Image Compression

Universidad Politécnica de Madrid 2010

Dr.-Ing. Henryk Richter
henryk.richter@comlab.uni-rostock.de

University of Rostock
Institute of Communications Engineering
http://www.int.uni-rostock.de
basics

- correlation of neighboring images in video sequences induces redundancy in time direction \(\rightarrow\) inefficient single image coding
origins of differences between images

- complex motion of camera and objects (a)
- changes in illumination
- scene changes
- disturbances (noise, recording errors, aliasing)
- changes in object topology (b)
- revelations and occlusions on objects and frame borders (c)

general observation: video sequences are in-stationary
techniques for decorrelation along the time axis

**spatial domain: correspondency methods**
- DPCM (special case of motion compensation for stationary images)
- block based local motion compensation
  - FSBM (fixed size block matching)
  - VSBM (variable size block matching)
  - OBMC (overlapping block motion compensation)
  - Object based block matching motion compensation

**transform domain:**
- 3D Wavelet-transform [Schwarz 2000]
- MCTF - motion compensated temporal filtering [Schäfer 2005]
translational block motion compensation

- each block receives a motion vector (MV)
- motion vector is a displacement of a block to predict in relation to that block’s position within the reference image
- predicted block is gathered from the reference image and used as source for difference coding
block matching techniques (1)

**various search metrics**
- mean square error (MSE), sum of square error (SSE)
- sum of absolute differences (SAD) or mean absolute difference (MAD)
- discrete cross correlation

- additionally: motion vector length

\[
MAD(d_x, d_y) = \frac{1}{B_x \cdot B_y} \sum_{j=y}^{y+B_y-1} \sum_{i=x}^{x+B_x-1} \left| I(i, j, t) - I(i - d_x, j - d_y, t - \Delta t) \right|
\]

\[
MSE(d_x, d_y) = \frac{1}{B_x \cdot B_y} \sum_{j=y}^{y+B_y-1} \sum_{i=x}^{x+B_x-1} \left( I(i, j, t) - I(i - d_x, j - d_y, t - \Delta t) \right)^2
\]

⇒ for constant block size, normalization (mean) not necessary
⇒ SAD/MAD favorable on architectures with hardware SAD support (e.g. Intel MMX2)
block matching techniques (2)

**workflow**

- for each candidate position
  - calculate displacement dependent prediction error metric
  - select best match

1) **exhaustive search ("full search")**
- scan all candidate positions in area
- guarantees optimal result for large search window
- problem: computational complexity (here: 256 positions per compensated block)
- variation: spiral search
block matching techniques (3)

2) three step search (TSS)
- find best match in radius 4
- use best match for next iteration with radius 2
- use best match as start for radius 1 to obtain final result
- 25 search positions but unreliable, small search window
- variation: logarithmic search

3) hexagon search (similar to TSS)
- as long as best match is not in the middle of the pattern, use best match as start for next iteration (ex. steps 1-3)
- in each iteration, only three new positions need to be calculated
- last step is radius 1 local search (ex. step 4)
- good small area results, heavily dependent on start MV, number of iterations not pre-determined
4) predictive search
- use predicted MV(s) from spatial/temporal neighbors
- perform small area exhaustive search around predicted MV(s)
- good complexity/reliability tradeoff, can cover large translations despite smaller search window
- variations: MVFAST, PMVFAST,

5) search patterns
- algorithms can be combined with different patterns of varying size

6) hierarchical block motion estimation
- calculate resolution pyramid
- use best vector from lowest layer for following layer
block motion compensation issues

- additional problems:
  - aperture problem
  - periodicity problem
  - block artifact problem, block structure not related to image content

- in most cases, true motion not required, just a good match

- optimize for smooth motion vector field
translational block motion compensation example

- **advantage of motion compensation is the drastic reduction of residual energy**
  - lower differences result in less significant data to be coded

- **issues:**
  - encoder complexity
  - in-stationary image parts, causality (e.g. frame borders)
  - motion vector transmission from encoder to decoder → tradeoff

---

frame difference w/o MC  frame difference with MC  motion vector field  block partitioning
sequences are divided into groups of pictures, where frames depend on each other.

