Unit 8: Video Encoding

Jun.-Prof. Dr. rer. nat. Marc Ritter

Juniorprofessur
MEDIA COMPUTING

TECHNISCHE UNIVERSITÄT CHEMNITZ

localizeit

funded by

Bundesministerium
für Bildung
und Forschung

INTENTA
ADVANCED RECOGNITION COMPONENTS

3D MICROMAC

3DInsight.de
Your Visualization Partner

IBS
IBS Software & Research GmbH

MPEG

- Moving Pictures Experts Group, active since 1988
- Official name: ISO/IEC JTC1/SC29/WG11 (International Organization for Standardization/International Electrotechnical Commission, Joint Technical Committee 1, Subcommittee 29, Working Group 11)
- Consists of more than 350 professionals from 200 companies in over 20 countries
- 3+ annual meetings, 5 days each
- Closely linked to the ITU (International Telecommunication Union)

The MPEG standards family

The MPEG standards family

MPEG-1 and MPEG-2: Standardization of audiovisual content. First standard to:

- Integrate audio and video into one solution
- Describe video encoding independent of video format
- Commonly developed by all involved industry branches
- Developed as Software package containing a complete implementation

MPEG-3:

- Originally planed to be used for HDTV: 1920x1080 px, Interlace, 30 fps, 20-40 Mbit/s
- Work was discontinued, since quality of MPEG-2 was sufficient for HDTV

MPEG-4: Content representation similar to MPEG-1&2, but with extended opportunities to integrate e.g. 3D models, synthetic audio

- Objective: arbitrary multimedia contents on arbitrary platforms
- Support of a variety of media, profiles and applications

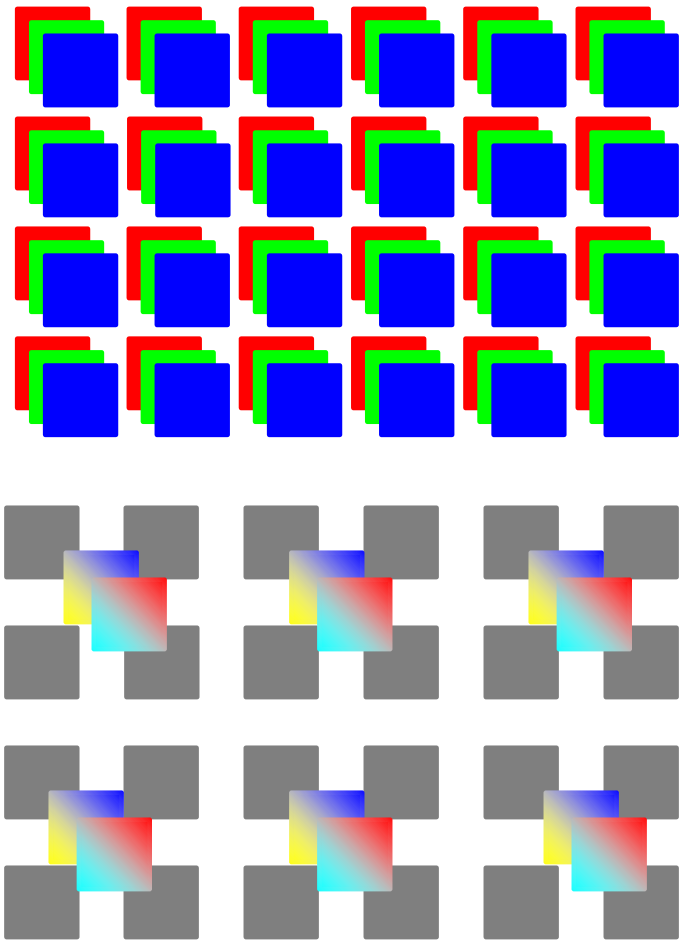
MPEG-7: Description of contents

MPEG-21: Interoperational multimedia infrastructure, especially Digital Rights Management (DRM)

MPEG-Standards

H.261

- Introduced in 1988 by the ITU
- Application: Video conferencing via ISDN
- Official ISDN bandwidths in Germany in 1989: 144 kBit/s, separated in
  - two B-channels allowing digital data transfer) à 64 kBit/s and
  - a D-channel with 16 kBit/s for steering operations
  - Join of multiple connectors possible to achieve n*64 kBit/s data transfer rates
- New technologies:
- Progressive processing (prediction over time)
- CIF- and QCIF-Frames
- Maximum delay: 150 ms
- YCbCr



(Quarter) Common Intermediate Format

Problem: Different resolutions in the most widely spread formats NTSC and PAL

NTSC:

- National Television System Committee
- 486 visible lines (often used are 480) at 29,97 Hz
- ½ NTSC at 352x240 px

PAL:

- Phase Alternating Line
- 576 visible lines at 25 Hz
- ½ PAL at 352x288 px

→ Definition of intermediate formats necessary, that are applicable to PAL and NTSC as well

Properties:

30 Hz using:

- CIF: 352x288 px (aka FCIF, Full CIF)
- QCIF: 174x144 px
- CIF uses the YCbCr color space at 4:2:0 subsampling (4:2:2 at odd lines; 4:0:0 at even lines)

Video Format	Resolution Luminance	Resolution Chrominance	Bit rate in Mbps with 30 fps
QCIF	176 x 144	88 x 72	9,1
CIF	352 x 288	176 x 144	36,5

MPEG-1 (1993)

- ISO/IEC 11172 Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s
- Part 1: Systems
  (Cor 1:1996, Cor 2:1999)
- Part 2: Video
  (Cor 1:1996, Cor 2:1999)
- Part 3: Audio (Cor 1:1996)
- Part 4 (1995): Compliance testing
- Part 5: Software simulation

- Software decoder, low quality
- Band width: 1,5 MBit/s
  - 1,25 MBit/s for video
  - Remainder for two audio channels
- Resolution:
  - Maximum 768 x 576 px
  - Europe: 360*288 px – 25 fps
  - USA: 352*240 px – 30fps
- Color coding: YCbCr (3 x 8 Bit)

MPEG-2 (ISO 13818): Generic coding of moving pictures and associated audio --- Properties and specifications

Juniorprofessur MEDIA COMPUTING
TECHNISCHE UNIVERSITÄT CHEMNITZ

- Functionality alike to MPEG-1, but with interlacing opportunity
- 720x480 px, 1280x820 px, 60 fps
- Audio in CD quality
- Hardware and software implementations
- Applicable for NTSC, HDTV, DVD

ISO/IEC 13818-1:2000
Information technology - Generic coding of moving pictures and associated audio information:
Part 1: Systems

ISO/IEC 13818-2:2000 **Part 2: Video**

ISO/IEC 13818-3:1998 **Part 3: Audio**

ISO/IEC 13818-4:1998 **Part 4: Conformance testing** (Cor 2:1998; Amd 1:1999 Advanced Audio Coding (AAC) conformance testing; Amd 2:2000 System target decoder model; Amd 3:2000 Additional audio conformance bitstreams)

ISO/IEC TR 13818-5:1997 **Part 5: Software simulation** (Amd 1:1999 Advanced Audio Coding (AAC))

ISO/IEC 13818-6:1998 **Part 6: Extensions for DSM-CC** [Digital storage media command and control] (Cor 1:1999; Amd 1:2000 Additions to support data broadcasting; Amd 2:2000 Additions to support synchronized download services, opportunistic data services and resource announcement in broadcast and interactive services)

