INT307 Multimedia Security System

Image & Video Compression

Shengchen Li

Xi'an Jiaotong-Liverpool University

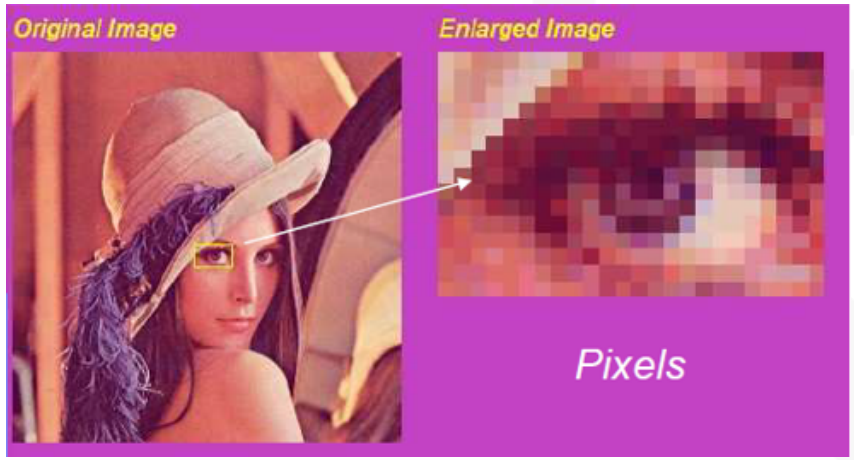
24th Aug 2022

Aims

- Master how images are presented in computer systems
- Understand how people perceive image and audio
- Master how images are compressed in computer systems
- Understand how video is compressed in computer systems
- Understand industrial compression standards

Image / Video Representation

- Images are composited by pixels
  - Each pixel has a colour
  - The density of pixel is known as ppi
  - Human visual system has a resolution of approximate 300 ppi
- Video is composited by frames
  - Each frame presents the image at a moment
  - The rate of frame known as fps
  - Human visual system approximates 60 fps



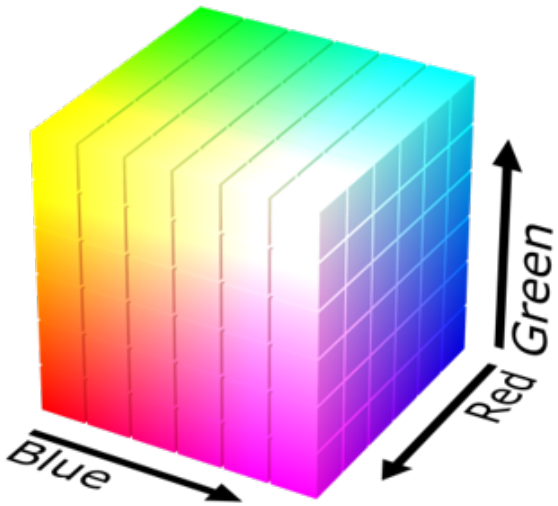
Colour Space

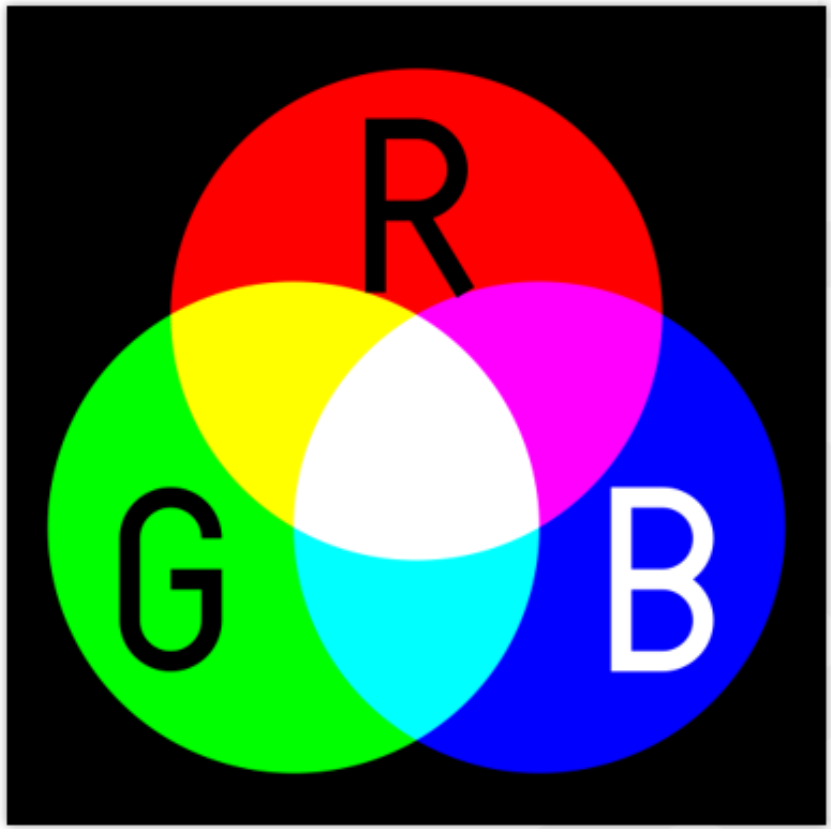
Usually a vector is used to represent the colour of a pixel, where each element in the vector represents a component

