Human Vision and Computer Vision

CSCE604133 Computer Vision

Fakultas Ilmu Komputer

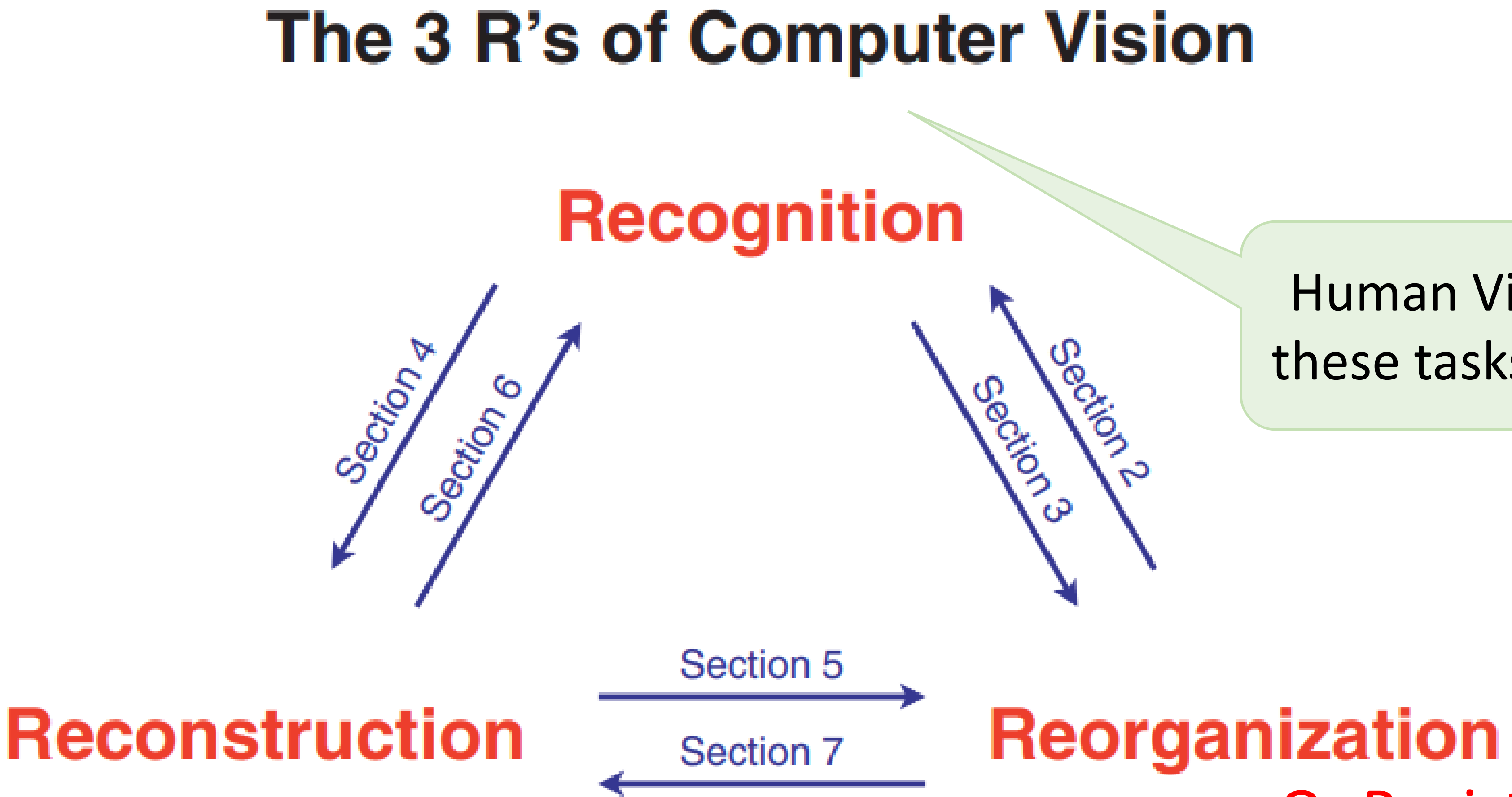
Universitas Indonesia

Acknowledgements

- These slides are created with reference to:
  - Computer Vision: Algorithms and Applications, 2nd ed., Richard Szeliski https://szeliski.org/Book/
  - Digital Image Processing, Gonzales and Woods, 3rd ed, 2008.
  - Course slides for CSCE604133 Image Processing – Faculty of Computer Science, Universitas Indonesia
  - Introduction to Computer Vision, Cornell Tech https://www.cs.cornell.edu/courses/cs5670/2024sp/lectures/lectures.html
  - Computer Vision, University of Washington https://courses.cs.washington.edu/courses/cse576/08sp/

The 3 R’s of Computer Vision

Malik, Jitendra, et al. "The three R’s of computer vision: Recognition, reconstruction and reorganization." *Pattern Recognition Letters* 72 (2016): 4-14.



Human Vision can do these tasks effortlessly

Or Registration

3D Scene vs 2D Image

Forward Process

- Models the physical process from 3D scenes (movement, light) that is projected onto a 2D image.

Inverse Process

- Taking the 2D image, and extracting the information, 3D scene, and other properties.

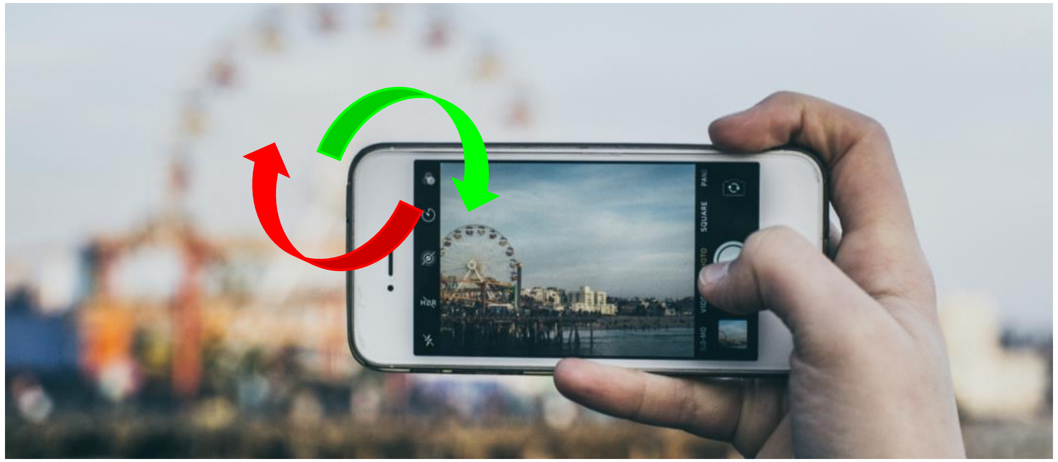
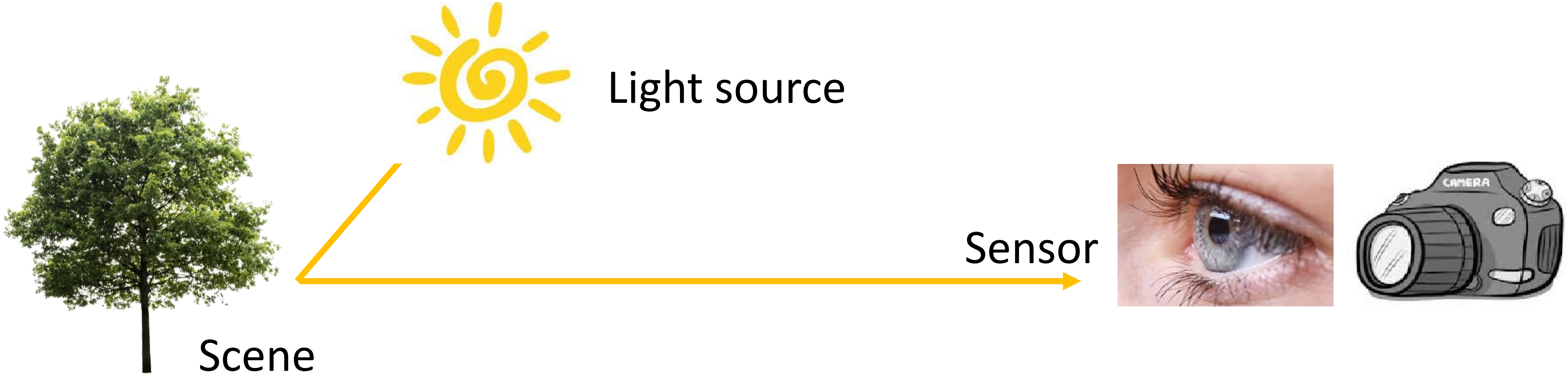


Image courtesy of Alina Fiene, Unsplash.

Image Acquisition

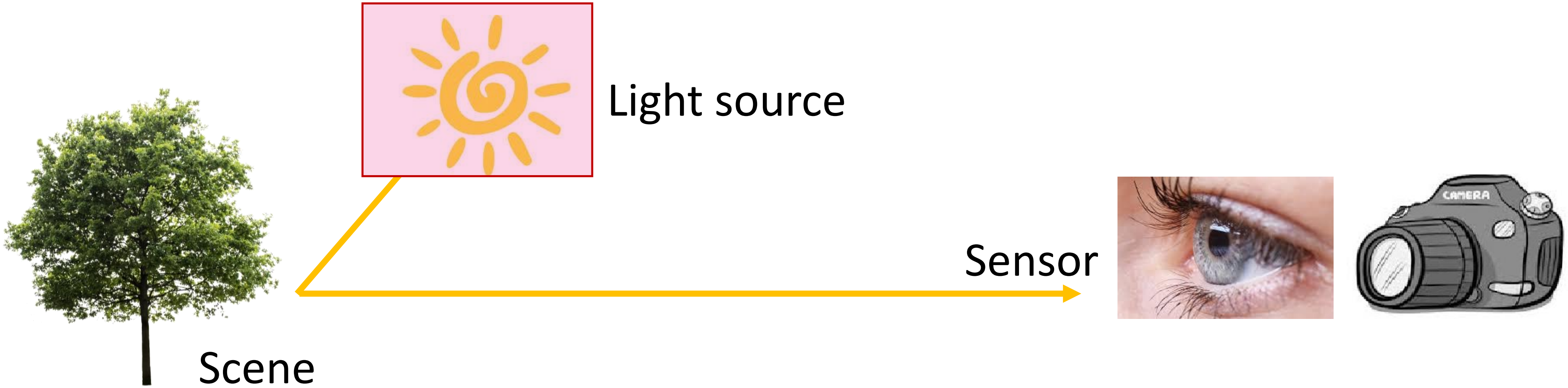
Image Acquisition

- Image acquisition requires 3 components:

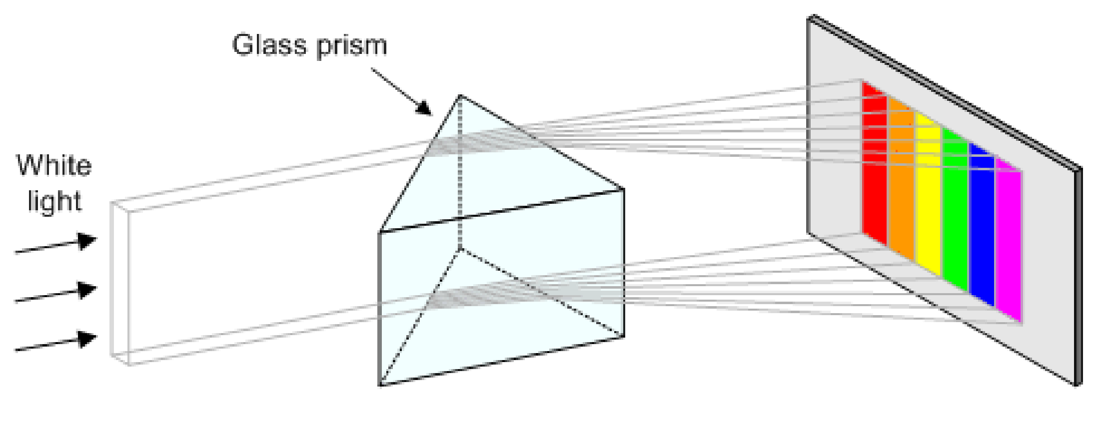


The Light Source

- Let’s look at the light



Newton’s Experiment

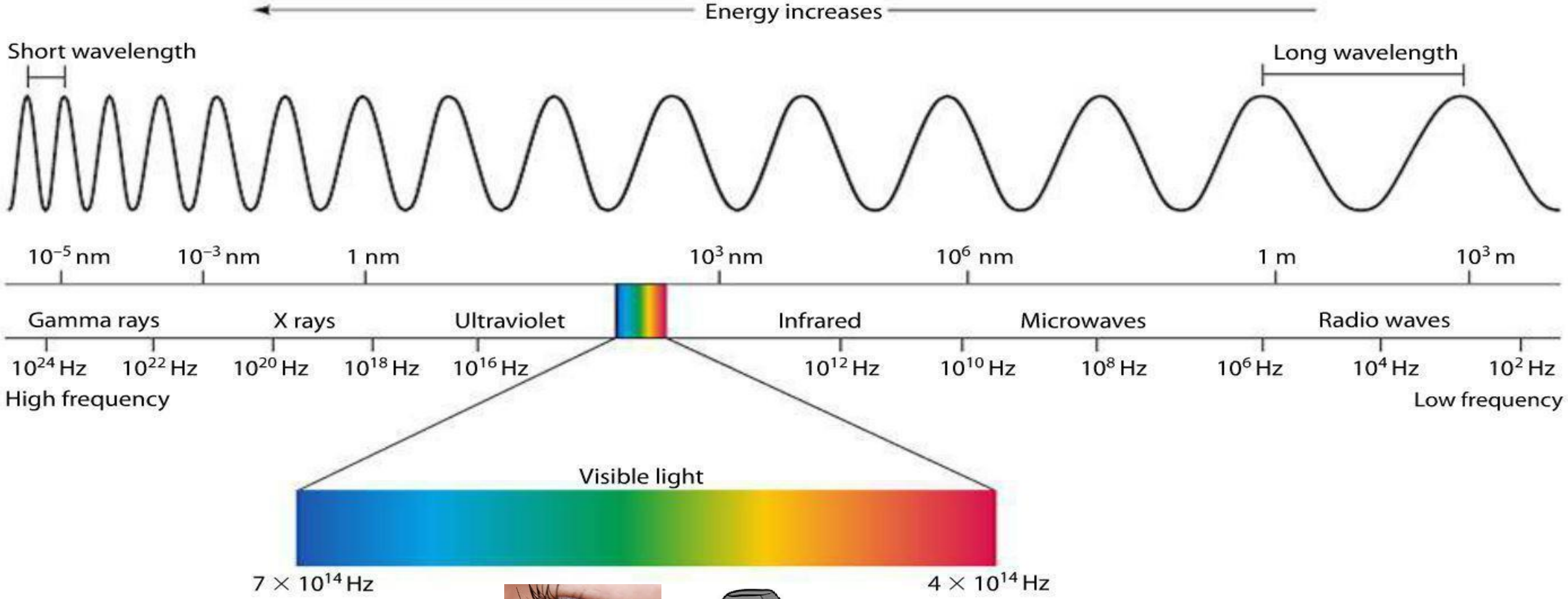


Newton’s experiment for splitting white light into a spectrum

- **Conclusion 1**: the white light was composed of a mixture of all colours in the spectrum
- **Conclusion 2**: the spectral colours were in fact the basic components (monochromatic lights) of the white light
- **Conclusion 3**: all the colours in the spectrum can be reunited to form the original white light again (by focusing the components back through a reversed prism).

