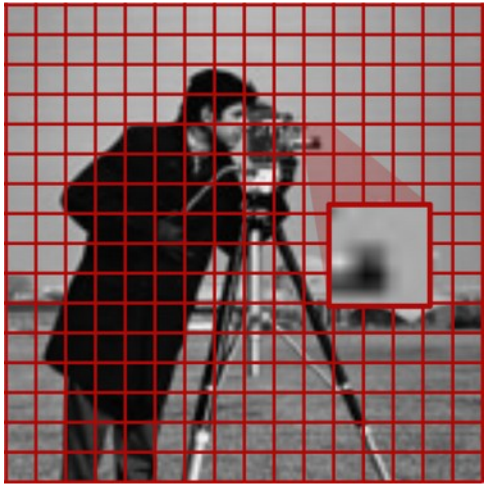
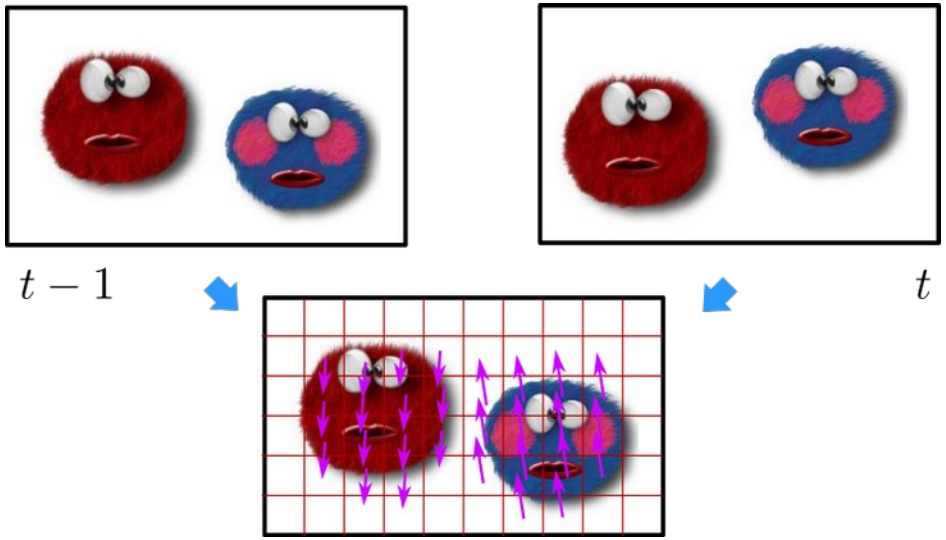
Encoding spatial correlation

- Processing 8x8 pixel blocks
  - Encoding using 2D DCT
  - Fast hardware implementation possible
- Motion JPEG
  - Simple to implement, well supported
  - Tolerance to rapid motion
  - Interframe encoding (each frame encoded separately)
  - Low compression ratios (e.g. 1:20, excellent quality 1:10)



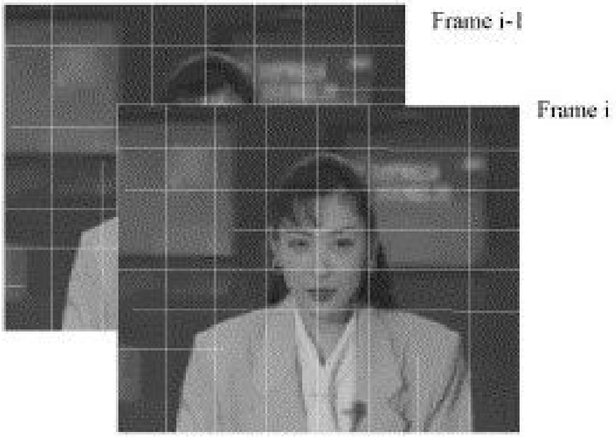
Encoding temporal correlation

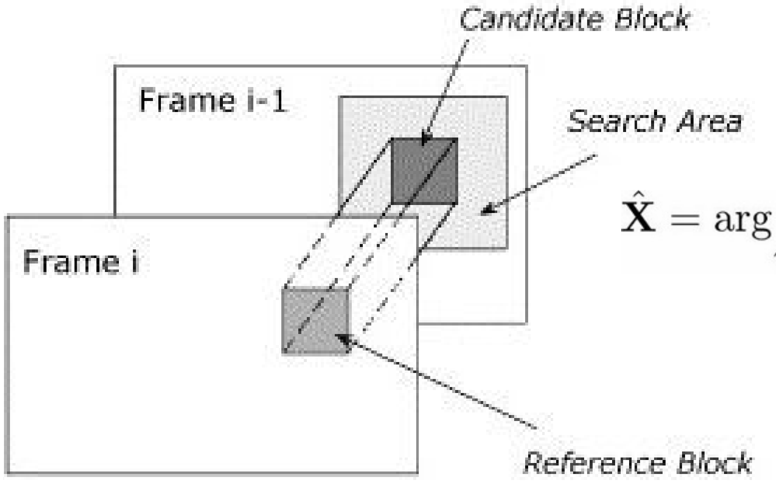
- Determining translation vector for each block
  - Translation for each block
  - Grouping blocks together (macroblocks)
- Frame reconstruction
  - Move blocks from previous frame
  - Visual difference encoded separately



Temporal correlation with blocks

- Determining block translation
  - Block matching method
  - Assume uniform motion within block
  - Searching for translation v, that minimizes distance between blocks





$$\hat{\mathbf{X}} = \arg\min_{v} \frac{1}{nm} \sum_{i,j} (x_{ij} - \hat{x}_{ij})^2$$

