

CMPT 365 Multimedia Systems

Media Compression - Video Coding Standards

Spring 2017

Video Coding Standards

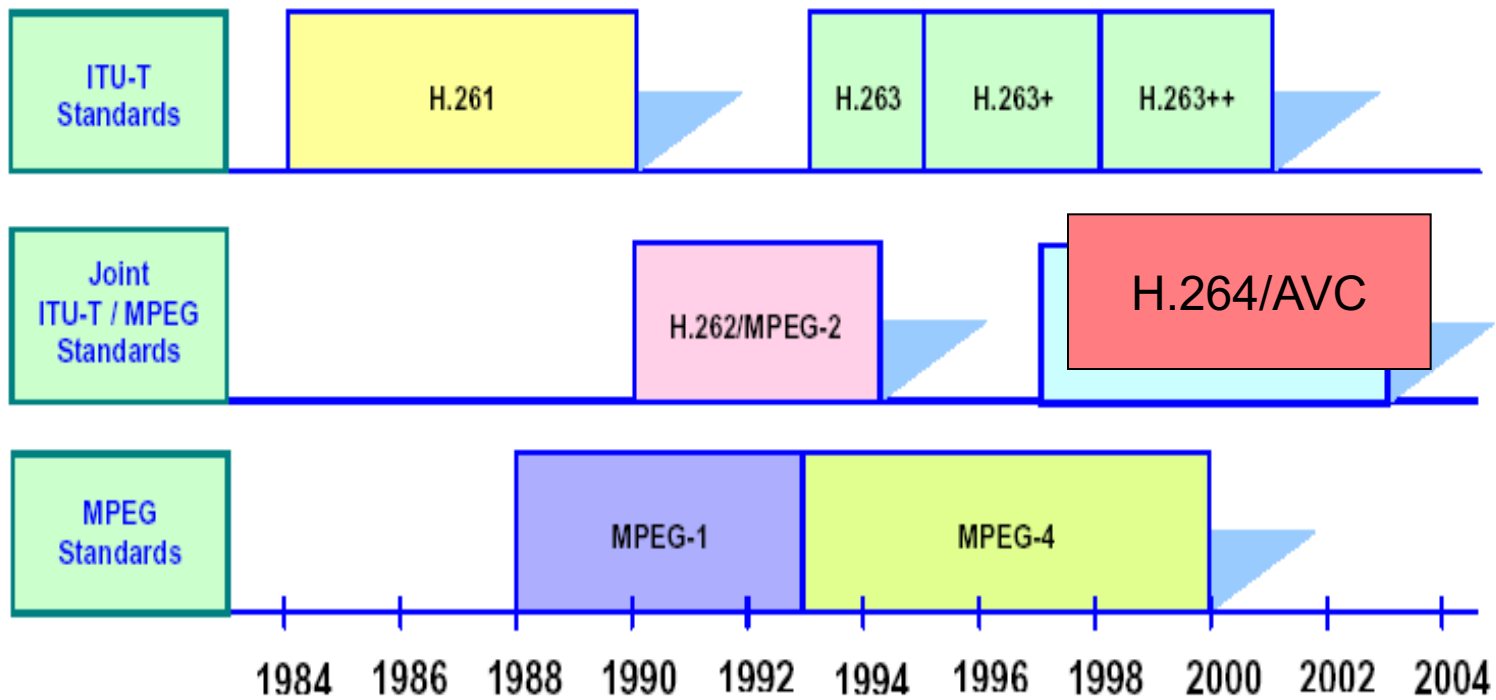
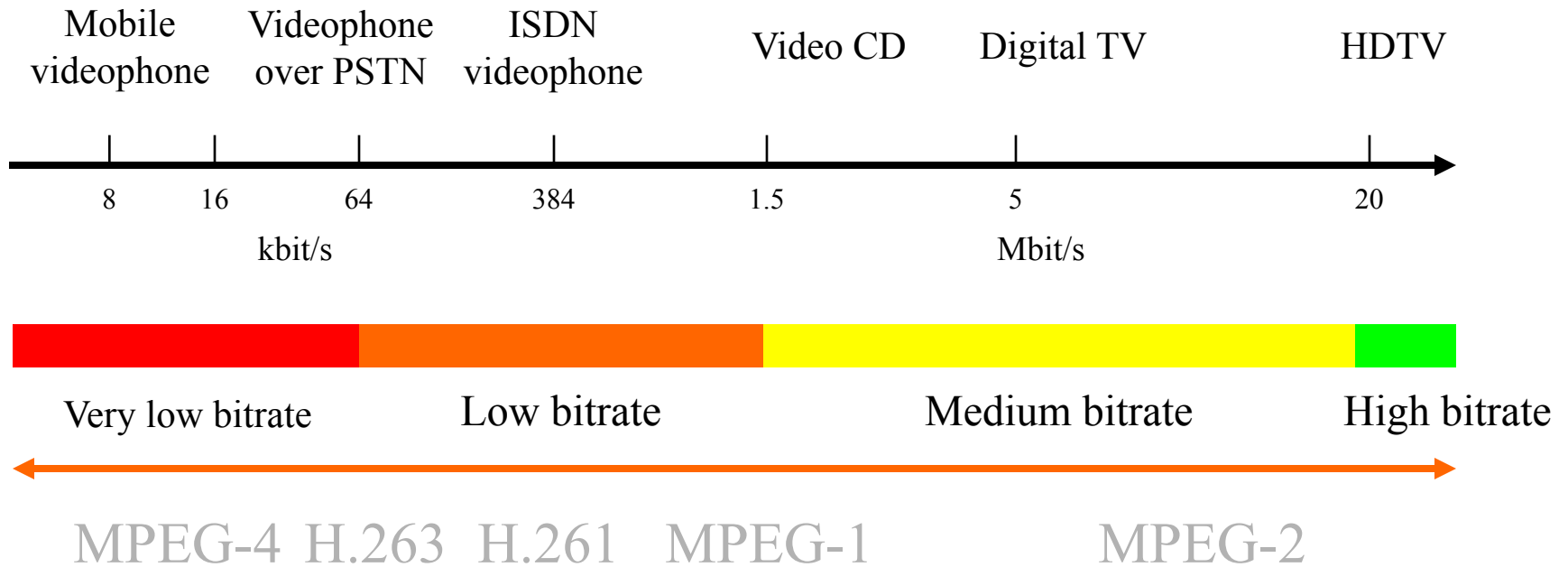


Figure 1. Progression of the ITU-T Recommendations and MPEG standards.

Coding Rate and Standards



Standardization Organizations

□ ITU-T VCEG (Video Coding Experts Group)

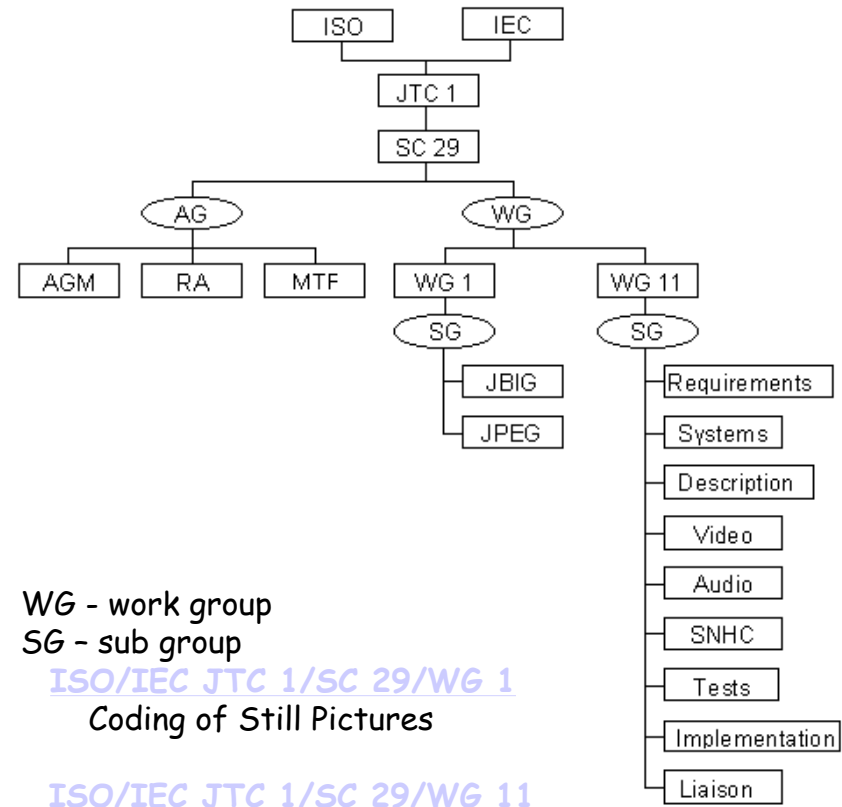
- standards for advanced moving image coding methods appropriate for conversational and non-conversational audio/visual applications.

□ ISO/IEC MPEG (Moving Picture Experts Group)

- standards for compression and coding, decompression, processing, and coded representation of moving pictures, audio, and their combination

□ Relation

- ITU-T H.262~ISO/IEC 13818-2(mpeg2)
Generic Coding of Moving Pictures and Associated Audio.
- ITU-T H.263~ISO/IEC 14496-2(mpeg4)



Introduction

- MPEG-1
- MPEG-2
- MPEG-4

Overview

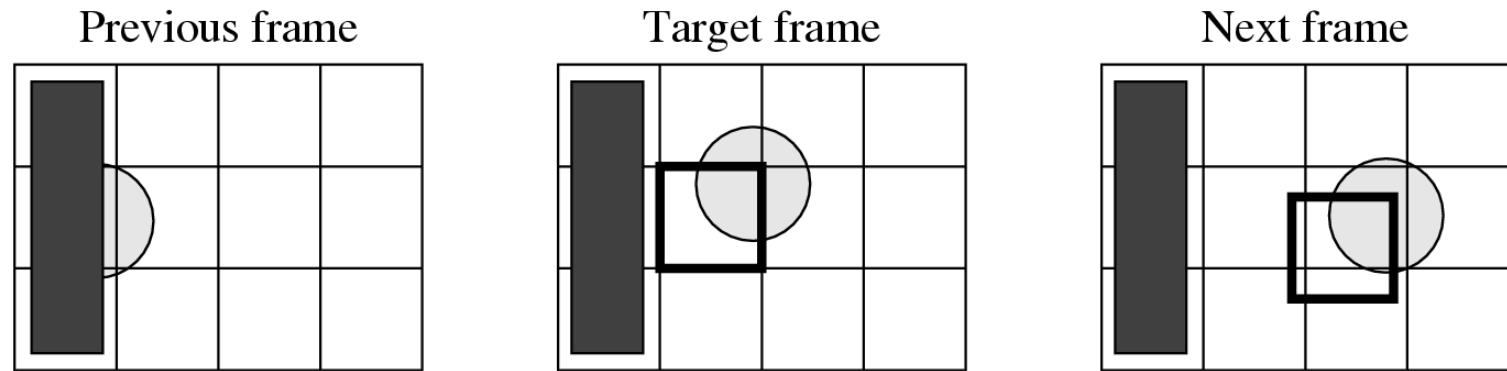
- **MPEG:** *Moving Pictures Experts Group*, established in 1988 for the development of digital video.
- It is appropriately recognized that proprietary interests need to be maintained within the family of MPEG standards:
 - Accomplished by defining only a compressed bitstream that implicitly defines the decoder.
 - The compression algorithms, and thus the encoders, are completely up to the manufacturers.

MPEG-1

- 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**.

