Video-Coding Basics

Univ.Prof. Dr.-Ing. Markus Rupp

LVA: 389.134 Video and Multimedia Transmissions over cellular Networks

LVA 389.054 Mobile Kommunikation, Vertiefung

LVA 389.168 Advanced Wireless Communications 1

Last change: Nov 25 2013

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Outline

- Basics on Video Sampling
  - Video Standards
- Data Rate Reduction
  - A brief overview of image cvompression
  - Video compression techniques
- Quality Improvements:
  - Deblocking, Error concealment
- Video over wireless

Video Sampling

- 3D Sampling:
  - 2D spatial domain (pixels)
  - 1D temporal domain (frame rate)
- Standards (Analogue):
  - National Television Systems Committee (NTSC) in use in Canada, Japan, South Korea, USA, and some other places in South America, working with 29.97 f/s (denoted commonly as 30 f/s)
  - In the rest of the world Phase Alternation by Line (PAL) and Sequentiel couleur a memoire (SECAM) are used, operating at a frame rate of 25 f/s.

Digital Television (DTV) (not yet for cellular)

- is the transmission of audio and video by digital signals, in contrast to the analog signals used by analog TV. Many countries are replacing broadcast analog television with digital television to allow other uses of the television radio spectrum.
- With DTV broadcasting, the range of formats can be broadly divided into two categories: high definition television (HDTV) for the transmission of high-definition video and standard-definition television (SDTV). These terms by themselves are not very precise, and many subtle intermediate cases exist.
- One of several different HDTV formats that can be transmitted over DTV is: 1280 × 720 pixels in progressive scan mode (abbreviated *720p*) or 1920 × 1080 pixels in interlaced video mode (*1080i*). Each of these utilizes a 16:9 aspect ratio. (Some televisions are capable of receiving an HD resolution of 1920 × 1080 at a 60 Hz progressive scan frame rate — known as 1080p.) HDTV cannot be transmitted over current analog television channels because of channel capacity issues.
- SDTV may use one of several different formats taking the form of various aspect ratios depending on the technology used in the country of broadcast. For 4:3 aspect-ratio broadcasts, the 640 × 480 format is used in NTSC countries, while 720 × 576 is used in PAL countries. For 16:9 broadcasts, the 704 × 480 format is used in NTSC countries, while 720 × 576 is used in PAL countries.

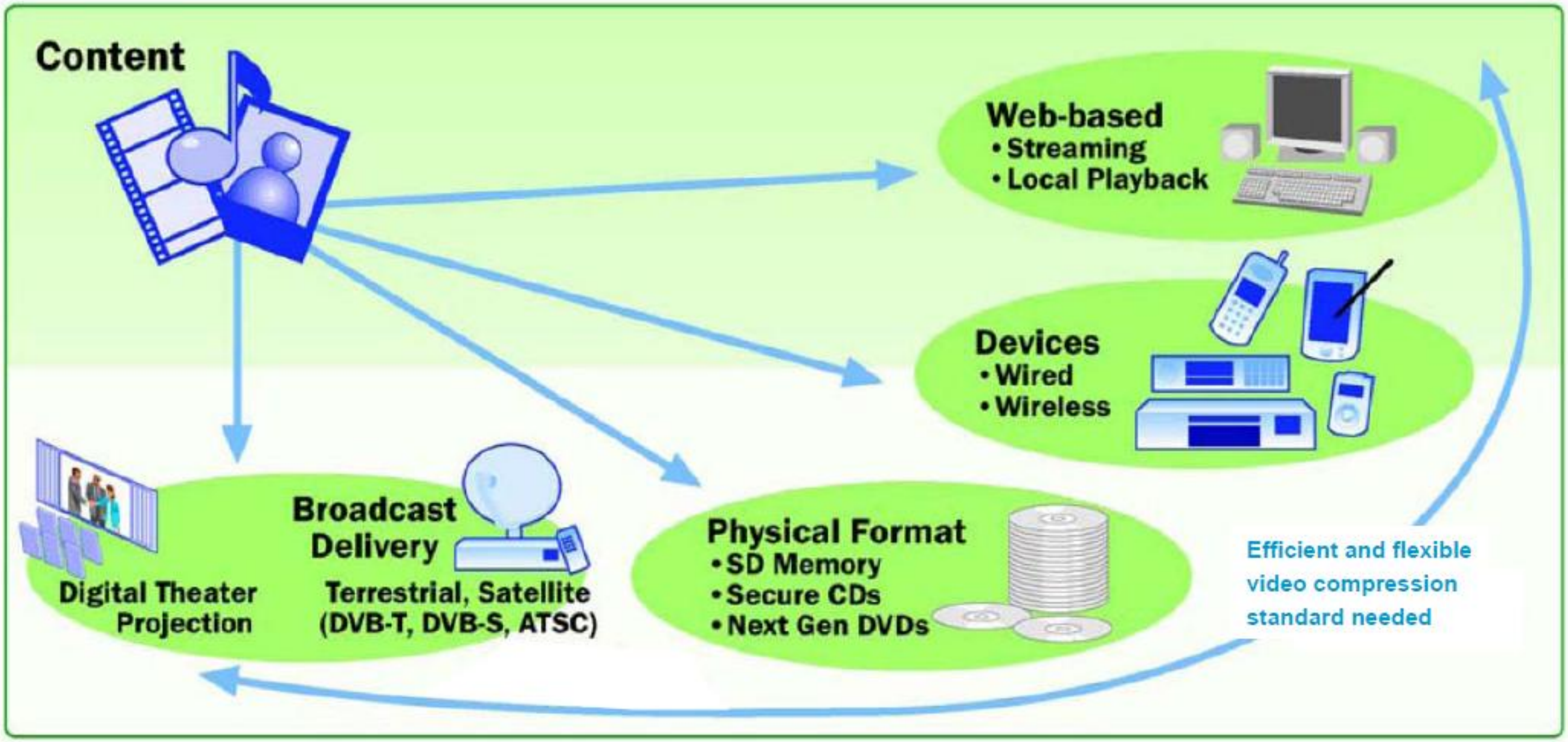
Digital Video coding standards also suited for cellular: applications and common structure

- ITU-T Rec. H.261 1988
- (ITU-T Rec. H.263) 1995
- ISO/IEC MPEG-1 1991
- ISO/IEC MPEG-2 1994
- (ISO/IEC MPEG-4) 1998
- State-of-the-art: H.264/AVC 2001
- New state of the art: H.265 2013

Applications of Video Compression

Digital television broadcasting	2 . . . 6 Mbps (10…20 Mbps for HD)	MPEG-2 (H264/AVC)
DVD video	5 . . . 8 Mbps	MPEG-2
Internet video streaming	20 . . . 300 kbps	MPEG-1, H.264/AVC, VC-1, or similar proprietary
Videoconferencing, videotelephony	20 . . . 2000 kbps	H.261, H.263, H.264/AVC
Video over 3G wireless	100 . . . 500 kbps	H.263, MPEG-4, H.264/AVC

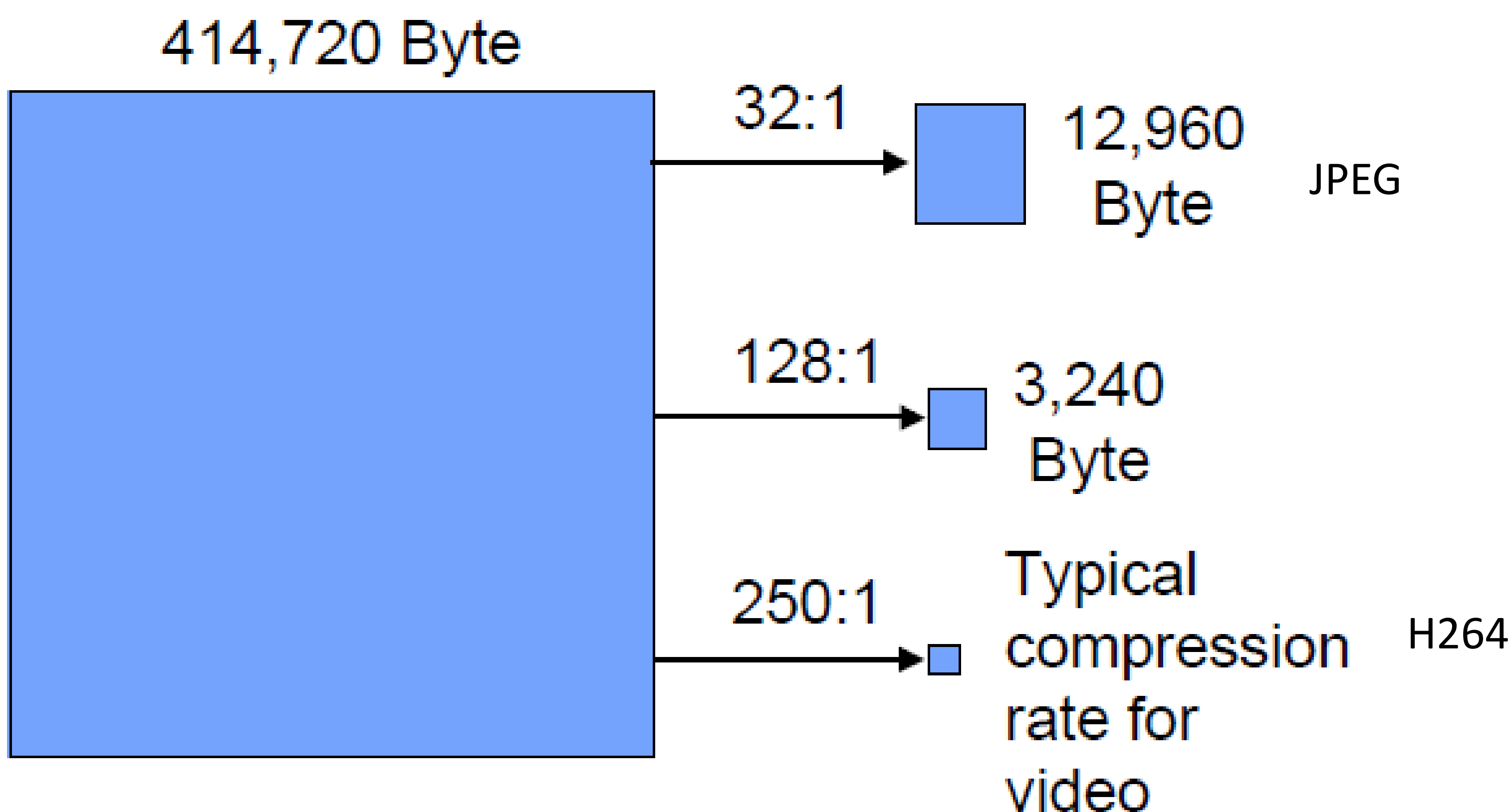
Applications of Video Compression



- Adapted from*[Srinivasanet al., 2004]*

Example

Geometric Interpretation



Luminance and Chrominance vs RGB

- Camera samples three colors: RGB

$$Y = k_r R + (1 - k_b - k_r) G + k_b B$$

$$C_b = \frac{0.5}{1 - k_b}(B - Y)$$

$$C_r = \frac{0.5}{1 - k_r}(R - Y)$$

$$k_b = 0.114, k_r = 0.229,$$

- Reason is better coding effect. For 4Y (luma) pixels, typically 1B and 1R are sufficient (YCrCb=4:2:0)

