An Introduction to MPEG

Based on B. Girod and M. van der Schaar's lecture notes

Motion JPEG

- JPEG system for compressing static images could be applied to a sequence of images, compressing each individually, this is called ***motion JPEG***
- Motion JPEG takes no advantage of any correlation between successive images
- In a typical scene there will be a great deal of similarity between nearby images of the same sequence.

In a typical scene there will be a great deal of similarity between nearby images of the same sequence.

Motion Compensation Approach

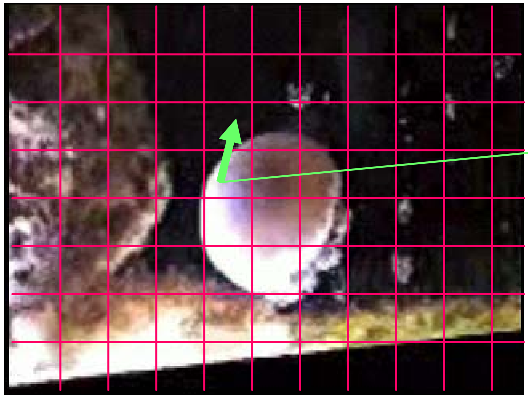
Basic idea of ***Motion Compensation***:

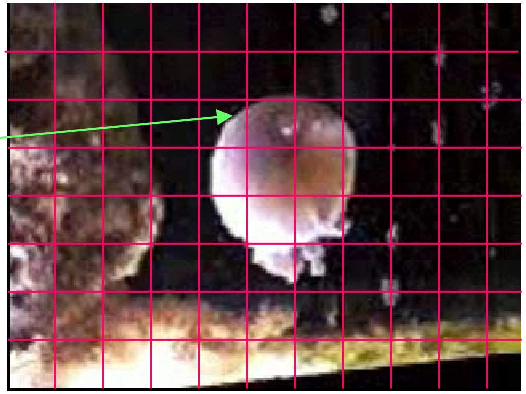
- Many “moving” images or image sequences consist of a static background with one or more moving foreground objects. We can get coding advantage from this.
- we code the first frame by baseline JPEG and use this frame as ***reference image***.
- Treat the second image block by block and compare each block with the same block in the reference image.
- For blocks that have identical block in reference image, we only send a special code instead of whole code.
- For other blocks, we just encode them as usual.

Motion Compensation Approach(cont.)

Motion Vectors

- static background is a very special case, we should consider the displacement of the block.
- Motion vector is used to inform decoder exactly where in the previous image to get the data.
- Motion vector would be zero for a static background.





Motion Compensation Approach(cont.)

Block Matching--how to find the matching block?

- Matching criteria:
  - In practice we couldn't expect to find the exactly identical matching block, instead we look for close match.
  - Most motion estimation schemes look for minimum mean square error(MMSE) between block.

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(I_n(x,y) - I_n^{'}(x,y))^2$$

- Matching block size:
  - How large the matching block will affect coding efficiency
  - block size MPEG used: 16×16

Motion Compensation Approach(cont.)

Search range:

- It's reasonable to consider an displacement of 360 pixels/s or about 60pixels/image in standard-definition television.
- In real-world scenes there is usually more or faster motion horizontally than vertically, generally the width of search area should be twice the height.
- Suggested search range: ± 60 pixles $\times$ 30 pixles

Motion Compensation Approach(cont.)

Residuals

- The differences between the block being coded and it's best match are known as residuals.
- The residuals maybe encoded and transmitted along with the motion vector, so the decoder will be able to reconstruct the block.
- We should compare the bits of transmitting the motion vector plus the residuals with the bits of transmitting the block itself and use the most efficient mechanism.

MPEG-1 Introduction

- MPEG: Moving Pictures Experts Group.
- MPEG-video is addressing the compression of video signals at about 1.5Mbits/s
- MPEG-1 is asymmetric system, the complexity of the encoder is much higher than that of the decoder.

