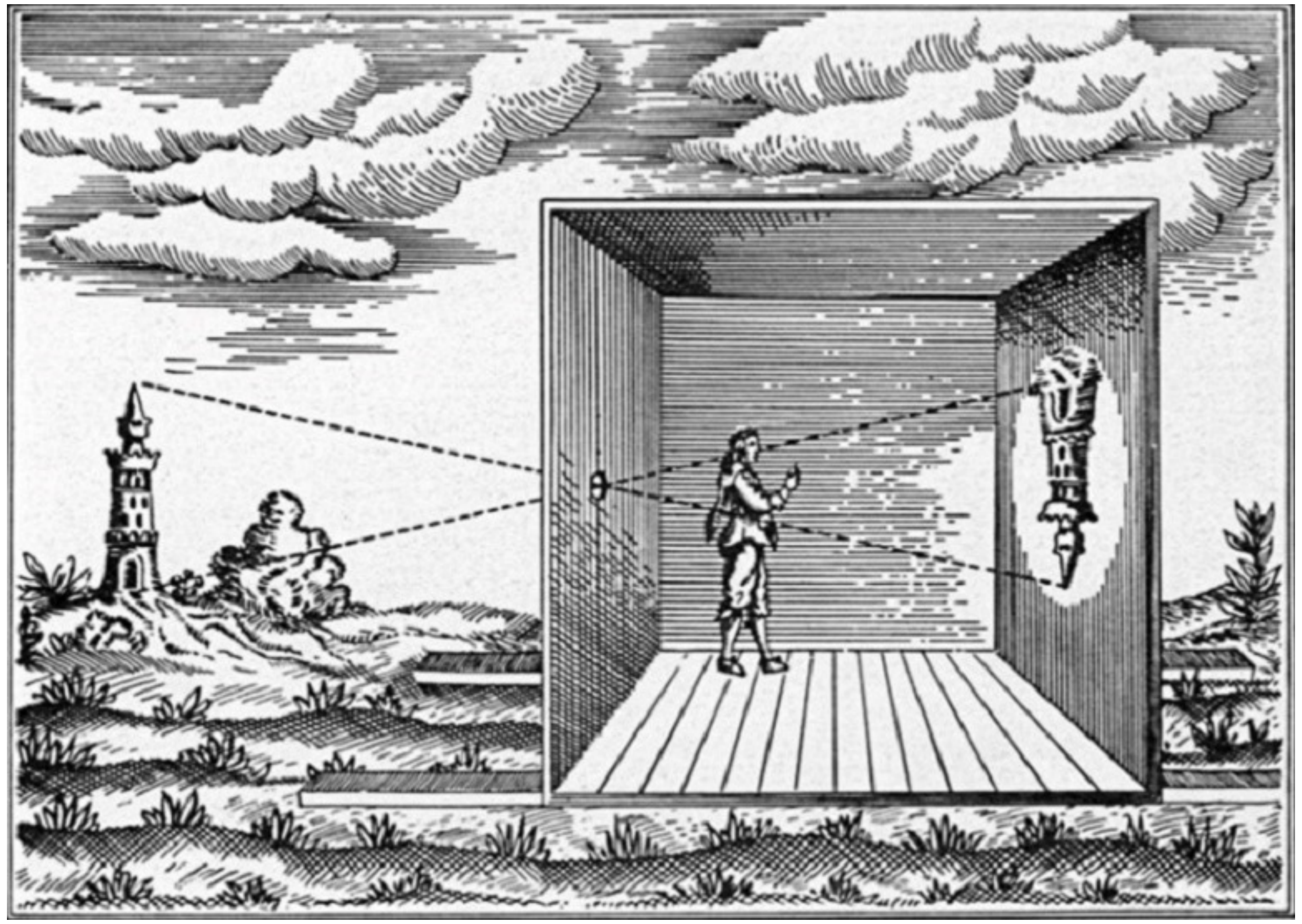
Lecture 2:
Image Formation

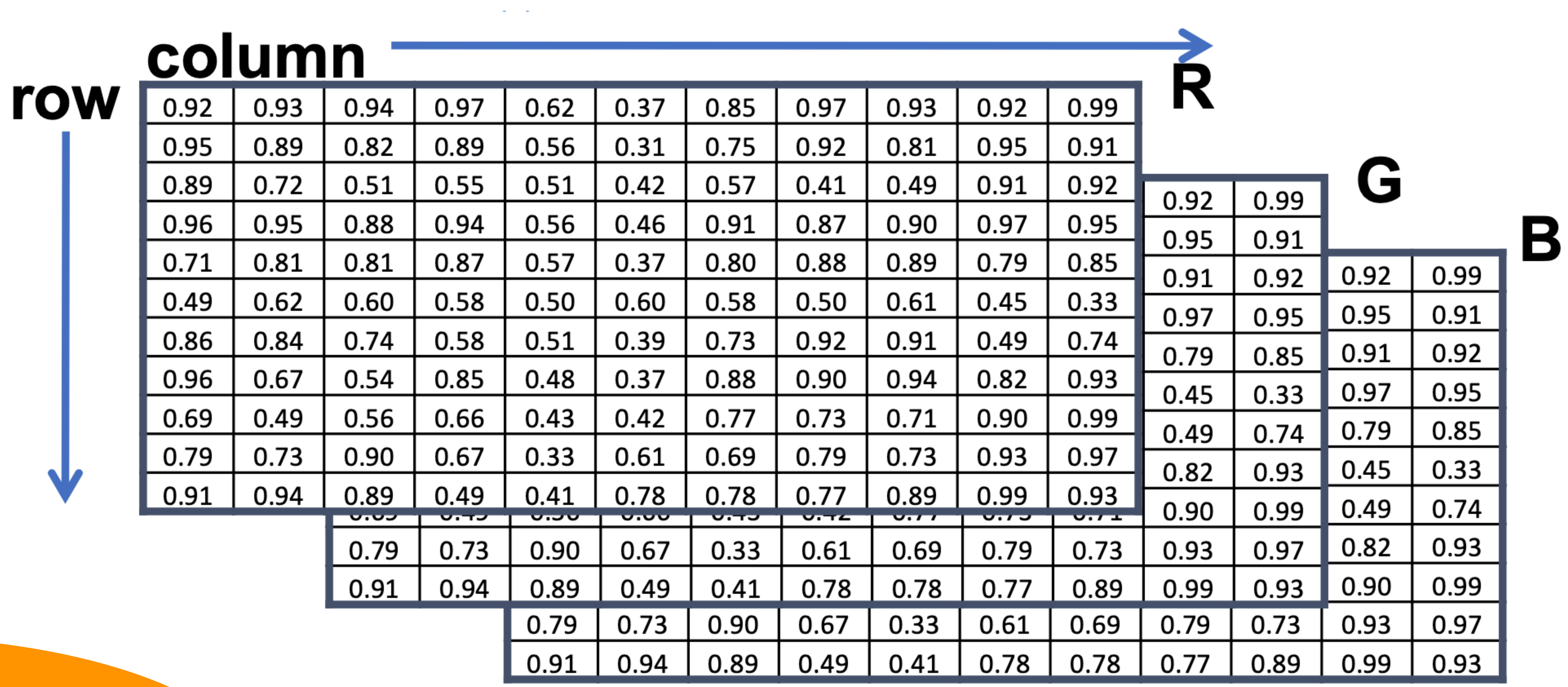


Laura Sevilla-Lara

Plan for today

- Image representation
- Light
- Color
- Cameras (Orthographic, Perspective)

Image Representation

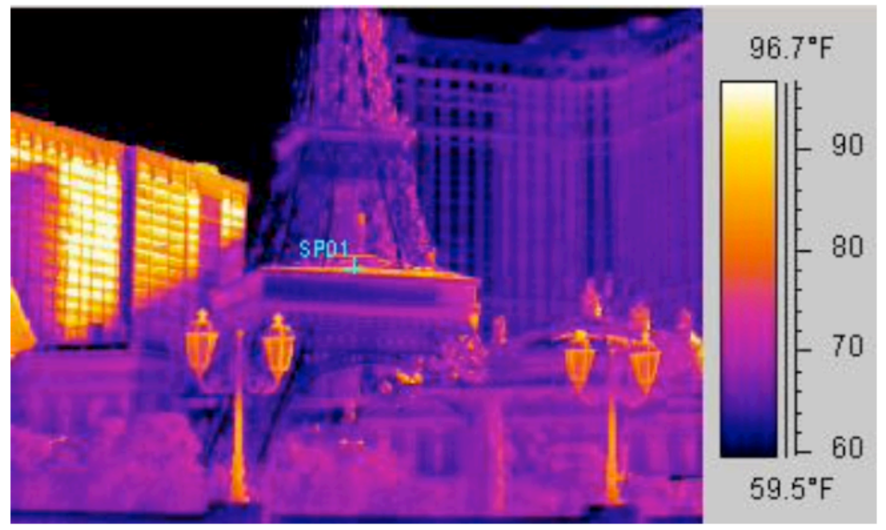


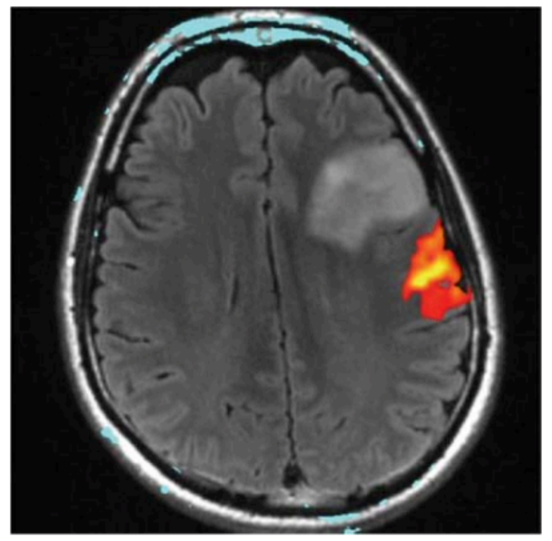
0.92	0.93	0.94	0.97	0.62	0.37	0.85	0.97	0.93	0.92	0.99
0.95	0.89	0.82	0.89	0.56	0.31	0.75	0.92	0.81	0.95	0.91
0.89	0.72	0.51	0.55	0.51	0.42	0.57	0.41	0.49	0.91	0.92
0.96	0.95	0.88	0.94	0.56	0.46	0.91	0.87	0.90	0.97	0.95
0.71	0.81	0.81	0.87	0.57	0.37	0.80	0.88	0.89	0.79	0.85
0.49	0.62	0.60	0.58	0.50	0.60	0.58	0.50	0.61	0.45	0.33
0.86	0.84	0.74	0.58	0.51	0.39	0.73	0.92	0.91	0.49	0.74
0.96	0.67	0.54	0.85	0.48	0.37	0.88	0.90	0.94	0.82	0.93
0.69	0.49	0.56	0.66	0.43	0.42	0.77	0.73	0.71	0.90	0.99
0.79	0.73	0.90	0.67	0.33	0.61	0.69	0.79	0.73	0.93	0.97
0.91	0.94	0.89	0.49	0.41	0.78	0.78	0.77	0.89	0.99	0.93

[James Hays]

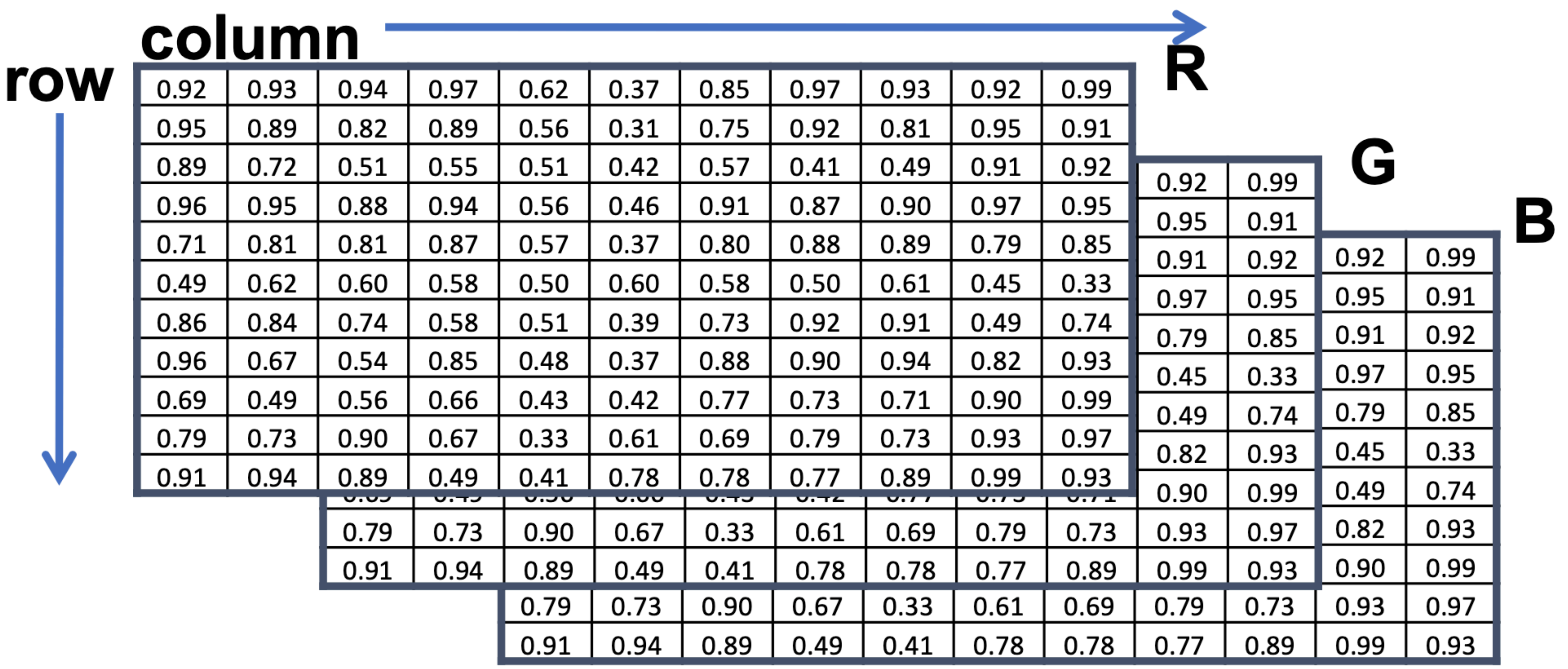
The 3D signal is sampled as a 2D function

Image Representation





[James Hays, Danny Alexander]



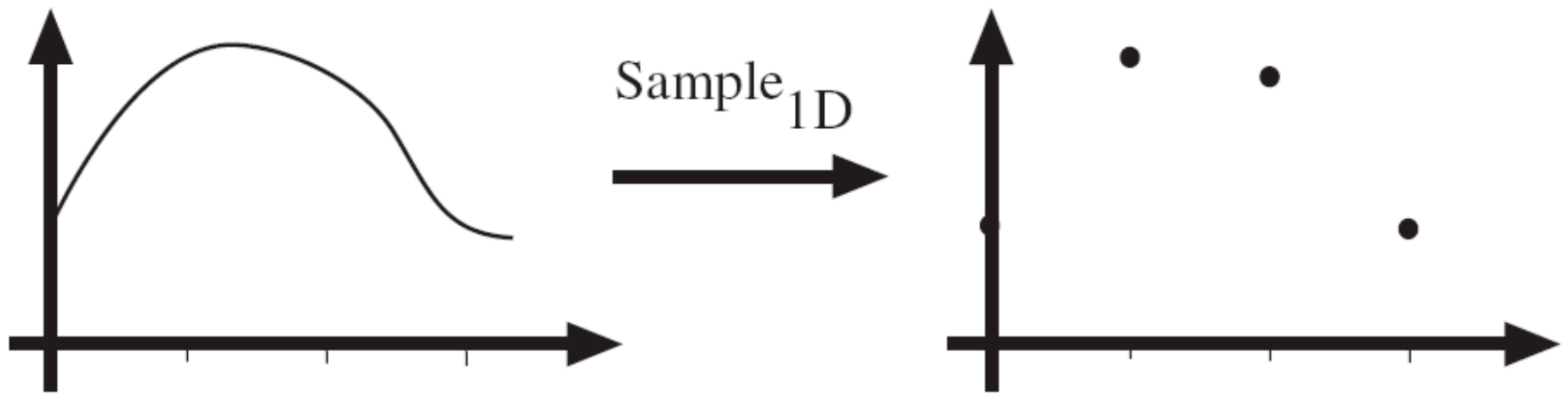
0.92	0.93	0.94	0.97	0.62	0.37	0.85	0.97	0.93	0.92	0.99
0.95	0.89	0.82	0.89	0.56	0.31	0.75	0.92	0.81	0.95	0.91
0.89	0.72	0.51	0.55	0.51	0.42	0.57	0.41	0.49	0.91	0.92
0.96	0.95	0.88	0.94	0.56	0.46	0.91	0.87	0.90	0.97	0.95
0.71	0.81	0.81	0.87	0.57	0.37	0.80	0.88	0.89	0.79	0.85
0.49	0.62	0.60	0.58	0.50	0.60	0.58	0.50	0.61	0.45	0.33
0.86	0.84	0.74	0.58	0.51	0.39	0.73	0.92	0.91	0.49	0.74
0.96	0.67	0.54	0.85	0.48	0.37	0.88	0.90	0.94	0.82	0.93
0.69	0.49	0.56	0.66	0.43	0.42	0.77	0.73	0.71	0.90	0.99
0.79	0.73	0.90	0.67	0.33	0.61	0.69	0.79	0.73	0.93	0.97
0.91	0.94	0.89	0.49	0.41	0.78	0.78	0.77	0.89	0.99	0.93

[James Hays]

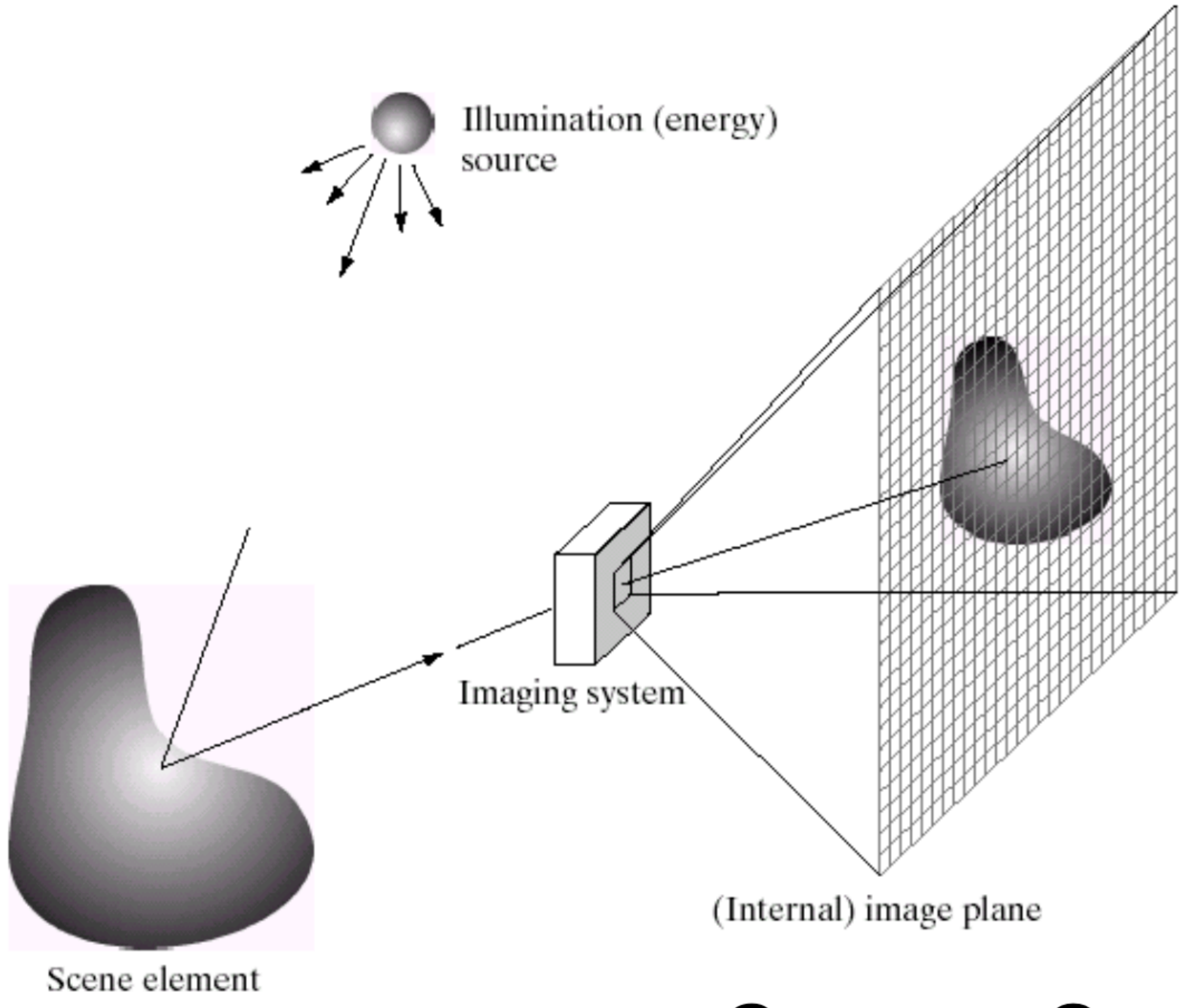
It can be any 3D signal with physical meaning
(ie, temperature, pressure, etc)

Sampling

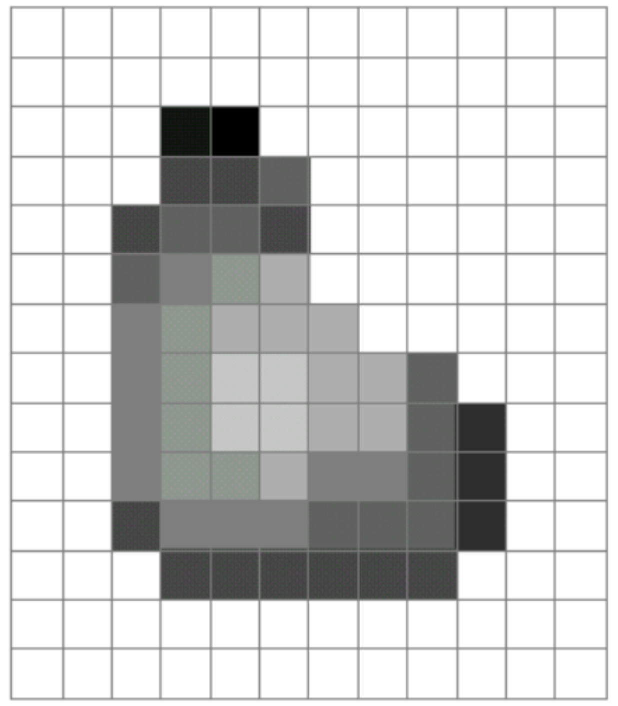
- Sampling in 1D takes a function, and returns a vector whose elements are values of that function at the sample points.



Sampling in Image Formation



Camera Sensor



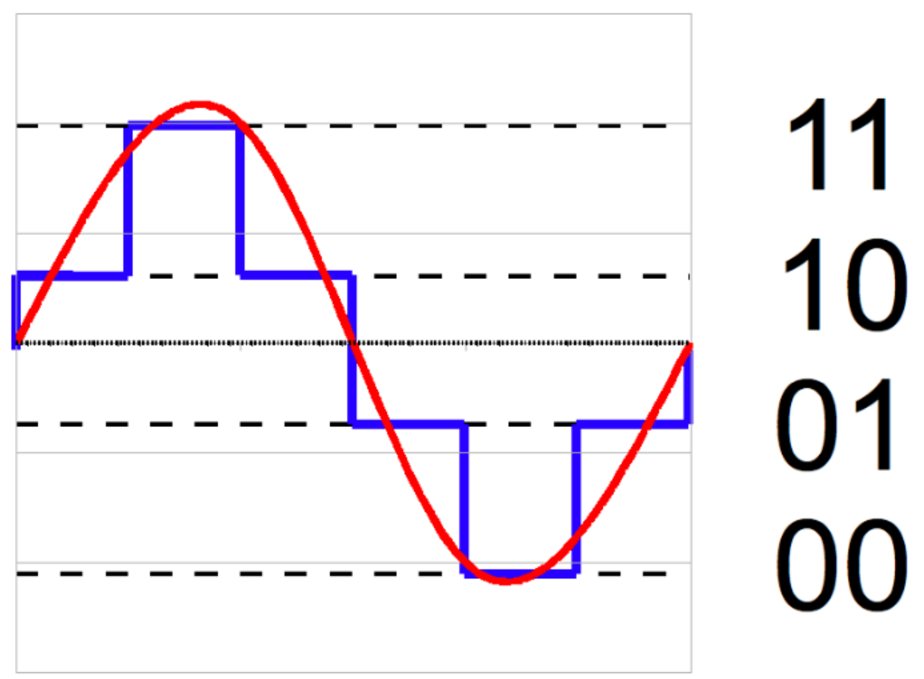
Output Image

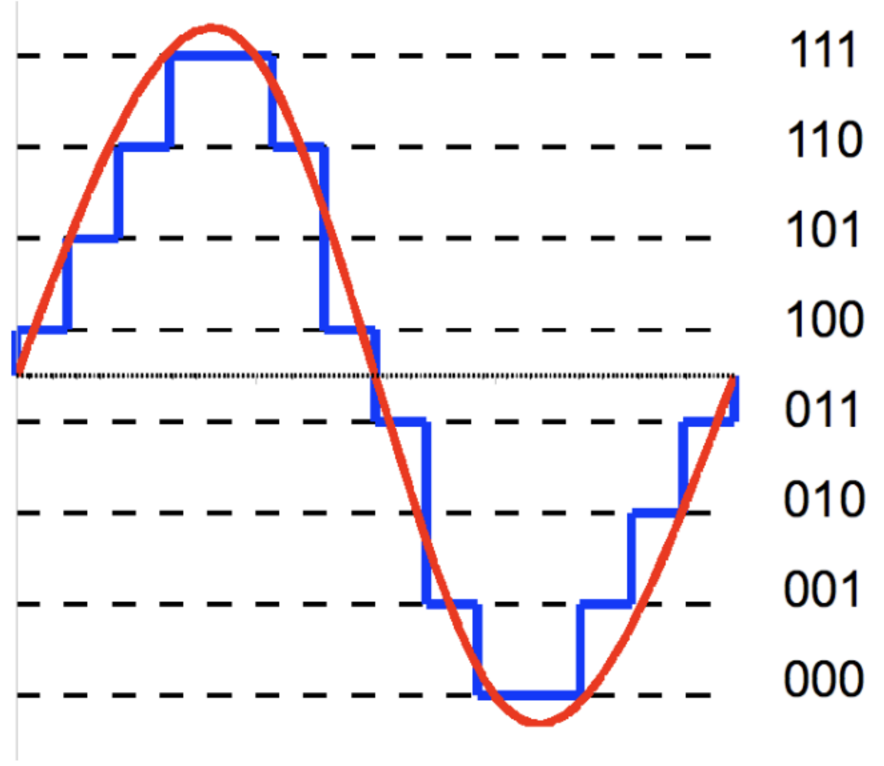
[James Hays]

Quantization

The measurement at each pixel is typically quantized to store the continuous value compactly.

