learning@cse.iitk

Kamlesh Tiwari

1 Introduction to Computer Graphics

Computer Graphics is a powerful tool for the rapid and economical production of pictures. Few of the major application areas of CG are as below.

- Computer aided design.
- Presentation graphics
- Computer art
- Entertainment
- Education and training
- Visualization
- Image processing
- Graphical user interface

2 Display Devices

The primary output device in a graphics system are video monitor. The operation of most video monitors is based on the standard cathode-ray tube (CRT)design, but several other technologies exist and solid-state monitors may eventually predominate.

2.1 CRT

A beam of electrons (cathode rays), emitted by an electron gun,passes through focusing and deflection systems that direct the beam toward specified positions on the phosphor coated screen. The phosphor then emits a small spot of light at each position contacted by the electron beam. Because the light emitted by the phosphor fades very rapidly, some method is needed for maintaining the screen picture. One way to keep the phosphor glowing is to redraw the picture repeatedly by quickly directing the electron beam back over the same points. This type of display is called a refresh CRT.

2.1.1 Persistence

For the phosphors coated on the screen, persistence represents the duration they continue to emit light after the CRT beam is removed. Persistence is defined as the time it takes the emitted light from the screen to decay to one-tenth of its original intensity. Lower-persistence phosphors require higher refresh rates to maintain a picture on the screen without flicker. A phosphor with low persistence is useful for animation; a high-persistence phosphor is useful for displaying highly complex, static pictures. Although some phosphors have a persistence greater than 1 second, graphics monitors are usually constructed with a persistence in the range from 10 to 60 microseconds.

2.1.2 Resolution

The maximum number of points that can be displayed without overlap on a CRT is referred to as the resolution. A more precise definition of resolution is the number of points per centimeter that can be plotted horizontally and vertically, although it is often simply stated as the total number of points in each direction. Two adjacent spot will appear distinct as long as their separation is greater than the diameter at which each spot has an intensity of about 60 percent of that at the center of the spot. Resolution of a CRT is dependent on the type of phosphor, the intensity to be displayed, and the focusing and deflection systems. Typical resolution on high-quality systems is 1280 by 1024. High- resolution systems are often referred to as high-definition systems. The physical size of a graphics monitor is given as the length of the screen diagonal, with sizes varying from about 12 inches to 27 inches or more.

2.1.3 Aspect ratio

This number gives the ratio of vertical points to horizontal points necessary to produce equal-length lines in both directions on the screen. An aspect ratio of 3 / 4 means that a vertical line plotted with three points has the same length as a horizontal line plotted with four points.

2.1.4 Raster scan

In this system, the electron beam is swept across the screen, one row at a time from top to bottom. As the electron beam moves across each row, the beam intensity is turned on and off to create a pattern of illuminated spots. Picture definition is stored in a memory area called the **refresh buffer** or **frame buffer**. Intensity range for pixel positions depends on the capability of the raster system. Refreshing on raster-scan displays is carried out at the rate of 60 to 80 frames per second, although some systems are designed for higher refresh rates. At the end of each scan line, the electron beam returns to the left side of the screen to begin displaying the next scan line. The return to the left of the screen, after refreshing each scan line, is called the **horizontal retrace** of the electron beam. And at the end of each frame, the electron beam returns (**vertical retrace**) to the top left comer of the screen to begin the next frame.

2.1.5 Random scan

In this type of display system the CRT has the electron beam directed only to the parts of the screen where a picture is to be drawn. Random scan monitors draw a picture one line at a time and for this reason are also referred to as **vector displays** (or **stroke-writing** or **calligraphic displays**). The component lines of a picture can be drawn and refreshed by a random scan system in any specified order. A pen plotter operates in a similar way andis an example of a random-scan, hard-copy device.

Refresh rate on a random-scan system depends on the number of lines to be displayed. Picture definition is now stored as a set of linedrawing commands in an area of memory refered to as the **refresh display file**. Sometimes the refresh display file is called the display list, display program, or simply the refresh buffer. To display a specified picture, the system cycles through the set of commands in the display file, drawing each component line in turn. After all line drawing commands have been processed, the system cycles back to the first line command in the list. Random-scan displays are designed to draw all the component lines of a picture 30 to 60 times each second.

Random-scan systems are designed for linedrawing applications and can not display realistic shaded scenes. Since pidure definition is stored as a set of linedrawing instructions and not as a set of intensity values for all screen points, vector displays generally have higher resolution than raster systems. Also, vector displays produce smooth line drawings because the CRT beam directly follows the line path. A raster system, in contrast, produces jagged lines that are plotted as discrete point sets.

2.1.6 interlacing

On some raster-scan systems (and in TV sets), each frame is displayed in two passes using an interlaced refresh pmedure. In the first pass, the beam sweeps across every other scan line fmm top to bottom. Then after the vertical retrace, the beam sweeps out the remaining scan lines.

2.2 Color monitors

A CRT monitor displays color pictures by using a combination of phosphors that emit different-colored light. By combining the emitted light from the different phosphors, a range of colors can be generated. The two basic techniques for producing color displays with a CRT are the beam-penetration method and the shadow-mask method.

2.2.1 Beam-penetration

The beam-penetration method for displaying color pictures has been used with random-scan monitors. Two layers of phosphor, usually red and green, are coated onto the inside of the CRT screen, and the displayed color depends on how far the electron beam penetrates into the phosphor layers. A beam of slow electrons excites only the outer red layer. A beam of very fast electrons penetrates through the red layer and excites the inner green layer. At intermediate beam speeds, combinations of red and green light are emitted to show two additional colors, orange and yellow. The speed of the electrons, and hence the screen color at any point, is controlled by the beam-acceleration voltage. Beam penetration has been an inexpensive way to produce color in random-scan monitors, but only four colors are possible, and the quality of pictures is not as good as with other methods.

2.2.2 Shadow-masking

Shadow-mask methods are commonly used in rasterscan systems because they produce a much wider range of colors than the beam-penetration method. A shadow-mask CRT has three phosphor color dots at each pixel position. One phosphor dot emits a red light, another emits a green light, and the third emits a blue light. This type of CRT has three electron guns, one for each color dot, and a shadow-mask grid just behind the phosphor-coated screen. The three electron beams are deflected and focused as a group onto the shadow mask, which contains a series of holes aligned with the phosphor-dot patterns. When the three beams pass through a hole in the shadow mask, they activate a dot triangle, which appears as a small color spot on the screen. The phosphor dots in the triangles are arranged so that each electron beam can activate only its corresponding color dot when it passes through the shadow mask.

We obtain color variations in a shadow-mask CRT by varying the intensity levels of the three electron beams. A sophisticated systems can set intermediate intensity levels for the electron beams, allowing several million different colors to be generated.

2.2.3 Flat-Panel Display

The term Flat-panel display refers to a class of video devices that have reduced volume, weight, and power requirements compared to a CRT. A significant feature of flat-panel displays is that they are thinner than CRTs, and we can hang them on walls or wear them on our wrists. We can separate flat-panel displays into two categories: *emissive displays* and *nonemissive displays.*

The emissive displays are devices that convert electrical energy into light. Plasma panels, thin-film electroluminescent displays, and Light-emitting diodes are examples of emissive displays. Flat CRTs have also been devised, in which electron beams arts accelerated parallel to the screen, then deflected 90' to the screen. Nonemmissive dis-

$\overset{o}{\kappa}$tiwari [at] cse.iitk.ac.in

plays use optical effects to convert sunlight or light from some other source into graphics patterns. The most important example of a nonemisswe flat-panel display is a liquid-crystal device.

3 Concept of Visual Information

The ability to see is one of the truly remarkable characteristics of living beings. It enables them to perceive and assimilate in a short span of time an incredible amount of knowledge about the world around them. The scope and variety of that which can pass through the eye and be interpreted by the brain is nothing short of astounding.

It is thus with some degree of trepidation that we introduce the concept of visual information, because in the broadest sense, the overall significance of the term is overwhelming. Instead of taking into account all of the ramifications of visual information; the first restriction we shall impose is that of finite image size, In other words, the viewer receives his or her visual information as if looking through a rectangular window of finite dimensions. This assumption is usually necessary in dealing with real world systems such as cameras, microscopes and telescopes for example; they all have finite fields of view and can handle only finite amounts of information.

The second assumption we make is that the viewer is incapable of depth perception on his own. That is, in the scene being viewed he cannot tell how far away objects are by the normal use of binocular vision or by changing the focus of his eyes. This scenario may seem a bit dismal. But in reality, this model describes an over whelming proportion of systems that handle visual information, including television, photographs, x-rays etc.

In this setup, the visual information is determined completely by the wavelengths and amplitudes of light that passes through each point of the window and reach the viewers eye. If the world outside were to be removed and a projector installed that reproduced exactly the light distribution on the window, the viewer inside would not be able to tell the difference.