- RGB $\rightarrow$ Screen display (Red, Green and Blue)
- YUV $\rightarrow$ MPEG / JPEG system: separate light and colour
- CMY(K) $\rightarrow$ Printer: Cyan, Magenta and Yellow (with Black)
- HSL $\rightarrow$ Hue, Saturation, Lightness (also known as value, i.e. HSV system)

RGB System

Additive Colour System

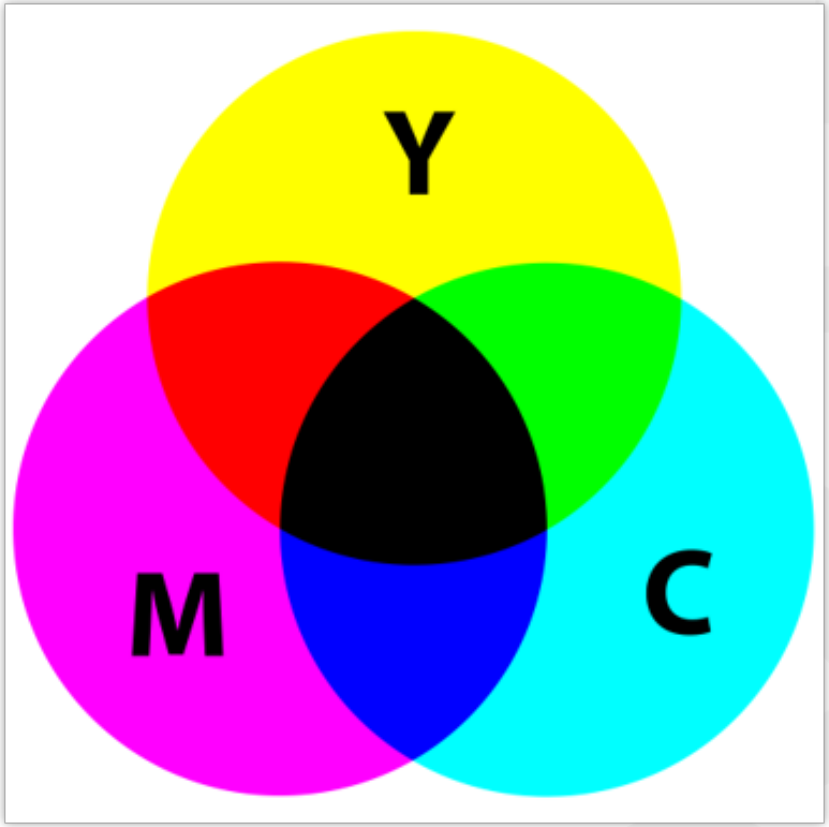




CMY(K) System

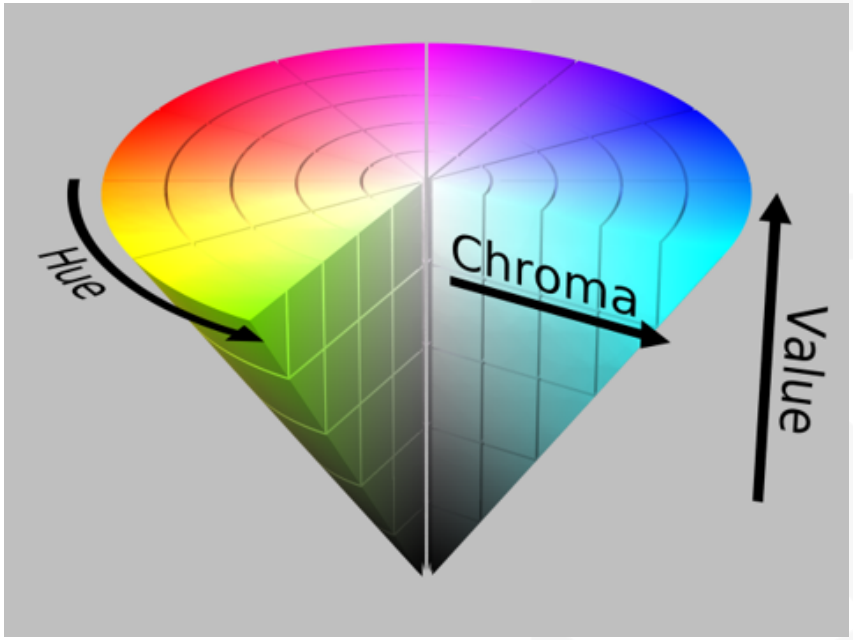
Absorptive colour system

- No perfect Black $\rightarrow$ K component for black
- Printer system



HSL / HSV System

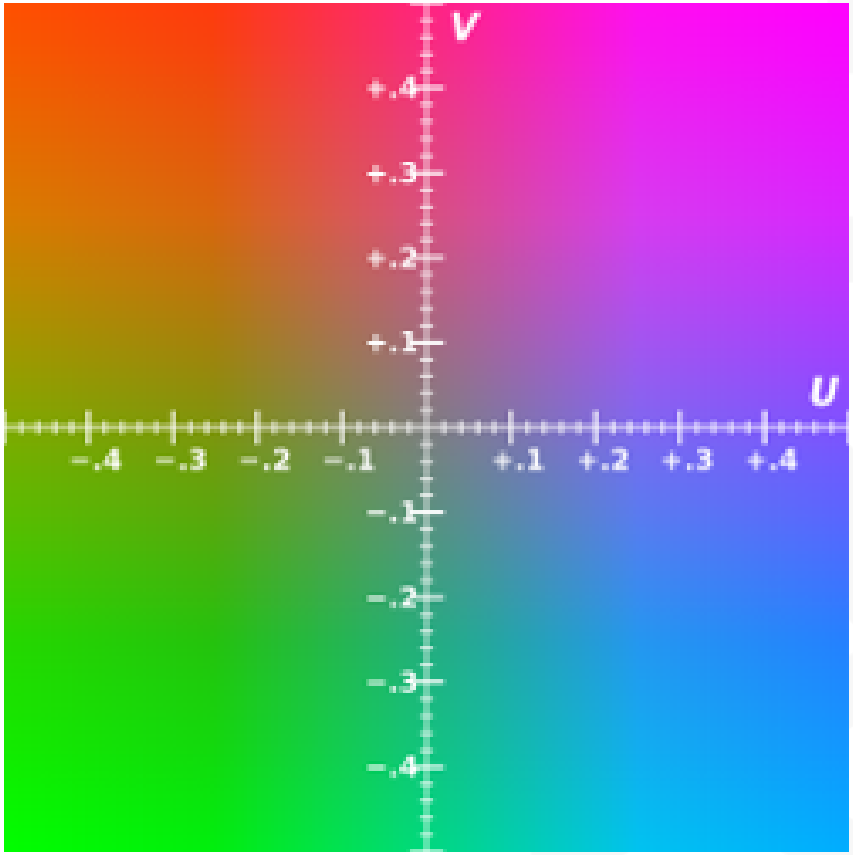
- HSL stands for Hue, Saturation and Light
- HSV stands for Hue, Saturation and Value



YUV System

- Separate Lightness and Colour
- Easier compression as human visual system is more sensitive for light

When $Y = 0.5$, the YUV system seems like



Calculation Question

