MPEG

Agenda

MPEG: standardised set of algorithms
Video Compression exploits temporal redundancy
Two main types of compression: spatial and temporal
Three types of frame:

- Intra Coded Frames (I)
- Predictive Coded Frames (P)
- Bi-directional Predictive Coded Frames (B)

MPEG 4 encodes media object
Other codecs

Introduction to MPEG

MPEG = Moving Pictures Experts Group (Standardised set of algorithms)

MPEG-1:

allow moving pictures and sound to be encoded into the bitrate of a Compact Disc (CD).
Commonly limited to about 1.5 Mbps (Megabits per second).
low bit requirement, MPEG-1 downsamples the images, and uses frame rates of only 24–30 Hz.
Only supports **progressive pictures** (progressive and interlaced)
Includes the popular MPEG-1 **Audio Layer III (MP3) audio compression format**.

MPEG-2:

Transport, video and audio standards for broadcast-quality television.
Supports **interlacing** and high definition
Chosen as the compression scheme for over-the-air digital television (e.g. DVB), digital satellite TV services (e.g. Dish Network), digital cable television signals, SVCD and DVD Video.

Video compression

rarely done in lossless form
Video: a sequence of pictures (or frames)
JPEG algorithm used for intra-frame coding
– e.g. moving JPEG (MJPEG) exploits only intra-frame coding (spatial redundancy)
High correlation between successive frames (temporal redundancy)
– motion estimation and motion compensation: inter-frame coding
– use a combination of actual frame contents and predicted frame contents

Inter-frame and intra-frame coding

Inter-frame compression: uses one or more earlier or later frames in a sequence to compress the current frame.
Intra-frame compression: uses only the current frame, which is effectively image compression.

MPEG 1 & 2 use both inter-frame and intra-frame coding

Temporal redundancy

Temporal redundancy occurs when, in a sequence of images, consecutive images contain similarities

With temporal compression, only the changes from one frame to the next are encoded. (a subtractive image can be used to describe the diffrences and similarities)

Possible causes of change: illumination | camera movement | noise | object motion | background change etc

Two main types of compression

Spatial compression (intra-frame coding)

Allow fast random access

Applied to key frames (e.g. every 12th frame, start of new scene, etc)

I-frames

Temporal compression (inter-frame coding)

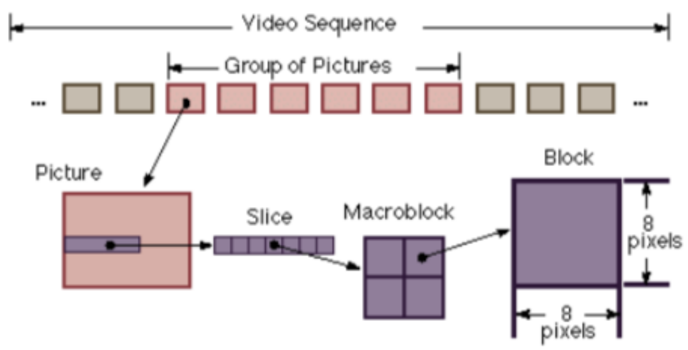
Allow **high compression**

Exploit temporal redundancy between frames

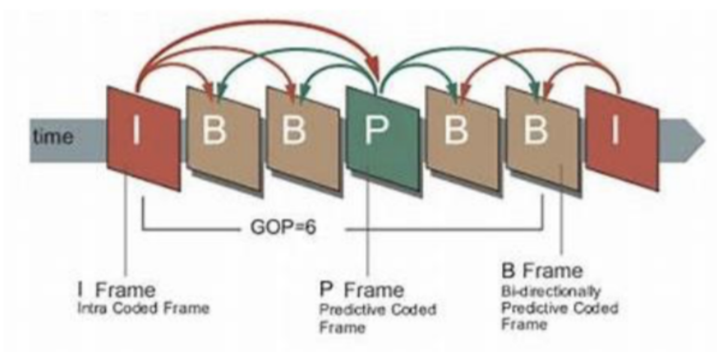
Applied to frames occurring between key frames – Based on motion detection

P-frames & B-frames

Structure of Digital Videos



Group of Pictures (GOP)



Typically, GOP have one I, several P and Bs (IBBPBBPBB)

MPEG video encoding (generalization)

- input frames are preprocessed

- color space conversion
- spatial resolution adjustment (chroma sub-sampling)

