CSE6367 Computer Vision

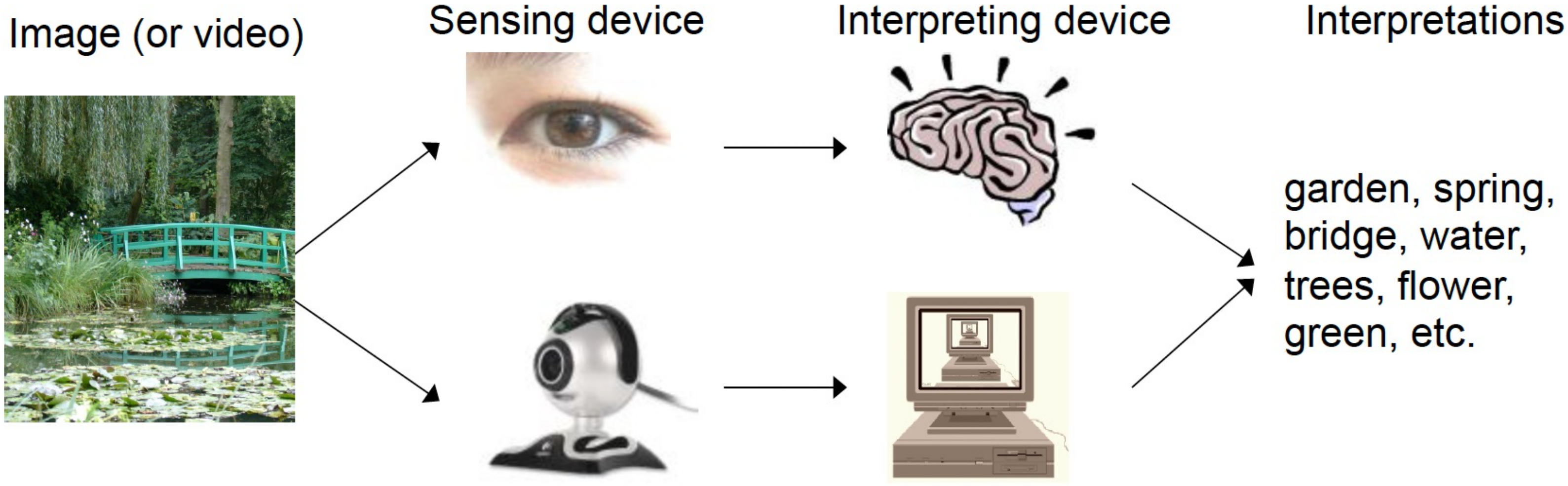
Instructor: Marnim Galib

Fall 2023

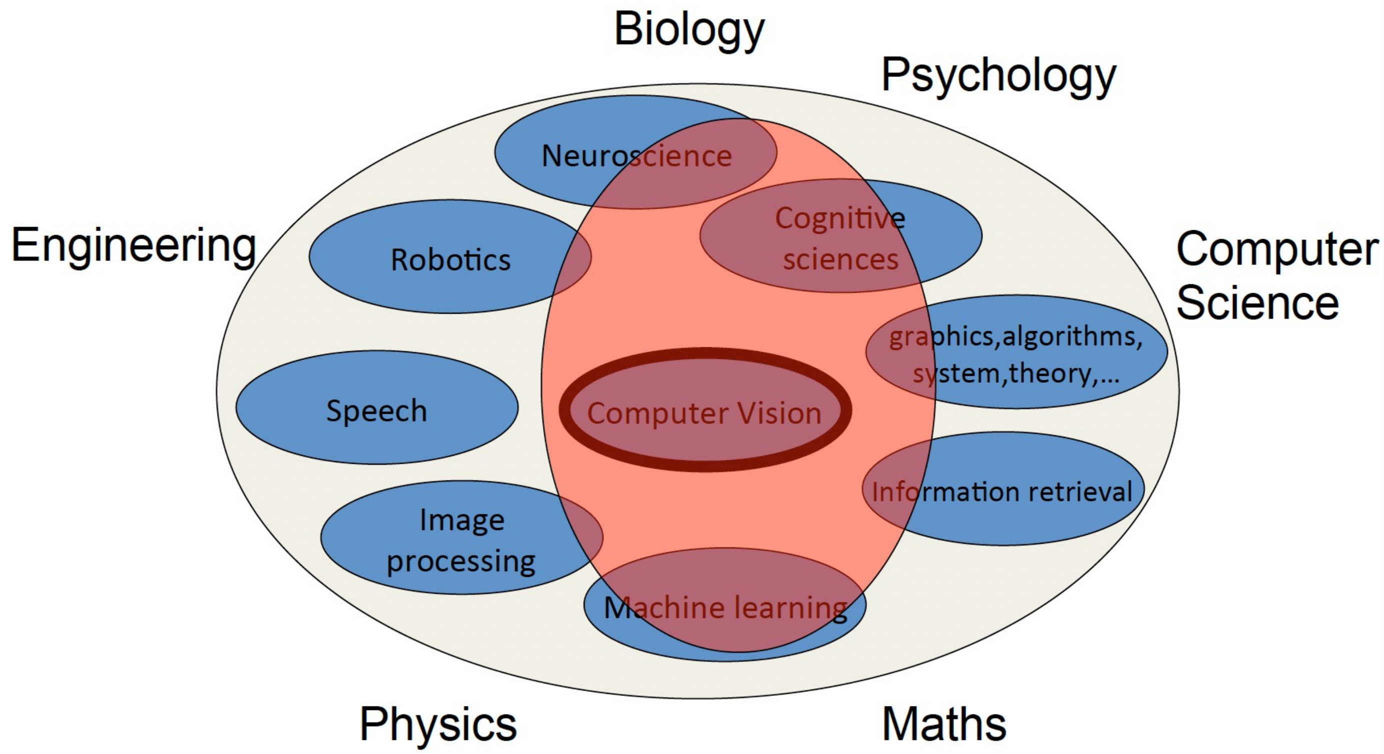
What is Computer Vision

- The goal of computer vision is to build machines that can see.
- While humans and animals naturally solve vision as a problem from a very young age, helping machines interpret and perceive their surroundings via vision remains a largely unsolved problem.

What is Computer Vision?



