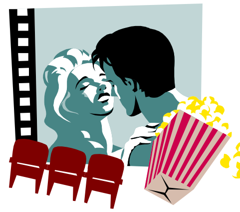
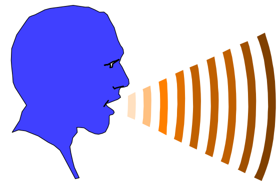





Video Processing & Communications

Video Coding Standards

Yao Wang
Polytechnic University, Brooklyn, NY11201
yao@vision.poly.edu

Based on: Y. Wang, J. Ostermann, and Y.-Q. Zhang, Video Processing and Communications, Prentice Hall, 2002.

Outline

- Overview of Standards and Their Applications
- ITU-T Standards for Audio-Visual Communications
  - H.261
  - H.263
  - H.263+, H.263++
- ISO Standards for
  - MPEG-1
  - MPEG-2
  - MPEG-4
  - MPEG-7

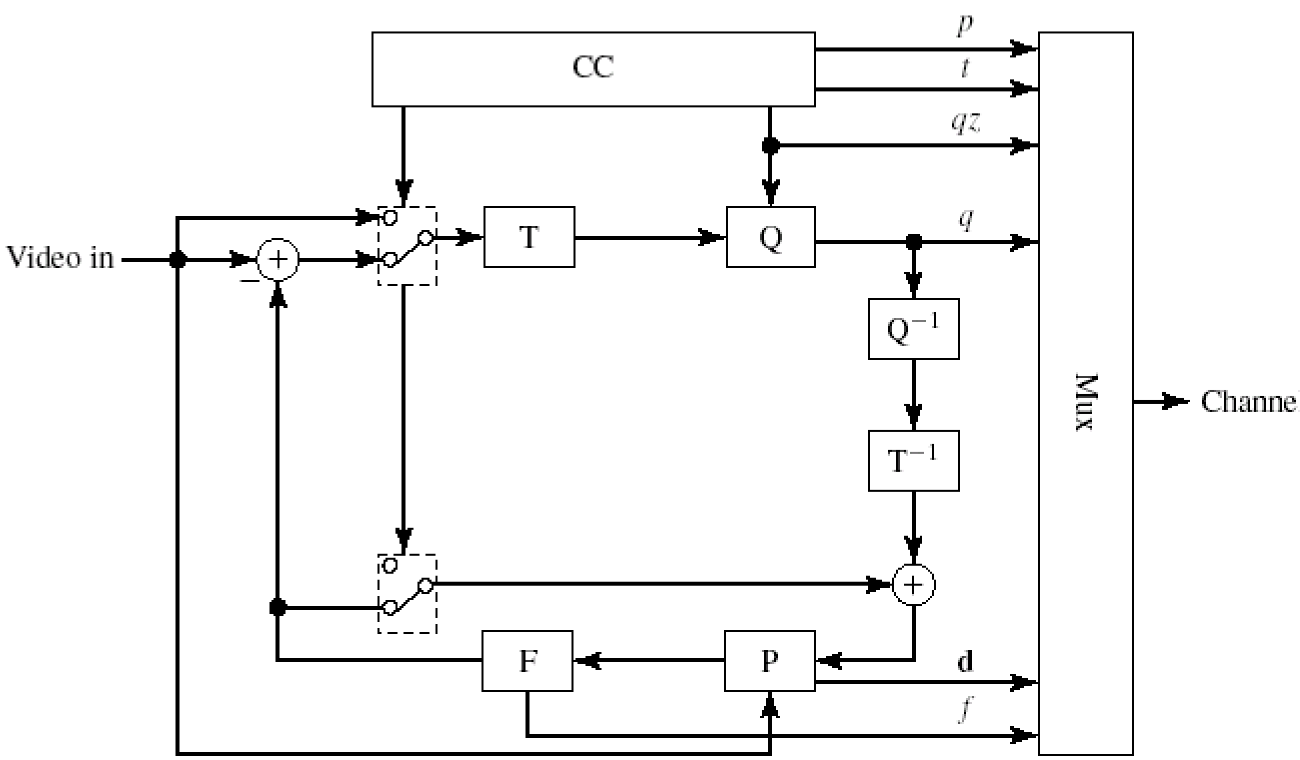
Multimedia Communications Standards and Applications

Standards	Application	Video Format	Raw Data Rate	Compressed Data Rate
H.320 (H.261)	Video conferencing over ISDN	CIF QCIF	37 Mbps 9.1 Mbps	>=384 Kbps >=64 Kbps
H.323 (H.263)	Video conferencing over Internet	4CIF/ CIF/ QCIF		>=64 Kbps
H.324 (H.263)	Video over phone lines/ wireless	QCIF	9.1 Mbps	>=18 Kbps
MPEG-1	Video distribution on CD/ WWW	CIF	30 Mbps	1.5 Mbps
MPEG-2	Video distribution on DVD / digital TV	CCIR601 4:2:0	128 Mbps	3-10 Mbps
MPEG-4	Multimedia distribution over Inter/Intra net	QCIF/CIF		28-1024 Kbps
GA-HDTV	HDTV broadcasting	SMPTE296/295	<=700 Mbps	18--45 Mbps
MPEG-7	Multimedia databases (content description and retrieval)			

H.261 Video Coding Standard

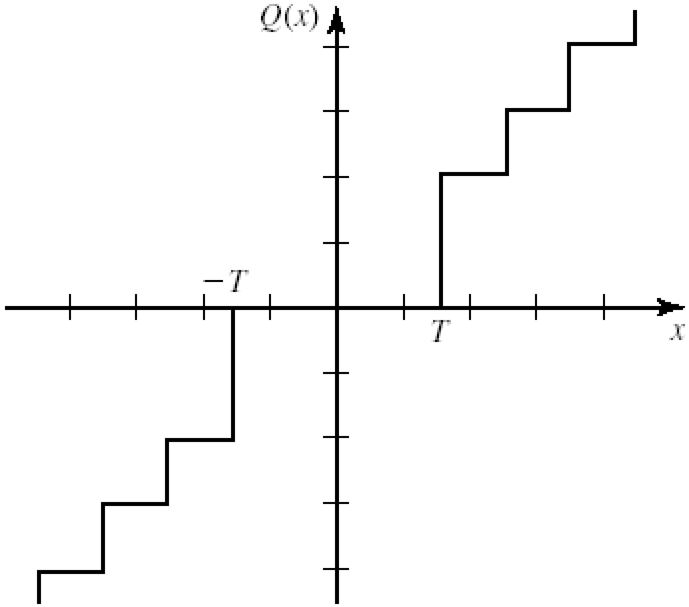
- For video-conferencing/video phone
  - Video coding standard in H.320
  - Low delay (real-time, interactive)
  - Slow motion in general
- For transmission over ISDN
  - Fixed bandwidth: px64 Kbps, p=1,2,…,30
- Video Format:
  - CIF (352x288, above 128 Kbps)
  - QCIF (176x144, 64-128 Kbps)
  - 4:2:0 color format, progressive scan
- Published in 1990
- Each macroblock can be coded in intra- or inter-mode
- Periodic insertion of intra-mode to eliminate error propagation due to network impairments
- Integer-pel accuracy motion estimation in inter-mode

H.261 Encoder



F: Loop filter; P: motion estimation and compensation

DCT Coefficient Quantization



DC Coefficient in Intra-mode:
Uniform, stepsize=8

Others:
Uniform with deadzone,
stepsize=2~64 (MQUANT)

Deadzone:
To avoid too many small coefficients being coded, which are typically due to noise

Motion Estimation and Compensation

- Integer-pel accuracy in the range [-16,16]
- Methods for generating the MVs are not specified in the standard
  - Standards only define the bitstream syntax, or the decoder operation)
- MVs coded differentially (DMV)
- Encoder and decoder uses the decoded MVs to perform motion compensation
- Loop-filtering can be applied to suppress propagation of coding noise temporally
  - Separable filter [1/4,1/2,1/4]
  - Loop filter can be turned on or off

Variable Length Coding

- DCT coefficients are converted into runlength representations and then coded using VLC (Huffman coding for each pair of symbols)
  - Symbol: (Zero run-length, non-zero value range)
- Other information are also coded using VLC (Huffman coding)

Parameter Selection and Rate Control

- MTYPE (intra vs. inter, zero vs. non-zero MV in inter)
- CBP (which blocks in a MB have non-zero DCT coefficients)
- MQUANT (allow the changes of the quantizer stepsize at the MB level)
  - should be varied to satisfy the rate constraint
- MV (ideally should be determined not only by prediction error but also the total bits used for coding MV and DCT coefficients of prediction error)
- Loop Filter on/off

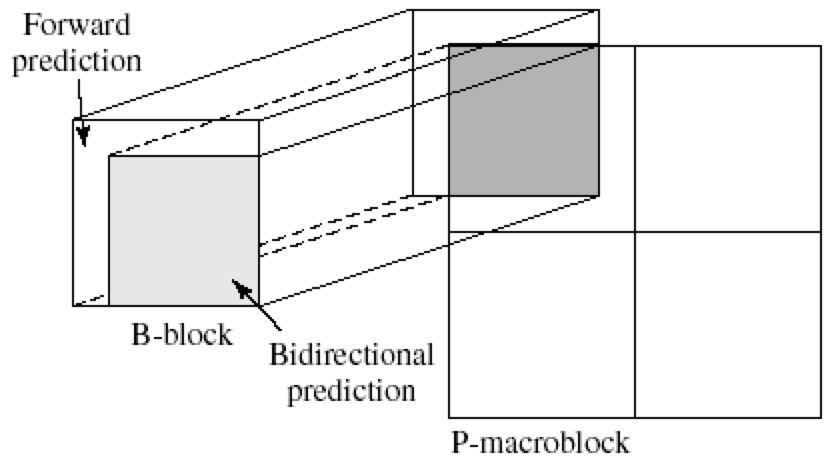
H.263 Video Coding Standard

- H.263 is the video coding standard in H.323/H.324, targeted for visual telephone over PSTN or Internet
- Developed later than H.261, can accommodate computationally more intensive options
  - Initial version (H.263 baseline): 1995
  - H.263+: 1997
  - H.263++: 2000
- Goal: Improved quality at lower rates
- Result: Significantly better quality at lower rates
  - Better video at 18-24 Kbps than H.261 at 64 Kbps
  - Enable video phone over regular phone lines (28.8 Kbps) or wireless modem

Improvements over H.261

- Better motion estimation
  - half-pel accuracy motion estimation with bilinear interpolation filter
  - Larger motion search range [-31.5,31], and unrestricted MV at boundary blocks
  - More efficient predictive coding for MVs (median prediction using three neighbors)
  - overlapping block motion compensation (option)
  - variable block size: 16x16 -> 8x8, 4 MVs per MB (option)
  - use bidirectional temporal prediction (PB picture) (option)
- 3-D VLC for DCT coefficients
  - (runlength, value, EOB)
- Syntax-based arithmetic coding (option)
  - 4% savings in bit rate for P-mode, 10% saving for I-mode, at 50% more computations
- The options, when chosen properly, can improve the PSNR 0.5-1.5 dB over default at 20-70 kbps range.

PB-Picture Mode

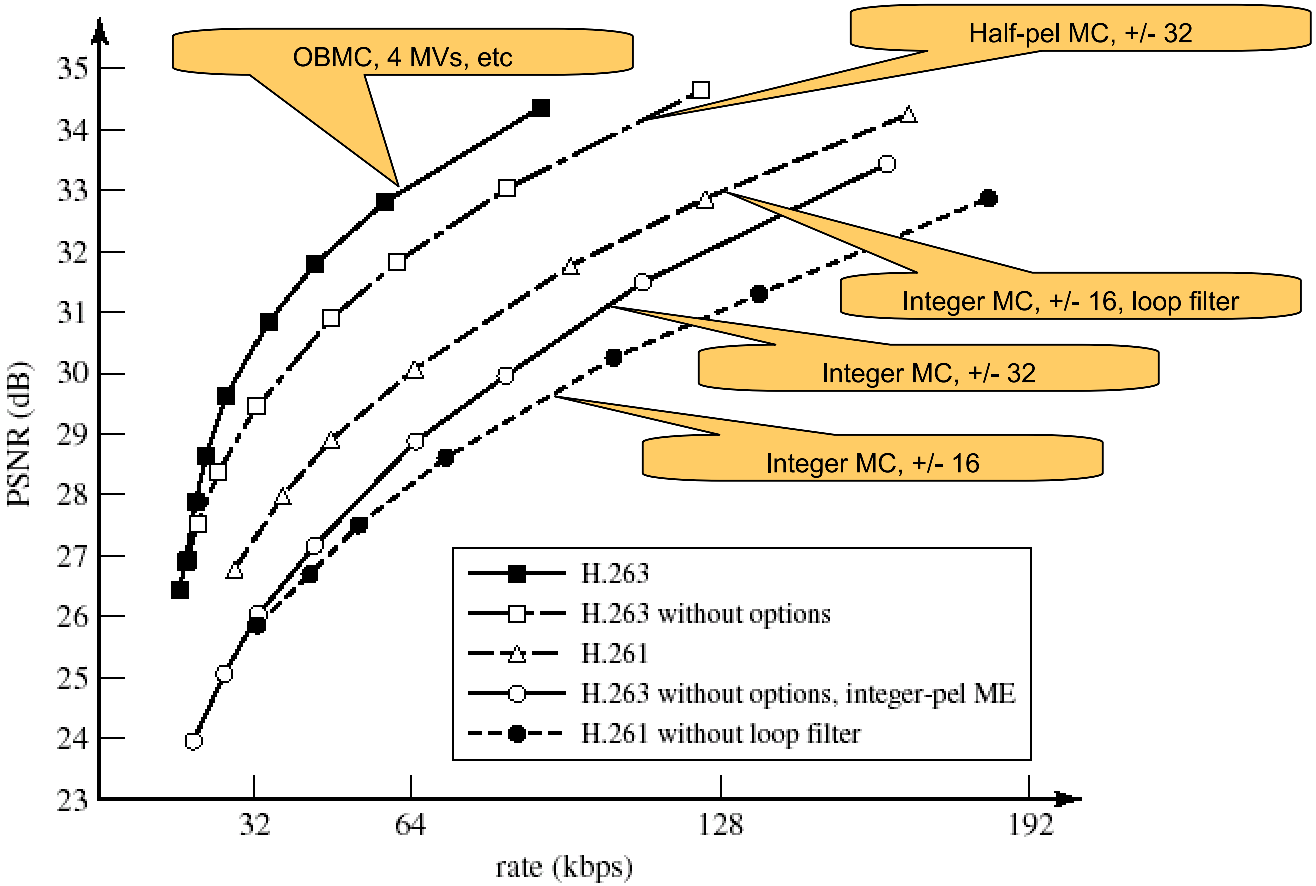


PB-picture mode codes two pictures as a group. The second picture (P) is coded first, then the first picture (B) is coded using both the P-picture and the previously coded picture. This is to avoid the reordering of pictures required in the normal B-mode. But it still requires additional coding delay than P-frames only.