Thus, the problem of numerically representing visual information is reduced to that of representing the distribution of light energy and wavelengths on the finite area of the window. We assume that the image perceived is "monochromatic" and static. It is determined completely by the perceived light energy (weighted sum of energy at perceivable wavelengths) passing through each point on the window and reaching the viewer's eye. If we impose Cartesian coordinates on the window, we can represent perceived light energy or "intensity" at point (x, y) by $f(x, y)$. Thus $f(x, y)$ represents the monochromatic visual information or "image" at the instant of time under consideration. As images that occur in real life situations cannot be exactly specified with a finite amount of numerical data, an approximation of $f(x, y)$ must be made if it is to be dealt with by practical systems. Since number bases can be changed without loss of information, we may assume $f(x, y)$ to be represented by binary digital data. In this form the data is most suitable for several applications such as transmission via digital communications facilities, storage within digital memory media or processing by computer.

3.1 Digital Image Definitions

A digital image $f[m, n]$ described in a 2D discrete space is derived from an analog image $f(x, y)$ in a 2D continuous space through a sampling process that is frequently referred to as digitization. The mathematics of that sampling process will be described in subsequent Chapters. For now we will look at some basic definitions associated with the digital image. The effect of digitization is shown in figure **??**.

The 2D continuous image $f(x, y)$ is divided into N rows and M columns. The intersection of a row and a column is termed a pixel. The value assigned to the integer coordinates $f[m, n]$ with $(m = 0, 1, 2, ..., M - 1)$ and $(n = 0, 1, 2, ..., N - 1)$ is $f[m, n]$. In fact, in most cases $f(x, y)$, which we might consider to be the physical signal that impinges on the face of a 2D sensor, is actually a function of many variables including depth (z),color (λ) and time (t). Unless otherwise stated, we will consider the case of 2D, monochromatic, static images in this module.

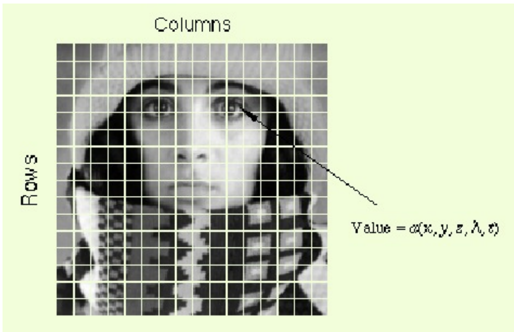


Figure 1: Digitization of a continuous image.

The pixel at coordinates $[m = 10, n = 3]$ has the integer brightness value 110.

The image shown in figure **??** has been divided into $N = 16$ rows and $M = 16$ The value assigned to every pixel is the average brightness in the pixel rounded to the nearest integer value. The process of representing the amplitude of the 2D signal at a given coordinate as an integer value with L different gray levels is usually referred to as amplitude quantization or simple quantization.

3.2 Common values

There are standard values for the various parameters encountered in digital image processing. These values can be caused by video standards, by algorithmic requirements, or by the desire to keep digital circuitry simple. Table **??** gives some comm

Quite frequently we see cases of $M = N = 2^k$ where $(k =$

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

Table 1: Common values of digital image parameters

8, 9, 10) .This can be motivated by digital circuitry or by the use of certain algorithms such as the (fast) Fourier transform.

The number of distinct gray levels is usually a power of 2, that is $L = 2^B$, where B is the number of bits in the binary representation of the brightness levels. When $B > 1$ we speak of a gray-level image; when $B = 1$ we speak of a binary image. In a binary image there are just two gray levels which can be referred to, for example, as "black" and "white" or "0" and "1".

Suppose that a continuous image $f(x, y)$ is approximated by equally spaced samples arranged in the form of an $N \times N$ array as:

$$\mathbf{f}(\mathbf{x},\mathbf{y}) = \begin{bmatrix} f(0,0) & \dots & f(0,N-1) \\ \vdots & & \vdots \\ f(N-1,0) & \dots & f(N-1,N-1) \end{bmatrix}_{N\times N}$$

Each element of the array refered to as "pixel" is a discrete quantity. The array represents a digital image.

The above digitization requires a decision to be made on a value for N a well as on the number of discrete gray levels allowed for each pixel.

It is common practice in digital image processing to let $N = 2^n$ and G = number of gray levels = 2^M. It is assumed that discrete levels are equally spaced between 0 to L in the gray scale.

Therefore the number of bits required to store a digitized image of size $N \times N$ is $b = N \times N \times m$ In other words a 128×128 image with 256 gray levels (ie 8 bits/pixel) required a storage of ≈ 17000 bytes.

The representation given above is an approximation to a continuous image.

Reasonable question to ask at this point is how many samples and gray levels are required for a good approximation? This brings up the question of resolution. The resolution (ie the degree of discernble detail) of an image is strangely dependent on both N and m. The more these parameters are increased, the closer the digitized array will approximate the original image.

Unfortunately this leads to large storage and consequently processing requirements increase rapidly as a function of N and large m.

3.3 Spatial and Gray level resolution

Sampling is the principal factor determining the spatial resolution of an image. Basically spatial resolution is the smallest discernible detail in an image.

As an example suppose we construct a chart with vertical lines of width W, and with space between the lines also having width W. A line-pair consists of one such line and its adjacent space. Thus width of line pair is and there are line-pairs per unit distance. A widely used definition of resolution is simply the smallest number of discernible line pairs per unit distance; for es 100 line pairs/mm.

Gray level resolution: This refers to the smallest discernible change in gray level. The measurement of discernible changes in gray level is a highly subjective process.

We have considerable discretion regarding the number of Samples used to generate a digital image. But this is not true for the number of gray levels. Due to hardware constraints, the number of gray levels is usually an integer power of two. The most common value is 8 bits. It can vary depending on application. When an actual measure of physical resolution relating pixels and level of detail they resolve in the original scene are not necessary, it is not uncommon to refer to an L-level digital image of size as having a spatial resolution of pixels and a gray level resolution of L levels.

3.4 Characteristics of Image Operations

There is a variety of ways to classify and characterize image operations. The reason for doing so is to understand what type of results we might expect to achieve with a given type of operation or what might be the computational burden associated with a given operation.

3.4.1 Type of operations

The types of operations that can be applied to digital images to transform an input image a[m, n] into an output image b[m, n] (or another representation) can be classified into three categories as shown in Table **??**

3.4.2 Types of neighborhoods

Neighborhood operations play a key role in modern digital image processing. It is therefore important to understand how images can be sampled and how that relates to the various neighborhoods that can be used to process an image.

Rectangular sampling - In most cases, images are sampled by laying a rectangular grid over an image as illustrated in Figure(1.1). This results in the type of sampling shown in Figure(1.3ab). Hexagonal sampling-An alternative sampling scheme is shown in Figure (1.3c) and is termed hexagonal sampling.

Both sampling schemes have been studied extensively and both represent a possible periodic tiling of the continuous image space. However rectangular sampling due to hardware and software and software considerations remains

Operation	Characterization	Generic Complexity/ Pixel
*Point	-the output value at a specific coordinate is dependent only on the input value at that same coordinate.	constant
*Local	-the output value at a specific coordinate is dependent on the input values in the neighborhood of that same coordinate.	P^2
*Global	–the output value at a specific coordinate is dependent on all the values in the input image..	N^2

Table 2: Types of image operations. Image size= $N \times N$ neighborhood size= $P \times P$. Note that the complexity is specified in operations per pixel.

the method of choice. Local operations produce an output pixel value based upon the pixel values in the neighborhood .Some of the most common neighborhoods are the 4-connected neighborhood and the 8-connected neighborhood in the case of rectangular sampling and the 6-connected neighborhood in the case of hexagonal sampling illustrated in Figure ??.

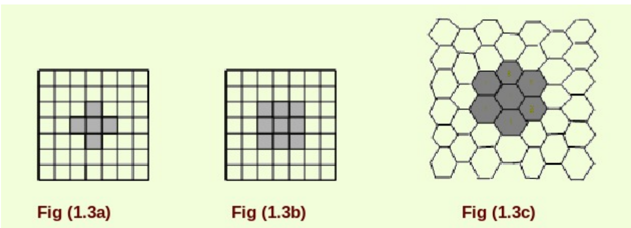


Figure 2: Types of neighborhoods

3.5 Video Parameters

We do not propose to describe the processing of dynamically changing images in this introduction. It is appropriate-given that many static images are derived from video cameras and frame grabbers-to mention the standards that are associated with the three standard video schemes that are currently in worldwide use- NTSC, PAL, and SECAM. This information is summarized in Table ??.

In a interlaced image the odd numbered lines (1, 3, 5.) are scanned in half of the allotted time (e.g. 20 ms in PAL) and the even numbered lines (2, 4, 6,.) are scanned in the remaining half. The image display must be coordinated with this scanning format. The reason for interlacing the

Standard Property	NTSC	PAL	SECAM
Images / Second	29.97	25	25
Ms / image	33.37	40.0	40.0
Lines / image	525	625	625
(horiz./vert.)= aspect radio	4:3	4:3	4:3
interlace	2:1	2:1	2:1
Us / line	63.56	64.00	64.00

Table 3: Standard video parameters

scan lines of a video image is to reduce the perception of flicker in a displayed image. If one is planning to use images that have been scanned from an interlaced video source, it is important to know if the two half-images have been appropriately ”shuffled” by the digitization hardware or if that should be implemented in software. Further, the analysis of moving objects requires special care with interlaced video to avoid ’Zigzag’ edges.