Horizontal picture size	$\leq$768 pixels
Vertical picture size	$\leq$ 576 lines
Number of macroblocks	$\leq$ 396
Number of macroblocks $\times$ picture rate	$\leq 396 \times 25 = 9900$
Picture rate	$\leq$ 30 pictures/s
VBV buffer size	$\geq$ 2,621,440 bits
Bit rate	$\leq$ 1,856,000 bits/s

Table 1: MPEG-1 Constraints

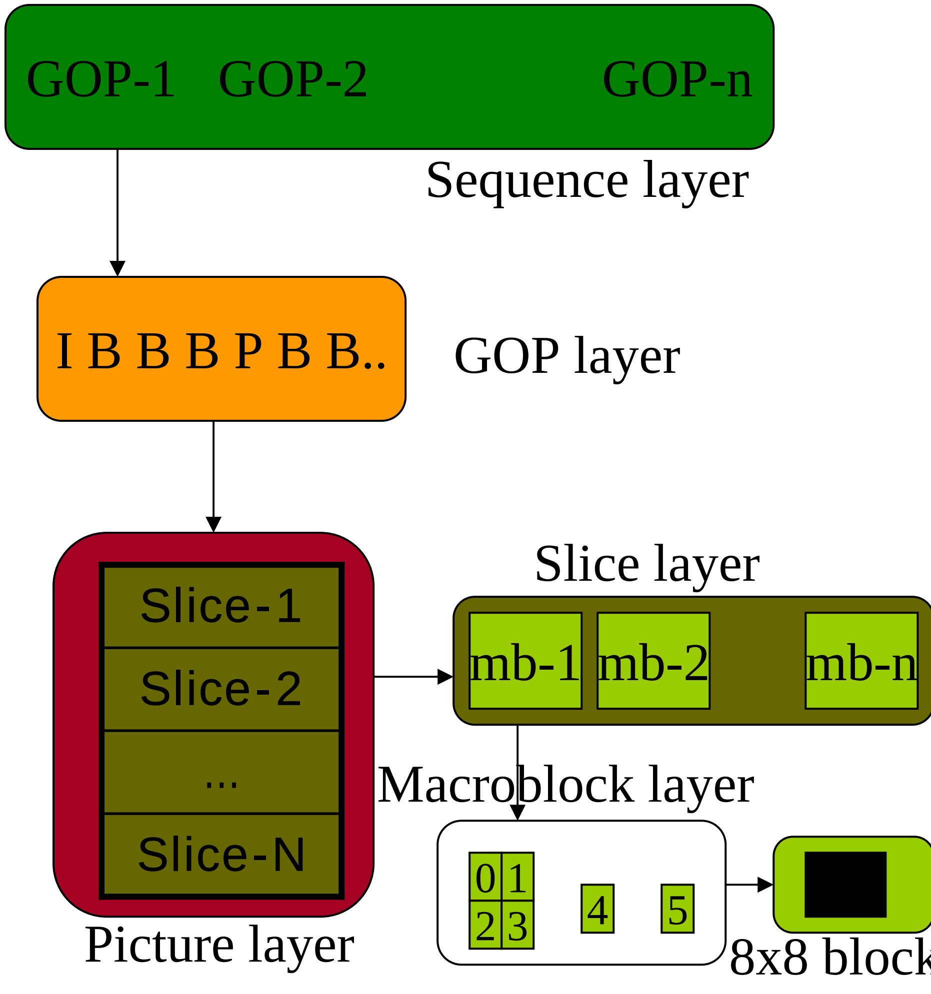
MPEG Hierarchy

The six layers of MPEG video bit stream

- Sequence Layer: video clip, complete program item.
- Group of Pictures Layer(GOP): include three different coding ways.
- Frame Layer
- Slice Layer: in case the data is lost or corrupted.
- Macroblock Layer: 16×16 luminance block.
- Block Layer(DCT unit)