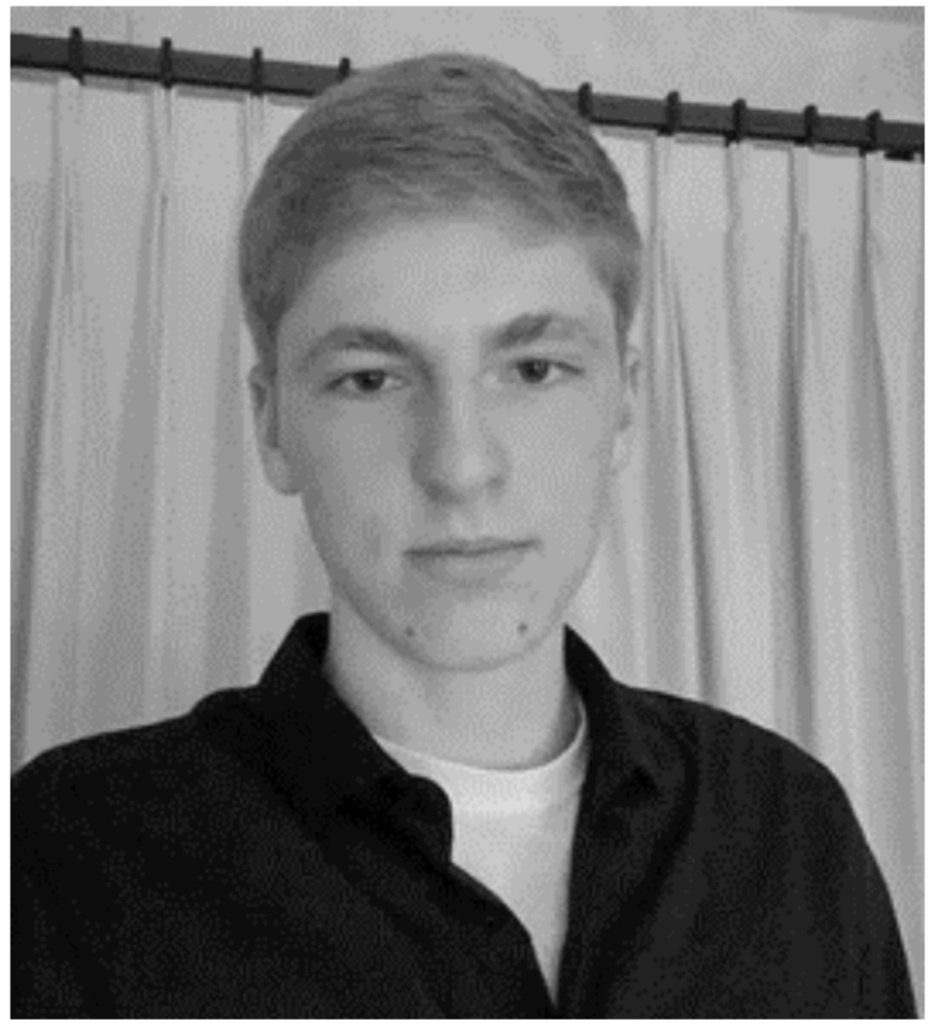
(Quantize: Approximate continuous by discrete.)





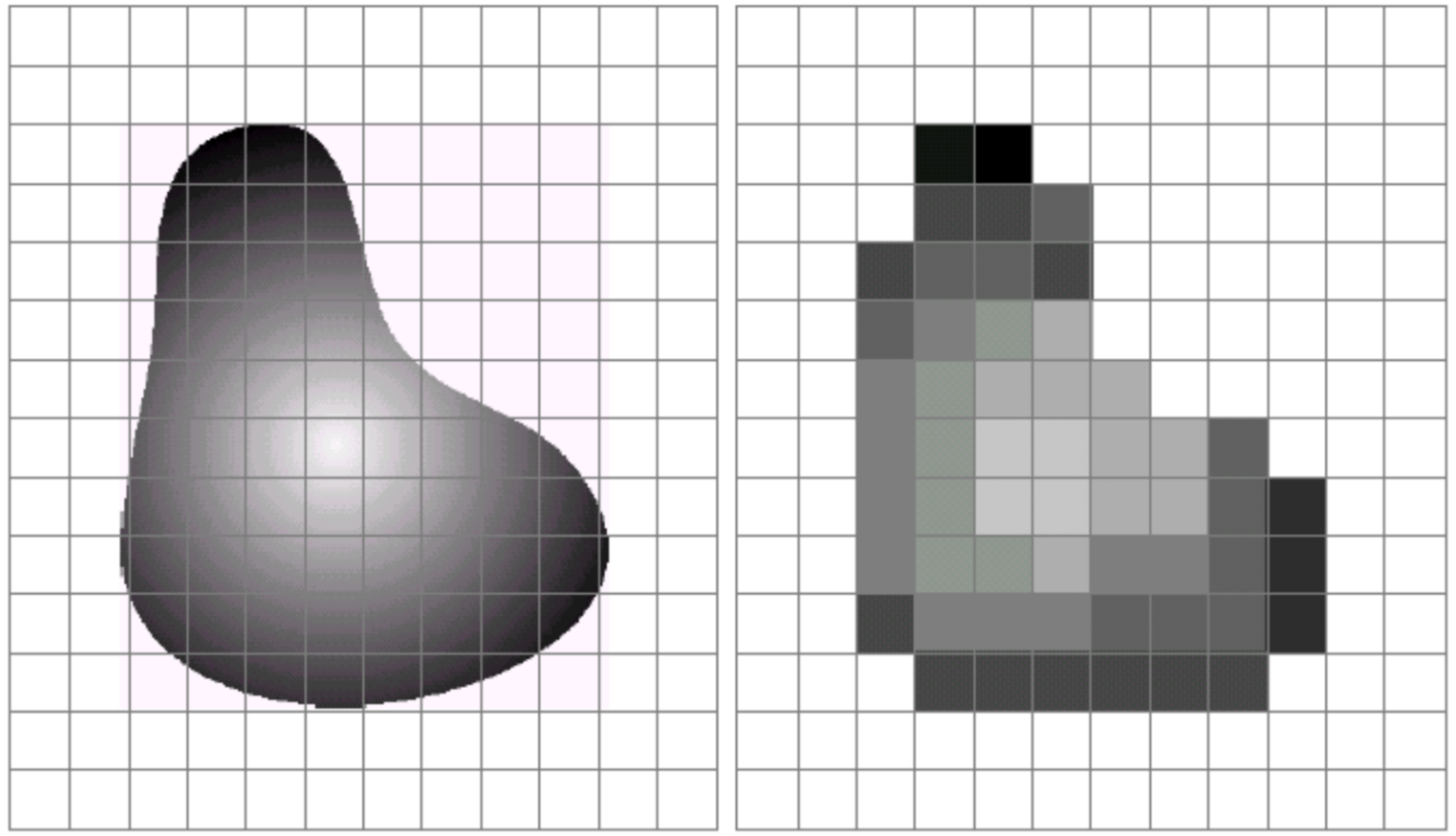
[Figures: Wikipedia]

Quantization Effects



8 bit – 256 levels 4 bit – 16 levels 2 bit – 4 levels 1 bit – 2 levels

Fidelity



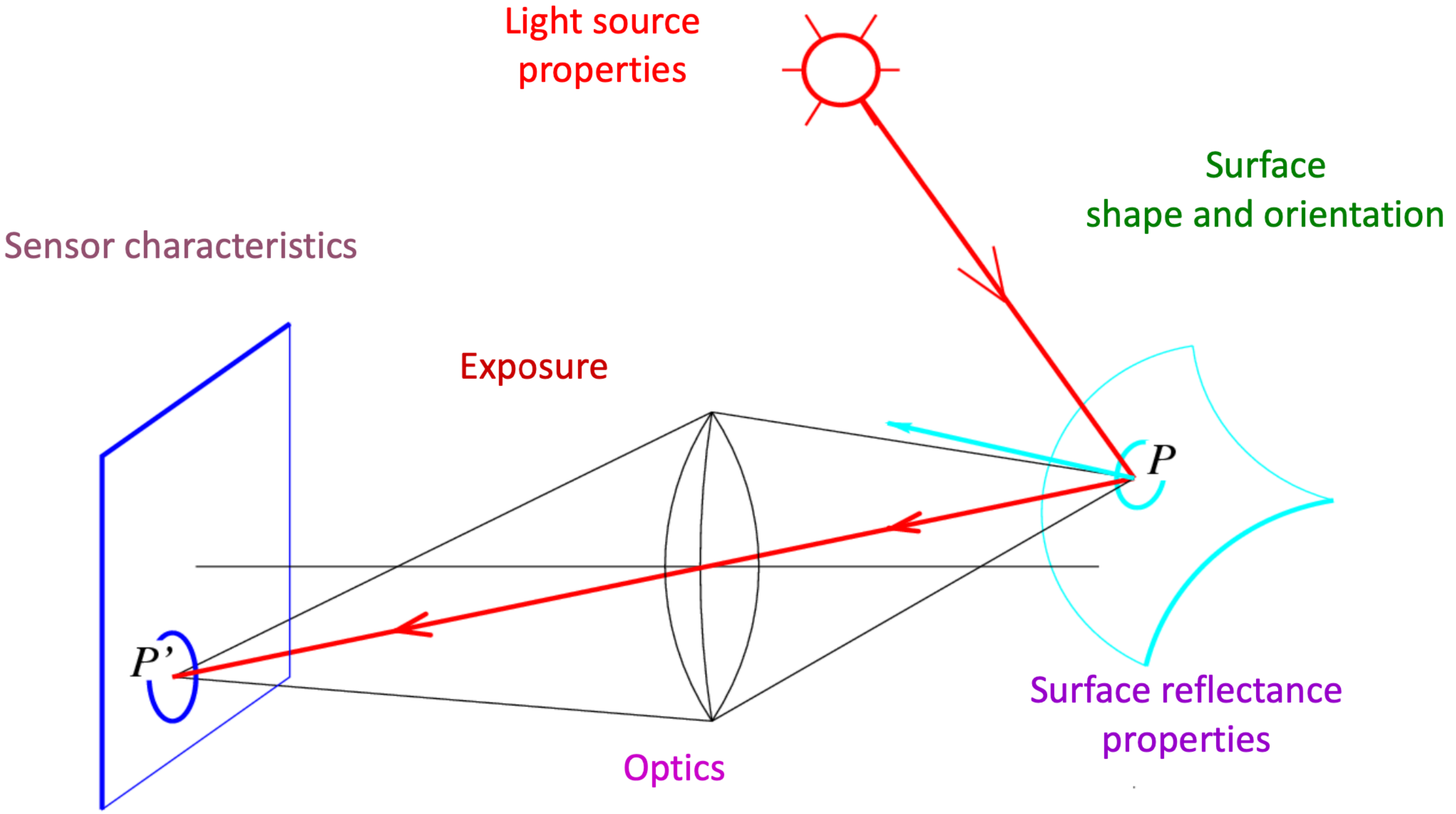
- Summary:
  - When we capture an image we sample light intensity in space (and time)
  - And quantize the result.
- Fidelity depends on sampling density and quantization strength
  - (Spatial resolution and radiometric resolution)

Plan for today

- Image representation
- Light
- Color
- Cameras (Orthographic, Perspective)

Image Formation

- What factors determine the brightness of an image pixel?



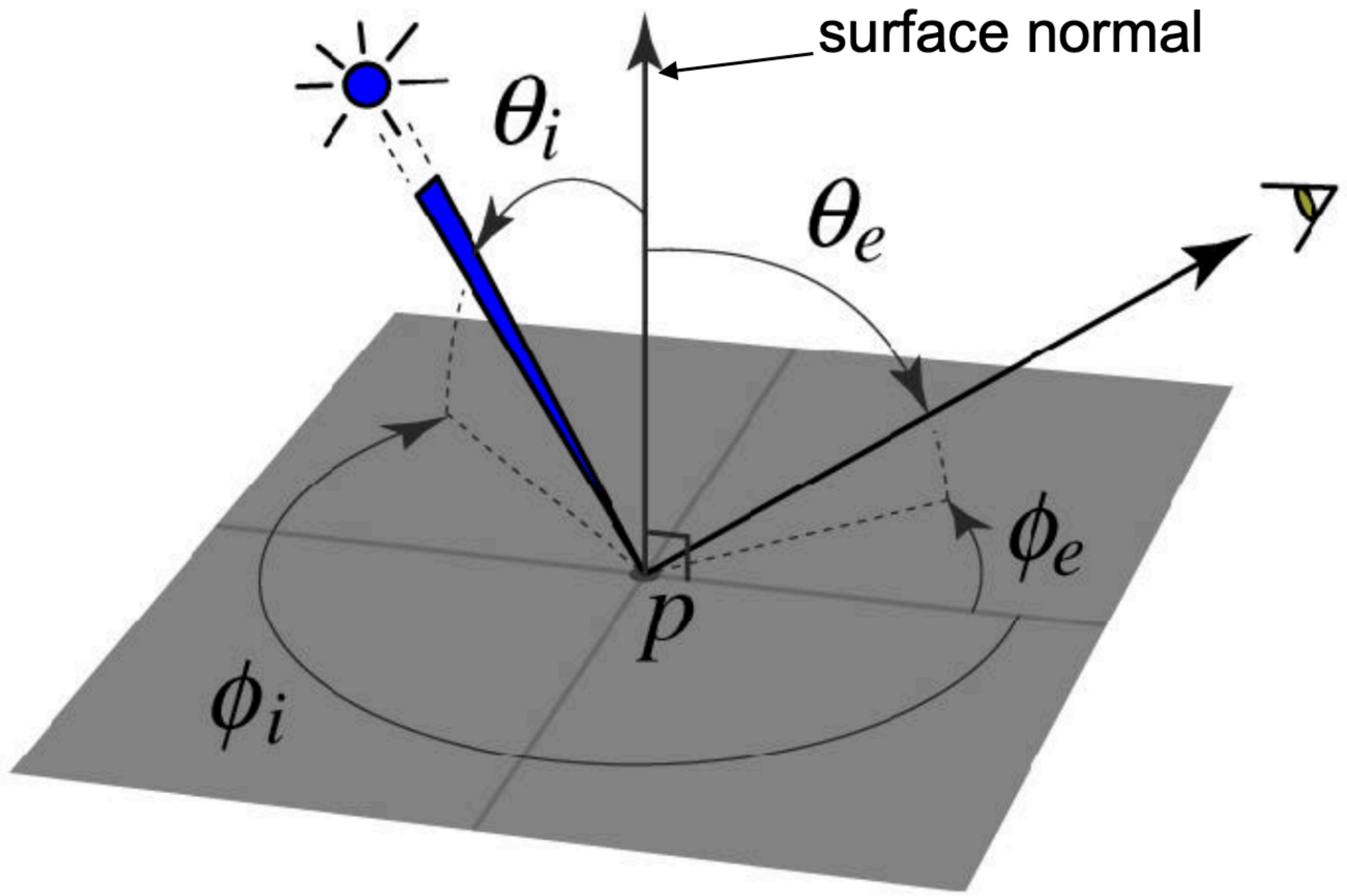
[Slide by L. Fei-Fei]

Interaction of Light and Matter

- What happens when a light ray hits an object?
  - Some light gets absorbed
    - converted to other forms of energy (e.g., heat)
  - Some gets transmitted through the object
    - possibly bent, through "refraction"
  - Some gets reflected
    - This is what a camera observes

BRDF

- The Bidirectional Reflection Distribution Function
  - Given an incoming ray (θ_i, ϕ_i) and outgoing ray (θ_e, ϕ_e) what proportion of the incoming light is reflected along outgoing ray?



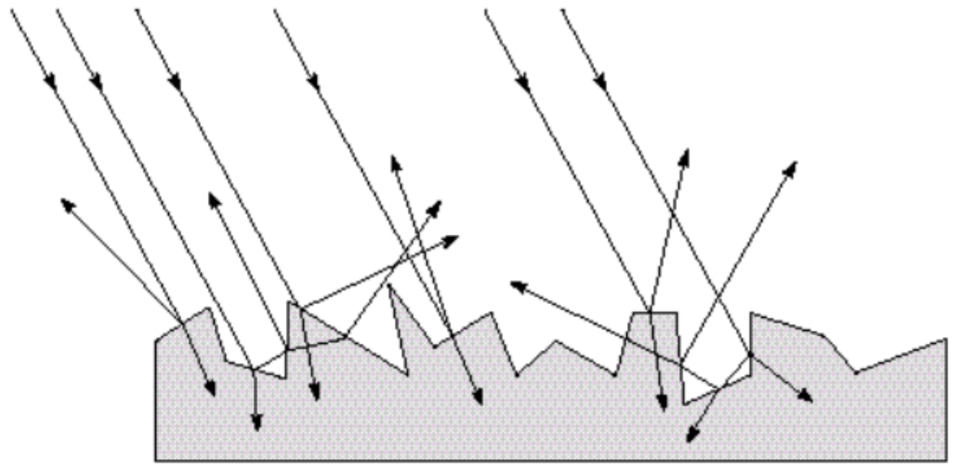
Answer given by the BRDF: $\rho(\theta_i, \phi_i, \theta_e, \phi_e)$

- Different kinds of surfaces have different BRDFs.

BRDFs Can Be Complicated…

Diffuse (Lambertian) Reflection

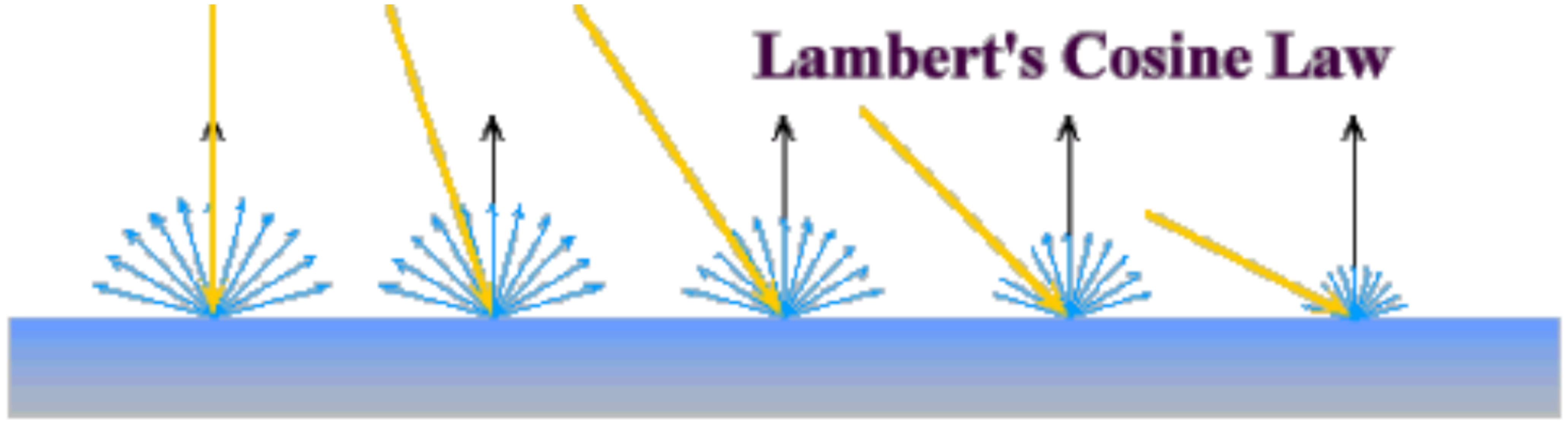
- Light is reflected equally in all directions
  - Dull, matte surfaces like chalk or latex paint
  - Microfacets scatter incoming light randomly
  - Effect is that light is reflected equally in all directions
  - Common assumption in computer vision
- Brightness of the surface depends on the incidence of illumination



Diffuse (Lambertian) Reflection

Lambertian surface BRDF:

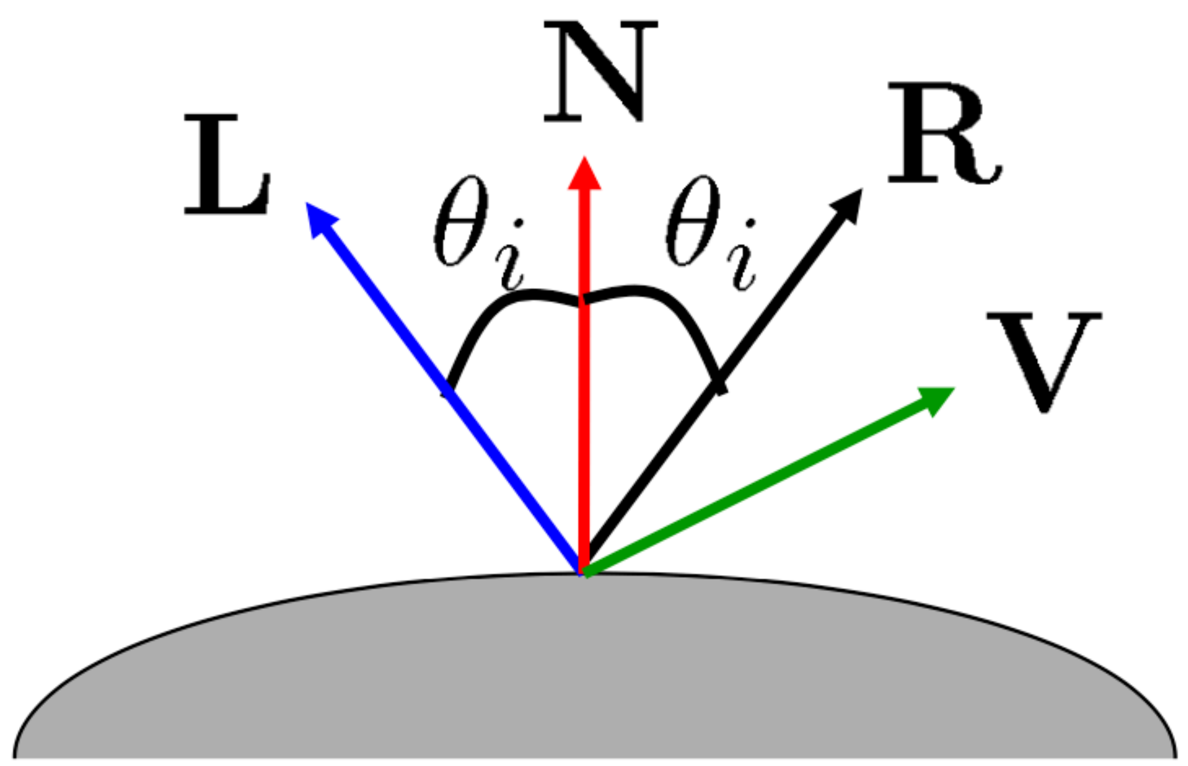
$$B = \rho(\mathbf{N} \cdot \mathbf{S})$$
$$= \rho \|\mathbf{S}\| \cos\theta$$



