Multimedia Communications

Multimedia Technologies & Applications

Prof. Dr. Abdulmotaleb El Saddik

Multimedia Communications Research Laboratory

School of Information Technology and Engineering
University of Ottawa

Ottawa, Ontario, Canada

elsaddik @ site.uottawa.ca

abed @ mcrlab.uottawa.ca

Content

- Motivation
- Requirements - General
- Fundamentals - Categories
- Coding
- Image Compression
  - JPEG
  - JPEG2000
- Video Compression
  - Video Compression Concepts
  - MPEG-1
  - MPEG-2
  - MPEG-4
  - MPEG-7
  - MPEG-21
- H.261 and related ITU Standards
- Conclusion

Motivation

Digital video in computing means for

- **Text:**
  - 1 page with 80 char/line and 64 lines/page and 2 Byte/Char
    - 80 x 64 x 2 x 8 = 80 kBit/page
- **Image:**
  - 24 Bit/Pixel, 512 x 512 Pixel/image
    - 512 x 512 x 24 = 6 MBit/Image
- **Audio:**
  - CD-quality, samplerate44,1 kHz, 16 Bit/sample
    - Mono: 44,1 x 16 = 706 kBit/s
    - Stereo: 706 x 2 = 1.412 MBit/s
- **Video:**
  - full frames with 1024 x 1024 Pixel/frame, 24 Bit/Pixel, 30 frames/s
    - 1024 x 1024 x 24 x 30 = 720 MBit/s
  - more realistic
    - 360 x 240 Pixel/frame = 60 MBit/s

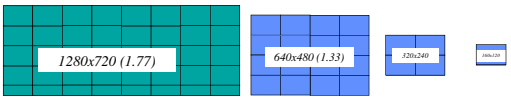
Hence compression is *necessary*

Video Data Size

size of uncompressed video in gigabytes

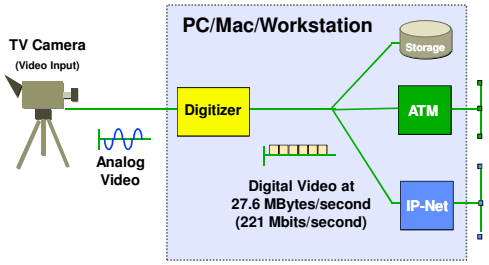
	1920x1080	1280x720	640x480	320x240	160x120
1 sec	0.19	0.08	0.03	0.01	0.00
1 min	11.20	4.98	1.66	0.41	0.10
1 hour	671.85	298.60	99.53	24.88	6.22
1000 hours	671,846.40	298,598.40	99,532.80	24,883.20	6,220.80

image size of video



4

Digitizing Video

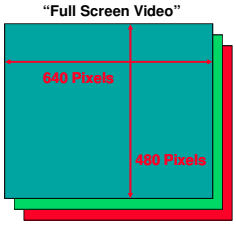


Transforming from Time domain into Frequency domain

Data in the transformed domain is easier to compress

5

Video Digitization Parameters



(Pixel Area) x (Color Depth) x (Frame Rate) = Bandwidth

(640*480) * 3 * 30 = 27,648,000 Bytes/s

Or

221,184,000 bits per second (221Mbps)

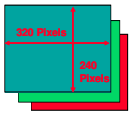
3-bytes (24-bit) Color

NTSC = 30 frames per second
PAL/SECAM = 25 frames per second

6

Modify Digitization Parameters

1/4 Screen Video



1-byte (8-bit) Color

10 frames per second

(Pixel Area) x (Color Depth) x (Frame Rate) = Bandwidth

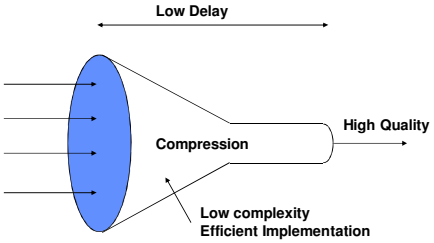
(320*240) * 1 * 10 = 768,000 Bytes/sec

Or

6,144,000 bits/sec (6.14Mbps)

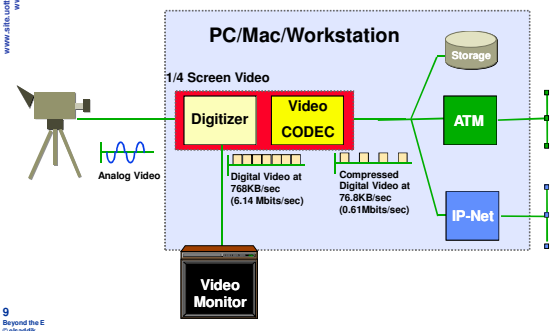
7

COMPRESSION



8

Video Compression



9

Video Bit Rate Calculation

$$\left[\frac{\text{width} * \text{height} * \text{depth} * \text{fps}}{\text{compression factor}}\right] = \text{bits/sec}$$

width = {160, 320, 640, 720, 1280, 1920, …} pixels
height = {120, 240, 480, 485, 720, 1080, …} pixels
depth = {1, 4, 8, 15, 16, 24, …} bits
fps = {5, 15, 20, 24, 30, …} frames per second
compression factor = {1, 6, 24, …}

Effects of Compression

storage for 1 hour of compressed video in megabytes

	1920x1080	*1280x720*	*640x480*	*320x240*	*160x120*
1:1	671,846	298,598	99,533	24,883	6,221
3:1	223,949	99,533	33,178	8,294	2,074
6:1	111,974	49,766	16,589	4,147	1,037
25:1	26,874	11,944	3,981	995	249
100:1	6,718	2,986	995	249	62

3 bytes/pixel, 30 frames/sec

Constraints for Multimedia Systems

- Compression/Decompression
  - Uncompressed video is rather tolerant to transmission errors, but compressed one is very sensitive to network transmission impairments
  - Decompression of images, video or audio should be possible without the knowledge of other data units. Media synchronization will take care of all presentational needs
- In conversational multimedia applications (Dialogue-based)
  - The overall end-to-end delay should not exceed 150 ms. However, the compression-decompression delay should be reduced to about 50 ms, to leave room for other network introduced delay components
- For presentational multimedia (Retrieval mode)
  - fast-forward,-backward, with simultaneous display, should be possible
  - random access to single images and audio frames of a data stream, with access time less than 0.5 sec

Coding Categories

Source coding
- based on data semantics
- often lossy

Entropy coding
- ignoring data semantics
- often lossless

Channel coding
- adaptation to comm. channel
- introduction of redundancy

Hybrid coding
- entropy
- source coding

13

Coding Categories and Techniques

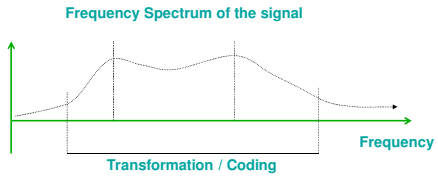
Entropy coding	Run-Length coding	
	Huffman coding	
	Arithmetic coding	
Source coding	Prediction	DPCM
		DM
	Transformation	FFT
		DCT
	Layered coding	Bit Position
		Sub-sampling
		Sub-band coding
	Vector Quantization	
Hybrid coding	JPEG	
	MPEG	
	H.261	
	Proprietary format: Quicktime, DVI RTV, DVI PLV,...	

