

# CAP 4453

# Robot Vision

Dr. Gonzalo Vaca-Castaño  
[gonzalo.vacacastano@ucf.edu](mailto:gonzalo.vacacastano@ucf.edu)



# Administrative details

- AI group announcement
- REU announcement
- Assignment Zero:
  - Due tomorrow
- Assignment 1:
  - Deadline: Next Friday

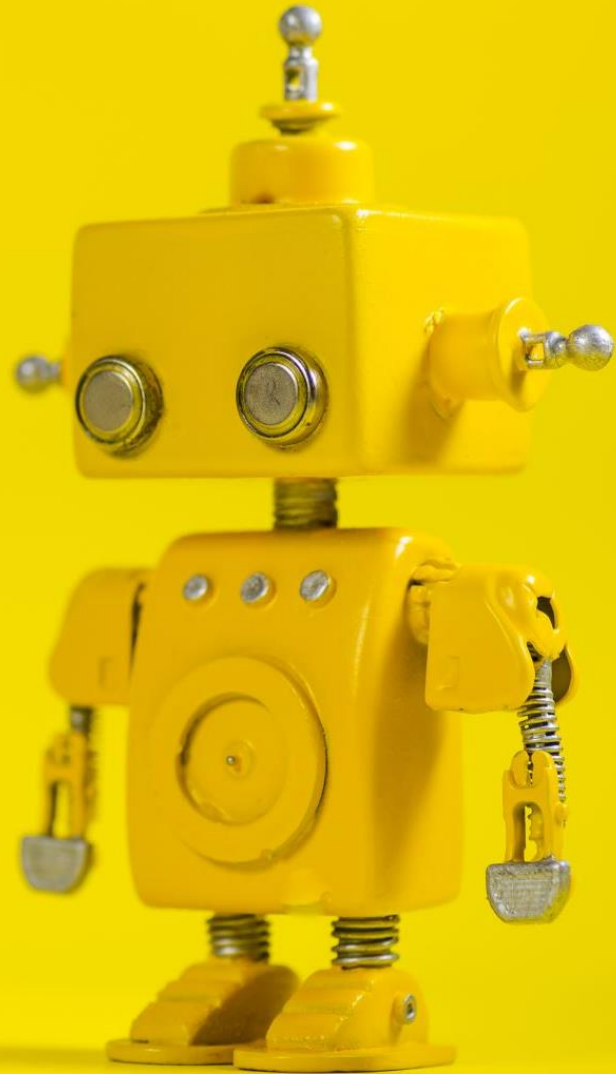


# Questions?



# Credits

- Some of this slides comes from:
  - Yogesh S Rawat (UCF)
  - Noah Snavelly (Cornell)
  - Ioannis (Yannis) Gkioulekas (CMU)
  - Mubarak Shah (UCF)
  - S. Seitz
  - James Tompkin
  - Ulas Bagci
  - L. Lazebnik



# Robot Vision

## 2. Basics of Images

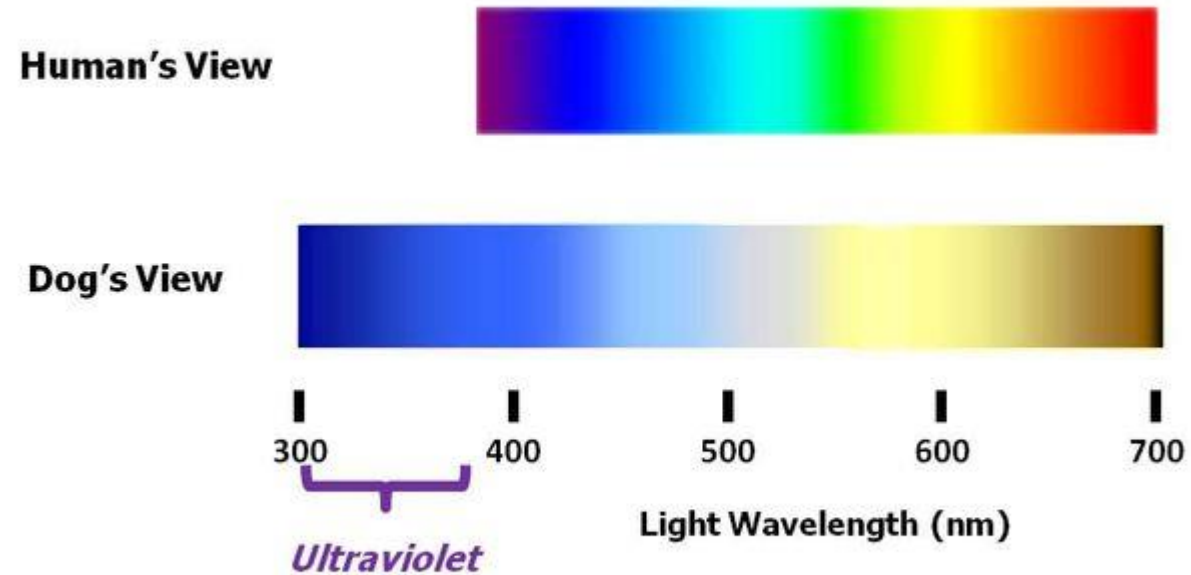


# Why there are 2 greens at BAYER

- As a cost-effective alternative, many manufacturers produce cameras that coat the CCD elements of their area arrays with a Bayer filter, named after Bryce E. Bayer, a scientist at Eastman Kodak who patented the concept in 1976. In Bayer-based color imagers, pixels on the image sensor are covered with a mosaic of red, green, and blue transmissive filters. In the Bayer pattern, 50% of the pixels are green, 25% are red, and 25% are blue (see image).
- Bayer noticed that the human eye obtains most of its sharpness information from green light, which is why the pattern has more green pixels. By using two green-filtered pixels for every red or blue, the Bayer pattern is designed to maximize the perceived sharpness in the luminance channel, composed mostly of green data

# From last class

## How the spectrum appears to people and dogs



# From last class



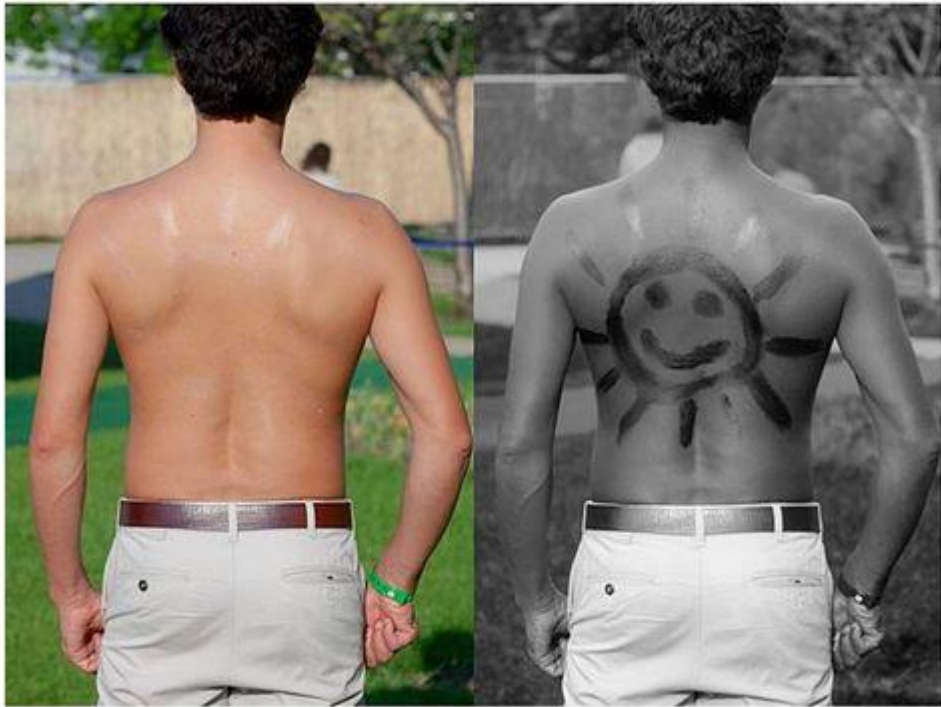
**Human View**  
(No UV Sensitivity)



**Dog View**  
(Some UV Sensitivity)



# From last class



**Sensitivity to UV makes targets that block or reflect these short wavelengths visible (like a drawing in sun screen lotion)**

# From last class

**The ability to see ultraviolet (UV) helps guide bees to the pollen containing parts of flowers**



**Human View  
(No UV Sensitivity)**



**Bee View  
(UV Sensitivity)**

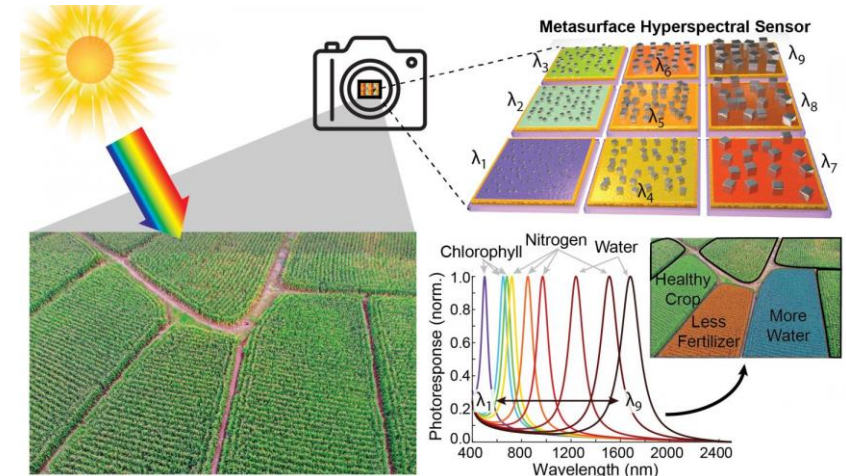
<https://www.psychologytoday.com/us/blog/canine-corner/201604/can-dogs-see-in-ultraviolet>



