Computer Graphics And Multimedia Applications SCS1302

Unit 5

Syllabus

- UNIT V MULTIMEDIA BASICS AND TOOLS
- **Introduction to multimedia** - Compression & Decompression - Data & File Format standards - Digital voice and audio - Video image and animation. **Introduction to Photoshop** – Workplace – Tools – Navigating window – Importing and exporting images – Operations on Images – resize, crop, and rotate. **Introduction to Flash** – Elements of flash document – flash environment – Drawing tools – Flash animations – Importing and exporting - Adding sounds – Publishing flash movies – Basic action scripts – GoTo, Play, Stop, Tell Target.

INTRODUCTION TO MULTIMEDIA

- Multimedia is a combination of text, graphic art, and sound, animation and video elements.
- The IBM dictionary of computing describes multimedia as "comprehensive material, presented in a combination of text, graphics, video, animation and sound. Any system that is capable of presenting multimedia, is called a multimedia system".
- A multimedia application accepts input from the user by means of a keyboard, voice or pointing device.
- Multimedia applications involve using multimedia technology for business, education and entertainment. Multimedia is now available on standard computer platforms.

APPLICATIONS

- **Business** - In any business enterprise, multimedia exists in the form of advertisements, presentations, video conferencing, voice mail, etc.

- **Schools** - Multimedia tools for learning are widely used these days. People of all ages learn easily and quickly when they are presented information with the visual treat.

- Home PCs equipped with CD-ROMs and game machines hooked up with TV screens have brought home entertainment to new levels. These multimedia titles viewed at home would probably be available on the multimedia highway soon.

- **Public places** - Interactive maps at public places like libraries, museums, airports and the stand-alone terminal

- **Virtual Reality (VR)** - This technology helps us feel a 'real life-like' experience. Games using virtual reality effect is very popular

- **Multimedia Elements :**
- Multimedia applications require dynamic handling of data consisting of a mix of text, voice, audio components, video components, and image animation.
- Integrated multimedia applications allow the user to cut sections of all or any of these components and paste them in a new document or in another application such as an animated sequence of events, a. desktop publishing system, or a spreadsheet.
- **Facsimile:**
- Facsimile transmissions were the first practical means of transmitting document images over telephone lines.
- to allow higher scanning density for better-quality fax
- **Document images :**
- Document images are used for storing business documents that must be retained for long periods of time
- Providing multimedia access to such documents removes the need far making several copies of the original for storage or distribution

- **<u>Photographic images :</u>**
- Photographic images are used for a wide range of applications . such as employee records for instant identification at a security desk, real estates systems with photographs of houses in the database containing the description of houses, medical case histories, and so on.
- **<u>Geographic information systems map (GIS):</u>**
- Map created in a GIS system are being used wildly for natural resources and wild life management as well as urban planning.
- **<u>Voice commands and voice synthesis:</u>**
- used for hands-free operations of a computer program.
- **<u>Audio message:</u>**
- Annotated voice mail already uses audio or voice message as attachments to memos and documents such as maintenance manuals.

- **Video messages:**
- Video messages are being used in a manner similar to annotated voice mail.
- **Holographic images:**
- Holographic images extend the concept of virtual reality by allowing the user to get "inside" a part, such as, an engine and view its operation from the inside.
- **Fractals:**
- This technology is based on synthesizing and storing algorithms that describes the information.

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.
- 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

- **Lossy compression:**
- 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.
- **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
- Eg: Text Compression

- 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:
- • Packpits encoding (Run-length encoding)
- • CCITT Group 3 I D
- • CCITT Group 3 2D
- • CCITT Group 4
- • Lempe l-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.
- 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.

Binary Image compression schemes

- A binary image containing black and white pixel is generated when a document is scanned in a binary mode.
- The schemes are applicable in office/business documents, handwritten text, line graphics, engineering drawings, and so on.
- 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.
- 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.
- 1. Pack pits Encoding(Run-Length Encoding)
- 2. CCITT Group 3 1-D Compression

- **1. Pack pits Encoding(Run-Length Encoding)**
- It is a scheme in which a consecutive repeated string of characters is replaced by two bytes.
- 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.

- **2. CCITT Group 3 1-D Compression**
- This scheme is based on run-length encoding and assumes that a typical scan line has long runs of the same color.
- This scheme was designed for black and white images only, not for gray scale or color images.

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.

Huffman code

- In computer science and information theory, a **Huffman code** is a particular type of optimal prefix **code** that is commonly used for lossless data compression.
- The output from **Huffman's** algorithm can be viewed as a variable-length **code** table for **encoding** a source symbol (such as a character in a file).

Huffman code

B	C	A	A	D	D	D	C	C	A	C	A	C	A	C

Initial string

Each character occupies 8 bits. There are a total of 15 characters in the above string. Thus, a total of `8 * 15 = 120` bits are required to send this string.

Using the Huffman Coding technique, we can compress the string to a smaller size.

Huffman coding first creates a tree using the frequencies of the character and then generates code for each character.

Once the data is encoded, it has to be decoded. Decoding is done using the same tree.

Huffman Coding prevents any ambiguity in the decoding process using the concept of **prefix code** ie. a code associated with a character should not be present in the prefix of any other code. The tree created above helps in maintaining the property.

Huffman code

Huffman coding is done with the help of the following steps.

1. Calculate the frequency of each character in the string.

1	6	5	3
B	C	A	D

Frequency of string

2. Sort the characters in increasing order of the frequency. These are stored in a priority queue `Q`.

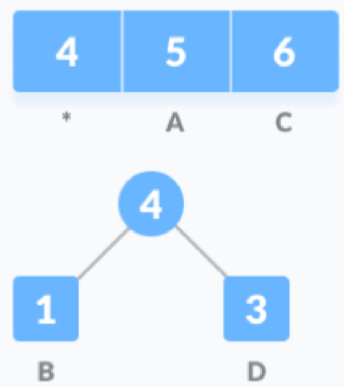
1	3	5	6
B	D	A	C

Characters sorted according to the frequency

Huffman code

3. Make each unique character as a leaf node.

4. Create an empty node `z`. Assign the minimum frequency to the left child of z and assign the second minimum frequency to the right child of `z`. Set the value of the `z` as the sum of the above two minimum frequencies.



