B. E.

Fourth Semester Examination, May-2009

MULTIMEDIA TECHNOLOGY

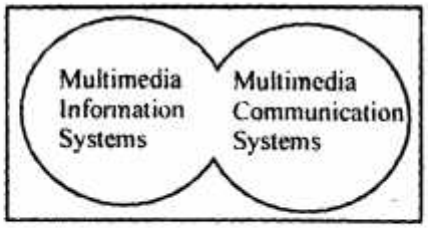
Note : Attempt any five questions. All questions carry equal marks.

Q. 1. What is multimedia? What are various requirements of creating a framework for multimedia systems?

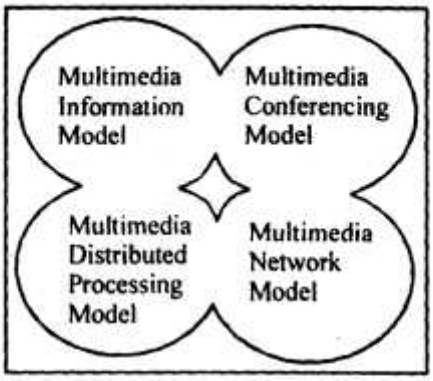
Ans. It is the simultaneous use of data in different media forces (voice, video, text, animations etc.), that is called multimedia. Digital video & audio media are the most demanding of the new media that are being added to the repertoire of computing & communication systems. Because of their time sampled nature, these types of media are frequently referred to as continuous media (M). The term multimedia computing commonly refers to the use of multimedia data types in computer application systems, & multimedia communications denotes communication systems, which supports the real-time transmission of continuous media.

The framework provides an overall picture of the development of distributed multimedia systems from which a system architecture can be developed. The framework highlights the dominant feature of multimedia systems, the integration of multimedia computing & communications, including traditional telecommunications & telephony functions.

Multimedia technology is facilitating the convergence of multimedia information processing systems & multimedia communication systems. The framework consists of four interrelated models. The information & distributed processing models constitute the Multimedia Information (MMIS). The conferencing & multiservice network models form the multimedia communication system (MCS).



(a)



(b)

Aspects of each model are discussed as follows :

(a) Multimedia Distributed Processing :

Model :

This model is a component of multimedia information system (MMIS) it includes system services, application tool kits & application frameworks. This model presents a layered view of a distributed environment where each layer provides services to the layers above significant application/layers are :

- Toolkits : GUI, Media/Service Control, Interchange, Object Access.

- Orchestration :

Admissibility, multichannel sync, meta scheduling.

- Media Service :

Object management, Device server, conferencing services, window server.

- System Services :

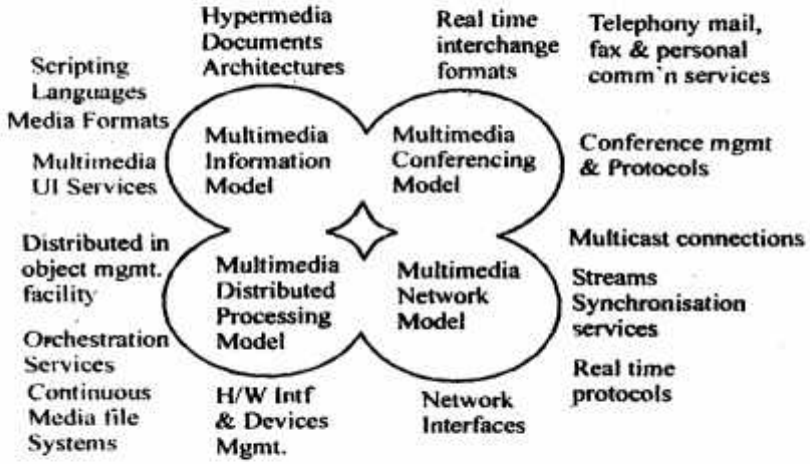
Continuous media file system, real time scheduling.

- **Hardware Services :**

Network Adapters, CODECs.

(b) Multimedia Information Model :

It includes data modelling for storage, retrieval & processing. This component relates to the abstractions & data models for organising multimedia documents, presentations & other information while there is no universal hypermedia document model, the ability to include multimedia content & describe association between document components (hyperlinking) is considered basic.