MPEG: Structure of the Coded Bit-Stream

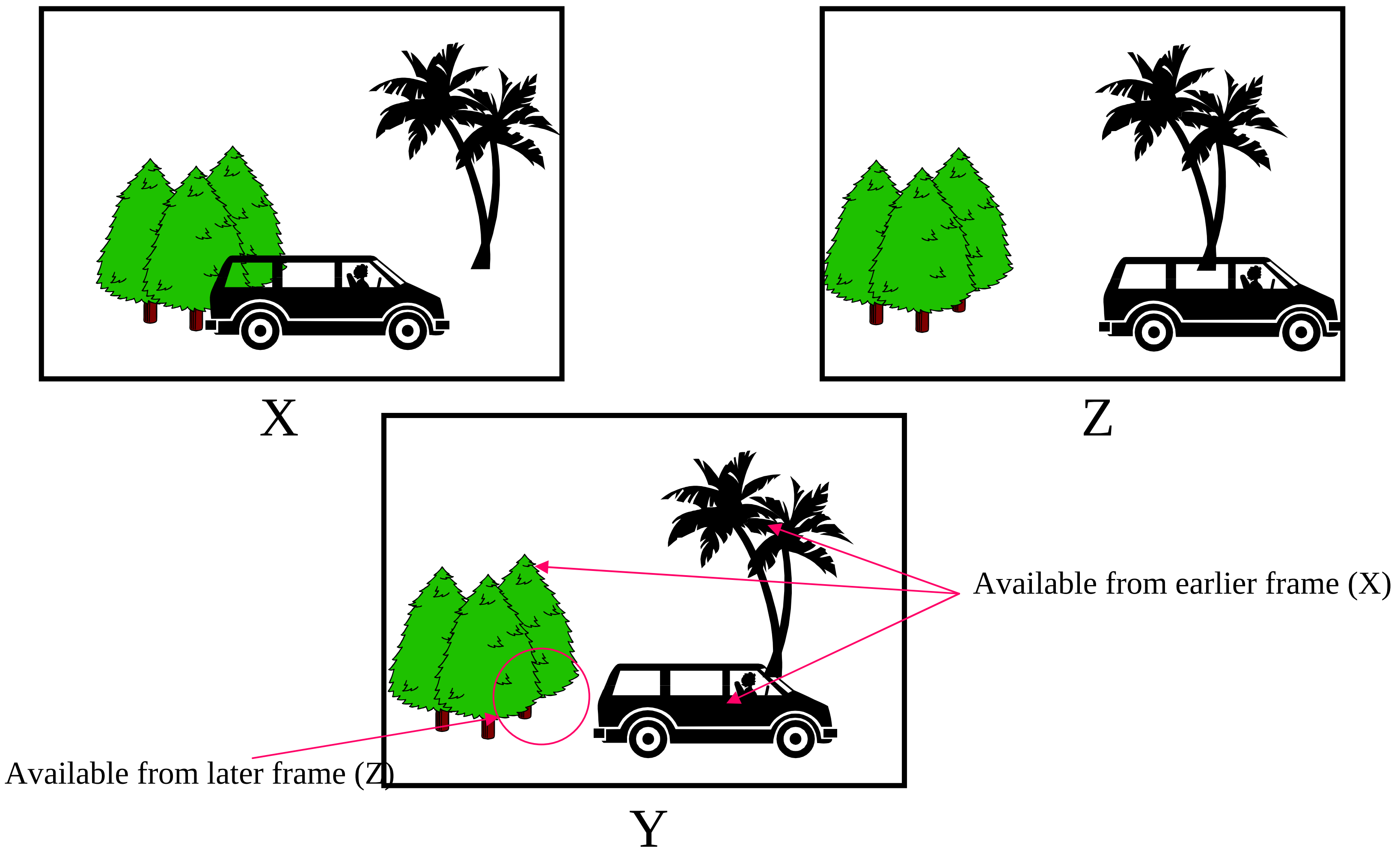
- **Sequence layer**: picture dimensions, pixel aspect ratio, picture rate, minimum buffer size, DCT quantization matrices
- **GOP layer**: will have one I picture, start with I or B picture, end with I or P picture, has closed GOP flag, timing info, user data
- **Picture layer**: temporal ref number, picture type, synchronization info, resolution, range of motion vectors
- **Slices**: position of slice in picture, quantization scale factor
- **Macroblock**: position, H and V motion vectors, which blocks are coded and transmitted