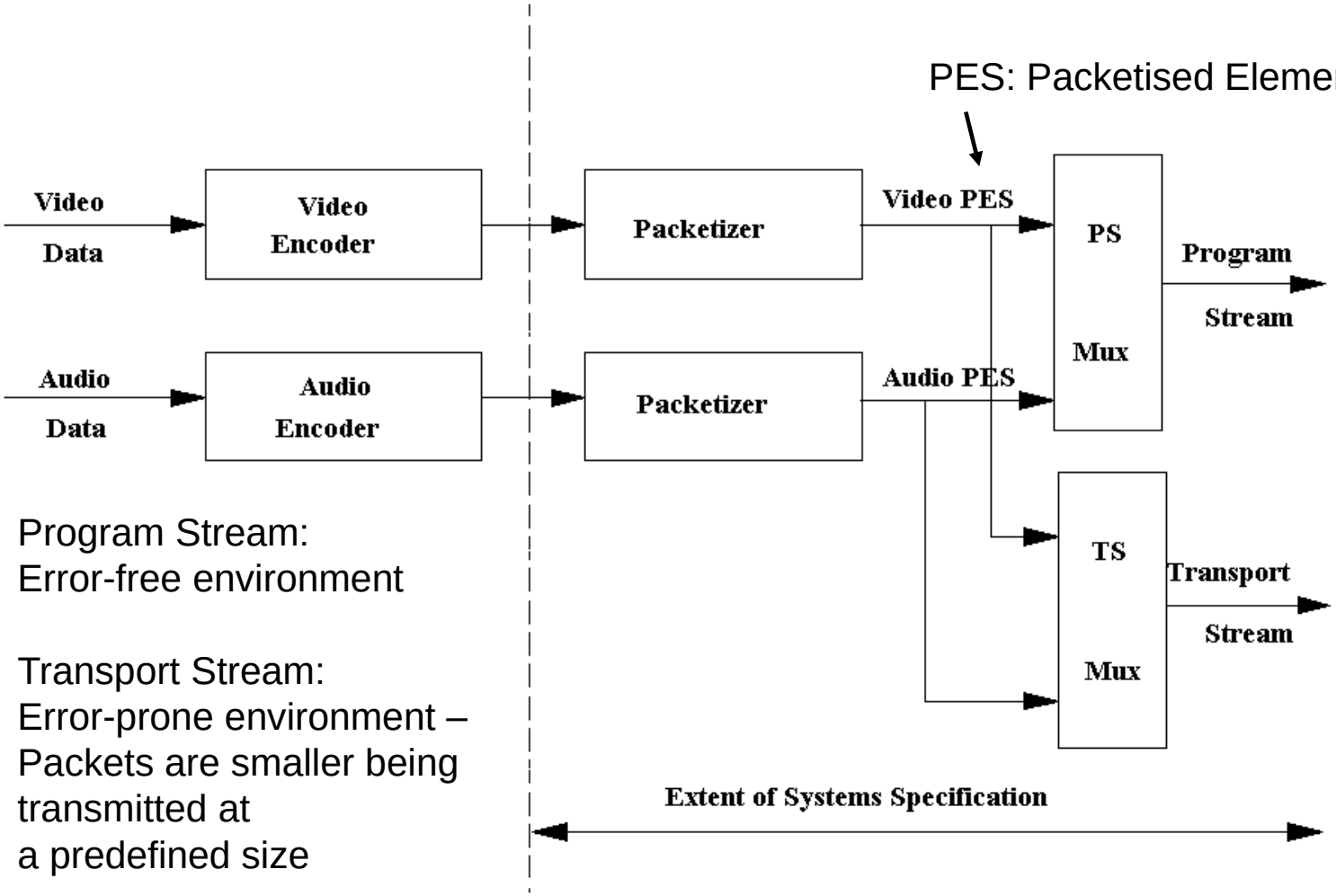
ISO/IEC 13818-7:1997 **Part 7: Advanced Audio Coding** (AAC) (Cor 1:1998)

Part 8: Discontinued, no interest from industry

ISO/IEC 13818-9:1996 **Part 9: Extension for real time interface for systems decoders**

ISO/IEC 13818-10:1999 **Part 10: Conformance extensions** for Digital Storage Media Command and Control (DSM-CC)

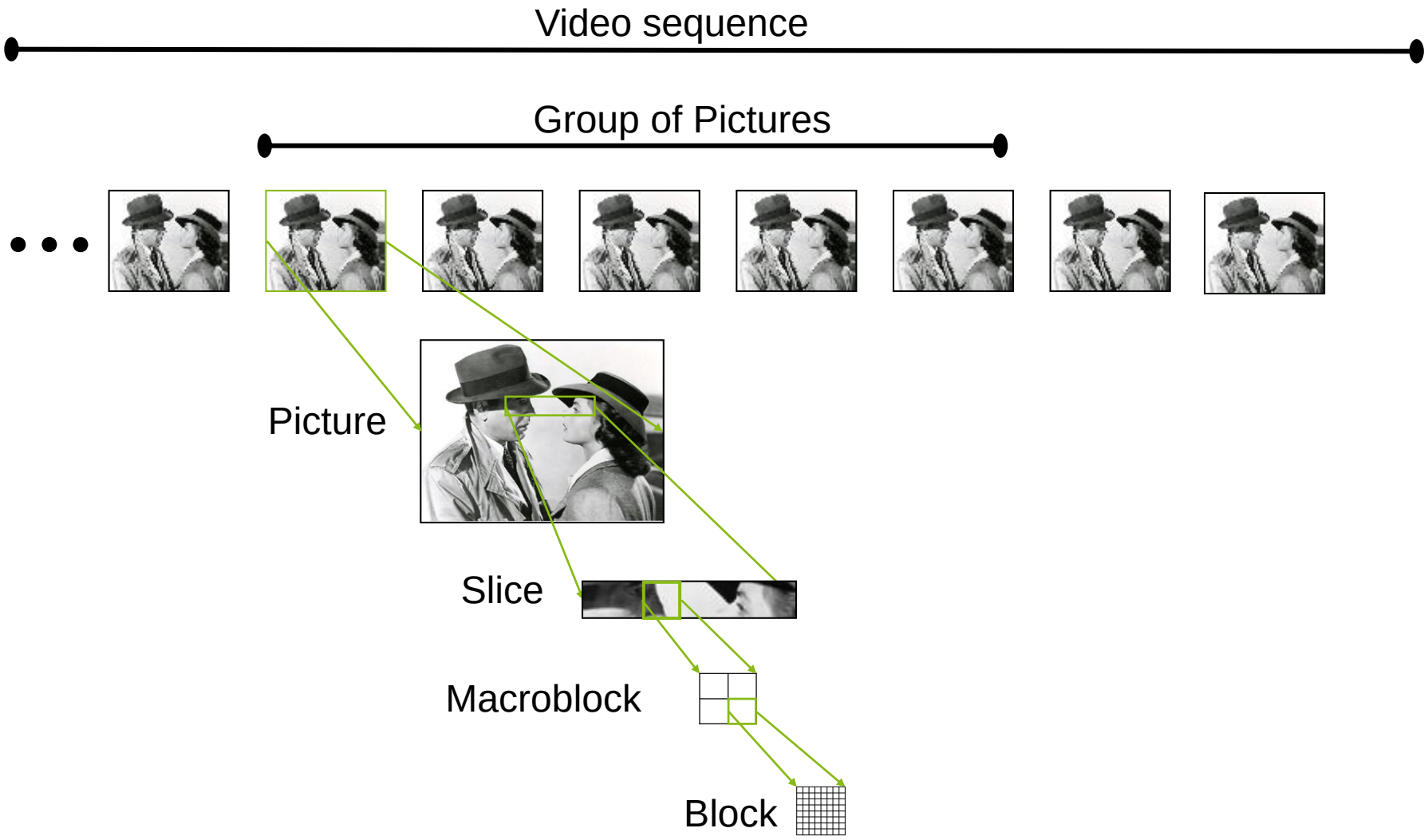
MPEG 2 (ISO 13818): Part 1 Systems



Program Stream:
Error-free environment

Transport Stream:
Error-prone environment –
Packets are smaller being transmitted at
a predefined size

Structure of a (MPEG) video data stream

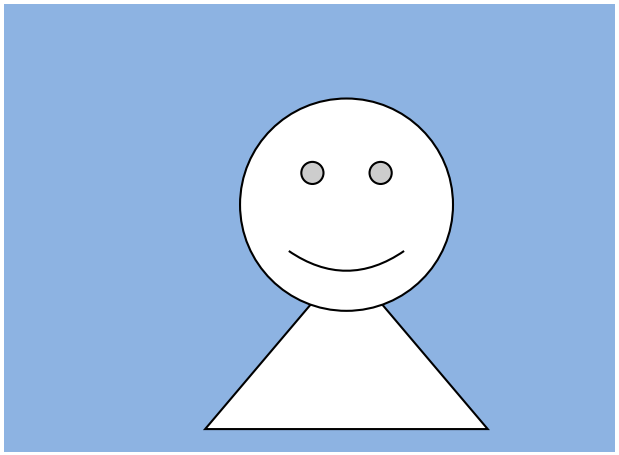


Coding with temporal prediction

- You already know about the capabilities of spatial prediction from JPEG and PNG encoding based on a neighbor basis.
- This principle is now applied to the temporal domain, i.e. code picture or image elements w.r.t. their appearance in the previous image.
- This is meaningful, especially for video conferencing. Why is this so?

