Internet Engineering Task Force (IETF)
Request for Comments: 8761
Category: Informational
ISSN: 2070-1721

A. Filippov
Huawei Technologies
A. Norkin
Netflix
J.R. Alvarez
Huawei Technologies

April 2020

Video Codec Requirements and Evaluation Methodology

Abstract

This document provides requirements for a video codec designed mainly for use over the Internet. In addition, this document describes an evaluation methodology for measuring the compression efficiency to determine whether or not the stated requirements have been fulfilled.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8761.

Table of Contents

1. Introduction
2. Terminology Used in This Document
   2.1. Definitions
   2.2. Abbreviations

3.1. Internet Video Streaming
3.2. Internet Protocol Television (IPTV)
3.3. Video Conferencing
3.4. Video Sharing
3.5. Screencasting
3.6. Game Streaming
3.7. Video Monitoring and Surveillance
4. Requirements
4.1. General Requirements
4.1.1. Coding Efficiency
4.1.2. Profiles and Levels
4.1.3. Bitstream Syntax
4.1.4. Parsing and Identification of Sample Components
4.1.5. Perceptual Quality Tools
4.1.6. Buffer Model
4.1.7. Integration
4.2. Basic Requirements
4.2.1. Input Source Formats
4.2.2. Coding Delay
4.2.3. Complexity
4.2.4. Scalability
4.2.5. Error Resilience
4.3. Optional Requirements
4.3.1. Input Source Formats
4.3.2. Scalability
4.3.3. Complexity
4.3.4. Coding Efficiency
5. Evaluation Methodology
6. Security Considerations
7. IANA Considerations
8. References
8.1. Normative References
8.2. Informative References
Acknowledgments
Authors' Addresses

1. Introduction

This document presents the requirements for a video codec designed mainly for use over the Internet. The requirements encompass a wide range of applications that use data transmission over the Internet, including Internet video streaming, IPTV, peer-to-peer video conferencing, video sharing, screencasting, game streaming, and video monitoring and surveillance. For each application, typical resolutions, frame rates, and picture-access modes are presented. Specific requirements related to data transmission over packet-loss networks are considered as well. In this document, when we discuss data-protection techniques, we only refer to methods designed and implemented to protect data inside the video codec since there are many existing techniques that protect generic data transmitted over networks with packet losses. From the theoretical point of view, both packet-loss and bit-error robustness can be beneficial for video codecs. In practice, packet losses are a more significant problem than bit corruption in IP networks. It is worth noting that there is an evident interdependence between the possible amount of delay and the necessity of error-robust video streams:

* If the amount of delay is not crucial for an application, then reliable transport protocols such as TCP that retransmit undelivered packets can be used to guarantee correct decoding of transmitted data.

* If the amount of delay must be kept low, then either data transmission should be error free (e.g., by using managed networks) or the compressed video stream should be error resilient.

Thus, error resilience can be useful for delay-critical applications to provide low delay in a packet-loss environment.

2. Terminology Used in This Document

2.1. Definitions

High dynamic range imaging
A set of techniques that allows a greater dynamic range of exposures or values (i.e., a wider range of values between light and dark areas) than normal digital imaging techniques. The intention is to accurately represent the wide range of intensity levels found in examples such as exterior scenes that include light-colored items struck by direct sunlight and areas of deep shadow [7].

Random access period
The period of time between the two closest independently decodable frames (pictures).

RD-point
A point in a two-dimensional rate-distortion space where the values of bitrate and quality metric are used as x- and y-coordinates, respectively.

Visually lossless compression
A form or manner of lossy compression where the data that are lost after the file is compressed and decompressed is not detectable to the eye; the compressed data appear identical to the uncompressed data [8].

Wide color gamut
A certain complete color subset (e.g., considered in ITU-R BT.2020 [1]) that supports a wider range of colors (i.e., an extended range of colors that can be generated by a specific input or output device such as a video camera, monitor, or printer and can be interpreted by a color model) than conventional color gamuts (e.g., considered in ITU-R BT.601 [17] or BT.709 [20]).

