Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding

Authors: Bart Masschelein(*), Jiangbo Lu(†), and Iole Moccagatta(*),

* Multimedia Group, IMEC, Kapeldreef 75, B-3001, Leuven, Belgium
† Department of Electrical Engineering, University of Leuven, Belgium

Presented by Bart Masschelein
2007 IEEE International Conference on Consumer Electronics (ICCE)
January 11th, 2007

Outline

• Introduction - The MPEG and ITU-T Standardization Bodies [7]
  » Overview and accomplishments (MPEG-* and H.26*)
  » Joint Video Team
  » The MPEG-4 video standards

• Chapter 1 - H.264/MPEG-4 AVC [24]
  » New (wrt. MPEG-4 Part 2) tools
  » Profiles and levels
  » Market acceptance and deployment

• Chapter 2 - H.264 Annex G/MPEG-4 Scalable Video Coding [56]
  » Why do we need scalable video
  » MPEG-4 SVC requirements and targeted apps.
  » Historical overview and standardization timeline
  » Tools enabling scalability
  » Media Aware Network Element (MANE)
  » Profiles and levels

• Chapter 3 - MPEG-4 Multi-view Video Coding [28]
  » History, context, and motivations
  » Typical multi-view system setups and application scenarios
  » Standardization and requirements on MVC
  » Basic concepts and principles of MVC
  » Coding tools currently considered

• Annex [15]
  » What else is going on as far as video coding standards (MPEG, ITU-T, SMPTE)
  » Bibliography and reference points
  » Acronyms and abbreviations
Introduction – The MPEG and ITU-T Standardization Bodies

The MPEG and ITU-T Standardization Bodies - Outline

- Overview and accomplishments (MPEG-* and H.26*)
- Joint Video Team
- The MPEG-4 video standards
Moving Picture Experts Group (MPEG):
ISO-IEC/JTC1/SC29/WG11

**MPEG Standards**
  - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s
  - Storage and retrieval of audio and video on compact disc
  - Efficient encoding of non-interlaced pictures with roughly VHS quality at 1.15 Mbit/s
  - Generic coding of moving pictures and associated audio information, interlace support
  - Digital video broadcasting (2-6 Mbps for SD, 10-20 Mbps for HD), DVD video (6-8 Mbps)
  - Coding of audio-visual objects (from frame-based to object-based coding)
  - Coding and integration of natural (aka, pixel based) still images and video sequences with synthetic content (aka, 2D and 3D graphic geometry, i.e. compression of wire grid parameters, synthetic text)
  - Efficient coding ranging from very low bit-rate (ex: 5 kbit/s) to more than 1 Gbit/s
  - Support progressive as well as interlaced video from sub-QCIF to 'Studio' resolutions (4k x 4k pixels)
  - Wide range of targets: from video streaming to real-time video conference, from DTV to interactive graphics applications
- **ISO/IEC 15938 MPEG-7 (2001)**
  - Multimedia content description interface
  - Different from the previous MPEG standards in the sense that what is represented is not the information itself, but the information about the information
- **ISO/IEC 21000 MPEG-21 (200x)**
  - Multimedia framework
  - Support environment to deliver and use all content types by different categories of users in multiple application domains
  - Define Users (anybody in the value network) and Digital Items (assembly of content) on which Users execute Actions that generate other Digital Items that can become object of Transactions.
Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding

Tutorial presented at ICCE 2007 © imec 2006

Video Coding Expert Group (VCEG):
ITU-T SG 16 Q.6

ITU

ITU-R

ITU-T

ITU-D

SG1

SG2

... Study Group 16 - Multimedia

WP1 – Modems and interface
V.34, V.25ter

WP2 – Systems
H.320 – ISDN
H.323 – LAN
H.324 – PSTN
T.120 - DATA

WP3 – Coding
G.7xx – Audio
H.261 – Video
H.263 - Video

ITU-T SG 16 Q.6 Standards

  - ISDN picture phones and for video conferencing systems
  - Image format: CIF (352 x 288 Y samples) or QCIF (176 * 144 Y samples), frame rate 7.5 to 30 fps
  - Bitrate multiple of 64 kbps (= ISDN-channel), typically 128 kbps including audio.
  - Stand-alone videoconferencing system or desk-top videoconferencing system, integrated with PC

- ITU-T Rec. H.262 = MPEG-2

  - International standard for picture phones over analog subscriber lines
  - Image format usually CIF, QCIF or Sub-QCIF, frame rate usually below 10 fps
  - Arbitrary bitrate (typically 20 kbps for PSTN)
  - Software-only PC video phone or TV set-top box
  - Compression engine for Internet video streaming
  - MPEG-4 Part 2 decoder decodes H.263 Baseline (short header mode)

- ITU-T Rec. H.264 = MPEG-4 AVC
  - Annex G = MPEG-4 SVC

[Bernd Girod: EE398B Image Communication II Video Coding Standards H.261&H.263 no. 7]
Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding

Tutorial presented at ICCE 2007 © imec 2006

Joint Video Team (JVT)

ISO/IEC
SC29/WG11

ITU-T
SG16/VCEG

MPEG-21

MPEG-4

MPEG-7

... JVT

H.262/MPEG-2

H.264/AVC

SVC

MVC

... H.261

H.263

H.263+

MPEG-21

MPEG-4

MPEG-7

... JVT

H.262/MPEG-2

H.264/AVC

SVC

MVC

... H.261

H.263

H.263+

The H.264/MPEG-4 Video Standards

• MPEG-4 Part 2
  - IS: 2004 (status: 3rd edition)