In a B-block, forward prediction (predicted from the previous frame) can be used for all pixels; backward prediction (from the future frame) is only used for those pels that the backward motion vector aligns with pels of the current MB. Pixels in the “white area” use only forward prediction.

An improved PB-frame mode was defined in H.263+, that removes the previous restriction.

Performance of H.261 and H.263

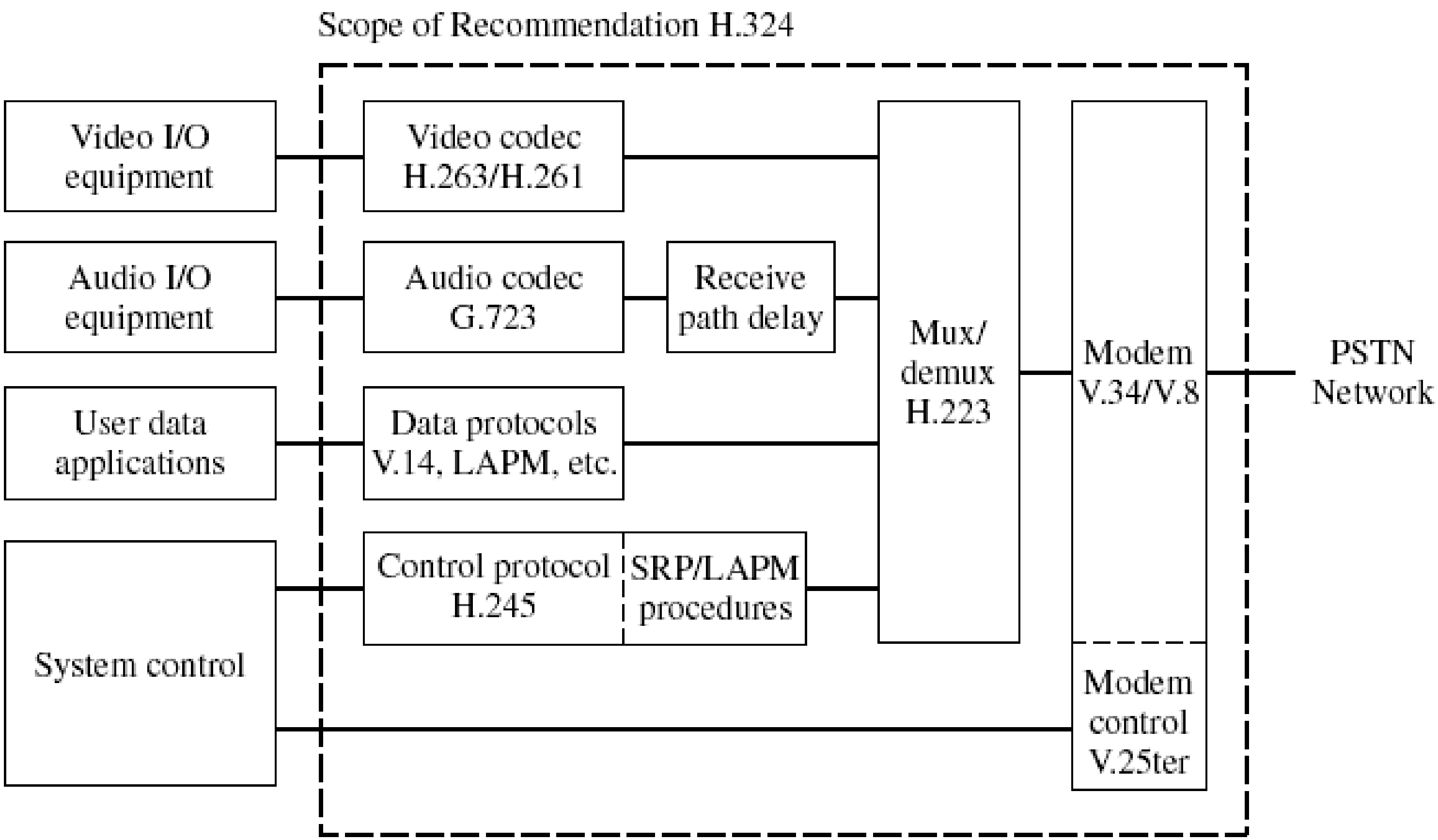


Forman, QCIF, 12.5 Hz

ITU-T Multimedia Communications Standards

Network	System	Video	Audio	Mux	Control
PSTN	H.324	H.261/3	G.723.1	H.223	H.245
N-ISDN	H.320	H.261	G.7xx	H.221	H.242
B-ISDN/ATM	H.321	H.261	G.7xx	H.221	Q.2931
	H.310	H.261/2	G.7xx/MPEG	H.222.0/1	H.245
QoS LAN	H.322	H.261/3	G.7xx	H.221	H.242
Non-QoS LAN	H.323	H.261/3	G.7xx	H.225.0	H.245

H.324 Terminal
(multimedia communication over PSTN)



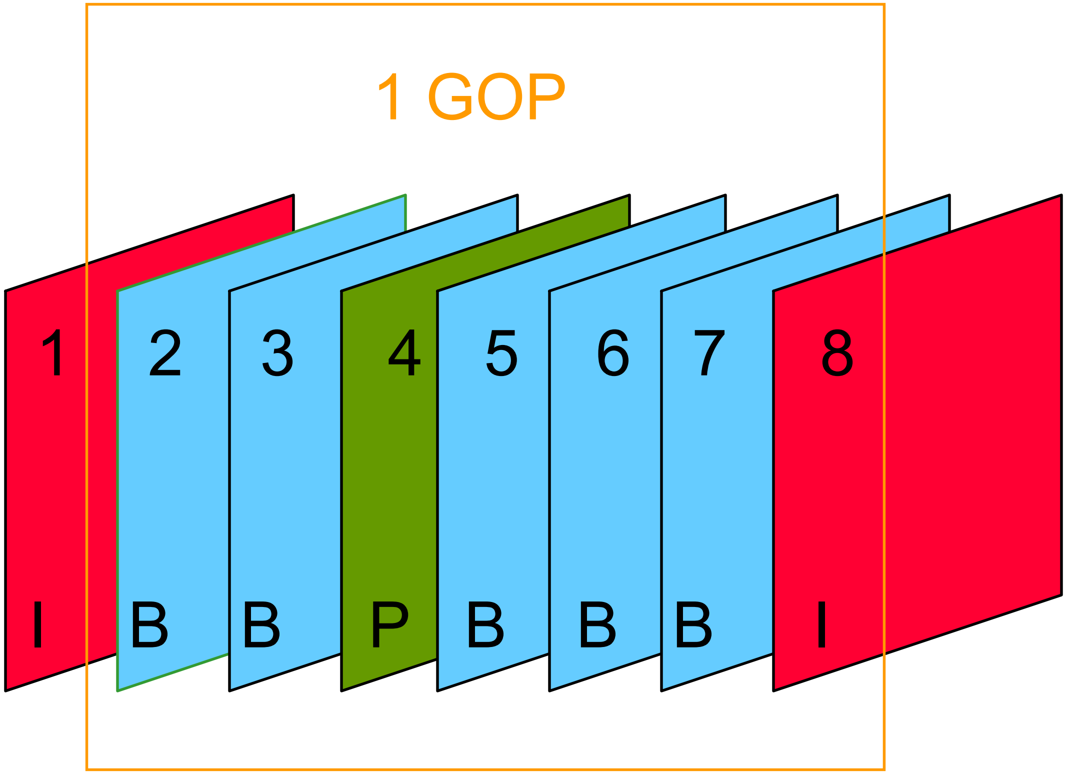
MPEG-1 Overview

- Audio/video on CD-ROM (1.5 Mbps, CIF: 352x240).
  - Maximum: 1.856 mbps, 768x576 pels
- Start late 1988, test in 10/89, Committee Draft 9/90
- ISO/IEC 11172-1~5 (Systems, video, audio, compliance, software).
- Prompted explosion of digital video applications: MPEG1 video CD and downloadable video over Internet
- Software only decoding, made possible by the introduction of Pentium chips, key to the success in the commercial market
- MPEG-1 Audio
  - Offers 3 coding options (3 layers), higher layer have higher coding efficiency with more computations
  - MP3 = MPEG1 layer 3 audio

MPEG-1 video vs H.261

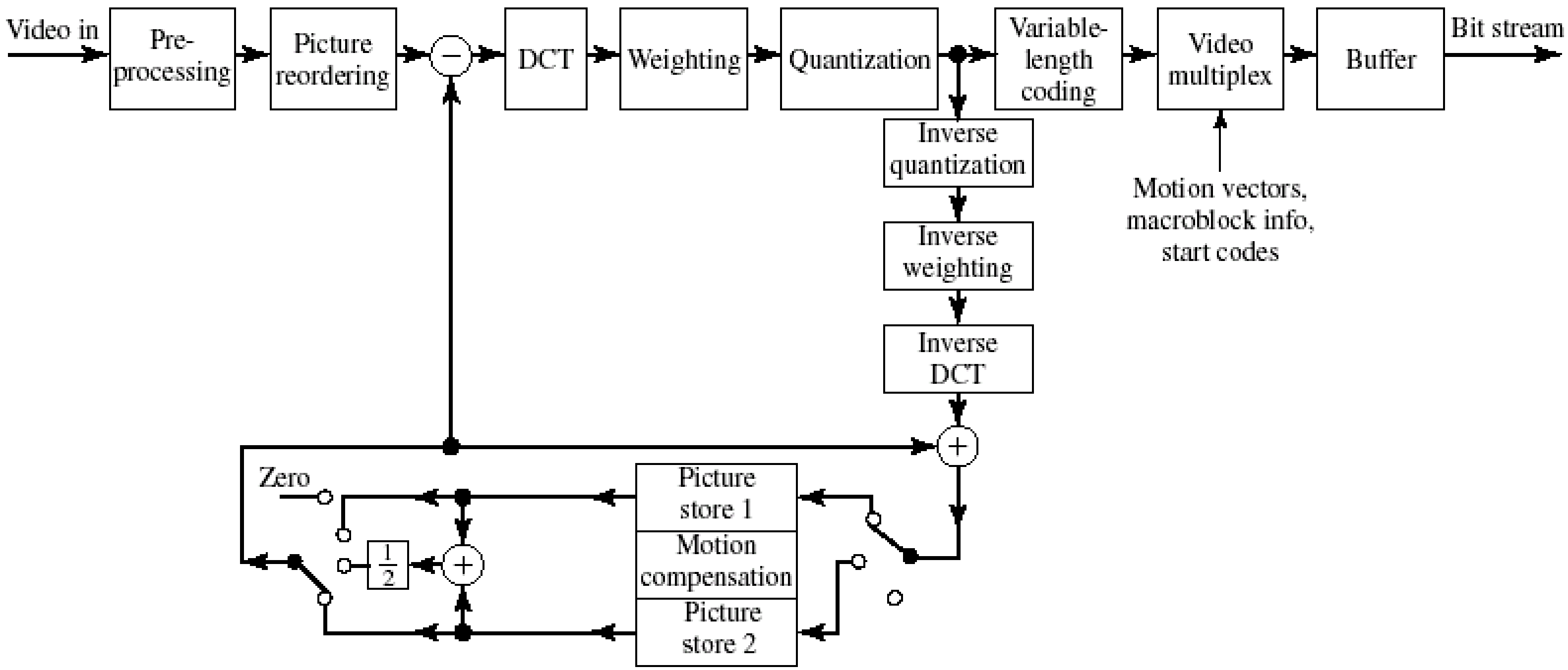
- Developed at about the same time
- Must enable random access (Fast forward/rewind)
  - Using GOP structure with periodic I-picture and P-picture
- Not for interactive applications
  - Do not have as stringent delay requirement
- Fixed rate (1.5 Mbps), good quality (VHS equivalent)
  - SIF video format (similar to CIF)
    - CIF: 352x288, SIF: 352x240
  - Using more advanced motion compensation
    - Half-pel accuracy motion estimation, range up to +/- 64
  - Using bi-directional temporal prediction
    - Important for handling uncovered regions
  - Using perceptual-based quantization matrix for I-blocks (same as JPEG)
    - DC coefficients coded predictively

