Investigation of Scalable HEVC and its Bitrate Allocation for UHD Deployment in the context of HTTP Streaming

Uma Sagar Madhugiri Dayananda
Department of Electrical Engineering
University of Texas, Arlington
Email: uma.sagar@mavs.uta.edu
Advisor: Dr. K.R. Rao
Outline

• Motivation
• Scalable High Efficiency Video Coding
• Investigation of SHVC for UHD deployment in the context of HTTP streaming
  • Evaluation Methodology
  • Summary of Experiments
• Evaluation Results
• SHVC Rate Allocation
• Conclusions
• Further Work
Motivation

Growing Over the top (OTT) market

4K devices and video content

State of the art video compression and streaming technologies

New Heterogeneous Client Distribution supporting UHD and HD content
# High Definition (HD) Vs Ultra-High Definition (UHD)

<table>
<thead>
<tr>
<th>Attributes</th>
<th>High Definition</th>
<th>Ultra-High Definition [17]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial</td>
<td>Resolution 1920x1080</td>
<td>3840x2160 (UHD-1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7680x4320 (UHD-2)</td>
</tr>
<tr>
<td>Temporal</td>
<td>Frame rate (fps) 60,50,30,25,24</td>
<td>120, 60, 50, 30, 25, 24</td>
</tr>
<tr>
<td></td>
<td>Scan Progressive, Interlaced</td>
<td>Progressive</td>
</tr>
<tr>
<td>Primary colors</td>
<td>Red primary (0.640, 0.300)</td>
<td>(0.708, 0.292)</td>
</tr>
<tr>
<td></td>
<td>Green Primary (0.150, 0.330)</td>
<td>(0.170, 0.797)</td>
</tr>
<tr>
<td></td>
<td>Blue Primary (0.600, 0.060)</td>
<td>(0.131, 0.046)</td>
</tr>
<tr>
<td>Coding format</td>
<td>Bit-depth 8 and 10 bit</td>
<td>10 and 12 bit</td>
</tr>
</tbody>
</table>
High Efficiency Video Coding (HEVC)

- State of the art, hybrid block based video coding technology, standardized in January 2013 [1]
- Provides bitrate reduction of around 50% when compared to previous MPEG-4 AVC standard at similar quality
- Supports resolutions up to 4K and 8k, bit depths of 8, 10, 12 and 16 bits per pixel
- In-built architecture support for parallel processing
- Compression obtained by exploiting redundancies in video sequences [9]:
  - Spatial redundancy removal - Intra prediction schemes, Block transforms based on Discrete Cosine Transform (DCT), Discrete Sine Transform (DST)
  - Temporal redundancy removal - Inter Prediction
  - Perceptual redundancy removal - Quantization
  - Statistical redundancy removal - Entropy coding using Context Adaptive Binary Arithmetic Coding (CABAC)
HTTP Streaming

• HTTP based adaptive bitrate streaming is a technique used for multimedia streaming over large distributed HTTP networks

• Works by detecting user's bandwidth and CPU capacity in real time and adjusting quality of a video stream accordingly

• Dynamic Adaptive Streaming over HTTP (DASH), also known as MPEG-DASH \([11]\) is an International standard developed by MPEG
Traditional Video Streaming

- Multiple versions of video are encoded and stored separately for all clients and connection types
Scalable video coding saves storage costs

- Basic version of the video is coded as base layer (BL) and improved versions of the video coded as enhancement layers (EL)
Why this Investigation?

• How can scalable video coding with SHVC be used for delivering UHD or 4K content for given heterogeneous client distribution?

• How much bitrate savings can be obtained using SHVC? Is there a overhead involved?

• How to allocate bit rate optimally for each layer based on client distribution?
SHVC – Scalable High Efficiency Video Coding

• Scalable Extension of HEVC, standardized in 2014 [2]
• Re-uses multiple single-layered HEVC cores to achieve high scalable coding efficiency
• Requires high level syntax (HLS) changes in HEVC at slice header level and above [12]
• Uses interlayer prediction to exploit redundancies between layers to gain coding efficiency by predicting enhancement layer using base layer
• Utilizes multi-loop coding framework, where base layer (BL) is decoded first, followed by decoding enhancement layer (EL) using interlayer prediction
SHVC – Scalability Options

- Additionally supports Bit-depth, Color Gamut, Interlaced to progressive and Hybrid Codec scalabilities
The original input picture of UHD resolution is down-sampled and encoded into base layer (BL) bit-stream.

For spatial enhancement layer (EL) picture coding, already coded data in the lower reference layer(s) are used for inter-layer prediction to improve EL coding efficiency [12].

