Compression

Compared to text files, other media files like image, audio + video take up a huge amount of disk space. It is extremely difficult working with such huge files ∵ of 2 reasons:-

(1) Due to their huge sizes, storage requirements increases rapidly with multiple such media files thereby ↑ing the cost. During mm dev. work, one can quickly run out of disk space

(2) Even if there is adequate storage space, such files require a large data transfer rate that may be beyond the capabilities of both the processor + hard disk.

Due to these facts, it is customary to subject media files to a process called compression.

Compression reduces their file sizes using mathematical algo, after which it becomes much easier to manipulate these files. The amt. of compression depends on both original media data as well as compression

Technique applied.

CODEC

After an analog quantity has been digitized, it is stored on the disk as a digital file. Such files are referred to as raw or uncompressed media data. To compress the file and reduce its size it needs to be filtered thus a specialized s/w called CODEC which stands for Compression/Decompression or Coder/Decoder. The s/w reads the media data + applies the mathematical algo to reduce its size.

Diag, - This algo works by finding the redundant info within the media files.

- Redundant info is that which can either be discared without affecting the media quality by appreciable measures of data can be expressed in more compact form.

- Amt of reduction depends on large no. of factors including both media data and CODEC used.

- The compressed media is then again stored on the disk or other storage media frequently as a diff. type of file with another extension.

- When the media needs to be played back, either as stand-alone or as part of presentation, the same CODEC is again used to decompress the data which was previously compressed. Hence, the compression process needs to be essentially reversible. However depending on type of CODEC used, the decompressed data may or may not be exactly identical to original uncompressed data.

- Decompression is necessary as the playback devices like the monitor or speakers cannot directly handle compressed data. After compression the raw digital data is usually converted to analog form before being played back.

Types of Compression

Based on methodologies compression can be div. into

Lossless vs Lossy compression

* Lossless compression :- means the original data is not changed permanently during compression. After decompression therefore the original data can be retrieved back exactly. CODECS used for lossless compression are called lossless CODECs.

* The algos within these CODECs attempts to represent the existing information in a more compact form without actually discarding any data.

- Adv :- Original data stays intact without degradation of quality and can be reused.

- Disadv :- Compression achieved is not very high + may be typically of order of 2 to 5 times.

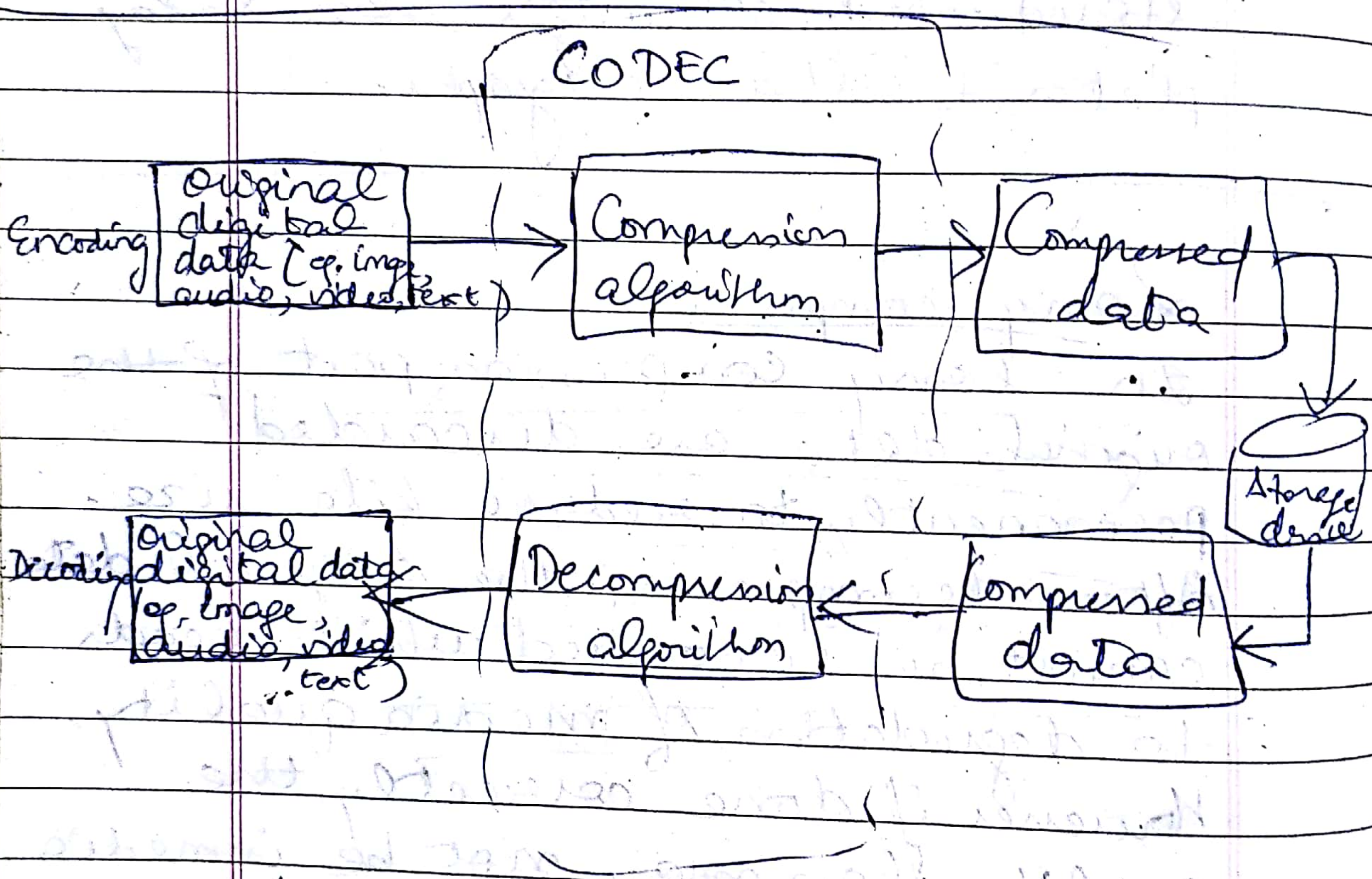
Lossless compression is used where media quality is of utmost importance and larger space requirements may be justified.
eg. quality plays an imp. part in diagnosing faults from digitally stored medical images like X-ray plates + ultra-sonographs.

Lossy compression

In lossy compression, parts of the original data are discarded permanently to reduce file size. After decompression, the original data cannot be recovered which leads to degradation of media quality. However if done correctly the quality loss may not be immediately apparent because of the limitations of human eye and ears + also because of tendency of human senses to bridge gaps in information.

