Q.2 a. Multimedia authoring tools are used for what purpose? Elaborate features of multimedia authoring tools.

Answer:

Multimedia authoring tools provide the important framework you need for organizing and editing the elements of multimedia like graphics, sounds, animations and video clips. Authoring tools are used for designing interactivity and the user interface, for presentation your project on screen and assembling multimedia elements into a single cohesive project.
Features of multimedia authoring tools are: (briefly explain any 6)

- Editing features
- Organizing features
- Programming features
- Interactive features
- Performance tuning features
- Playback features
- Delivery features
- Cross-Platform features
- Internet Playability

b. What does PNG format and TIFF stand for? Write special features of these two file formats.

Answer:

PNG format: standing for Portable Network Graphics - meant to supersede the GIF standard, and extend it in important ways.
Special features of PNG files include:

1. Support for up to 48 bits of color information - a large increase.
2. Files may contain gamma-correction information for correct display of color images, as well as alpha-channel information for such uses as control of transparency.
3. The display progressively displays pixels in a 2-dimensional fashion by showing a few pixels at a time over seven passes through each 8x8 block of an image.

TIFF: stands for Tagged Image File Format.

- The support for attachment of additional information (referred to as “tags") provides a great deal of flexibility.

1. The most important tag is a format designifier: what type of compression etc. is in use in the stored image.
2. TIFF can store many different types of image: 1-bit, grayscale, 8-bit color, 24-bit RGB, etc.
3. TIFF was originally a lossless format but now a new JPEG tag allows one to opt for JPEG compression.

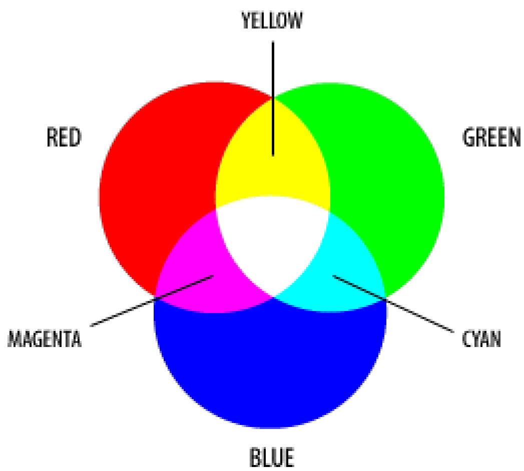
Q.3 a. Explain RGB and CMY(K) color models in images. Is transformation from RGB to CMY(K) possible? Give conversion matrix.

Answer:

RGB and its subset CMY form the most basic and well-known color model. This model bears closest resemblance to how we perceive color. It also corresponds to the principles of additive and subtractive colors.

RGB

Red, green, and blue are the primary stimuli for human color perception and are the primary additive colors. The relationship between the colors can be seen in this illustration:



The secondary colors of RGB, cyan, magenta, and yellow, are formed by the mixture of two of the primaries and the exclusion of the third. Red and green combine to make yellow, green and blue make cyan, blue and red make magenta.

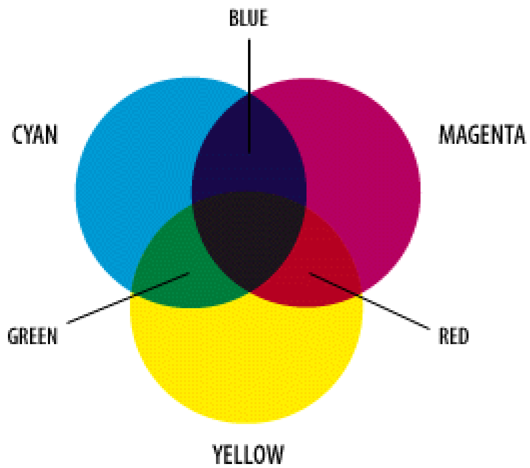
The combination of red, green, and blue in full intensity makes white. White light is created when all colors of the EM spectrum converge in full intensity.

The importance of RGB as a color model is that it relates very closely to the way we perceive color with the r g b receptors in our retinas. RGB is the basic color model used in television or any other medium that projects the color. It is the basic color model on computers and is used for Web graphics, but it cannot be used for print production.