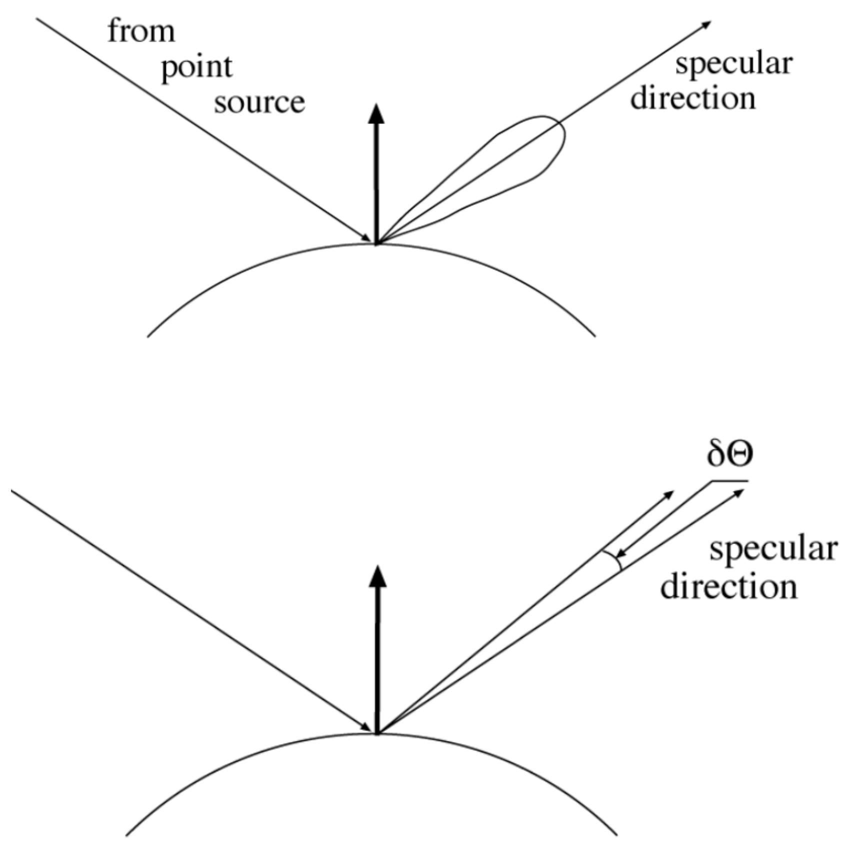
B: radiosity (total power leaving the surface per unit area)
ρ: albedo (fraction of incident irradiance reflected by the surface)
N: unit normal
S: source vector (magnitude proportional to intensity of the source)

Specular Reflection

For a perfect mirror, light is reflected about the normal **N**



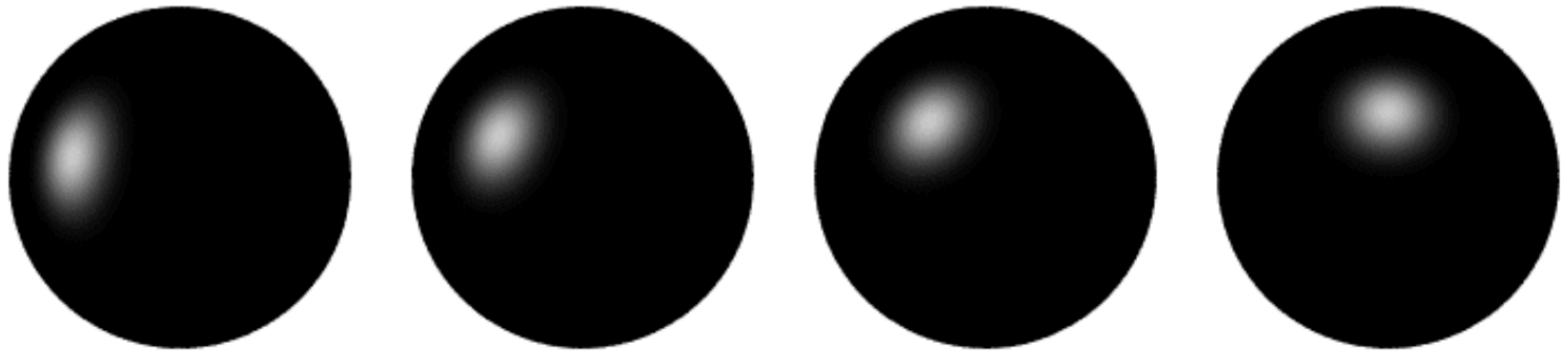
$$I_e = \begin{cases} I_i & \text{if } \mathbf{V} = \mathbf{R} \\ 0 & \text{otherwise} \end{cases}$$



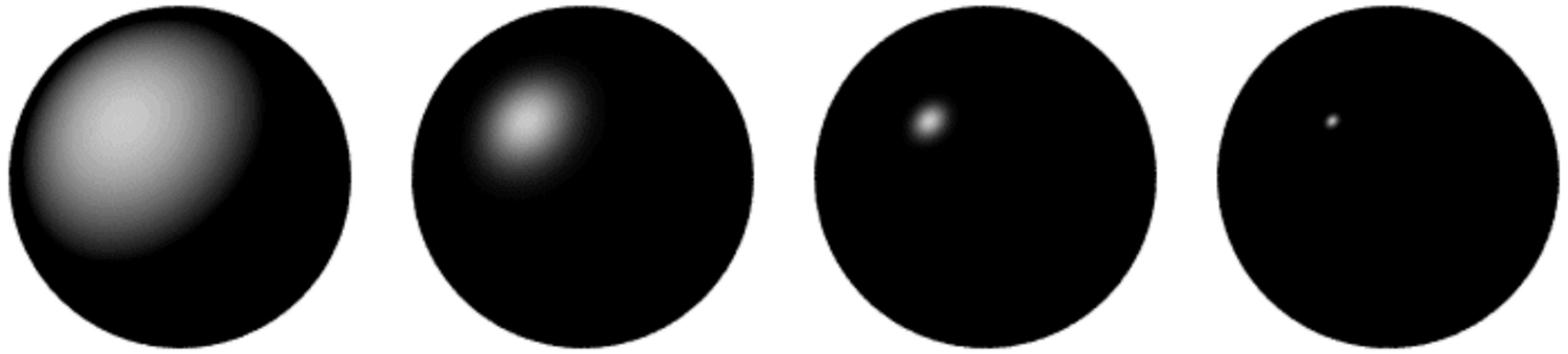
Near-perfect mirrors have a **highlight** around **R**

- common model: $$I_e = k_s(\mathbf{V} \cdot \mathbf{R})^{n_s} I_i$$

Specular Reflection

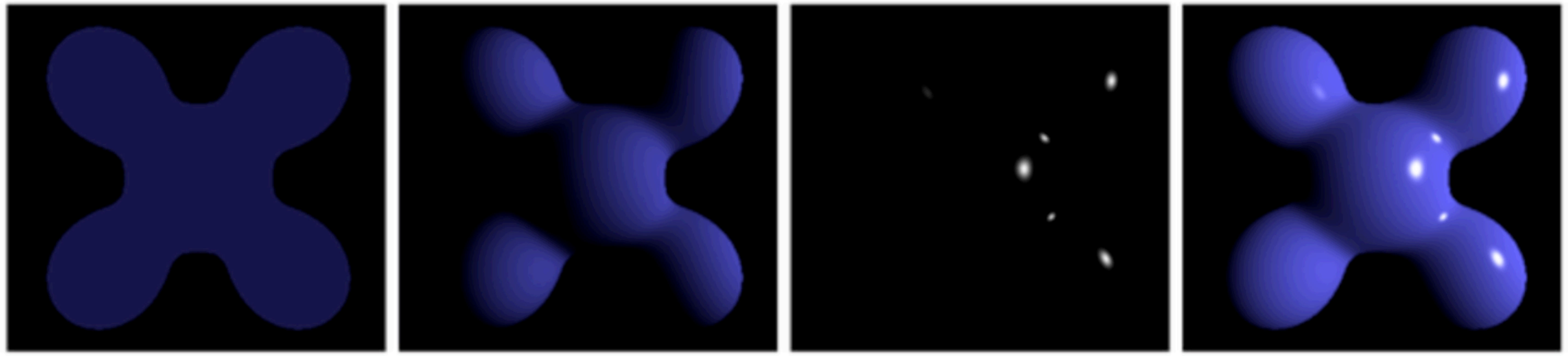


Moving the light source



Changing the exponent

Phong Reflection Model

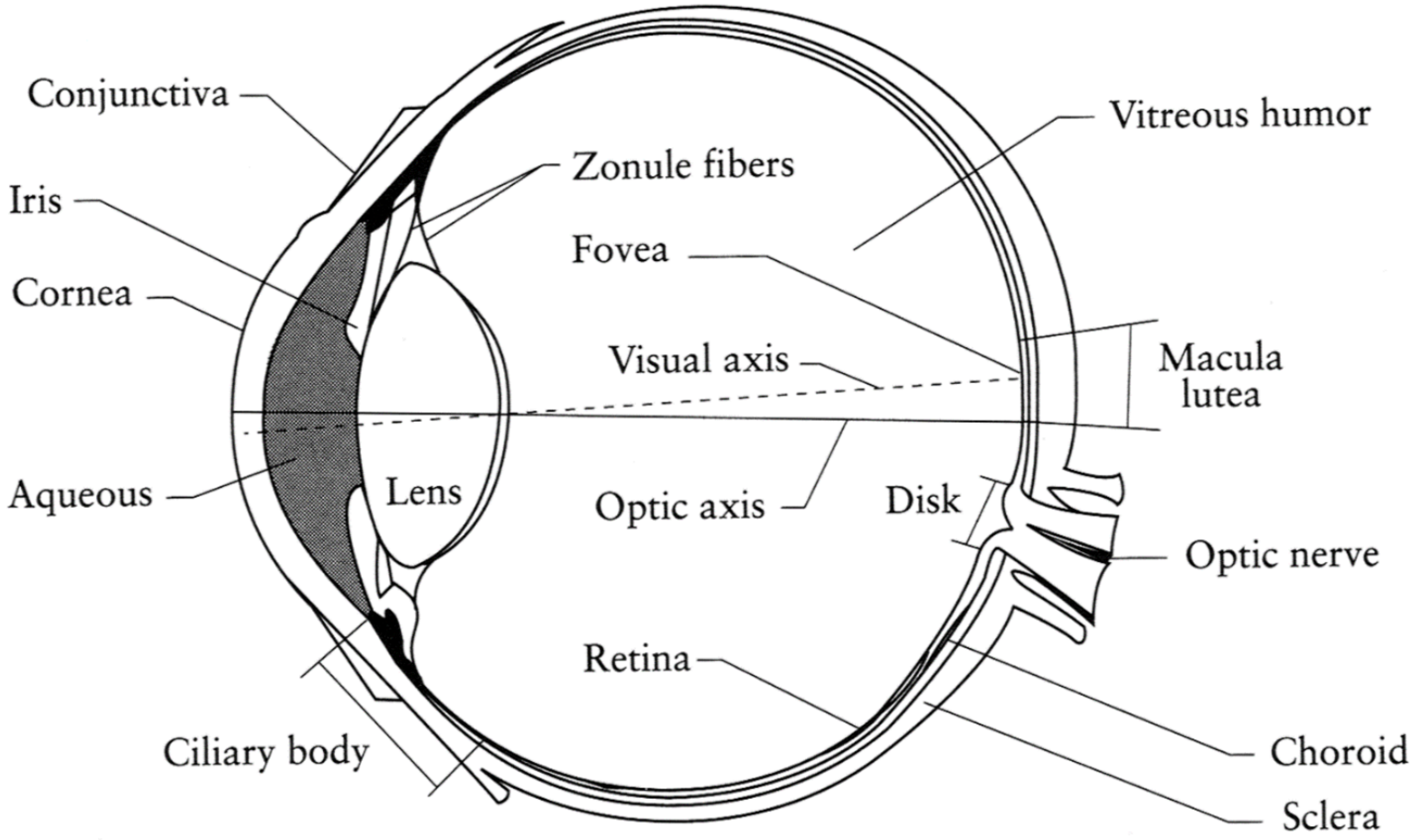


Ambient + Diffuse + Specular = Phong Reflection

Plan for today

- Image representation
- Light
- Color
- Cameras (Orthographic, Perspective)

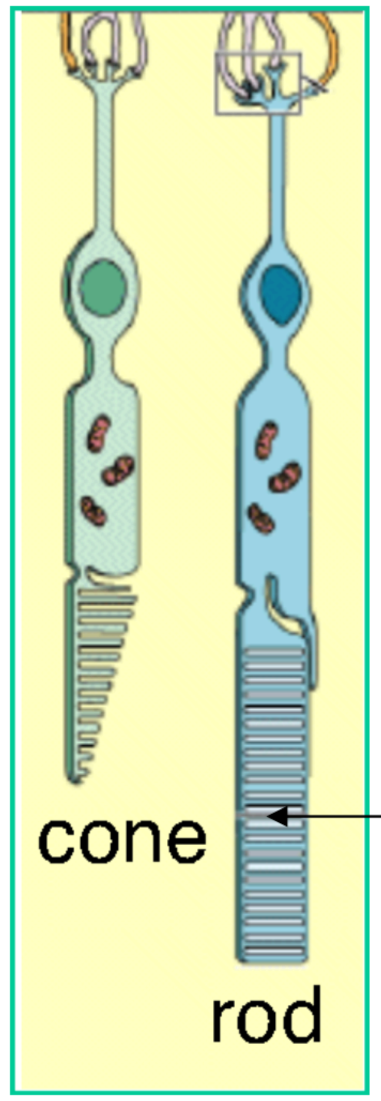
Color for Humans

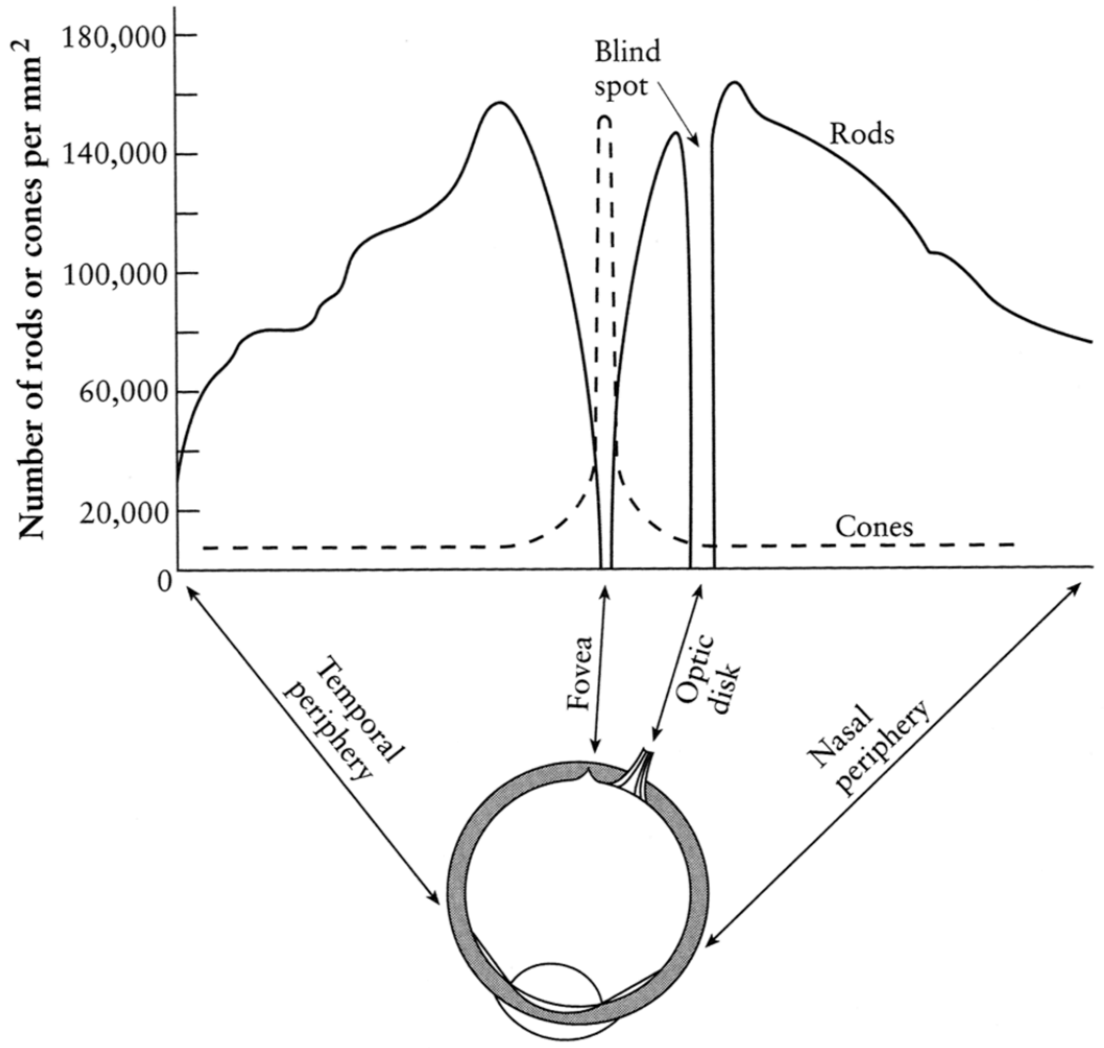


- The human eye is a camera
  - **Iris** - colored annulus with radial muscles
  - **Pupil** - the hole (aperture) whose size is controlled by the iris
  - What's the sensor?
    - photoreceptor cells (rods and cones) in the **retina**

Slide by Steve Seitz

Color for Humans

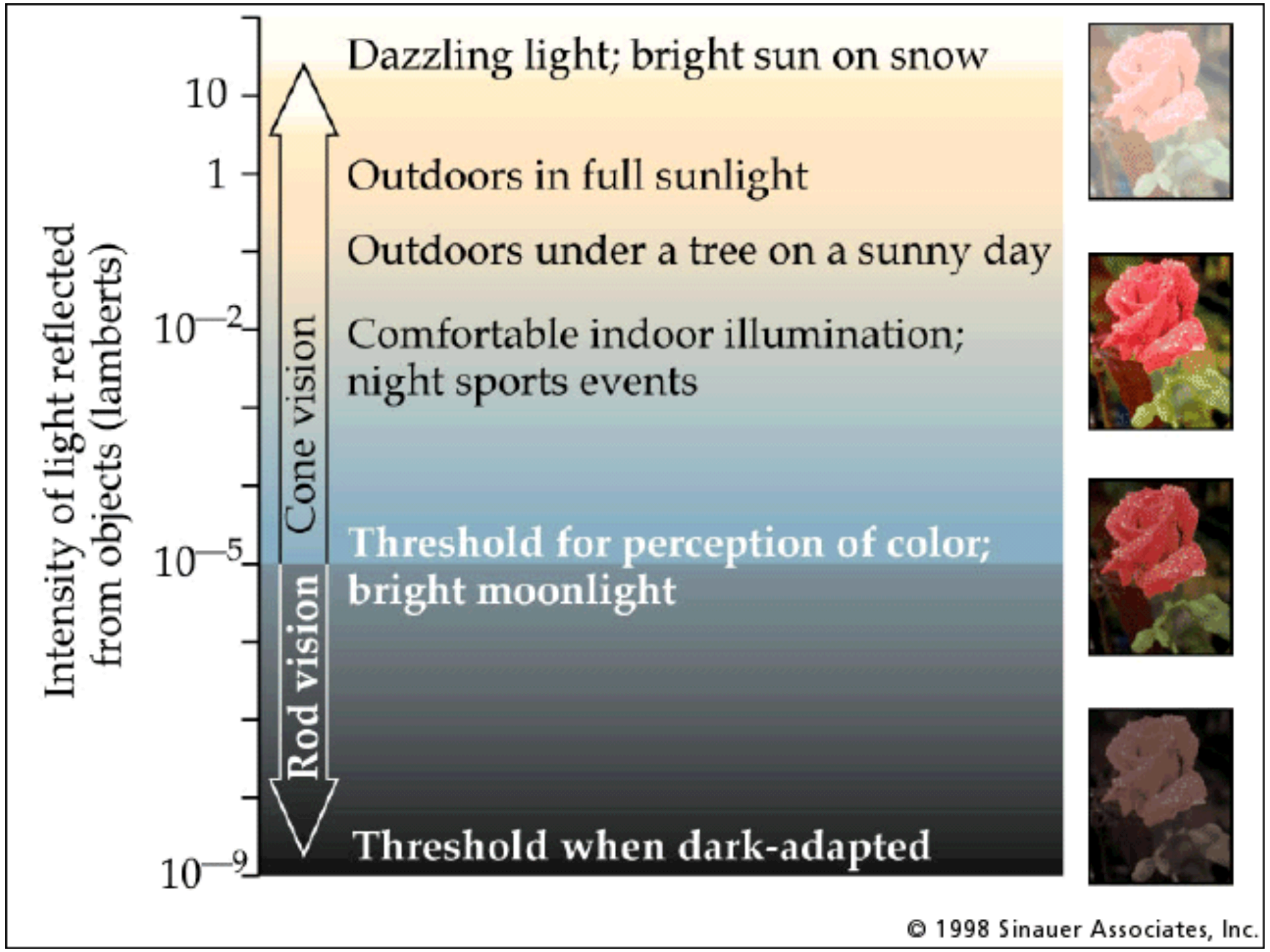




- **Rods**: Intensity perception. Highly sensitive, operate at night, grey-scale vision.
- **Cones**: Color perception. Less sensitive, operate in high light, color vision.
- Rods and Cones are **non-uniformly** distributed on the retina.
    - Less acuity in the periphery.
- Fovea - Small region (1 or 2°) at the center of the visual field containing the highest density of cones – and no rods

[Steve Seitz]

Color for Humans

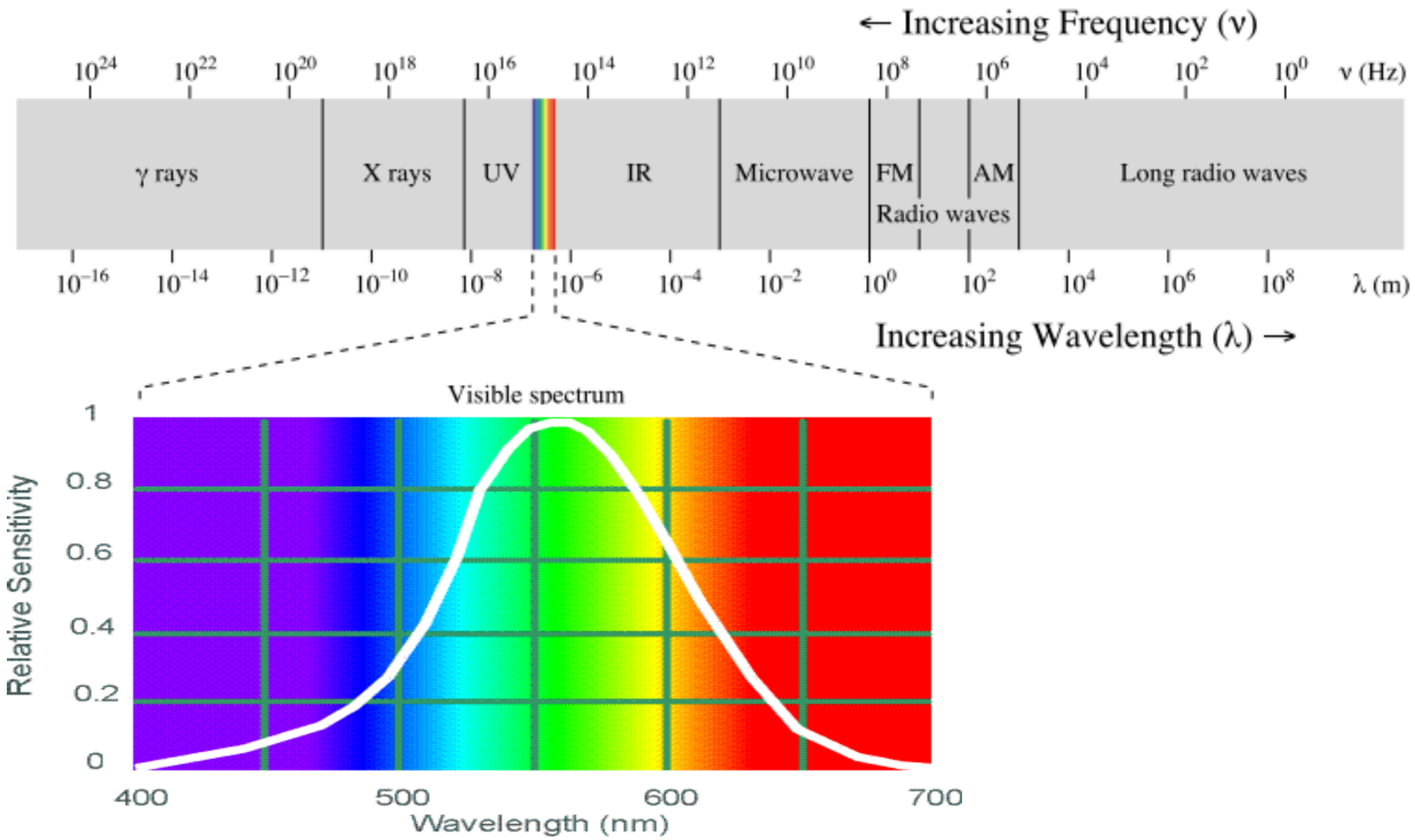


Color for Humans

- Our visual system has a large *dynamic range*
  - We can resolve both light and dark things at the same time
  - One mechanism is that we sense light intensity on a *logarithmic scale*
    - an exponential intensity ramp will be seen as a linear ramp
  - Another mechanism is *adaptation*
    - rods and cones adapt to be more sensitive in low light, less sensitive in bright light.