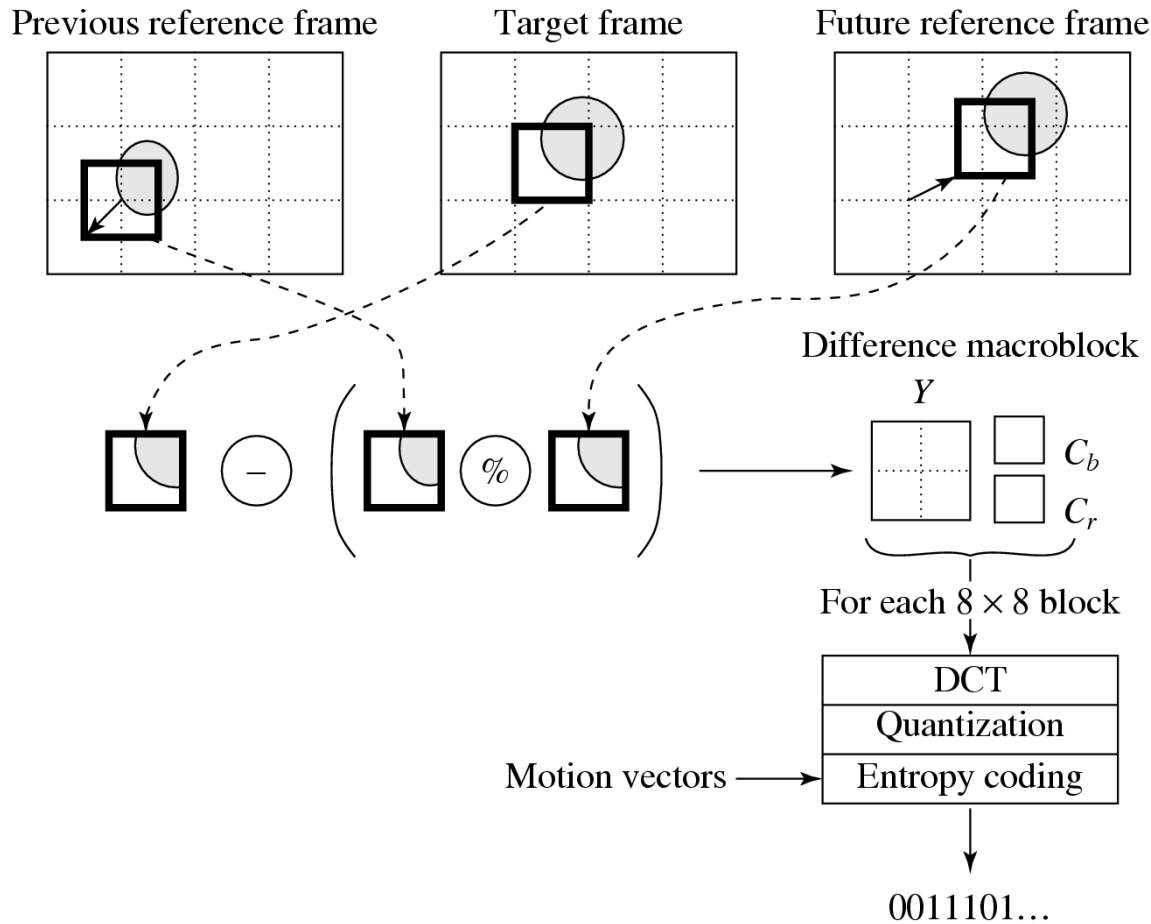


□ Fig 11.1: The Need for Bidirectional Search.

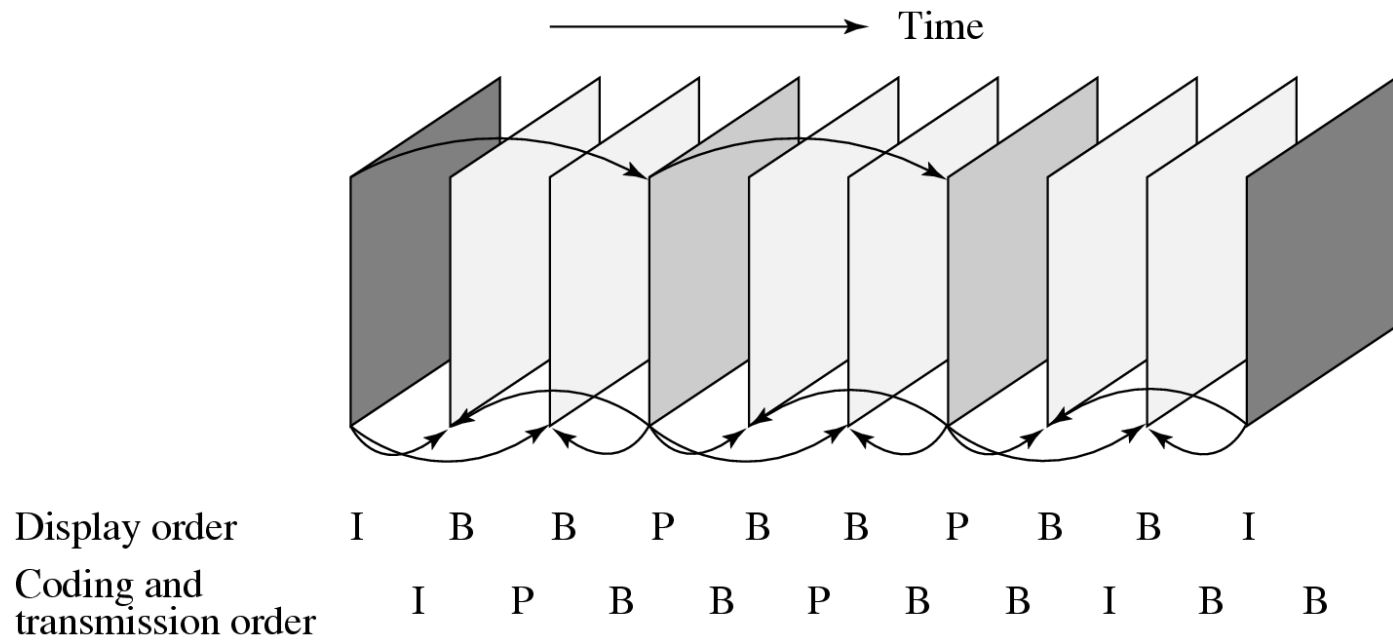
□ The MB containing part of a ball in the Target frame cannot find a good matching MB in the previous frame because half of the ball was occluded by another object. A match however can readily be obtained from the next frame.

Motion Compensation in MPEG-1 (Cont'd)

- MPEG introduces a third frame type — *B-frames*, and its accompanying bi-directional motion compensation.
- The MC-based B-frame coding idea is illustrated in Fig. 11.2:
 - 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.



■ Fig 11.2: B-frame Coding Based on Bidirectional Motion Compensation.



□ Fig 11.3: MPEG Frame Sequence.

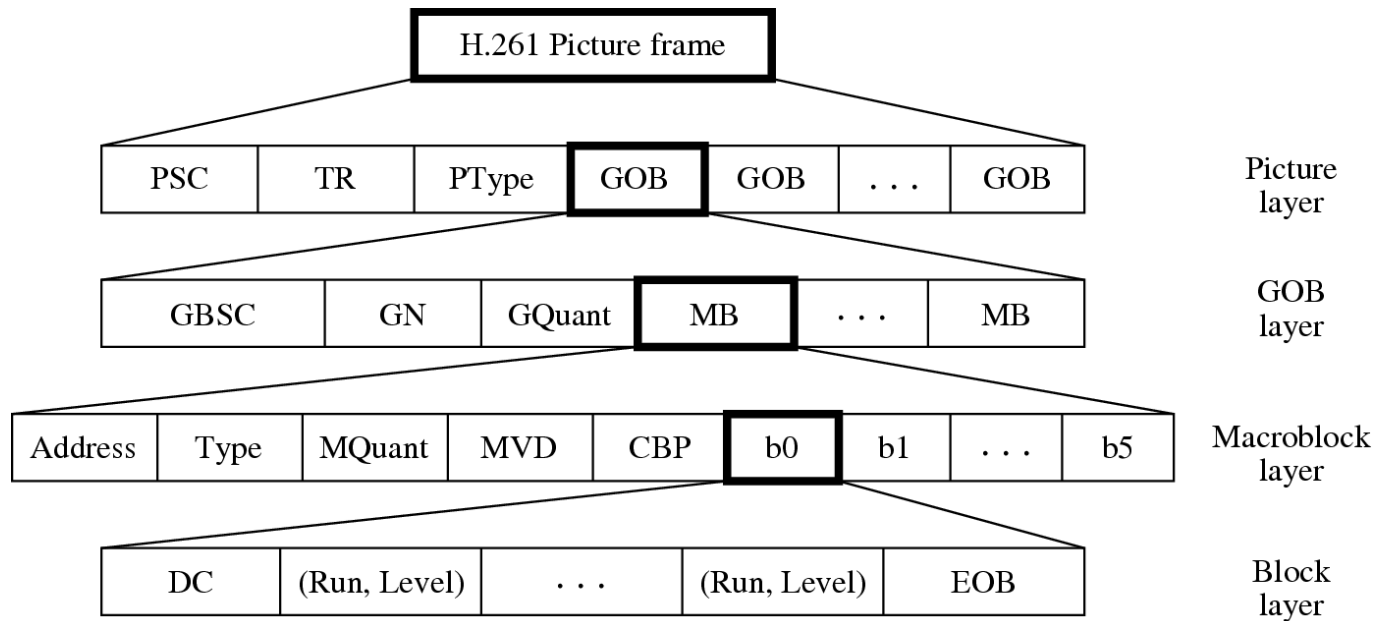
Other Major Differences from H.261

- 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) as shown in Table 11.1 is satisfied:

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

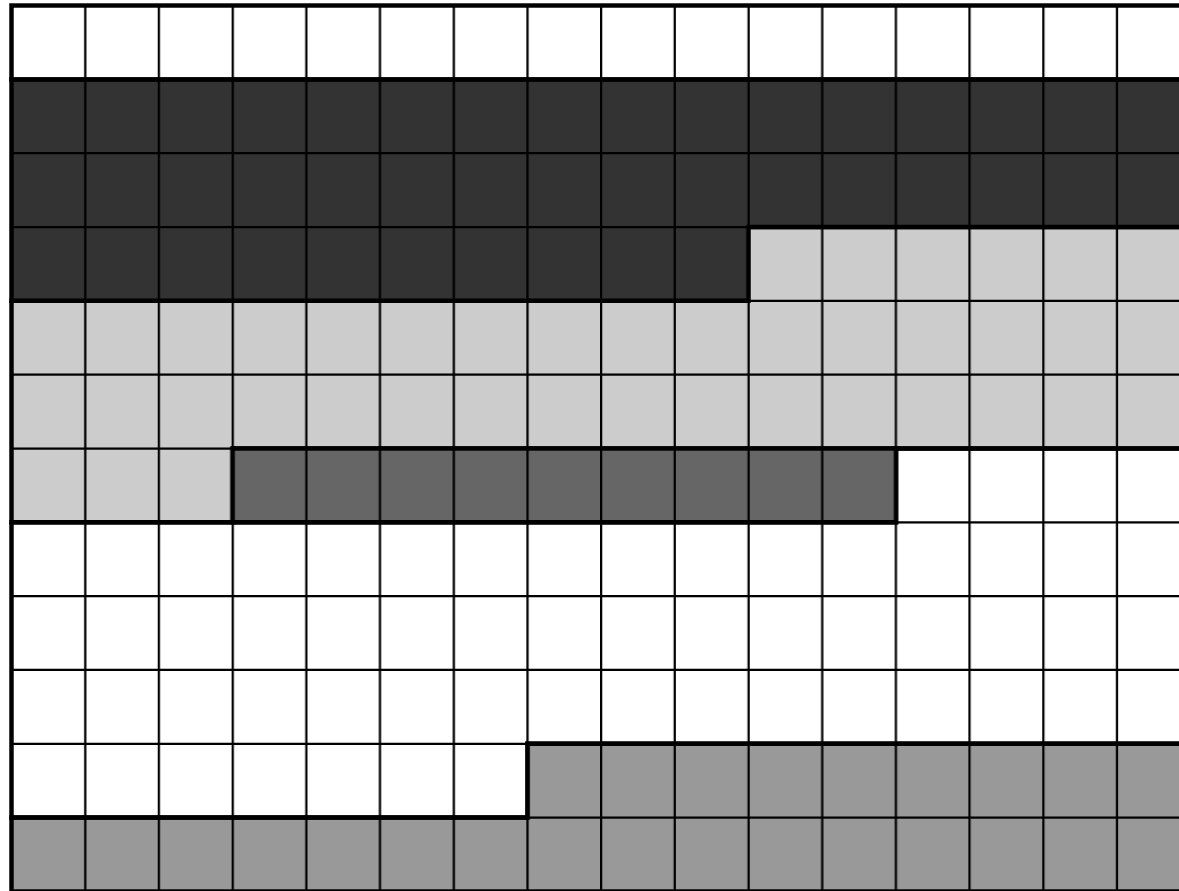
Other Major Differences from H.261 (Cont'd)

- Instead of *GOBs* as in H.261, an MPEG-1 picture can be divided into one or more **slices** (Fig. 11.4):
 - May contain variable numbers of macroblocks in a single picture.
 - May also start and end anywhere as long as they fill the whole picture.
 - Each slice is coded independently — additional flexibility in bit-rate control.
 - Slice concept is important for error recovery.



