CAP 4453
Robot Vision

Dr. Gonzalo Vaca-Castaño

gonzalo.vacacastano@ucf.edu

Course Website + resources

- CANVAS (webcourses)
- Alternative: CAP 4453 - Spring 2022 (ucf.edu)

- Colab notes:gonzo1978/CAP4453: Colab notes for CAP 4453 (github.com)

- Szeliski, Computer Vision: Algorithms and Applications

- Python for Computer Vision. A tutorial will be given in the class on PyTorch for deep learning.

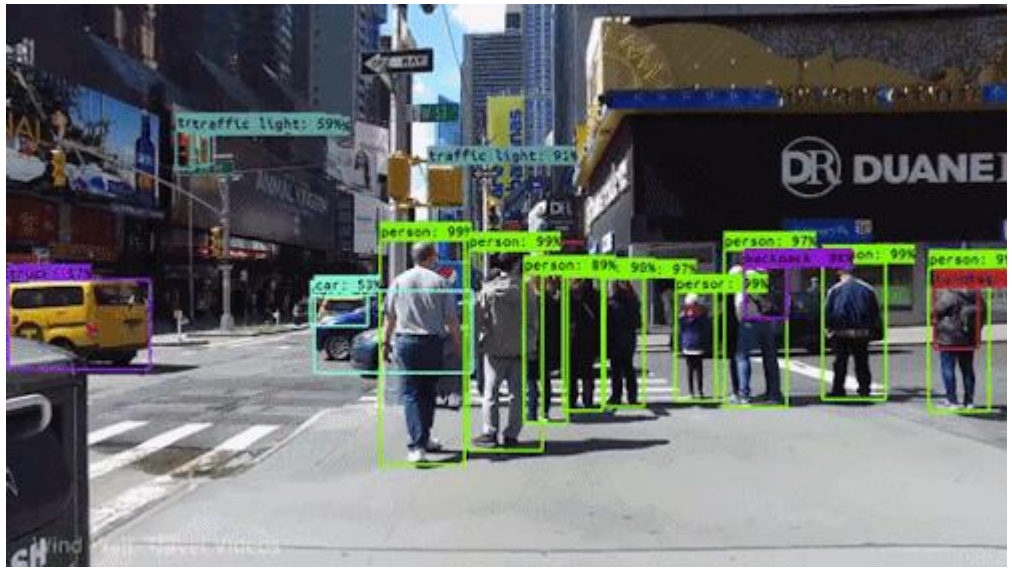
Course logistics

- Class time: Tuesday and Thursday 6:00pm – 7:15pm
    - Classroom: HEC O118

- Office hours [Zoom]
    - Wednesday 7:00pm-8:00 pm
    - By appointment (send me an email)

Course Outline

- Image Filtering
- Edge Detection
- Feature Extraction
- Image warping
- Optical Flow
- Basics of Neural Networks
- Deep Learning for Computer Vision
- Image Segmentation
- Image Classification
- Object Detection



Grading

- Weekly homework: 25%
- Mid term exam: 25%
- Final exam: 30%
- Programming project 1: 10%
- Programming project 2: 10%

Scores

- 95-100 = A
- 90-94 = A-
- 85-89 = B+
- 80-84 = B
- 75-79 = B-
- 70-74 = C+
- 65-69 = C
- 60-64 = C-
- 55-59 = D+
- 50-54 = D
- 45-50 = D-
- 0-44 = F

Grading criteria

- Programming: homework must be written in python. They should be written as colab notes. They should include explanatory/clear comments as well as a short report describing the approach, detailed analysis, and discussion/conclusion. Optionally, you could add GUI to your projects.
- Collaboration: Students are free to discuss ideas and technical concepts. However, students must submit original work for all assignments, projects and exams, and abide by UCF Golden Rule. Cheaters will not be tolerated.

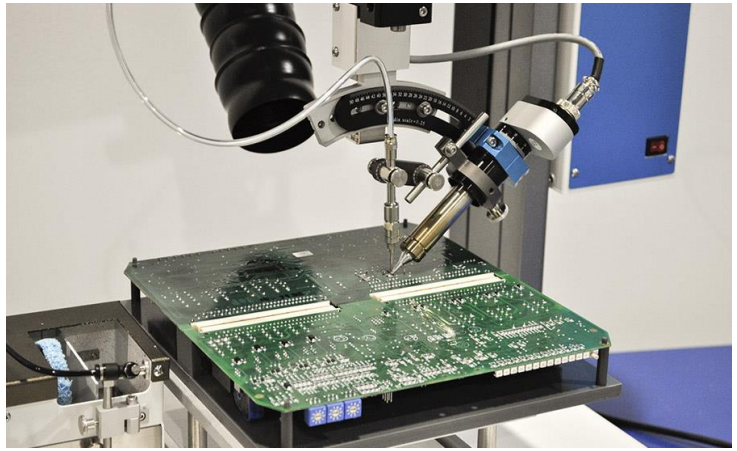
Questions?

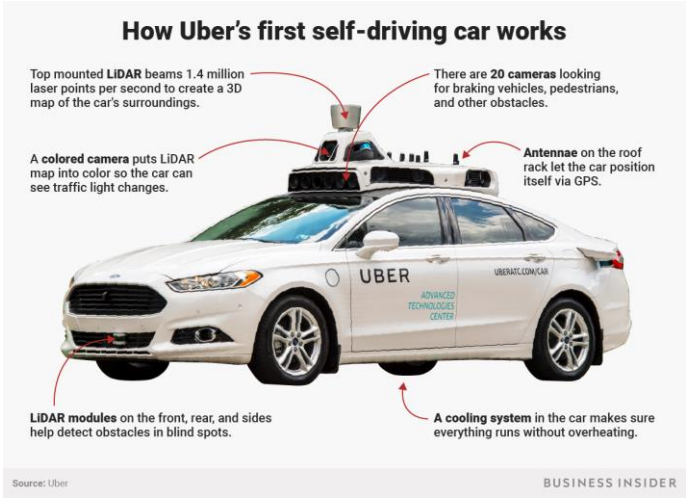
Robot Vision

I. Introduction

Robot vision

- Ability of robots to understand visual data from the world using:
  - Hardware: like 2D cameras, 3D stereo cameras
  - Computer algorithms
- Goal: automate task which human visual system can perform

