Digital video broadcasting module

MPEG VIDEO ENCODING

h.shirazi@qmul.ac.uk

Dr Hamid Shirazi

Table of Contents

Week3: Day-3

MPEG Standards

MPEG = Moving Pictures Expert Group

MPEG-1	**MPEG-2**	**MPEG-4**	**MPEG-7**	**MPEG-21**
Part1: systems ISO/IEC11172-1 "PES layer"	Part1: systems ISO/IEC13818-1 "Transportation"	Part1: systems ISO/IEC14496	Metadata, XML based ISO/IEC15938 "Multimedia Content Description Interface"	additional "tools" ISO/IEC21000
Part2: video ISO/IEC11172-2	Part2: video ISO/IEC13818-2	Part2: video ISO/IEC14496-2		
Part3: audio ISO/IEC11172-3	Part3: audio ISO/IEC13818-3	Part3: audio (AAC) ISO/IEC14496-3		
	Part6: DSM-CC ISO/IEC13818-6	Part10: video (AVC, H.264) ISO/14496-10		
	Part7: AAC ISO/IEC13818-7			

Transforing to and from Frequency Domain (brief)

- As a background to this section you should know the basic of methods which are used to to transform to and from Frequency Domain. The most important methods are as follows.
    - Fourier transform
    - Discrete Fourier Transform (DFT) and Inverse DFT (IDFT)
        - Discrete points sampling in the time domain at intervals Δt during a limited time window at N points
        - Simple but time consuming algorithm
        - Defined through field of complex numbers (time and frequency domain signals will have real and imaginary parts) – typically time domain are real
            - Performed in two input signals (real and imaginary part tables corresponding to sampled time or frequency domain)
    - Fast Fourier Transform (FFT)
        - It provides exactly the same result as a DFT but is more complex and much faster and is restricted to $N=2^x$ points
    - Discrete Cosine Transform (DCT)
        - FFT is a special case of DCT
        - Cosine-sine transform attempting to assemble a time domain signal by superposition of many different cosine and sine signals of different frequency and amplitude
        - also called Discrete Sine Transform (DST)
        - it is important for audio and video compression
- A thorough knowledge of these principles is of great importance to understanding the topics on video encoding, audio encoding and Orthogonal Frequency Division Multiplex (OFDM).
- DCT has became the basic algorithm for MPEG

$$H_n = \sum_{k=0}^{N-1} h_k e^{-j2\pi k\frac{n}{N}} = -\sum_{k=0}^{N-1} h_k \cos(2\pi k\frac{n}{N}) - j\sum_{k=0}^{N-1} h_k \sin(2\pi k\frac{n}{N});$$

$$h_k = \frac{1}{N}\sum_{n=0}^{N-1} H_n e^{j2\pi k\frac{n}{N}};$$

Figure 1: DFT and IDFT mathematical relationships

$$F_k = \sum_{z=0}^{N-1} f_z \cos(\frac{\pi k(z+\frac{1}{2})}{N}); \; F_k = \sum_{z=0}^{N-1} f_z \sin(\frac{\pi z k}{N});$$

Figure 2: formula for DCT and DST

SDTV and HDTV resolutions (brief)

- The aspect ratio for HDTV will normally be 16:9 which is also becoming the norm for SDTV.
- Initially HDTV was to be based on twice the number of lines and twice the number of pixels per line.
    - This would result in 1250 lines total with 1152 active lines and 1440 active pixels in a 625 line system and 1050 lines total with 960 active lines and 1440 active pixels in a 525 line system.
    - However, resolution used in
        - ATSC and HDTV in the US is 1280 x 720 pixels at 60 Hz.
        - In Australia it is usually 1920/1440 x 1080 pixels at 50 Hz.
        - The resolution of the European HDTV satellite channel EURO1080 is 1080 active lines x 1920 pixels at a field rate of 50 Hz.
    - Introduction of HDTV led to the use of MPEG-4 instead of MPEG-2 with efficiency factor of 2 to 3
- ITU has generally decided on a total number of 1125 lines in the 50 Hz and 60 Hz system, with **1080 active lines and 1920 pixels per line** both in the 50 Hz and the 60 Hz system.
- **An active image of 1080 lines x 1920 pixels is called the Common Image Format (CIF).**

