Video Codec Requirements and Evaluation Methodology

draft-ietf-netvc-requirements-02

Alexey Filippov (Huawei Technologies),
Andrey Norkin (Netflix),
Jose Alvarez (Huawei Technologies)

Contents

- An overview of applications
- Requirements
- Evaluation methodology
- Conclusions

Applications

- Internet Video Streaming
- Internet Protocol Television (IPTV)
- Video conferencing
- Video sharing
- Screencasting
- Game streaming
- Video monitoring / surveillance

Internet Video Streaming

- **Basic requirements:**
  - Significant improvements in compression efficiency between codec generations
  - Random access to pictures
    - Random Access Period (RAP) usually 2-5 seconds
  - Support of wide range of content types and formats
    - HDR and WCG
    - Gains on lower resolutions is important for adaptive streaming (many resolution)
    - Gains on easy content are also important for overall bitrate savings
    - Efficiency for film grain encoding which is present in a lot of content
  - Tools for perceptually optimized encoding
  - High encoding complexity can be tolerated in software encoders (up to 10x)
  - Bitstream should have a model allowing easy parsing and identification of components (frames, etc)
- **Optional requirements:**
  - Resolution, quality (SNR) and temporal (frame-rate) scalability

Internet Video Streaming

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 (2K), 1920x1080		RA
1080i, 1920x1080 *		RA
720p, 1280x720		RA
576p (EDTV), 720x576		RA
576i (SDTV), 720x576 *		RA
480p (EDTV), 720x480		RA
480i (SDTV), 720x480 *		RA
512x384		RA
QVGA, 320x240		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.

Internet Protocol Television (IPTV)

- **Basic requirements:**
  - Significant improvements in compression efficiency between codec generations

    Random access to pictures
    - Random Access Period (RAP) usually 0.5-1 seconds
  - Support of wide range of content types and formats
    - HDR and WCG
    - Efficiency for film grain encoding which is present in a lot of content
  - Tools for perceptually optimized encoding
  - Bitstream should have a model allowing easy parsing and identification of components (frames, etc)

- **Optional requirements:**
  - Resolution, quality (SNR) and temporal (frame-rate) scalability

Internet Protocol Television (IPTV)

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 end-to-end delay values should be less than 100 ms and 320 ms, respectively
  - Error robustness
  - Low-complexity encoder

- **Optional requirements:**
  - Temporal (frame-rate), 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
- **Optional requirements:**
  - 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 YCbCr 4:4:4 in addition to YCbCr 4:2:0 and YCbCr 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 4:4:4**		
5k, 5120x2880	15, 30, 60	AI, RA, FIZD
4k, 3840x2160	15, 30, 60	AI, RA, FIZD
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

Screencasting

Resolution	Frame-rate, fps	Picture access mode
Input color format: **YCbCr 4:4:4**		
5k, 5120x2880	15, 30, 60	AI, RA, FIZD
4k, 3840x2160	15, 30, 60	AI, RA, FIZD
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
  - Support of HDR
  - In some cases, high quality (fidelity) of a video signal is required after lossy compression

- **Optional requirements:**
  - Support of WCG
  - Support of a monochrome mode
    - e.g., for infrared cameras
  - 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

- General requirements
- Basic requirements
- Optional requirements

General requirements

- **Coding efficiency / compression performance**
  - Improvements over state-of-the-art video codecs such as HEVC/H.265 and VP9, at least, by 20-25%, preferably more
- **Good quality specification and well-defined profiles and levels:**
  - They are required to enable device interoperability and facilitate decoder implementations
- **High-level syntax should allow extensibility**
  - New features can be supported easily by using metadata
    - such as SEI messages, VUI, headers
- **Bit-stream should have a model that allows easy parsing and identification of components (such as frames)**
  - Similar to ISO/IEC14496-10, Annex B or ISO/IEC 14496-15
  - In particular, information needed for packet handling (e.g., frame type) should not require parsing anything below the header level.

General requirements (cont'd)

- **Support of perceptual quality tools**
  - such as adaptive QP and quantization matrices
- **The codec specification should define a buffer model**
  - Such as hypothetical reference decoder (HRD)
- **Specifications providing integration with system and delivery layers should be developed**

Basic requirements

- **Input source formats:**
  - Bit depth:
    - 8- and 10-bits per color component
    - Up to 12-bits for a high profile
  - Color sampling formats:
    - YCbCr 4:2:0
    - YCbCr 4:4:4, YCbCr 4:2:2 and YCbCr 4:0:0 (preferably in different profile(s))
  - Support of HDR and WCG
    - For profiles with bit depth of 10 bits per sample or higher

- **Support of arbitrary resolution (constrained to level limits) for such applications where a picture can have an arbitrary size**
  - e.g., in screencasting

