CAP 4453
Robot Vision

Dr. Gonzalo Vaca-Castaño

gonzalo.vacacastano@ucf.edu

Course Website + resources

- CANVAS (webcourses)
- Alternative: CAP 4453 - Fall 2023 (ucf.edu)

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

- Szeliski, Computer Vision: Algorithms and Applications

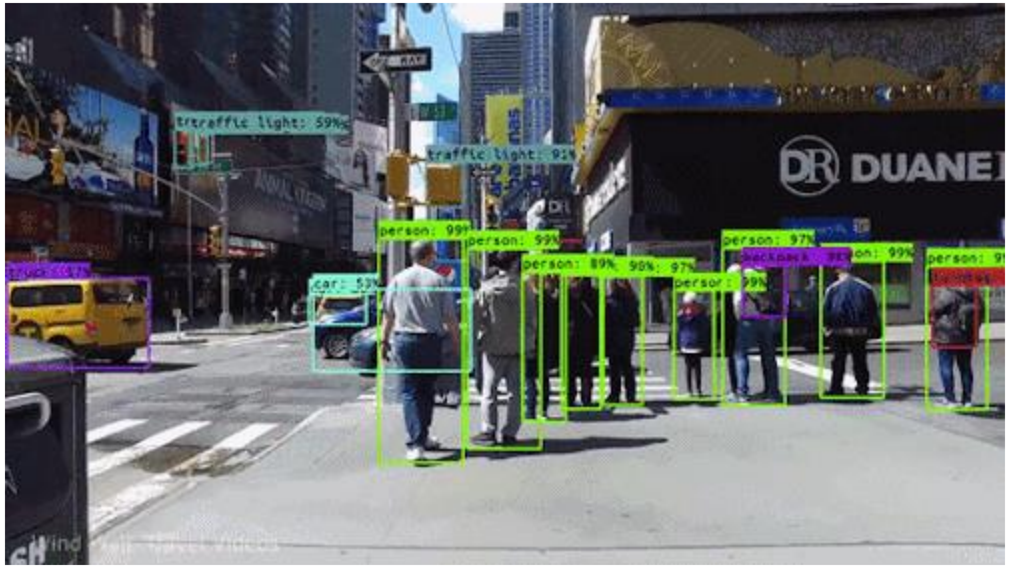
- Python for Computer Vision. A tutorial will be given in the class on PyTorch for deep learning.

Course logistics

- Class time: Monday and wednesday 12:00pm – 1:15pm
    - Classroom: MSB O360
- Office hours [Zoom]
    - Wednesday 8:00pm-9:00 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.

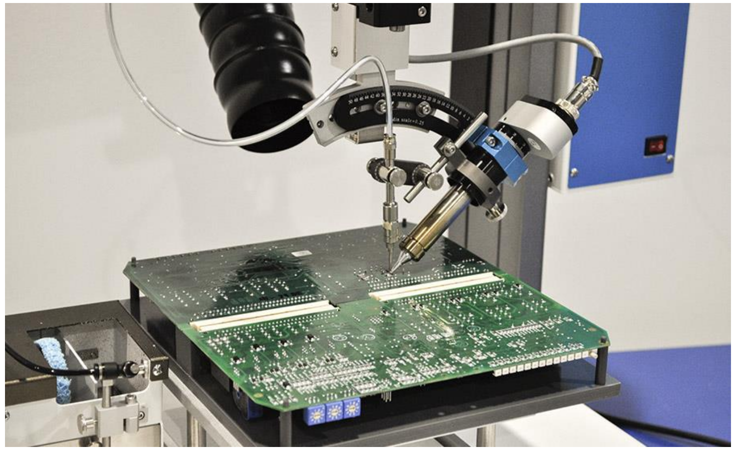
Questions?

