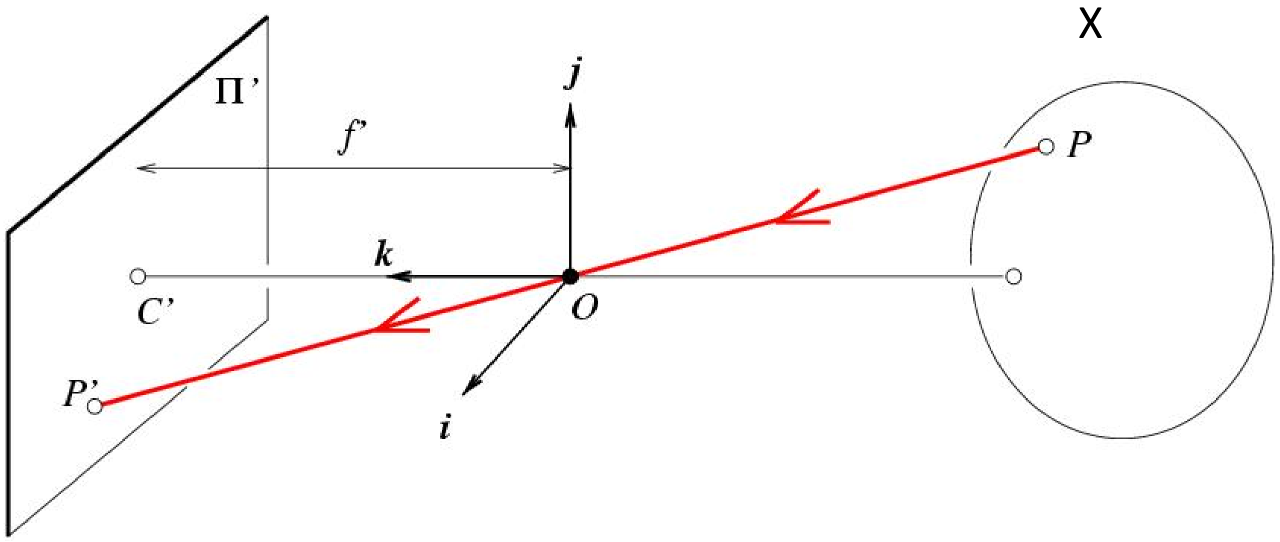
- Intrinsic Assumptions
  -
  -
  - No skew
- Extrinsic Assumptions
  - No rotation
  - Camera at (0,0,0)



$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \Rightarrow w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & 0 & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Projection (Camera) matrix

- Intrinsic Assumptions
- Extrinsic Assumptions
  - No rotation
  - Camera at (0,0,0)

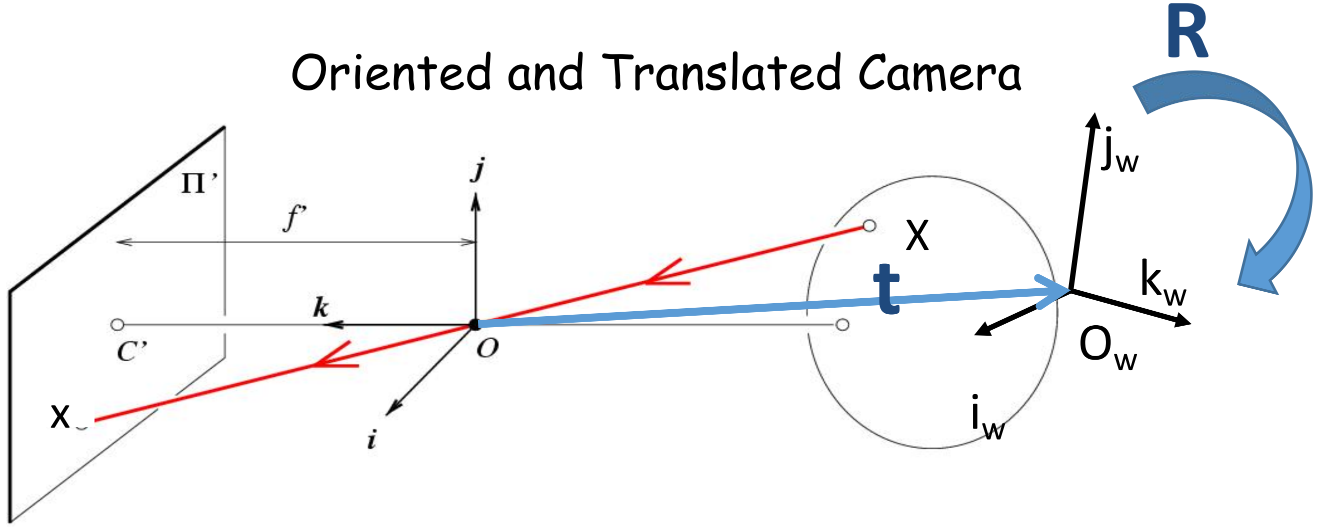


$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\mathbf{X} \Rightarrow w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & s & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

s encodes any possible skew between the sensor axes due to the sensor not being mounted perpendicular to the optical axis

