

MPEG - Motion Picture Expert Group

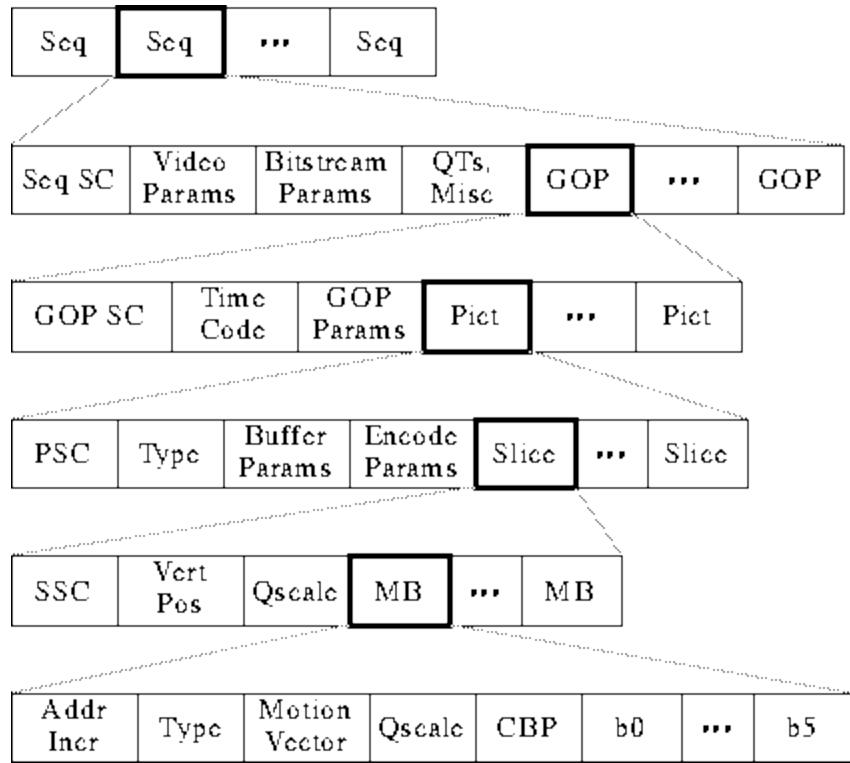
MPEG Standards

MPEG is an acronym for Moving Picture Experts Group, a committee formed by the ISO (International Organisation for Standardisation) to develop this standard. MPEG was formed in 1988 to establish an international standard for the coded representation of moving pictures and associated audio on digital storage media.

- MPEG-1
 - Medium Bandwidth (up to 1.5Mbits/sec)
 - 1.25Mbits/sec video 352 x 240 x 30Hz
 - 250Kbits/sec audio (two channels)
 - Non-interlaced video
 - Optimised for CD-ROM
- MPEG-2
 - Higher Bandwidth (up to 40Mbits/sec)
 - Up to 5 audio channels (i.e. surround sound)
 - Wider range of frame sizes (including HDTV)
 - Can deal with interlaced video
- MPEG-3
 - MPEG-3 was for HDTV application with dimensions up to 1920 x 1080 x 30Hz, however, it was discovered that the MPEG-2 and MPEG-2 syntax worked very well for HDTV rate video. Now HDTV is a part of MPEG-2 High-1440 Level and High Level toolkit.
- MPEG-4
 - Very Low Bandwidth (64Kbits/sec)
 - 176 x 144 x 10Hz
 - Optimised for videophones

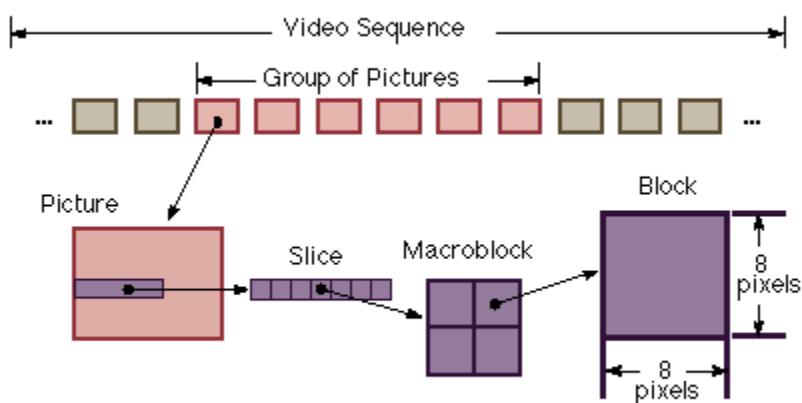
MPEG-2 Video

Video Stream Data Hierarchy



In the diagram above, we can see that the video bitstream consists of 5 layers:

- GOP,
- Pictures,
- Slice,
- Macroblock, and
- Block.