CMY(K)

Cyan, magenta, and yellow correspond roughly to the primary colors in art production: red, blue, and yellow. In the illustration below, you can see the CMY counterpart to the RGB model shown above:

Just as the primary colors of CMY are the secondary colors of RGB, the primary colors of RGB are the secondary colors of CMY. But as the illustrations show, the colors created by the subtractive model of CMY don't look exactly like the colors created in the additive model of RGB. Particularly, CMY cannot reproduce the brightness of RGB colors. In addition, the CMY gamut is much smaller than the RGB gamut



The CMY color space describes colors in terms of the subtractive primaries: cyan, magenta, and yellow. CMY is used mainly for hardcopy devices such as color printers. Generally, the conversion from RGB to CMY follows the equation

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

b. What is Chroma Subsampling? Where it is used? Briefly describe various chroma subsampling schemes. (6)

Answer:

Chroma subsampling is the practice of encoding images by implementing less resolution for chroma information than for luma information, taking advantage of the human visual system's lower acuity for color differences than for luminance.

It is used in many video encoding schemes — both analog and digital — and also in JPEG encoding.

Different chroma subsampling schemes:

(a) The chroma subsampling scheme \4:4:4" indicates that no chroma subsampling is used: each pixel's Y, Cb and Cr values are transmitted, 4 for each of Y, Cb, Cr.
(b) The scheme \4:2:2" indicates horizontal subsampling of the Cb, Cr signals by a factor of 2. That is, of four pixels horizontally labelled as 0 to 3, all four Ys are sent, and every two Cb's and two Cr's are sent, as (Cb0, Y0)(Cr0, Y1)(Cb2, Y2)(Cr2, Y3)(Cb4, Y4), and so on (or averaging is used).
(c) The scheme \4:1:1" subsamples horizontally by a factor of 4.
(d) The scheme \4:2:0" subsamples in both the horizontal and vertical dimensions by a factor

of 2. Theoretically, an average chroma pixel is positioned between the rows and columns.

c. Define HDTV. (3)

Answer:

(a) HDTV has a much wider aspect ratio of 16:9 instead of 4:3.
(b) HDTV moves toward progressive (non-interlaced) scan. The rationale is that interlacing introduces serrated edges to moving objects and
flickers along horizontal edges.

Q.4 a. State and briefly explain Nyquist Theorem, Signal-to-Noise Ratio(SNR), Signal-to-Quantization-Noise Ratio(SQNR). (6)

Answer:

Signal to Noise Ratio (SNR)
The ratio of the power of the correct signal and the noise is called the signal to noise ratio (SNR) | a measure of the quality of the signal.
The SNR is usually measured in decibels (dB), where 1 dB is a tenth of a bel. The SNR value, in units of dB, is defined in terms of base-10 logarithms of squared voltages, as follows:

$$SNR = 10 \log_{10} \frac{V_{signal}^2}{V_{noise}^2} = 20 \log_{10} \frac{V_{signal}}{V_{noise}}$$

Signal to Quantization Noise Ratio (SQNR)

The quality of the quantization is characterized by the Signal to Quantization Noise Ratio (SQNR).
(a) Quantization noise: the difference between the actual value of the analog signal, for the particular sampling time, and the nearest quantization interval value.
(b) At most, this error can be as much as half of the interval.

(c) For a quantization accuracy of N bits per sample, the SQNR can be simply expressed:

$$SQNR = 20\log_{10}\frac{V_{signal}}{V_{quan\_noise}} = 20\log_{10}\frac{2^{N-1}}{\frac{1}{2}}$$

$$= 20 \times N \times \log 2 = 6.02\,N(\text{dB})$$

b. State Adaptive Huffman Principle. Explain Initialization, coding, decoding and updating the tree part of Adaptive Huffman algorithm using a suitable example.

Answer:

Adaptive Huffman Principle
• In an optimal tree for n symbols there is a numbering of the nodes y1<y2<... <y2n-1 such that their corresponding weights x1,x2, ... , x2n-1 satisfy:
– x1 < x2 < ... < x2n-1
– siblings are numbered consecutively
• And vice versa
– That is, if there is such a numbering then the tree

Initialization
• The tree will encode up to m + 1 symbols including NYT.
• We reserve numbers 1 to 2m + 1 for node numbering.
• The initial Huffman tree consists of a single node

Coding Algorithm
1. If a new symbol is encountered then output the code for NYT followed by the fixed code for the symbol. Add the new symbol to the tree.
2. If an old symbol is encountered then output its code.
3. Update the tree to preserve the node number invariant.

Decoding Algorithm
1. Decode the symbol using the current tree.
2. If NYT is encountered then use the fixed code to decode the symbol. Add the new symbol to the tree.
3. Update the tree to preserve the node number invariant.

