Overview: Video Coding Standards

- Video coding standards: applications and common structure
- ITU-T Rec. H.261
- ISO/IEC MPEG-1
- ISO/IEC MPEG-2
- State-of-the-art: H.264/AVC
Applications of Video Compression

Efficient and flexible video compression standard needed

Adapted from [Srinivasan et al., 2004]
# Applications of Video Compression

<table>
<thead>
<tr>
<th>Application</th>
<th>Bit Rate</th>
<th>Encoding Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital television broadcasting</td>
<td>2 . . . 6 Mbps</td>
<td>MPEG-2 (H.264/AVC)</td>
</tr>
<tr>
<td></td>
<td>(10…20 Mbps for HD)</td>
<td></td>
</tr>
<tr>
<td>DVD video</td>
<td>5 . . . 8 Mbps</td>
<td>MPEG-2</td>
</tr>
<tr>
<td>Blu-ray Disk</td>
<td>up to 40 Mbps</td>
<td>MPEG-2, H.264/AVC, VC-1 (up to 1080p)</td>
</tr>
<tr>
<td>Internet video streaming</td>
<td>100 . . . 2000 kbps</td>
<td>MPEG-1, H.264/AVC, VC-1, or similar proprietary</td>
</tr>
<tr>
<td>Video over 3G wireless</td>
<td>100 . . . 500 kbps</td>
<td>H.263, MPEG-4, H.264/AVC</td>
</tr>
</tbody>
</table>
Motion-compensated Hybrid Coding
H.261, MPEG-1, MPEG-2, H.263, MPEG-4, H.264/AVC

Diagram:
- Coder Control
- Transform/Quantizer
- Decoder
- Intra/Inter
- Motion-Compensated Predictor
- Motion Estimator
- Deq./Inv. Transform
- Entropy Coding
- Control Data
- Quant. Transf. coeffs
- Motion Data

Flow:
- Control
- Data
- Quant.
- Transf. coeffs
- Motion
- Data
Video Compression Standards: Hierarchical Syntax
ITU-T Rec. H.261

- International standard for ISDN picture phones and for video conferencing systems (1990)
- Image format: CIF (352 x 288 Y samples) or QCIF (176 x 144 Y samples), frame rate 7.5 ... 30 fps
- Bit-rate: multiple of 64 kbps (= ISDN-channel), typically 128 kbps including audio
- Picture quality: for 128 kbps acceptable with limited motion in the scene
- Stand-alone videoconferencing system or desk-top videoconferencing system, integrated with PC
Macroblocks

- Macroblock (MB) of 16x16 pixels
- Sampling format: 4:2:0
- MB consists of 4 luminance and 2 chrominance blocks

![Macroblock Diagram]

- 16x16 luminance samples
- 8x8 Cb-samples
- 8x8 Cr-samples
H.261 Motion-Compensated Prediction

- Integer-pel accuracy
- One displacement vector per macroblock
- Maximum displacement vector range +/-16 horizontally and vertically
- Adaptive loop filter, separable in 1-D horizontal and vertical impulse response: \([\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]\)
- Differential encoding of motion vectors
H.261 Residual Coding

- 8x8 DCT
- Quantization
  - Uniform quantizer ($\Delta=8$) for intra-mode DC coefficients
  - Uniform threshold quantizer ($\Delta=2,4,\ldots,62$) for AC coefficients in intra-mode and all coefficients in inter-mode
- Zig-zag scan
- Run-level coding for entropy coding
  - (zero-run, value) symbols
  - zero-run: the number of coefficients quantized to zero since the last nonzero coefficient
  - value: the amplitude of the current nonzero coefficient
# H.261 Macroblock Types (VLC Table)

<table>
<thead>
<tr>
<th>Prediction</th>
<th>MQUANT</th>
<th>MVD</th>
<th>CBP</th>
<th>TCOEFF</th>
<th>VLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>0001</td>
</tr>
<tr>
<td>Intra</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>0000 001</td>
</tr>
<tr>
<td>Inter</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Inter</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>0000 1</td>
</tr>
<tr>
<td>Inter+MC</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>0000 0000 1</td>
</tr>
<tr>
<td>Inter+MC</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>0000 0001</td>
</tr>
<tr>
<td>Inter+MC</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>0000 0000 01</td>
</tr>
<tr>
<td>Inter+MC+FIL</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>Inter+MC+FIL</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>Inter+MC+FIL</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>0000 01</td>
</tr>
</tbody>
</table>
MPEG-1/2: GOP Structure

Group of Pictures = GOP

I-Picture

P-Picture

P-Picture

B-Pictures

time

1 3 4 2 6 7 8 5
MPEG-1/2 Encoder

Pre-processing → Picture reordering → 8x8 DCT → Weighting → Quantization → VLC → Video multiplex → Buffer

Inverse quantization → Inverse weighting → Inverse 8x8 DCT

Motion vectors, macroblock info, start codes

Video in → Bitstream

Picture store 1 → Motion compensation → Picture store 2

Motion compensation → 1/2 → Zero
MPEG-1: coding of I-pictures

- I-pictures: intraframe coded
- 8x8 DCT
- Arbitrary weighting matrix for coefficients
- Differential coding of DC-coefficients
- Uniform quantization
- Zig-zag-scan, run-level-coding
- Entropy coding
- Unfortunately, not quite JPEG
MPEG-1: coding of P-pictures

- Motion-compensated prediction from an encoded I-picture or P-picture (DPCM)
- Half-pel accuracy of motion compensation, bilinear interpolation
- One displacement vector per macroblock
- Differential coding of displacement vectors
- Coding of prediction error with 8x8-DCT, uniform threshold quantization, zig-zag-scan as in I-pictures
MPEG-1: coding of B-pictures

