

# **Computer Vision and Image Processing: Fundamentals and Applications**

**Professor M. K. Bhuyan**

**Indian Institute of Technology, Guwahati**

**Lecture - 2**

## **Introduction to Digital Image Processing**

Welcome to NPTEL MOOC's course on Computer Vision and Image Processing, Fundamentals and Application. So, in my last class, I discussed about some fundamental concepts of computer vision and also, I have highlighted some applications of computer vision.

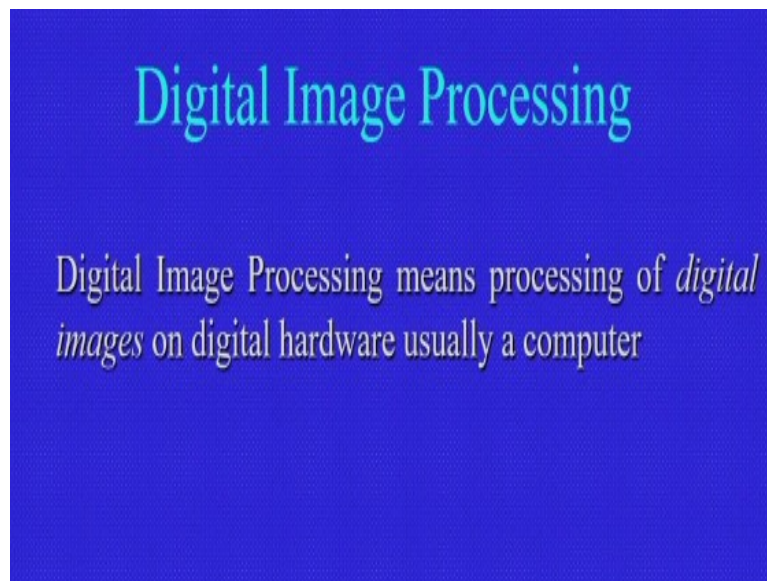
So, in my class today, I will discuss the introduction to Digital Image Processing. So, what do you mean by digital image? So, I can say, digital image means 2D array of quantize intensity values or maybe a 2D array of numbers. So digital image processing means how to manipulate these numbers.

Suppose, if I want to improve the visual quality of an image, then in this case, I have to manipulate these numbers. These numbers mean the quantized intensity values. And suppose, if I want to remove noises then also I have to manipulate these numbers. So, manipulation of these numbers is mainly the image processing.

In case of the image acquisition, the light is converted into electrical signal. So, I am getting analog signal, the analog image; the analog image can be converted into digital image by the process of sampling and the quantization. So, I have to do the sampling along the x direction and sampling along the y direction and that is called the spatial sampling.

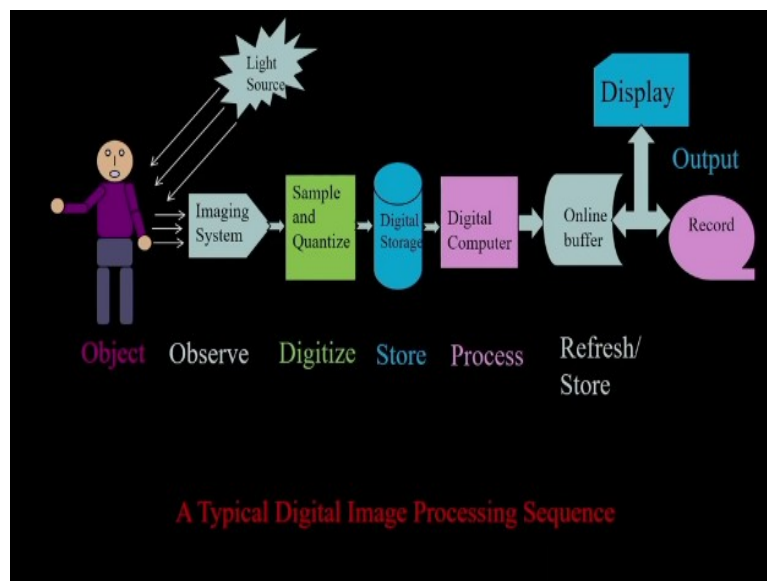
After sampling, I have to do the quantization of the intensity values. So, to convert an analog image into digital image, I have to do sampling and the quantization. So, what is digital image, I will show you.

(Refer Slide Time: 02:07)



Digital image processing means processing of digital images on digital hardware, usually a computer. So, this is the definition of digital image processing; that is the manipulation of the quantize intensity values, the manipulation of the numbers.

(Refer Slide Time: 02:22)



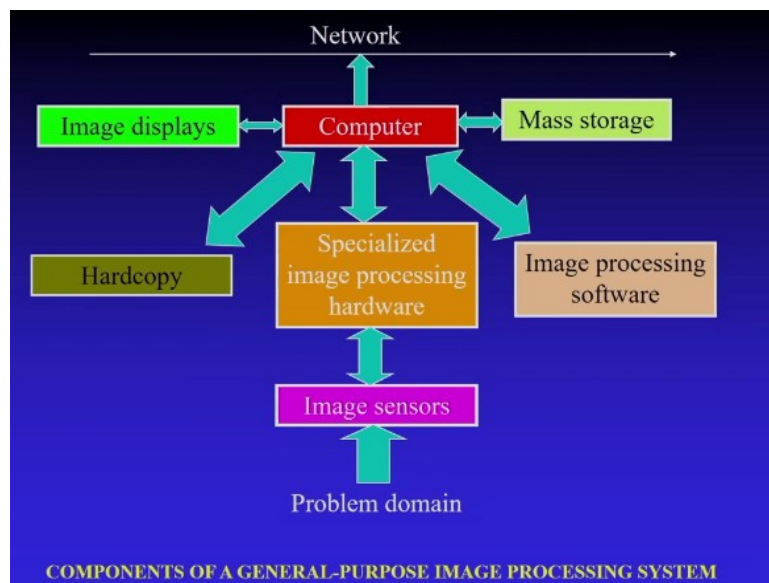
Now, here I have shown a typical digital image processing sequence. So, if you see, the light is coming from the source and it is reflected by the object and I have the imaging system. So, this imaging system converts the light photon into electrical signal.

And in this case, I am getting the analog image. So analog means it is a continuous function of time. To convert the analog signal, the analog image into digital, I have to do the sampling

and the quantization. So, sampling in the x direction, sampling in the y direction and that is called spatial sampling.

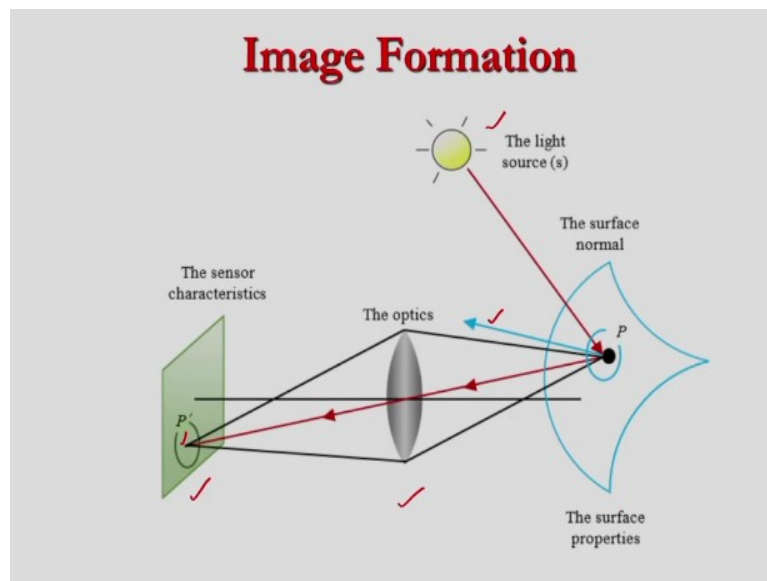
And after this, I have to do quantization of the intensity value. After this I am getting the digital image. This digital image I can store into digital storage, I can process the image in the digital computers, I can display the image, the processed image and I can store the image. So, this is a typical digital image processing sequence.

(Refer Slide Time: 03:14)



The same thing I have shown here in this figure. So, image sensors are available for image acquisition and after this, I will be getting the digital image and that digital image, I can process by image processing hardware. And I have also image processing software, so the process image can be displayed and it can be stored. This is the components of a general-purpose image processing system.

(Refer Slide Time: 03:40)



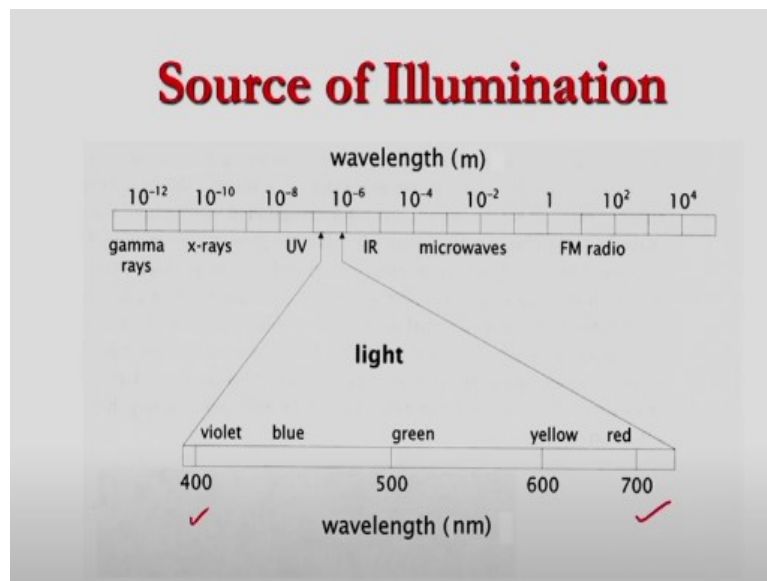
The concept of image formation here, I have shown and the light is coming from the source and it is reflected by the surface. So, I have shown the surface normal also. So, suppose, this is my surface normal and I have the optics, optics is mainly the camera.

So, light is reflected by the surface and I am getting the image in the image plane, that is, the sensors. So, sensor means it converts the light photon into electrical signal. This pixel intensity value depends on the amount of light coming from the surface.

So, if I consider one pixel at this point, this pixel value, the pixel intensity value depends on the amount of light reflected by the surface. So that means, that the pixel value actually depends on the surface property.

The surface property is called the reflectance property of the surface, that is, the amount of light reflected by the surface. So, this is a typical image formation system.

(Refer Slide Time: 04:40)



So, I have shown in this case, this is the electromagnetic spectrum. The source of illumination I have shown the gamma rays, X-rays, ultraviolet, infrared ray, microwaves, FM radio. And if you see the wavelengths from 400 to 700 nanometers, this is the visible range of the spectrum, the electromagnetic spectrum.

(Refer Slide Time: 05:08)

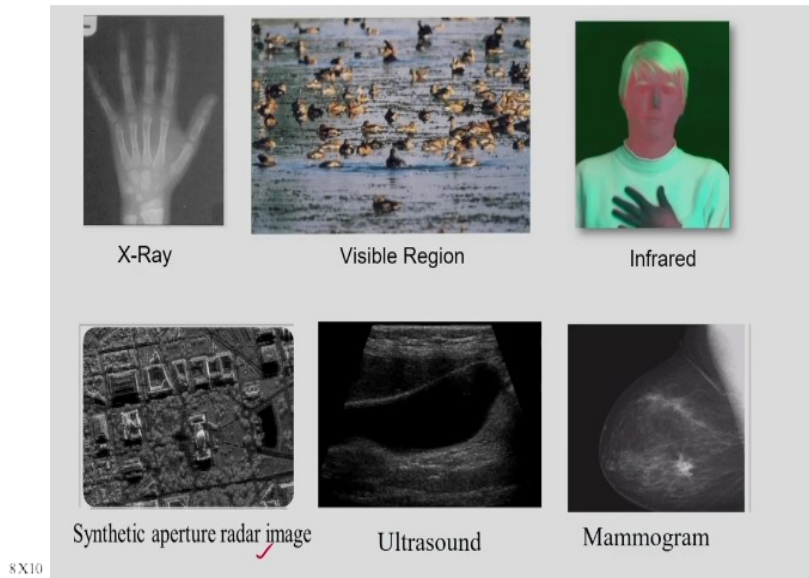
## Types of Digital Image

- Digital photos.
- Image sequences used for video broadcasting and playback.
- Multi-sensor data like satellite images in the visible, infrared and microwave bands.
- Medical images like ultra-sound, Gamma-ray images, X-ray images and radio-band images like MRI etc.
- Astronomical images.
- Electron-microscope images used to study material structure.

So, based on this source of illumination, I may have these types of images; may be the digital photo, the image sequence used for video broadcasting, and something like the multi-sensor data like satellite image, visible image, infrared images, and an image of them obtained from the microwave bands, medical images like ultrasound, gamma-ray images, X-ray images, and

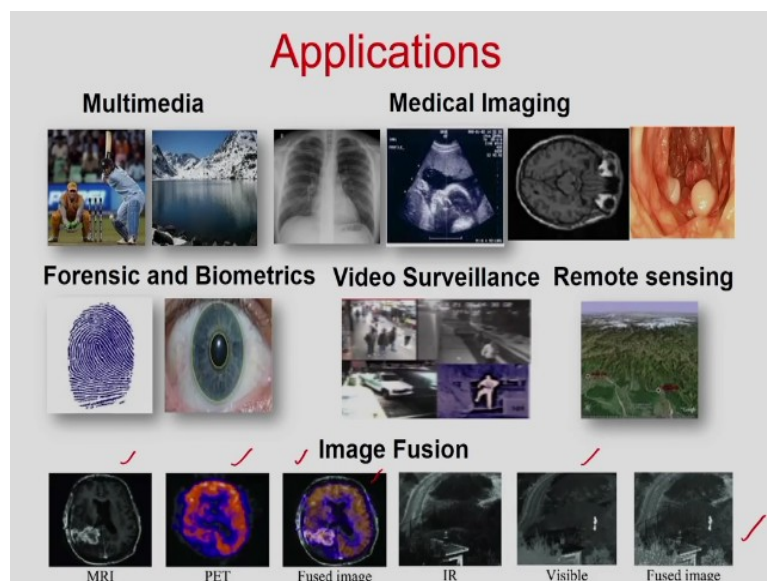
the radio band images like the MRI images, astronomical images. So, based on this electromagnetic spectrum, I may have these types of images.

