EFFECT OF DIFFERENCE IN 2D VIDEO QUALITY FOR LEFT AND RIGHT VIEWS ON OVERALL 3D VIDEO QUALITY

Kazuhisa Yamagishi[1], *Taichi Kawano*[1], *and Takanori Hayashi*[1]

[1]NTT Service Integration Laboratories, NTT Corporation, Tokyo 180-8585, Japan

ABSTRACT

Multi-view coding is expected to be used in next-generation encoders and decoders for stereoscopic video services. In multi-view coding, the bit rate for the right eye view is lower than that for the left eye view because the right view is encoded using inter-view technology. In addition, the right view is encoded at a much lower bit rate on the basis of binocular suppression. As a result, the video quality for the left view can be different from that for the right view. We explore here how such a video quality difference between the left and right views affects the overall 3D video quality. In addition, we model the quality characteristics. We conducted a subjective quality assessment to derive subjective quality characteristics. We show that the difference in 2D video quality between the left and right views has little influence on the overall 3D video quality and that the overall 3D video quality can be modeled using the 2D video quality for left and right views.

***Index Terms*—** 3D, 2D, Quality, AVC, MVC

1. INTRODUCTION

Due to advances in video acquisition, encoders and decoders (codecs), and displays, service providers have launched stereoscopic (hereafter, 3D) video services over cable, terrestrial, satellite, Internet protocol, and mobile networks [1, 2, 3]. Currently, many service providers provide a 3D video service using MPEG-2 or H.264/AVC (advanced video coding). In addition, in order to use the existing infrastructure for compression and transmission, the spatial resolution of the left and right views is usually reduced by half in the horizontal direction to maintain the spatial resolution of a full high definition (HD) 2D video sequence [4]. As a result, users perceive a degradation in quality due to the reduced spatial resolution [5]. Therefore, the use of full HD resolution per view is desirable to provide high-quality 3D video services.

There are basically two solutions to encode 3D full HD video. One solution is to apply H.264/AVC to each view. In such a case, the 2D videos for the left and right views usually have a symmetric quality. The other solution is to use H.264/MVC (multi-view coding) for encoding the 3D video. In H.264/MVC, the encoder usually encodes the left view at a higher bit rate than the right view on the basis of inter-view technology. In addition, service providers encode the right view at a much lower bit rate than the left on the basis of binocular suppression. As a result, the 2D video quality for the left view differs from that for the right view.

Subjective characteristics for 3D video quality were studied in [5, 6]. References [5, 6] show that the 3D video quality decreases as the coding artifacts increase, similarly to 2D video quality characteristics. Reference [5] showed that the quality of full HD resolution for left and right views is higher than that of half the size of the original sequence in the horizontal direction. In addition, L. Stelmach et al. showed the subjective quality characteristics for the asymmetric quality between the left and right views in the spatial resolution [7]. The right view was kept in full resolution, while the left view was down-sampled and then was up-sampled to the full resolution in the display. The result showed that there was a significant difference in the video quality between symmetric and asymmetric resolution. Thus, the effect of resolution on video quality has been well studied. On the contrary, P. Aflaki et al. show in Ref. [8] that the asymmetric quality has little impact on 3D quality when the quantization parameters for the left and right views are not significantly different. In addition, G.Saygili et al. compared the symmetric and asymmetric quality [9, 10]. The results showed that there was an impact of the asymmetric quality on the overall 3D video quality. However, Refs. [8, 9, 10] do not describe to what extent the 3D video quality is affected by the difference in 2D video quality between the left and right views. To develop an objective quality assessment model, it is necessary to investigate how the difference in 2D video quality between the left and right views affects 3D video quality and to model that characteristic.

We first describe here our subjective quality assessment to derive the effect of the difference in 2D video quality between the left and right views on the overall 3D video quality using conventional H.264/AVC, because it is important to separately control the video quality for the left and right views. Then, by using 2D video quality for the left and right views, we investigate how the difference in 2D video quality between these views affects the overall 3D video quality. Finally, we modeled the characteristics using subjective video quality for the left and right views because such subjective video quality can be estimated using objective quality assessment models

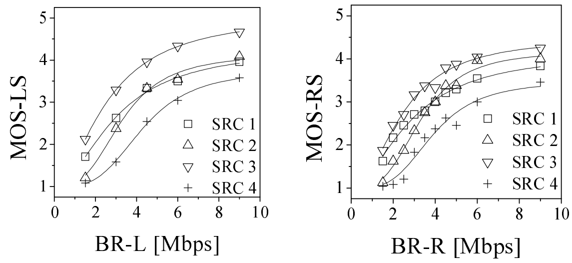
Table 1. Spatial and temporal information.

