IEICE TRANS. COMMUN., VOL.E96–B, NO.2 FEBRUARY 2013

PAPER Special Section on Quality of Communication Services Improving Quality of Life

Subjective Quality Metric for 3D Video Services

Kazuhisa YAMAGISHI[†a)], **Taichi KAWANO**[†], **Takanori HAYASHI**[†], *and* **Jiro KATTO**[††], *Members*

SUMMARY Three-dimensional (3D) video service is expected to be introduced as a next-generation television service. Stereoscopic video is composed of two 2D video signals for the left and right views, and these 2D video signals are encoded. Video quality between the left and right views is not always consistent because, for example, each view is encoded at a different bit rate. As a result, the video quality difference between the left and right views degrades the quality of stereoscopic video. However, these characteristics have not been thoroughly studied or modeled. Therefore, it is necessary to better understand how the video quality difference affects stereoscopic video quality and to model the video quality characteristics. To do that, we conducted subjective quality assessments to derive subjective video quality characteristics. The characteristics showed that 3D video quality was affected by the difference in video quality between the left and right views, and that when the difference was small, 3D video quality correlated with the highest 2D video quality of the two views. We modeled these characteristics as a subjective quality metric using a training data set. Finally, we verified the performance of our proposed model by applying it to unknown data sets.

key words: *QoE, quality, 3D, 2D, subjective quality metric, objective quality metric*

1. Introduction

Many content service providers offer three-dimensional (3D) video content over terrestrial, satellite, cable, or Internet protocol networks [1] because significant progress has recently been made in the development of displays and glasses for stereoscopic (hereafter, 3D) video. In general, the quality of experience (QoE) for 3D video service is mainly said to be composed of the spatio-temporal visual quality (i.e., 2D video quality), depth perception [2], [3], and visual comfort [4], [5]. In addition, new degradation factors such as crosstalk and the puppet theater effect [6] have an impact on QoE in 3D video services. 3D video QoE is affected by the 3D video processing chain [7]–[9] consisting of the video camera, encoder, transmission, decoder, display, and glasses. Although the video camera, display, and glasses affect the 3D video QoE, we mainly focus here on the effect of compression as a QoE factor. Compression affects the 3D video quality due to the limited transmission bandwidth (i.e., bit rate) despite the improving quality of video cameras and displays.

In terms of compression, 3D video quality depends on the employed coding scheme (e.g., MPEG-2 [10], H.264/AVC (advanced video coding) [11], or H.264/MVC (multi-view coding) [11]), codec implementation (e.g., profile and level), and encoding video format (e.g., frame-sequential or frame-compatible format). To use the existing infrastructure for codec and transmission, the spatial resolution of the left and right views, known as the side-by-side frame-compatible format [1], [12], 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 or H.264/AVC and is transmitted to the user's terminal such as a set-top box (STB). Finally, the side-by-side format video is decoded and up-converted to two full HD video signals for the left and right views, as shown in Fig. 1(a) (System A). Thus, users perceive a degradation in quality due to the reduced spatial resolution, where video quality for the left and right views is basically symmetric [13].

To avoid the degradation due to the reduced spatial resolution, the use of two full HD video signals for left and right views [13], which is called the frame-sequential format [12], is ideal. Two solutions that apply this format have been studied. One solution is to use H.264/MVC, as shown in Fig. 1(b) (System B). The H.264/MVC has 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. The other solution is to

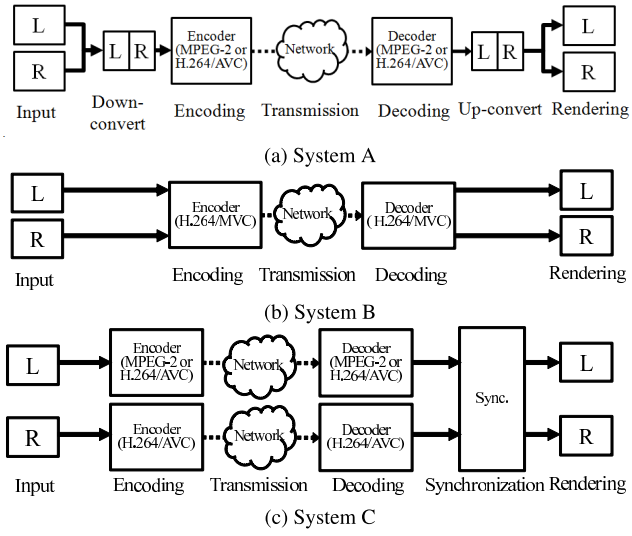


Fig. 1 3D video processing chains.

Manuscript received May 22, 2012.
Manuscript revised August 23, 2012.

[†]The authors are with NTT Service Integration Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.
[††]The author is with the Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo, 169-8555 Japan.
a) E-mail: yamagishi.kazuhisa@lab.ntt.co.jp

DOI: 10.1587/transcom.E96.B.410

use two encoders, as shown in Fig. 1(c) (System C). Two encoded video streams are separately transmitted and are then synchronized at the STB. As an example, the left-view video might be encoded by MPEG-2 or H.264/AVC and transmitted over a terrestrial or satellite network. At the same time, the right-view video might be encoded by H.264/AVC and transmitted over an IP network. Finally, the STB synchronizes both video signals and displays them as a 3D video. With systems B and C, service providers often encode the right-view video at a much lower bit rate than the left on the basis of the binocular suppression. In these systems, the two 2D video signals for the left and right views have full HD resolution, but have an asymmetric quality.

