

29. (c) Explain Apple's Quick Time. (8)

30. Explain Multimedia Storage and Retrieval, Massive Data Volumes, Storage Technologies and Multimedia Object Storage.

31. Explain about the involvement of DBMS and RDBMS Extensions for multimedia systems.

UNIT IV -MULTIMEDIA FILE HANDLING

Compression and decompression – Data and file format standards – Multimedia I/O technologies – Digital voice and audio – Video image and animation – Full motion video – Storage and retrieval technologies.

4.1 COMPRESSION AND DECOMPRESSION

Compression is the way of making files to take up less space. In multimedia systems, in order to manage large multimedia data objects efficiently, these data objects need to be compressed to reduce the file size for storage of these objects.

Compression tries to eliminate redundancies in the pattern of data.

For example, if a black pixel is followed by 20 white pixels, there is no need to store all 20 white pixels. A coding mechanism can be used so that only the count of the white pixels is stored. Once such redundancies are removed, the data object requires less time for transmission over a network. This in turn significantly reduces storage and transmission costs.

TYPES OF COMPRESSION

Compression and decompression techniques are utilized for a number of applications, such as facsimile system, printer systems, document storage and retrieval systems, video teleconferencing systems, and electronic multimedia messaging systems. An important standardization of compression algorithm was achieved by the CCITT when it specified Group 2 compression for facsimile system. .

When information is compressed, the redundancies are removed.

Sometimes removing redundancies is not sufficient to reduce the size of the data object to manageable levels. In such cases, some real information is also removed. The primary criterion is that removal of the real information should not perceptibly affect the quality of the result. In the case of video, compression causes some information to be lost; some information at a delete level is considered not essential for a reasonable reproduction of the scene. This type of compression is called lossy compression. Audio compression, on the other hand, is not lossy. It is called lossless compression.

Lossless Compression.

In lossless compression, data is not altered or lost in the process of compression or decompression. Decompression generates an exact replica of the original object. Text compression is a good example of lossless compression. The repetitive nature of text, sound and graphic images allows replacement of repeated strings of characters or bits by codes. Lossless compression techniques are good for text data and for repetitive data in images all like binary images and gray-scale images.

Some of the commonly accepted lossless standards are given below:

- ❖ **Packbits encoding (Run-length encoding)**
- ❖ **CCITT Group 3 1D**
- ❖ **CCITT Group 3 2D**
- ❖ **CCITT Group 4**
- ❖ **Lempel-Ziv and Welch algorithm LZW.**

Lossy compression is that some loss would occur while compressing information objects.

Lossy compression is used for compressing audio, gray-scale or color images, and video objects in which absolute data accuracy is not necessary.

The idea behind the lossy compression is that, the human eye fills in the missing information in the case of video.

But, an important consideration is how much information can be lost so that the result should not affect. For example, in a grayscale image, if several bits are missing, the information is still perceived in an acceptable manner as the eye fills in the gaps in the shading gradient.

Lossy compression is applicable in medical screening systems, video tele-conferencing, and multimedia electronic messaging systems.

Lossy compressions techniques can be used alone or in combination with other compression methods in a multimedia object consisting of audio, color images, and video as well as other specialized data types.

The following lists some of the lossy compression mechanisms:

- ✓ Joint Photographic Experts Group (JPEG)
- ✓ Moving Picture Experts Group (MPEG)
- ✓ Intel DVI
- ✓ CCITT H.261 (P * 24) Video Coding Algorithm
- ✓ Fractals.

Compression schemes are

Binary Image compression schemes

Binary Image Compression Scheme is a scheme by which a binary image containing black and white pixel is generated when a document is scanned in a binary mode.

The schemes are used primarily for documents that do not contain any continuous-tone information or where the continuous-tone information can be captured in a black and white mode to serve the desired purpose.

The schemes are applicable in office/business documents, handwritten text, line graphics, engineering drawings, and so on. Let us view the scanning process. A scanner scans a document as sequential scan lines, starting from the top of the page.

A scan line is complete line of pixels, of height equal to one pixel, running across the page. It scans the first line of pixels (Scan Line), then scans second "line, and works its way up to the last scan line of the page. Each scan line is scanned from left to right of the page generating black and white pixels for that scan line.

This uncompressed image consists of a single bit per pixel containing black and white pixels. Binary 1 represents a black pixel, binary 0 a white pixel. Several schemes have been standardized and used to achieve various levels of compressions. Let us review the more commonly used schemes.

1 .Packpits Encoding(Run-Length Encoding)

It is a scheme in which a consecutive repeated string of characters is replaced by two bytes. It is the simple, earliest of the data compression scheme developed. It need not to have a standard. It is used to compress black and white (binary) images. Among two bytes which are being replaced, the first byte contains a number representing the number of times the character is repeated, and the second byte contains the character itself.

In some cases, one byte is used to represent the pixel value, and the other seven bits to represents the run length.

2. CCITT Group 3 1-D Compression

This scheme is based on run-length encoding and assumes that a typical scanline has long runs of the same color.

This scheme was designed for black and white images only, not for gray scale or color images. The primary application of this scheme is in facsimile and early document imaging system.

Huffman Encoding

A modified version of run-length encoding is Huffman encoding.

It is used for many software based document imaging systems. It is used for encoding the pixel run length in CCITT Group 3 1-dGroup 4.

It is variable-length encoding. It generates the shortest code for frequently occurring run lengths and longer code for less frequently occurring run lengths.

Mathematical Algorithm for huffman encoding:

Huffman encoding scheme is based on a coding tree.

It is constructed based on the probability of occurrence of white pixels or black pixels in the run length or bit stream.

Table below shows the CCITT Group 3 tables showing codes or white run lengths and black run lengths.

White Run	Code	Black Run	Code
--------------	------	--------------	------

Length	Word	Length	Word
0	00110101	0	0000110111
1	000111	1	010
2	0111	2	11
3	1000	3	10
4	1011	4	011
5	1100	5	0011
6	1110	6	0010
7	1111	7	00011
8	10011	8	000101
9	10100	9	000100
10	00111	10	0000100

1

11	01000	10	0000100
11	01000	11	0000101
12	001000	12	0000111
13	000011	13	00000100
14	110100	14	00000111
15	110101	15	000011000
16	101010	16	0000010111
17	101011	17	0000011000
18	0100111	18	0000001000
19	0001100	19	0000 11 00 III
20	0001000	20	00001101000
21	0010111	21	00001101100
22	0000011	22	00000110111
23	0000100	23	00000101000
24	0101000	24	00000010111
25	0101011	25	00000011000
26	0010011	26	000011001010
27	0100100	27	000011001011
28	0011000	28	000011 00 11 00
29	00000010	29	000011001101
30	00000011	30	000001101000
31	00011010	31	000001101001
32	00011011	32	000001101010
33	00010010	33	000001101011
34	00010011	34	000011010010
35	00010100	35	000011 0 10011

For example, from **Table 2**, the run-length code of 16 white pixels is 101010, and of 16 black pixels 0000010111. Statistically, the occurrence of 16 white pixels is more frequent than the occurrence of 16 black pixels. Hence, the code generated for 16 white pixels is much shorter. This allows for quicker decoding. For this example, the tree structure could be constructed.

36	00010101	36	000011010100
37	00010110	37	000011010101
38	000101 II	38	000011010110

39	00101000	39	000011 0 1 0 1 1 1
40	00101001	40	000001101100
41	00101010	41	000001101101
42	00101011	42	000011011010
43	00101100	43	0000 11 0 1 1011
44	00101101	44	000001010100
45	00000100	45	000001010101
46	00000101	46	000001010110
47	00001010	47	000001010111
48	00001011	48	000001100100
49	01010010	49	000001100101
50	01010011	50	000001010010
51	01010100	51	000001010011
52	01010101	52	000000100100
53	00100100	53	000000110111

The codes greater than a string of 1792 pixels are identical for black and white pixels. A new code indicates reversal of color, that is, the pixel Color code is relative to the color of the previous pixel sequence.

Table 3 shows the codes for pixel sequences larger than 1792 pixels.

Run Length (Black and White)	Make-up Code
1792	00000001000
1856	00000001100
1920	00000001101
1984	000000010010
2048	000000010011
2112	000000010100
2176	000000010101
2240	000000010110
2304	000000010111
2368	000000011100
2432	000000011101
2496	000000011110
2560	000000011111

CCITT Group 3 compression utilizes Huffman coding to generate a set of make-up codes and a set of terminating codes for a given bit stream. Make-up codes are used to represent run length in multiples of 64 pixels. Terminating codes are used to represent run lengths of less than 64 pixels.

As shown in **Table 2**; run-length codes for black pixels are different from the run-length codes for white pixels. For example, the run-length code for 64 white pixels is 11011. The run length code for 64 black pixels is 0000001111. Consequently, the run length of 132 white pixels is encoded by the following two codes:

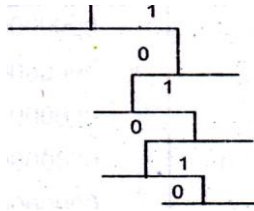
Makeup code for 128 white pixels - 10010

Terminating code for 4 white pixels - 1011

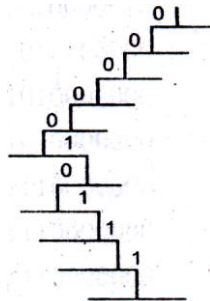
The compressed bit stream for 132 white pixels is 100101011, a total of nine bits. Therefore the compression ratio is 14, the ratio between the total number of bits (132) divided by the number of bits used to code them (9).

CCITT Group 3 uses a very simple data format. This consists of sequential blocks of data for each scanline, as shown in **Table 4**.

Coding tree for 16 white pixels



Coding tree for 16 black pixels



Note that the file is terminated by a number of EOLs (End of Line) if there is no change in the line [rom the previous line (for example, white space).

TABLE 4: CCITT Group 3 1D File Format

EOL	DATA	FILL	EOL	DATA	FILL	EOL...	DATA	FILL	EOL	EOL	EOL
LINE				LINE			LINE				
1				2			n				

Advantages of CCITT Group 3 ID

CCITT Group 3 compression has been used extensively due to the following two advantages:

It is simple to implement in both hardware and software .

It is a worldwide standard for facsimile which is accepted for document imaging application. This allows document imaging applications to incorporate fax documents easily.

- CCITT group 3 compressions utilizes Huffman coding to generate a set of make-up codes and a set of terminating codes for a give bit stream.
- CCITT Group 3 uses a very simply data format. This consists of sequential blocks of data for each scanline.

CCITT Group 3 2D Compression

It is also known as modified run length encoding. It is used for software based imaging system and facsimile.

It is easier to decompress in software than CCITT Group 4. The CCITT Group 3 2D scheme uses a "k" factor where the

image is divided into several group of k lines. This scheme is based on the statistical nature of images; the image data across the adjacent scanline is redundant.

If black and white transition occurs on a given scanline, chances are the same transition will occur within + or - 3 pixels in the next scanline.

Necessity of k factor

When CCITT Group 3 2D compression is used, the algorithm embeds Group 3 1 D coding between every k groups of Group 3 2D coding, allowing the Group 3 1 D coding to be the synchronizing line in the event of a transmission error.

Therefore when a transmission error occurs due to a bad communication link, the group 3 I D can be used to synchronize and correct the error.

Data formatting for CCIIT Group 3 2D

The 2D scheme uses a combination of additional codes called vertical code, pass code, and horizontal code to encode every line in the group of k lines.

The steps for pseudocode to code the code line are:

(i) Parse the coding line and look for the change in the pixel value. (Change is found at al location). (ii) Parse the reference line and look for the change in the pixel value. (Change is found at bl location). (iii) . Find the difference in location between bland a 1: delta = b1- al

Advantage of CCIIT Group 3 2D

- The implementation of the k factor allows error-free transmission . Compression ratio achieved is better than CCITT Group 3 1 D .
- It is accepted for document imaging applications.

DisadvantageIt doesn't provide dense compression

CCITT Group 4 2D compression

CCITT Group 4 compression is the two dimensional coding scheme without the k-factor.

In this method, the first reference line is an imaginary all-white line above the top of the image. The first group of pixels (scanline) is encoded utilizing the imaginary white line as the reference line.

The new coded line becomes the references line for the next scan line. The k-factor in this case is the entire page of line. In this method, there are no end-of-line (EOL) markers before the start of the compressed data

COLOR, GRAY SCALE AND STILL-VIDEO IMAGE COMPRESSION

Color:

Color is a part of life we take for granted. Color adds another dimension to objects. It helps in making things standout.

Color adds depth to images, enhance images, and helps set objects apart from -background.

