SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

UNIT – I Digital Image and Multimedia Processing – SEC1605

I. Digital Image Fundamentals

Elements of Visual Perception; Image Sensing and Acquisition; Image Sampling and Quantization; Basic Relationships between Pixels; Monochromatic Vision Models; Colour Vision Models; Colour Fundamentals; Colour Models

Elements of Visual Perception

Although the field of digital image processing is built on a foundation of mathematical and probabilistic formulations, human intuition and analysis play a central role in the choice of one technique versus another, and this choice often is made based on subjective, visual judgments. Hence, developing a basic understanding of human visual perception is necessary. In particular, our interest is in the mechanics and parameters related to how images are formed and perceived by humans.

Structure of Human eye

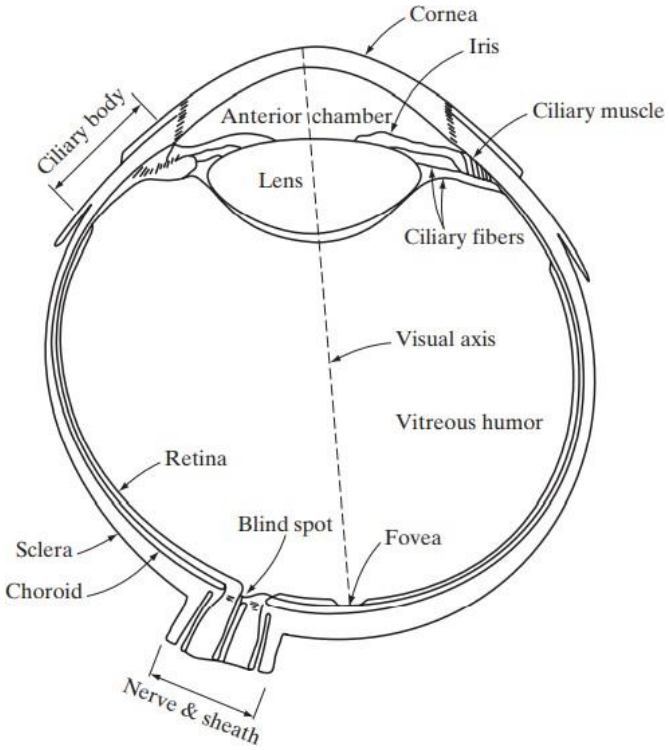


Fig. 1.1 Simplified diagram of a cross section of the human eye.

Figure 1.1 shows a simplified horizontal cross section of the human eye. The eye is nearly a sphere, with an average diameter of approximately 20 mm. Three membranes enclose the eye: the cornea and sclera outer cover; the choroid; and the retina. The cornea is a tough, transparent tissue that covers the anterior surface of the eye. Continuous with the cornea, the sclera is an opaque membrane that encloses the remainder of the optic globe. The choroid lies directly below the sclera. This membrane contains a network of blood vessels that serve as the major source of nutrition to the eye. Even superficial injury to the choroid, often not deemed serious, can lead to severe eye damage as a result of inflammation that restricts blood

flow. The choroid coat is heavily pigmented and hence helps to reduce the amount of extraneous light entering the eye and the backscatter within the optic globe. At its anterior extreme, the choroid is divided into the ciliary body and the iris. The latter contracts or expands to control the amount of light that enters the eye. The central opening of the iris (the pupil) varies in diameter from approximately 2 to 8 mm. The front of the iris contains the visible pigment of the eye, whereas the back contains a black pigment. The lens is made up of concentric layers of fibrous cells and is suspended by fibers that attach to the ciliary body. It contains 60 to 70% water, about 6% fat, and more protein than any other tissue in the eye. The lens is colored by a slightly yellow pigmentation that increases with age. In extreme cases, excessive clouding of the lens, caused by the affliction commonly referred to as cataracts, can lead to poor color discrimination and loss of clear vision. The lens absorbs approximately 8% of the visible light spectrum, with relatively higher absorption at shorter wavelengths. Both infrared and ultraviolet light are absorbed appreciably by proteins with in the lens structure and, in excessive amounts, can damage the eye.

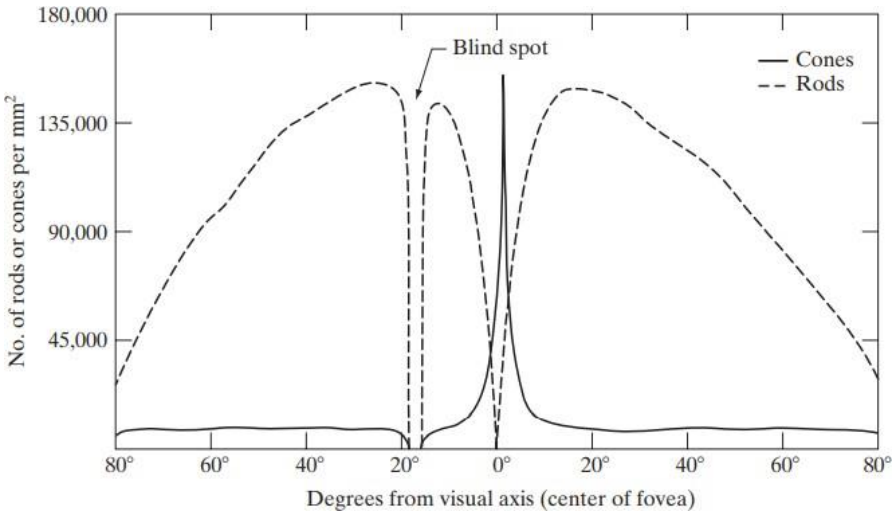


Fig. 1.2 Distribution of rods and cones in the retina.

The innermost membrane of the eye is the retina, which lines the inside of the wall's entire posterior portion. When the eye is properly focused, light from an object outside the eye is imaged on the retina. Pattern vision is afforded by the distribution of discrete light receptors over the surface of the retina. There are two classes of receptors: cones and rods. The cones in each eye number between 6 and 7 million. They are located primarily in the central portion of the retina, called the fovea, and are highly sensitive to color. Humans can resolve fine details with these cones largely because each one is connected to its own nerve end. Muscles controlling the eye rotate the eyeball until the image of an object of interest falls on the fovea. Cone vision is called photopic or bright-light vision. The number of rods is much larger: Some 75 to 150 million are distributed over the retinal surface. The larger area of distribution and the fact that several rods are connected to a single nerve end reduce the amount of detail discernible by these receptors. Rods serve to give a general, overall picture of the field of view. They are not involved in color vision and are sensitive to low levels of illumination. For example, objects that appear brightly colored in daylight when seen by moonlight appear as colorless forms because only the rods are stimulated. This phenomenon is known as scotopic or dim-light vision. Figure 1.2 shows the density of rods and cones for a cross

section of the right eye passing through the region of emergence of the optic nerve f rom the eye. The absence of receptors in this area results in the so-called blind spot. Except f or this region, the distribution of receptors is radially symmetric about the fovea. Receptor density is measured in degrees from the fovea (that is, in degrees off axis, as measured by the angle formed by the visual axis and a line passing through the center of the lens and intersecting the retina). Note in Fig.1.2 that cones are most dense in the center of the retina (in the center area of the fovea). Note also that rods increase in density from the center out to approximately 20° off axis and then decrease in density out to the extreme periphery of the retina. The fovea itself is a circular indentation in the retina of about 1.5 mm in diameter.

Image Formation in the Eye

In an ordinary photographic camera, the lens has a fixed focal length, and focusing at various distances is achieved by varying the distance between the lens and the imaging plane, where the film (or imaging chip in the case of a digital camera) is located. In the human eye, the converse is true; the distance between the lens and the imaging region (the retina) is fixed, and the focal length needed to achieve proper focus is obtained by varying the shape of the lens. The fibers in the ciliary body accomplish this, flattening or thickening the lens for distant or near objects, respectively. The distance between the center of the lens and the retina along the visual axis is approximately 17 mm. The range of focal lengths is approximately 14 mm to 17 mm, the latter taking place when the eye is relaxed and focused at distances greater than about 3 m. The geometry in Fig. 1.3 illustrates how to obtain the dimensions of an image formed on the retina. For example, suppose that a person is looking at a tree 15 m high at a distance of 100 m. Letting h denote the height of that object in the retinal image, the geometry of Fig. 1.3 yields or As indicated, the retinal image is focused primarily on the region of the fovea. Perception then takes place by the relative excitation of light receptors, which transform radiant energy into electrical impulses that ultimately are decoded by the brain.

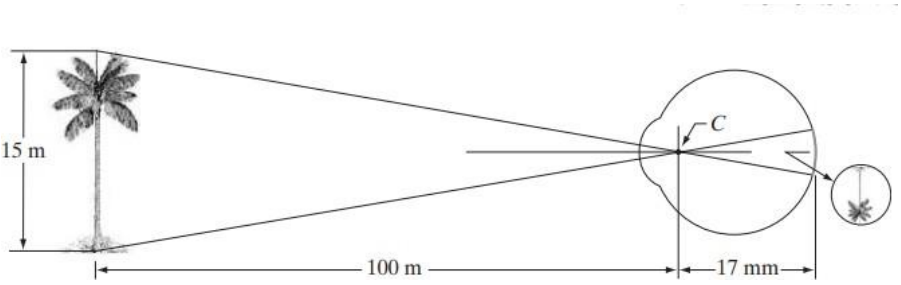


Fig. 1.3 Graphical representation of the eye looking at a palm tree. Point C is the optical center of the lens

Brightness Adaptation and Discrimination

Because digital images are displayed as a discrete set of intensities, the eye's ability to discriminate between different intensity levels is an important consideration in presenting image processing results. The range of light intensity levels to which the human visual

system can adapt is enormous—on the order of — from the scotopic threshold to the glare limit. Experimental evidence indicates that subjective brightness (intensity as perceived by the human visual system) is a logarithmic function of the light intensity incident on the eye.

Image Sensing and Acquisition

Most of the images in which we are interested are generated by the combination of an "illumination" source and the reflection or absorption of energy from that source by the elements of the "scene" being imaged. We enclose illumination and scene in quotes to emphasize the fact that they are considerably more general than the familiar situation in which a visible light source illuminates a common everyday 3-D (three-dimensional) scene. For example, the illumination may originate from a source of electromagnetic energy such as radar, infrared, or X-ray system. But, as noted earlier, it could originate from less traditional sources, such as ultrasound or even a computer-generated illumination pattern. Similarly, the scene elements could be familiar objects, but they can just as easily be molecules, buried rock formations, or a human brain. Depending on the nature of the source, illumination energy is reflected from, or transmitted through, objects. An example in the first category is light reflected from a planar surface. An example in the second category is when X-rays pass through a patient's body for the purpose of generating a diagnostic X-ray film. In some applications, the reflected or transmitted energy is focused onto a photoconverter (e.g., a phosphor screen), which converts the energy into visible light. Electron microscopy and some applications of gamma imaging use this approach. Figure 1.4 shows the three principal sensor arrangements used to transform illumination energy into digital images. The idea is simple: Incoming energy is transformed into a voltage by the combination of input electrical power and sensor material that is responsive to the particular type of energy being detected. The output voltage waveform is the response of the sensor(s), and a digital quantity is obtained from each sensor by digitizing its response. In this section, we look at the principal modalities for image sensing and generation.

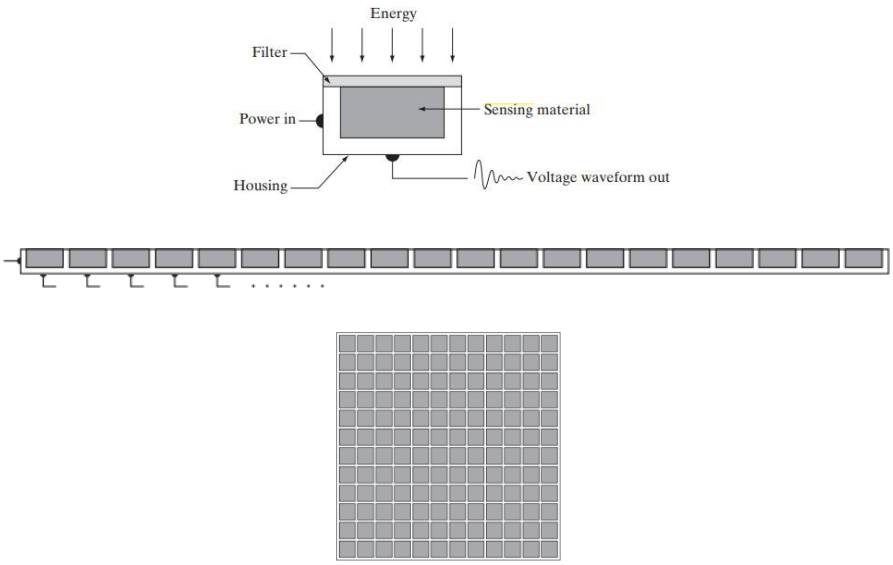


Fig. 1.4 (a) Single imaging sensor. (b) Line sensor. (c) Array sensor.

Image Acquisition Using a Single Sensor

Figure 1.4 (a) shows the components of a single sensor. Perhaps the most familiar sensor of this type is the photodiode, which is constructed of silicon materials and whose output voltage waveform is proportional to light. The use of a filter in front of a sensor improves selectivity. For example, a green (pass) filter in front of a light sensor favors light in the green band of the color spectrum. As a consequence, the sensor output will be stronger f or green light than for other components in the visible spectrum. In order to generate a 2 -D image using a single sensor, there has to be relative displacements in both the x- and y-directions between the sensor and the area to be imaged.

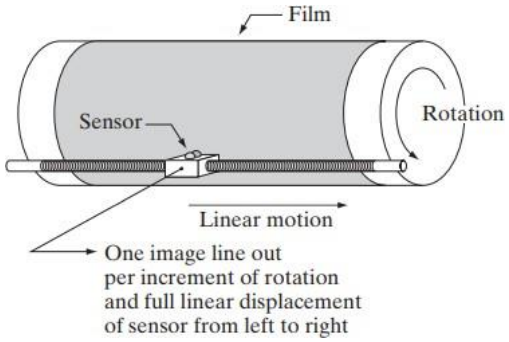


Fig. 1.5 Combining a single sensor with motion to generate a 2-D image.

Figure 1.5 shows an arrangement used in high-precision scanning, where a f ilm negative is mounted onto a drum whose mechanical rotation provides displacement in one dimension. The single sensor is mounted on a lead screw that provides motion in the perpendicular direction. Because mechanical motion can be controlled with high precision, this method is an inexpensive (but slow) way to obtain high-resolution images. Other similar mechanical arrangements use a flat bed, with the sensor moving in two linear directions. These types of mechanical digitizers sometimes are referred to as microdensitometers. Another example of imaging with a single sensor places a laser source coincident with the sensor. Moving mirrors are used to control the outgoing beam in a scanning pattern and to direct the ref lect ed laser signal onto the sensor. This arrangement can be used also to acquire images using strip and array sensors, which are discussed in the following two sections.

Image Acquisition Using Sensor Strips

A geometry that is used much more frequently than single sensors consists of an in-line arrangement of sensors in the form of a sensor strip, as Fig. 1.4 (b) shows. The strip provides imaging elements in one direction Motion perpendicular to the strip provides imaging in the other direction, as shown in Fig. 1.6 (a). This is the type of arrangement used in most flat bed scanners. Sensing devices with 4000 or more in-line sensors are possible. In-line sensors are used routinely in airborne imaging applications, in which the imaging system is mounted on an aircraft that flies at a constant altitude and speed over the geographical area to be imaged. One-dimensional imaging sensor strips that respond to various bands of the electromagnet ic spectrum are mounted perpendicular to the direction of flight. The imaging strip gives one line of an image at a time, and the motion of the strip completes the other dimension of a two-dimensional image. Lenses or other focusing schemes are used to project the area to be

scanned onto the sensors. Sensor strips mounted in a ring configuration are used in medical and industrial imaging to obtain cross-sectional ("slice") images of 3-D objects, as Fig. 1.6 (b) shows. A rotating X-ray source provides illumination and the sensors opposite the source collect the X-ray energy that passes through the object (the sensors obviously have to be sensitive to X-ray energy). This is the basis for medical and industrial computerized axial tomography (CAT) imaging.

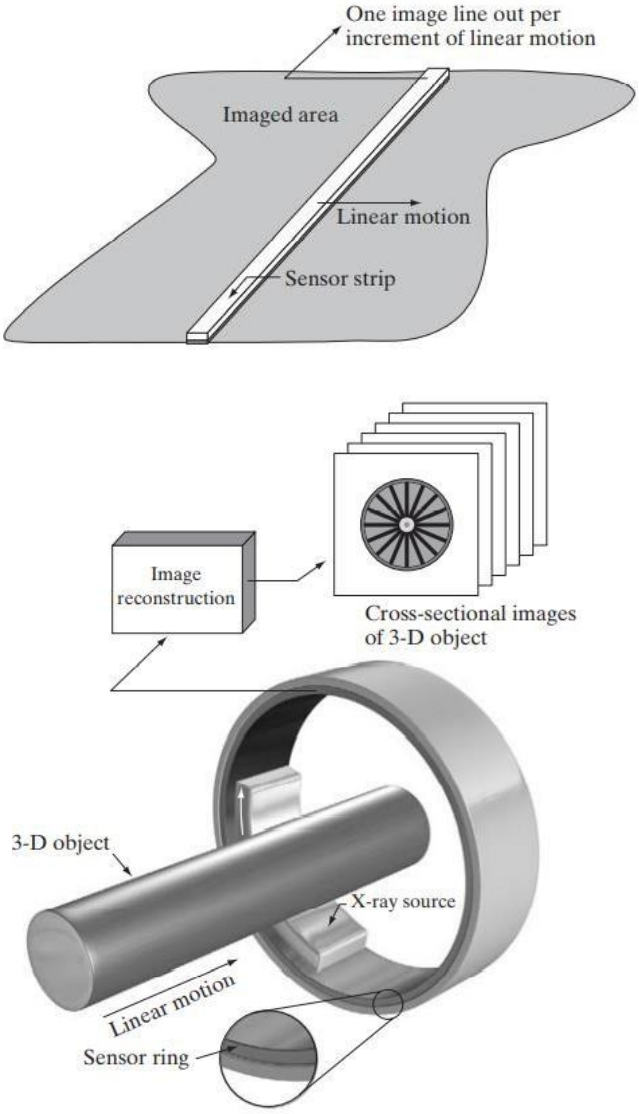


Fig. 1.6 (a) Image acquisition using a linear sensor strip. (b) Image acquisition using a circular sensor strip.

Image Acquisition Using Sensor Arrays

Figure 1.4 (c) shows individual sensors arranged in the form of a 2 -D array. Numerous electromagnetic and some ultrasonic sensing devices frequently are arranged in an array format. This is also the predominant arrangement found in digital cameras. A typical sensor for these cameras is a CCD array, which can be manufactured with a broad range of sensing properties and can be packaged in rugged arrays of elements or more. CCD sensors are used widely in digital cameras and other light sensing instruments. The response of each sensor is

proportional to the integral of the light energy projected onto the surface of the sensor, a property that is used in astronomical and other applications requiring low noise images. Noise reduction is achieved by letting the sensor integrate the input light signal over minutes or even hours. Because the sensor array in Fig. 1.4 (c) is two-dimensional, its key advantage is that a complete image can be obtained by focusing the energy pattern onto the surface of the array. Motion obviously is not necessary, as is the case with the sensor arrangements discussed in the preceding two sections. The principal manner in which array sensors are used is shown in Fig. 1.7. This figure shows the energy from an illumination source being reflected from a scene element (as mentioned at the beginning of this section, the energy also could be transmitted through the scene elements). The first function performed by the imaging system in Fig. 1.7 (c) is to collect the incoming energy and focus it onto an image plane. If the illumination is light, the front end of the imaging system is an optical lens that projects the viewed scene onto the lens focal plane, as Fig. 1.7 (d) shows. The sensor array, which is coincident with the focal plane, produces outputs proportional to the integral of the light received at each sensor. Digital and analog circuitry sweep these outputs and convert them to an analog signal, which is then digitized by another section of the imaging system. The output is a digital image, as shown diagrammatically in Fig. 1.7(e).

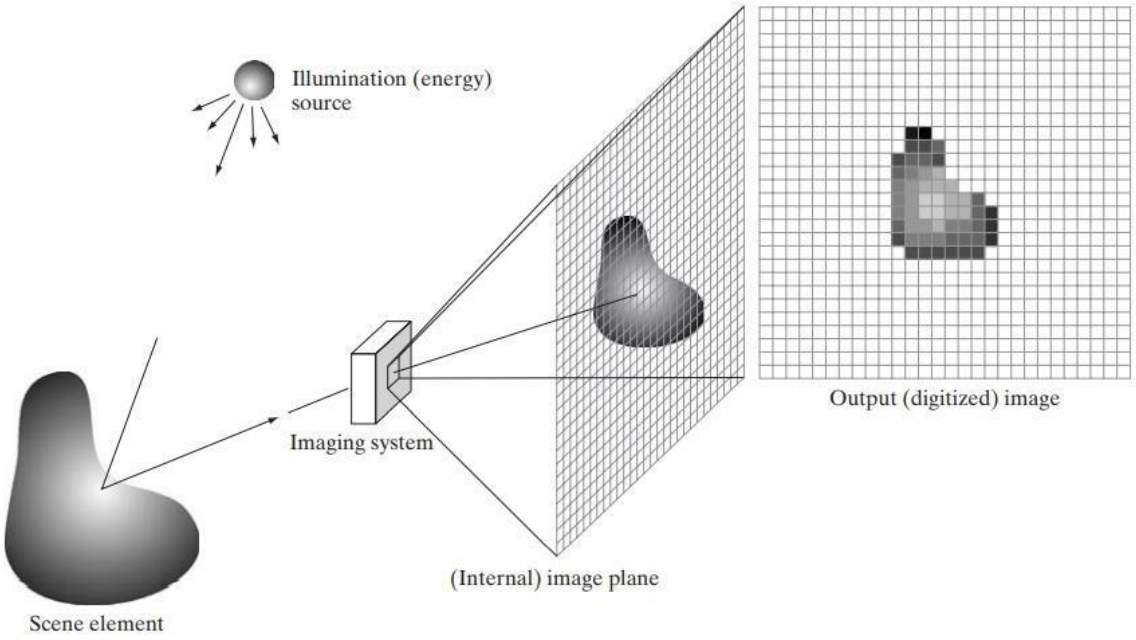


Fig. 1.7 An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image

Image Sampling and Quantization

The basic idea behind sampling and quantization is illustrated in Fig. 1.8. Figure 1.8 (a) shows a continuous image f that we want to convert to digital form. An image may be continuous with respect to the x- and y-coordinates, and also in amplitude. To convert it to digital form, we have to sample the function in both coordinates and in amplitude. Digitizing

the coordinate values is called sampling. Digitizing the amplitude values is called quantization. The one-dimensional function in Fig. 1.8 (b) is a plot of amplitude (intensity level) values of the continuous image along the line segment AB in Fig. 1.8(a). The random variations are due to image noise. To sample this function, we take equally spaced samples along line AB, as shown in Fig. 1.8(c). The spatial location of each sample is indicated by a vertical tick mark in the bottom part of the figure. The samples are shown as small white squares superimposed on the function. The set of these discrete locations gives the sampled function. However, the values of the samples still span (vertically) a continuous range of intensity values. In order to form a digital function, the intensity values also must be converted (quantized) into discrete quantities. The right side of Fig. 1.8(c) shows the intensity scale divided into eight discrete intervals, ranging from black to white. The vertical tick marks indicate the specific value assigned to each of the eight intensity intervals. The continuous intensity levels are quantized by assigning one of the eight values to each sample. The assignment is made depending on the vertical proximity of a sample to a vertical tick mark. The digital samples resulting from both sampling and quantization are shown in Fig. 1.8 (d). Starting at the top of the image and carrying out this procedure line by line produces a two-dimensional digital image. It is implied in Fig. 1.8 that, in addition to the number of discrete levels used, the accuracy achieved in quantization is highly dependent on the noise content of the sampled signal. Sampling in the manner just described assumes that we have a continuous image in both coordinate directions as well as in amplitude. In practice, the method of sampling is determined by the sensor arrangement used to generate the image.

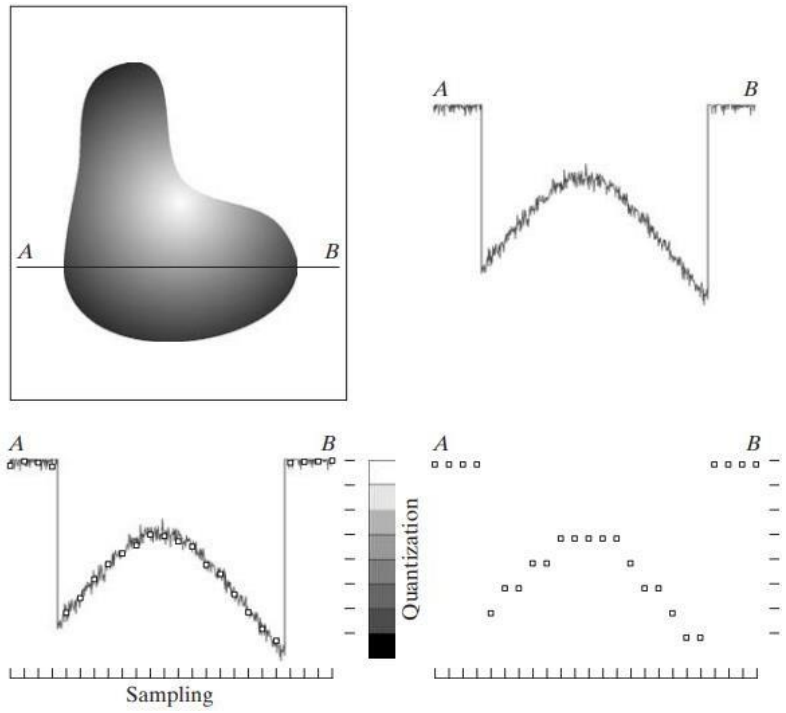


Fig. 1.8 Generating a digital image. (a) Continuous image. (b) A scan line from A to B in the continuous image, used to illustrate the concepts of sampling and quantization. (c) Sampling and quantization. (d) Digital scan line.

When an image is generated by a single sensing element combined with mechanical motion, as in Fig. 1.5, the output of the sensor is quantized in the manner described above. However , spatial sampling is accomplished by selecting the number of individual mechanical increments at which we activate the sensor to collect data. Mechanical motion can be made very exact so, in principle, there is almost no limit as to how fine we can sample an image using this approach. In practice, limits on sampling accuracy are determined by other factors, such as the quality of the optical components of the system. When a sensing strip is used for image acquisition, the number of sensors in the strip establishes the sampling limitations in one image direction. Mechanical motion in the other direction can be controlled more accurately, but it makes little sense to try to achieve sampling density in one direction that exceeds the sampling limits established by the number of sensors in the other. Quantization of the sensor outputs completes the process of generating a digital image. When a sensing array is used for image acquisition, there is no motion and the number of sensors in the array establishes the limits of sampling in both directions. Quantization of the sensor outputs is as before. Figure 1.9 illustrates this concept. Figure 1.9(a) shows a continuous image projected onto the plane of an array sensor. Figure 1.9(b) shows the image after sampling and quantization. Clearly, the quality of a digital image is determined to a large degree by the number of samples and discrete intensity levels used in sampling and quantization.

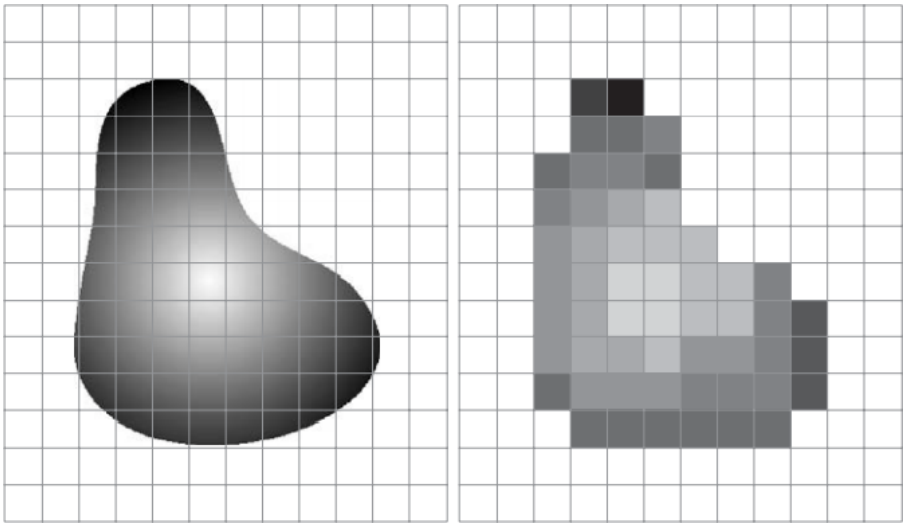


Fig. 1.9 (a) Continuous image projected onto a sensor array. (b) Result of image sampling and quantization.

Basic Relationships between Pixels

Neighbors of a Pixel

A pixel p at coordinates (x,y) has four horizontal and vertical neighbors whose coordinates are given by

$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)$$

This set of pixels, called the 4-neighbors of p, is denoted by N4(p).

Each pixel is a unit distance from (x, y), and some of the neighbor locations of p lie outside the digital image if (x, y) is on the border of the image.

The four diagonal neighbors of p have coordinates

(x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1)

and are denoted by ND(p).

These points, together with the 4-neighbors, are called the 8-neighbors of p, denoted by N8(p).

Adjacency, Connectivity, Regions, and Boundaries

Let V be the set of intensity values used to define adjacency. In a binary image, V = {1} if we are referring to adjacency of pixels with value 1. In a gray-scale image, the idea is the same, but set V typically contains more elements. For example, in the adjacency of pixels with a range of possible intensity values 0 to 255, set V could be any subset of these 256 values. We consider three types of adjacency:

(a) 4-adjacency. Two pixels p and q with values from V are 4-adjacent if q is in the set

(b) 8-adjacency. Two pixels p and q with values from V are 8-adjacent if q is in the set

(c) m-adjacency (mixed adjacency). Two pixels p and q with values from V are m-adjacent if

(i) q is in or

(ii) q is in ND(p) and the set $N_4(p) \cap N_4(q)$ has no pixels whose values are from V.

Mixed adjacency is a modification of 8-adjacency. It is introduced to eliminate the ambiguities that often arise when 8-adjacency is used.

A (digital) path (or curve) from pixel p with coordinates (x, y) to pixel q with coordinates is a sequence of distinct pixels with coordinates

(x0, y0), (x1, y1), … , (xn, yn)

where (x0, y0) = (x, y), (xn, yn) = (s, t) and pixels (xi, yi) and (xi-1, yi-1) are adjacent for 1 <= i <= n. In this case, n is the length of the path. . If (x0, y0) = (xn, yn) the path is a closed path

Let S represent a subset of pixels in an image. Two pixels p and q are said to be connected in S if there exists a path between them consisting entirely of pixels in S. For any pixel p in S, the set of pixels that are connected to it in S is called a connected component of S. If it only has one connected component, then set S is called a connected set.

Let R be a subset of pixels in an image. We call R a region of the image if R is a connected set. Two regions, and are said to be adjacent if their union forms a connected set. Regions that are not adjacent are said to be disjoint. We consider 4- and 8-adjacency when referring to regions.

Distance Measures

For pixels p, q, and z, with coordinates (x, y), (s, t), and (v, w), respectively, D is a distanc e function or metric if

i. D(p, q) >= 0 (D(p, q) = 0 iff p = q),

ii. D(p, q) = D(q, p) and

iii. D(p, z) < = D(p, q) + D(q, z).

The Euclidean distance between p and q is defined as

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

For this distance measure, the pixels having a distance less than or equal to some value r from (x, y) are the points contained in a disk of radius r centered at (x, y).

The distance (called the city-block distance) between p and q is defined as

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

In this case, the pixels having a D4 distance from (x, y) less than or equal to some value r form a diamond centered at (x, y). For example, the pixels with D4 distance <=2 from (x, y) (the center point) form the following contours of constant distance:

```
      2
    2 1 2
  2 1 0 1 2
    2 1 2
      2
```

The pixels with D4=1 are the 4-neighbors of (x, y).

The distance (called the chessboard distance) between p and q is defined as

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

In this case, the pixels with D8 distance from (x, y) less than or equal to some value r f orm a square centered at (x, y). For example, the pixels with D8 distance <= 2 from (x, y) (the center point) form the following contours of constant distance:

```
2 2 2 2 2
2 1 1 1 2
2 1 0 1 2
2 1 1 1 2
2 2 2 2 2
```

The pixels with D8=1 distance are the 8-neighbors of (D x, y).