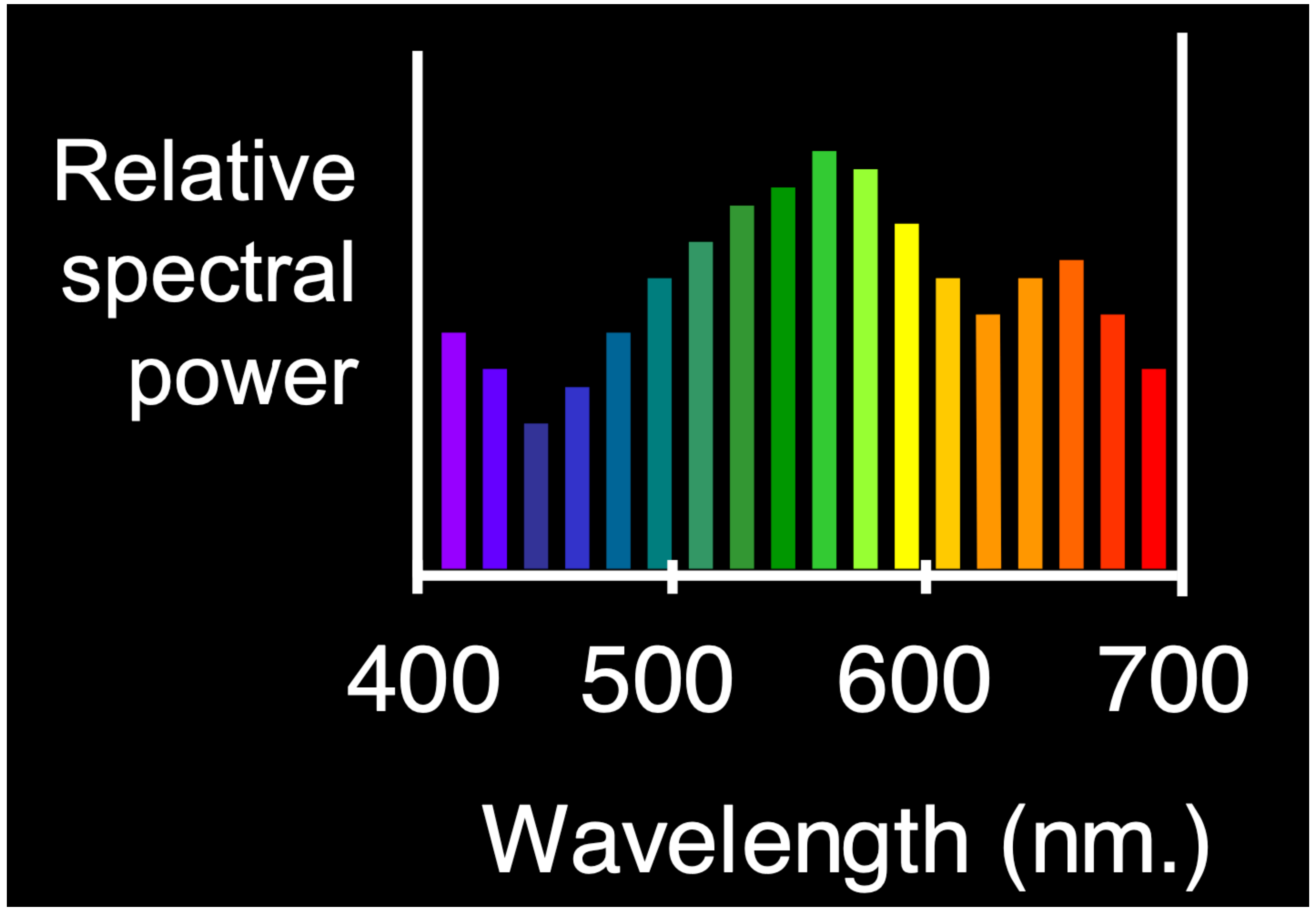
Background	Luminance (candelas per square meter)
Horizon sky	
Moonless overcast night	0.00003
Moonless clear night	0.0003
Moonlit overcast night	0.003
Moonlit clear night	0.03
Deep twilight	0.3
Twilight	3
Very dark day	30
Overcast day	300
Clear day	3,000
Day with sunlit clouds	30,000
Daylight fog	
Dull	300–1,000
Typical	1,000–3,000
Bright	3,000–16,000
Ground	
Overcast day	30–100
Sunny day	300
Snow in full sunlight	16,000

Electromagnetic Spectrum



Human Luminance Sensitivity Function

Describing a Source of Light

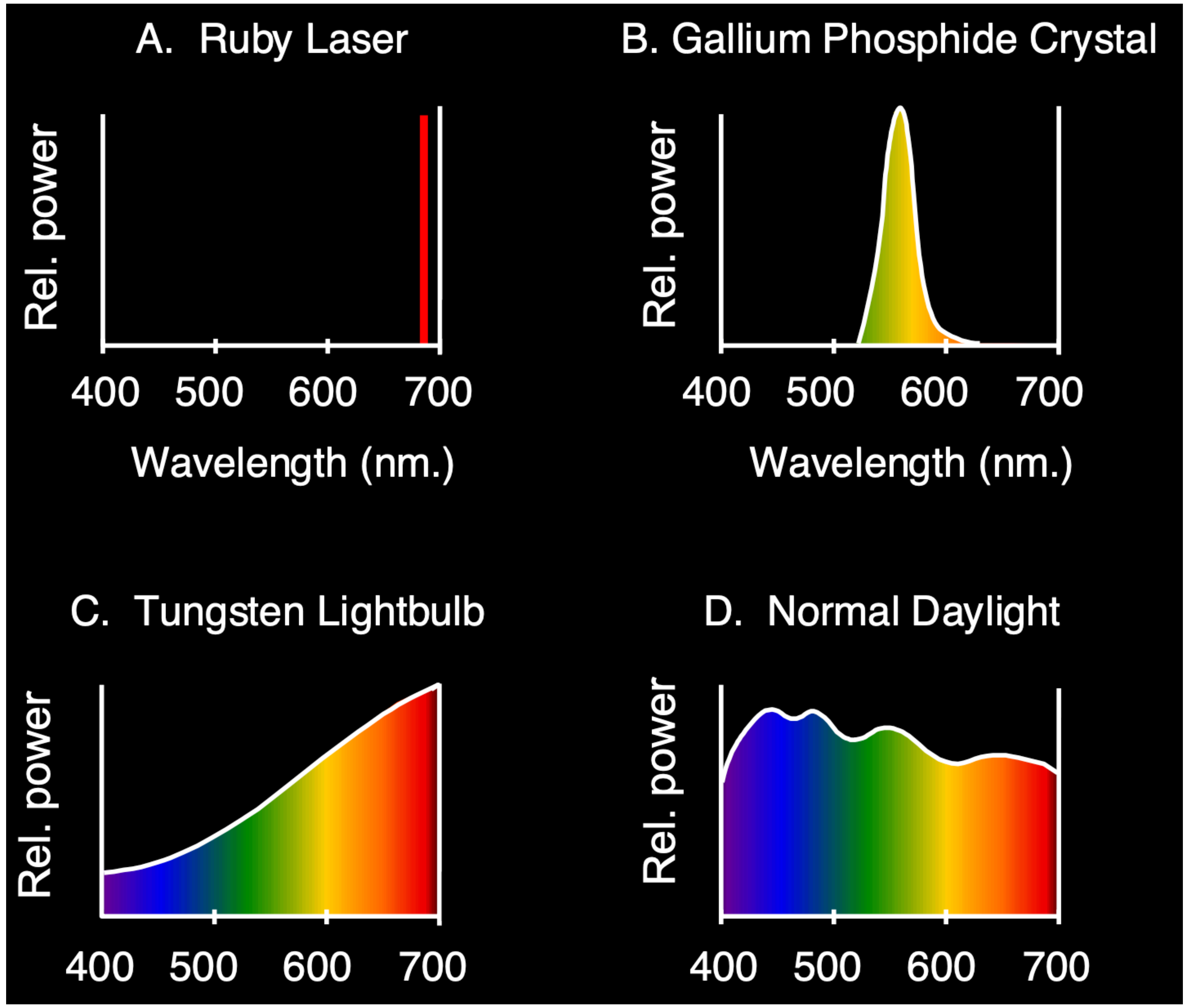


Any source of light can be completely described physically by its spectrum:

The amount of energy emitted (per time unit) at each wavelength 400-700nm

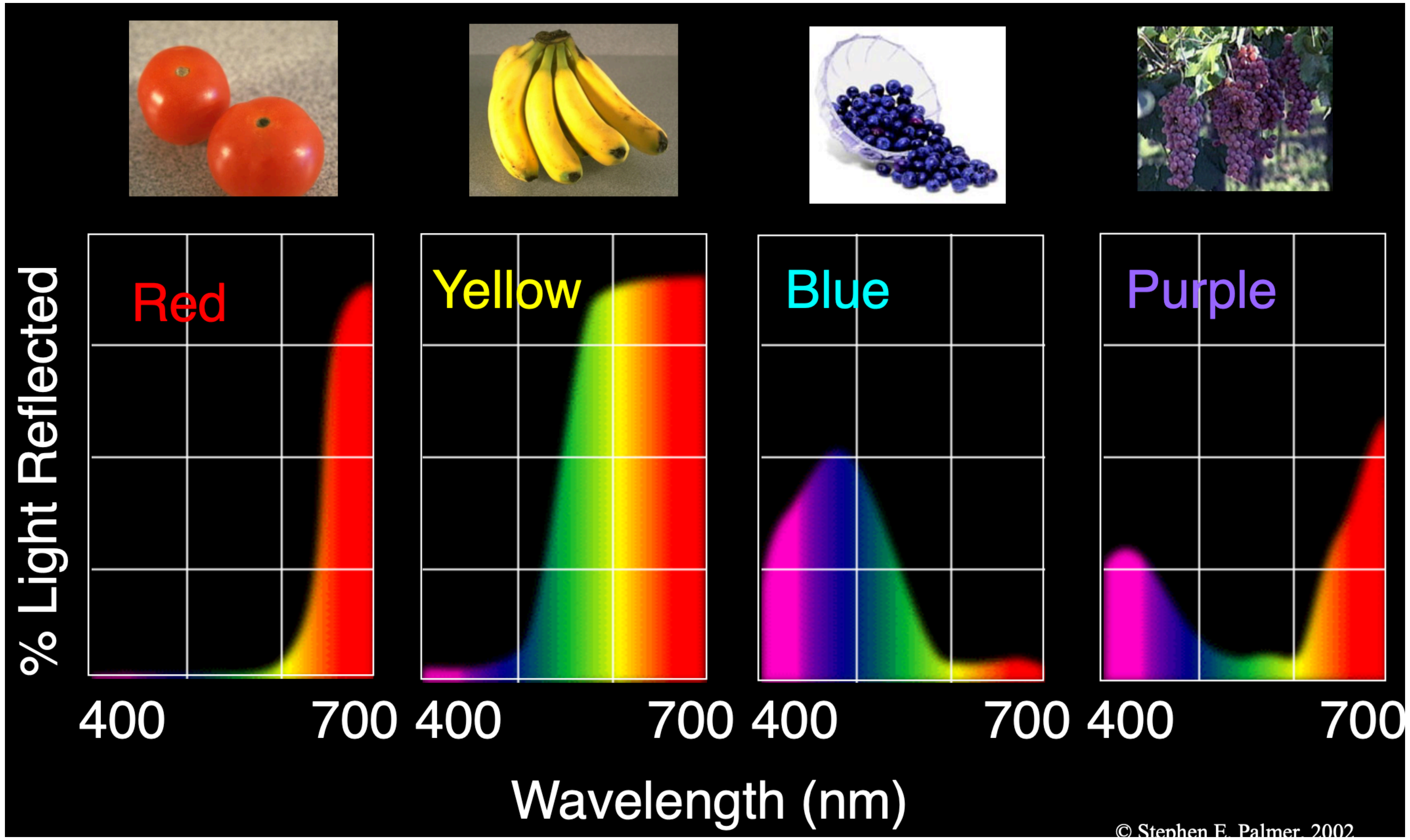
Spectra of Light Sources

Some examples of the spectra of light sources



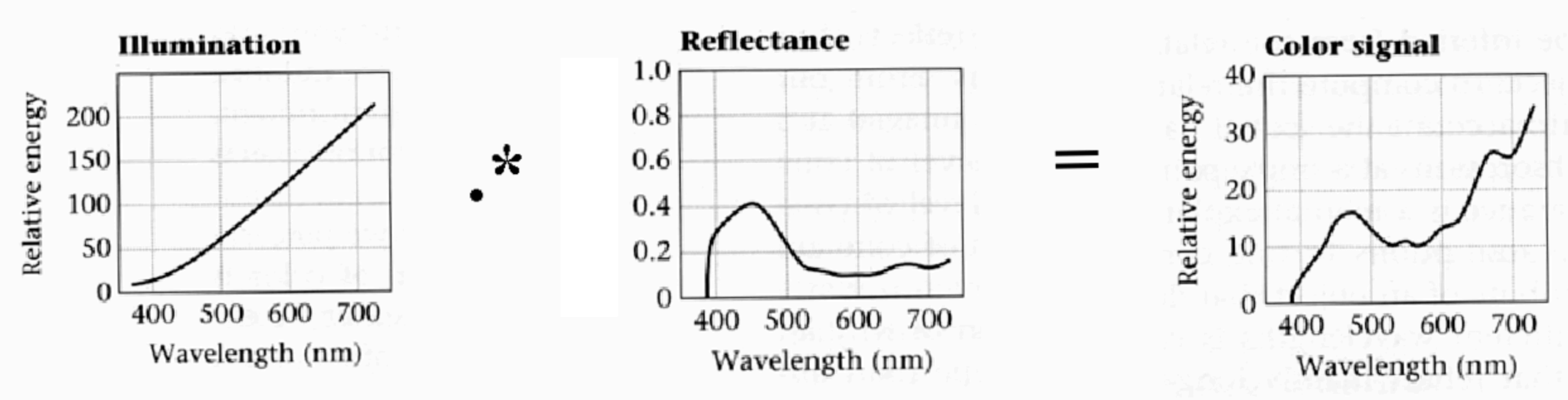
Spectra of Surfaces

Examples of the reflectance spectra of surfaces



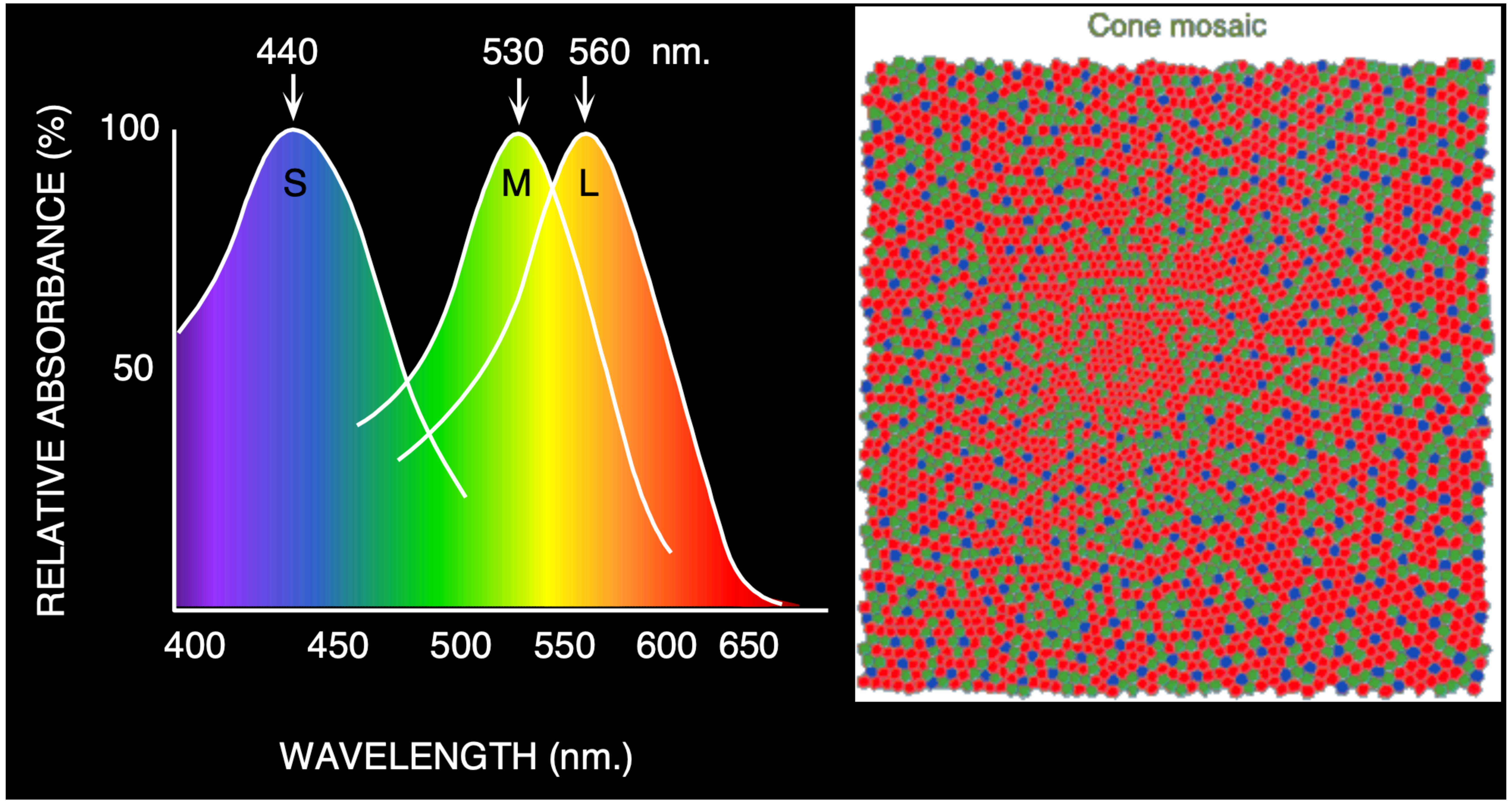
Interaction of Light and Surfaces

- Reflected color is the result of interaction of light source spectrum with surface reflectance

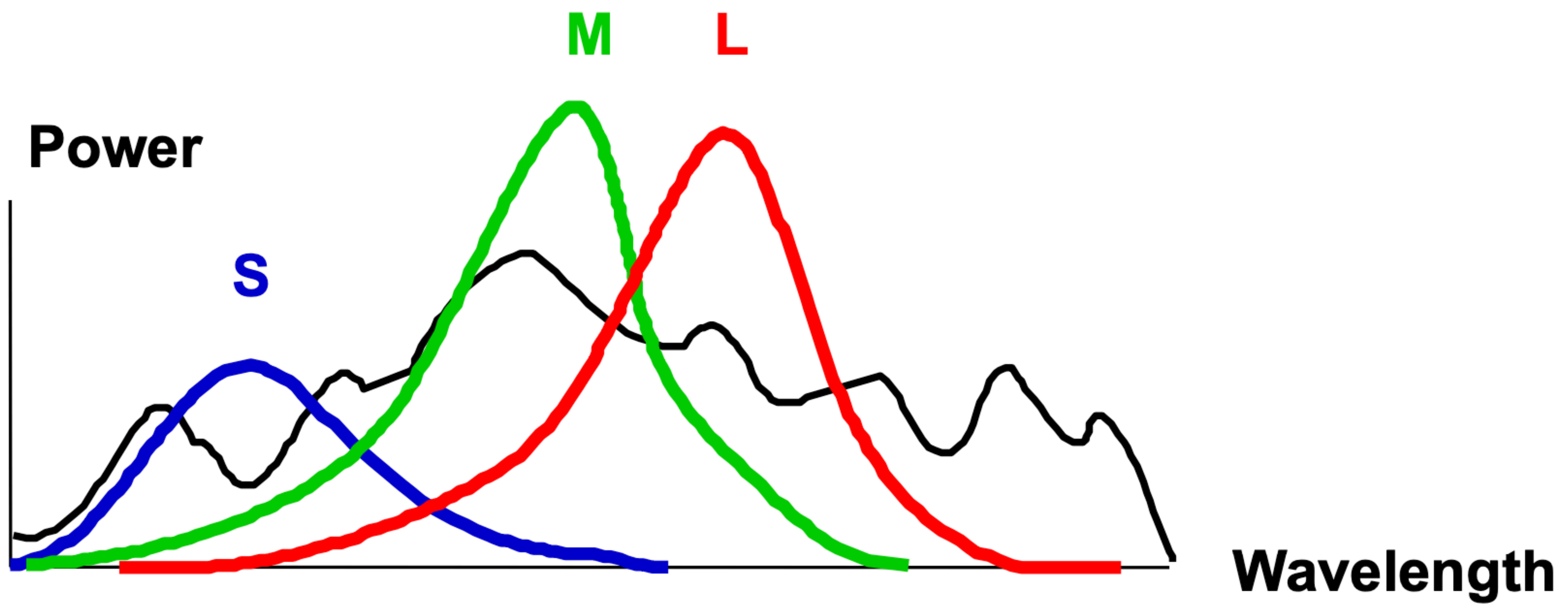


Back to Color in Humans

Humans have 3 types of cones:



Back to Color in Humans

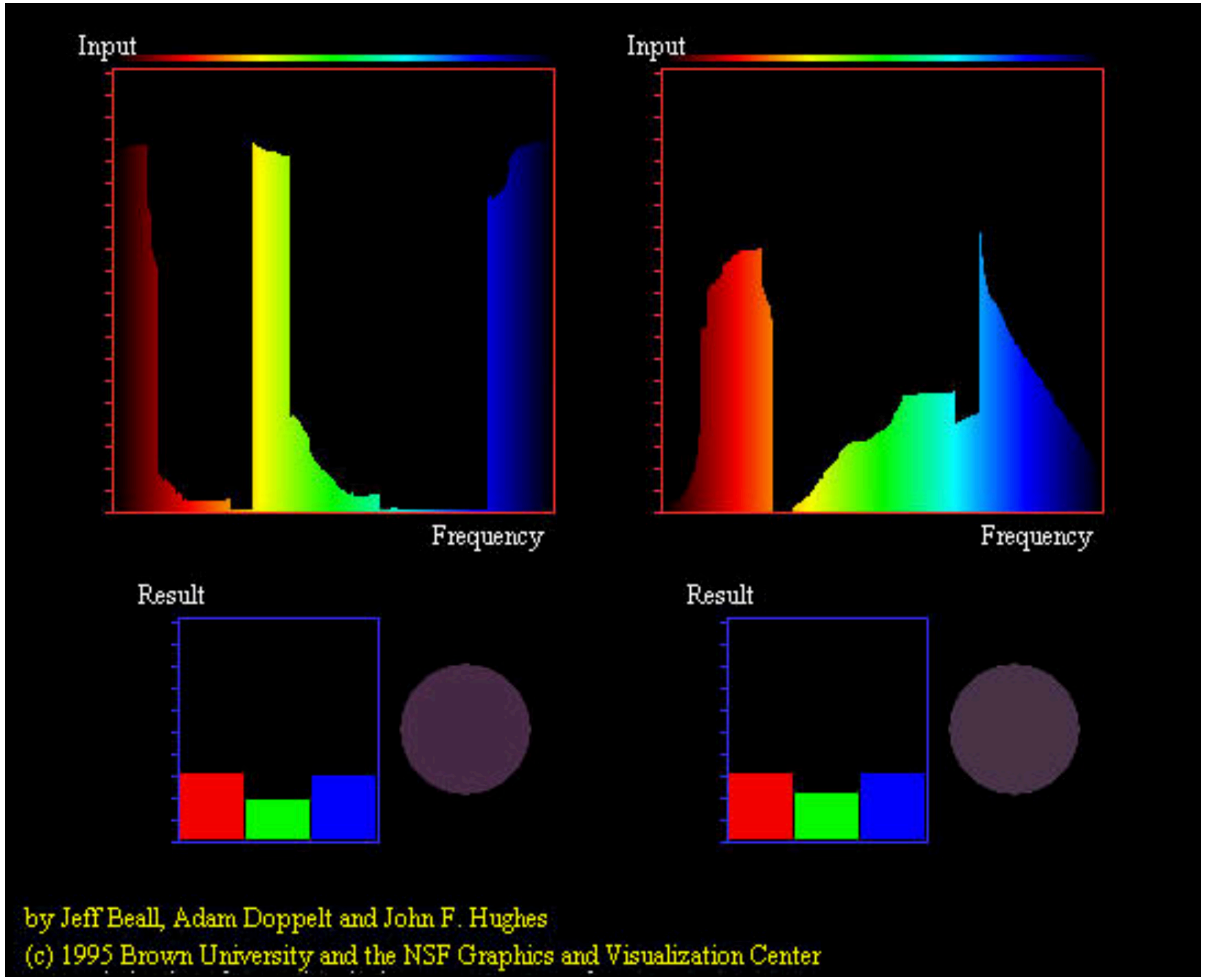


Rods and cones act as *filters* on the spectrum

- To get the output of a filter, multiply its response curve by the spectrum, integrate over all wavelengths
  - Each cone yields one number
  - => All colors are reduced to three real numbers.
- How can we represent an entire spectrum with 3 numbers?
- We can't! Most of the information is lost
  - As a result, two different spectra may appear indistinguishable
    - » such spectra are known as **metamers**

