



ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



Thị giác máy tính (computer vision)

Bài 1: Giới thiệu tổng quan

Giới thiệu

- Giảng viên:
 - TS. Nguyễn Thị Oanh, TS. Trần Nguyên Ngọc, TS. Đặng Tuấn Linh
 - Email:
 - oanhnt@soict.hust.edu.vn,
 - ngoctn@soict.hust.edu.vn
 - linhdt@soict.hust.edu.vn
 - Đơn vị công tác:
 - Trường Công nghệ thông tin và truyền thông - Đại học Bách khoa Hà Nội
- TA:
 - KhánhTQ, MinhNQN, TriNQ, TùngBT

Nội dung môn học

- Intro to CV
- Image Enhancement
- Frequency Domain
- Edge detection
- Feature and Image matching
- Camera model
- Multi-view
- Object recognition
- Object detection
- Semantic segmentation
- Motion and tracking
- Action Recognition
- Transfer learning

Cách thức triển khai

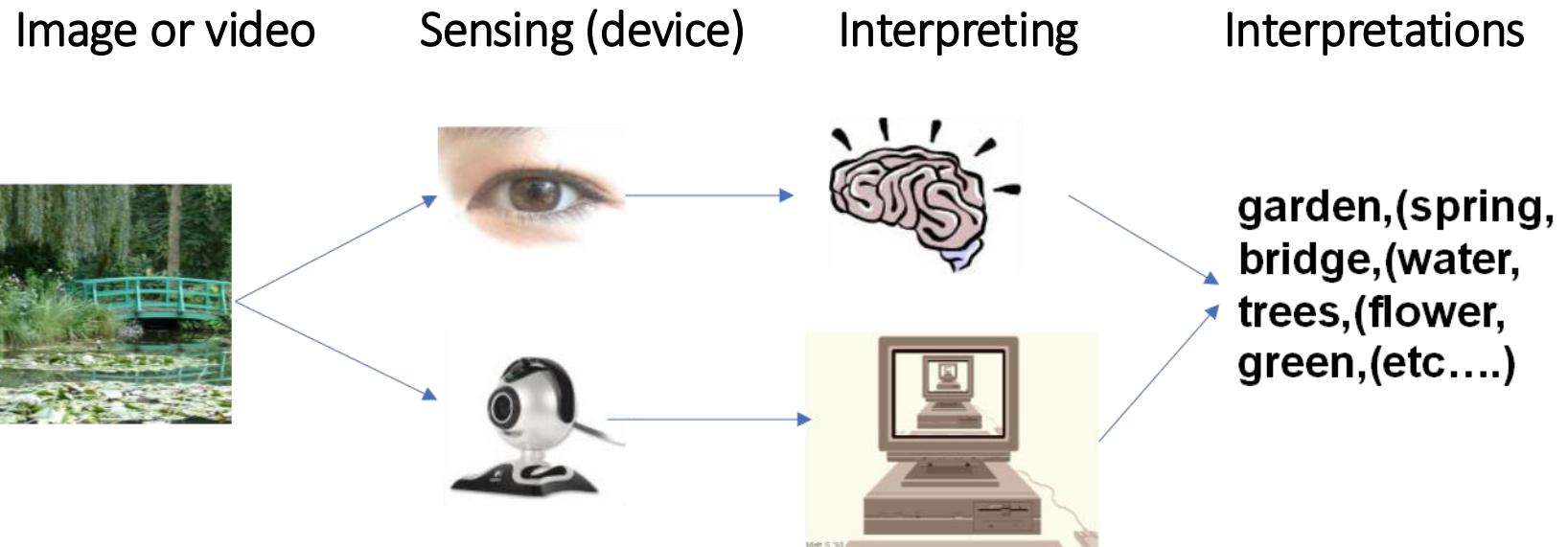
- Buổi học:
 - Trên lớp 3h = 2h lý thuyết (TA) + 1h thực hành (GV + TA)
 - BTVN: giao và nộp trên bkict
- Đánh giá giữa kỳ và cuối kỳ:
 - code + trắc nghiệm (bkict)
- Project: 4-6 SV/nhóm
 - 01 chủ đề/ nhóm: gợi ý + đề xuất
 - 01 buổi: bài toán + định hướng giải pháp
 - 01 buổi: bảo vệ kết quả cuối cùng

Nội dung Bài 1

- Giới thiệu chung
 - Khái niệm
 - Các cấp độ xử lý (Low level vision, Middle level vision, High level vision)
 - Lĩnh vực liên quan
 - Ứng dụng
- Quá trình hình thành và thu nhận ảnh
- Không gian màu
- Thực hành

Thị giác máy tính?

Human vision



Computer vision

From CS131 course “computer vision”,
Prof. Fei-Fei Li, Stanford 'Vision'Lab

Mục đích của thị giác máy tính

- Cầu nối giữa giá trị của các điểm ảnh (pixel) và "ngữ nghĩa" của bức ảnh



What we see

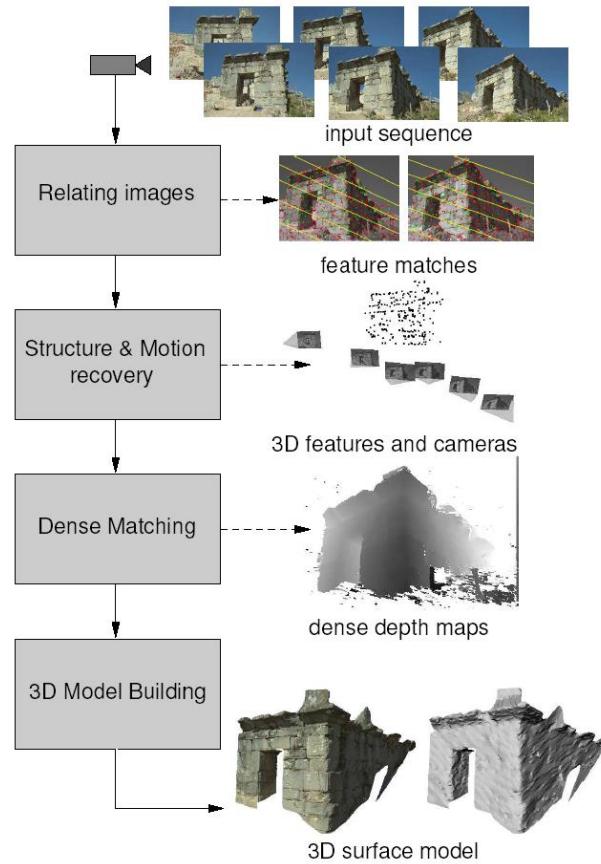
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

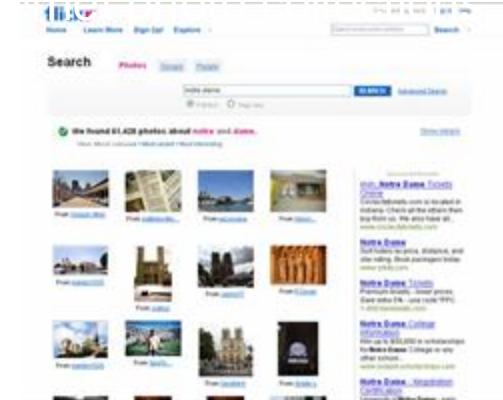
Loại thông tin có thể trích xuất từ ảnh?

- Thông tin 3D (ảnh coi như thiết bị đo lường)
- Thông tin ngũ nghĩa (ảnh: nguồn chứa ngũ nghĩa)

Vision as measurement device

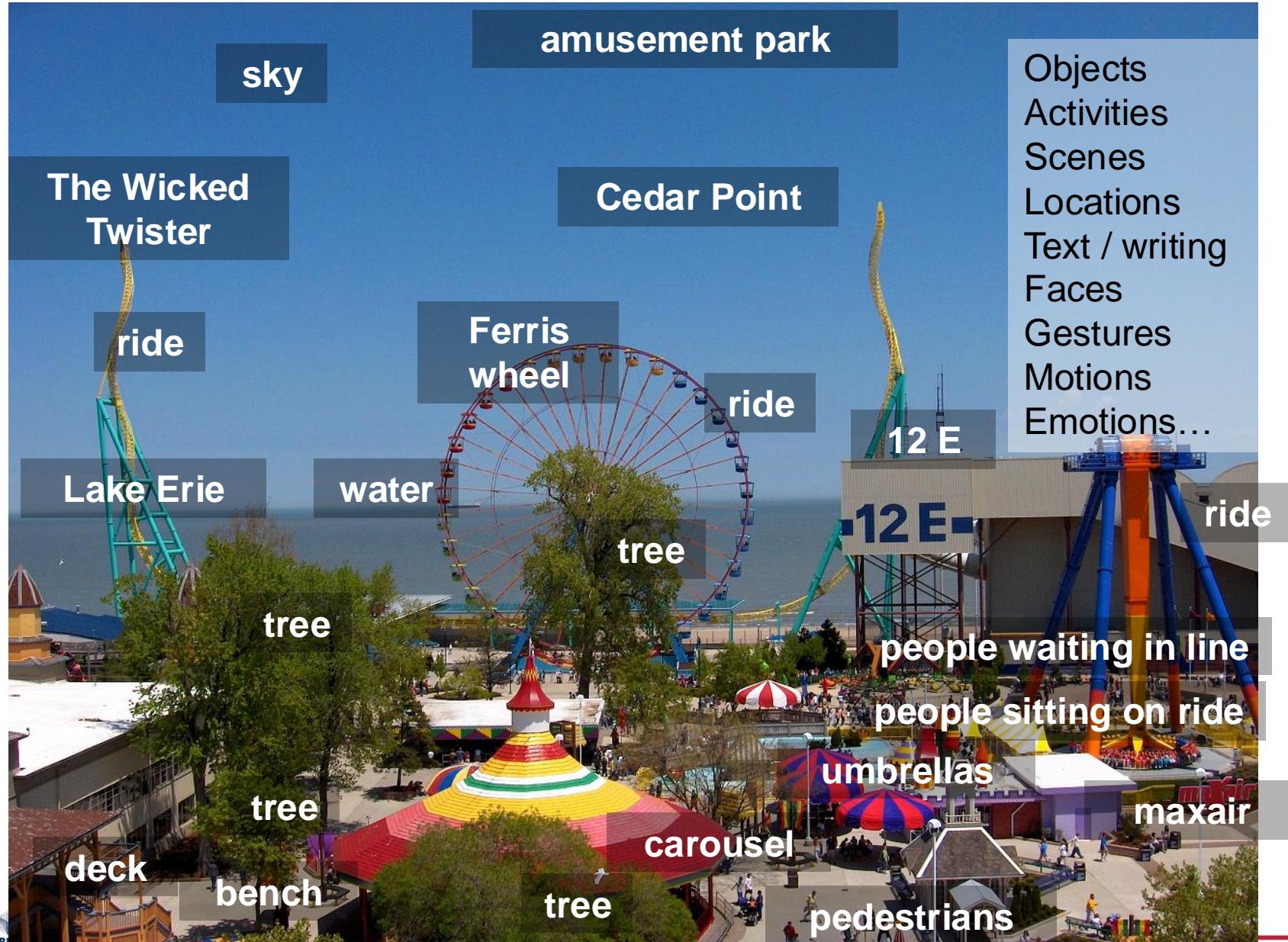


Pollefeys et al.



Goesele et al.

Vision as a source of semantic information



Thị giác máy tính?

- Là lĩnh vực khoa học **liên ngành** cho phép máy tính có thể **hiểu được nội dung bức ảnh/video ở mức cao**

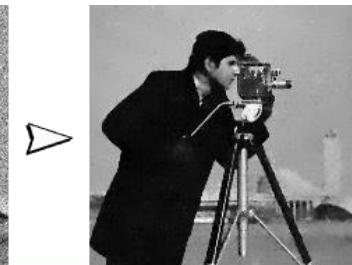
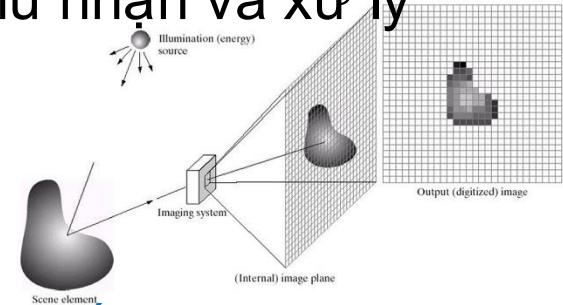


*What kind of scene?
Where are the cars?
How far is the building?
...*

Các mức độ xử lý

Levels of vision

- **Mức thấp (Low-level Vision):** tạo ảnh, thu nhận và xử lý ảnh
 - **Tạo ảnh:** quá trình tạo ra ảnh/video
 - **Thu nhận ảnh:**
 - Ảnh số được thu nhận bởi một số các **bộ cảm biến (sensors)**.
 - Tùy thuộc loại cảm biến, ảnh thu được có thể là ảnh 2D, 3D hoặc chuỗi ảnh.
 - **Xử lý ảnh** tập trung vào việc xử lý dữ liệu ảnh 2D như tăng cường độ cương phản, lọc nhiễu, biến đổi ảnh, ...
 - Được xem như giai đoạn tiền xử lý cần thiết trên dữ liệu đầu vào cho các ứng dụng thị giác máy tính
 - Làm việc với ảnh như là 1 ma trận
 - Đầu vào: ảnh → Đầu ra: ảnh

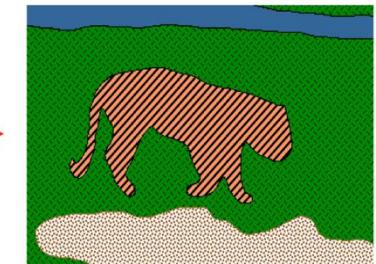
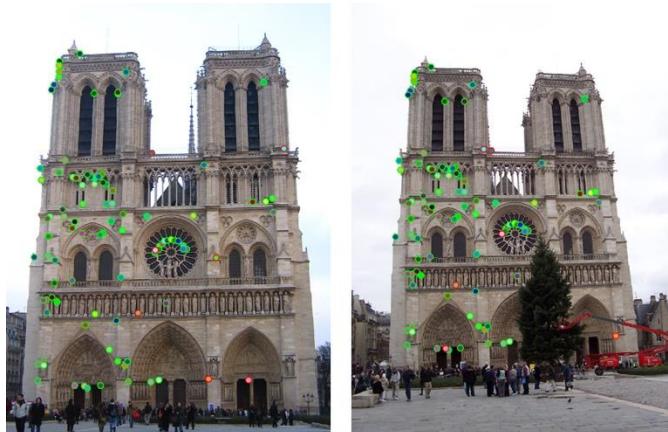


Các mức độ xử lý

Levels of vision

- **Mức giữa (Middle-level Vision):**

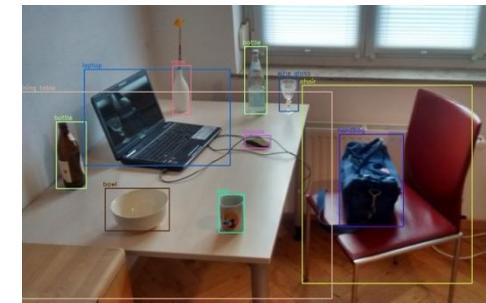
- Trích chọn đặc trưng: đặc trưng ảnh có mức độ phức tạp khác nhau được trích chọn từ ảnh. VD: Cạnh, góc, đường, kết cấu, hình dáng, ...
- So khớp ảnh
- Phân vùng ảnh



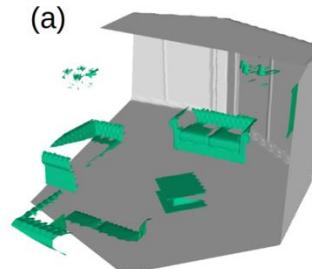
Các mức độ xử lý

Levels of vision

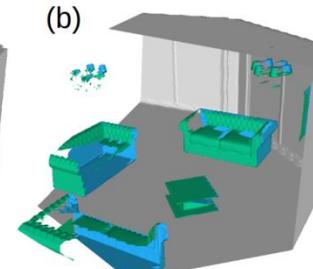
- **Mức cao (High-level Vision):** mức ngữ nghĩa cao.
VD: nhận dạng object, hiểu nội dung bức ảnh
 - Input: ảnh
 - Output: thông tin ngữ nghĩa, ra quyết định
- Một số chủ đề:
 - Nhận dạng (phân loại), định danh
 - Phát hiện
 - Phân tích chuyển động
 - Tái tạo cảnh, dựng 3D
 - ...