Framework

(c) Multiservice Network Model :

This model supports the communication model with the network architecture, network protocols & interfaces. Networks for distributed multimedia systems must support a wide range of traffic requirements, including traffic with real-time requirements. Such networks are described as multiservice. The requirements for network architecture include QOS guarantees that are sufficient for real time transport, multiway connections, & high performance.

(d) Multimedia Conferencing Model :

This model provides abstractions for multiparty conferencing. real-time interchange, electronic mail & telephony. Existing network architectures such as the OSI reference model & TCP/IP were not designed with the intent of supporting real time multiparty conferencing.

Q. 2. What is animation? What are various techniques associated with animation. What are various software tools available for animation in market today?

Ans. Animation makes static presentations come alive. It adds visual effect to our multimedia projects 7 webpages. Animation is more than wipes, fades & 200 ws.

Animation is an object moving across or into or out of the screen. Animation is possible because of a biological phenomenon such as persistence of vision & a psychological phenomenon such as phil.

When you create an animation, organise its execution in a series of logical steps :

(a) Creating animation :

1. Object definition : Gather up in your mind all the activities to provide in the animation or create a written script with list of activities & required objects & then create a story board to visualiset he animation.

2. Choose the animation tool best suited for the application/job.

3. Path specification (of object or camera)

4. Twinning : Then build & twine your sequences. This may involve creating objects, planning their movement, texturing the surfaces, adding lights.

5. Post process : Your animation, doing any special renderings & adding sound effects.

(b) Displaying animation sequence :

1. Displayed using raster animation, color table animation.
2. To achieve smooth animation, sequence of frames have to be presented on screen with a speed of atleast 30 per second.
3. Animation frames can be
   (a) Precomputed & preloaded in memory.

(b) Computed in real time.

Animation techniques in the today are :

1. Cell animation :

It uses a series of progressively different graphics on each frame of movie film & place 24 frames per second. The term cell derives from the clear celluloid sheets that were used for drawing each frame. Cell animation artwork begins with keyframes (the first & last frame of an actions). The series of frames in between the keyframes are drawn by a process called twinning.

2. Computer animation :

Using appropriate software & techniques, we can animate visual images in many ways :

- 2-D : In this, simple visual changes occur on the flat Cartesian x or y axes of the screen. e.g., a blinking word, a colour cycling logo.

- 2½ D : In this, an illusion of depth (2 axis) is added to an image through showing & highlighting.

- 3-D : In 3D, changes are calculated along all the 3 axes (x, y & z) allowing an image or object that itself is *created with a front, back, side, top & bottom view.*

3. Kinematics :

It is study of movement & motion of structures that have joints. Such as walking man. We need to calculate position, rotation, velocity of all the joints & articulated parts involved.

4. Morphine :

It involves transitions not only between still images but between moving images as well.

5. Stop-motion animation.

6. Drawn animation.

7. Cut-out animation.

Some software tools available for animation is market today are :

1. **Apple's hypercard** offers quicktime animation.

2. **Authorware** has a complete set of tools for incorporating & editing multimedia elements.

3. **Macromedia's director** provides a broad set of features to create animations.

4. **mTropolis** from quark provides an object oriented environment in which objects (animations) are assigned properties.

5. Quark Immedia Design Tool & Quark Immediate viewer.

Q. 3. What is JPEG? Enumerate its objectives, applications & architecture.

Ans. Everytime we open a JPEG image & edit it, then compress & save it as compressed. It is the first standard for image compression for still photographs. JPEG scheme compress image about 20%. The compressed image is divided into 8×8 pixels & resulting 64 pixels.

Some objectives of JPEG are :

1. To be at or near the state of art for degree of compression versus image quality.
2. To be parameterisable so that the suer can select the desired compression versus quality trade off.
3. To be applicable to any kind of source image, without regard to dimensions, image content, aspect ratio etc.
4. To have computational requirements that are reasonable for both hardware or software implementations.
5. To support four different modes of operations :

(a) Sequential Encoding :

Where each image component is encoded in the same order that it was scanned.

(b) Progressive Encoding :

Where image is encoded in multiple passes so that a coarse image is presented rapidly, followed by repeated images showing greater and greater detail.

(c) Lossless Encoding :

Where encoding guarantees exact reproduction of all the data in the source image.

