CS 598KN

Advanced Multimedia Systems Design

Lecture 3 – Image and Video (MPEG) Coding

Klara Nahrstedt

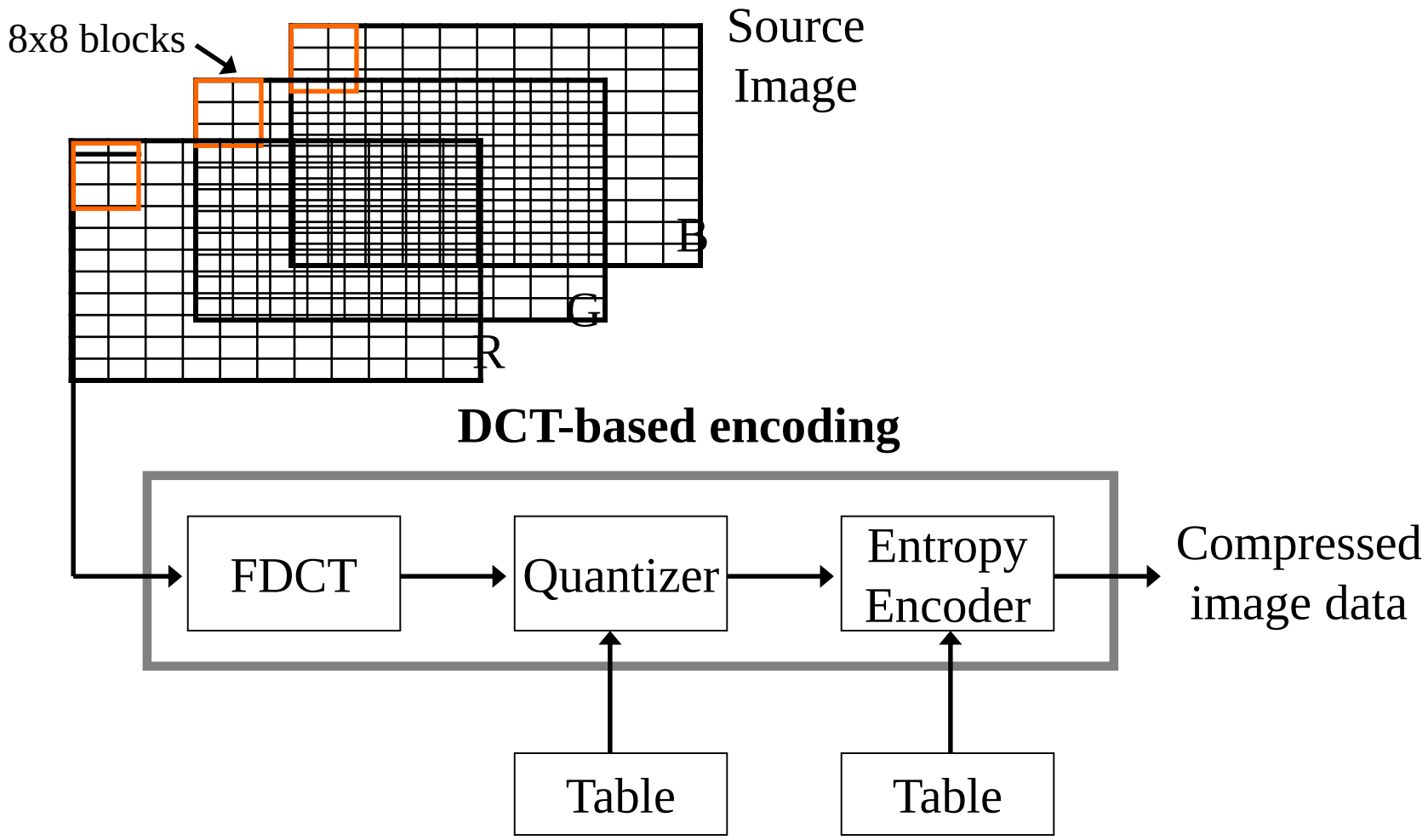
Fall 2017

Overview

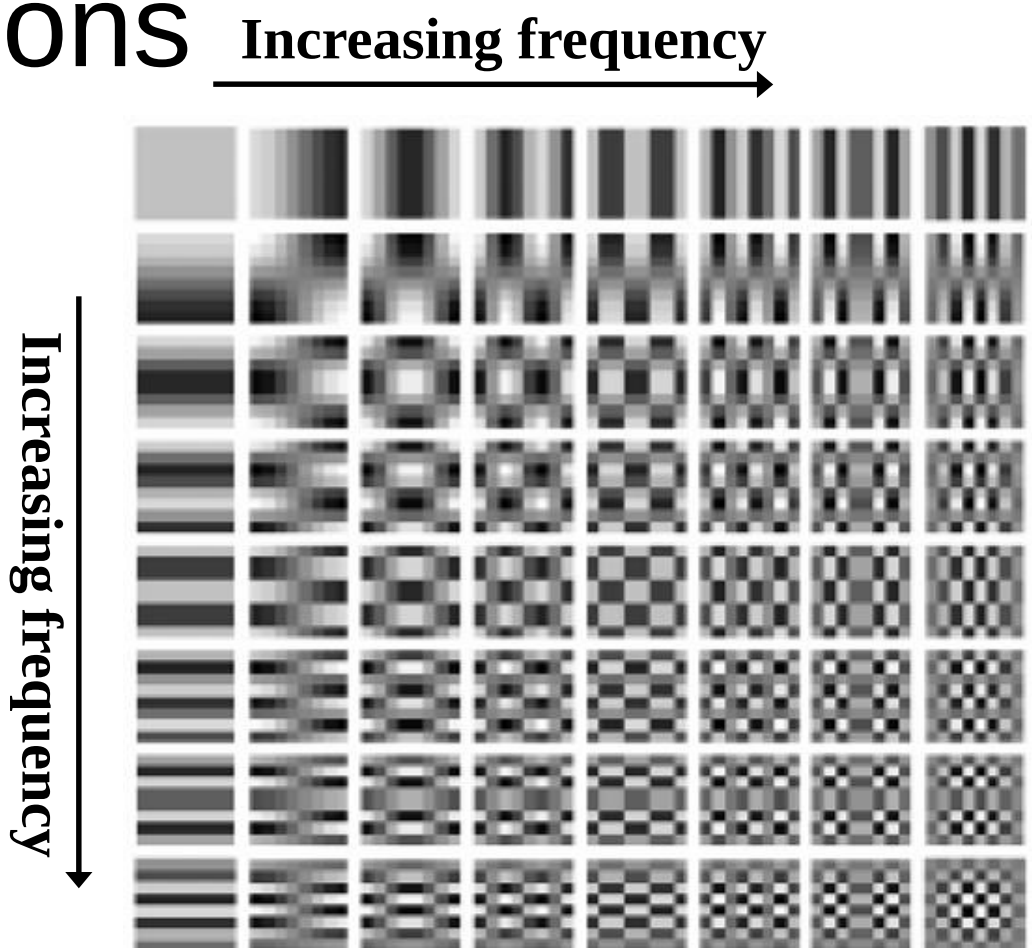
- JPEG Compression
- MPEG Basics
- MPEG-4
- MPEG-7

JPEG COMPRESSION

JPEG Compression



Visualization of Basic FDCT Functions



Coefficient Differentiation

- F(0,0)
  - includes the lowest frequency in both directions
  - is called **DC coefficient**
  - Determines fundamental color of the block
- F(0,1) …. F(7,7)
  - are called **AC coefficients**
  - Their frequency is non-zero in one or both directions

Quantization

- Throw out bits
- Consider example: 101101_2 = 45 (6 bits)
  - We can truncate this string to 4 bits: 1011_2 = 11
  - We can truncate this string to 3 bits: 101_2 = 5 (original value 40) or 110_2 = 6 (original value 48)
- Uniform quantization is achieved by dividing DCT coefficients by N and round the result (e.g., above we used N=4 or N=8)
- In JPEG – use quantization tables
  - Fq(u,v) = F(u,v)/Quv
  - Two quantization tables – one for luminance and one for two chrominance components

De facto Quantization Table

Eye becomes less sensitive

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

Eye becomes less sensitive

Entropy Encoding

- Compress sequence of quantized DC and AC coefficients from quantization step
  - further increase compression, without loss
- Separate DC from AC components
  - DC components change slowly, thus will be encoded using difference encoding

DC Encoding

- DC represents average intensity of a block
  - encode using difference encoding scheme
  - use 3x3 pattern of blocks

- Because difference tends to be near zero, can use less bits in the encoding
  - categorize difference into difference classes
  - send the index of the difference class, followed by bits representing the difference

