IEICE TRANS. COMMUN., VOL.E97–B, NO.4 APRIL 2014

PAPER

Performance Comparison of Subjective Assessment Methods for Stereoscopic 3D Video Quality

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

SUMMARY The International Telecommunication Union has standardized many subjective assessment methods for stereoscopic three-dimensional (3D) and 2D video quality. The same methods are used for 3D and 2D videos. The assessment time, stability, and discrimination ability, which means the ability to identify differences in video quality, are important factors in subjective assessment methods. Many studies on these factors have been done for 2D video quality. However, these factors for 3D video quality have not been sufficiently studied. To address this, we conduct subjective quality assessments for 3D and 2D videos using the absolute category rating (ACR), degradation category rating (DCR), and double stimulus continuous quality-scale (DSCQS) methods that are defined in ITU Recommendations. We first investigate the Pearson's correlation coefficients and Spearman's rank correlation coefficients between different pairings of the three methods to clarify which method is most efficient in terms of assessment time. The different pairings of the three methods exhibit high coefficients. These results indicate that the order relation of the mean opinion score (MOS) and the distance between the MOSs for these methods are almost the same. Therefore, for generally investigating the quality characteristics, the ACR method is most efficient because it has the shortest assessment time. Next, we analyze the stability of these subjective assessment methods. We clarify that the confidence intervals (CIs) of the MOSs for 3D video are almost the same as those for 2D video and that the stability of the DCR method is higher than that of the other methods. The DSCQS method has the smallest CIs for high-quality video. Finally, we investigate the discrimination ability of these subjective assessment methods. The results show that the DCR method performs better than the others in terms of the number of paired MOSs with a significant difference for low-quality video. However, we confirm that the DSCQS method performs better than the others for high-quality video.

key words: *ACR, DCR, DSQCS, 3D, 2D*

1. Introduction

Stereoscopic three-dimensional (3D) video has become widely popular. The International Telecommunication Union — Radiocommunications Sector (ITU-R) Recommendation BT.2021 [1] defines that the quality of experience (QoE) of 3D video is mainly affected by picture quality (video quality), depth quality, and visual comfort. We focus here on the video quality because it is the most important QoE factor. 3D video quality depends on the processing chain that consists of acquisition, coding (encoding and decoding), transmission, and display, as shown in Fig. 1. Because 3D video quality depends on the degradation factors present in the chain, service providers need to design their services based on the relationship between degradation factors and 3D video quality.

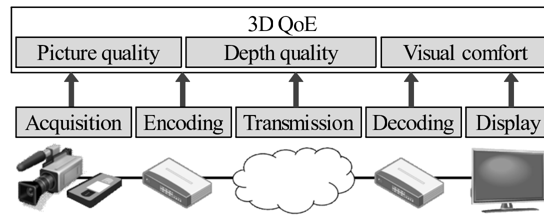


Fig. 1 Relationship between 3D video processing chain and QoE.

To design a service, a method to subjectively assess for 3D video quality is necessary. ITU has standardized many video quality assessment methods. ITU-R Recommendations BT.1438 [2] and BT.2021 [1] define the absolute category rating (ACR), degradation category rating (DCR), double stimulus continuous quality-scale (DSCQS), single stimulus continuous quality evaluation (SSCQE), and pair-comparison (PC) methods. These methods are the same as those for 2D video quality [3]. The assessment time, stability, and discrimination ability, which refers to the ability to identify differences in video quality, are important factors in these subjective assessment methods.

Several studies have compared these methods for 2D video quality in terms of the correlation of their mean opinion scores (MOSs) [4]–[6]. In addition, the assessment time and stability of these methods for 2D video quality have been extensively studied [7], [8]. In contrast, such results for 3D video quality have not been derived. However, it is also important to clarify the relationship between subjective assessment methods for 3D video quality because it is not clear whether the characteristics of 3D video such as the stereoscopy and difference in image quality between the left and right views affects subjective assessment methods for 3D video quality. In addition, it is important to investigate the discrimination ability of subjective assessment methods.

