The H.264 / MPEG-4 AVC Standard: Core Coding Technology and Recent Extensions

Detlev Marpe
Fraunhofer HHI, Berlin, Germany

H.264 / AVC: A Brief Historical Review

- August 1999: 1st Test model (TML-1) chosen among 4 technical proposals from Telenor, Nokia, Strathclyde University, and HHI
- December 2001: Formation of the Joint Video Team (JVT) between VCEG and MPEG (Moving Pictures Experts Group) to finalize H.26L as a joint standardization project – H.264 / MPEG-4 Part 10 (AVC: Advanced Video Coding)
- March 2003: Final Draft International Standard of Version 1
- May/July 2003: Approval by parent bodies
- Sept. 2004: Finalization of Fidelity Range Extensions ("FRExt") with a suite of 3 new "High" profiles
- January 2005: Scalable Video Coding (SVC) project launched

AVC: The Overall Conceptual Structure

Outline

0 History and Overall Conceptual Structure
1 Core Coding Technology
2 Fidelity Range Extensions
3 Scalable Video Coding Extensions
**Basic Macroblock Coding Structure**

- **Input Video Signal**
- **Split into Macroblocks 16x16 pixels**
- **Decoder**
- **Transform/Quant.**
- **Motion Estimation**
- **Deblocking Filter**
- **Entropy Coding**

**Main Innovative Features**

- Video coding layer of H.264/AVC is similar in spirit to other standards but with important differences.
- New key features are:
  - Enhanced motion compensation
  - Multiple reference pictures and generalized B pictures
  - Spatial intra prediction
  - Small blocks for transform coding
  - Integer block transform & adaptive transform block sizes
  - Improved deblocking filter
  - Enhanced entropy coding (CAVLC + CABAC)
  - Substantial bit rate savings (typically around 50%) relative to any other standard for the same perceptual quality
- Network abstraction layer takes lossy packet-switched networks into account

**Spatial Intra Prediction**

- **Input Video Signal**
- **Split into Macroblocks 16x16 pixels**
- **Decoder**
- **Transform/Quant.**
- **Motion Estimation**
- **Deblocking Filter**
- **Entropy Coding**

**Accuracy of Motion Description**

- **Input Video Signal**
- **Split into Macroblocks 16x16 pixels**
- **Decoder**
- **Transform/Quant.**
- **Motion Estimation**
- **Deblocking Filter**
- **Entropy Coding**

**Multiple Reference Frames**

- **Input Video Signal**
- **Split into Macroblocks 16x16 pixels**
- **Decoder**
- **Transform/Quant.**
- **Motion Estimation**
- **Deblocking Filter**
- **Entropy Coding**

**Integer Transforms**

- **Input Video Signal**
- **Transform/Quant.**
- **Motion Estimation**
- **Deblocking Filter**
- **Entropy Coding**

- 4x4 Block Integer Transform

\[
\begin{pmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 \\
1 & -2 & 2 & -1
\end{pmatrix}
\]

- Additional 2nd transform stage for resulting DCs in case of chroma blocks or luma 16x16 intra blocks by using a 2x2 Haar or 4x4 Hadamard transform
Deblocking Filter

- Deblocking within the loop by filtering of the edges of 4x4 blocks.
- Highly content adaptive filtering, where filtering strength is adjusted on three levels:
  - on slice level dependent on encoder (individual characteristics of the video)
  - on edge level dependent on intra/intra, motion estimation, motion compensation
  - on sample level with thresholds dependent on quantizer step size

Entropy Coding

Two alternatives:
- CAVLC (Context-Adaptive VLC)
  - Based on variable length codes
  - Switches between different VLCs
- CABAC (Context-Adaptive Binary Arithmetic Coding)
  - 10 – 20% bit rate savings vs. CAVLC
  - More complex than CAVLC

Example Coding Efficiency Comparison

60% bitrate savings vs. MPEG-2
35% bitrate savings vs. MPEG-4 ASP

Another Example

40% bitrate savings vs. MPEG-2

New Coding Tools in FReXt (1/2)

- 9 additional 8x8 spatial intra prediction modes (for luma only)
- Choice of intra prediction mode and block size (4x4, 8x8, or 16x16) on macroblock (MB) level

Outline

0 History and Overall Conceptual Structure
1 Core Coding Technology
2 Fidelity Range Extensions (FReXt)
3 Scalable Video Coding Extensions
New Coding Tools in FRExt (2/2)

- Additional 8x8 block size integer transform for luma
- Choice between 4x4 and 8x8 transform (on an MB level) under certain constraints imposed by the prediction block size
- Separable transform: $C_{8x8} = T_x \cdot B_y \cdot T_y$
- Easy implementation via fast butterfly operations using shifts and adds only
- Close approximation of length-8 DCT

$$\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 1 & 0 & -1 & -2 & -1 & 0 & 1 \\
3 & 2 & 1 & -1 & -2 & -1 & 0 & 1 \\
4 & 3 & 2 & 1 & -1 & -2 & -1 & 0 \\
5 & 4 & 3 & 2 & 1 & -1 & -2 & -1 \\
6 & 5 & 4 & 3 & 2 & 1 & -1 & -2 \\
7 & 6 & 5 & 4 & 3 & 2 & 1 & -1 \\
8 & 7 & 6 & 5 & 4 & 3 & 2 & 1
\end{pmatrix}$$

H.264/AVC Profiles and Related Tools

- High Profile (HP) is likely to replace Main Profile (MP) in typical consumer applications!