Bit Rates

- Typically an NxM image is sampled with 3q bits each pixel, resulting in 3qNM pixels per image
- Example: N=M=1000, q=6bits, FR=30f/s
  - 560 Mbit/s
- Rate reduction techniques:
  - Frame rate decimation (2,3,4,5) for low quality
  - Interlaced vs progressive (=non interlaced) scan

Mobile Video Standards

- Due to low quality and small screens, NxM can be selected a lot smaller:

Abbreviation	Size	Description
VGA	640×480	Video Graphics Array
QVGA	320×240	Quarter Video Graphics Array, called also Standard Interchange Format (SIF)
Q2VGA	160×120	
CIF	352×288	Common Intermediate Format (quarter of resolution 704×576 used in PAL)
QCIF	176×144	Quarter Common Intermediate Format

Smart phone (2011)

Cheap, low cost (2011)

- Mobile Video Example (4:2:0, QCIF): 1.5x25x8x176x144=7.6Mbit/s

Smart phone 2013: 2.560 x 1.440

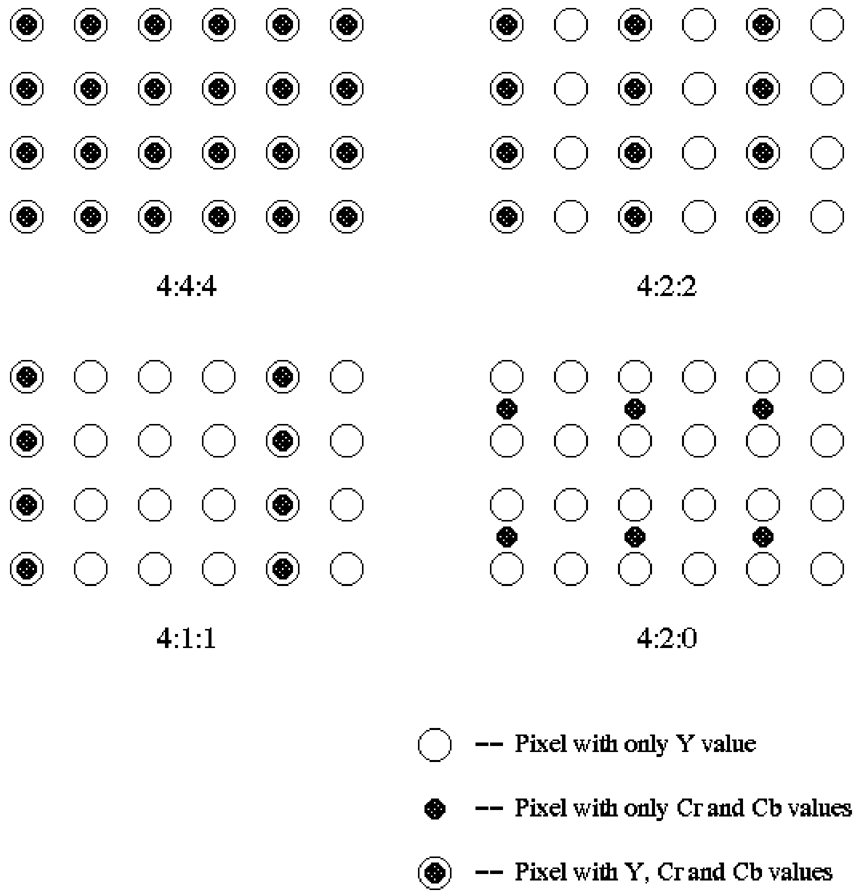
Chroma Subsampling

- Because of storage and transmission limitations, there is always a desire to reduce (or compress) the signal. Since the human visual system is much more sensitive to variations in brightness than color, a video system can be optimized by devoting more bandwidth to the luma component (usually denoted Y'), than to the color difference components Cb and Cr.
- The 4:2:2 Y'CbCr scheme for example requires two-thirds the bandwidth of (4:4:4) R'G'B'. This reduction results in almost no visual difference as perceived by the viewer.

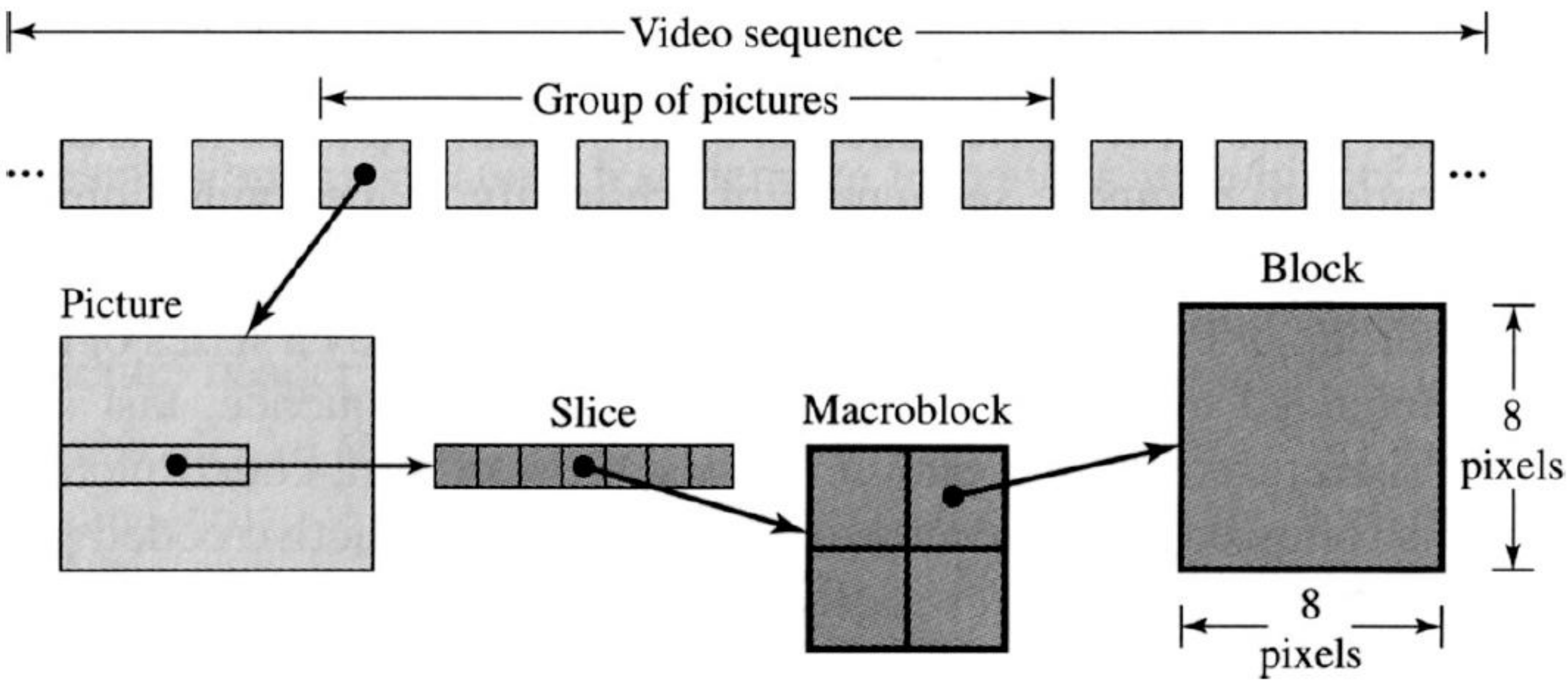
Chroma Subsampling

- The subsampling scheme is commonly expressed as a three part ratio *J*:*a*:*b* (e.g. 4:2:2), that describe the number of luminance and chrominance samples in a conceptual region that is *J* pixels wide, and 2 pixels high. The parts are (in their respective order):
  - **J** horizontal sampling reference (width of the conceptual region). Usually, 4.
  - **a** number of chrominance samples (Cr, Cb) in the first row of *J* pixels.
  - **b** number of (additional) chrominance samples (Cr, Cb) in the second row of *J* pixels.
- See also http://lea.hamradio.si/~s51kq/V-BAS.HTM

Subsampling YCrCB schemes



Video Compression Standards: Hierarchical Syntax

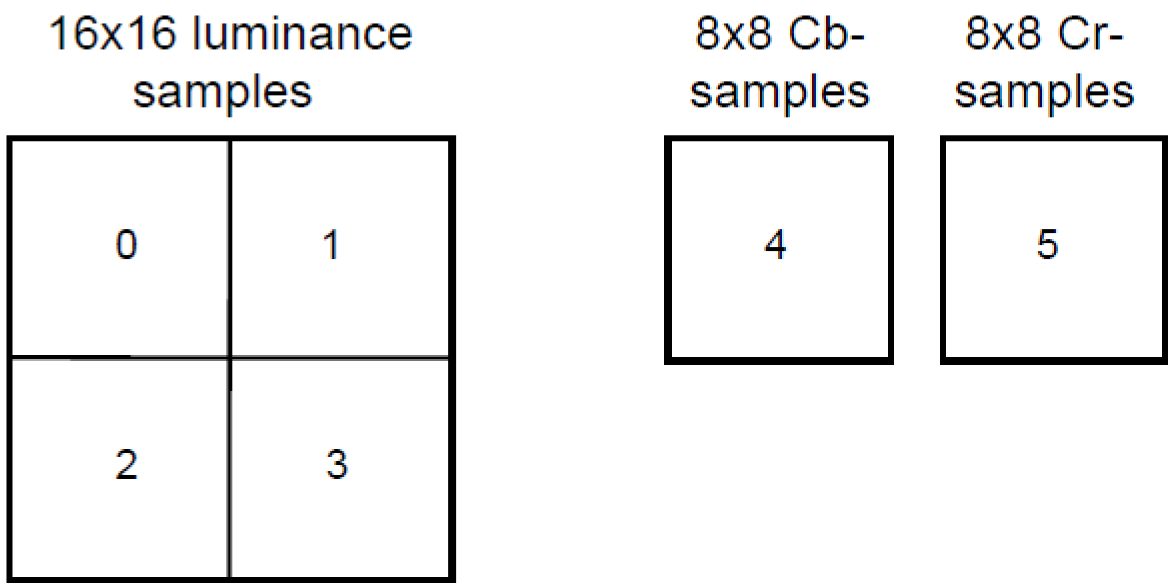


ITU-T Rec. H.261

- International standard for ISDN picture phones and for video conferencing systems (1990)
- Image format: CIF (352 x 288 Y samples) or
- QCIF (176 x 144 Y samples), frame rate 7.5 ... 30 f/s
- Bit-rate: multiple of 64 kbit/s (= ISDN-channel), typically 128 kbit/s including audio.
- Picture quality: for 128 kbit/s acceptable with limited motion in the scene
- Stand-alone videoconferencing system or
- desk-top videoconferencing system, integrated with PC

H.261 Macroblocks

- Macroblock (MB) of 16x16 pixels
- Sampling format: 4:2:0
- MB consists of 4 luminance and 2 chrominance blocks



TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

A bit of image compression basics

- Orthogonal Transform
- 2D DCT
- Laplacian Densities
- Coding: Zigzag, runlength, entropy

Coding gain of orthonormal transform

- Assume distortion rate functions for image samples

$$\boxed{d(R) \cong \varepsilon^2 \sigma_X^2 2^{-2R}}$$

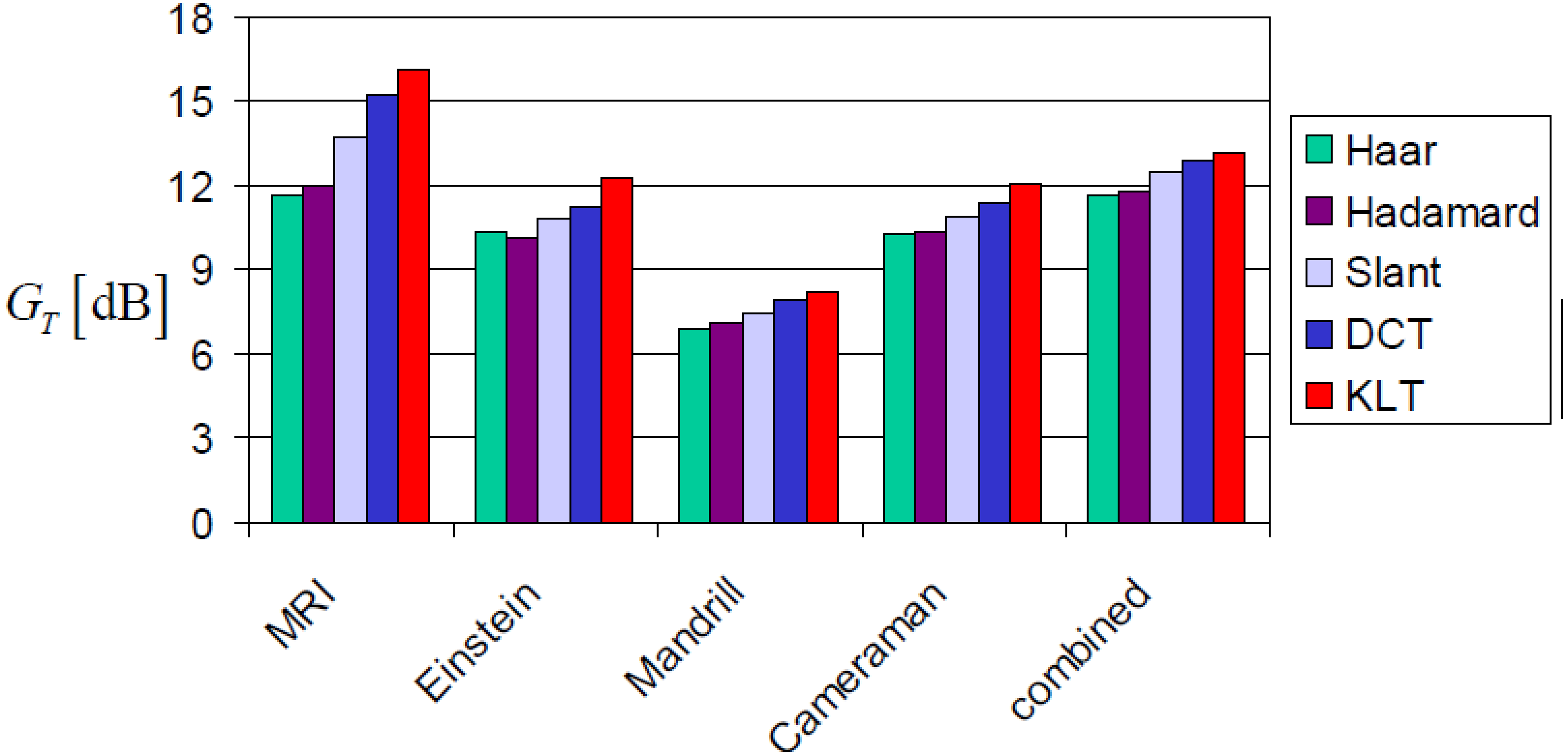
. . . and for encoding transform coefficients

$$\boxed{d^{XFORM}(R) = \frac{1}{N}\sum_{n=0}^{N-1} d_n(R_n) \cong \frac{1}{N}\sum_{n=0}^{N-1} \varepsilon^2 \sigma_{Y_n}^2 2^{-2R_n}; \qquad R = \frac{1}{N}\sum_{n=0}^{N-1} R_n}$$

- Transform coding gain

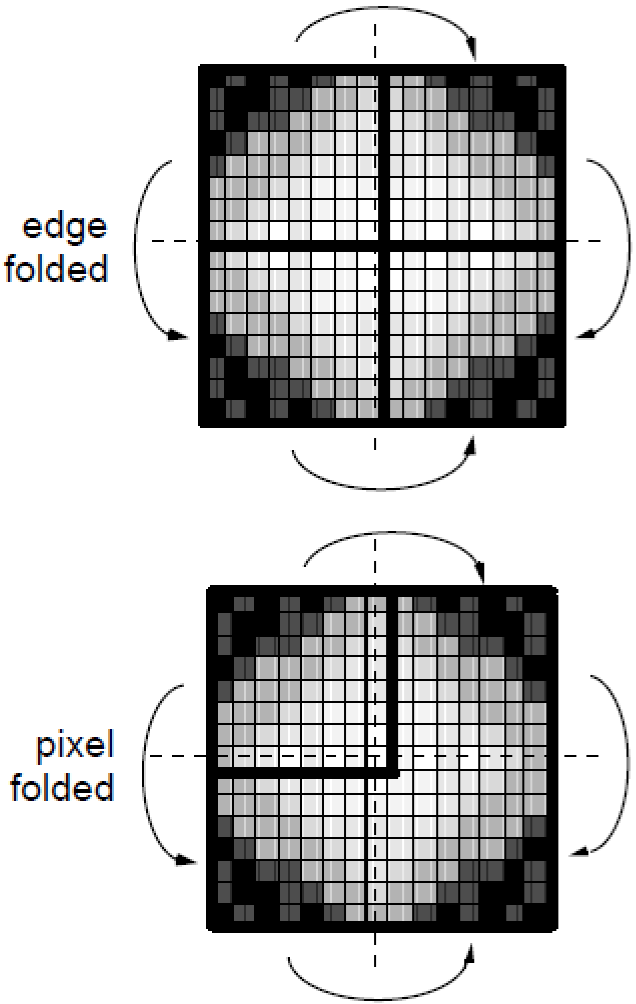
$$\boxed{G_T = \frac{d(R)}{d^{XFORM}(R)}}$$

Coding gain with 8x8 transforms



DCT vs DFT

- Transform coding of images using the Discrete Fourier Transform (DFT):
  - For stationary image statistics, the energy concentration properties of the DFT converge against those of the KLT for large block sizes.
  - <u>Problem</u> of blockwise DFT coding: blocking effects due to circular topology of the DFT and Gibbs phenomena.
  - <u>Remedy:</u> reflect image at block boundaries, DFT of larger symmetric block -> “DCT“



2D DCT

Type II-DCT of blocksize NxN is defined by transform matrix A containing elements

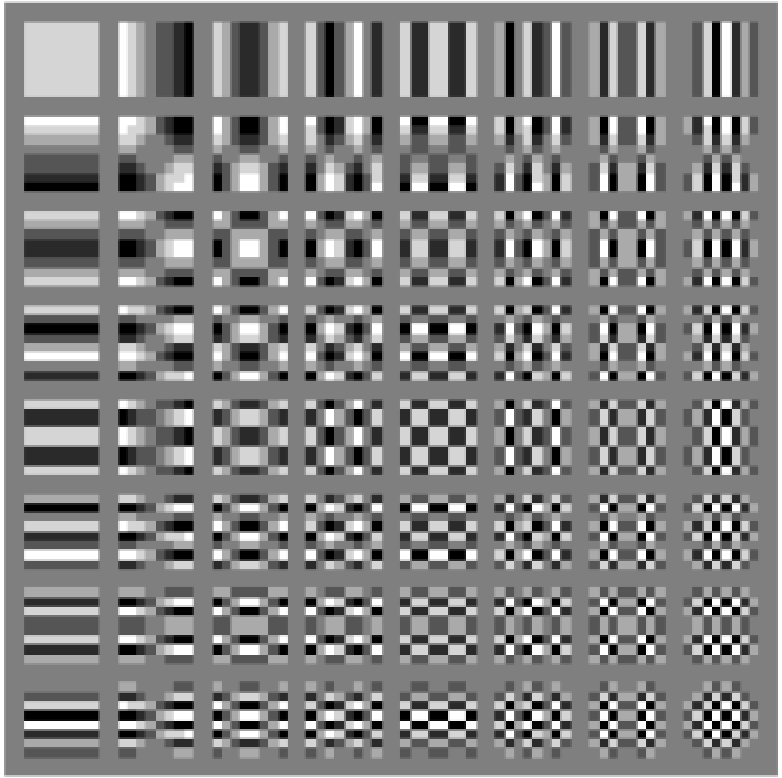
$$a_{ik} = \alpha_i \cos \frac{\pi(2k+1)i}{2N}$$

$$\text{for } \; i,k = 0,\ldots,N-1$$

$$\text{with} \quad \alpha_0 = \sqrt{\frac{1}{N}}$$

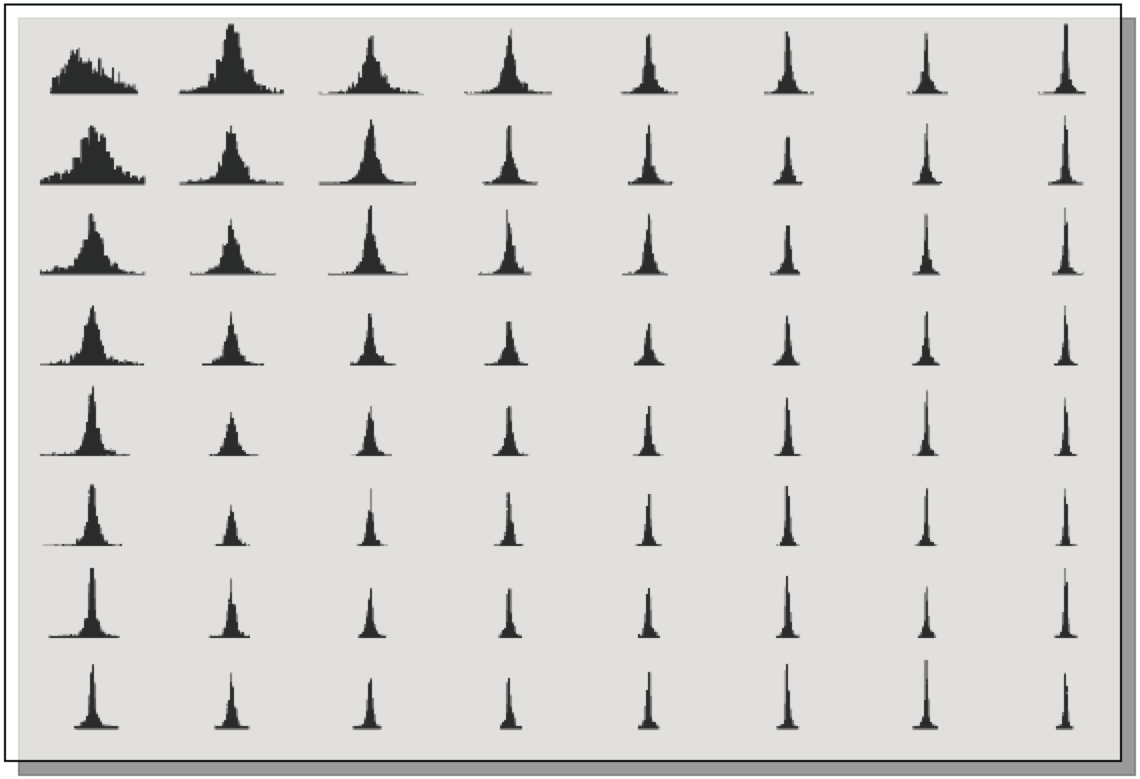
$$\alpha_i = \sqrt{\frac{2}{N}} \quad \forall i \neq 0$$