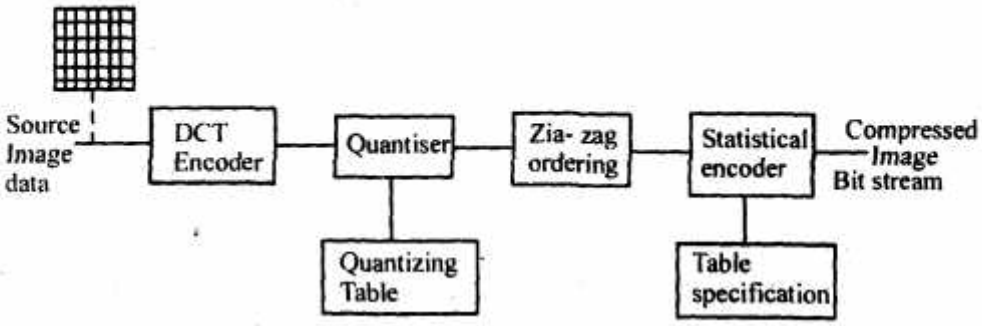
(d) Hierarchical Encoding :

Where image is encoded at multiple resolutions.

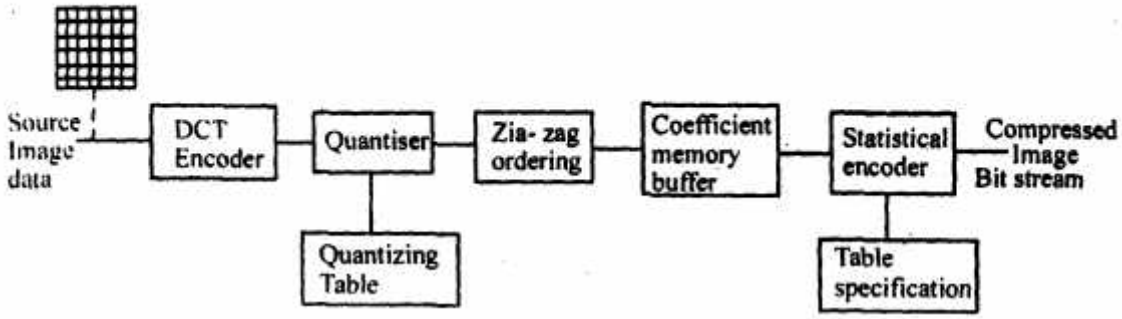
Architecture of JPEG :

The architectures apply to a single gray scale image or to one of the components scale image or two one of the component of a color image. To compress a color image, each block of image is divided into 8×8 pixel blocks, which acts as source image data & then one of following modes gets applied :

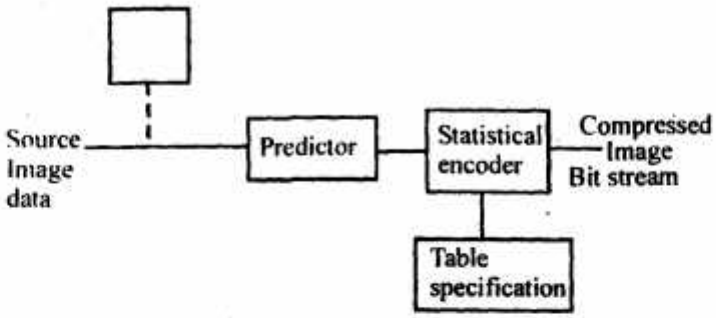
(a) Sequential Coding :



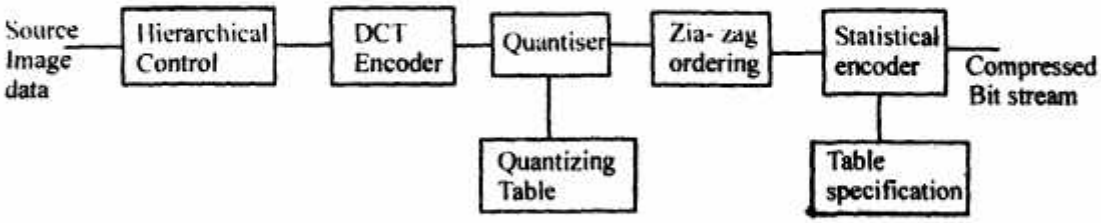
(b) Progressive Encoding :



(c) Lossless Coding :



