QoE Assessment Methodologies for 3D Video Services

Kazuhisa Yamagishi, Taichi Kawano, and Kimiko Kawashima

Abstract

Technological advances in three-dimensional (3D) video enable users to watch 3D video content. However, watching such content may cause visual discomfort and fatigue. This article gives an outline of the relationship between 3D video degradation factors and quality of experience (QoE) and introduces QoE assessment methodologies and standardization activities.

1. Introduction

Users can now easily enjoy watching three-dimensional (3D) content on terminals such as 3D televisions (TVs), personal computers, and smartphones due to the recent advances made in these terminals. Video quality has also been improved by the introduction of high definition (HD). In addition, by introducing depth perception to video, 3D video gives a new experience to users. However, some users complain of visual discomfort and fatigue from watching 3D video. To provide high-quality 3D video content, it is important to design and manage services based on the quality of experience (QoE), and to do this, a 3D video QoE assessment methodology is essential.

2. 3D video QoE

This section describes QoE for 3D video services. The International Telecommunication Union, Telecommunication standardization section (ITU-T) Recommendation BT.2021 defines 3D video QoE in terms of visual quality, depth perception, and visual discomfort [1], [2]. As described in section 1, fatigue [3] from 3D video content is also an important factor. In this article, we define QoE in terms of visual quality, depth perception, discomfort, and fatigue. These QoE factors are affected by the 3D video processing chain, as shown in **Fig. 1**.

There are differences between the processing chain and the human vision system (HVS) in 3D video

acquisition and display. For example, the position and angle of a camera do not match those of the human eyes. In addition, when a user views 3D video content, they see an image formed from two video images viewed separately by the left and right eye through stereoscopic glasses in the rendering phase of the processing chain. As a result, puppet theater[*1], cardboard[*2], and spatio-temporal asynchronous effects between the left and right views occur. Crosstalk[*3] due to the stereoscopic glasses also occurs.

3D video is downsized and/or encoded in order to reduce the network bandwidth and the amount of storage needed. To use the existing infrastructure for codec and transmission, the spatial resolution of the left and right views, which are arranged in a side-by-side frame-compatible format, is usually down-converted by half in the horizontal direction to maintain the spatial resolution of a full high-definition (HD) 2D video sequence. The video is encoded by MPEG-2 (Motion Picture Experts Group-2) or H.264/AVC (advanced video coding) and is transmitted to a user terminal such as a set-top box. Finally, the side-by-

---

*1 The puppet-theater effect makes a 3-D image look unnaturally small compared with the target image; people appear to be miniaturized puppets.

*2 The cardboard effect makes 3-D images look layered, i.e., consisting of flat objects against a flat background, though an observer can grasp the situation in front of and behind the shooting target.

*3 Crosstalk appears because of imperfect view separation when a small proportion of one eye's image is also perceptible by the other eye.
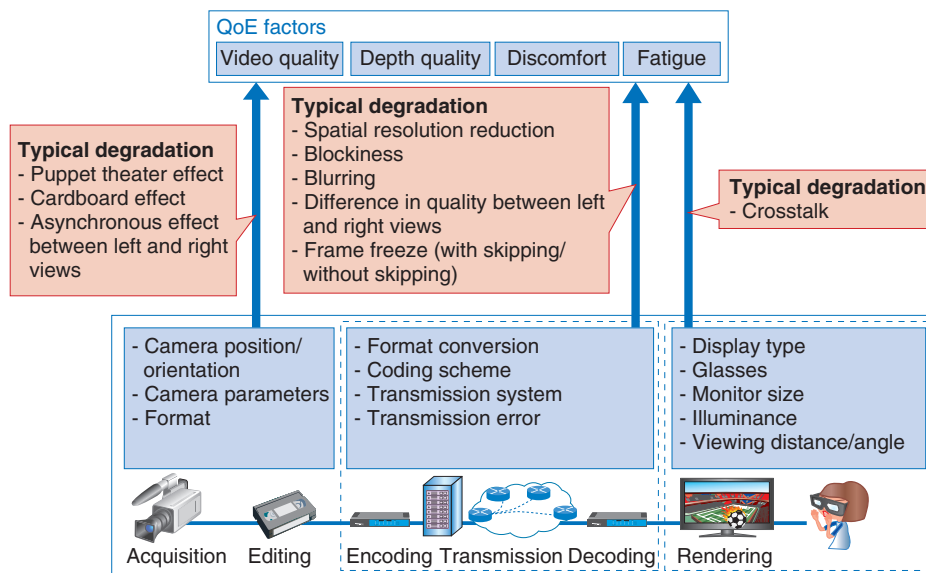
Fig. 1.   3D video processing chain.

side format video is decoded and up-converted to two full HD video signals for the left and right views. Thus, users perceive degradations in quality due to the reduced spatial resolution in addition to coding artifacts such as block noise. To prevent the degradation due to the reduced spatial resolution, the use of two full HD video signals for left and right views, which is called the frame-sequential format, is ideal. In this case, an H.264/MVC (multiview video coding) is often used, which involves an inter-view prediction technique that encodes the right-view video using both videos for left and right views in order to reduce the bit rate for the right-view video. With this system, service providers often encode the right-view video at a much lower bit rate than the left on the basis of binocular suppression. The two 2D video signals for the left and right views have full HD resolution, but they have an asymmetric quality in addition to coding artifacts such as block noise.

