Semester One 2021 ENGN 6528 Computer Vision

Week01-Lecture 02

Miaomiao Liu

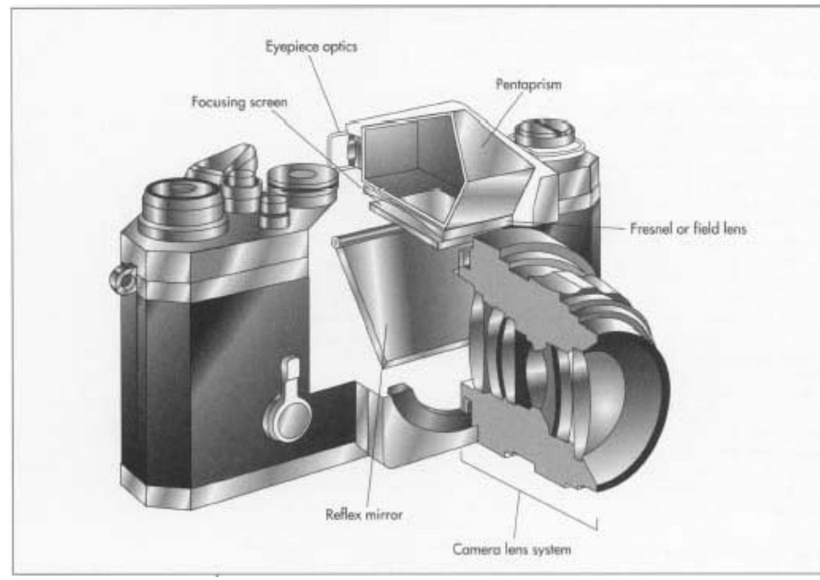
Week01-Lecture 02

Image Formation, Representation

Outline

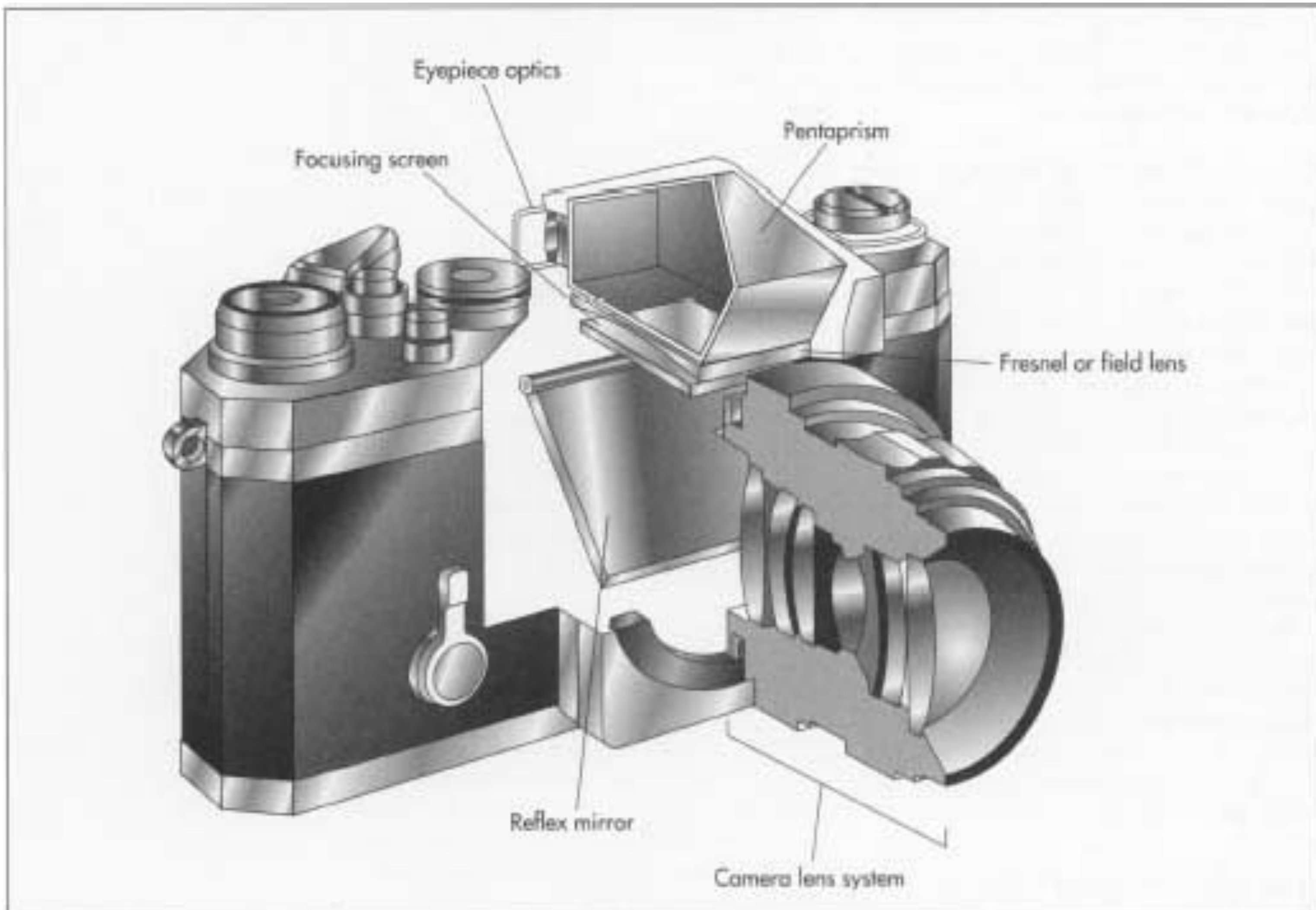
- What is a camera (sensor) ?
- Geometric image formation (details in 3D Vision, week 7 Lectures)
- Photometric image formation
- Image representation
  - Colour space

What is a camera ?



Slides from: F. Durand, S. Seitz, S. Lazebnik, S. Palmer

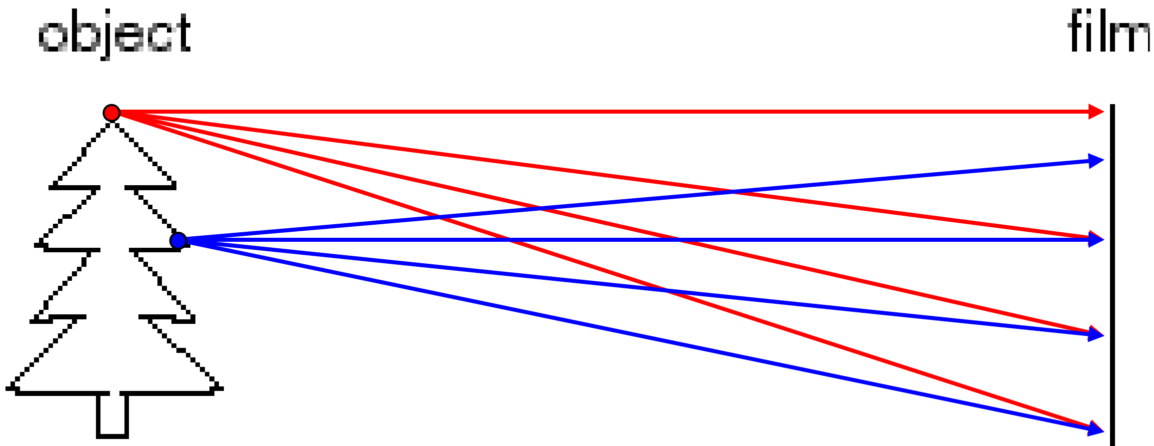
What is a camera ?



Slides from: F. Durand, S. Seitz, S. Lazebnik, S. Palmer

Let us design a camera

Let's design a camera

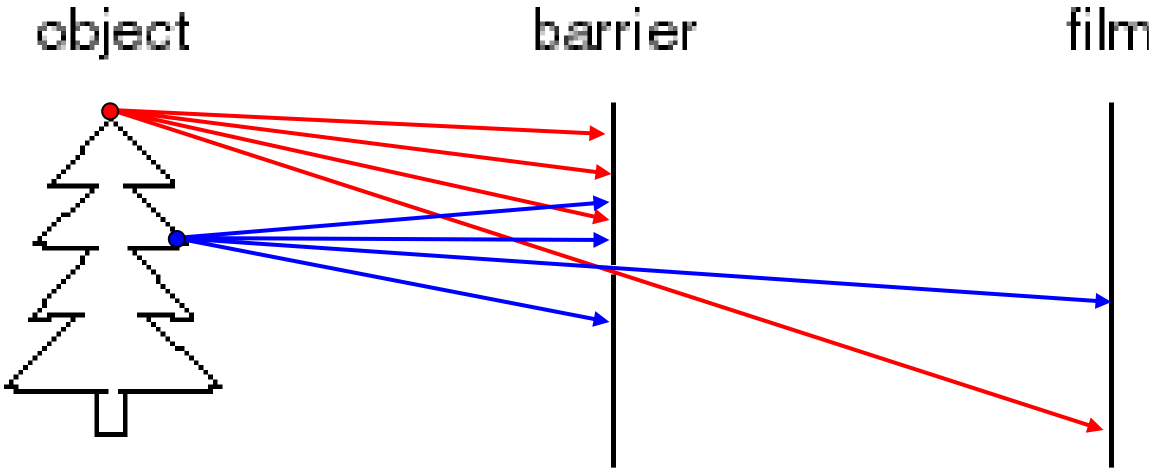


Idea 1: put a piece of film in front of an object

Will we get an image?

Slide by Steve Seitz

Pinhole camera

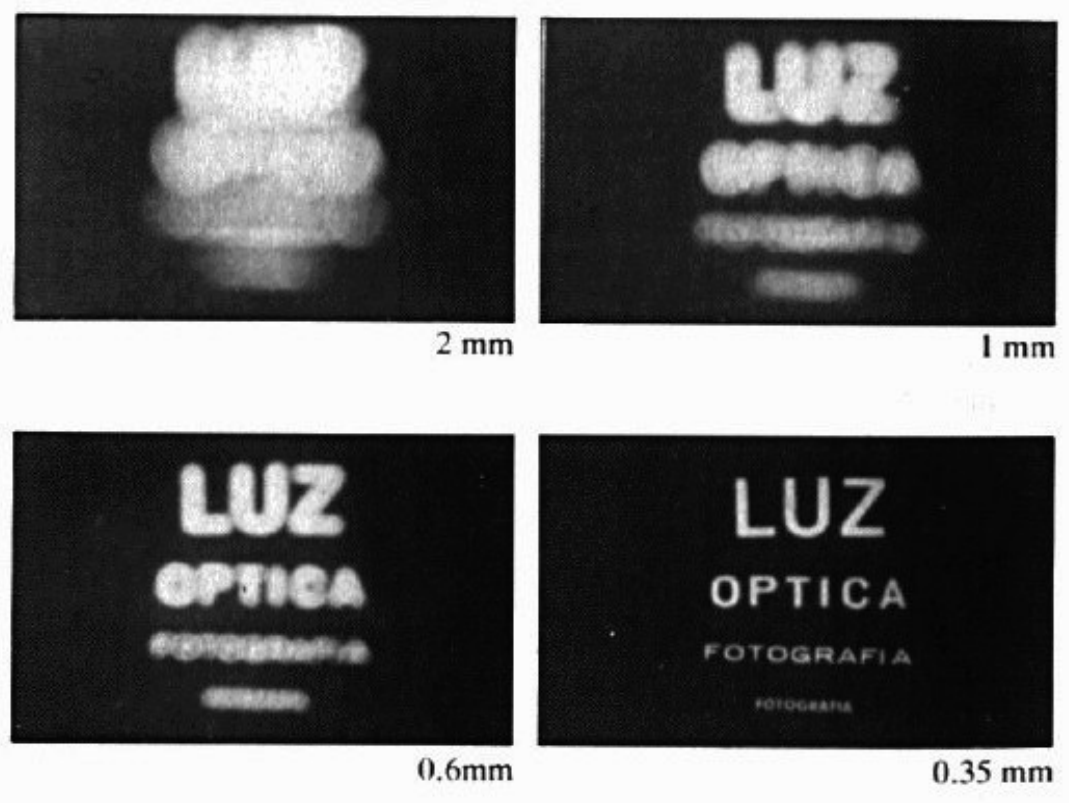


Add a barrier to block off most of the rays

- This reduces blurry effect
- The opening is known as the ***aperture***

Slide by Steve Seitz

Shrinking the aperture

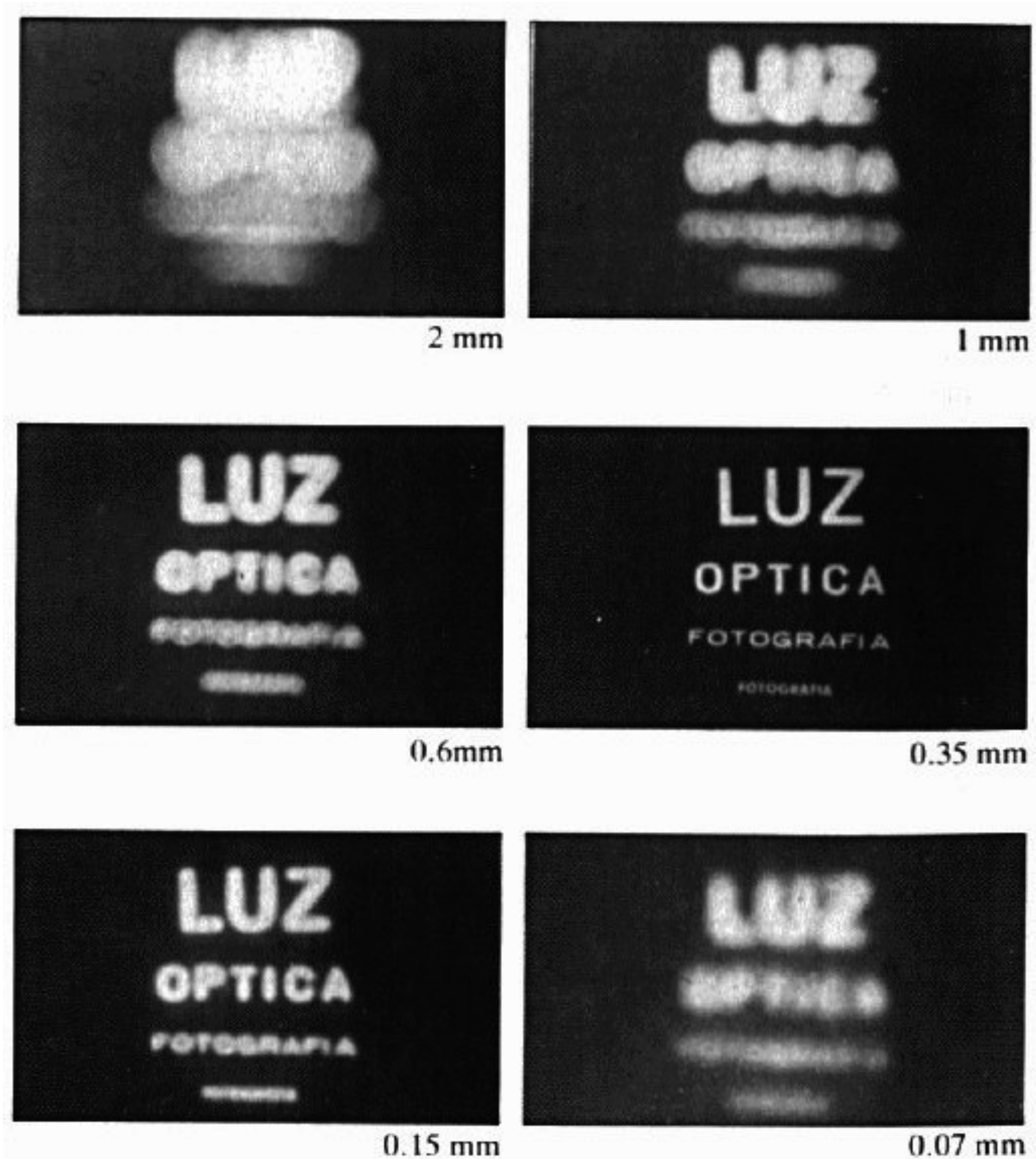


Why not make the aperture as small as possible?

