Cognitive-effect-based bit rate control to improve quality of experience for video streaming

Sayaka NISHIDE[†a)], Daichi KOMINAMI[††], Satoshi NISHIMURA[†], Tatsuya OTOSHI[††], Masaaki KUROZUMI[†]
Daiki FUKUDOME[†], Masao YAMAMOTO[†], and Masayuki MURATA[††]

SUMMARY In this study, we focus on the characteristics of the human cognitive bias and develop a method that does not reduce the quality of experience (QoE) of the viewer, even if the network conditions deteriorate during a video streaming playback. This method makes it difficult for the viewers to perceive any deterioration in the video quality by suppressing it through the implementation of a sudden decrease in the available bandwidth. To this end, we added a new stream with a gradual decrease in image quality to streams that can be selected for streaming distribution. As a result of a subjective evaluation experiment, it was found that the decrease in QoE was suppressed even when the transmission amount was suppressed.
keywords: *Adaptive bit rate streaming, QoE, MPEG-DASH, cognitive bias*

1. Introduction

In online video streaming distribution, adaptive streaming methods such as MPEG-DASH, which can change the bit rate according to the available bandwidth, have recently become the mainstream [1].

It is important to ensure the quality of experience (QoE) for viewers when video quality deteriorates because of the reduction in the available bandwidth. Here, QoE refers to the quality perceived by the user, which is a subjective index. Several studies have been conducted with the aim to enable the user to experience the highest QoE in a limited network bandwidth [2].

The QoE of the user is affected not only by the measurable quality of the streaming video but also by various biases that occur in the process of human cognition. For example, the video quality perceived by the viewer may significantly deteriorate during changing from a high bit rate stream to a low bit rate stream, and the difference between the bit rates before and after the change is large. In the process of cognition, one may make judgments that are contrary to statistical rationality. In the field of cognitive science, such a bias is called “cognitive bias.” Research on cognitive bias in video distribution is also underway[3][4], for example, Sackl A et al.[5] show that even if the video content is of the same quality, the QoE will differ depending on whether the user chooses the image quality himself. At this point, the user convinces himself that the image quality he chooses is better than the image quality he chooses automatically. This is due to one of the cognitive biases called cognitive dissonance, which is a bias that avoids contradictions in the brain when there is a contradiction in what one perceives. The presence of this bias indicates the need to consider the effects of cognitive bias on QoE.

Another example of cognitive bias is “anchor bias,” which influences subsequent decision-making by considering biases on the numbers and prior information. Considering this bias, when the image quality deteriorates, the larger the change is, the lower is the QoE of the user. Mok et al. found that users prefer mid-level footage in the event of quality degradation [6].

To reduce the impact of the changes in image quality, a large number of streams can be prepared with different bit rates and gradually change the streams from a high bit rate to a low bit rate. However, this increases the server load associated with creating and storing the stream.

In this study, we focus on cognitive bias, especially the anchoring bias, and propose a rate control method, in which sudden drops in video quality (due to throughput drops) can be suppressed by changing the bit rate gradually during streaming playback. The QoE of the viewer can thus be improved. On this basis, an experimental system was constructed, and a QoE evaluation experiment was performed on the subjects.

2. Methods

This section describes the rate control method, focusing on the anchoring bias. The current mainstream adaptive bit rate control technology determines the appropriate video bit rate for the current bandwidth from several prepared bit rates. Using this conventional method, the QoE of the viewer sharply decreases at the boundary of the segments when the bit rate decreases. Therefore, we propose a method to gradually reduce the quality in less than one segment (Fig. 1). This method is realized by inserting a “transition segment” between the high-bit rate segment and the low-bit rate segment while changing to a low-bit rate segment.

[†]The authors are with Japan Broadcasting Corporation (NHK), Kinuta, Setagaya-ku, Tokyo, 157-8510 Japan
[††]The authors are with Osaka University, Suita, Osaka, 565-0871 Japan.
[a)] E-mail: nishide.s-gs@nhk.or.jp

