COMP3600 Multimedia Systems

Data Compression

Wai Wong

Coding Requirements

Let us consider the general requirements imposed on most multimedia systems:

Storage — multimedia elements require much more storage space than simple text. For example, a full screen true colour image is

$$640 \times 480 \times 3 = 921600\,\text{bytes}.$$

The size of one second of uncompressed CD quality stereo audio is

$$44.1kHz \times 2 \times 2 = 176400\,\text{bytes}.$$

The size of one second of uncompressed PAL video is

$$384 \times 288 \times 3 \times 25\,\text{frames} = 8294400\,\text{bytes}.$$

Throughput — continuous media require very large throughput. For example, an uncompressed CD quality stereo audio stream needs $176400\,\text{bytes}/sec$. A raw digitised PAL TV signal needs

$$\begin{aligned}(13.5MHz + 6.75MHz + 6.75MHz) \times 8\text{bits} \\ = 216 \times 10^6\text{bits}/sec \\ = 27 \times 10^6\text{Bytes}/sec.\end{aligned}$$

Interaction — to support fast interaction, the end-to-end delay should be small. A 'face-to-face' application, such as video conferencing, requires the delay to be less than $50ms$. Furthermore, multimedia elements have to be accessed randomly.

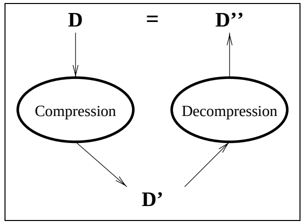
Conclusion:

- Multimedia elements are **very large**.
- We need to reduce the data size using **compression**.

Kinds of coding methods

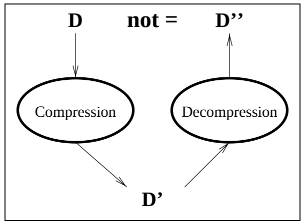
- *lossless* — the compression process does *not* reduce the amount of information.

  The original can be reconstructed exactly.



- *lossy* — the compression process reduces the amount of information.

  Only an approximation of the original can be reconstructed.



Categories of Compression Techniques

Entropy coding	Run-length coding	
	Huffman coding	
	Arithmetic coding	
Source coding	Prediction	DPCM
		DM
	Transformation	FFT
		DCT
	Vector Quantisation	
Hybrid coding	JPEG	
	MPEG	
	H.261	

- Entropy coding is lossless.
- Source coding and hybrid coding are lossy.

Coding techniques

- *Vector Quantisation* — a data stream is divided into blocks of n bytes (where $n > 1$). A predefined table contains a set of patterns is used to code the data blocks.
- *LZW* — a general compression algorithm capable of working on almost any type of data. It builds a data dictionary of data occurring in an uncompressed data stream. Patterns of data are identified and are matched to entries in the dictionary. When a match is found the code of the entry is output.

Pattern	Code
Begin	1
End	2
tion	3
...	...

Input function Begin x := 3; ...
Output `func@3 @1@4` ...

Since the code is shorter than the data pattern, compression is achieved. The popular zip application used this method to compress files.

- *Differential coding* — (also know as *prediction* or *relative* coding) The most known coding of this kind is DPCM (*Differential Pulse Code Modulation*). This method encodes the difference between the consecutive samples instead of the sample values. For example,

PCM	215	218	210	212	208	...
DPCM	215	3	-8	2	-4	...

DM (*Delta Modulation*) is a modification of DPCM. The difference is coded with a single bit.

Run-length coding

Digital images, audio and video data stream often contain sequences of the same bytes.

By replacing the sequence with its length, a substantial reduction of data can be achieved.

A special flag that does not occur in the data can be used to indicate the length of sequence.

For example:

$$ABCD\,\underbrace{CCCCCCCC}\,DBEC$$

compressed to

$$\downarrow$$

$$ABCD \quad C!8 \quad DBEC$$

In practice, there are many variances of this basic compression techniques.

One particular variance, known as *zero suppression*, assumes that only one symbol in the data stream appear very often, for example, a 0 in scanned text. A sequence of zeros is replaced by a flag and the number of occurrences.

Tabulators

A further variation is known as *tabulators*. A table of symbols is defined to replace sequences of different lengths of identical bytes/bits. For example, the CCITT Group 3 compression scheme defines the coding table shown.

