Multimedia Systems

Chapter 1 : Introduction to Multimedia Systems

Q. 1 Explain Properties of Multimedia system.

Ans. :

Properties of Multimedia System

1. **Independency**

Multimedia system consists of different media such as sound, graphics, text, and video. These media should be independent. Multimedia system requires several levels of independence. Computer controlled video recorder stores audio and video information, but there is tight connection between audio and video. Both the media are coupled through the common storage medium of the tape.

2. **Computer Support Integration**

It is important to know the tolerance and limits for each medium as integration will require knowledge of these for synchronization and indeed create further limits.

It is common (obvious) that media types are bundled together for ease of delivery, storage etc. Therefore, it is not surprising that formats have been developed to support, store and deliver media in an integrated form. By using multimedia system one should able to control media processing. Multimedia system should be programmable by professional user or end user (non professional user). When integrating different media you should also consider synchronization between media. The need for interchange between different multimedia applications probably running on different platforms has lead to the evolution of common interchange file formats. Many of these formats build on underlying individual media formats (MPEG, JPEG etc.)

3. **Communication Systems**

A Variety of multimedia applications running on different platforms will need to communicate with each other particularly if they are running on a distributed network. The distributed environment enables interesting multimedia application. Here information cannot be created, processed, presented and store but distributed above the single computer boundary. Until recently the lack of a common interchange file format was a serious impediment to development of a market of multimedia applications. A common interchange format needs to be widely adopted (be supported by many applications) and be sufficiently expressive to represent a wide variety of media content.

4. **Interactivity**

In a multimedia system, if the user has the ability to control what elements are delivered and when, the system is called an interactive system. Traditional mass media include television, film, radio, and newspapers. These are called mass media, since the communication processes are one way, originating from a source and being delivered to a mass audience. These technologies also combine audio, video, graphics, and text, but in a way that is inflexible.

For example, a film has a predefined beginning, middle, and end, irrespective of the audience watching it. With the power of the computer, the same media could be manipulated by the audience. In this manner, the audience does not need to remain passive, but becomes the user of the system. Thus, the key difference between mass media and multimedia is the shift from audience to users and one-way communication to two-way communication. This is accomplished through interactivity. To communicate with the system, the user can use a variety of devices such as the keyboard, mouse, tracking ball, touch screen, and pen-based mouse. Thus while designing a multimedia application; we have to decide the level of interactivity we wish to provide to the user of the system.

Q. 2 Explain Global structure of multimedia system.

Ans. : Global Structure of Multimedia

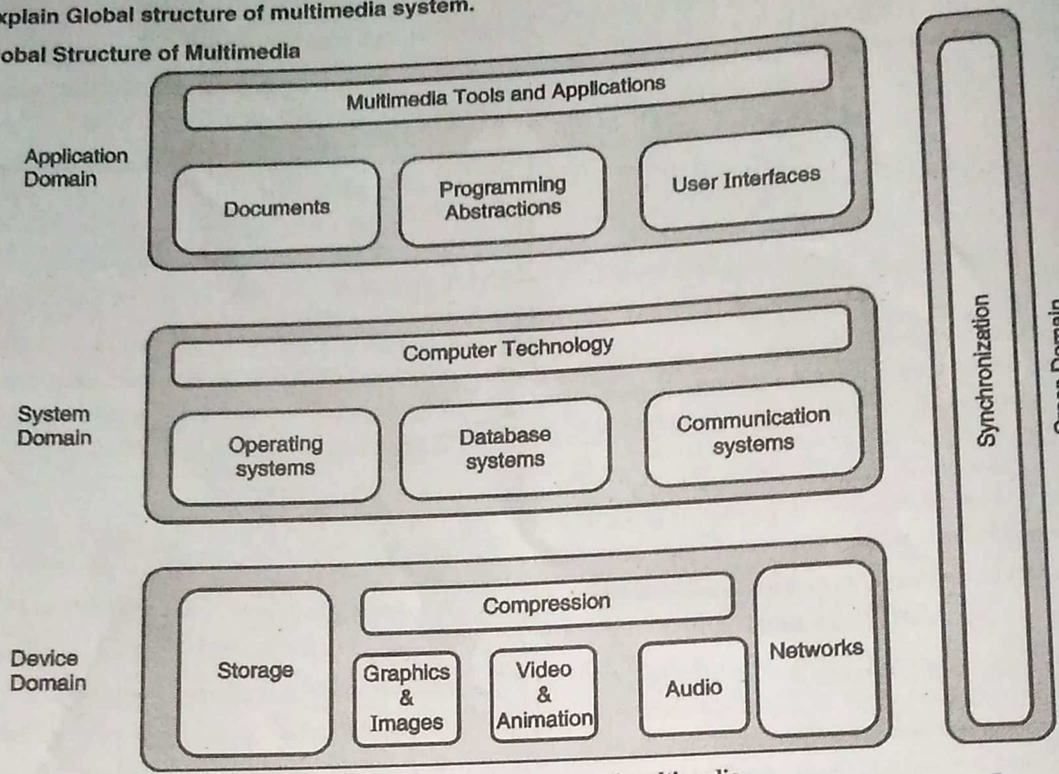


Fig. 1.1 : Global structure of multimedia

1. Device Domain

The device domain contains all multimedia elements including text, audio, video, image, animation, graphics etc. It also consists of the compression and storage of these elements. It also specifies how these elements are digitized and processed. An audio technique includes MIDI, speech generation, speech analysis and speech transmission. A video technique includes Chroma sub sampling, CCIR and HDTV. A multimedia networking allows for the data exchange of discrete and continuous media among computers. This communication requires proper services and protocol for data transmission. For compression of multimedia data various compression schemes are used such as Packbit encoding (RLE), CCITT group 3 and 4 compression, JPEG, JPEG 2000, MPEG 1,2,4,7 are used. Optical storage media offer a higher storage density at lower cost. Compact Disk - Digital Audio (CD-DA) allows the digital storage of stereo audio information at a high level of quality.

2. System Domain

Computer technology specifies the interface between device domain and the system domain. To utilize the device domain, three services exist. These services are mostly implemented in software. Multimedia database system includes text, audio, video and images. By using Multimedia database system, store, retrieve and manipulate these multimedia objects. The operating system serves as an interface between computer hardware and all other software components. It provides the user with a programming and computational environment, which should be easy to operate.

The operating system offers services related to the essential resources of computer : CPU, main memory, storage, and all input and output devices. The communication system is responsible for data transmission according to the timing and reliability requirements of the networked multimedia application. Multimedia communication requires proper services and protocol for data transmission.

3. Application Domain

A good user interface is defined as one of the perceived to the efficient and intuitive by most users. It is easy to learn and guide the user along by prompting action. The user interface should be responsive to user needs. Design a suitable user interface to support your data input and data validation requirements. A document consists of a set of structural information that can be in different forms of media, and during presentation can be generated or recorded. A document is aimed at the human perception of a human, and is accessible for computer representation. Many function of document handling and other application are accessible and presented to the user through a user interface.

Abstraction is the process of hiding the details and exposing only the essential features of a particular concept. Abstraction reduces the complexity of program generation and maintenance. Abstraction of multimedia data serves as the fundamental building block for programming different multimedia application, especially editor and other document processing tools. The services of the system domain are offered to application domain through proper programming abstraction.

4. Cross Domain

Synchronization in multimedia systems refers to the temporal relations between media objects in the multimedia system. Synchronization is addressed and supported by many system components including the operating system, communication system, database, documents and even often by application. Hence synchronization must be considered at several levels in multimedia system.

Q. 3 Explain multimedia workstation architecture in detail.

Ans. : Multimedia Workstation Architecture

Fig. 1.2 describes architecture of multimedia workstation environment. In this Fig. 1.2 left-hand side is very similar to non-multimedia system. To design and develop multimedia application, we need to extend the existing non multimedia hardware and software i.e. right side of Fig. 1.2 shows new entities required for supporting multimedia application.

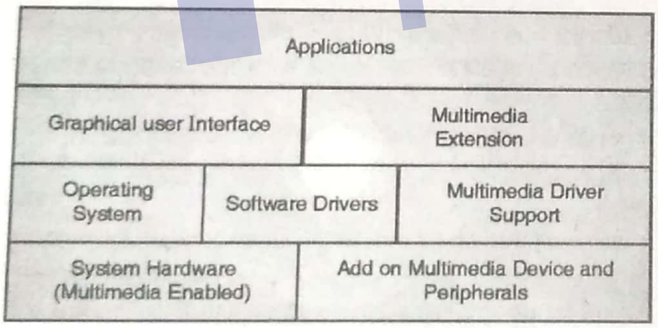


Fig. 1.2 : Multimedia workstation architecture

1. Application

In multimedia each object is compressed and then stored therefore while giving back it need to decompress. Application layer performs compression and decompression. Hence each architecture has application as the first layer. Multimedia applications such as educational, online training electronic messaging, multimedia repositories and video conferencing kind of applications are using at very first level of architecture.

2. General user Interface

On our computer system user access these applications which will provide graphical user interface for users those who will use the multimedia application. Multimedia application should interface with standard user interfaces such as Microsoft windows, X windows or with presentation manager.

3. System hardware

Multimedia enables means on board finicalities. i.e. system hardware present with the machine. Multimedia system should operate with or without special hardware required for multimedia such as DSP's with no change in application s/w and then how to get hardware interface with system those required for processing.

4. Multimedia extension

Extension is basically an extra finicality. Multimedia application like full-motion digital video requires extending the basic GUI provided by Windows. Example, due to web camera (add on) video - conferencing (extension) is possible.

5. Add on multimedia devices and peripherals

Printer, scanner, Microphone, Scanner, microphone, digital cameras each device require its own device controller and encoding hardware.

6. Software Driver and Multimedia Driver Support

Software drivers are required, so that the application can talk to devices. Use of software driver allows the user to interact with much wider range of peripherals and systems.

Q. 4 Which are the multimedia application? Explain in detail.

Ans. : Multimedia Applications

1. Document Imaging

Document Imaging is a technology that takes the hardcopy of a document and converts it into a digital format. Document imaging makes it possible to store, retrieve and manipulate very large volume of data such as documents. Documents may contain data, pictures, graphical representation of data which we are using for official use at various organizations, companies, firms, government offices. Document image systems use workflows that are customized for the purpose for which being used. The workflow defines sequence for scanning images, performing quality check, performing data entry based on contents of images, indexing them, and storing them on optical media.

Advantages of using Document imaging systems are :

(1) Extremely high speed of sharing

(2) Ease of availability

(3) Longer life of document.

2. Image Processing and Image Recognition

Digital image processing can be used to preprocessing (enhancement) of digital images, extracts the characteristics features of an image (region, boundary and skeleton). Image recognition is used in face recognition, fingerprint identification systems etc. **Digital image processing** is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modeled in the form of Multidimensional Systems.

3. Image Enhancement

Image enhancement will highlight some details of images by increasing sensitivity and contrast makes the picture darker by making borderline pixels black and increasing grey-scale level of pixels. Image enhancement capabilities might include following :

(a) **Image calibration :** Image pixels adjust at predefined level.

(b) **Real-time alignment :** Image is rotated by small angles in real time.

(c) **Grey scale normalization :** Overall grey-level of picture is evaluated.

(d) **RGB hue intensity adjustment :** Hue intensity we can adjust to predefined level to get the image with proper colors.

(e) **Frame averaging :** Intensity level of frame is averaged to overcome the effects of very dark and very light areas by adjusting the middle tones.

4. Image Animation

Image animation technology was developed by Walt Disney and brought into every home in the form of cartoon. Animation is the rapid display of sequence of images in order to create an illusion of motion. Images can be of two type : computer generated or captured (scanned). Computer created digital images are displayed sequentially at controlled speed. It is used successfully in designing moving parts such as automobile engines.

5. Image Annotation

Image annotation is the process by which a computer system automatically assigns meta-data in the form of captioning or keywords to a digital image.

There are two ways to do this :

1. A text file stored along with image.
2. Small image is stored with the original image.

6. Optical Character Recognition

It is the translation of image of hand written, typewritten or printed text into machine-editable digital text. OCR technology is used in bar code readers. OCR actually converts document image in editable word processing format. First character is scanned and predefined grid is applied and image is converted in digital form.

7. Digitizer

A graphics tablet or a digitizing tablet is computer input device that allows one to hand-draw images and graphics, just like we do with a pen on paper. Digitizer consists of a flat surface upon which we can draw images with an attached stylus.

Digitizer are of two types :

1. Active digitizer (Battery powered) : No need to touch.
2. Passive digitizer : Need to touch.

8. Full Motion Digital Video

Full motion video is used in gaming industry, training as well as in business world. It is also used in video conferencing, internet TV etc. There are two main issues that need to be addressed :

(a) **Time delay of the packet :** Network delay from source to destination must be < some Max. Delay (say t_1 seconds).

(b) **Buffer under run :** frame F_{n-1} must be in the playout buffer before frame F_n is rendered.

9. Multimedia E-mailing

Ordinary email is not multimedia application, but any attachment makes the e-mail message a hypermedia document. A hypermedia document is a multimedia document flowing over the internet. Multimedia E-mailing system requires a sophisticated infrastructure consisting of the following to support it :

(a) Electronic messaging system provides message store and forward facility.

(b) Message transfer agents to route messages to their final destinations.

(c) Message servers where user can store the message.

A document which contains image, audio, video is called multimedia document. This page conveys various messages regarding current news, job information, email, we can download music, photos, audio and various other documents. Electronic messaging means document content various kind of information as well as user send this document as a mail to other end.

Q. 5 Explain evolving technologies for multimedia systems.

Ans. :

Evolving Technologies for Multimedia Systems

Multimedia applications use a number of technologies generated for both commercial business applications as well as video game industry. Access of multimedia application on windows and other operating system cause rapid increase in use of these applications for various day-to-day uses. Because of that there interface is required between application software such as our multimedia system and operating system which users are using on machine.

1. Hypertext

Hypertext is an innovation to the paradigms of computing user interfaces that attempts to overcome the limitations of written text. Hypertext, instead of remaining static like traditional text, will dynamically "branch or perform on request". Thus hypertext makes possible the organization of material in ways that partially overcome the linearity inherent in written text. The prefix hyper - (Modern Greek term for over or beyond) signifies the overcoming of such constraints. The most frequently discussed form of hypertext document contains automated cross-references to other documents called hyperlinks. Selecting a hyperlink causes the computer to load and display the linked document. Documents referenced by hypertext can themselves be static (prepared and stored in advance) or dynamically generated (in response to user input). Therefore a well-constructed system using hypertext can encompass, incorporate or supersede other conventions of user-interface paradigms, such as menus and command lines, and can be used to access both static collections of cross-referenced documents and interactive applications.

The documents and applications can be local or can come from anywhere with the assistance of a computer network like the Internet. The most famous implementation of hypertext is the World Wide Web. Hypertext implements organization of non-sequential data by natural association of information rather than hierarchical filing structures as paper-based text documents.

2. Hypermedia document

Hypermedia is used as a logical extension of the term hypertext in which graphics, audio, video, plain text and hyperlinks intertwine to create a generally non-linear medium of information. This contrasts with the broader term multimedia, which may be used to describe non-interactive linear presentations as well as hypermedia. Hypermedia is a term created by Ted Nelson, and used in his 1965 article Complex information processing : a file structure for the complex, the changing and the indeterminate. It is used as a logical extension of the term hypertext, in which graphics, audio, video, plain text and hyperlinks intertwine to create a generally non-linear medium of information. This contrasts with the broader term multimedia, which may be used to describe non-interactive linear presentations as well as hypermedia. Hypermedia should not be confused with hypergraphics or super-writing which is not a related subject.

3. Hyper speech

A mechanism to connect telephony voice applications that are deployed at different sites. An underlying protocol, HSTP that provides synchronization of the telephony call with the application logic at the time of call transfer from one site to another. In the design of HSTP, we ensured that no change is required to existing communication standards, both in the IP world and in the circuit-switched PSTN network. Hyper speech is also supported without enforcing any changes to current standard voice programming languages such as Voice XML and SALT. Ability to provide a Hyper speech link to other voice applications can enable several cross-organizational applications such as a travel reservation and music purchase, where the payment is made by a secure voice application that is hosted by the bank.

Hyper speech can also support browsing of these voice applications across organizations. Increased connectivity across the different voice applications is likely to lead to a web of voice applications.

4. HDTV and UDTV

Requires substantial processing power. High performance microprocessors and DSPs and HDTV and UDTV (High Definition Television and Ultra Definition Television). Parallel development in the electronics industry is to raise the resolution levels of commercial television broadcasting. Better known TV broadcasting standards are, NTSC (National Television Standard Committee) PAL (Phase Alternate Line) SECAM (System Couleur Avec Memoire) NHK (Nippon Hoso Kyokai). Developments for HDTV and UDTV benefit the computer industry as the basic technologies for display and communications merge.

5. Fuzzy logic

Fuzzy logic is a form of many-valued logic derived from fuzzy set theory to deal with reasoning that is fluid or approximate rather than fixed and exact. In contrast with "crisp logic", where binary sets have two-valued logic, fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Put more simply, fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. FL requires some numerical parameters in order to operate such as what is considered significant error and significant rate-of-change-of-error, but exact values of these numbers are usually not critical unless very responsive performance is required in which case empirical tuning would determine them. Generally, FL is so forgiving that the system will probably work the first time without any tweaking.

Q. 6 Write short note on : Scanner.

Ans. : Scanner

A **scanner** is an acquisition peripheral for scanning documents, i.e. converting a paper document to a digital image. Scanner is one of the input device that basically, using for multimedia applications.

Document imaging is one of the multimedia application in that scanner is important input device to scan or to capture the images. Scanner acts as the camera eye and takes a photograph of document, creating an unaltered electronic pixel representation (of image) of the original.

Characteristics of a scanner

A scanner is generally characterized by the following elements :

1. **Resolution :** Expressed in dots per inch (referred to as dpi), the resolution defines the fineness of the scan. The order of magnitude of the resolution is around 1200 per 2400 dpi. The horizontal resolution is very much dependent on the quality and number of captors, whereas vertical resolution is closely linked to the accuracy of the drive motor.

   However it is important to distinguish the optical resolution, which is the actual resolution of the scanner, from the **interpolated resolution**. Interpolation is a technique involving defining intermediate pixels from among actual pixels, by calculating the mean of the colours of neighbouring pixels. This technology helps achieve good results but the interpolated resolution thus defined is in no way a criterion that can be used to compare scanners.

2. **The format of the document :** Depending on their size, scanners are able to accommodate documents of different sizes, generally A4 (21 × 29.7 cm), or more rarely A3 (29.7 × 42 cm).

3. **Acquisition speed :** Expressed in Pages Per Minute (ppm), the acquisition speed represents the scanner's ability to pick up a large number of pages per minute. The acquisition speed depends on the document format and the resolution chosen for the scan.

4. **Physical characteristics :** Other elements may be taken into account when choosing a scanner :
   (a) Size, in terms of the physical dimensions of the scanner.
   (b) Weight.
   (c) Electricity consumption, expressed in Watts (W).
   (d) Operating and storage temperatures.

5. **Noise level :** Scanners can be very noisy, and this may cause considerable disturbance.

5. **Accessories :** The drivers and user manual are usually provided, but you must check that connection cables are also provided; if not they must be purchased separately.

How a scanner works ?

The operating principle for a scanner is as follows :

1. The scanner moves over the document line by line.
2. Each line is broken down into "basic dots" which correspond to pixels.
3. A captor analyses the colour of each pixel.
4. The colour of each pixel is broken down into 3 components (red, green, blue).
5. Each colour component is measured and represented by a value. For 8-bit quantification, each component will have a value between 0 and 225 inclusive.
6. The rest of this article will specifically describe the operation of a flat scanner, although the operating mode for a hand scanner or sheet-fed scanner is exactly the same. The only difference is in the feeding of the document.
7. A flat scanner has a motor-driven lighted slot which scans the document line by line under a transparent glass panel on which the document is placed, with the scanning side face down.
8. The high-intensity light emitted is reflected by the document and converges towards a series of captors via a system of lenses and mirrors. The captors convert the light intensities received into electrical signals, which are in turn converted into digital data by an analogue-digital converter.

Types of scanners

Scanners come in wide variety of models with range of sizes, functions, speeds and resolutions.

1. Flatbed scanner
2. Rotary drum scanner
3. Handheld scanner

Q. 7 Explain Multimedia Database Management in detail.

Ans. : Multimedia Database Management

Multimedia database is a kind of database like any other database containing multimedia collections. Multimedia is defined as the combination of more than one media, they may be of two types-static and dynamic media. Text, graphics and images are categorized as static media; on the other hand, objects like animation, music, audio, speech, video are categorized as dynamic media. Graphic images may consist of cliparts, photographs, logos and custom drawings. Sound consists of voice narration, speech, music etc.

Video data encompasses sound as well as photos. To manage these data multimedia database management system is essential. Multimedia database management system can be defined as a software system that manages a collection of multimedia system can be defined as a software system that manages a collection of multimedia data and provides access to users to query and retrieve multimedia objects.

Generally, multimedia database contains text, image, animation, video audio, movie sound etc. But, all data are stored in the database in binary form.

Multimedia Database Management System

The main task of multimedia database management system is to abstract from the details of storage access and storage management. Multimedia Database Management System (MDBMS) is embedded between the application domain and the device domain. Location of MDBMS in multimedia system is shown in Fig. 1.3.

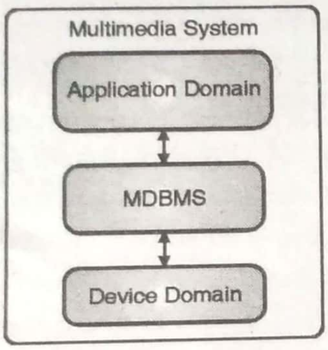


Fig. 1.3 : Location of MDBMS

MDBMS is part of system domain. It is integrated into system domain through the operating system and communication component. Integration of MDBMS into system is shown in Fig. 1.4.

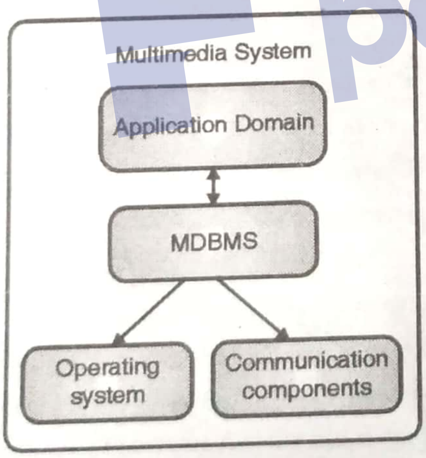


Fig. 1.4 : Integration of MDBMS into system

The design of a multimedia database management system is unlikely to follow in the footsteps of the design of a traditional database management system due to following characteristics of multimedia objects :

Multimedia objects are complex and therefore less completely captured in a MDBMS. Multimedia objects are audiovisual in nature. Multimedia objects are context dependence. Queries looking for multimedia objects are fuzzy in nature.

Why Multimedia Database ?

Multimedia database is capable of handling huge volume of multimedia objects which a general database fails to do effectively. Multimedia Database will help to create virtual museum. It will surely help to develop multimedia applications in various fields like teaching, medical sciences. Preserving decaying photographs, maps, films having got historical evidence or national importance.