Video codec selection

- Compression level
- Losing information during compression
- Transfer speed (bit-rate) vs. distortion/loss
- Algorithm complexity
- Fixed / variable bit rate
- Communication channel characteristics
- Standard compatibility

Motion Picture Expert Group (MPEG)

- Working group that proposes standards for video and audio compression and transmission
  - Established in 1988 (ISO + IEC)
- Proposes MPEG-X phases with multiple parts
  - Each part covers a certain aspect of the whole specification
  - Some standards are later revisited and amended

MPEG parts

- MPEG-1: Video and audio for digital media
- MPEG-2: Generic video coding
- MPEG-4: Interactive video, audio, 3D graphics, Web
- MPEG-H: High Efficiency Coding and Media Delivery in Heterogeneous Environments
- MPEG-I: Coded Representation of Immersive Media

MPEG profiles and levels

- Standard defines wide range of applications
  - In many cases support for entire standard is unrealistic or expensive
  - Profiles and levels formalize support constraints
- Profile – subset of specification
  - Picture coding
  - Chroma format
- Level – limit parameters
  - Maximum resolution
  - Framerate
  - Bitrate

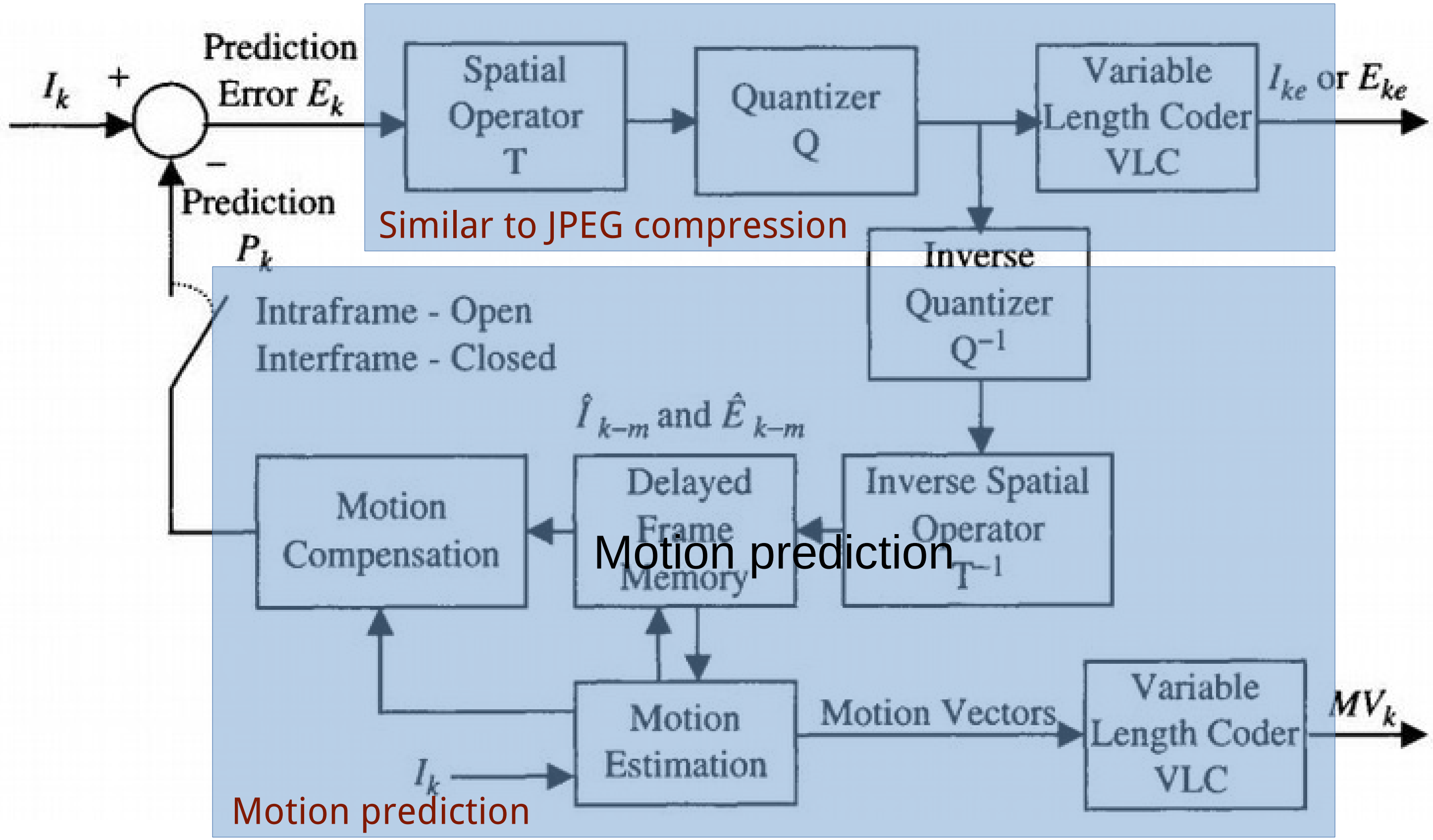
Video compression standards

Compression Standard	Publish Date	Application
H.261	1990	Video Conferencing, Video Telephony
MPEG 1 Part 2	1993	Video CD
H.262/MPEG 2 Part 2	1995	DVD Video, Blu-Ray, DVB, SVCD
H.263	1996	Video Conferencing, Videotelephony, Video on Mobile Phones(3GP)
MPEG 4 Part 2	1999	Video on the Internet
H.264/MPEG-4 AVC	2003	Blu-Ray, DVB, HD DVD
HEVC: *MPEG-H Part 2* ITU-T *H.265*	2013	Double data compression ratio compared to H.264
VVC: MPEG-I Part 3 ITU-T H.266	2020	Higher resolutions, omnidirectional, HDR

