Multimedia Computing

Video Compression Standards

MPEG

Video Compression

- A video consists of a time-ordered sequence of frames, i.e. images.
- An obvious solution to video compression would be predictive coding based on previous frames.
- Compression proceeds by subtracting images: subtract in time order and code the residual error.
- It can be done even better by searching for just the right parts of the image to subtract from the previous frame.

Video Compression with Motion Compensation

- Consecutive frames in a video are similar, i.e. temporal redundancy exists.
- The difference between the current frame and other frame(s) in the sequence will be coded – small values and low entropy, good for compression.
- Steps of Video compression based on Motion Compensation (MC):
  - 1. Motion Estimation (we have studied it in last lecture).
  - 2. MC-based Prediction.
  - 3. Derivation of the prediction error, i.e., the difference.

Topics

- H.261 and H. 263
- MPEG 1 and MPEG 2

H.261

- **H.261**: An earlier digital video compression standard, its principle of MC-based compression is retained in all later video compression standards.
  - The standard was designed for videophone, video conferencing and other audiovisual services over ISDN (Integrated Services Digital Network).
  - The video codec supports bit-rates of $p \times 64$ kbps, where p ranges from 1 to 30.
  - Require that the delay of the video encoder be less than 150 ms so that the video can be used for real-time bidirectional video conferencing.

H.261

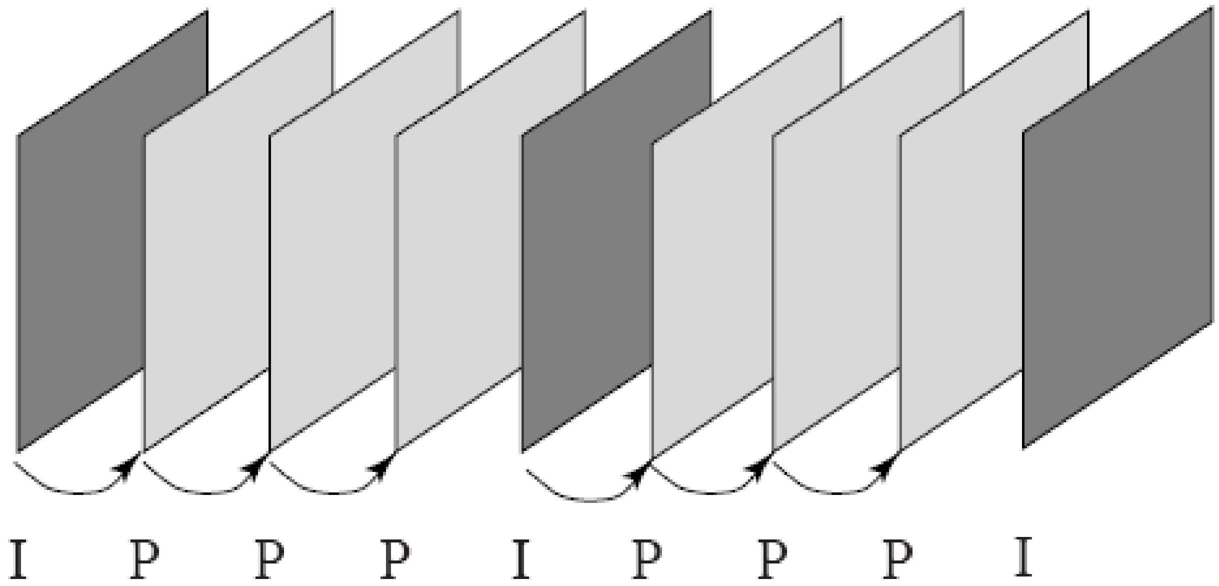
Video Formats Supported by H.261

Video format	Luminance image resolution	Chrominance image resolution	Bit-rate (Mbps) (if 30 fps and uncompressed)	H.261 support
QCIF	176×144	88×72	9.1	required
CIF	352×288	176×144	36.5	optional