Difference Coding applied to DC Coefficients

PREDICTOR

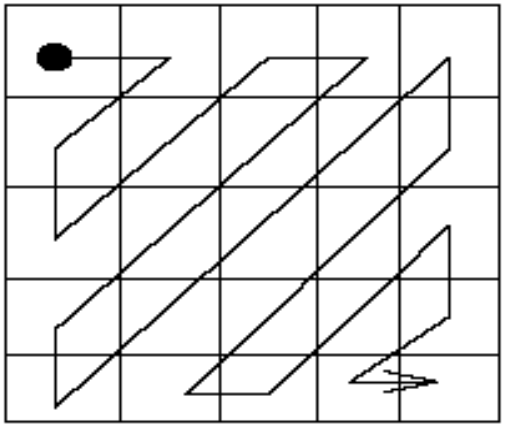
$$Diff_i = DC_i - DC_{i-1} \quad i>0$$

DC_0	DC_1	DC_2
DC_3	DC_4	DC_5
DC_6	DC_7	DC_8

DC_0	$Diff_1$	$Diff_2$
$Diff_3$	$Diff_4$	$Diff_5$
$Diff_6$	$Diff_7$	$Diff_8$

AC Encoding

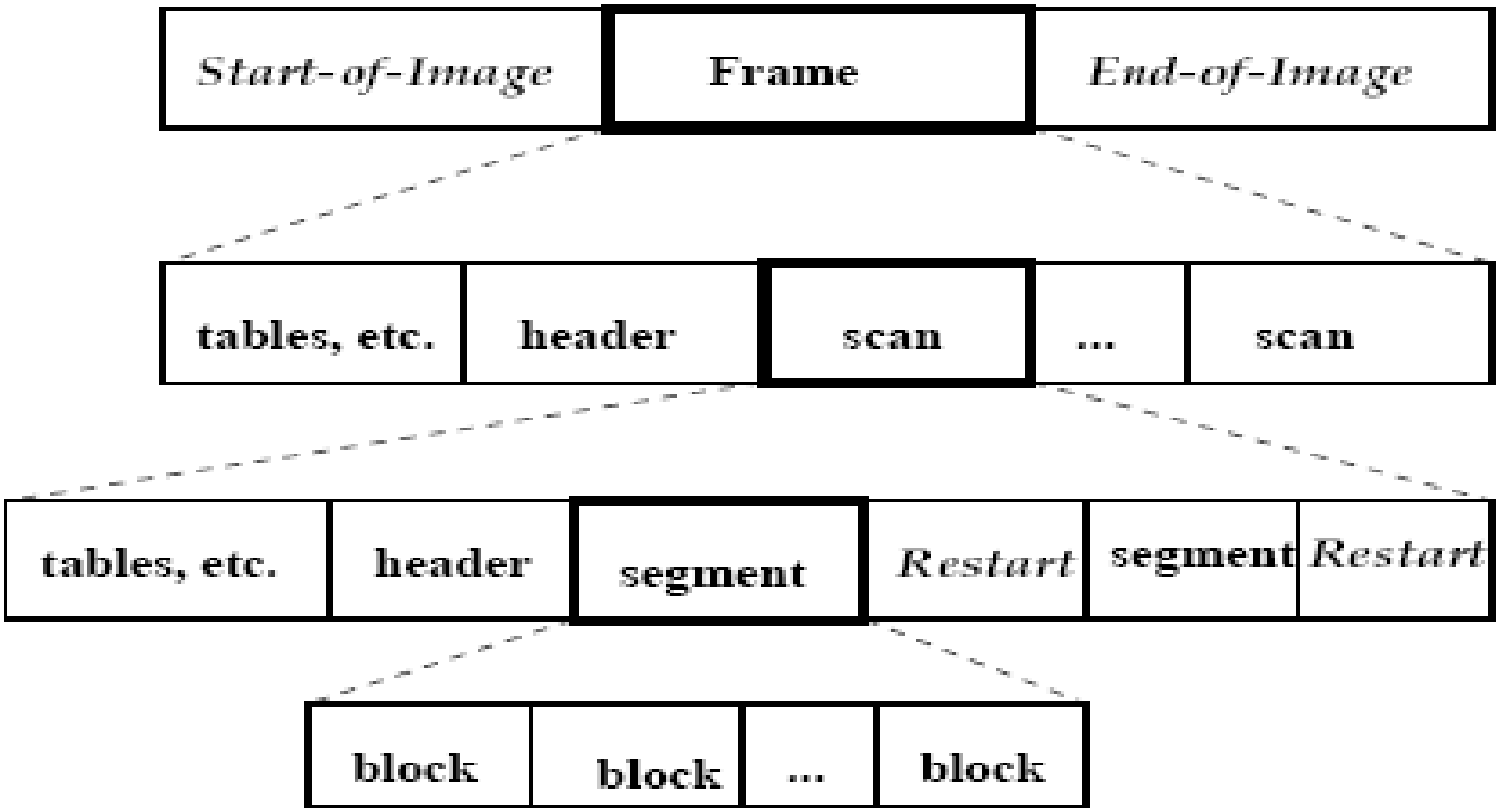
- Use zig-zag ordering of coefficients
  - orders frequency components from low->high
  - produce maximal series of 0s at the end
  - Ordering helps to apply efficiently entropy encoding
- Apply Huffman coding
- Apply RLE on AC zero values



Huffman Encoding

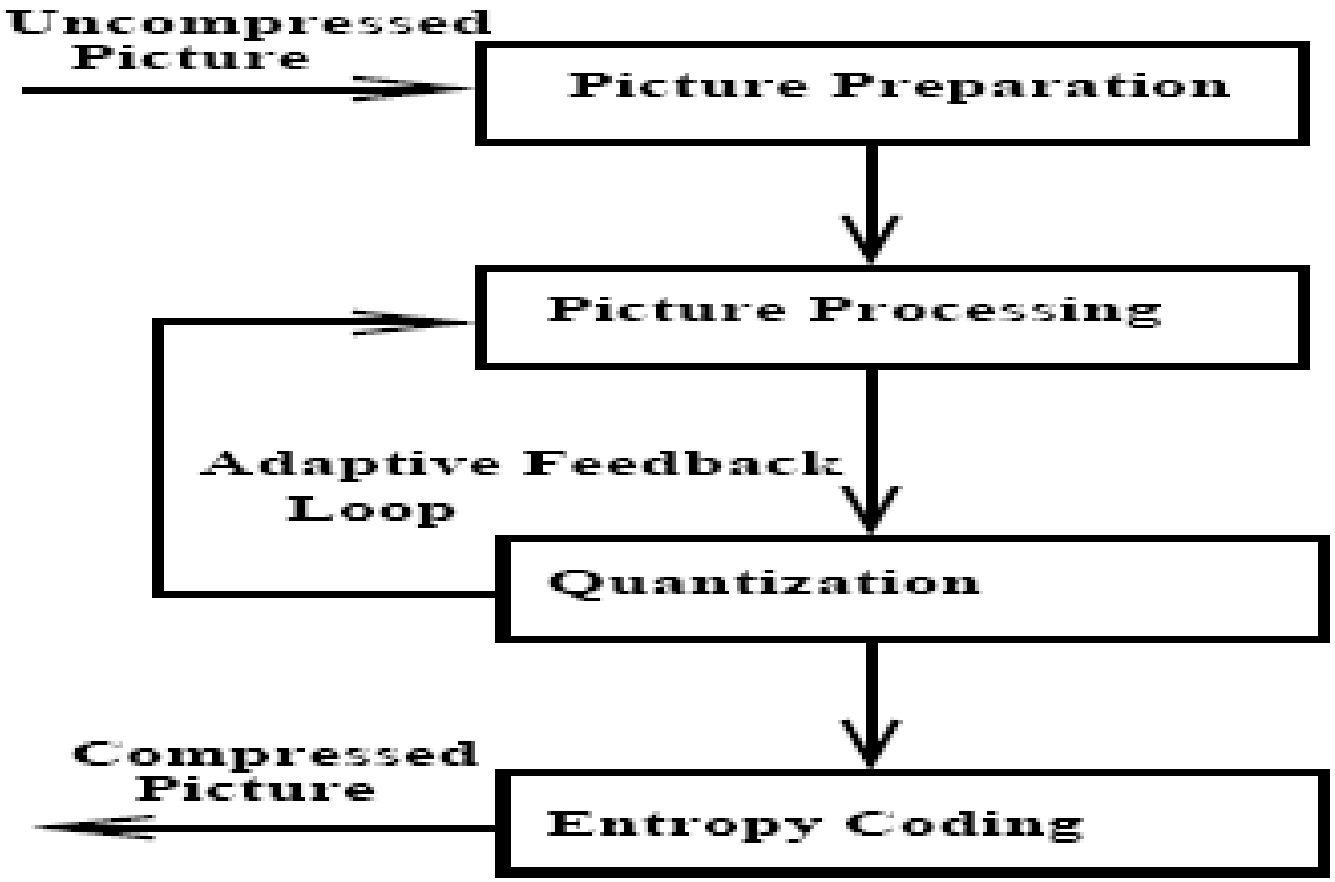
- Sequence of DC difference indices and values along with RLE of AC coefficients
- Apply Huffman encoding to sequence
- Attach appropriate headers
- Finally have the JPEG image!

Interchange Format of JPEG



VIDEO MPEG-2 COMPRESSION FORMAT (BASICS)

Video MPEG Compression is Hybrid Coding



RLE, Huffman Arithmetic Coding

MPEG General Information

- Goal: data compression
- MPEG (MPEG-1, MPEG-2, MPEG-4) defines video, audio coding and system data streams with synchronization
- MPEG information
  - Aspect ratios: 1:1 (CRT), 4:3 (NTSC), 16:9 (HDTV)
  - Refresh frequencies: 23.975, 24, 25, 29.97, 50, 59.94, 60 Hz

MPEG Image Preparation (Resolution and Dimension)

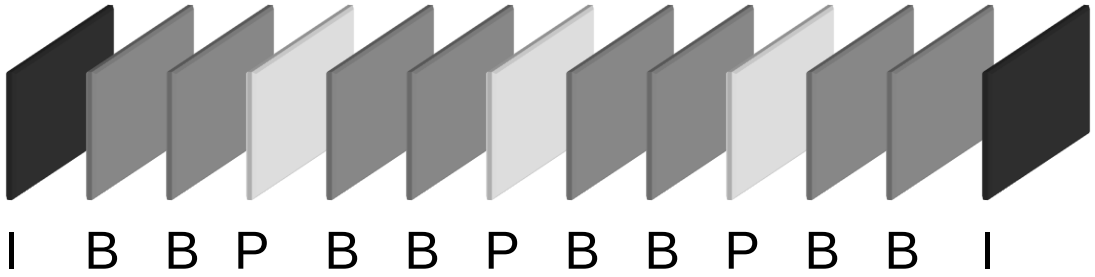
- MPEG defines exactly format
  - Three components: Luminance and two chrominance components (2:1:1) YCbCr
  - YCbCr is family of color spaces, Y is luma, Cb and Cr are blue-difference and red-difference chroma component
  - Resolution of luminance comp:X1 ≤ 768; Y1 ≤ 576 pixels
  - Pixel precision is 8 bits for each component
- Example of Video format: 352x240 pixels, 30 fps; chrominance components: 176x120 pixels

MPEG Image Preparation - Blocks

- Each image is divided into **macro-blocks**
- Macro-block : 16x16 pixels for luminance; 8x8 for each chrominance component
- Macro-blocks are useful for Motion Estimation