What is it related to?



The goal of Computer Vision

- To bridge the gap between what we see and what the computer sees.

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

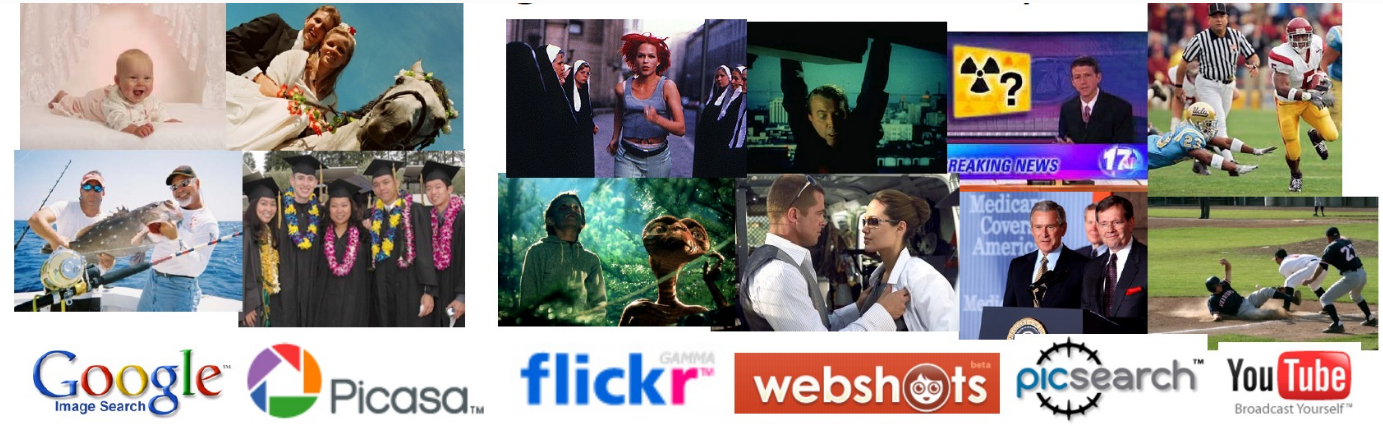
Human Vision

- Vision is our most powerful sense.
- It allows us to interact with the physical world without making any direct physical contact.
- It is believed that about 60% of the brain is, in one way or the other, involved in visual processing.