Analysis of the example video conference:

- Large areas in the images do not or change only a little bit over time.
- Background is static, remains the same.
- Person movement is reduced; mouth is alternating.



Structure of H.261 encoding

H.261 realizes two types of frames for prediction:

Intra-frame (I-frame):

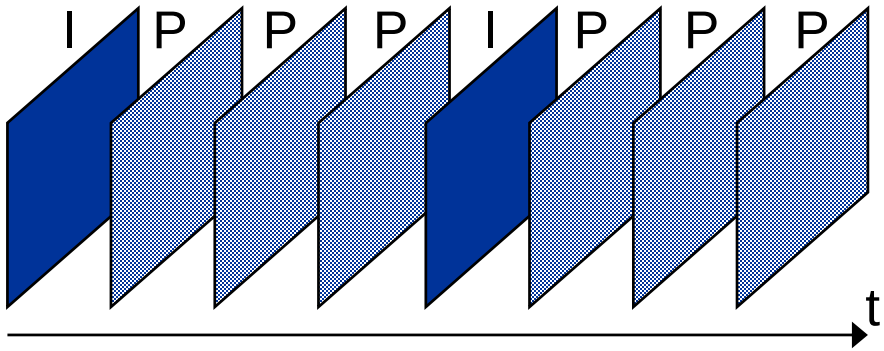
- Is a frame which uses JPEG encoding
- It uses YCbCr color space with 4x2x0 mode and consists of macro blocks.

Inter-frame (P-frame):

- Predicted from previous frame.
- Reason behind doing prediction is that there are temporal redundancies between consecutive frames.
- Also the color doesn't change much during frames.
- Thus predicted frame is derived from motion vector calculated from Mean Absolute Difference (MAD).

Method:

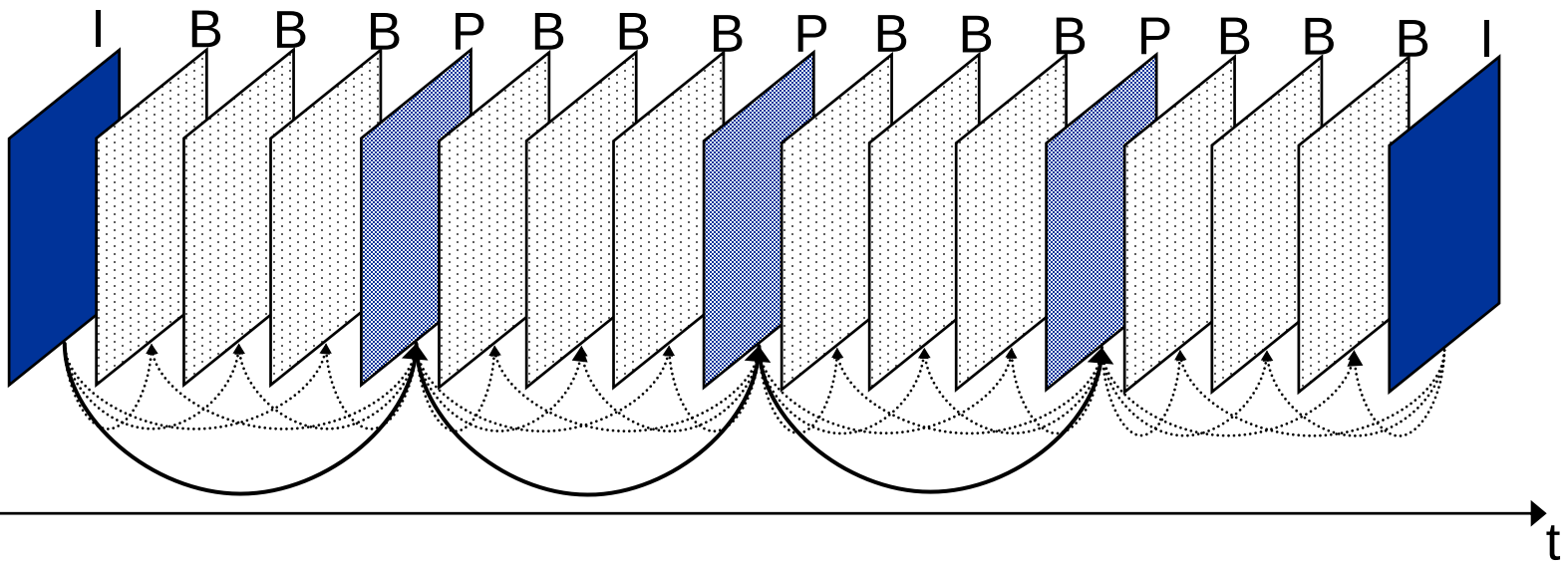
- In equal distances, I-Frames are inserted (@H.261 typically every fourth frame)
- Minimize predictive or reoccurring errors that would be endlessly repeated if only P-Frames are used.
- Allows the entries of new receivers while performing streaming applications.



Henning 2003: Figure 6.10

MPEG codec

Introduction of B-Frames

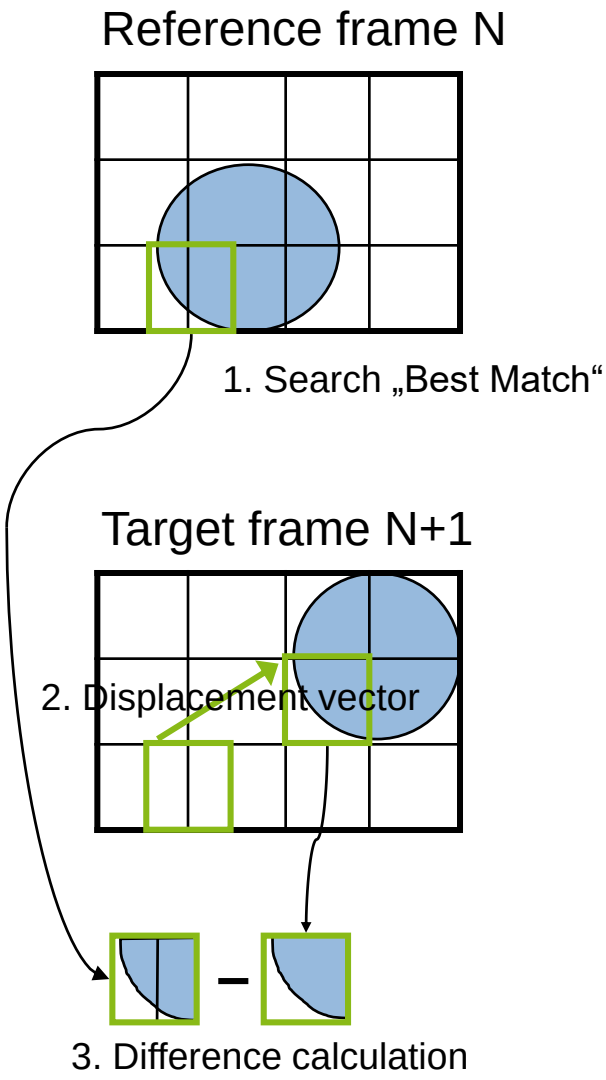


Visual appearance order																
Bild	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Typ	I	B	B	B	P	B	B	B	P	B	B	B	P	B	B	B	I
Transmission order																	
Bild	1	5	2	3	4	9	6	7	8	13	10	11	12	17	14	15	16
Typ	I	P	B	B	B	P	B	B	B	P	B	B	B	I	B	B	B