- For an image RGB system is used with a depth of 8 bits for each component, how many bits needed to represent a picture whose resolution is 1920×1080?
- With the same configuration of the previous question as a frame of a piece of video, how many bits needed to record a one-hour video whose resolution is 1920×1080 with a frame rate of 60 fps?

Compression

As a result, we need to compress the media

- Image Compression
- Video Compression
  - Intra-frame compression
  - Inter-frame compression

Perception of Image and Video

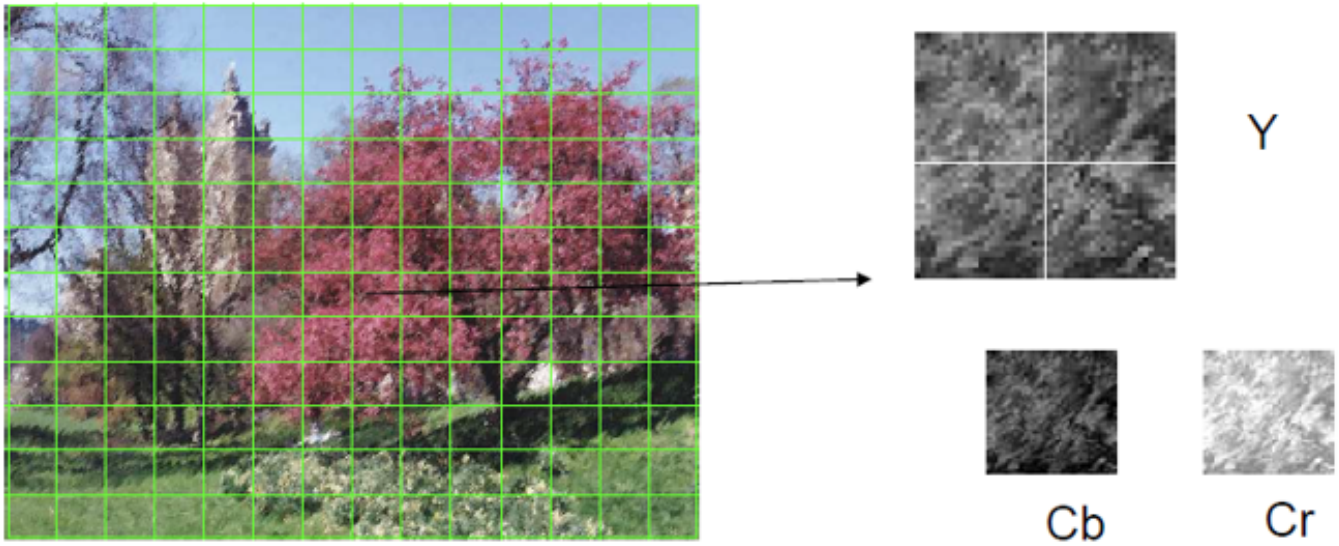
- Frequency of image
  - Rapid changes $\rightarrow$ High frequency
  - Minor changes $\rightarrow$ Low frequency
  - Usually the image is low frequency (redundancy in spatial)
- Redundancy in time domain
- Neighbour frames are likely to contain similar content

JPEG

- Divide Images into 8×8 (luminance) or 16×16 (chrominance) blocks
- DCT
- Quantisation
- Run-level Grouping
- VLC Encoding