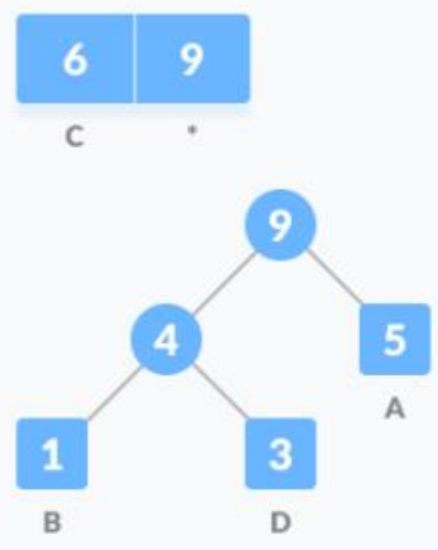
Getting the sum of the least numbers

5. Remove these two minimum frequencies from `Q` and add the sum into the list of frequencies (* denote the internal nodes in the figure above).

Huffman code

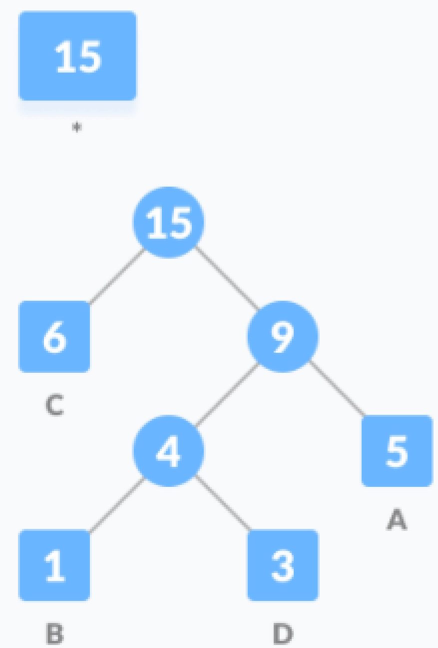
6. Insert node z into the tree.

7. Repeat steps 3 to 5 for all the characters.



Repeat steps 3 to 5 for all the characters.

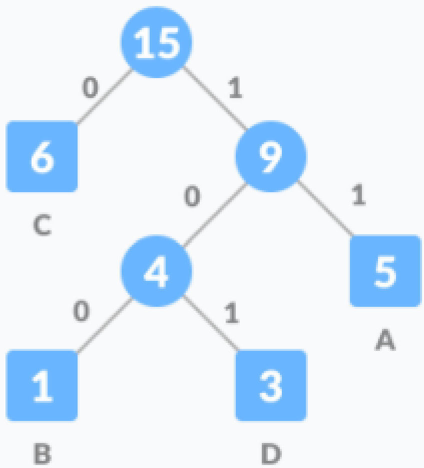
Huffman code



Repeat steps 3 to 5 for all the characters.

Huffman code

8. For each non-leaf node, assign 0 to the left edge and 1 to the right edge.



Assign 0 to the left edge and 1 to the right edge

Huffman code

For sending the above string over a network, we have to send the tree as well as the above compressed-code. The total size is given by the table below.

Character	Frequency	Code	Size
A	5	11	5*2 = 10
B	1	100	1*3 = 3
C	6	0	6*1 = 6
D	3	101	3*3 = 9
4 * 8 = 32 bits	15 bits		28 bits

Huffman code

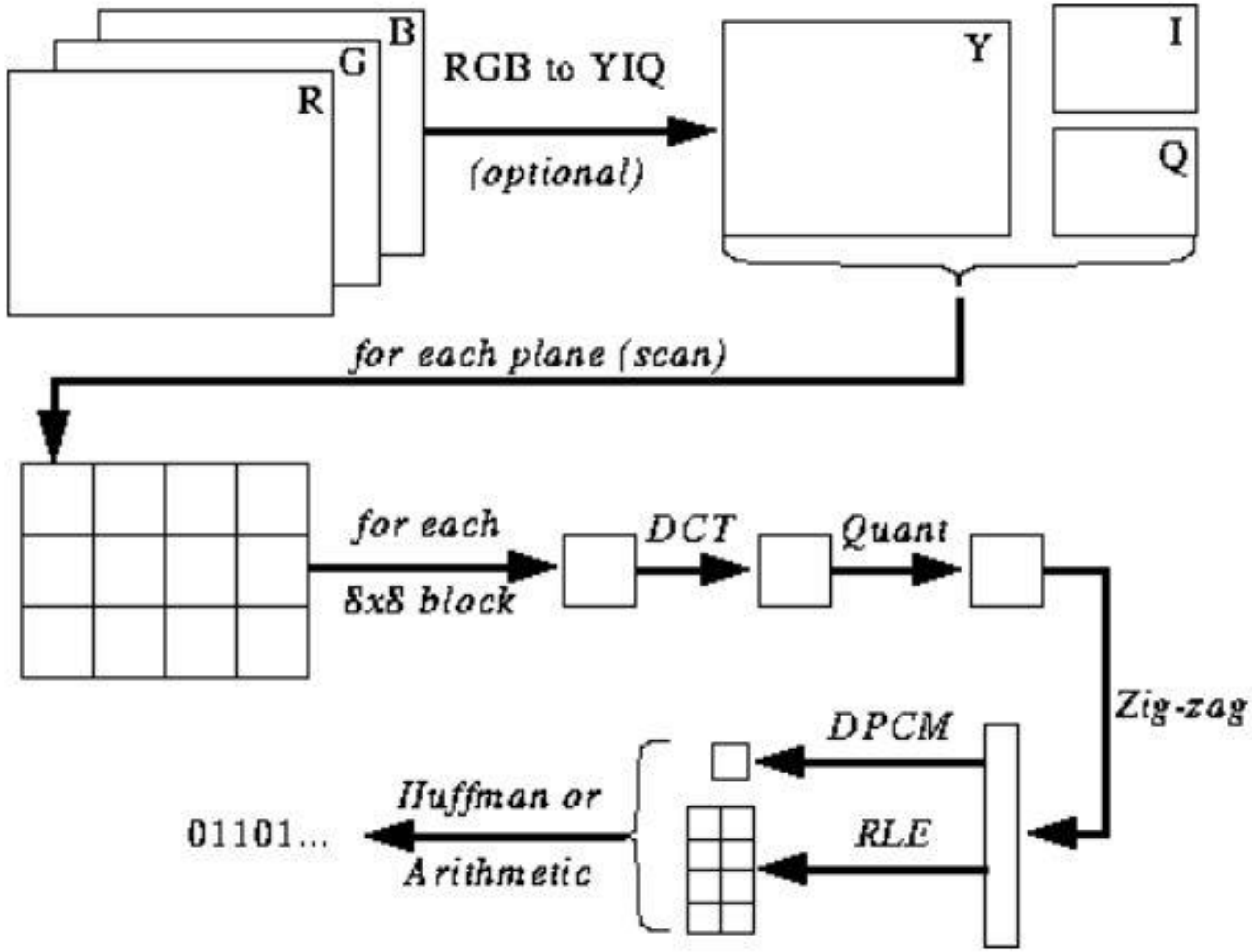
Without encoding, the total size of the string was 120 bits. After encoding the size is reduced to `32 + 15 + 28 = 75`.