Basic requirements (cont'd)

- **Coding delay**
  - Support of configurations with zero structural delay also referred to as "low-delay" configurations
    - **Note:** End-to-end delay should be up to 320 ms but its preferable value should be less than 100 ms
  - Support of configurations with non-zero structural delay
    - such as out-of-order or multi-pass encoding to provide additional compression efficiency improvements
- **Scalability**
  - Temporal (frame-rate) scalability

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
    - Real-time encoder should provide sufficient improvement in compression efficiency at reasonable encoder complexity increase
  - High-complexity software encoder implementations used by offline encoding applications
    - They can have 10x or more complexity increase compared to state-of-the-art video compression technologies such as HEVC/H.265 and VP9

Basic requirements (cont'd)

- **Error resilience**
  - Error resilience tools that are complementary to the error protection mechanisms implemented on transport level
  - The codec should support mechanisms that facilitate packetization of a bitstream for common network protocols
  - Packetization mechanisms should enable frame-level error recovery by means of retransmission or error concealment
  - The bitstream specification should support independently decodable sub-frame units similar to slices or independent tiles
    - It should be possible for the encoder to restrict the bit-stream to allow parsing of the bit-stream after a packet loss and to communicate it to the decoder

Optional requirements

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

- **Scalability:**
  - Resolution and quality (SNR) scalability
    - If they provide low compression efficiency penalty, they can be supported in the main profile
  - 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
- **Coding efficiency**
  - Compression efficiency on noisy content, content with film grain, computer generated content, and low resolution materials is desirable

Compression performance evaluation

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

Methodology of compression performance evaluation (cont'd)

- Objective evaluation in 3 ranges:
  - Low-bitrate range
  - Middle-bitrate range
  - High-bitrate range

- Points are selected using the reference codec quality levels

- Bjøntegaard Delta (BD)-rate should be computed:
  - An average value over all the 3 ranges should be provided
  - Values for each range should be provided as well

Quality assessment

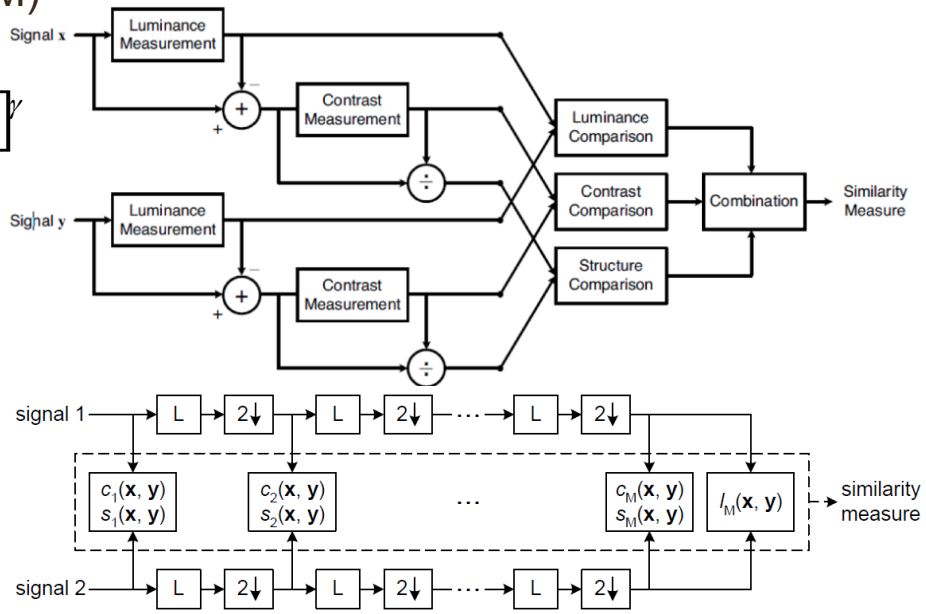
- Objective evaluation
  - Peak Signal-to-Noise Ration (PSNR)
    - where *B* is the bit depth of source signal
    - *R* and *T* are original and reconstructed signals, respectively
  - Multiscale Structural Similarity (MS-SSIM)

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

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

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

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



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

Methodology of compression performance evaluation

- **In this draft, just a high-level evaluation framework is proposed**
  - Further details (e.g., a list of video sequences, concrete bit-rates, etc) are described in the testing draft
  - The draft only encompasses an evaluation methodology for compression performance

- **Reference software**
  - Reference software provided to the NETVC WG for candidate codecs should comprise a fully operational encoder
    - that supports necessary rate controls, subjective quality optimization features and some degree of speed optimization and a “real-time” decoder

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
- The authors tried to take into account all the received comments
- An evaluation methodology for this codec is also proposed
- **We recommend to adopt this document**

Thank You