Macroblocks

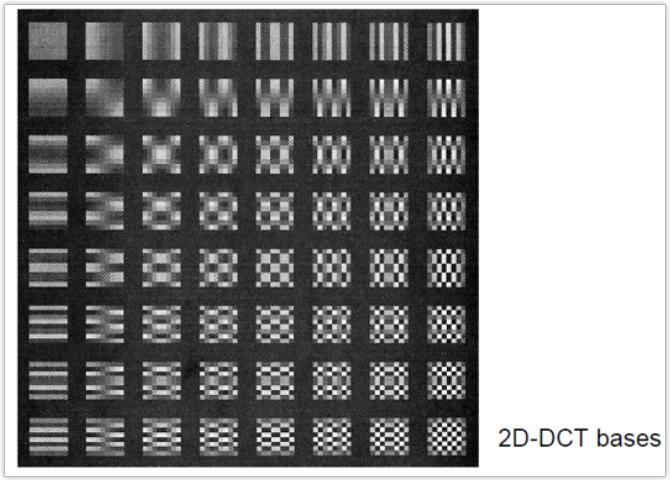
- An image is divided into 8×8 blocks for the luminance components
- An image is divided into 16×16 blocks for the chrominance components
- Chrominance blocks are down-sampled to 8×8 blocks (called 4:2:2 format)
- Zero-Padding on boundary blocks

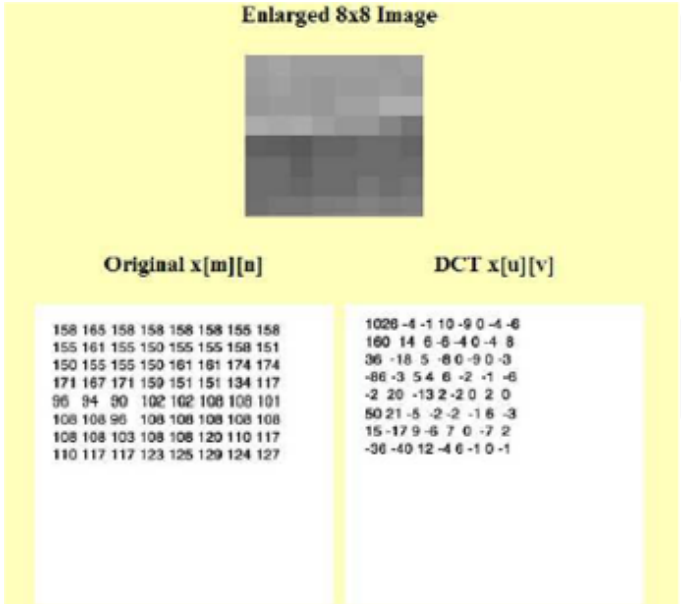


Discrete Frequency-Domain Analysis

$$F(\mu,\nu) = \frac{C(\mu)}{2}\frac{C(\nu)}{2}\sum_{y=0}^{7}\sum_{x=0}^{7} f(x,y)\cos[(2x+1)\mu\pi/16]\cos[(2x+1)\nu\pi/16] \tag{1}$$

$$C(\mu) = \begin{cases} \frac{1}{\sqrt{2}}, if\,\mu = 0 \\ 1, if\,\mu > 0 \end{cases} \tag{2}$$





Quantisation

Different Quality factor decides different quantisation table

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

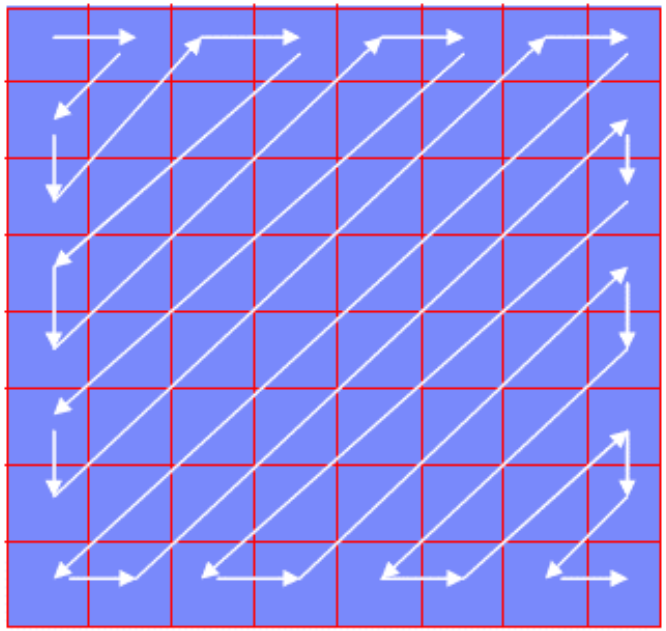
(a)

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

(b)