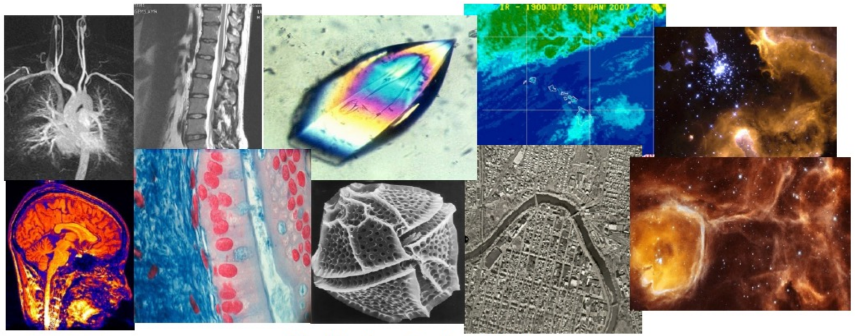
Why use machines to emulate vision

- To free up time to devote to more rewarding activities. Examples: tidying your home and driving to work.
- It is not particularly good at making precise measurements of things in the physical world.
- It can be designed to surpass the capability of human vision and extract information about the world that we simply cannot.

Applications of Computer Vision



Surveillance and security



Medical and scientific images

Image Processing

- Low-level vision operations are mainly concerned with processing the pixels of an image to extract basic features.
- These features include corners, blobs, edges, derivatives, etc.
- Filtering of the image is also performed at this stage, e.g. smoothing the image using a Gaussian kernel

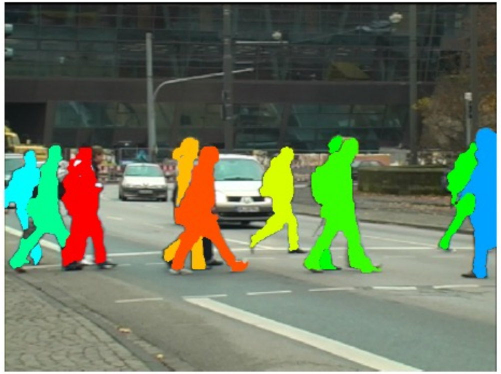
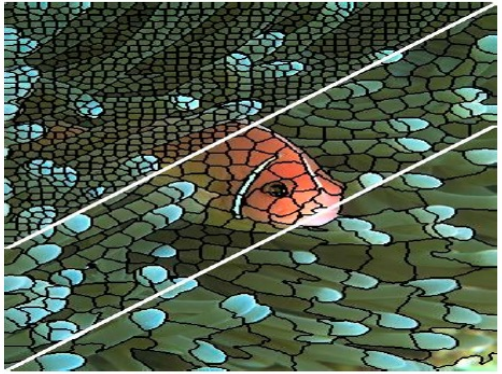
Edge Detection

Circle Detection

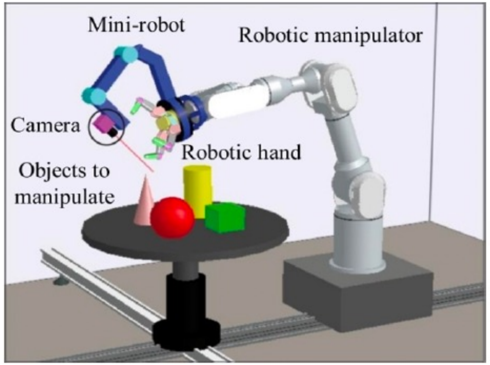
Finding Coherent Structure

- In mid-level vision operations, we want to find coherent structure in order to break the image into big units.
- These operations include segmentation (breaking images into useful pieces), tracking (keeping track of a moving object through a long sequence of views) etc.