An imp. consideration is how much info^n can be lost before human eye can tell the diff. Compression ratios of order of 10 to 50 times can be achieved.

Lossy compression is usually used where media quality may be sacrificed to a certain extent for reducing space requirements like in mm presentation & web page content.



How a CODEC works

Types of Compression

based on kind of redundancies exploited compression can be divⁿ into 2

Intraframe   Interframe

Intraframe compression:- It is applicable within a still image or a single video frame. Here, spatial redundancies are detected & exploited to reduce the file size. Such redundancies occur when diff. portion of an image are identical or similar to each other. Eg. an area of flat color in an image has pixels of identical values.

Interframe compression

It exploits the redundancies b/w adjacent frames in a video sequence referred to as temporal redundancy. Such redundancies occur when subsequent frames of a video sequence are also identical or very similar.

eg. in video of a news reader, most of the infoⁿ b/w adjacent frames remains unchanged with minor variations in lip movement & facial expression of reader.

Types of Redundancies

Statistical (Lossless) — Psycho-visual (Lossy)

Statistical redundancy

Here some statistical relationships exits b/w media data.
It may be related either to
i. actual infoⁿ content. eg. intensity values of adjacent pixels along a row in an image may be similar Or
it may be related to representation of information. eg. frequently occurring data may be represented by small no. of bits.

Statistical relationships within static media like images are spatial in nature, while those in time-varying media like sound are temporal in nature. In video clips, they may both be spatial-temporal.

Psycho-visual redundancy
It originates from the characteristics of the HVS. In the HVS, visual information is not perceived equally, some infoⁿ may be more imp. than other infoⁿ, so, if we apply fewer data to represent less important visual infoⁿ, perception will not be affected. eg, some colors in an image or specific sounds within an audio clip may not be detectable by human senses. Such infoⁿ can be eliminated to reduce file size, without an appreciable variation in quality perception.

Lossless / Statistical Compression Techniques.

Lossless compression techniques are also known as Entropy Coding. It refers to techniques which do not take into account the nature of infoⁿ to be compressed. They treat all data as sequence of bits + tries to achieve compression by statistical manipulation or rearrangement of bits i.e they ignore semantics of infoⁿ to be compressed.

Scanned by CamScanner

classmate

Date ______
Page ______

~~Entropy~~

Entropy

It is measure of energy in physical system that cannot be used to do useful work. It is also used to depict disorder in the system. The SI unit of entropy is Joules per Kelvin (J/K) same as heat capacity.

Entropy is used to depict randomness in a signal which in turn provides a measure of how much useful info^n can be carried by a signal.

eg. suppose a signal is used to transmit a text string from a source to a destination. It cannot be predicted with accuracy what the next char would be.

Entropy in terms of a discrete random event x having possible states 1 -- n can be defined as:-

$$\text{Entropy } H(x) = -\sum_{i=1}^{n} p(i) \log_2 p(i)$$

where n is the no. of diff. states and p(i) is the prob. of state i. This is called Shannon's law of entropy and can be used to provide a measure of min^n avg. no. of bits reqd. to transmit a particular byte stream.

The efficiency of a particular encoding scheme is often computed as a ratio of entropy of source to avg no. of bits/codeword that are reqd. with scheme.

Latter is computed acc. to relation:

Avg. no. of bits/codeword =

$$\sum_{i=1}^{n} N_i \, p(i)$$

where N_i is no. of bits, p(i) is prob. of occurrence of i^{th} symbol.

Run Length Encoding (RLE)

Here, any sequence of repetitive characters may be replaced by a more compact form. A series of n successive characters may be replaced by a single instance of char. and no. of occurences.
To depict the " " " that it has a special meaning + not part of normal text, a special char is used as a flag.

Eg. Uncompressed data:
UNNNNNN IMANNHEIM
Compressed data:
U!6NIMANNHEIM
↳ flag

NNNNNN $\xrightarrow[\text{by}]{\text{replaced}}$!6N

This method should only be used if no. of occurences is equal to or more than a specific no. (eg. in eg, it was 4).

Huffman Coding (or statistical coding)

* This method consists of identifying the most frequent bit or byte patterns in a given file + coding these patterns with fewer bits than initially represented.
* Less frequent patterns will be coded with more bits whereas most frequent patterns will use shorter codes.
* A table of correspondence called code-book b/w initial patterns + their new representations must be available at both encoding + decoding ends.
* This table is generated by analyzing frequencies of the occurences of each char in the encoding process.

AAAAAAAAAAAA BBCD

12 A's, 2 B's, 1 C, 1 D

~~Freq of~~

Prob. of occurrence of 4 char are:-

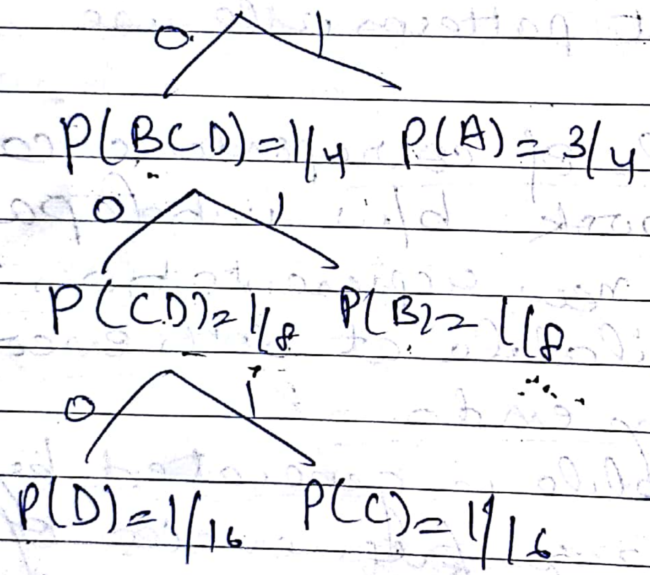
$P(A) = 3/4$, $P(B) = 1/8$,

$P(C) = P(D) = 1/16$

To transmit string without compression using 7-bit ASCII code requires total

$7 \times 16 = 112$ bits

↓ing order of freq.



Codewords are determined from paths from root node to leaf node. Codewords are:-

A : 1

B : 01

C : 001

D : 000

classmate Date Page

After compression, total no. of bit reqd. to represent document is,

$12 \times 1 + 2 \times 2 + 1 \times 3 + 1 \times 3 = 22$ bits.

