Video formats

David Bařina

April 10, 2013

Contents

1. Videa compression
2. Lossless formats
3. Lossy formats
4. Containers
5. Summary

Multimedia



- multimedia:

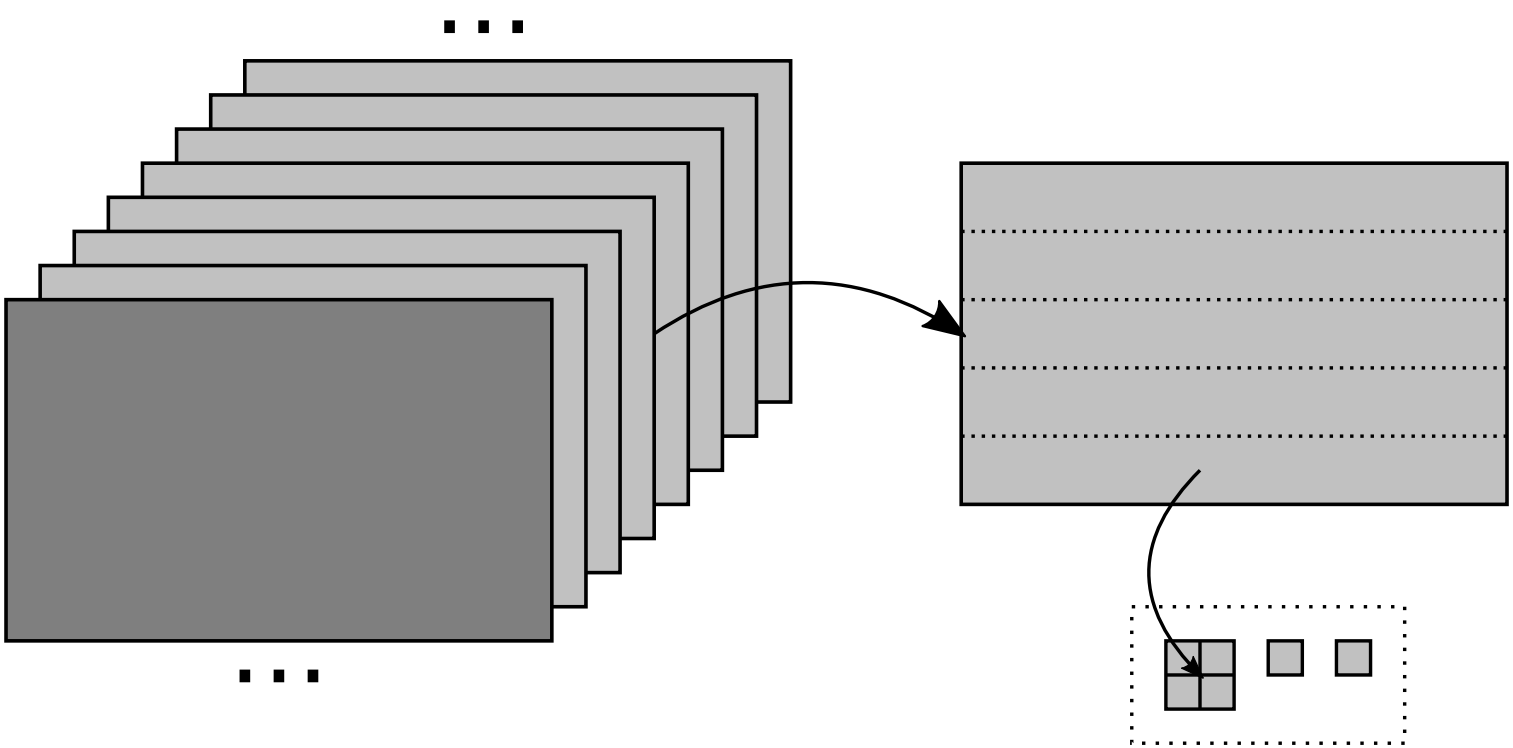
  text, audio, still image, **video**, metadata, ...
- needs:
  - acquire (camera),
  - store (hard drive, compression),
  - search (by description),
  - play,
  - edit (video editing), ...
- store: containers + codecs

Notions

- container
  - data streams/tracks (audio, video, subtitles, chapters)
  - e.g. AVI, Matroska, MPEG-TS
- video format
  - lossy, lossless compression
  - e.g. H.264, MPEG-4 ASP
- codec
  - compressor + decompressor
  - e.g. x264, DivX, libavcodec
- FourCC (4CC, four character code)
  - AVI
  - e.g. `DIVX` indicates DivX
- bitrate [bps]
- bitrate per segment:
  CBR (constant), VBR (fixed quality),
  ABR (long-term average, multipass encoding)
- postprocessing (deblocking)

Video

- 3D signal (x, y, t), color (Y, C_b, C_r)
- sequence of (similar) frames
- hierarchy:
  video, groups of pictures, frames, slices, macroblocks, blocks, samples

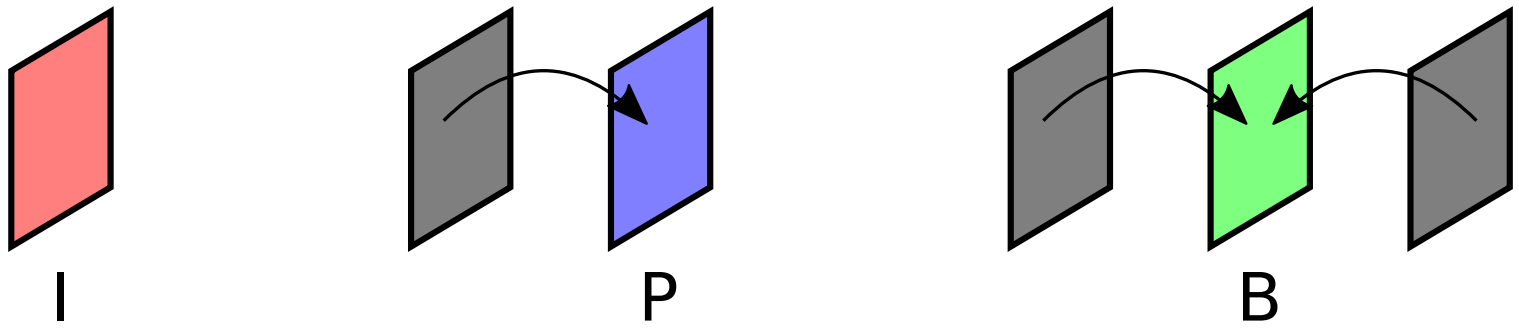


Frame

- type (key, delta)
- framerate [fps]
- timestamp
  - decoding time stamp (DTS)
  - presentation time stamp (PTS)
- index (by time, frame number)