PSC	Picture Start Code	TR	Temporal Reference
PType	Picture Type	GOB	Group of Blocks
GBSC	GOB Start Code	GN	Group Number
GQuant	GOB Quantizer	MB	Macroblock
MQuant	MB Quantizer	MVD	Motion Vector Data
CBP	Coded Block Pattern	EOB	End of Block

[Fig. 10.8: Syntax of H.261 Video Bitstream.](#)



□ Fig 11.4: Slices in an MPEG-1 Picture.

Other Major Differences from H.261 (Cont'd)

- Quantization:
 - MPEG-1 quantization uses different quantization tables for its Intra and Inter coding (Table 11.2 and 11.3).

- For DCT coefficients in Intra mode:

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

- For DCT coefficients in Inter mode:

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

Table 11.2: 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	25	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

Table 11.3: 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 (Cont'd)

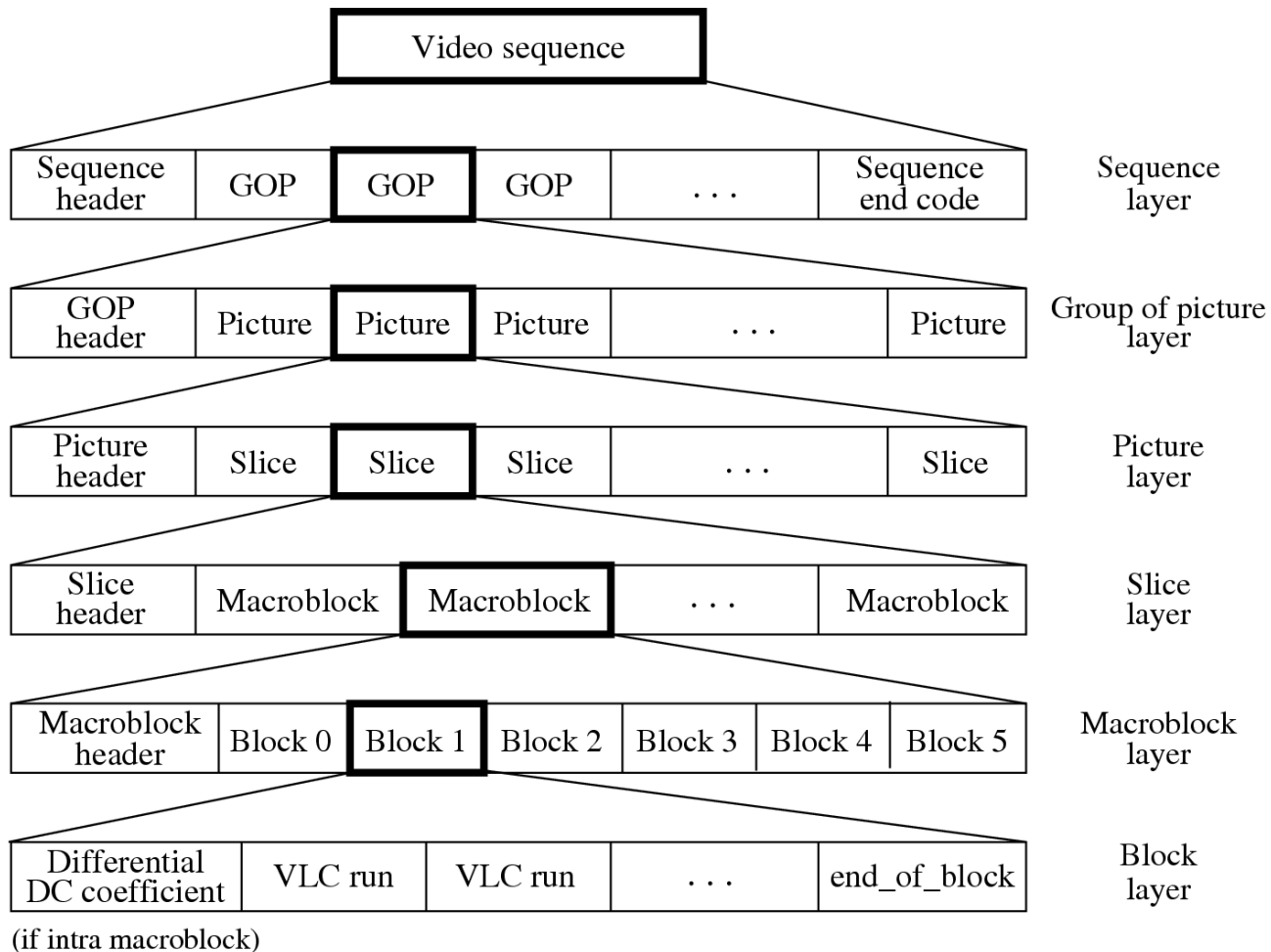
- MPEG-1 allows motion vectors to be of sub-pixel precision (1/2 pixel). The technique of "bilinear interpolation" for H.263 can be used to generate the needed values at half-pixel locations.
- Compared to the maximum range of ± 15 pixels for motion vectors in H.261, MPEG-1 supports a range of $[-512, 511.5]$ for half-pixel precision and $[-1,024, 1,023]$ for full-pixel precision motion vectors.
- The MPEG-1 bitstream allows random access — accomplished by GOP layer in which each GOP is time coded.

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.

□ **Table 11.4: Typical Compression Performance of MPEG-1 Frames**

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



□ Fig 11.5: Layers of MPEG-1 Video Bitstream.

11.3 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 (Table 11.5).
 - The DVD video specification allows only four display resolutions: 720×480, 704×480, 352×480, and 352×240
 - a restricted form of the MPEG-2 Main profile at the Main and Low levels.

□ **Table 11.5:** 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	*	*	*				

Table 11.6: Four Levels in the Main Profile of MPEG-2

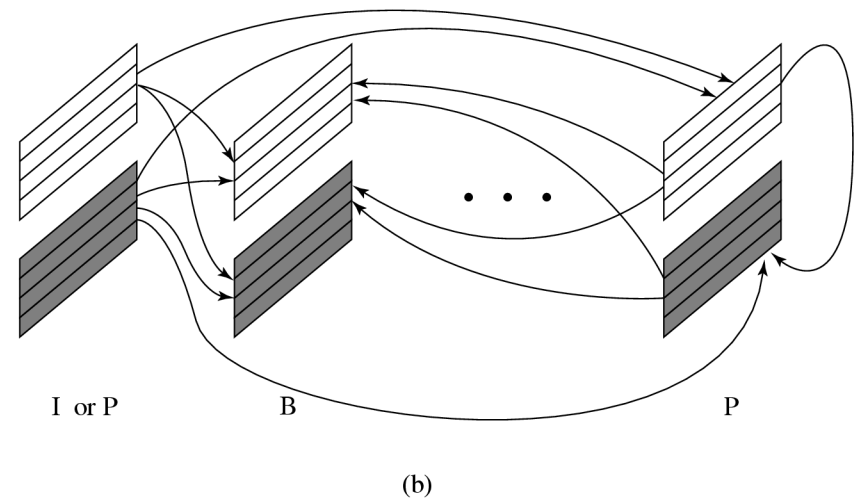
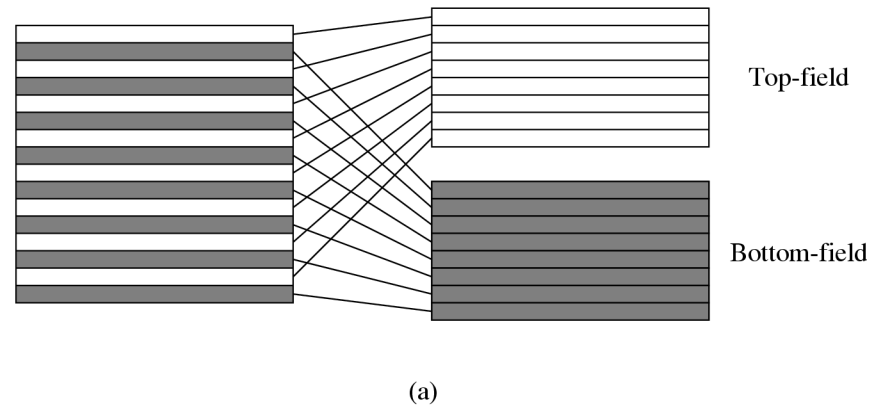
Level	Max. Resolution	Max fps	Max pixels/sec	Max coded Data Rate (Mbps)	Application
High	1920 × 1152	60	62.7 × 10 ⁶	80	film production
High 1440	1440 × 1152	60	47.0 × 10 ⁶	60	consumer HDTV
Main	720 × 576	30	10.4 × 10 ⁶	15	Studio TV
Low	352 × 288	30	3.0 × 10 ⁶	4	consumer tape equiv.

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*.
 - - 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*.

Fig. 11.6: Field pictures and Field-prediction for Field-pictures in MPEG-2.

- (a) Frame-picture vs. Field-pictures
- (b) Field Prediction for Field-pictures



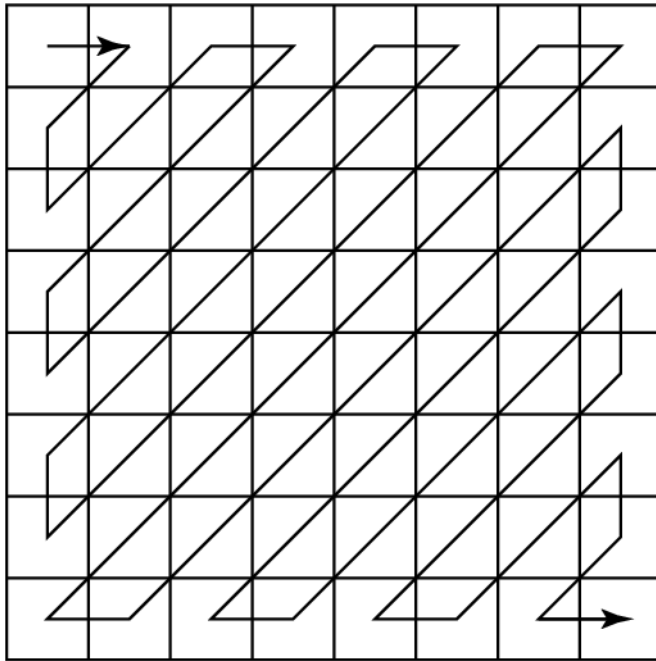
Five Modes of Predictions

- MPEG-2 defines **Frame Prediction** and **Field Prediction** as well as five prediction modes:
 1. **Frame Prediction for Frame-pictures:** Identical to MPEG-1 MC-based prediction methods in both P-frames and B-frames.
 2. **Field Prediction for Field-pictures:** A macroblock size of 16×16 from Field-pictures is used. For details, see Fig. 11.6(b).

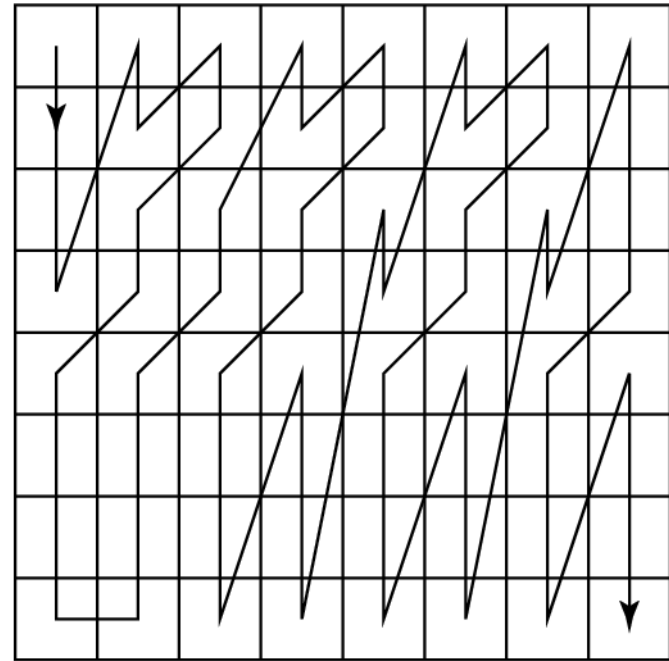
3. **Field Prediction for Frame-pictures**
4. **16×8 MC for Field-pictures**
5. **Dual-Prime for P-pictures**

Alternate Scan and Field DCT

- Techniques aimed at improving the effectiveness of DCT on prediction errors, 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, there exists less correlation between them than between the alternate rows.
 - 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.
- In MPEG-2, **Field_DCT** can also be used to address the same issue.

