Standardization in JVT: Scalable Video Coding

Jens-Rainer Ohm

RWTH Aachen University
Institute of Communications Engineering
ohm@ient.rwth-aachen.de
http://www.ient.rwth-aachen.de
Outline

- Introduction
- Summary on MPEG Call for Proposals on SVC and subsequent Core Experiments process
  - Results
  - Technical solutions
- Present status of SVC standard development in JVT
  - Temporal scalability and open-loop concepts
  - Spatial scalability
  - SNR and fine-granularity scalability
- Conclusions
Scalable Media – the Idea

- **Universal Media Access:**
  code once and then customize the stream to access content
  - “Anytime”
  - from “Anywhere” (i.e. using any access network - wireless, internet etc.)
  - and by “Anyone” (i.e. with any terminal complexity)
- **Compatibility** of different formats/resolutions

![Diagram showing Scalable Coded Content, Network, and Terminal with feedback loop]
MPEG/ITU-T Work Item on Scalable Video Coding

- **SVC Work item**
  - started as ISO/IEC 21000-13 (MPEG-21 Scalable Video Coding)
  - was moved to ISO/IEC 14496-10 January 2005

- **Timeline:**
  - October 2003: Call for Proposals
  - March 2004: Evaluation of proposals
  - January 2005: Working Draft (now within JVT)
  - October 2005: Committee Draft
  - March 2006 Final Committee Draft
  - July 2006 Final Draft International Standard
Scenario 2: Scalability over 2 spatial layers

MPEG Call for Proposals on SVC – Results

- Scenario 2 Overview

- Advanced scalable MC prediction

- MCTF+block transform

- MCTF+2DWT

<table>
<thead>
<tr>
<th>QCIF 1</th>
<th>QCIF 2</th>
<th>CIF 1</th>
<th>CIF 2</th>
<th>CIF 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOS x2 loss compared to anchor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- MOS x2 loss compared to anchor

- QCIF 1

- QCIF 2

- CIF 1

- CIF 2

- CIF 3

- A2

- J2

- K2

- B2

- H2

- D2

- I2

- E2

- L2

- C2

- G2

- F2
Following the Call

- Core experiments were run to further identify and develop SVC technology based on Scalable Video Model (SVM)
- Comparison made for wavelet-based and AVC/H.264-extending solutions
  - Full range of scalabilities tested (3x spatial, 3x temporal, medium-granularity SNR)
  - Formal visual tests (experts & non-experts) indicated superiority of AVC/H.264 solution, particularly at low rates and resolutions
JSVM and Working Draft

- **JSVM: Joint Scalable Video Model** (extending from MPEG Scalable Video Model 3.0)
- **Working Draft of 14496-10:2005/AMD1 Scalable Video Coding**
  - Extension of AVC / H.264 with compatible base layer
  - Motion-compensated temporal filtering (MCTF) with adaptive prediction and update steps for open-loop compression
  - Layered structure with "bottom-up" prediction from lower layers
  - One decoder loop
  - FGS functionality as extension of CABAC, modified interleaved scan order ("cyclic block coding")
  - New SVC-specific NAL unit types
  - Bitstream scalability at level of NAL packets
Motion-compensated Lifting Filters

- Temporal-axis lifting filters, MC/IMC in lifting flow
  - MC and IMC must be aligned
- H pictures similar to MC prediction P pictures
- Adaptive predict & update
  - Switching intra/forward/backward/bi-directional
- L pictures similar to I pictures when update turned down

H = A – MC(B)

Prediction

L = \frac{1}{2}[MC(A) + B]
MCTF as „Temporal-axis Wavelet Tree“

- No closed loop in prediction and update required
- Prediction loop over LLL.. possible
Layered Coding with Spatial Pyramids

- **Pyramid structure**
  - Basic building blocks from AVC / H.264
  - Spatial transform coding using integer DCT approximations (4x4 and 8x8)
Inter Layer Prediction