Video Sequence

Begins with a sequence header, includes one or more groups of pictures, and ends with an end-of-sequence code.

Group of Pictures (GOP)

A header and a series of one or more pictures intended to allow random access into the sequence.

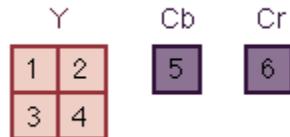
Picture

The primary coding unit of a video sequence. A picture consists of three rectangular matrices representing luminance (Y) and two chrominance (Cb and Cr) values. The Y matrix has an even number of rows and columns. The Cb and Cr matrices are one-half the size of the Y matrix in each direction (horizontal and vertical).

Slice

It consists one or more "contiguous" macroblocks. The order of the macroblocks within a slice is from left-to-right and top-to-bottom. Slice is important in the handling of errors. If the bitstream contains an error, the decoder can skip to the start of the next slice. Having more slices in the bitstream allows better error concealment, but uses bits that could otherwise be used to improve picture quality.

Macroblock



The basic coding unit in the MPEG algorithm. It is a 16 * 16 pixel segment in a frame. Since each chrominance component has one-half the vertical and horizontal resolution of the luminance component, a macroblock consists of four Y, one Cr, and one Cb block.

Block

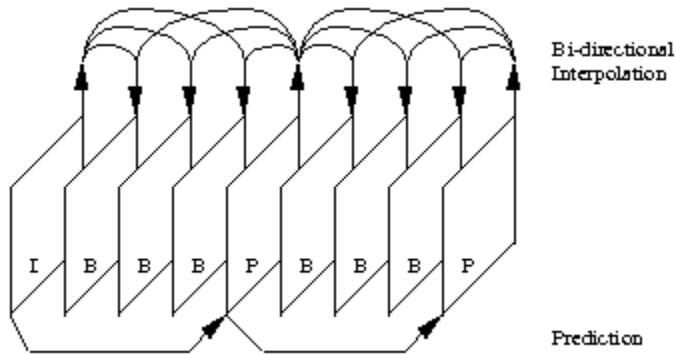
Block is the smallest coding unit in the MPEG algorithm. It consists of 8x8 pixels and can be one of three types: luminance(Y), red chrominance(Cr), or blue chrominance(Cb). The block is the basic unit in intra frame coding.

Picture Types

The MPEG standard specifically defines three types of pictures:

1. Intra Pictures (I-Pictures)
2. Predicted Pictures (P-Pictures)
3. Bidirectional Pictures (B-Pictures)

These three types of pictures are combined to form a group of picture.



Intra Pictures

- I-frames (Intra-coded Images) are self-contained without any references to other images. It is treated as a still image.
- MPEG uses JPEG for I-frames.
- This compression must be executed in real-time. Hence, the compression rate of I-frame is the lowest within MPEG.
- I-frames are points to random-access in MPEG stream.
- I-frames use 8 x 8 blocks defined within a macroblock. On these blocks, DCT is performed. Quantization is by constant value for all DCT coefficients. (i.e., no quantization table as in JPEG)

Intra pictures, or I-Pictures, are coded using only the information present in the picture itself, and provide potential random access points into the compressed video data. It uses only transform coding and provides moderate compression. Typically it uses about two bits per coded pixel.

Predicted Pictures

- Information of the previous I-frame and/or previous P-frames are required for encoding and decoding.
- The coding of P-frames utilizes successive images where their areas often do not change at all. The whole area is shifted (temporal redundancy).
- Temporal redundancy - It is required to determine the last P or I-frame that is most similar to the block under consideration.
- Motion Estimation Method is used at the encoder.