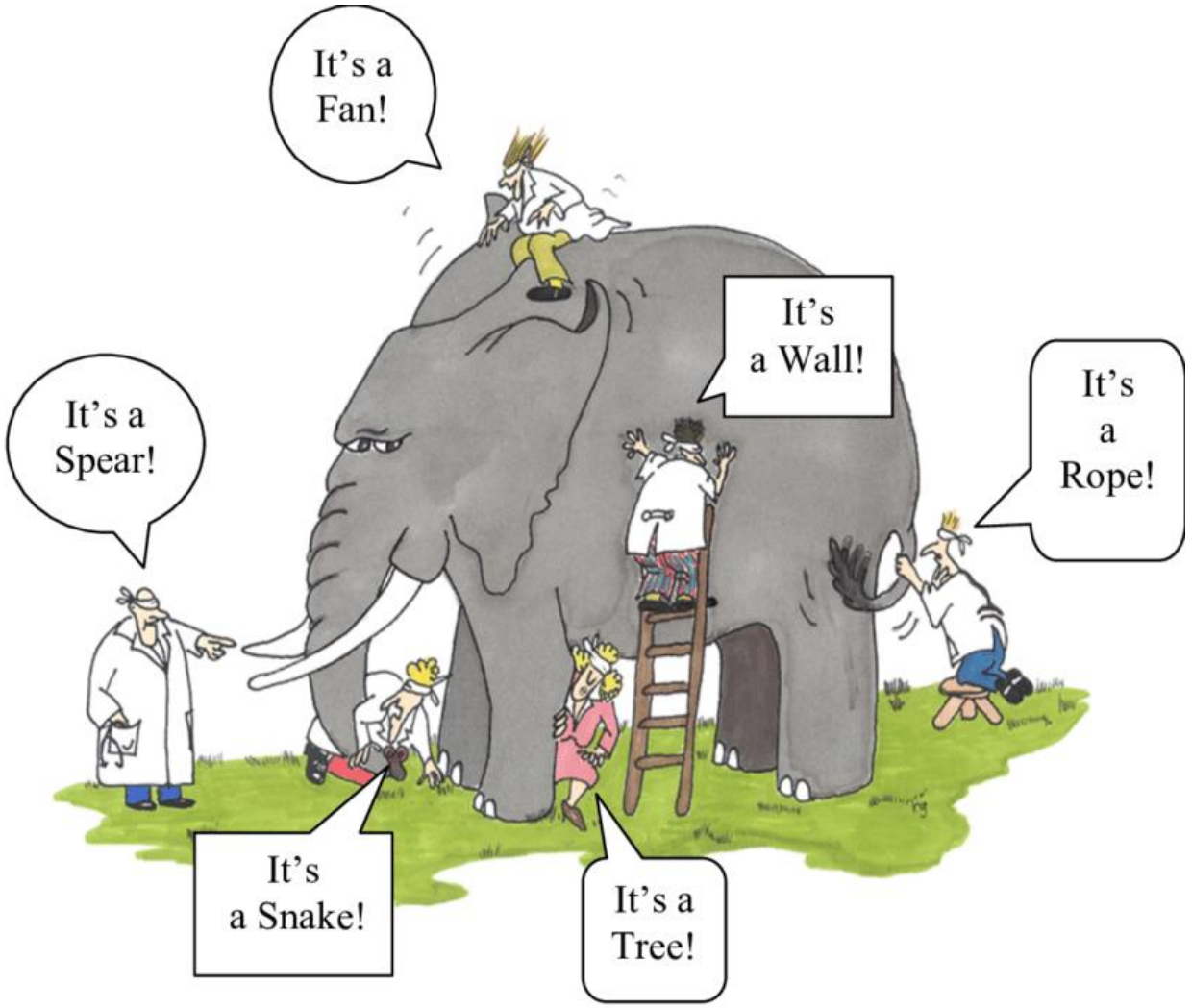




Vision and Image Understanding

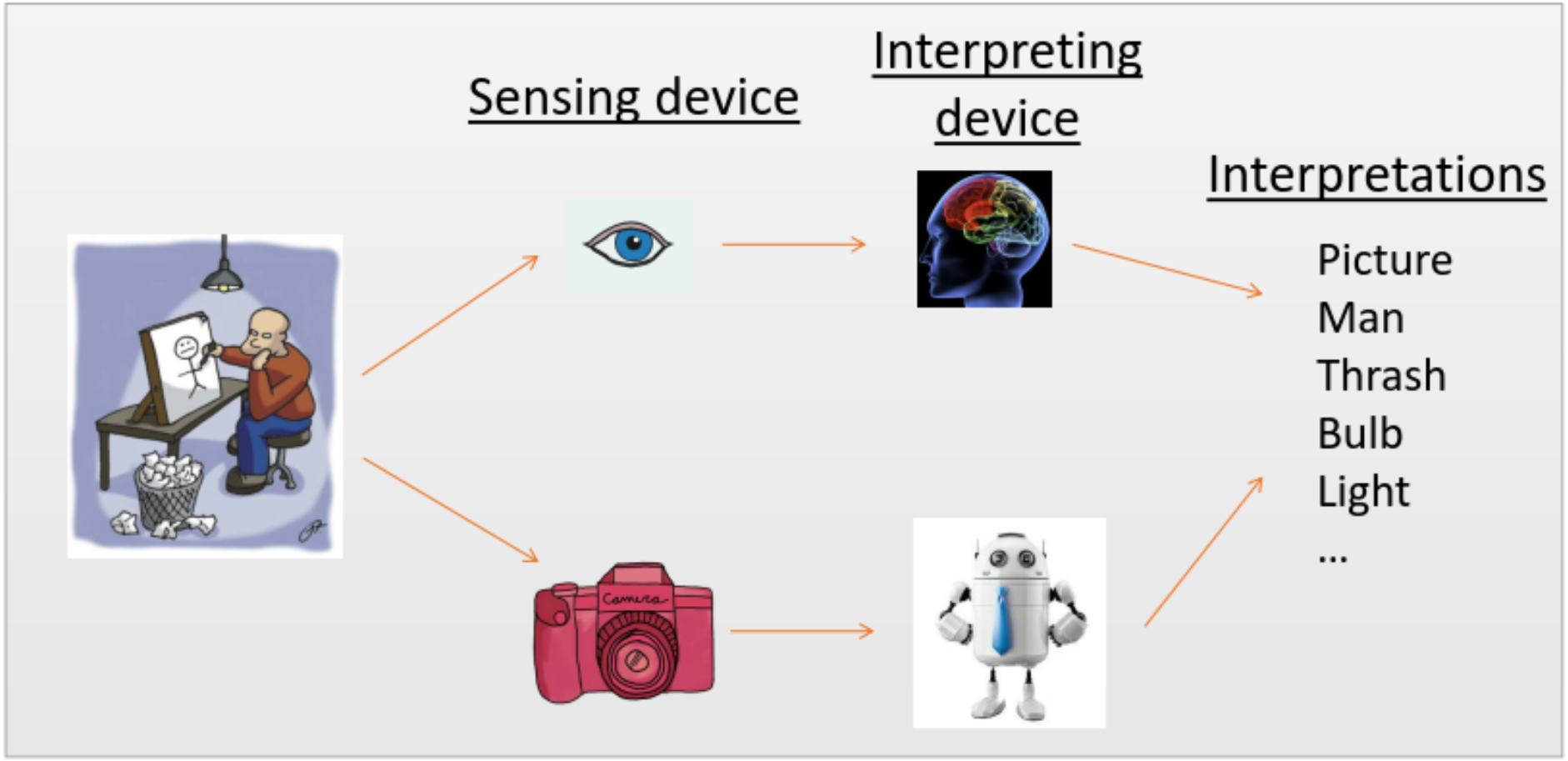
- Visual tasks: We use vision to interact with environments and survive
  - to navigate and avoid obstacles
  - to recognize and pick up objects
  - to identify food and danger
  - … friends and enemies
  - …