frames can be transmitted as partial packets, called slices.

basic processing unit of current video coding standards is the Macroblock, consisting of Y, Cb, Cr (ex.: 4:2:0)
Intra-frames are independent intrinsic frames without reference to other frames
- ("I-frames", "Intra-frames" or "keyframes")

required for coding/decoding start and random access to stream
I-only sequences allow frame-accurate editing (DV, Motion-JPEG)
low memory requirements in decoder
drawback: low coding efficiency without temporal de-correlation
basic frame types: P-frames

- P-frames (Predicted frames) are motion compensation predicted from frames transmitted earlier within the current GOP
- motion compensation limited to single direction per block (“forward prediction“)
- contain mostly difference information related to previous frames
- no random access to P-frames
- decoding error invalidates all subsequent predictive frames (until next I-frame)
B-frames (Bi-directional frames) are motion compensation predicted in **two** directions

- motion compensation either forward, backward or bidirectional
  - bidirectional by averaging forward and backward predictors

- contain mostly difference information related to previous frames

- in most standards not used as reference frames (disposable)
  - provide cost-effective means of fast-forward, random-access
  - coded with less residual information

*display order*
drawbacks:
- to ensure causality, display order and coding order are different
- residual information bits in B-frames do not contribute to other frames
- temporal distance between anchor P-frames leads to less correlation and higher bitrate in P-frames (successive B-frame limit)
- additional frame delay in encoder and decoder
  - (bad for streaming, conversational applications)
- increased encoder and decoder computational complexity and memory requirements
video coding standards overview

TV, HDTV Production
HDTV
DVB, DVD
CD-ROM
Streaming, Video Conf.
Video Phone


JPEG
MPEG-1
MPEG-2
DV
MPEG-4
ITU-MPEG (JVT)
H.264
JPEG2000

100 Mbit/s
20 Mbit/s
1 Mbit/s
64 kbit/s

ITU-T
H.261
H.263 ITU/MPEG (JVT)
H.264

HDTV, TV, HDTV
Production

ITU-T
H.261
H.263
H.264

MPEG-1
MPEG-2
MPEG-4

JPEG
H.261
H.263
H.264

MPEG-1
MPEG-2
MPEG-4

H.261
H.263
H.264

JVT

[HHI]
video coding standards

- **scope**
  - data stream representation (systems, audio, video)
  - mandatory decoder operations, **NOT** encoder

- **constraints**
  - video streams
    - image resolution and representation
    - frame rate
  - audio streams
    - sampling rates
    - bits per sample and sampling format
    - number of channels
  - general
    - data rates
    - allowed tools
Motion Picture Experts Group (http://www.mpeg.org)

- MPEG-1 (1993)
  - audio/video coding, bitrate ≤ 1.5 MBit/s
- MPEG-2 (1996)
  - audio/video coding, initial bitrate ≤ 15 MBit/s, various profiles
  - digital TV (DVB-S,-C,-T), DVD
- MPEG-3 (skipped)
  - planned as HDTV standard, features incorporated into MPEG-2
- MPEG-4 (2000, version 2)
  - audio/video coding for internet applications, mobile devices
  - object based coding, low bitrate coding
  - not limited to PCM data (structured audio, 3D objects)
- MPEG-7
  - media descriptions, meta information, search/retrieval
- MPEG-21
  - planned universal media framework
Hybrid Video Coding

Combination of motion compensated prediction and transform coding of the residual

- Lossless
  - Prediction step where the prediction error is retained
  - Transformation step
  - Coding of remaining data

- Lossy
  - Quantization

Encoder

Original image → prediction → transformation → quantization → pre-coding / entropy coding → storage / transmission

Decoder

Reconstructed image ← prediction ← inverse transformation ← inverse quantization ← decoding
MPEG-1

- Finalized 1993 (ISO/IEC 11172)
- first MPEG-Standard for digital video
  - primary target Video-CD, CD-I
    - ~1.5 MBit/s overall data rate, 64...192 kBit/s Audio
    - image format 352x288 pixel @25 FPS, 352x240 pixel @ 30FPS
    - up to 74 min video on single-speed CD