Projection (Camera) matrix

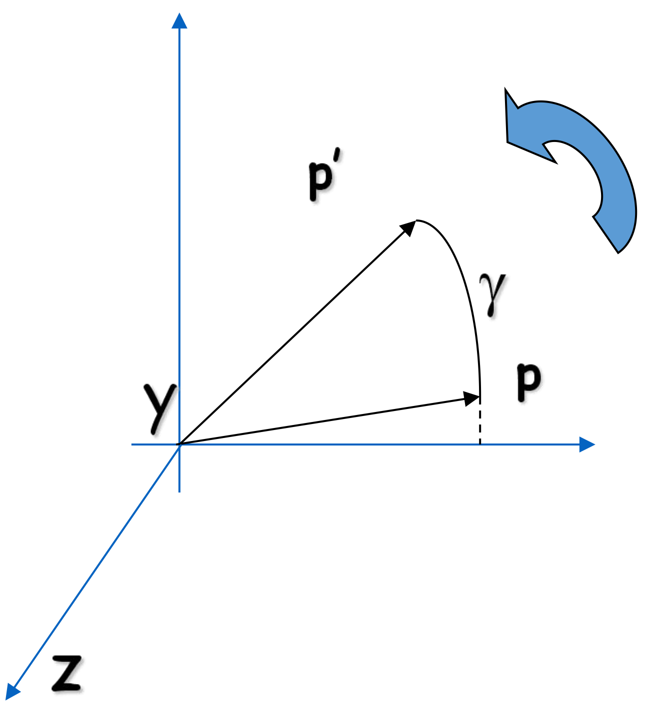
- Intrinsic Assumptions
  -
  -
  -
- Extrinsic Assumptions
  - No rotation
  -



$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{I} & \mathbf{t}\end{bmatrix}\mathbf{X} \Rightarrow w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & s & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Projection (Camera) matrix

- 3D Rotation of Points
  - ➢ Rotation around the coordinate axes, counter-clockwise:



$$R_x(\alpha)=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\alpha & -\sin\alpha\\ 0 & \sin\alpha & \cos\alpha\end{bmatrix}$$

$$R_y(\beta)=\begin{bmatrix}\cos\beta & 0 & \sin\beta\\ 0 & 1 & 0\\ -\sin\beta & 0 & \cos\beta\end{bmatrix}$$

$$R_z(\gamma)=\begin{bmatrix}\cos\gamma & -\sin\gamma & 0\\ \sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1\end{bmatrix}$$

Projection (Camera) matrix

- Allow camera rotation

$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\mathbf{X}$$

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & s & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

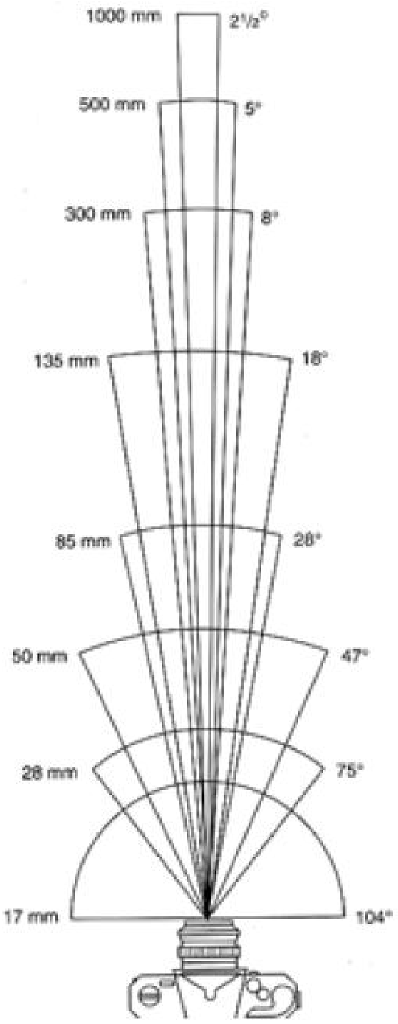
Projection (Camera) matrix

- Vanishing point = Projection from infinity

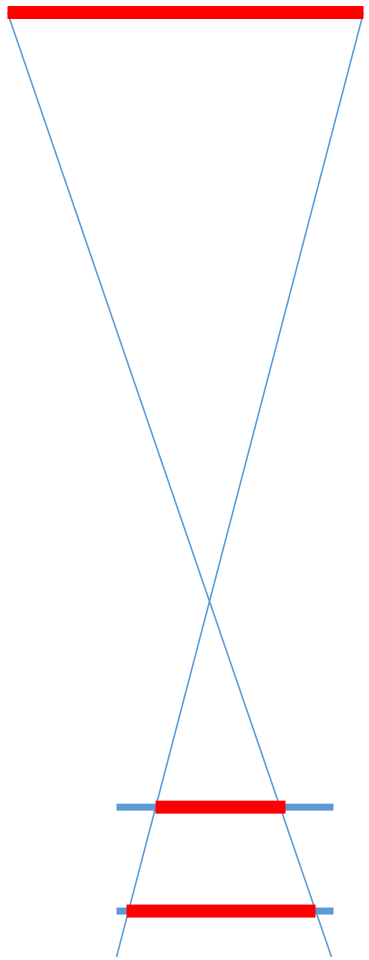
$$\mathbf{p} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 0 \end{bmatrix} \Rightarrow \mathbf{p} = \mathbf{KR}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \Rightarrow \mathbf{p} = \mathbf{K}\begin{bmatrix} x_R \\ y_R \\ z_R \end{bmatrix}$$

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_R \\ y_R \\ z_R \end{bmatrix} \Rightarrow \qquad u = \frac{f x_R}{z_R} + u_0 \qquad v = \frac{f y_R}{z_R} + v_0$$