Picture types

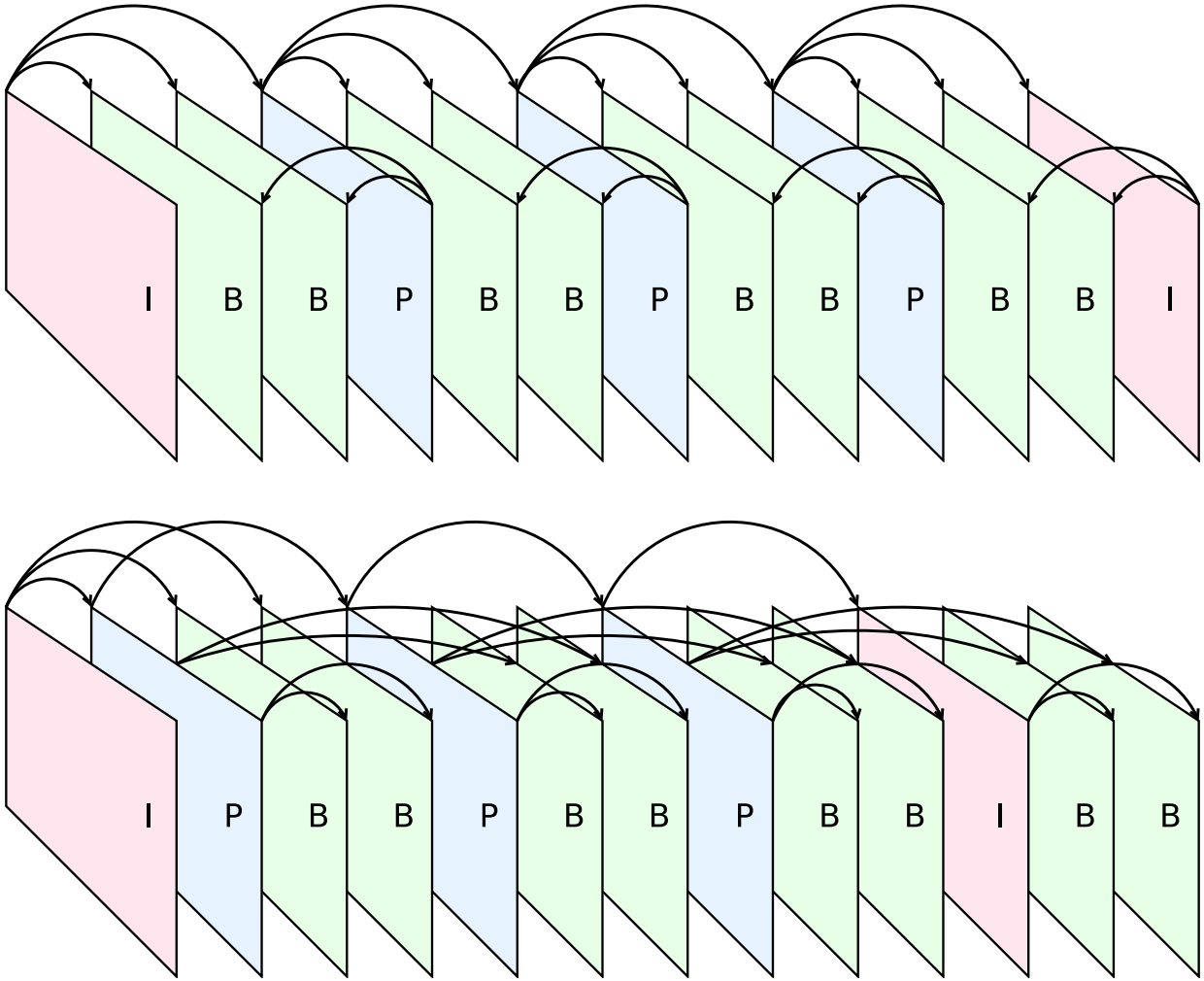
- key / intra / I frame
  - spatial redundancy
- delta / inter / P / B frame
  - temporal redundancy
  - B-frame problems
  - framedrop for B-frames
- other frames
  - D-frame
  - golden and altref frames
- frame reordering
  - B-frame after reference frames



Group of pictures

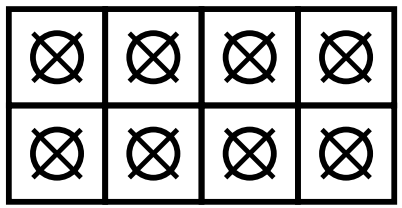
- GOP
- sequence of consecutive frames
- begins with I-frame
- open vs. closed GOP
- any frame decodable from its GOP (closed)
- error resilience
- GOP size
  - seeking takes long time
  - memory demanding
  - higher compression ratio
  - typically 12, max. 18

Group of pictures

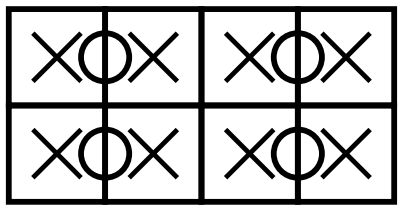


Chroma subsampling

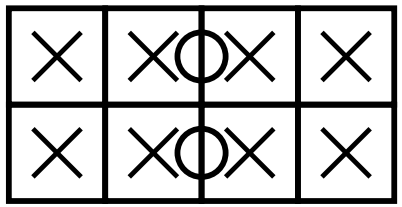
- Y vs. C_b, C_r
- unobservable by human eyes
- *J:a:b*, 4:4:4, 4:2:2, 4:1:1, 4:2:0
- centroids