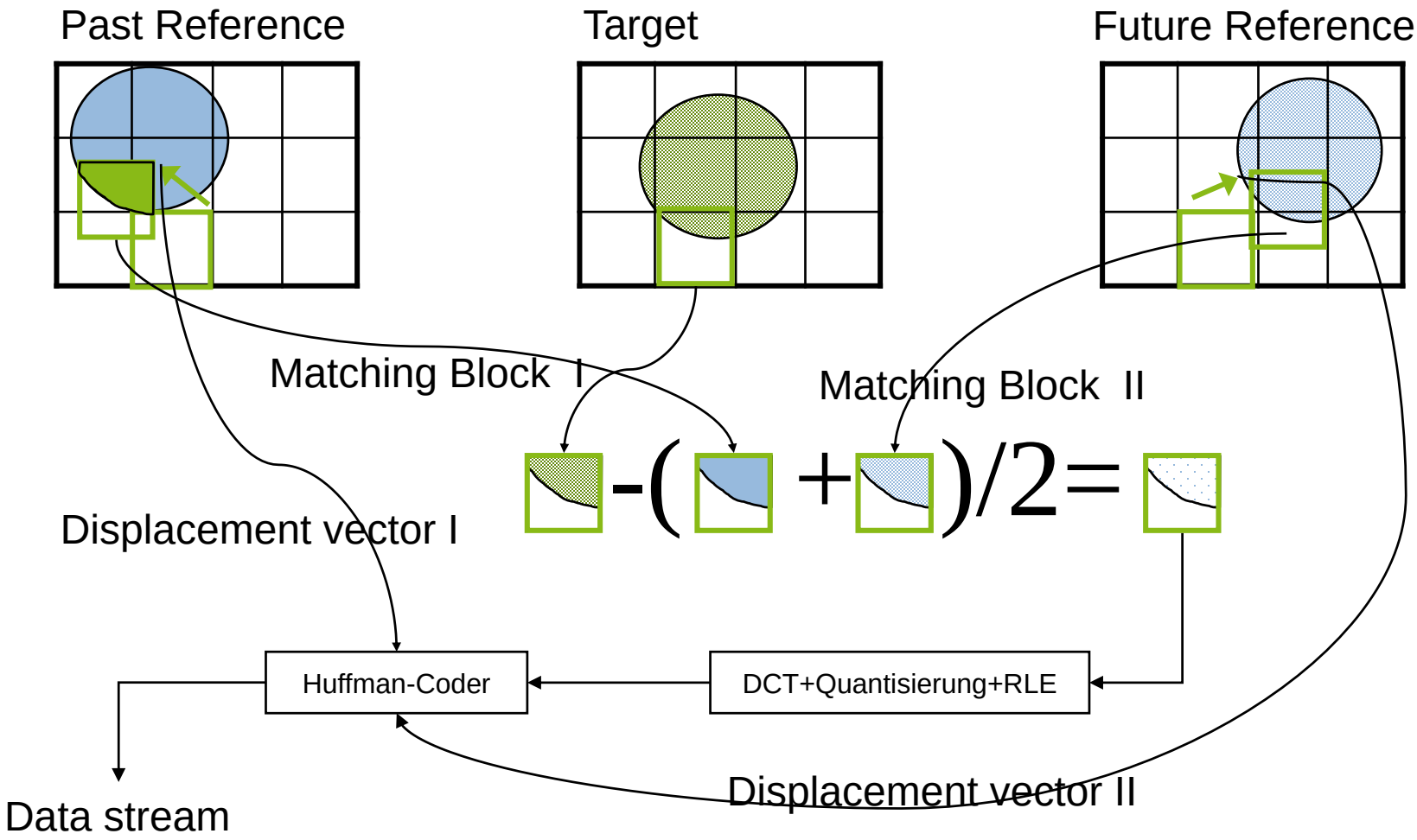
H.261 - Encoding of macroblocks by motion vectors

(i.e. movement vectors)

- Usage of macroblocks at 16x16 px
- Therein: 4 Y-and each 1 Cb + Cr blocks
- Coding of I-Frames via JPEG
- P-Frame: Need for a search of the corresponding area in the previous (reference) frame, what can be either an I-Frame or a P-Frame
- Reference matching is performed on fully “decoded” images, usually in the RGB color space
- Principle of motion vector calculation (see Figure)
- Reduced data is applied to JPEG steps: DCT, quantization and Huffman encoding

