https://doi.org/10.2352/ISSN.2470-1173.2021.11.HVEI-161

Extension of ITU-T P.1203 model to Tile-based Omnidirectional Video Streaming

Yuichiro Urata, Masanori Koike, Kazuhisa Yamagishi, Noritsugu Egi and Jun Okamoto; Nippon Telegraph and Telephone Corporation (NTT); Tokyo, Japan

Abstract

Omnidirectional video (ODV) streaming has become widespread. Since the data size of ODV is extremely large, tile-based streaming has been developed to compress the data. In this coding technology, high-quality tiles encoded at a higher bitrate for the users' viewing direction and low-quality tiles encoded at a lower bitrate for the whole environment are sent to the client, and a player decodes these tiles. As a result, quality degrades due to coding, streaming, and the client's buffer. Therefore, to provide high-quality tile-based ODV streaming services, quality-of-experience needs to be monitored by comprehensively evaluating the quality degradations. By taking into account the quality degradation due to the low-quality tiles, the ITU-T Recommendation P.1203 model, which can be used for monitoring the quality of 2D video streaming services, is extended to tile-based ODV streaming services. Our model is demonstrated to estimate quality with sufficiently high accuracy.

Introduction

Due to advances in camera, display, and video-processing technologies, omnidirectional video (ODV) streaming services have recently become widespread. Although ODV provides a highly immersive viewing experience, it has a larger amount of data than 2D video. Therefore, during the streaming, image-quality degradation due to coding, and quality adaptation and stalling often occur due to the throughput being reduced and the buffer being depleted. To monitor the normality of a service, the quality at end-clients must be monitored. To do that, a quality-estimation model needs to be developed.

To reduce the bitrate in ODV, tile-based streaming has been proposed [1, 2] and standardized [3], as shown in Fig. 1. In tile-based streaming, high- and low-quality tiles are streamed to a head-mounted display (HMD). The entire image is divided into multiple tiles for creating high-quality tiles, which are basically displayed on the HMD, and the entire image is downsized to low-quality tiles with a smaller resolution than the original and is displayed on the HMD when users change their viewing direction (i.e., viewpoint). The quality of high- and low-quality tiles depends on the employed resolution, framerate, and bitrate, like 2D video streaming or non-tile-based ODV [4, 5]. In addition, the display time of low-quality tiles (hereafter, the delay) also affects quality. Therefore, users perceive quality degradation due to encoding and upscaling [6, 7, 8, 9]. Like in 2D video streaming, MPEG-DASH [10, 11] is used in ODV streaming services. Video data with a suitable bitrate for the current throughput and buffer size is requested and downloaded. Therefore, the quality is adaptively changed due to the throughput and buffer fluctuation. Since the usage of the terminal buffer fluctuates, stalling sometimes occurs due to the throughput being reduced and the buffer being depleted [12, 13]. To develop a quality-estimation model, these quality-influencing factors need to be taken into account.

To monitor the quality of tile-based ODV streaming services at end-clients, this paper proposes extending ITU-T Recommendation P.1203 mode 0 model (hereafter, P.1203 model), which is used to estimate the quality of 2D video streaming services, to tile-based ODV streaming services by taking into account quality degradations due to the display of low-quality tiles. Concretely, three types of model are investigated: model A) information about high- and low-quality tiles and the delay is taken as input, model B) information about high and low-quality tiles is taken as input, and model C) information about only high-quality tiles is taken as input. Two subjective quality-assessment experiments are conducted to compare the quality-estimation accuracy of the three models.

Related work

In this section, issues of conventional quality-estimation models for tile-based ODV streaming are described, and ITU-T Recommendation P.1203, which is the base of the proposed model, is explained.

Conventional quality-estimation models

In 2D video, many quality-estimation models have been proposed [14, 15, 16, 17, 18] and ITU-T Recommendation P.1203 was standardized [19, 20, 21], where the P.1203 model is described in the next sub-section.

Alberti *et al.* [14] proposed a quality-estimation model that takes the bitrate, framerate, quantization parameter (QP), stalling frequency, stalling average duration, and quality change rate as input. The first and last three are the parameters used for estimating short- and long-term quality, respectively. Tran *et al.* [15] proposed a quality-estimation model that takes encoding parameters such as QP, frame rate, or resolution. Duamu *et al.* [16] proposed a quality-estimation model for estimating encoding quality by using full reference models and evaluating the effect of stalling and the temporal effect on quality by pooling strategies on the basis of an auto-regressive model. However, when quality is estimated at end-clients, the models using bitstream-layer information such as QP are not suitable because it is encrypted, and models using pixel information are not suitable because of their computational power.