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

JPEG Encoding



JPEG Encoding

- Decoding - Reverse the order for encoding

- The Major Steps in JPEG Coding involve:
- DCT (Discrete Cosine Transformation)
- Quantization
- Zigzag Scan
- DPCM on DC component
- RLE on AC Components
- Entropy Coding

Discrete Cosine Transform(DCT)

- Spatial domain ⟺ Frequency domain.
- Outputs DCT coefficients (containing spatial frequencies), which relate directly to how much the pixel values change as function of their position in the block:
- A lot of variations in pixel values: Represents an image with a lot of fine detail.
- Small variations in pixel values: Uniform color change and little fine detail.
- When there is little variations in pixel values, only a few data points are required to represent the image.
- DCT does not provide any compression:
- ◦ Rearranges the data into a form that allows another coding technique to compress the data more effectively

Quantization

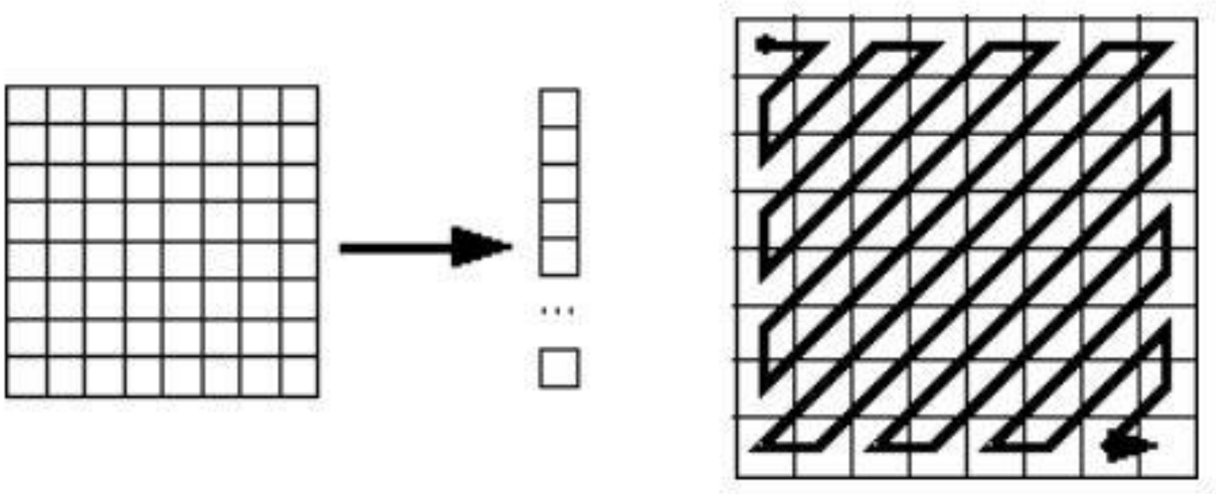
- Used to throw out bits
- ***Example***: 101101 = 45 (6 bits).
- Truncate to 4 bits: 1011 = 11.
- Truncate to 3 bits: 101 = 5.
- Quantization error is the main source of the Lossy Compression.
- **Quantization Tables**
- In JPEG, each F[u,v] is divided by a constant q(u,v).
- Table of q(u,v) is called *quantization table*.

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

- Eye is most sensitive to low frequencies (upper left corner), less sensitive to high frequencies (lower right corner)

Zig-zag Scan

- What is the purpose of the Zig-zag Scan:
- to group low frequency coefficients in top of vector.
- Maps 8 x 8 to a 1 x 64 vector
-



- **<u>Differential Pulse Code Modulation (DPCM) on DC component</u>**
- Here we see that besides DCT another encoding method is employed: DPCM on the DC component at least. Why is this strategy adopted:
- DC component is large and varied, but often close to previous value (like lossless JPEG).
- Encode the difference from previous 8x8 blocks – DPCM

- **<u>Run Length Encode (RLE) on AC components</u>**
- Yet another simple compression technique is applied to the AC component:
- 1x64 vector has lots of zeros in it
- Encode as (*skip*, *value*) pairs, where *skip* is the number of zeros and *value* is the next non-zero component.
- Send (0,0) as end-of-block sentinel value.

- **<u>Entropy Coding</u>**
- DC and AC components finally need to be represented by a smaller number of bits
- Categorize DC values into SSS (number of bits needed to represent) and actual bits.

