Computer Vision

Introduction

Dr. Pratik Mazumder

Syllabus

Introduction: The Three R's - Recognition, Reconstruction, Reorganization

Fundamentals: Formation, Filtering, Transformation, Alignment, Color

Image Restoration: Spatial Processing and Wavelet-based Processing

Geometry: Homography, Warping, Epipolar Geometry, Stereo, Structure from Motion, Optical flow

Segmentation: Key point Extraction, Region Segmentation (e.g., boosting, graph-cut and level-set), RANSAC

Feature Description and Matching: Key-point Description, handcrafted feature extraction (SIFT, LBP)

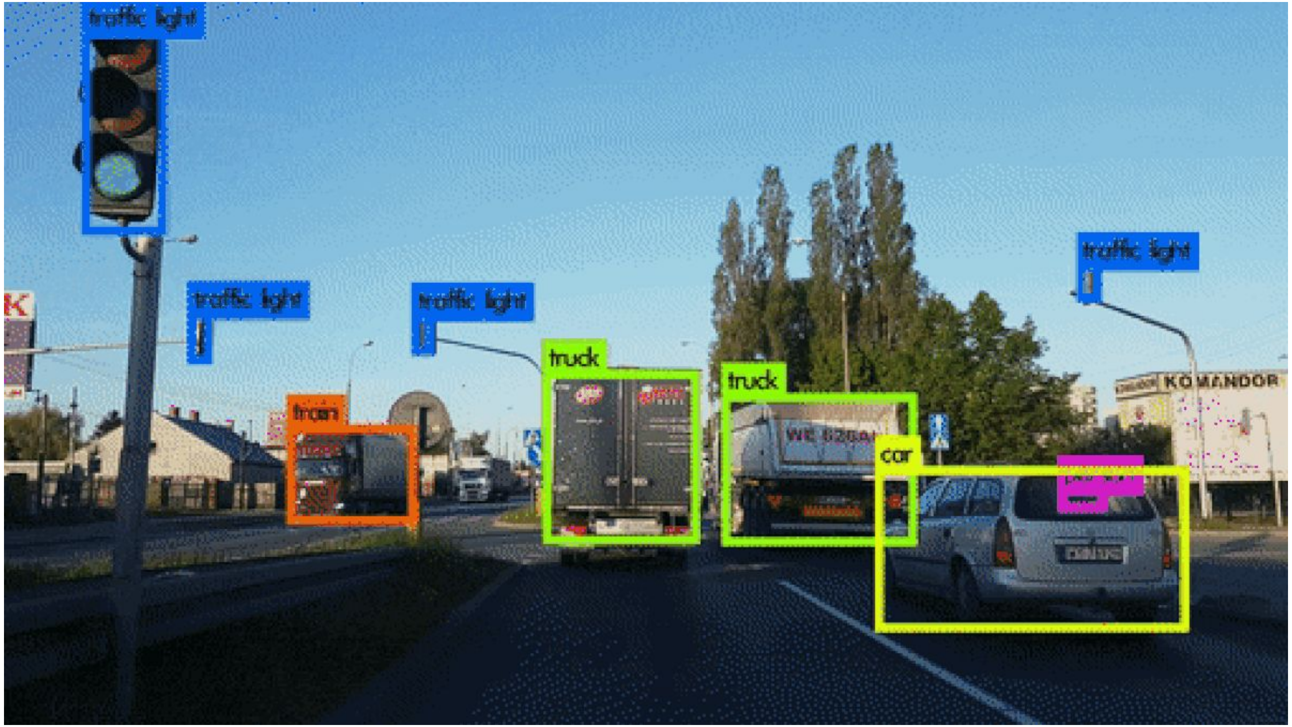
Deep Learning based Segmentation and Recognition: DL-based Object detection (e.g. Mask-RCNN, YOLO), Semantic Segmentation, Convolutional Neural Network (CNN) based approaches to visual recognition

Applications: Multimodal and Multitask Applications

Textbooks

- R. SZELISKI, (2010), Computer Vision: Algorithms and Applications, Springer-Verlag London.
- R. HARTLEY, A. ZISSERMAN (2004), Multiple View Geometry in Computer Vision, Cambridge University Press, 2nd Edition.

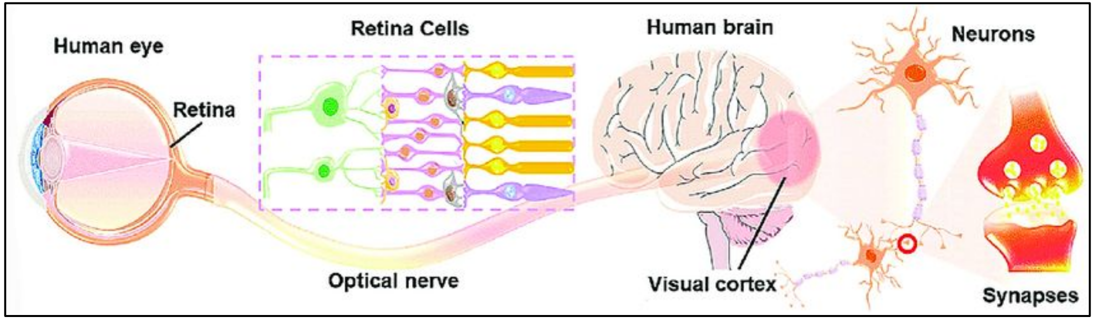
Let’s Begin



traffic light
traffic light
traffic light
traffic light
truck
truck
train
car
KOMANDOR

Biological Vision

- Ability to perceive the surrounding environment through the eyes.
- This involves
  - the reception of visual information by the eyes,
  - the transmission of signals to the brain, and
  - the subsequent interpretation of these signals to form a visual perception.



Computer Vision

- The image-capturing process of eyes is replicated through photography.
- Photography is not computer vision.
- Computer vision deals with how computers can be made to gain high-level understanding from digital images or videos.
- Field of artificial intelligence (AI) that
  - enables computers and systems to derive meaningful information from digital images, videos and other visual inputs
  - and take actions or make recommendations based on that information.
- Making Computers SEE

Why?

- For the same reason, we build machines
    - To make our lives easier
- Intellectual Challenge
- Give ~~Life~~ Vision


Figure credit: Hakan Bilen

Why?

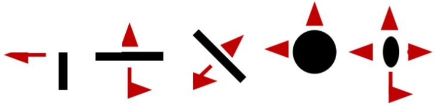
If we can give a computer the power of vision, can we not give us a better vision?