We investigated the correlation between the MOSs for different pairings of subjective assessment methods and the stability of these methods as regards 3D video quality [9]. We also compared these methods in terms of the discrimination ability in the present study. We conducted subjective tests for 3D and 2D videos using the widely used ACR, DCR, and DSCQS methods. We first investigated the Pearson's correlation coefficient (PCC) and Spearman's rank correlation coefficient (SRCC) for MOSs derived from pairing these methods to clarify which method is most efficient in terms of assessment time. Next, we compared the

Manuscript received June 28, 2013.
Manuscript revised November 8, 2013.

[†]The authors are with NTT Network Technology Laboratories, NTT Corporation, Musashino-shi,180-8585 Japan.
a) E-mail: kawano.taichi@lab.ntt.co.jp

DOI: 10.1587/transcom.E97.B.738

confidence interval (CI) of MOSs for 3D and 2D videos to investigate the stability of these methods when applied to assess 3D video quality. We also compared the stability of these methods in assessing 3D video quality. We then investigated the discrimination ability of these subjective assessment methods using paired MOSs that had a significant difference. Finally, from these results, we summarized which subjective assessment method for 3D video quality is most suitable in terms of the assessment time, stability, and discrimination ability.

This paper is organized as follows. Section 2 describes related work. Section 3 explains the three subjective assessment methods used for this study. Section 4 describes the experimental conditions, and Sect. 5 discusses the experimental results. Finally, Sect. 6 presents the conclusion and discusses further studies.

2. Related Work and Issues

Many studies have investigated the correlation between the MOSs of two subjective assessment methods for 2D video quality as well as the stability of these methods. Corriveau et al. [7] compared the DCR and DSCQS methods using standard definition 2D video. They found that the MOSs of these methods are highly correlated. Comparison studies of the DSCQS method with either the ACR or ACR with hidden reference (ACR-HR) [10] method using 2D video with low resolution have also been reported [4], [5]. They conclude that the correlation between the MOSs of these methods are high. Tominaga et al. [8] conducted several subjective tests for 2D video with quarter video graphics array resolution using the ACR, DCR, DSCQS, and ACR-HR methods. The results indicate that the MOSs of these methods are highly correlated. They also clarified that the CIs derived from these methods are almost the same. As described above, subjective assessment methods for 2D video quality have been sufficiently investigated.

Because a comparison study among subjective assessment methods for 3D video quality has not been sufficiently conducted, a user of Recommendation BT.2021 chooses a method based on his/her own criteria. For example, Yamagishi et al. [11] and Xing et al. [12] used the ACR method, Kim et al. [13] used the DCR method, and Yasakethu et al. [14] and Chen et al. [15] used the DSCQS method to investigate the quality characteristics of 3D video. However, it is difficult to compare and/or combine their research results because the relationship between these subjective assessment methods for 3D video quality is unclear. Therefore, it is necessary to clarify the suitable domain of applicability for each subjective assessment method for 3D video quality. Ozbek et al. [16] compared the stability between the DSCQS and their proposed method based on the subjective assessment methodology for video quality (SAMVIQ) [17] for 3D video quality in order to validate their proposed method. The results indicate that their proposed method has higher stability than the DSCQS method for low bit rate video. This research is of great value in clarifying the performance of subjective assessment methods. However, Ozbek et al. [16] did not use the ACR and DCR methods even though these methods are as widely used as the DSCQS method. That is, the correlation between the MOSs for different pairings of the three methods and the stability and discrimination ability of these methods has not been clarified. Therefore, it is necessary to investigate the correlation between MOSs derived using different pairings of these methods as well as their stability and discrimination ability.

3. Subjective Quality Assessment Methods

We discuss the ACR, DCR, and DSCQS methods. The procedures used to assess these methods are shown in Fig. 2. These methods are applied to assess both 3D and 2D videos.

In the ACR method, evaluators are shown once with the target video once for 10 seconds. Then, the evaluators assess the target video quality by scoring it based on a discrete scale of five categories within 10 seconds, as shown in Fig. 2 and Table 1. In our study, the assessment time was five seconds, so the assessment unit was 15 seconds.