14

Source Coding: Sub-Band

Assumption:

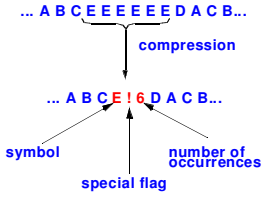
➢ Some frequency ranges are more important than others



15

Entropy Coding: Run-Length

- Assumption:
  - Long sequences of identical symbols
- Example:
  - Special variant: zero-length encoding
  - only repetition of zeroes count
  - in red, "symbol" not needed (i.e. "pays" for >3 repetitions)



16

Huffman Coding

Assumption:

- Some symbols occur more often than other
- E.g., characters in an English text

Principle:

- Frequently occuring symbols are coded with shorter strings

Table:

Symbol	Code
A	001
B	1
C	011
D	000
E	010

A	B	D	C	A	A	E
001	1	000	011	001	001	010

17

Basic Encoding Steps

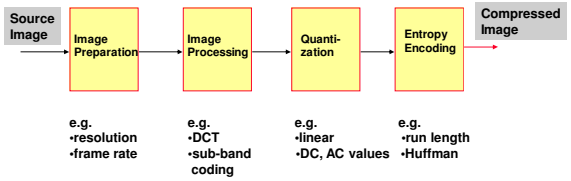


Image Preparation	Image Processing	Quanti-zation	Entropy Encoding
e.g. •resolution •frame rate	e.g. •DCT •sub-band coding	e.g. •linear •DC, AC values	e.g. •run length •Huffman

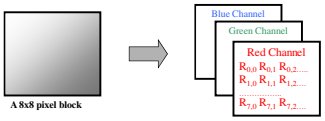
18

JPEG International Standard since 1992

- JPEG (Joint Photographic Expert Group) is a standard jointly developed by ISO/EIC JTC1/SC2/WG10 and ITU-TS
- Can compress still images 10:1 and 50:1, without visibly affecting image quality
- Uses the Discrete Cosine Transform (DCT) for coding and decoding
- Aimed at photographic material, thus it does not have embedded encoded/compressed audio signal
- Standard supports four modes
  - Sequential encoding
  - Progressive encoding
  - Lossless encoding
  - Hierarchical encoding
- 10 MHz JPEG chip can compress full-page, 300 dpi, 24-bit colour image in 1 sec or a 640 x 480 24-bit-colour image in 0.1 sec
- 25 MHz chip can do the same in 0.03 sec (Real-time video compr.)

Image Preparation

- Before performing DCT, divide image into blocks (usually 8x8 pixels)
- Each pixel in an image is represented by image components, usually 3 components (e.g. the RGB components)
- These are normally converted to a different format, such as the YUV, where the luminance and chrominance are separated.



Discrete Cosine Transform (DCT)

- The blocks of components are then transformed into frequency components.
- Each block is represented by 64 frequency components, one of which (the DC component) is the average colour of the 64 pixels in the block

- Applying FDCT to change the blocks from spatial domain to frequency domain
- Roughly speaking, F(0,0) is the average value within the block (called DC component).
- Other values stores the variation within a block (AC components)

Discrete Cosine Transform (DCT)

For $0 \le u \le 7, \quad 0 \le v \le 7$

$$F(u,v) = 0.25 C(u) C(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos\frac{(2x+1)u\pi}{16} \cos\frac{(2y+1)v\pi}{16}$$

WITH:

$$C(u) = C(v) = 1/\sqrt{2}, \quad for \ u,v = 0, \quad and \quad C(u) = C(v) = 1$$

Formula applied 64 times

- Blocks with 8x8 pixels result in 64 DCT coefficients
  - 1 DC coefficient (basic colour of block): most important
  - 63 AC coefficients: many zero or near-zero values

22

Example Fast DCT

Pixel block

-127	-128	-123	-119	-115	-111	-107	-103
-128	-123	-119	-115	-111	-107	-103	-100
-123	-119	0	-111	-107	-2	-100	-95
-119	-115	-111	-107	-103	-100	-95	-91
-111	-111	-107	-103	-100	-95	-91	15
-111	-15	-23	-30	-21	-20	-2	-83
-107	-103	-100	-95	-91	-87	-83	-79
-103	-100	-95	-91	-87	-83	-79	-75

FDCT

-711.25	-66.59	-0.02	-8.63	-22.00	-1.24	36.54	3.57
-105.04	0.21	-4.02	-0.92	-10.18	1.68	35.04	0.67
-81.72	-4.87	-26.86	0.47	-7.80	-0.50	-30.26	-2.16
58.90	3.08	19.46	4.13	27.20	-1.00	-56.33	-0.36
-58.00	-0.03	73.05	6.19	78.25	3.76	-5.90	-2.49
-42.73	-2.63	-39.20	-2.30	-32.02	-1.07	-3.90	-0.20
97.98	3.13	-66.26	-5.85	-79.57	-2.19	20.86	4.63
0.00	3.38	58.37	2.11	33.65	4.57	78.96	0.73

Block of DCT coefficients

23

Quantization Table

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

- DCT coefficients are divided by Quantization values and rounded off
- Q-values are chosen based on DCT co-eff. importance

24

Example

Quantization Table

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

-711.25 /16
= -44

-711.25	-66.59	-0.02	-8.63	-22.00	-1.24	36.54	3.57
-105.04	0.21	-4.02	-0.92	-10.18	1.68	35.04	0.67
-81.72	-4.87	-26.86	0.47	-7.80	-0.50	-30.26	-2.16
58.90	3.08	19.46	4.13	27.20	-1.00	-56.33	-0.36
-58.00	-0.03	73.05	6.19	78.25	3.76	-5.90	-2.49
-42.73	-2.63	-39.20	-2.30	-32.02	-1.07	-3.90	-0.20
97.98	3.13	-66.26	-5.85	-79.57	-2.19	20.86	4.63
0.00	3.38	58.37	2.11	33.65	4.57	78.96	0.73