• ITU-T Rec. H.264/MPEG-4 Part 10 Advanced Video Coding (AVC)
  - IS: 2005 (status: 3rd edition)

• AVC Amd. 1: Support of colour spaces
  - FDIS: October 2006 (status: FPDAM)

• AVC Amd. 2: Advanced 4:4:4 profiles
  - FDIS: January 2007 (status: PDAM)

• ITU-T Rec. H.264 Ann. G/AVC Amd. 3: Scalable Video Coding (SVC)
  - FDIS: January 2007 (status: PDAM)

• AVC Amd. 4: Multi-view Video Coding (MVC)
  - FDIS: January 2008 (status: WD)
Chapter 1 - H.264/MPEG-4 AVC

H.264/MPEG-4 AVC - Outline

- Focus on new (wrt. MPEG-4 Part 2) tools
- Video Coding Layer (VCL) and Network Adaptation Layer (NAL)
- Some coding tools in details:
  - FMO
  - Intra Prediction
  - Motion Compensated Prediction
  - Transform and Quantization
  - In-loop De-blocking Filter
  - Entropy Coding
- Profile and Levels
- Market Acceptance and Deployment
**H.264/MPEG-4 AVC Layer Structure**

- **Motivation:** network adaptation

- **Video Coding Layer**
  - Coded Macroblock
  - Data Partitioning
  - Coded Slice/Partition

- **Network Abstraction Layer**
  - H.320
  - MP4FF
  - H.323/IP
  - MPEG-2
  - etc.

---

**MPEG terminology**

- **Macroblock**
  - 16x16 luma, 2 8x8 chroma
  - macroblocks within a slice depend on each other
  - macroblock can be further partitioned

- **Slice**
  - a picture can be split into 1 or several slices
  - slices are self-contained, (only de-blocking may be performed across slices)
  - slices are a sequence of macroblock
  - sequence may not be consecutive
Flexible Macroblock Ordering (FMO)

- Transmit macroblocks in non-raster scan order
- Slice group
  - pattern of macroblock defined by a slice group map
  - a slice group may contain from 1 to several slices
- Slice group map types
  - interleaved
  - dispersed (checker board)
  - explicitly assign a slice group to each macroblock
  - one or more "foreground" slice group and a "leftover" slice group

Motivation: error resilience and ROI

The H.264/MPEG-4 AVC Codec Structure
Intra Prediction (1/2)

- Spatial domain prediction for Intra macroblocks
- 4x4 Intra Prediction (luma)
  - 9 modes
  - 13 available and reconstructed (no db) border samples of neighboring blocks
  - modes are coded w.r.t. "most probable" one
- Example: Diagonal-Left

Intra Prediction (2/2)

- 16x16 Intra Prediction (luma)
  - 4 modes
  - 16 vert. + 16 horiz. available and reconstructed (no db) border samples of neighboring macroblock
  - mode is transmitted
- 8x8 Intra Prediction (chroma)
  - 4 modes
  - 8 vert. + 8 horiz. available and reconstructed (no db) border samples of neighboring blocks
  - mode is transmitted
- Can predict from Intra or non-Intra macroblock
Quarter Sample Interpolation

- **Luma**
  - separable 6-tap filter for ½ pel
  - averaging integer and ½ pel position for ¼ pel
- **bilinear interpolation for chroma (down to 1/8)**

![Quarter Sample Interpolation Diagram]
### Multiple Reference Frames

- **Multiple picture buffer**
  - Decoded Picture Buffer
  - FIFO or sliding window
  - adaptive memory control
  - 16 pictures max (memory is constrained)

- **per-8x8 reference control**
- **Bi-predicted picture:** 2 sets of motion vector per block
New Types of Temporal Referencing

- Known dependencies (MPEG-1, MPEG-2, etc.)

- New dependencies
  - referencing order and display order are decoupled
    - IBBPBBP... vs. IBBPBBBBPBP...
  - referencing type and picture type are decoupled
    - B slices can be used as reference

Transform Coding (1/2)

- 4x4 Block Integer Transform
- Repeated transform of DC coeffs for 8x8 chroma and some 16x16 intra luma blocks
Transform Coding (2/2)

- Separable 4x4 transform
  - exact match inverse transform (integer arithmetic)
  - 16-bit arithmetic
  - easy implementation (adds and shifts, no multiplications)

- Hadamard transform
  - transform coeff. that covers the entire macroblock
  - 4x4 for Intra 16x16 MB (DCs only)
  - 2x2 for Chroma DCs
  - adds only

In-Loop De-blocking Filter

- filter edges of 4x4 block
- highly content adaptive
  - slice level: alpha and beta control global filter strength
  - edge level: MB type, motion, and # coded coeff. control filtering strength
  - sample level: quantizer dependent threshold control filtering for each sample
- filtering should be done after entire macroblock has been decoded
- vertical edges first, then horizontal edges
- filter both luma and chroma
- may cross slice boundary
Entropy Encoding

Variable Length Coding

- **Exp-Golomb code**
  - almost all symbols except transform coefficients
- **Context Adaptive VLC (CA-VLC)**
  - Context are built dependent of transform coefficients
  - Transform coefficients are scanned backward
  - TotalCoeff (replace end-of-block) and TrailingOnes
    - 4 VLC tables for luma and 1 VLC table for chroma
    - selection based on transform coeff. of 2 neighboring blocks (block above and on the left)
  - Coefficients levels
    - sign of TrailingOnes
    - bits spent to represent one level is function of previously encoded levels (max range is variable)
  - Run of zeros preceding each coefficient
There is more ....