If BL is encoded with HEVC, then HEVC decoder is used; similarly if BL is encoded with MPEG-4 AVC, then MPEG-4 AVC decoder is used.
SHVC – Encoder and Decoder Architecture with 3 Layers
Individual encoded bitstreams for each layer can be fragmented separately for HTTP streaming and decoded at the client side. When BL is encoded with HEVC, then HEVC decoder is used; similarly when BL is encoded with MPEG-4 AVC, then MPEG-4 AVC decoder is used.
Interlayer processing is applied to reference layer (RL) pictures to form Interlayer reference (ILR) pictures, when any parameters, including spatial resolution, bit depth and color gamut are different between RL and current EL [4].

Interlayer processing in SHVC includes three modules:

- Texture resampling
- Motion field resampling
- Color mapping
SHVC Investigation – Hybrid Codec

- UHD video and its services are increasing in multimedia ecosystem, but older devices are still supporting HD video. Also, few clients support early codecs such as MPEG-4 AVC and other clients support newer HEVC codecs.

- To provide backward compatibility and deliver UHD to a mix of this HD and UHD clients, the hybrid codec feature of SHVC is evaluated, where base layer can be HEVC or Non-HEVC (such as MPEG-4 AVC) [12].
**SHVC Investigation - Evaluation Methodology (1)**

- **Simulcast coding** – different versions of the video are coded independently and stored

<table>
<thead>
<tr>
<th>Simulcast</th>
<th>Low Resolution(HD)</th>
<th>SHVC</th>
<th>High Resolution(UHD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( B_{LR} = 6 \text{Mbps} )</td>
<td>( B_{BL} = 6 \text{Mbps} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( B_{HR} = 12 \text{Mbps} )</td>
<td>( B_{EL} = 6 \text{Mbps} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total Bitrate=18Mbps</td>
<td>Total Bitrate=12Mbps</td>
<td></td>
</tr>
</tbody>
</table>

- **Single Layer Coding**

<table>
<thead>
<tr>
<th>Single Layer Coding</th>
<th>High Resolution(UHD)</th>
<th>SHVC</th>
<th>Total Bitrate=12Mbps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( B_{HR} = 12 \text{Mbps} )</td>
<td>( B_{BL} = 6 \text{Mbps} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total Bitrate=12Mbps</td>
<td>( B_{EL} = 6 \text{Mbps} )</td>
<td></td>
</tr>
</tbody>
</table>

- **This is not a fair comparison as the bit rate savings are obtained at different quality levels**

- **R-D Plot for comparison of SHVC with Simulcast and Single Layer Coding**
SHVC investigation - Evaluation Methodology (2)

• SHVC encoding results in bitrate savings; however, it comes at a cost of scalability overhead. This overhead is mainly due to reduced compression efficiency resulting from separation of BL and EL, and additional protocol headers [14].

• In order to have a fair comparison, the bitrate differences between SHVC and simulcast coding are compared at same quality as represented below.

Bitrate comparison at Equal Quality
**Evaluation Metrics**

- **Bjontegaard Delta (BD) Metrics** [32,33]
  - Used to evaluate the average difference between two Rate-Distortion curves
    - **BD-PSNR**
      Compute average gain in PSNR between two rate-distortion curves.
    - **BD-Bitrate**
      Compute average per cent saving in bitrate between two rate-distortion curves.

- **Bitrate savings (BD-Bitrate)** obtained by comparing SHVC with Simulcast coding
- **Scalability overhead (BD-PSNR)** obtained by comparing SHVC with Single Layer Coding
Summary of Experiments

SHVC Encoding

HEVC BL

2 Layers, BL:EL = 50:50
HD BL + UHD EL
Spatial Scalability

3 Layers, BL:EL1:EL2 = 50:25:25
HD BL + UHD EL1 + Quality enhanced UHD EL2
Combined Scalability

MPEG-4 AVC BL

2 Layers, BL:EL = 50:50
HD BL + UHD EL
Spatial Scalability

3 Layers, BL:EL1:EL2 = 50:25:25
HD BL + UHD EL1 + Quality enhanced UHD EL2
Combined Scalability

Varying BL:EL from 10:90 to 90:10 in steps of 10

Fixed BL bitrate, varying total bitrate

Varying BL:EL from 10:90 to 90:10 in steps of 10
### Test Sequences

<table>
<thead>
<tr>
<th>Test Sequence</th>
<th>Frame rate (FPS)</th>
<th>HD Resolution</th>
<th>UHD Resolution</th>
<th>Total frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>PeopleOnStreet</td>
<td>30</td>
<td>1920x1080</td>
<td>3840x2160</td>
<td>150</td>
</tr>
<tr>
<td>Sintel39</td>
<td>24</td>
<td>1920x872</td>
<td>3840x1744</td>
<td>344</td>
</tr>
<tr>
<td>ParkJoy</td>
<td>50</td>
<td>1920x1080</td>
<td>3840x2160</td>
<td>500</td>
</tr>
</tbody>
</table>
Experimental Setup

- Rate control involves modifying the encoding parameters in order to maintain a target output bitrate. In order to encode the video at a given bitrate, rate control feature present in the reference software is used.