MPEG Video Processing

- Intra frames (same as JPEG)
  - typically about 12 frames between I frames
- Predictive frames
  - encode from previous I or P reference frame
- Bi-directional frames
  - encode from previous and future I or P frames

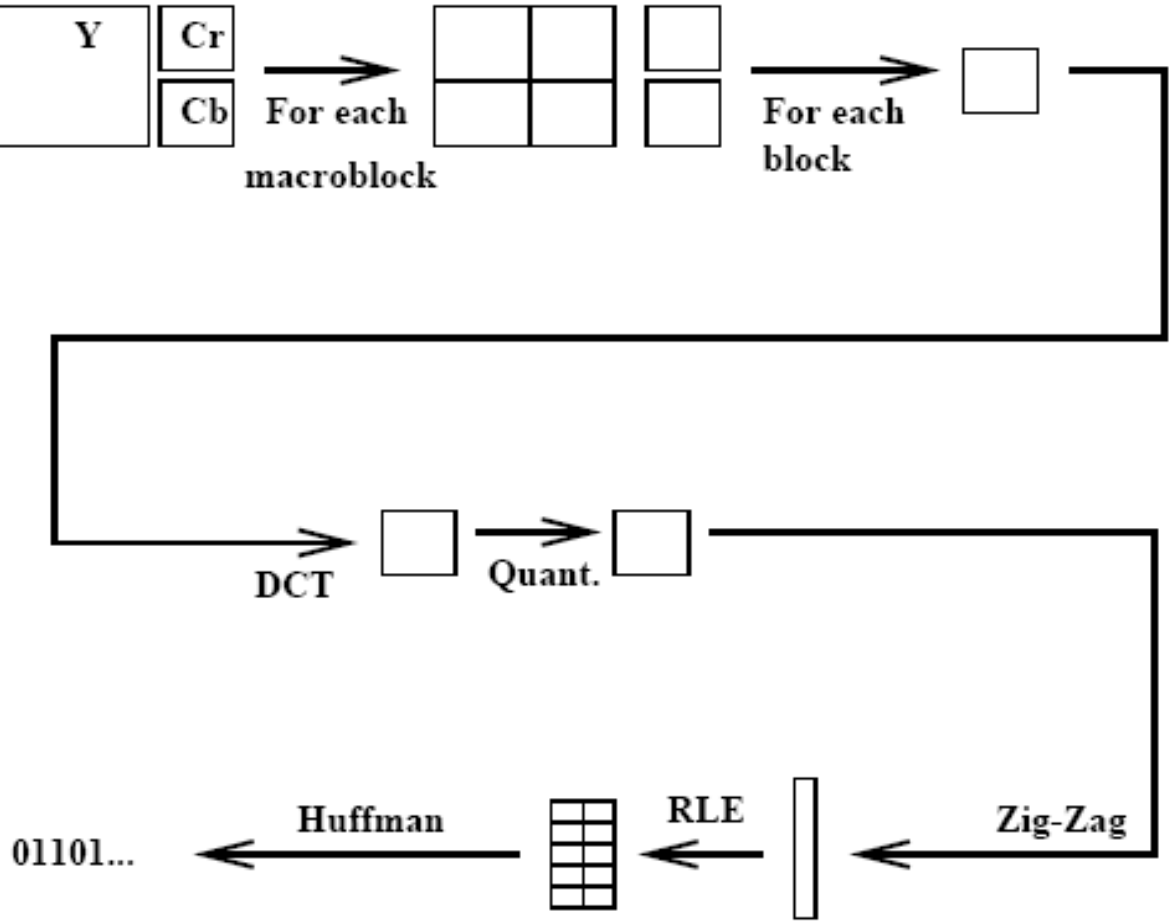


Selecting I, P, or B Frames

- Heuristics
  - change of scenes should generate I frame
  - limit B and P frames between I frames
  - B frames are computationally intense

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

MPEG Video I-Frames



Intra-coded images

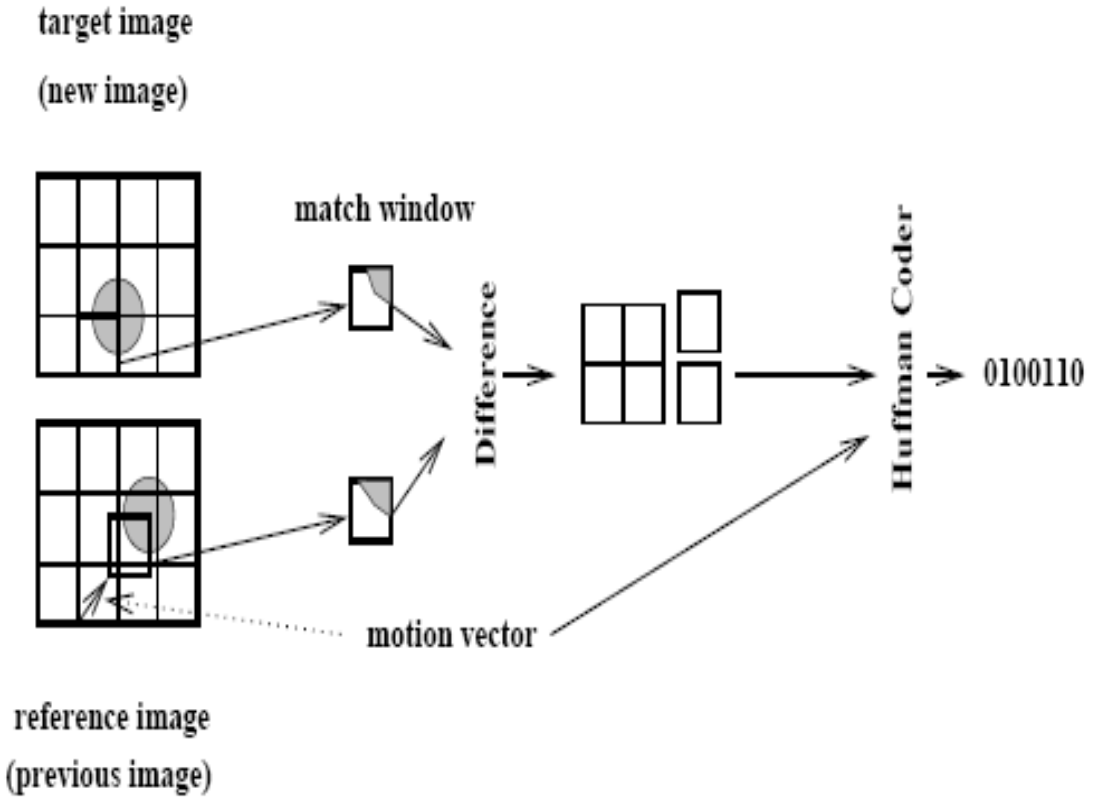
I-frames – points of random access in MPEG stream

I-frames use 8x8 blocks defined within Macro-block

No quantization table for all DCT coefficients, only quantization factor

MPEG Video P-Frames

Motion Estimation Method



Predictive coded frames require information of previous I frame and or previous P frame for encoding/decoding

For **Temporary Redundancy** we determine last P or I frame that is most similar to the block under consideration

Matching Methods

- SSD metric

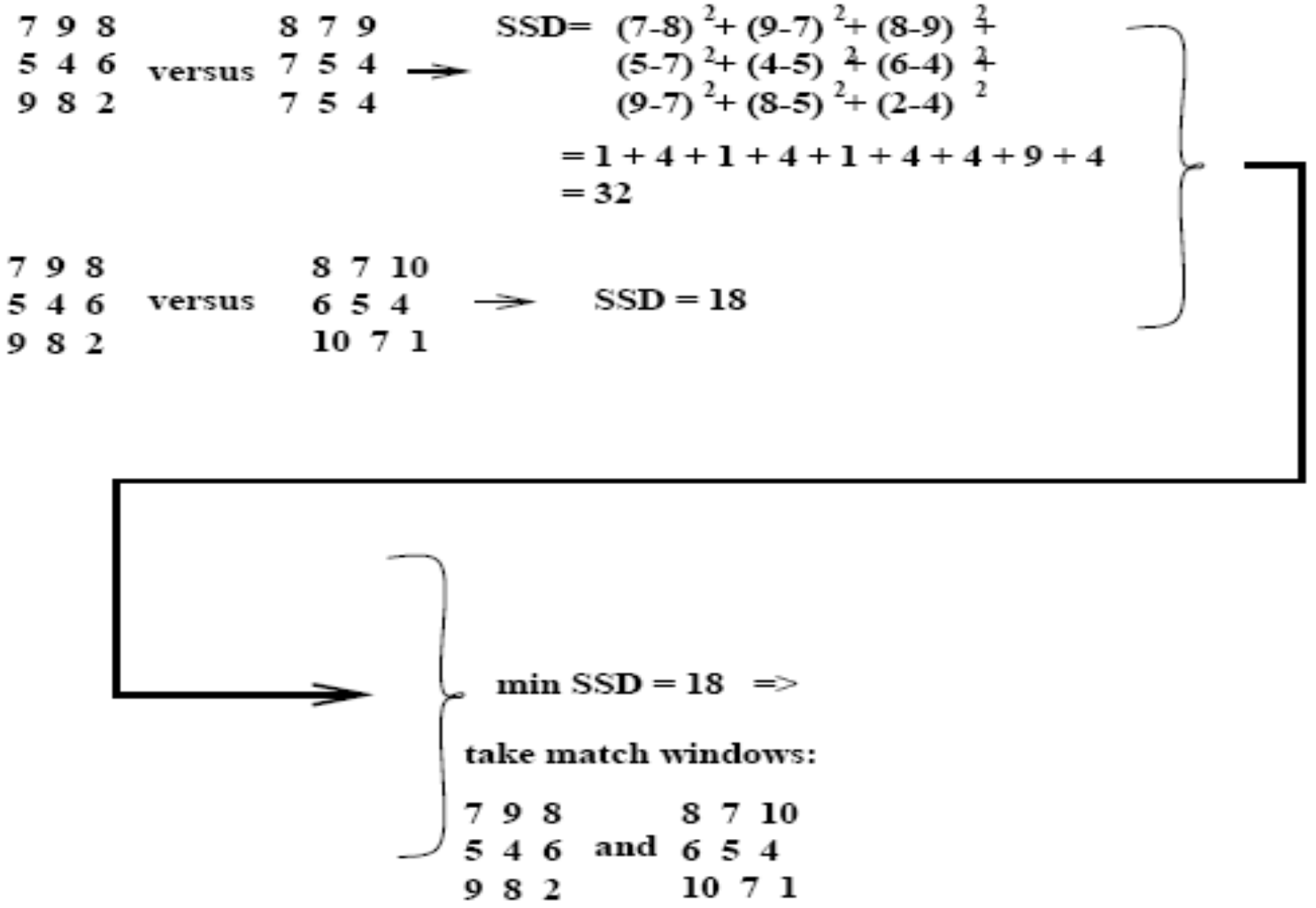
$$SSD = \sum_{i=0}^{N-1} (x_i - y_i)^2$$