Properties of MDBMS

Following are the properties of MDBMS :

1. **Persistence of data :** Data outlive processing programs and technologies, e.g. companies have to keep data in database for several decades.
2. **Consistent view of data :** Synchronization protocols provide a consistent view of data in a multi-user system.
3. **Security of data :** Transaction concepts ensure security and integrity protection in case of system failure. Recovery of lost data.
4. **Query and retrieval of data :** Query languages such as SQL (Structured Query Language) enable formulating database queries.

Characteristics of MDBMS

A MDBMS (Multimedia Database Management System) can be characterized based on its objective at the time of handling multimedia objects :

1. **Corresponding Storages Media :** Multimedia data should be stored according to characteristics of storage media.
2. **Descriptive Search Methods :** user should have content-based access to multimedia information.
3. **Device-independent Interface :** The interface to multimedia database should be device dependent.
4. **Format independent Interface :** Database queries should be independent of media format. MMDBMS should provide information in formats requested by the applications (GIF, TIFF, SUN Raster,...).
5. **View Specific and Simultaneous data Access :** Object sharing is the capability for different documents to share parts of their contents. Such a capability is especially necessary for multimedia documents as the amounts of storage space required to store a document might be quite large.
6. **Management of Large Amounts of Data :** MMDBMS must be capable of handling and managing large amounts of data
7. **Relational Consistency of Data Management :** This requirement ensures multimedia database consistency through rules, which impose some form of execution order on concurrent transactions.

8. **Real time data Transfer :** DBMS must perform read and write operations of continuous data in real time. Continuous data transfer should have higher priority than other action.

9. **Long Transaction :** The transfer of large amount of data will take a long time and must be done in a reliable fashion.

Multimedia Database Content

Multimedia database generally holds the following multimedia components like - text graphic, animation, sounds, video, image, speech etc.

Chapter 2 : Text and Digital Image

Q. 1 Explain Resource Interchange File Format.

Ans. :

RIFF (Resource Interchange File Format)

The Resource Interchange File Format (RIFF) is a generic file container format for storing data in tagged chunks. It was introduced in 1991 by Microsoft and IBM. The Resource Interchange File Format (RIFF), a tagged file structure is a general specification upon which many file formats can be defined. The main advantage of RIFF is extensibility, file format based on RIFF can be future proofed, as format changes can be ignored by existing application. The Resource Interchange File Format (RIFF), a tagged file structure, is a general specification upon which many file formats can be defined. The main advantage of RIFF is its extensibility; file formats based on RIFF can be future-proofed, as format changes can be ignored by existing applications.

The RIFF file format is suitable for the following multimedia tasks :

1. Playing back multimedia data.
2. Recording multimedia data.
3. Exchanging multimedia data between applications and across platforms.
4. RIFF files consist entirely of "chunks".

All chunks have the following format

1. 4 bytes : An ASCII identifier for this chunk, e.g. "fmt" or "data".
2. 4 bytes : An unsigned, little-endian 32-bit integer with the length of this chunk (except this field itself and the chunk identifier).
3. Variable-sized field : The chunk data itself, of the size given in the previous field.
4. A pad byte, if the chunk's length is not even.

RIFF specification defines following kinds of chunks :

1. **RIFF chunk :** Defines contents of RIFF file. It is the first chunk in RIFF file. In RIFF chunks first four bytes is ID, next four byte is data and next 4 bytes of RIFF chunk are allocated for the form type which contains four characters to identify the format of data stored. Various form types are WAVE, AVI, RMID, RDIB.

2. **List chunk :** A LIST chunk contains a list, or ordered sequence, of subchunks. A LIST chunk is defined as follows : LIST(<list-type> [<chunk>]...). The **<list-type>** is a four-character code that identifies the contents of the list. If an application recognizes the list type, it should know how to interpret the sequence of subchunks. Microsoft has only documented one list type called INFO. Each INFO list chunk consist of four character ASCII string ID called tag and four byte field containing size of chunk data, followed by data for that chunk. An 'INFO' list should contain only the following chunks. IARL (Archival Location) : Indicates where the subject of the file is archived. IART Artist. Lists the artist of the original subject of the file. For example, "Michaelangelo." ICMS (Commissioned). Lists the name of the person or organization that commissioned the subject of the file. For example, "Pope Julian II." ICMT (Comments) Provides general comments about the file or the subject of the file. If the comment is several sentences long, end each sentence with a period. Do not include newline characters. ICOP (Copyright). Records the copyright information for the file.

3. **Sub chunk :** Allows adding more information to primary chunk when primary chunk is not sufficient. It contains a four-character ASCII string i.e. tag ID to identify the type of data , four bytes of size containing the count of data value and data itself.

Fig. 2.1 shows organization of RIFF chunk, list and Sub chunk.

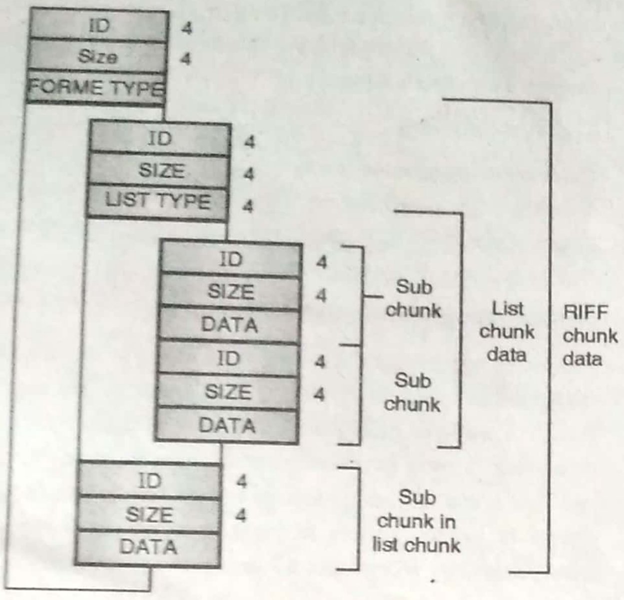


Fig. 2.1 : RIFF chunk with list chunk

Q. 2 Explain different types of images.

Ans. :

Types of Images

Images can be classified as follows :

(1) Monochrome Image

In this, each pixel is stored as a single bit (0 or 1). Here, 0 represents black while 1 represents white. It is a black and white image in the strictest sense. These images are also called bit mapped images. In such images, only black and white pixels and no other shades of grey.

(2) Grey Scale Image

Here each pixel is usually stored as a byte (8-bits). Due to this, each pixel can have values ranging from 0 (black) to 255(white). Grey scale images, as the name suggests have black, white and various shades of grey present in the image.

(3) Colour Image (24-bit)

Colour images are based on the fact that a variety of colours can be generated by mixing the three primary colours viz, Red, Green and Blue, in proper proportions. In colour images, each pixel is composed of RGB values and each of these colours require 8-bits (one byte) for its representation. Hence each pixel is represented by 24-bits [R(8-bits), G(8-bits), B(8-bits)].

A 24-bit colour image supports 16, 777, 216 different combination of colours.

Colour images can be easily converted to grey scale images using the equation

$$X = 0.30\,R + 0.59\,G + 0.11\,B \quad ...(1)$$

An easier formula that could achieve similar results is

$$X = \frac{R+G+B}{3} \quad ...(2)$$

(4) Half toning

Half-toning gives excellent results and perceive a grey level image just by using black pixels on a white background.

The logic to implement a half-toned image from a grey level image is given below.

(1) Define the size of the half-toned matrices based on the number of grey levels the original image has.

(2) Generate the matrices starting from all white pixels to all black pixels.

(3) Read the original image. For every grey level value read, plot the corresponding matrix.

Remember, the physical size of the half-toned image will always be bigger than the original image as for every single grey level value, we output an entire matrix.

Q. 3 Enlist and explain application of Images.

Ans. : Application of Images

The field of digital image processing has experienced continuous and significant expansion in recent years. The usefulness of this technology is apparent in many different disciplines covering medicine through remote sensing.

Medical applications	Restorations and enhancements
Digital cinema	Image transmission and coding
Color processing	Remote sensing
Robot vision	Hybrid techniques
Facsimile	Pattern recognition
Registration techniques	Multidimensional image processing
Image processing architectures and workstations	Video processing
Image sharpening and restoration	Medical field
Remote sensing	Transmission and encoding
Machine/Robot vision	Color processing

Pattern recognition	Video processing
Microscopic imaging	Programmable dsps for video coding
High-resolution display	High-quality color representation
Super-high-definition image processing	Impact of standardization on image processing.

Image sharpening and restoration

Image sharpening and restoration refers here to process images that have been captured from the modern camera to make them a better image or to manipulate those images in way to achieve desired result. It refers to do what Photoshop usually does. This includes Zooming, blurring, sharpening, gray scale to color conversion, detecting edges and vice versa, Image retrieval and Image recognition.

Medical field

The common applications of DIP in the field of medical is

1. Gamma ray imaging
2. PET scan
3. X Ray Imaging
4. Medical CT
5. UV imaging

UV imaging

In the field of remote sensing, the area of the earth is scanned by a satellite or from a very high ground and then it is analyzed to obtain information about it. One particular application of digital image processing in the field of remote sensing is to detect infrastructure damages caused by an earthquake.

As it takes longer time to grasp damage, even if serious damages are focused on. Since the area effected by the earthquake is sometimes so wide, that it not possible to examine it with human eye in order to estimate damages. Even if it is, then it is very hectic and time consuming procedure.

1. Transmission and encoding

This field does not only focus on transmission, but also on encoding. Many different formats have been developed for high or low bandwith to encode photos and then stream it over the internet or etc.

2. Machine/Robot vision

Apart from the many challenges that a robot face today, one of the biggest challenge still is to increase the vision of the robot. Make robot able to see things, identify them, and identify the hurdles etc. Much work has been contributed by this field and a complete other field of computer vision has been introduced to work on it.

3. Line follower robot

Most of the robots today work by following the line and thus are called line follower robots. This helps a robot to move on its path and perform some tasks. This has also been achieved through image processing.

4. Color processing

Color processing includes processing of colored images and different color spaces that are used. For example, RGB color model, YCbCr, HSV. It also involves studying transmission, storage and encoding of these color images.

5. Pattern recognition

Pattern recognition involves study from image processing and from various other fields that includes machine learning (a branch of artificial intelligence). In pattern recognition, image processing is used for identifying the objects in an images and then machine learning is used to train the system for the change in pattern. Pattern recognition is used in computer aided diagnosis, recognition of handwriting, recognition of images etc.

6. Video processing

A video is nothing but just the very fast movement of pictures. The quality of the video depends on the number of frames/pictures per minute and the quality of each frame being used. Video processing involves noise reduction, detail enhancement, motion detection, frame rate conversion, aspect ratio conversion, color space conversion etc.

Q. 4 Write short note on Data Compression

Ans. : Data Compression

Multimedia object consists of color image, photographic or video image, audio data and full motion video. These data object needs to be stored, retrieved, transmitted and displayed. This large data objects present two problems viz. storage and transmission.

Compression in computer terms means reducing the physical size of data such that it occupies less storage space and memory. Compressed files are, therefore, easier to transfer because there is a sizable amount of reduction in the size of data to be transferred. This results in a reduction in the time needed for file transfer as well as a reduction in the bandwidth utilization thus providing good sound quality even over a slow network.

Thus, the process of reducing the amount of data required to represent information is called compression.

Image files commonly contain considerable amount of data that is redundant and much that is irrelevant. Due to this, images are prime candidates for compression. Data compression techniques exploit inherent redundancies and irrelevancies by transforming a data file into a smaller file from which the original file can later be reconstructed exactly or approximately. There exists two different categories of data compression algorithms. They could either be Loss-less or they could be Lossy.

The loss-less algorithm eliminates only redundant information, so that one can recover the image exactly upon decompression of the file. In other words the restored image is identical to the original image.

f(x, y) —— compression —— decompression —— g(x, y)

For loss-less algorithms, $f(x, y) = g(x, y)$

Lossy algorithms eliminate redundant as well as irrelevant information and thus permit only an approximate reconstruction of the original image rather than an exact duplicate. Information is lost when lossy algorithms are used.

Loss-less algorithms are absolutely necessary for many types of data, for example : Executable codes (EXE's), word processing files, tabulated numbers etc.

f(x, y) —— compression —— decompression —— g(x, y)

For Lossy compression $f(x, y) \neq g(x, y)$. $g(x, y)$ is smaller compared to $f(x, y)$.

Compression Techniques

Compression is process of reducing the amount of data, required to represent given quantity of information. In most images neighboring pixels are correlated and there exists redundant information. Redundancy can be broadly classified into Statistical redundancy and Psycho visual redundancy.

Statistical redundancy can be classified into inter-pixel redundancy and coding redundancy. Inter-pixel can be further classified into spatial redundancy and temporal redundancy. Spatial redundancy or correlation between neighboring pixel values. Spectral redundancy or correlation between different color planes or spectral bands. Temporal redundancy or correlation between adjacent frames in a sequence of images in video applications.

In digital image compression, three basic data redundancies can be identified and exploited: Coding redundancy, Inter-pixel redundancy and Psychovisual redundancy.

Coding Redundancy : Some pixel values more common than others.

Inter-pixel Redundancy : Neighboring pixels have similar values.

Psychovisual Redundancy : Some color differences are imperceptible

Types of Data Compression

There are two types of compression :

(A) Lossless Compression (B) Lossy Compression

Lossless Compression

Lossless data compression is a class of data compression algorithms that allows the exact original data to be reconstructed from the compressed data. i.e. In lossless data compression, the compressed-then-decompressed data is an exact replication of the original data. Lossless data compression is used in many applications. For example, the zip file format produced by the Winzip program on a PC produces exact copies of the original material that was encoded within the zip file. It is also used in the Unix tool gzip. Lossless compression is used in cases where it is important that the original and the decompressed data be identical, or where deviations from the original data could be deleterious. Typical examples are executable programs, text documents and source code. Some image file formats, like PNG or GIF, use only lossless compression,

The following are some of the commonly used lossless standards :

1. Packbits encoding (run-length encoding)
2. CCITT Group 3 1-D (compression standard based on run-length encoding scheme)
3. CCITT Group 3 2-D (compression standard based on run-length encoding scheme modified by two-dimensional encoding)
4. CCITT Group 4 (compression standards based on two-dimensional compression)
5. Lempel - Ziv and Welch algorithm LZW (Techniques used by ARJ/PKZIP)

Lossy Compression

Lossy data compression, which only allows an approximation of the original data to be reconstructed, in exchange for better compression rates i.e. In lossy data compression, the decompressed data may be different from the original data. Typically, there is some distortion between the original and reproduced signal. **Lossy compression** is a data encoding method which discards (loses) some of the data, in order to achieve its goal, with the result that decompressing the data yields content that is different from the original, though similar enough to be useful in some way. Lossy compression is most commonly used to compress multimedia data (audio, video, still images), especially in applications such as streaming media and internet telephony. This method is used where absolute data accuracy is not essential. Lossy compression is the most commonly used compression type. This compression technique is used for image documents, audio, and video objects.

The following are some of the commonly used lossy standards :

1. Joint Photographic Experts Group (JPEG)
2. Motion Picture Experts Group (MPEG)
3. Adaptive Differential Pulse Code Modulation (ADPCM)
4. CCITT H.261 (Px64) Video Coding Algorithm
5. Intel DVI (Digital Video Interactive)

Q. 5 Compare Lossy and Lossless compression.

Ans. : Lossless Compression Vs. Lossy Compression

Sr. No.	Lossless compression	Lossy compression
1.	Lossless compression schemes are reversible, so	Lossy compression is irreversible. Original data

Sr. No.	Lossless compression	Lossy compression
	that the original data can be reconstructed.	cannot be reconstructed.
2.	Lossless compression schemes usually exploit statistical redundancy.	Lossy compression makes use of the fact, how humans perceive the data.
3.	Less compression ratio.	More compression ratio.
4.	Used for compression text and images.	Used for compressing audio, video and images.
5.	LZ77, LZW, PNG, CCITT Group 3 – 1 D CCITT Group 3 – 2 D CCITT Group 4 – 2 D	MPEG 1, MPEG 2, MPEG 4 Fractal compression H.261 video encoding GIF, JPEG Intel DVI

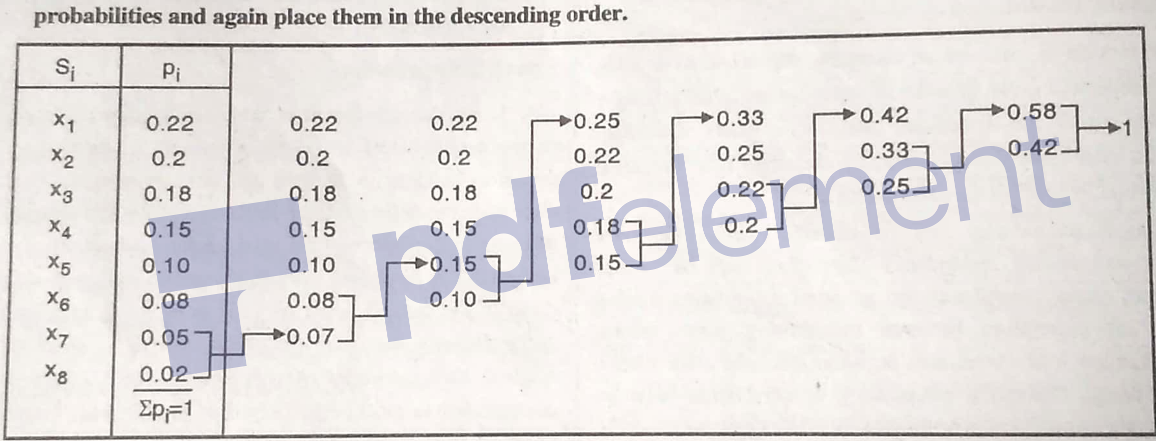
Q. 6 **Apply Huffman coding procedure to following message ensemble and determine average length of encoded message and coding efficiency (η). The symbols (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) are emitted with probabilities of (0.22, 0.2, 0.18, 0.15, 0.10, 0.08, 0.05, 0.02)**

Ans. :

Step 1 : The given information is as follows :

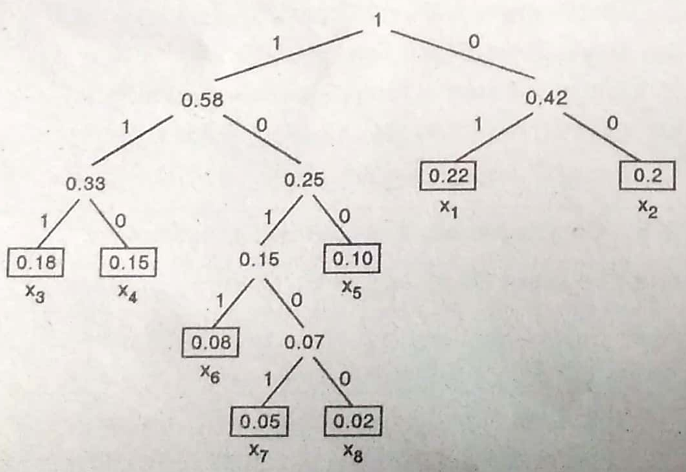
S_i Symbols	p_i Probability
x_1	0.22
x_2	0.2
x_3	0.18
x_4	0.15
x_5	0.10
x_6	0.08
x_7	0.05
x_8	0.02

Step 2 : **The given probabilities are already in the descending order. Ccreate the Huffman table. In this add the smallest 2 probabilities and again place them in the descending order.**



Step 3 :

From the Huffman table, we create the Huffman tree. We start from 1 and move backwards.



Each of the end nodes (boxes) in the original probability of the data.

Step 4 :

Labeling each branch. Let the left branches be 1 and the right branches be 0. (It could be the other way too). We talk along the branches to obtain the codes.

∴ Hence the codes obtained are

Symbol	Huffman code	No. of bits
x_1 →	01	2
x_2 →	00	2
x_3 →	111	3

Symbol	Huffman code	No. of bits
x_4 →	110	3
x_5 →	100	3
x_6 →	1011	4
x_7 →	10101	5
x_8 →	10100	5

Average length code = $\sum$ (Number of bits for each symbol) × (Its probability)

∴ Average length code = 2(0.22) + 2(0.2) + 3(0.18) + 3(0.15) + 3(0.10) + 4(0.08) + 5(0.05) + 5(0.02)

Average length code = 2.8 bits/symbol

Had we used the natural code, we would require 3 bits for each symbol since there are 8 symbols ($2^3 = 8$). Hence the average length code would have been 3 bits/symbol.

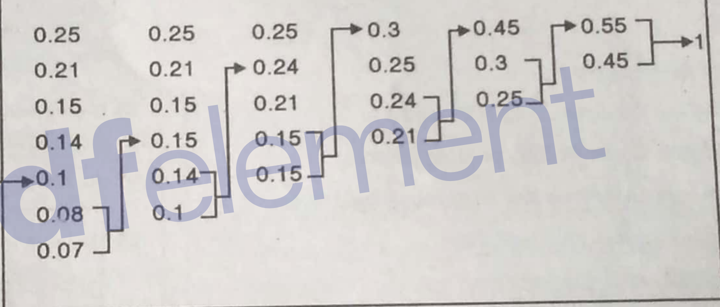
Q. 7 Given below is the table of 8 symbols and their frequency of occurrences. Give Huffman code for each symbol.

Symbol	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
Frequency	0.25	0.15	0.06	0.08	0.21	0.14	0.07	0.04

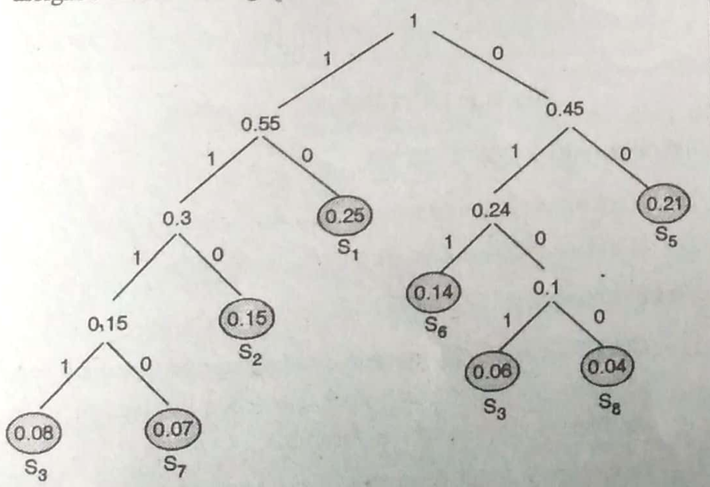
Ans. :

The Huffman table.

Symbol	Frequency	Descending order
S_1	0.25	0.25
S_2	0.15	0.21
S_3	0.06	0.15
S_4	0.08	0.14
S_5	0.21	0.08
S_6	0.14	0.07
S_7	0.07	0.06
S_8	0.04	0.04



Create the Huffman tree. Let the left side branches be assigned value 1 and right side branches be assigned value 0.



The final Huffman code is generated by traversing over the branches.