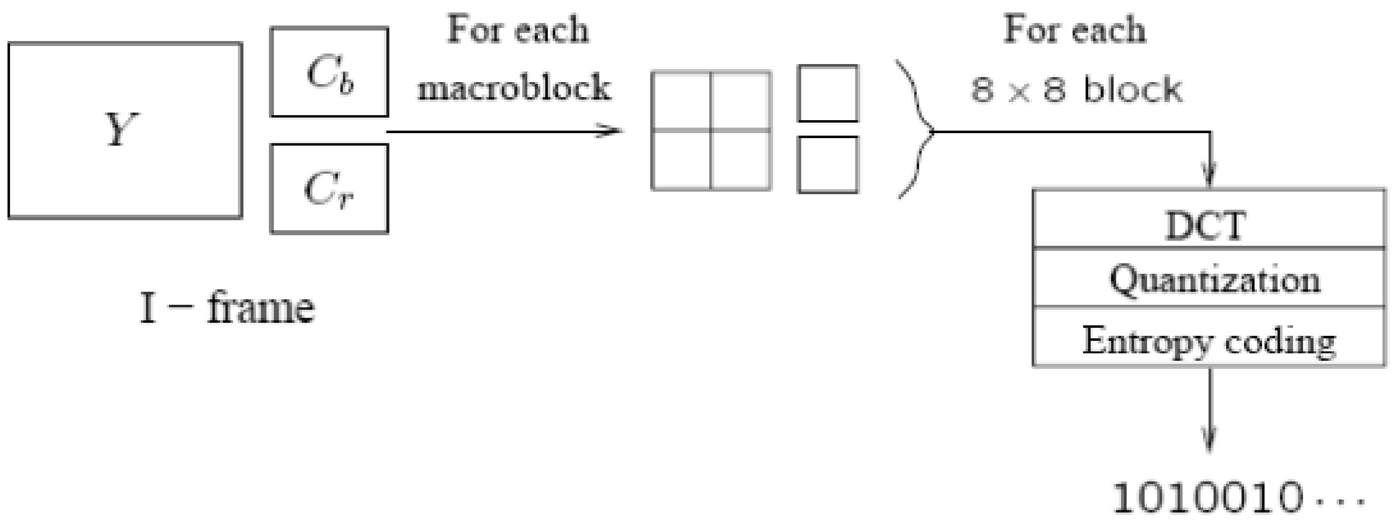
H.261 Frame Sequence

- Two types of image frames are defined: Intra-frames (**I-frames**) and Inter-frames (**P-frames**):
  - I-frames are treated as independent images. Transform coding method similar to JPEG is applied within each I-frame, hence “Intra".
  - P-frames are not independent: coded by a forward predictive coding method (prediction from a previous P-frame is allowed – not just from a previous I-frame).
  - Temporal redundancy removal is included in P-frame coding, whereas I-frame coding performs only spatial redundancy removal.
  - To avoid propagation of coding errors, an I-frame is usually sent a couple of times in each second of the video.
- Motion vectors in H.261 are always measured in units of full pixel and they have a limited range of 15 pixels, i.e., $p = 15$.

H.261 Frame Sequence

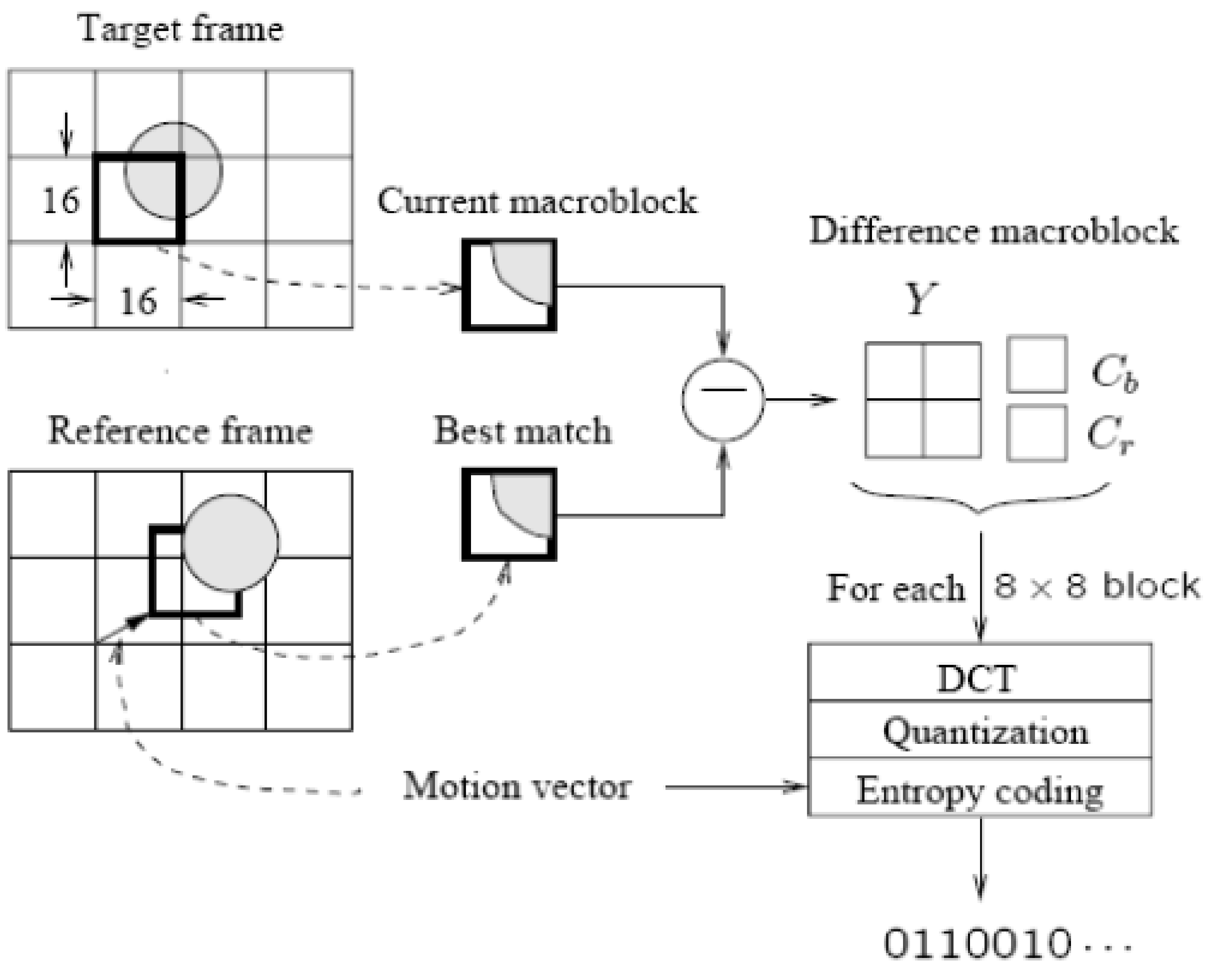


I-frame Coding



- Macroblocks are of size 16×16 pixels for the Y frame, and 8×8 for Cb and Cr frames (4:2:0 chroma subsampling is employed).
- A macroblock consists of four Y, one Cb, and one Cr 8×8 blocks.
- For each 8×8 block a DCT transform is applied, the DCT coefficients then go through quantization zigzag scan and entropy coding.