MPEG-1

- Audio-video storage and playback for VHS-quality video
  - Transfer rate ~1.5Mbit/s
  - Compression rate 26:1 (video), 6:1 (audio) without excessive loss
- Five parts: video, audio, compliance, reference implementation, system
- Asymmetric application
  - Compress once, decompress many times
  - Encoder can be complex, encoding slow
  - Decoded is simpler, decoding fast

MPEG compression overview

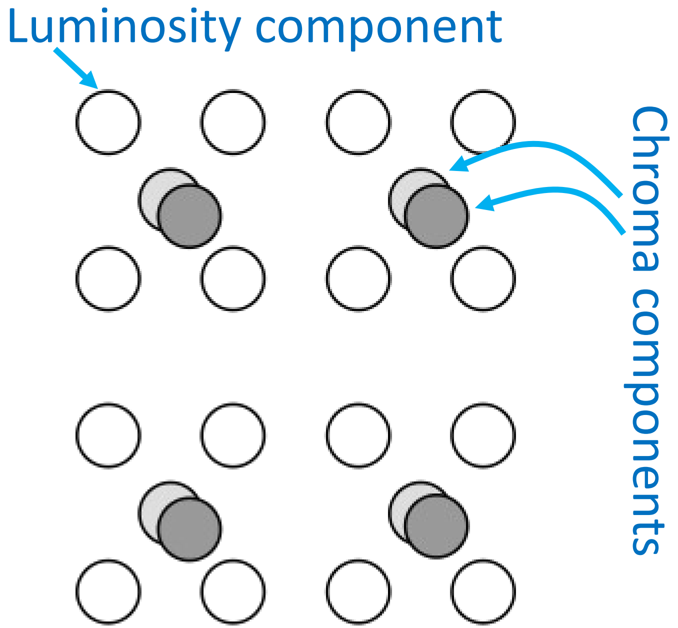


Requirements

- Normal playback with random access
- Support for video editing
- Reverse playback and fast playback
- Different resolution, frame rate
- Cheap hardware implementation of decoders

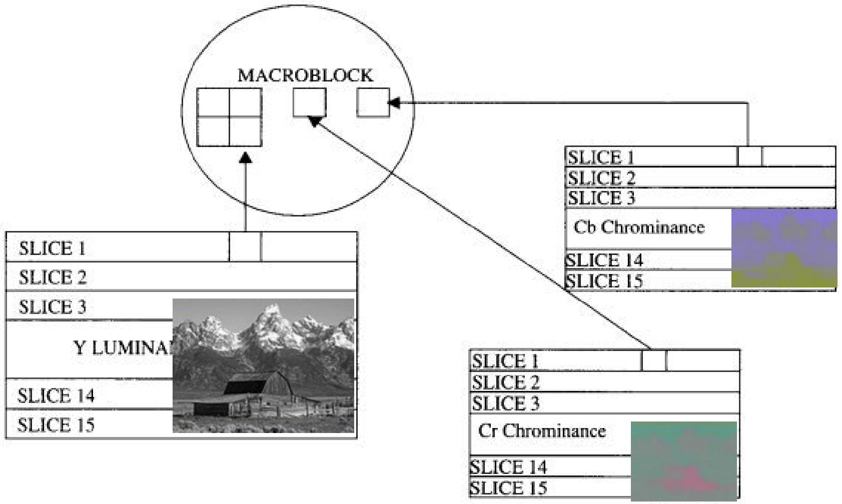
Video resolution

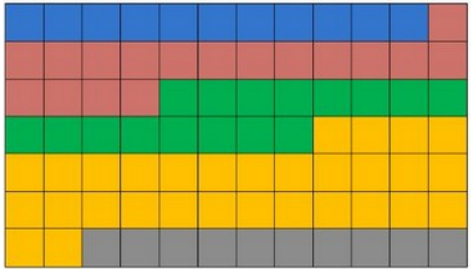
- Using YCrCb color space
  - Luminosity and two chroma channels
  - Chroma subsampling (Cr, Cb)
  - 8 bit values, 2:1:1 subsampling
- Common resolutions
  - 352x240, 352x288, 320x240
  - 352x240 (Y channel), 176x120 (Cr, Cb channels)



Macroblocks

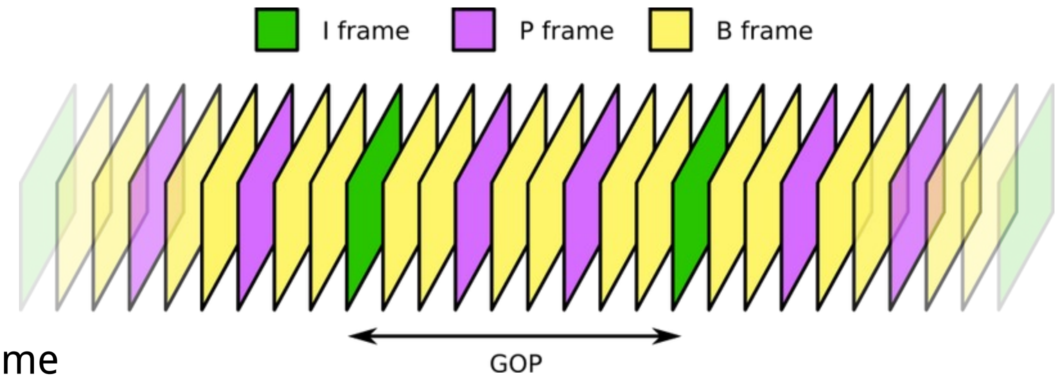
- YCrCb color space
  - Y channel - 16x16 blocks
  - Cr, Cb channels - 8x8 blocks
- Macroblock contains
  - Four 8x8 Y blocks
  - One 8x8 Cr block
  - One 8x8 Cb block
- Slice is sequence of macroblocks
  - Each slice can have different compression parameters
  - Easier error recovery (each slice encoded separately)