- 2D basis functions of the DCT:



Orthonormal Transforms results in Laplacian Densities

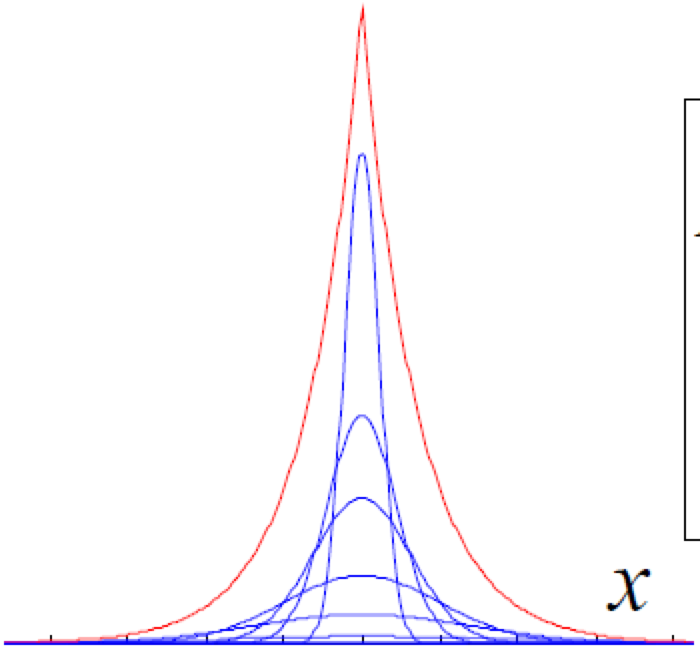
- Histograms for 8x8 DCT coefficient amplitudes measured for test image *[Lam, Goodman, 2000]*



Test image *Bridge*

- AC coefficients: Laplacian PDF
- DC coefficient distribution similar to the original image

Laplacian Density



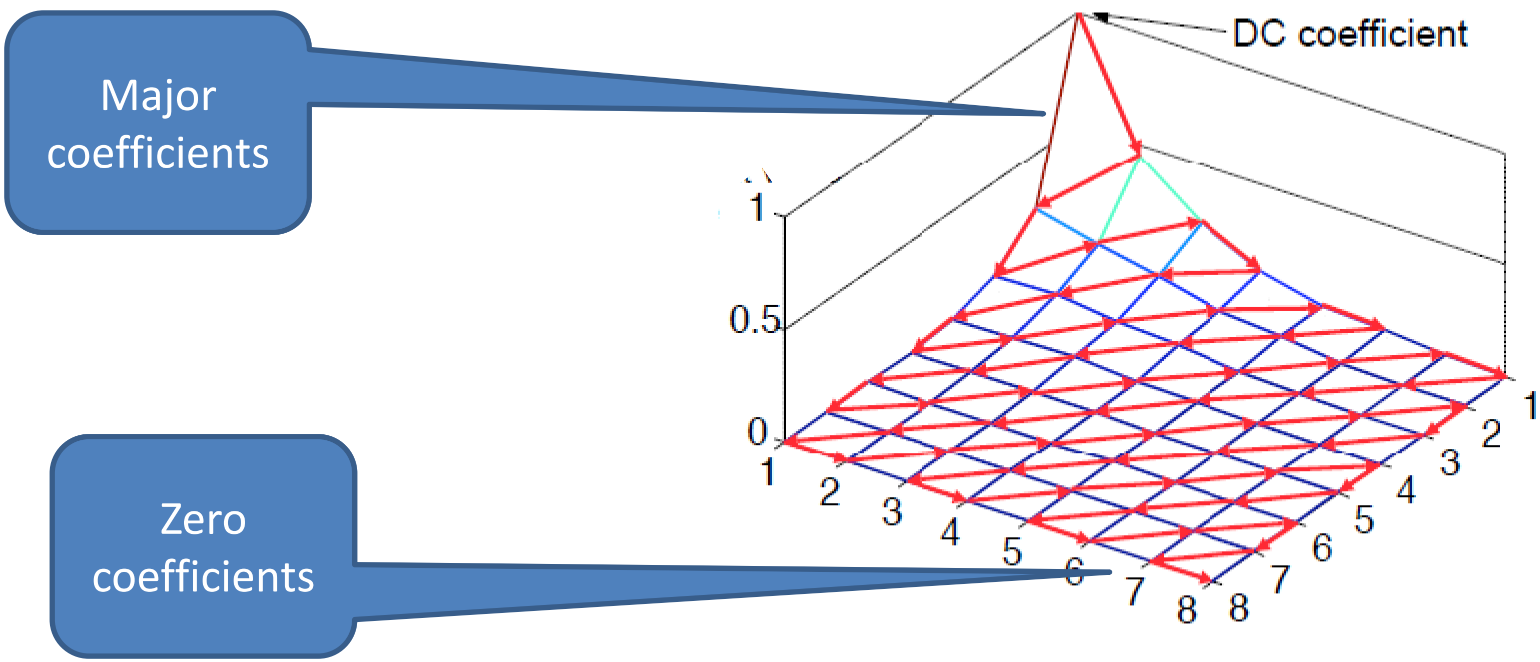
$$p_{Y_n}(y) = \int_0^\infty \frac{1}{\sqrt{2\pi v}} \cdot e^{-y^2/2v} \frac{1}{\sigma^2} e^{-v/\sigma_{y_n}^2} dv$$

$$= \sqrt{\frac{1}{2\sigma_{y_n}^2}} \cdot e^{-\sqrt{2}\cdot|y|/\sigma_{y_n}}$$

- For a given block variance, coefficient pdfs are Gaussian
- Gaussian mixture w/ exponential variance distribution yields a Laplacian
- Gaussian mixture w/ half-Gaussian variance distribution yields pdf very close to Laplacian *[Lam, Goodman, 2000]*
- Elegant explanation of Laplacian pdfs of DCT coefficients

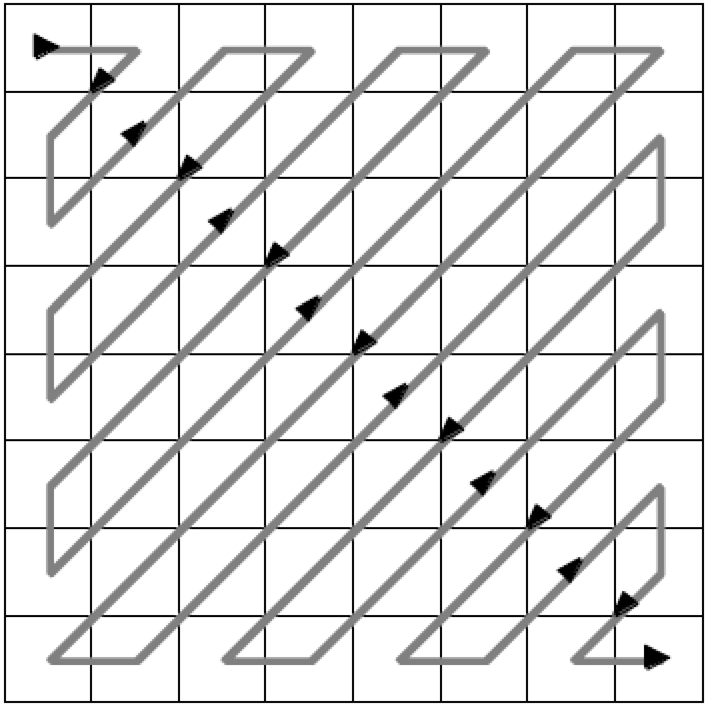
2D DCT

- 2D Discrete Cosine Transform maps an 8x8 block onto:



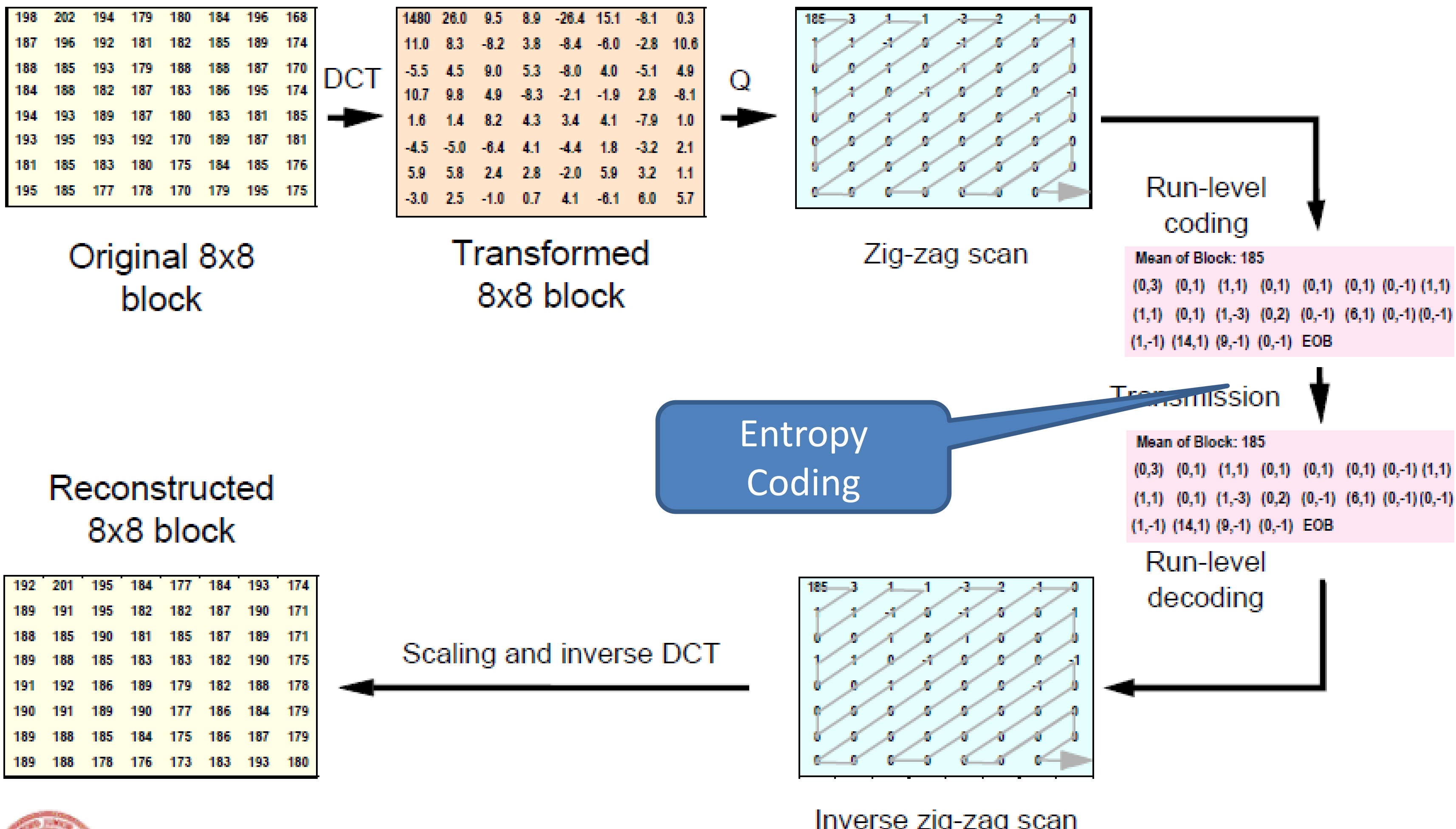
Zig-Zag-Scan

Efficient encoding of the position of non-zero transform coefficients: zig-zag-scan + run-level-coding