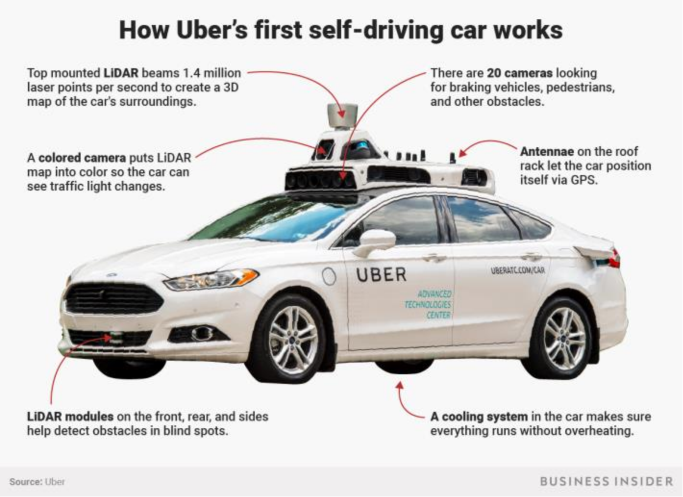
Robot Vision

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

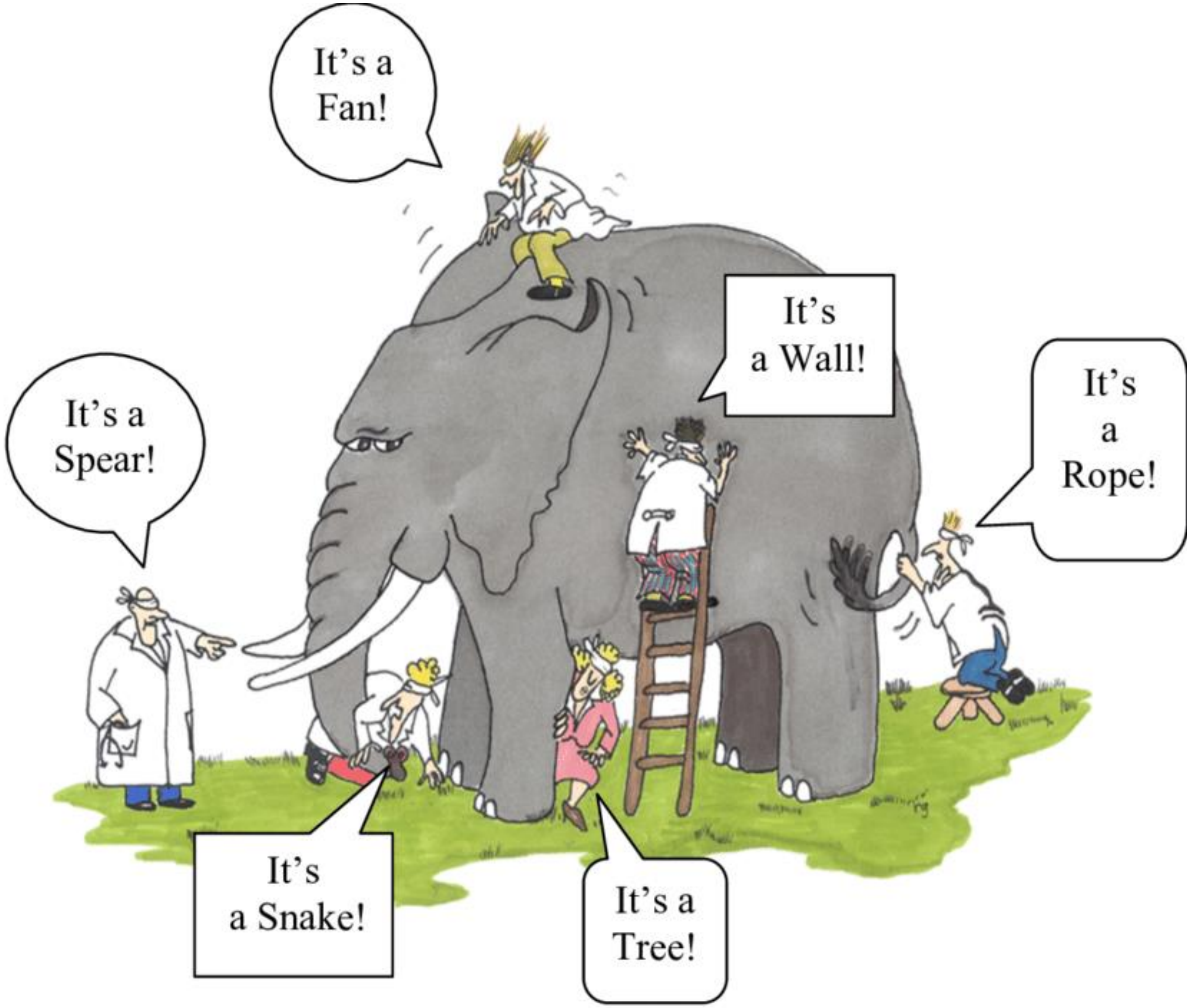




Vision and Image Understanding

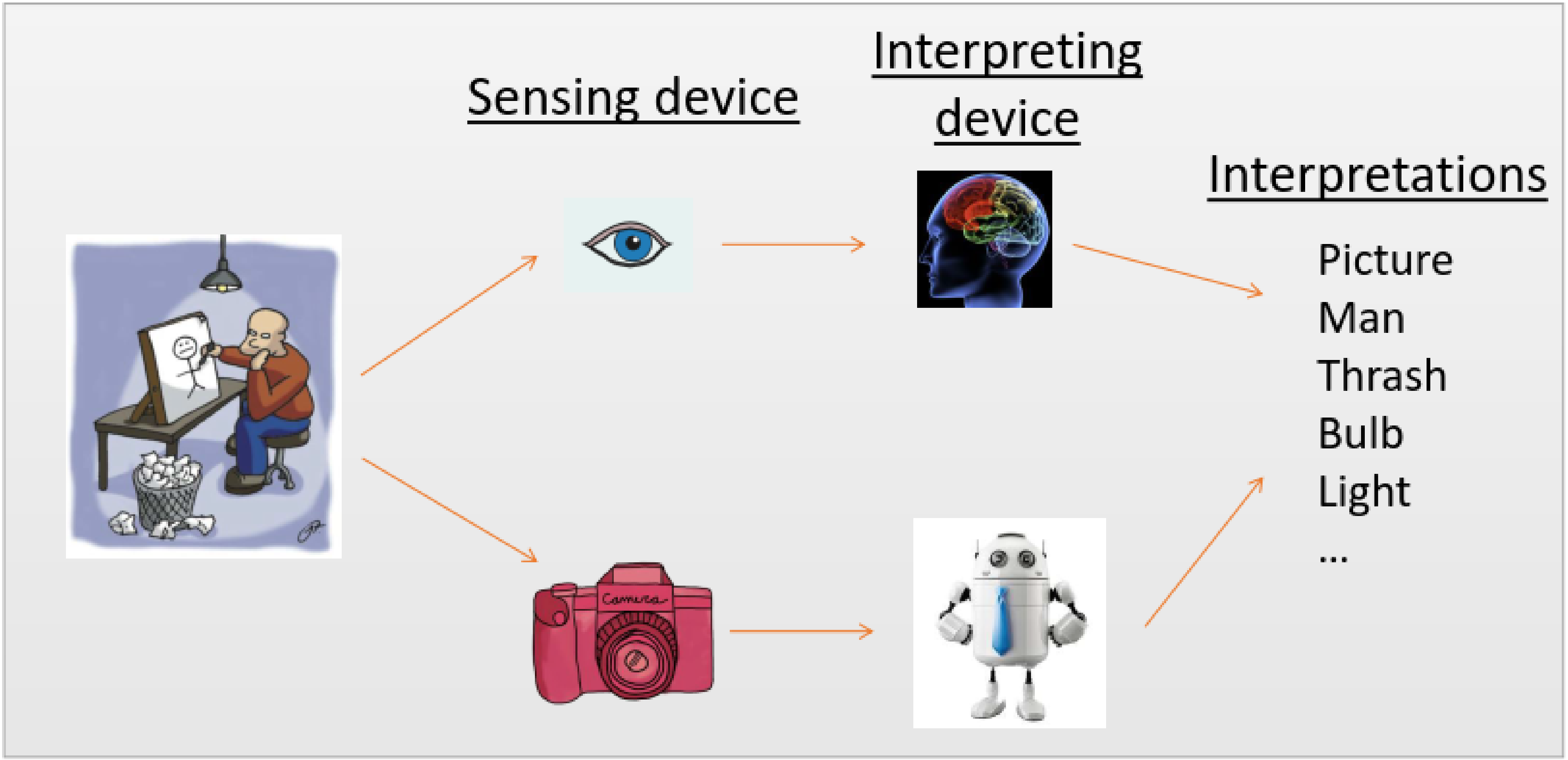
- Visual tasks: We use vision to interact with environments and survive
  - to navigate and avoid obstacles
  - to recognize and pick up objects
  - to identify food and danger
  - … friends and enemies
  - …

Visual Perception
(6 blind man and an elephant)



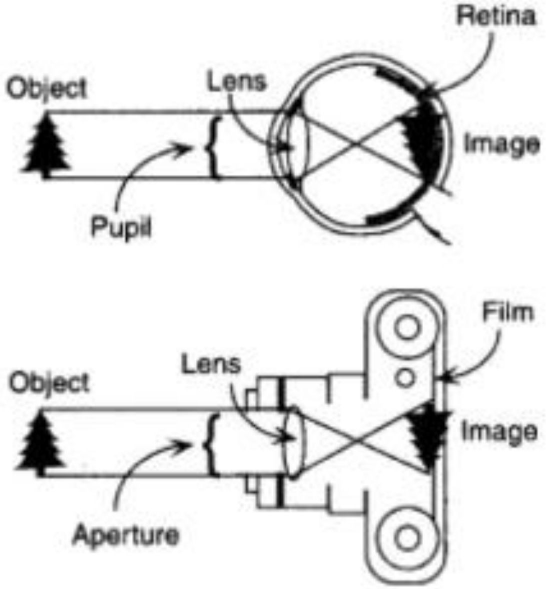
We need a Visual sensor

Robot vision Vs human vision



Visual Perception

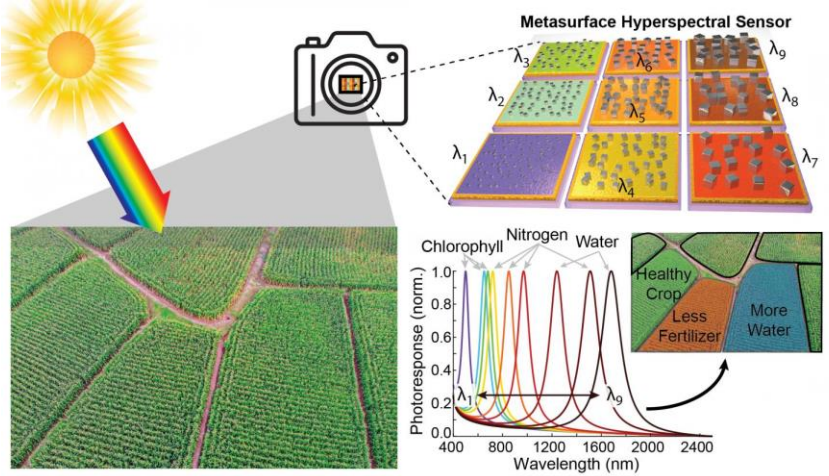
- Definition: *Process of acquiring knowledge about environmental objects and events by extracting information from the light they emit or reflect [Palmer, 2012].*



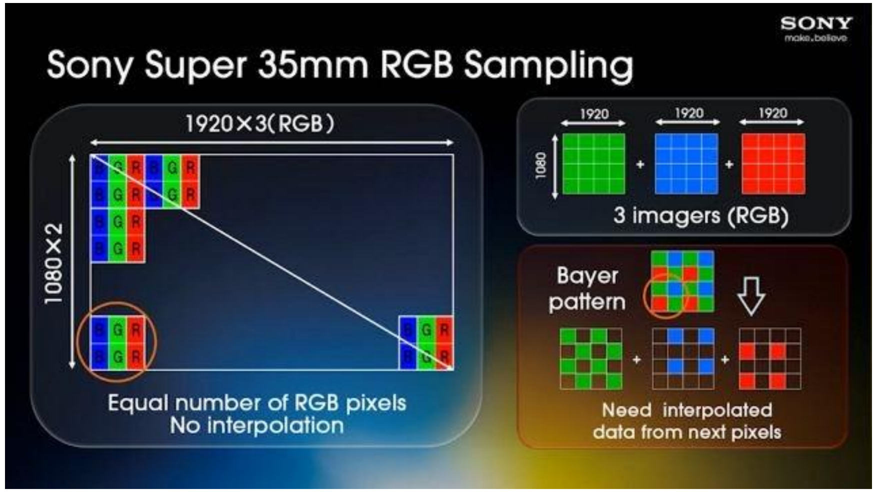
Perception is analogous to taking a picture!
(credit: Palmer, 2012)