- More coding tools
  - Motion vector prediction using motion vectors from 4 neighboring blocks
  - Adaptive Weighted prediction (generalized B slices)
    - each prediction sample can be weighted
    - an offset can be added
  - Interlaced coding
    - field or frame coding
    - macroblock adaptive frame/field coding
  - Context-adaptive Binary Arithmetic Codec (CABAC)
- More error resilience and network adaptation tools
  - Parameter set structure
  - Network Adaptation Layer (NAL) syntax structure
  - Arbitrary Slice Ordering (ASO)
  - Data Partitioning (DP)
  - Redundant Slices
  - SP/SI synchronization/switching slices
- More sideband information
  - Supplemental Enhancement Information (pan-scan, cropping, etc.)
  - Video Usability Information (aspect ratio of luma sample, overscan, etc.)

Profile and Levels (1/2)

- Main profile
- Extended profile
- Baseline profile

Profiles and Levels are de-coupled
- all profiles supports all levels
- levels supports from QCIF@64Kb/s to 2kx4k@240Mb/s
Profile and Levels (2/2)

- 4 High profiles (formally FRExt)
  - High (8-bit/sample video for 4:2:0)
  - High 10 (10-bit/sample for 4:2:0)
  - High 4:2:2 (10-bit/sample for 4:2:2)
  - 3 new features:
    - adaptive block-size transform
    - encode specified perceptual-based quantization scaling matrices
    - lossless coding ROI
- Few level and/or profile-dependent constraints
  - max number of motion vectors per two consecutive MBs for some levels (all profiles)
  - max number of slices for Main Profile (22 for SD-TV)
  - limit on motion vectors "spreading" for Simple Profile
  - etc.

H.264/MPEG-4 AVC Coding Efficiency Performance

Fig. 7: (a) – (e) Comparison of R-D curves for MPEG-2 (MP2), MPEG-4 Part 2 ASP (MP4 ASP) and H.264/AVC (MP4 AVC). I frames were inserted every 15 frames (N=15) and two non-reference B frames per reference I or P frame were used (M=3)

[Sullivan, SPIE, Aug. 2004]
**H.264/MPEG-4 AVC Market Acceptance (1/2)**

- **Disclaimers**
  - not intended to be comprehensive
  - public information, out-of-date at any point in time
  - major sources: Gary Sullivan and Thomas Wiegand

- **3GPP**
  - Baseline profile @ level 1b
  - recommended codec for PSS, MMS, and PSC in Rel. 6 (3GPP)
  - optional codec for streaming service (3GPP2)

- **Mobile TV**
  - Baseline profile @ levels 1b – 2.0
  - DVB: mandatory codec – Trial in Berlin
  - DAB: mandatory/single codec in T-DMB
  - ARIB: mandatory codec for mobile segment broadcast over ISDB-T
  - 3GPP: recommended codec for MBMS in rel

**H.264/MPEG-4 AVC Market Acceptance (2/2)**

- **SDTV and HDTV**
  - Main/High profiles @ levels 3.0 – 4.0
  - DVB: additional codec to MPEG-2 in services over MPEG-2 TS, single codec in services over IP
  - ISMA: mandatory codec in version 2

- **DVD Forum**: mandatory video decoding for HD-DVD
- **IPR Licensing**: MPEG-LA and Via Licensing Pool
- **Support in players**: Apple QuickTime (internet streaming)
H.264/MPEG-4 AVC Deployment

- Videoconferencing Systems: baseline profile @ levels 1b – 4.0 by 8 companies (Polycom, Tandberg, Sony, France Telecom, etc.)
- Consumer Electronics: Sony (PSP)
- ASIC
  - Mobilygen (wireless video)
  - Broadcom, Conexant, DG2L, Harmonic, Cradle Technologies, Modulus Video, Motorola, etc. (broadcast encoding and transmission)
- PC and DSP S/W Solutions: Intel, Envivio, Equator, FastVDO, Hantro, etc.
- Test Equipment: Tektronix and Vqual

→ AVC/H.264 gaining, but MPEG-2 still strong!

Chapter 2 –
H.264 Annex G/MPEG-4 Scalable Video Coding
H.264 Annex G/MPEG-4 Scalable Video Coding (SVC)

- Outline

- Why do we need scalable video
- MPEG-4 SVC requirements and targeted apps.
- Historical overview and standardization timeline
- Tools enabling scalability
- Media Aware Network Element (MANE)
- Profiles and levels

Traditional Video Coding
Increasing Demands

Simulcast: Encoding
Simulcast: Decoding

Dramatic Increase of Heterogeneous Devices
Solution: from Simulcast to Scalable Video Coding

MPEG-4 SVC: An Example Application

• Digital Video Surveillance
  – live monitoring in a multi-point surveillance system → low delay, pyramid scalability, temporal random access
  – storage → multipath scalability, temporal random access
  – interactive surveillance → random spatial access
  – mosaic display, image analysis, robust video streaming, etc.
  – Bosch, GE, Panasonic
Requirement: (Extended) Spatial scalability

Extended Spatial Scalability

- non-dyadic scaling
- cropping
- combination of cropping and non-dyadic scaling
Extended Spatial Scalability: Usage Scenario