ordering of the transform coefficients by zig-zag-scan

Threshold Coding



Original 8x8 block

198	202	194	179	180	184	196	168
187	196	192	181	182	185	189	174
188	185	193	179	188	188	187	170
184	188	182	187	183	186	195	174
194	193	189	187	180	183	181	185
193	195	193	192	170	189	187	181
181	185	183	180	175	184	185	176
195	185	177	178	170	179	195	175

Transformed 8x8 block

1480	26.0	9.5	8.9	-26.4	15.1	-8.1	0.3
11.0	8.3	-8.2	3.8	-8.4	-6.0	-2.8	10.6
-5.5	4.5	9.0	5.3	-8.0	4.0	-5.1	4.9
10.7	9.8	4.9	-8.3	-2.1	-1.9	2.8	-8.1
1.6	1.4	8.2	4.3	3.4	4.1	-7.9	1.0
-4.5	-5.0	-6.4	4.1	-4.4	1.8	-3.2	2.1
5.9	5.8	2.4	2.8	-2.0	5.9	3.2	1.1
-3.0	2.5	-1.0	0.7	4.1	-6.1	6.0	5.7

Reconstructed 8x8 block

192	201	195	184	177	184	193	174
189	191	195	182	182	187	190	171
188	185	190	181	185	187	189	171
189	188	185	183	183	182	190	175
191	192	186	189	179	182	188	178
190	191	189	190	177	186	184	179
189	188	185	184	175	186	187	179
189	188	178	176	173	183	193	180

Entropy of a Memoryless Source

- Let a memoryless source be characterized by an ensemble U_0 with:

$$\boxed{\begin{array}{l}\text{Alphabet } \{ a_0, a_1, a_2, \ldots, a_{K-1} \} \\ \text{Probabilities } \{ P(a_0), P(a_1), P(a_2), \ldots, P(a_{K-1}) \}\end{array}}$$

- Shannon: information conveyed by message "a_k":

$$\boxed{I(a_k) = -\log(P(a_k))}$$

- "Entropy of the source" is the <u>average</u> information contents:

$$\boxed{H(U_0) = E\{I(a_k)\} = -\sum_{k=0}^{K-1} P(a_k) * \log(P(a_k))}$$

- For „log" = „$\log_2$" the unit is bits/symbol

Redundant Codes: Example

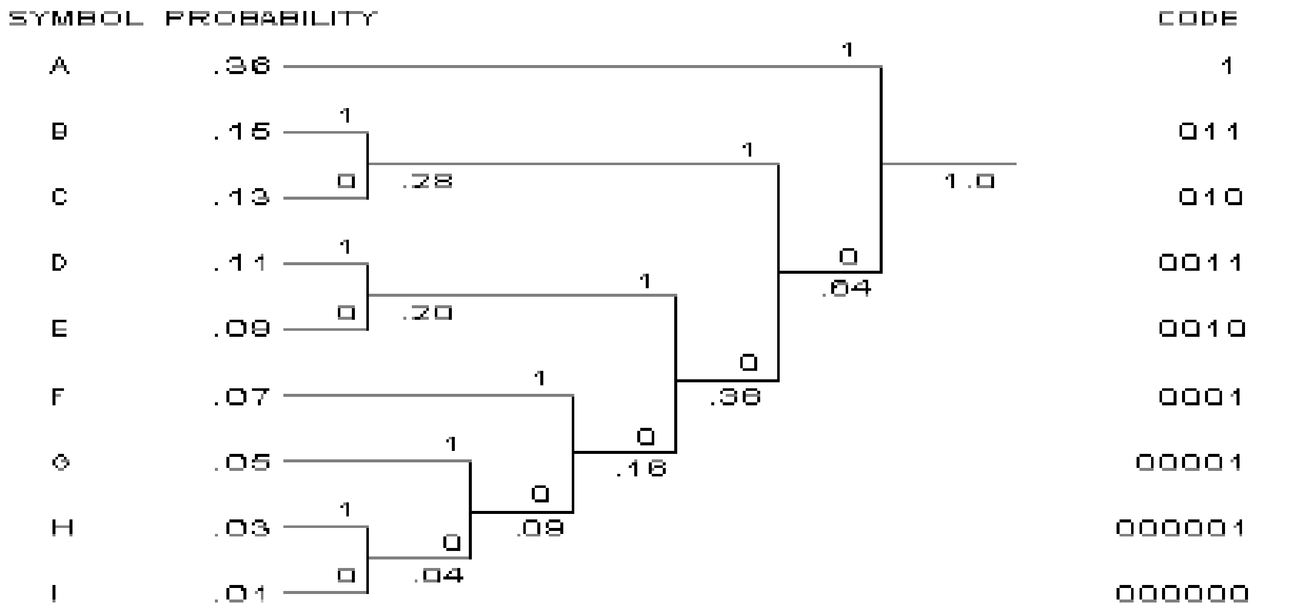
a_i	$P(a_i)$	redundant code	optimum code
a_1	0.500	00	0
a_2	0.250	01	10
a_3	0.125	10	110
a_4	0.125	11	111
$H(U_0)$ =1.75 bits		λ_{av}= 2 bits ρ = 0.25 bits	λ_{av}=1.75 bits ρ = 0 bits

Huffman Code

- Design algorithm for variable length codes proposed by D. A. Huffman (1952) always finds a code with minimum redundancy.
- Obtain code tree as follows:

1 Pick the two symbols with lowest probabilities and merge them into a new auxiliary symbol.
2 Calculate the probability of the auxiliary symbol.
3 If more than one symbol remains, repeat steps 1 and 2 for the new auxiliary alphabet.
4 Convert the code tree into a prefix code.

Hoffman Code Example



Fixed length coding: $R_{fixed} = 4$ bits/symbol
Huffman code: $R_{Huffman} = 2.77$ bits/symbol
Entropy $H(X) = 2.69$ bits/symbol
Redundancy of the Huffman code: $\rho = 0.08$ bits/symbol

Example: Morse vs. Huffman

	%	Morse Code	Huffman Code
A	6.22	.-	1011
B	1.32	-...	010100
C	3.11	-.-.	10101
D	2.97	-..	01011
E	10.53	.	001
F	1.68	..-.	110001
G	1.65	–.	110000
H	3.63	....	11001
I	6.14	..	1001
J	0.06	.—	01010111011
K	0.31	-.-	01010110
L	3.07	.-..	10100
M	2.48	–	00011
N	5.73	-.	0100
O	6.06	—	1000
P	1.87	.–.	00000
Q	0.10	–.-	0101011100
R	5.87	.-.	0111
S	5.81	...	0110
T	7.68	-	1101
U	2.27	..-	00010
V	0.70	...-	0101010
W	1.13	.–	000011
X	0.25	-..-	010101111
Y	1.07	-.–	000010
Z	0.06	–..	0101011101011

Figure 1: Morse and Huffman Codes for American-Roman Alphabet. The % column indicates the average probability (expressed in percent) of the letter occurring in English. The entropy $H(A)$ of the this source is 4.14 bits. The average Morse codeword length is 2.5 symbols. Adding one more symbol for the letter separator and converting to bits yields an average codeword length of 5.56 bits. The average Huffman codeword length is 4.35 bits.

Typical DCT Coding Artifacts

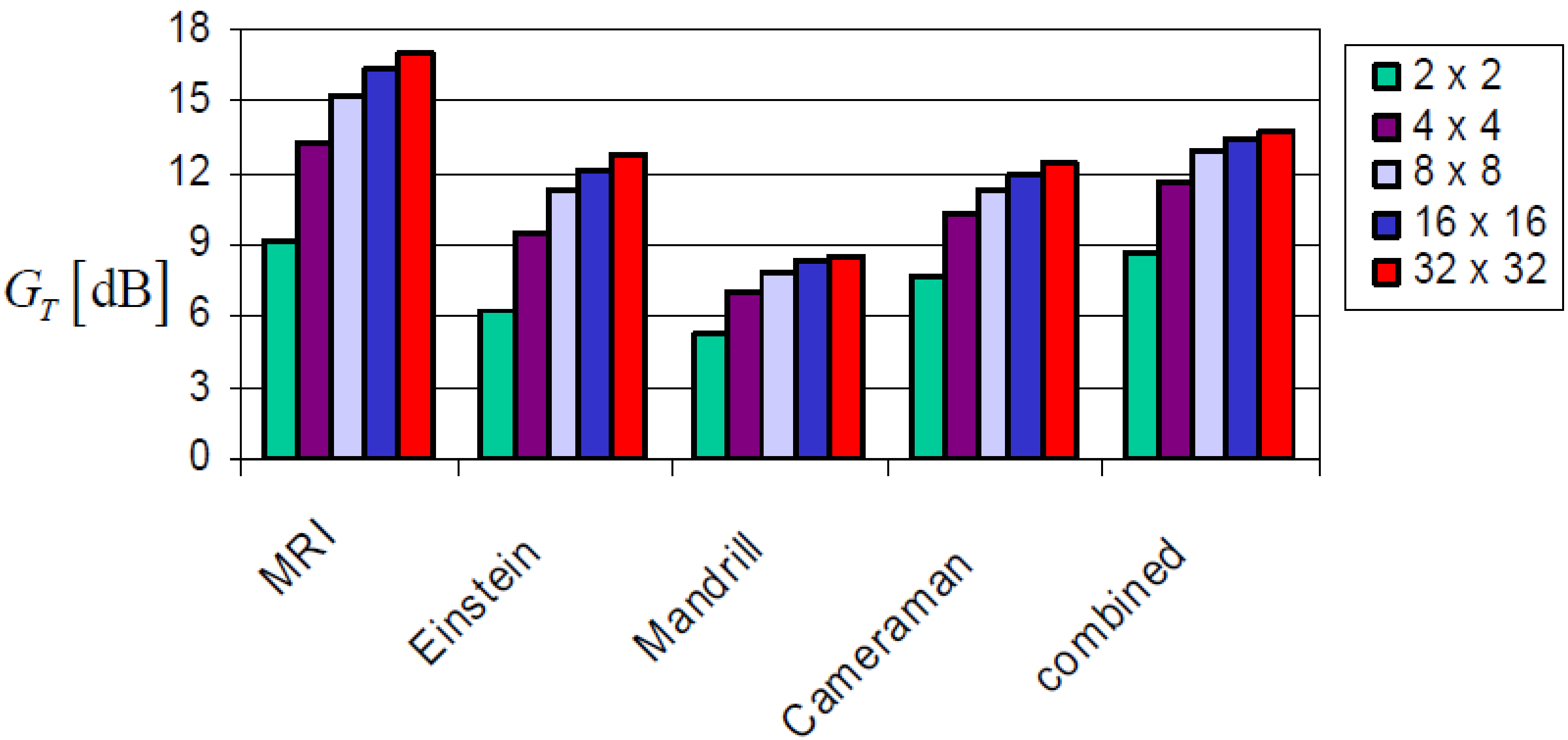
DCT coding with increasingly coarse quantization, block size 8x8