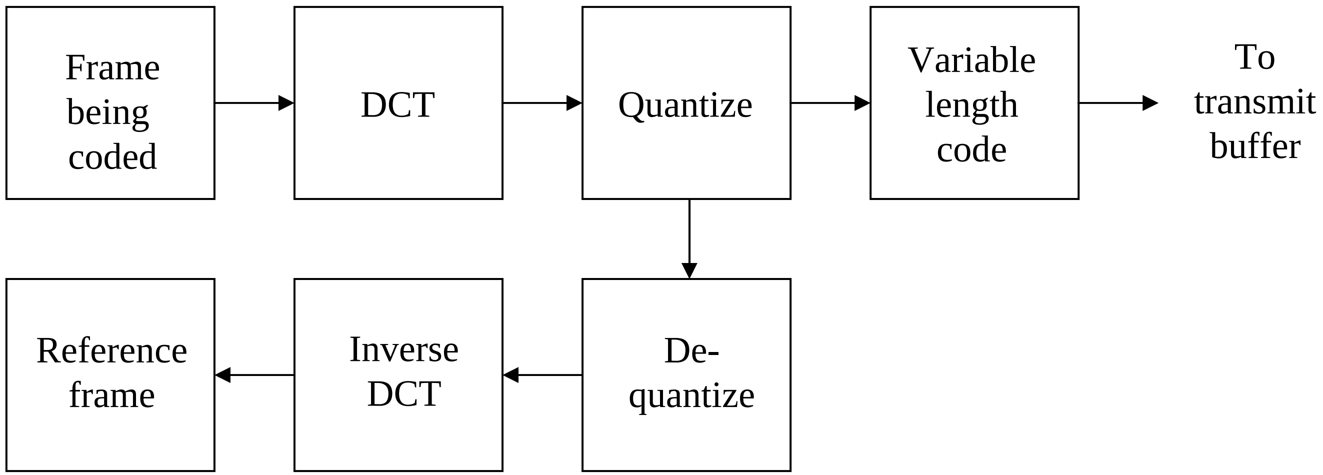
Frame Types in MPEG

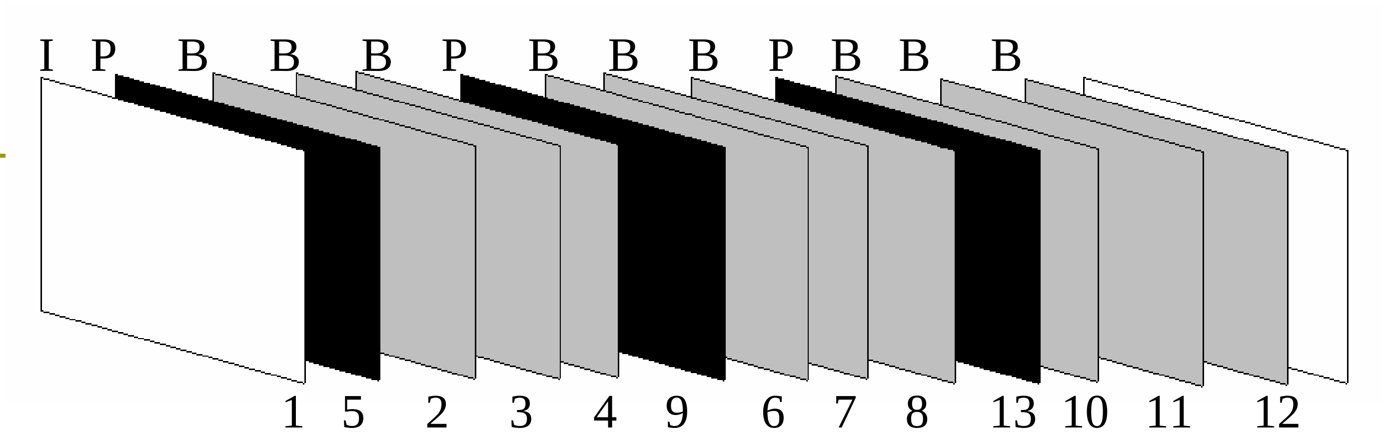
- Intra frames (I-frames)
  - A I-frame is encoded using only information from within that frame(intra coded) -- no temporal compression(inter coded).
- Non-intra frames (P-frames and B-frames)
  - motion compensated information will be used for coding.
  - P frame (predicted frame) use preceding frame as reference image
  - B frame (bidirectional frame) use both preceding frame and following frame as reference images

