Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Descartes research

Video Compression 101

Offenes Informatikkolloqium SS 2017

Lukas Iffländer

28.06.2017

http://se.informatik.uni-wuerzburg.de/

MOTIVATION

Back in Time

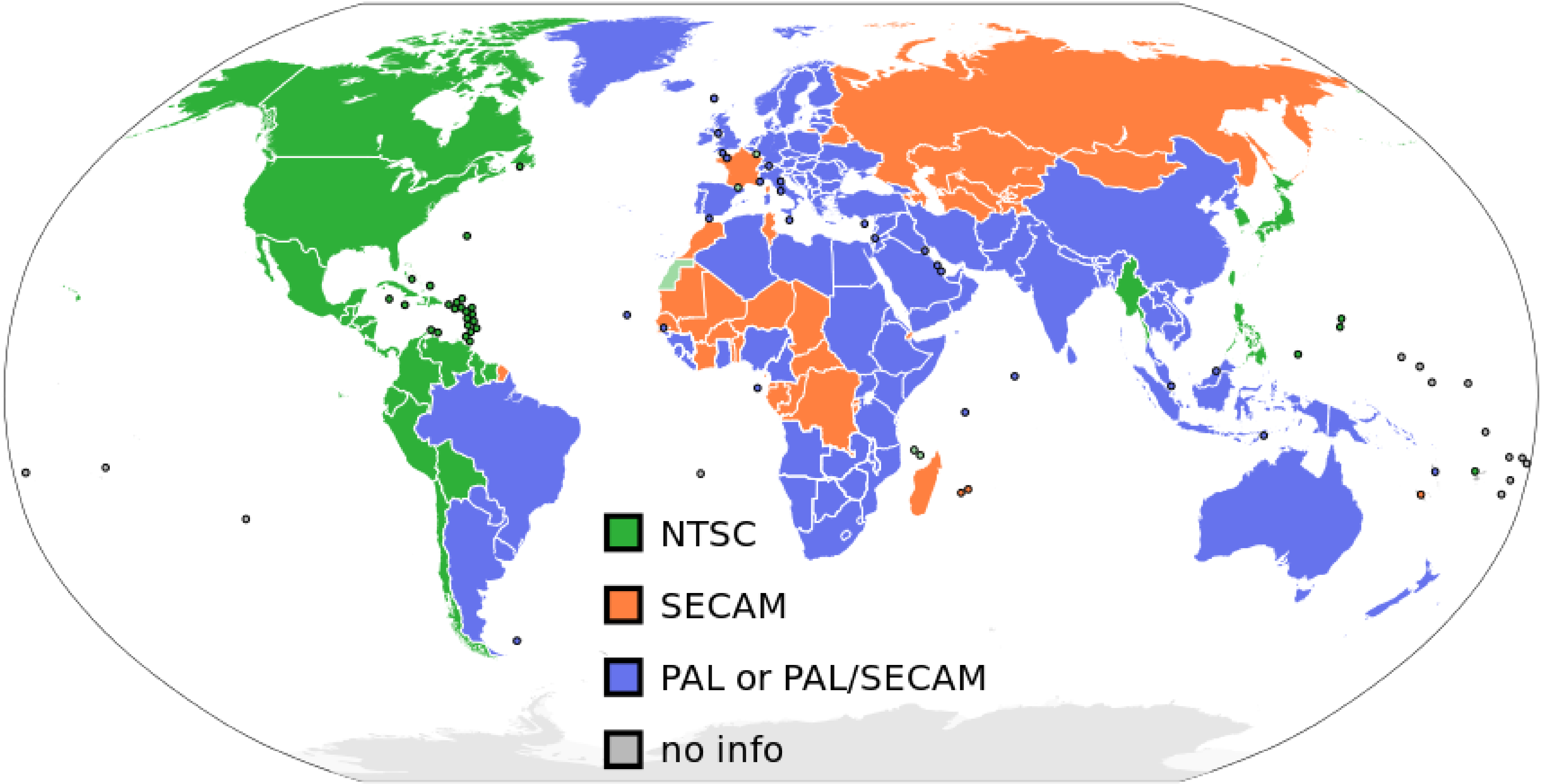
- Let‘s go back in time …
- … in a time, where nobody thought about streaming 4K via the internet
- … in a time, where TVs where not flat
- … in a time, where digital video was only a topic for few professionals
- … in a time, where renting movies was expensive
- … in a time, where a cheap solution was required to store videos recorded from TV
- … into my childhood

In the time of VHS Tapes

VHS

- Video Home System (VHS)
- Initial devices sold in 1976 (41 Years ago)
- Reading anlogue Images directly from the tape
- Line-wise reading
- Up to 430m of tape
- Playback time (E300/T240):
  - NTSC 250 min (4h 10min)
  - PAL 310 min (5h 10 min)

Excursion: PAL vs. NTSC



Excursion: PAL vs. NTSC

	NTSC	PAL
Abbreviation	National Television System Committee	Phase Alternation by Line
Video Bandwidth	4.2 MHz	5.0 MHz
Sound Carrier	4.5 MHz	5.5 MHz
Bandwidth	6 MHz	7 to 8 MHz
Vertical Frequency	60 Hz	50 Hz
Horizontal Frequency	15.734 kHz	15.625 kHz
Color Subcarrier Frequency	3.579545 MHz	4.433618 MHz
Lines/Field	525/60	625/50

How to „compress“ on VHS?

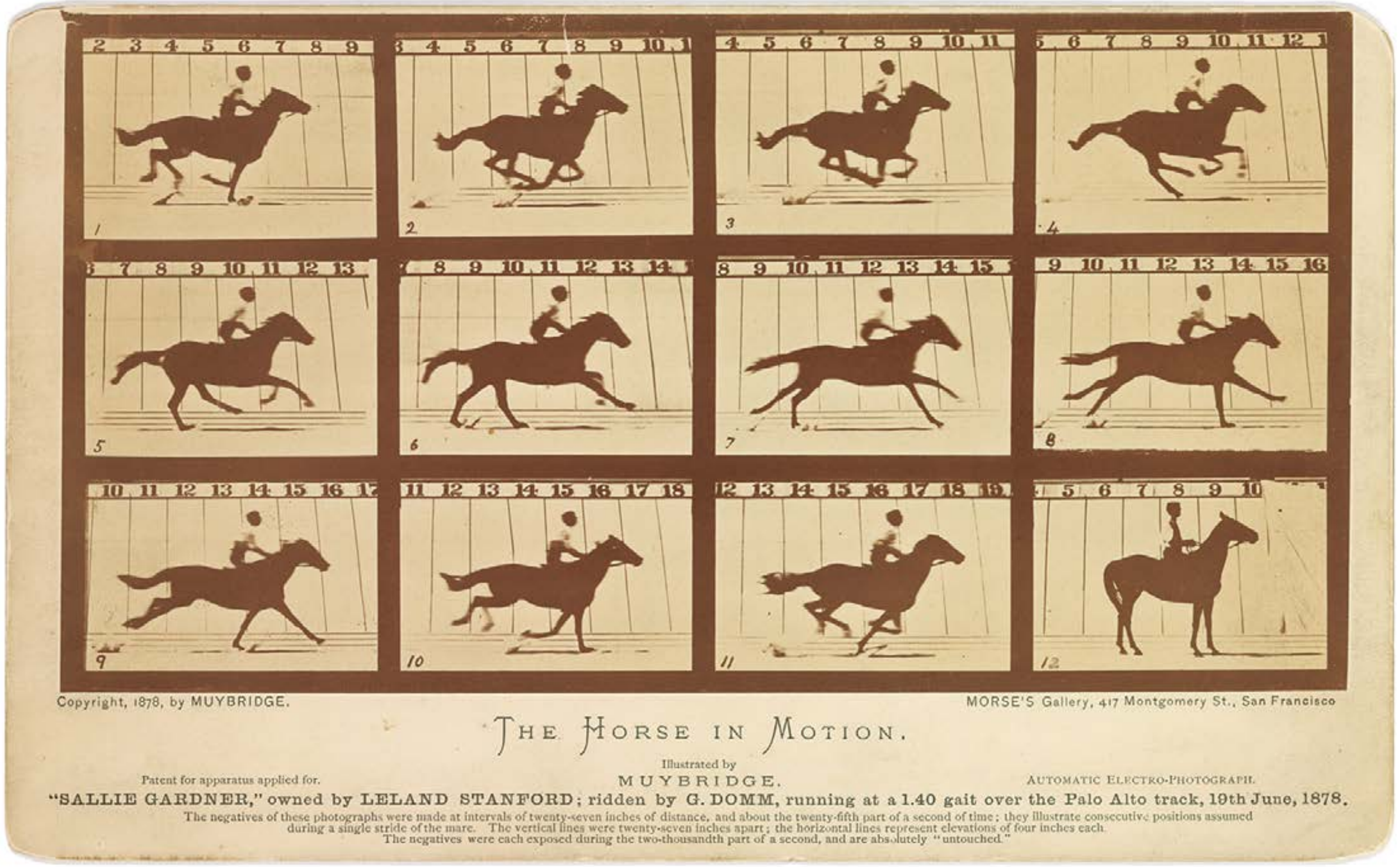
- Remember: Line-wise reading from a running analogue tape
- Idea: Let’s run the tape slower
- Long Play: ½ speed
  - NTSC 500 min (8h 20min)
  - PAL 620 min (10h 20 min)
- Extended Play / Super Long Play: 1/3 speed
  - NTSC 749 min (12h 29 min)
- Quality decrease:
  - Same information on only one half / one third of the space
  - Blurry screen
  - Problems with fast forward