Ghadiyaram *et al.* [17] focused on estimating continuous-time quality and evaluated the temporal effect of stalling by using the Hammerstein-Wiener models. Yamagishi and Hayashi [18] proposed a quality-estimation model that takes meta-data (such as the bitrate, resolution, and stalling information) as input and

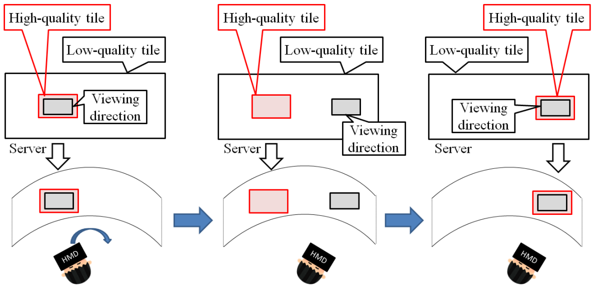


Figure 1. *Quality degradation due to changes in viewing area*

calculates the quality. With these 2D models, the effect of quality fluctuation and stalling can be evaluated, but tile-based VR image quality cannot. When these 2D quality-estimation models are applied to tile-based ODV, an issue remains: how to take into account the impact of the high- and low-quality tiles and the delay on quality.

A tile-based ODV quality-estimation model was proposed [8] that evaluates quality using latitude (viewing angle), tiling, stalls, and quality changes. However, it does not take bitrate account as input. Since the bitrate of each quality level varies among service providers, this model cannot be used for evaluating quality degradation affected by the change in bitrate. The display time of low-quality tiles (the delay) is also not taken into account.

From these investigations, a quality-estimation model needs to be developed that can be used for calculating the impact of the high- and low-quality tiles and the delay.

Recommendation P.1203 model

In 2D video streaming services, the P.1203 model has been standardized and is known to have sufficiently high quality-estimation accuracy. Therefore, the P.1203 model should be extended to a quality-estimation model for ODV streaming services. To do that, the P.1203 model is introduced in this section.

The P.1203 model consists of three modules. One is an audio-quality-estimation module that estimates short-term audio quality. Another is a video-quality-estimation module that estimates short-term video quality. The other is a quality-integration module that integrates time-series coding quality of video and audio, and this module takes the length of stalling time and position as input and estimates the quality for the media session.

The audio-quality-estimation module calculates audio quality per second, $O.21$, as follows.

$$\begin{aligned} O.21 &= MOSfromR(QA), && (1)\\ QA &= 100 - QcodA, && (2)\\ QcodA &= a1A \cdot \exp(a2A \cdot Bitrate) + a3A && (3) \end{aligned}$$

where $Bitrate$ is the audio bitrate in kbps and the coefficients $a1A$, $a2A$, and $a3A$ are constant. The variable $QcodA$ is the amount of quality degradation related to audio encoding.

The video-quality-estimation module calculates video quality per second, $O.22$, as follows.

$$\begin{aligned} O.22 &= MOSfromR(100 - \min(D, 100)), && (4)\\ D &= Dq + Du + Dt, && (5) \end{aligned}$$

$$\begin{aligned} Dq &= 100 - RfromMOS(MOSq), && (6)\\ MOSq &= q_1 + q_2 \cdot \exp(q_3 \cdot quant), && (7)\\ Du &= u_1 \cdot \log_{10}(u_2 \cdot (scaleFactor - 1) + 1), && (8)\\ scaleFactor &= \max\left(\frac{disRes}{codRes}, 1\right) && (9) \end{aligned}$$

where Dq is the amount of quality degradation related to the quantization calculated from $quant$, which is a variable related to the quantization parameter and calculated from the bit amount per pixel and bitrate. The coefficients q_{1-3} are constant. The variable Du is the amount of quality degradation related to the resolution of the encoding, $scaleFactor$ is a parameter capturing upscaling degradation, $disRes$ is display resolution, and $codRes$ is coding resolution. The coefficients u_1 and u_2 are constant. The variable Dt is the amount of degradation related to the frame rate. $MOSfromR$ converts the mean opinion score (MOS) from the psychological value R of $0-100$, and $RfromMOS$ converts R from MOS. The details of these two functions can be found in Annex E of ITU-T Recommendation P.1203.

In the quality-integration module, audio-visual (AV) quality $O.34$ is calculated first by using $O.21$ and $O.22$ as follows.

$$\begin{aligned} O.34_t = \; & av_1 + av_2 \cdot O.21_t + av_3 \cdot O.22_t \\ & + av_4 \cdot O.21_t \cdot O.22_t && (10) \end{aligned}$$

The subscript t is time, and av_{1-4} are coefficients.