Group of Picture Structure in MPEG



Encoding order: 4 2 3 8 5 6 7

MPEG-1 Video Encoder



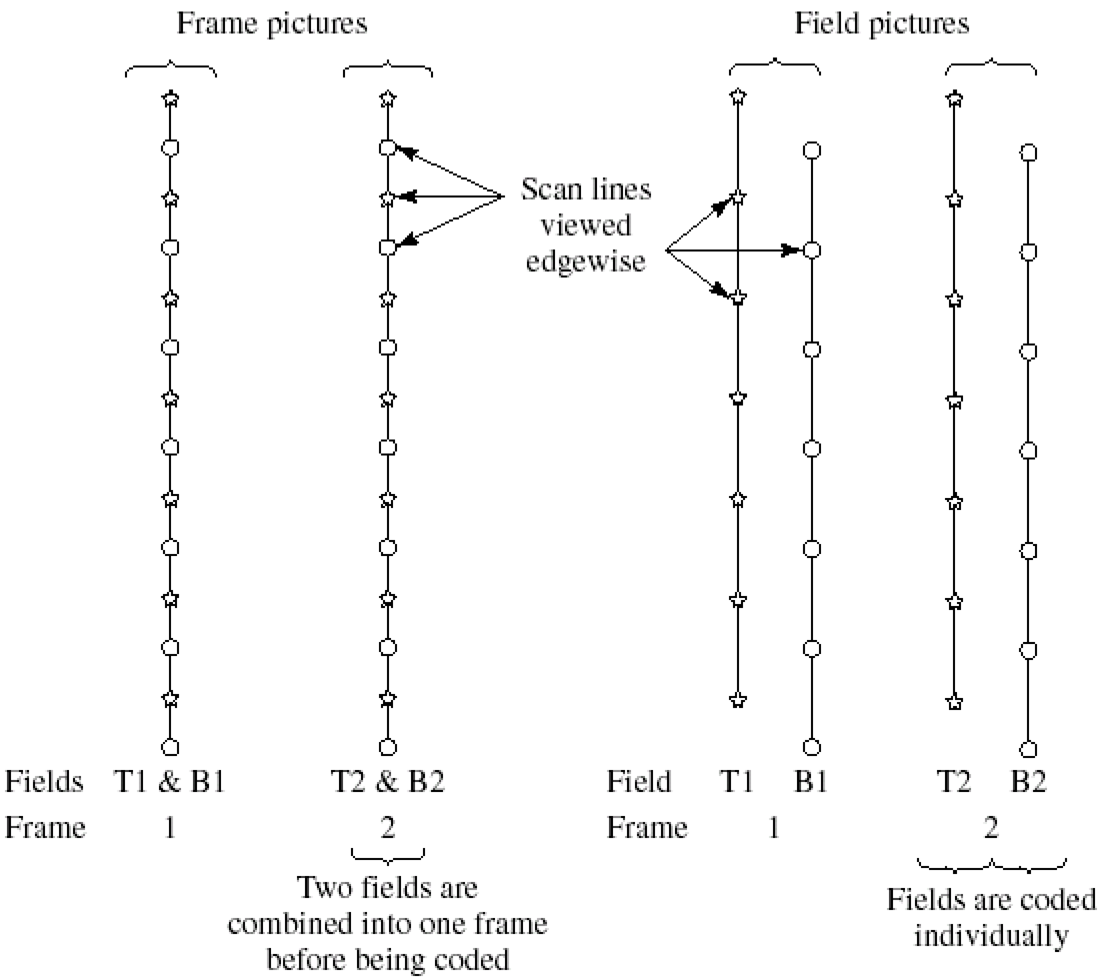
MPEG2 Overview

- A/V broadcast (TV, HDTV, Terrestrial, Cable, Satellite, High Speed Inter/Intranet) as well as DVD video
- 4~8 Mbps for TV quality, 10-15 for better quality at SDTV resolutions (BT.601)
- 18-45 Mbps for HDTV applications
  - MPEG-2 video high profile at high level is the video coding standard used in HDTV
- Test in 11/91, Committee Draft 11/93
- ISO/IEC 13818-1~6 (Systems, video, audio, compliance, software, DSM-CC)
- Consist of various profiles and levels
- Backward compatible with MPEG1
- MPEG-2 Audio
  - Support 5.1 channel
  - MPEG2 AAC: requires 30% fewer bits than MPEG1 layer 3

MPEG2 vs. MPEG1 Video

- MPEG1 only handles progressive sequences (SIF).
- MPEG2 is targeted primarily at interlaced sequences and at higher resolution (BT.601 = 4CIF).
- More sophisticated motion estimation methods (*frame/field prediction mode*) are developed to improve estimation accuracy for interlaced sequences.
- Different DCT modes and scanning methods are developed for interlaced sequences.
- MPEG2 has various scalability modes.
- MPEG2 has various profiles and levels, each combination targeted for different application

Frame vs. Field Picture

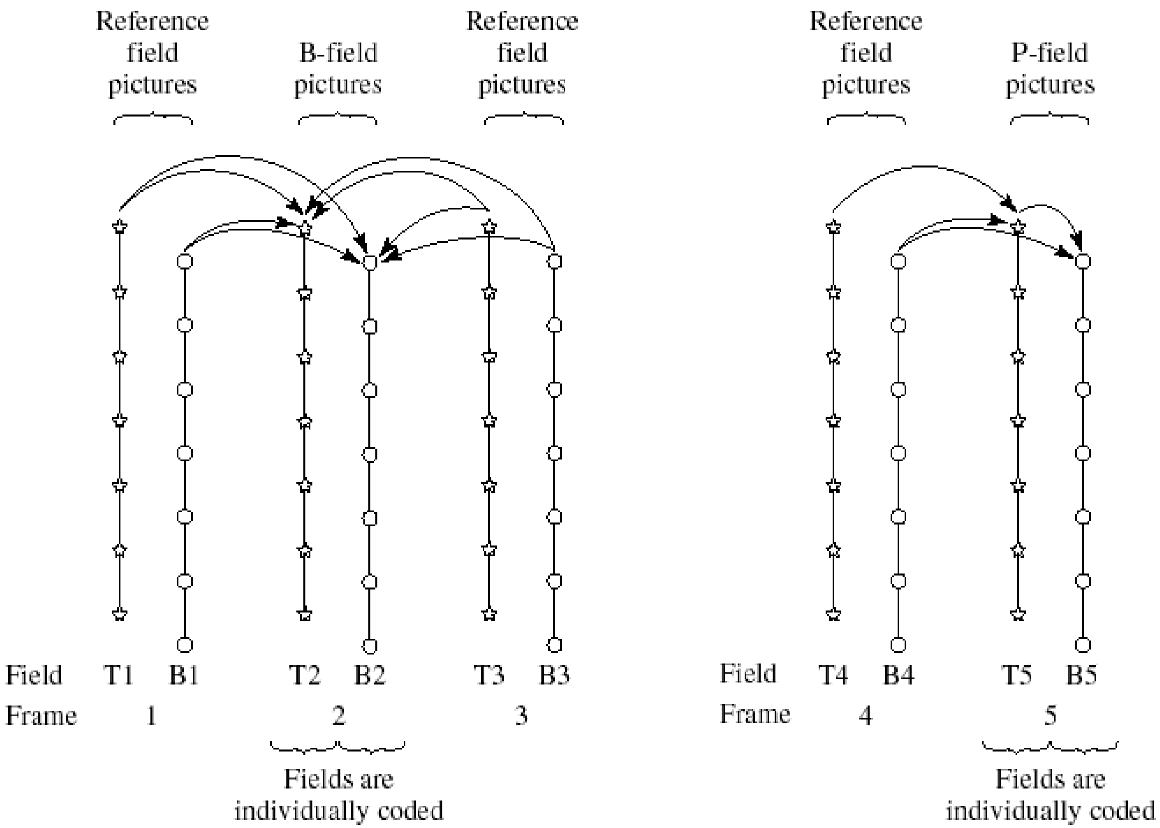


Motion Compensation for Interlaced Video

- Field prediction for field pictures
- Field prediction for frame pictures
- Dual prime for P-pictures
- 16x8 MC for field pictures

Field prediction for field pictures

- Each field is predicted individually from the reference fields
  - A P-field is predicted from one previous field
  - A B-field is predicted from two fields chosen from two reference pictures



Field Prediction for Frame Pictures

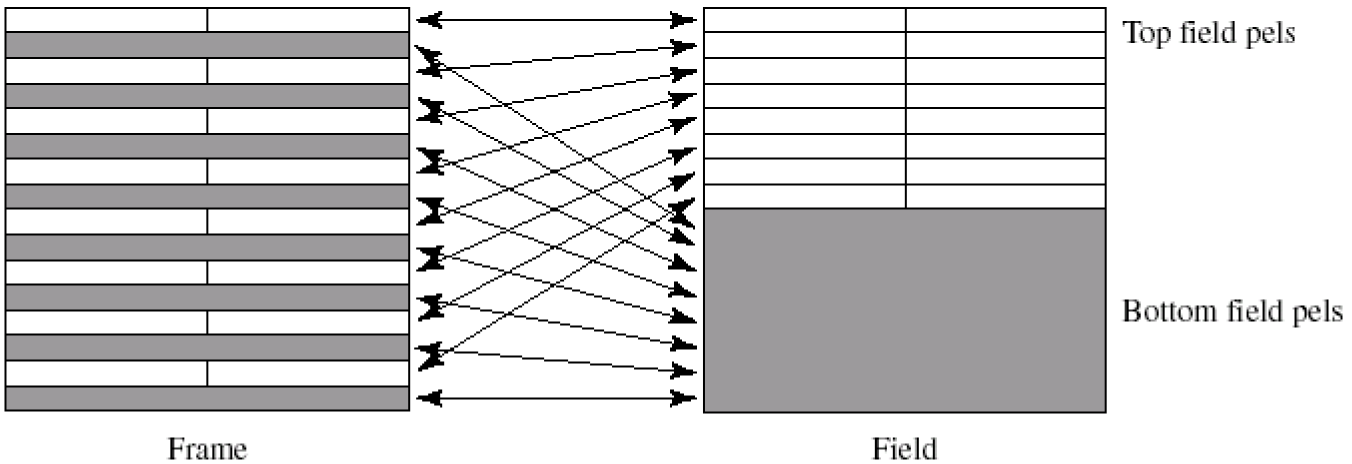


Figure 13.18 Field prediction for frame pictures: the MB to be predicted is split into top field pels and bottom field pels. Each 16×8 field block is predicted separately with its own motion vector (P-frame) or two motion vectors (B-frame).

DCT Modes

Two types of DCT and two types of scan pattern:

- **Frame DCT**: divides an MB into 4 blocks for Lum, as usual
- **Field DCT**: reorder pixels in an MB into top and bottom fields.

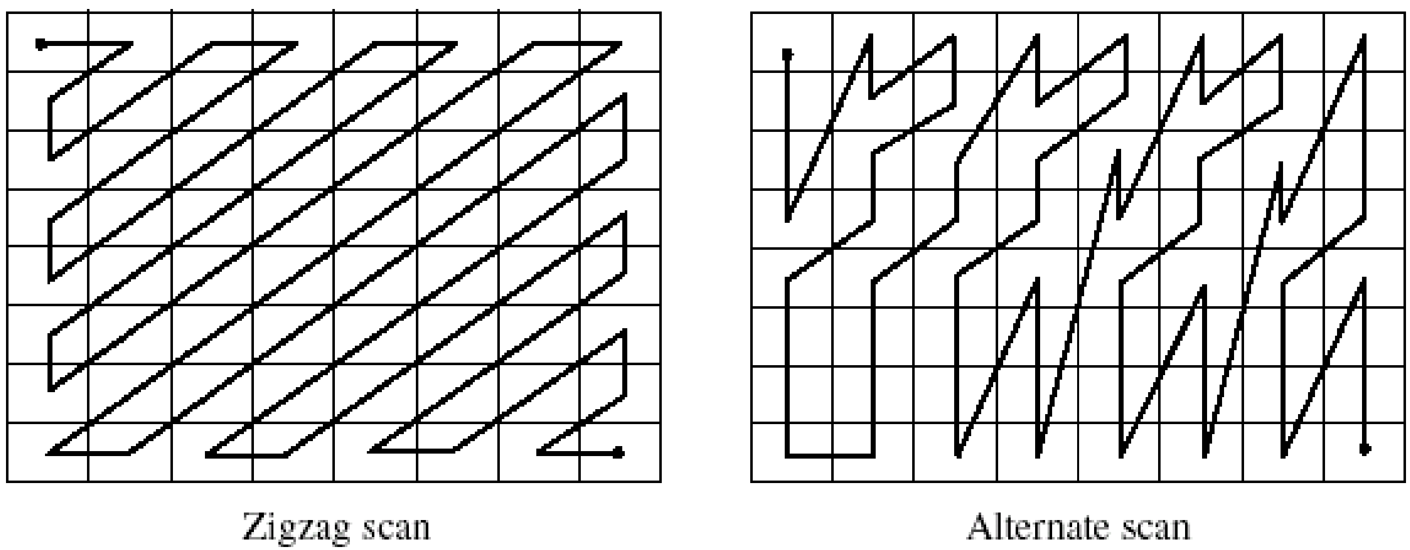
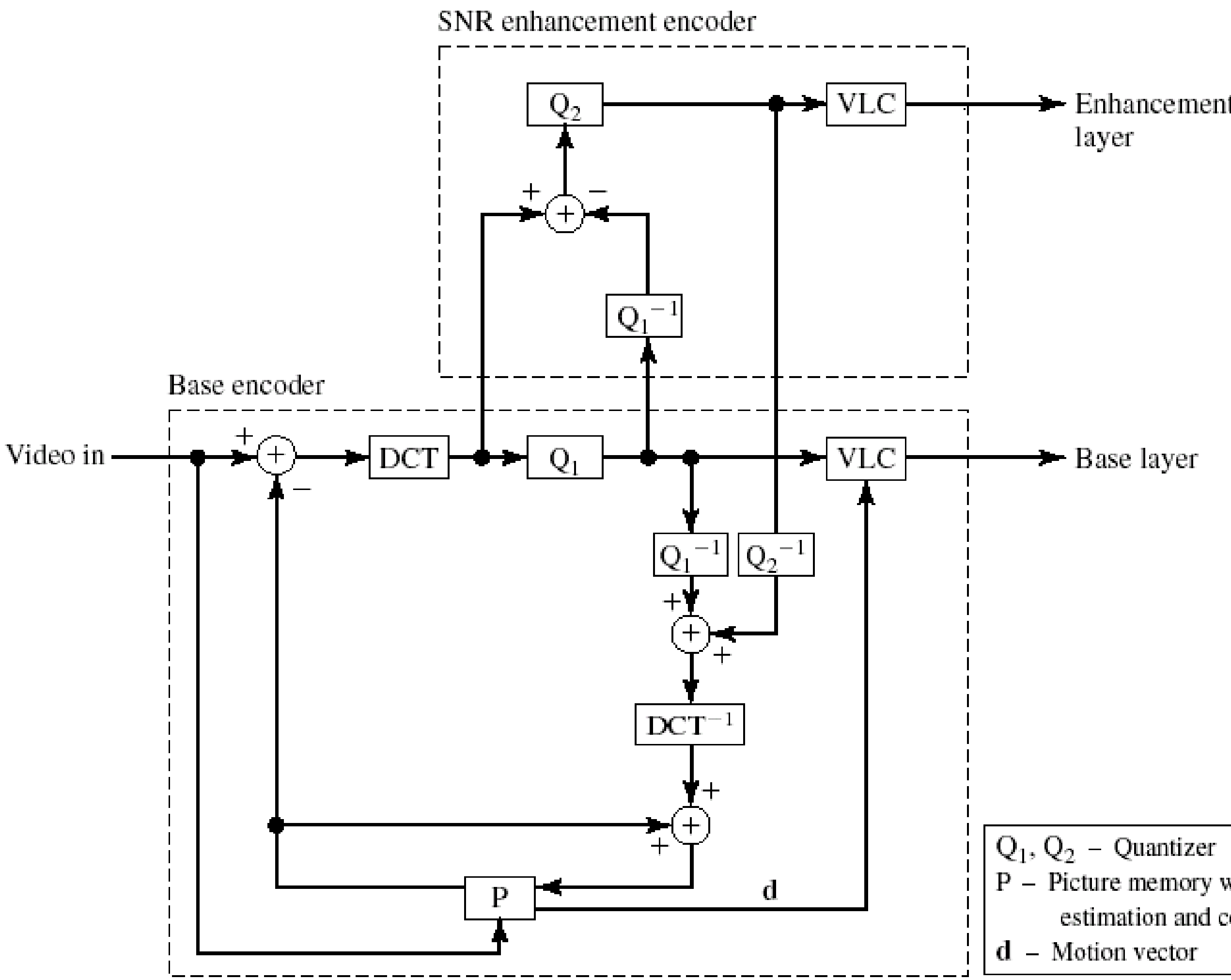