Symbol	Huffman code
S_1	10
S_5	00
S_2	110
S_6	011
S_4	1111
S_7	1110
S_3	0101
S_8	0100

Q. 8 Explain Group 3 One-Dimensional (G31D) in detail.

Ans. : Group 3 One-Dimensional (G31D)

Group 3 One-Dimensional encoding (G31D) is a variation of the Huffman keyed compression scheme. A bi-level image is composed of a series of black-and-white 1-bit pixel runs of various lengths (1 = black and 0 = white). A Group 3 encoder determines the length of a pixel run in a scan line and outputs a variable-length binary code word representing the length and color of the run. Because the code word output is shorter than the input, pixel data compression is achieved.

The run-length code words are taken from a predefined table of values representing runs of black or white pixels. This table is part of the T.4 specification and is used to encode and decode all Group 3 data.

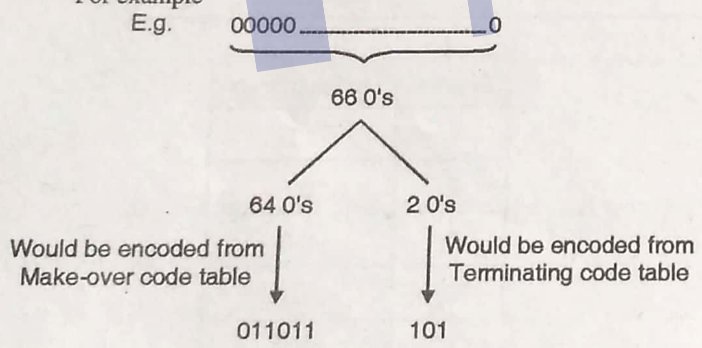
The size of the code words were originally determined by the CCITT, based statistically on the average frequency of black-and-white runs occurring in typical type and handwritten documents. Run lengths that occur more frequently are assigned smaller code words while run lengths that occur less frequently are assigned larger code words.

Run lengths are represented by two types of code words : makeup and terminating. An encoded pixel run is made up of zero or more makeup code words and a terminating code word. Terminating code words represent shorter runs, and makeup codes represent longer runs. There are separate terminating and makeup code words for both black and white runs.

Algorithm

1. Accept the input file.
2. Parse (scan) the input string of 0's and 1's.
3. Divide each run-length for encoding purpose.
4. Encode from make-over and terminating code tables.
5. The output is the compressed file.

For example



Resultant codeword : 011011 + 101 – 011011101

The above example shows how a run of 66 zeroes is broken up into a multiple of 64 zeroes plus two zeroes. The 64 0's being a multiple of 64 must be encoded from make-over code table. The remaining 2 0's must be encoded from terminating code tables.

Pixel runs with a length of 0 to 63 are encoded using a single terminating code. Runs of 64 to 2623 pixels are encoded by a single makeup code and a terminating code. Run lengths greater than 2623 pixels are encoded using one or more makeup codes and a terminating code. The run length is the sum of the length values represented by each code word.

Here are some examples of several different encoded runs :

a. A run of 20 black pixels would be represented by the terminating code for a black run length of 20. This reduces a 20-bit run to the size of an 11-bit code word, a compression ratio of nearly 2:1. This is illustrated in Fig. 2.2(a).

b. A white run of 100 pixels would be encoded using the makeup code for a white run length of 64 pixels followed by the terminating code for a white run length of 36 pixels (64 + 36 = 100). This encoding reduces 100 bits to 13 bits, or a compression ratio of over 7:1. This is illustrated in Fig. 2.2(b).

c. A run of 8800 black pixels would be encoded as three makeup codes of 2560 black pixels (7680 pixels), a makeup code of 1088 black pixels, followed by the terminating code for 32 black pixels (2560 + 2560 + 2560 + 1088 + 32 = 8800). In this case, we will have encoded 8800 run-length bits into five code words with a total length of 61 bits, for an approximate compression ratio of 144:1. This is illustrated in Fig. 2.2(c).

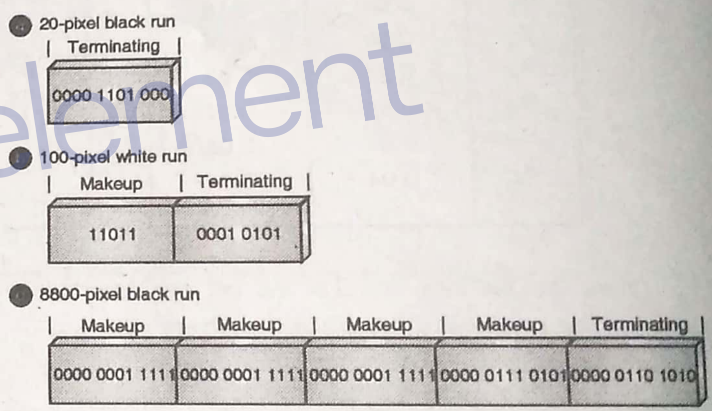


Fig. 2.2 : CCITT Group 3 encoding

Advantages of CCITT G3 1D

1) Very simple to implement.
2) It is standard for document imaging applications.

Disadvantages of CCITT G3 1D

1) CCITT Group 3 1D assumes a reliable communication link and does not provide any error protection mechanism when used for application such as facsimile.
2) In Group 3 – 1D technique, only horizontal run – length coding is done. But, redundancy is present vertically as well. So, in Group 3 – 2D technique, vertical redundancy is also tapped.

Q. 9 Explain CCITT Group 3 2D compression in detail.

Ans. :

CCITT G3 2D Compression

With Group 3 Two-Dimensional (G32D) encoding, the way a scan line is encoded may depend on the immediately preceding scan-line data. Many images have a high degree of vertical coherence (redundancy). By describing the differences between two scan lines, rather than describing the scan line contents, 2D encoding achieves better compression.

The first pixel of each run length is called a changing element. Each changing element marks a color transition within a scan line (the point where a run of one color ends and a run of the next color begins).

The position of each changing element in a scan line is described as being a certain number of pixels from a changing element in the current, coding line (horizontal coding is performed) or in the preceding, reference line (vertical coding is performed). The output codes used to describe the actual positional information are called Relative Element Address Designate (READ) codes.

Shorter code words are used to describe the color transitions that are less than four pixels away from each other on the code line or the reference line. Longer code words are used to describe color transitions lying a greater distance from the current changing element.

2D encoding is more efficient than 1-dimensional because the usual data that is compressed (typed or handwritten documents) contains a high amount of 2D coherence.

Because a G32D-encoded scan line is dependent on the correctness of the preceding scan line, an error, such as a burst of line noise, can affect multiple, 2-dimensionally encoded scan lines. If a transmission error corrupts a segment of encoded scan line data, that line cannot be decoded. But, worse still, all scan lines occurring after it also decode improperly.

To minimize the damage created by noise, G32D uses a variable called a K factor and 2-dimensionally encodes K-1 lines following a 1-dimensionally encoded line. If corruption of the data transmission occurs, only K-1 scan lines of data will be lost. The decoder will be able to resync the decoding at the next available EOL code.

The typical value for K is 2 or 4. G32D data that is encoded with a K value of 4 appears as a single block of data. Each block contains three lines of 2D scan-line data followed by a scan line of 1-dimensionally encoded data.

The K variable is not normally used in decoding the G32D data. Instead, the EOL code is modified to indicate the algorithm used to encode the line following it. If a 1 bit is appended to the EOL code, the line following is 1-dimensionally encoded; if a 0 bit is appended, the line following the EOL code is 2-dimensionally encoded. All other transmission code word markers (FILL and RTC) follow the same rule as in G31D encoding. K is only needed in decoding if regeneration of the previous 1-dimensionally encoded scan line is necessary for error recovery.

Algorithm

1) Accept the image file and divide it into blocks of size k.

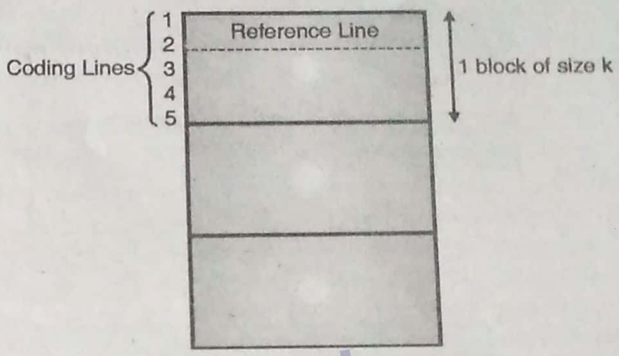


Fig. 2.3

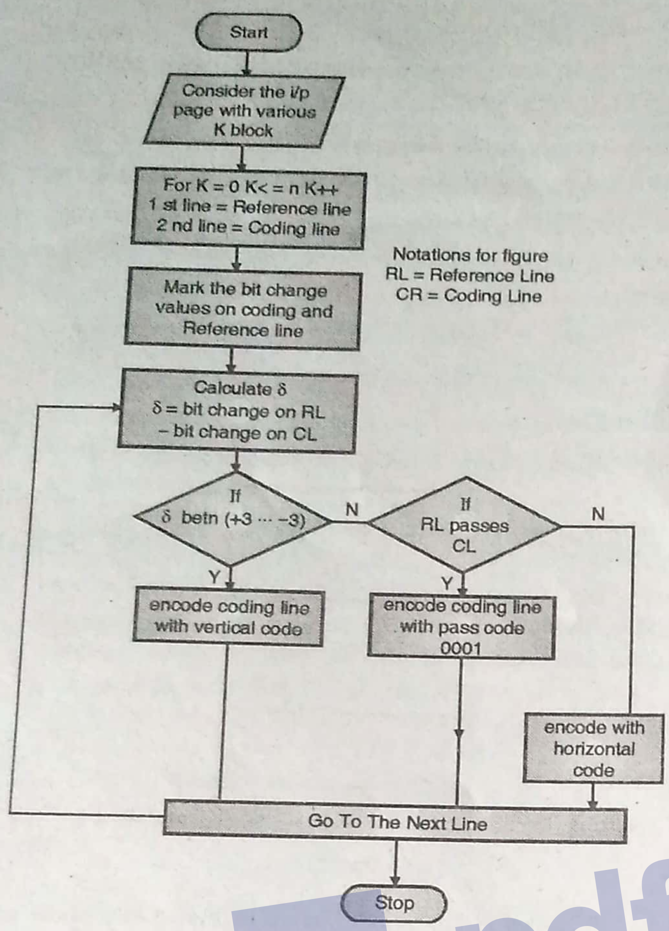
As shown in Fig. 2.3, the image is divided into blocks of size k (e.g., k = 3). In each block, the first line is taken as the reference line, and the remaining lines are taken as coding lines (2-5).

2) The reference line is encoded as Group 3 – 1D that is Huffman encoding (Make over code table and terminating code table).

3) In Group 3 – 2D, use 3 different coding techniques to exploit the vertical redundancy.
   (a) Vertical redundancy
   (b) Horizontal redundancy
   (c) Pass codes

4) Scan each coding line and compare it with the reference line. On the reference line take note of the index markers (b1, b2). On the coding line we record the index markers (a1, a2).

Flowchart for CCITT G32D



Q. 10 Give difference between CCITT GROUP 3 2D Vs CCITT GROUP 4 2D

Ans. : CCITT GROUP 3 2D Vs CCITT GROUP 4 2D

Sr. No.	CCITT GROUP 3 2D	CCITT GROUP 4 2D
1.	Entire Image is divided into several groups of K lines.	Entire image is treated as a single group. I.e. $K = \infty$.
2.	First line of each group is coded using CCITT group 31D.	First line of the image is coded with imaginary reference line of white pixels.
3.	Remaining lines of group are coded using first line as a reference line.	For coding i^{th} line (i-1) the line is used as a reference line.
4.	Because of K group error handling capabilities are existing.	If error occurs it will propagate in entire document.
5.	Lesser compression as compare to Group 4 2D.	Higher compression than group 3 2D.
6.	Simple as compared to group 4 2D.	Complex as compared to group 3 2D.
7.	Implemented only using software.	Need hardware implementation.
8.	Compression is faster.	Compression is slower.
9.	It uses End of line [EOL] signal.	It uses end of pages (EOP).
10.	At the end of fax it uses RTC [returned to control] signal.	It uses EOFB (End of FAX Block) signal.
11.	Average resolution of Fax.	Higher resolution of Fax.

Q. 6 Explain JPEG compression scheme.

Ans. :

JPEG compression involves

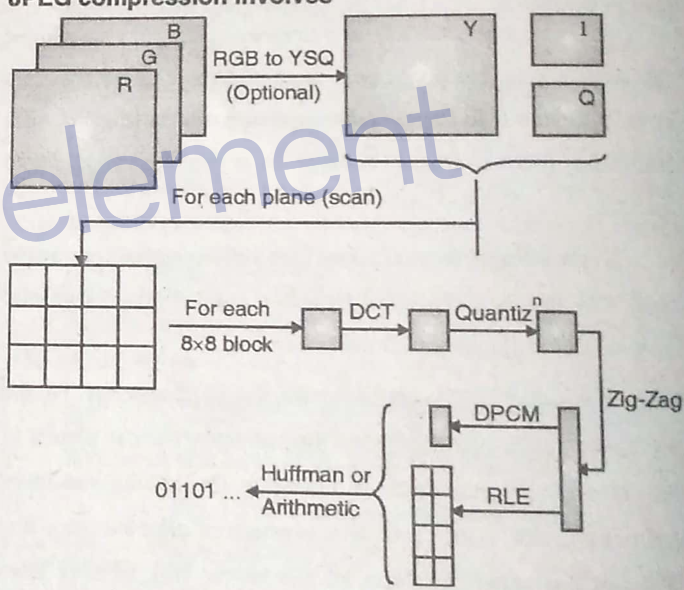


Fig. 2.4 : JPEG compression scheme

JPEG encoding

Major steps of JPEG encoding involves :

1) DCT (Discrete cosine Transformation)
2) Quantization
3) Zig-Zag
4) DPCM or DC component
5) RLE on AC components
6) Entropy coding

1) The Discrete Cosine Transform (DCT)

A particular conversion to $Y'C_BC_R$ is specified in the JFIF standard, and should be performed for the resulting JPEG file to have maximum compatibility. However, some JPEG implementations in "highest quality" mode do not apply this step and instead keep the color information in the RGB color model, where the image is stored in separate channels for red, green and blue brightness components. This results in less efficient compression, and would not likely be used when file size is especially important.

By applying DCT spatial representation is converted into frequency domain

DCT calculation

The formula for discrete cosine transform (creating DCT coefficient) is as follows :

$$\text{DCT}(i,j) = \frac{1}{\sqrt{2N}} C(i)\, C(j) \sum\sum \text{pixel}(x,y) \cos\left[\frac{(2x+1)i\pi}{2N}\right] \cos\left[\frac{(2y+1)j\pi}{2N}\right] \quad \ldots(1)$$

The formula for inverse discrete cosine transform (restoring original pixel information from a DCT coefficient) is :

$$\text{Pixel}(x,y) = \frac{1}{\sqrt{2N}} \sum\sum C(i)\, C(j)\, \text{DCT}(i,j) \cos\left[\frac{(2x+1)i\pi}{2N}\right] \cos\left[\frac{(2x+1)j\pi}{2N}\right] \quad \ldots(2)$$

Step 1 :

At this stage consider grey scale image with 8×8 block image is divided into 8-pixel by 8-pixel block. Then DCT coefficient are generated by applying DCT on 8×8 block with Equation (1).

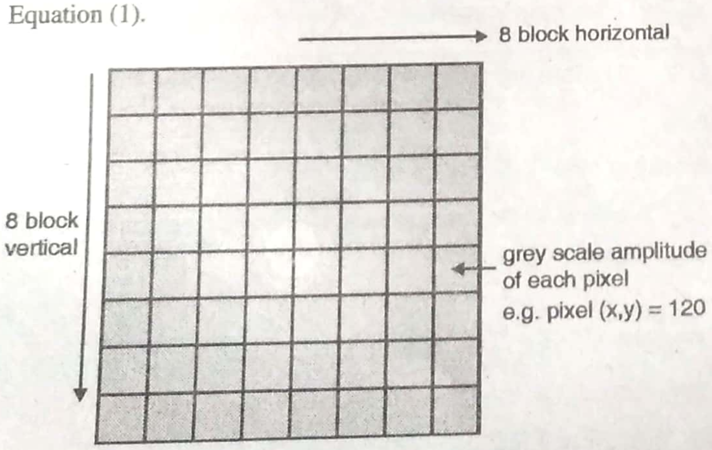


Fig. 2.5 : 8 × 8 image

Step 2 :

At this stage, after applying DCT Coefficient Equation (1), we will get 8×8 output matrix showing DCT coefficients.

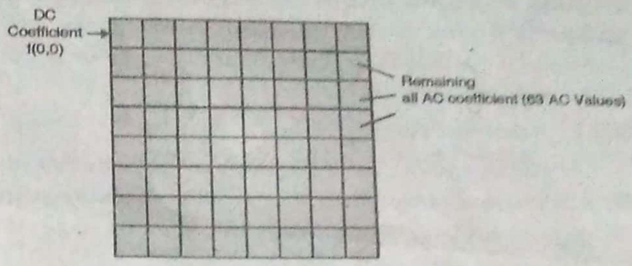


Fig. 2.6 : O/P matrix showing DCT coefficient

The top-left corner entry with the rather large magnitude. This is the DC coefficient. The remaining 63 coefficients are called the AC coefficients. The advantage of the DCT is its tendency to aggregate most of the signal in one corner of the result, as may be seen above. The quantization step to follow accentuates this effect while simultaneously reducing the overall size of the DCT coefficients, resulting in a signal that is easy to compress efficiently in the entropy stage.

Step 3 : Quantization

The human eye is good at seeing small differences in brightness over a relatively large area, but not so good at distinguishing the exact strength of a high frequency brightness variation. This allows one to greatly reduce the amount of information in the high frequency components. This is done by simply dividing each component in the frequency domain by a constant for that component, and then rounding to the nearest integer. This rounding operation is the only lossy operation in the whole process if the DCT computation is performed with sufficiently high precision. As a result of this, it is typically the case that many of the higher frequency components are rounded to zero, and many of the rest become small positive or negative numbers, which take many fewer bits to represent.

Thus The Quantized coefficient is given by

$$\text{Quantized coefficient}(i,j) = \frac{\text{DCT}(i,j)}{\text{Quantum}(i,j)}$$

Step 4 : Zig-Zag sequence

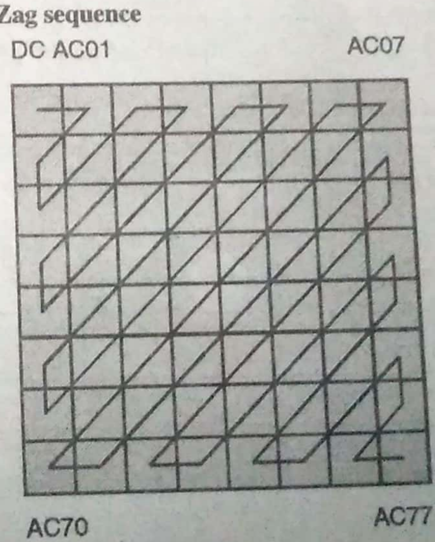


Fig. 2.7 : Use of Zig-Zag sequence for DCT coefficients

After the quantization step, most of the coefficients towards the lower right corner are zero. The Zigzag ordering rearranges the coefficients in one dimensional order, in result most of the zeroes will be placed at the end of the stream.

Step 5 : Entropy encoding

An entropy encoding is a lossless **data compression** scheme that is independent of the specific characteristics of the medium. Entropy is a lower bound on the average number of bits needed to represent the symbols (the data compression limit)

Formula for entropy is as follow :

$$\text{Entropy in number of bits} = \log_2 (\text{probability of object})$$

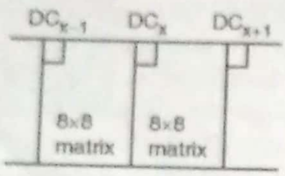


Fig. 2.8 : Successive blocks of quantized matrixes

Step 6 : DC coefficient coding

For DC coefficient coding, DC prediction is required i.e. DC prediction is nothing but DC coefficient of previous 8 × 8 block is subtracted from current 8 × 8 block. This generate differential DC coefficient. Typically small in value due to high degree of correlation between neighbouring 8 × 8 blocks.

$$\text{DC coefficient is delta D} = DC_x - DC_{x-1}$$

Each differential DC coefficient is encoded by utilizing 2 symbols.

Symbol – 1 and Symbol – 2

Symbol – 1 represents one piece of information called "size" the number of bits used to encode amplitude of DC coefficient.

Symbol – 2 represents amplitude of DC coefficient

Symbol – 1 encoded with VLC from Huffman table.

Symbol – 2 encoded with variable length Integer code with the help of following table :

Table 2.1 : DC coefficient values

Bit length	BCD	Differential DC coeff. values
0	0000	0
1	0001	-1,1
2	0010	-3,-2,2,3
3	0011	-7.... -4, 4.....7
4	0100	-15.....-8, 8.....15
5	0101	-31.....-16, 16.....31
6	0110	-63.....-32, 32.....63
7	0111	-127.....-33, 33.....127
8	1000	-255.....-128, 128...255
9	1001	-511.....256, 256.....511
10	1010	-1023.....-512, 512.....1023

Step 7 : AC coefficient coding

The remaining 63 coefs (the AC coefs) of each 64-element vector usually contain many zeros and so are coded with a combined run-amplitude Huffman code. The codeword represents the run-length of zeros before a non-zero coef and the Size of that coef. This is then followed by the Additional Bits which define the coef amplitude and sign precisely. Size and Additional Bits are defined just as for DC coefs. This 2-dimensional Huffman code (Run, Size) is efficient because there is a strong correlation between the Size of a coef and the expected Run of zeros which precedes it - small coefs usually follow long runs; larger coefs tend to follow shorter runs. No single 2-D event is so probable that the Huffman code becomes inefficient. In order to keep the code table size below 256, only the following Run and Size values are coded:

RUN 0 → 15 SIZE 1 → 10

These require 160 codes. Two extra codes, corresponding to (Run,Size) = (0,0) and (15,0) are used for EOB (End-of-block)and ZRL (Zero run length).

EOB is transmitted after the last non-zero coef in a 64-vector. It is the most efficient way of coding the final run of zeros. It is omitted in the rare case that the final element of the vector is non-zero. ZRL is transmitted whenever RUN>15, and represents a run of 16 zeros (15 zeros and a zero amplitude coef) which can be part of a longer run of any length. Hence a run of 20 zeros followed by -5 would be coded as (ZRL) (4,3) 010. When the code tables are defined in the image header, each codeword is assigned to a given (Run,Size) pair by making the decoded output byte Code Byte equal to (16 RUNS+ SIZE).

Q. 7 Explain working of JPEG2000 compression.

Ans. :

Working of JPEG2000 compression

The JPEG2000 compression standard is composed of the stages shown in the Fig. 2.9.