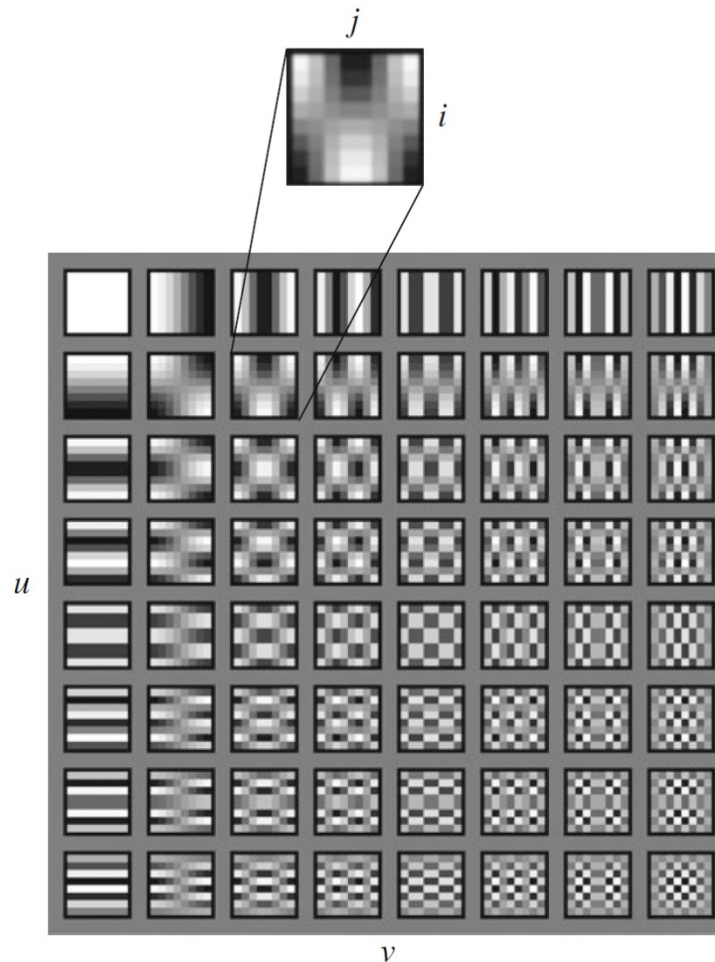


(a)



(b)

- **Fig 11.7:** Zigzag and Alternate Scans of DCT Coefficients for Progressive and Interlaced Videos in MPEG-2.



□ Fig. 8.9: Graphical Illustration of 8×8 2D DCT basis.

MPEG-2 Scalabilities

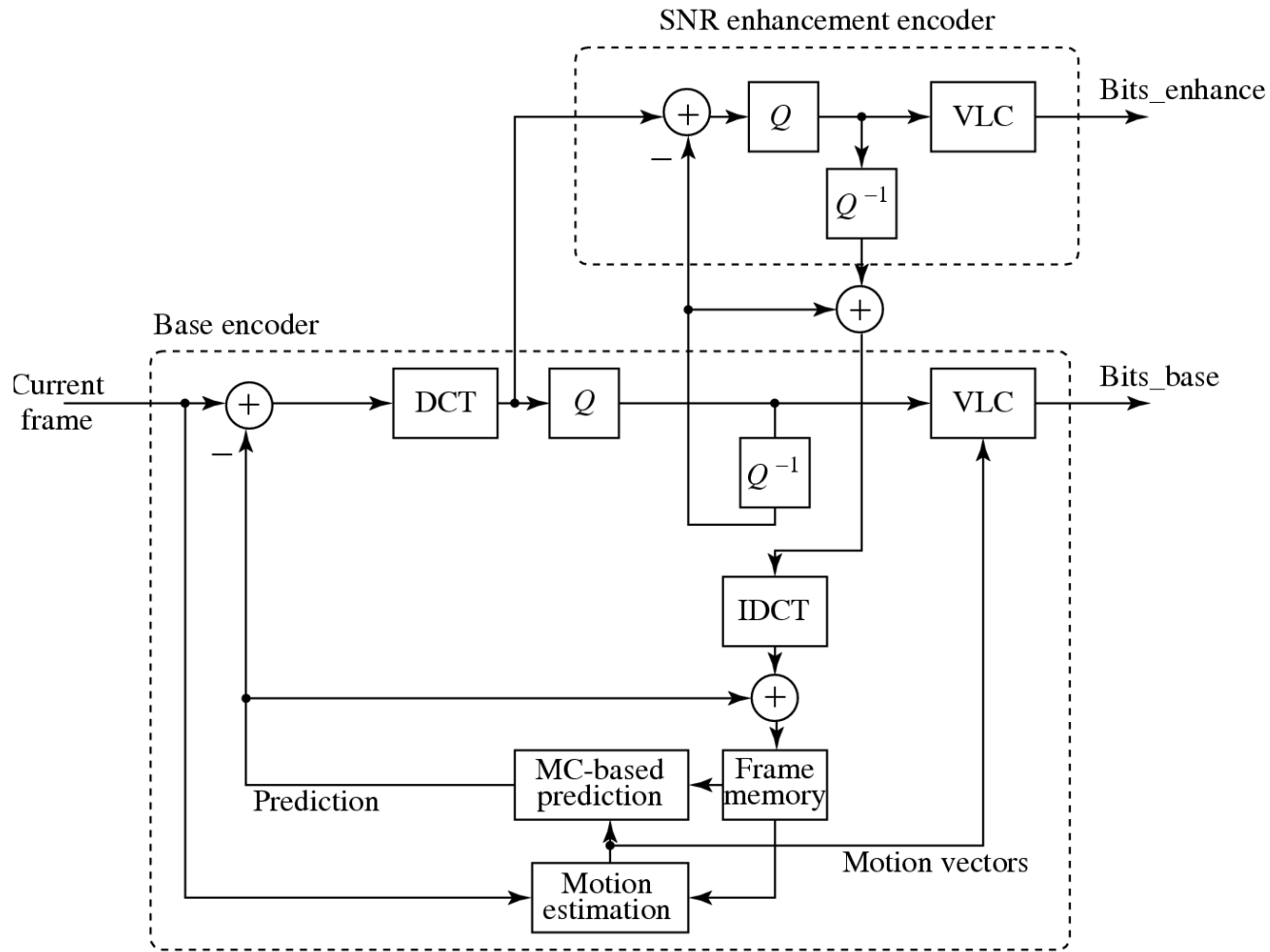
- The MPEG-2 **scalable coding**: A base layer and one or more enhancement layers can be defined — also known as **layered coding**.
 - The base layer can be independently encoded, transmitted and decoded to obtain basic video quality.
 - The encoding and decoding of the enhancement layer is dependent 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 (Cont'd)

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

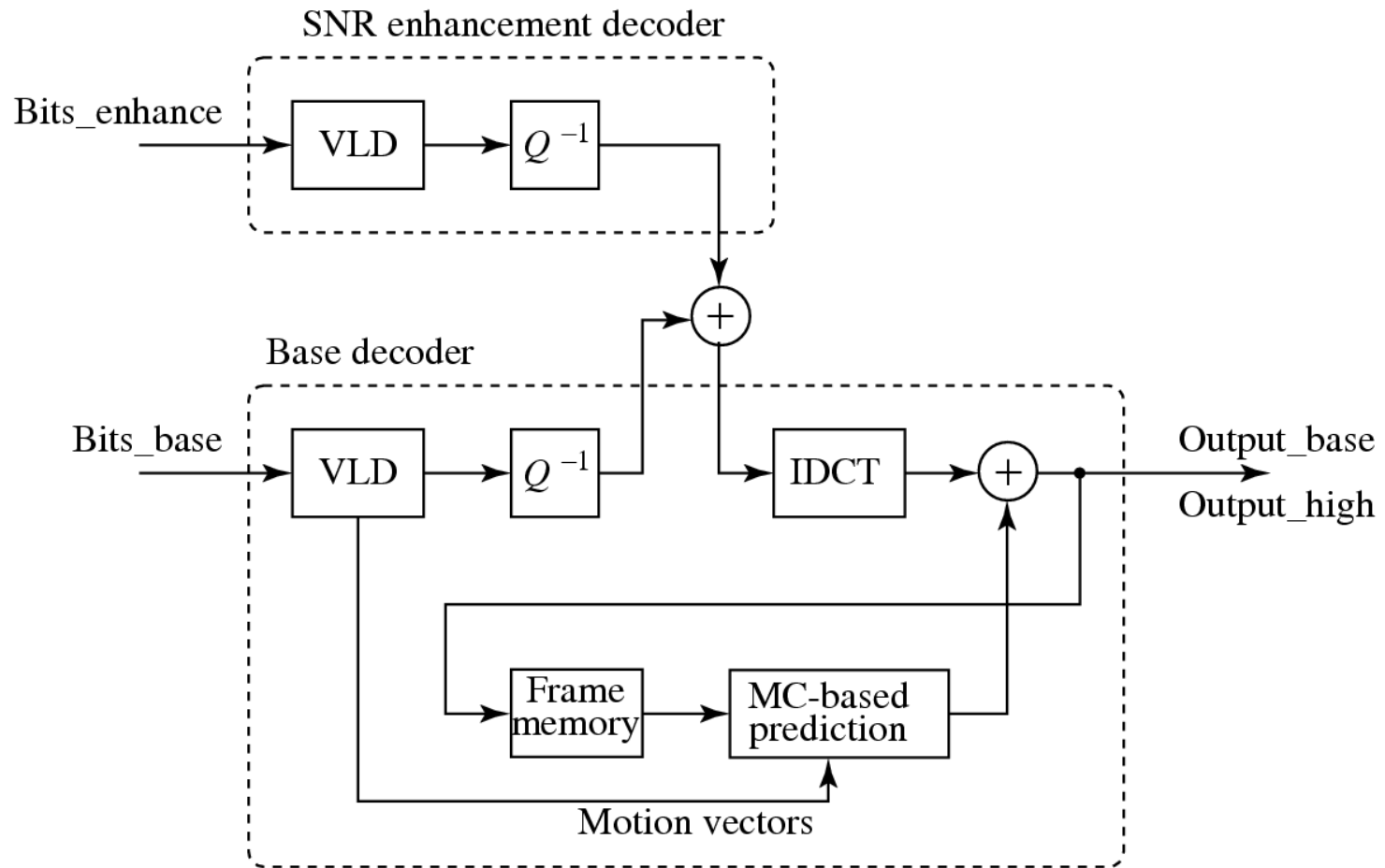
SNR Scalability

- **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*.



(a) Encoder

□ Fig 11.8 (a): MPEG-2 SNR Scalability (Encoder).