3.5.1 Tools

Certain tools are central to the processing of digital images. These include mathematical tools such as convolution, Fourier analysis, and statistical descriptions, and manipulative tools such as chain codes and run codes. We will present these tools without any specific motivation. The motivation will follow in later sections.

3.6 Convolution

There are several possible notations to indicate the convolution of two (multi-dimensional) signals to produce an output signal. The most common are:

$$c = a \otimes b = a * b$$

We shall use the first form $c = a \otimes b$ with the following formal definitions. In 2D continuous space:

$$C(x, y) = a(x, y) \otimes b(x, y) = \int_{-00}^{+00} \int_{-00}^{+00} a(\lambda, \zeta) b(x-\lambda, y-\zeta) d\lambda d\zeta$$

In 2D discrete space:

$$C(m, n) = a(m, n) \otimes b(m, n) = \sum_{j=-00}^{+00} \sum_{j=-00}^{+00} a(j, k) b(m-j, n-k)$$

3.6.1 Properties of Convolution

There are a number of important mathematical properties associated with convolution.

- Convolution is commutative.
  $c = a \otimes b = b \otimes a$
- Convolution is associative.
  $a \otimes (b \otimes c) = (a \otimes b) \otimes c = a \otimes b \otimes c$

- Convolution is distributive.
  $a \otimes (b + c) = (a \otimes b) + (a \otimes c)$

where a, b, c, and d are all images, either continuous or discrete.

3.7 Fourier Transforms

The Fourier transform produces another representation of a signal, specifically a representation as a weighted sum of complex exponentials. Because of Euler's formula:

$$e^{jq} = \cos(q) + j\sin(q)$$

where $j^2 = -1$, we can say that the Fourier transform produces a representation of a (2D) signal as a weighted sum of sines and cosines. The defining formulas for the forward Fourier and the inverse Fourier transforms are as follows. Given an image a and its Fourier transform A, then the forward transform goes from the spatial domain (either continuous or discrete) to the frequency domain which is always continuous.

$$Forward - A = F\{a\}$$

The inverse Fourier transform goes from the frequency domain back to the spatial domain.

$$Inverse - a = F^{-1}\{A\}$$

The Fourier transform is a unique and invertible operation so that:

In 2D continuous space:

$$Forward : A(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} a(x, y) e^{-j(ux+vy)} dx\, dy$$

$$Inverse : a(x, y) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} A(u, v) e^{+j(ux+vy)} du\, dv$$

In 2D Discrete space:

$$Forward : A(\Omega, \Psi) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} a(m, n) e^{-j(\Omega m+\Psi n)}$$

$$Inverse : a(m, n) = \frac{1}{4\pi^2} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} A(\Omega, \Psi) e^{+j(\Omega m+\Psi n)} d\Omega\, d\Psi$$

3.8 Properties of Fourier Transforms

There are a variety of properties associated with the Fourier transform and the inverse Fourier transform. The following are some of the most relevant for digital image processing.

- The Fourier transform is, in general, a complex function of the real frequency variables. As such the transform con be written in terms of its magnitude and phase.

  $$A(u, v) = |A(u, v)|e^{j\phi(u,v)}; A(\Omega, \Psi) = |A(\Omega, \Psi)|e^{j\phi(\Omega,\Psi)}$$

- A 2D signal can also be complex and thus written in terms of its magnitude and phase.

  $$a(x, y) = |a(x, y)|e^{j\theta(x,y)}; \quad a(m, n) = |a(m, n)|e^{j\theta(m,n)}$$

- If a 2D signal is real, then the Fourier transform has certain symmetries.[1]

  $$A(u, v) = A^*(-u, -v); \quad A(\Omega, \Psi) = A^*(-\Omega, -\Psi)$$

  For real signals equation leads directly to,

  $$|A(u, v)| = |A(-u, -v)|; \quad \phi(u, v) = -\phi(-u, -v)$$

  $$|A(\Omega, \Psi)| = |A(-\Omega, -\Psi)|; \quad \phi(\Omega, \Psi) = -\phi(-\Omega, -\Psi)$$

- If a 2D signal is real and even, then the Fourier transform is real and even

  $$A(u, v) = A(-u, -v); \quad A(\Omega, \Psi) = A(-\Omega, -\Psi)$$

- The Fourier and the inverse Fourier transforms are linear operations

  $$F(w_1 a + w_2 b) = F(w_1 a) + F(w_2 b) = w_1 a + w_2 b$$

  $$F^{-1}(w_1 a + w_2 b) = F^{-1}(w_1 a) + F^{-1}(w_2 b) = w_1 a + w_2 b$$

  where a and b are 2D signals(images) and w_1 and w_2 are arbitrary, complex constants.

- The Fourier transform in discrete space, $A(\Omega, \Psi)$, is periodic in both Ω and Ψ Both periods are 2π

  $$A(\Omega + 2\pi j, \Psi + 2\pi k) = A(\Omega, \Psi) \quad j\ k \in \text{ integers}$$

3.9 Importance of phase and magnitude

The definition indicates that the Fourier transform of an image can be complex. This is illustrated below in Figure **??**.

Figure (1.4a) shows the original image $a[m, n]$, Figure (1.4b) the magnitude in a scaled form as $\log(|A(\Omega, \Psi)|)$ and Figure (1.4c) the phase $\phi(\Omega, \Psi)$. Both the magnitude and the phase functions are necessary for the complete reconstruction of an image from its Fourier transform. Figure(1. 5a) shows what happens when Figure (1.4a) is restored solely on the basis of the magnitude information and Figure (1.5b) shows what happens when Figure (1.4a) is restored solely on the basis of the phase information.

Neither the magnitude information nor the phase information is sufficient to restore the image. The magnitude-only image Figure (1.5a) is unrecognizable and has severe dynamic range problems. The phase-only image Figure (1.5b) is barely recognizable, that is, severely degraded in quality.

[1] The symbol (*) indicates complex conjugation.

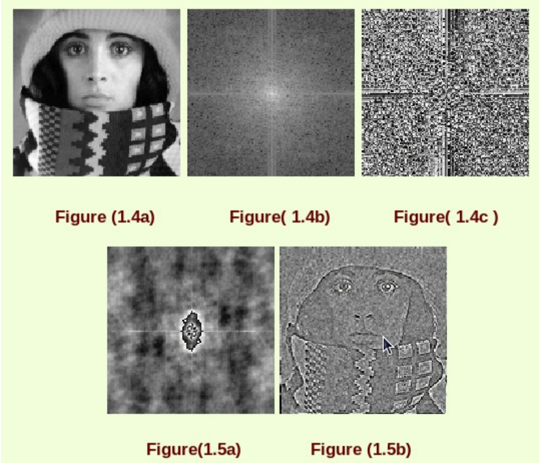


Figure 3: Importance of phase and magnitude

3.9.1 Circularly symmetric signals

An arbitrary 2D signal $a(x,y)$ can always be written in a polar coordinate system as $a(r,\theta)$. When the 2D signal exhibits a circular symmetry this means that:

$$a(x,y) = a(r,\theta) = a(r)$$

where $r^2 = x^2 + y^2$ and $\tan\theta = y/\pi$. As a number of physical systems such as lenses exhibit circular symmetry, it is useful to be able to compute an appropriate Fourier representation.

The Fourier transform $A(u,v)$ can be written in polar coordinates $A(\omega_r, \zeta$ and then, for a circularly symmetric signal, rewritten as a Hankel transform:

$$A(u,v) = F\{a(x,y)\} = 2\pi \int_0^\infty a(r) J_0(w_r r) dr = A(\omega_r) \tag{1}$$

where $\omega_r^2 = u^2 + v^2$ and $\tan\zeta = v/u$ and $J_0(*)$ is a Bessel function of the first kind of order zero.

The inverse Hankel transform is given by:

$$a(r) = \frac{1}{2} \int_0^\infty A(w_r r) w_r dw_r$$

The Fourier transform of a circularly symmetric 2D signal is a function of only the radial frequency w_r. The dependence on the angular frequency ζ has vanished. Further if $a(x,y) = a(r)$ is real, then it is automatically even due to the circular symmetry. According to equ (**??**), will then be real and even.

3.10 Statistics

In image processing it is quite common to use simple statistical descriptions of images and sub-images. The notion of a statistic is intimately connected to the concept of a probability distribution, generally the distribution of signal amplitudes. For a given region-which could conceivably be an entire image-we can define the probability distribution function of the brightnesses in that region and probability density function of the brightnesses in that region. We will assume in the discussion that follows that we are dealing with a digitized image $a(m,n)$.

Probability distribution function of the brightnesses

The probability distribution function $P(a)$, is the probability that a brightness chosen from the region is less than or equal to a given brightness value a. As a increases from $-\infty$ to $+\infty$, $P(a)$ increases from 0 to 1. $P(a)$ is monotonic, non-decreasing in a and thus $\frac{dP}{da} \geq 0$.

