A Video Output Method for H.265/HEVC Video and Audio IP Transmission and Its QoE

Toshiro Nunome
Department of Computer Science, Graduate School of Engineering,
Nagoya Institute of Technology, Nagoya 466–8555, Japan
Email: nunome@nitech.ac.jp

Abstract—**This paper exploits the idea of SCS (Switching between error Concealment and frame Skipping) for H.265/HEVC video and audio IP transmission. The method skips output of video frames with a larger ratio of lost slices than a predefined threshold. It continues frame skipping until new GOP. To assess the effectiveness of the method, we evaluate QoE through a subjective experiment. We employ two audiovisual contents and vary the number of slices in a video frame, GOP length, encoding bitrate, and the amount of background traffic. As a result, we find that the method exploiting the idea of SCS can enhance QoE for the content which has large movement.**

Index Terms—**H.265, audio and video IP transmission, frame skipping, subjective assessment**

I. Introduction

High definition video such as 4K and 8K has been attracting people's interests. Transmission of the high definition video requires efficient encoding techniques. Hence, H.265/HEVC (High Efficiency Video Coding) has been standardized.

In order to handle high definition video, H.265/HEVC advances technologies employed in H.264/AVC (Advanced Video Coding). Meanwhile, H.265/HEVC omits various error concealment mechanisms in H.264/AVC owing to requirements for high-speed encoding and decoding.

When we transmit H.265/HEVC over IP networks, we face quality degradation due to packet losses as in H.264/AVC transmission. The major open source encoder x265 [1] and decoder FFmpeg [2] have not implemented error concealment mechanisms for H.265/HEVC. Thus, when the packet loss occurs, in intra-coded slices, the missing area cannot be decoded. In inter-coded slices, substantial degradation of an image occurs because of incorrect moving vector information.

There are few studies on *QoE (Quality of Experience)* [3] of H.265/HEVC under network impairment, while many studies treat coding impairment. For example, References [4] and [5] compares the coding performance of several video coding standards include H.265/HEVC without network impairment.

Besides, Nightingale *et al.* have evaluated the effect of packet loss on subjective QoE in H.265/HEVC RTP (Real-time Transport Protocol)/UDP transmission [6],[7]. They have implemented two error concealment mechanisms to the HEVC reference software HM 8.0. On the other hand, in this paper, we deal with another approach for mitigating the effect of packet loss.

For H.264/AVC video and audio transmission, a QoE-based video output scheme *SCS (Switching between error Concealment and frame Skipping)* has been proposed [8]. SCS switches two video output schemes: error concealment and frame skipping. Video error concealment interpolates lost video slices due to packet drop with other information of the video stream. However, the spatial quality of the error-concealed video degrades compared to the original one since the scheme cannot perfectly interpolate the lost information. In addition, there is a problem that the degradation propagates to the succeeding frames in GOP (Group of Pictures). On the other hand, frame skipping does not output video frames which include lost slices. The scheme keeps the spatial quality of the output video original, while it degrades the temporal quality because of skipped frames. SCS defines the error concealment ratio R_c [%] as the ratio of the number of lost video slices to the total number of slices in a frame and introduces a threshold value T_h [%]. When R_c is larger than T_h in a video frame, the video frame is skipped, i.e., frame skipping. When an I (Intra) frame with $R_c \leq T_h$ comes out, the output scheme is switched to error concealment. An appropriate selection of the T_h value maximizes QoE.

Even when we cannot apply error concealment on H.265/HEVC transmission, the idea of SCS, i.e., switching to frame skipping instead of output damaged frames, can be effective on QoE enhancement.

In this paper, we enhance the idea of SCS for H.265/HEVC IP transmission. When the ratio of dropped slices exceeds a threshold, the receiver skips the rest of GOP. We assess QoE by a subjective experiment to evaluate the effectiveness. We also measure objective quality metrics of temporal quality (smoothness of output) and spatial quality (image quality). As the temporal quality measure, we employ the video frame loss ratio. We utilize PSNR (Peak Signal-to-Noise Ratio) for the spatial quality assessment. Then, we investigate the factors affecting QoE.