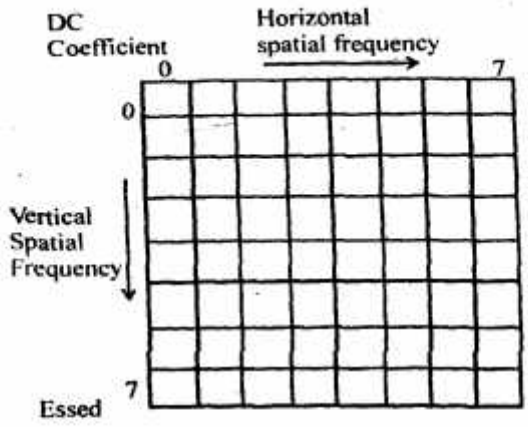
(d) Hierarchical Coding :



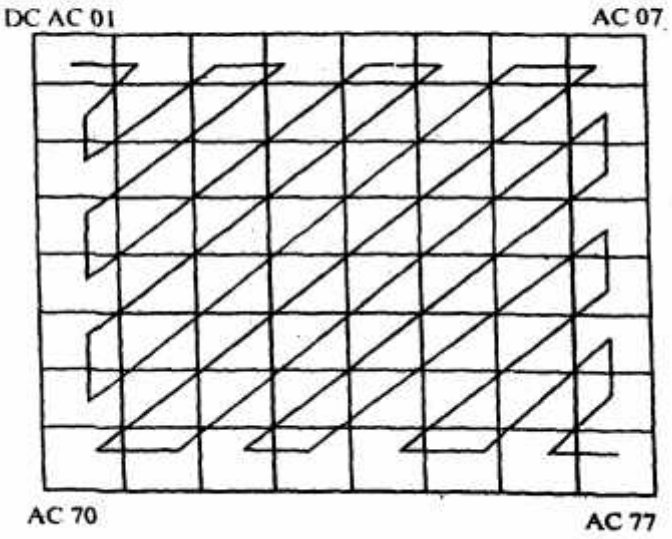
JPEG-DCT Encoding and Quantisation :

The output of DCT encoder (DCT coefficients) as a 2-D array with the DC coefficient in the upper left cover & AC coefficients arranged with increasing spatial frequency horizontally & vertically. These components are quantised according to 64-entry table, which must be specified to the encoder by the application. The quantisation table has 8 bits per entry. After quantisation, the DC coefficient is treated differently from AC coefficients. Zig-zag ordering is an important step, which arranges. The DCT coefficients so that statistical encoding will be more effective.

In order to create a bit stream where coefficients that are more likely to be non-zero (low frequency) are placed before coefficients that are more likely to be zero (high frequency) the zig-zag sequence is used to read the coefficients into the bitstream.



(a) 2-D matrix of DCT coefficients



(b) Zig-zag ordering

JPEG-Statistical coding :

The final encoder processing step is statistical coding. It achieves lossless compression by encoding it achieves lossless compressions by encoding the quantised DCT coefficients more efficiently based on their stastical characteristics. The JPEG standard allows & types of statistical coding Huffmann coding or arithmetic coding.

Huffman coding requires specification of huffman table. Huffman tables also become part of the image bitstream; the standard supports up to four Huffman tables perimage.

The arithmetic statistical coding does not require a separate table to be provided, but is does require a little more processing for implementation.

JPEG Predictive lossless coding :

The lossless compression option, does not use DCT. Instead, a simple predictor is used, but there is a choice of seven different kinds of predictions available. The different predictor choices specify how many and which adjacent pixels are used to predict next pixels. The statistical coding in the lossless mode can use either of the two methods specified for the DCT modes.

Q. 4. Write short notes on :

(a) Color palettes

(b) PNG

Ans.

Color palettes :

Palettes are mathematical tables that define the color of a pixel displayed on the screen. On Macintosh, these tables are called color look up tables (CLUTs). In windows, the term palette is used. The most common palettes are 1, 4, 8, 16 & 24 bits deep.

Color depth	Colors available
1-bit	Black & white (or any 2 colors)
4-bit	16 colors
8-bit	256 colors (good enough for color images)

16-bit	Thousands of colors
24-bit	More than 16 million colors (totally photorealistic)

For 256 color. 8 bit VGA system, the computer uses a color look-up table to determine which 256 colors. Out of the millions possible are available to you at any one time.