Compression ratio $= \frac{112}{22}$

= about 5 : 1

Q1

A 12, B 2, C 1, D 1

$$\begin{array}{lll} A\,12 & \longrightarrow A\,12 & \longrightarrow A\,12\,(1) \\ B\,2 & \longrightarrow B\,2\,(1) & \longrightarrow 4\,(0) \\ C\,1\,(1) & \longrightarrow 2\,(0) & \\ D\,1\,(0) & & \end{array}$$

Codewords from lt. to rt. + reversed

A : 1	A : 1
B : 10	B : 01
C : 100	C : 001
D : 000	D : 000

Reverse

Date ______
Page ______

Q. A document contains letters A thru F with frequencies as indicated:
A: 0.25, B = 0.1, C = 0.2,
D: 0.15, E = 0.26, F = 0.04

Use Huffman Coding to derive a codeword set.

A: 0.25
B: 0.1
C: 0.2
D: 0.15
E: 0.26
F: 0.04

↓ decreasing order

E: 0.26
A: 0.25
C: 0.20
D: 0.15
B: 0.10 (1)
F: 0.04 (0)

E: 0.26
A: 0.25
C: 0.20
D: 0.15 (1)
: 0.14 (0)

0.29

E 0.26

A 0.25 (1)

C 0.20 (0)

0.45

0.29 (1)

E 0.26 (0)

0.55 (1)

0.45 (0)

A : (1) (0) → 01

B : (1) (0) (1) (1) → 1101

C : (0) (0) → 00

D : (1) (1) (1) → 111

E : (0) (1) → 10

F : (0) (0) (1) (1) → 1100

E 0.26

A 0.25

C 0.20 (1)

0.29 (0)

E 0.26

A 0.25 (1)

0.49 (0)

E 0.26

Page ________

Lempel-Ziv (LZ) coding

* LZ coding uses codes document by considering a string of characters at a time instead of single characters.
* Both encoder + decoder holds a table containing all possible words in the document.
* As each word occurs in the text encoder stores the index of the word in table instead of actual word. The decoder then uses the index to access the corresponding word from table in order to reconstruct the document.
* Table is used as a dictionary and the LZ algo is called as dictionary-based compression algo.
* Most word processing packages have a dictionary associated with them which is used for both spell checking & for compression of text. Typically they contain in region of 25,000 words.
* To generate an index for each word would require a 15-bit no ($2^{15} = 32,768$).

Classmate
Date
Page

eg. to encode word 'compression', would just require 15 bits instead of 77 bits, corresponding to 11 7bit ASCII codes

- Basic req. of LZ algo is that a copy of dictionary is held by both encoder + decoder.

Lempel - Ziv - Welch (LZW) Coding
- Efficiency of LZ coding depends on how many of words in document to be encoded actually is included within the dictionary.
If small subset of words could be replaced by an index no. the overall compression factor might be quite low as remaining words would need to be stored without compression.

To address this problem a variation of the scheme has been developed known as LZW coding algo.

* This scheme allows dictionary to be built up dynamically by encoder + decoder as compressed text is stored.

* Initially dictionary contains only ASCII char set. The remaining entries are then dynamically added as each word is encountered.

eg. suppose dictionary is a 8-bit table + therefore containing 256 entries.

* The first 128 table cell are populated with ASCII char set, while remaining 128 is empty initially.

* When 1st word is encountered in document, at 1st the word is interpreted as a collection of characters + each char is replaced by its index in table.

* At end of word when white space char is encountered, encoder stores word in next available free slot of table + uses that index no. to represent word hence forth.

Scanned by CamScanner

* This procedure is repeated for each word, however prior to encoding characters individually encoders checks if word is already stored in table.

* If it is word is represented by its index no. only. Decoder has access to same table + uses index no to retrieve original word.

* LZW provides a technique for building a dynamic table length. Initially no. of entries is kept low, but if space is used up, the no of slots can be ↑ed.

eg. In above eg, if all 256 entries are used, encoder + decoder double size of table to 512 entries.

Q. LZ compression scheme is used to compress a text document. The dictionary table contains 1024 entries. The avg. char/word in document is 5. Calculate compression achieved.

1024 entries, so rep. by 10 bit codeword (as $2^{10} = 1024$)

Index of each word is 10 bit no.

If compression is not there, each char is rep. by 7

As 5 char/word, so each word rep. by 35 bit no. ($5 \times 7 = 35$)

Compression ratio $= \frac{35}{10} = 3.5:1$

Lossy / Perceptual Compression Technique

It is also known as Source Coding. Here, the human audio-visual capabilities + limitations are considered to isolate portions of the media that cannot be perceived by any human senses + then selectively discard them to reduce file size. If done correctly, the user will not be able to detect any perceptual change b/w original + changed one.
eg, in an image certain colours which cannot be perceived by human eye could be discarded while in sound clip certain freq. not audible by human ear may be filtered out

They produce higher compression ratios than entropy encoding.
The main processes involved in coding are:-
1. transform coding → which convert the i/p data into a diff. form from where it is easier to identify redundancies
2. Psycho-analysis:- which identifies the redundant portions according to the known characteristics of human sense organs.

JPEG ~~Image Coding standard~~

* JPEG (Joint Photographers Expert Group) is a compression standard for images.
* Its performance depends on the complexity of image. Normally image compression ratio is around 20:1.
* It does not perform well on line drawings.

Modes of operation

1. Sequential Encoding:- It is the simplest lossy mode. It involves a single scan, left to rt + top to bottom.

2. Progressive Encoding:- It is also a lossy mode. It also image to be built in multiple scans.

3. Lossless Encoding:- It does not have any data loss.

4. Hierarchical Encoding:- It has multiple resolution levels which can be decompressed separately.

Overview of Compression Steps

Compression process consists of foll^n steps:
1. Preparation of data blocks.
2. Source encoding steps involving forward DCT + quantization.
3. Entropy encoding steps involving RLE and Huffman encoding.

Decompression process is made up of foll^n

(1) Entropy decoding steps involving Huffman decoding + Run length decoding.

(2) Source decoding steps involving dequantization + inverse DCT.

Block Preparation

* An image is represented by 1 or more 2D array of pixel values.
  3 2D → colour

* For continuous tone grayscale image, using a CLUT, there will be a single array consisting of 8 bit values for each pixel of image.

* For color image without a CLUT, there will be 3 2D array of pixel values corresponding to R, G and B subcomponents.

* The block preparation step breaks each 2D array of image into individual blocks of 8×8 pixels / block.