- Numerous experiments are involved in SHVC investigation. Hence, encoding process is automated using Bash Scripts. Extraction of results, plotting and calculating BD-Metrics is performed using Python scripts.

- The experiments are performed on Linux cluster.

- Reference softwares used in the experiments and encoding parameters are given below:

<table>
<thead>
<tr>
<th>Video Coding Standard</th>
<th>Reference Software</th>
<th>Spatial Scalability</th>
<th>Bit-depth</th>
<th>Chroma Sampling</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHVC</td>
<td>SHM 6.1 [35]</td>
<td>2x with HD and UHD resolutions</td>
<td>8 bits per pixel</td>
<td>4:2:0 Format</td>
<td>Random Access Main Profile [31]</td>
</tr>
<tr>
<td>HEVC/H.265</td>
<td>HM 15.0 [36]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVC/H.264</td>
<td>JM 18.6 [37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results for PeopleOnStreet Test Sequence, 2 Layers

<table>
<thead>
<tr>
<th>Base Layer</th>
<th>BD-Bitrate (%)</th>
<th>BD-PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEVC</td>
<td>32.18</td>
<td>-0.76</td>
</tr>
<tr>
<td>MPEG-4 AVC</td>
<td>23.41</td>
<td>-1.3</td>
</tr>
</tbody>
</table>

2 Layers, BL:EL – 50:50
Summary of Results for 2 Layers

For UHD deployment with SHVC - 2 spatial layers
- HEVC as base layer gave 27% -32% bitrate savings, average scalability overhead of 0.7 dB
- MPEG-4 AVC as base layer gave in 11% -23% bitrate savings, average scalability overhead of 1.27 dB
Summary of Results for 3 Layers

- For UHD deployment with SHVC - 3 layers of combined spatial and quality scalability
  - HEVC as base layer gave 44%-49% bitrate savings, average scalability overhead of 1.02 dB
  - MPEG-4 AVC as base layer gave 32%-42% bitrate savings, average scalability overhead of 1.72 dB
Results - Varying Bitrate Allocation, PeopleOnStreet (1)

Bitrate Savings of SHVC when compared to Simulcast Coding for 2 Layers

Bitrate savings depend on the ratio of allocation of bit rates into layers.
Scalability overhead also depends on the ratio of bit rate allocation into various layers.
Increasing the bits allocated for the base layer does not necessarily increase the quality of resulting video. The quality is impacted by the resolutions of the layers as well.
Decoded video frame of PeopleOnStreet Sequence

Original UHD Test Sequence
HEVC UHD @ 20Mbps, Average PSNR = 36.89 dB
SHVC UHD (HEVC BL @10Mbps+HEVC EL@10Mbps)
Average PSNR = 36.10 dB
SHVC UHD (MPEG-4 AVC BL @ 10Mbps+HEVC EL@10Mbps)
Average PSNR = 35.66 dB
A bitrate allocation problem in the context of multiple layers or versions is formulated and a solution based on dynamic programming is presented in [25].

This rate allocation problem is adapted for the scenario of multilayer streaming with SHVC by setting the number of layers to 2.

Total number of clients in the context of HTTP streaming can be divided onto C classes based on bandwidth and each of this class has probability mass distribution given by \( f(c) \). The SHVC bit stream can be structured into 2 layers. The goal is to find an optimum structuring policy \( P^* = \{r_i, i=1, 2\} \), where \( r_i \) is the bitrate of each layer that yields maximum system wide utility \( U^*_o \) over the entire client class. This can be written as:

\[
U^*_o = \max_{P^*} U_0 = \sum_{k=1}^{C} f(k) \cdot u(b_k, b_k) \quad \text{such that} \quad r_1 < r_2
\]
The client utility function assumes that higher the effective rate that a client receives, the more satisfied that client will be, this is given by

\[ u_{rate}(b_c, b_c) = b_c \]

- \( b_c \) is the effective rate of the received stream
- \( b_c \) is the available bandwidth of the client class \( c \).