- SAD metric

$$SAD = \sum_{i=0}^{N-1} | x_i - y_i |$$

- Minimum error represents best match
  - must be below a specified threshold
  - error and perceptual similarity not always correlated

Example of Finding Minimal SSD



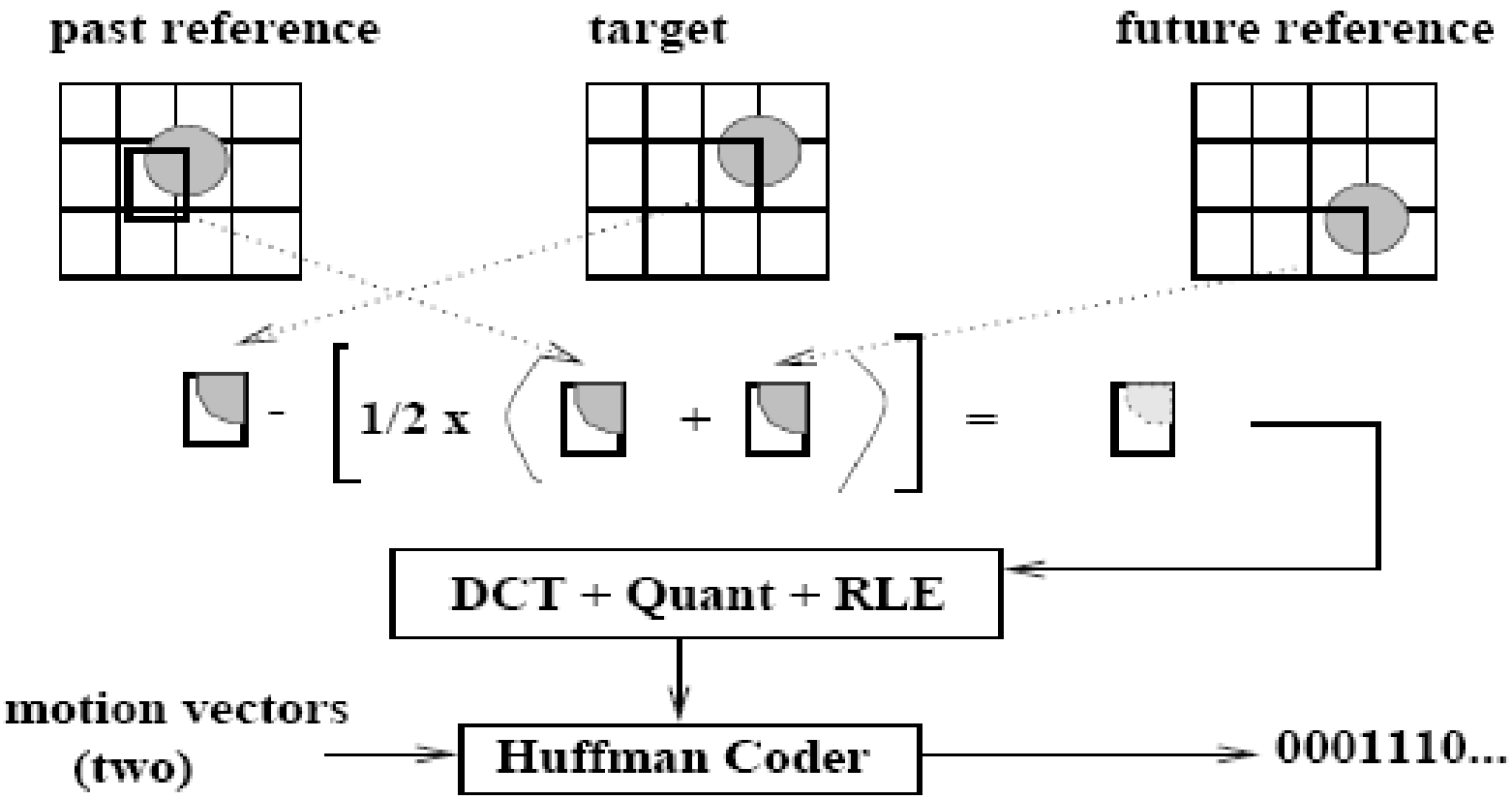
Syntax of P Frame

Addr	Type	Quant	Motion Vector	CBP	b0	b1	···	b5

Addr: address the syntax of P frame
Type: INTRA block is specified if no good match was found
Quant: quantization value per macro-block (vary quantization to fine-tune compression)
Motion Vector: a 2D vector used for motion compensation provides offset from coordinate position in target image to coordinates in reference image
CBP(Coded Block Pattern): bit mask indicates which blocks are present

MPEG Video B Frames

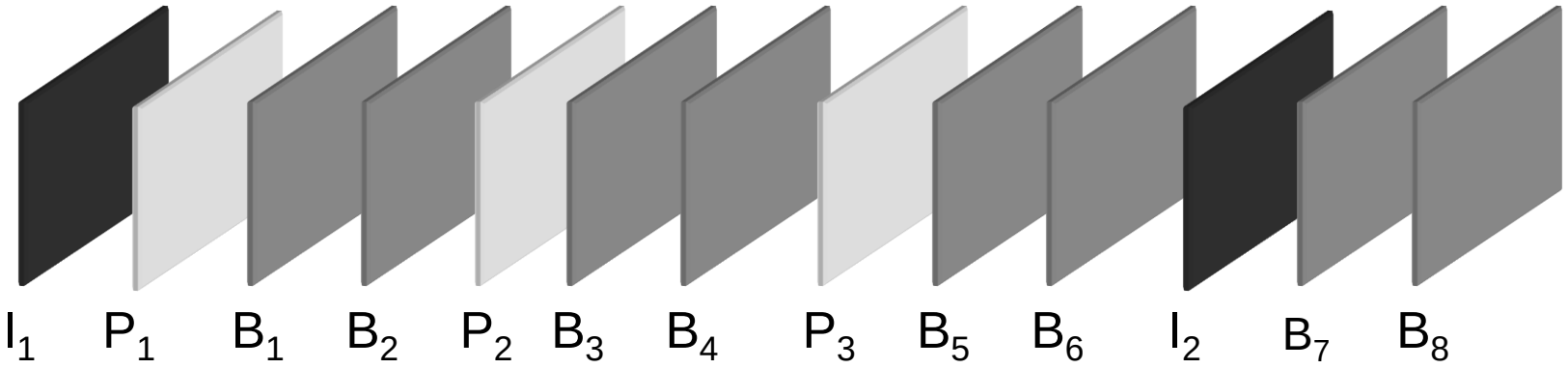
Bi-directionally Predictive-coded frames



MPEG Video Decoding

Display Order

Decoding Order

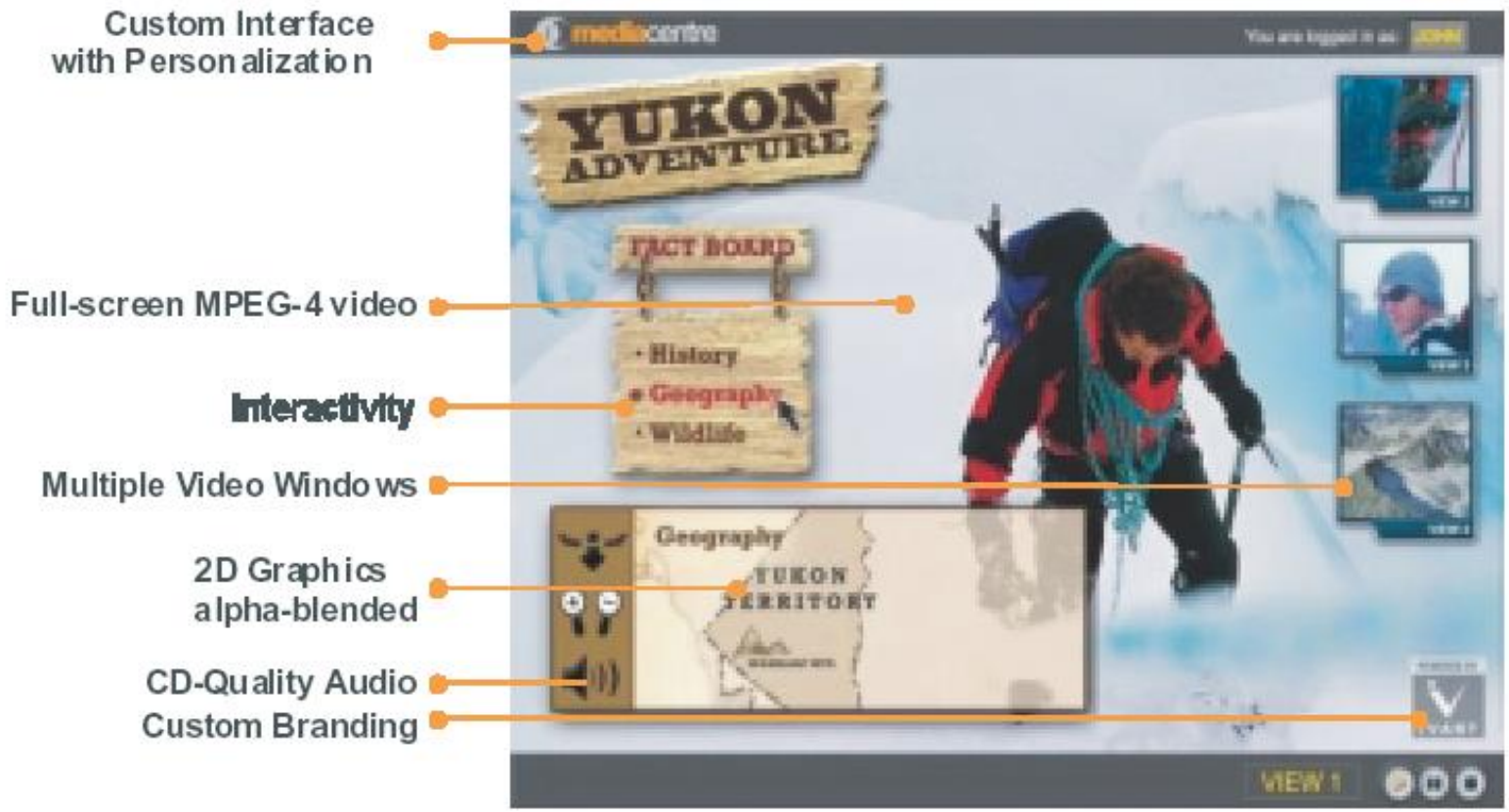


MPEG Video Quantization

- AC coefficients of B/P frames are usually large values, I frames have smaller values
  - Adjust quantization
- If data rate increases over threshold, then quantization enlarges step size (increase quantization factor Q)
- If data rate decreases below threshold, then quantization decreases Q

