Chapter - 1

04.02.17

Data representation :

→ Text → Unformatted
→ Text → formatted

→ Audio → Signal

→ Image → Pixel

→ Vedio → Sequence of images

When multiple data are required in one device it's called multimedia.

classification of medium :

→ Perception → Presentation → Transmission

→ Representation → Storage → Information exchang

⇒ See, hear → Perception

⇒ Input → ~~Camera~~ Camera, keyboard, Microphone

Output → Printer, Speaker, touch screen

Presentation ←

⇒ Hard drive, Memory card → Storage 2

⇒ Text → ASCII, EBSIDIC
Audio → MP3, MP4
Image → JPG, PNG
Vedio → MPEG
Animation → GIF
→ Representation

⇒ Wire, Wireless cable → Transmission

⇒ Email → Information exchange.

Criteria of media:

→ Space:

→ Values:

→ Dimensions: Time depended → Audio, text
Time independed → Vedeo, image.

25.02.19

Properties of multimedia system:

→ Qualitative rather than Quantative
→ Combination of media.
→ Level of independence.
→ Computer supported integration.
→ Communication system.

Data stream classification:

⇒ Transmission mode

→ Synchronous · [fixed time] → ethernet
→ Asynchronous · [changeable] → Email transfer
→ Isochronous · → Vedio transfere.

⇒ Time interval of consecutive packets

- Strongly periodic → TV transmission
- Weakly periodic → Email / SMS transfer
- Aperiodic → Live streaming.

⇒ Variation of packet amount

- Strongly regular.
- Weakly regular.
- irregular data stream.

⇒ Contigious packets

- Continuous.
- Discrete.

[Advantage & Disadvantage]

Chapter-2

03.03.19

Sound / Audio

Basic terms of sound:

→ Period.
→ Frequency
→ Amplitude.

Computer representation of sound:

→ Sampling
  → Sampling rate
    → Lossy.
    → Lossless.
  → Nyquist sampling theorem
→ Quantization
→ Sound hardware

Music:

⇒ Music Instrument Digital interface: [characteristics]
→ doesn't describe, Just command.
→ not a language
→ maintain data communication protocol.

→ digital interface, electrical connectors. 6

→ generates and transmits message among instruments.

→ relates computer and music.

Component:

⇒ MIDI Hardware
- → MIDI Ports
- → MIDI cable

MIDI Ports

→ five pin DIN connector

→ Three pin usable

→ Both ends are same

- MIDI In
- MIDI Out
- MIDI Thru

MIDI Cable

→ Serial transfer

→ Slow in bit transmission 31,250 bit/s

→ toal slow in sample transmission.

⇒ Data Format

→ Encodes travelling sound signal into a format contain
- basic frequency level.
- Sound volume.

→ creates small size file.
→ easy to modify and manipulate.
→ Wide choice of musical instruments.
→ Use digitaly sampled sound only.

11.03.19

MIDI Standered:

Properties:
→ Specifies 16 channel
→ A MIDI device mapped to a channel.
→ identifies 128 instruments with unique numbers.
→ Main property is to allow each synthesizer to recive max notes per channel.
→ 3-6 notes per channel.

Reception Modes:
→ Omni On / Poly.

→ ~~Omni~~ Omni On / mono

→ Omni off / Poly

→ Omni off / mono

⇒ Omni on / off

⇒ Poly / mono.

MIDI Devices:

→ Sound generator

→ Micro processor.

→ keyboard.

→ Control panel.

→ Auxiliary control panel.

→ Memory

→ Augmented Devices.

MIDI Message Format:

MIDI Message
- channel [Specific instrument]
  - Voice
  - Mode
- System [to all devices]
  - Real time
  - Common
  - exclusive

MIDI Software:

→ Music recording & performance application.
→ Musical notation and printing application.
→ Synthesizer patch editor and librarian.
→ Music education application.

18.03.19

Speech:

→ perceived/understood.
→ produced by human/machine.
→ Periodic behaviour during certain time interval.

→ Spectrum audio signal shows characteristic maxima - formats.

10

Speech processing system:

→ Speech generation
→ recognition
→ Von Kempelen's speaking machine (1791)
→ VODER (1939)
↓
[Voice Operation Demonstrator]

Speech

→ real time
→ understandable
→ natural

basic units/features

→ Fundamental frequency
→ Phone
→ Allophone
→ Morph
→ Sound — Voiced Sound
— Unvoiced sound

Reproduced speech output problems: 11

⇒ Time dependent sound concatenation:

→ Phone sound concatenation → Word crumb.

→ co-articulation

→ Prosody.

→ Syllable sound concatenation

- Constitutes diaphone

→ Word - sound concatenation

- co-articulation

⇒ Frequency dependent sound concatenation:

→ Formant synthesis → speech synthesis system

→ Prosody.

→ Co-articulation.

* Speech synthesis system:

→ Transcription

→ Synthesis

library
[rules to transfer word to phone and exceptions

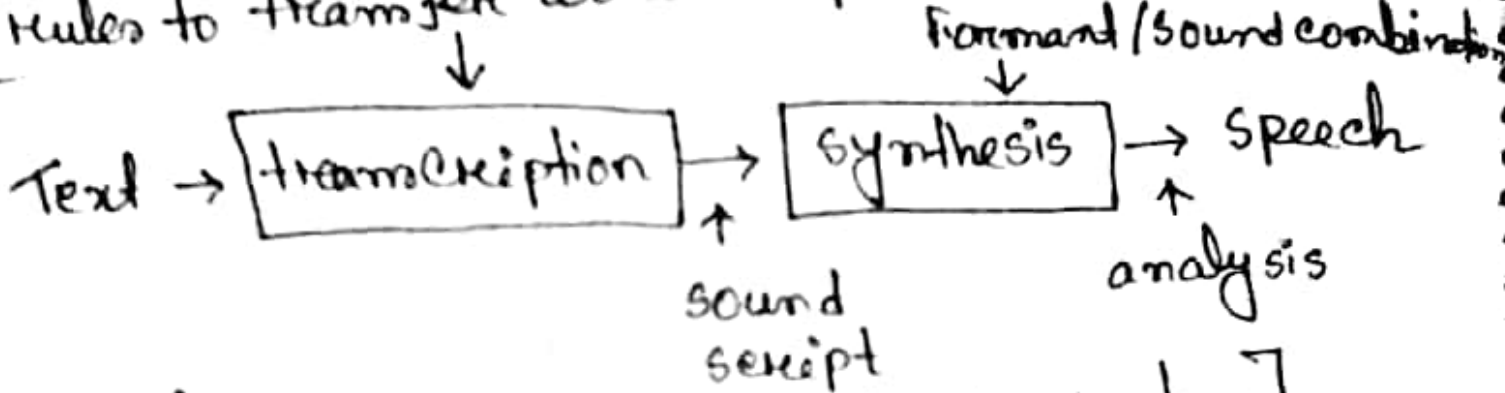
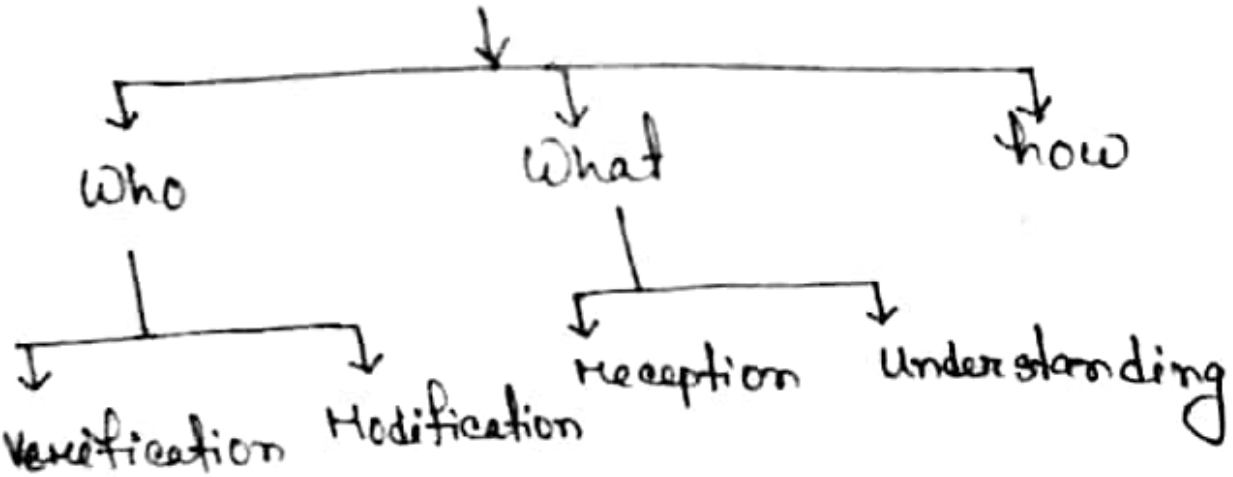


fig: Format [Speech Synthesis system]

⇒ Speech analysis system



24.03.17

13

Speech Recognition system:

→ Speech Analysis

- Data extracted from speech element.

→ Speech recognition

- Comparison of speech element with reference library

⇒ two types of speech recognition are:

→ Speaker dependent recognition [25,000 words] → Advanced training

→ Speaker independent recognition [500 words]

Problems in speech recognition:

→ environmental noise.

→ Pre determination of word boundary.

→ time normalization.

Chapter 4

→ Image:

→ Digital image: [Processing] self

→ Pixel:

→ Image format:
- → 1 color [gray scale value]
- → 3 color [RGB triad value]
- → 4 color [RGB triad value + gray scale value/ opacity]

→ Bitmap:

→ Un compressed

→ 3 dimensional array
- height x width x 24 (8 for each color)

→ Simple

→ Wide acceptance in windows application

→ .bmp, .tif / .tiff

⇒ JPG / JPEG (Joint Photographic experts group):

→ For continuous tone image.
→ 8 bits per color.
→ 16 millions of color.
→ Uses lossy compression.
→ Generational degration after ~~reap~~ repeated edition.

⇒ GIF (Graphical interchange format):

→ For large areas of same color.
→ moderate level of details.
→ Supports upto 256 colors.
→ Supports animation and image animation effects.
→ Uses loseless compression.
→ in effective for detailed image or dithered image.

⇒ PNG (Portable network Graphics):

→ free, open source successor of GIF.
→ loseless, portable and well compressed.

25.03.19

Graphics:

→ Pictorial synthesis of a real or imaginary object using computer aided models.
→ graphical primitives [Point, vertex, line/edge, region]
— attributes

→ no pixel represented.
→ direct manipulation is possible.

Dynamics:

→ Object moves in respect to other object.

→ Object and camera both moving.

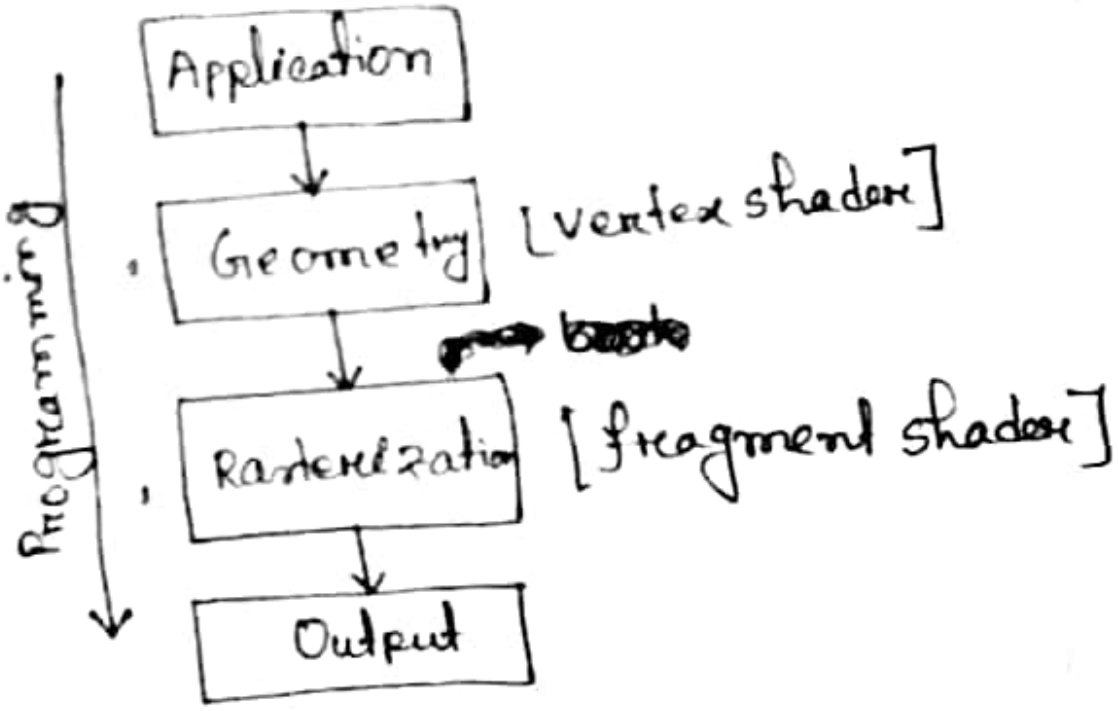
Update dynamics:

→ Colon change, background navigation etc.

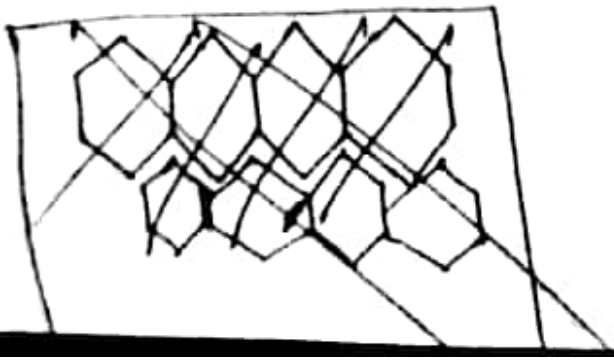
→ represent data in simple, and intelligent way.

→ Can represent multidimensional data.

Graphical pipeline:



⇒ Screen mosic tiled arrangement:



⇒ Progressive scanning

⇒ Interlace scanning

Dithering:

Raster display: [Book]

Image recognition:

⇒ Image analysis

⇒ Image recognition

Source image → formatting → Conditioning → Labeling → Grouping → Extracting → Matching ← template

target image

01.04.19

JPEG Compression:

→ Joint photographic Experts Group.

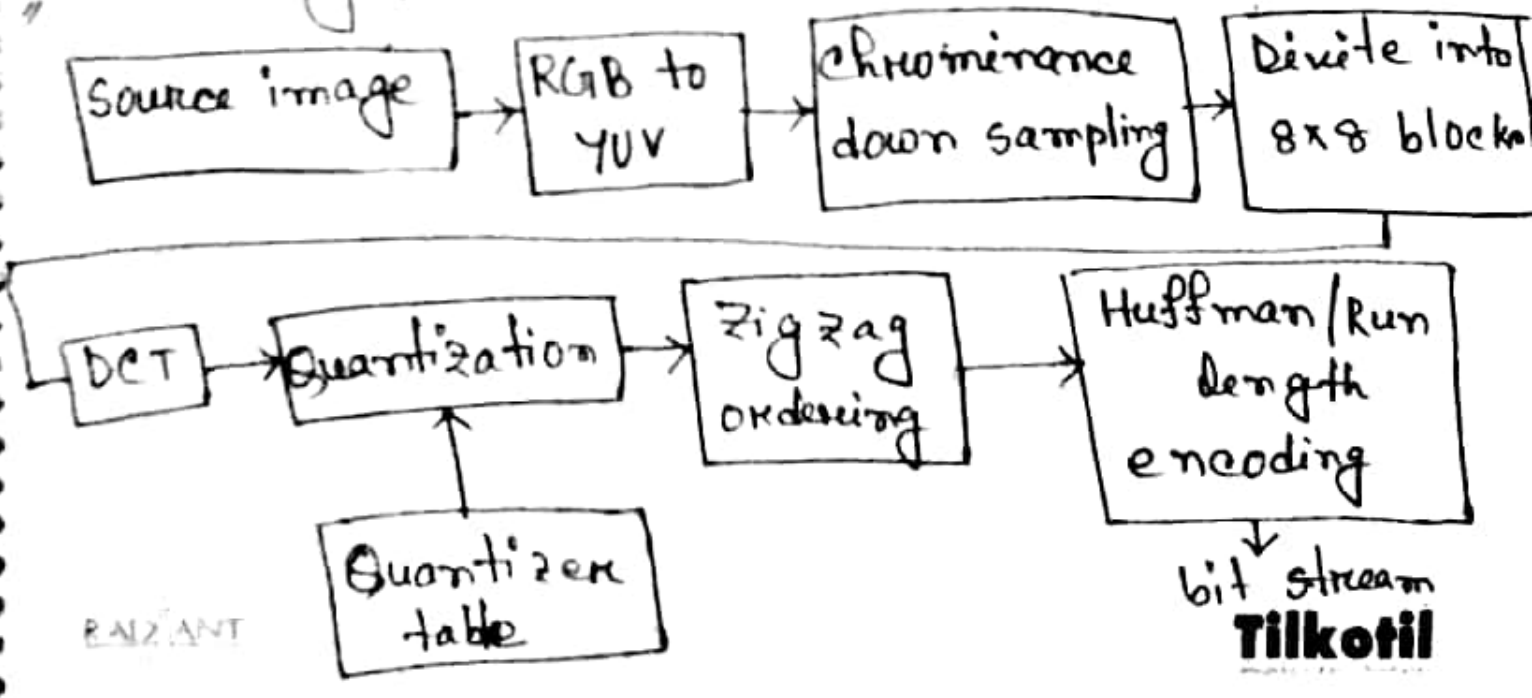
→ Lossy compression.

→ 24 bit color image.