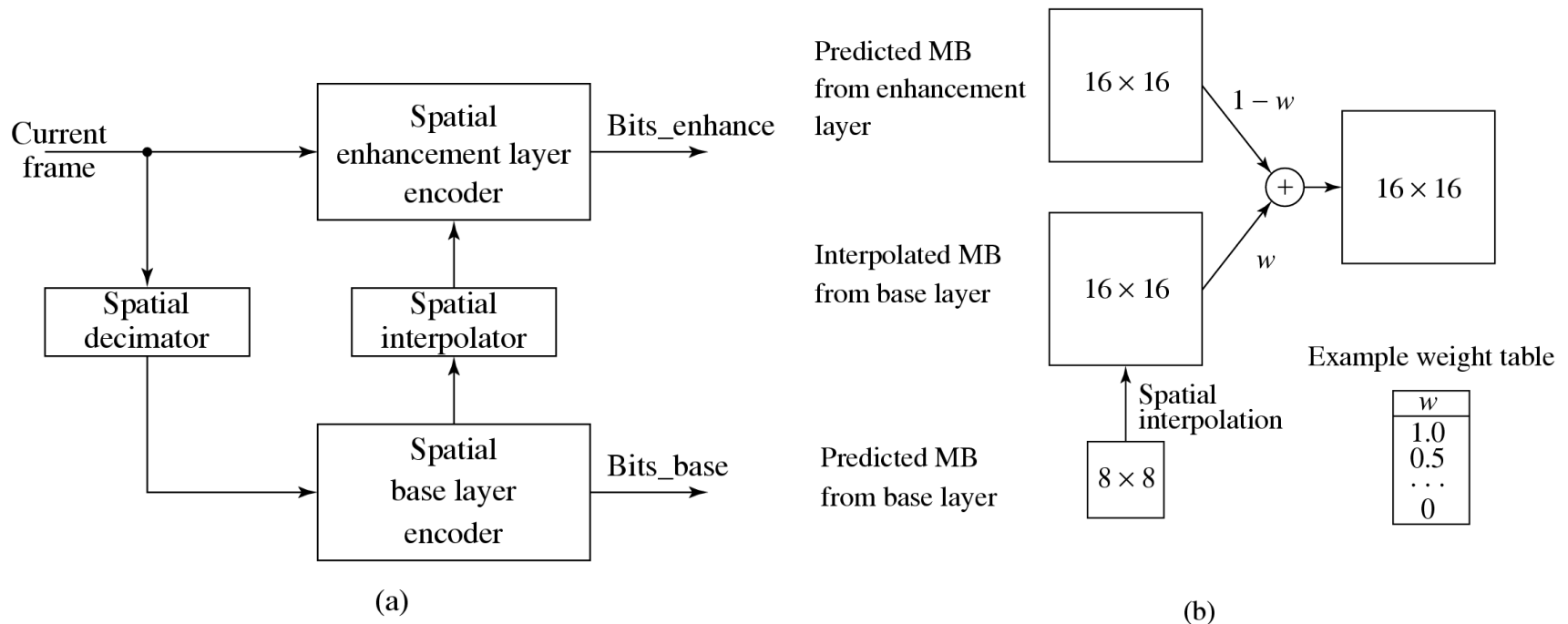


(b) Decoder

□ Fig 11.8 (b): MPEG-2 SNR Scalability (Decoder).

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.
- Fig. 11.9(a) shows a typical block diagram. Fig. 11.9(b) shows a case where temporal and spatial predictions are combined.

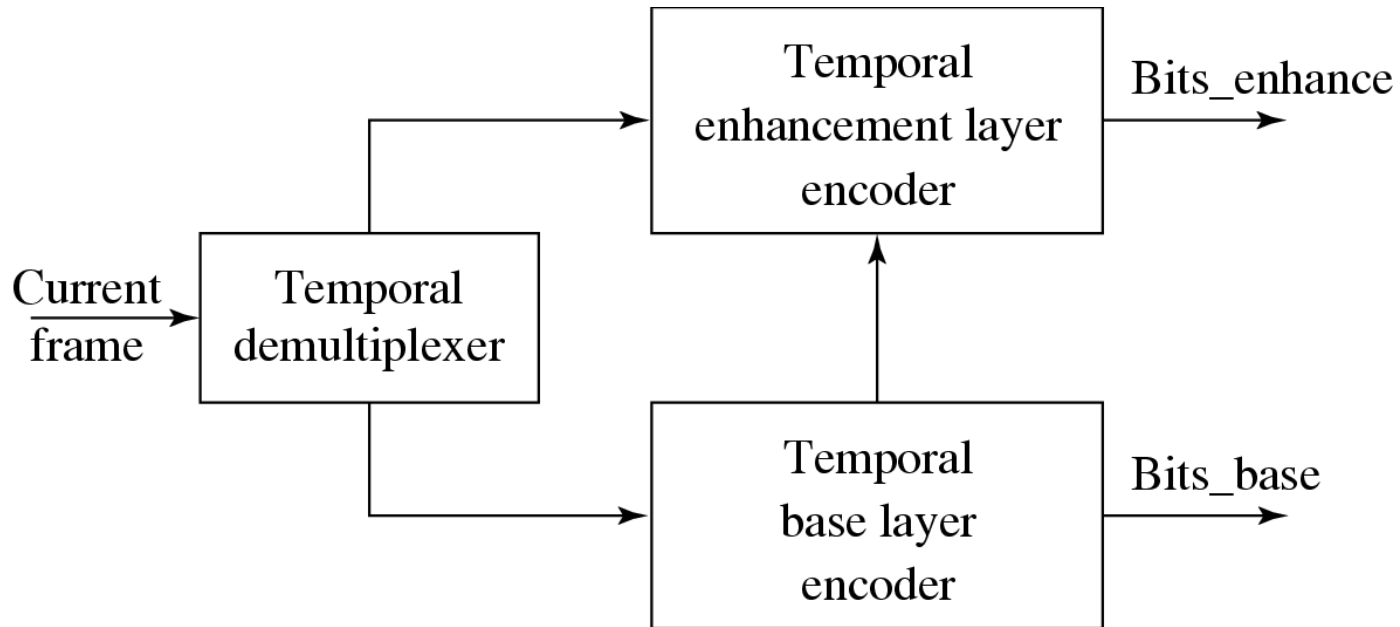


□ Fig. 11.9: Encoder for MPEG-2 Spatial Scalability.

□ (a) Block Diagram. (b) Combining Temporal and Spatial Predictions for Encoding at Enhancement Layer.

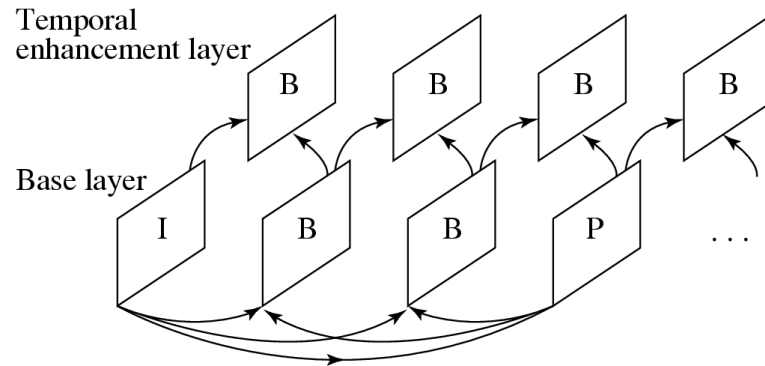
Temporal Scalability

- The input video is temporally demultiplexed into two pieces, each carrying half of the original frame rate.
- Base Layer Encoder carries out the normal single-layer coding procedures for its own input video and yields the output bitstream `Bits_base`.
- The prediction of matching MBs at the Enhancement Layer can be obtained in two ways:
 - Interlayer MC (Motion-Compensated) Prediction (Fig. 11.10(b))
 - Combined MC Prediction and Interlayer MC Prediction (Fig. 11.10(c))

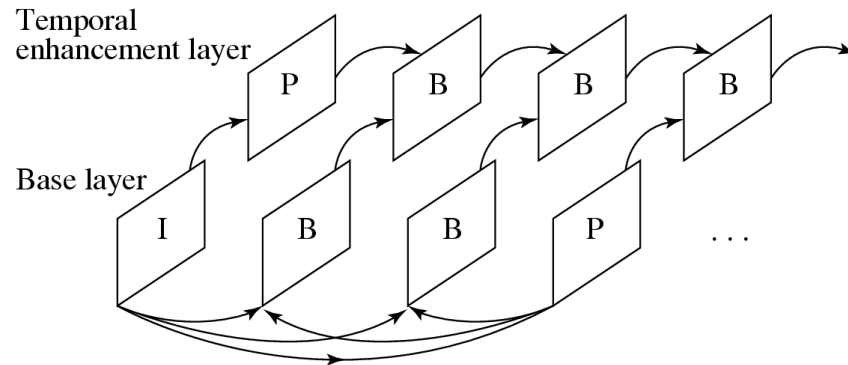


(a) Block Diagram

- Fig 11.10: Encoder for MPEG-2 Temporal Scalability.



(b) Interlayer Motion-Compensated (MC) Prediction



(c) Combined MC Prediction and Interlayer MC Prediction

□ Fig 11.10 (Cont'd): Encoder for MPEG-2 Temporal Scalability

Hybrid Scalability

- Any two of the above three scalabilities can be combined to form hybrid scalability:
 1. Spatial and Temporal Hybrid Scalability.
 2. SNR and Spatial Hybrid Scalability.
 3. SNR and Temporal Hybrid Scalability.
- Usually, a three-layer hybrid coder will be adopted which consists of Base Layer, Enhancement Layer 1, and Enhancement Layer 2.

Data Partitioning

- The *Base partition* contains lower-frequency DCT coefficients, enhancement partition contains high-frequency DCT coefficients.
- Strictly speaking, data partitioning is not layered coding, since a single stream of video data is simply divided up and there is no further dependence on the base partition in generating the enhancement partition.
- Useful for transmission over noisy channels and for progressive transmission.

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.

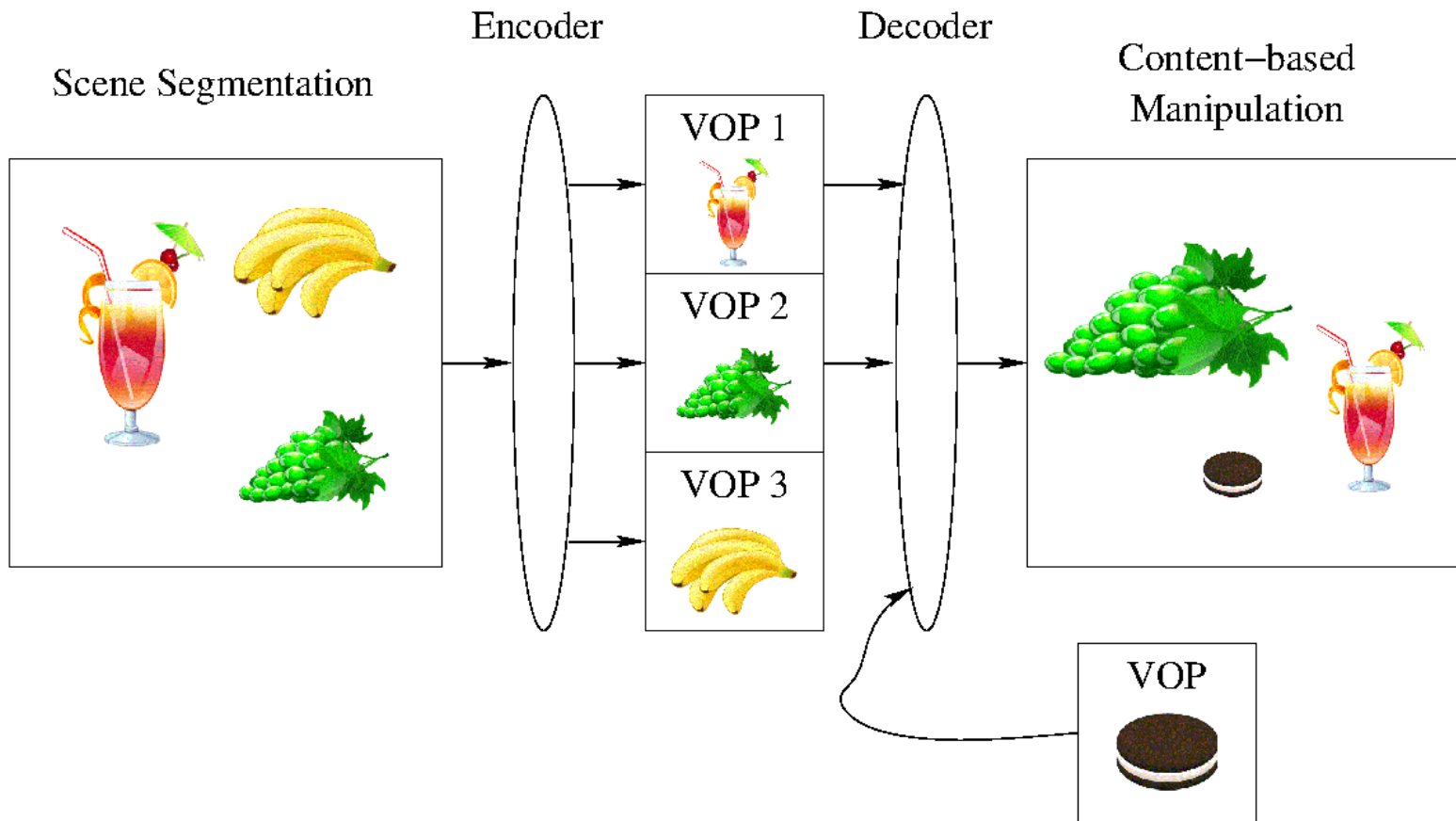
Other Major Differences from MPEG-1 (Cont'd)

- **Nonlinear quantization** — two types of scales are allowed:
 1. For the first type, scale is the same as in MPEG-1 in which it is an integer in the range of [1, 31] and $scale_i = i$.
 2. For the second type, a nonlinear relationship exists, i.e., $scale_i \neq i$. The i th scale value can be looked up from Table 11.7.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$scale_i$	1	2	3	4	5	6	7	8	10	12	14	16	18	20	22	24
i	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
$scale_i$	28	32	36	40	44	48	52	56	64	72	80	88	96	104	112	

Overview of MPEG-4

- **MPEG-4**: a newer standard. Besides compression, pays great attention to issues about user interactivities.
- MPEG-4 departs from its predecessors in adopting a new **object-based coding**:
 - Offering higher compression ratio, also beneficial for digital video composition, manipulation, indexing, and retrieval.
 - Figure 11.11 illustrates how MPEG-4 videos can be composed and manipulated by simple operations on the visual objects.
- The bit-rate for MPEG-4 video now covers a large range between 5 kbps to 10 Mbps.