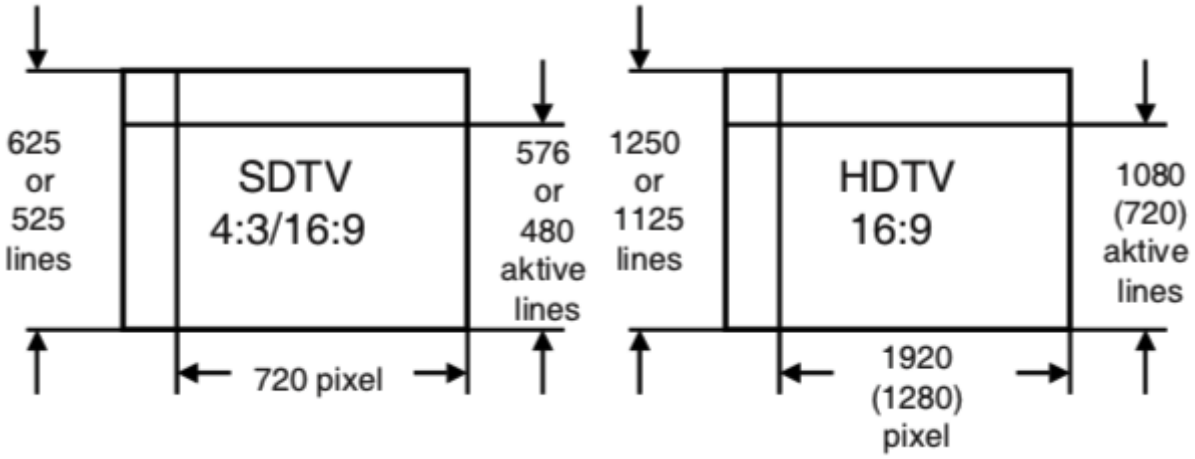


Figure 3: SDTV and HDTV resolution

MPEG Standards

MPEG = Moving Pictures Expert Group

MPEG-1	MPEG-2	MPEG-4	MPEG-7	MPEG-21
Part1: systems ISO/IEC11172-1 "PES layer"	Part1: systems ISO/IEC13818-1 "Transportation"	Part1: systems ISO/IEC14496	Metadata, XML based ISO/IEC15938 "Multimedia Content Description Interface"	additional "tools" ISO/IEC21000
Part2: video ISO/IEC11172-2	Part2: video ISO/IEC13818-2	Part2: video ISO/IEC14496-2		
Part3: audio ISO/IEC11172-3	Part3: audio ISO/IEC13818-3	Part3: audio (AAC) ISO/IEC14496-3		
	Part6: DSM-CC ISO/IEC13818-6	Part10: video (AVC, H.264) ISO/14496-10		
	Part7: AAC ISO/IEC13818-7			

Video coding basic

- SDTV raw signal (uncompressed) will have a date rate of 270 Mbit/s which is too much for broadcast purposes which is why they are subjected to a compression process before being processed for transmission.
- The 270 Mbit/s must be compressed to about 2...7 Mbit/s - a very high compression factor:
    - it is possible due to the use of a variety of redundancy and irrelevance reduction mechanisms.
- The data rate of an uncompressed HDTV signal is even higher than 1 Gbit/s and MPEG-2 coded HDTV signals have a data rate of about 15 Mbit/s.
    - This will be even less when MPEG-4 is applied.
- Compression in a nutshell means removing edundant or irrelevant information from the data stream
    - Redundant means superfluous, irrelevant means unnecessary.
    - Superfluous information is information which exists several times in the data stream, or information which has no information content, or simply information which can be easily and losslessly recovered by mathematical processes at the receiving end.
    - Redundancy reduction can be achieved, e.g. by variable-length (Huffman) coding.
        - Instead of transmitting ten zeroes, the information 'ten times zero' can be sent by means of a special code which is much shorter.
        - E.g. the alphabet of the Morse code (short and long code sequences depending on the frequency of the letters)
    - In video, irrelevant information can be video components which can not be seen by human eyes due to its anatomy (weak in recepting bright colours and thin/coarse structures):
        - Hence, sharpness in the color can be reduced which means a reduction in the bandwidth of the color information
        - Also means loss of information (irretrievable)

7-Step MPEG Data Reduction Process

- In MPEG following steps will be carried out to achieve **data reduction factor of 130**:
    1. 8 bits resolution instead of 10 bits (irrelevance reduction)
    2. Omitting the horizontal and vertical blanking interval (redundancy reduction)
    3. Reducing the color resolution also in the vertical direction (4:2:0) (irrelevance reduction)
    4. Differential pulse code modulation (DPCM) of moving pictures (redundancy reduction)
    5. Discrete cosine transform (DCT) followed by quantization (irrelevance reduction)
    6. Zig-zag scanning with variable-length coding (redundancy reduction)
    7. Huffman coding (redundancy reduction)

Figure 4: Data reduction

Reducing Quantisation from 10 to 8 bits

- In analogy TV video has a signal to noise ratio of 48dB, the noise component is just below the human eye perception
- In analog to digital conversion, the quantisation noise of 8 bits is already well below the human eye perception threshold (outside the studio), however if post processing is needed 10 bit will give a better quality in the studio for post-processing
- Reducing data rate from 10 to 8 bit means 20% reduction
    - Quantisation noise has risen by 12dB (S/N=N*6 dB)

Horizontal and Vertical Blanking Intervals

- According to ITU BT.R601, horizontal and vertical blanking intervals do not contain any relevant information but can contain supplementary information such as sound signal
    - MPEG does not concern with H/V blanking intervals and consider them to be coded seperately
    - The intervals can be regenerated again at the receivers
    - Figure 5 shows PAL system where 8% reduction can be achieved by omitting 50 vertical lines as only 575 lines are visible.
    - Further reduction of 19% can be achieved taking into account that active video area is only 52 µs out of 64 µs.
    - Considering the overlap, almost 25% reduction can be achived by omitting vertical blanking.

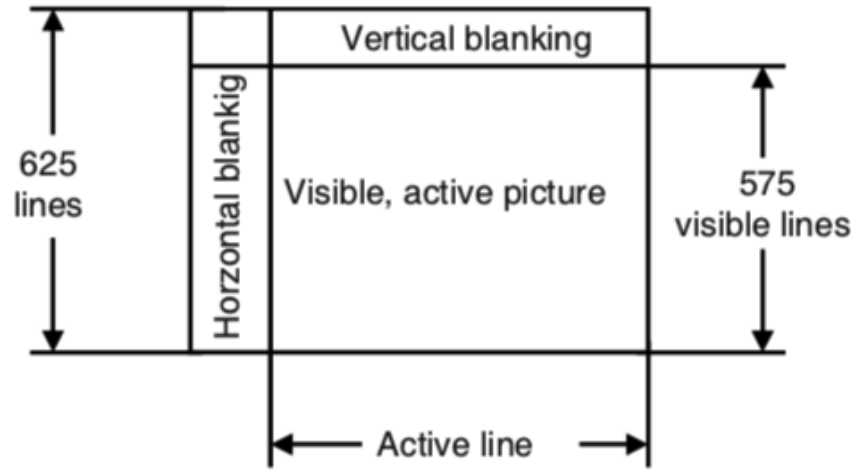


Figure 5: Horizontal and vertical blanking

Reduction in Vertical Colour Resolution (4:2:0)