For a scalable bit-stream having CGS layers, the effective rate of layer \( l \) (\( 1 \leq l \leq L \)) is defined as:

\[ r_l = \begin{cases} r_l, & l = 1 \\ r_{l-1} + \frac{(r_l - r_{l-1})}{1 + a(r_l)}, & 2 \leq l \leq L \end{cases} \]

\( a(r_l) \) is the scalability overhead function and it depends on the characteristics of the video sequence, granularity of scalable coding such as Coarse-Grain Scalability (CGS), and rate of the layer being encoded as well as the rates of the previous layers.

\[ b_c = r_l, \text{ when layers are CGS} \]
The SHVC bitrate allocation problem is evaluated for a sample client bandwidth distribution given below:

<table>
<thead>
<tr>
<th>Client Class</th>
<th>Bandwidth (Mbps)</th>
<th>Number of Clients</th>
<th>Probability Mass Function f(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>8</td>
<td>60</td>
<td>0.6</td>
</tr>
<tr>
<td>C2</td>
<td>12</td>
<td>10</td>
<td>0.1</td>
</tr>
<tr>
<td>C3</td>
<td>16</td>
<td>20</td>
<td>0.2</td>
</tr>
<tr>
<td>C4</td>
<td>20</td>
<td>10</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Evaluation for 2 layers resulted in: \( r_1 = 8 \text{ Mbps}, r_2 = 16 \text{ Mbps} \)
Conclusions

- Investigation of SHVC for UHD deployment in the context of HTTP streaming is performed by comparing it with Simulcast and Single Layer Coding.

- Bitrate savings and scalability overhead for the scenario of 2 and 3 layers of SHVC were determined. It can be concluded from these results, that UHD video can be efficiently deployed with scalable video coding SHVC.

- As the number of SHVC layers are increased, better bitrate savings are obtained, however, scalability overhead also increases.

- Previous investigation in Scalable Video Coding suggested that adding more bits into base layer reduces the scalability overhead [27]. However, the preliminary investigation conducted suggests that SHVC scalability overhead depends on the ratio of bit allocation in base and enhancement layers. Also, scalability overhead for spatial and quality enhancement layers needs to be computed separately.

- An existing bitrate allocation problem is adapted and evaluated in the context of 2 layered SHVC for a sample client distribution and optimal bitrates for base and enhancement layers are obtained.

- The research work is published as an IEEE conference paper [14].
Further Work

• Exploring bitrate allocation algorithms for optimal allocation of bits into layers of SHVC based on Game theory and other approaches can be done.

• Additional experiments to study the effect of scalability overhead and its modeling for different types of scalabilities (spatial, quality etc.,) on different test sequences can be performed.

• Detailed evaluation of computational complexity of SHVC for various types of scalabilities can be done.

• Application of parallel processing techniques such as algorithm level parallelization to high level parallelization for SHVC encoding and decoding of base and enhancement layers can be explored.
References (1)


References (4)


Acronyms and Abbreviations

AVC – Advanced Video Coding
BL – Base Layer
CABAC – Context Adaptive Binary Arithmetic Coding
CDN – Content Delivery Network
CGS – Coarse Grain Scalability
DASH – Dynamic Adaptive Streaming over HTTP
DCT – Discrete Cosine Transform
DST – Discrete Sine Transform
EL – Enhancement Layer
fps – Frame per second
HD – High Definition
HEVC – High Efficiency Video Coding
HLS – High Level Syntax
HTTP – Hyper Text Transfer Protocol
JCTVC – Joint Collaborative Team on Video Coding
Mbps – Megabits per second
MPEG – Moving Picture Experts Group
OTT – Over the Top
PSNR – Peak Signal to Noise Ratio
QP – Quantization Parameter
RD – Rate Distortion
RL – Reference Layer
SHVC – Scalable High Efficiency Video Coding
SNR – Signal to Noise Ratio
UHD – Ultra High Definition
Acknowledgements

• First and foremost, I take this opportunity to extend gratitude to my thesis advisor, Dr. K.R Rao, for introducing me to this fascinating area of research in video coding, his guidance, support, help rendered and continuous encouragement throughout the research.

• I am thankful to Adobe Systems Inc. for giving me an internship opportunity to carry out the research. I thank my mentor Dr. Vishy Swaminathan and the Video Systems Team at Adobe Research for guiding and supporting me throughout the work.

• I extend my thanks to Dr. Michael T. Manry and Dr. William E. Dillon for serving on the thesis committee. I thank my fellow members of Multimedia Processing Lab at UTA for their help and support during the stay. I also thank University of Texas at Arlington for providing necessary resources to carry out the research.

• Finally, I thank my parents, sisters, family members and friends, who always have been my backbone, giving unconditional support and encouragement throughout the journey.
Thank You!