To Digital (and beyond?)

- Introduction of the DVD to US Market in 1997
- At first only playback
- Long time frame until wide distribution of recording devices
- By 2005 still 94.5 million US Households equiped with VHS
- Mid 2000s retailer stopped selling VHS equipment and tapes
- Steady decline since then
- Still a standard staying undead
  - Family videos often in VHS
  - Some movies not re-released digitaly
  - 2009 Panasonic released a Combo-Player Blu-ray + VHS
  - Production of last VHS recorder: July 2016

What is a movie?

A series of pictures



Let‘s calculate size

- We use no compression
- We just add digital pictures after eachother

Resolution Horizontal	Resolution Vertical	Frame Rate	Duration [min]	Color Depth	Frames	MPixel/ Frame	MiB/ Frame	GiB total
1280	720	25	90	8	135.000	0,92	2,64	348
1920	1080	25	90	8	135.000	2,07	5,93	782
3840	2160	25	90	8	135.000	8,29	23,73	3.129
1280	720	48	90	8	259.200	0,92	2,64	667
1920	1080	48	90	8	259.200	2,07	5,93	1.502
3840	2160	48	90	8	259.200	8,29	23,73	6.007
1280	720	25	90	10	135.000	0,92	3,30	435
1920	1080	25	90	10	135.000	2,07	7,42	978
3840	2160	25	90	10	135.000	8,29	29,66	3.911
1280	720	48	90	10	259.200	0,92	3,30	834
1920	1080	48	90	10	259.200	2,07	7,42	1.877
3840	2160	48	90	10	259.200	8,29	29,66	7.508

Reason for compression

- Nobody wants to buy a 3 TB Harddrive for one 4K Movie
- We need a way to reduce the size
- ➔ Idea: Develop algorithms to reduce the size based on information about the image

GENERIC COMPRESSION

Compression Algorithms

- Lossless compression algorithms
  - Allow original data to be perfectly reconstructed
  - Typical processing steps:
    - Generate statistical model of the data
    - Use model to map input data to bit sequence so that frequently encountered data is more efficiently encoded
  - Typically low compression ratios for multimedia applications
- Lossy compression algorithms
  - Compressed data is not the same as the original data, but a close approximation of it
  - Trade-off distortion vs. required rate
  - Yields a much higher compression ratio than that of lossless compression

How to measure?

- How to measure the efficiency of a lossy compressions scheme?
- How to describe the difference between original and reconstructed data?

Rate-distortion theory

- Rate and Distortion as measures for efficiency of compression and difference between original and reconstructed data
- Aims at describing the minimum amount of distortion and the lowest rate possible
- Basic problem in rate-distortion theory:
  - Given a source distribution and a distortion measure, what is the minimum expected distortion achievable at a particular rate?
  - What is the minimum rate required to achieve a particular distortion?

Closeness/Fidelity of Reconstructed Sequence

- Perceived distortion strongly depends on specific use-case:
  - What is being compressed?
  - Who judges the distortion?
- Examples:
  - X-ray Image for diagnostics

    ➔Ask a doctor
  - Satellite image processed by a computer

    ➔ Investigate impact of distortion on machine behaviour
- Usually done by use surveys
- Mathematical model still lacking

Measuring Distortion

- Mean square error (MSE) σ^2, where x_n, y_n, and N are the input data sequenc, reconstructed data sequence, and length of the data sequence

$$\sigma^2 = \frac{1}{N}\sum_{n-1}^{N}(x_n - y_n)^2$$

- Signal to noise ratio (SNR), in decibel units (dB), where σ_x^2 is the average square value of the original data sequence and σ_d^2 is the MSE.