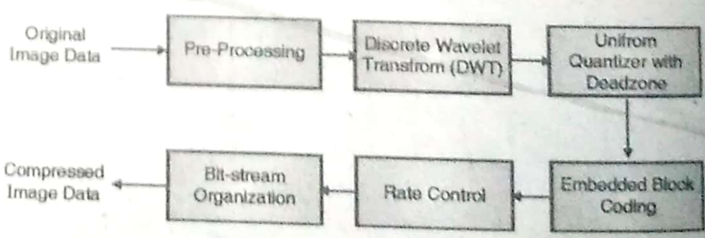


Fig. 2.9

1) Pre-processing

In the first stage, pre-processing is performed. Pre-processing actually contains three sub-stages, as shown in Fig. 2.10. These steps must be performed so that the discrete wavelet transformation can be properly performed.

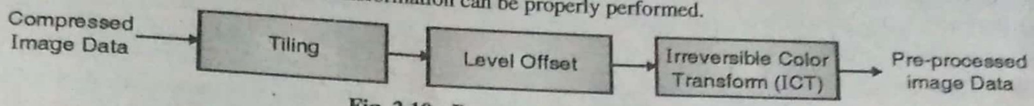


Fig. 2.10 : Pre-processing sub-stages

In tiling the image is split into so-called tiles, rectangular regions of the image that are transformed and encoded separately. Tiles can be any size, and it is also possible to consider the whole image as one single tile. Once the size is chosen, all the tiles will have the same size (except optionally those on the right and bottom borders). Dividing the image into tiles is advantageous in that the decoder will need less memory to decode the image and it can opt to decode only selected tiles to achieve a partial decoding of the image.

JPEG2000 require sample data to have a nominal dynamic range centered about zero. The level offset pre-processing stage ensures that this requirement is met

An irreversible color transform (ICT) is applied to convert RGB data into YCrCb data like this :

$$\begin{bmatrix} Y \\ C_r \\ C_b \end{bmatrix} = \begin{bmatrix} 0.299 & 0.586 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

2) Discrete Wavelet Transformation

JPEG2000 uses a discrete wavelet decomposition (DWT) to decompose each image tile into its high and low subbands as shown in Fig. 2.11. The DWT is performed by filtering each row and column of the pre-processed image tile with a high-pass and low-pass filter. Because this process results in double the number of samples, the output from each filter is down sampled by 2 (every other value is removed) so that the sample rate remains constant. Also, it does not matter if the rows or the columns of the component matrix are filtered first. The resulting DWT is the same.

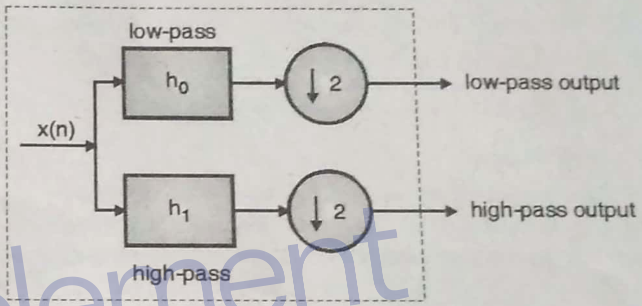


Fig. 2.11

In JPEG2000, multiple stages of the DWT are performed. The number of stages performed is implementation dependent. First, the image is decomposed into four subbands, LL_N, LH_N , HL_N and HH_N. The same 2-D subband decomposition is then applied to the lowest frequency subband (LL_N) to obtain subbands LL_{N-1} , LH_{N-1}, HL_{N-1} , and HH_{N-1}. The process is repeated N times, for an N-level decomposition. Commonly, the number of decomposition levels is around N = 5.

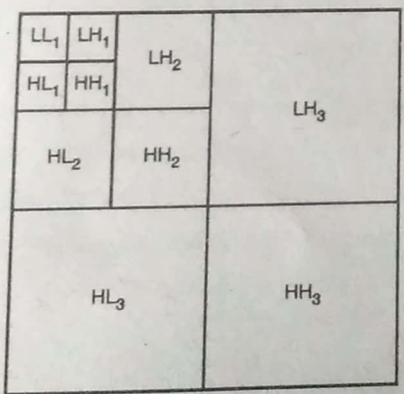


Fig. 2.12 : A three-level (N = 3) DWT decomposition

3) Uniform quantizer

The wavelet coefficients are quantized using a uniform quantizer with deadzone. For each subband b, a basic quantizer step size Δ_b is used to quantize all the coefficients in that subband according to :

$$q = \text{sign}(y)\left\lfloor \frac{|y|}{\Delta_b} \right\rfloor$$

Where y is the input to the quantizer, sign(y) denotes the sign of y, Δ_b is the step size, and q is the resulting quantizer index. Deadzone means that the quantization range about 0 is $2\Delta_b$. This ensures that more zeros result.

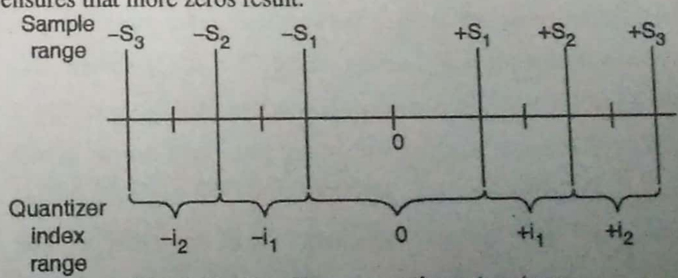


Fig. 2.13 : Deadzone quantizer structure

4) Embedded Block Coding

In JPEG2000, before coding is performed, the subbands of each tile are further partitioned into relatively small code-blocks (e.g. 64 × 64 or 32 × 32 samples) such that code blocks from a subband have the same size. Code-blocks are used to permit a flexible bit stream organization.

In JPEG2000, each code-block is encoded independently. The coding algorithm scans through the matrix of code block quantization indices in a striped manner as shown in Fig. 2.14. The code-block is partitioned into stripes with a nominal height of four samples. Then, the stripes are scanned from top to bottom, and the columns within a stripe are scanned from left to right.

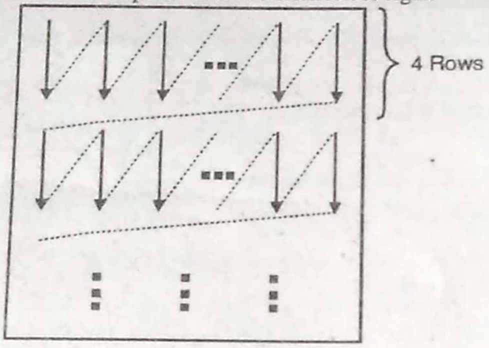


Fig. 2.14 : Stripe scan order within a block

5) Rate Control

Rate control is the process by which the code-stream is altered so that a target bit rate can be reached. Once the entire image has been compressed, a post-processing operation passes over all the compressed blocks and determines the extent to which each block's embedded bit stream should be truncated in order to achieve the target bit rate. The ideal truncation strategy is one that minimizes distortion while still reaching the target bit-rate.

6) Bit-stream Organization

In bit stream organization, the compressed data from the bit-plane coding passes are first separated into packets. One packet is generated for each precinct in a tile. A precinct is essentially a grouping of code blocks within a resolution level. Then, the packets are multiplexed together in an ordered manner to form one code-stream. The organization of the code-stream can be seen in Fig. 2.15.

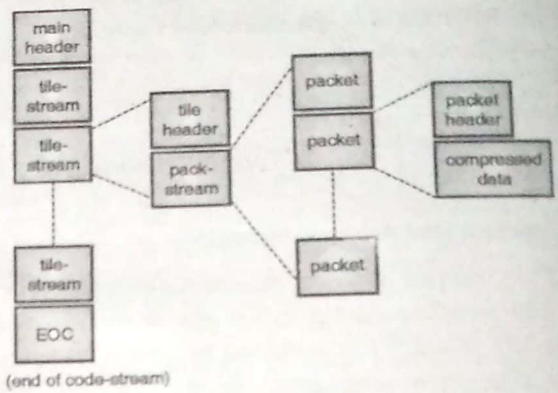


Fig. 2.15 : Code-stream organization

Chapter 3 : Digital Audio

Q. 1 Write short note on : Audio Formats.

Ans. :

Audio Formats

Sound can be stored in many different formats. Some popular audio formats are listed below and most widely used audio format MIDI, MPEG Audio and wave are described in detail.

The MIDI Format

The MIDI (Musical Instrument Digital Interface) is a format for sending music information between electronic music devices like synthesizers and PC sound cards. The MIDI format was developed in 1982 by the music industry. The MIDI format is very flexible and can be used for everything from very simple to real professional music making. MIDI files do not contain sampled sound, but a set of digital musical instructions (musical notes) that can be interpreted by your PC's sound card.

The downside of MIDI is that it cannot record sounds (only notes). Or, to put it another way : It cannot store songs, only tunes.

The upside of the MIDI format is that since it contains only instructions (notes), MIDI files can be extremely small. The example above is only 23K in size but it plays for nearly 5 minutes. The MIDI format is supported by many different software systems over a large range of platforms. MIDI files are supported by all the most popular Internet browsers. Sounds stored in the MIDI format have the extension .mid or .midi.

The RealAudio Format

The RealAudio format was developed for the Internet by Real Media. The format also supports video. The format allows streaming of audio (on-line music, Internet radio) with low bandwidths. Because of the low bandwidth priority, quality is often reduced. Sounds stored in the RealAudio format have the extension .rm or .ram.

The AU Format

The AU format is supported by many different software systems over a large range of platforms. Sounds stored in the AU format have the extension .au.

The AIFF Format

The AIFF (Audio Interchange File Format) was developed by Apple. AIFF files are not cross-platform and the format is not supported by all web browsers. Sounds stored in the AIFF format have the extension .aif or .aiff.

The SND Format

The SND (Sound) was developed by Apple. SND files are not cross-platform and the format is not supported by all web browsers. Sounds stored in the SND format have the extension .and. The WAVE Format. The WAVE (waveform) format is developed by IBM and Microsoft. It is supported by all computers running Windows, and by all the most popular web browsers. Sounds stored in the WAVE format have the extension .wav.

The MP3 Format (MPEG)

MP3 files are actually MPEG files. But the MPEG format was originally developed for video by the Moving Pictures Experts Group. We can say that MP3 files are the sound part of the MPEG video format. MP3 is one of the most popular sound formats for music recording. The MP3 encoding system combines good compression (small files) with high quality. Expect all future software systems to support it. Sounds stored in the MP3 format have the extension .mp3, or .mpga (for MPG Audio).

Q. 2 Explain WAVE File Format.

Ans. :

WAVE File Format

The WAVE file format is a subset of Microsoft's RIFF specification for the storage of multimedia files. A RIFF file starts out with a file header followed by a sequence of data chunks. A WAVE file is often just a RIFF file with a single "WAVE" chunk which consists of two sub-chunks - a "fmt " chunk specifying the data format and a "data" chunk containing the actual sample data.

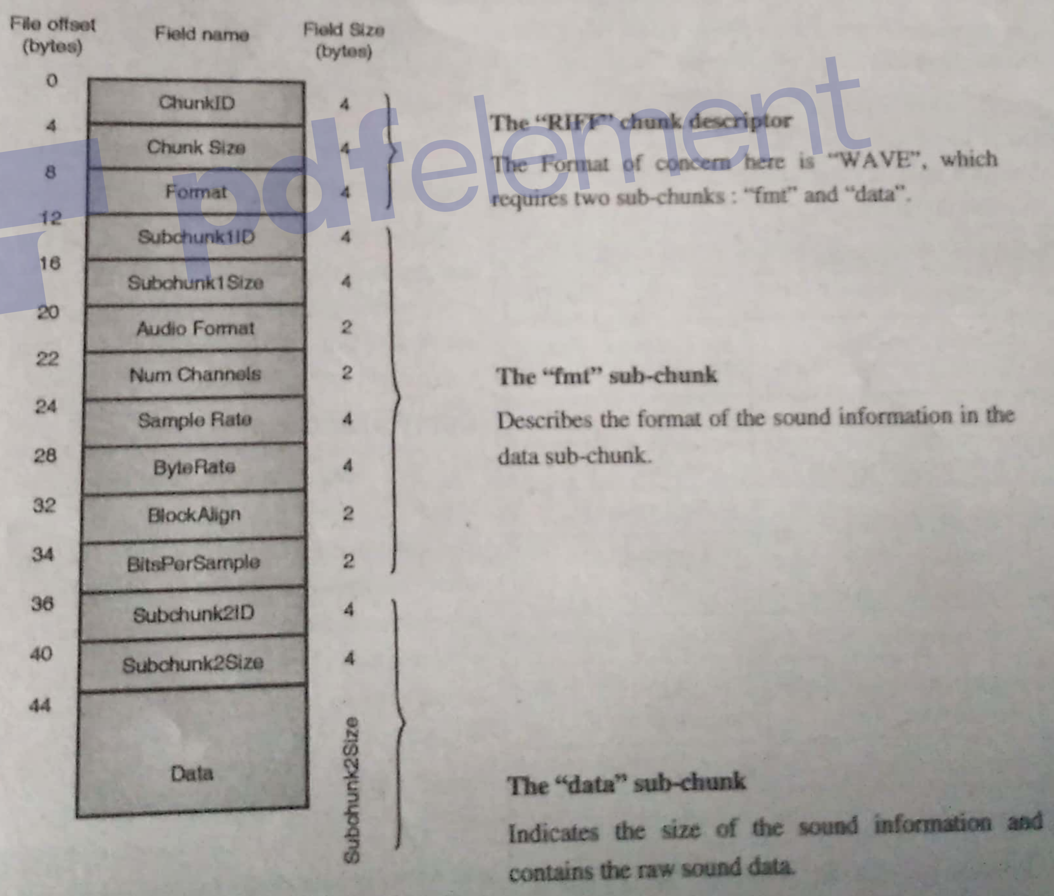


Fig. 3.1 : Wave file format

Wave File Header - RIFF Type Chunk

Wave file headers follow the standard RIFF file format structure. The first 8 bytes in the file is a standard RIFF chunk header which has a chunk ID of "RIFF" and a chunk size equal to the file size minus the 8 bytes used by the header. The first 4 data bytes in the "RIFF" chunk determines the type of resource found in the RIFF chunk. Wave files always use "WAVE". After the RIFF type comes all of the Wave file chunks that define the audio waveform.

Offset	Size	Description	Value
0x00	4	Chunk ID	"RIFF"
0x04	4	Chunk Data Size	(file size) - 8
0x08	4	RIFF Type	"WAVE"
0x10	Wave chunks		

Fig. 3.2 : RIFF Type Chunk Values

Wave File Chunks

There are quite a few types of chunks defined for Wave files. Many Wave files contain only two of them, specifically the Format Chunk and the Data Chunk. These are the two chunks needed to describe the format of the digital audio samples and the samples themselves.

Format Chunk - "fmt"

The format chunk contains information about how the waveform data is stored and should be played back including the type of compression used, number of channels, sample rate, bits per sample and other attributes.

Offset	Size	Description	Value
0x00	4	Chunk ID	"fmt " (0x666D7420)
0x04	4	Chunk Data Size	16 + extra format bytes
0x08	2	Compression code	1 - 65,535
0x0a	2	Number of channels	1 - 65,535
0x0c	4	Sample rate	1 - 0xFFFFFFFF
0x10	4	Average bytes per second	1 - 0xFFFFFFFF
0x14	2	Block align	1 - 65,535
0x16	2	Significant bits per sample	2 - 65,535
0x18	2	Extra format bytes	0 - 65,535

Chunk ID and Data Size

The chunk ID is always "fmt" (0x666D7420) and the size is the size of the standard wave format data (16 bytes) plus the size of any extra format bytes needed for the specific Wave format, if it does not contain uncompressed PCM data. The chunk ID string ends with the space character (0x20).

Compression Code

The first word of format data specifies the type of compression used on the Wave data included in the Wave chunk found in this "RIFF" chunk.

Number of Channels

The number of channels specifies how many separate audio signals that are encoded in the wave data chunk. A value of 1 means a mono signal, a value of 2 means a stereo signal, etc.

Sample Rate

The number of sample slices per second. This value is unaffected by the number of channels.

Average Bytes Per Second

This value indicates how many bytes of wave data must be streamed to a D/A converter per second in order to play the wave file. This information is useful when determining if data can be streamed from the source fast enough to keep up with playback. This value can be easily calculated with the formula:

AvgBytesPerSec = SampleRate * BlockAlign

Block Align

The number of bytes per sample slice. This value is not affected by the number of channels and can be calculated with the formula :

BlockAlign = SignificantBitsPerSample / 8 * NumChannels

Significant Bits Per Sample

This value specifies the number of bits used to define each sample. This value is usually 8, 16, 24 or 32. If the number of bits is not byte aligned (a multiple of 8) then the number of bytes used per sample is rounded up to the nearest byte size and the unused bytes are set to 0 and ignored.

Extra Format Bytes

This value specifies how many additional format bytes follow. It does not exist if the compression code is 0 (uncompressed PCM file) but may exist and have any value for other compression types depending on what compression information is need to decode the wave data. If this value is not word aligned (a multiple of 2), padding should be added to the end of this data to word align it, but the value should remain non-aligned.

Data Chunk - "data"

The Wave Data Chunk contains the digital audio sample data which can be decoded using the format and compression method specified in the Wave Format Chunk. If the Compression Code is 1 (uncompressed PCM), then the Wave Data contains raw sample values. This document explains how an uncompressed PCM data is stored, but will not get into the many supported compression formats. Wave files usually contain only one data chunk, but they may contain more than one if they are contained within a Wave List Chunk ("wavl").

Offset	Length	Type	Description	Value
0x00	4	char[4]	chunk ID	"data" (0x64617461)
0x04	4	dword	chunk size	depends on sample length and compression
0x08	sample data			

Fig. 3.3 : Data Chunk Format

Q. 3 Explain in detail MIDI.

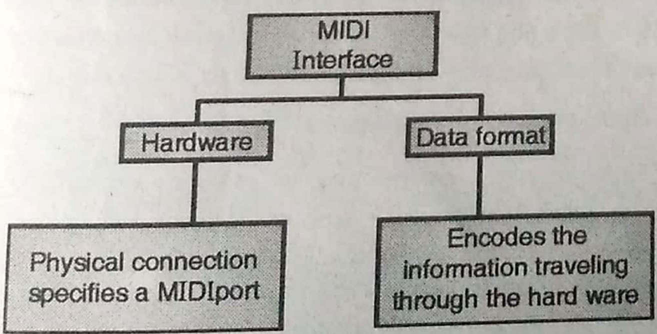
Ans. :

Musical Instrument Digital Interface (MIDI)

Musical Instrument Digital Interface is an industry-standard protocol that enables electronic musical instruments (synthesizers, drum machines), computers and other electronic equipment (MIDI controllers, sound cards, samplers) to communicate and synchronize with each other. Unlike analog devices, MIDI does not transmit an audio signal - it sends event messages about pitch and intensity, control signals for parameters such as volume, vibrato and panning, cues, and clock signals to set the tempo.

The Musical Instrument Digital Interface (MIDI) protocol has been widely accepted and utilized by musicians and composers since its conception in 1983. MIDI data is a very efficient method of representing musical performance information, and this makes MIDI an attractive protocol not only for composers or performers, but also for computer applications which produce sound, such as multimedia presentations or computer games.

However, the lack of standardization of synthesizer capabilities hindered applications developers and presented new MIDI users with a rather steep learning curve to overcome.



(a) Hardware Aspects of MIDI

Three 5-pin ports found on the back of every MIDI unit. MIDI IN : The connector via which the device receives all MIDI data. MIDI OUT : The connector through which the device transmits all the MIDI data it generates itself. MIDI THROUGH : The connector by which the device echoes the data receives from MIDI IN.

Fig. 3.4 illustrates a typical setup where :

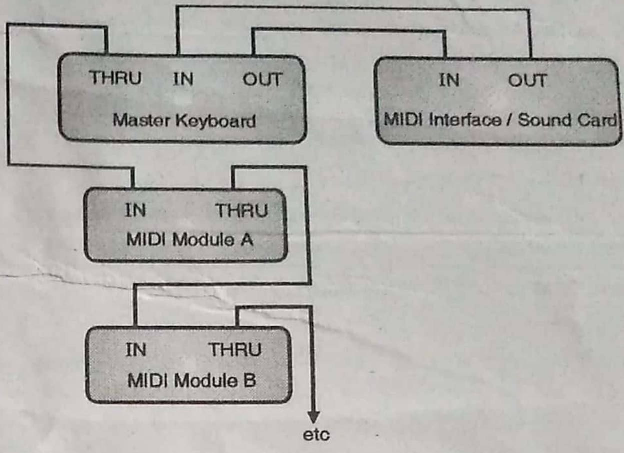


Fig. 3.4

A Typical MIDI Sequencer Setup

MIDI OUT of synthesizer is connected to MIDI IN of sequencer. MIDI OUT of sequencer is connected to MIDI IN of synthesizer and through to each of the additional sound modules. During recording, the keyboard-equipped synthesizer is used to send MIDI message to the sequencer, which records them. During play back : messages are send out from the sequencer to the sound modules and the synthesizer which will play back the music

(b) MIDI standard identifications

Bitstream encoding format for MIDI "messages" that, in the words of the standards document, "can be thought of as instructions which tell a music synthesizer how to play a piece of music."

Three levels have been established to manage player conformance :

General MIDI System Level 1 (GM1), designed to provide the minimum level of compatibility among MIDI hardware and software; includes 128 presets for instruments and 47 for percussion.

General MIDI System Level 2 (GM2), extensions to provide greater functionality, may not be as widely supported.

General MIDI Lite (GM lite), reduced performance, especially in mobile applications.

(c) MIDI Reception Modes

Functionality of Mode set-ON

ON mode is set then the MIDI device monitors all the channel and responds to all the messages irrespective of the channels through which it is transmitted. If further it is set to poly then the device can play several note at a time. On the other hand if it set to mono the device plays notes one at a time.

Functionality of Mode set-Off

Off mode is set then the MIDI device monitors and responds to that channel through which it is transmitted. If further it is set to poly then the device can play several note at a time. On the other hand if it set to mono the device plays notes one at a time.

Q. 4 Explain MIDI Messages.

Ans. : MIDI Messages

MIDI messages are used by MIDI devices to communicate with each other.

Structure of MIDI messages

MIDI message includes a status byte and up to two data bytes.

Status byte

The most significant bit of status byte is set to 1.

The 4 low-order bits identify which channel it belongs to (four bits produce 16 possible channels).

The 3 remaining bits identify the message.

The most significant bit of data byte is set to 0.

Classification of MIDI messages

```
                                                   ---- voice messages
                  --- channel messages----|
                 |                                 ---- mode messages
                 |
MIDI messages ----|
                                                   ---- common messages
                 |---- system messages ----|--- real-time messages
                                                   --- exclusive messages
```

A. Channel messages

Each MIDI cable sends 16 channels of information and the channel messages are messages that affect each channel independent of one another, the system messages are messages that affect the entire MIDI module . Thus Channel messages are messages that affect each channel independent of one another.

There are two types of channel messages :

1. Channel Voice Messages
2. Channel Mode Messages.

1. Channel voice messages

Almost all MIDI devices are equipped to receive MIDI messages on one or more of 16 selectable MIDI channel numbers. A device's particular voice (or patch, program, timbre) will respond to messages sent on the channel it is tuned to and ignore all other channel messages, analogous to a television set receiving only the station it is tuned to and rejecting the others. channel voice messages convey information about whether to turn a note on or off, what patch to change to, how much key pressure to exert (called aftertouch), etc.