Next, by using the time series data of $O.34$, the integrated quality $O.35$ is calculated as

$$\begin{aligned} O.35 = \; & O.35_{baseline} - negativeBias \\ & - oscComp - adaptComp, && (11) \end{aligned}$$

$$\begin{aligned} O.35_{baseline} &= \frac{\Sigma_t w_1(t) \cdot w_2(t) \cdot O.34_t}{\Sigma_t w_1(t) \cdot w_2(t)}, && (12)\\ w_1(t) &= t_1 - t_2 \cdot \exp\left(\left(\frac{t-1}{T}\right) / t_3\right), && (13)\\ w_2(t) &= t_4 - t_5 \cdot O.34_t, && (14) \end{aligned}$$

where the variables $negativeBias$, $oscComp$, and $adaptComp$ are the effect of the range and frequency of the quality change due to the throughput fluctuation on quality of experience (QoE), T is the duration of the media session, and t_{1-5} are coefficients.

The model has a machine-learning part that is a random forest and based on 14 features. The features are related to stalling duration, stalling frequency, $O.21$, and $O.22$ and are listed as follows. 1) Total number of stalling events occurring in the media session, excluding the initial stalling event. 2) The sum of the durations of all stalling events. 3) Frequency of stalling events: the number of stalling events (excluding the initial stalling) divided by the length of the media. 4) Ratio of stalling duration: The ratio of stallDur to the total media length. 5) The time elapsed from the start of the last stalling event to the end of a video. The initial stalling event is excluded from the calculation of this feature. The value of this feature is set to T if there is no stalling in the session. 6-8) The average of all the O.22 scores that correspond to the first, second, and last thirds of the O.22 score vector. The average of all the O.22 scores of the second third of the O.22 score vector. 9-11) The first, fifth, and tenth percentiles of O.22. 12)

All the O.21 scores corresponding to the first half of the session are averaged. 13) All the O.21 scores corresponding to the second half of the session are averaged.14) The length of the media. From the random forest outputs, quality values are estimated as $RFPrediction$.

Finally, the media session quality $O.46$ is calculated using $O.35$, stalling information, and $RFPrediction$.

$$O.46 = 0.02833052 + 0.98117059 \cdot O.46_{temp}, \tag{15}$$

$$\begin{aligned} O.46_{temp} &= 0.75 \cdot (1 + (O.35 - 1) \cdot SI) \\ &\quad + 0.25 \cdot RFPrediction, \end{aligned} \tag{16}$$

$$\begin{aligned} SI &= \exp\left(-\frac{numStalls}{s_1}\right) \cdot \exp\left(-\frac{tatalStallLen}{T \cdot s_2}\right) \\ &\quad \cdot \exp\left(-\frac{avgStallInterval}{T \cdot s_3}\right), \end{aligned} \tag{17}$$

where $numStalls$ is the number of stalling events, $totalStallLen$ is the total stalling time, $avgStallInterval$ is the average stalling interval, T is media session time, and s_{1-3} are coefficients.

Extensions of P.1203 model

This section explains the proposed video-quality-estimation modules, which take high- and low-quality tiles and delay related information. Since the P.1203 model calculates 2D video quality $O.22$, it can conceivably be used to estimate the quality of high- and low-quality tiles. When users change their viewing direction, they perceive quality degradations due to the quality of low-quality tiles and the delay. Therefore, the delay needs to be taken into account in the video-quality-estimation module.

To investigate the improvement of quality-estimation accuracy, three types of $O.22$ calculation models are developed. The simplest model (model C) takes the quality of the high-quality tiles as input. To take into account the quality degradation due to the low-quality tiles, the second model (model B) uses the weighted sum of video quality of high- and low-quality tiles. To take into account the quality degradation due to the delay, the third model (model A) takes the delay in addition to the input of model B.

A) High- and low-quality terms with delay (model A)

$$O.22 = \omega \cdot O.22_H + (1 - \omega) \cdot O.22_L, \tag{18}$$

$$\omega = d_1 \cdot delay^{-d_2} \tag{19}$$

B) High- and low-quality terms without delay (model B)

$$O.22 = \omega_1 \cdot O.22_H + (1 - \omega_1) \cdot O.22_L \tag{20}$$

C) Only high-quality term (model C)

$$O.22 = O.22_H \tag{21}$$

$O.22_H$ and $O.22_L$ are based on the video-quality estimation module of the P.1203 model and are calculated using the bitrate, resolution, and framerate of high- and low-quality tiles, respectively, where q_{1-3} in (7), ω_1, and d_{1-2} were derived using the experimental data described in next section.

