CS 598KN

# Advanced Multimedia Systems Design

# Lecture 3 – Image and Video (MPEG) Coding

Klara Nahrstedt

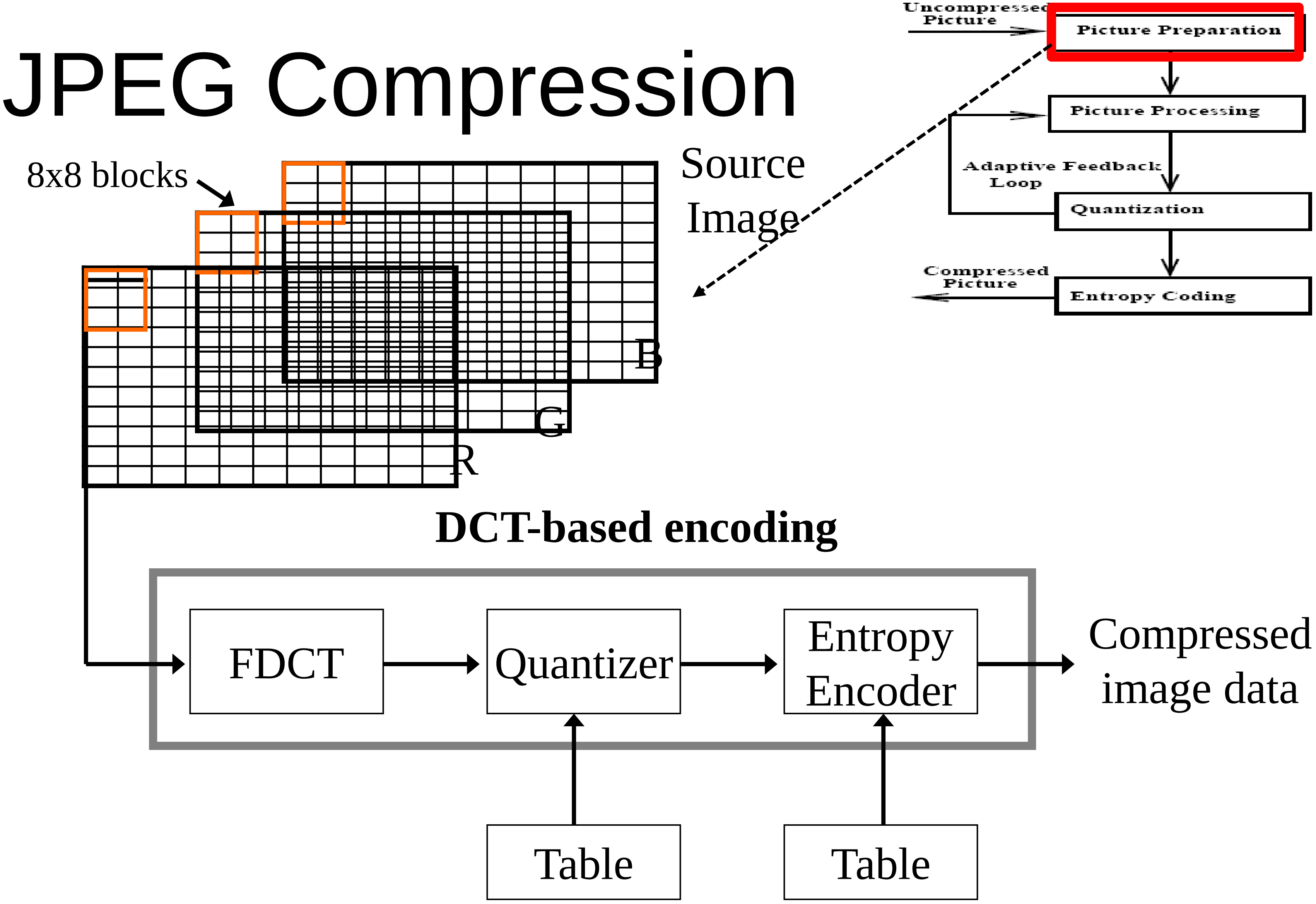
Fall 2019

# Overview

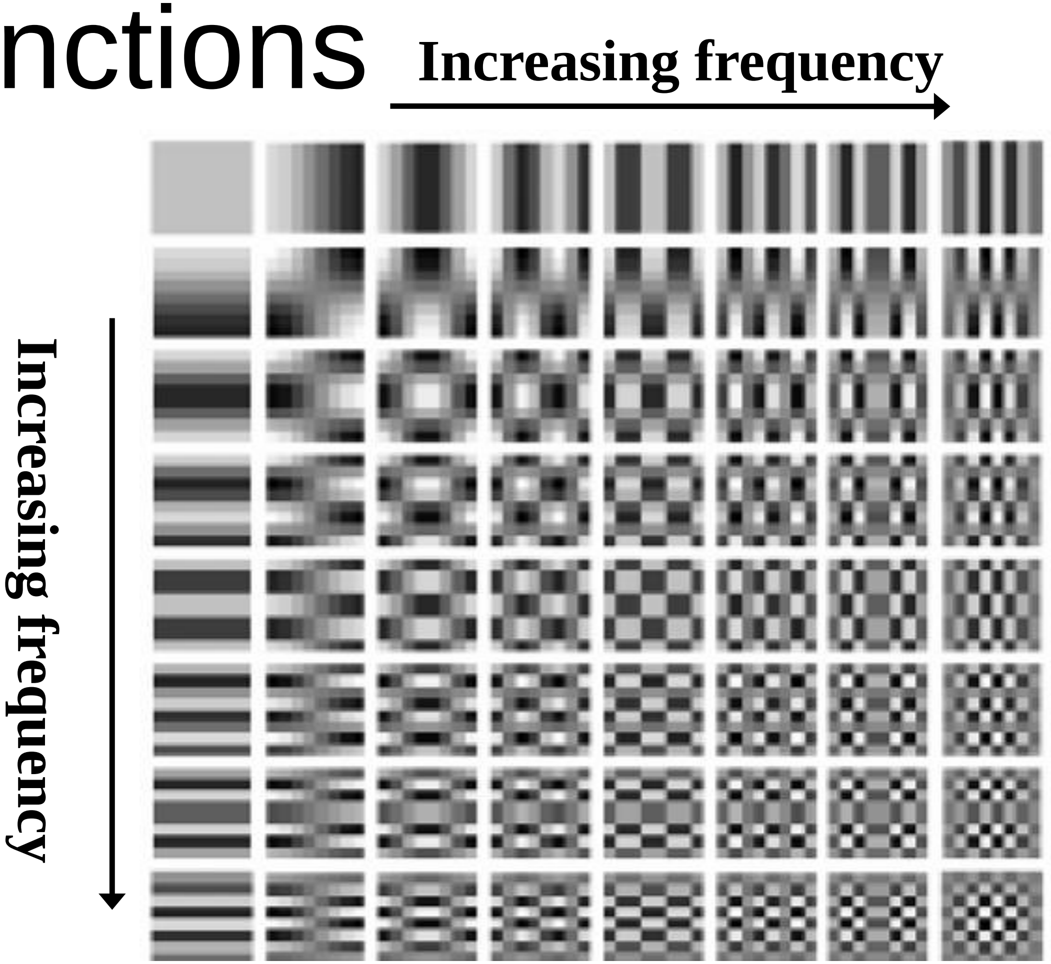
- JPEG and JPEG 2000 Compression
- MPEG Basics
- MPEG-4 for Broadcast Streaming

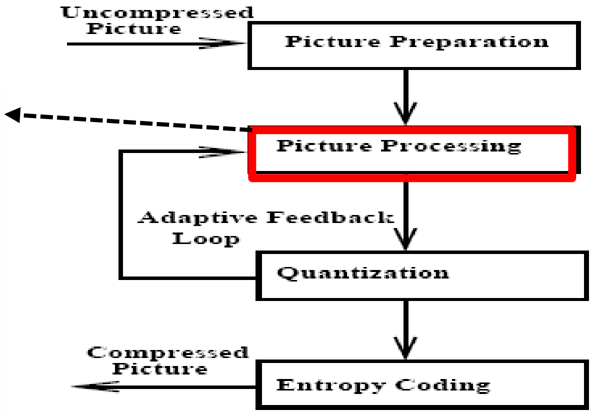
# JPEG COMPRESSION (BASIC CONCEPTS)

# JPEG Compression



# Visualization of Basic FDCT Functions





Uncompressed Picture

Picture Preparation

Picture Processing

Adaptive Feedback Loop

Quantization

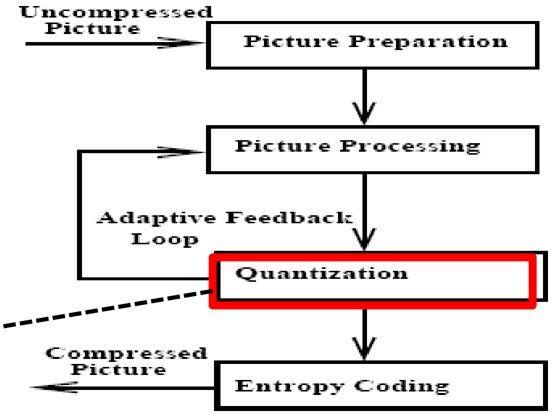
Compressed Picture

Entropy Coding

# 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

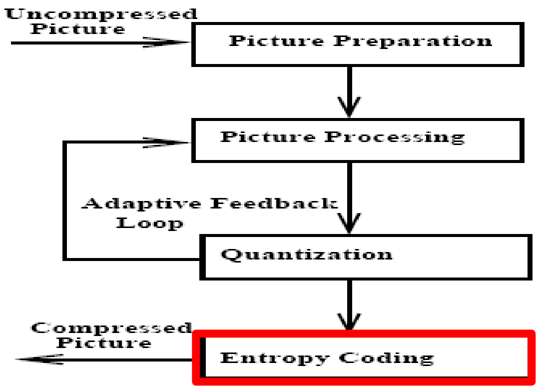
# De facto Quantization Table

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

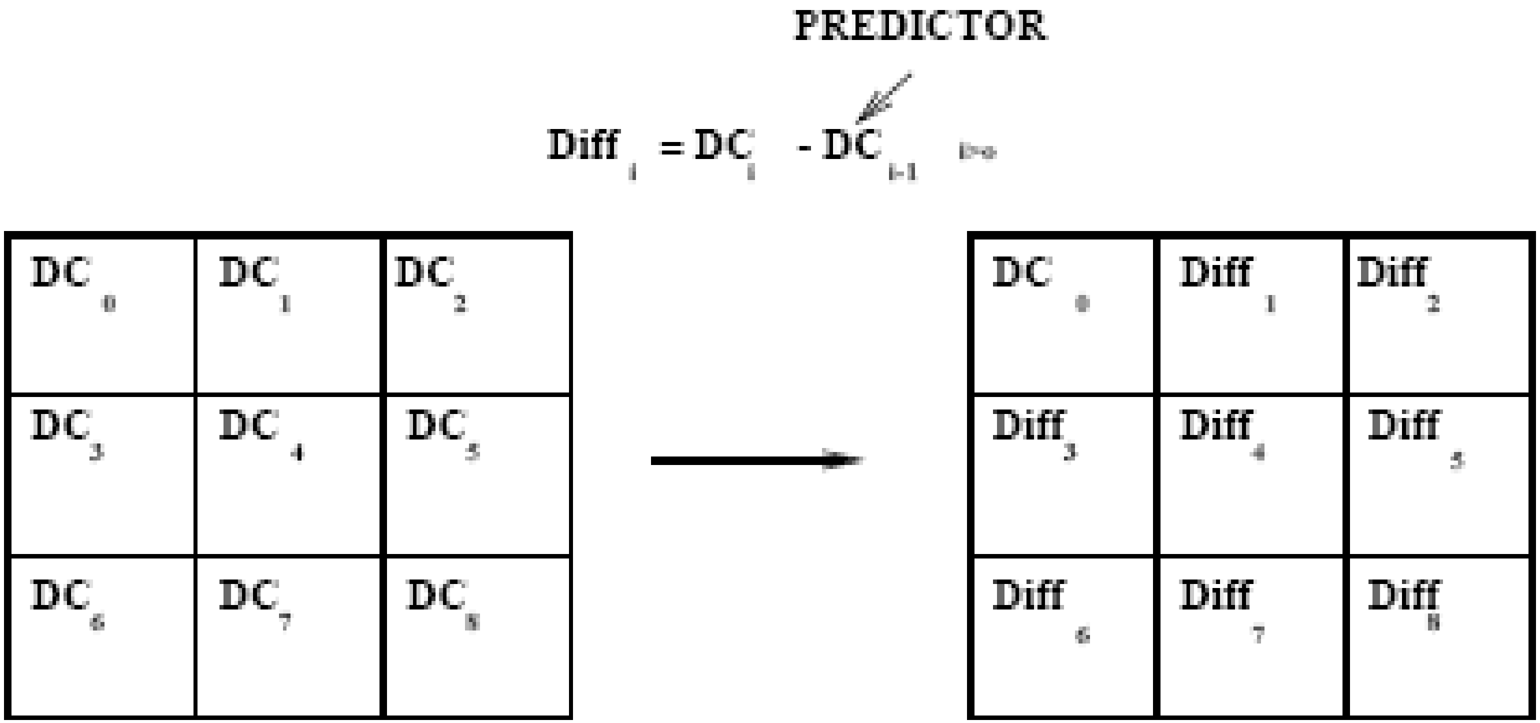
- Separate DC from AC components
  - DC components change slowly, thus will be encoded using difference encoding