Block of DCT coefficients

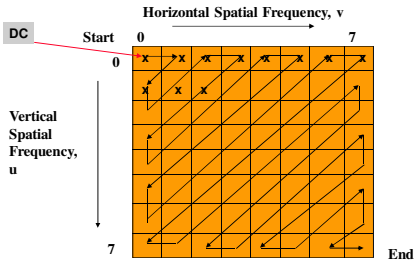
-44	-6	0	-1	-1	0	1	0
-9	0	0	0	0	0	1	0
-6	0	-2	0	0	0	0	0
4	0	1	0	1	0	-1	0
-3	0	2	0	1	0	0	0
-2	0	-1	0	0	0	0	0
2	0	-1	0	-1	0	0	0
0	0	1	0	0	0	1	0

Quantized Values

25

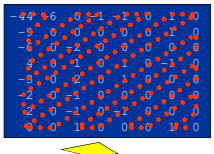
Zig-Zag sequence

- Use Zig-Zag walks to order the coefficients sequentially



26

Example

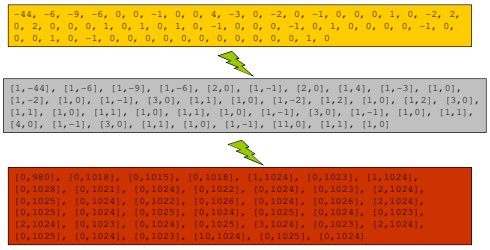


-44, -6, -9, -6, 0, 0, -1, 0, 0, 4, -3, 0, -2, 0, -1, 0, 0, 0, 1, 0, -2, 2, 0, 2, 0, 0, 0, 1, 0, 1, 0, 1, 0, -1, 0, 0, 0, -1, 0, 1, 0, 0, 0, 0, -1, 0, 0, 0, 1, 0, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0

27

Run-Length Encoding

- Coefficients are represented by [Repeats-1] [coefficient+1024]

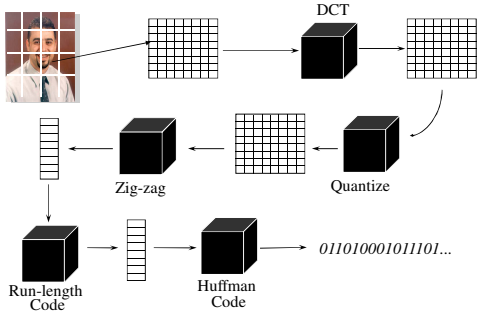


Run-Length Encoding

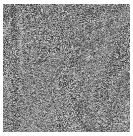
- *Repeats-1* has maximum value of 63, need 6 bits.
- *Quantized coefficient+1024* has range [0,2047], need 11 bits.

[0,980]	00000 0	01111 01010 0
[0,1018]	00000 0	01111 11101 0
[0,1015]	00000 0	01111 11011 1
[0,1018]	00000 0	01111 11101 0
. . .	. . .	
[0,1024]	00000 0	10000 00000 0

Block Transform Encoding



Examples



Uncompressed (262 KB)

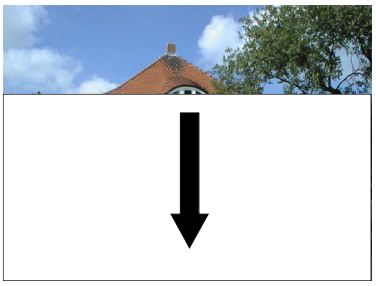
Compressed (22 KB, 12:1)

Compressed (6 KB, 43:1)

JPEG examples

Sequential image display:

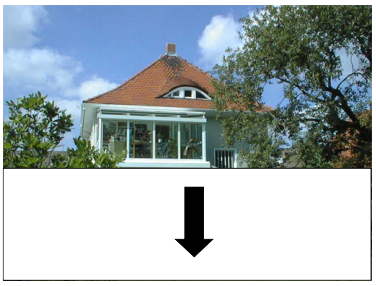
- Top to bottom
- Good for small images and fast processing



JPEG examples

Sequential image display:

- Top to bottom
- Good for small images and fast processing



JPEG examples

Sequential image display:

- Top to bottom
- Good for small images and fast processing

34

Progressive image display

- Coarse to fine
- Good for large and complicated images

35

Progressive image display

- Coarse to fine
- Good for large and complicated images

36

Progressive image display

- Coarse to fine
- Good for large and complicated images

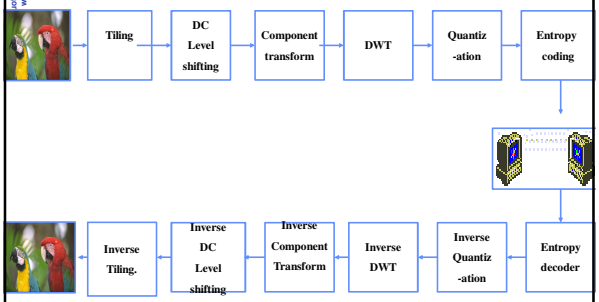
37

Image Compression – JPEG 2000

- **JPEG 2000 is aimed at emerging technology and multimedia such as**
  - Internet.
  - Mobile Multimedia.
  - Digital photography.
  - Scanning, e-commerce, Color facsimile, etc.
- **Distinguished Features**
  - Tiling capability
  - Scalability (quality, resolution…)
  - Progressive decoding by pixel and resolution accuracy
  - Error resilience (robustness to bit errors)
  - Improved performance at low bit rates
  - Lossless and lossy compression
  - Region of interest coding
  - Protective image security features

38

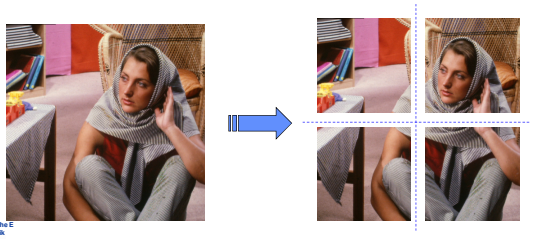
JPEG2000 Coding Scheme



39

Tiling

- The source image is partitioned into rectangular non overlapping blocks called tiles.
- Each Tile is treated independently and can be assigned its own compression parameters