The coefficients av_{1-4} in (10) were also re-optimized by using the experimental data because the video quality ($O.22$) and audio quality ($O.21$) may affect the AV quality ($O.34$) of ODV differently from that of 2D video streaming.

Since the impact of stalling on ODV quality might differ from that on 2D video quality, the coefficients for stalling (s_{1-3}) in (17) were also derived using the experimental data described in next section.

The calculation of $O.46_{temp}$ to $O.46$ in (15) is for adjusting the heterogeneity of the results of experiments conducted by multiple organizations. For this reason, (15) was not used in this study.

Subjective experiments

Two subjective quality assessment experiments (Experiments 1 and 2) were conducted to develop proposed models and to validate their quality-estimation accuracy.

Source reference sequences

Since the coding efficiency depends on source reference sequences (SRCs), 11 SRCs were used, as shown in Figure 2. In Experiment 1, SRC 11-14 and SRC C1 and C2 were used. In Experiment 2, SRC 21-25 and SRC C1 and C2 were used. SRC C1 and C2 are used as common sequences to compare results of the experiments. SRC C1 is a video shot from the side of a crane car with a small amount of movement. SRC C2 is shot in a botanical garden at a fixed position and has many edges and a medium amount of movement due to wind. SRC 11 is shot from inside a car driving on a mountain road with a large amount of movement. SRC 12 shows a women dancing in a dance studio and has few edges and a medium amount of movement. SRC 13 shows a man playing music in a room and has few edges and a small amount of movement. SRC 14 is shot from the side of a waterfall and has medium amounts of edges and movement. SRC 21 is shot from a sunroof or inside a car driving on a mountain road with a large amount of movement. SRC 22 shows a women dancing in a park and has medium amounts of edges and movement. SRC 23 is shot in a botanical garden at a fixed position and has many edges and a small amount of movement. SRC 24 shows men practicing soccer and has a small amount of movement. SRC 25 is shot in a church at a fixed position and has few edges and almost no movement. The resolution of the SRCs was 7680×3840, and the framerate was 30 fps. The SRCs lasted 60 seconds.

Experimental conditions and processed video sequences

To develop a quality-estimation model that can be used for calculating the impact of the high- and low-quality tiles and the delay, the bitrates of high- and low-quality tiles and the delay need to be varied.

Tile-based coding was used, as shown in Fig. 3. In these experiments, H.265/HEVC (Main Profile/Level 5.0, GoP: M=3, N=15, 1-pass encoding) was used for high- and low-quality tiles. The chunk size was 0.5 seconds. The SRCs were divided into 1920×1920 regions and encoded as high-quality tiles. The number of high-quality tiles was 60 (12×5). That is, tiles adjacent in the horizontal direction overlapped by 1280 ($= 1920 - 7680/12$) pixels, and tiles adjacent in the vertical direction overlapped by 1152 ($= 1920 - 3840/5$) pixels. The low-quality tiles were downsized from the source videos from 7680×3840 to 1920×1920, and those tiles were displayed at the original resolution on the HMD.