A large list of visual sensors

Monochrome cameras



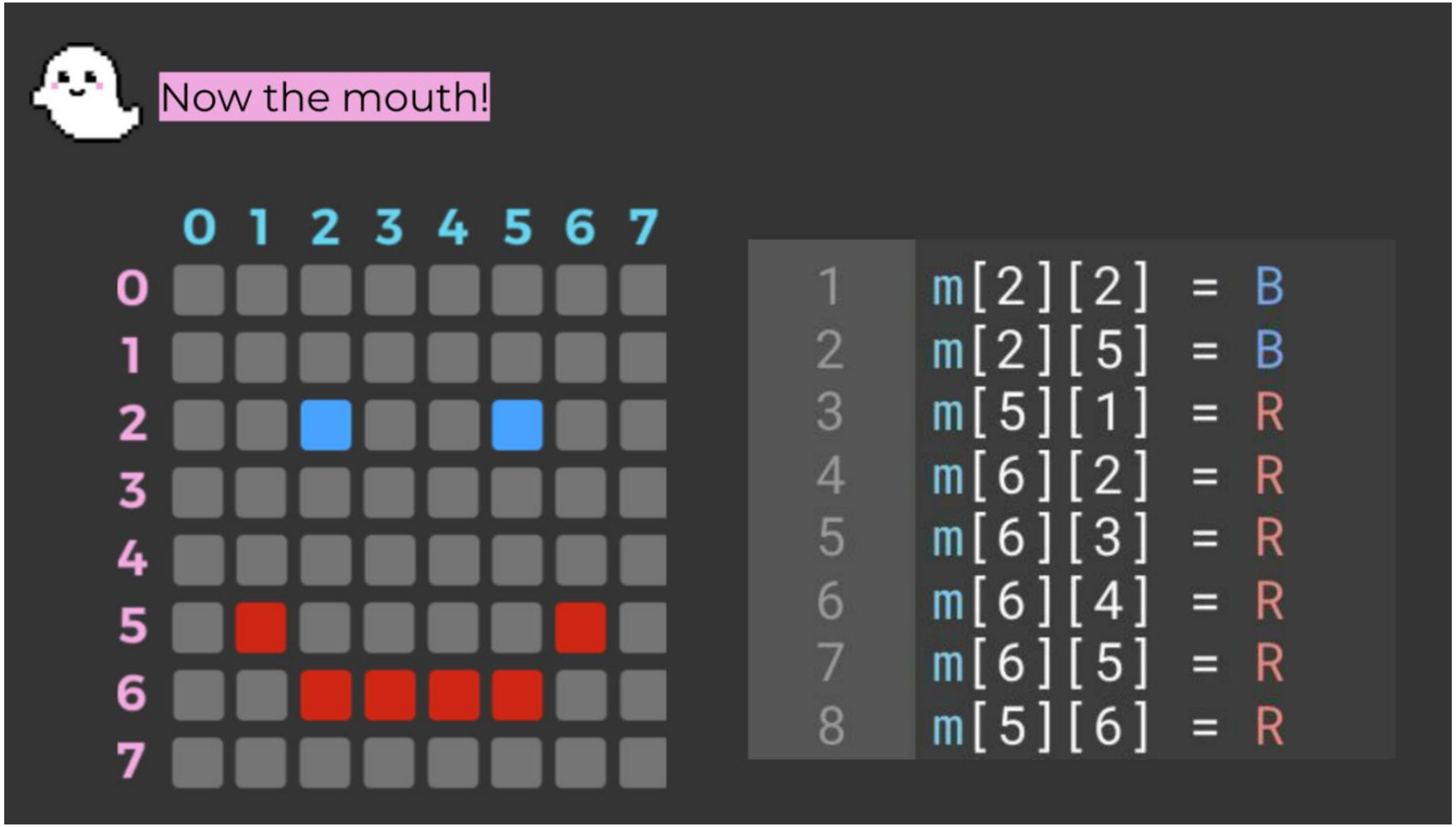
MultiSpectral cameras



RGB cameras

Lidar / Time of flight cameras

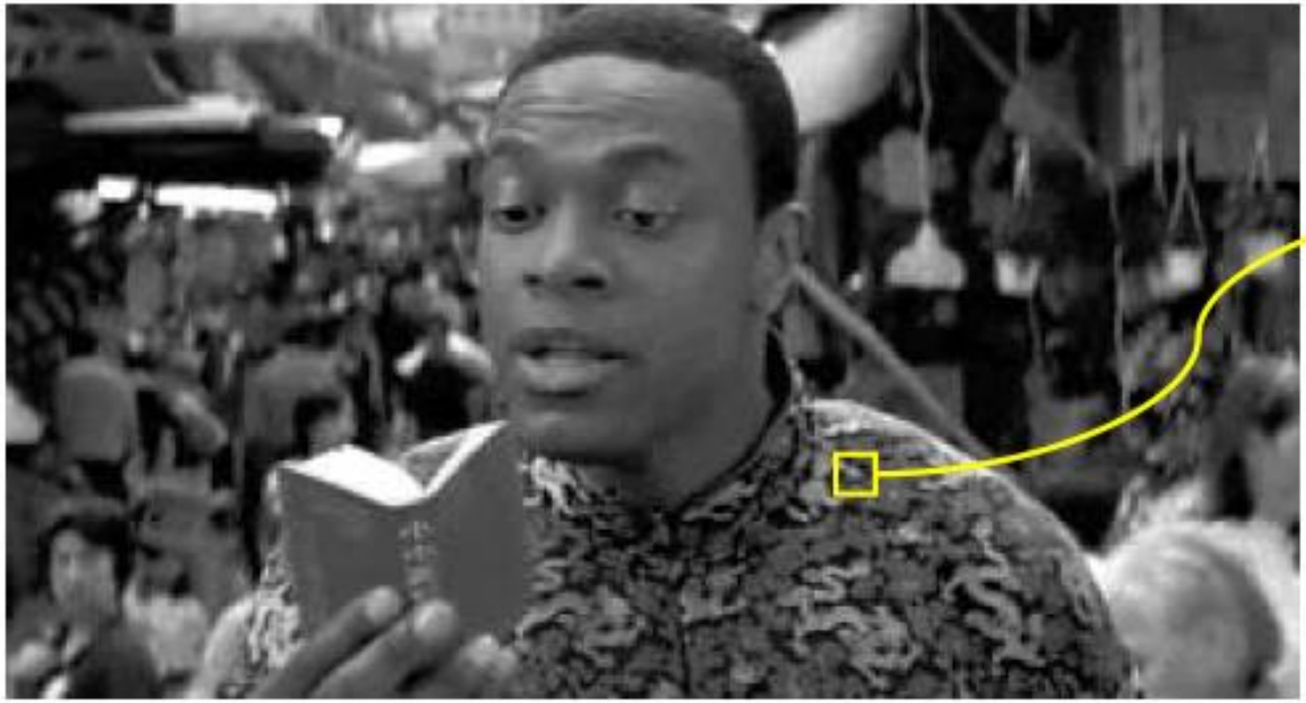
What is a (digital) image?



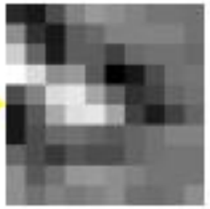
From imagilabs.com

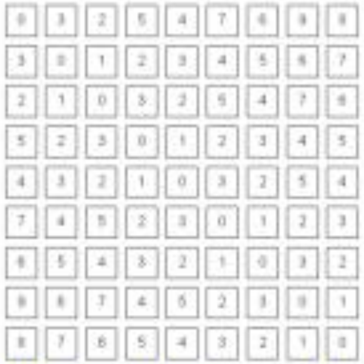
Goal of Robot Vision?

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



What we see!

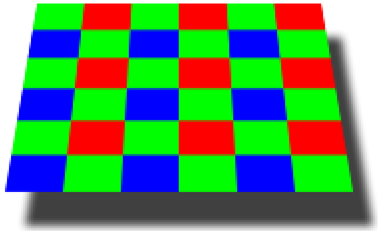


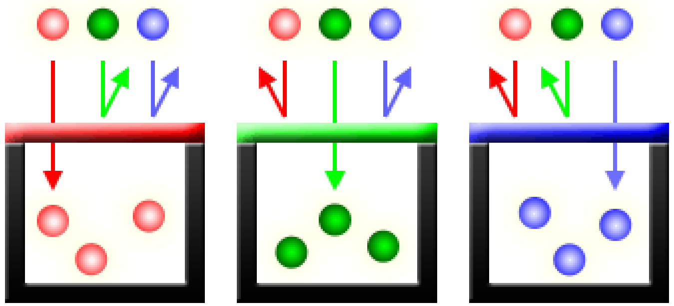


What robot sees!