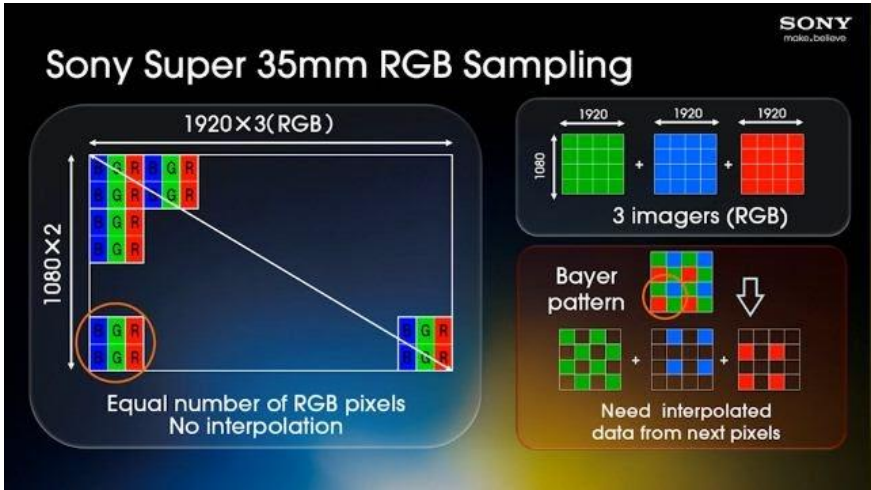
# A large list of visual sensors



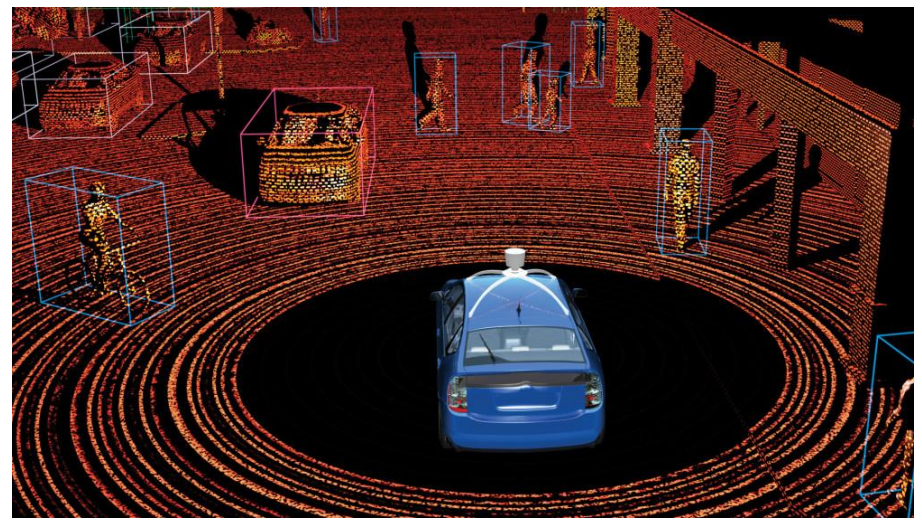
Monochrome cameras



MultiSpectral cameras

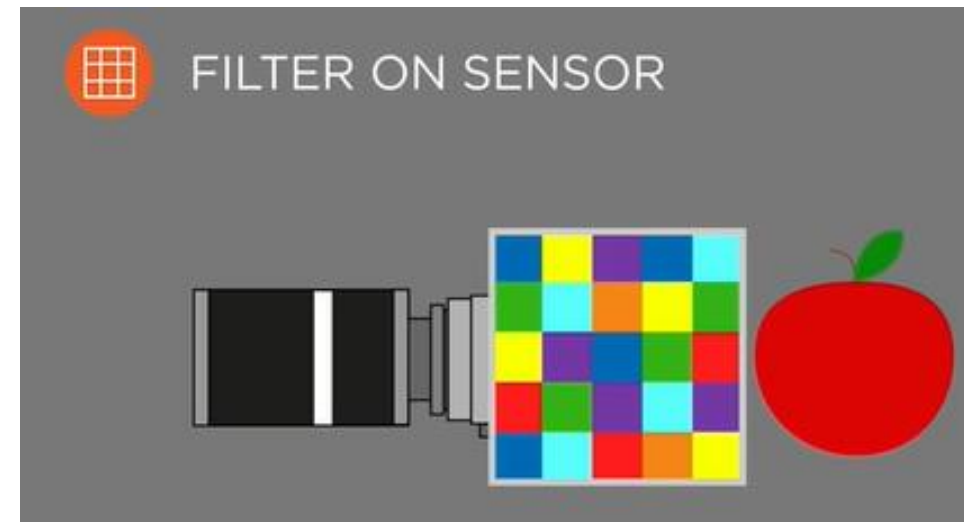
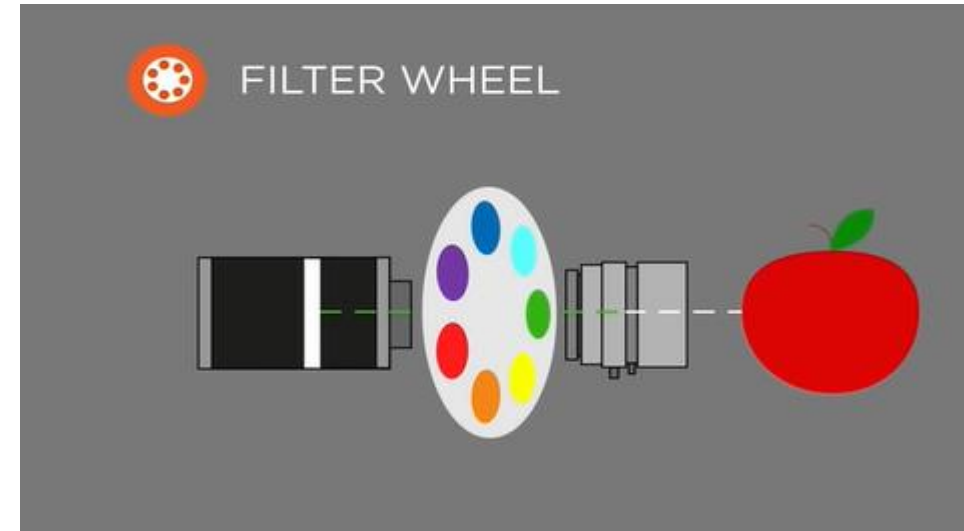
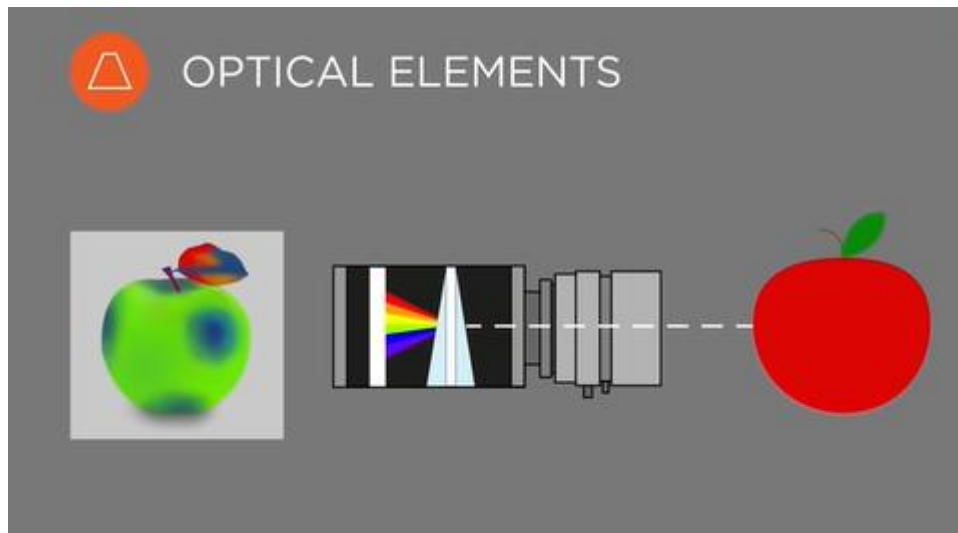


RGB cameras



Lidar / Time of flight cameras

# Multispectral cameras

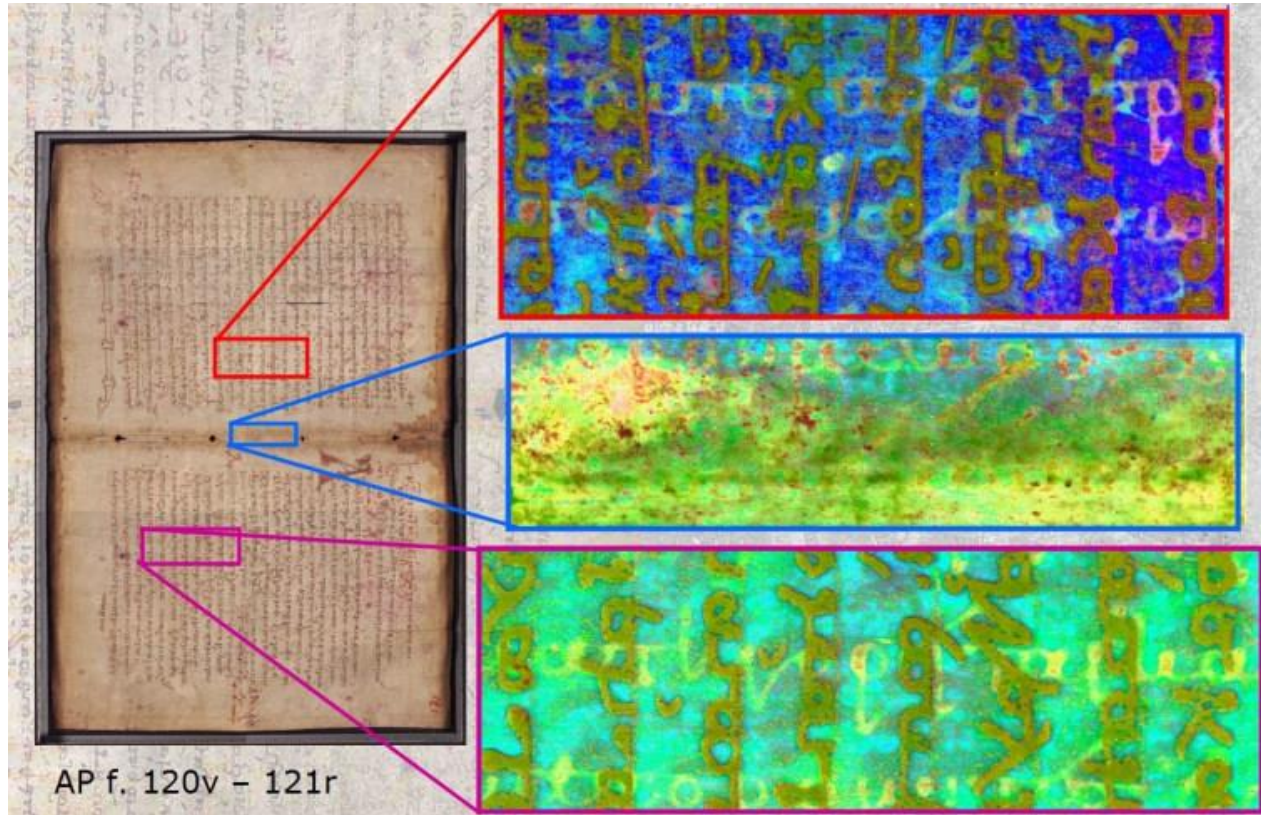


[Multispectral Cameras and Where Are They Used](#)





