Chapter-5

Multimedia System (Pokhara University)

Chapter 5 - Data Compression

Storage Space and Coding Requirements

Storage Requirement

Uncompressed graphics, audio and video data require considerable storage capacity. For e.g. an image represented in 640 by 480 resolutions with 8 bits per pixel require 300 kilobytes of storage space. So, in order to reduce the storage requirement of the multimedia data it should be compressed. The compression should be such that there should not be a loss of quality of the data and hence compression should not reduce the information content of the data. For e.g. 90% of the raw audio can be deleted without affecting the quality of the audio.

Bandwidth Requirement

Uncompressed data transfer requires greater bandwidth or data rate for its communication. If the data is compressed there can be a considerable reduction in the bandwidth requirement for the transmission of the data.

The following examples specify continuous media and derive the amount of storage required for one second of playback:

- ✓ An uncompressed audio signal of telephone quality sampled at 8 kHz and quantized with 8 bits per sample requires 64 kbits to store one second of playback and a bandwidth of 64 kbits/sec.
- ✓ An uncompressed stereo audio signal of CD quality is sampled at a rate of 44.1 kHz quantized at 16 bits per sample require 705.6 * 10^3 bits to store one second of playback and the bandwidth requirement is 705.6 * 10^3 bits/second.

As mentioned above compression in multimedia systems is subject to certain constraints.

- ✓ The quality of the compressed data should be as good as possible.
- ✓ To make cost-effective implementation possible, the complexity of the technique used should be minimal.
- ✓ The processing of the algorithm must not exceed certain time span.

In retrieval mode application, the following demands arise:

- ✓ *Fast forward and backward data retrieval* should be possible with simultaneous display.
- ✓ *Random access* to single images and audio frames of a data stream should be possible without extending the access time more than 0.5 second.
- ✓ Decompression of data should be independent of other data units.

For both dialogue and retrieval mode, the following requirements apply

- ✓ The format should be independent of frame size and video frame rate.
- ✓ The format should support various data rate.
- ✓ There should be synchronization between the audio and video.
- ✓ Compression and decompression should not require additional hardware.
- ✓ The compression of data in one system of multimedia should ensure the decompression in the other system.

Source, Entropy and Hybrid Coding

Source Coding

Source coding takes into account the semantics and the characteristics of the data. Thus the degree of compression that can be achieved depends on the data contents. Source coding is a lossy coding process in which there is some loss of information content. For e.g. in case of speech the speech is transformed from the time domain to frequency domain. In the psychoacoustic the encoder analyzes the incoming audio signals to identify perceptually important information by incorporating several psychoacoustic principles of the human ear. One is the critical-band spectral analysis, which accounts for the ear's poorer discrimination in higher frequency regions than in lower-frequency regions. The encoder performs the psychoacoustic analysis based on either a side-chain FFT analysis or the output of the filter bank.

E.g. Differential Pulse Code Modulation, Delta Modulation, Fast Fourier Transform, Discrete Fourier Transform, Sub-band coding etc

Entropy Coding

Entropy coding is used regardless of the media's specific characteristics. The data stream to be compressed is considered to be a simple digital sequence and the semantics of the data is ignored. It is concerned solely with how the information is represented.

Run Length Coding

Typical applications of this type of encoding are when the source information comprises long substrings of the same character or binary digit. Instead of transmitting the source string in the form of independent code-words or bits, it is transmitted in the form of a different set of code-words which indicate not only the particular character or bit being transmitted but also indication of the number of characters/bits in the substring. For e.g. if the string is AAAAABBBTTTTTTMMMMMMMM, it is encoded as A!5BBBT!6M!8 (In this case there is no point in encoding characters that repeats itself less than 4 times).

Diatomic encoding is a variation of run-length encoding based on a combination of two data bytes. This technique determines the most frequently occurring pairs of bytes. For e.g. in English language "E","T","TH","A","S","RE","IN" and "HE" occurs most frequently.

Huffman Coding

Huffman coding is an example of variable length coding. It is based in the concept that the probability of occurrence of the characters is not same so different number of bits is assigned for different character. Basically in variable length coding the characters that occur most frequently are assigned fewer numbers of bits. However in order to used variable length coding the destination must know the set of code-words being used by the source. In Huffman Coding the probability of occurrence of the characters are estimated and based on this estimation code-words are assigned to the characters.