Capturing a color image

[Understanding Digital Camera Sensors (cambridgeincolour.com)]

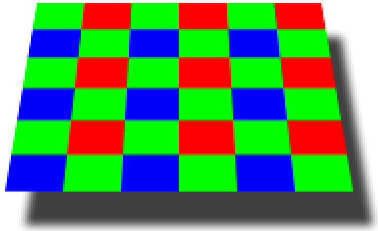




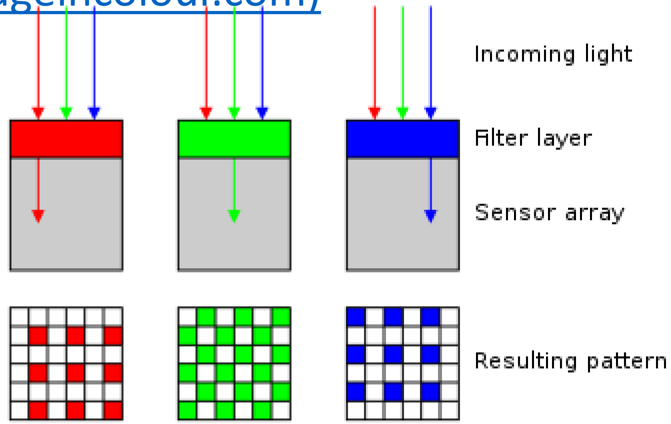
CCD Bayer sensor

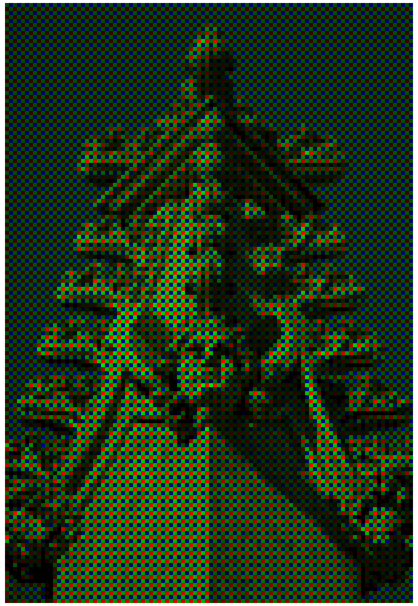
Capturing a color image

[Understanding Digital Camera Sensors (cambridgeincolour.com)](https://www.cambridgeincolour.com)



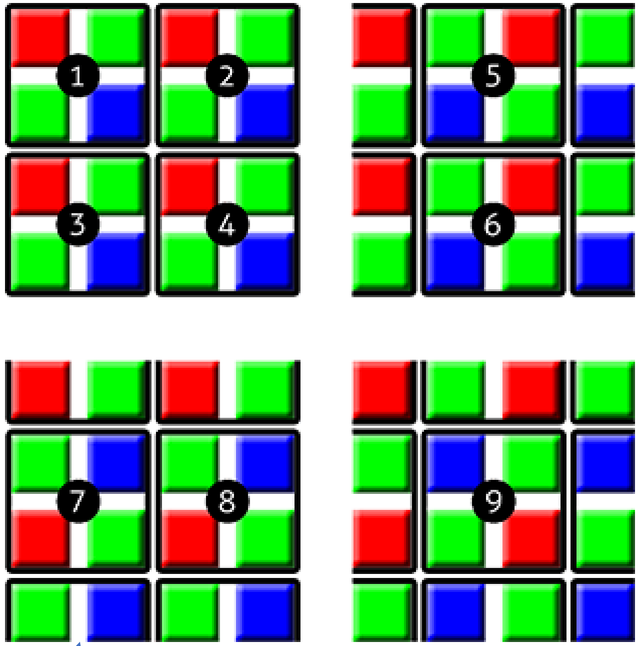
CCD Bayer sensor



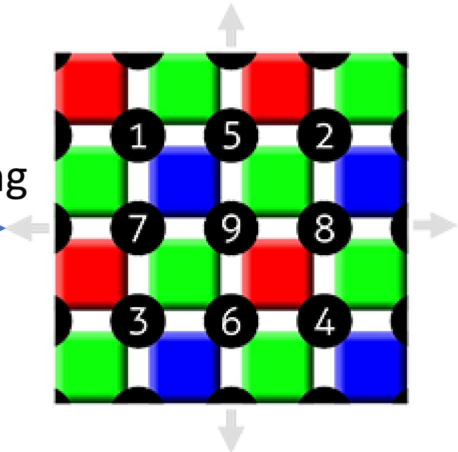


Capturing a color image

[Understanding Digital Camera Sensors (cambridgeincolour.com)](cambridgeincolour.com)



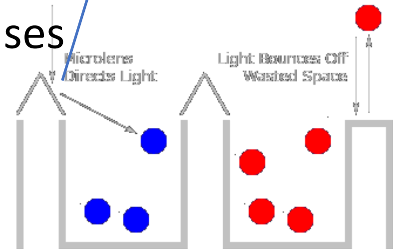
Bayer demosaicing



Next steps:

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

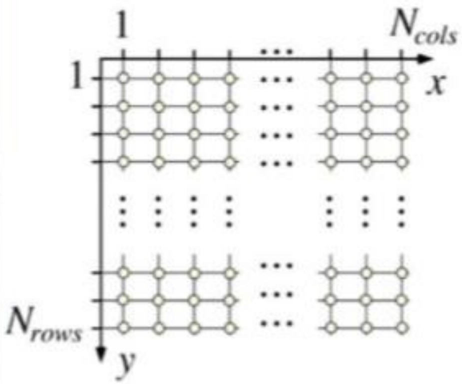
Micro-array lenses



buckets

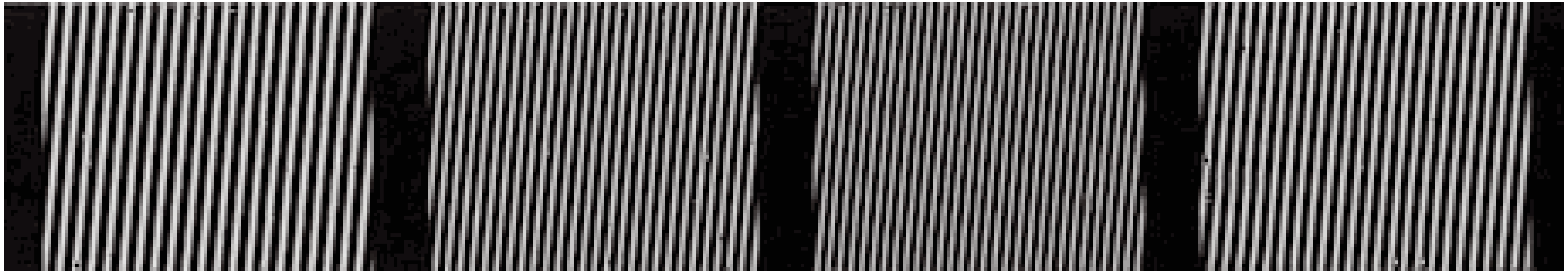
digitalization

8 bits → 255 levels



Left hand coordinate system

Anti-Aliasing



65% of Above Size

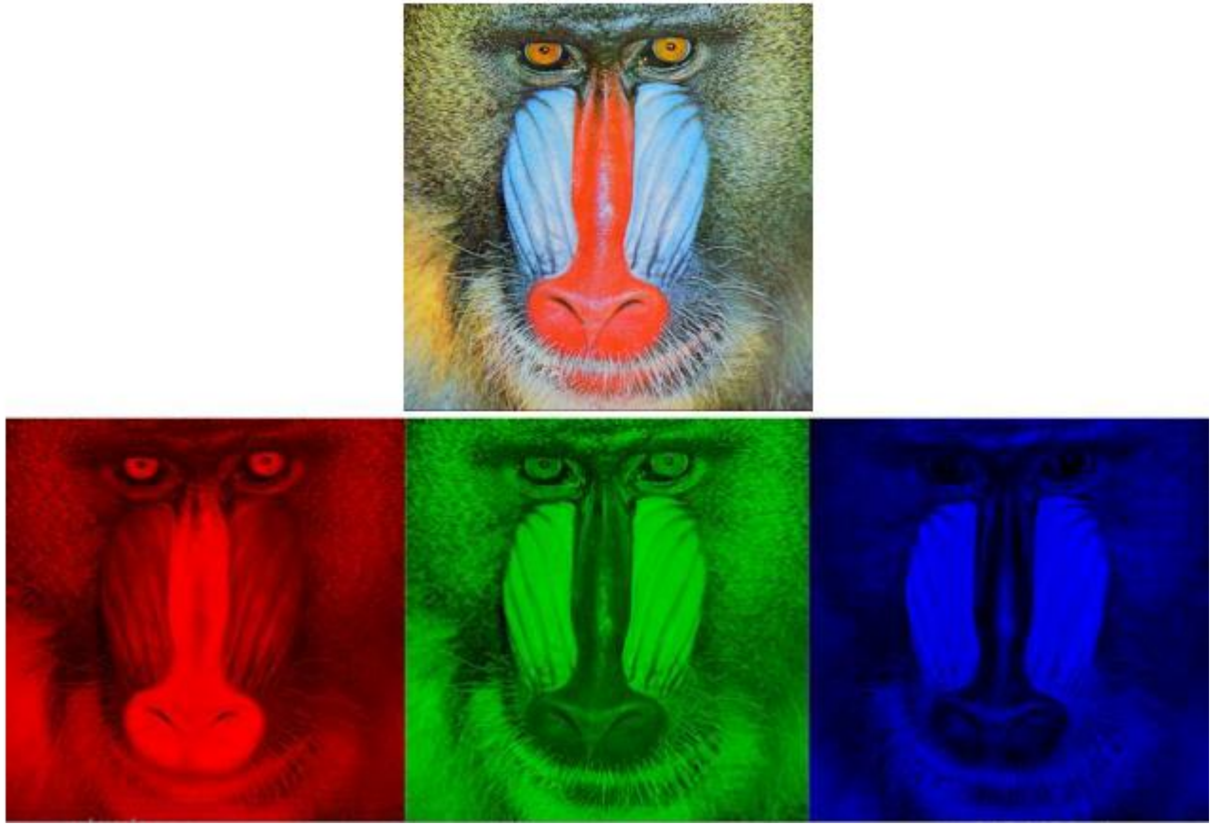
Typical CCD sensor sizes

Format:	1“	2/3“	1/1,8“	1/2“	1/2.5“	1/3“	1/4“
Diagonal:	16 mm	11 mm	9mm	8mm	7mm	6mm	4mm
	12,8 x 9,3mm	8,8 x 6,6mm	7,2 x 5,4mm	6,4 x 4,8mm	5.8 x 4,3mm	4,8 x 3,6mm	3,2 x 2,4mm
	$119mm^2$	$58mm^2$	$39mm^2$	$32mm^2$	$25mm^2$	$17mm^2$	$8mm^2$

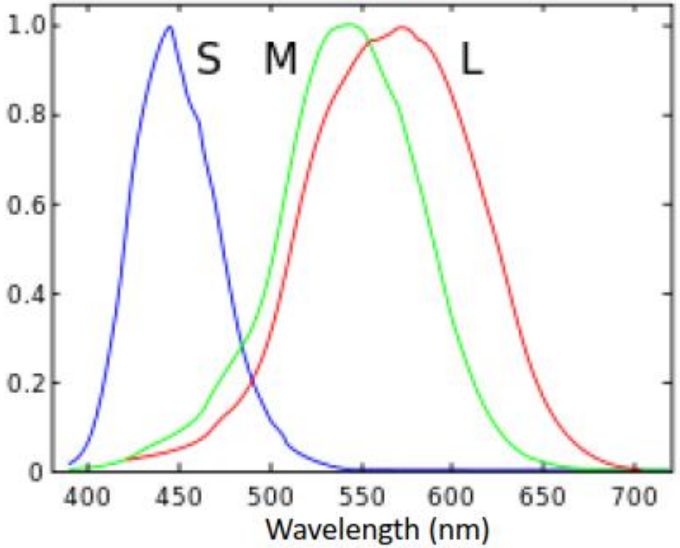
- Industrial cameras usually use 1/3" sensors in case of resolutions of 640 x 480 pixels, cameras with 1280 x 1024 pixels mainly 1/2".
- The quite popular camera resolution of 1600 x 1200 pixels often uses a somewhat larger sensor with 1/1.8" with the same pixel size.
- Sensors of consumer cameras (8 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

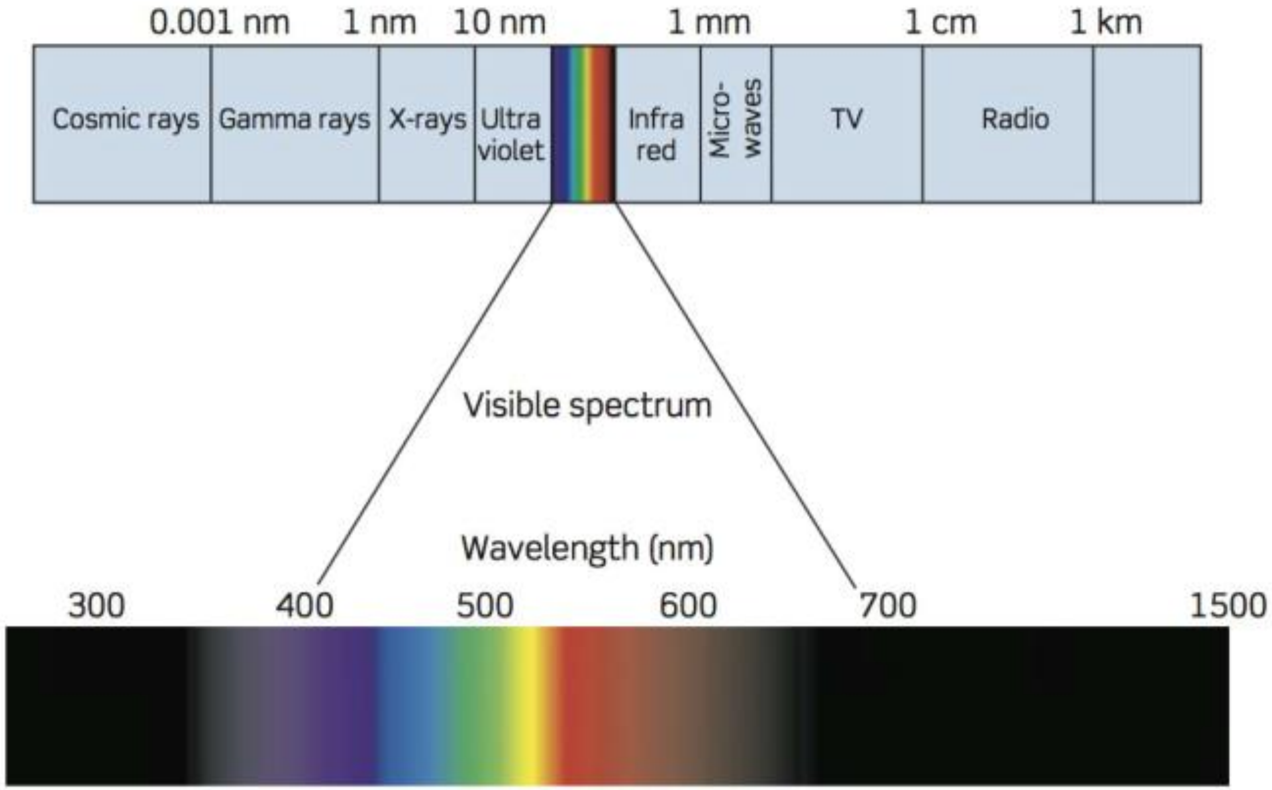


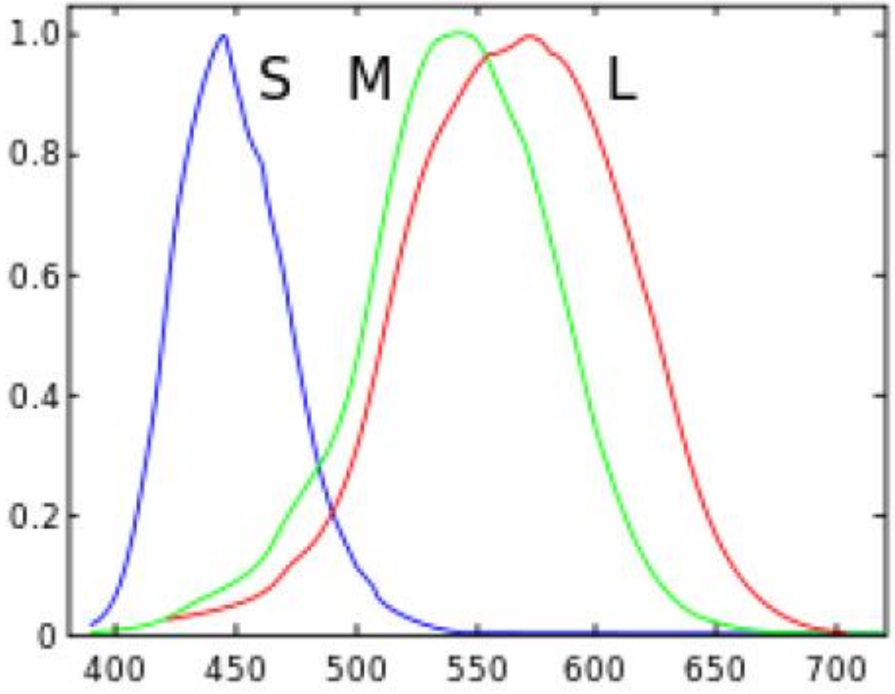
Some people might have 4 cone-types!