(Refer Slide Time: 05:44)



So here, I have given some examples. The first one is the X-ray, and next one is the photo that is obtained in the visible region of the electromagnetic spectrum, the next one is the infrared image, another one is the radar image; so, this is the radar image; ultrasound image, and the mammograms. So different types of images.

(Refer Slide Time: 06:06)



And some of the applications of image processing, here I have shown. The applications like multimedia, medical imaging, that is, medical image processing and medical image analysis.

I have many applications even in the forensic in the biometrics, like the fingerprint recognition, the iris recognition, video surveillance, remote sensing. There are many applications of image processing.

And one important application of image processing is image fusion. If you see, I have considered in the first example, I have two images, one is the MRI image another one is the PET image. So, these two images are fused and I am getting the fuse image.

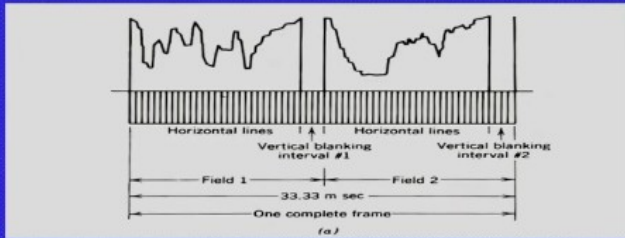
So that means the important information from MRI images and the PET images, we have considered and the redundant information I am neglecting, I am not considering. So, this is the fused image. In this fused image, I have more information as compared to the original source images. So, I have two images, one is the MRI image, another one is the PET image, and the fused image I am getting from these two images.

In a second example, I have considered the IR image and the image of the visible region of the spectrum. These two images I am fusing and I am getting the fuse image. So, it has more information as compared to the source images. And in this case, you will remember that I am neglecting the redundant information, I am not considering the redundant information. So, these are the applications of image processing.

(Refer Slide Time: 07:37)

### What is an analog image?

- Electrical Signal, for example, the output of a video camera, that gives the electric voltage at locations in an image



The diagram illustrates the timing of an analog video signal. It shows two fields, Field 1 and Field 2, each consisting of horizontal lines representing image data. Between the fields are vertical blanking intervals, labeled as Vertical blanking interval #1 and Vertical blanking interval #2. The total duration of one complete frame is indicated as 33.33 m sec.

So, what is analog image? Analog means it is a continuous function of time. This is the definition of the analog image. So here I have shown one analog signal that is obtained from the camera.

So, I have the horizontal lines and the vertical, the blanking signals; this is the, mainly the video signal. So, this analog signal can be converted into digital signal by the process of sampling and the quantization.

(Refer Slide Time: 08:09)

PAL → 625 lines  
 NTSC 25 frames/sec  
 SECAM

Bandwidth of Video signal = 4 MHz  
 Sampling frequency = 8 MHz  
 No. of pixels per frame =  $\frac{8 \times 10^6 \text{ samples/sec}}{25}$   
 $= 320000 \text{ pixels/frame}$   
 $\approx 512 \times 512$

$M \times N$

So, in TV standard, I know some of the standard like the PAL, the phase alternating by lines. So, this is one TV standard. Another standard is you know, NTSC. This is another standard. Another standard is SECAM. So, I am not going to discuss about this, but in India, we use the PAL systems.

So, in the PAL system, we have 625 lines; 625 lines. And in this case, we have considered the 25 frames per second; 25 frames per second. In phase PAL system, the phase alternating lines, this is a TV standard and we have considered 625 lines and 25 frames per second. In NTSC, generally we use the 30 frames per second. So, I am not going to discuss about the TV standard.

Now, let us consider, suppose a video signal, a video signal I am considering and suppose the bandwidth of the video signal, bandwidth of video signal I am considering, suppose it is 4 megahertz. Let us consider this example, 4 megahertz. So, what will be my sampling frequency then?

Bandwidth is 4 megahertz, so sampling frequency will be, sampling frequency, sampling frequency will be my 8 megahertz, as per the Nyquist sampling theorem, so my sampling frequency will be 8 megahertz.

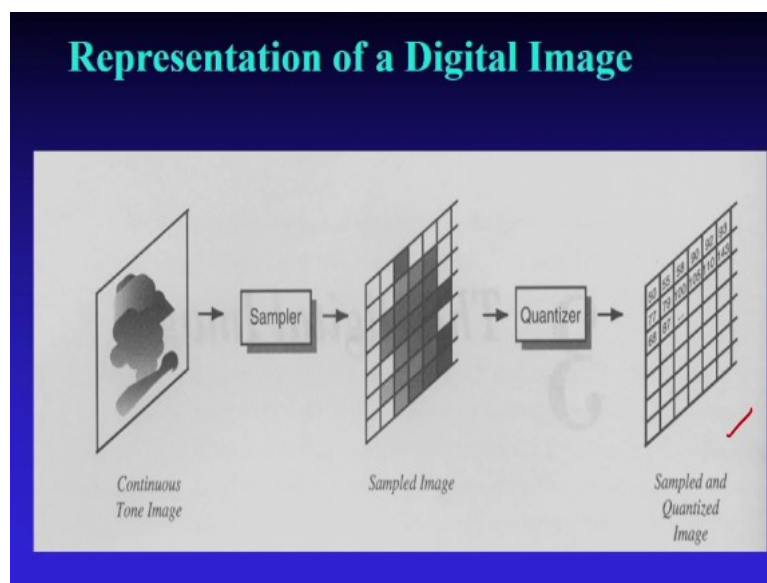


Now, I want to count the number of pixels per frame. So, I want to count the number of pixels per frame. So how many pixels per frame, that I want to determine. So, 8 megahertz,  $8 \times 10^6$ , that is samples per seconds, samples per seconds; divided by, I am considering 25 frames per second, 25 frames per second. So, this will be equal to 320000. This is the digital pixels per frame, pixels per frame, pixels per frame. So, this is approximately the  $512 \times 512$ .

So that means, the image size will be  $512 \times 512$ . The image is represented by  $M \times N$ . This is the size of the image. The image is represented by  $M \times N$ , this is the image because it is a 2D array of numbers. So, I have M number of rows and N number of columns. So, my rows are like this, these are my rows, rows and I have the columns, like this.

So, the image is represented by  $M \times N$ . So, in this example, I have seen that the size of the image will be  $512 \times 512$ . So, this I have shown this example.

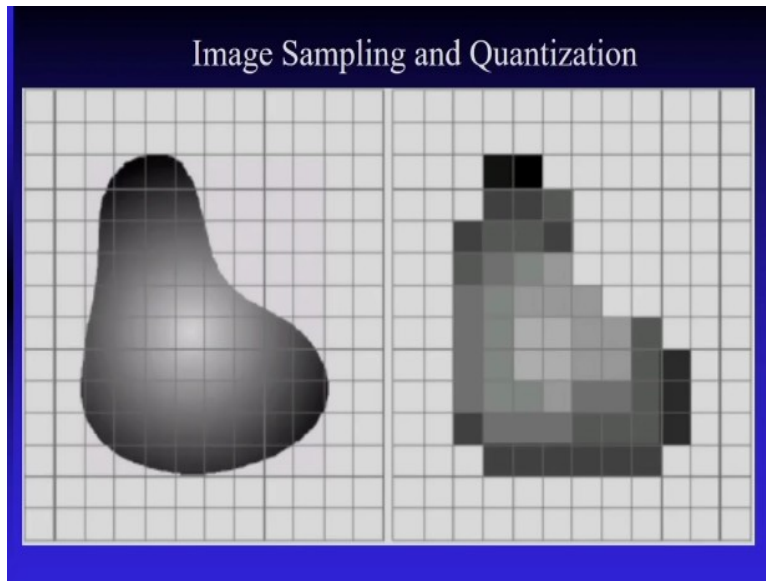
(Refer Slide Time: 11:39)



Now, this already I have explained how to get the digital image. So, in this diagram, you have seen that I have shown one continuous tone image that is the analog image and after this, I have to do the sampling; the sampling along the  $x$  direction and the sampling along the  $y$  direction I have to do.

And after this, I have to do the quantization; the quantization of the intensity values. So, I have done the quantization and after, this I am getting the 2D array of numbers. So, this is a sample and the quantized image I am getting. So, this is a representation of a digital image.

(Refer Slide Time: 12:13)



The same thing I am showing the sampling and the quantization. So, sampling along the  $x$  direction and the sampling along the  $y$  direction I have shown and after this, I am doing the quantization of the intensity value.

(Refer Slide Time: 12:25)

### What is a Digital Image ?

- **Sample** the 2D space on a regular grid
- **Quantize** each sample (round to nearest integer)

If our samples are  $\Delta$  apart, we can write this as:

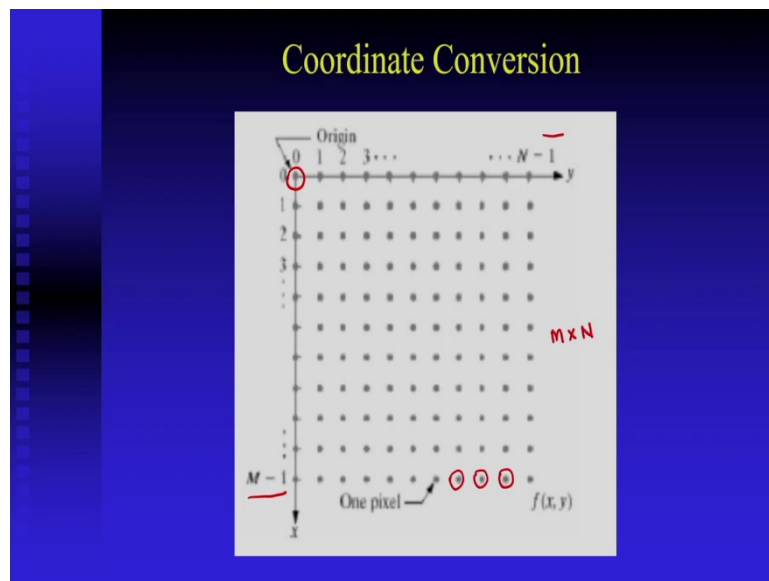
$$f[i, j] = \text{Quantize}\{f(i \Delta, j \Delta)\}$$

	$j \rightarrow$							
$i \downarrow$	62	79	23	119	120	105	4	0 ✓
	10	10	9	62	12	78	34 ✓	0
	10	58	197	46	46	0	0	48
	176	135	5	188	191	68	0	49
	2	1	1	29	26	37	0	77
	0	89	144	147	187	102 ✓	62	208 ✓
	255	252	0	166	123	62	0	31
	166	63	127	17	1	0	99	30

So, sampling the 2D space on a regular grid and after this I have to quantize each sample value. First, I have to do the sampling and after this I have to do the quantization. So, this this is first I am doing the sampling and after this, I am doing the quantization. So, it is 2D array of numbers and this is the digital image.

So, I have shown the values. The value is the 0 is the grayscale value. The 34 is the grayscale value, this 102 is the pixel value. So, I am considering this is one digital image.

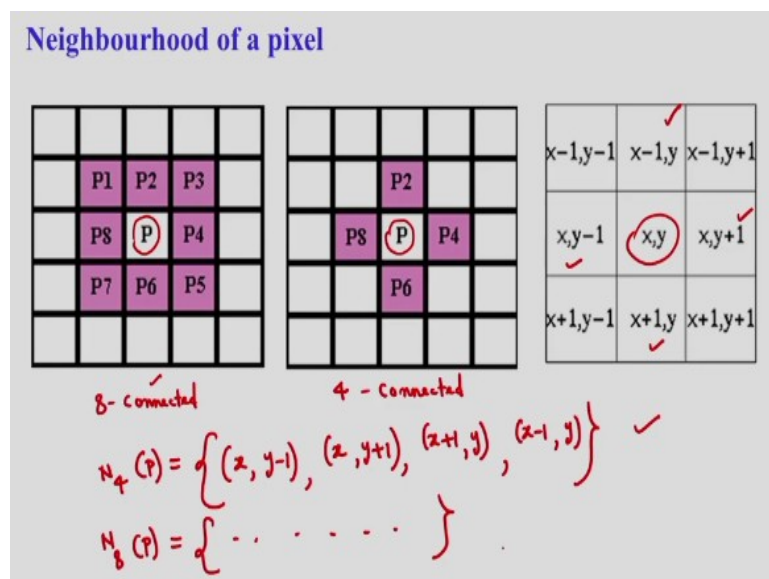
(Refer Slide Time: 12:58)



And in this case, already I have shown the image is represented by  $M \times N$ .  $M \times N$  is the size of the image. So,  $M \times N$  is the size of the image. So here I have shown the rows. The  $M$  number of rows I am considering and the  $N$  number of columns starting from 0 to  $N - 1$  and 0 to  $M - 1$ .

And this is the origin, the origin of the image, that is, the point is  $(0,0)$ . And I have shown the pixels; these are the pixels, the pixels of the image. These are the pixels of the image.

(Refer Slide Time: 13:34)



Now, let us consider the neighborhood of a pixel. So, in this figure, I have shown and a particular suppose pixel, the pixel is P. And this pixel I am considering P, I have shown two cases. One is the neighbourhood, the 8 neighborhood I have considered.

