b

UNIVERSITÄT
BERN

UNIVERSITÄT BERN

Advanced Networking and Future Internet

X. Image and Video Compression

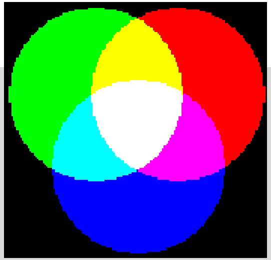
Prof. Dr. Torsten Braun, Institut für Informatik

Bern, 16.11.2020

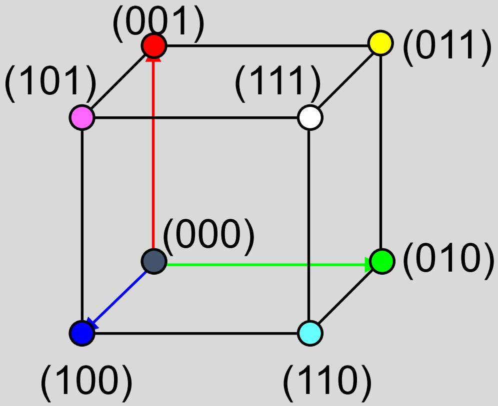
Advanced Networking and Future Internet: Image and Video Compression

Table of Contents

1. Image Compression



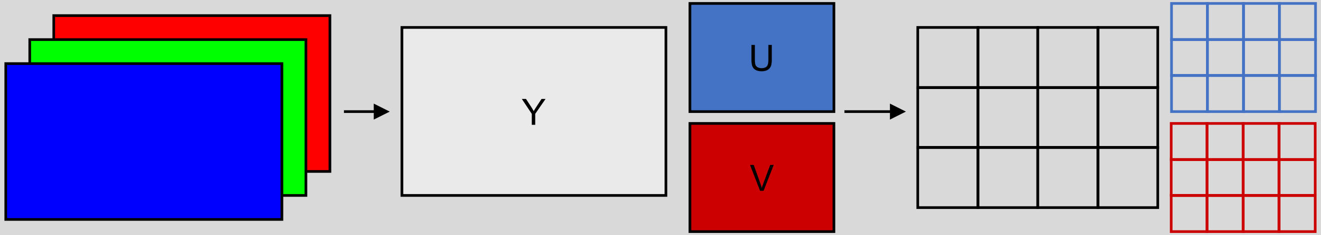
- Image consists of pixels (picture element)
- Sampling and quantization
- Representation of brightness or color information of a pixel by integers
  - Black / White: 0 or 1
  - Grey scale images: typically 8 bits:
    00 (black) – FF (white)
  - Color images: 3 numbers for red, green and blue
    $\rightarrow$ 24 bits



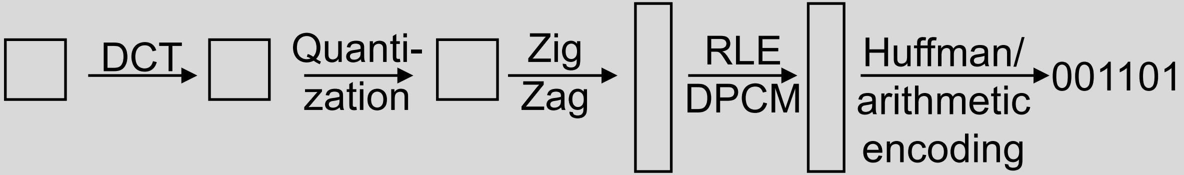
1. Image Compression

1. Joint Photographic Experts Group

Basic Mode:



for each block:



DPCM: Differential PCM, RLE: Run Length Encoding, DCT: Discrete Cosine Transform

UNIVERSITÄT BERN

1. Image Compression

2. YUV Color Encoding

- Compatible presentation for black/white and color TV required
- Separation of brightness (luminance, Y) and color information (chrominance, U/V)
  - Y = 0.299R + 0.587G + 0.114B
  - U = 0.493 (B-Y), blue chrominance
  - V = 0.877 (R-Y), red chrominance
- Human perception reacts more sensitively to brightness than to color information.
  → luminance encoding with higher bandwidth (sampling rate)

	Y	U	V
black	0	0	0
blue	0.11	0.439	-0.096
red	0.30	-0.148	0.614
magenta	0.41	0.291	0.517
green	0.59	-0.291	-0.517
cyan	0.70	0.148	-0.614
yellow	0.89	-0.439	0.096
white	1	0	0

UNIVERSITÄT BERN

1. Image Compression

3. DCT for Images

- Separation of an image into 8x8 pixel blocks with 8 bits per pixel [-128,127]
- Pixel values S_{xy}
- Transformation of each block into the frequency domain by DCT $\rightarrow$ S_{uv} (DCT coefficients)
- DC coefficient S_{00}: basic color tone
- AC coefficients: S_{ij} (i or j $\neq$ 0)
  - S_{07}: highest frequency in vertical direction
  - S_{70}: highest frequency in horizontal direction
  - S_{77}: highest frequency in both directions

1. Image Compression

3.1 Discrete Cosine Transformation

Basis functions b(x,y)

$$s_{uv} = \frac{1}{4} C_u C_v \sum_{x=0}^{7} \sum_{y=0}^{7} s_{xy} \boxed{\cos\frac{(2x+1)u\pi}{16} \cos\frac{(2y+1)v\pi}{16}},$$

$$C_u, C_v = \frac{1}{\sqrt{2}} \; for \;\; u,v=0;$$

$$C_u, C_v = 1 \;\; otherwise$$

Forward DCT

$$s_{xy} = \frac{1}{4} \sum_{u=0}^{7} \sum_{v=0}^{7} C_u C_v s_{uv} \cos\frac{(2x+1)u\pi}{16} \cos\frac{(2y+1)v\pi}{16},$$

