Computer Vision - Lecture notes all

Computer Vision (Amrita Vishwa Vidyapeetham)

COMPUTER VISION

Computer Vision is a field of Artificial Intelligence(AI) that enables the computer and systems to derive meaningful information from digital images, videos and other visual inputs - and take actions or make recommendations based on that information. If AI enables computers to think, computer vision enables them to see, observe and understand.

Eg, Computer vision is necessary to enable self-driving cars. Manufacturers such as Tesla, BMW, Volvo and Audi use multiple cameras, lidar, radar and ultrasonic sensors to acquire images from the environment so that their self-driving cars can detect objects, lane markings, signs and traffic signals to safely drive.

APPLICATIONS OF COMPUTER VISION

1. Optical Character Recognition
2. Machine Inspection
3. Retail
4. 3D modelling building(photogrammetry)
5. Automotive safety
6. Match move
7. Motion capture
8. Surveillance
9. Fingerprint recognition and biometrics

HISTORY OF COMPUTER VISION

- Larry Roberts is commonly accepted as the father of computer vision.
- Computer Vision came into existence during the 1960's

LEVELS OF HUMAN AND COMPUTER VISION SYSTEM :

Low Level Vision : Edge , Corner, Stereo reconstruction
Mid Level Vision : Texture, Segmentation and Grouping , illumination
High Level Vision :Tracking, Specific Object recognition , Category level object recognition

CAMERA PROJECTION

Projection : Projection is a technique or process which is used to transform a 3D object into a 2D object in a projection plane or view plane. (representing an n dimensional object into (n-1) dimension).

Projection is of two types:

1. Parallel projection

2. Perspective projection

Parallel projection are of two types orthographic and oblique. Orthographic projection can be multiview or axonometric view. Axonometric view can be isometric, dimetric or trimetric. Parallel projection another type is an oblique view, which is a general one. It is of two types: cavalier or cabinet.
Perspective projection can be represented in one point, two point or three point.

How to transform a 3D world object to 2D picture ? Try to form a ray between our eyes and a 3D object through a canvas. The ray hit the 3D object and bounced back to the canvas and painted colours of the 3D object to 2D surface.
Parts help in transforming picture from 3D to 2D world :
Extrinsic Parameters - It includes the camera's orientation i.e., Rotation(R) , Translation(T). The extrinsic parameters are the camera body configuration.
Intrinsic Parameters : Spatial relation between sensor and pinhole (K) , focal length (L). It gives the transformation of the optical parameters.

$$\begin{bmatrix} x \\ 1 \end{bmatrix} = L \left(K \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix} \right)$$

Lens configuration (internal parameter)

Process of forming a 3D world object to a 2D picture : Imagine that we are an artist who is going to draw a picture of the world. We're standing in front of the canvas and looking

into the world. The thing we do here is we try to form a ray between the eye and the point in the world that we are seeing and now we pick any points that we want to draw in our canvas. So, this ray that is being formed keeps the 3D world intersecting the points there and bounds that object back into the canvas and then paints the color of the 3D object into the 2D picture. So, that is the process of forming a 3D world object to a 2D picture.

Projection Equation : 3D to 2D image - f=focal length , c = focal distance { distance between observer and the object}

3D to 2D image: $x' = f\frac{X}{Z}$ $y' = f\frac{Y}{Z}$

PERSPECTIVE PROJECTION

Once those closest to us appear larger than the ones that are farther.

{Eg : Railway track }

Components Of Perspective :

Key things needed for the setup of perspective projection:

- Spectator (observer)
- Picture plane(plane that we are going to draw perspective image of object)
- Horizontal Line (The ray of light that finishes at horizontal line)
- Groundline

CENTER OF PROJECTION

It is the location of the eye on which projected light rays converge.

PINHOLE CAMERA

A pinhole camera is a simple camera without a lens but with a tiny aperture.

Camera that we use in daily life is not a canvas, it is a reverse canvas where the image plane is in the backside of us .This is a pinhole model.

How to create Image : By taking **Vanishing Point** : Vanishing point is obtained by the spectator looking parallel to whatever line he is looking at.

PROPERTIES OF LIGHT

Illumination :Property and effect of light. It is the amount of light incident on a surface

Luminance : Amount of visible light that come to eye from a surface

Refraction : bending of light rays when passing through a surface between one transparent material to another

Reflectance : A portion of incident light that is reflected from a surface.