Light as an Electromagnetic Wave

- Light is a wave: with wavelength (λ / lambda) and frequency f
- The electromagnetic spectrum:

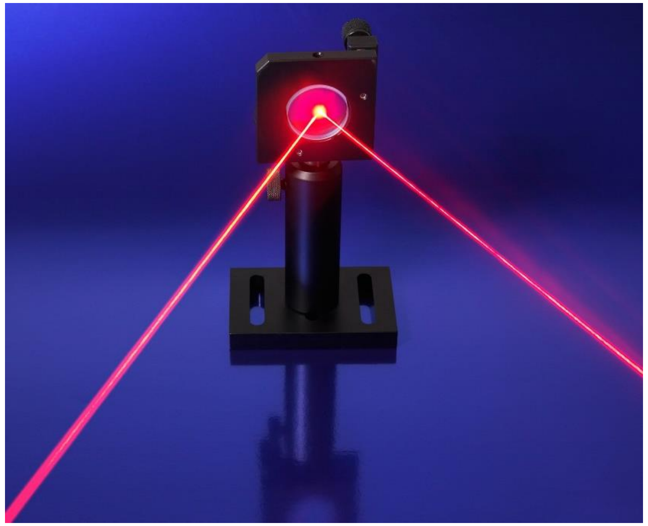


400 nm

700 nm

Light / Light sources / Illuminants

- Monochromatic light
  - At one wavelength only, from coherent light sources
  - Lasers

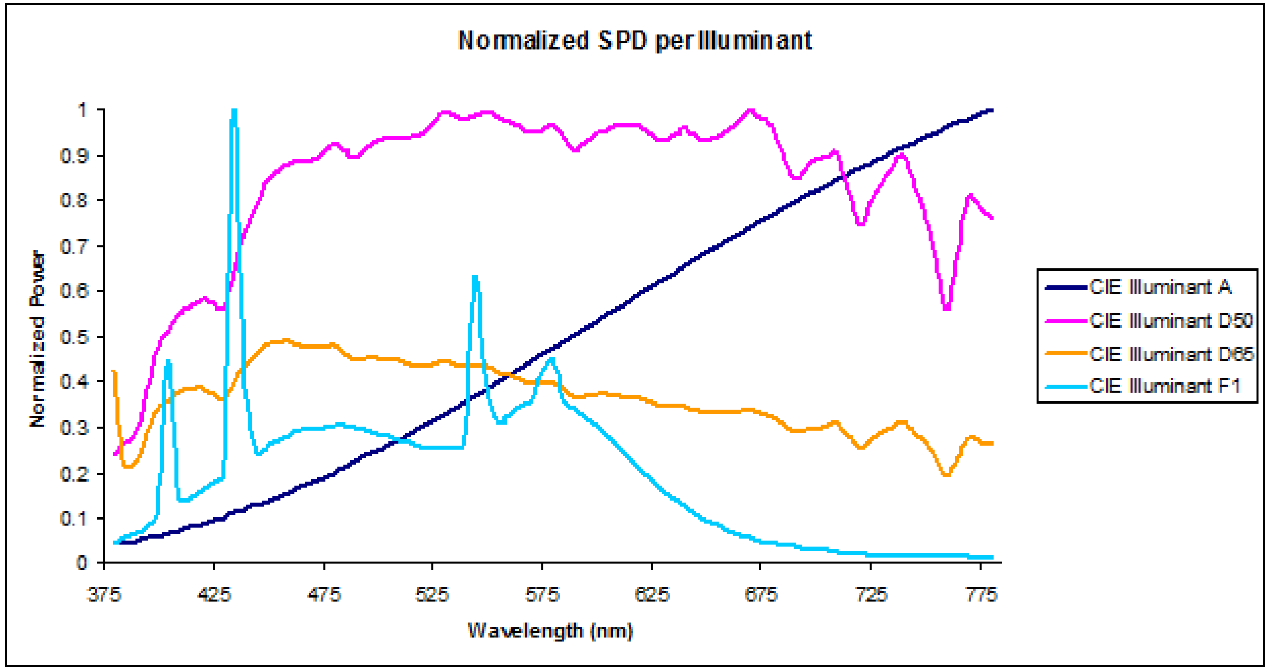


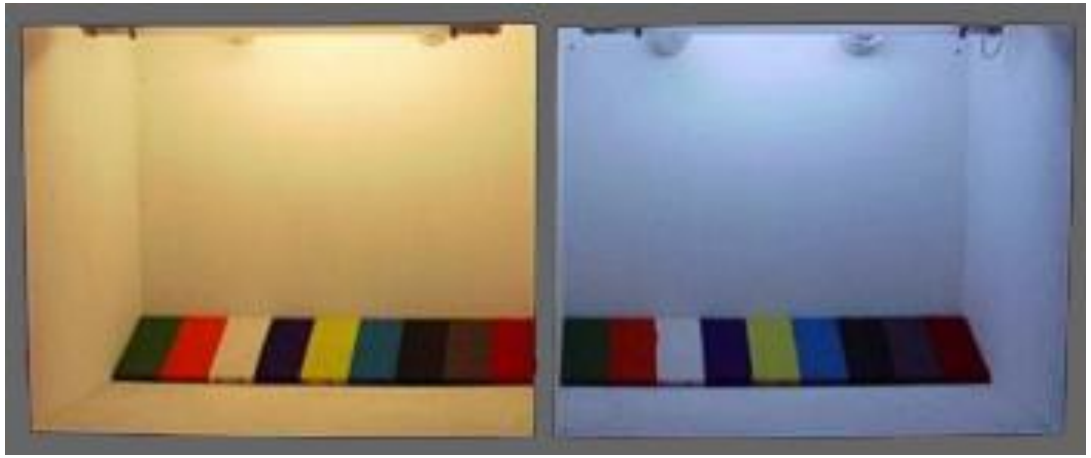
- Chromatic light (visible light)
  - The visible spectrum (λ 400-700 nm)
  - Sunlight, most light bulbs

Why are the illuminants important?

Spectral Power Distribution of Illuminants

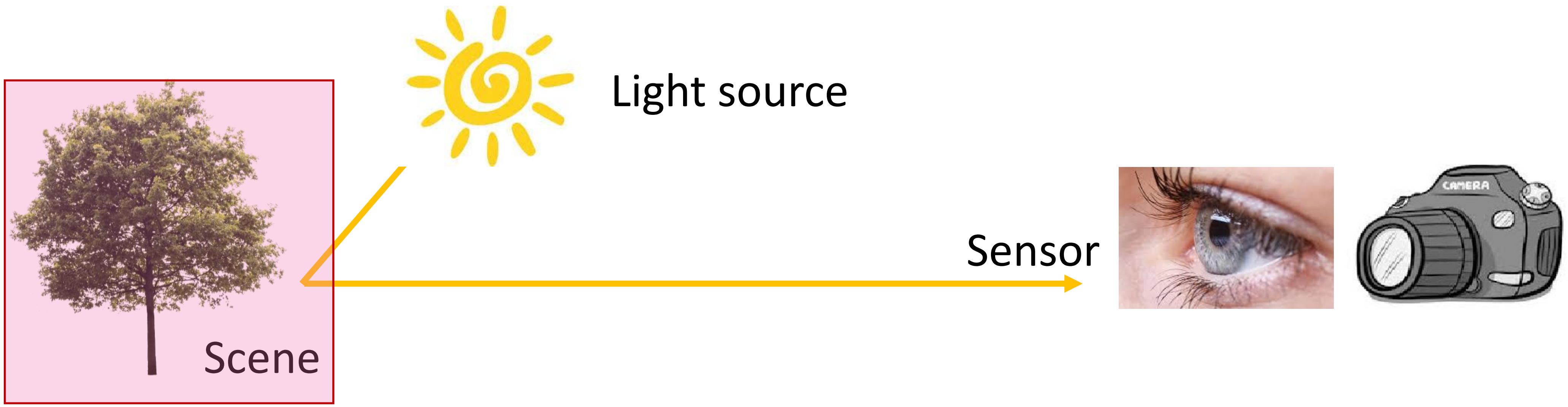
- Different lights are **spectrally** different and appear differently