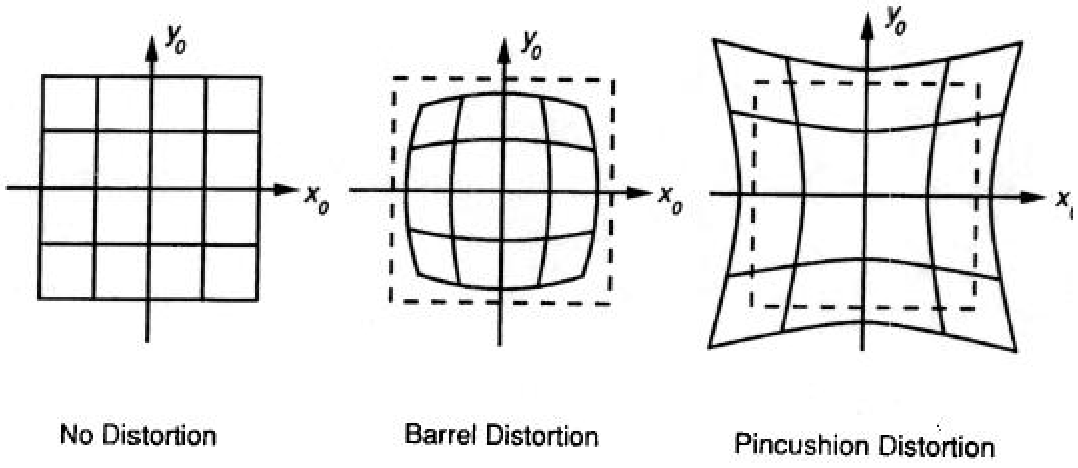
Field of View (Zoom, Focal Length)