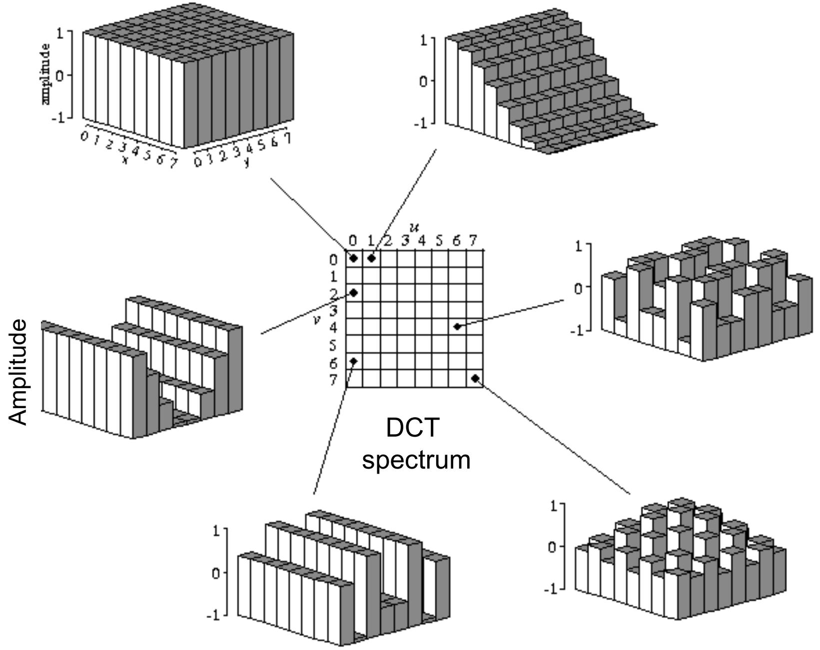
$$C_u, C_v = \frac{1}{\sqrt{2}} \; for \;\; u,v=0;$$

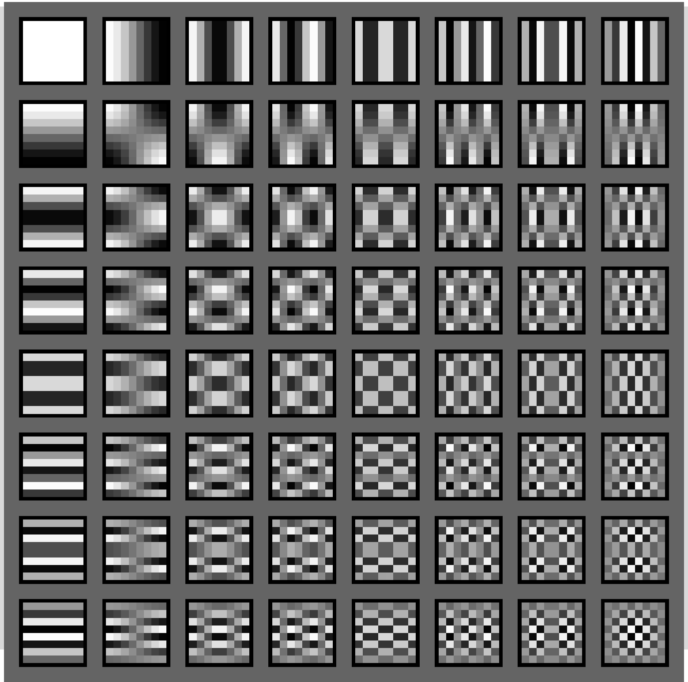
$$C_u, C_v = 1 \;\; otherwise$$

Inverse DCT

1. Image Compression

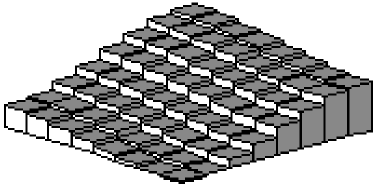
3.2 DCT Basis Functions



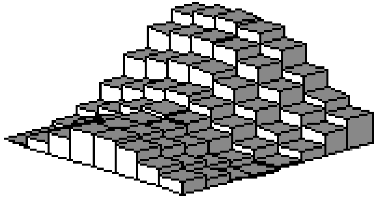


1. Image Compression

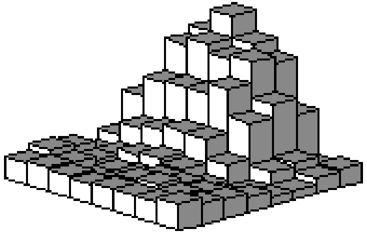
3.3 DCT Reconstruction with Low Frequencies



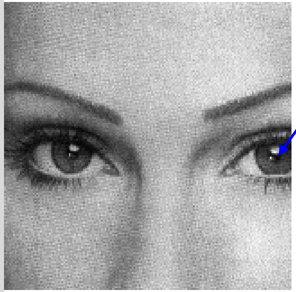
3 coefficients



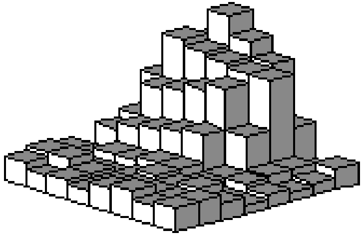
6 coefficients



15 coefficients



example: eye region



64 coefficients (correct image)