Segmentation

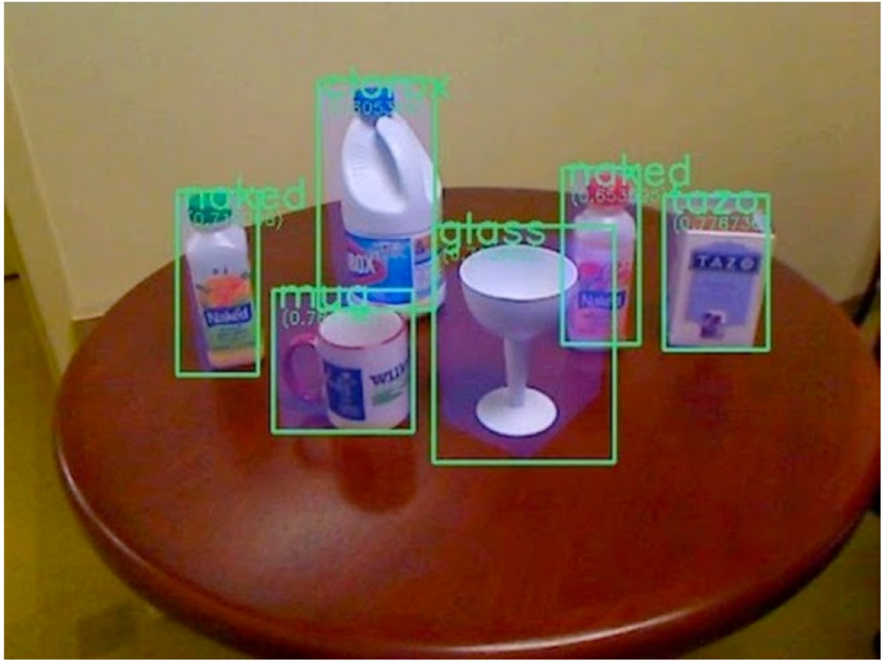
Tracking


Mini-robot
Robotic manipulator
Camera
Robotic hand
Objects to
manipulate

Object and Image Geometry

- In high-level vision operations, we want to determine the relations between object geometry and image geometry (e.g. find the position and orientation of known objects)
- We can use templates and classifiers to find objects that look the same from view to view (i.e. object recognition)
- Furthermore, we can create hierarchical frameworks to recognize the scene in which the objects reside (i.e. scene understanding)