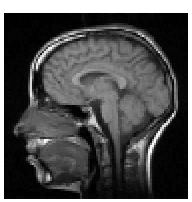
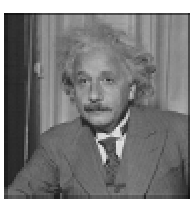
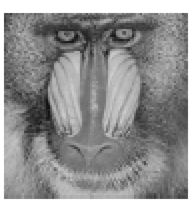
quantizer stepsize
for AC coefficients: 25

quantizer stepsize
for AC coefficients: 100

quantizer stepsize
for AC coefficients: 200

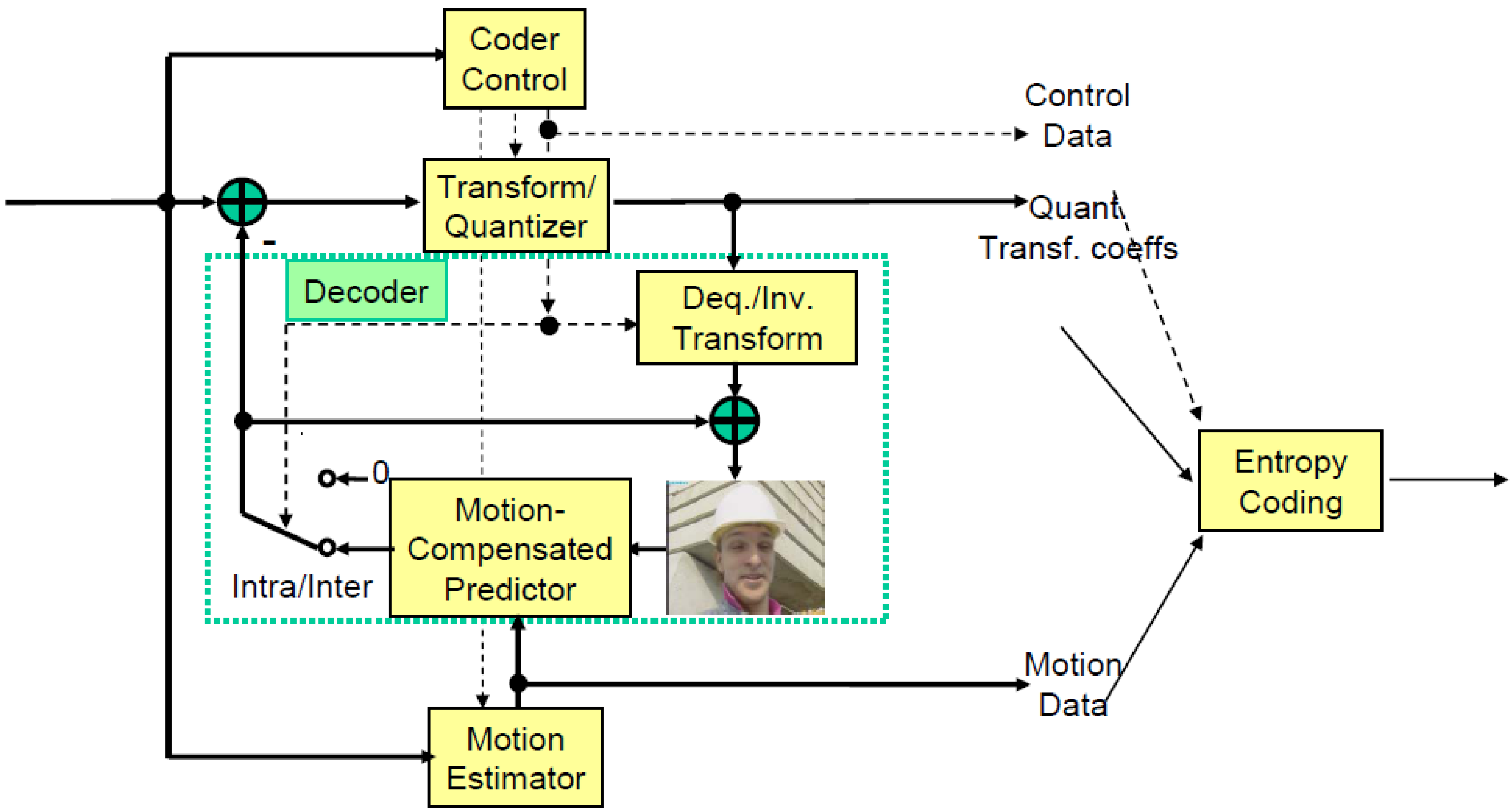
Influence of DCT Blocksize




Motion-compensated Hybrid Coding

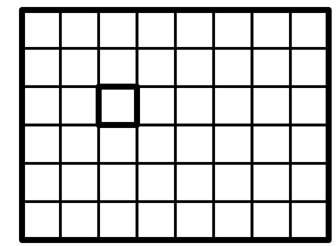
H.261, MPEG-1, MPEG-2, H.263, MPEG-4, H.264/AVC



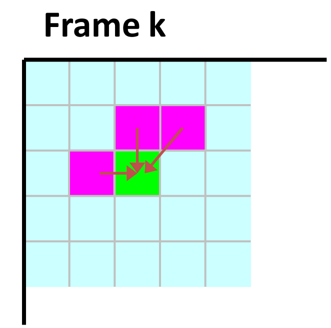
Prediction

- One sequence is encoded exploiting its **spatial** and **temporal** correlation

- As a first step, the picture is segmented into **macroblocks**

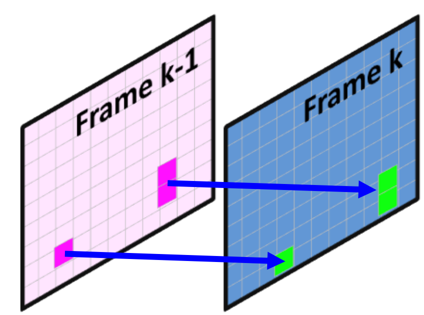


- A prediction is built for each macroblock



- The INTRA (spatial) encoding uses the neighboring macroblocks as source of prediction.

- The INTER (temporal) encoding uses the macroblocks belonging to the previous pictures as a source of prediction



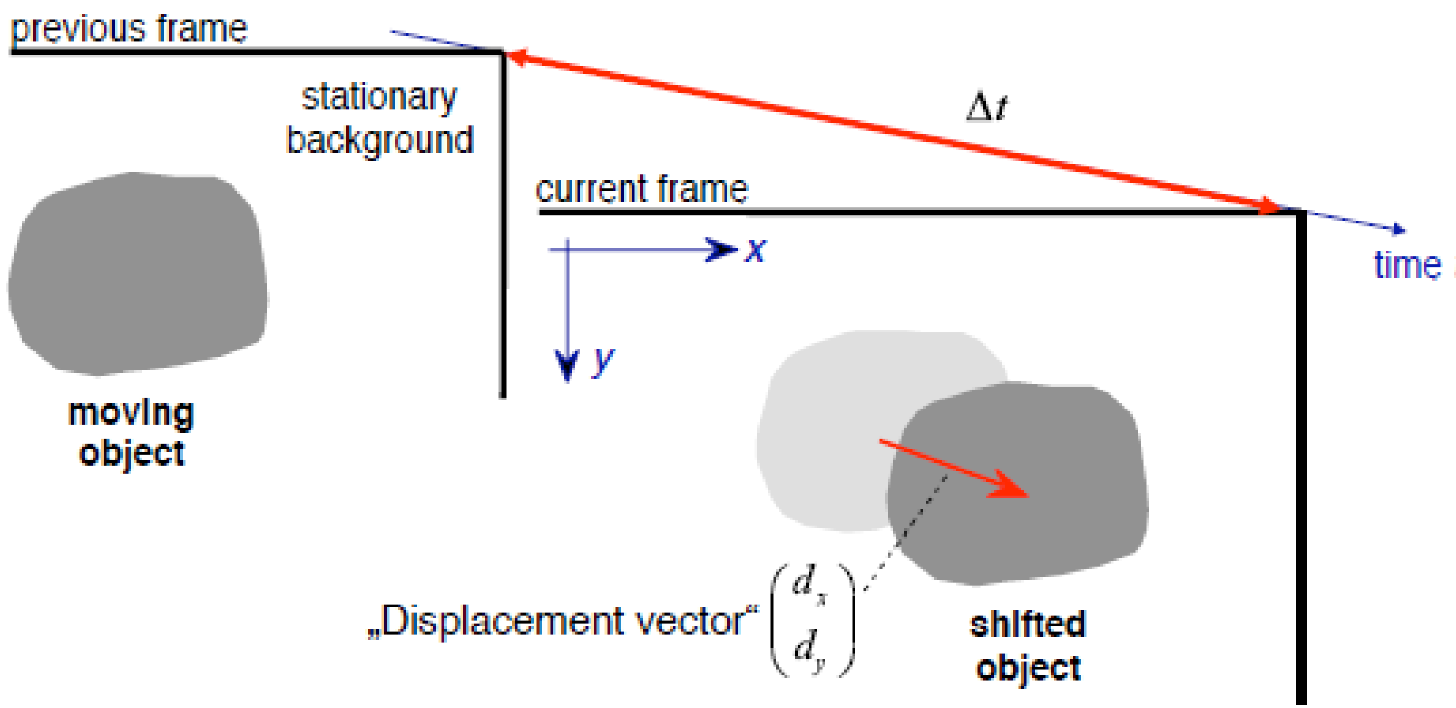
- The INTER encoding is much more performant than the INTRA encoding but...
  - Scene changes
  - ...