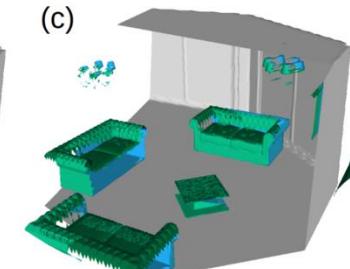
RGB Image



2.5D Object Surfaces

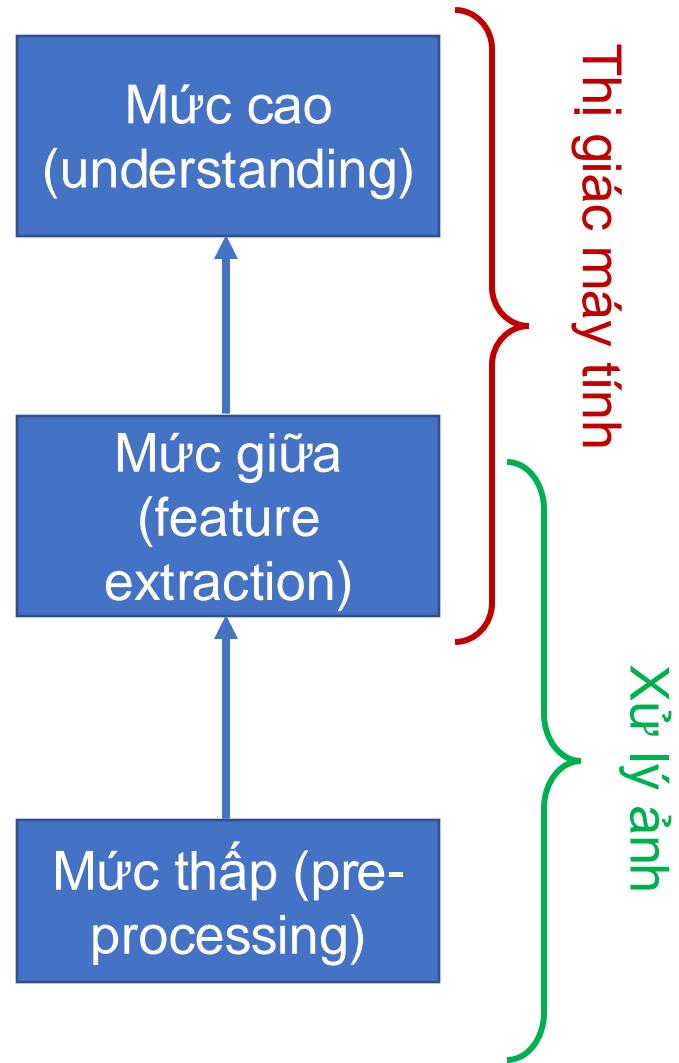
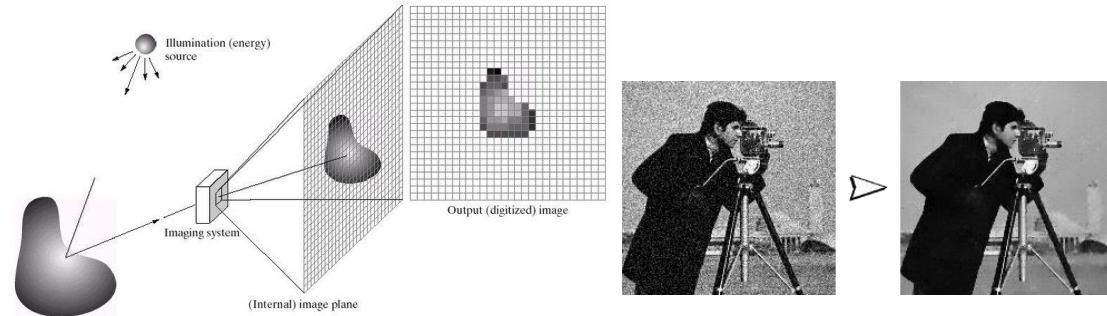
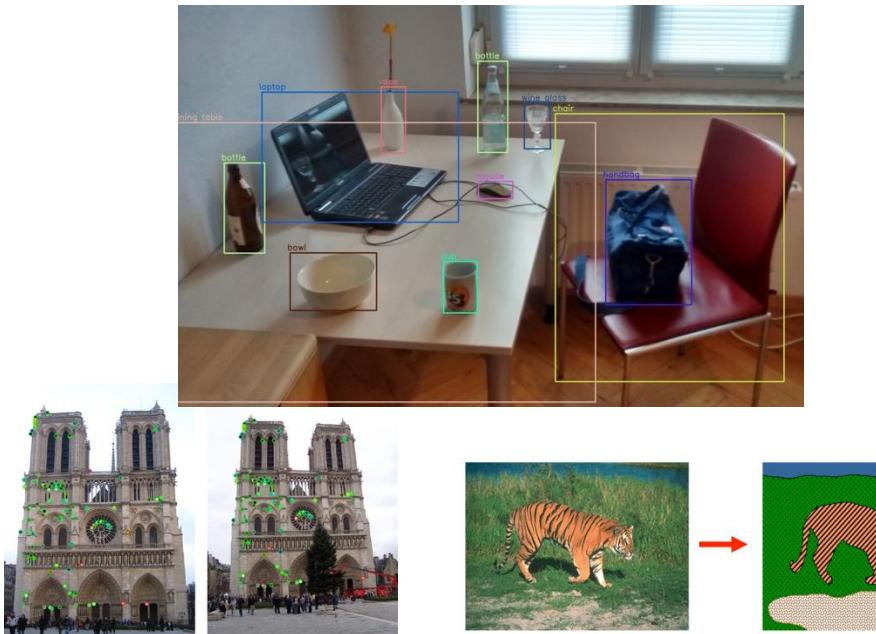


Multi-layer Surfaces



Multi-layer and
Virtual-view Surfaces

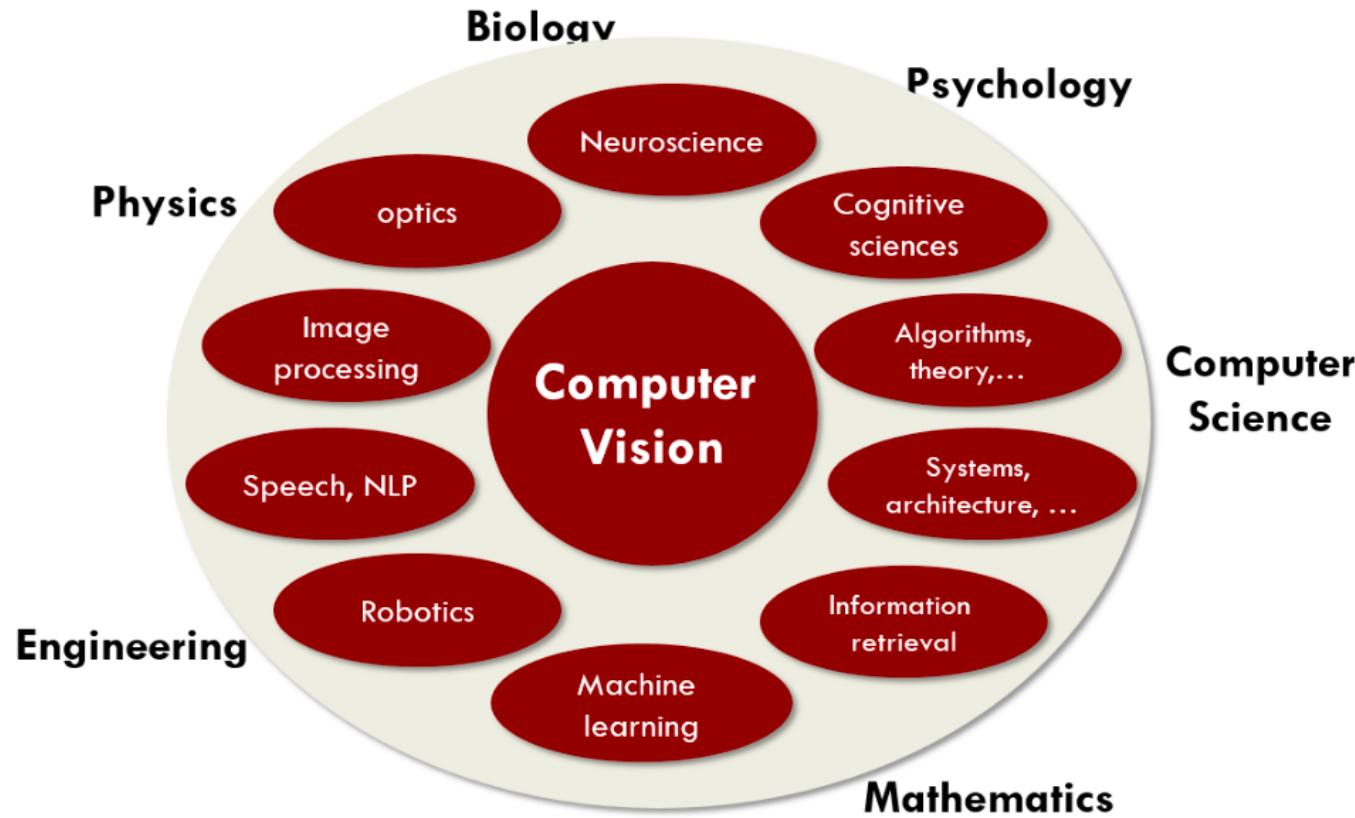
Xử lý ảnh vs. Thị giác máy tính



Một số vấn đề trong xử lý ảnh

- Thu nhận ảnh
- Nắn chỉnh biến dạng
- Khử nhiễu
 - Nhiễu hệ thống: có quy luật (dùng các phép biến đổi)
 - Nhiễu ngẫu nhiên (dùng bộ lọc)
- Chỉnh mức xám
- Nén ảnh

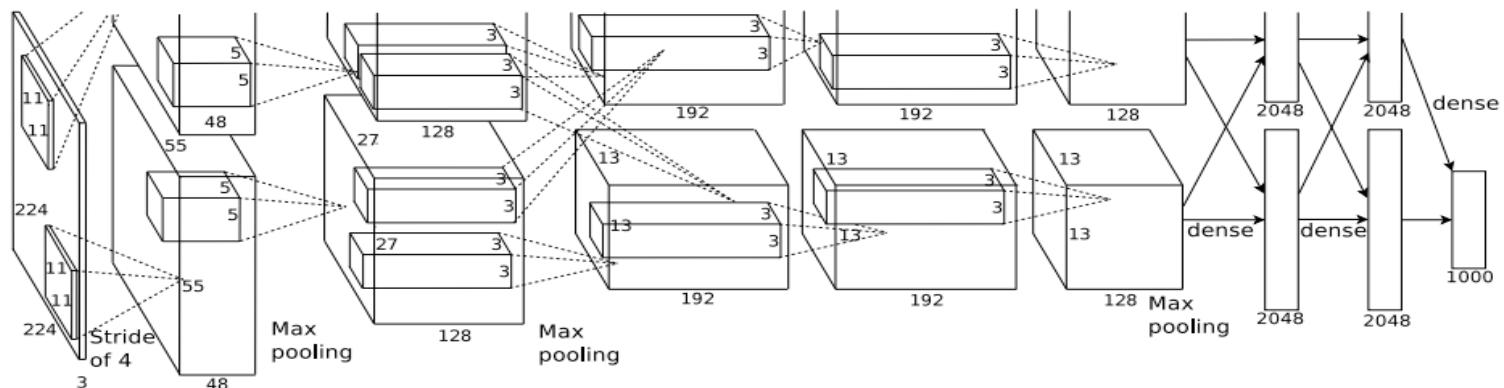
Lĩnh vực liên quan



Computer vision at the intersection of multiple scientific fields

Lĩnh vực liên quan

- Học máy: “The field of study that gives computers the ability to learn without being explicitly programmed.” – Arthur Samuel
 - Liên quan mật thiết, cùng chia sẻ chủ đề về nhận dạng mẫu và phương pháp học
 - Computer vision - Deep learning: Artificial Neural Networks with many layers (CNN: Convolutional Neural Network)



Miền ứng dụng

Robotics Application

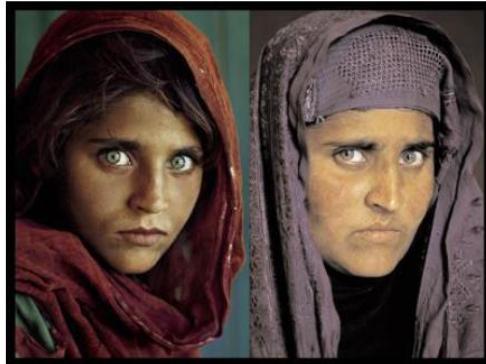
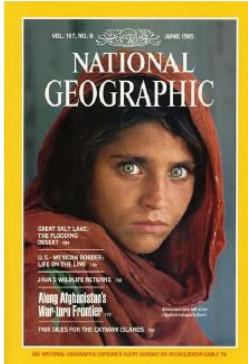
- Localization-determine robot location automatically
- Navigation
- Obstacles avoidance
- Assembly peg – in – hole, welding, painting
- Manipulation e. g. PUMA robot manipulator
- Human Robot Interaction HRI: Intelligent robotics to interact with and serve people

Miền ứng dụng

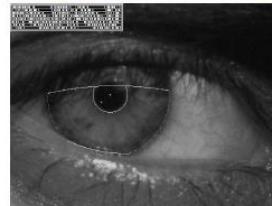
Security Application

- Biometrics iris, fingerprint, face recognition
- Surveillance-detecting certain suspicious activities or behaviors

— ...



[How the Afghan Girl was Identified by Her Iris Patterns](#)



Source: S. Seitz



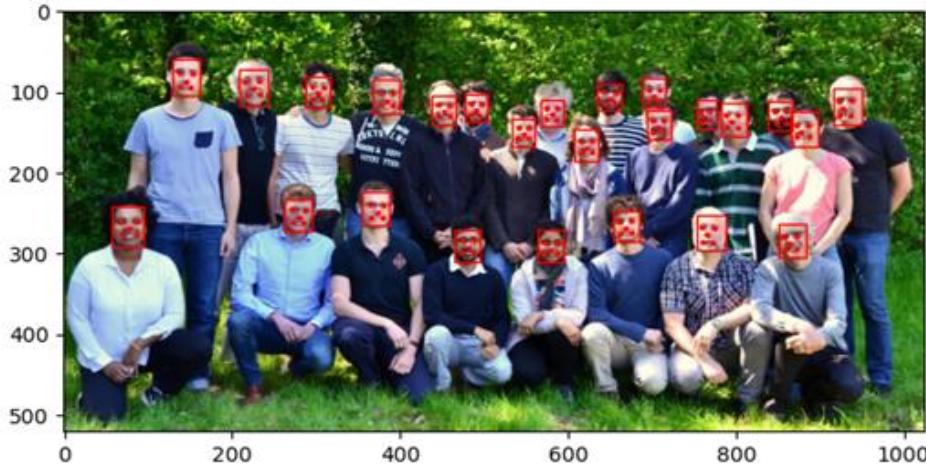
Fingerprint scanners



Face recognition systems

Source: from S. Seitz

Một số ví dụ



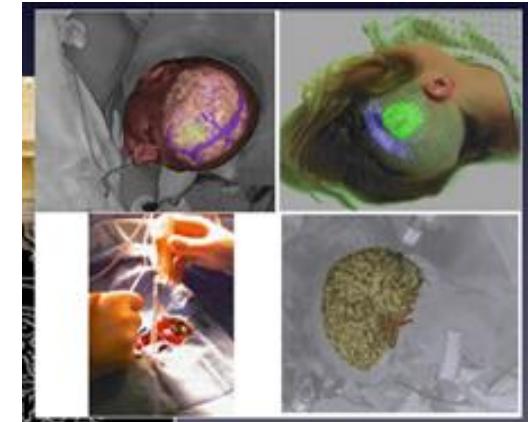
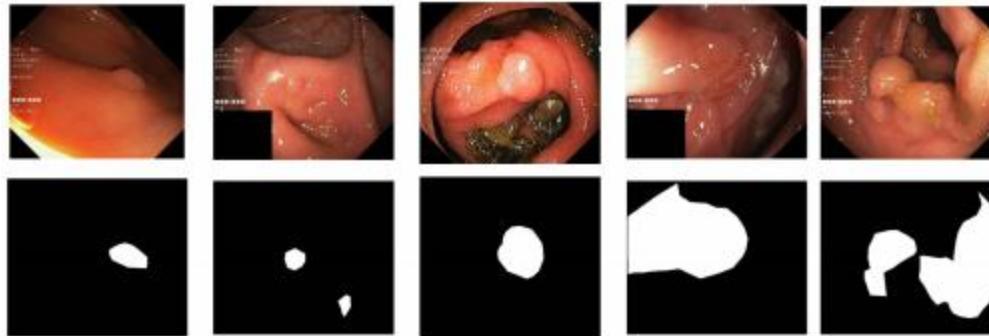
Facebook's suggestion



Miền ứng dụng

Medicine Application

- Classification and detection
- 2D/3D segmentation
- 3D human organ reconstruction MRI or ultrasound
- Vision-guided robotics surgery
- ...



