# MPEG (Image) Compression

SSIP 2006
László Czúni
Pannon University, Veszprém

## Aims of Image Coding

- Protect visual information:
  - Cryptography
  - Steganography
  - Watermarks
- Error resilient coding
- Content description
- Interactive applications
- Compress data size…

2

## Image Redundancy

Redundancy: exceeding the necessary data to represent a *proper* image

- Spatial
  - Neighboring pixels are similar
- Temporal
  - Succeeding image frames are similar
- Psychovisual
  - There can be information on an image what we don't perceive (depends on the viewing conditions)
- Code
  - Data representing the image can be simplified (see information/communication theory)

3

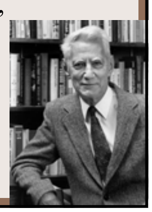
## Lossy vs. Lossless

- Lossless compression:
  - Does not utilize psychovisual redundancy
  - Typical compression rate for images: 1.5X - 3X
- Lossy compression:
  - It considers the psychovisual properties of the human visual system (HVS)
    - Frequency sensitivity
    - Adaptation (spatial or temporal)
    - Non-linearity
    - Frequency masking
  - Typical compression rate for stills: 10X-200X

4

## Claude Elwood Shannon

- April 30, 1916 – February 24, 2001
- A handsome American electrical engineer and mathematician,
- "the father of information theory",
- "Source Coding Theorem", 1948
- the founder of practical digital circuit design theory



## Noiseless Coding Theorem
(informally)

- Coding blocks, of size N, of source symbols into binary codewords
- S is an ergodic source with entropy H(S) and with ABC of size n.
- For any δ>0 there is N (large enough), that

$$H(S) \le \overline{L} < H(S) + \delta$$

where $\overline{L}$ is the average length of codewords

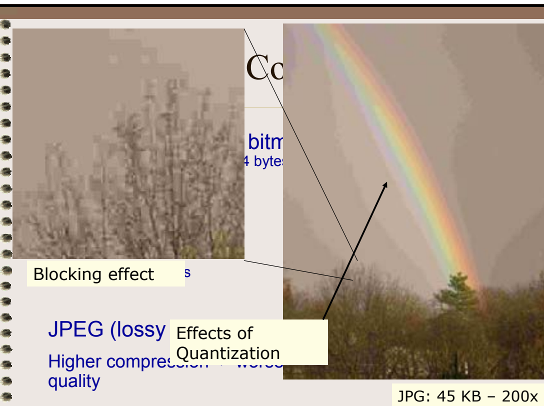
6

# Frequently Used Lossless Compression Methods

- Shannon-Fano
- Huffman coding
- Run-length coding
- Arithmetic coding
- LZW (TIFF, GIF)
- Lossless DPCM
- Etc.

7

Co

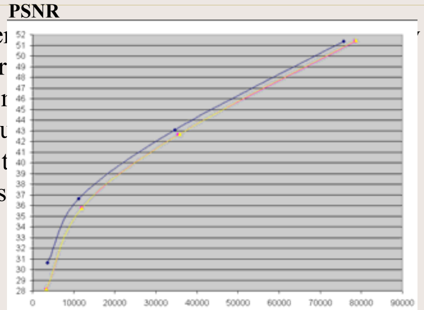



bitm
4 byte
s
JPEG (lossy
Higher compres... worse
quality

8

# Rate distortion (RD) Theorem, RD Curves

Deter
(or
cor
sou
at t
dis




9

# How to lose information? We should consider the HVS...

Both images contain the same amount of noise but they look different (have the same PSNR compared to the original)

10

# Prediction Methods

- DPCM (Differential Pulse Code Modulation)
  - The most probable values are expected, the prediction error is transmitted
- Spatial (from pixel to pixel)
- Temporal
  - Motion estimation and motion coding is necessary
- In lossy mode the prediction error (residual) is transformed, quantized, entropy coded then transmitted.

11

# Transform Methods

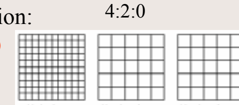
General model of transform coding:

1. Color transform (typically from RGB to YCrCb)
2. Undersampling of color channels (e.g. 4:2:2, 4:2:0)
3. Cutting into blocks (block based coding)
4. Transform (DCT, HD)
5. Quantization of transform coefficients
6. Run-length coding in zig-zag order
7. Entropy coding (Huffman, Arithmetic)

Popular transforms for compression:

- Discrete Cosine Transform (DCT)
- Hadamard Transform
- Wavelet Transform

4:2:0

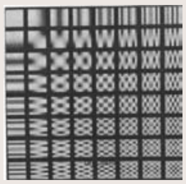


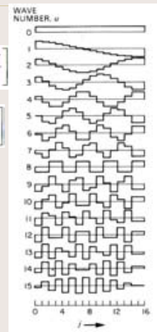
## Discrete Cosine Transform

$$f(u,v)=\frac{4C(u)C(v)}{N^2}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1}F(j,k)\cos\left[\frac{(2j+1)u\pi}{2N}\right]\cos\left[\frac{(2k+1)v\pi}{2N}\right]$$

$$F(j,k)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}C(u)C(v)f(u,v)\cos\left[\frac{(2j+1)u\pi}{2N}\right]\cos\left[\frac{(2k+1)v\pi}{2N}\right]$$

$$C(w)=\begin{cases}\frac{1}{\sqrt{2}} & w=0\\ 1 & w=1,2,\ldots,n-1\end{cases}$$






## Some Properties of DCT

- Invertible
- Linear
- Unitary
  - $\langle \boldsymbol{g} \mid \boldsymbol{h} \rangle = \langle U\boldsymbol{g} \mid U\boldsymbol{h} \rangle, \forall \boldsymbol{g}, \boldsymbol{h} \in V$
    $\|\boldsymbol{g}\|_2 = \langle \boldsymbol{g} \mid \boldsymbol{g} \rangle^{1/2} = \langle U\boldsymbol{g} \mid U\boldsymbol{g} \rangle^{1/2}$
  - $UU^{*T}=I$
  - rows and columns form orthonormal basis
- Separable: the 2D DCT transform can be computed by row and column operations

14

## Quantization of Transform Coefficients

- Transform coefficients are divided by the elements of the quantization matrix.

Intensity

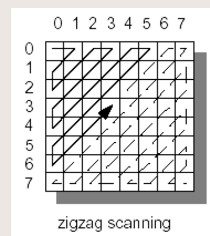
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 16 | 11 | 10 | 16 | 24 | 40 | 51 | 61 |
| 12 | 12 | 14 | 19 | 26 | 58 | 60 | 55 |
| 14 | 13 | 16 | 24 | 40 | 57 | 69 | 56 |
| 14 | 17 | 22 | 29 | 51 | 87 | 80 | 62 |
| 18 | 22 | 37 | 56 | 68 | 109 | 103 | 77 |
| 24 | 35 | 55 | 64 | 81 | 104 | 113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

Crominance

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 17 | 18 | 24 | 47 | 99 | 99 | 99 | 99 |
| 18 | 21 | 26 | 66 | 99 | 99 | 99 | 99 |
| 24 | 26 | 56 | 99 | 99 | 99 | 99 | 99 |
| 47 | 66 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |

## Processing of Transform Coefficients

- Zig-zag scanning
- Run-length encoding
- Entropy coding:
  - VLC
  - Arithmetic coding




16

## Hybrid Methods

- Both transform coding and prediction is applied:
  - MPEG-1,2: temporal prediction, DCT transform
  - MPEG-4: spatial and temporal prediction, DCT and Hadamard transform

17

## Standardization Organizations

- **ITU** (International Telecommunication Union)
  - ITU-T (ITU – Telecommunication)
    - VCEG (Video Coding Experts Group)
- **ISO** (International Standardization Organization)
  - JPEG (Joint Photographic Experts Group)
  - MPEG (Motion Picture Experts Group)

18

## History

1950-60 – DPCM (Differential Pulse Code Modulation)
1974 – Block-based DCT (Discrete Cosine Transform)
1985 – Hibrid MC/DCT (Motion-compensated DCT)
1986 – JPEG is founded, target: transmit a 720x576x16bit image in 10sec over a 64kbit/sec (ISDN) channel
1990 – ISO – JPEG standard
1990 – ITU-T: H.261 video telephony over ISDN
1993 – ISO: MPEG-1 (VideoCD, based on JPEG and H.261)

19

## History

- 1995 – ISO: MPEG-2
  - To compress PAL/NTSC signals (DVD, DVB-S, DVB-T)
- 1995 – ITU-T: H.263
- 2000 – ISO – MPEG-4 videókódolási szabvány
  - Popular Implementation DivX, XviD
- 2002 – ISO – JPEG 2000
  - DWT, incompatible with JPEG

20

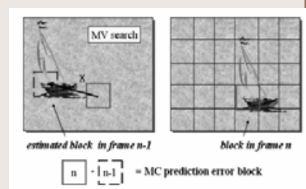
## History

- 2002 – ITU & ISO formed JVT (Joint Video Team) – H.264 / MPEG-4 part 10 AVC (Advanced Video Codec)
  - 4x4 MC/Integer Transformation
  - Tree structured motion estimation
  - ¼ pixel precision motion estimation
  - CAVLC (Context-Adaptive Variable Length Coding)
  - CABAC (Contect-Adaptive Binary Arithmetic Coding)
  - Full implementation Joint Model (JM) Reference Software

21

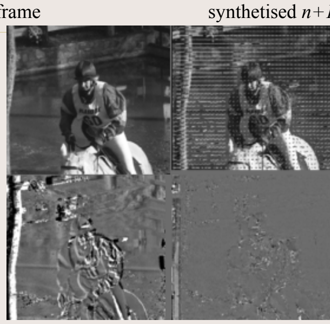
## Motion-based Coding

- Temporal correlation is reduced by motion estimation and motion vector coding
- Motion vectors are estimated and coded into the bitstream
- Motion vectors are used to synthetise the neighboring frames
- Residual image is transform coded




## Motion-based Coding

$n^{th}$ frame synthetised $n+1^{th}$ frame



difference between frames
$n$ and $n+1$ $n$ and predicted $n+1$

23

## MPEG

- Established in May 1988.
- Combine video and audio applications.
- Rather define decoding than coding.
- Motion oriented coding.