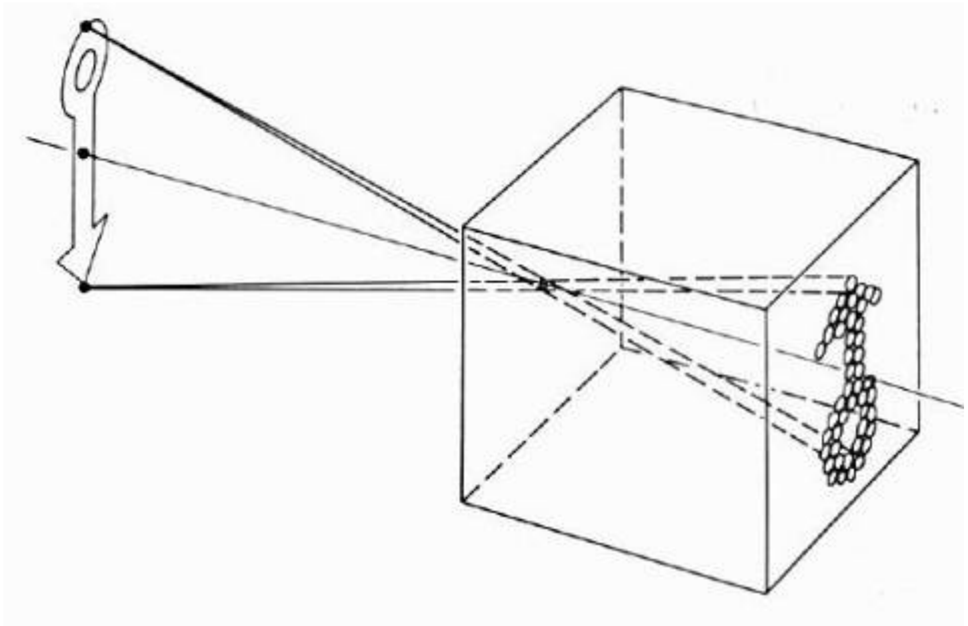
- Less light gets through
- Diffraction effects…

Slide by Steve Seitz

Shrinking the aperture



Pinhole camera model

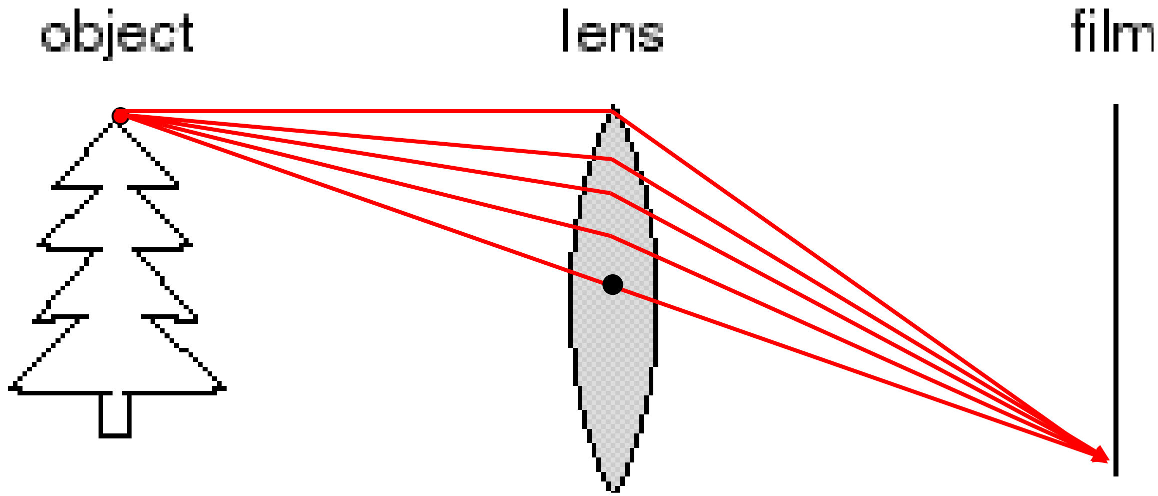


Pinhole model:

- Captures ***pencil of rays*** – all rays through a single point
- This point is called ***Center of Projection (focal point)***
- The image is formed on the **Image Plane**

Slide by Steve Seitz

Adding a lens… to capture more light

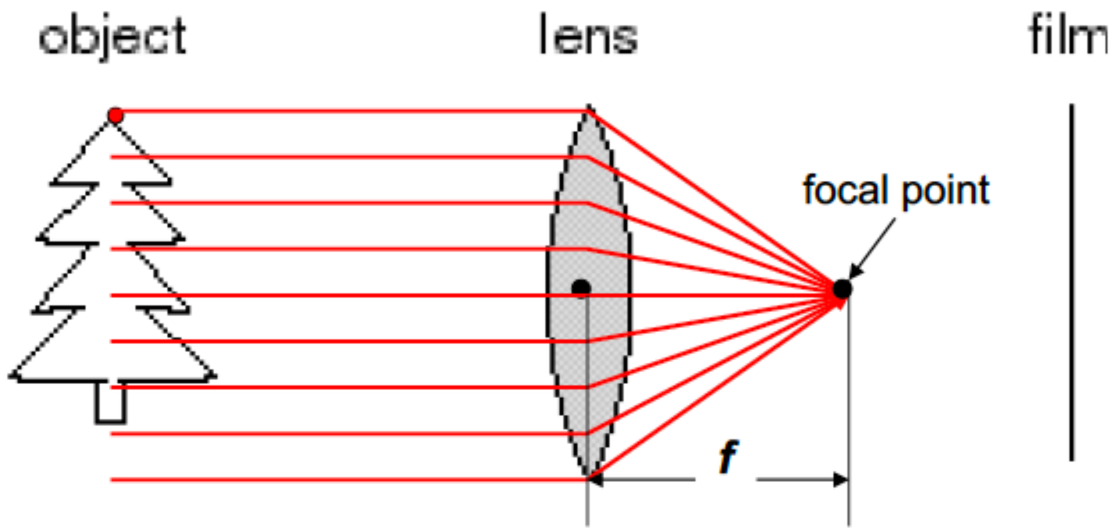


A lens focuses light onto the film

- Rays passing through the center are not deviated

Slide by Steve Seitz

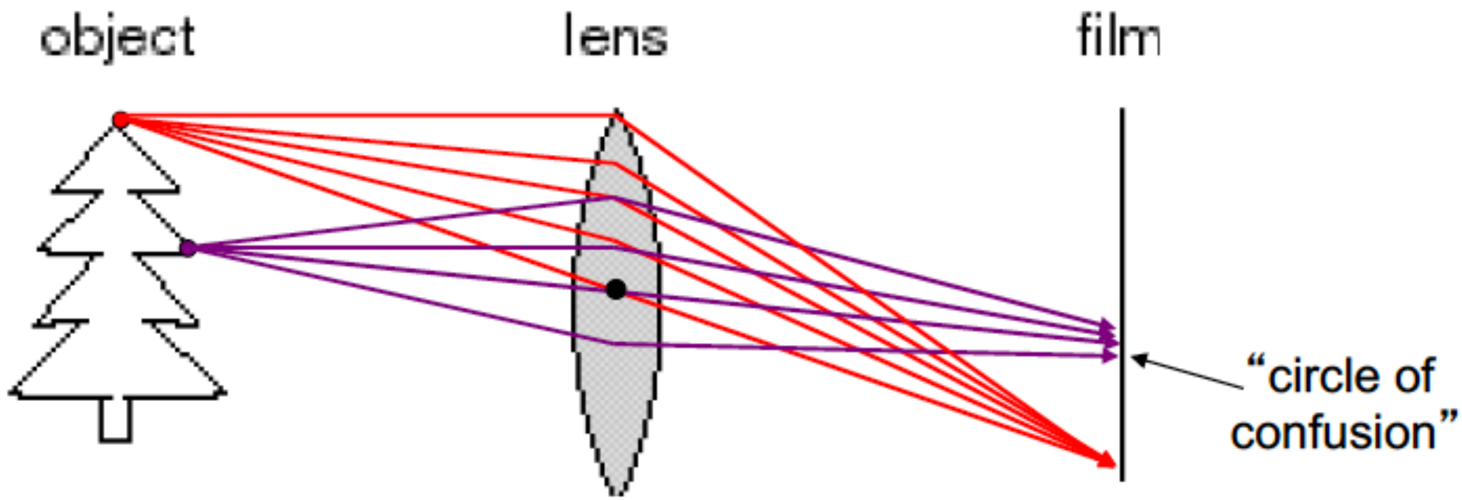
Adding a lens



- 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

Adding a Lens



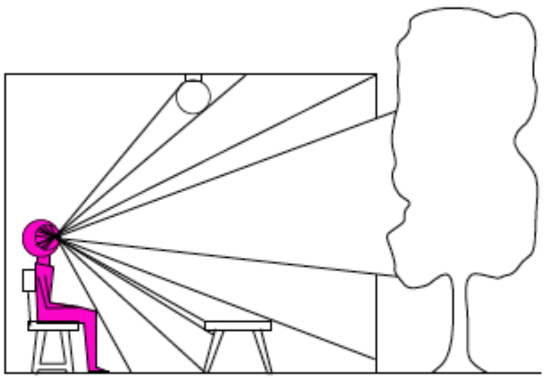
- A lens focuses light onto the film
  - There is a specific distance at which objects are "in focus"
    - other points project to a "circle of confusion" in the image

Slide by Steve Seitz

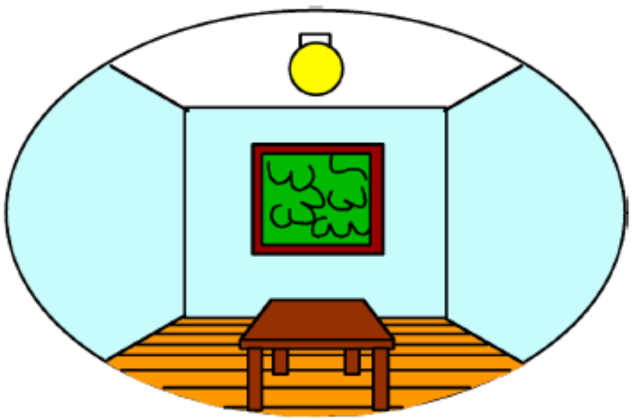
Geometric Image Formation: Pinhole Camera Model

(brief introduction, details to be covered in Multiple-view geometry)

3D world *2D image*



Point of observation



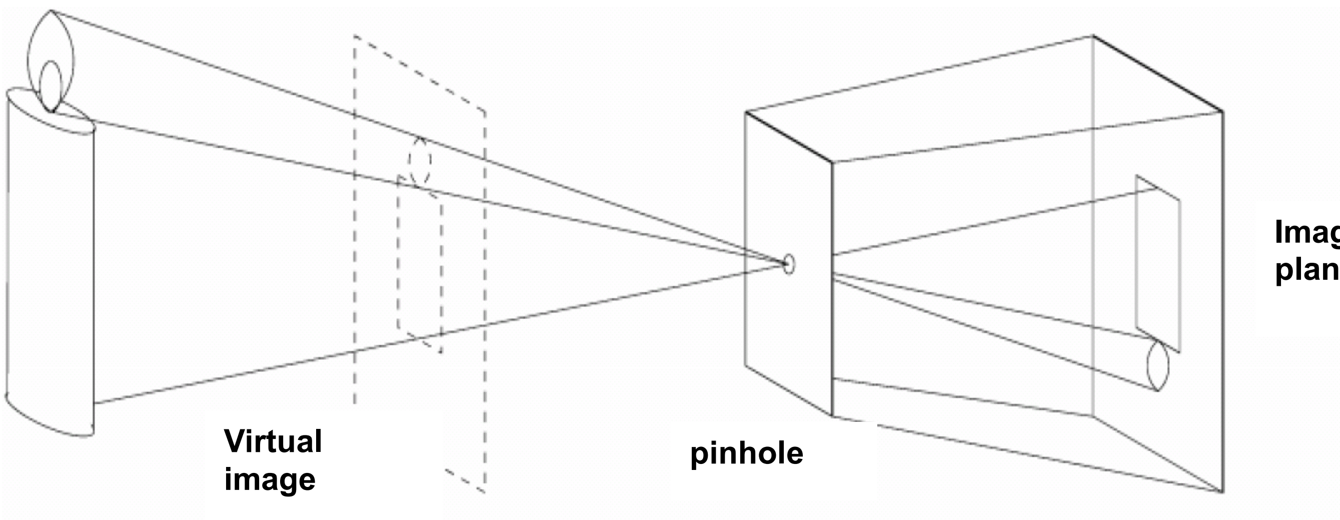
What have we lost?

- Angles
- Distances (lengths)

Slide by A. Efros

Pinhole camera

- Pinhole camera is a simple model to approximate imaging process: **perspective projection**.

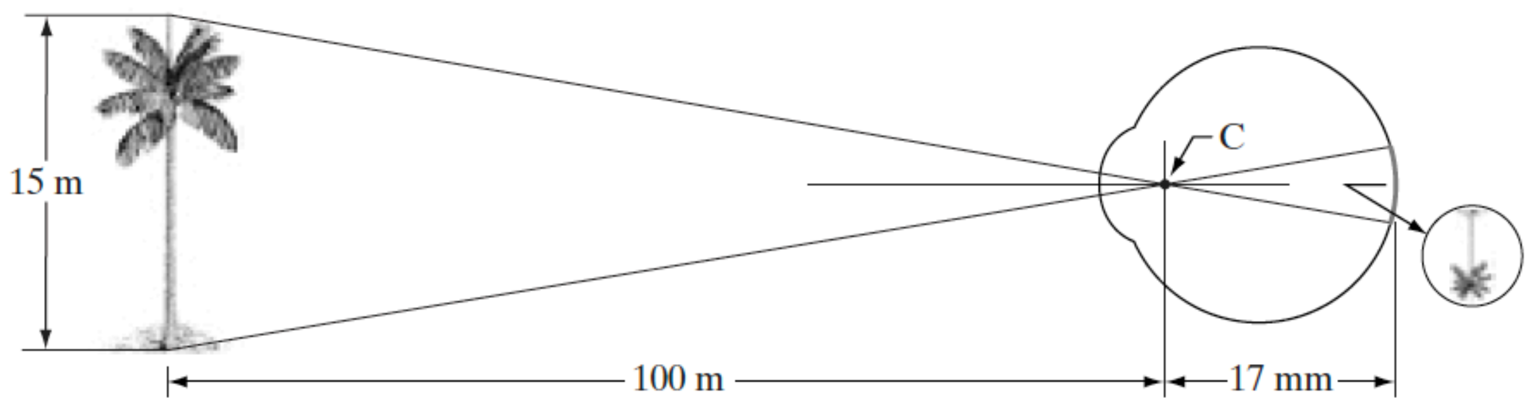


- If we treat pinhole as a point, only one ray from any given point can enter the camera.

Fig from Forsyth and Ponce

Image Formation in the Eye

FIGURE 2.3 Graphical representation of the eye looking at a palm tree. Point *C* is the optical center of the lens.



(reading pp37-38, Image formation in the eye)

Book: Digital Image Processing, Second Edition

Linear Algebra

① $P = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ $\quad P \in R^3$

② $p = \begin{pmatrix} u \\ v \end{pmatrix}$ $\quad p \in R^2$

③ P and p are vectors.

④ I: identity matrix

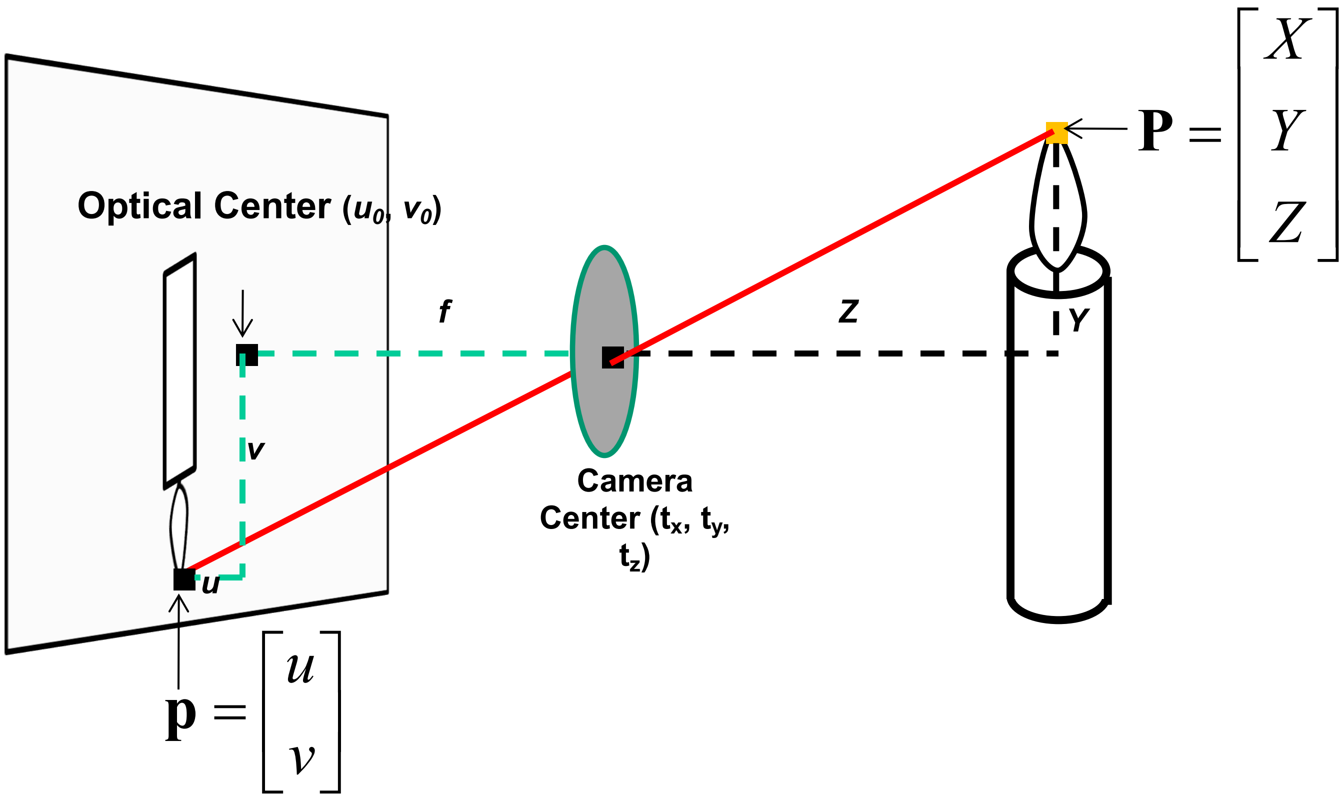
$[I \ \ 0]$

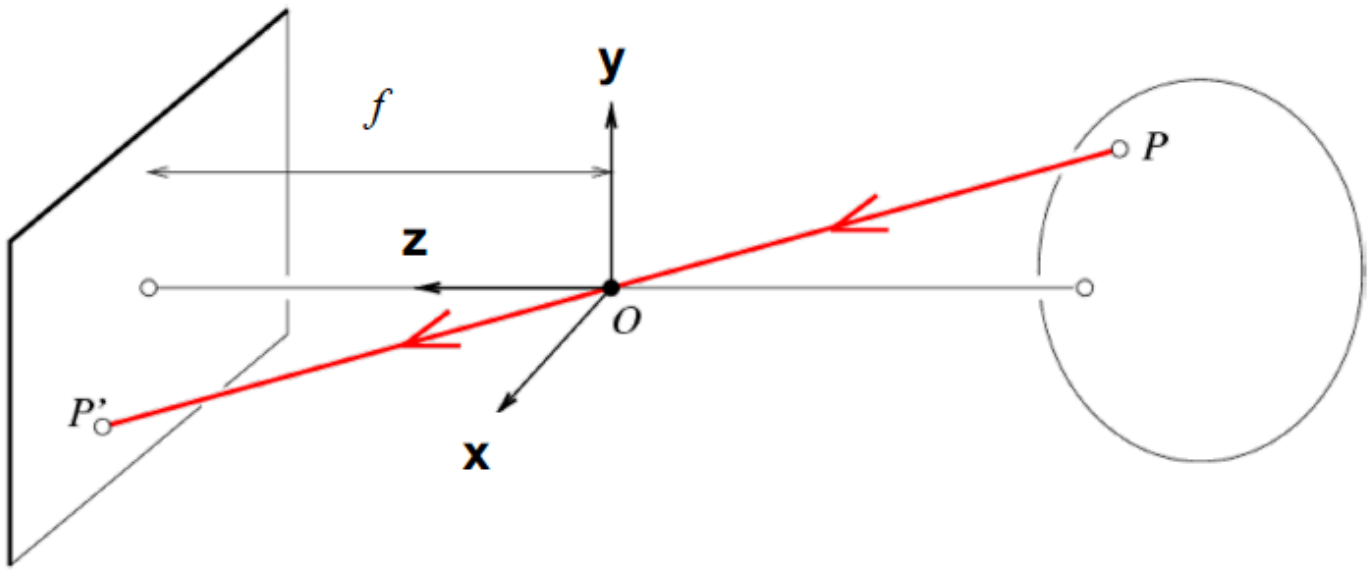
Assume $I \in R^{3\times3}$

$0 \in R^{3\times1}$

then: $[I \ \ 0]_{3\times4}$

Projection: world coordinates → image coordinates





- Projection equations
  - Compute intersection with image plane of ray from P = (x,y,z) to O
  - Derived using similar triangles