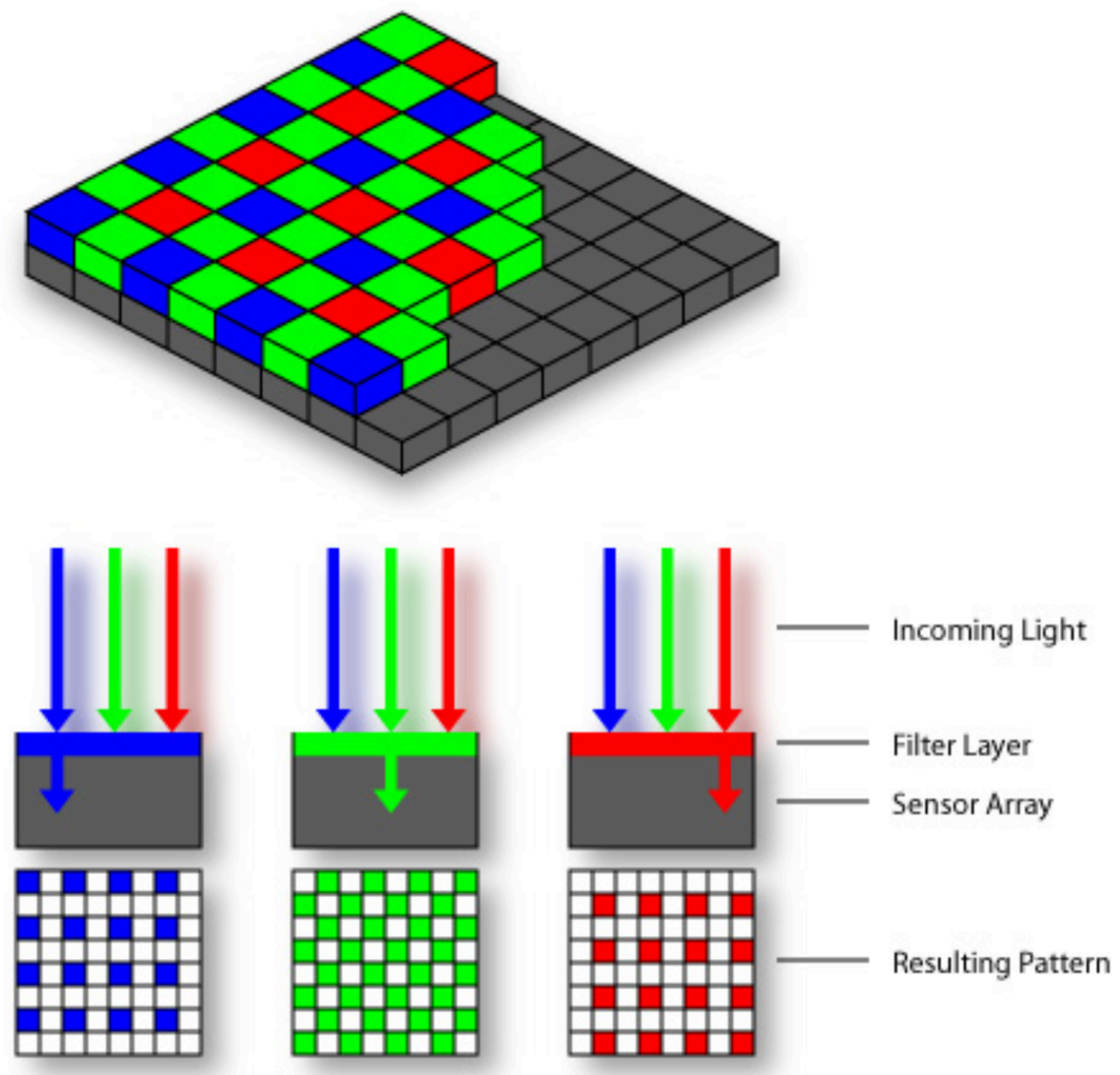
Slide by Steve Seitz

Metamers


Input
Frequency
Input
Frequency
Result
Result
by Jeff Beall, Adam Doppelt and John F. Hughes
(c) 1995 Brown University and the NSF Graphics and Visualization Center

Color in Cameras

- 1-chip: Bayer filter

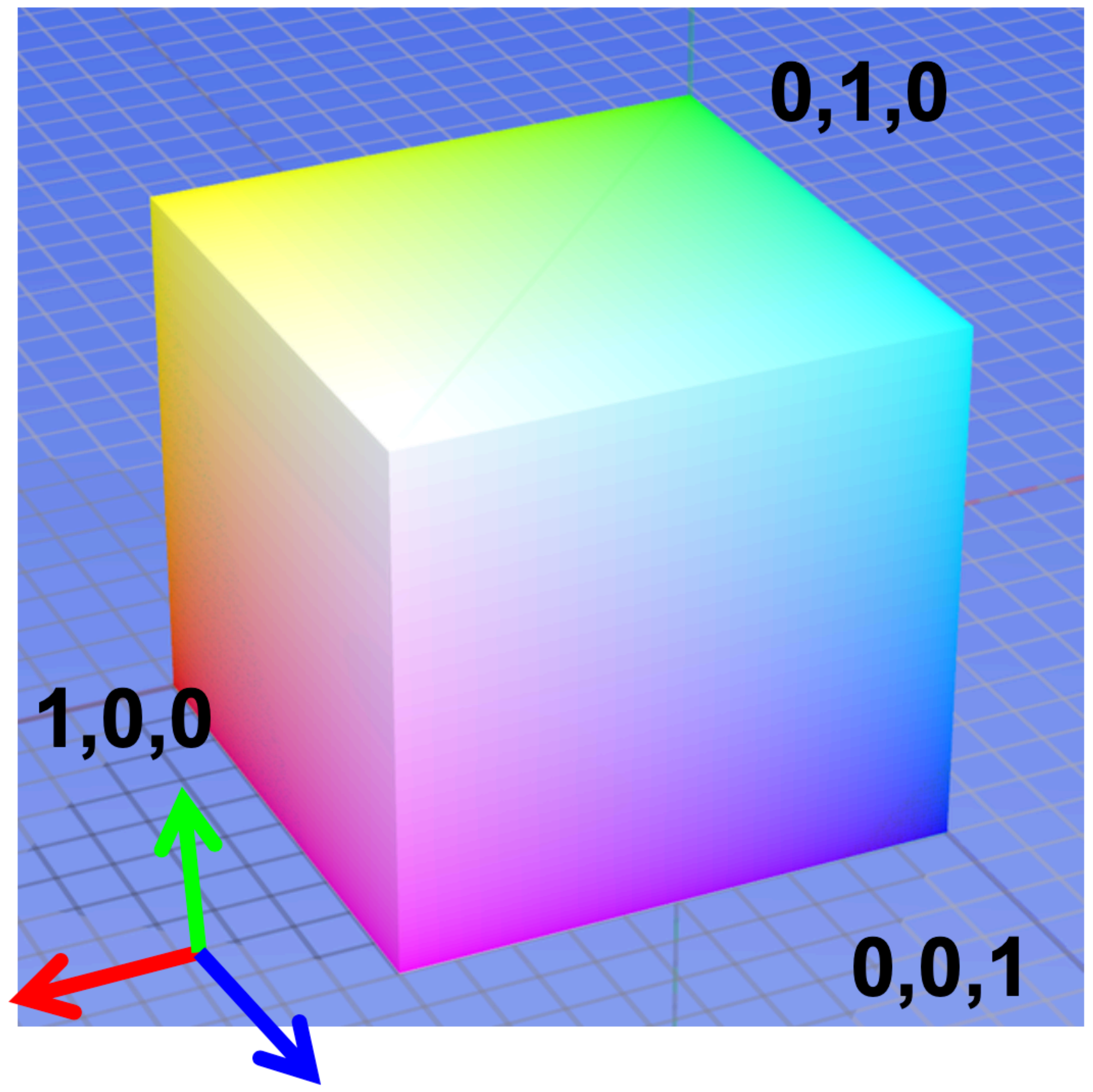


- Estimate RGB at ‘G’ cells from neighboring values

Slide by Steve Seitz

Color in Computers: RGB Space

Default color space



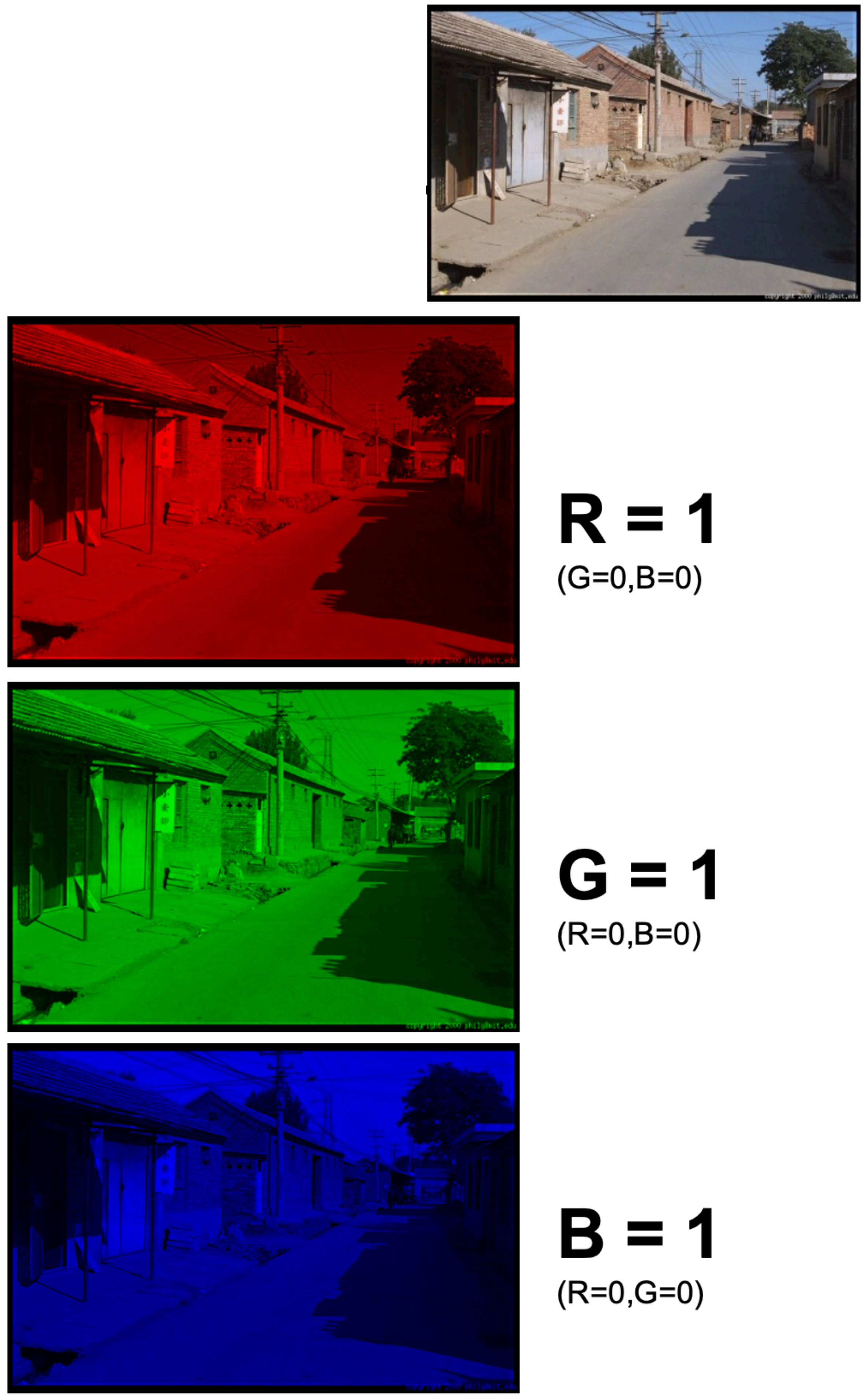
Any color: C = r*R + g*G + b*B

•Easy for devices

•Non-perceptual

- Strongly correlated channels
- Where is hue and saturation?

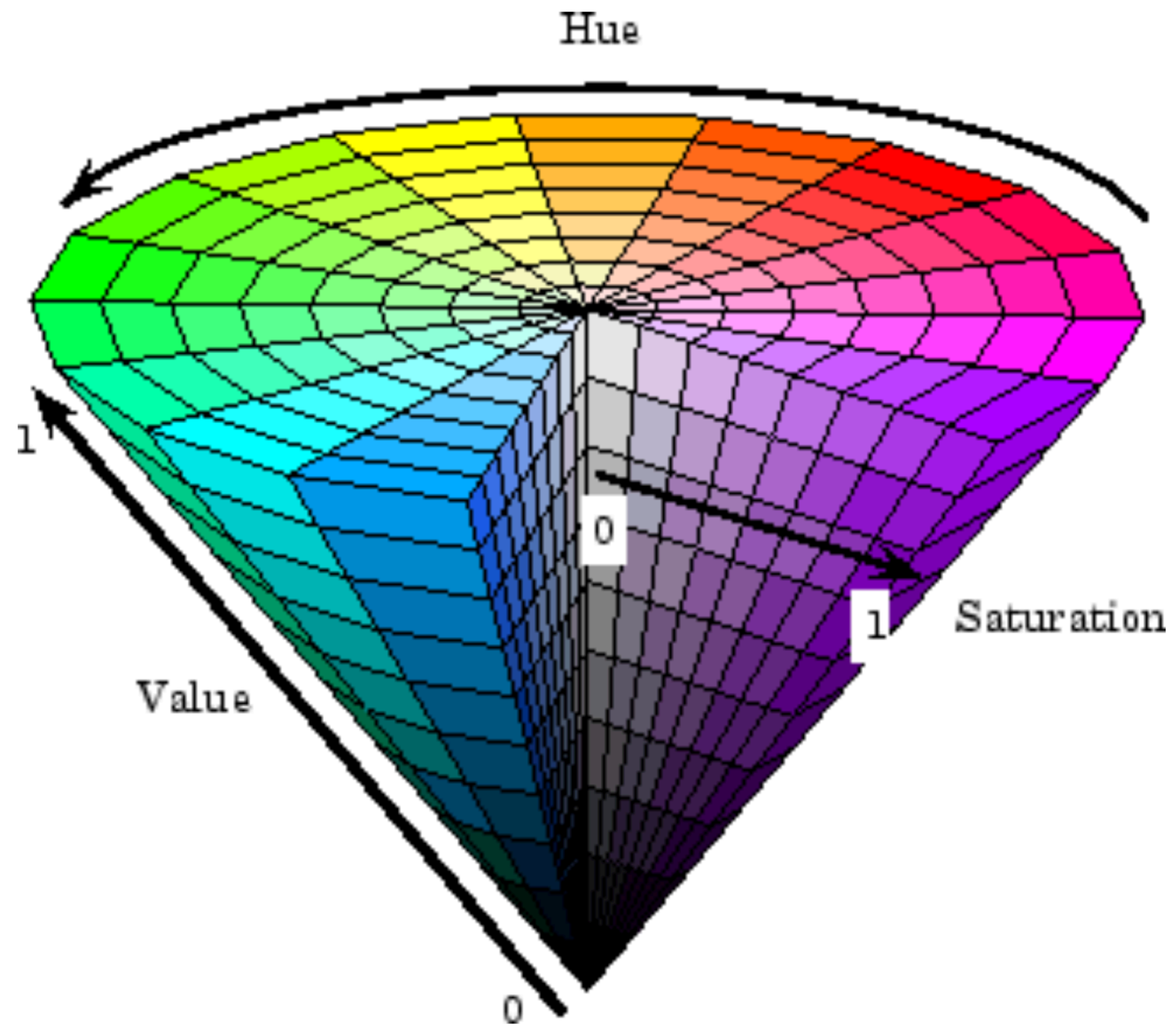
Linearly combine R, G, B to produce other colors.



[Cube Image: Wikipedia]

Color in Computers: HSV Space

Intuitive color space



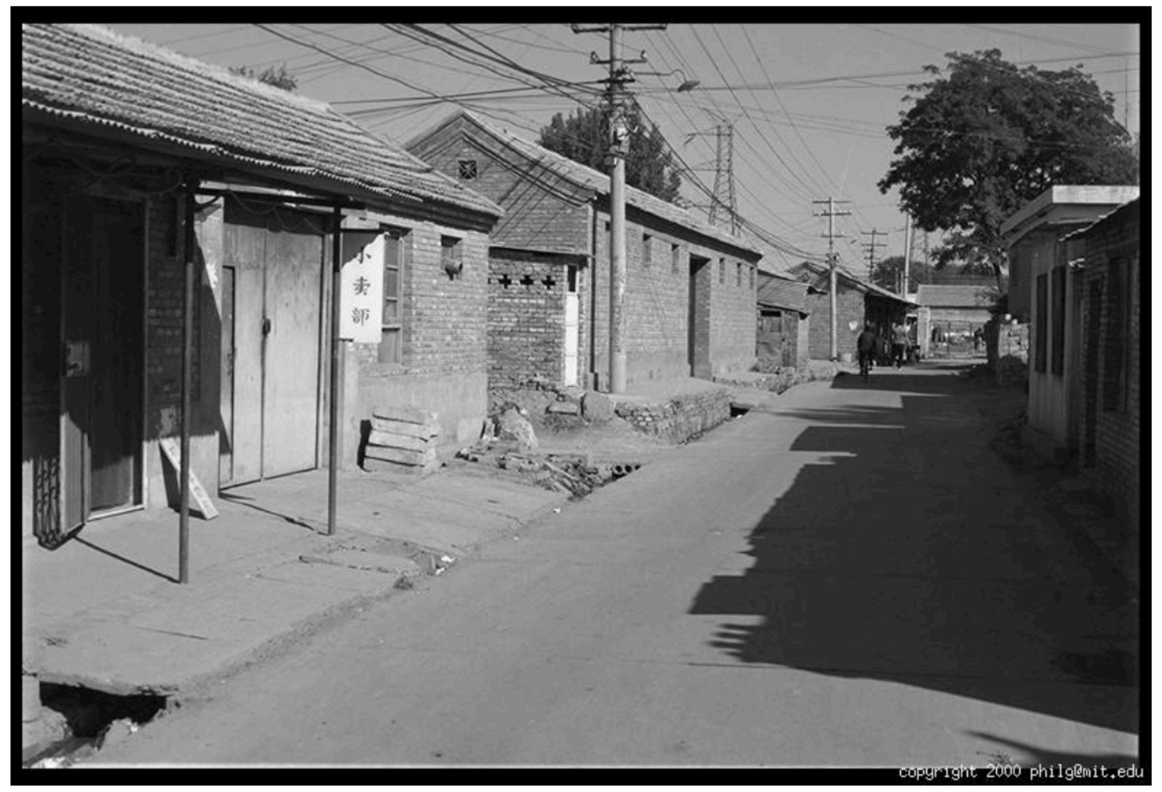
Fun Facts

- If you had to choose, would you rather go without:
  - intensity (‘value’), or
  - hue + saturation (‘chroma’)?

[James Hays]

Fun Facts

Intensity Only – Constant Color



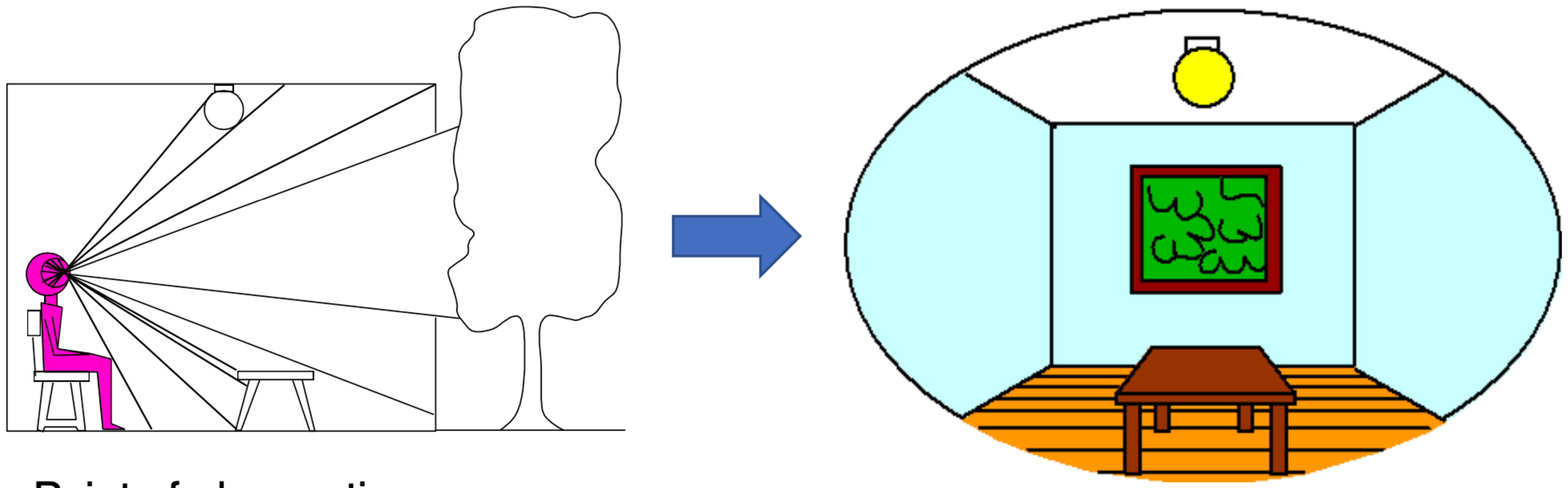
Color only – Constant Intensity

Plan for today

- Image representation
- Light
- Color
- Cameras (Orthographic, Perspective)

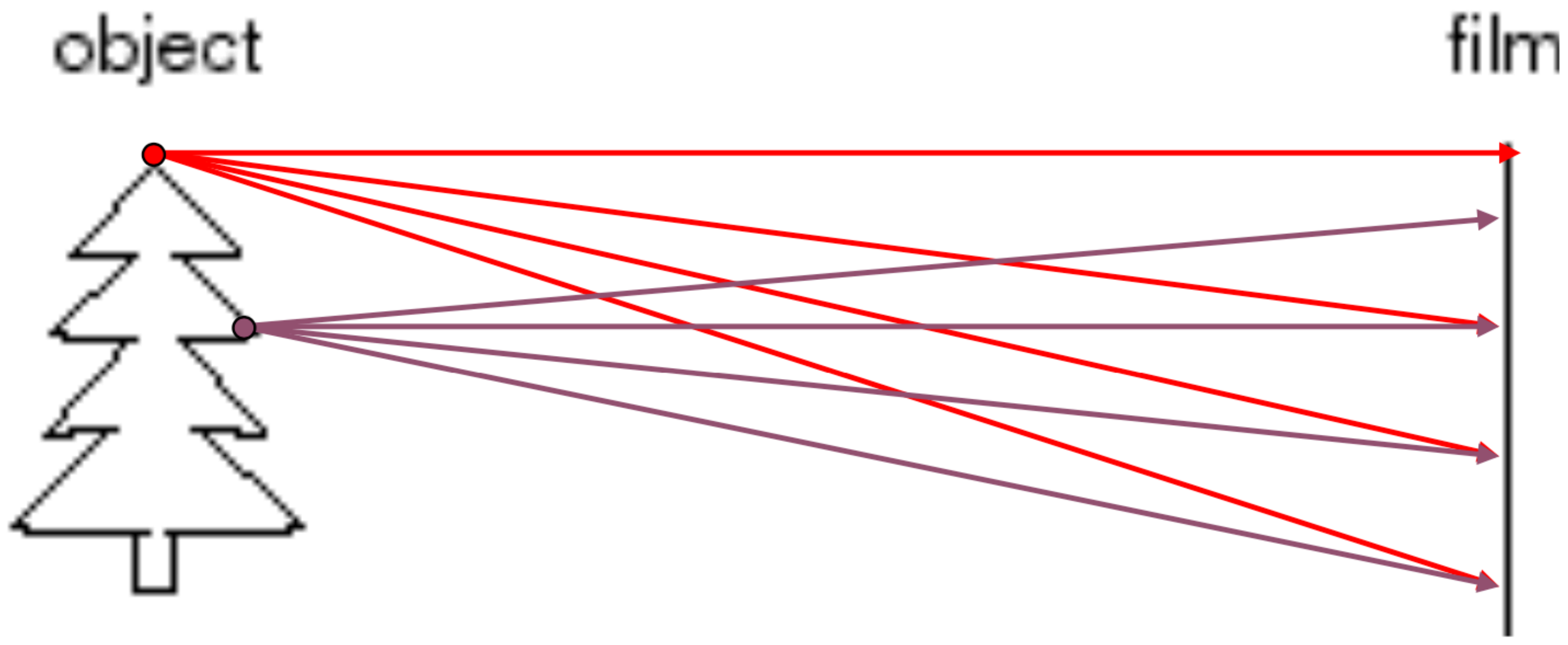
Dimensionality Reduction Machine (3D to 2D)

3D world *2D image*



Point of observation

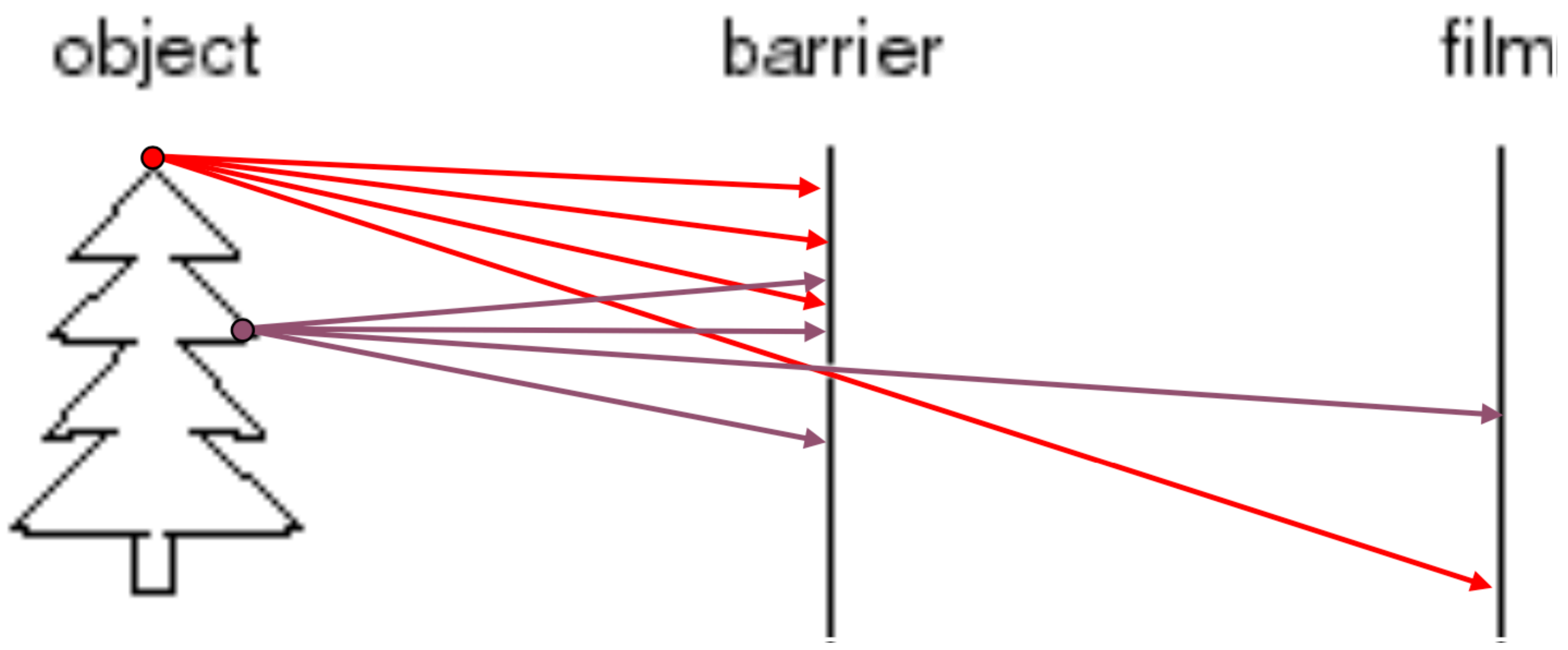
How to Capture the Visual World?