To vary the quality due to encoding, delay, adaptivity, and

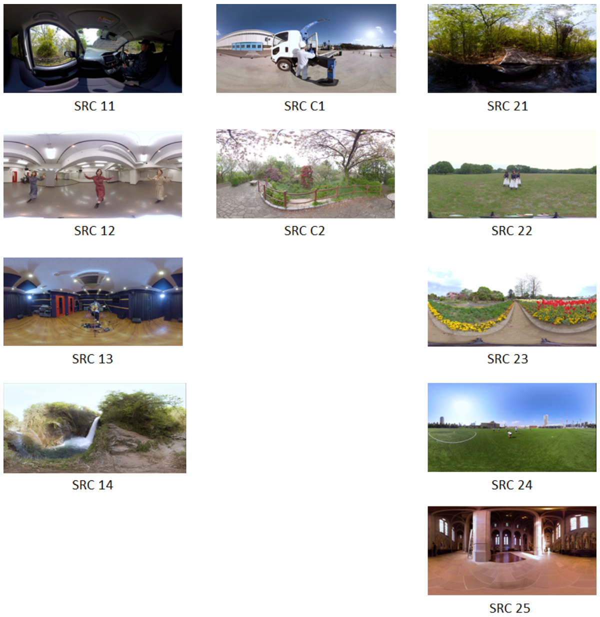


Figure 2. *Source contents*

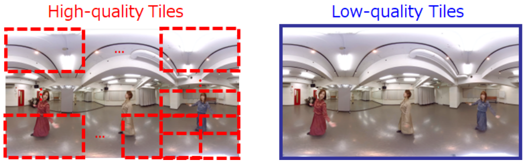


Figure 3. *Tile encoding*

stalling events, the parameters are used as follows. Eight quality levels (QL10-17) were used in Experiment 1, and ten (QL20-29) were used in Experiment 2. The bitrate pairs (high-quality tiles, low-quality tiles) for each quality level are shown in Table 1. To align the quality ranges of the experiments, the best quality levels (QL10 and QL20) and the worst (QL17 and QL29) were used as common conditions. The bitrate of high-quality tiles is the value for 1 tile out of 60, and low-quality tiles cover the whole environment. The delay was 1 to 8 seconds and was controlled by changing the player's buffer length. The experimental conditions were set so that the cases with and without bitrate fluctuation and stalling could be confirmed. The number of bitrate changes was 0 to 2, the number of stalls was 0 to 2, and stalling duration was 4 to 12 seconds per stalling event.

The number of processed video sequences (PVSs) was 48 in Experiments 1 and 2. In Experiment 1, SRC 11-14 and SRC C1 and C2 were used eight times. In Experiment 2, SRC 21-25 were used eight times, and SRC C1 and C2 were used four times.

The numbers of PVSs for combinations of stalling and quality changes are listed in Table 2. To show the same stalling events for all participants, the events were simulated by inserting frames that were stopped in the videos.

Table 1. Bitrate pairs (Mbps) for each quality level (QL)

QL	10	11	12	13	14	15	16	17		
High	16	16	8	8	4	4	2	2		
Low	16	8	8	4	4	2	2	1		
QL	20	21	22	23	24	25	26	27	28	29
High	16	9	9	7	7	5	5	3	3	2
Low	16	7	3	5	1	5	3	2	1	1

Table 2. The breakdown of the PVSs

Stalling	Quality change	Experiment 1	Experiment 2
-	-	16	16
x	-	8	8
-	x	12	12
x	x	12	12

Experimental environment

The participants used HTC Vive Pro, which is a virtual-reality headset with two glasses-like screens (1440 × 1600 pixels each), to watch ODVs. After the video is mapped onto a sphere, it is cropped and displayed in accordance with the viewing angle. The device displayed ODVs for both eyes with pseudo-parallax. The digital file level of -26dBov was set to an acoustic listening level of 18dBPa. The participants could freely change the viewing directions during the test.

Assessment method

Before the subjective test, participants took visual acuity and color vision tests, read the instructions (i.e., rating scale and voting procedure), and participated in a subjective test training that involved watching four videos. The video-quality-evaluation procedure followed the absolute category rating (ACR) methodology using a five-point scale. The participants wore a HMD, watched the videos, and evaluated the quality in a booth.

Participants take a short break (about 2 minutes) after watching the 4 videos (1 set) and took a longer break (about 10 minutes) after every three sets. The experiments lasted about 3 hours including instruction, visual acuity and color vision tests, training, and breaks. The presentation order of PVSs was randomized.

Participants

In both experiments, 32 participants took part: 16 males and 16 females with visual acuity of 1.0 or more with contact lenses or the naked eye. All the participants passed the visual acuity and color vision tests. They were naive participants who had not participated in subjective quality assessment experiments of ODV streaming in the previous six months. In Experiment 1, the participants were 18 to 35 years old (average age: 21.7). In Experiment 2, the participants were 18 to 26 years old (average age: 21.2).

Quality-estimation accuracy

Before investigating the quality-estimation accuracy, the stability of the subjective test was investigated on the basis of a 95% confidence interval (CI). Table 3 shows the mean, standard deviation, minimum values, and maximum values of the CIs. These mean CIs were almost the same as the mean CI (0.312) in Robitza et al. [20]. Since the CIs were not high, the stability can be said to

Table 3. Summary statistics of the confidence intervals

	Mean	Standard deviation	Minimum	Maximum
Experiment 1	0.318	0.043	0.193	0.402
Experiment 2	0.317	0.049	0.232	0.474

be high enough. Comparing the MOSs of the common sequences in the two experiments, the MOSs tended to be slightly higher in Experiment 2 than in Experiment 1. The linear mapping is as follows:

$$y = 0.9297x + 0.4184, \tag{22}$$

where x is MOS of Experiment 1 and y is MOS of Experiment 2. The Pearson correlation coefficient (PCC) was 0.957.