For example,

`11100000111100`

is coded as

`1011000110111`

The compression ratio can be up to $10:1$ to $20:1$. The codes for run-length of 0 to 64 are known as *terminating* codes. For run-length longer than 64 bits, it is encoded using more than one code. One or more *makeup* codes represent multiple of 64 bits of the run-length, and this is followed by a terminating code. For example, the run-length of 132 white pixels is encoded by the following two codes:

- makeup code for 128 white pixels — 10010
- terminating code for 4 white pixels — 1011

White	Codes	Black	Codes
0	00110101	0	0000110111
1	000111	1	010
2	0111	2	11
3	1000	3	10
4	1011	4	011
5	1100	5	0011
6	1110	6	0010
7	1111	7	00011
8	10011	8	000101
9	10100	9	000100
10	00111	10	0000100
...	...	...	...
64	11011	64	0000001111
128	10010	128	000011001000
192	010111	192	000011001001
...	...	...	...
1728	01001011	1728	0000001100101
EOL	000000000001	EOL	00000000000

Huffman coding

The principle of *Huffman coding* is to assign shorter code for symbol that has higher probability of occurring in the data stream.

The length of the Huffman code is optimal.

For example, a data stream has only five symbols $ABCDE$ with the following probabilities:

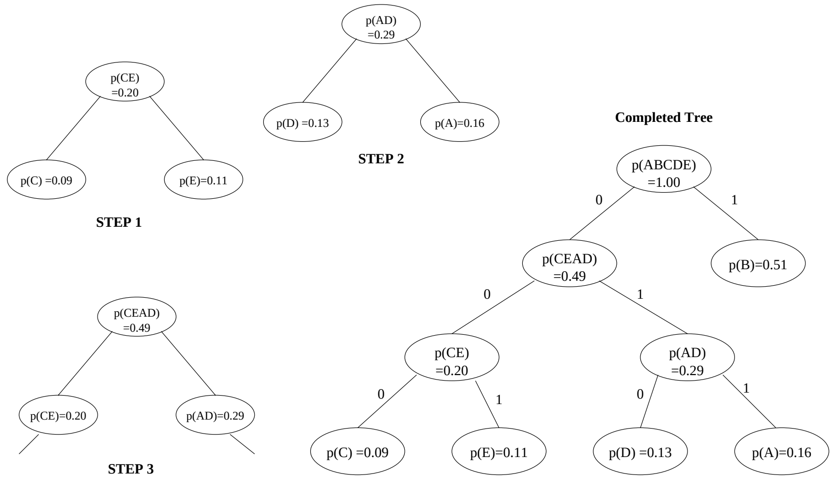
$$\begin{aligned} p(A) &= 0.16 \\ p(B) &= 0.51 \\ p(C) &= 0.09 \\ p(D) &= 0.13 \\ p(E) &= 0.11 \end{aligned}$$

A huffman code tree is created using the following procedures:

- two characters with the lowest probabilities are combined to form a binary tree.
- the two entries in the probability table is replaced by a new entry whose value is the sum of the probabilities of the two characters.
- repeat the two steps above until there is only one entry and a binary tree containing all characters is formed.
- assign 0 to the left branches and 1 to the right branches of the binary tree.
- the Huffman code of each character can be read from the tree starting from the root.

A	011
B	1
C	000
D	010
E	001

An example of Huffman code tree



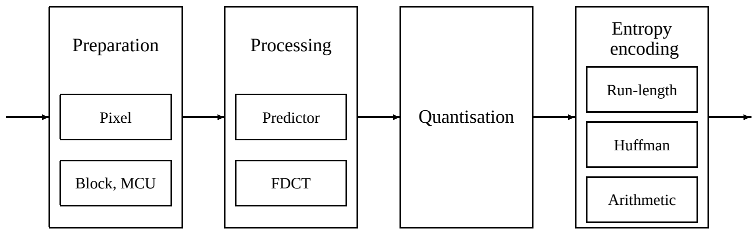
JPEG

- JPEG (stands for Joint Photographic Experts Group) is a joint ISO and CCITT working group for developing standards for compressing still images
- The JPEG image compression standard became an international standard in 1992
- JPEG can be applied to colour or gray-scale images
- By changing appropriate parameters, the user can select
  - the quality of the reproduced image
  - compression processing time
  - the size of the compressed image

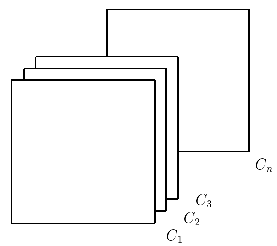
The JPEG standard have three levels of definition as follows:

- *Baseline system* — must reasonably decompress colour images, maintain a high compression ratio, and handle from 4bits/pixel to 16bits/pixel.
- *Extended system* — covers the various encoding aspects such as variable-length encoding, progressive encoding, and hierarchical mode of encoding.
- *Special lossless function* — ensures that at the resolution at which the image is compressed, decompression results in no loss of any detail the was in the original image.