FRExt: Intra Coding Performance

- Two popular 512 x 512 monochrome images

FRExt: Example R-D Performance

- 52% bitrate savings for AVC HP
- Perceptual quality is what matters!
- Subjective tests have shown a 3:1 nominal gain for AVC HP vs MPEG-2 (as shown in next slide)

AVC HP vs MPEG-2: Subjective Results

- Test conducted in 2004 by Blu-Ray Disc Association (BDA) with studio participants
- Three movie clips at 1080i/2400Hz (1080p)
- 16 Mbps AVC visually transparent
The H.264/MPEG-4 AVC Standard: Core Coding Technology and Recent Extensions – D. Marpe

**R-D Performance using Hierarch. B Pics**

*Night (1280x720@60Hz)*

![R-D Performance using Hierarch. B Pics](image)

60% bitrate savings for AVC HP

**New Types of Temporal Referencing**

![New Types of Temporal Referencing](image)

**New Types of Temporal Referencing**

![New Types of Temporal Referencing](image)

**New Types of Temporal Referencing**

![New Types of Temporal Referencing](image)

**Hierarchical B Pictures**

- Decoupling of display order and coding order
- Decoupling of reference ability and picture type (not possible with previous standards)
- Can provide temporal scalability
- Can also be used for SNR scalability
- Can improve coding efficiency by proper bit allocation

![Hierarchical B Pictures](image)
**Technical Contributions of HHI**

- More than 80 standard contributions to H.264/AVC since 1999
- HHI Proposals adopted as normative parts of H.264/AVC:
  - Context-based adaptive binary arithmetic coding (CABAC) including fast binary M coder
  - Tree-structured macroblock partition
  - Multiple-frame / multi-hypothesis prediction
  - Adaptive 8x8 transform / spatial intra prediction
- More Functionality
  - Rate-Distortion (R-D) optimized coder control
  - Modified R-D optimized coder control for error-prone packet-switched networks (together with Munich TU)
  - Improved coder control for hierarchical B pictures

**Administrative Support of HHI**

- Co-Chair of Joint Video Team and Associate Chair of MPEG Video
- Co-Editor of Standards
- Co-Editor of AVC, Version 1-3: D. Marpe
- Co-Editor of SVC Amendment: H. Schwarz
- Editor of the visual part of TS 102 005 (DVB-AVC): T. Wiegand
- Maintenance of the Reference Software
- Software Coordinator: K. Sühring
- Chairs of different JVT Ad-Hoc Groups (AHG)
  - Test Model and Ref. Software AHG: K. Sühring
  - Text and Editing AHG: T. Wiegand
  - CABAC AHG: D. Marpe

**H.264 / MPEG-4 AVC Adoption Status**

- Mobile TV (use of Baseline profile)
- Digital Video Broadcasting – Handheld (DVB-H)
- Digital Multimedia Broadcasting (DMB)
- Multimedia Broadcast/Multicast Service (MBMS)
- SDTV / HDTV Broadcast / IPTV (use of High/Main profile)
- DVB: revised Implementation Guide TS 101 154 (DVB-C/RI)
- Direct to Home broadcast satellite, e.g., DirectTV: Dish Network (USA)
  - Sky HD, BBC HD (UK and Ireland)
  - Premiere: ProSiebenSat.1 (Germany)
- Terrestrial HDTV pay TV services in France
- Media Storage (High profile)
- HD-DVD specification of the DVD Forum
- BD-ROM specification of the Blu-Ray Disc
- Gaming / Entertainment (Baseline/Main/High profile)
  - Sony PSP and PS 3
  - Apple iPod
  - Microsoft Xbox (announced for 2007)

**Outline**

1. History and Overall Conceptual Structure
2. Core Coding Technology
3. Fidelity Range Extensions
4. Scalable Video Coding Extensions

**More Functionality**

- More flexible source coding, i.e., scalability is needed
  - Simple adaptation to different bitrates, frame rates or spatial resolutions of the video content by removal of parts of a bit stream
  - However, scalability is difficult to realize and often causes severe penalties in coding efficiency
  - Idea: Re-use as much of the existing H.264 / MPEG-4 AVC components such as hierarchical motion-compensated prediction together with some new coding elements
  - Basic architecture designed by Fraunhofer HHI (Schwarz, Wiegand, Marpe; started in Summer 2003)

**Architecture of AVC Scalable Extension**

- Base layer bit-stream
- H.264/AVC compatible lower layers
- Rate-Distortion (R-D) optimized coder control
- Modified R-D optimized coder control for error-prone packet-switched networks (together with Munich TU)
- Improved coder control for hierarchical B pictures
- More flexible source coding, i.e., scalability is needed
Spatial Scalability

- Motion information for spatial scalable coding
  - The trade-off between motion and residual rate highly influences the coding efficiency in hybrid video coding
  - Separate motion fields should be used for each layer in spatial scalable video coding
- General approach for spatial scalable coding
  - Independent coding of spatial layers with layer-specific motion parameters, but a common coding structure
  - Additional inter-layer prediction mechanisms in order to employ base layer information for efficient spatial/SNR scalable coding
- Inter-layer prediction techniques
  - Inter-layer intra prediction (similar to older standards)
  - Inter-layer prediction of motion information
  - Inter-layer prediction of residual information (prediction error)

Current Status of Standardization

- HHI proposal was selected as the first Working Draft (WD1) of prospective Scalable Video Coding (SVC) standard (January 2005)
- Most components are reused as specified in the standard
- Support of spatial, temporal and (coarse/fine-grain) SNR scalability as well as their combination
- Slightly worse coding efficiency of scalable extension (on avg.) compared to H.264/MPEG-4 AVC single layer coding
- But: Still work in progress ...
- Standardization work is conducted in the Joint Video Team (JVT)
- SVC will be an Annex / Amendment to H.264 / MPEG-4 AVC
- Final draft SVC standard expected to be ready in April 2007

Questions & Comments ?