$$SNR = 10\log_{10}\frac{\sigma_x^2}{\sigma_d^2}$$

- Peak signal to noise ration (PSNR),

$$PSNR = 10\log_{10}\frac{x_{peak}^2}{\sigma_d^2}$$

IMAGE COMPRESSION

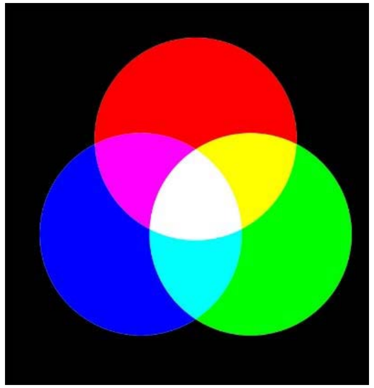
Idea

Compress the still images in our series of images

RGB and YC_bC_r

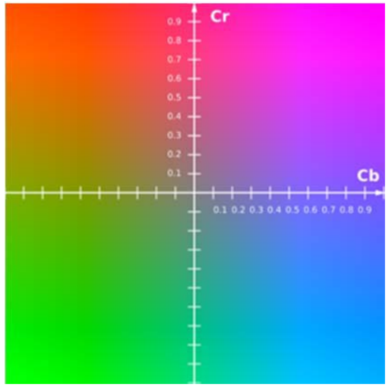
RGB

- Additive Color Space
  - Red
  - Green
  - Blue

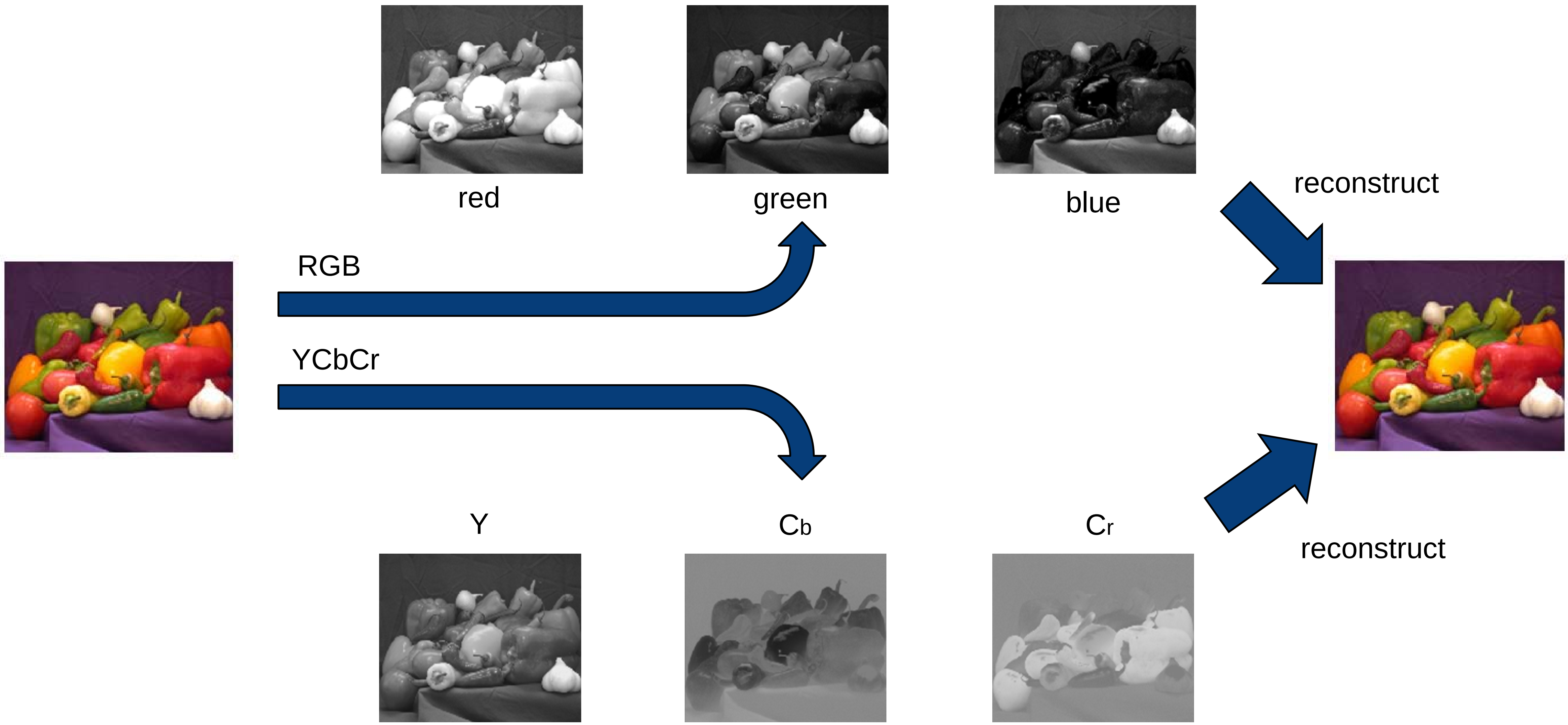


YC_bC_r

- Color information separated
  - Luma component (Y)
  - Chroma Components
    - Blue-difference C_b
    - Red-difference C_r

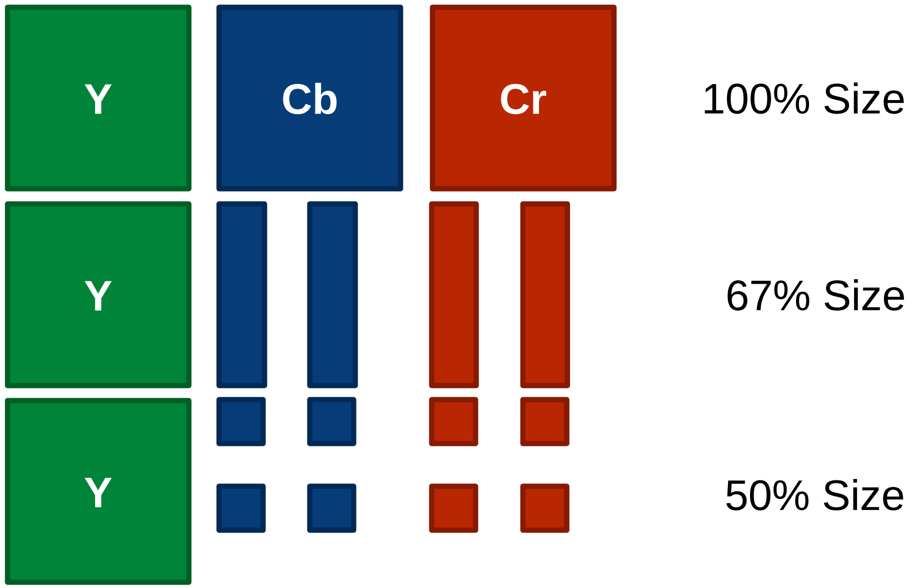


RGB and YC_bC_r



Downsampling

- Subsampling expressed as three part ratio J:a:b (e.g. 4:2:2)
  - J: horizontal sampling reference (width of region)
  - a: number of crominance samples in first row
  - b: number of crominance samples in second row
- Examples
  - 4:4:4
  - 4:2:2
  - 4:2:0



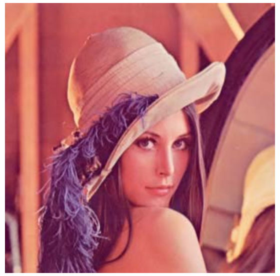
JPEG Transformation

- RGB to YC_bC_r Transformation
- 8x8 Direct Cosine Transformation
  - Puts most of spectral energy on few values
  - Quantization (represents samples to a given level of accuracy)
  - ZigZac-Scan
  - Huffman Encoding
- JPEG 2000
  - Use of Wavelets

Impact of Correlation on Compression

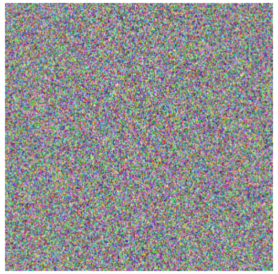
Original Image

768 KB



Original Image

768 KB



Original Image

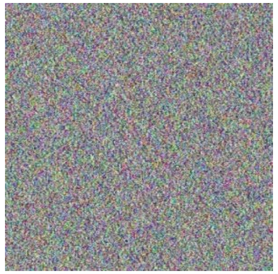
768 KB

Compressed Image

9 KB

Compressed Image

50 KB



Compressed Image

410 KB

1.17%

6.50%

53.32%

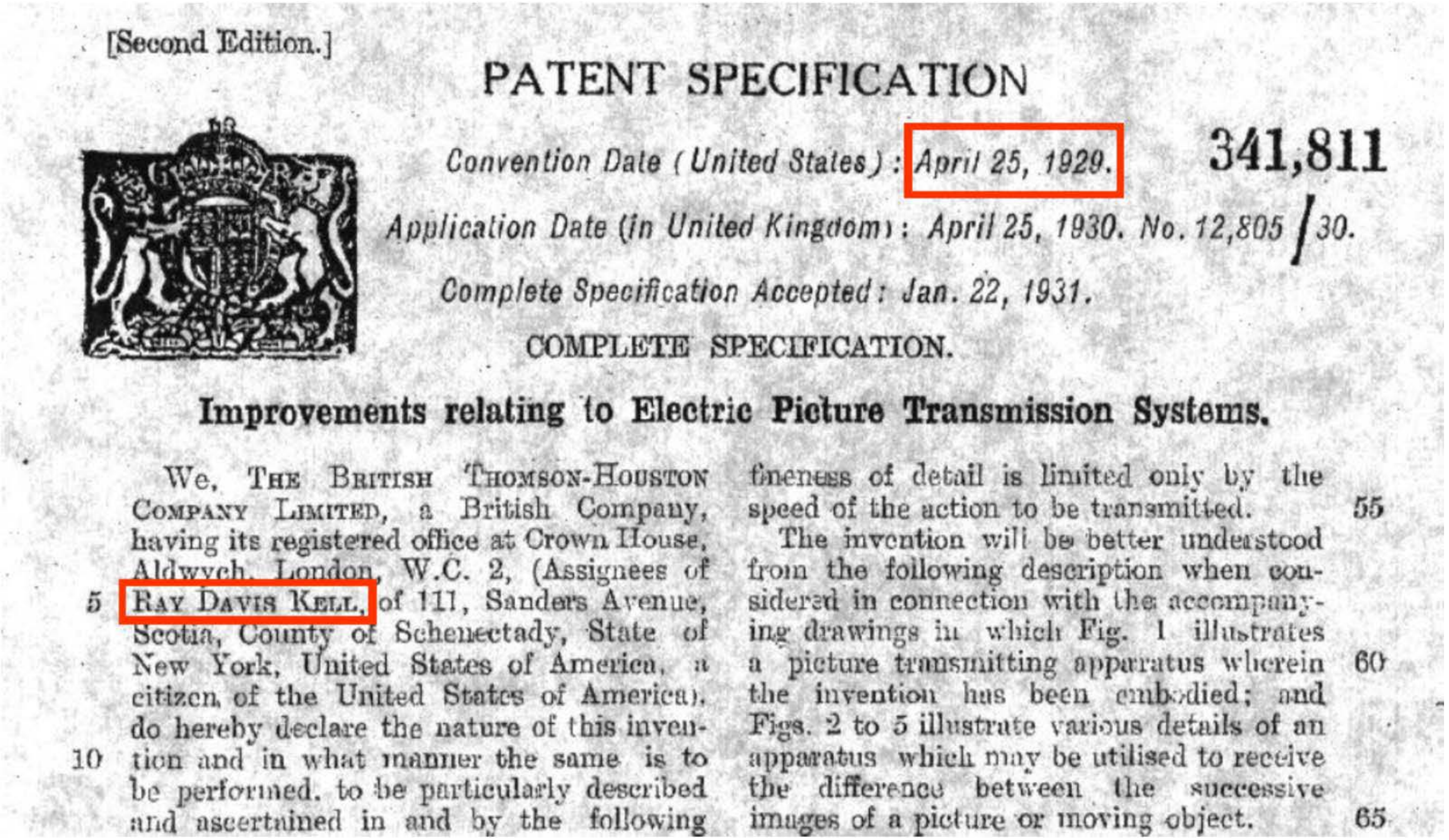
Where are we now?