Scanned by CamScanner

Discrete Cosine Transform (DCT)

* It is used to transform each block from spatial domain to the freq. domain.

* Each block is composed of 64 values which represents amplitude of sampled signal. This amplitude is a function of 2 spatial coordinates
  $a = f(x, y)$
  where x and y are 2 spatial dimensions and a represents the amplitude of pixel at sampled posn (x, y).

* After DCT, this fn is turned into another fn $c = g(F_x, F_y)$ where c is coefficient and F_x and F_y are respective spatial freq. for each direction.

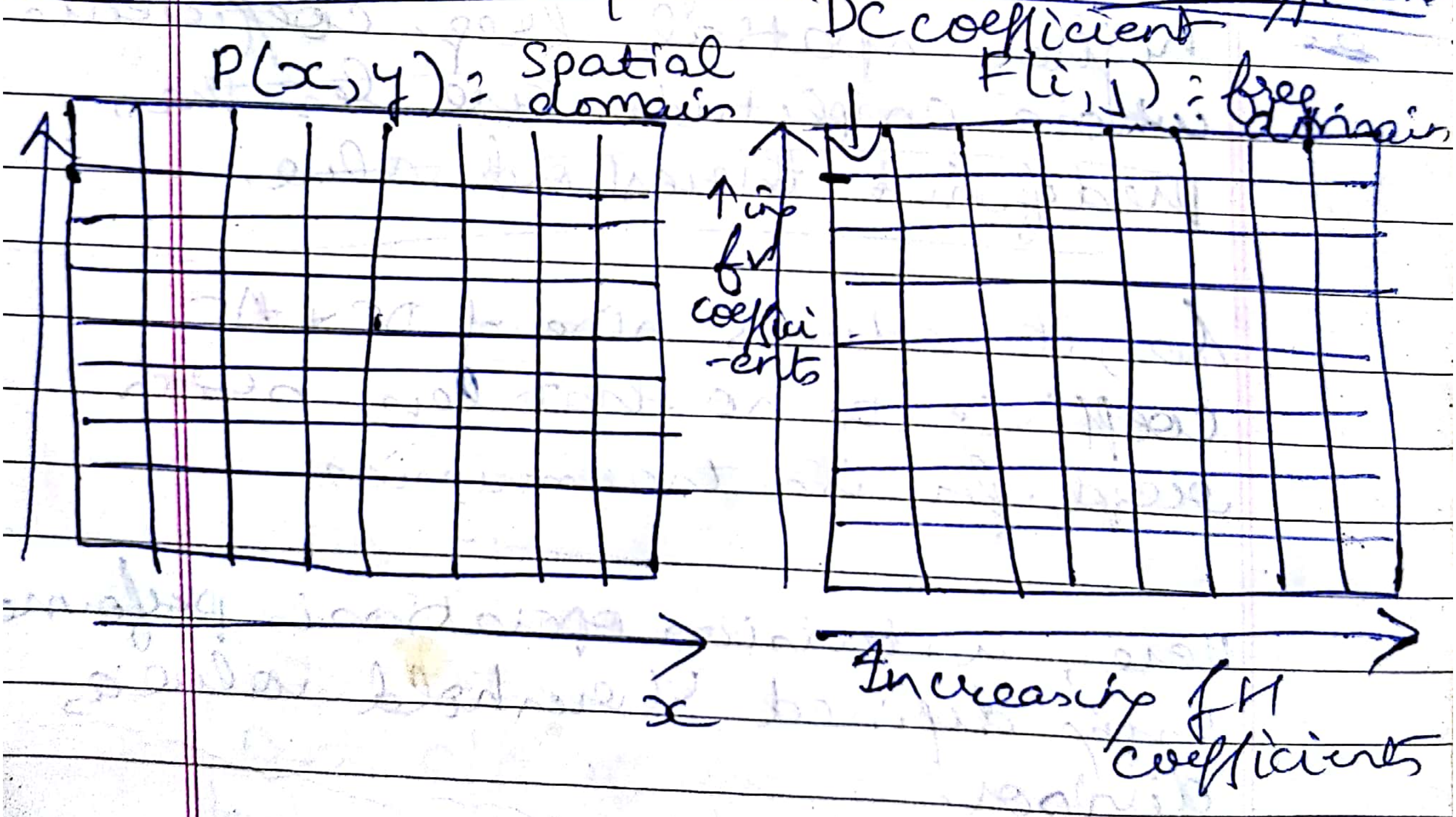
DCT of each 8×8 block is computed ~~using~~ as:

$$R[i,j] = \sum_{x=0}^{7} \sum_{y=0}^{7} P[x,y] \cdot \cos[\pi i(2x+1)/16] \cdot \cos[\pi j(2y+1)/16]$$

where x, y, i and j all vary from 0 to 7.

* For i = j = 0, two cosine terms become 0 and cos(0) = 1. This term obtained is called DC coefficient.

* Since values in other locations of transformed array have a freq. coefficient associated with them, they are called AC coefficient.



fv : Vertical freq.
fH : Horizontal "