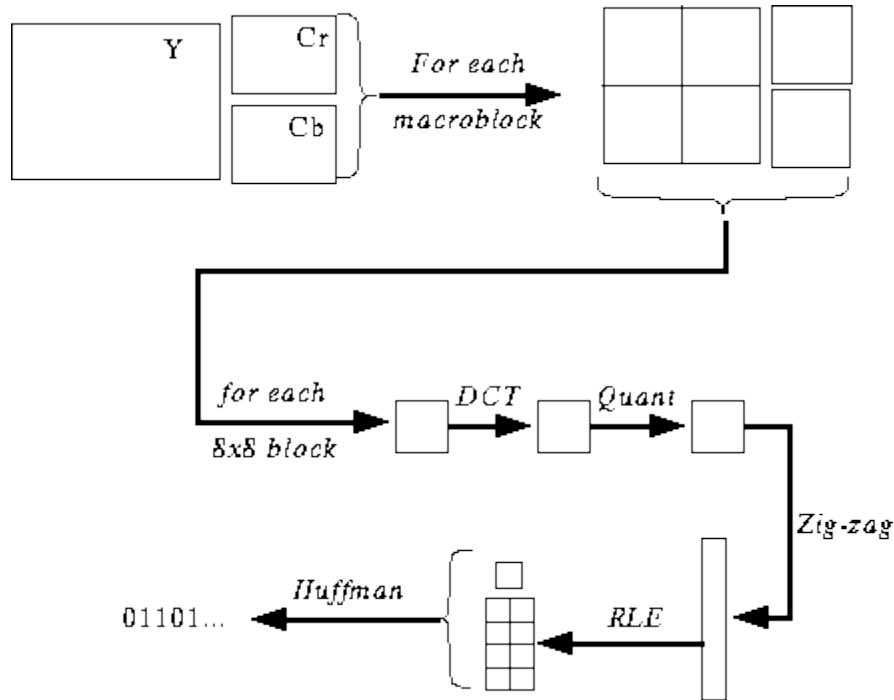
Predicted pictures, or P-pictures, are coded with respect to the nearest previous I- or P-pictures. This technique is called forward prediction and is illustrated in the above figure. Like I-pictures, P-pictures also can serve as a prediction reference for B-pictures and future P-pictures. Moreover, P-pictures use motion compensation to provide more compression than is possible with I-pictures.

Bidirectional Pictures

Bidirectional pictures, or B-pictures, are pictures that use both a past and future picture as a reference. This technique is called bidirectional prediction. B-pictures provide the most compression since it uses the past and future picture as a reference, however, the computation time is the largest.

Method of Encoding Pictures

Intra Pictures



The MPEG transform coding algorithm includes the following steps:

1. Discrete cosine transform (DCT)
2. Quantization
3. Run-length encoding

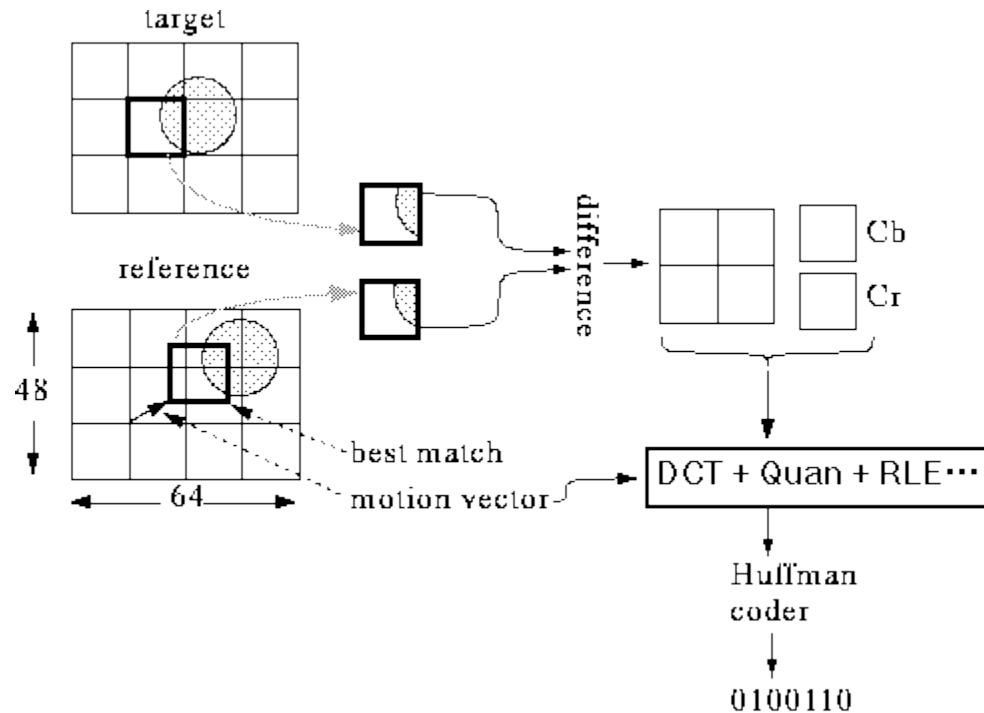
Both image blocks and prediction-error blocks have high spatial redundancy. To reduce this redundancy, the MPEG algorithm transforms 8x8 blocks of pixels or 8x8 blocks of error terms from the spatial domain to the frequency domain with the *Discrete Cosine Transform (DCT)*.