Updating the Tree
1. Let y be leaf (symbol) with current weight x.*
2. If y the root update x by 1, otherwise,
3. Exchange y with the largest numbered node with the same weight (unless it is the parent).**
4. Update x by 1
5. Let y be the parent with its weight x and go to 2.

Q.5 a. What is the rationale behind Transform coding? Write a note on DCT. Is DCT a linear transform. Explain.

Answer:

The rationale behind transform coding:
If Y is the result of a linear transform T of the input vector X in such a way that the components of Y are much less correlated, then Y can be coded more efficiently than X.
DCT stands for Discrete Cosine Transform. Spatial frequency indicates how many times pixel values change across an image block.
_ The DCT formalizes this notion with a measure of how much the image contents change in correspondence to the number of cycles of a cosine wave per block.
_ The role of the DCT is to decompose the original signal into its DC and AC components; the role of the IDCT is to reconstruct (re-compose) the signal.

Given an input function f(i; j) over two integer variables i and j (a piece of an image), the 2D DCT transforms it into a new function F(u; v), with integer u and v running over the same range as i and j. The general definition of the transform is:

$$F(u,v) = \frac{2\,C(u)\,C(v)}{\sqrt{MN}} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \cos\frac{(2i+1)\cdot u\pi}{2M} \cdot \cos\frac{(2j+1)\cdot v\pi}{2N} \cdot f(i,j)$$

where i; u = 0;1; : : : ; M − 1; j; v = 0;1; : : : ; N − 1; and the constants C(u) and C(v) are determined by

$$C(\xi) = \begin{cases} \frac{\sqrt{2}}{2} & if \quad \xi = 0, \\ 1 & otherwise. \end{cases}$$

2D Discrete Cosine Transform (2D DCT):

$$F(u,v) = \frac{C(u)\,C(v)}{4} \sum_{i=0}^{7} \sum_{j=0}^{7} \cos\frac{(2i+1)u\pi}{16} \cos\frac{(2j+1)v\pi}{16} f(i,j) \qquad (8.17)$$

where $i, j, u, v = 0, 1, \ldots, 7$, and the constants $C(u)$ and $C(v)$ are determined by Eq. (8.5.16).

2D Inverse Discrete Cosine Transform (2D IDCT):

The inverse function is almost the same, with the roles of $f(i,j)$ and $F(u,v)$ reversed, except that now $C(u)C(v)$ must stand inside the sums:

$$\tilde{f}(i,j) = \sum_{u=0}^{7} \sum_{v=0}^{7} \frac{C(u)\,C(v)}{4} \cos\frac{(2i+1)u\pi}{16} \cos\frac{(2j+1)v\pi}{16} F(u,v) \qquad (8.18)$$

where $i, j, u, v = 0, 1, \ldots, 7$.

1D Discrete Cosine Transform (1D DCT):

$$F(u) = \frac{C(u)}{2} \sum_{i=0}^{7} \cos\frac{(2i+1)u\pi}{16} f(i) \qquad (8.19)$$

where $i = 0, 1, \ldots, 7$, $u = 0, 1, \ldots, 7$.

1D Inverse Discrete Cosine Transform (1D IDCT):

$$\tilde{f}(i) = \sum_{u=0}^{7} \frac{C(u)}{2} \cos\frac{(2i+1)u\pi}{16} F(u) \qquad (8.20)$$

where $i = 0, 1, \ldots, 7$, $u = 0, 1, \ldots, 7$.

The DCT is a linear transform:
In general, a transform T (or function) is linear, if

$$\mathcal{T}(\alpha p + \beta q) = \alpha \mathcal{T}(p) + \beta \mathcal{T}(q)$$

where _ and _ are constants, p and q are any functions, variables or constants.
From the definition in Eq. 8.17 or 8.19, this property can readily be proven for the DCT because it uses only simple arithmetic operations.

b. What are design goals for JPEG2000 standard? Briefly highlight various problems this standard addresses.

Answer:

Design Goals:
– To provide a better rate-distortion tradeoff and improved subjective image quality.
– To provide additional functionalities lacking in the current JPEG standard.