From London and Upton



Beyond Pinholes: Radial Distortion

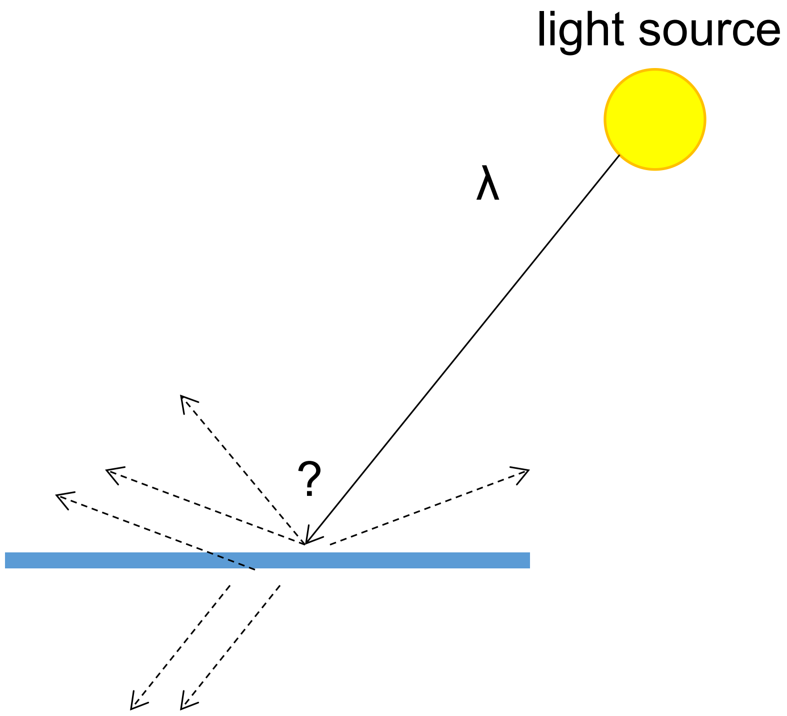


Corrected Barrel Distortion

Photometric image formation

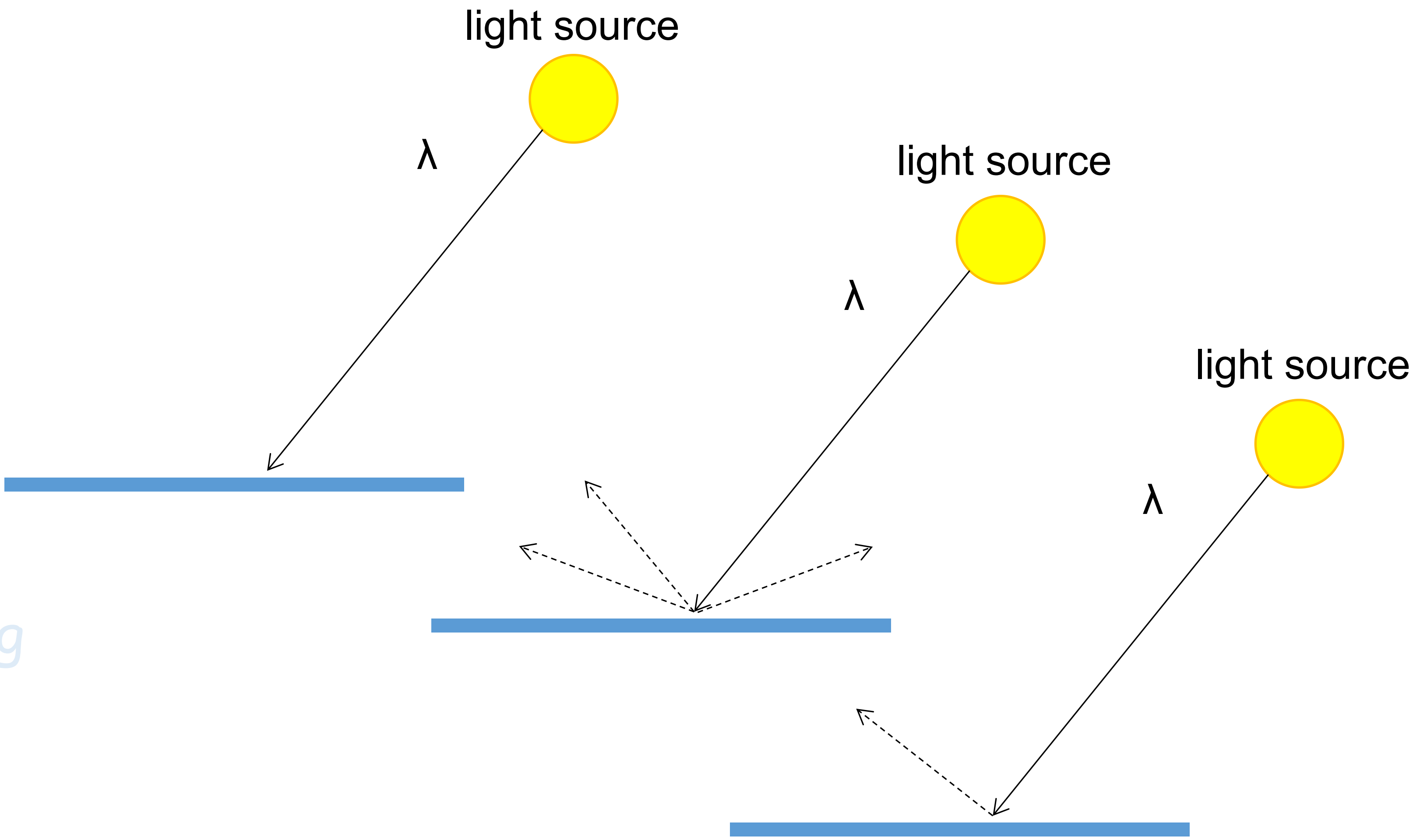
A photon's life choices

- Absorption 吸收
- Diffusion 漫射
- Reflection反射
- Transparency 透射
- Refraction 折射
- Fluorescence 荧光反应
- Subsurface scattering 次表面散射
- Phosphorescence 磷光
- Interreflection 相互反射



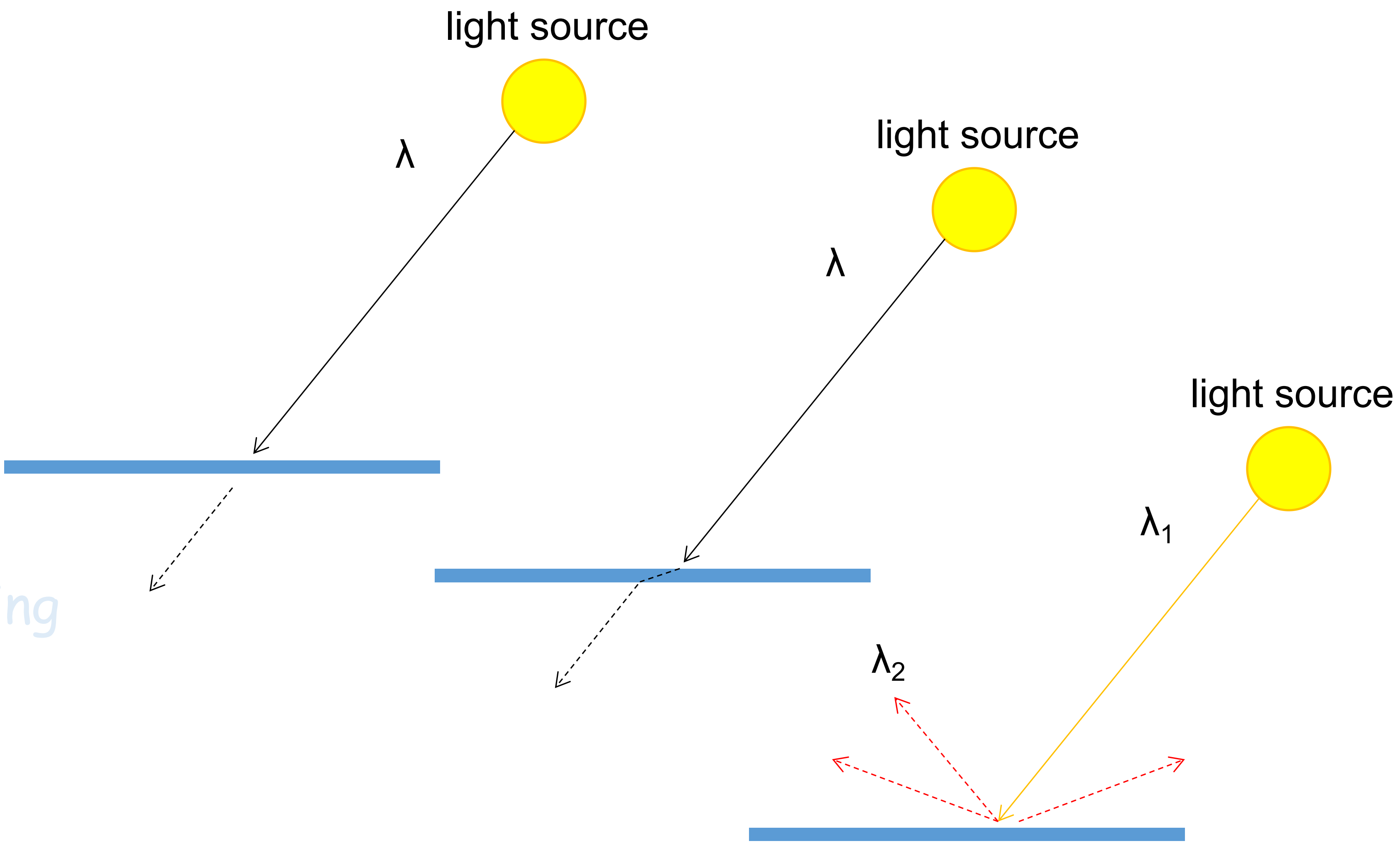
A photon's life choices

- Absorption
- Diffusion
- Reflection
- Transparency
- Refraction
- Fluorescence
- Subsurface scattering
- Phosphorescence
- Interreflection