1. Image Compression

4.1 Quantization

- Quantization: $K_{uv} = \lfloor S_{uv}/Q_{uv} + 0.5 \rfloor$
  - Quantization table Q_{uv} with 64 entries
  - Courser quantization for higher frequencies (Human eye is more sensitive to lower frequencies.)
- Inverse quantization: $R_{uv} = K_{uv} * Q_{uv}$
- Typical quantization tables Q_{uv}:

Luminance

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

Chrominance

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

1. Image Compression

4.2 Example: Quantization

120	60	40	30	4	3	0	0
70	48	32	3	4	1	0	0
50	36	4	4	2	0	0	0
40	4	5	1	1	0	0	0
5	4	0	0	0	0	0	0
3	2	0	0	0	0	0	0
1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Quantization

10	10	15	20	25	30	35	40
10	15	20	25	30	35	40	50
15	20	25	30	35	40	50	60
20	25	30	35	40	50	60	70
25	30	35	40	50	60	70	80
30	35	40	50	60	70	80	90
35	40	50	60	70	80	90	100
40	50	60	70	80	90	100	110

12	6	3	2	0	0	0	0
7	3	2	0	0	0	0	0
3	2	0	0	0	0	0	0
2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

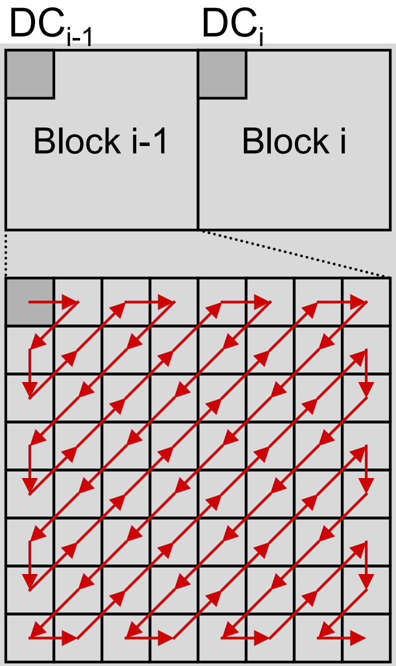
UNIVERSITÄT BERN

1. Image Compression

5. Entropy Encoding

- DC coefficients
  - Describe primary block color
  - Encoding of (usually) small differences between blocks (DPCM)
- AC coefficients
  - Zig zag sequence
  - Coefficients of lower frequencies have larger values.
  - Coefficients of higher frequencies have lower values.

  → Longer sequences of similar values allow good entropy encoding (RLE, Huffman, arithmetic).



1. Image Compression

6. Huffman Encoding

a	32	
b	29	
c	19	
d	11	
e	7	1
f	2	0

a	32	
b	29	
c	19	
d	11	1
ef	9	0

a	32	
b	29	
def	20	1
c	19	0

cdef	39	
a	32	1
b	29	0

ab	61	1
cdef	39	0

a	c	d	f	b	d	f
11	00	011	0100	10	011	0100

a	11
b	10
c	00
d	011
e	0101
f	0100

1. Image Compression

7.1 Arithmetic Encoding

- Characters are not encoded independently from each other,
  but always in relation to previous ones.
- higher efficiency than Huffman encoding, but not supporting random access.
- 1st step: estimation of frequencies of occurrence of each character
- Each string is mapped to a subinterval whose size is proportional to the symbol's occurrence probability.
- Encoding of a string by a number uniquely identifying the subinterval

Example:

- 3 symbols a, b, c with probabilities 0.4, 0.5, 0.1 divide interval [0,1) into subintervals [0,0.4), [0.4,0.9), [0.9,1).
- Next symbol divides respective interval further, e.g., for string "bb": [0.4 + 0.4 ▪ (0.9 - 0.4), 0.4 + (0.4 + 0.5) ▪ (0.9 - 0.4)) = [0.6, 0.85)

UNIVERSITÄT BERN

1. Image Compression

7.2 Example: Arithmetic Encoding

- Text: SWISS_MISS
- String is mapped to interval [0.71753375, 0.717535)
- Binary representation is: [0.10110111101100000100101010111,
  0.10110111101100000101111110110)
- Selection of number
  10110111101100000101000000000 (20bits !)

Character	Probability	Range
S	0.5	[0.5, 1.0)
W	0.1	[0.4, 0.5)
I	0.2	[0.2, 0.4)
M	0.1	[0.1, 0.2)
_	0.1	[0, 0.1)

1. Image Compression

8. Example: Lena

UNIVERSITÄT BERN

2. Video Compression

- Simple approach:
  Encode each image (= frame) using a still image compression mechanism,
  e.g., JPEG → Motion-JPEG
  (typical compression rate: 10:1)
- Typically: Redundancy between consecutive frames
- Only a portion of a frame is involved in a motion.
  - Examples:
    Eyes, lips, head, car, pedestrian

Approach

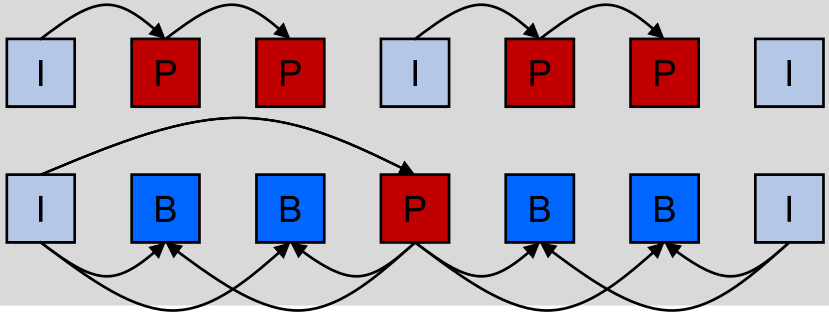
- Content prediction of frames based on previous and future frames
  → ***Motion Estimation***
- Predicted and actual positions of moving segments will slightly differ and must be described.
  → ***Motion Compensation***