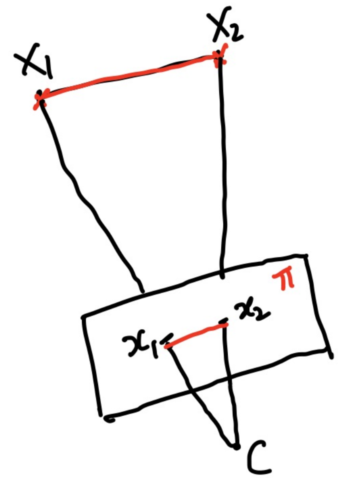
$$(x,y,z) \rightarrow (f\frac{x}{z}, f\frac{y}{z}, f)$$

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

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

Source: J. Ponce, S. Seitz

Explanation of scale and size of the object in the image



$$X_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$$

Assume: $z_1 = z_2 = z$, camera focal length f

then: $x_1 = \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} = \begin{pmatrix} f\frac{x_1}{z} \\ f\frac{y_1}{z} \end{pmatrix}$, $x_2 = \begin{pmatrix} u_2 \\ v_2 \end{pmatrix} = \begin{pmatrix} f\frac{x_2}{z} \\ f\frac{y_2}{z} \end{pmatrix}$

$$l_{12} = \begin{pmatrix} f\frac{x_2 - x_1}{z} \\ f\frac{y_2 - y_1}{z} \end{pmatrix}, \quad \|l_{12}\| = \frac{f}{z}\sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$$

Homogeneous coordinates Representation

Conversion

- Converting from Cartesian *to homogeneous* coordinates

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

homogeneous image coordinates

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

homogeneous scene coordinates

- Converting *from* homogeneous coordinates to cartesian

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

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

Homogeneous Coordinates

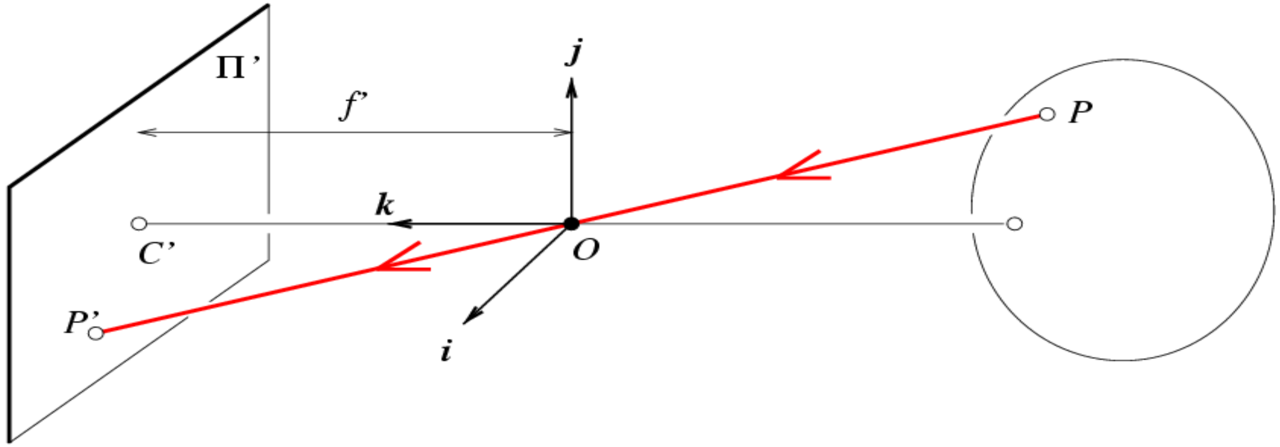
- Invariant to scaling

$$k\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} kx \\ ky \\ kw \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{kx}{kw} \\ \frac{ky}{kw} \end{bmatrix} = \begin{bmatrix} \frac{x}{w} \\ \frac{y}{w} \end{bmatrix}$$

Homogeneous Coordinates **Cartesian Coordinates**

- Point in cartesian is a ray in homogeneous

Projection matrix



Intrinsic Assumptions

- Unit aspect ratio
- Optical center at (0,0)
- No skew

Extrinsic Assumptions

- No rotation
- Camera at (0,0,0)

K

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

Slide Credit: Saverese

Remove assumption of known optical center

Intrinsic Assumptions

- Unit aspect ratio
- No skew

Extrinsic Assumptions

- No rotation
- Camera at (0,0,0)

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

Remove assumption: square pixels

Intrinsic Assumptions

- No skew

Extrinsic Assumptions

- No rotation
- Camera at (0,0,0)

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

Remove assumption: non-skewed pixels

Intrinsic Assumptions

Extrinsic Assumptions

- No rotation
- Camera at (0,0,0)

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

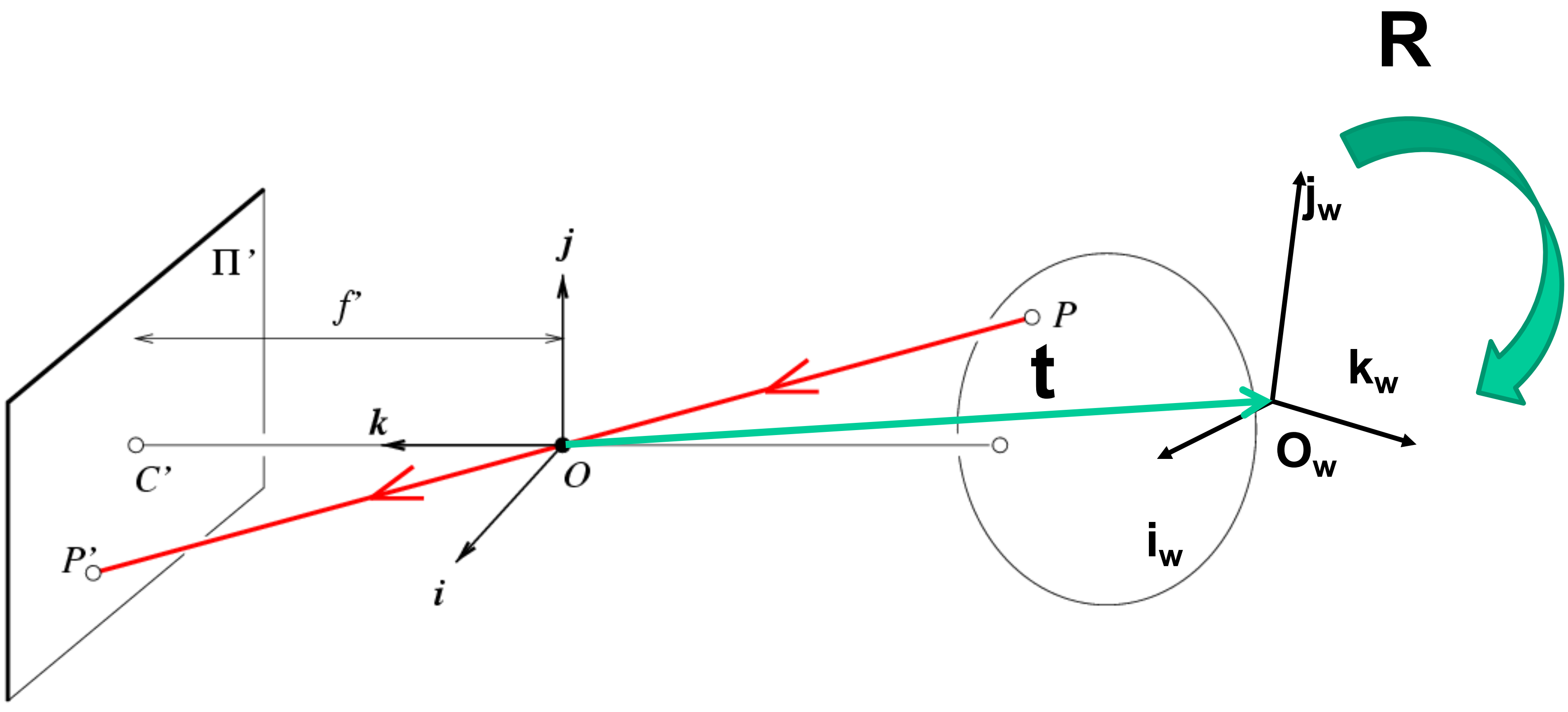
u0, v0: optical center

α, β: focal length on x, y direction, normally the same (f) for normal lens

s: distortion

Note: different books use different notation for parameters

Rotated and Translated Camera



Allow Camera Translation

Intrinsic Assumptions

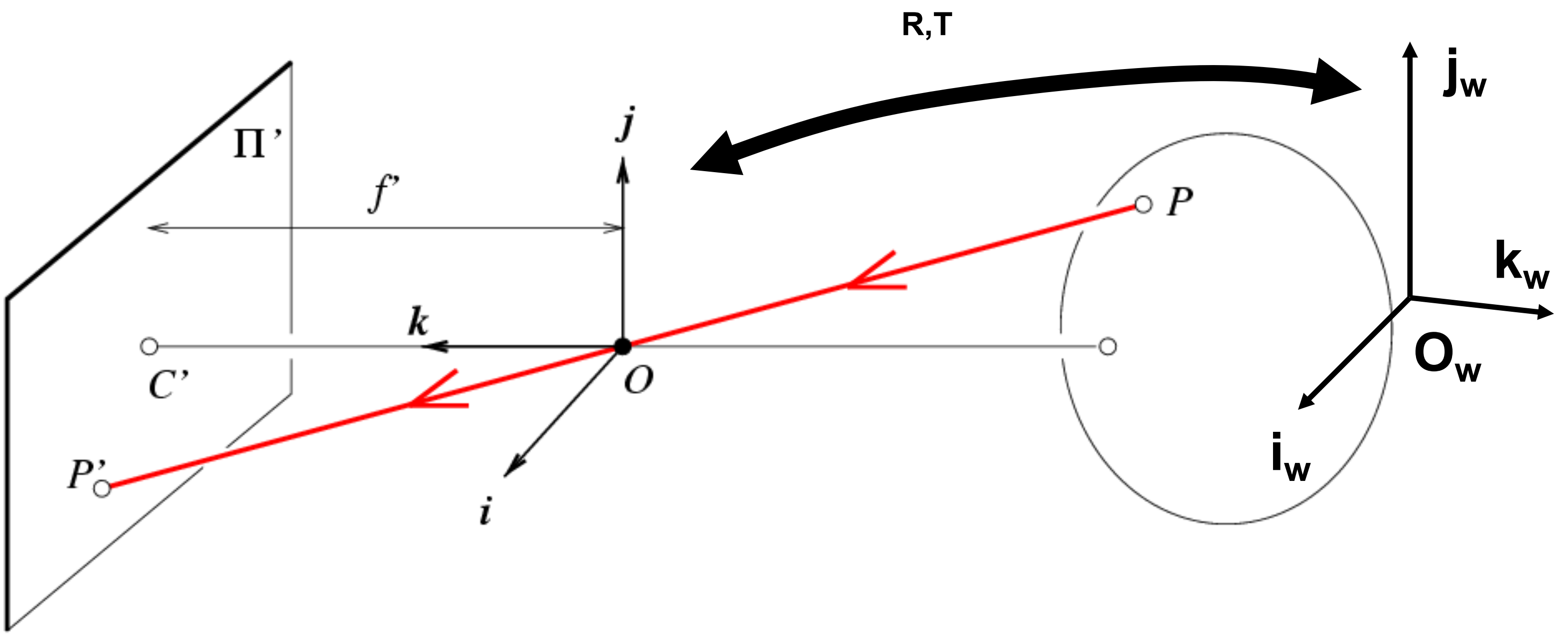
Extrinsic Assumptions

- No rotation

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

t: translation victor to describe the difference between the world coordinate system and camera coordinate system.

General Camera 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)

$$\mathrm{R} = \mathrm{R}_z(\alpha)\mathrm{R}_y(\beta)\mathrm{R}_x(\gamma)$$

General Camera Projection Matrix

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

$$w\begin{bmatrix}u\\ v\\ 1\end{bmatrix} = \underset{K}{\begin{bmatrix}\alpha & s & u_0\\ 0 & \beta & v_0\\ 0 & 0 & 1\end{bmatrix}}\left[\begin{matrix}\underbrace{\begin{matrix}r_{11} & r_{12} & r_{13}\\ r_{21} & r_{22} & r_{23}\\ r_{31} & r_{32} & r_{33}\end{matrix}}_{R} & \begin{matrix}t_x\\ t_y\\ t_z\end{matrix}\end{matrix}\right]\underset{X}{\begin{bmatrix}x\\ y\\ z\\ 1\end{bmatrix}}$$

Degrees of freedom

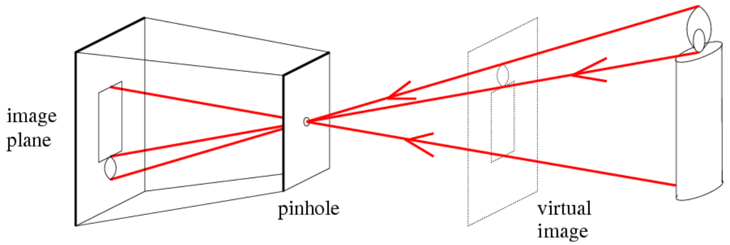
$$\mathbf{x} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\mathbf{X}$$

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \overset{5}{\begin{bmatrix} \alpha & s & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}} \overset{6}{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

A quick summary

- Pinhole camera model and camera projection matrix

- Homogeneous coordinates



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

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

Perspective effects

3D computer vision aims to solve an inverse problem of computer graphics

Photometric (radiometric) image formation

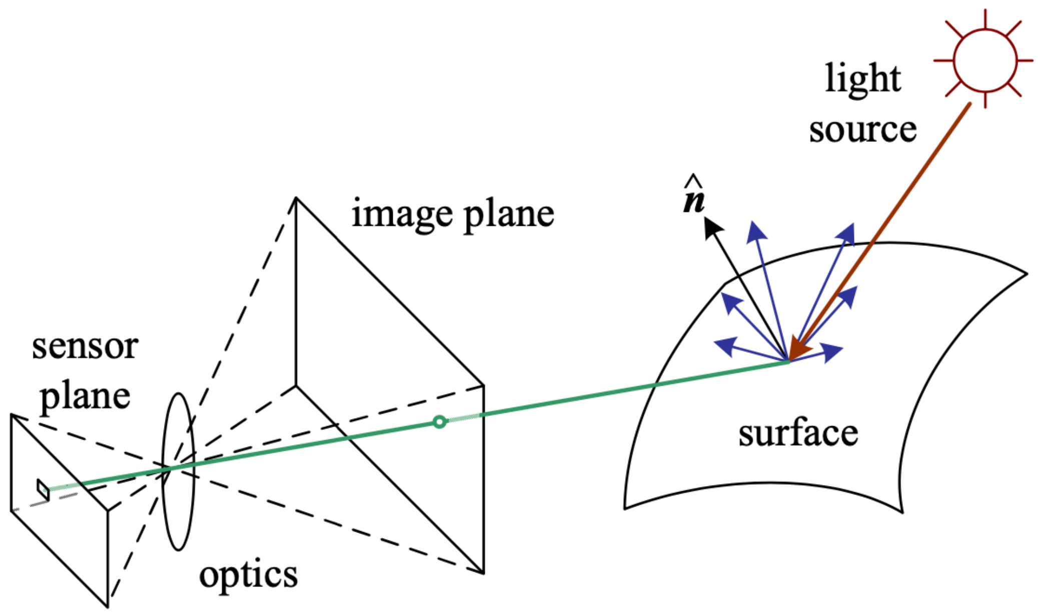
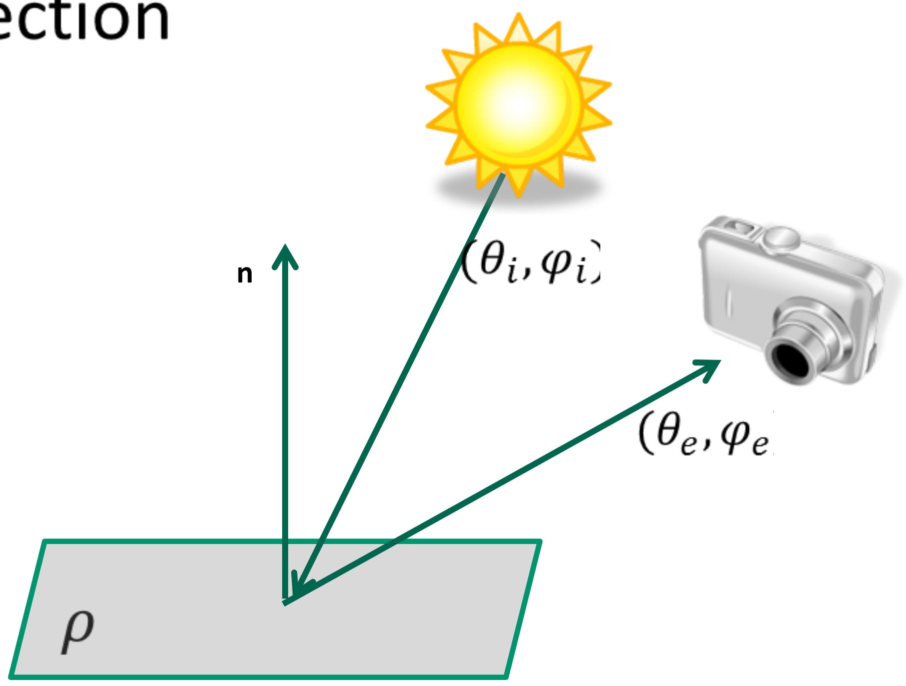


Figure 2.14 A simplified model of photometric image formation. Light is emitted by one or more light sources and is then reflected from an object's surface. A portion of this light is directed towards the camera. This simplified model ignores multiple reflections, which often occur in real-world scenes.