As mentioned above, 3D video quality depends on the 3D video processing chain. Therefore, to provide a high-quality 3D video service to users, it is important for service providers to optimize the video processing chain taking into account the quality balance between left and right-view video signals and to monitor the 3D video quality. To optimize the chain on the basis of subjective quality characteristics, it is necessary to investigate how the difference in 2D video quality between left and right views affects the 3D video quality. Moreover, to monitor 3D video quality in real time, it is necessary to model the characteristics and then to develop an objective quality metric for 3D video service.

We describe here some conventional studies on 3D video quality. We [13] 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 left and right views in the spatial resolution [14]. The right view was kept in full resolution, while the left view was down-sampled and then 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. However, neither of these studies describe to what extent 3D video quality was affected by the spatial resolution.

P. Aflaki et al. [15], G. Saygili et al. [16], [17], A. Vetro [18], and V.D. Silva et al. [19] showed subjective quality characteristics for the asymmetric quality between left and right views. To create an asymmetric quality, the quantization parameter (QP) for the left view differs from that for the right view [15]. As in the binocular rivalry phenomenon, in which the human visual system (HVS) fuses the two video signals so that the perceived 3D video quality is close to that of the higher quality view, Aflaki et al.'s results showed that the asymmetric quality has little impact on the 3D video quality. However, the quality difference between the left and right views was small because the difference in QP between left and right views was small in the study [15]. G. Saygili compared video quality for symmetric and asymmetric coding in terms of peak signal-to-noise ratio (PSNR) [16], [17]. The results showed that video quality for the asymmetric coding was lower than that for the symmetric coding. A. Vetro also compared video quality for systems B and C [18]. The results showed that video quality for the asymmetric coding was almost the same as that for the symmetric coding, when the bit rate ratio of the left view to the right view was 50% in H.264/MVC and that video quality for the asymmetric coding was lower than that for the symmetric coding, when the bit rate ratio of the left view to the right view was ranged from 35% to 5% in H.264/MVC. V. D. Silva investigated asymmetric quality for blocking and blurring artifacts by varying the QP and Gaussian filter [19]. The results showed that subjects identify the level of asymmetric encoding based on blocking artifacts, rather than the asymmetric level of blurring. Although these studies [15]–[19] describe that the difference in 2D video quality left and right views impacts 3D video quality, they do not describe to what extent 3D video quality was affected by this difference and do not model this.

P. Compisi et al. proposed an objective quality metric that can be used to evaluate 3D video quality using the average video quality between left and right views [20]. This metric is expected to work well when the difference in video quality between left and right views is small. However, it is thought that when there is a significant difference in 2D video quality between the left and right views, the model would overestimate the 3D video quality. A significant quality difference is expected in systems B and C, and therefore, a model that can evaluate 3D video quality when there is a significant difference in video quality between left and right views is required.

From these investigations, conventional studies do not show to what extent 3D video quality is affected by the difference in 2D video quality between left and right views and do not model this. Therefore, we need to describe these characteristics in detail, model them, and show that this model performs better than the conventional model. To address these issues, we propose an objective quality metric that can be used to optimize the quality balance between left and right views and to monitor 3D video quality. First, we conduct subjective quality assessments for 3D and 2D video to derive how a difference in 2D video quality between the left and right views affects the 3D video quality. Then, we model the characteristics as a subjective quality metric, where we focus on modeling 3D video quality using subjective quality for left and right views because objective quality metrics for 2D video services [21] have been standardized and thoroughly studied. Although our proposed model is based on the premise that the output of an objective quality metric for 2D video is used as the input of our proposed model, deriving the performance of our proposed model by using the output of the objective quality metric for 2D video is beyond the scope of this paper.

The remainder of this paper is structured as follows. The adopted subjective quality assessment methodology is described in Sect. 2. Subjective quality characteristics are shown in Sect. 3. The proposed model is described in Sect. 4. The performance of our proposed model is presented in Sect. 5. Finally, we summarize our findings and suggest possible directions for future studies in Sect. 6.

2. Subjective Quality Assessment

We conducted subjective quality assessments to derive subjective quality characteristics and to model the effect of the difference in 2D video quality between left and right views on 3D video quality. We used 12 sources (SRCs) and 2 video codecs in our experiments.