Object Recognition



Scene Understanding

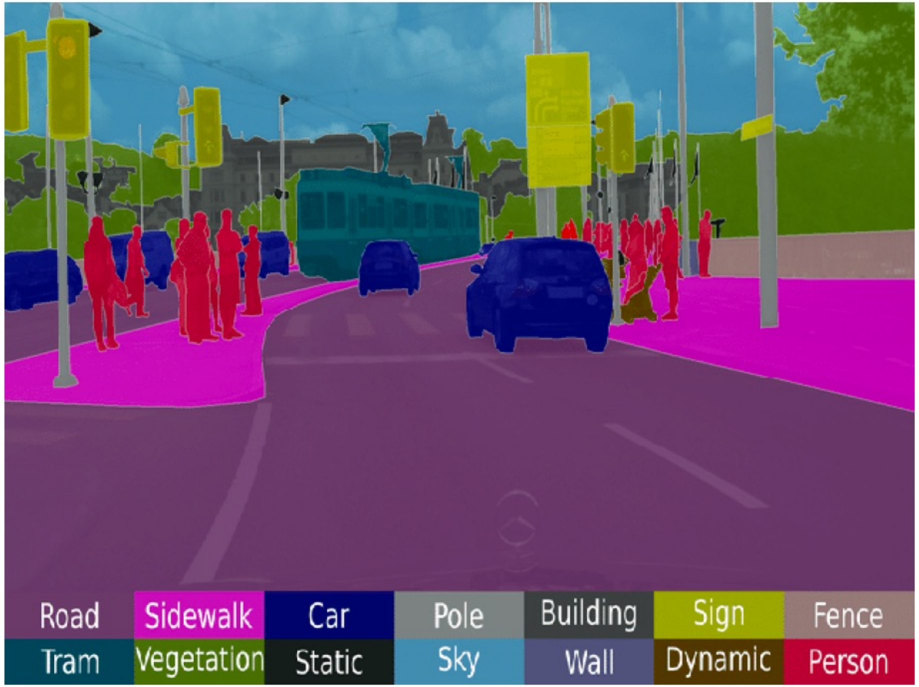


Image Formation

- Light is reflected off of objects and enters the camera through the lens.
- Each cell on the image sensor converts the light that hits it into an electrical signal.
- The electrical signals from all of the cells on the image sensor are then converted into digital data by the camera's processor.
- Finally, the digital data is displayed on a screen or printed out on a piece of paper.

Image Formation in Pinhole Camera

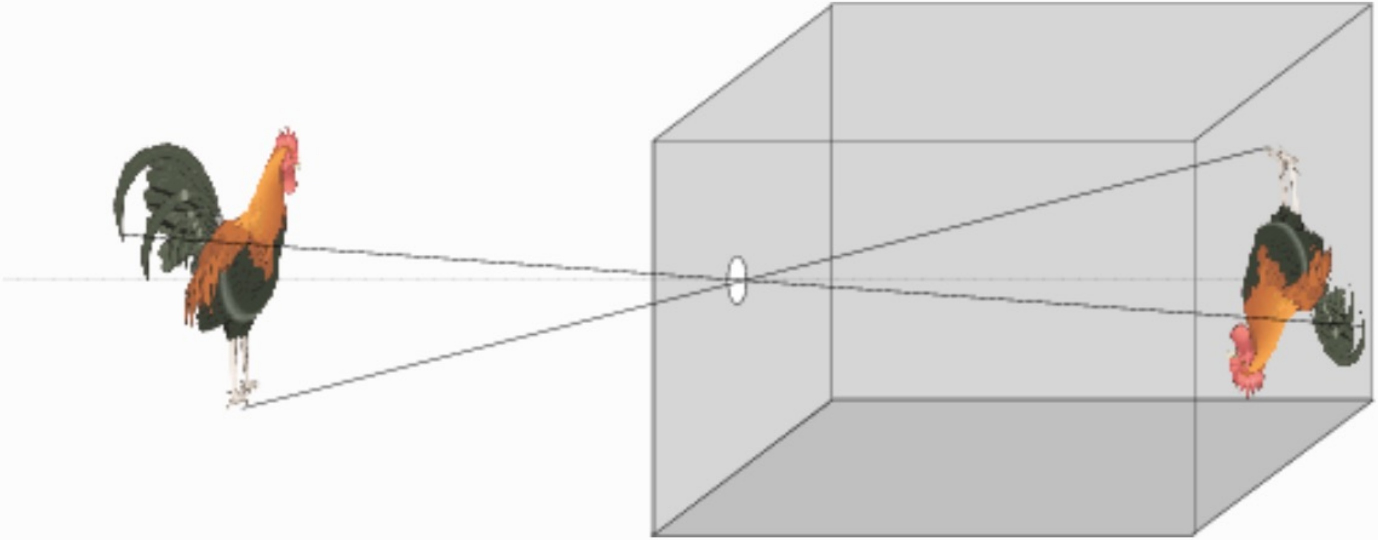


Image Formation

- In analog cameras, the image sensor is a strip of film that is coated with a light-sensitive emulsion. When light hits the film, it exposes the emulsion, which then undergoes a chemical reaction that creates a negative image.
- In digital cameras, the image sensor is a chip that is made up of millions of tiny light-sensitive cells.

Image Formation

Grayscale image:

- A 2D array of intensity values.
  - rows x columns.
- Usually, values range from 0 to 255, stored as 8-bit unsigned integers.
- Other ranges are possible.
  - E.g., 32-bit or 64-bit floats between 0.0 and 1.0.