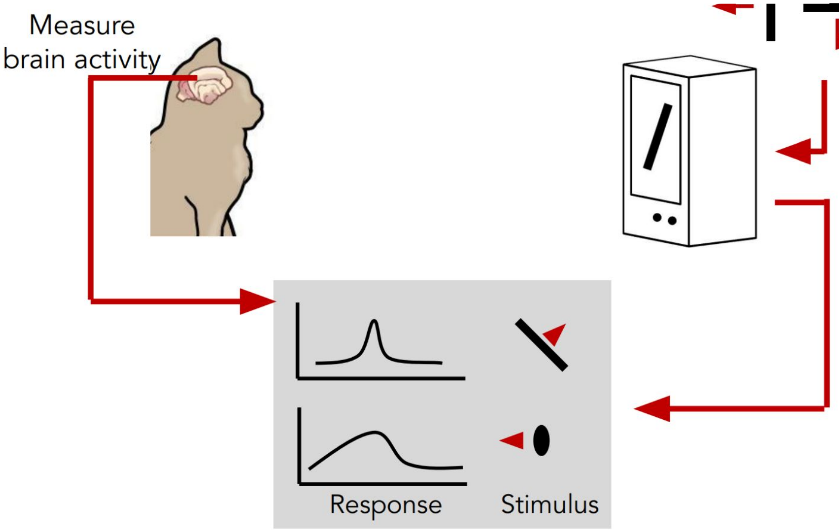
Credit: Marvel Studios

Understanding Biological Vision

- Hubel and Wilson (1960s)
  - Identified two main types of cells in the primary visual cortex
    - Simple cells and Complex cells.
    - Simple cells were found to have receptive fields that respond to specific oriented edges or bars of light.
    - Complex cells, on the other hand, displayed similar orientation selectivity but were less sensitive to the exact position of the stimulus within the receptive field.

Understanding Biological Vision





Simple cells:
Response to specific
rotation and orientation

Complex cells:
Response to light
orientation and
movement, some
translation invariance

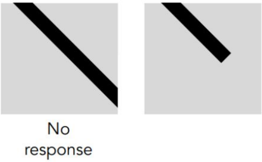


Figure credit: Fei Fei Li

Understanding Biological Vision

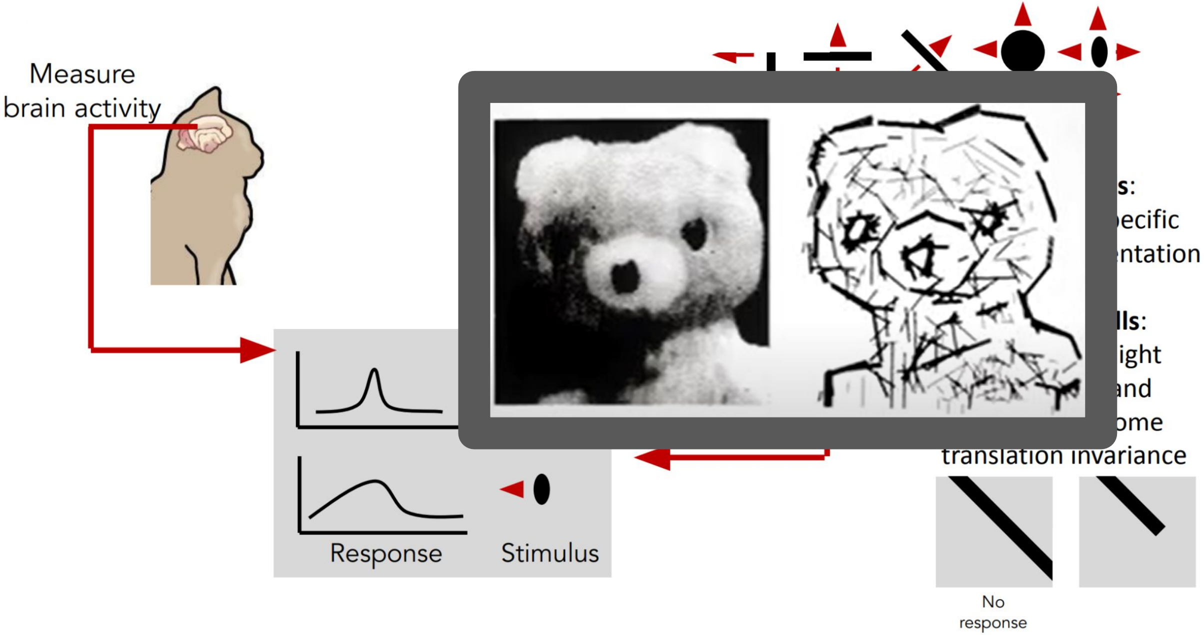


Figure credit: David Marr

Let’s do the same thing for a Computer

But is it that simple?

Issues with Computer Vision

- We can easily perceive the 3D structure of the world around us.
- Looking at flower,
  - you can tell the shape and translucency of each petal
    - through the subtle patterns of light and shading that play across its surface
    - and easily differentiate the flower from the background of the scene.

Issues with Computer Vision

- Looking at a framed family portrait, you can easily count and name all of the people in the picture and even guess at their emotions from their facial expressions.

- These abilities are difficult to replicate in computers.

- A major reason is the **inverse** problem.

- Before looking at that, what is the **forward** setup

    - Involves techniques like computer graphics which model how objects move and animate,
    - how light reflects off their surfaces,
    - is scattered by the atmosphere,
    - refracted through camera lenses (or human eyes), and
    - finally projected onto a flat (or curved) image plane.

Issues with Computer Vision

- The **inverse** problem
    - Trying to describe the world that we see in one or more images
        - and to reconstruct its properties, such as shape, illumination, and color distributions.
    - Trying to recover some unknowns given insufficient information to fully specify the solution, e.g.,
        - Given a 2D image trying to estimate the depth
        - Given a low resolution CCTV image of the crowd and trying to count the number of people in the crowd

Other factors that make the problem hard:

- Occlusion

Issues with Computer Vision

- low scene illumination, low resolution.

Issues with Computer Vision

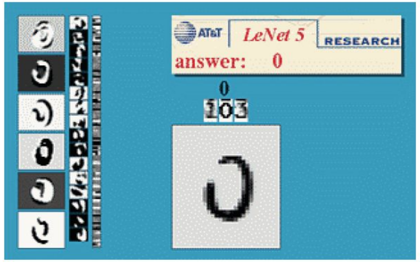
- Objects of the same category may look very different

Issues with Computer Vision

- Same object looks different from different viewpoints

Applications

- Optical character recognition (OCR)
- Machine inspection
- Retail
- Warehouse logistics
- Medical imaging
- Self-driving vehicles
- 3D model building (photogrammetry)
- Match move
- Motion capture (mocap)
- Surveillance
- Fingerprint recognition and biometrics