- We can compress our images using JPEG
- Gain depending on scenes between 99% and 0%. Average ~90%
- So 300GB for a 90 minutes 4K movie?
  - 430 CDs (720 MB)
  - 34x DVD (9 GB)
  - 12x BD Single Layer (25 GB)
  - 6x BD Double Layer (50 GB)
  - 3x BD XL (100/128 GB)
- Remember when you had to switch floppies / CDs when installing software?

➔ Nobody wants to do this when watching movies

VIDEO COMPRESSION

Interframe Coding of Signals



[Second Edition.]

PATENT SPECIFICATION

Convention Date (United States): April 25, 1929. 341,811

Application Date (in United Kingdom): April 25, 1930. No. 12,805/30.

Complete Specification Accepted: Jan. 22, 1931.

COMPLETE SPECIFICATION.

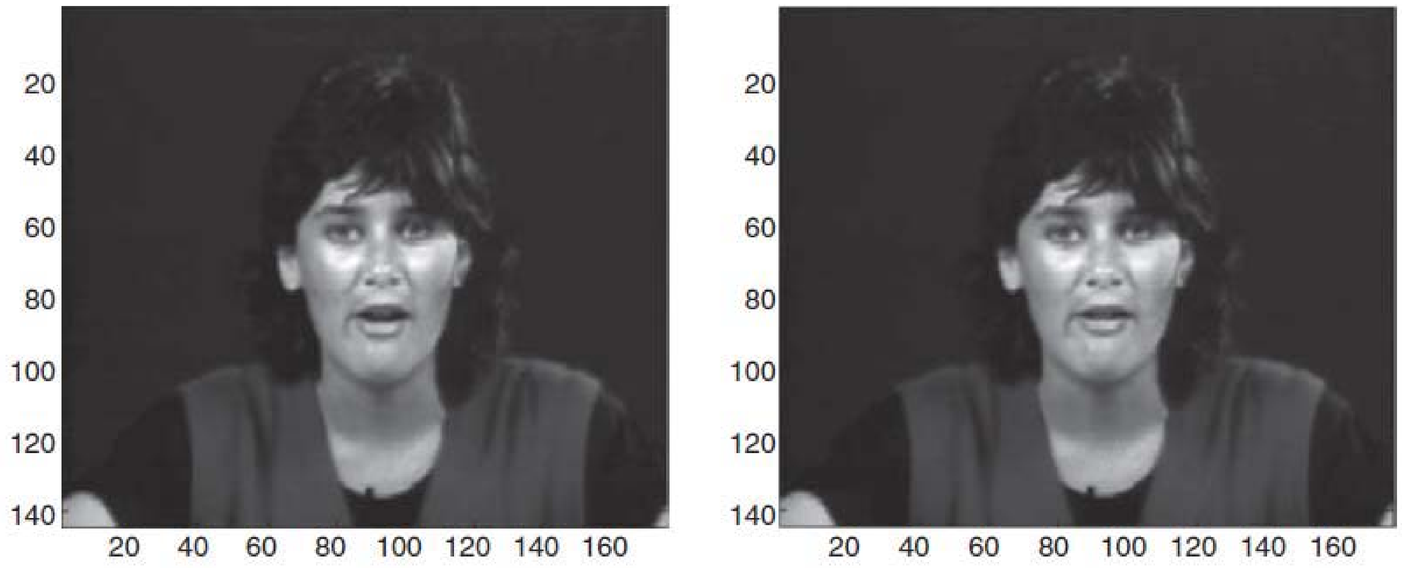
Improvements relating to Electric Picture Transmission Systems.

We, THE BRITISH THOMSON-HOUSTON COMPANY LIMITED, a British Company, having its registered office at Crown House, Aldwych, London, W.C. 2, (Assignees of RAY DAVIS KELL, of 111, Sanders Avenue, Scotia, County of Schenectady, State of New York, United States of America, a citizen of the United States of America), do hereby declare the nature of this invention and in what manner the same is to be performed, to be particularly described and ascertained in and by the following fineness of detail is limited only by the speed of the action to be transmitted.

The invention will be better understood from the following description when considered in connection with the accompanying drawings in which Fig. 1 illustrates a picture transmitting apparatus wherein the invention has been embodied; and Figs. 2 to 5 illustrate various details of an apparatus which may be utilised to receive the difference between the successive images of a picture or moving object.

- It has been customary in the past to transmit successive complete images of the transmitted picture.”[...] “In accordance with this invention, this difficulty is avoided by transmitting only the difference between successive images of the object.”

Temporal Redundancy



(a) Frame 24 (b) Frame 25

Yun Shi, Huifang Sun, Image and Video Compression for Multimedia Engineering

- Typically high correlation between pixels in successive frames
- Taking advantage of interframe correlation leads to video data compression

Frame Replenishment

- Pixels classified into changing and unchanging areas w.r.t. intensity
  - If difference threshold between previous and current pixel is exceeded, pixels are replenished, i.e., address and intensity of pixels are coded, stored in a buffer and then transmitted
  - Otherwise nothing is transmitted
  - Channel capacity depends on replenishment rate (→ threshold)
  - → Higher coding efficiency than encoding every pixel of every frame
  - → “Frame-difference prediction techniques”

Yun Shi, Huifang Sun, Image and Video Compression for Multimedia Engineering

Drawback:

- Handle of frame sequences containing rapid changes
  - Maintain transmission rate bit rate, threshold has to be raised
- Slow changes may not show up at receiver („dirty windows effect“)

Change Detection

- Frame n-1

$H_x = 7.23$

- Frame n

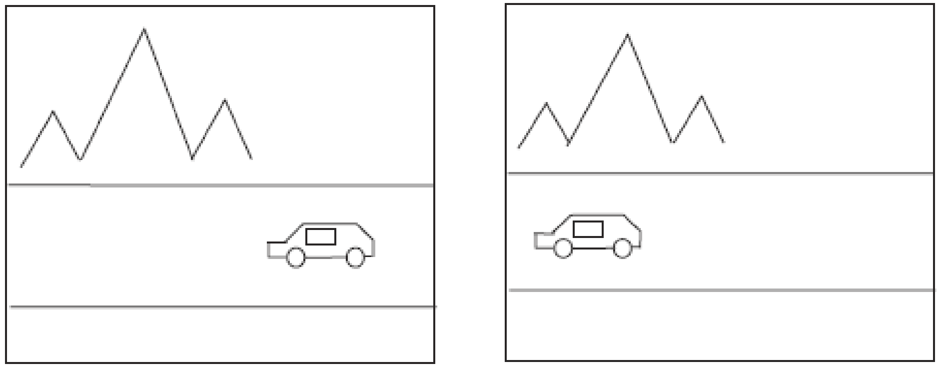
$H_x = 7.23$

- Frame n – Frame n-1

$H_x = 2.64$

Motion Compensated Coding

- Idea of a motion model
  - Changes between frames due to translation of moving objects
  - Estimation of „displacement“ vectors



(a) t_{n-1} (b) t_n

- Encoding of differential signals between intensity value of picture elements in the moving areas and counterparts in the previous frame

