Video Codec Requirements and Evaluation Methodology

draft-filippov-netvc-requirements-02

Alexey Filippov,
Jose Alvarez (Huawei Technologies)
Contents

• An overview of applications
• Requirements
• Evaluation methodology
• Conclusions
Applications

- Internet Protocol Television (IPTV)
- Video conferencing
- Video sharing
- Screencasting
- Game streaming
- Video monitoring / surveillance
Internet Protocol Television (IPTV) / IP-based over-the-top (OTT) video

- **Basic requirements:**
  - Random access to pictures
    - Random Access Period (RAP) should be kept small enough (approximately, 1-15 seconds);
  - Temporal (frame-rate) scalability;
  - Error robustness (for delay-critical OTT video transmission)

- **Optional requirements:**
  - resolution and quality (SNR) scalability
Internet Protocol Television (IPTV)

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Frame-rate, fps</th>
<th>Picture access mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>2160p (4K), 3840x2160</td>
<td>60</td>
<td>RA</td>
</tr>
<tr>
<td>1080p, 1920x1080</td>
<td>24, 50, 60</td>
<td>RA</td>
</tr>
<tr>
<td>1080i, 1920x1080 *</td>
<td>30 (60 fields per second)</td>
<td>RA</td>
</tr>
<tr>
<td>720p, 1280x720</td>
<td>50, 60</td>
<td>RA</td>
</tr>
<tr>
<td>576p (EDTV), 720x576</td>
<td>25, 50</td>
<td>RA</td>
</tr>
<tr>
<td>576i (SDTV), 720x576 *</td>
<td>25, 30</td>
<td>RA</td>
</tr>
<tr>
<td>480p (EDTV), 720x480</td>
<td>50, 60</td>
<td>RA</td>
</tr>
<tr>
<td>480i (SDTV), 720x480 *</td>
<td>25, 30</td>
<td>RA</td>
</tr>
</tbody>
</table>

**NB**: interlaced content can be handled at the higher system level and not necessarily by using specialized video coding tools. It is included in this table only for the sake of completeness as most video content today is in progressive format.
Video conferencing

• **Basic requirements:**
  - Delay should be kept as low as possible
    - The preferable and maximum delay values should be less than 100 ms and 350 ms, respectively
  - Temporal (frame-rate) scalability;
  - Error robustness

• **Optional requirements:**
  - resolution and quality (SNR) scalability
## Video conferencing

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Frame-rate, fps</th>
<th>Picture access mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1080p, 1920x1080</td>
<td>15, 30</td>
<td>JFPIX</td>
</tr>
<tr>
<td>720p, 1280x720</td>
<td>30, 60</td>
<td>JFPIX</td>
</tr>
<tr>
<td>4CIF, 704x576</td>
<td>30, 60</td>
<td>JFPIX</td>
</tr>
<tr>
<td>4SIF, 704x480</td>
<td>30, 60</td>
<td>JFPIX</td>
</tr>
<tr>
<td>VGA, 640x480</td>
<td>30, 60</td>
<td>JFPIX</td>
</tr>
<tr>
<td>360p, 640x360</td>
<td>30, 60</td>
<td>JFPIX</td>
</tr>
</tbody>
</table>
Video sharing

• **Basic requirements:**
  - Random access to pictures for downloaded video data
  - Temporal (frame-rate) scalability
  - Resolution and quality (SNR) scalability
  - Error robustness

• **Typical scenarios:**
  - GoPro camera
  - Cameras integrated into smartphones
### Video sharing*

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Frame-rate, fps</th>
<th>Picture access mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>2160p (4K), 3840x2160</td>
<td>24, 25, 30, 48, 50, 60</td>
<td>RA</td>
</tr>
<tr>
<td>1440p (2K), 2560x1440</td>
<td>24, 25, 30, 48, 50, 60</td>
<td>RA</td>
</tr>
<tr>
<td>1080p, 1920x1080</td>
<td>24, 25, 30, 48, 50, 60</td>
<td>RA</td>
</tr>
<tr>
<td>720p, 1280x720</td>
<td>24, 25, 30, 48, 50, 60</td>
<td>RA</td>
</tr>
<tr>
<td>480p, 854x480</td>
<td>24, 25, 30, 48, 50, 60</td>
<td>RA</td>
</tr>
<tr>
<td>360p, 640x360</td>
<td>24, 25, 30, 48, 50, 60</td>
<td>RA</td>
</tr>
</tbody>
</table>