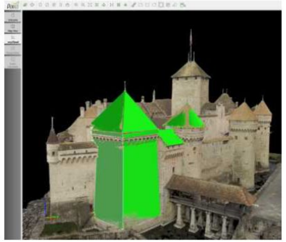


Figure credit: Szeliski

3R's of Vision

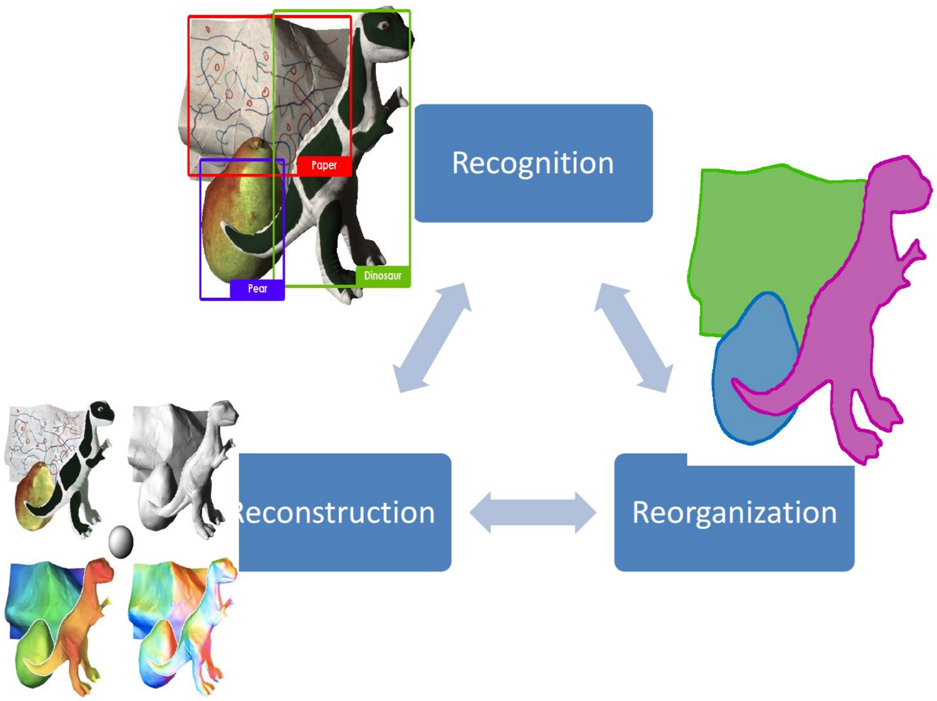


Figure credit: Jitendra Malik

3R’s of Vision

- Recognition
  - identifying and classifying objects or patterns within an image or a sequence of images.
  - assigning labels or categories to the visual content such as objects and scenes as well as to events and activities.
- Reconstruction
  - recovering three-dimensional geometry of the world from one or more of its images.
  - broadly “inverse graphics” – estimating shape, spatial layout, reflectance and illumination.
- Reorganization
  - refers to “perceptual organization”
  - organize/group/segment the different regions of the images.

3R's of Vision

Computer Vision can be considered as a result of the interaction of three processes:

- recognition, reconstruction and reorganization which operate in tandem,
- each provides input to the others and fruitfully exploits their output.

R’s Interactions

- **Recognition helps reconstruction**
  - Recognition can label objects and using the 3D model of the corresponding category, reconstruction can be improved.
- **Reconstruction helps reorganization**
  - Reconstruction can provide additional information like depth, contours, textures which can improve the reorganization.
- **Reconstruction helps recognition**
  - Additional information like depth can further improve recognition.
- **Reorganization helps reconstruction**
  - Segmented regions information can further improve reconstruction.