Slide from Jason Lawrence

Miền ứng dụng

Industrial Automation Application

- Industrial inspection defect detection
- Barcode and package label reading
- Object sorting
- Document understanding e. g. OCR
- ...

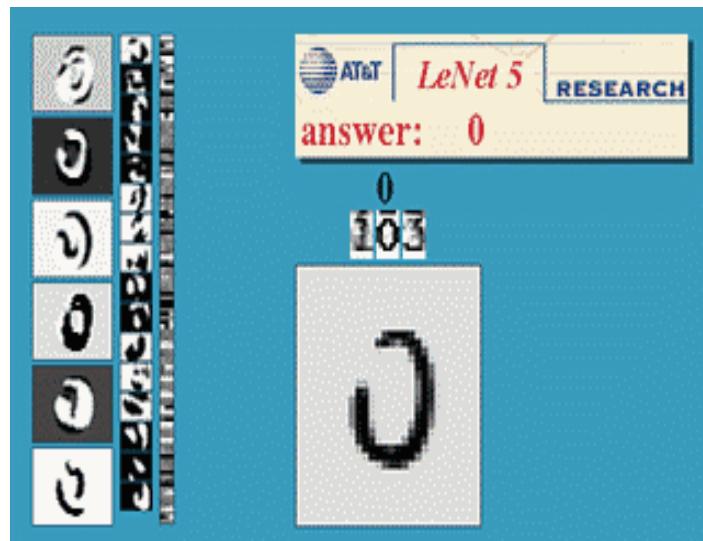
Transportation Application

- Autonomous vehicle
- Safety, e.g., driver vigilance monitoring
- ...

Một số ví dụ

Nhận dạng ký tự

Optical character recognition (OCR)



Digit recognition,(AT&T labs)



License plate readers

http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Source: from S. Seitz

Một số ví dụ

- Xe tự hành

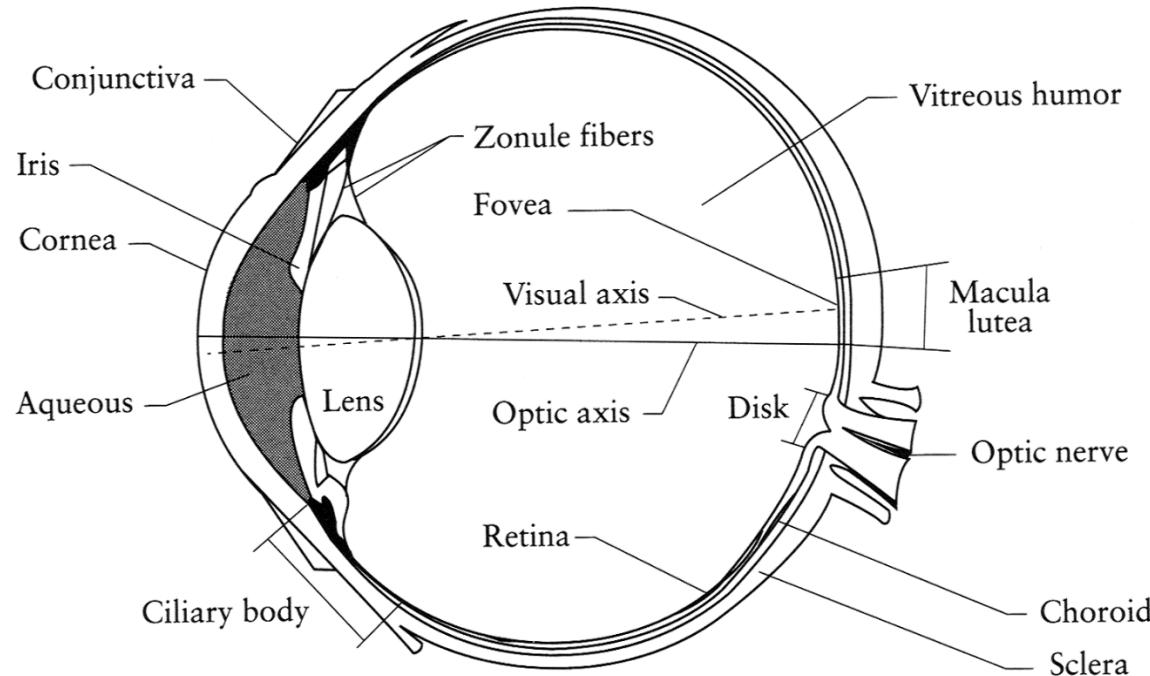


- Mobileye: Vision systems in high-end BMW, GM, Volvo models
 - “In mid 2010 Mobileye will launch a world's first application of full emergency braking for collision mitigation for pedestrians where vision is the key technology for detecting pedestrians.”

Nội dung

- Giới thiệu chung
 - Khái niệm
 - Các cấp độ xử lý (Low level vision, Middle level vision, High level vision)
 - Lĩnh vực liên quan
 - Ứng dụng
- Quá trình hình thành và thu nhận ảnh
- Không gian màu
- Thực hành

Thị giác người



- Mắt người như một máy ảnh
 - **Mống mắt (Iris)** – hình khuyên với các cơ hướng tâm
 - **Con ngươi (Pupil)** – "hole", độ mở được điều khiển bởi cơ mống mắt
 - Cảm biến ?
 - Tế bào thụ cảm quang (rods and cones) nằm trên **võng mạc**

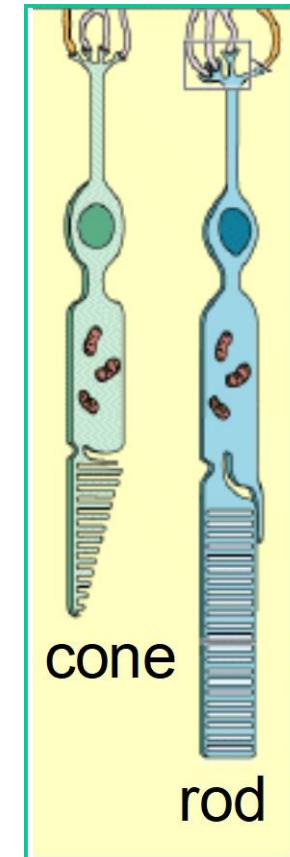
2 loại tế bào tiếp nhận ánh sáng

Cones

- tế bào hình nón
- ít nhạy cảm
- hoạt động trong ánh sáng mạnh
- cảm thụ màu sắc

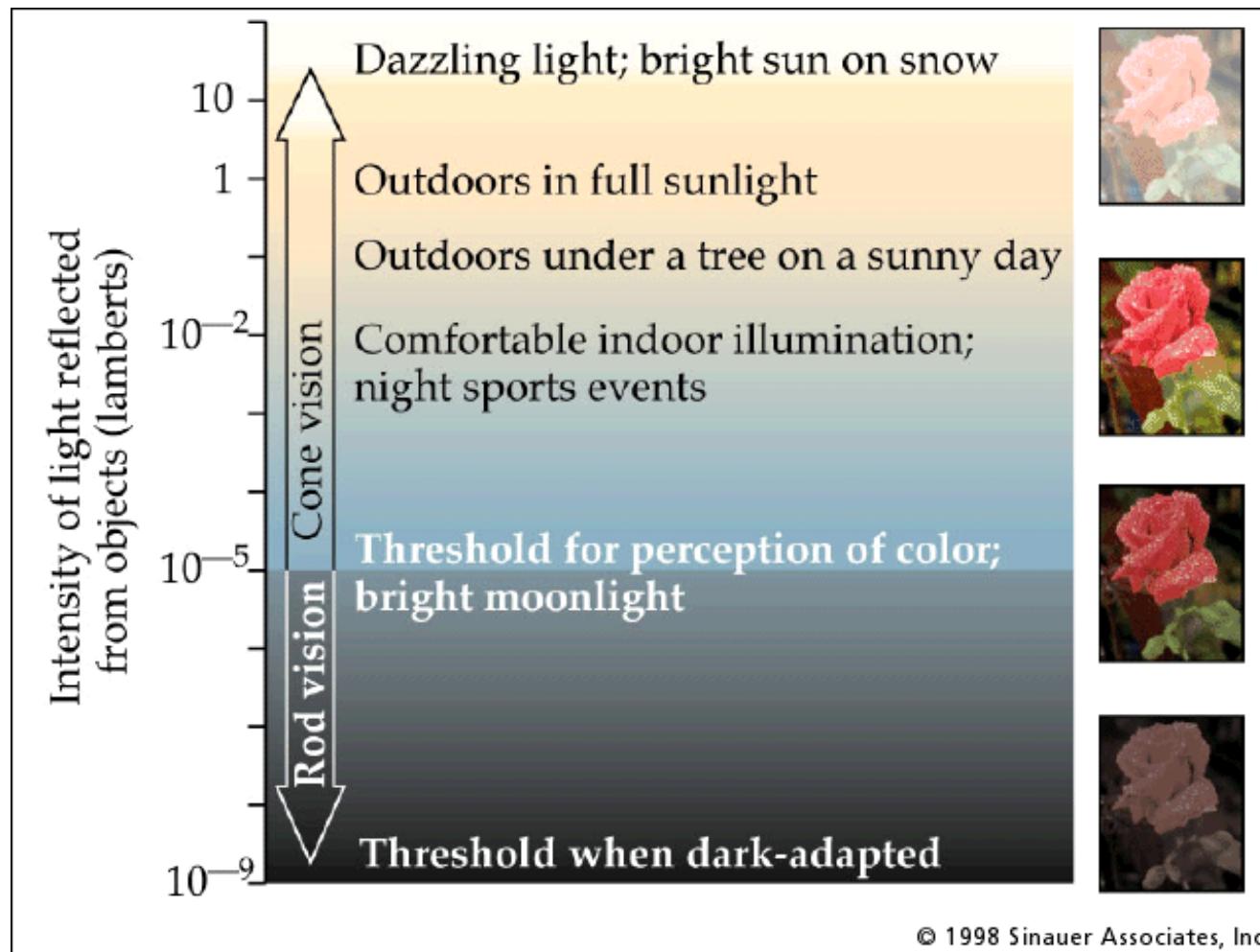
Rods

- tế bào hình que
- độ nhạy cao
- hoạt động trong ánh sáng yếu
- rất ít nhạy cảm với màu sắc



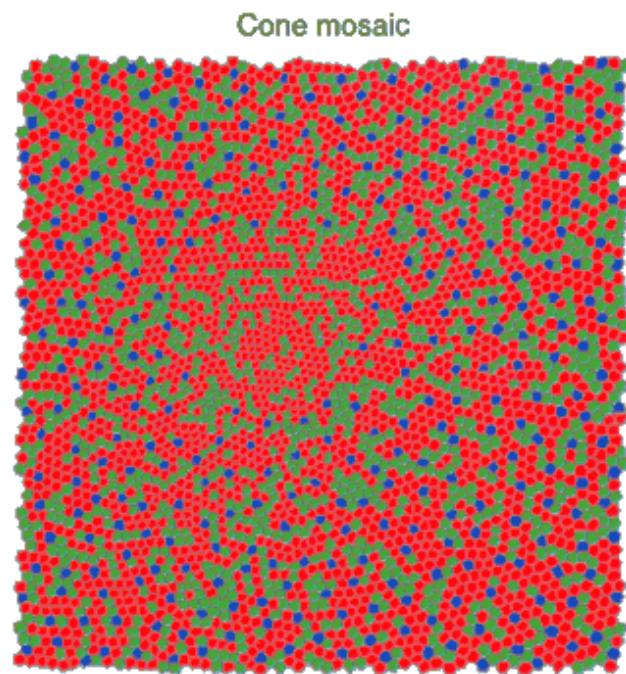
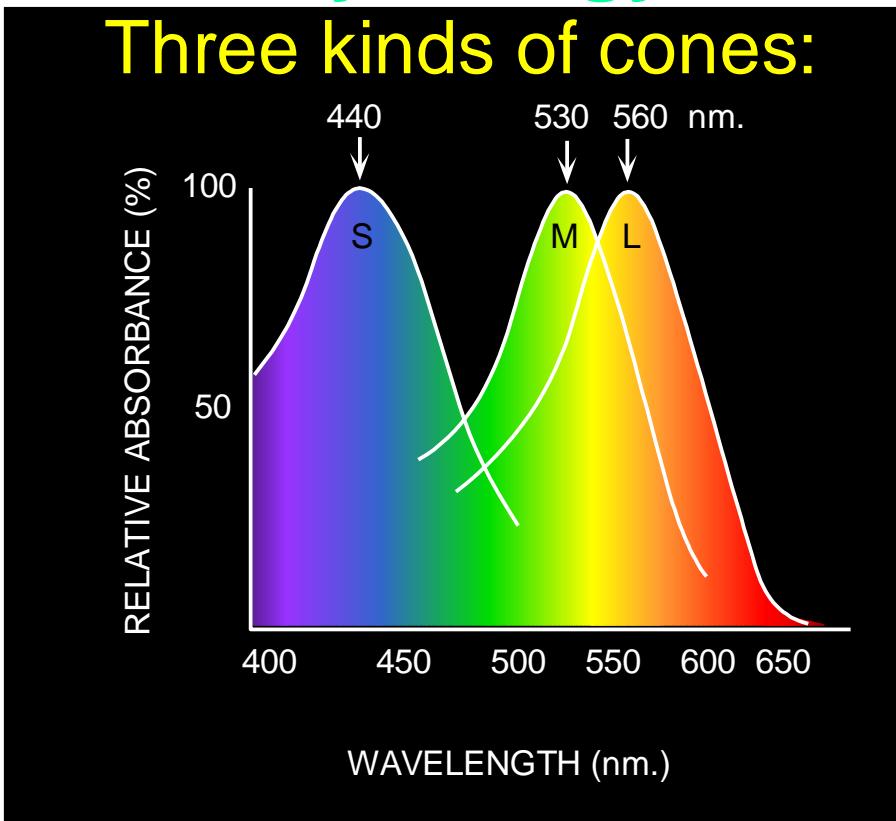
James Hays

Độ nhạy của Rod / Cone



Physiology of Color Vision

Three kinds of cones:



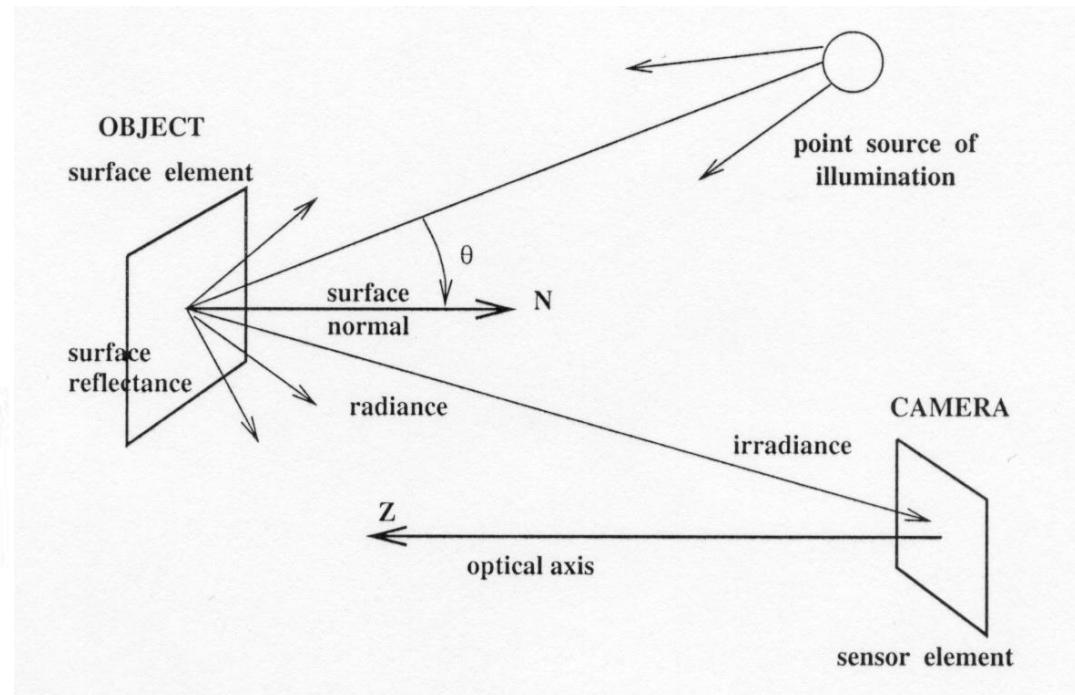
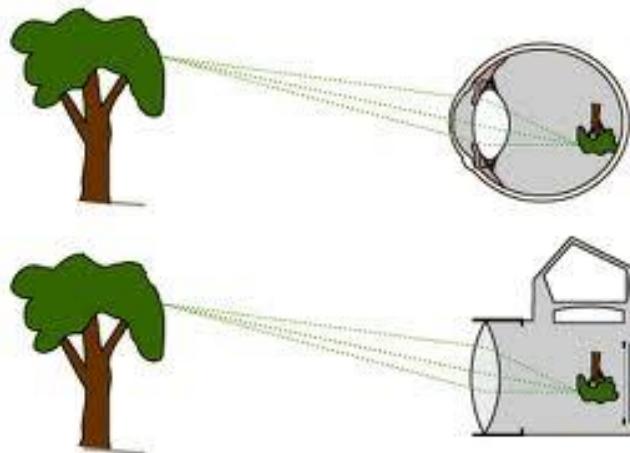
Cones : có 3 loại

- S: nhạy cảm với ánh sáng có bước sóng thấp (~440nm)
- M: nhạy cảm với ánh sáng có bước sóng trung bình (~530 nm)
- L: nhạy cảm với ánh sáng có bước sóng cao (~560 nm)

© Stephen E. Palmer, 2002

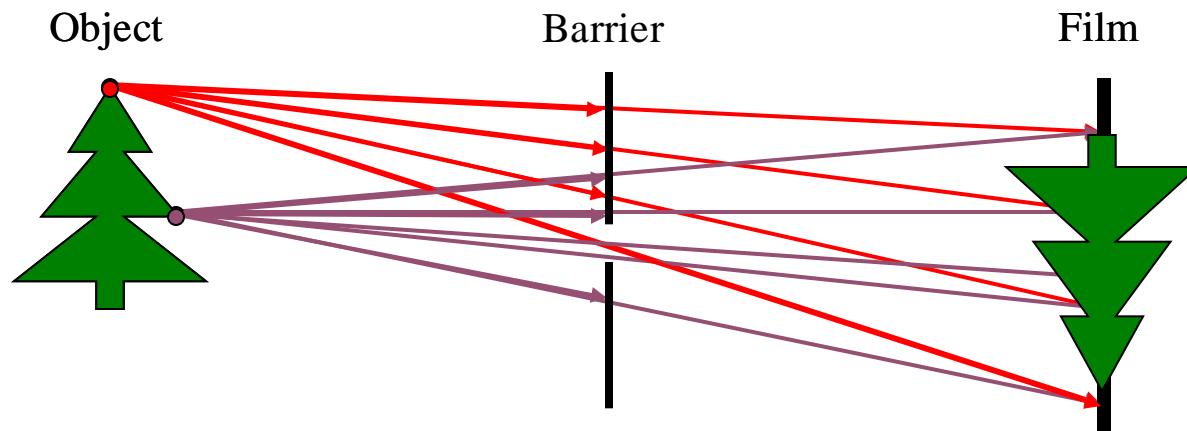
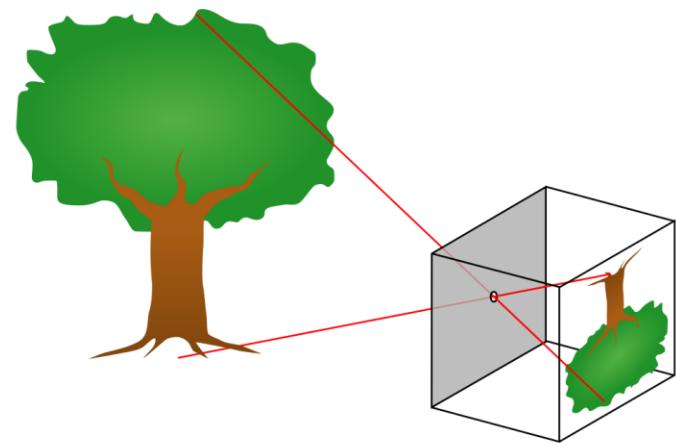
Quá trình hình thành ảnh

- Có sự tương đồng giữa mắt người và camera

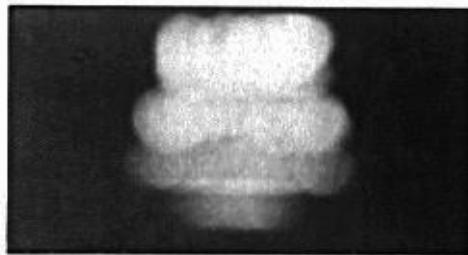


Mô hình pin-hole camera

- Camera without lenses
- Một hộp kín, một mặt có khoét 1 lỗ nhỏ ở giữa
- Độ mở của lỗ thủng (aperture) quyết định độ mờ (blur) của ảnh



Ảnh hưởng của độ mở (aperture size)



2 mm



1 mm



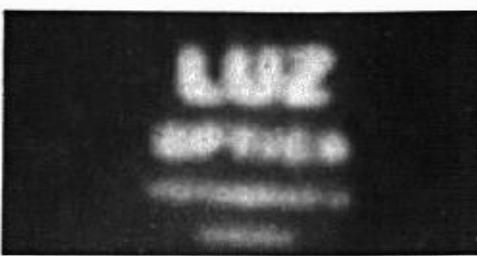
0.6mm



0.35 mm



0.15 mm

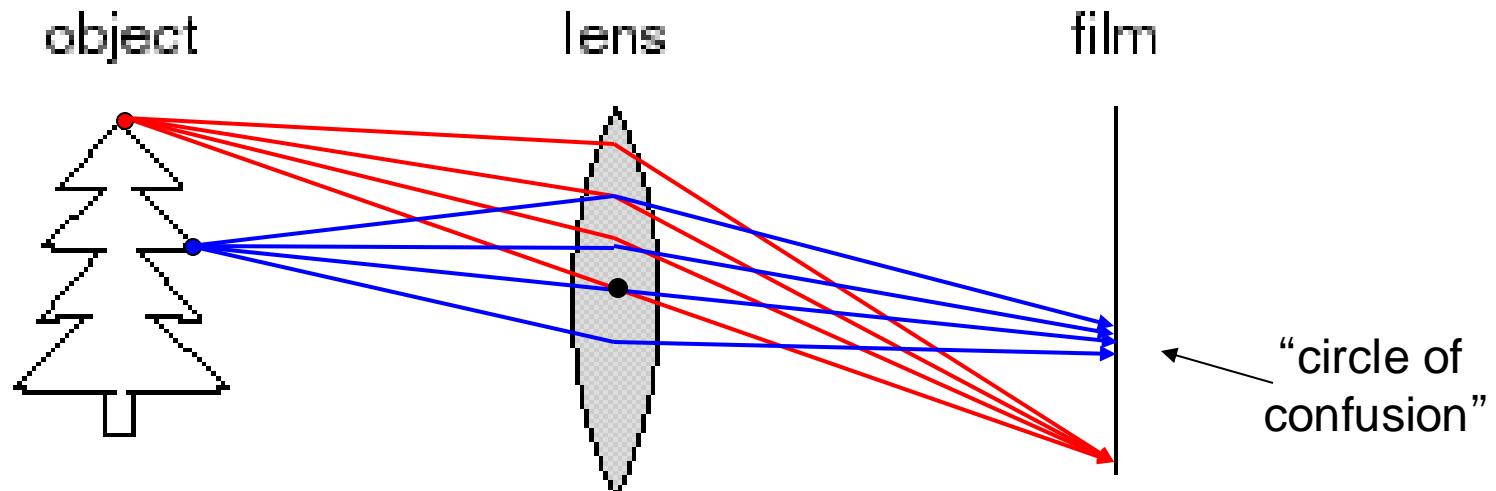


0.07 mm

- Thực tế: kích thước của aperture bị giới hạn do cần nguồn sáng đủ lớn
- Ảnh hưởng của hiệu ứng diffraction

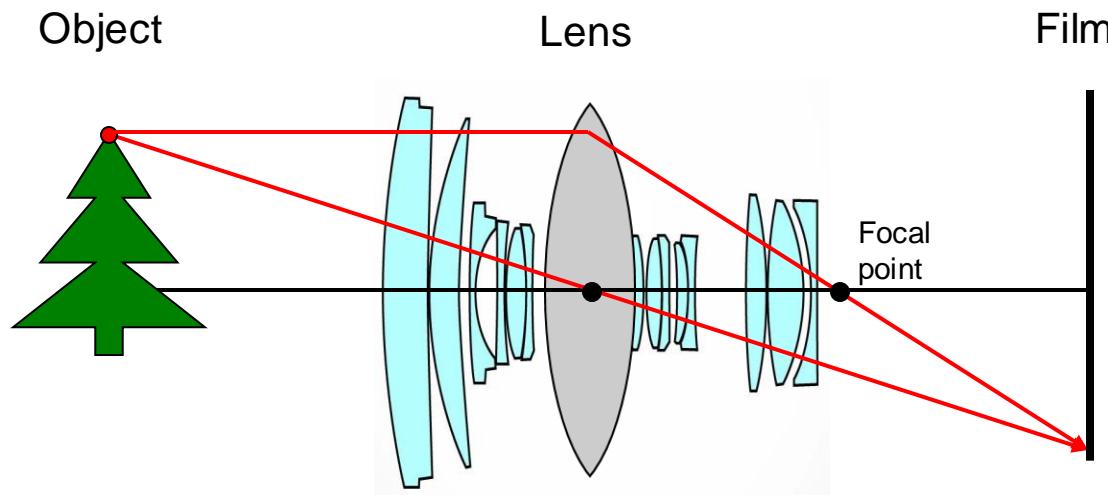
Camera thực tế

- Sử dụng ống kính quang học thay vì tấm barrier
 - Mỗi lens có một khoảng xác định tới film để đối tượng được chụp là “in focus”
 - Có thể ghép nối các lens với nhau



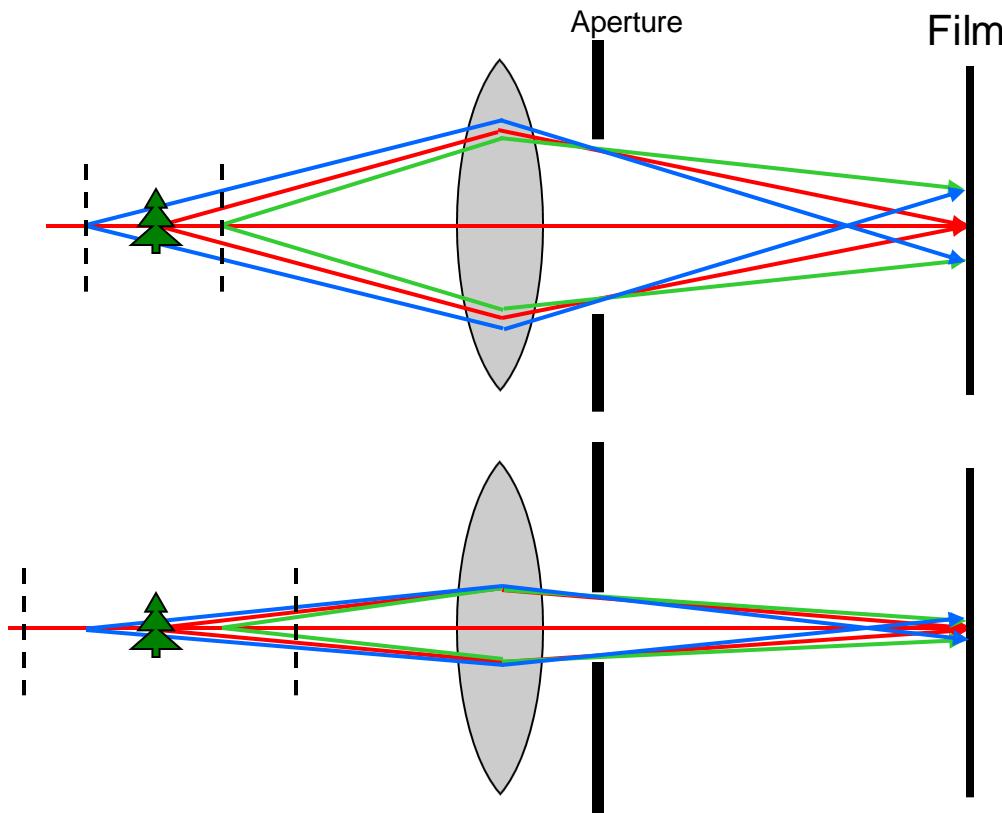
Ghép “thin” lens

Các hệ thống quang học của camera giả định các lens là mỏng (thin lens); không có độ dày



Bằng việc add thêm nhiều thin lens, khoảng cách từ lens đến film là được xác định là “in focus” đối với một scene khi đó nằm trên 1 mặt phẳng.