Let us review the physics of color. Visible light is a form of electromagnetic radiation or radiant energy, as are radio frequencies or x-rays. The radiant energy spectrum contains audio frequencies, radio frequencies, infrared, visible light, ultraviolet rays, x-rays and gamma rays.

Radian energy is measured in terms of frequency or wavelength. The relationship between the two is

$$\lambda = \frac{c}{f} \text{ meters}$$

where λ - is the wavelength in meters

c is the velocity of light in meters per second.

f is frequency of the radiation in hertz.

Since all electromagnetic waves travel through space at the velocity of light i.e., 3×10^8 meters/second - the wavelength is calculated by

$$\lambda = \frac{3}{f} \times 10^8 \text{ meters}$$

Color Characteristics

We typically define color by its brightness, the hue and depth of color.

Luminance or Brightness

This is the measure of the brightness of the light emitted or reflected by an object; it depends on the radiant, energy of the color band.

Hue This is the color sensation produced in an observer due to the presence of certain wavelengths of color. Each wavelength represents a different hue.

Saturation This is a measure of color intensity, for example, the difference between red and pink.

Color Models Several calm' models have been developed to represent color mathematically.

Chromacity Model It is a three-dimensional model with two dimensions, x and y, defining the color, and the third dimension defining the luminance. It is an additive model since x and yare added to generate different colors.

RGBModel RGB means Red Green Blue. This model implements additive theory in that different intensities of red, green and blue are added to generate various colors.

HSI Model The Hue Saturation and Intensity (HSI) model represents an artist's impression of tint, shade and tone. This model has proved suitable for image processing for filtering and smoothing images.

CMYK Model The Cyan, Magenta, Yellow and Black color model is used in desktop publishing printing devices. It is a color-subtractive model and is best used in color printing devices only.

YUV Representation The NTSC developed the YUV three-dimensional color model.

y -Luminance Component

UV -Chrominance Components.

Luminance component contains the black and white or grayscale information. The chrominance component contains color information where U is red minus cyan and V is magenta minus green.

YUV Model for JPEG

The *JPEG* compression scheme uses several stages.

The first stage converts the signal from the spatial RGB domain to the YUV frequency domain by performing discrete cosine transform. This process allows separating luminance or gray-scale components from the chrominance components of the image.

JOINT PHOTOGRAPHIC EXPERTS GROUP COMPRESSION (JPEG)

ISO and CCITT working committee joint together and formed Joint Photographic Experts Group. It is focused exclusively on still image compression.

Another joint committee, known as the Motion Picture Experts Group (MPEG), is concerned with full motion video standards.

JPEG is a compression standard for still color images and grayscale images, otherwise known as continuous tone images.

JPEG has been released as an ISO standard in two parts

- Part I specifies the modes of operation, the interchange formats, and the encoder/decoder specifies for these modes along with substantial implementation guide lines .
- Part 2 describes compliance tests which determine whether the implementation of an encoder or decoder conforms to the standard specification of part I to ensure interoperability of systems compliant with JPEG standards

Requirements addressed by JPEG

- The design should address image quality .
- The compression standard should be applicable to practically any kind of continuous-tone digital source image .
- It should be scalable from completely lossless to lossy ranges to adapt it.
- It should provide sequential encoding .
- It should provide for progressive encoding .
- It should also provide for hierarchical encoding .
- The compression standard should provide the option of lossless encoding so that images can be guaranteed to provide full detail at the selected resolution when decompressed.

Definitions in the JPEG Standard

The JPEG Standards have three levels of definition as follows:

- * Base line system
- * Extended system
- * Special lossless function.

The base line system must reasonably decompress color images, maintain a high compression ratio, and handle from 4 bits/pixel to 16 bits/pixel.

The extended system covers the various encoding aspects such as variable-length encoding, progressive encoding, and the hierarchical mode of encoding.

The special lossless function is also known as predictive lossless coding. It ensures that at the resolution at which the image is no loss of any detail that was there in the original source image.

Overview of JPEG Components JPEG Standard components are:

- (i) Baseline Sequential Codec
- (ii) OCT Progressive Mode
- (Hi) Predictive Lossless Encoding
- (iv) Hierarchical Mode.

These four components describe four different levels of *jpeg* compression.

The baseline sequential code defines a rich compression scheme the other three modes describe enhancements to this baseline scheme for achieving different results.

Some of the terms used in JPEG methodologies are:

Discrete Cosine Transform (OCT)

OCT is closely related to Fourier transforms. Fourier transforms are used to represent a two dimensional sound signal.

DCT uses a similar concept to reduce the gray-scale level or color signal amplitudes to equations that require very few points to locate the amplitude in Y-axis X-axis is for locating frequency.

DCT Coefficients

The output amplitudes of the set of 64 orthogonal basis signals are called OCT Co-efficients.

Quantization This is a process that attempts to determine what information can be safely discarded without a significant loss in visual fidelity. It uses OCT co-efficient and provides many-to-one mapping. The quantization process is fundamentally lossy due to its many-to-one mapping.

De Quantization This process is the reverse of quantization. Note that since quantization used a many-to-one mapping, the information lost in that mapping cannot be fully recovered

Entropy Encoder / Decoder Entropy is defined as a measure of randomness, disorder, or chaos, as well as a measure of a system's ability to undergo spontaneous change. The entropy encoder compresses quantized DCT co-efficients more compactly based on their spatial characteristics. The baseline sequential codec uses Huffman coding. Arithmetic coding is another type of entropy encoding

Huffman Coding Huffman coding requires that one or more sets of huff man code tables be specified by the application for encoding as well as decoding. The Huffman tables may be pre-defined and used within an application as defaults, or computed specifically for a given image.

Baseline Sequential codec

It consists of three steps: Formation of DCT co-efficients quantization, and entropy encoding. It is a rich compression scheme.

DCT Progressive Mode

The key steps of formation of DCT co-efficients and quantization are the same as for the baseline sequential codec. The key difference is that each image component is coded in multiple scans instead of a single scan.

Predictive Lossless Encoding

It is to define a means of approaching lossless continuous-tone compression. A predictor combines sample areas and predicts neighboring areas on the basis of the sample areas. The predicted areas are checked against the fully loss less sample for each area.

The difference is encoded losslessly using huffman or arithmetic entropy encoding .

Hierarchical Mode

The hierarchical mode provides a means of carrying multiple resolutions. Each successive encoding of the image is reduced by a factor of two, in either the horizontal or vertical dimension.

JPEG Methodology

The JPEG compression scheme is lossy, and utilizes forward discrete cosine transform (or forward DCT mathematical function), a uniform quantizer, and entropy encoding. The DCT function removes data redundancy by transforming data from a spatial domain to a frequency domain; the quantizer quantizes DCT co-efficients with weighting functions to generate quantized DCT co-efficients optimized for the human eye; and the entropy encoder minimizes the entropy of quantized DCT co-efficients.

The JPEG method is a symmetric algorithm. Here, decompression is the exact reverse process of compression.

Figure below describes a typical DCT based encoder and decoder.

Symmetric Operation of DCT based Codec

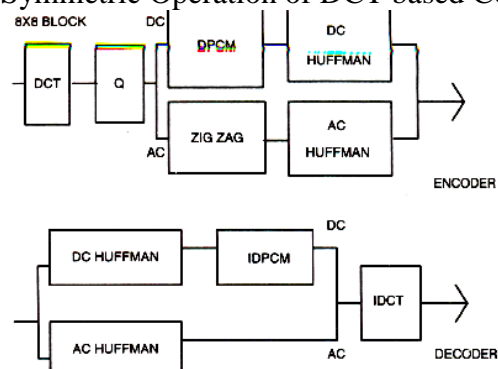


Figure below shows the components and sequence of quantization

5 * 8

Image blocks

Quantization

Quantization is a process of reducing the precision of an integer, thereby reducing the number of bits required to store the integer, thereby reducing the number of bits required to store the integer.

The baseline JPEG algorithm supports four color quantization tables and two huffman tables for both DC and AC DCT co-efficients. The quantized co-efficient is described by the following equation:

$$\text{Quantized Co-efficient (i, j)} = \frac{DCT(i, j)}{Quantum(i, j)}$$

ZigZag Sequence

Run-length encoding generates a code to represent the Count of zero-value OCT co-efficients. This process of run-length encoding gives an excellent compression of the block consisting mostly of zero values.

Further empirical work proved that the length of zero values in a run can be increased to give a further increase in compression by reordering the runs. JPEG came up with ordering the quantized OCT co-efficients in a ZigZag sequence

Entropy Encoding

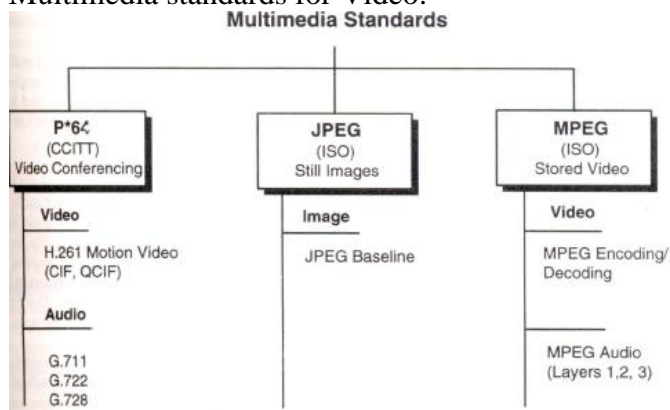
Entropy is a term used in thermodynamics for the study of heat and work. Entropy, as used in data compression, is the measure of the information content of a message in number of bits. It is represented as

Entropy in number of bits = \log_2 (probability of Object)

VIDEO IMAGE COMPRESSION

The development of digital video technology has made it possible to use digital video compression for a variety of telecommunications applications. Standardization of compression algorithms for video was first initiated by CCITT for teleconferencing and video telephony

Multimedia standards for Video:



Requirements for full-motion Video Compression

Applications using MPEG standards can be symmetric or asymmetric. Symmetric applications are applications that require essentially equal use of compression and decompression. Asymmetric applications require frequent decompression.

Symmetric applications require on-line input devices such as video cameras, scanners and microphones for digitized sound. In addition to video and audio compression, this standards activity is concerned with a number of other Issues concerned with playback of video clips and sound clips. The MPEG standard has identified a number of such issues that have been addressed by the standards activity. Let us review these Issues.

Random Access

The expectations generated for multimedia systems are the ability to play a sound or video clip from any frame with that clip, irrespective of on what kind-of media the information is stored

VCR paradigm

The VCR paradigm consists of the control functions typically found on a VCR such as play, fast forward, rewind, search forward and rewind search.

Multiplexing Multiple Compressed Audio and Video Bit Streams

It is a special requirement retrieved from different storage centers on a network. It may be received from different storage centers on a network. It may have to be achieved in a smooth manner to avoid the appearance of a jumpy screen.

Editability

Playback Device Flexibility

CCITT H.261 Video Coding Algorithms (P x 64)

The linear quantizer uses a step algorithm that can be adjusted based on picture quality and coding efficiency. The H.261 is a standard that uses a hybrid of OCT and OPCM (differential pulse Code Modulation) schemes with motion estimation.

It also defines the data format. Each MB contains the OCT coefficients (TCOEFF) of a block followed by an EOB (a fixed length end-of-block marker). Each MB consists of block data and an MB header. A GOB (Group of Blocks) consists of a GOB header. The picture layer consists of a picture header. The H.261 is designed for dynamic use and provides a fully contained organization and a high level of interactive control.

Moving Picture Experts Group Compression

The MPEG standards consist of a number of different standards.

The MPEG 2 suite of standards consist of standards for MPEG2 Video, MPEG - 2 Audio and MPEG - 2 systems. It is also defined at different levels, called profiles.

The main profile is designed to cover the largest number of applications. It supports digital video compression in the range of 2 to 15 M bits/sec. It also provides a generic solution for television worldwide, including cable, direct broadcast satellite, fibre optic media, and optical storage media (including digital VCRs).

MPEG Coding Methodology

The above said requirements can be achieved only by incremental coding of successive frames. It is known as interframe coding. If we access information randomly by frame requires coding confined to a specific frame, then it is known as intraframe coding.

The MPEG standard addresses these two requirements by providing a balance between interframe coding and intraframe coding. The MPEG standard also provides for recursive and non-recursive temporal redundancy reduction.

The MPEG video compression standard provides two basic schemes: discrete-transform-based compression for the reduction of spatial redundancy and block-based motion compensation for the reduction of temporal (motion) redundancy. During the initial stages of DCT compression, both the full motion MPEG and still image JPEG algorithms are essentially identical. First an image is converted to the YUV color space (a luminance/chrominance color space similar to that used for color television). The pixel data is then fed into a discrete cosine transform, which creates a scalar quantization (a two-dimensional array representing various frequency ranges represented in the image) of the pixel data.