We organize the remainder of this paper as follows. Section II introduces the slice structure on H.265/HEVC and our output method. Sections III and IV describe the experimental method and the QoE assessment method, respectively. Section V presents experimental results. Section VI concludes this paper.

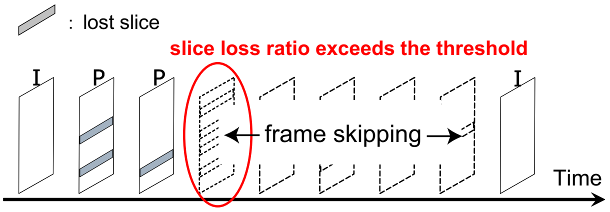

Fig. 1. Example of operation of our output method

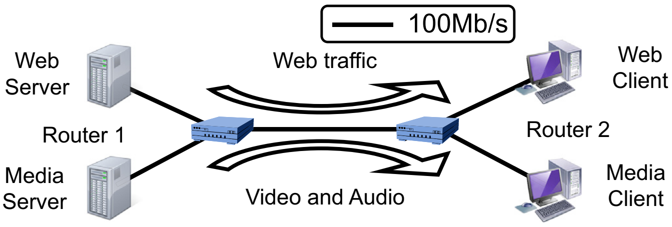

Fig. 2. Experimental network

II. Slices on H.265/HEVC and our Method

A slice of H.264/AVC consists of MBs (Macro Blocks) of 16 × 16 pixels information. On the other hand, the minimum unit consisting H.265/HEVC slices is a CTU (Coding Tree Unit), which size is from 16 × 16 to 64 × 64 pixels. In particular for high definition video, the CTU of 64 × 64 pixels can be employed for efficient coding. Thus, the number of pixels in a slice becomes large.

The CTU is recursively divided into CUs (Coding Units). Then, the CU is classified into a PU (Prediction Unit) or a TU (Transform Unit). This is for efficient encoding and decoding of high definition video. A region with small movement can be encoded with a large block, and a small block will be utilized for a region with large movement. On the other hand, the adjustable control makes error concealment difficult. The users can notice incongruity against simple interpolation techniques.

In this paper, we make a threshold as SCS has. The slice loss ratio is defined as the ratio of the lost slices to all the slices of a video frame. The receiver stops the output of video frames with the higher video slice loss ratio than the threshold. The frame skipping lasts the next intra-coded frame with the lower slice loss ratio than the threshold. Figure 1 shows an example of the operation.

III. Methodology of Experiment

Figure 2 shows the experimental system. All the links in the network are 100 Mb/s full-duplex Ethernet. Media Server transmits video and audio streams to Media Client through MMTP (MPEG Media Transport Protocol); it is an application-level protocol for multimedia transmission [9]. UDP is employed as the transport protocol under MMTP. For audio, each MMTP/UDP packet includes an MU (Media Unit), which is an information unit for media synchronization control. Each video MMTP/UDP packet consists of a video slice.

As the interference traffic of audio and video, Web Server transmits Web traffic to Web Client according to requests generated by WebStone 2.5 [10], which is a Web server benchmark tool. For the number of client processes, we employ 10 and 20.

We employ H.265/HEVC video and AAC-LC (Advanced Audio Coding-Low Complexity) CBR (Constant BitRate) stereo audio. Table I shows the specifications. The video encoding bitrate is set to about 3 Mb/s or 6 Mb/s. We utilize x265 ver. 2.1 as a video encoder. To stabilize the encoding bitrate, we use 2-pass encoding. We consider a video frame as a video MU. The MU rate is 29.97 MU/s. We deal with the two picture patterns: IPPPP (I+4P's) and IPPPPPPPPPPPPPP (I+14P's). The number of slices per picture frame is 4 or 16. The average bitrate and the MU rate of audio are 128 kb/s and 46.875 MU/s, respectively. We employ two contents: *drama* (a scene of historical drama) and *sport* (a scene of figureskating). The TI (Temporal perceptual Information) value except for scene changes of drama is 9.786, and that of sport is 36.415; the TI value indicates the amount of temporal changes of a video sequence [11]. Thus, in this experiment, sport has larger movement than drama.