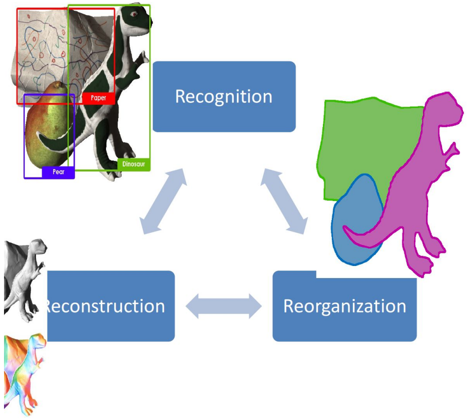


Image Formation and Filtering

Image Formation Process

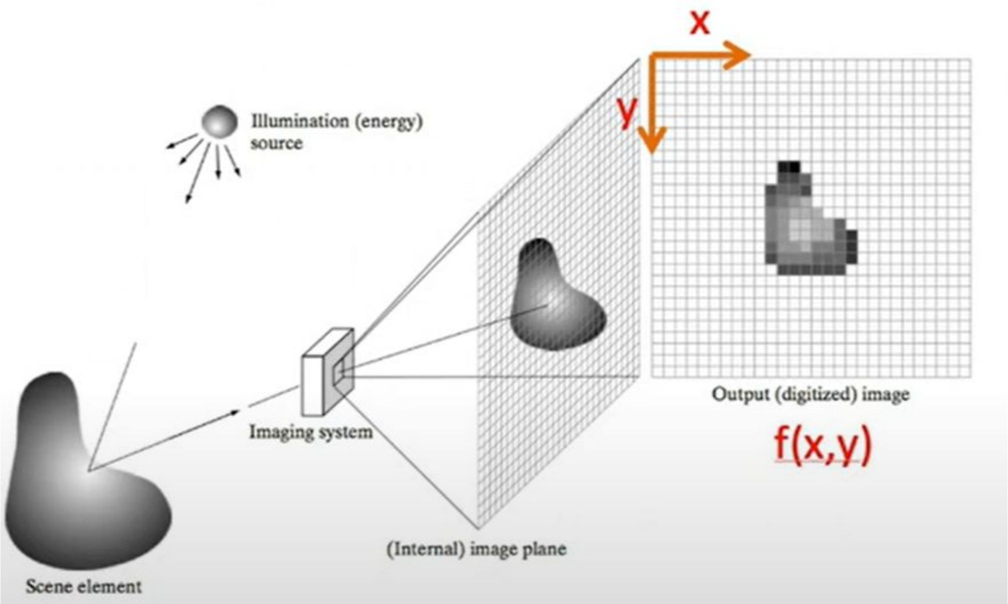


Image Credit Gonzalez and Woods

Image Formation Process

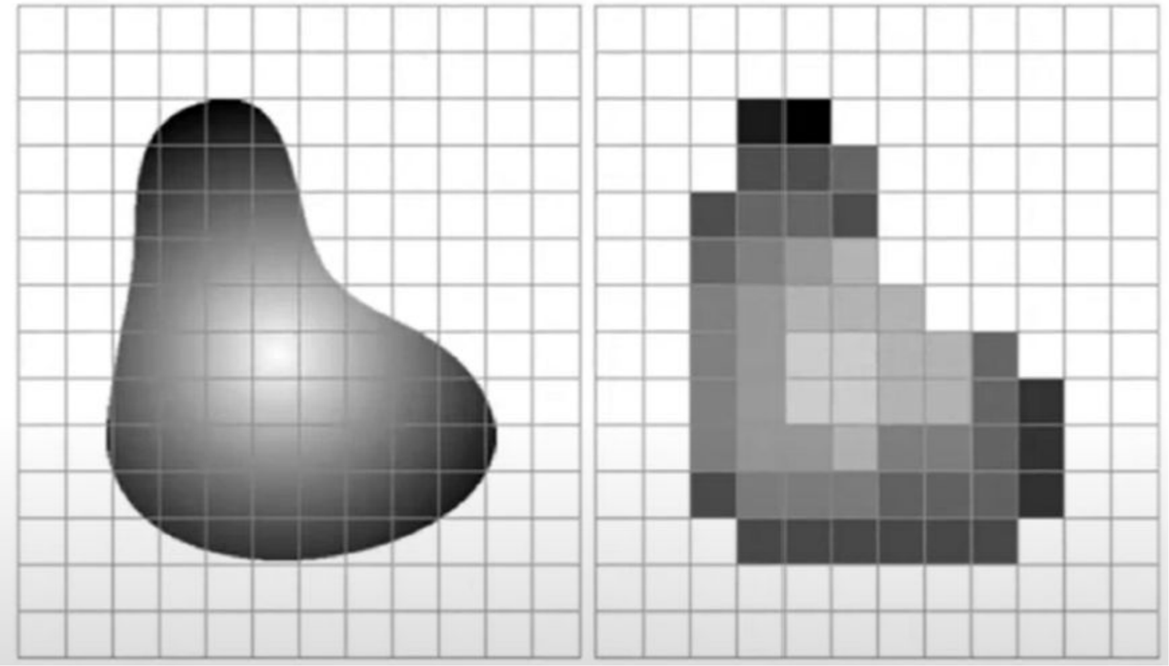


Image Credit Gonzalez and Woods

Digitization

Computers use discrete form of the images

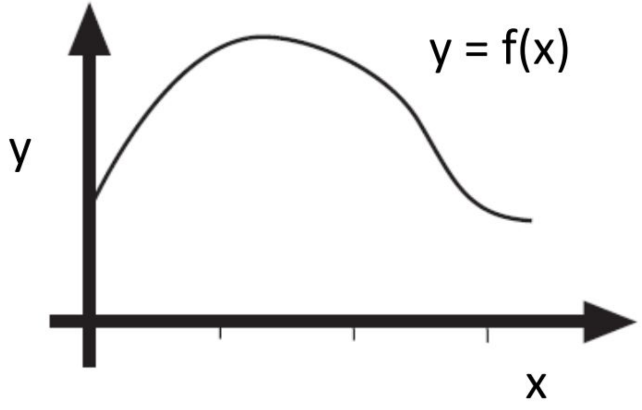
The process transforming continuous space into discrete space is called digitization.

Digitization

- Function

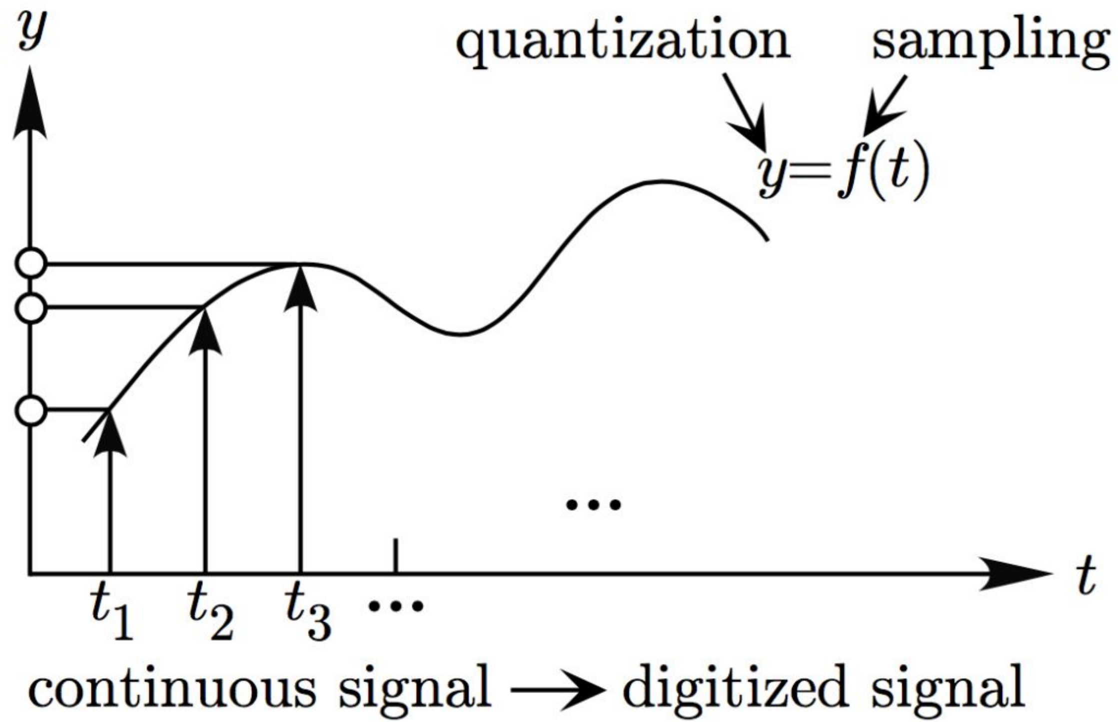
$$y = f(x)$$

- Domain of a function (x)
- Range of a function (y)
- Sampling
  - Discretization of domain
- Quantization
  - Discretization of range