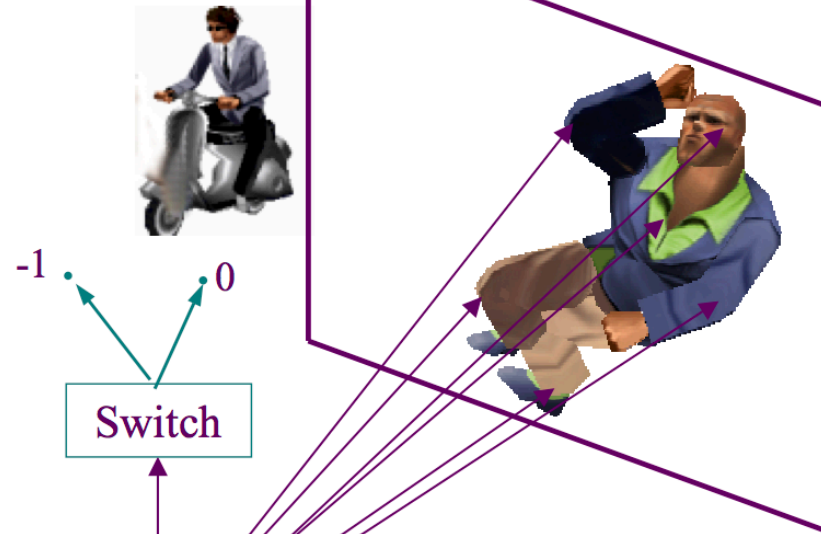
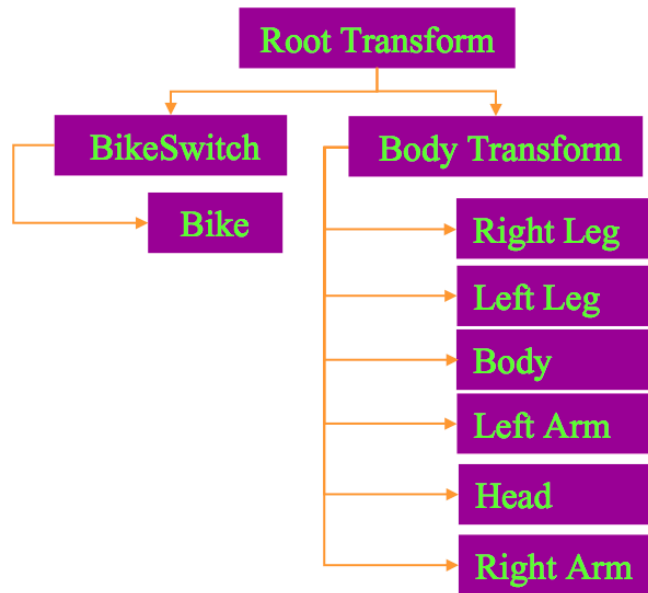


■ Fig. 11.11: Composition and Manipulation of MPEG-4 Videos. (VOP = Video object plane)

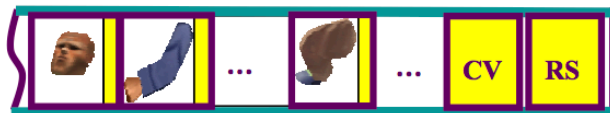


BIFS-Command

Scene Graph



BIFS-Command ES



Overview of MPEG-4 (Cont'd)

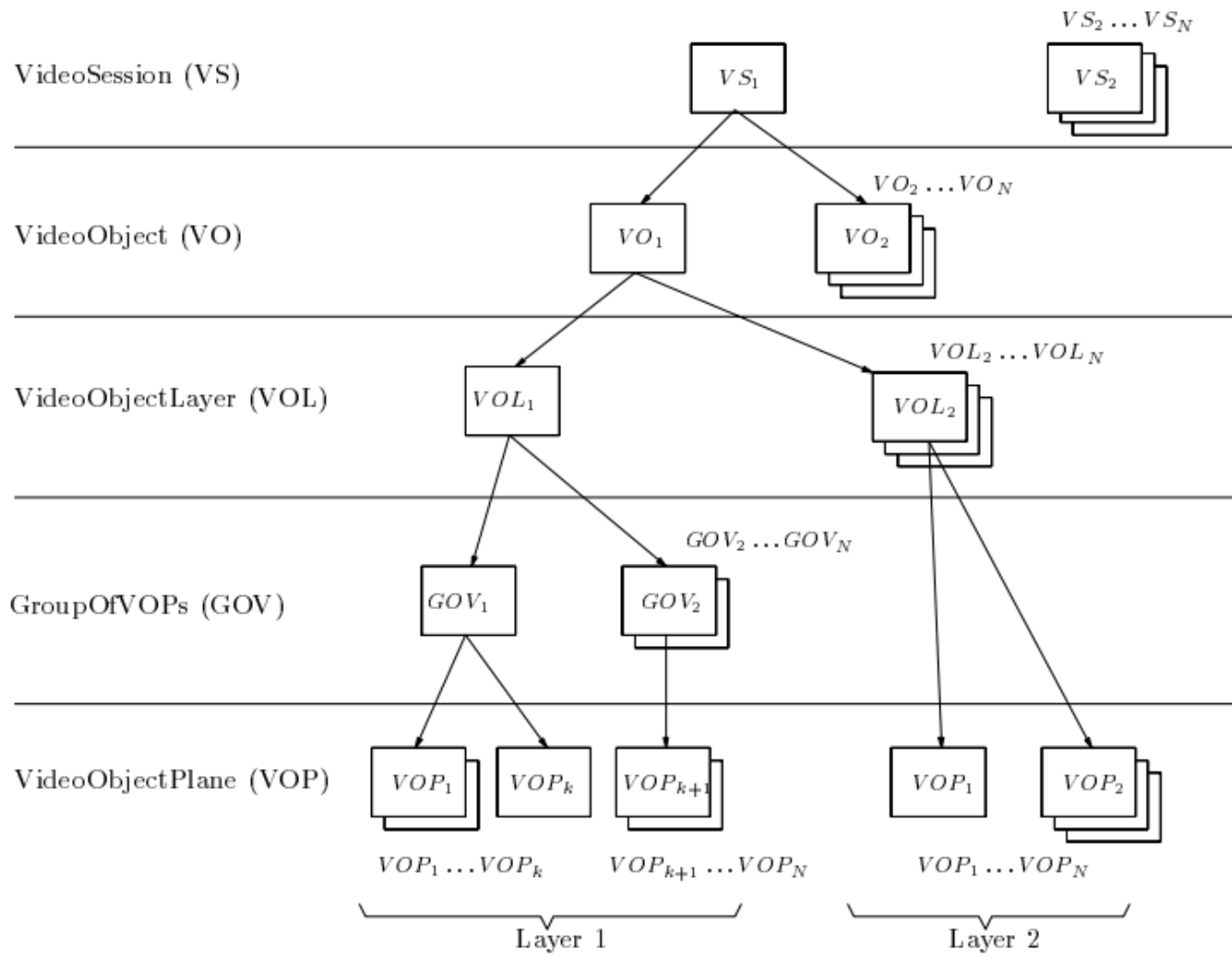
- MPEG-4 (Fig. 11.12(b)) is an entirely new standard for:
 - (a) Composing media objects to create desirable audiovisual scenes.
 - (b) Multiplexing and synchronizing the bitstreams for these media data entities so that they can be transmitted with guaranteed Quality of Service (QoS).
 - (c) Interacting with the audiovisual scene at the receiving end — provides a toolbox of advanced coding modules and algorithms for audio and video compressions.

Overview of MPEG-4 (Cont'd)

- The hierarchical structure of MPEG-4 visual bitstreams is very different from that of MPEG-1 and -2, it is very much video object-oriented.

Video-object Sequence (VS)
Video Object (VO)
Video Object Layer (VOL)
Group of VOPs (GOV)
Video Object Plane (VOP)

Fig. 11.13: Video Object Oriented Hierarchical Description of a Scene in MPEG-4 Visual Bitstreams.



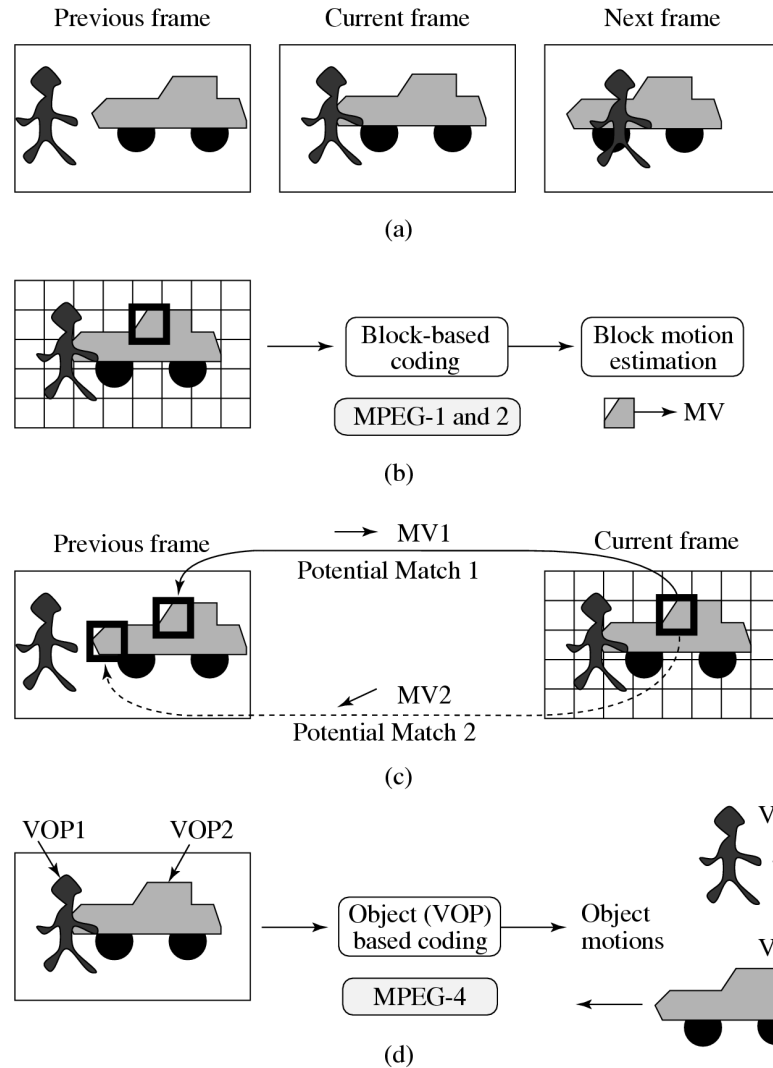
Overview of MPEG-4 (Cont'd)

1. **Video-object Sequence (VS)**—delivers the complete MPEG-4 visual scene, which may contain 2-D or 3-D natural or synthetic objects.
2. **Video Object (VO)** — a particular object in the scene, which can be of arbitrary (non-rectangular) shape corresponding to an object or background of the scene.
3. **Video Object Layer (VOL)** — facilitates a way to support (multi-layered) scalable coding. A VO can have multiple VOLs under scalable coding, or have a single VOL under non-scalable coding.
4. **Group of Video Object Planes (GOV)** — groups Video Object Planes together (optional level).
5. **Video Object Plane (VOP)** — a snapshot of a VO at a particular moment.