Motion estimation for different frames

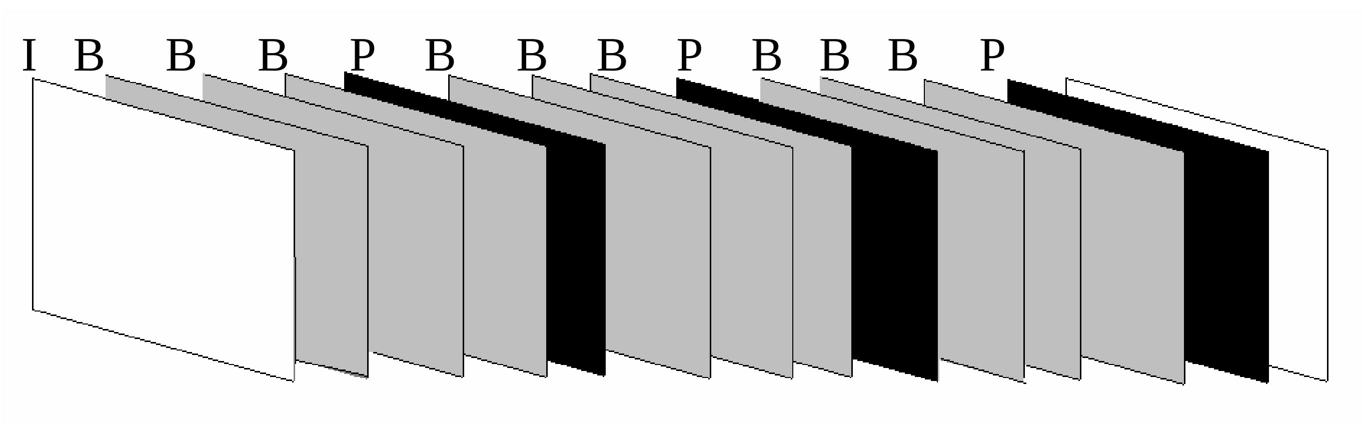


Coding I Frame





A typical group of pictures in coding order



A typical group of pictures in display order

Coding of Macroblock (cont.)

Intra coding of macroblocks

– just as what JPEG does

– MPEG has two default quantization tables, one for intra coding, another one for non-intra coding of residuals

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

MPEG quantization table(for intra coding)

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

JPEG quantization table(luminance)

Coding of Macroblock (cont.)

Non-intra coding of macroblocks

- The first step is to intra code the macroblock--just in case if we fail to find a reasonable match in motion estimation.

- Then we use motion estimation to find the nearest match and get the motion vector. Only luminance samples are used in motion estimation.

- Then each DCT block in macroblock will be treated separately. The residuals will be encode by DCT and quantization (use flat table) as in intra coding. DC along with AC.