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

Slide by Steve Seitz

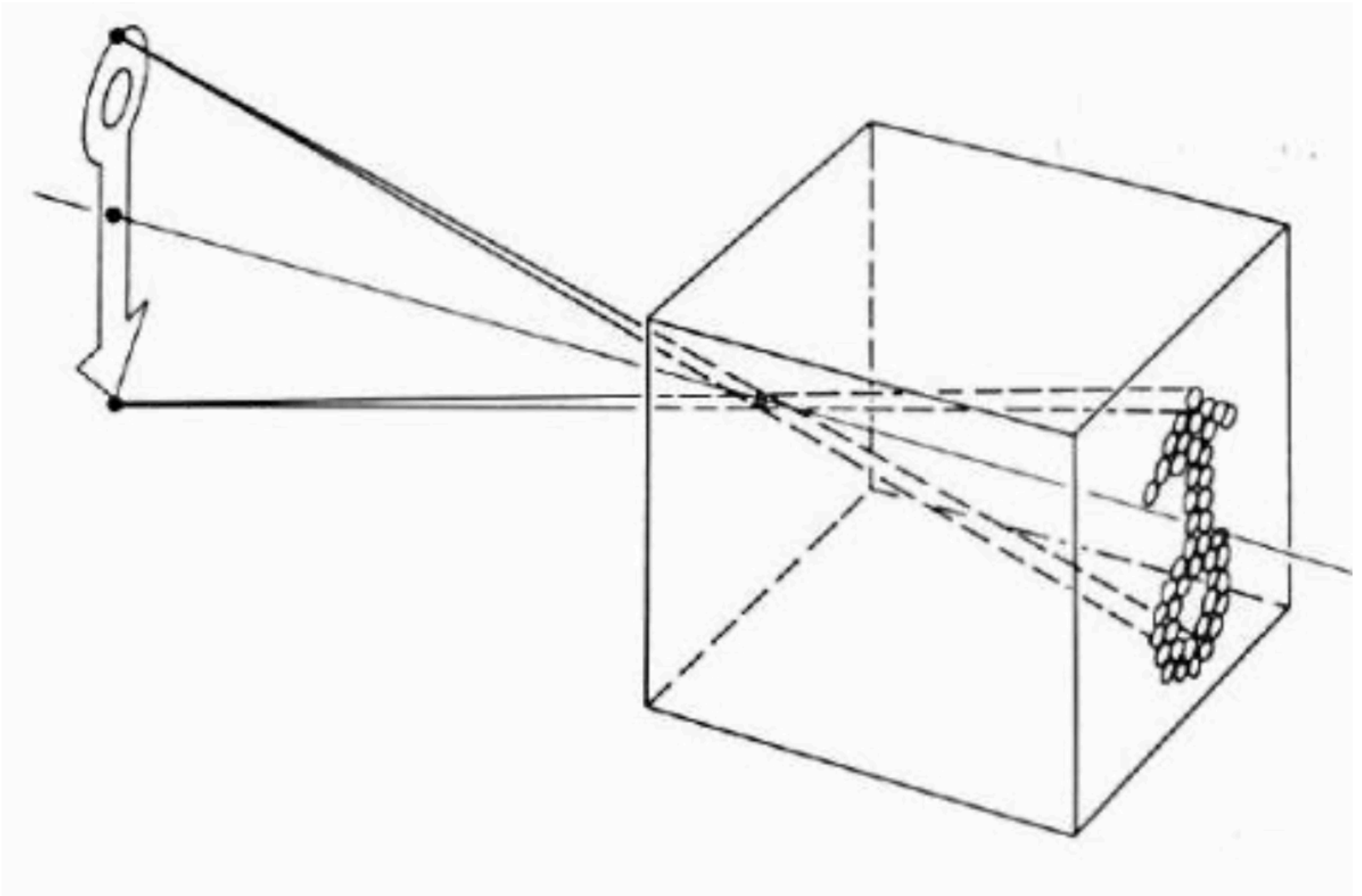
Pinhole Camera



- Idea 2: 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?

Slide by Steve Seitz

Pinhole Camera

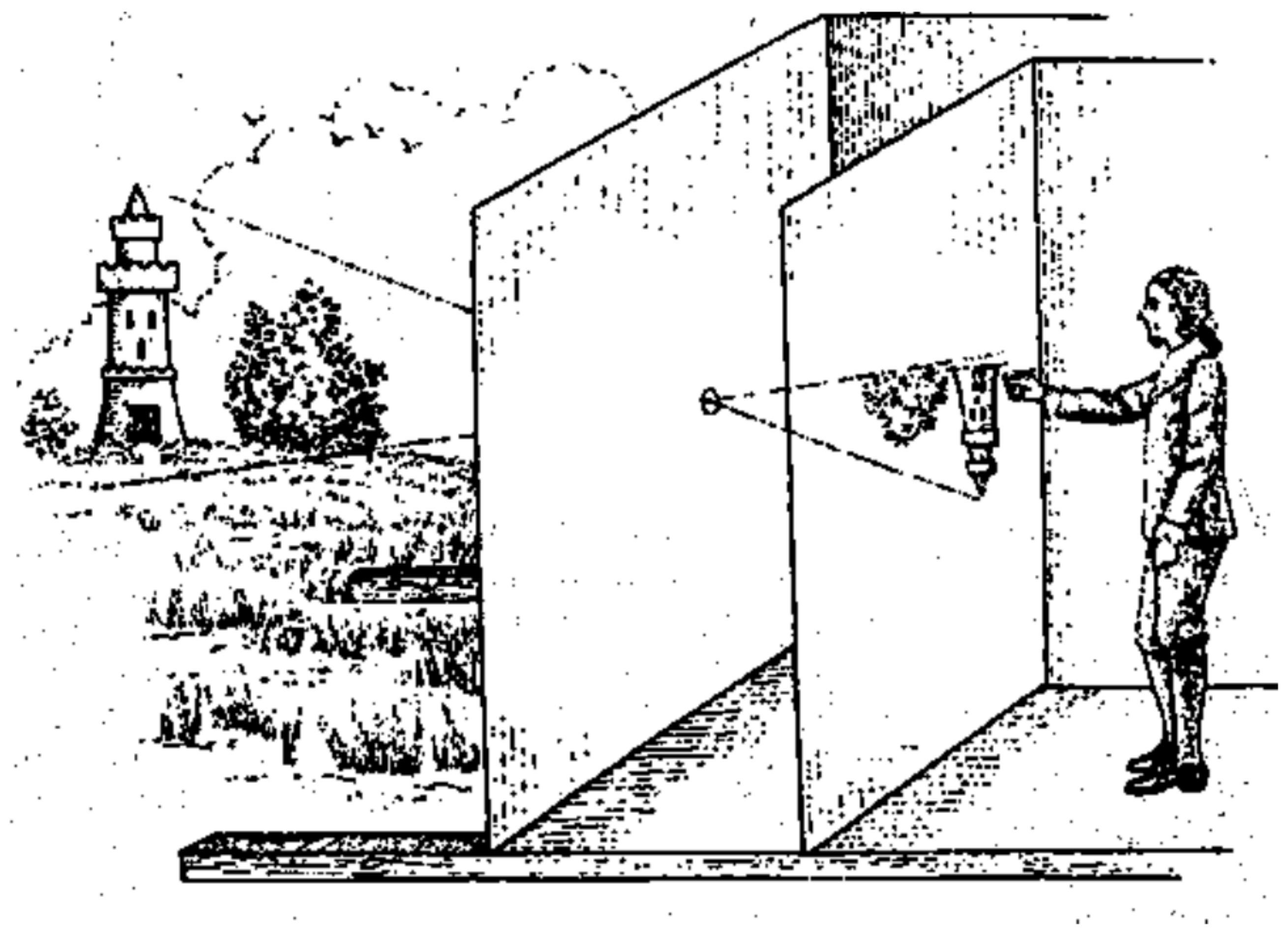


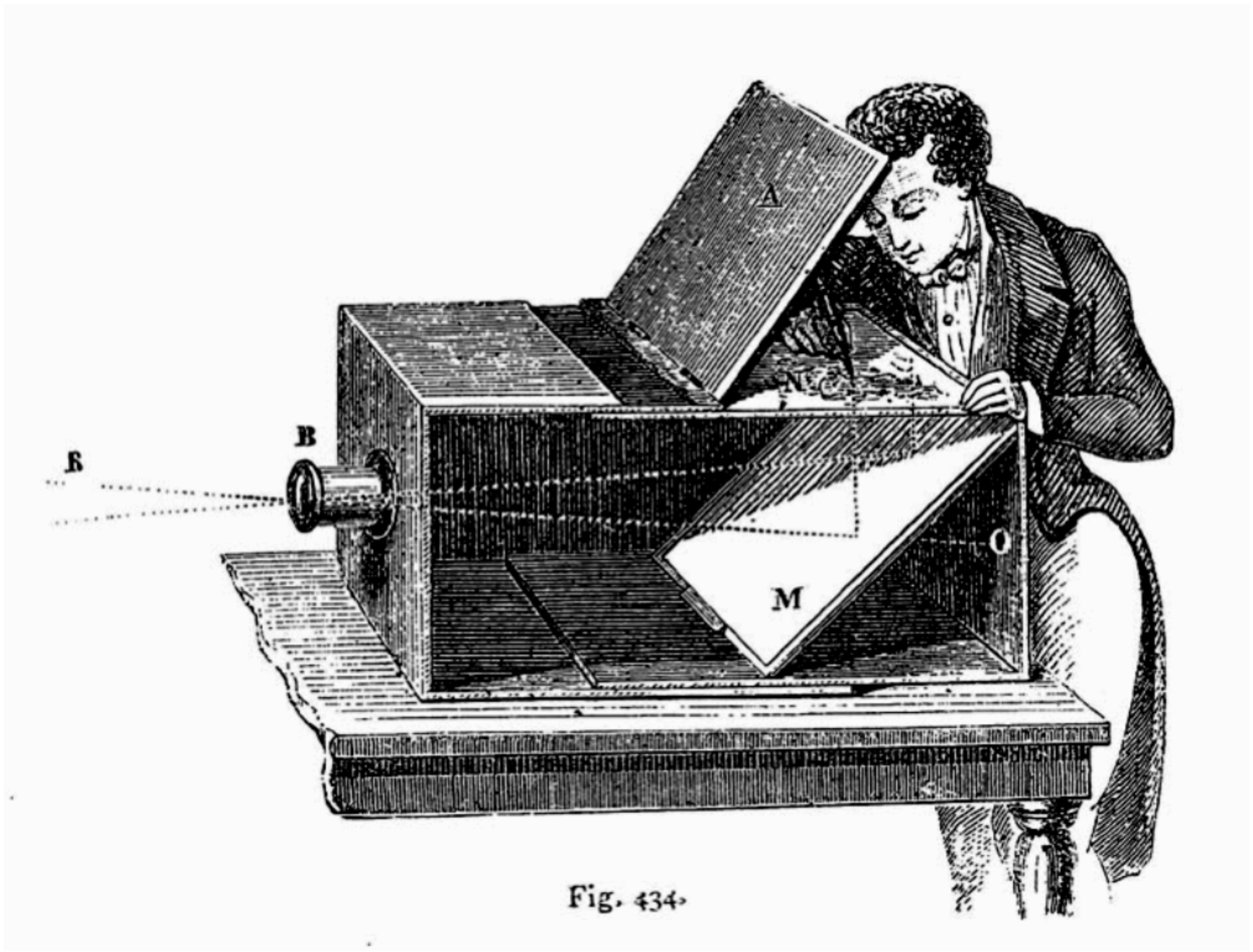
- Pinhole model:
  - Captures **pencil of rays** – all rays through a single point
  - The point is called **Center of Projection (COP)**
  - The image is formed on the **Image Plane**
  - **Effective focal length *f*** is distance from COP to Image Plane

Slide by Steve Seitz

First Cameras: Camera Obscura

- Camera Obscura known since classical period in China and Greece (e.g., Mo-Ti, China, 470BC to 390BC)
  - (Word ‘Camera’: Latin ‘room’. ‘Obscura’: ‘Dark’)
  - Exploiting the pinhole camera effect.



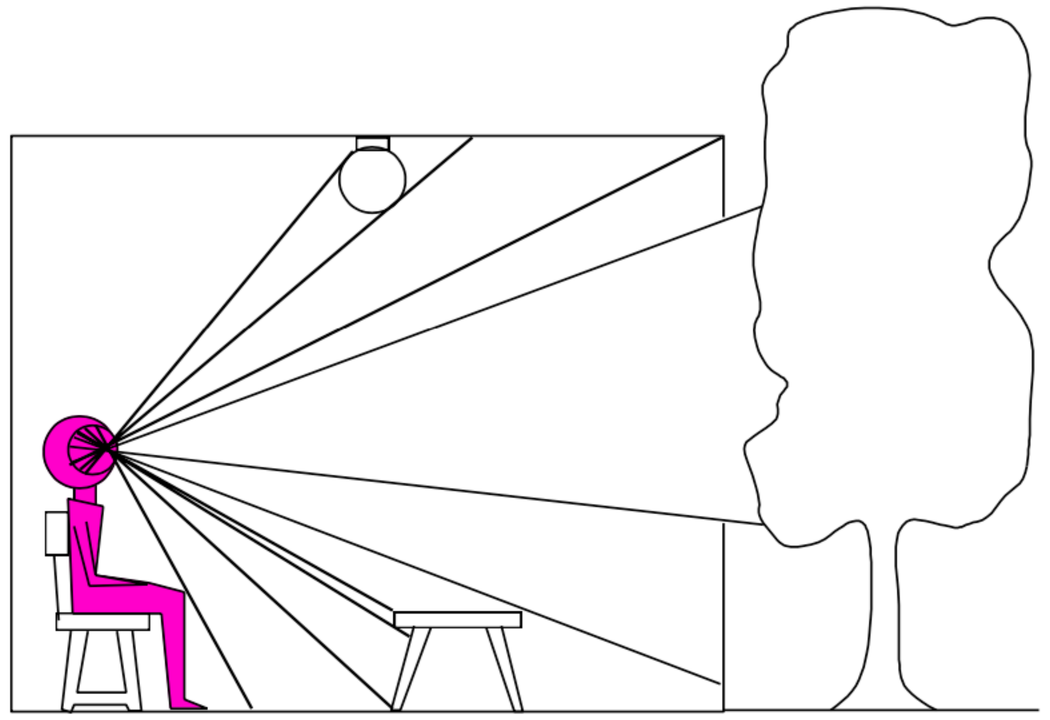


Lens Based Camera Obscura, 1568

Early pictures by tracing, later by film

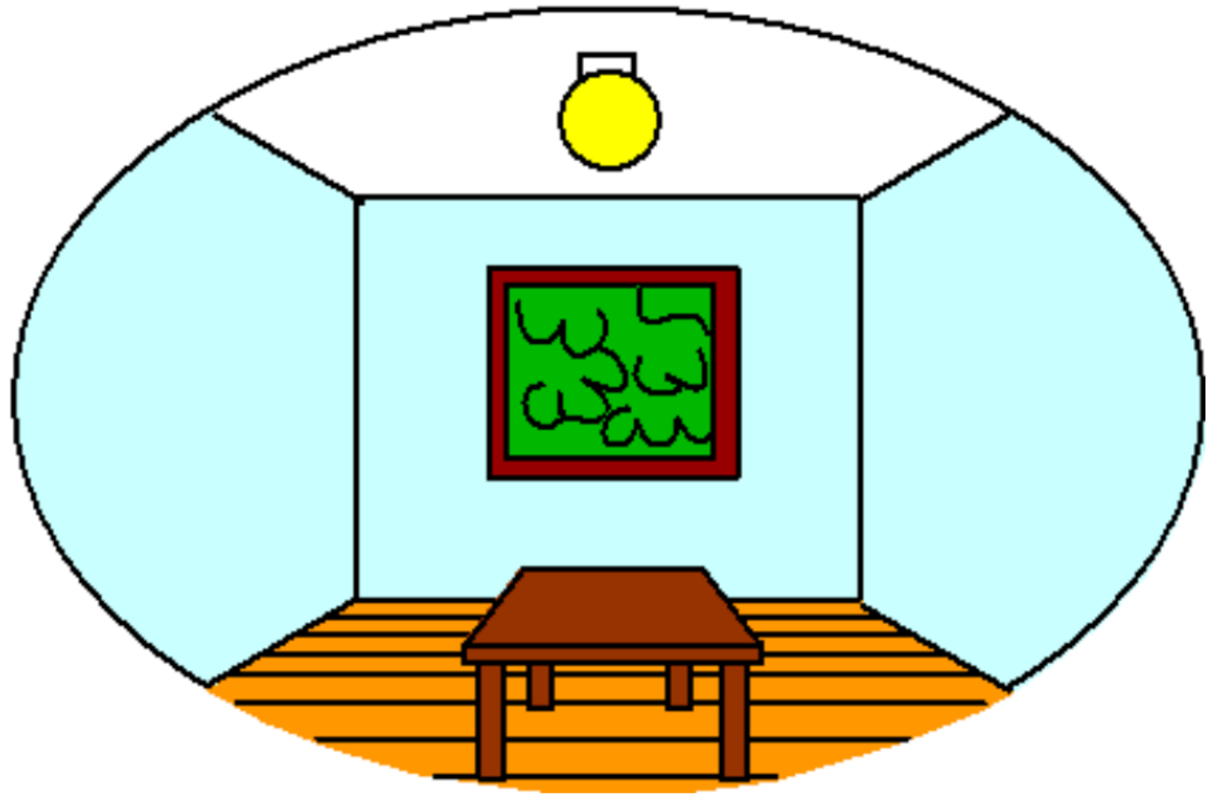
What is the Effect?

3D world



Point of observation

2D image

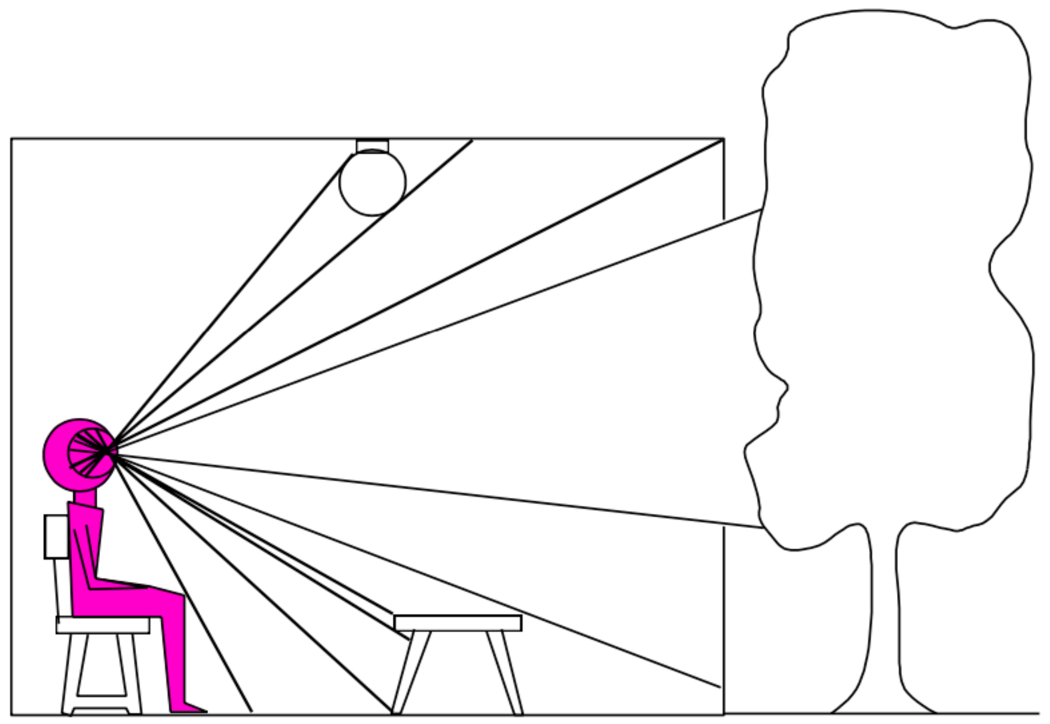


What is the effect of this dimensionality reduction?

- Some information is preserved, some is lost.

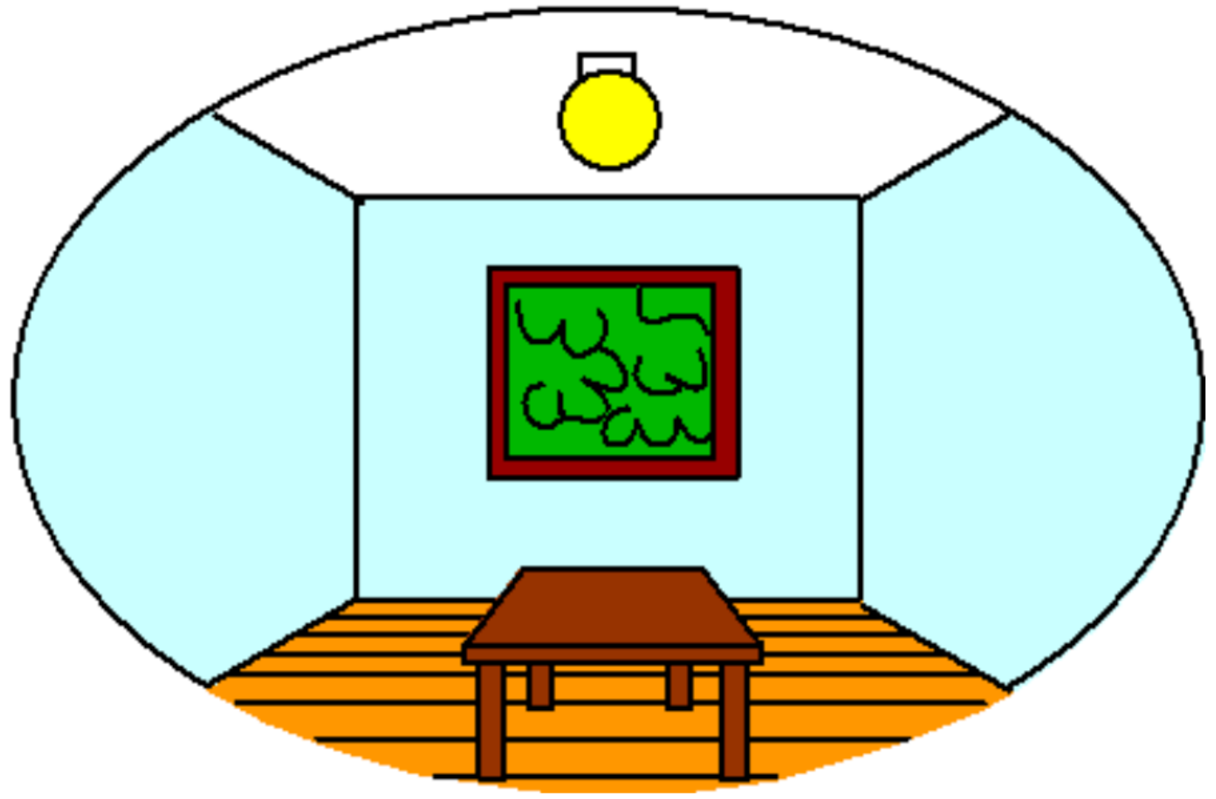
What is the Effect?