Figure 13.19 The zigzag scan as known from H.261, H.263, and MPEG-1 is augmented by the alternate scan in MPEG-2, in order to code interlaced blocks that have more correlation in the horizontal than in the vertical direction.

MPEG-2 Scalability

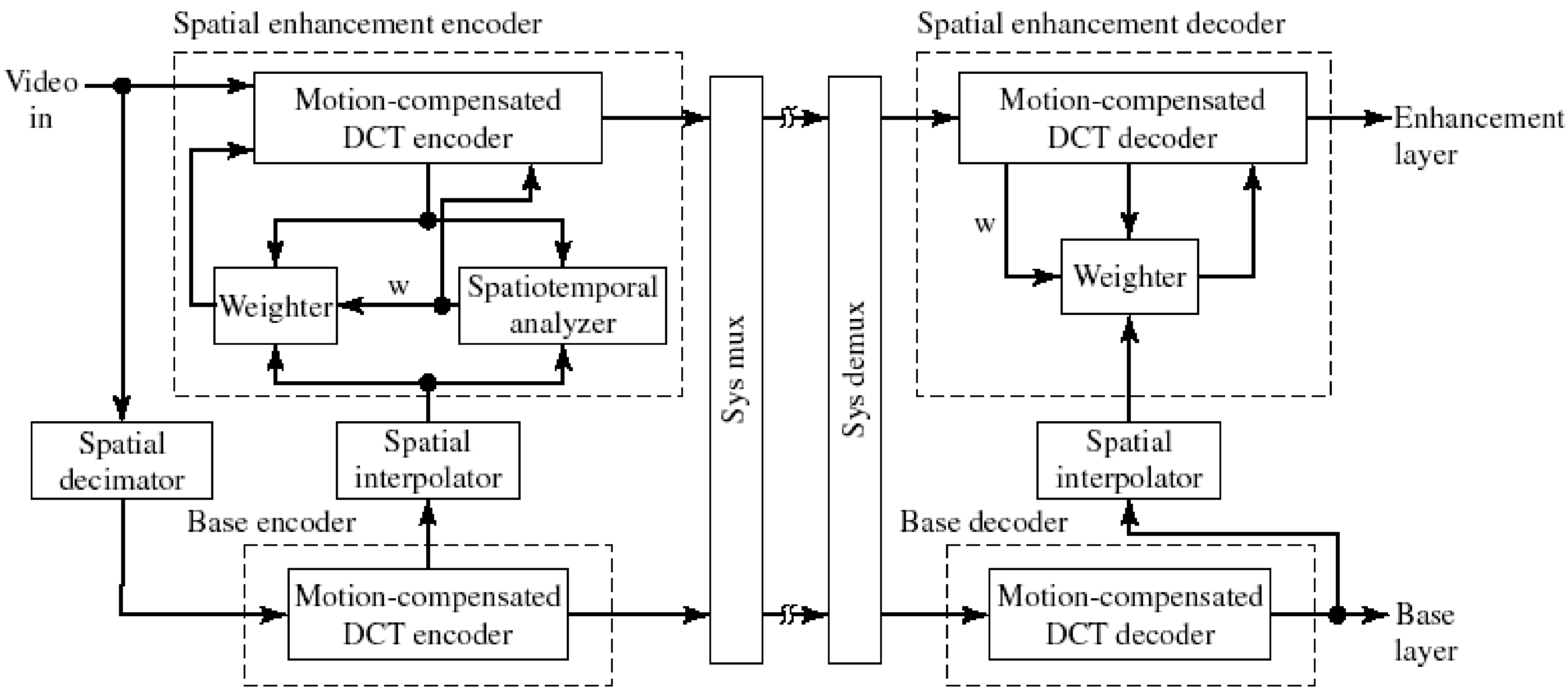
- Data partition
  - All headers, MVs, first few DCT coefficients in the base layer
  - Can be implemented at the bit stream level
  - Simple
- SNR scalability
  - Base layer includes coarsely quantized DCT coefficients
  - Enhancement layer further quantizes the base layer quantization error
  - Relatively simple
- Spatial scalability
  - Complex
- Temporal scalability
  - Simple
- Drift problem:
  - If the encoder's base layer information for a current frame depends on the enhancement layer information for a previous frame
  - Exist in the data partition and SNR scalability modes

SNR Scalability Encoder



Q_1, Q_2 – Quantizer
P – Picture memory with motion estimation and compensation unit
d – Motion vector

Spatial Scalability Codec



Temporal Scalability: Option 1

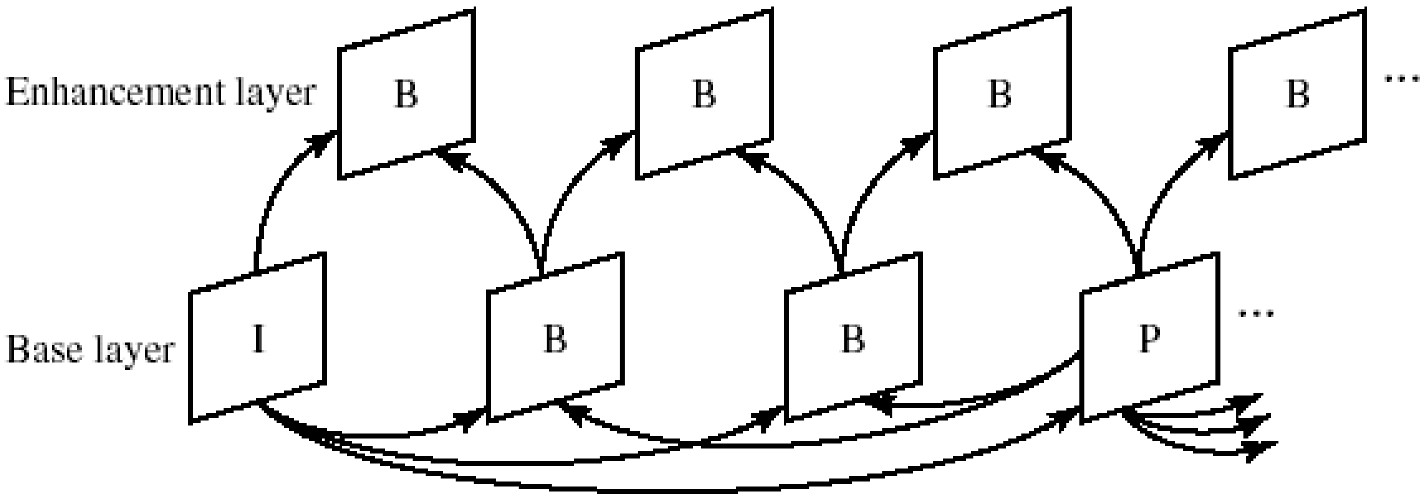


Figure 13.23 A configuration in which temporal scalability may use only the base layer to predict images in the enhancement layer. Obviously, errors in the enhancement layers to do not propagate over time.

Temporal Scalability: Option 2

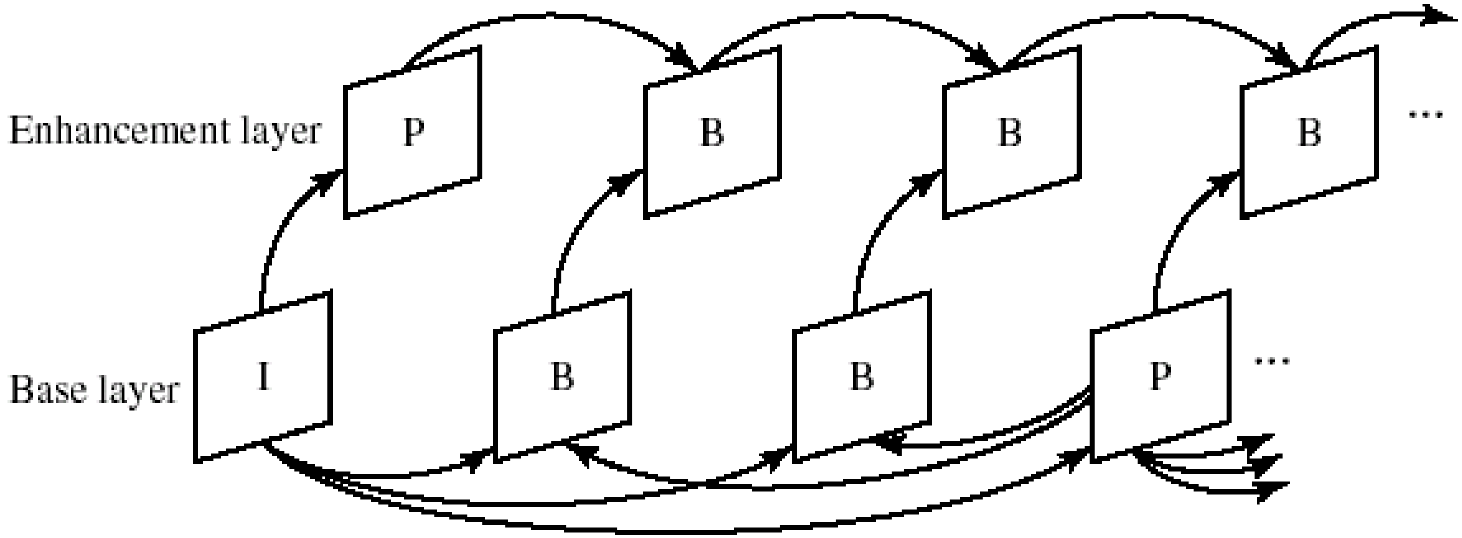


Figure 13.24 A configuration in which temporal scalability enhancement layer may use the base layer and the enhancement layer for prediction. This arrangement is especially useful for coding of stereoscopic video.

Profiles and Levels in MPEG-2

			Profile						
			Simple (I, P) (4:2:0)	Main (I, P, B) (4:2:0)	SNR (I, P, B) (4:2:0)	Spatial (I, P, B) (4:2:0)	High (I, P, B) (4:2:0; 4:2:2)	Multiview (I, P, B) (4:2:0)	4:2:2 (I, P, B) (4:2:0; 4:2:2)
Level	Low	Pels/line			352	352		352	
		Lines/frame			288	288		288	
		fps			30	30		30	
		mbps			4	4		8	
	Main	Pels/line	720	720	720		720	720	720
		Lines/frame	576	576	576		576	576	512/608
		fps	30	30	30		30	30	30
		mbps	15	15	15		20	25	50
	High-1440	Pels/line	1440		1440	1440	1440		
		Lines/frame	1152		1152	1152	1152		
		fps	60		60	60	60		
		mbps	60		60	80	100		
	High	Pels/line	1920			1920	1920	1920	
		Lines/frame	1152			1152	1152	1152	
		fps	60			60	60	60	
		mbps	80			100	130	300	

I, P, B: allowable picture types. Maximum bit rates include all layers in case of scalable bit streams.