2. Video Compression

1. Frame Types

Independently coded frames

– Intra-coded frames (I frames) are encoded without any reference to other frames, e.g., using JPEG for Y, C_b, C_r

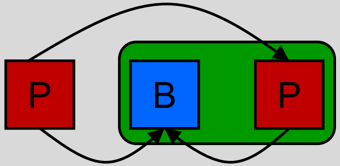


Predicted frames

– Predictive frames (P frames)

– Bidirectional frames (B frames)

– PB frames: P and B frame encoded in a single frame



b UNIVERSITÄT BERN

2. Video Compression

2. Group of Pictures

- I frames must appear periodically to correct corruptions.
- Group of pictures (GOP) = number of frames between successive I frames
- Propagation of errors from one P frame to the next → Number of P frames is limited.

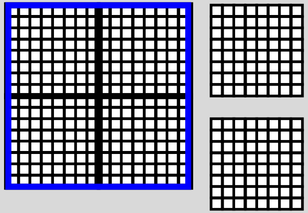
- B frames can be decoded when all (future) reference frames have been received.
  → Reordering required
  - Example: IBBPBBPBBI… → IPBBPBBIBB…

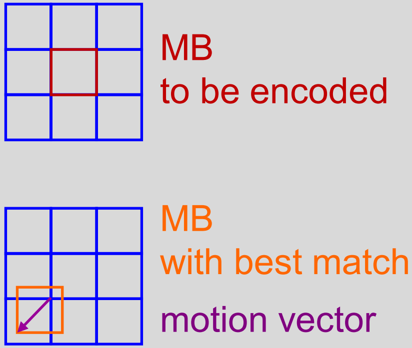
UNIVERSITÄT BERN

2. Video Compression

3. Motion Estimation

- Motion estimation is based on Macro-Blocks.
- 1 MB = 4 blocks (8x8) for Y, 1 block for Cb, 1 block for Cr
- Each MB of the target frame is compared with the corresponding MB in the preceding I or P frame (reference frame) on a pixel-by-pixel basis.
  - If a close match is found: encoding of MB address
  - If no close match is found:
    - search in a limited area around the MB in the reference frame
    - encoding of MB address, motion vector and prediction error
- Typically, only Y values are used for motion estimation.



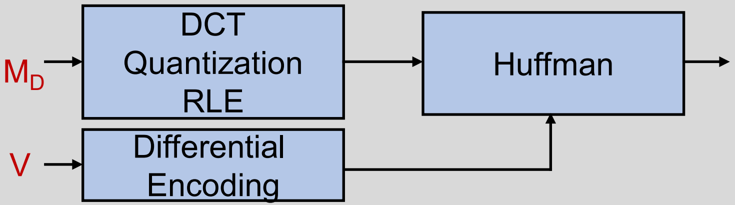


2. Video Compression

4.1 Motion Compensation in P Frames

Encoding of two parameters for best matching MB

1. Motion vector V = (x, y) based on MB or pixel boundaries
   – Most moving objects are larger than a MB. Multiple MBs are affected in the same way.
   → similar motion vectors for various MBs
   → differential encoding of motion vectors, Huffman encoding of resulting code words

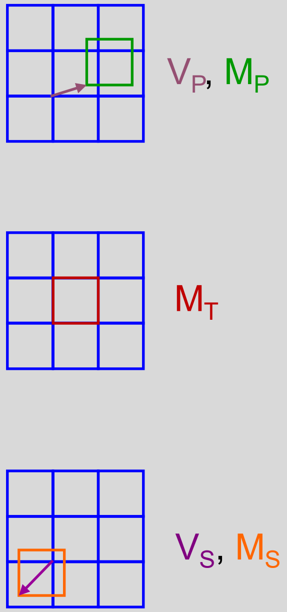


2. Prediction error comprises 3 matrices (1 for Y, C_b, C_r), each containing the difference values between the MB to be encoded (M_T) and the reference MB with best match (M_R): $M_D = M_T – M_R$

2. Video Compression

4.2 Motion Compensation in B Frames

- Motion estimation based on preceding and succeeding I or P frames
- Encoding options
  - $(M_T - M_P, V_P)$
  - $(M_T - M_S, V_S)$
  - $(M_T - \frac{1}{2}(M_P + M_S), V_P, V_S)$ (half pixel resolution)



2. Video Compression

5. H.261/H.263

H.261

- ITU standard (1993)
- Goals
  - Real-time video communications over p*64 kbps channels, p = 1-30
  - Delay by compression and decompression < 150 ms
- Typical picture formats: ratio height : width = 4 : 3
  - CIF (Common Intermediate Format)
    - 352 × 288 pixels (luminance)
    - 176 × 144 (chrominance)
  - QCIF (Quarter-CIF)
    - 176 × 144 (luminance)
    - 88 × 72 (chrominance)
- Use of I and P frames only, 3 P frames between 2 I frames