Following quantization, a number of compression algorithms are applied, including run-length and Huffman encoding. For full motion video (MPEG I and 2), several more levels of block based motion-compensated techniques are applied to reduce temporal redundancy with both causal and noncausal coding to further reduce spatial redundancy.

The MPEG algorithm for spatial reduction is lossy and is defined as a hybrid which employs motion compensation, forward discrete cosine transform (DCF), a uniform quantizer, and Huffman coding. Block-based motion compensation is *utilized* for *reducing temporal* redundancy (i.e. to reduce the amount of data needed to represent each picture in a video sequence). Motion-compensated reduction is a key feature of MPEG.

Moving Picture Types

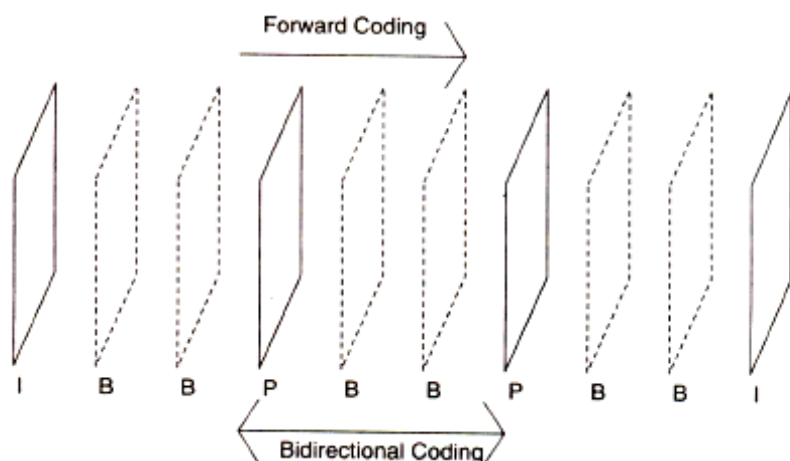
Moving pictures consist of sequences of video pictures or frames that are played back a fixed number of frames per second. To achieve the requirement of random access, a set of pictures can be defined to form a group of pictures (GOP) consisting of one or more of the following three types of pictures.

Intra pictures (I)

Unidirectionally predicted pictures (U) Bidirectionally predicted pictures (B)

A Gap consists of consecutive pictures that begin with an intrapicture. The intrapicture is coded without any reference to any other picture in the group.

Predicted pictures are coded with a reference to a past picture, either an intrapicture or a unidirectionally predicted picture. Bidirectionally predicted picture is never used as references. Motion Compensation for Coding MPEG



Let us review the concept of Macroblocks and understand the role they play in compression

MACRO BLOCKS

For the video coding algorithm recommended by CCITT, CIF and QCIF are divided into a hierarchical block structure consisting of pictures, groups of blocks (GOBs), Macro Blocks (MBs), and blocks. Each picture frame is divided into 16×16 blocks. Each Macroblock is composed of four 8×8 (Y) luminance blocks and two 8×8 (C_b and C_n) chrominance blocks. This set of six blocks, called a macroblock; is the basic hierarchical component used for achieved a high level of compression.

Motion compensation

Motion compensation is the basis for most compression algorithms for visual telephony and full-motion video. Motion compensation assumes that the current picture is some translation of a previous picture.

This creates the opportunity for using prediction and interpolation. Prediction requires only the current frame and the reference frame.

Based on motion vectors values generated, the prediction approach attempts to find the relative new position of the object and confirms it by comparing some block exhaustively. In the interpolation approach, the motion vectors are generated in relation to two reference frames, one from the past and the next predicted frame.

The best-matching blocks in both reference frames are searched, and the average is taken as the position of the block in the current frame. The motion vectors for the two reference frames are averaged.

Picture Coding Method

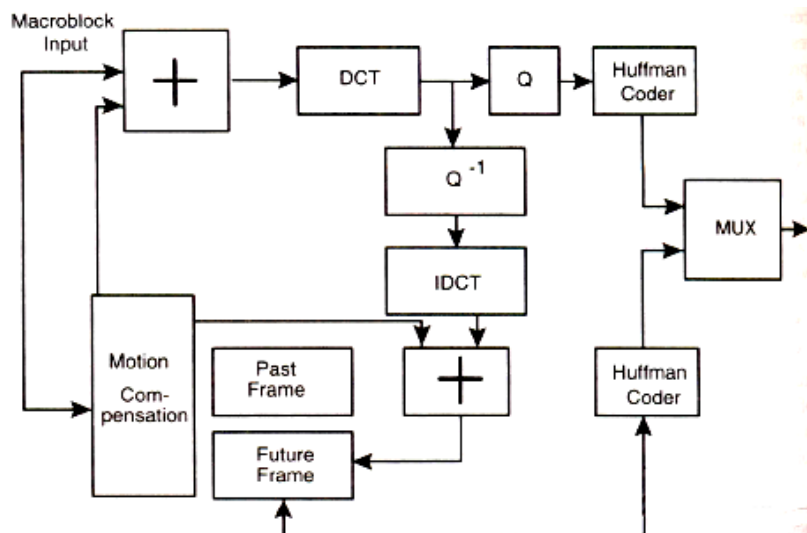
In this coding method, motion compensation is applied bidirectionally. In MPEG terminology, the motion-compensated units are called macro blocks (MBs).

MBs are 16×16 blocks that contain a number of 8×8 luminance and chrominance blocks. Each 16×16 macro block can be of type intrapicture, forward-predicted, backward predicted, or average.

MPEG Encoder

Figure below shows the architecture of an MPEG encoder. It contains DCT quantizer, Huffman coder and Motion compensation. These represent the key modules in the encoder.

Architecture of MPEG Encoder:



The Sequence of events for MPEG

First an image is converted to the YUV color space.

The pixel data is then fed into a DCT, which creates a scalar quantization of the pixel data.

Following quantization, a number of compression algorithms are applied, including run-length and Huffman encoding. For full-motion video, several more levels of motion compensation compression and coding are applied.

MPEG -2

It is defined to include current television broadcasting compression and decompression needs, and attempts to include hooks for HDTV broadcasting.

The MPEG-2 Standard Supports:

- 1.Video Coding: * MPEG-2 profiles and levels.
- 2.Audio Coding: *MPEG-1 audio standard for backward compatibility.
 - * Layer-2 audio definitions for MPEG-2 and stereo sound.
 - * Multichannel sound.
3. Multiplexing: MPEG-2 definitions

MPEG-2, "The Grand Alliance"

It consists of following companies AT&T, MIT, Philips, Sarnoff Labs, GI Thomson, and Zenith.

The MPEG-2 committee and FCC formed this alliance. These companies together have defined the advanced digital television system that include the US and European HDTV systems. The outline of the advanced digital television system is as follows:

- 1.Format: 1080/2: 1160 or 720/1.1160
 - 2.Video coding: MPEG-2 main profile and high level
 - 3.Audio coding: Dolby AC3
 - 4.Multiplexor: As defined in MPEG-2
- Modulation: 8- VSB for terrestrial and 64-QAM for cable.

Vector Quantization

Vector quantization provides a multidimensional representation of information stored in look-up tables, vector quantization is an efficient pattern-matching algorithm in which an image is decomposed into two or more vectors, each representing particular features of the image that are matched to a code book of vectors.

These are coded to indicate the best fit.

In image compression, source samples such as pixels are blocked into vectors so that each vector describes a small segment or sub block of the original image.

The image is then encoded by quantizing each vector separately

Intel's Indeo Technology

It is developed by Intel Architecture Labs Indeo Video is a software technology that reduces the size of uncompressed digital video files from five to ten times.

Indeo technology uses multiple types of 'lossy' and 'lossless' compression techniques.

AUDIO COMPRESSION

Audio consists of analog signals of varying frequencies. The audio signals are converted to digital form and then processed, stored and transmitted. Schemes such as linear predictive coding and adaptive differential pulse code modulation (ADPCM) are utilized for compression to achieve 40-80% compression.

FRACTAL COMPRESSION

A fractal is a multi-dimensional object with an irregular shape or body that has approximately the same shape or body irrespective of size. For example, if you consider 'stick' as your object, the fractal is defined, mathematically as

$$D = \lim_{L \rightarrow 0} \frac{\ln N(L)}{\ln 1/L}$$

where L - approaches 0,

N(L) ~ number of stick L, and L is the length of the stick.

4.2 DATA AND FILE FORMATS STANDARDS

There are large number of formats and standards available for multimedia system.

Let us discuss about the following file formats:

- ∴ Rich-Text Format (RTF)
- ∴ Tagged Image file Format (TIFF)

- ∴ Resource Image File Format (RIFF)
- ∴ Musical Instrument Digital Interface (MIDI)
- ∴ Joint Photographic Experts Group (JPEG)
- ∴ Audio Video Interleaved (AVI) Indeo file format
- ∴ TWAIN.

Rich Text Format

This format extends the range of information from one word processor application or DTP system to another.

The key format information carried across in RTF documents are given below:

Character Set: It determines the characters that supports in a particular implementation.

Font Table: This lists all fonts used. Then, they are mapped to the fonts available in receiving application for displaying text.

Color Table: It lists the colors used in the documents. The color table then mapped for display by receiving application to the nearer set of colors available to that applications.

Document Formatting: Document margins and paragraph indents are specified here.

Section Formatting: Section breaks are specified to define separation of groups of paragraphs.

Paragraph Formatting: It specifies style sheds. It specifies control characters for specifying paragraph justification, tab positions, left, right and first indents relative to document margins, and the spacing between paragraphs.

General Formatting: It includes footnotes, annotations, bookmarks and pictures.

Character Formatting: It includes bold, italic, underline (continuous, dotted or word), strike through, shadow text, outline text, and hidden text.

Special Characters: It includes hyphens, spaces, backslashes, underscore and so on

TIFF File Format

TIFF is an industry-standard file format designed to represent raster image data generated by scanners, frame grabbers, and paint/ photo retouching applications.

TIFF Version 6.0 .

It offers the following formats:

- (i) Grayscale, palette color, RGB full-color images and black and white.
- (ii) Run-length encoding, uncompressed images and modified Huffman data compression schemes.

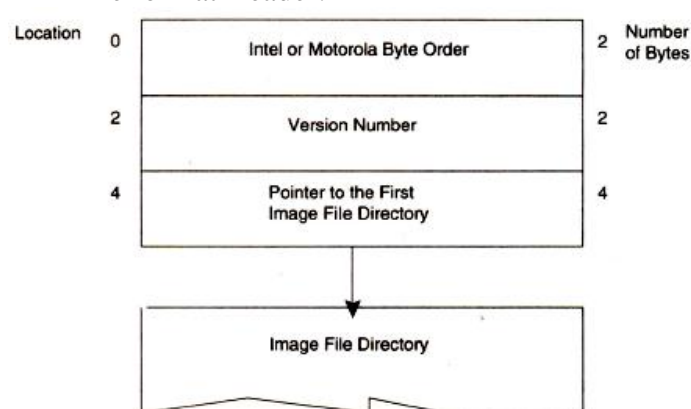
The additional formats are:

- (i) Tiled images, compression schemes, images using CMYK, YCbCr color models.

TIFF Structure

TIFF files consists of a header. The header consists of byteordering flag, TIFF file format version number, and a pointer to a table. The pointer points image file directory. This directory contains table of entries of various tags and their information.

TIFF file format Header:



TIFF Tags

The first two bytes of each directory entry contain a field called the Tag ID.

Tag IDs are grouped into several categories. They are Basic, Informational, Facsimile, Document storage and Retrieval.

TIFF Classes: (Version 5.0)

It has five classes

1. Class B for binary images
2. Class F for Fax
3. Class G for gray-scale images
4. Class P for palette color images
5. Class R for RGB full-color images.

Resource Interchange File Format (RIFF)

The **RIFF** file formats consist of blocks of data called chunks. They are

RIFF Chunk - defines the content of the RIFF file.

List Chunk - allows to embed archival location copy right information and creating date.

Subchunk - allow additional information to a primary chunk

The first chunk in a RIFF file must be a RIFF chunk and it may contain one or more sub chunk

The first four bytes of the RIFF chunk data field are allocated for the form type field containing four characters to identify the format of the data stored in the file: AVI, WAV, RMI, PAL and so

File type	Form typ	File extension
Waveform Audio File	WAVE	.WAV
Audio Video Interleaved file	AVI	.AVI
MIDI File	RMID	.RMI
Device Independent Bitmap file	RDIB	.RDI
Palette File	PAL	.PAL

The sub chunk contains a four-character ASCII string 10 to identify the type of data.