Profiles: tools

Levels: parameter range for a given profile

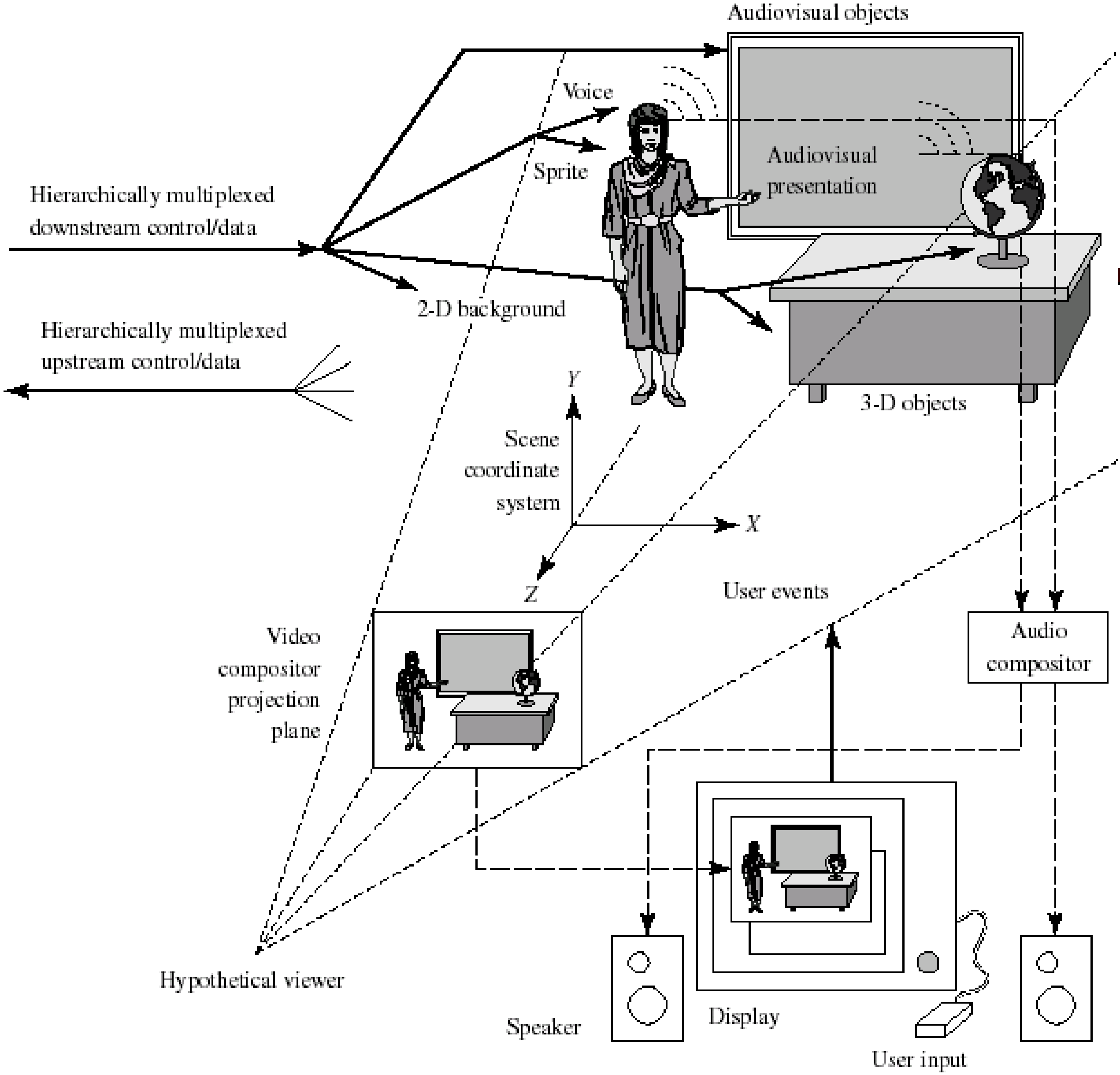
Main profile at main level (mp@ml) is the most popular, used for digital TV

Main profile at high level (mp@hl): HDTV

4:2:2 at main level (4:2:2@ml) is used for studio production

MPEG-4 Overview

- Functionalities beyond MPEG-1/2
  - Interaction with individual objects
    - The displayed scene can be composed by the receiver from coded objects
  - Scalability of contents
  - Error resilience
  - Coding of both natural and synthetic audio and video

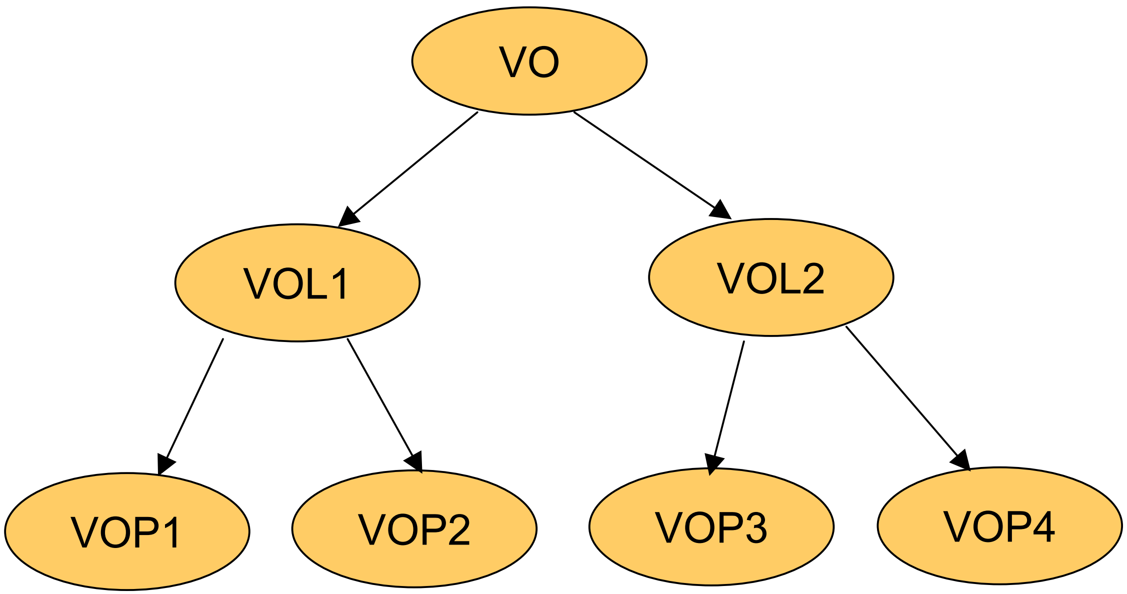


The displayed scene is composed by the receiver based on desired view angle and objects of interests

Object-Based Coding

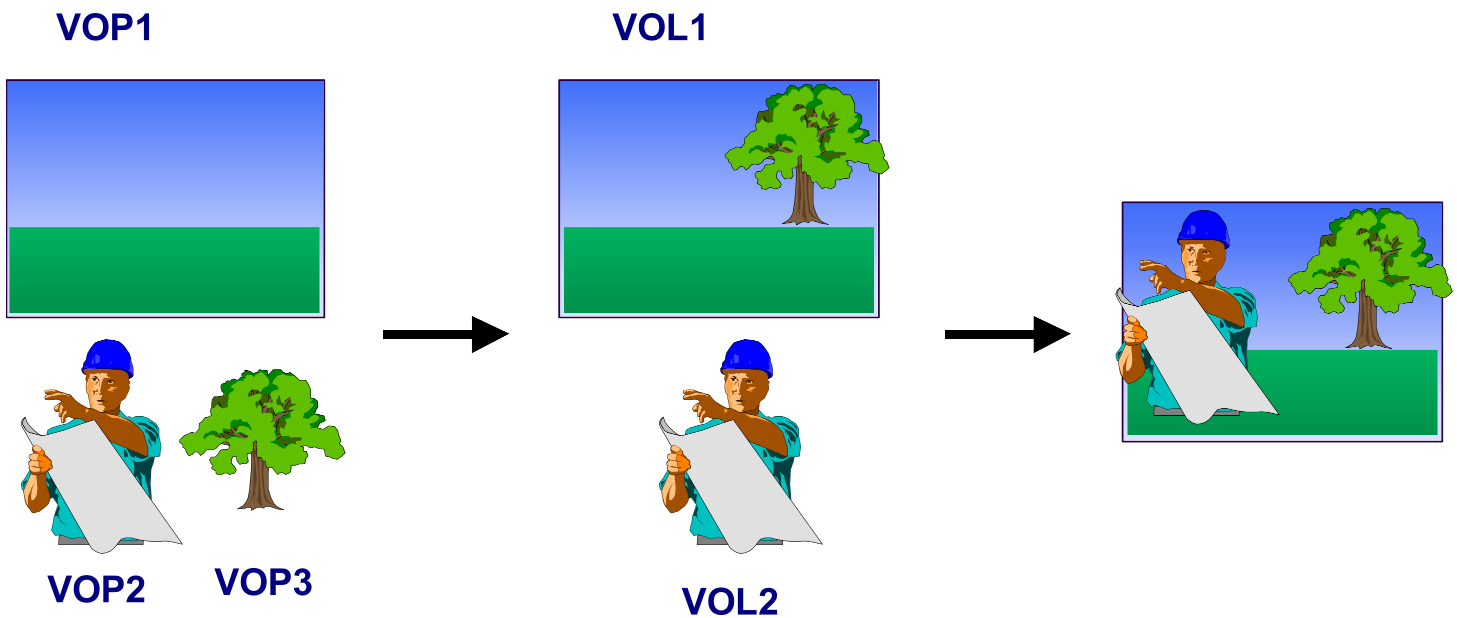
- Entire scene is decomposed into multiple objects
  - Object segmentation is the most difficult task!
  - But this does not need to be standardized ☺
- Each object is specified by its shape, motion, and texture (color)
  - Shape and texture both changes in time (specified by motion)
- MPEG-4 assumes the encoder has a segmentation map available, specifies how to code (actually decode!) shape, motion and texture

Object Description Hierarchy in MPEG-4



VO: video object
VOL: video object layer
(can be different parts of a VO or different rate/resolution representation of a VOL)
VOP: video object plane

Example of Scene Composition

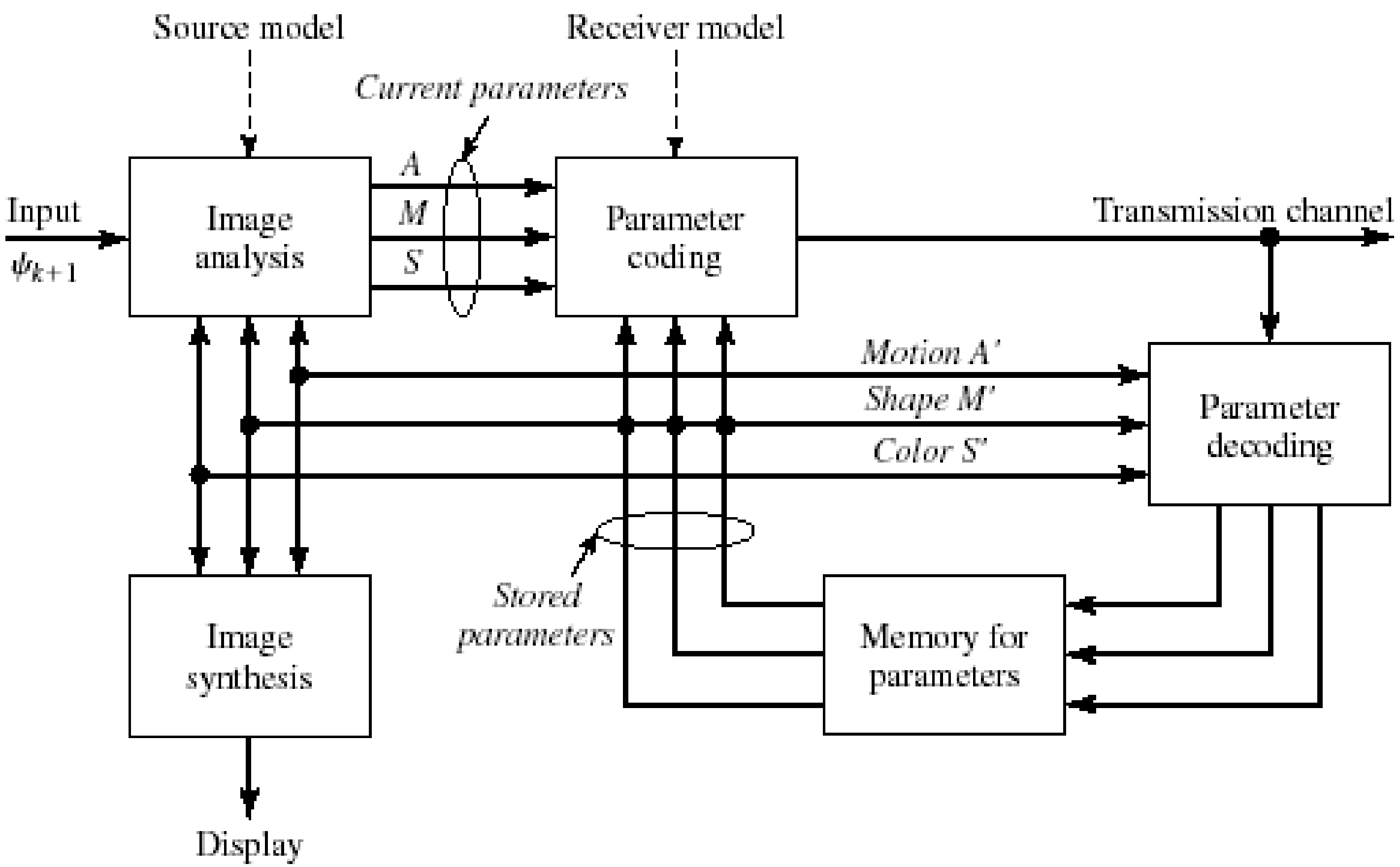


The decoder can compose a scene by including different VOPs in a VOL

Object-Based Coding Basics (Chap 10)

- Entire scene is decomposed into multiple objects
  - Object segmentation is the most difficult task!
  - But this does not need to be standardized ☺
- Each object is specified by its shape, motion, and texture (color)
  - Shape and texture both changes in time (specified by motion)

Generic Block Diagram for Object-Based Coding



Shape Coding Methods

- Shape is specified by alpha maps
  - Binary alpha map: specifies whether a pel belongs to an object
  - Gray scale alpha map: a pel belong to the object can have a transparency value in the range (0-255)
- Bitmap coding
  - Run-length coding
  - Pel-wise coding using context-based arithmetic coding
  - Quadtree coding
- Contour coding
  - Chain coding
  - Fourier descriptors
  - Polygon approximation

Context-based Arithmetic Coding

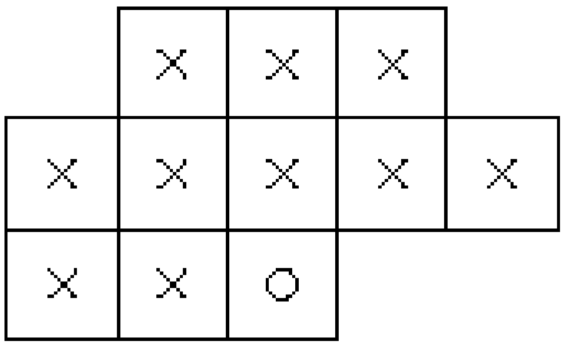


Figure 10.2 Template for defining the context of the pel to be coded (designated by "○").