11.4.2 Video Object-based Coding in MPEG-4

VOP-based vs. Frame-based Coding

- MPEG-1 and -2 do not support the VOP concept, and hence their coding method is referred to as **frame-based** (also known as **Block-based coding**).
- Fig. 11.14 (c) illustrates a possible example in which both potential matches yield small prediction errors for block-based coding.
- Fig. 11.14 (d) shows that each VOP is of arbitrary shape and ideally will obtain a unique motion vector consistent with the actual object motion.



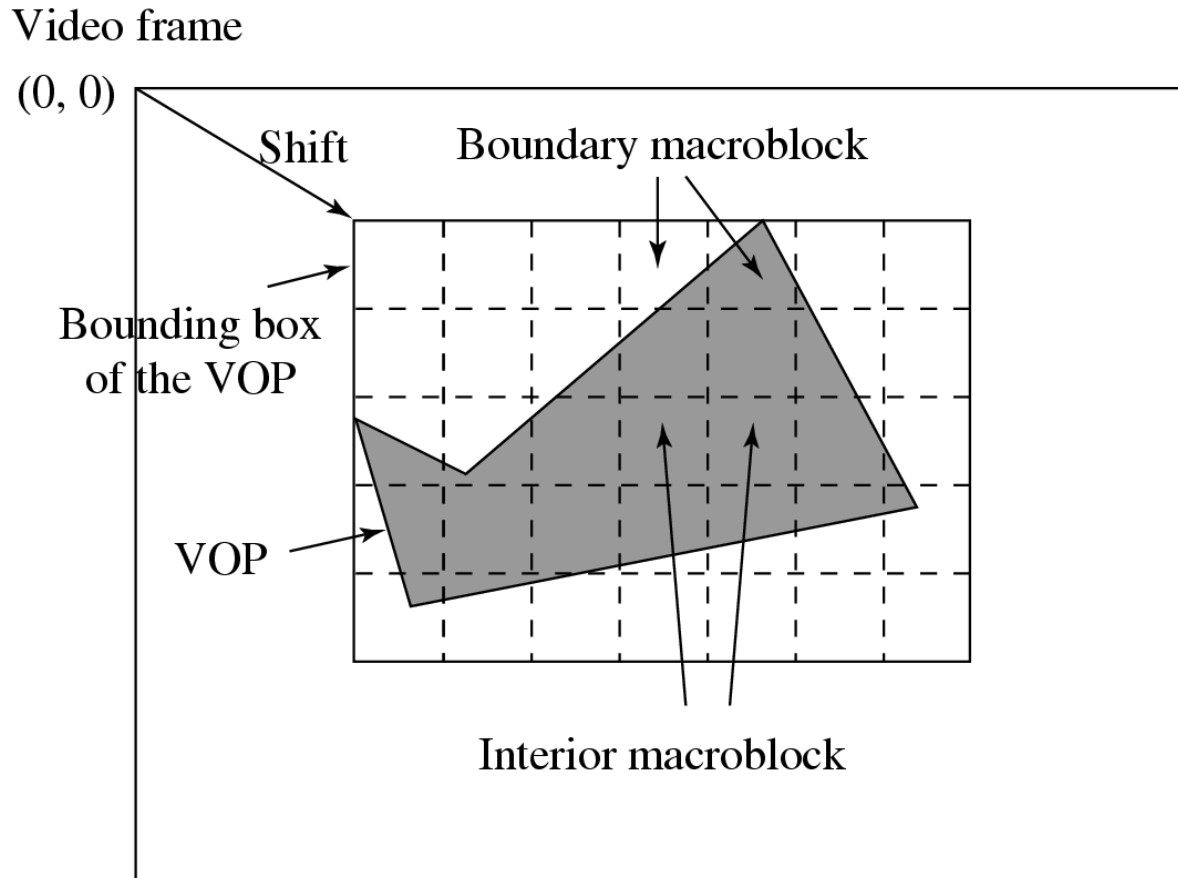
□ Fig. 11.14: Comparison between Block-based Coding and Object-based Coding.

VOP-based Coding

- MPEG-4 VOP-based coding also employs the Motion Compensation technique:
 - An Intra-frame coded VOP is called an **I-VOP**.
 - The Inter-frame coded VOPs are called *P-VOPs* if only forward prediction is employed, or *B-VOPs* if bi-directional predictions are employed.
 - The new difficulty for VOPs: may have arbitrary shapes, shape information must be coded in addition to the texture of the VOP.
- Note: *texture* here actually refers to the visual content, that is the gray-level (or chroma) values of the pixels in the VOP.

VOP-based Motion Compensation (MC)

- MC-based VOP coding in MPEG-4 again involves three steps:
 - (a) Motion Estimation.
 - (b) MC-based Prediction.
 - (c) Coding of the prediction error.
- Only pixels within the VOP of the current (Target) VOP are considered for matching in MC.
- To facilitate MC, each VOP is divided into many macroblocks (MBs). MBs are by default 16×16 in luminance images and 8×8 in chrominance images.



□ Fig. 11.15: Bounding Box and Boundary Macroblocks of VOP.

Texture Coding

- Texture coding in MPEG-4 can be based on:
 - DCT or
 - Shape Adaptive DCT (SA-DCT).

I. Texture coding based on DCT

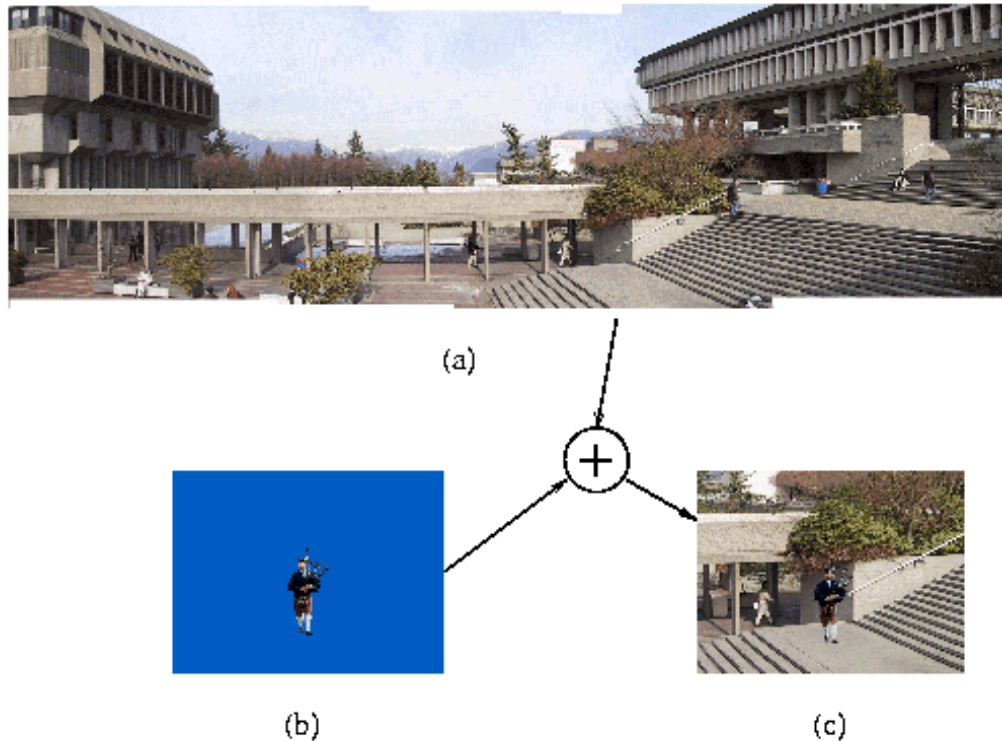
- In I-VOP, the gray values of the pixels in each MB of the VOP are directly coded using the DCT followed by VLC, similar to what is done in JPEG.
- In P-VOP or B-VOP, MC-based coding is employed — it is the prediction error that is sent to DCT and VLC.

Shape Coding

- MPEG-4 supports two types of shape information, **binary** and **gray scale**.
- Binary shape information can be in the form of a binary map (also known as *binary alpha map*) that is of the size as the rectangular bounding box of the VOP.
- A value '1' (opaque) or '0' (transparent) in the bitmap indicates whether the pixel is inside or outside the VOP.
- Alternatively, the gray-scale shape information actually refers to the *transparency* of the shape, with gray values ranging from 0 (completely transparent) to 255 (opaque).

Sprite Coding

- A **sprite** is a graphic image that can freely move around within a larger graphic image or a set of images.
- To separate the foreground object from the background, we introduce the notion of a **sprite panorama**: a still image that describes the static background over a sequence of video frames.
 - The large sprite panoramic image can be encoded and sent to the decoder only once at the beginning of the video sequence.
 - When the decoder receives separately coded foreground objects and parameters describing the camera movements thus far, it can reconstruct the scene in an efficient manner.
 - Fig. 12.10 shows a sprite which is a panoramic image stitched from a sequence of video frames.



□ **Fig. 11.20:** Sprite Coding. (a) The sprite panoramic image of the background, (b) the foreground object (piper) in a blue-screen image, (c) the composed video scene.

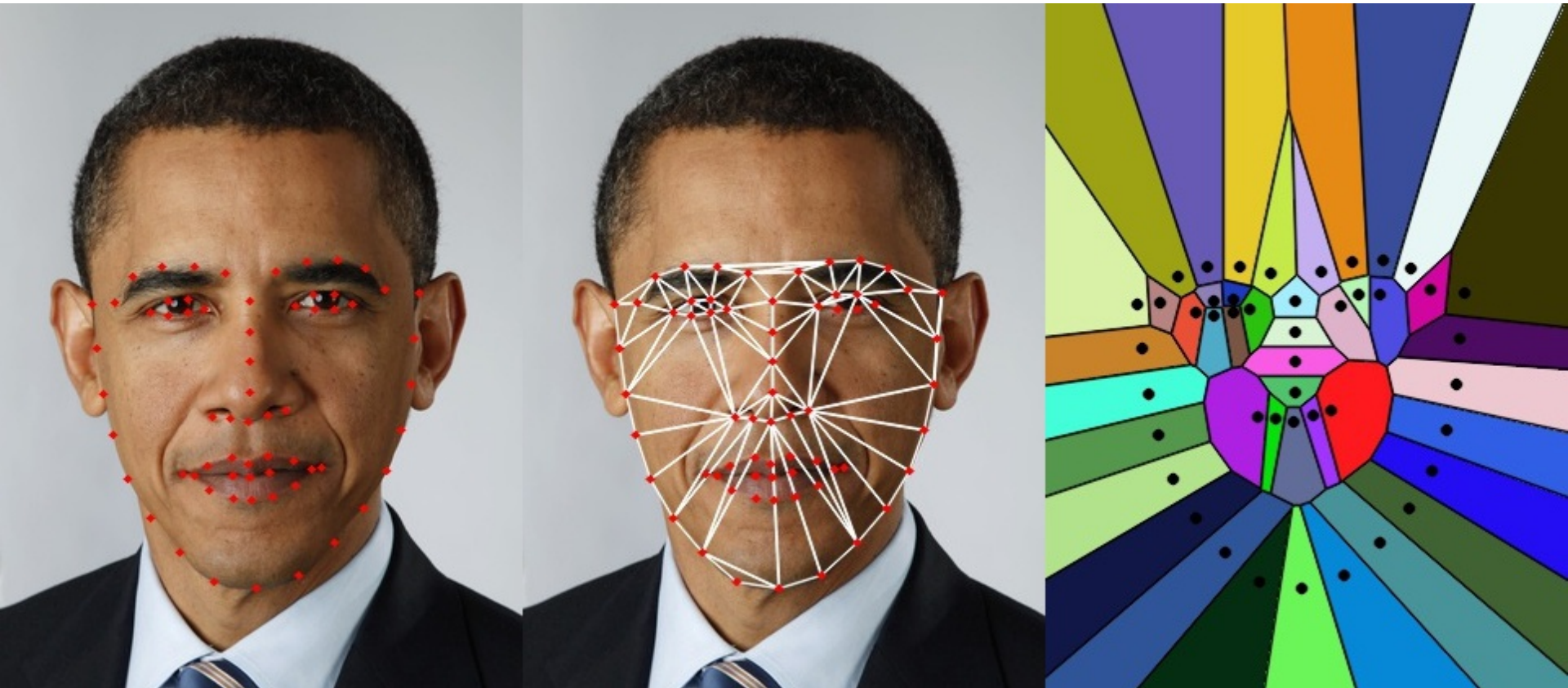
□ *Piper image courtesy of Simon Fraser University Pipe Band.*

Global Motion Compensation (GMC)

- “Global” - overall change due to camera motions (pan, tilt, rotation and zoom)
- Without GMC this will cause a large number of significant motion vectors
- There are four major components within the GMC algorithm:
 - Global motion estimation
 - Warping and blending
 - Motion trajectory coding
 - Choice of LMC (Local Motion Compensation) or GMC.