TABLE I
Specifications of Video and Audio

	video	audio
coding method	H.265/HEVC	MPEG-4 AAC-LC CBR
encoder	x265	qaac
image size [pixel]	1920 × 1080	-
number of slices per frame	4, 16	-
picture pattern	IPPPP, I+14P's	-
sampling rate [kHz]	-	48
channels	-	2
encoding bitrate [kb/s]	3000, 6000	128
average MU rate [MU/s]	29.97	46.875
playout buffering time [ms]	500	

Media Client outputs received audio and video after the playout buffering control. We set the playout buffering time to 500 ms. As the threshold value for the slice loss ratio to switch frame skipping, we employ 0% (pure frame skipping), 25%, and 100% (no frame skipping due to slice losses).

IV. QoE Assessment Method

In this paper, we assess QoE of the audio-video stream by a subjective experiment. It was conducted as follows.

We first made test samples for subjective assessment by actually outputting the audio and video MUs with the output timing obtained from the experiment. The test samples are called *stimuli*. Each stimulus lasted 10 seconds and was obtained by outputting the audio-video stream of the first 10 seconds in each experimental run. We set the duration of the stimulus owing to the assessors' burden.

TABLE II
FIVE CATEGORIES OF IMPAIRMENT

score	category
5	imperceptible
4	perceptible but not annoying
3	slightly annoying
2	annoying
1	very annoying

We put the stimuli in a random order and presented them to 19 assessors. They are 17 male students, a male faculty member, and a female faculty member. ITU-T Rec. P.911 describes that at least 15 subjects should participate in the experiment [11]. Thus, we employed the 19 assessors. On the other hand, we need to evaluate with more assessors; it is a future study issue. The total assessment time for an assessor is about 30 minutes.

A subjective score was measured by the *rating-scale method*, in which assessors classify each stimulus into one of a certain number of categories. We adopted the *five categories of impairment* as shown in Table II. We regard the integer value as a subjective score. We then calculate *MOS (Mean Opinion Score)* as the quantitative measure of perceptual quality.

V. EXPERIMENTAL RESULTS

A. Application-level QoS

In this paper, we treat the video MU loss ratio, the slice loss ratio of output video, and the PSNR as the application-level QoS parameters. The slice loss ratio of output video represents the percentage of lost slices in output video frames; the skipped frames are discarded in the calculation. The MU loss ratio is the ratio of the number of MUs not output at the recipient to the number of MUs transmitted by the sender.

We depict the audio MU loss ratio and the video MU loss ratio in Figs. 3 and 4, respectively. Figure 5 shows the slice loss ratio of output video. Figure 6 represents the PSNR of video luminance. Figures 4 through 6 are the results for drama with video encoding bitrate 6 Mb/s. In addition, we present the PSNR for drama with video encoding bitrate 3 Mb/s and that for sport with video encoding bitrate 6 Mb/s in Figs. 7 and 8, respectively.

In Fig. 3, we show the result of audio quality. The assessment of audio is a feature of our study; References [6] and [7] do not consider the effect of audio. We find in this figure that the audio MU loss ratio is smaller than 1 % even when the number of Web clients is 20. Thus, the audio quality does not degrade largely in this experiment. Furthermore, the differences among the threshold values are small.

We see in Fig. 4 that as the GOP length increases, the video MU loss ratio increases. This is because skipped frames due to the loss of reference frames increase when we employ the long GOP.