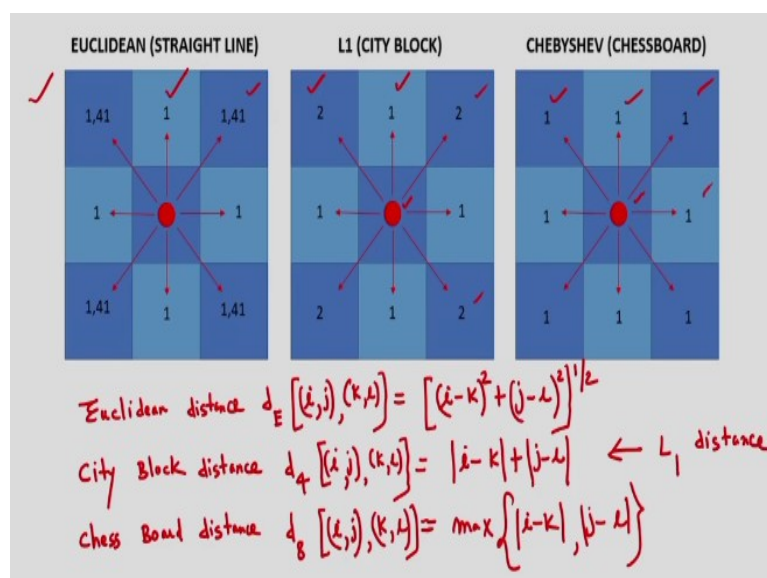
Corresponding to the first case, I am considering the neighborhood pixel the P1, P2, P3, P4, P5, P6, P7, P8, that is called the 8 connected neighborhood. In the second case, if you see the neighborhood pixel I am considering P2, P4, P6, P8. Corresponding to the center pixel, the center pixel is  $x, y$ ; this is center pixel.

So, if you see here, the first one is the 8-connected neighbourhood, this is the 8-connected neighbourhood, 8-connected neighbourhood. The second one is the 4-connected neighbourhood, the 4-connected neighborhood. So, this 4-connected neighborhood I can write like this N4 P, 4-connected neighborhood.

So, this pixel I have to consider, this pixel I have to consider, this pixel I have to consider, this pixel I have to consider that means it will be  $x, y - 1$ , that is the one pixel;  $\{(x, y + 1), (x + 1, y), (x - 1, y)\}$ . So, this is the 4-connected neighborhood.

Similarly, you can define the 8-connected neighborhood. So, in the, this is the neighborhood of the 4-connected neighborhood, the second one is the 8-connected neighborhood.

(Refer Slide Time: 15:31)



Now, you can see, corresponding to this neighborhood, I can determine the Euclidean distance, the city block distance, and also the chessboard distance. So how to define the distances?

This Euclidean distance, it is defined like this. Euclidean distance, Euclidean distance is suppose  $d_E$ ; Euclidean distance is  $d_E$ . So, suppose I am considering two points  $(i, j)$  and  $(k, l)$ .

So,  $[(i-k)^2 + (j-l)^2]^{\frac{1}{2}}$ , this is the Euclidean distance.

Another distance is L1 distance, that is called the city block distance, city block distance. The city block distance I can show it is suppose, the  $d_4$  distance, city block distance suppose. So, it is the distance between  $(i, j)$  and  $(k, l)$ . So, it is  $|i-k| + |j-l|$ , this distance is called the L1 distance. It is the City block distance or the, it is called the L1 distance, L1 distance.

Another distance is the chessboard distance or sometimes it is called a Chebyshev distance. So, I can define suppose the it is  $d_8$  between two pixels  $(i, j)$ ,  $(k, l)$ . So, I have to consider this distance; the distance is maximum value I have to determine,  $\max(|i-k|, |j-l|)$ .

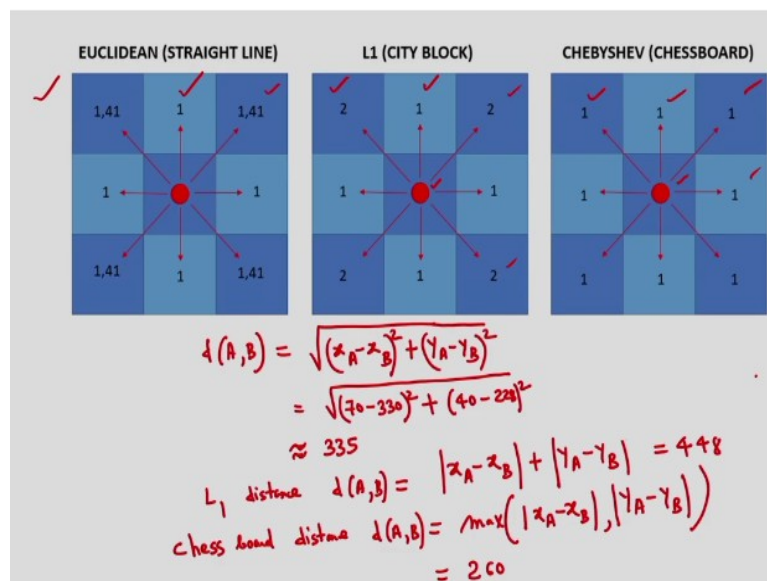
So, these distances I have defined. One is the Euclidean distance, one is the city block distance, another one is the chessboard distance.

So, corresponding to these distances you can see, in the first figure, this figure, I have determined the distance between the center pixel, the center pixel is this pixel and the neighborhood pixels. You can see the distance is 1.41 to the diagonal pixel and to this pixel it is the distance is 1, to this pixel it is 1.41; like this I am getting the distances. All the distances I can compute to the neighborhood pixel.

Similarly, if I consider the city block distance corresponding to the center pixel, the center pixel is this, I can find a distance between the neighborhood pixel. So, if I find the distance between the center pixel and this pixel the distance will be 2. This distance will be 1, this distance will be 2, like this, I am determining the distance. I am calculating the distance.

In the third case, I am considering the chessboard distance. So, corresponding to the center pixel, this center pixel, I am determining the distance between the neighborhood pixels and in this case, I am getting 1 1 1 1, like this distance I am getting. So, you can see these distances I can compute.

(Refer Slide Time: 19:26)



So I can give one example. Suppose, if I want to determine the distance between  $d$ , two points  $A$  and  $B$ , that is, the, suppose the Euclidean distance I can determine between two points,  $\sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$ .

So suppose this  $x_A$  is some,  $x_A$  is suppose 70 minus 330 suppose  $x_B$  is 330, I am taking one example, whole square plus suppose this  $y_A$  is 40 and  $y_B$  is 228 suppose whole square. So, this value I am getting it is 335 almost 335 this value. This value is 335. This is the Euclidean distance I can determine.

And the L1 distance, if you can compute L1 distance, then this distance will be  $d(A, B)$  will be you can see this is  $|x_A - x_B| + |y_A - y_B|$ . So, you will be getting the value something like 448. Also, you can determine the chessboard distance, also you can determine.

The Chessboard distance  $d(A, B)$  between two points, thus maximum you have to take and it is  $\max(|x_A - x_B|, |y_A - y_B|)$ . So, this value you will be getting something like 260. You can verify this one.

So, in this example, I have shown how to calculate the Euclidean distance, city block distance, and the chessboard distance. So, this is the important the neighborhood of the pixel a particular pixel.

(Refer Slide Time: 21:47)

## What is a digital Image.....?

- 2D array of numbers representing the sampled version of an image
- The image defined over a grid, each grid location being called a pixel.
- Represented by a finite grid and each intensity data is represented a finite number of bits.
- A binary image is represented by one bit. Gray-level image is represented by 8 bits.

■ Pixel and Intensities

16	18	19	20
255	19	18	20
21	21	22	30
21	33	22	23

So, what is digital image? The digital image is nothing but the 2D array of numbers representing the sample version of an image. The image defined over a grid and each grid location is called the pixels.

So, I have shown the pixels, these are the pixels, and represented by a finite grid and its intensity data is represented by a finite number of bits. So that concept I can explain like this.

(Refer Slide Time: 22:16)

$M \times N$   
 $[0 \dots L-1] \rightarrow [0 \dots 1]$   
 $L = 2^k$   
No. of bits reqd to store an image =  $M \times N \times k$

Dynamic range  
Resolution of an image

- Spatial resolution → no. of pairs of lines per unit distance
- Intensity resolution → smallest discernible changes in the intensity level  
 $2^8 = 256$

Video → frames/sec → Temporal resolution.

Suppose, the image is represented like this  $M \times N$ ; the size of the image is  $M \times N$ , that is,  $M$  means the number of rows,  $N$  number of columns. And the grayscale value suppose, I am

considering  $L$  number of grayscale values. So, 0 to  $L-1$ , grayscale value that can be normalized between 0 to 1.

Suppose, the  $L$  number of grayscale levels,  $L$  is equal to  $2^K$ . So now, the number of bits, number of bits required to store an image is equal to  $M \times N \times K$ . That is the number of bits required to store an image is equal to  $M \times N \times K$ .

Now, what is the definition of the dynamic range? Suppose, I have an image, the highest pixel value and a lowest pixel value I know. So, the difference between this is called a dynamic range. The contrast of an image depends on the dynamic range. So, for a good contrast image, the dynamic range will be more.

Now, the resolution of an image, resolution of an image. So how to define the resolution of the image? So, one is, one resolution is the spatial resolution, the spatial resolution. The spatial resolution means, I can say the number of pairs of lines per unit distance. This is the definition of the spatial resolution.

And another one is the intensity resolution. So, intensity resolution, I can say that the definition is smallest discernible, smallest discernible changes in the intensity level. So, suppose if I consider 8 bits, so,  $2^8=256$ . That means, how many intensity levels I am using? The 256 numbers of intensity level I am using.

So spatial resolution means the number of pairs of lines per unit distance. That means, so how many pixels, if I consider these are the pixels, so how many pixels per unit distance that is called spatial resolutions.

And intensity resolution means the smallest discernible changes in the intensity level. So how many intensity levels I am using that is the intensity resolution. And if I consider the video, in video I have to consider, suppose, if I consider a video, in video I have to consider how many frames, frames per second? How many frames per second that corresponds to the temporal resolution.

So, for video, I have temporal resolution and for the image, I have spatial resolution and the intensity resolution. And for the binary image, the binary image is represented by 1 bit. So that concept I can show you.



(Refer Slide Time: 26:22)

## What is a digital Image.....?

- 2D array of numbers representing the sampled version of an image
- The image defined over a grid, each grid location being called a pixel.
- Represented by a finite grid and each intensity data is represented a finite number of bits.
- A binary image is represented by one bit. Gray-level image is represented by 8 bits.

■ Pixel and Intensities

16	18	19	20
255	19	18	20
21	21	22	30
21	33	22	23

So here, I have shown the pixels and the intensities. This is the two dimensional array of numbers and here, I have shown the pixel value 255, 18, 19. So these are the pixel values.

(Refer Slide Time: 26:57)

$M \times N$   
 $[0 \ L-1] \rightarrow [0 \ 1]$   
 $L = 2^k$   
No. of bits reqd to store an image =  $M \times N \times k$

Dynamic range

Resolution of an image

- 0-255  $\rightarrow$  8 bits
- 0-127  $\rightarrow$  7
- 0-63  $\rightarrow$  6
- 0-31  $\rightarrow$  5
- 0-15  $\rightarrow$  4
- 0-7  $\rightarrow$  3
- 0-3  $\rightarrow$  2
- 0-1  $\rightarrow$  1  $\leftarrow$  Binary image

Spatial resolution  $\rightarrow$  no. of pairs of lines per unit distance

Intensity resolution  $\rightarrow$  smallest discernible changes in the intensity level

$2^8 = 256$

Video  $\rightarrow$  frames/sec  $\rightarrow$  Temporal resolution.

So already I have defined the resolution of an image, one is spatial resolution, another one intensity resolutions. So, suppose if I consider, suppose 0 to 255 levels; 0 to 255 levels. So how many, how many bits we need?

For this, we need 8 bits. And if I consider 0 to 127 then I need 7 bits; if I consider 0 to 63 then in this case, I need 6 bit; if I consider 0 to 31 then in this case, I need 5 bits; if I consider 0 to 15 then in this case, I will need 4 bits; if I consider 0 to 7 then I need 3 bits; if I consider

0 to 3, I need 2 bits; and if I consider 0 to 1, only two levels, 0 to 1 means two levels, so I need only one bit.

So, this is called the binary image, binary image. So, this is the binary image. So binary image needs only 1 bit. So, a binary image is represented by 1 bit and a grayscale, gray-level image is represented by 8 bits.

(Refer Slide Time: 28:19)

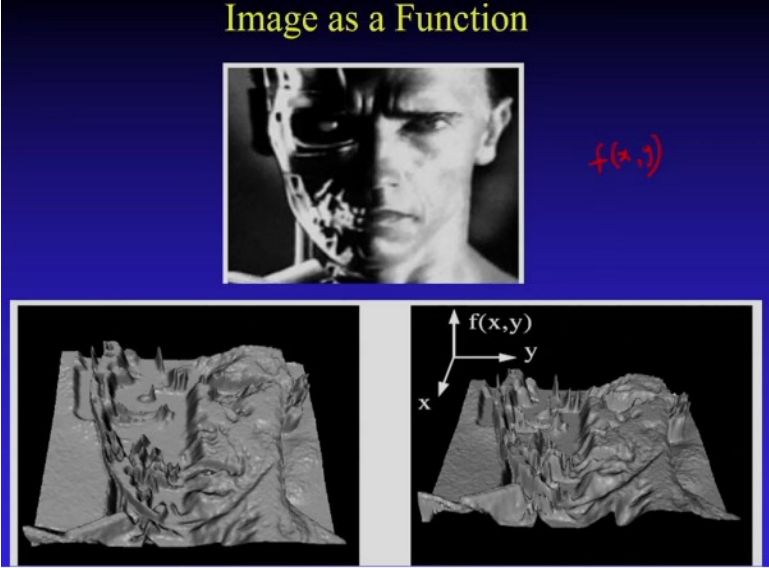
## Mathematically

- We can think of an **image** as a function,  $f$ , from  $\mathbb{R}^2$  to  $\mathbb{R}$ :
  - ◆  $f(x, y)$  gives the **intensity** at position  $(x, y)$
  - ◆ Realistically, we expect the image only to be defined over a rectangle, with a finite range:
    - ◆  $f: [a, b] \times [c, d] \rightarrow [0, 1]$

And the image is represented by  $f(x, y)$ . What is  $f(x, y)$ ?  $f(x, y)$  means the intensity at a position the position is  $(x, y)$ . This is the definition of image. The image is represented by  $f(x, y)$ .