To investigate the quality-estimation accuracy of the three models (Models A, B, and C), the coefficients (q_{1-3}, av_{1-4}, ω_1, d_{1-2}, s_{1-3}) were optimized by using the Experiment 1 results and Microsoft Excel Solver. Table 4 shows the root mean squared errors (RMSEs) and PCCs. By comparing models A, B, and C in Table 4, quality-estimation accuracy is shown to be improved by adding low-quality terms and the delay.

For proposed model A, cross-validation is conducted to assess the quality-estimation accuracy for test data by using the results of Experiments 1 and 2. To adjust the heterogeneity of the results of the experiments, the MOSs were transformed by using (22). Table 5 shows the PCCs and the RMSEs of model A for test data. In either case, the quality-estimation accuracy of model A was maintained and was better than that of model B in Table 4. Figures 4 and 5 show the scattered plots between estimated and subjective MOSs when using the results of Experiments 1 and 2 as test data, respectively. These results reveal proposed model A achieves a sufficient quality-estimation accuracy.

Some investigations should be noted to explain the results of proposed model A in detail. Table 6 shows the RMSEs for PVSs of each SRC. SRC 21 had higher RMSE than the others because SRC 21 is a video with a large amount of movement and many edges. In fact, the QP values of SRC 21 were higher than others for each bitrate. This estimation error is inevitable because the proposed model cannot take content features (e.g., QP or pixels) as input. To improve the quality-estimation accuracy, bitstream information such as QP or media signals could be required as input.

Next, the effects of stalling and quality changes on quality-estimation accuracy are investigated. Table 7 shows the RMSEs for four combinations of stalling and quality changes. Model A exhibited high quality-estimation accuracy even when there was stalling but low accuracy when there was only quality changes without stalling. As described above, since the impact of SRC on quality could not be calculated in the proposed model, the quality-estimation accuracy degraded when quality changed. Under the conditions with stalling, SRC had little effect and could be estimated with relatively high accuracy.

Conclusion

In this paper, to monitor the normality of tile-based omnidirectional video (ODV) services, an extension of the ITU-T Recommendation P.1203 mode 0 model to tile-based ODV streaming services is proposed. To evaluate its quality-estimation accuracy, subjective quality assessment experiments were conducted.

Table 4. PCCs and RMSEs of each model for Experiment 1

Model	PCC	RMSE
A) High- and low-quality terms with delay	0.85	0.40
B) High- and low-quality terms without delay	0.75	0.49
C) Only high-quality term	0.71	0.52

Table 5. PCCs and RMSEs of model A for test data

Training data	Test data	PCC	RMSE
Experiment 1	Experiment 2	0.77	0.47
Experiment 2	Experiment 1	0.76	0.45

Results show the quality-estimation accuracy can be improved by taking into account the delay and quality of high- and low-quality tiles. Cross-validation was conducted to assess the quality-estimation accuracy of the proposed model (model A) for test data. The quality-estimation accuracy was maintained and was better than the quality-estimation accuracy of simpler models with training data. The quality-estimation accuracy was high for several videos but low for the video with a large amount of movement and many edges. The proposed model's quality-estimation accuracy was high for stalling conditions but not for quality-change conditions without stalling.

In the future, subjective quality assessment tests with a large variety of video sources will need to be conducted to further optimize the coefficients and evaluate the proposed model because only 11 video sources were used in these experiments. In addition, if the impact of source on the quality needs to be calculated, a bitstream-based or pixel-based model will need to be investigated.

References

[1] D. Ochi, A. Kameda, Y. Kunita, A. Kojima, and S. Iwaki, "Live streaming system for omnidirectional video," in *Proc. of IEEE Virtual Reality (VR)*, Mar. 2015.

[2] V. R. Gaddam, M. Riegler, R. Eg, C. Griwodz, and P. Halvorsen, "Tiling in Interactive Panoramic Video: Approaches and Evaluation," in *IEEE Trans. on Multimedia*, vol. 18, no. 9, pp. 1819-1831, Sept. 2016.

[3] MPEG, "Omnidirectional Media Application Format," in https://mpeg.chiariglione.org/standards/mpeg-i/omnidirectional-media-format

[4] A. Singla, S. Fremerey, W. Robitza, P. Lebreton, and A. Raake, "Comparison of Subjective Quality Evaluation for HEVC Encoded Omnidirectional Videos at Different Bit-rates for UHD and FHD Resolution," in *Proc. of the on Thematic Workshops of ACM Multimedia 2017 (Thematic Workshops '17)*, Oct. 2017.

[5] A. Singla, W. Robitza, and A. Raake, "Comparison of subjective quality evaluation methods for omnidirectional videos with DSIS and Modified ACR," in *Proc. of Electronic Imaging, Human Vision and Electronic Imaging 2018*, Jan. 2018.