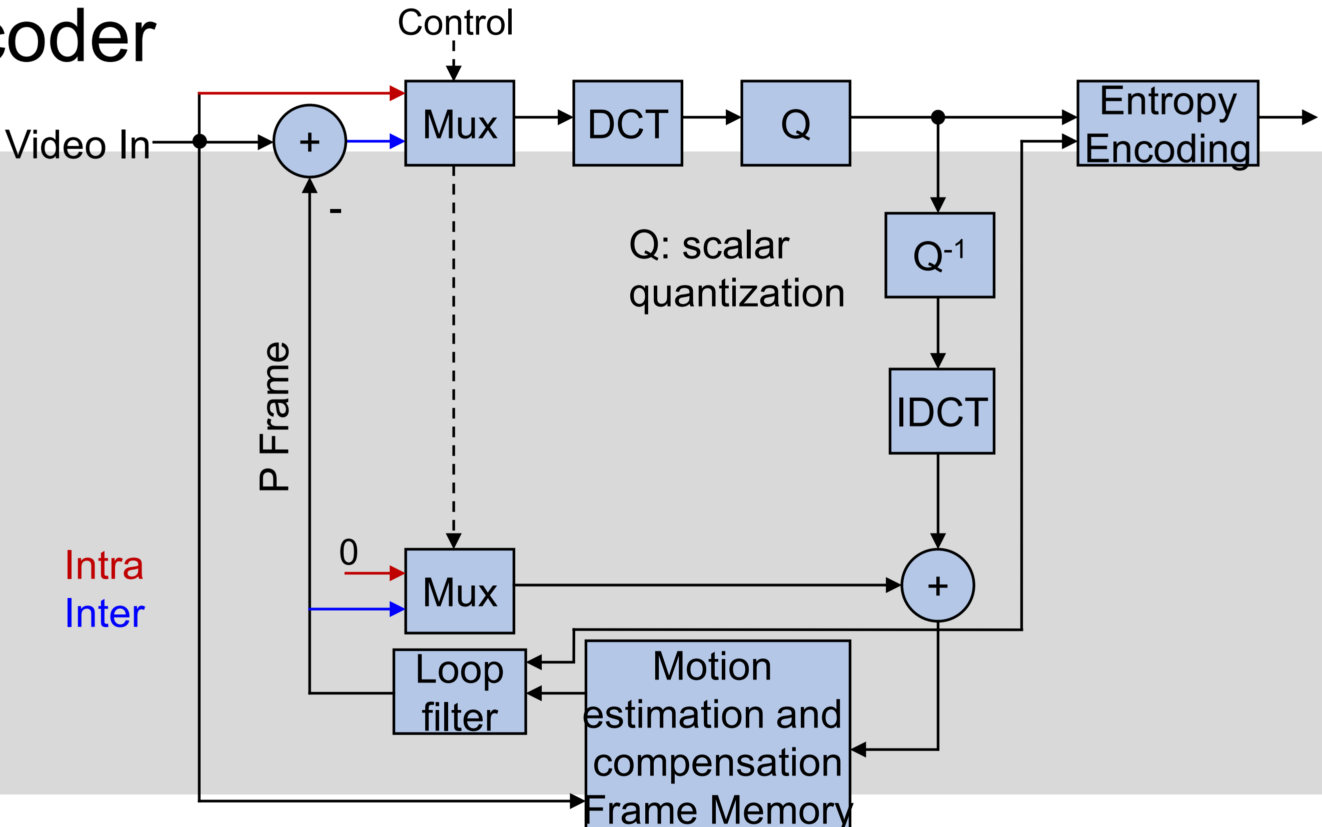
H.263

- ITU standard (1998)
- Goal
  - support of low bit rates (phone line modems and wireless networks)
- Additional resolutions:
  - sQCIF (128 × 96)
  - 4CIF(704 × 576)
  - 16CIF (1408 × 1152)
- PB frame mode
- Unrestricted Motion Vector Mode
  - Motion vectors can point outside of picture area.
  - useful for small pictures with movements towards edges
  - The next pixel inside the picture area is used, if referenced pixel is outside of the picture area.

b UNIVERSITÄT BERN

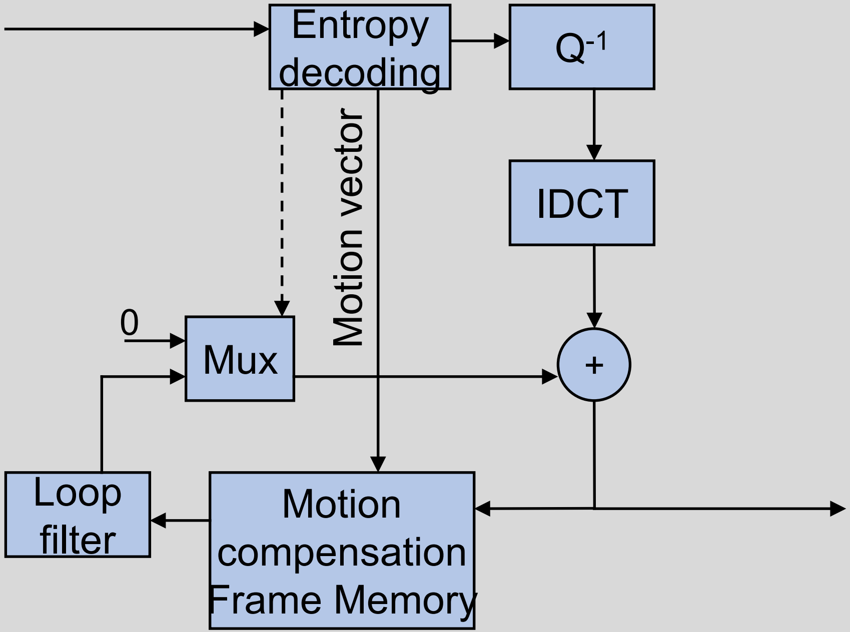
2. Video Compression

5.1 H.261 Encoder



2. Video Compression

5.2 H.261 Decoder



b UNIVERSITÄT BERN

2. Video Compression

4.6 Moving Pictures Expert Group

MPEG-1 (1992)

- Application: digital video streaming with random access, forward and backward search
- More expensive compression, simpler decompression → well-suited for stored video streaming
- Typical compression ratios: 10:1 for I, 20:1 for P, 50:1 for B
- Asymmetrical compression
- I frames
  - Recommendation: at least 1 I frame every 15 pictures
  - Complete pictures allow random access every 500 ms.
  - I frame compression is similar to JPEG, but lower compression rate due to real-time requirements.

