Network Working Group
Internet Draft
Intended status: Informational
Expires: January 6, 2016

A. Filippov
Huawei Technologies

July 6, 2015

Video Codec Requirements and Evaluation Methodology
draft-filippov-netvc-requirements-01.txt

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts can be accessed at http://datatracker.ietf.org/drafts/current/

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on January 6, 2016.

Copyright Notice

Abstract

This document provides requirements for a video codec designed mainly for use over the Internet. In addition, an evaluation methodology needed for measuring the parameters (compression efficiency, computational complexity, etc.) to ensure whether the stated requirements are fulfilled or not.

Table of Contents

1. Introduction

In this document, the requirements for a video codec designed mainly for use over the Internet are presented. The requirements encompass a wide range of applications that use data transmission over the Internet including IPTV (broadcasting over IP-based networks), peer-video conferencing, video sharing, screencasting, and video monitoring/ 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.

2. Applications

In this chapter, 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, and hence there are different types of communication channels involved (e.g., wired or wireless channels) that are characterized by different quality of service as well as bandwidth. The target platform and the channel bandwidth determine resolutions, frame-rates and quality or bit-rates for video streams to be encoded or decoded. By default, color formats YUV 4:2:0 and/or YUV 4:2:2 are assumed for the application scenarios listed below.

2.1. Internet Protocol Television (IPTV)

This is a type of pre-recorded broadcasting over IP-based networks. Typical content used in this application is movies, cartoons, series, TV shows, etc. The main requirements are as follows:

- Random access to pictures, i.e. random access period (RAP) should be kept small enough (approximately, 1-15 seconds);
- Temporal (frame-rate) scalability;
- Error robustness.

Support of resolution and quality (SNR) scalability is highly desirable. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 1.

Resolution	Frame-rate, fps	PAM
2160p (4K),3840x2160	60	RA
1080p, 1920x1080	24, 50, 60	RA
1080i, 1920x1080	30 (60 fields per second)	RA
720p, 1280x720	50, 60	RA
576p (EDTV), 720x576	25, 50	RA
576i (SDTV), 720x576	25, 30	RA

480p (EDTV), 720x480	50, 60	RA
480i (SDTV), 720x480	25, 30	RA

Table 1. IPTV: typical values of resolutions, frame-rates, and RAPs

2.2. 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 delay values should be less than 100 ms and 350 ms, respectively);
- Temporal (frame-rate) scalability;
- Error robustness.

Support of resolution and quality (SNR) scalability is highly desirable. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 2.

Resolution	Frame-rate, fps	PAM
1080p, 1920x1080	15, 30	JFPIC
720p, 1280x720	30, 60	JFPIC
4CIF, 704x576	30, 60	JFPIC
4SIF, 704x480	30, 60	JFPIC
VGA, 640x480	30, 60	JFPIC
360p, 640x360	30, 60	JFPIC
SIF, 352x240	30	JFPIC
CIF, 352x288	30	JFPIC

QVGA, 320x240	15, 30	JFPIC

Table 2. Video conferencing: typical values of resolutions, frame-rates, and RAPs

2.3. Video sharing

This is a service that allows people to upload and share video data (using live streaming or not) and to watch them. It is also known as video hosting. A typical scenario for this application is to capture video using mobile cameras such as GoPro 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;
- Resolution and quality (SNR) scalability;
- Error robustness.

For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 3.

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
240p, 426x240	24, 25, 30, 48, 50, 60	RA
144p, 256x144	24, 25, 30, 48, 50, 60	RA

Table 3. Video sharing: typical values of resolutions, frame-rates [8, 9], and RAPs

2.4. Screencasting

This is a service that allows users to record and distribute computer screen output. This service requires efficient compression of computer-generated content with high visual quality (up to visually and mathematically lossless) [1]. Currently, this application includes animation (cartoons, gaming content, data visualization, i.e. such a type of content that is characterized by fast motion, rotation, smooth shade, 3D effect, highly saturated colors with full resolution, clear textures and sharp edges with distinct colors [1]), 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 a wide range of video formats including RGB and YUV 4:4:4 in addition to YUV 4:2:0 and YUV 4:2:2 [1]. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 4.