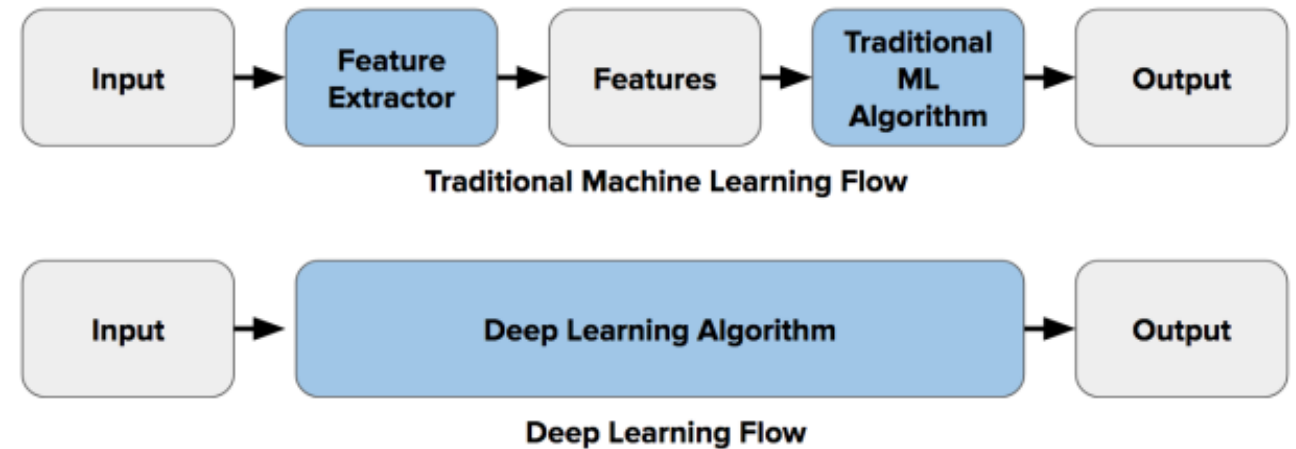
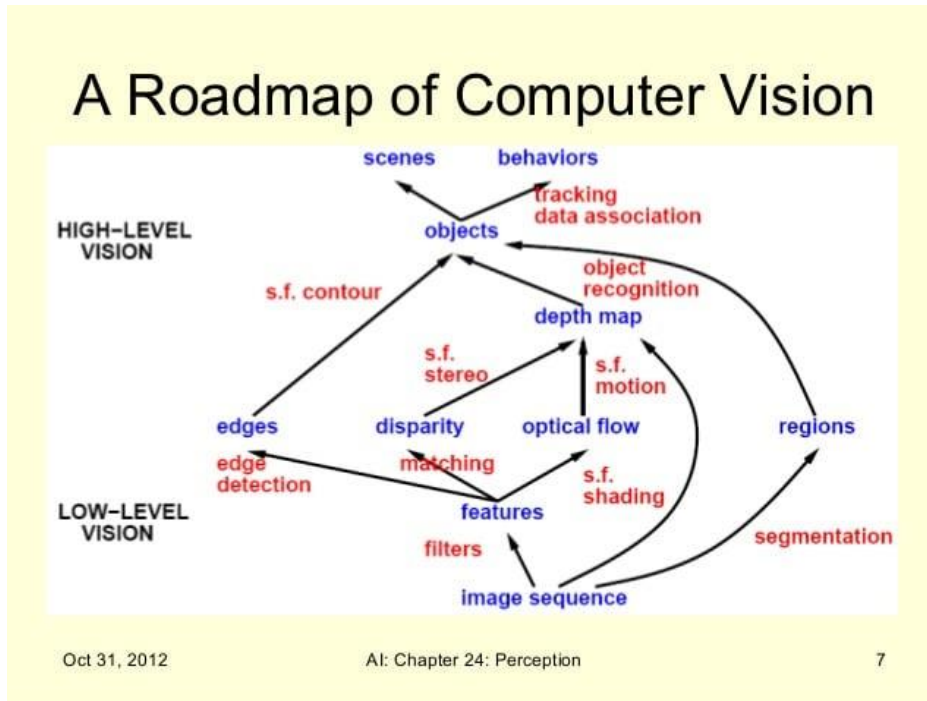
# MultiSpectral cameras



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

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

# 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)

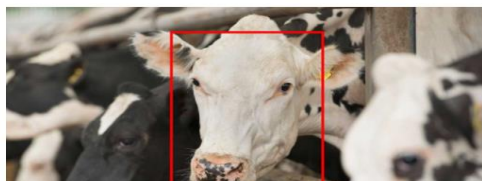
## Cargill brings facial recognition capability to farmers through strategic equity investment in Cainthus

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](#)

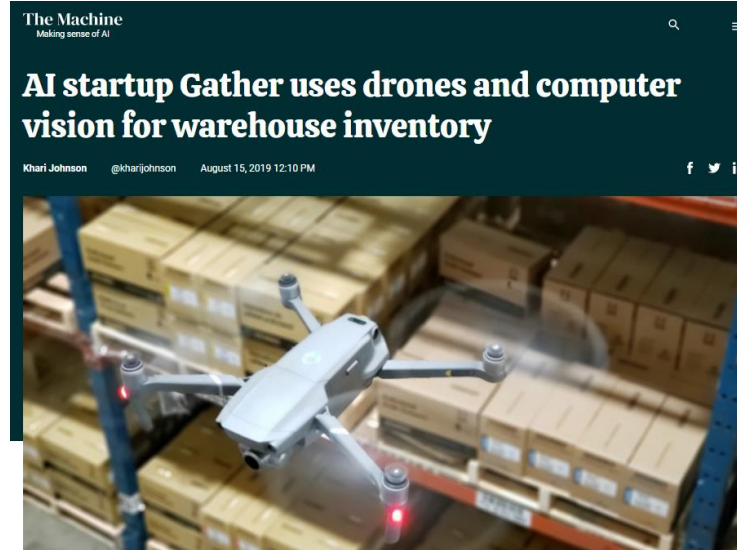
## Google's DeepMind Beats Doctors at Detecting 50 Eye Diseases Just by Looking at Scans

By KC Cheung

Last Modified Date - May 27, 2020



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



[AI startup Gather uses drones and computer vision for warehouse inventory | VentureBeat](#)

## 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.

[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



View comment

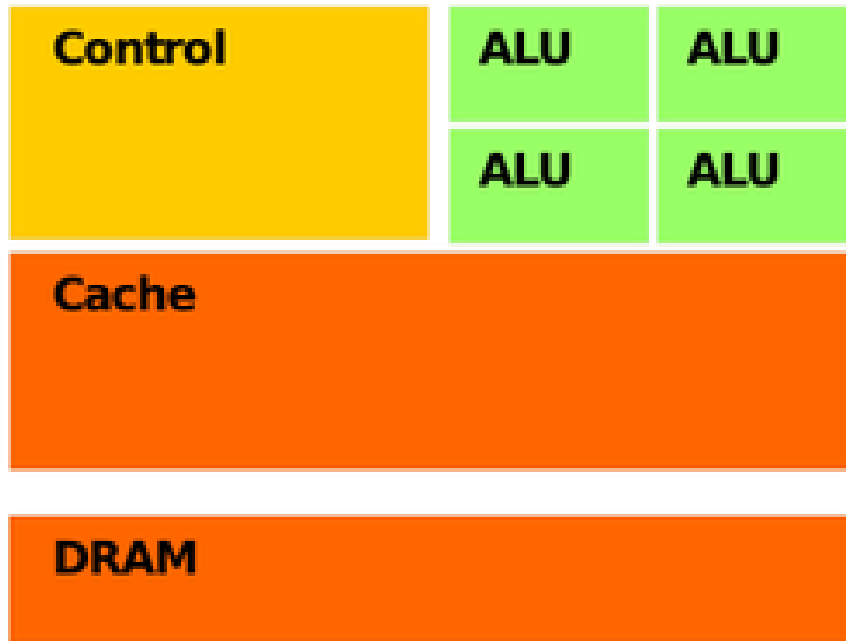
# What changed?

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

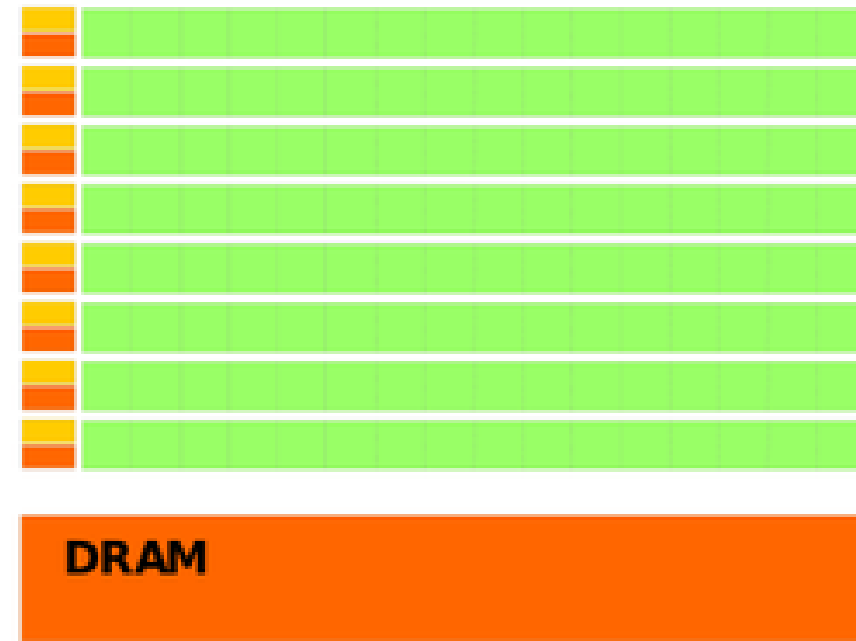




# CPU vs GPU



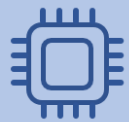
**CPU**



**GPU**

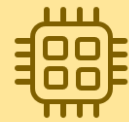
# Hardware

## Train



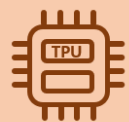
### CPU

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



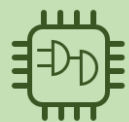
### GPU

- 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



# Outline

- Image as a function
- Extracting useful information from Images
  - **Histogram**
- Color spaces
  - RGB
  - HUE
  - CIE
- Homework 1

# Digitization

- Computers use discrete form of the images
- The process transforming continuous space into discrete space is called **digitization**