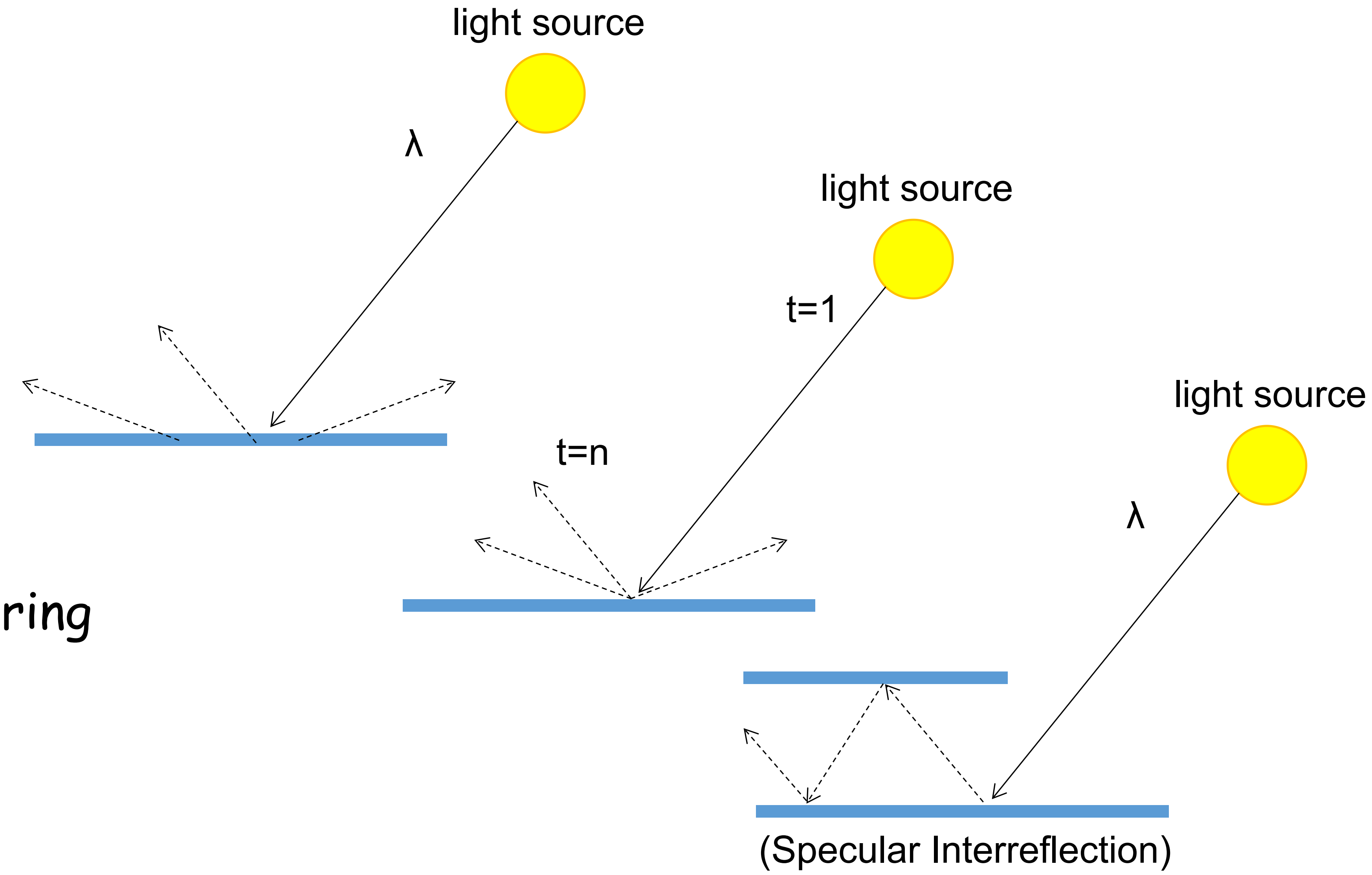
A photon's life choices

- Absorption
- Diffusion
- Reflection
- Transparency
- Refraction
- Fluorescence
- Subsurface scattering
- Phosphorescence
- Interreflection