- frame types (Intra or Inter) are decided for each frame/picture
- each picture is divided into macroblocks of 16 X 16 pixels
- macroblocks
  - are intracoded for I frames
  - are predictive coded or intracoded for P and B frames
  - are divided into six blocks of 8 X 8 pixels
    - – 4 luminance and 2 chrominance
    - – DCT is applied to each block →transform coefficients
      - » quantized
      - » zig-zag scanned
      - » variable-length coded

Three types of frame:

– Intra Coded Frames (I)
– Predictive Coded Frames (P)
– Bi-directional Predictive Coded Frames (B)

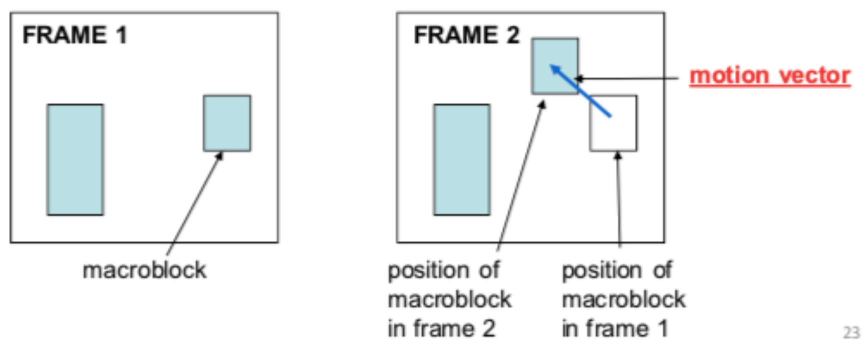
I-frames (Intra-coded frames)

- Self-contained (without reference to other frame)
- Treated as a still image and make use of JPEG
- Compression rate is the lowest
- Points for random access in video stream

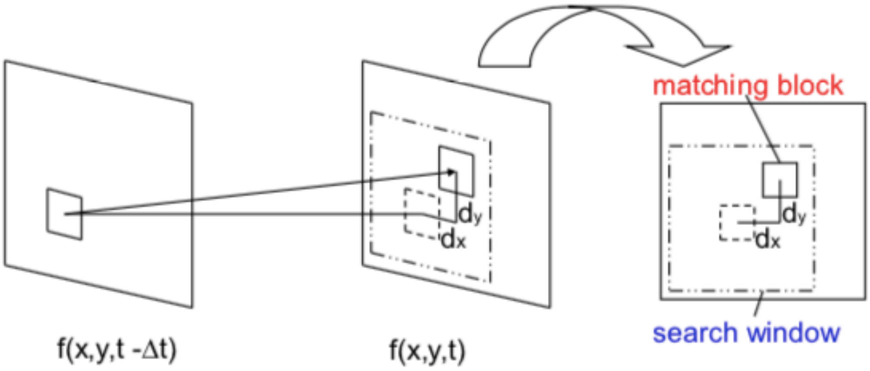
Motion estimation

Based on looking for matching macroblocks of data in successive frames, and measuring the difference between matching macroblocks in terms of:

– Position within the frame (**motion vectors**)
– Luminance and colour difference (**error term**)



For a given block in the target frame, finding a macroblock within the **search windows** of the reference frame



$$\min \sum_{B} \left\| f(x,y,t) - f(x-d_x, y-d_y, t-\Delta t) \right\|$$

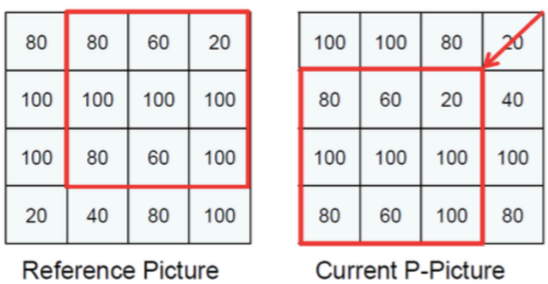
Matching macroblocks: only if a "close" match can be found
– Evaluate "closeness" with MSE or other metric
– If no suitable match found, just encode the macroblock as an I-block
Not too many P or B frames in a row
– Predictive error will keep propagating until next I frame
– Delay in decoding
Three situation:

1. One identical macroblock is found: no need to encode anything
2. No matching macroblock is found: use intra-frame encoding
3. One similar macroblock is found: encode difference values between the two macroblocks and a motion vector

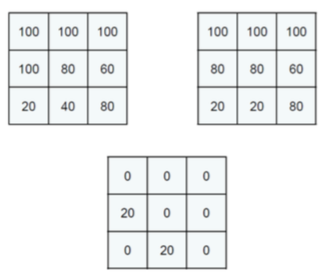
P-frames (Predictive Coded Frames)