P-frame MC-based Predictive Coding



P-frame Predictive Coding

- For each macroblock in the Target frame, a motion vector is allocated by one of the search methods discussed earlier.
- After the prediction, a *difference macroblock* is derived to measure the *prediction error*. Each of these 8×8 blocks go through DCT, quantization, zigzag scan and entropy coding procedures.
- The P-frame coding encodes the difference macroblock. If the prediction error exceeds a certain acceptable level. The macroblock itself is then encoded. This case is termed a *non-motion compensated macroblock*.
- For motion vector, the difference **MVD** is sent for entropy coding:

$$\mathrm{MVD} = \mathrm{MV}_{\mathrm{Preceding}} - \mathrm{MV}_{\mathrm{Current}}$$

Quantization in H.261

- The quantization in H.261 uses a constant step size, for all DCT coefficients within a macroblock.
- If we use *DCT* and *QDCT* to denote the DCT coefficients before and after the quantization, then for DC coefficients in Intra mode:

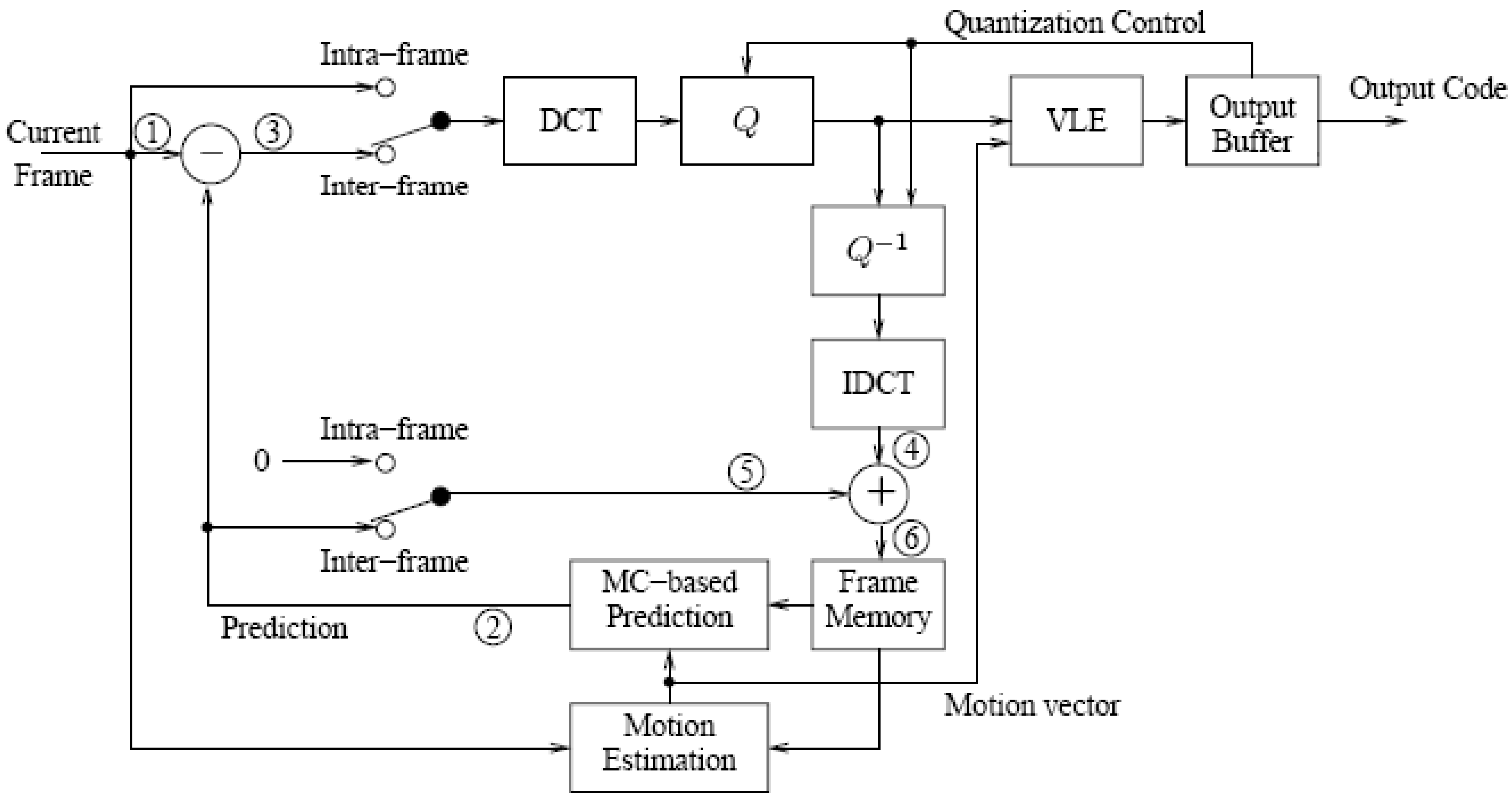
$$QDCT = round\left(\frac{DCT}{step\_size}\right) = round\left(\frac{DCT}{8}\right)$$

for all other coefficients:

$$QDCT = \left\lfloor\frac{DCT}{step\_size}\right\rfloor = \left\lfloor\frac{DCT}{2 * scale}\right\rfloor$$

scale is an integer in the range of [1, 31].