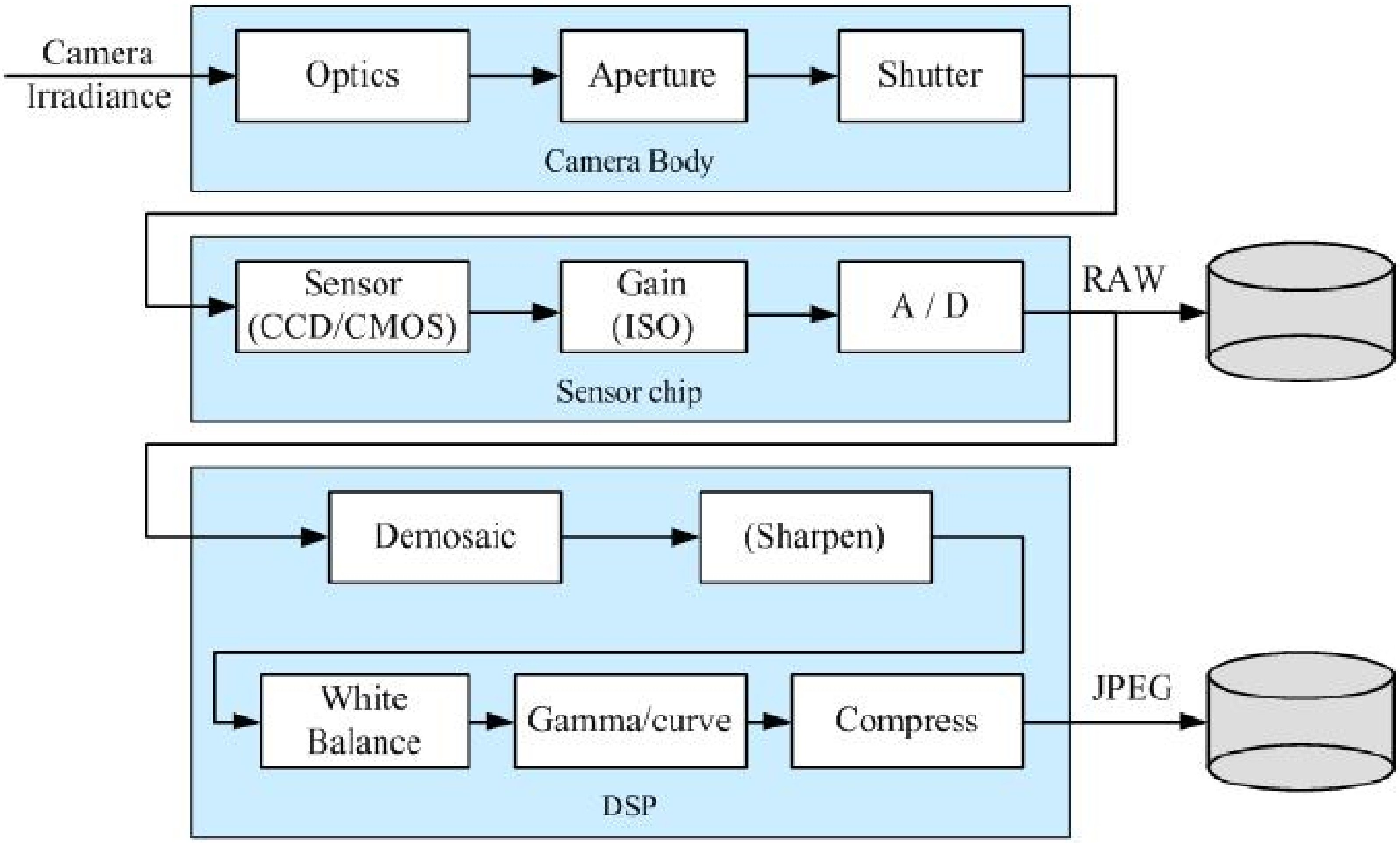
A photon's life choices

- Absorption
- Diffusion
- Reflection
- Transparency
- Refraction
- Fluorescence
- Subsurface scattering
- Phosphorescence
- Interreflection



The digital camera

Image sensing pipeline



Digital Camera

- A digital camera replaces film with a sensor array
  - Each cell in the array is light-sensitive diode （光敏二极管） that converts photons to electrons
  - Two common types
    - ✓Charge Coupled Device (CCD)
    - ✓CMOS

Sensor Array

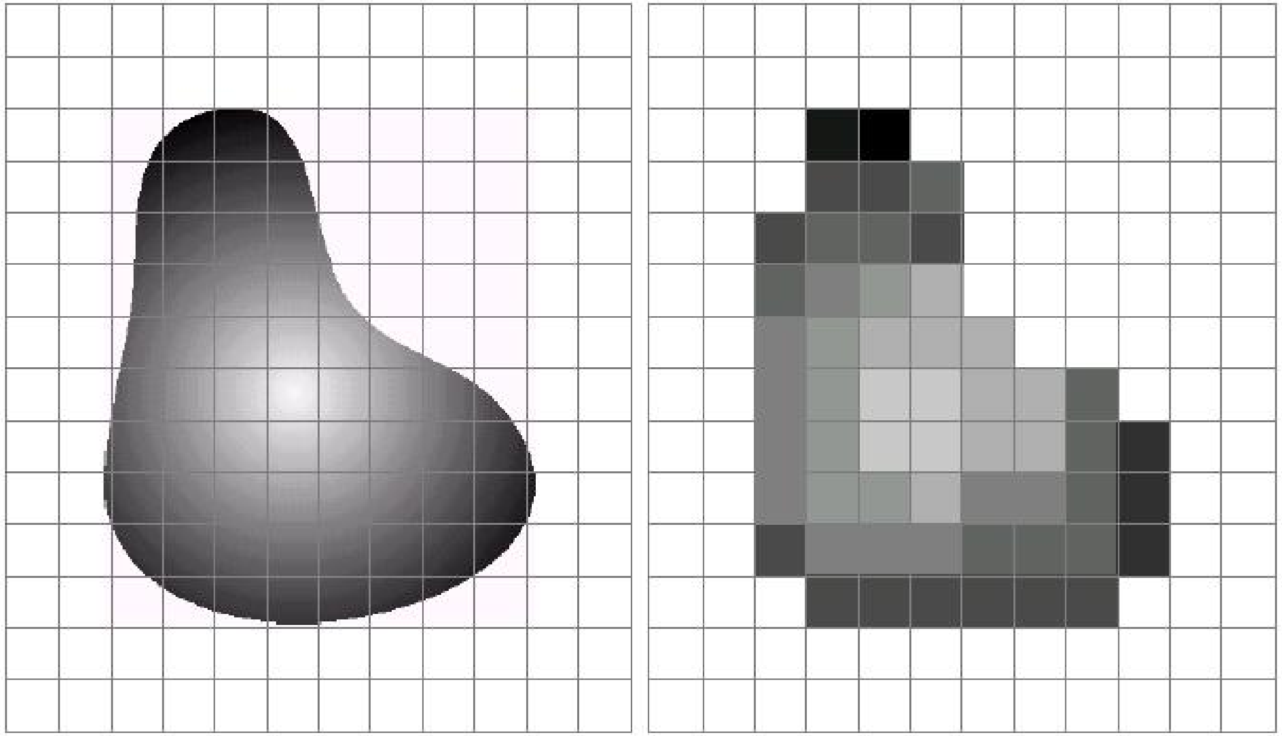
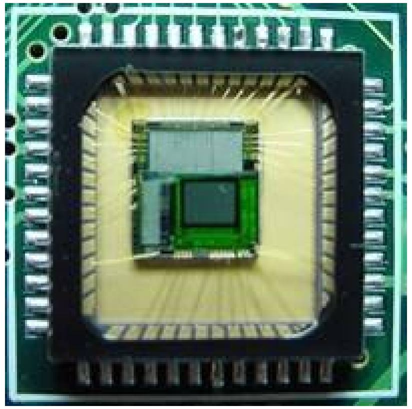


FIGURE 2.17 (a) Continuos image projected onto a sensor array. (b) Result of image sampling and quantization.