The combination of DCT and *quantization* results in many of the frequency coefficients being zeroed, especially the coefficients for high spatial frequencies. To take maximum advantage of this, the coefficients are organised in a *zigzag* order to produce long runs of zeros. The coefficients are then converted to a series of *run-amplitude pairs* each pair indicating a number of zero coefficients and the amplitude of a non-zero coefficient. These run-amplitude pairs are then coded with a *variable-length code (Huffman Encoding)*, which uses shorter codes for commonly occurring pairs and longer codes for less common pairs.

Some blocks of pixels need to be coded more accurately than others. For example, blocks with smooth intensity gradients need accurate coding to avoid visible block boundaries. To deal with this inequality between blocks, the MPEG algorithm allows

the amount of quantization to be modified for each macroblock of pixels. This mechanism can also be used to provide smooth adaptation to particular bit rate.

Predicted Pictures



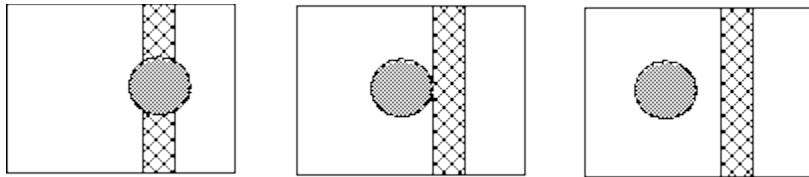
A P-picture is coded with reference to a previous image (reference image) which is an I- or P-Picture. From the above figure, the highlighted block in the target image (the image to be coded) is similar to the reference image except that it is shifted to the upper right. Most of the changes between the target and the reference image can be approximated as a translation of a small image region. Therefore a key technique called motion compensation prediction is used.

Motion compensation based prediction exploits the temporal redundancy of the image. Due to frames being closely related, it is possible to accurately represent or "predict" the data of one frame based on the data of a reference image, provided the translation is estimated. This process of prediction results in the number of bits being substantially reduced. In P-Pictures, each 16*16 sized macroblock is predicted from a macroblock of a previously encoded I picture. Since, frames are snapshots in time of a moving object, the macroblocks in the two frames may not be co-sited, i.e. correspond to the same spatial location. Hence, a search is conducted in the I-frame to find the

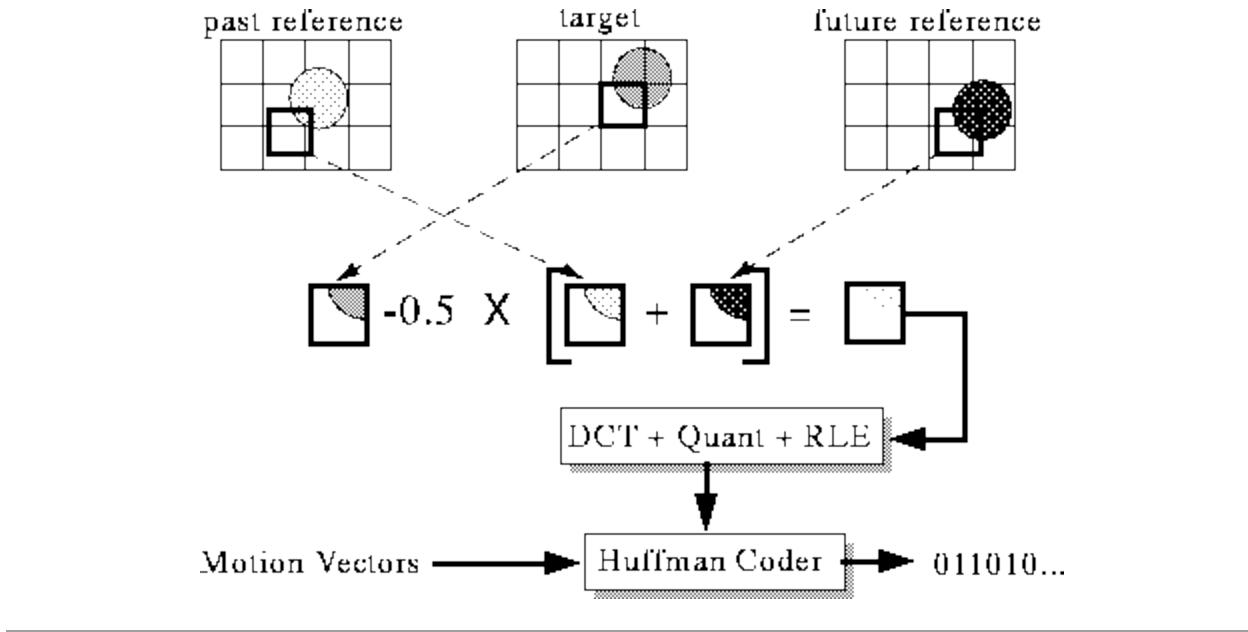
macroblock which closely matches the macroblock under consideration in the P frame. The difference between the two macroblocks is the prediction error. This error can be coded in the DCT domain. The DCT of the error results in few high frequency coefficients, which after the quantization process require a small number of bits for representation. The quantization matrices for the prediction error blocks are different from those used in an intra block, due to the distinct nature of their frequency spectrum. The displacements in the horizontal and vertical directions of the best matching macroblock from the cosited macroblock are called motion vectors. Differential coding is used because it reduces the total bit requirement by transmitting the difference between the motion vectors of consecutive frames. Finally run-length and Huffman encoding are used to encode the data.