The selection of video content is an important part of a subjective quality assessment when deriving subjective quality characteristics. The selected set of video sequences should span a wide range of spatial and temporal information. Twelve full HD 3D video sequences with a duration of 10 seconds each, described as follows, were used in the experiment: woman takes a photo of leaves in SRC 1 (Photo); a woman tries on some clothes in front of a mirror in SRC 2 (Mirror); a woman peddling a bicycle moves toward a camera in SRC 3 (Bicycle (front face)); a basketball player dribbles in SRC 4 (Basketball); a man and woman stand and look at flowers in a flower garden in SRC 5 (Flower garden); a woman paints flowers in a room in SRC 6 (Paint); a well-dressed woman walks in a living room in SRC 7 (Well-dressed woman); a woman peddling a bicycle moves to the right in a park in SRC 8 (Bicycle (side face)); a woman looks at maple leaves in SRC 9 (Woman and maple leaves); cheerleaders perform at a game in SRC 10 (Cheerleaders); a clown gives a balloon to a girl in SRC 11 (Clown); and some tropical fish swim in a tank at an aquarium in SRC 12 (Tropical fish). SRCs 1, 2, 3, 7, 8, and 9 were provided by the National Institute of Information and Communications Technology (NICT) Japan, SRCs 4, 5, 6, 10, and 11 were provided by the Digital Content Association of Japan, and SRC 12 was our own content. Spatial information (SI) and temporal information (TI) defined by ITU-T Recommendation P.910 [22] are listed in Table 1. The SRCs were classified into two groups (1 and 2) to verify whether a model was able to accurately evaluate 3D video quality of unknown video.

H.264/AVC was used to encode a side-by-side frame-compatible video and decode the video in Experiments A1 and A2 (system A). H.264/MVC was used to encode a frame-sequential video and decode the video in Experiments B1 and B2 (system B). We used H.264/AVC to separately encode a frame-sequential video and separately decode the video in Experiments C1 and C2 (system C), where the same H.264/AVC was used for both left and right-view video signals. Here, we denote the term Experiment XY, where X indicates the system (i.e., system A, B, or C) and Y indicates the SRC group (i.e., group 1 or 2). Encoding parameters can be found in Table 2.

The required overall bit rates for both left and right views were 1 to 16 Mbps in system A, 2 to 32 Mbps in system B, and 1.5 to 32 Mbps in system C, as listed in Table 3. In system B, the bit rate of the left view was about twice as large as that of the right view. In system C, to cover a wide range of 2D video quality for left and right views, the bit rate ratio of the left view to the right view ranged from 1:1 to 16:1.

Table 1 Spatial information (SI) and temporal information (TI).

Group 1		Left		Right	
SRC No.	Title	SI	TI	SI	TI
1	Photo	65	5	71	5
2	Mirror	32	15	32	15
3	Bicycle (front face)	54	16	57	17
4	Basketball	67	47	63	47
5	Flower garden	71	50	72	50
6	Paint	67	57	68	57

Group 2		Left		Right	
SRC No.	Title	SI	TI	SI	TI
7	Well-dressed woman	29	10	32	10
8	Bicycle (side face)	63	27	72	28
9	Woman and maple leaves	42	8	47	8
10	Cheerleaders	55	60	63	60
11	Clown	122	76	118	76
12	Tropical fish	65	21	64	21

Table 2 Experimental settings.

(a) Codecs.

Experiments A and C	H.264/AVC
Experiment B	H.264/MVC

(b) Coding parameters.

Profile	High profile
Video format	1920 × 1080p
Chroma format	4:2:0
Frame rate	24 fps
GoP	M=3, N=24 for H.264/AVC, M=3, N=22 for H.264/MVC

In total, 36 3D processed video sequences (PVSs) were used (6 HRCs (hypothetical reference circuits) × 6 SRCs) for Experiments A1, A2, B1, and B2, and 126 (21 HRCs × 6 SRCs) were used for Experiments C1 and C2. The total number of 2D PVSs was 72 (6 HRCs × 6 SRCs × 2 views) for Experiments A1, A2, B1, and B2, and 84 (5 HRCs × 6 SRCs for the left view + 9 HRCs × 6 SRCs for the right view) for Experiments C1 and C2. In subjective quality assessments for both 3D and 2D video, the uncompressed reference SRCs were used as an anchor, but scores for the uncompressed reference SRC were not used in the statistical analysis.

Here, subjective characteristics depend on the codec implementation, so we compare rate-distortion (RD) curves (i.e., bit rate vs. PSNR) for systems A, B, and C, where PSNR was computed using all frames for left and right views and where the bit rate and PSNR correspond to, respectively, the average bit rate and PSNR over all the considered 12 video sequences. In system A, we calculated the PSNR after up-sampling in the horizontal direction and post-filtering. Figure 2 shows that the RD curve for system B is higher than that of both systems A and C and that the RD curve for system A is almost the same as that for system C, where the curve for system C is depicted for each bit rate ratio of the left view to the right view, as shown in Fig. 2(b).

In the subjective quality assessment, the 3D video qual-

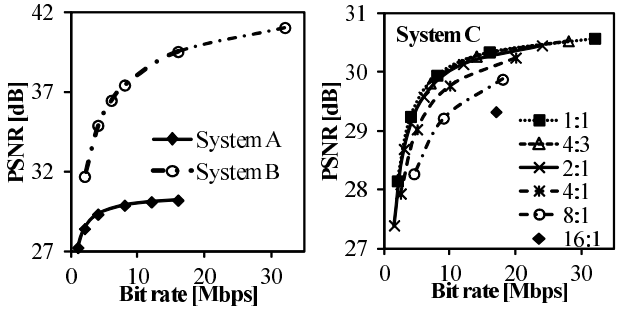
Table 3 Overall bit rates for systems A, B, and C.

(a) Systems A and B.