Quadtree Shape Coding

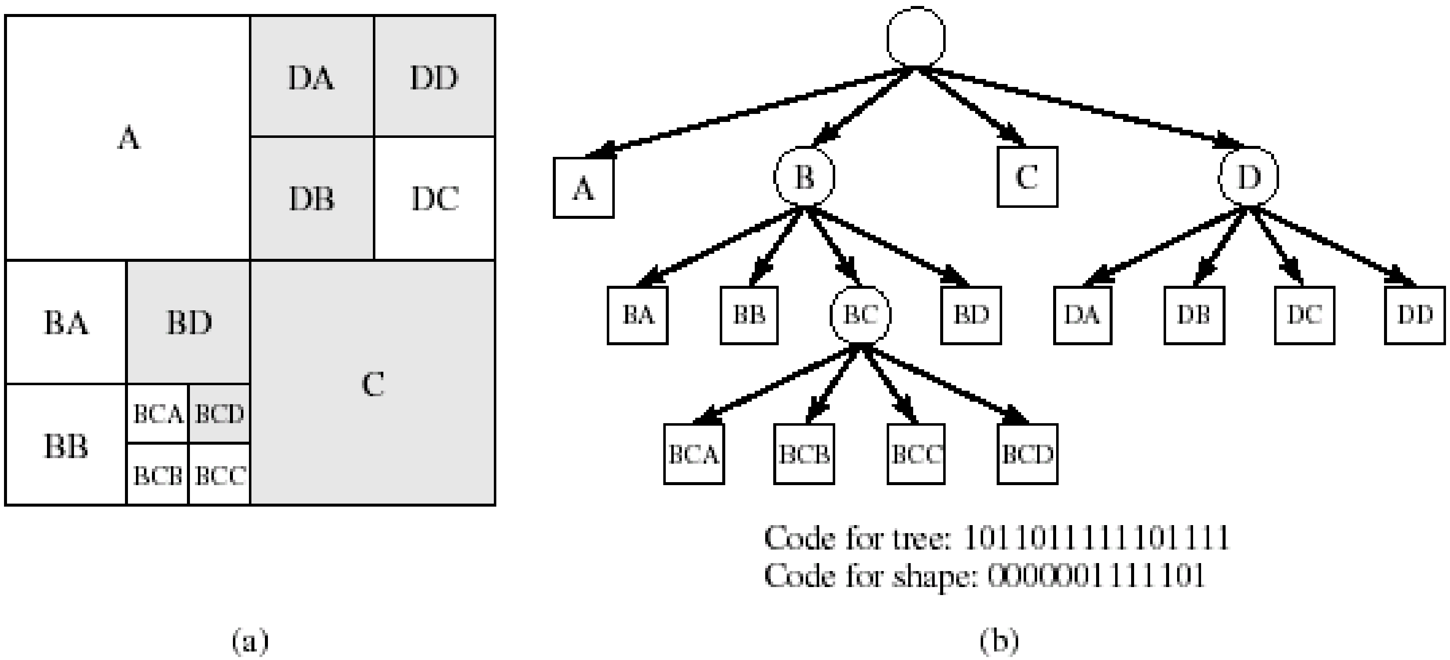


Figure 10.3 (a) An image is divided into a quad-tree that describes the object shape. (b) Related quad-tree—with code that describes the tree and labels for each square, such that the shape can be recovered.

Chain Coding and Differential Chain Coding

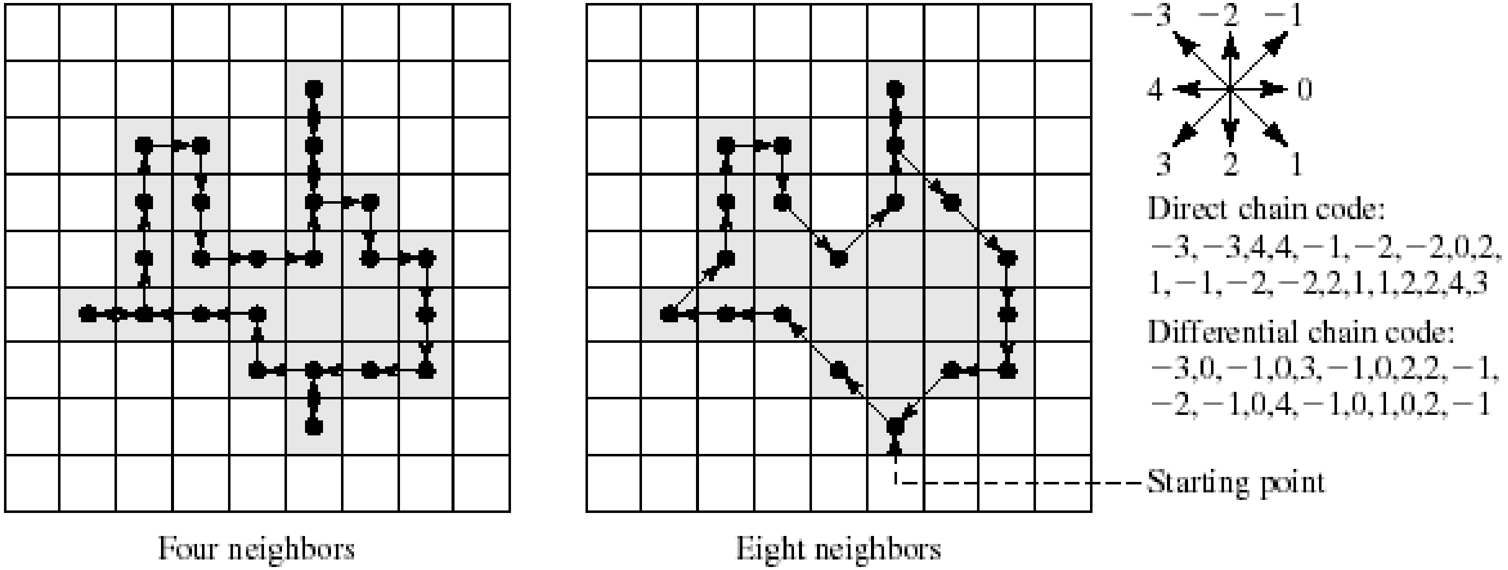
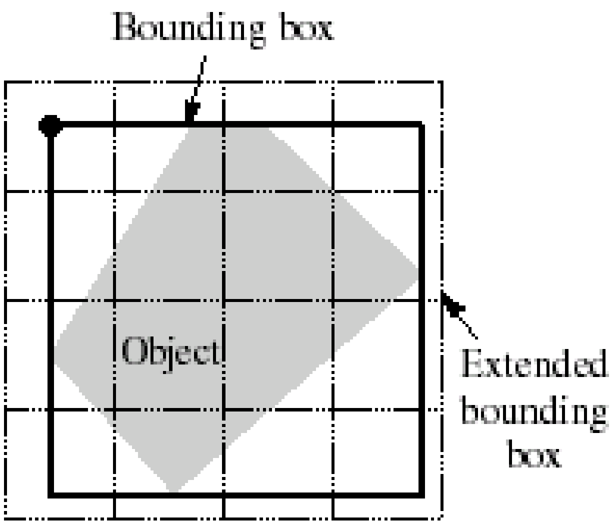


Figure 10.4 Chain code for pels with four and eight neighbors. We give examples for a direct chain code and for a differential code of the eight-connected chain. The first symbols of the two codes are identical and define the starting direction. The following symbols of the differential chain code are created by aligning direction 0 of the direction star with the direction of the last coded symbol.

Coding of Texture with Arbitrary Shape

- Texture extrapolation through padding
- Shape Adaptive DCT



Special color is assigned to pels not belonging to the object

Bounding box can be extended to multiples of 8x8 if the resulting box is to be coded using 8x8 DCT

Shape Adaptive DCT

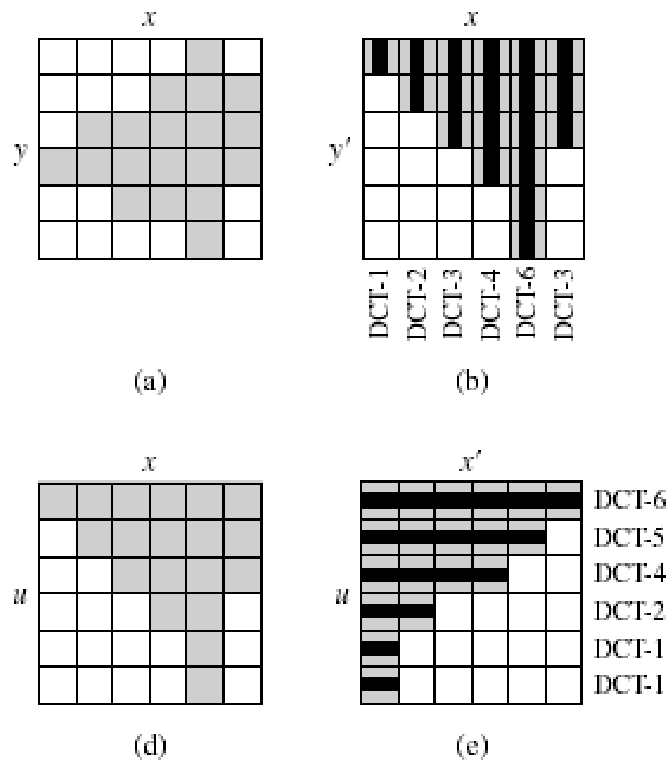
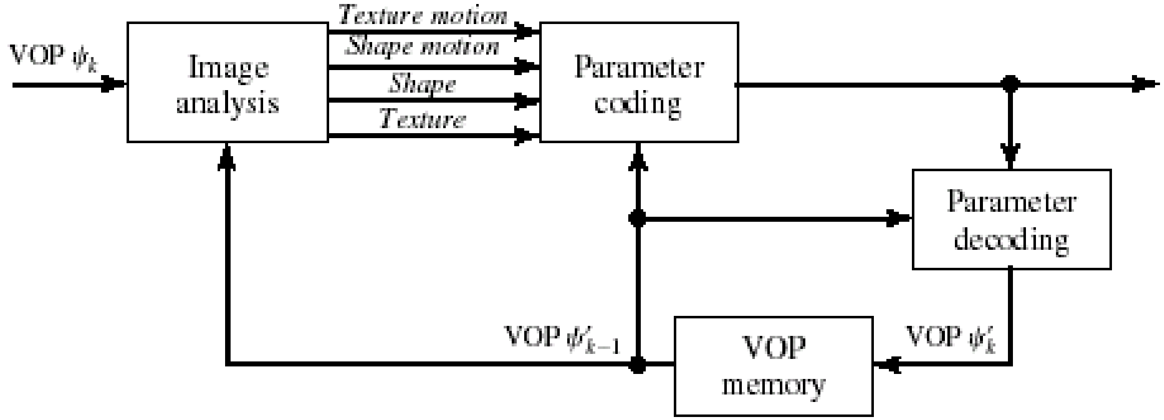


Figure 10.9 A shape-adaptive DCT requires transforms of length *n*: (a) original image segment, (b) pels shifted vertically, (c) location of DCT coefficients after vertical 1-D DCT, (d) location of DCT coefficients prior to horizontal 1-D DCT, (e) DCT coefficients shifted horizontally, (f) location of DCT coefficients after horizontal DCT.

MPEG-4 Shape Coding

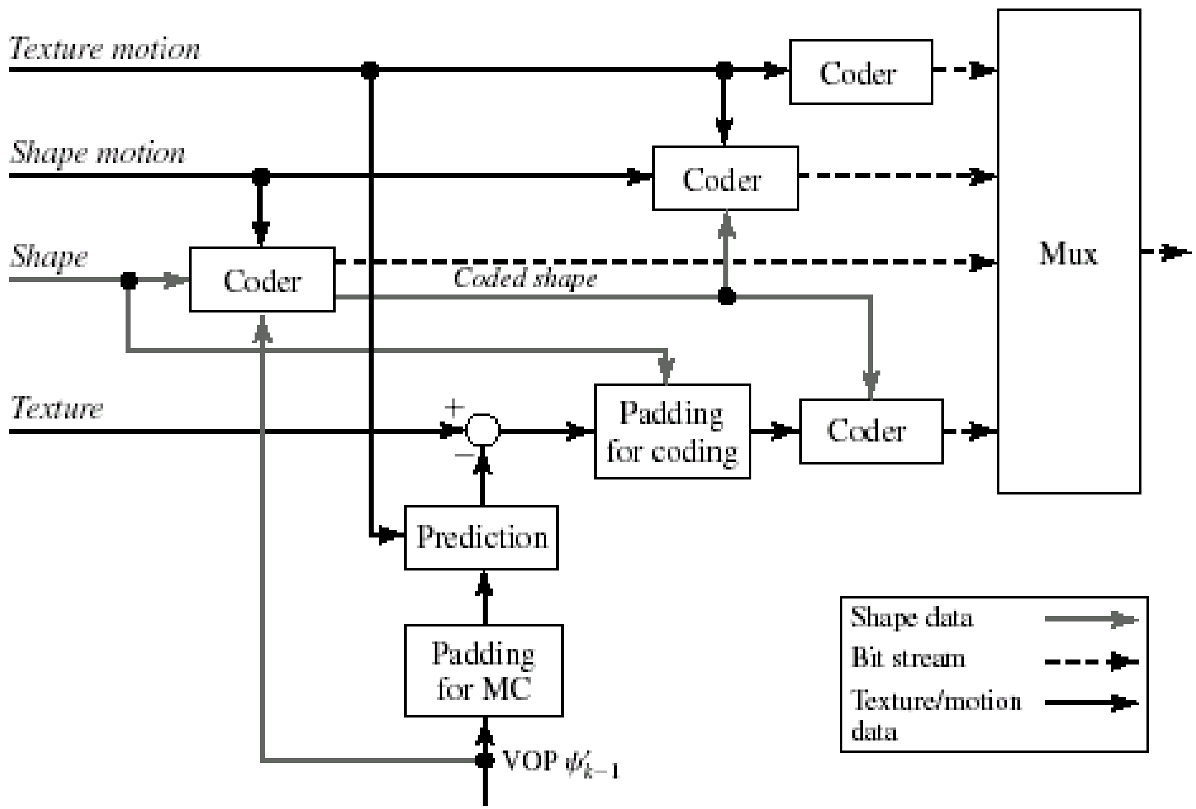
- Uses block-based approach (block=MB)
  - Boundary blocks (blocks containing both the object and background)
  - Non-boundary blocks: either belong to the object or background
- Boundary block’s binary alpha map (binary alpha block) is coded using context-based arithmetic coding
  - Intra-mode: context pels within the same frame
  - Inter-mode: context pels include previous frame, displaced by MV
    - Shape MV separate from texture MV
    - Shape MV predictively coded using texture MV
- Grayscale alpha maps are coded using DCT
- Texture in boundary blocks coded using
    - padding followed by conventional DCT
    - Or shape-adaptive DCT