- Use a reference frame for motion estimation
- Hold only the changes in the image from **a previous frame**
- Compression rate is higher

Each macroblock in a P-frame can be encoded either as an I-macroblock or as a P-macroblock.
P-macroblock: past reference frame -> motion vector + error term



Motion vector

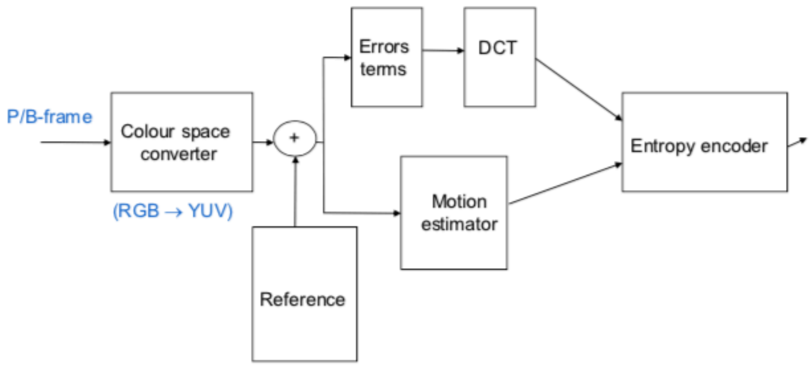


Error term

B-frames (Bi-directional Predictive Coded Frames)

- B-frame: search for matching MBs in both past and future frames
- typical pattern is IBBPBBPBB IBBPBBPBB IBBPBBPBB

P- and B- frames (summary)

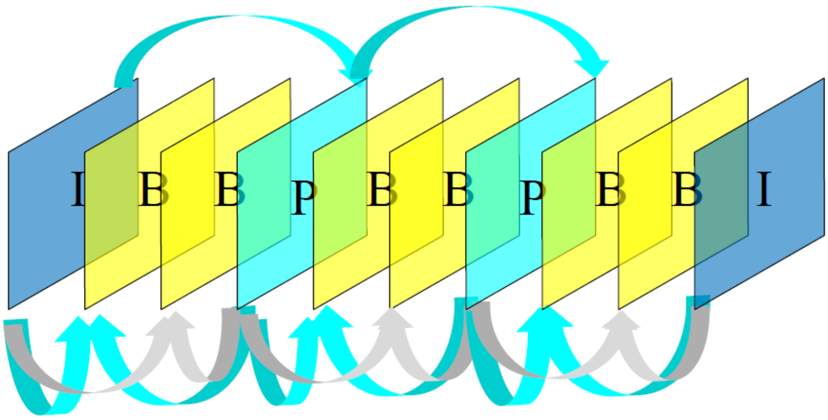


P-frame contains 1/3 data of I-frame
– motion vectors
– error macroblock
B-frame contains 1/2 to 1/5 data of P-frame
– least data but most computational
– **delay issue:** need to transfer future I or P frames before any dependent B-frames can be processed

Compressions rate: B>P>I

Group Of Pictures

I frames do not rely on other frames
P frames rely on previous I and P frames
B frames rely on previous and future I and P frames
No frames rely on B frames



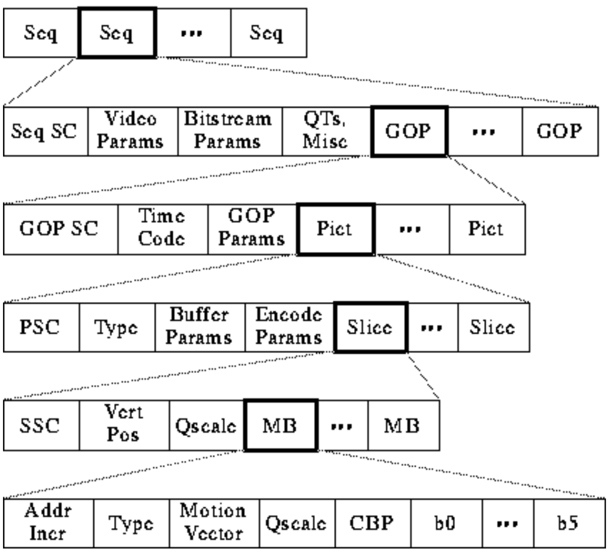
Transmissions order:
1 2 3 4 5 6 7 8 9 -> I B B B P B B B I
order: **1 5** 2 3 4 **9** 6 7 8
Question:

What can you say about the content of an MPEG video that contains very few I frames?
Solution: Since the appearance of I frame means that this frame is really different from the consecutive frames. In an MPEG video, very few I frames indicates that this video has only a few large changes, such as changing the background or main character in the scene. Mostly, the consecutive frames have changes such as in light, color, object position and etc.
(Solution: A video that contains very I frames has a lot of P and B frames. This means that the content is easily predictable, i.e. there is little change from one frame to the next. The video doesn't contain much action, it could be the video of an interview)

Bitrate Allocation

CBR–ConstantBitRate
– Streaming media uses this
– Easier to implement
VBR–VariableBitRate
– DVD's use this
– Usually requires 2-pass coding
– **Allocate more bits for complex scenes**

MPEG Video Bitstream



MPEG 1&2 summary

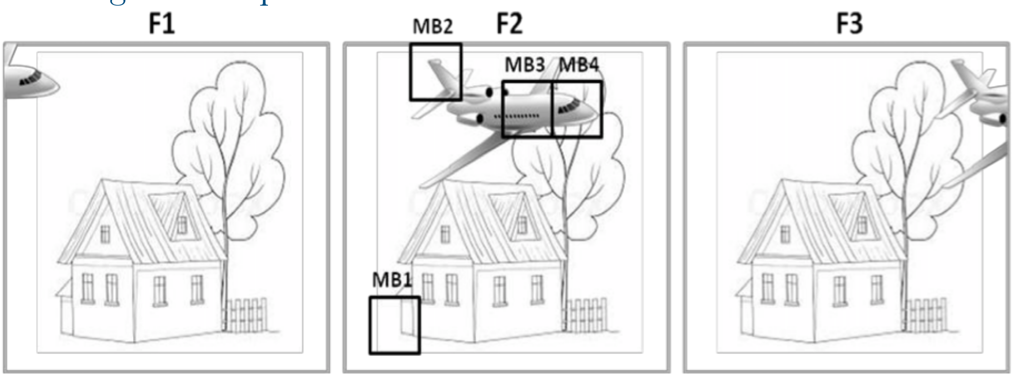
Images coded INTRA (I):
– Random access
– Error resiliance
Images coded INTER(P):
– Prediction from previous decoded image (I, P)
Images coded BI-INTER (B):
– Prediction from previous and / or future decoded image (I, P)

– allow effective prediction of **uncovered background** (areas of the current picture that were not visible in the past and visible in the future)

Question 1:
Explain how the choice of GOP structure affects the size of the MPEG video file
Solution: For a GOP structure, it typically have one I frames, several P and B frames. Since P frames are 1/2 size of I frames, B frames are 1/5 to 1/3 size of P frames (spatial compression offer large compression), more P frames and B frames in a GOP structure can reduce the size of the MPEG video file.
(Solution: When GOP structures are long (many P and B frames), the total number of I frames in the video will be small. Since I frames are the biggest, the video files containing a small number of them will tend to be smaller in size)

Question 2:
The frames of any GOP structure need to be re-ordered before transmission. Why?
Solution: Since for B-frames, it has a delay issue, which means B-frames need future I and P frames for reference. Therefore, the future I and P frames that are required by B-frames need to be transmitted first.
(Solution: The B frames contained in a GOP need two reference frames in order to be reconstructured. One of these references is a "future" frame, i.e. it normally comes later in the GOP structure. To make it available for decoding the B frame, it has to be sent before, hence the need to re-order the data stream)

Question 3:
The figure below shows three consecutive frames (F1, F2 and F3) of a video. In the middle frame (F2), four macroblocks (MB1, MB2, MB3 and MB4) are shown. For each of these four macroblocks, briefly explain how it should be encoded in MPEG1 to achieve the highest compression.



Solution: MB3 should be intra encoded, MB1 and MB3 should be bi-inter encoded, MB2 should be inter encoded in MPEG1 to achieve the highest compression. Since there is not a matching macroblocks in F1 and F3, MB1 should be intra encoded. A matching block for MB2 can be found in F3, so it should be bi-inter encoded. A matching block for MB4 can be found in F1, it should be inter encoded. ~~For MB1, matching blocks can be found in both F1 and F3, since we need to achieve the highest~~

~~compression, bi-inter encoding offer larger compression than inter encoding, it should be bi-inter encoded.~~
Solution:
MB1: this macroblock is exactly the same than in F1, nothing needs to be encoded.
MB2: a similar (although not identical) macroblock can be found in F3 so this should be a B (forward prediction) macroblock.
MB3: this macroblock cannot be matched with any other, it must be an I (intra coded) macroblock.
MB4: a similar (but not identical) macroblock can be found in F1 at a different position, this can be a P (predicted macroblock).