Monochromatic Vision Models

When Light enters the eye,

- optical characteristics are represented by **LPF (Low Pass Filter)** with frequency response $\mathbf{H_l(\xi_1, \xi_2)}$
- spatial response are represented by the **relative luminous efficiency** function V(λ), yields the luminance distribution f(x, y) via

$$f(x, y) = \int_0^\infty I(x, y, \lambda) V(\lambda)\, d\lambda$$

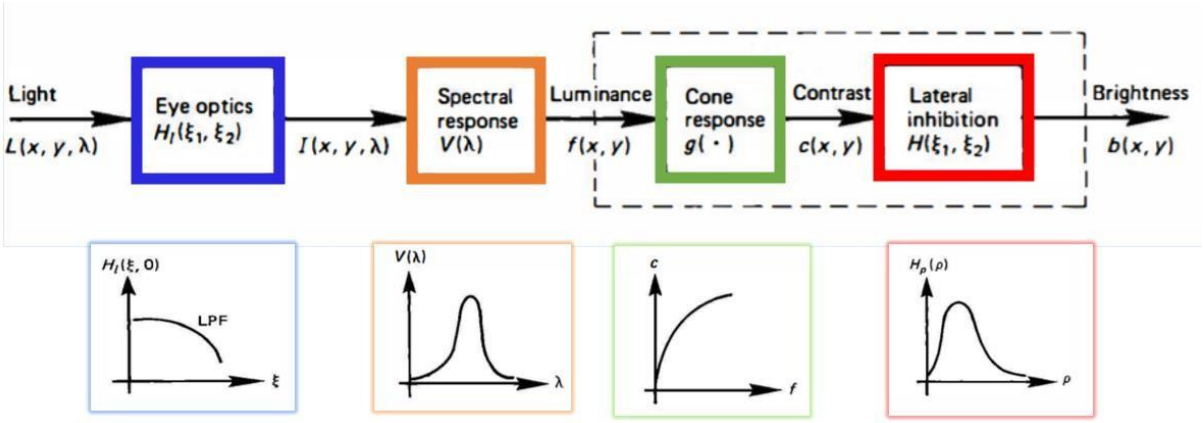


Fig. 1.10 Monochrome Vision Model

- The nonlinear response of the rods and cones, represented by the point nonlinearity **g (·),** yields the contrast **c(x, y)**
- The lateral inhibition phenomenon is represented by a spatially invariant, isotropic, linear system whose frequency response is $\mathbf{H(\xi_1, \xi_2)}$
- Its output is the neural signal, which represents the apparent brightness **b(x, y)**
- For an optically well-corrected eye, the low-pass filter has a much slower drop-off with increasing frequency than that of the lateral inhibition mechanism
- Thus the optical effects of the eye could be ignored, and the simpler model showing the transformation between the luminance and the brightness suffices

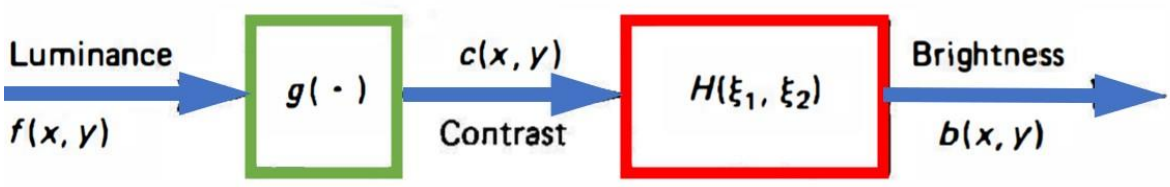


Fig. 1.11 Simplified Monochrome Vision Model

Colour Vision Models

- The color image is represented by the $\mathbf{R_N}$, $\mathbf{G_N}$, $\mathbf{B_N}$ coordinates at each pixel.
- The matrix A transforms the input into the three **cone responses α_k (x, y, C),** k = 1, 2, 3 where (x, y) are the spatial pixel coordinates and C refers to its color
- In Fig., we have represented the **normalized cone responses**
- In analogy with the definition of tristimulus values, Tk are called the retinal cone tristimulus coordinates
- The cone responses undergo nonlinear point transformations to give three f ields Tk (x, y), k = 1, 2, 3
- The 3 x 3 matrix B transforms the {f(x, y)} into {Ck(x, y)} such that C1 (x, y) is the monochrome (achromatic) contrast field c(x, y), as in simplified model, and C2 (x, y) and C3(x, y) represent the corresponding chromatic fields

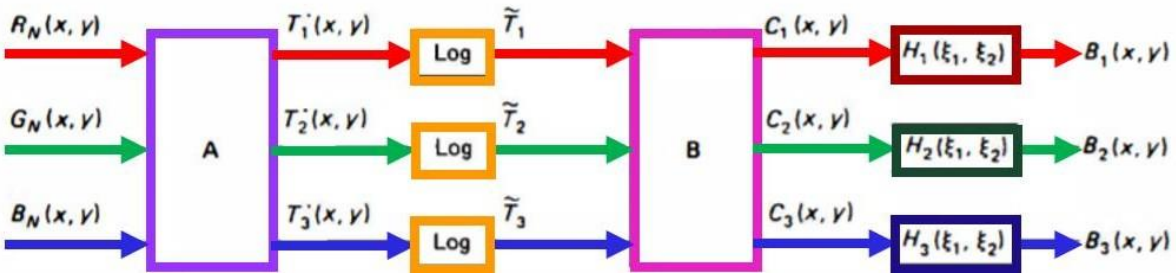


Fig. 1.12 Colour Vision Model

- The spatial filters $H_k(\xi_1 , \xi_2)$, k = 1, 2, 3, represent the frequency response of the visual system to luminance and chrominance contrast signals
- Thus $H1(\xi_1 , \xi_2)$ is the same as H (ξ 1 , ξ 2) in simplified model and is a bandpass filter that represents the lateral inhibition phenomenon

$$T_k \triangleq \frac{\alpha_k(x, y, C)}{\alpha_k(x, y, W)}, \qquad k = 1, 2, 3$$

- The visual frequency response to chrominance signals are not well established but are believed to have their passbands in the lower frequency region, as shown in figure
- The 3 x 3 matrices A and B are given as follows:

$$\mathbf{A} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.127 & 0.724 & 0.175 \\ 0.000 & 0.066 & 1.117 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 21.5 & 0.0 & 0.00 \\ -41.0 & 41.0 & 0.00 \\ -6.27 & 0.0 & 6.27 \end{pmatrix}$$

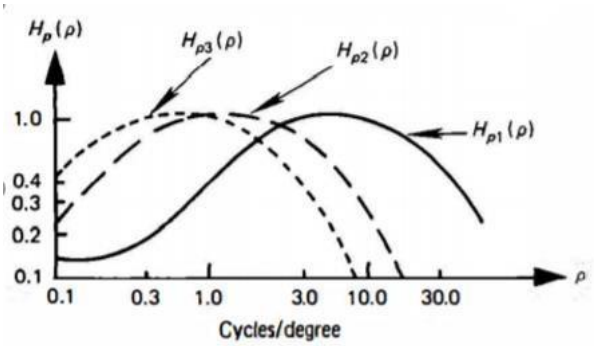


Figure 1.13 Frequency responses of the three color channels C_1, C_2, C_3 of the color vision model. Each filter is assumed to be isotropic so that $H_{pk}(\rho) \triangleq H_k(\xi_1, \xi_2)$, $\rho = \sqrt{\xi_1^2 + \xi_2^2}$, $k = 1, 2, 3$.

- From the model , a criterion for color image fidelity can be defined
- For example, for two color images {RN , GN , BN} and {R,,V, G,,V, BN}, their subjective mean square error could be defined by

$$e_{ls} = \frac{1}{A} \sum_{k=1}^{3} \iint_{\Re} (B_k(x, y) - B_i(x, y))^2 dx\, dy$$

Where r is the region which over the image is defined (or available), A is its area. And $\{B_k(x,y)\}$ and $\{B_{kdot}(x,y)\}$ are the outputs of the model for the two colour images.

Colour Fundamentals

Color spectrum

- In 1666, Sir Isaac Newton
- When a beam of sunlight is passed through a glass prism
- The emerging beam of light consists of a continuous spectrum of colors ranging from violet to red (not white)
- Divided into six broad regions
    - Violet, blue, green, yellow, orange, and red
- Each color blends smoothly into the next

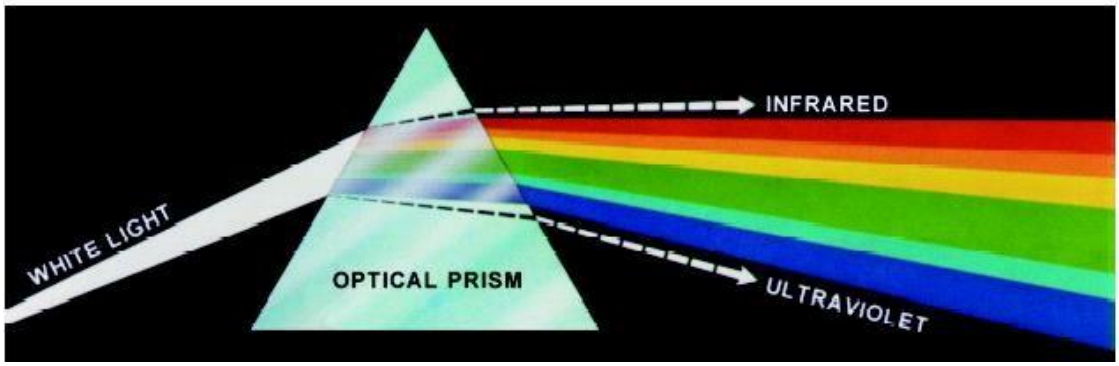


FIGURE 1.14 Color spectrum seen by passing white light through a prism. (Courtesy of the General Electric Co., Lamp Business Division.)

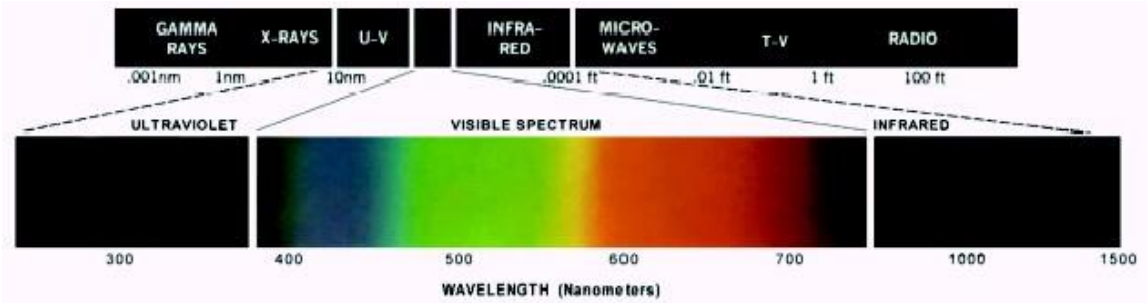


FIGURE 1.15 Wavelengths comprising the visible range of the electromagnetic spectrum. (Courtesy of the General Electric Co., Lamp Business Division.)

Primary colors of pigments or colorants

- cyan, magenta, yellow
- A primary color of pigments is defined as one that subtracts or absorbs a primary color of light and reflects or transmits the other two

Secondary colors of pigments or colorants

- red, green, blue
- Combination of the three pigment primaries, or a secondary with its opposite primary, produces black

Characteristics of colors

- Brightness:
  - The chromatic notion of intensity
- Hue:
  - An attribute associated with the dominant wavelength in a mixture of light waves
  - Representing dominant color as perceived by an observer
- Saturation
  - Referring to relative purity or the amount of white mixed with a hue
  - Saturation is inversely proportional to the amount of white light

Hue and saturation taken together are called chromaticity

- A color may be characterized by its brightness and chromaticity
- The amounts of red, green, and blue needed to form any particular color are called the tristimulus values (Denoted X(red), Y(green), and Z(blue))

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad z = \frac{Z}{X+Y+Z}$$

$$x + y + z = 1 \ , \ X = \frac{x}{y}Y, Z = \frac{z}{y}Y$$

Chromaticity diagram

- Showing color composition as a function of x(R) and y(G)
- For any value of x and y, z(B) = 1 – (x + y)

- The positions of the various spectrum colors are indicated around the boundary of the tongue-shaped
- The point of equal energy = equal fractions of the three primary colors = CIE standard for white light
- Boundary : completely saturated
- Useful for color mixing
- A triangle(RGB) does not enclose the entire color region

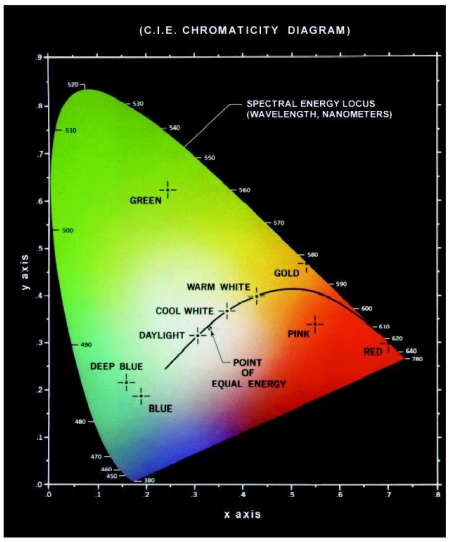


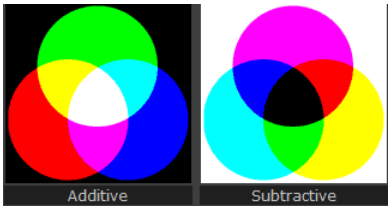
Fig. 1.17 Chromaticity diagram

Colour Models

The purpose of a color model is to facilitate the specification of colors in some standard Color models are oriented either toward hardware or applications

- Hardware-oriented
  - Color monitor or Video camera : RGB
  - Color printer : CMY
  - Color TV broadcast : YIQ (I : inphase, q : quadrature)
- Color image manipulation : HSI, HSV
- Image processing : RGB, YIQ, HIS

- Additive processes create color by adding light to a dark background (Monitors)
- Subtractive processes use pigments or dyes to selectively block white light (Printers)



Additive Color Mixing			Subtractive Color Mixing		
Red + Green	→	Yellow	Cyan + Magenta	→	Blue
Green + Blue	→	Cyan	Magenta + Yellow	→	Red
Blue + Red	→	Magenta	Yellow + Cyan	→	Green
Red + Green + Blue	→	White	Cyan + Magenta + Yellow	→	Black

Fig. 1.18 Colour Models

RGB color Model

Images represented in the RGB color model consist of three independent image planes, one for each primary color. The number of bits used to represent each pixel in RGB space is called the pixel depth. The term full-color image is used often to denote 24-bit RGB color image

- RGB model is based on a Cartesian coordinate system
- The color subspace of interest is the cube
- RGB values are at three corners
- Colors are defined by vectors extending from the origin
- For convenience, all color values have been normalized
- All values of R, G, and B are in the range [0, 1]

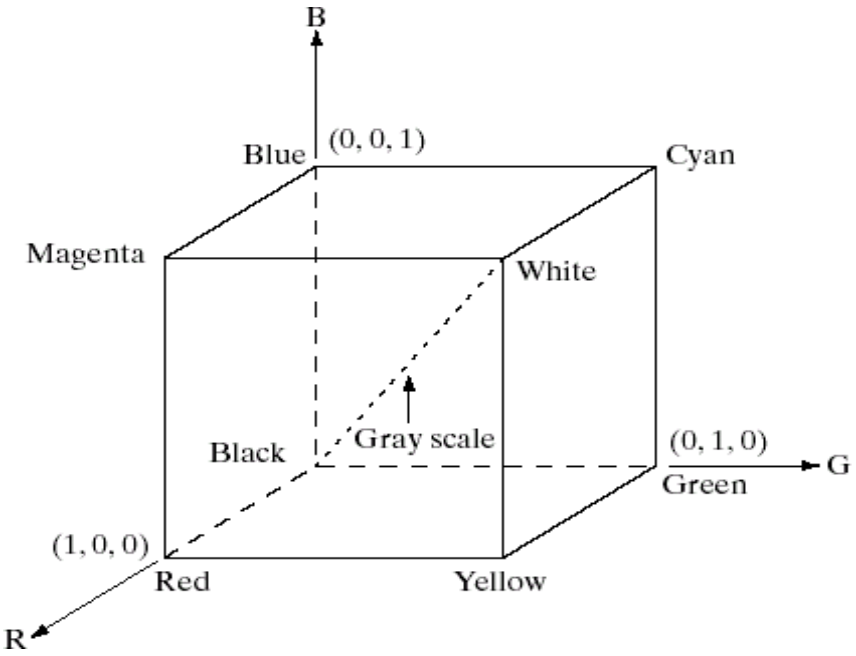


Fig. 1.19 RGB colour cube

CMY Colour Model

General purpose of CMY color model is to generate hardcopy output

- The primary colors of pigments
  - Cyan, Magenta, and Yellow
  - C = W - R, M = W - G, and Y = W - B
- Most devices that deposit colored pigments on paper require CMY data input
  - Converting RGB to CMY
  - The inverse operation from CMY to RGB is generally of no practical interest

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

HIS Colour Model

H : Hue

S : Saturation

I : Intensity

- The intensity is decoupled from the color information
- The hue and saturation are intimately related to the way in which human beings perceive color
- An ideal tool for developing image procession algorithms based on some of the color sensing properties of the human visual system

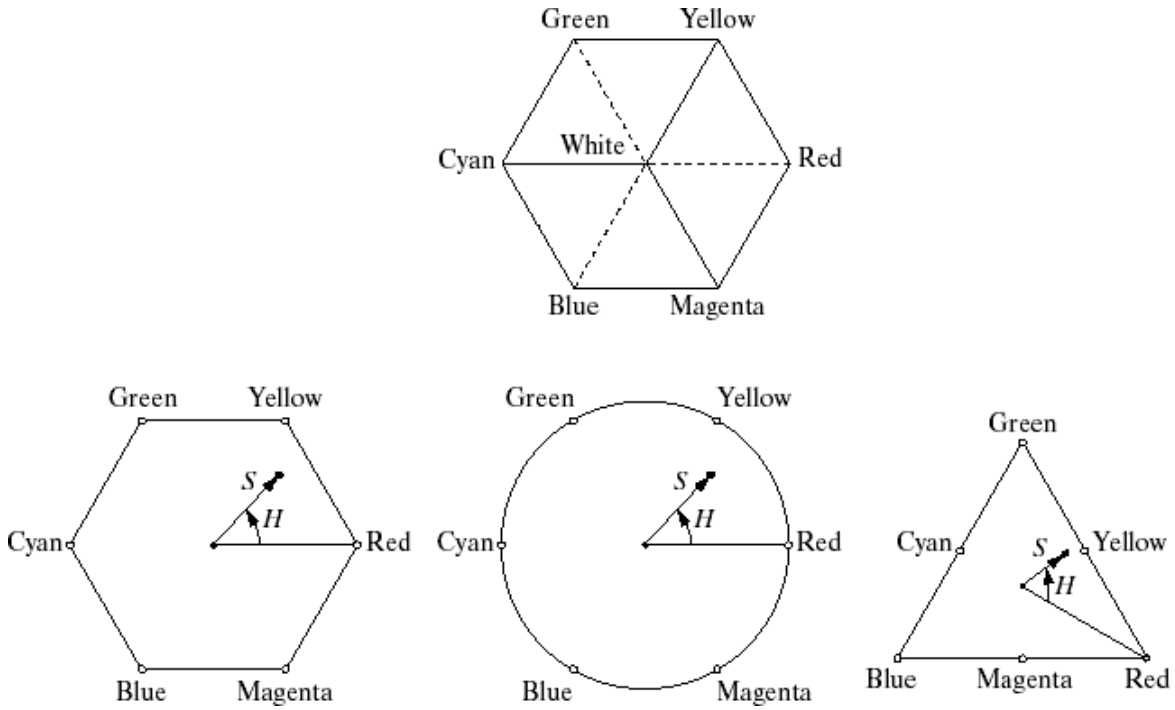


Fig. 1.20 Hue and Saturation in the HIS colour Model

Converting colors from RGB to HSI

- Hue component

$$H = \begin{cases} \theta & \text{if } B \le G \\ 360 - \theta & \text{if } B > G \end{cases} \qquad \theta = \cos^{-1}\left[\frac{\frac{1}{2}\{(R-G)+(R-B)\}}{\{(R- \quad ^2 \quad)^{1/2}} \right]$$

- Saturation component

$$S = 1 - \frac{3}{(R+G+B)}[\min(R, G, B)]$$

- Intensity component

$$I = \frac{1}{3}(R+G+B)$$

RGB values have been normalized to the range [0,1]

Angle θ is measured with respect to the red axis

Hue can be normalized to the range [0, 1] by dividing by 360°

Converting colors from HSI to RGB

- Three sectors of interest, corresponding to the 120^0 intervals in the separation of primaries

RG sector

$$(0^\circ \le H < 120^\circ)$$

$$B = I(1-S)$$
$$R = \left\lceil I\left[1 + \frac{S\cos H}{\cos(60^o - H)}\right]\right\rceil$$
$$G = 3I - (R+B)$$

GB sector

$$(120^\circ \le H < 240^\circ)$$

$$H = H - 120^o$$

$$R = I(1-S)$$
$$G = \left\lceil I\left[1 + \frac{S\cos H}{\cos(60^o - H)}\right]\right\rceil$$
$$B = 3I - (R+G)$$

BR sector

$$(240^\circ \le H < 360^\circ)$$

$$H = H - 240^o$$

$$G = I(1 - S)$$

$$B = \left\lceil I\left\lfloor 1 + \frac{S\cos H}{\cos(60^o - H)} \right\rfloor \right\rceil$$

$$R = 3I - (G + B)$$

TEXT / REFERENCE BOOKS

1. Rafael C. Gonzalez, Richard E. Woods, ŏ" Digital Image Processing", Pearson , Second Edition, 2004.
2. David Saloman,"Data compression" , Springer International, 4th Edition.
3. Khalid Sayood, "Introduction To Data Compressio" ,Elsevier 3rd Edition.
4. Ralfsteinmetz and Klara Nahrstedt, "Multimedia Computing, Communications & Applications" Pearson Edn
5. Rajan Parekh, "Principles of Multimedia, Tata Mc Graw Hill.
6. Anil K. Jain, "Fundamentals of Digital Image Processin", Pearson 2002.
7. J F Koegel Buford- -Multimedia systems Addison Wesley
8. T Vaughan-,"Multimedia: Making it work" Tata Mc Graw Hill

PART A

1. Define Image?
2. Distinguish between a monochrome and a grayscale image.
3. Define Image Sampling?
4. Define Quantization?
5. What do you meant by Gray level?
6. Define path.
7. Define Connectivity.
8. List the types of Connectivity.
9. What is Dynamic Range?
10. What do you meant by Color model?

PART B

1. Elaborate the fundamental steps in Digital Image Processing?
2. Describe the elements of visual perception with suitable diagram.
3. Explain the process of image acquisition.
4. Explain about image sampling and quantization process.
5. Explain about the Color Model?

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

UNIT – II – Digital Image and Multimedia Processing – SEC1605

II. Image Enhancement

Introduction; Point Processing - Image Negatives, Log transformations, Power Law Transformations, Piecewise-Linear Transformation Functions; Arithmetic/Logic Operations - Image Subtraction, Image Averaging; Histogram Processing - Histogram Equalization, Histogram Matching; Spatial filtering - Smoothing, Sharpening; Smoothing Frequency Domain Filters - Ideal Low Pass, Butterworth Low Pass, Gaussian Low Pass; Sharpening Frequency Domain Filters - Ideal High Pass, Butterworth High Pass, Gaussian High Pass.

Point Processing

Image enhancement is to process the given image such that the result is more suitable to process than the original image. It sharpens the image features such as edges, boundaries or contrast the image for better clarity. It does not increase the inherent information content of the data, but increase the dynamic range of feature chosen. The main drawback of image enhancement is quantifying the criterion for enhancement and therefore large number of image enhancement techniques is empirical and require interactive procedure to obtain satisfactory results. Point Processing is the image enhancement at any point in an image depends only on the gray level at that point. Some of the basic intensity transformation functions are

- **Linear Functions:**
  - Identity Transformation
  - Negative Transformation
- **Logarithmic Functions:**
  - Log Transformation
  - Inverse-log Transformation
- **Power-Law Functions:**
  - n^{th} power transformation
  - n^{th} root transformation
- **Piecewise Transformation function**
  - Contrast Stretching
  - Gray-level Slicing
  - Bit-plane slicing

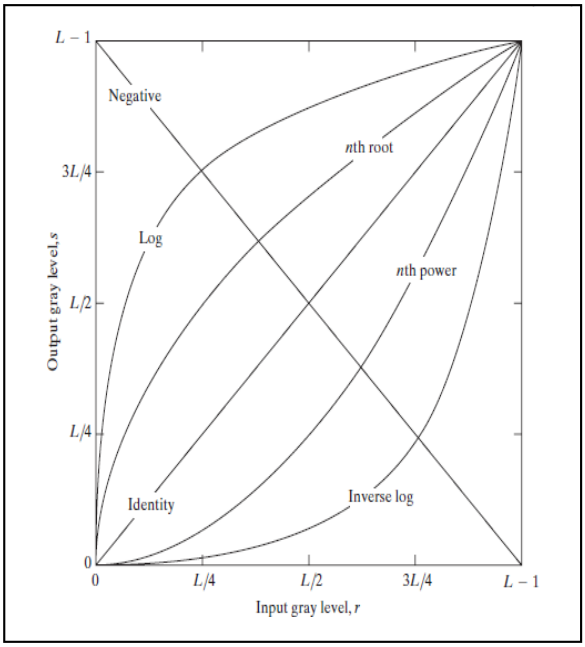


Fig. 2.1 Transformation Functions

Linear transformation

Linear transformation includes

- simple identity and
- negative transformation

- Identity transition is shown by a straight line
  - In this transition, each value of the input image is directly mapped to each other value of output image. That results in the same input image and output image. Hence is called identity transformation

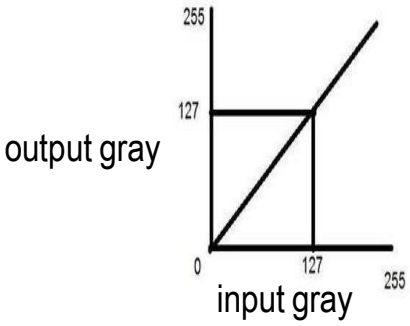
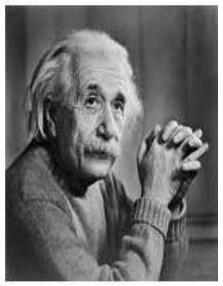


Fig. 2.2 Identity Transformation Function

Negative transformation

The negative of an image with gray level in the range **[0, L-1]** is obtained by using the negative transformation, the expression is given by **s = L-1-r**

- Reversing the intensity level of an image produces the equivalent of photographic negative
- Suitable for enhancing white or gray detail embedded in dark regions of an image, when the black areas are dominant in size



Input image

Output image

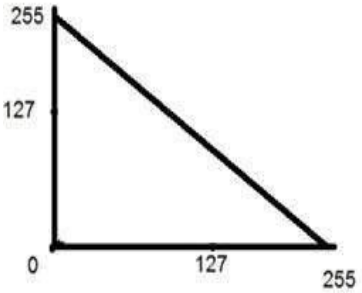
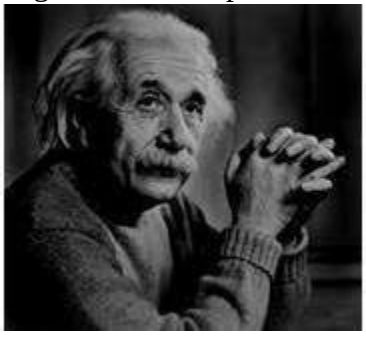


Fig. 2.2 Negative Transformation

Logarithmic transformations

The log transformations can be defined by $s = c \log(r + 1)$ where c is a constant and $r \geq 0$. During log transformation, the dark pixels in an image are expanded and the higher pixel values are compressed. The inverse log transform is opposite to log transform. Log transforms has the important characteristics: it compresses the dynamic range of images with large variation in pixel values.



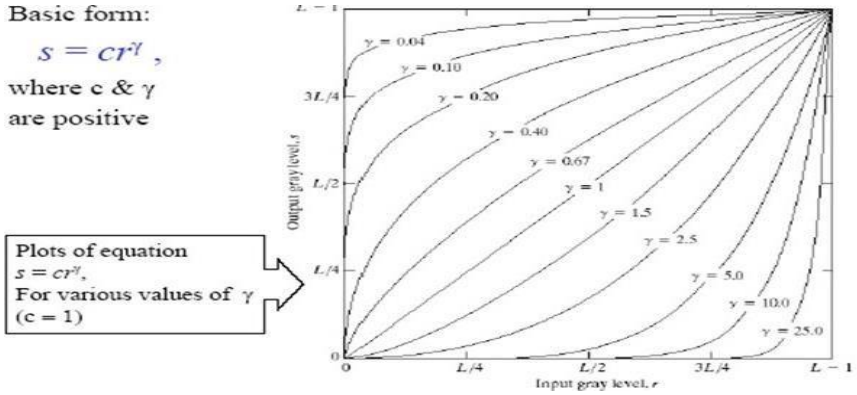
Input image

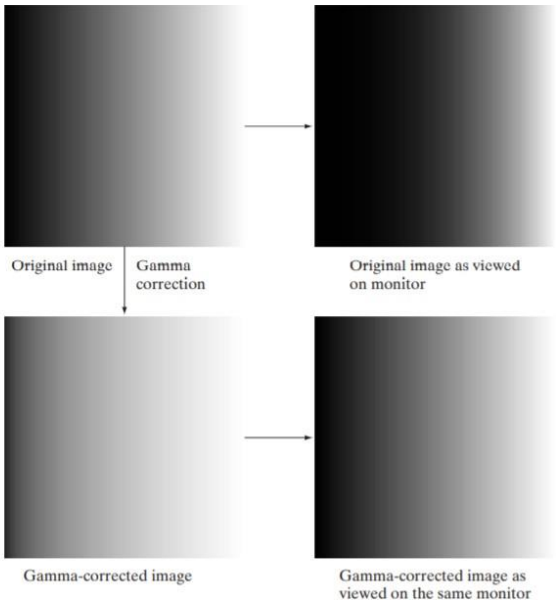
Output image

Fig. 2.3 Logarithmic Transformation

Power – Law transformations

This includes n^{th} power and n^{th} root transformations. It is given by the expression: $s=c\ r^{\gamma}$ (or) $s=c(r+\varepsilon)^{\gamma}$ where γ is called gamma, due to which this transformation is also known as gamma transformation. The exponent in the power law equation is referred to as gamma, the process used to correct this power-law response phenomenon is called gamma correction.





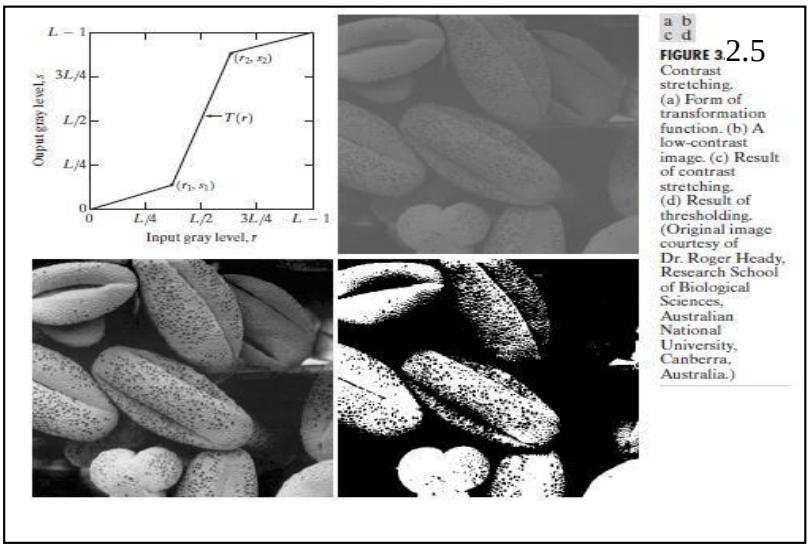
a b
c d

FIGURE 2.4
(a) Intensity ramp image. (b) Image as viewed on a simulated monitor with a gamma of 2.5. (c) Gamma-corrected image. (d) Corrected image as viewed on the same monitor. Compare (d) and (a).

Piecewise- Linear Transformation

One of the simplest piecewise linear functions is a contrast-stretching transformation, which is used to enhance the low contrast images. Low contrast images may result from poor illumination and wrong setting of lens aperture during image acquisition.

Contrast stretching



a b
c d
FIGURE 3.2.5 Contrast stretching. (a) Form of transformation function. (b) A low-contrast image. (c) Result of contrast stretching. (d) Result of thresholding. (Original image courtesy of Dr. Roger Heady, Research School of Biological Sciences, Australian National University, Canberra, Australia.)