H. 261 Encoder

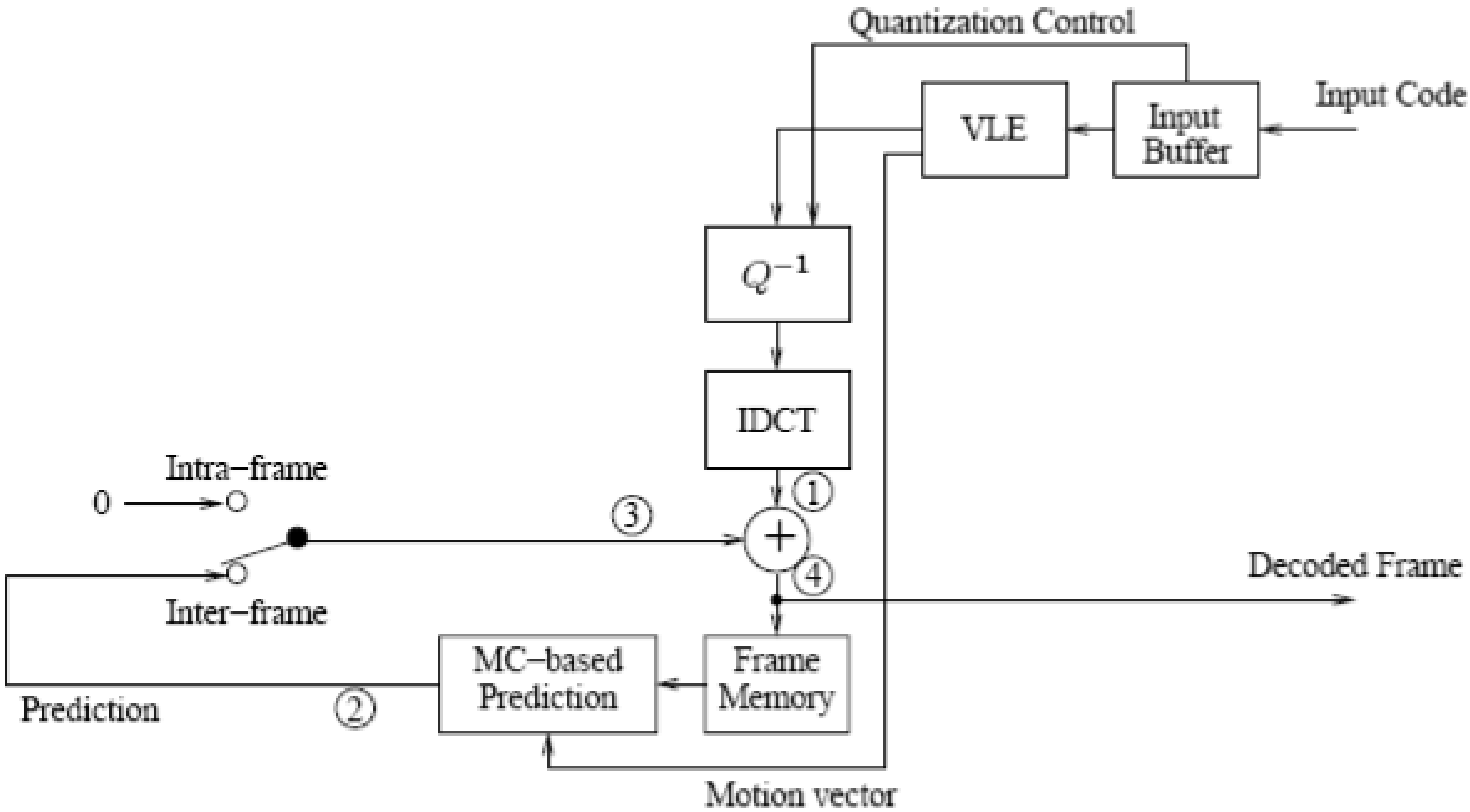


H. 261 Encoder

Current Frame	Observation Point					
	1	2	3	4	5	6
I	I			$\tilde{I}$	0	$\tilde{I}$
P_1	P_1	P_1'	D_1	$\tilde{D}_1$	P_1'	$\tilde{P}_1$
P_2	P_2	P_2'	D_2	$\tilde{D}_2$	P_2'	$\tilde{P}_2$

Data Flow at the Observation Points in H.261 Encoder

H. 261 Decoder



H. 261 Decoder

Current Frame	Observation Point			
	1	2	3	4
I	$\tilde{I}$		0	$\tilde{I}$
P_1	$\tilde{D}_1$	P'_1	P'_1	$\tilde{P}_1$
P_2	$\tilde{D}_2$	P'_2	P'_2	$\tilde{P}_2$

Data Flow at the Observation Points in H.261 Decoder

H. 263

- H.263 is an improved video coding standard for video conferencing and other audiovisual services transmitted on Public Switched Telephone Networks (PSTN).
  - Aims at low bit-rate communications at bit-rates of less than 64 kbps.
  - Uses predictive coding for inter-frames to reduce temporal redundancy and transform coding for the remaining signal to reduce spatial redundancy (for both intra-frames and inter-frame prediction).

H. 263

Video format	Luminance image resolution	Chrominance image resolution	Bit-rate (Mbps) (if 30 fps and uncompressed)	Bit-rate (kbps) BPPmaxKb (compressed)
sub-QCIF	128×96	64×48	4.4	64
QCIF	176×144	88×72	9.1	64
CIF	352×288	176×144	36.5	256
4CIF	704×576	352×288	146.0	512
16CIF	$1,408 \times 1,152$	704×576	583.9	1024

