The H.264/AVC Video Coding Standard
(ITU-T Rec. H.264 | ISO/IEC 14496-10)
Gary J. Sullivan, Ph.D.
ITU-T VCEG Rapporteur / Chair
ISO/IEC MPEG Video Rapporteur / Co-Chair
Microsoft Corporation Video Architect
November 2007

Video Coding Standardization Organizations
Two organizations have historically dominated general-purpose video compression standardization:
• ITU-T Video Coding Experts Group (VCEG)
  International Telecommunications Union —
  Telecommunications Standardization Sector (ITU-T,
  a United Nations Organization, formerly CCITT),
  Study Group 16, Question 6
• ISO/IEC Moving Picture Experts Group (MPEG)
  International Standardization Organization and
  International Electrotechnical Commission, Joint
  Technical Committee Number 1, Subcommittee 29,
  Working Group 11
Recently, the Society for Motion Picture and Television Engineers
(SMPTE) has also entered with “VC-1”, based on Microsoft’s WMV 9 but
this talk covers only the ITU and ISO/IEC work.

The Scope of Picture and Video Coding Standardization
Only the Syntax and Decoder are standardized:
• Permits optimization beyond the obvious
• Permits complexity reduction for implementability
• Provides no guarantees of Quality

The Advanced Video Coding Project
AVC / ITU-T H.264 / MPEG-4 part 10
§ History: ITU-T Q.6/SG16 (VCEG - Video Coding Experts Group)
  “H.26L” standardization activity (where the “L” stood for “long-term”)
§ Aug 1999: 1st test model (TML-1)
§ July 2001: MPEG open call for technology: H.26L demo’ed
§ Dec 2001: Formation of the Joint Video Team (JVT) between
  VCEG and MPEG to finalize H.26L as a new joint project (similar to
  MPEG-2/H.262)
§ July 2002: Final Committee Draft status in MPEG
§ Dec ’02 Technical freeze, FCD ballot approved
§ May ’03 Completed in both orgs
§ July ’04 Fidelity Range Extensions (FRExt) completed
§ Jan ’07 Professional Profiles completed

H.264/AVC Objectives
§ Primary technical objectives:
  • Significant improvement in coding efficiency
  • High loss/error robustness
  • “Network Friendliness” (carry it well on MPEG-2 or RTP or
    H.32x or in MPEG-4 file format or MPEG-4 systems or …)
  • Low latency capability (better quality for higher latency)
  • Exact match decoding
§ Initial extension objectives (in FRExt and Prof Profiles):
  • Professional applications (more than 8 bits per sample,
    4:4:4 color sampling, etc.)
  • Higher-quality high-resolution video
  • Alpha plane support (a degree of “object” functionality)
  • Extended color gamut support
A Comparison of Performance

- Test of different standards (ICIP 2002 study)
- Using same rate-distortion optimization techniques for all codecs
- Streaming test: High-latency (included B frames)
  - Four QCIF sequences coded at 10 Hz and 15 Hz (Foreman, Container, News, and Calendar)
  - Four QCIF sequences coded at 15 Hz and 30 Hz (Bugs, Flower Garden, Mobile and Tempete)

- Real-time communication test: No B frames
  - Four QCIF sequences encoded at 10 Hz and 15 Hz (Akiyo, Foreman, Mother and Daughter, and Silent Voice)

- Compare four codecs using PSNR measure:
  - H.263 (+) (high-latency profile, conversational high-compression profile, baseline profile)
  - MPEG-4 Visual (simple and advanced simple profiles with & without B pictures)
  - H.264 AVC version 1 (with & without B pictures)

Note: These test results are from a private study and are not an endorsed report of the JVT, VCEG or MPEG organizations.

Comparison to MPEG-2, H.263, MPEG-4p2

- Foreman QCIF 10Hz
- Bit-rate [kbit/s]
- Quality Y-PSNR [dB]

Motion Compensation Accuracy

- Motion vector accuracy 1/4 sample for luma
- 4 types for 8x8 chroma pred
- 4 types for 16x16 luma pred
- 9 types for 4x4 luma pred

Multiple Reference Frames

- Multiple Reference Frames
- Generalized Referencing Relationships
- Generalized Bi-Prediction

Intra Prediction

- Directional spatial prediction
  9 types for 4x4 luma pred,
  4 types for 16x16 luma pred,
  4 types for 8x8 chroma pred

- e.g., Mode 4:
  diagonal down/right prediction
  $a, b, f, k, p$ are predicted by
  $(A + 2M + I + 2) >> 2$
DCT-Like Transform Coding

| 1 | 1 | 1 | 1 |
| 2 | 1 | -1 | -2 |
| 1 | -3 | -1 | 1 |
| 1 | -2 | 2 | -1 |

Hierarchical transform of DC coefficients for 8x8 chroma and 16x16 intra luma blocks.

Entropy Coding

H.264/AVC Version 1 Profiles

Three profiles in version 1: Baseline, Main, and Extended

- Main Profile (esp. Broadcast)
  - All Baseline features except enhanced error resilience features
  - Interlaced video handling
  - Generalized B pictures
  - Adaptive weighting for B and P picture prediction
  - CABAC (arithmetic entropy coding)

- Extended Profile (esp. Streaming)
  - All Baseline features
  - Interlaced video handling
  - Generalized B pictures
  - Adaptive weighting for B and P picture prediction
  - More error resilience: Data partitioning
  - SP/SP switching pictures

Non-Baseline H.264/AVC Version 1 Profiles

- AVC standard finished May 2003, published as "twin text"
  - ITU-T Recommendation H.264
  - ISO/IEC 14496-10 MPEG-4 AVC