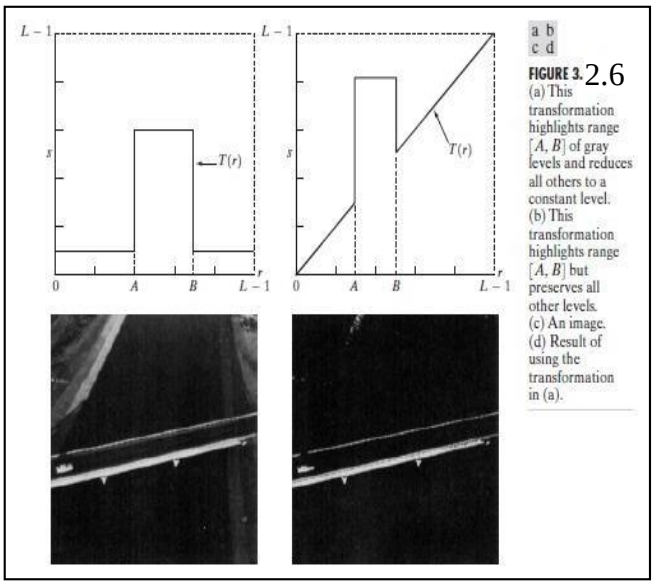
Figure shows a typical transformation used for contrast stretching. The locations of points (r1, s1) and (r2, s2) control the shape of the transformation function. If r1 = s1 and r2 = s2, the transformation is a linear function that produces no changes in gray levels. If r1 = r2, s1 = 0 and s2 = L-1, the transformation becomes a thresholding function that creates a binary image. Intermediate values of (r1, s1) and (r2, s2) produce various degrees of spread in the gray levels of the output image, thus affecting its contrast. In general, $r1 \leq r2$ and $s1 \leq s2$ is assumed, so the function is always increasing. Figure (b) shows an 8-bit image with low contrast. Fig. (c) shows the result of contrast stretching, obtained by setting (r1, s1) = (r_{min}, 0) and (r2, s2) = (r_{max},L-1) where r_{min} and r_{max} denote the minimum and maximum gray levels in the image, respectively. Thus, the transformation function stretched the levels linearly from their original range to the full range [0, L-1]. Finally, Fig. (d) shows the result of using the thresholding function defined previously, with r1=r2=m, the mean gray level in the image.

Gray-level Slicing

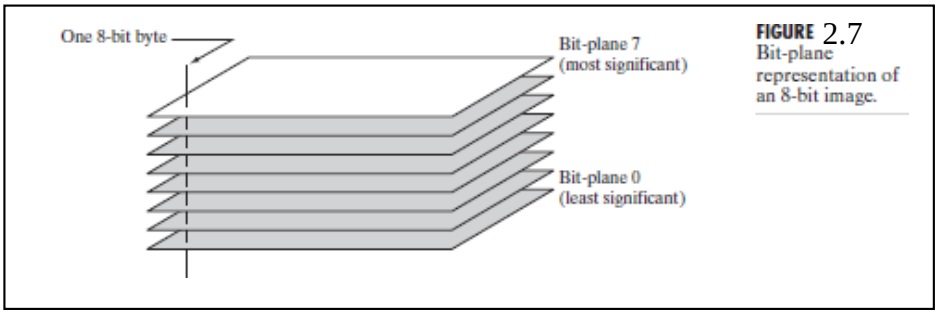
This technique is used to highlight a specific range of gray levels. It can be implemented in several ways, but the two basic themes are:

One approach is to display a high value for all gray levels in the range of interest and a low value for all other gray levels. This transformation, shown in Fig. (a), produces a binary image. The second approach, based on the transformation shown in Fig. (b), this brightens the desired range of gray levels but preserves gray levels unchanged Fig.(c)

shows a gray scale image, and fig.(d) shows the result of using the transformation in Fig.(a).



Bit-plane Slicing



Pixels are digital numbers, each one composed of bits. Instead of highlighting gray-level range, we could highlight the contribution made by each bit. This method is useful and used in image compression. Most significant bits contain the majority of visually significant data.

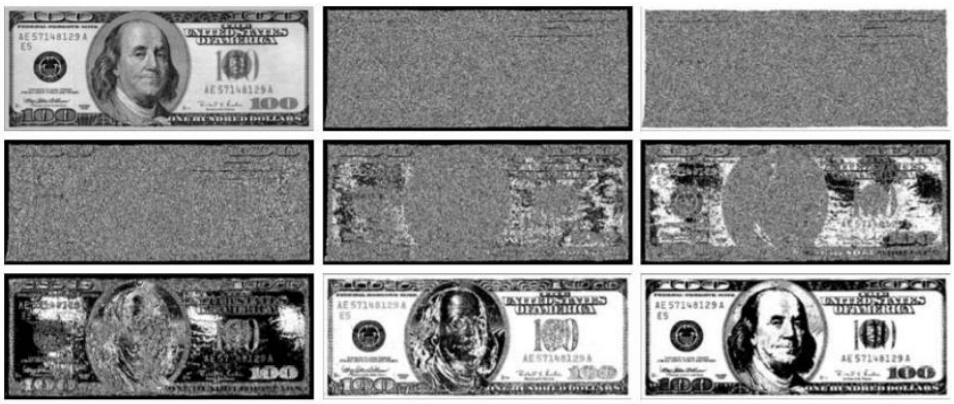


Fig. 2.8 Example for bit plane slicing

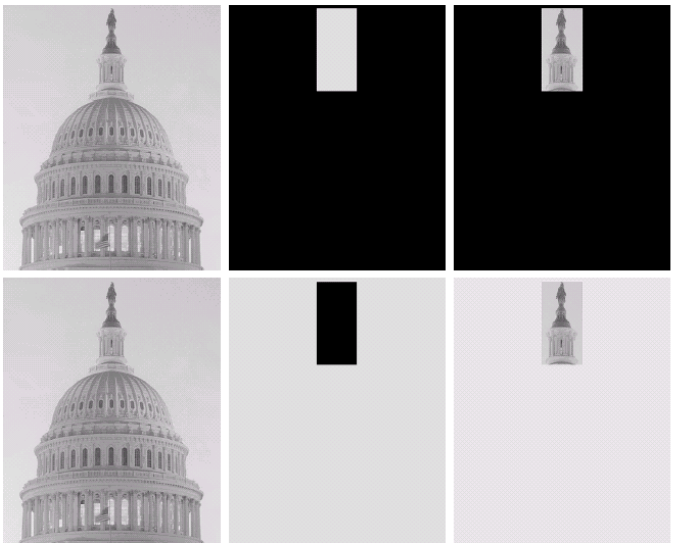
Arithmetic/Logic Operations

Image arithmetic applies one of the standard arithmetic operations or a logical operator to two or more images. The operators are applied in a pixel-by-pixel way, i.e. the value of a pixel in the output image depends only on the values of the corresponding pixels in the input images. Hence, the images must be of the same size. Although image arithmetic is the most simple form of image processing, there is a wide range of applications.

Logical operators are often used to combine two (mostly binary) images. In the case of integer images, the logical operator is normally applied in a bitwise way.

$$s(x, y) = f(x, y) + g(x, y)$$
$$d(x, y) = f(x, y) - g(x, y)$$
$$p(x, y) = f(x, y) \times g(x, y)$$
$$v(x, y) = f(x, y) \div g(x, y)$$

Arithmetic/logical operations are performed on pixel-by-pixel basis based on two or more images. When dealing with logical operations on gray-scale images, pixel values are processed as strings of binary numbers. In AND and OR image masks, light represents a binary 1 and dark represents a binary 0. Masking refers to as Region of Interest (ROI) processing.



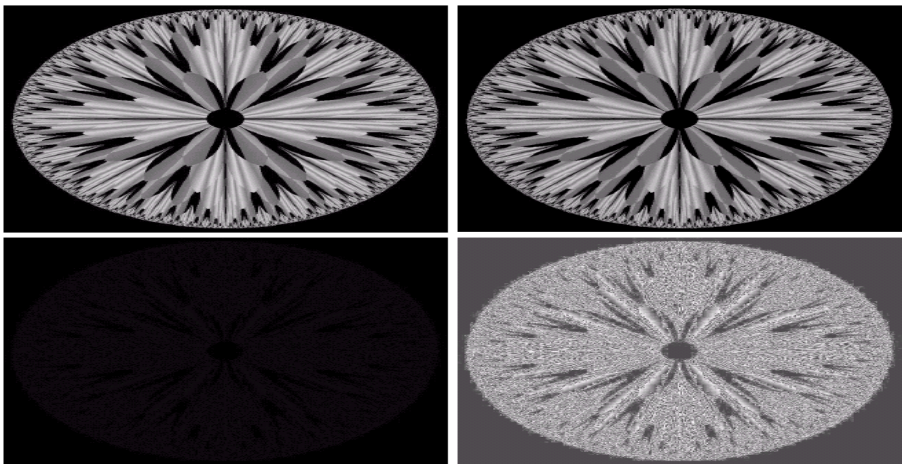
a b c
d e f

FIGURE 2.9 (a) Original image. (b) AND image mask. (c) Result of the AND operation on images (a) and (b). (d) Original image. (e) OR image mask. (f) Result of operation OR on images (d) and (e).

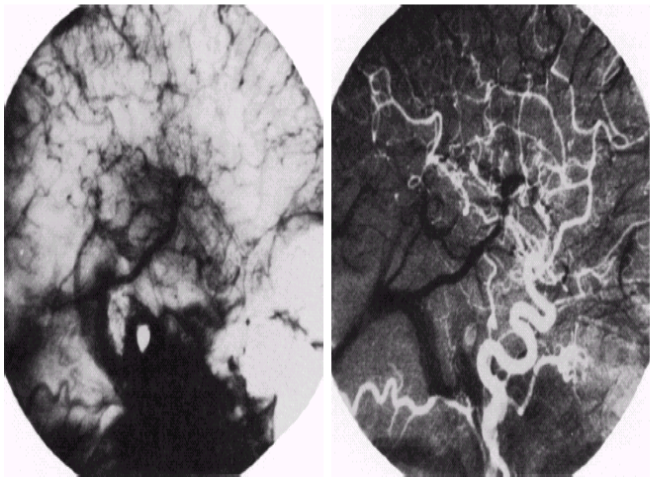
Image Subtraction

- Enhancement of differences between images
- Key usefulness of subtraction is the enhancement of differences between images.
- If the difference in the pixel value is small, then the image appears black when displayed in 8-bit display.
- To bring more detail contrast stretching can be performed

$$g(x, y) = f(x, y) - h(x, y)$$



a b
c d
FIGURE 2.10 (a) Original fractal image. (b) Result of setting the four lower-order bit planes to zero. (c) Difference between (a) and (b). (d) Histogram-equalized difference image. (Original image courtesy of Ms. Melissa D. Binde, Swarthmore College, Swarthmore, PA).



a b
FIGURE 2.11 Enhancement by image subtraction. (a) Mask image. (b) An image (taken after injection of a contrast medium into the bloodstream) with mask subtracted out.

Image Averaging

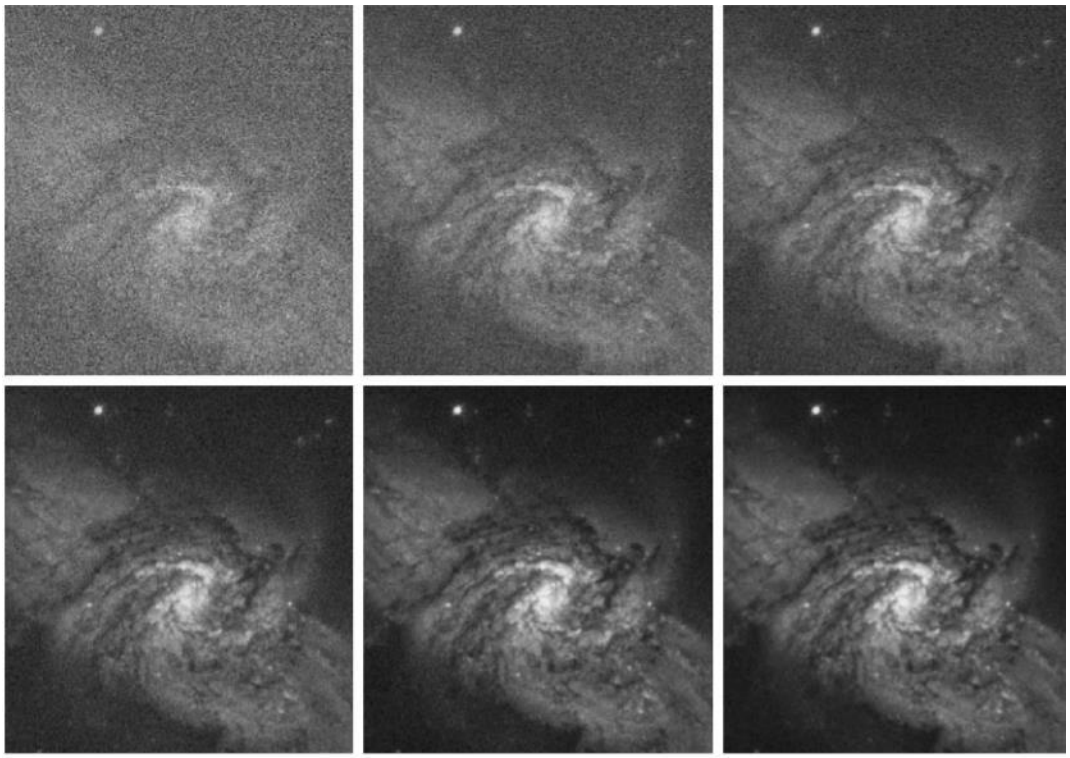
- Noisy image g(x,y) formed by the addition of noise

$$g(x, y) = f(x, y) + \eta(x, y)$$

- Averaging K different noisy images η(x,y) to an original image f(x,y)
- Objective is to reduce the noise content by adding a set of noisy images $\{g_i(x,y)\}$

$$\bar{g}(x,y) = \frac{1}{K}\sum_{i=1}^{K} g_i(x,y) \qquad E\{\bar{g}(x,y)\} = f(x,y)$$

$$\sigma^2_{\bar{g}(x,y)} = \frac{1}{K}\sigma^2_{\eta(x,y)} \qquad \sigma_{\bar{g}(x,y)} = \frac{1}{\sqrt{K}}\sigma_{\eta(x,y)}$$



a b c
d e f

FIGURE 2.12 (a) Image of Galaxy Pair NGC 3314 corrupted by additive Gaussian noise. (b)–(f) Results of averaging 5, 10, 20, 50, and 100 noisy images, respectively. (Original image courtesy of NASA.)

Histogram Processing

- Histogram is a graphical representation showing a visual impression of the distribution of data
- An Image Histogram is a type of histogram that acts as a graphical representation of the lightness/color distribution in a digital image
- It plots the number of pixels for each value

- The histogram of a digital image with gray levels in the range [0, L-1] is a discrete function $h(r_k) = n_k$ where r_k is the k^{th} gray level and n_k is the number of pixels in the image having gray level r_k

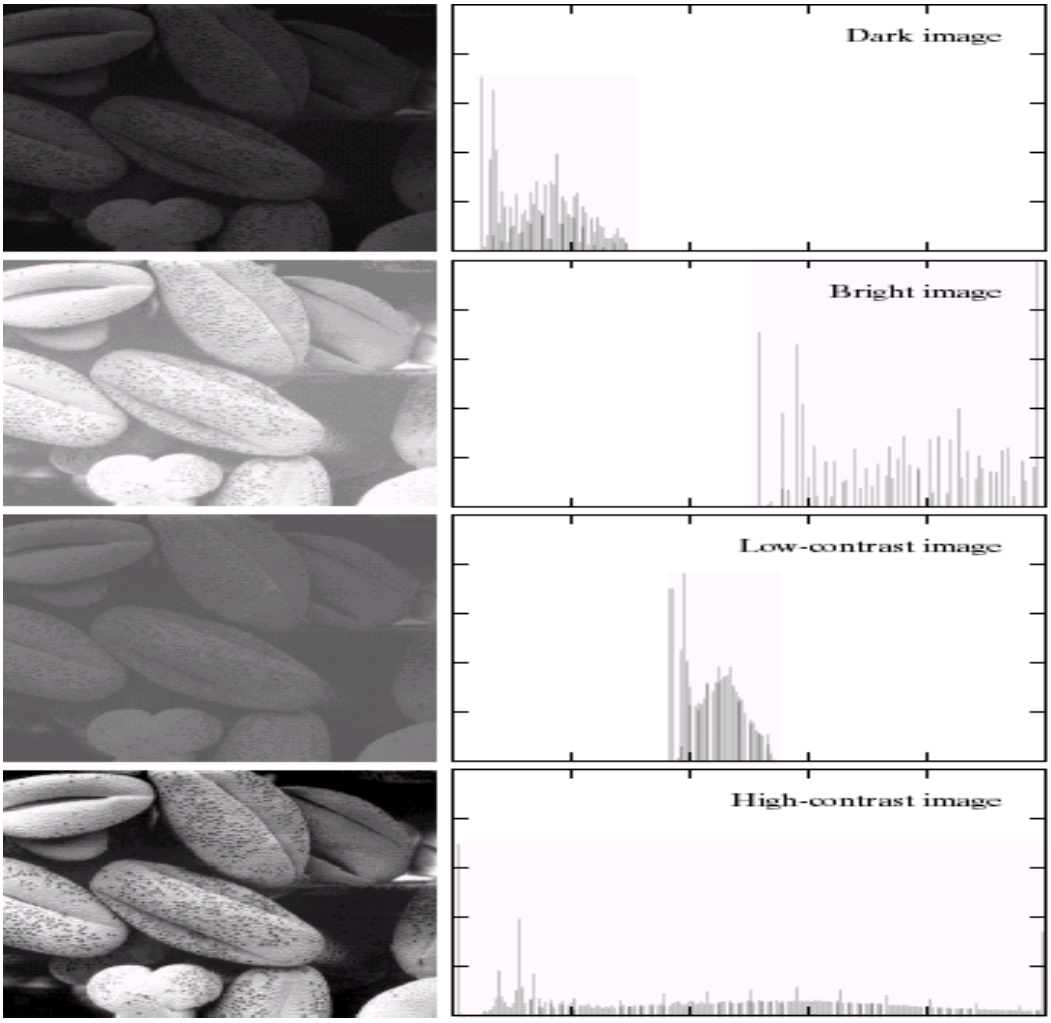


Fig. 2.13 Histogram of different types of images

It is common practice to normalize a histogram by dividing each of its values by the total number of pixels in the image, denoted by n.

A normalized histogram is given by

$$p(r_k) = n_k / n \qquad \text{for } k = 0, 1, \ldots, L-1$$

Thus, $p(r_k)$ gives an estimate of the probability of occurrence of gray level r_k

Note:

The sum of all components of a normalized histogram is equal to 1

Histogram Equalisation

The gray levels in an image is viewed as random variables in the interval[0,1]

Fundamental descriptors of a random variables is its probability density function

$p_s(s)$ and $p_s(r)$ are PDF of s and r

$$p_s(s) = p_r(r)\left|\frac{dr}{ds}\right|$$

Transformation has the particular importance in image processing

$$s = T(r) = \int_0^r p_r(\omega)d\omega$$

Discrete version of transformation- histogram equalization or histogram linearization

$$s_k = T(r_k) = \sum_{j=0}^{k}\frac{n_j}{n} = \sum_{j=0}^{k} p_r(r_j)$$

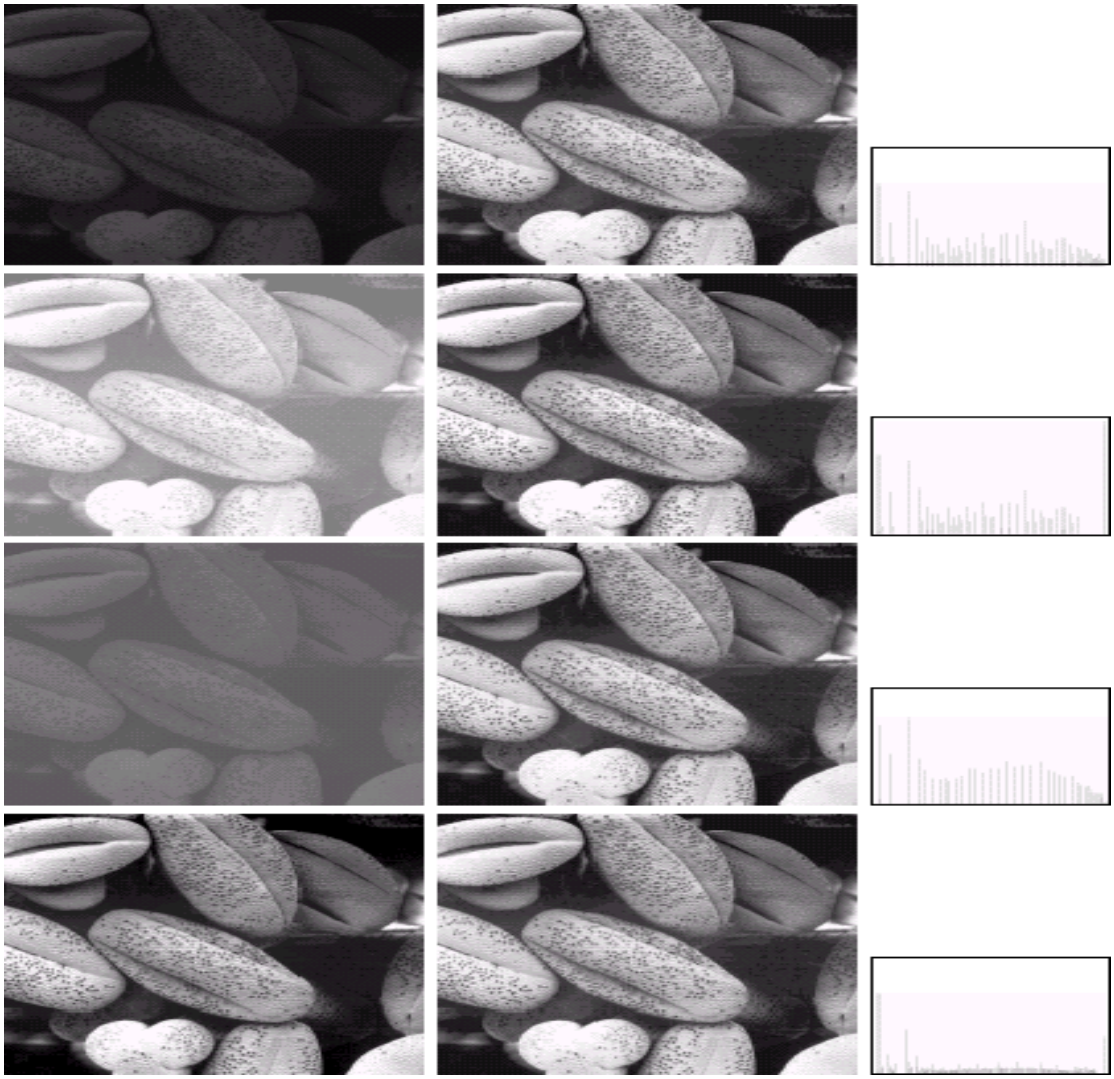


Fig. 2.14 Images after Histogram equalization

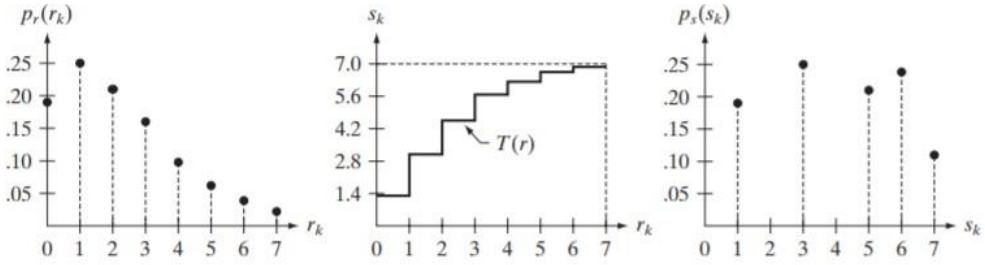
Assume the images have 64 x 64 = 4096 pixels in 8 gray levels. The following table shows the equalization process

Original Image Gray Level	No. of pixels (frequency)	Probability	Cumulative Probability	Multiply by Max. Gray Level	Rounding
0	790	0.19	0.19	1.33	1
1	1023	0.25	0.44	3.08	3

2	850	0.21	0.65	4.55	4
3	656	0.16	0.81	5.67	5
4	329	0.08	0.89	6.23	6
5	245	0.06	0.95	6.65	6
6	122	0.03	0.98	6.86	6
7	81	0.02	1.00	7	7

TABLE 2.1
Intensity distribution and histogram values for a 3-bit, 64 × 64 digital image.

r_k	n_k	$p_r(r_k) = n_k/MN$
$r_0 = 0$	790	0.19
$r_1 = 1$	1023	0.25
$r_2 = 2$	850	0.21
$r_3 = 3$	656	0.16
$r_4 = 4$	329	0.08
$r_5 = 5$	245	0.06
$r_6 = 6$	122	0.03
$r_7 = 7$	81	0.02



a b c

FIGURE 2.15 Illustration of histogram equalization of a 3-bit (8 intensity levels) image. (a) Original histogram. (b) Transformation function. (c) Equalized histogram.

- r is in the range with representing black and representing white
- For r satisfying these conditions, we focus attention on transformations (intensity mappings) of the form

$$s = T(r) \quad 0 \le r \le L - 1 \tag{3.3-1}$$

that produce an output intensity level s for every pixel in the input image having intensity r. We assume that:

(a) $T(r)$ is a monotonically[†] increasing function in the interval $0 \le r \le L - 1$; and

(b) $0 \le T(r) \le L - 1$ for $0 \le r \le L - 1$.

In some formulations to be discussed later, we use the inverse

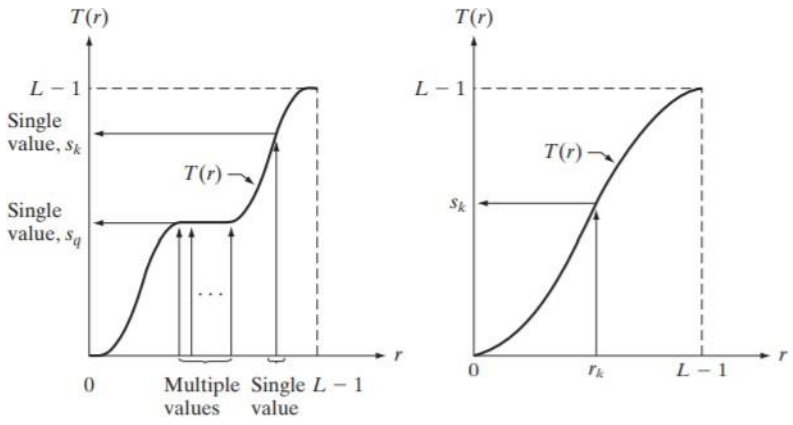
$$r = T^{-1}(s) \quad 0 \le s \le L - 1 \tag{3.3-2}$$

in which case we change condition (a) to

(a′) $T(r)$ is a strictly monotonically increasing function in the interval $0 \le r \le L - 1$.

a b

FIGURE 2.16 (a) Monotonically increasing function, showing how multiple values can map to a single value. (b) Strictly monotonically increasing function. This is a one-to-one mapping, both ways.



Histogram Matching (Specification)

Procedure for histogram matching:

- Obtain histogram of given image
- Use the equation to pre compute a mapped level s_k for each level r_k

$$s_k = T_r(k) = \sum_{j=0}^{k} p_r(r_j) = \sum_{j=0}^{k} n_j / n$$

- Obtain the transformation function G from the given $p_z(z)$ using the equation

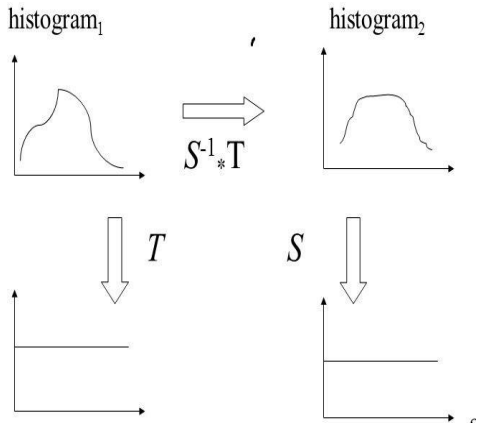
$$(G(\hat{z}) - sk) \ge 0; k = 0,1,2 \ \ldots L-1$$

- Precompute z_k for each value of s_k using the iterative scheme defined in connection with equation

$$(G(z\hat{}) - sk) \geq 0; k = 0,1,2 \ldots L-1$$

- For each pixel in the original image, if the value of that pixel is r_k, map this value to the corresponding level s_k; then map levels s_k into the final level z_k

Processed image that has a specified histogram is called histogram matching or histogram specification.



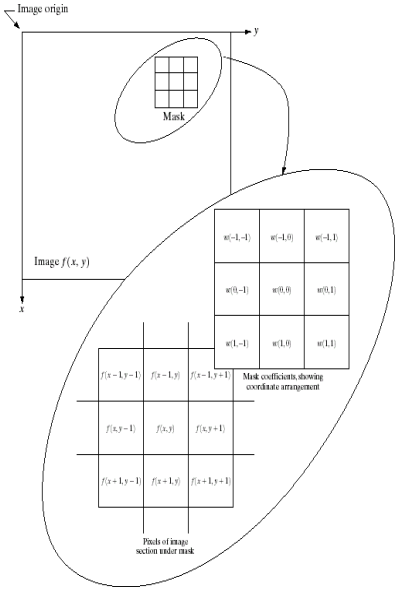
Spatial Filtering

- The output intensity value at (x,y) depends not only on the input intensity value at (x,y) but also on the specified number of neighboring intensity values around (x,y)
- Spatial masks (also called window, filter, kernel, template) are used and convolved over the entire image for local enhancement (spatial filtering)
- The size of the masks determines the number of neighboring pixels which inf luence the output value at (x,y)
- The values (coefficients) of the mask determine the nature and properties of enhancing technique.
- The mechanics of spatial filtering
- For an image of size *M* x *N* and a mask of size *m* x *n*
- The resulting output gray level for any coordinates x and y is given by

$$g(x, y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t)\, f(x+s, y+t)$$

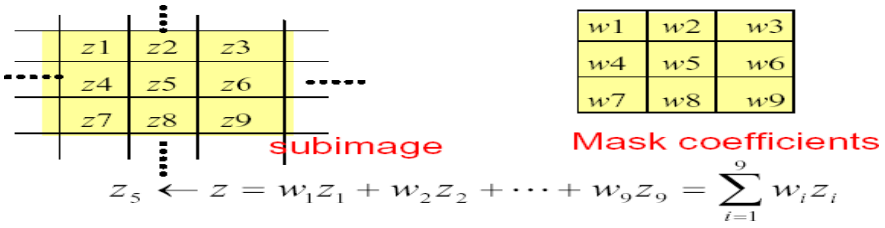
where $a = (m-1)/2, \quad b = (n-1)/2$

$x = 0, 1, 2, \ldots\ldots, M-1, \quad y = 0, 1, 2, \ldots\ldots, N-1,$



Given the 3×3 mask with coefficients: w_1, w_2,…, w_9

The mask cover the pixels with gray levels: z_1, z_2,…, z_9



z gives the output intensity value for the processed image (to be stored in a new array) at the location of z_5 in the input image

Mask operation near the image border

Problem arises when part of the mask is located outside the image plane; to handle the problem:

- Discard the problem pixels (e.g. $512x512_{input}$ $510x510_{output}$ if mask size is 3x3)
- Zero padding: expand the input image by padding zeros ($512x512_{input}$ $514x514_{output}$)
- Zero padding is not good - create artificial lines or edges on the border

- We normally use the gray levels of border pixels to fill up the expanded region (for 3x3 mask). For larger masks a border region equal to half of the mask size is mirrored on the expanded region.

Spatial Filtering for Smoothing

- For blurring/noise reduction;
- Smoothing/Blurring is usually used in preprocessing steps,

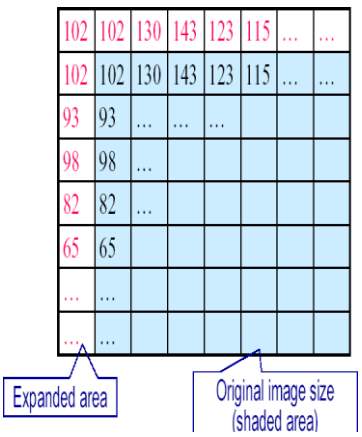
e.g., to remove small details from an image prior to object extraction, or to bridge small gaps in lines or curves

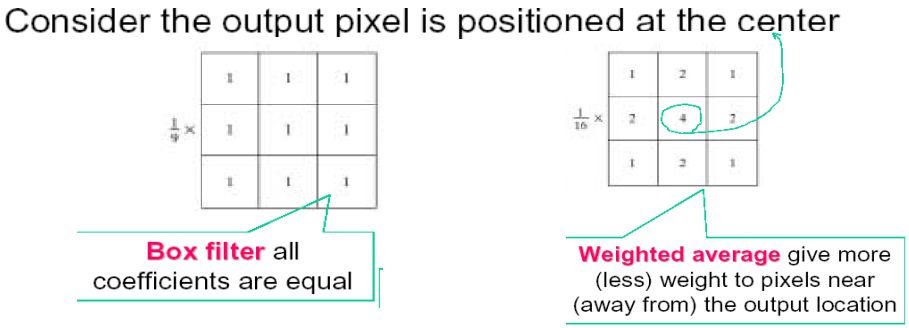
- Equivalent to Low-pass spatial filtering in frequency domain because smaller (high frequency) details are removed based on neighborhood averaging (averaging filters)

Implementation: The simplest form of the spatial filter for averaging is a square mask (assume $m \times m$ mask) with the same coefficients $1/m^2$ to preserve the gray levels (averaging).