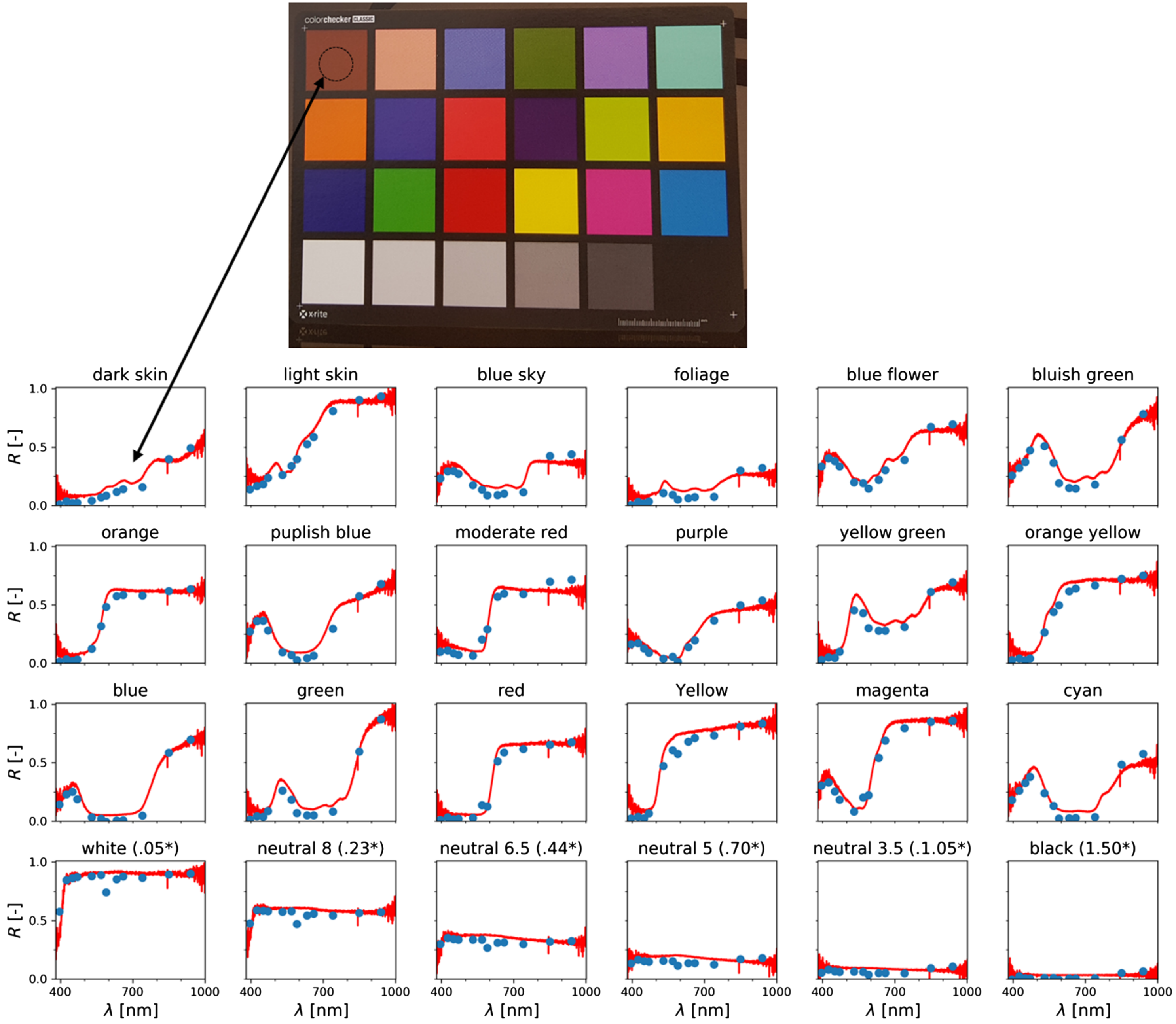


The Scene

- Let's look at the scene next

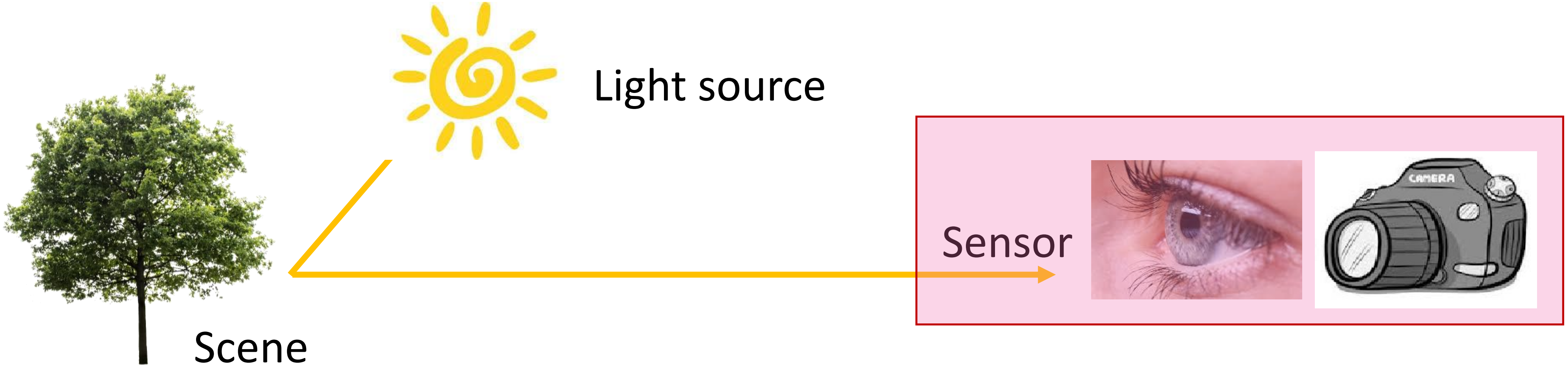


Spectral Reflectance of Objects



The Sensor

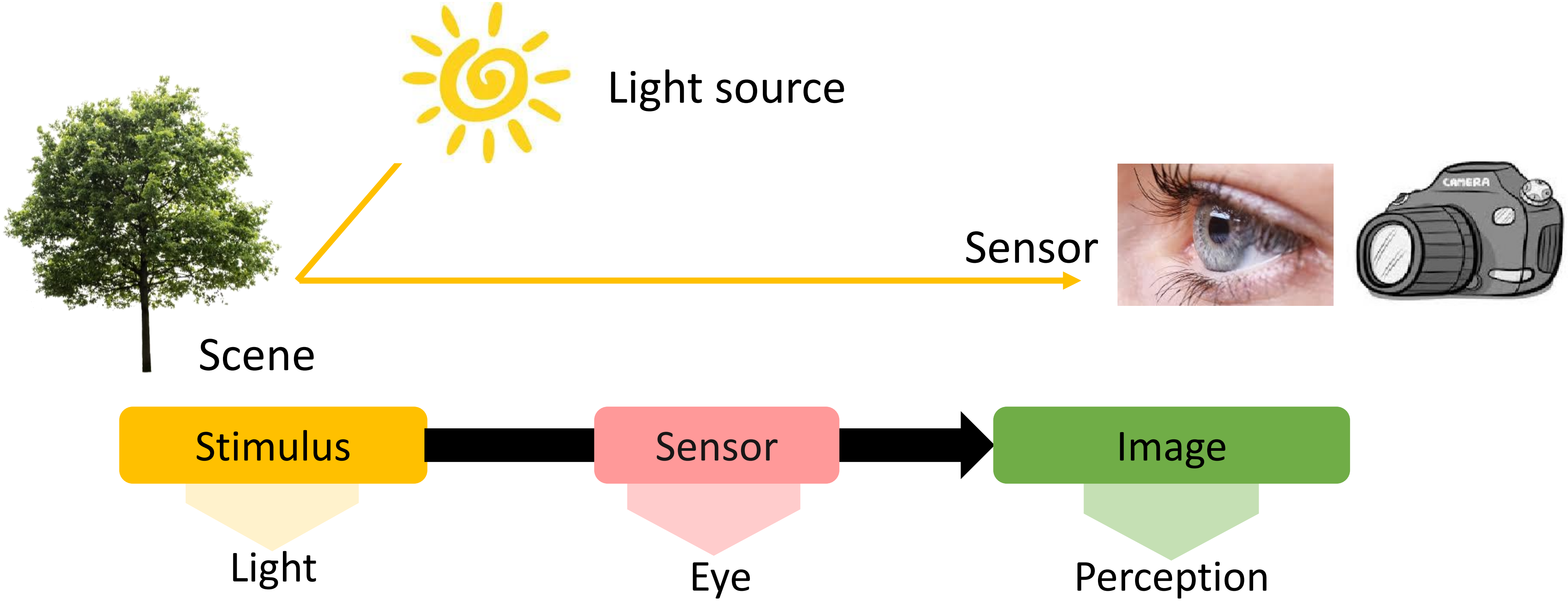
- Let's look at the sensor next



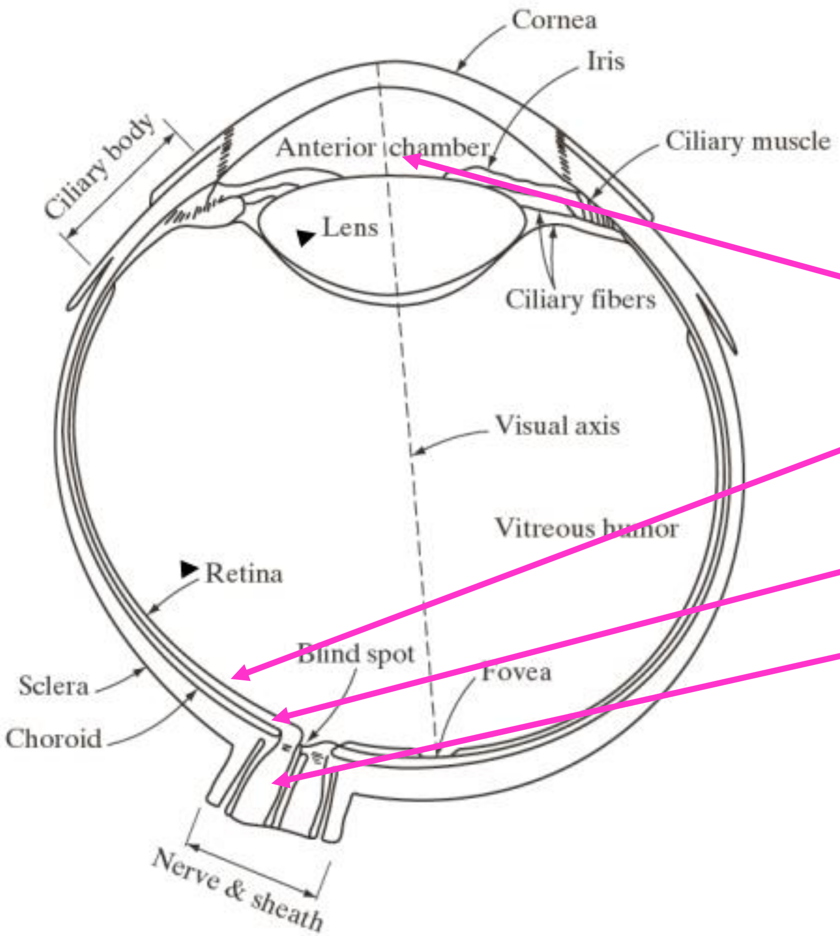
Human Vision

The Human Eye as a Sensor

- The image captured is not a physical image, but a **perception**



The Human Eye



- Light enters the eye
- An image is captured on the retina
- The image is translated into biological signals
- Signals are transmitted to the brain
- Processing in the brain
- Visual perception

Image Formation on the Retina

- The retina has optical sensors to perceive
- Cones (6-7 million sensors)
  - Cones are sensitive to color, resulting in **Photopic Vision**
- Rods (75-150 million sensors)
  - Rods create a more general overall image, resulting in **Scotopic Vision (Color Vision)**

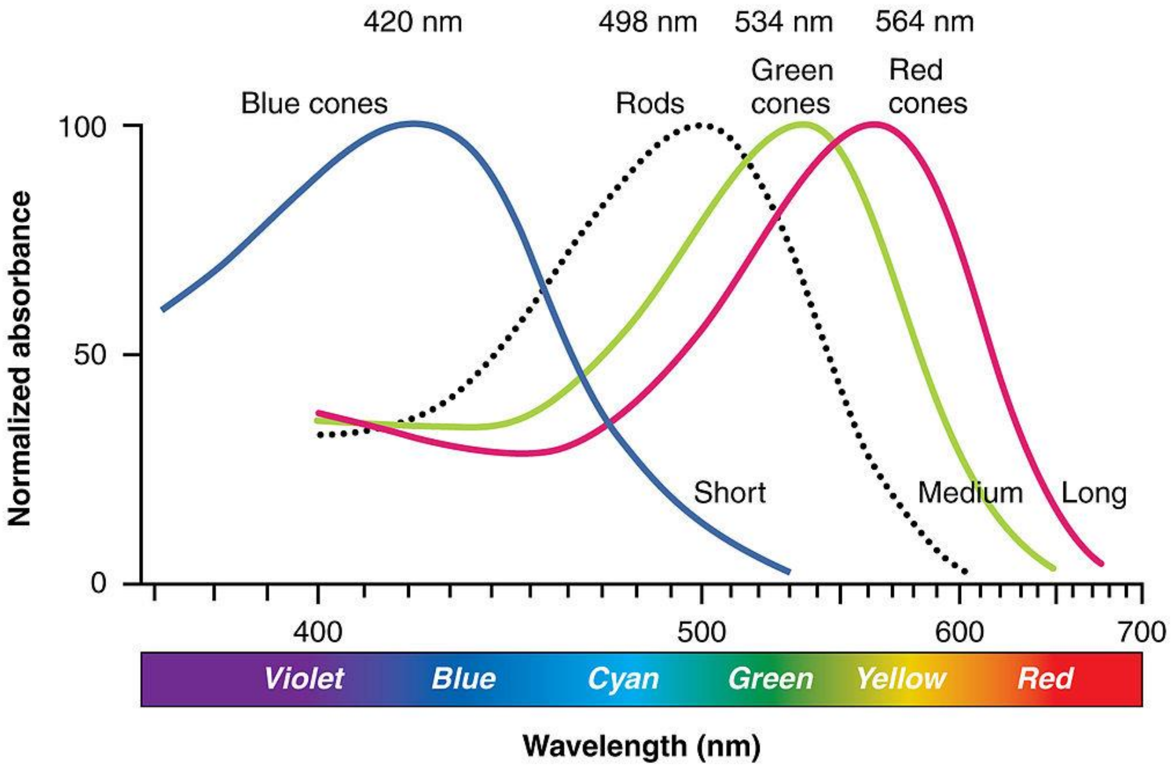
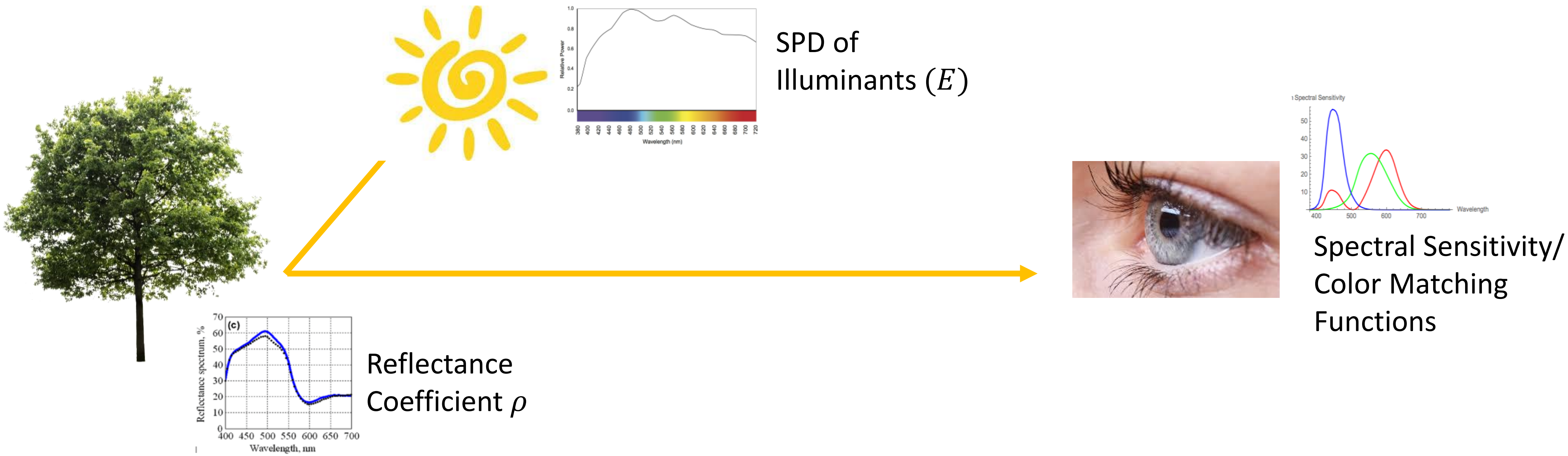


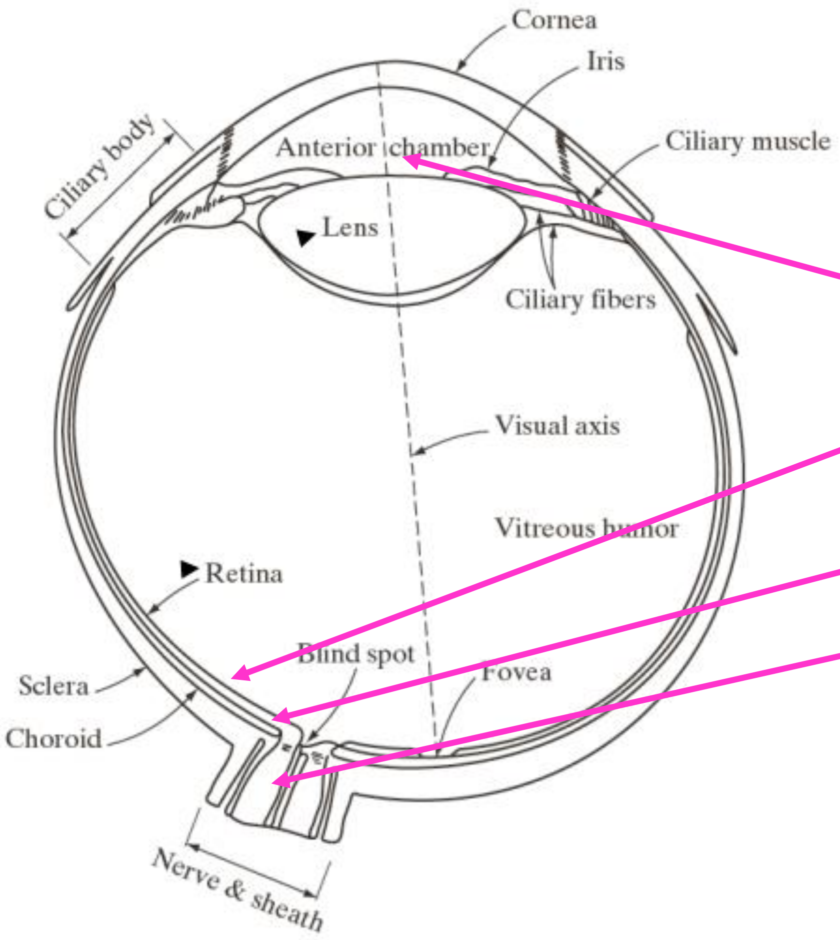
Image Formation on the Eye

- The **Spectral Power Distribution** and **Reflectance Coefficient** ρ change the light that arrives at the sensor



- The human eye has **spectral sensitivity**

Human Visual Perception



The Human Eye is Weird

Image capture by the human eye is not that simple!

- Light enters the eye
- An image is captured on the retina
- The image is translated into biological signals
- Signals are transmitted to the brain
- Processing in the brain
- Visual perception

gap yang gak bisa di cover

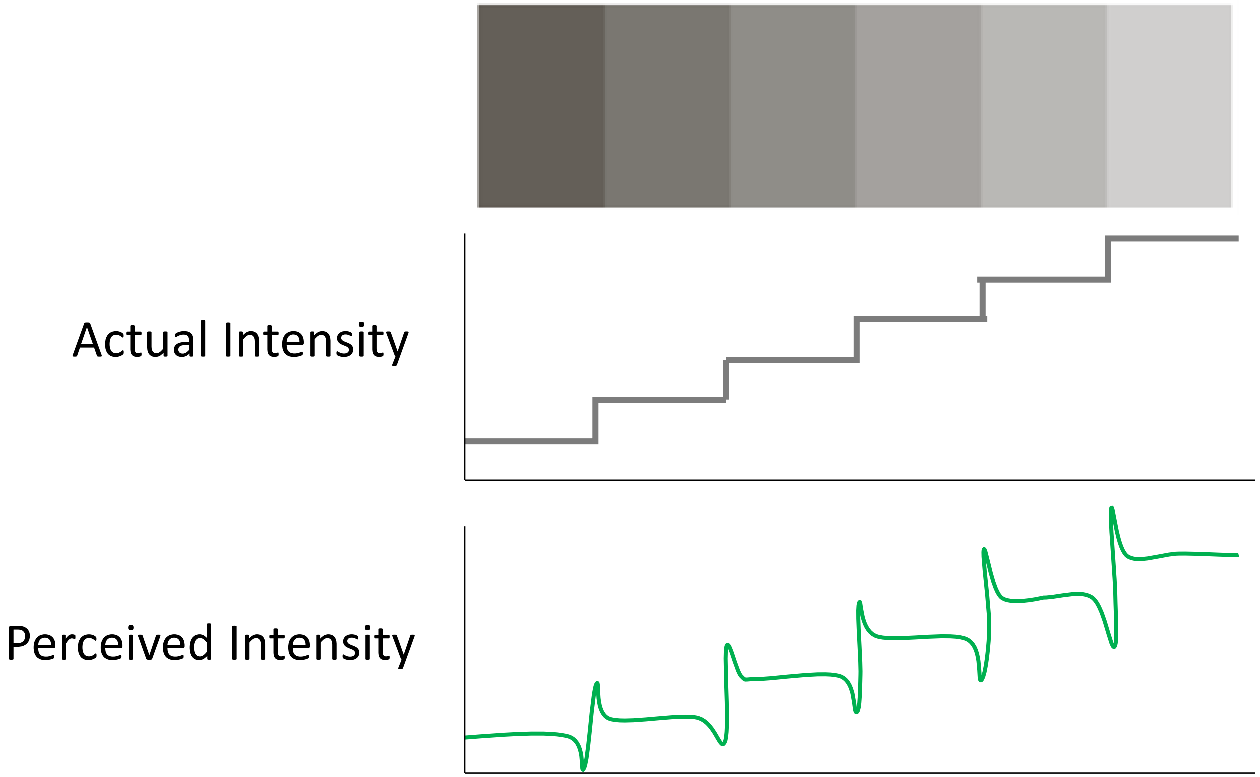
?