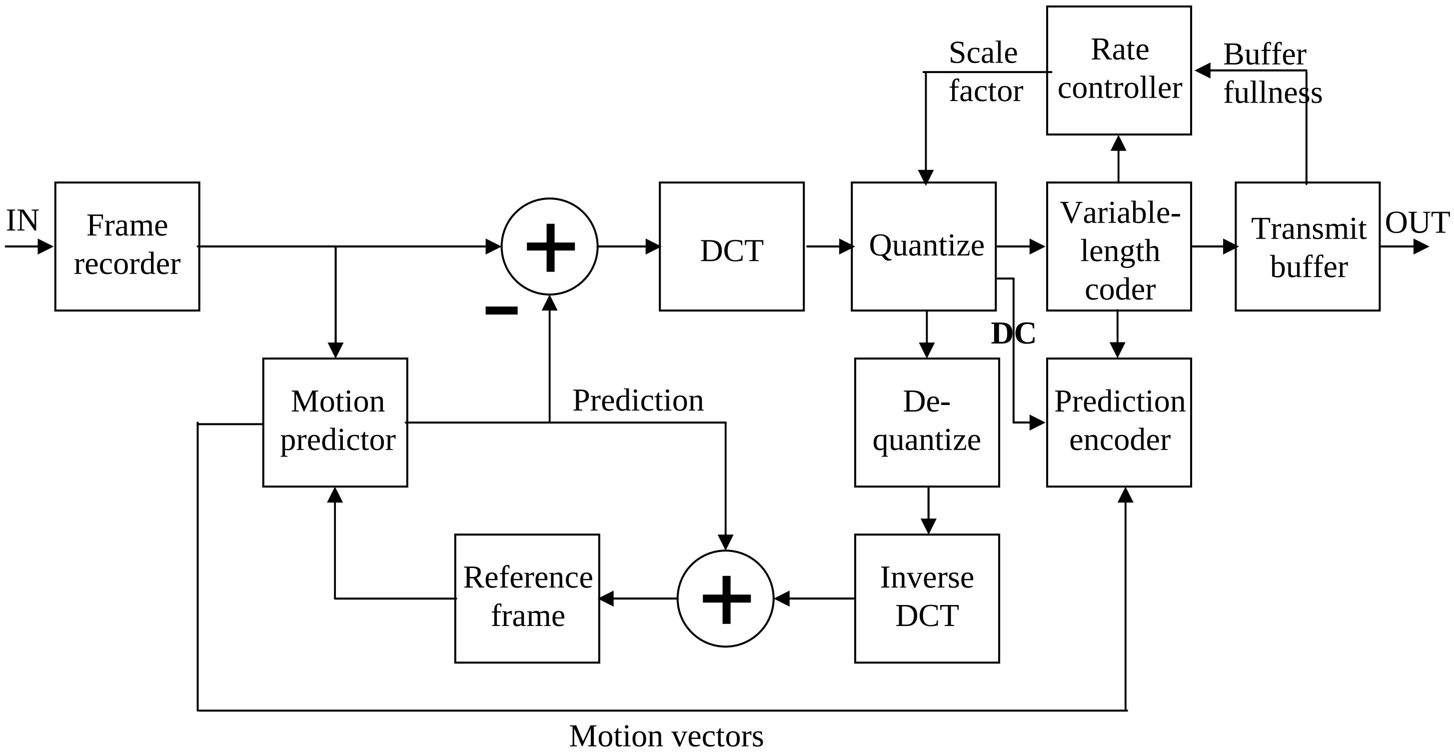
- Motion vectors are coded predictively

Coding of Macroblock (cont.)

- P-frames
  - If the block can be skipped, we just send a “skip” code
  - otherwise, we compare the number of total bits of inter and intra coding, choose the more efficient one. Mark this block accordingly.

- B-frames
  - Comparison among three methods of encoding

A Simplified MPEG encoder



MPEG-2

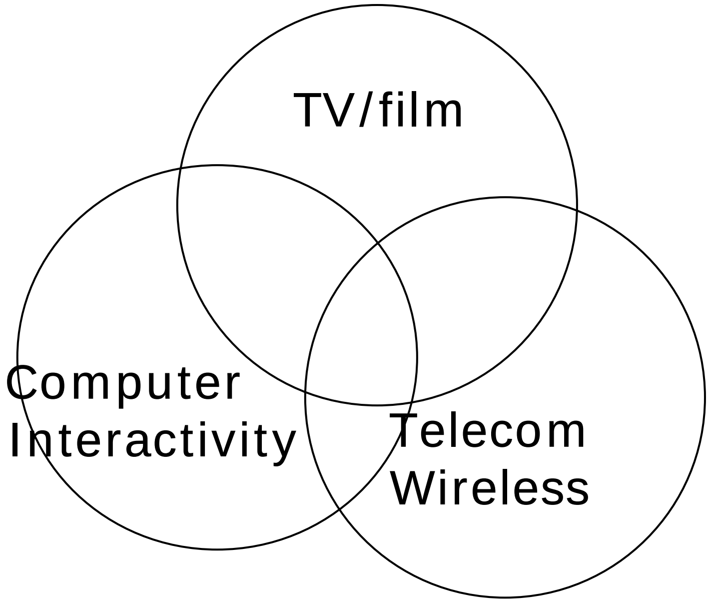
- Why another standard?
  - Support higher bit rates e.g., 80-100 Mbits/s for HDTV instead of the 1.15 Mbits/s for SIF
  - Support a larger number of applications
  - The encoding standard should be a toolkit rather than a flat procedure
    - Interlaced and non-interlaced frame
    - Different color subsampling modes e.g., 4:2:2, 4:2:0, 4:4:4
    - Flexible quantization schemes – can be changed at picture level
    - Scalable bit-streams
    - Profiles and levels

MPEG-2 Applications

- Digital Betacam: 90 Mbits/s video
- MPEG-2
  - Main Profile, Main Level, 4:2:0: 15 Mbits/s
  - High Profile, High Level, 4:2:0: adequate, expensive
  - Image quality preserved across generations of processing
  - Multiview Profile
    - Stereoscopic view – disparity prediction
    - Virtual walk-throughs composed from multiple viewpoints

