Multimedia

Advanced Video Coding Standards

Topics

- Optimizing Video Coding
  - Motion estimation with limited search window
  - Logarithmic motion estimation
  - Hierarchical search
- Support for interlacing
- Scalability
- H.264

Using Temporal Redundancy (Recap)

- To increase the compression rate in video we can encode the difference between frames instead of the frame itself.
- To find the difference between frames, we subtract them from each other.
- Problem:
  - There are some objects in the frame which change their position. This results in large difference values.

Motion Compensation

- Each image is divided into macroblocks of size N x N.
- By default, N = 16 for luminance images. For chrominance images, N = 8 if 4:2:0 chroma subsampling is adopted.
- Motion compensation is performed at the macroblock level.
- The current image frame is referred to as Target Frame.
- A match is sought between the macroblock in the Target Frame and the most similar area in previous and/or future frame(s) (referred to as Reference frame(s)).
- The displacement of the reference area to the target macroblock is called a motion vector MV.

Optimizing Motion Vector Search

- MV search is usually limited to a small immediate neighborhood
- The neighborhood is limited in both horizontal and vertical displacements to the range [−p, p].
- This makes a search window of size $(2p + 1) \times (2p + 1)$.

Search for Motion Vectors

- The difference between two macroblocks can then be measured by their *Mean Absolute Difference (MAD)*:

$$MAD(i,j)=\frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\left|C(x+k,y+l)-R(x+i+k,y+j+l)\right|$$

- The goal of the search is to find a vector (i, j) as the motion vector MV = (u, v), such that MAD(i, j) is minimum:

$$(u,v)=\left[\ (i,j)\,|\,MAD(i,j)\ is\ minimum,\ i\in[-p,p],j\in[-p,p]\right]$$

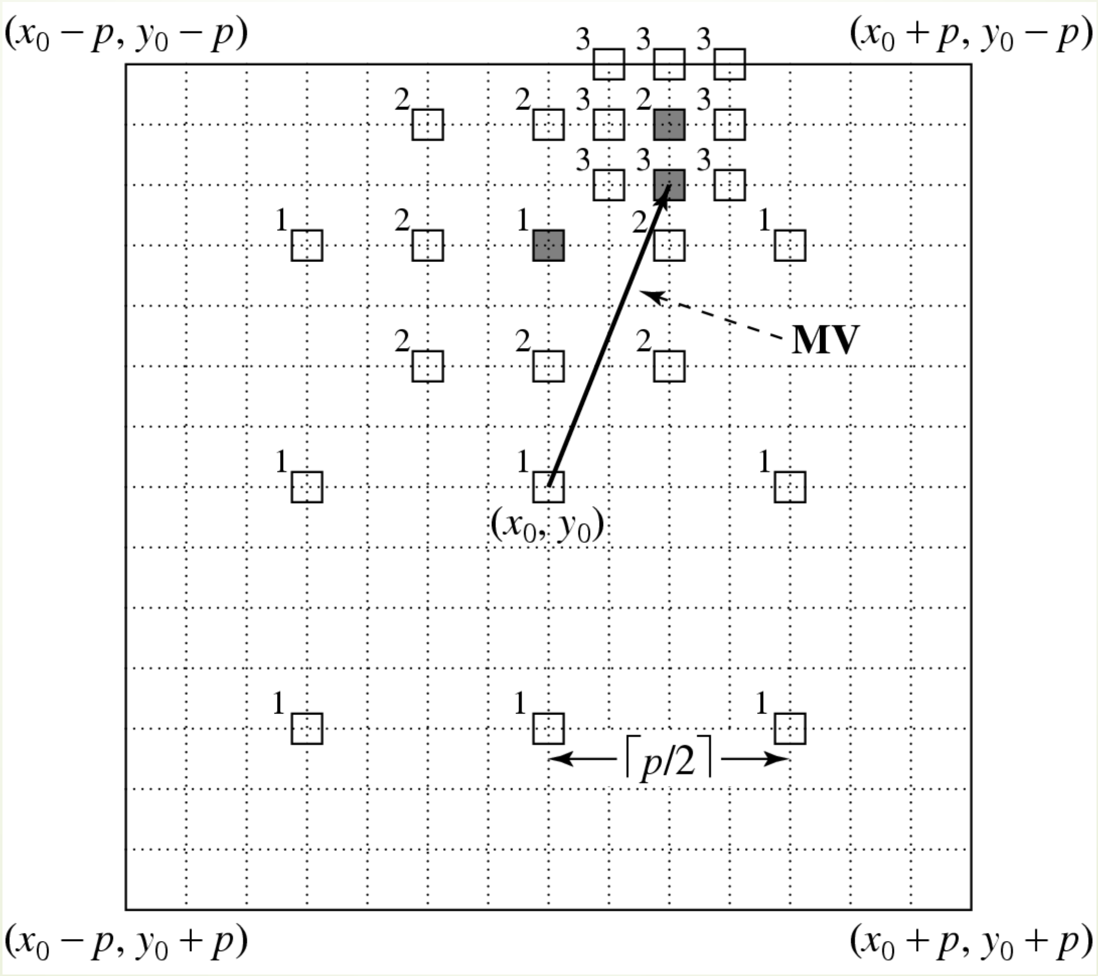
Sequential Search

- Sequentially search the whole (2p + 1) x (2p + 1) window in the Reference frame (also referred to as Full search).
  - A macroblock centered at each of the positions within the window is compared to the macroblock in the Target frame pixel by pixel and their respective MAD is then derived
  - The vector (i, j) that offers the least MAD is designated as the MV (u, v) for the macroblock in the Target frame.
  - Sequential search method is very time consuming.