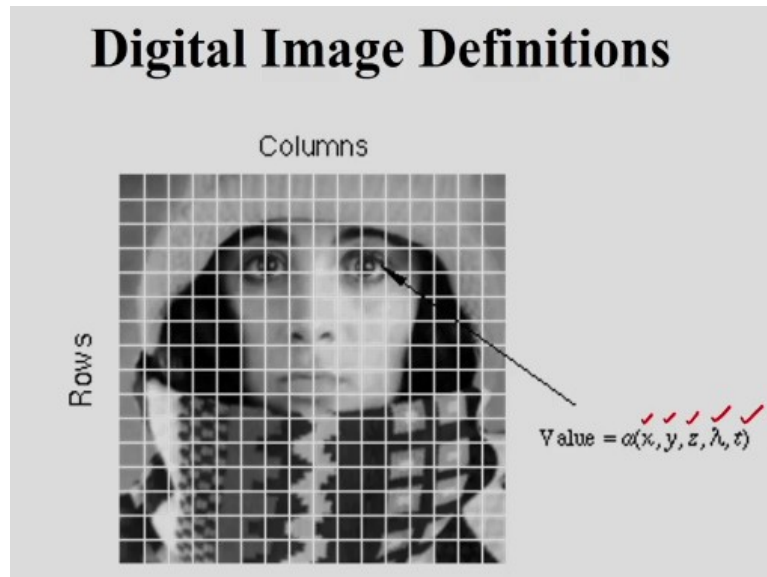
(Refer Slide Time: 28:33)

## Image as a Function



So here, I have shown the image as a function because image is nothing but this  $f(x, y)$ ;  $f(x, y)$ . So, the intensity value I have shown  $f(x, y)$  here.

(Refer Slide Time: 28:46)



And there is another definition of a digital image. In this definition, I have shown the rows and the columns. So, I have shown the pixels. So, in this definition, if you see, the image is represented by these parameters, one is  $x$  and  $y$  coordinate; that is the spatial coordinate. And if I consider the stereo image, then one important parameter is the depth information. So that corresponds to the depth information.

So, suppose if I have two images obtained by two cameras, so from two images, I can determine the depth information. So, in this case, the  $z$  represents the depth information. And if I consider the color image, so  $\lambda$  means the color information. And if I considered a video, video means the number of frames per second. So I have to consider temporal information, so  $t$  is the temporal information.

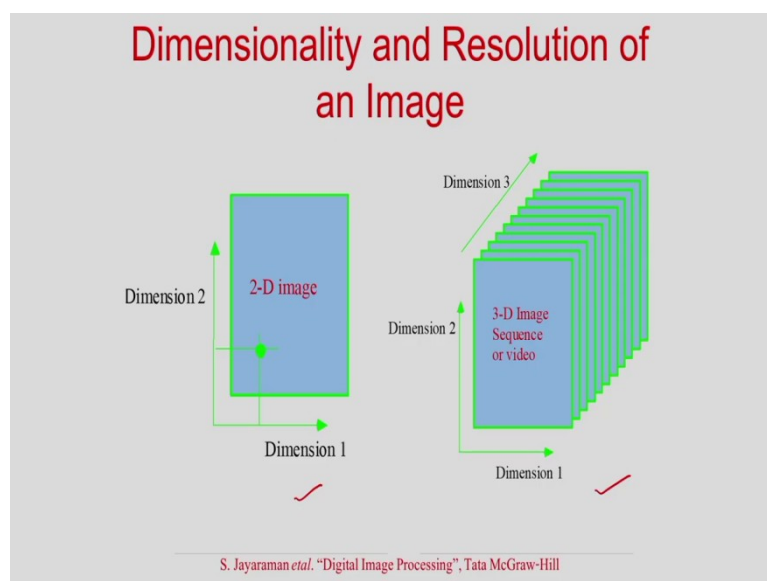
So image is represented by  $x, y, z, \lambda, t$ ;  $x$  and  $y$  is the spatial coordinate,  $z$  is the depth information,  $\lambda$  is the color information, and  $t$  is the temporal information. That is one definition of digital image.

(Refer Slide Time: 29:54)

Parameter	Symbol	Typical values
Rows	$N$	256,512,525,625,1024,1035
Columns	$M$	256,512,768,1024,1320
Gray Levels	$L$	2,64,256,1024,4096,16384

And some of the typical values, number of rows, number of columns. The 256, 512, 525, 625 like this, these are the number of rows; number of columns 256, 512; and gray-level values something like 2, 64, 256 like this, we have the values.

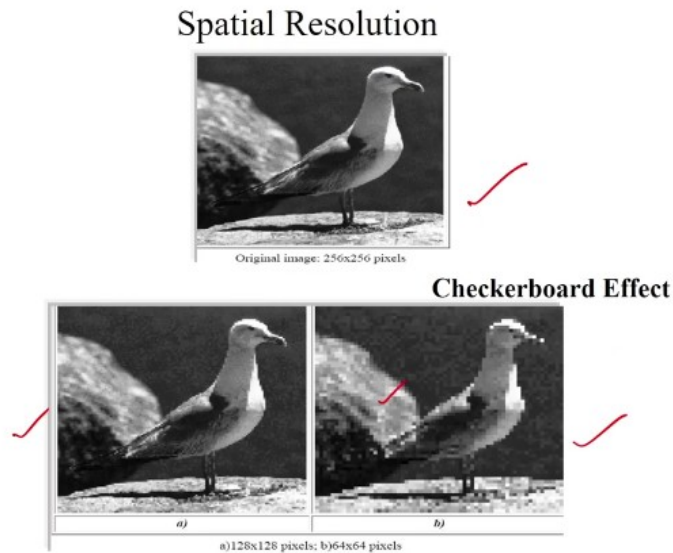
(Refer Slide Time: 30:15)



And in this case, I have shown the dimensionality and resolution of an image. So, in this case, I have shown the image, the image is 2 dimension and another case, I have shown 3 dimensions that is for the video, x and y coordinate and another one is number of frames per second, that is, the temporal information.

So here, I have shown that dimensionality and the resolution of an image. So. image is 2 dimension and video is 3 dimension.

(Refer Slide Time: 30:41)



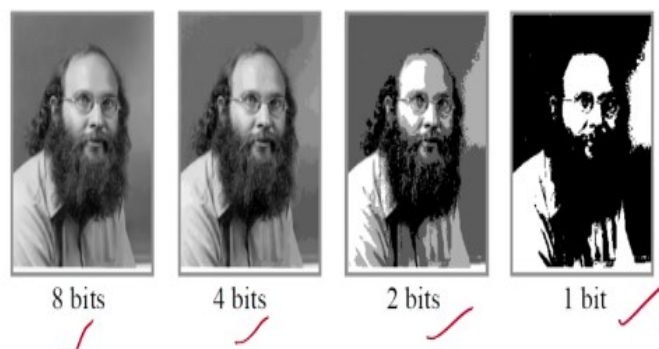
So in this example, I have shown, I have changed the spatial resolution. If you see the first image, the first image is  $256 \times 256$  pixels; the second image is  $128 \times 128$ , this image; the third image is  $64 \times 64$ .

So that means I am decreasing spatial resolution and because of this, I am getting some effects like checkerboard effect. This effect I have observed here and this is called a checkerboard effect because I am decreasing the spatial resolution of an image.

(Refer Slide Time: 31:13)

## Different Gray level Resolution

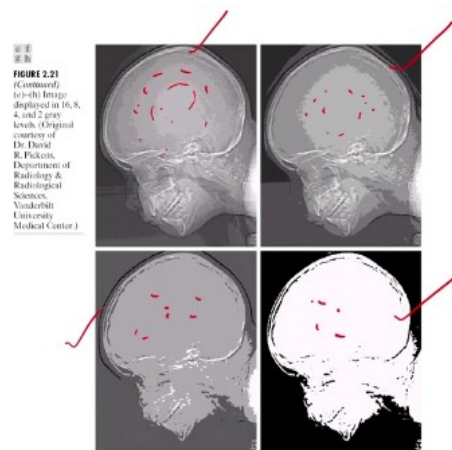
- Images with decreasing bits per pixel:



Now, I am considering changing the intensity resolution that is called the gray-level resolution. So, first example is I am considering the 8 bit image, next one is the 4 bit image, next one is 2 bit image, and 1 bit image is nothing but the binary image. So I am changing the gray-level resolution the intensity resolution.

(Refer Slide Time: 31:36)

**False contouring in image** (insufficient numbers of intensity levels)



This is another example. The initial, if you see the first image, this is 16 gray levels, second one is the 8 gray levels, the next one is 4 gray levels, and this is the 2 gray levels that means the binary image. So, if you see this, this portion, this portion, suppose if I consider, this portion is not properly represented in these images, in these images. So, this effect is called the false contouring effect.

So, I have two effects one is the checkerboard effect that is because of the spatial resolution, and another one is a false contouring effect because of insufficient number of intensity levels. So, I have given these two examples.

(Refer Slide Time: 32:17)

## Image Types

RGB

Grayscale

Black & White



The image types I have shown here. I have RGB image, the grayscale image, and the black and white image that is the binary image. So, in case of the color image, corresponding to a particular pixel, I have 3 components; one is the R component that is the red, another one is the G component that is the green; and the blue component that is the B.

So RGB value that is the primary colors. So, corresponding to a particular pixel, I have 3 components; R component, G component, and the Blue component. And this pixel, particular pixel is called the vector pixel, because I have 3 components.

(Refer Slide Time: 32:57)

### Binary Image


- Efficient in terms of storage
- Document processing, handwriting, fingerprint




1	1	0	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	1	1	0
0	0	0	0	0	1

So, this is one example of the binary image, basically the binary is 1 and 0 only. So if I apply the edge detection principle, the edge detection principle or edge detection, then in this case, I will be getting the binary image. So, this is one example of the binary image.

(Refer Slide Time: 33:16)

 Grayscale Image

- The range is usually a power of 2
- 256 levels are sufficient for most applications ✓



220	229	232	234	235	232	181
237	230	236	234	233	234	151
255	255	255	251	230	236	181
99	90	97	37	94	247	130
222	152	255	129	129	246	132
154	199	255	150	189	241	147
216	132	162	163	170	239	122

The next example is the grayscale image. So, in this case, I am considering 256 numbers of intensity levels and if you see, it is the grayscale image.

(Refer Slide Time: 33:26)

## What is a Colour Image?

- Three components:  
R, G, B each usually represented by 8 bits  
We call 24-bit video
- These three primary are mixed in different proportions to get different colours
- For different processing applications other formats (YIQ, YCbCr, HIS etc) are used

A color image is just a three – component function. We can write this as a “vector-valued” function:

$$f(x, y) = \begin{bmatrix} r(x, y) \\ g(x, y) \\ b(x, y) \end{bmatrix}$$

And after this, I have considered the color image. So, this is the color image because this is called the vector pixels. In the vector pixels, I have three components; R component, G component, and the blue component.



Now, I have different types of color models. I will discuss about the color models like RGB color model, YCbCr color model, HIS color model. So, all these color models I will discuss in the color image processing. Because I need 8 bits for the R components, 8 bits for the G component, 8 bits for the B component, so it is called 24 bit video, that is the color image.

(Refer Slide Time: 34:02)

**True Color Image**

- RGB image (red-green-blue)
- 16,777,216 different possible colors

Red	Green	Blue
48 55 58 57 52 58	84 78 82 79 79 78	48 80 77 82 87 77
56 63 63 69 58 57	53 53 51 51 58 66	51 53 56 59 69 66
50 54 54 53 55 56	65 62 66 60 60 69	52 53 54 54 62 60
63 78 72 69 69 69	128 128 133 138 111 133	105 128 128 132 107 106
66 91 91 86 83 82	137 136 132 138 138 133	141 139 139 137 135 131
69 76 82 79 76 75	158 158 154 154 158 153	95 99 109 108 112 109
61 69 73 79 76 76	98 103 112 108 111 107	86 93 107 111 105 102



And this is the example of the true color image, here you see corresponding to a particular pixel, I have the red pixels, the green pixels, and the blue pixels. So that means, pixel means the red information, green information, the blue information corresponding to a particular pixel. The red value, the green value and the blue value.

(Refer Slide Time: 34:22)

**Indexed Image**

- The image has an associated color map, which is simply a list of all the colors used in that image
- GIF, PNG formats, etc.

Indices	Colour map
4 5 5 5 5 5	0.1211 0.1211 0.1416
5 4 5 5 6	0.1807 0.2549 0.1729
5 5 5 0 8	0.2197 0.3447 0.1807
5 5 5 5 11 11	0.1611 0.1768 0.1924
5 5 5 8 16 20	0.2432 0.2471 0.1924
8 11 11 26 33 20	0.2119 0.1963 0.2002
11 20 33 33 33 37	0.2627 0.2588 0.2549
	0.2197 0.2432 0.2588
	⋮ ⋮ ⋮

*Handwritten red text: R G B, Look up table*

Another definition of indexed image. So, for the indexed image, I have one color map. So, this is a color map. So, this is something like a lookup table, lookup table. So, corresponding to these value, we have three values; one is the R value, this is R value, like G value, and a B value; three values are available.