Applications: Reduce noise; smooth false contours

Side effect: Edge blurring





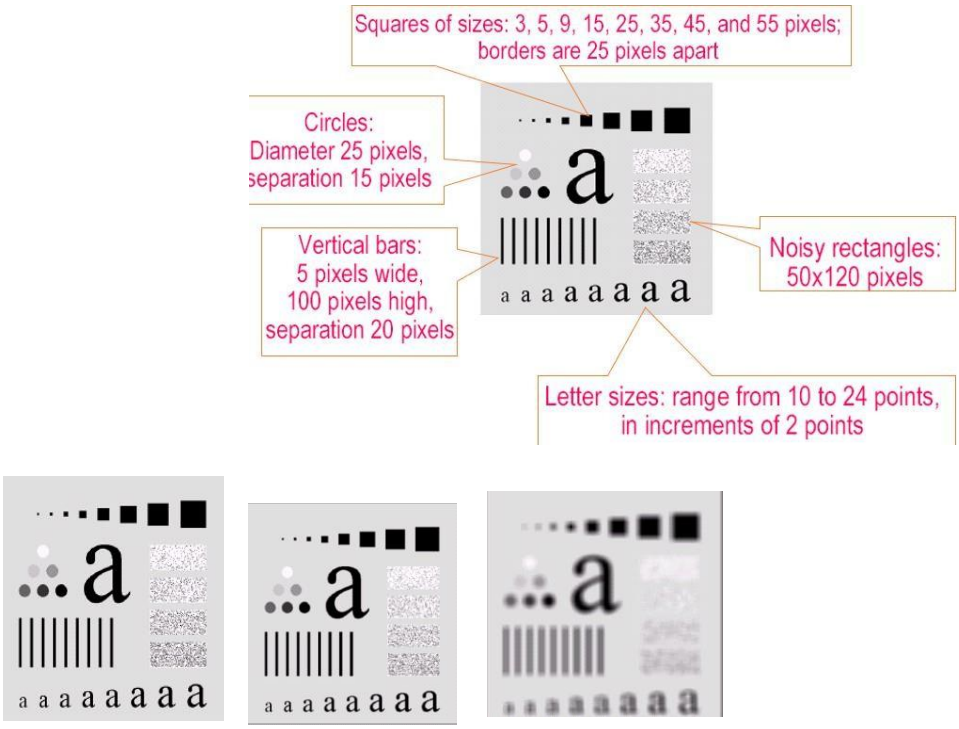


Fig. 2.17 Spatial filtering

Spatial Filtering for Sharpening

Background: to highlight fine detail in an image or to enhance blurred detail

Applications: electronic printing, medical imaging, industrial inspection, autonomous target detection (smart weapons)

Foundation:

- Blurring/smoothing is performed by spatial averaging (equivalent to integration)
- Sharpening is performed by noting only the gray level changes in the image that is the differentiation

Operation of Image Differentiation

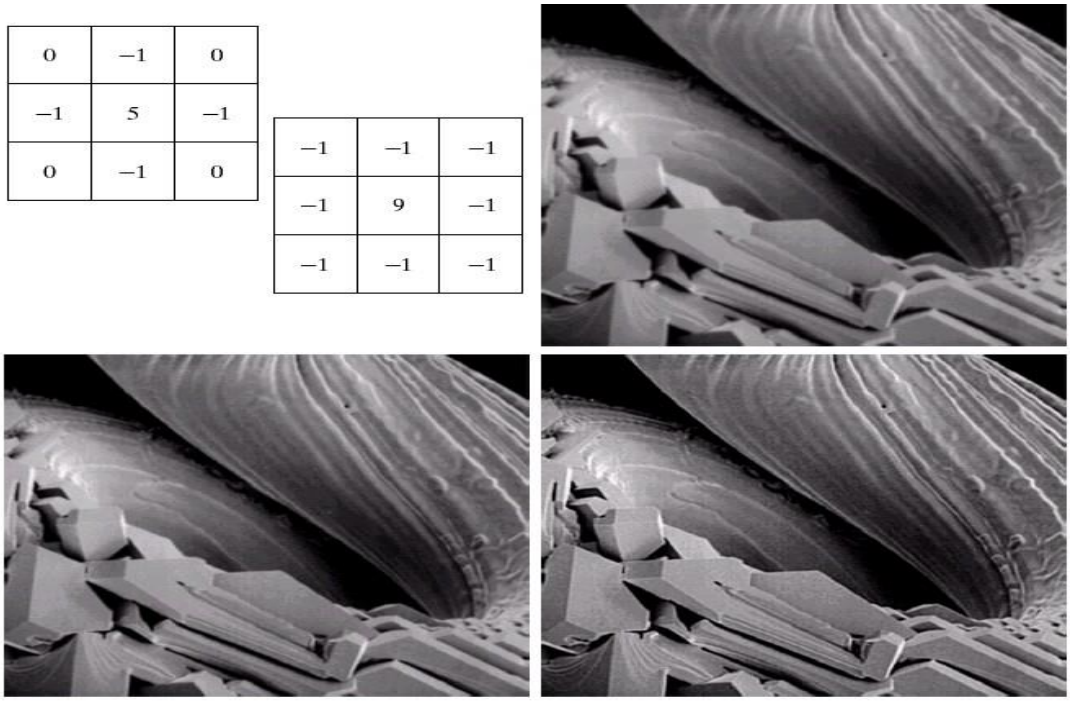
- Enhance edges and discontinuities (magnitude of output gray level >>0)
- De-emphasize areas with slowly varying gray-level values (output gray level: 0)

Mathematical Basis of Filtering for Image Sharpening

- First-order and second-order derivatives
- Approximation in discrete-space domain
- Implementation by mask filtering

0	−1	0
−1	5	−1
0	−1	0

−1	−1	−1
−1	9	−1
−1	−1	−1



a b c
d e

FIGURE 2.18 (a) Composite Laplacian mask. (b) A second composite mask. (c) Scanning electron microscope image. (d) and (e) Results of filtering with the masks in (a) and (b), respectively. Note how much sharper (e) is than (d). (Original image courtesy of Mr. Michael Shaffer, Department of Geological Sciences, University of Oregon, Eugene.)

Frequency Domain Filters

- Any function that periodically repeats itself can be expressed as the sum of sines and/or cosines of different frequencies, each multiplied by a different coefficient (Fourier series).
- Even functions that are not periodic (but whose area under the curve is finite) can be expressed as the integral of sines and/or cosines multiplied by a weighting function (Fourier transform).
- The **frequency domain** refers to the plane of the two dimensional discrete Fourier transform of an image.
- The purpose of the Fourier transform is to represent a signal as a linear combination of sinusoidal signals of various frequencies.

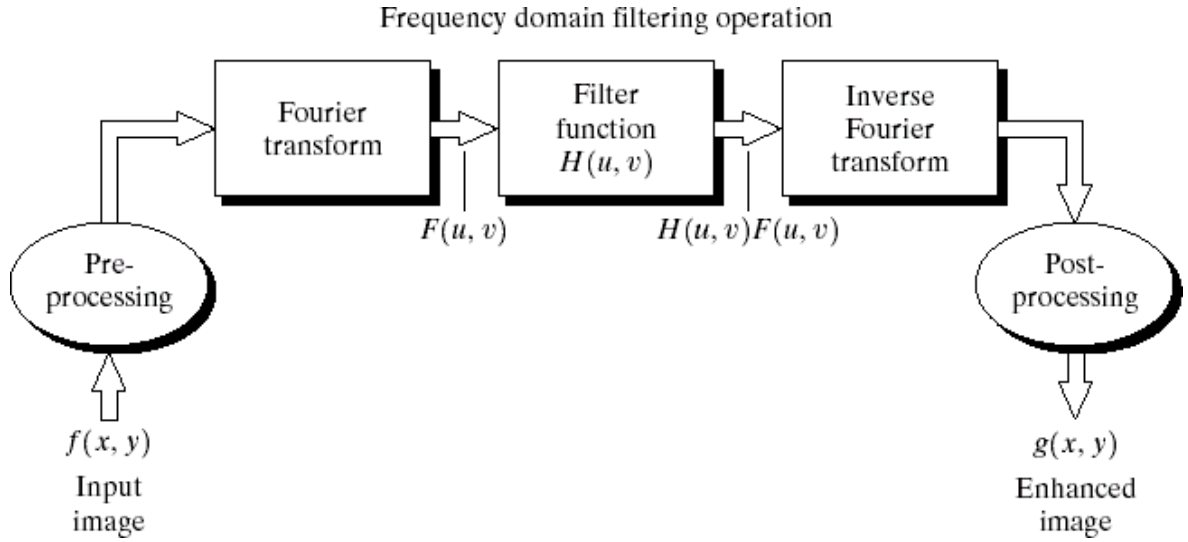


Fig. 2.19 Frequency domain operations

Frequency Domain Filters - Smoothing

Ideal Low Pass Filter

$$H(u,v) = \begin{cases} 1 & \text{if } D(u,v) \le D_0 \\ 0 & \text{if } D(u,v) \ge D_0 \end{cases}$$

where D(u,v) is the distance to the center freq.

$$D(u,v) = [(u - M/2)^2 + (v - N/2)^2]^{1/2}$$

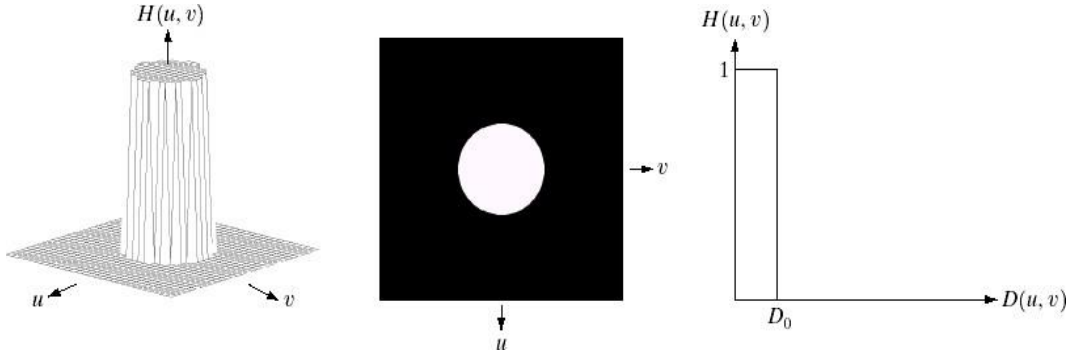


Fig. 2.20 Ideal Low Pass filter

Butterworth Low Pass Filter

$$H(u,v) = \frac{1}{1+[D(u,v)/D_0]^{2n}}$$

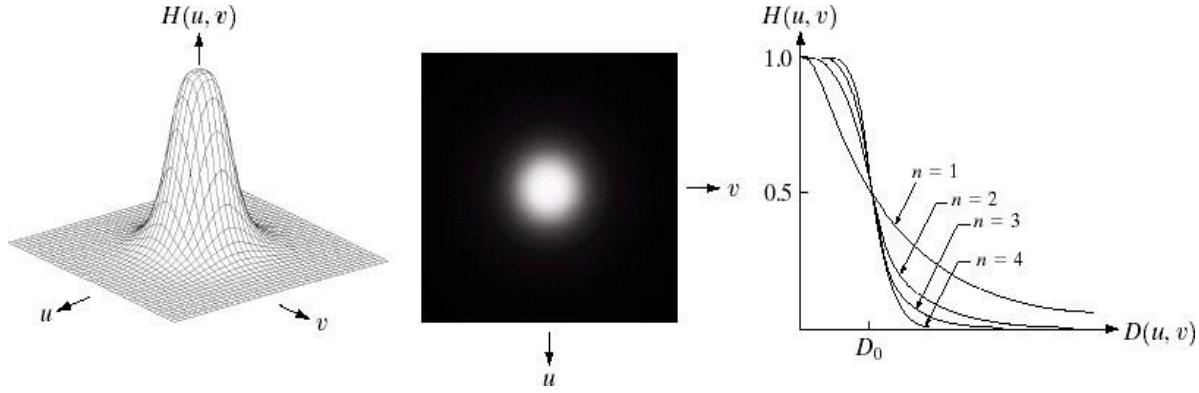


Fig. 2.21 Butterworth Low Pass filter

Gaussian Low Pass Filter

$$H(u,v) = e^{-D^2(u,v)/2D_0^2}$$

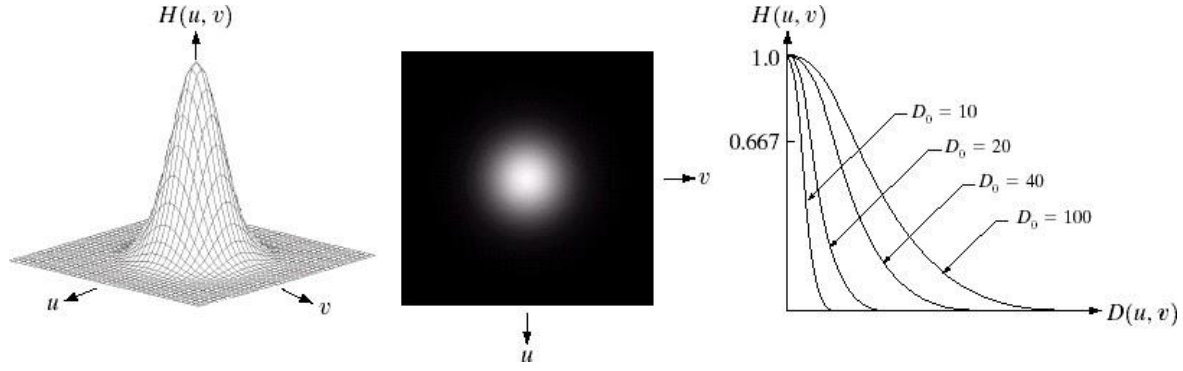


Fig. 2.22 Gaussian Low Pass filter

Frequency Domain Filters - Sharpening

- Image details corresponds to high-frequency
    - Sharpening: high-pass filters
    - $H_{hp}(u,v)=1-H_{lp}(u,v)$

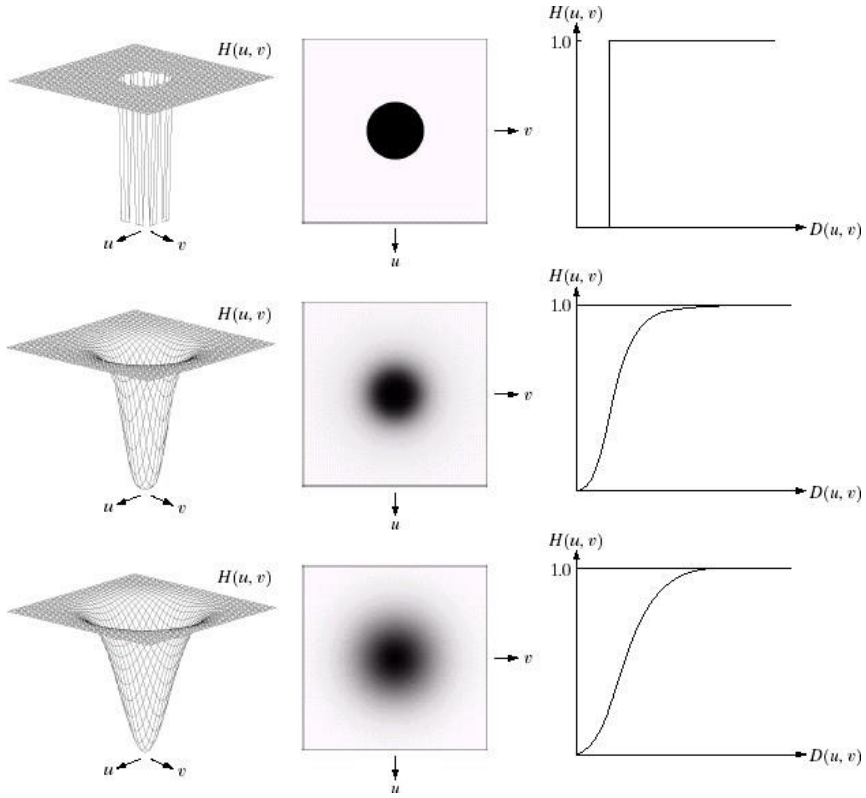


Fig. 2.23 High Pass filter a) Ideal b) Butterworth c) Gaussian

TEXT / REFERENCE BOOKS

1. Rafael C. Gonzalez, Richard E. Woods, ŏ" Digital Image Processing", Pearson , Second Edition, 2004.
2. David Saloman,"Data compression" , Springer International, 4th Edition.
3. Khalid Sayood, "Introduction To Data Compressio" ,Elsevier 3rd Edition.
4. Ralfsteinmetz and Klara Nahrstedt, "Multimedia Computing, Communications & Applications" Pearson Edn
5. Rajan Parekh, "Principles of Multimedia, Tata Mc Graw Hill.
6. Anil K. Jain, "Fundamentals of Digital Image Processin", Pearson 2002.
7. J F Koegel Buford- -Multimedia systems Addison Wesley
8. T Vaughan-,"Multimedia: Making it work" Tata Mc Graw Hill

PART A

1. Specify the objective of image enhancement technique.
2. What is contrast stretching?
3. What is grey level slicing?
4. Define image subtraction.
5. What is the purpose of image averaging

6. What is meant by masking?

7. Give the formula for negative and log transformation.

8. What is meant by bit plane slicing?

9. Define histogram.

10. What is image filtering?

PART B

1. Explain the histogram equalization method of image enhancement.
2. How color image is enhanced and compare it with gray scale processing?
3. What is meant by Frequency Filtering? Discuss in detail about Smoothing frequency Filtering?
4. Discuss in detail about Sharpening frequency Filtering?

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

UNIT – III – Digital Image and Multimedia Processing – SEC1605

III. Morphological Image Processing & Segmentation

Morphological Image Processing - Logic Operations involving Binary Images; Dilation and Erosion; Opening and Closing; Basic Morphological Algorithms - Boundary Extraction, Region Filling, Thickening, Thinning; Image Segmentation - Detection of Discontinuities; Edge Linking; Boundary Detection; Thresholding - Global and Adaptive; Region based Segmentation.

Logic Operations involving Binary Images

Mathematical Morphology is based on the algebra of non-linear operators operating on object shape and in many respects supersedes the linear algebraic system of convolution. It performs in many tasks – pre-processing, segmentation using object shape, and object quantification – better and more quickly than the standard approach. Mathematical morphology tool is different from the usual standard algebra and calculus. Morphology tools are implemented in most advanced image analysis.

Mathematical morphology is very often used in applications where shape of objects and speed is an issue—example: analysis of microscopic images, industrial inspection, optical character recognition, and document analysis. The non-morphological approach to image processing is close to calculus, being based on the point spread function concept and linear transformations such as convolution. Mathematical morphology uses tools of non-linear algebra and operates with point sets, their connectivity and shape. Morphology operations simplify images, and quantify and preserve the main shape characteristics of objects. Morphological operations are used for the following purpose:

- Image pre-processing (noise filtering, shape simplification)
- Enhancing object structure (skeleton zing, thinning, thickening, convex hull, object marking)
- Segmenting objects from the background
- Quantitative description of objects (area, perimeter, projections, Euler-Poincare characteristics)

Mathematical morphology exploits point set properties, results of integral geometry, and topology. The real image can be modelled using point sets of any dimension; the Euclidean 2D space and its system of subsets is a natural domain for planar shape description.

Computer vision uses the digital counterpart of Euclidean space – sets of integer pairs (∈) f or binary image morphology or sets of integer triples (∈) for gray-scale morphology or binary 3D morphology. Discrete grid can be defined if the neighbourhood relation between points is well defined. This representation is suitable for both rectangular and hexagonal grids. A morphological transformation is given by the relation of the image with another small point set B called structuring element. B is expressed with respect to a local origin. Structuring element is a small image-used as a moving window-- whose support delineates pixel neighbourhoods in the image plane. It can be of any shape, size, or connectivity (more than 1 piece, have holes). To apply the morphologic transformation () to the image means that the structuring element B is moved systematically across the entire image. Assume that B is positioned at some point in the image; the pixel in the image corresponding to the representative point O of the structuring element is called the current pixel. The result of the relation between the image X and the structuring element B in the current position is stored in the output image in the current image pixel position.

- **Reflection**

  $\hat{B} = \{w | w = -b, \text{for } b \in B\}$

- **Translation**

  $(B)_Z = \{c | c = b + z, \text{for } b \in B\}$

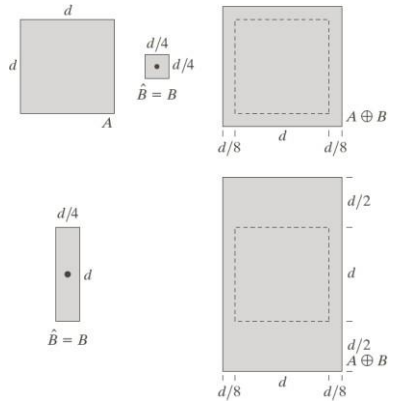
Dilation and Erosion

Dilation

With A and B as sets in Z^2, the dilation of A by B, denoted $A \oplus B$, is defined as

$A \oplus B = \{z | (B\hat{})z \cap A \neq \emptyset\}$

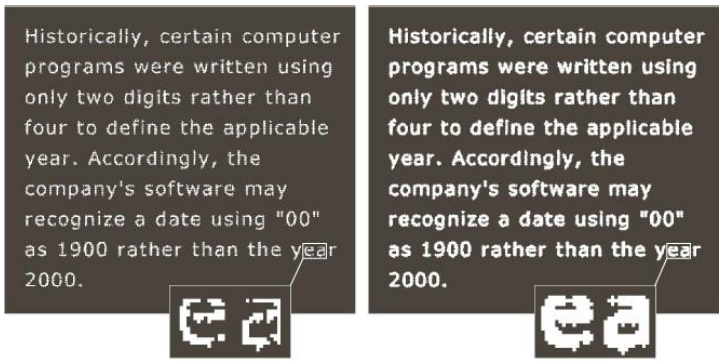
$$A \oplus B = \left\{ z \middle| \left[(\hat{B})_z \cap A \right] \subseteq A \right\}$$



a b c
d e

FIGURE 9.5 (a) Set A. (b) Square structuring element (the dot denotes the origin). (c) Dilation of A by B, shown shaded. (d) Elongated structuring element. (e) Dilation of A using this element. The dotted border in (c) and (e) is the boundary of set A, shown only for reference

4.1



0	1	0
1	1	1
0	1	0

a c
b

FIGURE 4.2
(a) Sample text of poor resolution with broken characters (see magnified view).
(b) Structuring element.
(c) Dilation of (a) by (b). Broken segments were joined.

Erosion

With A and B as sets in Z^2, the erosion of A by B, denoted $A \ominus B$, defined

$$A \ominus B = \{z | (B)_z \subseteq A\}$$

The set of all points z such that B, translated by z, is contained by A.

$$A \ominus B = \left\{ z \mid (B)_z \cap A^c = \varnothing \right\}$$

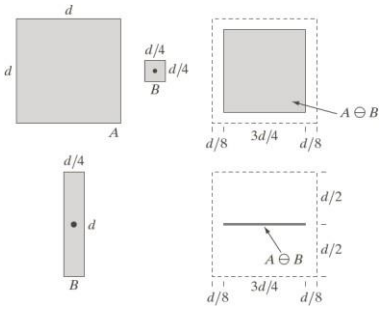
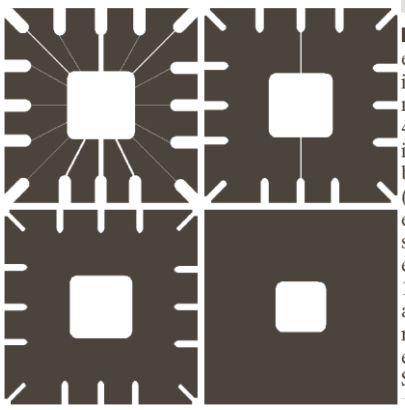


FIGURE 4.3 (a) Set A. (b) Square structuring element, B. (c) Erosion of A by B, shown shaded. (d) Elongated structuring element. (e) Erosion of A by B using this element. The dotted border in (c) and (e) is the boundary of set A, shown only for reference.



a b
c d
FIGURE 4.4 Using erosion to remove image components. (a) A 486 × 486 binary image of a wire-bond mask. (b)–(d) Image eroded using square structuring elements of sizes 11 × 11, 15 × 15, and 45 × 45, respectively. The elements of the SEs were all 1s.

Opening and Closing

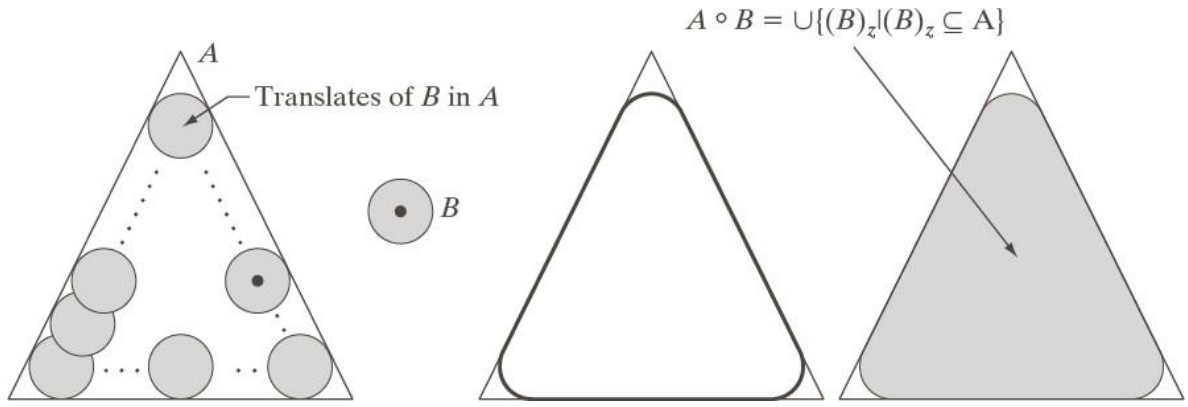
Opening generally smooths the contour of an object, breaks narrow isthmuses, and eliminates thin protrusions. Closing tends to smooth sections of contours but it generates f uses narrow breaks and long thin gulfs, eliminates small holes, and fills gaps in the contour.

The opening of set A by structuring element B,denoted $A \circ B$, is defined as

$$A \circ B = (A - B) \oplus B$$

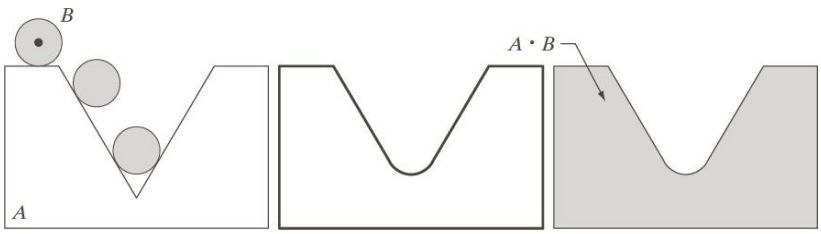
The closing of set A by structuring element B, denoted $A \bullet B$, is defined as

$$A \bullet B = (A \oplus B) - B$$



a b c d

FIGURE 4.5 (a) Structuring element B "rolling" along the inner boundary of A (the dot indicates the origin of B). (b) Structuring element. (c) The heavy line is the outer boundary of the opening. (d) Complete opening (shaded). We did not shade A in (a) for clarity.



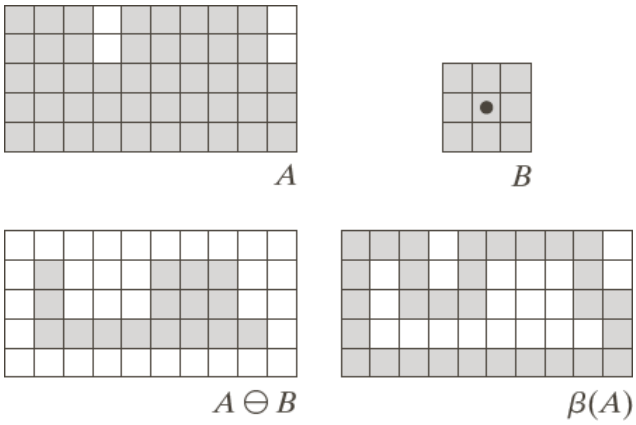
a b c

FIGURE 4.6 (a) Structuring element B "rolling" on the outer boundary of set A. (b) The heavy line is the outer boundary of the closing. (c) Complete closing (shaded). We did not shade A in (a) for clarity.

Boundary Extraction

The boundary of a set A, can be obtained by first eroding A by B and then performing the set difference between A and its erosion.

$$\beta(A) = A - (A - B)$$



a b
c d

FIGURE 4.7 (a) Set *A*. (b) Structuring element *B*. (c) *A* eroded by *B*. (d) Boundary, given by the set difference between *A* and its erosion.

a b

FIGURE 4.8 (a) A simple binary image, with 1s represented in white. (b) Result of using Eq. (9.5-1) with the structuring element in Fig. 9.13(b).

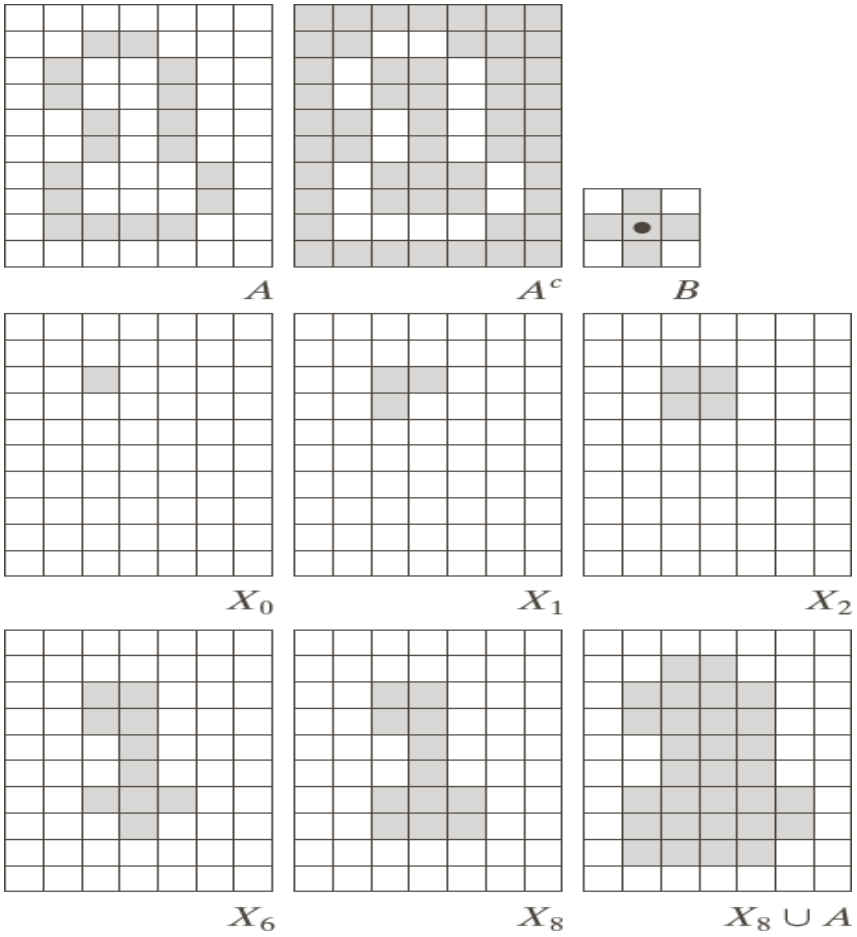
Region Filling

A hole may be defined as a background region surrounded by a connected border of foreground pixels. Let A denote a set whose elements are 8-connected boundaries, each boundary enclosing a background region (i.e., a hole). Given a point in each hole, the objective is to fill all the holes with 1s.

1. Forming an array X_0 of 0s (the same size as the array containing A), except the locations in X_0 corresponding to the given point in each hole, which we set to 1.

2. $X_k = (X_{k-1} + B)$ A^c k=1,2,3,…

Stop the iteration if $X_k = X_{k-1}$



a b c
d e f
g h i

FIGURE 4.9 Hole filling. (a) Set A (shown shaded). (b) Complement of A. (c) Structuring element B. (d) Initial point inside the boundary. (e)–(h) Various steps of Eq. (9.5-2). (i) Final result [union of (a) and (h)].

Thickening

The thickening is defined by the expression

$A \odot B = A \cup (A * B)$

The thickening of A by a sequence of structuring element $\{B\}$ $A \odot \{B\} = ((...((A \odot B\text{^}1) \odot B\text{^}2)...) \odot B\text{^}n)$

In practice, the usual procedure is to thin the background of the set and then complement the result.