16:9 HDTV

cropping and scaling

4:3 SDTV

http://products.sel.sony.com/hdtv/

Requirement: Temporal Scalability
Requirement: Quality Scalability

Requirement: Region of Interest scalability

- useful when bandwidth limitations
- real-time zooming and panning at receiver side
Additional Requirements

- Complexity scalability
  - Including a low complexity codec
- Coding efficiency performance
  - Coding penalty of max. 10% for same perceived quality
- Low end-to-end delay / fast random access
- Robustness to transmission errors
- Uniform way to manipulate and adapt scalable streams, to be mapped to widely used protocols

MPEG-4 SVC Targeted Applications (1/3)

- Wireless LAN video in home network
  - time-varying bandwidth \(\rightarrow\) fine grain SNR plus limited spatial scalability for managing wide bandwidth variations
  - device adaptation \(\rightarrow\) combined spatial/temporal/SNR, complexity scalability, arbitrary spatial resolution ratio’s, hybrid combinations of progressive and interlaced
  - Thomson, France Telecom, Philips, Siemens
- Broadband video distribution
  - ADSL clients, from DVB to DSL
  - client: display/processing adaptive video stream from low-quality (400Kbps@CIF) to HD-TV \(\rightarrow\) combined spatial/temporal/SNR/complexity with medium grain
  - operator/service: decoupling encoding from streaming, saving storage and bandwidth, content leverage \(\rightarrow\) simultaneous management of HD/SD (combined scalability)
  - France Telecom, Thomson, Philips
- Scalable video storage with erosion functionality
  - importance of the recorded video data, and thus its quality, decreases over time
  - Bosch, GE, Panasonic
MPEG-4 SVC Targeted Applications (2/3)

- MPEG-21 DIA based adaptation
  - technology to adapt scalable bit-streams to different devices and network → standardized system interface (e.g. bit-stream syntax)
  - Siemens AG, Deutsche Telekom, France Telecom

![Diagram of network topology]

Applications and Requirements for SVC [N6880, Jan 2005]

MPEG-4 SVC Targeted Applications (3/3)

- Video telephony/conferencing
  - Nokia, NTT, Siemens

- Mobile streaming/broadcast video
  - Thomson, Nokia, France Telecom, Orange, SFR, Samsung, Siemens

- Professional video manipulation
  - Thomson
Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding

**History**

**Exploration phase**

- **Temporal:** MCTF/Hierarch. B-frames
- **Spatial:** wavelets/DCT-based
- **Quality:** bitplane/layered coding

**Working draft:**

- Multiscale pyramid 2D+t
- Subjective and objective comparison

**Temporal:** MCTF

- Spatial: Layered approach
- Quality: Layered approach
Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding

Tutorial presented at ICCE 2007 © imec 2006

History

CfP [N5958]

Wavelet Exploration Group

WF 79, Marakech

WD

N8043

FDIS

Jan '07

79, Marakech

WAVELET VIDEO CODING

wavelets promising for high resolution, extended scal. and fine grain scal.

Three working modes:
T+2D: drift problems at full spatial resolutions
2D+T: compensate for shift-invariant nature of wavelets
2D+T+2D: intermediate solution

Status report on Wavelet Video Coding

subjective test indicate (compared to JSVM):
• significantly worse for combined scalability
• comparable/slightly worse for standalone SNR scal.

"Exploration discontinued until further proof of usefulness"
M11756: Proposed requirement modification

"The base layer of the standard shall be compliant with MPEG-4 AVC. The standard shall be capable of decoding MPEG-4 AVC bit streams"

MPEG-4 Part 10 Amd3

SVC continued in JVT

History

Further exploration

Core Experiments
AdHoc Groups

signaling layer dependencies
high-level syntax
FGS refinement
additional inter-layer modes

...
History

- Temporal scalability: Hierarchical B-frames
- Spatial scalability: Layered approach ESS
- Quality scalability: Layered approach for CGS MGS Bitplane coding for FGS

SVC Encoder Block Diagram
Temporal Scalability in MPEG-4 SVC

- Hierarchical B-frames
- Concept already existed in H.264/AVC

Hierarchical B-frames
Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding

Tutorial presented at ICCE 2007 © imec 2006

Hierarchical B-frames: performance

- High spatial detail
- Smooth motion
- Strong local motion

H.264/MPEG-4 AVC: classical prediction, i.e. 1st picture IDR, following ones Ps and Bs (IPP..., IBPBP..., IBBPBBP...) vs. hierarchical B-frames (GOP = 4, 8, 16, 32)

NOTE: encoder optimizations to improve coding eff. of hierarchical prediction


Spatial Scalability in Simulcast

Spatial decimation

MC & Intra prediction

texture motion

entropy coding

00101101110110111011011101001
01110101101110110111001101101
Spatial Scalability in MPEG-4 SVC

- Layered approach

- Three techniques:
  - Inter-layer Intra Texture Prediction
  - Inter-layer Motion Prediction
  - Inter-layer Residual Prediction

- Some concepts already existed in MPEG-2/4 for spatial scalability
Inter-layer Intra Texture Prediction

- provided in the Intra_Base macroblock mode
- constrained, allowing for single loop decoding at target layer (no multiple motion compensation processing)

Inter-layer Motion Prediction

- Base Layer Mode
  - macroblock partitioning, motion vectors and reference indices of base layer are used in enhancement layer
  - motion vectors are multiplied by the resolution ratio between the enhancement layer and its base layer

- Quarter Pel Refinement Mode
  - same as base layer mode
  - for each motion vector a quarter-sample motion vector refinement is additionally transmitted and added to the derived motion vectors
Inter-layer Residual Prediction

- only code the difference between current layer residual information and previous layer residual information
- adaptive, since only beneficial when motion vectors are (nearly) identical between both layers
- residual information is upsampled according to the spatial resolution ratio between the layers

Inter-layer Prediction Techniques: Performance

![Graph showing bit-rate vs. Y-PSNR for different prediction techniques]

This difference is the cost of spatial scalability!