E.g let the word to be encoded be AABABBEEEEFEFC
Length of string=14

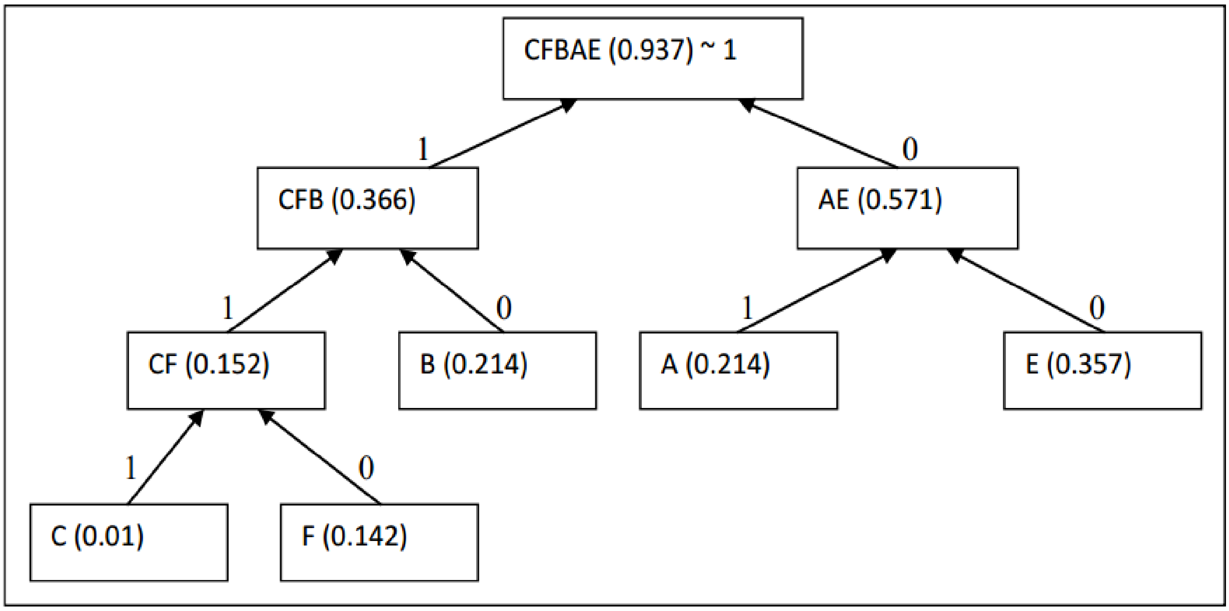
No. of A =3
Probability of occurrence of A=3/14=0.214;

No. of B=3
Probability of occurrence of B=3/14=0.214

No. of E=5
Probability of occurrence of E=5/14=0.357

No. of F=2
Probability of occurrence of F=2/14=0.142

No. of C=1
Probability of occurrence of C=1/14=0.01



C= 111
F= 110
B= 10
A=01
E=00
For videos, a Huffman table can be used for a single sequence of images, for a set of scenes or even for an entire film clip.

Arithmetic Coding

Unlike Huffman coding which used a separate codeword for each character, arithmetic coding yields a single codeword for each encoded string of characters. The first step is to divide the numeric range from 0 to 1 into a number of different characters present in the message to be sent – including the termination character – and the size of each segment by the probability of the related character.

Hybrid Coding

This type of coding mechanism involve the combine use of both the source coding and the entropy coding for enhancing the compression ratio still preserving the quality of information content. The example of Hybrid Coding includes MPEG, JPEG, H.261, DVI techniques.

Major Steps of Data Compression

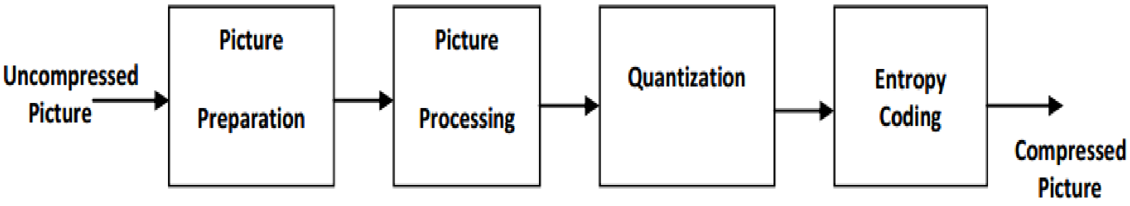


Figure: Major steps of data compression

Preparation:

Preparation involves analog to digital conversion of the picture where the image is divided into blocks of 4*4 or 8*8 pixels.