So, corresponding to this one, the corresponding to these RGB value, the index number is 6. Like this, I have all the index numbers. So, this is the indexed image, if you see. So, I need a color map and in the color map, you see the index number is associated with the color values, R value, G value, and the B values. So, corresponding to 6, I have this RGB values.

(Refer Slide Time: 35:14)



## Types of Digital Images

---

- Binary image – 1 bit/pixel
- Grayscale image – 8 bits/pixel
- True color or RGB image – 24 bits/pixel
- Indexed image – 8 bits/pixel

So, types of digital image. So binary image, 1 bit per pixels; grayscale image, 8 bit per pixels; true color or the RGB image, it is 24 bit per pixels; and for indexed image, 8 bit per pixels. So, this is the case.

(Refer Slide Time: 35:29)

## Image Processing

- ❖ Digital Image processing deals with manipulation and analysis of the digital image by a digital hardware, usually a computer.
- ❖ Emphasizing certain pictorial information for better clarity (human interpretation)
- ❖ Automatic machine processing of the scene data.
- ❖ Compressing the image data for efficient utilization of storage space and transmission bandwidth

So next one is what is image processing? The image processing is that manipulation and analysis of the digital image by digital hardware. And it is for the better clarity that is for the human interpretations and automatic machine processing of the scene data. So that is image processing.

(Refer Slide Time: 35:52)

## Image Processing Examples



So, I have given some examples here. So, one is image enhancement. Image enhancement means to improve the visual quality of an image. So first if you see, I am just the compare the two images, the input image and the output image. You can see that I am improving the visual quality of the image.

So, some of the examples like I can increase the contrast of an image, I can improve the brightness of an image. Like this, I can do many things, that is one example.

(Refer Slide Time: 36:27)

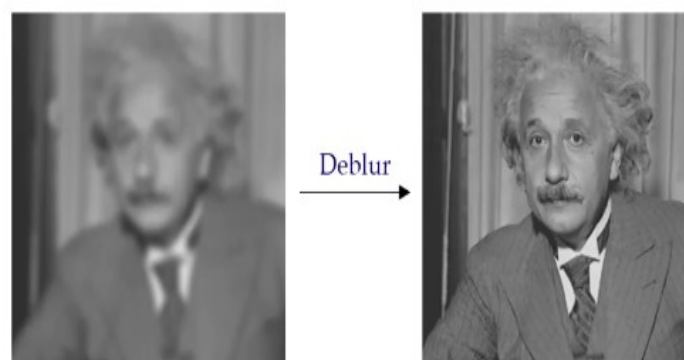
## Image Processing Examples



In this example, another example, I am doing the image enhancement. So first one is the input image, the second one is the better-quality image that is the enhanced image.

(Refer Slide Time: 36:39)

## Image Processing Examples

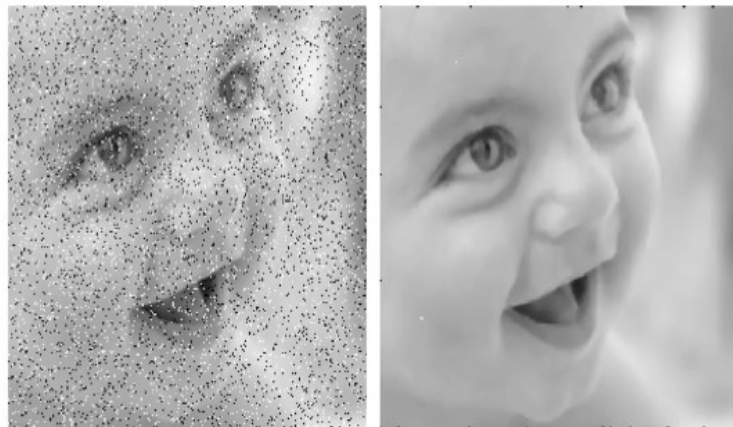


In this example, I am considering the blurred image. So blurred image maybe because of some motion blur. Suppose, the camera is moving, then in this case, I am getting the blurred image. And in this case, I can apply some image restoration technique to deblur the image.

So, in this example, I have the blurred image and I can apply some image restoration technique to get the deblur image.

(Refer Slide Time: 37:05)

## Image Processing Examples



In this example, I have considered a noisy image and after this, I am getting the filtered image, a better-quality image.

(Refer Slide Time: 37:14)

**Image Processing**

- An **image processing** operation typically defines a new image  $g$  in terms of an existing image  $f$ . We can transform either the range of  $f$

$$g(x, y) = t(f(x, y))$$

We can also transform the domain of  $f$ :

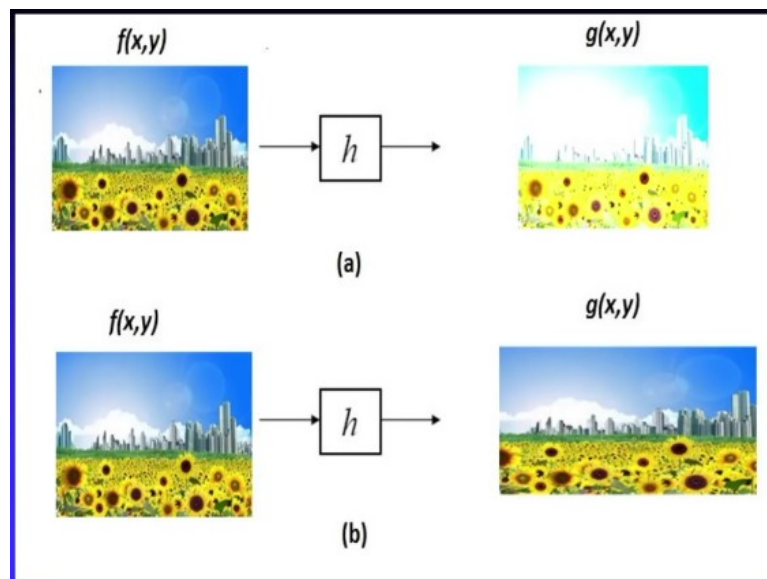
$$g(x, y) = f(t_x(x, y), t_y(x, y))$$

Now, the image processing, what is actually image processing now I can show. Because image is represented by  $f(x, y)$ , what is  $f(x, y)$ ? Intensity at the point, the point is  $x, y$ . Now, in this case, I have  $g(x, y)$ ;  $g(x, y)$  is the output image and  $f(x, y)$  is the input image. So, I can change the range of an image or I can change the domain of an image.

What do you mean by range of an image? The range of an image means, I can change the intensity value, the pixel value I can change by using some transformation. Also, I can change the domain of an image. Domain of an image means, I can change the spatial positions; x and y coordinate I can change.

One is just changing the range of an image and another one is changing the domain of an image. So, in this example, here you see, I am changing the range of the image. That means I am changing the intensity value. In the second case I am changing the spatial coordinates.

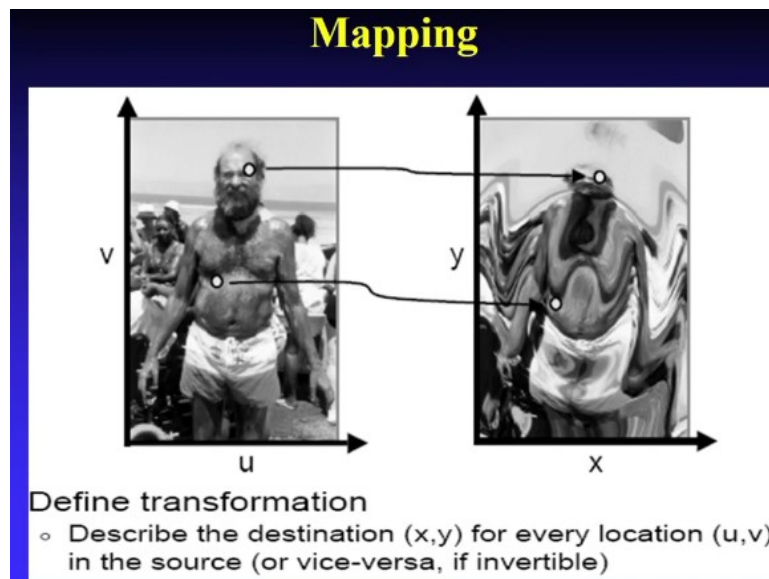
(Refer Slide Time: 38:14)



So, I have shown the two examples. The first one is the input image. This is the input image and I am changing the range of an image that is the pixel values.

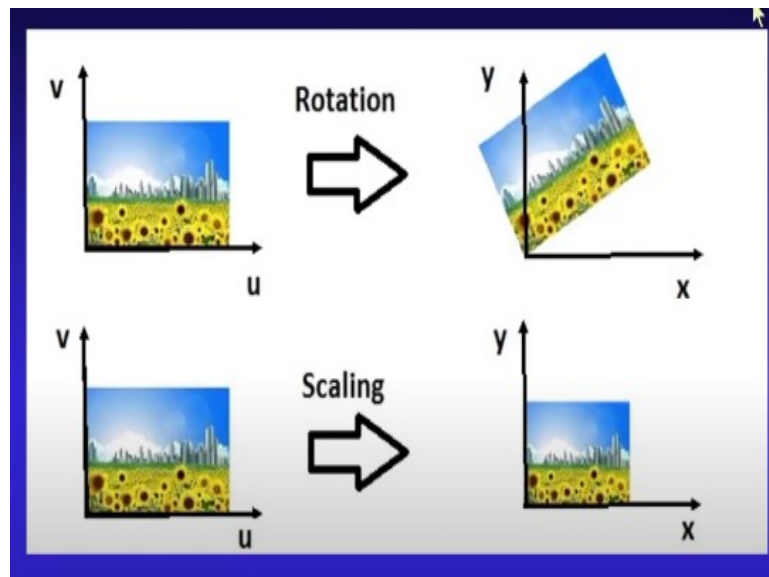
In a second example is the input image and I am changing the domain of the image. In this case I am doing the scaling of the image. So that means, this is the changing the domain of the image and these are changing of the range of an image.

(Refer Slide Time: 38:37)



This is one transformation, here if you see I am doing some mapping. This pixel is mapped to this point, this pixel is mapped to this point. So, this is my, this is my input image and this is my output image. So, I can do the mapping, that means, I am changing the spatial coordinates.

(Refer Slide Time: 38:55)



I can do rotation on an image, this is the rotation of an image. I can do the scaling of an image. That means, I can change the domain of an image, I can say the spatial position of the pixels.

(Refer Slide Time: 39:06)

## Linear Point Operation on Images

$$g(n) = h[f(n)]$$



The form of the function  $h$  is determined by the task at hand

$$g(n) = Pf(n) + K$$

**Image Negative**

Both scaling & Offset

$$g(n) = -f(n) + (L-1)$$
$$P = -1$$
$$K = L-1$$



Now, the linear point operation on an image. So, I can give one example here. Here you see  $g(n)$  is the output image,  $f(n)$  is the input image. And in this case, I am considering one operation, the operation is  $h$ , that is called a point operation, the linear point operation. The point operation is mainly the memoryless, we do not need any memory.

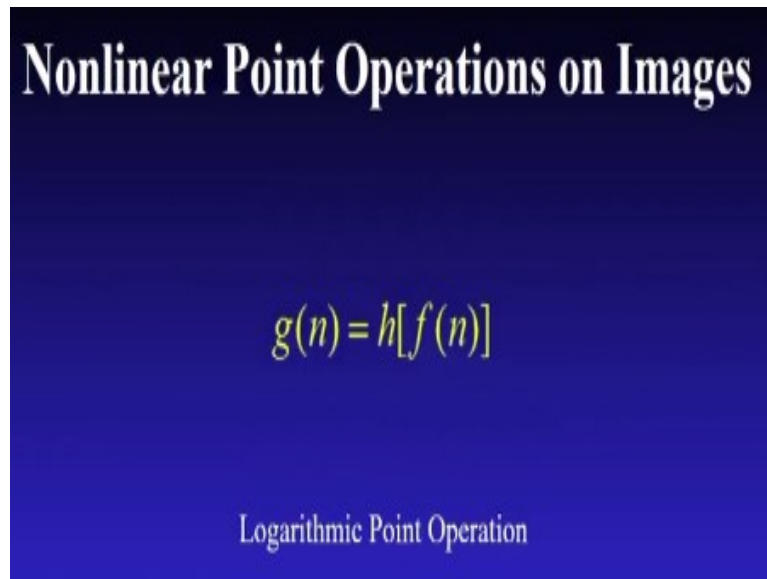
Now, in this case,  $g(n)$  is equal to I can consider  $g(n)$  is the output image,  $P \times f(n) + K$ . So, if I consider this expression,  $K$  is the offset and  $P$  is the scaling factor. Now, based on this the point operation, how to get that negative image?

So, in this case if I consider  $P$  is equal to minus one and  $K$  is equal to  $L$  minus 1,  $L$  means the number of gray level values, how many levels I am considering, from 0 to  $L$  minus 1. So, output is  $g(n) = -f(n) + L - 1$ .

So, if I consider this transformation, then in this case, I will be getting the negative image. So, this is my input image and then this is my output image.