# 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, we 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



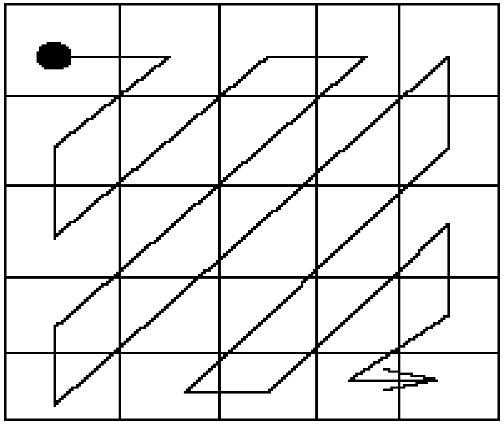
$$Diff_i = DC_i - DC_{i-1} \quad i>0$$

|  |  |  |
|---|---|---|
| $DC_0$ | $DC_1$ | $DC_2$ |
| $DC_3$ | $DC_4$ | $DC_5$ |
| $DC_6$ | $DC_7$ | $DC_8$ |

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

# AC Encoding

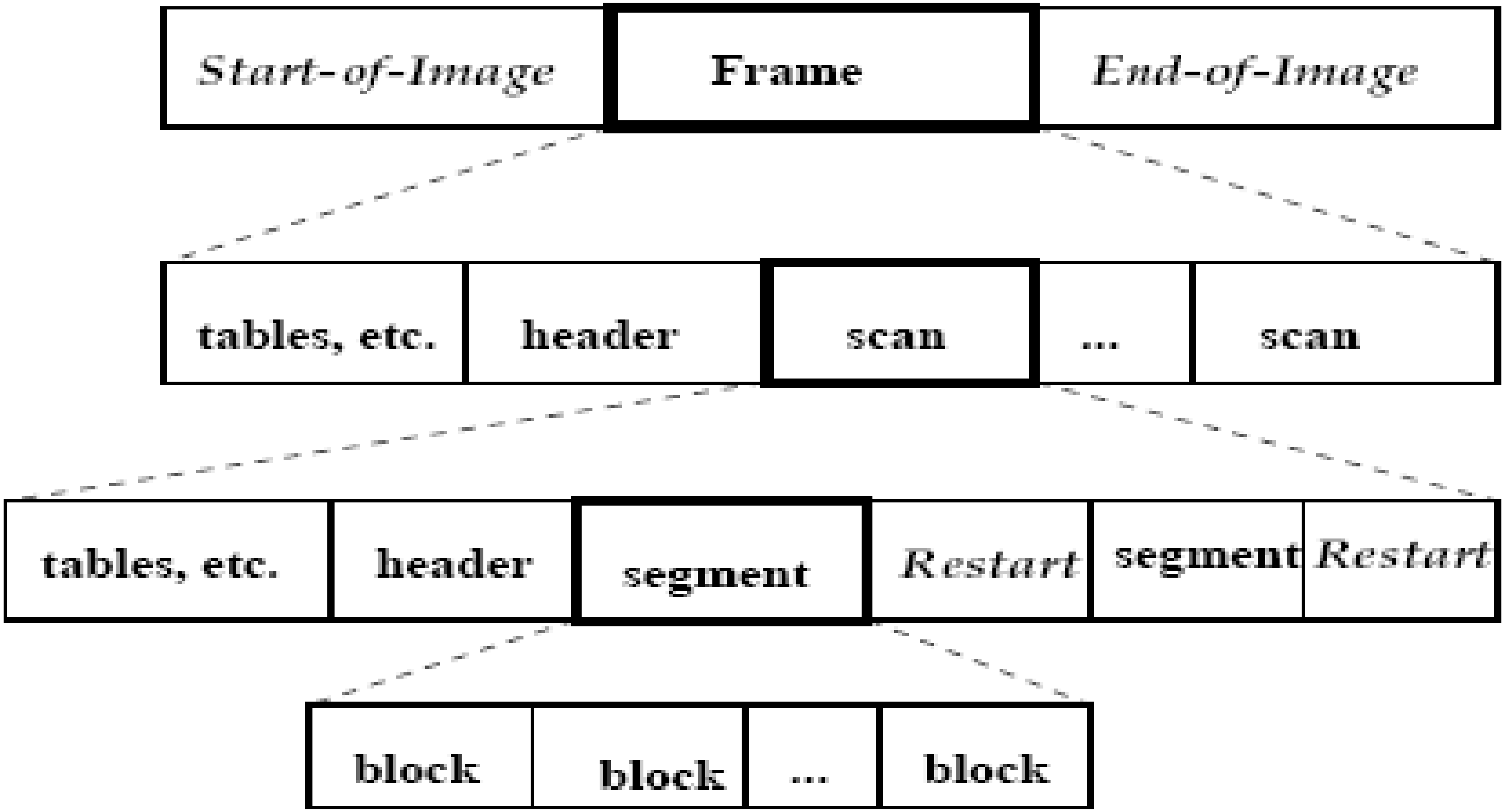
- Use zig-zag ordering of coefficients
  - orders frequency components from low->high
  - produce maximal series of 0s at the end
  - Ordering helps to apply efficiently entropy encoding
- Apply Huffman coding
- Apply RLE on AC zero values



# 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!

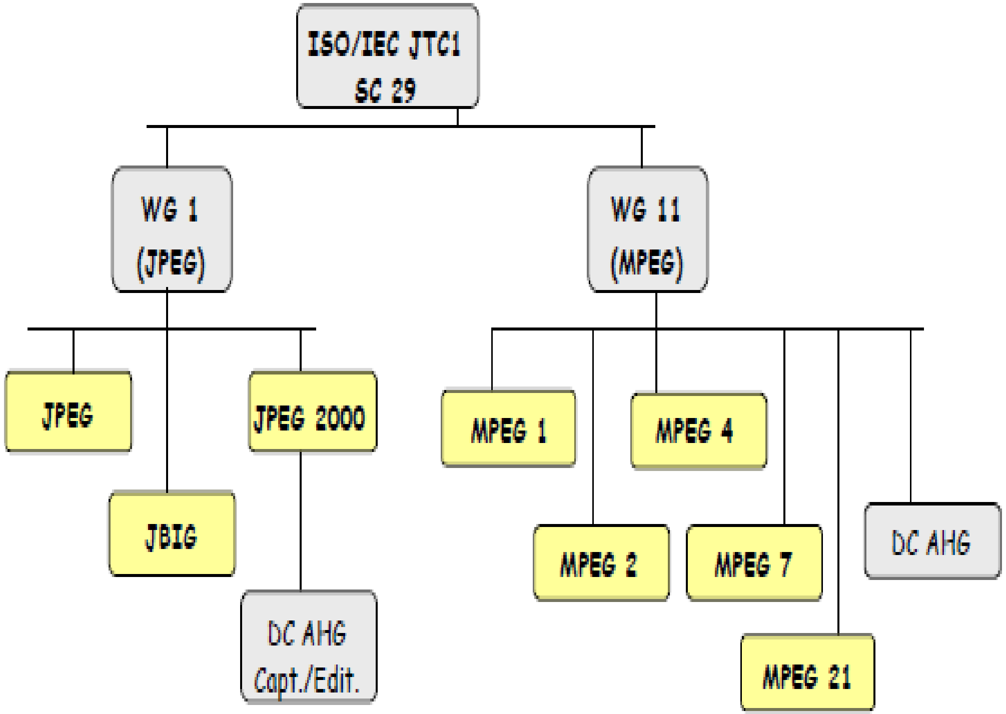
# Interchange Format of JPEG



# JPEG 2000

# From JPEG to JPEG 2000

- **Image Preparation**
  - Components Separation
  - Block Division (8x8 Blocks) of each Component
- **Image Processing**
  - Pixel Value Shifting
  - 2D DCT Transformation
  - Creation of DC and AC Coefficients
- **Quantization**
  - Quantization Tables
- **Entropy Coding**
  - Zig-Zag Ordering
  - DC Coefficients – Differential Coding
  - AC Coding - RLE & Huffman Coding



# JPEG-2000

Original (uncompressed TIF 116KB)

JPEG-2000 (8:1, 14KB)

JPEG (8:1, 14KB)

Created in 2000 by JPEG committee

http://www.photographical.net/jpeg2000.html

# JPEG-2000 Features

- Low bit rate compression performance
    - Current standards offer excellent rate-distortion performance in mid and high bit rates
    - Low bit rate distortions become unacceptable
- Lossless and lossy compression
    - Current standard does not provide superior lossless and lossy compression in a single code-stream

# JPEG-2000 Features

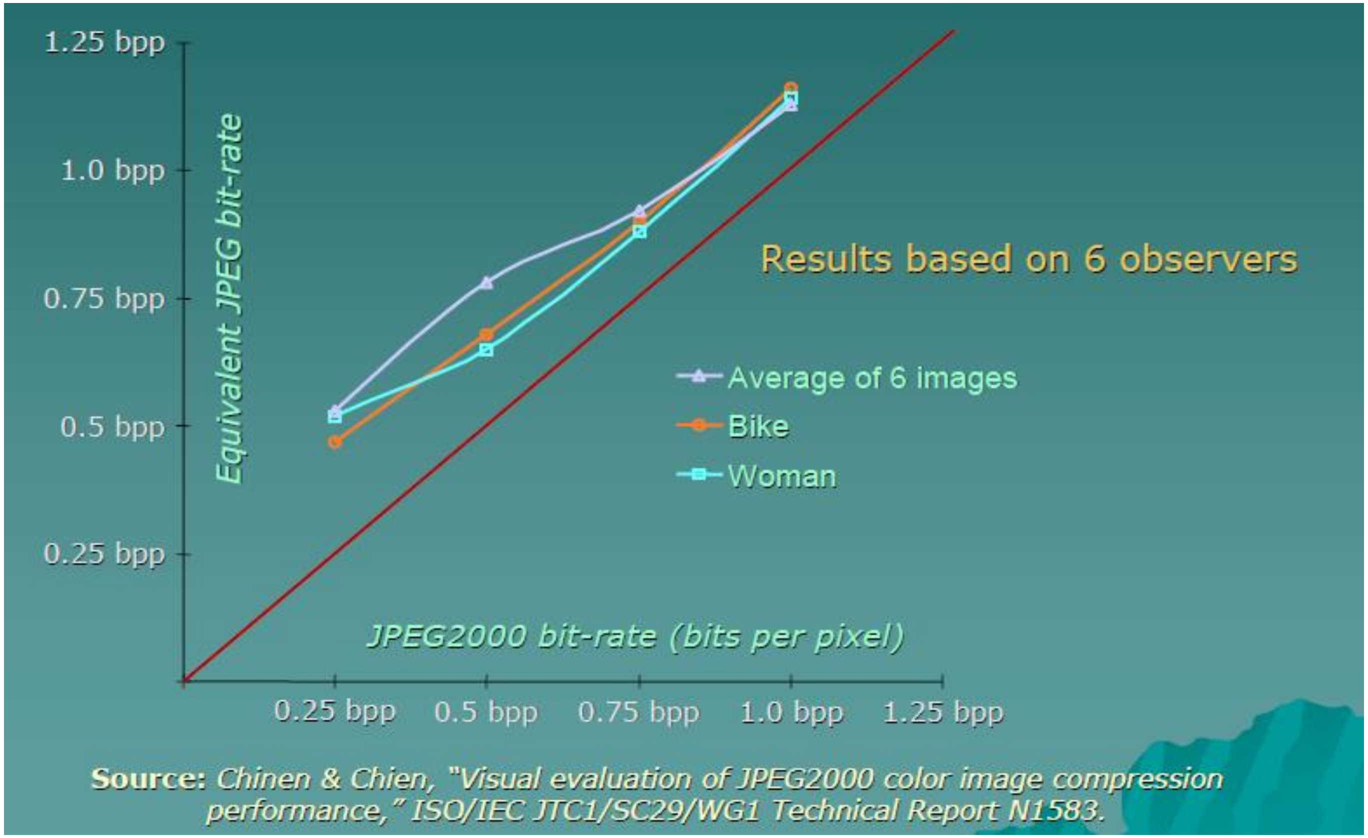
- Large Images
    - Current standard does not allow for images larger than 64Kx64K pixels without quality degradation
- Single decompression architecture
    - Current standard has 44 modes (application specific, and not used by majority JPEG coders)
    - Single common decompression architecture can provide greater interchange between applications

# JPEG-2000 Features

- Transmission in noisy environment
    - Current standard has provision for restart intervals, but image degrades badly when bit errors occur.
- Computer generated imagery (Graphics)
    - Current standard is optimized only for natural imagery
- Compound documents
    - Current standard is not applied to compound documents because of its poor performance when applied to text imagery

# JPEG-2000 Features

- Superior low bit rate performance
  - Below 0.25 bits per pixel for highly detailed grey-scale images



**Source:** *Chinen & Chien, "Visual evaluation of JPEG2000 color image compression performance," ISO/IEC JTC1/SC29/WG1 Technical Report N1583.*

# JPEG-2000 Features

- Lossless and lossy compression
  - Lossless compression uses progressive decoding (i.e., difference image encoding) for medical imaging
- Progressive transmission by pixel accuracy and resolution
  - Reconstruction of images is possible with different resolutions and pixel accuracy for different target devices

# JPEG-2000 Features

- Random code-stream access and processing
  - Needed in case images have parts that are more important than others
  - User defines "regions-of-interest" in the image to be randomly accessed and/or decompressed with less distortion than the rest of images
  - random code-stream processing allows operations: rotation, translation, filtering, feature extraction, scaling,…

# Methods of Compression

**DCT-based Artifact**

- DCT-based coder
  - New baseline JPEG algorithm required for backward compatibility with existing JPEG
- Wavelet-based coder
  - This method permits coding of still images with high coding efficiency as well as spatial and SNR (signal-to-noise ratio) scalability at fine granularity (see also tutorial – part1/part2/part3 - https://cseweb.ucsd.edu/~baden/Doc/wavelets/polikar_wavelets.pdf)

# Wavelet Transform

- DWT (Discrete Wavelet Transform) extracts information from the source image at different scales, locations and orientations
- JPEG-2000 uses two techniques in wavelet-based coder
  - 2D wavelets
  - multi-scale transforms
- Wavelet is defined as a set of basic functions, derived from the same prototype function
- Prototype function is known as “mother wavelet”
  - Examples: “Mexican Hat” wavelet, Haar wavelet