[Schwarz_IWSSIP_2005]
Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding

Tutorial presented at ICCE 2007 © imec 2006

Inter-layer Prediction Techniques: Performance

This difference is the cost of spatial scalability!

[Schwarz_IWSSIP_2005]

Quality Scalability in MPEG-4 SVC

3 types of quality scalability technologies:

- Coarse Grain Scalability
- Fine Grain Scalability
  - aka Progressive Refinement
- Medium Grain Scalability
Coarse Grain Scalability (1/3)

QP2 < QP1

QP = Quantization Parameter

Path: MC & Intra prediction \rightarrow entropy coding \rightarrow multiplexer

Texture: QP2
Motion: QP1

R2

Coarse Grain Scalability (2/3)

QP2 < QP1

QP = Quantization Parameter

Path: MC & Intra prediction \rightarrow entropy coding \rightarrow multiplexer

Texture: QP2
Motion: QP1

R2 + R1

Inter-layer prediction techniques

R1
Coarse Grain Scalability (3/3)

- Same concepts as with spatial scalability, but without the upsampling operations
- Coding efficiency is optimized for coarse rate graduations (factor 1.5-2 from one layer to the next)
Fine Grain Scalability (2/2)

- Bitplane coding
- Ordering of data from most to least significant

![Bitplane coding diagram](image)

- enables truncation of NAL units (bit packets) at any arbitrary (byte-aligned) point
- Useful when in need to reduce bit rate by a few per cent to adapt to current network conditions

Medium Grain Scalability (1/2)

- CGS is computationally simple, but only allows a discrete set of extraction points
- FGS allows a fine granular set of extraction points, but is computationally complex
- MGS increases the number of CGS extraction points by reusing the quality level and layer syntax used in FGS
Medium Grain Scalability (2/2)

![Diagram showing scalability and quality](image)

**Quality scalability: performance**

- Single/base layer: H.264/MPEG-4 AVC, 1st picture IDR, GOP = 16
- QP lowest PSNR layers - QP highest PSNR layers = 12 (factor 4 in bit-rate)
- DQP = QP PSNR layer_i - QP PSNR layer_(i+1)
- Motion = motion parameters of the enhancement layer are adaptively refined
Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding

Tutorial presented at ICCE 2007 © imec 2006

### Bringing it Together: Combined Scalability

**3D Scalability Space**

- **no strict notion of layer**

- **CIF@30Hz**

- **CIF@15Hz**

- **QCIF@30Hz**

(Schwarz, ICME’05)

---

**Embedded Scalability: Coding Efficiency Cost**

<table>
<thead>
<tr>
<th>Format</th>
<th>Bit rates (kbps/sec)</th>
<th>[Schwarz, ICME’05]</th>
</tr>
</thead>
<tbody>
<tr>
<td>QCIF 15Hz</td>
<td>96 112 128 160</td>
<td></td>
</tr>
<tr>
<td>CIF 7Hz</td>
<td>192 224 256 320 384</td>
<td></td>
</tr>
<tr>
<td>CIF 15Hz</td>
<td>256 320 384 480 512</td>
<td></td>
</tr>
<tr>
<td>CIF 30Hz</td>
<td>448 512 640 768 896</td>
<td></td>
</tr>
<tr>
<td>CIF 60Hz</td>
<td>1024 1280 1536 1920 2464</td>
<td></td>
</tr>
<tr>
<td>CIF 90Hz</td>
<td>1536 1780 2038 2560 3072</td>
<td></td>
</tr>
</tbody>
</table>

The difference between the two points is the cost of embedded scalability!

Note: the original target (from MPEG Req.) was 10% cod. eff. loss in exchange of embedded scalability.
Media Aware Network Element (MANE)

- A network element, such as a router or application layer gateway, that is capable of parsing certain aspects of the RTP payload headers or the RTP payload and reacting to the contents
- MANEs are aware of the bitstream signaling
  MANEs allow packets to be dropped according to the needs of the media coding.

Example: if a MANE has to drop packets due to congestion on a certain link, it can identify those packets whose dropping has the smallest negative impact on the user experience and remove them in order to remove the congestion and/or keep the delay low.

Bitstream Adaptation (scenario 1)
Profiles and Levels: not there yet!