40

Component Transformation

- Color transformation is applied to each tile
- JPEG 2000 supports Irreversible and Reversible Component Transformation (ICT) for lossy and lossless coding

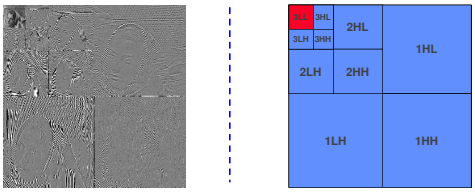
RGB Color Space

YC_bC_r Color Space

41

Discrete Wavelet Transform

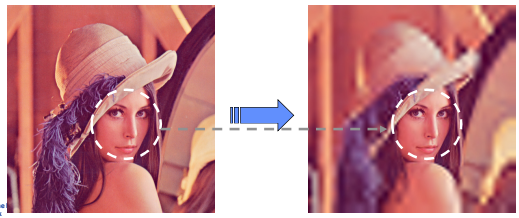
- JPEG2000 uses a Discrete Wavelet Transform in place of the DCT used by JPEG.
- Each tile component is transformed into the wavelet domain using 9/7 or 5/3 filter transformations
- Wavelets decomposes the images into high frequency and low frequency *sub bands*.



42

Quantization

- Quantization is applied on the sub bands coefficients
- Embedded scalar quantization allow for quantization at different layers of quality
- **Realizes the concept of Region of Interest**



43

Comparison with JPEG

Criteria	JPEG 2000	JPEG
Compression performance at		
High bit rates	✔	
Low bit rates	✔	
Functionality i.e. features provided within a unified algorithm	✔	
Rate-distortion at the same bit rate	✔	
Computational Complexity		✔
Lossless performance	✔	

44

Video Compression

In video streams, there are 2 types of redundancy that can be explored:

- **Spatial redundancy**
- **Temporal redundancy**

Recall that spatial redundancy is what JPEG and other still image algorithms use.

- **There are two groups of video compression products:**
  - **Based purely on spatial redundancy**
  - **Based on both spatial and temporal redundancy**

45

Spatial-Redundancy-Only Video Compression

- Called motion JPEG
- Compress each frame individually, without reference to any other frames in the sequence
  - → thus does not consider inter-frame redundancies
- audio is not supported in an integrated fashion
- Motion JPEG Hardware (Chips, boards) for near real-time compression/ decompression available, but storage and retrieval from a hard disc still takes a second or more.
  - High quality video requires fast SCSI discs or cashing of short video sequences in large memory buffers.

46

JPEG for full-motion video

- Advantages:
  - Loss of frames does not affect other frames
  - Less encoding complexity and delay
  - Easier editing
- Disadvantages:
  - network-based JPEG applications unlikely, since it is bandwidth-intensive
    - Typical rate for studio quality TV: 10 ~ 20 Mbps

Basically, lower compression rates is needed

47

Spatial and temporal redundancy video compression – MPEG

We have seen with JPEG how spatial redundancy can be explored. MPEG utilises, as well as spatial redundancy, the fact that frames in a sequence are similar to each other. This is what is known as temporal redundancy.

A few definitions are required here:

- Macroblocks
  - This is a 16x16 pixel block, composed of 4 times 8x8 luminance blocks and 2 colour difference blocks
- Motion Vectors
  - Indicates the spatial translation of a macroblock between two frames.

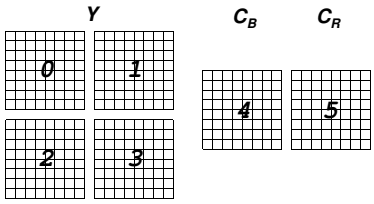
48

Macroblocks

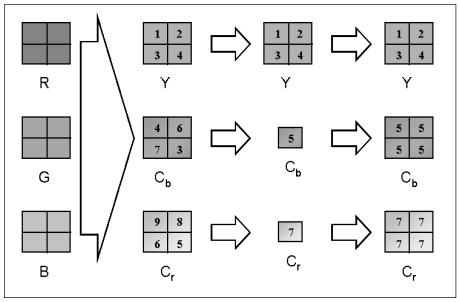
$$Y = 0{,}2125 \cdot r + 0{,}7154 \cdot g + 0{,}0721 \cdot b$$

$$c_r = b - Y$$

$$c_b = r - Y$$



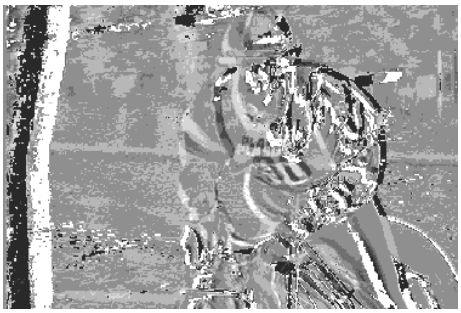
Macroblocks



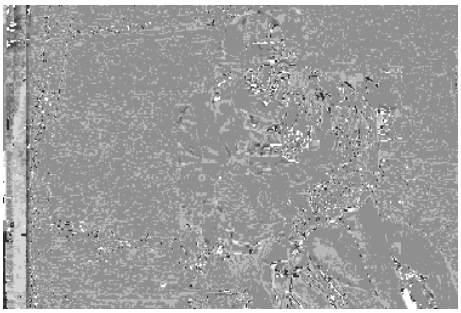
Motion Vectors: Basis image

Motion Vectors: 2nd Image with motion

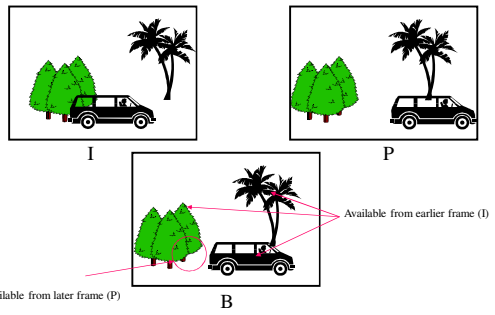
Motion Vectors: Difference without motion compensation



Motion Vectors: Difference with motion compensation



Motion estimation for different frames



55

MPEG

- Motion Picture Expert Group (MPEG)
  - ISO/IEC working group(s)
  - ISO/IEC JTC1/SC29/WG11
  - ISO IS 11172 since 3/93
- coding of combined:
  - video and audio information
- Starting point: MPEG-1
  - Audio/video at about 1.5 Mbit/s
  - Based on experiences with JPEG and H.261
- Follow-up standards
  - MPEG-2
  - MPEG-4
  - MPEG-7
  - MPEG-21