- Motion-compensated prediction from two consecutive P- or I-pictures
  - either
    - only forward prediction (1 vector/macroblock)
  - or
    - only backward prediction (1 vector/macroblock)
  - or
    - Average of forward and backward prediction = interpolation (2 vectors/macroblock)

- Half-pel accuracy of motion compensation, bilinear interpolation

- Coding of prediction error with 8x8-DCT, uniform quantization, zig-zag-scan as in I-pictures
MPEG-2 vs. MPEG-1

- Efficiently compress interlaced digital video at broadcast quality
  - Frame pictures or field pictures
  - Adaptive frame/field prediction
  - Adaptive frame/field DCT
- Improved coding efficiency by different quantization, VLC tables, and additional coefficient scan patterns
- Spatial, temporal and SNR scalability profiles (rarely used)
Frame = Both Fields Combined
Adaptive Frame/Field DCT
Adaptive Frame/Field Motion Compensation

**Frame Prediction**

**Field Prediction**
JVT Project

- **August 1999**: 1st test model (TML-1) of H.26L
- **December 2001**: Formation of the Joint Video Team (JVT) between VCEG and ISO/IEC JTC 1/SC 29/WG 11 (MPEG) to establish a joint standard project - **H.264 / MPEG4-AVC**
- **ITU-T Approval**: May 2003
- **ISO/IEC Approval**: October 2003
H.264/AVC Coder

Input Video Signal

Split into Macroblocks 16x16 pixels

Coder Control

Transform/Scal./Quant.

Decoding

Scaling & Inv. Transform

Deblocking Filter

Intra-frame Prediction

Motion Compensation

Motion Estimation

Intra/Inter

Output Video Signal

Entrophy Coding

Control Data

Quant. Transf. coeffs

Output Data

[source: G. Sullivan, VCEG]
Common Elements with other Standards

- Macroblocks: 16x16 luma + 2 x 8x8 chroma samples
- Block-wise motion compensation
- Variable block-size motion compensation
- Block transform of prediction error
- Scalar quantization
- I, P, and B coding types

[source: G. Sullivan, VCEG]
H.264 Motion Compensation Accuracy

Input Video Signal

Split into Macroblocks 16x16 pixels

Decoder

Coder Control

Transform/Scal./Quant.

Scaling & Inv. Transform

Intra/Inter Coder Control

Motion Estimation

Intra-frame Prediction

Motion Compensation

16x16
16x8
8x16
8x8

Motion vector accuracy 1/4 (6-tap filter)

[source: G. Sullivan, VCEG]
H.264 Multiple Reference Frames

- Multiple Reference Frames
- Generalized B Frames
- Weighted Prediction

[source: G. Sullivan, VCEG]
H.264 Intra Prediction

- Directional spatial prediction (9 types for luma, 1 chroma)

\[
\begin{array}{cccccccc}
Q & A & B & C & D & E & F & G \\
I & a & b & c & d & & & \\
J & e & f & g & h & & & \\
K & i & j & k & l & & & \\
L & m & n & o & p & & & \\
\end{array}
\]

- e.g., Mode 3: diagonal down/right prediction
  \[a, f, k, p \text{ are predicted by } (A + 2Q + I + 2) >> 2\]

[source: G. Sullivan, VCEG]
H.264 4x4 Transform

- 4x4 Block Integer Transform
  \[
  A = \begin{bmatrix}
  1 & 1 & 1 & 1 \\
  2 & 1 & -1 & -2 \\
  1 & -1 & -1 & 1 \\
  1 & -2 & 2 & -1
  \end{bmatrix}
  \]

- Repeated transform of DC coeffs for 8x8 chroma and some 16x16 Intra luma blocks

[source: G. Sullivan, VCEG]
Deblocking Filter

One dimensional visualization of an edge position

Filtering of \( p_0 \) and \( q_0 \) only takes place if:

1. \(|p_0 - q_0| < \alpha(QP)\)
2. \(|p_1 - p_0| < \beta(QP)\)
3. \(|q_1 - q_0| < \beta(QP)\)

Where \( \beta(QP) \) is considerably smaller than \( \alpha(QP) \)

Filtering of \( p_1 \) or \( q_1 \) takes place if additionally:

1. \(|p_2 - p_0| < \beta(QP) \) or \(|q_2 - q_0| < \beta(QP)\)

\( (QP = \text{quantization parameter}) \)

[source: G. Sullivan, VCEG]
Deblocking: Subjective Result for Intra

Highly compressed first decoded intra picture at 0.28 bit/sample

[source: G. Sullivan, VCEG]

Without Filter

With H264/AVC Deblocking
Deblocking: Subjective Result for Inter

Highly compressed decoded inter picture

Without Filter  With H264/AVC Deblocking

[source: G. Sullivan, VCEG]
Example Streaming Test Result

[Wiegand, et al. 2003]
Example Streaming Test Result

Tempete CIF 15Hz

Rate saving relative to MPEG-2

Y-PSNR [dB]

[H.264/AVC MP]

[MPEG-4 ASP]

[H.263 HLP]

[Wiegand, et al. 2003]
Example Entertainment-Quality Applications Result

![Graph showing Y-PSNR vs Bit-rate for Entertainment SD (720x576i) 25Hz with MPEG-2 and H.264/AVC MP]

[Wiegand, et al. 2003]

Bernd Girod: EE398A Image and Video Compression

Video Coding Standards no. 34
Example Entertainment-Quality Applications Result

![Graph showing rate saving relative to MPEG-2 vs. Y-PSNR (dB) for Entertainment SD (720x576i) 25Hz. The graph illustrates the performance of H.264/AVC MP.]
Further reading