2.2. Abbreviations

AI	All-Intra (each picture is intra-coded)
BD-Rate	Bjontegaard Delta Rate

FIZD	just the First picture is Intra-coded, Zero structural Delay
FPS	Frames per Second
GOP	Group of Picture
GPU	Graphics Processing Unit
HBR	High Bitrate Range
HDR	High Dynamic Range
HRD	Hypothetical Reference Decoder
HEVC	High Efficiency Video Coding
IPTV	Internet Protocol Television
LBR	Low Bitrate Range
MBR	Medium Bitrate Range
MOS	Mean Opinion Score
MS-SSIM	Multi-Scale Structural Similarity quality index
PAM	Picture Access Mode
PSNR	Peak Signal-to-Noise Ratio
QoS	Quality of Service
QP	Quantization Parameter
RA	Random Access
RAP	Random Access Period
RD	Rate-Distortion
SEI	Supplemental Enhancement Information
SIMD	Single Instruction, Multiple Data
SNR	Signal-to-Noise Ratio
UGC	User-Generated Content
VDI	Virtual Desktop Infrastructure
VUI	Video Usability Information
WCG	Wide Color Gamut

3. Applications

In this section, an overview of video codec applications that are currently available on the Internet market is presented. It is worth noting that there are different use cases for each application that define a target platform; hence, there are different types of communication channels involved (e.g., wired or wireless channels) that are characterized by different QoS as well as bandwidth; for instance, wired channels are considerably more free from error than wireless channels and therefore require different QoS approaches. The target platform, the channel bandwidth, and the channel quality determine resolutions, frame rates, and either quality or bitrates for video streams to be encoded or decoded. By default, color format YCbCr 4:2:0 is assumed for the application scenarios listed below.

3.1. Internet Video Streaming

Typical content for this application is movies, TV series and shows, and animation. Internet video streaming uses a variety of client devices and has to operate under changing network conditions. For this reason, an adaptive streaming model has been widely adopted. Video material is encoded at different quality levels and different resolutions, which are then chosen by a client depending on its capabilities and current network bandwidth. An example combination of resolutions and bitrates is shown in Table 1.

A video encoding pipeline in on-demand Internet video streaming typically operates as follows:

* Video is encoded in the cloud by software encoders.

* Source video is split into chunks, each of which is encoded separately, in parallel.

* Closed-GOP encoding with intrapicture intervals of 2-5 seconds (or longer) is used.

* Encoding is perceptually optimized. Perceptual quality is important and should be considered during the codec development.

Resolution *	PAM	Frame Rate, FPS **
4K, 3840x2160	RA	24/1.001, 24, 25, 30/1.001, 30, 50, 60/1.001, 60, 100, 120/1.001, 120
2K (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	
512x384	RA	
QVGA, 320x240	RA	

Table 1: Internet Video Streaming: Typical Values of Resolutions, Frame Rates, and PAMs

*Note: 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 the progressive format.

**Note: The set of frame rates presented in this table is taken from Table 2 in [1].

The characteristics and requirements of this application scenario are as follows:

* High encoder complexity (up to 10x and more) can be tolerated since encoding happens once and in parallel for different segments.

* Decoding complexity should be kept at reasonable levels to enable efficient decoder implementation.

* Support and efficient encoding of a wide range of content types and formats is required:

  - High Dynamic Range (HDR), Wide Color Gamut (WCG), high-resolution (currently, up to 4K), and high-frame-rate content are important use cases; the codec should be able to encode such content efficiently.

  - Improvement of coding efficiency at both lower and higher resolutions is important since low resolutions are used when streaming in low-bandwidth conditions.

  - Improvement on both "easy" and "difficult" content in terms of

compression efficiency at the same quality level contributes to the overall bitrate/storage savings.

- Film grain (and sometimes other types of noise) is often present in movies and similar content; this is usually part of the creative intent.

* Significant improvements in compression efficiency between generations of video standards are desirable since this scenario typically assumes long-term support of legacy video codecs.

* Random access points are inserted frequently (one per 2-5 seconds) to enable switching between resolutions and fast-forward playback.

* The elementary stream should have a model that allows easy parsing and identification of the sample components.

* Middle QP values are normally used in streaming; this is also the range where compression efficiency is important for this scenario.

* Scalability or other forms of supporting multiple quality representations are beneficial if they do not incur significant bitrate overhead and if mandated in the first version.

3.2. Internet Protocol Television (IPTV)

This is a service for delivering television content over IP-based networks. IPTV may be classified into two main groups based on the type of delivery, as follows:

* unicast (e.g., for video on demand), where delay is not crucial; and

* multicast/broadcast (e.g., for transmitting news) where zapping (i.e., stream changing) delay is important.