		Left		Right	
SRC No.	Title	SI	TI	SI	TI
1	Flamenco	39	24	39	24
2	Dolphin	57	58	59	58
3	Woman and maple leaves	50	10	56	10
4	Aquarium	65	21	64	21

Table 2. Overall bit rates for 3D video.

BR-O* (Mbps)	3.0	4.5	5.0	5.5	6.0
BR-L* (Mbps)	1.5	3.0	3.0	3.0	3.0
BR-R* (Mbps)	1.5	1.5	2.0	2.5	3.0
BR-O (Mbps)	6.0	7.0	8.0	9.0	8.0
BR-L (Mbps)	4.5	4.5	4.5	4.5	6.0
BR-R (Mbps)	1.5	2.5	3.5	4.5	2.0
BR-O (Mbps)	9.0	10.0	11.0	12.0	18.0
BR-L (Mbps)	6.0	6.0	6.0	6.0	9.0
BR-R (Mbps)	3.0	4.0	5.0	6.0	9.0

*: BR-O, BR-L, and BR-R represent bit rate for 3D, left, and right views, respectively.



(a) BR-L vs. MOS-LS (b) BR-R vs. MOS-RS

Fig. 1. 2D video quality for left and right views.

for 2D video (e.g., ITU-T Recommendation J.341).

The remainder of this paper is structured as follows. The adopted subjective assessment methodology is described in Section 2. Subjective quality characteristics are presented in Section 3. Finally, we summarize our findings and suggest possible directions for future studies in Section 4.

2. SUBJECTIVE QUALITY ASSESSMENT

We used H.264/AVC and four sources (SRCs) to investigate the effect of the difference in 2D video quality between the left and right views on the overall 3D video quality.

The selection of video content is an important factor in deriving the subjective quality characteristics. The selected set of video sequences should span a wide range of spatial and temporal information. Four full HD 3D video sequences with a duration of 10 seconds each, were used in the experiment. Spatial information (SI) and temporal information (TI) defined by ITU-T Recommendation P.910 [11] are listed in Table 1. A man and woman dance in SRC 1 (Flamenco); a dolphin swims and jumps a pool in SRC 2 (Dolphin); a woman looks at maple leaves in SRC 3 (Woman and maple leaves); and several tropical fish swim in a tank in SRC 4 (Aquarium).

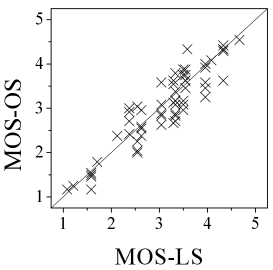
To cover a wide range of 2D video quality for the left and right views and to investigate how the quality difference between the left and right views affects the 3D video quality in H.264/MVC, the required bit rates for the left view (BR-L) were 1.5, 3.0, 4.5, 6.0, and 9.0 Mbits per second (Mbps), and those for the right view (BR-R) were 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, and 9.0 Mbps. The overall bit rates for 3D video (BR-O) are listed in Table 2. A total of 60 3D processed video sequences (PVSs) were used (15 conditions $\times$ 4 SRCs), while there were 20 2D PVSs for the left view (5 conditions $\times$ 4 SRCs), and 40 2D PVSs for the right view (10 conditions $\times$ 4 SRCs).

In the subjective quality assessment, the 3D video quality was evaluated using an absolute category rating (ACR) with a five-grade scale (Excellent, Good, Fair, Poor, and Bad) [12]. The subjects were required to rate the quality within 5 seconds after a video sequence was presented. The subjective score was represented as a mean opinion score (MOS), where MOS-OS, MOS-LS, and MOS-RS represent subjective MOS for the overall 3D video, subjective MOS for the left view, and subjective MOS for the right view, respectively. Before we started the subjective test, we conducted screening tests. We used two tests indicated in ITU-R Recommendation BT. 1438 [13]: Coarse and Fine stereopsis tests. We also screened the subjects for visual acuity and color vision. Twenty-four subjects (twelve males and twelve females) passed the screening tests and participated in the subjective test. They ranged in age from 20 to 39 years old (average: 30 years old). The subjects viewed each video sequence with polarized glasses at a distance of 3H (about 150 cm), where H indicates the picture height. The encoded videos were displayed at the 1920 $\times$ 1080 native resolution on a 40-inch monitor. We used 20 lux for the room illumination as the laboratory environment. In the 2D video test, subjects viewed each video sequence without the glasses.