Paint programs provide a palette tool for displaying available colors. Most color pickers & selectors also provide a mechanism for specifying a palette color numerically. Palettee display & color picking tools are not uniform among applications or across platforms.

In 24-bit color systems, our graphics adapter works with 3 channels of 256 discrete shaded of each color (red, green & blue) represents as the 3 axes of a cube.

There are some color techniques to avoid. When the destination of our work is a video tape or TV monitor :

- Avoid using a pattern or mosaic.
- Avoid thin horizontal lines.
- Avoid extremely bright or intense colors that may flare up on a TV screen.
- Avoid some reds that may true brown on T.V.

Dithering :

It is a process whereby the colour value of each pixel is changed to the closest matching color value in the target palettee using a mathematical algorithm. Since there are now only 256 colors available to represent the thousands or even million of colors in the original image, pixels using the 256 remaining colors are intermixed & the eye perceives a colornot in the palette, created by blending the colors mixed together. Thus, any given pixel might not be mapped to its closest palette entry, but instead to the average over some area of the image, this average will be closer to the correct color than a substitute color would be.

How well the dithered image renders a good approximation of original depends upon algorithm used.

2. PNG : It stands for portable network graphics & is pronounced as ping, is the newest & most flexible of these 3 graphic file formats. It is an open source graphics file format that could be freely developed on the web. For e.g., each time an on line weather station wants to generate a new graphic with the local forecast. It's supposed to buy a license first to develop in GIF format.

For this reason, PNG was created to replace the GIF file format on the web. PNG is an extension to GIF.

- It replaces GIF.
- It uses lossless compression.
- It allows you to choose between 3 bit depths that is, 8, 4, 32 bit depth & more the depth, more the size of image.

PNG supports only transparency but do not support interlacing & animation feature.

By looking at the characteristics of PNG, it light think of PNG as being the best of both GIF & JPEG formats. But PNG does not replaces JPEG format, it does after & expand on some of its features.

PNG supports multiple levels & multiple colors of transparency. While PNGs using 1-bit transparency are smaller in file size to GIFs, those with multilevel transparency are substantially larger.

Characteristics	Description
- Color mode	Can be stored in 8 bit, 24 bit or 32 bit
- Compression method	Lossless
- Animation	Not supported
- Transparency	Supported (only in 8 bit or 3 bit mode)
- Interlacing	Not supported
- File type	Bitmap

The additional benefits of PNG is its gamma corrections. The PNG file format has the capability to correct for different in how computers & monitors interpret color values. All these characteristics make a PNG well suited for almost any type of web graphic. Only some of the newest browsers support it, a plug-in must be downloaded to view graphics in PNG format on older browsers.

Q. 5. (a) Explain the concept of MPEG in detail.

Ans. Motion Picture Expert Group is a digital motion video & associated audio compression standard. It is a bitstream specification standard where pixels are converted into bitstream. It is a method of choice for encoding motion images because it is widely accepted for internet & DVD video.

Some of its objectives are :

1. The standard will deliver acceptable video quality at compressed data rates between 1.0 & 1.5 Mbps.
2. It will support either symmetric.
3. Audio/video synchronisation will be maintained.
4. When it is required, compression-decompression delay can be controlled.
5. Catastrophic behaviour in the presence of data errors should be avoidable.

Architecture of MPEG is as follows :

There are 4 different kinds of pictures, depending on how each picture is to be decoded.

- I pictures are intracoded, meaning that they are coded independent of any other picture.
- P pictures are predicted pictures.
- B pictures are interpolated pictures, which are coded by interpolating between a previous and a future I or P picture.
- D pictures are special format that is only used for implementing fast search modes.

Bit Stream Syntax :

The six levels of hierarchy supports all features of MPEG :

1. Sequence layer : It is an independent video stream.

2. Group of picture layer : This is a dip of video that begins with a ran down access entry point.

3. Picture layer : It represents/displayable image.

4. Slice layer : Provides resynchronisation of decoder in the event of an error.

5. Macroblock layer : It is a 16 × 16 pixel motion compensation unit.

6. Block layer : 8 × 8 pixel block that can be intracoded, interpolated or motion compensated.

MPEG standards are :

1. MPEG 1 :

It can deliver 1.2 Mbps of video, 250 kbps of 2 channel stereo audio using CD-ROM technology. Software decompression algorithm for MPEG1 are slower. It requires dedicated hardware for playback but better than TV screen.