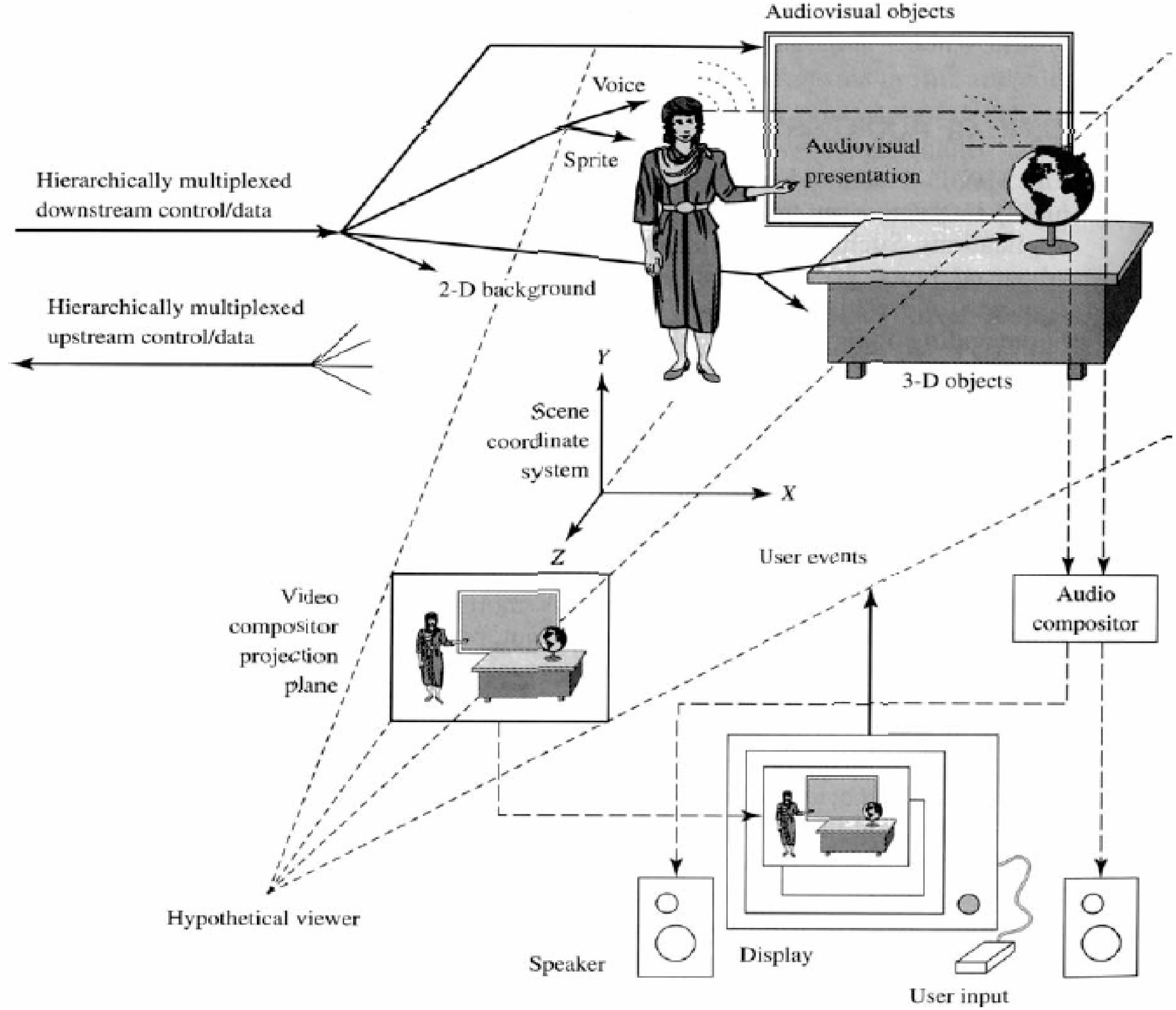
MPEG-4

- Support highly interactive multimedia applications as well as traditional applications
- Advanced functionalities: interactivity, scalability, error resilience, …
- Coding of natural and synthetic audio and video, as well as graphics
- Enable the multiplexing of audiovisual objects and composition in a scene.



- Video on LANs, Internet video
- Wireless video
- Video database
- Interactive home shopping
- Video e-mail, home video
- Virtual reality games, multi-viewpoint training

MPEG-4



Audiovisual objects
Voice
Sprite
Audiovisual presentation
Hierarchically multiplexed downstream control/data
2-D background
Hierarchically multiplexed upstream control/data
3-D objects
Y
Scene coordinate system
X
Z
User events
Audio compositor
Video compositor projection plane
Hypothetical viewer
Speaker
Display
User input