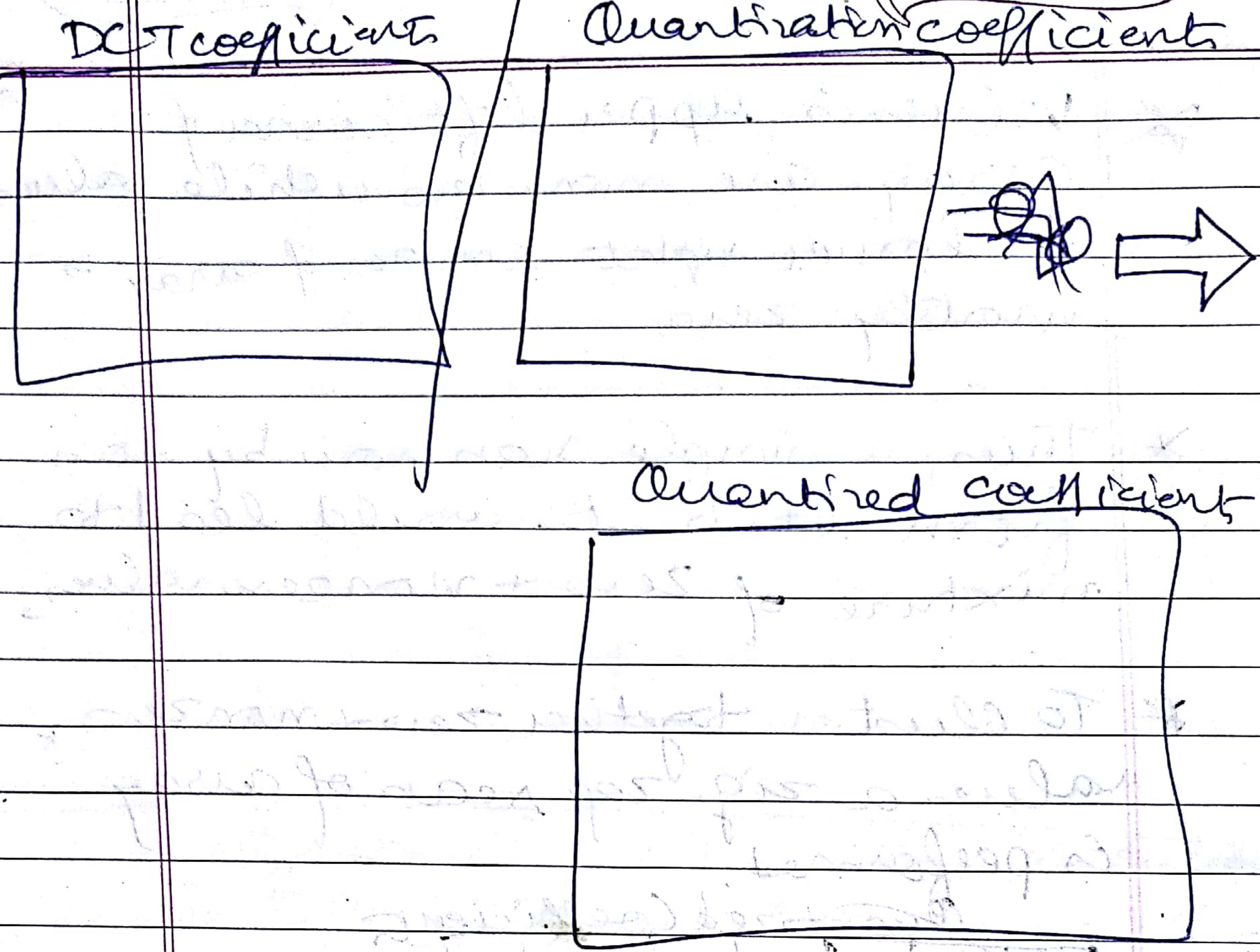
Quantization

So far no info is lost. If the image is stored as such + inverse Transform is applied, initial image is recovered in exact form. The human eye responds primarily to DC coefficients. If magnitude of a higher freq. coefficient is below a certain threshold, eye will not detect it. This property is exploited in quantization by dropping the higher spatial freq. coefficients whose amplitudes are less than predefined threshold value.

So, it reduces size of DC + AC coefficients so that less b/w is reqd. for its transmission.

Here, a division operation is performed using defined threshold value as divisor.

If resulting quotient is zero the coefficient is less than threshold value while if it is non zero this indicates no. of times the coefficient is greater than threshold. These values are stored in 8X8 Table called quantization



Zig-Zag Scan

After DCT stage, ~~the~~ remaining stage include lossless entropy encoding. The encoding coding operate on 1D string of values ie vector. But o/p of quantization is however a 2D array. So array is converted to 1D array. This is called vectoring.

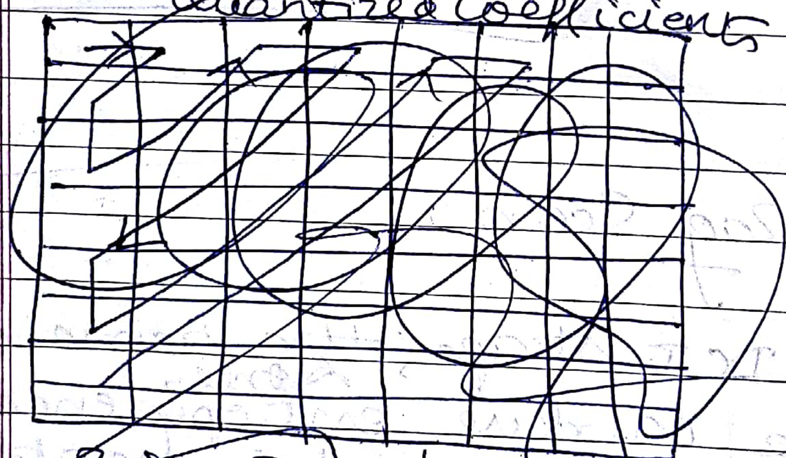
Table

* Values in upper left corner of array are non-zero while values in lower right corner of array is mostly zero.

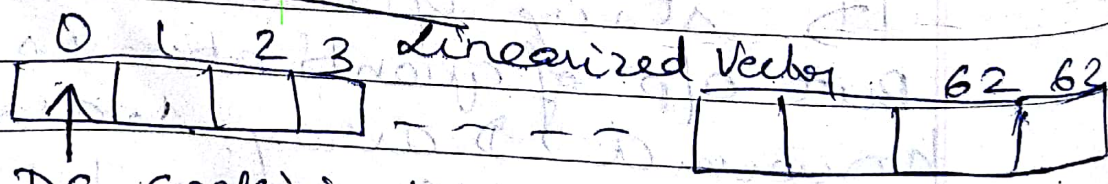
* Thus, a simple scan row by row from lt. to rt. would lead to mixture of zero + non zero values.

* To cluster together zero + nonzero values a zig-zag scan of array is performed.

Quantized Coefficients



Zig-Zag Scan



DC coefficient

AC coefficient in ↑ing order of freq.

Quantized Coefficients

Zig-Zag Scan

(5) Differential Pulse code modulation
DPCM Encoding

* There is one DC coefficient / block.
* DC coefficient is measure of mean value of 64 values of that block.
* The sequence of DC coefficients is encoded in DPCM mode.
* This means that the difference b/w the DC coefficient of each block and adjacent block is computed + stored.
* This reduces the no. of bits reqd. to encode the relatively large magnitude of DC coefficient.

RLE Encoding

* After quantization steps, most of the coefficients have been reduced to zero values.
* The surviving values of each block is to be stored. Many of these values will be zero and to take adv. of this, they are run length encoded.
* Due to zig-zag scan of AC coefficients of each block have been grouped in a way that zero values have been clustered together. To exploit this, for each string of repeated zero values, a single value is stored along with count of how many times values is to be repeated.