Visual Perception
(6 blind man and an elephant)



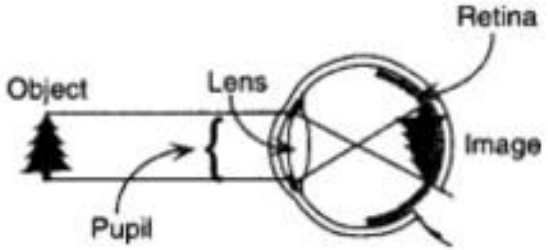
We need a Visual sensor

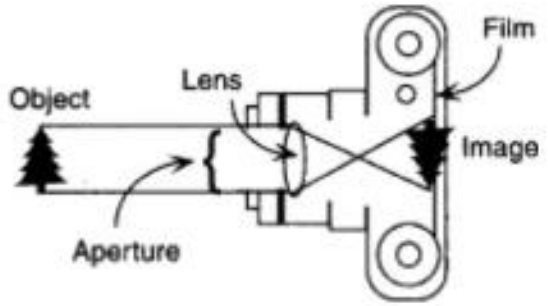
Robot vision Vs human vision



Visual Perception

- Definition: *Process of acquiring knowledge about environmental objects and events by extracting information from the light they emit or reflect [Palmer, 2012].*

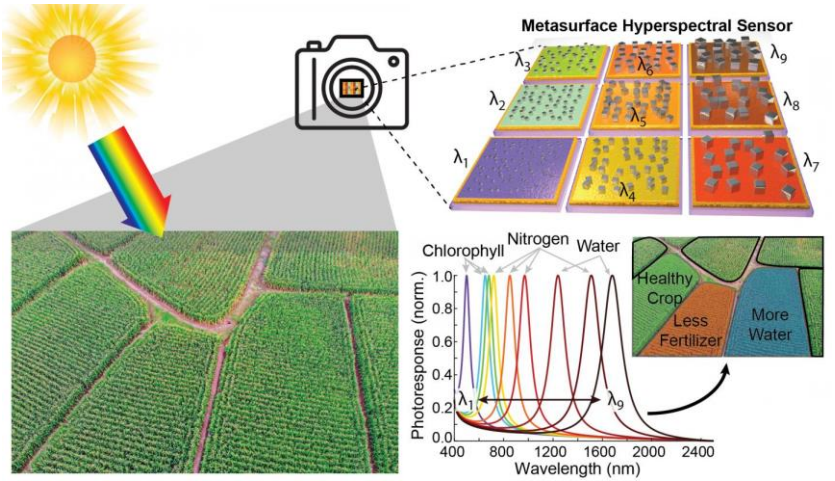




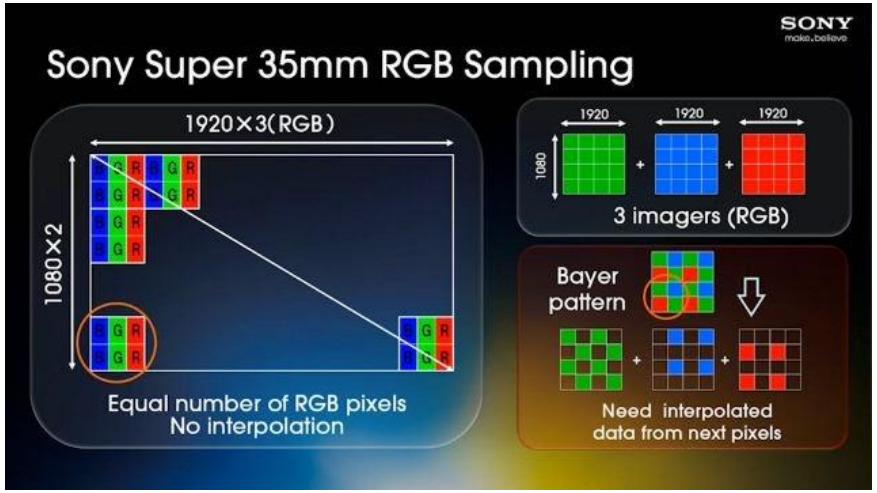
Perception is analogous to taking a picture!
(credit: Palmer, 2012)