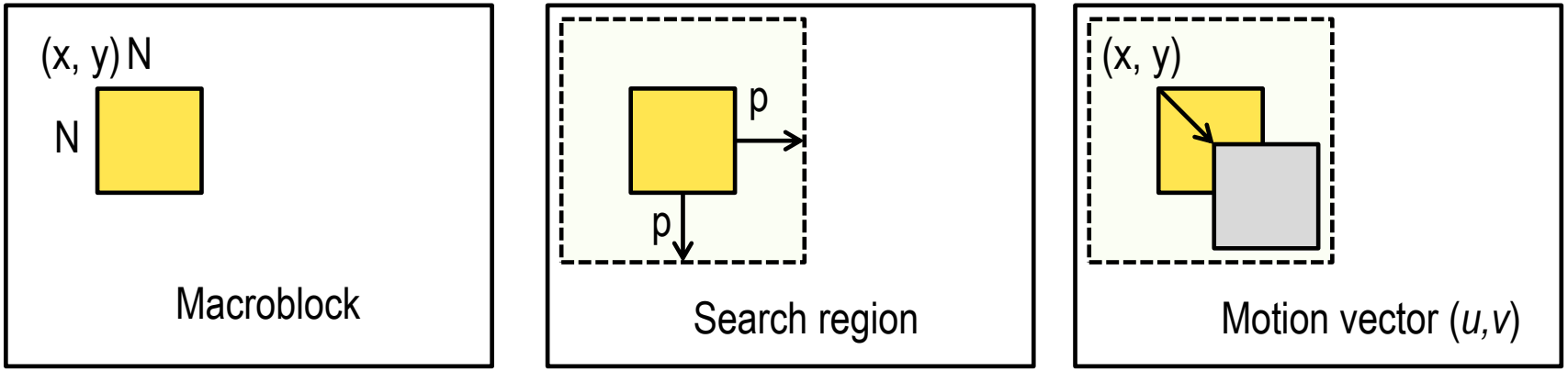


MPEG – Prediction of B-Frames

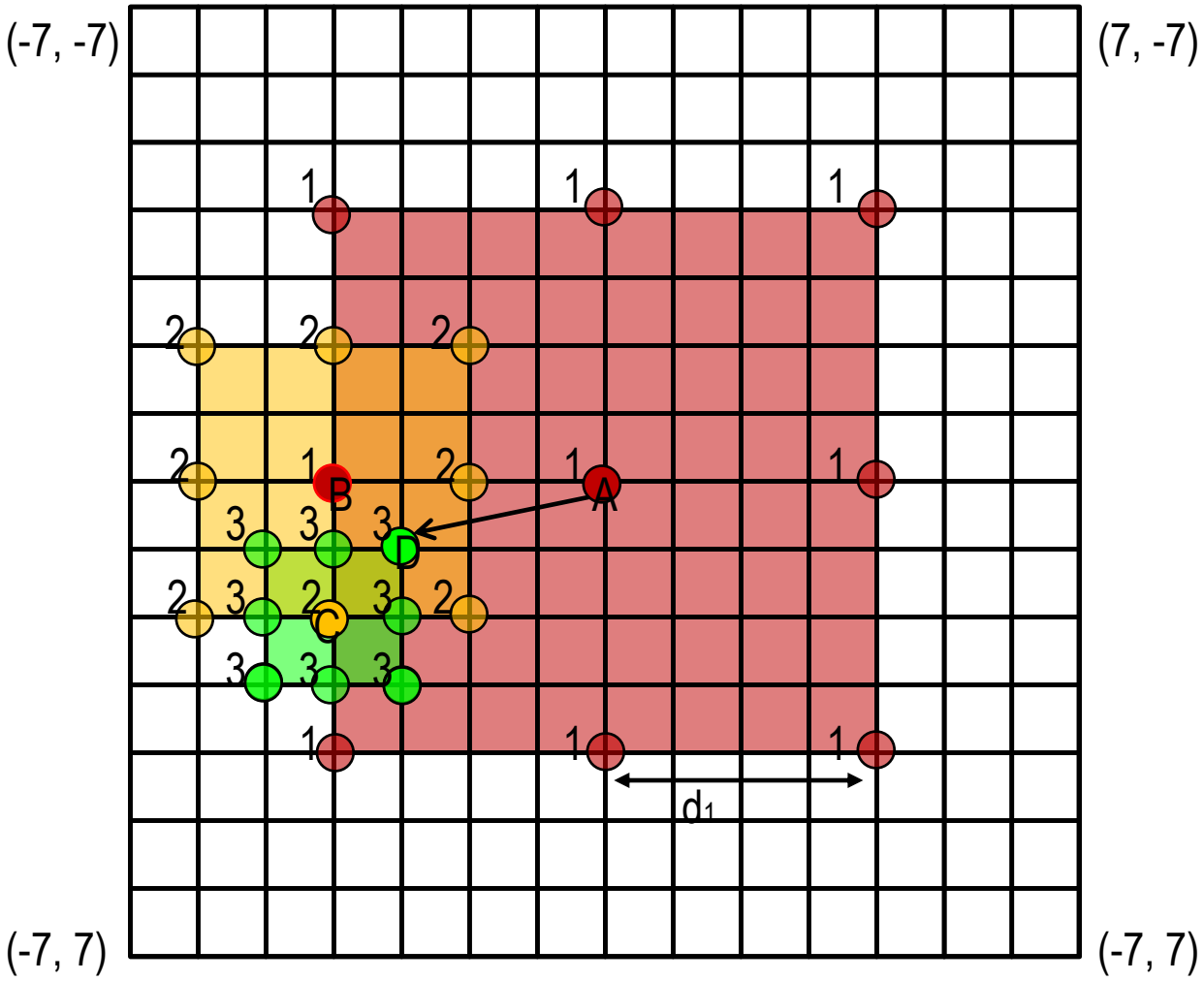


Henning 2003: Fig. 6.16

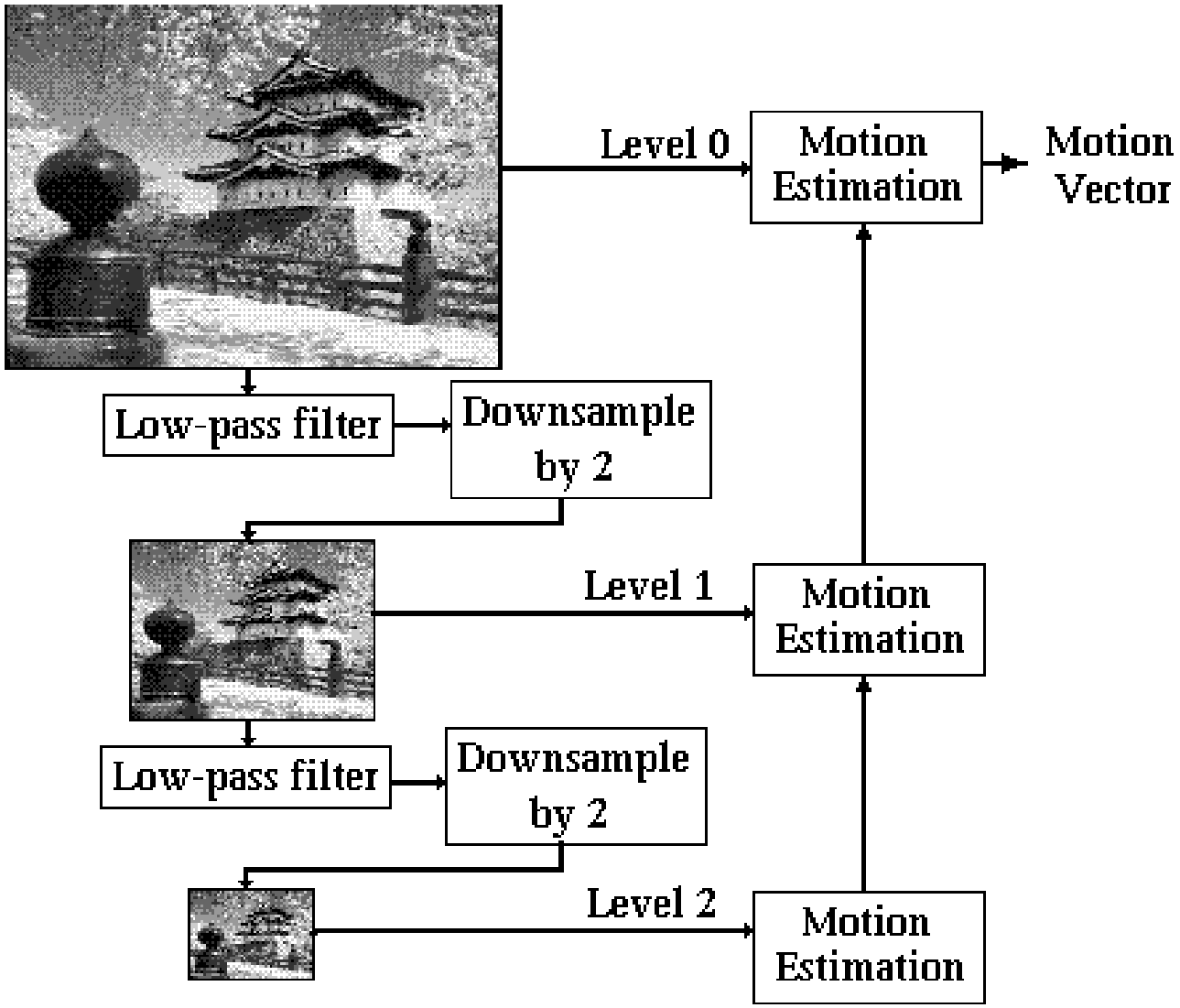
Another illustration of the principles for motion estimation