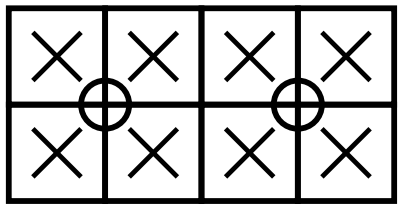
4:4:4



4:2:2



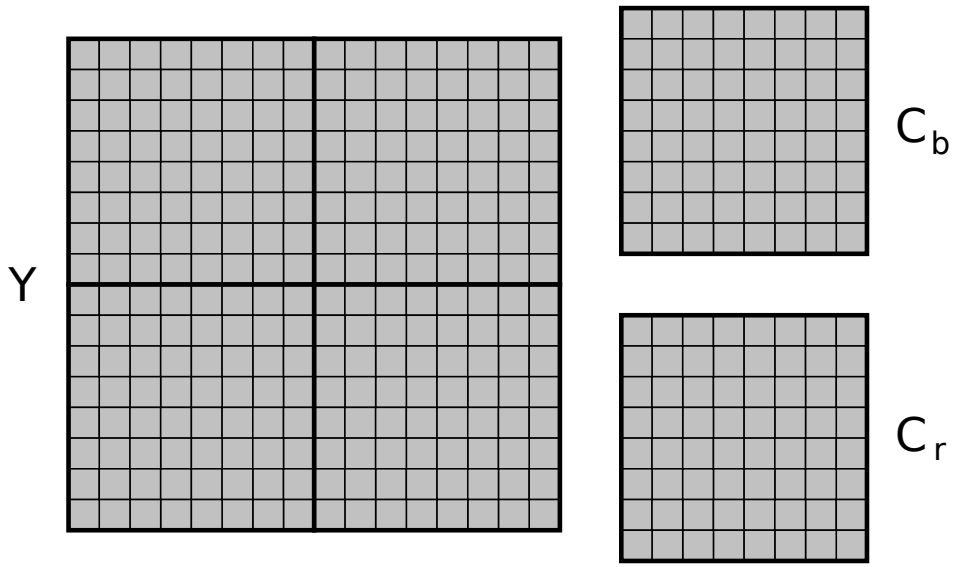
4:1:1



4:2:0

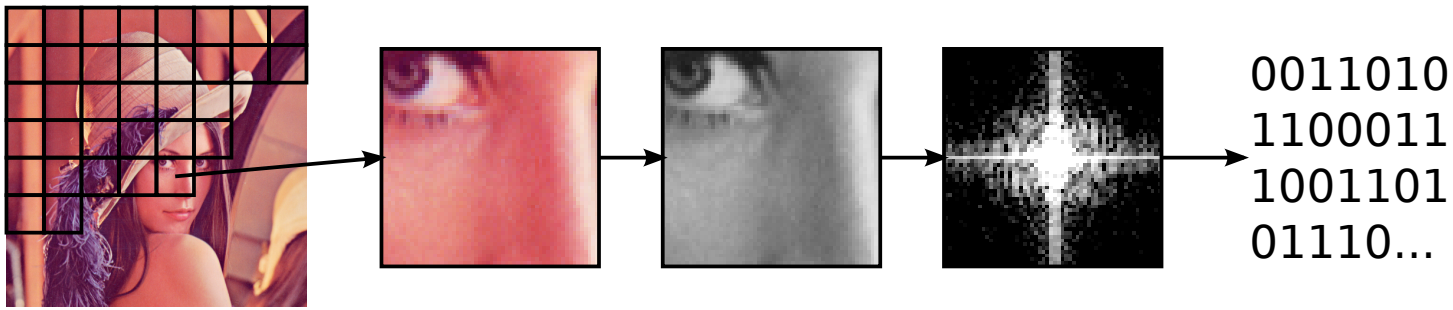
Blocks, macroblocks

- block $= 8 \times 8$ samples, DCT
- in newer standards smaller blocks (4–16)
- macroblock $= 6$ blocks (4:2:0)
  $2 \times 2\ Y + 1\ C_b + 1\ C_r = 6$
  $16 \times 16 + 8 \times 8 + 8 \times 8$
- macroblocks can be divided into subblocks (H.264)
- fundamental element of motion compensation

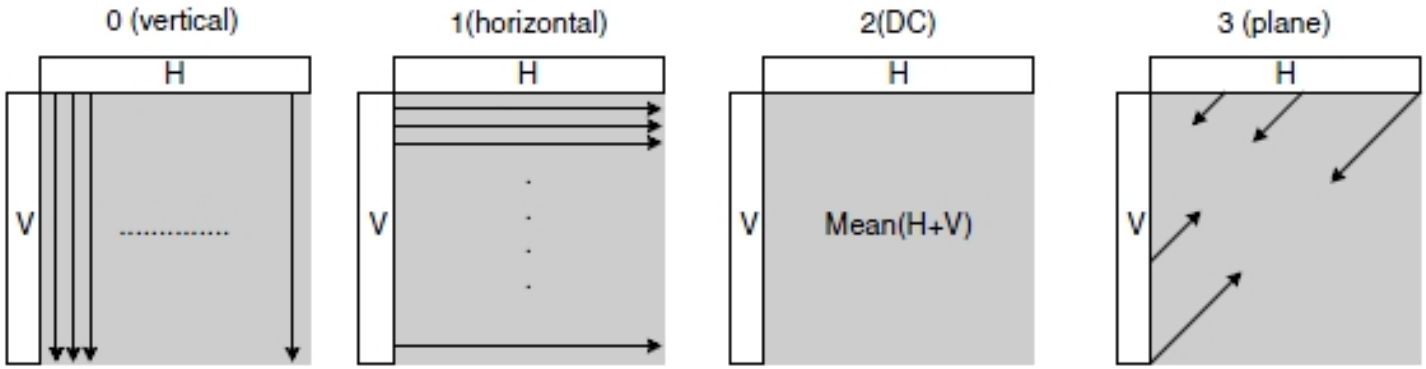


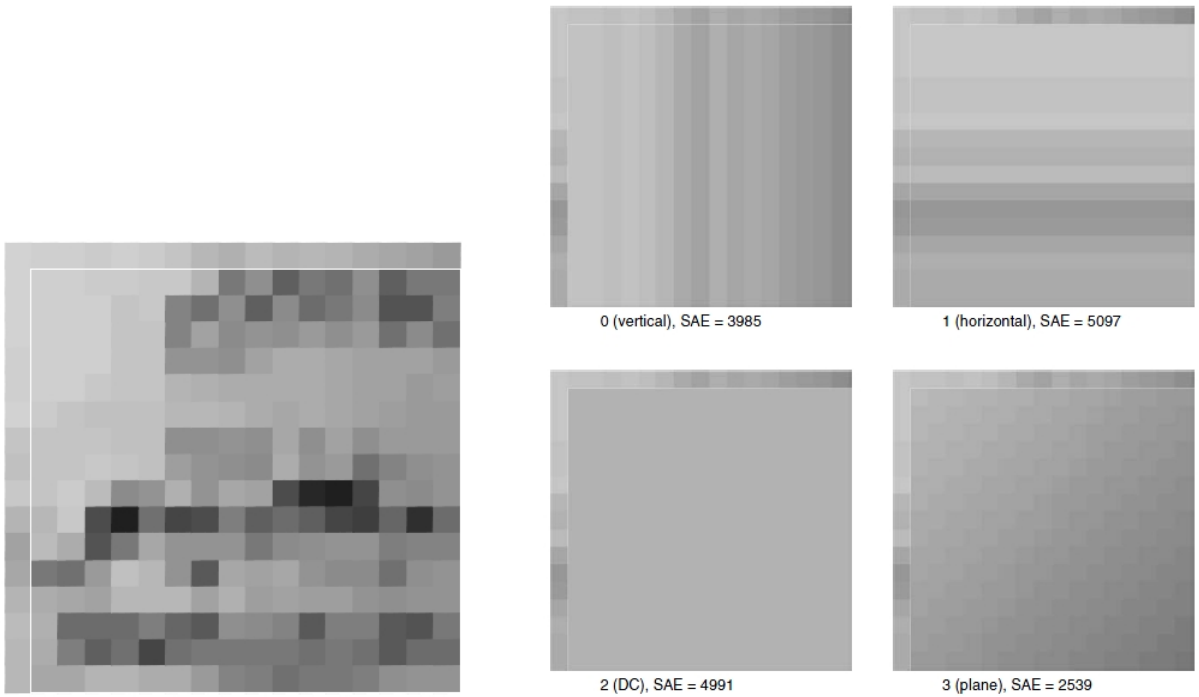
Intra-frame coding

- spatial redundancy
- like JPEG
- macroblocks, or parts of macroblocks
- prediction (several modes)
- transform (DCT, DWT)
- quantization (adaptive $q \cdot Q$), zigzag
- coding (RLE-0, VLC)