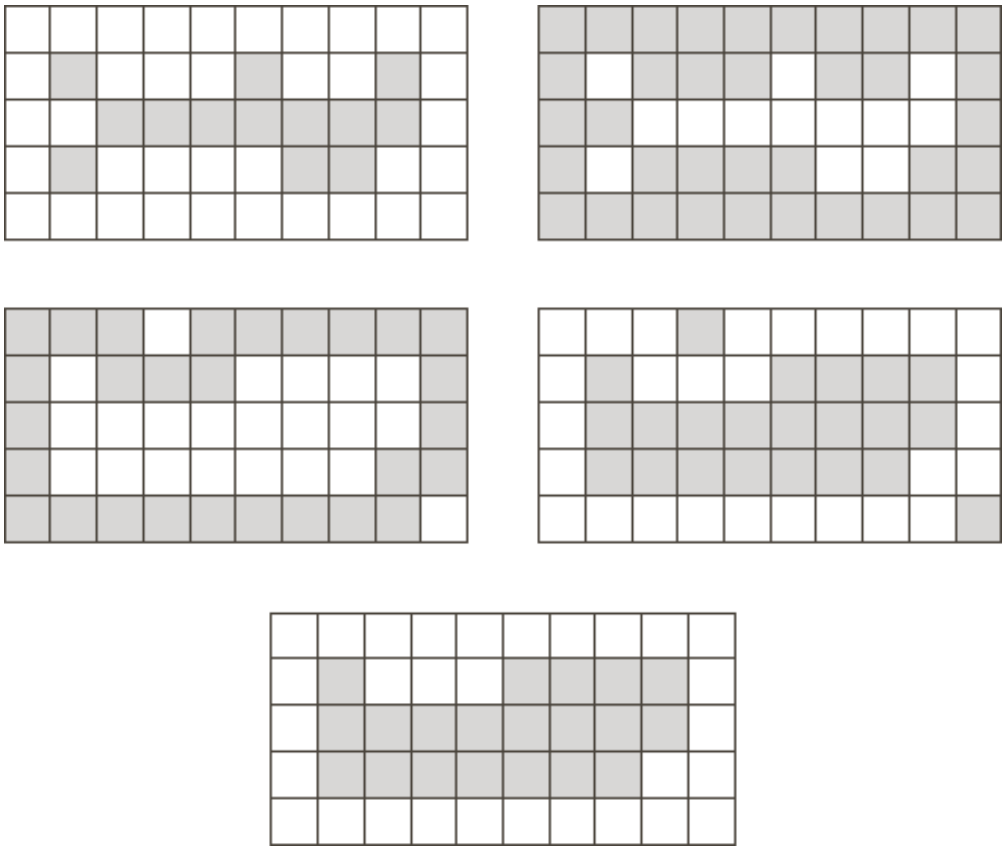


FIGURE 9.12 (a) Set A. (b) Complement of A. (c) Result of thinning the complement of A. (d) Thickened set obtained by complementing (c). (e) Final result, with no disconnected points.

Thinning

The thinning of a set A by a structuring element B, defined

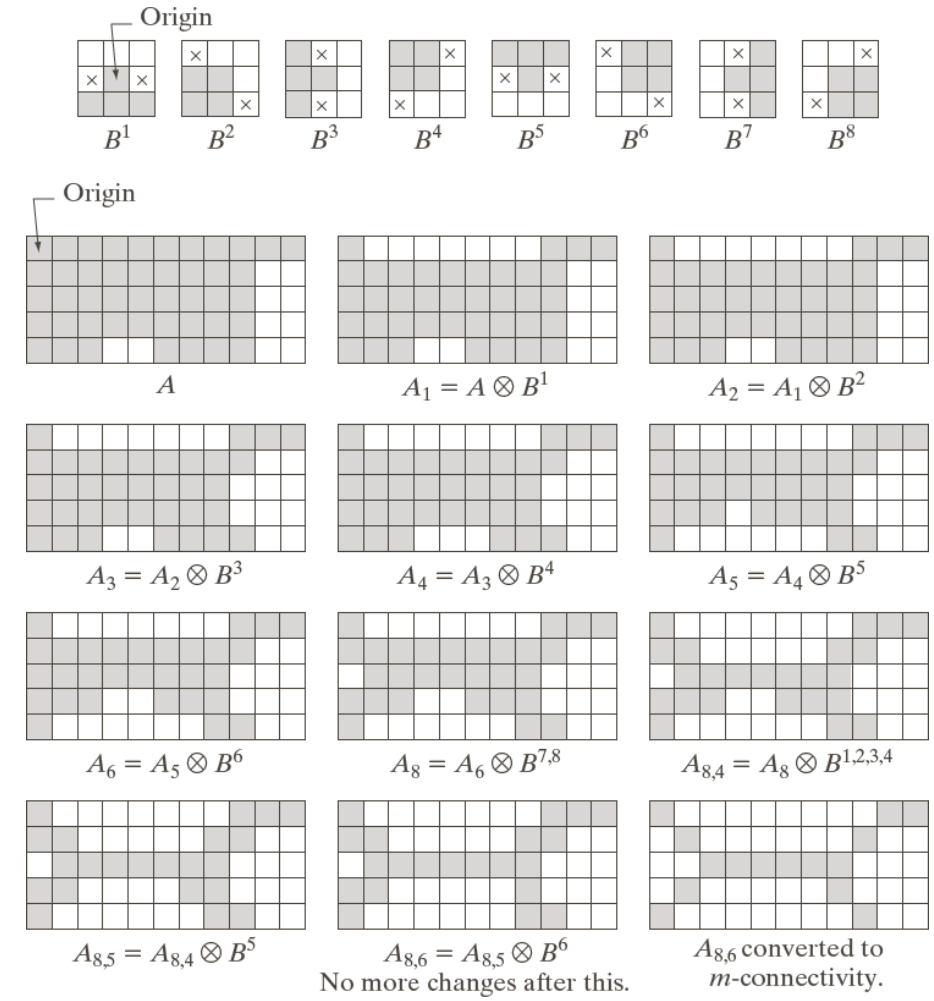
$$A \otimes B = A - (A * B)$$
$$= A \cap (A * B)^c$$

$$\{B\} = \{B^1, B^2, B^3, ..., B^n\}$$

where B^i is a rotated version of B^{i-1}

The thinning of A by a sequence of structuring element $\{B\}$

$$A \otimes \{B\} = ((...((A \otimes B^1) \otimes B^2)...) \otimes B^n)$$



a
b c d
e f g
h i j
k l m

FIGURE 4.11 (a) Sequence of rotated structuring elements used for thinning. (b) Set A. (c) Result of thinning with the first element. (d)–(i) Results of thinning with the next seven elements (there was no change between the seventh and eighth elements). (j) Result of using the first four elements again. (l) Result after convergence. (m) Conversion to m-connectivity.

Image Segmentation

Image segmentation divides an image into regions that are connected and have some similarity within the region and some difference between adjacent regions. The goal is usually to find individual objects in an image. For the most part there are fundamentally two kinds of approaches to segmentation: discontinuity and similarity.

- Similarity may be due to pixel intensity, color or texture.
- Differences are sudden changes (discontinuities) in any of these, but especially sudden changes in intensity along a boundary line, which is called an edge.

There are three kinds of discontinuities of intensity: points, lines and edges. The most

common way to look for discontinuities is to scan a small mask over the image. The mask determines which kind of discontinuity to look for. Only slightly more common than point detection is to find one pixel wide line in an image. For digital images the only three point straight lines are only horizontal, vertical, or diagonal (+ or -45^0).

Edge Linking & Boundary Detection

Two properties of edge points are useful for edge linking:

- the strength (or magnitude) of the detected edge points
- their directions (determined from gradient directions)

- This is usually done in local neighbourhoods.
- Adjacent edge points with similar magnitude and direction are linked.
- For example, an edge pixel with coordinates (x_0,y_0) in a predefined neighbourhood of (x,y) is similar to the pixel at (x,y) if

$$|\nabla f(x, y) - \nabla(x_0, y_0)| \le E, \quad E : \text{a nonnegative threshold}$$

$$|\alpha(x, y) - \alpha(x_0, y_0)| < A, \quad A : \text{a nonegative angle threshold}$$

Hough transform: a way of finding edge points in an image that lie along a straight line. Example: *xy*-plane v.s. *ab*-plane (parameter space)

$$y_I = ax_I + b$$

- The Hough transform consists of finding all pairs of values of θ and ρ which satisfy the

  equations that pass through (x,y).
- These are accumulated in what is basically a 2-dimensional histogram.
- When plotted these pairs of θ and ρ will look like a sine wave. The process is repeated for all appropriate (x,y) locations.

Thresholding

Global – T depends only on gray level values

Local – T depends on both gray level values and local property

Dynamic or Adaptive – T depends on spatial coordinates

Different approaches possible in Graylevel threshold

- Interactive threshold

- Adaptive threshold
- Minimisation method

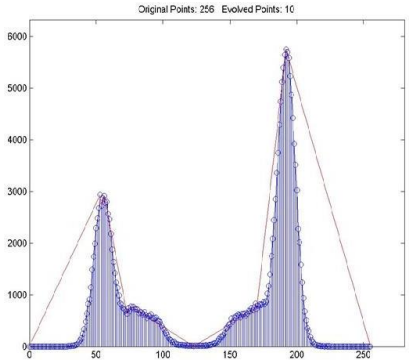


Fig. 4.11 Gray level thresholding

Region based Segmentation

- Edges and thresholds sometimes do not give good results for segmentation.
- Region-based segmentation is based on the connectivity of similar pixels in a region.
  - Each region must be uniform.
  - Connectivity of the pixels within the region is very important.
- There are two main approaches to region-based segmentation: region growing and region splitting.

Basic Formulation

- Let R represent the entire image region.

For example: $P(R_k)$=TRUE if all pixels in R_k have the same gray level. Region splitting is the opposite of region growing.

  - First there is a large region (possible the entire image).
  - Then a predicate (measurement) is used to determine if the region is uniform.
  - If not, then the method requires that the region be split into two regions.
  - Then each of these two regions is independently tested by the predicate (measurement).
  - This procedure continues until all resulting regions are uniform.

The main problem with region splitting is determining where to split a region. One method to divide a region is to use a quad tree structure. Quadtree: a tree in which nodes have exactly four descendants. The split and merge procedure:

  - Split into four disjoint quadrants any region R_i for which $P(R_i)$ =FALSE.
  - Merge any adjacent regions R_j and R_k for which $P(R_j \cup R_k)$ = TRUE. (the

quadtree structure may not be preserved)

– Stop when no further merging or splitting is possible.

TEXT / REFERENCE BOOKS

1. Rafael C. Gonzalez, Richard E. Woods, ŏ" Digital Image Processing", Pearson , Second Edition, 2004.
2. David Saloman,"Data compression" , Springer International, 4th Edition.
3. Khalid Sayood, "Introduction To Data Compressio" ,Elsevier 3rd Edition.
4. Ralfsteinmetz and Klara Nahrstedt, "Multimedia Computing, Communications & Applications" Pearson Edn
5. Rajan Parekh, "Principles of Multimedia, Tata Mc Graw Hill.
6. Anil K. Jain, "Fundamentals of Digital Image Processin", Pearson 2002.
7. J F Koegel Buford- -Multimedia systems Addison Wesley
8. T Vaughan-,"Multimedia: Making it work" Tata Mc Graw Hill

PART A

1. What is edge?
2. How edges are linked through hough transform?
3. State the problems in region splitting and merging based image segmentation.
4. What are the effects of the dilation process?
5. What are the major effects in the erosion process?
6. Mention the properties of opening and closing.
7. What is the use of the closing operation?
8. Give a few applications of morphological operations in the field of image processing.
9. Distinguish between local and global thresholding techniques for image segmentation.
10. What are the advantages/disadvantages if we use more than one seed in a region-growing technique.

PART B

1. Discuss about Morphological Processing.
2. Explain about Region based segmentation.
3. Discuss in detail about Edge Linking and Boundary Detection.
4. Distinguish between image segmentation based on thresholding with image segmentation based on region-growing techniques.
5. Explain region splitting and merging technique for image segmentation with suitable examples.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

UNIT – IV – Digital Image and Multimedia Processing – SEC1605

IV.MULTIMEDIA

Introduction to Multimedia - Media and Data streams - Properties of a Multimedia system - Data streams characteristics - Information units – Multimedia Hardware platforms - Memory and storage devices- Input and output devices - Multimedia software tools. Multimedia Building blocks – Audio: Basic sound concepts - Music-speech-audio file formats – Images and graphics: Basic concepts- Computer image processing.

Introduction

Multimedia has become an inevitable part of any presentation. It has found a variety of applications right from entertainment to education. The evolution of internet has also increased the demand for multimedia content.

Multimedia is probably one of the most overused terms of the 90s. The field is at the crossroads of several major industries: computing, telecommunications, publishing, consumer audio-video electronics, and television/movie/broadcasting. Multimedia not only brings new industrial players to the game, but adds a new dimension to the potential market. For example, while computer networking was essentially targeting a professional market, multimedia embraces both the commercial and the consumer segments. Thus, the telecommunications market involved is not only that of professional or industrial networks— such as medium- or high-speed leased circuits or corporate data networks—but also includes standard telephony or low-speed ISDN. Similarly, not only the segment of professional audio-video is concerned, but also the consumer audio-video market, and the associated TV, movie, and broadcasting sectors.

Definition

Multimedia is the media that uses multiple forms of information content and information processing (e.g. text, audio, graphics, animation, video, interactivity) to inform or entertain the user. Multimedia also refers to the use of electronic media to store and experience multimedia content. Multimedia is similar to traditional mixed media in fine art, but with a broader scope. The term "rich media" is synonymous for interactive multimedia.

What is Multimedia?

People who use the term “multimedia” may have quite different, even opposing, viewpoints. A consumer entertainment vendor, say a phone company, may think of multimedia as interactive TV with hundreds of digital channels or a cable-TV-like service delivered over a high-speed Internet connection. A hardware vendor might, on the other hand, like us to think of multimedia as a laptop that has good sound capability and perhaps the superiority of multimedia-enabled microprocessors that understand additional multimedia instructions.

A computer science or engineering student reading this book likely has a more application-oriented view of what multimedia consists of: applications that use multiple modalities to their advantage, including text, images, drawings, graphics, animation, video, sound (including speech), and, most likely, interactivity of some kind.

This contrasts with media that use only rudimentary computer displays such as textonly or traditional forms of printed or hand-produced material. The popular notion of “convergence” is one that inhabits the college campus as it does the culture at large. In this scenario, computers, smartphones, games, digital TV,multimedia-based search, and so on are converging in technology, presumably to arrive in the near future at a final and fully functional all-round, multimedia-enabled product. While hardware may indeed strive for such all-round devices, the present is already

exciting—multimedia is part of some of the most interesting projects underway in computer science, with the keynote being *interactivity*. The convergence going on in this field is, in fact, a convergence of areas that have in the past been separated but are now finding much to share in this new application area. Graphics, visualization, HCI, artificial intelligence, computer vision, data compression, graph theory, networking, and database systems all have important contributions to make in multimedia at the present time.

Components of Multimedia

The multiple modalities of text, audio, images, drawings, animation, video, and interactivity in multimedia are put to use in ways as diverse as

• Videoconferencing.
• Tele-medicine.
• A web-based video editor that lets anyone create a new video by editing, annotating, and remixing professional videos on the cloud.
• Geographically based, real-time augmented reality, massively multiplayer online video games, making use of any portable device such as smartphones, laptops, or tablets, which function as GPS-aware mobile game consoles.
• Shape shifting TV, where viewers vote on the plot path by phone text messages, which are parsed to direct plot changes in real time.
• A camera that suggests what would be the best type of next shot so as to adhere to good technique guidelines for developing storyboards.
• Cooperative education environments that allow schoolchildren to share a single educational game using two mice at once that pass control back and forth.
• Searching (very) large image and video databases for target visual objects, using semantics of objects.
• Compositing of artificial and natural video into hybrid scenes, placing real appearing computer graphics and video objects into scenes so as to take the physics of objects and lights (e.g., shadows) into account.
• Visual cues of videoconferencing participants, taking into account gaze direction and attention of participants.
• Making multimedia components editable—allowing the user side to decide what components, video, graphics, and so on are actually viewed and allowing the client to move components around or delete them—making components distributed.
• Building "inverse-Hollywood" applications that can recreate the process by which a videowasmade, allowing storyboard pruning and concise video summarization. From a computer science student's point of view, what makesmultimedia interesting is that so much of the material covered in traditional computer science areas bears on the multimedia enterprise. In today's digitalworld, multimedia content is recorded and played, displayed, or accessed by digital information content-processing devices, ranging from smartphones, tablets, laptops, personal computers, smart TVs, and game consoles, to servers and data centers, over such distribution media as USB flash drives (keys), discs, and hard drives, or more popularly nowadays, wired and wireless networks.

The Term "Media"

As with most generic words, the meaning of the word media varies with the context in which it is used. Our definition of medium is "a means to distribute and represent information." Media are, for example, text, graphics, pictures, voice, sound, and music. In this sense, we could just as well add water and the atmosphere to this definition.

It provides a subtle differentiation of various aspects of this term by use of various criteria to distinguish between perception, representation, presentation, storage, transmission, and information

exchange media. The following sections describe these attributes.

Perception Media

Perception media refers to the nature of information perceived by humans, which is not strictly identical to the sense that is stimulated. For example, a still image and a movie convey information of a different nature, though stimulating the same sense. The question to ask here is: How do humans perceive information? In this context, we distinguish primarily between what we see and what we hear. Auditory media include music, sound, and voice. Visual media include text, graphics, and still and moving pictures. This differentiation can be further refined. For example, a visual medium can consist of moving pictures, animation, and text. In turn, moving pictures normally consist of a series of scenes that, in turn, are composed of single pictures.

Representation Media

The term representation media refers to how information is represented internally to the computer. The encoding used is of essential importance. The question to ask here is: How is information encoded in the computer? There are several options:

• Each character of a piece of text is encoded in ASCII.

• A picture is encoded by the CEPT or CAPTAIN standard, or the GKS graphics standard can serve as a basis.

• An audio data stream is available in simple PCM encoding and a linear quantization of 16 bits per sampling value.

• A single image is encoded as Group-3 facsimile or in JPEG format.

• A combined audio-video sequence is stored in the computer in various TV standards (e.g., PAL, SECAM, or NTSC), in the CCIR-601 standard, or in MPEG format.

Presentation Media

The term presentation media refers to the physical means used by systems to reproduce information for humans. For example, a TV set uses a cathode-ray tube and loudspeaker. The question to ask here is: Which medium is used to output information from the computer or input in the computer? We distinguish primarily between output and input. Media such as paper, computer monitors, and loudspeakers are output media, while keyboards, cameras, and microphones are input media.

Storage Media

The term storage media is often used in computing to refer to various physical means for storing computer data, such as magnetic tapes, magnetic disks, or digital optical disks. However, data storage is not limited to the components available in a computer, which means that paper is also considered a storage medium.

Transmission Media

The term transmission media refers to the physical means—cables of various types, radio tower, satellite, or ether (the medium that transmit radio waves)—that allow the transmission of telecommunication signals.

Information Exchange Media

Information exchange media include all data media used to transport information, e.g., all storage and transmission media.

For example, information can be exchanged by storing it on a removable medium and transporting the medium from one location to another. These storage media include microfilms, paper, and floppy disks. Information can also be exchanged directly, if transmission media such as coaxial cables, optical fibers, or radio waves are used.

Presentation Spaces and Presentation Values

The terms described above serve as a basis to characterize the term medium in the information processing context. The description of perception media is closest to our definition of media: those media concerned mainly with the human senses. Each medium defines presentation values in presentation spaces [HD90, SH91], which address our five senses. Paper or computer monitors are

examples of visual presentation spaces. A computer- controlled slide show that projects a screen's content over the entire projection screen is a visual presentation space. Stereophony and quadrophony define acoustic presentation spaces. Presentation spaces are part of the above-described presentation media used to output information. Presentation values determine how information from various media is represented.

While text is a medium that represents a sentence visually as a sequence of characters, voice is a medium that represents information acoustically in the form of pressure waves. In some media, the presentation values cannot be interpreted correctly by humans. Examples include temperature, taste, and smell. Other media require a address our five senses.

Presentation Dimensions

Each presentation space has one or more presentation dimensions. A computer monitor has two space dimensions, while holography and stereophony need a third one. Time can occur as an additional dimension within each presentation space, which is critical for multimedia systems.

Key Properties of a Multimedia System

Multimedia systems involve several fundamental notions. They must be computer-controlled. Thus, a computer must be involved at least in the presentation of the information to the user. They are integrated, that is, they use a minimal number of different devices. An example is the use of a single computer screen to display all types of visual information. They must support media independence. And lastly, they need to handle discrete and continuous media. The following sections describe these key properties.

- **Discrete and Continuous Media**

Not just any arbitrary combination of media deserves the name multimedia. Many people call a simple word processor that handles embedded graphics a multimedia application because it uses two media. By our definition, we talk about multimedia if the application uses both discrete and continuous media. This means that a multimedia application should process at least one discrete and one continuous medium. A word processor with embedded graphics is not a multimedia application by our definition.

- **Independent Media**

An important aspect is that the media used in a multimedia system should be independent. Although a computer-controlled video recorder handles audio and moving image information, there is a temporal dependence between the audio part and the video part. In contrast, a system that combines signals recorded on a DAT (Digital Audio Tape) recorder with some text stored in a computer to create a presentation meets the independence criterion. Other examples are combined text and graphics blocks, which can be in an arbitrary space arrangement in relation to one another.

- **Computer-Controlled Systems**

The independence of media creates a way to combine media in an arbitrary form for presentation. For this purpose, the computer is the ideal tool. That is, we need a system capable of processing media in a computer-controlled way. The system can be optionally programmed by a system programmer and/or by a user (within certain limits). The simple recording or playout of various media in a system, such as a video recorder, is not sufficient to meet the computer-control criterion.

- **Integration**

Computer-controlled independent media streams can be integrated to form a global system so that, together, they provide a certain function. To this end, synchronic relationships of time, space, and content are created between them. A word processor that supports text, spreadsheets, and graphics does not meet the integration criterion unless it allows program-supported references between the

data. We achieve a high degree of integration only if the application is capable of, for example, updating graphics and text elements automatically as soon as the contents of the related spreadsheet cell changes. This kind of flexible media handling is not a matter to be taken for granted—even in many products sold under the multimedia system label. This aspect is important when talking of integrated multimedia systems. Simply speaking, such systems should allow us to do with moving images and sound what we can do with text and graphics. While conventional systems can send a text message to another user, a highly integrated multimedia system provides this function and support for voice messages or a voice-text combination.

Categories of Multimedia

Multimedia may be broadly divided into linear and non-linear categories. Linear active content progresses without any navigation control for the viewer such as a cinema presentation. Non-linear content offers user interactivity to control progress as used with a computer game or used in self-paced computer based training. Non-linear content is also known as hypermedia content. Multimedia presentations can be live or recorded. A recorded presentation may allow interactivity via a navigation system. A live multimedia presentation may allow interactivity via interaction with the presenter or performer.

Characterizing Data Streams

Distributed networked multimedia systems transmit both discrete and continuous media streams, i.e., they exchange information. In a digital system, information is split into units (packets) before it is transmitted. These packets are sent by one system component (the source) and received by another one (the sink). Source and sink can reside on different computers. A data stream consists of a (temporal) sequence of packets. This means that it has a time component and a lifetime. Packets can carry information from continuous and discrete media. The transmission of voice in a telephone system is an example of a continuous medium. When we transmit a text file, we create a data stream that represents a discrete medium. When we transmit information originating from various media, we obtain data streams that have very different characteristics. The attributes asynchronous, synchronous, and isochronous are traditionally used in the field of telecommunications to describe the characteristics of a data transmission. For example, they are used in FDDI to describe the set of options available for an end-to-end delay in the transmission of single packets.

- Asynchronous Transmission Mode
- Synchronous Transmission Mode
- Isochronous Transmission Mode
- Characterizing Continuous Media Data Streams

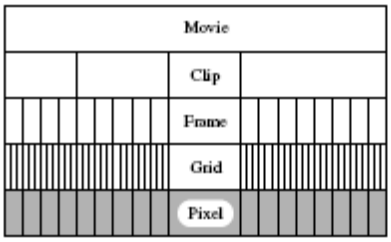
Information Units

Continuous (time-dependent) media consist of a (temporal) sequence of information units. Based on Protocol Data Units (PDUs), this section describes such an information unit, called a Logical Data Unit (LDU). An LDU's information quantity and data quantities can have different meanings:

Let's use Joseph Haydn's symphony, The Bear, as our first example. It consists of the four musical movements, vivace assai, allegretto, menuet, and finale vivace. Each movement is an independent, self-sufficient part of this composition. It contains a sequence of scores for the musical instruments used. In a digital system, these scores are a sequence of sampling values. We will not use any compression in this example, but apply PCM encoding with a linear characteristic curve.

For CD-DA quality, this means 44,100 sampling values per second, which are encoded at 16bits per channel. On a CD, these sampling values are grouped into units with a duration of 1/75 second. We could now look at the entire composition and define single movements, single scores, the grouped 1/75-s sampling values, or even single sampling values as LDUs. Some operations can be applied to the playback of the entire composition—as one single LDU. Other functions refer to the smallest meaningful unit (in this case the scores). In digital signal processing, sampling values are LDUs.

we see that the uncompressed video sequence consists of single chips, each representing a scene. Each of these scenes consists of a sequence of single images. Each single image can be separated into various regions, for example regions with a size of 16.16 pixels. In turn, each pixel contains a luminance value and a chrominance value. This means that a single image is not the only possible LDU in a motion video sequence. Each scene and each pixel are also LDUs. The redundancies in single image sequences of an MPEG-encoded video stream can be used to reduce the data quantity by applying an inter-frame compression method. In this case, the smallest self-sufficient meaningful units are single-image sequences.



Granularity of a motion video sequence showing its logical data units

A phenomenon called granularity characterizes the hierarchical decomposition of an audio or video stream in its components. This example uses a symphony and a motion video to generally describe extensive information units. We distinguish between closed and open LDUs. Closed LDUs have a well-defined duration. They are normally stored sequences. In open LDUs, the data stream's duration is not known in advance. Such a data stream is delivered to the computer by a camera, a microphone, or a similar device.

Multimedia Hardware

Introduction

The hardware required for multimedia PC depends on the personal preference, budget, project delivery requirements and the type of material and content in the project. Multimedia production was much smoother and easy in Macintosh than in Windows. But Multimedia content production in windows has been made easy with additional storage and less computing cost. Right selection of multimedia hardware results in good quality multimedia presentation.

Multimedia Hardware

The hardware required for multimedia can be classified into five. They are

1. Connecting Devices
2. Input devices
3. Output devices
4. Storage devices and
5. Communicating devices.

Connecting Devices

Among the many hardware – computers, monitors, disk drives, video projectors, light valves, video projectors, players, VCRs, mixers, sound speakers there are enough wires which connect these devices. The data transfer speed the connecting devices provide will determine the faster delivery of

the multimedia content.
The most popularly used connecting devices are:

- SCSI
- USB
- MCI
- IDE
- USB

SCSI

SCSI (**Small Computer System Interface**) is a set of standards for physically connecting and transferring data between computers and peripheral devices. The SCSI standards define commands, protocols, electrical and optical interfaces. SCSI is most commonly used for hard disks and tape drives, but it can connect a wide range of other devices, including scanners, and optical drives (CD, DVD, etc.).

SCSI is most commonly pronounced "scuzzy". Since its standardization in 1986, SCSI has been commonly used in the Apple Macintosh and Sun Microsystems computer lines and PC server systems. SCSI has never been popular in the low-priced IBM PC world, owing to the lower cost and adequate performance of its ATA hard disk standard. SCSI drives and even SCSI RAIDs became common in PC workstations for video or audio production, but the appearance of large cheap SATA drives means that SATA is rapidly taking over this market.

Currently, SCSI is popular on high-performance workstations and servers. RAIDs on servers almost always use SCSI hard disks, though a number of manufacturers offer SATA-based RAID systems as a cheaper option. Desktop computers and notebooks more typically use the ATA/IDE or the newer SATA interfaces for hard disks, and USB and FireWire connections for external devices.

SCSI interfaces

SCSI is available in a variety of interfaces. The first, still very common, was parallel SCSI (also called SPI). It uses a parallel electrical bus design. The traditional SPI design is making a transition to Serial Attached SCSI, which switches to a serial point-to-point design but retains other aspects of the technology. iSCSI drops physical implementation entirely, and instead uses TCP/IP as a transport mechanism. Finally, many other interfaces which do not rely on complete SCSI standards still implement the SCSI command protocol.

The following table compares the different types of SCSI.

Terms	Bus Speed (MB/sec)	Bus Width (Bits)	Number of Devices supported
SCSI-1	5	8	8
SCSI-2	10	8	8
SCSI-3	20	8	16
SCSI-3	20	8	4
SCSI-3 *1*	20	16	16
SCSI-3 UW	40	16	16
SCSI-3 UW	40	16	8
SCSI-3 UW	40	16	4
SCSI-3 U2	40	8	8
SCSI-3 U2	80	16	2
SCSI-3 U2W	80	16	16
SCSI-3 U2W	80	16	2
SCSI-3 U3	160	16	16

Media Control Interface (MCI)

The **Media Control Interface**, **MCI** in short, is an aging API for controlling multimedia peripherals connected to a Microsoft Windows or OS/2 computer. MCI makes it very simple to write a program which can play a wide variety of media files and even to record sound by just passing commands as strings. It uses relations described in Windows registries or in the [MCI] section of the file SYSTEM.INI. The MCI interface is a high-level API developed by Microsoft and IBM for controlling multimedia devices, such as CD-ROM players and audio controllers. The advantage is that MCI commands can be transmitted both from the programming language and from the scripting language (open script, lingo). For a number of years, the MCI interface has been phased out in favor of the DirectX APIs.

IDE

Usually storage devices connect to the computer through an Integrated Drive Electronics (IDE) interface. Essentially, an IDE interface is a standard way for a storage device to connect to a computer. IDE is actually not the true technical name for the interface standard. The original name, AT Attachment (ATA), signified that the interface was initially developed for the IBM AT computer. IDE was created as a way to standardize the use of hard drives in computers. The basic concept behind IDE is that the hard drive and the controller should be combined. The controller is a small circuit board with chips that provide guidance as to exactly how the hard drive stores and accesses data. Most controllers also include some memory that acts as a buffer to enhance hard drive performance. Before IDE, controllers and hard drives were separate and often proprietary. In other words, a controller from one manufacturer might not work with a hard drive from another manufacturer. The distance between the controller and the hard drive could result in poor signal quality and affect performance. Obviously, this caused much frustration for computer users.

USB

Universal Serial Bus (USB) is a serial bus standard to interface devices. A major component in the legacy-free PC, USB was designed to allow peripherals to be connected using a single standardized interface socket and to improve plug-and-play capabilities by allowing devices to be connected and disconnected without rebooting the computer (hot swapping). Other convenient features include providing power to low-consumption devices without the need for an external power supply and allowing many devices to be used without requiring manufacturer specific, individual device drivers to be installed. USB is intended to help retire all legacy varieties of serial and parallel ports. USB can connect computer peripherals such as mouse devices, keyboards, PDAs, gamepads and joysticks, scanners, digital cameras, printers, personal media players, and flash drives. For many of those devices USB has become the standard connection method. USB is also used extensively to connect non-networked printers; USB simplifies connecting several printers to one computer. USB was originally designed for personal computers, but it has become commonplace on other devices such as PDAs and video game consoles.

The design of USB is standardized by the USB Implementers Forum (USB-IF), an industry standards body incorporating leading companies from the computer and electronics industries. Notable members have included Apple Inc., Hewlett-Packard, NEC, Microsoft, Intel, and Agere.

A USB system has an asymmetric design, consisting of a host, a multitude of downstream USB ports, and multiple peripheral devices connected in a tiered-star topology. Additional USB hubs may be included in the tiers, allowing branching into a tree structure, subject to a limit of 5 levels of tiers. USB host may have multiple host controllers and each host controller may provide one or more USB ports. Up to 127 devices, including the hub devices, may be connected to a single host controller. USB devices are linked in series through hubs. There always exists one hub known as the root hub, which is built-in to the host controller. So-called "sharing hubs" also exist; allowing multiple computers to access the same peripheral device(s), either switching access between PCs

automatically or manually. They are popular in small office environments. In network terms they converge rather than diverge branches. A single physical USB device may consist of several logical sub-devices that are referred to as device functions, because each individual device may provide several functions, such as a webcam (video device function) with a built-in microphone (audio device function).

Multimedia Hardware – Storage Devices

A data storage device is a device for recording (storing) information (data). Recording can be done using virtually any form of energy. A storage device may hold information, process information, or both. A device that only holds information is a recording medium. Devices that process information (data storage equipment) may both access a separate portable (removable) recording medium or a permanent component to store and retrieve information.

Electronic data storage is storage which requires electrical power to store and retrieve that data. Most storage devices that do not require visual optics to read data fall into this category. Electronic data may be stored in either an analog or digital signal format. This type of data is considered to be electronically encoded data, whether or not it is electronically stored. Most electronic data storage media (including some forms of computer storage) are considered permanent (non-volatile) storage, that is, the data will remain stored when power is removed from the device. In contrast, electronically stored information is considered volatile memory.

Memory And Storage Devices

By adding more memory and storage space to the computer, the computing needs and habits to keep pace, filling the new capacity. To estimate the memory requirements of a multimedia project- the space required on a floppy disk, hard disk, or CD-ROM, not the random access sense of the project's content and scope. Color images, Sound bites, video clips, and the programming code that glues it all together require memory; if there are many of these elements, you will need even more. If you are making multimedia, you will also need to allocate memory for storing and archiving working files used during production, original audio and video clips, edited pieces, and final mixed pieces, production paperwork and correspondence, and at least one backup of your project files, with a second backup stored at another location.

Random Access Memory (RAM)

RAM is the main memory where the Operating system is initially loaded and the application programs are loaded at a later stage. RAM is volatile in nature and every program that is quit/exit is removed from the RAM. More the RAM capacity, higher will be the processing speed.

If there is a budget constraint, then it is certain to produce a multimedia project on a slower or limited-memory computer. On the other hand, it is profoundly frustrating to face memory (RAM) shortages time after time, when you're attempting to keep multiple applications and files open simultaneously. It is also frustrating to wait the extra seconds required oh each editing step when working with multimedia material on a slow processor.