Video Formats Supported by H.263

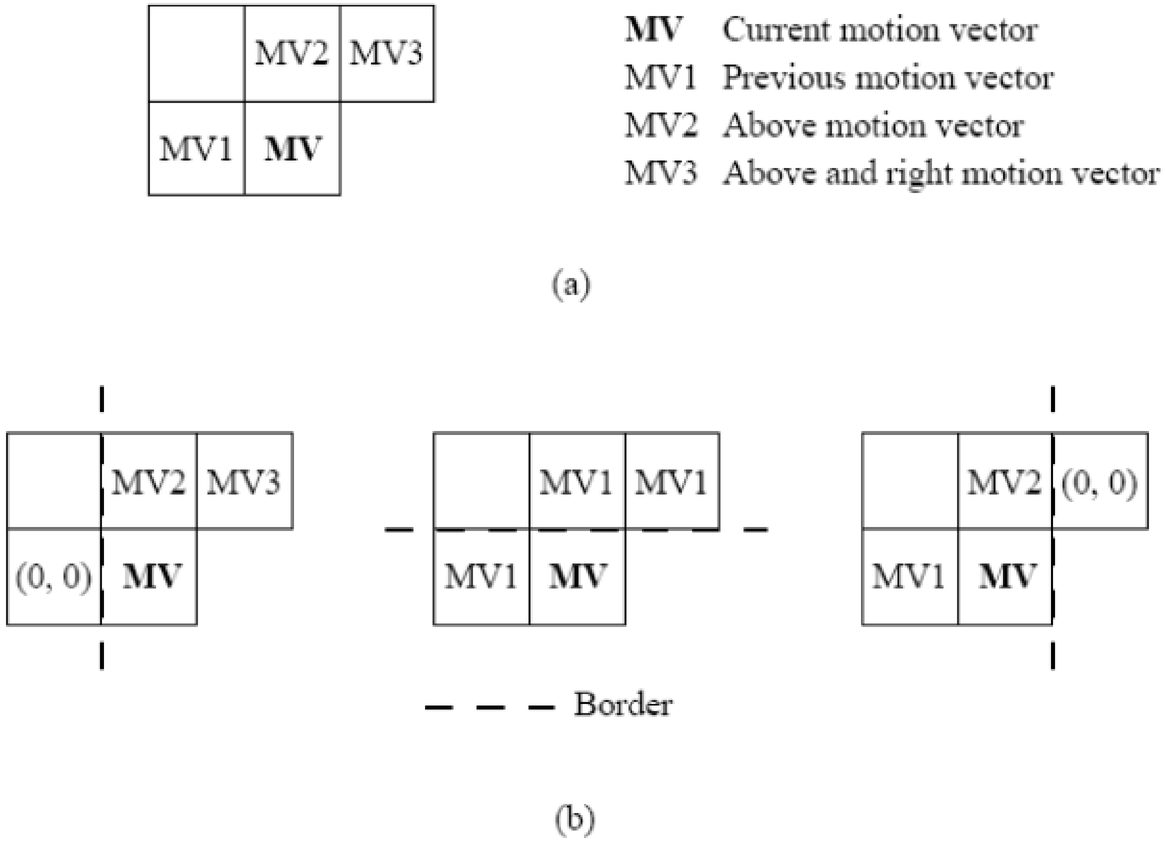
Motion Compensation in H.263

- The horizontal and vertical components of the MV are predicted from the median values of the horizontal and vertical components, respectively, of MV1, MV2, MV3 from the "previous", "above" and "above and right" blocks.
- For the Macroblock with MV(u, v):

$$
\begin{aligned}
u_p &= median(u_1, u_2, u_3) \\
v_p &= median(v_1, v_2, v_3).
\end{aligned}
$$

Instead of coding the MV(u, v) itself, the error vector (δu, δv) is coded, where $\delta u = u - u_p$ and $\delta v = v - v_p$.

Motion Compensation in H.263

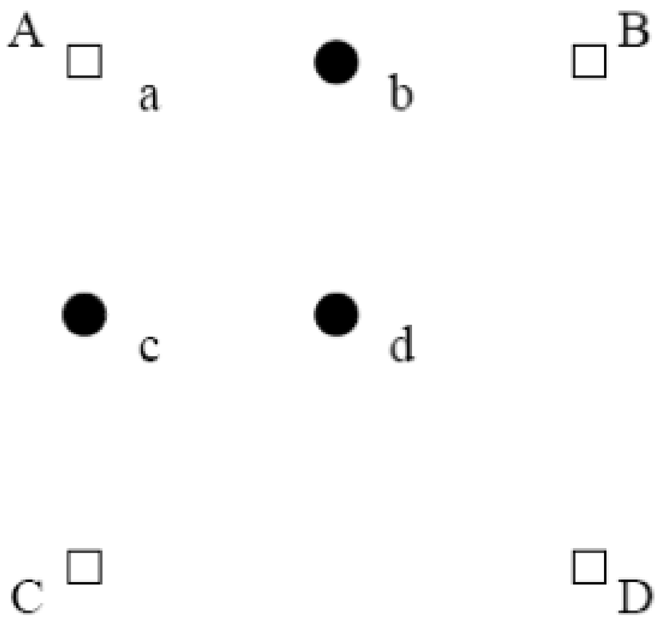


Prediction of Motion Vector in H.263.

Half-Pixel Precision

- In order to reduce the prediction error, **half-pixel precision** is supported in H.263 vs. full-pixel precision only in H.261.
  - The default range for both the horizontal and vertical components *u* and *v* of MV(*u, v*) are now [−16*;* 15.5].
  - The pixel values needed at half-pixel positions are generated by a simple bilinear interpolation method.