In the DCR method, the source video is presented first for 10 seconds, and then the target video to be rated is presented for 10 seconds. The evaluators know that the former is the source video. A three-second gray video is inserted between the source and target videos. Then, the evaluators assess the impairments in the target video in relation to the source video by scoring it based on a discrete scale of five

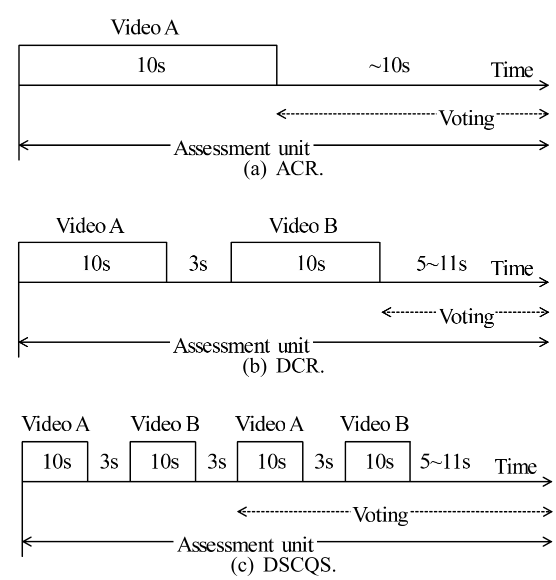


Fig. 2 Subjective assessment methods.

Table 1 Rating scale in ACR.

Score	Quality scale
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Table 2 Rating scale in DCR.

Score	Impairment scale
5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

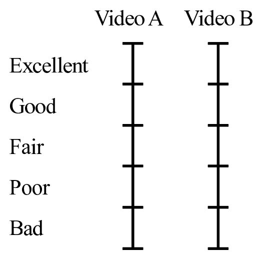


Fig. 3 Rating scale with DSCQS.

categories for five to 11 seconds, as shown in Fig. 2 and Table 2. In our study, the assessment time was seven seconds, so the assessment unit was 30 seconds.

In the DSCQS method, each pair of source and target videos is presented twice. Each video is presented for 10 seconds, and a three-second gray video is inserted between them. The display order of the source and target videos is random, and the evaluators do not know which of these videos is the source video. The source and target video quality is individually assessed by inserting a mark on a continuous scale from zero to 100, which is divided into five equally spaced categories for five to 11 seconds, as shown in Figs. 2 and 3. For our study, the assessment time was 11 seconds, so the assessment unit was 60 seconds. The opinion scores of the DSCQS method were calculated as the absolute difference between the opinion scores for the source and target videos.

For our study, the rating scales for the ACR, DCR, and DSCQS were given in Japanese.

To analyze the subjective test results, the opinion scores derived using each method are averaged over all evaluators as follows:

$$\mathrm{MOS}(i, j) = \frac{1}{K} \times \sum_{k=1}^{K} \mathrm{OS}(i, j, k), \tag{1}$$

where i is the index of the subjective assessment method (i.e., ACR, DCR, or DSCQS method), j is the index of the processed video sequence (PVS), k is the index of the evaluator, K is the number of evaluators, and $\mathrm{OS}(i, j, k)$ is the opinion score.

4. Experimental Conditions

We conducted subjective quality assessment for 3D and 2D videos using the ACR, DCR, and DSCQS methods and then compared these methods.

We used eight 3D videos (10 seconds each) provided by the National Institute of Information and Communications Technology (NICT) [18] in Japan and the Digital Content Association of Japan (DCAJ) [19], as listed in Table 3. We simulated a frame-sequential 3D video format consisting of a sequence of alternating left and right frames because this format is widely used, e.g., in 3D Blu-ray. The video resolution for each view was Full-HD (1920 × 1080p). We used the left view of the uncompressed 3D video as the source video in the 2D video quality assessment. The temporal perceptual information measurement (TI) and spatial perceptual information measurement (SI) [10] for these eight videos (left view of uncompressed 3D video) are also listed in Table 3.

The codec and coding parameters listed in Table 4 were used to vary the video quality. Each view of 3D video was individually encoded in this experiment. We used six different bit rates for the 3D video and three for the 2D video, as listed in Table 5. Five out of six bit-rate conditions for 3D video were asymmetric between both views in order to simulate the asymmetric coding that is generically used for frame-sequential 3D video formats to improve compression efficiency.

Although uncompressed source videos were assessed using the ACR method, they were not evaluated using the DCR and DSCQS methods (i.e., the uncompressed source

Table 3 Video content.