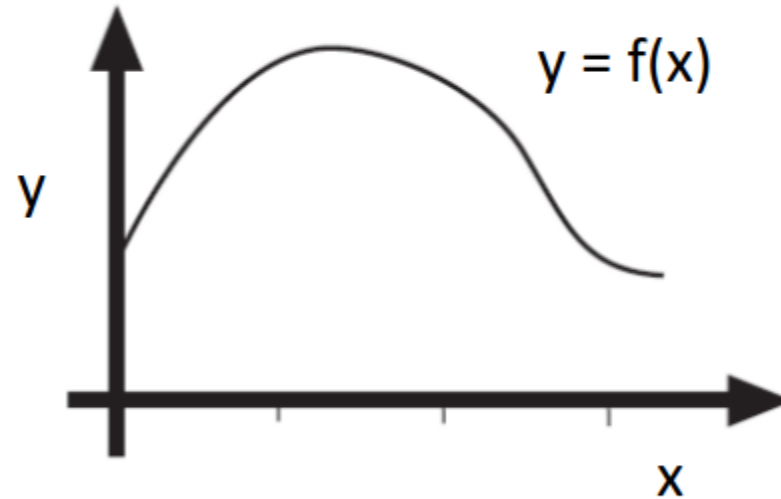


# Digitization

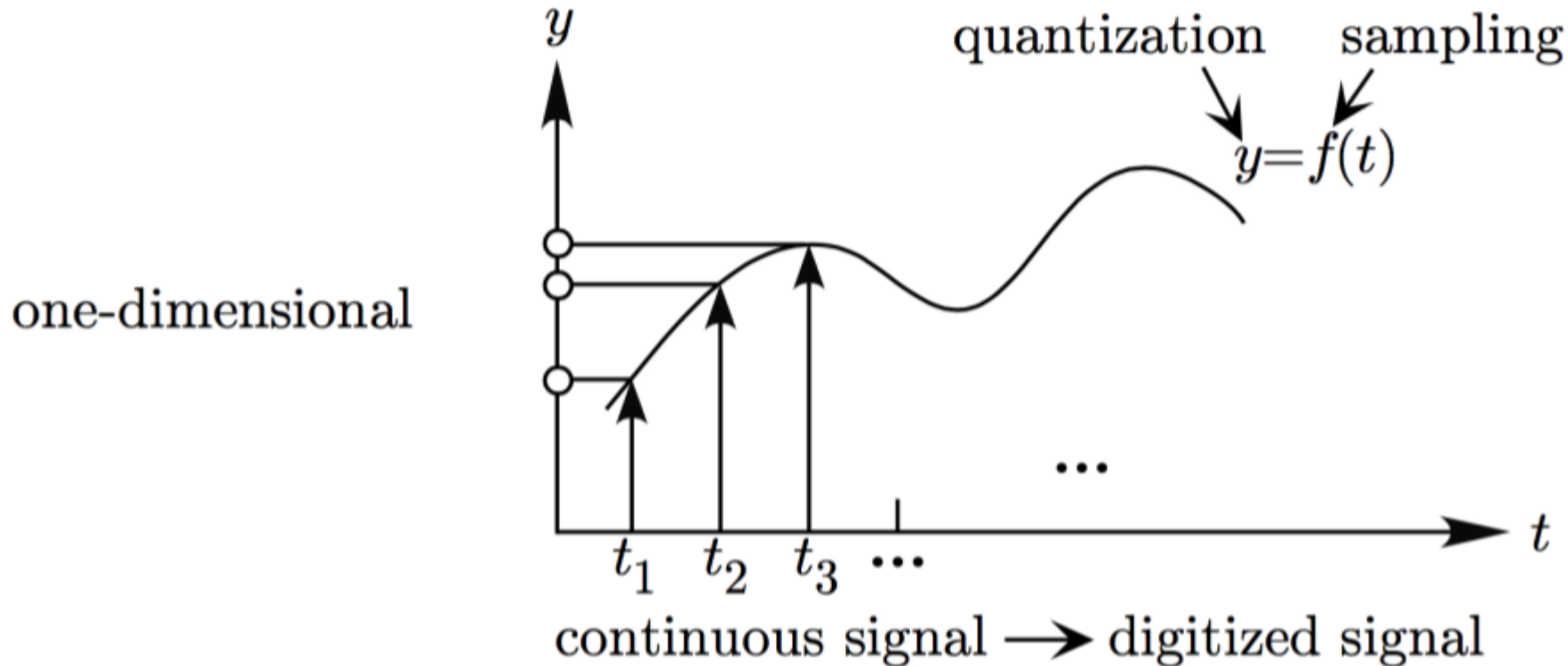
- Function

$$y = f(x)$$

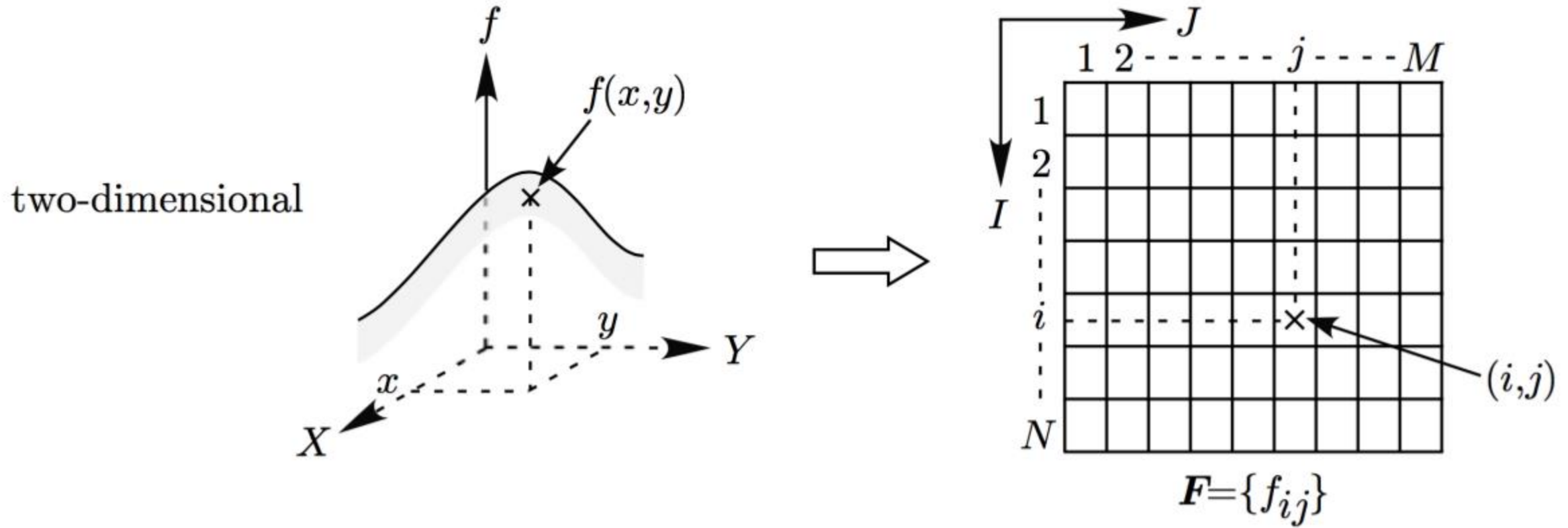
- Domain of a function
- Range of a function
- Sampling
  - Discretization of domain
- Quantization
  - Discretization of range