3D world



Point of observation

2D image



What have we lost?

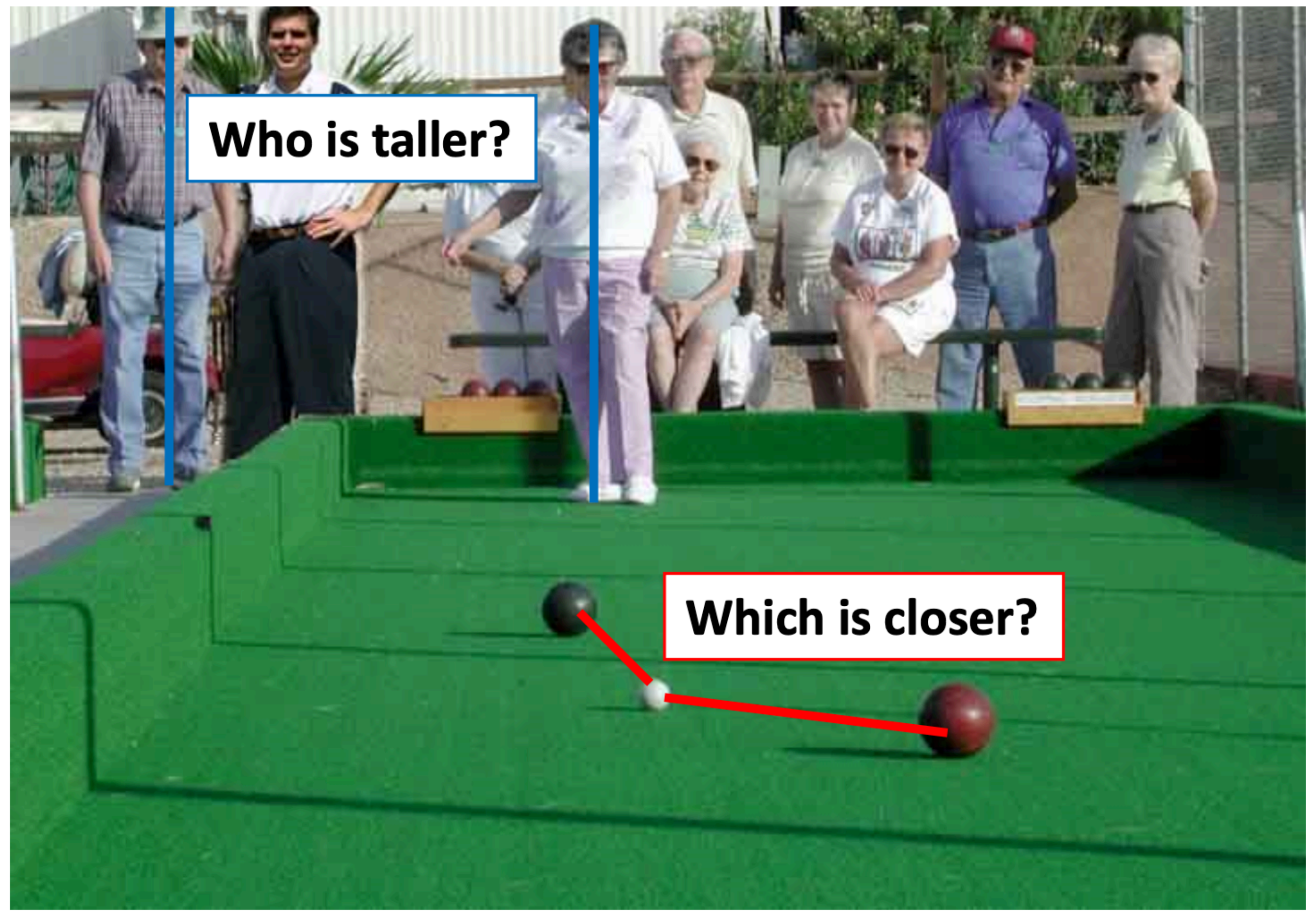
- Angles
- Distances (lengths)

What have we preserved?

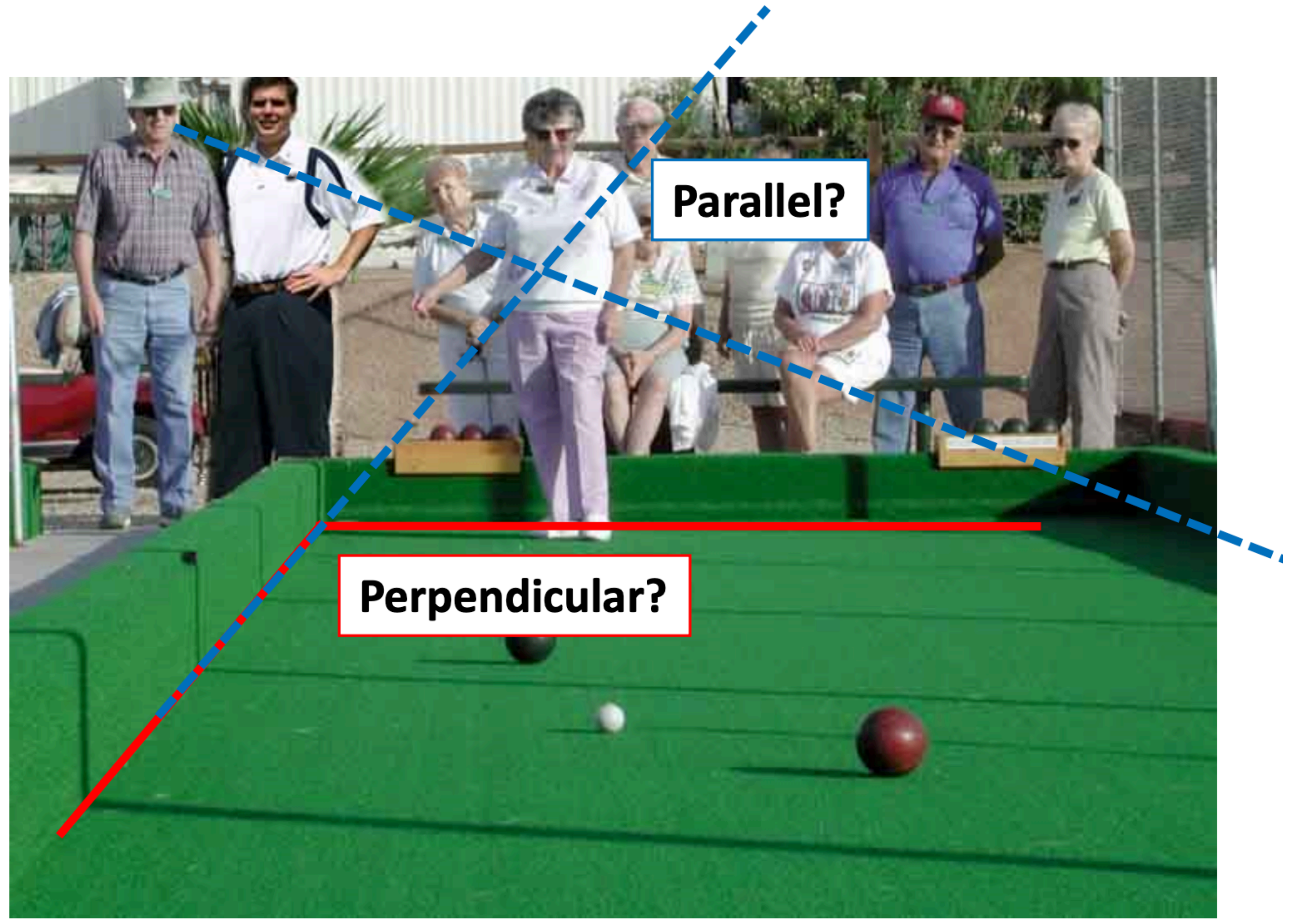
- Straight Lines
- Incidence

What is the Effect?

Length (and so area) is lost.

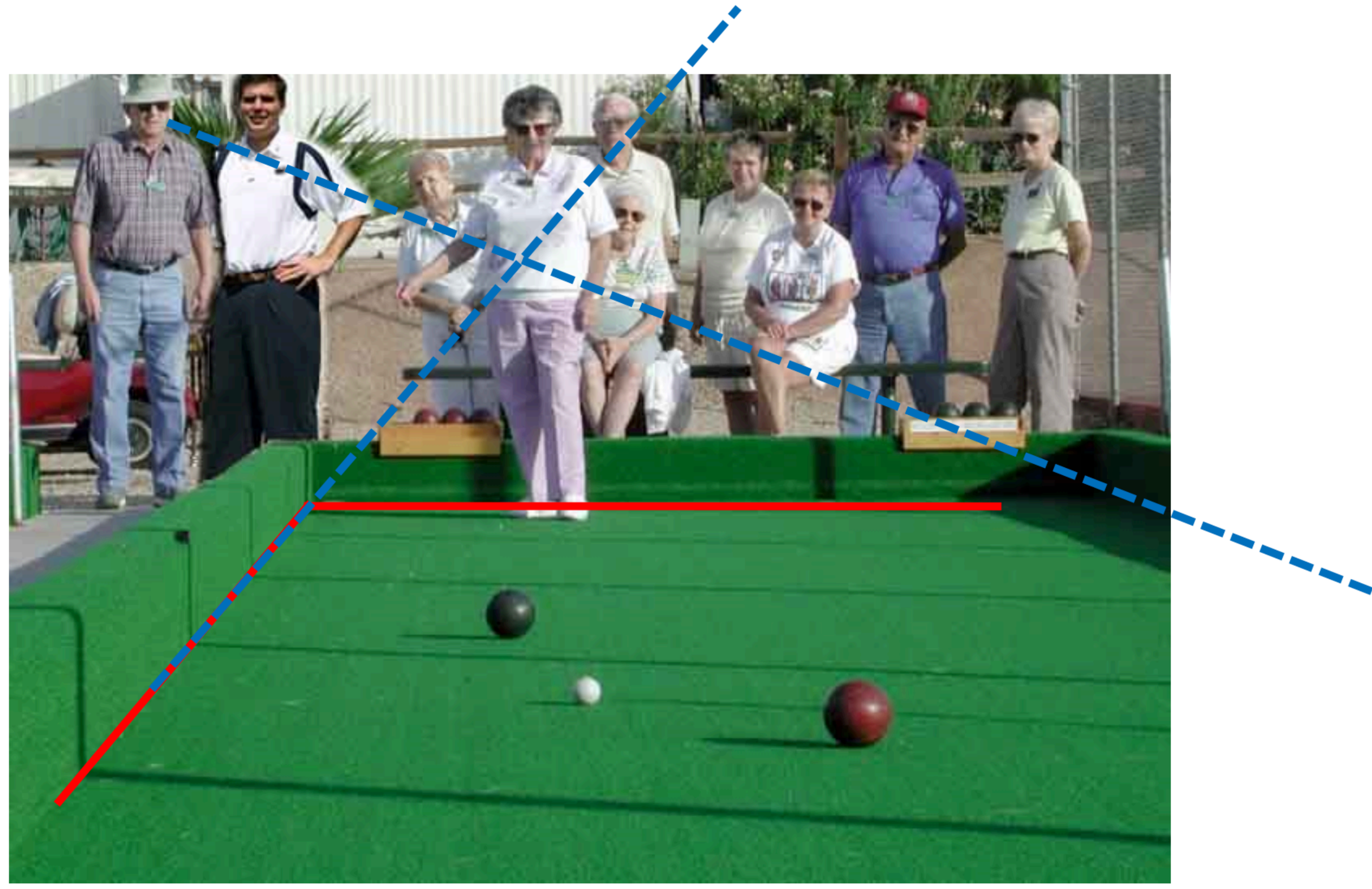


Angles are lost.



What is preserved?

- Straight lines are still straight.

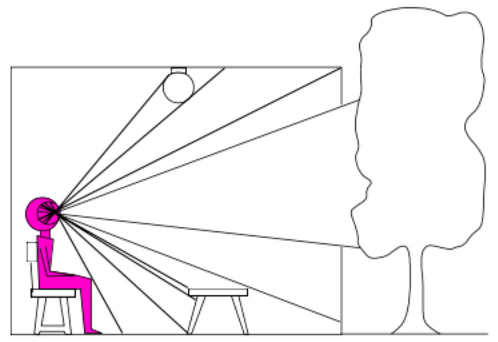


What is the Effect?

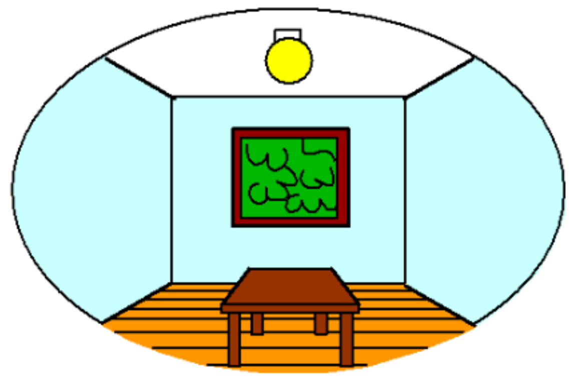
- Many-to-one:
  - All points along the same visual ray map to the same point in an image
  - An infinite number of 3D world scenes correspond to any single image.
    - We (our brains and computer vision algorithms) need to use prior knowledge about likely scenes to disambiguate.

Modeling Projection

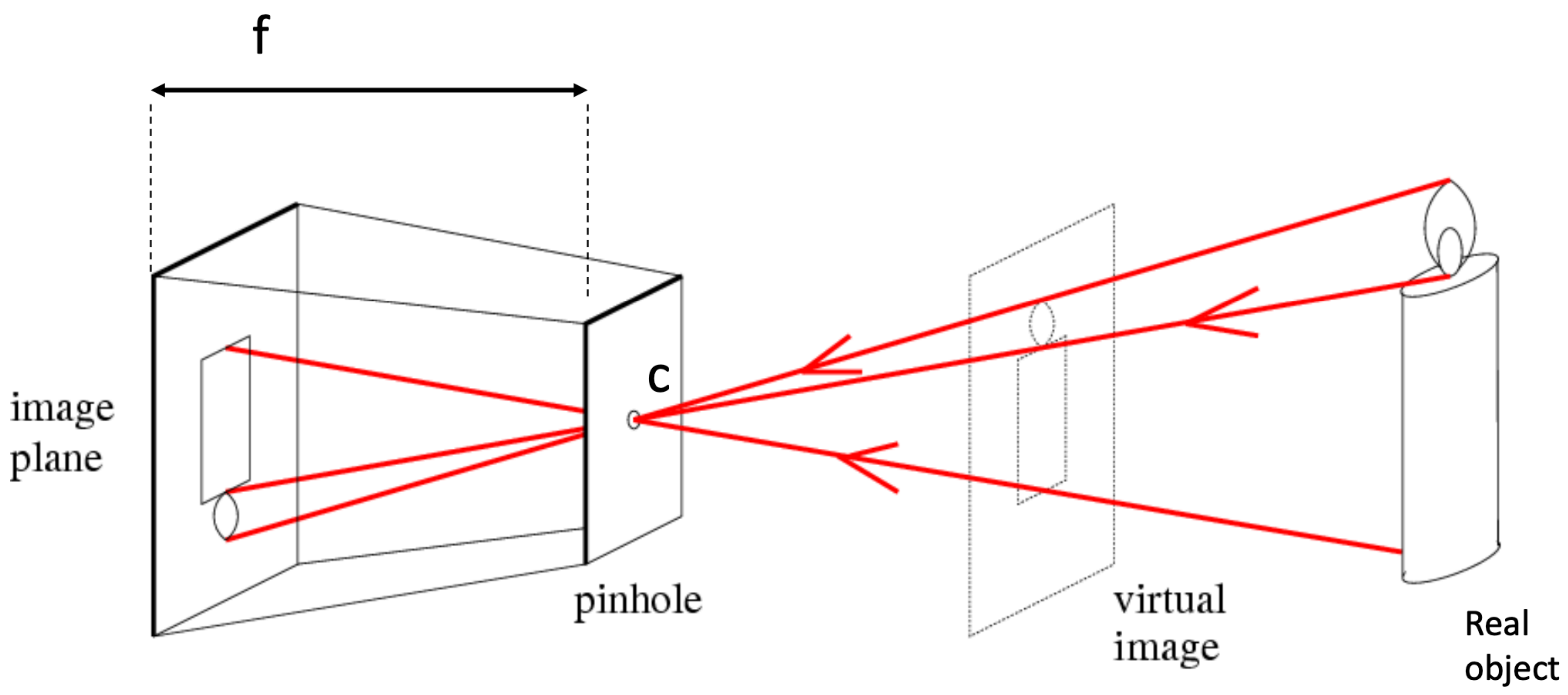
- Camera creates an image…
  - I(x,y) measures how much light is captured at pixel (x,y)
- We want to know:
  - Where does a point (x,y,z) in the world get measured in the camera image?
  - Two models:
    - Perspective projection
    - Orthographic projection



Point of observation



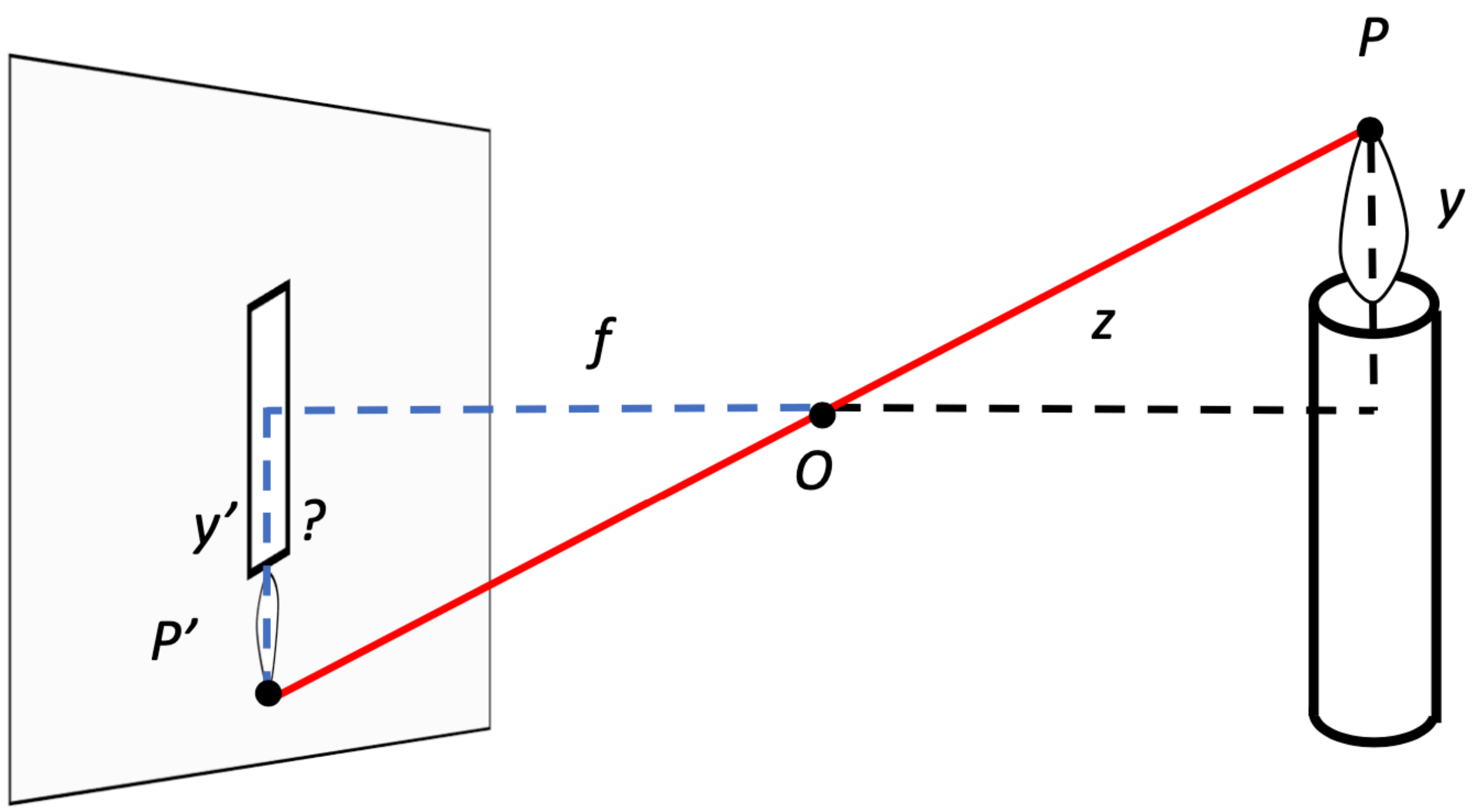
Perspective Projection



f = Focal length
c = Optical center of the camera

Figure from Forsyth

Perspective Projection



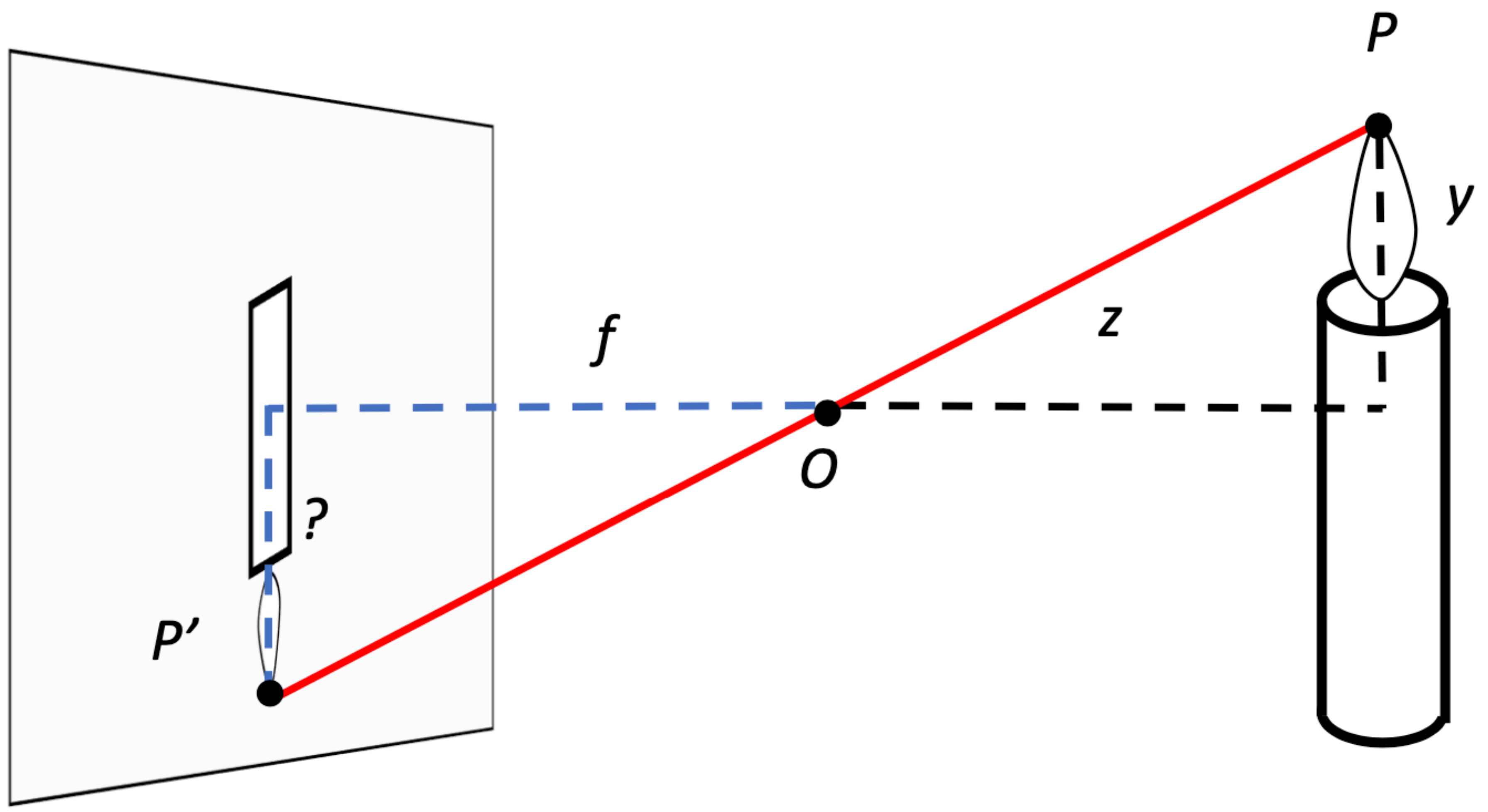
Given a point *P=(x,y,z)* in the world, where does it appear in the image plane?