(a)

MPEG4 video coder overview

Details of parameter coding

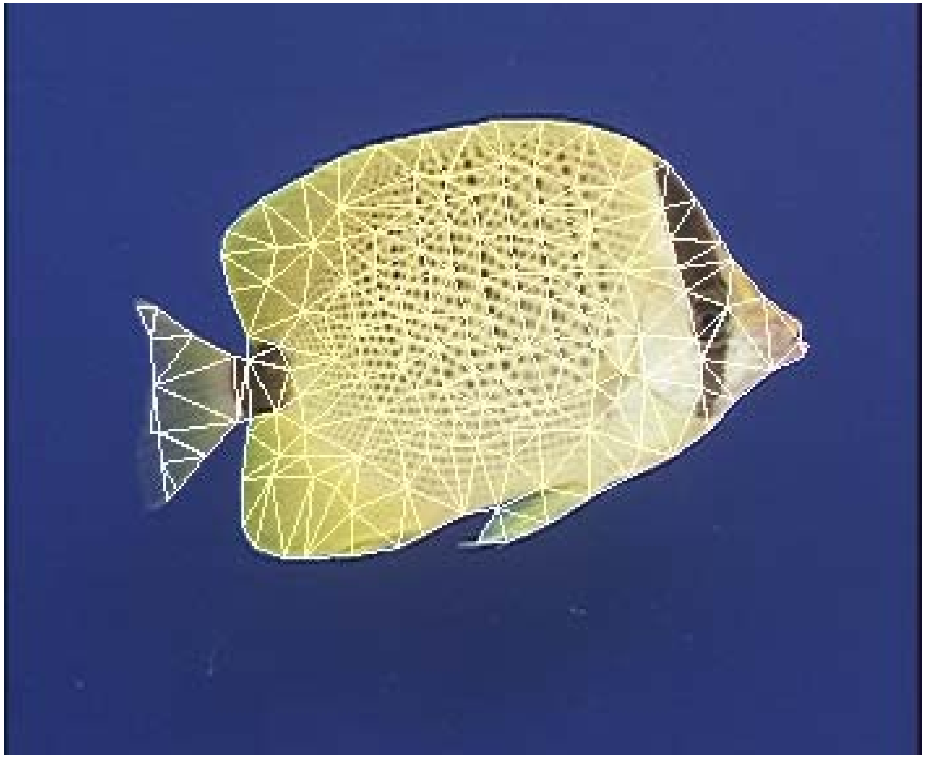


Still Texture Coding

- MPEG-4 defines still texture coding method for intra frame, sprite, or texture map of an mesh object
- Use wavelet based coding method

Mesh Animation

- An object can be described by an initial mesh and MVs of the nodes in the following frames
- MPEG-4 defines coding of mesh geometry, but not mesh generation



Body and Face Animation

- MPEG-4 defines a default 3-D body model (including its geometry and possible motion) through body definition table (BDP)
- The body can be animated using the body animation parameters (BAP)
- Similarly, face definition table (FDP) and face animation parameters (FAP) are specified for a face model and its animation

Face Animation Through FAP

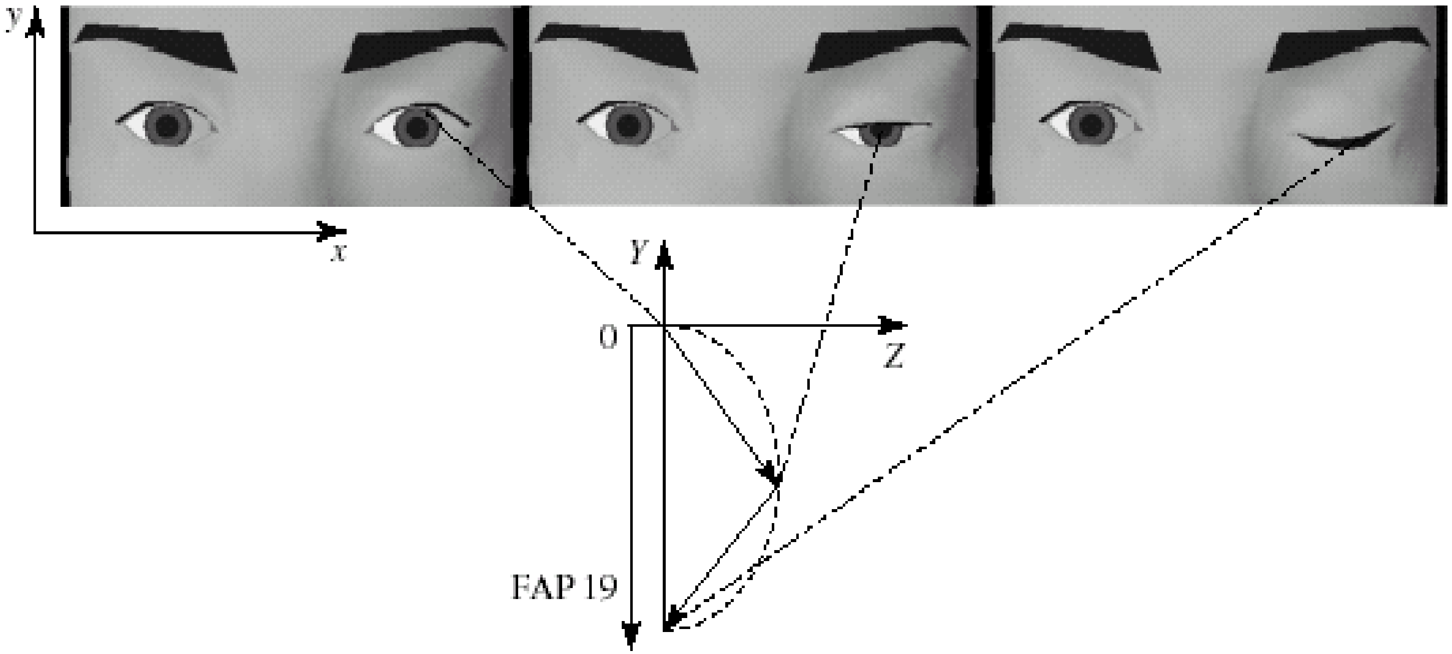


Figure 13.36 Neutral state of the left eye (left) and two deformed animation phases for the eye blink (FAP 19). The FAP defines the motion of the eyelid in the negative y direction; the faceDefTable defines the motion in one of the vertices of the eyelid in the x and z directions.

Text-to-Speech Synthesis with Face Animation

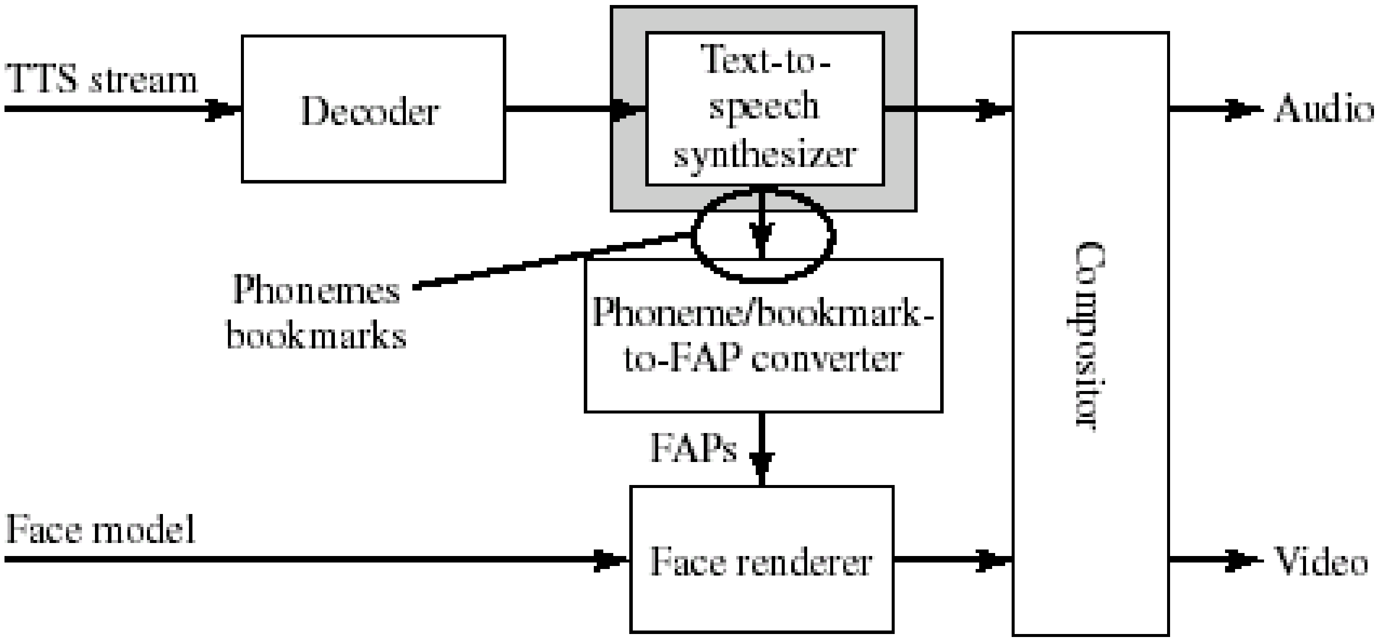


Figure 13.37 MPEG-4 architecture for facial animation, allowing synchronization of facial expressions and speech generated by a proprietary text-to-speech synthesizer.

MPEG-4 Profiles

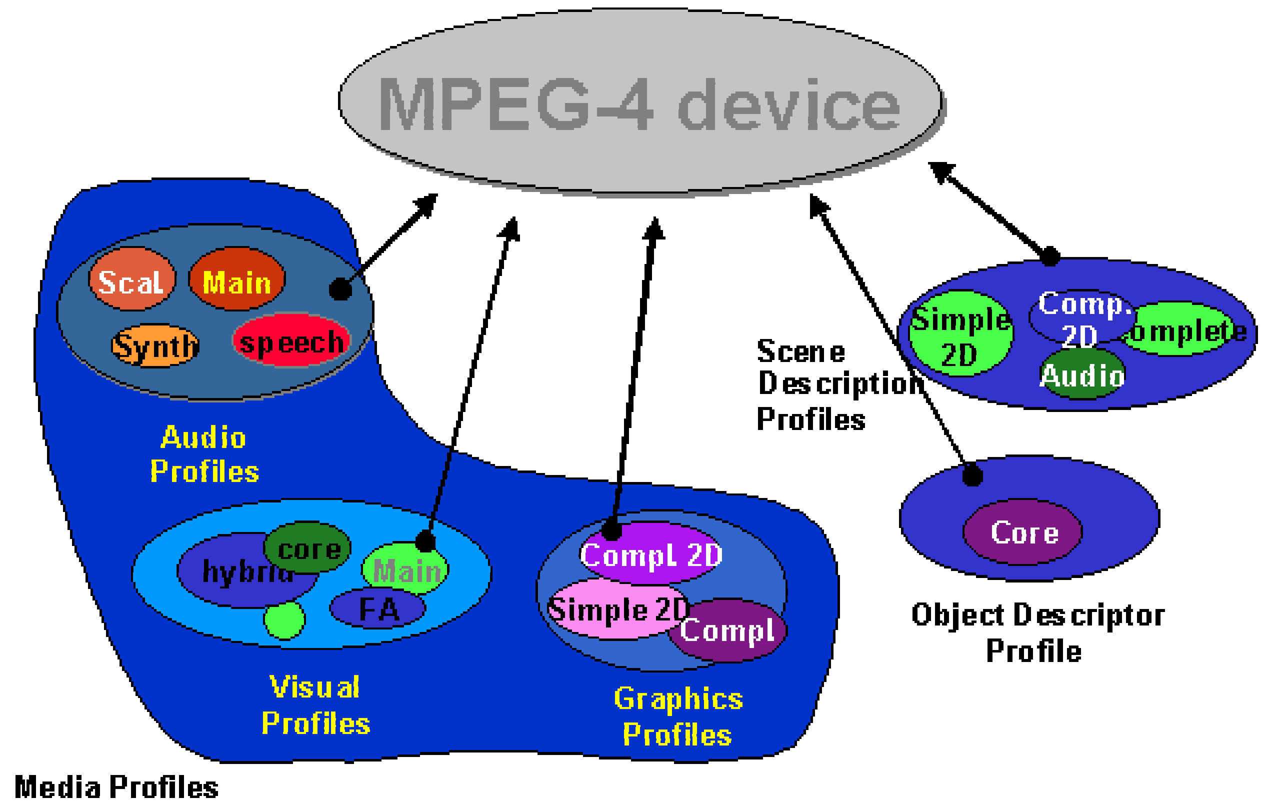


Figure 2 - an overview of the profile structure in MPEG-4. Note that the figure does not list all profiles, notably in the visual area. It is not meant to reflect hierarchical relationships.