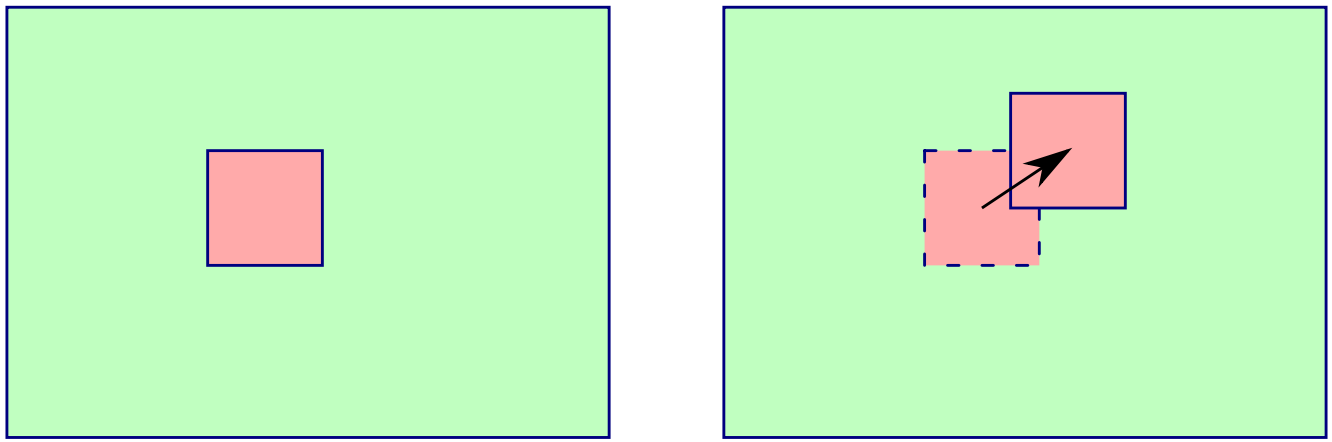
Macroblock prediction



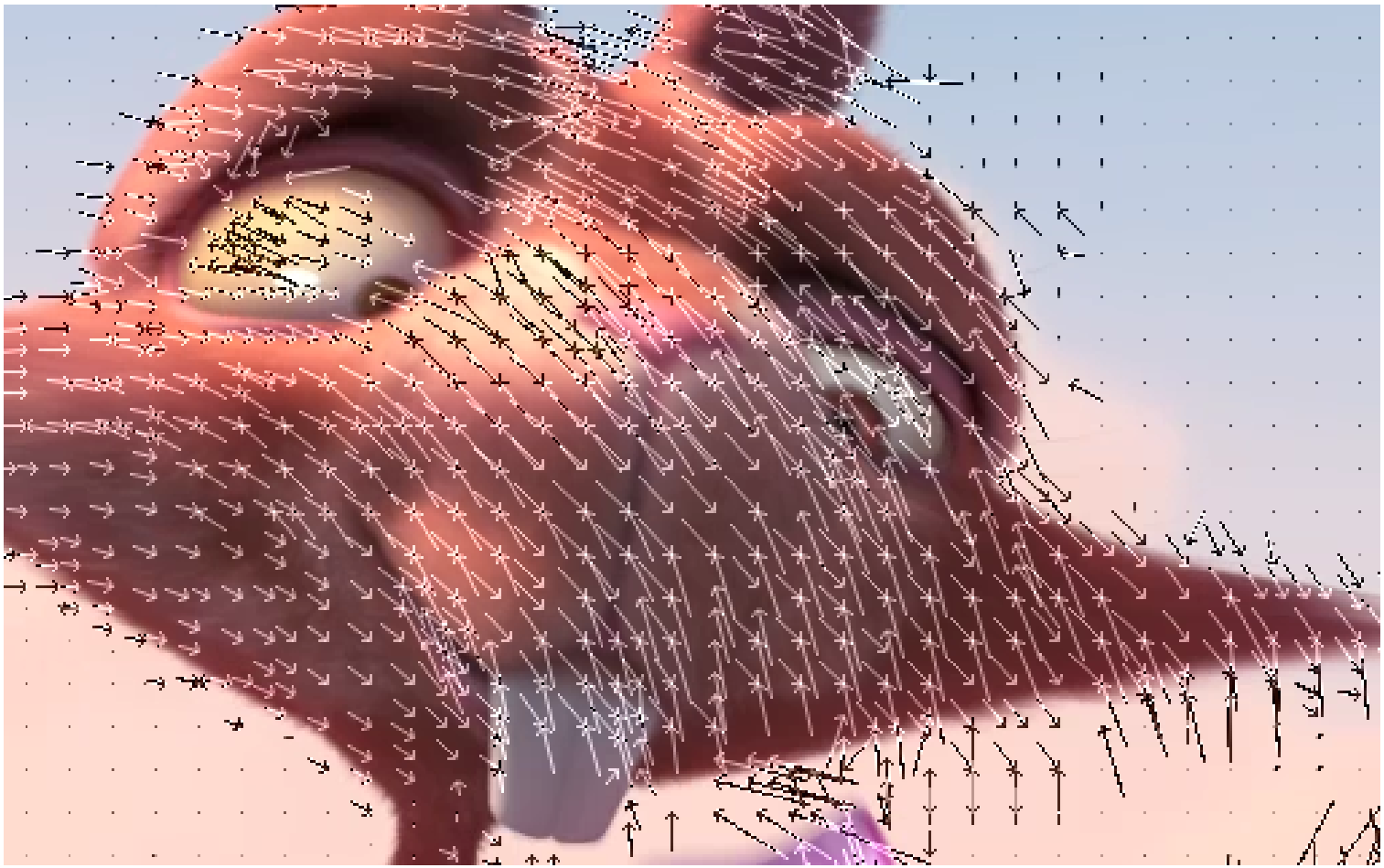


Motion estimation and compensation

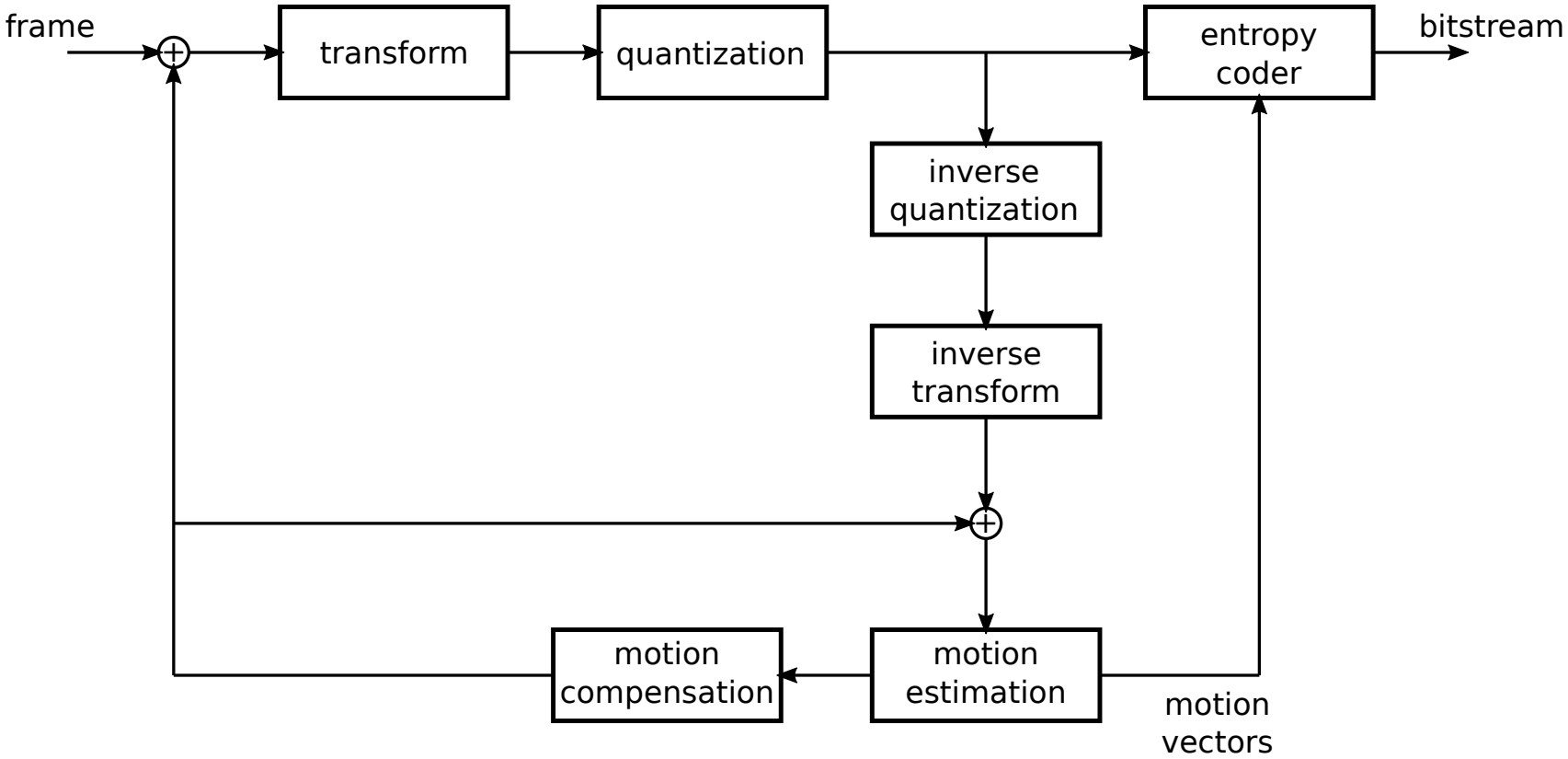
- temporal redundancy
- macroblocks, or parts of macroblocks
- motion estimation (ME)
- motion vector (MV) – accuracy (half-pel, Qpel), differentially encoded
- motion compensation (MC)
  - global (GMC) – affine transformation
  - block (BMC) – motion vector
  - overlapped block (OBMC) – computationally demanding