Probability density function of the brightnesses

The probability that a brightness in a region falls between a and $a + \triangle a$,given the probability distribution function $P(a)$ can be expressed as $P(a)\triangle a$ where $P(a)$ is the probability density function.

$$P(a)\triangle a = (\frac{dP(a)}{da})\triangle a$$

Because of monotonic, non-decreasing character of $P(a)$ we have $P(a) \geq 0$ and $\int_{-\infty}^{+\infty} P(a)\, da = 1$. For an image with quantized (integer) brightness amplitudes, the interpretation of $\triangle a$ is the width of a brightness interval. We assume constant width intervals. The brightness probability density function is frequently estimated by counting the number of times that each brightness occurs in the region to generate a histogram, $h[a]$.The histogram can then be normalized so that the total area under the histogram is 1. Said another way, the $P(a)$ for region is the normalized count of the number of pixels, N, in a region that have quantized brightness a:

$$p[a] = \frac{a}{N} h[a] \text{ with } N = \sum_{\alpha} h[a]$$

The brightness probability distribution function for the image is shown in Figure(1. 6a). The (unnormalized) brightness histogram which is proportional to the estimated brightness probability density function is shown in Figure(**??**). The height in this histogram corresponds to the number of pixels with a given brightness.

Both the distribution function and the histogram as measured from a region are a statistical description of that region. It must be emphasized that both $P(a)$ and $p(a)$ should be viewed as estimates of true distributions when they are computed from a specific region. That is, we view an image and a specific region as one realization of the various random processes involved in the formation of that image and that region . In the same context, the statistics defined below must be viewed as estimates of the underlying parameters.

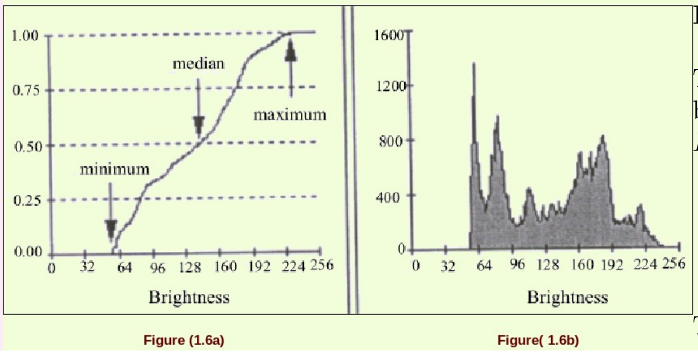


Figure 4: (a) Brightness distribution function of Figure(1. 4a) with minimum, median, and maximum indicated. (b) Brightness histogram of Figure (1.4a).

3.11 Average

The average brightness of a region is defined as sample mean of the pixel brightnesses within that region. The average m_a of the brightness over the N pixels within a region is given by:

$$m_a = \frac{a}{N_{m,n \in R}} \sum a[m,n]$$

Alternatively, we can use a formulation based upon the (unnormalized) brightness histogram, ,with discrete brightness values a. This gives:

$$m_a = \frac{1}{N} \sum_a a.h(a)$$

The average brightness m_a is an estimate of the mean brightness u_a, of the underlying brightness probability distribution.

Standard deviation

The unbiased estimate of the standard deviation, of the brightnesses within a region with N pixels is called the sample standard deviation and is given by:

$$\begin{aligned} S_a &= \sqrt{\frac{1}{N-1} \sum_{m,n \in R} (a[m,n] - m_a)^2} \\ &= \sqrt{\frac{\sum_{m,n \in R} a^2[m,n] - N m_a^2}{N-1}} \end{aligned}$$

Using the histogram formulation gives

$$S_a = \sqrt{\frac{\sum_a a^2 - h[a] - N.m_a^2}{N-1}}$$

The standard deviation S_a is an estimate of σ_a of the underlying brightness probability distribution.

3.12 Coefficient-of-variation

The dimensionless coefficient-of-variation, CV, is defined as:

$$CV = \frac{S_a}{m_a} \times 100\%$$

Percentiles

The percentile, p%, of an unquantized brightness distribution is defined as that value of the brightness such that: $P(a) = p\%$ or equivalently

$$\int_{-\infty}^{a} p(\alpha)\, d\alpha = p\%$$

Three special cases are frequently used in digital image processing.

- 0% the minimum value in the region
- 50% the median value in the region
- 100% the maximum value in the region.

All three of these values can be determined from Figure **??** (1.6a).

Mode

The mode of the distribution is the most frequent brightness value. There is no guarantee that a mode exists or that it is unique.

Signal to noise ratio

The signal-to-noise ratio, SNR, can have several definitions. The noise is characterized by its standard deviation, .The characterization of the signal can differ. If the signal is known to lie between two boundaries, then the SNR is defined as:

- Bounded signal

$$SNR = 20 \log_{10} \left(\frac{s_a}{s_n}\right) dB \tag{2}$$

- Stochastic signal:
  If the signal is not bounded but has a statistical distribution then two other definitions are known:
  S & N inter-dependent $SNR = 20 \log_{10} \left(\frac{s_a}{s_n}\right) dB$ and S & N independent $SNR = 20 \log_{10} \left(\frac{s_a}{s_n}\right) dB$ where m_a and s_a are defined above.

Statistics from Figure ??

A SNR calculation for the entire image based on equ (**??**) is not directly available. The variations in the image brightnesses that lead to the large value of s (=49.5) are not, in general, due to noise but to the variation in local information. With the help of the region there is a way to estimate the SNR. We can use the S_R (=4.0) and the dynamic range $a_{max} - a_{min}$, for the image (=241-56) to calculate a global SNR (=33.3 dB). The underlying assumptions are that

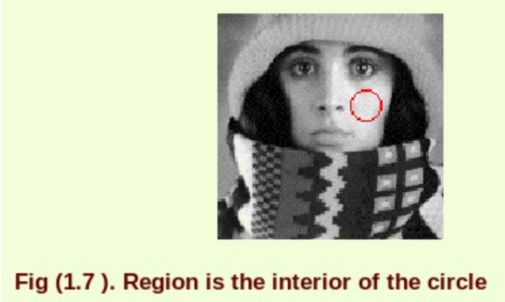


Figure 5: Statistics from region interior of the circle

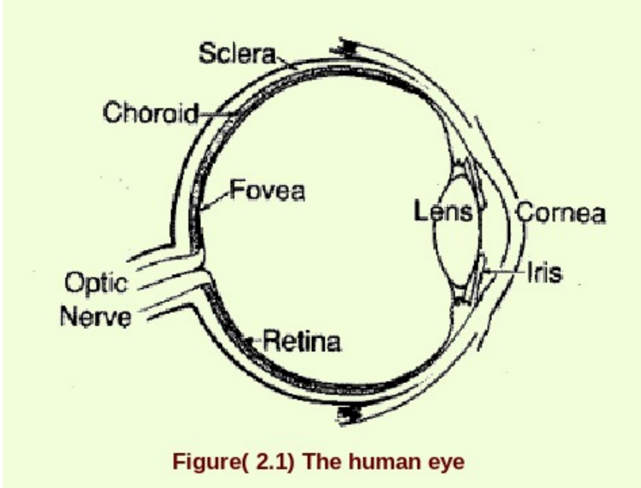


Figure 6: Elements of Human Visual Perception.

Statistic	Image	ROI
Average	137.7	219.3
Standard Deviation	49.5	4.0
Minimum	56	202
Median	141	220
Maximum	241	226
Mode	62	220
SNR (db)	NA	33.3

Table 4: Statistics from Figure **??**

1. The signal is approximately constant in that region and the variation in the region is therefore due to noise, and
2. That the noise is the same over the entire image with a standard deviation given by $S_n = S_R$.

4 Perception

Many image processing applications are intended to produce images that are to be viewed by human observers. It is therefore important to understand the characteristics and limitations of the human visual system to understand the "receiver" of the 2D signals. At the outset it is important to realise that (1) human visual system (HVS) is not well understood; (2) no objective measure exists for judging the quality of an image that corresponds to human assessment of image quality, and (3) the typical human observer does not exist Nevertheless, research in perceptual psychology has provided some important insights into the visual system [stock ham].

The first part of the visual system is the eye. This is shown in figure(**??**). Its form is nearly spherical and its diameter is approximately 20 mm. Its outer cover consists of the 'cornea' and 'sclera'

The cornea is a tough transparent tissue in the front part of the eye. The sclera is an opaque membrane, which is continuous with cornea and covers the remainder of the eye. Directly below the sclera lies the "choroids", which has many blood vessels.At its anterior extreme lies the iris diaphragm. The light enters in the eye through the central opening of the iris, whose diameter varies from 2mm to 8mm, according to the illumination conditions. Behind the iris is the "lens" which consists of concentric layers of fibrous cells and contains up to 60 to 70% of water. Its operation is similar to that of the man made optical lenses. It focuses the light on the "retina" which is the innermost membrane of the eye.

Retina has two kinds of photoreceptors: cones and rods. The cones are highly sensitive to color. Their number is 6-7 million and they are mainly located at the central part of the retina. Each cone is connected to one nerve end.

Cone vision is the photopic or bright light vision. Rods serve to view the general picture of the vision field. They are sensitive to low levels of illumination and cannot discriminate colors. This is the scotopic or dim-light vision. Their number is 75 to 150 million and they are distributed over the retinal surface. Several rods are connected to a single nerve end. This fact and their large spatial distribution explain their low resolution.

Both cones and rods transform light to electric stimulus, which is carried through the optical nerve to the human brain for the high level image processing and perception.