What determines the scene radiance?

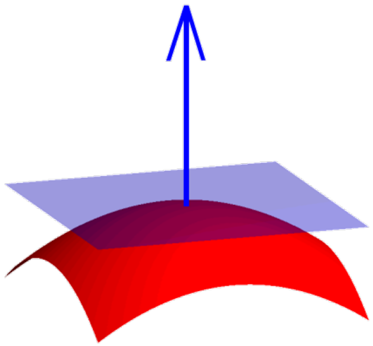
- The amount of light that falls on the surface
- The fraction of light that is reflected (albedo)
- Geometry of light reflection
  - Shape of surface
  - Viewpoint



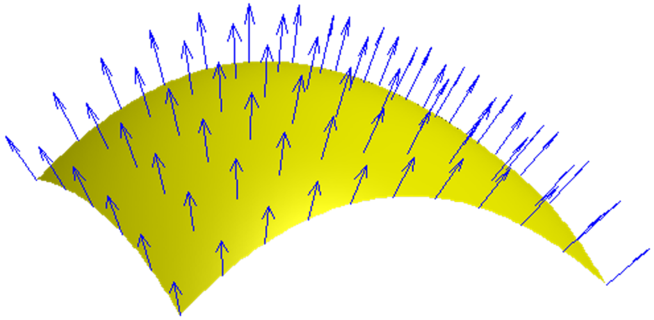
Surface Normal

Convenient notation for surface orientation

A smooth surface has a tangent plane at every point



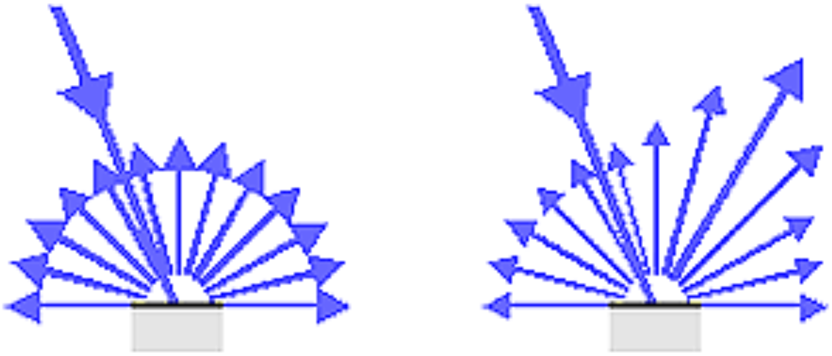
We can model the surface using the normal at every point



Lambertian surface

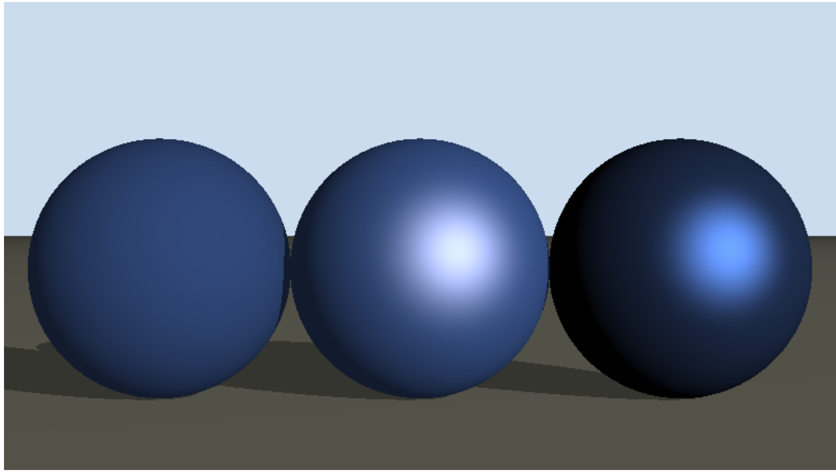
- Ideal diffuse reflectors.

- The apparent brightness of such a surface to an observer is the same regardless of the observer's angle of view.



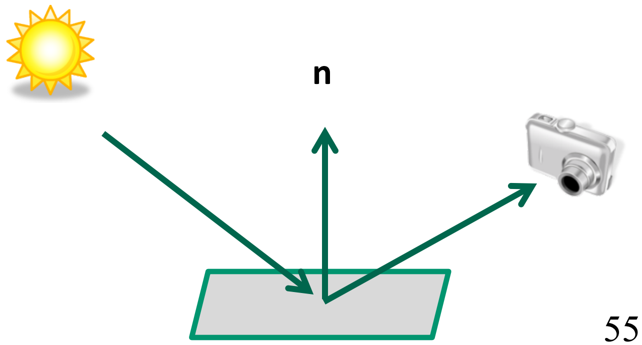
Ideal diffuse reflection (Lambertian surface)

Diffuse reflection with directional component

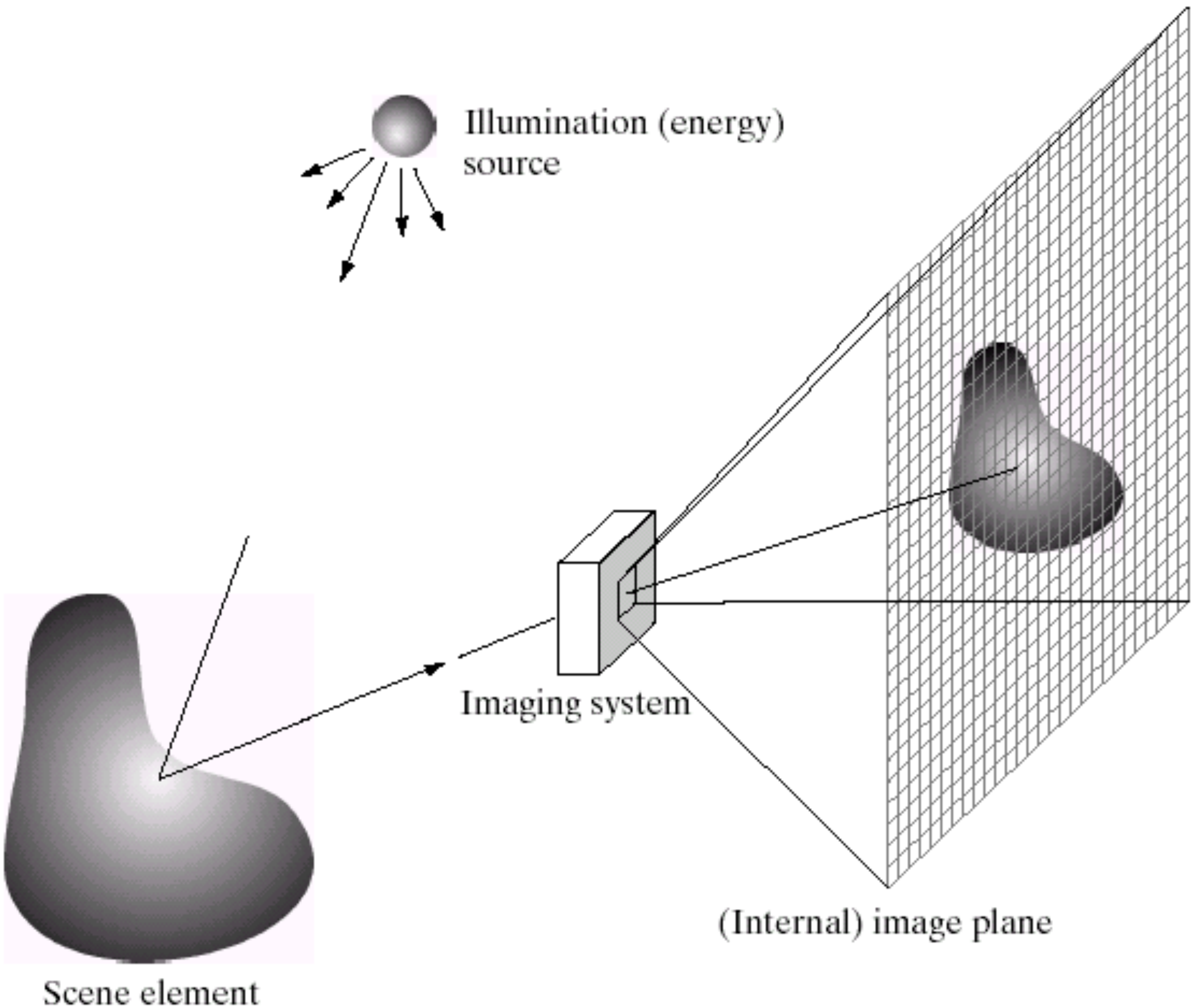


Lambertian Surface

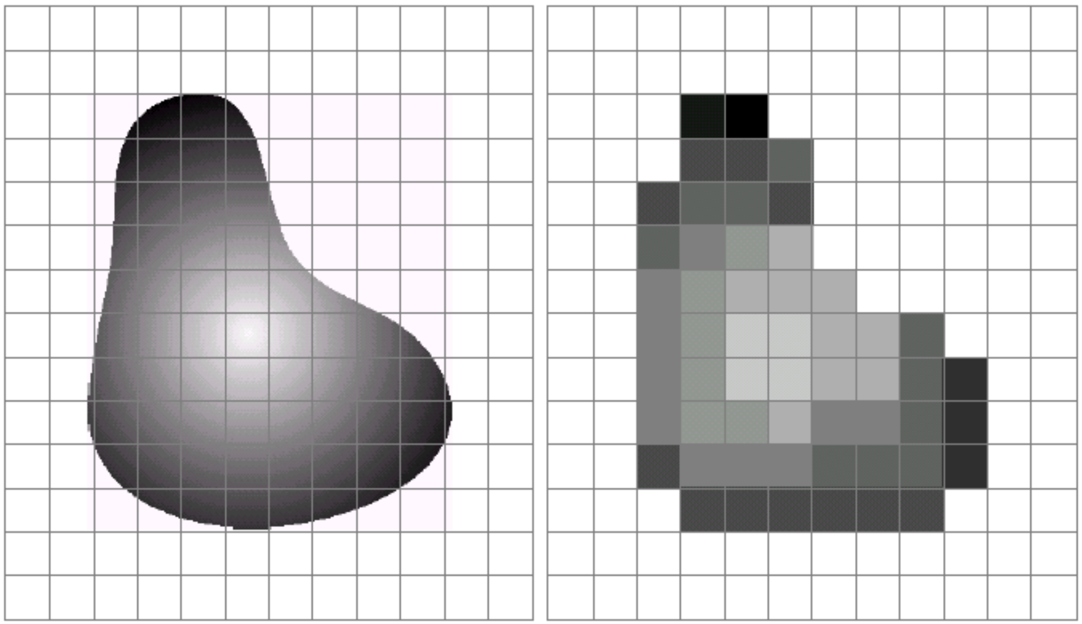
- Appears equally bright from all viewing directions
- Reflects all light without absorbing
- Matte surface, no “shiny” spots
- Brightness of the surface as seen from camera is linearly correlated to the amount of light falling on the surface



Photometric Image Formation

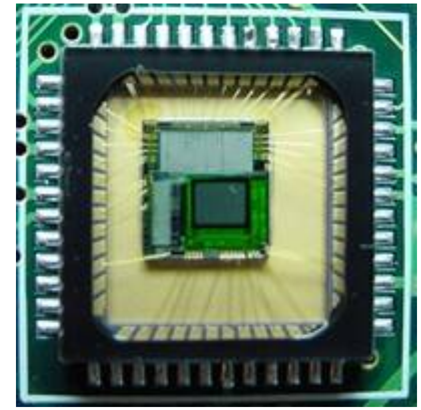


Sensor Array



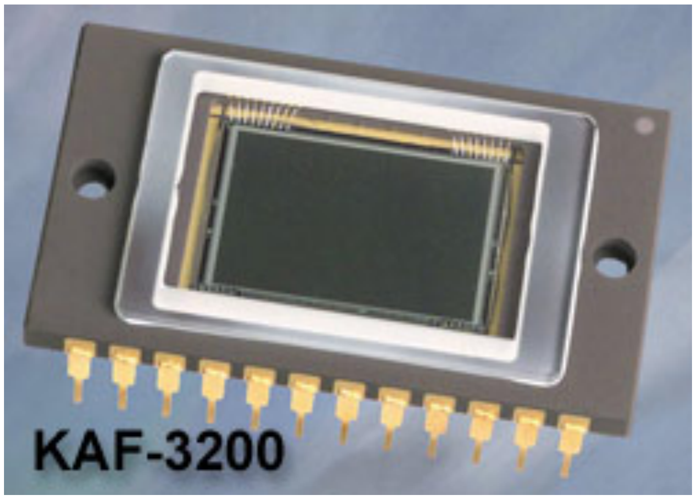
a b

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



CCD/CMOS sensor

Image Sensors : Array Sensor



CCD KAF-3200E from Kodak.