Resolution	Frame-rate, fps	PAM
Input color format: RGB		
WQXGA, 2560x1600	15, 30, 60	AI, RA, JFPIC
WUXGA, 1920x1200	15, 30, 60	AI, RA, JFPIC
WSXGA+, 1680x1050	15, 30, 60	AI, RA, JFPIC
WXGA, 1280x800	15, 30, 60	AI, RA, JFPIC
XGA, 1024x768	15, 30, 60	AI, RA, JFPIC
SVGA, 800x600	15, 30, 60	AI, RA, JFPIC
VGA, 640x480	15, 30, 60	AI, RA, JFPIC
Input color format: YUV 4:4:4		
1440p (2K), 2560x1440	15, 30, 60	AI, RA, JFPIC

1080p, 1920x1080	15, 30, 60	AI, RA, JFPIC
720p, 1280x720	15, 30, 60	AI, RA, JFPIC

Table 4. Screencasting for RGB and YUV 4:4:4 format: typical values of resolutions, frame-rates, and RAPs

2.5. Game streaming

This is a service that provides game content over the Internet to different local devices such as notebooks, gaming tablets, etc. In this category of applications, server renders 3D games in cloud server, and streams the game to any device with a wired or wireless broadband connection [2]. This allows anyone to play (or resume) full featured games from anywhere in the Internet [2]. An example of this application is Nvidia Grid [2]. Another category application is broadcast of video games played by people over the Internet in real time or for later viewing [2]. There are many companies such as Twitch, YY in China enable game broadcasting [2]. Games typically contain a lot of sharp edges and large motion [2]. The main requirements are as follows:

- Random access to pictures for downloaded video data;
- Temporal (frame-rate) scalability;
- Error robustness.

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

2.6. Video monitoring / 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 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, respectively. 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 high dynamic range as well as resolution and quality (SNR) scalability is an essential requirement

for video surveillance. For this application, typical values of resolutions, frame-rates, and RAPs are presented in Table 5.

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

Table 5. Video monitoring / surveillance: typical values of resolutions, frame-rates, and RAPs

3. Requirements

The most basic requirement is coding efficiency, i.e. compression performance. It should be better than for state-of-the-art video codecs such as HEVC/H.265 and VP9. Levels to be supported by the new codec are presented in Table 6.

Level	Example picture resolution at highest frame rate
1	128x96@30.0 176x144@15.0
2	176x144@100.0 352x288@30.0
3	352x288@60.0 640x360@30.0
4	640x360@60.0 960x540@30.0
5	720x576@75.0 960x540@60.0 1280x720@30.0

	1,280x720@68.0
6	2,048x1,080@30.0
	1,280x720@120.0
7	2,048x1,080@60.0
	1,920x1,080@120.0
8	3,840x2,160@30.0
	4,096x2,160@30.0
	1,920x1,080@250.0
9	4,096x2,160@60.0
	1,920x1,080@300.0
10	4,096x2,160@120.0
	3,840x2,160@120.0
11	8,192x4,320@30.0
	3,840x2,160@250.0
12	8,192x4,320@60.0
	3,840x2,160@300.0
13	8,192x4,320@120.0

Table 6. Codec levels

3.1. Basic requirements

3.1.1. Input source formats:

- Bit depth: 8- and 10-bits per color component;
- Color sampling formats: YUV 4:2:0, YUV 4:2:2

3.1.2. Coding delay:

- Support of “low-delay” configurations (delay should be up to 350 ms but its preferable value should be less than 100 ms)

3.1.3. 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

3.1.4. Scalability:

- Temporal (frame-rate) scalability

3.1.5. Error resilience:

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

3.2. Optional requirements

3.2.1. Input source formats

- Bit depth: up to 16-bits per color component;
- Color sampling formats: YUV 4:4:4 and RGB;
- Auxiliary channel (e.g., alpha channel) support;
- Support of high dynamic range

3.2.2. Scalability:

- Resolution and quality (SNR) scalability;
- Computational complexity scalability, i.e. computational complexity is decreasing along with degrading picture quality

3.2.3. Complexity:

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

- 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

4. Evaluation methodology

4.1. Compression performance evaluation

As shown in Fig.1, compression performance testing is performed in 3 ranges :

- Low bit-rate range (LBR) is the range that contains the 4 lowest bit-rates of the 6 specified bit-rates;
- Medium bit-rate range (MBR) is the range that contains the 4 medium bit-rates of the 6 specified bit-rates;
- High bit-rate range (HBR) is the range that contains the 4 highest bit-rates of the 6 specified bit-rates.

To avoid any rate control mechanisms that can significantly impact compression performance, the deviation between bit-rates of reference and tested codecs should be less than the threshold value that should be defined in a separate document. This deviation is calculated as follows:

$$D = abs((BRr - BRt)/ BRr)*100\%,$$

where BRr and BRt are bit-rates of reference and tested codecs.