4.1 Model of the Human Eye

Based on the anatomy of the eye, a model can be constructed as shown in Figure(2.2).Its first part is a simple optical system consisting of the cornea, the opening of iris, the lens and the fluids inside the eye. Its second part consists of the retina, which performs the photo electrical transduction, followed by the visual pathway (nerve) which performs simple image processing operations and carries the information to the brain.

Image Formation in the Eye

The image formation in the human eye is not a simple phenomenon. It is only partially understood and only some of the visual phenomena have been measured and understood. Most of them are proven to have non-linear characteristics. Two examples of visual phenomena are:Contrast sensitivity , Spatial Frequency Sensitivity

Contrast sensitivity

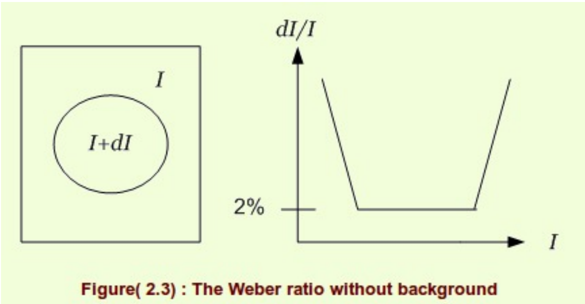


Figure 7: The Weber ratio without background

Let us consider a spot of intensity I+dI in a background having intensity I, as is shown in Figure (2.3) ; dI is increased from 0 until it becomes noticeable. The ratio dI/I, called Weber ratio, is nearly constant at about 2% over a wide range of illumination levels, except for very low or very high illuminations, as it is seen in Figure (2.3). The range over which the Weber ratio remains constant is reduced considerably, when the experiment of Figure (2.4) is considered. In this case, the background has intensity I_0 and two adjacent spots have intensities I and I+dI, respectively. The Weber ratio is plotted as a function of the background intensity in Figure (2.4). The envelope of the lower limits is the same with that of Figure (2.3). The derivative of the logarithm of the intensity I is the Weber ratio:

$$d[\log(I)] = \frac{dl}{I}$$

Thus equal changes in the logarithm of the intensity result in equal noticeable changes in the intensity for a wide range of intensities. This fact suggests that the human eye performs a pointwise logarithm operation on the input image.

Another characteristic of HVS is that it tends to "overshoot" around image edges (boundaries of regions having different intensity). As a result, regions of constant intensity, which are close to edges, appear to have varying intensity. Such an example is shown in Figure (2.5). The stripes appear to have varying intensity along the horizontal dimension, whereas their intensity is constant. This effect is called Mach band effect. It indicates that the human eye is sensitive to edge information and that it has high-pass characteristics

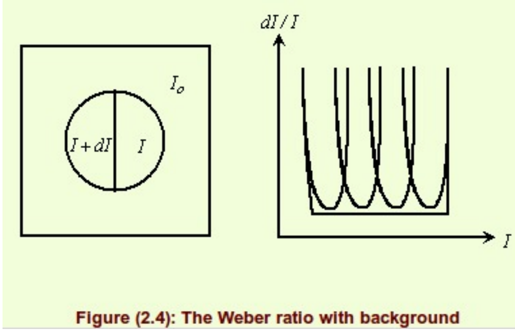


Figure 8: The Weber ratio with background

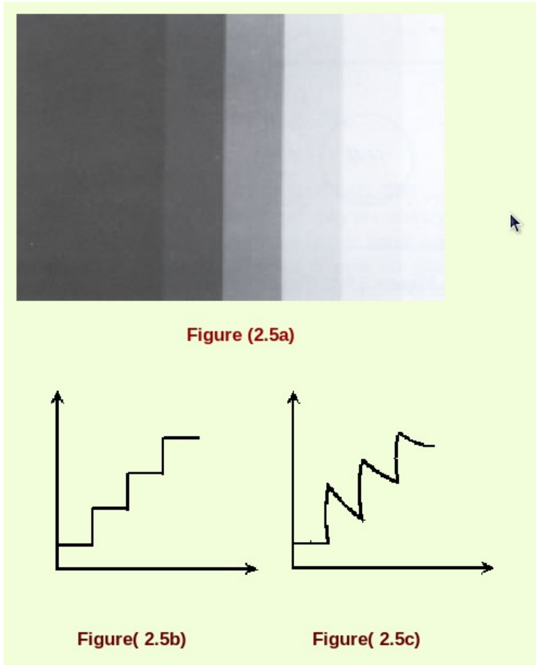


Figure 9: The Mach-band effect: (a) Vertical stripes having constant illumination (b) Actual image intensity profile (c) Perceived image intensity profile

Spatial Frequency Sensitivity

If the constant intensity (brightness) I0 is replaced by a sinusoidal grating with increasing spatial frequency (Figure2.6a), it is possible to determine the spatial frequency sensitivity. The result is shown in Figure (2.6a, 2.6b).

To translate these data into common terms, consider an "ideal" computer monitor at a viewing distance of 50 cm. The spatial frequency that will give maximum response is at 10 cycles per degree. (See figure above) The one degree at 50 cm translates to 50 tan (1 deg.) =0.87 cm on the computer screen. Thus the spatial frequency of maximum response f_{max}==10 cycles/0.87 cm=11.46 cycles/cm at this viewing distance. Translating this into a general formula gives:

$$f_{max} = \frac{10}{d * \tan(1^o)} = \frac{572.9}{d} \text{ Cycles/cm}$$

where d=viewing distance measured in cm.

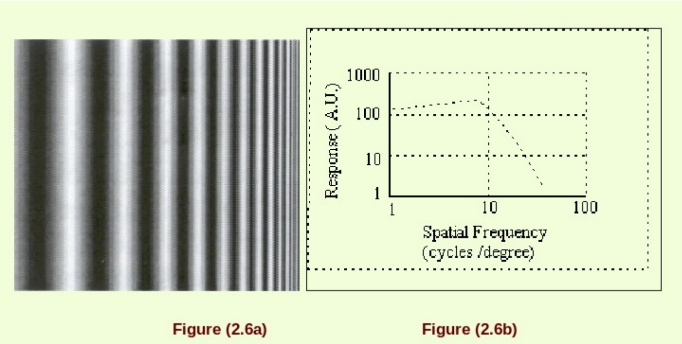


Figure 10: b) shows Sinusoidal test grating ; spatial frequency sensitivity

4.2 Fundamentals of Color Images

Light is a form of electromagnetic (em) energy that can be completely specified at a point in the image plane by its wavelength distribution. Not all electromagnetic radiation is visible to the human eye. In fact, the entire visible portion of the radiation is only within the narrow wavelength band of 380 to 780 nms. Till now, we were concerned mostly with light intensity, i.e. the sensation of brightness produced by the aggregate of wavelengths. However light of many wavelengths also produces another important visual sensation called "color". Different spectral distributions generally, but not necessarily, have different perceived color. Thus color is that aspect of visible radiant energy by which an observer may distinguish between different spectral compositions.

A color stimulus therefore specified by visible radiant energy of a given intensity and spectral composition.Color is generally characterised by attaching names to the different stimuli e.g. white, gray, back red, green, blue. Color stimuli are generally more pleasing to eye than "black and stimuli" .Consequently pictures with color are widespread in TV photography and printing.

Color is also used in computer graphics to add "spice" to the synthesized pictures. Coloring of black and white pictures by transforming intensities into colors (called pseudo colors) has been extensively used by artist's working in pattern recognition. In this module we will be concerned with questions of how to specify color and how to reproduce it. Color specification consists of 3 parts:

1. Color matching
2. Color differences
3. Color appearance or perceived color

We will discuss the first of these questions in this module

4.3 Representation of color for human vision

Let $S(\lambda)$ denote the spectral power distribution (in watts $/m^2$ /unit wavelength) of the light emanating from a pixel of the image plane, and λ the wavelength. The human retina contains pre-dominantly three different color receptors (called cones) that are sensitive to 3 overlapping areas of the visible spectrum. The sensitivities of the receptors peak at approximately 445. (Called blue), 535 (called green) and 570 (called red) nanometers.

Each type of receptors integrates the energy in the incident light at various wavelengths in proportion to their sensitivity to light at that wavelength. The three resulting numbers are primarily responsible for color sensation. This is the basis for trichromatic theory of color vision, which states that the color of light entering the eye may be specified by only 3 numbers, rather than a complete function of wavelengths over the visible range. This leads to significant economy in color specification and reproduction for human viewing. Much of the credit for this significant work goes to the physicist Thomas Young.

The counterpart to trichromacy of vision is the Trichromacy of Color Mixture.

This important principle states that light of any color can be synthesized by an appropriate mixture of 3 properly chosen primary colors.

Maxwell in 1855 showed this using a 3-color projecting system. Several development took place since that time creating a large body of knowledge referred to as colorimetry.

Although trichromacy of color is based on subjective & physiological finding, these are precise measurements that can be made to examine color matches.

4.4 Color matching

Consider a bipartite field subtending an angle (<) of 2^o at a viewer's eye. The entire field is viewed against a dark, neutral surround. The field contains the test color on left and an adjustable mixture of 3 suitably chosen primary colors on the right as shown in Figure (2.7).