56

MPEG

- MPEG
  - allows coding comparison across multiple frames and therefore can yield compression ratios of 50:1 to 200:1
  - MPEG chips
    - provide VHS quality at 1.2-1.5 Mbps and 200:1
    - can also give 50:1 and broadcast video quality at 6 Mbps
- algorithm asymmetrical:
  - more complex to compress than decompress

57

MPEG - Video: Processing Step

4 types of frames:

- I-frames (intra-coded frames):
  - Real-time decoding demands and sometimes in encoding too
  - Compression of I frames the lowest in MPEG
  - I-frames are points for random access in MPEG streams
  - coding and decoding like JPEG
  - Structured in 8x8 blocks, within macroblocks of 16x16, that are DCT coded, quantized and entropy coded

58

MPEG - Video: Processing Step

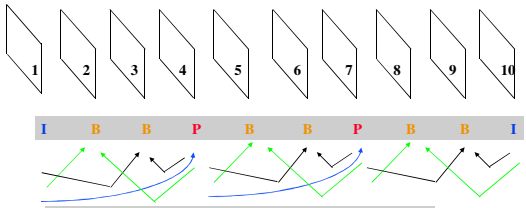
- P-frames (predictive coded frames):
  - Require about 1/3 of data of I-frames
  - Reference to previous I- or P-frames
  - Motion vector calculated
    - MPEG does not define how to determine the motion vector
    - difference of similar macroblocks is DCT coded
  - DC and AC coefficients are runlength coded
- B-frames (bi-directional predictive coded frames):
  - Reference to previous and subsequent (I or P) frames
  - One or two motion vectors are encoded
  - Interpolation between matching macroblocks allowed (both directions)
- D-frames (DC-coded frames):
  - Only DC-coefficients are DCT coded
  - For fast forward and rewind

59

MPEG Video-frame sequence

- I frame: Intra frame
- P frame: Predicted frame
- B frame: Bidirectionally interpolated frame

Sequence
- Defined by application



MPEG coded sequence will be transmitted in different order:

I P B B P B B I B B
1 4 2 3 7 5 6 10 8 9

60

MPEG in a Nutshell

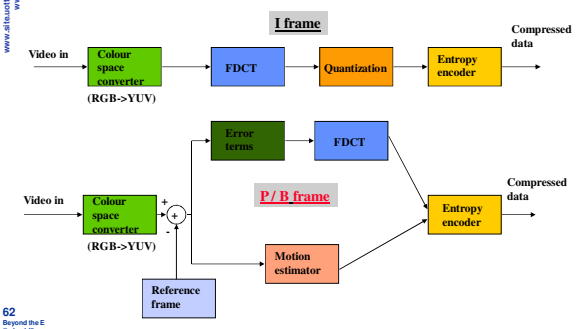
- I-Frames are self contained but less compressed than P and B Frames.
- B-Frames are the most compressed frames.

Typical sequences of frames are:

- I BBB P BBB I…
- I BB P BB P BB I…
- I BB P BB P BB P BB I...

61

MPEG Video-Coding Procedure



62

MPEG- Audio Coding

- Sampling compatible to encoding of CD-DA and DAT:
  - Sampling rates:
    - 32 kHz, 44,1 kHz, 48 kHz
  - Sampling precision:
    - 16 bit/sample
- Audio channels:
  - Mono (single, 1 channel)
  - Stereo (2 channels)
    - dual channel mode (independent, e.g., bilingual)
    - optional: joint stereo (exploits redundancy and irrelevancy)

63

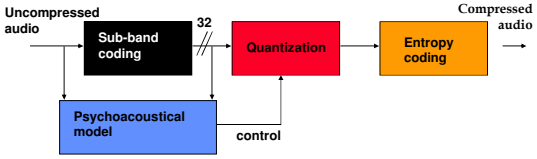
MPEG Audio

- Application Example: DAB Digital Audio Broadcasting
- uses MPEG layer 2 (compression also known as "MUSICAM" =
  - (Masking pattern adapted Universal Subband Integrated Coding And Multiplexing)
- delays, for VLSI implementation:
  - max. 30 ms encoding
  - max. 10 ms decoding
- SW codec delays vary for different layers, implementations, computers (rule-of-thumb may be 50/100/150 ms for layer 1/2/3, which makes MP3 rather inappropriate for real-time conversation)

64

MPEG-Audio Coding

- FFT applied to audio and spectrum is split into 32 non-interleaved sub-bands
  - for each sub-band, amplitude of audio signal is calculated
  - also, noise level is determined simultaneously with FFT, using a "psychoacoustic" model
    - Rough quantization at low noise level and fine one at high-level



65

MPEG- Audio Coding

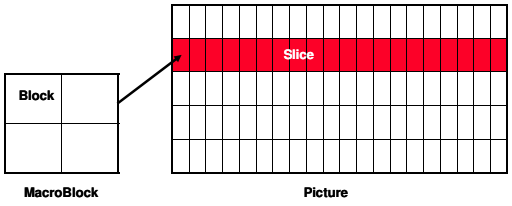
- Defines 3 layers of quality, with different complexity of encoder/ decoder
  - "higher layer" means "more complex" & "can handle lower layers"
- Data rates
  - 14 fixed data rates per layer, between 32 kbps-448 kbps
    - In steps of 16 kbit/s
  - Layer 1: max. 448 Kbit/s (ca. 1:4 compression, e.g. used as PASC in DCC "Digital Compact Cassette")
    - PASC= Precision Adaptive Subband Coding. It is a lossy compression method for audio
  - Layer 2: max. 384 Kbit/s (ca. 1:6-8, common, e.g. as MUSICAM in DAB)
  - Layer 3: max. 320 Kbit/s (ca. 1:10-12, the famous MP3)
  - Higher data rates are allowed for the modes:
    - "stereo"
    - "joint stereo"
    - "dual channel"

66

Codec	Bit Rate	Compression Ratio
Layer 1 (MP1)	384 kbps	3.7
Layer 2 (MP2)	192 kbps	7.3
Layer 3 (MP3)	128 kbps	11.0
AAC	96 kbps	14.7

MPEG Layers

- Each picture is divided to m horizontal slices
- Each slices contains n macroblocks
- Each macroblock contains of 16x16 pixels with the total of 256 pixels
- Each block composed of 8x8 pixels which is 64 total pixels



MPEG - Fellow up

- MPEG-2:
  - Higher data rates for high-quality audio/video
  - Multiple layers and profiles
  - Studio quality TV and CD quality audio channels. 4 to 6 Mbps typically.
- MPEG-3
  - Initially HDTV
  - MPEG-2 scaled up to subsume MPEG-3
- MPEG-4:
  - Initially, lower data rates for e.g. mobile communication
  - then: focus coding & additional functionalities based on image contents
  - Video conferencing at very low bit rates: 4.8 to 64 Kbps, with 10fps.
- MPEG-7 (EC = "experimental core" status):
  - Content description
  - Basis for search and retrieval
  - See section on databases
- MPEG-21 (upcoming):
  - Framework for multimedia business, delivery... what's missing?
  - maybe eCommerce focus --> e.g., security, watermarking?