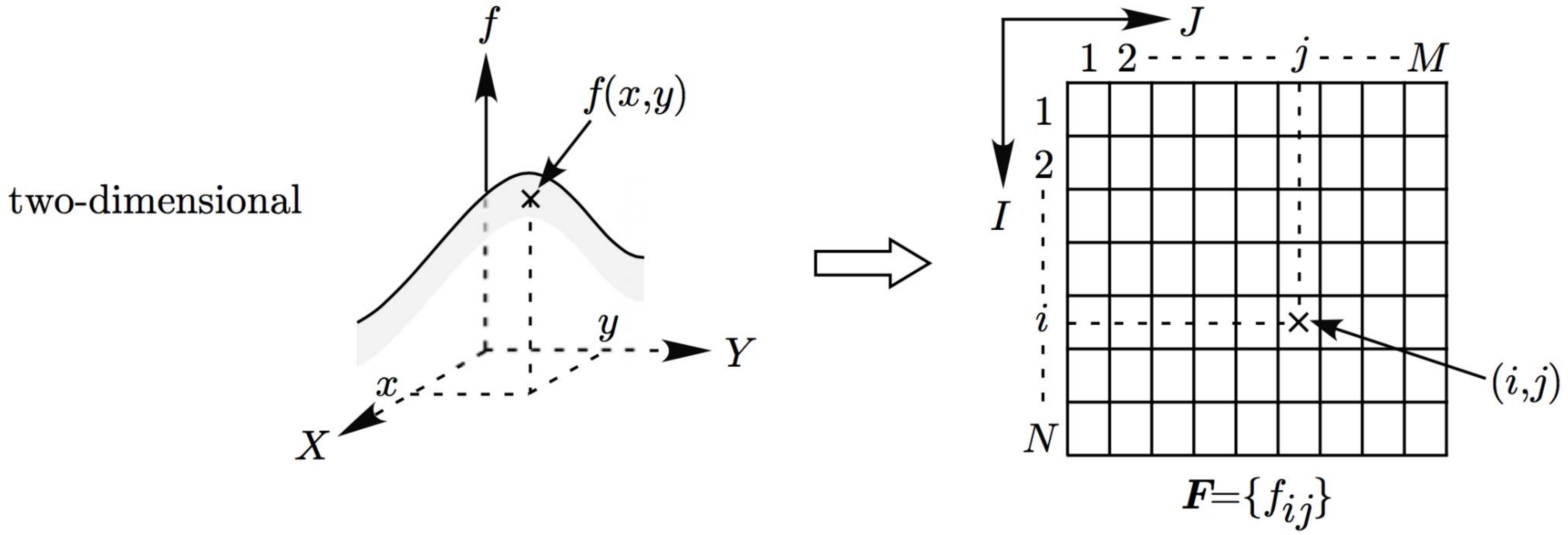


Digitization of a 1D function

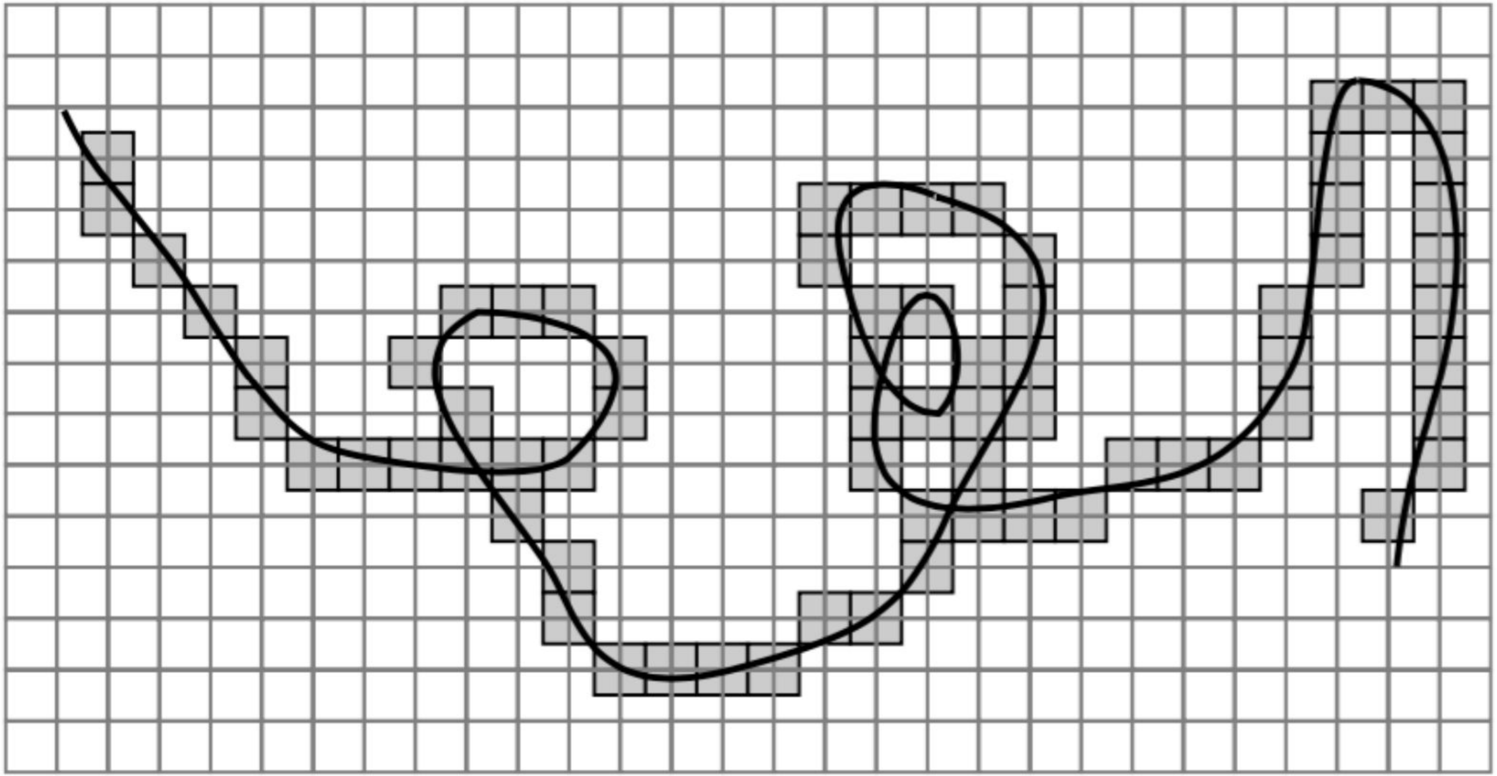
one-dimensional



Digitization of a 2D function



Digitization of an arc



Gray scale digital image

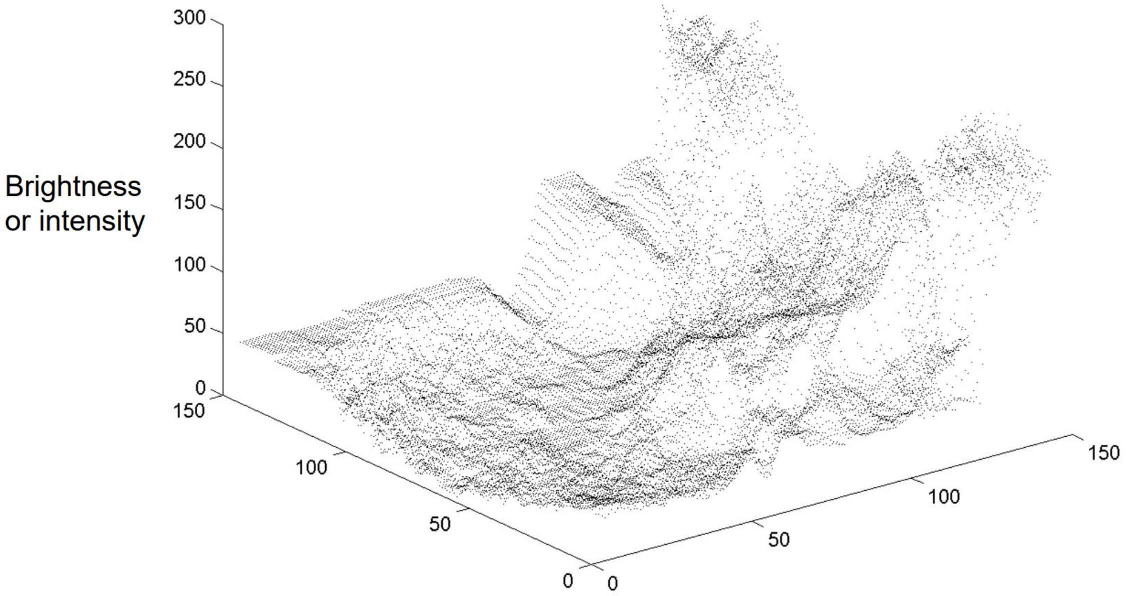
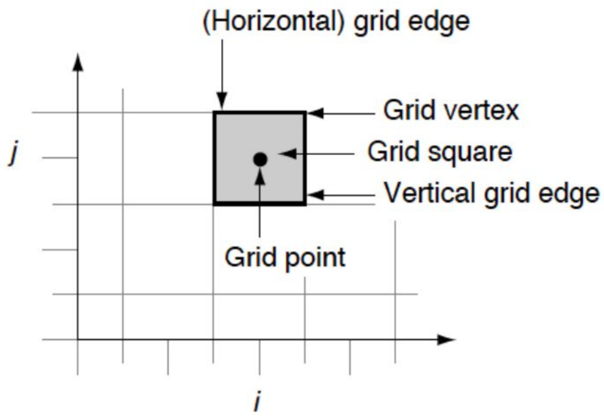


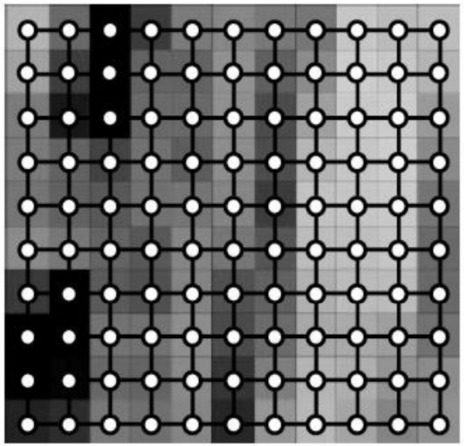
Image Credit Danny Alexander

Images

- An image P is a function defined on a (finite) rectangular subset G also called (2D) grid, and an element of G is called a pixel (picture element).