Processing:

This involves the conversion of the information from the time domain to the frequency domain by using DCT.

Quantization:

It defines discrete level or values that the information is allowed to take. This process involves the reduction of precision. The quantization process may be uniform or it may be differential depending upon the characteristics of the picture.

Entropy Encoding:

This is the lossless compression method where the semantics of data is ignored but only its characteristics are considered. It may be run length coding or entropy coding.
After compression, the compressed video stream contains the specification of the image starting point and an identification of the compression technique may be the part of the data stream. The error correction code may also be added to the stream. Decompression is the inverse process of compression.

JPEG (Joint Photographic Experts Group)

The JPEG standard for compressing continuous –tone still pictures (e.g. photographs) was developed by photographic experts working under the joint auspices of ITU, ISO and IEC. JPEG is significant in compression because MPEG or the standard for motion picture compression is just the JPEG encoding applied to each frame separately.

There are some requirements of JPEG standard and they are:

- ✓ The JPEG implementation should be independent of image size.
- ✓ The JPEG implementation should be applicable to any image and pixel aspect ratio.
- ✓ Color representation itself should be independent of the special implementation.
- ✓ Image content may be of any complexity, with any statistical characteristics.
- ✓ The JPEG standard specification should be state of art(or near) regarding the compression factor and achieved image quality.
- ✓ Processing complexity must permit a software solution to run on as many available standard processors as possible. Additionally, the use of specialization hardware should substantially enhance image quality.

Steps in JPEG Compression

Step 1: (Block Preparation)

This step involves the block preparation. For e.g. let us assume the input to be 640*480 RGB image with 24 bits/pixel. The luminance and chrominance component of the image is calculated using the YIQ model for NTSC system.

$$Y=0.30R + 0.59G + 0.11B$$
$$I= 0.60R – 0.28G – 0.32B$$
$$Q= 0.21R – 0.52G + 0.31B$$

For PAL system YUV model is used. Separate matrices are constructed for Y, I and Q each elements in the range of 0 and 255. The square blocks of four pixels are averaged in the I and Q matrices to reduce them to 320*240. Thus the data is compressed by a factor of two. Now, 128 is subtracted from each element of all three matrices to put 0 in the middle of the range. Each image is divided up into 8*8 blocks. The Y matrix has 4800 blocks; the other two have 1200 blocks.

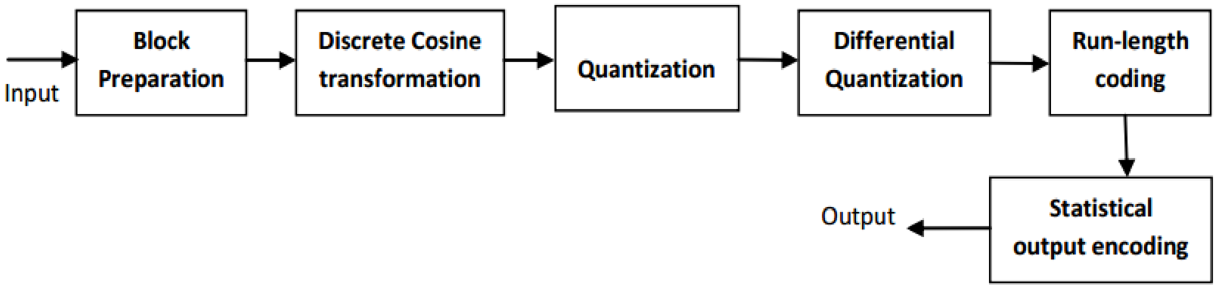

Input
Block Preparation
Discrete Cosine transformation
Quantization
Differential Quantization
Run-length coding
Statistical output encoding
Output

Figure: The operation of JPEG in lossy-sequential mode.

Step 2: (Discrete Cosine Transformation)
Discrete Cosine Transformation is applied to each 7200 blocks separately. The output of each DCT is an 8*8 matrix of DCT coefficients. DCT element (0,0)is the average value of the block. The other element tells how much spectral power is present at each spatial frequency.

Step 3: (Quantization)
In this step the less important DCT coefficients are wiped out. This transformation is done by dividing each of the coefficients in the 8*8 DCT matrix by a weight taken from a table. If all the weights are 1 the transformation does nothing however, if the weights increase sharply from the origin, higher spatial frequencies are dropped quickly.