Visual Perception in The Brain

The image is a **perception** based on the process on the brain.

Brightness Adaptation

- Mach Bands: Perceived intensity is not a simple function of actual intensity



human perception

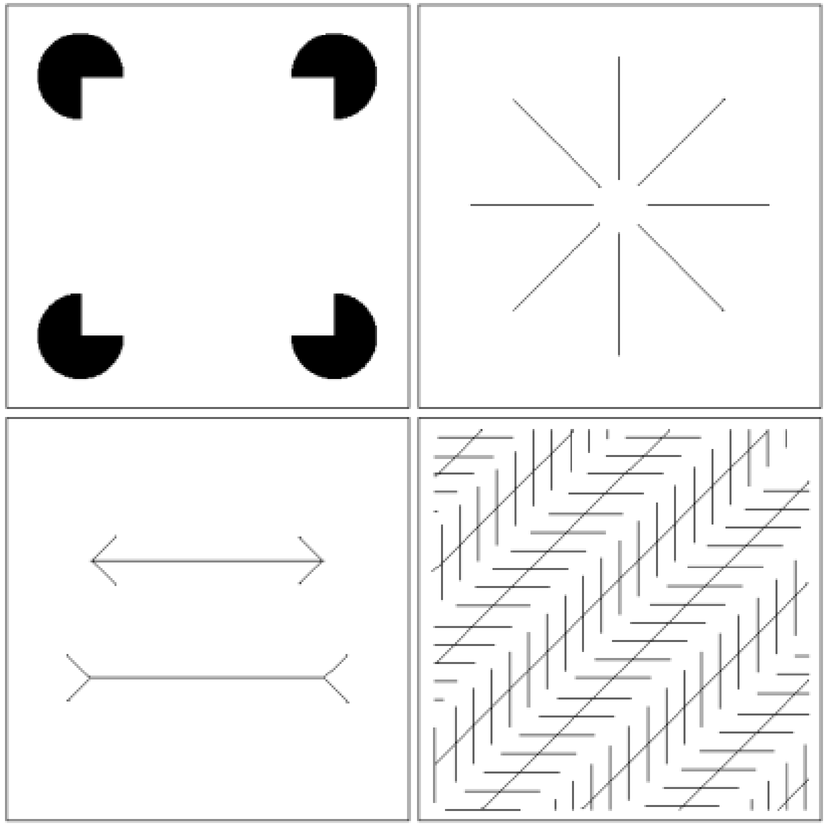
Brightness Adaptation

- Simultaneous Contrast: a region's perceived brightness does not depend simply on its intensity.

di komputer semua warna itu sama

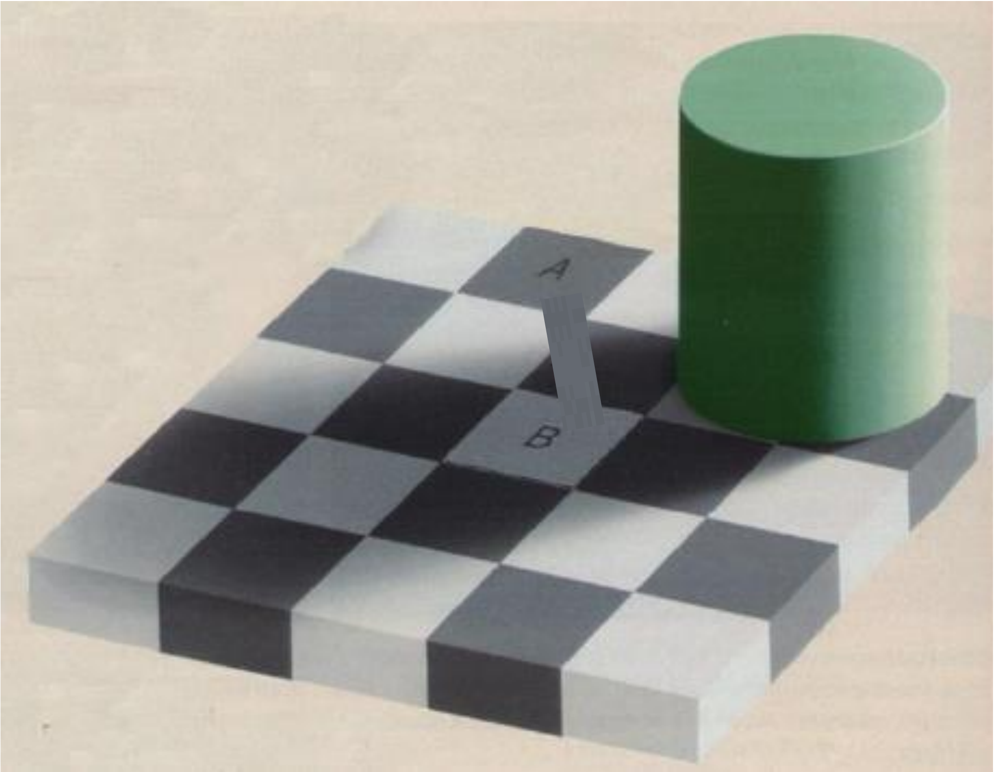
Other Curious Things on the Human Eye

- The eye fills in non-existing info or wrongly perceives geometrical properties of objects



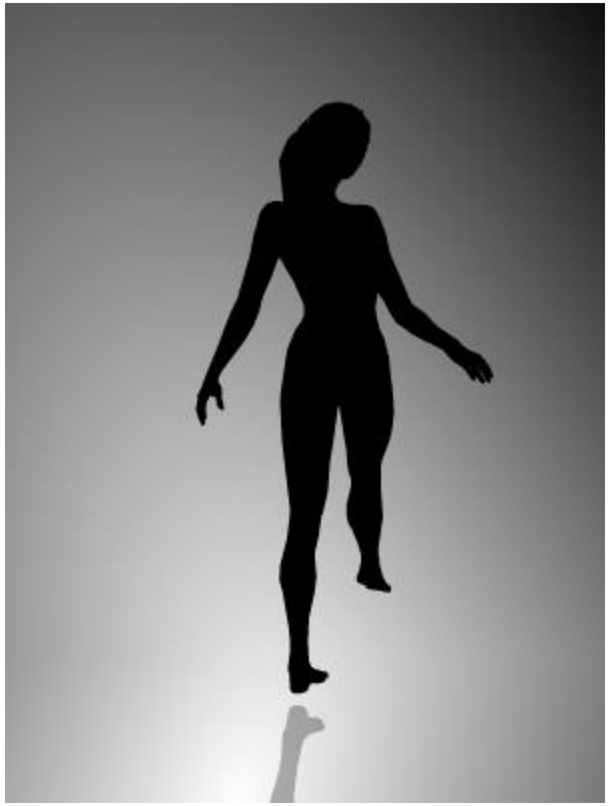
What about with Color?

- Is A and B the same color?



Color Constancy

What about with Motion?



The Human Eye

- Becomes the basis on which the digital camera is built on.
- We try to obtain an image similar to the perception we obtain in our eyes.
- What can we replicate in digital images?
- What information can we recover from digital images with computer vision?

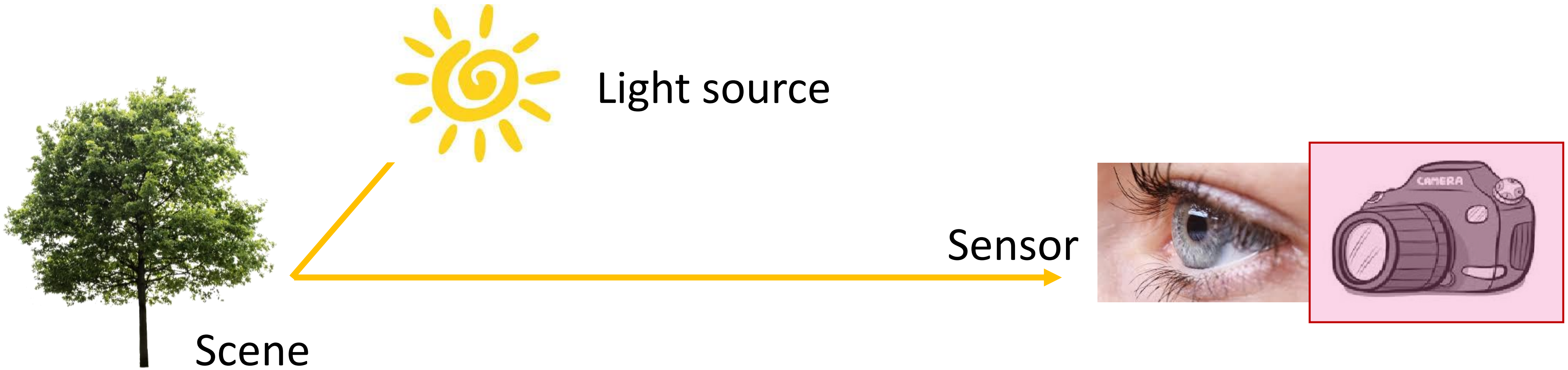
There will be some things that computer vision simply can not do.

for example the perecived or how color looks different towarsds human even though its the same.

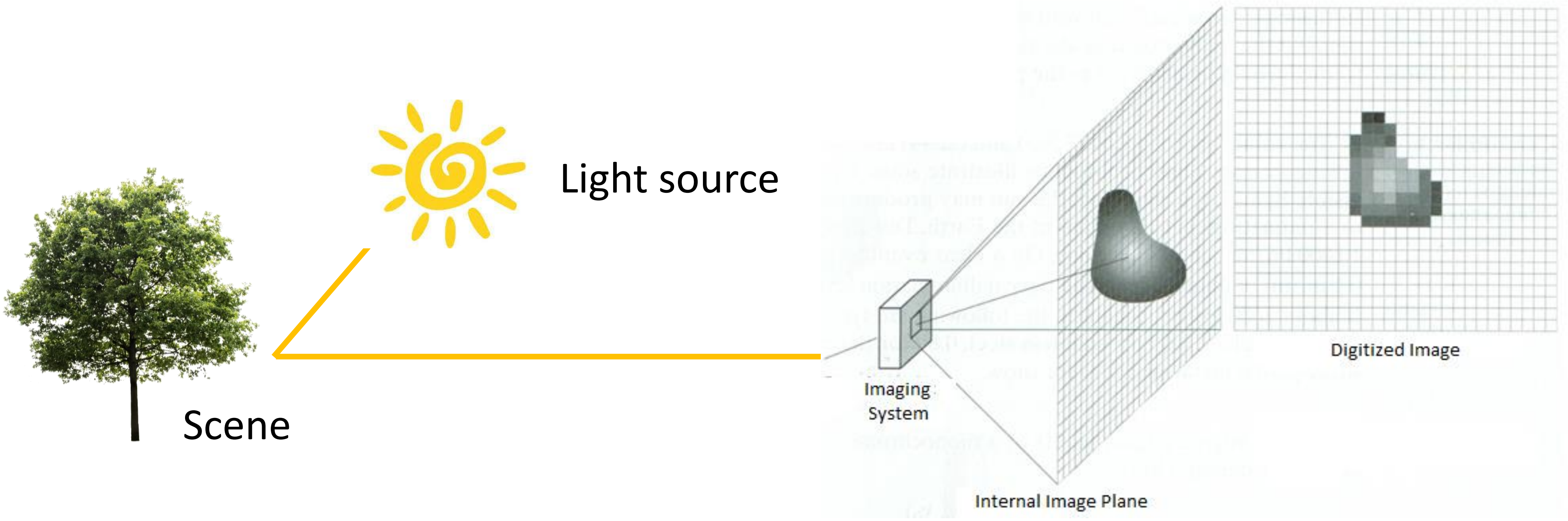
Digital Image Acquisition

Digital Cameras as Sensors

- The camera has a digital sensor



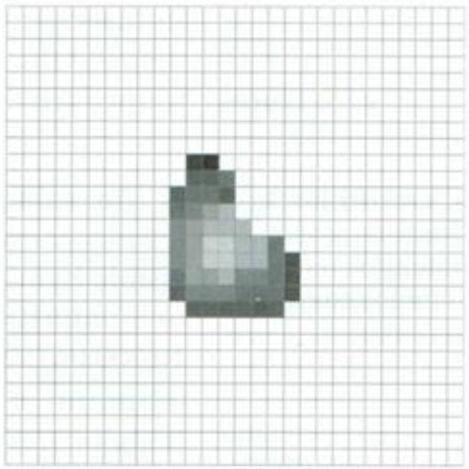
Digital Images



Digital Images

1000 x 1000

- A matrix where each matrix element the gray level intensity $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$
- Intensity function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$: x and y is the spatial coordinate on the matrix and the function value of $f(x, y)$ is the intensity level at that location
- Intensity function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$ is obtained through
  - Spatial discretization (sampling)
  - Intensity discretization (quantization);