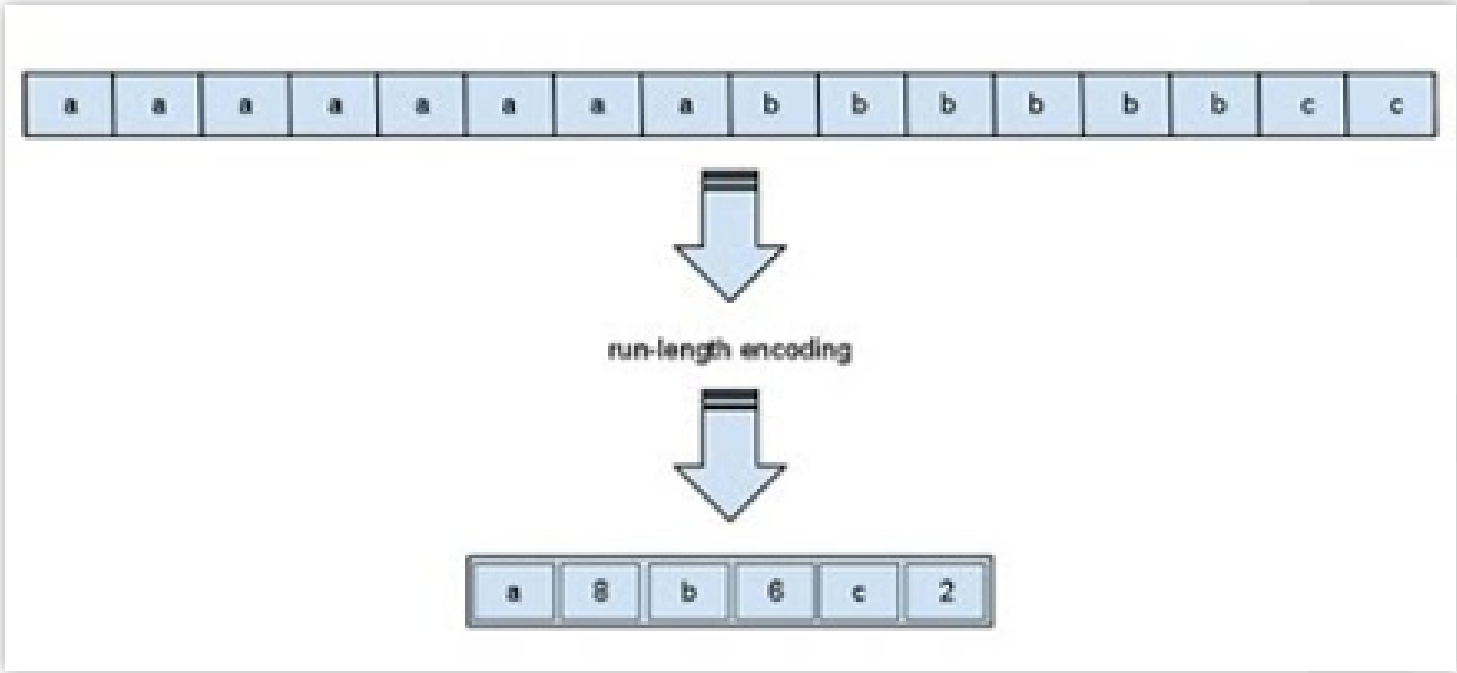
Zig-Zag Order

An order to sort the 2D DCT coefficients to 1D signals



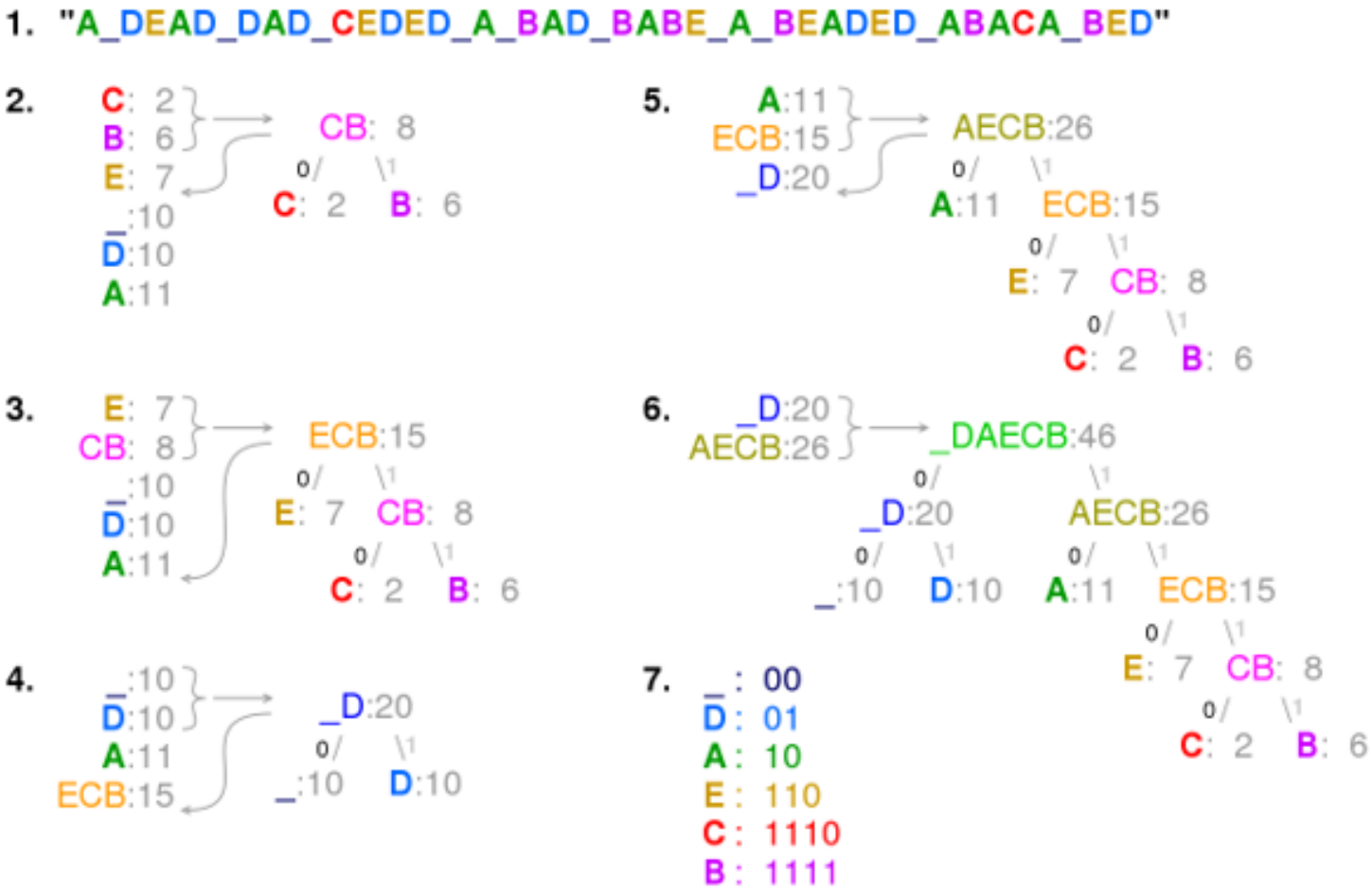
Lossless Compression

Run Length Coding



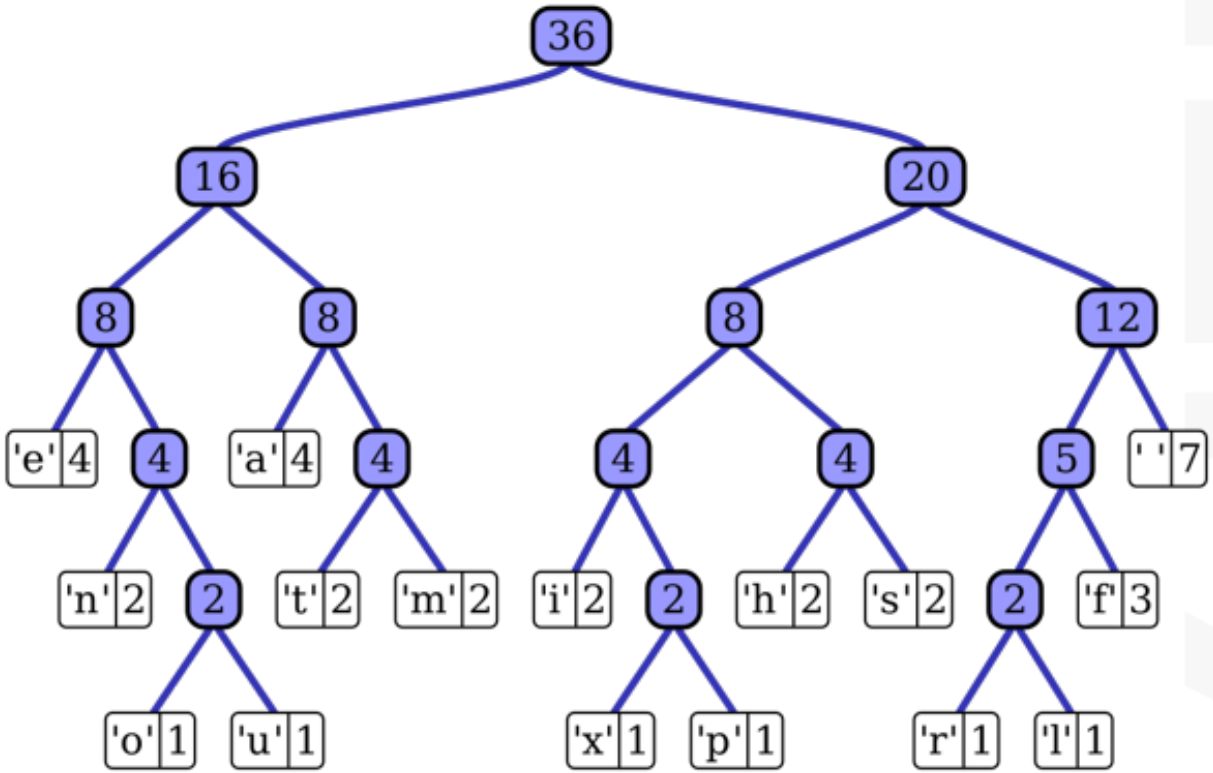