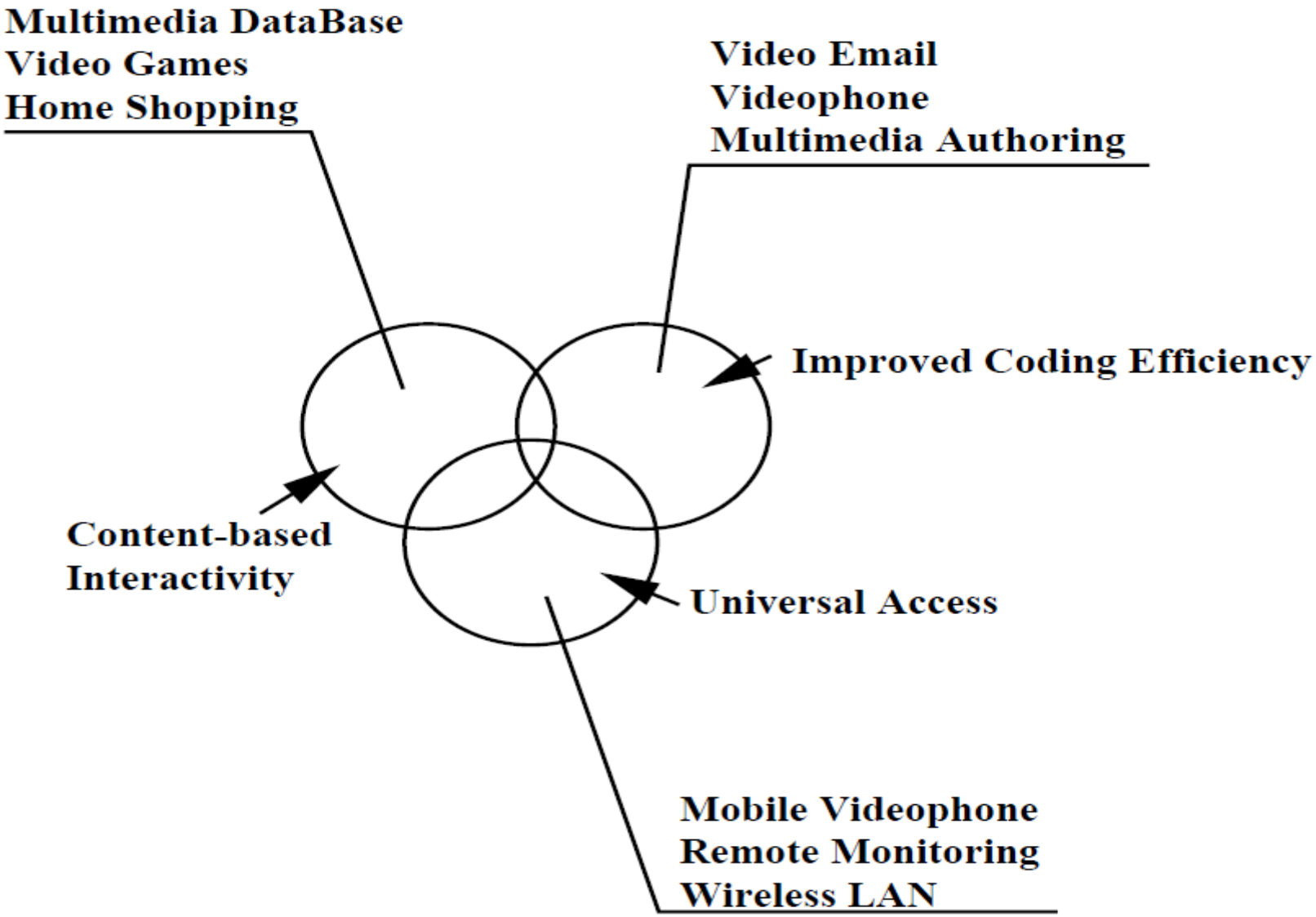
MPEG-4

MPEG-4 Example



ISO N3536 MPEG4

MPEG-4 Characteristics and Applications



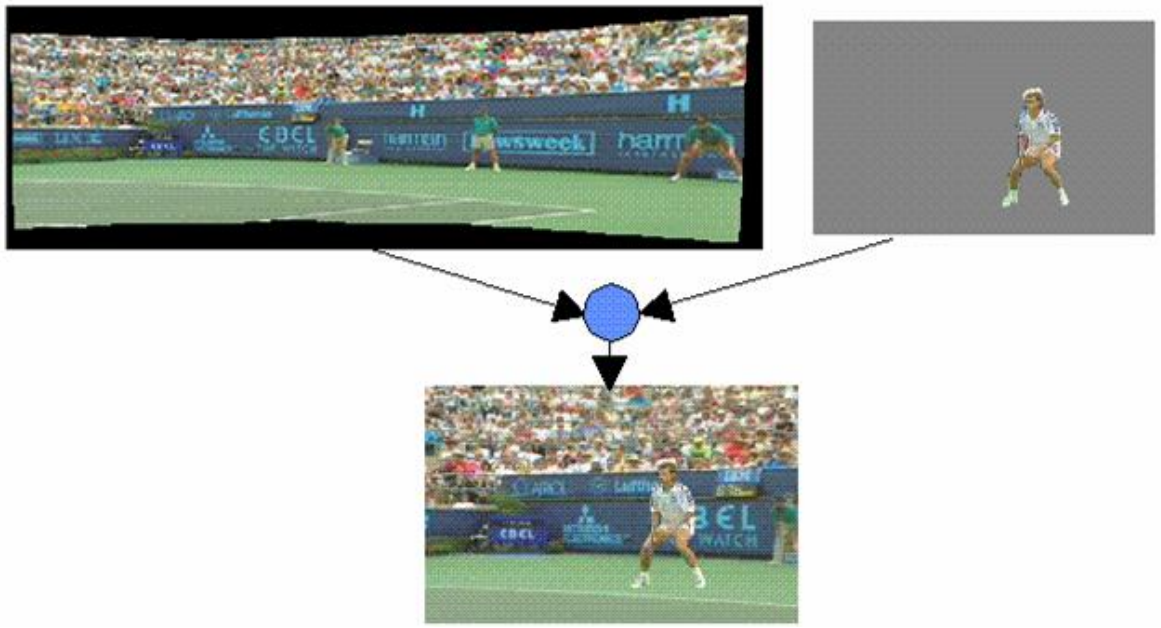
Media Objects

- An object is called a *media object*
  - real and synthetic images; analog and synthetic audio; animated faces; interaction
- Media objects have
  - Spatial relationships
  - Temporal relationships
- Compose media objects into a hierarchical representation
  - form compound, dynamic scenes

MPEG-4 Example

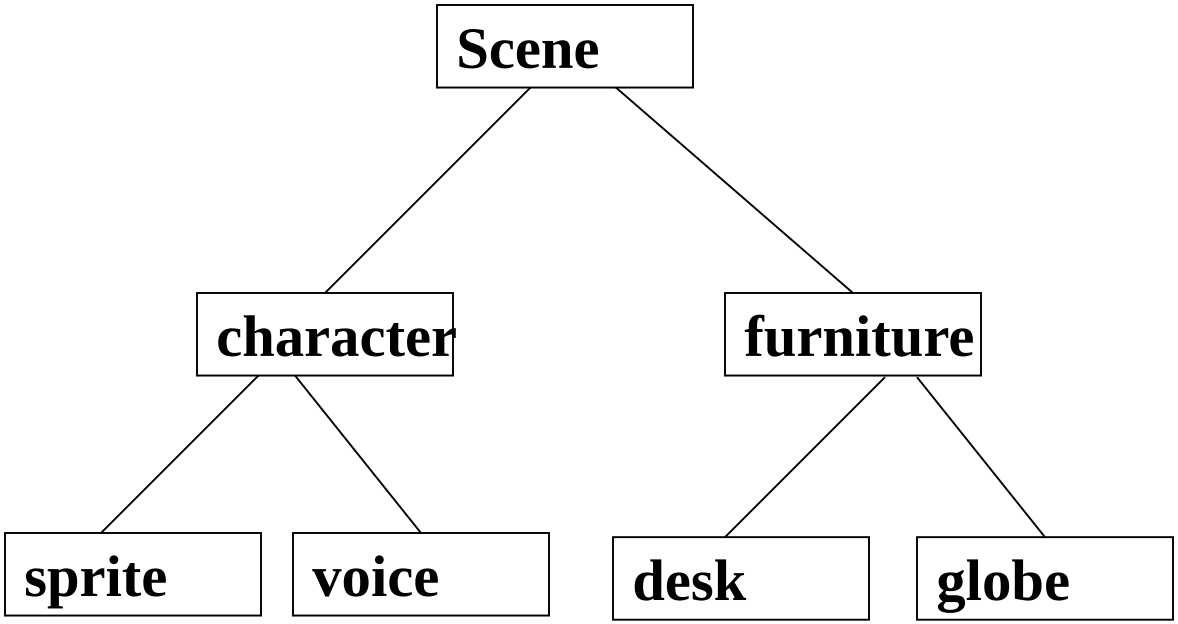
Media Object - Background

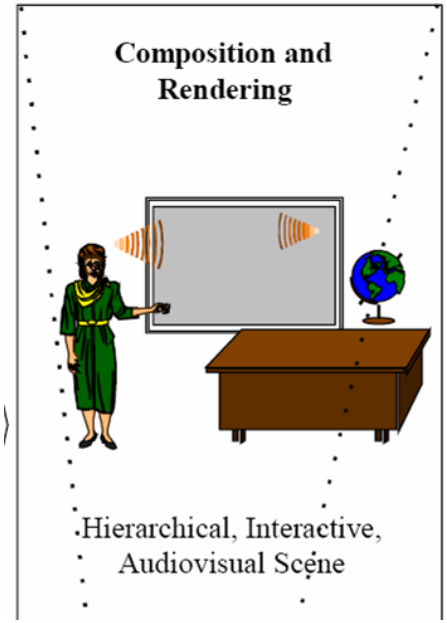
Media Object -Player



ISO N3536 MPEG4

Spatial Relationship - Composition (Scene Graph)