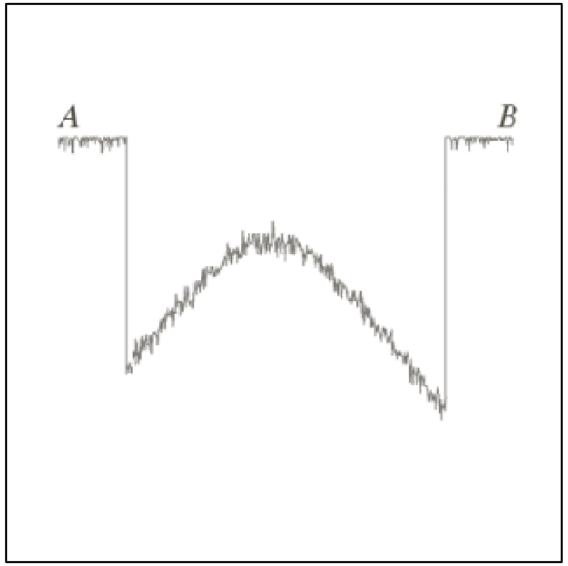
Digitized Image

angka yang dimasukkin tidak

Sampling and Quantization

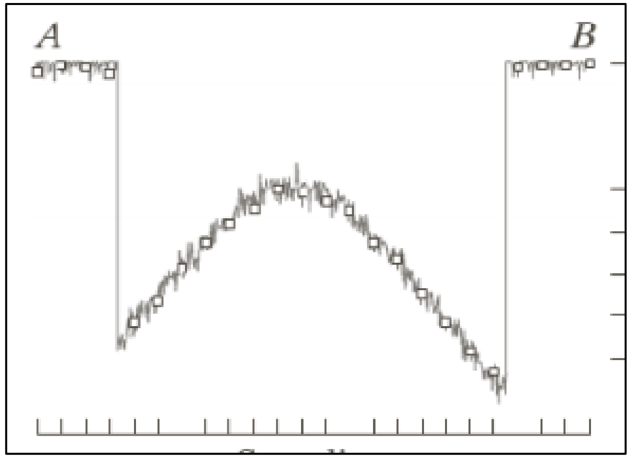
- We only have limited “spots” that can be used to store the intensities
- We only have limited “values” that can be used to represent the intensities

dari sinyal ini

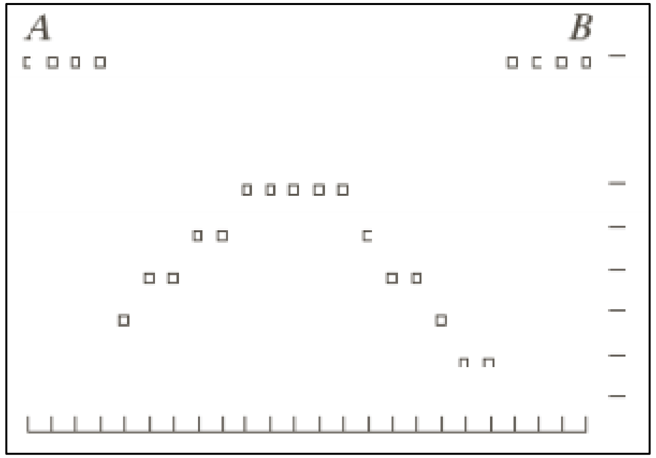


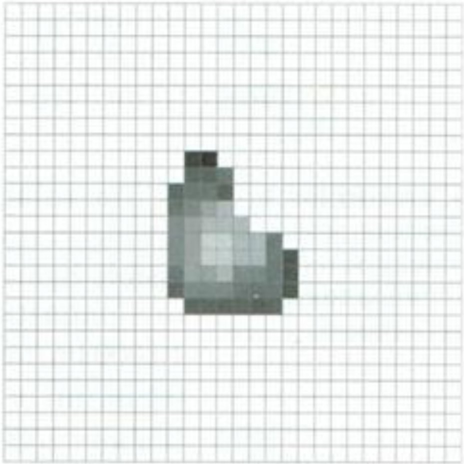
harus disampling karena terlalu banyak

Sampling



Quantization



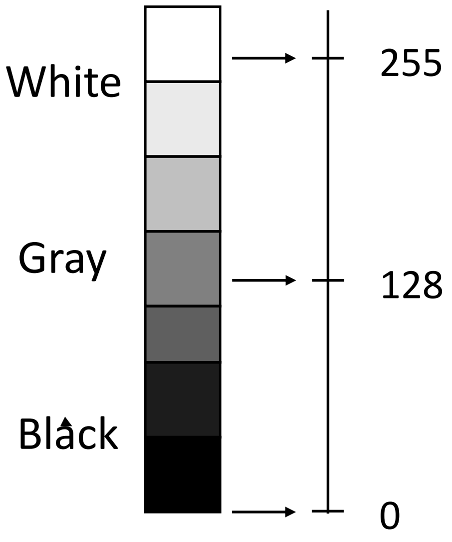


Digitized Image

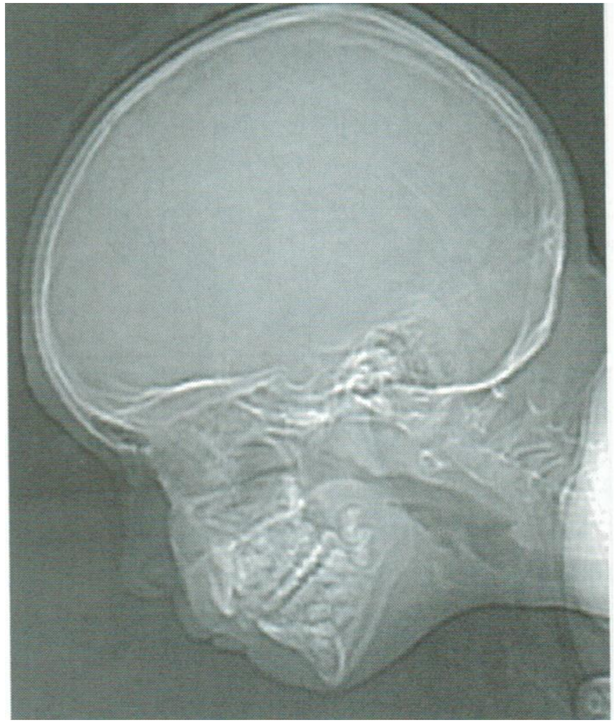
Image Intensity Resolution

citra dibuat 8 bit
bisa merepresentasikan 2^8 nilai = 256 yang berbeda.

- How smooth / rough is the division of the intensity level values represented in each pixel. Transforming the continuous signal to discrete intensities in a digital image matrix is quantization.



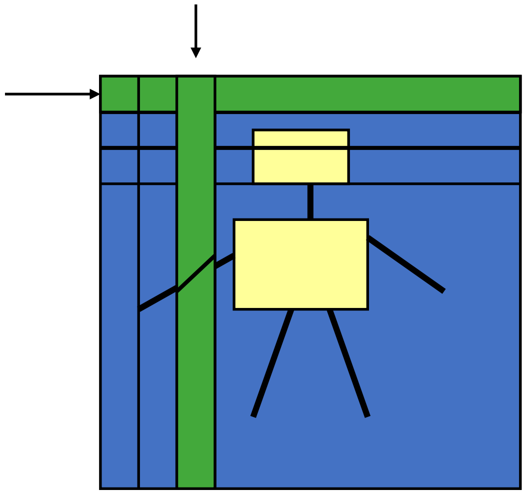
Example: 8-bit image



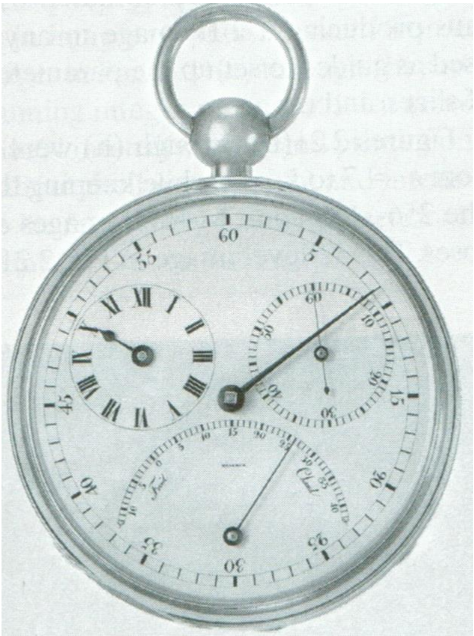
1 buah pixel cuman bisa di representasikan oleh 2^8 bit

Image Spatial Resolution

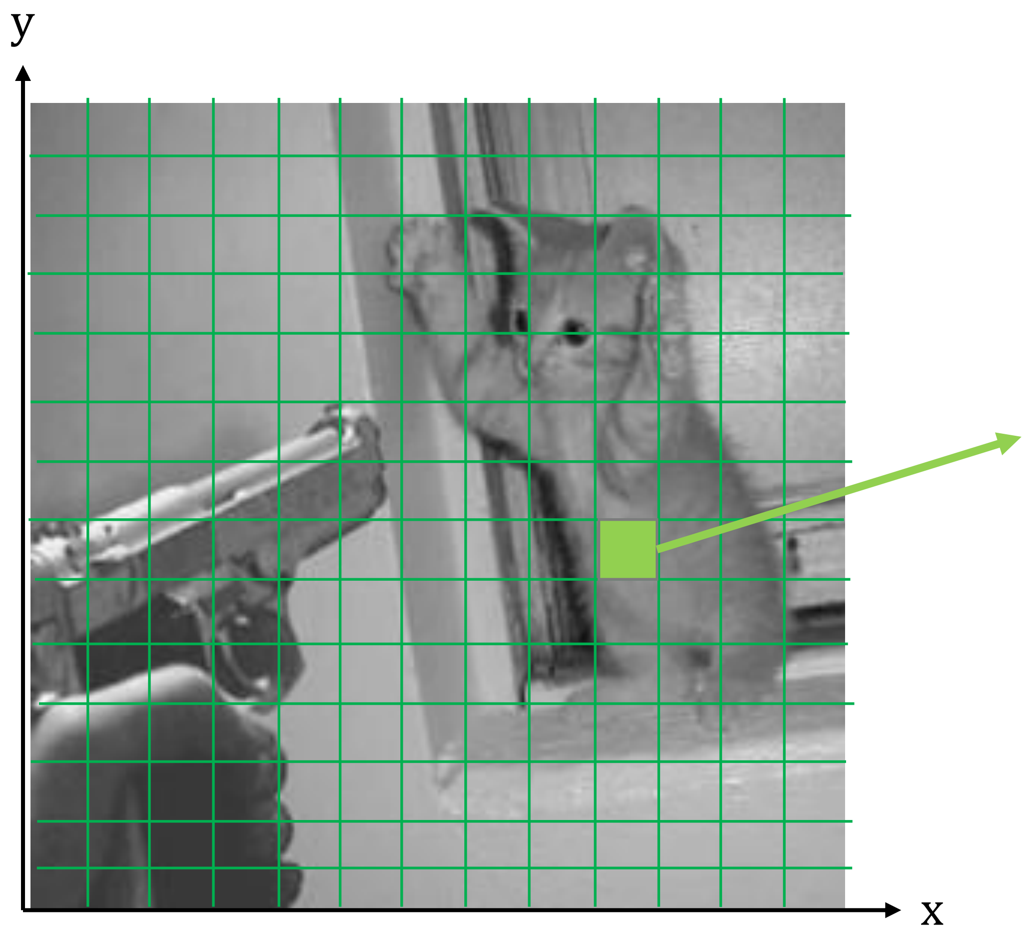
- How smooth / rough is the division of the grid (rows and columns) of pixels.
- Transforming the continuous signal to a limited number of values in a grid is digitization / sampling.



Picture

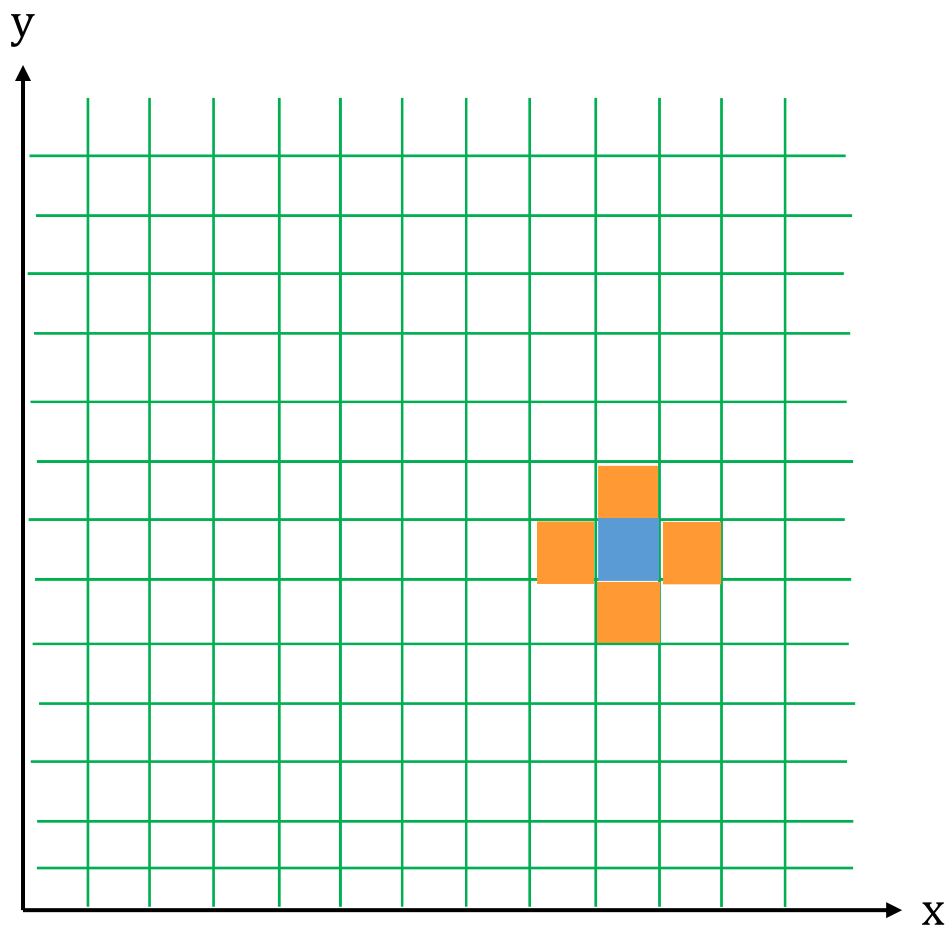


Digital Images



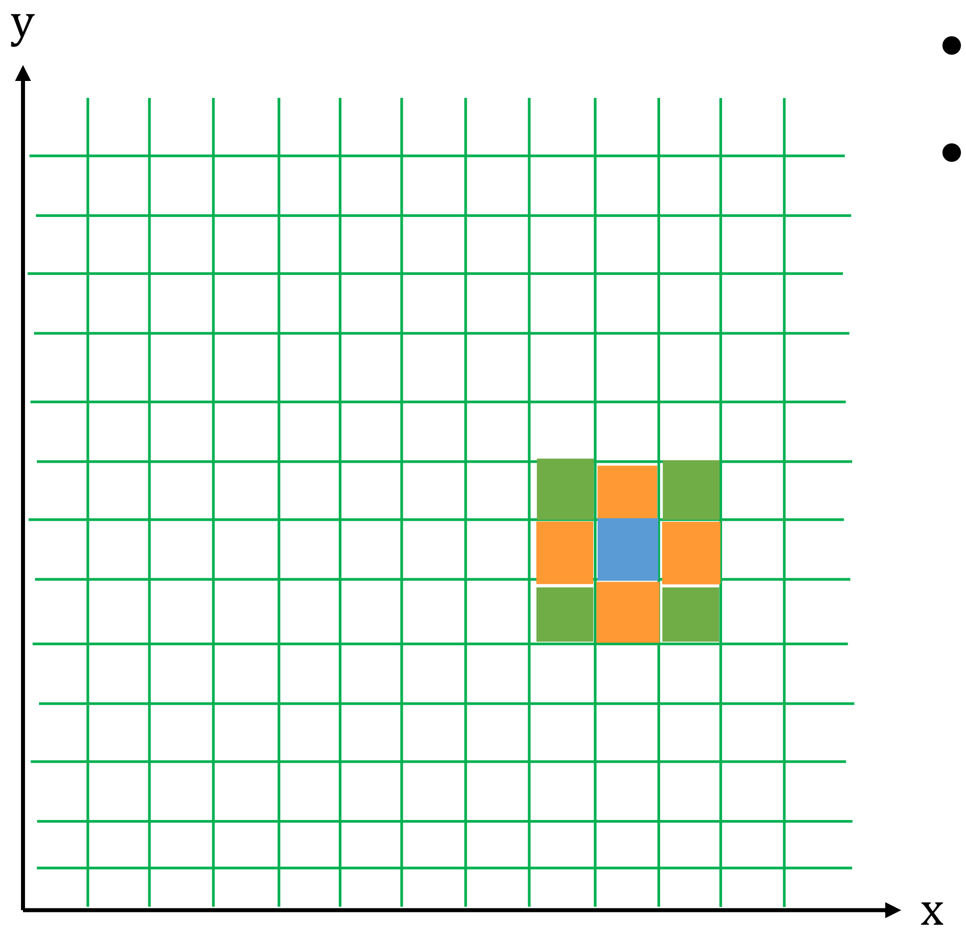
- Intensity function $f(x, y)$
- Discretization of..
  - Spatial resolution *(Sampling)*
  - Intensity resolution *(Quantization)*
- Picture elements (pixels)

Pixel Neighbors



- A pixel $p(x, y)$
- Has 4 horizontal and vertical neighbors
  - **4-neighbors $\boldsymbol{N_4(p)}$**
  - $(x+1, y), (x-1, y),$
  - $(x, y+1), (x, y-1)$

Pixel Neighbors (2)



- A pixel $p(x, y)$
- Has 4 *more* diagonal neighbors
  - **8-neighbors $\boldsymbol{N_8(p)}$**
  - $(x+1, y+1), (x+1, y-1)$
  - $(x-1, y+1), (x-1, y-1)$

cahaya yang dateng bisa dari segalaa arah, makanya kita pakai 8 pixel neighbor

relevan apabila kita mau ekstrak fitur-fitur/bentuk-bentuk dalam sebuah citra

Adjacency and Connectivity

ketika mau ekstrak objek-objek

- kita bisa pakai rule adjacency,
ketika mau batasin objek yang diambil dari gambar, berapa limit yang perlu di set untuk stop ambil gambarnya.