Image Formation

Color image:

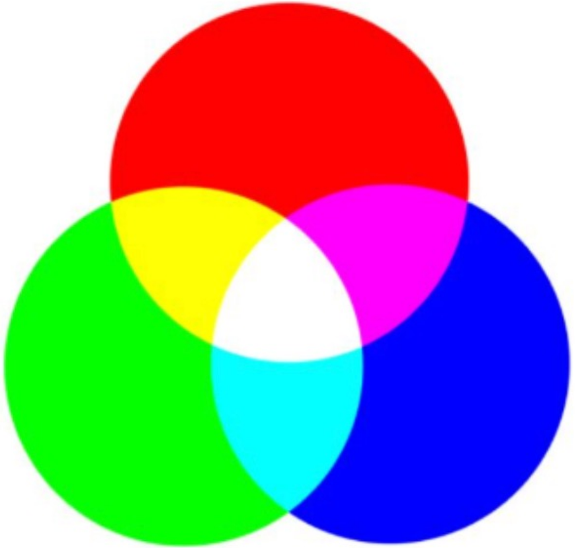
- Three 2D arrays of intensity values.
  - Red, green, blue.
- Can also be seen as a 3D array.

Color Models

- RGB
- CYMK
- HSV

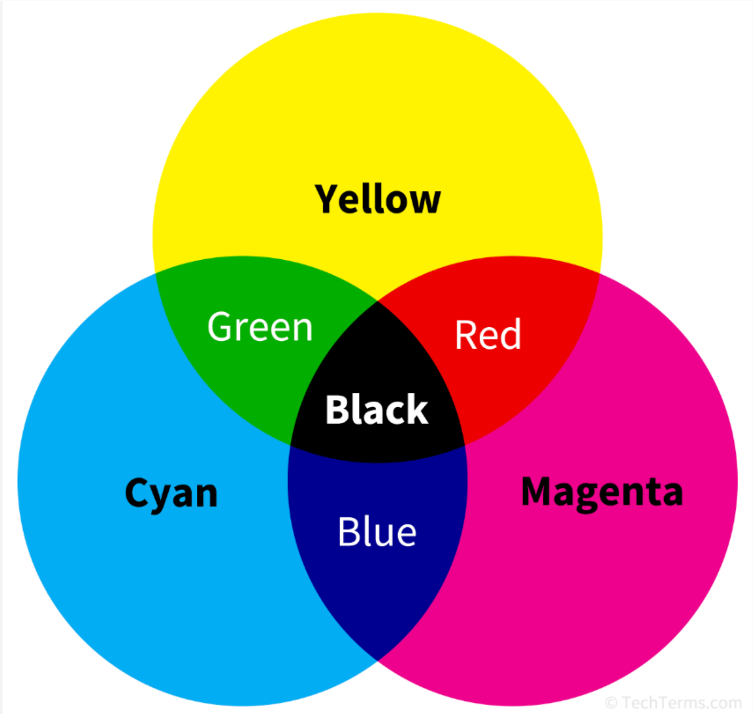
RGB

- This is an additive color model. 3 main channels: Red, Green and Blue.
- Color combination:
  - Green(255) + Red(255) = Yellow
  - Green(255) + Blue(255) = Cyan
  - Red(255) + Blue(255) = Magenta
  - Red(255) + Green(255) + Blue(255) = White