System A	16, 12, 8, 4, 2, and 1 Mbps
System B	32, 16, 8, 6, 4, and 2 Mbps

(b) System C.

Overall bit rate (Mbps)	Bit rate for left view (Mbps)	Bit rate for right view (Mbps)
32	16	16
28	16	12
24	16	8
20	16	4
18	16	2
17	16	1
16	8	8
14	8	6
12	8	4
10	8	2
9	8	1
8	4	4
7	4	3
6	4	2
5	4	1
4.5	4	0.5
4	2	2
3	2	1
2.5	2	0.5
2	1	1
1.5	1	0.5



a) Systems A and B. b) System C.

Fig. 2 RD curves.

Table 4 Five-grade quality scale.

Score	Quality scale (in Japanese)
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

ity was evaluated using an absolute category rating (ACR) with a five-grade scale [23], as listed in Table 4. The subjects were required to rate the quality within 5 seconds after a video sequence was presented. The presentation order of PVSs was randomized in these tests. 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, left view, and right view, respectively. Before starting the subjective test, we conducted screening tests. We used two tests indicated in ITU-R Recommendation BT. 1438 [24]: Coarse and Fine stereopsis tests. We also screened the subjects for visual acuity and color vision. Thirty-two male and female subjects who passed the screening tests participated in the subjective test. They ranged in age from 20 to 39 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 1920×1080 native resolution on a 40-inch LCD 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

First, we describe the stability of our subjective quality assessment. We calculated the 95% confidence interval (CI) using the t-student distribution for the 32 subjects. The CI was calculated as follows:

$$CI = 2.040\frac{\sigma}{\sqrt{32}},$$

where σ represents the standard deviation of the score per subject. Figure 3 shows the relationship between MOS and CI. The average CI for the 3D video was 0.261, the maximum CI was 0.377, and the minimum CI was 0.000[†]. The average CI for the 2D video was 0.253, the maximum CI was 0.396, and the minimum CI was 0.000. Brotherton et al. obtained average CIs for two subjective 2D video quality tests of 0.36 and 0.31 [25]. Because the average CIs in our tests were lower than these values, it can be said that stability was not an issue in our subjective quality assessment of 3D and 2D video.

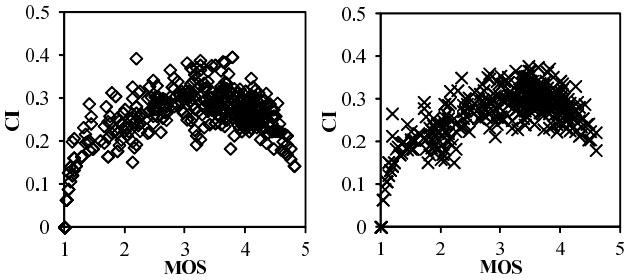
From this point, we use subjective data for video group 1 because characteristics are not distinguishable when there are so many plots in a figure. We confirmed that the following described findings were almost the same as those for video group 2.

We investigated the relationship between MOS_OS and either MOS_LS or MOS_RS. In system A, MOS_OS highly correlated with both MOS_LS and MOS_RS (Figs. 4(a) and(b)). In systems B and C, MOS_OS correlated with both MOS_LS and MOS_RS in some plots, while it did not correlate with either MOS_LS or MOS_RS in the other plots.

We also investigated in detail whether the absolute value of the difference between MOS_LS and MOS_RS, dMOS_LR ($ABS(MOS\_LS - MOS\_RS)$[††]), affected MOS_OS. Figures 5(a) and (b) show the relationship between MOS_OS, MOS_LS, and MOS_RS for respective bit rates of 32 Mbps and 17 Mbps in system C. As shown in Fig. 5(a), when dMOS_LR was small, MOS_OS was almost the same as MOS_LS and MOS_RS. However, when dMOS_LR was significantly large, as shown in Fig. 5(b),

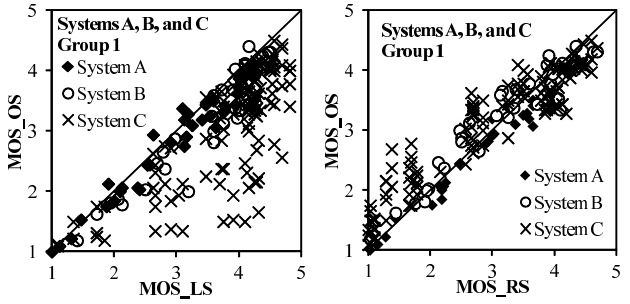
[†] $CI = 0$ means that all subjects rated the same score.

[††] $ABS()$ function calculates the absolute value.