24

# MPEG-1

- Issued in 1992.
- Developed on the basis of JPEG and H.261.
- Achieved the goal of storing moving pictures and audio on a CD with quality comparable to that of VHS.
- What matters in a television signal is not the number of lines or the number of fields per second but the bandwidth of the signal in the analogue domain and the number of pixels in the digital domain. Result: normative definition of the constrained parameter set (CPS) with no references to television standards.



# MPEG-1

- Important features: random access of video, reverse playback, editability of the compressed bit stream, audio-video synchronization.
- Motion compensation of macroblocks (MB) (16x16 pixels, 4 luminance blocks and 2 chrominance blocks).
- Three types of MB:
  - Skipped MB: predicted from the previous frame with no motion vector.
  - Inter MB: motion compensated from the previous frame.
  - Intra MB: no prediction from the previous frame.
- Motion compensation error is coded with DCT.



# MPEG-1

- Types of frames
  - Intraframe (I): no reference to other frames.
  - Predictive (P): from the nearest previously coded I or P frame with motion compensation.
  - Bi-directional predicted or interpolated frames (B): never used as reference for other frames.
- Rate control:
  - Defining quantization levels for MBs.
  - Defining the ratio for I,P and B frames.
- Interlaced mode is not directly supported: only the subsampled top field is coded, then at the decoder the even field is predicted and horizontally interpolated from the decoded odd field.



# MPEG-2

- Starting of MPEG2 project in July 1990 in Porto, Portugal mainly for solving the problems of novel television coding techniques.
- Specification of international standard in 1994.
- Provides all components necessary for interactive television, e.g. client-server protocol to allow a user to interact with the content of the server, encryption support, etc.
- Applications: DVD, Advanced Audio Coding (AAC) over the Internet, etc.
- Introduced motion compensation of *frames* and *fields*.
- Scales and levels. Three types of scalability: spatial, temporal, SNR.
- Error resilient coding by data partitioning (not part of the standard).



# MPEG-2 Levels

**Table 1. Upper bound of parameters at each level of a profile.**

| Level | Parameters |
|---|---|
| HIGH | 1920 samples/line<br>1152 lines/frame<br>60 frames/s<br>80 Mbit/s |
| HIGH 1440 | 1440 samples/line<br>1152 lines/frame<br>60 frames/s<br>60 Mbit/s |
| MAIN | 720 samples/line<br>576 lines/frame<br>30 frames/s<br>15 Mbit/s |
| LOW | 352 samples/line<br>288 lines/frame<br>30 frames/s<br>4 Mbit/s |

# MPEG-2: Profiles

**Table 2. Algorithms and functionalities supported with each profile.**

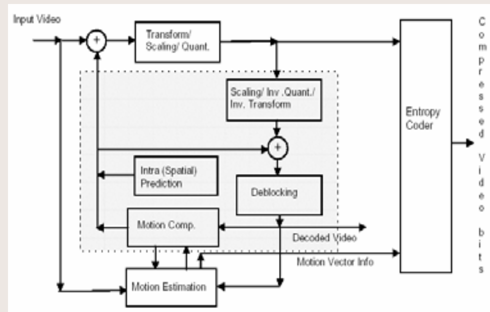
| Profile | Algorithms |
|---|---|
| HIGH | Supports all functionality provided by the Spatial Scalable Profile plus the provision to support:<br>• 3 layers with the SNR and Spatial scalable coding modes<br>• 4:2:2 YUV-representation for improved quality requirements |
| SPATIAL Scalable | Supports all functionality provided by the SNR Scalable Profile plus an algorithm for:<br>• spatial scalable coding (2 layers allowed):<br>• 4:0:0 YUV-representation |
| SNR Scalable | Supports all functionality provided by the MAIN Profile plus an algorithm for:<br>• SNR scalable coding (2 layers allowed)<br>• 4:2:0 YUV-representation |
| MAIN | Nonscalable coding algorithm supporting functionality for:<br>• coding interlaced video<br>• random access<br>• B-picture prediction modes<br>• 4:2:0 YUV-representation |
| SIMPLE | Includes all functionality provided by the MAIN Profile but:<br>• does not support B-picture prediction modes<br>• 4:2:0 YUV-representation |

# MPEG-4

- Started in 1994, international standard in 1999 (some work, on extensions, is still in progress).
- Deals with „audio-visual objects" (AVOs) rather than „bit streams" produced by encoding audio & video.
- More interaction with content.
- BIFS - Binary Format for Scene Description: composition technology.
- Contents intellectual property rights management infrastructure:
  - Content identification, Automatic monitoring & tracking of AVOs, Tracking of AVO modification history, etc.
- DMIF - Delivery Multimedia Integration Format: Hide the delivery technology details from the DMIF user and ensure the establishment of end to end connections.
- Effects:
  - High quality for audio & video over very low bit rate channels (like 28.8 kbit/s)
  - Real-time interpersonal communication.
  - High level personalization due to AVOs.
  - a lot more