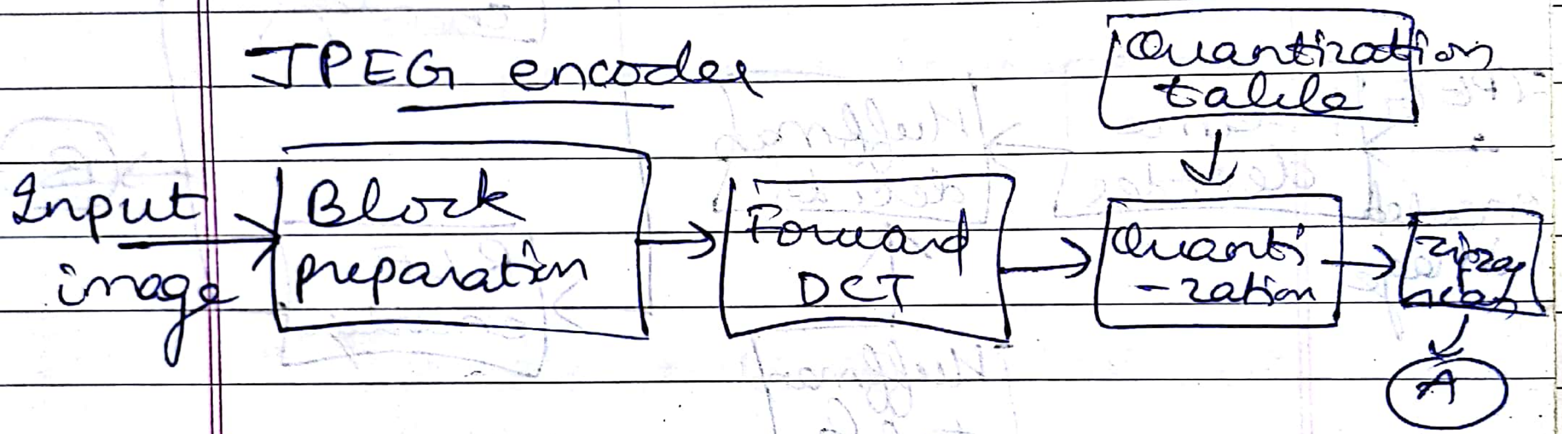
Huffman Encoding

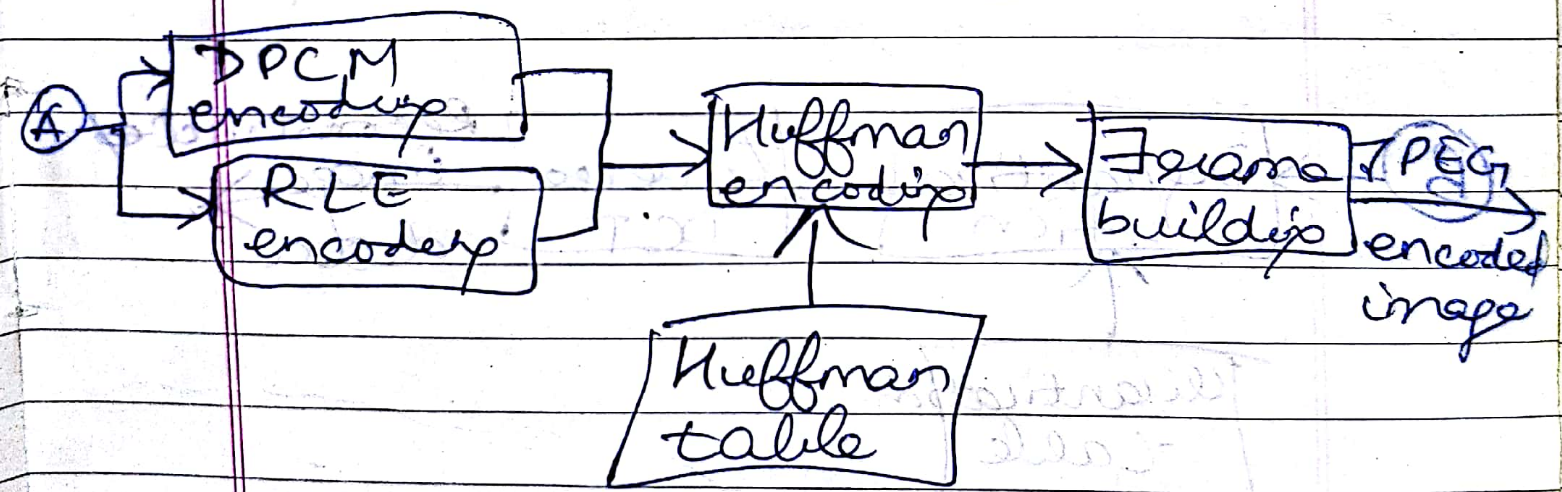
The final step is to apply huffman coding which allocates codes & is applied both to DC and AC coefficients.

Frame building

The frame packing block does the final assembling of data and adds additional error checking codes before sending data to O/p as an encoded data stream