2. Channel mode messages

Channel mode messages are a special case of the Control Change message (&HBx or 1011nnnn). The difference between a Control message and a Channel Mode message, which share the same status byte value, is in the first data byte. Data byte values 121 through 127 have been reserved in the Control Change message for the channel mode messages.

Channel mode messages determine how an instrument will process MIDI voice messages. Mode Messages are used to assign voice relationship for up to 16 channels that is to set the device to MONO mode or POLY mode. Omni mode on enables device to receive voice messages on all channels.

B. System Messages

System messages carry information that is not channel specific, such as timing signal for synchronization, positioning information in pre-recorded MIDI sequences, and detailed setup information for the destination device.

1. System real-time messages

These messages are related to synchronization. These messages are used for setting values of real time parameters of a system such as start, stop sequencer.

2. System common messages

These messages are common to complete system. These messages provide for function such as selecting a song, setting song position point with number of beats and sending a tune request to an analog synthesizer.

3. System exclusive message

These messages contain manufacturer specific data such as identification, serial number, model number and other information. System exclusive message are related to things that cannot be standardized. It is just a stream of bytes, all with their high bits set to 0, bracketed by a pair of system exclusive start and end messages (&HF0 and &HF7).

Q. 5 Explain different Schemes of Audio Compression.

Ans. : Schemes of Audio Compression

The most commonly used compression schemes for audio are :

1. Pulse Code Modulation

A signal is pulse code modulated to convert its analog information into a binary sequence, i.e., 1s and 0s. The output of a PCM will resemble a binary sequence.

The transmitter section of a Pulse Code Modulator circuit consists of **Sampling, Quantizing** and **Encoding**, which are performed in the analog-to-digital converter section. The low pass filter prior to sampling prevents aliasing of the message signal.

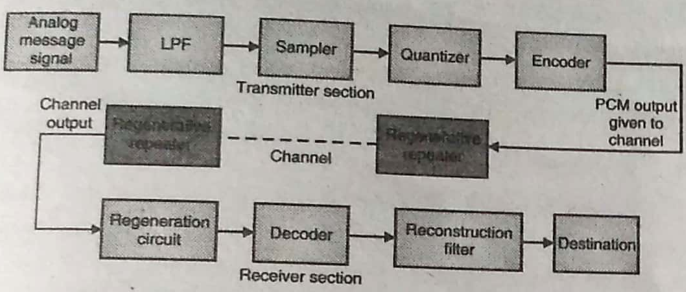


Fig. 3.5

The basic operations in the receiver section are **regeneration of impaired signals, decoding,** and **reconstruction** of the quantized pulse train. Following is the block diagram of PCM which represents the basic elements of both the transmitter and the receiver sections.

(a) **Low Pass Filter :** This filter eliminates the high frequency components present in the input analog signal which is greater than the highest frequency of the message signal, to avoid aliasing of the message signal.

(b) **Sampler :** This is the technique which helps to collect the sample data at instantaneous values of message signal, so as to reconstruct the original signal. The sampling rate must be greater than twice the highest frequency component **W** of the message signal, in accordance with the sampling theorem.

(c) **Quantizer :** Quantizing is a process of reducing the excessive bits and confining the data. The sampled output when given to Quantizer, reduces the redundant bits and compresses the value.

(d) **Encoder :** The digitization of analog signal is done by the encoder. It designates each quantized level by a binary code. The sampling done here is the sample-and-hold process. These three sections (LPF, Sampler, and Quantizer) will act as an analog to digital converter. Encoding minimizes the bandwidth used.

(e) **Regenerative Repeater :** This section increases the signal strength. The output of the channel also has one regenerative repeater circuit, to compensate the signal loss and reconstruct the signal, and also to increase its strength.

(f) **Decoder :** The decoder circuit decodes the pulse coded waveform to reproduce the original signal. This circuit acts as the demodulator.

(g) **Reconstruction Filter :** After the digital-to-analog conversion is done by the regenerative circuit and the decoder, a low-pass filter is employed, called as the reconstruction filter to get back the original signal.

2. Adaptive Differential Pulse Code Modulation (ADPCM)

ADPCM stands for Adaptive Differential Pulse Code Modulation. **ADPCM** is a very widely used lossy coding scheme. The acronym refers to waveform codecs which instead of quantizing the sound signal directly, like PCM codecs, quantize the difference between the sound signal and a prediction that has been made of the sound signal.

If the prediction is accurate then the difference between the real and predicted sound samples will have a lower variance than the real sound samples, and will be accurately quantized with fewer bits than would be needed to quantize the original sound samples. At the decoder the quantized difference signal is added to the predicted signal to give the reconstructed sound signal. Using this technique one can achieve about 40-80% compression.

For the samples that are highly correlated, when encoded by PCM technique, leave redundant information behind. To process this redundant information and to have a better output, it is a wise decision to take a predicted sampled value, assumed from its previous output and summarize them with the quantized values.

Such a process is called as **Differential PCM (DPCM)** technique. DPCM differs from PCM because it quantizes the difference of the actual sample and predicted value. That is the reason it is called as differential PCM.

3. Delta Modulation

The type of modulation, where the sampling rate is much higher and in which the stepsize after quantization is of a smaller value Δ, such a modulation is termed as **delta modulation.**

Following are some of the features of delta modulation :

(a) An over-sampled input is taken to make full use of the signal correlation.

(b) The quantization design is simple.

(c) The input sequence is much higher than the Nyquist rate.

(d) The quality is moderate.

(e) The design of the modulator and the demodulator is simple.

(f) The stair-case approximation of output waveform.

(g) The step-size is very small, i.e., Δ (delta).

(h) The bit rate can be decided by the user.

(i) This involves simpler implementation.

Delta Modulation is a simplified form of DPCM technique, also viewed as **1-bit DPCM scheme**. As the sampling interval is reduced, the signal correlation will be higher.

Delta Modulator

The Delta Modulator comprises of a 1-bit quantizer and a delay circuit along with two summer circuits.

Delta Demodulator

The delta demodulator comprises of a low pass filter, a summer, and a delay circuit. The predictor circuit is eliminated here and hence no assumed input is given to the demodulator.

Chapter 4 : Digital Video

Q. 1 Explain different Types of Video Signals.

Ans. :

Types of Video Signals

Video signals can be organized in three different ways : Component video, Composite video and S-video.

1. Component Video

Component video is a video signal that has been split into two or more components. In popular use, it refers to a type of analog video information that is transmitted or stored as three separate signals. component analog video signals do not use R, G, and B components but rather a colorless component, termed luma, which provides brightness information (as in black-and-white video). This combines with one or more color-carrying components, termed chroma, that give only color information.

In component video, the luminance (Y) and two color difference signals (U and V or I and Q) are separated into three separate analog signals that can be transmitted over three separate wires or stored in three separate tracks on an analog tape, or digitized separately. Component video is used in professional video production and provides the best quality and the most accurate reproduction of colors. The professional Betacam SP video cameras use component video.

2. Composite Video

Composite video signals are analog signals that combine luminance and chrominance (color) information in a single analog signal that can be transmitted over a single wire or stored in a single track on an analog magnetic tape. The NTSC video signals used by commercial television sets in the United States and Japan are an example of composite signals. Composite video is particularly prone to errors in reproducing exact colors due to the overlap of the color and luminance signals.

Most of the analog home video equipment records a signal in the composite format and a composite video interface is used to connect a VHS tape player, game consoles or a DVD player to the television.

In composite video, three source signals are combined with sync pulses to form a composite video signal. The three source signals are referred as YUV in which Y represents the brightness of the picture and it also includes the synchronizing pulses. The color information is carried between U and V.

Two orthogonal phases of a color carrier signal are mixed with them in the first place to form a signal called as chrominance. The Y signal and the UV signal are then combined together and this is equivalent to the frequency-division multiplexing. The signals are compressed and then channeled through a single wire. Comb filter present in the television set is used to separate the signals. This results in degradation of signal quality.

3. S-video

S-video is one of a number of methods of separating a video signal into different components for transmission from a video cassette recorder or playback machine to a television set or video monitor. The technology was introduced to the market by JVC in 1987 as "separate video," which was quickly shortened to "s-video." S-video cables are the cables that connect two devices that are equipped with s-video capability to transfer the signal from one to another.

S-video was one of a number of enhancements in bringing the signal from the video cassette player to the television, and separates the video signal into luma, or luminescence, and chroma, or color.

S-video cables carry four or more wires wrapped together in an insulated sleeve, with S-video connectors at either end. The most common S-video connector has four pins – one for the chroma signal, one for the luma, and two ground wires, one for each signal.

S-Video is commonly used throughout the world with relative popularity. It is found on consumer TVs, DVD players, high-end video cassette recorders, digital TV receivers, DVRs, game consoles, and graphics cards. S-Video cables are used for computer-to-TV output for business or home use.

Q. 2 Explain Chroma Subsampling.

Ans. : Chroma Subsampling

Chroma subsampling is the process whereby the color information in the image is sampled at a lower resolution than the original. A reduced color resolution in digital component video signals. To accommodate storage and bandwidth limitations, the two color components (Cb, Cr) in digital video signals are compressed by sampling them at a lower rate than the brightness (Y). Color information is actually discarded.

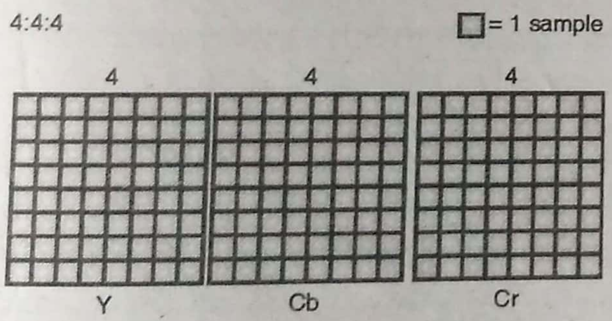
YCbCr is the YUV color space recorded digitally. Y is brightness (luma), and Cb and Cr are the U and V color difference signals. In chroma subsampling, only the colors are compressed, not the luma because the eye is more sensitive to brightness than to the color components.

Various levels of YCbCr subsampling

YCbCr is designated as 4:n:n. The 4 represents a sampling rate of 13.5 MHz, which is the standard frequency for digitizing analog NTSC, PAL and SECAM. The next two digits represent the Cb and Cr rate. Each 8x8 matrix represents a "macroblock" of 64 pixels in a video frame. The pink squares are the pixel locations where the sample is taken. Sony's HDCAM uses a different notation because it compresses both the luma and the colors.

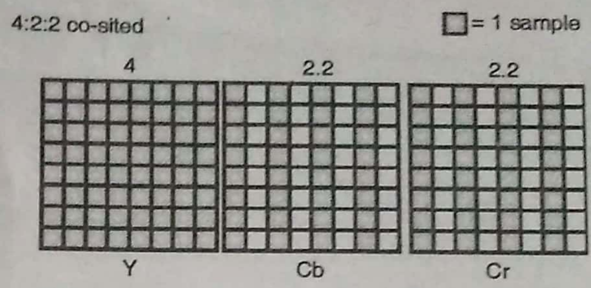
4:4:4 (Cb/Cr Same as Luma)

The resolution of chrominance information (Cb & Cr) is preserved at the same rate as the luminance (Y) information also known as 1x1 (or subsampling disabled). Cb and Cr are sampled at the same full rate as the luma. MPEG-2 supports 4:4:4 coding, but having the same number of color difference samples as the luma is considered overkill and not worth the additional bandwidth to transmit it. When video is converted from one color space to another, it is often resampled to 4:4:4 first.



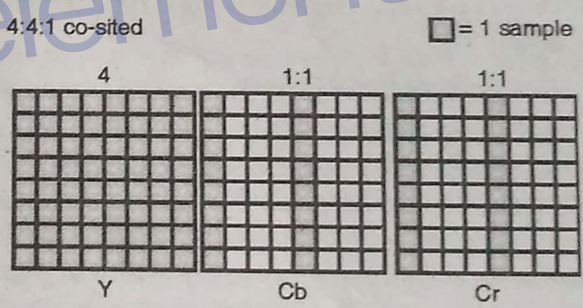
4:2:2 (1/2 the Luma Samples)

Half of the horizontal resolution in the chrominance is dropped (Cb & Cr), while the full resolution is retained in the vertical direction, with respect to the luminance. This is also known as 2 × 1 chroma subsampling, and is quite common for digital cameras. Cb and Cr are sampled at half the horizontal resolution of Y. Co-sited means that Cb/Cr samples are taken at the same time as Y. 4:2:2 color sampling is widely used and considered very high quality. It is used for prosumer and professional digital video recording, including DV (at 50 Mbps), Digital Betacam and DVCPRO 50 and is an option in MPEG-2.



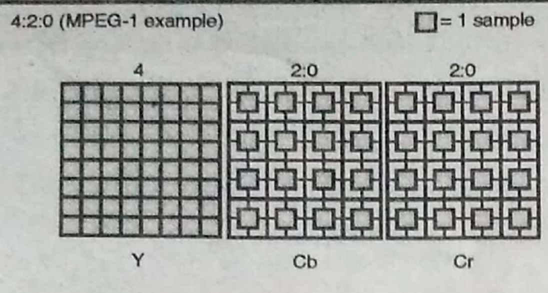
4:1:1 (1/4 the Luma Samples)

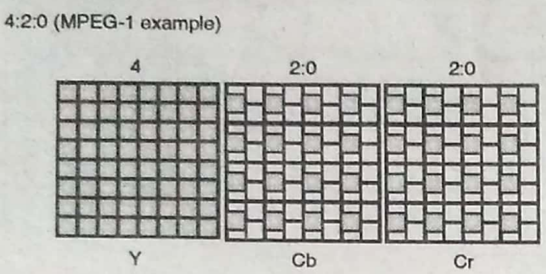
Only a quarter of the chrominance information is preserved in the horizontal direction with respect to the luminance information. I don't think this format is nearly as common as the other variations. Cb and Cr are sampled at one quarter the horizontal resolution. Co-sited means that Cb/Cr samples are taken at the same time as the Y. Co-sited 4:1:1 is used in DV, DVCAM and DVCPRO formats.



4:2:0 (1/4 the Luma Samples)

With respect to the information in the luminance channel (Y), the chrominance resolution in both the horizontal and vertical directions is cut in half. This form is also known as 2 × 2 chroma subsampling. The zero in 4:2:0 means that Cb and Cr are sampled at half the vertical resolution of Y. MPEG-1 and MPEG-2 use 4:2:0, but the samples are taken at different intervals. By the time MPEG-2 came along, it was known that 4:2:2 coding was often converted to 4:2:0, which is why MPEG-2 sampling more closely lines up with the 4:2:2 pattern. H.261/263 also uses 4:2:0.





Q. 3 Explain different Video Format.

Ans. : Video Format

Video can be stored in many different formats. Some popular video formats are as follow :

1. The AVI Format

The AVI (Audio Video Interleave) format was developed by Microsoft. The AVI format is supported by all computers running Windows, and by all the most popular web browsers. It is a very common format on the Internet, but not always possible to play on non-Windows computers. Videos stored in the AVI format have the extension .avi.

2. The Windows Media Format

The Windows Media format is developed by Microsoft. Windows Media is a common format on the Internet, but Windows Media movies cannot be played on non-Windows computer without an extra (free) component installed. Some later Windows Media movies cannot play at all on non-Windows computers because no player is available. Videos stored in the Windows Media format have the extension .wmv.

3. The MPEG Format

The MPEG (Moving Pictures Expert Group) format is the most popular format on the Internet. It is cross-platform and supported by all the most popular web browsers. Videos stored in the MPEG format have the extension .mpg or .mpeg.

4. The QuickTime Format

The QuickTime format is developed by Apple. QuickTime is a common format on the Internet, but QuickTime movies cannot be played on a Windows computer without an extra (free) component installed. Videos stored in the QuickTime format have the extension .mov.

5. The RealVideo Format

The RealVideo format was developed for the Internet by Real Media. The format allows streaming of video (on-line video, Internet TV) with low bandwidths. Because of the low bandwidth priority, quality is often reduced. Videos stored in the RealVideo format have the extension .rm or .ram.

6. The Shockwave (Flash) Format

The Shockwave format was developed by Macromedia. The Shockwave format requires an extra component to play. This component comes preinstalled with the latest versions of Netscape and Internet Explorer. Videos stored in the Shockwave format have the extension .swf.

Q. 4 Explain Block Matching Algorithm.

Ans. : Block Matching Algorithm

Block Matching Algorithm (BMA) is the most popular motion estimation algorithm. BMA calculates motion vector for an entire block of pixels instead of individual pixels. The same motion vector is applicable to all the pixels in the block. This reduces computational requirement and also results in a more accurate motion vector since the objects are typically a cluster of pixels.

BMA algorithm is illustrated in Fig. 4.1. The current frame is divided into pixel blocks and motion estimation is performed independently for each pixel block. Motion estimation is done by identifying a pixel block from the reference frame that best matches the current block, whose motion is being estimated. The reference pixel block is generated by displacement from the current block's location in the reference frame. The displacement is provided by the Motion Vector (MV). MV consists of is a pair (x, y) of horizontal and vertical displacement values.

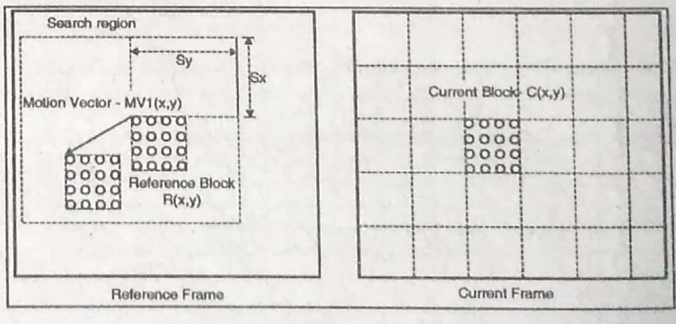


Fig. 4.1 : Block matching algorithm

There are various criteria available for calculating block matching.

Two popular criteria are listed below :

$$\text{Sum of Square Error (SSE)} = \sum_{x=1}^{x=N} \sum_{y=1}^{y=N} (C(x, y) - R(x, y))^2$$

Sum of Absolute Difference (SAD)

$$= \sum_{x=1}^{x=N} \sum_{y=1}^{y=N} |C(x, y) - R(x, y)|$$

SSE provides a more accurate block matching, however requires more computations. SAD provides fairly good match at lower computational requirement. Hence it is widely used for block matching. There are various other criteria also available such as cross correlation, maximum matching pixel count etc. The reference pixel blocks are generated only from a region known as the search area. Search region defines the boundary for the motion vectors and limits the number of blocks to evaluate. The height and width of the search region is dependent on the motion in video sequence. The available computing power also determines the search range. Bigger search region requires more computation due to increase in number of evaluated candidates.

Typically the search region is kept wider (i.e. width is more than height) since many video sequences often exhibit panning motion. The search region can also be changed adaptively depending upon the detected motion. The horizontal and vertical search range, Sx & Sy, define the search area (+/-Sx and +/- Sy) as illustrated in Fig. 4.1.

Q. 5 Write short note on H.261.

Ans. : H.261

H.261 is video coding standard by the ITU. It was designed for data rates which are multiples of 64 Kbit/s, and is sometimes called p × 64 K bit/s (p is in the range 1-30). These data rates suit ISDN lines, for which this video codec was originally designed for. H.261 transport video stream using the real-time transport protocol, RTP, with any of the underlying protocols that carry RTP.

H261 supports motion compensation in the encoder as an option. In motion compensation a search area is constructed in the previous (recovered) frame to determine the best reference macroblock.

H261 supports two image resolutions, QCIF (Quarter Common Interchange format) which is (144 × 176 pixels) and CIF (Common Interchange format) which is (288 × 352)

The video multiplexer structures the compressed data into a hierarchical bitstream that can be universally interpreted. The hierarchy has four layers :

1. **Picture layer :** Corresponds to one video picture (frame).
2. **Group of blocks :** Corresponds to 1/12 of CIF pictures or 1/3 of QCIF.
3. **Macroblocks :** Corresponds to 16x16 pixels of luminance and the two spatially corresponding 8 × 8 chrominance components.
4. **Blocks :** Corresponds to 8 × 8 pixels.

H.261 Encoder

The source coder operates on only non-interlaced pictures. Pictures are coded as luminance and two color difference components(Y, Cb, Cr). The Cb and Cr matrices are half the size of the Y matrix.

The three main elements in an H.261 encoder as illustrated in Fig. 4.2 are :

Prediction

H261 defines two types of coding. INTRA coding where blocks of 8x8 pixels each are encoded only with reference to themselves and are sent directly to the block transformation process. On the other hand INTER coding frames are encoded with respect to another reference frame

A prediction error is calculated between a 16 × 16 pixel region (macroblock) and the (recovered) correspondent macroblock in the previous frame. Prediction error of transmitted blocks (criteria of transmission is not standardized) are then sent to the block transformation process.

Blocks are inter or intra coded.

Intra-coded blocks stand alone.

Inter-coded blocks are based on predicted error between the previous frame and this one.

Intra-coded frames must be sent with a minimum frequency to avoid loss of synchronisation of sender and receiver.

Block Transformation

H261 supports motion compensation in the encoder as an option. In motion compensation a search area is constructed in the previous (recovered) frame to determine the best reference macroblock. Both the prediction error as well as the motion vectors specifying the value and direction of displacement between the encoded macroblock and the chosen reference are sent.

INTRA coding where blocks of 8 × 8 pixels each are encoded only with reference to themselves and are sent directly to the block transformation process. On the other hand INTER coding frames are encoded with respect to another reference frame. The inter-picture prediction removes temporal redundancy. The transform coding removes the spatial redundancy. Motion vectors are used to help the codec compensate for motion. To remove any further redundancy in the transmitted bitstream, variable length coding is used.

Quantization and Entropy Coding

The purpose of this step is to achieve further compression by representing the DCT coefficients with no greater precision than is necessary to achieve the required quality. The number of quantizers are 1 for the INTRA dc coefficients and 31 for all others.

Entropy coding involves extra compression (non-lossy) is done by assigning shorter code-words to frequent events and longer code-words to less frequent events. Huffman coding is usually used to implement this step.

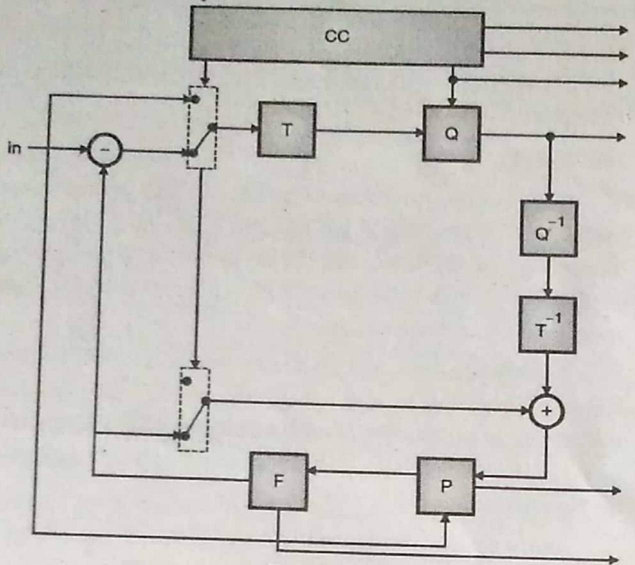


Fig. 4.2

In other words, for a given quality, we can lose coefficients of the transform by using less bits than would be needed for all the values This leads to a "coarser" picture. We can then entropy code the final set of values by using shorter words for the most common values and longer ones for rarer ones.