Motion vectors

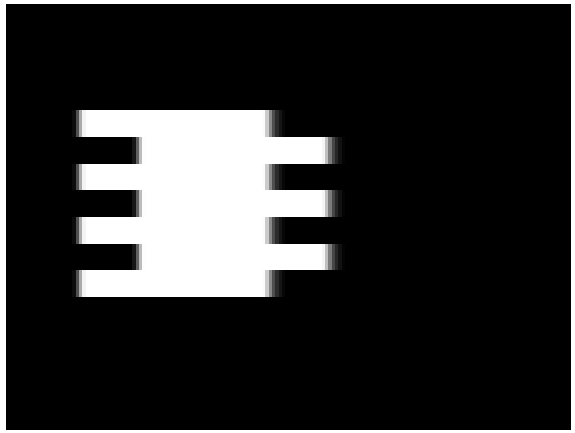


Hybrid coder



Interlacing

- only even or odd lines at one time
- visible artifacts on moving objects

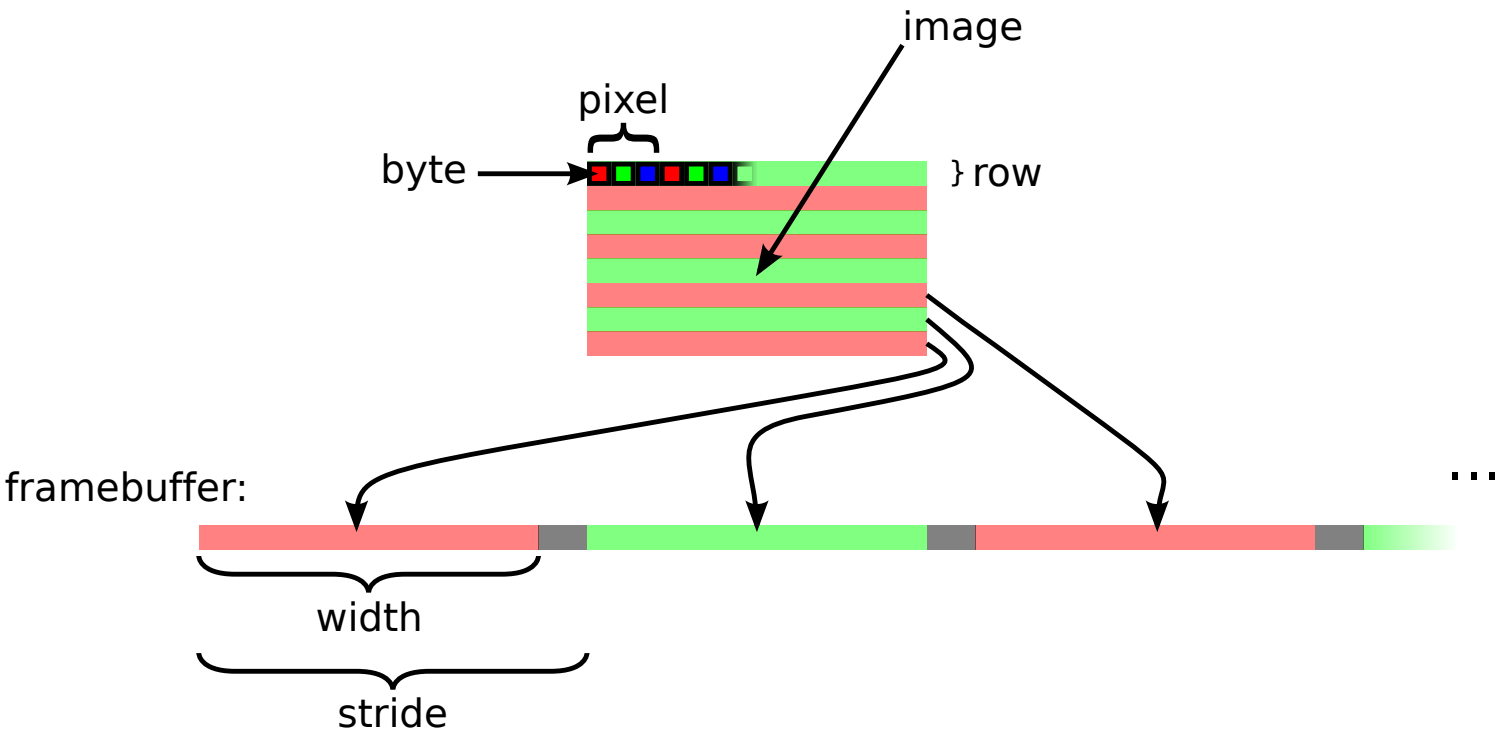


- deinterlacing (algorithms)
- digital video, MPEG
  - odd and even lines in separate pictures (field)
  - odd and even lines interlaced in single picture (frame)

Standards

- organization: ISO/IEC, ITU-T (CCITT)
- standardization: bitstream syntax + decoder
- profiles (algorithms)
  - encoder
  - decoder
- levels (technical limitations)