- The two-color difference signals C_B and C_R are sampled at half the data rate compared with the luminance signal Y:
    - (Y:C_B:C_R) format is (4:2:2).
- The sampling rate of the luminance signal is 74.25 MHz.
- The sampling rate of the color difference signals is 0.5 x 74.25 MHz = 37.125 MHz.
- ITU-R BT.709 provided a sampling rate of 72 MHz for the luminance and 36 MHz for the chrominance (bandwidth reduction).
    - To avoid aliasing, the luminance signal bandwidth is limited to 30 MHz and that of the chrominance signals to 15 MHz by low-pass filtering them before they are sampled.
    - In the 1125/60 system, and with a 10 bit resolution, this results in a gross physical data rate of:
        - Y: 74,25 x 10 Mbit/s = 742.5 Mbit/s
        - CB: 0.5 x 74,25 x 10 Mbit/s = 371.25 Mbit/s
        - CR: 0.5 x 74,25 x 10 Mbit/s = 371.25 Mbit/s
            - In total: 1.485 Gbit/s → gross data rate (1125/60)
    - Because of the slightly lower sampling rates in the 1250/50 system, the gross data rate, with 10 bit resolution, is then:
        - Y: 72 x 10 Mbit/s = 720 Mbit/s
        - CB: 0.5 x 72 x 10 Mbit/s = 360 Mbit/s
        - CR: 0.5 x 7,2 x 10 Mbit/s = 360 Mbit/s
            - In total: 1.44 Gbit/s → gross data rate (1250/50)

- The color resolution of this 4:2:2 signal is only reduced in the horizontal direction.
- The vertical color resolution corresponds to the full resolution resulting from the number of lines in a television frame.
- Human eye cannot distinguish between horizontal and vertical as far as color resolution is concerned.
    - It is possible, therefore, to also reduce the color resolution to one half in the vertical direction without perceptible effect.
    - MPEG applies irrelevance reduction by adopting (4:2:0) format associating four Y pixels to only one C_B value and one C_R value each resulting in further 25% data rate reduction
        - e.g. one line has C_B samples for every other pixel, and the next line has C_R samples for every other pixel.

Differential Pulse Code Modulation of Moving Pictures

- Further data reduction techniques should be applied to achieve greater savings in compression.
- Adjoining moving pictures differ only very slightly from each other. There are stationary areas which do not change from frame to frame hence subject to redundancy reduction techniques.
    - Processing the picture areas and identify the difference and tranfer only the delta value from one frame to the next.
    - This particular method of redundancy reduction using the reference value is called differential pulse code modultion (DPCM)

➢ What is DPCM?

- If a continuous analog signal is sampled and digitized, discrete values, i.e. values which are no longer continuous, are obtained at equidistant time intervals .
- These values can be represented as pulses spaced apart at equidistant intervals, which corresponds to a pulse code modulation.
- The height of each pulse carries information in discrete, non-continuous form about the current state of the sampled signal at precisely this point in time.
- In reality the difference between adjacent samples are not large because of the previous band limiting, however, if only the difference is transmitted, the data rate can be reduced.
    - Issue: in case of transmission error, it takes a long time till the demodulator can match the original signal and recover.
    - This can be addressed by sending regularly the complete samples then a few differences followed again by a complete sample (adopted by MPEG-1/2)

❖ This led to group of pictures (GoP) and dividing pictures to smaller sections of blocks and macroblocks for better processing of picture areas, as shown in Figure 6