Interframe Coding of Video Signals

. . . exploits similarity of successive pictures

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Motion Compensated Prediction
Block Matching Algorithm



Prediction for the luminance signal $S(x,y,t)$ within the moving object:

$$\boxed{\hat{S}(x,y,t) = S(x-d_x, y-d_y, t-\Delta t)}$$

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Performance Indicators

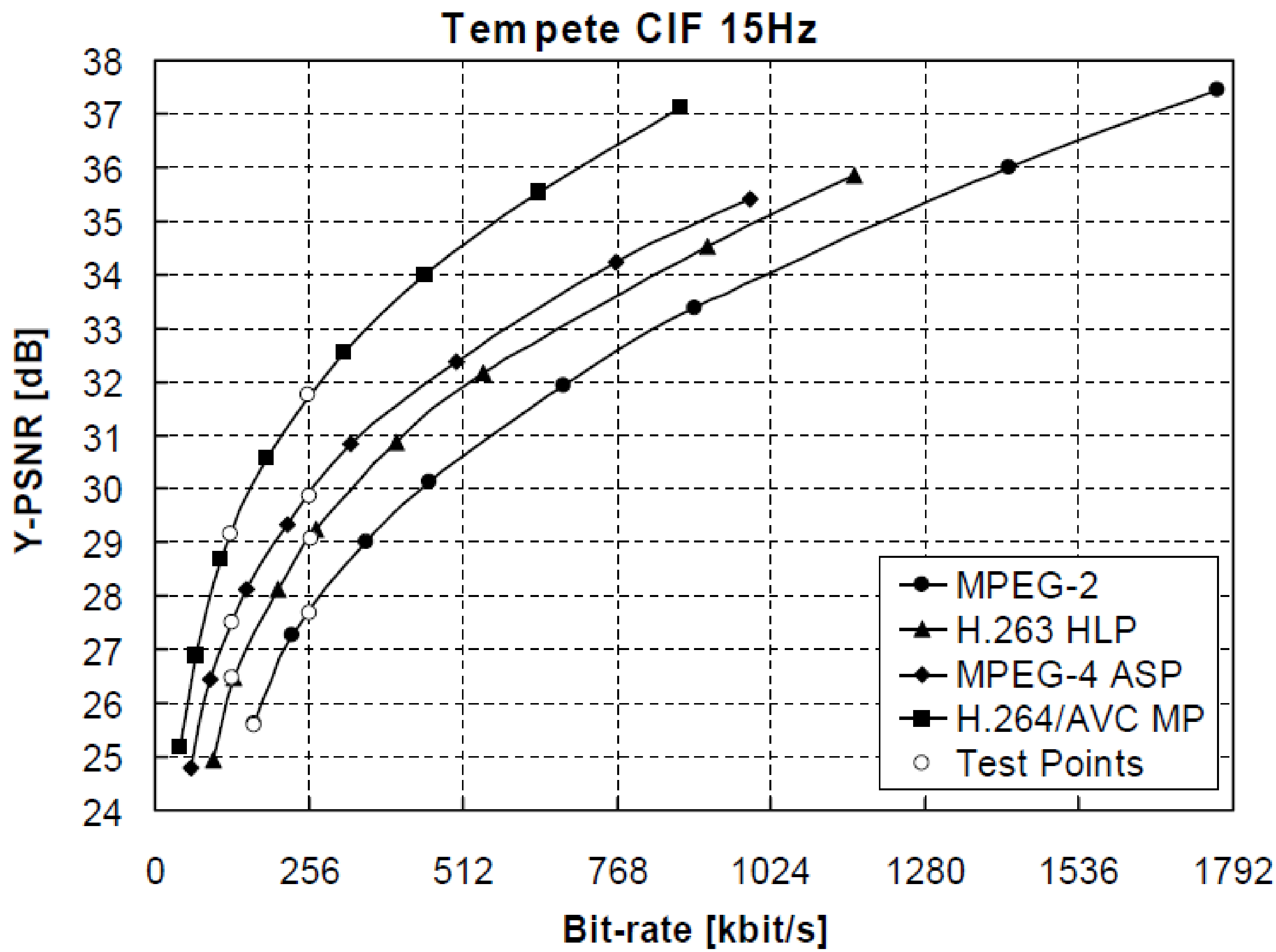
- Minimum Mean Square Error (MMSE)

$$\mathrm{MSE}[n] = \frac{1}{M \cdot N \cdot |\mathcal{C}|} \sum_{c \in \mathcal{C}} \sum_{i=1}^{N} \sum_{j=1}^{M} \left[\mathbf{F}_n^{(c)}(i,j) - \mathbf{R}_n^{(c)}(i,j) \right]^2$$

- Peak Signal to Noise Ratio (PSNR)

$$\mathrm{PSNR}[n] = 10 \cdot \log_{10} \frac{(2^q - 1)^2}{\mathrm{MSE}[n]} \; [\mathrm{dB}]$$

Video Coding Test Results



[Wiegand, et al. 2003]

Soccer Video Sequences

- Soccer video streaming is one of the preferred contents
- The quality (as appreciated by the users) suffers from
  - Resolution downsampling (to fit the mobile device display)
  - High compression ratio (to match the available data rate)

Original uncompressed sequence

Compressed sequence

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

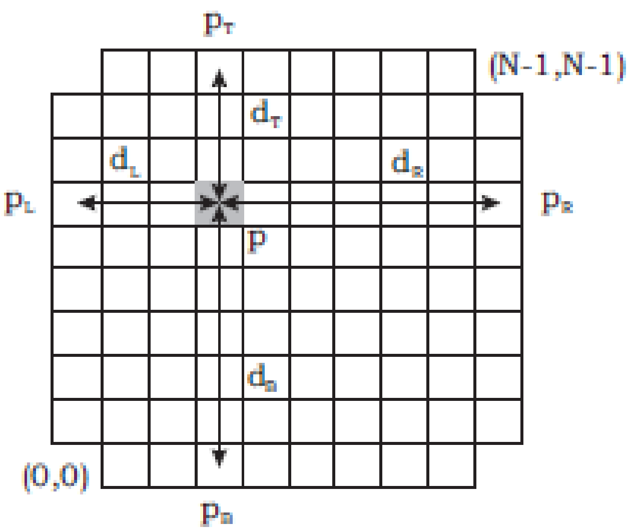
Quality Improvement: Deblocking Filter

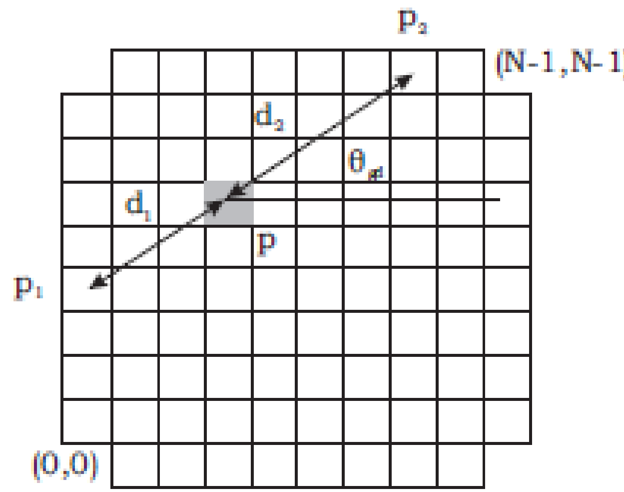
Without Filter With H264/AVC Deblocking

[source: G. Sullivan, VCEG]

Error Concealment Methods (not standardized)

- Spatial Concealment by interpolation





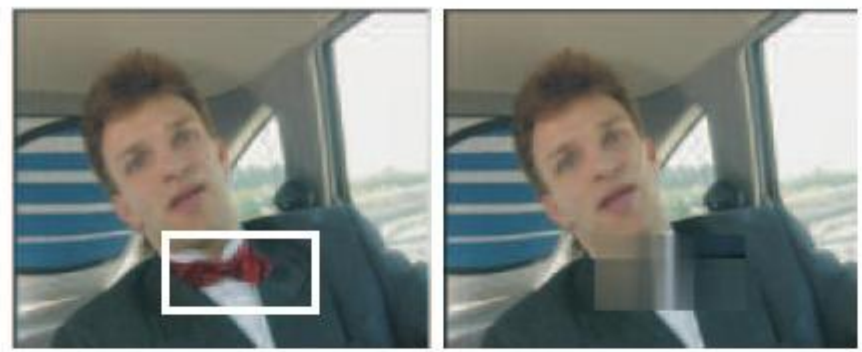
- weighted averaging — directional interpolation

Dissertation O.Nemethova

$$f_{i,j} = \frac{1}{d_1 + d_2}\left[d_2 f_{i_1,j_1} + d_1 f_{i_2,j_2}\right],$$

Spatial Interpolation

weighted averaging

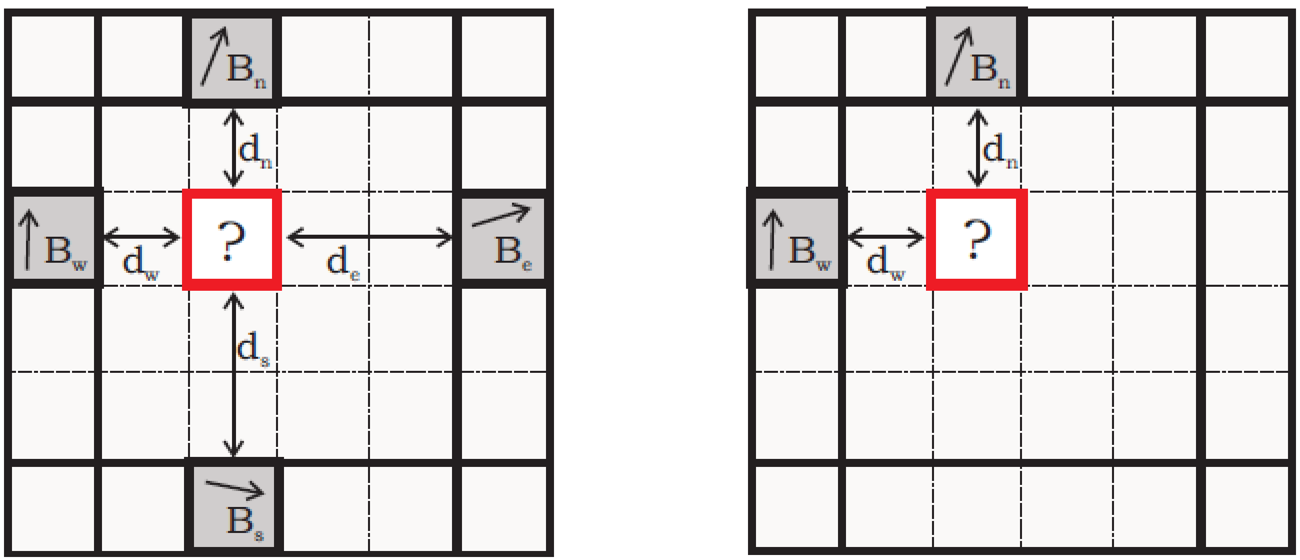


Directional interpolation

Dissertation O.Nemethova

Temporal Concealment

- Motion Vector estimation:



$$\widehat{\underline{mv}}^{(i,j)} = \frac{d_{\mathrm{e}}\underline{mv}_{\mathrm{w}}^{(j)} + d_{\mathrm{w}}\underline{mv}_{\mathrm{e}}^{(j)} + d_{\mathrm{n}}\underline{mv}_{\mathrm{s}}^{(i)} + d_{\mathrm{s}}\underline{mv}_{\mathrm{n}}^{(i)}}{d_{\mathrm{e}} + d_{\mathrm{w}} + d_{\mathrm{n}} + d_{\mathrm{s}}}$$