Frame memory organization



Uncompressed (RAW) frames

- FourCC: `0x00000000`
- uncompressed DIB, BGR24

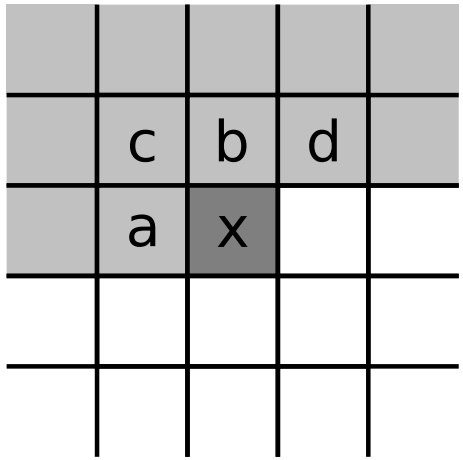
Microsoft RLE

- FourCC: `mrle`
- MSDN: 4, 8 bpp; FFmpeg: 16, 24, 32
- pairs of bytes A, B

A	B	meaning
0	0–2	escape code (end of picture)
0	3–255	copy next B bytes
1–255	0–255	fill next A bytes with value B

Huffyuv

- fast, GPL, fork Lagarith
- FourCC: `HFYU`
- predictor + entropy coder
- predictors: left a, plane p, median $\mathrm{med}(a, b, p)$



$$p = a + b - c$$

- color model: $R - G$, G, $B - G$
- each channel has its own Huffman table (static)
- Lagarith: predictor (median), RLE, arithmetic coder, null frames

MJPEG

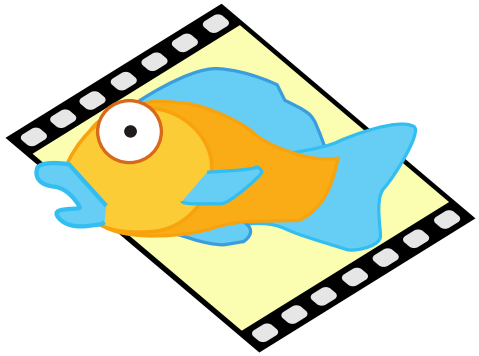
Motion JPEG

- not a standard, incompatibility, cameras, many FourCCs
- I-frames, JPEG
- also Motion JPEG 2000

VP3

- short specifications (about 1000 lines), 2004, 4CC `VP30`, `VP31`
- YCbCr 4:2:0, 8×8 DCT, kvantization, RLE-0, zig-zag, DC differences, Huffman coding
- block motion compensation (BMC), golden(I)/inter(P)-frames
- fragment = block 8×8
  superblock = $4 \times 4 = 16$ fragments (channel independent)
  makroblock = blocks 2×2 Y + 1 U + 1 V = 6 (over channels)
- Hilbert curve order for blocks in superblock
- golden-frames: like JPEG
- inter-frames: 8 modes for each macroblock
  (I-frame, difference from previous or I-frame, motion vector)

Theora

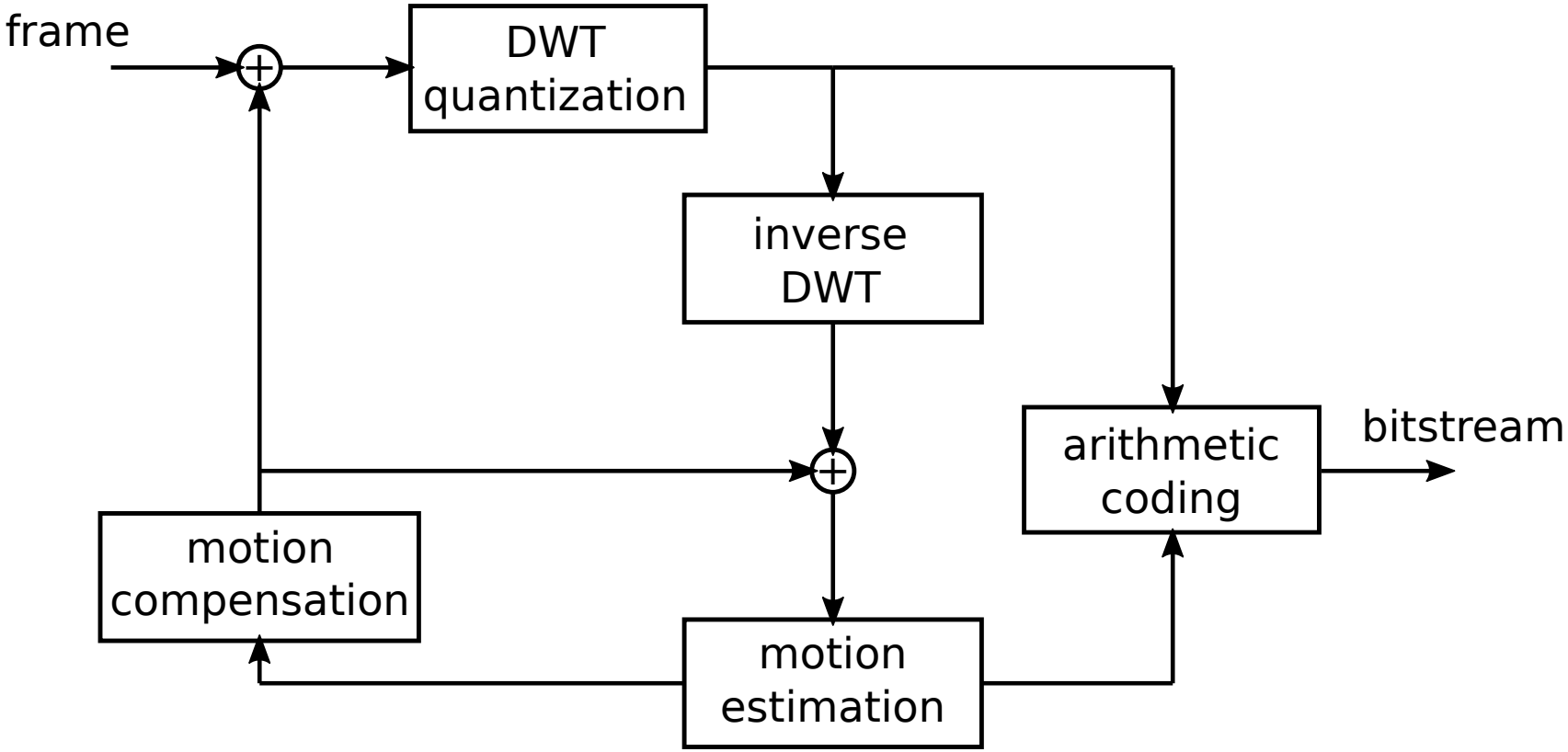


- for Ogg, Xiph.org Foundation, FourCC `theo`
- based on VP3
- I/P-frames only
- 8×8 DCT, chroma subsampling, BMC, max. 8 bits per channel
- libtheora

Dirac

- lossy, open source, BBC, FourCC: `BBCD`
- YCC models; interlacing; 4:4:4, 4:2:2, 4:2:0 subsampling
- DWT, choice of 7 wavelets
- implementation
  - Schrödinger in C (optimized)
  - dirac-research (originally Dirac) in C++
- GOP, I/P/B-frames, motion compensation (motion vectors)
- context arith. coder or exp-Golomb codes
- subset as VC-2 (Dirac Pro) – I-frames only
- no block effect as with the MPEG
- specifications 134 pages, 2008

Dirac


frame
DWT
quantization
inverse
DWT
arithmetic
coding
bitstream
motion
compensation
motion
estimation