- 4-adjacency: p and q that fulfill V are 4-adjacent if $q \in N_4(p)$

$$\begin{array}{ccc} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}$$

- 8-adjacency: p and q that fulfill V are 8-adjacent if $q \in N_8(p)$

$$\begin{array}{ccc} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}$$

- m-adjacency: p and q that fulfill V are m-adjacent if
  - $q \in N_4(p)$, or
  - $q \in N_D(p)$ and $N_4(p) \cap N_4(q)$ has no pixels that fulfil V

$$\begin{array}{ccc} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}$$

Connected Components

- For $V = \{1\}$, we obtain the connected components the image.

4-neighbors

0	0	0	0	0	0
0	0	1	1	0	0
0	1	1	0	0	0
0	1	1	0	1	1
0	0	0	1	1	1
0	0	0	0	0	0

8-neighbors

0	0	0	0	0	0
0	0	1	1	0	0
0	1	1	0	0	0
0	1	1	0	1	1
0	0	0	1	1	1
0	0	0	0	0	0

- Within 1 connected component, all the pixels make up a *connected set.*

- R is a *region* if R is a connected set
- Regions are adjacent if their union becomes a connected set, otherwise they are disjoint.

$\boldsymbol{R_1}$

0	0	0	0	0	0
0	0	1	1	0	0
0	1	1	0	0	0
0	1	1	0	1	1
0	0	0	1	1	1
0	0	0	0	0	0

$\boldsymbol{R_2}$

ini pakai rule 4 neighbor, makanya dia gak diambil sebagai 1 area.

ini berguna untuk segmentation

***Are* R_1 and R_2 adjacent?

Mathematical Operators on Images

+ - * /

- Examples:
  - Substracting two temporal images can be used to detect a change in an area:
    - No change : difference is 0
    - Some change: difference is not 0

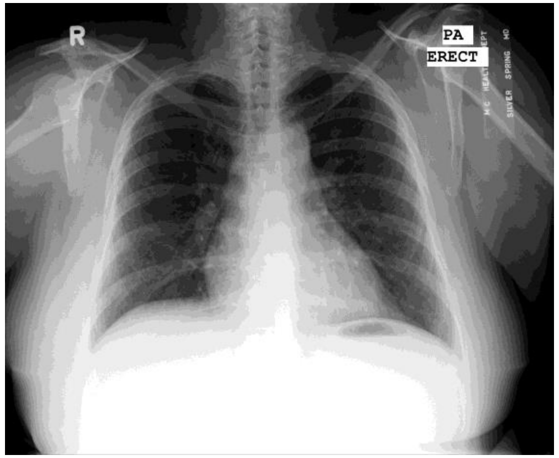
di minus aja untuk ngitung apabila terjadi kebakaaran

Logic Operators on Images

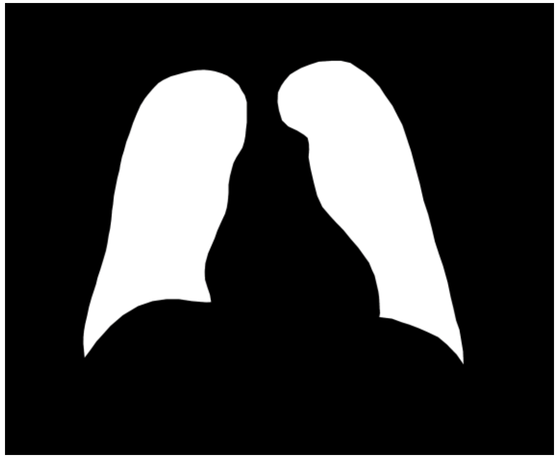
$$OR \quad AND \quad NOT$$

- Example:
  - Masking (AND) operation to separate the object and background of a biomedical images

Kalau xray, kalau mau liat paru-paru aja, kita bisa pakai AND untuk lihat paru-paru yang kita memang mau liat aja.



XRay



Mask

AND

Note: Operators on Images

- Operations
  - Array operations: *pixel-by-pixel*

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{bmatrix}$$

  - Matrix operations

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

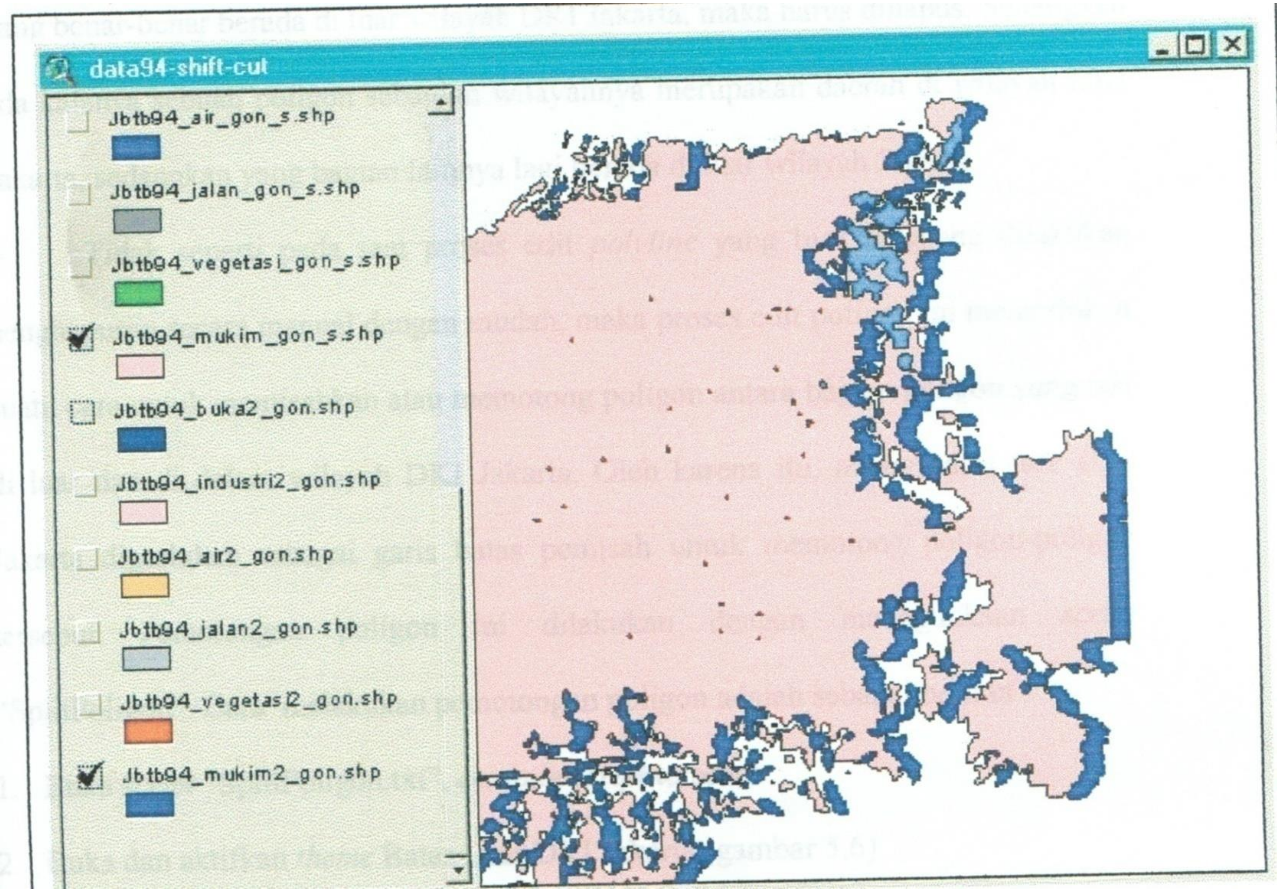
- Mathematical operations on images → assume array operations*)

mathematical operation on images, assume array operation. jadi a11xa11, a12 b12

Distortions in Digital Images

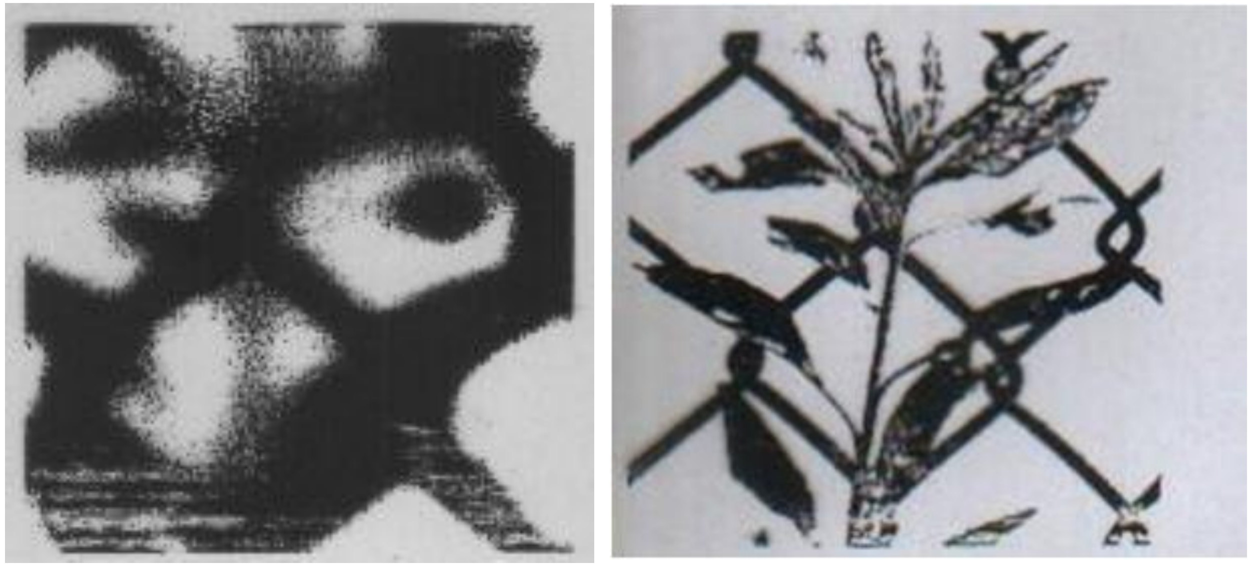
(Source: Ira Hastitu *et. al,* 2002)

- Geometric Distortions
  - Spatial distortion
  - Result of a change in position and direction because of movement of the person capturing the image of the object captured
  - *) Could also be a result of a fault of the internal camera sensor



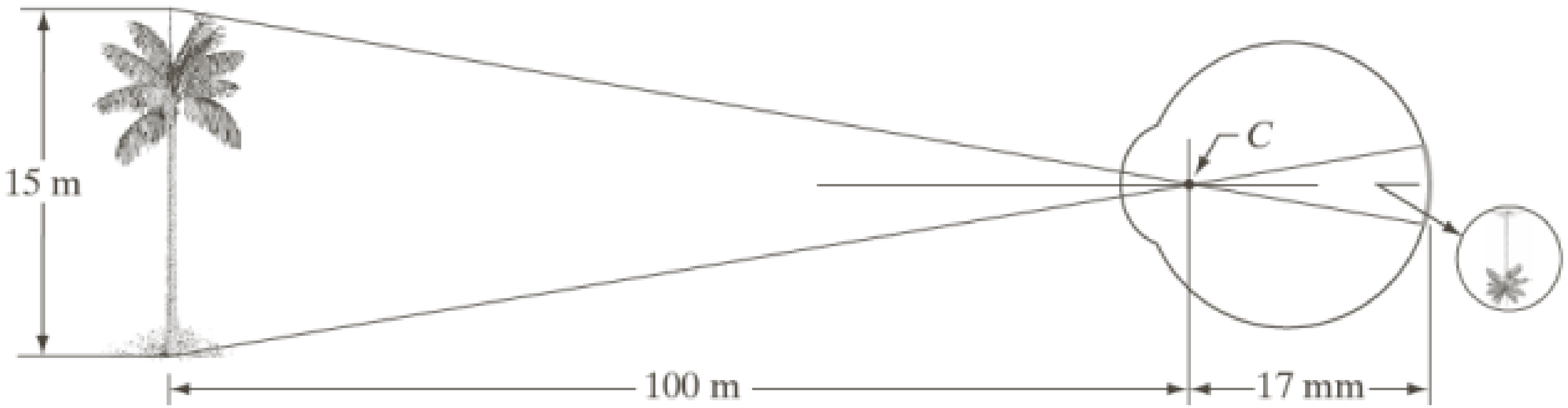
Distortions in Digital Images (2)

- Radiometric Distortions
  - Due to an incorrect intensity distribution
  - Result of a different atmosphere / environment condition (haze, etc) resulting in a different gray level captured
  - *) Could also be a result of a fault of the internal camera sensor
  - Can be corrected digitally through filters



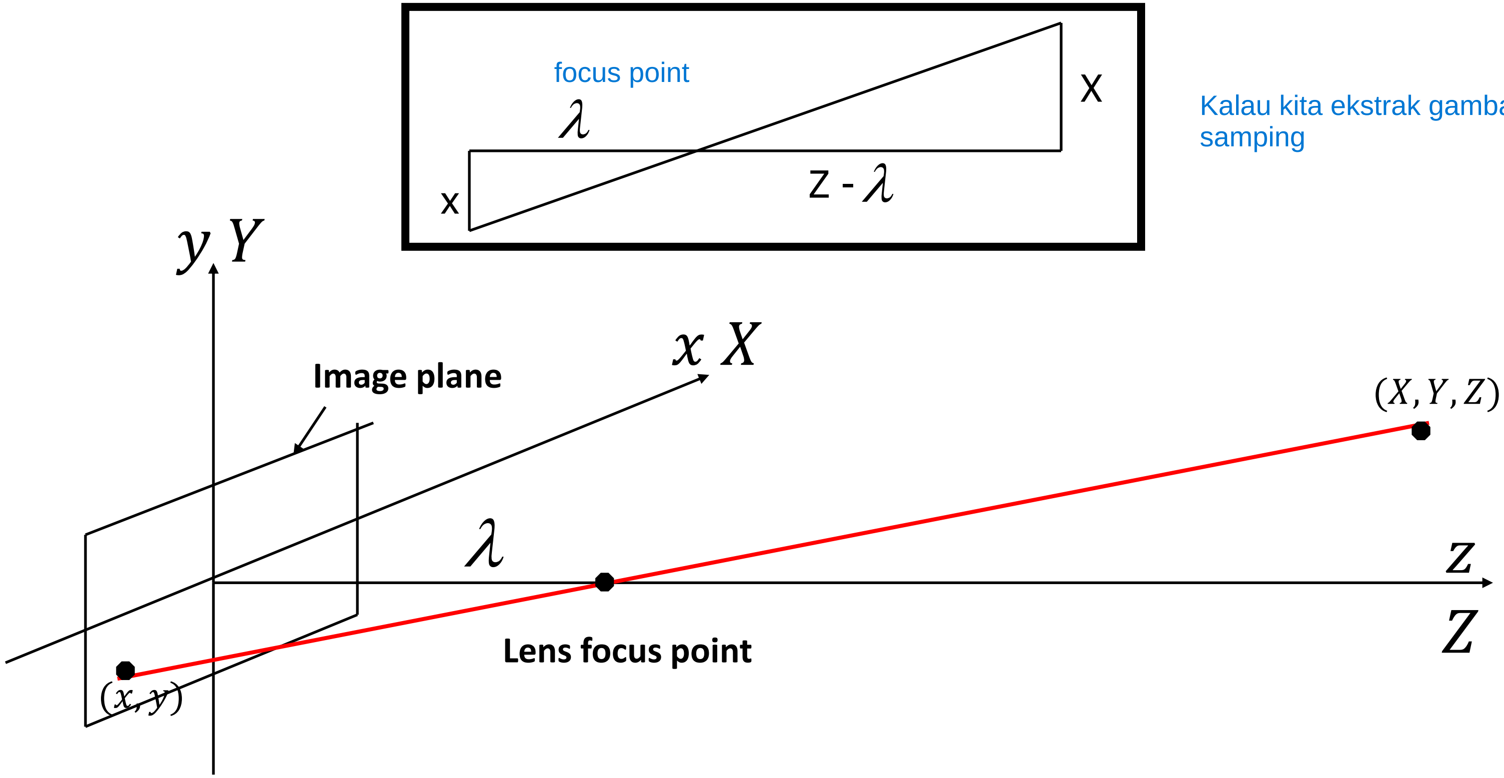
Source: Leaves on a branch (MSU, 1990)

Image Formation in the Eye vs the Camera



- Photo camera: lens has fixed focal length. We can focus at various distances by varying the distance between lens and imaging plane (location of film or chip)
- Human eye: Distance lens-imaging region (retina) is fixed. Focal length for proper focus obtained by varying the shape of the lens.