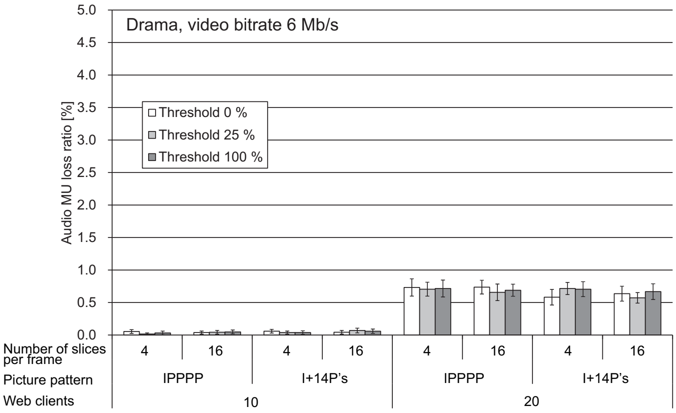

Fig. 3. Audio MU loss ratio (drama, video encoding bitrate 6 Mb/s)

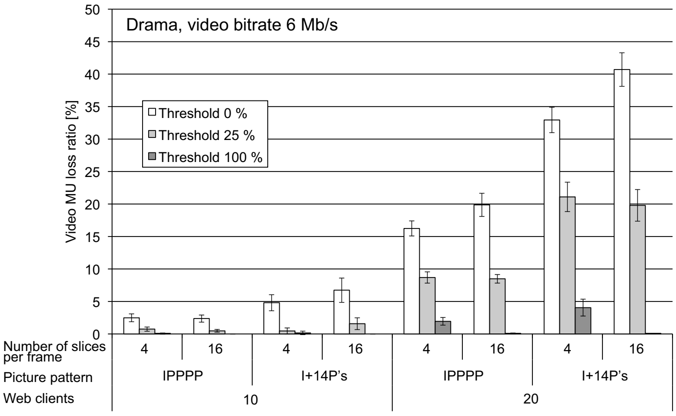

Fig. 4. Video MU loss ratio (drama, video encoding bitrate 6 Mb/s)

We also notice in Fig. 4 that the MU loss ratio increases as the slices per frame increase with the 0 % method when the number of Web clients is 20. On the other hand, with the larger threshold values than 0 %, the MU loss ratio decreases as the number of slices increases. The 0 % method performs frame skipping when a slice drops. As the number of slices increases, the overhead increases. In addition, the number of slices affects the efficiency of predictive coding methods. The size of I picture for 16 slices per frame is larger than that for 4 slices. Thus, the frame skipping occurs more frequently for 16 slices per frame in the 0 % method.

In Fig. 5, the slice loss ratio in the 0 % method is always 0 because the method does not output damaged frames. Besides, the GOP patterns scarcely affect the slice loss ratio.

We can observe in Fig. 5 that the slice loss ratio of output video becomes small as the number of slices per frame increases. This is because the ratio of damaged slices becomes small as the slice size decreases.

In Figs. 6 through 8, we notice that the PSNR values become small as the threshold and the GOP length increase. This is because the image quality degradation due to lost slices propagates. In addition, the number of slices per frame slightly reduces the PSNR values. As we find in Figs. 4 and

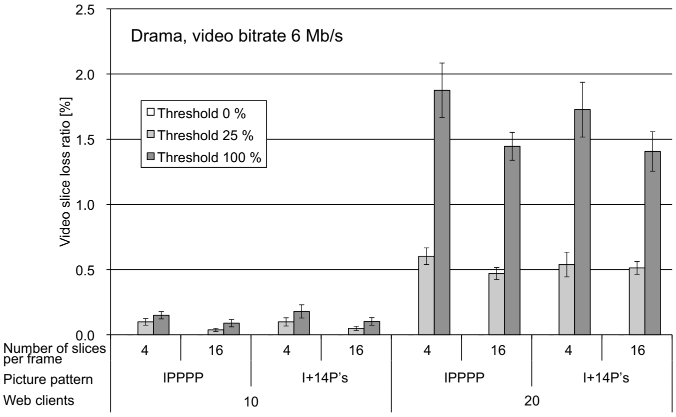

Fig. 5. Slice loss ratio of output video (drama, video encoding bitrate 6 Mb/s)