Value	SSS
0	0
-1,1	1
-3,-2,2,3	2
-7..-4,4..7	3

- *Example*: if DC value is 4, 3 bits are needed.
- Send off SSS as Huffman symbol, followed by actual 3 bits.
- For AC components (*skip*, *value*), encode the composite symbol (*skip*,*SSS*) using the Huffman coding.
- Huffman Tables can be custom (sent in header) or default.

Summary of the JPEG bit stream

- JPEG components have described how compression is achieved at several stages. Let us conclude by summarizing the overall compression process:
- A "Frame" is a picture, a "scan" is a pass through the pixels (e.g., the red component), a "segment" is a group of blocks, a "block" is an 8x8 group of pixels.
- Frame header: sample precision (width, height) of image number of components unique ID (for each component) horizontal/vertical sampling factors (for each component) quantization table to use (for each component)
- Scan header Number of components in scan component ID (for each component) Huffman table for each component (for each component)
- Misc. (can occur between headers) Quantization tables Huffman Tables Arithmetic Coding Tables Comments Application Data

Moving Picture Experts Group Compression

- MPEG stands for "Moving Picture Experts Group.“
- MPEG is an organization that develops standards for encoding digital audio and video.
- It works with the International Organization for Standardization (ISO) and the International Electro technical Commission (IEC) to ensure media compression standards are widely adopted and universally available.
- The MPEG organization has produced a number of digital media standards since its inception in 1998. Examples include:

- **MPEG-1** – Audio/video standards designed for digital storage media (such as an MP3 file)
- **MPEG-2** – Standards for digital television and DVD video
- **MPEG-4** – Multimedia standards for the computers, mobile devices, and the web
- **MPEG-7** – Standards for the description and search of multimedia content
- **MPEG-MAR** – A mixed reality and augmented reality reference model
- **MPEG-DASH** – Standards that provide solutions for streaming multimedia data over HTTP (such as servers and CDNs)

- Using MPEG compression, the file size of a multimedia file can be significantly reduced with little noticeable loss in quality.
- This makes transferring files over the internet more efficient, which helps conserve Internet bandwidth.
- MPEG compression is so ubiquitous that the term "MPEG" is commonly used to refer to a video file saved in an MPEG file format rather than the organization itself.
- These files usually have a ".mpg" or ".mpeg" file extension.
- **File extensions: .MP3, .MP4, .M4V, .MPG, .MPE, .MPEG**

MPEG compression removes two types of redundancies:

1. **Spatial redundancy:**

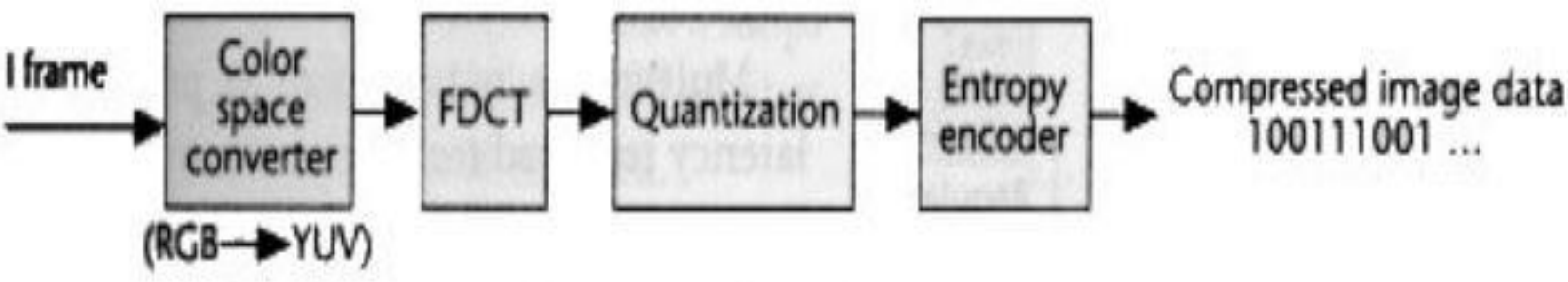
- the value of a pixel is predictable given the values of neighboring pixels.
- It is removed with the help of DCT compression.

2. **Temporal redundancy:**

- Pixels in two video frames that have the same values in the same location .
- It is removed with the help of Motion compensation technique.

MPEG constructs three types of pictures namely:

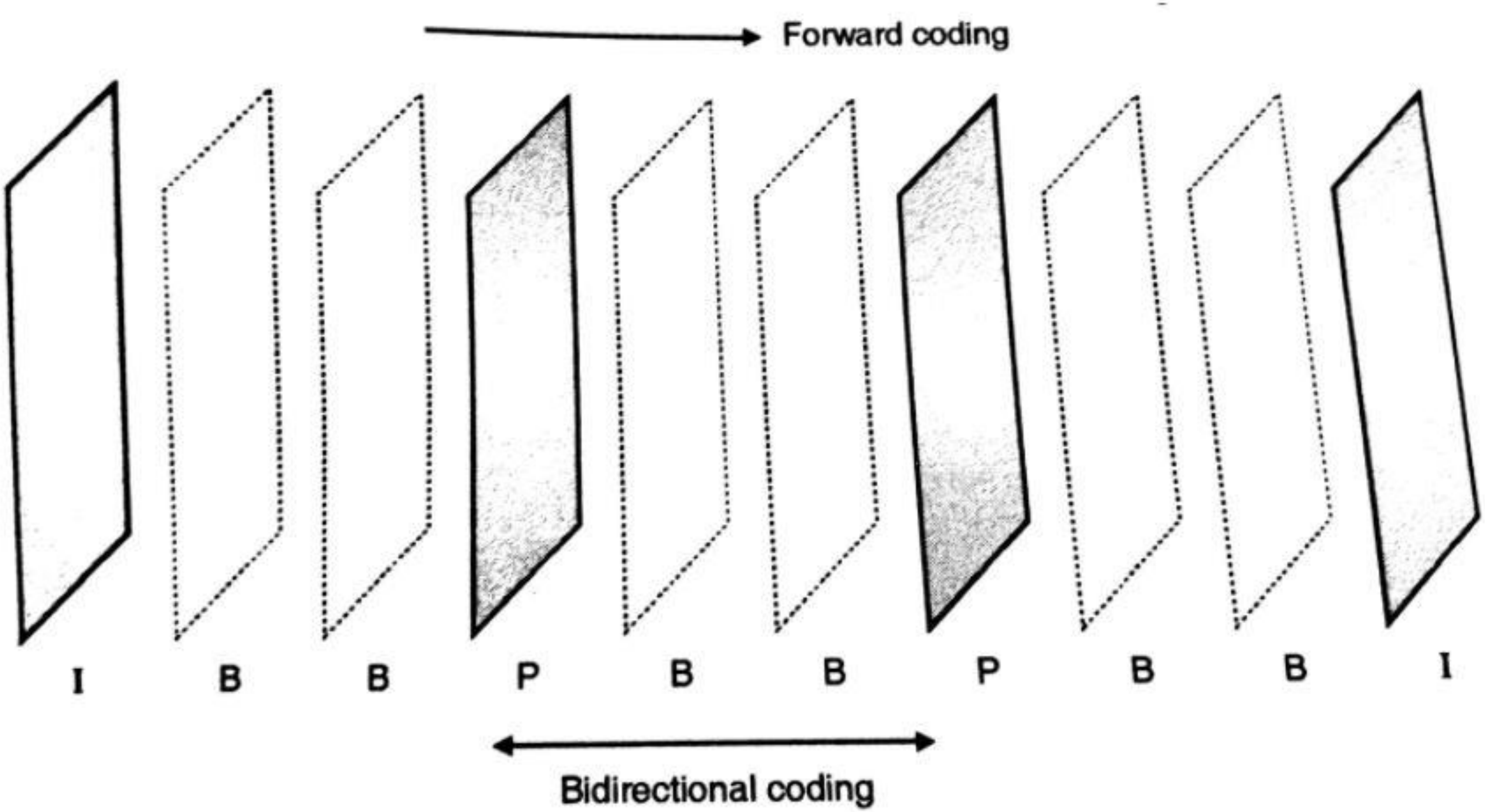
- Intra pictures (I-pictures)
- Predicted pictures (P-pictures)
- Bidirectional predicted pictures (B-pictures)
- **The MPEG algorithm employs following steps:**
- **Intra frame DCT coding (I-pictures):**
- The I-pictures are compressed as if they are JPEG images.



- **Motion-compensated inter-frame prediction (P-pictures):**

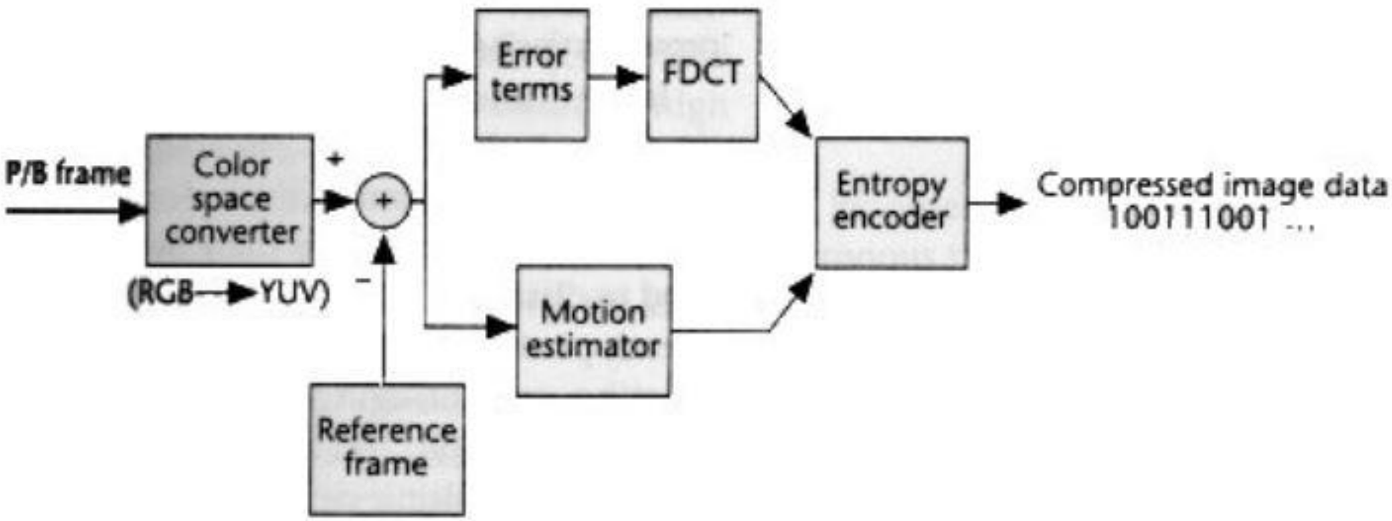
- In most video sequences there is a little change in the contents of image from one frame to the next.
- Most video compression schemes take advantage of this redundancy by **using the previous frame to generate a prediction of current frame.**
- This is based on current value to predict next value and code their difference called as **prediction error.**
- The frame to be compared is split in to blocks first and then best matching block is searched.
- Each block uses previous picture for estimating prediction.
- This search process is called as prediction.

- **B-frame (Bidirectional predictive frame):**
- -Frames can also be predicted from future frames.
- Such frames are usually predicted from two directions, i.e. from the I- or P-frames that immediately precede or follow the predicted frame.
- -These bidirectionally predicted frames are called B-frames.
- A coding scheme could, for instance, be IBBPBBPBBPBB.
- -B-pictures uses the previous or next I-frame or P-frame for motion compensation and offers the highest degree of compression.
- Each block in a B-picture can be forward, backward or bidirectionally predicted.


Forward coding
I
B
B
P
B
B
P
B
B
I
Bidirectional coding

Bidirectional predicted pictures (B):

- Bidirectional predicted pictures utilize three types of motion compression techniques.
- Forward motion compensation - uses past picture information.
- Backward motion compensation - uses future picture information .
- Bidirectional compensation - uses the average of the past and future picture information.