# Digitization of 1D function

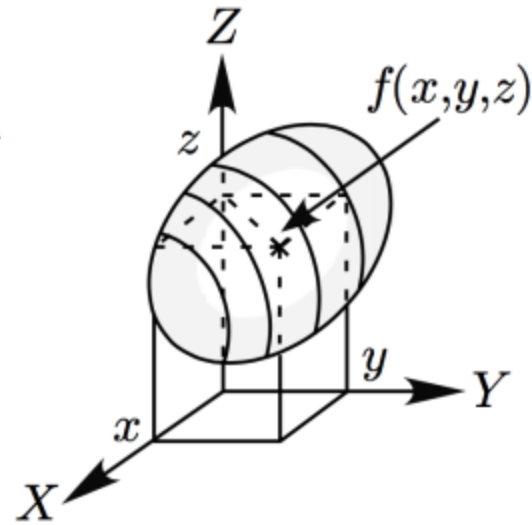


# Digitization of 2D function

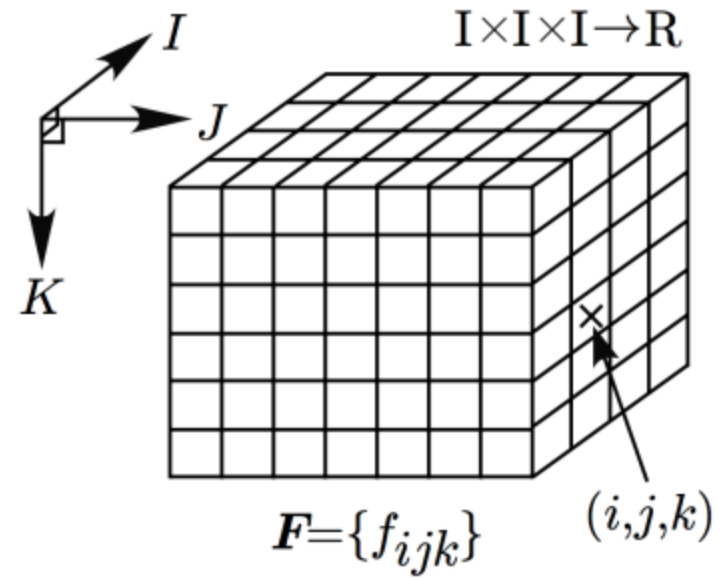


# Digitization of 3D function

three-dimensional



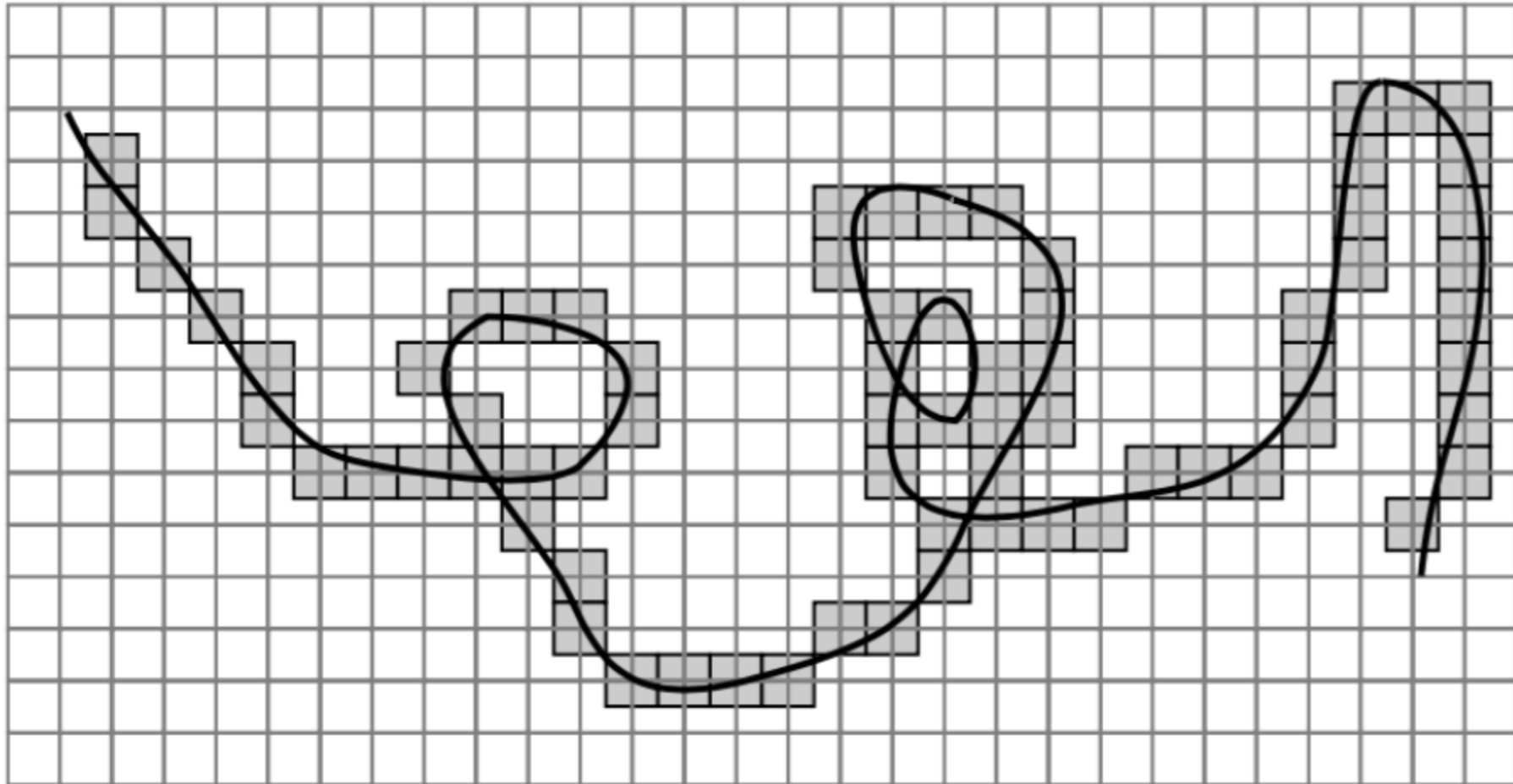
continuous image



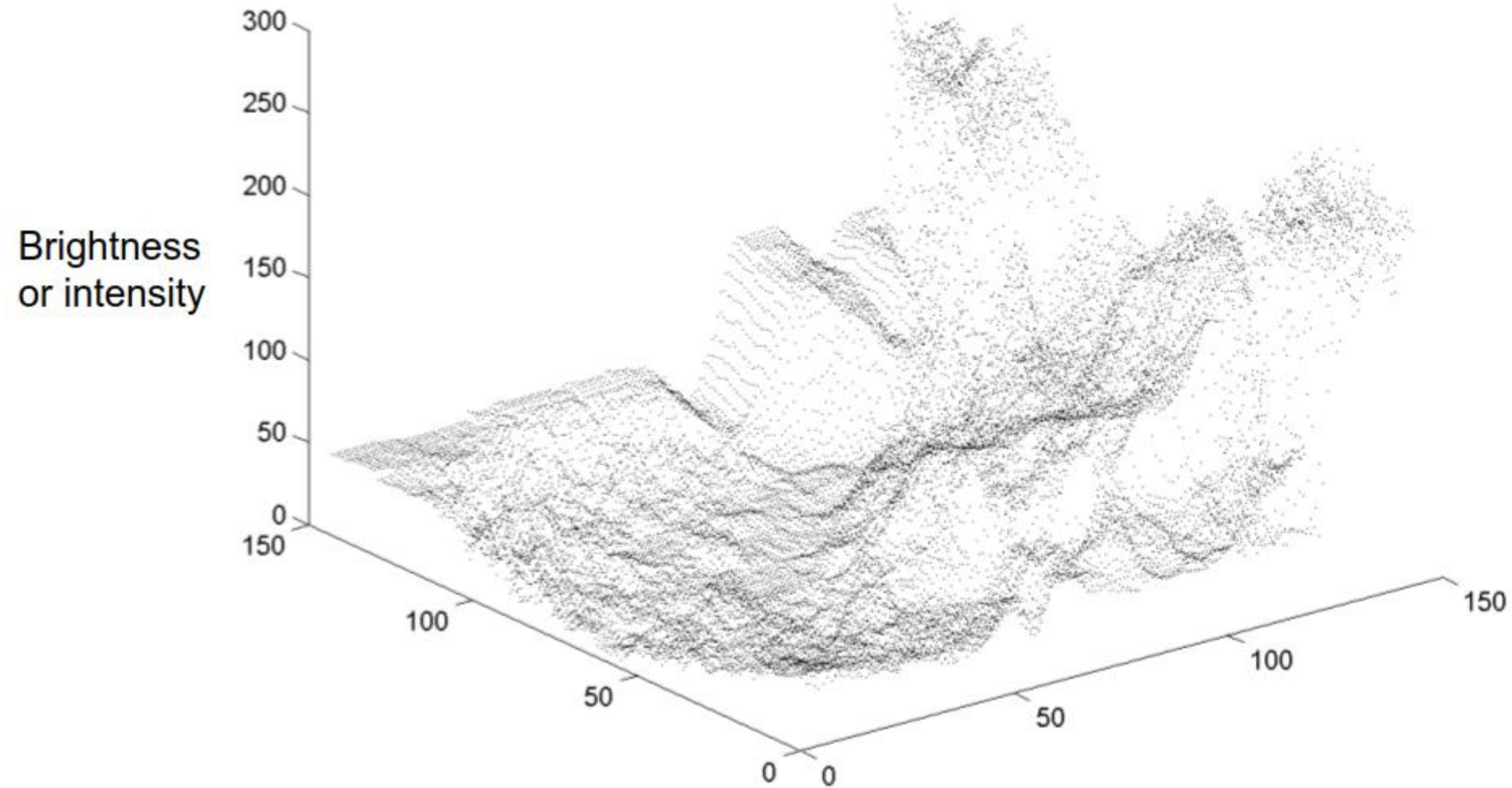
digitized image



# Digitization of an arc

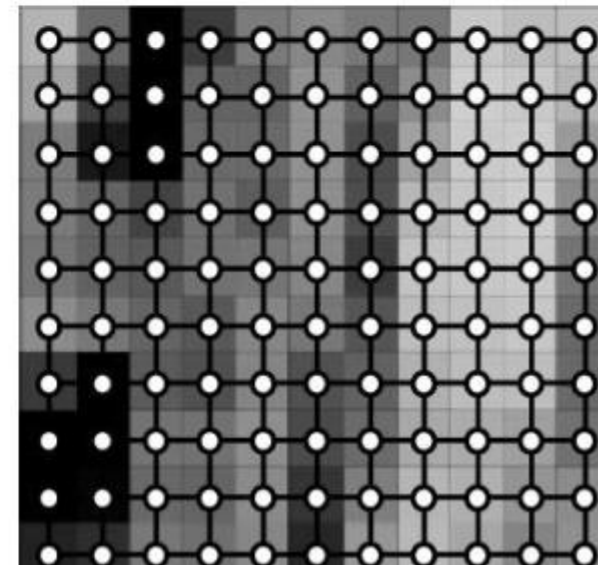
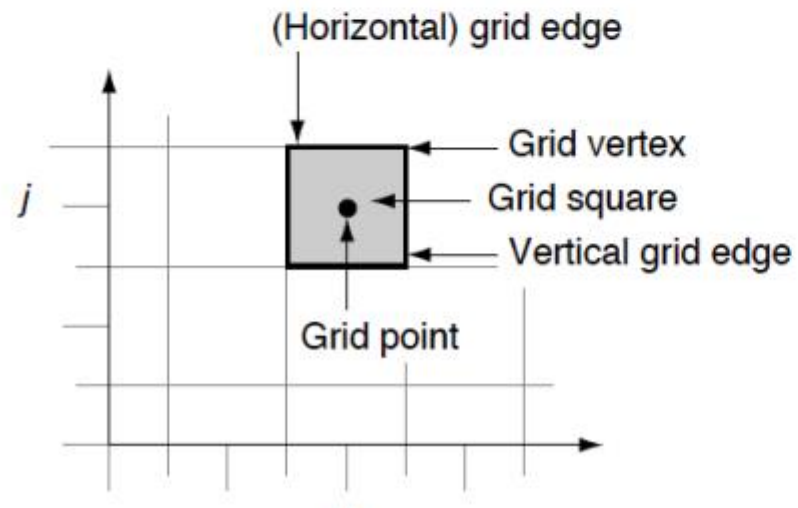


# Gray scale digital image



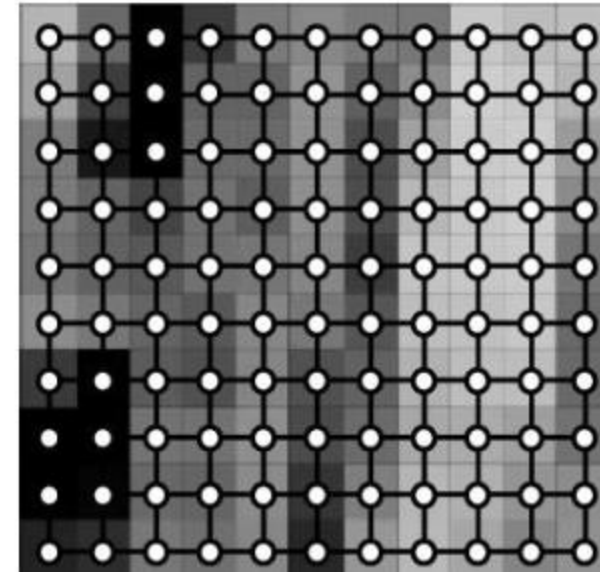
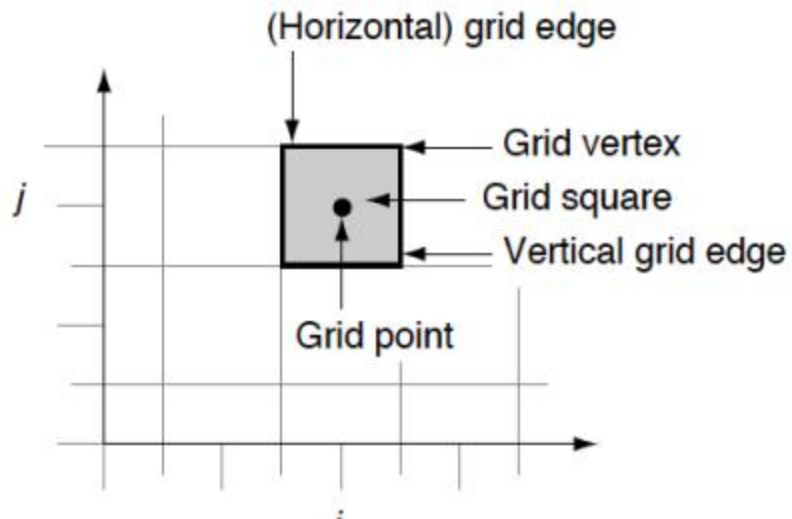
# Definition

- An image  $P$  is a function defined on a (finite) rectangular subset  $G$  of a regular planar orthogonal array.
- $G$  is called (2D) **grid**, and **an element of  $G$  is called a pixel**.
- $P$  assigns a value of  $P(p)$  to each  $p \in G$