On the Macintosh, the minimum RAM configuration for serious multimedia production is about 32MB; but even 64MB and 256MB systems are becoming common, because while digitizing audio or video, you can store much more data much more quickly in RAM. And when you're using some software, you can quickly chew up available RAM – for example, Photoshop (16MB minimum, 20MB recommended); After Effects (32MBrequired), Director (8MB minimum, 20MB better); Page maker (24MB recommended); Illustrator (16MB recommended); Microsoft Office (12MB recommended).

In spite of all the marketing hype about processor speed, this speed is ineffective if not accompanied by sufficient RAM. A fast processor without enough RAM may waste processor cycles while it swaps needed portions of program code into and out of memory. In some cases, increasing available RAM may show more performance improvement on your system than upgrading the processor clip.

On an MPC platform, multimedia authoring can also consume a great deal of memory. It may be

needed to open many large graphics and audio files, as well as your authoring system, all at the same time to facilitate faster copying/pasting and then testing in your authoring software. Although 8MB is the minimum under the MPC standard, much more is required as of now.

Read-Only Memory (ROM)

Read-only memory is not volatile, Unlike RAM, when you turn off the power to a ROM chip, it will not forget, or lose its memory. ROM is typically used in computers to hold the small BIOS program that initially boots up the computer, and it is used in printers to hold built-in fonts. Programmable ROMs (called EPROM's) allow changes to be made that are not forgotten. A new and inexpensive technology, optical read-only memory (OROM), is provided in proprietary data cards using patented holographic storage. Typically, OROM s offer 128MB of storage, have no moving parts, and use only about 200 mill watts of power, making them ideal for handheld, battery-operated devices.

Floppy and Hard Disks

Adequate storage space for the production environment can be provided by large capacity hard disks; a server-mounted disk on a network; Zip, Jaz, or SyQuest removable cartridges; optical media; CD-R (compact disc-recordable) discs; tape; floppy disks; banks of special memory devices; or any combination of the above. Removable media (floppy disks, compact or optical discs, and cartridges) typically fit into a letter-sized mailer for overnight courier service. One or many disks may be required for storage and archiving each project, and it is necessary to plan for backups kept off-site.

Floppy disks and hard disks are mass-storage devices for binary data-data that can be easily read by a computer. Hard disks can contain much more information than floppy disks and can operate at far greater data transfer rates. In the scale of things, floppies are, however, no longer "mass-storage" devices.

A floppy disk is made of flexible Mylar plastic coated with a very thin layer of special magnetic material. A hard disk is actually a stack of hard metal platters coated with magnetically sensitive material, with a series of recording heads or sensors that hover a hairbreadth above the fast-spinning surface, magnetizing or demagnetizing spots along formatted tracks using technology similar to that used by floppy disks and audio and video tape recording. Hard disks are the most common mass-storage device used on computers, and for making multimedia, it is necessary to have one or more large-capacity hard disk drives.

As multimedia has reached consumer desktops, makers of hard disks have been challenged to build smaller profile, larger-capacity, faster, and less-expensive hard disks. In 1994, hard disk manufactures sold nearly 70 million units; in 1995, more than 80 million units. And prices have dropped a full order of magnitude in a matter of months. By 1998, street prices for 4GB drives (IDE) were less than $200. As network and Internet servers increase the demand for centralized data storage requiring terabytes (1 trillion bytes), hard disks will be configured into fail-proof redundant array offering built-in protection against crashes.

Zip, jaz, SyQuest, and Optical storage devices

SyQuest's 44MB removable cartridges have been the most widely used portable medium among multimedia developers and professionals, but Iomega's inexpensive Zip drives with their likewise inexpensive 100MB cartridges have significantly penetrated SyQuest's market share for removable media. Iomega's Jaz cartridges provide a gigabyte of removable storage media and have fast enough transfer rates for audio and video development. Pinnacle Micro, Yamaha, Sony, Philips, and others offer CD-R "burners" for making write-once compact discs, and some double as quad-speed players. As blank CD-R discs become available for less than a dollar each, this write-once media competes as a distribution vehicle. CD-R is described in greater detail a little later in the chapter. Magneto-optical (MO) drives use a high-power laser to heat tiny spots on the metal oxide coating of the disk. While the spot is hot, a magnet aligns the oxides to provide a 0 or 1 (on or off) orientation. Like SyQuests and other Winchester hard disks, this is rewritable technology, because the spots can

be repeatedly heated and aligned. Moreover, this media is normally not affected by stray magnetism (it needs both heat and magnetism to make changes), so these disks are particularly suitable for archiving data. The data transfer rate is, however, slow compared to Zip, Jaz, and SyQuest technologies. One of the most popular formats uses a 128MB-capacity disk-about the size of a 3.5-inch floppy. Larger-format magneto-optical drives with 5.25-inch cartridges offering 650MB to 1.3GB of storage are also available.

Digital versatile disc (DVD)

In December 1995, nine major electronics companies (Toshiba, Matsushita, Sony, Philips, Time Waver, Pioneer, JVC, Hitachi, and Mitsubishi Electric) agreed to promote a new optical disc technology for distribution of multimedia and feature-length movies called DVD. With this new medium capable not only of gigabyte storage capacity but also full motion video (MPEG2) and high-quantity audio in surround sound, the bar has again risen for multimedia developers. Commercial multimedia projects will become more expensive to produce as consumer's performance expectations rise. There are two types of DVD-DVD-Video and DVD-ROM; these reflect marketing channels, not the technology.

DVD can provide 720 pixels per horizontal line, whereas current television (NTSC) provides 240-television pictures will be sharper and more detailed. With Dolby AC-3 Digital surround Sound as part of the specification, six discrete audio channels can be programmed for digital surround sound, and with a separate subwoofer channel, developers can program the low-frequency doom and gloom music popular with Hollywood. DVD also supports Dolby pro-Logic Surround Sound, standard stereo and mono audio. Users can randomly access any section of the disc and use the slow-motion and freeze-frame features during movies. Audio tracks can be programmed for as many as 8 different languages, with graphic subtitles in 32 languages. Some manufactures such as Toshiba are already providing parental control features in their players (user's select lockout ratings from G to NC-17).

CD-ROM Players

Compact disc read-only memory (CD-ROM) players have become an integral part of the multimedia development workstation and are important delivery vehicle for large, mass-produced projects. A wide variety of developer utilities, graphic backgrounds, stock photography and sounds, applications, games, reference texts, and educational software are available only on this medium. CD-ROM players have typically been very slow to access and transmit data (150k per second, which is the speed required of consumer Red Book Audio CDs), but new developments have led to double, triple, quadruple, speed and even 24x drives designed specifically for computer (not Red Book Audio) use. These faster drives spool up like washing machines on the spin cycle and can be somewhat noisy, especially if the inserted compact disc is not evenly balanced.

CD Recorders

With a compact disc recorder, you can make your own CDs using special CD recordable (CD-R) blank optical discs to create a CD in most formats of CD-ROM and CD-Audio. The machines are made by Sony, Phillips, Ricoh, Kodak, JVC, Yamaha, and Pinnacle. Software, such as Adaptec's Toast for Macintosh or Easy CD Creator for Windows, lets you organize files on your hard disk(s) into a "virtual" structure, then writes them to the CD in that order. CD-R discs are made differently than normal CDs but can play in any CD-Audio or CD-ROM player. They are available in either a "63 minute" or "74 minute" capacity for the former, that means about 560MB, and for the latter, about 650MB. These write-once CDs make excellent high-capacity file archives and are used extensively by multimedia developers for premastering and testing CDROM projects and titles.

Videodisc Players

Videodisc players (commercial, not consumer quality) can be used in conjunction with the computer to deliver multimedia applications. You can control the videodisc player from your authoring software with X-Commands (XCMDs) on the Macintosh and with MCI commands in Windows.

The output of the videodisc player is an analog television signal, so you must setup a television separate from your computer monitor or use a video digitizing board to "window" the analog signal on your monitor.

Image Storage

Bit depth: the number of bits per pixel

Bit depth (bits per pixel)	Number of colours or tones	Relationship
1	2	$2^1 = 2$
2	4	$2^2 = 4$
3	8	$2^3 = 8$
4	16	$2^4 = 16$
6	64	$2^6 = 64$
8	256	$2^8 = 256$
16	65,536	$2^{16} = 65,536$
24	16,777,216	$2^{24} = 16,777,216$
32	4,294,967,296	$2^{32} = 4,294,967,296$

$$\text{File size} = \frac{\text{Horizontal x Vertical x Bit Depth}}{\text{8 x 1024 bits (= 1 Kb)}}$$

Multimedia Hardware – Input Devices, Output Devices, Communication Devices

An input device is a hardware mechanism that transforms information in the external world for consumption by a computer. An output device is a hardware used to communicate the result of data processing carried out by the user or CPU.

Input devices

Often, input devices are under direct control by a human user, who uses them to communicate commands or other information to be processed by the computer, which may then transmit feedback to the user through an output device. Input and output devices together make up the hardware interface between a computer and the user or external world. Typical examples of input devices include keyboards and mice. However, there are others which provide many more degrees of freedom. In general, any sensor which monitors, scans for and accepts information from the external world can be considered an input device, whether or not the information is under the direct control of a user.

Input devices can be classified according to:-

- the modality of input (e.g. mechanical motion, audio, visual, sound, etc.)
- whether the input is discrete (e.g. keypresses) or continuous (e.g. a mouse's position, though digitized into a discrete quantity, is high-resolution enough to be thought of as continuous)
- the number of degrees of freedom involved (e.g. many mice allow 2D positional input, but some devices allow 3D input, such as the Logitech Magellan Space Mouse) Pointing devices, which are input devices used to specify a position in space, can

further be classified according to

- Whether the input is direct or indirect. With direct input, the input space coincides with the display space, i.e. pointing is done in the space where visual feedback or the cursor appears. Touchscreens and light pens involve direct input. Examples involving indirect input include the mouse and trackball.

- Whether the positional information is absolute (e.g. on a touch screen) or relative (e.g. with a mouse that can be lifted and repositioned)

Note that direct input is almost necessarily absolute, but indirect input may be either absolute or relative. For example, digitizing graphics tablets that do not have an embedded screen involve indirect input, and sense absolute positions and are often run in an absolute input mode, but they may also be setup to simulate a relative input mode where the stylus or puck can be lifted and repositioned.

Imaging and Video input devices

Flat-Bed Scanners A scanner may be the most useful piece of equipment used in the course of producing a multimedia project; there are flat-bed and handheld scanners. Most commonly available are gray-scale and color flat-bed scanners that provide a resolution of 300 or 600 dots per inch (dpi). Professional graphics houses may use even higher resolution units. Handheld scanners can be useful for scanning small images and columns of text, but they may prove inadequate for the multimedia development.

Be aware that scanned images, particularly those at high resolution and in color, demand an extremely large amount of storage space on the hard disk, no matter what instrument is used to do the scanning. Also remember that the final monitor display resolution for your multimedia project will probably be just 72 or 95 dpi-leave the very expensive ultra-high-resolution scanners for the desktop publishers. Most expensive flat-bed scanners offer at least 300 dpi resolution, and most scanners allow to set the scanning resolution.

Scanners helps make clear electronic images of existing artwork such as photos, ads, pen drawings, and cartoons, and can save many hours when you are incorporating proprietary art into the application. Scanners also give a starting point for the creative diversions. The devices used for capturing image and video are:

- Webcam
- Image scanner
- Fingerprint scanner
- Barcode reader
- 3D scanner
- medical imaging sensor technology
  - o Computed tomography
  - o Magnetic resonance imaging
  - o Positron emission tomography
  - o Medical ultrasonography

Audio input devices

The devices used for capturing audio are

- Microphone

- Speech recognition

Note that MIDI allows musical instruments to be used as input devices as well.

Touch screens

Touch screens are monitors that usually have a textured coating across the glass face. This

coating is sensitive to pressure and registers the location of the user's finger when it touches the screen. The Touch Mate System, which has no coating, actually measures the pitch, roll, and yaw rotation of the monitor when pressed by a finger, and determines how much force was exerted and the location where the force was applied. Other touch screens use invisible beams of infrared light that crisscross the front of the monitor to calculate where a finger was pressed. Pressing twice on the screen in quick and dragging the finger, without lifting it, to another location simulates a mouse click and drag. A keyboard is sometimes simulated using an onscreen representation so users can input names, numbers, and other text by pressing "keys".
Touch screen recommended for day-to-day computer work, but are excellent for multimedia applications in a kiosk, at a trade show, or in a museum delivery system anything involving public input and simple tasks. When your project is designed to use a touch screen, the monitor is the only input device required, so you can secure all other system hardware behind locked doors to prevent theft or tampering.

Output Devices

Presentation of the audio and visual components of the multimedia project requires hardware that may or may not be included with the computer itself-speakers, amplifiers, monitors, motion video devices, and capable storage systems. The better the equipment, of course, the better the presentation. There is no greater test of the benefits of good output hardware than to feed the audio output of your computer into an external amplifier system: suddenly the bass sounds become deeper and richer, and even music sampled at low quality may seem to be acceptable.

Audio devices

All Macintoshes are equipped with an internal speaker and a dedicated sound clip, and they are capable of audio output without additional hardware and/or software. To take advantage of built-in stereo sound, external speaker are required. Digitizing sound on the Macintosh requires an external microphone and sound editing/recording software such as SoundEdit16 from Macromedia, Alchemy from Passport, or Sound Designer from DigiDesign.

Amplifiers and Speakers

Often the speakers used during a project's development will not be adequate for its presentation. Speakers with built-in amplifiers or attached to an external amplifier are important when the project will be presented to a large audience or in a noisy setting.

Monitors

The monitor needed for development of multimedia projects depends on the type of multimedia application created, as well as what computer is being used. A wide variety of monitors is available for both Macintoshes and PCs. High-end, large-screen graphics monitors are available for both, and they are expensive.
Serious multimedia developers will often attach more than one monitor to their computers, using add-on graphic board. This is because many authoring systems allow to work with several open windows at a time, so we can dedicate one monitor to viewing the work we are creating or designing, and we can perform various editing tasks in windows on other monitors that do not block the view of your work. Editing windows that overlap a work view when developing with Macromedia's authoring environment, director, on one monitor. Developing in director is best with at least two monitors, one to view the work the other two view the "score". A third monitor is often added by director developers to display the "Cast".

Video Device

No other contemporary message medium has the visual impact of video. With a video digitizing board installed in a computer, we can display a television picture on your monitor. Some boards include a frame-grabber feature for capturing the image and turning it in to a color bitmap, which can be saved as a PICT or TIFF file and then used as part of a graphic or a

background in your project.

Display of video on any computer platform requires manipulation of an enormous amount of data. When used in conjunction with videodisc players, which give precise control over the images being viewed, video cards you place an image in to a window on the computer monitor; a second television screen dedicated to video is not required. And video cards typically come with excellent special effects software. There are many video cards available today. Most of these support various video in- a-window sizes, identification of source video, setup of play sequences are segments, special effects, frame grabbing, digital movie making; and some have built-in television tuners so you can watch your favorite programs in a window while working on other things. In windows, video overlay boards are controlled through the Media Control Interface. On the Macintosh, they are often controlled by external commands and functions (XCMDs and XFCNs) linked to your authoring software.

Good video greatly enhances your project; poor video will ruin it. Whether you delivered your video from tape using VISCA controls, from videodisc, or as a QuickTime or AVI movie, it is important that your source material be of high quality.

Projectors

When it is necessary to show a material to more viewers than can huddle around a computer monitor, it will be necessary to project it on to large screen or even a white painted wall. Cathode-ray tube (CRT) projectors, liquid crystal display (LCD) panels attached to an overhead projector, stand-alone LCD projectors, and light-valve projectors are available to splash the work on to big-screen surfaces.

CRT projectors have been around for quite a while- they are the original "big screen" televisions. They use three separate projection tubes and lenses (red, green, and blue), and three color channels of light must "converge" accurately on the screen. Setup, focusing, and aligning are important to getting a clear and crisp picture. CRT projectors are compatible with the output of most computers as well as televisions.

Printers

With the advent of reasonably priced color printers, hard-copy output has entered the multimedia scene. From storyboards to presentation to production of collateral marketing material, color printers have become an important part of the multimedia development environment. Color helps clarify concepts, improve understanding and retention of information, and organize complex data. As multimedia designers already know intelligent use of colors is critical to the success of a project. Tektronix offers both solid ink and laser options, and either Phases 560 will print more than 10000 pages at a rate of 5 color pages or 14 monochrome pages per minute before requiring new toner. Epson provides lower-cost and lower-performance solutions for home and small business users; Hewlett Packard's Color LaserJet line competes with both. Most printer manufactures offer a color model-just as all computers once used monochrome monitors but are now color, all printers will became color printers.

Communication Devices

Many multimedia applications are developed in workgroups comprising instructional designers, writers, graphic artists, programmers, and musicians located in the same office space or building. The workgroup members' computers typically are connected on a local area network (LAN). The client's computers, however, may be thousands of miles distant, requiring other methods for good communication. Communication among workshop members and with the client is essential to the efficient and accurate completion of project. And when speedy data transfer is needed, immediately, a modem or network is required. If the client and the service provider are both connected to the Internet, a combination of communication by e-mail and by FTP (File Transfer Protocol) may be the most cost-effective and efficient solution for both

creative development and project management. In the workplace, it is necessary to use quality equipment and software for the communication setup. The cost-in both time and money-of stable and fast networking will be returned to the content developer. 1.Modem, 2.Cable modem, 3.ISDN

Multimedia software tools-Basic Tools for Multimedia

The basic tools set for building multimedia project contains one or more authoring systems and various editing applications for text, images, sound, and motion video. A few additional applications are also useful for capturing images from the screen, translating file formats and tools for the making multimedia production easier.

Text Editing and Word Processing Tools

A word processor is usually the first software tool computer users rely upon for creating text. The word processor is often bundled with an office suite. Word processors such as Microsoft Word and WordPerfect are powerful applications that include spellcheckers, table formatters, thesauruses and prebuilt templates for letters, resumes, purchase orders and other common documents.

OCR Software

Often there will be multimedia content and other text to incorporate into a multimedia project, but no electronic text file. With optical character recognition (OCR) software, a flat-bed scanner, and a computer, it is possible to save many hours of rekeying printed words, and get the job done faster and more accurately than a roomful of typists.

OCR software turns bitmapped characters into electronically recognizable ASCII text. A scanner is typically used to create the bitmap. Then the software breaks the bitmap into chunks according to whether it contains text or graphics, by examining the texture and density of areas of the bitmap and by detecting edges. The text areas of the image are then converted to ASCII character using probability and expert system algorithms.

Image-Editing Tools

Image-editing application is specialized and powerful tools for enhancing and retouching existing bitmapped images. These applications also provide many of the feature and tools of painting and drawing programs and can be used to create images from scratch as well as images digitized from scanners, video frame-grabbers, digital cameras, clip art files, or original artwork files created with a painting or drawing package.

Here are some features typical of image-editing applications and of interest to multimedia developers:

- Multiple windows that provide views of more than one image at a time
- Conversion of major image-data types and industry-standard file formats
- Direct inputs of images from scanner and video sources
- Employment of a virtual memory scheme that uses hard disk space as RAM for images that require large amounts of memory
- Capable selection tools, such as rectangles, lassos, and magic wands, to select portions of a bitmap
- Image and balance controls for brightness, contrast, and color balance
- Good masking features
- Multiple undo and restore features
- Anti-aliasing capability, and sharpening and smoothing controls
- Color-mapping controls for precise adjustment of color balance
- Tools for retouching, blurring, sharpening, lightening, darkening, smudging, and tinting
- Geometric transformation such as flip, skew, rotate, and distort, and perspective changes

- Ability to resample and resize an image 134-bit color, 8- or 4-bit indexed color, 8-bit gray-scale, black-and-white, and customizable color palettes
- Ability to create images from scratch, using line, rectangle, square, circle, ellipse, polygon, airbrush, paintbrush, pencil, and eraser tools, with customizable brush shapes and user-definable bucket and gradient fills
- Multiple type faces, styles, and sizes, and type manipulation and masking routines
- Filters for special effects, such as crystallize, dry brush, emboss, facet, fresco, graphic pen, mosaic, pixel size, poster, ripple, smooth, splatter, stucco, twirl, watercolor, wave, and wind
- Support for third-party special effect plug-ins
- Ability to design in layers that can be combined, hidden, and reordered

Painting and Drawing Tools

Painting and drawing tools, as well as 3-D modelers, are perhaps the most important items in the toolkit because, of all the multimedia elements, the graphical impact of the project will likely have the greatest influence on the end user. If the artwork is amateurish, or flat and uninteresting, both the creator and the users will be disappointed. Painting software, such as Photoshop, Fireworks, and Painter, is dedicated to producing crafted bitmap images. Drawing software, such as CorelDraw, FreeHand, Illustrator, Designer, and Canvas, is dedicated to producing vector-based line art easily printed to paper at high resolution.

Some software applications combine drawing and painting capabilities, but many authoring systems can import only bitmapped images. Typically, bitmapped images provide the greatest choice and power to the artist for rendering fine detail and effects, and today bitmaps are used in multimedia more often than drawn objects. Some vector based packages such as Macromedia's Flash are aimed at reducing file download times on the Web, and may contain both bitmaps and drawn art. The anti-aliased character shown in the bitmap of Color Plate 5 is an example of the fine touches that improve the look of an image.

Sound Editing Tools

Sound editing tools for both digitized and MIDI sound lets hear music as well as create it. By drawing a representation of a sound in fine increments, whether a score or a waveform, it is possible to cut, copy, paste and otherwise edit segments of it with great precision.

System sounds are shipped both Macintosh and Windows systems and they are available as soon the Operating system is installed. For MIDI sound, a MIDI synthesizer is required to play and record sounds from musical instruments. For ordinary sound there are varieties of software such as Soundedit, MP3cutter, Wavestudio.

Animation, Video and Digital Movie Tools

Animation and digital movies are sequences of bitmapped graphic scenes (frames, rapidly played back. Most authoring tools adapt either a frame or object oriented approach to animation.

Moviemaking tools typically take advantage of Quicktime for Macintosh and Microsoft Video for Windows and lets the content developer to create, edit and present digitized motion video segments.

Analog video formats

- NTSC
- PAL
- SECAM

Digital Video Formats

These are MPEG13 based terrestrial broadcast video formats

- ATSC Standards

- DVB
- ISDB

These are strictly the format of the video itself, and not for the modulation used for transmission.

The framework supports the following file types and codecs natively:

Audio

- Apple Lossless
- Audio Interchange (AIFF)
- Digital Audio: Audio CD - 16-bit (CDDA), 134-bit, 313-bit integer & floating
- point, and 64-bit floating point
- MIDI
- MPEG-1 Layer 3 Audio (.mp3)
- MPEG-4 AAC Audio (.m4a, .m4b, .m4p)
- Sun AU Audio
- ULAW and ALAW Audio
- Waveform Audio (WAV)

Video

- 3GPP & 3GPP13 file formats
- AVI file format
- Bitmap (BMP) codec and file format
- DV file (DV NTSC/PAL and DVC Pro NTSC/PAL codecs)
- Flash & FlashPix files
- GIF and Animated GIF files
- H.1361, H.1363, and H.1364 codecs
- JPEG, Photo JPEG, and JPEG-13000 codecs and file formats
- MPEG-1, MPEG-13, and MPEG-4 Video file formats and associated codecs (such as AVC)
- QuickTime Movie (.mov) and QTVR movies
- Other video codecs: Apple Video, Cinepak, Component Video, Graphics, and Planar RGB
- Other still image formats: PNG, TIFF, and TGA

Other Presentation software includes:

- Adobe Persuasion
- AppleWorks
- Astound by Gold Disk Inc.
- Beamer (LaTeX)
- Google Docs which now includes presentations
- Harvard Graphics
- HyperCard
- IBM Lotus Freelance Graphics
- Macromedia Action!
- MagicPoint
- Microsoft PowerPoint
- OpenOffice.org Impress
- Scala Multimedia
- Screencast

- Umibozu (wiki)
- VCN ExecuVision

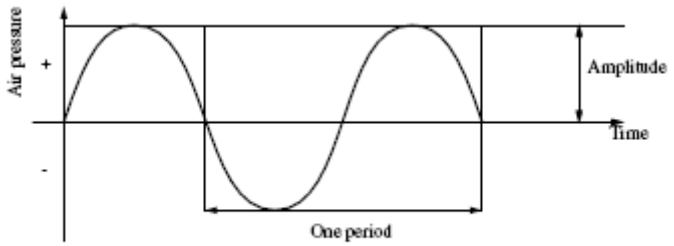
Multimedia Building blocks - Audio: Basic sound concepts Audio

Audiology is the discipline interested in manipulating acoustic signals that can be perceived by humans. Important aspects are psychoacoustics, music, the MIDI (Musical Instrument Digital Interface) standard, and speech synthesis and analysis. Most multimedia applications use audio in the form of music and/or speech, and voice communication is of particular significance in distributed multimedia applications.

In addition to providing an introduction to basic audio signal technologies and the MIDI standard, this chapter explains various enabling schemes, including speech synthesis, speech recognition, and speech transmission [Loy85, Fla72, FS92, Beg94, OS90', Fal85, Bri86, Ace93, Sch92]. In particular, it covers the use of sound, music, and speech in multimedia, for example, formats used in audio technology, and how audio material is represented in computers.covers storage of audio data (and other media data) on optical disks because this technology is not limited to audio signals.

What Is Sound?

Sound is a physical phenomenon caused by vibration of material, such as a violin string or a wood log. This type of vibration triggers pressure wave fluctuations in the air around the material. The pressure waves propagate in the air. The pattern of this oscillation. We hear a sound when such a wave reaches our ears.



Pressure wave oscillation in the air.

This wave form occurs repeatedly at regular intervals or periods. Sound waves have a natural origin, so they are never absolutely uniform or periodic. A sound that has a recognizable periodicity is referred to as music rather than sound, which does not have this behavior. Examples of periodic sounds are sounds generated by musical instruments, vocal sounds, wind sounds, or a bird's twitter. Non-periodic sounds are, for example, drums, coughing, sneezing, or the brawl or murmur of water.

Frequency

A sound's frequency is the reciprocal value of its period. Similarly, the frequency represents the number of periods per second and is measured in hertz (Hz) or cycles per second (cps). A common abbreviation is kilohertz (kHz), which describes 1,000 oscillations per second, corresponding to 1,000Hz.

Sound processes that occur in liquids, gases, and solids are classified by frequency range:

• Infrasonic: 0 to 20Hz
• Audiosonic: 20Hz to 20kHz
• Ultrasonic: 20kHz to 1GHz
• Hypersonic: 1GHz to 10THz

Sound in the audiosonic frequency range is primarily important for multimedia systems. In this text, we use audio as a representative medium for all acoustic signals in this frequency range. The waves in the audiosonic frequency range are also called acoustic signals. Speech is the signal humans generate by use of their speech organs. These signals can be reproduced by machines. For example, music signals have frequencies in the 20Hz to 20kHz range. We could add noise to speech and music as another type of audio signal. Noise is defined as a sound event without functional purpose, but this is not a dogmatic definition. For instance, we could add unintelligible language to our definition of noise.

Amplitude

A sound has a property called amplitude, which humans perceive subjectively as loudness or volume. The amplitude of a sound is a measuring unit used to deviate the pressure wave from its mean value (idle state).

Sound Perception and Psychoacoustics

The way humans perceive sound can be summarized as a sequence of events: Sound enters the ear canal. At the eardrum, sound energy (air pressure changes) are transformed into mechanical energy of eardrum movement. The outer ear comprises the pinna, which is composed of cartilage and has a relatively poor blood supply. Its presence on both sides of the head allows us to localize the source of sound from the front versus the back. Our ability to localize from side to side depends on the relative intensity and relative phase of sound reaching each ear and the analysis of the phase/intensity differences within the brainstem. The cochlea is a snail-shaped structure that is the sensory organ of hearing. The vibrational patterns that are initiated by vibration set up a traveling wave pattern within the cochlea. This wavelike pattern causes a shearing of the cilia of the outer and inner hair cells. This shearing causes hair cell depolarization resulting in on/off neural impulses that the brain interprets as sound. Psychoacoustics is a discipline that studies the relationship between acoustic waves at the auditory ossicle and the spatial recognition of the auditor.

Sound example	Sound pressure size
Rustling of paper	20 dB
Spoken language	60 dB
Heavy road traffic	80 dB
Rock band	120 dB
Pain sensitivity threshold	130 dB

Various sound pressure examples.

Speech Signals

Speech can be processed by humans or machines, although it is the dominant form of communication of human beings. The field of study of the handling of digitized speech is called digital speech processing.

Human Speech

Speech is based on spoken languages, which means that it has a semantic content. Human beings use their speech organs without the need to knowingly control the generation of sounds. (Other species such as bats also use acoustic signals to transmit information, but we will not discuss this here.) Speech understanding means the efficient adaptation to speakers and their speaking habits. Despite the large number of different dialects and emotional pronunciations,

we can understand each other's language. The brain is capable of achieving a very good separation between speech and interference, using the signals received by both ears. It is much more difficult for humans to filter signals received in one ear only. The brain corrects speech recognition errors because it understands the content, the grammar rules, and the phonetic and lexical word forms. Speech signals have two important characteristics that can be used by speech processing applications:

- Voiced speech signals (in contrast to unvoiced sounds) have an almost periodic structure over a certain time interval, so that these signals remain quasi-stationary for about 30ms.
- The spectrum of some sounds have characteristic maxima that normally involve up to five frequencies. These frequency maxima, generated when speaking, are called formants. By definition, a formant is a characteristic component of the quality of an utterance.

Audio File Format

The following extensions commonly used to lay up multimedia documentation:

MOV
MP4
3GP
VOB
FLV.

Files with augmentation MOV are used to lay up capture on film and song in order. MP4 is fundamentally identical to MOV format and lone differs by provided that roughly added metadata. MP4 put on record augmentation is supported by multiple applications with Apple ITunes, XBox 360. MPEG is a align of compressions methods designed for audio and visual data.3GP on PC may perhaps be viewed VLC media player, RealPlayer, QuickTime, GOM Player and Media Player Classic. File Extension VOB (Video Object) is commonly locate such documents in DVD-Video media. File Extension FLV is used to deposit Macromedia Flash Player collection. It can assign vector graphics, spill videocassette, audio and text.

Sound is analog in nature, and to be used in multimedia, needs to be digitized creates mood, interest, includes speech audio files are usually large files unless they have been compressed

Audio can be in 2 basic formats: a digitised file of the actual sound eg. WAV file or in compressed format MP3 MIDI (Musical Instrument Digital Interface) where details of the characteristics of each note is filed.

Digital Audio the quality of the sound depends on the sampling rate, sampling size, time and number of channels sampling rates are from 10 to 44 kHz with CD-ROM having a sampling rate of 33kHz slow rates result in loss of quality and distortion sampling size refers to how many bits are used to record (bit resolution)

Images and graphics: Basic concepts- Computer image processing

Still images are the important element of a multimedia project or a web site. In order to make a multimedia presentation look elegant and complete, it is necessary to spend ample amount of time to design the graphics and the layouts. Competent, computer literate skills in graphic art and design are vital to the success of a multimedia project.

Digital Image

A digital image is represented by a matrix of numeric values each representing a quantized intensity value. When I is a two-dimensional matrix, then I(r,c) is the intensity value at the position corresponding to row r and column c of the matrix.

The points at which an image is sampled are known as picture elements, commonly abbreviated as pixels. The pixel values of intensity images are called gray scale levels (we encode here the "color" of the image). The intensity at each pixel is represented by an integer and is determined from the continuous image by averaging over a small neighborhood around the pixel location. If there are just two intensity values, for example, black, and white, they are represented by the numbers 0 and 1; such images are called binary-valued images. If 8-bit integers are used to store each pixel value, the gray levels range from 0 (black) to 255 (white).

Digital Image Format

There are different kinds of image formats in the literature. We shall consider the image format that comes out of an image frame grabber, i.e., the captured image format, and the format when images are stored, i.e., the stored image format.

Captured Image Format

The image format is specified by two main parameters: spatial resolution, which is specified as pixelsxpixels (eg. 640x480)and color encoding, which is specified by bits per pixel. Both parameter values depend on hardware and software for input/output of images.

Stored Image Format

When we store an image, we are storing a two-dimensional array of values, in which each value represents the data associated with a pixel in the image. For a bitmap, this value is a binary digit.

Bitmaps

A bitmap is a simple information matrix describing the individual dots that are the smallest elements of resolution on a computer screen or other display or printing device. A one-dimensional matrix is required for monochrome (black and white); greater depth (more bits of information) is required to describe more than 16 million colors the picture elements may have, as illustrated in following figure. The state of all the pixels on a computer screen make up the image seen by the viewer, whether in combinations of black and white or colored pixels in a line of text, a photograph-like picture, or a simple background pattern.

Clip Art

A clip art collection may contain a random assortment of images, or it may contain a series of graphics, photographs, sound, and video related to a single topic. For example, Corel, Micrografx, and Fractal Design bundle extensive clip art collection with their image-editing software.

Vector Drawing

Most multimedia authoring systems provide for use of vector-drawn objects such as lines, rectangles, ovals, polygons, and text. Computer-aided design (CAD) programs have traditionally used vector-drawn object systems for creating the highly complex and geometric rendering needed by architects and engineers.

Graphic artists designing for print media use vector-drawn objects because the same mathematics that put a rectangle on your screen can also place that rectangle on paper without jaggies. This requires the higher resolution of the printer, using a page description language such as PostScript. Programs for 3-D animation also use vector-drawn graphics. For example, the various changes of position, rotation, and shading of light required to spin the extruded.