- parts
  - Part1: Systems (packet stream, a/v multiplex)
  - Part2: Visual (MPEG-1 video coding)
  - Part3: Audio (MPEG-1 Audio Layers I,II,III)
  - Part4: Compliance testing
  - Part5: Reference Software Model
“closed loop coding“:
- decoder already part of encoder
- reference frames in encoder buffer match the frames available to the decoder
MPEG-1 quantization of DCT coefficients

- **Intra DC coefficient**
  - fixed (de-)quantizer $F'(0,0) = F_Q(0,0) \cdot 8$
  - reason: block artifacts

- **AC coefficients**
  - intra quantizer matrix $W(u,v)$ perceptually modeled
  - inter quantizer matrix flat
  - linear quantizer scale $q_{\text{scale}} = 1,2,...,31$
  - quantization formula
    $$F_Q(u,v) = \frac{(16 F(u,v) + l)}{(2 W(u,v) q_{\text{scale}})}$$
    - intra: $l=\text{sgn}(F(u,v)) \cdot W(u,v) \cdot q_{\text{scale}}$, inter: $l=0$
  - reconstruction formula
    $$F'(u,v) = \frac{(2 F_Q(u,v) + k)}{16} \cdot W(u,v) \cdot q_{\text{scale}}$$
    - intra: $k=0$, inter: $k=\text{sgn}(F_Q(u,v))$
MPEG-1 quantization (2)

- intra: rounding
- inter: truncation (dead zone)

$q_{\text{scale}}$ is the main parameter for rate control

$q_{\text{scale}}$ can be changed between macroblocks (diff. transmission)

intra quantization is tuned for accuracy

inter quantization is tuned for suppression of small coefficients
MPEG-1 block scanning

- frequency progressive zigzag scanning
  - DC coefficient coded separately
    - difference to previously coded DC
  - key concept: diagonal frequency progression
  - low frequency coefficients prioritized
  - compare to intra quantization matrix
MPEG-1 VLC

- **group symbols into „run“ and „level“**
  - „run“ is the number of zeroes before each significant coefficient in block-scan order

- **jointly transmit run and level as complex code**
  - uniform distribution of sign assumed and appended to all codes
  - finish transmitted block with EOB (end of block) marker
  - unlikely symbols prefixed with ESCAPE code and sent as FLC

<table>
<thead>
<tr>
<th>code</th>
<th>run</th>
<th>level</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>EOB</td>
<td>-</td>
</tr>
<tr>
<td>1s*</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11s*</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>011s</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0100s</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0101s</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>00101s</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>00110s</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>00111s</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>000110s</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>000111s</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>ESCAPE</td>
<td>-</td>
</tr>
</tbody>
</table>

symbol-code-mapping (exemplary)

*different code of first symbol in inter blocks*
MPEG-1 Motion Compensation

- **16x16, half pel resolution**
  - *virtual* half pixel positions gained by interpolation
  - rules for virtual pixels b,h,j
    - $b = (A+B+1)/2$
    - $h = (A+C+1)/2$
    - $j = (A+B+C+D+2)/4$

- algorithm applied to luma and chroma, latter vectors scaled

- differential coding of motion vectors
  - predictor from last macroblock
  - VLC
MPEG-1 Mode Decision

- **I-frames**
  - Intra Macroblocks only

- **P-frames**
  - Forward (last P/I-frame) 16x16 + MV
  - Intra

- **B-frames**
  - Forward (last P/I-frame) 16x16 + MV
  - Backward (next P/I-frame) 16x16 + MV
  - Bi-directional (last and next frame) 16x16 + 2 MV
  - Intra

**optimal mode not trivial: rate-distortion problem**
Finalized 1996 (ISO/IEC 13818), joint recommendation ITU-T H.262

**primary application digital TV**
- from MPEG-1 up to HDTV (~1 MBit/s ... 80 MBit/s)
- initial target: CCIR 601 interlaced video at 3-9 MBit/s (DVB, DVD)