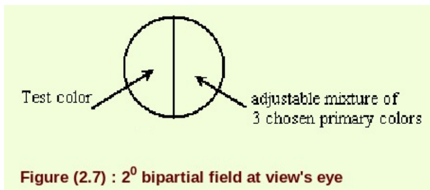


Figure 11: 2^o bipartial field at view's eye

It is found that most test colors can be matched by a proper mixture of 3 primary colors as long as the primary colors are independent. The primary colors are usually chosen as red, green & blue or red, green & violet. The "tristimulus values" of a test color are the amount of 3 primary colors required to give a match by additive mixture.They are unique within an accuracy of the experiment. Much of colorimetry is based on experimental results as well as rules attributed to Grassman. Two important rules that are valid over a large range of observing conditions are "linearity " and "additivity". They state that,

1. The color match between any two color stimuli holds even if the intensities of the stimuli are increased or decreased by the same multiplying factor, as long as their relative spectral distributions remain unchanged. As an example, if stimuli $s_1(\lambda)$ and $s_2(\lambda)$ match, and stimuli $s_3(\lambda)$ and $s_4(\lambda)$ also match, then additive mixtures $s_1(\lambda)+s_3(\lambda)$ and $s_2(\lambda)+s_4(\lambda)$ will also match.

2. Another consequence of the above rules of Grassman trichromacy is that any four colors cannot be linearly independent. This implies tristimulus value of one of the 4 colors can be expressed as linear combination of tristimulus values of remaining 3 colors.. That is, any color C is specified by its projection on 3-axes R, G, B corresponding to chosen set of primaries. This is shown in Figure ??

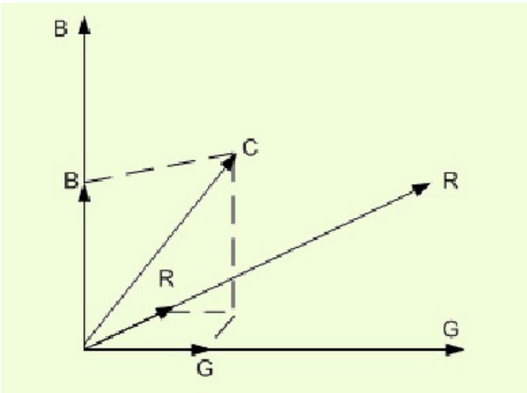


Figure 12: The color-matching functions for the 20 Standard Observer , using primaries of wavelengths 700(red), 546.1 (green), and 435.8 nm (blue), with units such that equal quantities of the three primaries are needed to match the equal energy white, E

Consider a mixture of two colors S1 and S2 i.e S=S1+S2 If S1 is specified by (Rs1, Gs1, Bs1) and S2 is specified by (Rs2, Gs2, Bs2) This implies, S is specified by (Rs1,+Rs2,Gs1,+Gs2,Bs1,+Bs2)

The constraint of color matching experiment is that only non-ve amounts of primary colors can be added to match a test color. In practice this is not sufficient to effect a match. In this case, since negative amounts of primary cannot be produced, a match is made by simple transposition i.e. by adding positive amounts of primary to the test color

a test color S might be matched by S+3G=2R+B or S=2R-3G+B

⇒The negative tristimulus values (2,-3,1) present no special problem.

By convention, tristimulus values are expressed in normalized form. This is done by a preliminary color experiment in which left side of the split field shown in Fig (2.7), is allowed to emit light of unit intensity whose spectral distribution is constant wrt λ i.e. (equal energy white E).Then the amount of each primary required for a match is taken by definition as one "unit".

The amount of primaries for matching other test colors is then expressed in terms of this unit. In practice equal energy white 'E' is matched with positive amounts of each primary.

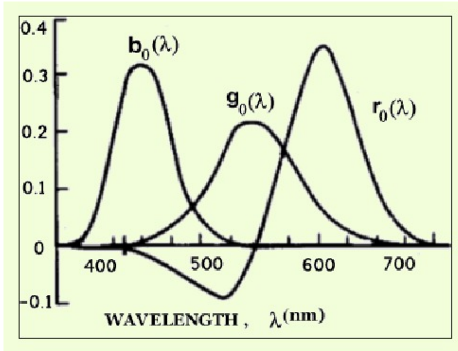


Figure 13: 2^o bipartial field at view's eye

4.5 Color-Coordinate Systems.

4.6 CIE System of Color Specification

4.7 Chromaticity coordinates in CIE-XYZ system.

4.8 Color Mixtures

4.9 Polar Coordinate Representation of color

4.10 Color Transformations

5 Sampling

It is generally true that all discrete sequences are formed in an attempt to represent some underlying continuous signal. Although many discrete representations of continuous signals are possible, periodic sampling is by far the representation mostly used due to the simplicity of its implementation. We consider, in this section, the relationships between continuous signals and the discrete sequences which are obtained from them by periodic sampling. In particular, we first consider the specific case of rectangular periodic sampling, and then a more general case of periodic sampling with arbitrary sampling geometries.

 ktiwari [at] cse.iitk.ac.in

5.0.1 Two dimensional rectangular sampling

We discuss 2D rectangular sampling of a stillerriage $x_a(t_1, t_2)$ in two spatial coordinates. In spatial rectangle sampling, we sample at the locations.

$$\begin{aligned} t_1 &= n_1 T_1 \\ t_2 &= n_2 T_2 \end{aligned} \tag{3}$$

Where T_1 and T_2 are sampling distances in the t_1 and t_2 directions, respectively. The 2D rectangular sampling grid is depicted in figure **??** below. The sampled signal can

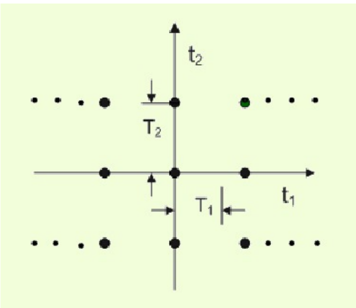


Figure 14: 2D rectangular sampling grid

be expressed in terms of the unitless coordinate variables (n_1, n_2) as:

$$x(n_1, n_2) = x_a(n_1T_1, n_2T_2) \qquad \forall (n_, n_2) \in Z^2$$

In some cases it is convenient to define an intermediate sampled signal in terms of continuous coordinate variables given by,

$$\begin{aligned} x_p(t_1, t_2) &= x_a(t_1, t_2) \sum_{n_1} \sum_{n_2} \delta(t_{\_} n_1T_1, t_2 - n_2T_2) \\ &= \sum_{n_1} \sum_{n_2} x_a(n_1t_1, n_2t_2) \delta(t_{\_} n_1T_1, t_2 - n_2T_2) \\ &= \sum_{n_1} \sum_{n_2} x(n_1, n_2) \delta(t_{\_} n_1T_1, t_2 - n_2T_2) \end{aligned}$$

Note $x_p(t_1, t_2)$ is indeed a sampled signal because of the presence of 2D Dirac delta functions.

5.0.2 Spectrum of the sampled signal

We now relate the Fourier $X_p(\Omega_1, \Omega_2)$ transform or $X(w_1, w_2)$ of the sampled signal to that of the continuous signal $x_a(t_1, t_2)$. As given earlier, the 2D continuous space Fourier transform $X_a(\Omega_1, \Omega_2)$ of a $x_a(t_1, t_2)$ signal with continuous variables (t_1, t_2) is given by,

$$X_a(\Omega_1, \Omega_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_a(t_1, t_2) exp^{-j2\pi(t_1\Omega_1 + t_2\Omega_2)} dt_1 \, dt_2$$

where $(\Omega_1, \Omega_2) \in R^2$ and the inverse Fourier transform is given by,

$$x_a(t_1, t_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} X_a(\Omega_1, \Omega_2) exp^{-j2\pi(t_1\Omega_1 + t_2\Omega_2)} d\Omega_1 \, d\Omega_2$$

Here the spatial frequency variables Ω_1, Ω_2 have the units in cycles/mm and are related to radian frequencies by a scale factor of 2π. In order to evaluate the 2D FT $X_p(\Omega_1, \Omega_2)$ of $x_p(t_1, t_2)$ after substitution of equations and exchange the order of function and summation to obtain,

$$\begin{aligned} X_p(\Omega_1, \Omega_2) &= \sum_{n_1} \sum_{n_2} x_a(n_1T_1, n_2T_2) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\\ &\quad \delta(t_1 - n_1T_1, t_2 - n_2T_2) \times \\ &\quad exp^{-j2\pi(\Omega_1 t_1, \Omega_2 t_2)}) \, dt_1 \, dt_2 \end{aligned}$$

which simplifies as,

$$X_p(\Omega_1, \Omega_2) = \sum_{n_1} \sum_{n_2} x_a(n_1T_1, n_2T_2) e^{-j2\pi(\Omega_1 n_1 T_1, \Omega_2 n_2 T_2)}$$

Note that $X_p(\Omega_1, \Omega_2)$ is periodic with the fundamental period given by the region $\Omega_2 < |\frac{1}{2T_1}|$ and $\Omega_1 < |\frac{1}{2T_2}|$. Letting $w_1 = \Omega_1 T_1$ and $w_2 = \Omega_2 T_2$ and using above equations, we obtain the discrete space Fourier transform relation, in terms of unitless frequency variables w_1 and w_2 as,

5.1 Evaluation

- One MidSem Exam
- Assignment (Metlab based)
- **Term Paper** understand the paper, fillup the gapes, demonstration/seminar. Will be done in groups (of two).
- EndSem Exam

5.2 Questions

Q1. Why do we process images?