Encoded 3D video is packetized and transmitted over a network such as an IP (Internet protocol) or terrestrial network. Packet loss sometimes occurs in networks. Block noise occurs if there is no packet-loss concealment (PLC) technique applied in the user terminal. In contrast, if a PLC technique is applied in the user terminal, freezing artifacts will be introduced when the PLC scheme of the receiver replaces the erroneous frames (either due to packet loss or error propagation) with the previous error-free frame until a decoded picture without errors has been received.

This type of artifact is also called *freezing with skipping*. The rebuffering artifacts come from rebuffering events at the receiver, which could be the result of a stream arriving late. This type of artifact is also called *freezing without skipping.*

Degradation perceptions may change due to the display size, room illuminance, viewing distance, and angle.

Therefore, as mentioned previously, because 3D video QoE is affected by many factors, methodologies are needed to assess 3D video QoE.

3.   3D video subjective assessment methodology

Subjective assessment, in which users subjectively evaluate 3D video QoE, is a fundamental quality assessment technique. ITU-R Recommendation BT.2021 was standardized for the 3D video quality subjective assessment method. As described in section 2, it is important to develop methodologies that assess depth perception, discomfort, and fatigue since 3D video QoE is affected by these factors, in addition to visual quality. The Video Quality Experts Group (VQEG) is currently investigating subjective assessment methodologies concerning depth perception and discomfort. However, it is often difficult to evaluate discomfort and fatigue using questionnaire-based subjective assessment because the levels of these indicators are sometimes low in questionnaires. Therefore, it is important to supplement information

obtained from questionnaires with biological information such as heart rate, breathing rate, pupil changes, and eye-blink responses when assessing discomfort and fatigue. Our group has investigated the relationship between fatigue and biological information. The relationship between fatigue and the difference in video quality between left and right views is shown in **Fig. 2**. In this figure, the lower the number on the fatigue axis, the greater the amount of fatigue; i.e., a value of 1 represents high fatigue, whereas 5 represents low fatigue. The relationship between the eye-blink rate and the difference in video quality is shown in **Fig. 3**. As shown in Fig. 2, fatigue score increases as the difference in video quality increases. As shown in Fig. 3, the eye-blink rate increases as the difference in video quality increases. These results suggest that fatigue can be evaluated using the eye-blink rate.

4. 3D video quality objective estimation methodology

Developing an objective quality estimation model that can be used to estimate QoE using information such as 3D video signals is essential for monitoring QoE.

It is important to take into account the degradation factors described in section 2 in order to develop such a model. In principle, block noise, blurring, and freezing due to encoding and transmission also occur in 3D video services. Therefore, the 2D video quality objective estimation model can be applied to 3D video quality estimation. However, since quality degradation factors such as the difference in video quality between left and right views, the asynchronous effect between left and right views, and crosstalk do not occur in 2D video services, these factors need to be taken into account in 3D video quality estimation.

Our group has been developing an objective quality estimation model that takes 2D video quality for left and right views, which is derived from a 2D video quality objective estimation model, as input. Video quality is denoted as a mean opinion score (MOS), where 2D video quality for the left view is denoted as MOS-L, 2D video quality for the right view is denoted as MOS-R, and the difference in video quality for left and right views is denoted as

dMOS-LR (= ABS(MOS-L – MOS-R)).

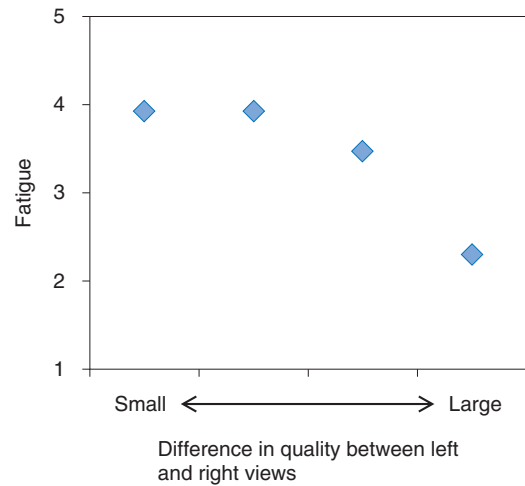We compared the performance of our model with