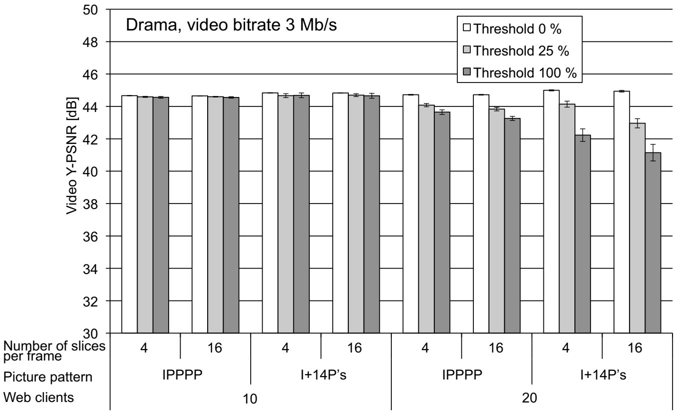

Fig. 7. Video PSNR (drama, video encoding bitrate 3 Mb/s)

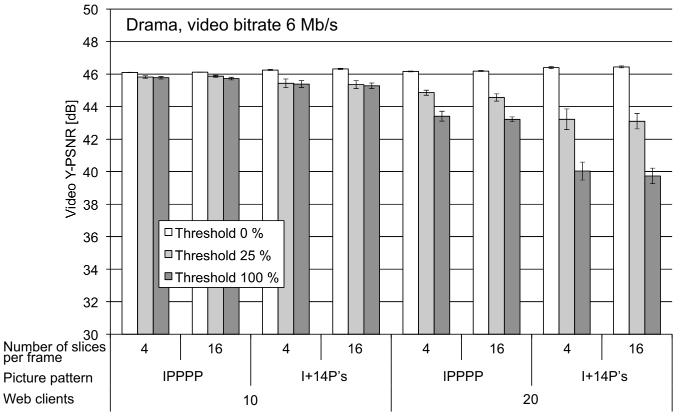

Fig. 6. Video PSNR (drama, video encoding bitrate 6 Mb/s)

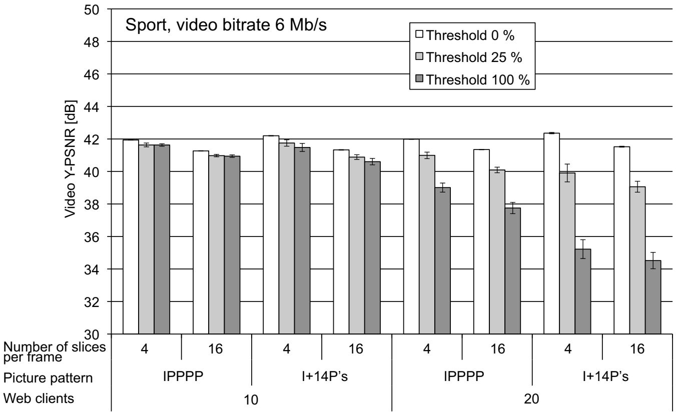

Fig. 8. Video PSNR (sport, video encoding bitrate 6 Mb/s)

5, the damaged MUs affect the image quality of output video although the output MUs increase.

In addition, we see in Figs. 6 and 8 that sport has the small PSNR values than drama. This is owing to the characteristics of the contents.

B. QoE

We present QoE assessment results in Figs. 9 through 12. Figures 9 and 10 are the results for drama, while Figs. 11 and 12 are for sport.

In Fig. 9, we find that for the number of Web clients 10, we can achieve good MOS values around 5 irrespective of the threshold values. Under the lightly loaded condition with the smaller encoding bitrate, the number of lost packets is not many, and then the output of damaged MUs merely affects users perception.