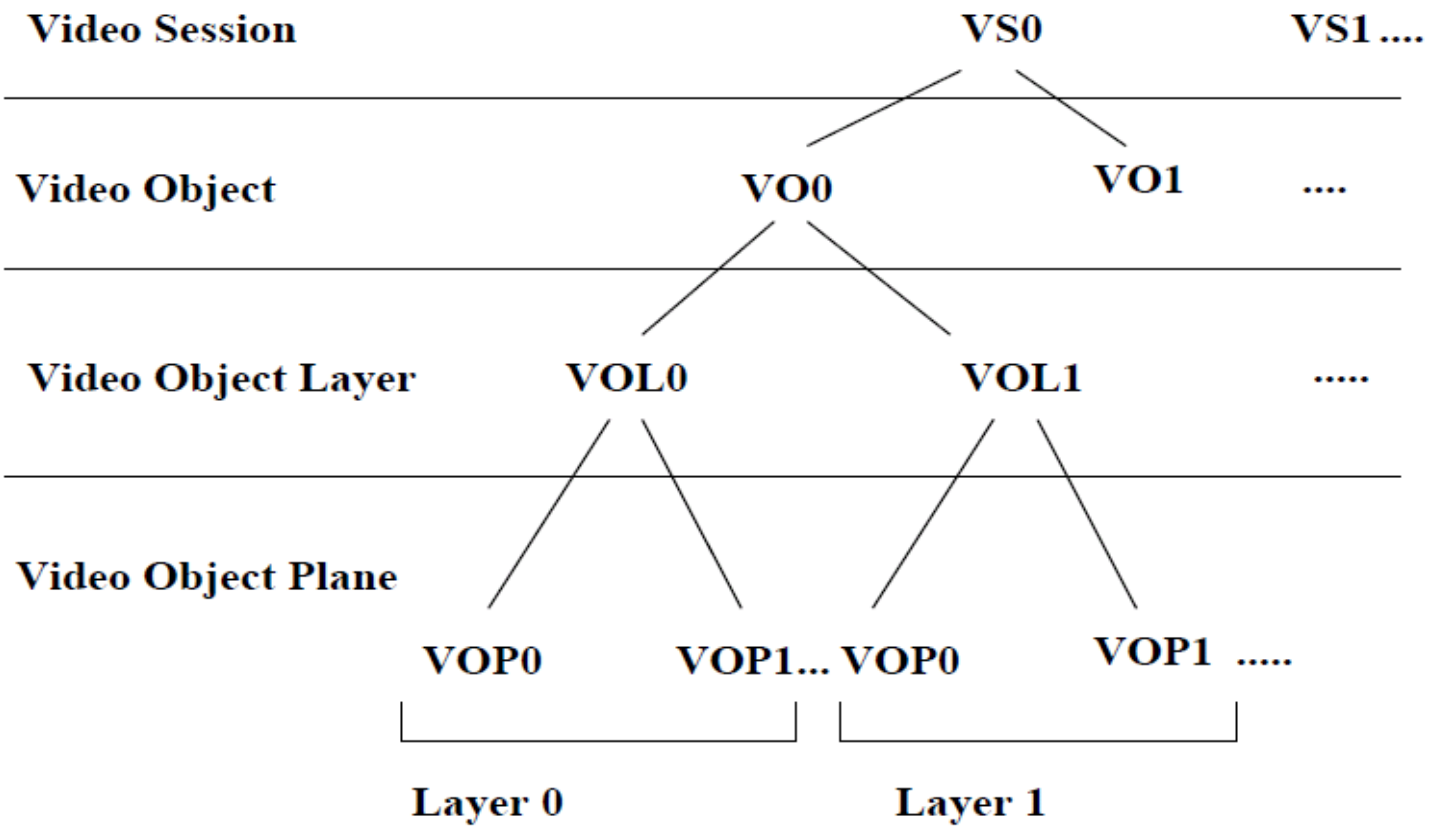


ISO N3536 MPEG4

Temporal Relationships

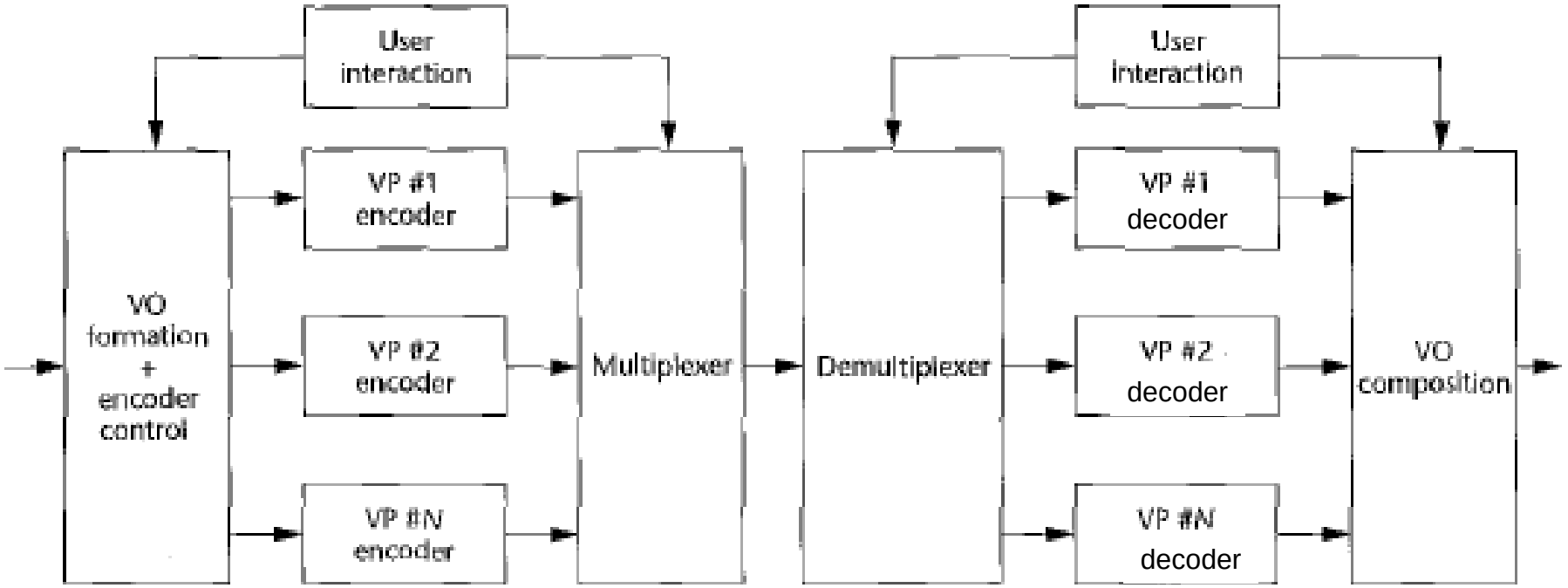
- Composition stream (BIFS) has its own associated time base
  - **Composition timestamps** specify at what time access units for composition must be ready at input of composition information decoder
- Timestamps are attached to each elementary stream
  - **Decoding timestamp (DTS)** specifies at which time the access unit for media object should be ready at decoder input
  - **Composition timestamp (CTS)** specifies time when object should be ready at the composition unit (compositor input).