- Picture digitization and coding - for transmission, storage, and printing
- Picture enhansment and restoration
- Picture segmentation and description - for machine vision, image understanding.

Q2. What is image?

- Panchromatic - gray scale, 2D light intensity function *f(x,y)*
- Multispectral - color image, *f(x,y)* is a vector (R,G,B)

Q3. What is digital image?

Descritize both in special and intensity function.

$$\mathbf{g} = \begin{bmatrix} f(1,1) & \ldots & f(1,N) \\ \vdots & & \vdots \\ f(N,1) & \ldots & f(N,N) \end{bmatrix}_{N \times N}$$

$0 \le f(x,y) \le G-1$, where G is number of gray levels and is bounded by $G = 2^m$

Number of bits required to represent an image is $N \times N \times m$ typically it is 512*512*8

Q4. What is resolution of an image?
Minute details observable in image. Checherboard pattern.
Reducing m but keeping N constant is called **false contouring**.

6 Quantization

6.1 Sensitivity

For color images we use three set of sensors as below.

Book:*Image Processing, The Fundamentals* Maria Petrou, Panaglota Bosdogianni, *Publisher:* John Willey & Sons

Purpose of Image processing

- Picture degitization and coding
- Picture enhansment and restoration
- Picture segmentation and description

Monochrome Imageis a function $f(x, y)$, which is 2D having light intensity values[2]. An image point at position x, y is called *pixel*.

An image of size $N \times N$ with 2^m different gray levels needs $N \times N \times m$ bits.

chackerboard effectkeeping m constant and decreasing N produces this effect.

False contouring[3]Keeping N constant and reducing m. For more detailed picture (like picture of croud) false contouring have less effects.

Resolutionexpresses how much details we can see in picture.

Brightness of a pixel is the light intensity value recorded at sensor from corresponding physical object part.

Image processing is doneby using image tramsformations. Which in turn done using operators. Operator takes an image (called input image) and produces another image (called output image).

Linear operator '$\circledast$' have following property

$$\circledast[af + bg] = a \circledast [f] + b \circledast [g]$$

Point spread function of an operator is what we get out if we apply the operator on a point source

$$\circledast[\text{Point source}] = \text{point spread function}$$

$$\circledast[\delta(x - \alpha, y - \beta)] = h(x, \alpha, y, \beta)$$

where $\delta(x - \alpha, y - \beta)$ is a point source of brightness 1 centred at point (α, β).

In other words point spread function $h(x, \alpha, y, \beta)$ expresses how much the input value at position (x, y) influences the value at point (α, β).

For a linear operator $\circledast[a\delta(x-\alpha, y-\beta)] = a\circledast[\delta(x-\alpha, y-\beta)]$.

Operator are defined in terms of **point spread functions**

Effect of an operator characterized by $h(x, \alpha, y, \beta)$ on an image $f(x, y)$ can be written as

$$g(\alpha, \beta) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) h(x, \alpha, y, \beta)$$

Shift invarient PSF do not have influance based on actual pixel position rather it depends on relative positions.

$$h(x, \alpha, y, \beta) = h(\alpha - x, \beta - y)$$

Convolutionunder assumption of shift invarient PSF, following is convolution

$$g(\alpha, \beta) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) h(\alpha - x, \beta - y)$$

Seprable PSFWhen columns are influenced independently from the rows of the image the PSF is calles seprable.

$$h(x, \alpha, y, \beta) \equiv h_c(x, \alpha) h_r(y, \beta)$$

For such case we can write as below

$$g(\alpha, \beta) = \sum_{x=0}^{N-1} h_c(x, \alpha) \sum_{y=0}^{N-1} f(x, y) h_r(y, \beta)$$

When PSF is both *shift invarient* and *seprable*, then

$$g(\alpha, \beta) = \sum_{x=0}^{N-1} h_c(\alpha - x) \sum_{y=0}^{N-1} f(x, y) h_r(\beta - y)$$

Define an extended source of constant brightness

$$\delta_n(x, y) \equiv n^2 rect(nx, ny)$$

Where n is a positive constant

$$rect(nx, ny) \equiv \begin{cases} 1 & \text{inside a rectangle } |nx| \le \frac{1}{2}, |ny| \le \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

The total brightness of this source is given by

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta_n(x, y)\, dx\, dy = n^2 \underbrace{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} rect(nx, ny)\, dx\, dy}_{\text{area of rectangle}} = 1$$

Dirac delta function

$$\delta(x, y) = \begin{cases} \neq 0 & \text{for } x = y = 0 \\ 0 & \text{elsewhere} \end{cases}$$

[2] $0 \le f(x, y) \le G - 1$, where G is the maximum possible intensity value. G is generally in the form 2^m

[3] **Contour:** the outline of a figure or body

With the property

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(x,y)\ dx\ dy = 1$$

It has an interesting property

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(x,y)g(x,y)\ dx\ dy = g(0,0)$$

Shifting property[4]

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta_n(x-a,y-b)g(x,y)\ dx\ dy = g(a,b)$$

Fundamental equation of linear image processing

$$\boxed{\text{g} = \text{H f}}$$

Where

- $H = h_{\alpha\beta}$
- $h_{\alpha\beta}^T \equiv [h(0,\alpha,0,\beta), h(1,\alpha,0,\beta), ..., h(N-1,\alpha,0,\beta), h(0,\alpha,1,\beta), h(1,\alpha,1,\beta), ..., h(N-1,\alpha,1,\beta), ..., h(0,\alpha,N-1,\beta), h(1,\alpha,N-1,\beta), ..., h(N-1,\alpha,N-1,\beta)]$
- Image $f^T \equiv [f(0,0), f(1,0), ..., f(N-1,0), f(0,1), f(1,1), ..., f(N-1,1), ..., f(0,N-1), f(1,N-1), ..., f(N-1,N-1)]$

V$_n$it is $N \times 1$ column matrix having all elements except n^{th} set as zero.

Image processing refers to processing of two dimensional picture by a digital computer. Basic classes of image processing applications

- Image representation and modeling
  Concerns with charqcterization of the quantity that each picture-element represents. Image can represent luminance of object in the scene, absorption characteristics of the body tissue, the radar cross section of a target, temperature profile of a region ... or anything.
- Image enhencement
  The goal is to accenture certain image feature for sbusequent analysis or for image display.
- Image restoration
  Refers to removal or minimization of known degradation in an image. It includes deblurring, noise filtering, geometric distortion corrections .. etc.
- Image analysis
  Concerns with making quantative measurements from an image to produce its discription.

[4] $\delta_n(x-a,y-b) = n^2 rect[n(x-a), n(y-b)]$

- Image reconstruction
  This is a special class of image restoration problems where a two-(or higher) dimensional object is reconstructed from sveral one-dimensional projections.
- Image data compression.
  Reduces the amount of storage required for same visual information.

j is $\sqrt{-1}$, z^* is the complex conjugate of z

Seprable form for several well known one-dimensional functions their two-dimensional versions are seprable form

$$f(x,y) = f_1(x)f_2(y)$$

Dirac Delta $\delta(x) = 0$ for $x \neq 0$

$$\lim_{\epsilon\to 0}\int_{-\epsilon}^{+\epsilon} \delta(x)\ dx = 1$$

Shifting Property

$$\int_{-\infty}^{+\infty} f(x')\delta(x-x')\ dx' = f(x)$$

Scaling Property

$$\delta(ax) = \frac{\delta(x)}{|a|}$$

Kronecker Delta

$$\delta(n) = \begin{cases} 0 & n \neq 0 \\ 1 & n = 0 \end{cases}$$

Shifting Property of Kronecker Delta

$$\sum_{m=-\infty}^{+\infty} f(m)\delta(n-m) = f(n)$$

Rectangle

$$rect(x) = \begin{cases} 1 & |x| \leq \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases}$$

Signum

$$sgn(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

Sinc

$$sinc(x) = \frac{\sin \pi x}{\pi x}$$

Comb

$$comb(x) = \sum_{n=-\infty}^{\infty} \delta(x-n)$$

ktiwari [at] cse.iitk.ac.in

Triangle

$$tri(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & |x| > 1 \end{cases}$$

Linear Syatem

$$\mathcal{H}[a_1 x_1(m,n) + a_2 x_2(m,n)] = a_1 \mathcal{H}[x_1(m,n)] + a_2 \mathcal{H}[x_2(m,n)]$$

Impulse Response When input is Kronecker delta function at location (m', n') the output at location (m, n) is defined as

$$h(m,n,m',n') = \mathcal{H}[\delta(m - m', n - n')]$$

PSFimpulse response is called the Point Spread Function when input and output represent a positive quantity.