2. MPEG 2 :

It requires higher data rate of 2 to 1.5 Mbps. It delivers high resolution, picture quality, stability, interlaced video formats. It uses DCT algorithm.

3. MPEG 4 :

It is in development & it is a content based method for assimilating multimedia elements. It offers indexing, hyperlinking, browsing, uploading, downloading. It allows virtual reality working.

4. MPEG 7 :

It integrates information about image, video, sound elements which is already possible is MPEG 4, along with class for facial expressions, personality characteristics.

Q. 5. (b) Explain DVI technology. Explain its working.

Ans. Intel corporation & IBM corporation have developed a proprietary programmable intel chipset which implements the compression/decompression technology in a co-processor environment on any type of computer platform. These chipsets consist of 2 VLSI chips to separate image processing & display functions DVI is an important engine for present & future compression developments.

DVI technology can do both symmetric & asymmetric motion video compression/decompression. A symmetric approach is PLV while symmetric approach is RTV.

DVI-production level compression :

Video for PLV must be sent to central compression facility which uses large computers & special interface equipments. For compression but must be decompressed or played back by the DVI systems itself.

Its picture quality is the highest :

- It is an interframe compression technique where compression is done frame by frame.
- It is block oriented & involves multiple compression techniques.
- PLV has high resolution & frame rate than RTV.
- It is very costly to perform PLV at reasonable speeds.

DVI-Real time compression :

It can be implemented on any DVI system where Action Media capture board is installed. Compression is done with computing power available in dVI system as well as decompression or playing back is also on the same system. Picture quality is not good. It is a compression process that is done in real time, on a DVI development system. With RTV, the developer may compress his/her own video & audio to same file size as PLV & then used in same way as PLV files are used. RTV is not frame to frame compressed. So can be started or stopped at any point in time.

Q. 6. (a) Explain any 2 techniques used for encoding analog signals?

Ans. Real world requirements may make it impossible to handle the full bit stream of, for e.g., CD-quality audio. One solution is

Data Modulation :

Delta modulation is to encode not the value of each sample, but the difference between one sample & the next. In most cases, fewer bits per sample need to be transmitted, but delta modulation has problems. Delta modulation is a modification of DPCM. In Differential Pulse Code Modulation (DPCM), it is not necessary to store the whole number of bits of each sample. It is sufficient to represent only the first PCM-coded sample as a whole & all following samples as the difference from the previous one. Whereas in delta modulation. When coding the differences, it uses exactly one bit, which indicates whether the signal increases or decreases. This leads to an inaccurate coding of steep edges. This technique is particularly profitable if the coding does not depend on 8-bit grid units. If the differences are small, a small number of bits is sufficient. A non-typical sequence of characters will not be compressed using these methods.

A more prominent variant & practical adaptive compression technique is adaptive.

Delta pulse code modulation (ADPCM) : In order to handle both signals that change quickly as well as signals that change slowly, the step size encoded between adjacent samples varies according to the signal itself. In other words, if the waveform is changing rapidly, large steps can be quantised. This is the method used in CD-I (compact disc-interactive). It is a successful development of DPCM. Here, differences are encoded using a small number of bits only (e.g., 4 bits). Therefore, either rough transitions are coded correctly (these bits represent bits with a higher significance) or small changes are coded exactly (DPCM-Encoded values are the less significant bits). In the first case, the resolution of low audio signals would not be sufficient & in the second case, a loss of high frequencies would occur. ADPCM adapts to this significance for a particular data stream as follows : the coder divides the value of DPCM samples by a suitable coefficient & the decoder multiplies the compressed data by the same coefficient. i.e., the step size of signal changes. For A-low transmission, signal is encoded according to,

$$y = \begin{cases} \dfrac{AX}{1+LnA} & \left(0 \le x \le \dfrac{1}{A}\right) \\ \dfrac{1+Ln(AX)}{1+LnA} & \left(\dfrac{1}{A} \le x \le 1\right) \end{cases}$$

For μ -low encoding, formula is,

$$Y = \frac{Ln(1+\mu x)}{Ln(1+\mu)} \qquad (0 \le x \le 1)$$

Q. 6. (b) Enumerate various applications of multimedia.

Ans. 1. Entertainment :

The use of interactive media for entertainment is not a new phenomenon. A drive for VOD, interactive TV in VOD and move from highly controlled audience mode to openeded collaboration group have increased the usage of multimedia.