Lossless Compression

Huffman Coding



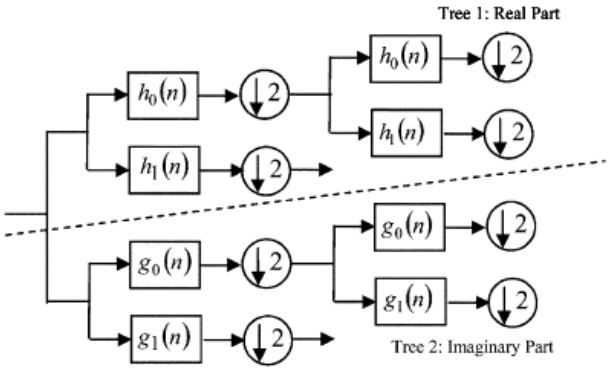
Exercise

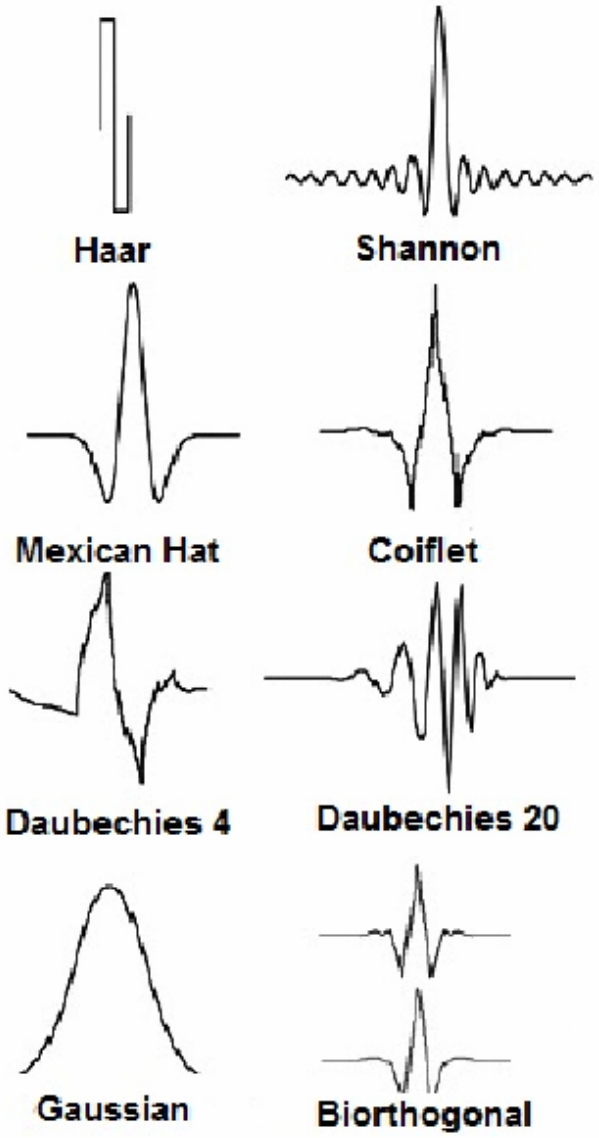
Please use Huffman coding to code "`this is an example of a huffman tree`"



JPEG2000 Standard

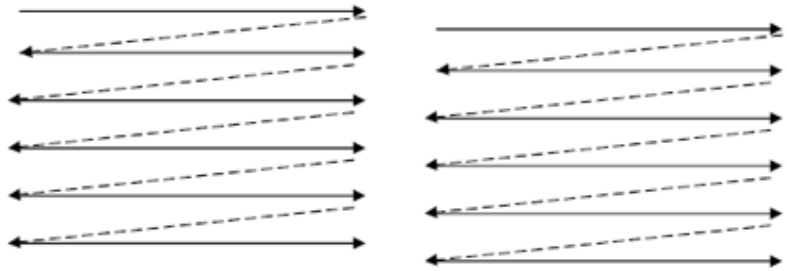
- Based on Wavelet transform
- Support Scalable Coding
  - Decoding can be terminated at any point
  - Coarse-image is transmitted first, then more and more details coming





