MPEG

- MPEG video is broken up into a hierarchy of layer
- From the top level, the first layer is known as the video sequence layer, and is any self contained bitstream, for example a coded movie.
- The second layer down is the group of pictures, which is composed of 1 or more groups of intra (I) frames and/or non-intra (P and/or B).
- The third layer down is the picture layer itself, and the next layer beneath it is called the slice layer.
- Each slice is a contiguous sequence macroblocks, which are 16x16 arrays of luminance pixels, or picture data elements, with 2 8x8 arrays of associated chrominance pixels.
- The macroblocks can be further divided into distinct 8x8 blocks, for further processing such as transform coding.

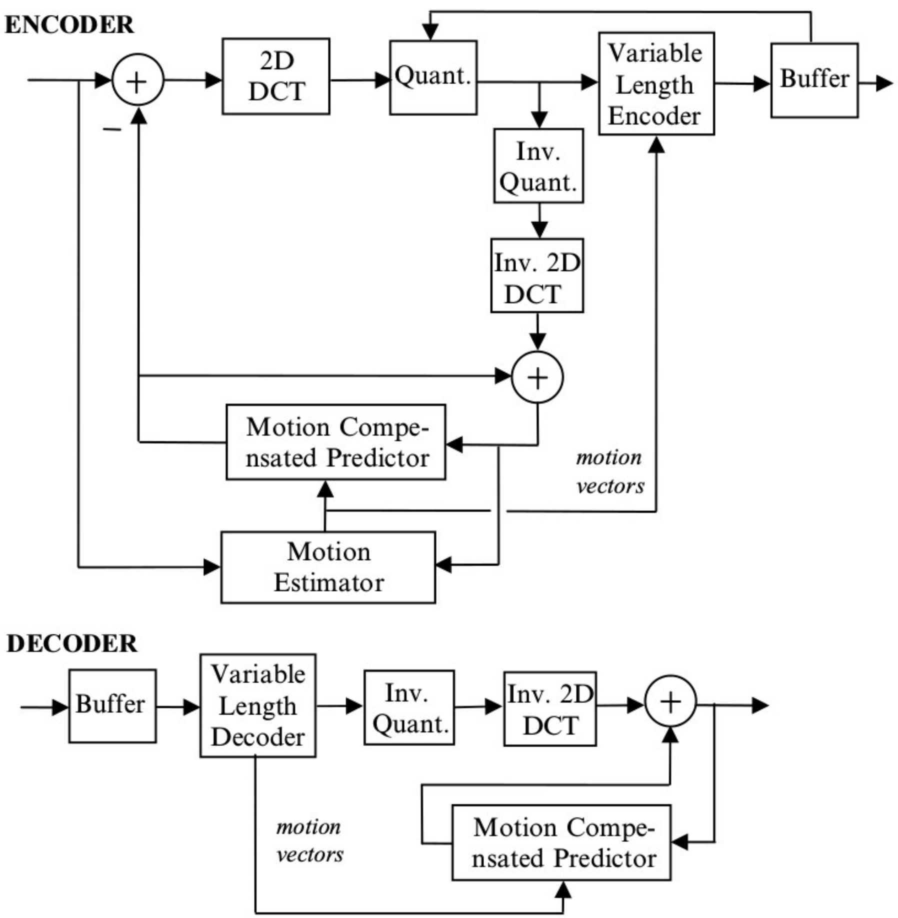
Interframe coding

- A video scene captured as a sequence of frames can be efficiently coded by estimating and compensating for motion between frames prior to generating inter-frame difference signal for coding.
- For the ease of processing, each frame of video is uniformly partitioned into smaller units called Macroblocks where each macroblock consists of a 16 × 16 block of luma, and corresponding chroma blocks.
- Each block of pixels (say 16 × 16 luma block of a MB) in the current frame is compared with a set of candidate blocks of same size in the previous frame to determine the one that best predicts the current block.
- The set of blocks includes those within a search region in previous frame centered on the position of current block in the current frame.