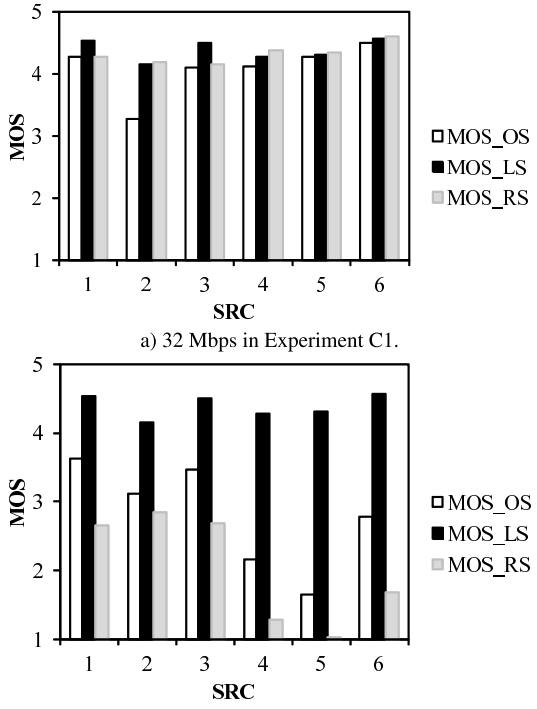
a) 3D video. b) 2D video.

Fig. 3 MOS vs. CI.



a) MOS_LS vs. MOS_OS. b) MOS_RS vs. MOS_OS.

Fig. 4 Relationship between MOS_OS and either MOS_LS or MOS_RS.



a) 32 Mbps in Experiment C1.

b) 17 Mbps in Experiment C1.

Fig. 5 Relationship among MOS_OS, MOS_LS, and MOS_RS.

MOS_OS decreased from the highest MOS of two views due to the asymmetric 2D video quality. This characteristic implies that dMOS_LR affects MOS_OS. Next, as also shown in Fig. 5(b), even when the overall bit rate per SRC was the same, the MOS_OS for the SRCs differed greatly. In addition, the MOS_OS varied with dMOS_LR.

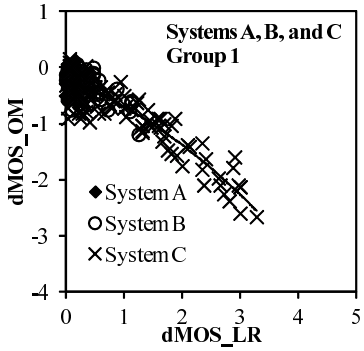


Fig. 6 dMOS_LR vs. dMOS_OM.

Next, we investigated how the dMOS_LR affected MOS_OS. Here, we define dMOS_OM as the value representing the difference between MOS_OS and the maximum value between MOS_LS and MOS_RS ($MOS\_OS - MAX(MOS\_LS, MOS\_RS)$[†]). Figure 6 plots the relationship between dMOS_LR and dMOS_OM. With system A, the plots were scattered in the upper left side of the graph because MOS_LS was basically almost the same as MOS_RS. In addition, when MOS_LS = MOS_RS, dMOS_OM was nearly equal to 0 (zero) because MOS_OS was almost the same as MOS_LS and MOS_RS. With systems B and C, dMOS_OM varied with dMOS_LR. In addition, the characteristic was expressed by a quadratic function, as shown in Fig. 6.

We can summarize subjective quality characteristics as follows. One characteristic showed that when dMOS_LR was small, MOS_OS highly correlated with either MOS_LS or MOS_RS due to binocular suppression [15]–[19]. The other characteristic suggested that dMOS_OM varied with dMOS_LR, and the characteristic was expressed by a quadratic function. As described in Sect. 1, we showed to what extent 3D video quality is affected by the difference in 2D video quality between left and right views in detail and modeled this as the novelty of this study.

4. Proposed Objective Quality Metric

We propose an objective quality metric that can be used for optimizing encoded 3D video quality and for monitoring 3D video quality (Fig. 7). As mentioned in Sect. 1, we developed the 3D video quality subjective metric taking subjective 2D video quality for left and right views as input in this study because conventional objective quality metrics for 2D video (e.g., ITU-T Recommendation J.341) can be used to estimate the 2D video quality for left and right views.

As described in Sect. 3, 3D video quality is affected by the video quality difference between left and right views, and when the difference is small, 3D video quality can be expressed by 2D video quality for either left or right view.

[†]$MAX()$ function calculates the maximum value.

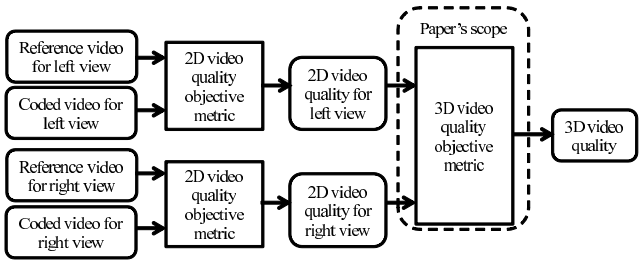


Fig. 7 Proposed objective quality metric.

Therefore, we model these characteristics as follows;

$$\begin{aligned} MOS\_OO &= a + b \cdot MAX(MOS\_LS, MOS\_RS) \\ &\quad + c \cdot ABS(MOS\_LS - MOS\_RS) \\ &\quad + d \cdot (MOS\_LS - MOS\_RS)^2, \\ &\approx a + b \cdot MAX(MOS\_LO, MOS\_RO) \\ &\quad + c \cdot ABS(MOS\_LO - MOS\_RO) \\ &\quad + d \cdot (MOS\_LO - MOS\_RO)^2, \end{aligned} \tag{1}$$

where, a, b, c, and d are coefficients that can be optimized using a training data set and multi-regression analysis, and MOS_LS and MOS_RS represent subjective 2D video quality for the left and right views, MOS_OO represents objective 3D video quality, and MOS_LO and MOS_RO represent objective 2D video quality that can be estimated by using a conventional 2D video quality metric.