P/B frame
Color space converter
(RGB → YUV)
+
−
Error terms
FDCT
Motion estimator
Reference frame
Entropy encoder
Compressed image data
100111001 ...

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) 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.

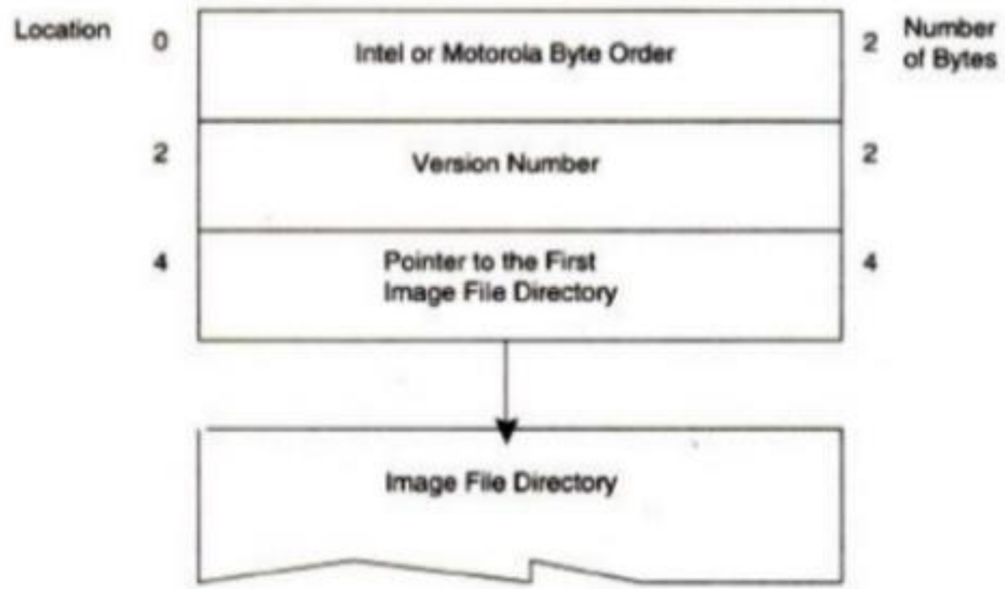
The additional formats are:

(i) Tiled images, compression schemes, images using CMYK, $YCbC_r$ color models.

TIFF Structure

TIFF files consists of a header. The header consists of byte ordering 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 arc grouped into several categories. They are Basic, Informational, Facsimile, Document storage and Retrieval.

TIFF Classes

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 Inter change 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 typl	File extension
Waveform Audio File	WAVE	.WAV
Audio Video Interleaved file	AVI	.AVI
MIDI File	RMID	.RMI
Device Independent Bitmap file	RDIB	.RDI
Pallette 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.

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 syn~hronized 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 sent 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, ressuming the sequencer from a stopped position, and resetting the system.

System Exclusive messages

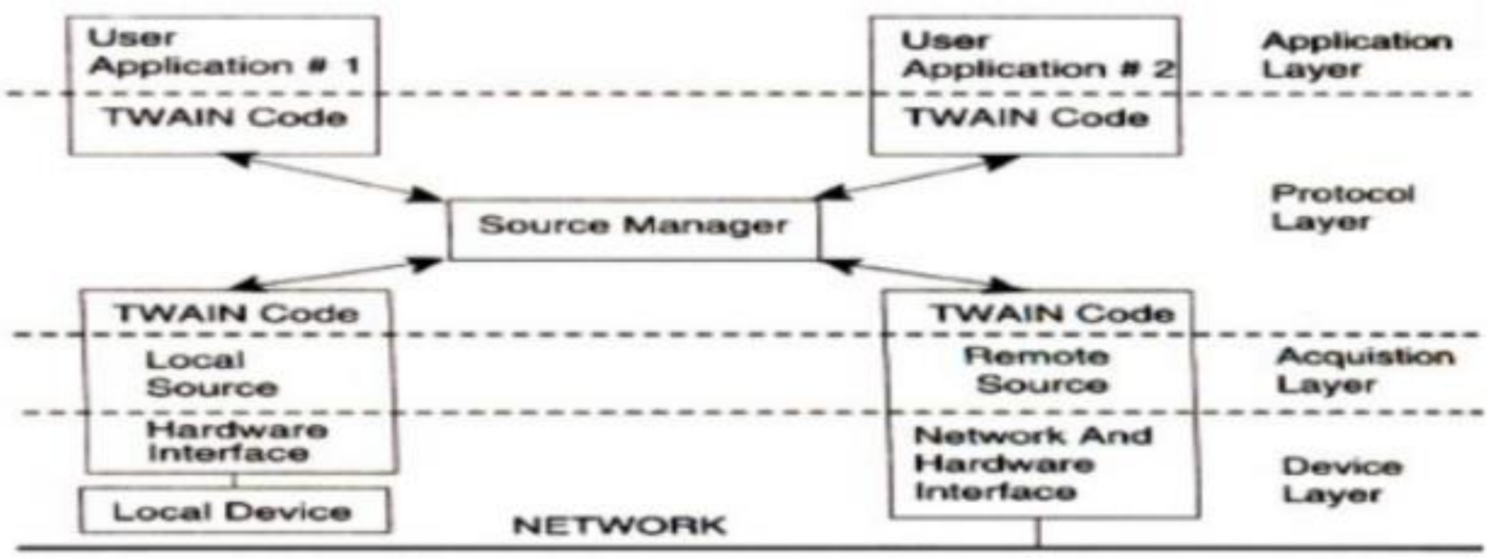
These messages contain manufacturer-specific data such as identification, serial number, model number, and other information. Here, a standard file format is generated which can be moved across platforms and applications.

TWAIN

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:



TWAIN

- 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.

TWAIN

- The Device Layer: The device layer receives software commands and controls the device hardware. NEW WAVE RIFF File Format: This format contains two sub chunks:
- (i) Fmt (ii) Data.