31

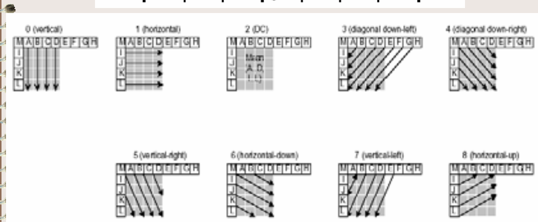
# MPEG-4 part 10 / AVC (Advanced Video Codec) / H.264



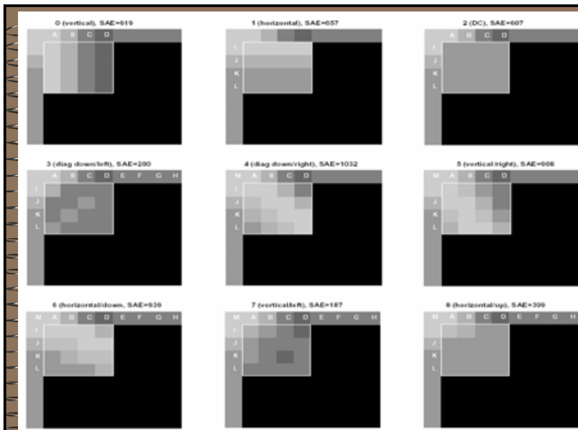

32

# H.264 – 4x4 intra prediction

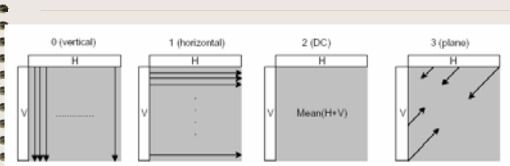
| M | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| I | a | b | c | d | | | | |
| J | e | f | g | h | | | | |
| K | i | j | k | l | | | | |
| L | m | n | o | p | | | | |

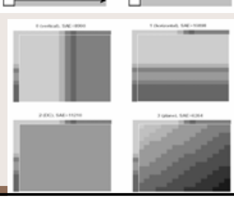



33



# H.264 – 16x16 intra prediction

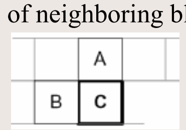


35

# H.264 – Intra prediction

- For color channels only the 16x16 modes can be applied but for 8x8 blocks
- Prediction modes of neighboring blocks can be estimated

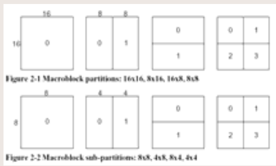


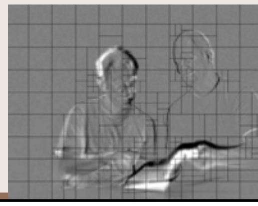

- If blocks A and B have the same mode then only 1 bit is enough to code the mode of block C

36

## H.264 – Inter prediction

- Tree-structured motion compensation: 16x16 macroblock are further partitioned
- Motion information of 4x4 blocks can be groupped into larger areas



## H.264 – Inter prediction

I blocks: intra coded

P blocks:

- one motion vector to a preceding or to a subsequent frame
- can be used on P or B frames

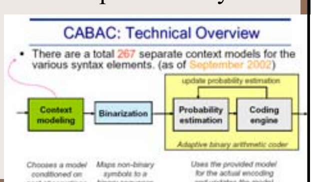
B blocks:

- two motion vectors to preceding or subsequent frames
- the two prediction results are weighted
- can be used on B frames

38

## Context adaptive entropy coding

- CAVLC (Context-based Adaptive Variable Length Coding)
  - Huffman-based
  - Simple implementation
- CABAC (Context-based Adaptive Binary Arithmetic Coding)
  - More complex
  - More efficient




## H.264 Frame Types

- I frames:
  - Contains only I blocks
  - Special frame: IDR-frame (Instantaneous Decoder Refresh), to reset the decoder
  - Reference for P and B blocks
- P frames:
  - Contains P and I blocks
  - Reference for P and B blocks
- B frames:
  - Contains P and B blocks
  - Not a reference type

40

## H.264 – Integrated „In-loop" filter

- MPEG-4 Part 2: optional post filtering to smooth block boundaries to decrease blocking effects
- H.264: smoothing in integrated form: prediction of P- and B-frames are based on filtered images
- Optional
- Increases complexity and image quality

41

## H.264 – Integrated „In-loop" filter

Figure 6 Reconstructed, QP=36 (no filter)