2D Logarithmic Search

- Logarithmic search is a more efficient version, that is suboptimal but still usually effective.
- The procedure for 2D Logarithmic Search of motion vectors takes several iterations and is similar to a binary search.
  - Initially only nine locations in the search window are used as seeds for a MAD -based search.
  - After the one that yields the minimum MAD is located, the center of the new search region is moved to it and the step-size ("offset") is reduced to half.
  - In the next iteration, MAD is calculated for the nine new locations, and so on.

2D Logarithmic Search for Motion Vectors

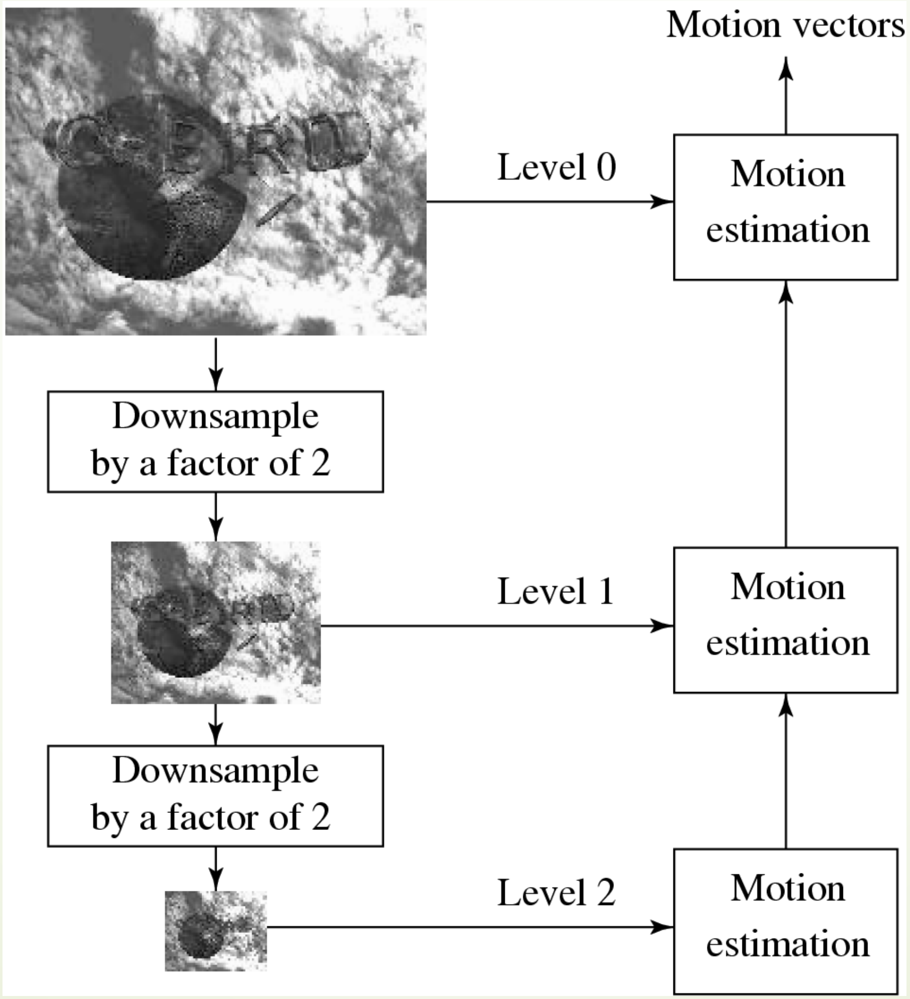


Hierarchical Search

- The search benefits from a hierarchical (multiresolution) approach in which initial estimation of the motion vector can be obtained from images with a significantly reduced resolution.

- Since the size of the macroblock is smaller and p can also be proportionally reduced, the number of operations required is greatly reduced.

A Three-level Hierarchical Search for Motion Vectors.