- Profiles
  - Restrict which tools to be used
- Levels
  - Restrict on implementation issues (memory usage, resolution, ...)
- Profile and Level discussion started in previous meeting (October 2006)
  - Expected results at next meeting
- New issues introduced by scalability
  - Traditional conformance point (profile@level) not sufficient, need to explain how scalability layers interact
  - Base and enhancement layers may belong to different profiles, how to express that to limit decoder complexity?
Chapter 3 - MPEG-4 Multi-view Video Coding

MPEG-4 Multi-view Video Coding - Outline

- History, context, and motivations
- Typical multi-view system setups and application scenarios
- Standardization and requirements on MVC
- Basic concepts and principles of MVC
- Coding tools currently considered
MPEG-4 Multi-view Video Coding - History, Context, and Motivations

2D color TV

3D color TV

2D → 3D

Passive single-view video

Interactive multi-view video

1-view → N-view

3D Video/Stereo Video [JVT-T100]

• Generating 3D depth impression from separate views for each eye
• Not necessarily limited to one user: multiple users can watch it simultaneously from a wide range of positions, with multiple inputs/outputs
Free Viewpoint Video (FVV) [JVT-T100]

- Aiming at free navigation and interactivity
- Some challenges to tackle:
  - Multiple video streams need to be synchronized
  - Huge amount of data needs to be compressed and transmitted
  - Inter-viewpoint redundancy to be exploited (effective only for dense camera setups, not for dome-type arrangements)

Typical Multi-view Camera Configuration

- Parallel view
- Convergent view
- Divergent view

Stanford Multi-Camera Array with 128 synchronized CMOS cameras [Levoy et al., SIGGRAPH'96]

8 high resolution PtGrey cameras along a 1D arc [Zitnick et al., SIGGRAPH'04]

Dodeca 1000 camera – 11 video cameras in a dodecahedral array
Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding

Tutorial presented at ICCE 2007 © imec 2006

Interactive Multi-view Realization: Free Viewpoint Video (FVV)

- A classical trade-off between costs (e.g. camera#, processors) and benefits (e.g. navigation range, virtual views quality)
- Depth (or disparity) map to be computed using computer vision techniques

High-quality Video Interpolation using a Layered Representation [Zitnick, SIGGRAPH’04]
High-quality Video Interpolation using a Layered Representation: Set-up and Results

Virtual Viewpoint Video
Capture multiple synchronized video streams

balletB-half.avi  day1-lazy0-half.avi

Application Scenarios [N5877]

Entertainment e.g. concert, sport, movie, game...
Education e.g. instruction video, cultural archives
Medical surgery
Surveillance
Event broadcasting
Immersive video conference
Advertisements
FVV system
MVC
**FVV System and FVV Decoder [N8064]**

- Basic components of an example FVV system

- Example architecture of a FVV decoder

**Standardization Progress [JVT-T100]**

MVC - MPEG-4 Part 10, Amd.4

- Evaluation of call for proposals
- Call for proposals on MVC (along with requirements)
- Call for comments on 3DAV
- Exploration reports

- MPEG 3DAV AHG established
- JVT activity & WD
- FDIS
- PDAM
- FPDA
### Requirements for MVC (1/2) [N8064]

#### Compression related requirements

<table>
<thead>
<tr>
<th>Features</th>
<th>Shall</th>
<th>Should</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compression efficiency</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>View scalability</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Free viewpoint scalability</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Spatial/temporal/SNR scalability</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Resource consumption</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Low delay</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Robustness (or error resilience)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Resolution, bit depth, chroma sampling format</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Picture quality among views</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

### Requirements for MVC (2/2)

<table>
<thead>
<tr>
<th>Features</th>
<th>Shall</th>
<th>Should</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal random access</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>View random access</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Spatial random access</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Resource management for decoders</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Parallel processing</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

#### System support related requirements

<table>
<thead>
<tr>
<th>Features</th>
<th>Shall</th>
<th>Should</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synchronization</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>View generation</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Non-planar imaging and display systems</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Camera parameters</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Synchronized Multi-view Video Streams

Inter-viewpoint Correlation and Prediction
[Merkle, ICME’06], [m12945]

- T (temporal)
- S (inter-view)
- L/R (combined)

- Normally, T better than S better than L/R

- Influencing factors:
  - Frame rate
  - Inter-camera distance (baseline)
  - Content complexity (motion intensity, illumination effects, etc)
Promising Coding Tools Currently Considered for MVC Standard [m13394]

- Hierarchical B pictures for temporal dependencies and an adapted prediction scheme (HHI proposal)
- MVC encoder design recommendations
- Block level illumination compensation (5 competing proposals)
- View synthesis prediction (2 different proposals)
Hierarchical B Pictures for MVC: Overview
[Merkle, ICME’06], [m12945], [W8019], [JVT-P014]

- Uses hierarchical B-pictures combined in temporal and interview dimension, proposed by FhG-HHI
- Proved best performance out of 8 proposals in response to a CfP in formal subjective tests
- Current MVC reference model
- Fully compatible to H.264/MPEG4-AVC

- Reorganization of input images into a single stream prior to encoding
- Decoder needs to de-interleave the decoded pictures into MV video streams

Hierarchical B Pictures for MVC: Coding Structure (GOP size 8)

[Smolić, JVT-T100, July 2006]
Hierarchical B Pictures for MVC: Coding Results

- Inter-view prediction: hierarchical B pictures for temporal and inter-view dimensions
- Simulcast: hierarchical B pictures for temporal dimension only
- Anchor: H.264/AVC simulcast coding without hierarchical B pictures

Hierarchical B Pictures for MVC: Summary

- Hierarchical B pictures (accounts for about 50% coding gain)
- Sophisticated reference frame selection (coding structure)
- Decoding Picture Buffer Size increases to $2 \times GOP\_length + \text{views} #$
- Coding efficiency vs low-delay decoding: the coding efficiency drops significantly as $GOP\_length$ decreases.
- In general, significant coding gain over simulcast anchor and simulcast hierarchical B picture coding, when $GOP\_length$ is set to 12 and 15.
Promising Coding Tools Currently Considered for MVC Standard [m13394]