Four bytes of size contains the count of data values, and the data. The data structure of a chunk is same as all other chunks.

RIFF Chunk The first 4 characters of the RIFF chunk are reserved for the "RIFF" ASCII string. The next four bytes define the total data size.

The first four characters of the data field are reserved for form type. The rest of the data field contains two subchunk:

- (i) fmt ~ defines the recording characteristics of the waveform.
- (ii) data ~ contains the data for the waveform.

LIST Chunk

RIFF chunk may contains one or more list chunks.

List chunks allow embedding additional file information such as archival location, copyright information, creating date, description of the content of the file.

RIFF MIDI FILE FORMAT

RIFF MIDI contains a *RIFF* chunk with the form type "RMID" and a subchunk called "data" for MIDI data.

The 4 bytes are for ID of the *RIFF* chunk. 4 bytes are for size 4 bytes are for form type

4 bytes are for ID of the subchunk data and 4 bytes are for the size of MIDI data.

RIFF DIBS (Device-Independent Bit Maps) .

DIB is a Microsoft windows standard format. It defines bit maps and color attributes for bit maps independent of devices. DIBs are normally embedded in .BMP files, .WMF meta data files, and .CLP files.

DIB Structure

BITMAPINFOHEADER	RGBQUAD	PIXELS
------------------	---------	--------

BIT MAP INFOHEADER is the bit map information header.

RGBEQUAD is the color table structure.

PIXELS are the array of bytes for the pixel bit map.

The following shows the DIB file format

BITMAPINFOHEADER	BITMAPINFO=BITMAPINF O	PIXELS
------------------	---------------------------	--------

A RIFF DIB file format contains a RIFF chunk with the Form Type "RDIB" and a subchunk called "data" for DIB data.

4 bytes denote ID of the RIFF chunk

4 bytes refer size of XYZ.RDI 4 bytes define Forum Type

4 bytes describe ID of the sub chunk data 4 bytes define size of DIB data.

RIFF PALETTE File format

The RIFF Palette file format contains a RIFF chunk with the Form Type "RP AL" and a subchunk called "data" for palette data. The Microsoft Windows logical palette structure is enveloped in the RIFF data subchunk. The palette structure contains the palette version number, number of palette entries, the intensity of red, green and blue colours, and flags for the palette usage. The palette structure is described by the following code segment:

```
typedef struct tagLOGPALETTE {
```

```
    WORD palVersion; // Windows version number for the structure
```

```
    // Number of palettes color entries
```

```
    PALETTEENTRY palpalEntry []; // array of PALETTEENTRY data } LOGPALETTE; structure
```

the form type "AVI" and two mandatory list chunks, "hdr 1" and "n10vi".

The "hdr 1" defines the format of the data "Movi" contains the data for the audio-video streams. The third list chunk called "id xl", is an optional index chunk.

Boundary condition Handling for AVI files

Each audio and video stream is grouped together to form a rec chunk. If the size of a rec chunk is not a multiple of 2048 bytes, then the rec chunk is padded to make the size of each rec chunk a multiple of 2048 bytes. To align data on a 2048 byte boundary, dummy data is added by a "JUNK" data chunk. The JUNK chunk is a standard RIFF chunk with a 4 character identifier, "JUNK," followed by the dummy data.

MIDI File Format

The MIDI file format follows music recording metaphor to provide the means of storing separate tracks of music for each instrument so that they can be read and synchronized when they are played.

The MIDI file format also contains chunks (i.e., blocks) of data. There are two types of chunks: (i) header chunks (ii) track chunks.

Header Chunk

It is made up of 14 bytes .

- The first four-character string is the identifier string, "MThd" .
- The second four bytes contain the data size for the header chunk. It is set to a fixed value of six bytes .
- The last six bytes contain data for header chunk.

Track chunk

The Track chunk is organized as follows:

∴ The first 4-character string is the identifier.

∴ The second 4 bytes contain track length.

MIDI Communication Protocol

This protocol uses 2 or more bytes messages.

The number of bytes depends on the types of message. There are two types of messages:

- (i) Channel messages and (ii) System messages.

Channel Messages

A channel message can have up to three bytes in a message. The first byte is called a status byte, and other two bytes are called data bytes. The channel number, which addresses one of the 16 channels, is encoded by the lower nibble of the status byte. Each MIDI voice has a channel number; and messages are sent to the channel whose channel number matches the channel number encoded in the lower nibble of the status byte. There are two types of channel messages: voice messages and the mode messages.

Voice messages

Voice messages are used to control the voice of the instrument (or device); that is, switch the notes on or off and send key pressure messages indicating that the key is depressed, and send control messages to control effects like vibrato, sustain, and tremolo. Pitch wheel messages are used to change the pitch of all notes .

Mode messages

Mode messages are used for assigning voice relationships for up to 16 channels; that *is*, to set the device to MOWO mode or POLY mode. Omny Mode on enables the device to receive voice messages on all channels.

System Messages

System messages apply to the complete system rather than specific channels and do not contain any channel numbers. There are three types of system messages: common messages, real-time messages, and exclusive messages. In the following, we will see how these messages are used.

Common Messages These messages are common to the complete system. These messages provide for functions such as select a song, setting the song position pointer with number of beats, and sending a tune request to an analog synthesizer.

System Real Time Messages

These messages are used for setting the system's real-time parameters. These parameters include the timing clock, starting and stopping the sequencer, resuming the sequencer from a stopped position, and resetting the system.

System Exclusive messages

These messages contain manufacturer-specific data such as identification, serial number, model number, and other information. Here, a standard file format is generated which can be moved across platforms and applications.

JPEG Motion Image:

JPEG Motion image will be embedded in A VI RIFF file format.

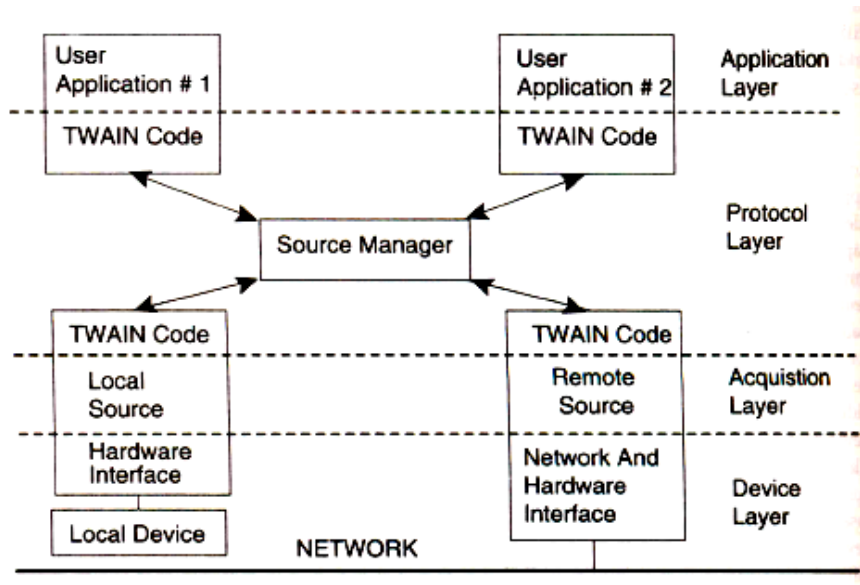
There are two standards available:

- (i) MPEG ~ In this, patent and copyright issues are there.
- (ii) MPEG 2 ~ It provide better resolution and picture quality.

TWAIN

To address the problem of custom interfaces, the TWAIN working group was formed to define an open industry standard interface for input devices. They designed a standard interface called a generic TWAIN . interface. It allows applications to interface scanners, digital still cameras, video cameras.

TWAIN ARCHITECHTURE:



- The Twain architecture defines a set of application programming interfaces (APIs) and a protocol to acquire data from input devices.
- It is a layered architecture.
- It has application layer, the protocol layer, the acquisition layer and device layer.
- Application Layer: This layer sets up a logical connection with a device. The application layer interfaces with protocol layer.
- Protocol Layer: This layer is responsible for communications between the application and acquisition layers.
- The main part of the protocol layer is the source Manager.

- Source manager manages all sessions between an application and the sources, and monitors data acquisition transactions. The protocol layer is a complex layer.

It provides the important aspects of device and application interfacing functions.

The Acquisition Layer: It contains the virtual device driver.

It interacts directly with the device driver. This layer is also known as source.

It performs the following functions:

1. Control of the device.
2. Acquisition of data from the device.
3. Transfer of data in agreed format.
4. Provision of user interface to control the device.

The Device Layer: The device layer receives software commands and controls the device hardware.

NEW WAVE RIFF File Format: This format contains two subchunks:

(i) Fmt (ii) Data.

It may contain optional subchunks:

- (i) Fact
- (ii) Cue points
- (iii) Play list
- (iv) Associated datalist.

Fact Chunk: It stores file-dependent information about the contents of the WAVE file. **Cue Points**

Chunk: It identifies a series of positions in the waveform data stream. **Playlist Chunk:** It specifies a play order for series of cue points. **Associated Data Chunk:** It provides the ability to attach information, such as labels, to sections of the waveform data stream. **Inst Chunk:** The file format stores sampled sound synthesizer's samples.

4.3 MULTIMEDIA INPUT/OUTPUT TECHNOLOGIES

Multimedia Input and Output Devices

Wide ranges of Input and output devices are available for multimedia.

Image Scanners: Image scanners are the scanners by which documents or a manufactured part are scanned. The scanner acts as the camera eye and take a photograph of the document, creating an unaltered electronic pixel representation of the original.

Sound and Voice: When voice or music is captured by a microphone, it generates an electrical signal. This electrical signal has analog sinusoidal waveforms. To digitize, this signal is converted into digital voice using an analog-to-digital converter.

Full-Motion Video: It is the most important and most complex component of Multimedia System. Video Cameras are the primary source of input for full-motion video.

Pen Driver: It is a pen device driver that interacts with the digitizer to receive all digitized information about the pen location and builds pen packets for the recognition context manager. Recognition context manager: It is the main part of the pen system. It is responsible for co-ordinating windows pen applications with the pen. It works with Recognizer, dictionary, and display driver to recognize and display pen drawn objects.

Recognizer: It recognizes hand written characters and converts them to ASCII.

Dictionary: A dictionary is a dynamic link library (DLL); The windows form pen computing system uses this dictionary to validate the recognition results.

Display Driver: It interacts with the graphics device interface and display hardware.

When a user starts writing or drawing, the display driver paints the ink trace on the screen.

Video and Image Display Systems Display System Technologies

There are variety of display system technologies employed for decoding compressed data for displaying. Mixing and scaling technology: For VGA screen, these technologies are used.

VGA mixing: Images from multiple sources are mixed in the image acquisition memory.

VGA mixing with scaling: Scalar ICs are used to sizing and positioning of images in predefined windows.

Dual buffered VGA mixing/Scaling: If we provide dual buffering, the original image is prevented from loss. In this technology, a separate buffer is used to maintain the original image.

Visual Display Technology Standards

MDA: Monochrome Display Adapter.

It was introduced by IBM .

- :. It displays 80 x 25 rows and columns .
- :. It could not display bitmap graphics .
- :. It was introduced in 1981.

CGA: Color Graphics Adapter .

- :. It was introduced in 1981.
- :. It was designed to display both text and bitmap graphics.
- it supported RGB color display,
- :. It could display text at a resolution of 640 x 200 pixels .
- :. It displays both 40 x 25 and 80 x 25 row! and columns of text characters.

MGA: Monochrome Graphics Adapter .

- :. It was introduced in 1982 .
- :. It could display both text and graphics .
- :. It could display at a resolution 720 x 350 for text and 720 x 338 for Graphics . MDA is compatible mode for this standard.

EGA: Enhanced Graphics Adapter .

- :. It was introduced in 1984 .
- :. It emulated both *MDt*. and CGA standards .
- :. It allowed the display of both text and graphics in 16 colors at a resolution of 640 x 350 pixels.

PGA: Professional Graphics Adapter.

- :. It was introduced in 1985 .
- :. It could display bit map graphics at 640 x 480 resolution and 256 colors .
- :. Compatible mode of this standard is CGA.

VGA: Video Graphics Array . :. It was introduced by IBM in 1988 .

- :. It offers CGA and EGA compatibility .
- :. It display both text and graphics .
- :. It generates analog RGB signals to display 256 colors .
- :. It remains the basic standard for most video display systems.

SVGA: Super Video Graphics Adapter. It is developed by VESA (Video Electronics Standard Association) . It's goal is to display with higher resolution than the VGA with higher refresh rates with minimize flicker.

XGA: Extended Graphics Array

It is developed by IBM . It offers VGA compatible mode . Resolution of 1024 x 768 pixels in 256 colors is offered by it. XGA utilizes an interlace scheme for refresh rates.