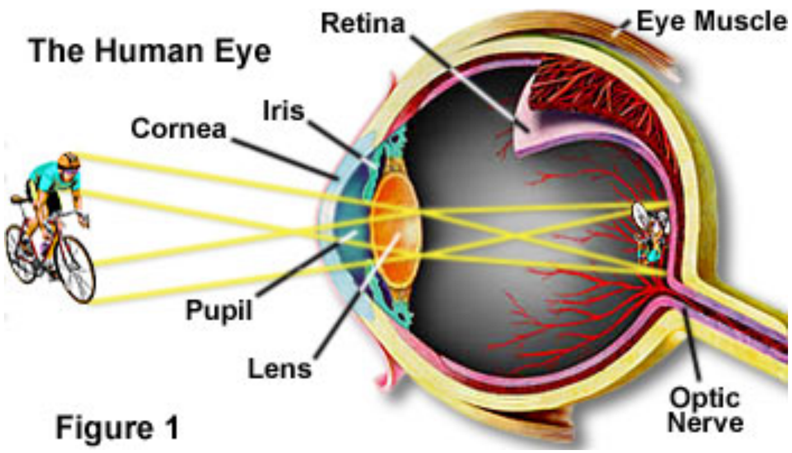
(2184 x 1472 pixels,

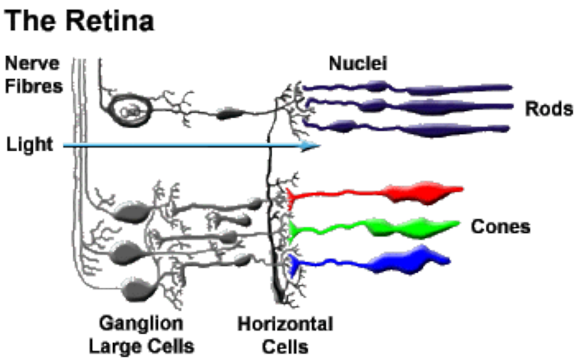
Pixel size 6.8 microns^2)

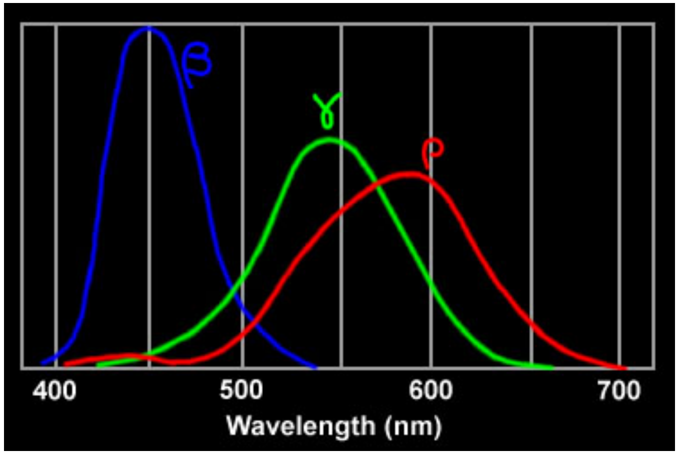
Charge-Coupled Device (CCD)

- Used for convert a continuous image into a digital image
- Contains an array of light sensors
- Converts photon into electric charges accumulated in each sensor unit

Human Color Perception







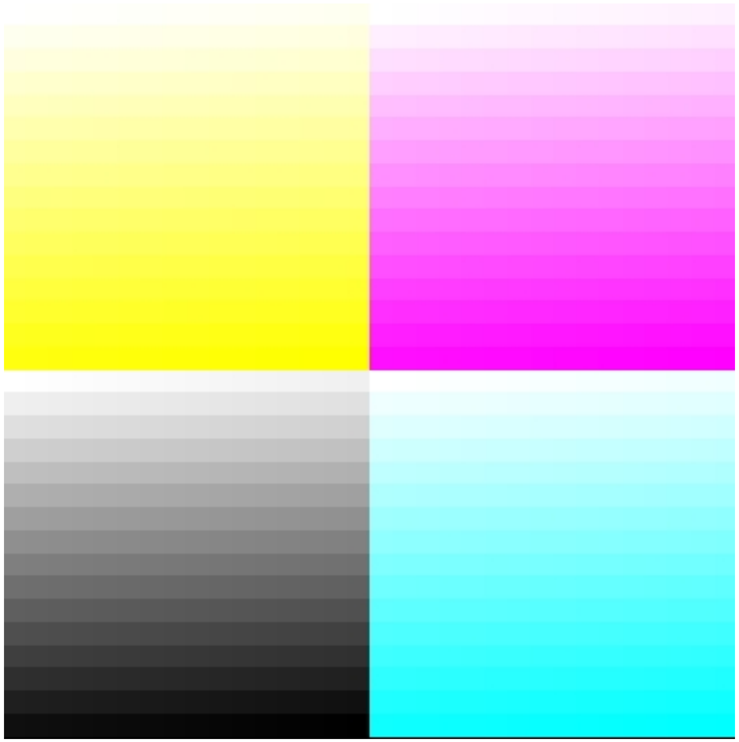
A website about human color perception:

- http://www.photo.net/photo/edscott/vis00010.htm

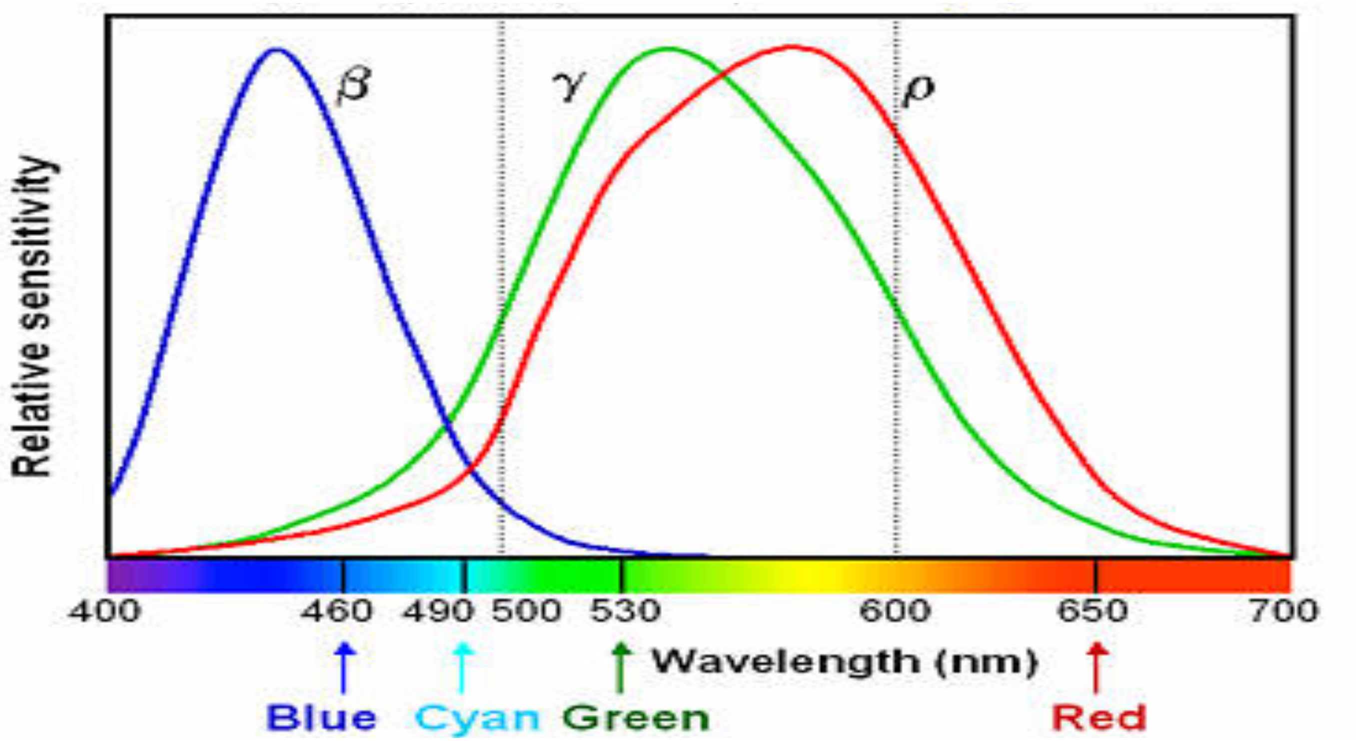
How many Colors can be Seen?

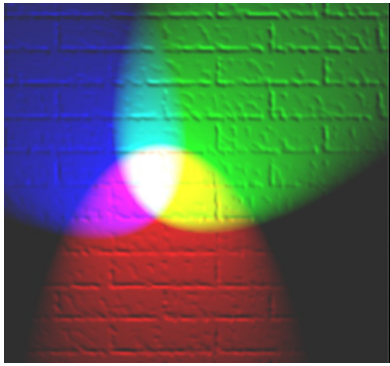
- Human eyes can distinguish about
  - 128 different hues
  - 130 different saturation levels

- We can distinguish about 380,000 colours!!



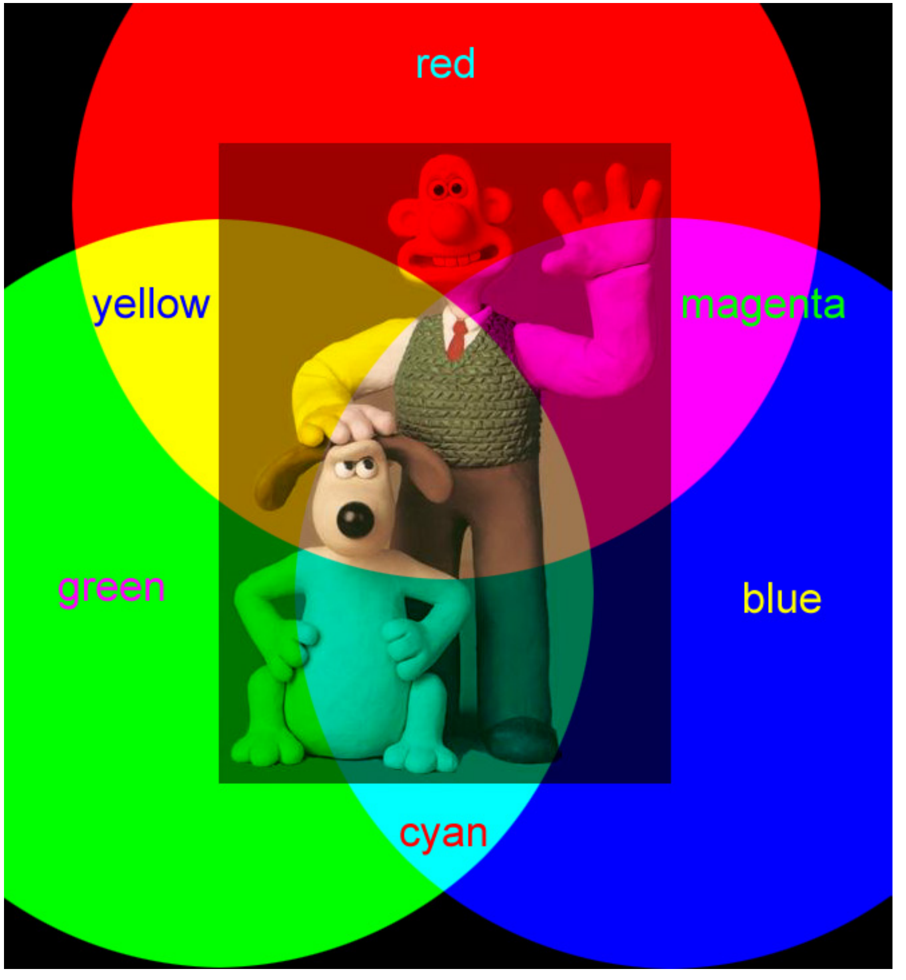
- Wavelength of the light





Colour Images

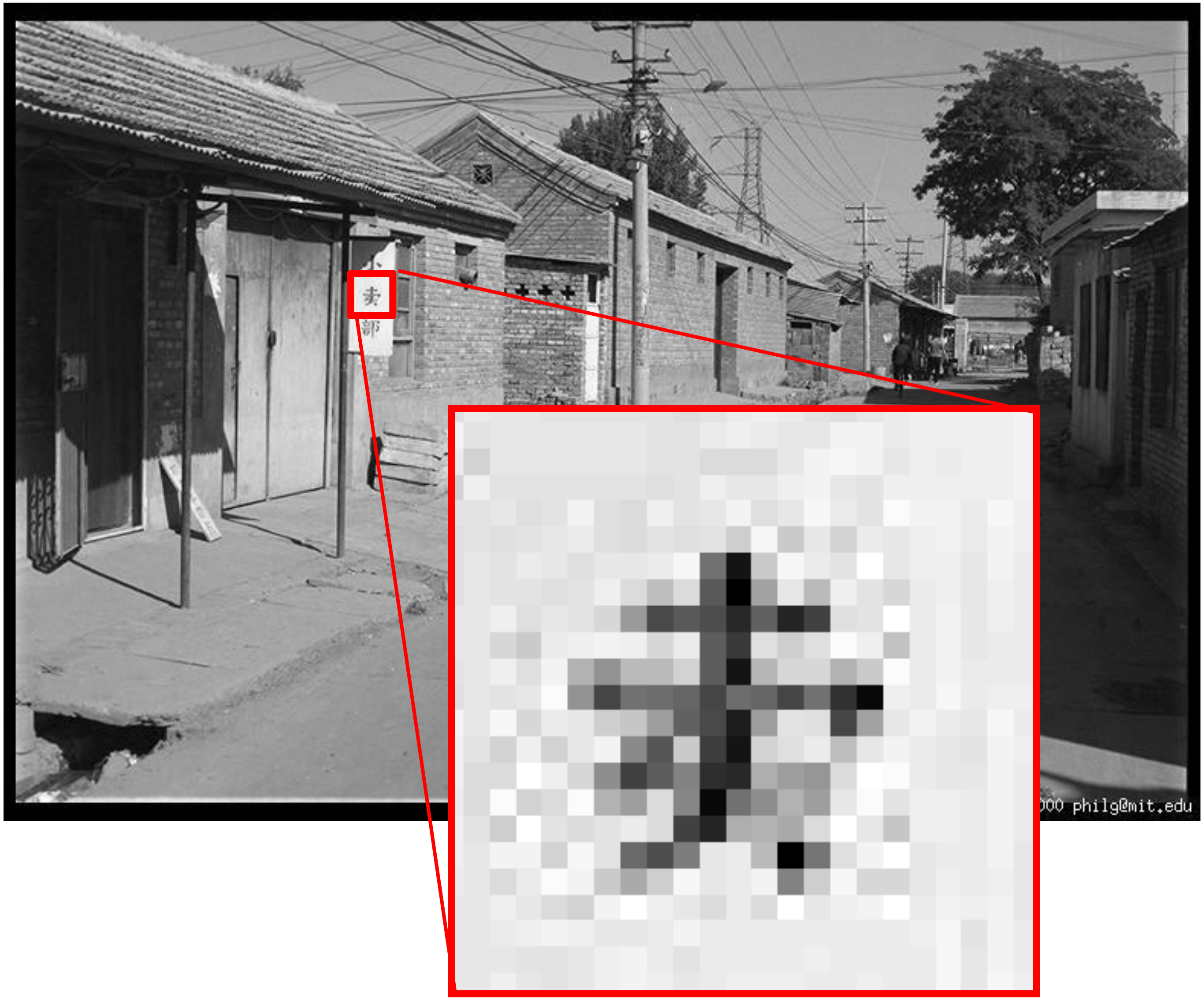
- are constructed from three intensity maps.
- Each intensity map is projected through a colour filter (*e.g.,* red, green, or blue, or cyan, magenta, or yellow) to create a monochrome image.
- The intensity maps are overlaid to create a color image.
- Each pixel in a color image is a three dimensional vector.



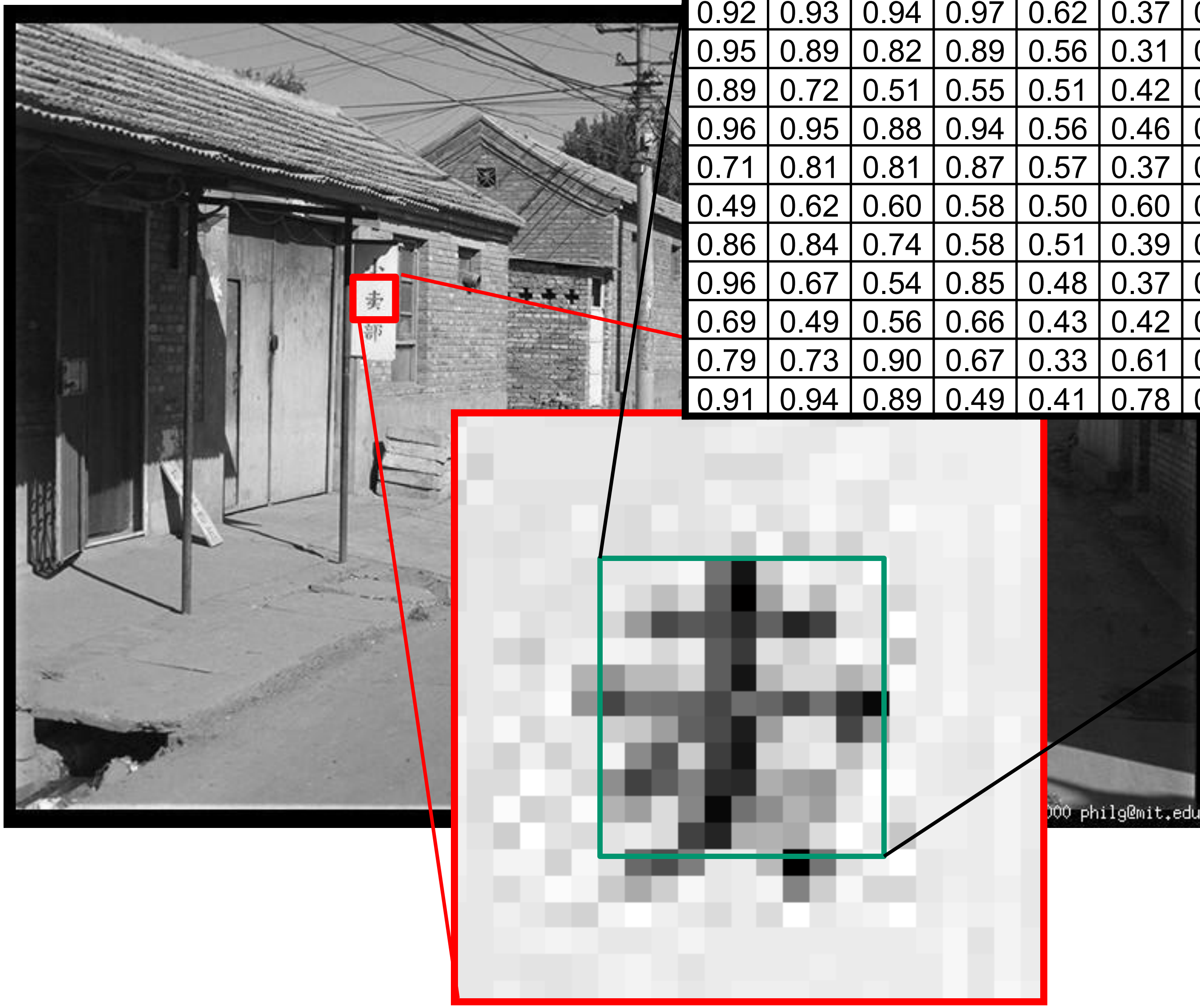
62

Digital image representation

The digital image (pixel matrix)



The digital image (pixel matrix)