No.	Name	TI	SI	Source
1	Camera	5	65	NICT
2	Maple	8	42	NICT
3	Dress	10	29	NICT
4	Bicycle	27	63	NICT
5	Basketball	47	67	DCAJ
6	Field of flowers	50	71	DCAJ
7	Cheerleader	60	55	DCAJ
8	Pierrot	76	122	DCAJ

Table 4 Coding conditions.

Parameter	Value
Codec type	H.264/AVC
Profile/level	High profile/level 4
Video format	1920 × 1080p
Frame rate	24 fps
GOP	M = 3, N = 15

Table 5 Bit rate conditions.

	Bit-rate [Mbps]		
	Total	Left view	Right view
3D	3	2	1
	4.5	3	1.5
	6	3	3
	6	4	2
	6	4.5	1.5
	12	8	4
	Uncompressed (for ACR only)		
2D	2	-	-
	3	-	-
	6	-	-
	Uncompressed (for ACR only)		

video was used only as a reference).

We used 56 PVSs for the 3D video and 32 PVSs for the 2D video in the ACR method. We used 48 PVSs for the 3D video and 24 PVSs for the 2D video in the DCR and DSCQS methods.

A subjective test of 3D video quality was carried out before that of 2D video quality. Before the subjective test was started, we conducted two screening tests specified in ITU-R Recommendation BT. 1438 [2]: Coarse and Fine stereopsis tests. We also screened the subjects for visual acuity and color vision. Forty-two evaluators who passed the screening tests participated in both experiments. They ranged in age from 20 to 39. All evaluators assessed all the PVSs for each method over a two-day period. The 42 evaluators were divided into six groups. Three subjective quality assessments were used in a different order for each group. Each group was divided into two subgroups that viewed PVSs in a different order. The presentation order of PVSs was randomized to reduce the order effect on the opinion score. Two evaluators viewed a single monitor simultaneously. Evaluators were centered evenly in front of the monitor and viewed each video at a distance of 3 H (3 H is about 150 cm; H indicates the picture height) from a 40-inch 3D monitor. The evaluators viewed the 3D video with polarized glasses and 2D video without polarized glasses. The room illuminance was 20 lux.

5. Experimental Results

5.1 Correlation Analysis for Subjective Assessment Methods

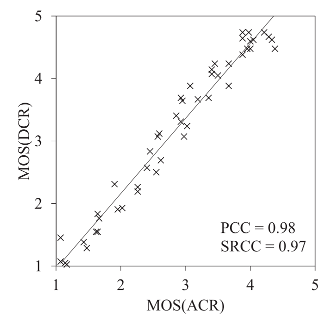
We investigated the PCCs and SRCCs between the MOSs for different pairings of the ACR, DCR, and DSCQS methods for 3D and 2D videos to clarify which method is most efficient in terms of assessment time, as shown in Figs. 4 and 5. All PCCs and SRCCs for the DSCQS method were negative because the MOS for the DSCQS method decreased as video quality increased. The figures show that the PCCs and SRCCs were high. These results indicate that the order relation of the MOSs and the distance between them for these methods are almost the same. Therefore, if the investigation objective is to clarify quality characteristics, the ACR method is more efficient than the others because it has the shortest total assessment time.

5.2 Stability of Subjective Assessment Methods

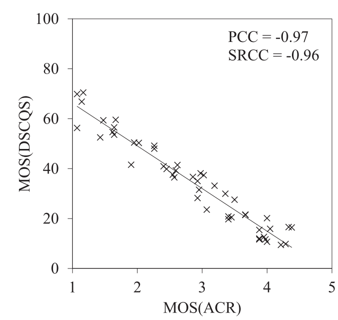
We investigated the stability of the subjective assessment methods for 3D video quality by comparing the 95% CI of the MOS for each method.

5.2.1 Comparison of Stability between 3D and 2D Videos

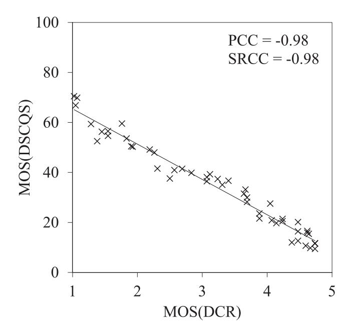
We compared the CI between 3D and 2D videos in order to clarify whether the stability of the subjective assessment methods for 3D video quality is the same as that for 2D video quality. The CI is calculated as follows,



(a) ACR vs. DCR.