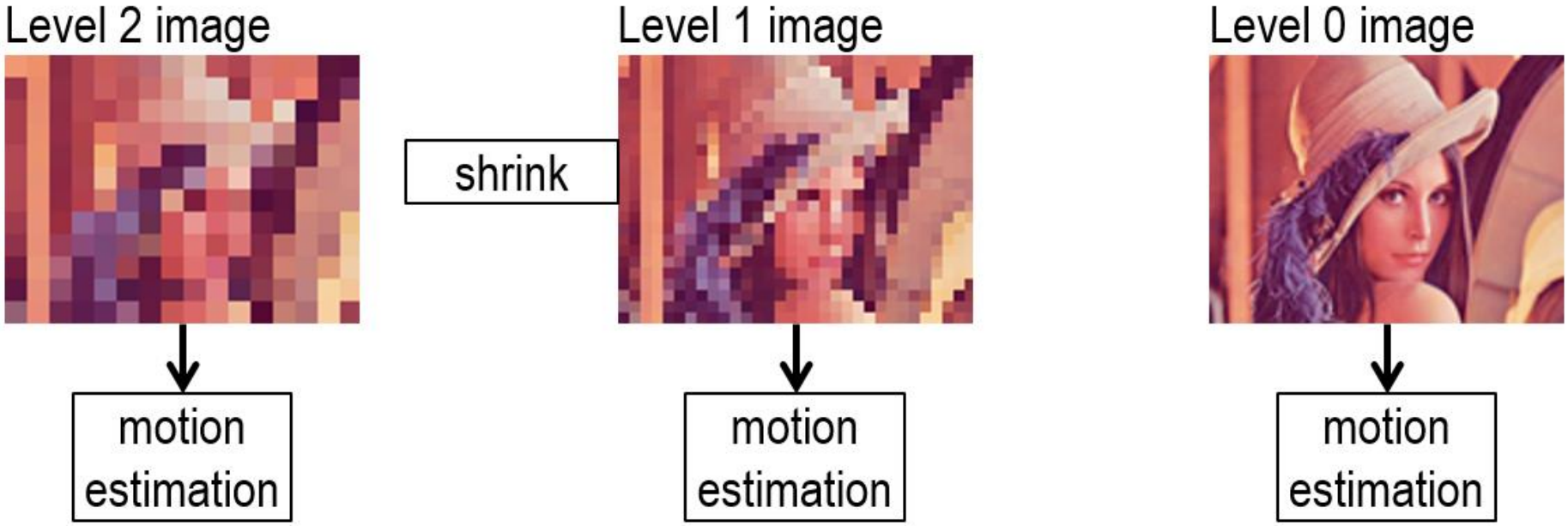
Logarithmic Search



Hierarchical Search I

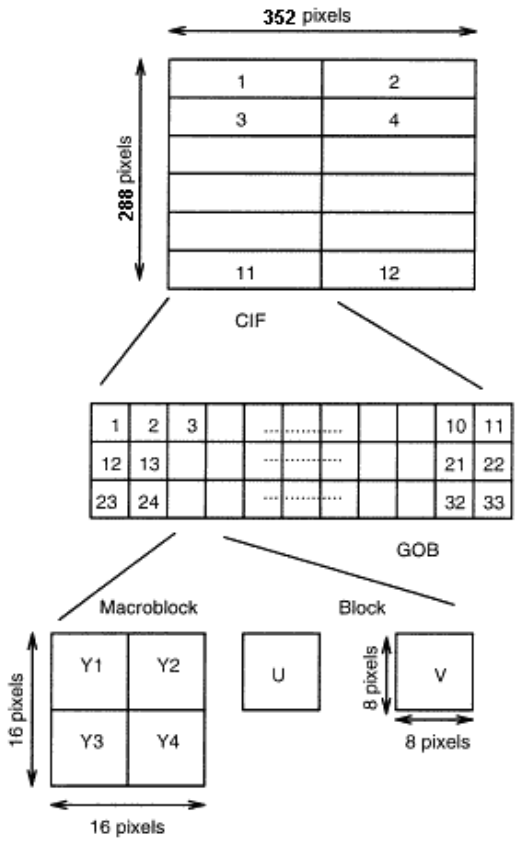


Hierarchical Search II

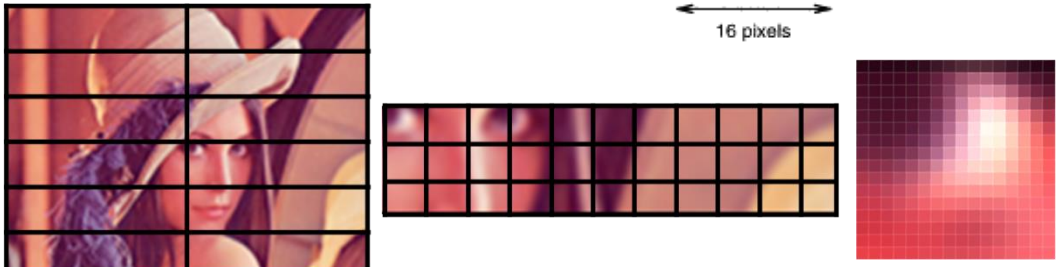


Introducing the Group of Blocks (GOB) concept

- The group of blocks are image grids of 11x3 blocks, which in turn consist of 16x16 pixels.
- The overall picture consists entirely of 2x6 GOBs. (= 352x288 pixels, see CIF)
- By dividing the GOB the encoding and playback is optimized. For example, if an error in a p-frame occurs, it is passed through until the occurrence of the next I-frame.
- Errors are reduced to the GOB.
- In addition, a data reduction is possible, since fixed areas are omitted during the encoding.



From: http://www.sciencedirect.com/science/article/pii/S0169755297001232



Structure of a H.261 Bitstream

PSC: Single image starts with the picture start code

TR: Temporal reference (aka time stamp) also in order to synchronize the image with the audio stream

PType: Picture type shows whether we have an I-Frame or a P-Frame to decode

Followed by 12 GOB with CIF or 3 with QCIF

GOB SC: Start code within a GOB

Grp#: Number of the group in the GOB (can be omitted with no changes to the previous reference image

Quant: Individual quantization table per GOB or do you prefer a global one? Make use of stronger quantization in video conferencing at the borders

MB: Macroblock with address in the image (can be omitted again when unchanged).

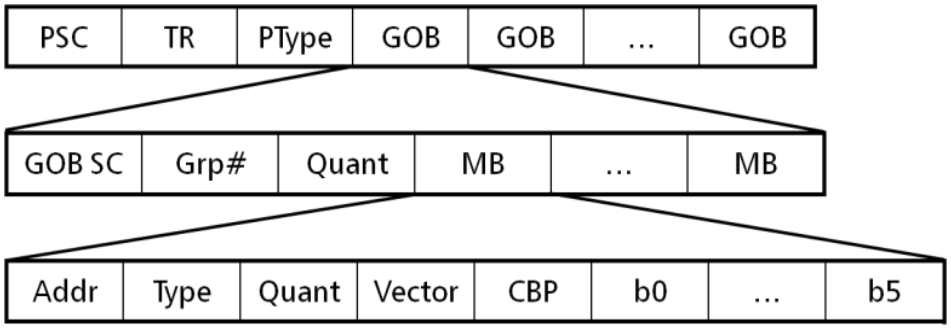
Type: Intra or inter-frame encoding of reference block.

Vector: Motion vector values

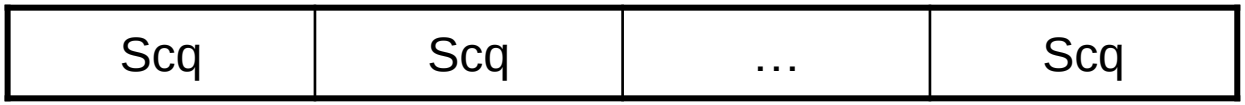
Quant: Vary quantization?!

CBP: Coded block pattern shows if the block should be omitted for encoding

bN: the blocks (4 Y, 1 Cr, 1 Cb)