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

Fundamentals of Digital Images

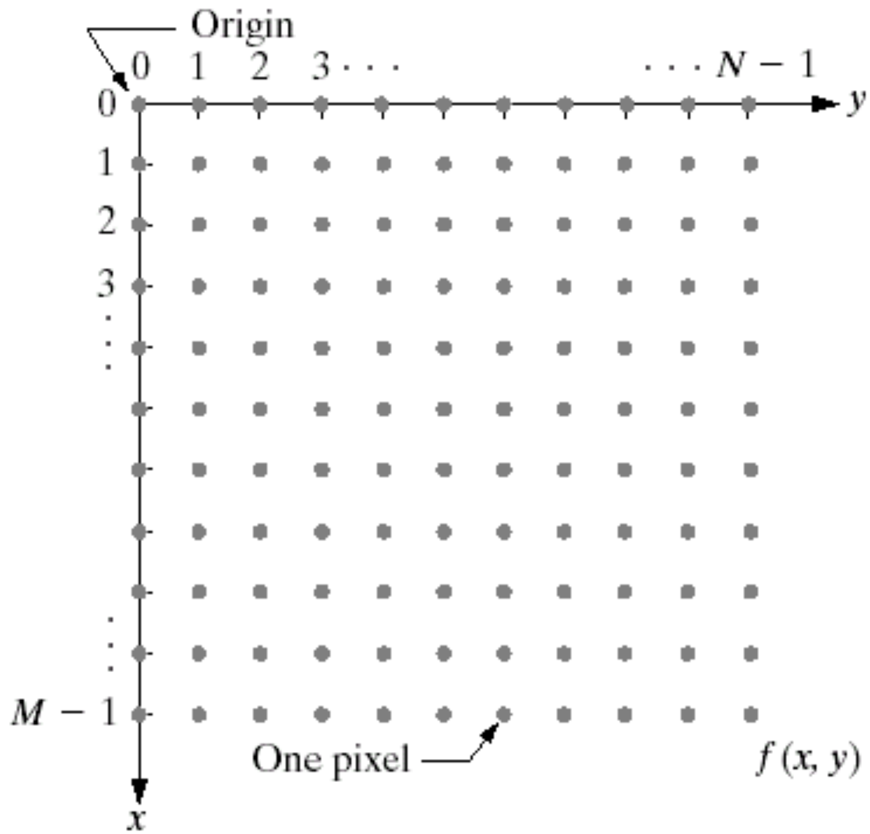
Image "After snow storm"

- An image = a function of spatial coordinates.
- Spatial coordinate: (x,y) for 2D case such as photograph,

  (x,y,z) for 3D case such as CT scan images

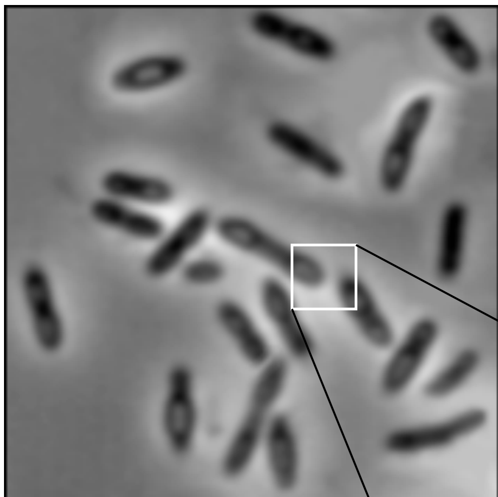
  (x,y,t) for video.
- The function f may represent the intensity (for greyscale images) or color (for color images) or other associated values.

Conventional Coordinate for Image Representation



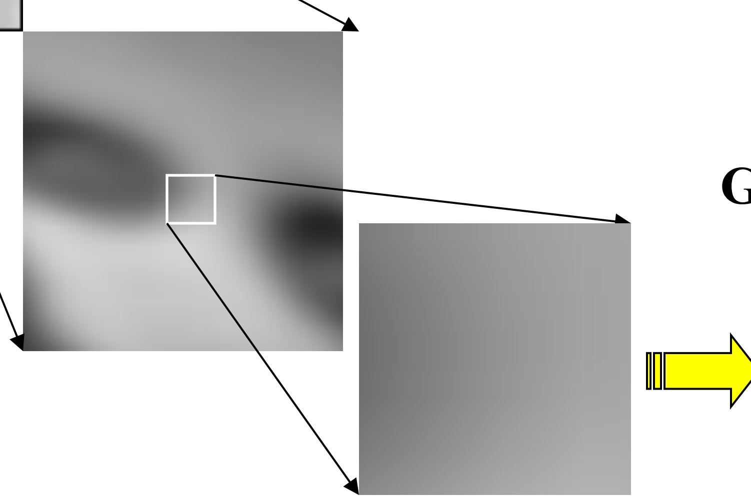
(Images from Rafael C. Gonzalez and Richard E. Wood, Digital Image Processing, 2nd Edition.

Grey-scale Image



Intensity image or grey-scale image

Each pixel corresponds to light intensity normally represented in gray scale (gray level).



Gray scale values

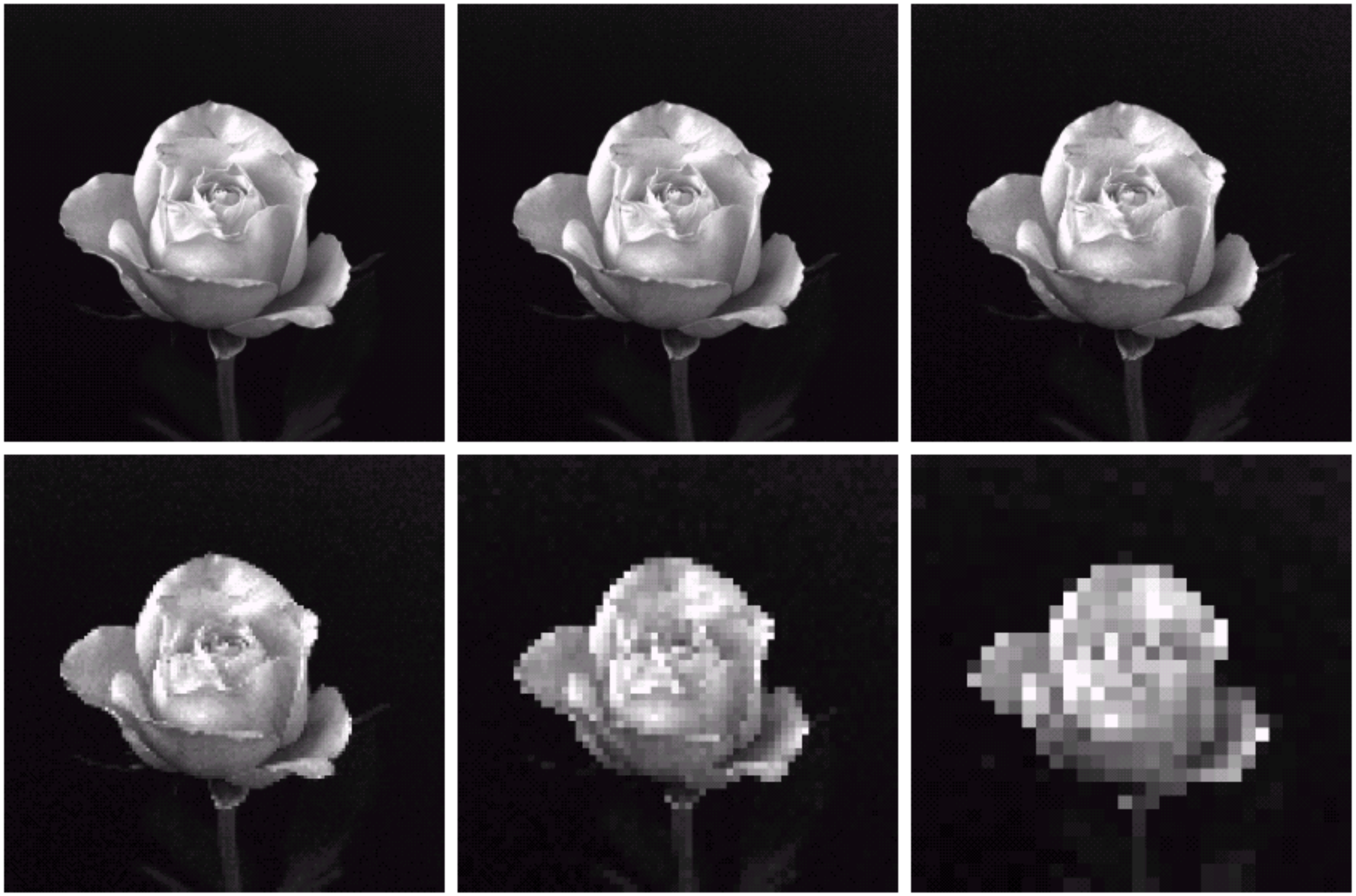
$$\begin{bmatrix} 10 & 10 & 16 & 28 \\ 9 & 6 & 26 & 37 \\ 15 & 25 & 13 & 22 \\ 32 & 15 & 87 & 39 \end{bmatrix}$$

Effect of Spatial Resolution

FIGURE 2.19 A 1024 × 1024, 8-bit image subsampled down to size 32 × 32 pixels. The number of allowable gray levels was kept at 256.

(Images from Rafael C. Gonzalez and Richard E. Wood, Digital Image Processing, 2nd Edition.

Effect of Spatial Resolution



a b c
d e f

FIGURE 2.20 (a) 1024 × 1024, 8-bit image. (b) 512 × 512 image resampled into 1024 × 1024 pixels by row and column duplication. (c) through (f) 256 × 256, 128 × 128, 64 × 64, and 32 × 32 images resampled into 1024 × 1024 pixels.

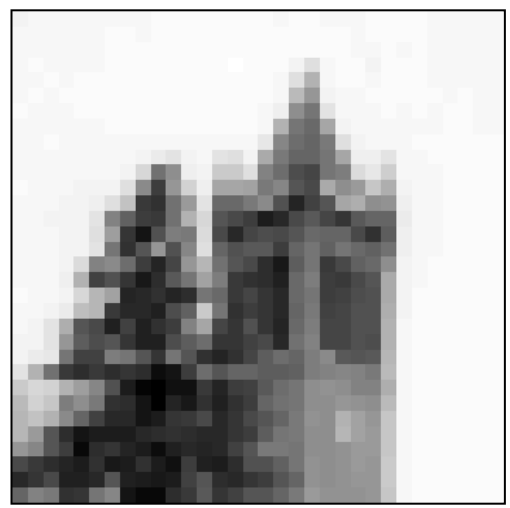
(Images from Rafael C. Gonzalez and Richard E. Wood, Digital Image Processing, 2nd Edition.

Effect of Spatial Resolution

256x256 pixels

128x128 pixels

64x64 pixels

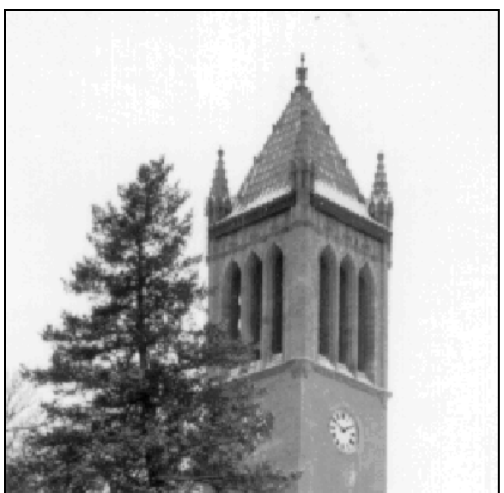


32x32 pixels

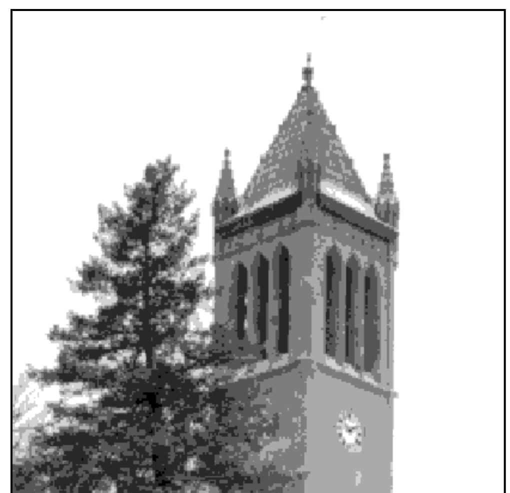
Effect of Quantization Levels

256 levels

128 levels

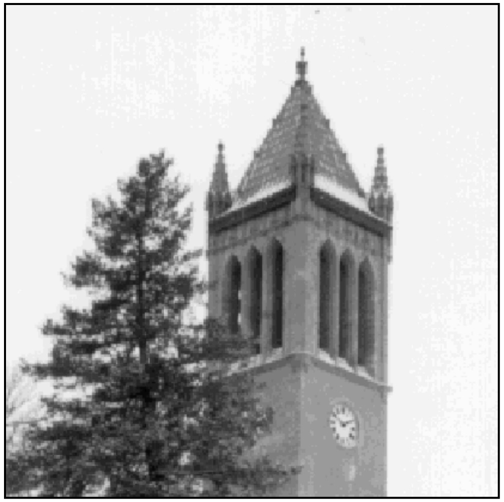


64 levels



32 levels

Effect of Quantization Levels (cont.)

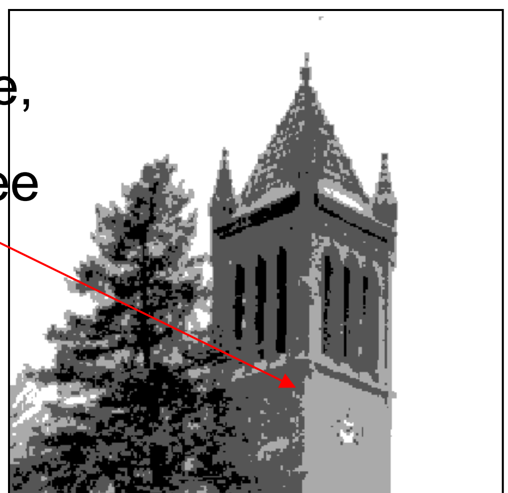

16 levels

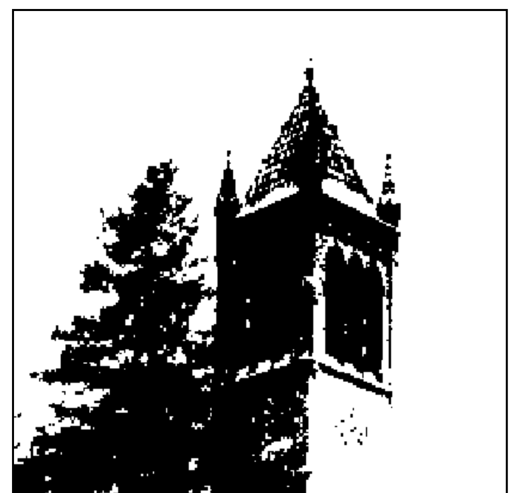

8 levels

In this image,

it is easy to see

false contour.