Flat Panel Display system

Flat panel displays use a fluorescent tube for backlighting to give the display a sufficient level of brightness. The four basic technologies used for flat panel display are:

1. Passive-matrix monochrome
2. Active-matrix monochrome
3. Passive-matrix color
4. Active-matrix color.

LCD (Liquid Crystal Display)

Construction: Two glass plates each containing a light polarizer at right angles to the other plate, sandwich the nematic (thread like) liquid crystal material.

Liquid crystal is the compounds having a crystalline arrangement of molecules. But it flow like a liquid. Nematic liquid crystal compounds are tend to keep the long axes of rod-shaped molecules aligned.

Rows of horizontal transparent conductors are built into one glass plate, and columns of vertical conductors are put into the other plate. The intersection of two conductors defines a pixel position.

Passive Matrix LCD

Working: Normally, the molecules are aligned in the 'ON' state.

Polarized light passing through the materials is twisted so that it will pass through the opposite polarizer. The light is then reflected back to the viewer. To turn off the pixel, we have to apply a voltage to the two intersecting conductors to align molecules so that the light is not twisted.

ACTIVE Matrix LCD

In this device, a transistor is placed at each pixel position, using thin-film transistor technology. The transistors are used to control the voltage at pixel locations and to prevent charge from gradually leaking out of the liquid crystal cells.

PRINT OUTPUT TECHNOLOGIES

There are various printing technologies available namely Dot matrix, inkjet, laser print server and ink jet color. But, laser printing technology is the most common for multimedia systems.

To explain this technology, let us take Hewlett Packard Laser jet-III laser printer as an example. The basic components of the laser printer are

∴ Paper feed mechanism ∴ Paper guide ∴ Laser assembly ∴ Fuser ∴ Toner cartridge.

Working: The paper feed mechanism moves the paper from a paper tray through the paper path in the printer. The paper passes over a set of corona wires that induce a change in the paper .

- The charged paper passes over a drum coated with fine-grain carbon (toner), and the toner attaches itself to the paper as a thin film of carbon .The paper is then struck by a scanning laser beam that follows the pattern of the text on graphics to be printed . The carbon particles attach themselves to the pixels traced by the laser beam . The fuser assembly then binds the carbon particles to the paper.

Role of Software in the printing mechanism:

The software package sends information to the printer to select and control printing features .

Printer drivers (files) are controlling the actual operation of the printer and allow the application software to access the features of the printer.

IMAGE SCANNERS

In a document imaging system, documents are scanned using a scanner. \The document being scanned is placed on the scanner bed or fed into the sheet feeder of the scanner .The scanner acts as the camera eye and takes a photograph of the document, creating an image of the original. The pixel representation (image) is recreated by the display software to render the image of the original document on screen or to print a copy of it.

Types of Scanners

A and B size Scanners, large form factor scanners, flat bed scanners, Rotary drum scanners and hand held scanners are the examples of scanners.

Charge-Coupled Devices All scanners use charge-coupled devices as their photosensors. CCDs consists of cells arranged in a fixed array on a small square or rectangular solid state surface. Light source moves across a document. The intensity of the light reflected by the mirror charges those cells. The amount of charge is depending upon intensity of the reflected light, which depends on the pixel shade in the document.

Image Enhancement Techniques

HalfTones In a half-tone process, patterns of dots used to build .scanned or printed image create the illusion of continuous shades of gray or continuous shades of color. Hence only limited number of shades are created. This process is implemented in news paper printers.

But in black and white photograph or color photograph, almost infinite levels of tones are used.

Dithering

Dithering is a process in which group of pixels in different patterns are used to approximate halftone patterns by the scanners. It is used in scanning original black and white photographs.

Image enhancement techniques includes controls of brightness, deskew (Automatically corrects page alignment), contrast, sharpening, emphasis and cleaning up blacknoise dots by software.

Image Manipulation

It includes scaling, cropping and rotation.

Scaling: Scaling can be up or down, the scaling software is available to reduce or enlarge. This software uses algorithms.

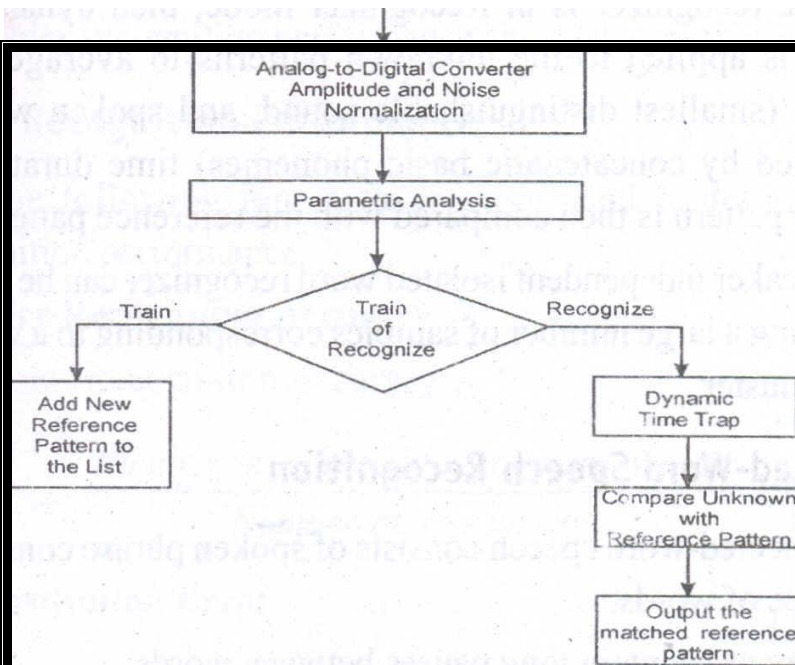
Cropping: To remove some parts of the image and to put the rest of the image as the subset of the old image.

Rotation: Image could be rotated at any degree for displaying it in different angles.

4.4 DIGITAL VOICE AND AUDIO

Digital Audio

Sound is made up of continuous analog sine waves that tend to repeat depending on the music or voice. The analog waveforms are converted into digital format by analog-to-digital converter (ADC) using sampling process.



Sampling process

Sampling is a process where the analog signal is sampled over time at regular intervals to obtain the amplitude of the analog signal at the sampling time.

Sampling rate

The regular interval at which the sampling occurs is called the sampling rate.

Digital Voice

Speech is analog in nature and is converted to digital form by an analog-to-digital converter (ADC). An ADC takes an input signal from a microphone and converts the amplitude of the sampled analog signal to an 8, 16 or 32 bit digital value.

The four important factors governing the

ADC process are

- **sampling rate**
- **resolution**
- **linearity and**
- **conversion speed.**

Sampling Rate: The rate at which the ADC takes a sample of an analog signal.

Resolution: The number of bits utilized for conversion determines the resolution of ADC.

Linearity: Linearity implies that the sampling is linear at all frequencies and that the amplitude truly represents the signal.

Conversion Speed: It is a speed of ADC to convert the analog signal into Digital signals. It must be fast enough.

VOICE Recognition System

Voice Recognition Systems can be classified into three types.

1. Isolated-word Speech Recognition.
2. Connected-word Speech Recognition.
3. Continuous Speech Recognition.

1. Isolated-word Speech Recognition.

It provides recognition of a single word at a time. The user must separate every word by a pause. The pause marks the end of one word and the beginning of the next word.

Stage 1: Normalization

The recognizer's first task is to carry out amplitude and noise normalization to minimize the variation in speech due to ambient noise, the speaker's voice, the speaker's distance from and position relative to the microphone, and the speaker's breath noise.

Stage2: Parametric Analysis

It is a preprocessing stage that extracts relevant time-varying sequences of speech parameters. This stage serves two purposes: (i) It extracts time-varying speech parameters. (ii) It reduces the amount of data of extracting the relevant speech parameters.

Training mode In training mode of the recognizer, the new frames are added to the reference list.

Recognizer mode If the recognizer is in Recognizer mode, then dynamic time warping is applied to the unknown patterns to average out the phoneme (smallest distinguishable sound, and spoken words are constructed by concatenating basic phonemes) time duration. The unknown pattern is then compared with the reference patterns.

A speaker independent isolated word recognizer can be achieved by grouping a large number of samples corresponding to a word into a single cluster.

2 Connected-Word Speech Recognition Connected-word speech consists of spoken phrase consisting of a sequence of words. It may not contain long pauses between words.

The method using Word Spotting technique

It Recognizes words in a connected-word phrase. In this technique, Recognition is carried out by compensating for rate of speech variations by the process called dynamic time warping (this process is used to expand or compress the time duration of the word), and sliding the adjusted connected-word phrase representation in time past a stored word template for a likely match.

Continuous Speech Recognition

This system can be divided into three sections:

- (i) A section consisting of digitization, amplitude normalization, time nonnalization and parametric representation.
- (ii) Second section consisting of segmentation and labeling of the speech segment into a symbolic string based on a knowledgebased or rule-based systems.
- (iii) The final section is to match speech segments to recognize word sequences.

Voice Recognition performance

It is categorized into two measures: Voice recognition performance and system performance.

The following four measures are used to determine voice recognition performance.

1. Voice Recognition Accuracy

$$\text{Voice Recognition Accuracy} = \frac{\text{Number of correctly recognized words}}{\text{Number of test words}} \times 100$$

2. Substitution Error

$$\text{Substitution error} = \frac{\text{Number of substituted words}}{\text{Number of test words}} \times 100$$

3. No Response Error

$$\frac{\text{Number of no responses}}{\text{Number of test words}} \times 100$$

4. Insertion Error

$$\text{Insertion error} = \frac{\text{Number of insertion error}}{\text{Number of test words}} \times 100$$

Voice Recognition Applications

Voice mail integration: The voice-mail message can be integrated with e-mail messages to create an integrated message.

DataBase Input and Query Applications

A number of applications are developed around the voice recognition and voice synthesis function.

The following lists a few applications which use Voice recognition.

- Application such as order entry and tracking

It is a server function; It is centralized; Remote users can dial into the system to enter an order or to track the order by making a Voice query.

- Voice-activated rolodex or address book

When a user speaks the name of the person, the rolodex application searches the name and address and voice-synthesizes the name, address, telephone numbers and fax numbers of a selected person. In medical emergency, ambulance technicians can dial in and register patients by speaking into the hospital's centralized system.

- Police can make a voice query through central data base to take follow-up action ifhe catch any suspect.
- Language-teaching systems are an obvious use for this technology. The system can ask the student to spell or speak a word. When the student speaks or spells the word, the systems performs voice recognition and measures the student's ability to spell. Based on the student's ability, the system can adjust the level of the course. This creates a self-adjustable learning system to follow the individual's pace.
- Foreign language learning is another good application where" an individual student can input words and sentences in the system. The system can then correct for pronunciation or grammar.

Musical Instrument Digital Interface (MIDI)

MIDI interface is developed by Daver Smith of sequential circuits, inc in 1982. It is an universal synthesizer interface

MIDI Specification 1.0

MIDI is a system specification consisting of both hardware and software components which define inter-connectivity and a communication protocol for electronic synthesizers, sequences, rhythm machines, personal computers, and other electronic musical instruments. The inter-connectivity defines the standard cabling scheme, connector type and input/output circuitry which enable these different MIDI instruments to be interconnected. The communication protocol defines standard multibyte messages that allow controlling the instrument's voice and messages including to send response, to send status and to send exclusive.

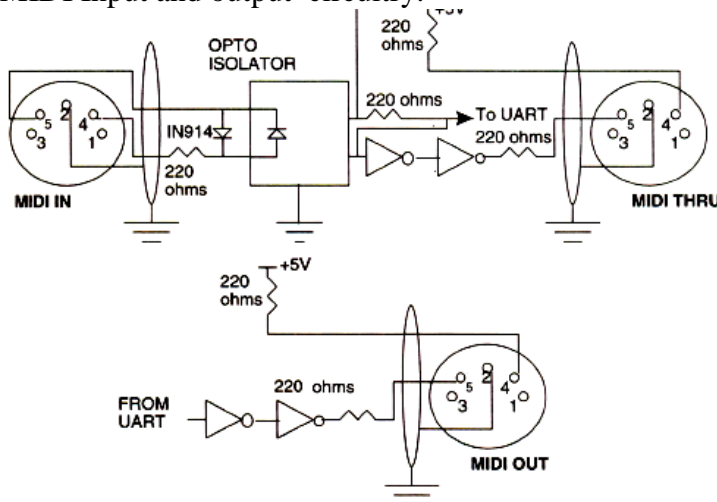
MIDI Hardware Specification

The MIDI hardware specification requires five pin panel mount receptacle DIN connectors for MIDI IN, MIDI OUT and MIDI THRU signals. The MIDI IN connector is for input signals. The MIDI OUT is for output signals. The MIDI THRU connector is for daisy-chaining multiple MIDI instruments.