In the IPTV scenario, traffic is transmitted over managed (QoS-based) networks. Typical content used in this application is news, movies, cartoons, series, TV shows, etc. One important requirement for both groups is that random access to pictures (i.e., the random access period (RAP)) should be kept small enough (approximately 1-5 seconds). Optional requirements are as follows:

* Temporal (frame-rate) scalability; and

* Resolution and quality (SNR) scalability.

For this application, typical values of resolutions, frame rates, and PAMs are presented in Table 2.

Resolution *	PAM	Frame Rate, FPS **
2160p (4K),	RA	24/1.001, 24, 25, 30/1.001, 30, 50,

3840x2160		60/1.001, 60, 100, 120/1.001, 120
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	

Table 2: IPTV: Typical Values of Resolutions, Frame Rates, and PAMs

*Note: 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 a progressive format.

**Note: The set of frame rates presented in this table is taken from Table 2 in [1].

3.3. Video Conferencing

This is a form of video connection over the Internet. This form allows users to establish connections to two or more people by two-way video and audio transmission for communication in real time. For this application, both stationary and mobile devices can be used. The main requirements are as follows:

* Delay should be kept as low as possible (the preferable and maximum end-to-end delay values should be less than 100 ms [9] and 320 ms [2], respectively);

* Temporal (frame-rate) scalability; and

* Error robustness.

Support of resolution and quality (SNR) scalability is highly desirable. For this application, typical values of resolutions,

frame rates, and PAMs are presented in Table 3.

Resolution	Frame Rate, FPS	PAM
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

Table 3: Video Conferencing: Typical Values of Resolutions, Frame Rates, and PAMs

3.4. Video Sharing

This is a service that allows people to upload and share video data (using live streaming or not) and watch those videos. It is also known as video hosting. A typical User-Generated Content (UGC) scenario for this application is to capture video using mobile cameras such as GoPros or cameras integrated into smartphones (amateur video). The main requirements are as follows:

* Random access to pictures for downloaded video data;

* Temporal (frame-rate) scalability; and

* Error robustness.

Support of resolution and quality (SNR) scalability is highly desirable. For this application, typical values of resolutions, frame rates, and PAMs are presented in Table 4.

Typical values of resolutions and frame rates in Table 4 are taken from [10].

Resolution	Frame Rate, FPS	PAM
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

Table 4: Video Sharing: Typical Values of Resolutions, Frame Rates, and PAMs

3.5. Screencasting

This is a service that allows users to record and distribute video data from a computer screen. This service requires efficient compression of computer-generated content with high visual quality up to visually and mathematically (numerically) lossless [11]. Currently, this application includes business presentations (PowerPoint, Word documents, email messages, etc.), animation (cartoons), gaming content, and data visualization. This type of content is characterized by fast motion, rotation, smooth shade, 3D effect, highly saturated colors with full resolution, clear textures and sharp edges with distinct colors [11], virtual desktop infrastructure (VDI), screen/desktop sharing and collaboration, supervisory control and data acquisition (SCADA) display, automotive/navigation display, cloud gaming, factory automation display, wireless display, display wall, digital operating room (DiOR), etc. For this application, an important requirement is the support of low-delay configurations with zero structural delay for a wide range of video formats (e.g., RGB) in addition to YCbCr 4:2:0 and YCbCr 4:4:4 [11]. For this application, typical values of resolutions, frame rates, and PAMs are presented in Table 5.

Resolution	Frame Rate, FPS	PAM
Input color format: RGB 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
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

Table 5: Screencasting for RGB and YCbCr 4:4:4 Format: Typical Values of Resolutions, Frame Rates, and PAMs

3.6. Game Streaming

This is a service that provides game content over the Internet to different local devices such as notebooks and gaming tablets. In this category of applications, the server renders 3D games in a cloud server and streams the game to any device with a wired or wireless broadband connection [12]. There are low-latency requirements for transmitting user interactions and receiving game data with a turnaround delay of less than 100 ms. This allows anyone to play (or resume) full-featured games from anywhere on the Internet [12]. An example of this application is Nvidia Grid [12]. Another application scenario of this category is broadcast of video games played by people over the Internet in real time or for later viewing [12]. There are many companies, such as Twitch and YY in China, that enable game broadcasting [12]. Games typically contain a lot of sharp edges and large motion [12]. The main requirements are as follows:

* Random access to pictures for game broadcasting;

* Temporal (frame-rate) scalability; and

* Error robustness.