⇒ Encoding: (step)

1. RGB to YUV model.
2. Divide into 8x8 blocks.
3. DCT.
4. Quantization.
5. Encoding.

⇒ Encoding process:



→ Luminance → Light shed

Chrominance → color

YUV model
- → chrominance → hue (color variable)
- → Luminance → saturation

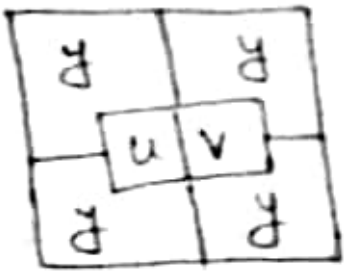
Step: 1 →

1(a): $Y = 0.299 * R + 0.587 * G + 0.114 * B$

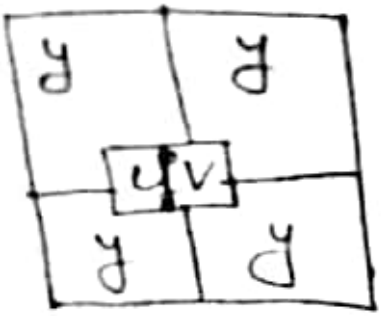
$U = 0.1687 * R - 0.3313 * G + 0.5 * B + 128$

$V = 0.5 * R - 0.4187 * G - 0.813 * B + 128$

1(b):



4:1:1



4:2:0

chrominance
down sampling

→ Color variation কাহাকাছি হাকলে down sampling করা.

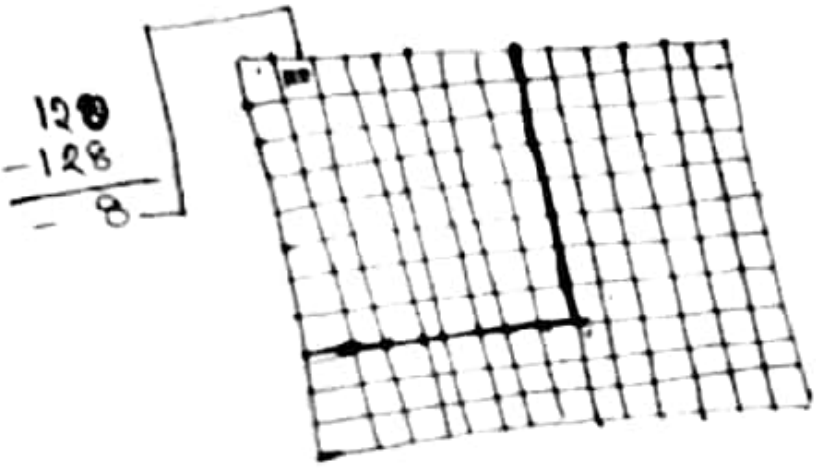
Step:2 → Devide into 8x8 blocks [8x8 = 64bit]

240 x 320

= 76800 / 64

= 1200 blocks

→ DCT করার আগে shift left করতে হবে.



Step:3 → DCT → (Source image)

$$T(u,v) = \sum_{x=0}^{N} \sum_{y=0}^{N} f(x,y)\, g(x,y,u,v)$$

Inverse DCT → (Destination image)

$$f(x,y) = \sum_{u=0}^{N} \sum_{v=0}^{N} T(u,v)\, g(x,y,u,v)$$

$f(x,y)$, $T(u,v)$

NXN NXN

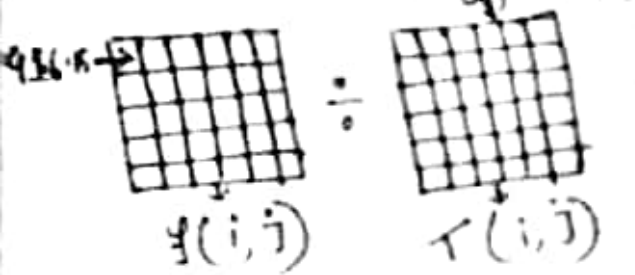
$$g(x,y,u,v) = \alpha(u)\alpha(v) \cos\left[\frac{(2x+1)u\pi}{2N}\right] \cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

$$\alpha(x) = \begin{cases} \frac{1}{\sqrt{N}} & x = 0 \\ \sqrt{\frac{2}{N}} & x = 1 \dots N \end{cases}$$

→ Sampling frequency generate করার জন্য DCT প্রয়োজন ..... , ...

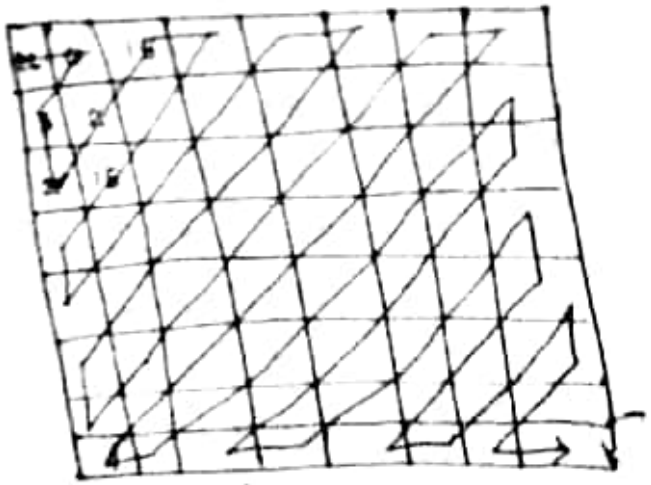
Step: (1→

$$B(i,j) = \frac{f(i,j)}{T(i,j)}$$ → Luminance table / Chrominance table



16 | 416.15 | 26.0597
32
96
96
150

Step: 5(a) →



Video:

→ Visual representation.
→ Transmission.
→ Digitalization.
→ Video formats.

Characteristics:

1. Visual representation:
   → Vertical detail [Viewing distance]
   → Horizontal detail [Picture width]
   → Total detail content
   → Perception of depth
   → Luminance & chrominance.

⇒ Temporal aspects:
   → Motion resolution.
   → Continuity of motion [15 frame in 1S]
   → Flicker
   → Video bandwidth selection

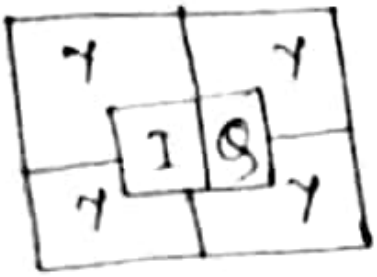
2. Transmission :

RGB

YUV

YIQ → TV transmission

- Chrominance
- Luminance



$Y = 0.30R + 0.59G + 0.11B$

$I = 0.60R - 0.28G - 0.32B$

$Q = 0.21R - 0.52G + 0.31B$

3. Digitalization :

Two steps are : 1. Sampling 2. Quantization.