Half-pixel precision by bilinear interpolation in H.263



□ Full-pixel position

● Half-pixel position

$a = A$

$b = (A + B + 1) / 2$

$c = (A + C + 1) / 2$

$d = (A + B + C + D + 2) / 4$

Topics

- H.261 and H. 263
- MPEG 1 and MPEG 2

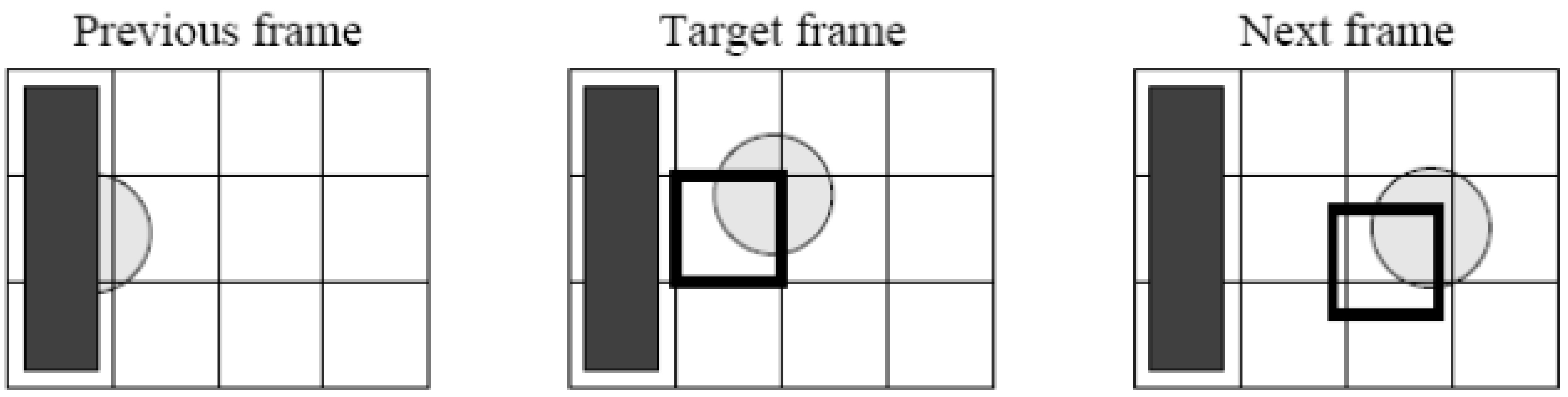
MPEG-1

- **MPEG**: *Moving Pictures Experts Group*, established in 1988 for the development of digital video.
- MPEG-1 adopts the CCIR601 digital TV format also known as SIF (*Source Input Format*).
- MPEG-1 supports only non-interlaced video. Normally, its picture resolution is:
  - 352×240 for NTSC video at 30 fps
  - 352×288 for PAL video at 25 fps
  - It uses 4:2:0 chroma subsampling
- The MPEG-1 standard is also referred to as ISO/IEC 11172. It has five parts: 11172-1 Systems, 11172-2 Video, 11172-3 Audio, 11172-4 Conformance, and 11172-5 Software.

Motion Compensation in MPEG-1

- Motion Compensation (MC) based video encoding in H.261 works as follows:
  - In Motion Estimation (ME), each macroblock (MB) of the Target P-frame is assigned a best matching MB from the previously coded I or P frame - **prediction**.
  - **prediction error:** The difference between the MB and its matching MB, sent to DCT and its subsequent encoding steps.
  - The prediction is from a previous frame – **forward prediction**.

The Need for Bidirectional Search

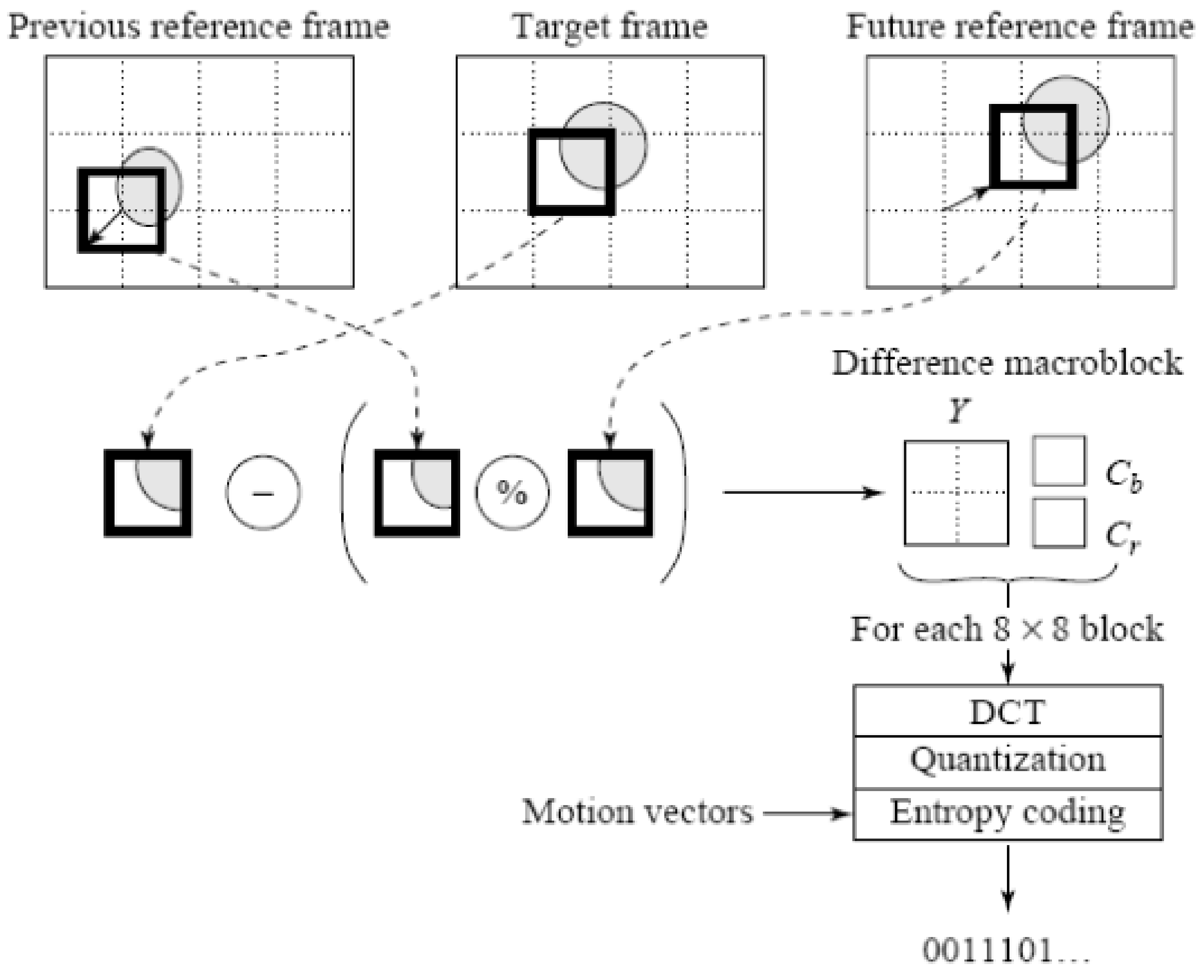


- The MB containing part of a ball in the Target frame cannot find a good matching MB in the previous frame because half of the ball was occluded by another object. A match however can readily be obtained from the next frame.