(b) ACR vs. DSCQS.

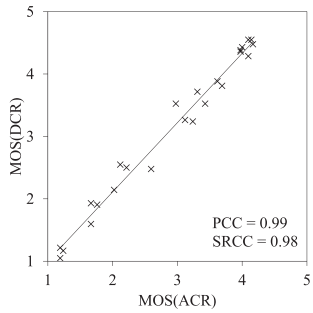


(c) DCR vs. DSCQS.

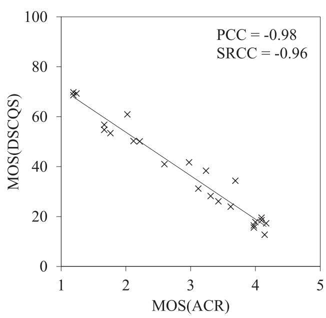
Fig. 4 Relationship between MOSs of subjective assessment methods for 3D video quality.

$$\mathrm{CI}(i, j) = t_{K-1}(0.95)\frac{s(i, j)}{\sqrt{K}}, \tag{2}$$

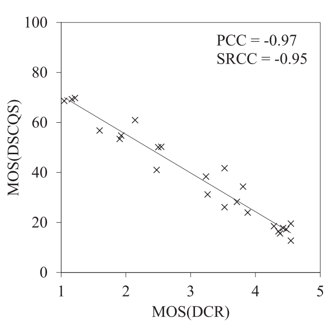
where $t_\phi()$ is the Student's t-distribution for ϕ degrees of freedom, i is the index of a subjective assessment method, j is the index of the PVS, $s(i, j)$ is the standard deviation of opinion scores of method i for PVS j, and K is the number of evaluators. The relationships between the MOSs and CIs for the ACR, DCR, and DSCQS methods for 3D and 2D videos are plotted in Fig. 6. As shown in the figure, the CIs for 3D videos were similar to those for 2D video regardless of the method. Tables 6, 7, and 8 list the mean CIs (MCIs) for 3D or 2D videos. Since the MCIs bear a non-linear relationship with the MOSs, we equally divide the MOSs into four ranges and calculate the MCI for each MOS range in order to treat them as a linear approximation. There are no



(a) ACR vs. DCR.

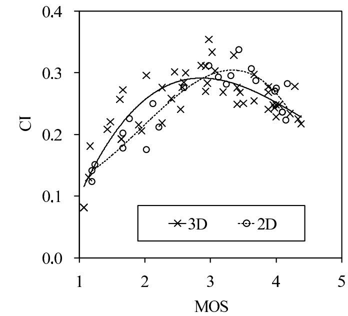


(b) ACR vs. DSCQS.

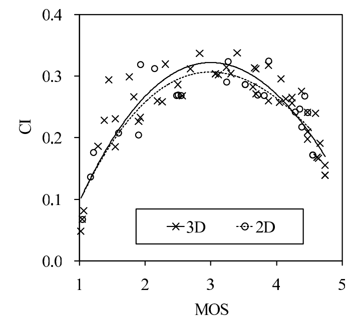


(c) DCR vs. DSCQS.

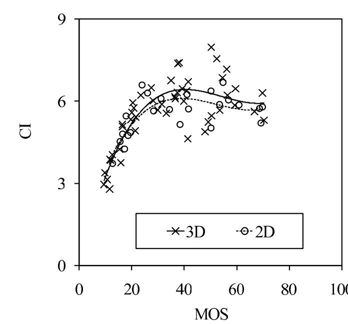
Fig. 5 Relationship between MOSs of subjective assessment methods for 2D video quality.



(a) ACR.



(b) DCR.



(c) DSCQS.

Fig. 6 Relationship between MOSs and CIs for 3D and 2D videos.

data in the lower MOS range for the DSCQS method. The Student's t-tests with a level of significance of 5% show that there is no significant difference in MCI between 3D and 2D videos for all subjective assessment methods regardless of the MOS range. These results indicate that the ACR, DCR, and DSCQS methods can be used to assess 3D video quality with the same stability as for 2D video quality.