- P assigns a value of P(p) to each p∈G

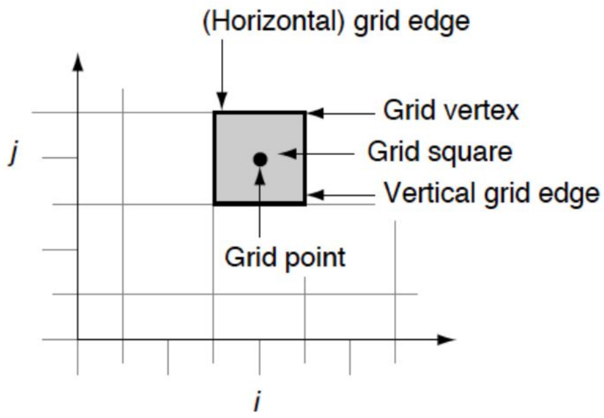


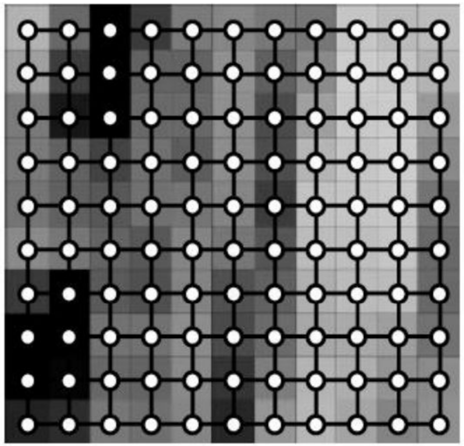


Images

- 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, …

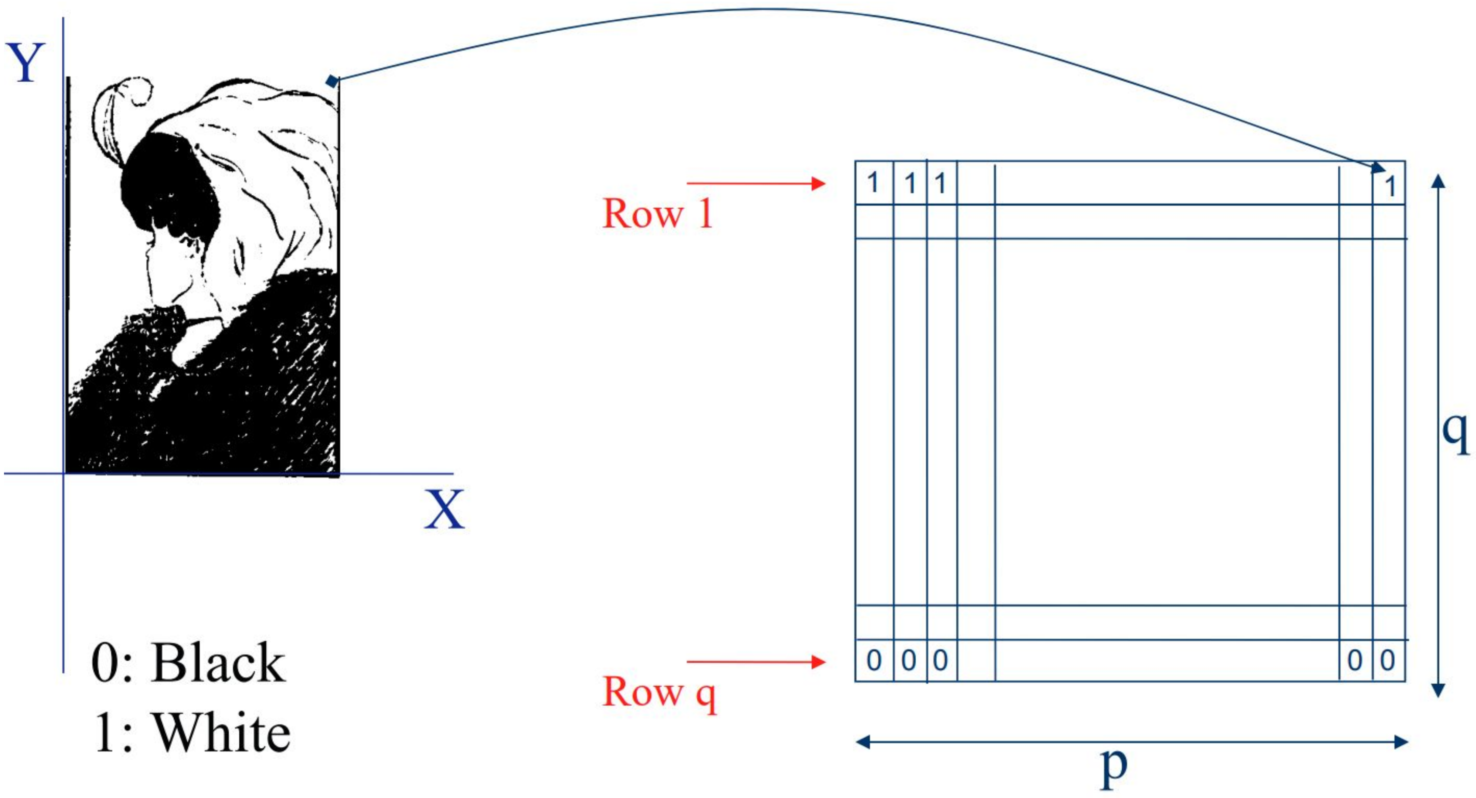




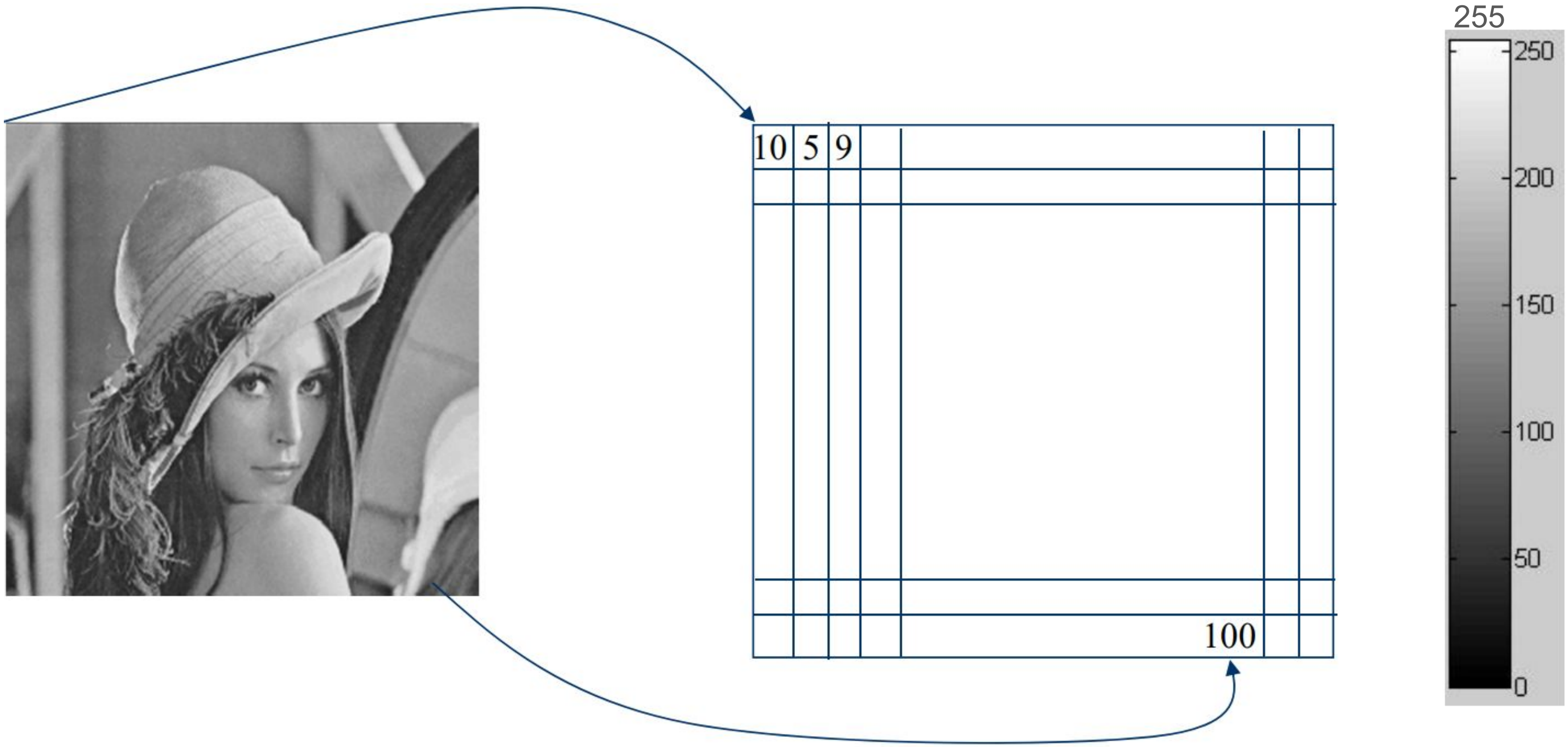
Images

- Binary
- Grayscale
- Color

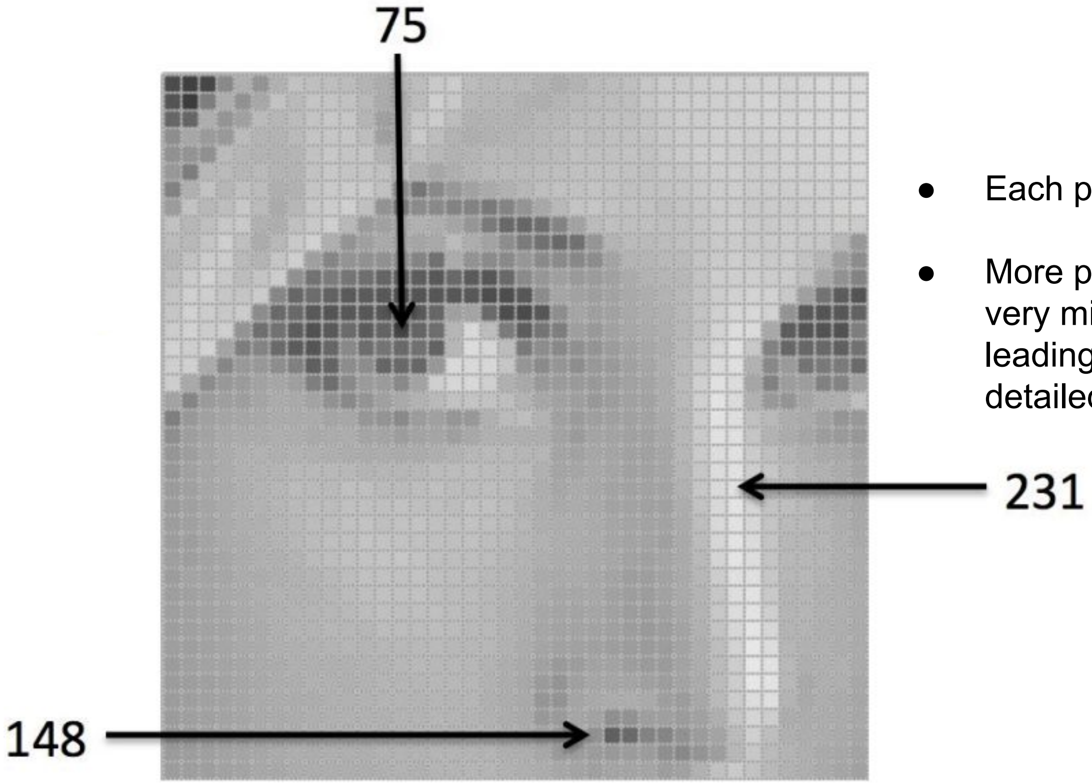
Binary Images



Grayscale Images

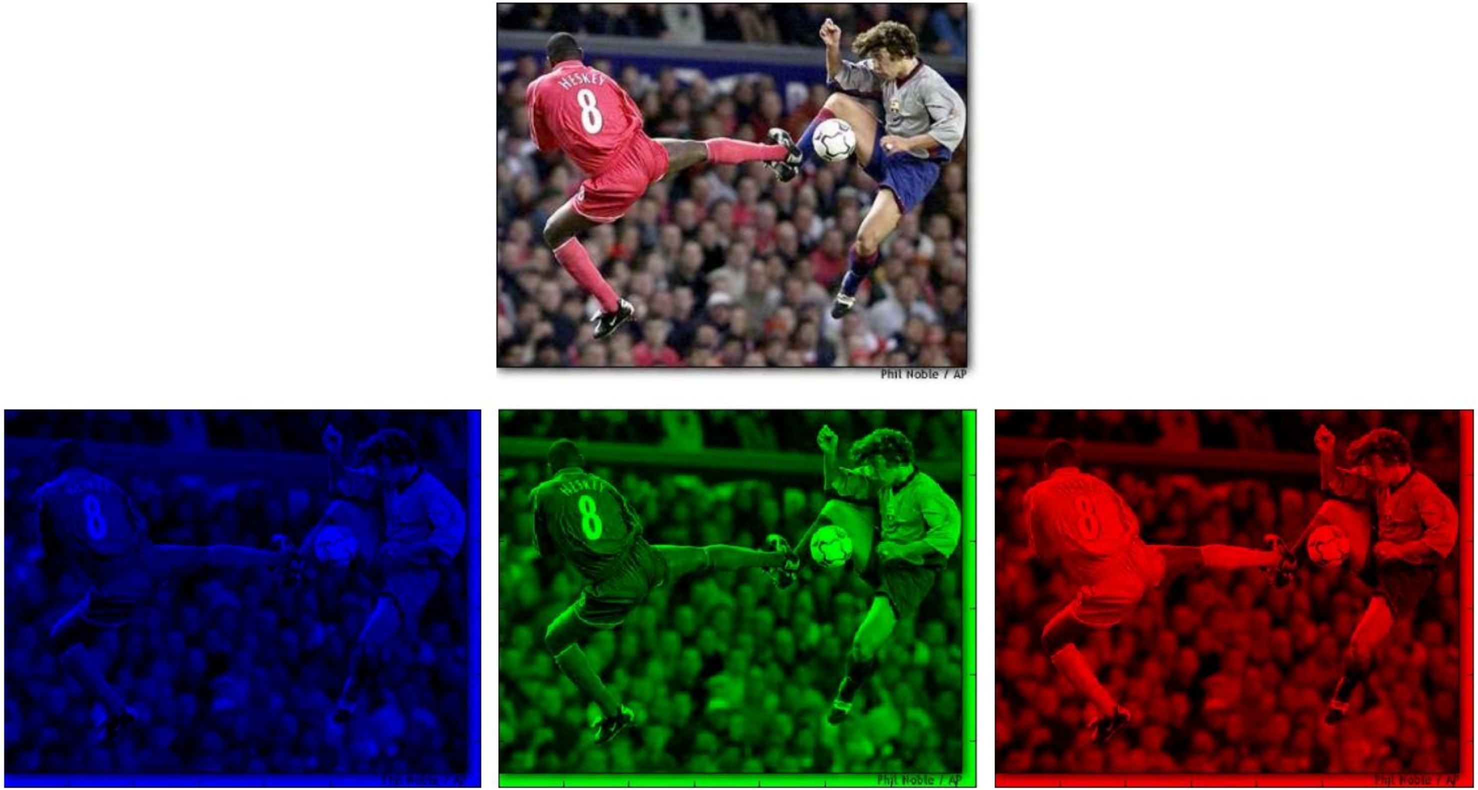


Grayscale Images

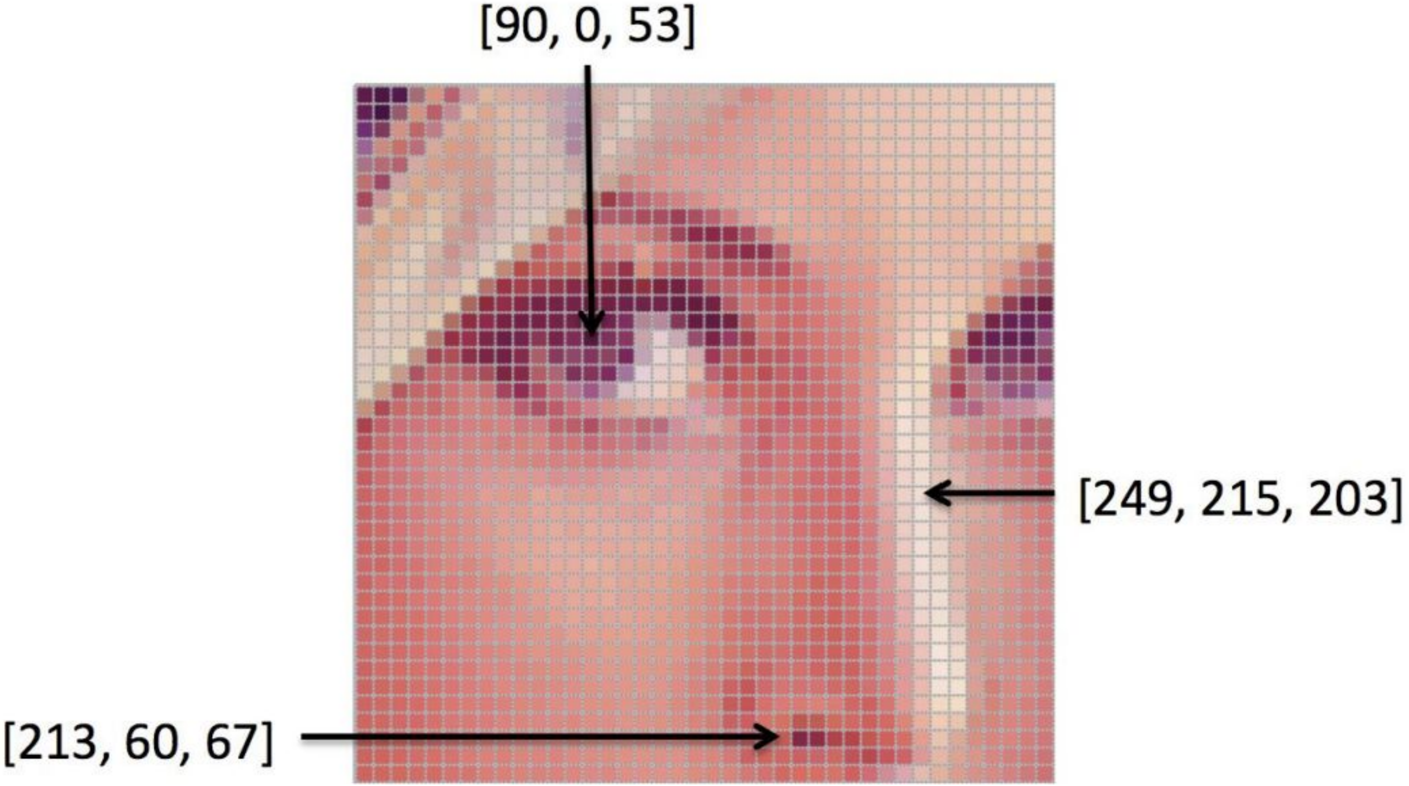


- Each pixel has a uniform color
- More pixels allows capturing very minute color variations leading to smoother and detailed image

Color Images



Color Images



Sampling

- Sampling along the x and y axes of the image decides the resolution of the image.
    - More the sampling levels - more the resolution, better image quality, more details captured

- Trade-off: details vs storage

Quantization

- Quantization of the intensity levels decides leads to fixing the levels of intensity used
    - More intensity levels - more detailed images

- Trade-off: details vs storage

Original (256 colors) | 8 colors | 4 colors

Quantization

- 8 bits to represent pixel intensity - 0 to 255 intensity levels

8 bits per pixel　　4 bits per pixel　　2 bits per pixel　　1 bit per pixel

Images

- Resolution becomes especially important in some applications where very minute details are important.

- Imagine using a pixelated low resolution image to check if you have a tumor in the brain!!