Supporting Interlaced Video

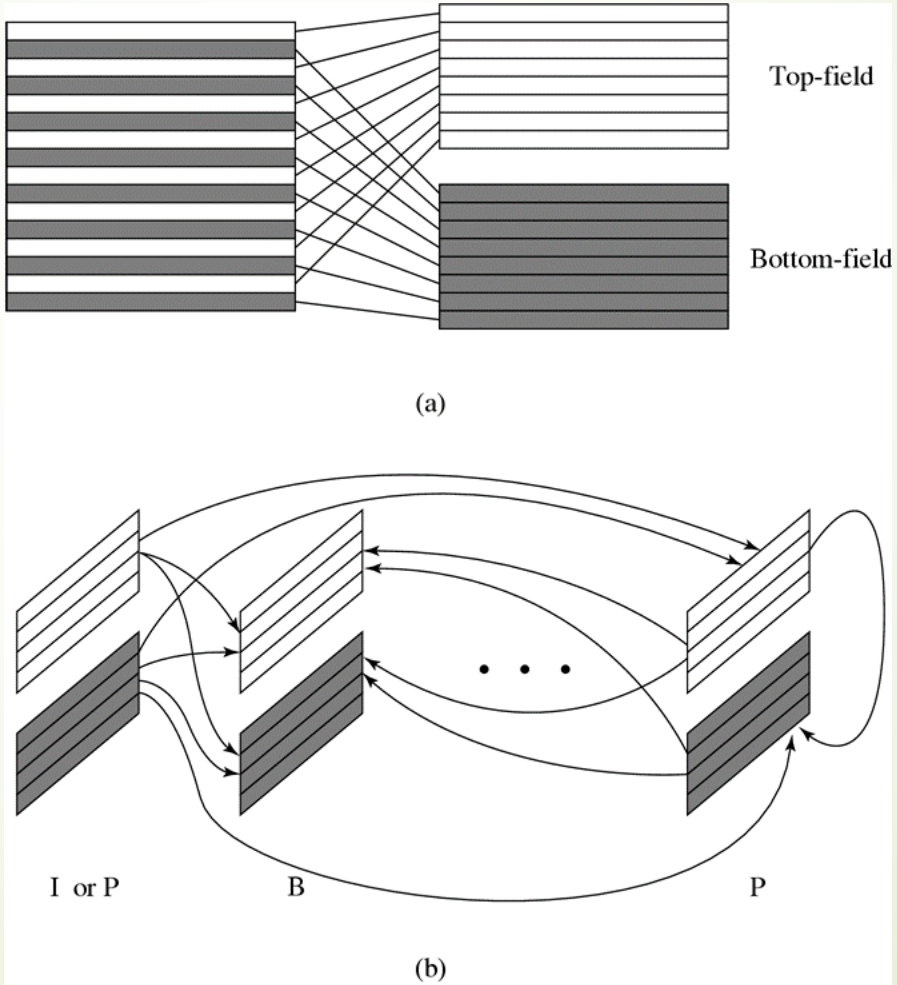
- MPEG must support interlaced video as well since this is one of the options for digital broadcast TV and HDTV.

- In interlaced video each frame consists of two fields, referred to as the top-field and the bottom-field.

- In a **Frame-picture**, all scanlines from both fields are interleaved to form a single frame, then divided into 16×16 macroblocks and coded using MC.

- If each field is treated as a separate picture, then it is called **Field-picture**.

Frame Picture vs. Field Pictures
Field Prediction vs. Field Prediction



Modes of Predictions

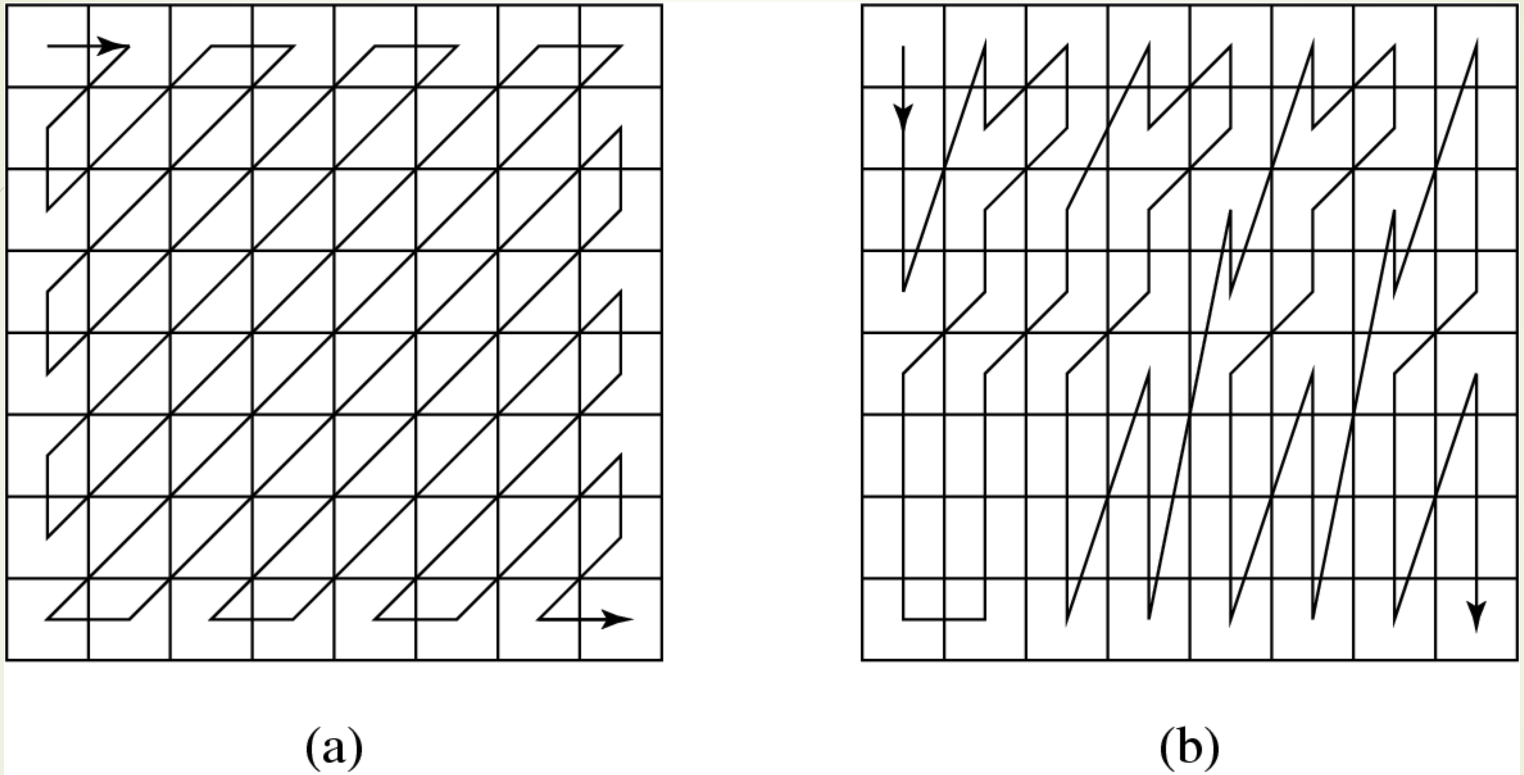
- MPEG-2 defines Frame Prediction and Field Prediction :

  - Frame Prediction for Frame-pictures: Identical to MPEG-1 MC-based prediction methods in both P-frames and B-frames.