5.2.2 Comparison of Stability among Subjective Assessment Methods

To clarify which subjective assessment method for 3D video quality is more suitable in terms of stability, we compared the CI, among the methods. To use the same scale for these methods, the MOSs of the ACR and DCR methods were converted to those of the DSCQS method using a linear function because there were high correlations among the MOSs of these methods, as stated in Sect. 5.1. Figure 7 plots the relationship between MOS(DSCQS(ACR)), MOS(DSCQS(DCR)), and the MOSs of the DSCQS method and CIs. MOS(DSCQS(ACR)) and MOS(DSCQS(DCR)) represent the MOSs of the DSCQS method converted from those of the ACR and DCR methods, respectively. Since the CIs of these methods vary depending on the MOS, we discuss the results for high-quality video (MOS(DSCQS) < 20) and low-quality video (MOS(DSCQS) >= 20)). As shown in the graph, the CI of the DCR method was smaller than that of the others regardless of video quality, and the ACR method was slightly worse than the DCR method in terms of CI. Furthermore, the CI of the DSCQS method was small for high-quality video; otherwise, it was large. Therefore, these results indicate that for low-quality video, the DCR method is most suitable in terms of stability and that the

Table 6 MCIs for ACR.

	MCI for 3D	MCI for 2D
MOS(ACR)≥4	0.24	0.22
4>MOS(ACR)≥3	0.28	0.29
3>MOS(ACR)≥2	0.28	0.25
2>MOS(ACR)	0.19	0.17

Table 7 MCIs for DCR.

	MCI for 3D	MCI for 2D
MOS(DCR)≥4	0.21	0.22
4>MOS(DCR)≥3	0.30	0.29
3>MOS(DCR)≥2	0.28	0.28
2>MOS(DCR)	0.20	0.18

Table 8 MCIs for DSCQS.

	MCI for 3D	MCI for 2D
MOS(DSCQS)≤25	4.34	4.84
25<MOS(DSCQS)≤50	6.14	5.83
50<MOS(DSCQS)≤75	6.17	5.84
75<MOS(DSCQS)	-	-

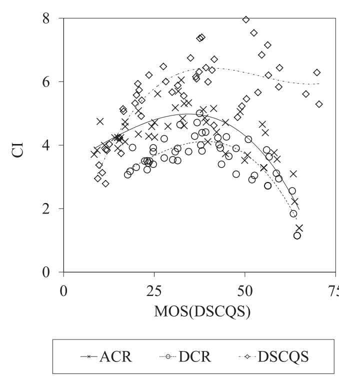


Fig. 7 Relationship between MOSs and CIs.

stability of the DSCQS and DCR methods is high for high-quality video. Furthermore, when the assessment time is factored in, the DCR method is found to be most suitable for all quality ranges. These findings were the same even when the MOSs of the ACR and DSCQS methods were converted to those of the DCR method, or the MOSs of the DCR and DSCQS methods were converted to those of the ACR method.

5.3 Discrimination Ability of Subjective Assessment Methods

We performed the Student's t-test with a level of significance of 5% among MOSs for each method to investigate their discrimination ability for 3D video quality. We used the number of significant differences (N) between MOS(i) and MOS(l) (l = 1, 2, $\cdots$, I, $l \neq i$), where i and l represent the PVS index and I is 48, the number of PVSs used for the analysis (i.e., eight uncompressed sources for the ACR method were not used). The MOS of a certain PVS was compared with those of the others. If N of a certain subjective assessment method is higher than that of the others, the method with higher N has a greater ability to identify differences in video quality.

We analyzed the relationship between the discrimination ability of the methods and video quality. We investigated the relationship between MOS and N using the MOSs of the DSCQS method since the MOSs of all methods were highly correlated, as explained in Sect. 5.1. Figure 8 shows the relationship between the MOSs of the DSCQS method and the Ns of the ACR, DCR, and DSCQS methods. These curves were expressed by a convex downward function because when a high MOS is compared with a low MOS and vice versa, there is a significant difference.

Since the Ns of these methods vary depending on the MOS, we discuss the results for high-quality video (MOS(DSCQS) < 20) and low-quality video (MOS(DSCQS) >= 20)). For high-quality video, the N of the DSCQS method was larger than that of the ACR or DCR method, and for low-quality video, the N of the DSCQS method was smaller than the others, as shown in Fig. 8. In other words, the discrimination ability of the DSCQS method was higher than the others for high-quality video and lower than the other methods for low-quality video.