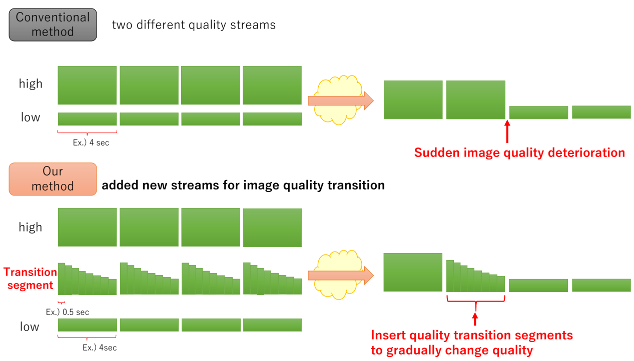


Fig. 1 Method to realize gradual video quality deterioration

3. Experiment

3.1 Experiment outline

A subjective evaluation experiment was conducted to compare the proposed method with conventional methods. Three types of streams were used in this experiment. A “high-rate” stream with 6 Mbps, a “low-rate” stream with 1.4 Mbps, and a "transition-segment" stream. The transition segment is composed of ten 0.5-s segments that have a bit rate of 0.85 times the previous bit rate (Fig. 2). It was inserted between the high-rate segment and the low-rate segment so that it would gradually deteriorate in 5 s. The resolution of the videos was 1920 × 1080. The flame rate of these videos was 29.97 fps, and both videos were encoded by H.264.

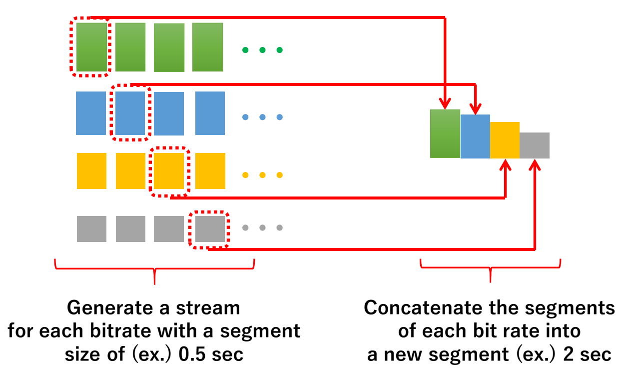


Fig. 2 Approaches to create a quality transition stream

The results were aggregated by a two-choice questionnaire method to determine which of the *A* and *B* with different quality degradation methods achieved better impression. *A* is a pattern using the proposed method, in which "transition segment" is played for the first 5 s, and then the low-rate stream is played for 10 s. *B* is a pattern using the conventional method, in which the high-rate stream is played for the first 5 s and then the low-rate stream is played for 10 s. The subjects of the experiment were 21 participants (15 male, 6 female) from Osaka University students in their teens to 20 s, who were gathered by open recruitment. After receiving the explanation of the experiment contents and oral explanation from the person in charge of the experiment, the participants submitted the consent form for participation in the experiment and the use of the experimental data. In addition, the experiment was conducted after obtaining a research ethics approval from the Graduate School of Information Science and Technology, Osaka University (ethics approval number 202008). The experimental environment was set based on ITU-R BT.500, and the size of the display was 27 inches.

The procedure of the experiment was as follows. Two types, i.e., *A* and *B* mentioned above, of 15 s with the same content were played, and there was a 2-s break between the two. Then, the type with the best impression was selected during an 8-s interval, denoted as one period. A break of 1 min was taken for every eight periods. The sequence was repeated for 90 periods (Fig. 3). The order of *A* and *B* was randomized, and the genres of the scenes used were not biased.

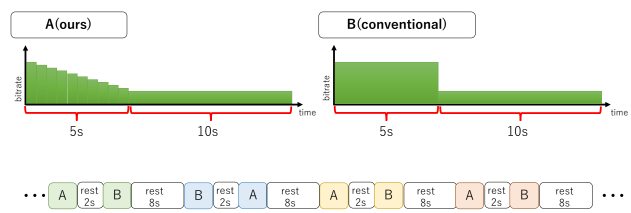


Fig. 3 Experimental design
(the order of *A* and *B* is random)

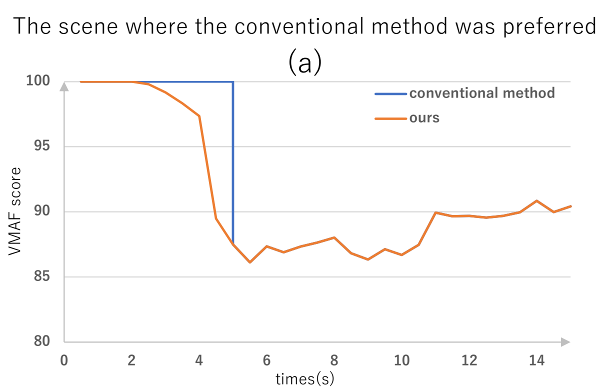
3.2 Results

The results of the questionnaire after the experiment showed no significant differences for most subjects. As for the details, there were ten participants who liked the conventional method, eight participants who liked the proposal method, and three participants who liked both methods. There were 41 scenes that were preferred to be viewed with the proposed method and 49 scenes that were preferred to be viewed with the conventional method.