5. Performance Evaluation of Proposed Model

5.1 Performance Requirements

This section describes our target quality-estimation accuracy. We compared the performance of our proposed model to that of a conventional model using the following conventional model [20], which estimates 3D video quality using the average 2D video quality between left and right views:

$$\begin{aligned} MOS\_OO &= e + f \cdot (MOS\_LS + MOS\_RS)/2 \\ &\approx e + f \cdot (MOS\_LO + MOS\_RO)/2, \end{aligned} \tag{2}$$

where e and f are coefficients that can be optimized using a training data set and regression analysis and the definitions of MOS_OO, MOS_LS, MOS_RS, MOS_LO, and MOS_RO are the same as those of Eq. (1).

A criterion is that the performance of our proposed model optimized by using the training data set (i.e., Experiment C1) should be better than that of the conventional model. That is, we investigated whether our proposed model could more accurately evaluate the 3D video quality for unknown SRCs (i.e., video group 2) and unknown systems (i.e., systems A and B) than the conventional model.

We compared the performance using the Pearson's correlation coefficient (PCC), root mean square error (RMSE), and outlier ratio (OR), which is based on whether the difference between MOS_OS and MOS_OO is larger than the CI described in Sect. 3.

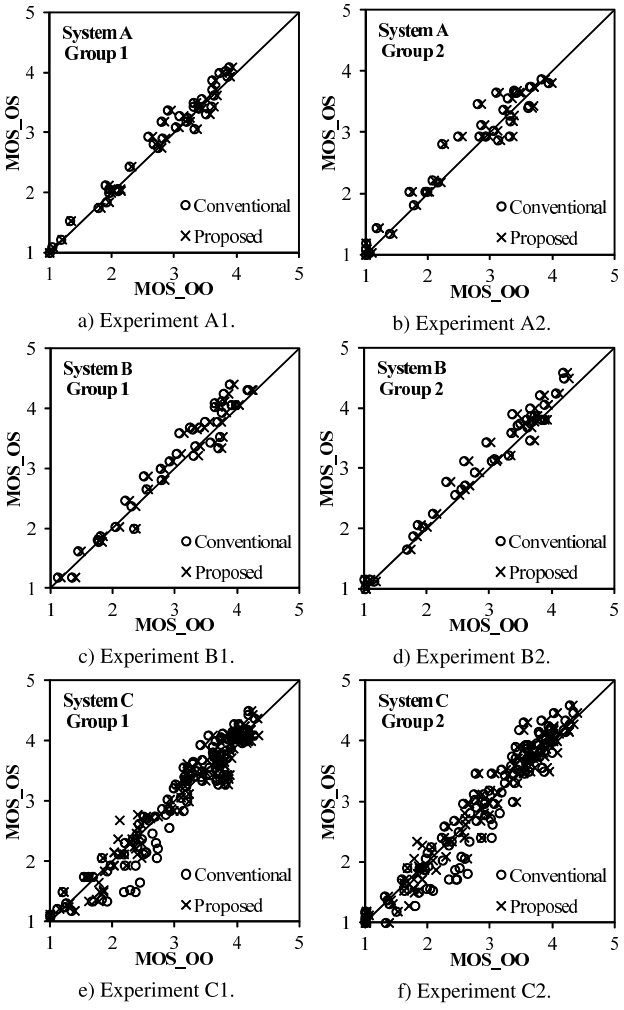
Table 5 Coefficients of proposed and conventional models.

(a) Proposed model

Coefficient	Value
a	0.000
b	0.922
c	-0.329
d	-0.104

(b) Conventional model

Coefficient	Value
e	0.000
f	0.912



a) Experiment A1. b) Experiment A2.
c) Experiment B1. d) Experiment B2.
e) Experiment C1. f) Experiment C2.

Fig. 8 MOS_OO vs. MOS_OS.

5.2 Performance of Proposed Model

We compared the performance of our proposed model with that of the conventional model by calculating the coefficients a, b, c, and d of our proposed model and coefficients e and f of the conventional model using the training data set (Experiment C1), as listed in Table 5. The coefficients of the proposed and conventional models were significant at the 5% level, where coefficients a and e were set to be 0 because coefficients a and e were not significant at the 5% level. We used Experiments A1, B1, A2, B2, and C2 as the unknown data set.

Table 6 PCCs for proposed and conventional models.

Experiment	Proposed	Conventional
A1	0.98	0.98
A2	0.97	0.97
B1	0.97	0.97
B2	0.99	0.98
C1	0.97	0.96
C2	0.98	0.95
A1, B1, and C1	0.97	0.96
A2, B2, and C2	0.98	0.96
All PVSs	0.97	0.96

Table 7 RMSEs for proposed and conventional models.

Experiment	Proposed	Conventional
A1	0.16	0.18
A2	0.24	0.26
B1	0.24	0.27
B2	0.21	0.25
C1	0.23	0.28
C2	0.23	0.30
A1, B1, and C1	0.22	0.26
A2, B2, and C2	0.23	0.29
All PVSs	0.23	0.27

Table 8 ORs for proposed and conventional models.