**parts**
- Part1: Systems (program stream, transport stream), a.k.a. ITU-T H.222
- Part2: Visual (MPEG-2 video coding)
- Part3: Audio (MPEG Audio Layers I, II, III, multi-channel extensions)
- Part4: Compliance testing
- Part5: Reference Software Model
- Part6: Digital Storage Media Command and Control (DSM-CC)
- Part7: Advanced Audio Coding (AAC), formally added in 1997
- Part9: Extension for real time interfaces
- Part10: Compliance extensions for DSM-CC
MPEG-2 specifics

- **decoders required to decode legacy MPEG-1 streams**
  - extended syntax compared to MPEG-1, mostly unchanged semantics

- **main changes in video coding** [ISO00b] [MPFL97]
  - range extension for width, height, bitrate, etc.
  - interlace coding
    - picture structure (frame/field, adaptive frame/field)
    - motion compensation
    - DCT interlacing, coefficient scanning
  - scalable extension
  - error concealment
  - profiles and levels
  - extended aspect ratio information for non square pixels, „anamorphic“
  - extended IDCT mismatch control
  - stuffing macroblocks
  - AC run/level syntax
MPEG-2 specifics (2)

main changes (cont)

- intra DC quantization: dividers $8,4,2,1$ for precision $8,9,10,11$ Bit
- AC quantization with doubled accuracy, linear or nonlinear QP scale
- separate quantization tables for luma/chroma, custom matrices
## MPEG-2 Profiles / Levels

<table>
<thead>
<tr>
<th>profile</th>
<th>simple</th>
<th>main</th>
<th>4:2:2</th>
<th>high</th>
<th>spatial scalable</th>
<th>SNR scalable</th>
</tr>
</thead>
<tbody>
<tr>
<td>short desc.</td>
<td>SP</td>
<td>MP</td>
<td>422P</td>
<td>HP</td>
<td>SCP</td>
<td>SNR</td>
</tr>
<tr>
<td>B-frames</td>
<td>no</td>
<td></td>
<td></td>
<td>yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scalable</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>SNR or spatial</td>
<td>SNR</td>
<td></td>
</tr>
<tr>
<td>subsampling</td>
<td>4:2:0</td>
<td>4:2:0</td>
<td>4:2:2</td>
<td>4:2:0</td>
<td>4:2:0</td>
<td>4:2:0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>level</th>
<th>Low Level</th>
<th>Main Level</th>
<th>High 1440</th>
<th>High Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>short desc.</td>
<td>LL</td>
<td>ML</td>
<td>H-14</td>
<td>HL</td>
</tr>
<tr>
<td>width (pixels)</td>
<td>352</td>
<td>720</td>
<td>1440</td>
<td>1920</td>
</tr>
<tr>
<td>height (pixels)</td>
<td>288</td>
<td>576</td>
<td>1152</td>
<td>1152</td>
</tr>
<tr>
<td>framerate (Hz)</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>bitrate (MBit/s)</td>
<td>4</td>
<td>15</td>
<td>60</td>
<td>80</td>
</tr>
</tbody>
</table>
## MPEG-2 Profile @ Level combinations

<table>
<thead>
<tr>
<th>profile@level</th>
<th>resolution (pixels)</th>
<th>max. framerate</th>
<th>subsampling</th>
<th>bitrate (MBit/s)</th>
<th>application</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP@LL</td>
<td>176x144</td>
<td>15</td>
<td>4:2:0</td>
<td>0.096</td>
<td>mobile phones</td>
</tr>
<tr>
<td>SP@ML</td>
<td>352x240</td>
<td>24</td>
<td>4:2:0</td>
<td>0.384</td>
<td>PDA, streaming</td>
</tr>
<tr>
<td></td>
<td>352x288</td>
<td>15</td>
<td>4:2:0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MP@LL</td>
<td>352x288</td>
<td>30</td>
<td>4:2:0</td>
<td>4</td>
<td>(S)VCD</td>
</tr>
<tr>
<td>MP@ML*</td>
<td>720x576</td>
<td>25</td>
<td>4:2:0</td>
<td>15</td>
<td>DVD, DVB</td>
</tr>
<tr>
<td></td>
<td>720x480</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MP@H-14</td>
<td>1440x1080</td>
<td>30</td>
<td>4:2:0</td>
<td>60</td>
<td>HDV (25 MBit/s)</td>
</tr>
<tr>
<td></td>
<td>1280x720</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MP@HL</td>
<td>1920x1080</td>
<td>30</td>
<td>4:2:0</td>
<td>80</td>
<td>ATSC, HD-DVB</td>
</tr>
<tr>
<td></td>
<td>1280x720</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>422P@ML</td>
<td>720x576</td>
<td>25</td>
<td>4:2:2</td>
<td>50</td>
<td>Sony IMX</td>
</tr>
<tr>
<td></td>
<td>720x480</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Main Level note: for PAL/NTSC compatibility: either 720x576@25 Hz or 720x480@29,97 Hz
analog TV is based on half frames
- one line offset between "top" field and "bottom" field
- 50/60 fields per second, temporal offset between fields is 1/50, 1/60 s
- trade-off between smooth motion and vertical flickering, constrained by bandwidth