4. Video formats :

CGA	EGA	VGA	8514/A display adapter	XGA
1. Color graphic adapter	Enhanced graphics adapter	Video graphics adapter		Extended graphics adapter
2. 320x200 Pixel	640x350 pixel	640x480 pixels	1024x768 pixels	640x480 pixel
3. 4 color	16 color	256 color	256 color	65000 colors
4. Storage capacity	Storage capacity	Storage capacity	Storage capacity	Storage capacity
320x200x 2byte/Pixel x 8	640x350x4x8	640x480x8byte x8bit		

SVGA
1. Super VGA
2. 1024x768 pixels
3. 24bit color
4. Storage capacity

06.06.13

Television and Animation

⇒ Television Standards:

⇒ Analog Tv: → NTSC
→ SECAM
→ PAL

⇒ NTSC:

→ National Television Standard committee

→ Widely used in Japan and mostly other asian countries, south and north america

→ uses QAM

→ 4.429/3.57 MHz (BW), for motion frequency 30Hz

→ Picture consists 525 lines (486 is visible)

→ 4.2MHz for luminance and 1.5 MHz chrominance

→ Manual color correction.

→ resolution 720×480

→ refresh rate 60Hz

→ 30 frame/s

⇒ SECAM:

→ Sequential couleur Avec Memoire
(sequential color with memory)

→ Uses in France, ~~eate~~ eastern europe.

→ Uses FM.

→ Motion frequency 25 Hz

→ 625 lines, 576 visible

→ auto color correction

→ refresh rate 50Hz

→ $\frac{1}{2}$ the resolution after transmission.

→ 25 frame/s.

⇒ PAL:

→ Phase Alternating line

→ Uses in Western Europe (UK, sweden)
.... Australia...

→ Uses QAM

→ automatic color correction

→ 625 lines, 576 is visible

→ refresh rate 50Hz

→25 frames/S

⇒ Digital TV:

→ IDTV

→ D_2MAC

⇒ IDTV:

→ Improved definition television

→ Improved version of NTSC

→ Doubles line from 520 to 1050.

→ Doubles bandwidth to transmit data.

⇒ D_2MAC: (improved version of PAL)

→ Duo Binary Multiplexed Analogue components.

→ data rate is 10.125 Mbit/S

→ reduced vision bandwidth using DMAC algorithm.

→ QAM + FM.

⇒ HDTV:

→ High definition Television system.

→ Resolution - Twice of conventional system.
- In creament - Vertical definition - added new 1000 lines.
- Luminance details by employing 5th times of BW of conventional system
- Additional BW applies to maintain color quality.

→ Aspect ratio : 16/9 = 1.777 ...

→ Viewing distance is closer than conventional system.

⇒ Digital Coding:

→ Composite coding:

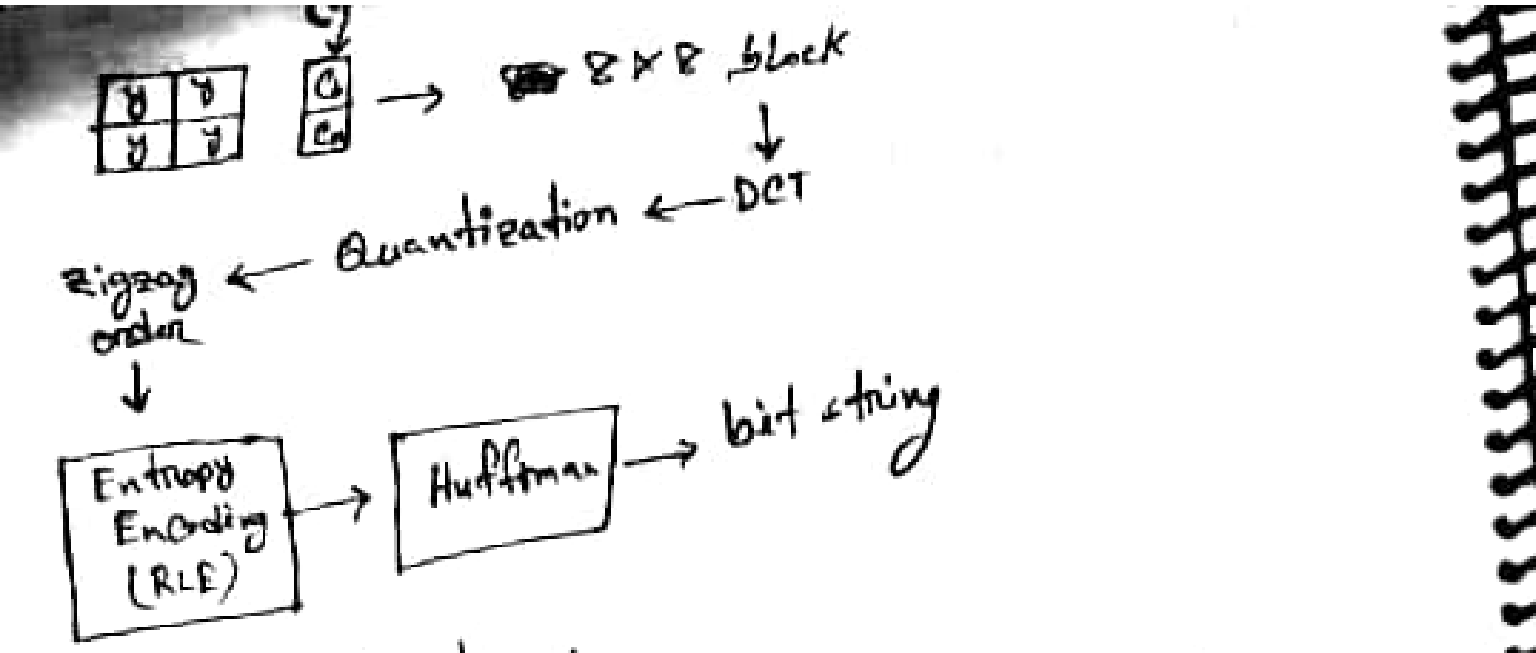
→ Simples in digitizing video signal.