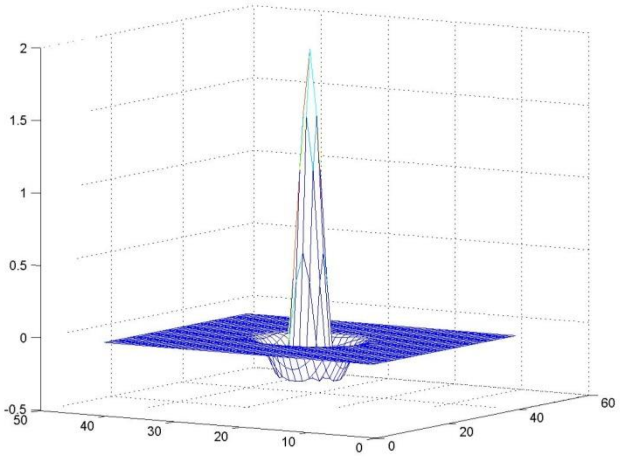
# 2D Mexican Hat ‘analyzing wavelet’

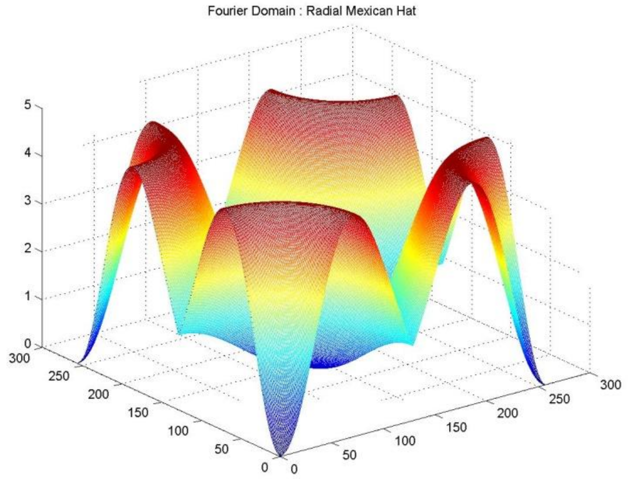
Time domain

Frequency domain

$$\Psi(x,y) = (x^2 + y^2 - 2)e^{-\frac{1}{2}(x^2+y^2)}$$

$$\Phi(w1,w2) = -2\pi(w1^2 + w2^2)e^{-\frac{1}{2}(w1^2+w2^2)}$$

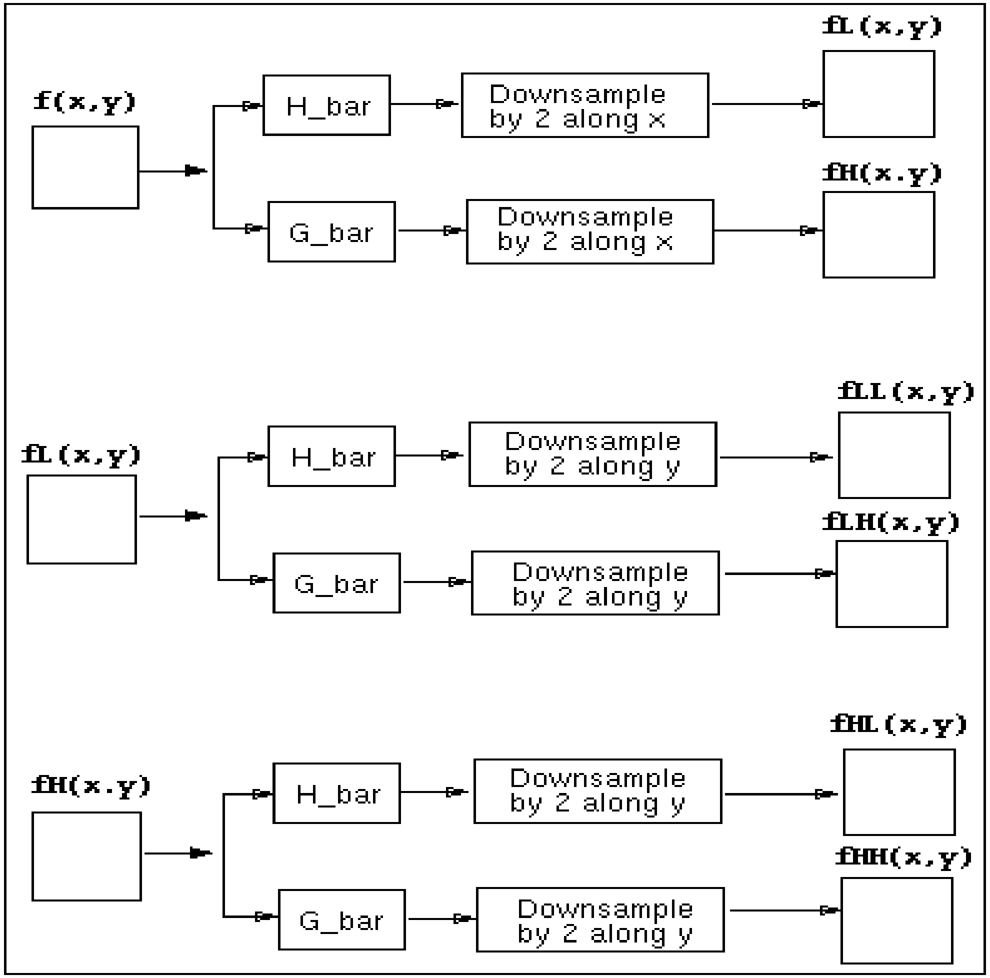




# Wavelet Transform Properties

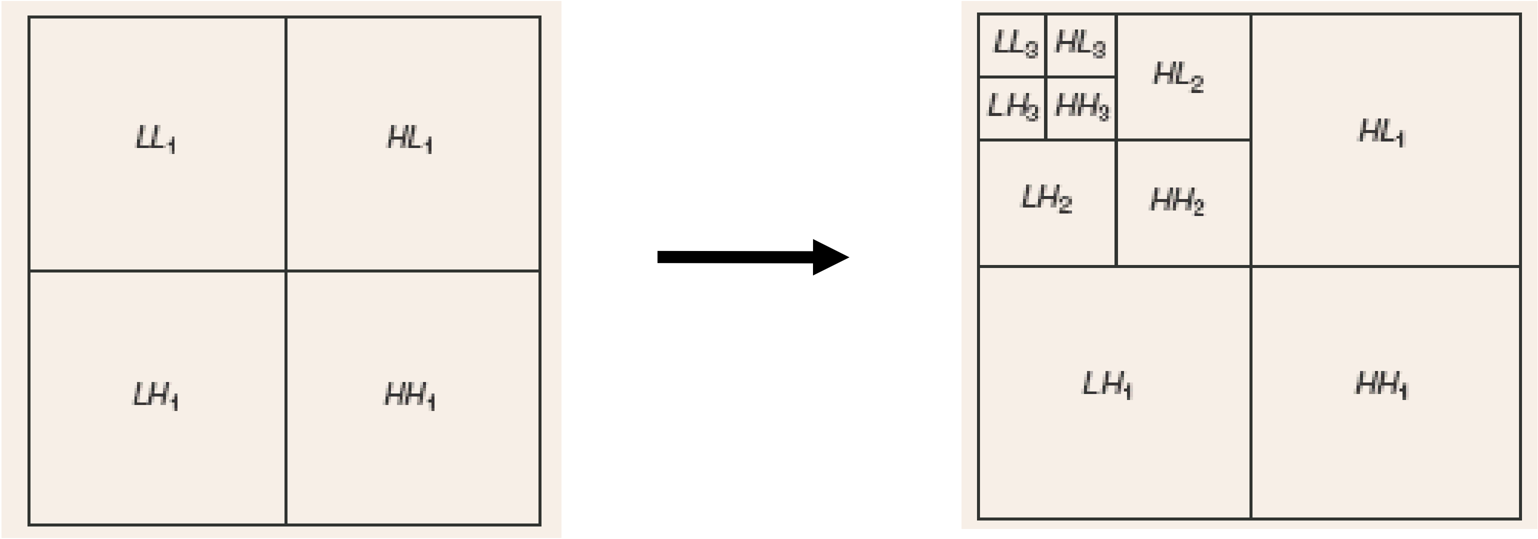
- Wavelet transform coders process high and low frequency parts of image independently
  - DCT methods have difficulties with high-frequency information
- Wavelet method transforms image as a whole (not subdivided into pixel blocks)
  - No blocking artifacts occur
  - Wavelet coders degrade gracefully

# Forward Wavelet Transform

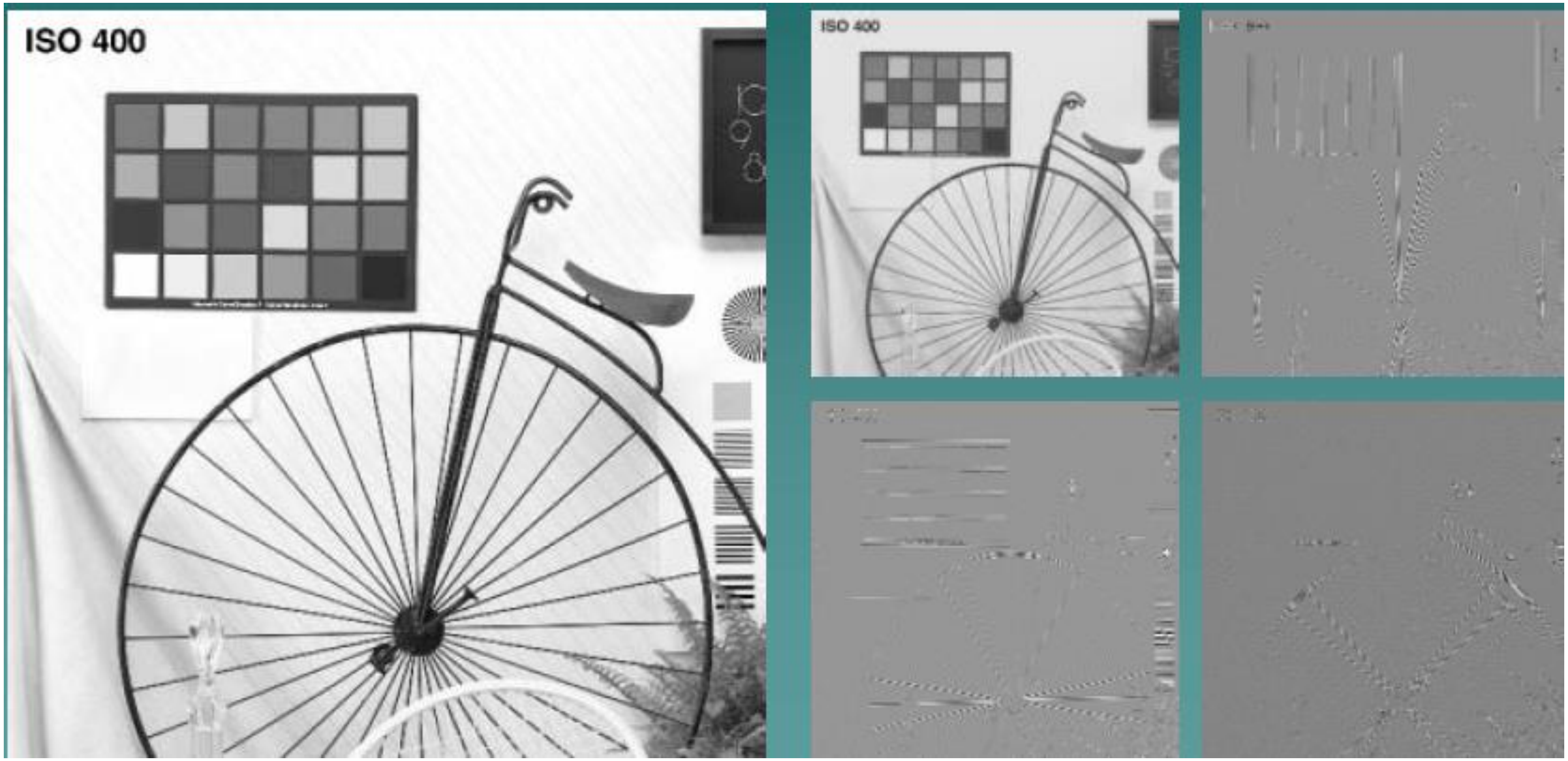


- Image is first filtered along the x dimension, resulting in low-pass and high-pass image
- Since bandwidth of both low pass and high pass image is now half that of the original image, both filtered images can be down-sampled by factor 2 without loss of information
- Then both filtered images are again filtered and down-sampled along the y dimension resulting in four sub-images