Video Syntax Structure

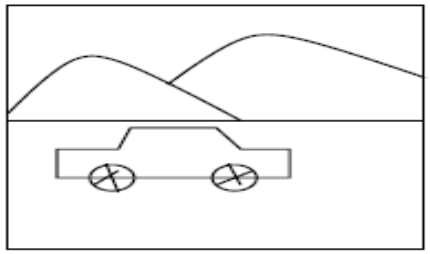
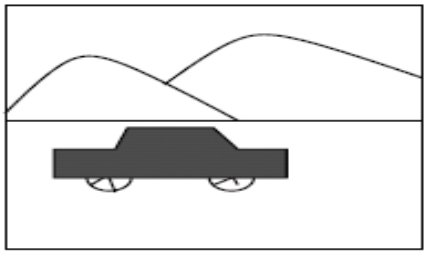
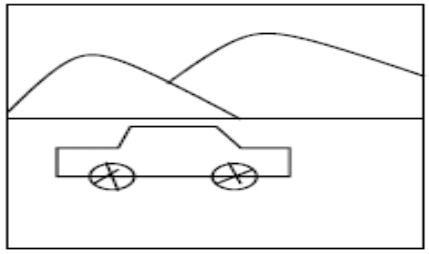
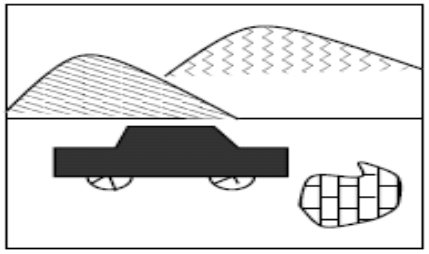
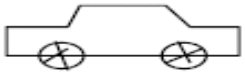


New MPEG-4 Aspect: Object-based layered syntactic structure

MPEG-4 Coding Architecture



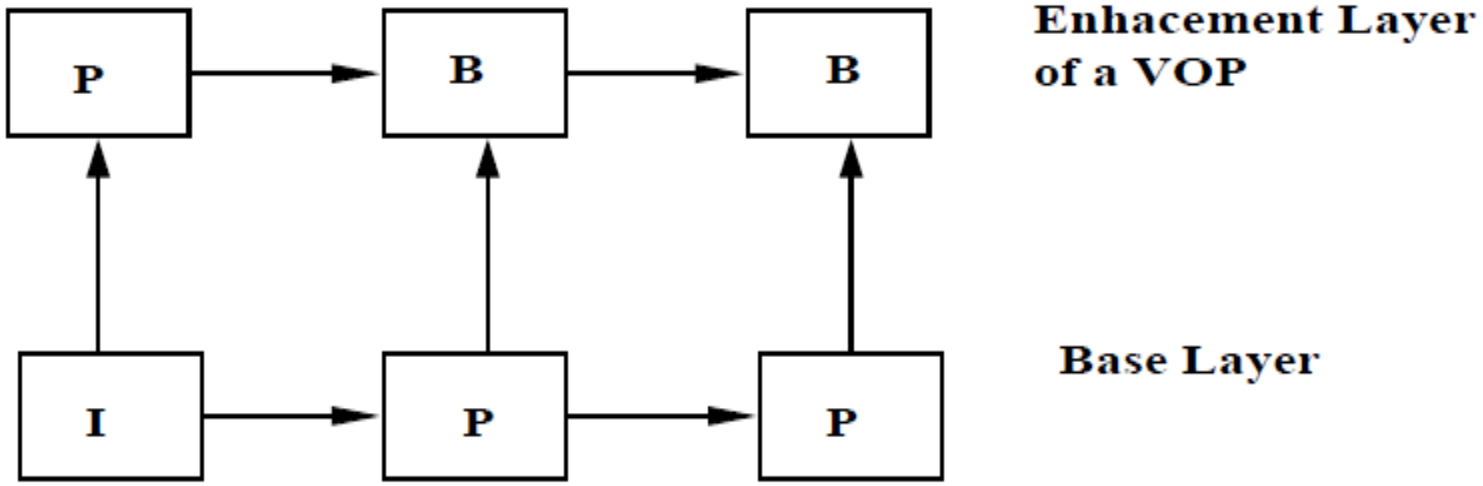
Examples of Base and Enhancement Layers

	Base Layer	Enhacement
Enhacement Type = 1	 VOL0: Entire Frame	 VOL1: Car
Enhancement Type = 2	 VOL0: Entire Frame	 VOL1: Entire Frame
	 VOL0: car	 VOL1: car

Coding of Objects

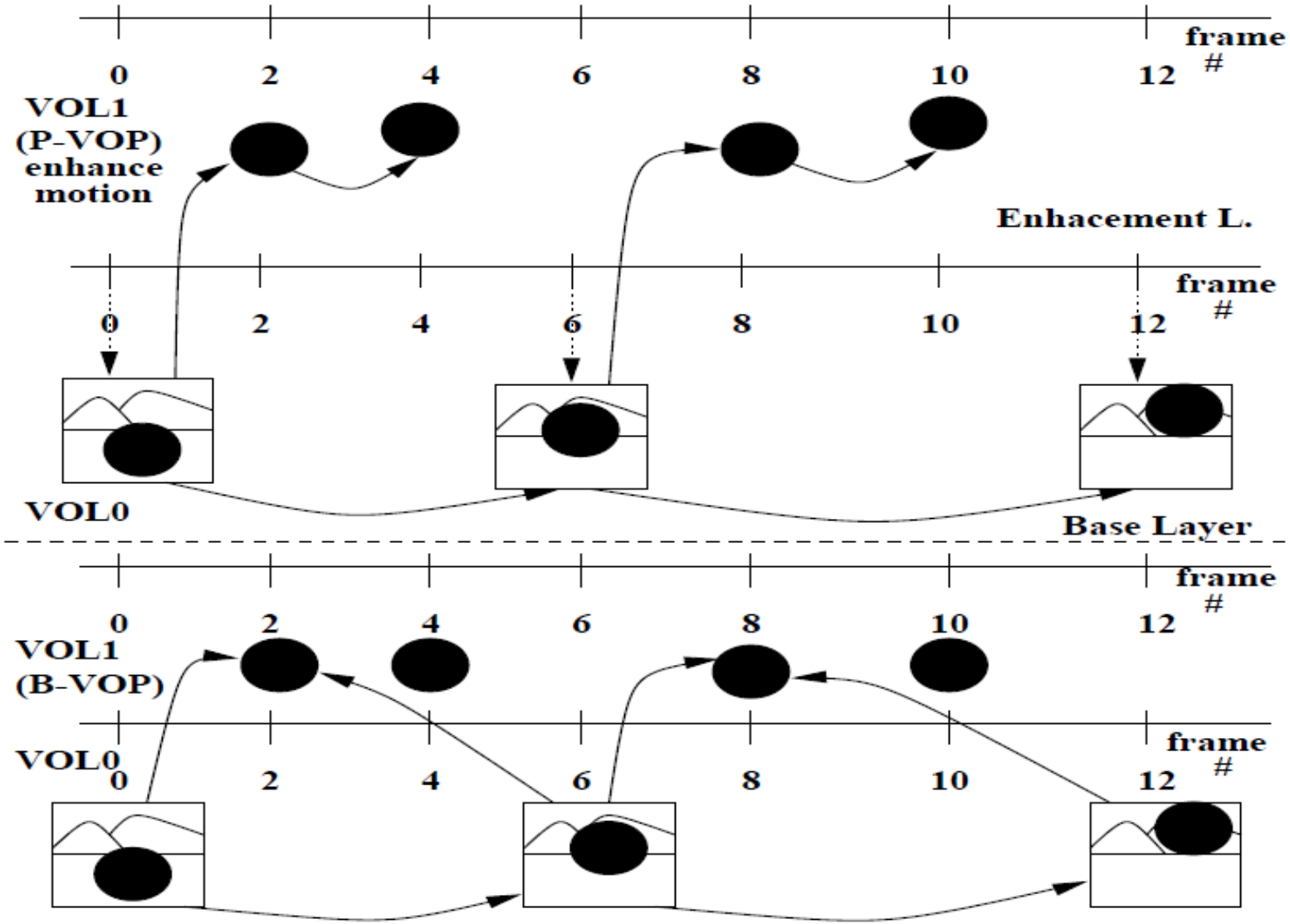
- Each VOP corresponds to an entity that after being coded is added to the bit stream
- Encoder sends together with VOP
  - Composition information where and when each VOP is to be displayed
- Users are allowed to change the composition of the entire scene displayed by interacting with the composition information

Spatial Scalability



VOP which is temporally coincident with **I-VOP** in the base layer, is encoded as **P-VOP in the enhancement layer**.
VOP which is temporally coincident with **P-VOP** in the base layer is encoded as **B-VOP in the enhancement layer.**

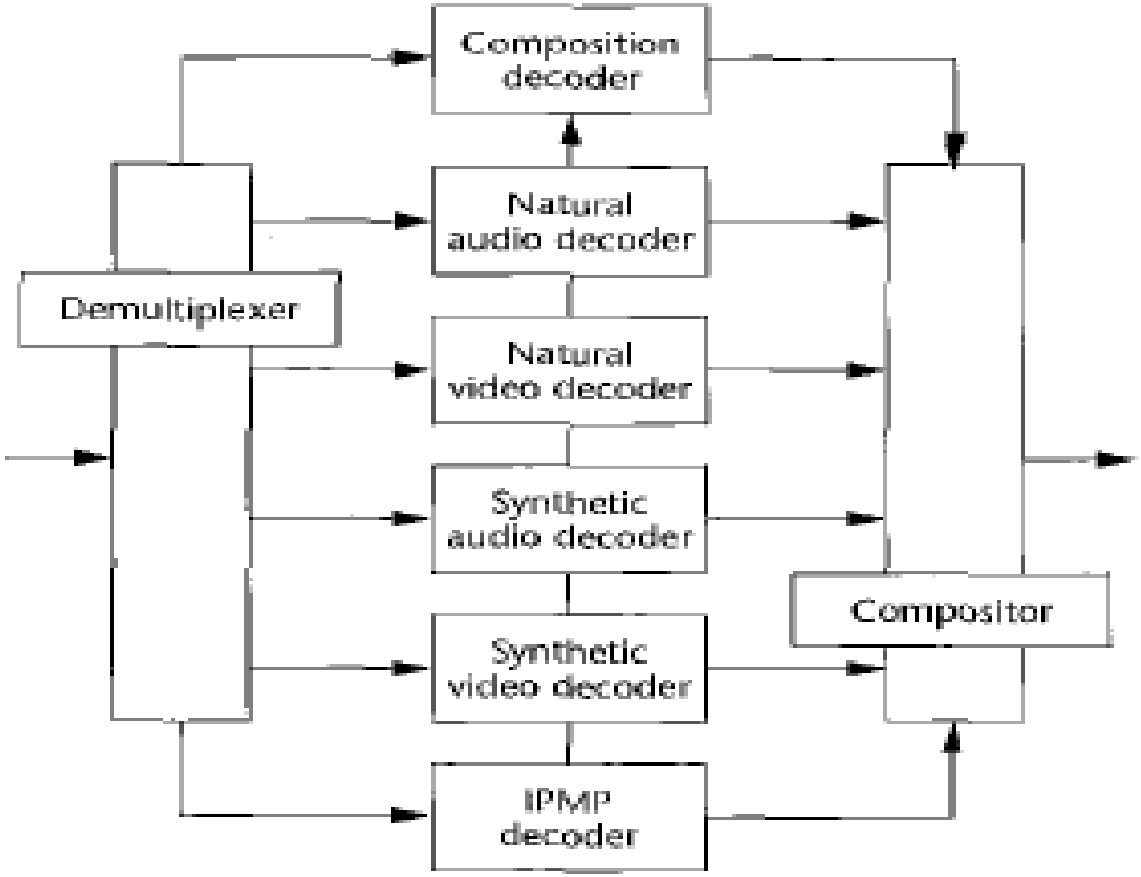
Temporal Scalability



Composition (cont.)