- Hierarchical B pictures for temporal dependencies and an adapted prediction scheme (HHI proposal)
- **MVC encoder design recommendations**
  - Block level illumination compensation (5 competing proposals)
  - View synthesis prediction

MVC encoder design recommendations

- **Low delay to enable fast temporal random access**
  - Ex: prediction structure corresponding to GOP with 1 frame ("Short")
- **Simplified prediction structures to reduce the number of reference candidates**
  - Ex: inter-view prediction is only applied to anchor frames
- **Alternative prediction structures (e.g. placing reference I in middle) to improve coding efficiency**
Promising Coding Tools Currently Considered for MVC Standard [m13394]

- Hierarchical B pictures for temporal dependencies and an adapted prediction scheme (HHI proposal)
- MVC encoder design recommendations
- Block level illumination compensation (5 competing proposals)
- View synthesis prediction

Block Level Illumination Compensation: Illumination/Color Mismatch Problem [JVT-T100: CE2]

- Imperfectly calibrated cameras
- Different perspective projection direction
- Different reflection effects
Block Level Illumination Compensation [JVT-T100: CE2]:
Current status

- Weighted prediction (WP) existing in H.264/AVC for slice level illumination compensation
- 5 competing proposals aiming at compensate the illumination differences using various local techniques (see details in [JVT-T110], [JVT-T113], [JVT-T114], [JVT-T115], [JVT-T117])
  - Some proposals use both scale and offset; others use offset only
  - Three proposals employ predictive coding of DC residue; one proposal combines slice-level WP with block-level delta-WP; the other proposal uses 2D (W,O) vector quantization
  - Some proposals support only 16x16; others include smaller block sizes for IC
  - All proposals introduce new syntaxes
  - Encoder aspects: all proposal involve illumination invariant matching cost calculation in motion/disparity search; in general IC requires extra cost in mode decision
- Comparison against reference w/o WP 0.1 to 0.6 dB PSNR gain
- As of July ’06, no convergence has been reached

Promising Coding Tools Currently Considered for MVC Standard [m13394]

- Hierarchical B pictures for temporal dependencies and an adapted prediction scheme (HMI proposal)
- MVC encoder design recommendations
- Block level illumination compensation (5 competing proposals)
- **View synthesis prediction**
View Synthesis Prediction

- Require camera parameters and multi-view geometry
- Exploit camera parameters and multi-view geometry (i.e. corresponding points across different views lie on the epipolar line) to generate intermediate views which may be a more accurate reference candidate
- Ex in picture: MERL [JVT-T123] approach

Conclusions
Conclusions

- MPEG and ITU-T Standardization Bodies
- MPEG-* and H.26*
- The H.264/MPEG-4 video standards
- H.264/MPEG-4 Part 10 AVC
- SVC extension of H.264/MPEG-4 AVC (FDIS Jan ’07)
- MVC extension of H.264/MPEG-4 AVC (FDIS Jan ’08)

Annex
Annex - Outline

- What else is going on as far as video coding standards (MPEG, ITU-T, SMPTE)
- Bibliography and reference points
- Acronyms and abbreviations

Other MPEG Video Activities (1/2)

- ISO/IEC 23002-1 and 23002-2 MPEG Visual Technologies
  - Accuracy specification for implementation of integer-output IDCT (Part 1) to replace IEEE 1180 (same text), issue CORs
  - CIP for fixed-point implementation (i.e. no drifting) of DCT/IDCT passing IEEE 1180 and causing no problem with existing codec and streams
  - 9 proposals on October 2005
  - interesting dynamic between MPEG-2 "old timers" and the proponent of this activity (ex: how much precision is needed?, do we really need this work?)
  - SW test-bed and test plan to evaluate proposals
  - WD 2.0 of ISO/IEC 23002-2 Integer DCT Transforms

- ISO/IEC 23002-3 MPEG Visual Technologies: Auxiliary Video Data Representation
  - stereoscopic applications (ex: stereo video display)
  - encode depth/disparity map re-using existing SoA (low hanging fruit)
  - ISO/IEC 23002-3 timeline:
    - WD on Jan 2006
    - CD on April 2006
    - FDIS on January 2007
  - current status: FCD
Other MPEG Video Activities (2/2)

- **Reconfigurable Video Coding (RVC)**
  - formally known as VCTR
  - toolbox of standard video coding tools with standard interfaces (FU)
  - mechanism to describe decoders in machine readable format
  - a specific decoding solution is implemented as a set of FUs connected to a global control unit
  - benefits:
    - reconfigurable decoders → comply with multiple standards at a minimal implementation cost
    - componentization of codec design and reusability
    - higher level of flexibility in interoperability
    - fast standardization of new coding tools → new standard codecs
  - status: requirements definition and draft of CfP
  - time line:
    - CfP April 2006,
    - proposals by July 2006
    - FDIS July 2007
    - current status: WD 1.0

Other JVT Activities

- **Advanced 4:4:4 Profiles**
  - replace High 4:4:4 (FRExt H444P)
  - 14496-10:2006 Amd. 2 Advanced 4:4:4 Profiles
  - Two 4:4:4 profiles (one for Intra only and one for Intra+Inter): tentative names Hi444Intra (only containing I slices) and Hi444Predictive (further input requested at the next meeting)
  - Current status : FPDAM2