VP8

- Google (On2), similar to H.264
- implementation libvpx, support in browsers (HTML5)
- exact DCT, WHT (Walsh–Hadamard transform)
- only 8-bit YUV 4:2:0
- each frame into macroblocks (16×16 Y, 8×8 U a V pixels), macroblocks in raster order
- macroblocks into 4×4 subblocks (16 Y, 4 U, 4 V), subblocks in raster order
- DCT/WHT for 4×4
  - DCT for subblocks of 16Y, 4U, 4V
  - WHT for DC coefficients of 4×4 DCT Y blocks
- no frame reordering
- boolean entropy coder = variant of arith. coder
- specification using sections of C source code; 303 pages, readable

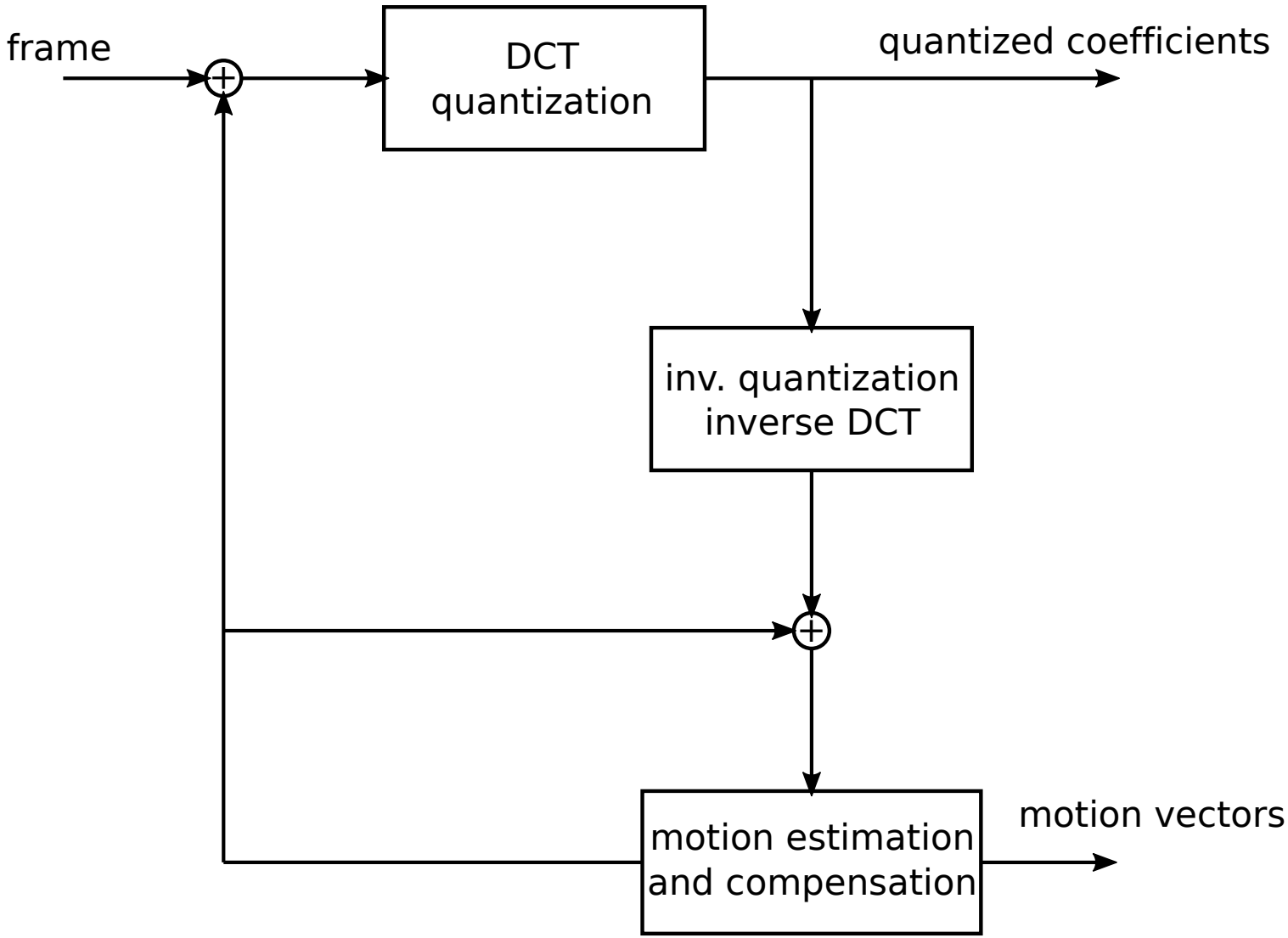
VP8

- only I(intra)/P(inter)-frames, no B-frames
- P-frames from any previous frame in GOP
- defines golden and altref frames: every I-frame is golden and altref
- any P-frame can replace last golden or altref frame
- macroblocks of P-frames are predicted from:
  previous, last golden, last altref, decoded part of current frame
- I(intra)-prediction: predict macroblock from already decoded row or column above or on the left; 4 methods
- P(inter)-prediction: motion vectors (Qpel), motion vectors for chroma subblocks as average from 4 corresponding Y subblocks
- frame decoding followed by loop filter – remove the blocking artifacts (do not confuse with postprocessing)

H.261

- ITU-T, 1988, 29 pages
- CIF, QCIF, YCbCr 4:2:0
- macroblock (16×16 Y + 8×8 Cb + 8×8 Cr)
- 8×8 DCT, kvantization, zigzag, EC (RLE-0 + VLC)
- inter/intra-frames
- motion compensation (MC)
  - one vector per macroblock
  - against the previous frame
  - motion vectors (integer only + possible blur)
- loop filter

H.261


frame
DCT
quantization
quantized coefficients
inv. quantization
inverse DCT
motion estimation
and compensation
motion vectors