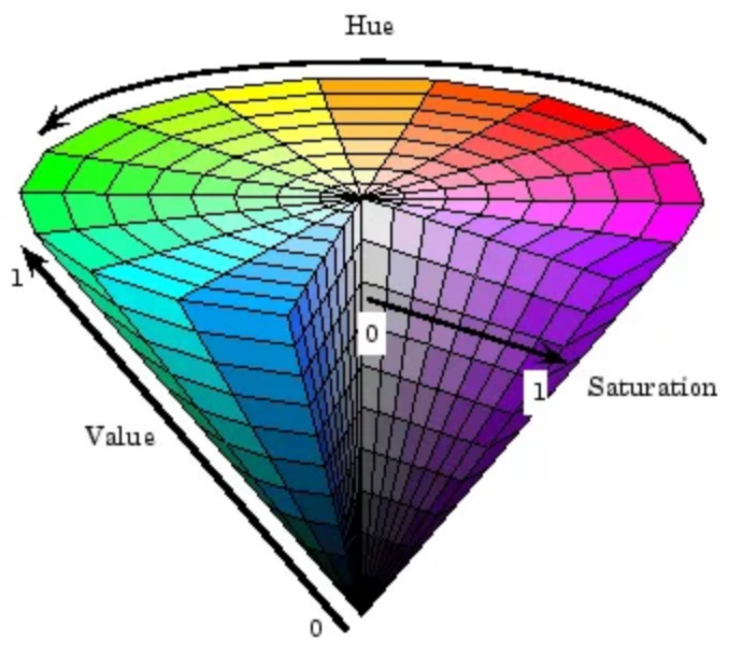
CMYK

- Subtractive color model. It stands for Cyan, Magenta, Yellow and Black (key).
- Widely used in printers.
- Cyan is negative of Red. Magenta is negative of Green. Yellow is negative of Blue.



HSV

- The image consists of three channels: Hue, Saturation and Value.
- Hue represents different colors in different angle ranges.
- Saturation describes the percentage of the color.
- The value represents the intensity of the color.



Raster vs Vector Graphics

Raster Formats:

- Raster formats store images as a grid of pixels.
- This means if we enlarge a raster image, we see jagged edges of the pixels.
- Best suited to store photographs.

Raster vs Vector Graphics

Vector Graphics:

- Vector formats store part of an images as a geometric formula.
- This means if we scale up a vector image it's always sharp.
- Vector formats are better suited for logos, texts which require sharpness even when you zoom in.

Reducing File Size

- Most raster formats apply compression.
- Lossy compression discards data, so reduces image quality.
- Non-lossy compression reduces file size whilst retaining all information.

Alpha Channels

- Some formats store pixel transparency in an ”alpha channel” so RGB becomes RGBA.
- When an image is overlaid onto another image, the alpha value of the source color is used to determine the resulting color.
- If the alpha value is opaque, the source color overwrites the destination color; if transparent, the source color is invisible, allowing the background color to show through.

Image File Formats

- Image Format describes how data related to the image will be stored.
- Data can be stored in compressed, Uncompressed, or vector format.
- Examples: JPEG, PNG, GIF, TIFF, RAW

JPEG

- Created by the Joint Photographic Experts Group in 1992. Extension is .jpg or .jpeg
- JPEGs store raster images, either in RGB or CMYK, with lossy compression.
- Due to compression, some data is lost but that loss is very less. No alpha channel and maximum 24 bit color.
- It is a very common format and is good for digital cameras, nonprofessional prints, E-Mail, Powerpoint, etc., making it ideal for web use.

GIF

- Graphics Interchange Format created in 1987.
- Raster based non-lossy compression.
- Small file size as miximum 8 bit color.
- Useful for web graphics, including animations.

PNG

- Portable Network Graphic - Created in 1994
- Stores raster images using non-lossy compression.
- Can include an alpha channel
- It was designed to replace gif format as gif supported 256 colors unlike PNG which support 16 million colors.
- Either 8 bit or 24 bit color. (PNG8, PNG24)

TIFF

- The Tagged Image File Format was released in 1986, and last updated in 2002.
- It does not perform any compression on images, and a high-quality image is obtained but the size of the image is also large, which is good for printing, and professional printing.
- Supports RGB, CMYK, B&W, with 8, 16 or 32 bpc.
- Includes an alpha channel.

RAW

- Native raster digital camera file format.
- RGB with 12, 14 or 16 bpc (camera dependent).
- Direct binary data dump from camera's sensors.
- These images are the equivalent of a digital negative, meaning that they hold a lot of image information.
- These images need to be processed in an editor such as Adobe Photoshop or Lightroom.

Slide Courtesy

- Dr. Fei Fei Li, Stanford Vision Lab
- Dr. Shree K. Nayar, Columbia University
- Dr. Vassilis Athitsos, CSE@UTA
- Dr. William Beksi, CSE@UTA
- GeeksforGeeks