We investigated why the discrimination ability of the DSCQS method decreases for low-quality video. Figure 9 shows a comparison between the MOSs of the DSCQS method and the standard variation of opinion scores for each PVS among all three methods. The opinion scores of the ACR and DCR methods are converted to those of the DSCQS method using a linear function in order to use the same scale for all methods. As shown in the graph, the standard variation of the DSCQS method was large for low-quality video. This is due to the additivity of variance (i.e., the variance of the difference between two sets is equal to the sum of each variance when two sets are independent). Therefore, when the quality of a certain PVS is high, the independence decreases in the DSCQS method. This is attributed to the fact that the quality of the PVS is similar to that of the source video. Therefore, the discrimination ability of the DSCQS method gradually increases as video quality decreases.

For low-quality video, N was larger in the DCR method than that of the others, as shown in Fig. 8. This is due to the fact that there are no issues with respect to the additivity of variance in the DCR method. For high-quality video, N of the DCR method was smaller than that of the DSCQS method.

In the ACR method, N was smaller than that for the DCR method regardless of the MOS, as shown in Fig. 8. This is attributed to the fact that the variance of the MOS increases due to the evaluators assessing the PVS without the source video in the ACR method, as shown in Fig. 9.

These results indicate that for high-quality video, the DSCQS method has a higher discrimination ability than the others; otherwise, the DCR method had the best performance.

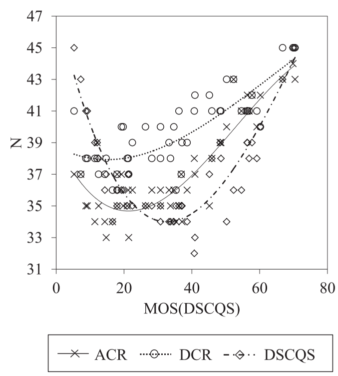


Fig. 8 Relationship between MOS of DSCQS method and N.

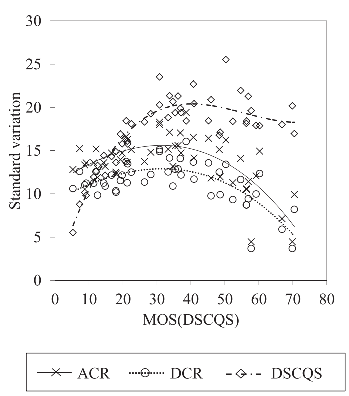


Fig. 9 Relationship between MOS of DSCQS method and standard variation.

5.4 Discussion

We can summarize the results as follow. As described in Sect. 5.1, the PCCs and SRCCs between different pairings of the three methods were high. Therefore, if the quality characteristics need to be investigated, the ACR method is most efficient because it has the shortest total assessment time. As discussed in Sect. 5.2, the CI of the DCR method was smaller than those of the others. However, the CI of the DSCQS for high-quality video was small. In other words, when the highest priority is stability, the DCR method is suitable, regardless of video quality, and the DSCQS method is also suitable for high-quality video. As explained in Sect. 5.3, the N of the DCR method was larger than those of the others for low-quality video, and the N of the DSCQS method was larger than those of the others for high-quality video. Therefore, when the highest priority is discrimination ability, the DCR method is suitable for low-quality video and the DSCQS method is suitable for high-quality video.

6. Conclusion

We conducted subjective quality assessments of 3D and 2D videos using the ACR, DCR, and DSCQS methods in order to evaluate the PCCs, SRCCs, stability, and discrimination ability of these methods.

We first investigated the PCCs and SRCCs between different pairings of the three methods to clarify which method is most efficient in terms of the assessment time. The results indicated that the PCCs and SRCCs among the MOSs of these methods are high. Therefore, if the objective of an experiment is to clarify quality characteristics, the ACR method is most efficient because the total assessment time of the ACR method is the shortest.

We clarified that the stability of these subjective assessment methods for 3D video quality is the same as that for 2D quality because the CIs of the MOSs of these methods for 3D video is similar to that for 2D video regardless of video quality. In addition, we found that the CIs of the MOSs of the DCR and DSCQS methods are smallest for high-quality video and that the CI of the DCR method is smaller than that of the others for low-quality video. That is, the stability of the DCR and DSCQS methods are higher than that of the ACR method for high-quality video, and the stability of the DCR method is higher than that of the others for low-quality video. Furthermore, the DCR method is most suitable for all quality ranges in terms of stability and assessment time.