Interframe coding

- When the best matching block is found, a motion vector is determined, which specifies the reference block
- The key idea is to combine transform coding (in the form of the Discrete Cosine Transform (DCT) of 8 × 8 pixel blocks) with predictive coding (in the form of differential Pulse Code Modulation (PCM)) in order to reduce storage and computation of the compressed image

Interframe coding



- Since motion compensation is difficult to perform in the transform domain, the first step in the interframe coder is to create a motion compensated prediction error in the pixel domain
- For each block of current frame, a prediction block in the reference frame is found using motion vector found during motion estimation, and differenced to generate prediction error signal.
- The resulting error signal is

  transformed using 2D DCT, quantized by an adaptive quantizer, entropy encoded using a Variable Length Coder (VLC) and buffered for transmission

MPEG1

- The MPEG-1 standard is the first true multimedia standard with specifications for coding, compression, and transmission of audio, video, and data streams in a series of synchronized, mixed Packets
- The driving focus of the standard was storage of
- multimedia content on a standard CDROM, which supported data transfer rates of 1.4 Mb/s and a total storage capability of about 600 MB
- MPEG-1 was intended to provide VHS VCR-like video and audio quality, along with VCR-like controls.
- MPEG-1 is formally called ISO/IEC 11172.

MPEG1/H.261

- MPEG1 employes key techniques used by the H.261 standard
- In H.261, only two picture formats, common intermediate format (CIF) and quarter-CIF
- (QCIF), are allowed.
- CIF pictures are made of three components: luminance Y and color differences Cb and Cr.
- The CIF picture size for Y is 352 pels 4 per line by 288 lines per frame.
- The two color signals are subsampled to 176 pels per line and 144 lines per frame.
- The image aspect ratio is 4(horizontal):3(vertical), and the picture rate is 29.97 frames per second.
- A picture frame is partitioned into 8 line × 8 pel image blocks.
- A Macroblock (MB) is defined as four 8 × 8 (or one 16 × 16) Y block/s, one Cb
- block, and one Cr block at the same location.
- The compressed H.261 video bit stream contains several layers.
- They are picture layer, group of blocks (GOB) layer, Macroblock (MB) layer, and block layer.
- The higher layer consists of its own header followed by a number of lower layers.

MPEG1/H.261

- Picture Layer: In a compressed video bit stream, we start with the picture layer. Its header contains:
  - Picture start code (PSC) a 20-bit pattern.
  - Temporal reference (TR) a 5-bit input frame number.
  - Type information (PTYPE) such as CIF/QCIF selection.
  - Spare bits to be defined in later versions.

MPEG1/H.261

- GOB Layer: At the GOB layer, a GOB header contains:
  - Group of blocks start code (GBSC) a 16-bit pattern
  - Group number (GN) a 4-bit GOB address
  - Quantizer information (GQUANT)
  -

MPEG1/H.261

- Macroblock (MB) Layer: At the MB layer, the header contains:
  - Macroblock address (MBA) location of this MB relative to the previously coded MB inside the GOB
  - Type information (MTYPE) 10 types in total.
  - Quantizer (MQUANT) information
  - Motion vector data (MVD) differential displacement vector
  - Coded block pattern (CBP) indicates which blocks in the MB are coded

MPEG1/H.261

- Block layer
  - The lowest layer is the block layer, consisting of quantized transform coefficients (TCOEFF), followed by the end of block (EOB) symbol. All coded blocks have the EOB symbol.
  - There are essentially four types of coded MBs as indicated by MTYPE:
    - Intra – original pels are transform-coded.
    - Inter – frame difference pels (with zero-motion vectors) are coded.
    - Inter_MC – displaced (nonzero-motion vectors) frame differences are coded
    - Inter_MC_with_filter – the displaced blocks are filtered by a predefined loop filter, which may help reduce visible coding artifacts at very low bit rates.

MPEG1/H.261

- H.261 coding semantics (brief outline)
  - A single-motion vector (horizontal and vertical displacement) is transmitted for one Inter_MC MB. That is, the four Y blocks, one Cb, and one Cr block all share the same motion vector.
  - The range of motion vectors is +− 15 Y pels with integer values.
  - The transform coefficients of either the original (Intra) or the differential (Inter) pels are ordered according to a zigzag scanning pattern. These transform coefficients are selected and quantized at the encoder, and then coded using variable-length code- words (VLCs) and/or fixed-length codewords (FLC), depending on the values. Just as with JPEG, successive zeros between two nonzero coefficients are counted and called a RUN.
  - The value of a transmitted nonzero quantized coefficient is called a LEVEL. The most likely occurring combinations of (RUN, LEVEL) are encoded with a VLC, with the sign bit terminating the RUN-LEVEL VLC codeword.
-