Figure 7 Reconstructed, QP=36 (with filter

## H.264 – Integrated „In-loop" filter

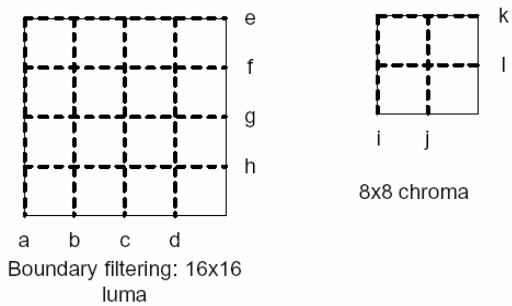



**Figure 1 Edge filtering order in a macroblock**

## H.264 – Integrated „In-loop" filter

| | |
|---|---|
| p or q is intra coded **and** boundary is a macroblock boundary | Bs=4 (strongest filtering) |
| p or q is intra coded and boundary is **not** a macroblock boundary | Bs=3 |
| neither p or q is intra coded; p or q contain coded coefficients | Bs=2 |
| neither p or q is intra coded; neither p or q contain coded coefficients; p and q have different reference frames **or** a different number of reference frames **or** different motion vector values | Bs=1 |
| neither p or q is intra coded; neither p or q contain coded coefficients; p and q have same reference frame and identical motion vectors | Bs=0 (no filtering) |

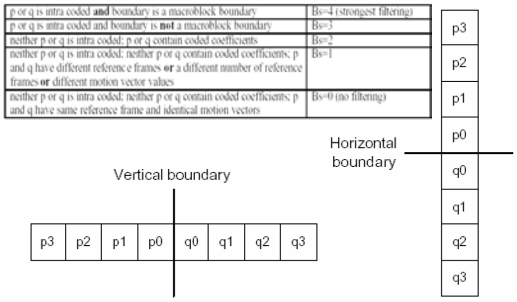



**Figure 2 Pixels adjacent to vertical and horizontal boundaries**

## MPEG-4 FRExt

- Fidelity Range Extension
- Problems with H.264
  - 8 bit 4:2:0 format (YCrCb)
  - RGB-YCrCb-RGB conversion causes error

$$Y = K_R * R + (1 - K_R - K_B) * G + K_B * B, \quad Cb = \frac{1}{2}\left(\frac{B-Y}{1-K_B}\right), \quad Cr = \frac{1}{2}\left(\frac{R-Y}{1-K_R}\right).$$

45

## MPEG-4 FRExt

- YCgCo representation
  - RGB -> YCgCo

$$Co = R - B; \quad t = B + (Co >> 1); \quad Cg = G - t; \quad Y = t + (Cg >> 1);$$

  - YCgCo -> RGB

$$t = Y - (Cg >> 1); \quad G = t + Cg; \quad B = t - (Co >> 1); \quad R = B + Co.$$

- „Residual Color Transform" (4:4:4 format only)
  - Prediction is in RGB format
  - Residual is color transformed

46

## MPEG-4 FRExt

- 8x8 integer transform
- 4x4 Hadamard transform
- 4:2:2 and 4:4:4 formats
- 9, 10, 11, 12 bit-depth
- 8x8-as intra prediction
- B-frames can be reference frames
- Lossless modes
  - Intra PCM blocks
  - Entropy coded macroblocks
- Macroblock based quantization
- Dinamic quantization tables

47

## MPEG-4 FRExt

New profiles

| Coding Tools | High | High 10 | High 4:2:2 | High 4:4:4 |
|---|---|---|---|---|
| Main Profile Tools | X | X | X | X |
| 4:2:0 Chroma Format | X | X | X | X |
| 8 Bit Sample Bit Depth | X | X | X | X |
| 8x8 vs. 4x4 Transform Adaptivity | X | X | X | X |
| Quantization Scaling Matrices | X | X | X | X |
| Separate Cb and Cr QP control | X | X | X | X |
| Monochrome video format | X | X | X | X |
| 9 and 10 Bit Sample Bit Depth | | X | X | X |
| 4:2:2 Chroma Format | | | X | X |
| 11 and 12 Bit Sample Bit Depth | | | | X |
| 4:4:4 Chroma Format | | | | X |
| Residual Color Transform | | | | X |
| Predictive Lossless Coding | | | | X |