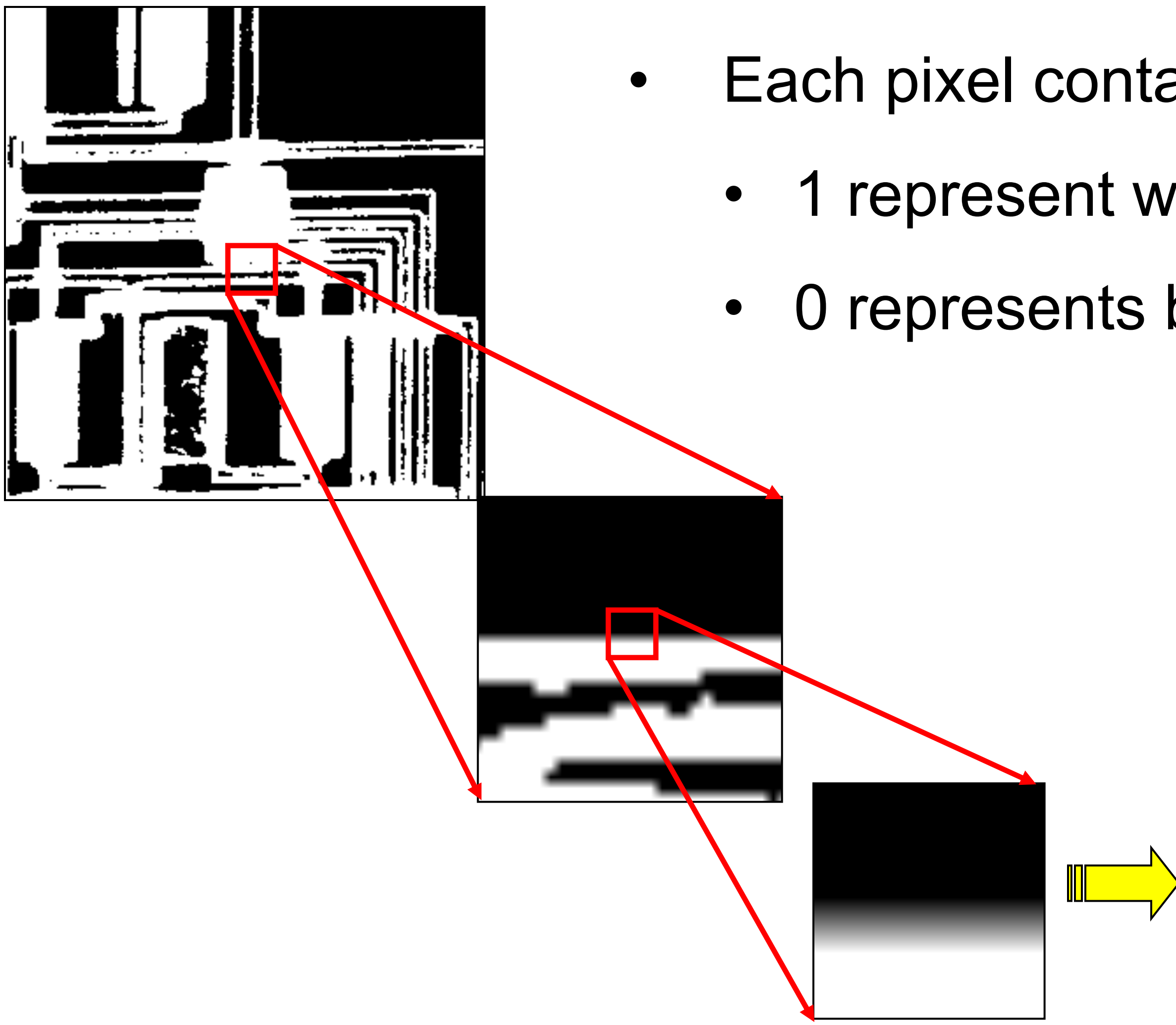

4 levels


2 levels

Two-level image: binary image

Binary image or black and white image

- Each pixel contains one bit :
  - 1 represent white
  - 0 represents black



Binary data

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Colour Images

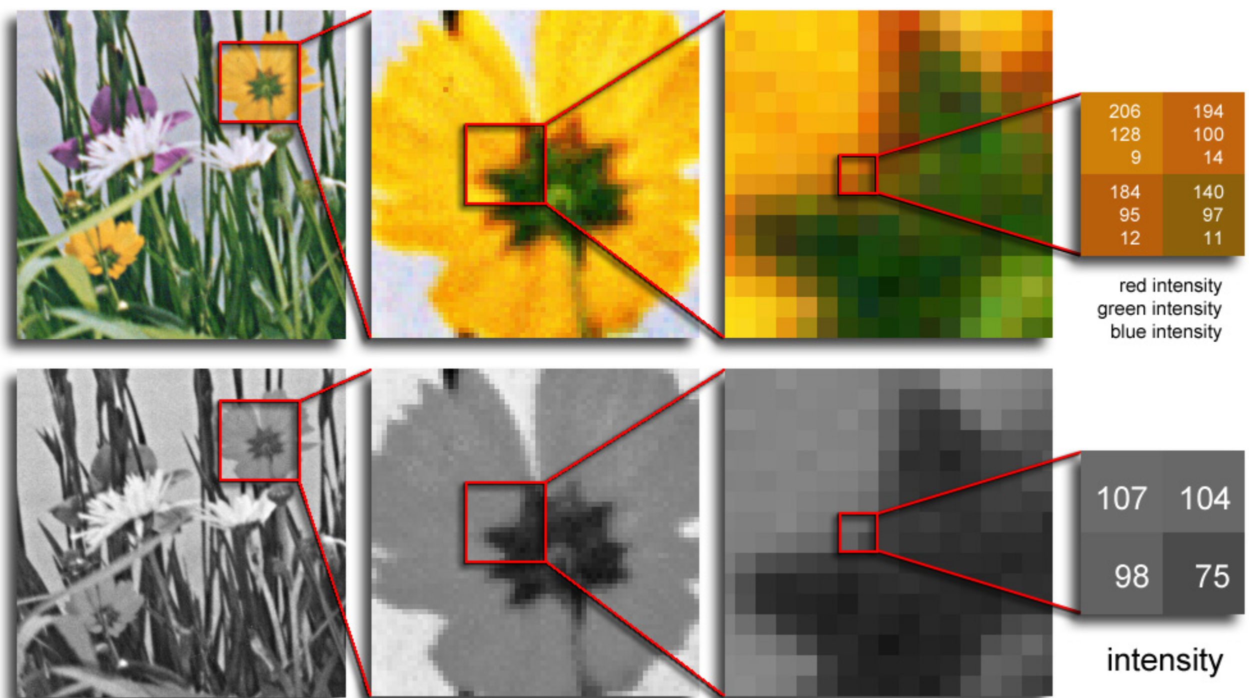
A colourful image

Colour Image

- Colour images have 3 values per pixel;
- Greyscale images have 1 value per pixel.

a grid of squares, each of which contains a single colour

each square is called a pixel (for *picture element*)



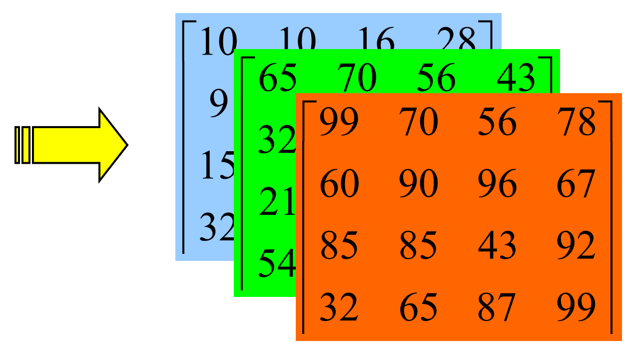
78

Colour Image

Color (RGB) image:

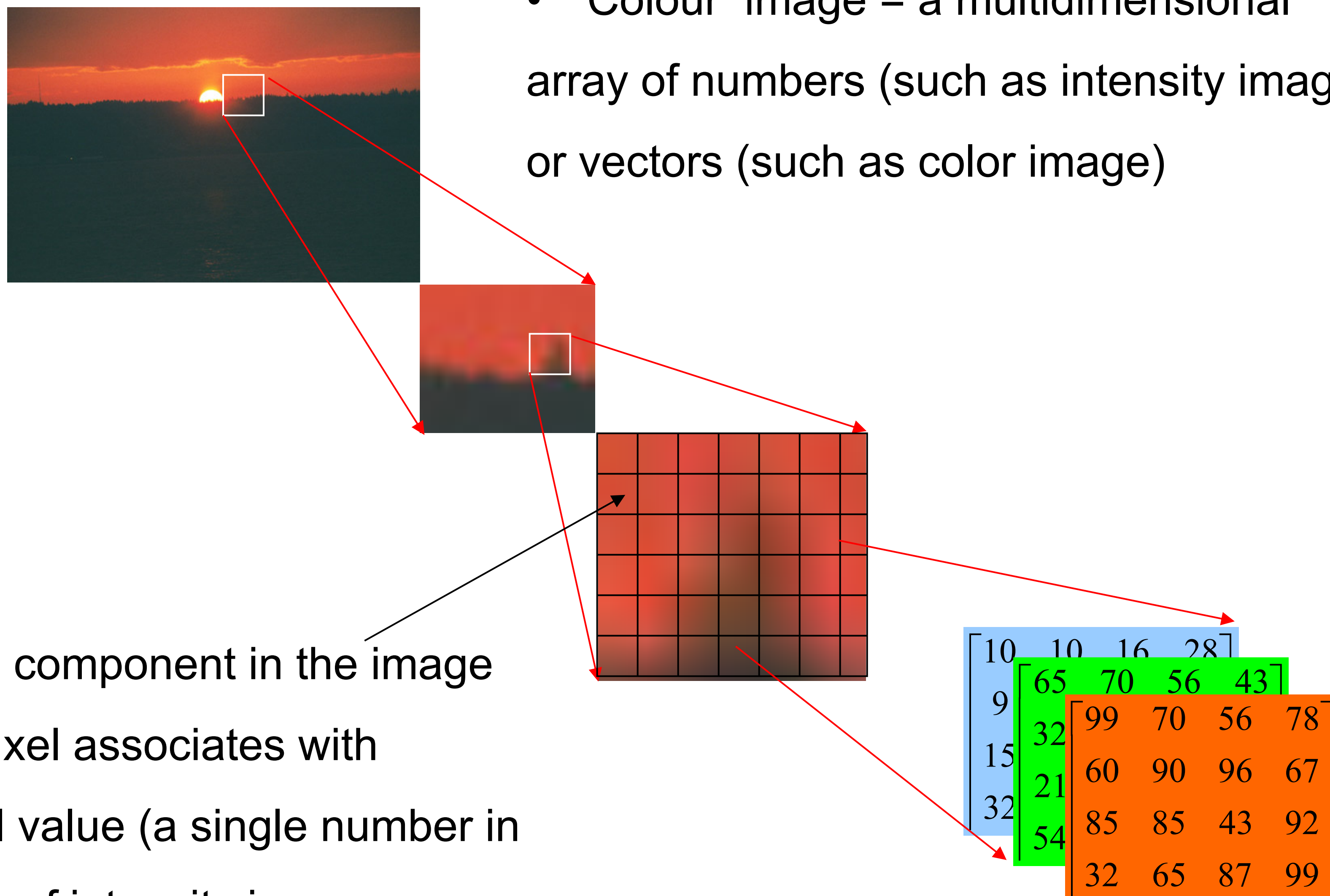
- each pixel contains a vector representing red, green and blue components.

RGB components



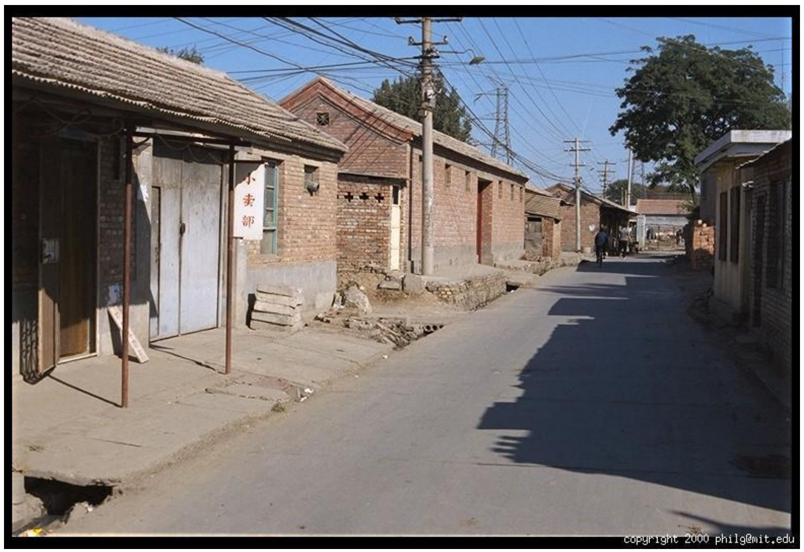
Colour Image

- Colour image = a multidimensional array of numbers (such as intensity image) or vectors (such as color image)



- Each component in the image called pixel associates with the pixel value (a single number in the case of intensity images or a vector in the case of color images).

Colour Image



Colour Images in Matlab

Images represented as a matrix

Suppose we have a NxM RGB image called "im"

- im(1,1,1) = top-left pixel value in R-channel
- im(y, x, b) = y pixels down, x pixels to right in the bth channel

imread(filename) returns a uint8 image (values 0 to 255)

- **Convert to double format (values 0 to 1) with im2double(). (important !)**

column

row

R

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

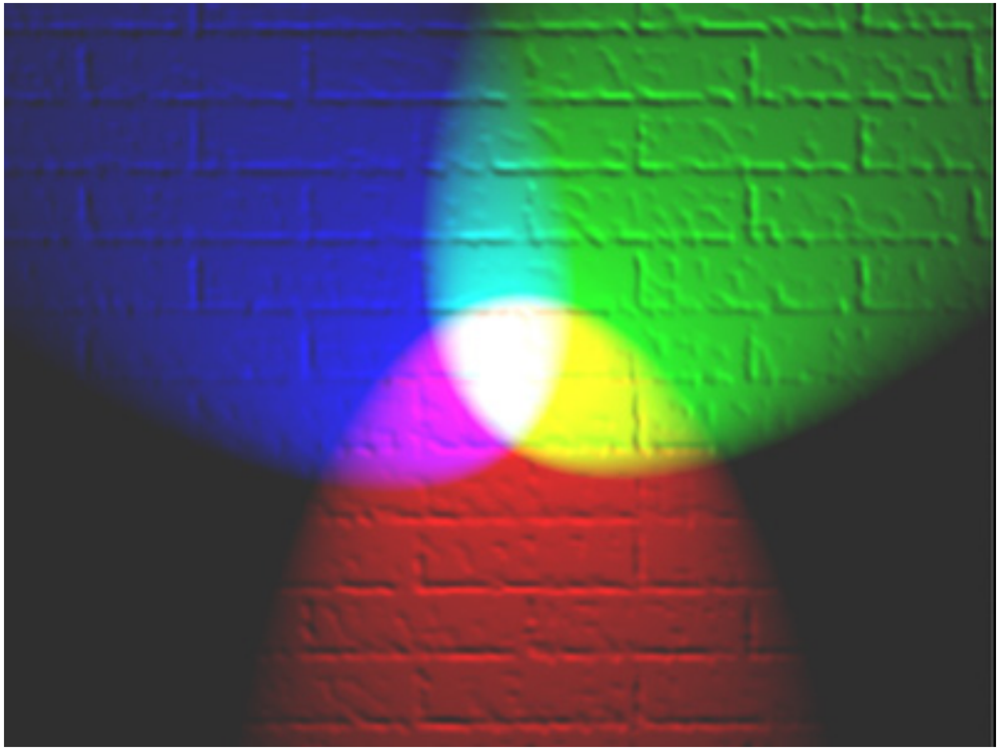
G

									0.92	0.99
									0.95	0.91
									0.91	0.92
									0.97	0.95
									0.79	0.85
									0.45	0.33
									0.49	0.74
									0.82	0.93
									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

B

									0.92	0.99
									0.95	0.91
									0.91	0.92
									0.97	0.95
									0.79	0.85
									0.45	0.33
									0.49	0.74
									0.82	0.93
									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

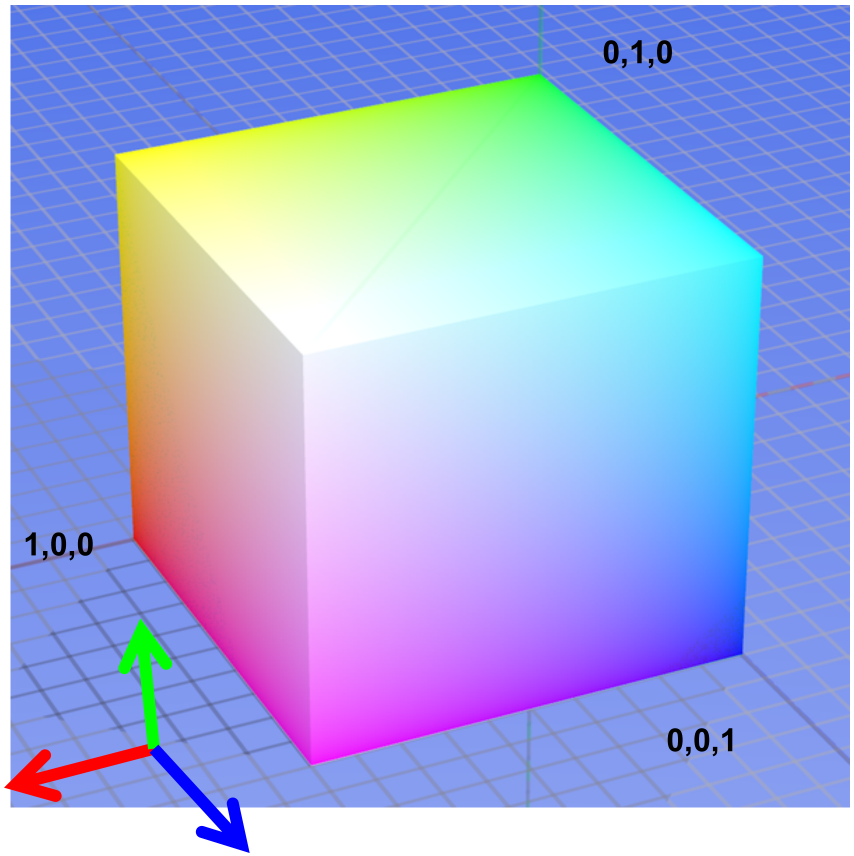
Color spaces



http://en.wikipedia.org/wiki/File:RGB_illumination.jpg

RGB space

Default color space



R

(G=0,B=0)

G

(R=0,B=0)

B

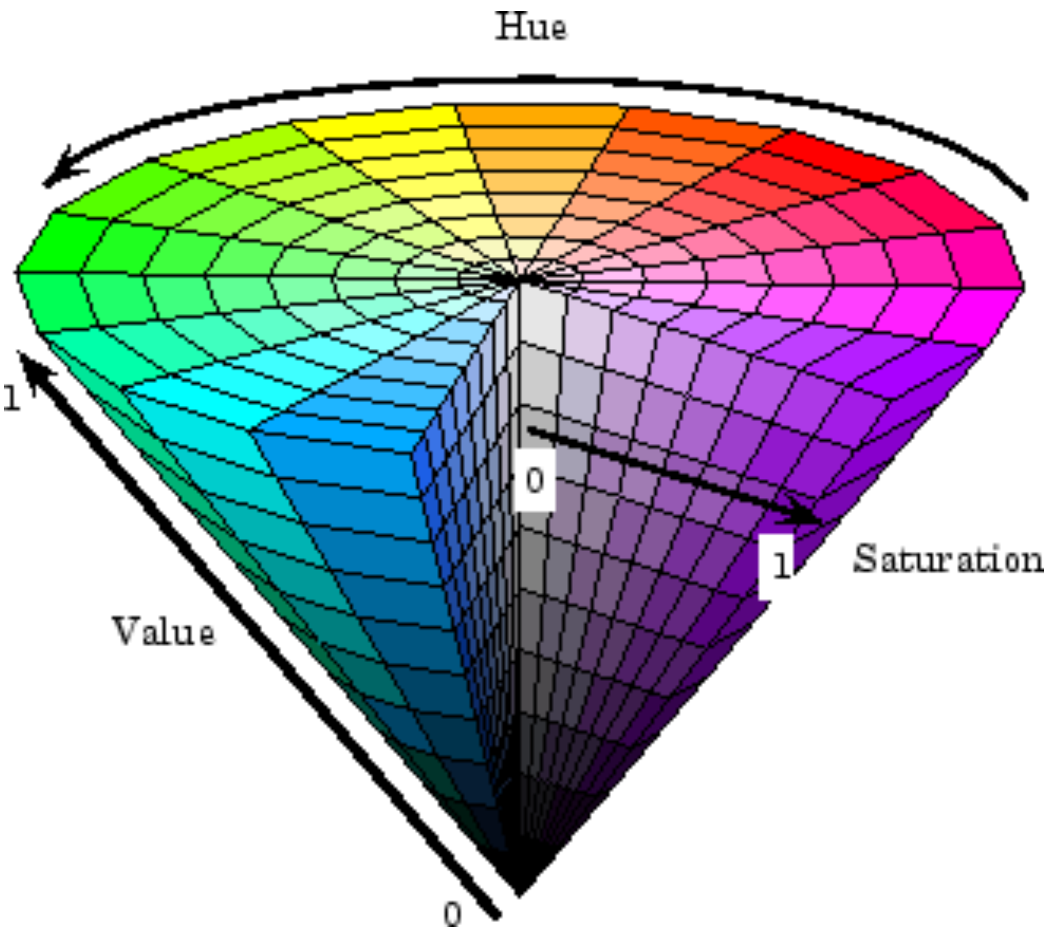
(R=0,G=0)

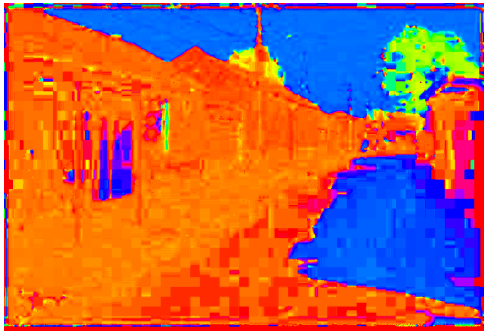
Some drawbacks

- Strongly correlated channels
- Not perceptually meaningful.