  - Field Prediction for Field-pictures: A macroblock size of 16 × 16 from Field-pictures is used.

Alternate Scan and Field DCT

- Techniques aimed at DCT are only applicable to Frame-pictures in interlaced videos:
  - Due to the nature of interlaced video the consecutive rows in the 8×8 blocks are from different fields,
  - As a result, there exists less correlation between rows than between the alternate columns.
  - Alternate scan recognizes the fact that in interlaced video the vertically higher spatial frequency components may have larger magnitudes and thus allows them to be scanned earlier in the sequence.



- Zigzag and Alternate Scans of DCT Coefficients for Progressive and Interlaced Videos in MPEG-2.

Why Scalability?

- Scalable coding is especially useful for MPEG-2 video transmitted over networks with following characteristics:
  - Networks with very different bit-rates.
  - Networks with variable bit rate (VBR) channels.
  - Networks with noisy connections.

MPEG-2 Scalabilities

- A base layer and one or more enhancement layers can be defined — also known as layered coding.
  - The base layer can be independently encoded, transmitted and decoded to obtain basic video quality.
  - The encoding and decoding of the enhancement layer is dependent on the base layer or the previous enhancement layer.

MPEG-2 Scalabilities (Cont’d)

- MPEG-2 supports the following scalabilities:
  - SNR Scalability—enhancement layer provides higher SNR.
  - Spatial Scalability — enhancement layer provides higher spatial resolution.
  - Temporal Scalability—enhancement layer facilitates higher frame rate.
  - Hybrid Scalability — combination of any two of the above three scalabilities.
  - Data Partitioning — quantized DCT coefficients are split into partitions.

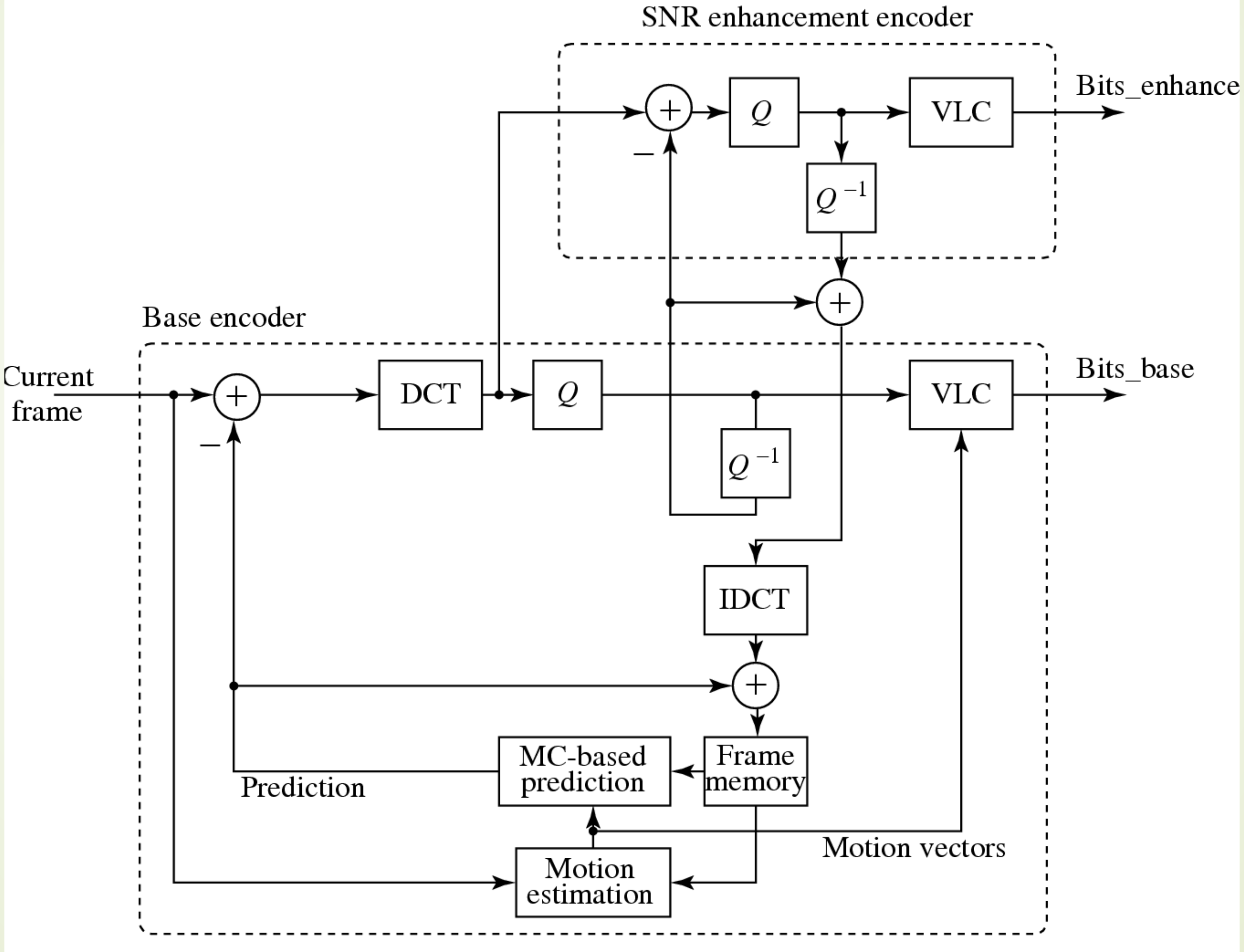
SNR Scalability

- **SNR scalability**: Refers to the enhancement/refinement over the base layer to improve the Signal-Noise-Ratio (SNR).