Video Coding Efficiency Tools

- Sprite
  - Code a large background in the beginning of the sequence, plus affine mappings, which map parts of the background to the displayed scene at different time instances
  - Decoder can vary the mapping to zoom in/out, pan left/right
- Global motion compensation
  - Using 8-parameter projective mapping
  - Effective for sequences with large global motion
- DC and AC prediction: can predict DC and part of AC from either the previous and block above
- Quarter-pel motion estimation
- Similar to H.263
  - 3D VLC
  - Four MVs and Unrestricted MVs
  - OBMC not required

MPEG-4 vs. MPEG-1 Coding Efficiency

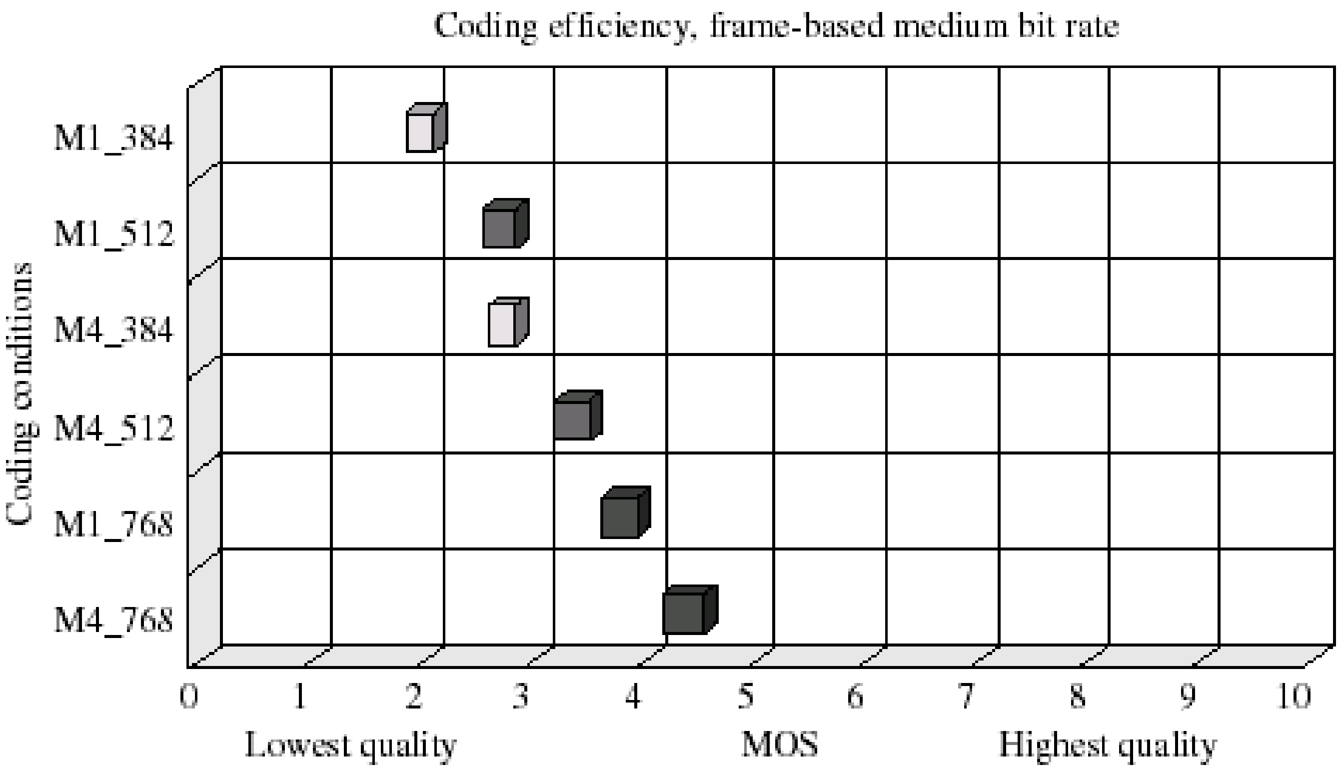
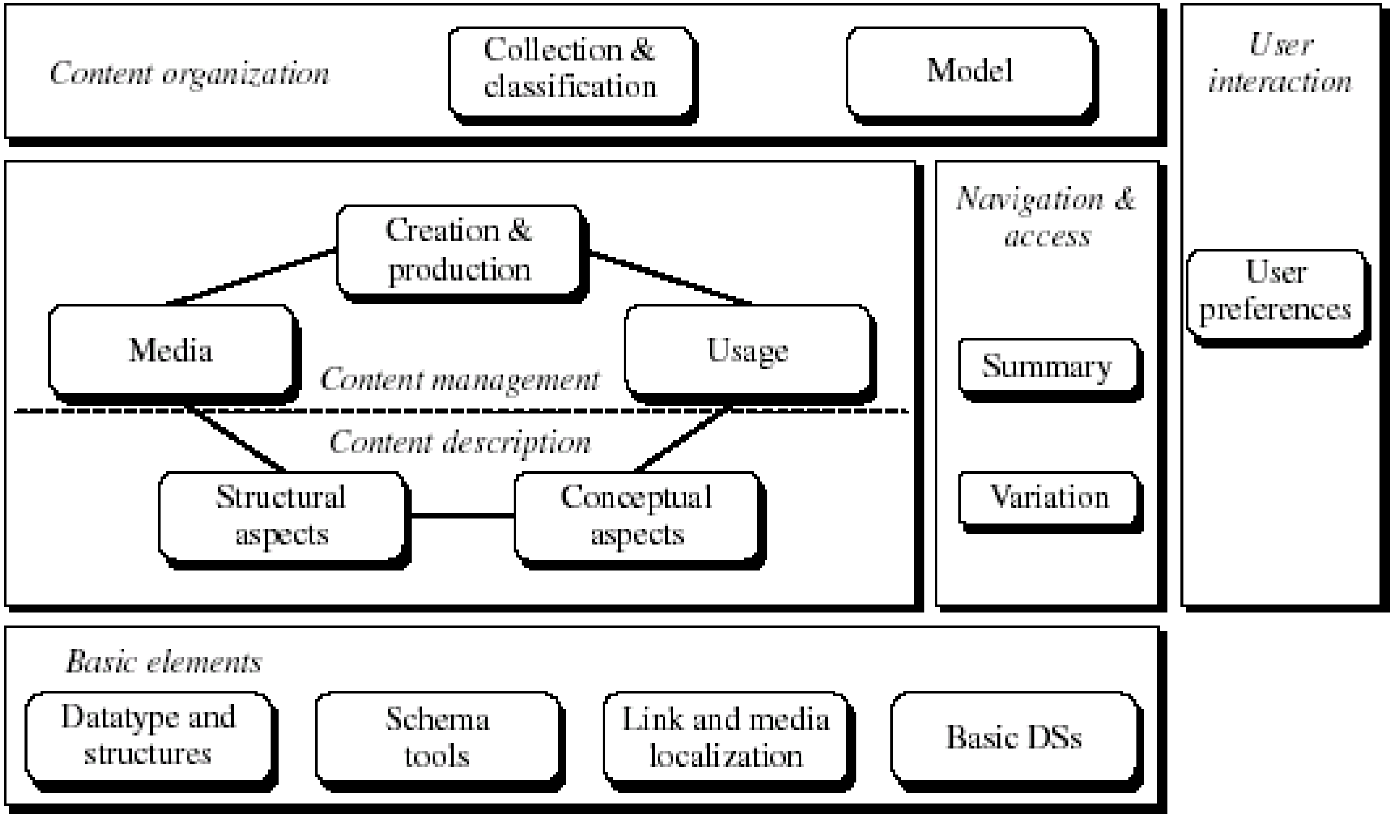


Figure 13.39 Subjective quality of MPEG-4 Main profile versus MPEG-1. M4_x is an MPEG-4 coder operating at the rate of x kbps; M1_x is an MPEG-1 encoder operating at the given rate [27].

MPEG-7 Overview

- MPEG-1/2/4 make content available, whereas MPEG-7 allows you to find the content you need!
  - Enable multimedia document indexing, browsing, and retrieval
  - Define the syntax for the metadata (e.g. index and summary) attached to the document
  - Generation of index and summary is not part of the standard!
- Content description in MPEG-7
  - Descriptor (D): describing low-level features
  - Description scheme (DS): combining Ds to describe high-level features/structures
  - Description definition language (DDL): define how Ds and DSs can be defined or modified
  - System tools

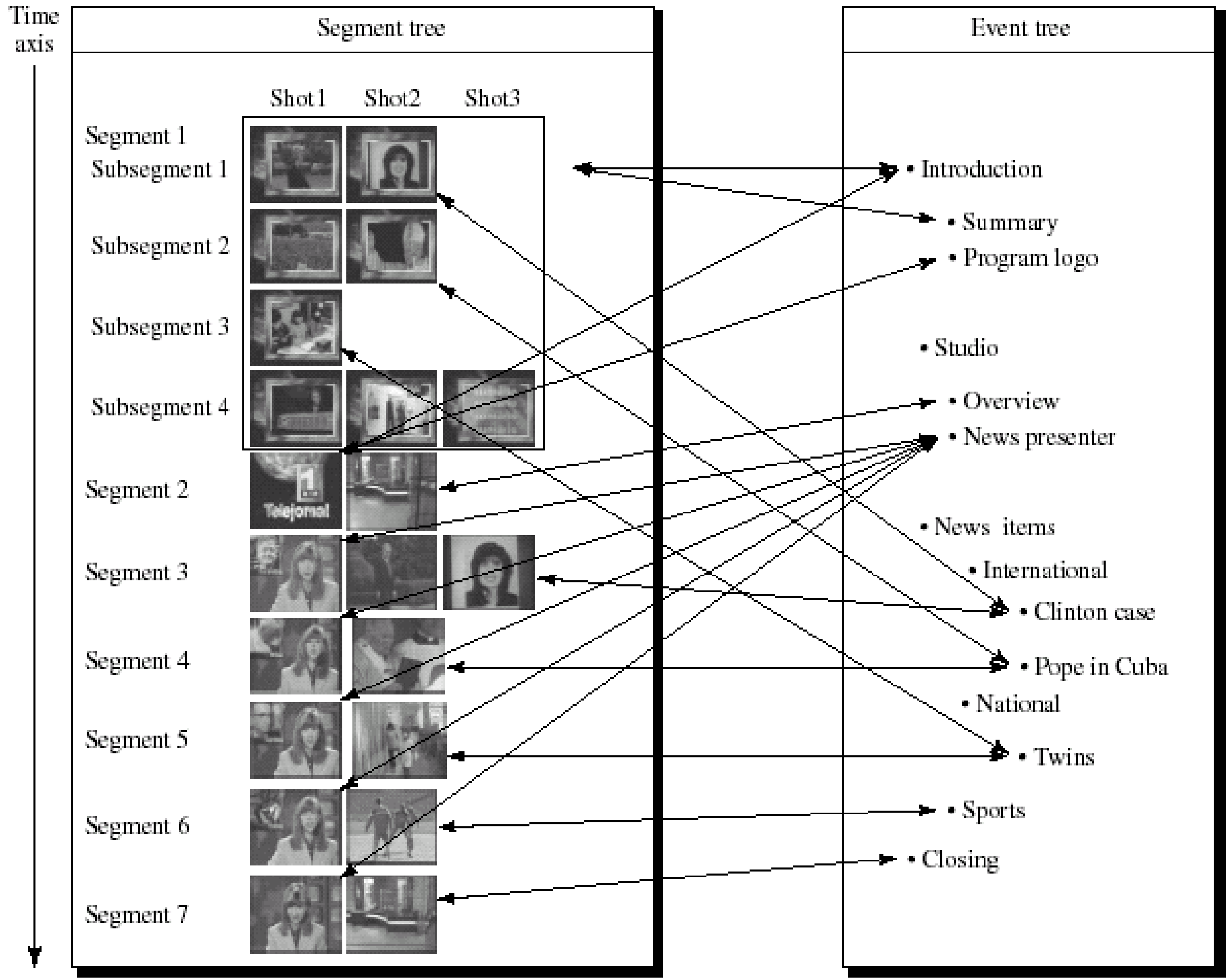
Multimedia Description Scheme Overview



Content description: segment tree and event tree

Segment tree
= table of contents

Event tree
= index



MPEG-7 Visual Descriptors

- Color
  - Histogram, dominant color, etc.
- Texture
  - Homogeneity: energy in different orientation and frequency bands (Gabor transform)
  - Coarseness, directionarity, regularity
  - Edge orientation histogram
- Motion
  - Camera motion
  - Motion trajectory of feature points in non-rigid object
  - Motion parameters of a rigid object
  - Motion activity
- Shape
  - Boundary-based vs. region-based

Summary

- H.261:
  - First video coding standard, targeted for video conferencing over ISDN
  - Uses block-based hybrid coding framework with integer-pel MC
- H.263:
  - Improved quality at lower bit rate, to enable video conferencing/telephony below 54 bkps (modems or internet access, desktop conferencing)
  - Half-pel MC and other improvement
- MPEG-1 video
  - Video on CD and video on the Internet (good quality at 1.5 mbps)
  - Half-pel MC and bidirectional MC
- MPEG-2 video
  - TV/HDTV/DVD (4-15 mbps)
  - Extended from MPEG-1, considering interlaced video

Summary (Cnt'd)

- MPEG-4
  - To enable object manipulation and scene composition at the decoder -> interactive TV/virtual reality
  - Object-based video coding: shape coding
  - Coding of synthetic video and audio: animation
- MPEG-7
  - To enable search and browsing of multimedia documents
  - Defines the syntax for describing the structural and conceptual content
- Newer standards
  - H.264: improved coding efficiency (by having more options for optimization)
  - MPEG-21: beyond MPEG-7, considering intellectual property protection, etc.

Homework

- Reading assignment: Chap. 13