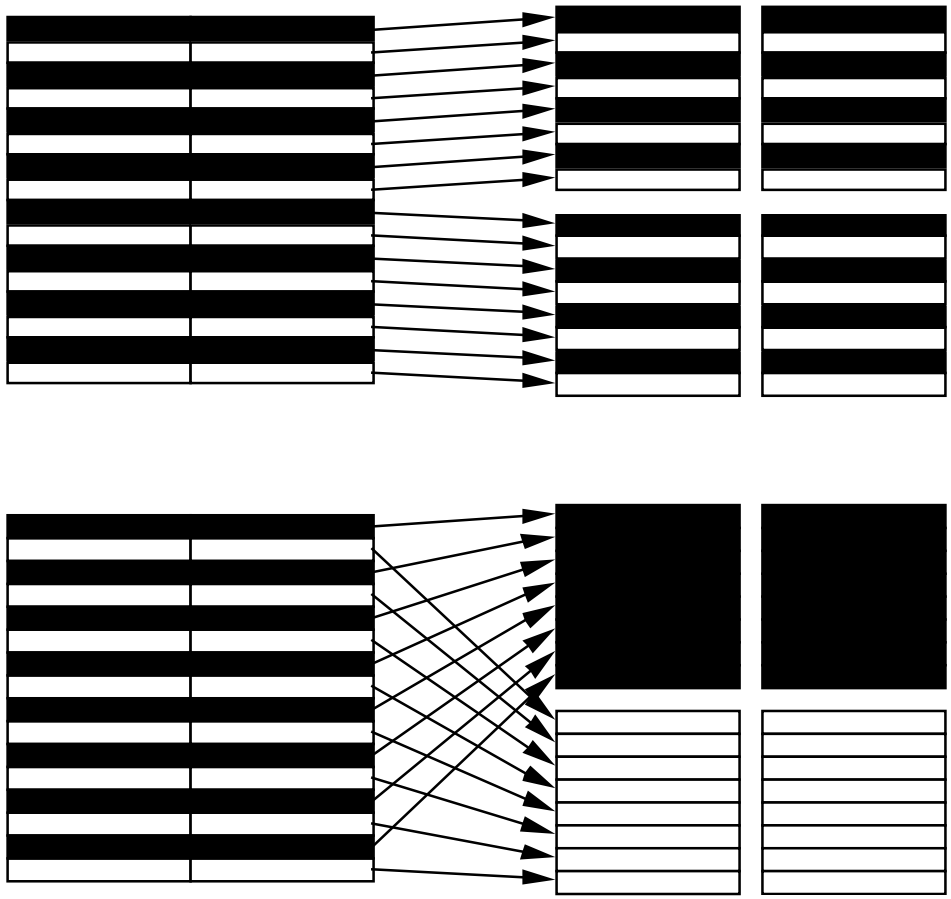
MPEG-1 Part 2

- ISO/IEC 11172-2, 1993, based on H.261 and JPEG
- VCD, (SVCD), (DVD)
- YCbCr 4:2:0
- I/P/B/D-frames
- GOP (15–18)
- I-frames like JPEG
- P-frames against I/P-frames
- motion vectors for each macroblock (half-pel)
- P-frames 1 vector / macroblock; B-frames 2 vectors
- DTS (decoding time stamps) in container
- D-frames DC coefficients
- blocks 8×8, macroblocks $16 \times 16 = 6$ blocks (4Y + Cb + Cr)

MPEG-2 Part 2 / H.262

- ISO/IEC 13818-2, 1995, improves MPEG-1, but more complex
- interlaced video
- 4:2:2, 4:2:0, 4:4:4 subsampling
- I/P/B-frames (not D)
- macroblocks, 1 motion vector (MV) per macroblock
- standard 243 pages
  - profiles (algorithms, image formats)
  - levels (bitrate, frame size)
  - e.g. SP do not allow B-frames
- SVCD, DVD, HD DVD, BD, digital TV (DVB-T)
- VBR

MPEG-2 Part 2 / H.262



MPEG-4 Part 2

- ISO/IEC 14496-2, 1999, based on H.263
- YCbCr 4:2:0, I/P/B-frames, DCT, MV
- no loop filter (!)
- standard: profiles, 517 pages, complex (3D shape)
  - SP: no B-frames
  - ASP: B-frames, interlacing, Qpel, GMC (global motion compensation)
- Qpel
- GMC
- many implementations (DivX, Xvid, 3ivx, FFmpeg), many 4CCs

MPEG-4 Part 10 / H.264

- ISO/IEC 14496-10, ITU-T H.264, AVC (Advanced Video Coding)
- 2003, many new features
- 4CC `AVC1`, `H264`
- 4:0:0 (monochromatic), 4:2:0, 4:2:2, 4:4:4
- BD, HD DVD, digital TV (DVB – HD stations)
- I/P/B-frames, DCT a WHT (Walsh–Hadamard transform)
- WHT for DC coefficients
- loop filter
- standard 680 pages

MPEG-4 Part 10 / H.264

- various sizes of blocks
- for prediction of P/B-frames up to 16 reference frames (instead of 1 or 2)
- motion compensation for blocks 4×4 to 16×16
- macroblock divided into subblocks
- several motion vectors (MV) per macroblock (subblocks)
- weighted prediction between two reference frames
- intra-coded macroblocks predicted from surrounding blocks
- entropy coders: exp-Golomb; CABAC or CAVLC
- implementation: x264, FFmpeg, DivX, CoreAVC

Other formats

- H.265 / HEVC
  High Efficiency Video Coding, ISO/IEC 23008-2, MPEG-H Part 2
  in development, successor for H.264
- VP9
- WMV + VC-1
- ...

AVI (Audio Video Interleave)

- `.avi` extension
- Microsoft, since 1992
- RIFF (Resource Interchange File Format)
- to identify the codec uses FourCC (e.g. `DIVX` = DivX)
- cannot contain subtitles, chapters, menus (third party)
- formerly restriction on 1 GB
- index by frame number (not time) at the end of file (not suitable for streaming)
- wide support
- does not support B-frames

Matroska

- `.mkv` extension
- open format, announced 2002
- supports absolutely everything, unlimited number of tracks
- based on EBML (Extensible Binary Meta Language)
- codec identified by CodecID (text string)
  - e.g. `V_MPEG4/ISO/AVC` = H.264
  - code `V_MS/VFW/FOURCC` encapsulates FourCC
  - `http://haali.su/mkv/codecs.pdf`
- WebM (`.webm` extension) is profile Matroska format
  - Google, 2010, intended for the web (HTML5)
  - only VP8 + Vorbis codecs

MP4 (MPEG-4 Part 14)

- `.mp4` extension
- ISO standard, since 2003 (version 2)
- based on MPEG-4 Part 12, which is based on QuickTime
- to identify objects (also codecs) uses four-character identifiers
  - e.g. `avc1` for MPEG-4 AVC
  - code-points, `http://www.mp4ra.org/codecs.html`
- supports any codec (priváte or registered)

ASF (Advanced Systems Format)

- `.asf`, `.wma`, `.wmv` extensions
- Microsoft
- objekts (also used codecs) identified by GUID
  - encapsulates FourCC, e.g.
    `DEFINE_GUID(CLSID_XVID, mmioFOURCC('D','I','V','X'), ...`
- theoretically any codec
- index for keyframes, can be by time, frame number or ref. timestamp

QuickTime, Ogg

QuickTime

- `.mov` extension
- Apple, also QuickTime File Format (QTFF)
- MPEG-4 Part 12 is based on QT

Ogg

- `.ogv`, `.oga`, `.ogx`, `.ogg` extensions
- open format
- part of the Ogg project, Xiph.Org Foundation
- formerly in the HTML5 specification
- hack OGM (Ogg Media)

FLV, 3GP

Flash Video

- `.flv`, `.f4v` extensions
- Adobe Systems (originally Macromedia)
- two versions: FLV (SWF) and F4V (MPEG-4 Part 12)
- limited codec support

3GP, 3G2

- `.3gp`, `.3g2` extensions
- based on MPEG-4 Part 12
- for mobile devices
- limited codec support

MPEG

MPEG-PS

- ISO, specified in MPEG-1 Part 1 and MPEG-2 Part 1
- VOB, EVO (with limitations) – DVD-Video and HD DVD
- limited codec support + private streams
- ES (audio, video), PES (packets),
  PS (combines multiple PES – e.g. audio a video)

MPEG-TS

- ISO, 1995, specified in MPEG-2 Part 1
- can contain multiple PS (Program Streams)
- limited codec support + private streams
- broadcasting
- TS encapsulates several PES

Summary

- notions: format, codec, container, FourCC, frame, GOP, macroblock
- methods: intra-coding, inter-coding, hybrid coder
- video formats:
  MS RLE, Huffyuv,
  VP3/Theora, Dirac, VP8, H.261, MPEG-{1,2,4}, H.264
- container formats:
  AVI, MKV, MP4 (ISO, 3GP, MOV), MPEG-{PS,TS}