➔ Compression w.r.t. motion (“side information”)

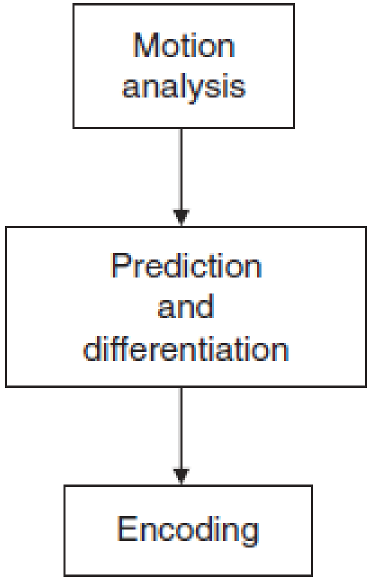
➔ More efficient than frame-difference prediction techniques

- Problem: Introduction of prediction errors
  - Rotations, camera zooming
  - Disappearance of picture areas, appearance of new areas

Prediction Error for Motion Compensated Coding

- Predicted frame and original frame may differ significantly
  - Prediction error leads to disturbed video frames
- Solution: Compute and transmit prediction error additionally to achieve good quality frames
- Higher coding efficiency, but also higher computational complexity
- Three stage coding:
  - Motion analysis; estimation of displacement vectors
  - Prediction error computed using predicted and original frame
  - Encoding of prediction error and motion vectors

➔ 22-50% reduced bit rate compared to frame-difference prediction techniques

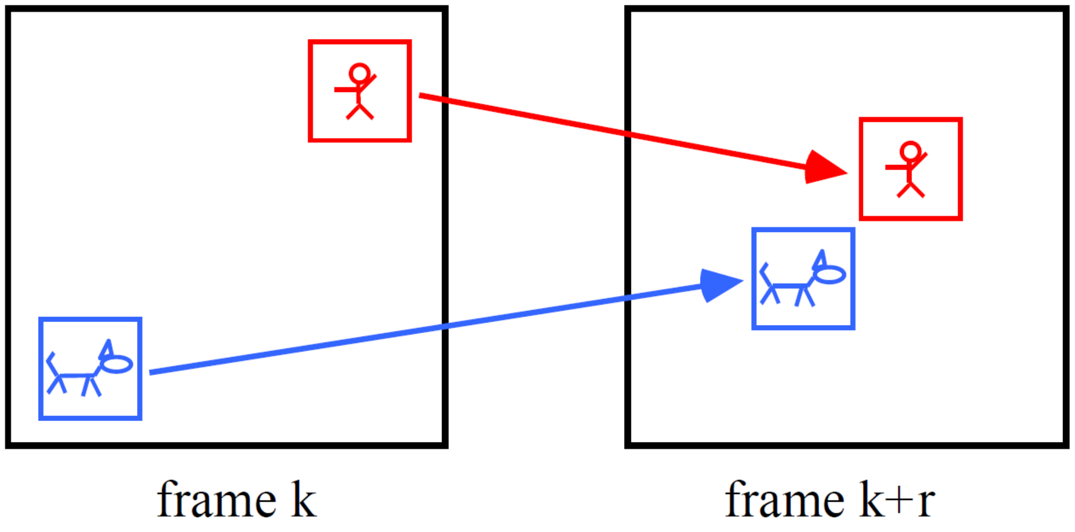


Block Matching

- Partitioning of frame into non overlapped, equally spaced, fixed size, small rectangular blocks
- Trade-off block size: The smaller the block size,
  - the better approximation of rotations, zooms, etc.,
  - the more motion vectors have to be computed, i.e., more cpu cycles, more side information
  - 16x16 blocks used in MPEG-1, MPEG2,…
- Find one displacement vector for each block
- Within a search range, find the best math that minimizes an error measure
- Intelligent search strategies can reduce computation

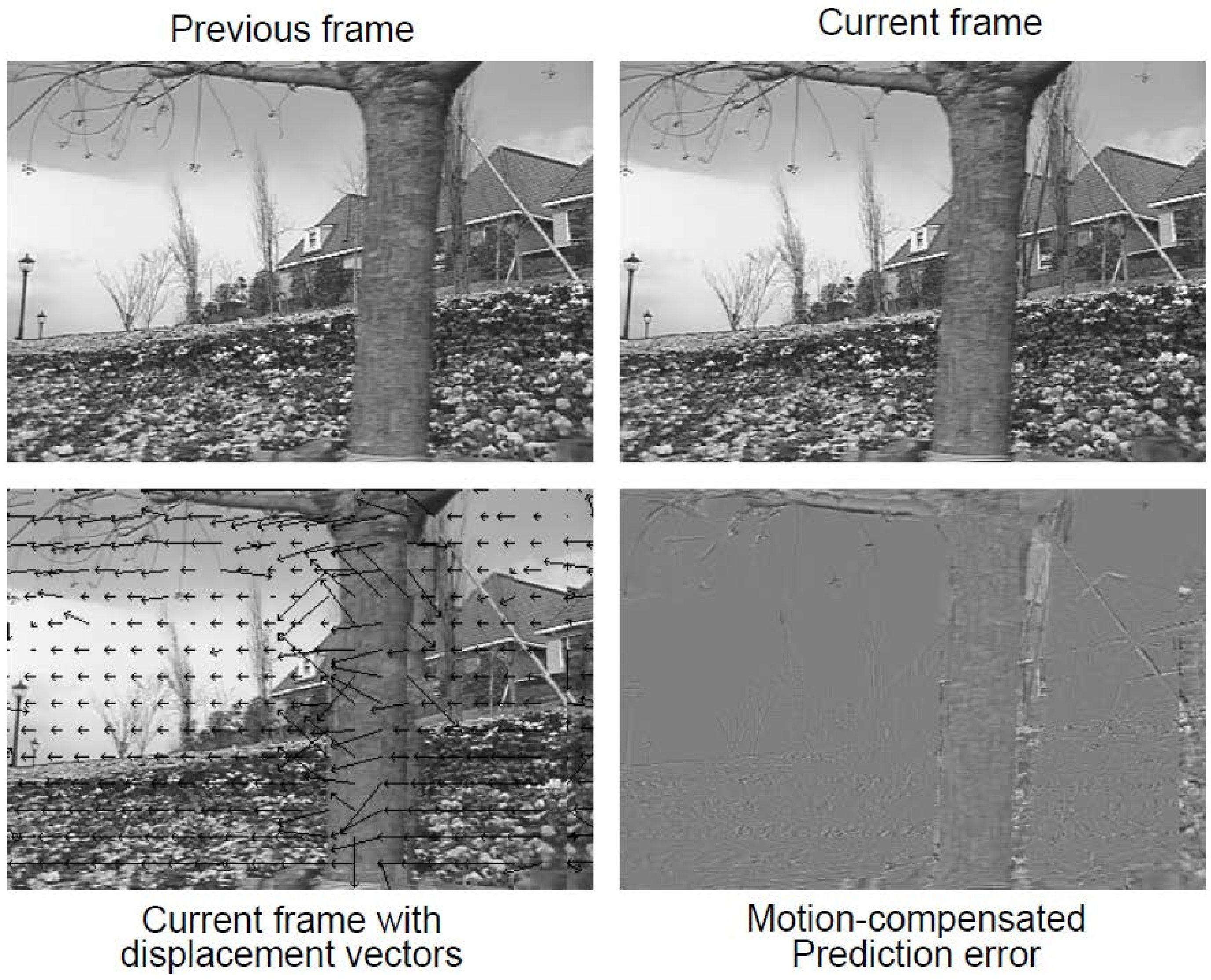
Block Matching

- Estimate Motion
  - Equivalently: Find correspondence between frames
- Block Matching (template matching)
  - For every block in one frame, find the best match in another
- Requires search