Output of any linear systemcan be obtained from its impulse response and the input by applying the superposition rule

$$\begin{aligned} y(m,n) &= \mathcal{H}[x(m,n)] \\ &= \mathcal{H}\big[\sum_{m'}\sum_{n'} x(m',n')\delta(m-m',n-n')\big] \\ &= \sum_{m'}\sum_{n'} x(m',n')\mathcal{H}\big[\delta(m-m',n-n')\big] \\ &= \sum_{m'}\sum_{n'} x(m',n')\ h(m,n,m',n') \end{aligned}$$

Shift invarient (or spatially invarient) system if translation of input causes the translation of output.

$$\mathcal{H}[\delta(m,n)] = h(m,n;0,0)$$

By definition

$$h(m,n,m',n') \triangleq \mathcal{H}[\delta(m - m', n - n')]$$

$$h(m,n,m',n') = h(m - m', n - n')$$

Convolution for shift invarient system the output becomes

$$y(m,n) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h(m - m', n - n')x(m', n')$$

7 Fourier Transform

The fourier transform of a complex function $f(x)$ is fedined as

$$F(\xi) \triangleq \mathcal{F}[f(x)] \triangleq \int_{-\infty}^{\infty} f(x)\ exp(-j2\pi\xi x)\ dx$$

The inverse Fourier Transform $F(\xi)$ is defined as

$$f(x) \triangleq \mathcal{F}^{-1}[F(\xi)] \triangleq \int_{-\infty}^{\infty} F(\xi)\ exp(j2\pi\xi x)\ dx$$

Two dimensional Fourier trnsform is defined in similar way as

$$F(\xi_1, \xi_2) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ exp[-j2\pi(x\xi_1 + y\xi_2)]\ dx\ dy$$

Properties of the Fourier Transform.

1. **Spatial Frequencies**
   If $f(x,y)$ is luminance and x, y the spatial coordinates, then ξ_1, ξ_2 are the spacial frequencies that represent luminance changes with respect to spatial distances.
2. **Uniqueness**
   For continuous functions $f(x,y)$ and $F(\xi_1, \xi_2)$ are unique with one another.
3. **Seprability**
   By definition of FT kernal is seprable, so that it can be written as a seprable transform in x and y that is

   $$F(\xi_1, \xi_2) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ exp[-j2\pi(x\xi_1 + y\xi_2)]\ dx\ dy$$
4. **Frequency response and eigenfunction of shift invarient system1**

8 Image Perception

Lightis the electromagnetic radiation that stimulates our visual response. Light received from an oblect can be written as

$$I(\lambda) = \rho(\lambda)L(\lambda)$$

Where $\rho(\lambda)$ represents the reflectivity or transmissivity of the object, $L(\lambda)$ is the incident energy distribution.

Human Photoreceptors

Property	Rods	Cones
No.	100 million	6.5 million
Vision	Scotopic	Photopic
Color	No Color	color vision
Nerves	one for group	one for every
Concentration	outer to fovea	near to fovea

Luminance or intensityof a spatially distributed object with light distribution $I(x, y, \lambda)$ is defined as

$$f(x,y) = \int_{0}^{\infty} I(x,y,\lambda)\ V(\lambda)\ d\lambda$$

where $V(\lambda)$ is *relative luminous efficiency function* of the vidual system. For human eye $V(\lambda)$ is bell-shaped curve.

Brightness is the perceived luminance and depends on the luminance of surround. Objects with same luminance can have different brightness.

Simultaneous ContrastSince our perception is sensitive to luminus contrast rather tham the absolute value, therefore two squares of same luminance value embedded between

$\overset{o}{\kappa}$tiwari [at] cse.iitk.ac.in

square of different darkness will appear of different luminance.

Weber's LawIf the luminance f_o of an object is just noticeably different from the luminance of sorround f_s, then

$$\frac{|f_s - f_o|}{f_o} = \text{ Constant}$$

when $f_o = f$ and $f_s = f + \triangle f$ we can say

$$\frac{\triangle f}{f} = d(\log f) = \text{ Constant}$$

Value of Constant is found to be 0.02

Mach BandsThe effect shows that brightness is not a monotonic function of luminance. Consider any two adjacent different gray level bars; The apparent brightness is not uniform alonf the width of the bar. Transition of the gray level at the bar appears brighter at darker side and diaker at lighter side. The overshoot and undershoot illustrates the Match band effect. (ref page 75 of book)

Modulation Transfer Function (MTF) also known as spatial frequency response is a metric which characterizes sharpness of a photographic imaging system or of a component of the system (lens, film, image sensor, scanner, enlarging lens, etc.)

9 Image Sampling and Quantization

Bandlimited signalA signal is said to be a band limited signal if all of it's frequency components are zero above a certain finite frequency. A function $f(x, y)$ is *bandlimited* if its Fourier transform $F(\xi_1, \xi_2)$ is zero outside a bounded region in frequency plane, that is $F(\xi_1, \xi_2) = 0$, $|\xi_1| > \xi_{x0}$ and $|\xi_2| > \xi_{y0}$. Quantities ξ_{x0}, and ξ_{y0} are called the x and y bandwidth of the image.

Fourier transform of an arbitary sampled functionis a scaled, periodic replication of the Fourier transform of the original function.

Ideal image sampling function is a two dimensional infinite array of Dirac delta function situated on a rectangular grid with spacing $\Delta x, \Delta y$

$$\text{comb}(x, y, \Delta x, \Delta y) \triangleq \sum\sum_{m,n=-\infty}^{\infty} \delta(x - m\Delta x, y - n\Delta y)$$

Samples Imageis defined as

$$\begin{aligned} f_s(x, y) &= f(x, y)\text{comb}(x, y; \Delta x, \Delta y) \\ &= \sum\sum_{m,n=-\infty}^{\infty} f(m\Delta x, n\Delta y)\delta(x - m\Delta x, y - n\Delta y) \end{aligned} \quad (4)$$

Fourier Transform of a comb function with spacing $\Delta x, \Delta y$ is another comb function with spacing $(\frac{1}{\Delta x}, \frac{1}{\Delta y})$ namely

$$\begin{aligned} \text{COMB}(\xi_1, \xi_2) &= \mathbb{F}\{\text{comb}(x, y, \Delta x, \Delta y)\} \\ &= \xi_{xs}\xi_{ys} \sum\sum_{k,l=-\infty}^{\infty} \delta(\xi_1 - k\xi_{xs}, \xi_2 - k\xi_{ys}) \\ &= \xi_{xs}\xi_{ys} \sum\sum_{k,l=-\infty}^{\infty} \text{comb}(\xi_1 - k\xi_{xs}, \xi_2 - k\xi_{ys}) \end{aligned} \quad (5)$$

Where $\xi_{xs} \triangleq \frac{1}{\Delta x}$, and $\xi_{ys} \triangleq \frac{1}{\Delta y}$, Finally Fourier transform of the sampled image $f_s(x, y)$ is given by convolution

$$\begin{aligned} F_s(\xi_1, \xi_2) &= F(\xi_1, \xi_2) \circledast \text{COMB}(\xi_1, \xi_2) \\ &= \xi_{xs}\xi_{ys} \sum\sum_{k,l=-\infty}^{\infty} F(\xi_1, \xi_2) \circledast \delta(\xi_1 - k\xi_{xs}, \xi_2 - k\xi_{ys}) \\ &= \xi_{xs}\xi_{ys} \sum\sum_{k,l=-\infty}^{\infty} F(\xi_1 - k\xi_{xs}, \xi_2 - k\xi_{ys}) \end{aligned} \quad (6)$$

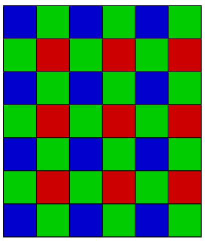
Therefore the Fourier transform of a sampled image is, within a scaler factor, a periodic replication of the Fourier transform of the input image on a grid whose spacing is (ξ_{xs}, ξ_{ys})

Reconstruction of the image from its samples If the x, y sampling frequencies are grater than twice the bandwidth, that is $\xi_{xs} > 2\xi_{x0}, \xi_{ys} > 2\xi_{y0}$ or equivalently $\Delta x < \frac{1}{2\xi_{x0}}, \Delta y < \frac{1}{2\xi_{y0}}$ Then $F(\xi_1, \xi_2)$ can be reconstructed by a low-pass filter with frequency response

$$H(\xi_1, \xi_2) = \begin{cases} \frac{1}{\xi_{xs}\xi_{ys}} & \text{if } \xi_1, \xi_2 \in \Re \\ 0 & \text{Otherwise} \end{cases}$$

Where $\Re$ is any region whose boundary $\partial\Re$ is contained within the annular ring between the rectangles $\Re_1$ and $\Re_2$.

Bayer filter mosaicis a color filter array for arranging RGB color filters on a square grid of photosensors. The filter pattern is 50% green, 25% red and 25% blue to mimic the physiology of the human eye[5]. Bryce Bayer's patent in 1976 called the green photosensors *luminance-sensitive* elements and the red and blue ones *chrominance-sensitive elements.*



aliasingwraparound error

9.1 References

- A. K. Jain - Fundamental of Digital Image Processing.

[5]The retina has more rod cells than cone cells and rod cells are most sensitive to green light.

- Gonzales - DIP
- Netravdi & Haskell - Digital Pictures.
- J R Ohms - Multimedia Communication Technology.
- Bovik - Handbook of Image and Video Processing
- Sayood - Data Compression
- NPTEL notes.

$\overset{o}{\kappa}$tiwari [at] cse.iitk.ac.in