A large list of visual sensors

Monochrome cameras



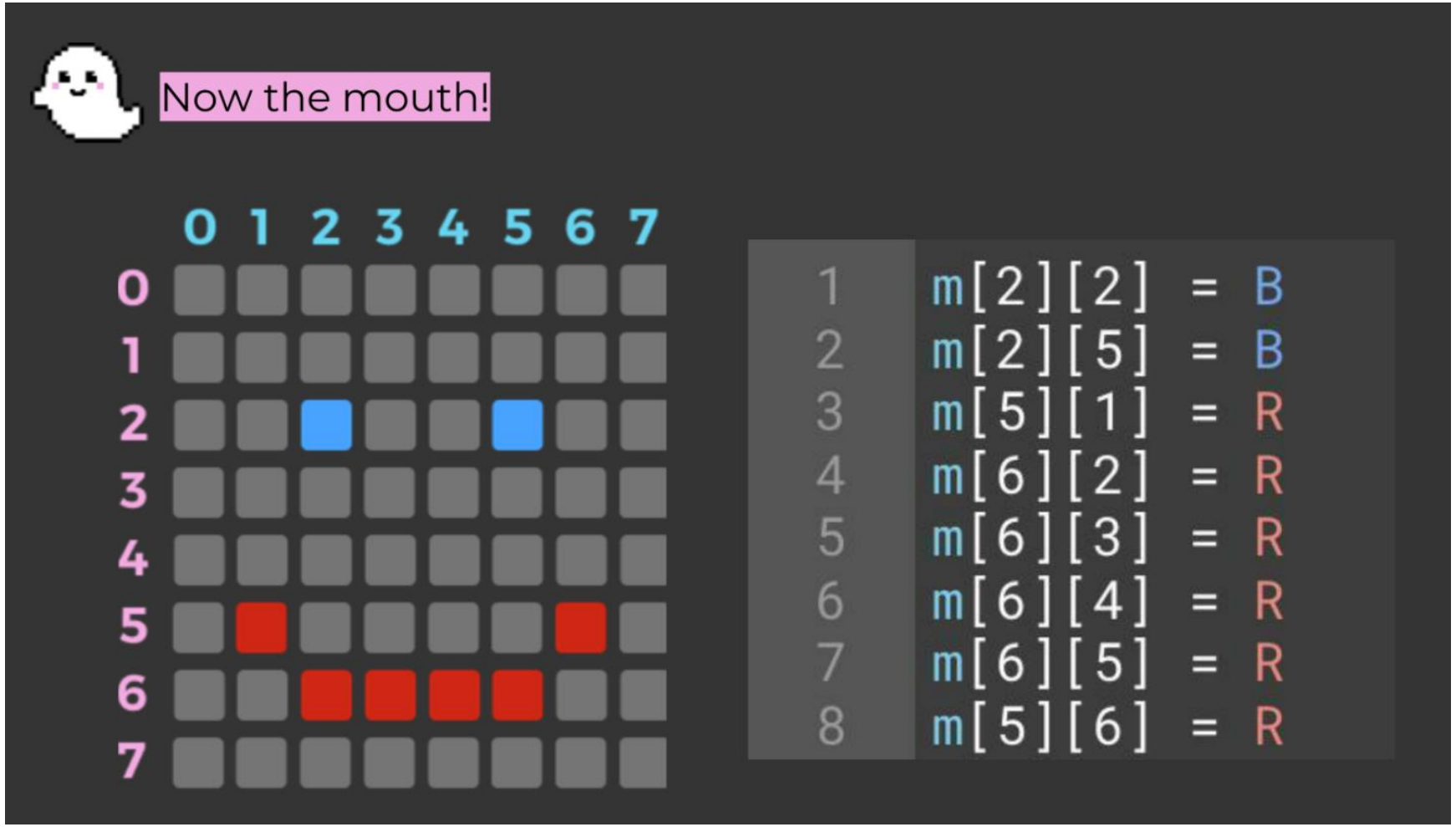
MultiSpectral cameras



RGB cameras

Lidar / Time of flight cameras

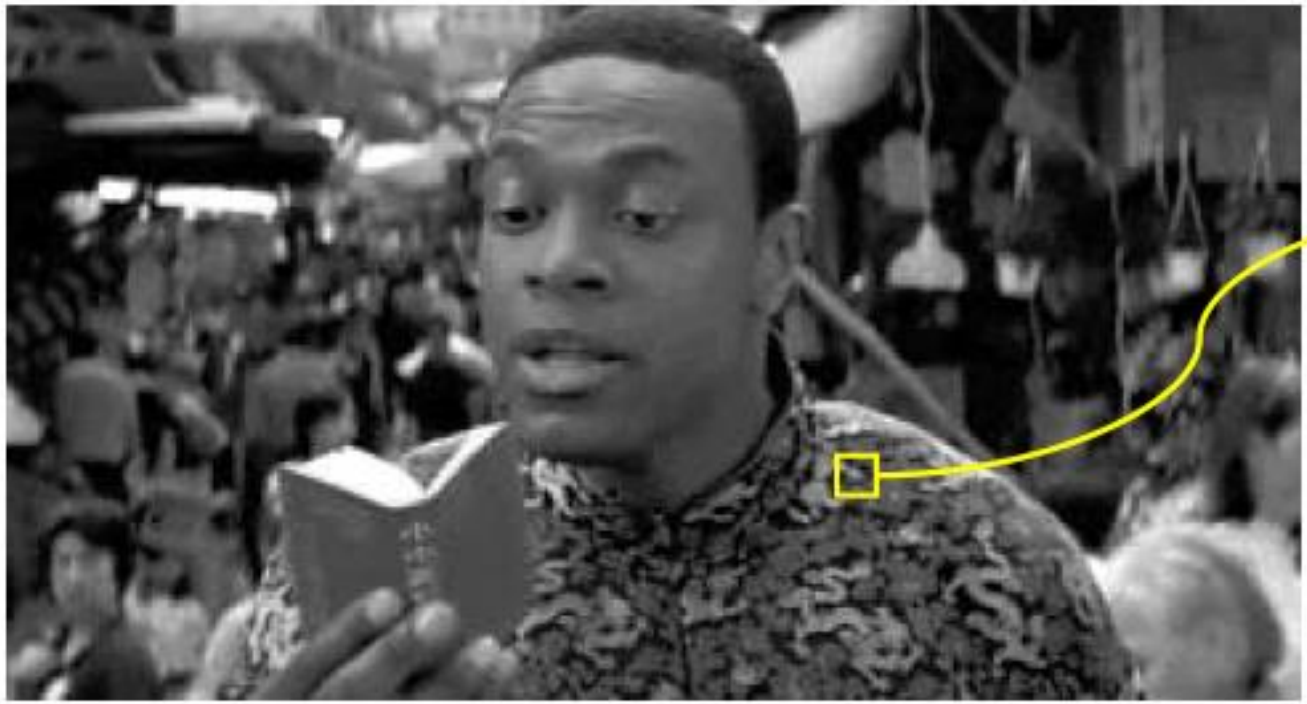
What is a (digital) image?



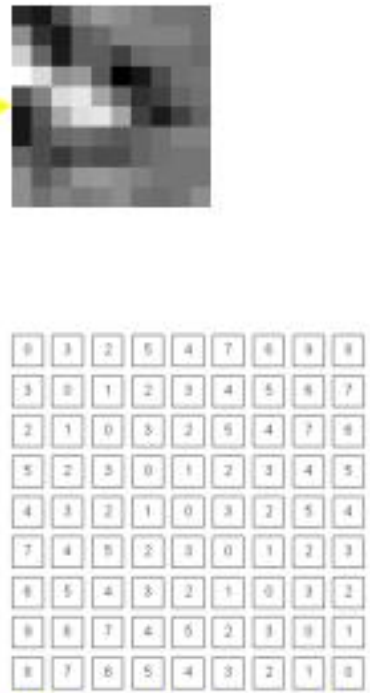
From imagilabs.com

Goal of Robot Vision?

- To bridge the gap between
  - image pixels and “meaning” (semantic)!