- **SVC File Format (FF)**
Video Standards on the Horizon

- MPEG Video: exploration activity for technologies to increased video compression efficiency (workshops)

- ITU-T VCEG: H.265 on hold to send a clear message

- Society of Motion Picture and Television Engineers (SMPTE)
  - Tech. Committee on Television Video Compression (C24), AhG on Video Codec (VC) - 9 Compressed Video Bitstream Format and Decoding Process
  - MS push to standardize WMV-9 (failed in MPEG) in preparation for HD-DVD battle
  - Simple and Main profiles: "rubber stamp", as forced to be WMV-9 backward compatible
  - Advanced profile: more open standardization approach, as innovation is required to support interlaced content
  - Current status: completed
    - VC-1 specification – SMPTE 421M
    - VC-1 conformance – SMPTE RP228
    - VC-1 transport – SMPTE RP227

Some Interesting Papers Relevant to MVC

- 3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes [Matusik, SIGGRAPH'04]
- Disparity compensated prediction [Zitnick, SIGGRAPH’04]
- Video transcoding for special visual effects in MVC, e.g. frozen moment, view switching [Lou, MM’05]
- 3D voxel model based MVC [Gao, ICIP’05]
- MVC based on global motion model [Guo, PCM’04]
- A framework for MVC using layered depth images [Yoon, PCM’05]
- Ensuring color consistency across multiple cameras [Ilie, ICCV’05]
- Dependent bit allocation for MVC [Kim, ICIP’05]
- Predictive fast motion/disparity search for MVC [Lai, VCIP’06]
- Epipolar geometry assisted motion estimation for MVC [Lu, ICIP’06]
- Multicast of real-time MVC [Zuo, ICME’06]
- Interactive transport of MVC for 3DTV [Kurutepe, PV’06]
### MPEG and JVT Documents (1/2)


### MPEG and JVT Documents (2/2)

Recent Developments in Video Compression Standards: from Scalable to Multi-view Video Coding
Tutorial presented at ICCE 2007
© imec 2006

References (1/4)


References (2/4)

References (3/4)


Online resources:
- MPEG site: http://mpeg.nist.gov/welcome.php
- JVT site: http://ftp3.itu.ch/av-arch/jvt-site/
- MPEG-4 working documents: http://www.chiariglione.org/mpeg/working_documents.htm#MPEG-4
- HHI SVC page: http://ip.hhi.de/imagecom_G1/savce/

References (4/4)

Acronyms and Abbreviations (1/3)

- SVC: Scalable Video Coding (Amd3 to MPEG-4 Part 10)
- AVC: Advanced Video Coding (MPEG-4 Part 10)
- MPEG: Moving Picture Expert Group
- ITU-T: International Telecommunication Union - Telecommunication (formally CCITT)
- VCEG: Video Coding Expert Group
- JVT: Joint Video Team
- SMPTE: Society of Motion Picture and Television Engineers
- AIG: Ad hoc-Group
- CIP: Call for Proposal
- SoA: State-of-the-Art
- CA-VLC: Context Adaptive Variable Length Coding
- FReX: Fidelity Range Extension
- CoR: Corrigendum
- Amd: Amendment
- WD: Working Draft
- CD: Committee draft
- FDIS: Final Draft International Standard
- IS: International Standard
- PDAM: Preliminary Draft Amendment
- FPDAM: Final Preliminary Draft Amendment
- FDAM: Final Draft Amendment
- CE: Core Experiment
- DMB: Digital Audio Broadcasting
- T/S-DMB: Terrestrial/Satellite Digital Multimedia Broadcasting
- ISMA: Internet Streaming Media Alliance

Acronyms and Abbreviations (2/3)

- MPEG-21 DIA: MPEG-21 Digital Item Adaptation
- DVB: Digital Video Broadcasting
- DVB-T: Digital Video Broadcasting Terrestrial (terrestrial television)
- DVB-H: Digital Video Broadcasting Handheld (terrestrial television for handhelds)
- 3GPP: 3rd Generation Partnership Project
- 3GPP2: 3rd Generation Partnership Project 2
- ARIB: Association of Radio Industries and Businesses (standardization organization in Japan)
- ISDB-T: Integrated Services Digital Broadcasting-Terrestrial
- IPR: Intellectual Property Right
- VRC: Reconfigurable Video Coding (formally VCTR)
- VCTR: Video Coding Tools Repository
- MVC: Multi-View Coding
- FF: File Format
- WHV-0: Windows Media Version 0
- AAP: Alternative Approval Process (ITU)
- ISO: International Standards Organization
- IEA: International Electrotechnical Commission
- DSL/ADSL: Digital Subscriber Line/Asynchronous Digital Subscriber Line
- DCT/IDCT: Discrete Cosine Transform/Inverse Discrete Cosine Transform
- Vidwav: Wavelet Video Coding
- PSP: PlayStation Portable (Sony)
- MMS: Multimedia Messaging (3GPP)
- PSS: Packet Switched Stream (3GPP)
- PSC: Primary Synchronization Code (3GPP)
Acronyms and Abbreviations (3/3)

- MVC: Multi-view Video Coding (in the Annex of the full-set slides already)
- FVV: Free Viewpoint Video
- 3DAV: 3D Audio and Video
- GOP: Group of Pictures
- DPB: Decoded Picture Buffer
- RPL: Reference Picture Lists
- WP: Weighted Prediction
- VSP: View Synthesis Prediction
- DCVP: Disparity Compensated View Prediction
- VPS: View Parameter Set