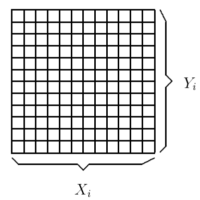
JPEG Compression Process



JPEG — Preparation



- A source image consists of at least one and at most 255 planes.
- Each plane C_i may have different number of pixels in the horizontal (X_i) and vertical (Y_i) dimension.
- The resolution of the individual plane may be different.
- Each pixel is represented by a number of bits p where $2 \leq p \leq 12$.
- The meaning of the value in these planes is not specified in the standard.
- The image is divided into 8×8 blocks.



JPEG — Discrete Cosine Transform

- DCT transforms the data from a spatial domain to a frequency domain.
- It removes redundancy in the data.
- It is proven to be the optimal transform for large classes of images.
- The DCT algorithm is symmetrical, and an inverse DCT algorithm can be used to decompress an image.

The DCT coefficients of each 8×8 blocks are calculated using the formula below.

$$DCT(i,j) = \frac{1}{\sqrt{2N}} C_i C_j \sum_{x=0}^{7} \sum_{y=0}^{7} p_{xy} \cos\left[\frac{(2x+1)i\pi}{2N}\right] \cos\left[\frac{(2y+1)j\pi}{2N}\right]$$

where $C_i, C_j = \frac{1}{\sqrt{2}}$ for $i, j = 0$, otherwise $C_i, C_j = 1$.

Here is an example. On the left is the 8×8 block of pixels, and on the right is the DCT coefficients.

132	126	138	140	144	145	147	155
136	140	140	147	140	148	155	156
140	143	144	148	150	152	154	155
144	144	146	145	149	150	153	160
150	152	155	156	150	145	144	140
144	145	146	148	143	158	150	140
150	156	157	156	140	146	156	145
148	145	146	148	156	160	140	145

172	-18	15	-8	23	-9	-14	19
21	-34	24	-8	-10	11	14	7
-9	-8	-4	6	-5	4	3	-1
-10	6	-5	4	-4	4	2	1
-8	-2	-3	5	-3	3	4	6
4	-2	-4	6	-4	4	2	1
4	-3	-4	5	6	3	1	1
0	-8	-4	3	2	1	4	0

JPEG — Quantisation

- The DCT output matrix is quantised to reduce the precision of the coefficients
- This increases the compression
- $DCT(0,0)$ is known as the *DC coefficient* which represents the basic colour, i.e., wave-length, of the image block
- The other DCT coefficients are known as *AC coefficients* which represent the frequency components of the data block
- AC coefficients further away from the DC coefficient can be dropped to reduce the data size
- JPEG baseline algorithm defines a set of quantisation tables
- Each element q in the table, known as *quantum* is used in the following formula to calculate the quantised coefficients Q:

$$Q_{ij} = \frac{DCT_{ij}}{q_{ij}}$$

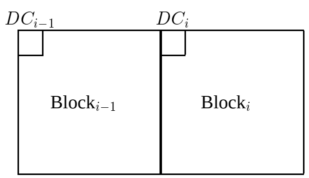
On the left is quantum matrix for quality level 1, and on the right the result of quantising the example from previous page.

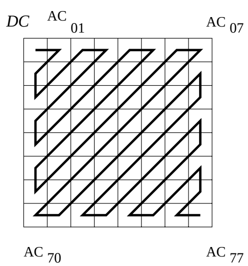
4	7	10	13	16	19	22	25
7	10	13	16	19	22	25	28
10	13	16	19	22	25	28	31
13	16	19	22	25	28	31	34
16	19	22	25	28	31	34	37
19	22	25	28	31	34	37	40
22	25	28	31	34	37	40	43
25	28	31	34	37	40	43	46

43	3	2	0	0	0	0	0
3	3	2	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

JPEG — Entropy Encoding

- The DC-coefficients are treated separately from the AC-coefficients
- A DC-coefficient is encoded as the difference between the current coefficient and the previous one
- The AC-coefficients are encoded using Huffman and/or arithmetic encoding
- They are put in a zig-zag order as shown in the diagram below





MPEG

- MPEG (stands for Moving Picture Experts Group) is also a joint ISO and CCITT working group for developing standards for compressing still images
- The MPEG video compression standard became an international standard in 1993
- MPEG uses technology defined in other standards, such as JPEG and H.261
- It defines a basic data rate of 1.2Mbits/sec
- It is suitable for symmetric as well asymmetric compression It follows the reference scheme that consists of four stages of processing:
  1. Preparation
  2. Processing
  3. Quantisation
  4. Entropy Encoding

In the preparation stage, unlike JPEG, MPEG defines the format of the images.