Bidirectional Pictures

Example:



In the above pictures, there is some information which is not in the reference frame. Therefore B pictures are coded like P-pictures except that the motion vectors can refer either to the previous reference picture, the next picture, or both. The following is the mechanism of B-picture coding.



Profiles and Levels

MPEG-2 is designed to support a wide range of applications and services of varying bit rate, resolution, and quality. The MPEG-2 standard defines 4 profiles and 4 levels for ensuring inter-operability of these applications.

The profile defines the

- colorspace resolution, and
- scalability of the bitstream.

The levels define the

- maximum and minimum for image resolution,
- Y (Luminance) samples per second,
- the number of video and audio layers supported for scalable profiles, and
- the maximum bit rate per profile.

Upon decoding, the video decoder will use the profiles and levels defined depending on its availability and the need to handle the particular bitstream.

MPEG-2 Levels

Level	Max. Sampling dimensions fps	Pixels/sec	Max. bitrate	Significance
Low	352 x 240 x 30	3.05 M	4 Mb/s	CIF, consumer tape equiv.
Main	720 x 480 x 30	1 □ 0.40 M	15 Mb/s	CCIR 601, studio TV
High 1440	1440 x 1152 x 30	47.00 M	60 Mb/s	4x 601, consumer HDTV
High	1920 x 1080 x 30	62.70 M	80 Mb/s	production SMPTE 240 std

MPEG-2 Profiles

Profile	Comments
Simple	Same as Main, only without B-pictures. Intended for software applications, perhaps CATV.
Main	Most decoder chips, CATV satellite. 95% of users.
Main+	Main with Spatial and SNR scalability.
Next	Main+ with 4:2:2 macroblocks.

Relationship among profiles and levels

Level \ Profile	Simple	Main	Main+	Next
High	illegal	/	illegal	4:2:2 chroma
High - 1440	illegal	/	With spatial Scalability	4:2:2 chroma

Main	/	90% of users	Main with SNR scalability	4:2:2 chroma
Low	illegal	/	Main with SNR scalability	illegal

Interlaced Video and Picture Structures

MPEG-2 supports two scanning methods, one is progressive scanning and the other is interlaced scanning. Interlaced scanning scans odd lines of a frame as one field (odd field), and even lines as another field (even field). Progressive scanning scans the consecutive lines in sequential order.

An interlaced video sequence uses one of two picture structures: frame structure and field structure. In the frame structure, lines of two fields alternate and the two fields are coded together as a frame. One picture header is used for two fields. In the field structure, the two fields of a frame may be coded independently of each other, and the odd field is followed by the even field. Each of the two fields has its own picture header.

The interlaced video sequence can switch between frame structure and field structures on a picture-by-pictures basics. On the other hand, each picture in a progressive video sequence is a frame picture.

MPEG-2 Audio

MPEG-2 provides a low bit rate coding for multichannel audio. There are five full bandwidth channels (left, right, center, and two surround channels), plus an additional low frequency enhancement channel, and/or up to seven commentary/multilingual channels. The MPEG-2 Audio Standard will also extend the stereo and mono coding of MPEG-1 Audio Standard (ISO/IEC IS 11172-3) to half sampling rates (16 kHz, 22.05 kHz and 24 kHz), for improved quality for bit rate at or below 64 kbits/s, per channel