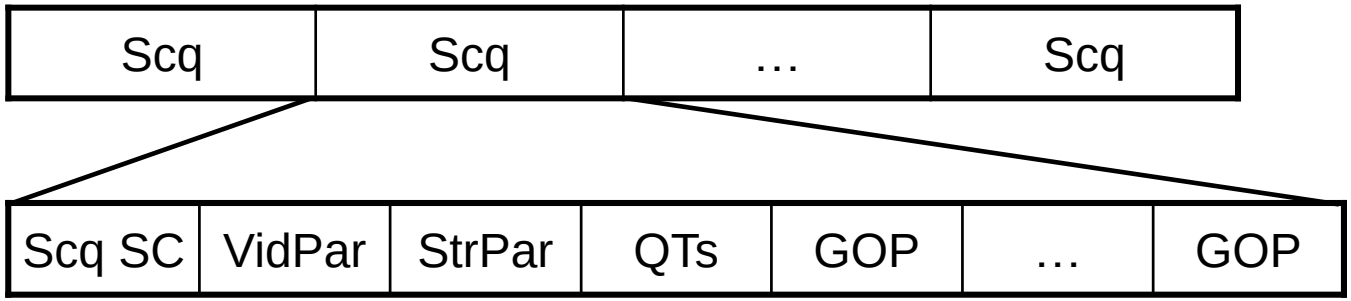


MPEG-Bitstream



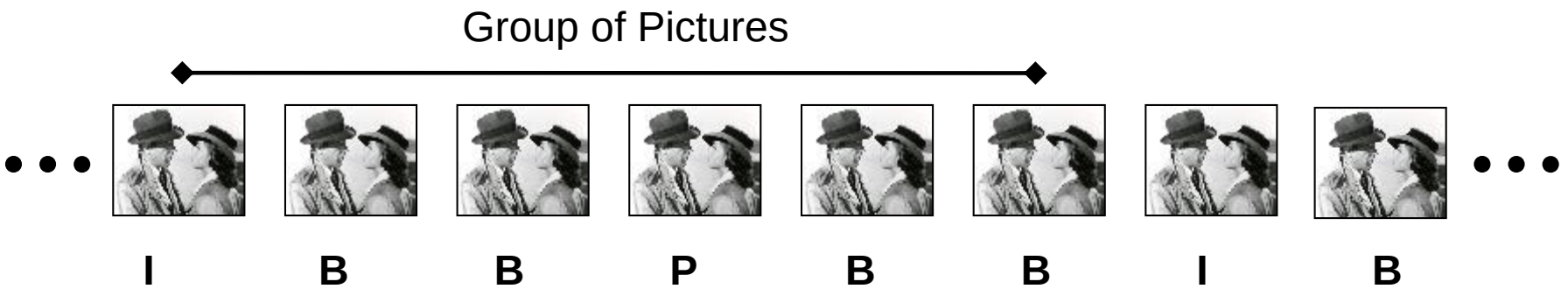
Sequence Layer

MPEG-Bitstream

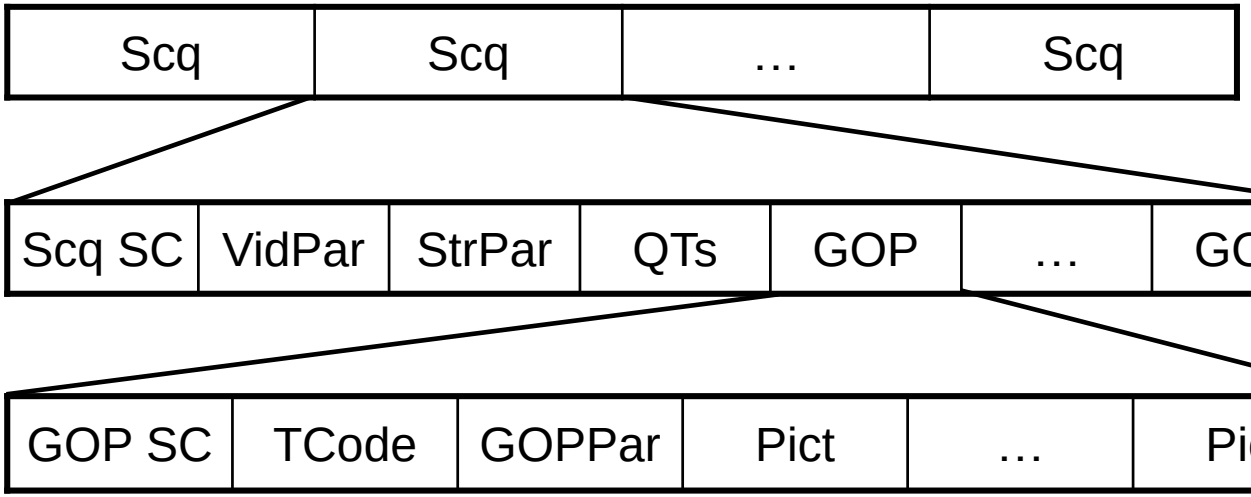


Sequence Layer

Group of Pictures Layer



MPEG-Bitstream

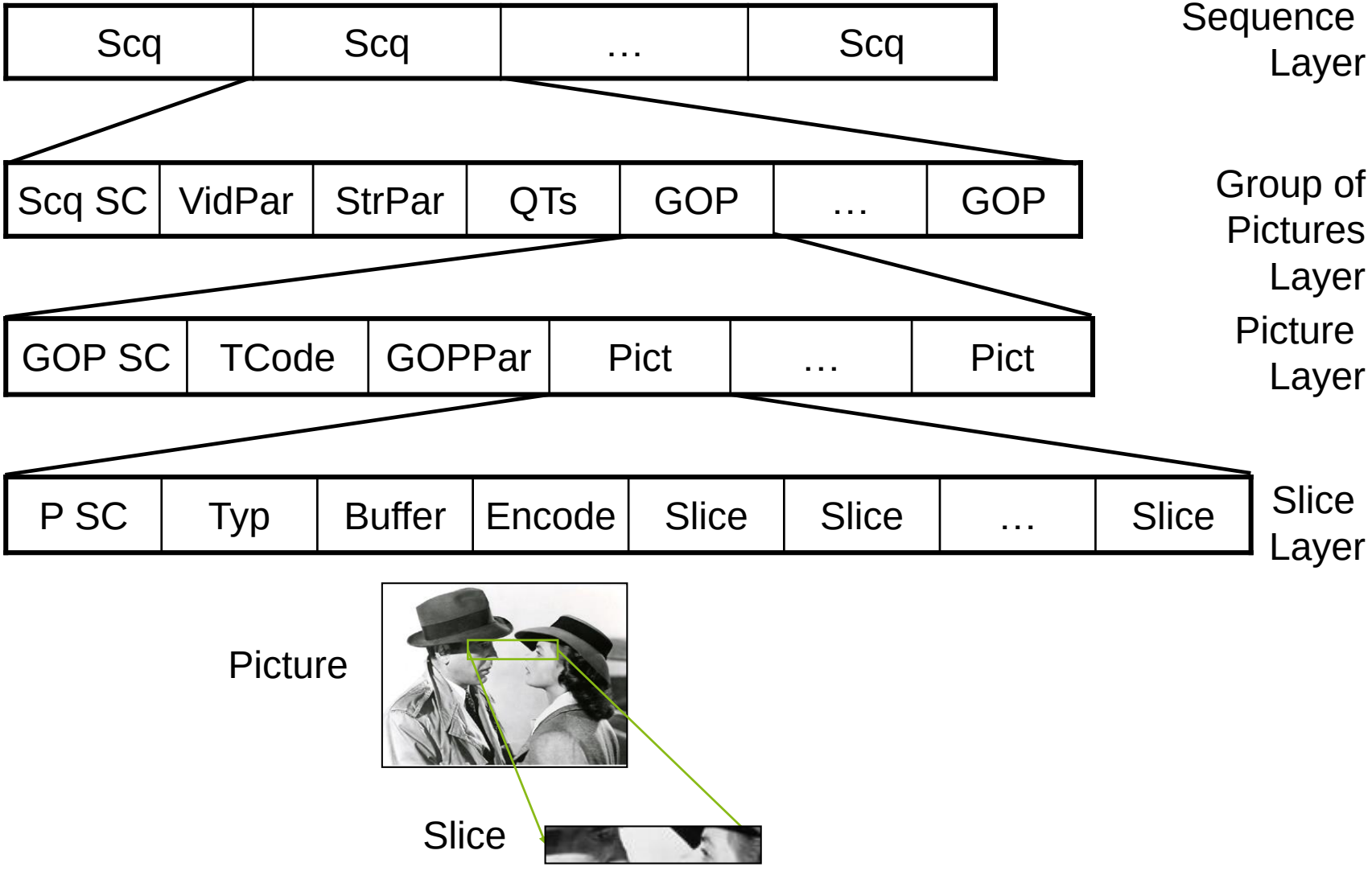


Sequence Layer

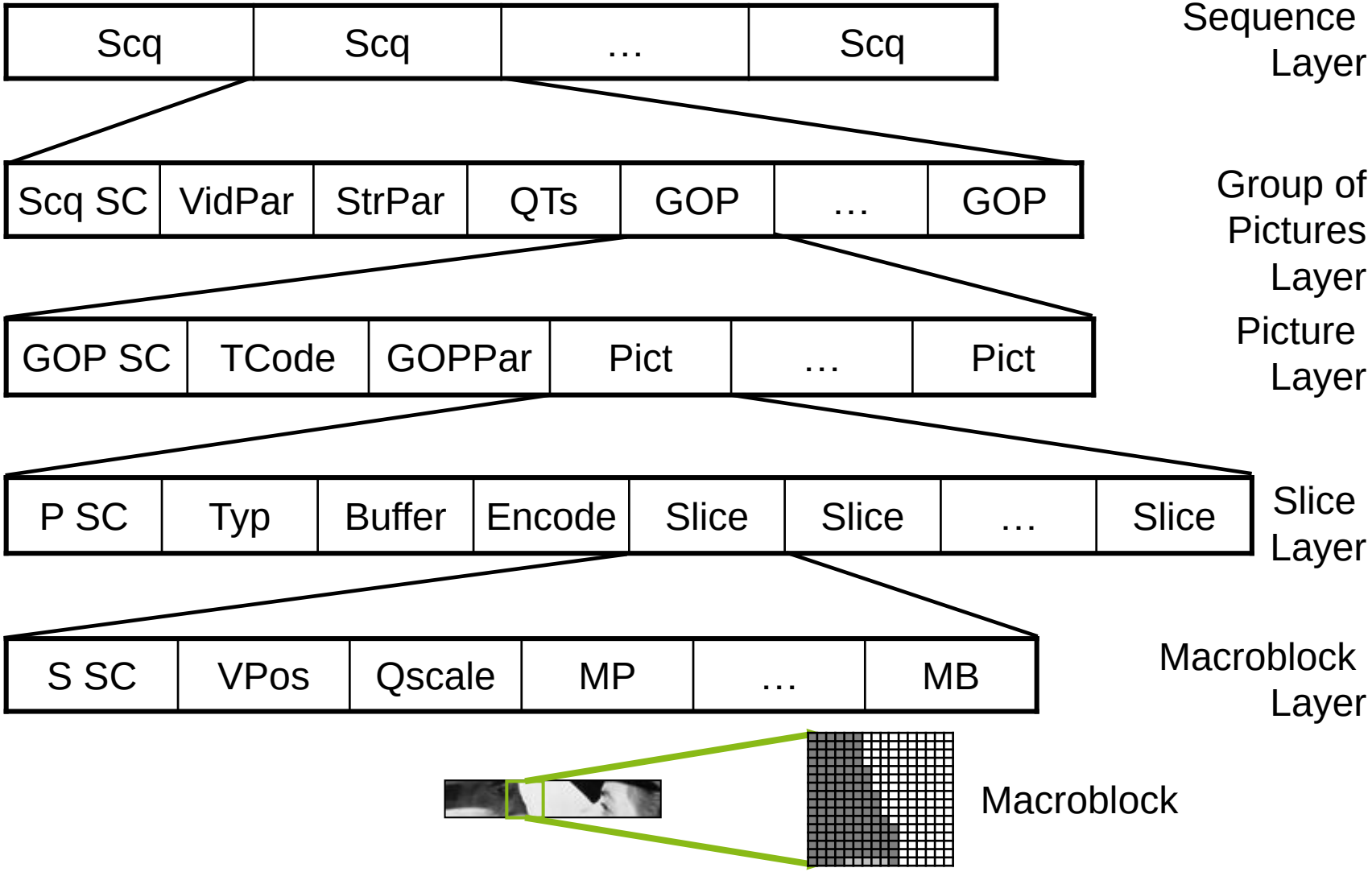
Group of Pictures Layer

Picture Layer

MPEG-Bitstream

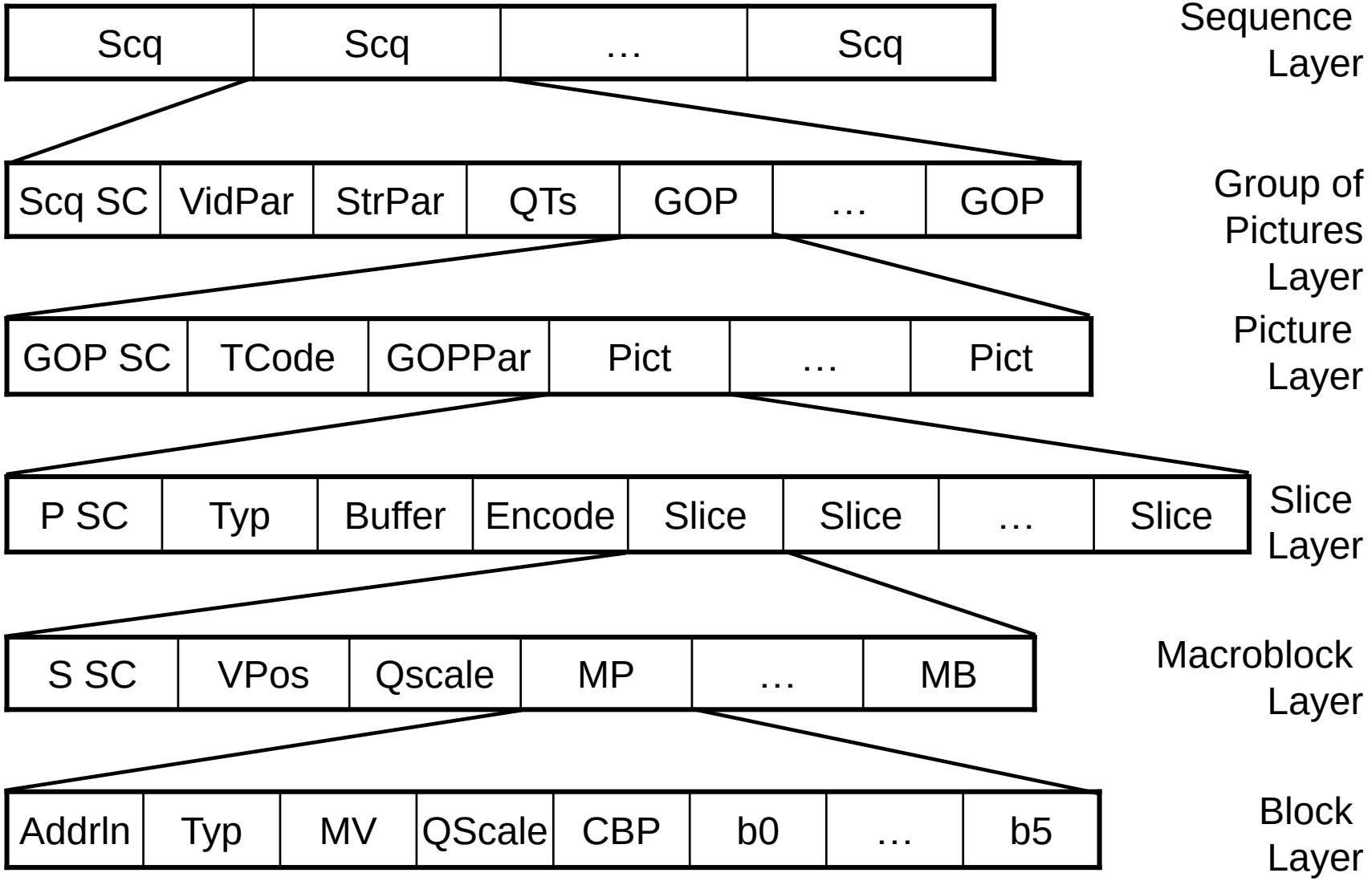


MPEG-Bitstream



MPEG-Bitstream

Juniorprofessur
MEDIA COMPUTING
TECHNISCHE UNIVERSITÄT CHEMNITZ



Examples (Please visit the page yourself!)

http://www.cybersite.de/german/service/Tutorial/mpeg/

The original sequence consists of 82 images, 352*240 pixel, 24 Bit RGB

Format	Size	Compression ratio	Remarks
BMPs	20,3MB		Standard image format, uncompressed
YUVs	10 MB	1:2	Raw information
FLI/FLC	4,3 MB	1:4,7	Delta encoding, Difference to the predecessor
I	238 kB	1:88	No motion estimation
IB	128 kB	1:163	
IBBIBB	55 kB	1:378	
IBBBBPBBBBPBBBB	42 kB	1:490	
IBBBBB...	25,3 kB	1:822	

Summary

H.261 is a little bit outdated and overcome by new methods with great compression ratios, but…

… it shows great potential to introduce to the principles of video encoding that are extended and optimized by current compression codecs! ☺

The relevant principles are:

- Split brightness and color information
- Use image information from predecessor image
- Adjustment of motion between frames
- Individual quantization
- While creating a bit stream many options are available to disable transfer of image parts

Coming up next…

(Presence sessions!)

~~01.06. Video coding~~

End of this week: Perform tutorial exercise *Video coding*!

08.06. Audio coding

15.06. Hexagonal coding

22.06. Complex tutorial:
Video, Audio, and hexagons

29.06. Consultation!

11.08. 8.00-9.00, 1/316
Bilingual written exams
Good luck and success!