Support of resolution and quality (SNR) scalability is highly desirable. For this application, typical values of resolutions, frame rates, and PAMs are similar to ones presented in Table 3.

3.7. Video Monitoring and Surveillance

This is a type of live broadcasting over IP-based networks. Video streams are sent to many receivers at the same time. A new receiver may connect to the stream at an arbitrary moment, so the random access period should be kept small enough (approximately, 1-5 seconds). Data are transmitted publicly in the case of video monitoring and privately in the case of video surveillance. For IP cameras that have to capture, process, and encode video data, complexity -- including computational and hardware complexity, as well as memory bandwidth -- should be kept low to allow real-time processing. In addition, support of a high dynamic range and a monochrome mode (e.g., for infrared cameras) as well as resolution and quality (SNR) scalability is an essential requirement for video surveillance. In some use cases, high video signal fidelity is required even after lossy compression. Typical values of resolutions, frame rates, and PAMs for video monitoring and

surveillance applications are presented in Table 6.

Resolution	Frame Rate, FPS	PAM
2160p (4K), 3840x2160	12, 25, 30	RA, FIZD
5Mpixels, 2560x1920	12, 25, 30	RA, FIZD
1080p, 1920x1080	25, 30	RA, FIZD
1.23Mpixels, 1280x960	25, 30	RA, FIZD
720p, 1280x720	25, 30	RA, FIZD
SVGA, 800x600	25, 30	RA, FIZD

Table 6: Video Monitoring and Surveillance: Typical Values of Resolutions, Frame Rates, and PAMs

4. Requirements

Taking the requirements discussed above for specific video applications, this section proposes requirements for an Internet video codec.

4.1. General Requirements

4.1.1. Coding Efficiency

The most fundamental requirement is coding efficiency, i.e., compression performance on both "easy" and "difficult" content for applications and use cases in Section 3. The codec should provide higher coding efficiency over state-of-the-art video codecs such as HEVC/H.265 and VP9, at least 25%, in accordance with the methodology described in Section 5 of this document. For higher resolutions, the improvements in coding efficiency are expected to be higher than for lower resolutions.

4.1.2. Profiles and Levels

Good-quality specification and well-defined profiles and levels are required to enable device interoperability and facilitate decoder implementations. A profile consists of a subset of entire bitstream syntax elements; consequently, it also defines the necessary tools for decoding a conforming bitstream of that profile. A level imposes a set of numerical limits to the values of some syntax elements. An example of codec levels to be supported is presented in Table 7. An actual level definition should include constraints on features that impact the decoder complexity. For example, these features might be as follows: maximum bitrate, line buffer size, memory usage, etc.

Level	Example picture resolution at highest frame rate

1	128x96(12,288*)@30.0 176x144(25,344*)@15.0
2	352x288(101,376*)@30.0
3	352x288(101,376*)@60.0 640x360(230,400*)@30.0
4	640x360(230,400*)@60.0 960x540(518,400*)@30.0
5	720x576(414,720*)@75.0 960x540(518,400*)@60.0 1280x720(921,600*)@30.0
6	1,280x720(921,600*)@68.0 2,048x1,080(2,211,840*)@30.0
7	1,280x720(921,600*)@120.0
8	1,920x1,080(2,073,600*)@120.0 3,840x2,160(8,294,400*)@30.0 4,096x2,160(8,847,360*)@30.0
9	1,920x1,080(2,073,600*)@250.0 4,096x2,160(8,847,360*)@60.0
10	1,920x1,080(2,073,600*)@300.0 4,096x2,160(8,847,360*)@120.0
11	3,840x2,160(8,294,400*)@120.0 8,192x4,320(35,389,440*)@30.0
12	3,840x2,160(8,294,400*)@250.0 8,192x4,320(35,389,440*)@60.0
13	3,840x2,160(8,294,400*)@300.0 8,192x4,320(35,389,440*)@120.0

Table 7: Codec Levels

*Note: The quantities of pixels are presented for applications in which a picture can have an arbitrary size (e.g., screencasting).

4.1.3. Bitstream Syntax

Bitstream syntax should allow extensibility and backward compatibility. New features can be supported easily by using metadata (such as SEI messages, VUI, and headers) without affecting the bitstream compatibility with legacy decoders. A newer version of the decoder shall be able to play bitstreams of an older version of the same or lower profile and level.