2. VOD :

This system allows users to select & watch video content over a network as a part of an interactive television system. VOD provides users with a large subset of VCR functionality including pause, fast forward, fast rewind etc.

3. Collaborative computer supported games :

The collaborative computer supported games such as 'The Next President' is based on simulated election

campaign in which thousand of users participate.

4. Interactive cinema :

In this, the next scene or action of actor/actress is decided by the spectators. One such example is the movie 'I'm Your man.'

5. Manufacturing :

Many organisations have adopted concurrent engineering for the design & manufacturing of complex system & products. Role of multimedia in concurrent engineering :

(a) Engineering & manufacturing as a visually rich domain for with information processing.

(b) Group communication have facilitated participants to refer to on-line materials.

6. Business :

MM is being used in real estates sales. e.g., Home vision allows users to browser a photo database of homes.

7. MM communication for health care :

Use of imaging techniques (X-Ray, MRI), on line medical record system, imaging for various hospital facilities have increased in health care owing to multimedia.

8. Geographic information systems :

Databases are searched for information whose user interfaces must be visually oriented, showing maps blueprints of the facilities. On line support is provided for management of roads, building & power lines.

9. Education :

It can be made more interactive using multimedia & recasted by aspiring communicators. Which is hard to express, can be easily conveyed by pictures, video, audio etc. Visual AC wave is a project of Apple's multimedia lab project.

Q. 7. What is virtual reality operating system? How virtual environment display & orientation making takes place & how these are useful to society?

Ans. The term Virtual Reality (VR) is used by many different people with many meanings. There are some people to whom VR is a specific collection of technologies i.e., a Head Mounted Display, Glove Input Device

& Audio some other people stretch the term to include conventional books, movies or pure fantasy & imagination.

Virtual Reality is a way for humans to visualize, manipulate and interact with computers and extremely complex data. It is also called as oxymoron, cyberspace, Artificial Reality, Synthetic environment, Simulator Technology. The visualisation part refers to the computer generating visual, auditory or other sensual outputs to the user of a world within the computer. This world may be a CAD model, a scientific simulation or a view into a database. Some systems uses a conventional computer monitor to display the visual world. This sometimes is called desktop VR or a Window on World (wow).

Immersive Systems :

The ultimate VR systems completely immerse the user's personal viewpoint inside the virtual world. These immersive VR systems are often equipped with a head Mounted Display (HMD). This is a helmet or a face mask that holds the visual & auditory displays. A nice variation of the immersive systems use multiple large projection displays to create a 'cave' or room in which the viewer stand.

Virtual Environment Displays :

One hardware device closely associated with VR is the Head Mounted Device (HMD). These use some sort of helmet or goggles to place small video displays in front of each eye, with special optics to focus & stretch the perceived field of view. Most HMDS use 2 displays and can provide stereoscopic imaging. Other use a single larger display to provide higher resolution, but without the stereoscopic vision. Most lower cost HMDs use LCD displays, while others use small CRTs. The wave expensive HMDs use special CRTs. A HMD requires a position tracker in addition to the helmet. Alternatively, the display can be mounted an armature for support and tracking (a Boom, display).

Orientation Tracking :

1. Mechanical armatures can be used to provide fast & very accurate tracking such armatures may look like a desk camp. The drawbacks of mechanical sensors are the encumbrance of device & its restrictions on motion. Shooting star system makes a low cost armature system for head tracking.

2. Ultrasonic sensors can be used to track position and orientation. A set of emitters & receivers are used with a known relationship between the emitters & between the receivers. Drawbacks are low resolution, long lag times and interference from echoes. Logitech & Transition state are 2 companies that provide ultrasonic tracking systems.

3. Magnetic trackers use sets of coil that are pulsed to produce magnetic fields. Limitations are high latency for the measurement & processing, range limitations, interference from ferrous materials. The 2 primary companies selling magnetic trackers are pathemus & ascension.

4. Optical Orientation :

Tracking systems have been developed. On method uses a ceiling grid LEDs & a head mounted camera. Two problems are limited space & lack of full motion (rotations). One company selling an optical tracker is origin instruments.