Bernd Girod:Image Communication II

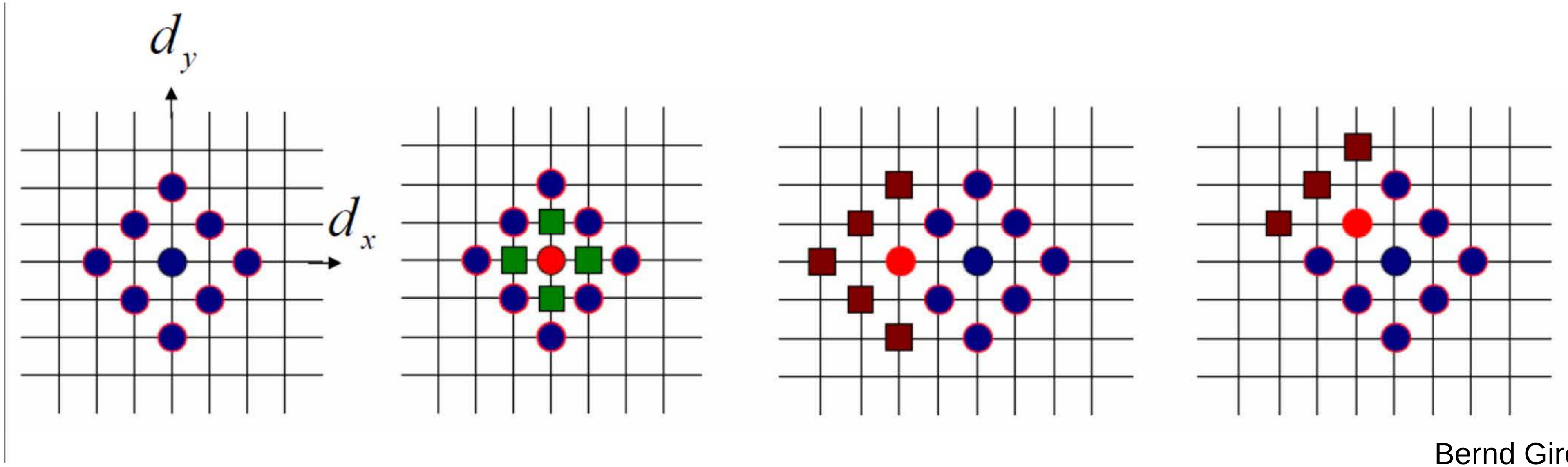
Motion-Compensated Prediction: Example



Bernd Girod:Image Communication II

Block Matching: Search Strategies

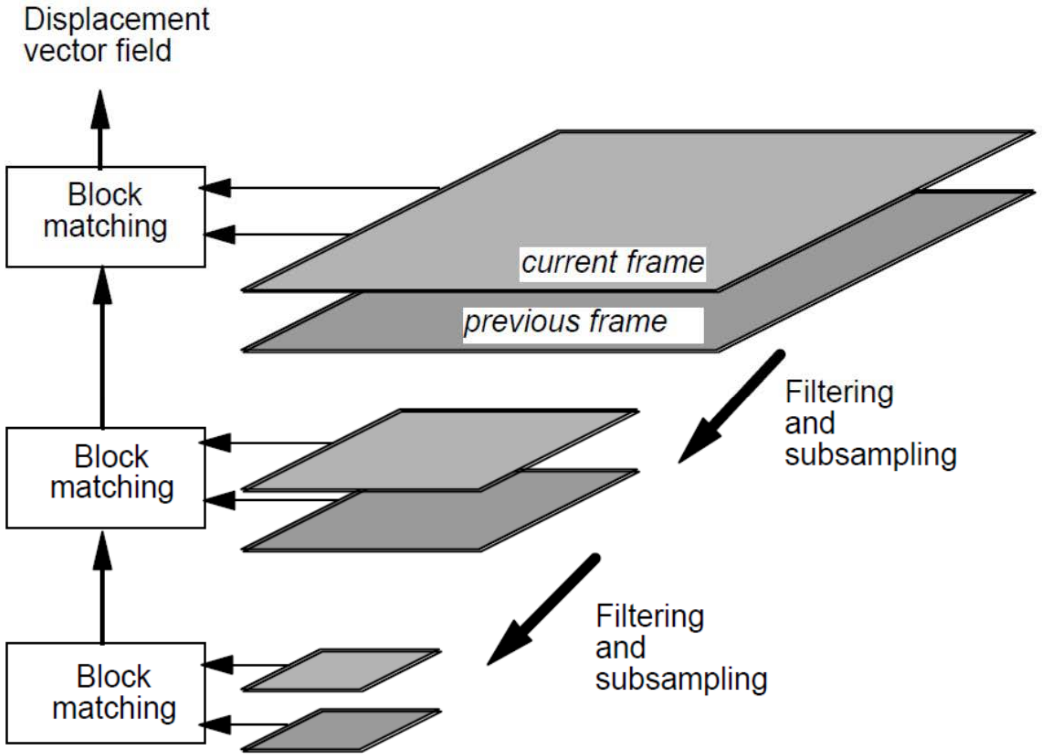
- Full Search
- Logarithmic Search
- Diamond Search



Bernd Girod:Image Communication II

Hierarchical Block Matching

- Large blocks:
  - More likely to track actual motion
  - Less likely to converge on local minima
- Small blocks:
  - Better quality of matches
- Hierarchical block matching
  - Use motion vectors of large blocks as starting points for searches for small blocks



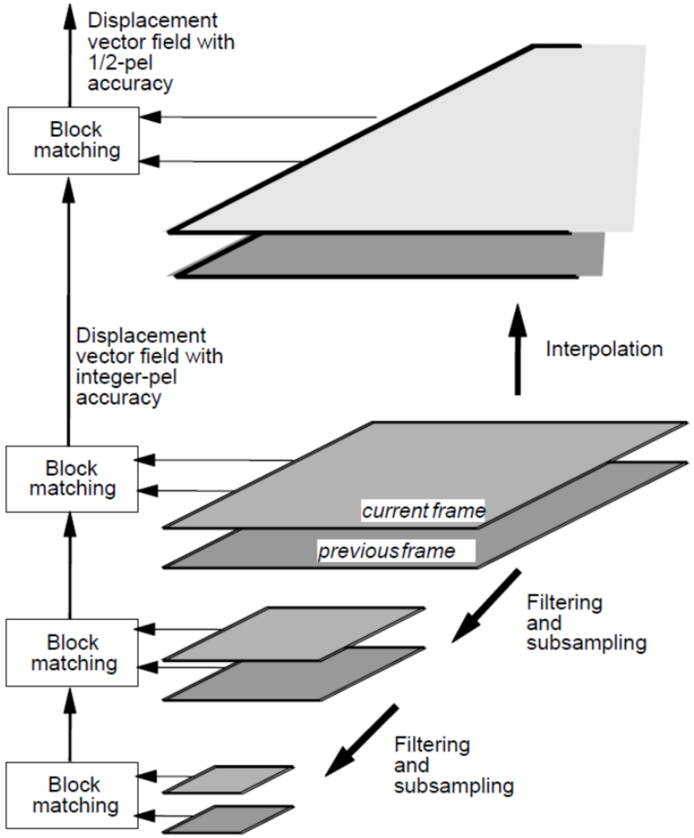
Bernd Girod:Image Communication II

Sub Pixel Accuracy

- Interpolate pixel raster of the reference frame to desired fractional pixel accuracy

  (→ interpolation)

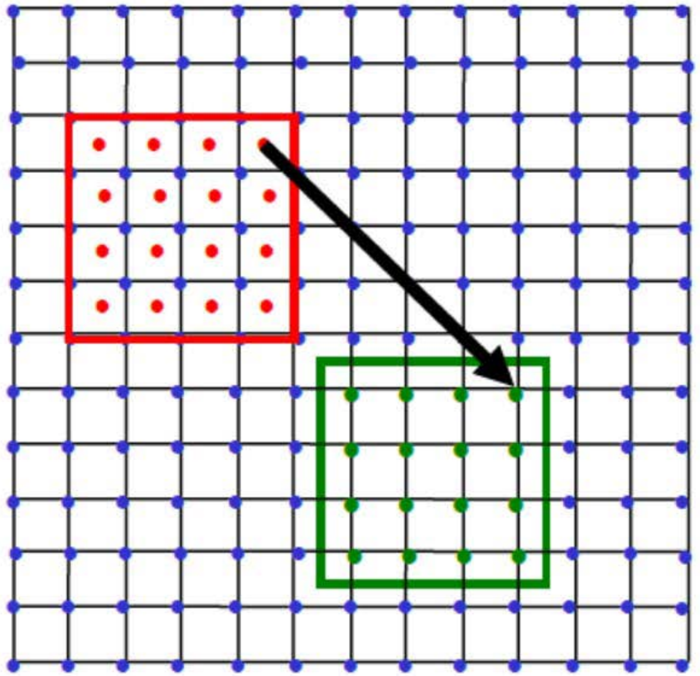
- Extension of displacement vector search to fractional accuracy