- The MPEG-2 SNR scalable encoder will generate output bitstreams *Bits_base* and *Bits_enhance* at two layers:

  1. At the Base Layer, a coarse quantization of the DCT coefficients is employed which results in fewer bits and a relatively low quality video.
  2. The coarsely quantized DCT coefficients are then inversely quantized (Q^{-1}) and fed to the Enhancement Layer to be compared with the original DCT coefficient.
  3. Their difference is finely quantized to generate a **DCT coefficient refinement**, which, after VLC, becomes the bitstream called Bits_enhance.

MPEG-2 SNR Scalability (Encoder)

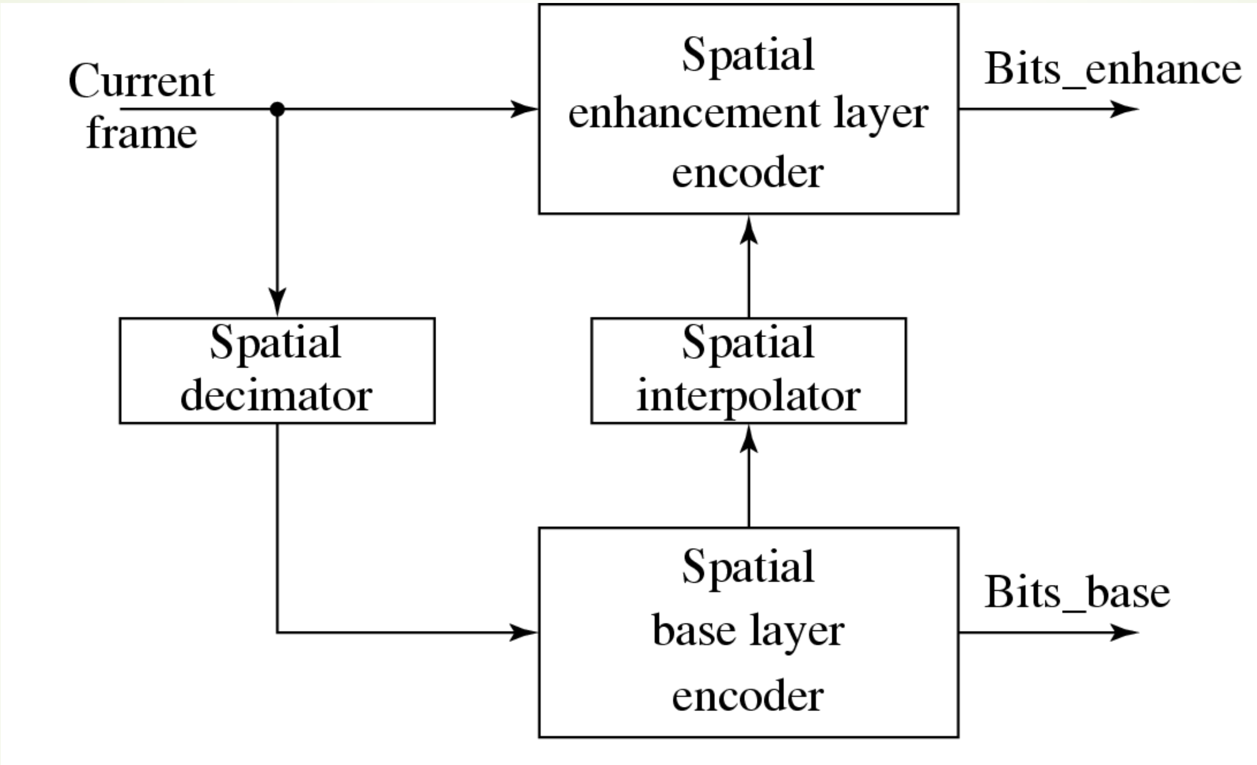


(a) Encoder

Spatial Scalability

- The base layer is designed to generate bitstream of reduced resolution pictures.
- When combined with the enhancement layer, pictures at the original resolution are produced.
- The Base and Enhancement layers for MPEG-2 spatial scalability are not as tightly coupled as in SNR scalability.