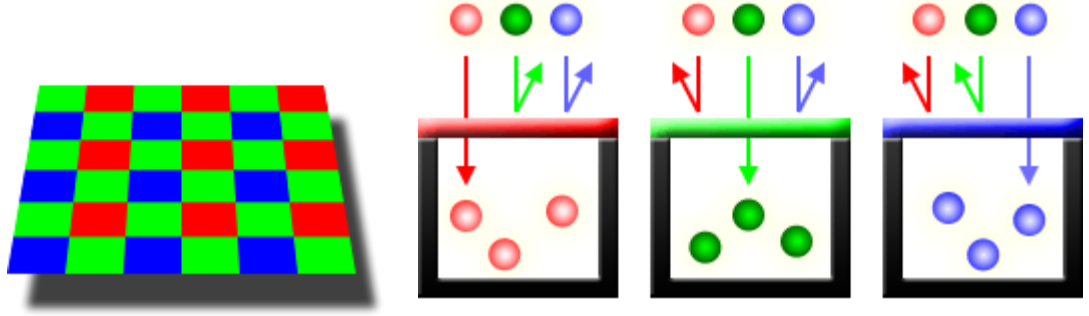
What we see!



What robot sees!

Capturing a color image

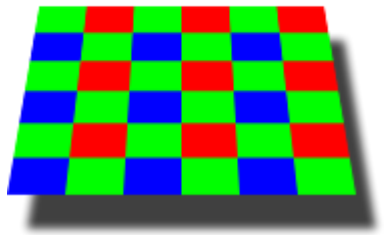
Understanding Digital Camera Sensors (cambridgeincolour.com)



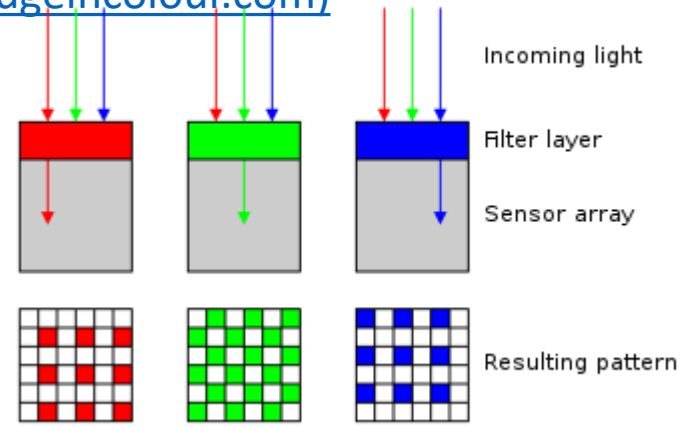
CCD Bayer sensor

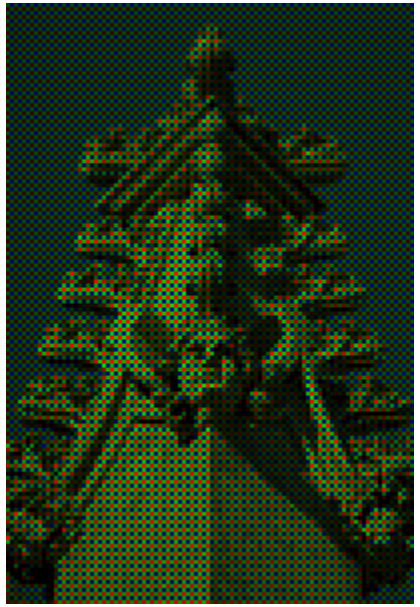
Capturing a color image

[Understanding Digital Camera Sensors (cambridgeincolour.com)]



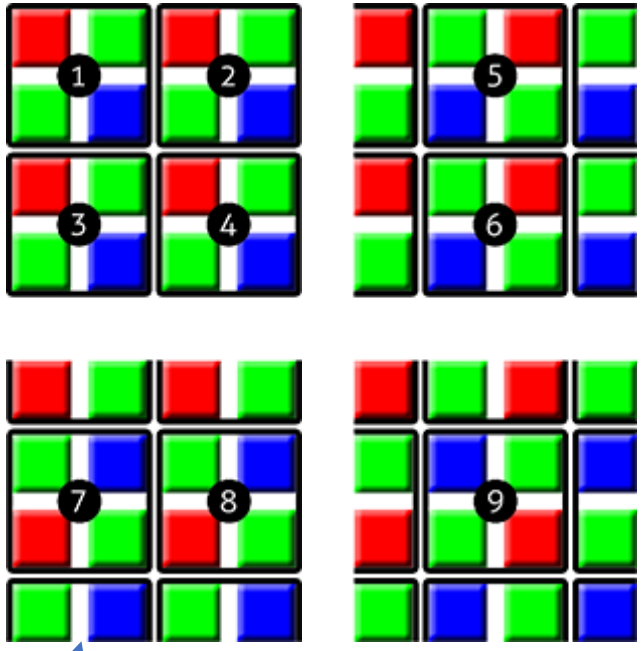
CCD Bayer sensor



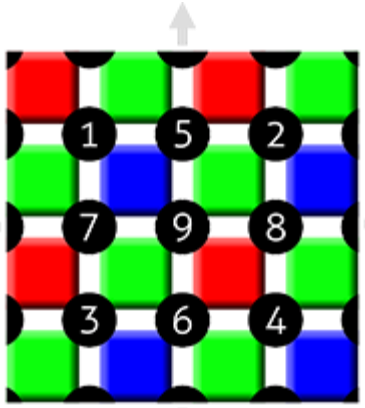


Capturing a color image

[Understanding Digital Camera Sensors (cambridgeincolour.com)]



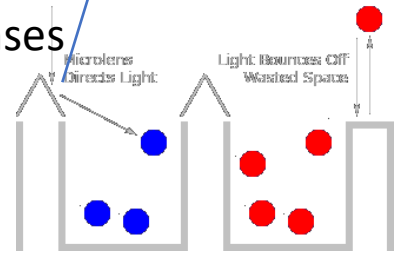
Bayer demosaicing



Next steps:

- Demosaicing artifacts
  - Optical low pass filter
  - Anti-aliasing filtering

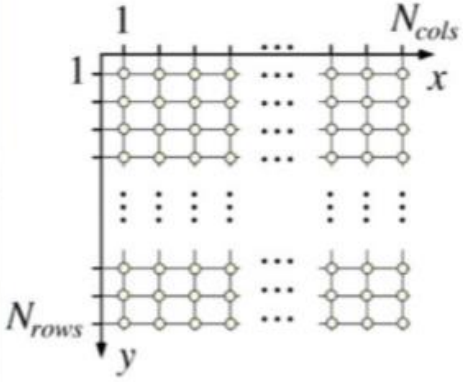
Micro-array lenses



buckets

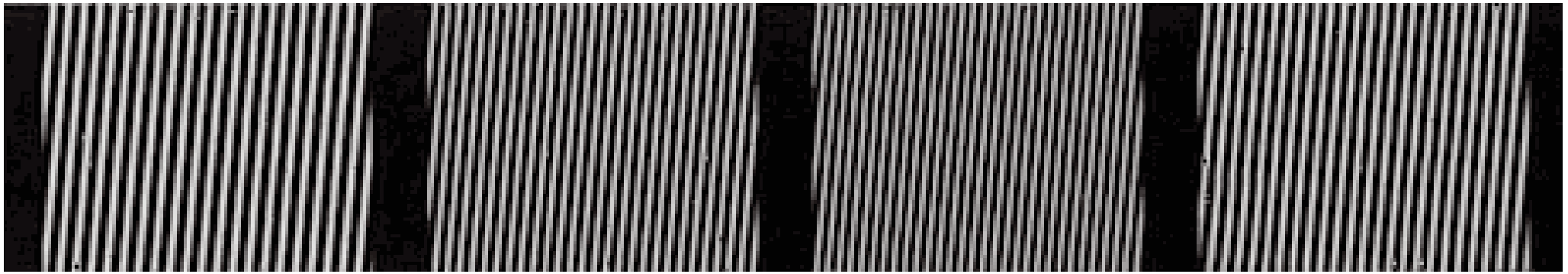
digitalization

8 bits → 255 levels



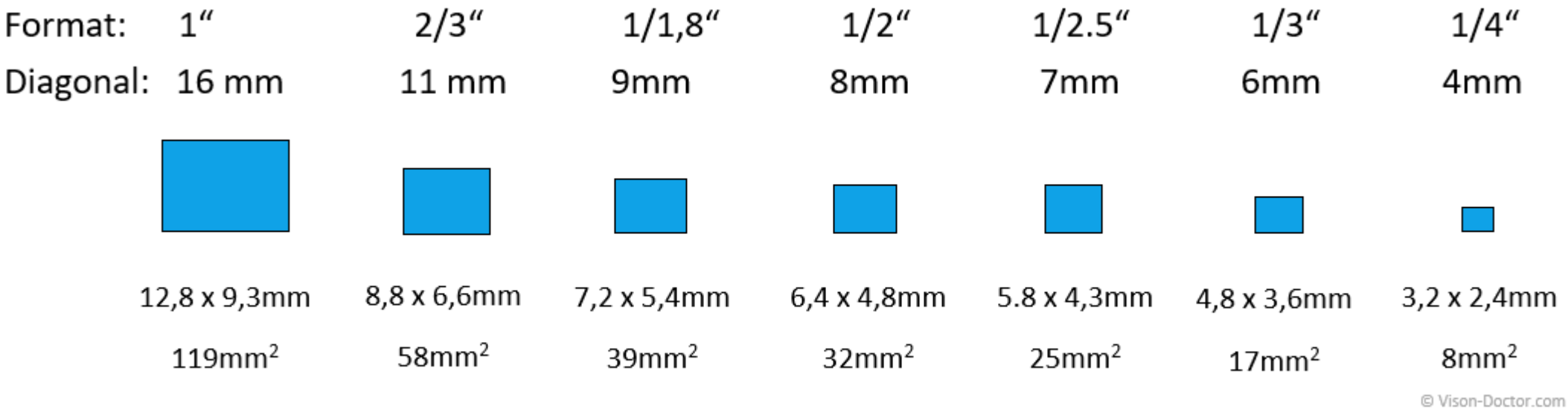
Left hand coordinate system

Anti-Aliasing



65% of Above Size

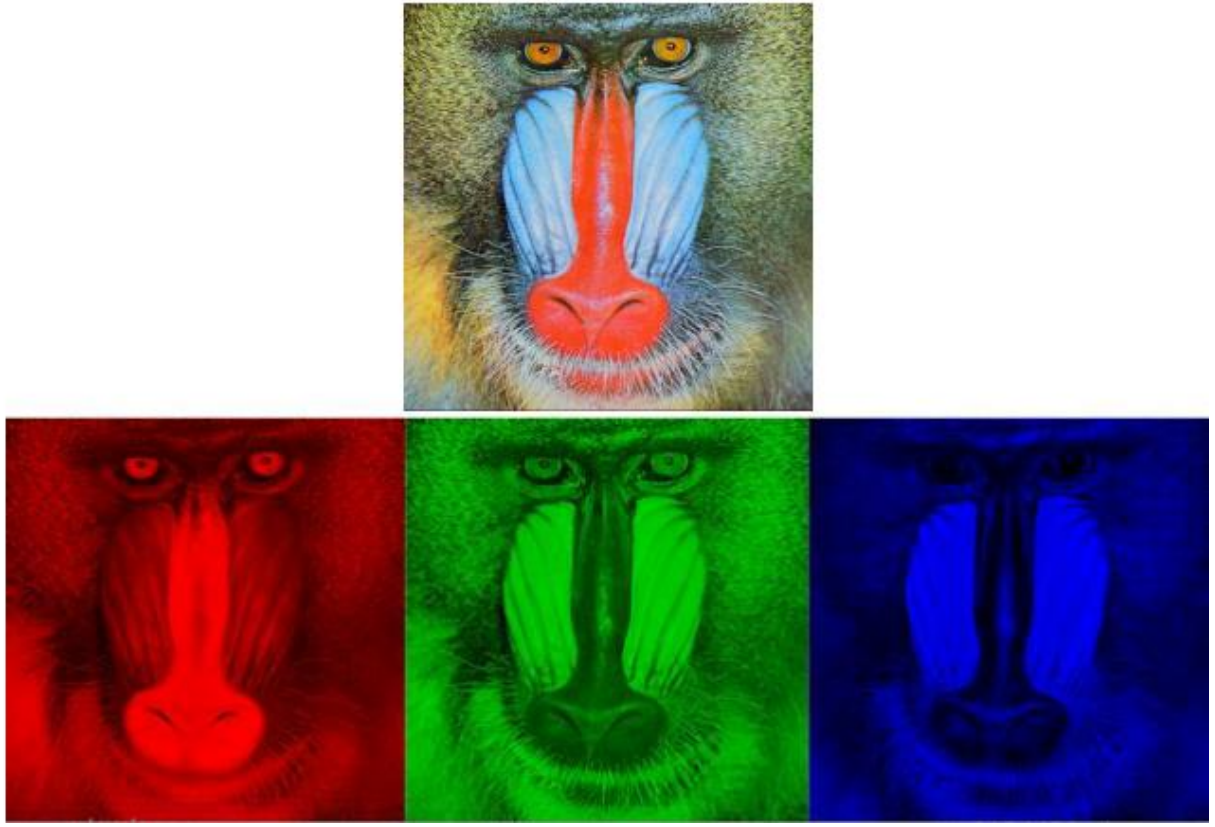
Typical CCD sensor sizes



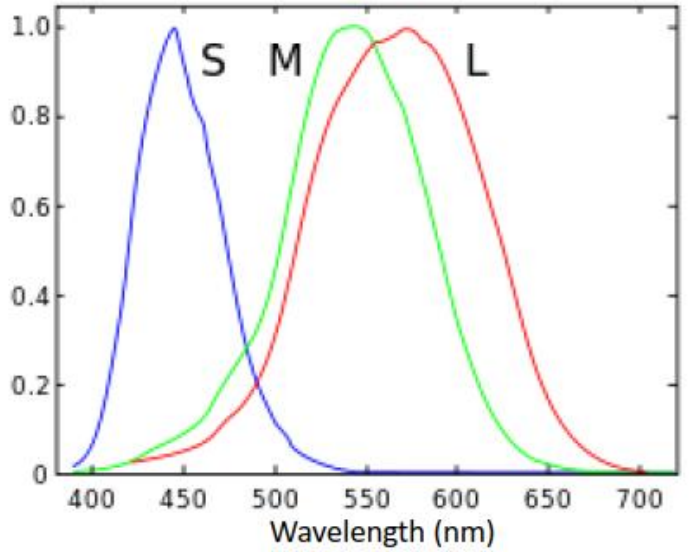
- Industrial cameras usually use 1/3" sensors in case of resolutions of 640 x 480 pixels, cameras with 1280 x 1024 pixels mainly 1/2".
- The quite popular camera resolution of 1600 x 1200 pixels often uses a somewhat larger sensor with 1/1.8" with the same pixel size.
- Sensors of consumer cameras (8 to 12 megapixels for 200 euros) have pixel sizes of mostly 1.7 µm today