JPEG encoder

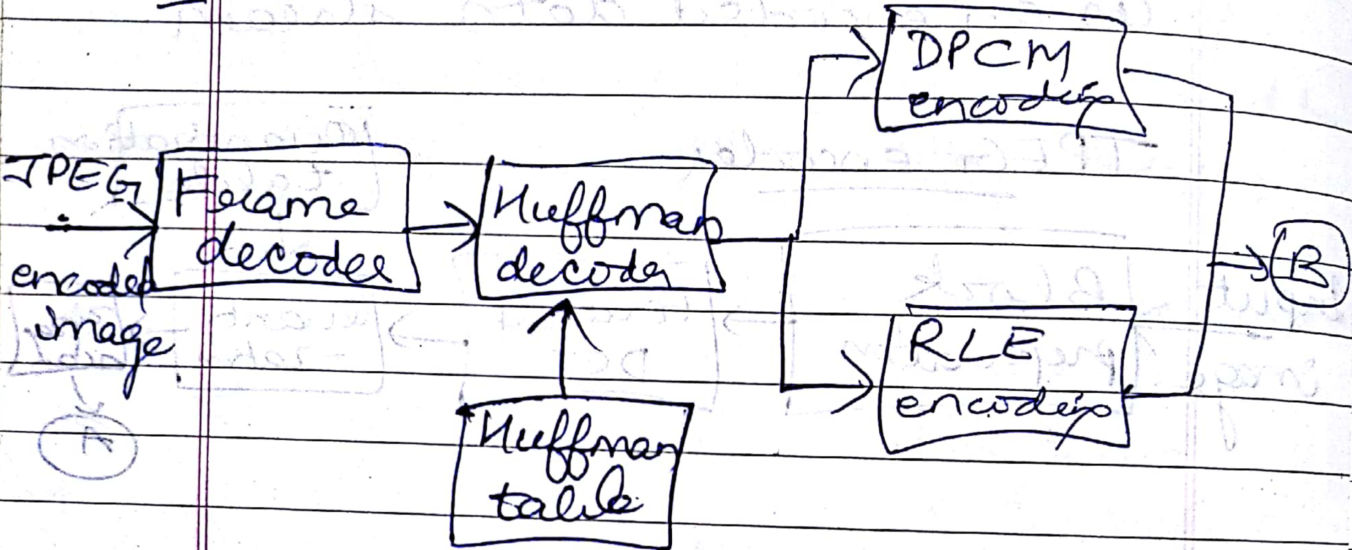


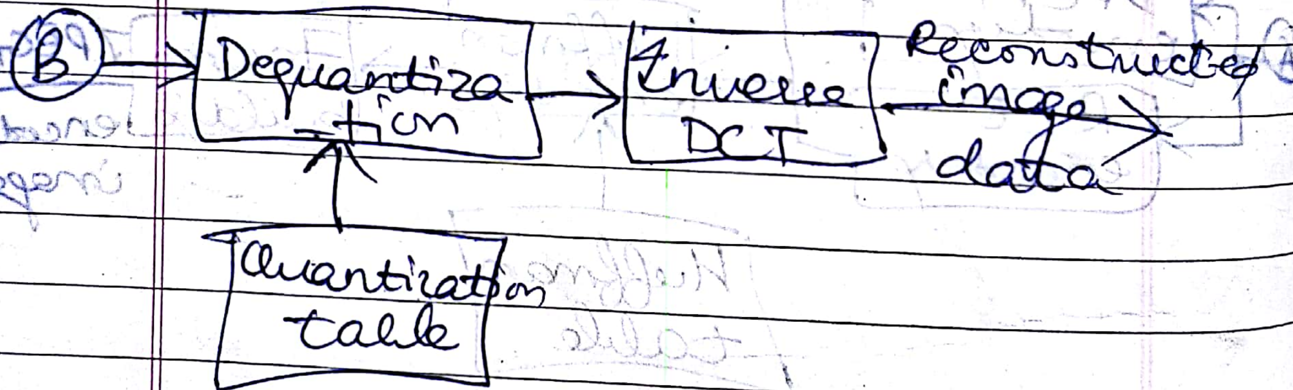


Decoder

— JPEG decoder is made up of no. of stages which are corresponding decoder section of those used in encoder.

—





JPEG decoder

JPEG does not work well with black + white images. To address this limitation, another group called Joint Bi-level Experts Group (JBIG) was formed whose main goal is to define lossless compression algo for these images.

MPEG (Motion Pictures Expert Group)

MPEG is working gp. under ISO/IEC (International Standards Organization/ International Electrotechnical Commission) set up to formulate a set of standards relating to a range of mm applications involving audio + video.

MPEG-1

MPEG dev. the ISO/IEC international standard 11172 for reduced data rate coding of video + audio signals. This standard is called MPEG-1. It is used for Video CD + includes popular MP3 audio format. It has 5 diff. parts:-

Part-1 Systems (ISO/IEC 11172-1: 1993):- It addresses problem of combining 1 or more data streams from video + audio parts of MPEG-1 standard with timing info to form single stream.

Part 2 - Video (ISO/IEC 11172-2: 1993)
It is used for compressing video sequences at bit rates about 1.5 Mbit/s.

Part 3 - Audio (ISO/IEC 11172-3: 1993)
It is used for compressing audio sequences, both mono + stereo.

Part 4 - Compliance Testing
(ISO/IEC 11172-4: 1995) :- It specifies how tests can be designed to verify whether bitstream + decoders meet the requirements as specified in Parts 1, 2 and 3 of MPEG-1 standard.

Part 5 - Software simulation
(ISO/IEC TR 11172-5: 1998)
It is a technically report giving full s/w implementation of 1st 3 parts of standard.

MPEG-2

Video + audio storage + transmission standard for broadcast quality television. Used for digital TV Tx, digital satellite TV services, digital cable TV signals + for DVD applications. It has series of document which are subsets of ISO Recommendation 13818. It has 8 parts.

MPEG-3

Designed for HDTV, but abandoned when it became apparent that MPEG-2 was sufficient for that purpose.

MPEG-4

- The primary uses for MPEG-4 standard are web + CD distribution, conversational (videophone) + broadcast television.
- It absorbs many features of MPEG-1 and MPEG-2.
- It adds new features such as VRML support.

It has 21 parts.

- Most of features included in MPEG-4 are left to individual developers to decide whether to implement them.

MPEG-7

* Also called 'Multimedia Content Description Interface' which provides a set of descriptors (D) ie quantitative measures of audio-visual features + DS (description schemes) (DS) ie structure of descriptors + their relationship. * A language called DDL (Description Definition language) can be used to specify scheme.
* The images, audio, video that have MPEG-7 data associated with them can be indexed + searched for.
* It has 10 parts.

MPEG-21

- Defined as ISO/IEC 21000 -
It defines an open framework for mm delivery + consumption
- It is based on 2 essential concepts :-

classmate

Date

Page

→ 9 parts

(a) definition of fundamental unit of distribution + transn

(4) Concept of users interacting with digital items.

MPEG-1 Audio

It is standard for digital audio.
It has 3 diff. layers pertaining to 3 levels of complexities for algorithms.
Layer 1 is basic mode + layer II + layer III has ↑ing levels of complexities associated with them which in turn produces a corresponding ↑ in level of compression.

Layer I: 192 Kbps/channel compression 4:1, quality same as digital audio cassette.

Layer II: 128 Kbps/channel, compression 6:1 to 8:1, quality same as digital audio broadcasting.

Layer III: 64 Kbps/channel, compression 10:1 to 12:1, quality same as audio CD quality.

Bit rates are for a single audio channel. There are 4 diff. audio channel formats. They are :-

- Monophonic (mono), where only a single channel carries audio signal.
- Dual mono - where same mono signal is carried over 2 channels.
- Stereophonic (stereo) - 2 channels carry different signals
- Joint stereo - stereo signals are combined & carried over a single channel.

For stereo & dual mono format, bitrates are doubled.

MPEG-1 Audio layer I (MP1)

- ~~Lea~~ Least sophisticated method
- Requires high data rate of 192 kbps/channel.
- Compression ratio is about 4:1.
- Quality same as digital audio cassette.