We also see in Fig. 9 that the MOS value decreases as the number of slices per frame increases when the number of Web clients is 20. This is owing to the degradation of encoding efficiency in many slices per frame.

In Figs. 9 and 10, we notice that the MOS value decreases as the GOP length increases for the number of Web clients 20. This is owing to the error propagation of predictive coding.

We also see in Fig. 10 that the MOS values in the 25 % method and the 100 % method increase as the number of video slices per frame increases. This is because we can restrict the effect of the slice loss in the small region.

We find in Fig. 11 that the 25 % method has larger MOS value than the 0% and 100% methods for the long GOP when the number of Web clients 20. We also see the same tendency in Fig. 12. Thus, the idea of SCS can enhance QoE in the H.265/HEVC video transmission.

VI. Conclusions

This paper assessed the effect of video output methods on QoE in H.265/HEVC video and audio transmission in order to investigate the tradeoff relationships between temporal quality and spatial quality. In the H.265/HEVC transmission, the effect of a lost slice becomes larger than that in the H.264/AVC because of treating high definition video; the slice size in pixels can increase in H.265/HEVC. In addition,

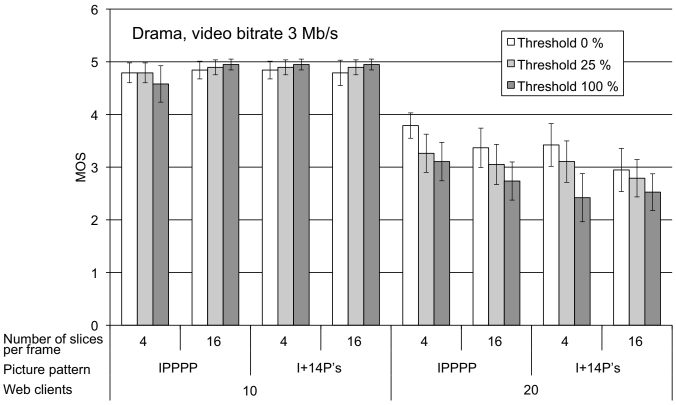

Fig. 9. MOS (drama, video encoding bitrate 3 Mb/s)

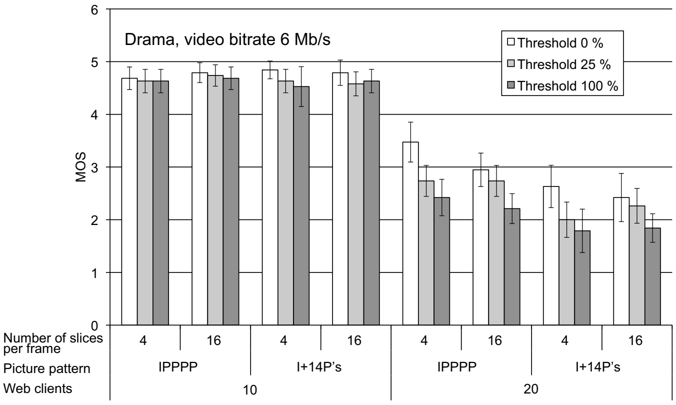

Fig. 10. MOS (drama, video encoding bitrate 6 Mb/s)

H.265/HEVC omits some error concealment techniques in H.264/AVC. However, we found in the results that the idea of SCS can enhance QoE in the content with large movement.

In future work, we need to investigate the appropriate output methods for H.265/HEVC. We will also devise a QoE enhancement method exploiting the functions in MMTP. The evaluation with more diverse contents and assessors are needed.