Some might have just 2!

Color

- Color vision has evolved over millions of years.

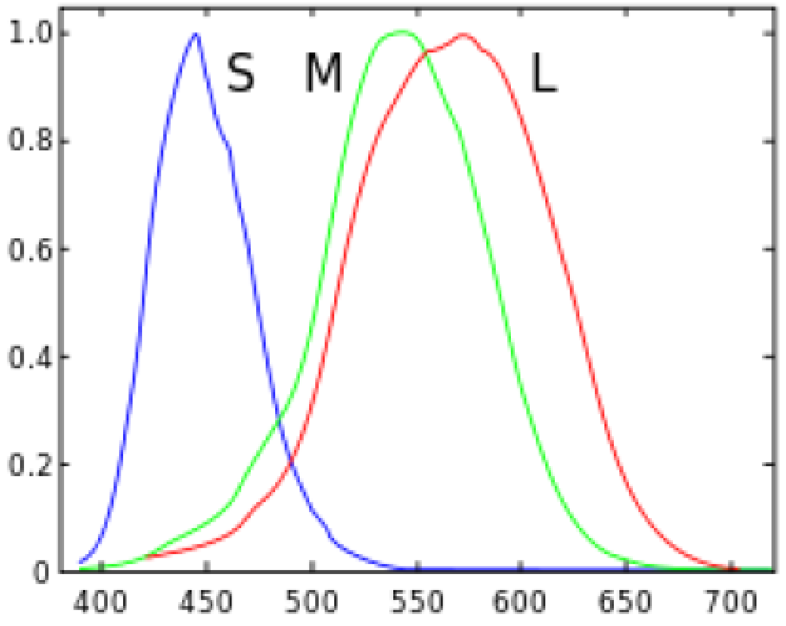




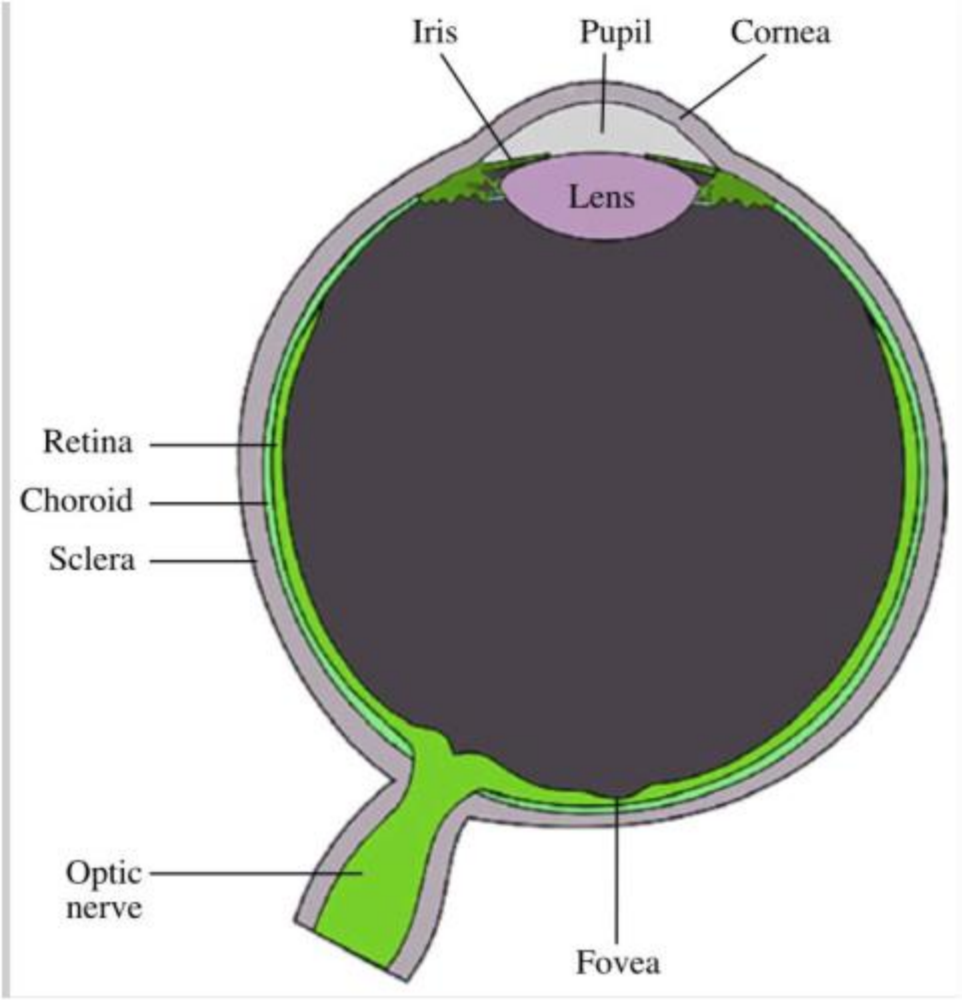
Color

- 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



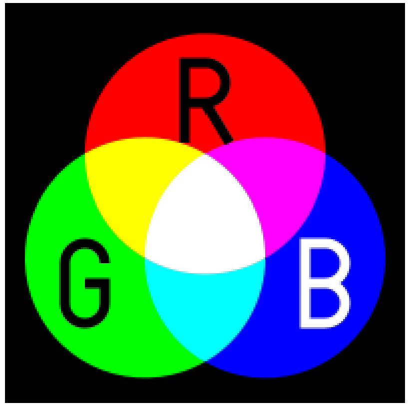
There are three different types of color-sensitive cones corresponding to (roughly)

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

6-7 million cones
120 million rods

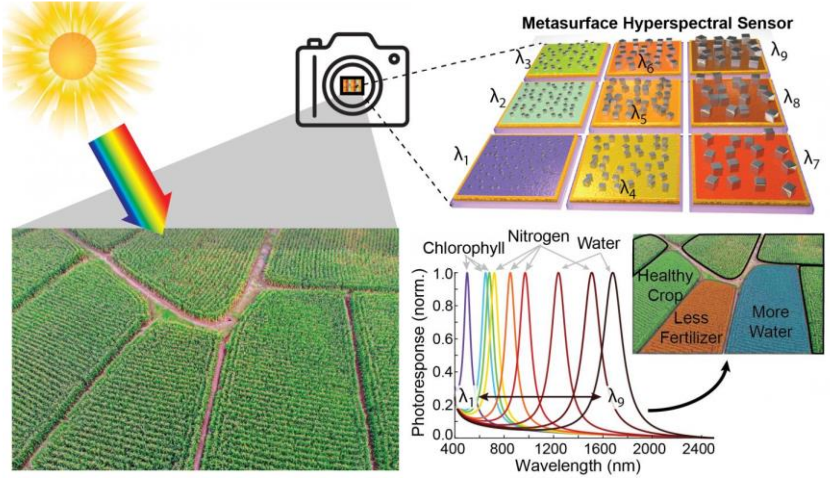
Credit: Klette, 2012.

Demo: Color is in your head !