TWAIN

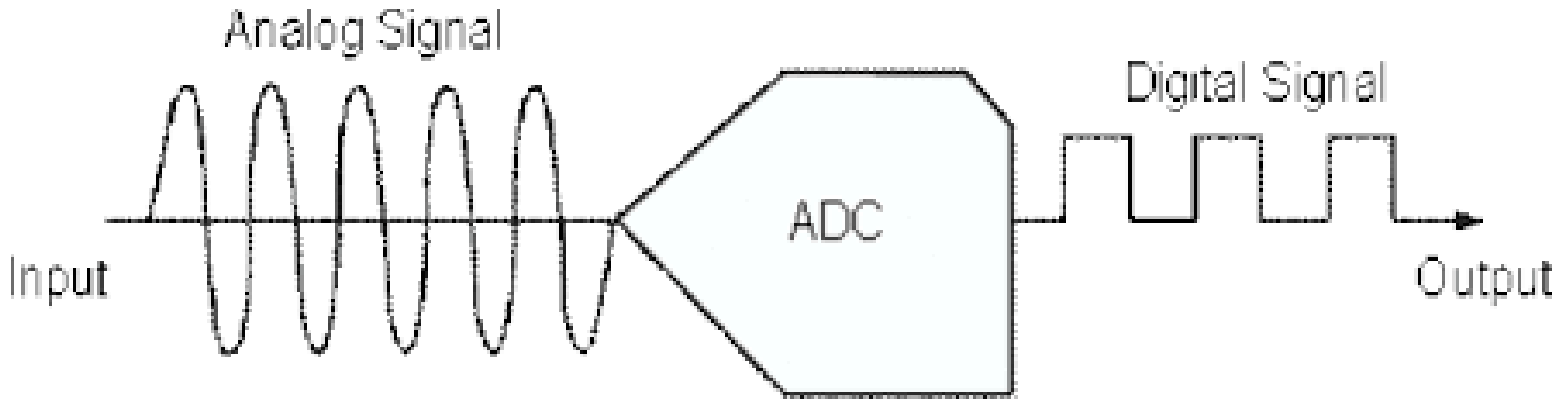
It may contain optional subchunks:

(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.

DIGITAL VOICE AND AUDIO
Digital Audio

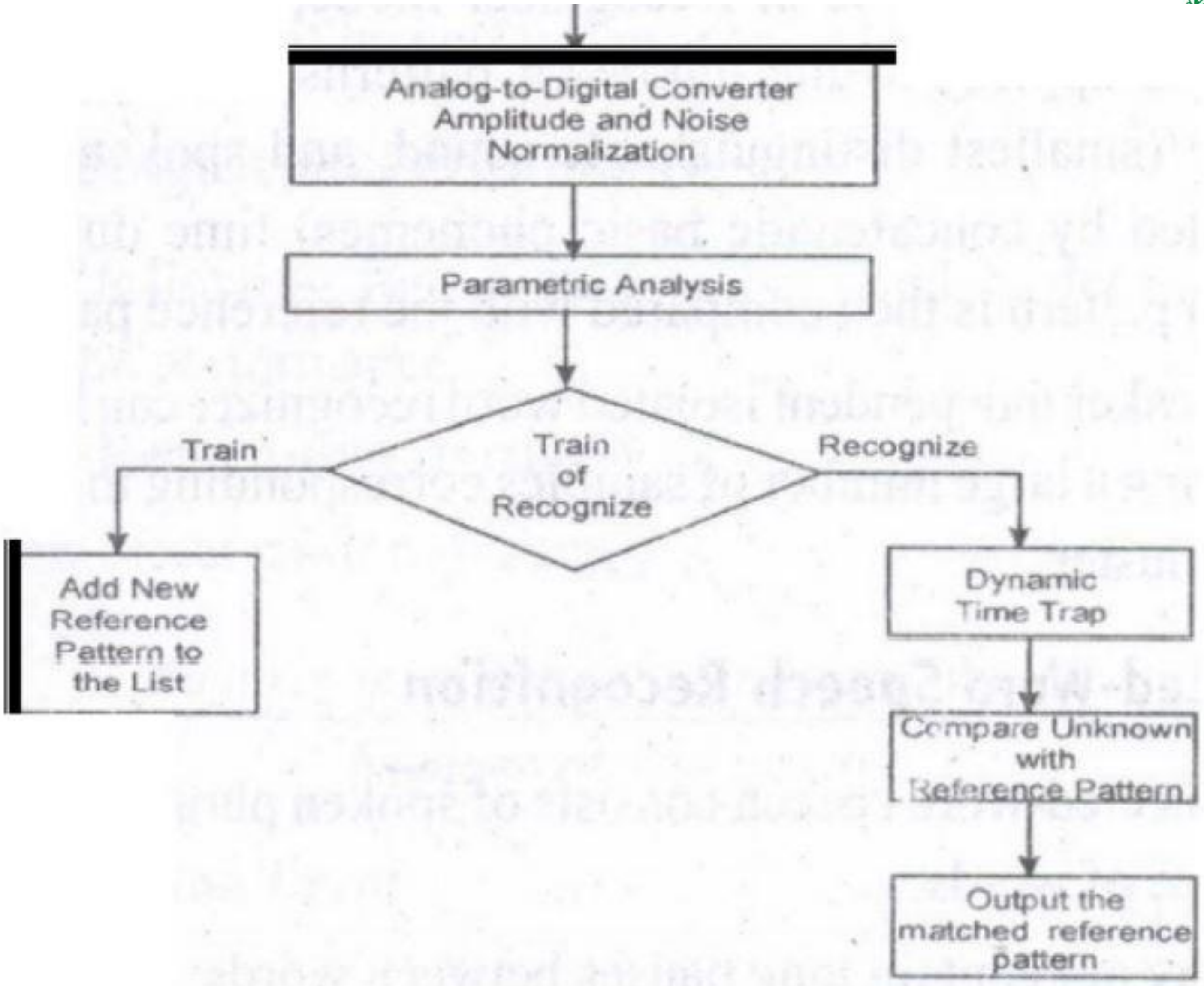
- Sound is made up of continuous analog sine waves that tend to repeat depending on the music or voice.
- The analog waveforms are converted into digital format by analog-to-digital converter (ADC) using sampling process.
- **Sampling process :**Sampling is a process where the analog signal is sampled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


Analog Signal
Digital Signal
ADC
Input
Output

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 tmly 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 System

- Voice Recognition Systems can be classified into three types.
- 1.Isolated-word Speech Recognition.
- 2.Connected-word Speech Recognition.
- 3.Continuous Speech Recognition.