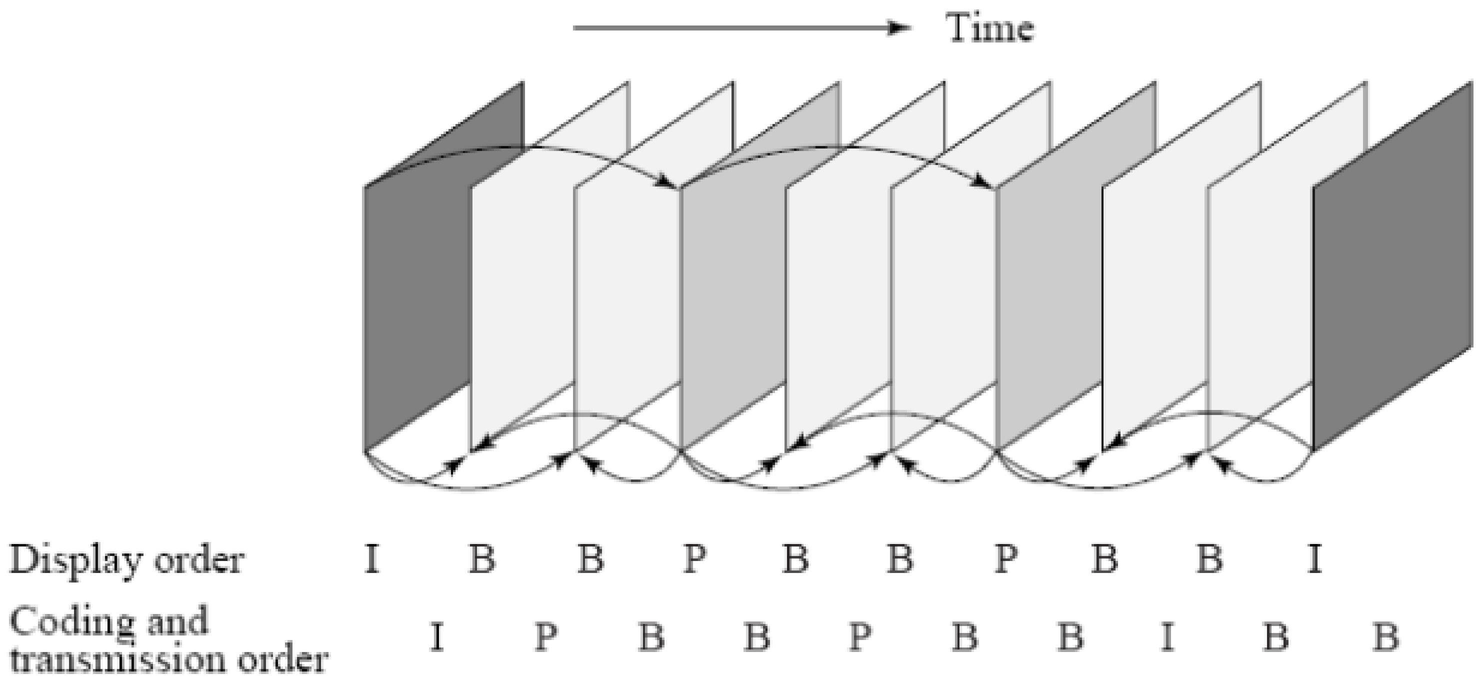
Motion Compensation in MPEG-1

- MPEG introduces a third frame type – *B-frames,* and its accompanying bi-directional MC.
  - Each MB from a B-frame will have up to *two* motion vectors (MVs) (one from the forward and one from the backward prediction).
  - If matching in both directions is successful, then two MVs will be sent and the two corresponding matching MBs are averaged (indicated by "%" in the figure) before comparing to the Target MB for generating the prediction error.
  - If an acceptable match can be found in only one of the reference frames, then only one MV and its corresponding MB will be used from either the forward or backward prediction.

B-frame Coding Based on Bidirectional Motion Compensation



MPEG Frame Sequence



Other Major Differences from H.261

- Source formats supported:
  - H.261 only supports CIF (352×288) and QCIF (176×144) source formats, MPEG-1 supports SIF (352×240 for NTSC, 352×288 for PAL).
  - MPEG-1 also allows specification of other formats as long as the Constrained Parameter Set (CPS) shown below is satisfied:

Parameter	Value
Horizontal size of picture	≤ 768
Vertical size of picture	≤ 576
No. of MBs / picture	≤ 396
No. of MBs / second	$\leq 9,900$
Frame rate	≤ 30 fps
Bit-rate	$\leq 1,856$ kbps

Other Major Differences from H.261

- Quantization: MPEG-1 quantization uses different quantization tables for its Intra and Inter coding (see next slides).

For DCT coefficients in Intra mode:

$$QDCT[i,j] = round\left(\frac{8 \times DCT[i,j]}{step\_size[i,j]}\right) = round\left(\frac{8 \times DCT[i,j]}{Q_1[i,j] * scale}\right)$$

For DCT coefficients in Intra mode:

$$QDCT[i,j] = \left\lfloor\frac{8 \times DCT[i,j]}{step\_size[i,j]}\right\rfloor = \left\lfloor\frac{8 \times DCT[i,j]}{Q_2[i,j] * scale}\right\rfloor$$

Quantization Tables

Default Quantization Table (Q_1) for Intra-Coding

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Default Quantization Table (Q_2) for Inter-Coding

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

Other Major Differences from H.261

- MPEG-1 allows motion vectors to be of sub-pixel precision (1/2 pixel). The technique of “bilinear interpolation" for H.263 can be used to generate the needed values at half-pixel locations.
- Compared to the maximum range of 15 pixels for motion vectors in H.261, MPEG-1 supports a range of [−512, 511.5] for half-pixel precision and [−1024, 1023] for full-pixel precision motion vectors.
- The MPEG-1 bitstream allows random access.