Blindspot : Spot where your optic nerve connects your retina

DIGITIZATION

Digitization is the process of converting information into a digital format.

There is a processed pipeline to convert analog to digital image.

- SAMPLING : Digitization wrt Coordinate values
  The sampling rate determines the spatial resolution of digitized image
- QUANTIZATION : Digitization wrt amplitude.
  The quantization level determines the number of grey levels in the digitized image.

REPRESENTING DIGITAL IMAGE

b = M*N*K

b of bits required to store a digitized image of size MXN

IMAGE TYPES

Binary Image : (b&w image) : Each pixel contain 1 bit (1 : black , 0: white)

Digital Image :

Monochromatic / GrayScale /Intensity Image: Pixel value can be in range 0-255

Each pixel corresponds to light intensity normally represented in gray scale.

Colour Image / RGB : Each pixel contains a vector representing Red, Green, Blue components.

Index Image : Construct a look up table and each image is denoted by an Index number and each index number has its own RGB value.

IMAGE INTERPOLATION

Interpolation : constructing new data points within the range of a discrete set of known data points.

Image Interpolation : is a tool which is used to zoom, shrink and geometric corrections of an image (re-sampling of images).

Image interpolation refers to the "guess" of intensity values at missing locations

Why Image Interpolation ?

If we want to see an image bigger - When we see a video clip on a PC, we like to see it in full screen mode.

If we want a good image : If some block of an image gets damaged during the transmission, we want to repair it

If we want a cool image - Manipulate images digitally can render fancy artistic effects as we often see in movies

ZOOMING

Zooming tells us that you are trying to expand the size of the image.

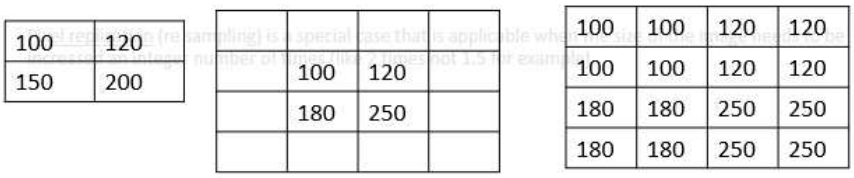
Two step procedure

- Creation of new pixel location
- Assigning gray levels to those new location

Methods :

- **Nearest Neighbourhood Interpolation**
- **Pixel Replication**
- **Bilinear Interpolation**

Nearest Neighbour Interpolation :



•Suppose an Image of Size 2x2 pixels image will be enlarged 2 times .

•Lay an imaginary 4*4 grid over the original image.

•For any point in the overlay, look for the closest pixel in the original image, and assign its gray level to the new pixel in the grid.

•When all the new pixels are assigned values, expand the overlay grid to the original specified size to obtain the zoomed image.

Limitation : it creates a checkerboard effect . When you are trying to replicate neighbourhood pixel values, sharpness of the image decreases.

Pixel Replication :

It is a special case that is applicable when the size of the image needs to be increased an integer number of times.(Eg: 5 times)

•Double the size of the image

•Duplicate each column

Bilinear Interpolation :

Resampling method that uses the distance weighted average of the four nearest pixel values to estimate a new pixel value.

DISTANCE MEASURE

If we have 3 pixels: p,q,z: *p* with (*x*,*y*) *q* with (*s*,*t*) and *z* with (*v*,*w*)

Then: D is to be distance metric iff

•$D(p,q) \geq 0$ [D(p,q)=0 iff p = q]
•$D(p,q) = D(q,p)$ (symmetry)
•$D(p,z) \leq D(p,q) + D(q,z)$ (triangular inequality)

Euclidean Distance

$$D_e(p,q) = [(x-s)^2 + (y-t)^2]^{\frac{1}{2}}$$

City Block Distance

$$D_4(p,q) = |(x-s)| + |(y-t)|$$

Chess board Distance

$$D_8(p,q) = \max(|(x-s)|, |(y-t)|)$$

COLOUR MODEL

Specification of a coordinate system and a subspace within that system where each colour is represented by a single point

Eg : RGB, CMY,HSI

RGB :

- ➔ It is the process of mixing three primary colours red, green and blue together in different proportions to make more different colours. Secondary
- ➔ It is used for digital works.
- ➔ 16,77,261 colours

CMY :

- ➔ Cyan , Magenta, yellow

[CMY] = 1- [R G B]

- ➔ It is used for print works.