[6] A. Ghosh, V. Aggarwal, and F. Qian, "A rate adaptation algorithm for tile-based 360-degree video streaming," in *Proc. of the 7th International Conference on Multimedia Systems (ACM MMSys)*, May 2016.

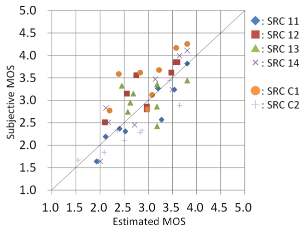


Figure 4. *Estimation results of model A trained by using Experiment 2 for Experiment 1*

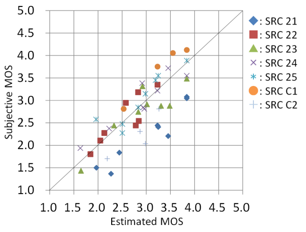


Figure 5. *Estimation results of model A trained by using Experiment 1 for Experiment 2*

Table 6. RMSEs for PVSs of each SRC

Training	Experiment 2	Training	Experiment 1
Test	Experiment 1	Test	Experiment 2
SRC	RMSE	SRC	RMSE
11	0.30	21	0.83
12	0.41	22	0.25
13	0.49	23	0.34
14	0.39	24	0.25
-	-	25	0.28
C1	0.63	C1	0.41
C2	0.42	C2	0.64

[7] A. Singla, S. Göring, A. Raake, B. Meixner, R. Koenen, and T. Buchholz, “Subjective quality evaluation of tile-based streaming for omnidirectional videos,” in *Proc. of the 10th ACM Multimedia Systems Conference (ACM MMSys)*, June 2019.

[8] J. Li, R. Feng, Z. Liu, W. Sun, and Q. Li, “Modeling QoE of virtual reality video transmission over wireless networks,” in *Proc. of 2018 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018.

[9] R. Schatz, A. Zabrovskiy, and C. Timmerer, “Tile-based Streaming of 8K Omnidirectional Video: Subjective and Objective QoE Evaluation,” *Proc. of 2019 11th International Conference on Quality of Multimedia Experience (QoMEX)*, June 2019.

[10] I. Sodagar, “The MPEG-DASH Standard for Multimedia Streaming Over the Internet,” *IEEE MultiMedia*, vol. 18, no. 4, pp. 62–67, Apr. 2011.

[11] J. L. Feuvre and C. Concolato, “Tiled-based adaptive streaming using MPEG-DASH,” in *Proc. of the 7th International Conference on Multimedia Systems (ACM MMSys)*, May 2016.

[12] W. Zhang, W. Zou, and F. Yang, “The Impact of Stalling on the Perceptual Quality of HTTP-based Omnidirectional Video Streaming,” in *Proc. of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019.

[13] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo, “Towards subjective quality of experience assessment for omnidirectional video streaming” in *Proc. of 2017 9th International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017.

[14] C. Alberti, D. Renzi, C. Timmerer, C. Mueller, S. Lederer, S. Battista, and M. Mattavelli, “Automated QoE evaluation of Dynamic Adaptive Streaming over HTTP,” in *Proc. of 2013 5th International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2013.

[15] H. T. T. Tran, N. P. Ngoc, A. T. Pham, and T. C. Thang, “A multi-factor QoE model for adaptive streaming over mobile networks,” *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016.

Table 7. RMSEs for conditions

	Training	Experiment 1	Experiment 2
	Test	Experiment 2	Experiment 1
Stalling	Quality changes		
-	-	0.50	0.51
x	-	0.30	0.21
-	x	0.57	0.47
x	x	0.39	0.47

[16] Z. Duanmu, A. Rehman, K. Zeng, and Z. Wang, “Quality of experience prediction for streaming video,” in *Proc. of 2016 IEEE International Conference on Multimedia and Expo (ICME)*, July 2016.

[17] D. Ghadiyaram, J. Pan, and A. C. Bovik, “Learning a Continuous-Time Streaming Video QoE Model,” *IEEE Trans. Image Processing*, vol. 27, no. 5, pp. 2257–2271, May 2018.

[18] K. Yamagishi and T. Hayashi, “Parametric Quality-Estimation Model for Adaptive-Bitrate Streaming Services,” *IEEE Trans. Multimedia*, vol. 19, no. 7, pp. 1545–1557, 2017.

[19] W. Robitza, M. Garcia, and A. Raake, “A modular HTTP adaptive streaming QoE model—Candidate for ITU-T P. 1203 (“P. NATS”),” in *Proc. of 2017 9th International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017.