- Machine vision cameras (C-mount) with resolutions from VGA to 2 megapixels normally have pixels of 4.6 to 6.5 µm with a 10 - 15 times larger light-active surfaces and thus clearly better signal results. If you need images as noise-free as possible and precise measuring results, look for preferably large sensor pixels, even if these cameras are more expensive!
- Well capacity: This specification describes how many electrons a pixel element can hold before it is completely saturated. A pixel of 5.5 µm structure size can accumulate approximately 20,000 electrons, a 7.4 µm pixel 40,000 electrons.
- The larger the full well capacity, yet the better the maximum signal-noise ratio. Consumer cameras with pixel sizes of 1.7 µm require only about 1,000 photons for the pixel saturation.

Image Type: RGB (red, green, blue)

- Image has three channels (bands)
- Each channel spans a-bit values.



Human Cone-cells (normalized) responsivity spectra

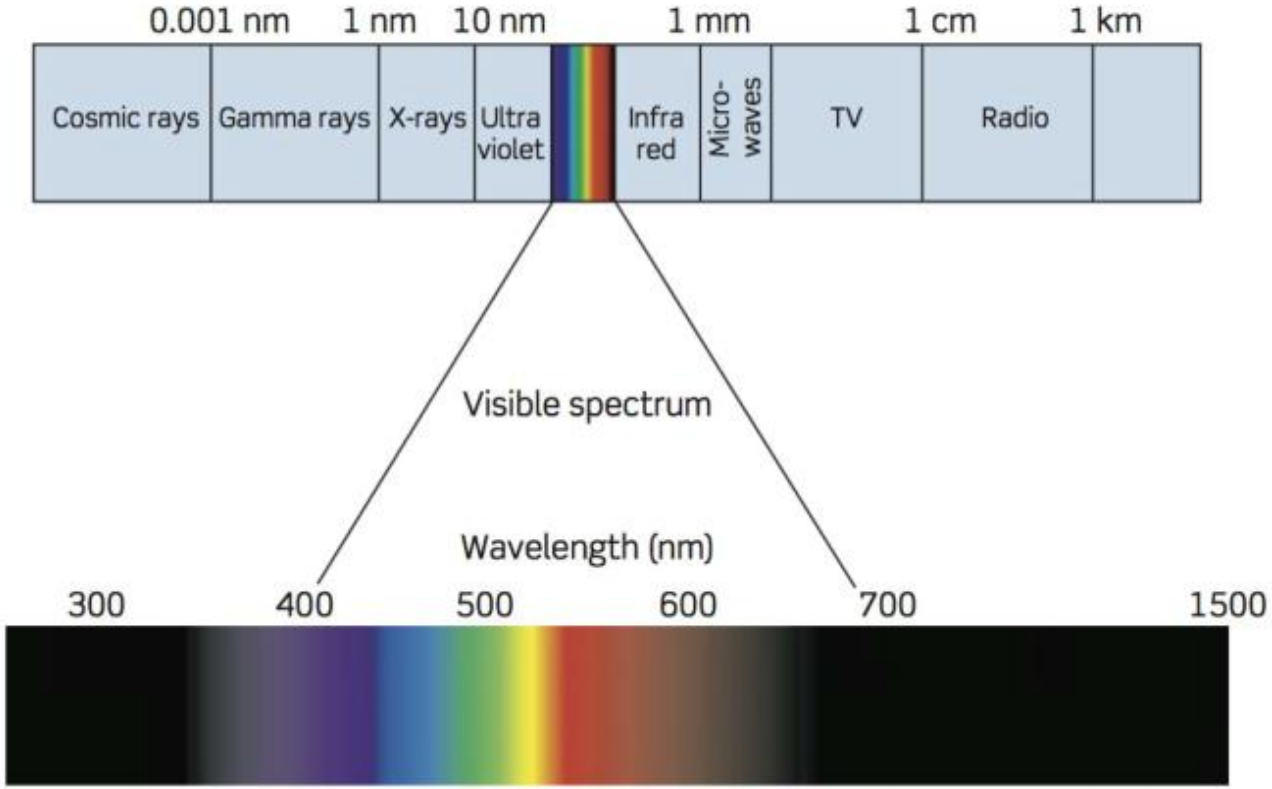


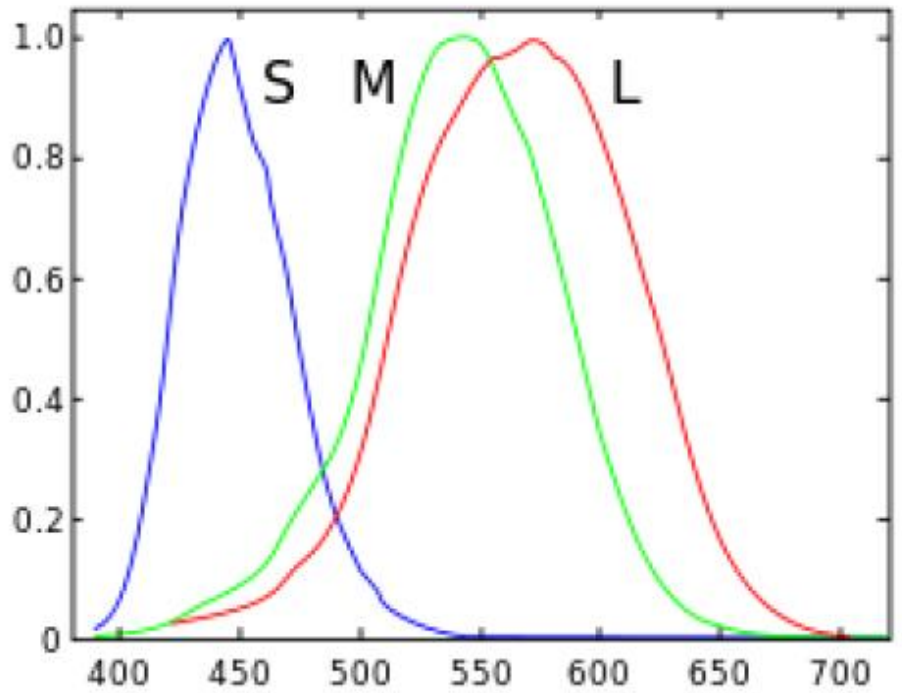
Some people might have 4 cone-types!