MPEG 2

- From MPEG-1 to MPEG-2
  - Improvement in quality
    - from VCR to TV to HDTV
- No CD-ROM based constraints
  - higher data rates
    - MPEG-1: about 1.5 Mbit/s
    - MPEG-2: 2-100 Mbit/s
- Prominent role for digital TV in DVB (digital video broadcasting)
  - commercial MPEG-2 realizations available

MPEG 2

- an international standard (1994)
- CBR (constant bit rate) and VBR video (Variable bit rate)
- Picture quality higher than that of current NTSC, PAL and SECAM broadcast systems
- Compression to bit rates in the range of:
  - 60 Mbps for HDTV
  - 15 Mbps for NTSC, PAL and SECAM
  - 4-15 Mbps for TV signals conforming to CCIR 601
- MPEG-2 consists of five *profiles:* (*Simple* ➔ does not support B frames, *Main, Next, ..*) each having four *levels :*
  - High level Type 1: 1152 lpf, 1920 ppl, 60 fps -> 80 Mbps
  - High level Type 2: 1152 lpf, 1440 ppl, 60 fps -> 60 Mbps
  - Main: 576 lpf, 720 ppl, 30 fps -> 15 Mbps
  - Low: 288 lpf, 352 ppl, 30 fps -> 4 Mbps
    - Lpf = line per frame, ppl = pixels per line
- Video industry will use Main level/ Main profile
  - cable TV may use the Simple profile
- HDTV: High level / Main profile (USA), Next(Europe)

MPEG-2 Video Profiles and Levels

	Simple Profile	Main Profile	SNR Scalable Profile	Spatial Scalable Profile	High Profile
High Level 1920 pixels/line 1152 lines		80 Mbps			100 Mbps
High-1440 Level 1440 pixels/line 1152 lines		60 Mbps		60 Mbps	80 Mbps
Main Level 720 pixels/line 576 lines	15 Mbps	15 Mbps	15 Mbps		20 Mbps
Low Level 352 pixels/line 288 lines		4 Mbps	4 Mbps		
LAYERS and PROFILES	No B-frames Not Scalable	B-frames Not Scalable	B-frames SNR Scalable	B-frames SNR Scalable or Spatial Scalable	B-frames SNR Scalable or Spatial Scalable

Signal to Noise (SNR) scaling : noise introduced by quantization errors and block structures

MPEG 2 Audio

(two modest) extension to MPEG-1 audio:

1. "low sample rate extension" LSE:
   - ❖ 1/2 of all MPEG-1 rates: 16, 22.05, 24kHz
   - ❖ quantization down to 8 bits/sample
2. "multichannel extension": more channels, i.e. up to
   - ❖ 5 full bandwidth channels (surround system)
     - left and right front
     - center (in front)
     - left and right back
   - ❖ "multilingual extension": 7 more, i.e. up to 12 channels (multiple languages, commentary)

➢ Backward compatibility with MPEG-1 audio
   - ❖ Only three MPEG-2 audio codecs will not provide backward compatibility (in the range of 256- 448 kbps)

MPEG - Audio

➢ MPEG-1 (ISO/IEC 11172-3) provides
   - ❖ single-channel ('mono') and two-channel ('stereo' or 'dual mono') coding at 32, 44.1, and 48 kHz sampling rate.
   - ❖ The predefined bit rates range from 32 to 448 kbit/s for Layer I, from 32 to 384 kbit/s for Layer II, and from 32 to 320 kbit/s for Layer III.

➢ MPEG-2 BC (ISO/IEC 13818-3) provides
   - ❖ a backwards compatible multichannel extension to MPEG-1; up to 5 main channels plus a 'low frequent enhancement' (LFE) channel can be coded; the bit rate range is extended up to about 1 Mbit/s;
   - ❖ an extension of MPEG-1 towards lower sampling rates 16, 22.05, and 24 kHz for bitrates from 32 to 256 kbit/s (Layer I) and from 8 to 160 kbit/s (Layer II & Layer III).

➢ MPEG-2 AAC (ISO/IEC 13818-7) provides (AAC = Advanced Audio Coding)
   - ❖ a very high-quality audio coding standard for 1 to 48 channels at sampling rates of 8 to 96 kHz, with multichannel, multilingual, and multiprogram capabilities.
   - ❖ AAC works at bit rates from 8 kbit/s for a monophonic speech signal up to in excess of 160 kbit/s/channel for very-high-quality coding that permits multiple encode/decode cycles.
   - ❖ Three profiles of AAC provide varying levels of complexity and scalability.

➢ MPEG-4 (ISO/IEC 14496-3) will provide
   - ❖ Coding and composition of natural and synthetic audio objects at a very wide range of bit rates

MPEG-2 System Definition

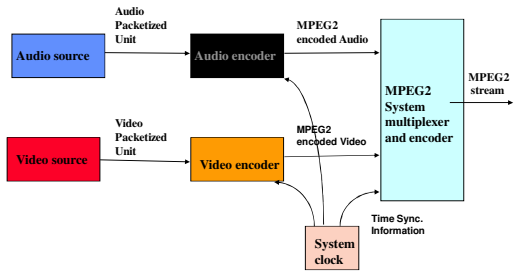
➢ Steps
   - ❖ audio and video combined to "Packetized Elementary Stream"
   - ❖ PES combined to "Program Stream" or "Transport Stream"

➢ Program Stream
   - ❖ Error-free environment
   - ❖ Packets of variable length
   - ❖ One single stream with one timing reference

➢ Transport Stream
   - ❖ Designed for "noisy" (lossy) media channels
   - ❖ Multiplex of various programs with one or more time bases
   - ❖ Packets of 188 bytes