MPEG-1 Audio layer II (MP2)

- MP2 files 1st appeared on Internet & were played back using Xing MPEG player & on UNIX, in a program called MA Play.
- Bitrate is 128 Kbps/channel
- Compression ratio ranges from 6:1 to 8:1.
- Quality is same as that of digital audio broadcasting.

MPEG-1 Audio layer III (MP3)

- ~~It~~ is used to compress files prior to electronic distribution. The MP3 CODEC is lossy in nature.
- The crucial point is which data to keep and which to discard
- To decide MP3 depends on the perceptual characteristics of human acoustic system.
- Specific sounds may not be audible to human ear, it is called masking.
- So, masked sounds are removed & remaining data forms the compressed file.

- MP3 encoder 1st convert input audio ~~to~~ the freq. domain using DFT (Discrete Fourier transform) after which a psycho acoustic block (containing infoⁿ about characteristic of human auditory system) identifies + removes masked audio compnent.

- The remaining data is sent to quantizer + coding block for code word generation.
- The frame packing block adds error checking codes before storing compressed data as an MP3 file.

- Data must be parsed. thus as MP3 decoder for playback dequantizes codewords + converts them back to temporal domain from freq. domain.

- Bit rates of 2A8, 64, 112, 128 192 and 320 Kbps may be allowed.

ID3 Tag → It is format for storing textual metadata for MP3 files such as title, artist, album, track no etc.
The various variants are:-

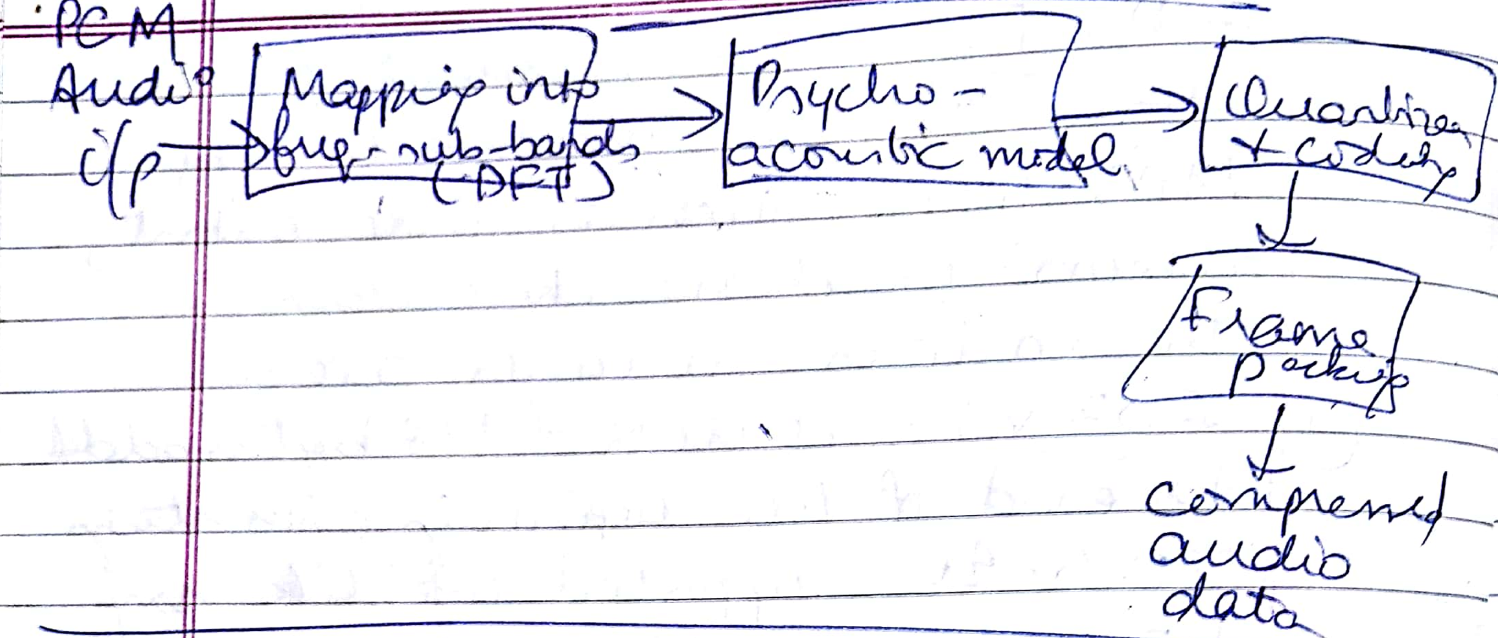
(1) ID3V1: consists of 128 bytes added to end of file beginning with string TAG. It supported info^n like song title, artist, album, year etc.

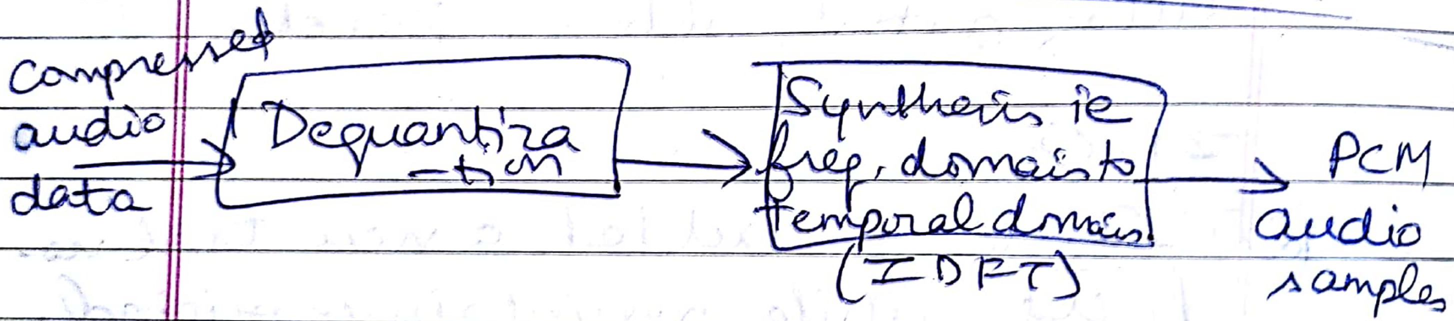
~~ID3~~

(2) ID3V1.1 → added a new track no. field while maintaining original size of 128 bytes.

(3) ID3V2:- expanded ID3 format with add^n of many more fields such as storing of lyrics within file. The tags are included at beginning instead of the end which has an advantage that related info^n can be read without reaching end of file.

MPEG-1 audio encoder





MPEG-1 audio decoder