The JPEG2000 standard addresses the following problems:
Lossless and Lossy Compression: There is currently no standard that can provide superior lossless compression and lossy compression in a single bitstream.
Low Bit-rate Compression: The current JPEG standard offers excellent rate-distortion performance in mid and high bit-rates. However, at bit-rates below 0.25 bpp, subjective distortion becomes unacceptable. This is important if we hope to receive images on our web-enabled ubiquitous devices, such as web-aware wristwatches and so on.
large Images: The new standard will allow image resolutions greater than 64K by 64K without tiling. It can handle image size up to 232 − 1.
– **Single Decompression Architecture:** The current JPEG standard has 44 modes, many of which are application specific and not used by the majority of JPEG decoders.
Transmission in Noisy Environments: The new standard will provide improved error resilience for transmission in noisy environments such as wireless networks and the Internet.
– **Progressive Transmission:** The new standard provides seamless quality and resolution scalability from low to high bit-rate. The target bit-rate and reconstruction resolution need not be known at the time of compression.
– **Region of Interest Coding:** The new standard allows the specification of Regions of Interest (ROI) which can be coded with superior quality than the rest of the image.
One might like to code the face of a speaker with more quality than the surrounding furniture.
Computer Generated Imagery: The current JPEG standard is optimized for natural imagery and does not perform well on computer generated imagery.
– **Compound Documents:** The new standard offers metadata mechanisms for incorporating additional non-image data as part of the file. This might be useful for including text along with imagery, as one important example.

Q.6 a. Describe briefly four optional H.263 Coding Modes.

Answer:

1. Unrestricted motion vector mode :{The pixels referenced are no longer restricted to be within the boundary of the image.
{When the motion vector points outside the image boundary, the value of the boundary pixel that is geometrically closest to the referenced pixel is used.
{The maximum range of motion vectors is [-31.5, 31.5].

2. Syntax-based arithmetic coding mode:
{ As in H.261, variable length coding (VLC) is used inH.263 as a default coding method for the DCT coefficients.
{ Similar to H.261, the syntax of H.263 is also structured as a hierarchy of four layers. Each layer is coded using a combination of _xed length code and variable length code.

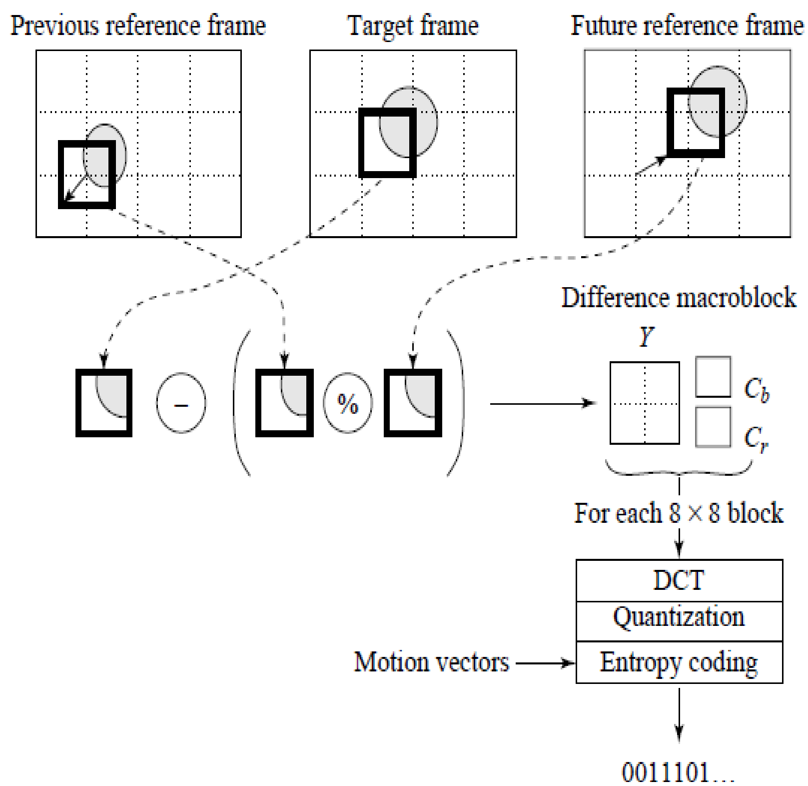
3. Advanced prediction mode:
{ In this mode, the macroblock size for MC is reduced from 16 to 8.
{ Four motion vectors (from each of the 8_8 blocks) are generated for each macroblock in the luminance image.
4. PB-frames mode:
{ In H.263, a PB-frame consists of two pictures being coded as one unit.
{ The use of the PB-frames mode is indicated in PTYPE.
{ The PB-frames mode yields satisfactory results for videos with moderate motions.
{ Under large motions, PB-frames do not compress as well as B-frames and an improved new mode has been developed in Version 2 of H.263.

b. How does Motion Compensation (MC) based video encoding works in H.261? Explain MC-based B-frame coding idea of MPEG-1.