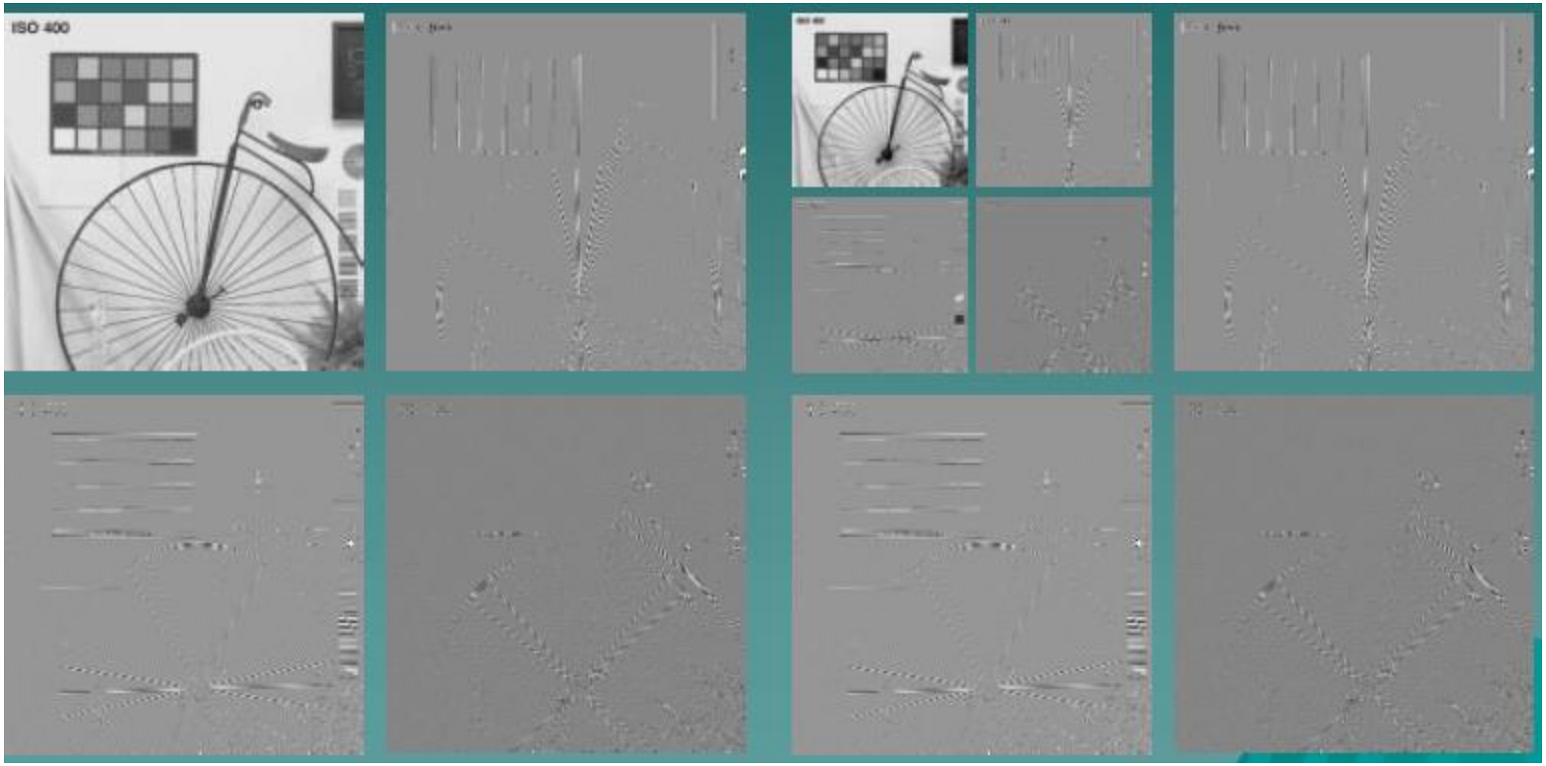
# Wavelet Transform



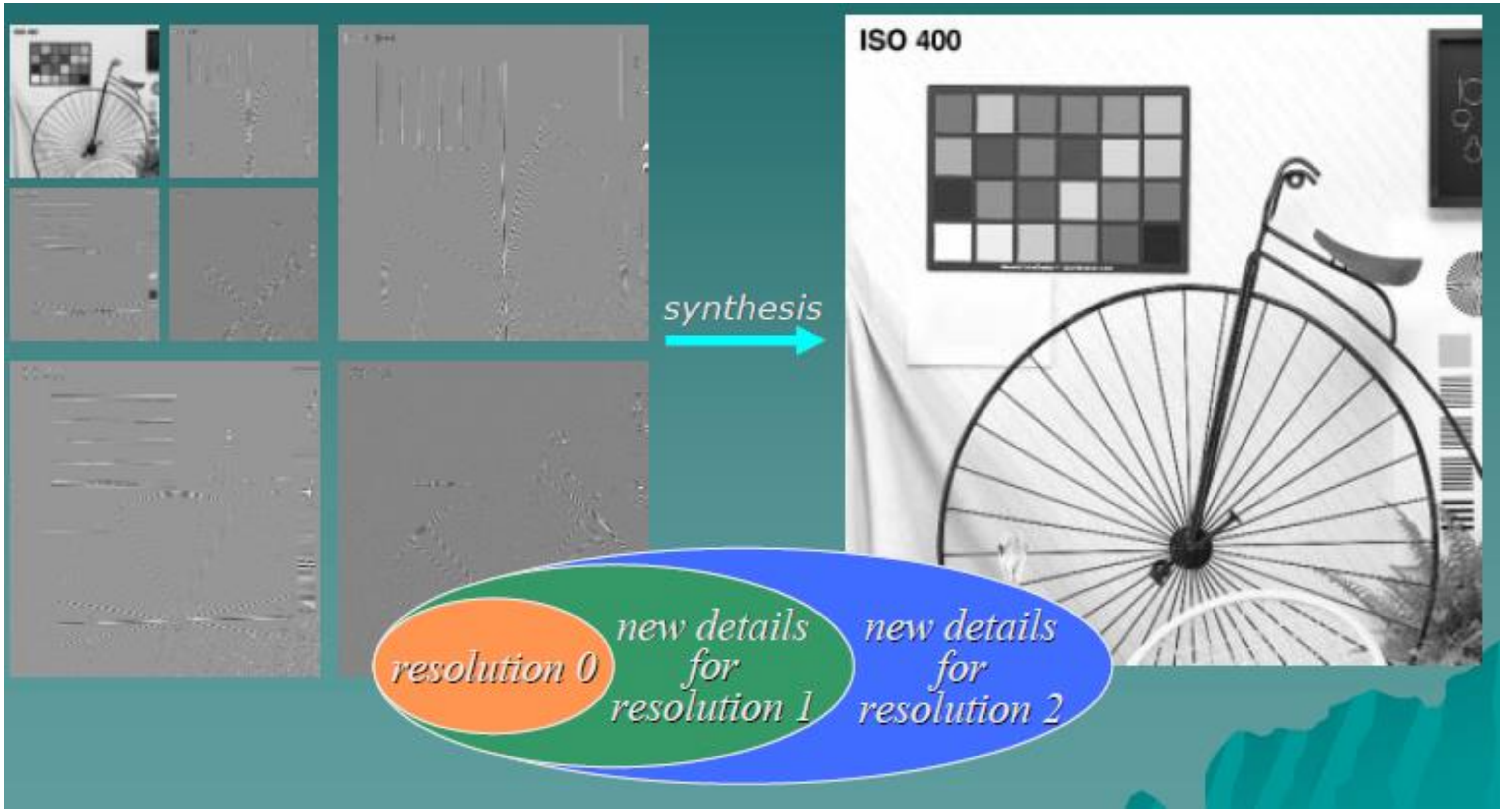
# Wavelet Transform (1)



# Wavelet Transform (2)



# JPEG-2000 Resolution Scalability



Source: http://www.ee.unsw.edu.au/~taubman/seminars_files/IEEE_IEA_J2K.pdf

# JPEG-2000 Scalability

- Scalable in both SNR and resolution

More bits

Example of resolution progressive bit-stream ordering

# JPEG-2000 Performance

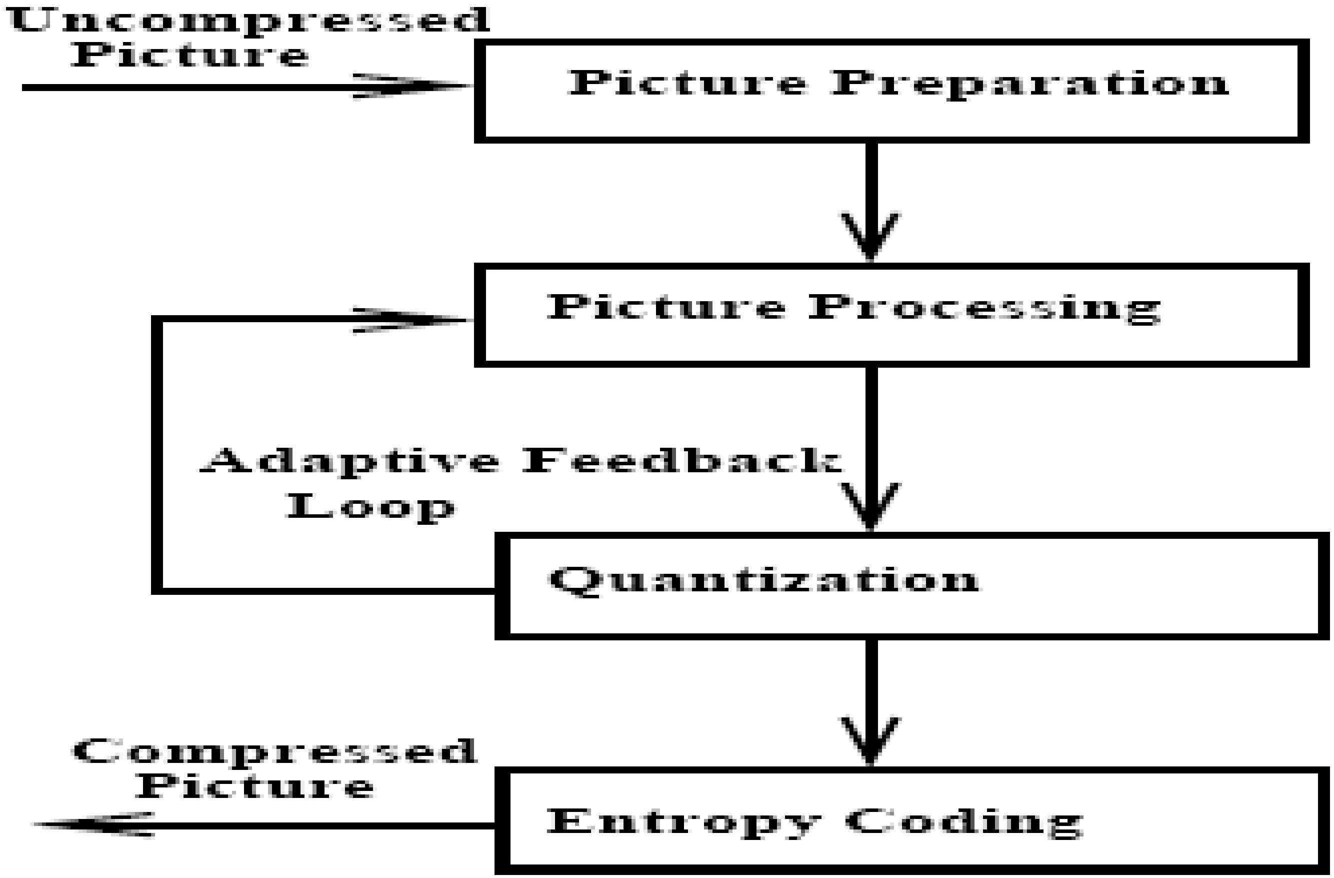
- Gain up to about 20% compression performance to the first JPEG standard
- Applications of JPEG-2000
  - Large images
  - Images with low-contrast edges (e.g., medical images
  - In printers, scanners, facsimile
  - HD satellite images

# Conclusion - Artifacts of JPEG-2000 Compression

- Compression 1/20 size is without incurring visible artifacts
- If artifacts occur they can be seen as Smoothing rather than squares or mosquito noise

# VIDEO MPEG COMPRESSION FORMAT (BASIC CONCEPTS)

# 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)

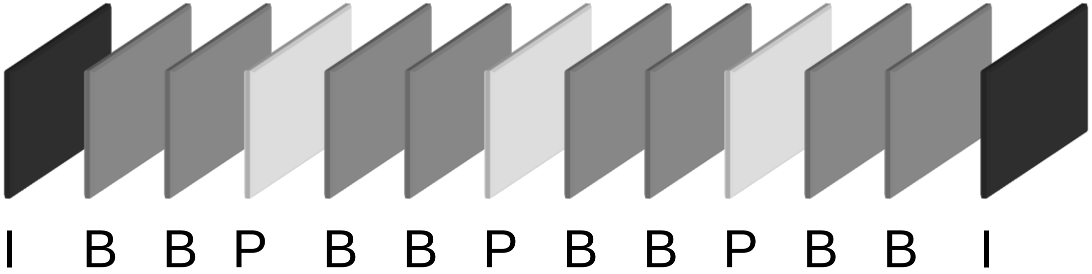
- MPEG defines exactly format
  - Three components: Luminance and two chrominance components (2:1:1) YCbCr
  - YCbCr is family of color spaces, Y is luma, Cb and Cr are blue-difference and red-difference chroma component
  - Resolution of luminance comp:X1 ≤ 768;  Y1 ≤ 576 pixels
  - Pixel precision 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