compatibility of MPEG-2 with interlaced video is key feature for digital TV with existing TV-sets
MPEG-2 3:2 pulldown

- TV cameras provide interlaced material, cinema cameras do not
- 24 Hz progressive cinema film conversion
  - 29.97 Hz NTSC DVD by 3:2 pulldown after 0.1% deceleration
  - 25 Hz PAL DVD by 4% acceleration, two fields from one frame
MPEG-2 frame pictures / field pictures

- Two picture structures allowed in MPEG-2
  - Field picture
    - Fields coded as frames of half overall height
  - Frame picture
    - Support for progressive frames (like MPEG-1)
    - Interleaving of fields
MPEG-2 interlaced DCT support

- alternate scan order on frame level, interlace DCT on macroblock level
MPEG-2 Scalability

- Graceful degradation of image quality in error-prone environments
- Singular stream support for different transmission bandwidths
- Singular support for devices of different capabilities

- SNR scalability (related to data partitioning)
  - Base layer with coarse quantization
  - Enhancement layer(s) with finer quantization

- Spatial scalability
  - Base layer at low spatial resolution
  - Enhancement layer with higher spatial resolution

- Temporal scalability
  - Lower frame rate in base layer

- Hybrid scalability
ITU-T Rec. H.263

- "Video coding for low bit rate communication"
- H.263 (version 1) finalized in 1996 [H263]
- H.263+ (version 2) finalized in 1998
- H.263++ (version 3) finalized in 2000

Key design choices/requirements
- use of available technology
- low complexity (for low cost devices)
- interoperability and/or coexistence with other standards (H.320,H.261)
- error robustness
- hardware or software implementation
- QOS parameters: resolution, delay, frame rate, color performance
- application: video phone or multimedia terminal
H.263 version 1 provides the base syntax for MPEG-4 Part 2
- H.263 baseline profile must be supported by all compliant MPEG-4 video decoders
- after adoption of H.263v1, divergent development of H.263 and MPEG-4

Popular applications (today)
- video functionality of mobile phones
  - 3GPP standard: MP4 container, H.263 content
- video streaming
  - initial video codec for Adobe® Flash Video

H.263+, H.263++ added features like quarter pel motion estimation, long term prediction, scalability
- effectively testbed for new technologies ⇒ H.26L project (JVT/H.264)
## H.263 Profiles