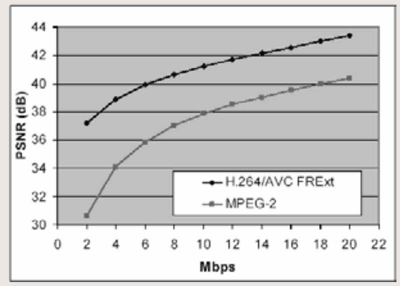
48

# MPEG-4 FRExt

New levels

| Level Number | Typical Picture Size | Typical frame rate | Maximum compressed bit rate (for VCL) in Non-FRExt profiles | Maximum number of reference frames for typical picture size |
|---|---|---|---|---|
| 1 | QCIF | 15 | 64 kbps | 4 |
| 1b | QCIF | 15 | 128 kbps | 4 |
| 1.1 | CIF or QCIF | 7.5 (CIF) / 30 (QCIF) | 192 kbps | 2 (CIF) / 9 (QCIF) |
| 1.2 | CIF | 15 | 384 kbps | 6 |
| 1.3 | CIF | 30 | 768 kbps | 6 |
| 2 | CIF | 30 | 2 Mbps | 6 |
| 2.1 | HHR (480i or 576i) | 30 / 25 | 4 Mbps | 6 |
| 2.2 | SD | 15 | 4 Mbps | 5 |
| 3 | SD | 30 / 25 | 10 Mbps | 5 |
| 3.1 | 1280x720p | 30 | 14 Mbps | 5 |
| 3.2 | 1280x720p | 60 | 20 Mbps | 4 |
| 4 | HD Formats (720p or 1080i) | 60p / 30i | 20 Mbps | 4 |
| 4.1 | HD Formats (720p or 1080i) | 60p / 30i | 50 Mbps | 4 |
| 4.2 | 1920x1080p | 60p | 50 Mbps | 4 |
| 5 | 2kx1k | 72 | 135 Mbps | 5 |
| 5.1 | 2kx1k or 4kx2k | 120 / 30 | 240 Mbps | 5 |

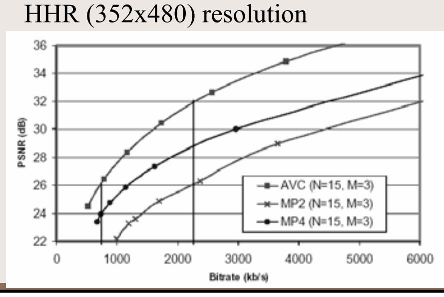
# MPEG-4 FRExt




50

# Compression Efficiency

- Mobil and Calendar video sequence at HHR (352x480) resolution




51

# Video Quality

- Objective vs. subjective
- Difficult to evalute the subjective quality
- Video Quality Experts Group
  - www.vqeg.org
  - PSNR is still a good tool

ITU measurement standard

- ITU-R BT.500
- Double Stimulus Continuous Quality Scale (DSCQS)
- (bad, poor, fair, good, excellent)

52

- Thank you for your attention.

53

ITU-R BT.500 Recommendation and SAMVIQ (ITU-R BT.700)

| | Methodology | | | | | |
|---|---|---|---|---|---|---|
| Parameter | DSIS | DSCQS | SS | SSCQE | SDSCE | SAMVIQ |
| Explicit reference | yes | no[1] | no | no | yes | yes (uncompressed) |
| Hidden reference | no | yes[1] | no | no | no | yes |
| High anchor | no | yes[1] | no | no | no | no (hidden ref) |
| Low anchor | no | yes[1] | no | no | no | yes |
| Scale | 5 grades | bad -excellent (continuous quality scale) | 5 grades | bad - excellent (continuous quality scale) | bad - excellent (continuous quality scale) | bad - excellent (continuous quality scale) |
| Sequence length | 10 s | 10 s | 10 s | ≥5 min | 10 s | 10 s |
| Picture format | all | all | all | all | all | all |
| | | | | | | |
| 2 simultaneous stimulus | no | no | no | no | yes | no |
| Presentation of test material | variant I: once variant II: twice in succession | twice in succession (double stimulus) | once | once | once | as often as user likes (multi stimuli) |
| Videos per trial | 2 | 2 | 1 | 1 | 2 | max 10[2] |
| | | | | | | |
| Voting | only test sequence | only test sequence and reference | only test sequence | only test sequence | difference between the test sequence and ref. simultaneously shown | test sequence and reference |
| Possibility to change the vote before proceeding | no | no | no | no | no | yes |
| Continuous quality evaluation | no | no | no | yes (moving slider in a continuous way) | yes (moving slider in a continuous way) | no |
| Minimum accepted votes | 15 | 15 | 15 | 15 | 15 | 15 |
| Rejection criteria | | yes, but not stable | | | | yes |
| | | | | | | |
| Observers per display | 1 to many | 1 to many | 1 to many | 1 to many | 1 to many | 1 |
| Display | all (mainly TV) | all (mainly TV, DLP) | all (mainly TV) | all (mainly TV) | all (mainly TV) | all (mainly PC, PDA) |
| Quality results | relative, depending on reference quality | relative, depending on compared sequence | relative | relative | relative, depending on ref quality | absolute measure of video quality |
| | | | | | | |
| Standard | ITU-R BT.500-11 | ITU-R BT.500-11 | ITU-R BT.500-11 | ITU-R BT.500-11 | ITU-R BT.500-11 | ITU-R 6Q6 WP 6Q (new ITU-R BT.700) |

[1] not mandatory (could be any test sequence), [2] different bit rates in one trial to avoid contextual effects

DSIS: Double Stimulus Impairment Scale (= Degradation Category Rating DCR in ITU-T P.910)
DSCQS: Double Stimulus Continuous Quality Scale
SS: Single Stimulus (= Absolute Category Rating ACR in ITU-T P.910)
SSCQE: Single Stimulus Continuous Quality Evaluation
SDSCE: Simultaneous Double Stimulus for Continuous Evaluation
SAMVIQ: Subjective Assessment Methodology for Video Quality

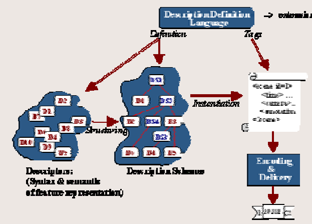
# MPEG-7

- Multimedia content description standard (ISO/IEC 2004)
- It uses XML to store metadata, and can be attached to timecode in order to tag particular events, or synchronise lyrics to a song, for example.
- It was designed to standardise:
  - a set of description schemes and descriptors
  - a language to specify these schemes, called the Description Definition Language (DDL)
  - a scheme for coding the description
- The combination of MPEG-4 and MPEG-7 has been referred to as MPEG-47.