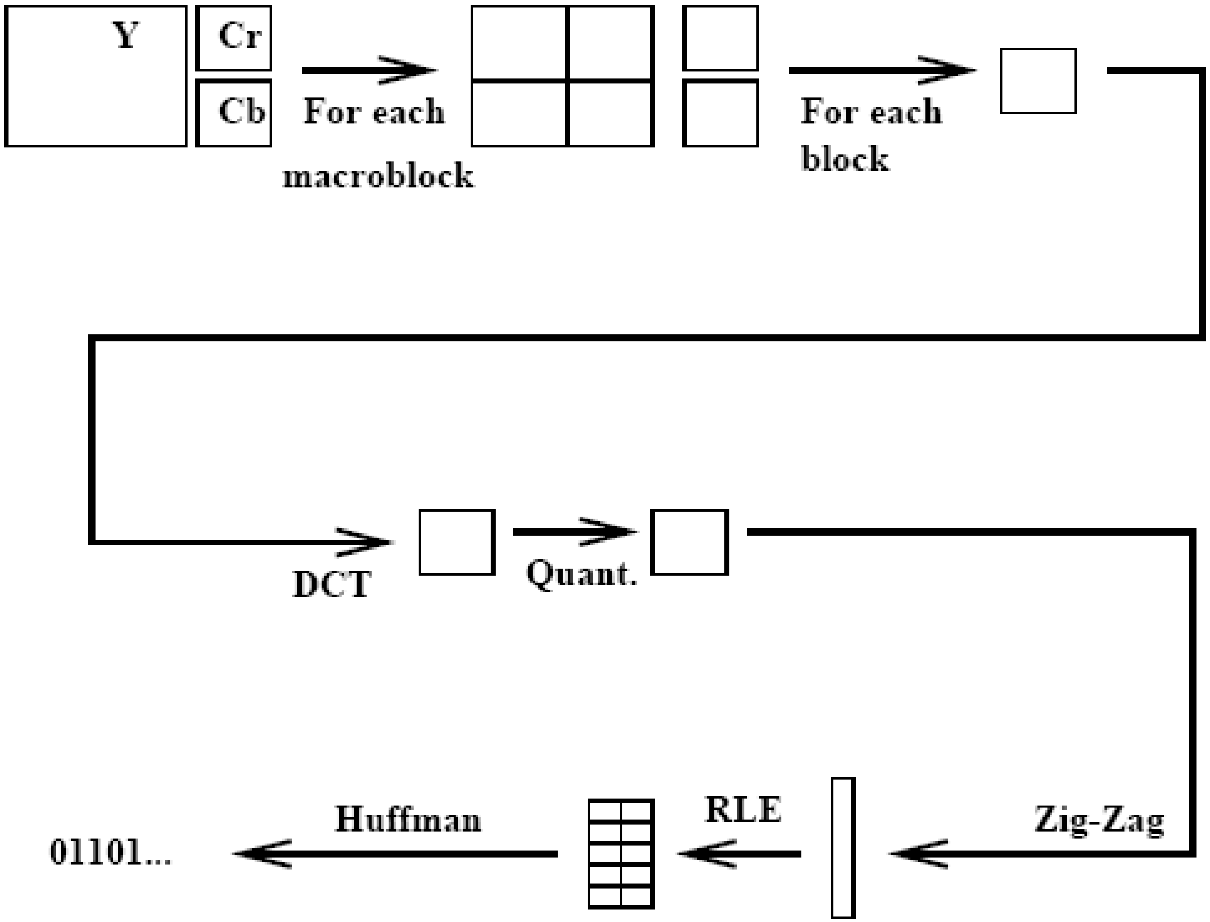


# Selecting I, P, or B Frames

- 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

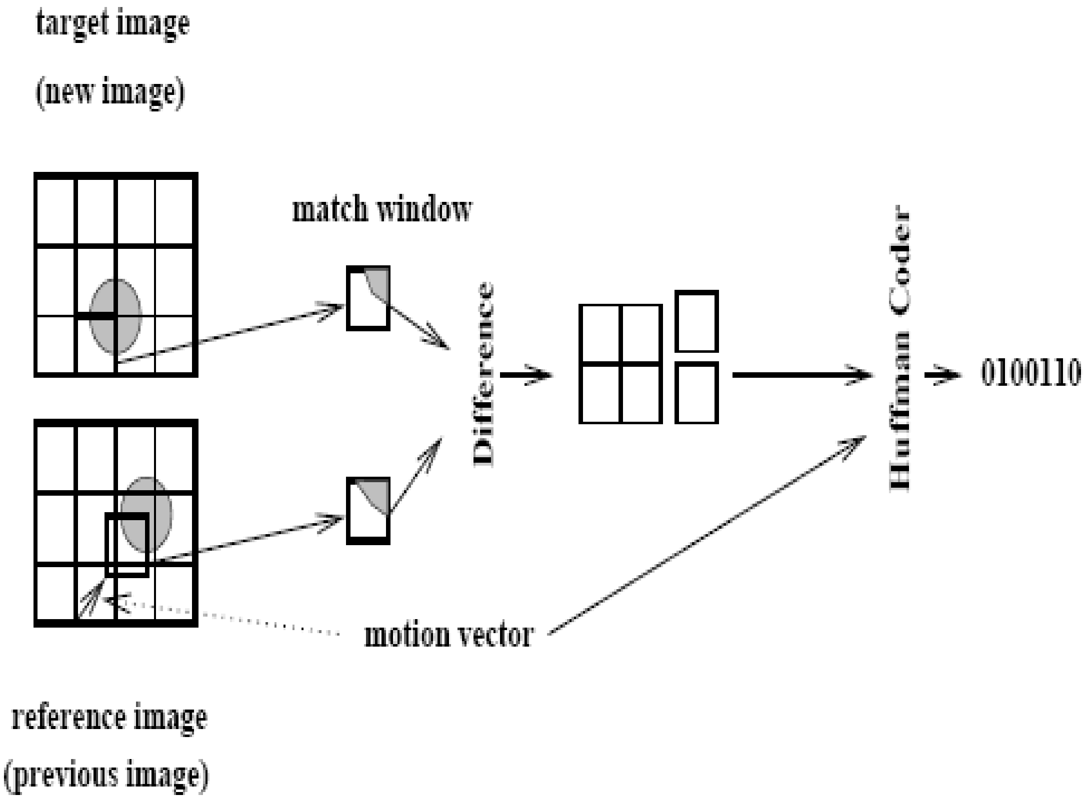
I-frames – points of random access in MPEG stream

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

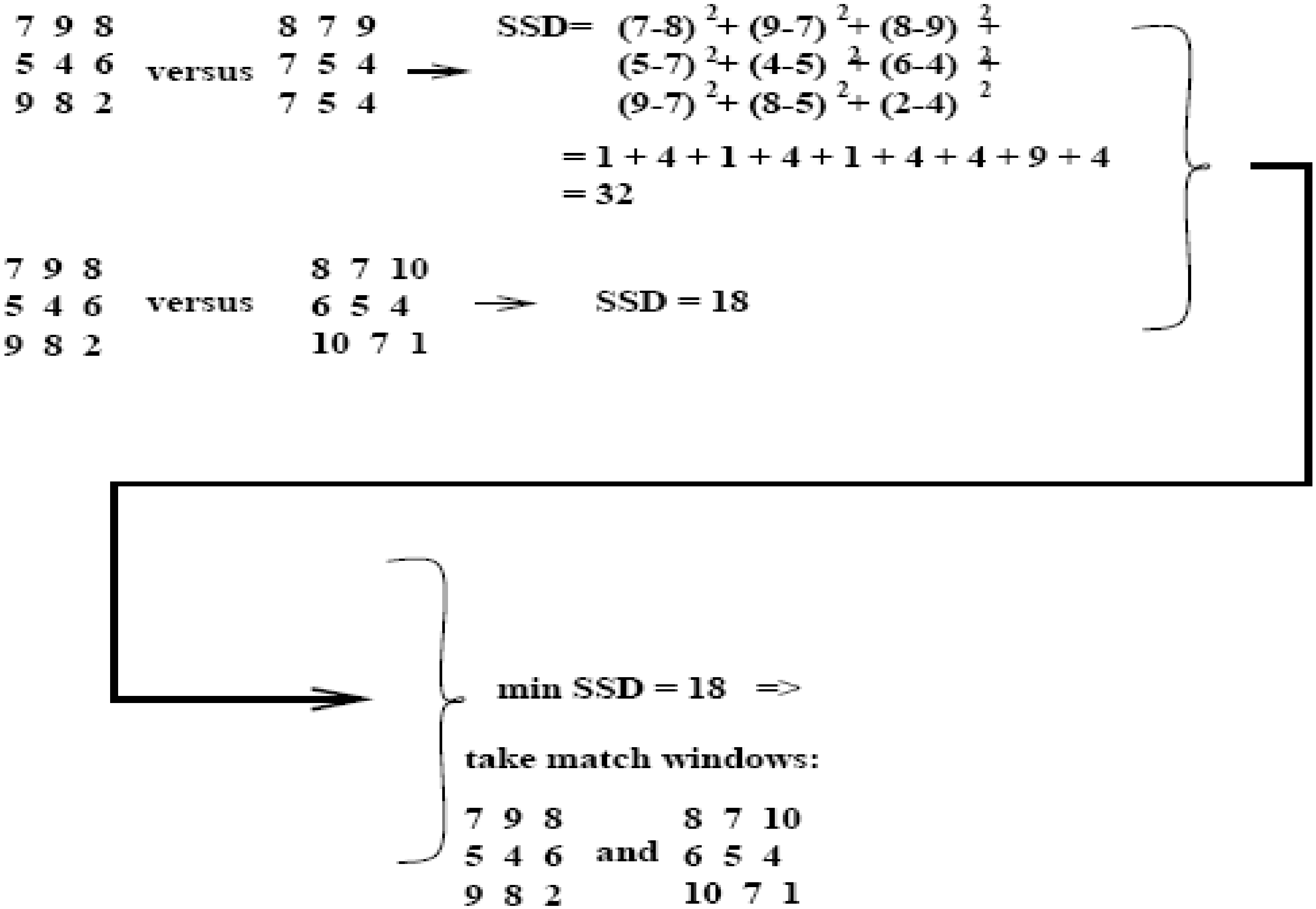
$$SSD = \sum_{i=0}^{N-1} (x_i - y_i)^2$$

- 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



$$\begin{matrix} 7 & 9 & 8 \\ 5 & 4 & 6 \\ 9 & 8 & 2 \end{matrix} \quad \text{versus} \quad \begin{matrix} 8 & 7 & 9 \\ 7 & 5 & 4 \\ 7 & 5 & 4 \end{matrix} \longrightarrow$$

$$\begin{aligned} SSD = \; & (7-8)^2 + (9-7)^2 + (8-9)^2 + \\ & (5-7)^2 + (4-5)^2 + (6-4)^2 + \\ & (9-7)^2 + (8-5)^2 + (2-4)^2 \\ & = 1 + 4 + 1 + 4 + 1 + 4 + 4 + 9 + 4 \\ & = 32 \end{aligned}$$

$$\begin{matrix} 7 & 9 & 8 \\ 5 & 4 & 6 \\ 9 & 8 & 2 \end{matrix} \quad \text{versus} \quad \begin{matrix} 8 & 7 & 10 \\ 6 & 5 & 4 \\ 10 & 7 & 1 \end{matrix} \longrightarrow \quad SSD = 18$$

min SSD = 18 =>

take match windows:

$$\begin{matrix} 7 & 9 & 8 \\ 5 & 4 & 6 \\ 9 & 8 & 2 \end{matrix} \quad \text{and} \quad \begin{matrix} 8 & 7 & 10 \\ 6 & 5 & 4 \\ 10 & 7 & 1 \end{matrix}$$

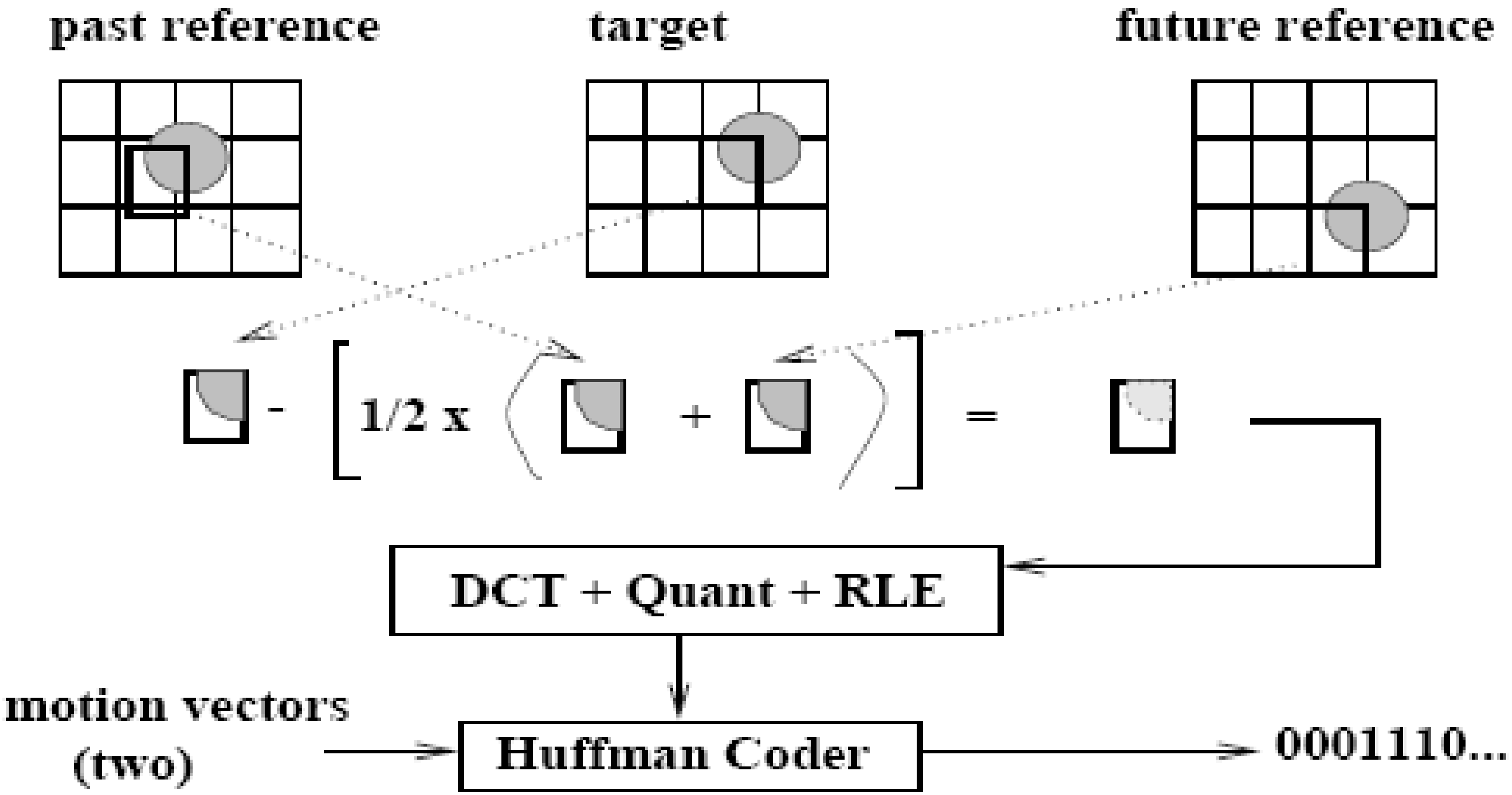
# Syntax of P Frame

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 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**