Độ sâu vùng quan sát (Depth of field - DOF)



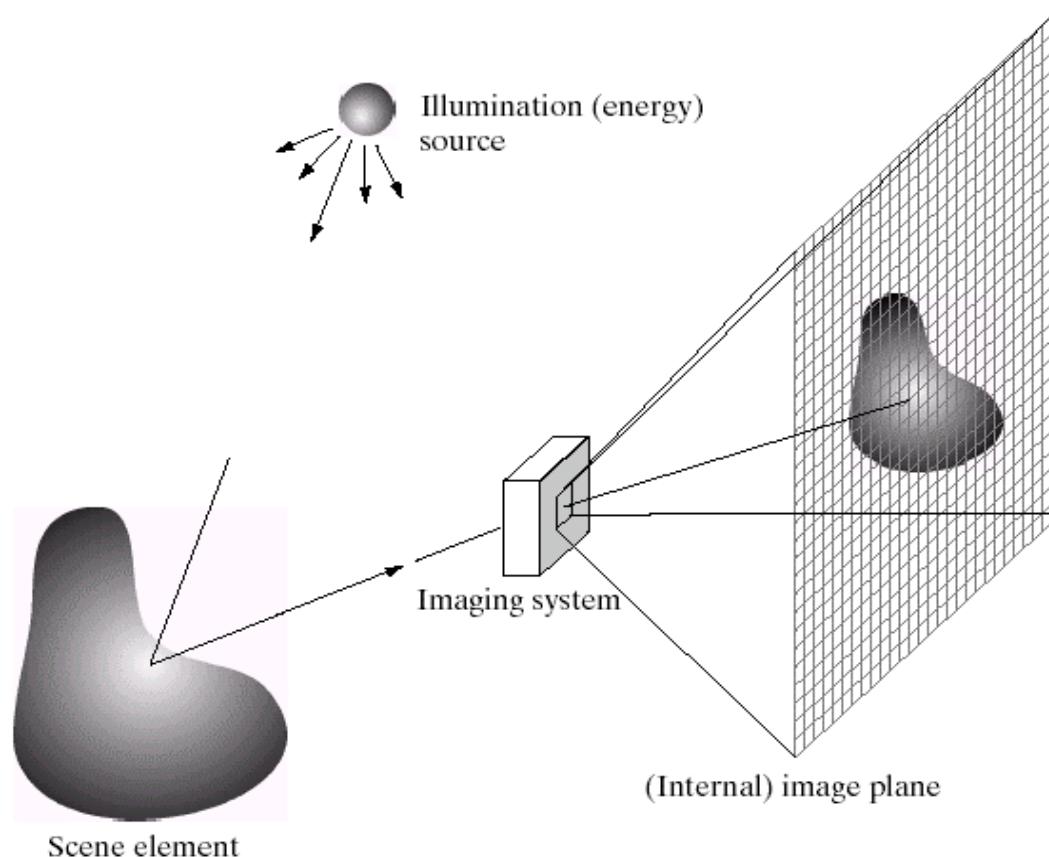
$f/5.6$



$f/32$

- Aperture size dẫn đến vùng “in focus” thay đổi
 - Aperture size tỉ lệ nghịch với khoảng hội tụ (DOF)

Thu nhận ảnh

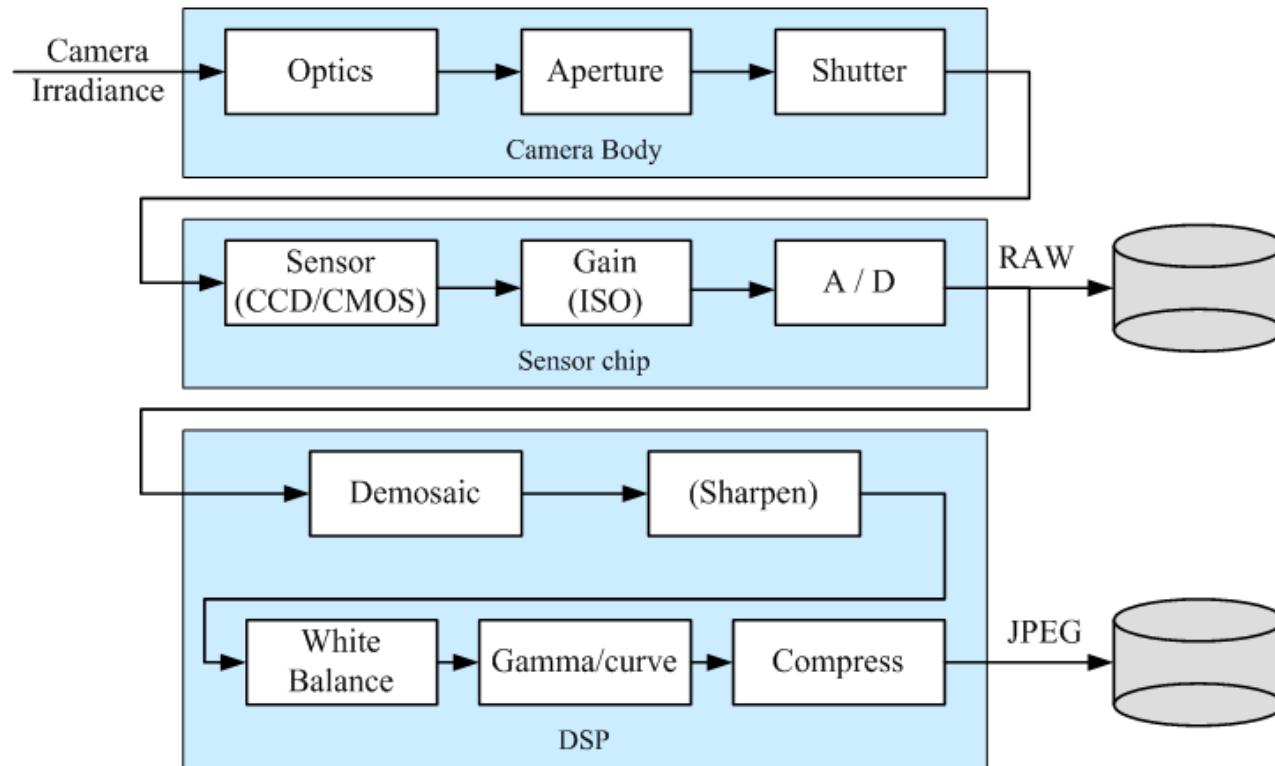


Adapted from S. Seitz

Thu nhận ảnh

- Nguồn sáng
- Đo quang:
 - Đo **độ sáng** cảm nhận được từ **năng lượng điện tử** thấy được của tia sáng
- Hệ thống quang (lenses):
 - Vật được chiếu sáng bởi các tia sáng từ nguồn sáng.
 - Tia sáng tới vật bị phản xạ, phát tán ra xung quanh tùy thuộc vào chất liệu bề mặt của vật.
 - Tia sáng từ vật đi tới hệ thấu kính → tạo ảnh
- Cảm biến hình ảnh: cảm biến **CCD (charge-coupled device)** or **CMOS** cung cấp tín hiệu 2D thu nhận được
- Camera số: Tín hiệu 2D thu được được đưa qua bộ chuyển đổi **analog-to-digital (ADC)** => tạo ảnh số

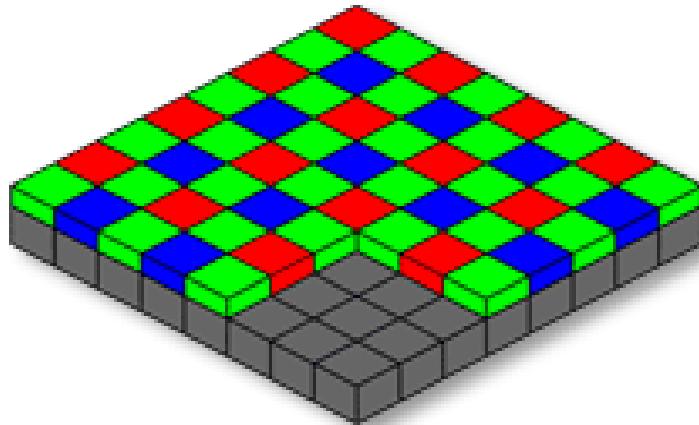
Camera số: thu nhận và số hóa



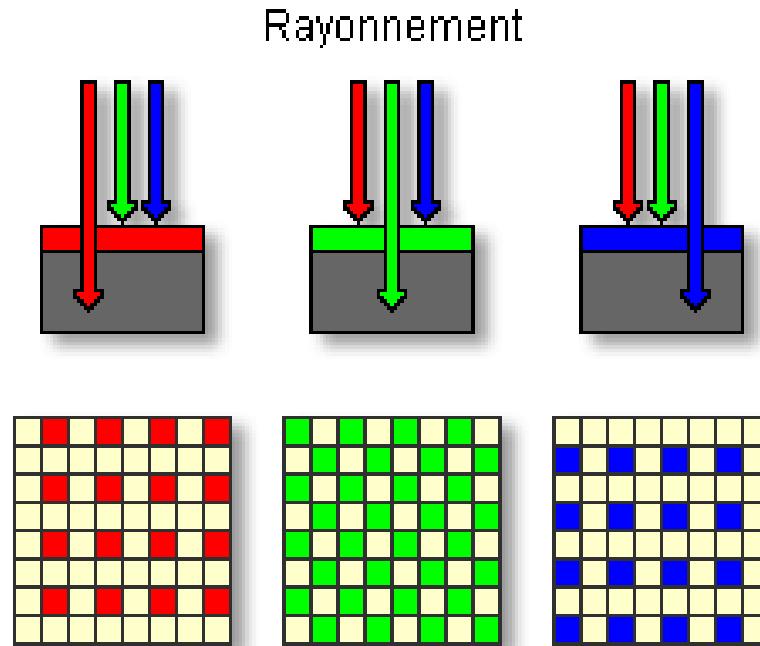
Digital camera: Image sensing and processing pipeline

Adapted from S. Seitz

Cảm biến hình ảnh: ví dụ



Capteur photosensible
recouvert d'une grille de Bayer



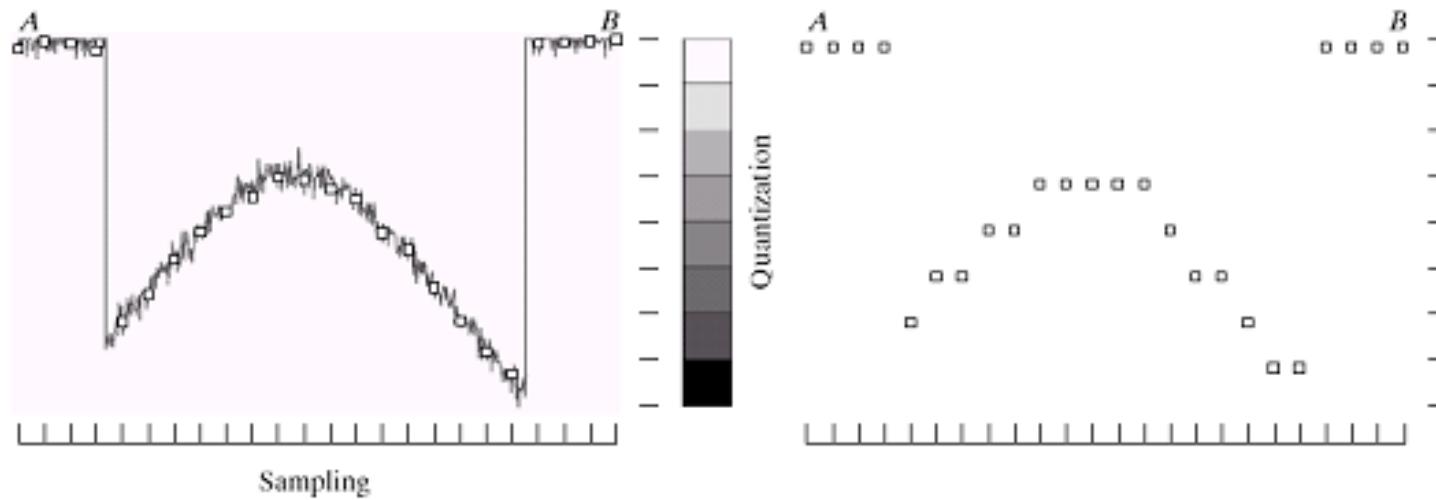
Cảnh thật -> Ảnh số



Digitization = **Sampling (lấy mẫu)**
+ **Quantization (Lượng tử hóa)**

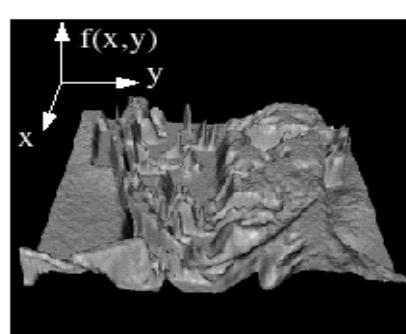
Quá trình số hóa: lấy mẫu và lượng tử hóa

- **Lấy mẫu** đều trên không gian 2D
- **Lượng tử** cho từng điểm được lấy mẫu (làm tròn đến số nguyên gần nhất)



Quá trình số hóa: lấy mẫu và lượng tử hóa

- **Lấy mẫu** đều trên không gian 2D
- **Lượng tử** cho từng điểm được lấy mẫu (làm tròn đến số nguyên gần nhất)

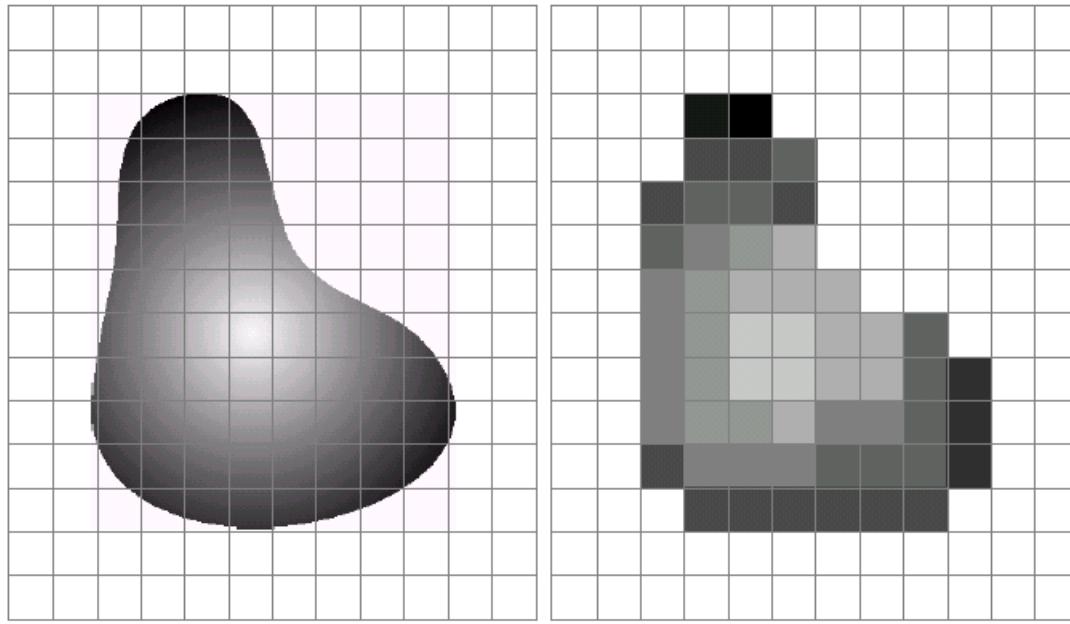


j →
↓ i

62	79	23	119	120	105	4	0
10	10	9	62	12	78	34	0
10	58	197	46	46	0	0	48
176	135	5	188	191	68	0	49
2	1	1	29	26	37	0	77
0	89	144	147	187	102	62	208
255	252	0	166	123	62	0	31
166	63	127	17	1	0	99	30

2D

Digital image



a b

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

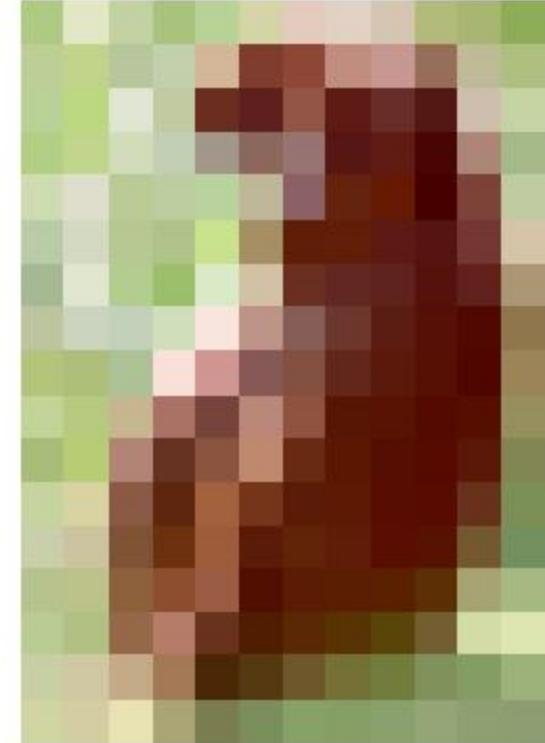
Độ phân giải [không gian] (sampling)



200 X 278



50 X 70



12 X 18

[Độ phân giải] mức xám (Quantization)