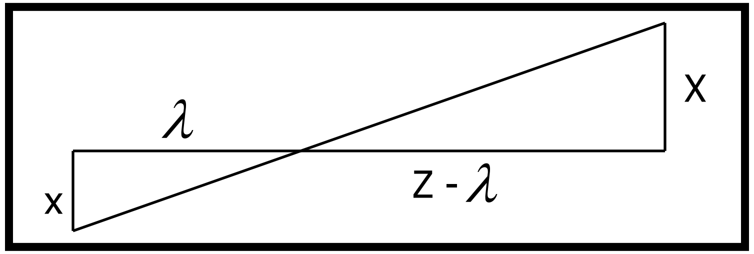
Camera and World Coordinate System



Kalau kita ekstrak gambar ini, kita dapet gambar yang samping

- The camera coordinate system (x, y, z) aligned with the world coordinate system (X, Y, Z)

Camera and World Coordinate System (3)



- With both axes (camera and world coordinates) aligned, the object in the world and the image on the image plane with create similar triangles (*segitiga sama dan sebangun*)

$$\frac{-x}{\lambda} = \frac{X}{Z-\lambda}$$

rumus untuk mengtranslate dari letak dan ukuran di dunia nyata vs letak dan ukuran di dunia maya (komputer)

$$x = \frac{\lambda X}{\lambda - Z}$$

$$y = \frac{\lambda Y}{\lambda - Z}$$

Homogeneous Coordinate System

- Object coordinates in the *world system* are commonly written in the *homogeneous coordinate system*
- For the Cartesian coordinates (W_c) dan homogeneous coordinates (W_h):

$$W_C = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \qquad W_H = \begin{pmatrix} kX \\ kY \\ kZ \\ k \end{pmatrix}$$

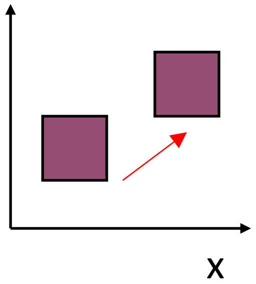
k is *a non-zero constant,* usually $k = 1$.

- Thus, how can the Cartesian coordinate $W_c\ (X, Y, Z)$ be obtained from W_h?

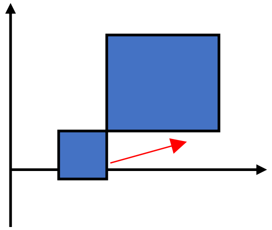
Geometric Transformations in the Homogeneous Coordinate System

- Why?
  - The need of a uniform representation for various cameras and setups (homogeneous representation)
  - To enable a composite transformation efficiently
  - To save the normalization factor of the coordinates due to a series of transformations

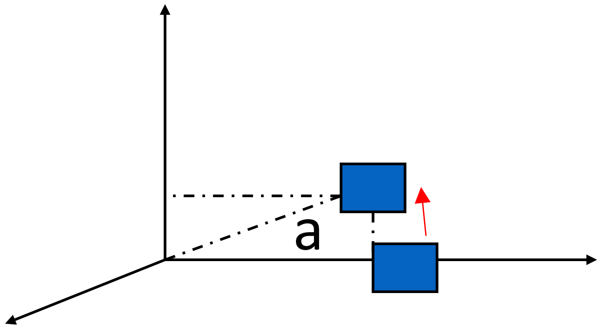
Geometric Transformations in Images



Translation



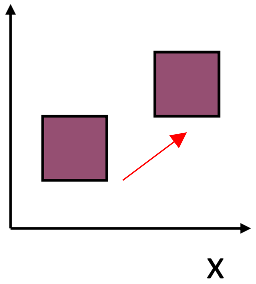
Scale / Dilatation



Rotation

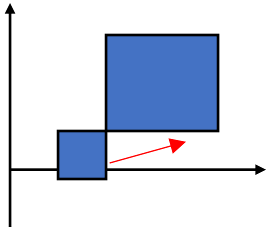
- Recall linear geometric transformation through matrix multiplication

Geometric Transformations in Images



Translation

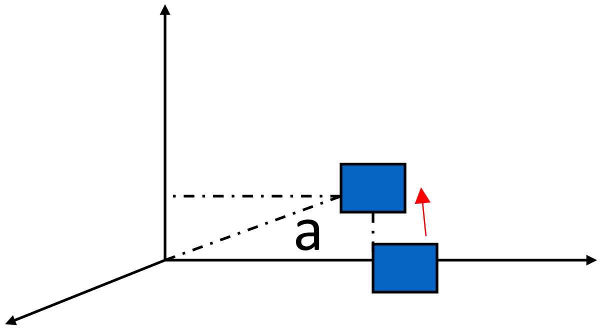
$$X' = X + Tx \quad Y' = Y + Ty$$



Scale / Dilatation

rumus skalar

$$X' = Sx.X \qquad Y' = Sy.Y$$



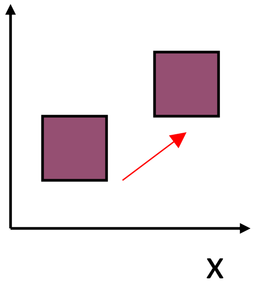
Rotation

rumus rotasi

$$X' = X\cos(a) \quad Y' = X\sin(a)$$

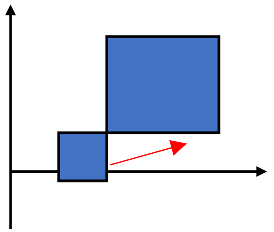
Transformation Matrices for Geometric Transformations

perkalian matriks



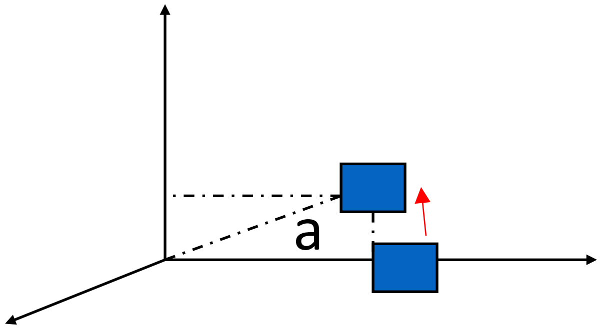
Translation

$$\begin{bmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



Scale / Dilatation

$$\begin{bmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



Rotation

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha & 0 \\ 0 & -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Perspective Transformation

perspective: awal2 2d jadi 1d

- Perspective transformation matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{\lambda} & 1 \end{bmatrix}$$

Kita bisa mengira-ngira letak dari orang

- Kita bisa memetakan kalau seseorang di X, Y tapi bukan Z axisnya

Tapi kalau kita foto seseorang dalam angle lain maka kita punya X dan Y yang kita ketahui

Jadi kita punya 2 SPL (SIstem persamaan linear), nah kita bisa dapet Z nya.

  - Minus means the image is upside down, λ is the lens focus length, and $\frac{1}{\lambda}$ is the scale factor.
  - The object coordinate on the camera can be obtained from the object coordinate in the world using the perspective transformation

Homogeneous Coordinate System on the Camera

- The object coordinates in the camera : C_C for C_h the Cartesian and homogeneous coordinates, respectively
- Homogeneous coordinate system on the camera:

$$C_h = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{\lambda} & 1 \end{bmatrix} \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix} = \begin{bmatrix} kX \\ kY \\ kZ \\ -(\frac{kZ}{\lambda}) + k \end{bmatrix}$$

- Cartesian coordinate on the camera $C_C(X, Y, Z)$ can thus be obtained from the homogeneous coordinates C_h :

Basic Camera Mathematical Model

- The Cartesian coordinates of the camera systems

$$C_c = \begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} kX/(-(kZ/\lambda)+k) \\ kY/(-(kZ/\lambda)+k) \\ kZ/(-(kZ/\lambda)+k) \end{vmatrix} = \begin{vmatrix} \lambda X/(\lambda - Z) \\ \lambda Y/(\lambda - Z) \\ \lambda Z/(\lambda - Z) \end{vmatrix}$$

The relation between the (x, y, z) and (X, Y, Z) is the Basic Camera Mathematical Model

Color Image Formation

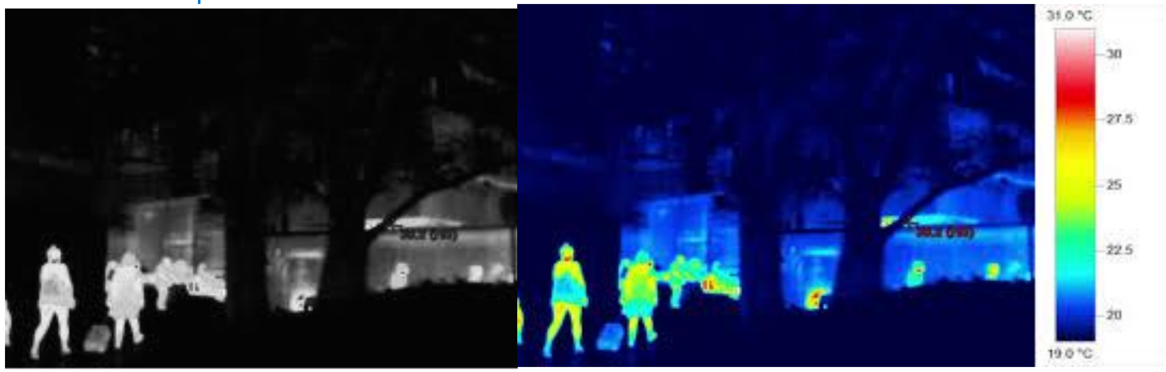
Pseudocolor Images

infrared is not really red

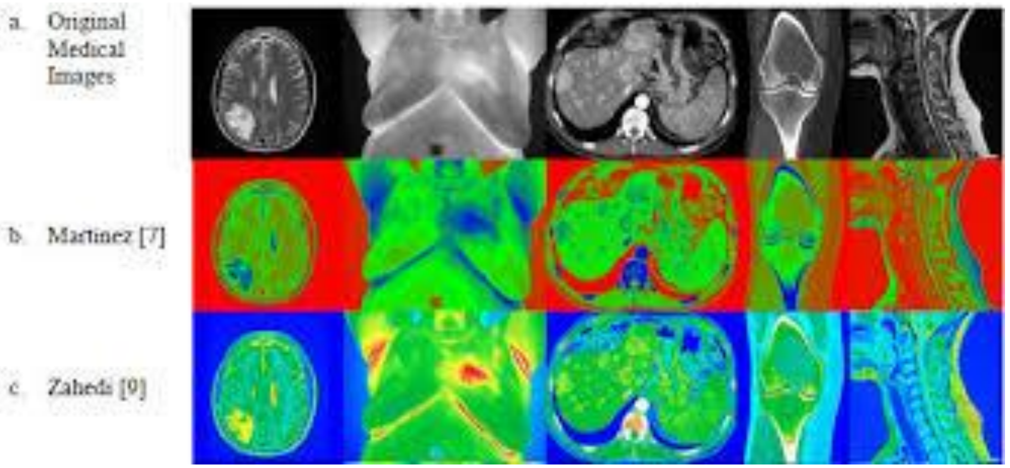
But what is the one that being caught is that the intensity of the picture

computer

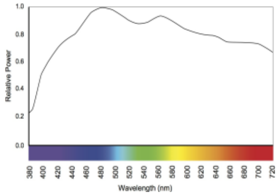
human want this



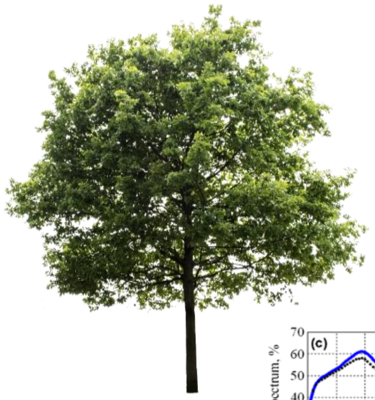
- Pseudocolor / *false color* : The color in the image is obtained by manually assigning a color to a certain range of gray level intensities for visualization purposes.
- Why pseudocolor?
  - Human vision can distinguish **thousands** of colors
  - Human vision can only distinguish **less than a hundred** grayscale values

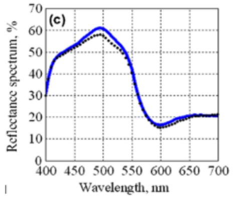


Color Image Acquisition

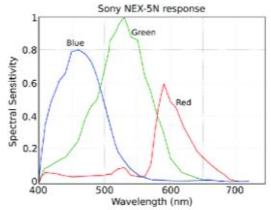


SPD of Illuminants (E)