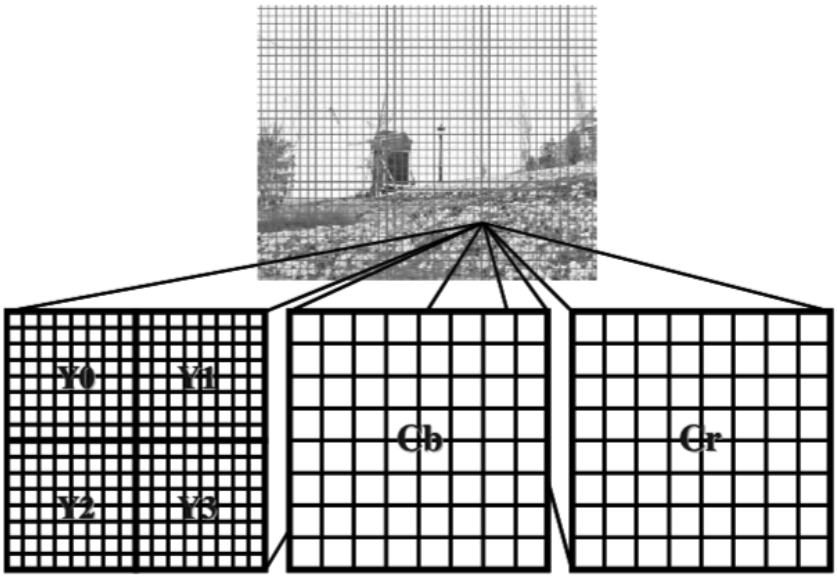
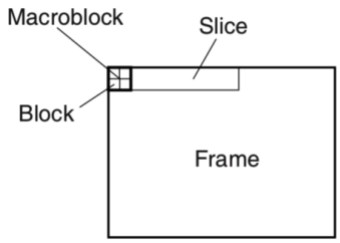
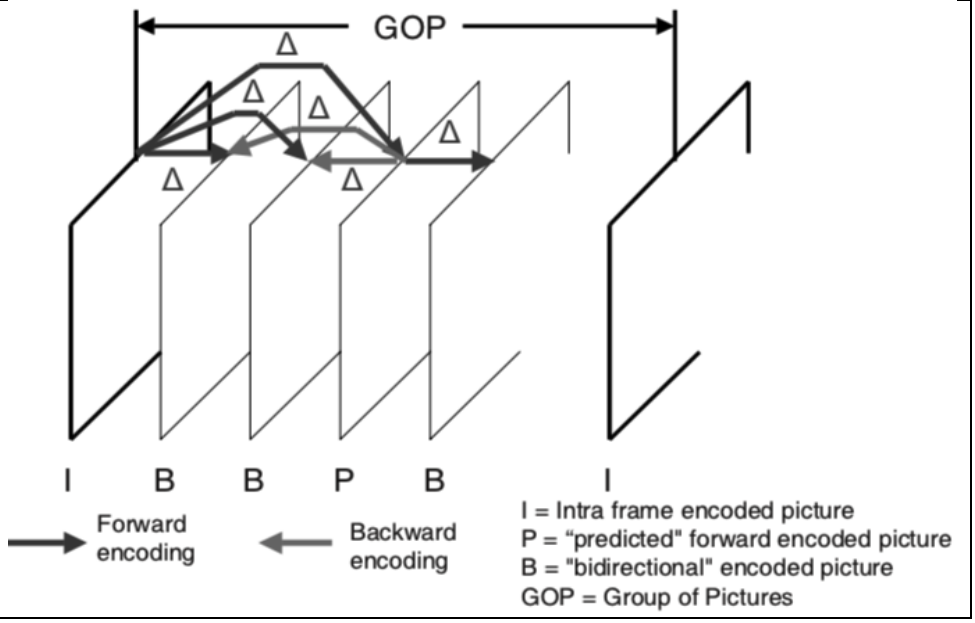


Figure 6: dividing a picture into blocks and macroblocks

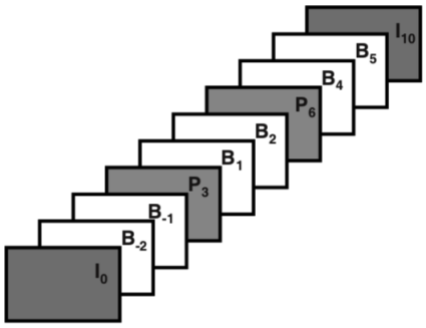
Structure of a Pictures

- The smallest unit of the video stream is a **block** consisting of 8 x 8 pixels
- In the case of a 4:2:0 profile, four luminance blocks and one CB block and one CR block in each case together form one **macroblock**.
    - Horizontal and vertical pixels are seleted to be divisable by 16 and also by 8 (Y: 720 x 576 pixels)
- A certain number of macroblocks in a row form a **slice**. Each slice starts with a header which is used for resynchronization, e.g. in the case of bit errors.
- Many slices together will then form a **frame** (picture). A frame, too, starts with a header, the picture header.
- There are different types of frames, called:
    - I (intraframe) frame
    - P (predicted) frame
    - B (bidirectionally predicted) frame
- A certain number of frames corresponding to a coding pattern of the I, P and B frame coding predetermined by the encoder, form a group of pictures (GOP). Each GOP has a GOP header.
    - In broadcasting, relatively short GOPs are used which, as a rule, have a length of about 12-15 frames, i.e about half a second.
    - The MPEG decoder can only lock to the signal and begin to reproduce pictures when it receives the start of a GOP, i.e. the first I-frame.
    - Longer GOPs can be chosen for mass storage devices such as the DVD since it is easy to position their read head on the first I frame.

Block, macroblock, slice and frame	Bidirectionally predicted delta frames
	

Frames and GOP

- The picture analysis is performed at the macroblocks.
- Based on the differences in corresponding macroblocks of the consequent frames, the frames are grouped into three encoding categories (I, P and B)
- At certain intervals, the complete reference frames are transmitted without forming the difference. These frames are called Intra-coded frames (I-frame) which are interspersed between them the delta frames (P and B frames – Interframes)
- The diffrence are analysed at the macro-block level forming P and B frames:
    - The respective macroblock of a following frame is always compared with the macroblock of the preceding frame
    - The difference could be caused by displacement then motion vector (shift in any direction) and difference with respect to the preceding macroblock is transmitted.
    - Otherwise, if there is not change found, nothing will be sent, and if a new macroblock is found (no correlation), the complete macroblock will be transmitted.
    - Apart from unidirectionally forward predicted frames there are also bidirectionally, i.e. forward and backward, predicted delta frames, so- called B frames.
        - Data rate is much lower in B frames (pictures) than I and P frames.
- The arrangement of frames occurring between two I-pictures, i.e. complete pictures, is called a group of pictures (GOP)
- A GOP consists of a particular number and a particular structure of B pictures and P pictures arranged between two I pictures.
    - A GOP usually has a length of about 12 frames (according to MPEG the length can be variable) and corresponds to the following order
        - {I, B, B, P, B, B, P, ..., I}
- For decoding B picture, information of preceding I and P pictures and that of the following I and P pictures are needed.
    - Because of the bidirectional differential coding, the order of the frames does not correspond to the original order and the headers and especially the PES headers, therefore, carry a time stamp so that the original order can be restored (DTS).
- GOP structure must be altered during the transmission so that the respective backward prediction information is already available before the actual B pictures.
    - For this reason, the frames are transmitted in an order which no longer corresponds to the original order as shown below (DTS in PES can be used to re-order the frames)