8 bits



4 bits

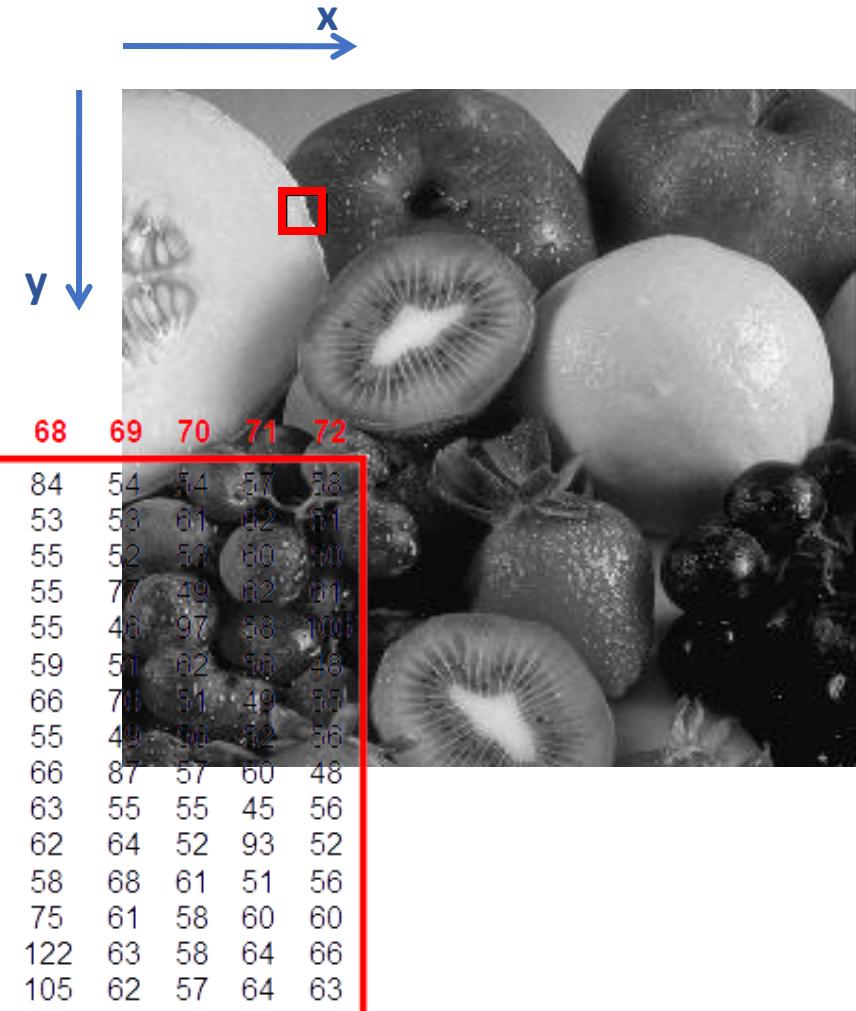


2 bits

Ảnh số

- Ảnh I Nx M: ~ ma trận N x M

- Chỉ số (0,0): góc trên trái
- $I(x,y)$: giá trị điểm ảnh tại (x,y)



Một số loại ảnh phổ biến

- Ảnh nhị phân:

- $I(x,y) \in \{0, 1\}$
- 1 pixel: 1 bit



- Ảnh đa mức xám:

- $I(x,y) \in [0..255]$: mức xám, cường độ sáng
- 1 pixel: 8 bits (1 byte)



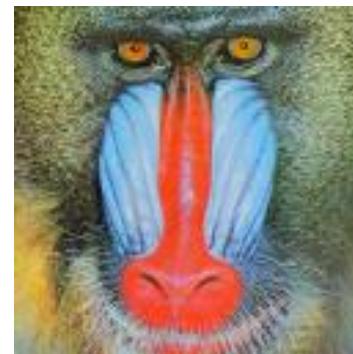
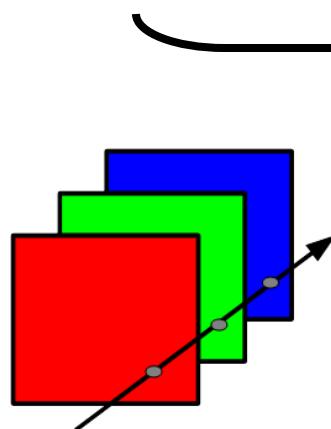
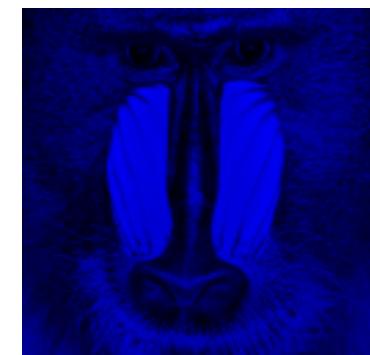
- Ảnh màu:

- $I_R(x,y), I_G(x,y), I_B(x,y) \in [0..255]$
- 1 pixel: 24 bits (3 bytes)



- Khác: ảnh đa phẳng, độ sâu,...

Ảnh màu trong không gian màu RGB



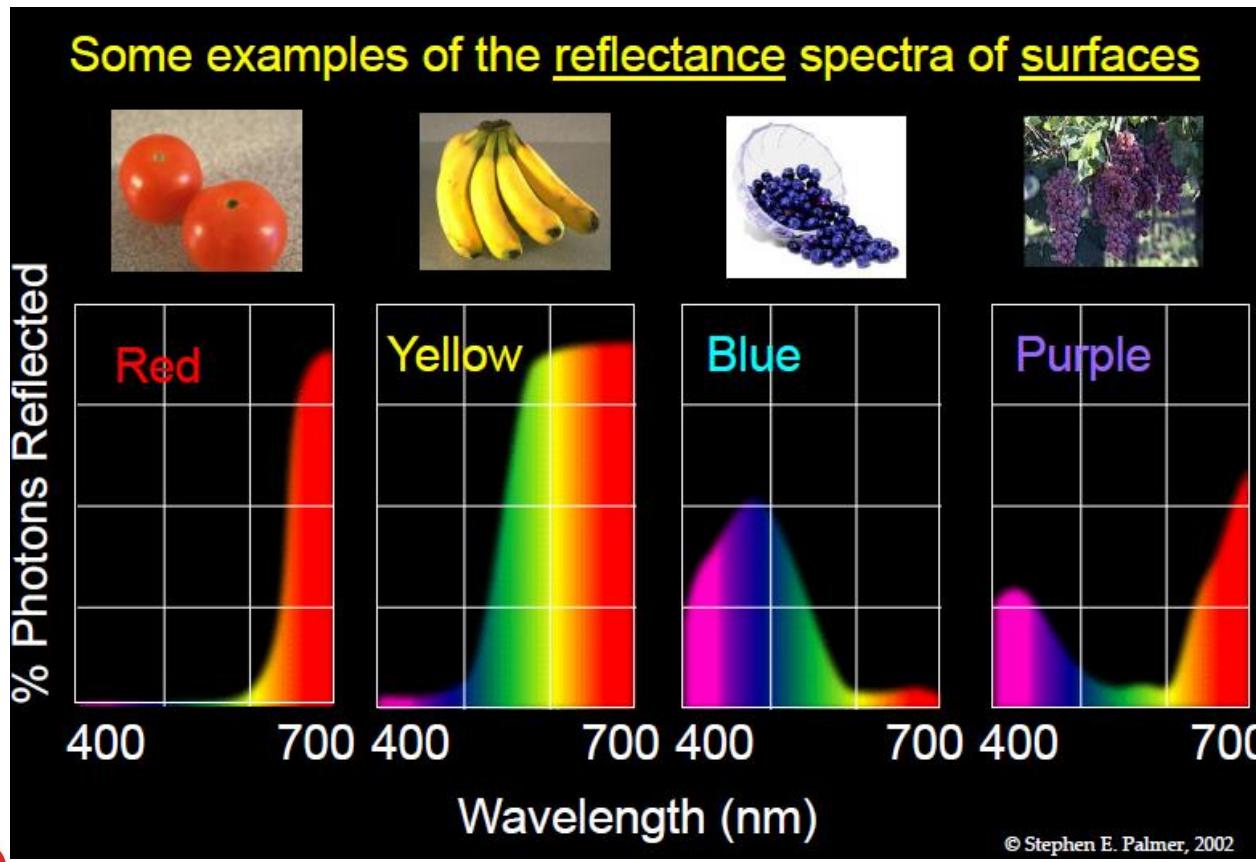
It exists other color spaces:
Lab, HSV, ...

Nội dung

- Giới thiệu chung
 - Khái niệm
 - Các cấp độ xử lý (Low level vision, Middle level vision, High level vision)
 - Lĩnh vực liên quan
 - Ứng dụng
- Quá trình hình thành và thu nhận ảnh
- Không gian màu
- Thực hành

Màu sắc trong ảnh

- Màu nhận được do đáp ứng của bề mặt với chùm sáng chiếu tới

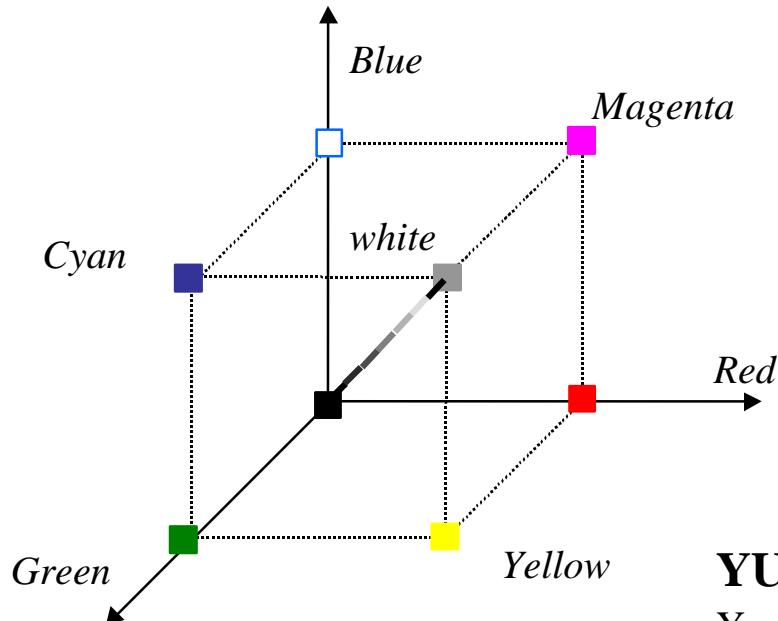


Không gian màu

- Định nghĩa:
 - Là không gian gồm các thành phần màu (primary color) được kết hợp để biểu diễn màu sắc
- Các không gian màu thường sử dụng:
 - RGB, HSV, CMY (Cyan, Magenta, Yellow), Lab, Luv...
- Tại sao có nhiều hơn một không gian màu (?)
 - Do đặc thù của ứng dụng (in ấn màu, hiển thị (monitor) màu)
 - Một số không gian màu độc lập hoặc phụ thuộc thiết bị
 - Một số không gian màu tuyến tính, một số khác không tuyến tính với cảm nhận của mắt người
 - etc

Không gian màu

M μ s³/c



Y Cb Cr (JPEG)

$$Cb = U/2 + 0.5$$

$$Cr = V/1.6 + 0.5$$

RGB

Black	(0, 0, 0)
Red	(255, 0, 0)
Green	(0, 255, 0)
Yellow	(255, 255, 0)
Blue	(0, 0, 255)
Magenta	(255, 0, 255)
Cyan	(0, 255, 255)
White	(255, 255, 255)

YUV (Luminance)

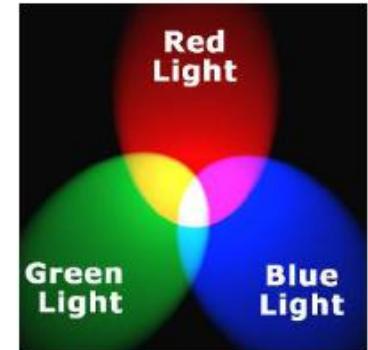
$$Y = 0.299R + 0.587G + 0.114B$$

$$U = R - Y$$

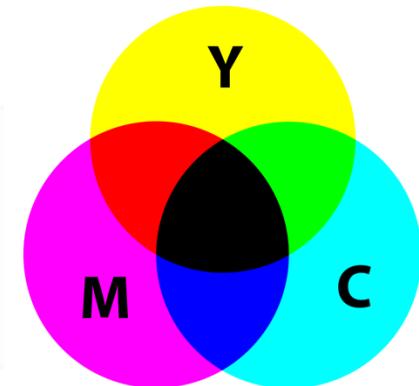
$$V = B - Y$$

Không gian màu RGB và CMY

- RGB (Red – Green – Blue):
 - phổ biến, được dùng trong các thiết bị hiển thị
 - hệ thống phối màu cộng : chồng 3 kênh R, G, B → tạo nên một màu
- CMY (Cyan-Magenta-Yellow):
 - Thường dùng trong in ấn, photo
 - Hệ thống không gian màu trừ: ánh sáng đi qua các bộ lọc C,M,Y để tạo màu
- Đặc điểm:
 - Không độc lập với thiết bị
 - Không gian màu RGB thường **không tuyến tính** với việc cảm nhận màu của mắt người

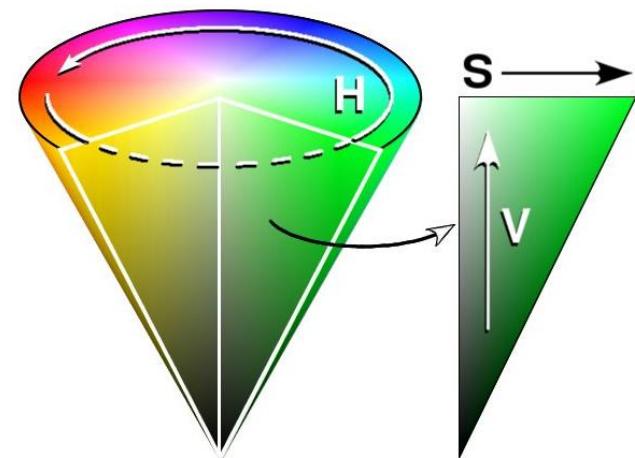


$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$



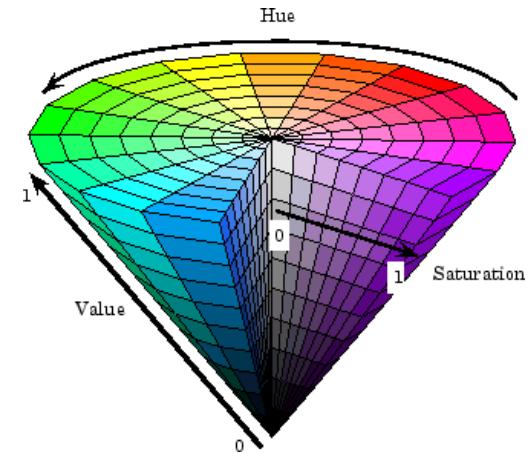
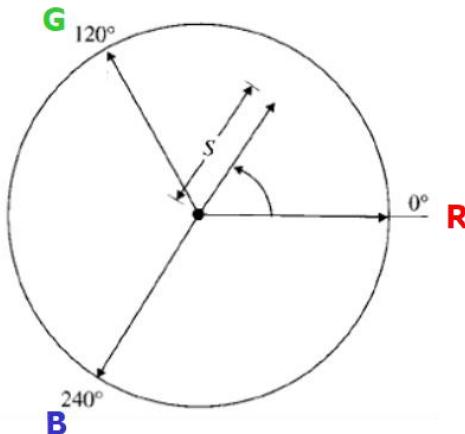
HSV (Hue – Saturation- Value)

- Hue-Saturation-Value / Lightness
 - Sắp xếp lại màu dễ hình dung hơn
 - Phù hợp cảm nhận màu của thị giác người
 - Hay dùng cho phân vùng/nhận dạng
- Giá trị tại 1 điểm ảnh:
 - **Màu sắc** (H + S) (chroma):
 - H: sắc tố, tông màu
 - S: độ bão hòa, sắc độ
 - **Cường độ sáng** (V)
- RGB không có sự phân tách này



HSV

- Hue: $0 \rightarrow 360$
- Saturation: $0 \rightarrow 1$
- Value: $0 \rightarrow 255$



- colour cone

- $H = \text{hue} / \text{colour in degrees} \in [0, 360]$
- $S = \text{saturation} \in [0, 1]$
- $V = \text{value} \in [0, 1]$