MPEG1

- Video coding as per MPEG-1 uses coding concepts similar to H.261, namely spatial coding by taking the DCT of 8 × 8 pixel blocks, quantizing the DCT coefficients, storing the DCT coefficients for each block in a zigzag scan, and doing a variable run length coding of the resulting DCT coefficient stream.
- Temporal coding is achieved by using the ideas of uni- and bi- directional motion compensated prediction, with three types of pictures resulting:
  - I or Intra pictures which were coded independently of all previous or future pictures
  - P or Predictive pictures which were coded based on previous I or previous P pictures
  - B or Bi-directionally predictive pictures which were coded based on either the next and/or the previous pictures

MPEG1

- Because video is a sequence of still images, it is possible to achieve some compression using techniques similar to JPEG
- Such methods of compression are called
- intraframe coding techniques, where each picture of video is individually and independently compressed or encoded.
- Intraframe coding exploits the spatial redundancy that exists between adjacent pels of a picture.
- Pictures coded using only intraframe coding are called I-pictures.

MPEG1

- Because video is a sequence of still images, it is possible to achieve some compression using techniques similar to JPEG
- Such methods of compression are called
- intraframe coding techniques, where each picture of video is individually and independently compressed or encoded.
- Intraframe coding exploits the spatial redundancy that exists between adjacent pels of a picture.
- Pictures coded using only intraframe coding are called I-pictures.

MPEG1

- As in JPEG and H.261, the MPEG-1 video-coding algorithm employs a block based two-dimensional DCT
- A picture is first divided into 8 × 8 blocks of pels, and the two-dimensional DCT is then applied independently on each block
- This operation results in an 8 × 8 block of DCT coefficients in which most of the energy in the original (pel) block is typically concentrated in a few low-frequency coefficients
- A quantizer is applied to the DCT coefficients, which sets many of them to zero. This quantization is responsible for the lossy nature of the compression algorithms in JPEG, H.261 and MPEG-1 video. Compression is achieved by transmitting only the coefficients that survive the quantization operation and by entropy-coding their locations and amplitudes.

MPEG1

- Exploiting temporal redundancy
- Temporal redundancy results from a high degree of correlation between adjacent pictures. The MPEG-1 algorithm exploits this redundancy by computing an interframe difference signal called the prediction error. In computing the prediction error, the technique of motion compensation is employed to correct for motion.
- A Macroblock (MB) approach is adopted for motion compensation.

MPEG1

- In unidirectional or Forward Prediction, 16 × 16 luma block of each macroblock in the current picture to be coded is matched with a block of the same size in a previous picture called the Reference picture.
- As in H.261 blocks of the Reference picture that “best match” the 16 × 16 luma blocks of current picture, are called the Prediction blocks.
- The prediction error is then computed as the difference between the Target block and the Prediction block.
- The position of this best-matching Prediction block is indicated by a motion vector that describes the displacement between it and the Target block.
- Unlike H.261 where each motion vector is specified at “integer pel” accuracy, in MPEG-1 each motion vector is specified at “half-pel” accuracy, thus allowing improved prediction.
- The motion vector information is also encoded and transmitted along with the prediction error. Pictures coded using Forward Prediction are called P-pictures.

MPEG1

- Bidirectional temporal prediction
- Bidirectional temporal prediction, also called Motion-Compensated Interpolation, is a key feature of MPEG-1 video.
- Pictures coded with Bidirectional prediction use wo Reference pictures, one in the past and one in the future.
- A Target 16 × 16 luma block in bidirectionally coded pictures can be predicted by a 16 × 16 block from the past Reference picture (Forward Prediction), or one from the future Reference picture (Backward Prediction), or by an average of two 16 × 16 luma blocks, one from each Reference picture (Interpolation).
- In every case, a Prediction 16 × 16 block from a Reference picture is associated with a motion vector.
- Pictures coded using Bidirectional Prediction are called B-pictures.
- Pictures that are Bidirectionally predicted are never themselves used as Reference pictures, i.e., Reference pictures for B-pictures must be either P-pictures or I-pictures.
- Similarly, Reference pictures for P-pictures must also be either P-pictures or I-pictures.
- Bidirectional prediction provides a number of advantages.
  - The primary one is that the compression obtained is typically higher than can be obtained from Forward (unidirectional) prediction alone.
- However, Bidirectional prediction does introduce extra delay in the encoding process, because pictures must be encoded out of sequence.
- Further, it entails extra encoding complexity because block matching (the most computationally intensive encoding procedure) has to be performed twice for each Target block, once with the past Reference picture and once with the future Reference picture