- Fidelity-Range Extensions (FRExt)
  - Work item initiated in July 2003
  - More than 8 bits, color other than 4:2:0
  - Alpha coding
  - More coding efficiency capability
  - Also new supplemental information

- Professional Profiles
  - Work item initiated in October 2005
  - Focus initially on 4:4:4 (replacing prior FRExt 4:4:4 profile)
  - Later work on all-intra and new supplemental information
Technical Features – Part 1

- Larger transforms
  - 8x8 transform (as was in older standards)
  - Drop 4x8, 8x4, or larger, 16-point...
- Filtered intra prediction modes for 8x8 block size
- Quantization matrix
  - 4x4, 8x8, intra, inter trans. coefficients weighted differently
  - Old idea, dating to JPEG and before (circa 1986?)
  - Full capabilities not yet explored (visual weighting)
- Coding in various color spaces
  - 4:2:2, 4:2:0, Monochrome, with/without Alpha
  - New integer color transform (a VUI-message item)

Technical Features – Part 2

- Efficient lossless interframe coding
- Film grain characterization for analysis/synthesis representation
- Stereo-view video support
- Deblocking filter display preference

8x8 16-Bit (Bossen) Transform

```
[ 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 ]
[ 12 10 6 3 -3 -6 -10 -12 ]
[ 8 4 -4 -8 -8 -4 4 8 ]
[ 10 -3 -12 -6 6 12 3 -10 ]
[ 8 -8 -8 8 -8 -8 -8 8 ]
[ 6 -12 3 10 -10 -3 12 -6 ]
[ 4 -8 8 -4 -4 8 -8 4 ]
[ 3 -6 10 -12 12 -10 6 -3 ]
```

8x8 Transform Advantage

<table>
<thead>
<tr>
<th>Sequence</th>
<th>% BD bit-rate reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movie 1</td>
<td>11.59</td>
</tr>
<tr>
<td>Movie 2</td>
<td>12.71</td>
</tr>
<tr>
<td>Movie 3</td>
<td>12.97</td>
</tr>
<tr>
<td>Movie 4</td>
<td>11.06</td>
</tr>
<tr>
<td>Movie 5</td>
<td>13.46</td>
</tr>
<tr>
<td>Crawford</td>
<td>10.93</td>
</tr>
<tr>
<td>Riverbed</td>
<td>15.65</td>
</tr>
<tr>
<td>Average</td>
<td>12.48</td>
</tr>
</tbody>
</table>

Quantization Matrix

- Similar concept to MPEG-2 design
- Vary step size based on frequency
- Adapted to modified transform structure
- More efficient representation of weights
- Eight downloadable matrices (at least 4:2:0)
  - Intra 4x4 Y, Cb, Cr
  - Intra 8x8 Y
  - Inter 4x4 Y, Cb, Cr
  - Inter 8x8 Y

New Profiles Created by FRExt

- 4:2:0, 8-bit: “High” (HP)
- 4:2:0, 10-bit: “High 10” (H10)
- 4:2:2, 10-bit: “High 4:2:2” (H422)
- Effectively the same tools, but acting on different input data
- The High Profile has been a major force in recent industry developments (HD DVD, Blu-ray Disc, DBS, Terrestrial Broadcast, IPTV, etc.)
- The others are emerging in professional applications (e.g., content acquisition, editing, studios, recording)
A Performance Test for High Profile
(from JVT-L033 - Panasonic)

Subjective tests by Blu-Ray Disk Founders of FRExt HP
- 4:2:0/8 (HP) 1920x1080x24p (1080p), 3 clips.
- Nominal 3:1 advantage to MPEG-2
  - 8 Mbps HP scored better than 24 Mbps MPEG-2
- Apparent transparency at 16 Mbps

Some Notes on Quality Testing

- Use recent reference software (if using ref software)
- Use rate-distortion optimization in encoder
- Use large-range good-quality motion search
- Use appropriate “High” profile (incl. adaptive transform)
- If testing for PSNR, use “flat” quant matrices
- Otherwise, use “non-flat” quant matrices
- Use CABAC entropy coding
- Use more than 1 or 2 reference pictures
- Use hierarchical B reference frames coding structure
- Use bi-predictive search optimization (see JVT-N014)
- If testing high-quality PSNR, use adaptive thresholding

AVC Profile Overview

New Profiles for Professional Apps (2007)

New Scalable Video Coding Profiles

Work-in-progress: Multi-view Video Coding
For Further Information

JT, MPEG, and VCEG management team members:
- Gary J. Sullivan (garysull@microsoft.com)
- Jens-Rainer Ohm (ohm@ei.rwth-aachen.de)
- Ajay Luthra (aluthra@motorola.com)
- Thomas Wiegand (wiegand@hhi.de)

H.264-AVC literature references:
- IEEE Transactions on Circuits and Systems for Video Technology Special Issue on H.264-AVC (July 2003): Includes several highly-referenced papers
  (Luthra, Sullivan, Wiegand, Eds.)
- Paper in Proceedings of IEEE Jan 2005 (Sullivan & Wiegand)
- Overview incl. FRExt: SPIE Aug 2004 (Sullivan, Topiwala, & Luthra)
- Paper at SPIE VCIP 2005: Meta-overview and deployment (Sullivan)
- Paper in IEEE Communications Magazine, Aug 2006 (Marpe, Wiegand, Sullivan)
- Paper on Professional Extensions, IEEE ICIP, Sept 2007 (Sullivan et al.)
- Wikipedia H.264/MPEG-4 AVC page
- IEEE Transactions on Circuits and Systems for Video Technology Special Issue on Scalable Video Coding – Standardization and Beyond (Sept 2007)
  (Wiegand, Sullivan, Ohm, Luthra, Eds.)