MIDI Interconnections

The MIDI IN port of an instrument receives MIDI messages to play the instrument's internal synthesizer. The MIDI OUT port sends MIDI messages to play these messages to an external synthesizer. The MIDI THRU port outputs MIDI messages received by the MIDI IN port for daisy-chaining external synthesizers.

MIDI Input and output circuitry:



Communication Protocol

The MIDI communication protocol uses multibyte messages; There are two types of messages:

- (i) Channel messages
- (ii) System messages.

The channel message has three bytes. The first byte is called a status byte, and the other two bytes are called data bytes.

The two types of channel messages:

- (i) Voice messages
- (ii) Mode messages.

System messages: The three types of system messages.

Common message: These messages are common to the complete system. These messages provide for functions.

System real-time messages: These messages are used for setting the system's real-time parameters. These parameters include the timing clock, starting and stopping the sequencer, resuming the sequencer from a stopped position and restarting the system.

System exclusive message: These messages contain manufacturer specific data such as identification, serial number, model number and other information.

SOUND BOARD ARCHITECTURE A sound card consists of the following components:

MIDI Input/Output Circuitry, MIDI Synthesizer Chip, input mixture circuitry to mix CD audio input with

LINE IN input and microphone input, analog-to-digital converter with a pulse code modulation circuit to convert analog signals to digital to create WAVfiles, a decompression and compression chip to compress and decompress audio files, a speech synthesizer to synthesize speech output, a speech

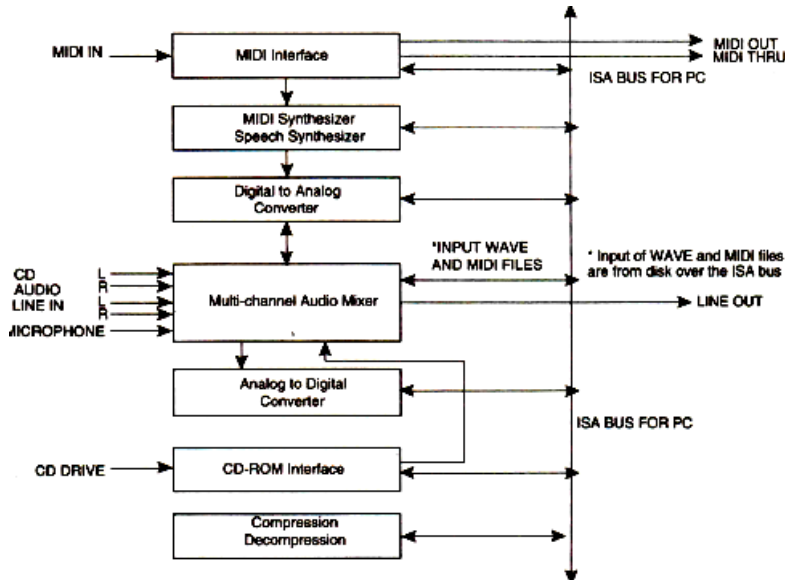
recognition circuitry to recognize speech input and output circuitry to output stereo audio OUT or LINEOUT.

AUDIO MIXER

The audio mixer component of the sound card typically has external inputs for stereo CD audio, stereo LINE IN, and stereo microphone MICIN.

These are analog inputs, and they go through analog-to-digital conversion in conjunction with PCM or ADPCM to generate digitized samples.

SOUND BOARD ARCHITECTURE:



Analog-to-Digital Converters: The ADC gets its input from the audio mixer and converts the amplitude of a sampled analog signal to either an 8-bit or 16-bit digital value.

Digital-to-Analog Converter (DAC): A DAC converts digital input in the form of WAVE files, MIDI output and CD audio to analog output signals.

Sound Compression and Decompression: Most sound boards include a codec for sound compression and decompression.

ADPCM for windows provides algorithms for sound compression.

CD-ROM Interface: The CD-ROM interface allows connecting a CD ROM drive to the sound board.

VIDEO IMAGES AND ANIMATION

VIDEO FRAME GRABBER ARCHITECTURE

A video frame grabber is used to capture, manipulate and enhance video images.

A video frame grabber card consists of video channel multiplexer, Video ADC, Input look-up table with arithmetic logic unit, image frame buffer, compression-decompression circuitry, output color look-up table, video DAC and synchronizing circuitry.

Video Channel Multiplexer:

A video channel multiplexer has multiple inputs for different video inputs. The video channel multiplexer allows the video channel to be selected under program control and switches to the control circuitry appropriate for the selected channel in a TV with multi-system inputs.

Analog to Digital Converter: The ADC takes inputs from video multiplexer and converts the amplitude of a sampled analog signal to either an 8-bit digital value for monochrome or a 24 bit digital value for colour.

Input lookup table: The input lookup table along with the arithmetic logic unit (ALU) allows performing image processing functions on a pixel basis and an image frame basis. The pixel image-processing functions are histogram stretching or histogram shrinking for image brightness and contrast, and histogram sliding to brighten or darken the image. The frame-basis image-processing functions perform logical and arithmetic operations.

Image Frame Buffer Memory: The image frame buffer is organized as a 1024 x 1024 x 24 storage buffer to store image for image processing and display.

Video Compression-Decompression: The video compression/decompression processor is used to compress and decompress still image data and video data.

Frame Buffer Output Lookup Table: The frame buffer data represents the pixel data and is used to index into the output look up table. The output lookup table generates either an 8 bit pixel value for monochrome or a 24 bit pixel value for color.

SVGA Interface: This is an optional interface for the frame grabber. The frame grabber can be designed to include an SVGA frame buffer with its own output lookup table and digital-to-analog converter.

Analog Output Mixer: The output from the SVGA DAC and the output from image frame buffer DAC is mixed to generate overlay output signals. The primary components involved include the display image frame buffer and the display SVGA buffer. The display SVGA frame buffer is overlaid on the image frame buffer or live video, This allows SVGA to display live video.

Video and Still Image Processing

Video image processing is defined as the process of manipulating a bit map image so that the image can be enhanced, restored, distorted, or analyzed.

Let us discuss about some of the terms using in video and still image processing.

Pixel point to point processing: In pixel point-to-point processing, operations are carried out on individual pixels one at a time.

Histogram Sliding: It is used to change the overall visible effect of brightening or darkening of the image. Histogram sliding is implemented by modifying the input look-up table values and using the input lookup table in conjunction with arithmetic logic unit.

Histogram Stretching and Shrinking: It is to increase or decrease the contrast.

In histogram shrinking, the brighter pixels are made less bright and the darker pixels are made less dark.

Pixel Threshold: Setting pixel threshold levels set a limit on the bright or dark areas of a picture. Pixel threshold setting is also achieved through the input lookup table.

Inter- frame image processing

Inter- frame image processing is the same as point-to-point image processing, except that the image processor operates on two images at the same time. The equation of the image operations is as follows:

Pixel output (x, y) = (Image 1(x, y)

Operator (Image 2(x, y)

Image Averaging: Image averaging minimizes or cancels the effects of random noise.

Image Subtraction: Image subtraction is used to determine the change from one frame to the next .for image comparisons for key frame detection or motion detection.

Logical Image Operation: Logical image processing operations are useful for comparing image frames and masking a block in an image frame.

Spatial Filter Processing The rate of change of shades of gray or colors is called spatial frequency. The process of generating images with either low-spatial frequency-components or high frequency components is called spatial filter processing.

Low Pass Filter: A low pass filter causes blurring of the image and appears to cause a reduction in noise.

High Pass Filter: The high-pass filter causes edges to be emphasized. The high-pass filter attenuates low-spatial frequency components, thereby enhancing edges and sharpening the image.

Laplacian Filter: This filter sharply attenuates low-spatial-frequency components without affecting and high-spatial frequency components, thereby enhancing edges sharply.

Frame Processing Frame processing operations are most commonly for geometric operations, image transformation, and image data compression and decompression Frame processing operations are very compute intensive many multiply and add operations, similar to spatial filter convolution operations.

Image scaling: Image scaling allows enlarging or shrinking the whole or part of an image.

Image rotation: Image rotation allows the image to be rotated about a center point. The operation can be used to rotate the image orthogonally to reorient the image if it was scanned incorrectly. The operation can also be used for animation. The rotation formula is:

pixel output-(x, y) = pixel input (x cos Q + y sin Q, - x sin Q + Y cos Q)

where, Q is the orientation angle

x, y are the spatial co-ordinates of the original pixel.

Image translation: Image translation allows the image to be moved up and down or side to side. Again, this function can be used for animation.

The translation formula is:

Pixel output (x, y) =Pixel Input (x + Tx, y + Ty) where

T_x and T_y are the horizontal and vertical coordinates. x, y are the spatial coordinates of the original pixel.
Image transformation: An image contains varying degrees of brightness or colors defined by the spatial frequency. The image can be transformed from spatial domain to the frequency domain by using frequency transform.

Image Animation Techniques

Animation: Animation is an illusion of movement created by sequentially playing still image frames at the rate of 15-20 frames per second.

Toggling between image frames: We can create simple animation by changing images at display time. The simplest way is to toggle between two different images. This approach is good to indicate a "Yes" or "No" type situation.

Rotating through several image frames: The animation contains several frames displayed in a loop. Since the animation consists of individual frames, the playback can be paused and resumed at any time.

4.5 FULL MOTION VIDEO

Most modem cameras use a CCD for capturing the image. HDTV video cameras will be all-digital, and the capture method will be significantly different based on the new NTSC HDTV Standard.

Full-Motion Video Controller Requirements

Video Capture Board Architecture: A full-motion video capture board is a circuit card in the computer that consists of the following components:

- (i) Video input to accept video input signals.
- (ii) S- Video input to accept RS 170 input.
- (iii) Video compression-decompression processor to handle different video compression-decompression algorithms for video data.
- (iv) Audio compression-decompression processor to compress and decompress audio data.
- (v) Analog to digital converter.
- (vi) Digital to analog converter.
- (vii) Audio input for stereo audio LINE IN, CD IN. (viii) Microphone.

A video capture board can handle a variety of different audio and video input signals and convert them from analog to digital or digital to analog.

Video Channel Multiplexer: It is similar to the video grabber's video channel multiplexer.

Video Compression and Decompression: A video compression and decompression processor is used to compress and decompress video data.

The video compression and decompression processor contains multiple stages for compression and decompression. The stages include forward discrete cosine transformation and inverse discrete cosine transformation, quantization and inverse quantization, ZigZag and Zero run-length encoding and decoding, and motion estimation and compensation.

Audio Compression: MPEG-2 uses adaptive pulse code modulation (ADPCM) to sample the audio signal. The method takes a difference between the actual sample value and predicted sample value. The difference is then encoded by a 4-bit value or 8-bit value depending upon the sample rate

Analog to Digital Converter: The ADC takes inputs from the video switch and converts the amplitude of a sampled analog signal to either an 8-bit or 16-bit digital value.

4.6 STORAGE AND RETRIVAL TECHNOLOGY

Multimedia systems require storage for large capacity objects such as video, audio and images.

Another requirement is delivery of audio and video objects. Storage technologies include battery powered RAM, Nonvolatile flash, rotating magnetic disk drives, and rotating optical disk drives: Let us discuss these technologies in detail.

MAGNETIC MEDIA TECHNOLOGY

Magnetic hard disk drive storage is a mass storage medium.

It has advantages of continual reduction in the price per mega byte of high-capacity storage.

It has high capacity and available in low cost.

In this section let us concentrate on magnetic disk I/O subsystems most applicable to multimedia uses.

HARD DISK TECHNOLOGY

Magnetic hard disk storage remains a much faster mass storage to play an important role in multimedia systems.

It remains a much faster mass storage medium than any other mass storage medium.

ST506 and MFM Hard drives: ST506 is an interface that defines the signals and the operation of signals between a hard disk controller and the hard disk. It is developed by seagate. It is used to control platter speed and the movement of heads for a drive. Parallel data is converted to a series of encoded pulses by using a scheme called MFM (modified frequency modulation). The MFM encoding scheme offers greater packing of bits and accuracy than the FM encoding scheme. Other encoding scheme is Run-Length-Limited. Its drive capacity varies from 20 M Bytes to 200 M Bytes.

ESDI Hard Drive: ESDI (Enhanced Small Device Interface) was developed by a consortium of several manufacturers. It converts the data into serial bit streams.

It uses the Run-Length-Limited Scheme for encoding. The drive has data separator circuitry Drive capacity varies from 80 M Bytes to 2 GB. ESDI interface has two ribbon cables: (i) 36 pin cable for control signals. (ii) 20 pin cable for data signals.