Dissertation O.Nemethova

Concealment Methods

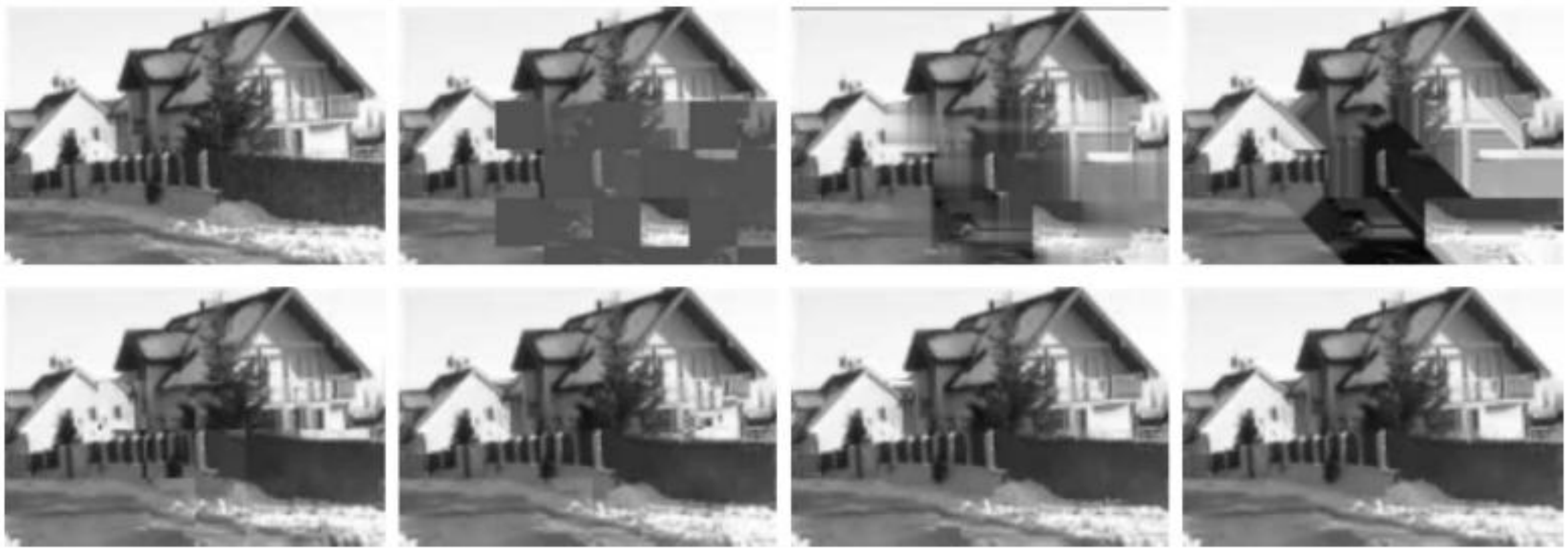


Figure 5.6 Screenshots of a part of an I frame in the 'panorama' sequence: compressed original (Y-PSNR= 35.86 dB), error pattern (Y-PSNR= 10.45 dB), weighted averaging (Y-PSNR= 18.09 dB), directional interpolation (Y-PSNR= 16.57 dB), copy-paste (Y-PSNR= 22.76 dB), boundary matching (Y-PSNR= 26.27 dB), 8 × 8 block matching with (Y-PSNR= 30.27 dB), 2 × 2 block matching with (Y-PSNR= 30.74 dB).

Dissertation O.Nemethova

Concealment Methods

Method/case	1	2	3	4	5	6	7
Weighted averaging	31.71	41.52	36.86	36.03	37.63	40.08	37.52
Maximal smoothness FOD-US	41.90	41.66	36.71	36.81	37.61	38.41	**37.83**
Maximal smoothness FOD-EA	43.46	40.81	37.40	**37.16**	38.28	40.42	37.53
Maximal smoothness SOD-US	42.21	40.43	36.26	36.79	37.74	39.60	37.49
Directional interpolation	42.60	30.26	27.32	22.30	21.27	41.55	35.51
Segmented dir. int.	42.60	30.26	38.12	28.87	23.31	43.28	34.87
POCS, WA, 5 it.	42.85	33.28	28.33	21.25	23.39	40.86	35.95
POCS, dir. int. initial, 5 it.	42.86	31.12	**38.39**	20.30	22.41	40.92	35.03
Copy-paste	40.41	38.17	36.93	32.67	39.09	53.62	30.86
MV interpolation	43.56	43.94	35.58	28.16	44.32	48.06	30.12
MV interpolation SM	43.92	44.72	35.77	29.82	47.74	57.46	30.75
Boundary matching	43.92	44.28	37.02	33.21	47.86	57.46	30.68
Block matching	44.14	**47.65**	36.86	33.21	**48.39**	57.46	30.75
Model based (PCA)	**45.24**	41.11	37.72	25.53	43.21	**59.88**	31.41

Dissertation O.Nemethova

Adaptive Method Selection

```
if scene change
   if clear edges AND enough neighbours
      method = directional interpolation
   else
      method = weighted averaging
else
   if I frame
      method = block matching
   else
      if MV correct
         method = decod without residuals
      else
         method = MV interpolation
```

Dissertation O.Nemethova

Video Transmission

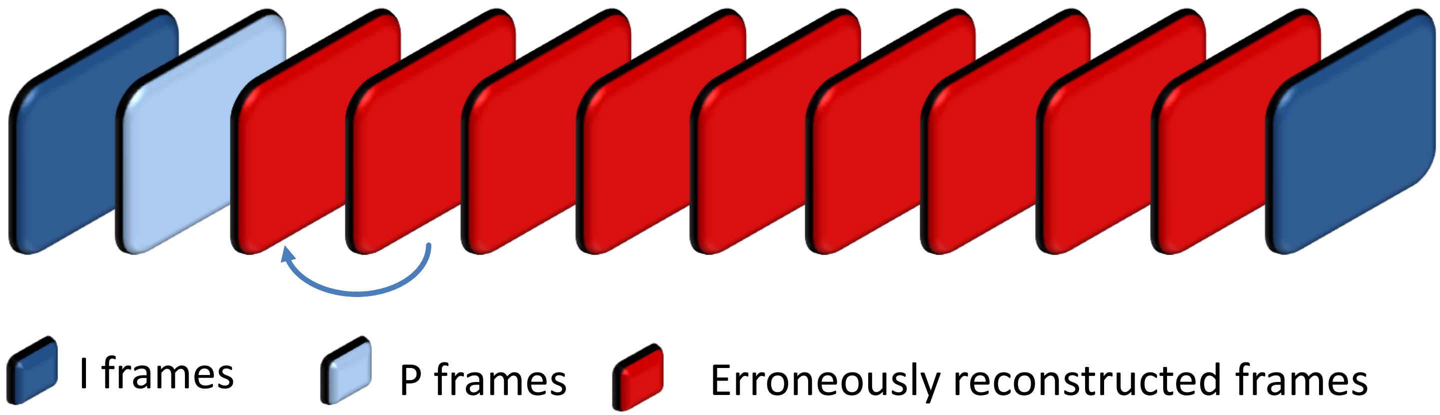
- All the information needed for reconstructing the frame are stored
  - Type of encoding
  - Element used for prediction
  - Coefficients
  - ...

- The bitstream has to be segmented into smaller chunks (packets)

IP	UDP	RTP	Video Payload

- Each data chunk is further encapsulated into a protocol stack

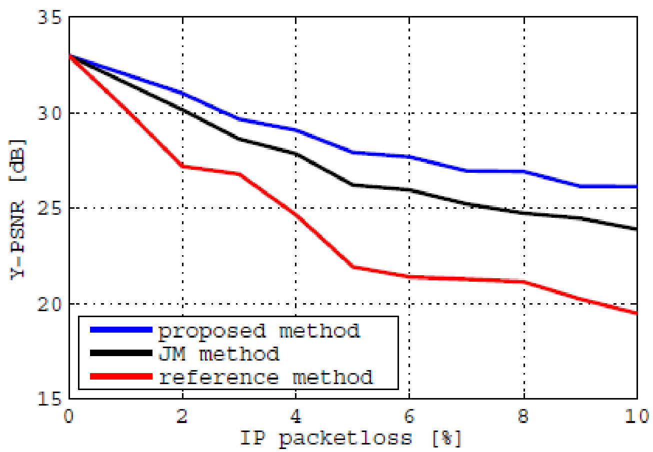
Effect of Errors at Sequence Level



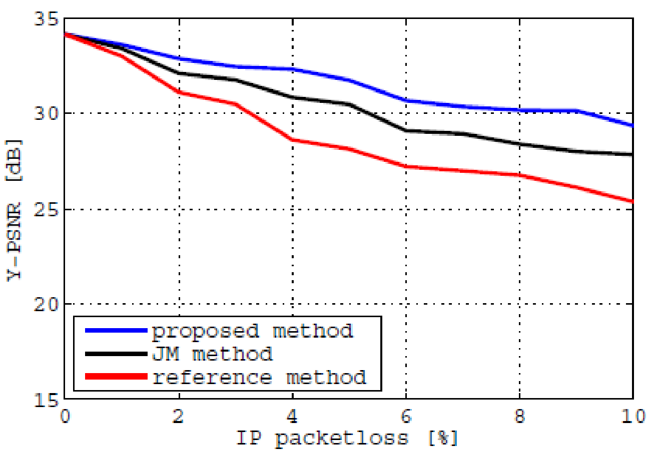
- If one packet is corrupted, the picture is incorrectly reconstructed
- The following pictures are using the damaged frame for temporal prediction
  - Even though their packets are correctly received, the corresponding frames are incorrectly reconstructed
- This effect, temporal error propagation, lasts until the following Intra frame

Comparison

- Reference: weighted averaging (spatial) and copy paste (temporal)
- JM: implemented in Joint Model software



(a) 'mobile'



(b) 'soccer'

Dissertation O.Nemethova

Further Reading

- Ming Liou, “Overview of the 64 kbit/s video coding standard,” Com. of the ACM, vol. 34, no. 4, pp. 59-63, April 1991.
- D. LeGall, “MPEG: a video compression standard for multimedia applications,” Com. of the ACM, vol. 34, no. 4, pp. 46-58, April 1991.
- IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on the H.264/JVC Video Coding Standard, July 2003.
- V. K. Goyal, “Theoretical foundations of transform coding,”IEEE Signal Processing Magazine, vol. 18, no. 5, pp. 9-21, Sept. 2001
- W.-H. Chen, W. Pratt, “Scene Adaptive Coder,”IEEE Transactions on Communications, vol. 32, no. 3, pp. 225-232, March 1984.
- E. Y. Lam, J. W. Goodman, “A Mathematical Analysis of the DCT Coefficient Distributions for Images,” IEEE Transactions on Image Processing, vol. 9, no. 10, pp. 1661-1666, October 2000.
- O.Nemethova, Error Resilient Transmission of Video Streaming over Wireless Mobile Networks, PhD thesis, 2007.