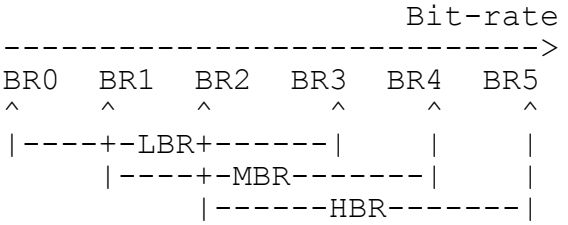

Figure 1 Bit-rate ranges for the CBR mode

To assess the quality of output (decoded) sequences, two indexes, PSNR [3] and MS-SSIM [3,11], should be separately calculated for each color plane. For obtaining an integral estimation, BD-rate [4] should be computed for each range and each quality index. Finally, 18 values should be obtained for a color format, which contains 3 color planes (e.g., for YUV or RGB). A list of video sequences that should be used for testing as well as the 6 values of bit-rates are defined in a separate document. It should use the information on the codec applications presented in this document. As the reference for evaluation, the HEVC/H.265 codec [5,6] must be used. The reference source code of the HEVC/H.265 codec can be found at [7]. The HEVC/H.265 codec must be configured according to [8] and Table 9.

Intra-period, second	HEVC/H.265 encoding mode according to [7]
AI	Intra Main or Intra Main10
RA	Random access Main or Random access Main10
JFPIC	Low delay Main or Low delay Main10

Table 9. Intra-periods for different HEVC/H.265 encoding modes according to [8]

In addition to the objective quality measures defined above, subjective evaluation must also be performed before adopting any new tool and a final codec standard. 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.

5. Security Considerations

This document itself does not have any security considerations. However, it is worth noting that a codec implementation (for both an encoder and a decoder) should cover the worst case of computational complexity, memory bandwidth, and physical memory size (e.g., for decoded pictures used as references). Otherwise, it can be considered as a security vulnerability and lead to denial-of-service (DoS) in the case of attacks.

6. Conclusions

In this document, an overview of Internet video codec applications and typical use cases as well as a prioritized list of requirements for an Internet video codec are presented. An evaluation methodology for this codec is also proposed.

7. References

7.1. Normative References

[1] H. Yu, K. McCann, R. Cohen, and P. Amon, "Requirements for future extensions of HEVC in coding screen content", ISO/IEC JTC1/SC29/WG11 MPEG2013/N14174, San Jose, USA, Jan. 2014

[2] Manindra Parhy, "Game streaming requirement for Future Video Coding," MPEG Contribution m36771, June 2015, Warsaw, Poland.

[3] ISO/IEC PDTR 29170-1: Information technology -- Advanced image coding and evaluation methodologies -- Part 1: Guidelines for codec evaluation.

[4] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves (VCEG-M33)," in VCEG Meeting (ITU-T SG16 Q.6), Austin, Texas, USA, Apr. 2-4 2001.

[5] ISO/IEC 23008-2:2015. Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding

[6] Recommendation ITU-T H.265: High efficiency video coding, 2013.

[7] https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

[8] F. Bossen, "Common test conditions and software reference configurations," JCTVC-L1100, Geneva, Switzerland, Jan. 2013.

7.2. Informative References

[9] "Recommended upload encoding settings (Advanced)" https://support.google.com/youtube/answer/1722171?hl=en

[10] "YouTube introduces 144p resolution on some videos" http://www.youtube.com/watch?v=2ZMRkGbrB3c

[11] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," Invited Paper, IEEE Asilomar Conference on Signals, Systems and Computers, Nov. 2003, Vol. 2, pp. 1398-1402.

[12] http://www.digitizationguidelines.gov/term.php?term=compressionvisuallylossless)

8. Acknowledgments

The author would like to thank Jiantong Zhou, Paul Coverdale, Vasily Rufitskiy, Dr. Haitao Yang, Viktor Stepin, Maxim Sychev, and Sergey Ikonin for many useful discussions on this document and their help while preparing it as well as Mr. Mo Zanaty, Mr. Thomas Daede, and Mr. Greg Coppa for their valuable comments on the first revision of this document.

This document was prepared using 2-Word-v2.0.template.dot.

Appendix A. Abbreviations used in the text of this document

Abbreviation	Meaning
AI	All-Intra (each picture is intra-coded)
BD-Rate	Bjontegaard Delta Rate
GOP	Group of Picture
HBR	High Bit-rate Range
PAM	Picture Access Mode
RA	Random Access
RAP	Random Access Period
IPTV	Internet Protocol Television
JFPIC	Just the First Picture is Intra-Coded
LBR	Low Bit-rate Range
MBR	Medium Bit-rate Range
MOS	Mean Opinion Score
MS-SSIM	Multi-Scale Structural Similarity quality index
PSNR	Peak Signal-to-Noise Ratio

Appendix B. Used terms

Term	Meaning
Random access period	is the period of time between two closest independently decodable frames (pictures).
Visually lossless compression	is 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 appearing identical to the uncompressed data [12].

Authors' Addresses

Alexey Filippov
Huawei Technologies

Email: alexey.filippov@huawei.com