Image from: http://en.wikipedia.org/wiki/File:RGB_color_solid_cube.

HSV space





H

(S=1,V=1)

S

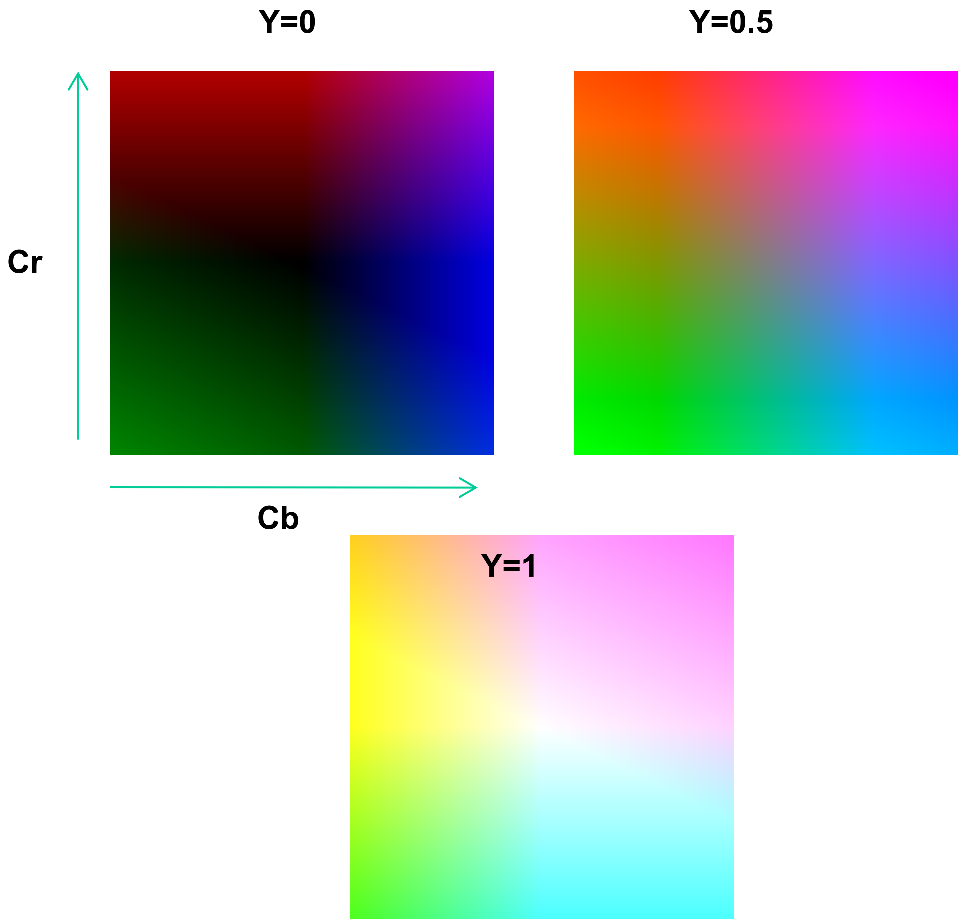
(H=1,V=1)

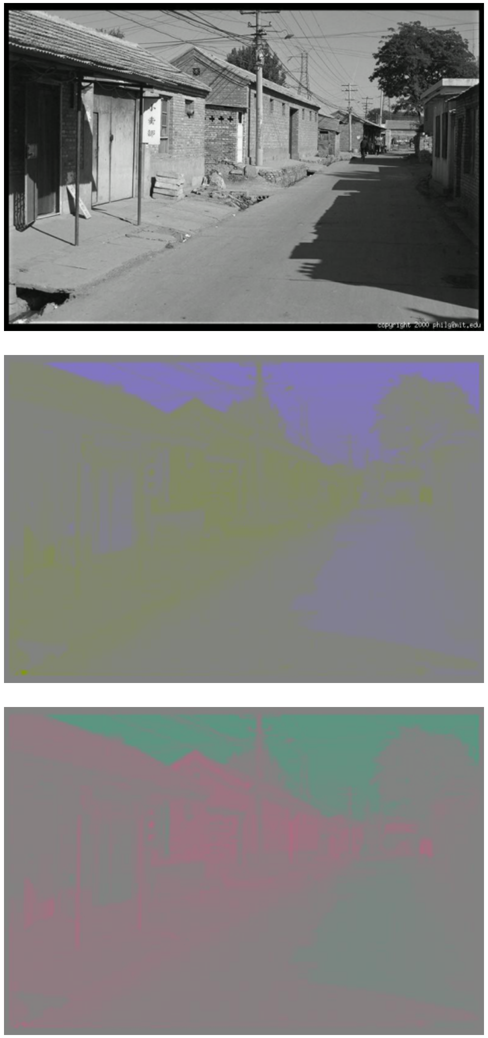
V

(H=1,S=0)

Intuitive color space

YCbCr (YUV) space





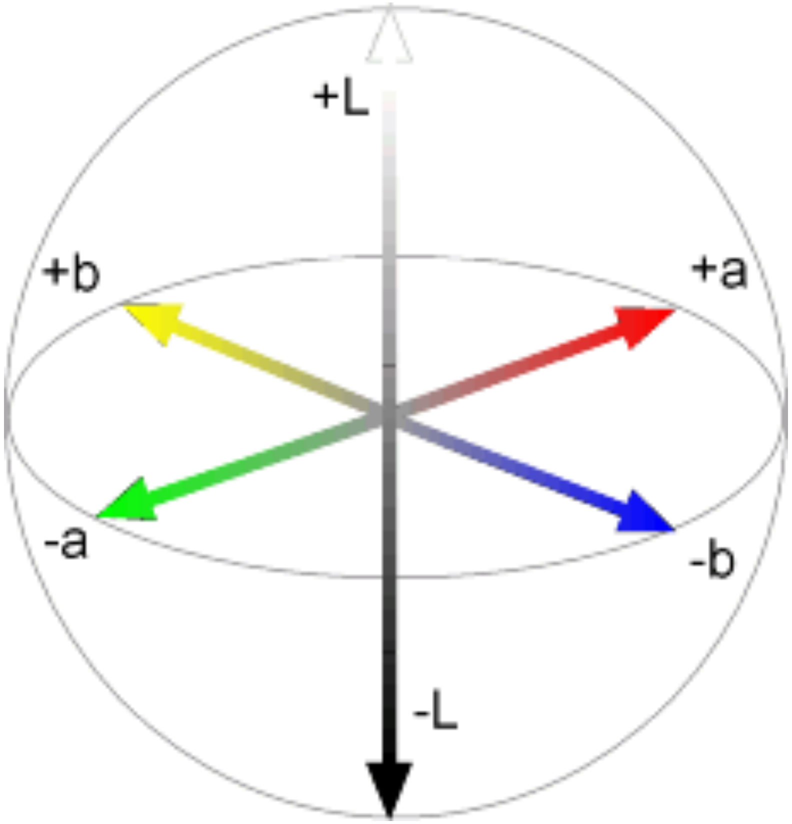
Y
(Cb=0.5,Cr=0.5)

Cb
(Y=0.5,Cr=0.5)

Cr
(Y=0.5,Cb=05)

Fast to compute, good for compression, used by TV

CIE-L*a*b* space



L

(a=0,b=0)

a

(L=65,b=0)

b

(L=65,a=0)

"Perceptually uniform" color space

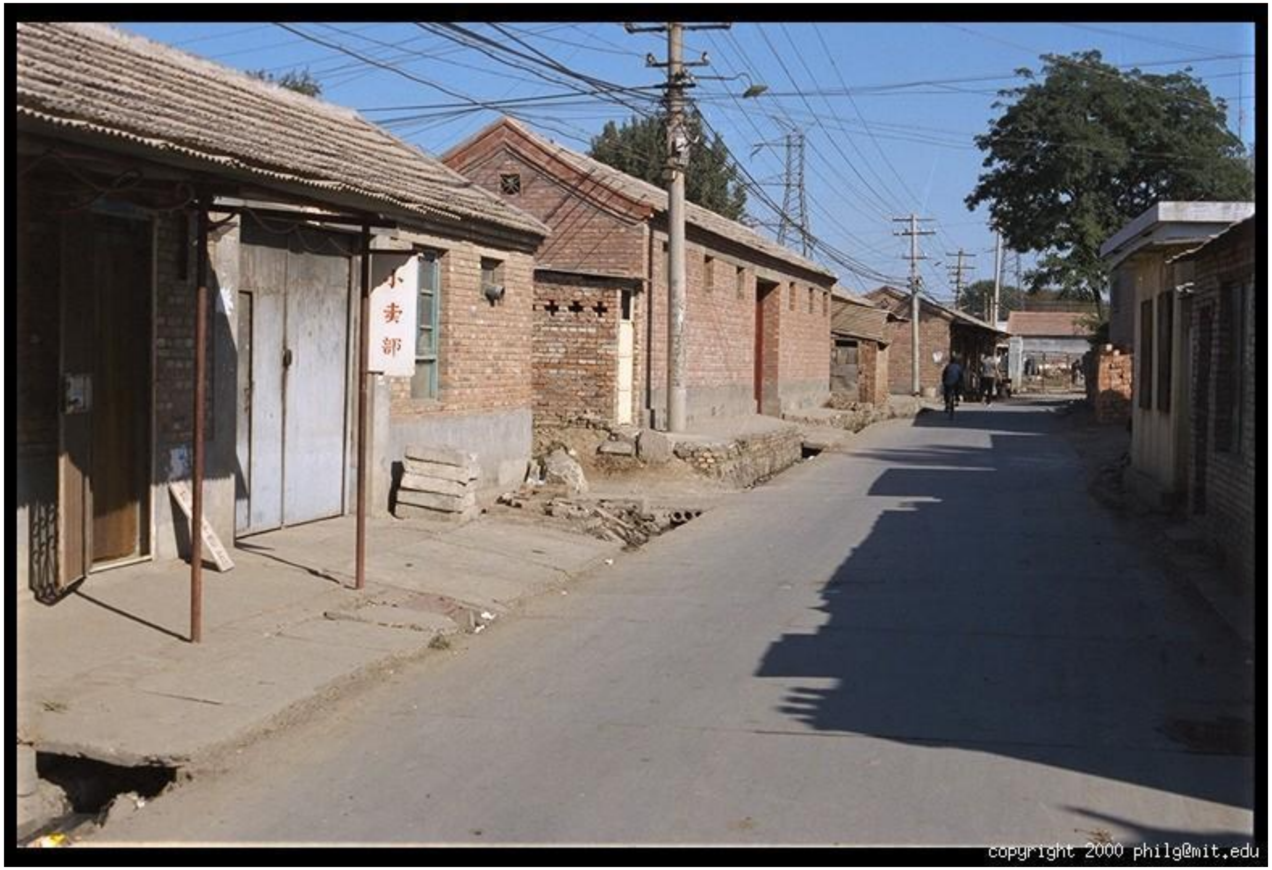
https://knowledge.ulprospector.com/10780/pc-the-cielab-lab-system-the-method-to-quantify-colors-of-coatings/

Conversions between different colour spaces

$$
\begin{aligned}
H &= \arccos \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B))}} \\
S &= 1-3\frac{\min(R,G,B)}{R+G+B} \\
V &= \frac{1}{3}(R+G+B)
\end{aligned}
$$

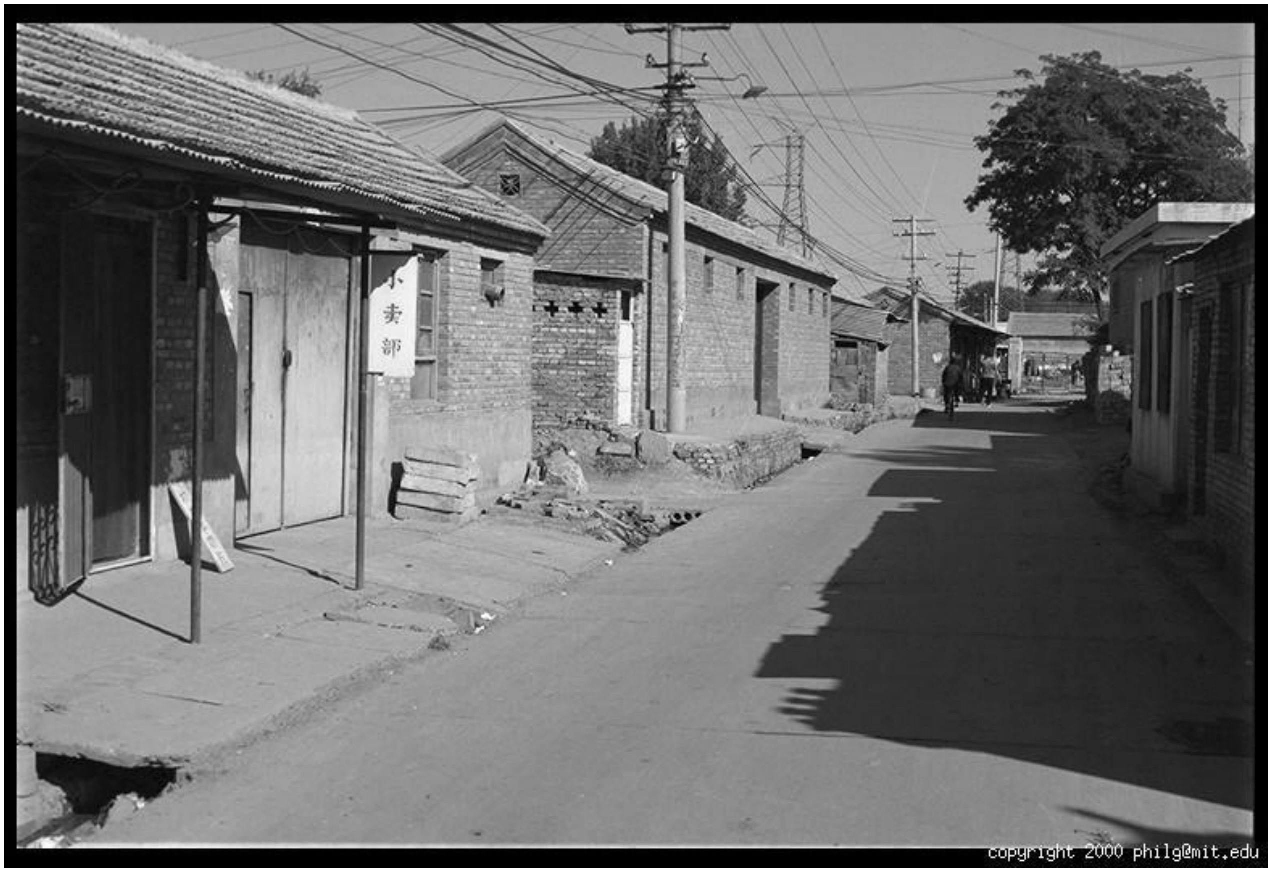
$$
\begin{aligned}
Y &= 0.299R+0.587G+0.114B \\
C_r &= R-Y \\
C_b &= B-Y
\end{aligned}
$$

Most semantic information is however contained in the intensity band



Original image

Most information is contained in intensity channel



Summary

- Geometric image formation describes where the 3D objects are projected in the image. (location)

- Photometric image formation describes the appearance of the objects. (intensity, color, appearance.)

Readings

Computer Vision: Algorithms and Applications

2nd Edition, Chapter-2 (2.1.4- perspective, camera intrinsics, camera matrix, 2.3.2 color)

Digital image processing, Chapter 2.4.2, 2.4.3

Take home message:

- Have a general idea of geometric image formation and photometric image formation process.
- Understand the color image representation.