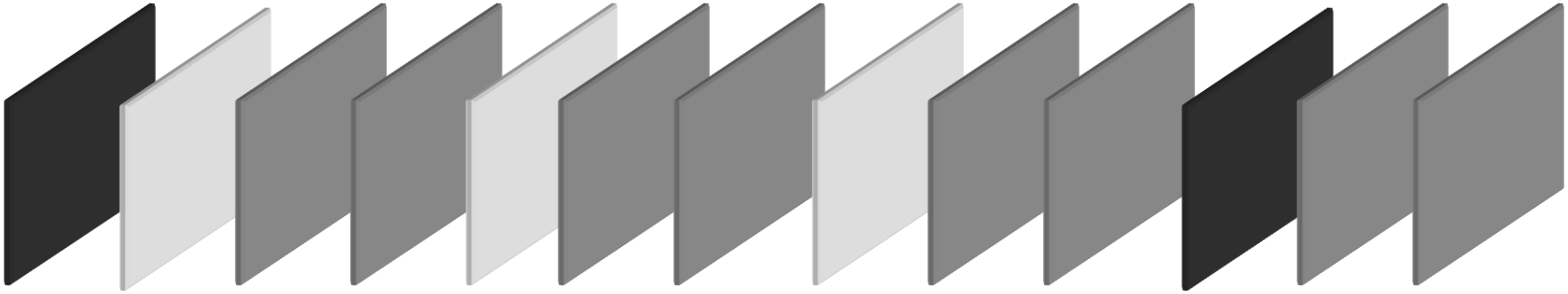


# MPEG Video Decoding

Display Order

$I_1$ $B_1$ $B_2$ $P_1$ $B_3$ $B_4$ $P_2$ $B_5$ $B_6$ $P_3$ $B_7$ $B_8$ $I_2$

Decoding Order



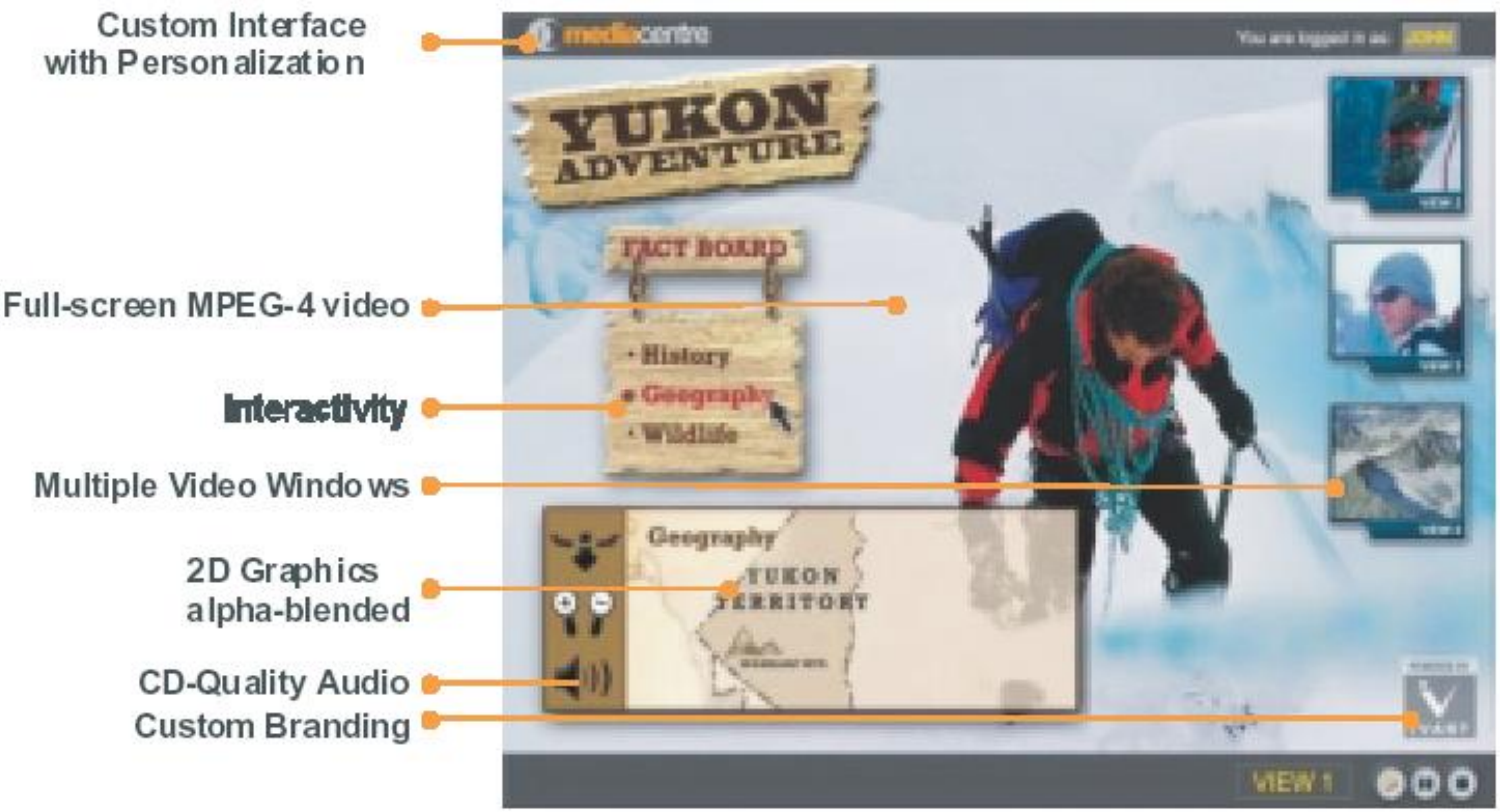
$I_1$ $P_1$ $B_1$ $B_2$ $P_2$ $B_3$ $B_4$ $P_3$ $B_5$ $B_6$ $I_2$ $B_7$ $B_8$

# 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

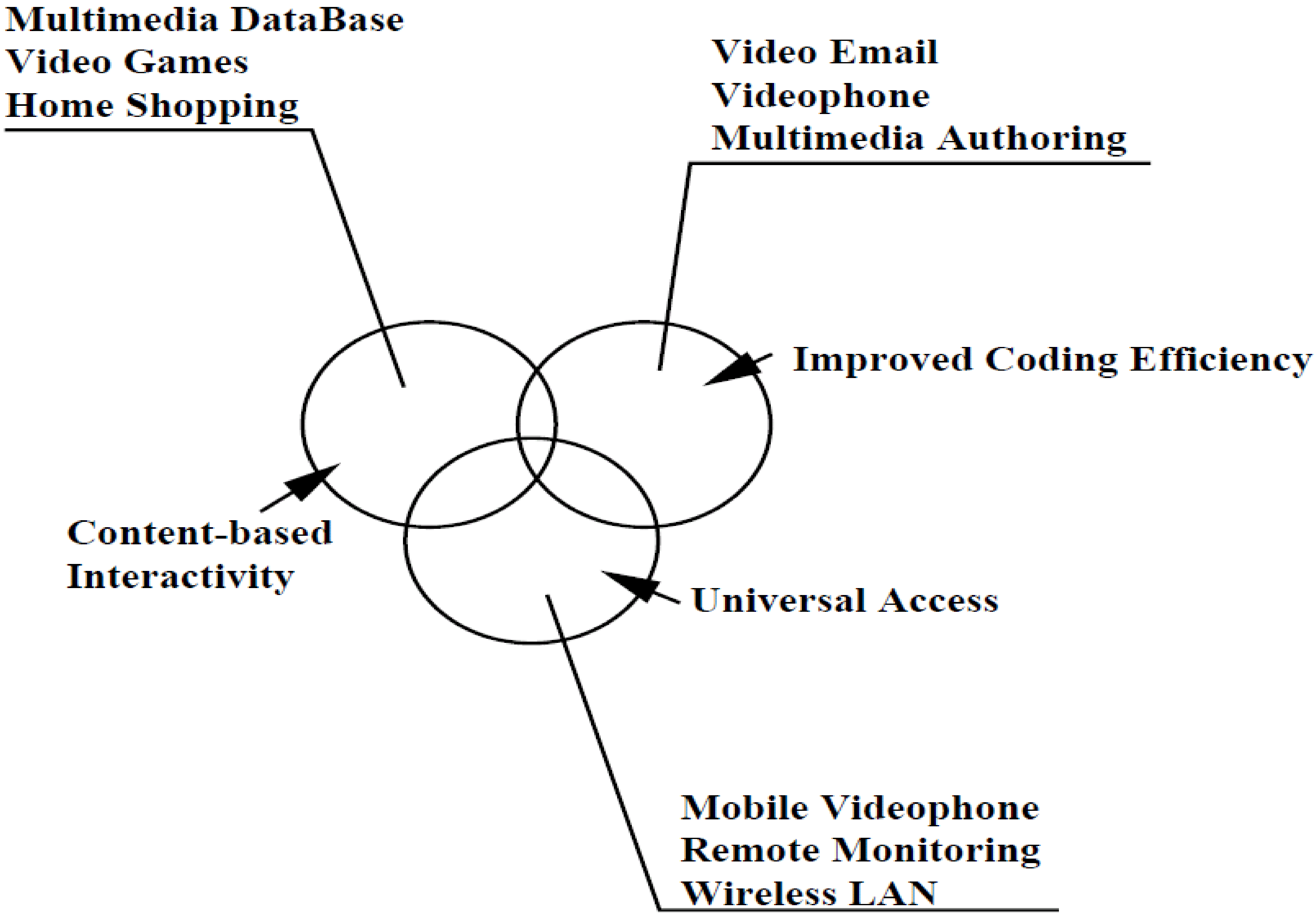
# MPEG-4

# MPEG-4 Example



ISO N3536 MPEG4

# MPEG-4 Characteristics and Applications



Multimedia DataBase
Video Games
Home Shopping

Video Email
Videophone
Multimedia Authoring

Improved Coding Efficiency

Content-based
Interactivity

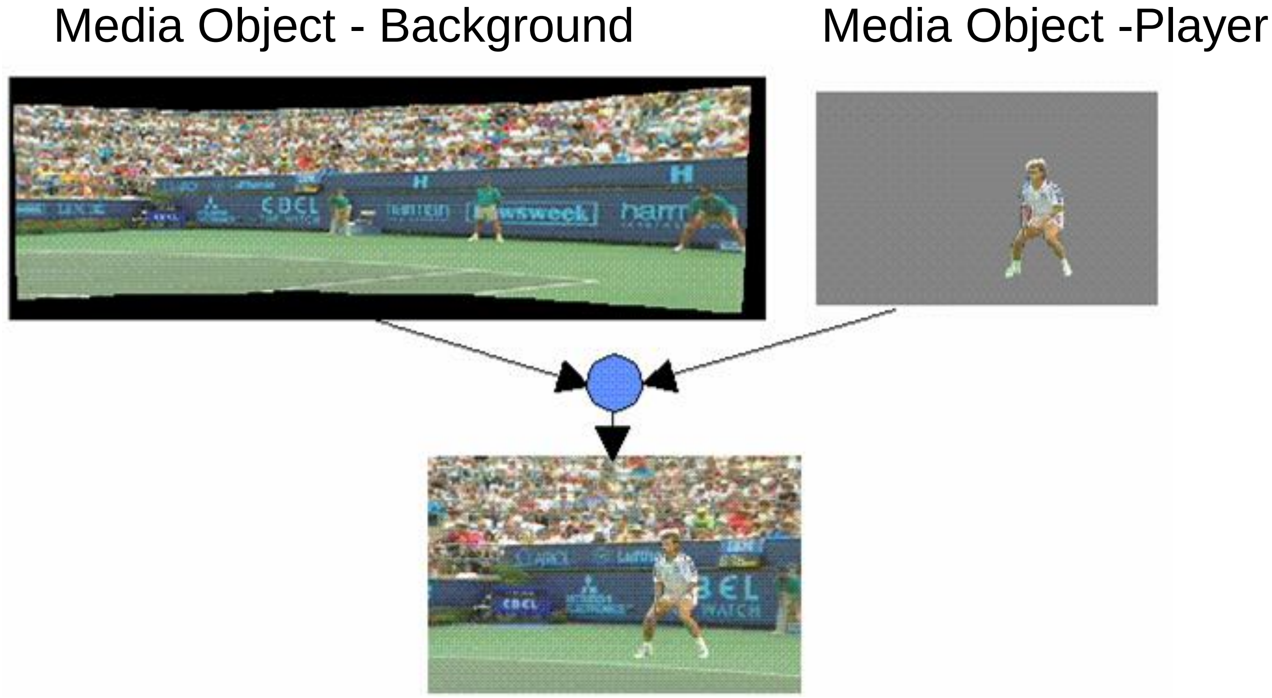
Universal Access

Mobile Videophone
Remote Monitoring
Wireless LAN

# Media Objects

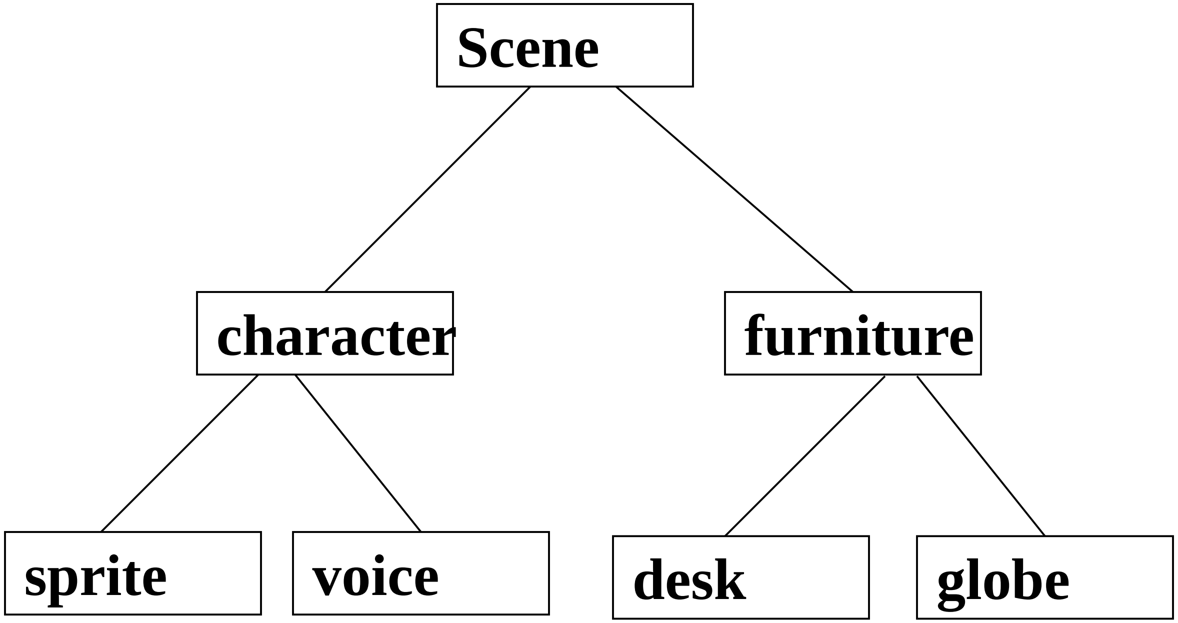
- An object is called a *media object*
  - real and synthetic images; analog and synthetic audio; animated faces; interaction
- Media objects have
  - Spatial relationships
  - Temporal relationships
- Compose media objects into a hierarchical representation
  - form compound, dynamic scenes