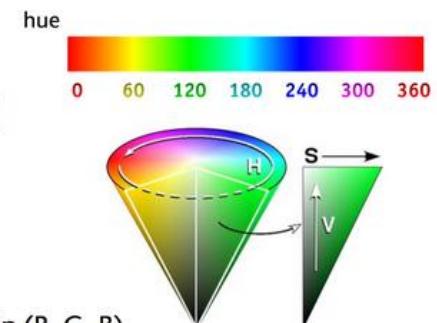
- conversion RGB → HSV

- $V = \max = \max(R, G, B), \quad \min = \min(R, G, B)$

- $S = (\max - \min) / \max \quad (\text{or } S = 0, \text{ if } V = 0)$

- $H = 60 \times \begin{cases} 0 + (G - B) / (\max - \min), & \text{if } \max = R \\ 2 + (B - R) / (\max - \min), & \text{if } \max = G \\ 4 + (R - G) / (\max - \min), & \text{if } \max = B \end{cases}$

$$H = H + 360, \text{ if } H < 0$$

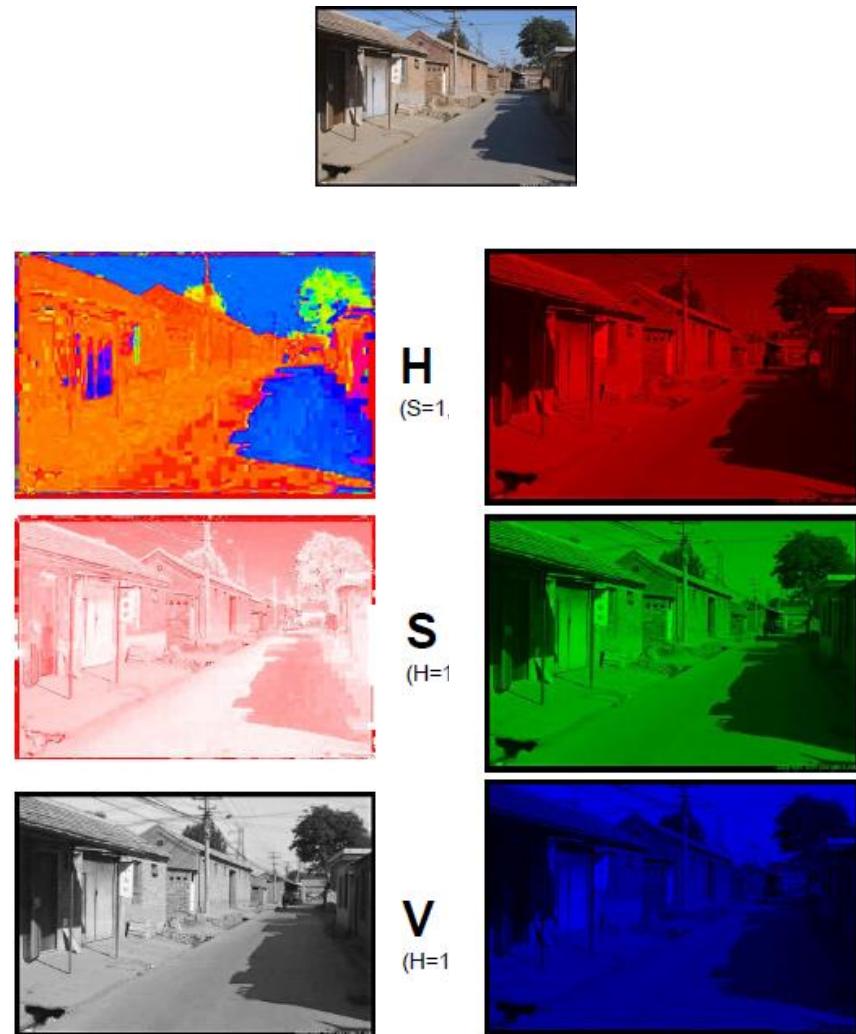
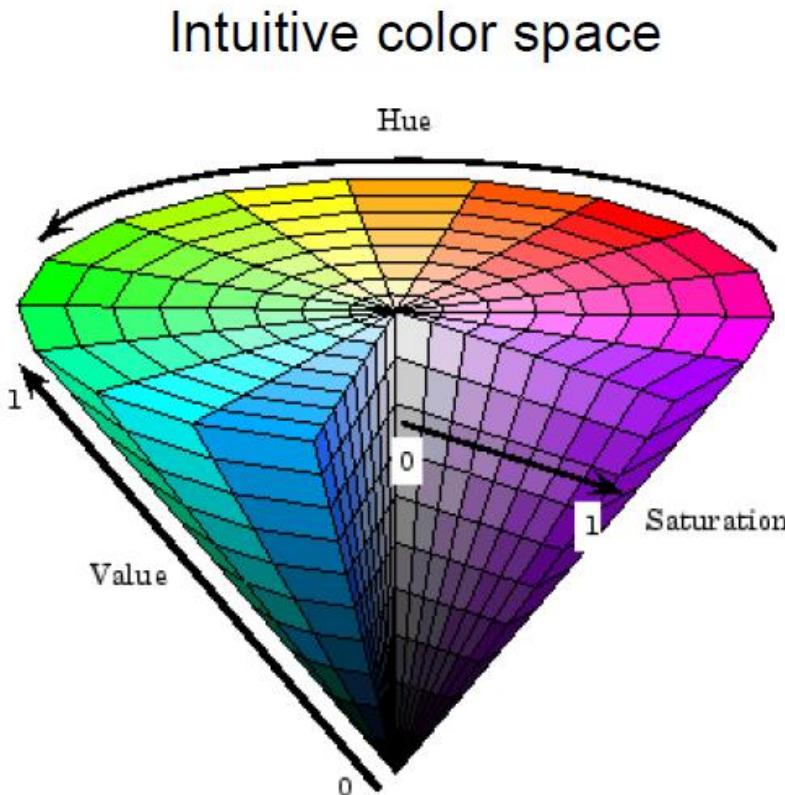


HSV

- Lưu ý:
 - màu của đối tượng có thể biểu diễn bằng **1 khoảng trên Hue**
 - Hue < 60° không có nghĩa!
 - Định nghĩa khoảng : $40^\circ < \text{Hue} < 80^\circ$
 - Khoảng này chỉ có nghĩa nếu Saturation > threshold (nếu không là xám)
 - Màu object độc lập với gáy **Value**, V: nhạy cảm với điều kiện chiếu sáng

Ví dụ minh họa

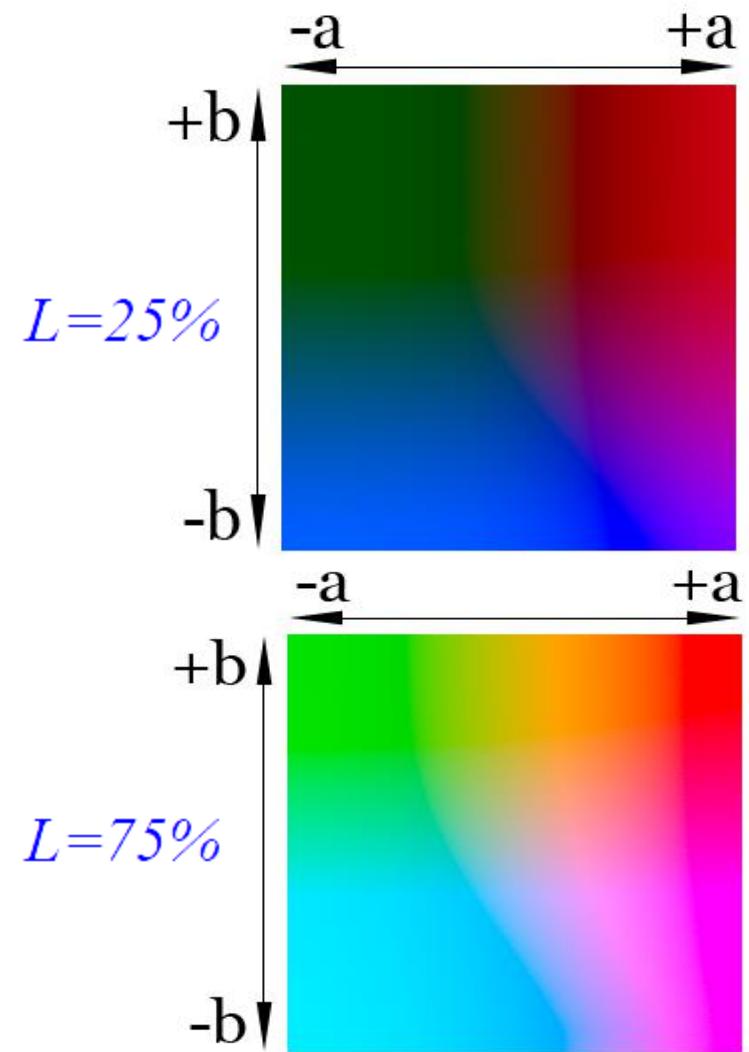
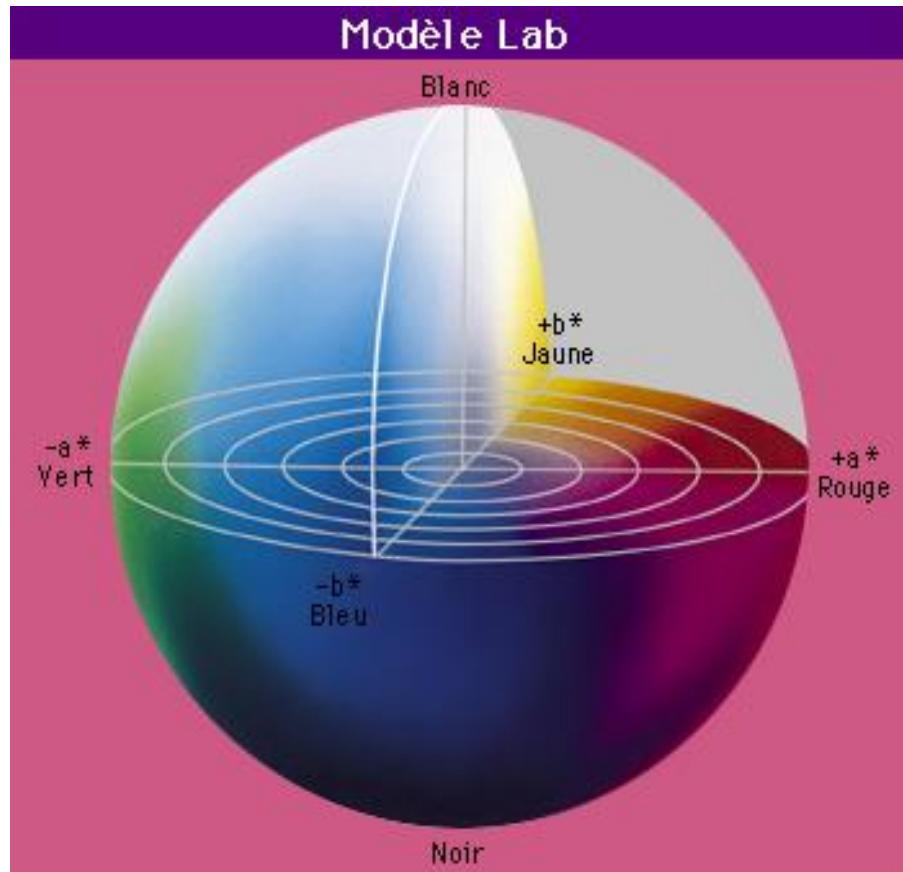
Color spaces: HSV



Không gian màu CIE Luv/Lab

- **Lab** (còn gọi $L^*a^*b^*$) và Luv được nghiên cứu dựa trên thị giác người
 - Độc lập thiết bị
 - Tuyến tính với cảm nhận màu của mắt người
- Thành phần:
 - **L** : độ sáng, 0% (black) → 100% (white)
 - **a***: trục biểu diễn màu green (negative value, -127) đến màu red (positive value, +127)
 - **b***: trục biểu diễn từ blue (negative value, -127) đến yellow (positive value, +127)

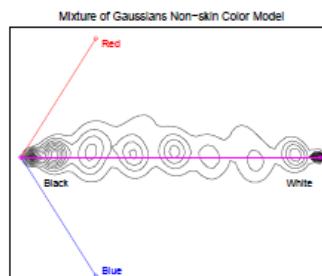
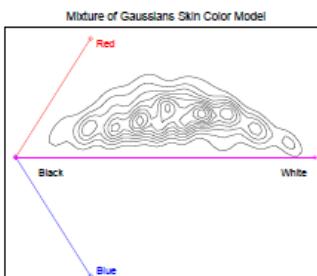
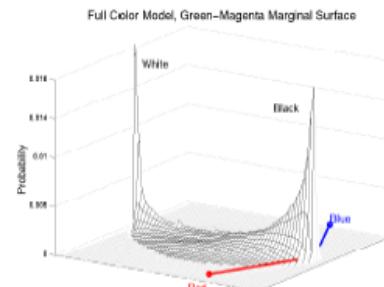
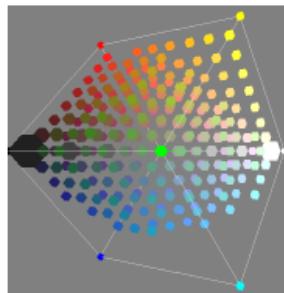
Lab



Ứng dụng không gian màu

- RGB không phân tách Chrominance và intensity (luminance)

$$P(rgb|skin) = \frac{s[rgb]}{T_s}, \quad P(rgb|\neg skin) = \frac{n[rgb]}{T_n}$$



$Cr > 150 \ \&\& Cr < 200 \ \&\& Cb > 100 \ \&\& Cb < 150.$

[Shaik_ICRTC-2015]

Không gian màu vs thay đổi điều kiện chiếu sáng

- collected 10 images of the cube under varying illumination conditions
- separately cropped every color to get 6 datasets for the 6 different colors



Changes in color due to varying Illumination conditions

- Compute the density plot: Check the distribution of a particular color say, blue or yellow in different color spaces. The density plot or the 2D Histogram gives an idea about the variations in values for a given color

Source: Vikas Gupta, Learn OpenCV

Không gian màu vs thay đổi điều kiện chiếu sáng

- Điều kiện chiếu sáng tương đồng: very compact

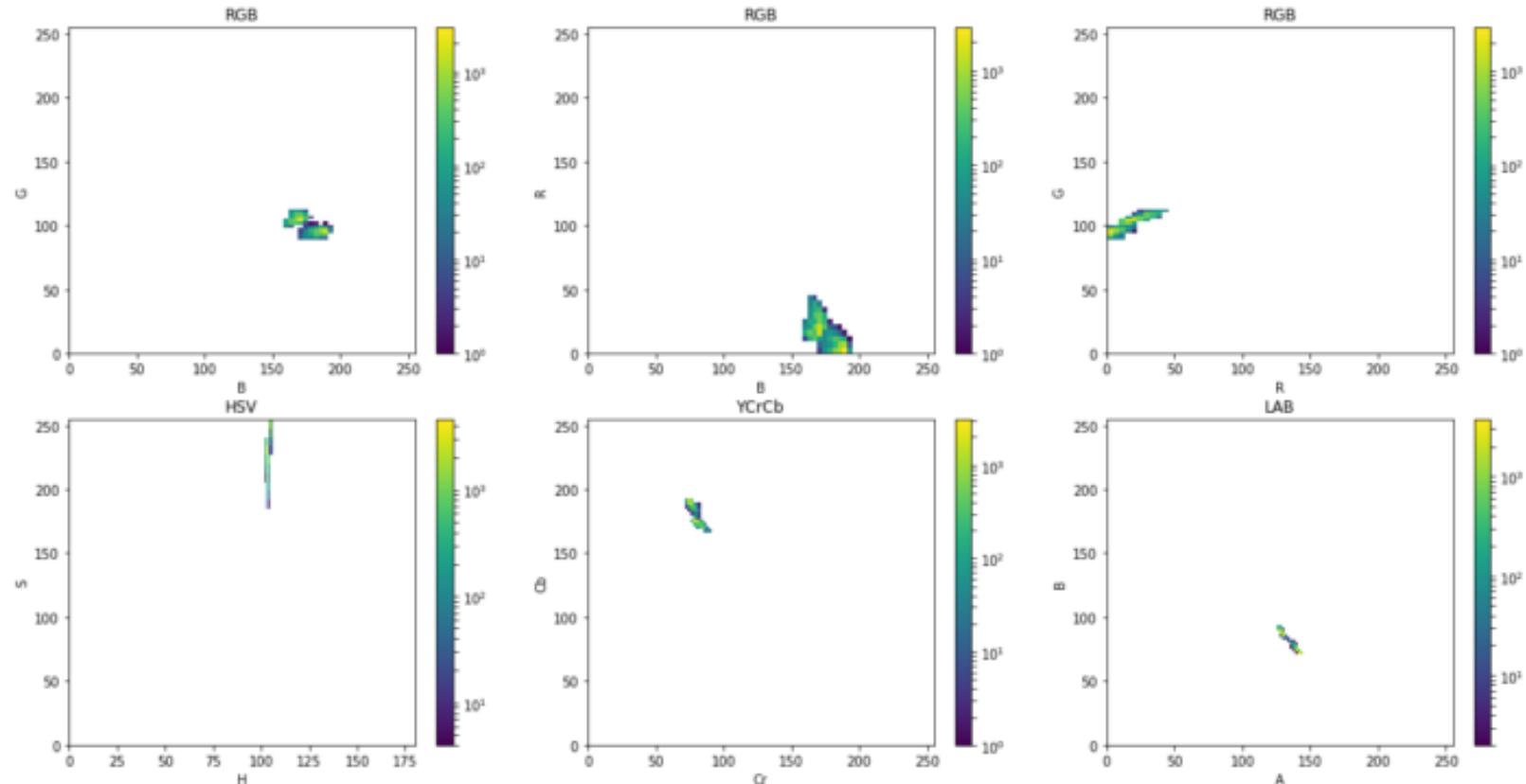


Fig.: Density Plot showing the variation of values in color channels for 2 similar bright images of **blue color**