Finally, to compare the discrimination ability of the subjective assessment methods, we conducted a Student's t-test between the MOSs for each method. The results showed that N in the DSCQS method is larger than those in the others for high-quality video and that N in the DCR method is higher than those in the others for low-quality video. Thus, the DSCQS method had the highest discrimination ability for high-quality video, and the DCR method had the highest for low-quality video.

The following issues call for further study. In this paper, we compared the subjective assessment methods for 3D video with compression artifacts only. The study will be extended to include other artifacts such as packet loss and re-buffering. In addition, the performance of these methods not only for quality but also depth perception and visual comfort needs to be evaluated.

References

[1] ITU-R Recommendation BT.2021, "Subjective methods for the assessment of stereoscopic 3DTV systems," Aug. 2012.
[2] ITU-R Recommendation BT.1438, "Subjective assessment of stereoscopic television pictures," March 2000.
[3] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," Jan. 2012.
[4] ITU-R Document 6Q/131-E, "Technical report: Comparison of DSCQS and ACR," Oct. 2005.
[5] Q. Huynh-Thu and M. Ghanbari, "A comparison of subjective video quality assessment methods for low-bit rate and low-resolution video," IASTED Signal Image Process, pp.70–76, Aug. 2005.
[6] C. Lee, H. Choi, E. Lee, S. Lee, and J. Choe, "Comparison of various subjective video quality assessment methods," SPIE Image Quality and System Performance III, vol.6059, pp.605906-1–7, June 2006.
[7] P. Corriveau, C. Gojmerac, B. Hughes, and L. Stelmach, "All subjective scales are not created equal: The effects of context on different scales," Signal Process., vol.77, no.1, pp.1–9, Aug. 1999.
[8] T. Tominaga, T. Hayashi, J. Okamoto, and A. Takahashi, "Performance comparisons of subjective quality assessment methods for

mobile video," IEEE QoMEX 2010, pp.82–87, June 2010.
[9] T. Kawano, K. Yamagishi, and T. Hayashi, "Performance comparisons of subjective assessment methods for 3D video quality," IEEE QoMEX 2012, pp.218–223, July 2012.
[10] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," April 2008.
[11] K. Yamagishi, T. Kawano, T. Hayashi, and J. Katto, "Subjective quality metric for 3D video services," IEICE Trans. Commun., vol.E96-B, no.2, pp.410–418, Feb. 2013.
[12] L. Xinga, T. Ebrahimia, and A. Perkisa, "Subjective evaluation of stereoscopic crosstalk perception," SPIE Proc. Visual Communications and Image Processing, vol.7744, pp.774410-1–9, July 2010.
[13] W. Kim, S. Kim, and J. Kim, "Analysis on the spectrum of a stereoscopic 3-D image and disparity-adaptive anti-aliasing filter," IEEE Trans. Circuits Syst. Video Technol., vol.19, no.10, pp.1561–1565, Oct. 2009.
[14] S.L.P. Yasakethu, C.T.E.R. Hewage, W.A.C. Fernando, and A.M. Kondoz, "Quality analysis for 3D video using 2D video quality models," IEEE Trans. Broadcast., vol.54, no.4, pp.1969–1976, Nov. 2008.
[15] M. Chen, D. Kwon, L.K. Cormack, and A.C. Bovik, "Optimizing 3D image display using the stereoacuity function," IEEE ICIP 2012, pp.617–620, Sept. 2012.
[16] N. Ozbek, G. Ertan, and O. Karakus, "Interactive quality assessment for asymmetric coding of 3D video," 3DTV Conference: The True Vision — Capture, Transmission and Display of 3D Video, pp.1–4, May 2011.
[17] ITU-R Recommendation BT.1788, "Methodology for the subjective assessment of video quality in multimedia applications," Jan. 2007.
[18] http://3d-contents.nict.go.jp/ (in Japanese)
[19] http://www.dcaj.org/project/3d/evaluation.html (in Japanese)

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.

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 the 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.

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 in NTT Laboratories. He received the Telecommunication Advancement Foundation Award in Japan in 2008 and the Telecommunication Technology Committee Award in Japan in 2012.