T	Transform	P	Flag for INTRA/INTER
Q	Quantiser	t	Flag for transmitted or not
P	Picture memory with motion compensated variable delay	qz	Quantiser indication
F	Loop filter	q	Quantizing index for transform co-efficients
CC	Coding control	v	Motion vector
		f	Switching on/off the loop filter

H261 Encoder

Q. 6 Explain MPEG-1 Compression Algorithm.

Ans. :

MPEG-1 Compression Algorithm

The MPEG-1 algorithm is based around two key techniques : Temporal compression and spatial compression. Temporal compression relies upon similarity between successive pictures using prediction and motion compensation whereas spatial compression relies upon redundancy within small areas of a picture and is based around the DCT transform, quantization and entropy coding techniques.

1) Temporal Compression

Inter (B and P) pictures are coded using motion compensation, primarily prediction and interpolation.

Prediction

The predicted picture is the previous picture modified by motion compensation. Motion vectors are calculated for each macroblock.

The motion vector is applied to all four luminance blocks in the macro block. The motion vector for both chrominance blocks is calculated from the luma vector. This technique relies upon the assumption that within a macroblock the difference between successive pictures can be represented simply as a vector transform (i.e. there is very little difference between successive pictures, the key difference being in position of the Pixels)

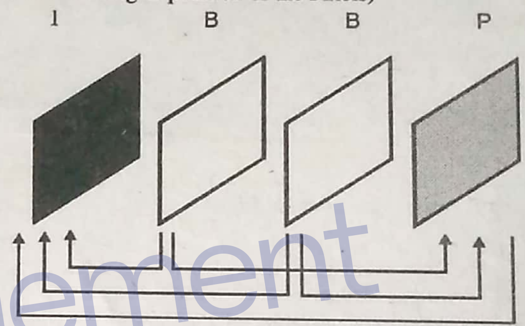


Fig. 4.3 : Make up of I, B and P pictures

Interpolation

Interpolation (or bidirectional prediction) generates high compression in that the picture is represented simply as an interpolation between the past and future I or P pictures (again this is performed on a macroblock level). Pictures are not transmitted in display order but in the order in which the decoder requires them to decode the bitstream (the decoder must of course have the reference picture(s) before any interpolated or predicted pictures can be decoded).

2) Spatial Compression

The spatial compression techniques are similar to those of JPEG, DCT, Quantization and entropy coding. The compression algorithm takes advantage of the redundancy within each block (8×8 Pixels).

The resulting compressed data stream is made up of a combination of spatial and temporal compression techniques which best suit the type of picture being compressed. Decoding is controlled through the use of MPEG system codes which are put into the data stream explaining how to reconstruct specific areas of picture.

Through a combination of techniques, MPEG-1 compression is designed to give good quality (typically similar or better quality to VHS) images from such storage media as CD-ROM. The quality is however, dependent upon the type of picture compressed and the level of redundancy within the sequence coded. Picture quality will also depend upon how well the sequence has been coded and which features are required - For Example : For fast random access, N will tend towards zero hence the quality of compression will deteriorate, if random access is not required then the number of P and B frames can increase, hence increasing the potential quality. The standard does not specify a method of compression but a syntax for the compressed data, this allows for differing compression techniques depending upon differing requirements. The decoding techniques are defined due to the nature of the compressed data stream.

This method allows for true flexibility in coding whilst retaining the format and hierarchy ensuring compatibility in the data stream and hence uniform readability.

Q. 7 Explain MPEG-7.

Ans. : MPEG-7

MPEG-7 is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group), the committee that also developed the Emmy Award winning standards known as MPEG-1 and MPEG-2, and the MPEG-4 standard. MPEG-1 and MPEG-2 standards made interactive video on CD-ROM and Digital Television possible. MPEG-4 is the multimedia standard for the fixed and mobile web enabling integration of multiple paradigms.

MPEG-7, formally named "Multimedia Content Description Interface", is a standard for describing the multimedia content data that supports some degree of interpretation of the information meaning, which can be passed onto, or accessed by, a device or a computer code. MPEG-7 is not aimed at any one application in particular; rather, the elements that MPEG-7 standardizes support as broad a range of applications as possible.

MPEG-7 offers a comprehensive set of audiovisual Description Tools (the metadata elements and their structure and relationships, that are defined by the standard in the form of Descriptors and Description Schemes) to create descriptions (i.e., a set of instantiated Description Schemes and their corresponding Descriptors at the users will), which will form the basis for applications enabling the needed effective and efficient access (search, filtering and browsing) to multimedia content. This is a challenging task given the broad spectrum of requirements and targeted multimedia applications, and the broad number of audiovisual features of importance in such context.

MPEG-7 will also standardize ways to define other descriptors as well as structures (description schemes) for the descriptors and their relationships. This description (i.e., the combination of descriptors and description schemes) will be associated with the content itself, to allow fast and efficient searching for material of a user's interest. AV material that has MPEG-7 data associated with it can be indexed and searched. This 'material' may include still pictures, graphics, 3D models, audio, speech, video, and information about how these elements are combined in a multimedia presentation ("scenarios" composition information). Special cases of these general data types may include facial expressions and personal characteristics.

Because the descriptive features must be meaningful in the context of the application, they will be different for different user domains and different applications. This implies that the same material can be described using different types of features, tuned to the area of application. All these descriptions will be coded in an efficient way - efficient for search, that is.

MPEG-7 data may be physically located with the associated AV material, in the data stream or on the same storage system, but the descriptions could also live somewhere else on the globe. When the content and its descriptions are not colocated, mechanisms that link AV material and their MPEG-7 descriptions are useful; these links should work in both directions.

Q. 8 Explain DVI file format.

Ans. : DVI Technology

Intel is the current owner of DVI, which was one of the first systems that provided practical full-motion video incorporating real-time decompression technology.

DVI technology has defined a file format for storing audio/video objects. Applications should use this and other industry standard file formats, to increase interoperability with other applications such as media editing and manipulation tools. The DVI multimedia file format is particularly appropriate for motion video objects that use the compression algorithms, and media objects that use ActionMedia II board pixel formats.

The DVI multimedia file format was designed to grow into a general purpose repository for complex multimedia objects, including information that might be added by media object editors.

DVI is actually both the name of the Digital Video Interactive hardware system sold by Intel and the file format associated with that system. DVI technology is essentially a PC-based interactive audio/video system used for multimedia applications. The DVI system consists of a board for use in an Intel-based PC, drivers, and associated software. The four components of DVI technology are :

1. DVI hardware chipset
2. Run-time software interface
3. Data compression and decompression schemes
4. Data file formats

The heart of the DVI system is the hardware architecture based on the video display processor (VDP) chipset. DVI technology was originally designed for implementation on the IBM

PC AT platform. A collection of three 16-bit, ISA-bus DVI interface boards (audio, video, and CD-ROM) were plugged into the AT, and all of the hardware capabilities were accessed through the run-time software interface. The functions in the interface were called by writing a software program using a programming language such as assembly or C.

Today, Intel distributes licenses to third-party developers who want to incorporate DVI technology into their platforms and multimedia products. All of IBM's multimedia hardware platforms (such as the Action Media II boards) and software applications are based upon DVI technology.

DVI is a major competitor of QuickTime, AVI, and MPEG for market share in digital audio/video applications.

DVI allows the storage and playback of audio and video information. All DVI images have a 5:4 pixel aspect ratio and are 256 × 240 pixels in size. DVI is also capable of storing still images and supports both a lossy and a lossless native compression method for such images. DVI works across MS-DOS, Microsoft Windows, and OS/2 platforms and supports the capability of using its own proprietary compression scheme, or using user-definable algorithms, such as JPEG, as well. Audio compression is achieved using either the ADPCM or PCM8 algorithms.

File Organization

The DVI file format is extremely flexible in its design and is used to store a wide variety of data. This format is capable of storing both still-image and motion-video/audio data. A common practice of DVI is to store each color plane of an image in a separate disk file. This allows the easy reading and writing of bitmap information, without the need to buffer data to read or write a single file.

A still image is saved using three color-channel files and possibly a colormap and alpha-channel file as well. Motion-video/audio data is stored using the Audio/Video Support System (AVSS) file format. AVSS (pronounced "avis") allows audio and video data to be stored in the same file and played back in a synchronized manner. All AVSS files have the extension .AVS or the file type AVSS.

Chapter 5 : Multimedia Network Communication and Representation

Q. 1 Write short note on : Quality of Service.

Ans. : Quality of Service

The International Standards Organization (ISO) uses Quality of Service (QoS), a concept for specifying how "good" networking services are, to define parameterization. Researchers have yet to determine the "best" set of QoS parameters for multimedia systems (or benchmarks to compare the different approaches).

Quality of Service Provide guarantees on the ability of a network to deliver predictable results.

QoS (Quality of Service) refers to a broad collection of networking technologies and techniques. The goal of QoS is to provide guarantees on the ability of a network to deliver predictable results. Elements of network performance within the scope of QoS often include availability (uptime), bandwidth (throughput), latency (delay), and error rate.

QoS involves prioritization of network traffic. QoS can be targeted at a network interface, toward a given server or router's performance, or in terms of specific applications. A network monitoring system must typically be deployed as part of QoS, to insure that networks are performing at the desired level.

QoS is especially important for the new generation of Internet applications such as VoIP, video-on-demand and other consumer services. Some core networking technologies like Ethernet were not designed to support prioritized traffic or guaranteed performance levels, making it much more difficult to implement QoS solutions across the Internet.

Resource Management Architecture

Various components of a resource management subsystem are shown in Fig. 5.1. The main goal of resource management-guaranteed delivery of multimedia data.

Goal of resource management architecture implies three main actions :

Reserving and allocating resources (end-to-end) during multimedia call establishment so that traffic can flow according to the QoS specification,

Adhering to resource allocation during multimedia delivery using proper service disciplines,

Adapting to resource changes during an ongoing multimedia session.

The resource management subsystem includes resource managers at the hosts as well as at the network nodes. Resource management protocols are used to exchange information about resources among the resource managers.

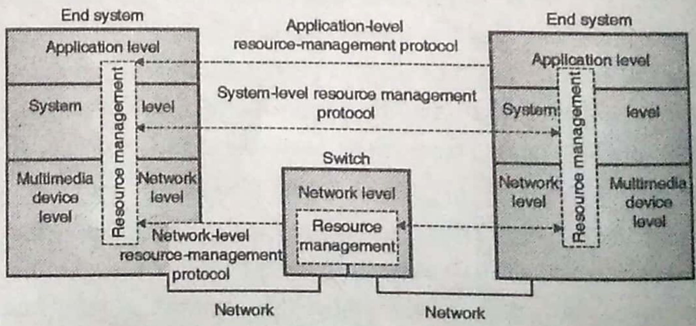


Fig. 5.1

Establishing and Closing Multimedia Call

Any multimedia user expects the application to provide a certain level of quality. Before any multimedia data is transmitted, these user-defined requirements must be communicated to the resource management entities of all involved system components. The QoS parameters are then negotiated and, where specifications differ, translated between layers. Finally, the required resources must be admitted, reserved, and allocated along the path between the senders and receiver(s). These basic steps are performed during multimedia call establishment. The call close-down procedure, from the management point of view, concerns resource deallocation.

QoS Negotiation and Translation

A general architecture for communicating QoS parameters comprises two services : negotiation and translation of QoS parameters.

QoS negotiation

Two parties are always encountered in the generic QoS negotiation process. The negotiation can be peer-to-peer (for example, application-to-application) or layer-to-layer (for example, application-to-system or human user-to-application). The peer-peer negotiation is also known as the caller-to-callee negotiation and the layer-to-layer negotiation is called the service user-to-service provider negotiation.

The purpose of the negotiation is (1) to establish common QoS parameter values among the service users and providers (that is, negotiate a contract) and (2) to capitalize scarce resource capacities by reserving only the real demand at any point in time.

The most significant variations of negotiation among the service users (caller/callee) and the service provider are bilateral peer-to-peer and layer-to-layer negotiations and triangular negotiation for a contractual value.

QoS translation

Since different NMS layers operate on different objects to provide or use services, they may require different QoS parameters. For example, the mean loss rate of packet networks has no meaning for a video capture device. Likewise, frame quality, the number of pixels in both axes used to initialize frame capture buffers, is of little use to network layer services.

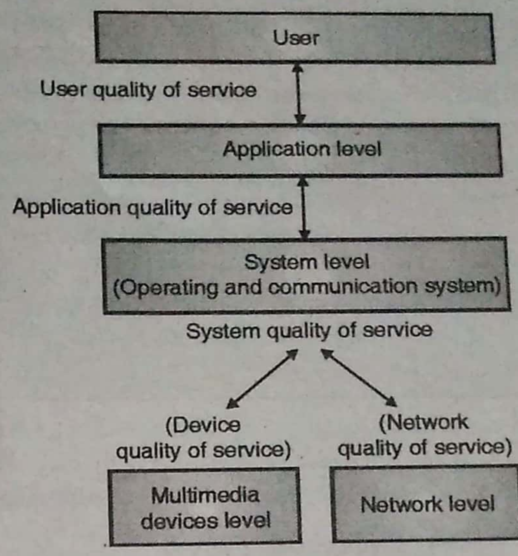


Fig. 5.2

Coordinating services in different NMS layers requires translating QoS parameters between layers. The translation often requires additional knowledge stored with the specific component.

Q. 2 Explain in detail Real-Time Transport Protocol.

Ans. : RTP - Real-Time Transport Protocol

The Real-time Transport Protocol (RTP) provides end-to-end delivery services for data with real-time characteristics, such as interactive audio and video or simulation data, over multicast or unicast network services. Applications typically run RTP on top of UDP to make use of its multiplexing and checksum services; both protocols contribute parts of the transport protocol functionality. However, RTP may be used with other suitable underlying network or transport protocols. RTP supports data transfer to multiple destinations using multicast distribution if provided by the underlying network.

RTP itself does not provide any mechanism to ensure timely delivery or provide other quality-of-service guarantees, but relies on lower-layer services to do so. It does not guarantee delivery or prevent out-of-order delivery, nor does it assume that the underlying network is reliable and delivers packets in sequence. The sequence numbers included in RTP allow the receiver to reconstruct the sender's packet sequence, but sequence numbers might also be used to determine the proper location of a packet, for example in video decoding, without necessarily decoding packets in sequence.

RTP consists of two closely-linked parts

The Real-time Transport Protocol (RTP), to carry data that has real-time properties. The RTP Control Protocol (RTCP), to monitor the quality of service and to convey information about the participants in an on-going session. The latter aspect of RTCP may be sufficient for "loosely controlled" sessions, i.e., where there is no explicit membership control and set-up, but it is not necessarily intended to support all of an application's control communication requirements.

V	P	X	CC	M	PT	Sequence Number
Timestamp						
Synchronization Source(SSRC)						
Content Source(CSRC) (0 -15)						

Fig. 5.3

Protocol Structure - RTP (Real-Time Transport Protocol)

1. V - Version. Identifies the RTP version.
2. P - Padding. When set, the packet contains one or more additional padding octets at the end which are not part of the payload.
3. X - Extension bit. When set, the fixed header is followed by exactly one header extension, with a defined format.
4. CSRC count - Contains the number of CSRC identifiers that follow the fixed header.
5. M - Marker. The interpretation of the marker is defined by a profile. It is intended to allow significant events such as frame boundaries to be marked in the packet stream.
6. Payload type - Identifies the format of the RTP payload and determines its interpretation by the application. A profile specifies a default static mapping of payload type codes to payload formats. Additional payload type codes may be defined dynamically through non-RTP means.
7. Sequence number - Increments by one for each RTP data packet sent, and may be used by the receiver to detect packet loss and to restore packet sequence.
8. Timestamp - Reflects the sampling instant of the first octet in the RTP data packet. The sampling instant must be derived from a clock that increments monotonically and linearly in time to allow synchronization and jitter calculations.
9. SSRC - Synchronization source. This identifier is chosen randomly, with the intent that no two synchronization sources within the same RTP session will have the same SSRC identifier.
10. CSRC - Contributing source identifiers list. Identifies the contributing sources for the payload contained in this packet.

Q. 4 Write short note on : Design Issues for Multimedia Authoring.

Ans. : Design Issues for Multimedia Authoring

Enterprise wide standards should be set up to ensure that the user requirements are fulfilled with good quality and made the objects transferable from one system to another. So standards must be set for a number of design issues :

1. Display Resolution
2. Data formula for capturing data
3. Compression algorithms
4. Network Interfaces
5. Storage Formats

Display Resolution

Display Resolution should be at various levels since user's requirement differs accordingly. For example, for television, people will watch it at distance of atleast 10 to 15 feet's away. But in the case of computer (i.e. Desktop), people will watch only at the distance of 12-18 inches away. So, display resolution of computer must be more than for TV. Set the display resolution as per user's requirement. In a large organization, there are a number of different display types with large variations in resolution and display technologies. Hence, the display resolution becomes an important consideration. A number of design issues must be considered for handling different display outputs.

They are :

Levels of standardization on display resolutions.

Display control standardization.

Corporate norms for service degradations.

Corporate norms for network traffic degradations as they relate to resolution issues.

Setting norms will be easy, if the number of different work station types, window managers, and monitor resolutions are limited in number.

File Format and Data Compression Issues

There are varieties of data formats available for image, audio and full motion video objects. Since the varieties are so large, controlling them becomes difficult. So, standardize on a single format. Instead, select a set for which reliable conversion application tools are available.

Another key design issue is to standardize on one or two compression formula for each type of data object. For example for facsimile machines, CCITT group 3 and 4 should be included in the selected standard. Similarly for full motion video, the selected standard should include MPEG and its derivatives such as MPEG 2.

Standardizing protocols in WAN's is essential to ensure that multimedia objects can be transformed from one LAN via a WAN to another LAN without loss of information. All hubs, routers, switches, etc. must be standardized.

Storing and retrieving of very large objects is an important consideration. Storing should be done quickly and retrieving should be done efficiently. While doing storage, it is useful to have some information about the object itself available outside the object to allow a user to decide if they need to access the object data.

Thus select a small set of file formats for which reliable conversion tools are available. Standardize on one or two compression method for each type of data objects, Consider what is the more convenient and efficient means of accessing large object, such as video clips

Service Degradation Policies

This issue concerns with what will happen if resources is insufficient either temporarily or locally. For example, in a distributed video application, video is sent to a remote client to be played back. If the network is overloaded, a number of policies may be possible :

Decline further requests with a message to try later and give proper reasons. Provide the playback service but at a lower resolution. Provide the playback service at full resolution but drop intermediate frames. Provide service at full resolution and frame rate in blocks.

Q. 4 What are Design Approach to Authoring Systems?

Ans. : Design Approach to Authoring Systems

Designing an authoring system spans a number of design issues. They include :

1. Hypermedia application design specifies.
2. User Interface aspects.
3. Embedding/linking streams of objects to a main document or presentation.
4. Storage of and access to multimedia objects.
5. Playing back combined streams in a synchronized manner.

A good user interface design is more important to the success of hypermedia applications. User interface presents a window to the user to control storage and retrieval, to insert objects in the documents and to specify the exact point of insertion, and to index marks for combining different multimedia streams and rules for playing them back. The authoring system must allow playing several streams in a coordinated manner to produce a final product.

Hypermedia Application Design Considerations

1. **Integration of applications :** The appearance of the applications and the ability of the applications to exchange data.
2. **Structuring the information :** Is to identify the information objects and to develop an information model to define the relationships among these objects.
3. **Object types and object hierarchies :** How the various attributes and representations of real-world objects are related.
4. **Object representation :** Consists of display/playback requirements and timing information for each object and its sub-objects.
5. **Object connections :** Describes the relationship between objects and helps to navigate among the objects.

User Interface Aspects

User interface design for multimedia application is more involved than for other application due to the number of types of interaction with the user. Consequently, rather than a simple user interface dialogue editor multimedia application needs to use four different kinds of user interface development tools.

Classification of these user interface development tools :

1. Media editors
2. An authoring application
3. Hypermedia object creation
4. Multimedia object locator and browser

A media editor is an application responsible for the creation and editing of a specific multimedia object such as an image, voice, or video object.

Common UI and Application Integration

In all tools we have buttons, scroll bars, toolbars, menus, file, edit, view, insert, format options. Additionally textbox, check box, text area, drop down menu are the main GUI components.

Navigation through the application

1. **Navigation :** Refers to the sequence in which the application progresses.
2. **Direct navigation :** Completely predefined.
3. **Free-form navigation :** The user determines the sequence of actions.
4. **Browse navigation :** The user is provided a large number of choices.

   An important aspect of any multimedia system is to maintain a clear perspective and the relationship between those objects.
5. **Designing user interface :** In designing GUI fist plan the overall structure of application, the its content and interactive behaviors and finally look and feel of application.
6. **Special metaphors for Multimedia application :** User interface metaphors are designs based on real world objects. For user interface metaphors that were not known to users, a new design took sometimes to take hold.

7. **The organiser metaphor :** Lotus Organiser was the first to use a screen representation of the ubiquitous office diary-type organizer. This is a clear example of a close adaptation of an existing user interface to a GUI.
8. **The telephone metaphor :** combines a well-known user interface with other GUI element to provide a more convenient means of access information and perform communication
9. **Aural user interface :** the real challenge in designing AUI systems is to create an aural desktop that substitutes voice and ear for the keyboard and display, and be able to mix and match them
10. **The VCR metaphor :** this is one of the most common user interfaces

Linking and Embedding : Definition

OLE is technology developed by Microsoft to make different programs interact with each other to make each program more useful and visually appealing. OLE object can be embedded into different application; they however retain original format and link to originating program. For example, an image can be embedded or linked into text document

Linking files. When link a file : The file remains in its original location.

Anything user change in the original file affects all the files to which it is linked. If, double-click a SolidWorks image in a Microsoft Word document, the SolidWorks software launches for user to edit the original file.

Embedding files. When you embed one file in another :

The original file becomes part of the file in which user embedded it. Any changes user make to the embedded SolidWorks document affect only that document. Any changes user make to the original SolidWorks file do not affect the document embedded in the Word file. If user double-click a SolidWorks image in a Word document, the SolidWorks software launches for user to edit the file.

Information Access

Access structure defines the way objects can be accessed and how navigation takes place through the information objects. The following describes common forms of navigation for information access.

1. **Direct** : This requires that the user has knowledge of the specific object that need to be accessed, e.g. object ID, object name.
2. **Indexed** : An attribute of an object, e.g., ID, may be used as an index. Indexed access may result in multiple copies of the same object.
3. **Random selection (browsing)** : The user can pick one of several possible items that are not arranged in any logical sequence.
4. **Path selection of guided tour** : The application guides the user through a predefined path across a number of objects.

Multimedia Objects Display and Playback Issues

Multimedia document contains several object such as Image, Audio, Text, Video. High volume multimedia document with multiple objects we can store in less space with the help of various compression methods. Whenever we are downloading such a document from network to PC hard-disk, there is need have a system compatibility for that particular multimedia document and because of that require display and playback issues for various objects on multimedia document.