Source: Vikas Gupta, Learn OpenCV

Không gian màu vs thay đổi điều kiện chiếu sáng

- Điều kiện chiếu sáng tương đồng: very compact

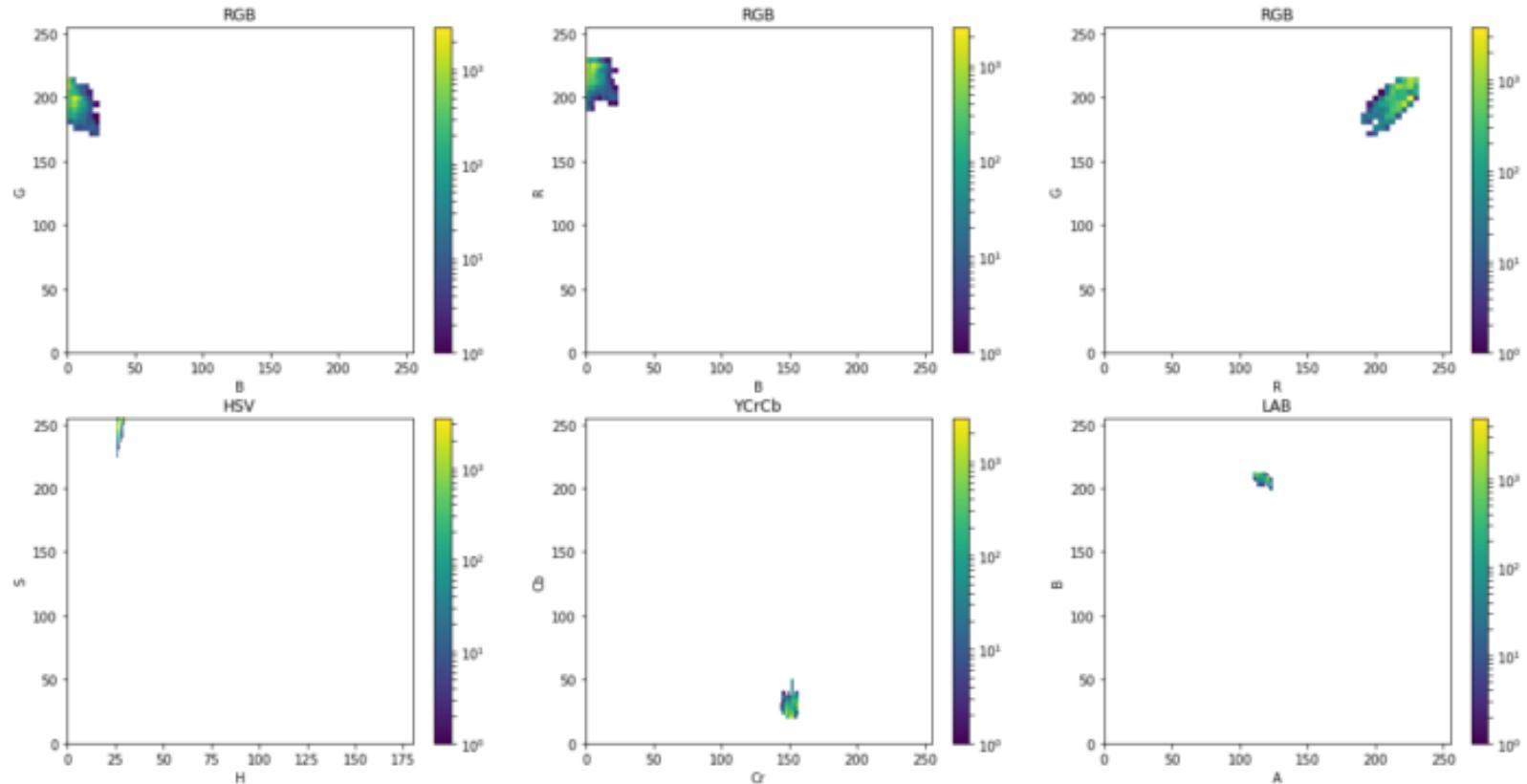


Fig.: Density Plot showing the variation of values in color channels for 2 similar bright images of **yellow color**

Source: Vikas Gupta, Learn OpenCV

Không gian màu vs thay đổi điều kiện chiếu sáng

- Điều kiện chiếu sáng khác nhau:

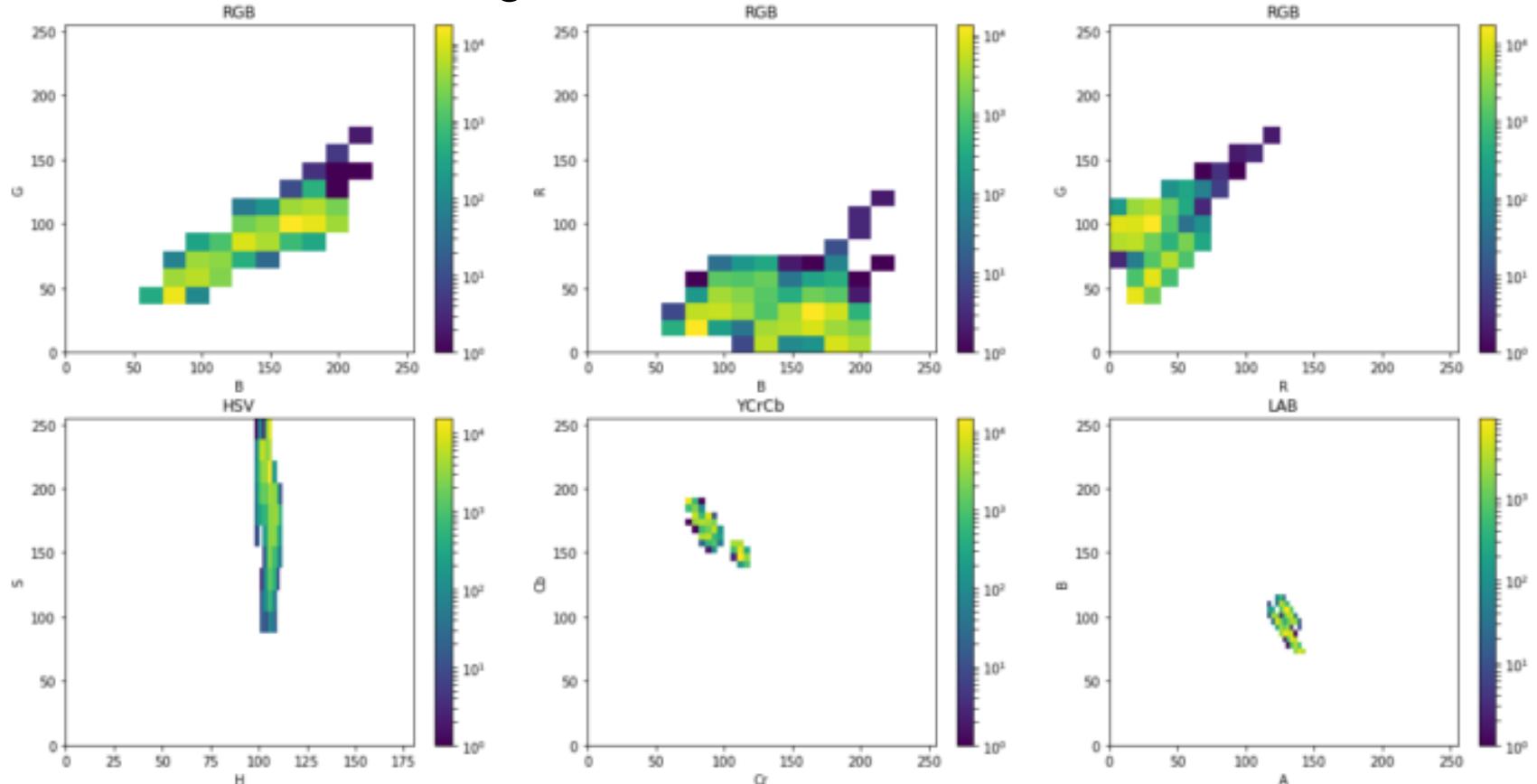


Fig.: Density Plot showing the variation of values in color channels under varying illumination for the **blue color**

Source: Vikas Gupta, Learn OpenCV

Không gian màu vs thay đổi điều kiện chiếu sáng

- Điều kiện chiếu sáng khác nhau

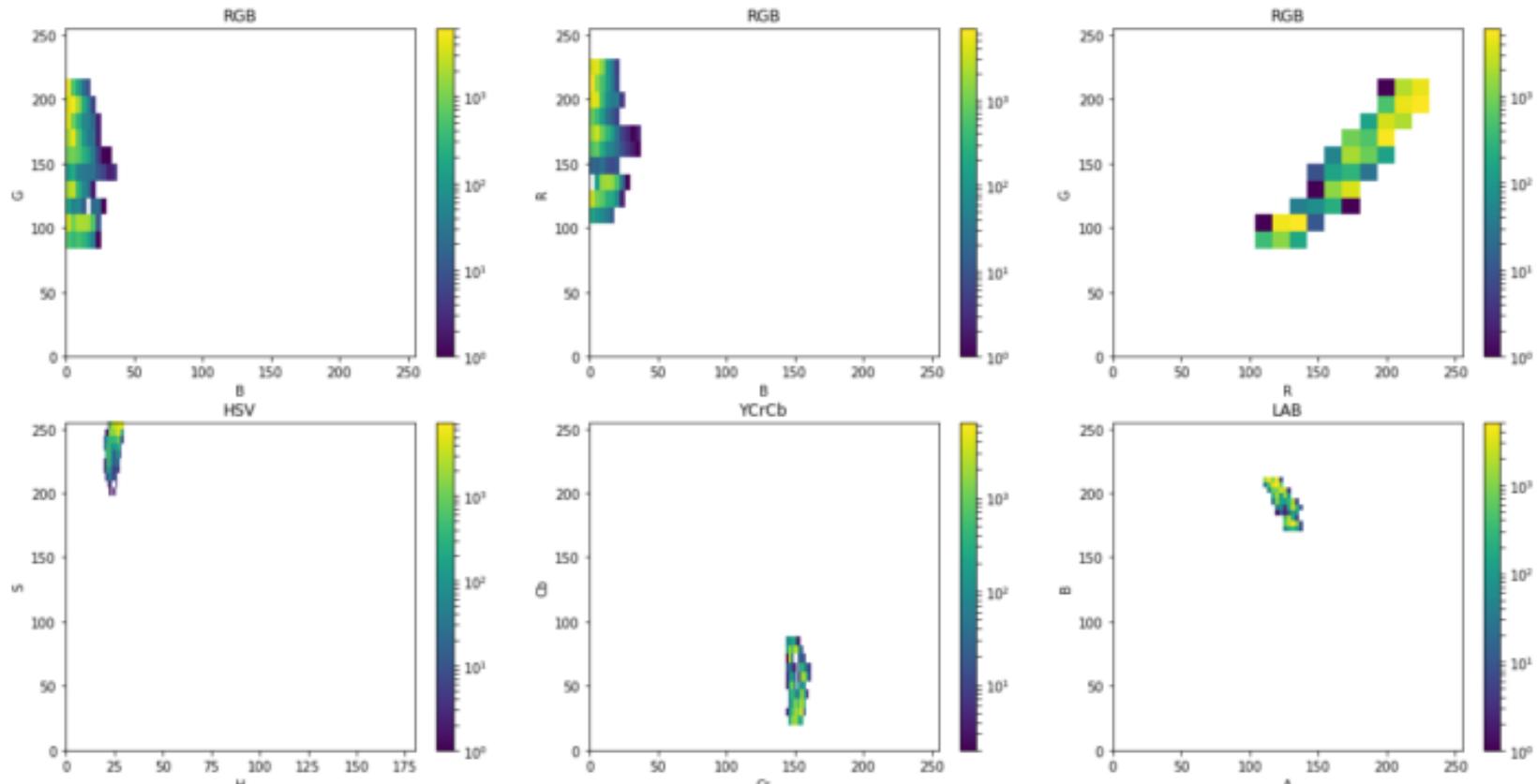


Fig.: Density Plot showing the variation of values in color channels under varying illumination for the **yellow color**

Source: Vikas Gupta, Learn OpenCV

Không gian màu vs thay đổi điều kiện chiếu sáng

- Điều kiện chiếu sáng khác nhau, cùng màu:
 - RGB: biến thiên giá trị ở các kênh lớn
 - HSV: giá trị tập trung (compact) trên kênh **H**. Chỉ kênh H chứa trị tuyệt đối về màu → 1 lựa chọn k tồi
 - YCrCb, LAB: compact trên kênh **CrCb** và trên kênh **AB**
 - Mức độ tập trung trên LAB tốt hơn
- Chuyển đổi giữa các không gian màu (OpenCV):
 - cvtColor(bgr, ycb, COLOR_BGR2YCrCb);
 - cvtColor(bgr, hsv, COLOR_BGR2HSV);
 - cvtColor(bgr, lab, COLOR_BGR2Lab);

Image histogram

- Histogram (lược đồ xám / màu)
 - Lược đồ trên ảnh xám có giá trị nằm trong dải $[0, L-1]$ là một hàm rời rạc: $h(r_k) = n_k$
 - r_k là mức xám thứ k
 - n_k : số điểm ảnh có mức xám thứ k r_k
 - Là biểu diễn phân bố mức xám/màu sắc của các điểm ảnh trên ảnh

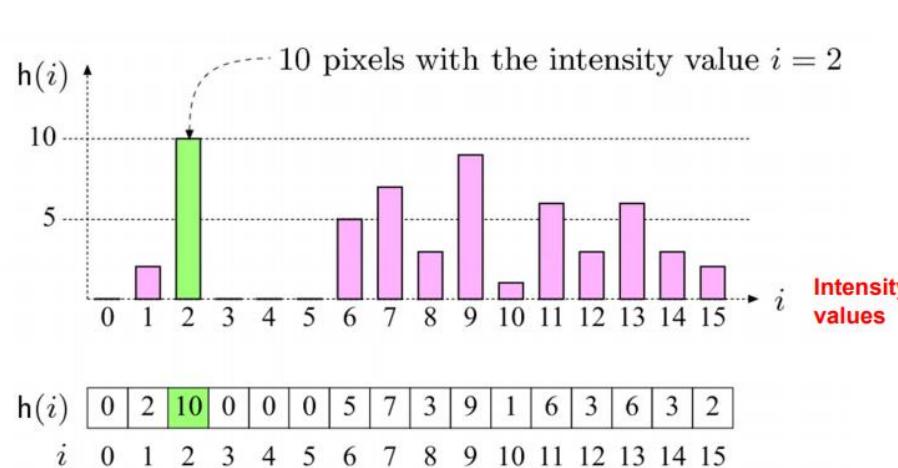
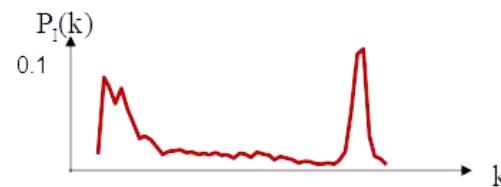
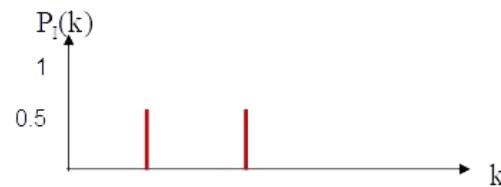


Image histogram

- Histogram thường được chuẩn hóa

– Chia cho tổng số điểm ảnh trong ảnh (n) : $h(r_k) = \frac{n_k}{n}$



Dải động = [min intensity, max intensity]



25
YEARS ANNIVERSARY
SOICT

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

**Thank you for
your attention!**