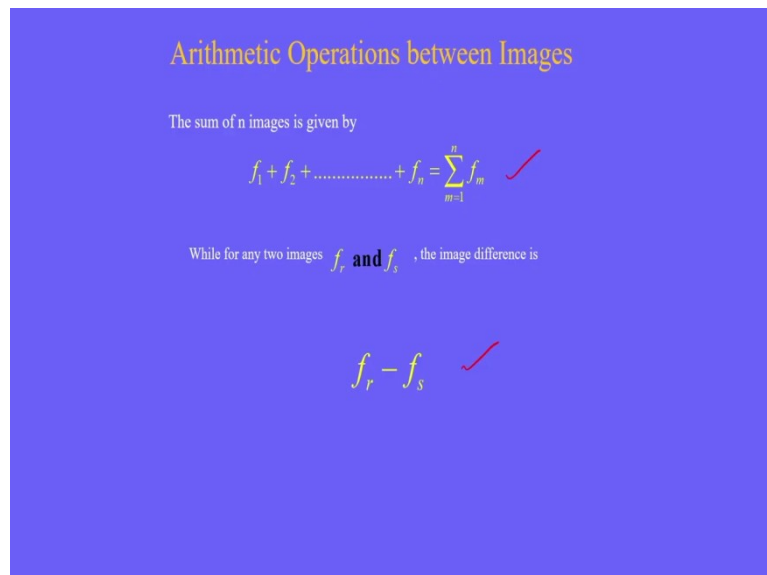


(Refer Slide Time: 40:11)



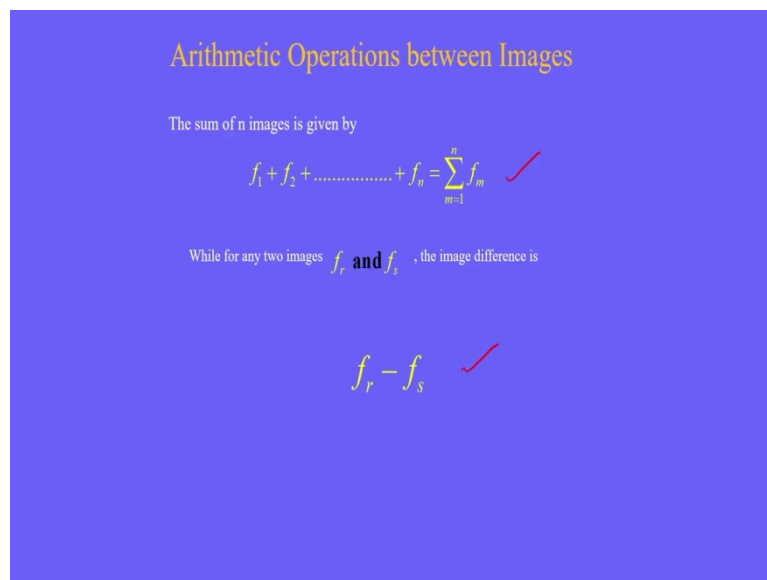
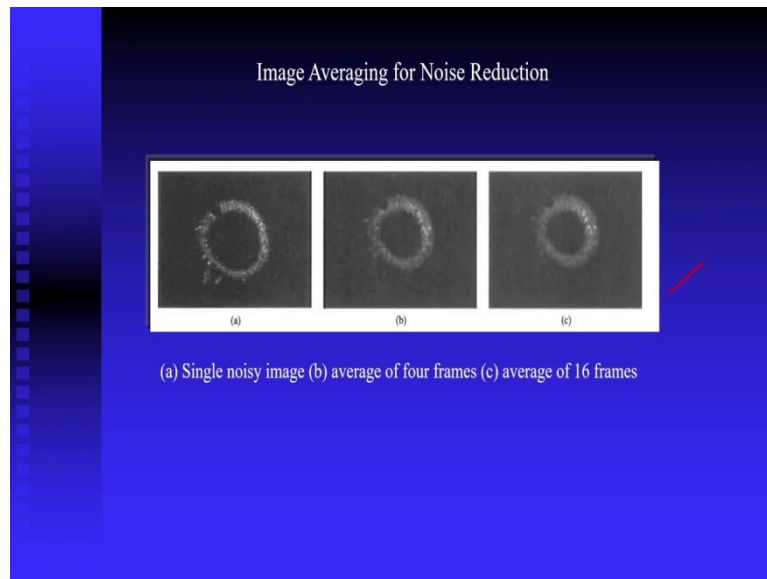
Also, I can apply the nonlinear point operation, maybe the logarithmic operation I can apply. So, this point operation I am going to discuss after 2, 3 lectures, that is, when I will be discussed about the image enhancement technique, that time I will be discussed about the point operation. So, this is about the nonlinear point operation and the linear point operations on image.

(Refer Slide Time: 40:36)



Here, I am giving one example, the addition of images. So, n number of images I am considering. After this, I am doing the addition, addition of n number of images. The second example is the subtraction of two images.

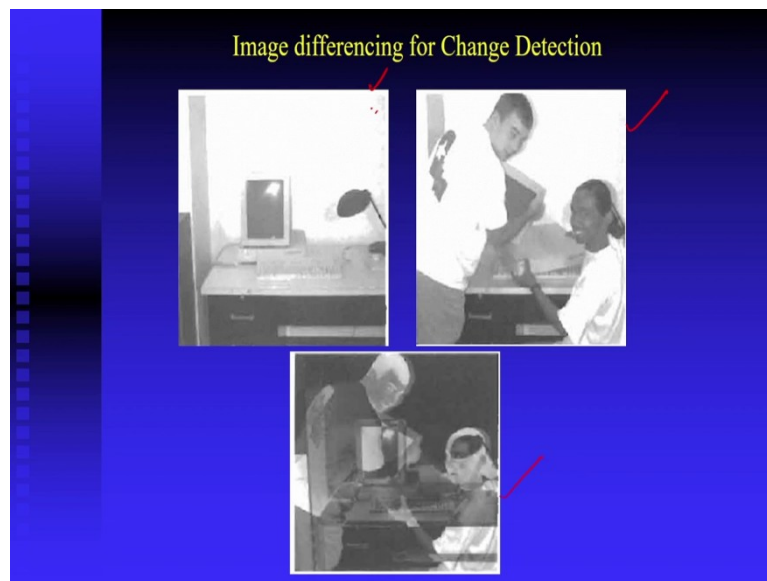
(Refer Slide Time: 40:51)



If I do addition of the images what will happen, you see I have the noisy image. Suppose I have number of noisy frames. If I do the averaging, then in this case, the noise can be reduced, image averaging for noise reduction.

So just I am adding this one that means, I am actually averaging the images, that is, I am doing to sum of n input images and corresponding to this I am getting this this example. So, in this example, the image averaging for noise reduction.

(Refer Slide Time: 41:22)



And already I have explained that how to find a difference between two images. So, one example is, suppose if I consider these two images, the first image is this. In this case, the background is available; in the second case the background and the foreground is available.

Suppose the moving objects are available in the image, in the first case, there is no moving objects. And in the second case, suppose, the background is also there, the moving object is there.

So, in this case, if I do the subtraction of these images from the first image to second image, then in this case, I will be getting the foreground, something like this, moving objects I can determine. That is called a change detection.

So, this is called a principle of change detections. Thus, I have to subtract the images then in this case I will be getting the moving objects in the image.

(Refer Slide Time: 42:11)

The slide is titled "Geometric Image Operations" and contains the following content:

- Equation:  $g(n) = f(n') = f[a(n)]$
- Text: **where, mapping function**  $a(n) = [a_1(n_1, n_2), a_2(n_1, n_2)]$
- Image Translation:  $a_1(n_1, n_2) = n_1 - b_1$ ,  $a_2(n_1, n_2) = n_2 - b_2$  (with a red checkmark)
- Image Zooming: The image zoom either magnifies or minifies the input image according to the mapping functions.  
 $a_1(n_1, n_2) = n_1 / c$ ,  $a_2(n_1, n_2) = n_2 / d$  (with red checkmarks and a handwritten box containing  $\begin{bmatrix} c_1 & -c_2 \\ -c_2 & c_1 \end{bmatrix}$ )
- Image Rotation:  $a_1(n_1, n_2) = n_1 \cos \theta - n_2 \sin \theta$ ,  $a_2(n_1, n_2) = n_1 \sin \theta + n_2 \cos \theta$  (with red checkmarks)

Like this, we can do some geometric image operations. So, I can do image translation like this, I can change the coordinate; coordinate is  $n_1$ , another one is  $n_2$ . So, I can do the translation of the image. This the coordinate is  $x$  and  $y$  actually, in this case, I am showing  $n_1$  and  $n_2$ . So  $n_1$  minus  $b_1$  and  $n_2$  minus  $b_2$ , that means I am doing the image translation.

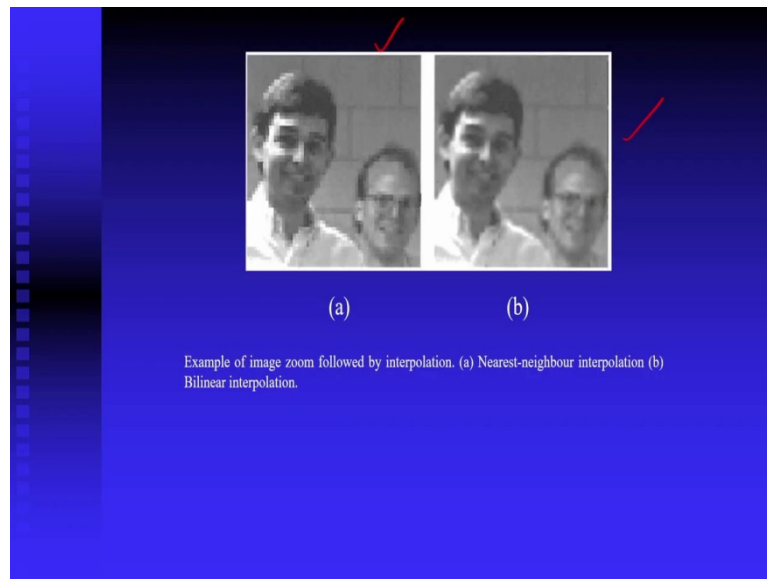
The second example is also if you see, the coordinate  $n_1$  is divided by  $c$ , the coordinate  $n_2$  is divided by  $d$ . So that means I can zoom the image, zoom in and zoom out I can do. I will explain this one zoom later on.

So, this is also a geometric image operation and this is one example of the image rotation that

I can rotate the image. So, for the rotation, the transformation matrix is  $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

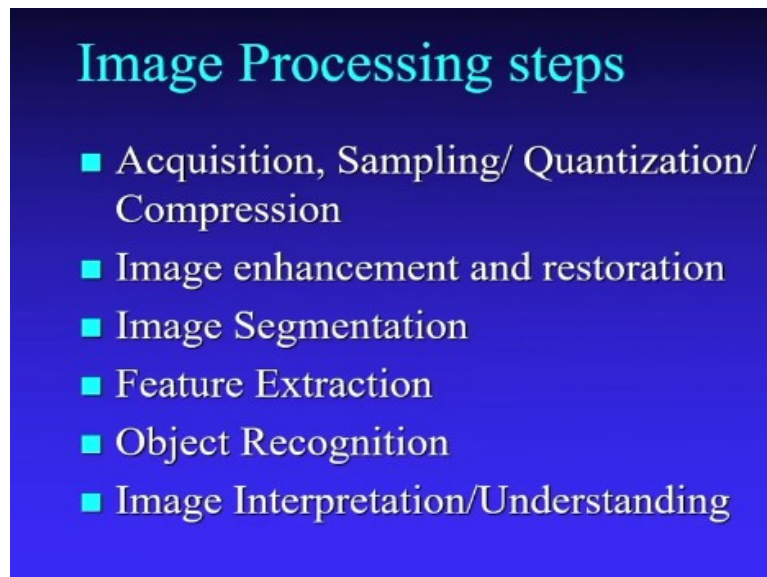
. So, by using this transformation matrix you can rotate an image.

(Refer Slide Time: 43:21)



Here I have, in this example I have shown the zoom. After zooming, I have done some interpolation, nearest neighbour interpolation it is done. The second case is the bilinear interpolation that is more accurate. So, I have shown two interpolation techniques, one is the nearest neighbour interpolation, another one is the bilinear interpolation after zooming.

(Refer Slide Time: 43:40)



So, image processing steps I have. One is the image acquisition. After this, I have to get the digital image. So, for this, I have to do the sampling and the quantization. Compression is important because if I considered the raw image or the raw video, that is very bulky. So that is why for storing the image, for storing the video, or for the transmission of the image or for

the transmission of the video, I have to do image compression, I have to do video compression.

Next one is the image enhancement. Image enhancement is to improve the visual quality of an image. So, one example is, I can improve the contrast of an image or maybe I can improve the brightness of an image. That is image enhancement.

And restoration is also the meaning is same. To improve the visual quality of an image. So, there is a basic difference between enhancement and the restoration. In image enhancement that is mainly the subjective. It depends on the subjective preference of an observer.

So, I can say, the image is not good, the contrast is not good, and based on my observation I can improve the visual quality of an image. That is subjective and there is no mathematical theory for image enhancement.

But in case of the image restoration, I have mathematical models. So, I can give one or two examples like this. Suppose I am taking some image and camera is moving, then in this case, I am getting the blurred image. So, in this case, I can deblur the image by using the restoration principle, the image restoration principle.

So, for this I have to develop a mathematical model and based on this mathematical model, I can improve the visual quality of the image. So, for this, I have to apply image restoration. That is image restoration is objective and image enhancement is subjective, the subjective preference of an observer.

I can give another example of the image restoration. Suppose the object is not properly focused then in this case, I am getting the blurred image. Then also I can improve the visual quality of the image. I can de-blur the image.