Motion Estimation Process

- Starting with a delta frame to be encoded
- The system looks in the preceding frame (forward prediction P) and possibly also in the subsequent frame (bidirectional prediction B) for suitable macroblock information in the environment of the macroblock to be encoded
    - This is done by using the principle of block matching within a certain search area around the macroblock.
- If a matching block is found in front, and also behind in the case of bi- directional coding, the motion vectors are determined forward and backward and transmitted.
- In addition, any additional block delta which may be necessary can also be transmitted, both forward and backward.
- However, the block delta is coded separately by DCT with quantization.

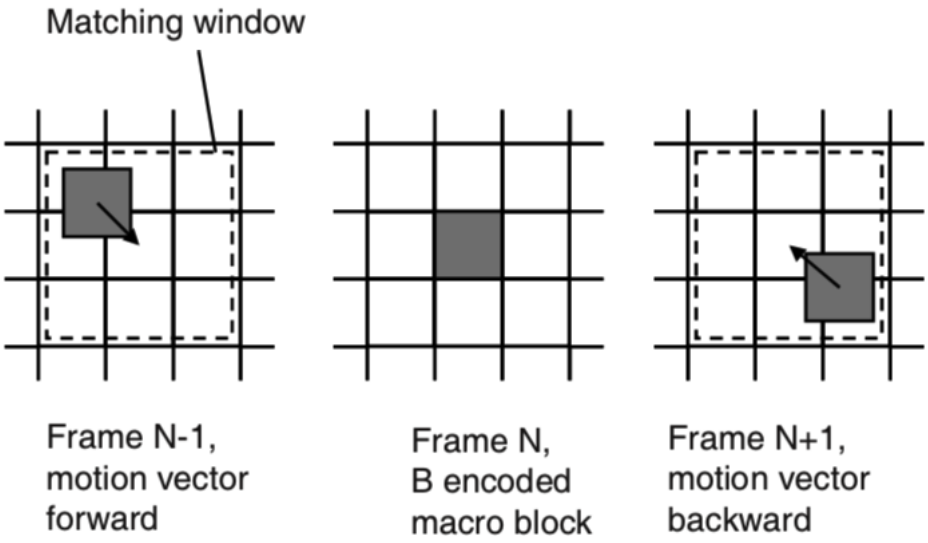
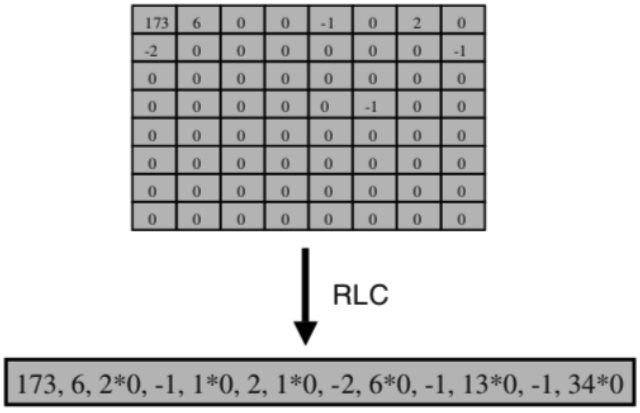


Figure 7: Motion vectors

Further Encoding Techniques

1. **Discrete Cosine Transform (DCT)** Followed by **Quantization**:
    - DCT is the basic and successful algorithm used in JPEG which also forms the central algorithm for MPEG video coding method
    - Human eye anatomy: fine structure vs coarse structures (low frequency disturbance percieved more than high frequency disturbance) therefore S/N is measured weighted based on eye sensitivity
        - Low-frequency, coarse image components are coded with finer quantization and fine image components are coded with coarser quantization in order to save data rate.

    ❖ How to separate coarse from fine components?
    - By means of transform coding
        - Transition from time domain to frequency domain
        - DCT is applied (8 samples in video line transformed to frequency domain corresponding to 8 power values in frequency domain.
        - Q (quantisation) is applied on DCT coefficients in frequency domain
            - Q-factor will be increased in direction of finer image structures
2. **Zig-Zag Scanning** with Run-Length Coding of Zero Sequences
    - Application of run-length coding (RLC) saves the transmission of large number of adjacent zeros resulted from zig-zag scanning of quantised DCT coefficients
    - This type of redundancy reduction, in conjunction with DCT and quantization, provides the main gain in the data compression.