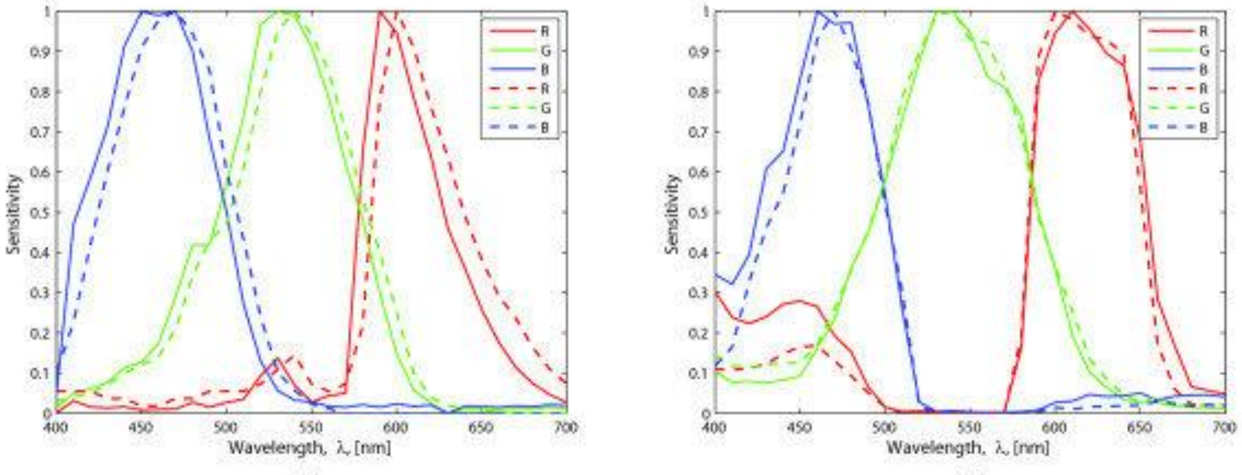


Reflectance Coefficient ρ

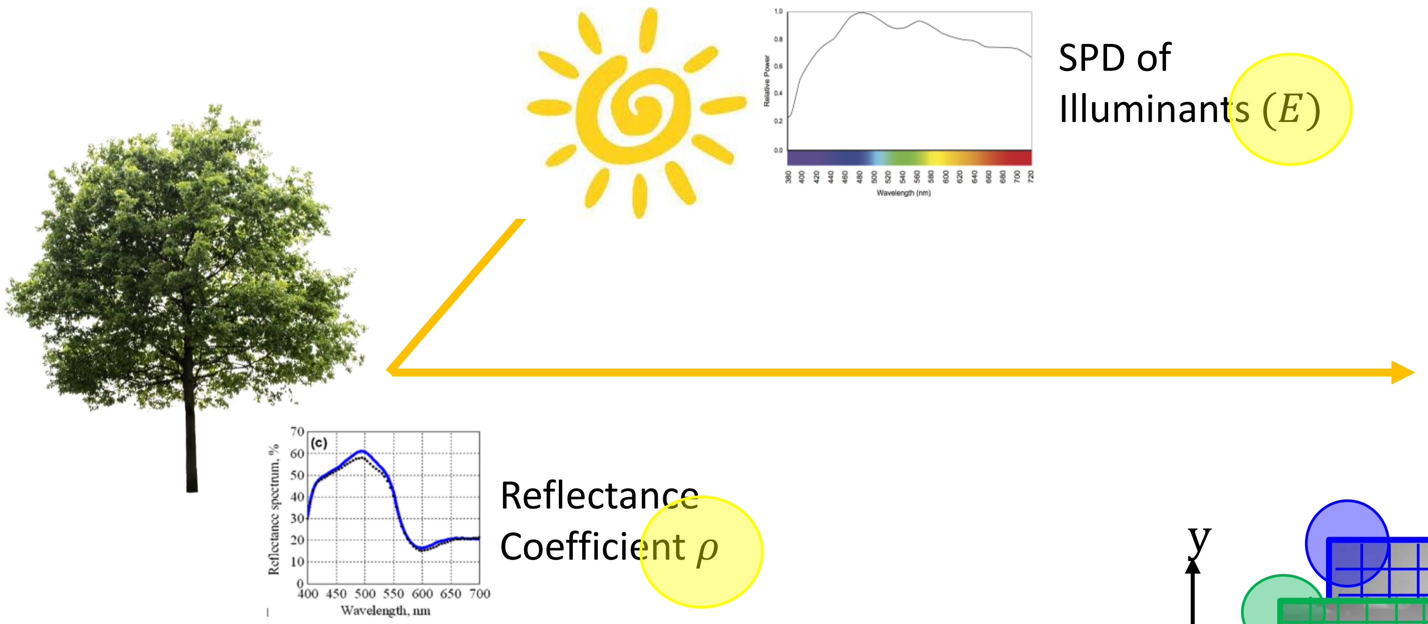


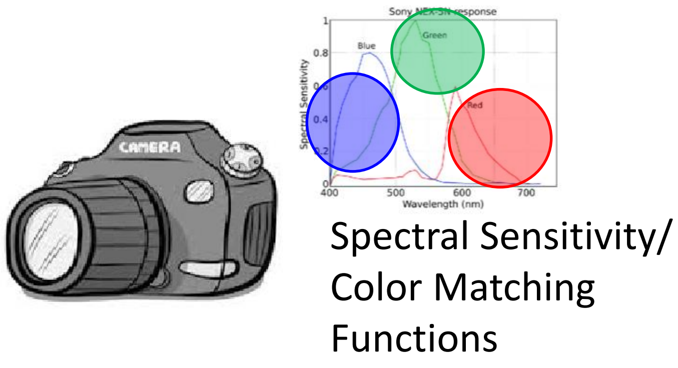
Spectral Sensitivity/ Color Matching Functions

- **Camera Sensor Sensitivities** for RGB Capture: a Nikon D70 (solid) and Canon 20D (dotted). b Fujifilm 3D (Left - solid, Right - dotted).



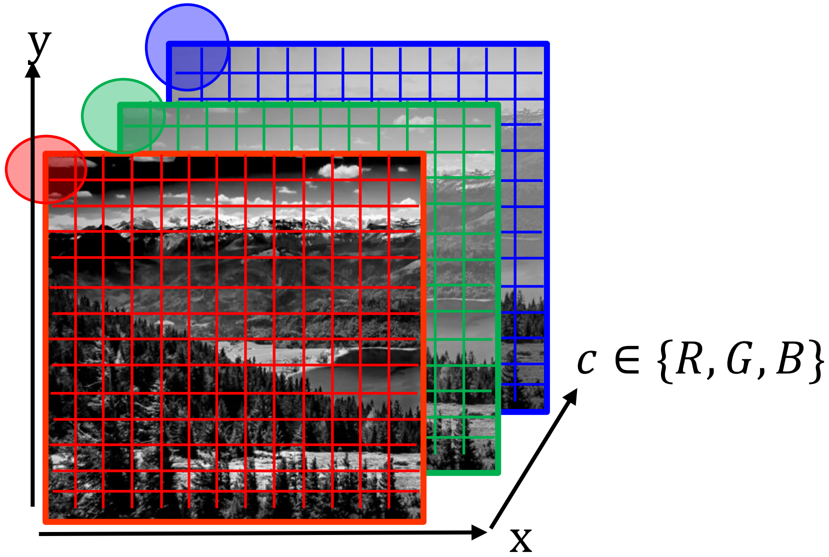
Digital Color Images



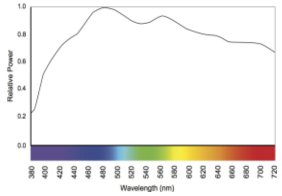


- The final image captured is the result of matrix multiplication of Illumination E, reflectance (ρ), and sensor sensitivities of $(R.G.B)$
- Resulting in the RGB Image

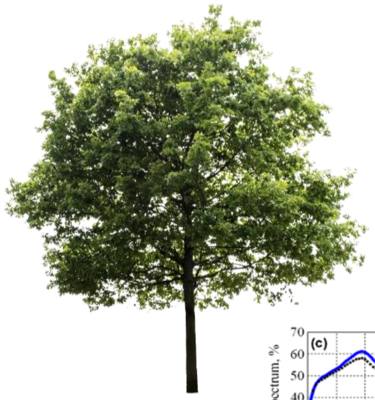
disetiap titik pixel ada R, G, B nya.

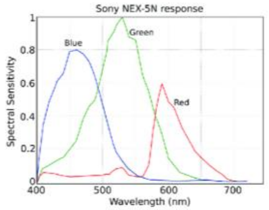


Color is a Perception / Sensation

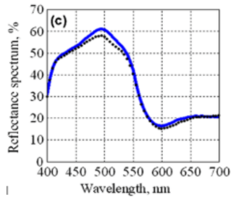


SPD of Illuminants (E)





Spectral Sensitivity/ Color Matching Functions



Reflectance Coefficient ρ

hasil erkalian cahaya (SPD) dikali dengan reflectance (p).

warna yang kita lihat di mata sama kamera beda.

- Each sensor (eye, camera) will have a specific spectral sensitivity (R,G,B)
- **You may not all see the same color**
- How is color defined/communicated?

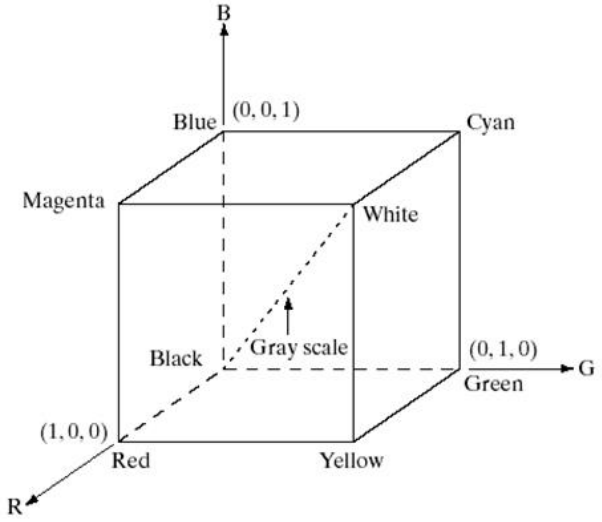
A Bit of Color Trivia

- How many colors can humans distinguish?
  - The estimate is more than 1 million unique colors
- How do we communicate with each other and describe a certain color perception?
  - Samples. Human vocabulary is very limited
- We need some type of **system** that can order and organize the colors
  - To facilitate the specification of color
  - Based on some attributes that were agreed upon

Color Order Systems / Color Model

RGB Color Model

- Standard model for color monitor
- Based on tri-stimulus theory of color vision
- There are so many variables among different systems / hardware, so we need a reliable subset of colors that can always be reproduced reliably on any system
  - Safe RGB colors / all-systems safe colors / safe Web colors/ safe browser colors.
- The variety of color reproducing systems process colors differently
  - We have 216 colors that can be reproduced by all systems *de facto*



CIE 15: Technical Report: Colorimetry, 3rd edition. 2014

CIELAB Color Space

RGB itu tidak bebas linear.

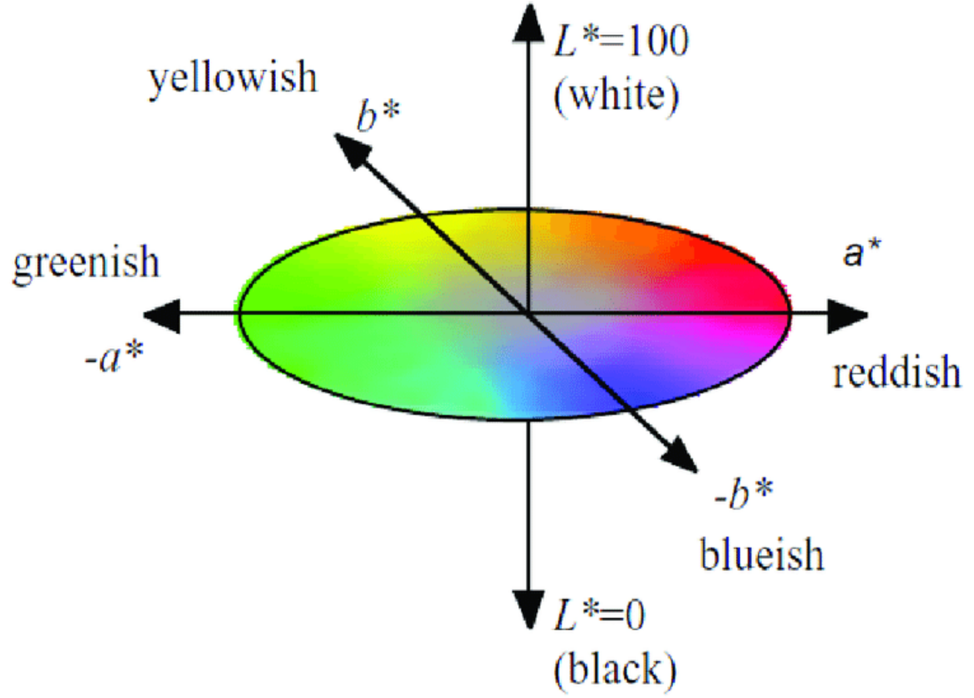
Mending ubah Intensity

RGB = physical representation of color

L tidak mempengaruhi warna nya

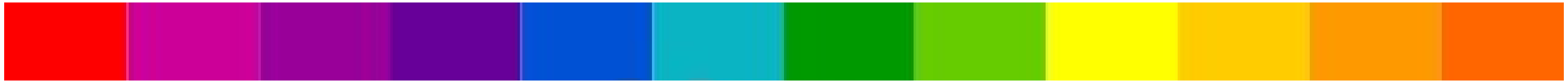
kita gedein L ataupun kecilin tidak akan mempengaruhi warnanya

- The nonlinear relations for $L*$, $a*$, and $b*$ are intended to mimic the nonlinear response of the eye
- Components $L, a*, b*$
  - $\boldsymbol{L}*$ for the lightness
  - $\boldsymbol{a}*$ and $\boldsymbol{b}*$ for the green–red and blue–yellow components

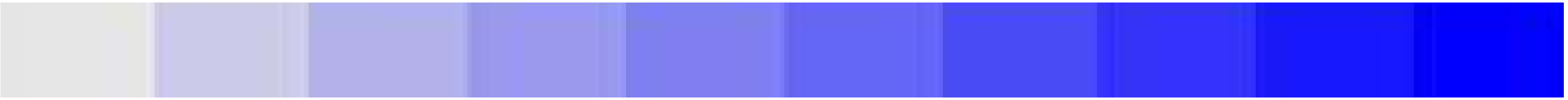


HSI (also known as HSV / HSL)

- Hue



- Saturation



- Intensity

Look it up yourselves..

- **Color systems:**
  - CIEXYZ, CIELAB, CIELUV,…
  - RGB variants: sRGB, Adobe RGB, …
  - YIQ, YUV, YCbCr,…

V · T · E	Color space [hide]
	List of color spaces · Color models
CAM	CIECAM02 · iCAM
CIE	XYZ (1931) · RGB (1931) · CAM (2002) · YUV (1960) · UVW (1964) · **CIELAB (1976)** · CIELUV (1976)
RGB	RGB color space · sRGB · rg chromaticity · Adobe · Wide-gamut · ProPhoto · scRGB · DCI-P3 · Rec. 709 · Rec. 2020 · Rec. 2100
YUV	YUV (PAL) · YDbDr (SECAM · PAL-N) · YIQ (NTSC) · YCbCr (Rec. 601 · Rec. 709 · Rec. 2020 · Rec. 2100) · ICtCp (Rec. 2100) · YPbPr · xvYCC · YCoCg
Other	CcMmYK · CMYK · Coloroid · LMS · Hexachrome · HSL, HSV · HCL · Imaginary color · OSA-UCS · PCCS · RG · RYB · HWB
Color systems and standards	ACES · ANPA · Colour Index International (CI list of dyes) · DIC · Federal Standard 595 · HKS · ICC profile · ISCC–NBS · Munsell · NCS · Ostwald · Pantone · RAL (list)
	For the vision capacities of organisms or machines, see Color vision.