# MPEG-4 Example

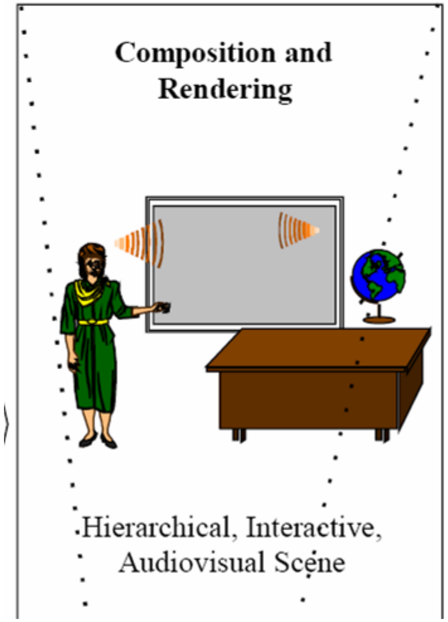


ISO N3536 MPEG4

# Spatial Relationship - Composition (Scene Graph)



- Scene
  - character
    - sprite
    - voice
  - furniture
    - desk
    - globe

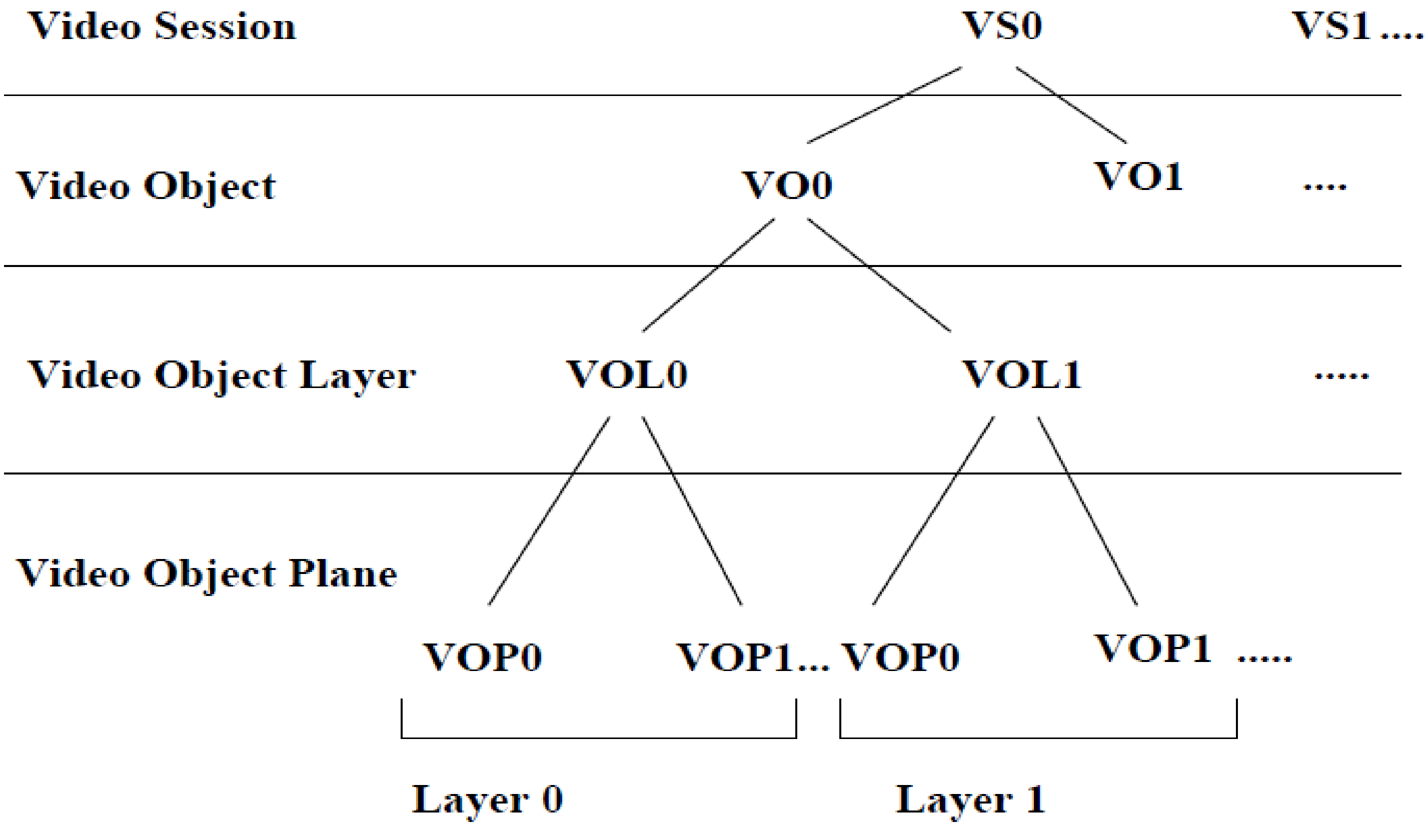


ISO N3536 MPEG4

# Temporal Relationships

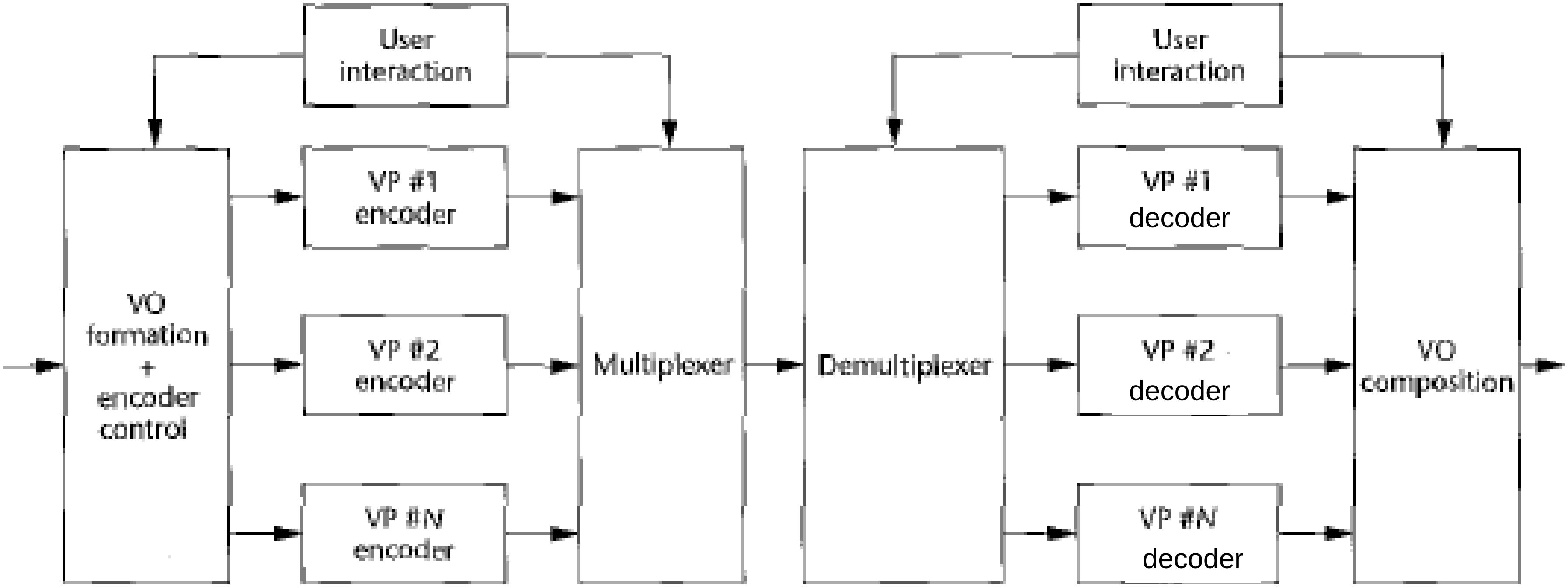
- Composition stream (BIFS) has its own associated time base
  - **Composition timestamps** specify at what time access units for composition must be ready at input of composition information decoder
- Timestamps are attached to each elementary stream
  - **Decoding timestamp (DTS)** specifies at which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