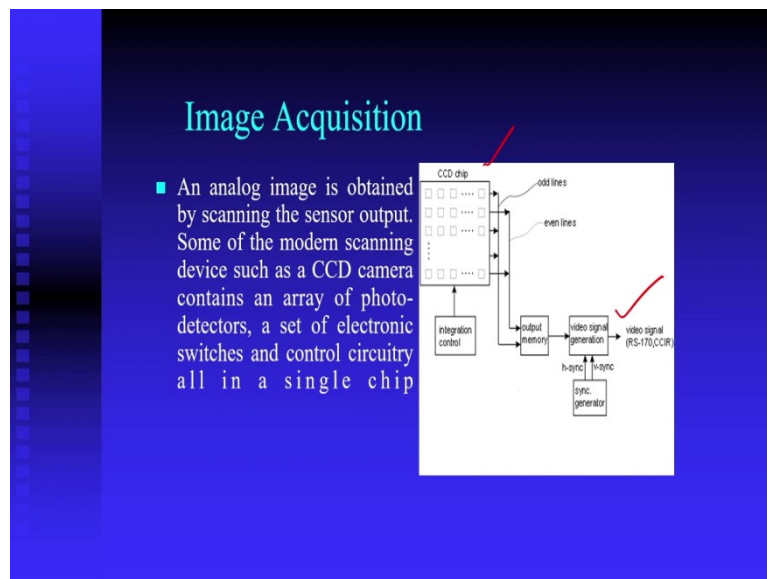
And suppose if I take one image in the foggy environment, then in this case, I am also getting not good quality image. Then by using this image enhancement principle or the restoration principle, I can improve the visual quality of an image. So, you see the fundamental difference between image enhancement and the restoration.

In image enhancement there is no mathematical model, general model but in image restoration, I can develop some mathematical model and by using this mathematical model, I can improve the visual quality of an image.

Next one is image segmentation. The image segmentation is the partitioning of an image into connected homogeneous region. So, I can later on, I will discuss about the image segmentation. So, that means I am doing the partitioning of an image into connected homogeneous region.

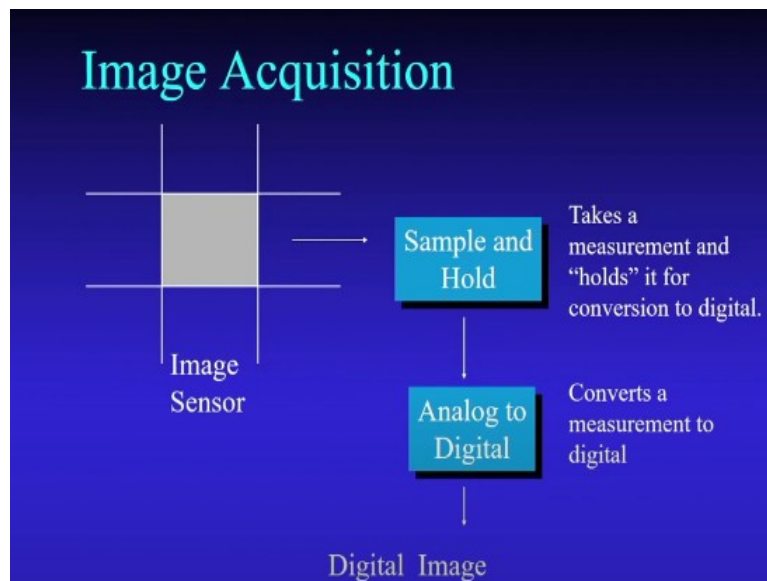
And after this, for computer vision, I have to extract features. And after this, based on the features, I have to go for recognition, the object recognition that is called the image interpretation or the understanding.

(Refer Slide Time: 46:57)



So, this is a typical image acquisition systems. So, we have the CCD, the search coupled device camera. Here you see the CCD chips. So the photon is converted into electrical signal. After this, I am getting the video signals, that is, the analog signals. The analog signal can be converted into digital, already I have explained about this.

(Refer Slide Time: 47:16)



So, this is the image sensor. The light photon is converted into electrical signal and after this, we have to do the sampling and the whole. And after this, you can convert the analog signal into digital image.

(Refer Slide Time: 47:30)

### Sampling/ Quantization/ Compression

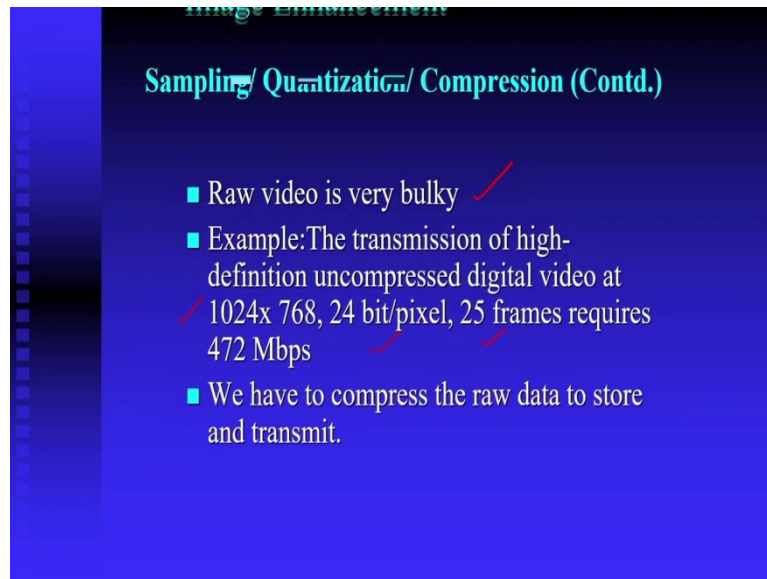
- A digital image is obtained by sampling and quantizing an analog image.
- The analog image signal is sampled at rate determined by the application concerned
- Still image 512X512, 256X256 ✓ ✓
- Video: 720X480, 360X240, 1024x 768 (HDTV)
- The intensity is quantized into a fixed number of levels determined human perceptual limitation
- 8 bits is sufficient for all but the best applications  
10 bits – Television production, printing
- 12-16 bits – Medical imagery ✓

So, these are the, for the still image, these are the 512 × 512 some typical values. 256 × 256, one typical values. For video like this 720 × 480, 1024 × 768 that is the high definition TV, like this. And regarding the intensity resolution, 8 bits is sufficient for all but the best applications needs 10 bits, television, production, printing.



And for medical imaging generally, the 12 to 16 bits images we consider. So, this is the case and one important application is the medical imaging. So, we need more number of bits we need more number of intensity levels.

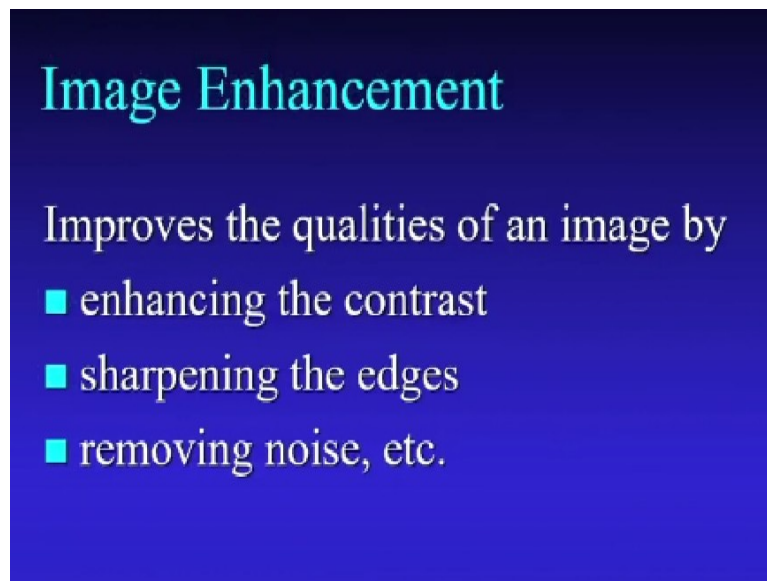
(Refer Slide Time: 48:16)



And if you see why we need the compression. Suppose a raw video is very bulky. So suppose in this example, if I considered a digital video, uncompressed video,  $1024 \times 768$  that is the size of the image frame.

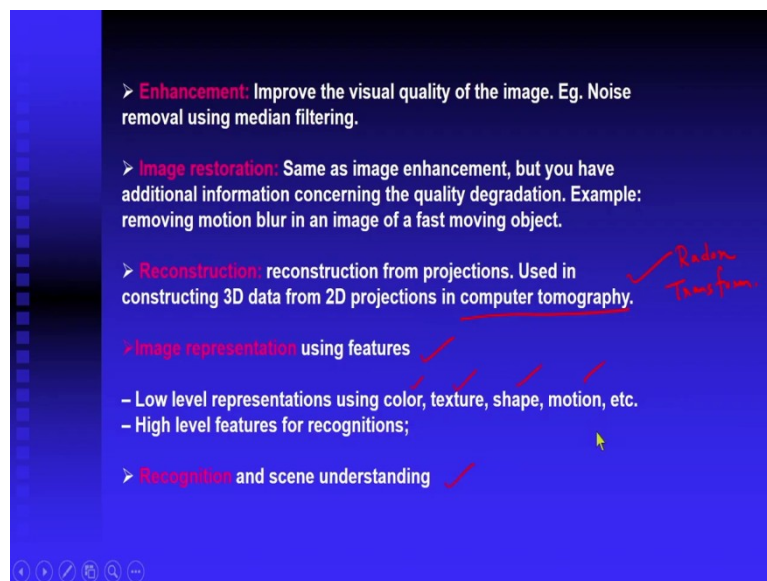
Suppose if I consider 24 bit per pixel, why it is 24 per pixel? That is the, if I consider the color, color image that is the 24 bit per pixels and 25 frames per second. That means 472 Mbps. So that is very difficult to store and also for transmission. So, that is why we have to compress the image, the compression of the image is important in this case. So, I have given this example.

(Refer Slide Time: 49:00)



And image enhancement, this already I have explained that to improve the visual quality of an image, like this we can improve the contrast, we can sharpen the edges, we can remove the noise, the image enhancement.

(Refer Slide Time: 49:16)



This is already I have explained. One is enhancement; restoration, image restoration, I have already explained. And reconstruction, mainly we have the 2D images. So, if I want to get the 3D information, the reconstruction of the 3D information from the 2D projections, that is called a reconstruction.

This concept I am going to explain maybe after 2, 3 lectures that is image reconstruction from 2D projections. That concept I will explain the radon transform. Radon transform I will explain later on. So, construction of 3D information from the 2D projections that is used in CT-scan, the computer tomography. So, this concept I am going to explain when I will discuss the radon transform.

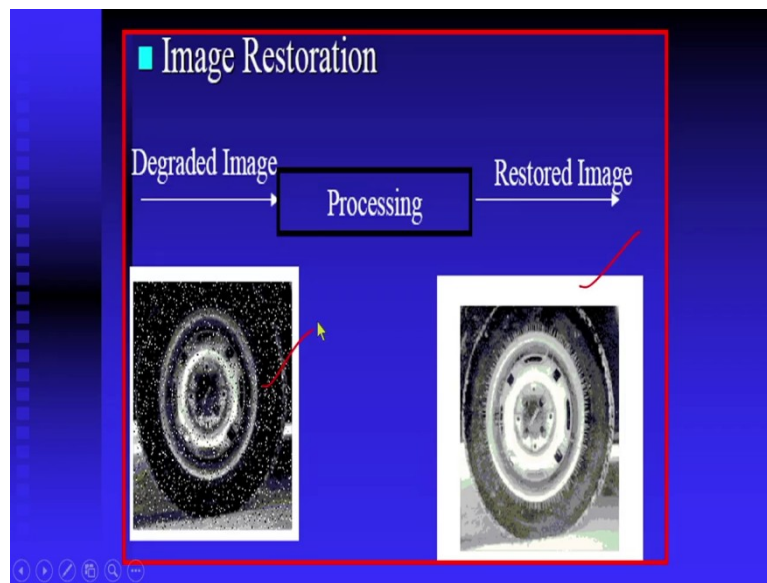
After this, we have to extract the features. Image features I have to extract, maybe something like the color feature, I can consider the texture, I can consider shape features, motion features I can consider. And finally, we will go for the recognition or the classification of the objects, that is the machine learning. This is the machine learning, feature extraction and recognitions.

(Refer Slide Time: 50:30)



So, these are some examples, the image filtering. Original image, the noisy image, and the filtered image. So, this is one example.

(Refer Slide Time: 50:36)



Second example is the image restoration. Degraded image after processing I am getting the restored image. That is a very good quality image, this is the noisy image.

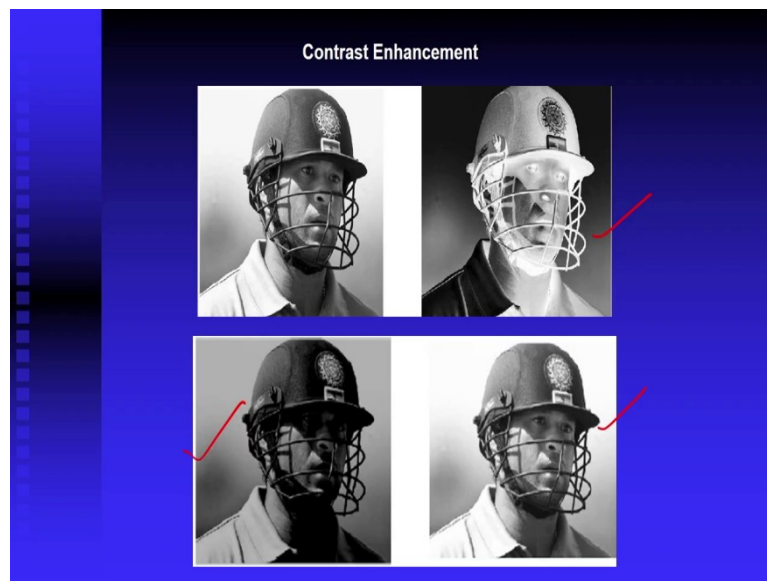
(Refer Slide Time: 50:47)

The slide has a blue background and is titled "Histogram Equalization" in a light blue font. It contains three bullet points, each preceded by a small blue square:

- Enhance the contrast of images by transforming the values in an intensity image to its normalized histogram
- The histogram of the output image is uniformly distributed.
- Contrast is better

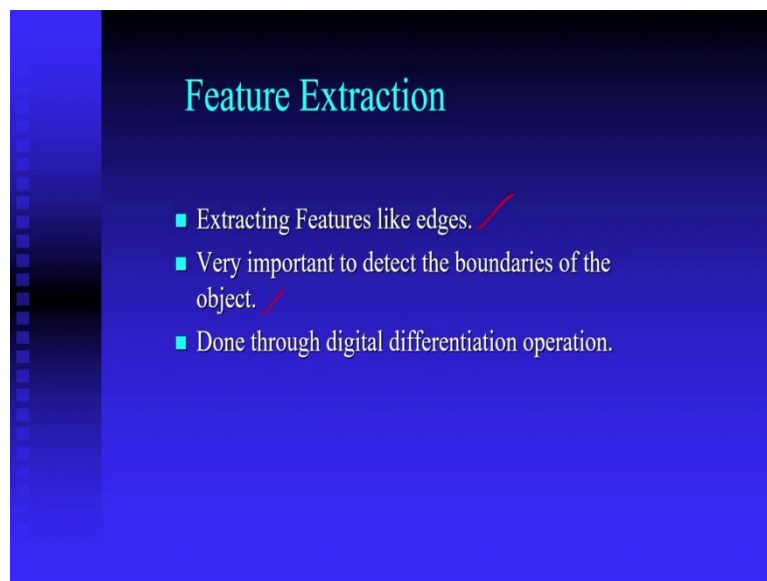
And for contrast improvement, there is a technique. This technique also I am going to explain later on. That is by histogram equalization technique, I can improve the visual quality of an image. That means I can improve the contrast of an image. So, this concept I am going to explain when I will discuss the image enhancement principles.