- Derive the answer using similar triangles:

$$\frac{f}{-y\prime} = \frac{z}{y} \Rightarrow y' = \frac{-fy}{z}$$

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

Perspective Projection

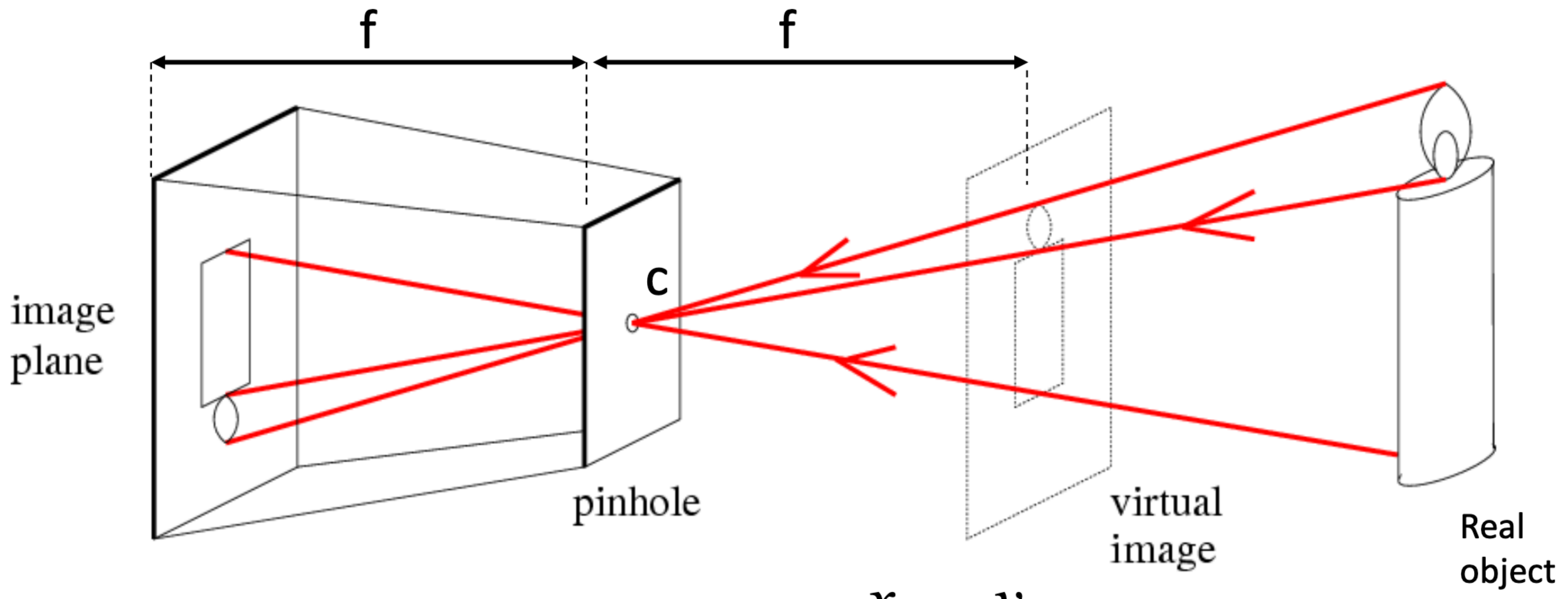


Projection from world to imaging plane:

- All points along one ray project to the same point on the plane
- The projection of the point at O is undefined.
- The image is inverted vertically and horizontally (Kepler 1604)

Perspective Projection

- Dealing with the inverted image:
  - Brain: No problem.
  - Computer: Flip the measured image horiz + vert.
  - Conceptually: We can imagine the imaging plane is in front of the optical center.

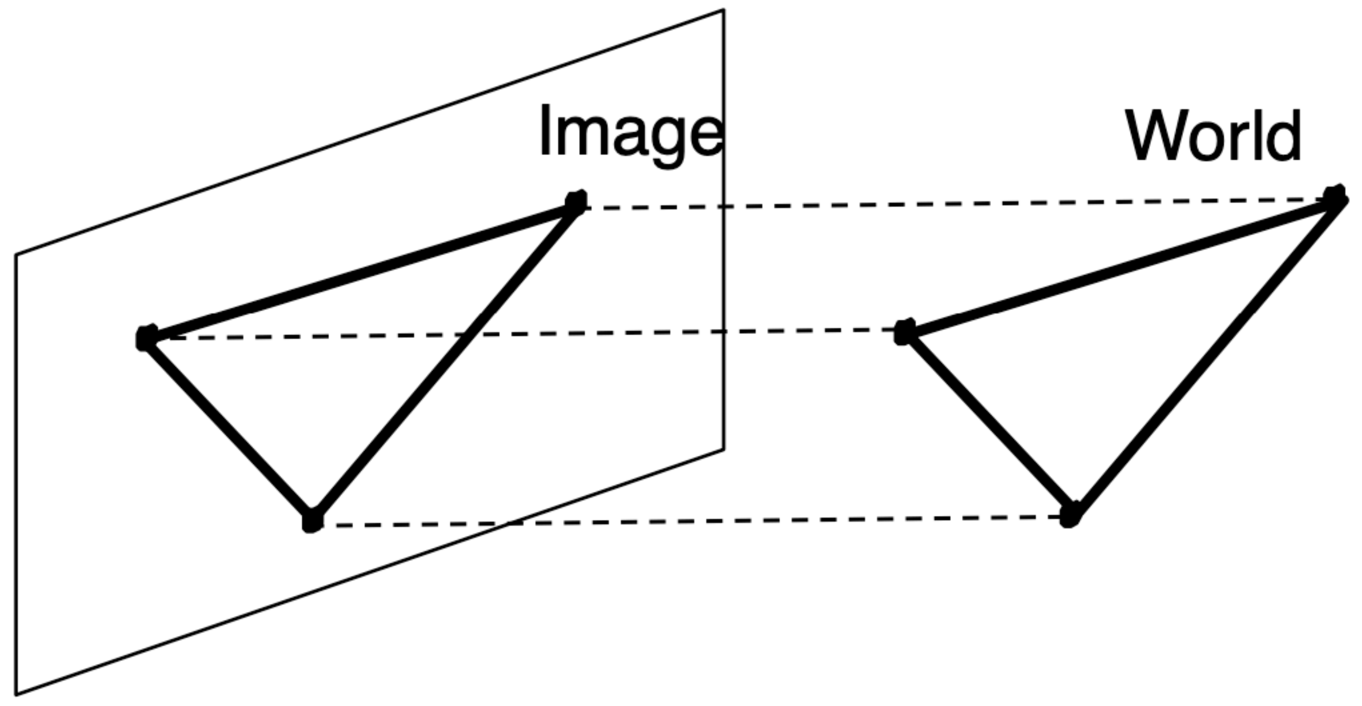


f = Focal length
c = Optical center of the camera

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

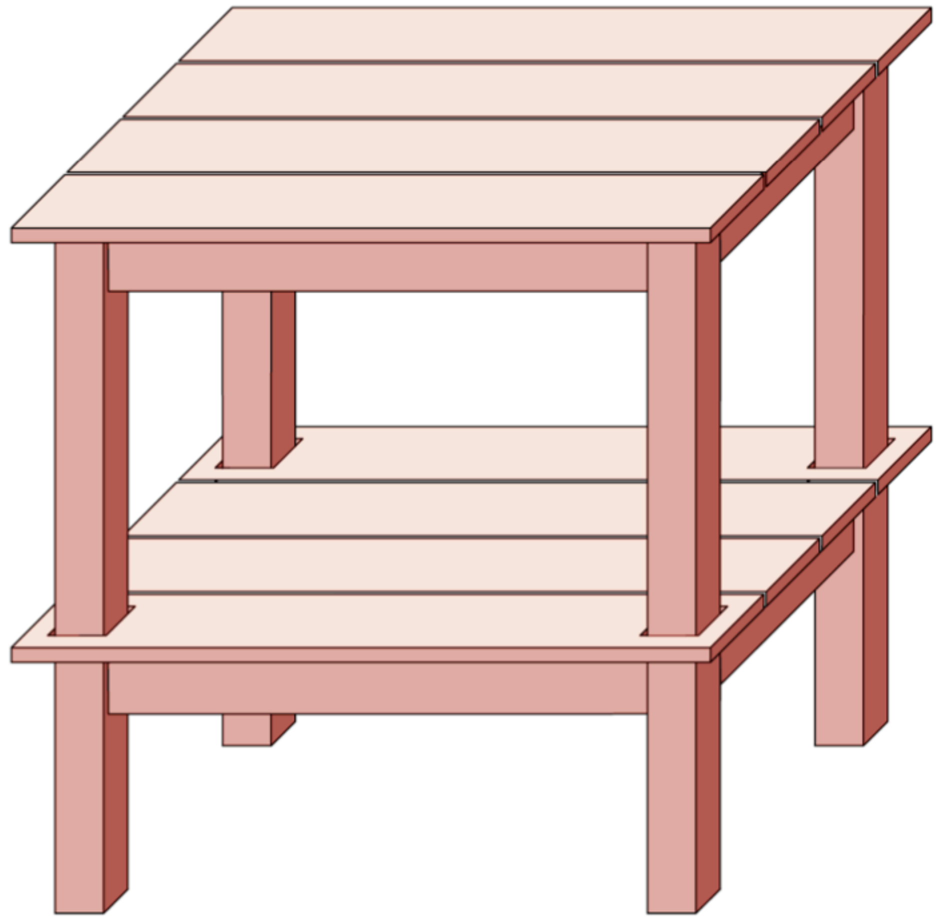
Orthographic Projection

- Special case of perspective projection
  - Distance from the COP to the PP is infinite

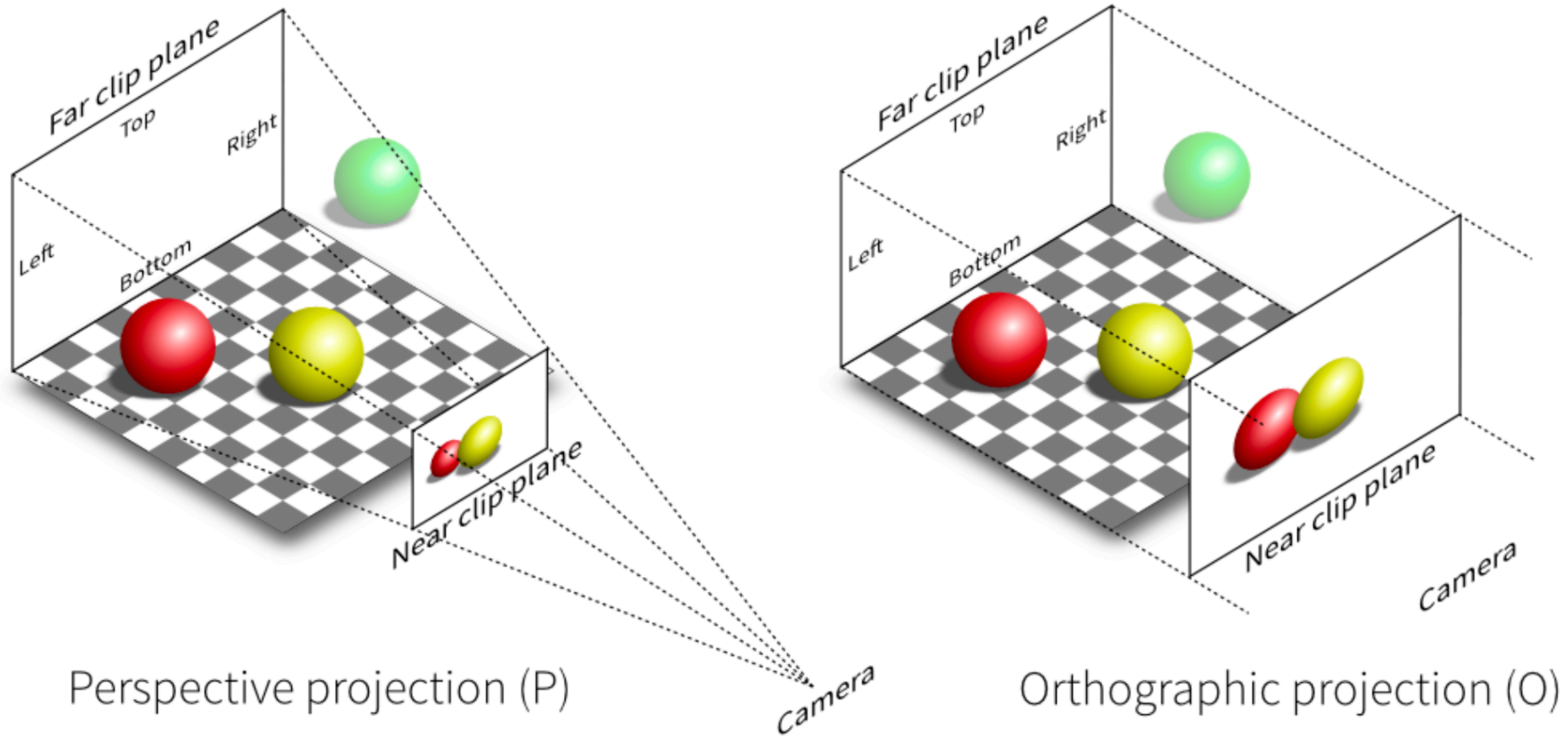


  - Good approximation for telephoto optics
  - Also called "parallel projection": (x, y, z) → (x, y)

Orthographic Projection

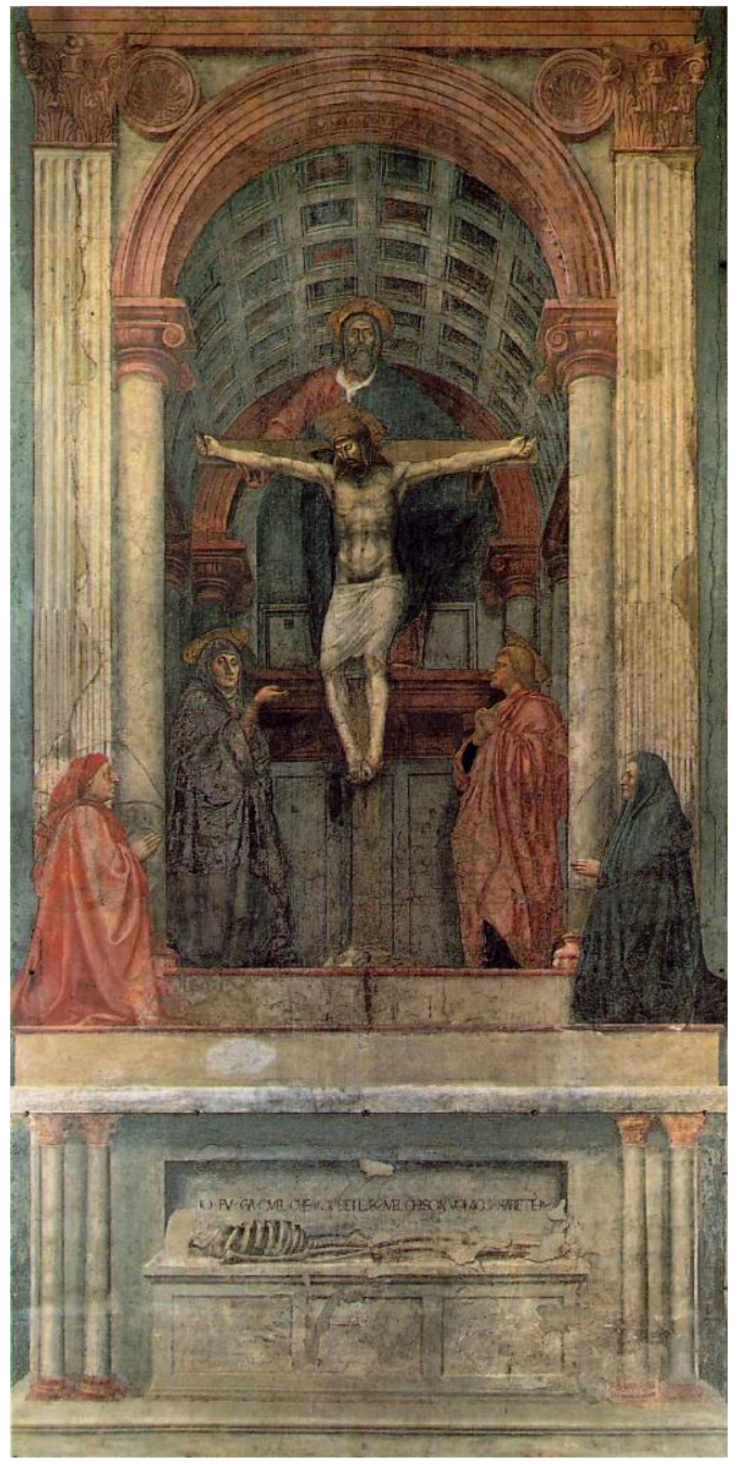


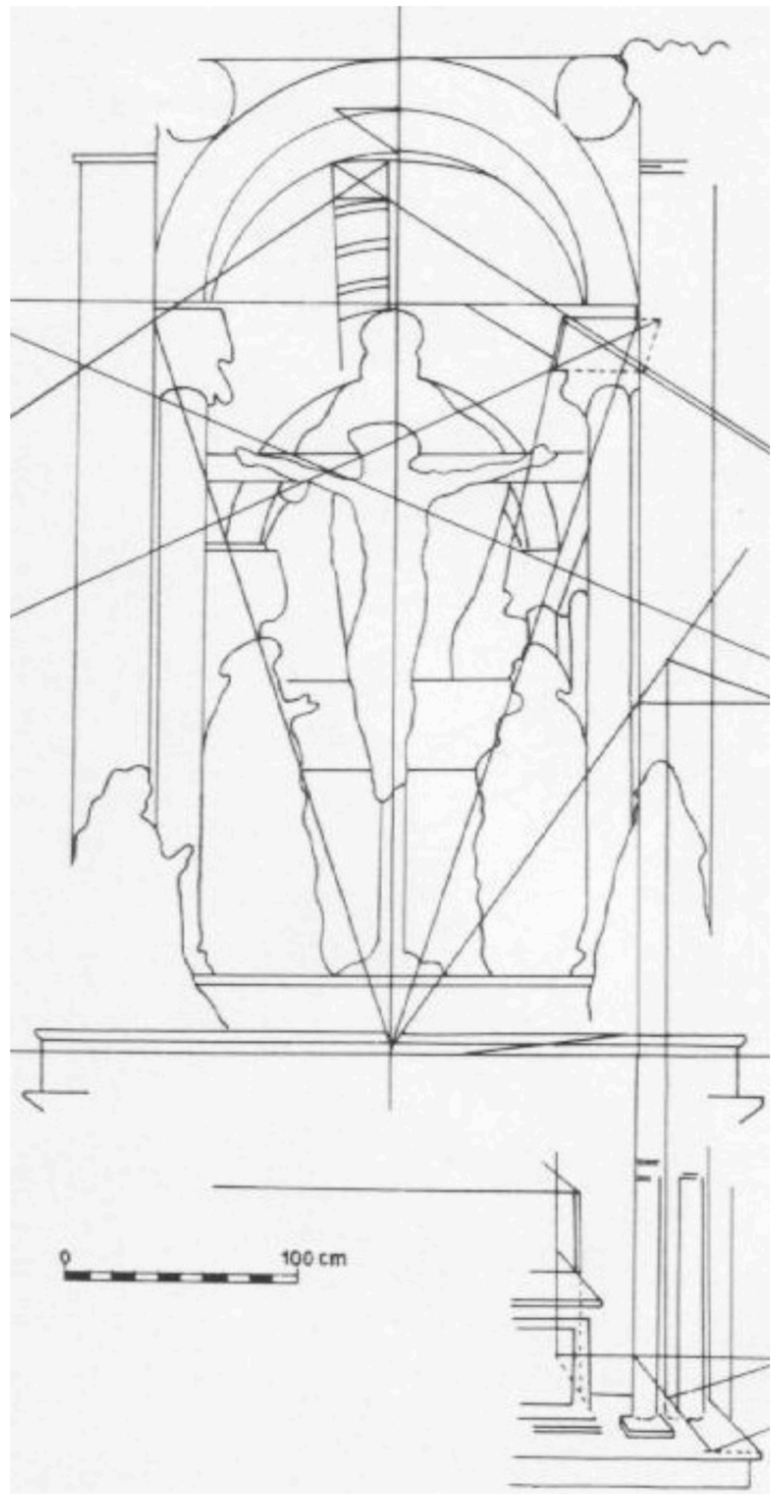
Perspective vs. Orthographic Projection



Perspective vs. Orthographic Projection

The Kangxi Emperor’s Southern Inspection Tour (1691-1698) by Wang Hui





- Masaccio, *Trinity*, Santa Maria Novella, Florence, 1425-28
- One of the first consistent uses of perspective in Western art

Cameras as Linear Systems

- We would like to model the image formation projection as a linear transformation (matrix multiply).
  - Transformation is non-linear due to division by *z*.

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

Cameras as Linear Systems

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

- Is this a linear transformation?
  - no—division by z is nonlinear

Trick: add one more coordinate:

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

homogeneous image coordinates

homogeneous scene coordinates

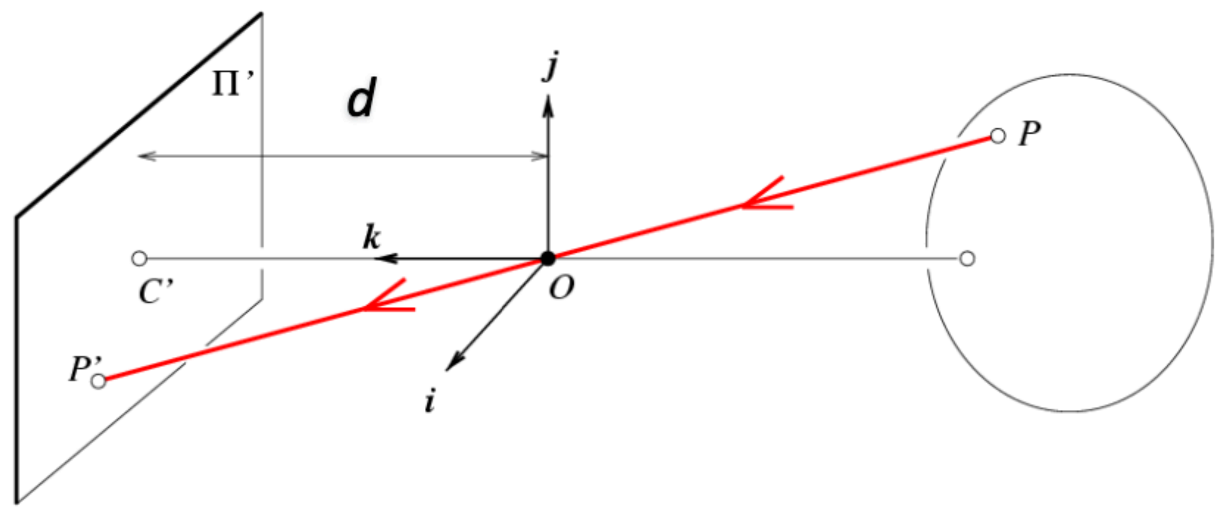
Converting *from* homogeneous coordinates

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

Slide by Steve Seitz

Cameras as Linear Systems

- Relating a real-world point to a point on the camera



- Cartesian Coordinates: $P = (x, y, z) \rightarrow P' = (-d\frac{x}{z}, -d\frac{y}{z})$
- Homogeneous Coordinates:
  - Perspective projection is a matrix multiply $P' = M\ P$

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

"Projection Matrix" *M*

divide by third coordinate

Example

- 1. Object point at (10, 6, 4), d=2

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 6 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 & 6 & -2 \end{bmatrix}$$

$$\boxed{\Rightarrow x' = -5,\ y' = -3}$$

2. Object point at (25, 15, 10)

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} 25 \\ 15 \\ 10 \\ 1 \end{bmatrix} = \begin{bmatrix} 25 & 15 & -5 \end{bmatrix}$$

$$\boxed{\Rightarrow x' = -5,\ y' = -3}$$

Perspective projection is not 1-to-1!

[L Shapiro]

Application: Recovering 3D

J. Vermeer, *Music Lesson*, 1662

A. Criminisi, M. Kemp, and A. Zisserman, [Bringing Pictorial Space to Life: computer techniques for the analysis of paintings](), *Proc. Computers and the History of Art*, 2002

http://research.microsoft.com/en-us/um/people/antcrim/ACriminisi_3D_Museum.wmv

[S. Lazebnik]

Cameras as Linear Systems

- So far we looked at an ideal, simplified camera model
  - Intrinsic: Unit aspect ratio, no skew, optical centre at (0,0),
  - Extrinsic: No rotation, Camera at (0,0,0).

$$P' = M\ P \qquad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ -z/d \end{bmatrix}$$

- More generally, lots of transformations to model a real camera….

$$\begin{pmatrix} \text{2D} \\ \text{point} \\ \text{(3x1)} \end{pmatrix} = \begin{pmatrix} \text{Camera to} \\ \text{pixel coord.} \\ \text{trans. matrix} \\ \text{(3x3)} \end{pmatrix} \begin{pmatrix} \text{Perspective} \\ \text{projection matrix} \\ \text{(3x4)} \end{pmatrix} \begin{pmatrix} \text{World to} \\ \text{camera coord.} \\ \text{trans. matrix} \\ \text{(4x4)} \end{pmatrix} \begin{pmatrix} \text{3D} \\ \text{point} \\ \text{(4x1)} \end{pmatrix}$$

Summary

- **Image representation**
  - Sampling
  - Quantization
- **Light**
  - BRDF
  - Lambertian, Specular, Phong
- **Color**
  - In humans
  - In cameras
- **Cameras**
  - Pinhole Cameras
  - Orthographic and Projective Geometry
  - Cameras as linear systems