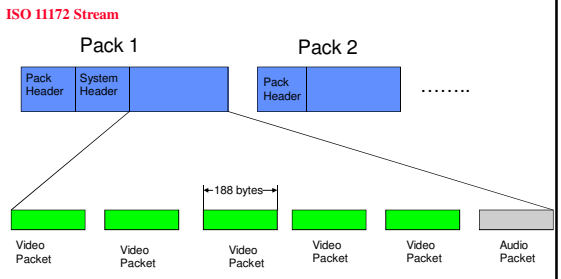
➢ Conversion between Program and Transport Streams possible

MPEG 2 Elementary Streams



76

MPEG 2 Streams



77

MPEG - Fellow up

- MPEG-2:
  - Higher data rates for high-quality audio/video
  - Multiple layers and profiles
  - Studio quality TV and CD quality audio channels. 4 to 6 Mbps typically.
- MPEG-3
  - Initially HDTV
  - MPEG-2 scaled up to subsume MPEG-3
- **MPEG-4:**
  - **Initially, lower data rates for e.g. mobile communication**
  - **then: focus coding & additional functionalities based on image contents**
  - **Video conferencing at very low bit rates: 4.8 to 64 Kbps, with 10fps.**
- **MPEG-7 (EC = "experimental core" status):**
  - **Content description**
  - **Basis for search and retrieval**
  - **See section on databases**
- **MPEG-21 (upcoming):**
  - **Framework for multimedia business, delivery... what's missing?**
  - **maybe eCommerce focus --> e.g., security, watermarking?**

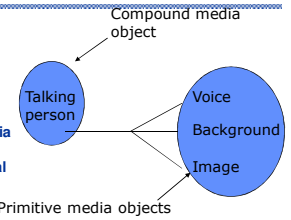
78

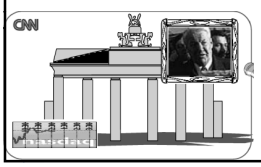
MPEG 4

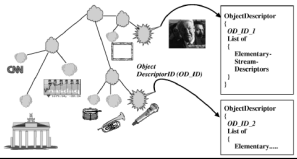
- **MPEG-4 (ISO 14496) originally:**
  - **Targeted at systems with very scarce resources**
  - **To support applications like**
    - **Mobile communication**
    - **Videophone and E-mail**
  - **Max. data rates and dimensions (roughly):**
    - **VLBV "Very Low Bit-rate Video"**
      - Between 4800 and 64000 bits/s
    - **176 columns x 144 lines x 10 frames/s**
      - Largely covered by H.263 (QCIF)
- **therefore re-orientation:**
  - **Goal to provide enhanced functionality**
  - **to allow for analysis and manipulation of image contents**

MPEG 4 - Technical information

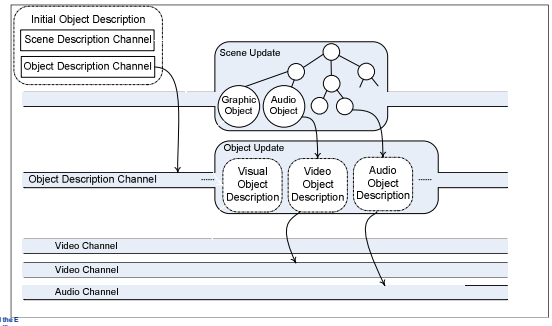
- **Objects are organized in a hierarchal fashion.**
- **Each object has its own description element.**
  - **Allows handling of the object**
- **One or more primitive media objects can be combined.**
- **Techniques from the Virtual Reality model language.**





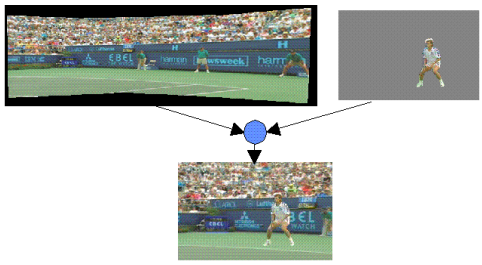


Content Access Management



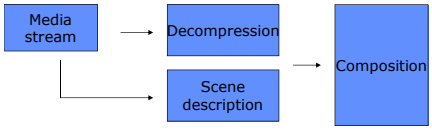
Video objects

- Divide video components
  - Person and background
- Camera position information



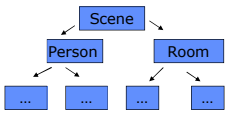
MPEG 4 -- Media streams

- One or more media streams
- Descriptors for the objects and the stream



Scene description

- Grouping of the objects
  - Directed acyclic graph
- Positioning the objects
  - Special attributes



MPEG7

- Increasing availability of Multimedia content
- Increasing creation of Multimedia content
- Increasing use of Multimedia content by machines
- The need for searching, categorizing, describing, managing and filtering

→ Great need for Standard Description

- MPEG-7 proposing such a standard
- MPEG-7 does not deal with implementation
  - (Great for Master Thesis)

MPEG 21

- MPEG-21 Multimedia Framework
  - The vision for MPEG-21 is:

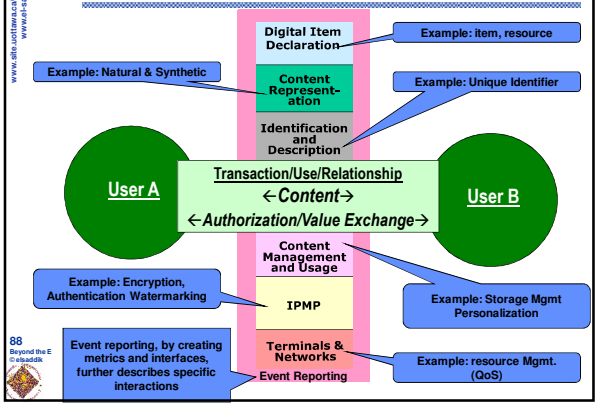
to define a multimedia framework *to enable transparent and augmented use of multimedia resources across a wide range of networks and devices* used by different communities

MPEG 21

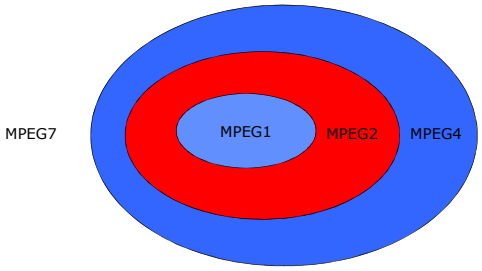
Seven Architectural 'Elements' in the Multimedia Framework:

1. Digital Item Declaration
2. Digital Items Representation
3. Digital Item Identification and Description
4. Content Management and Usage
5. Intellectual Property Management and Protection
6. Terminals and Networks
7. Event Reporting

MPEG 21



MPEG relations

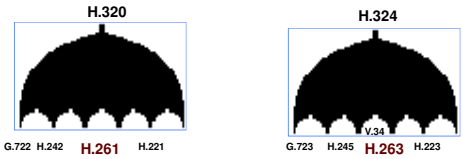


Source Image Formats

Format	Pixels	H261 Encoder/Decoder	H263 Encoder/Decoder
SQCIF Sub Quarter Common Intermediate Format	128 x 96	optional	required
QCIF Quarter Common Intermediate Format	176 x 144	required	required
CIF Common Intermediate Format	352 x 288	optional	optional
4QCIF 4 Times Quarter Common Intermediate Format	704 x 576	not defined	optional
16QCIF 16 Times Quarter Common Intermediate Format	1408 x 1152	not defined	optional

ITU-Standards for Narrow-Band Videoconferencing

- H.320:
  - Standard for videoconferencing over ISDN lines
- H.324:
  - Standard for videoconferencing over POTS (Plain Old Telephone Service)
- H.32x's umbrella specification structure:



H.261 and related ITU Standards

- Video codec for audiovisual services at p x 64kbit/s
  - ("p-times-sixtyfour", where p means "multiples-of"):
  - ITU- CCITT standard from 1990
    - ITU = International Telecommunication Union
    - CCIT = Consultative Committee for International Telegraph and Telephone
  - For ISDN
  - With p=1,..., 30
- Technical issues:
  - Real-time encoding/decoding
  - Max. signal delay of 150ms
  - Constant data rate
  - Implementation in hardware (main goal) and software

H.261 – Resolution Format

- Unlike JPEG, H.261 defines a very precise image format
  - Image components:
    - Luminance signal (Y)
    - Two color difference signals (Cb,Cr)
  - Subsampling according to CCIR 601 (4:1:1)
    - ITU-R 601: (formerly CCIR) designates a "raw" digital video format with 704 x 480 pixels
      - CCIR = International Radio Consultative Committee

Two resolution formats are specified:

- Optional
  - Common Intermediate format (CIF) resolution
    - Y: 352 x 288 pixel
    - At 29.97 frames/s app. 36.46 Mbps (uncompressed) i.e. ~ 570 * 64kbps
- Mandatory
  - Quarter Common Intermediate Format (QCIF) resolution (has half of CIF resolution)
    - Y: 176 x 144 pixel
    - At 29.97 frames/s app. 9.115 Mbps (uncompressed)
- all H.261 implementations must be able to encode and decode QCIF ; CIF is optional

H.261 (p x 64) Video Compression

- DCT-based compression algorithm, like JPEG, with
  - differential PCM (DPCM) with motion estimation for interframe coding and
  - variable word-length entropy coding (such as Huffman)
- very high-compression ratios for full-color, real-time motion video transmission
- combines intraframe and interframe coding
- optimized for applications such as
  - video-conferencing, which are not motion-intensive
- limited motion search and estimation strategies
- compression ratios from 24:1 to 200:1
- covers the entire ISDN channel capacity (p x 64 kbps, p=1,2,...,30)
  - for p=1 or 2: videophone, desk-top video-conferencing applications
  - for p=6 or higher, more complex pictures are transmitted. Good for group video-conferencing

H.261

- Intraframe coding takes no advantage of redundancy between frames.
  - Intraframe coding: yields "reference frame" f0
  - each 8x8 block is transformed by DCT
  - DCT with same quantization factor for all AC values
  - this factor may be adjusted by loopback filter
  - intraframes rare (bandwidth!, main application videophone)
- Interframe coding (corresponds to P frames of MPEG)
  → Motion estimation
  - interframes: f1,f2,f3,... relative to f0 (differential encoding)
  - Search of similar macroblock (16x16) in previous image
  - Position of this macroblock defines motion vector
  - Search range is up to the implementation:
    - max. ± 15 pixel
    - but: motion vector may also always be 0 ("bad" software encoder)
      - e.g. H.261 also allows simple implementation, considering only the differences between macroblocks located in the same position, thus a zero motion vector

Main Differences between H.261 and H.263

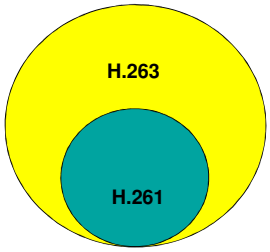
- Extension to H.261
- max. bitrate: H.263 approx. 2.5 x H.261; lowest bitrates suitable f. modem

Main Differences between H.261 and H.263

- Base Level Differences (always ON)
  - No filter for HF (high frequency) noise in feedback loop
  - Motion vectors produced with 1/2-pixel resolution
  - Picture format for sub-QCIF (128x96)
  - Huffman tables designed specifically for low bit rate.
  - JPEG is the still picture mode
- Optional Level Differences (Negotiated)
  - Unlimited search space for motion vector
    → fast encoder can do better
  - Syntax-based Arithmetic coding
  - Advanced prediction mode
  - PB-frames (2 combined pictures: 1 B- & 1 P-Frame)

Main Differences between H.261 and H.263

- N.B. H.261 is fully contained within H.263



97

Conclusion

JPEG

- Very general format with good compression ratio
- SW and HW for baseline mode available

H.261/ H.263

- Established standard by telecom world
- Preferable hardware realization

MPEG-1, MPEG-2, MPEG-4, MPEG-7

- MPEG-2 with data rates between 2 and 100 Mbps
- MPEG-4, MPEG-7: objects coding, content descr.

Proprietary Systems: Quicktime, DVI, CD-I,...

- Product that use of other standards
- Migration to use the standards

98

Encoding Rates of Various Standards

Standard	Data Rate	Compression
JPEG (for video)	10-20 Mbps	7-27 times
MPEG-1	1.2-2.0 Mbps	100 times
H.261	64kbps-2Mbps	24 times
DVI	1.2-1.5 Mbps	160 times
CD-I	1.2-1.5 Mbps	100 times
MPEG-2	4-60 Mbps	30-100 times
CCIR 723	32-45 Mbps	3-5 times
CCIR 601/D-1	140-270 Mbps	Reference
PictureTel SG3	0.1-1.5 Mbps	100 times
Software compression (small window)	~2 Mbps	6 times

NB. For JPEG , it was assumed 640 x 480 x 24-bit colour, 15 fps

99