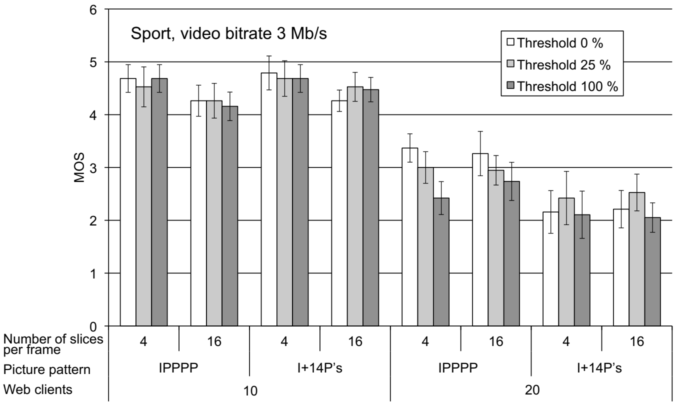

Fig. 11. MOS (sport, video encoding bitrate 3 Mb/s)

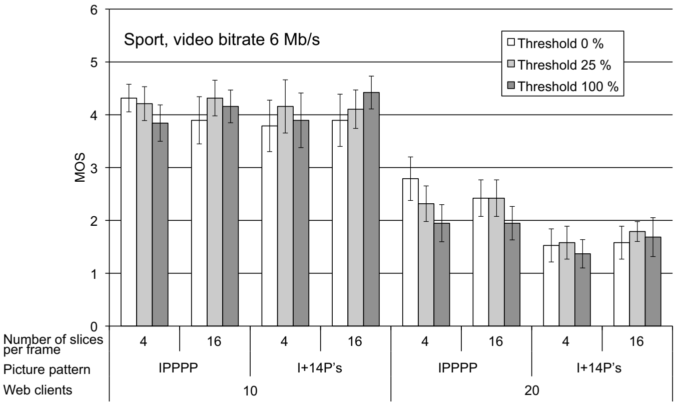

Fig. 12. MOS (sport, video encoding bitrate 6 Mb/s)

ACKNOWLEDGMENT

We thank Mr. Yuya Kumazaki for his support for the experiment.

REFERENCES

[1] "x265 HEVC Encoder / H.265 Video Codec," *http://x265.org*.

[2] "FFmpeg, " *http://www.ffmpeg.org/*

[3] ITU-T Rec. P.10/G.100, "Amendment 5: New definitions for inclusion in Recommendation," July 2016.

[4] V. Bajčinovci, M. Vranješ, D. Babić, B. Kovačević, "Subjective and objective quality assessment of MPEG-2, H.264 and H.265 videos," *Proc. 59th International Symposium ELMAR-2017*, pp. 73-77, Sept. 2017.

[5] M. A. Layek, N. Q. Thai, M. A. Hossain, N. T. Thu, L. P. Tuyen, A. Talukder, T. Chung, and E. -N. Huh, "Performance analysis of H.264, H.265, VP9 and AV1 video encoders," *Proc. APNOMS 2017*, pp. 322-325, Sept. 2017.

[6] J. Nightingale, Q. Wang, C. Grecos, and S. Goma, "Subjective evaluation of the effects of packet loss on HEVC encoded video streams," *Proc. IEEE ICCE-Berlin 2013*, pp. 358-359, Sept. 2013.

[7] J. Nightingale, Q. Wang, C. Grecos, and S. Goma, "The impact of network impairment on quality of experience (QoE) in H.265/HEVC video streaming," *IEEE Trans. on Consumer Electronics*, vol. 60, no. 2, pp. 242-250, May 2014.

[8] S. Tasaka, H. Yoshimi, A. Hirashima and T. Nunome, "The effectiveness of a QoE-based video output scheme for audio-video IP transmission," *Proc. ACM Multimedia2008*, pp. 259-268, Oct. 2008.

[9] ISO/IEC 23008-1, "Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 1: MPEG media transport (MMT)," Second edition, Aug. 2017.

[10] Mindcraft Inc, "WebStone benchmark information," *http://www.mindcraft.com/webstone/*.

[11] ITU-T Rec. P.911, "Subjective audiovisual quality assessment methods for multimedia applications," Dec. 1998.