Step 4: (Differential Quantization)
This step reduces the (0,0) value of each block by replacing it with the amount it differs from the corresponding element in the previous block. Since these elements are the averages of their respective blocks, they should change slowly, so taking the differential values should reduce most of them to small values. The (0,0) values are referred to as the DC components; the other values are the AC components.

Step 5: (Run length Encoding)
This step linearizes the 64 elements and applies run-length encoding to the list. In order to concentrate zeros together, a zigzag scanning pattern is used. Finally run length coding is used to compress the elements.

Step 6: (Statistical Encoding)
Huffman encodes the numbers for storage or transmission, assigning common numbers shorter codes than uncommon ones.

JPEG produces a 20:1 or even better compression ratio. Decoding a JPEG image requires running the algorithm backward and thus it is roughly symmetric: decoding takes as long as encoding.

Lossy Sequential DCT-based Mode of JPEG

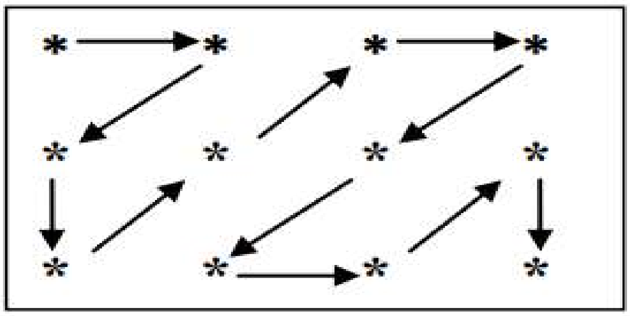
Image Processing
It basically involves the block preparation where the image samples are grouped into 8*8 pixels and passed to the encoder. Then Discrete Cosine Transformation is applied to the blocks where the pixel values are shifted into the range [-128,127] with zero as the center. Each of these values is then transformed using *Forward DCT (FDCT).* DCT is similar to *Discrete Fourier Transformation* as it maps the values from the time to the frequency domain.

Quantization

The JPEG application provides a table of 64 entries. Each entry will be used for the quantization of one of the 64 DCT-coefficients.

Entropy Encoding

During the initial step of entropy encoding, the quantized DC-coefficients are treated separately from the quantized AC-coefficients.



- ✓ The DC-coefficient determines the basic color of the data units.
- ✓ The DCT processing order of the AC coefficients involves the zigzag sequence to concentrate the number of zeros.

JPEG specifies Huffman and arithmetic encoding as entropy encoding methods. However, as this is lossy sequential DCT-based mode, only Huffman encoding is allowed. In lossy sequential mode the framework of the whole picture is not formed but parts of it are drawn i.e. sequentially done.

Expanded Lossy DCT-based Mode

It differs from the sequential mode in terms of number of bits per sample. Here 12 bits along with 8 bits per sample can be used.

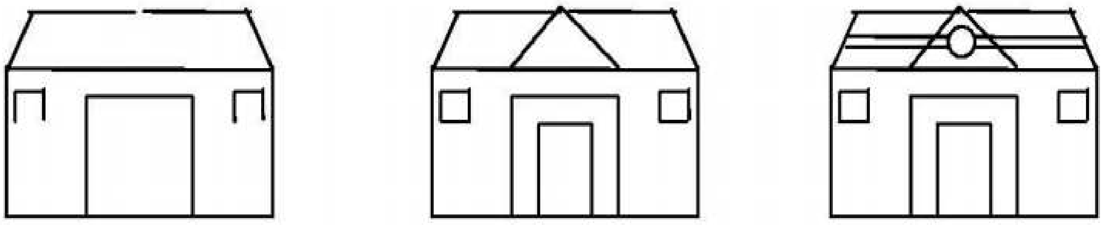


Figure: Progressive Picture Presentation

For the expanded lossy DCT-based mode, JPEG specifies *progressive encoding* in addition to sequential encoding. At first, a very rough representation of the image appears which is progressively refined until the whole image is formed. This progressive coding is achieved by layered coding.

Progressiveness is achieved in two different ways:

- ✓ By using a *spectral selection* in the first run only, the quantized DCT coefficients of low frequencies of each data unit are passed in the entropy encoding. In successive runs, the coefficients of higher frequencies are processed.
- ✓ *Successive approximation* transfers all of the quantized coefficients in each run, but single bits are differentiated according to their significance. The most-significant bits are encoded first, then the less-significant bits.