Equal amount of CMY produce black

HSI : Uses 3 measures to describe colour

Hue : Indicate dominant wavelength in mixture of lightwaves

Saturation : Give a measure of degree to which RGB is diluted with white light

Intensity : Brightness is nearly impossible to measure and use to describe colour sensation

Pseudo Colour Image Processing : Assigning colour to gray level values based on Intensity slicing and gray level to colour transformation.

GEOMETRIC PRIMITIVES AND TRANSFORMATIONS

- Geometric primitives form the basic building blocks used to describe 2D and 3D shapes

- Basic geometric primitives - points, lines, conics, etc

POINTS

- Points lying on an Euclidean 2D plane(like the image plane) are usually described as vectors:

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$$

  - This is a common way to reason about points but has some limitations. For example, we cannot define points at infinity

2D PROJECTIVE SPACE

- Since the imaging apparatus usually behaves like a pinhole camera model, many of the transformations that can happen can be described as projective transformations.
- This offers a general and powerful way to work with points,lines and conics.
- The 2D projective space is simply defined as

$$\mathbb{P}^2 = \mathbb{R}^3 - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

POINT OPERATORS

Consider two images f and g defined on the same domain. Then their pixel wise addition is denoted as f+g. Or consider a positive valued image f and a, the image log(1+f) that we got by taking the logarithm of all pixel values in the image.
These two operations are examples of point operators.
Point operations refer to running the same conversion operation for each pixel in a grayscale image.

DIFFERENT POINT PROCESSING TECHNIQUES

Thresholding - select pixels with given values to produce binary images.
Adaptive Thresholding - like Thresholding except choose values locally
Contrast Stretching - spreading out gray level distribution.
Histogram Equalization - general method of modifying intensity distribution.
Logarithm Operator - reducing contrast of brighter regions.

Exponential Operator - enhancing contrast of brighter regions.

COMPOSITING AND MATTING

- In many photo editing and visual effects applications, it is often desirable to cut a foreground object out of one scene and put it on top of a different background. The process of extracting the object from the original image is often called matting , while the process of inserting it into another image (without visible artifacts) is called compositing .
- Compositing equation $C = (1 - \alpha)B + \alpha F$. This operator attenuates the influence of the background image B by a factor $(1 - \alpha)$ and then adds in the color (and opacity) values corresponding to the foreground layer F.
- The intermediate representation used for the foreground object between these two stages is called an alpha-matted color image. In addition to the three colour RGB channels, an alpha matted image contains a fourth alpha channel α (or A) that describes the relative amount of opacity or fractional coverage at each pixel . Pixels within the object are fully opaque ($\alpha = 1$), while pixels fully outside the object are transparent ($\alpha = 0$). Pixels on the boundary of the object vary smoothly between these two extremes.

HISTOGRAM EQUALIZATION

- Histogram equalization is an image processing technique that adjusts the contrast of an image by using its histogram.
- It is used to improve contrast in images. It accomplishes this by spreading out the most frequent intensity values, i.e., stretching out the intensity range of the image.

TONAL ADJUSTMENT

- One of the most used application of pointwise image processing
- Tonal adjustments are those adjustments and changes we make to the brightness and contrast of our image.

FEATURE DETECTION

Feature detection is a low-level image processing operation. That is, it is usually performed as the first operation on an image, and examines every pixel to see if there is a feature present at that pixel.

EDGES

- An edge in an image is a sharp variation of the intensity function. In grayscale images this applies to the intensity or brightness of the pixels. In colour images it can also refer to the sharp variations of colour.
- An edge can be defined as a set of connected pixels that forms a boundary between two disjoint regions. There are three types of edges:
    1. Horizontal edges.
    2. Vertical edges.
    3. Diagonal edges.
- An edge is distinguished from noise by possessing long range structure.
- Properties of edges include gradient and orientation.
- Edge detection is an important task in object recognition
- When two parallel edges meet, they form a corner.