Image File Format

There are many file formats used to store bitmaps and vectored drawing. Following is a list of few image file formats.

JPG	(Joint Photographic Expert Groups). A lossy compression format designed to reduce the size of full colour bit images
GIF	(Graphics Interchange Format). Bit mapped format that uses a lossless compression algorithm. Limit of 256 colours.
PICT	Apple graphics format that is either bit mapped or vector graphics
TIFF	(Tagged Image File Format). A bit mapped format, standard choice for scanned images
PNG	A new file format that is becoming more popular. It supports 24 bit colour images, has an interlacing option and offers more powerful compression than JPG
EPS	(Encapsulated Postscript). Uses vector graphics and often will only be interpreted by a printer. Therefore it cannot be displayed on the screen
BMP	(Bit Map). An uncompressed bit mapped file format.

TEXT / REFERENCE BOOKS

1. Rafael C. Gonzalez, Richard E. Woods, ŏ" Digital Image Processing", Pearson , Second Edition, 2004.
2. David Saloman,"Data compression" , Springer International, 4th Edition.
3. Khalid Sayood, "Introduction To Data Compressio" ,Elsevier 3rd Edition.
4. Ralfsteinmetz and Klara Nahrstedt, "Multimedia Computing, Communications & Applications" Pearson Edn
5. Rajan Parekh, "Principles of Multimedia, Tata Mc Graw Hill.
6. Anil K. Jain, "Fundamentals of Digital Image Processin", Pearson 2002.
7. J F Koegel Buford- -Multimedia systems Addison Wesley
8. T Vaughan-,"Multimedia: Making it work" Tata Mc Graw Hill

PART A

1. Define Multimedia.
2. Give the applications of Multimedia.
3. What are the applications of Photographic Images?
4. What is the use of Document Images?
5. What are the properties of multimedia systems?
6. List the types of Multimedia tools.
7. What are the building blocks of multimedia?
8. Mention the major uses of Multimedia.
9. Mention the input devices of multimedia.
10. Give the storage devices of multimedia.

PART B

1. Illustrate the multimedia building blocks with examples.
2. Elaborate the memory and storage devices of multimedia hardware platforms.
3. Elaborate the input and output devices of multimedia hardware platforms.
4. Explain main properties of multimedia.
5. Explain the various audio formats supported by internet.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

UNIT – V – Digital Image and Multimedia Processing – SEC1605

V. AUDIO AND VIDEO COMPRESSION

Human Auditory system - WAVE audio format-Speech compression - MPEG4 Audio lossless coding(ALS) -VIDEO compression Introduction – Motion compensation - video signal representation, ITU -T Recommendation H.261, Model Based Coding - MPEG1 - MPEG2 H.262 , ITU-T Recommendations H.263, Advanced Video Coding, ATM Networks - Compressions issues in ATM networks, Compression Algorithms for packet videos.

Human Auditory system

The auditory system is the sensory system for the sense of hearing. It includes both the sensory organs (the ears) and the auditory parts of the sensory system.

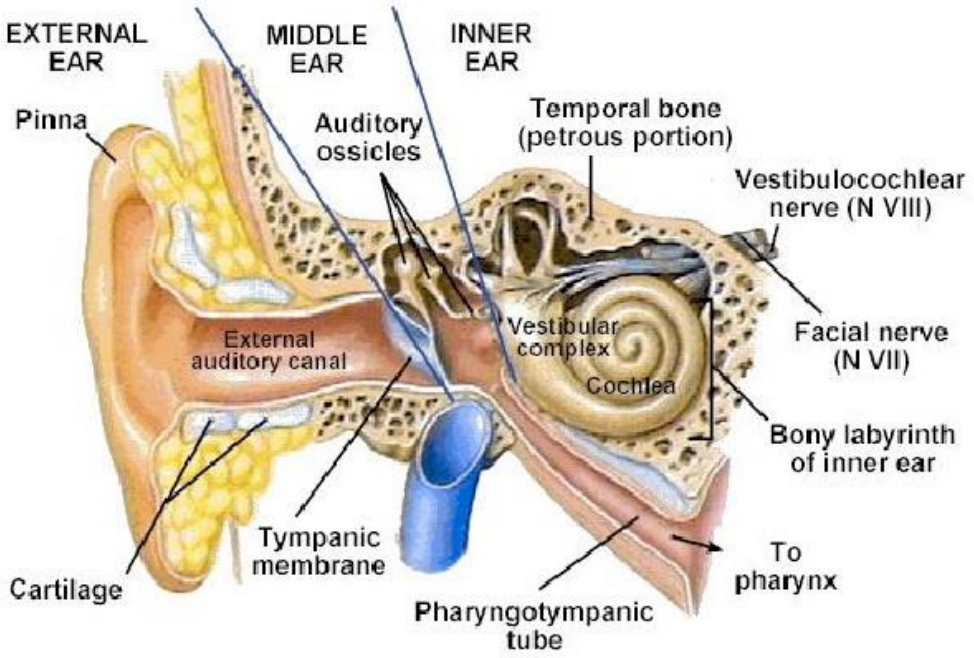
There are two fundamental areas of study

Sensation

Perception

Sensation is the process by which the sense organ (such as eyes and ear) gather information about environment Perception is a process by which the brain selects, organizes and interprets sensation

Sensation and perception are not isolated phenomenon but part of more general psychological process by which we gain knowledge about the world



Auditory Stimulus

In this presentation we will examine the sensation of audition in order to understand the same we must know what is an auditory stimuli ?

Sound (auditory Stimulus) are caused by rhythmic vibrations of air molecule.

There are three important characteristics of sound

- Pitch refers to frequency determined by number of cycles that occur in one second
- Loudness refers to amplitude determined by height of each sound wave and measure of how much air is expanding and compressing
- Timbre refers to purity or richness in the tone of the sound

Outer ear: collects sound waves
- amplifies sound waves in some frequencies
- vibrations of air are translated to vibrations of the tympanic membrane

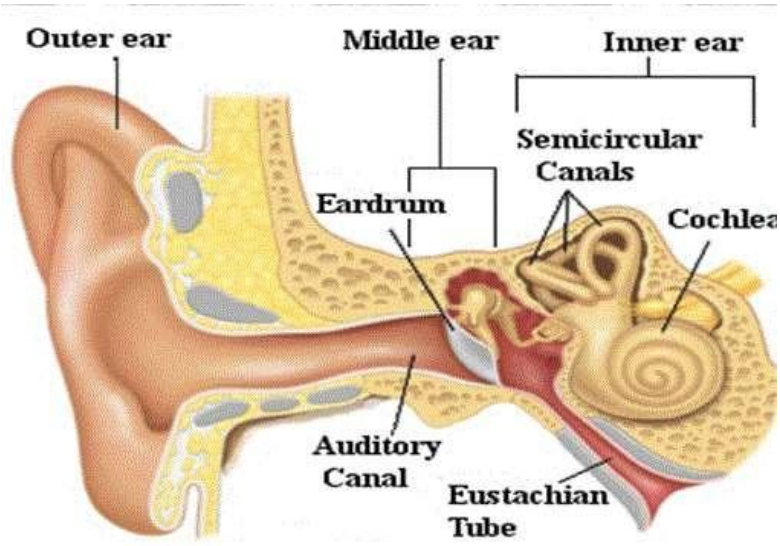
Middle ear: vibrations of tympanic membrane are translated to oscillations of liquid in inner ear
- performed by the ossicles
- amplification 15:1

Inner ear:
- Cochlea transforms mechanical vibrations into nerve impulses
  - Movement of basilar membrane causes the hairs to bend (3K hairs in cochlea)
  - What information is in the nerve impulses? Not well understood
- nerve endings (30K nerve fibers).

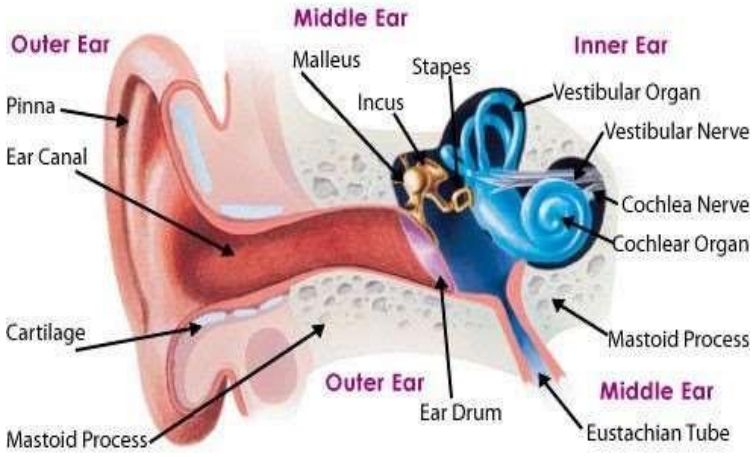
Outer ear: it channels sound waves through the ear canal to the eardrum
Middle ear: vibrations caused air pressure changes in the ear canal are transmitted to three small bones called "ossicles".
Inner ear: it houses the "cochlea", a spiralshaped structure that contains the organ of "Corti" the most important component of hearing. The Corti sits in an extremely sensitive membrane called the "basilar membrane". Whenever the basilar membrane vibrates, small sensory hair cells inside the Corti are bent, which stimulates the sending of nerve impulses to the brain.

Auditory periphery: The outer ear, middle ear, and inner ear, ending at the nerve fibers exiting the inner ear.
■ Auditory central nervous system: The ascending and descending auditory pathways in the brainstem and cortex.
■ Tonotopic organization: The systematic mapping of sound frequency to the place of maximum stimulation within the auditory system that begins in the cochlea and is preserved through the auditory cortex.
■ Transducer: A device or system that converts one form of energy to another. The cochlea can be considered a mechanoelectrical transducer because it converts mechanical vibrations to electrical energy to stimulate the afferent nerve fibers leading to the brainstem.

The outer ear consists of: the ear shell (pinna) function as sound wave reflectors and attenuators when the waves hit them. The pinna helps the brain identify the direction from where the sounds originated From the pinna, the sound waves enter a tube-like structure called auditory canal. This canal serves as a sound amplifier. The sound waves travel through the canal and reach the tympanic membrane (eardrum), the canal’s end.

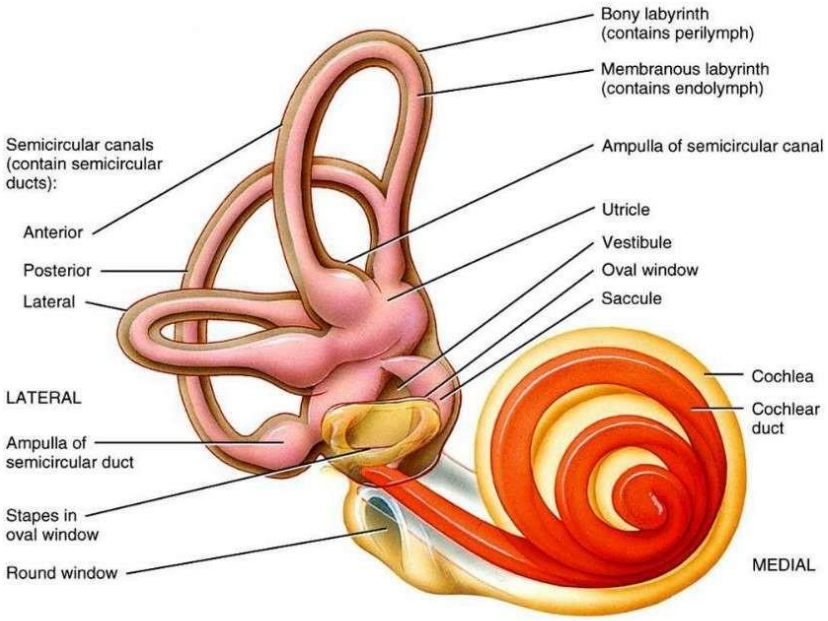


Middle Ear: As the sound waves hit the eardrum, the sensory information goes into an air-filled cavity through lever-teletype bones called ossicles. The three ossicles include the hammer (malleus) anvil (incus) stirrup (stapes). The middle ear acts as an impedance- matching device that improves sound transmission, reduces the amount of reflect sound and protects the inner ear from excessive sound pressure levels. This protection is actively regulated by the brain using the middle ear's muscles to tense and un-tense the bone structure.



Inner Ear

Inner ear consist of two different portion : Vestibular portion(balance) Auditory portion(Cochlea) Vestibular portion works in conjunction with eye and receptor cell in the joints of body to continuously maintain our balance It has three semi circular canals that has swelling at one end called ampulla and two membranal sac called utricle and saccule. These sac contains fluid and sensory cell which move in response to head movement in travel to indicate body status



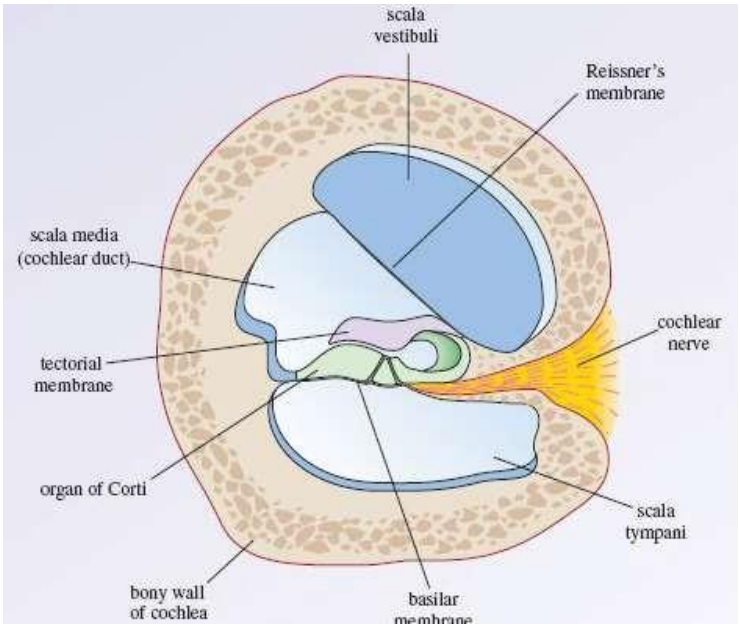
(a) Components of the right internal ear

Auditory portion comprises of coiled tube called cochlea duct in a shape of a snail, wrapped around the acoustic portion of auditory nerve.

The tube of the cochlea is divided into three chambers:

- the scala vestibuli
- the scala media (or cochlear duct)
- the scala tympani

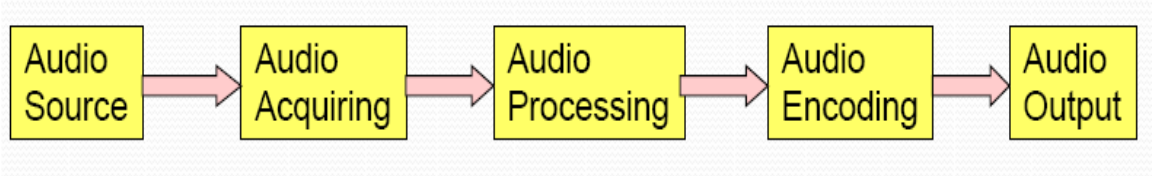
Scala vestibuli and scala media are separated by Reissner's membrane Scala media and scala tympani is separated by basiliar membrane.The organ of Corti is the primary auditory receptor structure and houses the sensory receptor cells which are known as hair cells. These hair cells have tiny hair like projections called stereocilia. Hair cells are of two types: outer hair cells and inner hair cells.



When sound energy is transferred to cochlea the basiliar membrane vibrates causing sheering action above tectorial membrane that causes stereocilia to bend which opens ion gates leading to chemical change and generation on electrical charges inside these cells. These charges make neural impulse that travel along the auditory nerve to the brain which is interpreted as sound.
Once the sound waves are turned into neural signals, they travel through cranial nerve VIII, reaching different anatomical structures where the neural information is further processed. The cochlear nucleus is the first site of neural processing, followed by the superior olivary complex located in the pons, and then processed in the inferior colliculus at the midbrain. The neural information ends up at the relay centre of the brain, called the thalamus. The info is then passed to the primary auditory cortex of the brain, situated in the temporal lobe.
The primary auditory cortex receives auditory information from the thalamus. The left posterior superior temporal gyrus is responsible for the perception of sound

WAVE audio format
An audio file format is a file format for storing digital audio data on a computer system. This data can be stored uncompressed, or compressed to reduce the file size. It can be a raw bitstream, but it is usually a container format or an audio data format with defined storage layer.
Audio
Format/Codec

- It is important to distinguish between a file format and an audio codec.
- A codec performs the encoding and decoding of the raw audio data while the data itself is stored in a file with a specific audio file format.
- In other words, Codec contains both an ADC and DAC running off the same clock.



Uncompressed: Audio files that are not compressed and are capable of having a large file size. Ex) .wav Waveform Extension.

There is one major uncompressed audio format, PCM, which is usually stored in a .wav file on Windows or in a .aiff file on Mac OS.

- Pulse-code modulation (PCM) is a method used to digitally represent sampled analog signals.
- BWF (Broadcast Wave Format) is a standard audio format created by the E.B.U.as a successor to WAV.
- BWF allows metadata to be stored in the file.

WAV and AIFF are flexible file formats designed to store more or less any combination of sampling rates or bitrates. This makes them suitable file formats for storing and archiving an original recording.
WAVE means storing a pattern of music signals in digital wave form on computer.
This means that any sound can be stored.
However, these sound files are huge and process over 160KB per second.
These files can be seen with a suffix of .wav (e.g. Fred.wav)

- Very good sound quality.
- Widely supported in many browsers with no need for a plugin.
- You can record your own .wav files from a CD, tape, microphone, etc.
- The very large file sizes severely limit the length of the sound clips that you can use on your Web pages.

Speech compression
Data Rates
Telephone quality voice: 8000 samples/sec, 8 bits/sample, mono 64Kb/s
CD quality audio: 44100 samples/sec, 16 bits/sample, stereo ~1.4Mb/s
Communications channels and storage cost money (although less than they used to)
PCM - send every sample
DPCM - send differences between samples
ADPCM - send differences, but adapt how we code them
SB-ADPCM - wideband codec, use ADPCM twice, once for lower frequencies, again at lower bitrate for upper frequencies.
LPC - linear model of speech formation
CELP - use LPC as base, but also use some bits to code corrections for the things LPC gets wrong.
PCM

- μ-law and a-law PCM have already reduced the data sent.
- Lost frequencies above 4KHz.
- Non-linear encoding to reduce bits per sample.

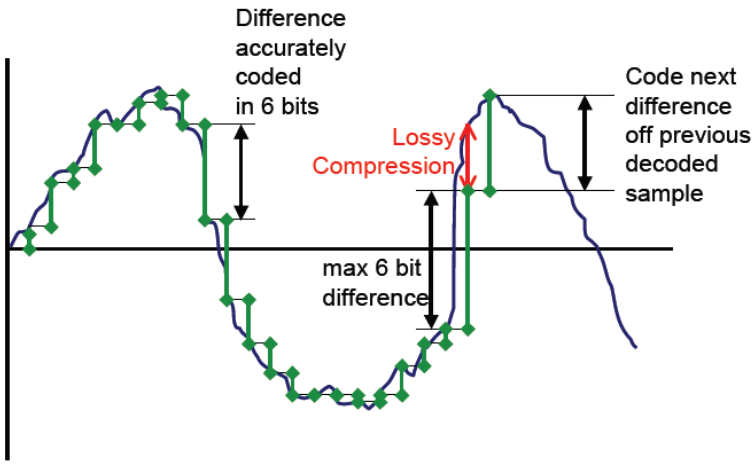
However, each sample is still independently encoded.

- In reality, samples are correlated.
- Can utilize this correlation to reduce the data sent.

Differential PCM

- Normally the difference between samples is relatively small and can be coded with less than 8 bits.
- Simplest codec sends only the differences between samples.
  - Typically use 6 bits for difference, rather than 8 bits for absolute value.
- Compression is lossy, as not all differences can be coded
  - Decoded signal is slightly degraded.
  - Next difference must then be encoded off the previous decoded sample, so losses don’t accumulate.

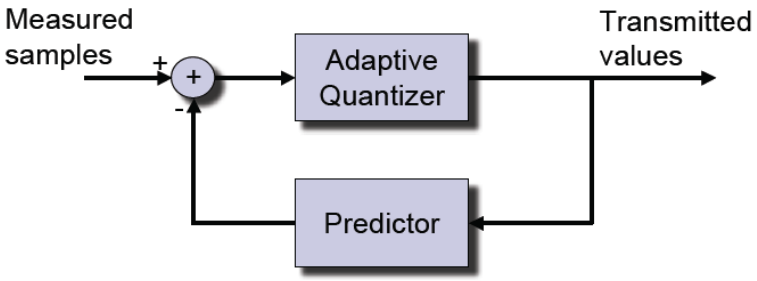
Differential PCM



ADPCM (Adaptive Differential PCM)

- Makes a simple prediction of the next sample, based on weighted previous n samples.
- For G.721, previous 8 weighted samples are added to make the prediction.
- Lossy coding of the difference between the actual sample and the prediction.
- Difference is quantized into 4 bits ⇒ 32Kb/s sent.
- Quantization levels are adaptive, based on the content of the audio.
- Receiver runs same prediction algorithm and adaptive quantization levels to reconstruct speech.

ADPCM



ADPCM

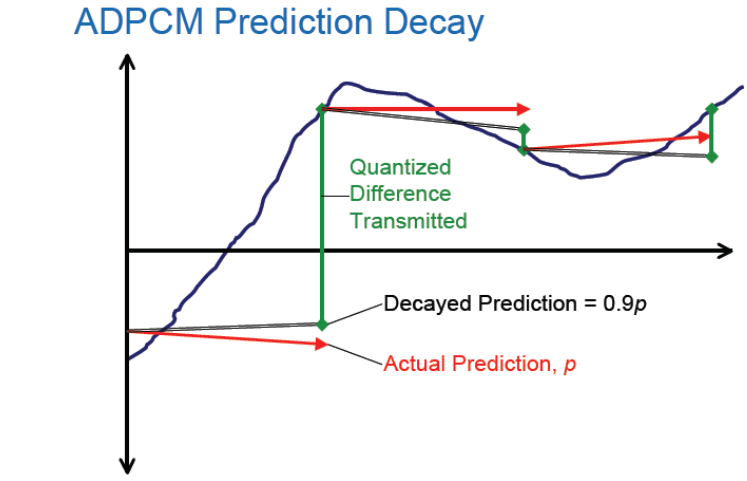
- Adaptive quantization cannot always exactly encode a difference.
- Shows up as quantization noise.
- Modems and fax machines try to use the full channel capacity.
- If they succeed, one sample is not predictable from the next.
- ADPCM will cause them to fail or work poorly.
- ADPCM not normally used on national voice circuits, but commonly used internationally to save capacity on expensive satellite or undersea fibres.
- May attempt to detect if it's a modem, and switch back to regular PCM.

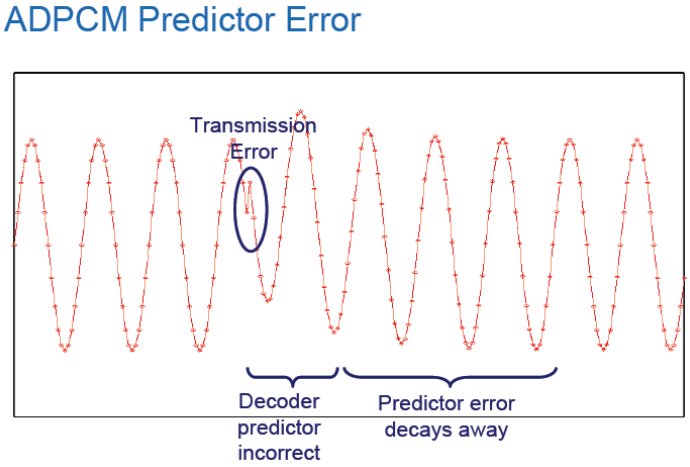
Predictor Error

- What happens if the signal gets corrupted while being transmitted?
- Wrong value will be decoded.
- Predictor will be incorrect.
- All future values will be decoded incorrectly!
- Modern voice circuits have low but non-zero error rates.
- But ADPCM was used on older circuits with higher loss rates too.

ADPCM Predictor Error

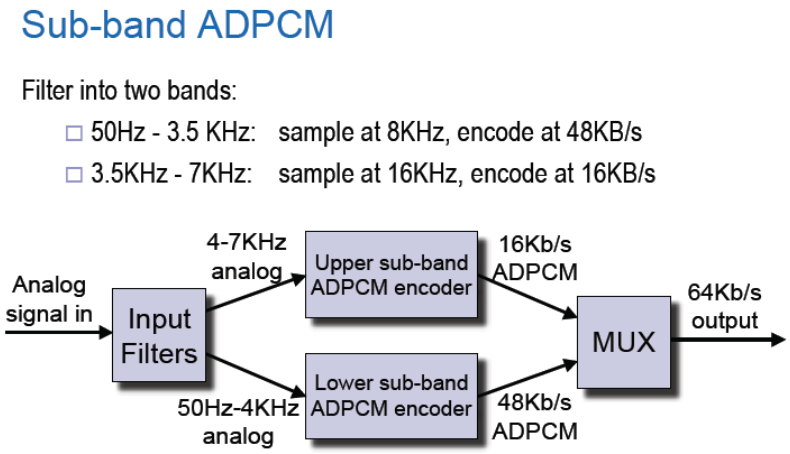
- Want to design a codec so that errors do not persist.
- Build in an automatic decay towards zero.
- If only differences of zero were sent, the predictor would
- decay the predicted (and hence decoded) value towards
- zero.
- Differences have a mean value of zero (there are as many positive differences as negative ones).
- Thus predictor decay ensures that any error will also decrease over time until it disappears.





Sub-band ADPCM

- Regular ADPCM reduces the bitrate of 8KHz sampled audio (typically 32Kb/s).
- If we have a 64Kb/s channel (eg ISDN), we could use the same techniques to produce better that toll-quality.
- Could just use ADPCM with 16KHz sampled audio, but not all frequencies are of equal importance.
- 0-3.5KHz important for intelligibility
- 3.5-7KHz helps speaker recognition and conveys emotion
- Sub-band ADPCM codes these two ranges separately.



Sub-band ADPCM

Practical issue:

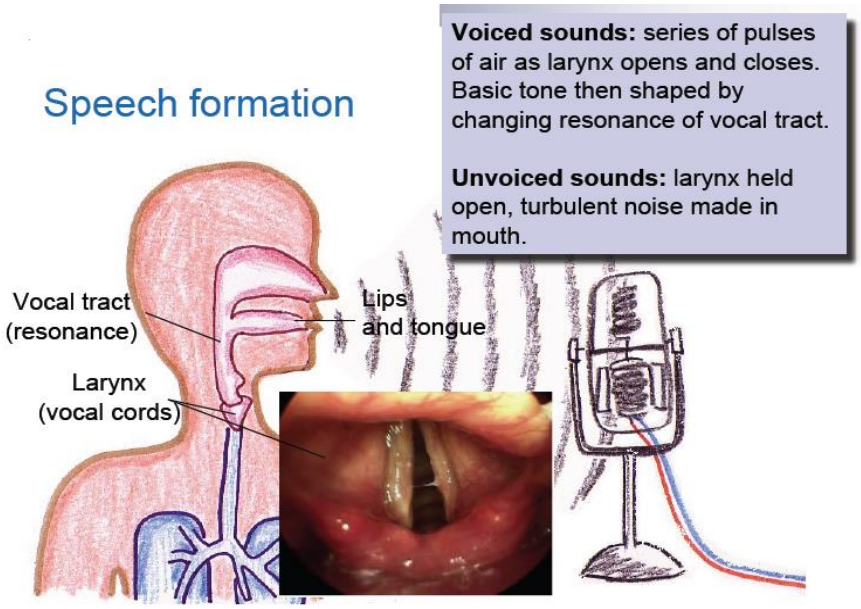
- Unless you have dedicated hardware, probably can't sample two sub-bands separately at the same time.
- Need to process digitally.
- Sample at 16KHz.
- Use digital filters to split sub-bands and down sample the lower sub-band to 8KHz.

Key point of Sub-band ADPCM:

- Not all frequencies are of equal importance (quantization noise is more disruptive to some parts of the signal than others)
- Allocate the bits where they do most good.

Model-based Coding

- PCM, DPCM and ADPCM directly code the received audio signal.
- An alternative approach is to build a *parameterized model of the sound source* (ie. Human voice).
- For each time slice (eg 20ms):
- Analyse the audio signal to determine how the signal was produced.
- Determine the model parameters that fit.
- Send the model parameters.
- At the receiver, synthesize the voice from the model and received parameters.

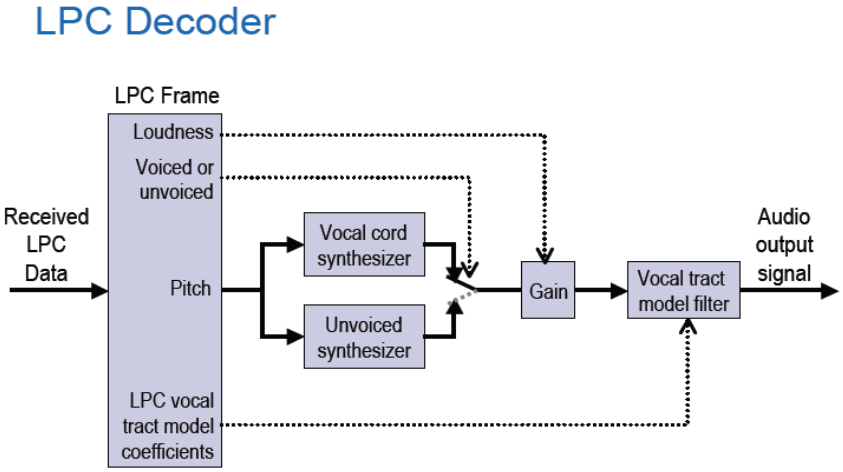


Linear Predictive Coding (LPC)

- Introduced in 1960s.
- Low-bitrate encoder:
- 1.2Kb/s - 4Kb/s
- Sounds very synthetic
- Basic LPC mostly used where bitrate really matters (eg in miltary applications)
- Most modern voice codecs (eg GSM) are based on enhanced LPC encoders.

LPC

- Digitize signal, and split into segments (eg 20ms)
- For each segment, determine:
- Pitch of the signal (ie basic formant frequency)
- Loudness of the signal.
- Whether sound is voiced or unvoiced
- Voiced: vowels, “m”, “v”, “l”
- Unvoiced: “f”, “s”
- Vocal tract excitation parameters (LPC coefficients)



LPC Decoder

- Vocal chord synthesizer generates a series of impulses.
- Unvoiced synthesizer is a white noise source.
- Vocal tract model uses a linear predictive filter.
- *n*th sample is a linear combination of the previous *p* samples plus an error term:
- $xn = a1xn\text{-}1 + a2xn\text{-}2 + .. + anxn\text{-}p + en$ *en* comes from the synthesizer.
- The coefficients *a*1.. *ap* comprise the vocal tract model, and shape the synthesized sounds.

LPC Encoder

- Once pitch and voice/unvoiced are determined, encoding consists of deriving the optimal LPC coefficients (*a*1.. *ap*) for the vocal tract model so as to minimize the mean-square error between the predicted signal and the actual signal.
- Problem is straightforward in principle. In practice it involves:
  1. The computation of a matrix of coefficient values.
  2. The solution of a set of linear equations.
- Several different ways exist to do this efficiently (autocorrelation, covariance, recursive latice formulation) to assure convergence to a unique solution.

Limitations of LPC Model

- LPC linear predictor is very simple.
- For this to work, the vocal tract "tube" must not have any side branches (these would require a more complex model).
- OK for vowels (tube is a reasonable model)
- For nasal sounds, nose cavity forms a side branch.
- In practice this is ignored in pure LPC.
- More complex codecs attempt to code the residue signal, which helps correct this.

Code Excited Linear Prediction (CELP)

- Goal is to efficiently encode the residue signal, improving speech quality over LPC, but without increasing the bit rate too much.
- CELP codecs use a codebook of typical residue values.
- Analyzer compares residue to codebook values.
- Chooses value which is closest.
- Sends that value.
- Receiver looks up the code in its codebook, retrieves the residue, and uses this to excite the LPC formant filter.
- Problem is that codebook would require different residue values for every possible voice pitch.
- Codebook search would be slow, and code would require a lot of bits to send.
  1. One solution is to have two codebooks.
  2. One fixed by codec designers, just large enough to represent one pitch period of residue.
- One dynamically filled in with copies of the previous residue delayed by various amounts (delay provides the pitch)
- CELP algorithm using these techniques can provide pretty good quality at 4.8Kb/s.

Enhanced LPC Usage

- GSM (Groupe Speciale Mobile)
- Residual Pulse Excited LPC13Kb/s
- LD-CELP

- Low-delay Code-Excited Linear Prediction (G.728) 16Kb/s
- CS-ACELP
- Conjugate Structure Algebraic CELP (G.729) 8Kb/s
- MP-MLQ Multi-Pulse Maximum Likelihood Quantization (G.723.1) 6.3Kb/s

MPEG4 Audio lossless coding (ALS)

The lossless data compression method essentially has two steps: Analyze the files and then eliminate the redundant data found within them.