Experiment	Proposed	Conventional
A1	0.08	0.14
A2	0.22	0.25
B1	0.31	0.36
B2	0.25	0.28
C1	0.20	0.25
C2	0.22	0.33
A1, B1, and C1	0.20	0.25
A2, B2, and C2	0.23	0.30
All PVSs	0.21	0.28

Figure 8 plots the relationships between MOS_OO and MOS_OS. PCCs, RMSEs, and ORs for our proposed and conventional models are listed in Tables 6, 7, and 8, where all performance comparisons were done under the condition that models were trained using Experiment C1. The validity of our proposed model for unknown systems and SRCs was verified because the performance of the proposed model was better than that of the conventional model in all experiments. We can conclude from these results that our proposed model can be applied to optimize the 3D video processing chain and monitor 3D video quality.

5.3 Considerations

Some considerations should be noted to explain the results in detail. Some scattered plots indicated low performance in the conventional model, especially in systems B and C. To investigate this trend, we list the RMSE per SRC in Table 9. Basically, the RMSEs for our proposed model were lower than those for the conventional model, excluding SRCs 2 and 3. We investigated why the improvement depends on SRC by defining dMOS_LRavg, which represents the dMOS_LR averaged over all HRCs. Table 10 lists the dMOS_LRavg per SRC. As listed in Tables 9 and 10, when dMOS_LRavg was large, the RMSE improvement was also large.

Table 9 RMSE per SRC for proposed and conventional models.

SRC	Proposed	Conventional	Improvement ratio
1	0.23	0.27	15%
2	0.27	0.26	−4%
3	0.21	0.19	−10%
4	0.22	0.29	24%
5	0.16	0.27	41%
6	0.20	0.27	25%
7	0.26	0.29	11%
8	0.21	0.24	14%
9	0.16	0.26	36%
10	0.20	0.28	30%
11	0.19	0.30	38%
12	0.33	0.34	5%

Table 10 dMOS_LRavg per SRC.

SRC	dMOS_LRavg	SRC	dMOS_LRavg
1	0.59	7	0.44
2	0.50	8	0.53
3	0.66	9	0.92
4	0.74	10	0.75
5	0.78	11	0.86
6	0.72	12	0.71

Table 11 Performance comparison.

	Proposed	Conventional
$0 \leq$ dMOS_LR ≤ 1	0.22	0.24
$1 <$ dMOS_LR	0.23	0.35

Next, we investigated why our proposed model improves the RMSE performance, in comparison with the conventional model. Table 11 lists the RMSE for the proposed and conventional models in the range of $0 \leq$ dMOS_LR ≤ 1 and $1 <$ dMOS_LR. The results show that our proposed model is effective in improving the RMSE performance when there is a significant difference in video quality between the left and right views in the PVSs.

6. Conclusion

We proposed a novel metric that can be applied to calculate the difference in 2D video quality between left and right views. We pointed out several issues in conventional studies regarding the effect of the difference in 2D video quality between left and right views on 3D video quality. We investigated the effect of this difference on 3D video quality by conducting subjective quality assessments. The results showed that the 3D video quality did not depend on 2D video quality of either the left or right view only; rather, the 3D video quality was affected by the difference in 2D

video quality between left and right views; additionally, the dMOS_OM was expressed by the quadratic function with the variable dMOS_LR.

We compared the performance of our proposed model to that of the conventional model, which is based on the average 2D video quality for left and right views. The results indicated that our proposed model performed better than the conventional model. In particular, we found that when the difference in 2D video quality between left and right views was large, our proposed model was able to more accurately evaluate the 3D video quality than the conventional model.

The following issue calls for further study. We used subjective quality of left and right views as the input of our proposed model in this work. To objectively estimate 3D video quality without use of subjective quality of left and right views, we need to incorporate an objective quality metric for 2D video (e.g., ITU-T Recommendation J.341) into the proposed model and to verify the validity of the resultant model.