MPEG 4 encodes media objects

MPEG 4: provides standardised ways of representing units of aural, visual or audio-visual content, **as discrete "media objects"**
Coding of audio-visual objects:
– Objects can be coded and decoded independently
– Object can be flexibly composed to create different scenes
– Natural and synthetic object are treated in the same way
– Excellent error resilience

Other codecs

Codec: a method for encoding and decoding data, more specifically, a protocol for compressing data, especially video
Container: what holds the grouping of compressed video as defined by the codecs. The container takes care of packaging, transport and presentation
– Usually represented by a file extension

Part 10 and H.264 AVC:
- Approx. half the bit rate of MPEG2/MPEG4 Part 2
- 16 reference frames can be used (limit of 2 for B-frames)
- VBSMC (block size from 16 x 16 to 4 x 4)
- supporting a wide range of bit rates and a wide range of resolutions.

QuickTime
- developed by Apple Computer
- The QuickTime Moving (.mov) file format, an open source media container

AVI
- Video for Windows multimedia container format
- Microsoft

WMV
- Windows Media Video
- Adopted as format for Blu-ray discs

Flash Video
Container file format used to deliver video over the Internet : FLV or F4V

HTML5 Video
A HTML 5 tag introduced for the purpose of playing videos or movies

Questions in test

a) This question is about MPEG.

[14 marks]

i) What type of MPEG frame makes no prediction?

(1 mark)

ii) What type of MPEG frame is never used as a reference frame during decompression? Why?

(3 marks)

iii) Suppose an MPEG encoder uses the nine-frame sequence IBBPBBPBB. Draw a diagram showing the dependencies between the first 12 frames of a compressed clip produced by this encoder.

(3 marks)

iv) Briefly explain the Block Matching Algorithm (BMA) for motion estimation when a maximum motion displacement of x pixels is used.

(4 marks)

v) Given a maximum motion displacement of 8 pixels, how many evaluations of the matching criterion are required in the BMA?

(1 mark)

vi) Briefly explain the basic principle of fast motion estimation techniques.

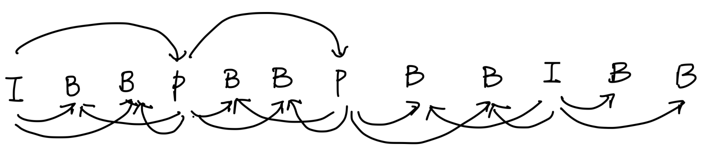
(2 marks)

Solution:

(1) I-frame

(2) B-frame is never used as a reference frame during decompression. Since for B-frames, they are coded on the basis of P-frames and I-frames. As the result, only the difference with the reference frames are stored. In other words, B-frames are not reference frames for any other frames during compression. During decompression, we need to recover the inter-coded frames by adding the difference and the reference frames. In this case, since B-frames are not used as reference during coding process, it is also never used as reference frames during decompression.

(3)



iii) In MPEG, why are B frames typically smaller than I and P frames?
Solution: Since B frames are coded based on both previous and future I and P frames, the spatial redundancy is large in B frames. Ultimately, only the difference between B frames and reference frames need to be stored. Therefore, B frames are typically smaller than I and P frames.

c) This question is about MPEG1 video compression. Consider the three consecutive video frames shown in Figure 2. The middle frame is a B frame. It contains 4 macroblocks.

[8 marks]

i) What predictions can be made about the top left macroblock? How will it be encoded?

(3 marks)

ii) What predictions can be made about the bottom right macroblock? How will it be encoded?

(5 marks)

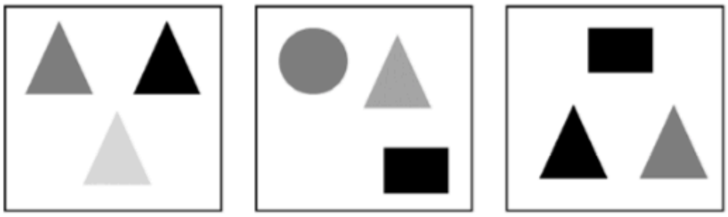


Figure 2: Three consecutive video frames

(1) For the top left macroblock, since we cannot find a matching macroblock in previous consecutive frame or future consecutive frame. It will be intra encoded.
(2) For the bottom right macroblock, we can find a matching macroblock in the future consecutive frame. From the future matching macroblock, it shows that the current macroblock goes left and up. It will be bi-direction predictional encoded.

Why are B-frames typically more compressed than P-frames?
Solution: because they can make better predictions thanks to using two reference frames rather than one