Image Display Issue

Images on image server we are storing in compressed form and therefore it is necessary to decompresses the images properly on PC at the time of display / download.

1. **Scalling :** Image is called to fit in user defined window and for that proper scalling factor is required.
2. **Zooming :** Zooming shows the proper details regarding selected part of image user can zoom the image by selecting zooming factor.
3. **Rubber banding :** With the help of rubber banding user can cut the part image and after copying to clipboard user can zoom that portion of image.
4. **Panning :** Panning is used to find out details of image those are not visible in the full image.

Video Playback Issues

Whenever we are playing audio and video object, for that synchronization between audio and video component is necessary at the time of playbull.

1. **Video frame interleaving :** It defines layout of audio-video file playback control for video file can be exercised at the time of decompression.
2. **Program degradation :** Whenever, server is sending audio-video file to client workstation it will separate out both components at the time of streaming and then it will send but client workstation is unable to receive the data properly which is send by the server and this will results in program degradation.

Q. 5 Explain User Interface Design Issues.

Ans. :

User Interface Design Issues

While desigining user interface following design issues are considered.

Presentation design process should involve sequential flow of actions and parallel interactive action. That is there should be extensive feedback going on between the components making decision about media and modalities. As shown in Fig. 5.4 a conceptual architecture with knowledge base (lower part of Fig. 5.4) is used by an intelligent multimedia presentation system (upper part of Fig. 5.4).

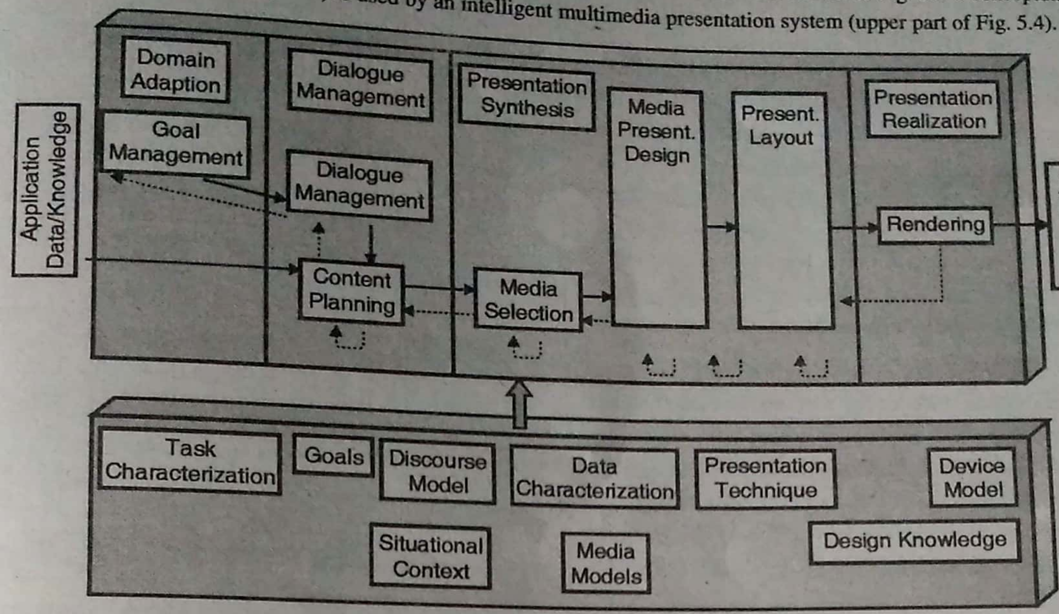


Fig. 5.4 : Architeture of multimedia presentation

Information Characteristics for Presentation

A complete set of nformation charactristics makes knowledge defination and representarion easier because it allows for appropriate mapping between information presentaton techniques.

The information charactristics specify following :

1. **Types (ordering information)**

Coordinates vs. amount (specify points in time, space or other domains). Intervals vs. ratio (suggest the type of comprisons meaningful among elements of co-ordinate and amount data types.)

2. **Relation Structures**

Functional dependencies (e.g. bar chart) Non-functional dependencies (e.g. entry in a relational darabase)

3. **Multi-domain Relations**

Multiple attributes of a single object set (e.g. position, colors ...) Multiple object sets (e.g. graphical symbols on a map) Multiple display (e.g. multiple windows)

4. **Large Data sets**

Numerous attributes of collecitons of heterogeneous objects (e.g presentation of semantic networks)

Presentation Design Knowledge

Issues like content selection, media and presentation technique selection, and presentation coordination. When creating a presentation, some design considerations should be followed. Issues like content selection, media and presentation technique selection, and presentation coordination must be considered. The effective application of these considerations will make presentations easier to follow and understand and will make the presenter appear more professional.

Content selection

In selecting content, consider a variety of information types : statistics, testimony, cases, illustrations, history, and particularly narratives that help convey the goal you have for your presentation. You will also want to choose information that will appeal to your audience particularly their attitudes, interests, biases, and prejudices about the topic.

Media selection

Multimedia is an excellent way to enhance a presentation. Graphics, sound, video, animation, and charts can all add to the message. Multimedia files should be kept small, since they will have to load into the presentation computer's RAM before being displayed and therefore may cause delays or pauses in the presentation. Additional media should be added to a presentation when they improve the quality, increase the impact of the message, or present information better than text alone.

Coordination

Multimedia presentation design involves more than just merging output in different media; it also requires a fine-grained coordination of different media.

This includes distributing information onto different generators, tailoring the generation results to each other, and integrating them into a multimedia output.

Thus coordination needs following mechanisms

Encoding techniques,

Presentation objects that represent facts.

Multiple display

Effective Human-Computer Interaction (HCI)

Human Computer Interaction (HCI) is the study of human interaction with computer software applications. Human-computer interaction is defined as "a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them". Human-computer interaction is the behavioral study of human factors based on the design of a UI within a computer application.

A good example of effective HCI can be seen in the current air traffic control systems. Users of these systems are responsible for ensuring plans do not collide with other plans in the sky. The user screen of the air traffic control system will change in appearance to alert controllers to problems. Typically, the color of objects will become brighter and bolder as the situation worsens. In addition, sounds can be added to draw the attention of the user to the problem.

Modern cell phones also include human-computer interaction design. Using the phone to find contacts, send messages, or retrieve voice mails requires a well- designed user interface. Typically, users of modern products expect a device to be easy to learn.

One of the most important issues regarding effective HCI is user – friendliness. Therefore during the design of effective of HCI the following main issues must be considered.

1. Context
2. Linkage to the world
3. Evaluation of the interface
4. Interactive capabilities
5. Separability of the user interface from the application.

Q. 6 Write short note on : Design Steps for Multimedia System.

Ans. :

Design Steps for Multimedia System

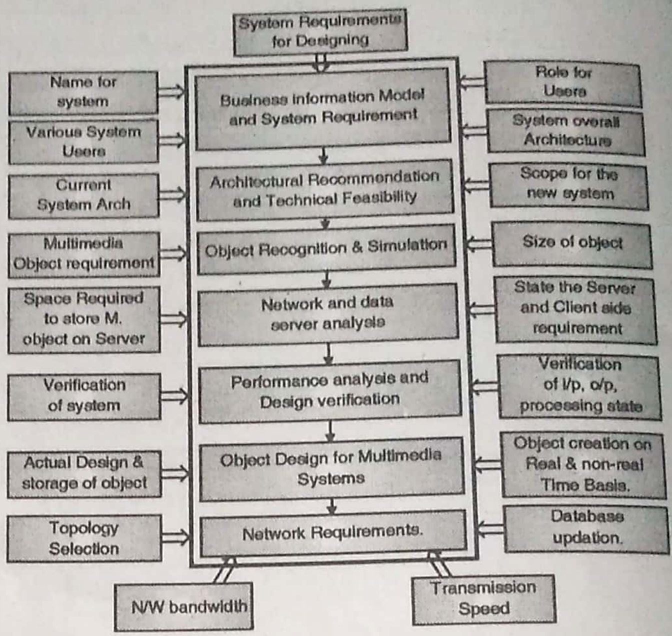


Fig. 5.5 : System design steps for multimedia system

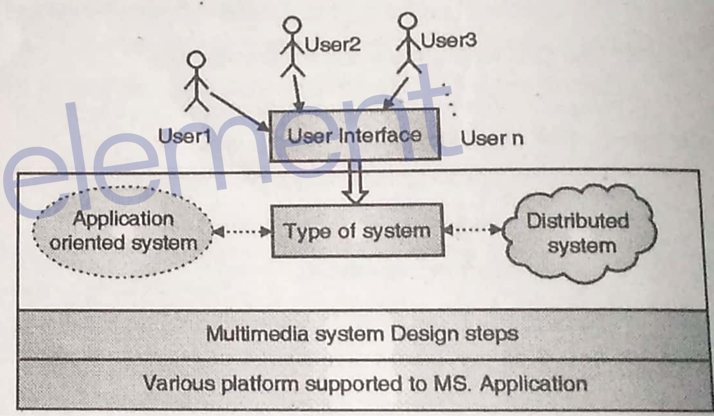


Fig. 5.6 : Overall system design concept with platform support, types of system, users, GUI and design constraint etc.

Fig. 5.6 describes essential design steps of our system design methodology for multimedia system design. Successful phased approach based on this system design methodology for advanced custom information system design consisting of the following steps in design process.

1. **Developing the project plan (Project organization and Management).**

   The key to efficient and reliable system design is well planning.

2. **Developing business information model and document system requirements.**

   The business information model describes the information objects that the business collects and stores, and details how these information objects are used. A number of factors contribute to success in developing a business information model that truly reflects the information system requirements of the business and is adaptable, as business needs change. User applications requirements can be derived from the business information model. The business information model becomes a critical document for current and future management of information needs for the corporation. It should be maintained as a living document that is maintained current so that applications can be enhanced as the model is updated in response to changes in business need.

3. **Performing technology assessment and preparing an architectural recommendation and technology feasibility report.**

It is important to understand the current architecture and feasibility of implementing the new system on this architecture. It is essential to determine how far the architecture can carry the applications, whether it is adaptable, if it can be made, an adjunct to the new technology, or it has to be replaced completely The architectural recommendation and technology feasibility report defines the proposed architecture, especially as it relates to the networked components of the system. Technology recommendations are based on this architecture and feasibility report. The report defines the major sources of information, where primary information stores should be located on the basis of usage patterns, and how information will be accessed, replicated, migrated, and used across the network. Most important, the hardware platforms and the target database system are determined. The feasibility analysis determines to what extend the realized functionalities can be achieved by expanding the current system as well as by replacing / enhancing it, and at what cost.

4. **Modelling database and application using simulation techniques and prototypes as needed.**

Prototype serves as the guide for the sequence of implementation of database servers, system software, and application components. Use of prototyping during design provides an interactive means for user to view the user interface in terms of content as well as the sequence of function and keystroke required to perform these function.

5. **At this stage there is enough information to determination the data server needs and types of networks available to serve them.**

Issues such as what kind of performance is acceptable to users (data sever need), methods for making a slow system appear as if the system is responding in an acceptable manner (network) and user interface issues to address user concerns during slower activities are discussed.

6. **Performance analysis and design verification at this stage goes a long way in ensuring that designed system will meet its goals and performs at expected levels.**

Performance analysis is a careful study of all functions performed by all users on a network, their work patterns, the volumes and types of data accessed by the users, and the relative frequencies for accessing the network. Performance monitoring and tuning have always been complex. Benchmarks have been used as a real quantitative means of measuring performance. The application design is using user interface features to hide performance inadequacy. While there is no real improvement in performance, some operations are being performed in parallel, and the attempt is to give the user a perception of higher performance.

Any potential new system or application software must be analyzed to determine performance and throughput requirements and the hardware and software required to provide that performance. If the proposed or available hardware and software configuration, including LAN and WAN, is unable to provide the required performance at peak loads, the degradation of performance against load must be determined.

7. **Designing information system that is all data object, multimedia objects and services objects can be designed at this stage.**

Information system model consist of models of function and data element that present the user level perspective i.e. what the user sees from a perspective of function and data element entered, displayed and reported for each function. Information system model presents a description of information system in a pictorial and intelligible manner.

8. **Design all object servers and network protocol for each segment of network.**

Network transportation is an important performance issue. Overloaded networks deteriorate in performance very rapidly. An important aspect of network management is predicting when and how networks become overloaded. Constant network monitoring is essential to determine network-loading patterns. Network monitoring is also useful in setting up routing algorithms to balance network loads. An important component of routing network traffic is the location of objects.

Q. 7 Continuous education program : one of the most important application using both technologies, network and multimedia, is distance learning, computer offer the chance for new educational procedures, which is combination to network reach levels beyond imagination. Anyone could be educated by greatest teachers of the world. Computer cannot replace the teacher, but they can bring him closer to the student. Assume required data if any and specify clearly. You are appointed as consultant to implement this application.

(1) **Design performance requirements if this application is to be used in distributed environment.**

(2) **Give the workflow design.**

(3) **Model various objects and design special multimedia user interface.**

OR

Design Web pages for e-learning application in detail.

Ans. :

Step 1 : Business model :

E-learning, distance learning , online-campus.

Provide virtual learning.

Resolve space and time gap between teacher and student.

Step 2 : Business information model :

(a) **Application user**

1. Teacher
2. Student

(b) **Role of users**

1. Student
2. Browsing
3. Register
4. Fetch information
5. Download white papers
6. Choose course and spiritualised subjects.
7. Pay online
8. Enroll for multiple courses.

(c) **Performance Requirement**

Assumptions

(1) **Online university offers 20 courses / subjects / technical / literature based enrollments.**

400 students takes there tutorial daily. Every tutorial has a interactive class room session objects based test papers, containing video clips, audio to support lecture, text to support literature, about current topic.

(2) **Amount of space required for object.**

Assume that there are 10 users are active at any time for online users.

(a) **Video content in tutorial (Video object)**

Frame $= 320 \times 200$ @ 30 frames / sec. and 8 bit color.

$= 1.92$ MB / sec.

For 30 m of lecture (in video)

$= 1.92 \times 30 \times 60 = 3.456$ GB

If compression ratio is 30 : 1

Space required / user $= \dfrac{3.456 \text{ GB}}{30} = 115.2$ MB

For 10 users space required $= 115.2 \times 10$ MB $= 11523$ MB

Network bandwidth required

For compressed video (lectures) for 10

Users $=$ 1152 MB / sec.

(b) **Audio content in tutorial (Audio Object)**

For 30 min lecture, 8 kHz / sample , 8 bits/sample

$= 8$ KB/sec $\times 30 \times 60 = 14.4$ MB

Space required for 1 session $= 14.4$ MB

Space required for 10 active users $= 14.4$ MB $\times 10 = 144$ MB

Networks bandwidth (Download BW) required if compressed with, 30 : 1 Ratio

$= \dfrac{144 \text{ MB}}{30} = 4.8$ MB

bandwidth required for compressed audio = 4.8 MB/sec.

(c) **Images Object**

Average compressed image = 30 KB

(Assume 10 images / session downloads) = 30 MB × 10 = 0.3 MB

Compression (20 : 1) $= \dfrac{0.3 \text{ MB}}{20} =$ **15 MB**

For 10 users $= 15$ KB $\times 10 =$ **150 KB**

Network bandwidth $=$ 150 kB/sec. (download BW)

∴ **Total space required on content server end.**

$=$ 1152 MB + 144 MB + 0.15 MB = **1296.15 MB**

∴ Total bandwidth to be supported by WAN connectivity.

$=$ 1296.15 MB/sec.

Therefore 1 GB buffer (working memory) space of server end sufficient enough to cater 6 users taking virtual classroom coaching.

Total bandwidth (Download) to be supported by WAN connecting to content server.

$=$ 1296.15 MB/sec.

But FDM / TDM (Multiplexing with degree 10)

$=$ B.W. requirements can be reduced .

$= \dfrac{1296.15}{10} = 129.6$ MB / sec. $\approx$ 130 MB/sec.

Step 3 : Architectural recommendation to support above performance requirements :

(a) Topological Layout

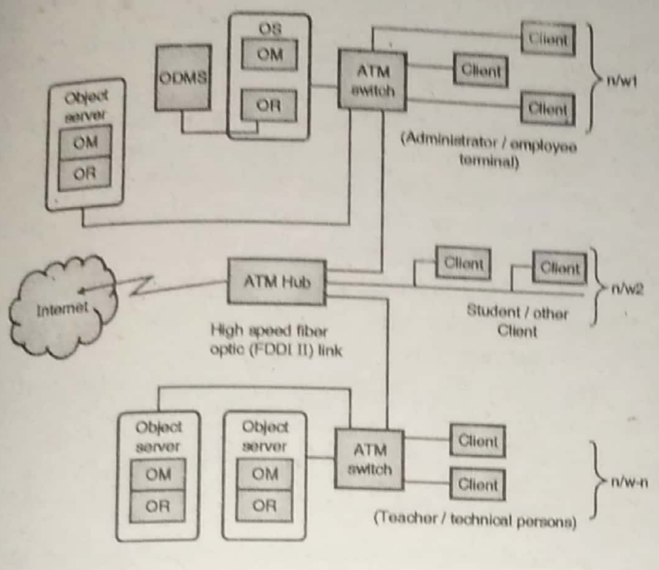


Fig. 5.7 : Teacher / technical WS Terminals at co. premises

Step 4 : Performance attribute achieved :

Good response time.

Quick and efficient search delete to indexing.

Good availability and failsafe operation as object server replication done.

User friendliness.

Step 5 : Workflow design :

Capture stage : The video conversation or a stored video mail message are recorded as part of doc or database record. Captured information is recorded directly on an object server for multi objects .

User access stage

Archiving and purging

I/P and O/P component of workflow are completely independent of each other and may or may not be reside on same network.

Size of DB is large and need for Indexing every component in an information repository is a must.

Various objects which will be used for distance learning will be :

(1) **Text** : Lecture notes, tutorials, textbook, references, technical papers, journals, manuals.

(2) **Images** : Book cover (covering name and author), Lecturer image, flash, images at the start and end of lecture, image of student, images of student-lecture interaction, various images related to lecture subject like diagrams, drawing, map etc.

(3) **Audio** : The lectures talking, student questions, sound of flash images indicating start and end of lectures.

(4) **Video** : Video conferencing overall, display of some classroom virtual tour.

Step 6 :

a) User interface for home page

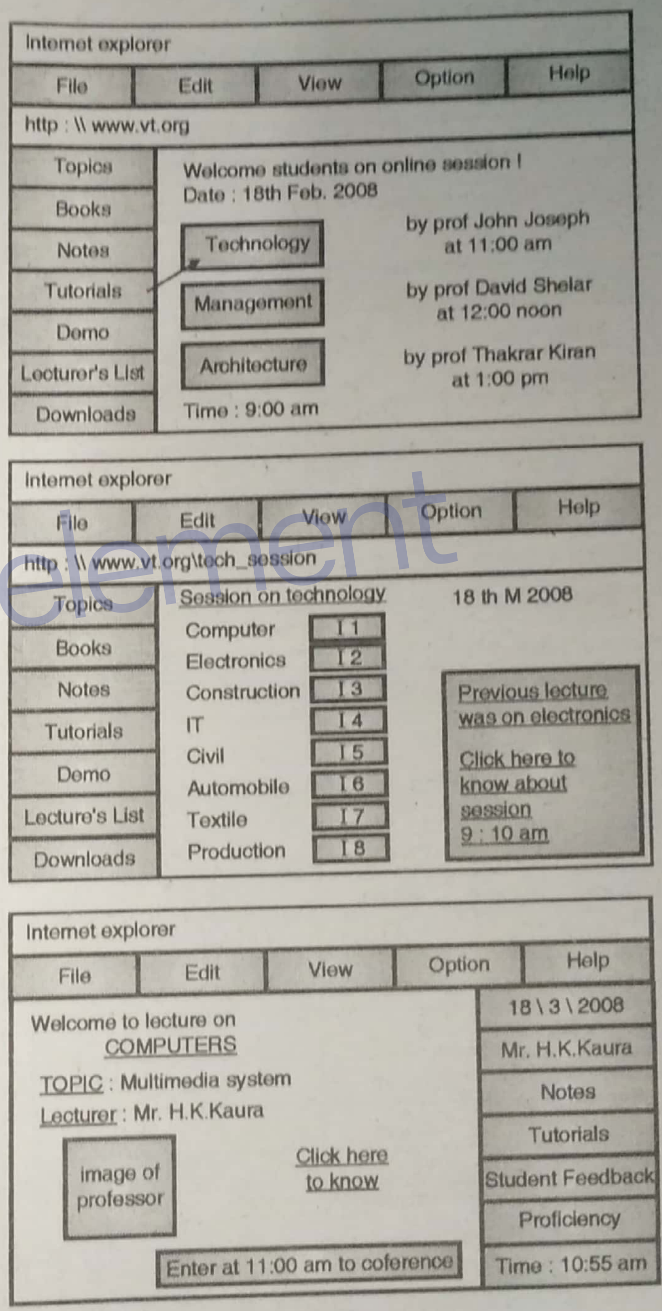


Fig. 5.8 : GUI for a particular course

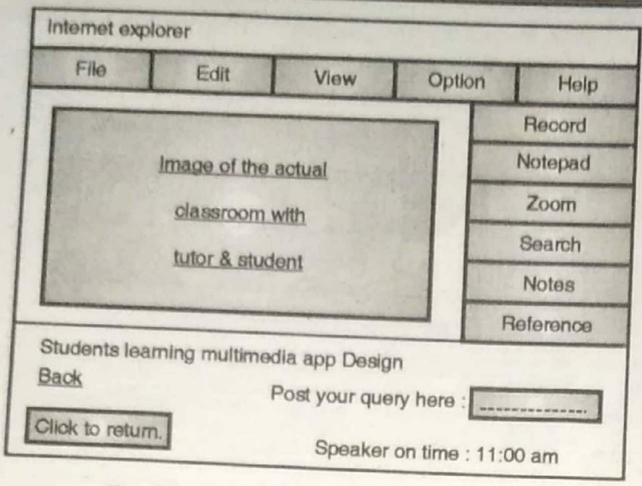


Fig. 5.9 : GUI of actual classroom session

Step 7 : **Schema Design**

a) **Student _ db**

Attribute	Type	Remark
Student _ id	Numeric	pK
Student _ name	String	
Student _ address	String	
DOB	Date	
Course	String	
Previous degree	String	
Photo/Image	Image	[DIB , bmp, GIF]
Date of Admission	Date	
Credit card no	Numeric [encrypted]	Comp. Algo.

b) **Teacher _ db**

Attribute	Type	Remark
Teacher _ id	Numeric	pK
Teacher _name	String	
Address	String	
Qualification	String	
Trade	String	
Photo/Image	Image	[DIB , BMP, GIF]
Session detail	Time	
Subjects taught	String	

c) **Course _ offered**

Attribute	Type	Remark
Course _ id	Numeric	pK
Course _name	String	
No. of subjects	Numeric	
Related Teacher	String	

duration	Numeric	
Session Time	Time	
Date of admission	Date	
Fee_	Numeric	
Last date of payment	Date	

Q. 8 "Dictionary of the living word" is an ideal type of multimedia presentation using at maximum text, sound, images and video. You are appointed as a consultant to implement this application in distributed environment.
Assume required data if any and specify clearly.
(a) How can you manage multimedia object servers ?
(b) Give the design for managing distributed objects ?