CMOS sensor

Sampling and Quantization

- Shannon's Sampling Theorem

$$f_s \geq 2f_{\max}$$

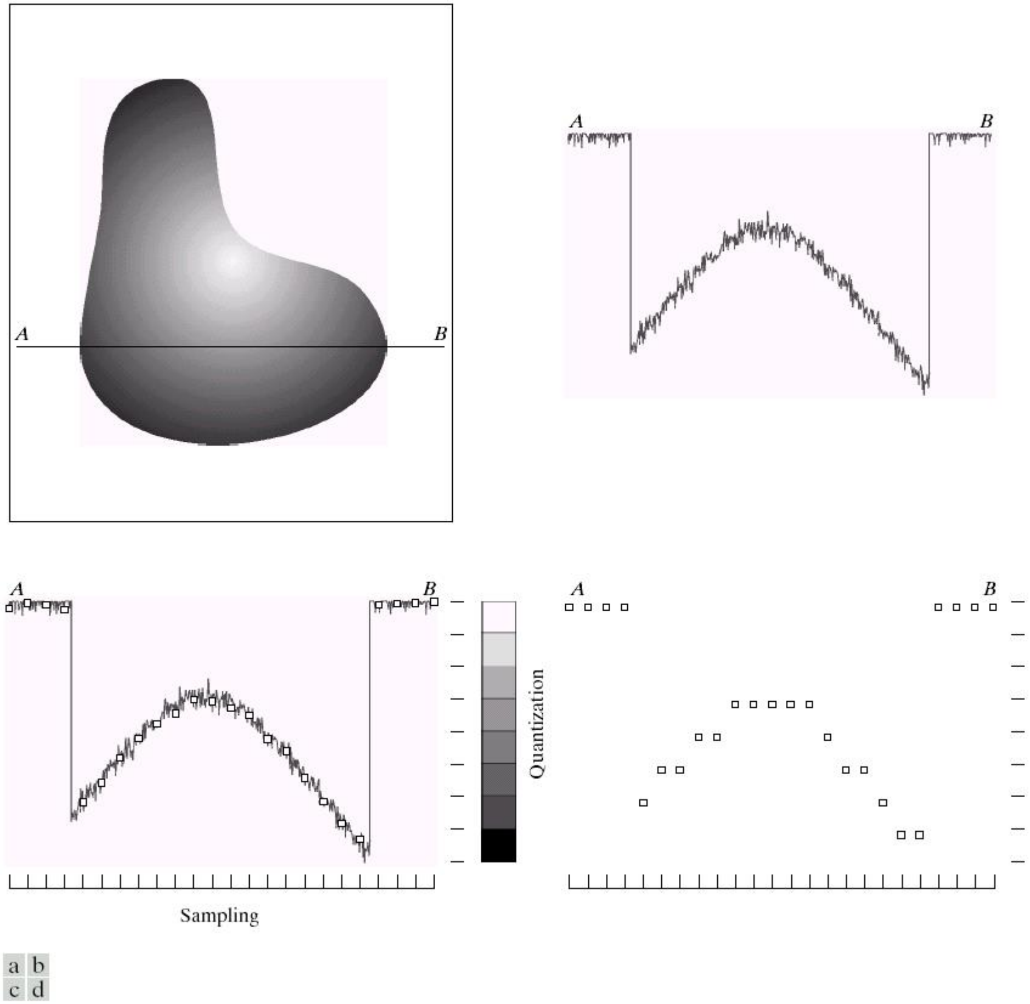
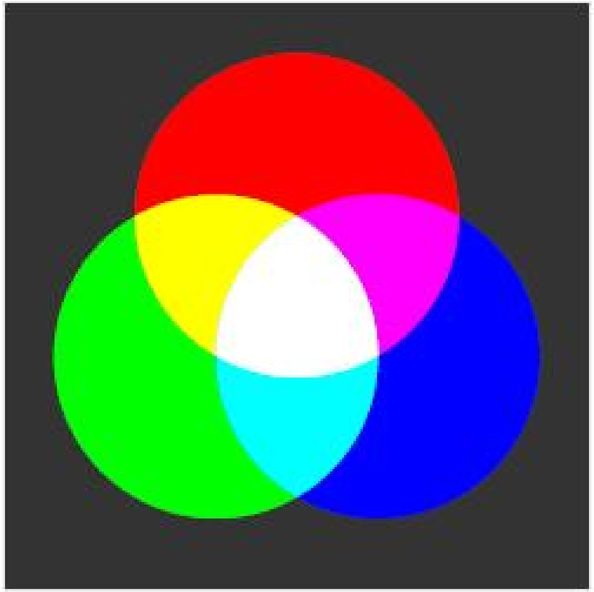


FIGURE 2.16 Generating a digital image. (a) Continuous image. (b) A scan line from A to B in the continuous image, used to illustrate the concepts of sampling and quantization. (c) Sampling and quantization. (d) Digital scan line.