Hierarchical Mode

This mode uses either the lossy DCT-based algorithms or the lossless compression technique. The main feature of this mode is the encoding of an image at different resolutions, i.e., the encoded data contains images at several resolutions. This involves the sampling of images at higher resolutions at first and then subsequently reducing the horizontal and vertical resolution. The compressed image is subtracted from the previous result. The process is repeated until the full resolution is reduced considerably and no further compression is possible.

This requires more storage but as images at different resolution are available the systems using low resolution need not decode the image as the low resolution image is easily available.

MPEG (Motion Picture Expert Group)

The Motion Pictures Expert Group was formed by the ISO to formulate a set of standards relating to a range of multimedia applications that involve the use of video with sound. The coders associated with the audio compression part of these standards are known *MPEG audio coders* and a number of these use perceptual coding. MPEG can deliver a data rate of at most 1856000 bits/second, which should not be exceeded. Data rates for audio are between 32 and 448 Kbits/second; this data rate enables video and audio compression of acceptable quality.

The MPEG standard exploits the other standards and they are

- ✓ JPEG:
  JPEG is used as the motion picture is a continuous sequence of still image.

- ✓ H.261:
  H.261 video compression standard has been defined by the ITU-T for the provision of video telephony and videoconferencing services over an Integrated Services Digital Network (ISDN). It specifies two resolution formats with an aspect ratio of 4:3 are specified. *Common Intermediate Format* (CIF) defines a luminance component of 288 lines, each with 352 pixels. The chrominance components have a resolution with a rate of 144 lines and 176 pixels per line. *Quarter CIF* (QCIF) has exactly half of the CIF resolution i.e., 176*144 pixels for the luminance and 88*72 pixels for the other components.

MPEG video uses video compression algorithms called *Motion-Compensated Discrete Cosine Transform* algorithms. The algorithms use the following basic algorithm

- ✓ *Temporal Prediction:* It exploits the temporal redundancy between video pictures.
- ✓ *Frequency Domain Decomposition:* It uses DCT to decompose spatial blocks of image data to exploit statistical and perceptual spatial redundancy.
- ✓ *Quantization:* It reduces bit rate while minimizing loss of perceptual quality.
- ✓ *Variable-length Coding:* It exploits the statistical redundancy in the symbol sequence resulting from quantization as well as in various types of side information.

As far as audio compression is concerned the time-varying audio input signal is first sampled and quantized using PCM, the sampling rate and number of bits per sample being determined by the specific application. The bandwidth that is available for transmission is divided into a number of *frequency subbands* using a bank of *analysis function* which because of their role, are also known as *critical-band filters.*

Video Encoding

Video is nothing but simply a sequence of digitized pictures. The video that MPEG expects to process is composed of a sequence of frames or fields of luma and chroma.

Frame-Based Representation:

MPEG-1 is restricted to representing video as a sequence of frames. Each frame consists of three rectangular arrays of pixels, one for the luma (Y, black and white) component, and one each for the chroma (Cr and Cb, color difference) components. The chroma arrays in MPEG-1 are sub-sampled by a factor of two both vertically and horizontally relative to the luma array.

Field-Based Representation:

MPEG-2 is optimized for a wider class of video representations, including, most importantly, field-based sequences. *Fields* are created by dividing each frame into a set of two interlaced fields, with odd lines from the frame belonging to one field and even lines to the other. The fields are transmitted in interlaced video one after the other, separated by half a frame time.

MPEG provides 14 different image aspect ratios per pixel which are coded in the data stream. The image refresh frequency is also encoded in the data stream. Eight frequencies are defined: 23.976Hz, 24 Hz, 25 Hz, 29.97 Hz, 30 Hz, 50 Hz, 59.94 Hz and 60 Hz.

The technique that is used to exploit the high correlation between successive frames is to predict the content of many of the frames. Instead of sending the source video as a set of individually-compressed frames, just a selection is sent in this form and, for the remaining frames, only the differences between the actual frame contents and the predicted frame contents are sent. This operation is known as ***motion estimation*** and, since to indicate any small differences between the predicted and actual positions of the moving segments involved. The latter is known as ***motion compensation.***

Types of Image Frames in MPEG

There are two basic types of compressed frame: those that are encoded independently and those that are predicted. The first are as *intracoded frames* or *I frames.* In practice, there are two types of predicted frames: *predictive* or *P-frames* and *bidirectional* or *B-frames* and because of the way they are derived, the latter are also known as *intercoded* or *interpolation frames.*