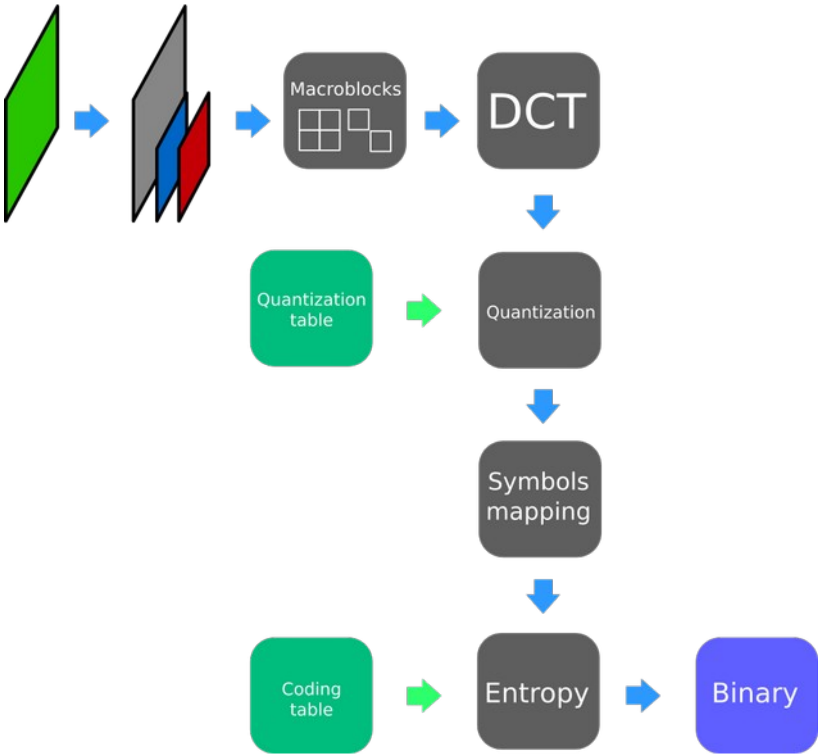
Frame types

- Group of pictures (GOP)
- Intra frame (I frame)
  - Encoded similarly than JPEG image
  - Approximately 12 frames apart
- Predictive frame (P frame)
  - Encoded with motion from previous I or P frame
- Bi-directional frames (B frames)
  - Encoded with prediction from previous and next I or P frames
- Low-resolution frame (frame D)
  - Fast video navigation
  - Only DC coefficients of DCT
  - Rarely used



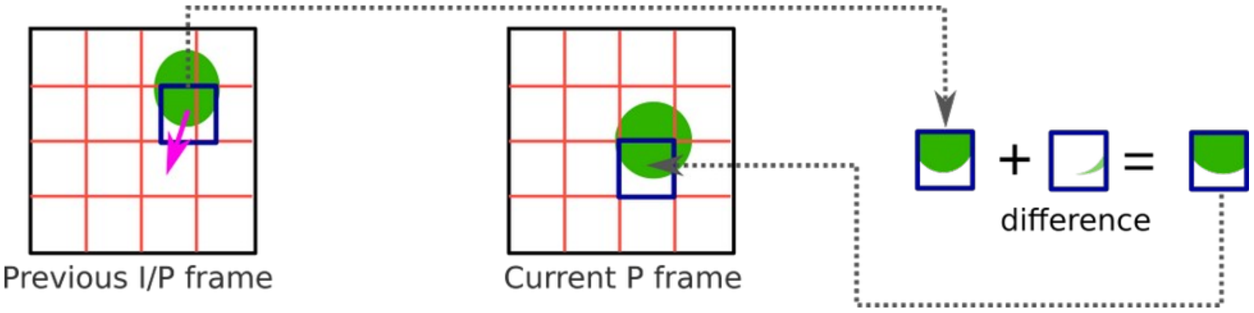
Type I frames

- Each macroblock is encoded statically
  - Chroma subsampling
  - Macroblocks
  - Similar approach than JPEG (different quantization tables)
- Enables random access to video content



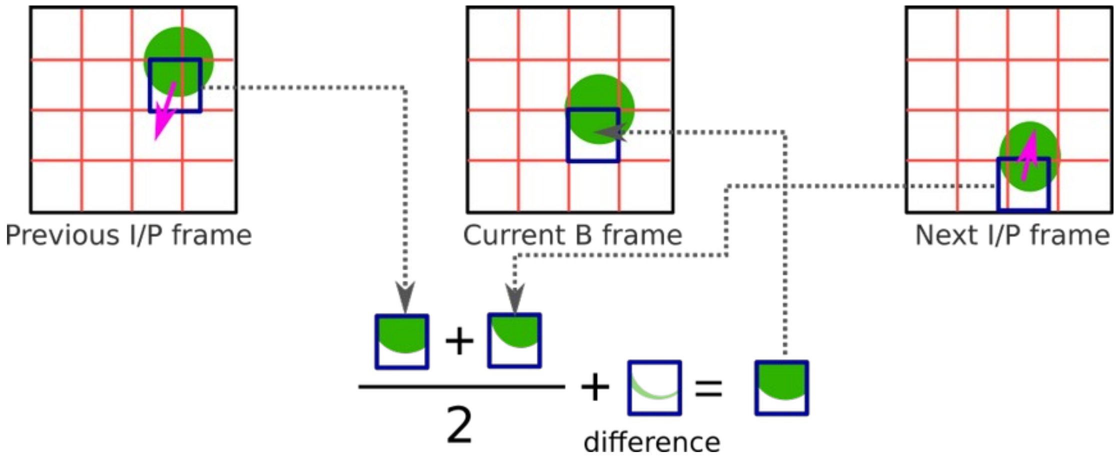
Type P frames

- Use information from previous I or P frame
  - Motion vector of each macroblock computed based on Y channel
  - Encode visual difference between the previous and current macroblock
  - If difference too big encode entire macroblock