- **Motion**: Upsampled partitioning and motion vectors for prediction
- **Residual**: Block-wise upsampled residual (bi-linear)
- **Intra**: Upsampled intra MB (H.264 / AVC 1/2-pel filter)
Extended Spatial scalability

- Cropping and non-dyadic up-sampling
- Macroblocks not aligned between layers
SNR Scalability

- **Layered representation of L/H pictures**
- **Non-scalable ‘base-layer’** (BL)
  - Motion information
  - Coarsest representation of intra and residual data
  - Minimum acceptable reconstruction quality
- **Quality scalable enhancement layers** (EL)
  - Residue between original and SNR-BL representation
  - Doubled quantizer precision per SNR-EL
  - FGS: truncation of EL packets at arbitrary points
  - Refinement in the transform domain
  - Single inverse transform at the decoder side
- **“Dead substream” concept** allows flexible increase of quality beyond the limitations of layered coding
SNR & Spatial scalability: 2 spatial layer example
SNR & Spatial scalability: 2 spatial layer example

Video → 5/3 MCTF

Quality Layer = OR

AVC (with Hierarchical B pictures)

Residual Coding FGS

ProgressiveRefinement Quality Layer

Residual Coding CGS

Motion Coding (Opt.)

N^1 x Quality Layer

N^0 x Quality Layer

Texture

Motion (M_p^0)

Texture

Motion (M_p)

Bitstream

Multiplex
Realisation of FGS SNR Scalability

- **Three scans for FGS coding of transform coefficients**
  1. Non-significant transform coefficients of significant blocks
  2. Refinement symbols for already significant coefficients
  3. Coefficients of non-significant blocks

- **Scanning pattern for each scan**
  - Scanning from low to high frequency bands (zig-zag)
  - Inside each band: first luma, then chroma in raster scan

- **Coding of transform coefficient levels**
  - Non-significant coefficients
    - Coded Block Pattern, Coded Block Bit, DeltaQP, Transform Size
    - CABAC symbols (SIG, LAST, SIGN, ABS)
  - Significant coefficients
    - Refinement symbol (-1, 0, +1)
Bitstream structure

- Data cube model and layered stream representation

---

Layer 1
- T0, S1, B0

Layer 1
- T1, S1, B0

Base-Layer
- T1, S0, B0

Layer 2
- T2, S1, B0

Layer 2
- T2, S0, B0

Layer 3
- T0, S2, B0

Layer 3
- T1, S2, B0

Layer 3
- T2, S2, B0

Temporal
- 7.5 fps
- 15 fps
- 30 fps

Spatial
- QCIF
- CIF
- 4CIF

SNR
New SVC NAL unit types can be ignored by conventional AVC parser
Performance Example

CREW 176x144 15

Rate [kbit/s]

PSNR [dB]

- JSVM2 CombinedScalability
- JSVM2 SingleLayer
- JSVM2 SnrScalability
- JSVM2 SpatialScalability
Performance Example

CREW 352x288 30

Rate [kbit/s]

PSNR [dB]

- JSVM2 CombinedScalability
- JSVM2 SingleLayer
- JSVM2 SnrScalability
- JSVM2 SpatialScalability
Performance Example

CREW 704x576 60

Rate [kbit/s]

PSNR [dB]

- JSVM2 CombinedScalability
- JSVM2 SingleLayer
- JSVM2 SnrScalability
- JSVM2 SpatialScalability
Summary

- **SVC to become an extension of H.264 / AVC**
  - Residual coding using existing tools
  - MCTF open loop structure with update step
  - Pyramid structure with inter layer prediction
  - FGS quality layers plus a minimum quality base layer

- **Good compression performance**
  - tradeoffs by the amount of flexibility in scalability

- **Further improvements expected e.g. by**
  - adaptive MCTF structures
  - combinations closed/open loop
  - improvement of quantization & FGS schemes
  - cross-layer dependencies of texture and motion
  - joint RD optimization over multiple layers