References

[1] DVB Document A154, "Digital video broadcasting (DVB); Frame compatible plano-stereoscopic 3DTV (DVB-3DTV)," Feb. 2011.
[2] W.J. Tam and L.B. Stelmach, "Psychovisual aspects of viewing stereoscopic video sequences," Proc. SPIE, vol.3295, pp.226–235, Jan. 1998.
[3] G. Leon, H. Kalva, and B. Furht, "3D video quality evaluation with depth quality variations," 3DTV-CON, pp.301–304, May 2008.
[4] M.T.M. Lambooij, W.A. IJsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: A review," Proc. SPIE, vol.6490, pp.64900I-1–13, Jan. 2007.
[5] K. Ukai, "Human factors for stereoscopic images," IEEE ICME, pp.1697–1700, July 2006.
[6] H. Yamanoue, M. Okui, and F. Okano, "Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images," IEEE Trans. Circuits Syst. Video Technol., vol.16, no.6, pp.744–752, June 2006.
[7] B.F. Col and K. O'Connell, "3DTV at home: Status, challenges and solutions for delivering a high quality experience," International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), Jan. 2010.
[8] Q. Huynh-Thu, P. Le Callet, and M. Barkowsky, "Video quality assessment: From 2D to 3D - challenges and future trends," International Conference on Image Processing (ICIP), pp.4025–4028, Sept. 2010.
[9] P. Merkle, K. Muller, and T. Wiegand, "3D Video: Acquisition, coding, and display," IEEE Trans. Consum. Electron., vol.56, no.2, pp.946–950, Jan. 2010.
[10] ITU-T Recommendation H.262, "Information technology — Generic coding of moving pictures and associated audio information: Video," Feb. 2000.
[11] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services," March 2010.
[12] K. Kim, J. Lee, D. Suh, and G. Park, "Efficient stereoscopic contents file format on the basis of ISO base media file format," Proc. SPIE, vol.7256, pp.72560N-1–9, Jan. 2009.
[13] K. Yamagishi, L. Karam, J. Okamoto, and T. Hayashi, "Subjective characteristics for stereoscopic high definition video," International Workshop on Quality of Multimedia Experience (QoMEX), pp.37–42, Sept. 2011.
[14] L. Stelmach, W. Tam, D. Meegan, and A. Vincent, "Stereo image quality: Effects of mixed spatio-temporal resolution," IEEE Trans. Circuits Syst. Video Technol., vol.10, no.2, pp.188–193, March 2000.
[15] P. Aflaki, M.M. Hannuksela, J. Hakkinen, P. Lindroos, and M. Gabbouj, "Subjective study on compressed asymmetric stereoscopic video," International Conference on Image Processing (ICIP), pp.4021–4024, Sept. 2010.
[16] G. Saygili, C.G. Gurler, and A.M. Tekalp, "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming," IEEE Trans. Broadcast., vol.57, no.2, pp.593–601, June 2011.
[17] G. Saygili, C.G. Gurler, and A.M. Tekalp, "Quality assessment of asymmetric stereo video coding," Int. Conf. on Image Processing (ICIP), pp.4009–4012, Sept. 2010.
[18] A. Vetro, A.M. Tourapis, K. Müller, and T. Chen, "3D-TV content storage and transmission," IEEE Trans. Broadcast., vol.57, no.2, pp.384–394, June 2011.
[19] V.D. Silva, H.K. Arachchi, E. Ekmekcioglu, A. Fernando, S. Dogan, A. Kondoz, and S. Savas, "Psycho-physical limits of interocular blur suppression and its application to asymmetric stereoscopic video delivery," IEEE 19th Int. Packet Video Workshop (PV2012), pp.184–189, May 2012.
[20] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment," Proc. 15th European Signal Processing Conference (EURASIP), pp.2110–2114, Sept. 2007.
[21] ITU-T Recommendation J.341, "Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference," Jan. 2011.
[22] ITU-R Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," April 2008.
[23] ITU-R Recommendation BT.500–12, "Methodology for the subjective assessment of the quality of television pictures," Sept. 2009.
[24] ITU-R Recommendation BT.1438, "Subjective assessment of stereoscopic television pictures," March 2000.
[25] M.D. Brotherton, Q. Huynh-Thu, D. Hands, and K. Brunnström, "Subjective multimedia quality assessment," IEICE Trans. Fundamentals, vol.E89-A, no.11, pp.2920–2932, Nov. 2006.

Kazuhisa Yamagishi received his B.E. degree in Electrical Engineering from Tokyo University of Science and M.E. degree in Electronics, Information, and Communication Engineering from Waseda University in Japan in 2001 and 2003. He joined NTT Laboratories in 2003. He has been engaged in subjective quality assessment of multimedia telecommunications and image coding. Currently, he is 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 Investigators' Award (IEICE) in Japan in 2007 and the Telecommunication Advancement Foundation Award in Japan in 2008.

Taichi Kawano received his B.E. and M.E. degrees in Engineering from the University of Tsukuba, Ibaraki, in 2006 and 2008. He joined NTT Laboratories in 2008. He has been engaged in subjective and objective 2D/3D video quality assessment for Internet protocol television (IPTV) services. He received the Young Investigators' Award (IEICE) in Japan in 2011.

Takanori Hayashi received his B.E., M.E., and Ph.D. degrees in Engineering from the University of Tsukuba, Ibaraki, in 1988, 1990, and 2007. He joined NTT Laboratories in 1990 and has been engaged in the quality assessment of multimedia telecommunication and network performance measurement methods. Currently, he is the manager of the Service Assessment Group at NTT Laboratories. He received the Telecommunication Advancement Foundation Award in Japan in 2008.

Jiro Katto received his B.S., M.E., and Ph.D. degrees in electrical engineering from the University of Tokyo in 1987, 1989 and 1992. He worked for NEC Corporation from 1992 to 1999. He was also a visiting scholar at Princeton University, NJ, USA, from 1996 to 1997. He then joined Waseda University in 1999, where he is now a professor at the Department of Computer Science, School of Fundamental Science and Engineering. His research interest is in the field of multimedia signal processing and multimedia communication systems such as the Internet and mobile networks. He received the Best Student Paper Award at SPIE's conference of Visual Communication and Image Processing in 1991, and received the Young Investigator Award of IEICE in 1995. He is a member of the IEEE and the IPSJ.