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 relevent 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 concatenatic 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.

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 normalization and parametric representation.
- (ii) Second section consisting of segmentation and labeling of the speech segment into a symbolic string based on a knowledge based 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

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

2. Substitution Error

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

3. No Response Error

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

4. Insertion Error

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

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**

- • 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.
- **Foreign language learning**

Musical Instrument Digital Interface (MIDI)

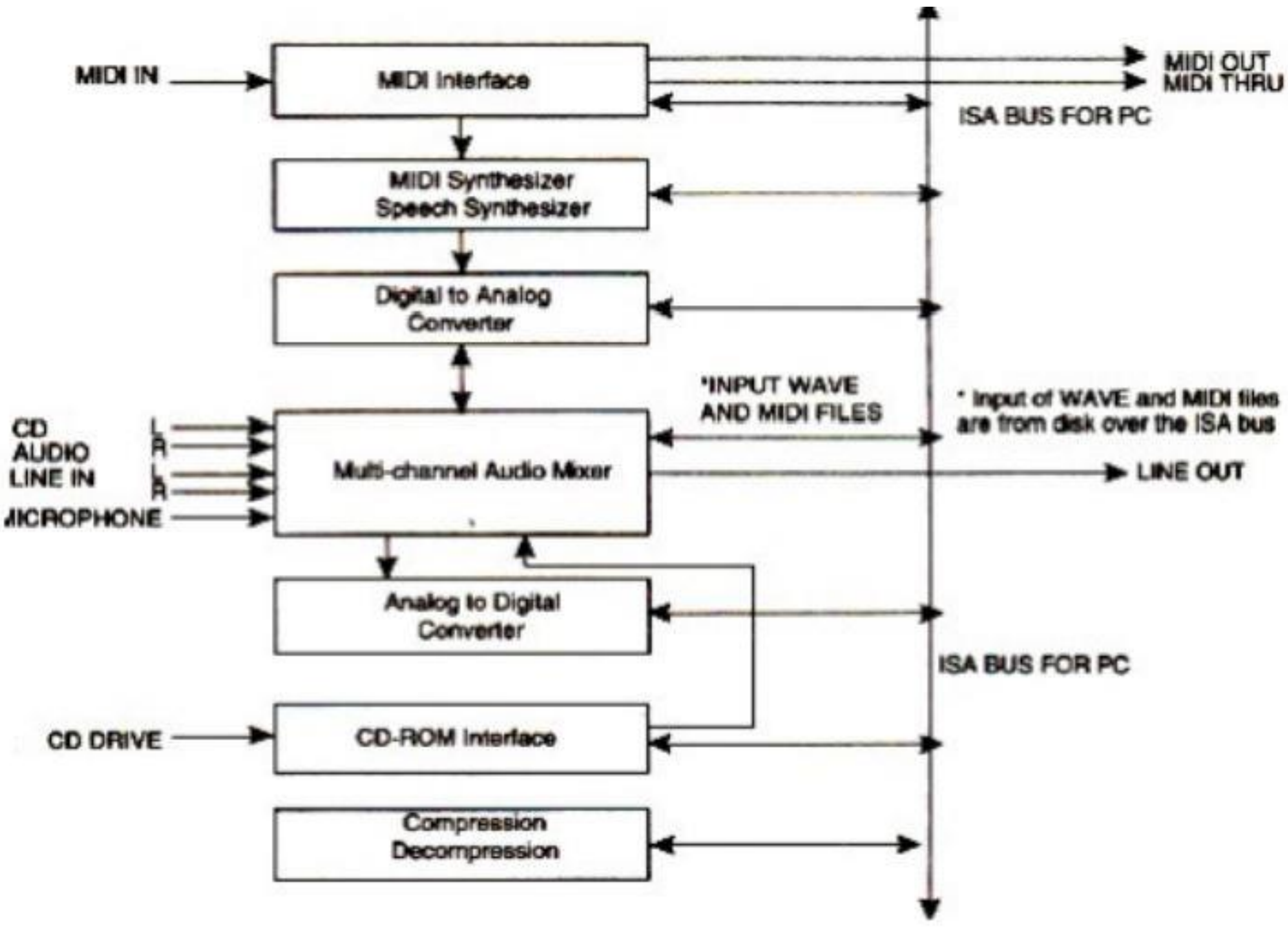
- MIDI interface is developed by Daver Smith of sequential circuits, inc in 1982. It is an universal synthesizer interface .
- **MIDI Specification 1.0**
- MIDI is a system specification consisting of both hardware and software components which define inter-connectivity and a communication protocol for electronic synthesizers, sequences, rythm machines, personal computers, and other electronic musical instruments.

- **MIDI Hardware Specification**
- The MIDI hardware specification require five pin panel mount requires five pin panel mount receptacle DIN connectors for MIDI IN, MIDI OUT and MIDI THRU signals.
- MIDI IN connector is for input signals The MIDI OUT is for output signals 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 OUT port sends MIDI messages to play these messages to an external synthesizer

Communication Protocol

- The MIDI communication protocol uses multibyte messages; There are two types of messages:
- (i) Channel messages
- (ii) System messages
- **The channel messages** have 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



SOUND BOARD ARCHITECTURE

- A sound card consist 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.
- **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 'foml of W AVE 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 u CD ROM drive to the sound board.

SIST

JUSTICE PEACE REVOLUTION

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY

DEEMED UNIVERSITY