EDGE DETECTION

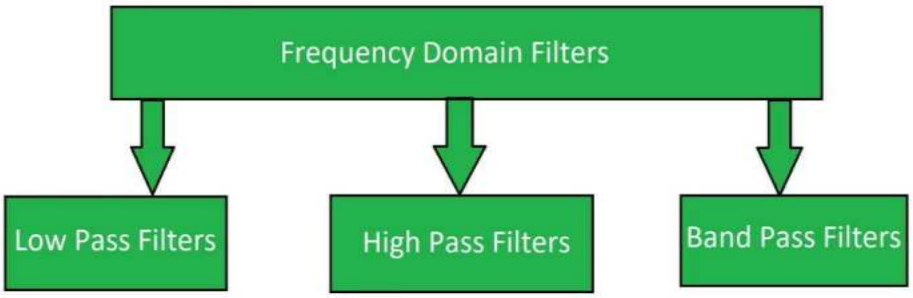
Edge detection is a method of segmenting an image into regions of discontinuity. It is widely used technique in digital image processing like:

- ➔ Pattern recognition.
- ➔ Image morphology.
- ➔ Feature extraction.

USE: Edge detection allows the users to observe the features of an image for a significant change in the gray level. This texture indicates the end of one region in the image and the beginning of another. It reduces the amount of data in an image and preserves structural properties of an image.

FREQUENCY DOMAIN FILTERS AND ITS TYPES

- Frequency domain filters are used for smoothing and sharpening of image by removal of high or low frequency components.
- Frequency domain filters are different from spatial domain filters as it basically focuses on the frequency of the images.
- It is basically done for two basic operations i.e., smoothing and sharpening



Classification of Frequency Domain Filters

Low Pass Filter : Low pass filter removes the high frequency components that means it keeps the low frequency components. It is used for smoothing the image by attenuating high frequency components and preserving low frequency components.

Mechanism of low pass filtering in frequency domain is given by:

$G(u,v) = H(u,v) . F(u,v)$,

Where F(u,v) is the Fourier Transform of the original image and H(u,v) is the Fourier transform of filtering mass.

High Pass filter : High pass filter removes the low frequency components that means it keeps the high frequency components. It is used for sharpening the image. It is used to sharpen the image by attenuating low frequency components and preserving high frequency components.

Mechanism of high pass filtering in frequency domain is given by:

$H(u,v) = 1 - H'(u,v)$,

Where H(u,v) is the Fourier Transform of high pass filtering and H' (u,v) is the Fourier transform of low pass filtering .

Band pass filter : Band pass filter removes the very low frequency and very high frequency components that means it keeps the moderate range band of frequencies. Band pass filtering is used to enhance edges while reducing the noise at the same time.

SIFT

Scale Invariant Feature Transform : is a feature detection algorithm in computer vision to detect and describe local features (KeyPoints) in images

These keypoints are scale & rotation invariant that can be used for various computer vision applications, like image matching, object detection, scene detection, etc

SIFT Detector

SIFT Descriptor

Features to consider while performing Matching

If an image has a rich content, Well define signature, Well define position in image, Should be invariant to rotation and scaling, Should be insensitive to light.

If you are locating an edge/line , a similar image is found in many places. You will not get correct patching

Techniques :

BLOB

In computer vision, **blob detection** methods are aimed at detecting regions in a digital image that differ in properties, such as brightness or color, compared to surrounding regions. Informally, a blob is a region of an image in which some properties are constant or approximately constant;

RANSAC

Random Sample Consensus - developed by Fischler and Bolles

Application : to separate inliers and outliers

Basic Idea of Ransac : We try to find the best partition of points in Inlier and Outlier set and estimate the model from Inlier set.

There will be points :

Red points : points without a "good" match in the other image In this image, the goodness of the match is decided by looking at the ratio of the distances to the second nearest neighbor and first nearest neighbor. If this ratio is high (above some threshold), it is considered a "good" match.
Blue points : These are points with a "good" match in which the match was wrong, meaning it connected two points that did not actually correspond in the world.
Yellow points : These are correct matches. We needed to run RANSAC until it randomly picked 4 yellow points from among the blue and yellow points (the matches estimated to be "good").

Randomly pick a set of data points and identify how many points are closely related to the points.

Ransac Steps :

Find the best partition of point in inlier and outlier and estimate the mode from inlier.