4.1.4. Parsing and Identification of Sample Components

A bitstream should have a model that allows easy parsing and identification of the sample components (such as Annex B of ISO/IEC 14496-10 [18] or ISO/IEC 14496-15 [19]). In particular, information needed for packet handling (e.g., frame type) should not require parsing anything below the header level.

4.1.5. Perceptual Quality Tools

Perceptual quality tools (such as adaptive QP and quantization matrices) should be supported by the codec bitstream.

4.1.6. Buffer Model

The codec specification shall define a buffer model such as hypothetical reference decoder (HRD).

4.1.7. Integration

Specifications providing integration with system and delivery layers should be developed.

4.2. Basic Requirements

4.2.1. Input Source Formats

Input pictures coded by a video codec should have one of the following formats:

* Bit depth: 8 and 10 bits (up to 12 bits for a high profile) per color component.

* 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))

* For profiles with bit depth of 10 bits per sample or higher, support of high dynamic range and wide color gamut.

* Support of arbitrary resolution according to the level constraints for applications in which a picture can have an arbitrary size (e.g., in screencasting).

Exemplary input source formats for codec profiles are shown in Table 8.

Profile	Bit depths per color component	Color sampling formats
1	8 and 10	4:0:0 and 4:2:0
2	8 and 10	4:0:0, 4:2:0,

		and 4:4:4
3	8, 10, and 12	4:0:0, 4:2:0, 4:2:2, and 4:4:4

Table 8: Exemplary Input Source Formats for Codec Profiles

4.2.2. Coding Delay

In order to meet coding delay requirements, a video codec should support all of the following:

* Support of configurations with zero structural delay, also referred to as "low-delay" configurations.
  - Note: End-to-end delay should be no more than 320 ms [2], but it is preferable for its value to be less than 100 ms [9].
* Support of efficient random access point encoding (such as intracoding and resetting of context variables), as well as efficient switching between multiple quality representations.
* Support of configurations with nonzero structural delay (such as out-of-order or multipass encoding) for applications without low-delay requirements, if such configurations provide additional compression efficiency improvements.

4.2.3. Complexity

Encoding and decoding complexity considerations are as follows:

* Feasible real-time implementation of both an encoder and a decoder supporting a chosen subset of tools for hardware and software implementation on a wide range of state-of-the-art platforms. The subset of real-time encoder tools should provide meaningful improvement in compression efficiency at reasonable complexity of hardware and software encoder implementations as compared to real-time implementations of state-of-the-art video compression technologies such as HEVC/H.265 and VP9.
* High-complexity software encoder implementations used by offline encoding applications can have a 10x or more complexity increase compared to state-of-the-art video compression technologies such as HEVC/H.265 and VP9.

4.2.4. Scalability

The mandatory scalability requirement is as follows:

* Temporal (frame-rate) scalability should be supported.

4.2.5. Error Resilience

In order to meet the error resilience requirement, a video codec should satisfy all of the following conditions:

* Tools that are complementary to the error-protection mechanisms implemented on the transport level should be supported.

* 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 codec should support effective mechanisms for allowing decoding and reconstruction of significant parts of pictures in the event that parts of the picture data are lost in transmission.

* The bitstream specification shall support independently decodable subframe units similar to slices or independent tiles. It shall be possible for the encoder to restrict the bitstream to allow parsing of the bitstream after a packet loss and to communicate it to the decoder.

4.3. Optional Requirements

4.3.1. Input Source Formats

It is a desired but not mandatory requirement for a video codec to support some of the following features:

* Bit depth: up to 16 bits per color component.

* Color sampling formats: RGB 4:4:4.

* Auxiliary channel (e.g., alpha channel) support.

4.3.2. Scalability

Desirable scalability requirements are as follows:

* Resolution and quality (SNR) scalability that provides a low-compression efficiency penalty (increase of up to 5% of BD-rate [13] per layer with reasonable increase of both computational and hardware complexity) can be supported in the main profile of the codec being developed by the NETVC Working Group. Otherwise, a separate profile is needed to support these types of scalability.

* Computational complexity scalability (i.e., computational complexity is decreasing along with degrading picture quality) is desirable.

4.3.3. Complexity

Tools that enable parallel processing (e.g., slices, tiles, and wave-front propagation processing) at both encoder and decoder sides are highly desirable for many applications.