Bernd Girod:Image Communication II

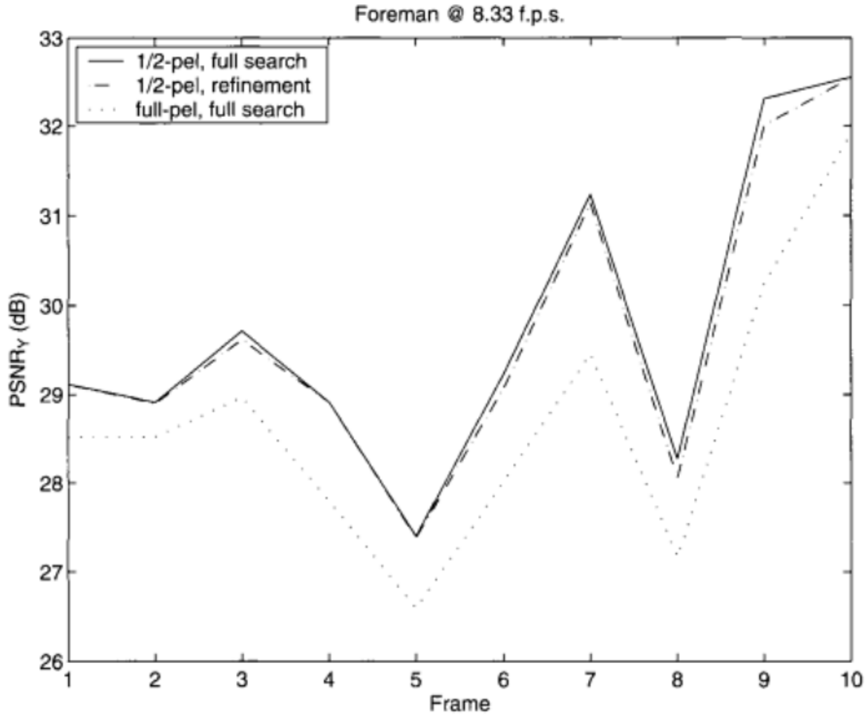
Sub Pixel Accuracy

- Half-pixel accurate displacements
  - Number of candidate blocks increases w.r.t. maximum distance
- PSNR as quality metric, different search algorithmns



$$\begin{pmatrix} d_x \\ d_y \end{pmatrix} = \begin{pmatrix} 4.5 \\ 4.5 \end{pmatrix}$$

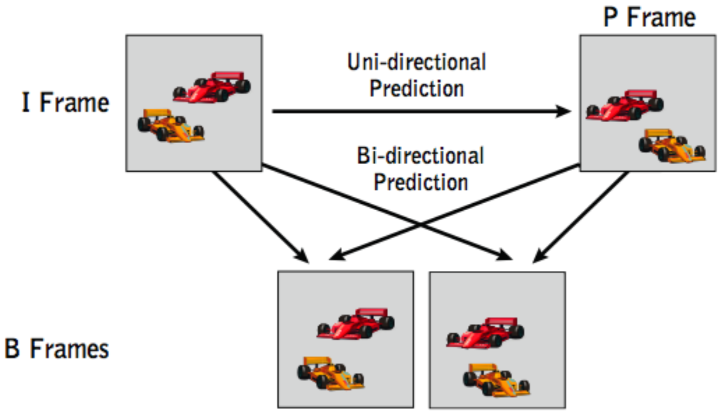
Bernd Girod:Image Communication II

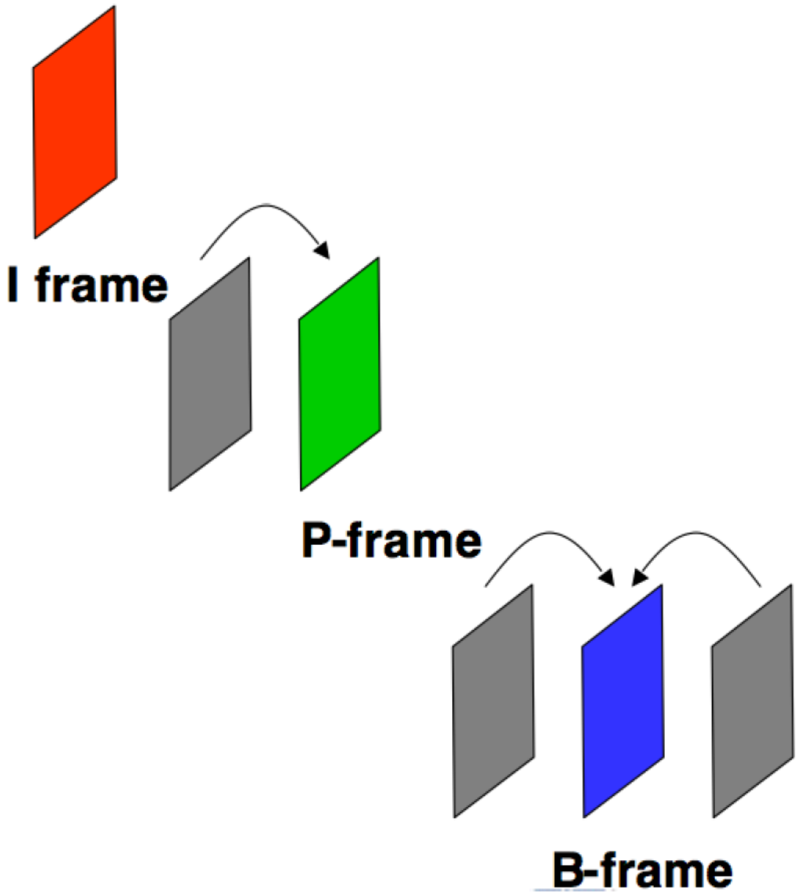


M. Al-Mualla, C. Canagarajah, D. Bull, Video Coding for Mobile Communications

Frame Types

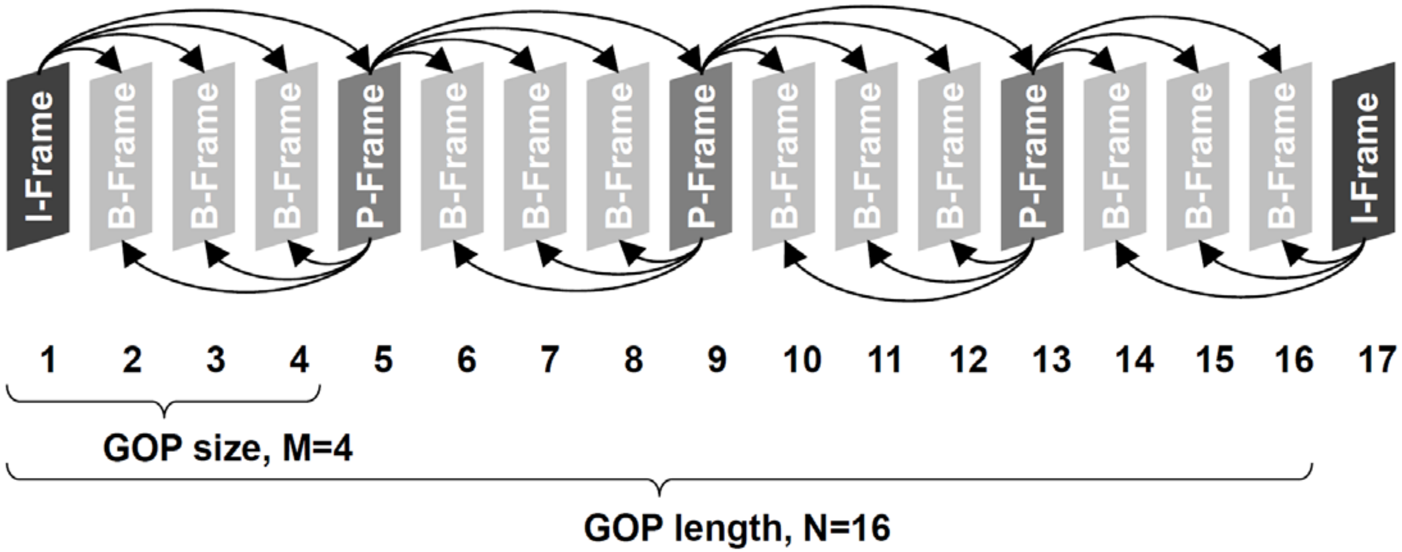
- Different frame types:
  - I-frame: „Intra-coded frame“
    - Independent of other frames
    - One intracoded frame (JPEG)
  - P-frame: „Predictively coded frame“
    - Depends on previous frame
  - B-frame: „Bidirectionally predicted frame“
    - Depends on previous and subsequent frame