(Refer Slide Time: 51:08)



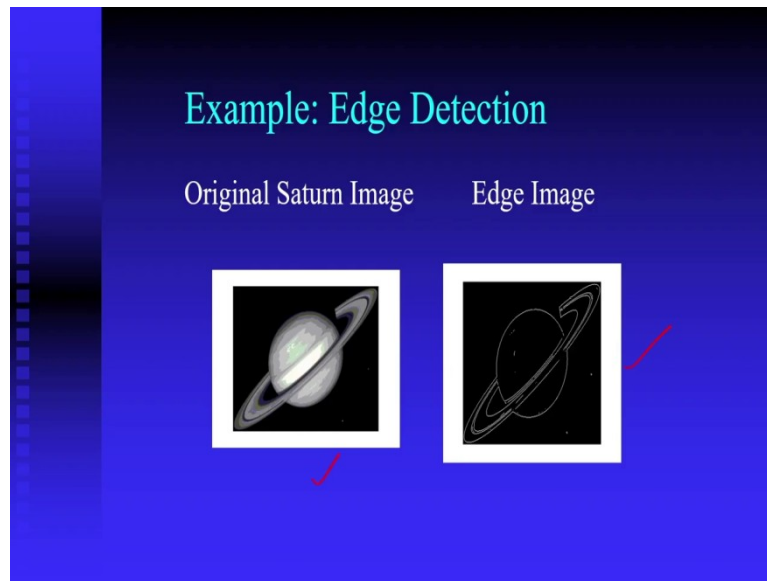
This is one example of the contrast enhancement. This is not a good contrast image, this is not a good contrast image but I think this is also good contrast image I can say. So, by using the histogram equalization technique, I can improve the contrast of an image.

(Refer Slide Time: 51:23)



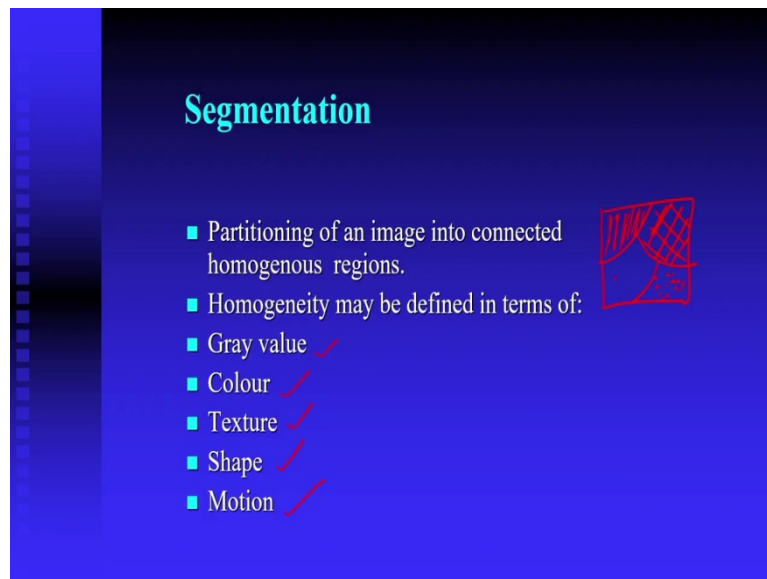
And after this, I have to extract features. The features may be something like edges I can consider, the boundary of an objects I can consider. So, I can extract many, many features from the image or maybe from the video.

(Refer Slide Time: 51:38)



And after feature extraction this one example is the edge detection technique. So, this is my input image and I am extracting the edges of the image.

(Refer Slide Time: 51:45)



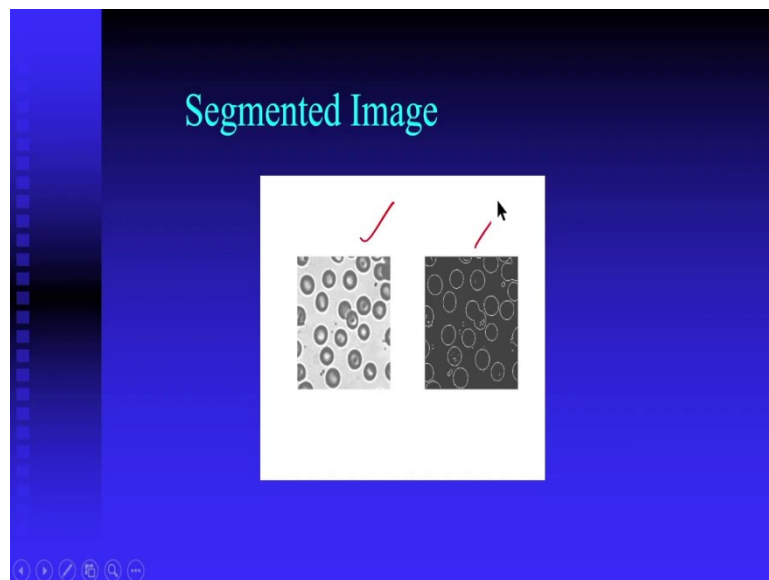
And the segmentation is the partitioning of an image into connected homogeneous region. This homogeneity may be defined in terms of the gray value, color value, texture value, shape information, and the motion information. So, based on this principle, based on these values, I can define the homogeneity.

So again, I am going to discuss about the segmentation principle. So, suppose in my next classes, so suppose if I want to do the partitioning of an image, so this is one partition, this is one partition, this is one partition.

So, this portion is homogeneous, this portion is also homogeneous. If I consider this, portion is homogeneous, this portion is homogeneous. Like this, this portion is also homogeneous. So, homogeneity I can define in terms of gray value, color, texture, like this, I can define.

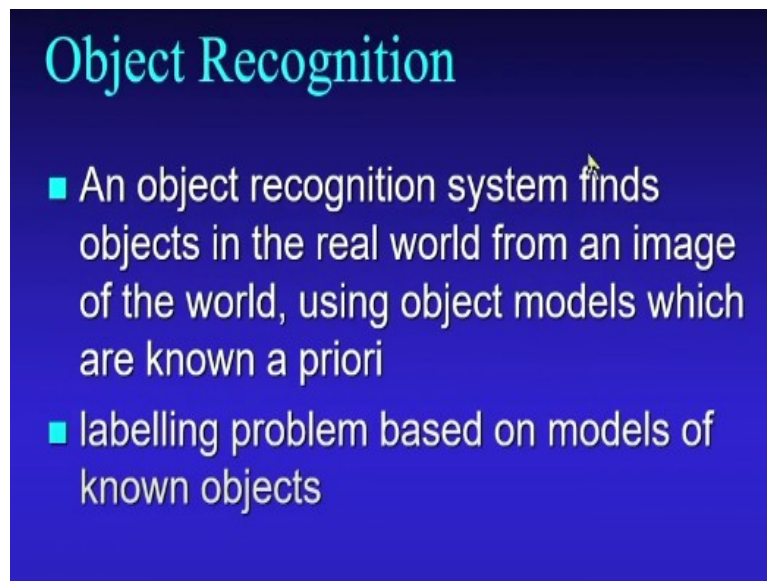
So, corresponding to this region, the gray value is almost same or maybe the color may be same. So, like this I can do the partitioning of an image.

(Refer Slide Time: 52:47)



So, this is one example of the segmentation, image segmentation. This is my image, input image and after this, I am doing the segmentation.

(Refer Slide Time: 52:54)

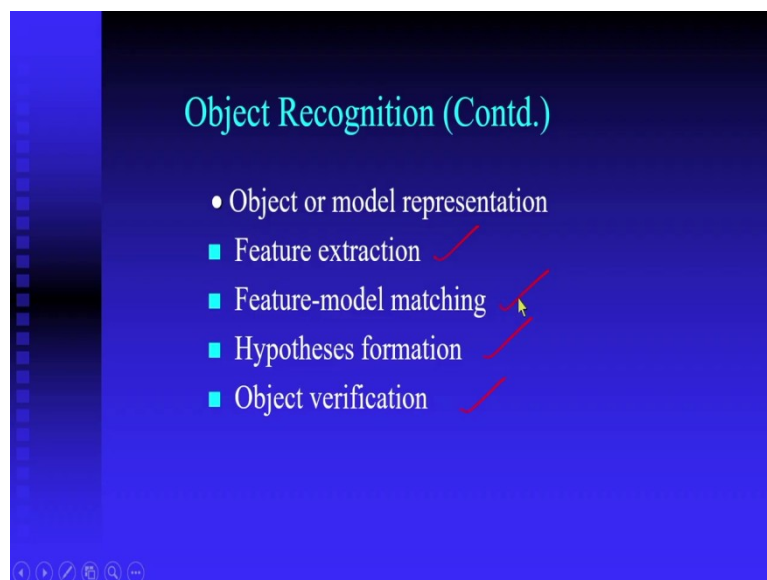


## Object Recognition

- An object recognition system finds objects in the real world from an image of the world, using object models which are known a priori
- labelling problem based on models of known objects

And finally, after feature extraction, I will go for the object recognition. So, this is my last step that is I have to extract the features. The features maybe the color is a features, texture maybe the features. So, I can extract different types of features from the image and after this, I will go for the object recognition.

(Refer Slide Time: 53:16)



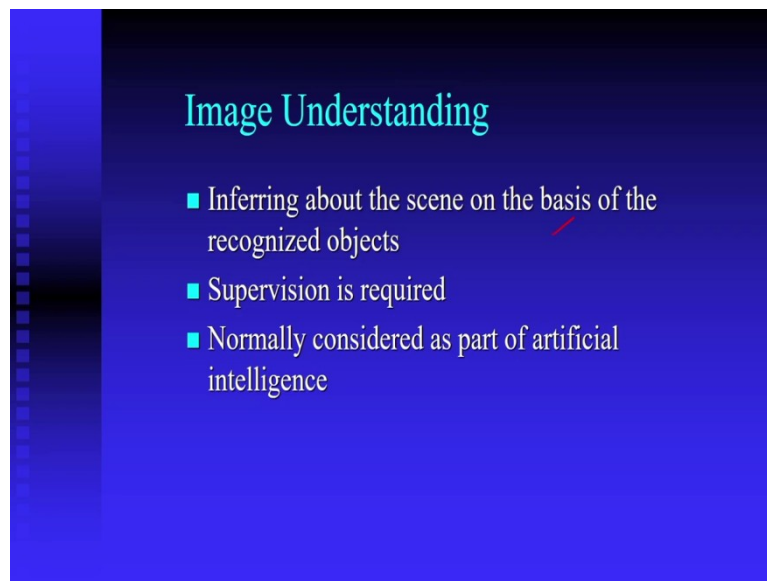
## Object Recognition (Contd.)

- Object or model representation
- Feature extraction ✓
- Feature-model matching ✓
- Hypotheses formation ✓
- Object verification ✓

So first one is the feature extraction, feature model matching, hypothesis formation, and object verification, I have to do this.



(Refer Slide Time: 53:27)



And finally, I will go for the image understanding. So maybe I can apply some supervised technique and that is the part of the artificial intelligence or maybe the machine learning. So, I am going to discuss about this in my last lecture the machine learning. So, this is the final step the image understanding.

So, in this class, I have discussed about the concept of image processing, what is the meaning of the digital image. And after this, I have discussed about the types of digital image, the RGB image, the grayscale image, the binary image, index image. So, I have discussed about this.

And also, the concept of the resolution, resolution of an image. So first, I discussed about the spatial resolution, after this, I discussed the intensity resolution. After this, I discussed about the image processing techniques. So, I can change the domain of an image, I can change the range of an image. So, I have given some examples I can do the rotation of an image, I can do the scaling of an image. Like this I can do many operations, I can do zooming also.

After this, I have considered the typical image processing system, the computer vision systems. So first, I have to do some image processing and in this case, before image processing, I have to get the image. The analog image can be converted into digital image by sampling and the quantization.

After this, I can improve the visual quality of the image by using the enhancement and the restoration techniques. And after this, I can extract the features from the image. And finally, I

will go for object recognition. That is the pattern classification or the machine learning techniques I can apply.

So, this is about my, the second lecture. So, next class, I am going to explain the concept of the image formation principle and also, I will discuss about the radiometry. Radiometry means the measurement of light. So, let me stop here today. Thank you.