* High-level multicore parallelism: encoder and decoder operation, especially entropy encoding and decoding, should allow multiple

frames or subframe 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

4.3.4. Coding Efficiency

Compression efficiency on noisy content, content with film grain, computer generated content, and low resolution materials is desirable.

5. Evaluation Methodology

As shown in Figure 1, compression performance testing is performed in three overlapped ranges that encompass ten different bitrate values:

* Low bitrate range (LBR) is the range that contains the four lowest bitrates of the ten specified bitrates (one of the four bitrate values is shared with the neighboring range).

* Medium bitrate range (MBR) is the range that contains the four medium bitrates of the ten specified bitrates (two of the four bitrate values are shared with the neighboring ranges).

* High bitrate range (HBR) is the range that contains the four highest bitrates of the ten specified bitrates (one of the four bitrate values is shared with the neighboring range).

Initially, for the codec selected as a reference one (e.g., HEVC or VP9), a set of ten QP (quantization parameter) values should be specified as in [14], and corresponding quality values should be calculated. In Figure 1, QP and quality values are denoted as "QP0"-"QP9" and "Q0"-"Q9", respectively. To guarantee the overlaps of quality levels between the bitrate ranges of the reference and tested codecs, a quality alignment procedure should be performed for each range's outermost (left- and rightmost) quality levels Qk of the reference codec (i.e., for Q0, Q3, Q6, and Q9) and the quality levels Q'k (i.e., Q'0, Q'3, Q'6, and Q'9) of the tested codec. Thus, these quality levels Q'k, and hence the corresponding QP value QP'k (i.e., QP'0, QP'3, QP'6, and QP'9), of the tested codec should be selected using the following formulas:

```
Q'k =   min { abs(Q'i - Qk) },
      i in R

QP'k = argmin { abs(Q'i(QP'i) - Qk(QPk)) },
       i in R
```

where R is the range of the QP indexes of the tested codec, i.e., the candidate Internet video codec. The inner quality levels (i.e., Q'1, Q'2, Q'4, Q'5, Q'7, and Q'8), as well as their corresponding QP values of each range (i.e., QP'1, QP'2, QP'4, QP'5, QP'7, and QP'8), should be as equidistantly spaced as possible between the left- and rightmost quality levels without explicitly mapping their values

using the procedure described above.

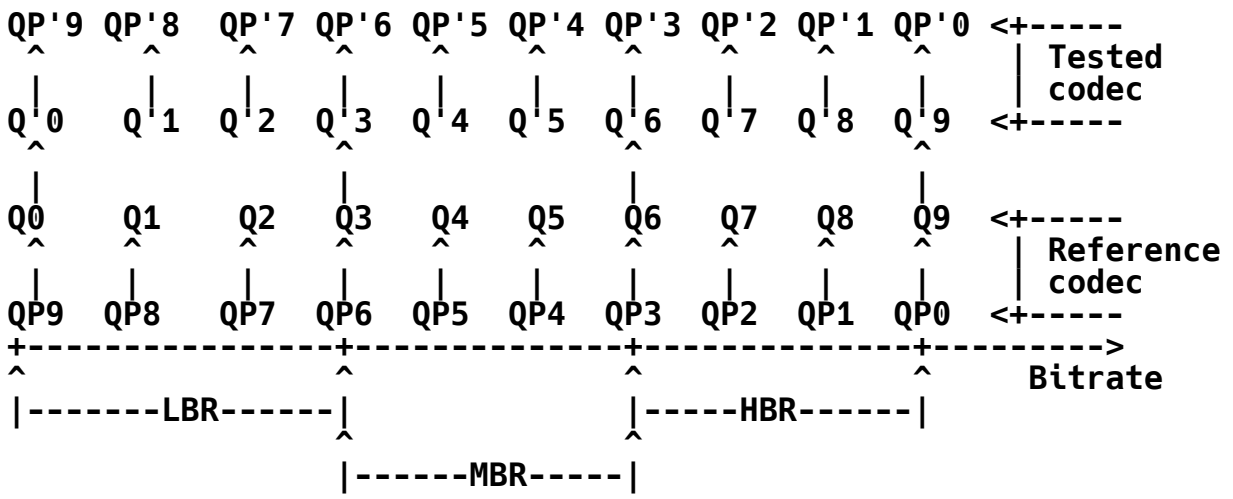

Figure 1: Quality/QP Alignment for Compression Performance Evaluation

Since the QP mapping results may vary for different sequences, this quality alignment procedure eventually needs to be performed separately for each quality assessment index and each sequence used for codec performance evaluation to fulfill the requirements described above.