<table>
<thead>
<tr>
<th>profile</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>unrestricted motion compensation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>extended motion vector range</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>arithmetic coding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>8x8 motion compensation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>overlapped block motion compensation (OBMC)</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>PB-frames</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>advanced intra coding</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>deblocking filter</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>slice structured coding (without submodes)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>slice structured coding with arbitrary slice order</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Temporal (B-frames), Temporal, SNR, Spatial Scalability, B-frames for temporal scalability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>data partitioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>reference picture resampling (warping)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Enhanced Reference Picture Selection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**H.263** | **H.263+ (version 2)** | **H.263++ (version 3)**
## H.263 Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>45</th>
<th>50</th>
<th>60</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max picture format</td>
<td>QCIF 176x144</td>
<td>CIF 352x288</td>
<td>CIF 352x288</td>
<td>CIF 352x288</td>
<td>QCIF 176x144 CPFMT for H.263+</td>
<td>CIF, CPFMT</td>
<td>720x288 CPFMT</td>
<td>720x576 CPFMT</td>
</tr>
<tr>
<td>Min. picture interval</td>
<td>2002/(30000)s for CIF 1001/(30000)s for QCIF, sub-QCIF</td>
<td>1001/(30000)s</td>
<td>1001/(30000)s</td>
<td>2002/(30000)s CPCFC for H.263+</td>
<td>1/50s for CIF or lower 1001/(60000)s at 352x240 or smaller CPCFC</td>
<td>1/50s at 720x288 or lower 1001/(60000)s at 720x240 or smaller CPCFC</td>
<td>1/50s at 720x576 or lower 1001/(60000)s at 720x480 or smaller CPCFC</td>
<td></td>
</tr>
<tr>
<td>Max bitrate in kBit/s</td>
<td>64</td>
<td>128</td>
<td>384</td>
<td>2048</td>
<td>128</td>
<td>4096</td>
<td>8192</td>
<td>16384</td>
</tr>
</tbody>
</table>

H.263+, H.263++

- **predefined set of spatial resolutions**
  - sub-QCIF, QCIF, CIF, 4CIF, 16CIF
  - custom sizes only in H.263+, H.263++

CPCFC = Custom Picture Clock Frequency
CPFMT = Custom Picture Format

---

MP4 42
Henryk Richter
H.263 Motion Compensation (1)

- 16x16, half pel resolution
  - one vector per macroblock

- 8x8, half pel resolution (APM)
  - four vectors per macroblock

- Interpolation rounding control
  - traditional MPEG-1/-2 method biased towards "+1"
  - rules for virtual pixels b, h, j
    - \( b = (A+B+1-RC)/2 \)
    - \( h = (A+C+1-RC)/2 \)
    - \( j = (A+B+C+D+2-RC)/4 \)
  - RC recommended to alternate values 0 and 1 between consecutive P-frames
H.263 OBMC

- **overlapped block motion compensation**
  - motion vectors of surrounding blocks applied in addition to current block
  - each pixel predicted from 3 sources
  - luma only

- **prediction weighting**
  - prediction weights depend on distance to block center
  - overlapping 4 pixels deep

- **goals**
  - reduced motion uncertainty, quantization noise
  - reduced block artifacts
H.263 unrestricted motion compensation

- In MPEG-1/-2 and H.263 baseline, all motion compensated blocks must point inside image area.
- Unrestricted MC allows motion vectors point partially or completely beyond image borders.

Basic rule for any referenced pixel outside the image area:
Replicate nearest border pixel of image.

⇒ Main use: camera motion.
H.263 motion vector prediction

- Motion vectors predicted from surrounding blocks
- Component-wise \((x,y)\) median of three predictors
- Difference to median prediction is MVD, coded in bitstream
- Special rules when macroblock is on image or GOB (group of blocks) border
- Special rules for INTER4V (8x8) mode

\[
\text{MEDIAN}(a, b, c) = a + b + c - \max(a, b, c) - \min(a, b, c)
\]
H.263 quantization

- **Intra DC coefficient**
  - fixed (de-)quantizer $F'(0,0) = F_Q(0,0) \cdot 8$

- **AC coefficients**
  - uniform quantization matrix
  - quantizer indices $QP = 1, 2, 3, \ldots, 31$
  - even quantizer scale $q_{scale} = 2, 4, 6, \ldots, 62$
  - quantization formula
    \[ F_Q(u, v) = F(u, v) / q_{scale} \quad 0 \leq u, v \leq 7 \]
  - reconstruction formula
    \[ F'(u, v) = \text{sgn}(F_Q(u, v)) \cdot \frac{(2|F_q(u, v)| + 1) \cdot q_{scale}}{2} \quad 0 \leq u, v \leq 7 \]
  - low complexity quantization with dead-zone around zero
H.263 coefficient coding

- significant coefficients (level) associated with number of zeroes (run) between them
- same 8x8 zigzag scan order as in JPEG, MPEG-1
- symbols grouped into complex codes of run, absolute level, sign and indicator if last significant coefficient in block
- entropy coding by VLC tables (distinct Intra/Inter tables, ESCAPE symbol for improbable combinations via FLC)