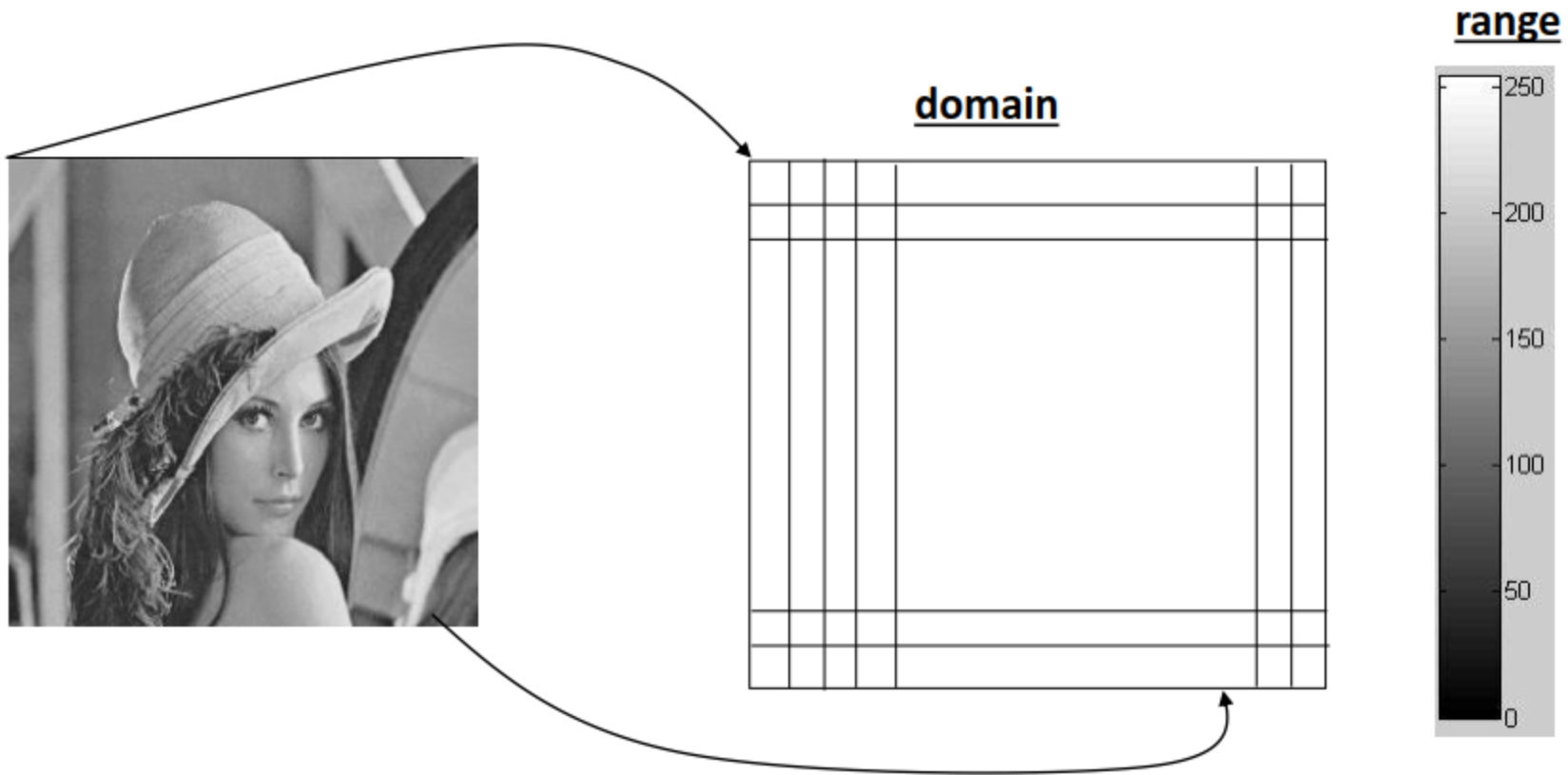


# Definition

- Pictures are not only sampled
- They are also quantized
  - they may have only a finite number of possible values
  - i.e., 0 to 255, 0-1, ...



# Digitization





# Sampling



# Quantization



Original  
(256 colors)



8 colors



4 colors

# About the picture



[Lena Forsen - playmate, who became the "mother" of JPEGs \(fotoblogia.pl\)](http://fotoblogia.pl)



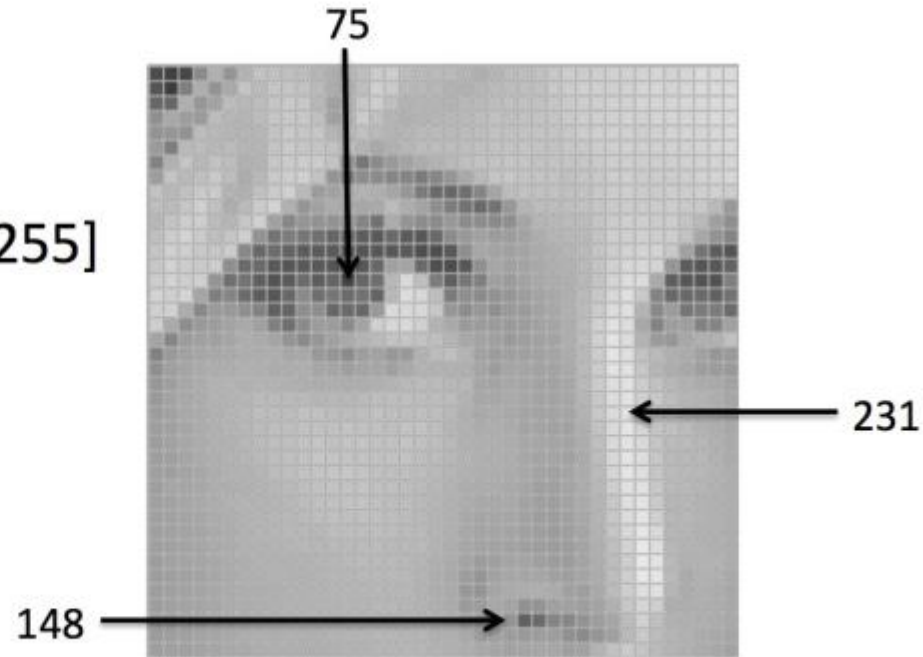
# Resolution

- Also a display parameter
  - defined in **dots per inch (DPI)** or
  - measure of spatial pixel density
  - standard value for recent screen technologies is 72 dpi.
  - Recent printer resolutions are in 300 dpi and/or 600 dpi.



# Gray scale image

- An image contains discrete number of pixels
  - A simple example
  - Pixel value:
    - “grayscale”  
(or “intensity”): [0,255]



# Color image

- An image contains discrete number of pixels

- A simple example

- Pixel value:

- “grayscale”

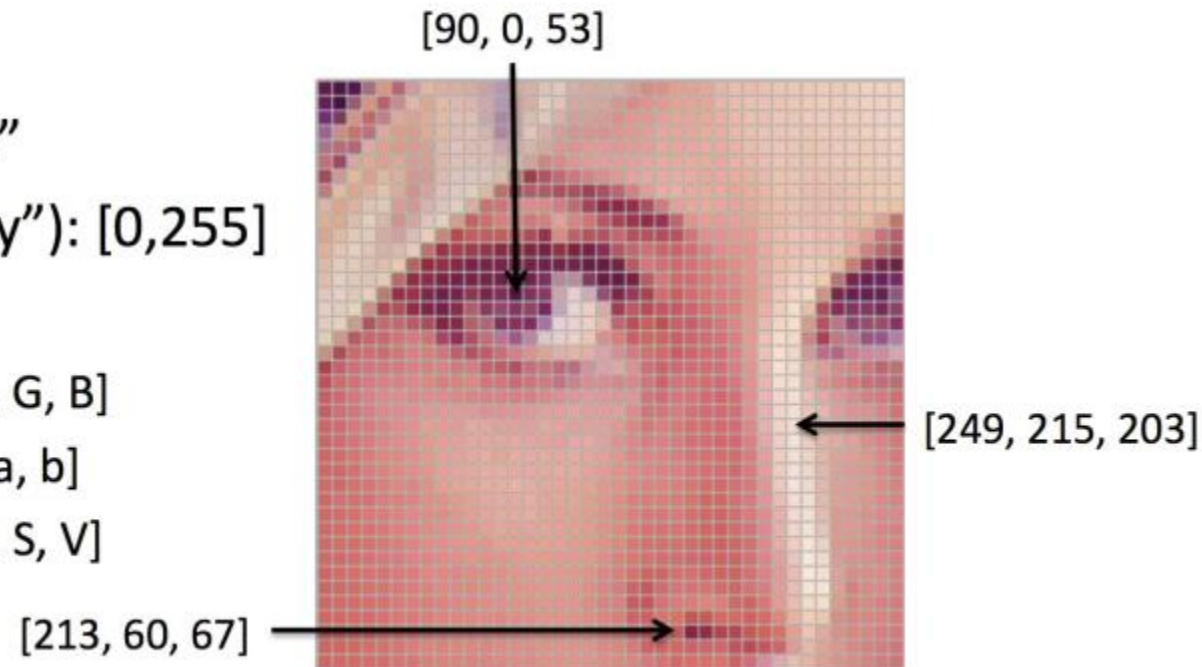
(or “intensity”): [0,255]

- “color”

- RGB: [R, G, B]

- Lab: [L, a, b]

- HSV: [H, S, V]



Source: F.F. Li

# RGB Channels





# RGB Channels

640



480

How many pixels do you need to represent this image?



# RGB Channels

640



480

How many bytes do you need to represent this image?



# RGB Color Space

Compression technique

$$R = \int_{300}^{830} S(\lambda) R(\lambda) d\lambda$$

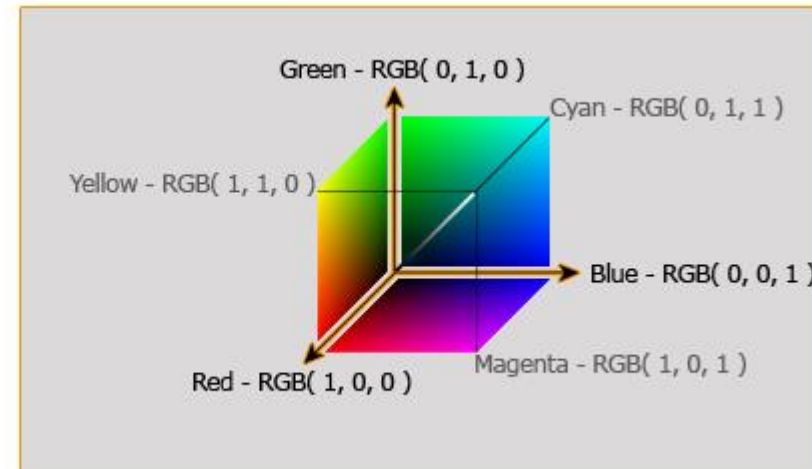
$$G = \int_{300}^{830} S(\lambda) G(\lambda) d\lambda$$

$$B = \int_{300}^{830} S(\lambda) B(\lambda) d\lambda$$

$S(\lambda)$  is the light spectrum,

$R(\lambda)$ ,  $G(\lambda)$  and  $B(\lambda)$  are the sensitivity functions

Color Cube  
RGB Color Format



www.equasys.de

- These are colours with different spectra but with same perceptual values
- RGB colour space is the basic colour space
- Device-dependant colour space