Type B frames

- Encoded using previous and next I or P frame
  - Averaging motion-compensated images
  - Encoding only difference
  - Delayed decoding



Example of decoding a frame

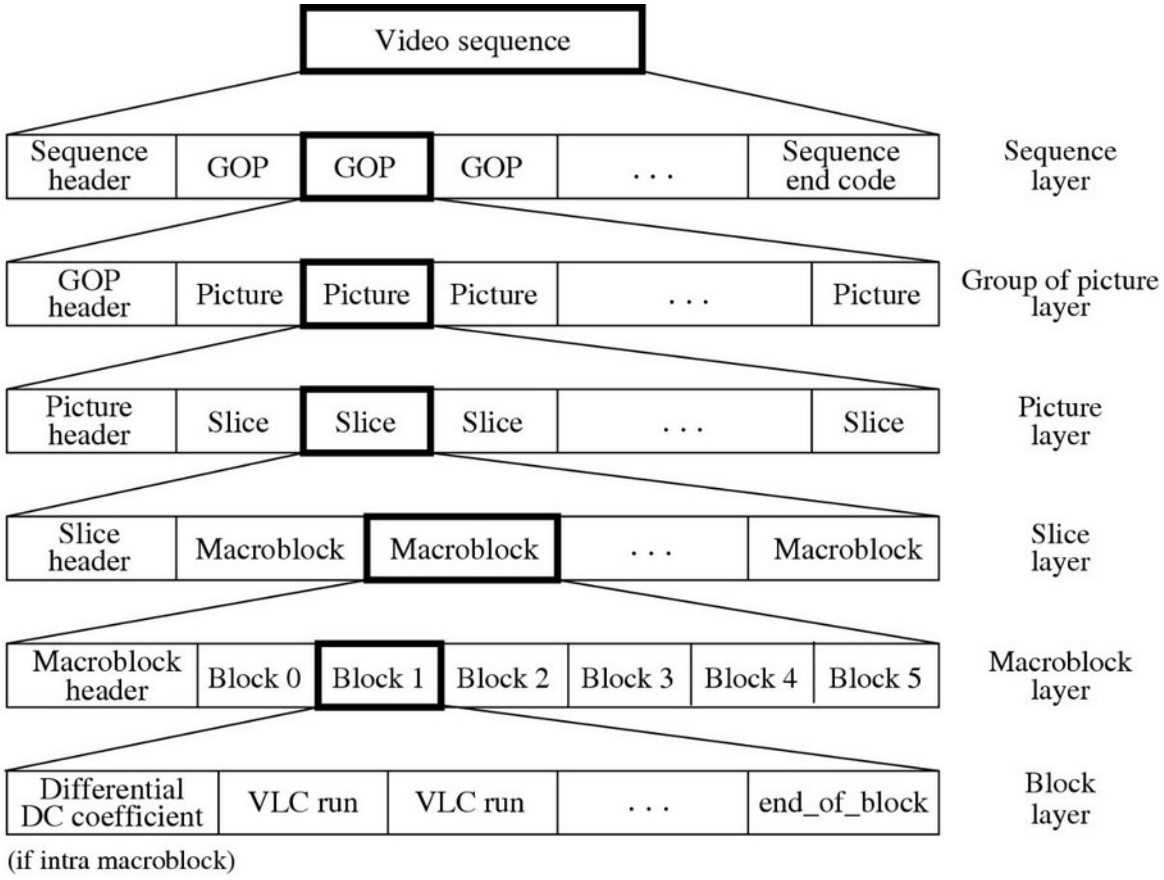
Our video stream contains frames

$$\{I_1, B_1, B_2, P_1, B_3, B_4, P_2, B_5, B_6, I_2, B_7, B_8, P_3, B_9, B_{10}, P_4, B_{11}, B_{12}, I_3\}$$

Which frames do we have to read to reconstruct the following frames

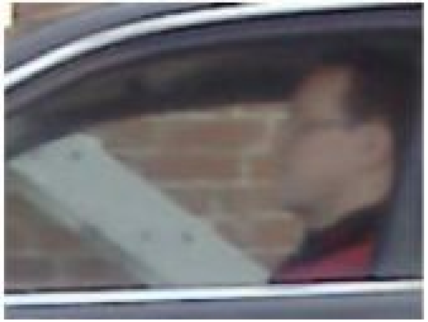
- I2 – only I2
- P2 – I1, P1, P2
- P3 – I2, P3
- B4 – I1, P1, P2, B4