5. Inertial trackers have been developed that are small & accurate enough for VR use. These only provide notational measurements & are also not accurate for slow orientation changes.

In society, the applications being developed for VR run a wide spectrum, from games to architectural and business planning. Many applications are worlds that are very similar to our own, like CAd or architectural modelling.

Some other applications for society involve visualising the web and flow of word's financial markets, navigating a large corporate information base etc.

A number of companies made low cost LCD shutter glasses for use with TVs. There are circuits and code for hooking these up to a computer available on many of the online systems, BBSs and internet FTP sites.

Q. 8. Write short notes on :

(a) Speech recognition and generation

(b) Intelligent virtual reality software system.

Ans. (a) Speech Recognition :

A speech recognition system starts by breaking speech down into a parametric representation. The first step is to isolate speech segments in time. The speech signal is paremeterised as the outputs of a bank of bandpass filters or as LPC coefficients. Another algorithm for improving identification involves Hidden Markow Models (HMM). The ore of a HMM identification system is a finite-state machine, with probabilities associated with the transition from one state to another. But the states of the machine cannot be directly observed. Instead, a finite number of observations can be made about the current state of the state machine. There is an algorithm for deriving the probability that a given sequence of observations was generated by a given sequence of states. There is also an algorithm for deciding which of a set of models produced the speech being analyses. Recently, neural networks have come to be used for speaker identification & speaker verification.

Obviously, speech recognition systems have an easier job if all speakers speak the same text. Isolated words are easier, connected speech is harder. Handling any arbitrary speaker from the general population is harder.

Speech generation :

The organs involved in speech include larynx which encloses loose flaps of muscle called vocal cords. The puffs of air that are released create a waveform, which can be approximated by a series of rounded pulses. The waveform created by vocal cords propagates through a series of irregularity shaped tubes, including throat, the mouth & the nasal passages. At the lips & other points in the tract, part of the waveform is transmitted further & part is reflected. The flow can be significantly constricted or completely interrupted by the uvula, the teeth and the lips.

A voiced sound occurs when the vocal cords produce a more or less regular waveform vowels are voiced sounds produced without any major abstraction in the vocal cavity. Consonants arise when the vocal tract is more or less obstruct sounds at the level of consonants & vowels are collectively known as phone news, the most basic unit of speech differentiation, analysis & synthesis. The next level up from phonemes is the dipthony and the syllable, then the word.

A major driving force in speech synthesis has come from text-to-speech (TTS). A TTS system assumes that the text already exists in machine readable form, such as an ASCII file. The machine readable form is possible obtained from optical character recognition. TTS converts text symbols to a parameter stream representing sounds. Other parts of stream are broken down into morphemies the syntactic basic units of the

language. The sounds are coucäterated. Then higher-level elements of speech such as prosody (rise & fall of pitch), overall emphasis, & glottal stops are added. Syntactical analysis provides the basis for adjustments in clause ends or sentence ends. e.g., the rise in pitch for a question.

(b) Intelligent multimedia system consists of an integrated work environment with a human computer interface designed as an intelligent agent with the ability to conduct dialogue with the user in coordinated multiple media/modalities. The human-computer interaction is modeled on the wanner in which 2 or more people naturally communicate in coordinated multiple modalities when working with graphics, video & other devices at hand.

The system should have the ability to :

1. Conduct dialogue with the user :
   - (a) Adhere to respected principles of conversation.
   - (b) Adhere to respected human factor guidelines.
2. Maintain knowledge & belief models to enable the system to understand user inputs & compose system outputs.
   - (a) Track & model the dynamic focus of the dialogue in order to maintain context during the dialogue.
   - (b) Model the user's task & the state of the user's accomplishments & progress with respect to the task.
3. Maintain knowledges bases of information about :
   - (a) Modalities & user interaction.
   - (b) World knowledge
   - (c) Application-specific knowledge.
4. Act as an intelligent agent for accessing & using application systems, through such activities as :
   - (a) Assisting the user with finding, selecting & accessing appropriate tools to apply to the task.
   - (b) Assisting & guiding the user in the accomplishment of tasks.
5. Decide how information and responses are to be presented to the user.
   - (a) Compose the output in multiple modalities.
   - (b) Present the multimedia output in a coordinated manner.