To assess the quality of output (decoded) sequences, two indexes (PSNR [3] and MS-SSIM [3] [15]) are separately computed. In the case of the YCbCr color format, PSNR should be calculated for each color plane, whereas MS-SSIM is calculated for the luma channel only. In the case of the RGB color format, both metrics are computed for R, G, and B channels. Thus, for each sequence, 30 RD-points for PSNR (i.e., three RD-curves, one for each channel) and 10 RD-points for MS-SSIM (i.e., one RD-curve, for luma channel only) should be calculated in the case of YCbCr. If content is encoded as RGB, 60 RD-points (30 for PSNR and 30 for MS-SSIM) should be calculated (i.e., three RD-curves, one for each channel) are computed for PSNR as well as three RD-curves (one for each channel) for MS-SSIM.

Finally, to obtain an integral estimation, BD-rate savings [13] should be computed for each range and each quality index. In addition, average values over all three ranges should be provided for both PSNR and MS-SSIM. A list of video sequences that should be used for testing, as well as the ten QP values for the reference codec, are defined in [14]. Testing processes should use the information on the codec applications presented in this document. As the reference for evaluation, state-of-the-art video codecs such as HEVC/H.265 [4][5] or VP9 must be used. The reference source code of the HEVC/H.265 codec can be found at [6]. The HEVC/H.265 codec must be configured according to [16] and Table 9.

Intra-period, second	HEVC/H.265 encoding mode according to [16]
AI	Intra Main or Intra

	Main10
RA	Random access Main or Random access Main10
FIZD	Low delay Main or Low delay Main10

Table 9: Intraperiods for Different HEVC/H.265 Encoding Modes According to [16]

According to the coding efficiency requirement described in Section 4.1.1, BD-rate savings calculated for each color plane and averaged for all the video sequences used to test the NETVC codec should be, at least,

* 25% if calculated over the whole bitrate range; and

* 15% if calculated for each bitrate subrange (LBR, MBR, HBR).

Since values of the two objective metrics (PSNR and MS-SSIM) are available for some color planes, each value should meet these coding efficiency requirements. That is, the final BD-rate saving denoted as S is calculated for a given color plane as follows:

S = min { S_psnr, S_ms-ssim }

where S_psnr and S_ms-ssim are BD-rate savings calculated for the given color plane using PSNR and MS-SSIM metrics, respectively.

In addition to the objective quality measures defined above, subjective evaluation must also be performed for the final NETVC codec adoption. For subjective tests, the MOS-based evaluation procedure must be used as described in Section 2.1 of [3]. For perception-oriented tools that primarily impact subjective quality, additional tests may also be individually assigned even for intermediate evaluation, subject to a decision of the NETVC WG.

6. Security Considerations

This document itself does not address any security considerations. However, it is worth noting that a codec implementation (for both an encoder and a decoder) should take into consideration the worst-case computational complexity, memory bandwidth, and physical memory size needed to process the potentially untrusted input (e.g., the decoded pictures used as references).

7. IANA Considerations

This document has no IANA actions.

8. References

8.1. Normative References

[1] ITU-R, "Parameter values for ultra-high definition television systems for production and international programme exchange", ITU-R Recommendation BT.2020-2, October 2015, <https://www.itu.int/rec/R-REC-BT.2020-2-201510-I/en>.

[2] ITU-T, "Quality of Experience requirements for telepresence services", ITU-T Recommendation G.1091, October 2014, <https://www.itu.int/rec/T-REC-G.1091/en>.

[3] ISO, "Information technology -- Advanced image coding and evaluation -- Part 1: Guidelines for image coding system evaluation", ISO/IEC TR 29170-1:2017, October 2017, <https://www.iso.org/standard/63637.html>.

[4] ISO, "Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding", ISO/IEC 23008-2:2015, May 2018, <https://www.iso.org/standard/67660.html>.

[5] ITU-T, "High efficiency video coding", ITU-T Recommendation H.265, November 2019, <https://www.itu.int/rec/T-REC-H.265>.