I-frames (Intracoded Images):

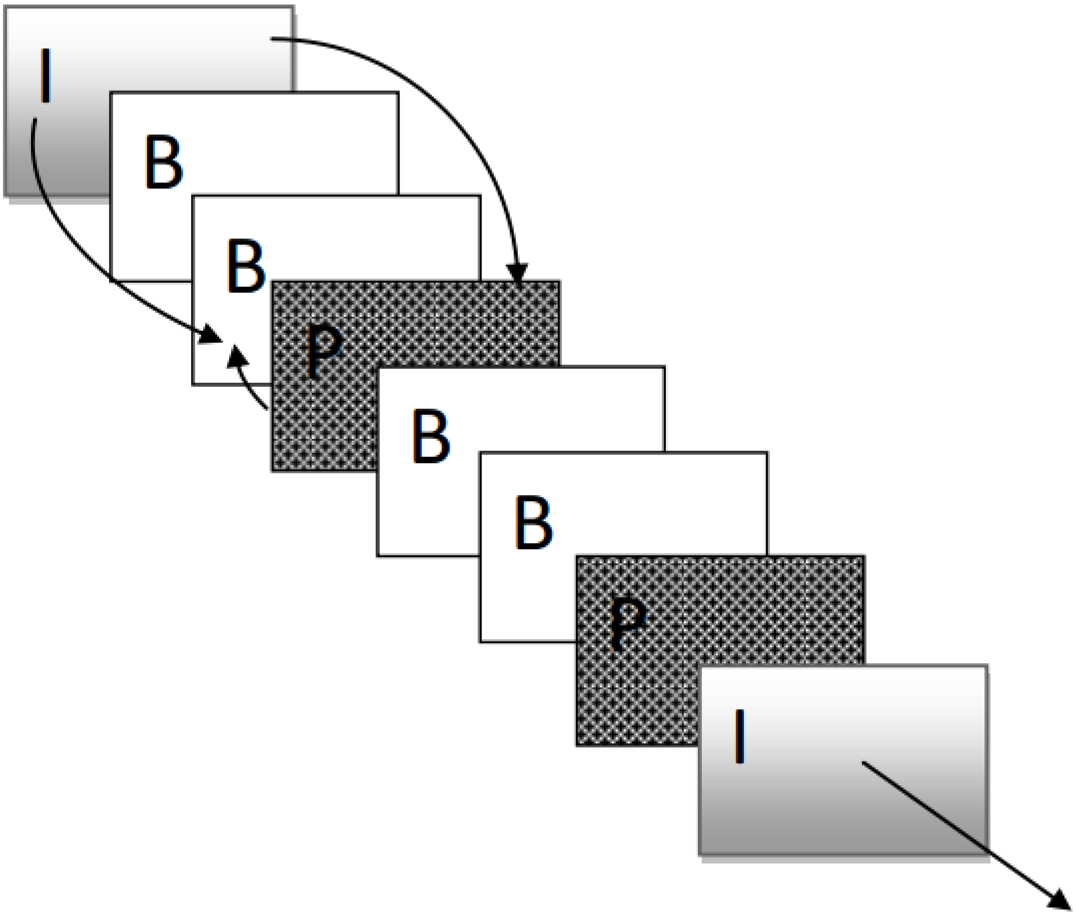
I-frames are encoded without reference to any other frames. Each frame is treated as a separate (digitized) picture and the Y, C_b , and C_r matrices are encoded independently using the JPEG algorithm.

The level of compression obtained with I-frames is relatively small. I-frames must be present in the output stream at regular intervals in order to allow for the possibility of the contents of an encoded I-frame being corrupted during transmission. The number of frames/pictures between successive I-frames is known as a *group of pictures* or *GOP.*

P-frames (Predictive-Coded Frames):

The encoding of a P-frame is relative to the contents of either a preceding I-frame or a preceding P-frame. As indicated, P-frames are encoded using a combination of motion estimation and motion compensation and hence significantly higher levels of compression can be obtained. In practice, the number of P-frames between each successive pair of I-frames is limited since any errors present in the first P-frame will be propagated to the next. The number of frames between a P-frame and the immediately preceding I- or P-frame is called the *prediction span.* It is given the symbol M and typical values range from 1 through 3.