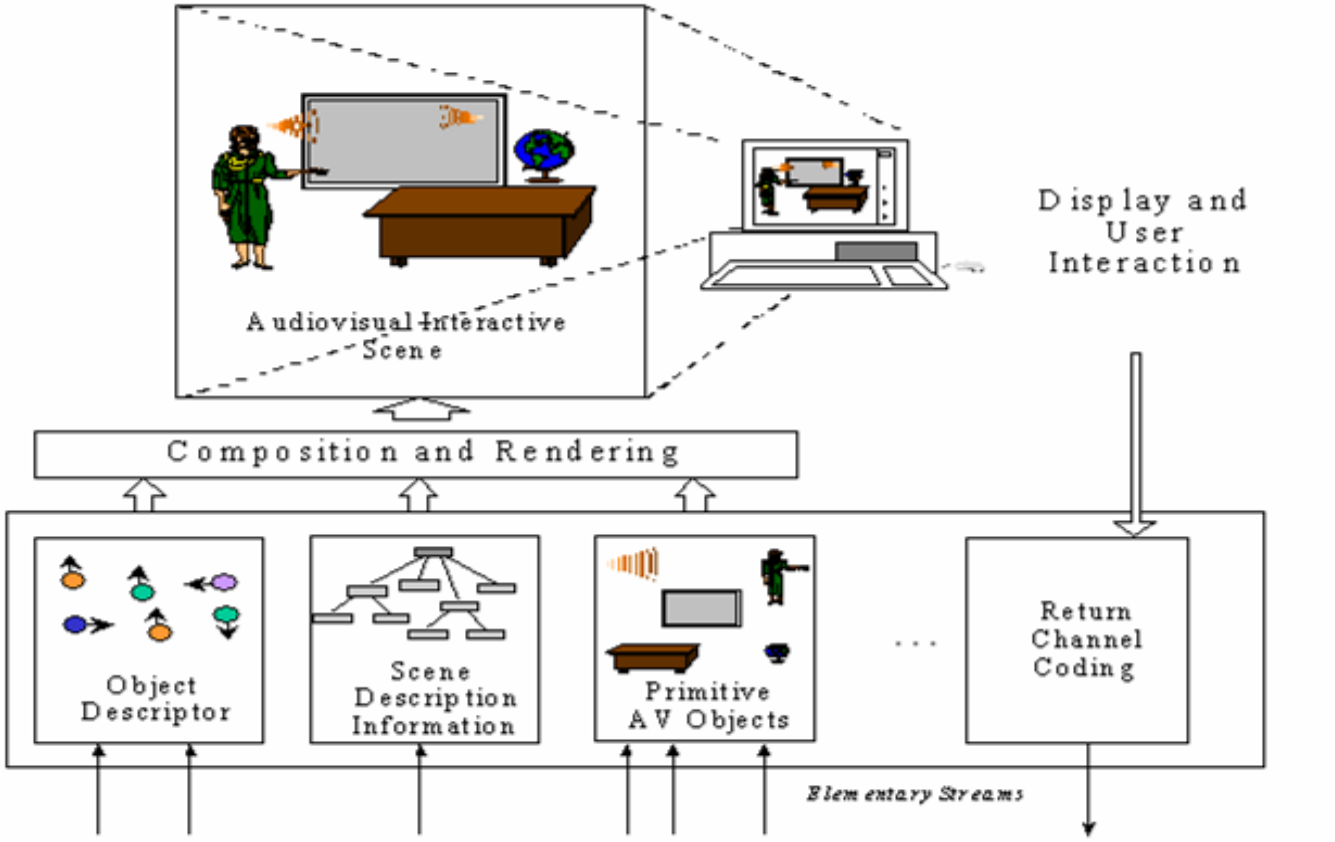
- Encode objects in separate channels
  - encode using most efficient mechanism
  - transmit each object in a separate *stream*
- Composition takes place at the decoder, rather than at the encoder
  - requires a binary scene description (BIFS)
- BIFS is low-level language for describing:
  - hierarchical, spatial, and temporal relations

MPEG-4 Decoder Architecture



Source: Batista et al, “MPEG-4:A mulitmedia standard for the thrid millenium: Part 1”, IEEE Multimedia, 1999

MPEG-4 Rendering



ISO N3536 MPEG4

Synchronization Layer (timing and synchronization)

- Synchronization of elementary streams
  - Elementary streams consist of access units which correspond to portions of stream with DTS and CTS
- Global timeline (high-resolution units)
  - e.g., 600 units/sec
- Each continuous track specifies relation
  - e.g., if a video is 30 fps, then a frame should be displayed every 33 ms.
- Others specify start/end time

MPEG-7

MPEG-7 Motivation

- Need for searching for audio-visual information
- Goal:
  - input a picture of motorbike from “Terminator II” and get search results
  - Efficiently search for video sequence where King Lear congratulates his assistants
  - Input a whistle of a melody and find a song

MPEG-7 Description

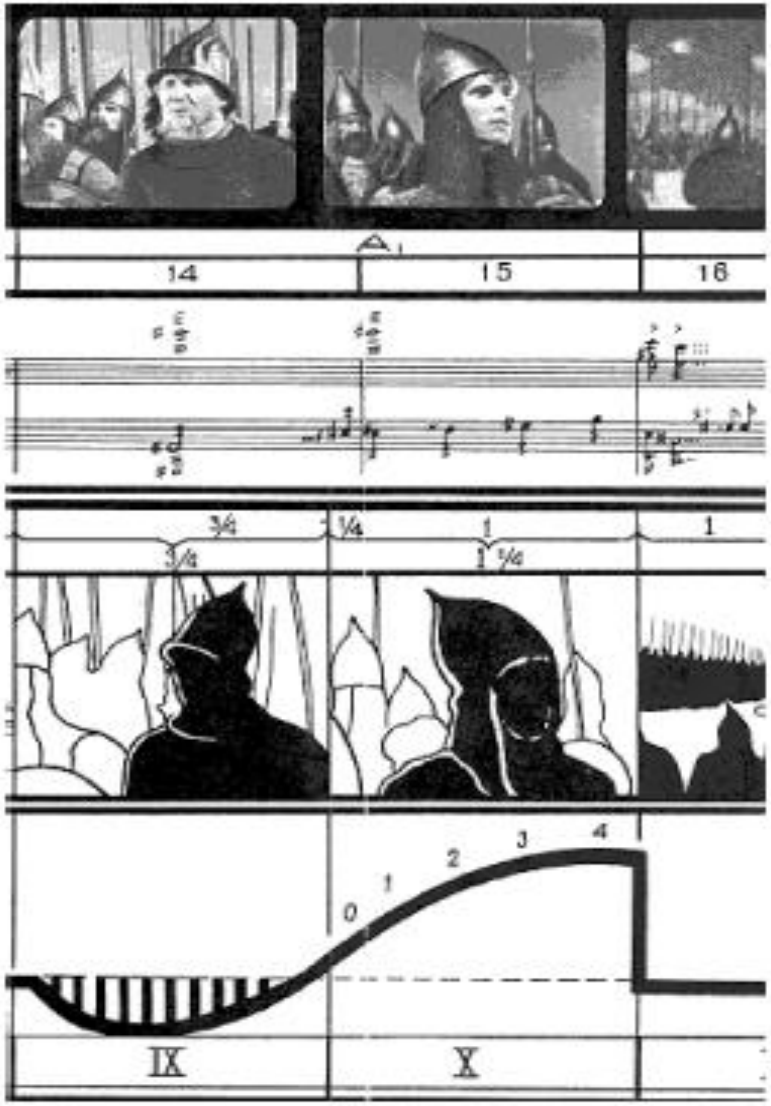
- Standardization of
  - core set of quantitative measures of audio-visual features, called Descriptors (D)
  - structures of descriptors and their relationships, called Description Schemes (DS)
  - A language, called Description Definition Language (DDL)
- Result: one can index and search for audio-visual material that has MPEG-7 metadata associated with it

MPEG-7 Features

- **Multimedia content description**
  - Set of methods and tools for different classes of multimedia content description
  - Types of descriptions:
    - Medium-based: Sampling rate, location of shot cuts, camera's focal length, etc
    - Physical: physical loudness of sound, ..
    - Perceptual: perceptual 'pitch'
    - Transcription: dialog transcriptions
    - Architectural: domain of document structuralists
    - Annotative: domain of human annotations

MPEG-7 Features

- **Flexibility in data management**
  - Referencing - to parts of document, who document, series of documents
  - Multimodality – description of multimedia content in such a way as to allow queries based on visual descriptions to retrieve audio data and visa versa
  - Applications – description of stored, streamed, real-time, non-RT apps



 Source: Nack et all, “MPEG-7”, IEEE Multimedia 1999

MPEG-7 Features

- **Globalization of data resources**
  - MPEG-7 resources can be
    - Co-located with audio-visual material or
    - Remote - live somewhere else
  - Need mechanisms to link audio-video material and MPEG-5 descriptions

MPEG-7 Features

- MPEG-7 does not extract descriptions/features automatically
- MPEG-7 concentrates on representations that can be used for descriptions
- MPEG-7 uses describing text documents such as SGML, XML (Extensible Markup Language), RDF (Resource Description Language)
- MPEG-4 and MPEG-7 are a pair of standards complementing each other

Summary

- MPEG is a very prevalent video coding format, especially in video-on-demand
- MPEG-1, MPEG-2 and MPEG-4 are content distribution systems
- MPEG-7 (and MPEG-21) are description-specific systems to assist in representing metadata of e.g., MPEG-4 coded data and querying for MPEG coded data