Encoder of MPEG-2 Spatial Scalability

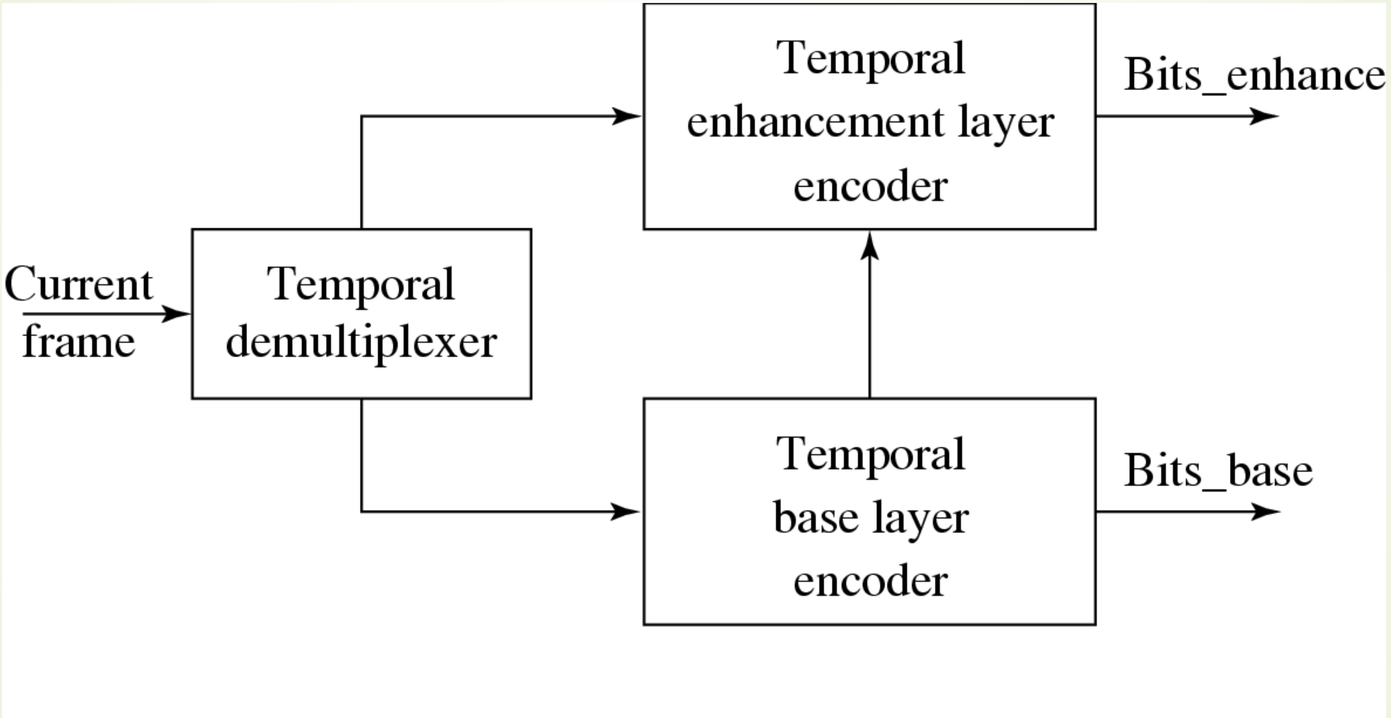


(a)

Temporal Scalability

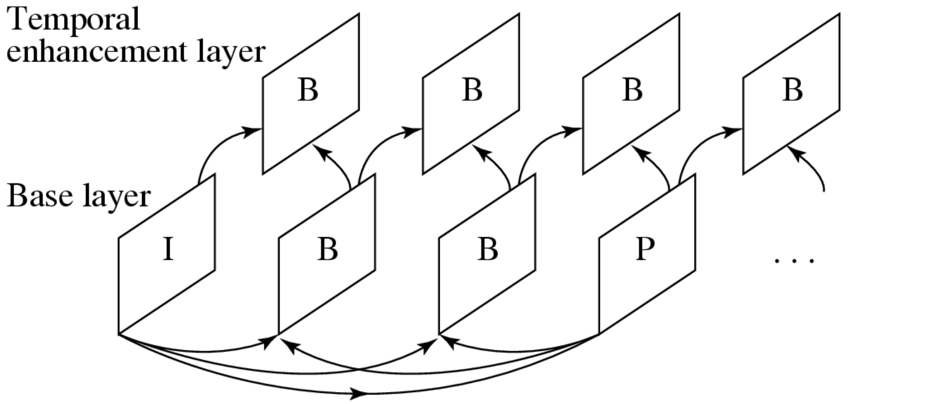
- The input video is temporally split into two pieces, each carrying half of the original frame rate.
- Base Layer Encoder carries out the normal single-layer coding procedures.
- The prediction of matching MBs at the Enhancement Layer can be obtained in two ways:
  - Interlayer MC (Motion-Compensated) Prediction
  - Combined MC Prediction and Interlayer MC Prediction

Encoder of MPEG-2 Temporal Scalability

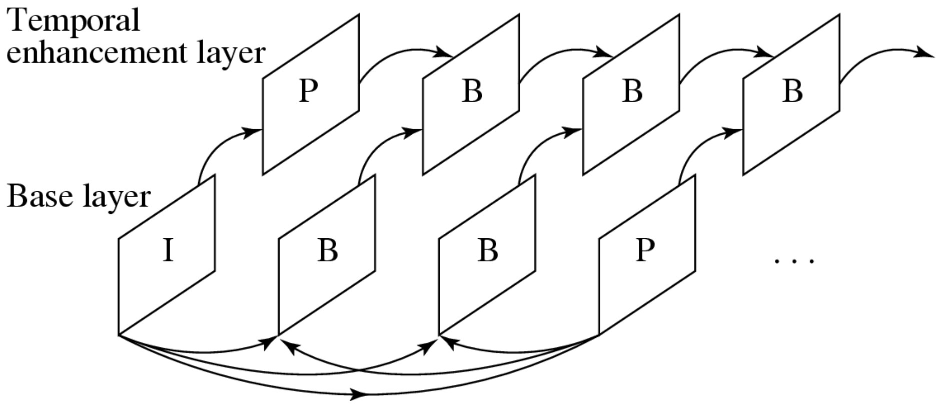

Temporal enhancement layer encoder
Bits_enhance
Current frame
Temporal demultiplexer
Temporal base layer encoder
Bits_base

(a) Block Diagram

Encoder of MPEG-2 Temporal Scalability



(b) Interlayer Motion-Compensated (MC) Prediction



(c) Combined MC Prediction and Interlayer MC Prediction

Overview of H.264

- H.264 is known as MPEG-4 Part 10, AVC (Advanced Video Coding). It is often referred to as the H.264/AVC (or H.264/MPEG-4 AVC) video coding standard.
- H.264 provides a higher video coding efficiency, up to 50% better compression than MPEG-2 and up to 30% better than H.263 and MPEG-4, while maintaining the same quality of the compressed video.