→ combines all separate components.
- One luminance
- two chrominance

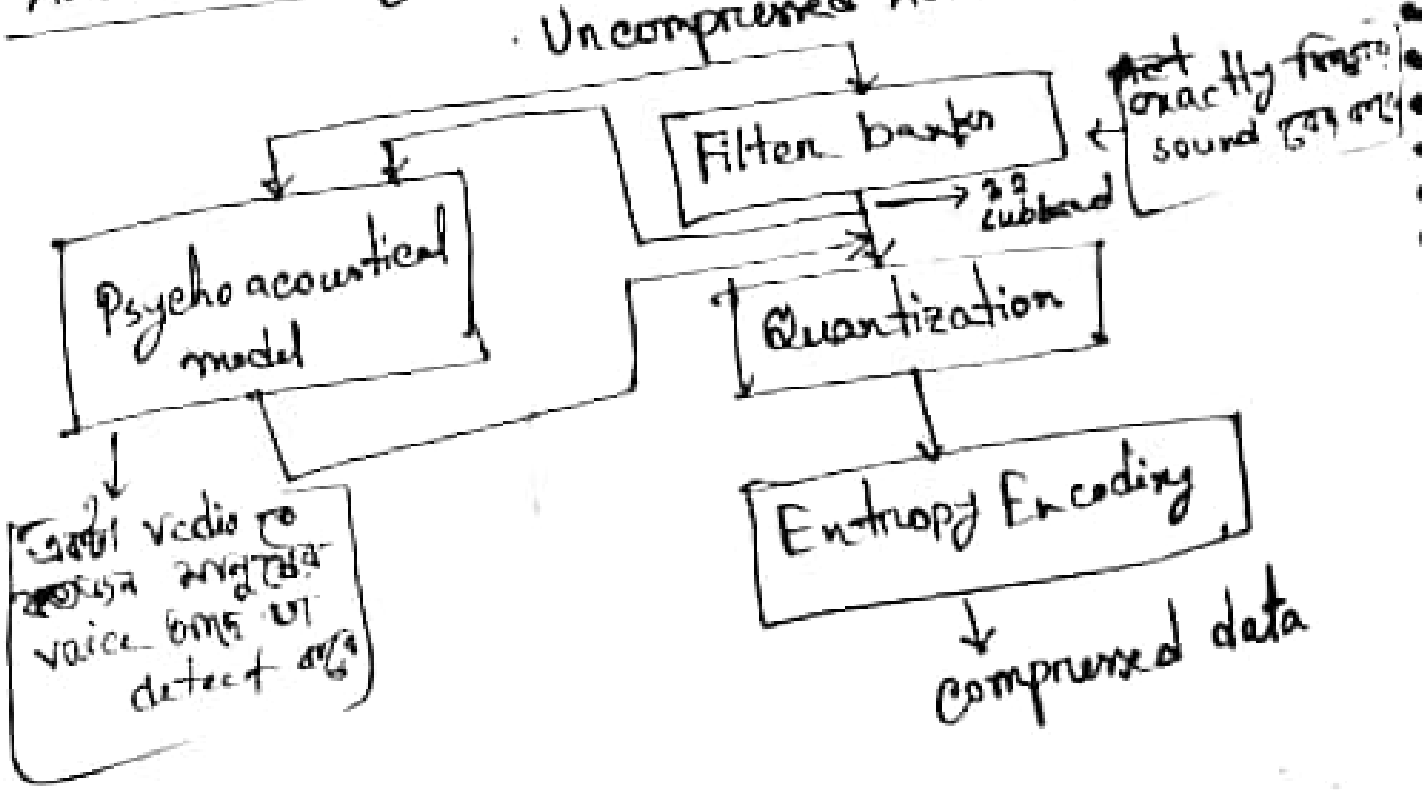
→ Crosstalk between elements.
→ depends on television standards.
→ Data reduction can be allocated.
→ BW allocation is not sufficient.

⇒ Component coding:
→ Separate digitization of each component.
→ transmit together through multiplexing.
→ Luminance sampled with 13.5 MHz.
→ Chrominance - 6.75 Hz

⇒ Computer based animation:
→ Full explicit control.
→ Procedural control.
→ Comtraint based control.
→ Tracking live Action / Rotoscoping.
→ Kinetics & Dynamics.



Audio Encoding:



06.05.19

Data Compression:

⇒ Compression techniques:

→ JPEG - image
→ H.261 - Vedio
→ HPEG - Video + audio
→ DVI - image + audio + Video

⇒ Coding requirements:

→ Text:
→ 2bytes
→ 8x8 pixels.

→ Image: (Vector Array)
→ 500 lines - horizontal details
- Vertical details
- 8 bit attribute field
→ Horizontal axis (10 bits)

Vertical axis (9bits)

→ 1 bytes/ pixel - 256 colors

→ Audio:

→ Sample rate 44.1 KHz

→ Quantization rate 16 bit / sample

→ Storage $(44.1 \times 16) \times 10^3 \times$ bit

→ BW $(44.1 \times 16) \times 10^3$ Hz

→ Video:

→ Image resolution 640 * 480

→ 3 bytes per pixel (luminance 1 Chrominance 2)

→ Storage : 921,000 bytes

→ Refresh rate 25 frame/s

⇒ Dialogue Mode:

→ Human to human conversation.

→ Retrieval Mode:

→ Human to device / cloud communication

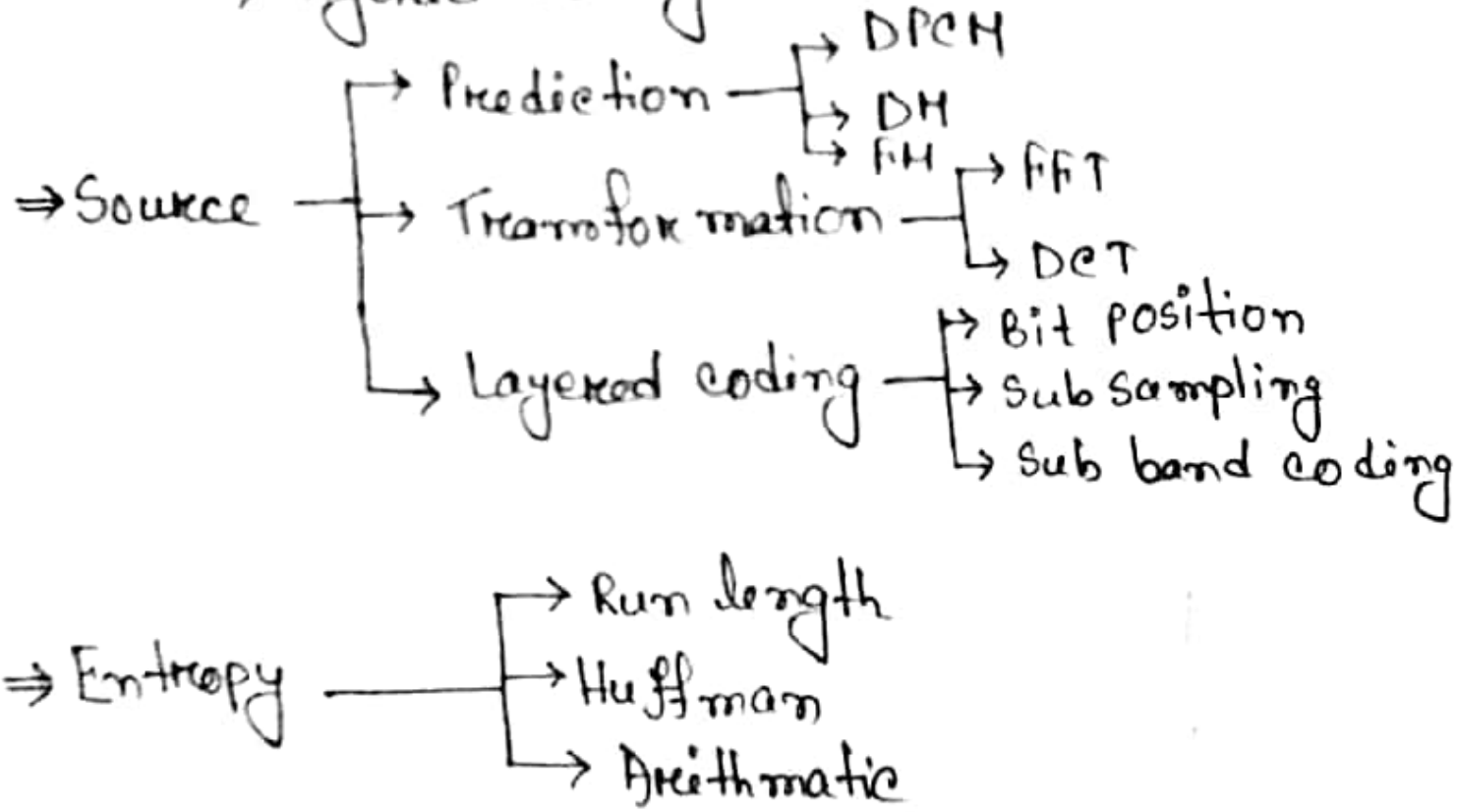
⇒ Both dialogue and retrieval mode:

Coding Categories:

→ Source coding

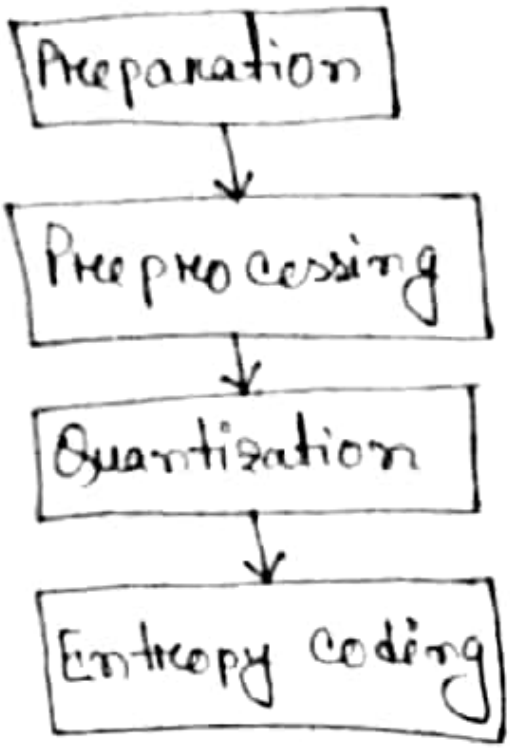
→ Entropy Coding

→ Hybrid coding



⇒ Hybrid
- → JPEG
- → HPEG
- → H·261
- → DVI

Compression steps: [~~Encoding~~]



Decompression:

⇒ Symmetrical Decompression:

→ Same cost

→ lossy data

Asymmetrical decompression:

→ lossless data

→ Cost higher

09.05.19

⇒ H.261 Video compression standard:

→ Published by ITU-T in 1990.

→ internation video compression standard.

→ Used for conjuction with other control and framing standards.

→ describes only video compression in a audio visual service.

→ Targeted for circuit switched network.

→ two imge formal

- CIF (352X288)
- QCIF (176X144)

→ Combination of

- inter picture prediction [temporal redundancy]
- transform coding [spatial redundancy]
- motion vectors [motion compensation]

→ based on JPEG.

⇒ Design details:

→ YCbCr format.

→ must be able to compress QCIF format, CIF is optional.

→ Source image contains 16×16 macroblocks.