55

# MPEG-7

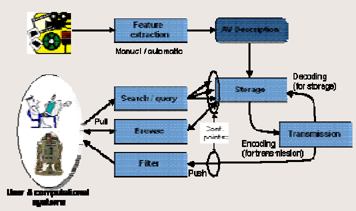
Main Elements




56

# MPEG-7

Possible Applications




57

# MPEG-21

- The MPEG-21 standard, from the Moving Picture Experts Group aims at defining an open framework for multimedia applications. ISO 21000.
- Specifically, MPEG-21 defines a "Rights Expression Language" standard as means of sharing digital rights/permissions/restrictions for digital content from content creator to content consumer. As an XML-based standard, MPEG-21 is designed to communicate machine-readable license information and do so in an "ubiquitous, unambiguous and secure" manner.

58

# MP3

- MPEG-1 Layer 3
- Lossy compression format
- Standardized by ISO/IEC in 1991
- The MP3 format uses a hybrid transformation to transform a time domain signal into a frequency domain signal:
  - 32-band polyphase quadrature filter
  - 36 or 12 tap MDCT; size can be selected independent for sub-band 0...1 and 2...31
  - Aliasing reduction postprocessing
- In terms of the MPEG specifications, AAC (Advanced audio coding) from MPEG-4 is to be the successor of the MP3 format

59

# Windows Media Video

- WMV version 7 (WMV1) was built upon Microsoft's own non-standard version of MPEG-4 Part 2.
- WMV version 9 standardized as an independent SMPTE standard (421M, also known as VC-1)
- There are currently (April 2006) 16 companies in the VC-1 patent pool.
- Microsoft is also one of the members of the MPEG-4 AVC/H.264 patent pool.

60

## DCT factorization (in H.264)

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos\frac{\pi(2m+1)p}{2M} \cos\frac{\pi(2n+1)q}{2N}, \quad 0 \le p \le M, \; 0 \le q \le N \qquad a = \frac{1}{2}$$

$$b = \sqrt{\frac{1}{2}} \cos\left(\frac{\pi}{8}\right)$$

$$\alpha_p = \begin{cases} 1/\sqrt{M}, & p = 0 \\ \sqrt{2/M}, & 1 \le p \le M-1 \end{cases} \qquad \alpha_q = \begin{cases} 1/\sqrt{N}, & q = 0 \\ \sqrt{2/N}, & 1 \le q \le N-1 \end{cases} \qquad c = \sqrt{\frac{1}{2}} \cos\left(\frac{3\pi}{8}\right)$$

$$\mathbf{Y} = \mathbf{AXA}^T = \begin{bmatrix} a & a & a & a \\ b & c & -c & -b \\ a & -a & -a & a \\ c & -b & b & -c \end{bmatrix} \quad \mathbf{X} \quad \begin{bmatrix} a & b & a & c \\ a & c & -a & -b \\ a & -c & -a & b \\ a & -b & a & -c \end{bmatrix}$$

d = c/b= 0.414

$$\mathbf{Y} = \left(\mathbf{CXC}^T\right) \otimes \mathbf{E} = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & d & -d & -1 \\ 1 & -1 & -1 & 1 \\ d & -1 & 1 & -d \end{bmatrix} \quad \mathbf{X} \quad \begin{bmatrix} 1 & 1 & 1 & d \\ 1 & d & -1 & -1 \\ 1 & -d & -1 & 1 \\ 1 & -1 & 1 & -d \end{bmatrix}\right) \otimes \begin{bmatrix} a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \\ a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \end{bmatrix}$$

## DCT factorization (in H.264)

Inverse DCT, where E is the pre-scaling matrix

$$\mathbf{X}' = \mathbf{C}_i^T (\mathbf{Y} \otimes \mathbf{E}_i) \mathbf{C}_i = \begin{bmatrix} 1 & 1 & 1 & 1/2 \\ 1 & 1/2 & -1 & -1 \\ 1 & -1/2 & -1 & 1 \\ 1 & -1 & 1 & -1/2 \end{bmatrix} \left( \left[ \quad \mathbf{Y} \quad \right] \otimes \begin{bmatrix} a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \\ a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \end{bmatrix} \right) \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1/2 & -1/2 & -1 \\ 1 & -1 & -1 & 1 \\ 1/2 & -1 & 1 & -1/2 \end{bmatrix}$$