3. **Huffman** coding (variable length coding)
    - Codes occuring frequently in RLC are also subject to Huffman coding resulting in further redundancy reduction
    - Most frequently used code words are encoded using shorter codes (Morse corde)

MPEG-2 Profiles and Levels

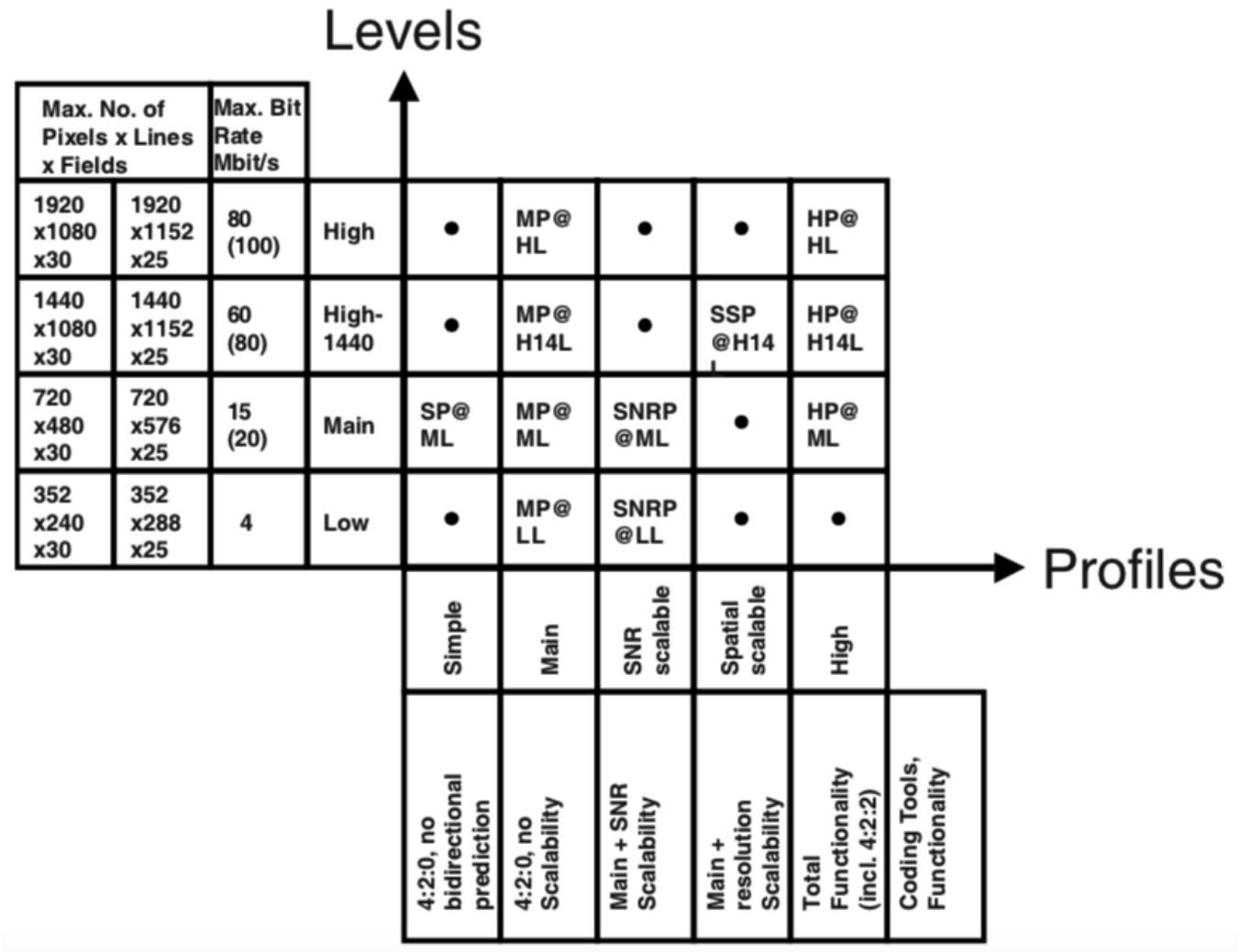


Figure 8: MPEG-2 profiles and levels (Main Profile@Main Level and Main Profile@High Level)