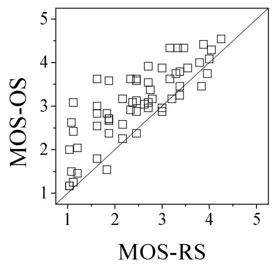
3. SUBJECTIVE QUALITY CHARACTERISTICS

To ensure that the range of MOS-LS and MOS-RS was widely spread, we first show the relationship between the bit rate and either MOS-LS or MOS-RS, as shown in Fig. 1. The 2D video quality is widely scattered in the range from 1 to 5. In addition, the MOS per bit rate depends on the SRC.

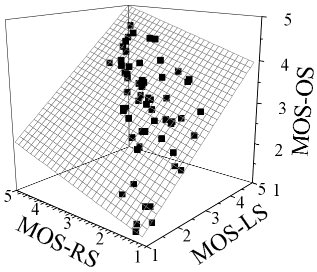
Next, we describe the relationship among MOS-OS, MOS-LS, and MOS-RS (Fig. 2). The Pearson's correlation coefficient (PCC) and root mean square error (RMSE) between MOS-OS and either MOS-LS or MOS-RS for each



(a) MOS-LS vs. MOS-OS



(b) MOS-RS vs. MOS-OS



(c) MOS-OS vs. MOS-LS vs. MOS-RS

Fig. 2. 2D video quality vs. 3D video quality.

Table 3. RMSE.

(a) Left view

SRC No.	1	2	3	4
PCC	0.84	0.91	0.87	0.92
RMSE	0.34	0.34	0.37	0.38

(b) Right view

SRC No.	1	2	3	4
PCC	0.79	0.74	0.80	0.83
RMSE	0.55	1.04	0.85	0.73

SRC are listed in Table 3. MOS-LS highly correlates with MOS-OS in most of the plots, while MOS-RS does not. In addition, although the PCC and RMSE between MOS-OS and MOS-LS did not depend on the SRC, the PCC and RMSE between MOS-OS and MOS-RS depended on the SRC, as indicated in Table 3. That is, the dominant factor for the MOS-OS was the highest 2D video quality of the two views (i.e., the dominant factor for the MOS-OS was MOS-LS because MOS-LS was higher than MOS-RS in this experiment).

We investigated how the MOS-RS affected the MOS-OS. Fig. 3 shows the relationship between deltaMOS-OL and deltaMOS-LR, where deltaMOS-OL denotes the difference between MOS-OS and MOS-LS ($MOS\text{–}OS - MOS\text{–}LS$), and deltaMOS-LR denotes the difference between MOS-LS and MOS-RS ($MOS\text{–}LS - MOS\text{–}RS$). The deltaMOS-OL decreases as deltaMOS-LR increases. This result implies that although the dominant factor of the MOS-OS was the 2D video quality for the left view in this experiment, the MOS-OS was further reduced by the deltaMOS-LR. However,

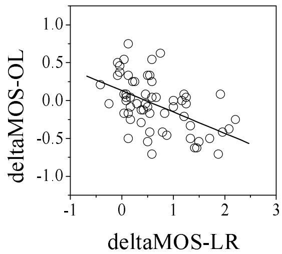


Fig. 3. deltaMOS-OL vs. deltaMOS-LR.

compared with the decrease in the MOS-LS, the increase in the deltaMOS-LR did not affect the MOS-OS much because even when the deltaMOS-LR was about 2.0, the deltaMOS-OL was about -0.5. We believe that this result is attributed to the binocular rivalry theory [14], which states that the perceived 3D video quality is close to that of the higher quality view. Note that MOS-LS was basically higher than MOS-RS because BR-L is equal to or larger than BR-R in this experiment. In addition, MOS-LS correlates with MOS-OS because there was not much significant impact of deltaMOS-LR on deltaMOS-OL. Therefore, if MOS-RS is higher than MOS-LS, the characteristic would be reversed.

From the investigations previously described, we estimated the MOS-OS using multiple regression, where the independent values are MOS-LS and MOS-RS. The objective overall 3D video quality, MOS-OO, can be modeled as follows:

$$\begin{aligned} MOS\text{–}OO &= 0.16 + 0.99 \cdot MOS\text{–}LS \\ &\quad -0.28 \cdot (MOS\text{–}LS - MOS\text{–}RS) \quad (1) \\ &= 0.16 + 0.71 MOS\text{–}LS \\ &\quad +0.28 MOS\text{–}RS, \quad (2) \end{aligned}$$

where the coefficients of multiple regression are significant at the 1% level. The coefficient of the left view is higher than that of the right view because the quality for the left view was higher than that for the right view in this experiment. If the quality for the right view is higher than that for the left view, the coefficient for the right view would be larger than that for the left view. Incidentally, even if the interaction term $MOS\text{–}LS \times MOS\text{–}RS$ is added to the Eq. (2), such an Eq. is not properly validated in terms of the significance level because the coefficient of $MOS\text{–}LS \times MOS\text{–}RS$ was not significant at both the 1 and 5% levels. Fig. 4 plots the relationship between MOS-OS and MOS-OO. For all four video sequences, the PCC and RMSE between MOS-OS and MOS-OO, which is derived by Eq. (2), is better than that between MOS-OS and MOS-LS, as listed in Tables 3 and 4. From the coefficients of MOS-LS and MOS-RS, the impact of MOS-LS on MOS-OS was about 2.5 times as large as that of MOS-RS on MOS-OS. From these results, we can conclude that the MOS-OS is composed of not only MOS-LS but also MOS-RS, and that MOS-OS can be modeled by the multiple-regression model that has MOS-LS and MOS-RS as