Soln. :

Step 1 : **Business model :**

This is collection of words with 3 associated meanings / picture / text explaination and image including sound and video. This is not possible even with a (Hard copy) version. Integration with number of technical / non-technical / literature based information or tutorial sites.

Step 2 : **Business-information model :**

(a) Types of user

(1) Browsers/client.

(2) **Administrator :** Who will update, querying. Modify data regularly.

(3) Other application linked to 'online dictionary.'

(4) Application embedding (As OLE) the "online dictionary".

(b) Availability

Should be good as user may access any time. The only revenue possible is through advertisement, therefore should always be up and running.

(c) Performance requirements

Assume that there are 10000 bits/day as each term is explained with help of text + image format on clicking the icon a short video + audio clip (less than 2 min) is played demonstrating usage/reference of the same.

Storage space requirement at content server.

(a) Video content [for max 2 min clip]

Frame = 640 × 480 @ 30 frames /sec, 8-bit color

Resolution = 73.728 MB

Compression 30 : 1 = 2.4576 MB

Assuming that there are 10 active user / application Browsing, Integrating, linking with online dictionary then buffering required at playback server.

For 10 active users = 10×2.4576 MB

= 24.576 MB

Download b/w required to be supported by ISP. (Internet service provider) 25 MB/sec + (20%)

= 30 MB/sec (for peak loads)

Audio content [For max 2 min clip]

8 kHz/sample, 8 bits/sample

$8 \text{ KB} \times 2 \times 60 = 960$ KB

Storage required for 10 users

= 9.6 MB

If Compressed (20 : 1) = $\frac{9.6 \text{ MB}}{20} = 0.48 \text{ MB} = 480$ KB

Network b/w = 480 KB + 20%

= 576 KB/sec [At peak load]

(d) Architectural recommendation and technological feasibility

(1) Topological layout

Horizontally distributed and replicated web servers for parallel processing of incoming reg.

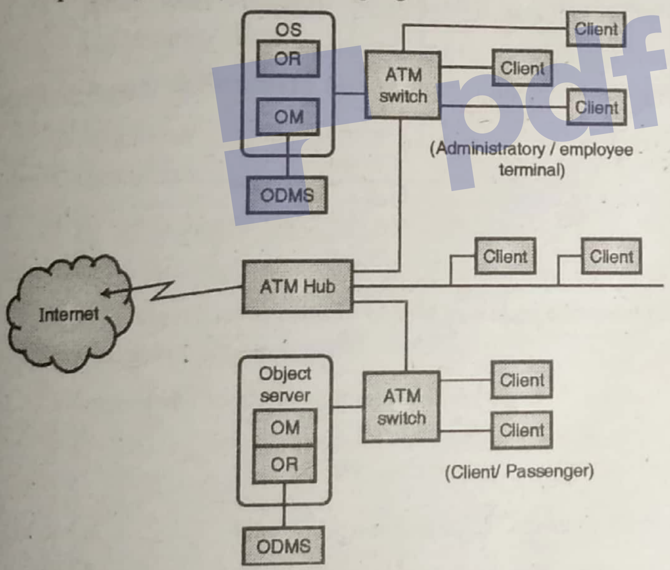


Fig. 5.10 : Topological Layout

Step 3 : Design issues for managing distributed object :

(1) How objects are located.

(2) How to retrieved and manage it in multiuser environment.

(3) Object replication.

(4) Object distributor.

(5) Object recompilation.

Following lists the type of communication that one server may make to another server. Search object class directory for current location of object. Replicate copy of object, update object name server directory. Pause relieved object.

Step 4 : The architecture above contains following key elements needed to manage the distributed object.

(1) Common object management application.

(2) ORB

(3) Object name server.

(4) Object directory manager.

(5) Object server.

(6) Object manager.

(7) Network manager.

(8) Object-Data store.

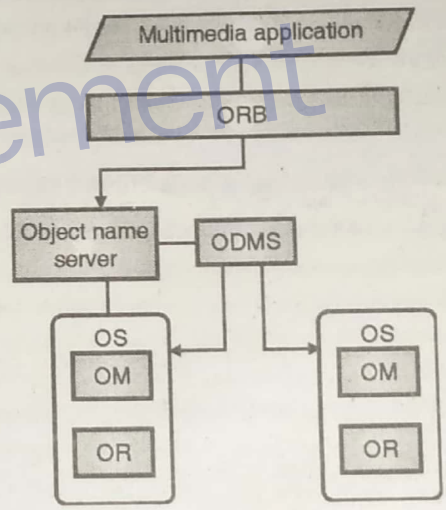


Fig. 5.11 : Object Server

Chapter 6 : Multimedia Security

Q. 1 Explain requirements of multimedia system.

Ans. :

Requirements of multimedia system

Some of the important security requirements that have to be implemented in ensuring an approximate fool-proof security are :

Confidentiality :

This requirement emphasizes the permit of only authorized access to the document or content and prevents the unauthorized access of resources. Cryptographic mechanisms are used to prevent the unauthorized access of resources. Private-key and public key crypto systems are use to achieve confidentiality. However, after decryption, unauthorized access cannot be prevented.

Data Integrity :

It is required by this security requirement that the data be identically maintained from its source to destiny, and has not been accidentally or maliciously modified, altered, or destroyed, and remain unchanged right throughout the operations such as transfer, storage, and retrieval.

The modification of the information can be detected by using one-way hashing function and digital watermarking. In one-way hashing function encrypted information can be easily calculated from the given input by applying a hash-function. But in reality, it is virtually not possible to compute the actual information from the encrypted one.

Also, the hash function is non-inverseable and hence this combined with other techniques can be used for maintaining data integrity. In digital watermarking technique, the embedded message is robust to alterations, i.e. even if the original document is altered; the watermark embedded in the document remains as such. If it is altered, then it will affect the actual information also.

Data Origin Authenticity :

When a document is found, the origin of that resource should be traceable. The origin of the resource can be traceable by using message authentication code, digital signature and watermarking, from which the exact identity of the person, to whom the document belongs, can be identified.

Entity Authenticity :

Entities participating in the communication should prove that they are the one they claim to be. This proves that the communication is carried out between the correct entities. response protocol is used to ensure that the communication is carried out between the correct entities. In this protocol, the time-variant challenge is provided to the participating entity, which proves its identity by giving proper response using the secrete key associated that entity.

Non-Repudiation It should be possible to detect and prove the rightful ownership of that document. Many authors are worried about distributing their works in fear that it may be copied illegally or represented as another's work. Non-repudiation facilitates the identification of the end users who have copied the document.

The rightful ownership of that document should be detected and proven. Non-repudiation facilitates the identification of the end users who have copied the document. The rightful ownership of that document can be detected and proven.

Q. 2 Explain the working of digital signature.

Ans. : Digital Signatures

A digital signature is basically a way to ensure that an electronic document (e-mail, spreadsheet, text file, etc.) is authentic. Authentic means that you know who created the document and you know that it has not been altered in any way since that person created it.

A Digital Signature is a type of signature, but the only difference is that it involves the use of mathematical pin or algorithm to sign and validate the authenticity of a document, file or software instead of pen and paper. Digital signatures rely on certain types of encryption to ensure authentication.

A digital signature is used to make sure that the file(s) sent digitally belongs to a designated source and reaches the intended receiver in its original format without any tampering. In simple terms, a digital signature works in the same way as an envelope seal does.

Imagine wanting to send a physically signed document from one country to another. User would need to send the documents by means of a courier. This process involves loads of paperwork and thereby wasting invaluable time. Instead, if user had just used a digital signature, the documents could have been sent electronically in a matter of minutes. This way user can save time as well as money. Numerous studies conducted around the world show that using digital signatures can save a whole working week for any working professional. The time saved combined with the undeniable savings in money is surely going to fuel the rapid acceptance of digital signatures around the world.

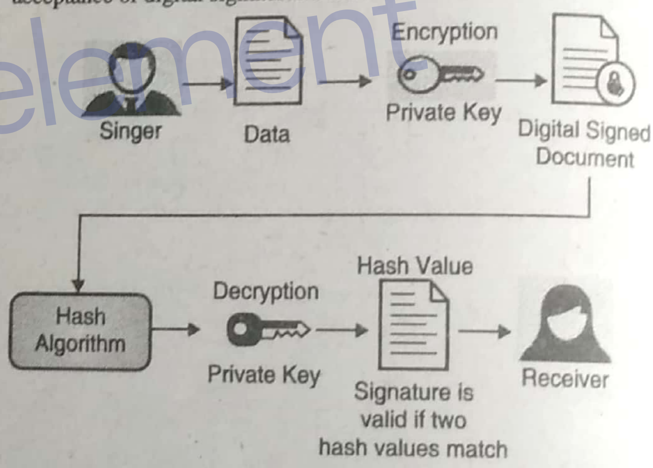


Fig. 6.1

How does a Digital Signature Work?

Digital signatures are based on Public Key infrastructure. By this mechanism, two keys are generated, a Public Key and Private Key. The private key is kept by the signer and it should be kept securely. On the other hand, the receiver must have the public key to decrypt the message.

For example, a person named A wants to send an encrypted message to B. As stated above, A must have a private key to sign the message digitally.

Before encrypting the message using the private key, an algorithm named 'MD algorithm' encrypts the message to be sent by A into a 128/256-bit format known as a hash value. Then A's private key encrypts this hash value. On completion of both the processes, A's message is said to be digitally signed.

On the side of B, the digitally signed message is decrypted with the help of the signer's public key. The public key decrypts the message and converts it into another hash value. Then the program which is used to open the message (e.g., MS Word, Adobe Reader etc.) compares this hash value to the original hash value which was generated on A's side. If the hash value on B's side matches with the hash value generated on A's side, then the program will allow the message to open up and displays the message "The document has not been modified since this signature was applied." The program will not allow the document to open if both the hash values don't match.

Q. 3 Write short note on : Types of Steganography.

Ans. :

Types of Steganography

The process of steganography technique can be defined into 4 categories such as Text, Image, Audio and Video. In text steganography text files are used to hide data. Here in this system the confidential data can be of any format like text, audio, video or image.

1. Image Steganography

The image steganography is used to hide a secret message inside an image. The most widely used technique to hide secret bit inside the LSB of the cover image. Because this method uses bits of each pixel in the image, it is necessary to use a lossless compression format, otherwise the hidden information will get lost in the transformations of a lossy compression algorithm. When using a 24 bit color image, a bit of each of the red, green and blue color components can be used, so a total of 3 bits can be stored in each pixel in this way we can use more secret bit to hide data in it.

In image steganography image files like BMP, PNG, JPEG, TIFF, GIF etc. are used to hide data. Technique used to achieve confidentiality is LSB, spread spectrum etc. In Spatial Domain Embedding steganography algorithm is based on modifying the least significant bit layer of images. This technique makes use of the fact that the least significant bits in an image could be thought of random noise and changes to them would not have any effect on the image.

The message bits are permuted before embedding, this has the effect of distributing the bits evenly, thus on average only half of the LSB will be modified. Popular steganographic tools based on LSB embedding vary in their approach for hiding information. Some algorithms change LSB of pixels visited in a random walk, others modify pixels in certain areas of images, or instead of just changing the last bit they increment or decrement the pixel value.

2. Audio Steganography

The sender embeds secret data of any type using a key in a digital cover file to produce a stego file, in such a way that an observer cannot detect the existence of the hidden message. In many schemes a method of audio steganography based on modification of least significant bits (LSB) the audio samples in the temporal domain or transform domain have been proposed.

Some of these methods employ LSB technique and combine it with other techniques such as error diffusion, minimum error replacement (MER) and temporal masking effect. Another method embeds covert messages in the LSB of wavelet transform and recently in the integer transform. The main objectives of the LSB based schemes are to raise payload or the maximum amount of the information to be embedded and to prevent audio quality degradation. The best format is wave format since the reading of the bits of data is easier in wave file, also the compression is good in wave files and the distortion of data is very less in wave files.

In Phase coding technique the technique encodes the message bits as phase shifts in the phase spectrum of a digital signal, achieving an inaudible encoding in terms of signal-to-perceived noise ratio. Phase coding relies on the fact that the phase components of sound are not as perceptible to the human ear as noise is. The basic Spread Spectrum (SS) method attempts to spread secret information across the audio signal's frequency spectrum as much as possible. The SS method spreads the secret message over the sound file's frequency spectrum, using a code that is independent of the actual signal.

3. Video Steganography

Video files are generally a collection of images and sounds, so most of the presented techniques on images and audio can be applied to video files too. The great advantages of video are the large amount of data that can be hidden inside and the fact that it is a moving stream of images and sounds. The video steganography is nothing but a combination of image and audio steganography.

Q. 4 Explain in detail Distributed Multimedia Systems.

Ans. :

Distributed Multimedia Systems

A distributed multimedia system comprises several types of components, such as media servers, meta-databases, proxies, routers and clients. In addition, a large number of adaptation possibilities exist, from simple frame dropping up to virtual server systems that dynamically allocate new resources on demand. The main problem is determining which kind of component can best be used for each kind of adaptation.

The following non-exhaustive list gives an overview of the tools and their major features :

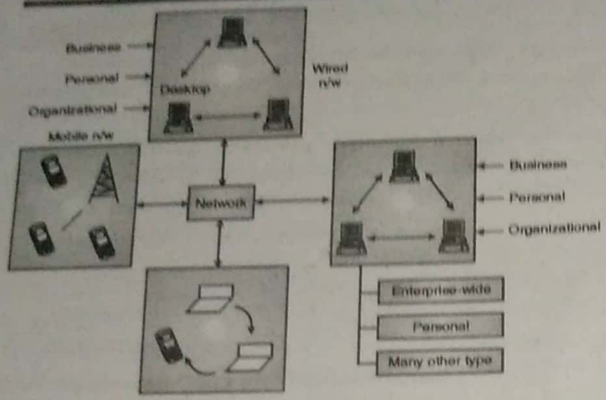


Fig. 6.2 : Distributed multimedia system

1. Media Server

Standard compliant media streaming by using RTSP and RTP/UDP. Communicates terminal capabilities of the client device and user preferences using standardized MPEG-21 descriptors. Supports real-time adaptation of media content according to the clients' terminal capabilities, the user preferences, and the available network resources; for example, mobile devices get a lower stream quality than high-performance workstations with good network access.

2. Proxy Server

Incorporates both a server and a client implementation (since a proxy must act as a server to the client and a client to the server). Caches elementary streams in different quality versions. Implements quality-aware replacement strategies. Can be dynamically relocated in the vicinity of requesting clients.

3. Meta-database

Multimedia database schema based on the MPEG-7 standard. Multimedia indexing framework. Cost-based query optimization for range and k-nearest neighbour searches. Application-level libraries for content-based image retrieval systems, audio recognition tools, video browsing tools, and quality aware MPEG-4 proxies.

4. Media Player

Standard compliant control of RTP-based media streams by using RTSP. Supports parallel presentation of many videos in different viewers, in different qualities.

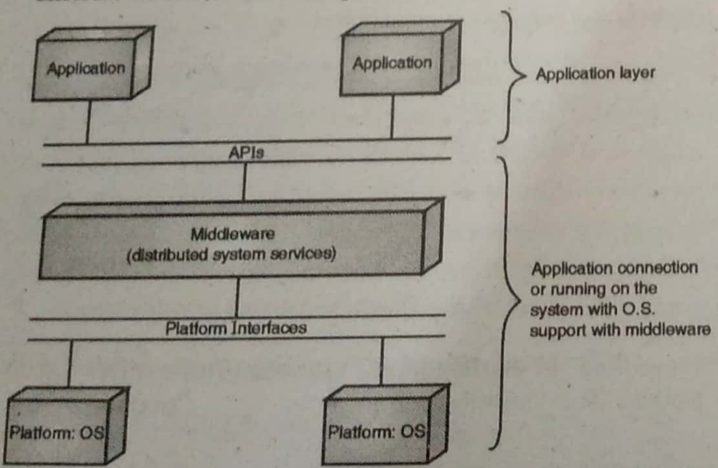


Fig. 6.3 : Various layer of DMS

Components of a Distributed Multimedia System

Distributed multimedia system consists of three different basic components: an Information server, a wide area network and a multimedia client on the user site. The user interface or the multimedia client deals with the issues related to presentation and manipulation of multimedia objects and the interaction with the user.

Multimedia client consists of a computer with a special hardware such as a microphone, high-resolution graphics display, stereo speakers, and a network interface. The user interacts with the system via a computer keyboard, mouse or a hand held remote control. The network provides the communication mechanism between the user and the server. The multimedia traffic requires transfer of large volumes of data at very high speeds, even when the data is compressed. Continuous media as video and audio require guarantees of minimum bandwidth and maximum end-to-end delay (jitter).

The server is responsible for managing multimedia databases and also composing general multimedia objects for the user. The composition of the object is a complex process of integrating and synchronizing multimedia data for transport, display and manipulation. The system usually consists of multiple users, servers and networks as shown below :

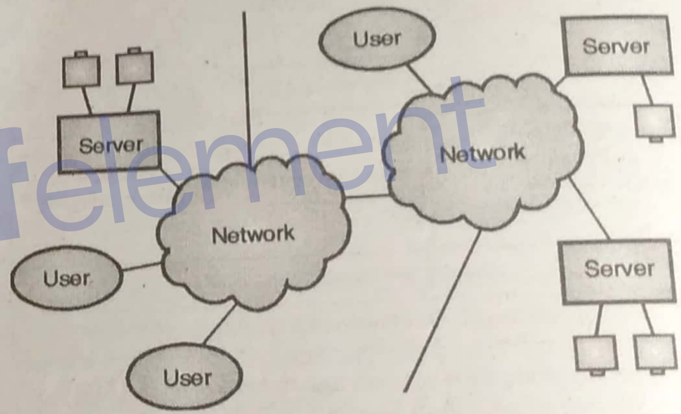


Fig. 6.4 : Component of distributed multimedia system

Now let us discuss each of these components in some detail.

i. User Terminal

A Multimedia terminal consists of a computer with a special hardware such as a microphone, high-resolution graphics display, stereo speakers, and a network interface. The user interacts with the system via a computer keyboard, mouse or a hand held remote control. Many of the user terminals still resemble traditional computers. Because of this, additional development work is required before the terminals can meet the requirements of the multimedia data and the user.

Because of the large size of the multimedia objects and real - time requirements the multimedia terminal or the network should include large data buffers. To restore the temporal relationship of a data stream, stream handlers should be connected to the data buffers. To synchronize the possible multiple data streams and to control the stream handlers, a synchronization and streaming manager is required. Since multimedia data objects are large, the terminal should also include compression and decompression hardware.

ii. Network and Communication

Multimedia communication differs from the traditional communication. The multimedia traffic requires transfer of large volumes of data at very high speeds, even when the data is compressed.

Especially for interactive multimedia communication the network must provide low latency. Continuous media as video and audio require guarantees of minimum bandwidth and maximum end-to-end delay. The variation in delay referred to as jitter, and loss of data must also be bound.

iii. Multimedia Server

Current personal computers, workstations and servers are designed to handle traditional forms of data. Their performance is optimized for a scientific or transaction – oriented type of workload. These systems do not perform well for multimedia data, requiring fast data retrieval and guaranteed real time capabilities. The I/O capacity is usually a severe bottleneck.

Q. 5 Write short note on : Virtual Reality.

Ans. :

Virtual Reality

Is a term that applies to computer-simulated environments that can simulate physical presence in places in the real world, as well as in imaginary worlds. Most current virtual reality environments are primarily visual experiences, displayed either on a computer screen or through special stereoscopic displays, but some simulations include additional sensory information, such as sound through speakers or headphones.

Some advanced, haptic systems now include tactile information, generally known as force feedback, in medical and gaming applications. Furthermore, virtual reality covers remote communication environments which provide virtual presence of users with the concepts of telepresence and telexistence.

Users can interact with a virtual environment or a Virtual Artifact (VA) either through the use of standard input devices such as a keyboard and mouse, or through multimodal devices such as a wired glove, the Polhemus, and omnidirectional treadmills.

Virtual reality is often used to describe a wide variety of applications commonly associated with immersive, highly visual, 3D environments. The development of CAD software, graphics hardware acceleration, head mounted displays, database gloves, and miniaturization have helped popularize the notion.

Virtual Reality data

VR addresses the construction of artificial worlds, with clear spatial dimensions databases for VR can structure and store data using methods beyond the conventional abstractions of GIS.

Virtual Reality tools

Under computer control allowing access to the artificial worlds with internet viewers, VR navigators and dedicated stand-alone hardware stations. The hardware components of a multimedia and/or a Virtual Reality PC or workstation Multimedia requires perception and interaction with use of visual and auditory participation, i.e. the production of vision and sound, Virtual Reality additionally requires tactile and vestibular participation.

A Virtual Reality system may be considered to be an expansion of a multimedia system into a multi-sensory system.

The additional components of a Virtual Reality PC or workstation may include any of the following :

Tactile interaction

Head Mounted Display (HMD) - wide field of view, an amoriphically projected stereo. Tactile feedback devices, vibrotactile displays - teletactile feedback glove, virtual joystick.

Force feedback

Teleoperation systems - force feedback joystick, remote manipulator arm, joystring.

Vestibular

Motion platforms - flight simulators, motion simulators.

Other interactive devices

2 degrees of freedom (DOF) - mouse, joystick, 2-d tablet with gesture recognition, touch screen. 6 degrees of freedom - wand, 6 DOF mouse, data glove, force ball. Wired clothing - datasuit. Biological input (biosensor) - voice recogniser, skin temperature probe, myoelectric (muscle) sensor, cerebroelectric (brain) sensor.

Applications

VR can be used in a GIS in two ways :

A tool for purely viewing three dimensional models of data. This can be purely in an office situation or in the field overlaying three dimensional data on top of real world data. Applications of the latter in underground pipe work, user can 'see' network under their feet. The whole user interface to the GIS dataset, allowing for the display of VR, MM and standard data in three dimensions. This would envolve the creation of a virtual interface. Possibility of viewing any data easily from any angle.

Education

Self-led interaction with the real world, especially for children. Introducing geographical concepts, displaying distant 'realities' possibilities, using MM and VR, of the 'virtual fieldtrip'. Use of MM for local studies and global geography knowledge building, whilst integrating with other National Curriculum subjects such as history, economics, biology, geology and information technology.

Scientific research

Creating three and four dimensional views of spatial data. Preliminary views of integrated data sets prior to verification of data linkages and casualties MM integration and overlay of datasets, for example; vector data with attribute information on raster satellite imagery. Physical geography data, e.g. meteorological, geological, oceanographic data ideally suited to its four dimensional nature, VR applications include environmental monitoring, hazard and risk assessment, atmospheric modelling, planning and forecasting, pollution analysis, terrain visualisation, multi-variate analysis.

Military for training purposes and scenario building, particularly VR representations of terrain. Entertainment improving realism of interaction with spatial data. Built environment VR applications in architectural simulation, urban planning, resource modelling. Archival of geographic information MM storage of the disparate range of data which can convey geographical information.