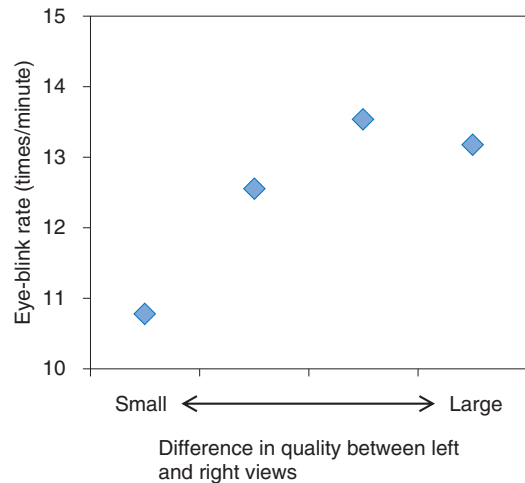Fig. 2. Difference in quality between left and right views vs. fatigue.

Fig. 3. Difference in quality between left and right views vs. eye-blink rate.

that of a conventional model used to calculate the average 2D video quality for left and right views. **Table 1** lists the performance values of our model and the conventional model, i.e., the root mean square errors in the range of

$0 \leq$ dMOS-LR ≤ 1 and $1 <$ dMOS-LR.

The results show that our model can estimate 3D video quality in the range of

$1 <$ dMOS-LR,

more accurately than the conventional model can.

VQEG is also investigating a 3D video quality objective estimation model and discussing a test plan that will be used to verify the validity of such future models.

Table 1. Quality estimation accuracy.

	Proposed model	Conventional model
0 ≤ dMOS-LR ≤ 1	0.22	0.24
1 < dMOS-LR	0.23	0.35

5. Conclusion

The introduction of 3D video has enabled service providers to provide a new visual experience, e.g., depth perception, to users. However, some users have complained of visual discomfort and fatigue from watching 3D video. Therefore, it is important to clarify factors that affect QoE and to develop a model to estimate QoE in order to provide high-QoE 3D video. Our group has been investigating subjective assessment methodologies for 3D video quality and fatigue as well as an objective quality estimation model. We plan to propose our model to VQEG in the future. We will also develop subjective and objective quality assessment methods for QoE other than that for 3D video quality using biological information. We will promote the practical use of these methods, which will contribute to providing a safe and pleasant 3D video streaming service.
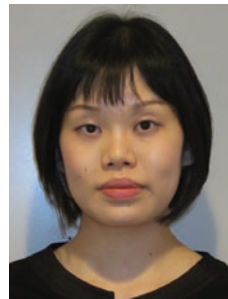
References

[1] ITU-R Recommendation BT.2021, "Subjective methods for the assessment of stereoscopic 3DTV systems," Aug. 2012.
[2] M. Lambooij, W. Ijsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: a review," Journal of Imaging Science and Technology, Vol. 53, No. 3, pp. 030201–14, 2009.
[3] K. Yamagishi, T. Kawano, and T. Hayashi, "Effect of Difference in 2D Video Quality for Left and Right Views on Overall 3D Video Quality," Proc. of 2012 IEEE International Conference on Image Processing (ICIP 2012), Orlando, Florida, USA, 2012.

Kazuhisa Yamagishi
Research Engineer, Service Assessment Group, Communication Traffic & Service Quality Project, NTT Network Technology Laboratories.
He received the B.E. degree in electrical engineering from Tokyo University of Science and the M.E. degree in electronics, information, and communication engineering from Waseda University, Tokyo, in 2001 and 2003, respectively. He joined NTT in 2003. He has been engaged in subjective quality assessment of multimedia telecommunications and image coding. He is currently working on quality assessment of multimedia services over IP networks. He has been contributing to ITU-T SG12 since 2006. From 2010 to 2011, he was a visiting researcher at Arizona State University. He received the Young Investigator's Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2007 and the Telecommunication Advancement Foundation Award in Japan in 2008. He is a member of IEICE.

Taichi Kawano
Researcher, Service Assessment Group, Communication Traffic & Service Quality Project, NTT Network Technology Laboratories.
He received the B.E. and M.E. degrees in engineering from Tsukuba University, Ibaraki, in 2006 and 2008, respectively. Since joining NTT in 2008, he has been engaged in research on 2D/3D video quality assessment. He received the Young Investigator's Award from IEICE in 2011. He is a member of IEICE.

Kimiko Kawashima
Researcher, Service Assessment Group, Communication Traffic & Service Quality Project, NTT Network Technology Laboratories.
She received the B.E. and M.E. degrees in engineering from Keio University, Kanagawa, in 2008 and 2010, respectively. Since joining NTT in 2010, she has been engaged in research on quality assessment of visual communication services. She is currently working on quality assessment of 3D services. She is a member of IEICE.