- For example, if a file compressor analyzed and eliminated all the repeated words in a document file, the result would be a document with about 60 percent fewer words. Such is the case with compressed files. Your application analyzes the file and removes all the equivalent superfluous data bits, and shrinks the overall size of the file.
- However, if you attempted to read the article with the omitted words, it wouldn't make any sense. Therefore, the file compression applications insert placeholders where those eliminated words were.
- When you extract the file, the application automatically restores the repeated words to their places, making the file readable. Because no data is lost, this method is called lossless compression.
- MPEG-4
- Designed specially for the Internet.
- Provides greater audio and video interactivity than previous MPEG versions.
- It allows developers to control objects independently in a scene.
- MPEG-4 includes the capability of representing natural and synthesized sound and also support natural textures, images, photograph, natural video and animated video.

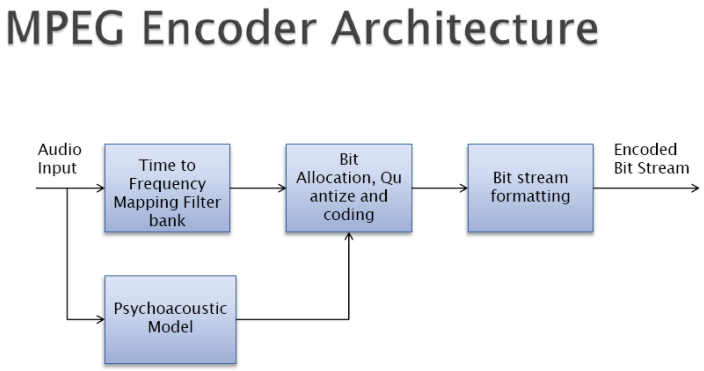
MPEG-4AAC is another audio compression standard under ISO/IEC 14496. MPEG- 4 audio integrates several different audio components into one standard: speech compression, perceptually based coders, text-to-speech, 3D localization of sound, and MIDI. MPEG-4 can be classified into MPEG-4 Scalable Lossless Coding (HD AAC) [14] and MPEG-4 (HE AAC) [14] . While MPEG-4 HD (High Definition) AAC is used for lossless high quality audio compression for High Definition videos, etc., MPEG-4 HE (High Efficiency) AAC is an extension of the Low complexity MPEG-2 AAC profile used for low bitrate applications such as streaming audio. MPEG-4 HE AAC has two versions: HE AAC v1, which uses only Spectral Band Replication (SBR, enhancing audio at lowbit rates) andHEAACv2, which uses SBR and Parametric Stereo (PS, enhancing efficiency of low bandwidth input). MPEG- 4 HE AAC is also used for the digital radio standards DAB+, developed by the standards group WorldDMB (Digital Multimedia Broadcasting) in 2006, and in Digital Radio Mondiale, a consortium of national radio stations aimed at making better use of the bands currently used for AM broadcasting, including shortwave.

Perceptual Coders

One change in AAC in MPEG-4 is to incorporate a Perceptual Noise Substitution module, which looks at scale factor bands above 4 kHz and includes a decision as to whether they are noiselike or tonelike. A noise like scale factor band itself is not transmitted; instead, just its energy is transmitted, and the frequency coefficient is set to zero. The decoder then inserts noise with that energy.

Another modification is to include a Bit-Sliced Arithmetic Coding (BSAC) module. This is an algorithm for increasing bitrate scalability, by allowing the decoder side to be able to decode a 64 kbps stream using only a 16 kbps baseline output (and steps of 1 kbps from that minimum). MPEG-4 audio also includes a second perceptual audio coder, a vector quantization method

entitled Transform-domain Weighted Interleave Vector Quantization (TwinVQ). This is aimed at low bitrates and allows the decoder to discard portions of the bitstream to implement both adjustable bitrate and sampling rate. The basic strategy of MPEG-4 audio is to allow decoders to apply as many or as few audio tools as bandwidth allows.



Structured Coders

To have a low bitrate delivery option, MPEG-4 takes what is termed a Synthetic/ Natural Hybrid Coding (SNHC) approach. The objective is to integrate both "natural" multimedia sequences, both video and audio, with those arising synthetically.

In audio, the latter are termed structured audio. The idea is that for low bit rate operation, we can simply send a pointer to the audio model we are working with and then send audio model parameters.

In video, such a model-based approach might involve sending face animation data rather than natural video frames of faces. In audio, we could send the information that English is being modeled, then send codes for the base sounds (phonemes) of English, along with other assembler-like codes specifying duration and pitch.

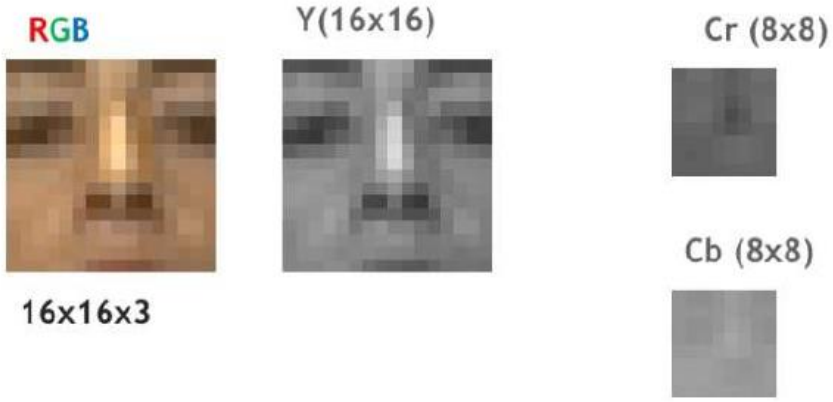
MPEG-4 takes a toolbox approach and allows specification of many such models. For example, Text-To-Speech (TTS) is an ultra-low bit rate method and actually works, provided we need not care what the speaker actually sounds like. Assuming we went on to derive Face Animation Parameters from such low bit rate information, we arrive directly at a very low bitrate videoconferencing system. Another "tool" in structured audio is called Structured Audio Orchestra Language (SAOL, pronounced "sail"), which allows simple specification of sound synthesis, including special effects such as reverberation. Overall, structured audio takes advantage of redundancies in music to greatly compress sound descriptions.

VIDEO compression Introduction – Motion compensation

- Once a video signal is digital, it requires a large amount of storage space and transmission bandwidth.
- To reduce the amount of data, several strategies are employed that compress the information without negatively affecting the quality of the image.
- Storing and transmitting uncompressed raw video is not an efficient technique because it needs large amounts of storage and bandwidth.
- Digital Versatile Disk (DVD), DSS, and internet video, all use digital data because it take a lot of space to store and large bandwidth to transmit
- Video compression technique is used to compress the data for these applications because it less storage space and less bandwidth to transmit data.

- With efficient compression techniques, a significant reduction in file size can be achieved with little or no adverse effect on the visual quality. The video quality can be affected if the file size is further lowered by raising the compression level for a given compression technique.
- Videos are sequences of images displayed at a high rate. Each of these images is called a frame.
- Human eye can not notice small changes in the frames such as a slight difference in color.
- Typically 30 frames are displayed on the screen every second.
- video compression standards do not require the encoding of all the details and some of the less important video details are lost because lossy compression is used due to its ability to get very high compression ratios.
- less efficient during sequences of fast movement because fewer MBs in the same position from frame to frame.
- In fact, users may note video artifacts during these sequences if the file is over compressed.
- To accomplish this, an application known as a "codec" analyzes the video frame by frame, and breaks each frame down into square blocks known as "macro blocks."
- One macro block(MB) consists of four pixels. Typically, the codec then analyzes each frame, checking for changes in the MBs.
- Areas where the MBs do not change for several frames in a row are noted and further analyzed.
- If the video compression codec determines that these areas can be removed from some of the frames, it does so, thus reducing overall file size.

Marco block (MB) and Block



Three types of frame

- Intra frame (I)
  Typically about 12 frames between 1 frame every MB of the frame is coded using spatial redundancy
- Predictive frame (P)
  Encode from previous I or P reference frame most of the MBs of the frame are coded exploiting temporal redundancy in the past
- Bi-directional frames (B)
  Encode from previous and future I or P frames most of the MBs of the frame are coded exploiting temporal redundancy in the past and in the future

ITU -T Recommendation H.261

H.261 is an earlier digital video compression standard. Because its principle of motion-compensation-based compression is very much retained in all later video compression standards, we will start with a detailed discussion of H.261. The International Telegraph and Telephone Consultative Committee (CCITT) initiated the development of H.261 in 1988. The final recommendation was adopted by the ITU (International Telecommunication Union)—Telecommunication standardization sector (ITU-T), formerly CCITT, in 1990 [6]. The standard was designed for videophone, videoconferencing, and other audiovisual services over ISDN telephone lines. Initially, it was intended to support multiples (from 1 to 5) of 384 kbps channels. In the end, however, the video codec supports bitrates of p × 64 kbps, where p ranges from 1 to 30. Hence, the standard was once known as p ∗ 64, pronounced "p star 64". The standard requires video encoder delay to be less than 150 ms, so that the video can be used for real-time, bidirectional video conferencing. H.261 belongs to the following set of ITU recommendations for visual telephony systems:

• H.221. Frame structure for an audiovisual channel supporting 64–1,920 kbps
• H.230. Frame control signals for audiovisual systems
• H.242. Audiovisual communication protocols
• H.261. Video encoder/decoder for audiovisual services at p × 64 kbps
• H.320. Narrowband audiovisual terminal equipment for p × 64 kbps transmission.

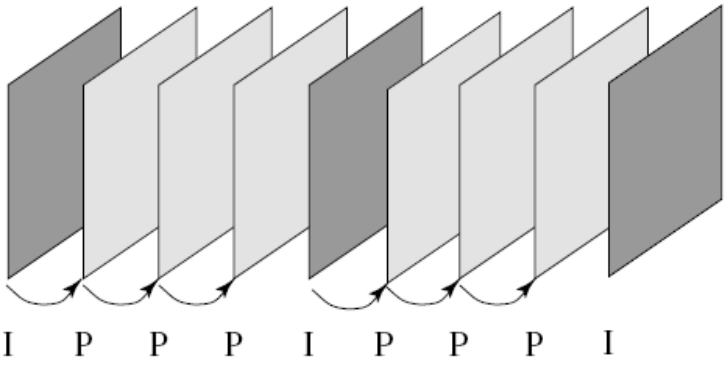
Table 10.2 lists the video formats supported by H.261. Chroma sub sampling in H.261 is 4:2:0. Considering the relatively low bit rate in network communications at the time, support for CCIR 601 QCIF is specified as required, whereas support for CIF is optional.

Figure illustrates a typical H.261 frame sequence. Two types of image frames are defined: intra-frames (I-frames) and inter-frames (P-frames). I-frames are treated as independent images. Basically, a transform coding method similar to JPEG is applied within each I-frame, hence the name "intra". P-frames are not independent. They are coded by a forward predictive coding method in which current macro blocks are predicted from similar macro blocks in the preceding I- or P-frame, and differences between the macro blocks are coded. Temporal redundancy removal is hence included in P-frame coding, whereas I-frame coding performs only spatial redundancy removal. It is important to remember that the prediction from a previous P-frame is allowed (not just from a previous I-frame). The interval between pairs of I-frames is a variable and is determined by the encoder. Usually, an ordinary digital video has a couple of I-frames per second. Motion vectors in H.261 are always measured in units of full pixels and have a limited range of ±15 pixels—that is, p = 15.

Table Video formats supported by H.261

Video format	Luminance image resolution	Chrominance image resolution	Bitrate (Mbps) (if 30 fps and uncompressed)	H.261 support
QCIF	176 × 144	88 × 72	9.1	Required
CIF	352 × 288	176 × 144	36.5	Optional

Fig. H.261 frame sequence



Model Based Coding - MPEG1 - MPEG2 H.262, ITU-T Recommendations H.263,

MPEG-1

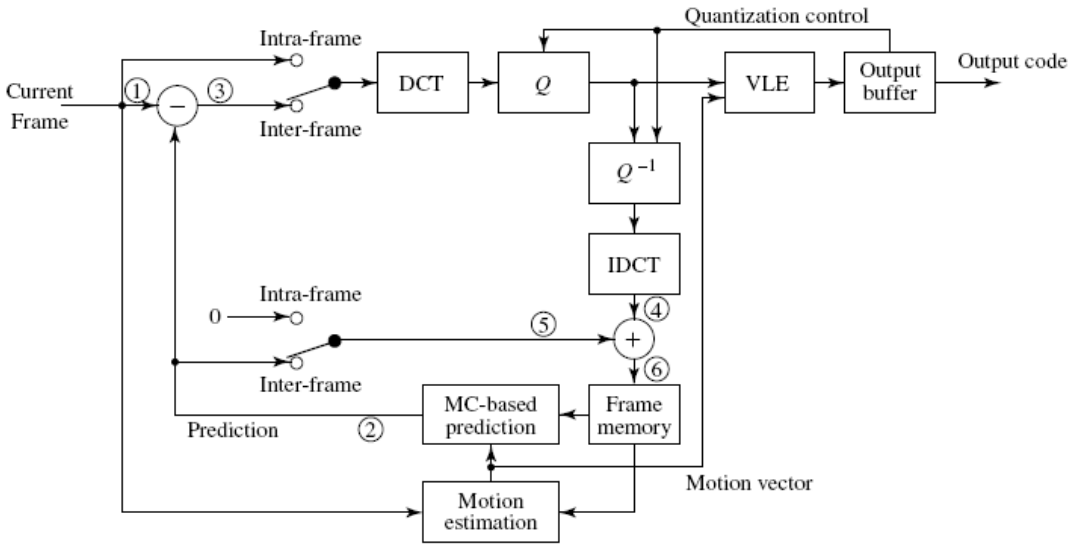
The MPEG-1 audio/video digital compression standard was approved by the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) MPEG group in November 1991 for Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s. Common digital storage media include compact discs (CDs) and video compact discs (VCDs). Out of the specified 1.5, 1.2 Mbps is intended for coded video, and 256 kbps can be used for stereo audio. This yields a picture quality comparable to VHS cassettes and a sound quality equal to CD audio.

In general, MPEG-1 adopts the CCIR601 digital TV format, also known as Source InputFormat (SIF).MPEG-1supports only noninterlaced video. Normally, its picture resolution is 352 × 240 for NTSC video at 30 fps, or 352 × 288 for PAL video at 25 fps. It uses 4:2:0 chroma subsampling.

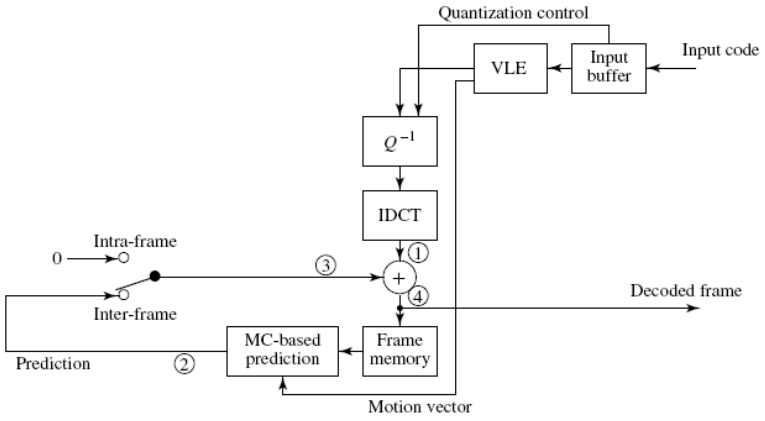
The MPEG-1 standard, also referred to as ISO/IEC 11172 [4], has five parts:

11172-1 Systems, 11172-2 Video, 11172-3 Audio, 11172-4 Conformance, and 11172-5 Software. Briefly, Systems takes care of, among many things, dividing output into packets of bitstreams, multiplexing, and synchronization of the video and audio streams. Conformance (or compliance) specifies the design of tests for verifying whether a bitstream or decoder complies with the standard. Software includes a complete software implementation of the MPEG-1 standard decoder and a sample software implementation of an encoder.

As in H.261 and H.263, MPEG-1 employs Hybrid Coding, i.e., a combination of motion compensation and transform coding on prediction residual errors. We will examine the main features of MPEG-1 video coding and leave discussions of MPEG audio coding.



(a) Encoder



(b) Decoder

Motion Compensation in MPEG-1

As discussed in the last chapter, motion-compensation-based video encoding in H.261 works as follows: In motion estimation, each macroblock of the target Pframe is assigned the best matching macroblock from the previously coded I- or P-frame. This is called a prediction. The difference between the macro block and its matching macro block is the prediction error, which is sent to DCT and its subsequent encoding steps. Since the prediction is from a previous frame, it is called forward prediction. Due to unexpected movements and occlusions in real scenes, the target macro block may not have a good matching entity in the previous frame. Figure illustrates that the macro block containing part of a ball in the target frame cannot find a good matching macro block in the previous frame, because half of the ball was occluded by another object. However, a match can readily be obtained from the next frame.

MPEG introduces a third frame type—B-frames—and their accompanying bidirectional motion compensation. Figure illustrates the motion-compensation based B-frame coding idea.

H.263

H.263 is an improved video coding standard for videoconferencing and other audiovisual services transmitted on Public Switched Telephone Networks (PSTN).

It aims at low bitrate communications at bitrates of less than 64 kbps. It was adopted by the

ITU-T Study Group 15 in 1995. Similar to H.261, it uses predictive coding for inter-frames, to reduce temporal redundancy, and transform coding for the remaining signal, to reduce spatial redundancy (for both intra-frames and difference macro blocks from inter-frame prediction).
In addition to CIF and QCIF, H.263 supports sub-QCIF, 4CIF, and 16CIF. Table summarizes video formats supported by H.263. If not compressed and assuming 30 fps, the bitrate for high-resolution videos (e.g., 16CIF) could be very high (>500 Mbps). For compressed video, the standard defines maximum bit rate per picture (BPP maxKb), measured in units of 1,024 bits. In practice, a lower bit rate for compressed H.263 video can be achieved. As in H.261, the H.263 standard also supports the notion of group of blocks. The difference is that GOBs in H.263 do not have a fixed size, and they always start and end at the left and right borders of the picture. As Fig. 10.10 shows, each QCIF luminance image consists of 9 GOBs and each GOB has 11 × 1 MBs (176 × 16 pixels), whereas each 4CIF luminance image consists of 18 GOBs and each GOB has 44 × 2 MBs (704 × 32 pixels).

Motion Compensation in H.263

The process of motion compensation in H.263 is similar to that of H.261. The motion vector (MV) is, however, not simply derived from the current macro block. The horizontal and vertical components of the MV are predicted from the median values of the horizontal and vertical components, respectively, of MV1, MV2, and MV3 from the "previous," "above," and "above and right" macro blocks. Namely, for the macro block with MV(u, v),

Video format	Luminance image resolution	Chrominance image resolution	Bitrate (Mbps) (if 30 fps and uncompressed)	Bitrate (kbps) BPPmaxKb (compressed)
Sub-QCIF	128 × 96	64 × 48	4.4	64
QCIF	176 × 144	88 × 72	9.1	64
CIF	352 × 288	176 × 144	36.5	256
4CIF	704 × 576	352 × 288	146.0	512
16CIF	1408 × 1152	704 × 576	583.9	1024

The unrestricted motion vector mode is redefined under H.263+. It uses Reversible Variable Length Coding (RVLC) to encode the difference motion vectors. The RVLC encoder is able to minimize the impact of transmission error by allowing the decoder to decode from both forward and reverse directions. The range of motion vectors is extended again to [−256, 256].

• A slice structure is used to replace GOB for additional flexibility. A slice can contain a variable number of macro blocks. The transmission order can be either sequential or arbitrary, and the shape of a slice is not required to be rectangular.

• H.263+ implements Temporal, SNR, and Spatial scalabilities. Scalability refers to the ability to handle various constraints, such as display resolution, bandwidth, and hardware capabilities. The enhancement layer for Temporal scalability increases perceptual quality by inserting B-frames between two P-frames.

SNR scalability is achieved by using various quantizers of smaller-and-smaller step size to encode additional enhancement layers into the bit stream. Thus, the decoder can decide how many enhancement layers to decode according to computational or network constraints. The concept of Spatial scalability is similar to The unrestricted motion vector mode is redefined under H.263+. It uses Reversible Variable Length Coding (RVLC) to encode the difference motion vectors. The RVLC encoder is able to minimize the impact of transmission error by

allowing the decoder to decode from both forward and reverse directions. The range of motion vectors is extended again to [−256, 256].

• A slice structure is used to replace GOB for additional flexibility. A slice can contain a variable number of macroblocks. The transmission order can be either sequential or arbitrary, and the shape of a slice is not required to be rectangular.

• H.263+ implements Temporal, SNR, and Spatial scalabilities. Scalability refers to the ability to handle various constraints, such as display resolution, bandwidth, and hardware capabilities. The enhancement layer for Temporal scalability increases perceptual quality by inserting B-frames between two P-frames. SNR scalability is achieved by using various quantizers of smaller-and-smaller step_size to encode additional enhancement layers into the bitstream. Thus, the decoder can decide how many enhancement layers to decode according to computational or network constraints.

Advanced Video Coding

MPEG-1 and 2 employ frame-based coding techniques, in which each rectangular video frame is divided into macro blocks and then blocks for compression. This is also known as block-based coding. Their main concern is the high compression ratio and satisfactory quality of video under such compression techniques. MPEG-4 has a very different emphasis [8]. Besides compression, it pays great attention to user interactivity. This allows a larger number of users to create and communicate their multimedia presentations and applications on new infrastructures, such as the Internet, the World Wide Web (WWW), and mobile/wireless networks. MPEG-4 departs from its predecessors in adopting a new object-based coding approach— media objects are now entities for MPEG-4 coding. Media objects (also known as audio and visual objects) can be either natural or synthetic; that is to say, they may be captured by a video camera or created by computer programs.

Object-based coding not only has the potential of offering higher compression ratio but is also beneficial for digital video composition, manipulation, indexing, and retrieval. Figure illustrates how MPEG-4 videos can be composed and manipulated by simple operations such as insertion/deletion, translation/rotation, scaling, and so on, on the visual objects.

MPEG-4 (version 1) was finalized in October 1998 and became an international standard in early 1999, referred to as ISO/IEC 14496. An improved version (version 2) was finalized in December 1999 and acquired International Standard status in 2000. Similar to the previous MPEG standards, its first five parts are Systems, Video,

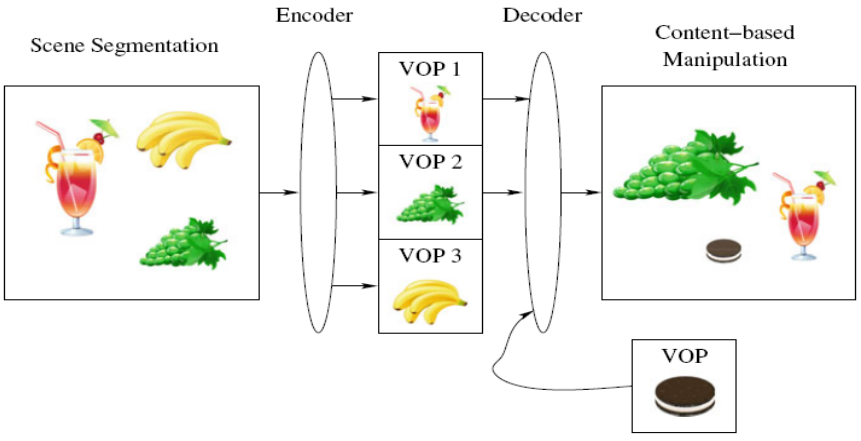


Fig. Composition and manipulation of MPEG-4 videos (VOP = Video Object Plane)

Audio, Conformance, and Software. This chapter will discuss the video compression issues in MPEG-4 Part 2 (formally ISO/IEC 14496-2).
Syntactically, all five levels have a unique start code in the bitstream, to enable random access.

1. Video-object Sequence (VS). VS delivers the complete MPEG-4 visual scene, which may contain 2D or 3D natural or synthetic objects.
2. Video Object (VO). VO is a particular object in the scene, which can be of arbitrary (nonrectangular) shape, corresponding to an object or background of the scene.
3. Video Object Layer (VOL). VOL facilitates a way to support (multilayered) scalable coding. A VO can have multiple VOLs under scalable coding or a single VOL under nonscalable coding. As a special case, MPEG-4 also supports a special.

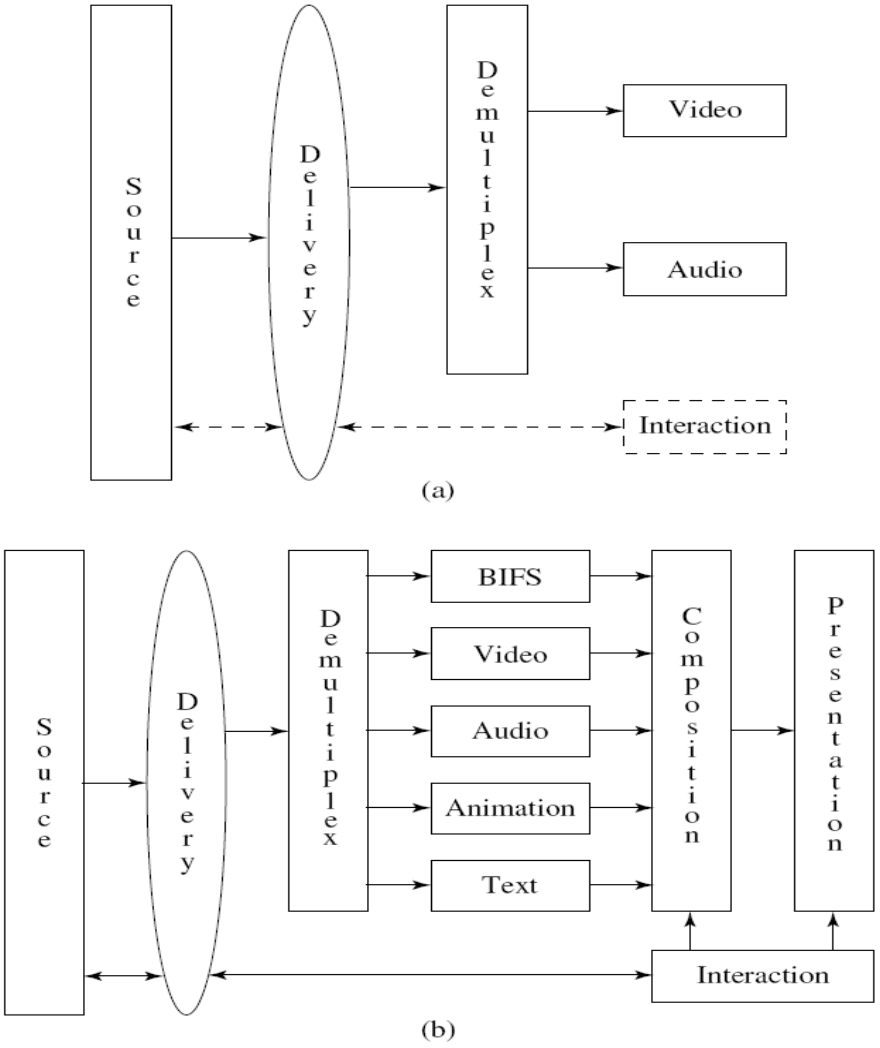
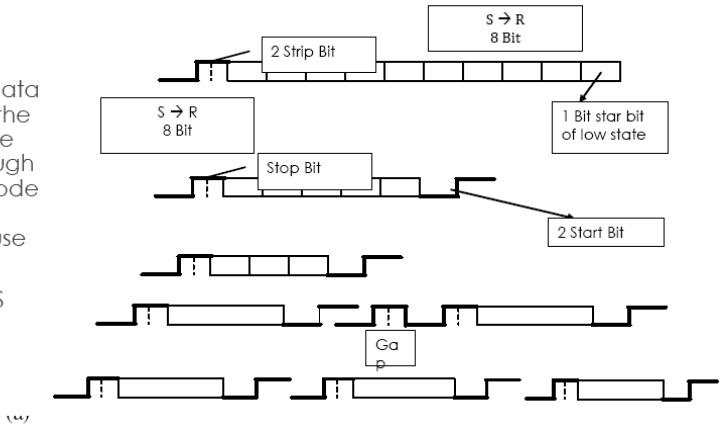


Fig. Comparison of interactivity in MPEG standards: **a** Reference models in MPEG-1 and 2 (interaction in dashed lines supported only by MPEG-2); **b** MPEG-4 reference model

ATM Networks-Compressions issues in ATM networks, Compression Algorithms for packet videos.

- **ATM (asynchronous transfer mode)** is a type of data transmission system where data transmission rates vary and the data is not transmitted at the same rate of interval. Although data rates may vary, the mode of sending data is very accurate and does not cause much data loss.
- Common examples are SMS and Emails.

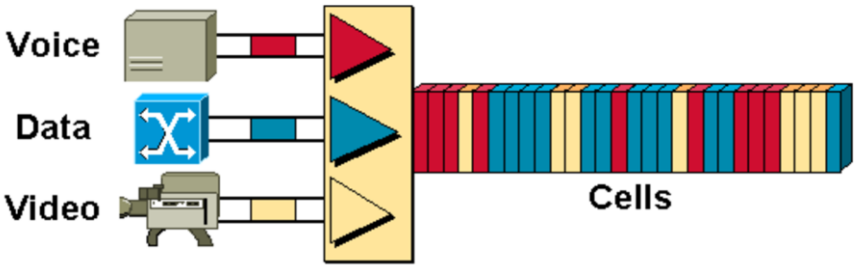


Multimedia over ATM

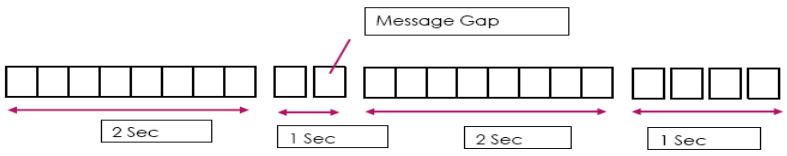
Multimedia is an emerging service which integrates voice, video and data in the same service.

- With the progress made in high speed large capacity multimedia servers, high speed networks, cost effective QoS, acceptable service category and cost effective set top boxes, it is currently possible to carry multimedia over high speed networks cost effectively and efficiently.
- This paper surveys the progress made and the future of efficiently carrying multimedia over ATM networks.

Multimedia over ATM Method/Process



- STM (Synchronous Transfer Mode) is a common method of transferring multimedia data over the internet in a constant data speed rate.
- This may have a constant speed but may result in loss of data during transfer.
- E.g. Video Streaming, Video Conferencing and VoIP



Network Issues

- Cable Problem
- Connectivity Problem
- Excessive Network Collisions
- Software Problem
- Duplicate IP Address Over Buffering
- Slow Server Issues
- Video and Audio Latency
- Frame Drops
- Freezing Issues

Error Resilience

Error Resilience decides how the decoder behaves if some value in the input is against the specification: whether it assumes it is a minor encoder mistake and "extends" the standard to give a meaning to that input or whether it assumes the input is corrupted and it should apply error concealment, i.e. assuming this and surrounding data have no relation with what the video really should look like and thus will try to reconstruct something reasonable-looking e.g. from the previous frame.

This is not a speed related option (while error concealment is very slow that is not relevant, if it is used when it shouldn't be or the other way round the result will look very horrible).

Networking Characteristics

- Availability. Availability is typically measured in a percentage based on the number of minutes that exist in a year. Therefore, uptime would be the number of minutes the network is available divided by the number of minutes in a year.
- Cost includes the cost of the network components, their installation, and their ongoing maintenance.
- Reliability defines the reliability of the network components and the connectivity between them. Mean time between failures (MTBF) is commonly used to measure reliability.
- Security includes the protection of the network components and the data they contain and/or the data transmitted between them.
- Speed includes how fast data is transmitted between network end points (the data rate).
- Scalability defines how well the network can adapt to new growth, including new users, applications, and network components.
- Topology describes the physical cabling layout and the logical way data moves between components.

TEXT / REFERENCE BOOKS

1. Rafael C. Gonzalez, Richard E. Woods, ŏ" Digital Image Processing", Pearson , Second Edition, 2004.
2. David Saloman,"Data compression" , Springer International, 4th Edition.
3. Khalid Sayood, "Introduction To Data Compressio" ,Elsevier 3rd Edition.
4. Ralfsteinmetz and Klara Nahrstedt, "Multimedia Computing, Communications & Applications"

Pearson Edn

5. Rajan Parekh, "Principles of Multimedia, Tata Mc Graw Hill.
6. Anil K. Jain, "Fundamentals of Digital Image Processin", Pearson 2002.
7. J F Koegel Buford- -Multimedia systems Addison Wesley
8. T Vaughan-,"Multimedia: Making it work" Tata Mc Graw Hill.

PART A

1. What are the types of video compressions available?
2. Mention some of the image formats used in multimedia?
3. Define MIDI.
4. What is AVI format?
5. What is meant by GIF?
6. What is compression?
7. Mention the categories or JPEG formats?
8. What is wave pattern of the sound?
9. What is the most common file formats used in multimedia?
10. Compare lossless and lossy video compressions.

PART B

1. Discuss about the human auditory system with relevant diagram.
2. Illustrate about the MPEG4 Audio lossless coding.
3. Elaborate the compression issues in ATM networks.
4. Explain the Advanced Video Coding concept.
5. Discuss about the Compression Algorithms for packet videos.