IDE: Integrated Device Electronics (IDE) contains a,n integrated controller with drive.

The interface is 16 bit parallel data interface. The IDE interface supports two IDE drives. One is master drive and other is slave drive. Here, Jumper setting is required. The transfer rate is 8 MHz at bus speed.

New Enhanced IDE Interface

This new interface has a transfer rate of 9-13 M Bytes/See with maximum capacity around 8 GB. It supports upto four drives CD ROM and tape drives.

SCSI (Small Computer System Interface)

It is an ANSI X3T9.2 standard which supports SCSI and SCSI2 Standards. The Standard defines both software and hardware.

SCSI-I:It defines an 8-bit parallel data path between host adapter and device.

Here, host adapter is known as initiator and the device is known as target. There are one initiator and seven targets.

Nine control signals define the activity phases of the SCSI bus during a transaction between an initiator and a target. The phases are:

(i) arbitration phase (ii) selection phase (iii) command phase (iv) data phase (v) status phase (vi) message phase (vii) bus free phase.

Arbitrary Phase: In this phase an initiator starts arbitration and tries to acquire the bus.

Selection Phase: In this phase, an initiator has acquired the bus and selects the target to which it needs to communicate.

Command Phase: The target now enters into this phase. It requests a command from the initiator. Initiator places a command on the bus. It is accepted by the target.

Data Phase: The target now enters in this phase. It requests data transfer with the initiator. The data is placed on the bus by the target and is then accepted by the initiator.

Status Phase: Now, the target enters in status phase. It indicates the end of data transfer to the initiator.

Message Phase: This is the last phase. It is to interrupt the initiator signaling completion of the read message. The bus free phase is a phase without any activity on the bus so that the bus can settle down before the next transaction. SCSI-I transfers data in 8-bit parallel form, and the transfer rate vades rom I M Bytes/See to 5 M Bytes/Sec. SCSI-I'drive capacity varies from 20 M bytes to 2 GB. SCSI-1 has over 64 commands specified to carry out transactions.

Commands include read, write, seek, enquiry, copy, verify, copy and verify, compare and so on.

SCSI-2

It has the same aspects of SCSI -1, But with faster data transfer . rates, and wider data width.

It includes few more new commands, and vender-unique command sets for optical drives, tape drives, scanners and so on. To make the bus wider, a system designer uses a second 68-pin connector in addition to the standard 50 pin connector.

Magnetic Storage Densities and Latencies

The Latency is divided into two categories: seek latency and rotational latency. Data management provides the command queuing mechanism to minimize latencies and also set-up the scatter-gather process to gather scattered data in CPU main memory.

Seek Latencies: There are three seek latencies available. They are· overlapped seek latency, Mid-transfer seek and Elevator seek.

Rotational Latencies: To reduce latency, we use two methods. They are:

(i) Zero latency read/write: Zero latency reads allow transferring data immediately after the head settles. It does not wait for disk revolution to sector property.

(ii) Interleaving factor: It keeps up with the data stream without skipping sectors. It determines the organization of sectors.

Transfer Rate and I/O per Second: I/O transfer rate varies from 1.2 M bytes/Sec. to 40 M bytes/Sec. Transfer rate is defined as the rate at which data is transferred from the drive buffer to the host adapter memory.

Data Management: It includes Command queueing and Scattergather. Command queueing allows execution of multiple sequential commands with system CPU intervention. Scatter is a process of setting the data for best fit in available block of memory or disk. Gather is a process which reassembles data into contiguous blocks on memory or disk ..

Figure below shows the relationship between seek latency, Rotational latency and Data transfer

It is a method of attaching multiple drives to a single host adapter. The data is written to the first drive first, then after filling it, the controller, allow the data to write in second drive, and so on. Meantime Between Failure (MTBF) = MTBF of single/drive/ Total no. of dr

RAID (Redundant Array of Inexpensive Disks)

It is an alternative to mass storage for multimedia systems that combines throughput speed and reliability improvements.

RAID is an array of multiple disks. In RAID the data is spread across the drives. It achieves fault tolerance, large storage capacity and performance improvement.

If we use RAID as our hot backups, it will be economy. A number of RAID schemes have been developed:

1. Hot backup of disk systems
2. Large volume storage at lower cost
3. Higher performance at lower cost
4. Ease of data recovery
5. High MTBF.

There are six levels of RAID available.

(i) RAID Level 0 Disk Striping

It spreads data across drives. Data is striped to spread segments of data across multiple drives. Data striping provides high transfer rate. Mainly, it is used for database applications.

RAID level 0 provides performance improvement. It is achieved by overlapping disk reads and writes. Overlapping here means, while segment 1 is being written to drive 1, segment 2 writes can be initiated for drive 2.

RAID Level 1 Disk Mirroring

The Disk mirroring causes two copies of every file to be written on two separate drives. (Data redundancy is achieved).

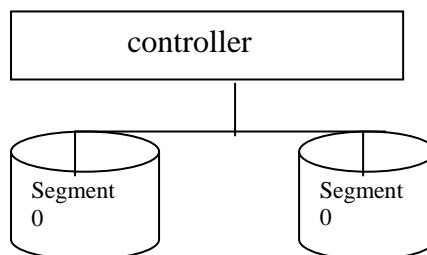
These drives are connected to a single disk controller. It is useful in mainframe and networking systems. Apart from that, if one drive fails, the other drive which has its copy can be used.

Performance:

- Writing is slow.
- Reading can be speeded up by overlapping seeks.
- Read transfer rate and number of I/O per second is better than a single drive.

I/O transfer rate (Bandwidth) = No. of drives x drive I/O transfer rate

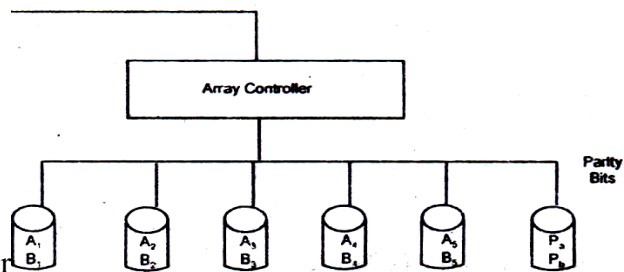
$$\text{No of I/O's Per second} = \frac{I/O \text{ transfer rate}}{\text{Average size of transfer}}$$



Disk controller arrangement for RAID Level 1

RAID Level 2, - Bit interleaving of Data: It contains arrays of multiple drives connected to a disk array controller.

Data (written one bit at a time) is bit interleaved across multiple drives. Multiple check disks are used to detect and correct errors.



Host adapter

organization of bit interleaving for RAID level 2

It provides the ability to handle very large files, and a high level of integrity and reliability. It is good for multimedia system. RAID Level 2 utilizes a hamming error correcting code to correct single-bit errors and doublebit errors.

Drawbacks:

(i) It requires multiple drives for error correction (ii) It is an expensive approach to data redundancy. (iii) It is slow.

Uses: It is used in multimedia system. Because we can store bulk of video and audio data.

RAID Level-3 Parallel Disk Array: RAID 3 subsystem contains an array of multiple data drives and one parity drive, connected to a disk array controller.

The difference between RAID 2 and RAID 3 is that RAID 3 employs only parity checking instead of the full hamming code error detection and correction. It has the advantages of high transfer rate, cost effective than RAID 2, and data integrity.

RAID Level-4 Sector Interleaving: Sector interleaving means writing successive sectors of data on different drives.

As in RAID 3, RAID 4 employs multiple data drives and typically a single dedicated parity drive.

Unlike RAID 3, where bits of data are Written to successive disk drives, an RAID 4, the first sector of a block of data is written to the first drive, the second sector of data is written to the second drive, and so on. The data is interleaved at the data level.

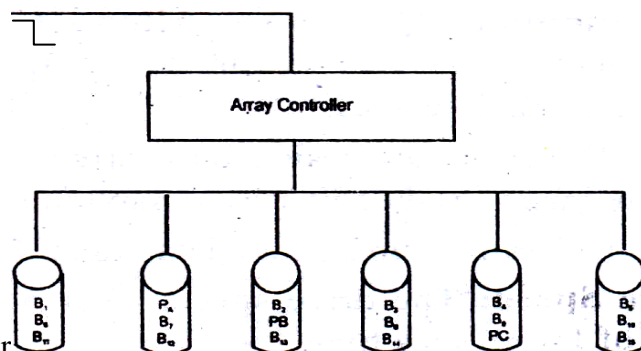
RAID Level-4 offers *cost-effective* improvement in performance with data.

RAID Level-5 Block Interleaving: In RAID Levels, as in all the other RAID systems, multiple drives are connected to a disk array controller.

The disk array controller contains multiple SCSI channels.

A RAID 5 system can be designed with a single SCSI host adapter with multiple drives connected to the single SCSI channel.

Unlike RAID Level-4, where the data is sector-interleaved, in RAID Level-5 the data is block-interleaved.



Host adapter

RAID LEVEL 5 DISK ARRAYS

Optical Media

CD ROM, WORM (Write once, Read many) and rewritable optical systems are optical drives.

•CD-ROMs have become the primary media of choice for music due to the quality of ,sound.

WORMs and erasable optical drives both use lasers to pack information densely on a removable disk.

Optical Media can be classified by technology as follows:

- CD-ROM - Compact Disc Read Only Memory
- WORM - Write Once Read Many
- Rewritable - Erasable

- Multifunction - WORM and Erasable.

-

CD-ROM

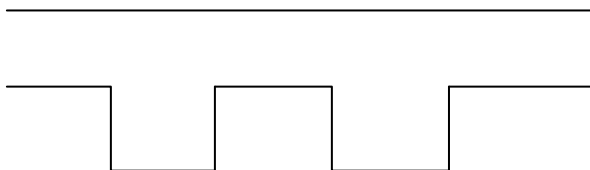
Physical Construction of CD ROMs:

It consists of a polycarbonate disk. It has 15 mm spindle hole in the center. The polycarbonate substrate contains lands and pits.

The space between two adjacent pits is called a land. Pits, represent binary zero, and the transition from land to pits and from pits to land is represented by binary one.

The polycarbonate substrate is covered by reflective aluminium or aluminium alloy or gold to increase the reflectivity of the recorded surface. The reflective surface is protected by a coat of lacquer to prevent oxidation. A CD-ROM consists of a single track which starts at the center from inside and spirals outwards. The data is encoded on this track in the form of lands and pits. A single track is divided into equal length sectors and blocks.

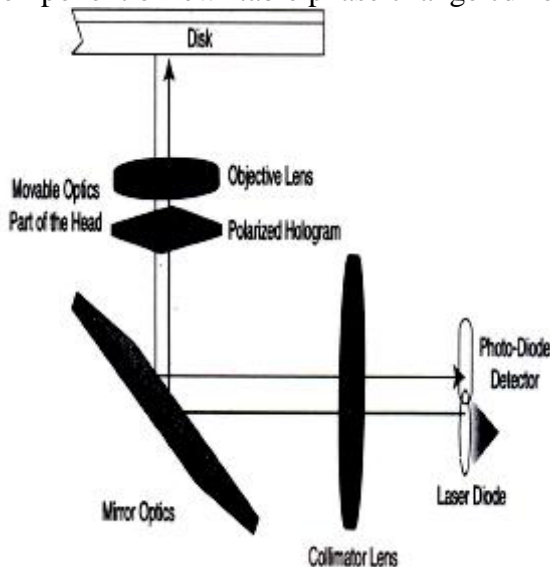
CD-ROM Physical Layers



Each sector or block consists of 2352 bytes, also called a frame. For Audio CD, the data is indexed on addressed by hours, minutes, seconds and frames. There are 75 frames in a second.

Magnetic Disk Organization: Magnetic disks are organized by CYlinder, track and sector. Magnetic hard disks contain concentric circular tracks. They are divided into sector.

Component of rewritable phase change cd-rom



Organization of magnetic media

CD-ROM Standards : A number of recording standards have emerged for CD-ROMs.

They are:

CD-DA (DD-Digital Audio) Red Book: CD-ROM is developed by philips and sony to store audio information. CD-DA is the basic medium for the music industry.

The standard specifies multiple tracks, with one song per track. One track contains one frame worth of data: 2352 bytes. There are 75 frames in a second. Bandwidth = 176 KB/s.

CD-ROM Mode 1 Yellow Book: The Mode 1 Yellow Book Standard was developed for error correction. The Yellow Book Standard dedicates 288 bytes for error detection codes (EDCs) and error correction codes (ECCs).