* - **Sources of these data:**
  - "Recommended upload encoding settings (Advanced)"
    [https://support.google.com/youtube/answer/1722171?hl=en](https://support.google.com/youtube/answer/1722171?hl=en)
Screencasting

• **Basic requirements:**
  - Support of a wide range of input video formats
    - RGB and YUV 4:4:4 in addition to YUV 4:2:0 and YUV 4:2:2
  - High visual quality
    - up to visually and mathematically lossless

• **Optional requirements:**
  - Error robustness
## Screencasting

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Frame-rate, fps</th>
<th>Picture access mode</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input color format: RBG</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WQXGA, 2560x1600</td>
<td>15, 30, 60</td>
<td>AI, RA, JFPIC</td>
</tr>
<tr>
<td>WUXGA, 1920x1200</td>
<td>15, 30, 60</td>
<td>AI, RA, JFPIC</td>
</tr>
<tr>
<td>WSXGA+, 1680x1050</td>
<td>15, 30, 60</td>
<td>AI, RA, JFPIC</td>
</tr>
<tr>
<td>WXGA, 1280x800</td>
<td>15, 30, 60</td>
<td>AI, RA, JFPIC</td>
</tr>
<tr>
<td>XGA, 1024x768</td>
<td>15, 30, 60</td>
<td>AI, RA, JFPIC</td>
</tr>
<tr>
<td>VGA, 640x480</td>
<td>15, 30, 60</td>
<td>AI, RA, JFPIC</td>
</tr>
<tr>
<td><strong>Input color format: YUV 4:4:4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1440p (2K), 2560x1440</td>
<td>15, 30, 60</td>
<td>AI, RA, JFPIC</td>
</tr>
<tr>
<td>1080p, 1920x1080</td>
<td>15, 30, 60</td>
<td>AI, RA, JFPIC</td>
</tr>
<tr>
<td>720p, 1280x720</td>
<td>15, 30, 60</td>
<td>AI, RA, JFPIC</td>
</tr>
</tbody>
</table>
Game streaming

• **Basic requirements:**
  - Random access to pictures
  - Temporal (frame-rate) scalability
  - Error robustness

• **Optional requirements:**
  - Resolution and quality (SNR) scalability

• **Specific features:**
  - This content typically contains many sharp edges and large motion
Video monitoring / surveillance

• **Basic requirements:**
  - Random access to pictures for downloaded video data
    - Random Access Period (RAP) should be kept in the range of 1-5 seconds
  - Low-complexity encoder

• **Optional requirements:**
  - Support of high dynamic range
  - Temporal, resolution and quality (SNR) scalability
# Video monitoring / surveillance

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Frame-rate, fps</th>
<th>Picture access mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>2160p (4K), 3840x2160</td>
<td>12</td>
<td>RA</td>
</tr>
<tr>
<td>5Mpixels, 2560x1920</td>
<td>12</td>
<td>RA</td>
</tr>
<tr>
<td>1080p, 1920x1080</td>
<td>25</td>
<td>RA</td>
</tr>
<tr>
<td>1.3Mpixels, 1280x960</td>
<td>25, 30</td>
<td>RA</td>
</tr>
<tr>
<td>720p, 1280x720</td>
<td>25, 30</td>
<td>RA</td>
</tr>
<tr>
<td>SVGA, 800x600</td>
<td>25, 30</td>
<td>RA</td>
</tr>
</tbody>
</table>
Requirements

- Basic requirements
- Optional requirements
Basic requirements

• **Coding efficiency / compression performance**
  - It should be better than for state-of-the-art video codecs such as HEVC/H.265 and VP9

• **Input source formats:**
  - Bit depth:
    - 8- and 10-bits per color component
  - Color sampling formats:
    - YUV 4:2:0 and YUV 4:4:4

• **Coding delay**
  - Support of “low-delay” configurations
    - Delay should be up to 320 ms but its preferable value should be less than 100 ms
Basic requirements (cont’d)

- **Complexity**
  - Feasible real-time implementation of both an encoder and a decoder for hardware and software implementation based on a wide range of state-of-the-art platforms

- **Scalability**
  - Temporal (frame-rate) scalability

- **Error resilience**
  - Error resilience tools that are complementary to the error protection mechanisms implemented on transport level
Optional requirements