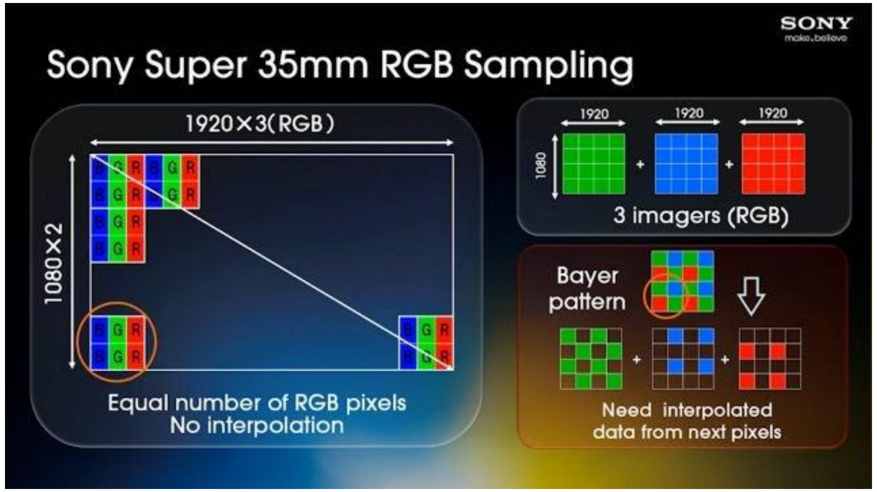


A large list of visual sensors

Monochrome cameras



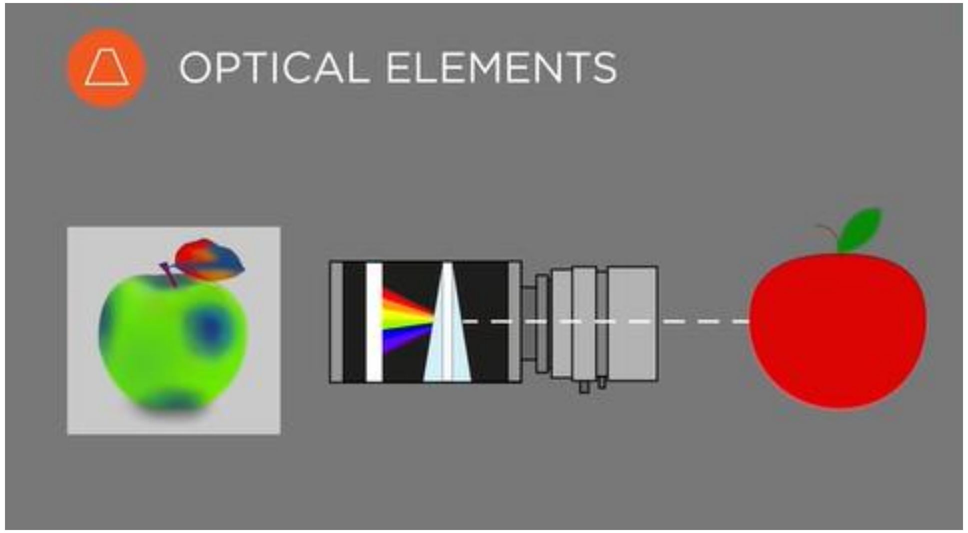
MultiSpectral cameras

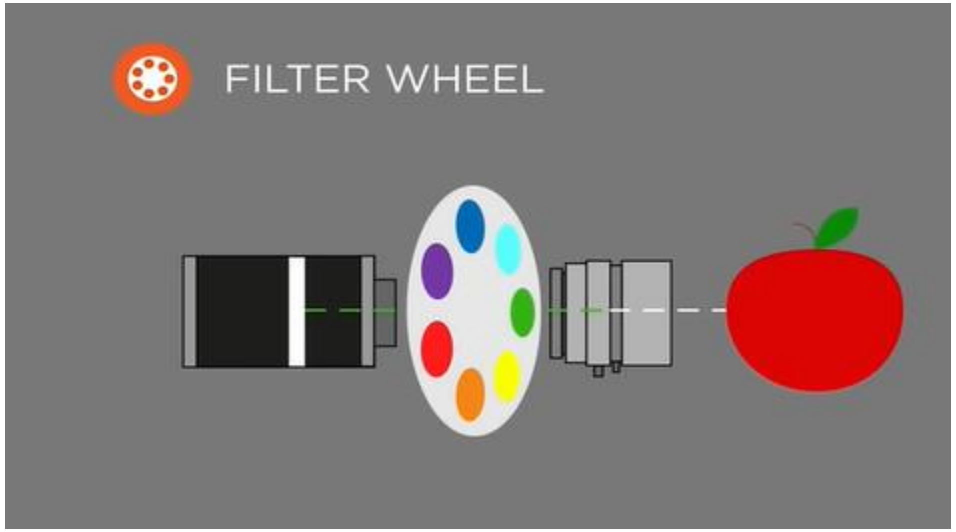


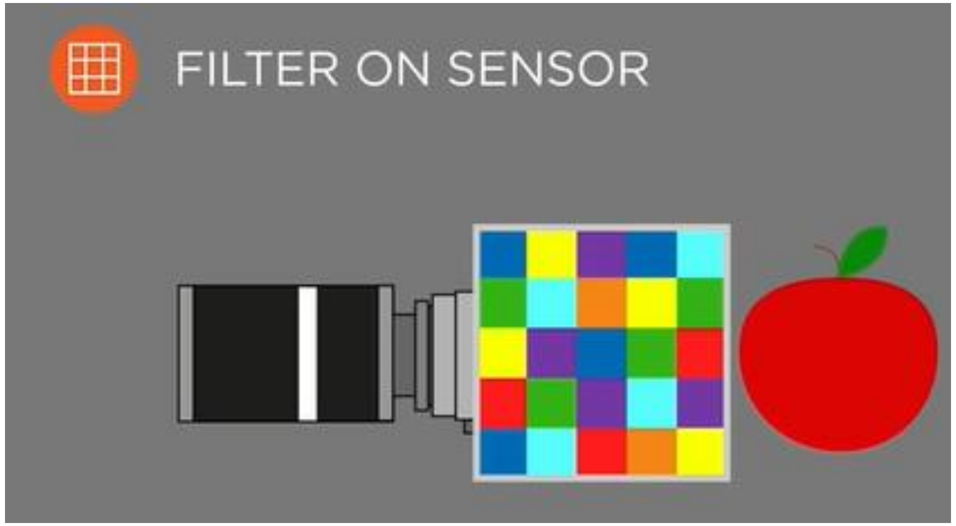
RGB cameras

Lidar / Time of flight cameras

Multispectral cameras

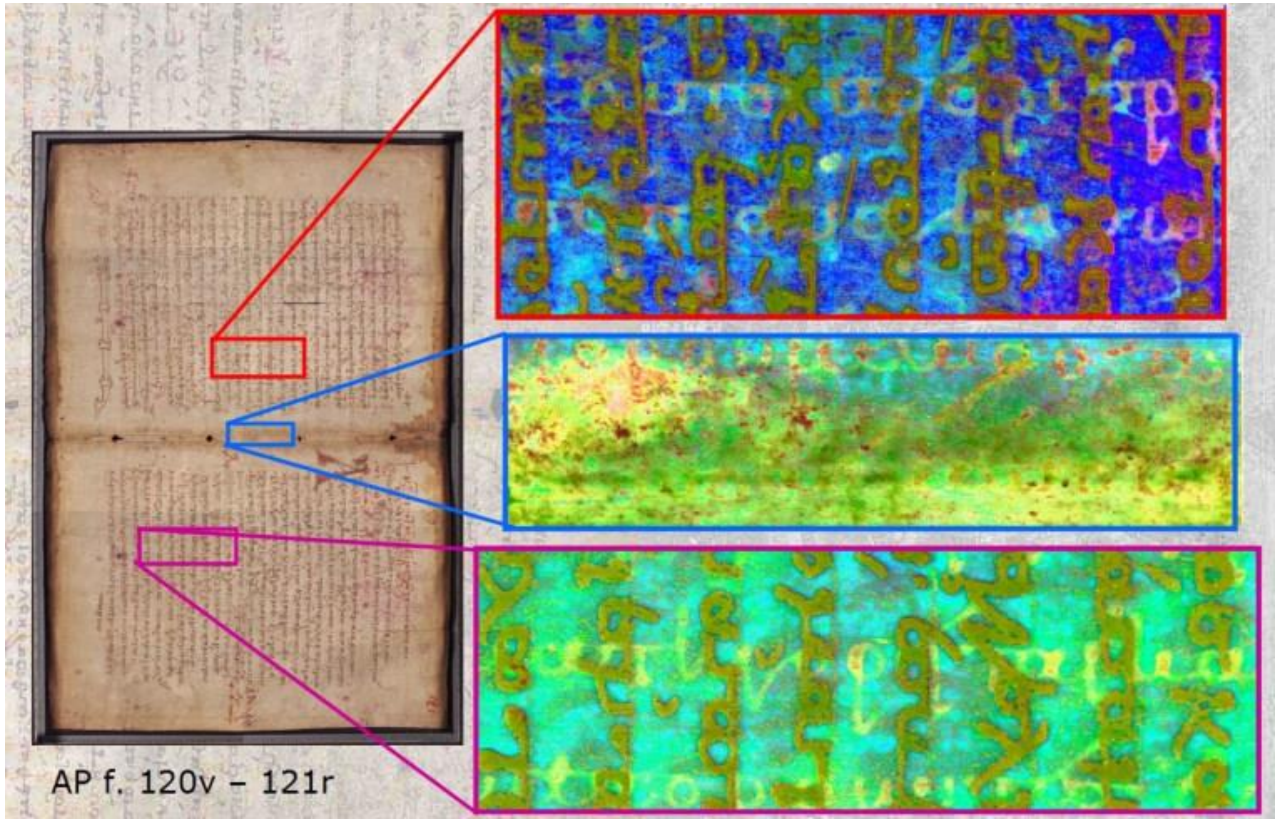






Multispect… and Where Are They Used

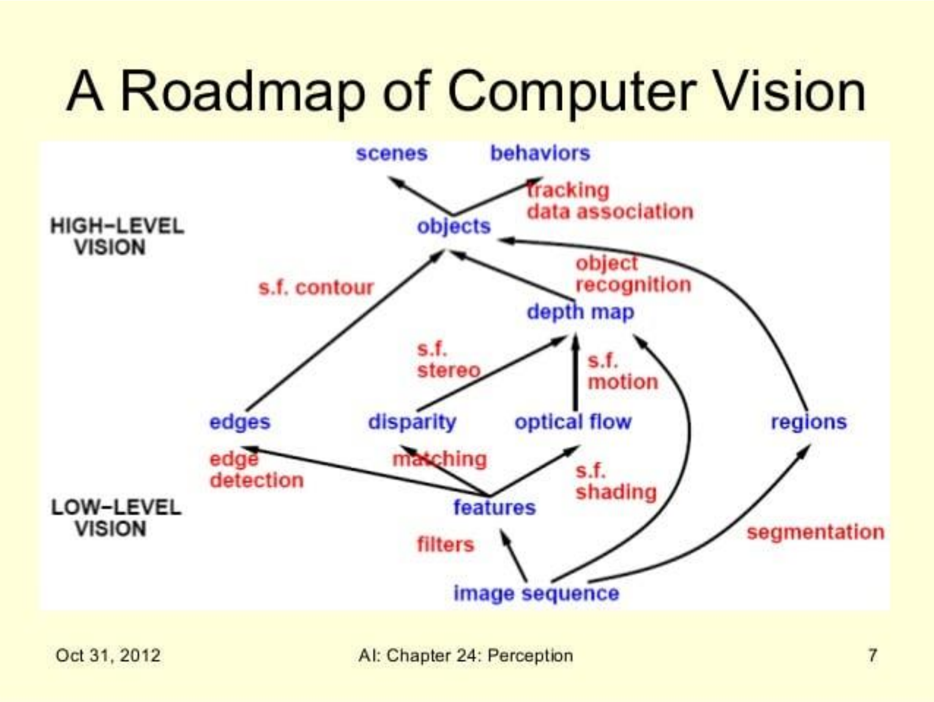
MultiSpectral cameras



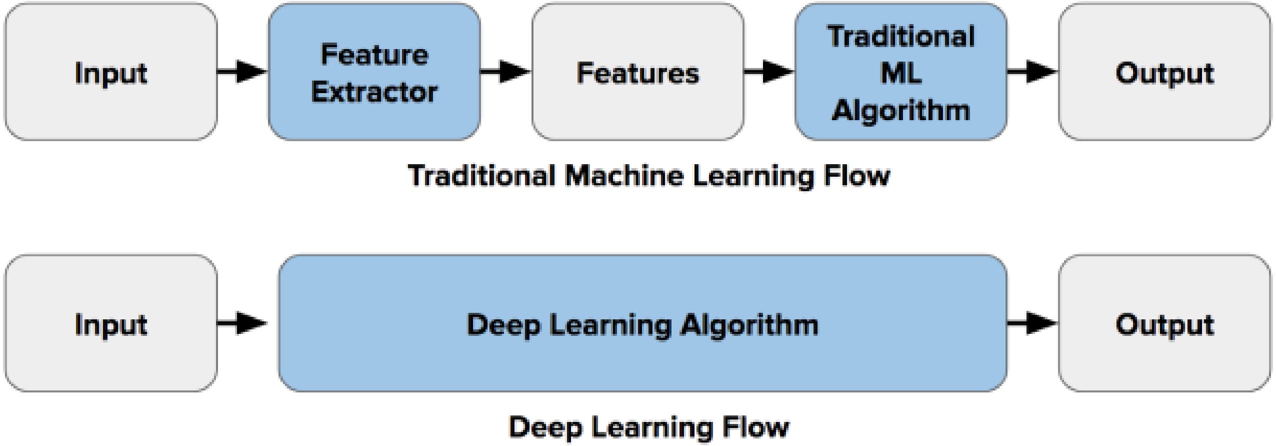
[Multispectral Imaging: New Technology Resurrects Centuries-Old Texts (nbcnews.com)]

- PCB inspection
- Skin characterization
- Food inspection
- Agriculture
  - Analyzing crops
- Military

Computer Vision 2010 vs Today



In 2010



Applications

- Self Driving cars
- Biometric verification
- Healthcare (Gauss Surgical)
- Medical diagnostics
- Retail (amazon go)
- Industrial monitoring
- Agriculture (remove weeds)

Google's DeepMind Beats Doctors at Detecting 50 Eye Diseases Just by Looking at Scans

By KC Cheung

Last Modified Date - May 27, 2020

Share

Google's DeepMind Beats Doctors at Detecting 50 Eye Diseases Just by Looking at Scans - Algorithm-X Lab (algorithmxlab.com)

Forbes

Walmart Expands Use of Bossa Nova's Robots from 50 to 350 Stores

Steve Banker Contributor
Transportation
I cover logistics and supply chain management.

Follow

Walmart Expands Use of Bossa Nova's Robots from 50 to 350 Stores (forbes.com)

This article is more than 2 years old.

Bossa Nova Robotics at Being Used at Walmart Stores BOSSA NOVA ROBOTICS

It was recently announced that Walmart has expanded its use of autonomous mobile robots (AMRs) used for real-time, on-shelf product data. Walmart will put these AMRs in 350 stores. The robots come from Bossa Nova Robotics.

Cargill brings facial recognition capability to farmers through strategic equity investment in Cainthus

Sensors give farmers clear picture of animal health and well-being

(MINNEAPOLIS) January 31, 2018 — Cargill and Cainthus, a Dublin-based machine vision company, are reshaping how animal producers make decisions for their livestock through a strategic partnership that will bring facial recognition technology to dairy farms across the world. The deal includes a minority equity investment from Cargill. Terms were not disclosed.

Cainthus uses breakthrough predictive imaging to monitor the health and wellbeing of livestock. Their proprietary software uses images to identify individual animals based on hide patterns and facial recognition, and tracks key data such as food and water intake, heat detection and behavior patterns. The software then delivers analytics that drive on-farm decisions that can impact milk production, reproduction management and overall animal health.