11.4.3 Synthetic Object Coding in MPEG-4

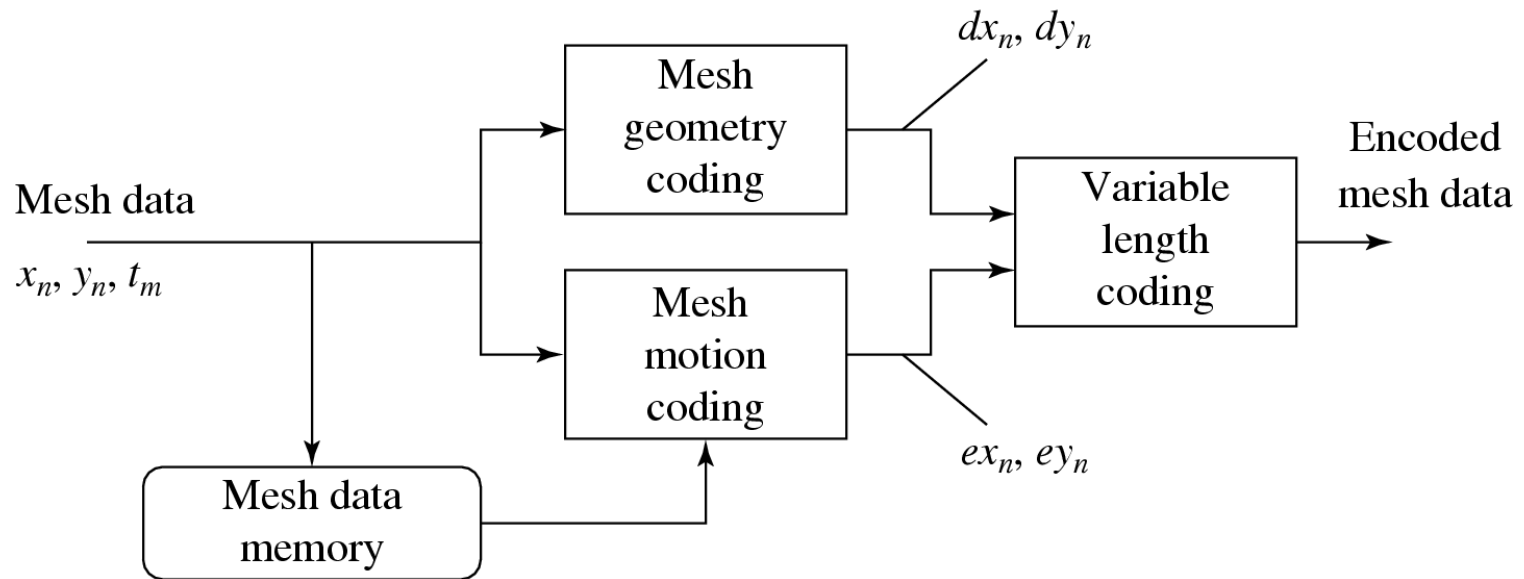
2D Mesh Object Coding



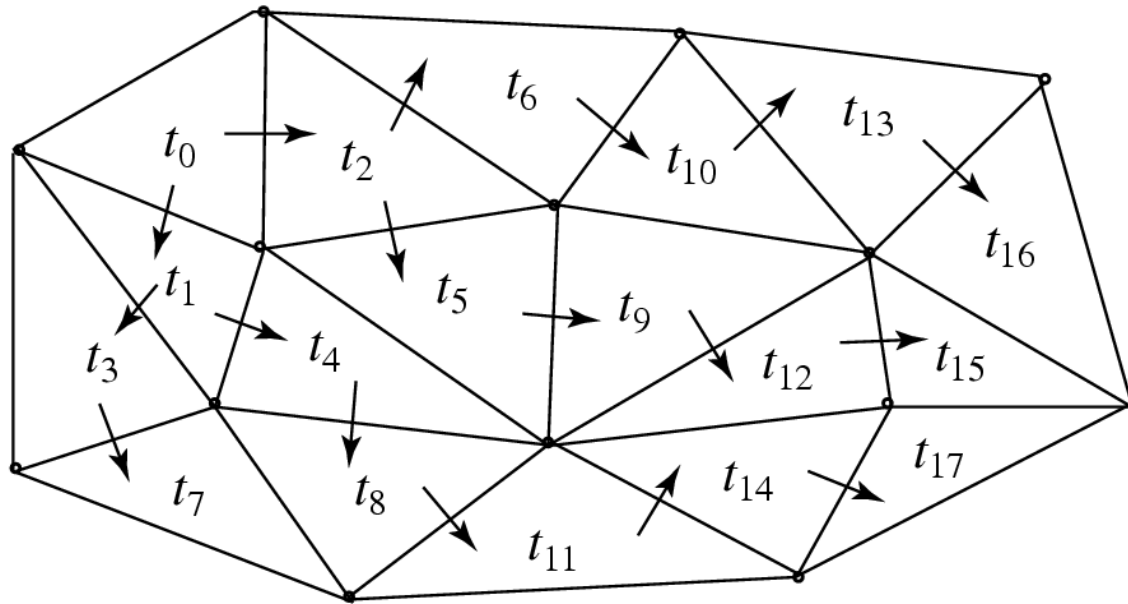
11.4.3 Synthetic Object Coding in MPEG-4

2D Mesh Object Coding

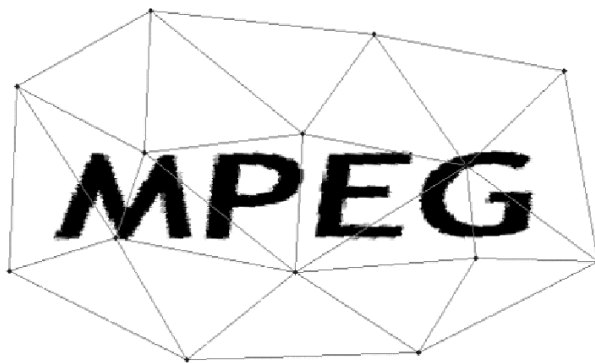
- **2D mesh:** a tessellation (or partition) of a 2D planar region using polygonal patches:
 - The vertices of the polygons are referred to as *nodes* of the mesh.
 - The most popular meshes are *triangular meshes* where all polygons are triangles.
 - The MPEG-4 standard makes use of two types of 2D mesh: **uniform mesh** and **Delaunay mesh**
 - 2D mesh object coding is compact. All coordinate values of the mesh are coded in half-pixel precision.
 - Each 2D mesh is treated as a *mesh object plane (MOP)*.



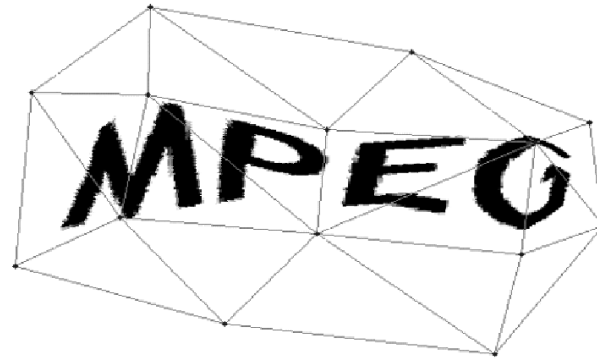
□ Fig. 11.21: 2D Mesh Object Plane (MOP) Encoding Process



□ Fig. 11.24: A breadth-first order of MOP triangles for 2D mesh motion coding.



(a)



(b)

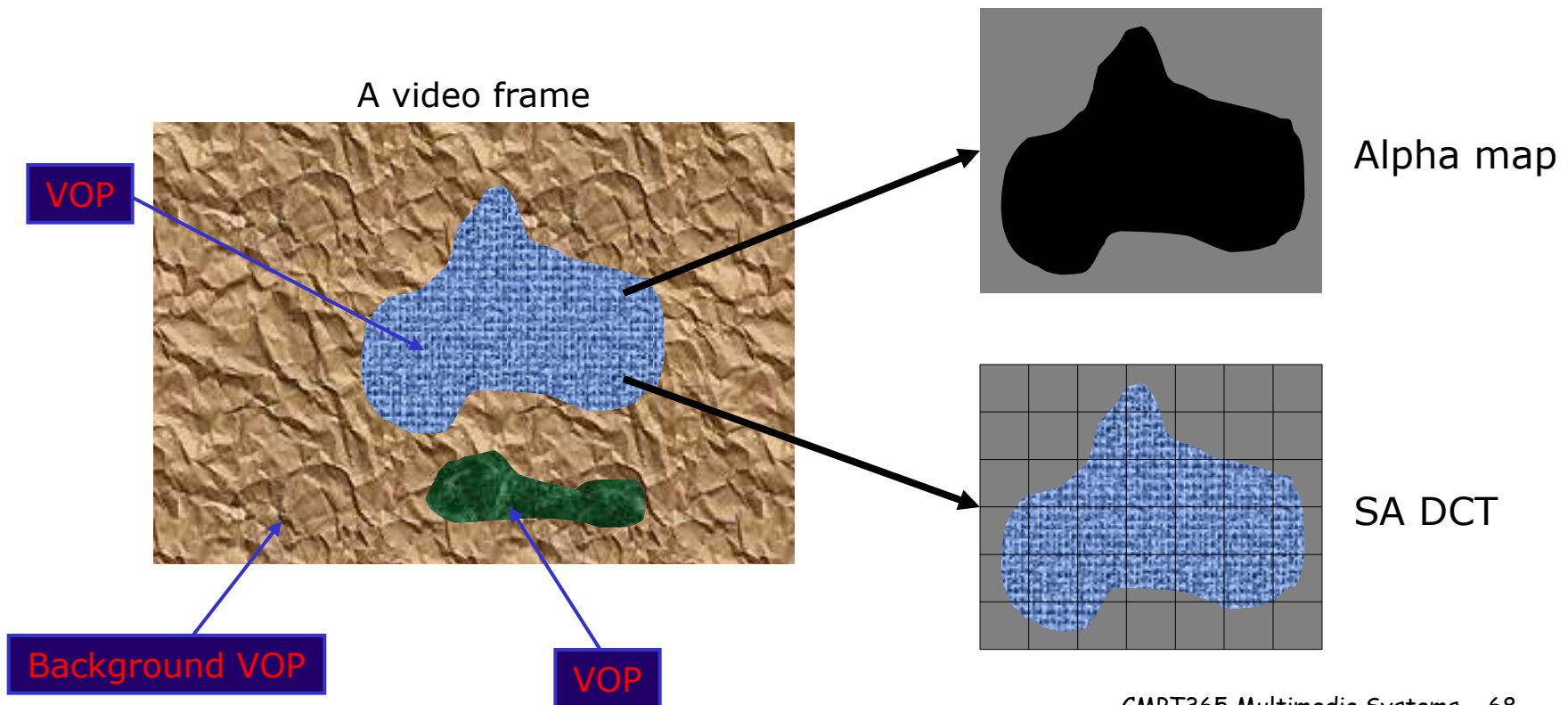
- Fig. 11.25: Mesh-based texture mapping for 2D object animation.

3D Model-Based Coding

- MPEG-4 has defined special 3D models for **face objects** and **body objects** because of the frequent appearances of human faces and bodies in videos.
- Some of the potential applications for these new video objects include teleconferencing, human-computer interfaces, games, and e-commerce.
- MPEG-4 goes beyond wireframes so that the surfaces of the face or body objects can be shaded or texture-mapped.

MPEG-4 Example

- ❑ Instead of "frames": Video Object Planes
- ❑ Shape Adaptive DCT



Example



Example