Some might have just 2!

Color

- Color vision has evolved over millions of years.

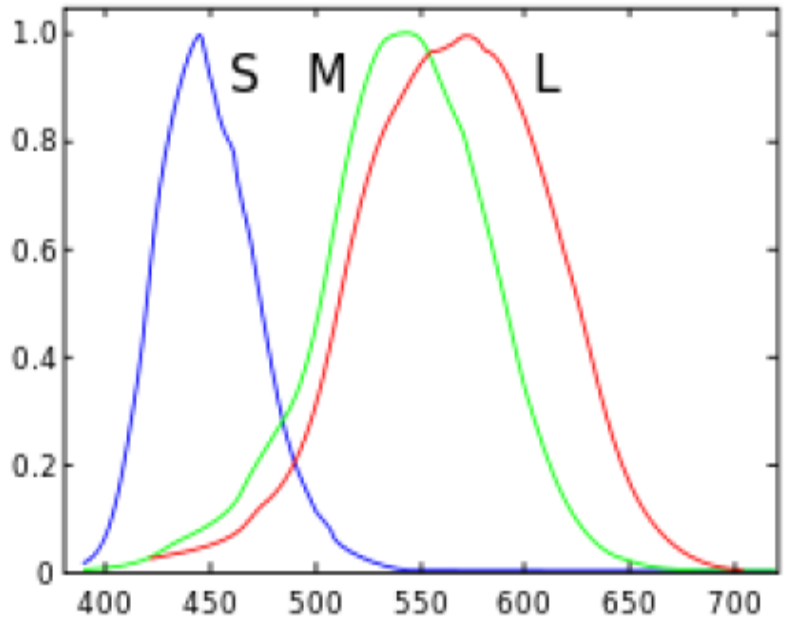




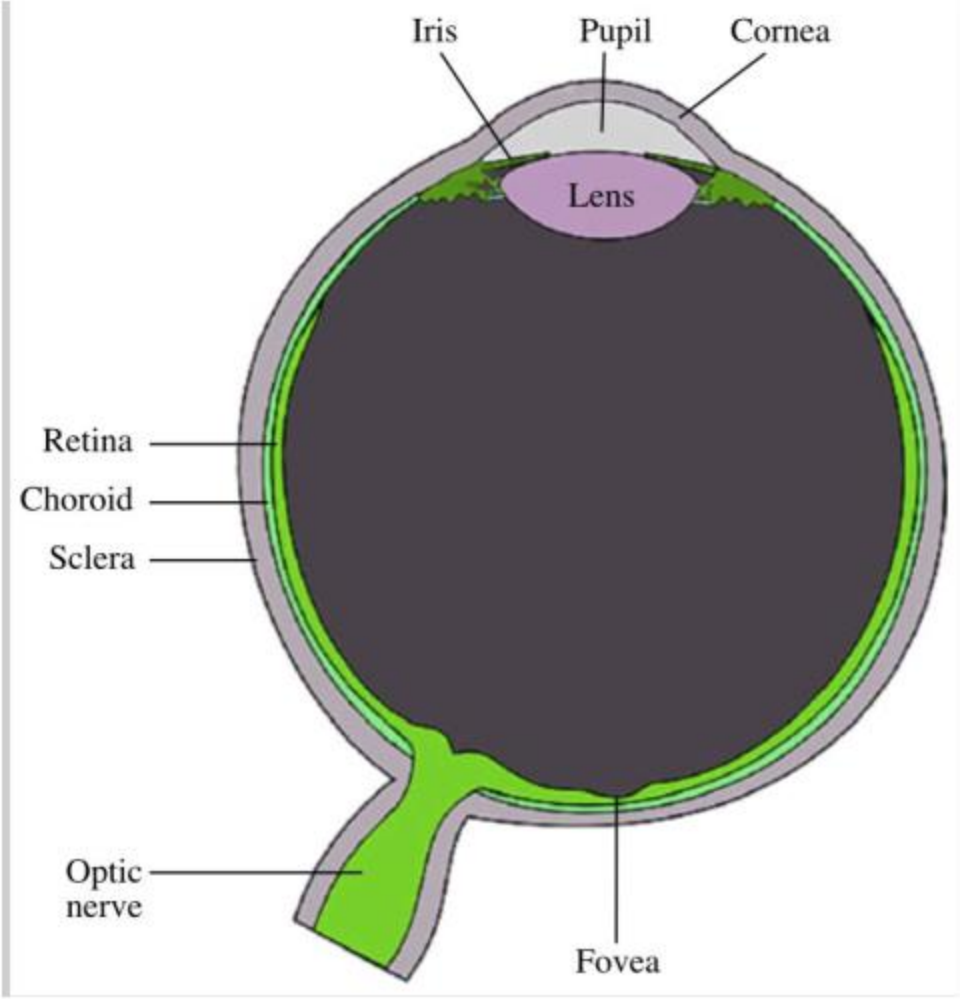
Color

- If there is no light, there is no color!
- Human vision can only discriminate a few dozens of grey levels on a screen, but hundreds of thousands of different colors.

- RED -> ~625 to 780 nm [long wavelength]
- ORANGE -> ~ 590 to 625 nm [long wavelength]
- YELLOW -> ~565 to 590 nm [middle range wavelength]
- GREEN -> ~ 500 to 565 nm [middle range wavelength]
- CYAN -> ~485 to 500 nm [middle range wavelength]
- BLUE -> ~440 to 485 nm [short wavelength]
- VIOLET -> ~330 to 440 nm [very short wavelength]



Retina of Human Eye



There are three different types of color-sensitive cones corresponding to (roughly)

- RED (64% of the cones)
- GREEN (about 32%), and
- BLUE (about 2%).

6-7 million cones
120 million rods

Credit: Klette, 2012.

Demo: Color is in your head !

Questions?