Datastream hierarchy



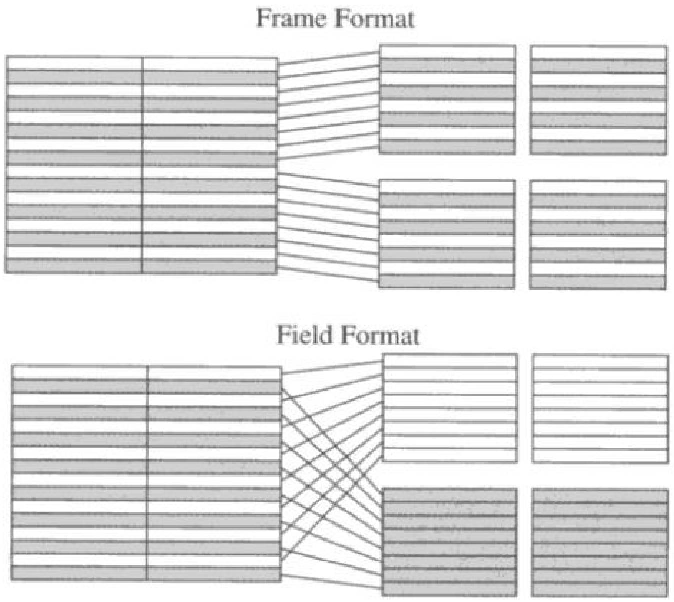
MPEG-2

- Higher resolution video (HDTV) for 4-15Mb/s channels
  - Suitable for digital broadcast TV
  - Support for interlaced video
  - Better robustness (errors)
  - Scalability (various bitrates)
- MPEG-2 decoders should also decode MPEG-1 streams
- Excellent quality compression rate 1:30
- Two interlaced video formats
  - Frame format
  - Field format



Field/frame format

- Frame format
  - Odd and even lines are encoded as single image
  - Image encoded using single header
- Field format
  - Odd and even lines encoded separately, every field encoded as separate image
  - Each field has its own header
- Interleaved format can switch between two modes for each image
- Progressive mode uses only frame format



Scalability

- Different profiles for various applications and quality levels (e.g., DVD, recordings from two cameras)
- Video separated in multiple layers
  - Base layer: information for coarse image reconstruction
  - Enhancement layers: information for improving coarse image quality
  - Separate transmission of basic (high priority) and enhancement (lower priority) layers

Spatial and frequency scalability

- Spatial scalability (image size)
  - Base layer: Images encoded with lower resolution
  - Enhancement layers: Missing high-resolution information, base layer used for prediction
- Frequency scalability (DCT coefficients)
  - Base layer: lower-frequency DCT (e.g. DC + motion vectors)
  - Enhancement layers: higher frequencies DCT (AC)

SNR and temporal scalability

- SNR scalability (signal-to-noise ratio)
  - Base layer: Heavily quantized intensity, coded in original resolution
  - Enhancement layers: residual information
- Temporal scalability
  - Base layer: images with lower framerate (e.g. I frames)
  - Enhancement layers: missing images in between, using motion prediction to reconstruct them (e.g. B frames)

MPEG-4

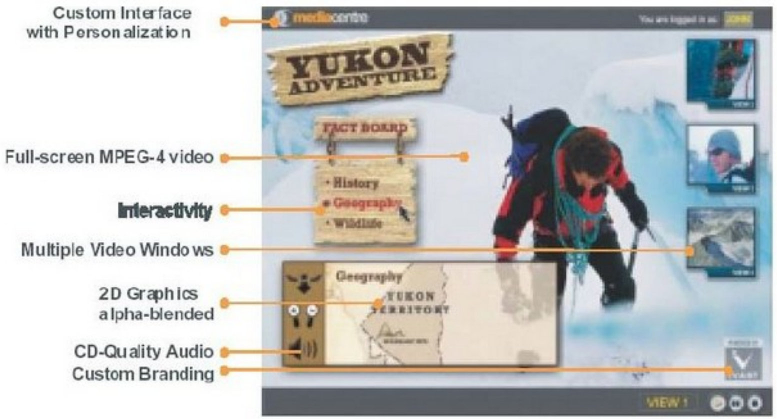
- Focus on interactive multimedia content
  - Better compression rate with higher throughput (bitrates).
  - Robust in environments with frequent errors
- Object oriented content representation – media objects
  - Interactivity on object level
  - Random access to objects

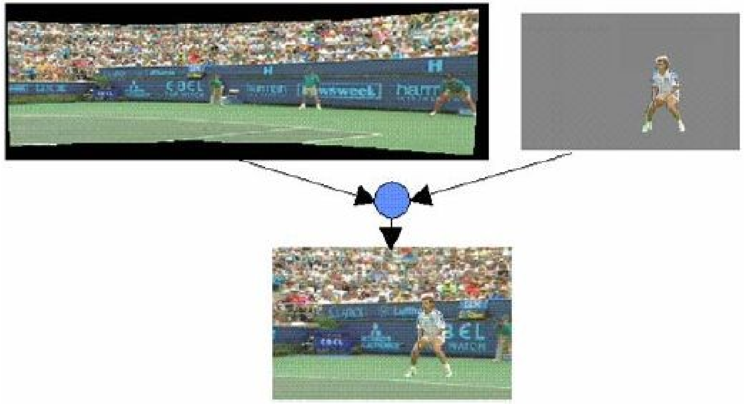
Media objects

- Different media sources
  - Real and synthetic images, sound, graphics
  - Interaction, manipulation, indexing, retrieval
- Hierarchical organization of media objects
  - Objects form dynamic/composed scenes
  - Objects separated in channels and encoded separately
  - Objects are transferred in separate streams
  - Composition occurs on the decoding device
  - Scene description language (BIFS)

Scene hierarchy

- Video-object Sequence (VS)
  - Complete visual scene
- Video Object (VO)
  - Arbitrary non-rectangular shape
- Video Object Layer (VOL)
  - Scalable coding support
- Video Object Plane (VOP)
  - Snapshot of VO at time t
  - MPEG1, 2: Entire frame is one VOP
- Group of Video Object Planes (GOV)
  - Optional group of VOPs