Overview of H.264 (Cont'd)

- Main features of H.264/AVC are:
  - Integer transform in 4 × 4 blocks.
  - Block-size motion compensation in luma images.
  - Quarter-pixel accuracy in motion vectors.
  - Multiple reference picture motion compensation.
  - Robust to data errors and data losses.

Motion compensation

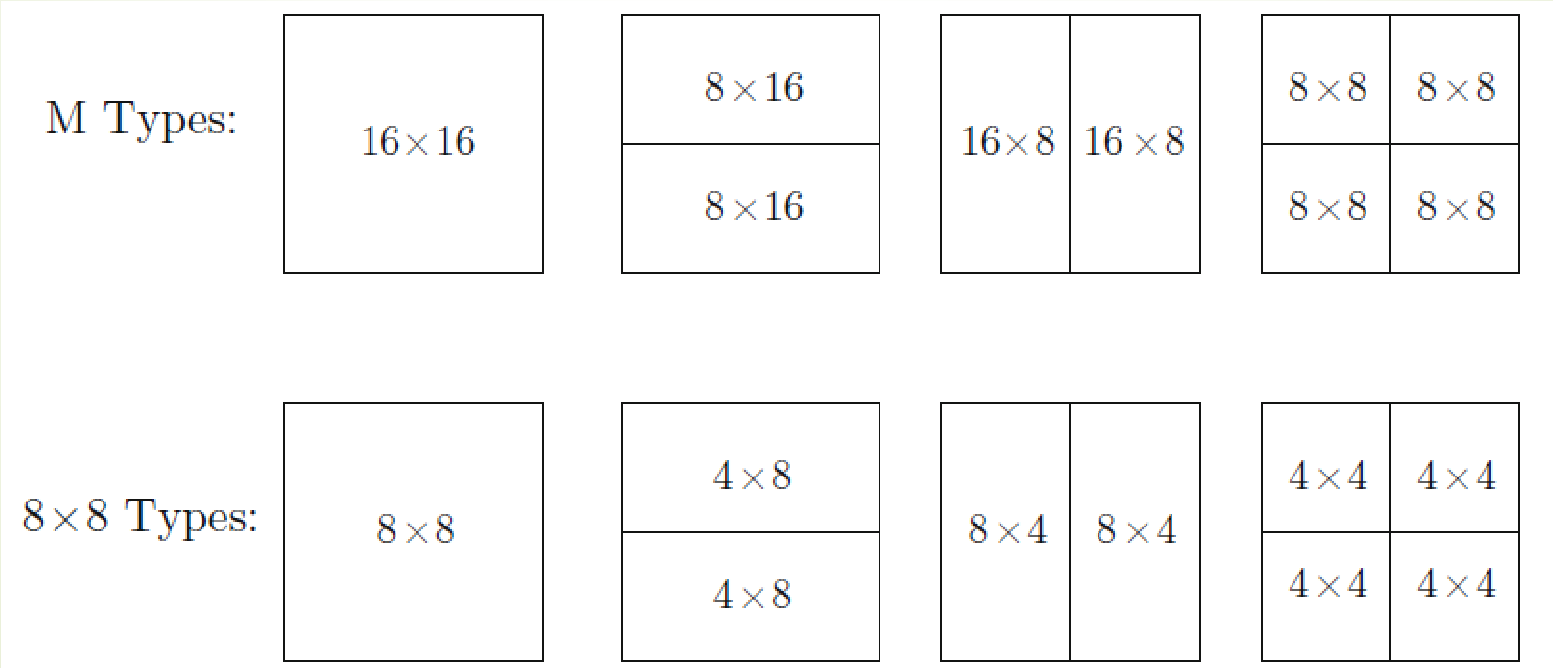
- H.264 employs the technology of *hybrid coding*, i.e.,
  - a combination of inter-picture motion predictions and intra-picture spatial prediction, and transform coding on residual errors.

Intra-Frame Motion Prediction

- H.264EVC offers the option where the pixel values from the same video frame as the current block, are used for prediction.
- This kind of prediction is called spatial or intra-frame prediction (Intra).
- Thus, the term “hybrid” refers to the use of the two possible ways of eliminating temporal or spatial redundancy in videos at the same time.
- H.264 is therefore called Highly Efficient Video Coding (HEVC)

Variable Block-Size Compensation

- Initial macroblock size is 16×16, but could be further divided into small blocks