3.3 Discussion

On a large scale, there was no significant difference between the experimental results of the proposed method and those of the conventional method. Upon comparing the video multi-method assessment fusion (VMAF) [7] scores between the same scenes of patterns *A* and *B*, no significant features were obtained to compare the video quality by itself. The VMAF scores are an image quality evaluation index that is also known to have a high correlation with the subjective evaluation value by human perception of multiple videos with different aggregation results. Figure 4 shows a graph of the VMAF scores for the streams with these methods relative to the high-rate streams of the scene for which the conventional method was the most popular and the scene for which the proposed method was the most popular among the programs featuring nature, for which the difference in the aggregated results for each video was large. Figure 4 shows that the VMAF score of conventional method is constant until 5 s after the start and drops sharply after that point, while the VMAF score of ours gradually decreases. After 5 s, both equations show the same value. Further, comparing the data size of the first 5 s of video, for example the scene in Fig.4b, the data size when using the proposed method is approximately 52% smaller than that when using the conventional method.

On the basis of this result, it is considered that in case of the proposed method, the decrease in the QoE is suppressed even if the transmission capacity is reduced. In addition, while comparing the scenes of the same genre with relatively large differences in results, a scene with relatively few features (e.g., medium shot) is preferred to be viewed using the proposed method. Contrarily, scenes such as close-ups and bird's-eye views where changes in quality are easily noticed are preferred to be viewed using the conventional method. This tendency may be utilized for the distribution control, depending on the future research.



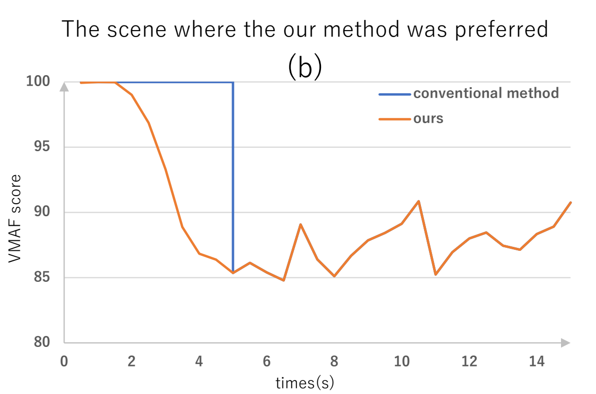


Fig. 4 Comparison of changes in the VMAF score over time

4. Conclusions

In this paper, we propose a bit rate control method that suppresses the decline in the QoE. This prevents a sudden decrease in the video quality caused by a decrease in the usable bandwidth due to network congestion when the conventional method is used. This was achieved by the proposed method, which inserts segments to gradually reduce the video quality. In the future, we plan to determine the appropriate experimental parameters and verify the effects accurately.

References

[1] Hirabayashi M, "Next-generation video distribution technology 'MPEG-DASH' Technology overview and standardization/related technology trends" , Journal of the Institute of Image Information and Television Engineers, 67, 2013, 109–115

[2] NTT DOCOMO, Website, https://www.nttdocomo.co.jp/info/news_release/2014/05/29_00.html

[3] Nishizawa N, Tatsuya O, Masayuki M, "Modeling of Cognitive Bias before and after Image Quality Change for Video Viewing Users", IEICE Technical Report CQ2020-24, 2020, 13-18

[4] Mussweiler, T., Strack, F, "The semantics of anchoring", Organizational Behavior and Human Decision Processes, 86, 2001, 234–255

[5] A. Sackl, P. Zwickl, S. Egger-Lampl, and P. Reichl, "The role of cognitive dissonance for QoE evaluation of multimedia services" , 2012 IEEE Globecom Workshops, GC Wkshps 2012, 2012, 1352–1356

[6] Ricky K. P. Mok, Xiapu L, Edmond W. W. C, and Rocky K. C. C, "QDASH: A QoE-aware DASH system", MMSys '12, Proceedings of the 3rd Multimedia Systems Conference, 2012

[7] Netflix, Website, Zhi L, Anne A, Ioannis K, Anush M and Megha M, Toward A Practical Perceptual Video Quality Metric, https://netflixtechblog.com/