H.263/MPEG 4 Part 2

- 21 profiles – various applications
- Quarter Pixel motion compensation (Qpel)
  - Quarter pixel size for encoding motion vectors
  - Interpolation of blocks
- Global motion compensation (GMC)
  - Motion of scene based on affine transformation
  - Each macroblock encoded using global or local transformation
- Notable implementations: DivX and Xvid

H.264/MPEG-4 Part 10

- First version finished in 2003
  - Intended for storing high-definition video (Blu-Ray)
- Most widely used video codec
  - Used for HDTV, DVB-T
  - Included in Adobe Flash Player, Microsoft Silverlight
  - De-facto standard in HTML5
- Increased compression against MPEG-2
- Number of profiles available, from video-conference to high-resolution stereoscopic streams

H.264 improvements

- Variable Length entropy coding of blocks
  - Context-adaptive variable length coding (CAVLC)
  - Context-adaptive binary arithmetic coding (CABAC)
- Flexible block size
- Motion compensation
  - Larger range
  - More accurate (Qpel)
- Intra-frame prediction
  - Prediction of blocks using neighboring blocks
- Signal-adaptive deblocking filters
  - Smoothing edges of blocks
  - Increases subjective quality

Excellent quality compression ratio 1:50

H.265 – High Efficiency Video Coding

- Part of MPEG-H phase (heterogeneous environments)
- Boost video streaming
  - Various block sizes, organized hierarchically (CTU)
  - Better intra- and inter-frame prediction
  - Parallel processing (WPP)
- 50% smaller streams than H.264
- Limited acceptance
  - Patents
  - Higher computational requirements (10x)

https://sonnati.wordpress.com/2014/06/20/h265-part-i-technical-overview/

H.266 - Versatile Video Coding

- Part of MPEG-I phase
- Resolutions from 4K to 16K
- Support for 360° videos
- High dynamic range support
- Auxiliary channels (depth, transparency)
- Variable frame rate (0 to 120 Hz)
- Encoding complexity up to ten times that of H.265

Evaluating video/image compression

- Qualitative evaluation
- Quantitative evaluation
  - Evaluation datasets (Kodak, Xiph, CDVL)
  - Evaluation measures (PSNR, SSIM)

Peak Signal-to-Noise Ratio (PSNR)

- Based on Mean Squared Error
  - Measured in dB
  - MAX = 255 for 8 bit image

$$
\begin{aligned}
PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\
&= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\
&= 20 \cdot \log_{10}\left(MAX_I\right) - 10 \cdot \log_{10}\left(MSE\right)
\end{aligned}
$$

Structural Similarity Index (SSIM)

- Perceived change in structural information
  - Luminance, contrast, structure
- Neighborhood pixels have strong dependencies
- MS-SSIM – multi scale variant
- Calculated on a fixed local window (m x n)