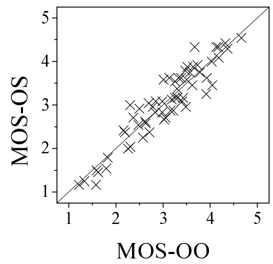


Fig. 4. Objective 3D quality vs. subjective 3D quality.

Table 4. RMSE between MOS-OS and MOS-OO.

SRC No.	1	2	3	4
PCC	0.89	0.93	0.92	0.96
RMSE	0.28	0.33	0.26	0.31

the independent values.

4. CONCLUSION

We conducted a subjective quality assessment to derive the subjective quality characteristics between the 2D video quality for the left and right views and the 3D video quality, and we modeled the quality characteristics.

We first showed that the 3D video quality mainly depended on the highest 2D video quality of the two views (i.e., the 2D video quality for the left view in this experiment). Next, we found that the difference in 2D video quality between the left and right views had little impact on 3D video quality. Finally, we found that the subjective 3D video quality can be modeled by the multiple regression function, where the independent values are the 2D video quality for the left and right views.

Further work is suggested in order to extend the results reported in this paper. First of all, our proposed multiple regression model needs to be validated using other sources. Next, it would be interesting to investigate whether the difference in 2D video quality between the left and right views has an effect on depth perception. Furthermore, it would be interesting to develop a video-signal-based objective quality assessment model because we used 2D video subjective quality for the left and right views as input of the model for this paper.

5. REFERENCES

[1] B. F. Col and K. O'Connell, "3DTV at home: status, challenges and solutions for delivering a high quality experience," *Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Jan. 2010.

[2] P. Merkle, K. Muller, and T. Wiegand, "3D video: Acquisition, coding, and display," *IEEE Trans. on Consumer Electronics*, vol. 56, no. 2, pp. 946–950, Jan. 2010.

[3] Q. Huynh-Thu, P. Le Callet, and M. Barkowsky, "Video quality assessment: From 2D to 3D - challenges and future trends," *Int. Conf. on Image Processing (ICIP)*, pp. 4025–4028, Sept. 2010.

[4] K. Kim, J. Lee, D. Suh, and G. Park, "Efficient stereoscopic contents file format on the basis of ISO base media file format," *Proc. of the SPIE*, vol. 7256, pp. 72560N–1–9, Jan. 2009.

[5] K. Yamagishi, L. Karam, J. Okamoto, and T. Hayashi, "Subjective characteristics for stereoscopic high definition video," *Int. Workshop on Quality of Multimedia Experience (QoMEX)*, Sept. 2011.

[6] J. Seo, D. Kim, and K. Sohn, "Compressed stereoscopic video quality metric," *Proc. of the SPIE*, vol. 7237, pp. 723712–1–11, Jan. 2007.

[7] L. Stelmach, W. Tam, D. Meegan, and A. Vincent, "Stereo image quality: Effects of mixed spatio-temporal resolution," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 2, Mar. 2000.

[8] P. Aflaki, M. M. Hannuksela, J. Hakkinen, P. Lindroos, and M. Gabbouj, "Subjective study on compressed asymmetric stereoscopic video," *Int. Conf. on Image Processing (ICIP)*, pp. 4021–4024, Sept. 2010.

[9] G. Saygili, C. G. Gurler, and A. Murat Tekalp, "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3d video streaming," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 593–601, June 2011.

[10] G. Saygili, C. G. Gurler, and A. Murat Tekalp, "Quality assessment of asymmetric stereo video coding," *Int. Conf. on Image Processing (ICIP)*, pp. 4009–4012, Sept. 2010.

[11] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Apr. 2008.

[12] ITU-R Recommendation BT.500, "Methodology for the subjective assessment of the quality of television pictures," Sept. 2009.

[13] ITU-R Recommendation BT.1438, "Subjective assessment of stereoscopic television pictures," Mar. 2000.

[14] R. Blake, "Threshold conditions for binocular rivalry," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 3, no. 2, pp. 251–257, May 1977.