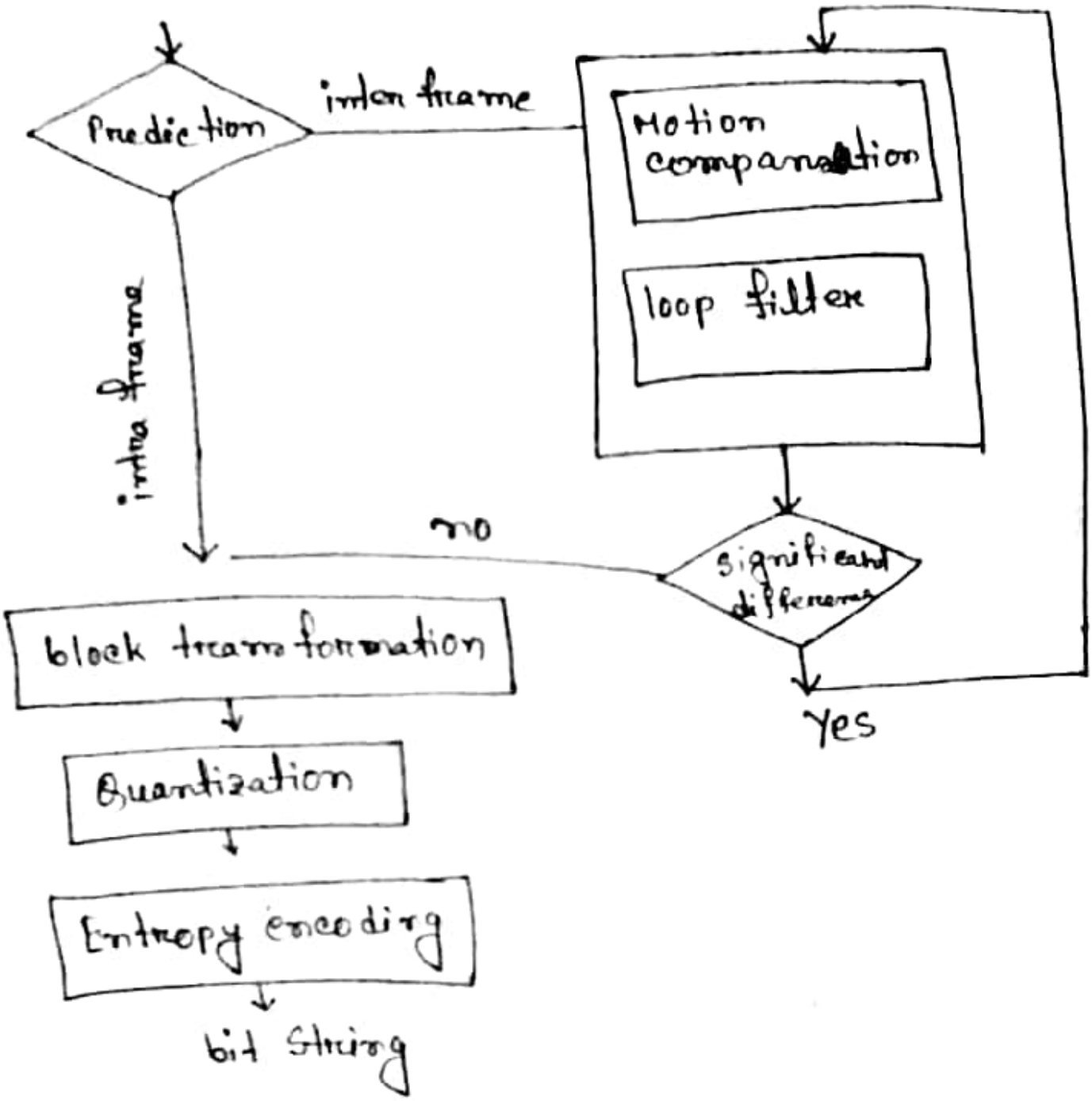
→ each macroblocks contains 8×8 block.

→ each block contains 2×2 pixel for YCbCr value.

→ main elements

- Predictor [inter frame / intra frame]
- Block transformation [DPCM / DCT]
- Quantization and Entropy encoding.

+ ITU-T, ISDN, CIF, QCIF, DPCM, DCT...



⇒ Difference between interframe and intraframe?

→ I-Frame
- intermediate frame (background)

→ P-Frame / Delta frame. (difference create frame)

⇒ Entropy Encoding:

→ Run length Encoding.

→ Huffman Encoding.

→ Arithmetic Encoding.

→ Run length Encoding: (4-259) → 3 bits

aa ff bbb ccccccc dd a ff [fixed length storag]

↓

aa ff bbb c!6 dd a ff

→ Huffman Encoding: [* Variable length storag]

aa ff bbb ccaccc dd a ff

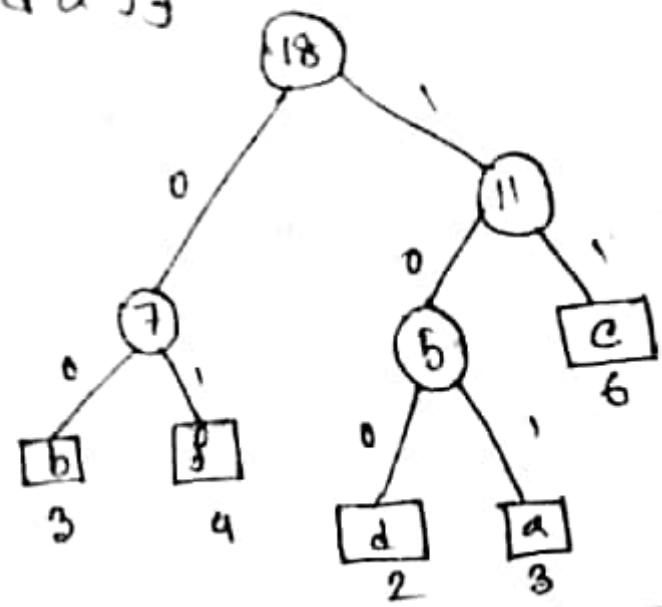
a	b	c	d	f
3	3	6	2	4

a = 101 f = 001

b = 00

c = 11

d = 100



→ Arithmatic Encoding: 10

aa ff bbb ccccccc dd a ff

↓ ↓

K

K = 5

Assignment: [next sunday]

1. Digital image processing [9 step detail]

2. OSI 7 layer effects on BHS.

CT-2: Graphics + Vedio formats পড়বে.

25.03.19 - 08.04.19

MPEG

06.05.13
(Sunday)

Audio + Video Encoding:

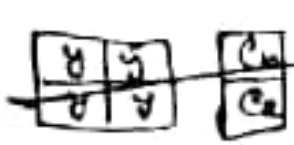
Vedio Encoding:

- → Sequence layer
- → GOP layer
- → Picture layer
- → Slice layer
- → Macro block layer
- → block layer

* Difference between JPEG & MPEG

→ $Y C_b C_r$ model

→ Three frames
  - I frame (Intra frame)
  - P frame (Predictive frame)
  - B frames (Bilateral frame)



~~→ 8×8 block → DCT → Quantization →~~

(P.T.O)