- Take a sample .consider the number of data points required to fit the model,
- Compute model parameter using the sample datapoint
- Score by the fraction of inliers within a preset threshold of the model. Check how many samples are agreeing with model parameters
  {agreeing samples- inliers , and not agreeing samples- outliers

  Outliers : Points which do not have a perfect match.
  Inliers : Points which are having a perfect match

SEGMENTATION

- Image segmentation focuses on partitioning an image into different parts according to their features and properties.
- The primary goal of image segmentation is to simplify the image for easier analysis. Segmentation groups together similar-looking pixels for efficiency of further processing.
    1. Full segmentation: Individual objects are separated from the background and given individual ID numbers (labels)
    2. Partial segmentation: The amount of data is reduced(separating objects from background) to speed the further processing.
- Segmentation technique can be either classified as contextual or non contextual
    1. Contextual : additionally exploit the relationships between image features. Thus, a contextual technique might group together pixels that have similar grey levels and are close to one another.

2. Non Contextual : ignores the relationship that exists between features in an image, pixels are simply grouped together on the basis of some global attribute, such as grey level.

HIERARCHICAL CLUSTERING

- Hierarchical clustering is an unsupervised machine learning algorithm which is used to group the unlabeled data sets into a cluster.
- In this algorithm, we develop the hierarchy of clusters in the form of a tree, and this tree-shaped structure is known as the dendrogram.
- The hierarchical clustering technique has two approaches to build a tree from the input set S:
    1. Agglomerative :
        - ➔ Agglomerative is a bottom-up approach, in which the algorithm starts with taking all data points as single clusters and merging them until one cluster is left. (i.e until S is achieved as the root)
        - ➔ It is the most common approach.
    2. Divisive :
        - ➔ Divisive algorithm is the reverse of the agglomerative algorithm as it is a top-down approach.
        - ➔ Recursively partitioning S until singleton sets are reached.
- Why hierarchical clustering? ,
  In hierarchical clustering algorithms, we don't need to have knowledge about the predefined number of clusters.
- Advantages ,
    1. Dendrograms are great for visualization
    2. Provides hierarchical relation between clusters
    3. Shown to be able to capture concentric clusters
- Disadvantages ,
    1. It is not easy to define levels for clusters.
    2. Experiments showed that other clustering techniques outperform hierarchical clustering.

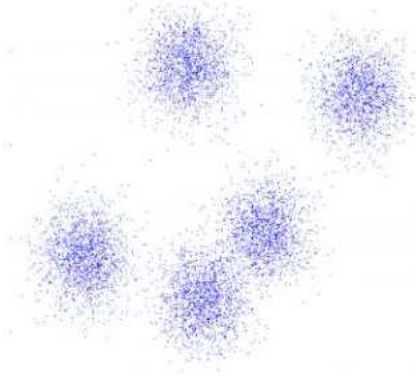
K-MEAN CLUSTERING

- K-Mean clustering is an unsupervised learning algorithm, which groups the unlabeled dataset into different clusters in such a way that each dataset belongs to only one group that has similar properties.

- Here K defines the number of predefined clusters that need to be created in the process, as if K=2,there will be two clusters, and for K=3, there will be three clusters, and so on.
- It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distance between the data point and their corresponding clusters.
- The algorithm takes the unlabeled data as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.
- The K-Mean clustering algorithm mainly performs two tasks:
    1. Determines the best value for K center points or centeroids by an iterative process.
    2. Assigns each datapoint to its closest k-center. These data points which are near to the particular k-center creates a cluster. It stops when no points' asignmentents change.

K-Mean Clustering

- Iterative Approach
    - Initialize: Pick *K* random points as cluster centers
    - Alternate:
        1. Assign data points to closest cluster center
        2. Change the cluster center to the average of its assigned points
    - Stop when no points' assignments change

