Video Codec Requirements and Evaluation Methodology

draft-ietf-netvc-requirements-01

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

IPTV / OTT video

Resolution	Frame-rate, fps	Picture access mode
2160p (4K),3840x2160	24/1.001, 24, 25, 30/1.001, 30, 50, 60/1.001, 60, 100, 120/1.001, 120 (Table 2 in ITU-R BT-2020)	RA
1080p, 1920x1080		RA
1080i, 1920x1080 *		RA
720p, 1280x720		RA
576p (EDTV), 720x576		RA
576i (SDTV), 720x576 *		RA
480p (EDTV), 720x480		RA
480i (SDTV), 720x480 *		RA

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 320 ms, respectively
  - Temporal (frame-rate) scalability;
  - Error robustness

- **Optional requirements:**
  - resolution and quality (SNR) scalability

Video conferencing

Resolution	Frame-rate, fps	Picture access mode
1080p, 1920x1080	15, 30	FIZD
720p, 1280x720	30, 60	FIZD
4CIF, 704x576	30, 60	FIZD
4SIF, 704x480	30, 60	FIZD
VGA, 640x480	30, 60	FIZD
360p, 640x360	30, 60	FIZD

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*

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

* **- Sources of these data:**
  - "Recommended upload encoding settings (Advanced)" 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

Resolution	Frame-rate, fps	Picture access mode
Input color format: **RBG**		
WQXGA, 2560x1600	15, 30, 60	AI, RA, FIZD
WUXGA, 1920x1200	15, 30, 60	AI, RA, FIZD
WSXGA+, 1680x1050	15, 30, 60	AI, RA, FIZD
WXGA, 1280x800	15, 30, 60	AI, RA, FIZD
XGA, 1024x768	15, 30, 60	AI, RA, FIZD
SVGA, 800x600	15, 30, 60	AI, RA, FIZD
VGA, 640x480	15, 30, 60	AI, RA, FIZD
Input color format: **YUV 4:4:4**		
1440p (2K), 2560x1440	15, 30, 60	AI, RA, FIZD
1080p, 1920x1080	15, 30, 60	AI, RA, FIZD
720p, 1280x720	15, 30, 60	AI, RA, FIZD

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

Resolution	Frame-rate, fps	Picture access mode
2160p (4K),3840x2160	12	RA
5Mpixels, 2560x1920	12	RA
1080p, 1920x1080	25	RA
1.3Mpixels, 1280x960	25, 30	RA
720p, 1280x720	25, 30	RA
SVGA, 800x600	25, 30	RA

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
- **End-to-end delay**
  - Support of configurations with zero structural delay also referred to as “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 and wide color gamut

- **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

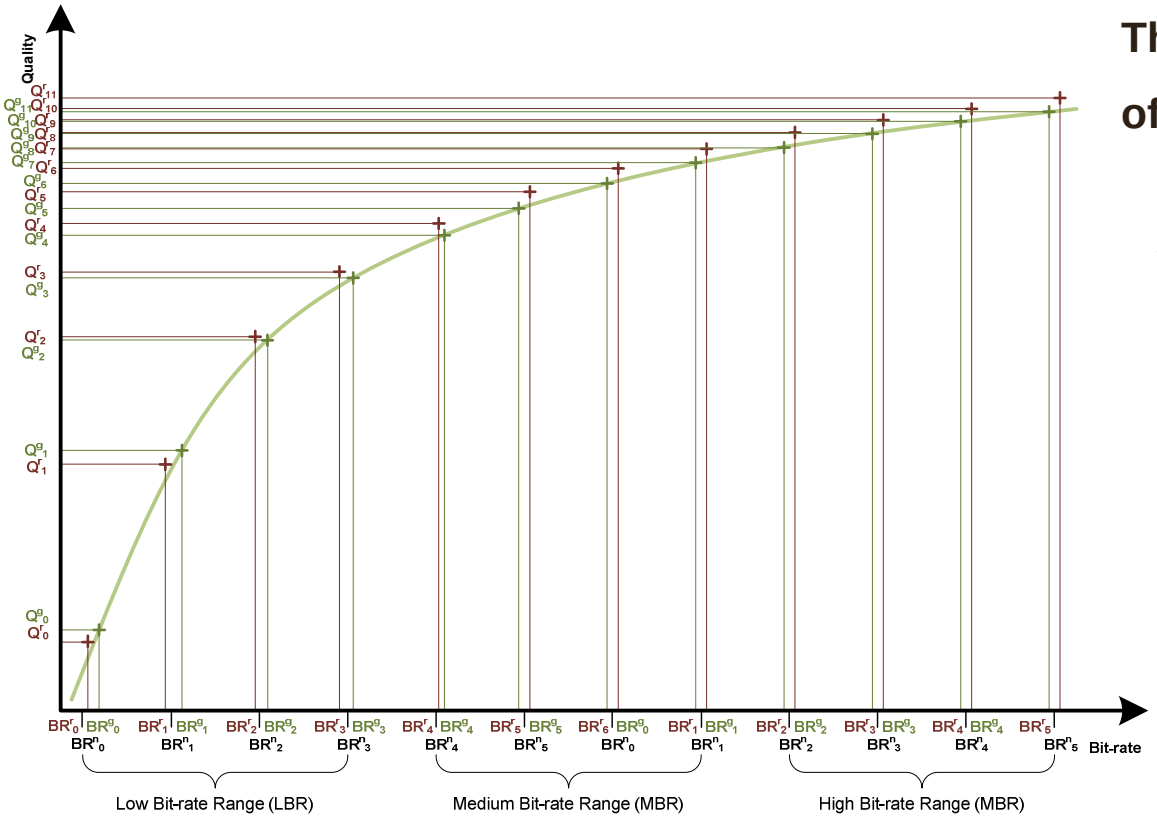
Compression performance evaluation

- **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:

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

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

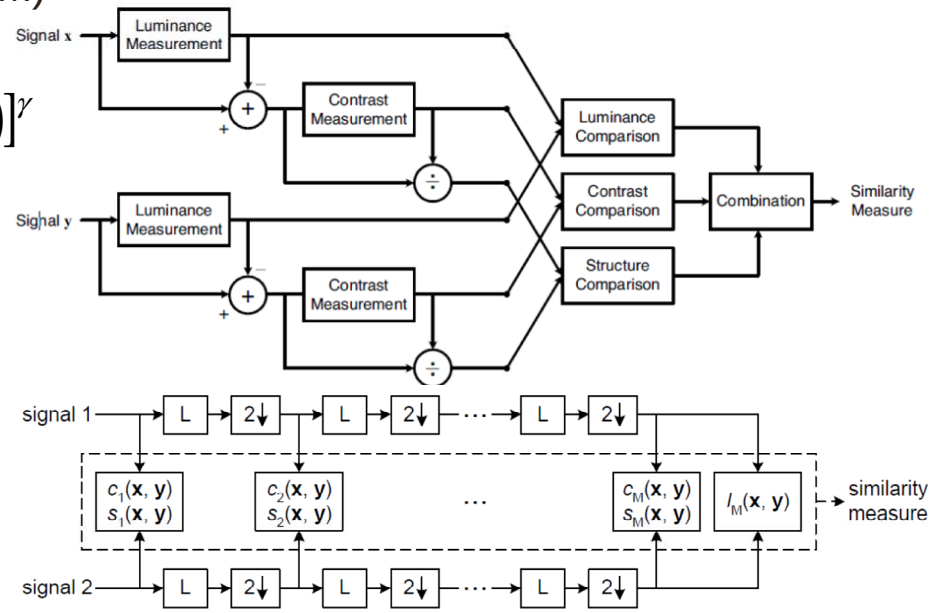
- Objective evaluation
  - Peak Signal-to-Noise Ration (PSNR)
    - where *B* is the bit depth of source signal
    - *R* and *T* are original and reconstructed signals, respectively
  - Multiscale Structural Similarity (MS-SSIM)

$$PSNR = 20\,\mathrm{Log}\left(\frac{\left(2^{B}-1\right)}{\sqrt{\frac{1}{MN}\sum_{y=1}^{M}\sum_{x=1}^{N}\left(R(x,y)-S(x,y)\right)^{2}}}\right)$$

$$ssim(x_i, y_i) = \left[l(x_i, y_i)\right]^{\alpha} \cdot \left[c(x_i, y_i)\right]^{\beta} \cdot \left[s(x_i, y_i)\right]^{\gamma}$$

$$ssim(x_i, y_i) = \frac{\left(2\mu_{xi}\mu_{yi} + C_1\right)\left(2\sigma_{xiyi} + C_2\right)}{\left(\mu_{xi}^2 + \mu_{yi}^2 + C_1\right)\left(\sigma_{xi}^2 + \sigma_{yi}^2 + C_2\right)}$$

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



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

An overview of received comments

- The changes include description of the Internet video streaming use case, which is very relevant for Netflix, and proposed modifications and suggestions to the general requirements to NETVC.
- IPTV and Internet Video Streaming (OTT) use cases are separated
- The following basic requirements are proposed for the Internet Video Streaming use case:
  - Support of HDR, WCG and high frame-rate
  - Low-complexity decoder
  - Frequent RAP
- General requirements for the NETVC codec:
  - Good quality specification and well-defined profiles and levels are required to enable device interoperability and facilitate decoder implementations
  - High-level syntax shall allow extensibility
  - Elementary stream shall have a model that allows easy parsing and identification of the sample components (such as ISO/IEC14496-10, Annex B or ISO/IEC 14496-15).
  - Perceptual quality tools, such as adaptive QP and quantization matrices, shall be supported
  - Small penalty for resolution and quality (SNR) scalability support
    - less than 5% of bit-rate increase per layer
  - Compression efficiency on video with film grain noise (for movie and TV content)
  - Input source formats (basic requirements)
    - HDR and WCG shall be supported.
    - Bit depth: 8- and 10-bits per color component from the start, preferably 12-bit support as well.
    - Color sampling formats: YCbCr 4:2:0, YCbCr 4:4:4,YCbCr 4:2:2

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 overview of received comments to be taken into account by the next IETF NETVC meeting

- 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