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. IEEE TIP 2004

PSNR and SSIM examples

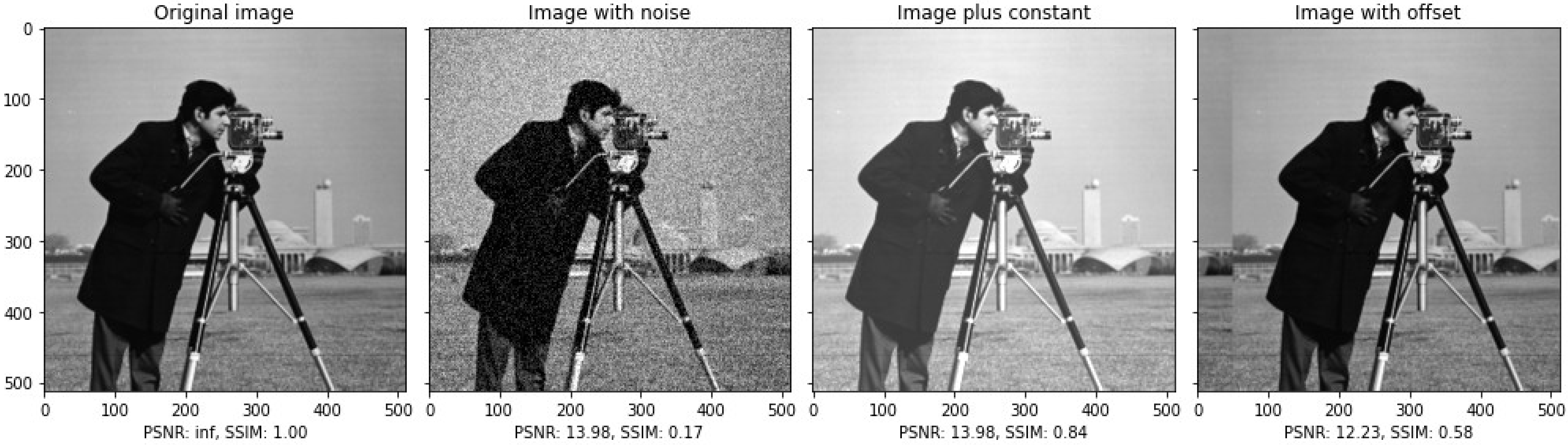
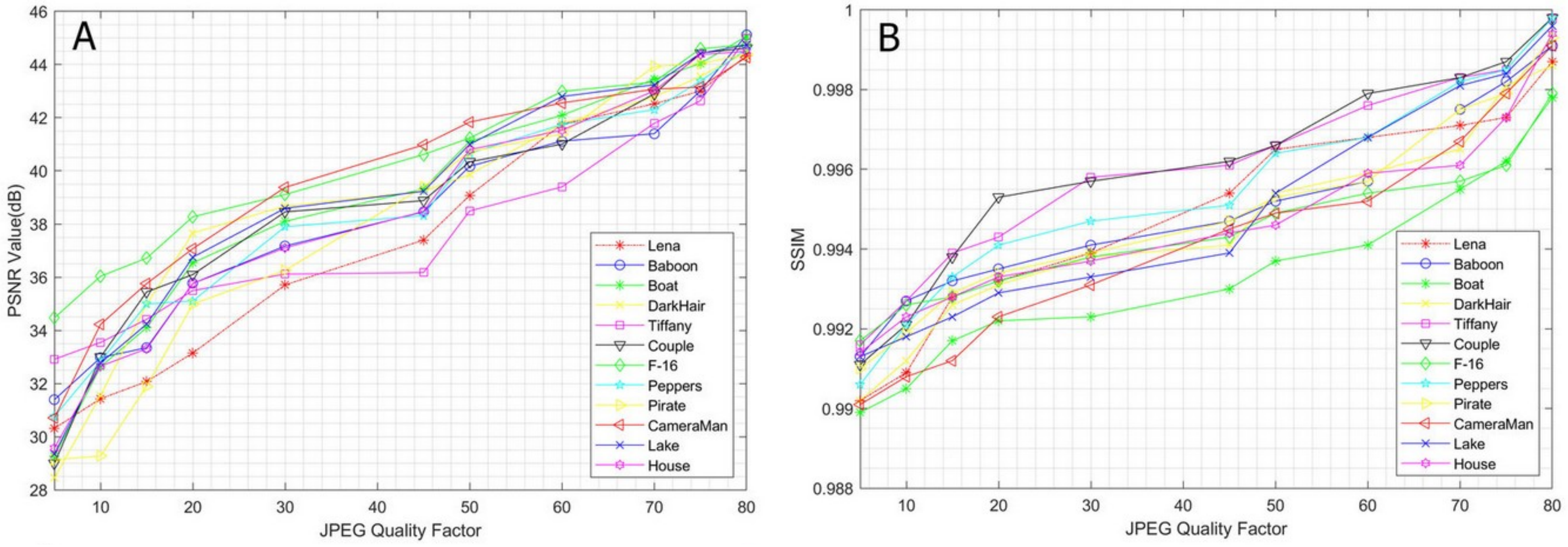
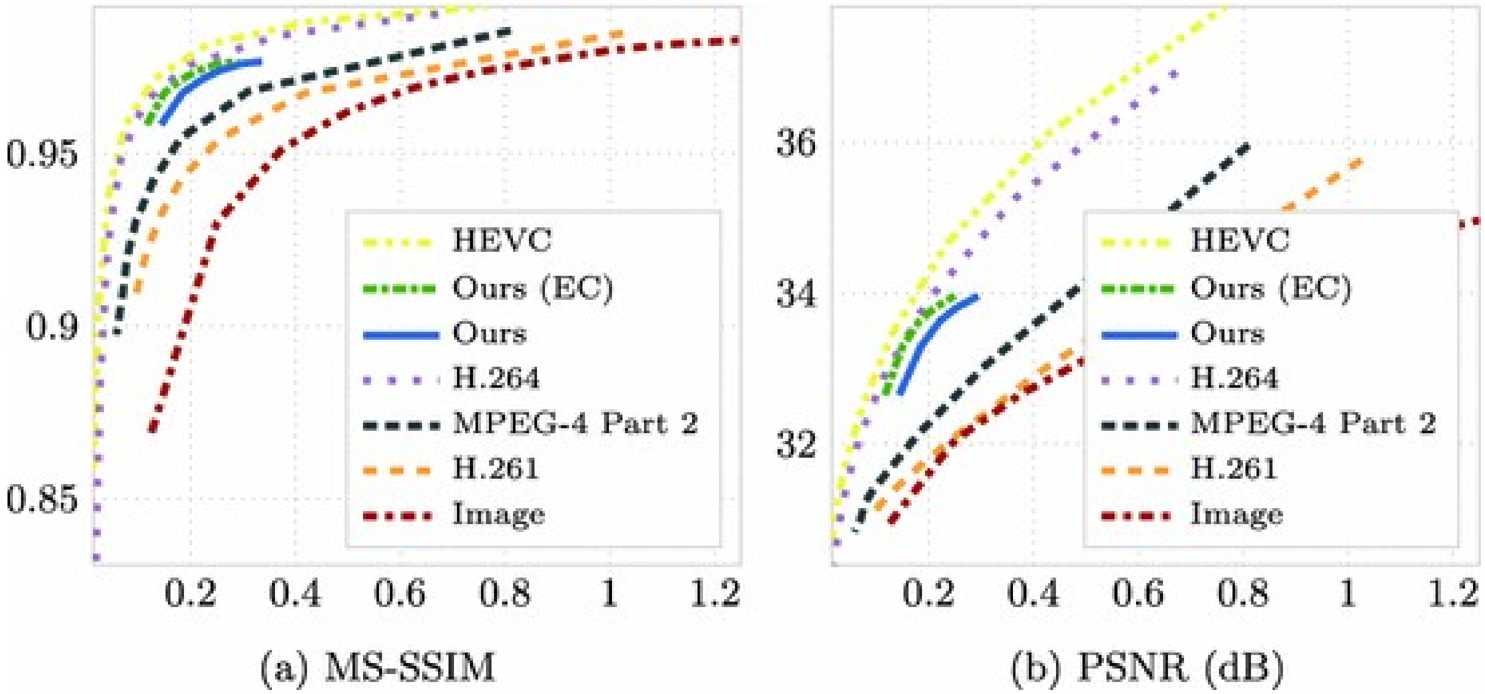


Image compression (JPEG)



Video compression comparison



Wu, Chao-Yuan, Singhal, Nayan, Krahenbuhl, Philipp Video Compression Through Image Interpolation, ECCV 2018