P-frames consist of I-frame macro blocks and six predictive macro blocks. The coder must determine if a macro block should be coded predicatively or as a macro block of an I-frame and furthermore, if there is a motion vector that must be encoded. A P-frame can contain macro blocks that are encoded using the same technique as I-frames.



B-frames (Bi-directionally predictive-coded frames):
Motion estimation involves comparing small segments of two consecutive frames for differences and, should a difference be detected, a search is carried out to determine to which neighboring segment the original segments has moved. In order to minimize the time for each search, the search region is limited to just a few neighboring segments. Some applications may involve very fast moving objects; however, it is possible for a segment to have moved outside the search region. To allow for this possibility, in applications such as movies, in addition to P-frames, second types of frames are used called B-frames.

The content of the B-frames are predicted using search regions in both past and future frames. In addition to allowing for occasional fast moving objects, this also provides better motion estimation when, for example, an object moves in front of or behind or another object. B-frames provide the highest level of compression and, because they are not involved in the coding of other frames, they do not propagate errors.

[To perform the decoding operation, the received information relating to I-frames can be decoded immediately it is received in order to recreate the original frame. With P-frames, the received information is first decoded and the resulting information is then used, together with the decoded contents of the preceding I- or P-frame, to derive the decoded frame contents. In the case of B-frames, the received information is first decoded and the resulting information is then used, together with both the immediately preceding I- or P-frame contents and the immediately succeeding P- or I-frame contents, to derive the decoded frame contents.]

D-frames (DC-Coded Frames):
D-frame has been defined for use in movie/video-on-demand applications. D-frames are used for display in fast-forward or fast-rewind modes. D-frames are inserted at regular intervals throughout the

stream. These are highly compressed frames and are ignored during the decoding of P- and B-compression algorithm, the DC coefficient associated with each 8 * 8 block pixels- both for the luminance and the two chrominance signals- is the mean of all the values in the related block. Hence by using only encoded DC coefficients of each block of pixels in the periodically inserted D-frames, a low-resolution sequence of frames is provided each of which can be decoded at the higher speeds that are expected with the rewind and fast-forward operations.

Audio Encoding:
The time-varying audio input signal is first sampled and quantized using PCM, the sampling rate and number of bits per sample being determined by the specific application. The bandwidth that is available for transmission is divided into a number of *frequency subbands* using a bank of *analysis filters* which, because of their role, are also known as *critical-band filters.* Each frequency subbands is of equal width and, essentially, the bank of filters maps each set of 32 PCM samples into an equivalent frequency samples, one per subband, Hence each is known as a *subband sample* and indicates the magnitude of each of the 32 frequency components that are present in a segment of the audio input signals of a time duration equal to 32 PCM samples.