[20] W. Robitza, S. Göring, A. Raake, D. Lindegren, G. Heikkilä, J. Gustafsson, P. List, B. Feiten, U. Wüstenhagen, M. Garcia, K. Yamagishi, and S. Broom. “HTTP adaptive streaming QoE estimation with ITU-T rec. P. 1203: open databases and software,” in *Proc. of 2018 9th ACM Multimedia Systems Conference (MMSys ’18)*, June 2018.

[21] ITU-T Recommandation P.1203, “Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport,” *ITU-T*, 2017.

Author Biography

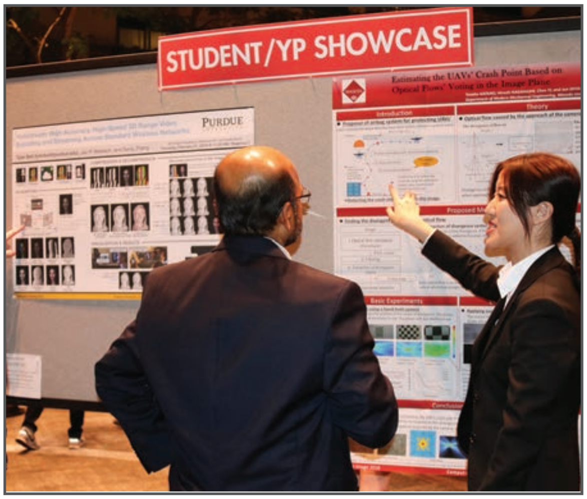
Yuichiro Urata received his B.E. and M.E. degrees in Engineering from University of Electro-Communications, Tokyo, Japan in 2009 and 2011. Since then he has worked in NTT Network Technology Laboratories in Tokyo. His work has focused on the Quality of Experience for videos.

Masanori Koike received his Bachelor (2015) in Engineering and Master (2017) in Information science and technology from the University of Tokyo, Japan. Since 2017, he has worked in NTT Network Technology Laboratories in Tokyo. His main research interest is the Quality of Experience for VR videos.

Kazuhisa Yamagishi received his B.E. degree in Electrical Engineering from the Tokyo University of Science, Japan, in 2001 and his M.E. and Ph.D. degrees in Electronics, information, and Communication Engineering from Waseda University, Japan, in 2003 and 2013. Since joining NTT Laboratories in 2003, he has been engaged in the development of objective quality-estimation models for multi-media telecommunications.

Noritsugu Egi received his B.E. and M.E. degrees in Electrical Communication Engineering from Tohoku University, Japan in 2003 and 2005. He joined NTT Laboratories, Tokyo, Japan, in 2005. Currently, he is researching speech and audio quality assessment.

Jun Okamoto received the B.E. and M.E. degrees in electrical engineering from the Tokyo University of Science in 1994 and 1996. In 1996, he joined NTT Laboratories, where he has been involved in the quality assessment of multimedia telecommunication and network performance measurement methods.


STUDENT/YP SHOWCASE
PURDUE

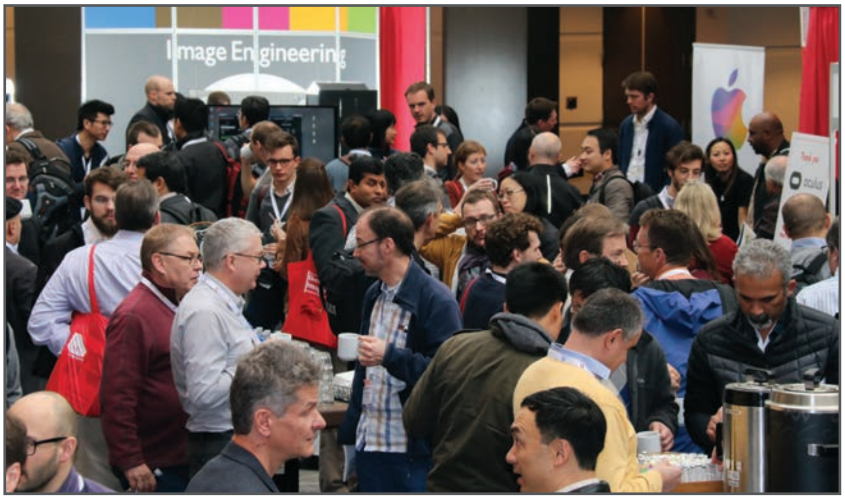


Image Engineering

AR is a fundamental and transformative interface and platform.
20+ years from my initial working demo to today's AR displays
Optimistic that it won't take another 20 to reach ubiquitous consumer AR!
My first working AR prototype
Today's platforms
Ubiquitous consumer AR
Optimis... this is << 20