MPEG-2 (1995)

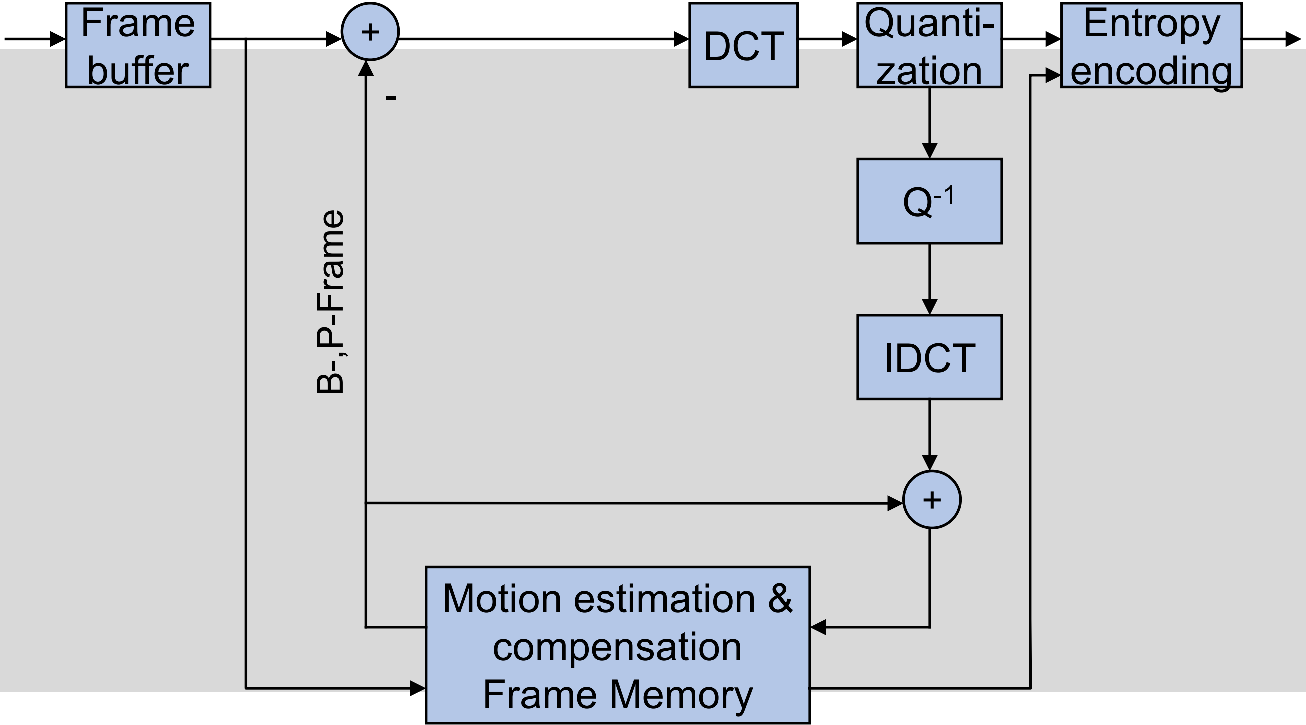
- Application: digital TV (4-100 Mbps), Digital Video Broadcast

MPEG-4 (1999)