MPEG-2 Encoder Block Diagram

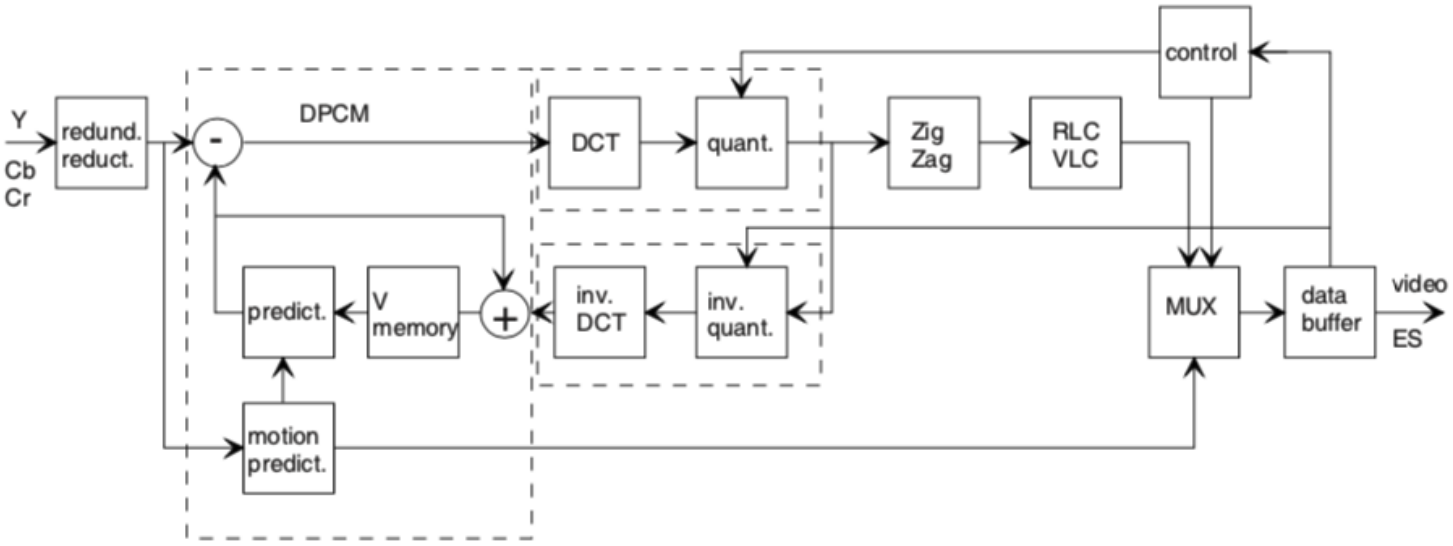


Figure 9: MPEG-2 encoder block diagram

- In contrast to analogue TV, in digital TV no fields are transmitted but only frames and fields are then recreated at the receiver from the frame buffer.
- There is a special DCT method which result in a better quality for interlaced scanning method where macroblocks are first recorded line by line before applying DCT.
    - There are two scanning methods to draw a picture on TV screen:
        - Progressive (all HD screens are progressive): entire picture is drawn at once resulting in reduction of flickering
        - Interlaced: half of the picture appears on the screen at a time followed by the other half in 1/60sec later.

Structure of MPEG-2 video ES:

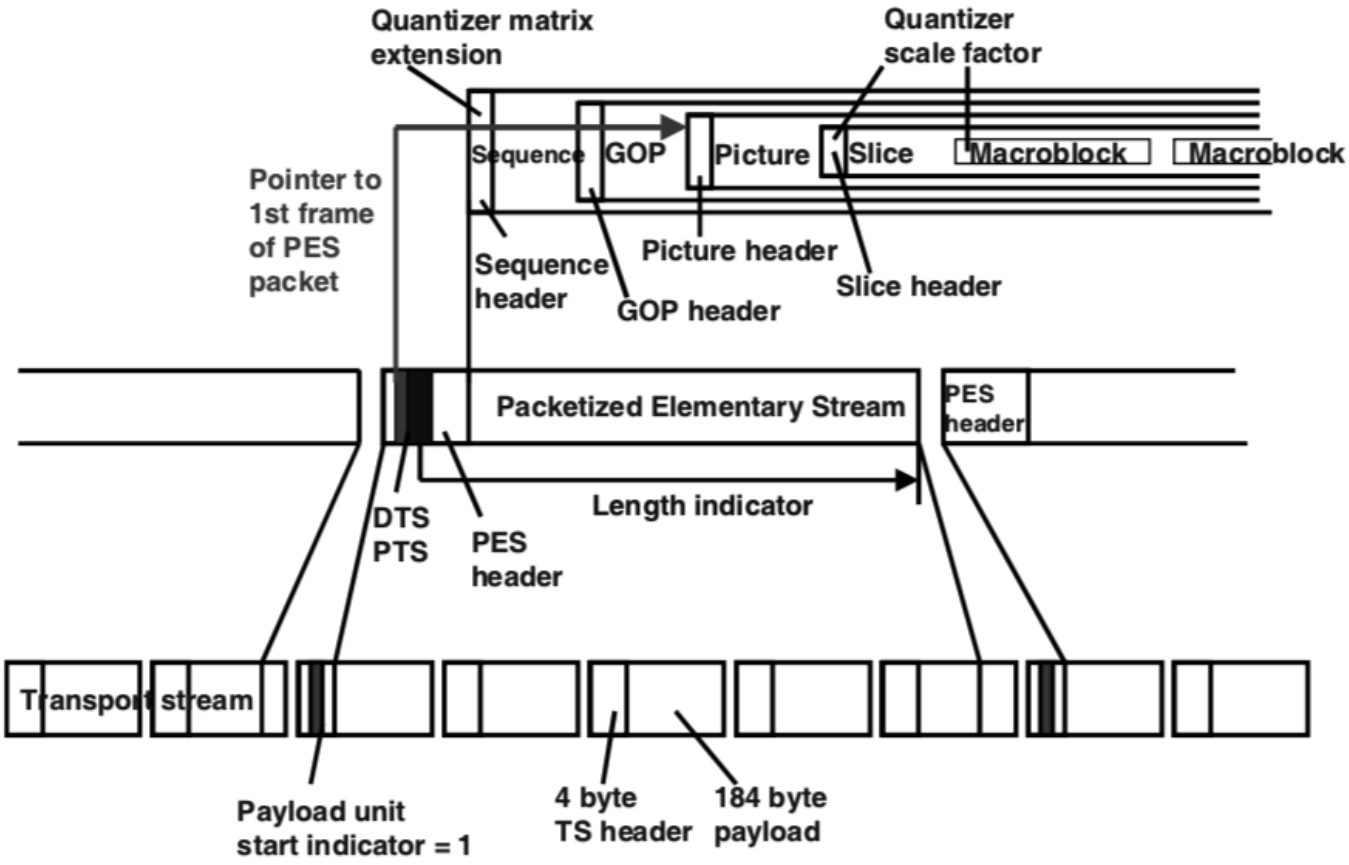


Figure 10: Structure of MPEG-2 video ES

History of developing video coding

- Basic principle of video coding has not changed but much more efficient video coding is now available such as MPEG-4 Part-10 Advanced Video Coding (AVC) (H.264) with data rates to be decreased by 30 to 50%
    - SDTV signal can now be compressed to approx. 1.5-3 Mbit/s compared with a data rate of 2-7 Mbit/s (comparing with original 270 Mbit/s)
- Smaller but variable transform block sizes are used
- DCT was replaced by a similar transform (integer transform) in MPEG-4 Part 10