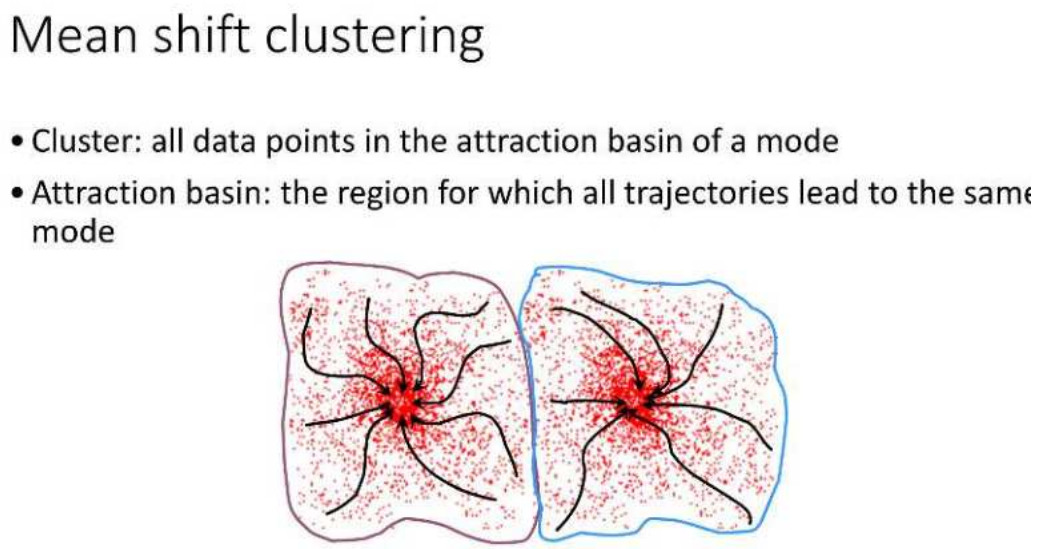


- Advantages ,
    1. Very simple method
    2. Converges to a local minimum
- Disadvantages ,
    1. Memory-intensive
    2. Need to pick K
    3. Sensitive to initialization
    4. Sensitive to outliers

5. Only find "spherical clusters"

MEAN-SHIFT CLUSTERING

- Mean-shift is falling under the category of a clustering algorithm in contrast to unsupervised learning that assigns data points to the clusters iteratively by shifting points towards the mode(mode is the highest density of data points in the region, in the context of the Mean-shift).



- Advantages ,
    1. Does not assume spherical clusters
    2. Just a single parameter(window size)
    3. Find variable number of modes
    4. Robust to outliers
- Disadvantages ,
    1. Output depends on window size
    2. Computationally expensive
    3. Does not scale well with dimension of feature space

Active Contour

Active contour is a **type of segmentation technique** which can be defined as use of energy forces and constraints for segregation of the pixels of interest from the

image for further processing and analysis. Active contour is described as an active model for the process of segmentation.

VIVA QUESTIONS

1. Quantization and Sampling

DIGITIZATION

Digitization is the process of converting information into a digital format.

There is a processed pipeline to convert analog to digital image.

- SAMPLING : Digitization wrt Coordinate values
  The sampling rate determines the spatial resolution of digitized image
- QUANTIZATION : Digitization wrt amplitude.
  The quantization level determines the number of grey levels in the digitized image.

2. Bits for representing color pixel

REPRESENTING DIGITAL IMAGE

b = M*N*K

b of bits required to store a digitized image of size MXN

IMAGE TYPES

Binary Image : (b&w image) : Each pixel contain 1 bit (1 : black , 0: white)

Digital Image :

Monochromatic / GrayScale /Intensity Image: Pixel value can be in range 0-255

Each pixel corresponds to light intensity normally represented in gray scale.

Colour Image / RGB : Each pixel contains a vector representing Red, Green, Blue components.

Index Image : Construct a look up table and each image is denoted by an Index number and each index number has its own RGB value.

1bit Image : (Binary Image) Each pixel is stored as a single bit (0 or 1) .

8bit gray level Image : Each pixel has gray value b/w 0-255.
Each pixel is represented by 1 bit.

24bit colour image : Each pixel is represented by 3 bytes representing RGB.

256 x 256 x 256 colors (16,777,216 colours)

For colour pixel : 8bits

3. why we will get different colors by having different intensities of RGB

Colour Mixing Theory

Additive color mixing theory deals with mixing of light. The primary colors red, blue and green can be paired to form white (red, blue and green), magenta (red and blue), yellow (red and green) and cyan (green and blue).

4. Interpolation

Interpolation : constructing new data points within the range of a discrete set of known data points.

Image Interpolation : is a tool which is used to zoom, shrink and geometric corrections of an image (re-sampling of images).

Image interpolation refers to the "guess" of intensity values at missing locations

Why Image Interpolation ?

If we want to see an image bigger - When we see a video clip on a PC, we like to see it in full screen mode.

If we want a good image : If some block of an image gets damaged during the transmission, we want to repair it

If we want a cool image - Manipulate images digitally can render fancy artistic effects as we often see in movies