- Interactive audio and video over wireless networks and Internet
- HDTV

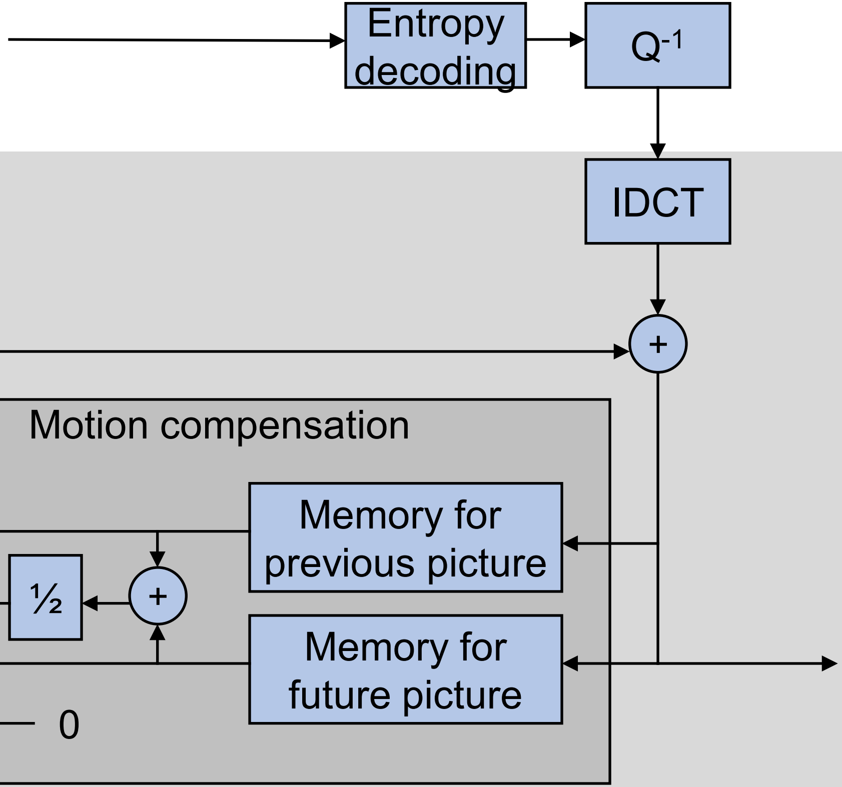
2. Video Compression

6.1.1 MPEG-1 Encoder



2. Video Compression

6.1.2 MPEG-1 Decoder



UNIVERSITÄT BERN

2. Video Compression

6.2 MPEG-2

- **Scalability**: dividing bit stream into layers
  - Reception of a few layers for minimum quality
  - Additional layers for higher quality
- Several levels up to 1920 × 1152 pixels @ 30 Hz

Scalability Schemes

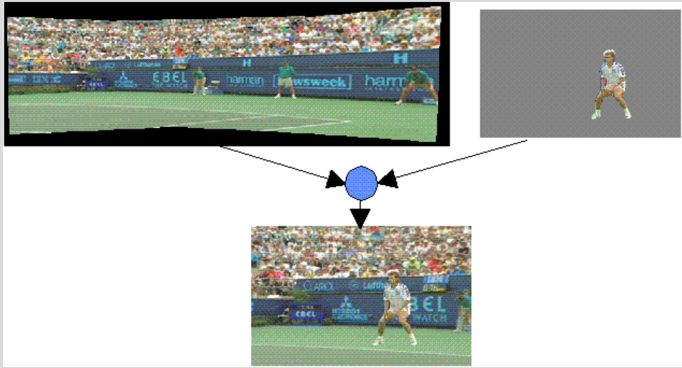
- Data partitioning
  1. partition: critical information (header, DCT coefficients of lower frequencies, motion vectors)
  2. partition: remaining information
- SNR scalability
  - Encoding in different layers with same resolutions but different quality by refinement of DCT coefficients → Transmission of layers over channels with different quality (basic layers: good quality)
- Spatial scalability
  - Layers with different spatial resolution (from H.261 to HDTV studio quality)
  - Enhancement layer can use spatially interpolated lower layer.
- Temporal scalability
  - Encoding of layers with different frame rates: Basic layer is encoded with lower rate, enhancement layers include frames for higher rates.

UNIVERSITÄT BERN

2. Video Compression

6.3 MPEG-4

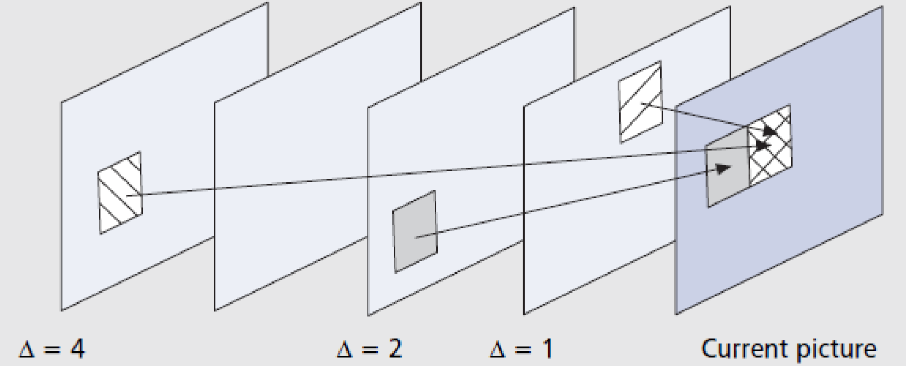
- Introduction of object-oriented video encoding: audio-visual objects
  - Examples: person, vehicle, background, graphics, animated bodies
  - Objects may constitute of sub-objects
- Object descriptions
  - to describe objects and to allow objects to be manipulated by the user before viewing
- Scene descriptions
  - to describe relations between objects of a single scene.
- Separation of video frames into video object planes; separate encoding



2. Video Compression

7. H.264/MPEG-4 Advanced Video Coding

- Joint standard of MPEG and ITU
- Applications: Videoconferencing, Digital Video Broadcast
- Better efficiency (bit rate: 40 % lower than for MPEG-2) by many small improvements over MPEG-1/2
  - Smaller macro-blocks (4 x 4 pixels) possible
  - Association of MBs to slices with independent coding (I, P, B)
  - Motion vectors with ¼ pixel accuracy
  - Multiple reference pictures
  - Weighted averaging of reference pictures
  - Spatial and temporal prediction
  - Motion vectors over picture boundaries
  - Improved entropy encoding



UNIVERSITÄT BERN

2. Video Compression

7.1 H.264/AVC Scalable Video Coding

Scalability

= removal of parts of the video bit stream to adapt to needs / preferences of end users, terminal capabilities, network conditions

- In a multicast scenario, terminals with different capabilities can be served from a single source.
- Stronger protection for more important parts of the bit stream
- Example application: surveillance with different devices