Rounding for simplification: $a = \frac{1}{2}$, $b = \sqrt{\frac{2}{5}}$, $d = \frac{1}{2}$

$$\mathbf{Y} = \mathbf{C}_f \mathbf{X} \mathbf{C}_f^T \otimes \mathbf{E}_f = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \quad \mathbf{X} \quad \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{bmatrix}\right) \otimes \begin{bmatrix} a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \\ a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \end{bmatrix}$$

$$\mathbf{X}' = \mathbf{C}_i^T (\mathbf{Y} \otimes \mathbf{E}_i) \mathbf{C}_i = \begin{bmatrix} 1 & 1 & 1 & 1/2 \\ 1 & 1/2 & -1 & -1 \\ 1 & -1/2 & -1 & 1 \\ 1 & -1 & 1 & -1/2 \end{bmatrix} \left( \left[ \quad \mathbf{Y} \quad \right] \otimes \begin{bmatrix} a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \\ a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \end{bmatrix} \right) \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1/2 & -1/2 & -1 \\ 1 & -1 & -1 & 1 \\ 1/2 & -1 & 1 & -1/2 \end{bmatrix}$$

62

## H.264 Profiles

- Baseline
  - Minimal complexity
  - Error resiliency for unreliable networks
- Main
  - Best coding efficiency
- Extended
  - The combination of the two

| Coding Tools | Baseline | Extended | Main |
|---|---|---|---|
| I, P Slices | X | X | X |
| CAVLC | X | X | X |
| Error Resilience | X | X | |
| SP and SI Slices | | X | |
| B Slices | | X | X |
| Interlaced Coding | | X | X |
| CABAC | | | X |

## H.264 Levels

- Complexity
- Memory requirements
- 15 levels

| Level Number | Example of Typical Picture Size | Maximum Frame Rate for Typical Picture Size | Maximum Compressed bit rate (for VCL) | Maximum number of reference frames for typical picture size |
|---|---|---|---|---|
| 1 | QCIF | 15 | 64 kbps | 4 |
| 1.1 | 320x240<br>QCIF | 10<br>30 | 192 kbps | 3<br>9 |
| 1.2 | CIF | 15 | 384 kbps | 6 |
| 1.3 | CIF | 30 | 768 kbps | 6 |
| 2 | CIF | 30 | 2 Mbps | 6 |
| 2.1 | HHR | 30/25 | 4 Mbps | 6 |
| 2.2 | SD | 15 | 4 Mbps | 5 |
| 3 | SD | 30/25 | 10 Mbps | 5 |
| 3.1 | 1280X720P | 30 | 14 Mbps | 5 |
| 3.2 | 1280X720P | 60 | 20 Mbps | 4 |
| 4 | HD Formats (720P, 1080I) | 60P/30I | 20 Mbps | 4 |
| 4.1 | HD Formats (720P, 1080I) | 60P/30I | 50 Mbps | 4 |
| 4.2 | 1920X1080P | 60P | 50 Mbps | 4 |
| 5 | 2kx1k | 72 | 135 Mbps | 5 |
| 5.1 | 2kx1k<br>4kx2k | 120<br>30 | 240 Mbps | 5 |

64

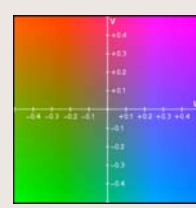
## JPEG – Veszteséges kódolás

- RGB-YUV konverzió
  - Krominancia csatornák alulmintavételezése
  - Történelmi okok: Fekete-fehér TV-kompatibilitás
  - Objektív ok: Szemünkben kevesebb a színi információkat feldolgozó receptor (csap)!
  - Pl: 4:2:0 (8 bit)


65

## RGB – YUV konverzió

- **YUV** color space: luminancia, 2 krominancia komponens
- PAL és NTSC televíziós rendszerekben
- RGB-ből átszámítása:
  - $Y = 0.299R + 0.587G + 0.114B$
  - $U = 0.492(B - Y) = -0.147R - 0.289G + 0.436B$
  - $V = 0.877(R - Y) = 0.615R - 0.515G - 0.100B$
- Fix pontos közelítés:
  - Y := min(abs(r * 2104 + g * 4130 + b * 802 + 4096 + 131072) >> 13,235)
  - U := min(abs(r * -1214 + g * -2384 + b * 3598 + 4096 + 1048576) >> 13,240)
  - V := min(abs(r * 3598 + g * -3013 + b * -585 + 4096 + 1048576) >> 13,240)




Y = 0.5

# References


- [1] Martin Fiedler – **Seminar Paper -** Implementation of a basic H.264/AVC Decoder
- [2] Ajay Luthra, Pankaj Topiwala – **Overview of the H.264/AVC video coding standard**
- [3] Iain E G Richardson – **H.264 / MPEG-4 Part 10 White Paper** (*www.vcodex.com*)
- [4] Dirk Farin – **Introduction to Video Compression Introduction to Video Compression H.261**
- [5] Gary J. Sullivan, Pankaj Topiwala, and Ajay Luthra – **The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions**
- [6] THOMAS SIKORA, SENIOR MEMBER, IEEE – **Trends and Perspectives in Image and Video Coding**
- [7] Majid Rabbani – **JPEG-2000: Background, Scope, And Technical Description**