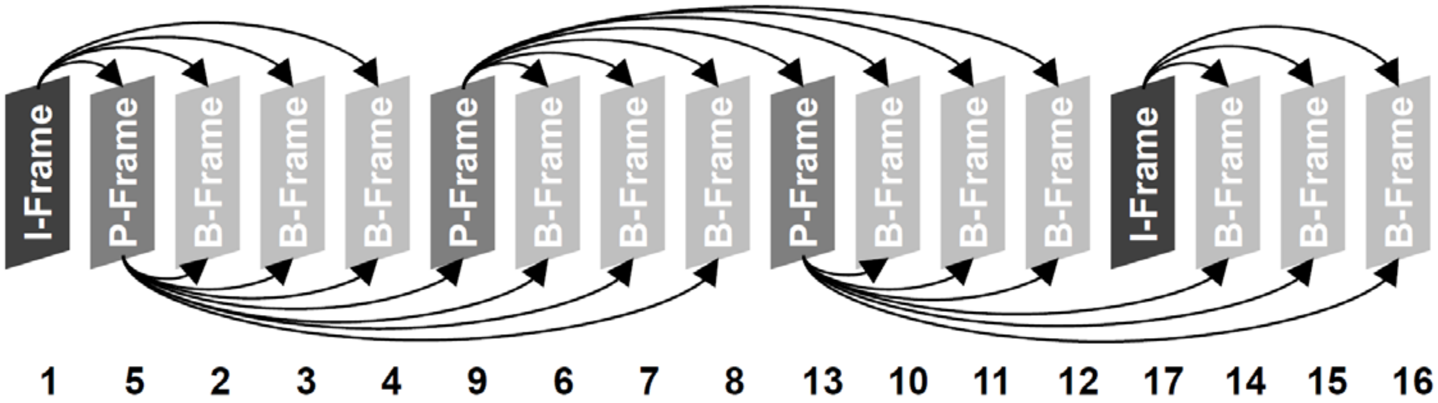


Group-of-Pictures

- Display Order



- Encoding Order



VIDEO ENCODING TECHNOLOGIES

MPEG-1/2 Video

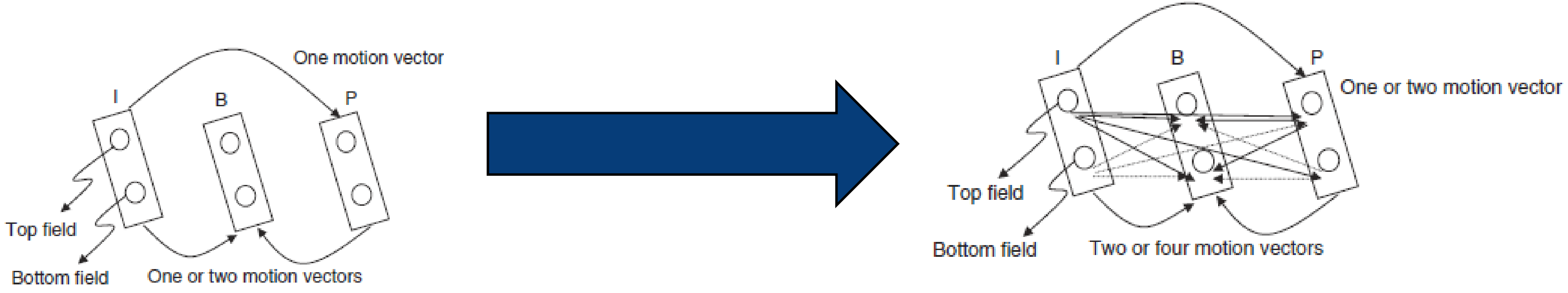
- Developed for multimedia CD-ROM applications
- Important Features:
  - Frame-based random access of video
  - Fast forward/backward searches
  - Reverse playback of video
- Important: Standards provide knowledge for how to design decoders that are able to successfully decode the bitstream
  - They DO NOT specify the means of generating the bit stream, e.g., how to generate video stream with best picture quality for a given bit rate?

MPEG-1 Features

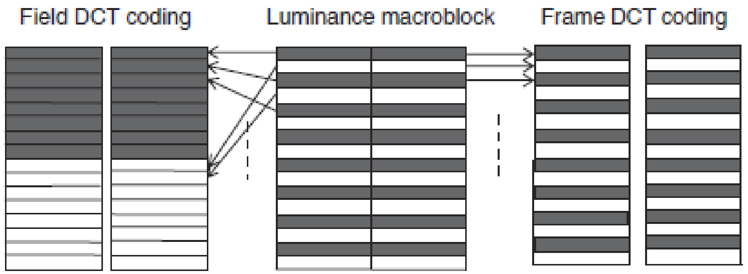
- Achieves high compression ratio by exploiting intraframe and interframe redundancies
  - Mainly based on DCT coding and interframe motion compensation
  - JPEG pictures are used from time to time to enable random access
- Full-motion compensated DCT & DPCM hybrid coding

MPEG-2 Enhancements

- Basic coding structure similar to MPEG-1, i.e., intraframe and interframe DCT with I-,P-. B-pictures.
- Most important features:
  - Frame prediction modes for supporting interlaced video input

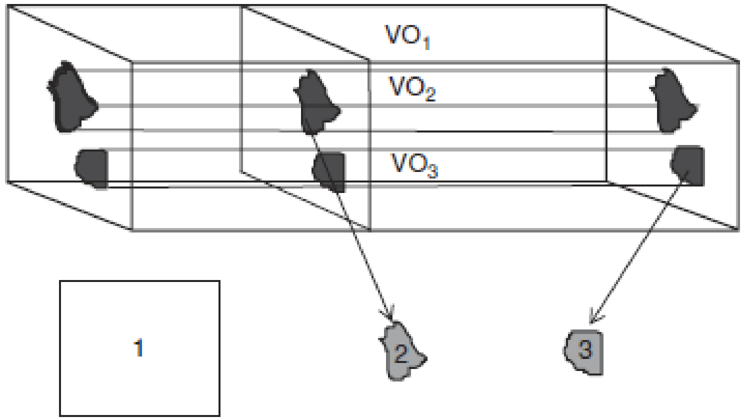


  - Frame DCT coding syntax (interlaced video)
    - Adaptive selection of either field or frame DCT coding
    - Efficiency depends on motion
  - Downloadable quantization Matrix



MPEG-4: Content Based Video Coding

- Considers syntetic and natural Videos
- Main Feature: Object based Compression



H.264 / AVC

- H.264 or MPEG 4 Part 10 Advanced Video Coding (AVC)
  - High coding efficiency
  - ~50% bit rate savings for similar quality relative to the performance of earlier standards
- Improved accuracy
  - Sub-Pixels
  - Variable block size
  - Multiple reference pictures for motion compensation (up to 15)

H.265 / HEVC

- High Efficiency Video Coding (HEVC), also known as H.265 and MPEG-H Part 2
- Supports up to 8K
- Areas up to 64x64
  - Realized via tree-Structures
- Pattern comparison
- New motion prediction approaches (range increased by 4 powers of 2)
- Optimized for high parallelization

Acknowledgement

- Thank you Wikipedia!
- A lot of Information taken from „Future Internet Applications“ lecture by Thomas Zinner

Thank you for your attention!