Typical Sizes of MPEG-1 Frames

- The typical size of compressed P-frames is significantly smaller than that of I-frames, because temporal redundancy is exploited in inter-frame compression.
- B-frames are even smaller than P-frames because of (a) the advantage of bi-directional prediction and (b) the lowest priority given to B-frames.

Typical Compression Performance of MPEG-1 Frames

Type	Size	Compression
I	18 kB	7:1
P	6 kB	20:1
B	2.5 kB	50:1
Avg	4.8 kB	27:1

MPEG-2

- MPEG-2: For higher quality video at a bit-rate of more than 4 Mbps.
- Defined seven profiles aimed at different applications:
  - Simple, Main, SNR scalable, Spatially scalable, High, 4:2:2, Multiview.
  - Within each profile, up to four levels are defined (see the Tables in next slide).
  - The DVD video specification allows only four display resolutions: 720×480, 704×480, 352×480, and 352×240, which is a restricted form of the MPEG-2 Main profile at the Main and Low levels.

Profiles and Levels in MPEG-2

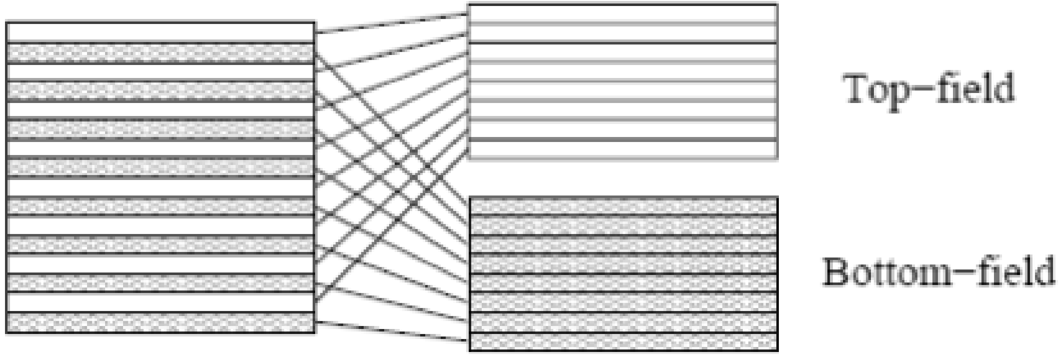
Level	Simple Profile	Main Profile	SNR Scalable Profile	Spatially Scalable Profile	High Profile	4:2:2 Profile	Multiview Profile
High		*			*		
High 1440		*		*	*		
Main	*	*	*		*	*	*
Low		*	*				

Four Levels in the Main Profile of MPEG-2

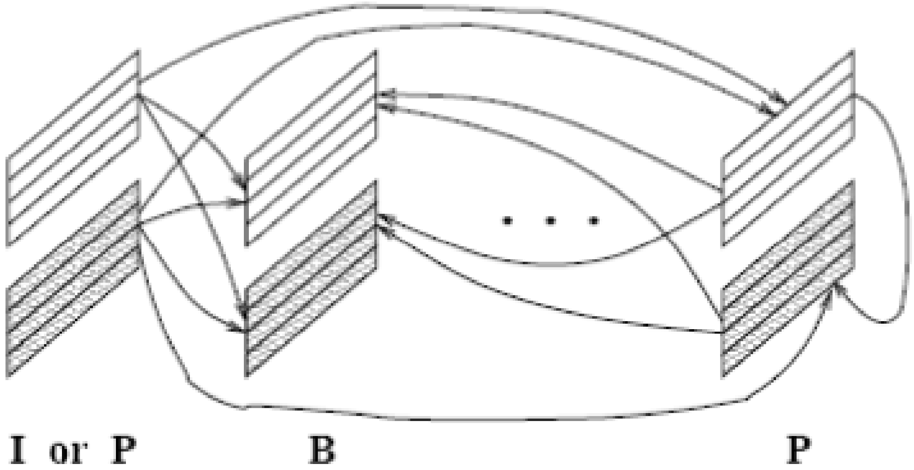
Level	Max Resolution	Max fps	Max Pixels/sec	Max coded Data Rate (Mbps)	Application
High	$1,920 \times 1,152$	60	62.7×10^6	80	film production
High 1440	$1,440 \times 1,152$	60	47.0×10^6	60	consumer HDTV
Main	720×576	30	10.4×10^6	15	studio TV
Low	352×288	30	3.0×10^6	4	consumer tape equiv.

Supporting Interlaced Video

- MPEG-2 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* (see next slide).
  - 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*.



(a) Frame−picture vs. Field−pictures



(b) Field Prediction for Field−pictures

Five Modes of Predictions

- MPEG-2 defines Frame Prediction and Field Prediction as well as five prediction modes:
    I. Frame Prediction for Frame-pictures
    II. Field Prediction for Field-pictures
    III. Field Prediction for Frame-pictures
    IV. 16×8 MC for Field-pictures
    V. Dual-Prime for P-pictures
- Please refer to textbook for detailed information of the five prediction modes.

MPEG-2 Scalabilities

- The MPEG-2 scalable coding: 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.
- 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

- 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.
- Please refer to the textbook for detailed information.

Other Major Differences from MPEG-1

- Better resilience to bit-errors**:** In addition to *Program Stream*, a *Transport Stream* is added to MPEG-2 bit streams.
- Support of 4:2:2 and 4:4:4 chroma subsampling.
- More restricted slice structure: MPEG-2 slices must start and end in the same macroblock row. In other words, the left edge of a picture always starts a new slice and the longest slice in MPEG-2 can have only one row of macroblocks.
- More flexible video formats: It supports various picture resolutions as defined by DVD, ATV and HDTV.

References

- Ze-Nian Li and Marks S. Drew, *Fundamentals of Multimedia,* Pearson Education, Inc.