CD-ROM Mode 2 Yellow Book

The Mode 2 Yellow Book standard was developed for compressed audio and video applications where, due to lossy compression, data integrity is not quite as important. This standard maintains the frame structure but it does not contain the *ECC/EDC* bytes. Removing the *ECC/EDC* bytes allows a frame to

contain an additional 288 bytes of data, resulting in an increase of 14% more data. The frame structure is shown in the Table below:

Synchronization	Header	Data
12 Bytes	4 Bytes	2336 Bytes
0-11	13-15	16-2351

CD-ROMXA

XA stands for Extended Architecture. The standard was created for extending the present CD-ROM format.

CD-ROM XA contains multiple tracks. Each track's content is described by mode. CD-ROM XA also allows interleaving audio and video objects with data for synchronized playback. It does not support video compression. It supports audio compression. It uses Adaptive differential pulse Code Modulation algorithms.

CD-MO Orange Book Part 1

This standard defines an optional pre-mastered area conforming to the Red, Yellow or Green book standards for read-only, and a recordable area. It utilizes a read/write head similar to that found in magneto-optical drives. We can combine the pre-master multimedia objects as the base and develop their own versions.

CD-R Orange Book Part 2

This standard allows writing data once to a writeable disk. Here, the CD contains a polycarbonate substrate with pits and lands.

The polycarbonate layer is covered with an organic dye recording layer.

As in CD-ROM construction, the track starts from the center and spirals outwards. CD-R uses a high powered laser beam. The laser beam alters the state of the organic dye such that when the data is read, the altered state of dye disperses light instead of reflecting it. The reflected beam is measured for reading the state of each bit on the disk.

Mini-Disk

Mini-Disk for Data is known as MD-Data. It was developed by Sony Corporation. It is the data version of the new rewritable storage format. It can be used in three formats to support all users.

- A premastered optical disk.
- A recordable magneto-optical disk.
- A hybrid of mastered and recorded.

Its size is 2.5 inch. It provides large capacity. It is low cost. It is used in multimedia applications.

WORM Optical Drives

It records data using a high power laser to create a permanent burnt-in record of data. The laser beam makes permanent impressions on the surface of the disk.

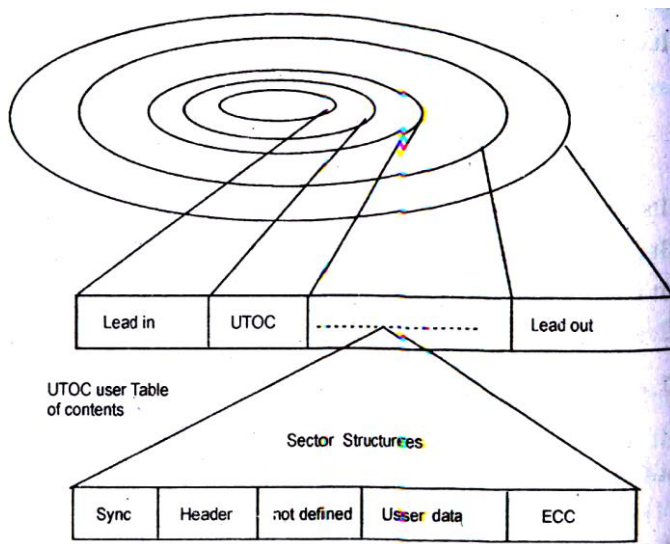
It creates pits. Information is written once. It cannot be written over and cannot be erased. i.e., Here data cannot be edited.

Recording of information: During recording, the input Signal is fed to a laser diode. The laser beam from the laser diode is modulated by the input signal. It switches the laser beam on and off. If the beam is on, it strikes the three recording layers.

The beam is absorbed by the bismuth-tellurium layer. Heat is generated within the layer. This heat diffuses the atoms in the three recording layers. It forms four-element alloy layer. Now, the layer becomes recorded layers.

Reading Information from disk:

During disk read, a weaker, laser beam is focused on to the disk. It is reflected back. The beam splitter mirror and lens arrangement sends the reflected beam to the photo detector. The photo sensor detects the beam and converts it into an electrical signal.



WORM DRIVE Applications

- On-line catalogs
- Large-volume distribution
- Transaction logging
- Multimedia archival.

Rewritable Optical Disk Technologies

This technology allows erasing old data and rewriting new data over old data. There are two types of rewritable technology: (i) Magneto-optical (ii) Phase change.

Magneto-Optical Technology

It uses a combination of magnetic and laser technology to achieve read/write capability. The disk recording layer uses a weak magnetic field to record data under high temperature. High temperature is achieved by laser beam.

When the beam is on, it heats the spot on the magneto optical disk to its curie temperature. The rise in temperature makes the spot extra sensitive to the magnetic field of bias field.

Magneto-optical drives require two passes to write data; in the first pass, the magneto optical head goes through an erase cycle, and in the second pass, it writes the data.

During the erase cycle, the laser beam is turned on and the bias field is modulated to change the polarity of spots to be erased. During the write cycle, the bias field is turned on and the laser beam is modulated to change the polarity of some spots to 1 according to the bit value.

Phase change Rewritable optical Disk

In phase change technology the recording layer changes the physical characteristics from crystalline to amorphous and back under the influence of heat from a laser beam.

To read the data, a low power laser beam is transmitted to the disk. The reflected beam is different for a crystalline state than for an amorphous state. The difference in reflectivity determines the polarity of the spot.

Benefits: it requires only one pass to write.

Dye Polymer Rewritable Disk

There is no need of magnetic technology here.

This technology consists of giant molecules formed from smaller molecules of the same kind with light-sensitive dye. This technology is also used in WORM drives.

HIERARCHICAL STORAGE MANAGEMENT

multi-function drive is a single drive unit. It is capable of reading and writing a variety of disk media.

Three types of technologies are used for multi-function drives. They are:

- Magneto-optical disk for both rewritable and WORM capability.
- Magneto-optical disk for rewritable and dye polymer disk for WORM capability.
- Phase change technology for both rewritable and WORM capability.

The storage hierarchies described in the pyramid consist of random access memory (RAM), on-line fast magnetic hard disks, optical disks and juke boxes, diskettes, and tapes.

Permanent Vs. Transient Storage issues

The process of moving an object from one level in the storage hierarchy to another level in that hierarchy is called migration. Migration to objects to off-line media and removal of these objects from on-line media is called archiving. Migration Can be set up to be manual or automatic.

Manual migration requires the user or the system administrator to move objects from one level of storage to another level. Systems with automatic migration perform this task automatically. In document-imaging systems, compressed image files are created in magnetic cache areas on fast storage devices when documents are scanned.

Optical Disk Library (Juke box)

An optical juke box stacks disk platters to be played. In the optical disk library, the platters are optical and contain objects such as data, audio, video, and images.

An optical disk library has one or more optical drives. An optical disk library uses a very-high-speed and accurate server-controlled electromechanical robotics elevator mechanism for moving the optical platters between their slots on a disk stacks and the drives. The robotics mechanism removes disk platter from a drive and returns it to its slots on the stack after the disk has finished playing (usually when the drive is required for another disk). The robotics device operates and manages multiple drives under program control. .

A juke box may contain drives of different types, including WORM, rewritable, or multifunction. Juke boxes contain one or more drives. A juke box is used for storing large volumes of multimedia information in one cost effective store.

Juke box-based optical disk libraries can be networked so that multiple users can access the information. Optical disk libraries serve as near-line storage for infrequently used data.

Hierarchical Storage Applications: Banks, insurance companies, hospitals, state and federal governments, manufacturing companies and a variety of other business and service organizations need to permanently store large volumes of their records, from simple documents to video information, for audit trail use.

CACHE MANAGEMENT FOR STORAGE SYSTEMS

Disk caches are an integral part of a hierarchical storage management architecture. Hierarchical storage consists of a number of media classes, ranging from high speed and expensive on-line fast cache storage to low-cost off-line storage.

Role of on-line caches: The primary role of on-line caches as used in document-imaging systems is to provide high speed on-line storage for documents currently in use that may be accessed in the future. This role can be extended to multimedia systems.

Hierarchical Organization of Caches: Caches are used at various storage levels.

The following lists representative storage systems using cache storage.

- ~ Hardware disk caches and system memory caches for stand alone systems.
- ~ Disk storage caches for optical disk libraries.
- ~ Disk storage caches for networked systems.

Low-level Disk Caching

Disk Caching Controllers: Two approaches to implement disk caching controllers are: in hardware and software. A hardware caching controller is designed with its own on-board CPU and private memory

The private memory is used for storing disk data temporarily. It is known as disk cache.

When an I/O request is received by the caching controller, the CPU on the caching controller initiates sectors of data, including sectors that contain the requested data.

Disk writes managed through a disk cache can be delayed writes or write-throughs. For delayed through, data is written by the host CPU to the disk cache, and the caching controller writes the data from the disk cache to the disk when read activity is low.

Cache Organization for Hierarchical Storage Systems: The hierarchical, storage management system consists of at least three or four types of storage as follows:

- System memory Cache
- On-line high speed magnetic disk storage
- Near-line optical disk libraries
- Off-line optical tape storage
-

Many Cache designs use a high-water mark and a low-water mark to trigger cache management operations. When the cache storage fills up to the high-water mark, the cache manager starts creating more space in cache storage. Space is created by discarding objects.

- The cache manager maintains a data base of objects in the cache. Cache areas containing updated objects are frequently called dirty cache.
- Objects in dirty cache are written back at predetermined time intervals or before discarding an object.

2 - MARKS

1. What do you mean by compression and decompression.
2. State the types of compression.
3. What is Huffman Encoding?
4. Write short notes on Packbits Encoding?
5. Write down about CCITT group 3-I-D compression.
6. What is the necessity of k factor in CCITT Group-3 2D compression.
7. What are the steps involved for pseudo code to code line in data formatting for CCITT Group-3 2D?
8. State the advantages of CCITT Group-3 2D.
9. Write short notes on YUV representation.
10. What is Discrete Cosine Transform?
11. Write short notes on Quantization.
12. What are the color characteristics.
13. What is predictive lossless encoding for?
14. What is the role of entropy encoding in data compression.
15. What is Macroblock in MPEG?
16. Define the term "Motion Compensation".
17. Write short notes on MPEG encoder.
18. What do you mean by vector quantization in MPEG standard.
19. What do you mean by audio compression.
20. Write short notes on fractal compression.
21. List the key formats available in Rich Text Format.
22. What is TIFF File Format?
23. What is TIFF tag ID?
24. What is full motion video?
25. What are the advantages of dye sublimation printer.
26. List the features of scanner.
27. What is ADC stands for?
28. Write short notes on histogram sliding. What is dithering?
29. What do you mean by RIFF-chunk
30. Define LIST chunk
31. Describe TWAIN architecture
32. How do scanner work.
33. Give some of the visual display technology standards
34. State the four important factors that governs ADC process.
35. State the three types of Voice Recognition system.
36. Write short notes on MIDI.
37. What is digital camera? State its advantages.
38. What do you mean by Spatial Filter Processing.
39. Write short notes on disk spanning.
40. What is RAID?
41. State the uses of magnetic storage in multimedia.
42. Give brief notes on CD-ROM.
43. What are the three formats of minidisk?
44. What is juke box. Give other name used for juke box.
45. What are the four types of storage in cache organization for hierarchical storage systems.

16 Marks

1. Explain briefly about binary image compression schemes. (16)
- 2.(a) Explain the characteristics of color in detail. (10)
(b) State the relationship between frequency and wave length! in measuring radian energy.(6)
3. Write about JPEG in detail. (16)
- 4.(a) Explain DCT in detail. (12)
(b) Write short notes on zig zag sequence. (4)
5. State the requirements for full motion video compression in detail. (16)
6. What is MPEG? Discuss it in detail. (16)
7. Explain all data and file format standards in detail. (16)
8. Give a detailed description about voice recognition
(b) What is DIB stands for? (4)
9. Explain the different types of messages that are used with MIDI communication protocol. (16)
10. Explain the TWAIN architecture with neat diagram. (16)
11. Explain some of the video and image display system technique in detail. (16)
12. Give short notes for the following standards: (i) MDA, (3), (ii) CGA (3), (iii) MGA (3); (iv) VGA (4), (v) XGA (3).
- 13.(a) Discuss CRT with neat diagram. (10)
(a) Discuss flat panel display system briefly. (8)
(b) Give a short description about laser printer technology. (6)
© State the role of software in the printing mechanism. (2)
14. Explain in detail about voice recognition system. (12)
(b) Write down the formula for voice recognition accuracy and substitution error.
16. a) Describe hierarchical storage management in detail.
(b) What is migration? Write short notes about it.
17. (i) What is Optical Disk Library? Explain it.
(ii) Discuss about Cache Management for storage system.