The Latest Development on MPEG Video Coding Standards

National Chiao Tung University

蔡淳仁 Chun-Jen Tsai
03/8/2004
Outline

- Introduction
- MPEG-4 Part 10: AVC (a.k.a. H.264)
  - Design Concept
  - What Makes the Differences
  - Licensing Issues
- MPEG-21 Part 13: Scalable Video Coding
  - MPEG SVC Call-for-Proposal
  - Potential Technologies
- Conclusion
Introduction

- MPEG-2 is an extremely successful video codec
- MPEG-4 SP and ASP is gaining momentum, but does it fulfill the demands of new applications?
- New tricks in the MPEG toolbox:
  - Back to the basics (well, sort of): MPEG-4 AVC
  - Jump into the future: MPEG-21 SVC
MPEG-4 AVC/ITU-T H.264

- Developed by the Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG
- Became an International Standard in May, 2003
- Goal: 50% coding efficiency gain over MPEG-4 ASP
- Three layer design:
  - Video Coding Layer (VCL):
    - How to compress source video
  - Network Abstraction Layer (NAL):
    - How to pack bitstream data error-resiliently
  - Transport Encapsulation Layer (TEL):
    - How to carry the NAL units over real transport systems
    - This layer was a “spin-off” from the original NAL design
Video Coding Layer Concept

- Compression efficiency was the top priority
  - Bit-error resilience is no longer an important issue
- Slice-based operation ("picture-based" is obsolete?!)  
  - Packet-based transport systems dominate the future
- Information provided by the transport (system) shall not present in VCL (e.g. timestamps)
VCL Key Features

- Predictive coding
  - 13-mode luma, 4-mode chroma intra prediction
  - Gazillion modes for inter prediction
- 16-bit integer combined transform/quantization
  - Exact forward-inverse transform pair is used
  - Transform block size is 4x4
- Two types of entropy coding methods
  - Universal VLC and Context Adaptive VLC
  - Context Adaptive Binary Arithmetic Coding
- In-loop filter
Controversial VCL Features

- Flexible Macroblock Ordering (FMO), Arbitrary Slice Ordering (ASO), Data-Partitioning (DP) are controversial “VCL” features
- For example, FMO causes hardware design problem:
  - MBs in a frame can be assigned to different slice groups
  - Each slice group is further divided into slices

![Diagram of slice groups](image-url)
Network Abstraction Layer Concept

- NAL is a packet-based compressed data format mainly designed for error-resilience.
- NAL is suitable for both packet-oriented and bitstream-oriented transports.
- Each NAL unit carries some video or system data.
- In bitstream-oriented transports, each NAL can be preceded by a start code prefix → requires a start code emulation prevention mechanism.
Not Much News in the Big Picture

- For example, the encoder diagram:
It’s All in the Details

- For example, the gazillions of coding modes:

- Or, the complicated references frame patterns:
So, How Much Do We Gain?

- AVC Baseline vs. SP

- MPEG also conducted a verification test. AVC achieves 2x coding efficiency gain over competing codecs in roughly 77% of the test cases (from 48 kbps to 20Mbps)
Potential AVC Killer: Licensing Issue

- JVT original goal was to make Baseline free
  - Some “better” technology was kicked out for this reason
  - Baseline tools used in Main Profile are not free
- MPEGLA does not allow different licensing terms for same technology
- Current term is especially harsh for the digital broadcasting community. EBU already published an open statement to boycott the new standard if licensing terms don’t change
Scalable Video Coding

- There are three factors (dimensions) that determines the perceptual quality of a video presentation:
  - Picture Resolution
  - Frame rate
  - Bitrate
- Traditionally, these parameters was fixed once the coding is done
- For SVC, we want to be able change these parameters on-the-fly during the presentation
Application Scenario

On the road

Cellular Network (RAN)

In the office

Internet (WAN)

802.11 (WLAN)
Theoretical Goal of SVC

- Change quality smoothly in real time

- Million dollar question: how to measure “quality” in 3D scalable space?
# MPEG SVC Call-for-Proposal

## Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec. 1, 2003</td>
<td>Preliminary intention to participate</td>
</tr>
<tr>
<td>Dec. 31, 2003</td>
<td>Deadline for Pre-registration</td>
</tr>
<tr>
<td>Feb. 1, 2004</td>
<td>Formal registration (€ 1,500-2,000)</td>
</tr>
<tr>
<td>Feb. 16, 2004</td>
<td>Coded test material at the test site</td>
</tr>
<tr>
<td>Feb. 20, 2004</td>
<td>Subjective assessment starts</td>
</tr>
<tr>
<td>March 1, 2004</td>
<td>Registration/submission of documents</td>
</tr>
<tr>
<td>March 9, 2004</td>
<td>Report of the subjective test results</td>
</tr>
<tr>
<td>March 15-19, ‘04</td>
<td>MPEG 68th meeting, München, DE</td>
</tr>
</tbody>
</table>
Status of the CfP

- 21 registrations of proposals (from 9 academic institutes/ 12 companies)
- 11 proposals are based on inter-frame wavelet, 5 proposals are based on DCT, and 5 proposals are based on undisclosed technologies.
- Two test scenarios:
  - Scenario 1: fully scalable codecs from 64kbps to 6Mbps; and from QCIF to 4CIF (13 proposals)
  - Scenario 2: limited scalable codecs from 48kbps to 1Mbps; and from OCIF to CIF (14 proposals)
Potential Technologies

- **DCT-based:**
  - MPEG-4 SSP, FGS are well known, but lacks full-dimensional scalability and coding efficiency.
  - New DCT-based techniques are based on AVC, use multiple prediction loops, and progressive layer approach to increase both coding efficiency and scalability.

- **Wavelet-based:**
  - Most candidates are based on Prof. John Woods’ 3D Subband Coding approach (and his reference code).
  - Full-dimensional scalability with decent high bitrate coding efficiency.
  - The performance under low bitrate (below 512kbps) requires some improvements.
DCT-based Example: SRFGS


- RFGS Concept:

  - EL (Enhancement Layer)
  - BL (Base Layer)

  ![Diagram](image)
SRFGS

ELₙ (Last Enhancement Layer)

ELₜ (Second Enhancement Layer)

ELₐ (First Enhancement Layer)

BL (Base Layer)

\[ E_{N, n-1} \rightarrow QE_{N, n} \rightarrow E_{N, n} \]

\[ E_{B, n-1} \rightarrow QE_{B, n} \rightarrow E_{B, n} \]

\[ E_{A, n-1} \rightarrow QE_{A, n} \rightarrow E_{A, n} \]

\[ B_{n-1} \rightarrow O_{n} \rightarrow B_{n} \]

--- Prediction ---

\[ \text{Quantization Error} \]

--- Reconstruction ---
Rate-Distortion Plot for Mobile Seq.
Wavelet-based Example: MC-EZBC

MCTF: Motion Compensated Temporal Filtering
EZBC: Embedded Zeroblock Coding
AC: Arithmetic Coding
Temporal Subband Decomposition

GOP

Stage 1

Stage 2
MCTF Concept
Spatial Subband Decomposition

- **Decomposition**
  - $x(m,n) \rightarrow h_0(m) \downarrow 2 \rightarrow h_0(n) \rightarrow LL$
  - $h_1(m) \downarrow 2 \rightarrow h_1(n) \rightarrow LH$
  - $2 \rightarrow h_0(n) \rightarrow HL$
  - $2 \rightarrow h_1(n) \rightarrow HH$

- **Spatial Scalability**
  - $LH_2$ $HH_2$ $HL_1$ $HH_1$
Rate-Distortion Plot for Mobile Seq.

Mobile Calendar

PSNR (dB) vs. Rate (Kbps) for different encoding methods:
- FGS
- MC_EZBC
- AVC
Motion Information Scalability

- A Proposal from NCTU:
  S.S Tsai, H.-M. Hang, Tihao Chiang, “Motion Information Scalability for MC-EZBC,” MPEG Document M9723

- Compressed data = motion info. + residual wavelet-transformed image data

- Partition of motion vector by size
  - Base: 64-by-64 to 16-by-16
  - Enhancement: 8-by-8 to 4-by-4
Rate-Distortion Plot for Harbour Seq.
Conclusions

- It is still unclear whether AVC and/or SVC can be as successful as MPEG-2 due to various non-technical reasons.
- In the long run (5 to 10 years), SVC may be more important than AVC since channel bandwidth/storage space is increasing rapidly. On the other hand, content authoring cost and device adaptability demand are also increasing rapidly.