- **Input source formats:**
  - Bit depth:
    - up to 16-bits per color component
  - Color sampling formats:
    - YUV 4:2:2 and RGB
  - Support of auxiliary channel:
    - e.g., alpha channel
  - Support of high dynamic range

- **Scalability:**
  - Resolution and quality (SNR) scalability
  - Computational complexity scalability
    - Computational complexity is decreasing along with degrading picture quality
Optional requirements (cont’d)

• **Complexity**
  - Tools that enable parallel processing at both encoder and decoder sides are highly desirable for many applications
    - E.g., slices, tiles, wave front propagation processing

  - **High-level multi-core parallelism**
    - Encoder and decoder operation, especially entropy encoding and decoding, should allow multiple frames or sub-frame regions (e.g. 1D slices, 2D tiles, or partitions) to be processed concurrently, either independently or with deterministic dependencies that can be efficiently pipelined

  - **Low-level instruction set parallelism**
    - Favor algorithms that are SIMD/GPU friendly over inherently serial algorithms
Compression performance evaluation

• Methodology of compression performance evaluation

• Quality assessment
  - Objective evaluation
  - Subjective evaluation
Methodology of compression performance evaluation

• Requirements do not make sense if a way of how to check them is not defined
  ▪ In this draft, just a high-level evaluation framework is proposed
    □ Further details (e.g., a list of video sequences, concrete bit-rates, etc) should be described in a separate document
  ▪ The draft only encompasses an evaluation methodology for compression performance
    □ However, evaluation procedure should be proposed for each requirement if checking its fulfillment is not evident
Methodology of compression performance evaluation (cont’d)

The deviation between bit-rates of reference and tested codecs:

\[ D = \text{abs} \left( \frac{BR_r - BR_t}{BR_r} \right) \cdot 100\% < D_{THR} \]

where \( BR_r \) and \( BR_t \) are bit-rates of reference and tested codecs

- Nominal value of bit-rate
- Value of bit-rate for the 1st codec
- Value of bit-rate for the 2nd codec

For obtaining an integral result in each range, **Bjøntegaard Delta (BD)-rate** should be computed.
Quality assessment

- Objective evaluation
  - Peak Signal-to-Noise Ration (PSNR)
    - where $B$ is the bit depth of source signal
    - $R$ and $T$ are original and reconstructed signals, respectively
  - Multiscale Structural Similarity (MS-SSIM)

$$ssim(x_i, y_i) = [l(x_i, y_i)]^r \cdot [c(x_i, y_i)]^g \cdot [s(x_i, y_i)]^b$$

$$ssim(x_i, y_i) = \frac{2\mu_{x_i}\mu_{y_i} + C_1}{\mu_{x_i}^2 + \mu_{y_i}^2 + C_1} \cdot \frac{2\sigma_{x_iy_i} + C_2}{\sigma_{x_i}^2 + \sigma_{y_i}^2 + C_2}$$

$$SSIM(X,Y) = \frac{1}{N} \sum_{i=1}^{N} ssim(x_i, y_i)$$

$$PSNR = 20 \log \frac{2^B - 1}{\sqrt{\frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} (R(x, y) - S(x, y))^2}}$$
Quality assessment (cont’d)

- Subjective evaluation
  - Final and some intermediate decisions should be made using subjective evaluation
  - Mean Opinion Score (MOS)
    - MOS provides a numerical indication of the perceived quality of a picture or a picture sequence after a process such as compression, quantization, transmission and so on.
    - The MOS is expressed as a single number in the range 1 to 5 in the case of a discrete scale (resp., 1 to 100 in the case of a continuous scale)
      - where 1 is the lowest perceived quality, and 5 (resp., 100) is the highest perceived quality
    - Confidence interval can be calculated
    - Some outliers can be rejected
      - This rejection allows us to correct influences induced by the observer's behavior, or bad choice of test pictures or picture sequences
Conclusions

- This document contains
  - an overview of Internet video codec applications and typical use cases
  - a prioritized list of requirements for an Internet video codec

- An evaluation methodology for this codec is also proposed
  - We strongly recommend to the NETVC WG to include an evaluation framework into the requirements output document
  - Since in the previous meeting, one of the main goals was formulated as to be “better than state-of-the-art compression”, we suggest performing comparison with the reference model of HEVC/H.265
    - In the future, even with the Joint Exploration Model (JEM) software
Thank You