(exemplary) 4x4-block with coefficients
H.263 macroblock flow

- improved coding efficiency through skip flag and conditional elements
- complex codes instead of flag bits allow optimized VLC tables

**Macroblock Layer**
- COD
- MCBPC
- CBPY
- DQUANT
- MVD
- MVD 2-4

**Block Layer**
- CBP
- Intra DC
- TCOEFF

- CBPY, CBPC, CBPB
- INTRA only
- Run/Level/Sign/Last

Inter-Frames only
PB-Frames only
INTER4V only
INTRA/INTER+Q
### H.263 MB types

<table>
<thead>
<tr>
<th>picture type</th>
<th>MB type</th>
<th>Name</th>
<th>COD</th>
<th>MCBPC</th>
<th>CBPY</th>
<th>DQUANT</th>
<th>MVD</th>
<th>MVD_2-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRA</td>
<td>3</td>
<td>INTRA</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRA</td>
<td>4</td>
<td>INTRA+Q</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTRA</td>
<td>Stuffing</td>
<td>-</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTER</td>
<td>Not coded</td>
<td>SKIP</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTER</td>
<td>0</td>
<td>INTER</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>INTER</td>
<td>1</td>
<td>INTER+Q</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>INTER</td>
<td>2</td>
<td>INTER4V</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>INTER</td>
<td>3</td>
<td>INTRA</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTER</td>
<td>4</td>
<td>INTRA+Q</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTER</td>
<td>5</td>
<td>INTER4V+Q</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>INTER</td>
<td>Stuffing</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **mode decision optimization** essential for good coding efficiency
- **in low complexity image areas, COD flag reduces overall data amount for the respective macroblock to a single bit**
Title
- Information technology — Coding of audio-visual objects

Application areas
- Internet Multimedia
- Interactive Video Games
- Interactive Storage Media
- Multimedia Mailing
- Networked Database Services
- Remote Emergency Systems
- Remote Video Surveillance
- Wireless Multimedia
- IPC Interpersonal Communications (videoconferencing, videophone)
MPEG-4 Parts

MPEG4 (ISO/IEC 14496)

- Part1 Systems (1999) (includes extensions to MPEG transport streams)
- Part2 Visual (2001) (incorporates large parts from H.263)
- Part3 Audio (2001) (includes new modes to AAC, CELP, HVXC speech codecs)
- Part4 Conformance (2001)
- Part5 Reference Software (2001)
- Part6 Delivery Multimedia Integration Framework (2000)
- Part7 Optimized Visual Reference Software
- Part8 Carriage of MPEG-4 Contents over IP (2002) (esp. RTP)
- Part9 Reference Hardware Description (2003)
- Part11 Scene description and Application engine, (BIFS)
- Part12 ISO Base Media File Format: A file format for storing media content.
- Part13 Intellectual Property Management and Protection (IPMP)
- Part14 MPEG-4 File Format (based on Part12, derived from Quicktime file format)
- Part15-23: Text representation/compression, texture compression etc.
MPEG-4 concepts

intended as generic audiovisual framework including
- audio
- text
- still images
- video
- 2D/3D mesh objects

visual representation
- scene graph, based on elementary objects
- client-side compositing
- arbitrary shaped objects
  - non-rectangular natural images and video, optionally with alpha-channel
  - synthesized objects (VRML)
- interactivity, back channel information
MPEG-4 terminal, receiver side

Network Layer

- Transmission on variable number of channels

Demultiplex

- Elementary streams

Decompression

- A/V objects

Scene graph

- Scene graph

Upstream channel for user interaction, quality reports, data requests

Compositing

- Final rendered scene, interactivity options

Die Schlagzeilen
## MPEG-4 Visual Profiles

<table>
<thead>
<tr>
<th>profile / tools</th>
<th>Simple</th>
<th>Core</th>
<th>Main</th>
<th>Advanced Real Time Simple</th>
<th>Advanced Coding Efficiency</th>
<th>Advanced Simple</th>
<th>Fine Granularity Scalable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic (I-VOP, P-VOP, AC/DC Prediction, 4MV, unrestricted MV)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Error resilience (Slice Resync, Data Partitioning, RVLC)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Short Header (H.263)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B-VOP</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>H.263 + MPEG quantization</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>P-VOP based temporal scalability</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Binary Shape</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Grey Shape (Alpha channel)</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Interlace</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sprite</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Dynamic Resolution Conversion</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NewPred</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Global Motion Compensation</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>1/4 Pel Motion Compensation</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SA-DCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fine Granularity Scalability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FGS Temporal Scalability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