MPEG-4 Video Coding

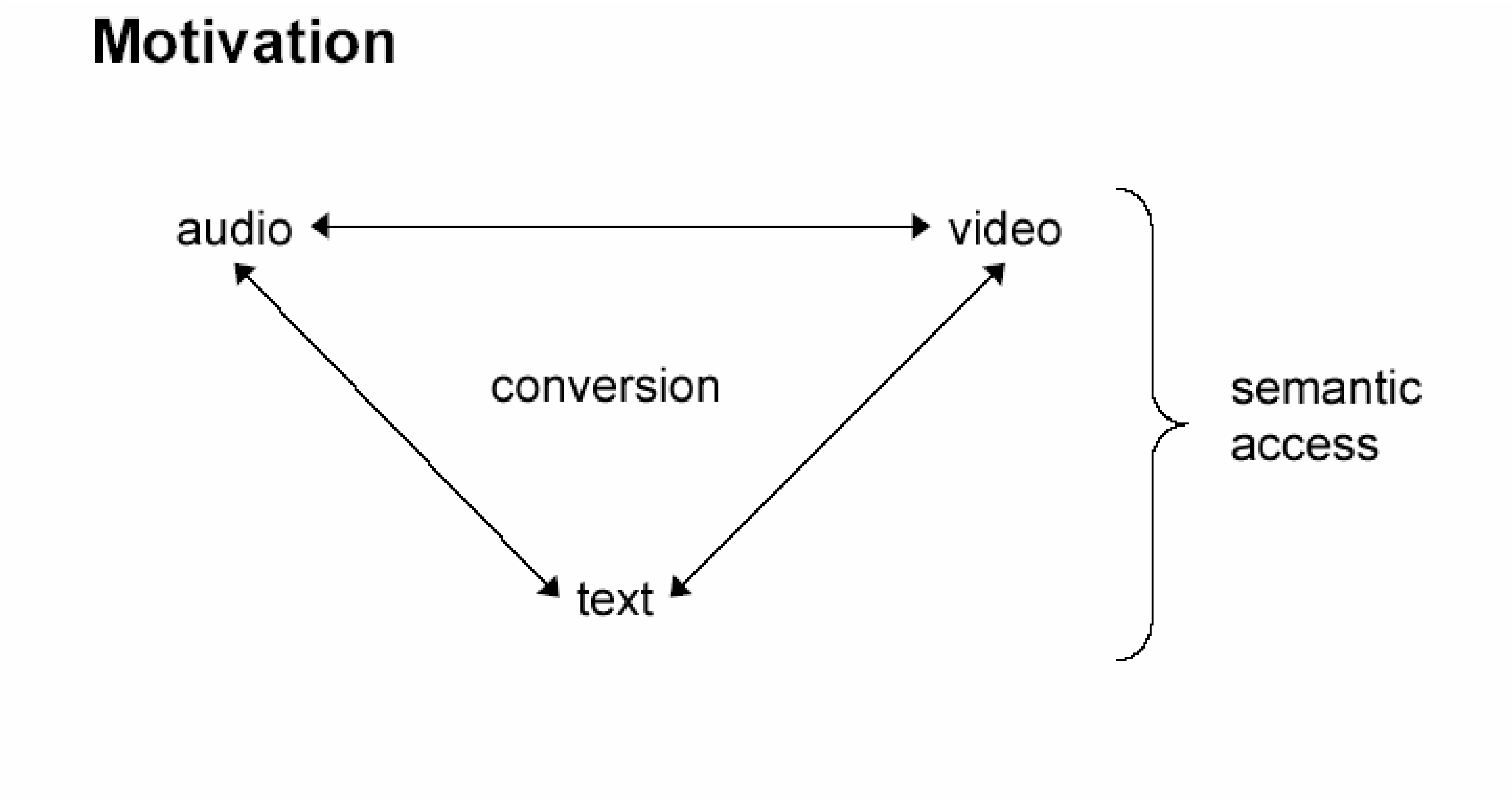
- Basic video coding
  - Definition of Video Object (VO), Video Object Layer (VOL), Video Object Plane (VOP)
  - Improved coding efficiency vs. MPEG-1 and MPEG-2
  - Based on H.263 baseline
  - Global motion compensation
  - Sprite
  - Quarter pixel motion compensation

MPEG-4: Video coding

- Object-based video coding
  - Binary shape coding
  - Greyscale shape coding
  - Padding for block-based DCT of texture
  - Shape-adaptive DCT
- DWT for still texture coding
- Mesh animation, face and body animation

MPEG-7

- Semantic Annotationfor Motion Pictures



Application Scenarios

- Image Understanding (surveillance)
- Intelligent Vision
- Smart Camera/VCRs
- Information Retrieval
- Information Filtering

- Computer-based training

MPEG-7

MPEG Family

- MPEG-1
- MPEG-2

} frame-based encoding of waveforms

- MPEG-4 object-based representations (interactive representations)

Note: MPEG-3, 5 & 6 do not exist

MPEG-7

- Think of MPEG-7 as a meta structure to allow the decoder to access to the semantic the videos.