Answer: Motion Compensation (MC) based video encoding in H.261 works as follows:

{ In Motion Estimation (ME), each macroblock (MB) of the Target P-frame is assigned a best matching MB from the previously coded I or P frame - prediction.
{ prediction error: The di_erence between the MB and its matching MB, sent to DCT and its subsequent encoding steps.
{ The prediction is from a previous frame | forward prediction.
MPEG introduces a third frame type | B-frames, and its accompanying bi-directional motion compensation.
_ The MC-based B-frame coding idea is illustrated in Figure:
{Each MB from a B-frame will have up to two motion vectors (MVs) (one from the forward and one from the backward prediction).
{If matching in both directions is successful, then two MVs will be sent and the two corresponding matching MBs are averaged (indicated by `%' in the figure) before comparing to the Target MB for generating the prediction error.
{ If an acceptable match can be found in only one of the reference frames, then only one MV

and its corresponding MB will be used from either the forward or backward prediction.



Q.7 a. Give an overview of MPEG-4. Explain object-based visual coding in MPEG-4.

Answer: MPEG-4: a newer standard. Besides compression, pays great attention to issues about user interactivities.

• MPEG-4 departs from its predecessors in adopting a new object-based coding:
– Offering higher compression ratio, also beneficial for digital video composition, manipulation, indexing, and retrieval.
• The bit-rate for MPEG-4 video now covers a large range between 5 kbps to 10 Mbps.
MPEG-4 is an entirely new standard for:
(a) Composing media objects to create desirable audiovisual scenes.
(b) Multiplexing and synchronizing the bitstreams for these media data entities so that they can be transmitted with guaranteed Quality of Service (QoS).
(c) Interacting with the audiovisual scene at the receiving end
— provides a toolbox of advanced coding modules and algorithms for audio and video compressions.
The hierarchical structure of MPEG-4 visual bitstreams is very different from that of MPEG-1 and -2, it is very much video object-oriented.
Object-based Visual Coding in MPEG-4

MPEG-4 VOP-based coding also employs the Motion Compensation technique:
– An Intra-frame coded VOP is called an I-VOP.
– The Inter-frame coded VOPs are called P-VOPs if only forward prediction is employed, or B-VOPs if bi-directional predictions are employed.
• The new difficulty for VOPs: may have arbitrary shapes, shape information must be coded in addition to the texture of the VOP.
Note: texture here actually refers to the visual content, that is the gray-level (or chroma) values of the pixels in the VOP.

b. What is Vocoder? For what purpose it is used? Explain LPC Vocoder.

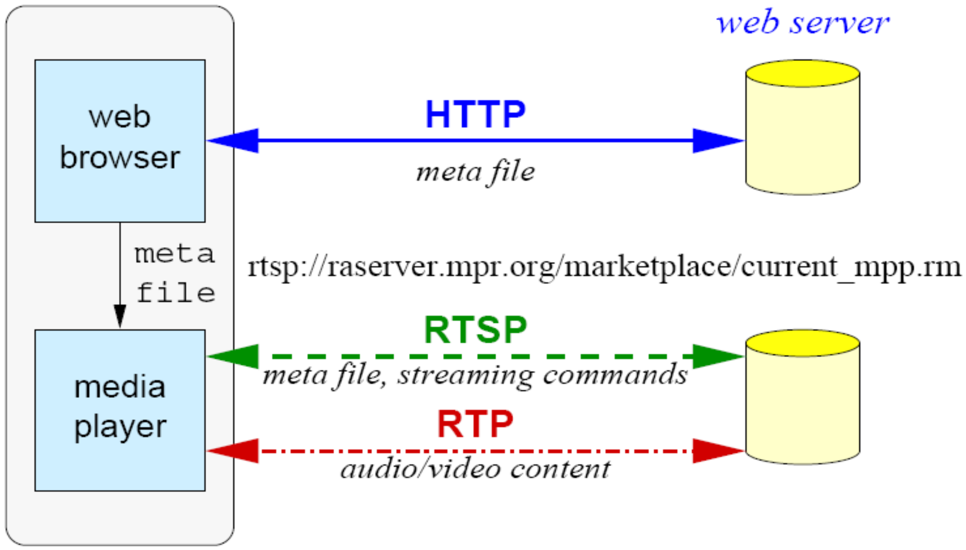
Answer:

Vocoders – voice coders, which cannot be usefully applied when other analog signals, such as modem signals, are in use.
◦ concerned with modeling speech so that the salient features are captured in as few bits as possible.
Linear Predictive Coding (LPC)
 LPC vocoders extract salient features of speech directly from the waveform, rather than transforming the signal to the frequency domain
 LPC Features:
◦ uses a time-varying model of vocal tract sound generated from a given excitation
◦ transmits only a set of parameters modeling the shape and excitation of the vocal tract, not actual signals or differences
 small bit-rate
 About "Linear": The speech signal generated by the output vocal tract model is calculated as a function of the current speech output plus a second term linear in previous model coefficients
LPC Coding Process
 LPC starts by deciding whether the current segment is voiced or unvoiced:
◦ For unvoiced: a wide-band noise generator is used to create sample values f(n) that act as input to the vocal tract simulator
◦ For voiced: a pulse train generator creates values f(n)
◦ Model parameters ai: calculated by using a least squares set of equations that minimize the difference between the actual speech and the speech generated by the vocal tract model, excited by the noise or pulse train generators that capture speech parameters

Q.8 a. What do you understand by RTSP? Give its architecture and state its various properties.

Answer:

Real-Time Streaming protocol (RTSP): Real Time Streaming Protocol, RTSP, is an application-level protocol. Its goal is to offer robust protocol that can stream multimedia over multicast and unicast in "one to many"-applications. It also supports interoperation between clients and servers from different vendors.



The Real Time Streaming Protocol boasts of very many properties. It is clear that IPv6 has been considered in these:

- **Extendable** (new methods and parametres are easy to add)
- **Easy to parse** (standard HTML or MIME parser can be used)
- **Secure** (HTTP authentication methods, transport and network layer security mechanisms applicable)
- **Transport-independent** (protocols such as UDP, RDP and TCP applicable)
- **Multi-server capable** (there can be media streams from different servers in one presentation)
- **Control of recording devices** (both playback and recording control possible)
- **Separation of stream control and conference initiation** (The only requirement is that the conference initiation protocol either provides or can be used to create a unique conference identifier)
- **Suitable for professional applications** (frame-level accuracy through SMPTE time stamps is supported to allow remote digital editing)
- **Presentation description neutral** (no particular format imposed, the presentation description must contain at least one RTSP URI, however)
- **Proxy and firewall friendly** (protocol should be readily handled by both application and transport-layer firewalls)
- **HTTP-friendly** (RTSP reuses HTTP concepts, where sensible)
- **Appropriate server control** (i.e. servers should not start streaming to clients in such a way that clients cannot stop the stream)
- **Transport negotiation** (transport method can be negotiated just before streaming)
- **Capability negotiation**(client must have a way to find out if some of the basic features are disabled in the server)

b. Describe briefly Internet Telephony.

Answer: Refer Page 455-456 from Text Book-I

Q.9 a. Define animation and list various principles of animation and describe how it can be used in multimedia?

Answer:

By definition, animation is the act of making something come alive.
Depending on the size of the project, you can animate the whole thing or you can just animate parts of it.

Principles of animation

Animation is possible because of a biological phenomenon known as persistence of vision and
a psychological phenomenon called phi.
With animation, a series of images are changed very slightly and very rapidly, one after the other, seemingly blending together into a visual illusion of movement.
Digital display video builds 24, 30, or 60 entire frames or pictures every second. Movies on film are typically shot at a shutter rate of 24 frames per second.
Visual effects such as wipes, fades, zooms, and dissolves, available in most authoring packages, are a simple form of animation. Animation is an object actually moving across, into, or out of the screen.

b. What is a Magneto Optical (MO) drive and how does it work?

Answer:

An MO disk drive is a computer storage device which utilizes both an optical laser and magnetic field to record data on a special removable optical disk. This recording technology has a number of benefits for the user including high reliability and low cost per megabyte of storage.

Magneto Optical disks are coated with a special material which can be magnetized, but only at a relatively high temperature of approximately 300 degrees, called the Curie Point. Data is recorded by changing the magnetic field of the spot being heated by the laser. When the spot cools, it retains the orientation of the magnetic field, which represents a data bit.

TEXT BOOKS

I. Fundamentals of Multimedia, Ze-Nian Li and Mark S. Drew, Pentice Hall, Edition – 2007
II. Principles of Multimedia, Ranjan Parekh, Tata McGraw-Hill, Edition 2006