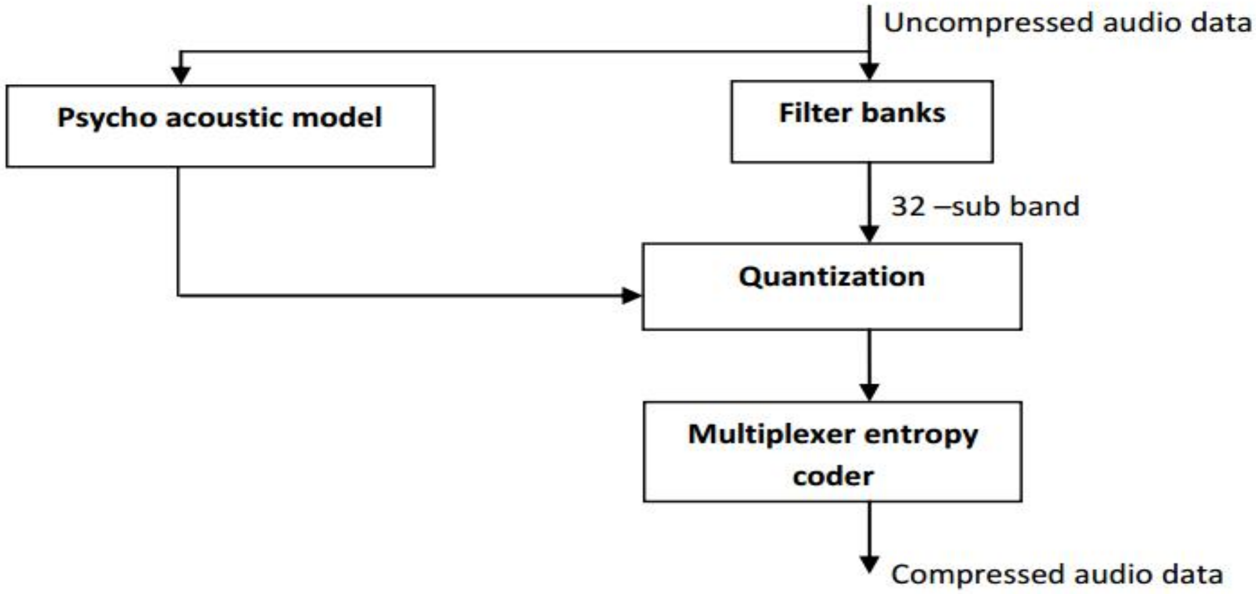


Figure: MPEG basic steps for audio encoding

In addition to filtering the input samples into separate frequency subbands, the analysis filter bank also determines the maximum amplitude of the 12 subband samples in each subband. Each is known as the *scaling factor* for the subband and these are passed both to the psychoacoustic model and together with the set of frequency samples in each subband, to the corresponding quantizer block. The 12 sets of 32 PCM samples are first transformed into an equivalent set of frequency components using a mathematical technique known as the *discrete Fourier transform (DFT).* In addition, the set of scaling factors are used to determine the quantization accuracy- and hence bit allocations- to be used for each of the audible components.

Three different layers of encoder and decoder complexity and performance are defined. DCT is applied for frequency domain transformation of audio. At a higher noise level, a rough quantization is performed, and at a lower noise level, a finer quantization is applied. The quantized spectral portions of layers one and two are PCM-encoded and those of layer three are Huffman-encoded. The layers support different maximal bit rates: layer 1 allows for a maximal bit rate of 448 Kbits/second, layer 2 for 384 Kbits/second and layer 3 for 320 Kbits/s.

DVI (Digital Video Interactive)

DVI is a technology that includes coding algorithms. The fundamental components are a VLSI chip set for the video subsystem, a well specified data format for audio and video files, an application user interface to the audio-visual kernel and compression, as well as decompression, algorithms. For encoding audio standard signal processor is used. Processing of images and video is performed by a video processor.

Audio and Still Image encoding:

Audio signals are digitized using 16-bits per sample. Audio signals may be PCM-encoded or compressed using the adaptive differential pulse coded modulation (ADPCM) technique. Supported sampling frequencies are: 11025Hz, 22050Hz and 44100 Hz for one or two PCM-coded channels. And 8268Hz, 31129Hz, 33075Hz for ADPCM.

For Still Images, DVI assumes an internal digital YUV format for image preparation. Any video input signal must first be transformed into this format. The color of each pixel is split into luminance component and the two chrominance components (U and V). The luminance represents the gray scale image. With RGB, DVI computes the YUV signal using the following relationship.

$$Y=0.30R+0.59G+0.11B$$
$$U=B-Y$$
$$V=R-Y$$

It leads to:

$$U=-30R-0.59G+0.89B$$
$$V=0.70R-059G-0.11B$$

DVI Determines the components YUV according to the following:

$$Y=0.299R+0.587G+0.144B+16$$
$$U=0.577B-0.577Y+137.23$$
$$V=0.730R-0.730Y+139.67$$

DVI is able to process image in the 16-bit YUV format and the 24-bit YUV format. The 24-bit YUV format uses 8 bits for each component. The 16-bit YUV format coded the Y components of each pixel with 6 bits and the color difference components with 5 bits each.

There are 2 bitmap formats: Planer and Packed.

Planer:

All data of the Y component are stored first, followed by the U component values and then all V values.

Packed:

For the packed bitmap format, the Y, U, and V information of each pixel is stored together by the data of the next pixel.

References:

- ✓ Multimedia: Computing, Communications and Applications", Ralf Steinmetz and Klara Nahrstedt, Pearson Education Asia
- ✓ "Multimedia Communications, Applications, Networks, protocols ad Standards", Fred Halsall, Pearson Education Asia
- ✓ "Multimedia Systems", John F. Koegel Buford, Pearson Education Asia

Assignments:

(1) How is source coding different from entropy encoding? Describe about the MGPEG video compression.

(2) What are the different types of compression technique used? Explain in detail any one Source encoding technique used for data compression.

(3) How is source coding different from entropy encoding? Describe about the JPEG compression.

(4) What is data compression? Why multimedia data should be compressed? Describe the JPEG compression with its different modes.