# RGB Color Space

Compression technique

$$R = \int_{300}^{830} S(\lambda) R(\lambda) d\lambda$$

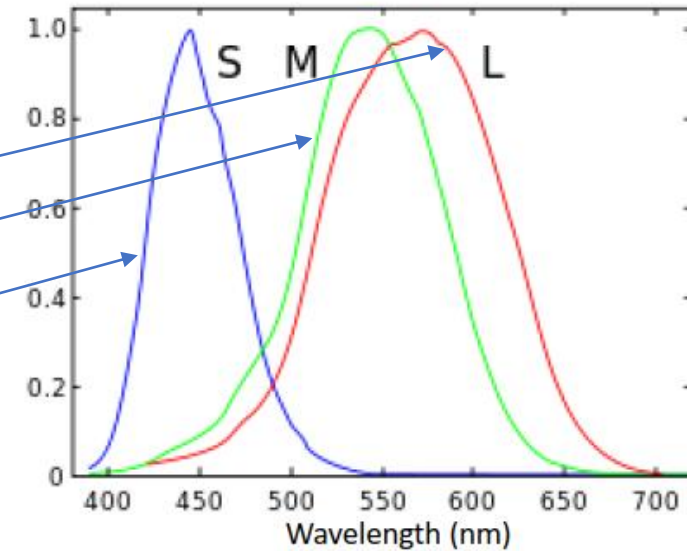
$$G = \int_{300}^{830} S(\lambda) G(\lambda) d\lambda$$

$$B = \int_{300}^{830} S(\lambda) B(\lambda) d\lambda$$

$S(\lambda)$  is the light spectrum,

$R(\lambda)$ ,  $G(\lambda)$  and  $B(\lambda)$  are the sensitivity functions

Color Cube  
RGB Color Format



Red - RGB( 1, 0, 0 )

Magenta - RGB( 1, 0, 1 )

- These are colours with different spectra but with same perceptual values
- Device-dependant colour space
- RGB colour space is the basic colour space



# RGB Color Space

Compression technique

$$R = \int_{300}^{830} S(\lambda) R(\lambda) d\lambda$$

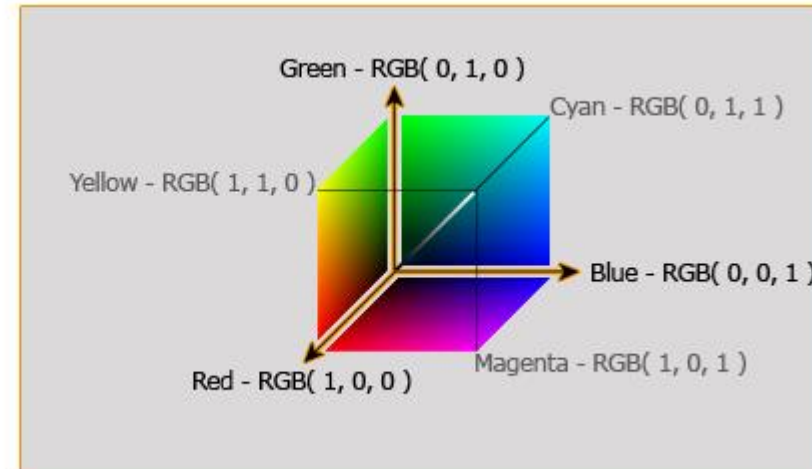
$$G = \int_{300}^{830} S(\lambda) G(\lambda) d\lambda$$

$$B = \int_{300}^{830} S(\lambda) B(\lambda) d\lambda$$

$S(\lambda)$  is the light spectrum,

$R(\lambda)$ ,  $G(\lambda)$  and  $B(\lambda)$  are the sensitivity functions

Color Cube  
RGB Color Format



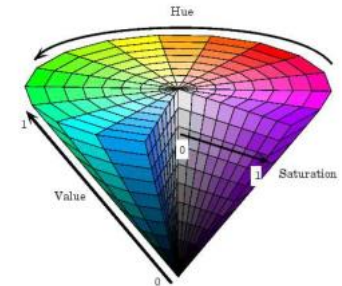
www.equasys.de

## Disadvantages

- a high correlation between its components
  - about 0.78 for rBR (cross correlation between the Band R channel)
  - 0.98 for rRG
  - 0.94 for rGB
- It is psychologically non-intuitive
- Perceptual non-uniformity (add a value have different effect for every color)

# Phenomenal color spaces

- Most natural way for humans of describing colors
- Described by 3 attributes
  - Hue: the colour is red, green, yellow, blue, purple ...
  - Saturation: the level of non-whiteness
  - Brightness is a measure of the intensity of light.
- HSL color space. Hue, Saturation, Luminance
  - transformations from the RGB space.
  - inherit all the short-comings of RGB space.
  - There is usually a hue discontinuity around 360 degrees.
  - This makes difficult to make arithmetic operations in such a color space.



$$I = \frac{R + G + B}{3}$$

$$S = 1 - \left( \frac{3}{R + G + B} \right) \min(R, G, B)$$

$$H = \cos^{-1} \left( \frac{0.5(R - G) + (R - B)}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right)$$

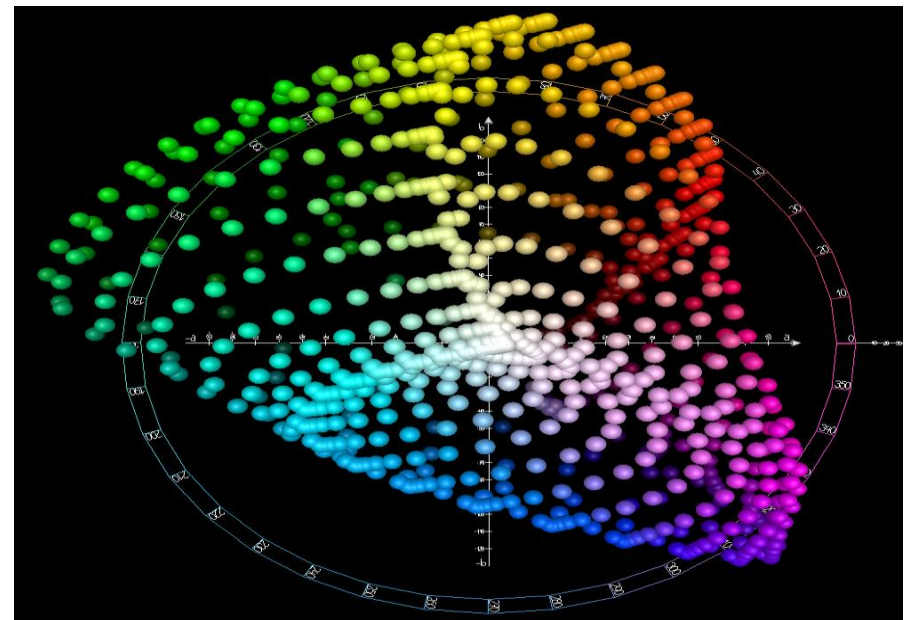
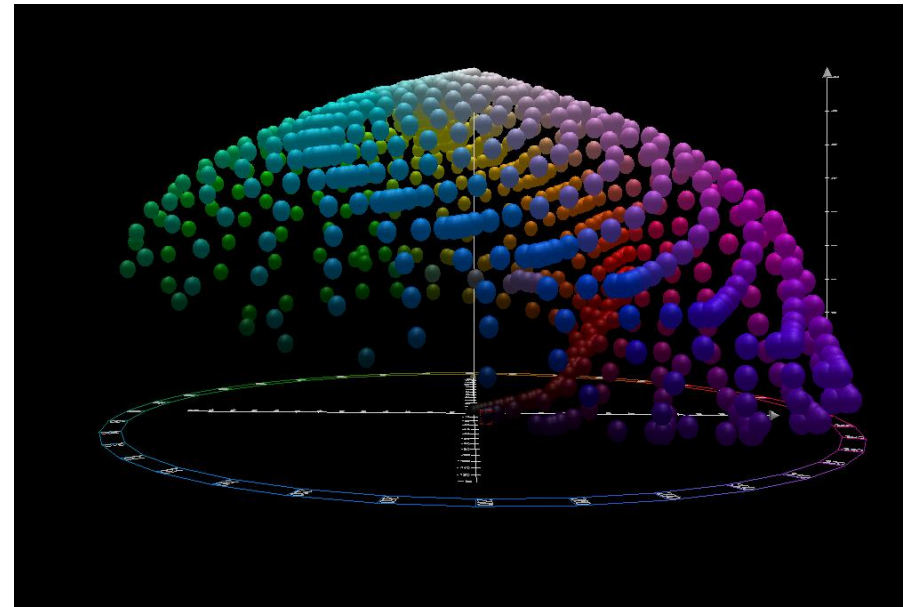
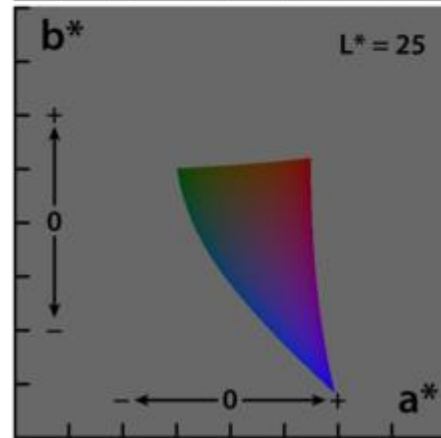
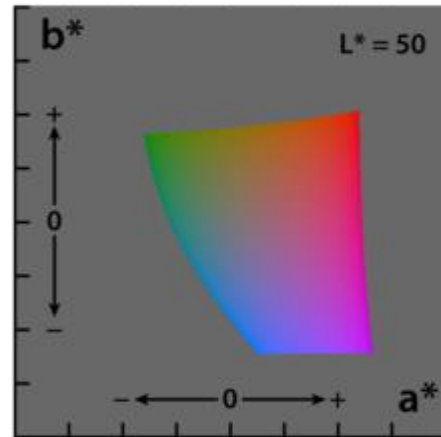
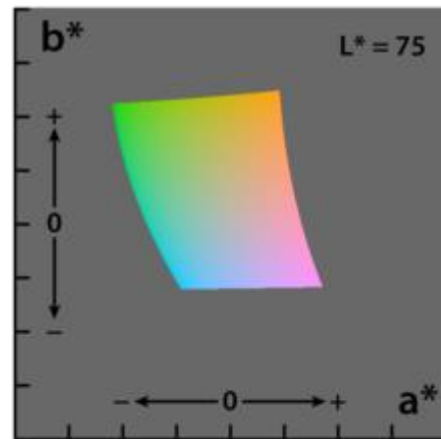
where  $I$  (intensity) is used instead of  $V$  (value).