Color

- Primary and secondary colors



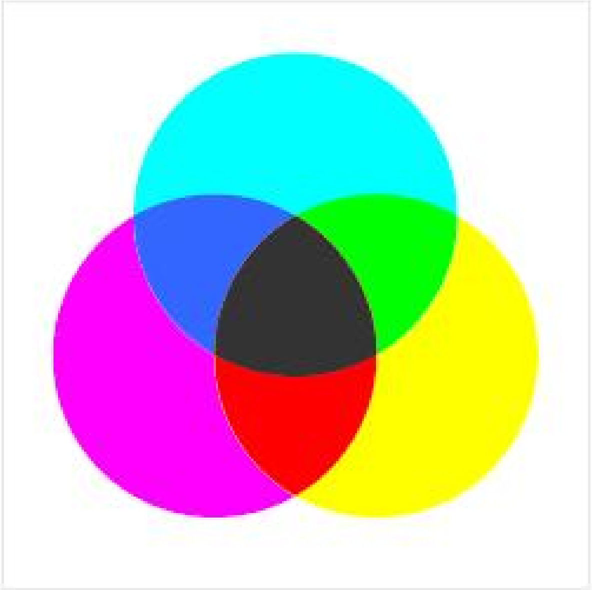
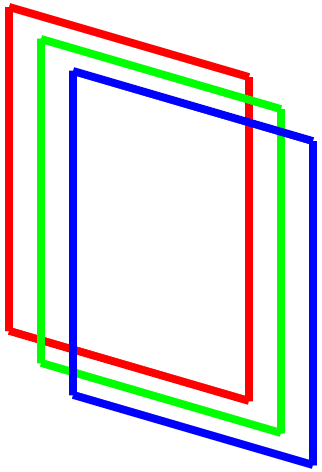
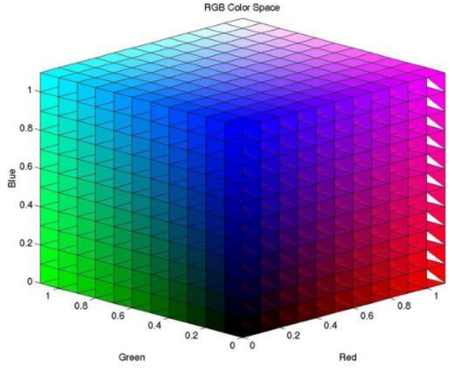


Image: three matrices



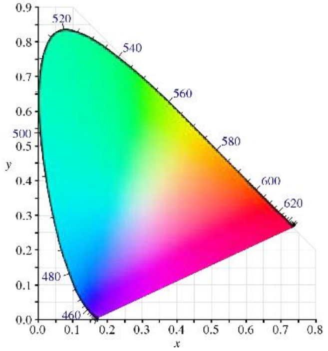
Color Spaces

- RGB



$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

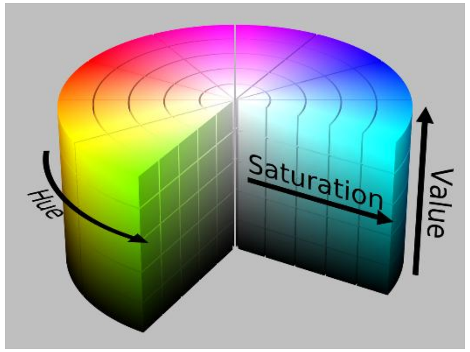
- CIE XYZ



Luminance

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad z = \frac{Z}{X+Y+Z}$$

- HSV
  - Hue
  - Saturation
  - Value



$$C = V \times S_{HSV}$$

$$H' = \frac{H}{60^\circ}$$

$$X = C \times (1 - |H' \mod 2 - 1|)$$

$$m = V - C$$

$$(R_1, G_1, B_1) = \begin{cases} (0,0,0) & \text{if } H \text{ is undefined} \\ (C,X,0) & \text{if } 0 \leq H' \leq 1 \\ (X,C,0) & \text{if } 1 < H' \leq 2 \\ (0,C,X) & \text{if } 2 < H' \leq 3 \\ (0,X,C) & \text{if } 3 < H' \leq 4 \\ (X,0,C) & \text{if } 4 < H' \leq 5 \\ (C,0,X) & \text{if } 5 < H' \leq 6 \end{cases}$$

$$(R, G, B) = (R_1 + m, G_1 + m, B_1 + m)$$

Color Filter Arrays

- Color filter array layout
- Interpolated pixel values
  - ➢ The luminance signal is mostly determined by green values
  - ➢ The visual system is much more sensitive to high frequency detail in luminance than in chrominance

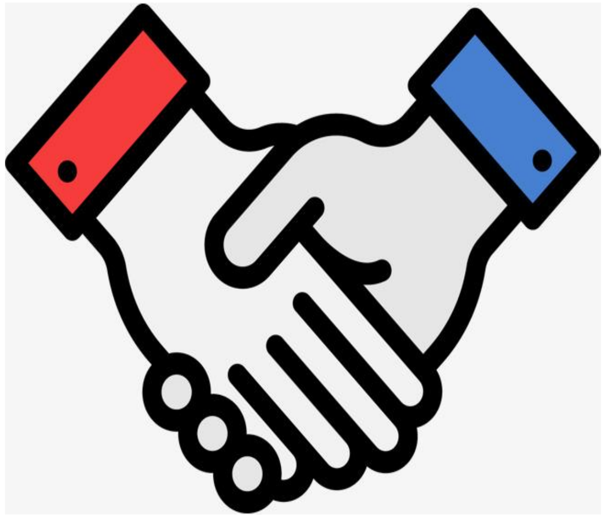
G	R	G	R
B	G	B	G
G	R	G	R
B	G	B	G

rGb	Rgb	rGb	Rgb
rgB	rGb	rgB	rGb
rGb	Rgb	rGb	Rgb
rgB	rGb	rgB	rGb

Conclusions

Conclusions

- Vanishing points and vanishing lines
- Pinhole camera model and camera projection matrix
- Homogeneous coordinates
- Digital camera



Thanks

zhengf@sustc.edu.cn