- **Visual object types for raster image sources (i.e. natural images)**
- **additional object types for synthetic images**
  - (2D Mesh, Texture, Face Coding)
MPEG-4 popular implementations / derivates

**DivX :-) 3.11**
- Microsoft proposal MPEG-4 V3, rejected by ISO due to late entry
- Microsoft released that codec as „MP43“ within their own media framework, artificially limited to ASF streams as target systems layer
- codec is **not** MPEG-4 compliant, though lossless conversion feasible
- feature set comparable to H.263, stunning image quality at that time
- codec was hacked to allow inclusion into AVI files
- DivX name originated from „Digital Video Express“, a rental service with limited disc lifetime or limited play counts, respectively

**DIVX 4.0+, XVID, 3IVX**
- MPEG-4 V.2 ASP codec framework, initially based on reference software „MoMuSys“, open-source release by DIVX forked into XVID
- bitstreams are MPEG-4 Part 2 compliant, although proper interoperability is version-dependent with regressions in early versions

**RealVideo (certain versions), Quicktime 6**
MPEG-4 AC/DC prediction

- improved spatial de-correlation by prediction in frequency domain
- adaptive decision whether vertical or horizontal DC prediction
- flag for optional AC prediction
  - direction derived from DC prediction decision
  - either first row or column of AC coefficients from upper or left block
  - predictor coefficients scaled to match QP or current MB

**effect in spatial domain:**
- horizontal/vertical block average copied to current block

\[
if(|F_B(0, 0) - F_A(0, 0)| < |F_B(0, 0) - F_C(0, 0)|) \\
PRED = V \\
else \\
PRED = H
\]
MPEG-4 B-Frames

**DIRECT Mode**
- in addition to forward, backward and bi-directional prediction
- predict bi-directional motion vector pairs \((MVB,MVP)\) from
  - motion between backward reference I/P-frame and future reference P-frame
  - co-located macroblocks in the reference frames
  - temporal distance between the reference frames
  - implicitly assume 8x8 blocks using the default accuracy of 1/4 or 1/4 pel
- optionally MV prediction error as difference motion vector \((MVD)\) transmitted

![Diagram showing B-frames and reference frames](image-url)
new feature to MPEG-4 is 1/4 pixel motion compensation accuracy
8-tap filter for approximating half pel positions
high accuracy operations limited to 8-tap filter stage only, downscaling/clipping directly after half pel approximation
quarter pel positions calculated by averaging half pel positions and nearest raster pixel
mirrored block extension limits picture data bandwidth to the same range as in conventional interpolation methods
Global Motion Compensation

- allowed in S-frames (sprite frames) only, no bidirectional prediction
- special case of sprites (static, one frame life time)
- 1 extra bit per macroblock to signal global/local motion model
- global model
  - either 2, 4 or 6 parameters (translation, scaling, affine)
  - mapping on 16x16 macroblocks, per-macroblock global/local decision
  - warping parameters transmitted with points of pre-defined reference location
  - bilinear interpolation, accuracy up to 1/64
- residual coding identical to other motion compensated macroblocks

1 warp point

2 warp points

3 warp points

4 warp points

(regular sprites only, not GMC)

average MV calculated for local MC
Sprites
- non-rectangular objects
- intended for panoramic background images, where foreground motion is added as individual video objects
- for displayed scenes, a parameter set is transmitted for the visible part
- incremental sprite update possible
- up to 8 warping parameters (perspective model) applied

[Smolic 2001]
motion based video segmentation

- affine parametric motion fields for object detection and segmentation
- intended for 3D wavelet coding [Schwarz 2000]
- useful for MPEG-4 shape coding as well

original image

motion field
References

References