- Each image consists of three components — YUV
- The luminance component has twice as many samplesin the horizontal and vertical axes as the other two components (known as *colour sub-sampling*
- The resolution of the luminance component should not exceed 768 pixels
- for each component, a pixel is coded with eight bits

How MPEG encode the video stream

In order to achieve higher compression ratio, MPEG uses the fact the image on consecutive frames differ relative small. It uses a *temporal prediction* technique to encode the frame so that the storage requirement is greatly reduced.

Common MPEG data stream consists of four kinds of frames:

- *I-frame* (Intra-frame) — it is a self contained frame, and it is coded without reference to any other frames.
- *P-frame* (Predictive-coded frame) — It is coded using the predictive technique with reference to the previous I-frame and/or previous P-frame.
- *B-frame* (Bi-directionally predictive-coded frame) — It requires information of the previous and following I- and P-frames for encoding and decoding.
- *D-frame* (DC-coded frame) Only the lowest frequency component of image is encoded. It is used in fast forward or fast rewind.

Group of Pictures

The I-, P- and B-frames are often arranged into groups, known as *Group of pictures* (GOP). A typical GOP consists of twelve frames in the following sequence:

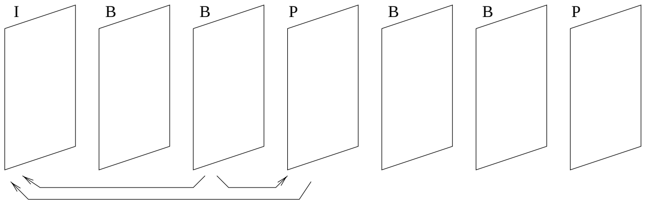
IBBPBBPBBPBB

When the video is transfered and decoded, the frames are in the following order:

I P B B P B B P B B B B

This is known as the transmission sequence.

When random access function is required, the I-frames will be used as index, and the accessing to the video will be in a resolution of twelve frames.



I-frame encoding

I-frames are compressed using JPEG compression technique, however, the compression/decompression processing must be carried out in real-time, i.e., on average $1/30sec \simeq 33ms$ for 30 frame/sec video stream.

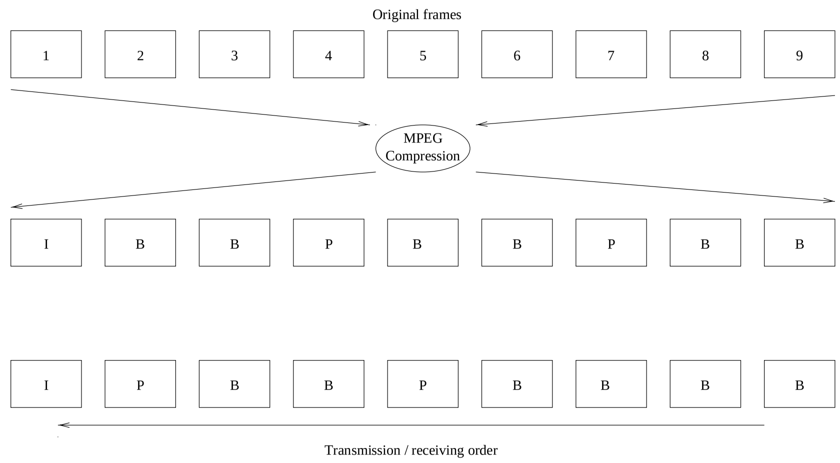
MPEG-2

- MPEG-2 is a newer video encoding standard which builds on MPEG-1
- It supports higher video quality and higher data rate (up to 80 Mbits/sec)
- It supports several resolutions:

pixels/line	line/frame	frames/sec
352	288	30
720	576	30
1920	1152	60

Summary

- Compression methods — lossless vs. lossy
- Entropy coding — run-length encoding, Huffman encoding
- Source coding — prediction (DPCM, DM), transformation (DCT)
- hybrid coding — JPEG, MPEG


Original frames
1
2
3
4
5
6
7
8
9
MPEG
Compression
I
B
B
P
B
B
P
B
B
I
P
B
B
P
B
B
B
B
Transmission / receiving order