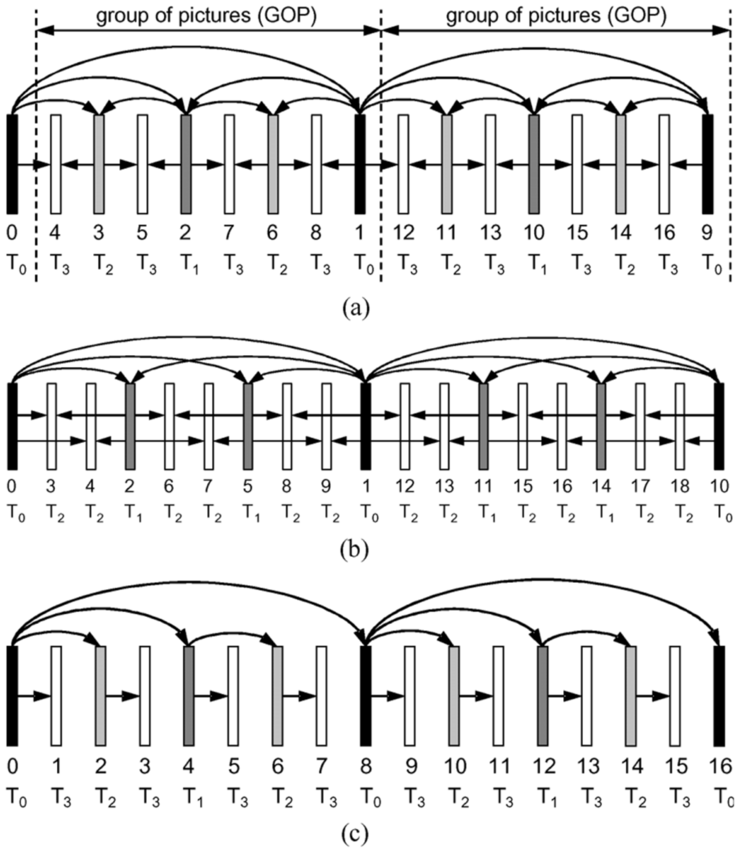
Scalability schemes

1. Quality scalability
   - Enhancement layers with smaller quantization steps
2. Temporal scalability
3. Spatial scalability

2. Video Compression

7.2 Temporal Scalability

- Temporal base layer and one or more enhancement layers (encoded as B frames)
- Motion-compensation prediction with references to pictures in layers with equal or lower identifier
  - a) Temporal layers T_0-T_4
  - b) Temporal layers T_0-T_2
  - c) Temporal layers T_0-T_3 without structural delay

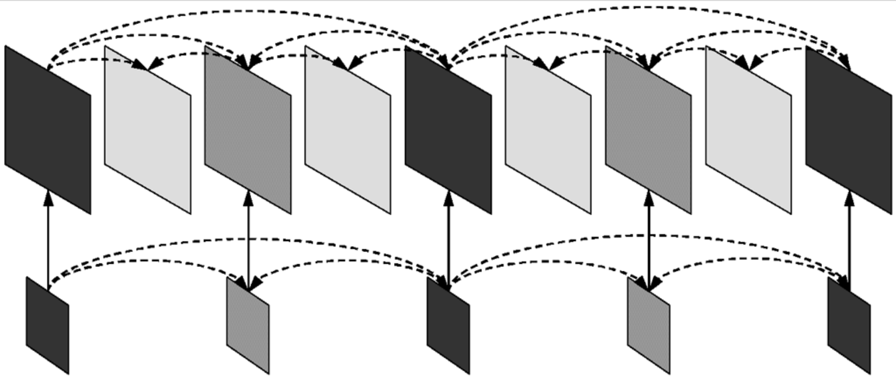


UNIVERSITÄT BERN

2. Video Compression

7.3 Spatial Scalability

- Different layers with different spatial resolution
- in each layer: motion-compensated prediction and intra-layer prediction
- Inter-layer prediction to improve efficiency

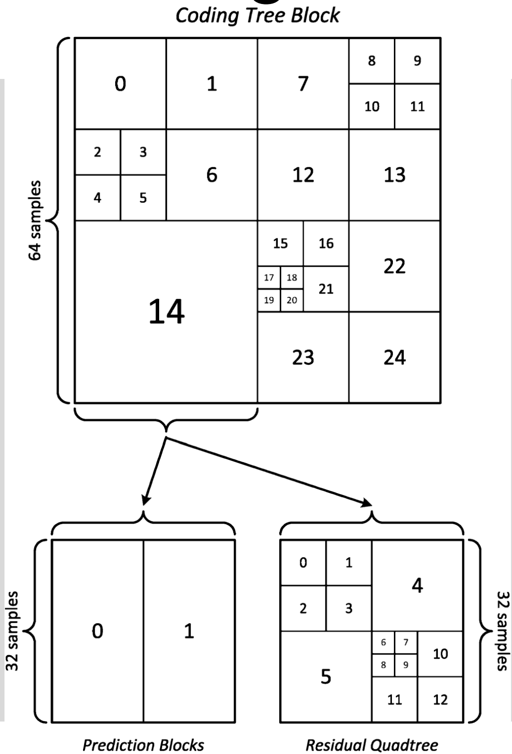


b UNIVERSITÄT BERN

2. Video Compression

8. H.265/MPEG-H High Efficiency Video Coding

- HEVC can encode video files twice as efficiently as H.264/MPEG-4.
- Block-based spatial and temporal prediction with transform coding of the prediction error (residual)
- Flexible partitioning of pictures into different size N x N coding tree blocks
- Coding Blocks as a leaf of the coding quadtree is further subdivided into Prediction Blocks (≦ 64 x 64 pixels) and Transform Blocks
- Higher precision of motion vectors



Thanks

for Your Attention

UNIVERSITÄT
BERN

Prof. Dr. Torsten Braun, Institut für Informatik

Bern, 16.11.2020