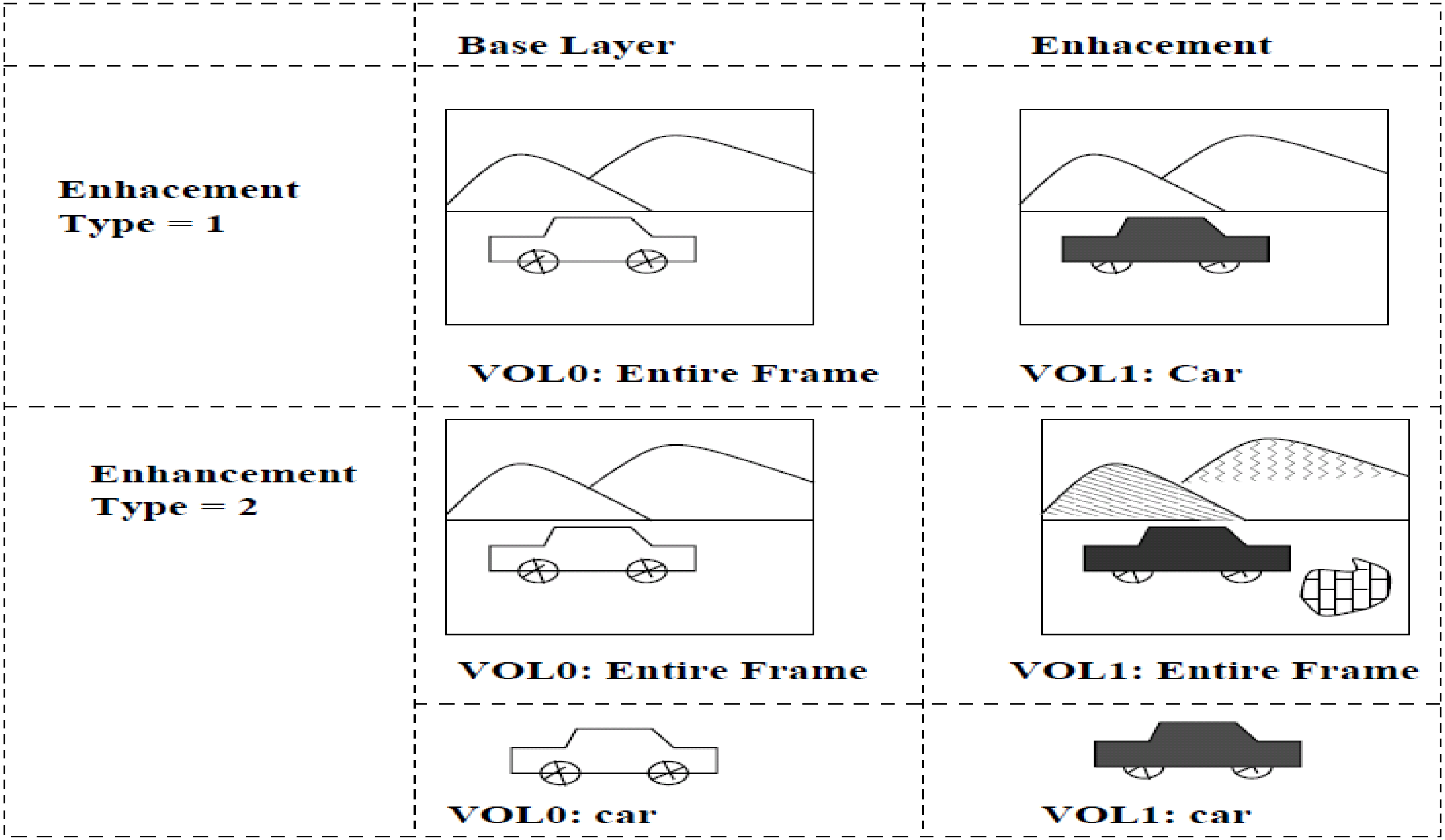


**New MPEG-4 Aspect: Object-based layered syntactic structure**

# MPEG-4 Coding Architecture



# Examples of Base and Enhancement Layers

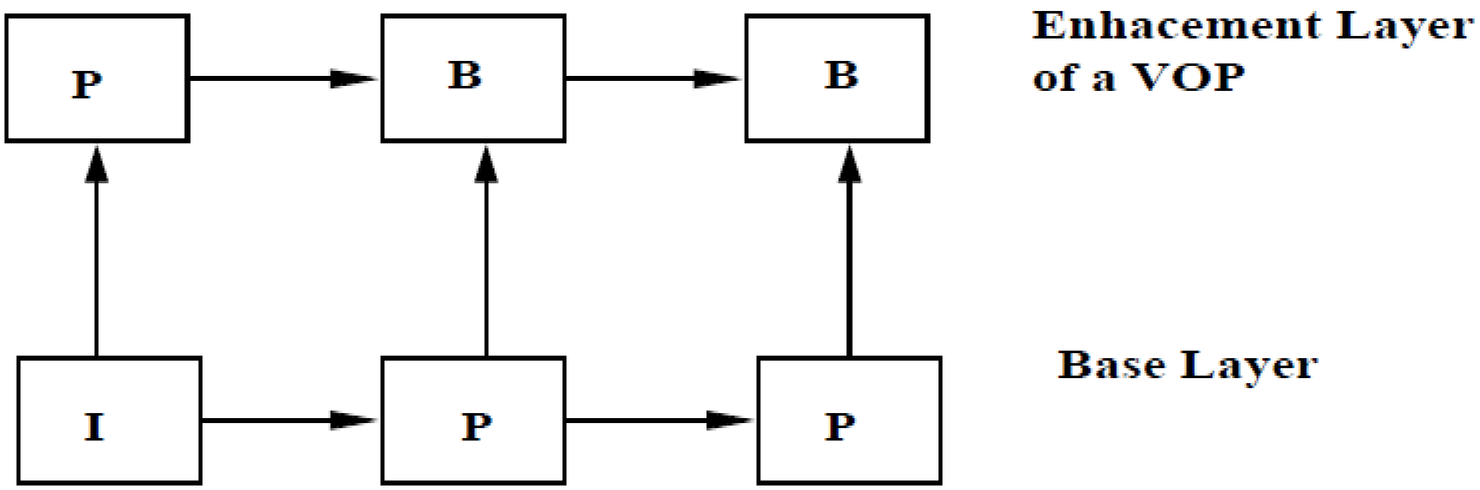


| | Base Layer | Enhacement |
|---|---|---|
| Enhacement Type = 1 | VOL0: Entire Frame | VOL1: Car |
| Enhacement Type = 2 | VOL0: Entire Frame | VOL1: Entire Frame |
| | VOL0: car | VOL1: car |

# Coding of Objects

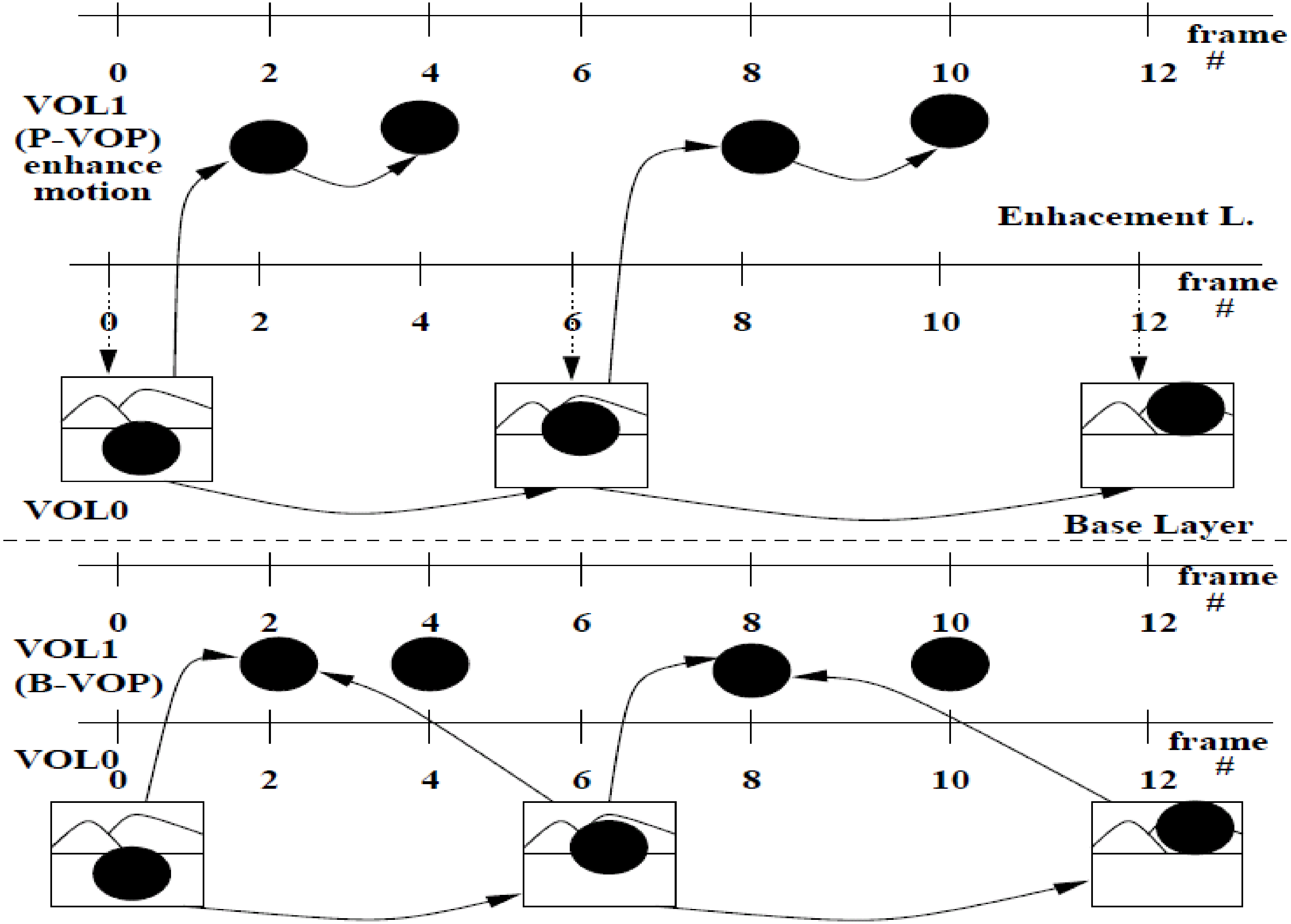
- Each VOP corresponds to an entity that after being coded is added to the bit stream
- Encoder sends together with VOP
  - Composition information where and when each VOP is to be displayed
- Users are allowed to change the composition of the entire scene displayed by interacting with the composition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

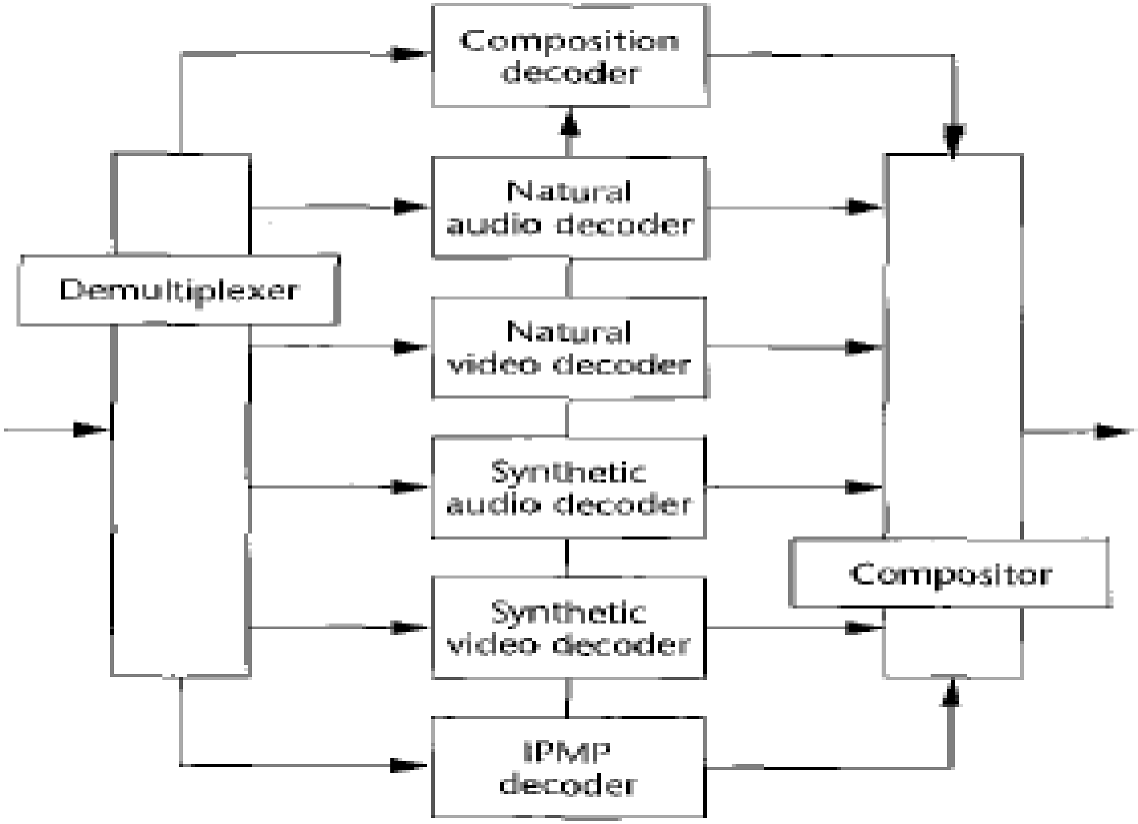
# Temporal Scalability



# Composition (cont.)

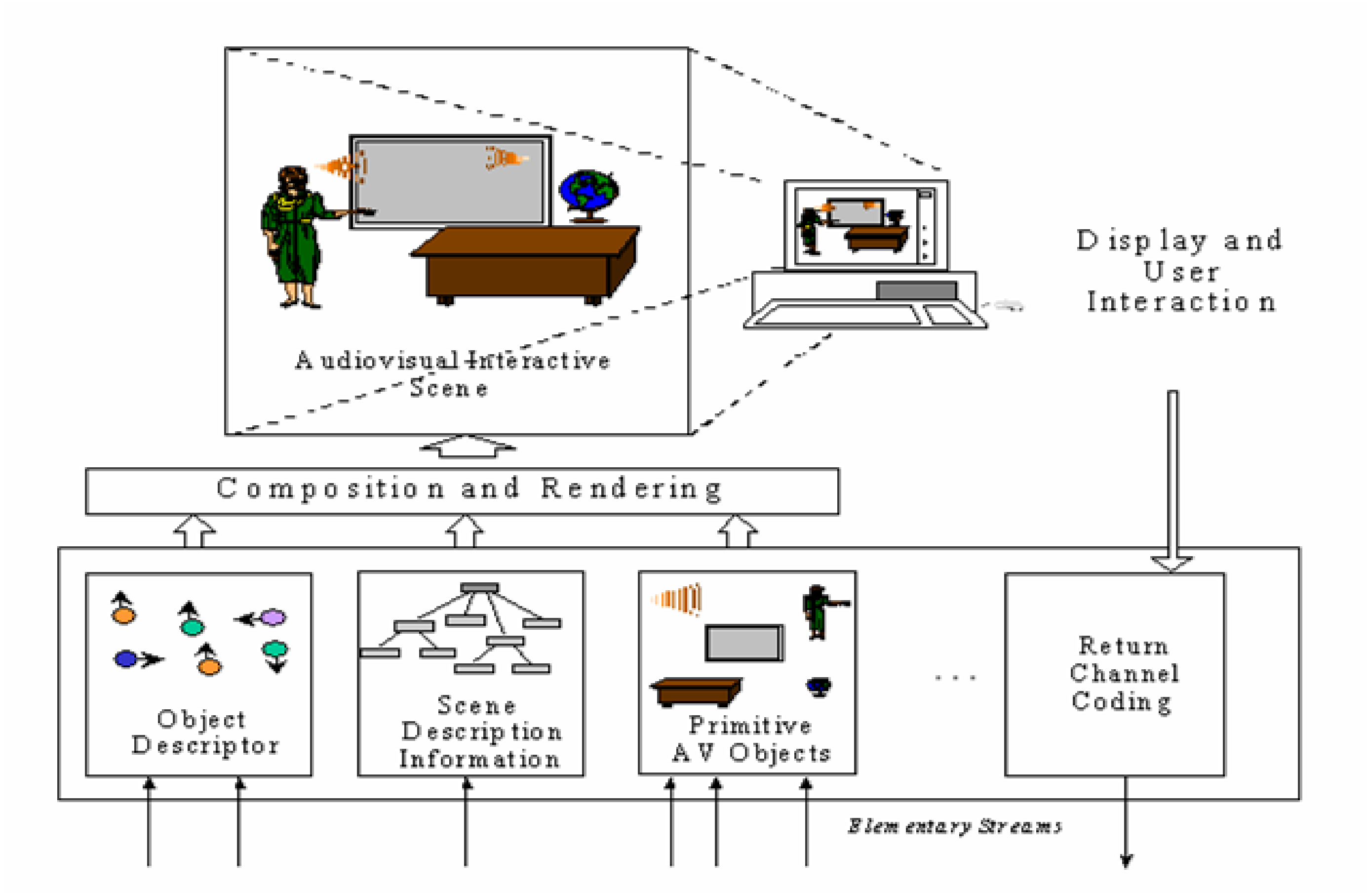
- Encode objects in separate channels
  - encode using most efficient mechanism
  - transmit each object in a separate *stream*
- Composition takes place at the decoder, rather than at the encoder
  - requires a binary scene description (BIFS)
- BIFS is low-level language for describing:
  - hierarchical,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

# MPEG-4 Decoder Architecture



Source: Batista et al, “MPEG-4:A multimedia standard for the third milenium: Part 1”, IEEE Multimedia, 1999

# MPEG-4 Rendering



ISO N3536 MPEG4

# Summary

- MPEG is a very prevalent video coding format, especially in video-on-demand
- MPEG-2 and MPEG-4 are in content distribution systems
- Next lecture:
  - H.264/AVC (MPEG-4/AVC)
  - H.265/ HEVC
  - VP9, AV1, Future Trends