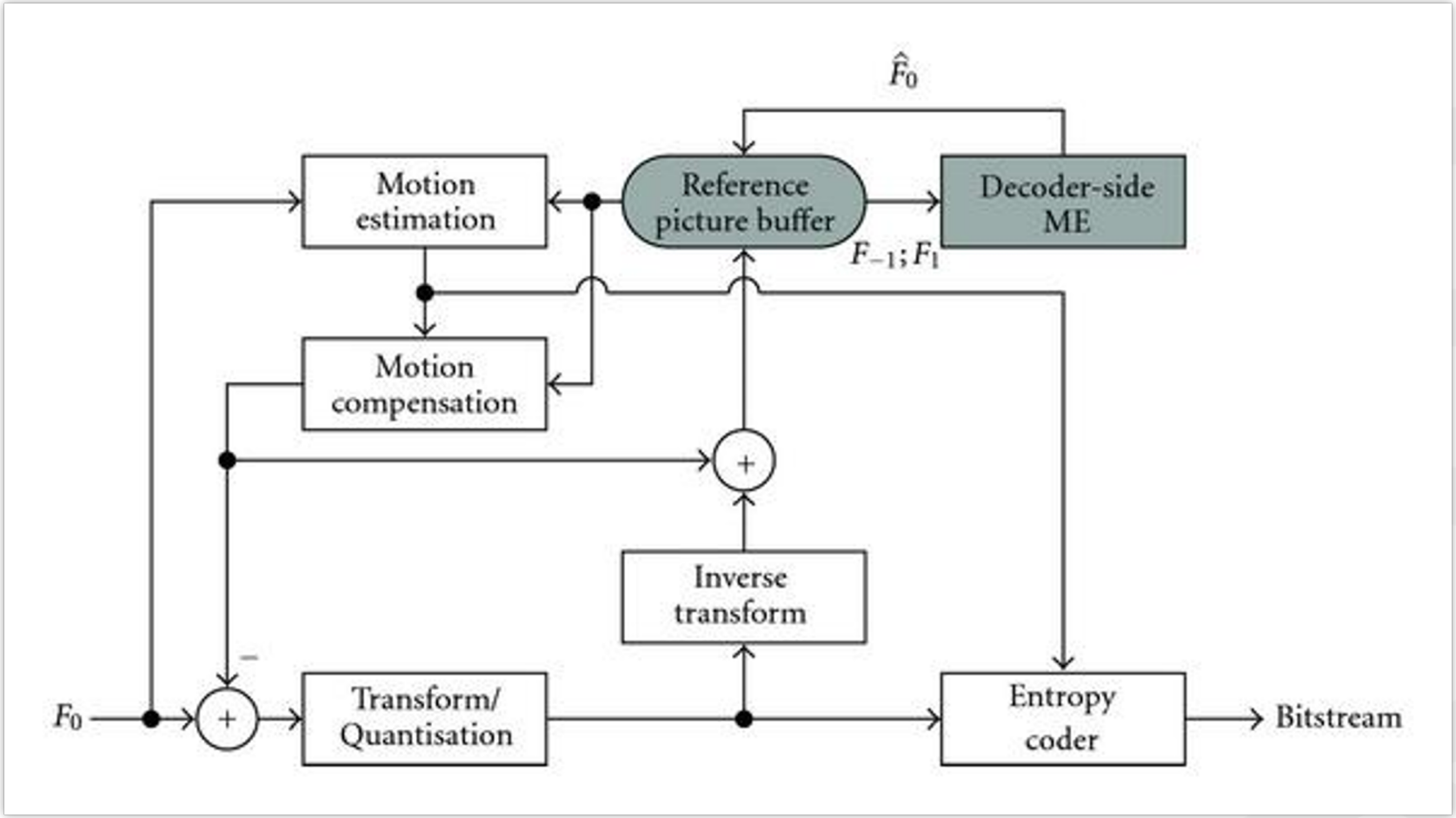
Video Format Standards

- PAL vs NTSC
- Interlaced vs Progressive



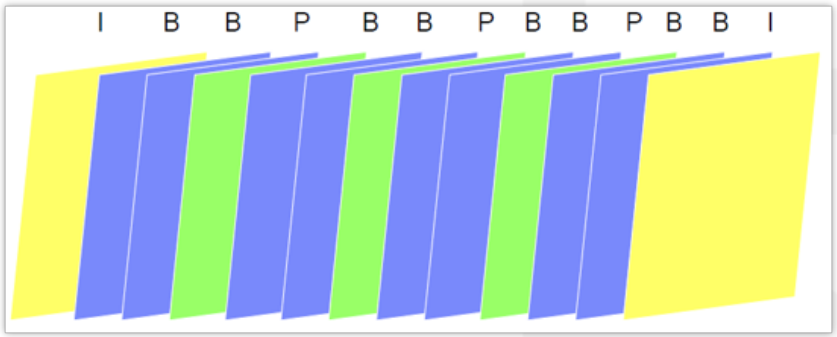
- 1080i, 1080p
- 4K $\rightarrow$ Ultra High Definition

MPEG Video Compression Framework



Group of Pictures

- A group of pictures containing three types of frames
  - I frames: Intra-frame coding frames
  - P frames: Prediction frames
  - B frames: Bi-directional prediction frames
- Play order: IBBPBBPBB
- Transmission order: IPBBPBBBB



I frames

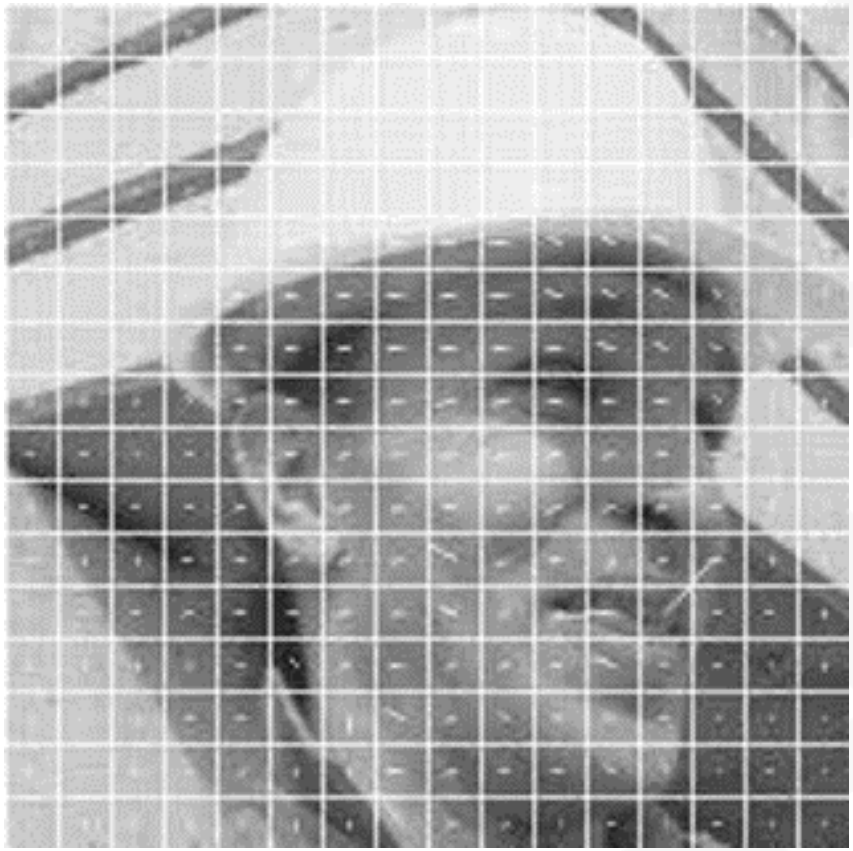
JPEG compression standard with minor differences

P & B frames

- Motion Estimation
- Motion Compensation: Additional intra-frame coding blocks for microblocks with significant differences (P frames ONLY)
- Key problems
  - Identify motion compensation
  - Efficient motion estimation

Motion Estimation Example

- Move the original block and compare the distance with it's the corresponding blocks in the target frame
- Searching from nearest neighbors to outer locations. If the error distance between the original block and the shifted block in the target frame < a threshold, then quite searching
- Search can be based on the motion vector of spatially adjacent blocks and previous (time = t – 1) motion estimation



Video Compression Standards

- MPEG-1
- MPEG-2 / H. 262
- MPEG-4 Part 10
  - Flexible microblock size
  - Flexible motion compensation
  - Efficient motion estimation