# CIE Colour Spaces

- CIE (Commission Internationale de l'Éclairage (illumination))
  - In 1931 laid down the CIE 1931 standard colorimetric observer.
  - CIE XYZ: CIE standardized the XYZ values as tristimulus values that can describe any color that can be perceived by an average human observer
    - XYZ are positives
    - It is device dependent
  - CIELuv and CIELab: proposed in 1976
  - Goal: provide a perceptually equal space
  - CIElab
    - $L^*$  closely matches human perception of lightness ( black at 0 and white at 100)
    - $a^*$  : green–red opponent colors, with negative values toward green and positive values toward red. Unbounded, usually bounded from -128 to 127
    - $b^*$  : blue–yellow opponents, with negative numbers toward blue and positive toward yellow. Unbounded

# CIE Lab





# Cielab

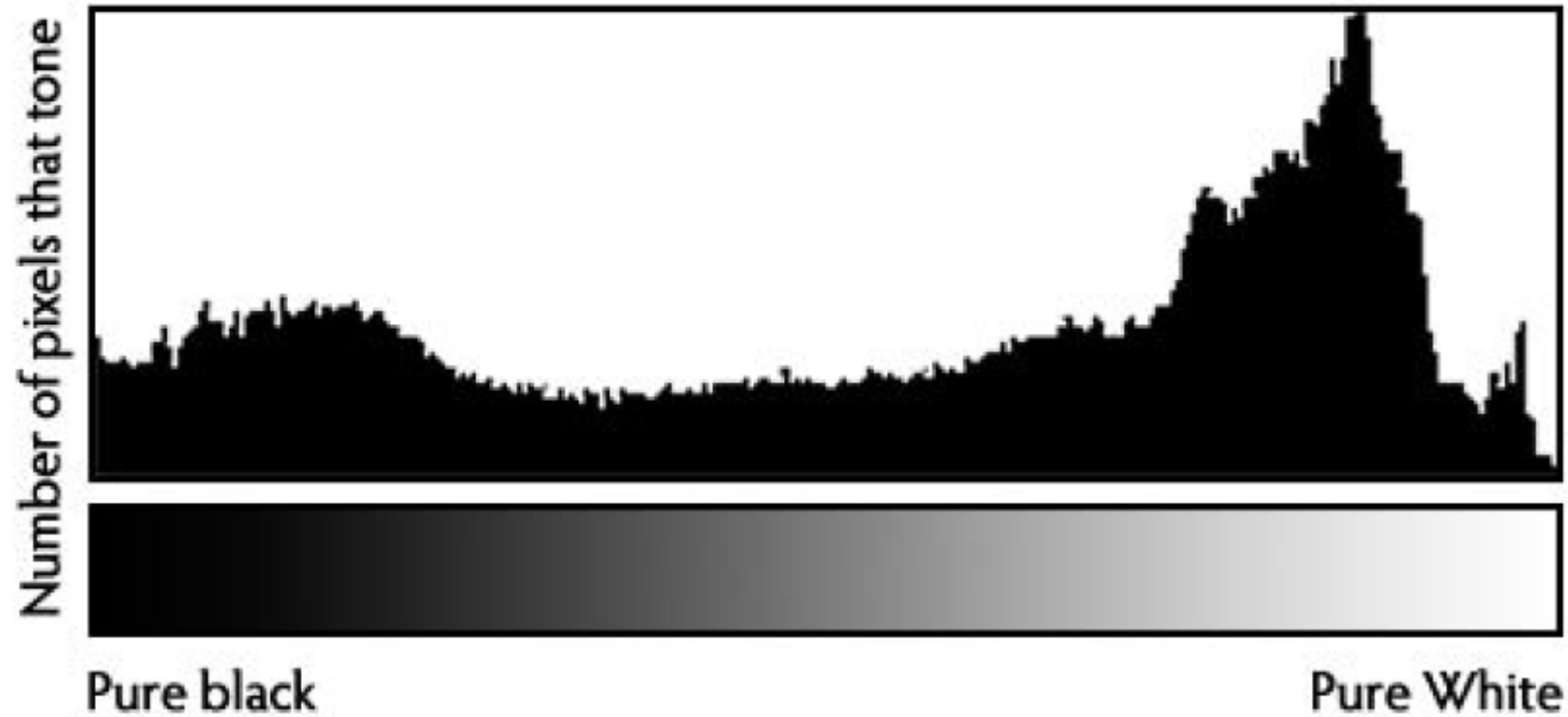


LAB color enhancement in Photoshop.

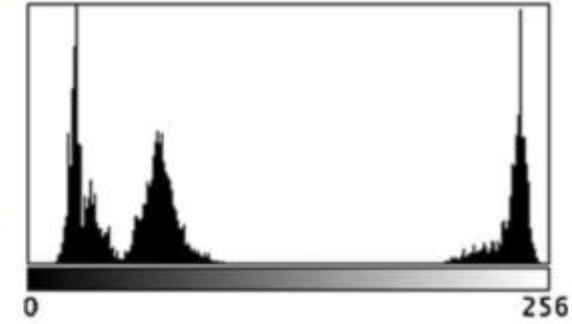
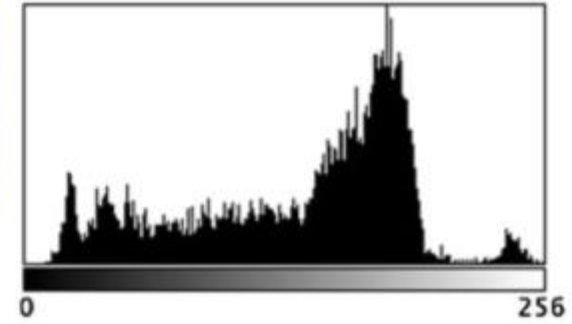
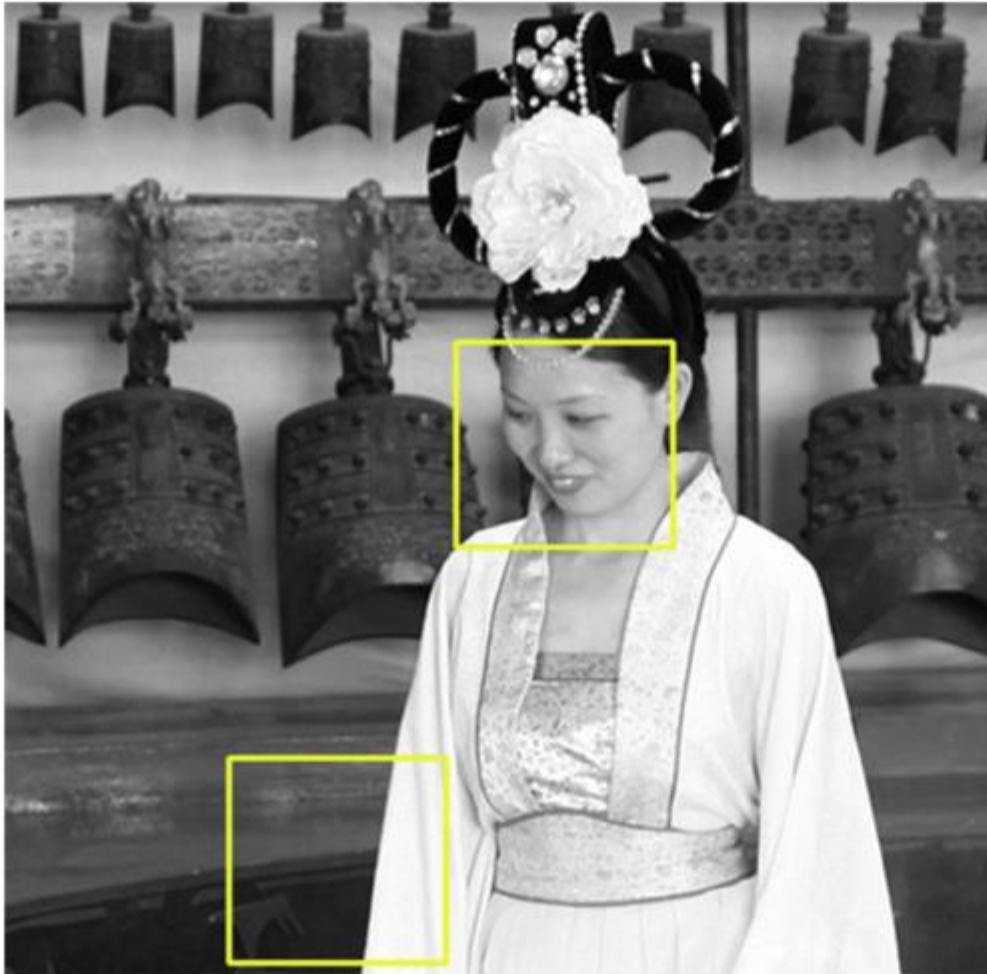
Normal Image



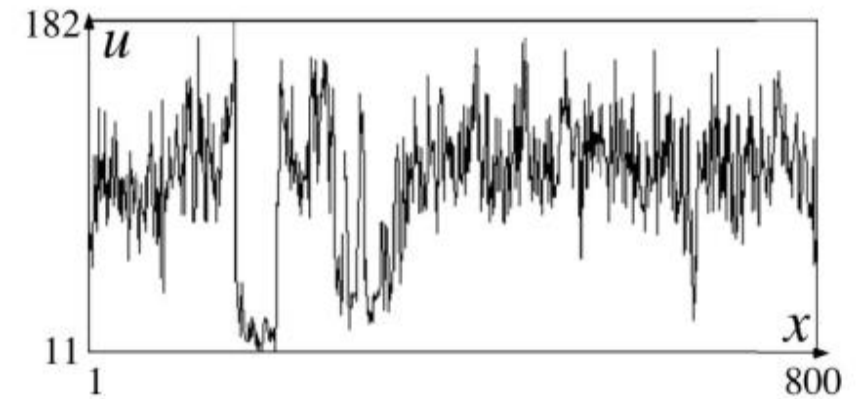
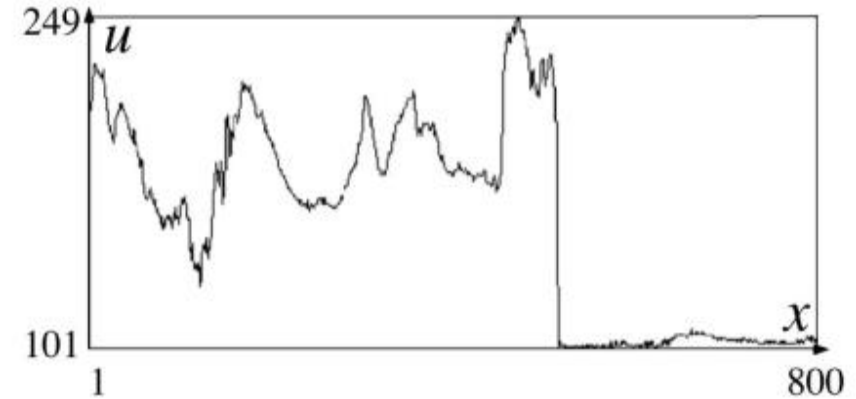
# Image Histogram



# Histogram Example



# Intensity profiles for selected (two) rows





# Questions?



# Coding homeworks

- Presented as a notebook using colab
  - <https://colab.research.google.com/>
- Homeworks are posted at webcouses as a link to:
  - [gonzo1978/CAP4453: Colab notes for CAP 4453 \(github.com\)](https://github.com/gonzo1978/CAP4453)