[6] Fraunhofer Institute for Telecommunications, "High Efficiency Video Coding (HEVC) reference software (HEVC Test Model also known as HM)", <https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/>.

8.2. Informative References

[7] Federal Agencies Digital Guidelines Initiative, "Term: High dynamic range imaging", <http://www.digitizationguidelines.gov/term.php?term=highdynamicrangeimaging>.

[8] Federal Agencies Digital Guidelines Initiative, "Term: Compression, visually lossless", <http://www.digitizationguidelines.gov/term.php?term=compressionvisuallylossless>.

[9] Wenger, S., "The case for scalability support in version 1 of Future Video Coding", SG 16 (Study Period 2013) Contribution 988, September 2015, <https://www.itu.int/md/T13-SG16-C-0988/en>.

[10] YouTube, "Recommended upload encoding settings", <https://support.google.com/youtube/answer/1722171?hl=en>.

[11] Yu, H., Ed., McCann, K., Ed., Cohen, R., Ed., and P. Amon, Ed., "Requirements for an extension of HEVC for coding of screen content", ISO/IEC JTC 1/SC 29/WG 11 Moving Picture Experts Group MPEG2013/N14174, San Jose, USA, January 2014, <https://mpeg.chiariglione.org/standards/mpeg-h/high-efficiency-video-coding/requirements-extension-hevc-coding-screen-content>.

[12] Parhy, M., "Game streaming requirement for Future Video Coding", ISO/IEC JTC 1/SC 29/WG 11 Moving Picture Experts Group N36771, Warsaw, Poland, June 2015.

[13] Bjontegaard, G., "Calculation of average PSNR differences between RD-curves", SG 16 VCEG-M33, April 2001, <https://www.itu.int/wftp3/av-arch/video-site/0104_Aus/>.

[14] Daede, T., Norkin, A., and I. Brailovskiy, "Video Codec Testing and Quality Measurement", Work in Progress, Internet-Draft, draft-ietf-netvc-testing-09, 31 January 2020, <https://tools.ietf.org/html/draft-ietf-netvc-testing-09>.

[15] Wang, Z., Simoncelli, E.P., and A.C. Bovik, "Multiscale structural similarity for image quality assessment", IEEE Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, DOI 10.1109/ACSSC.2003.1292216, November 2003, <https://ieeexplore.ieee.org/document/1292216>.

[16] Bossen, F., "Common HM test conditions and software reference configurations", Joint Collaborative Team on Video Coding (JCT-VC) of the ITU-T Video Coding Experts Group (ITU-T Q.6/SG 16) and ISO/IEC Moving Picture Experts Group (ISO/IEC JTC 1/SC 29/WG 11) , Document JCTVC-L1100, April 2013, <http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=7281>.

[17] ITU-R, "Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios", ITU-R Recommendation BT.601, March 2011, <https://www.itu.int/rec/R-REC-BT.601/>.

[18] ISO/IEC, "Information technology -- Coding of audio-visual objects -- Part 10: Advanced video coding", ISO/IEC DIS 14496-10, <https://www.iso.org/standard/75400.html>.

[19] ISO/IEC, "Information technology -- Coding of audio-visual objects -- Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format", ISO/IEC 14496-15, <https://www.iso.org/standard/74429.html>.

[20] ITU-R, "Parameter values for the HDTV standards for production and international programme exchange", ITU-R Recommendation BT.709, June 2015, <https://www.itu.int/rec/R-REC-BT.709>.

Acknowledgments

The authors would like to thank Mr. Paul Coverdale, Mr. Vasily Rufitskiy, and Dr. Jianle Chen for many useful discussions on this document and their help while preparing it, as well as Mr. Mo Zanaty, Dr. Minhua Zhou, Dr. Ali Begen, Mr. Thomas Daede, Mr. Adam Roach,

Dr. Thomas Davies, Mr. Jonathan Lennox, Dr. Timothy Terriberry, Mr. Peter Thatcher, Dr. Jean-Marc Valin, Mr. Roman Danyliw, Mr. Jack Moffitt, Mr. Greg Coppa, and Mr. Andrew Krupiczka for their valuable comments on different revisions of this document.

Authors' Addresses

Alexey Filippov
Huawei Technologies

Email: alexey.filippov@huawei.com

Andrey Norkin
Netflix

Email: anorkin@netflix.com

Jose Roberto Alvarez
Huawei Technologies

Email: j.alvarez@ieee.org