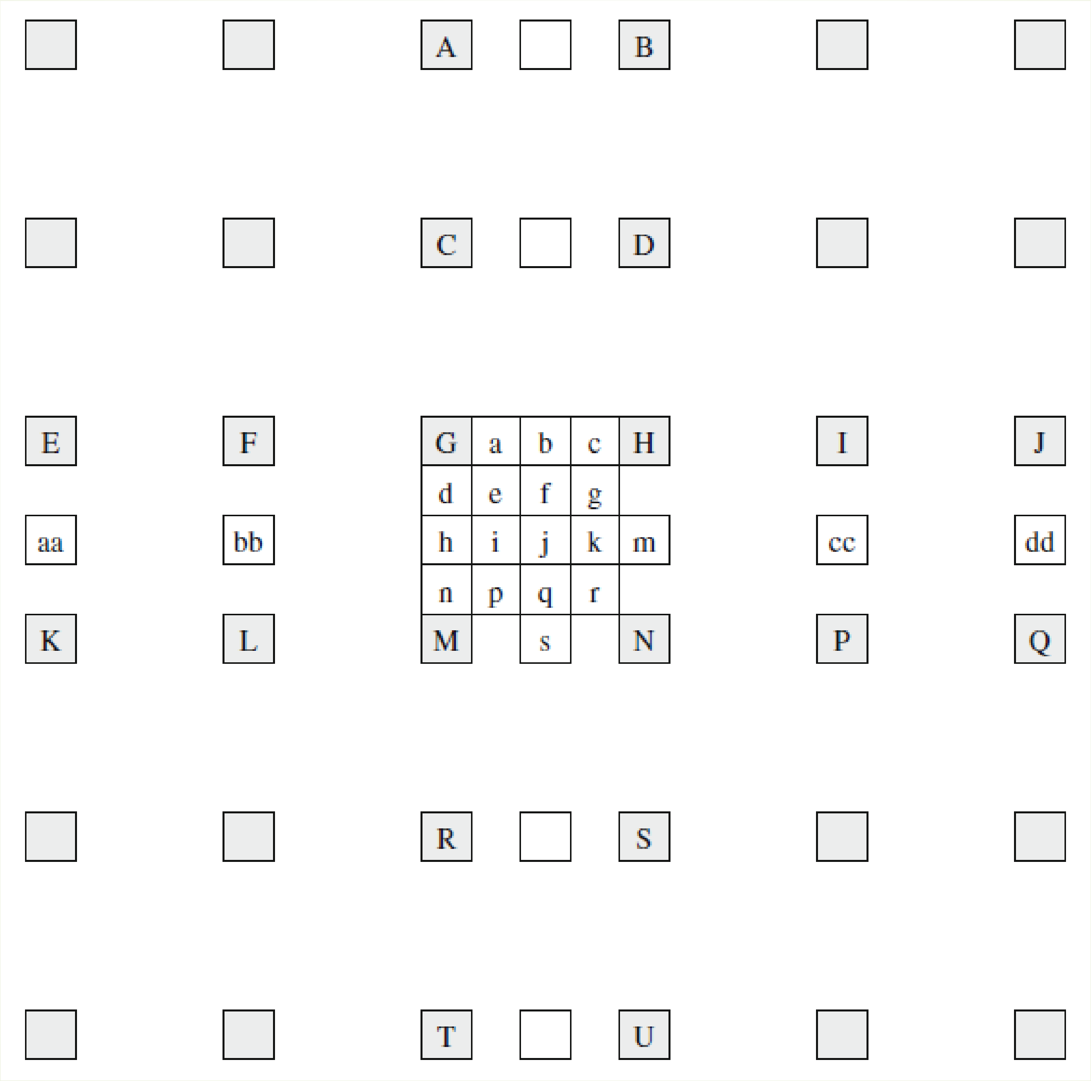


Quarter-Pixel Precision

- The accuracy of motion compensation is quarter-pixel precision in luma images.
- Pixel values at half-pixel and quarter-pixel positions can be derived by interpolation
- The half pixel b and h could be calculated by

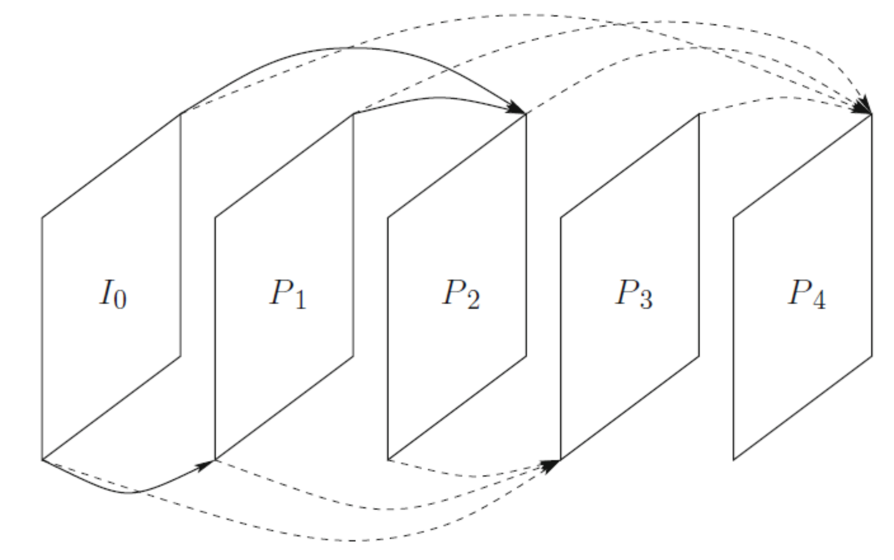
$$b_1 = E - 5F + 20G + 20H - 5I + J$$
$$h_1 = A - 5C + 20G + 20M - 5R + T$$
$$b = (b_1 + 16) \gg 5$$
$$h = (h_1 + 16) \gg 5.$$

H.264 Fractional Sample Interpolation



Additional Options in Group of Pictures

- No B-frames
- Multiple reference frames



- Hierarchical prediction structure

Questions?