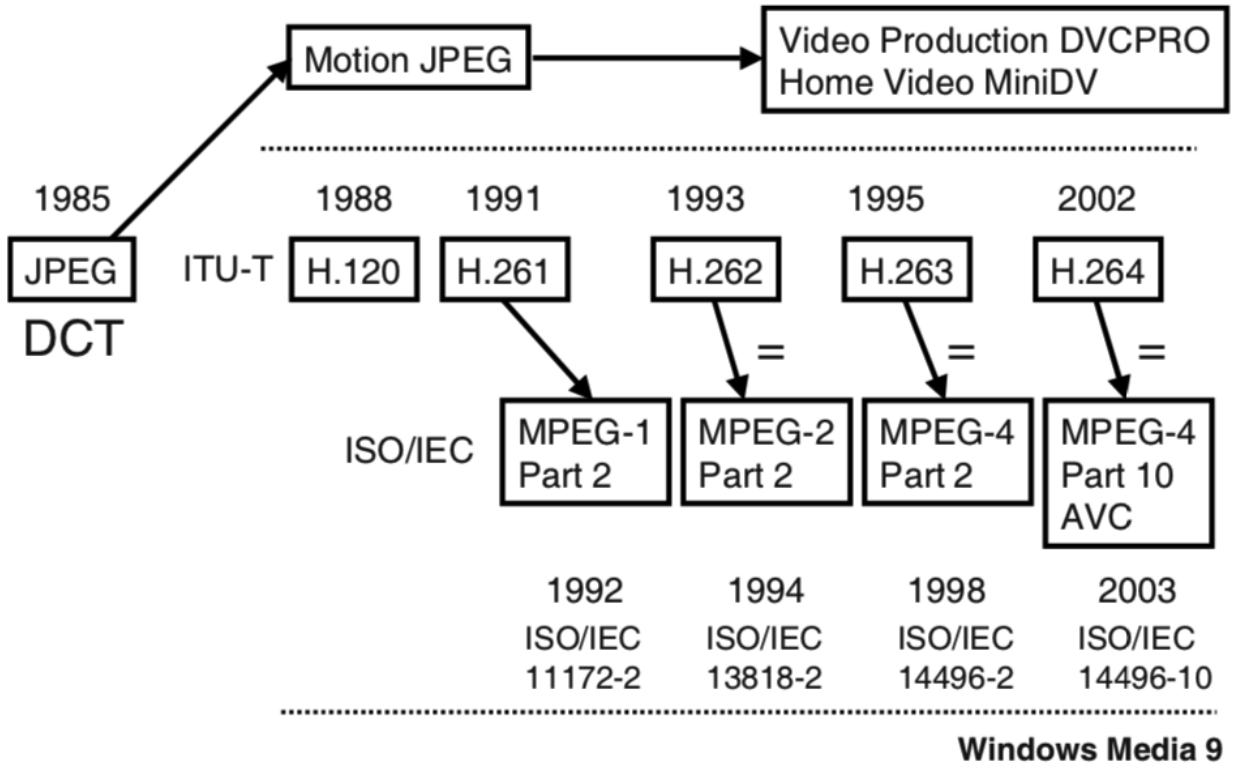


Figure 11: History of video encoding techniques

MPEG-4 Part 10 Main Features

- Formats 4:2:0, 4:2:2 and 4:4:4 are supported
- Up to 16 reference frames maximum
- Improved motion compensation (1/4 pixels accuracy)
- Switching P (SP) and Switching I (SI) frames
- Higher accuracy due to 16 bit implementation
- Flexible macroblock structure (16x16, 16x8, 8x16, 8x4, 4x8, 4x4)
- 52 selectable sets of quantization tables
- Integer or Hadamard transform instead of a DCT (block size 4x4 or 2x2 pixels, resp.)
- In-loop deblocking filter (eliminates blocking artefacts)
- Flexible slice structure (better bit error performance)
- Entropy encoding; variable length coding (VLC) and context
- adaptive binary arithmetic coding (CABAC)
- Figure 12 illustrates high level of video encoding block diagram in MPEG-4 highlighting de-**blocking filter**
    - MPEG-4 also uses a deblocking filter which is intended to additionally suppress the visibility of **blocking artefacts**. This is also aided by the smaller block size and the variable macroblock and slice size.

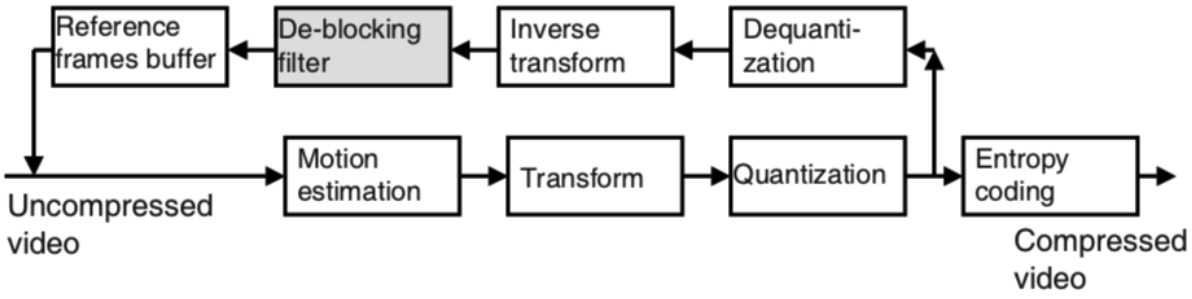


Figure 12: MPEG-4 video encoding block diagram