"We are enthused about what this partnership will mean for farmers across the world," said David Hunt, president and co-founder, Cainthus. "Cargill is a natural partner for us, given their focus on bringing a world-class digital capability to the market and their understanding of how technology will truly help farmers succeed. We think this partnership will be a game changer for farmers because it will allow them to efficiently scale their business."

Cargill brings facial recognition capability to farmers through strategic equity investment in Cainthus | Cargill

The Machine
Making sense of AI

AI startup Gather uses drones and computer vision for warehouse inventory

Khari Johnson @kharijohnson August 15, 2019 12:10 PM

AI startup Gather uses drones and computer vision for warehouse inventory | VentureBeat

Disney's facial recognition AI watches you watch movies | Daily Mail Online

Disney reveals AI that can monitor cinemagoer's faces to reveal what they REALLY thought of the movie

- Disney is using neural networks to learn about stereotypical reactions to films
- After observing viewers for only a few minutes, it can predict facial expressions for the remainder of the movie
- The system 'taught' itself to recognize smiling and laughing all on its own

By SAGE LAZZARO FOR DAILYMAIL.COM
PUBLISHED: 17:58 EDT, 24 July 2017 | **UPDATED:** 11:10 EDT, 25 July 2017

Share 90 shares View comme

What changed?

- Emergence of deep learning
- Advancement in hardware
- Availability of large-scale data
  - ImageNet
  - OpenImages
  - YFCC100M
  - Youtube-8M
  - Kinetics
  - AVA
  - ...

Hardware

Train

CPU

- Small models
- Small datasets
- Useful for design space exploration

GPU

- Medium-to-large models, datasets
- Image, video processing
- Application on CUDA or OpenCL

TPU

- Matrix computations
- Dense vector processing
- No custom TensorFlow operations

FPGA

- Large datasets, models
- Compute intensive applications
- High performance, high perf./cost ratio

Inference

Device	Pros	Cons
CPU	Cost effective, fit for general purpose, powerful cores, high memory capacity	Don't fully exploit parallelism, low throughput performance
GPU	High throughput performance, a good fit for modern architectures (ConvNets)	Expensive, energy-hungry, has IO latency, memory limitations
Custom AI Chips (ASIC, SoC)	Potential to significantly boost inference performance	Expensive and hard to develop
FPGA	Chip, energy efficient, flexible	Extremely difficult to use, not always better than CPU/GPU

How to Choose Hardware for Deep Learning Inference | Deci

CPU, GPU, FPGA or TPU: Which one to choose for my Machine Learning training? – InAccel

Questions?