PAPER Special Section on Image Media Quality

A Histogram-Based Quality Model for HTTP Adaptive Streaming

Huyen T. T. TRAN[†], **Nam PHAM NGOC**[††], **Yong Ju JUNG**[†††], ***Nonmembers***, **Anh T. PHAM**[†], ***and*** **Truong Cong THANG**[†a)], ***Members***

SUMMARY HTTP Adaptive Streaming (HAS) has become a popular solution for multimedia delivery nowadays. Because of throughput variations, video quality fluctuates during a streaming session. Therefore, a main challenge in HAS is how to evaluate the overall video quality of a session. In this paper, we explore the impacts of quality values and quality variations in HAS. We propose to use the histogram of segment quality values and the histogram of quality gradients in a session to model the overall video quality. Subjective test results show that the proposed model has very high prediction performance for different videos. Especially, the proposed model provides insights into the influence factors of the overall quality, thus leading to suggestions to improve the quality of streaming video.

key words: *video quality model, adaptive streaming, subjective test, histogram*

1. Introduction

HTTP Adaptive Streaming (HAS) has become a popular solution for multimedia delivery nowadays. In HAS, a video is encoded into multiple versions with different bitrates (and so different quality levels) [1]. Each version is further divided into short segments. Based on the estimated throughput, a client downloads a series of segments with suitable versions. Because of throughput variations, segment quality values fluctuate drastically during a session. Therefore, a main challenge in HAS is how to evaluate the overall quality of a session with strong quality variations.

There have been many studies on video quality models. Some studies find out and quantify the impacts of different factors on video quality, such as quantization parameter (QP) [2]–[4], resolution [5], frame rate [6], PSNR [7], and motion activity [8]. It should be noted that, in video encoding, QP is used to control the quantization [9] and is considered to be a key factor affecting the video quality. More specifically, the higher the QP is, the lower the video quality becomes.

For adaptive streaming sessions with quality variations, the overall quality is generally estimated based on the instant quality values (here after referred to as segment quality) [10]–[12]. A segment quality value can be a subjective score [11], [13] or an objective score which can be predicted (sometimes represented) using encoding parameters such as bitrate [14], [15], and QP [12], [16].

To predict the overall quality of a session, most of previous quality models use the average [11], [17], the median [12], the minimum [12], the standard deviation [11] of the (segment) quality values as well as the switching frequency [11], [14]. However, the impacts of the quality values and the switch amplitudes on the overall quality are still not fully understood yet.

In this paper, we propose a quality model to predict the overall quality of a session in HAS. The statistics of the different quality values and the statistics of the quality changes in a session, which are respectively represented by the histogram of quality values and the histogram of quality gradients, are considered to be the key features of the overall quality. Besides, the quality of a segment, which is assumed to be constant, is computed using the average QP of that segment. Through subjective test results and comparison with two reference models, we show that the overall quality can be predicted well by our proposed quality model. In addition, based on model parameters, we can quantify the impacts of different factors, namely segment quality, switch amplitude, and content on the overall quality.

This paper is organized as follows. In Sect. 2, we highlight the related work and our contributions. Section 3 presents our proposed model in detail. Subjective test results and evaluation of the quality model are presented in Sect. 4. Section 5 discusses influence factors on the overall quality. Finally, Sect. 6 concludes the paper.

2. Related Work

In general, the quality of experience (QoE) of a streaming session is affected by the perceptual quality, initial delay, and stalling (or interruptions). The perceptual quality is in turn determined by 1) the quality amplitude (i.e., high or low) and 2) the quality variations of the session. Recent studies have investigated, both qualitatively and quantitatively, different factors that impact the quality of a session in HAS [10]–[20].

Regarding the impact of the total time on a certain quality level on human perception, an observation in [13] is that the time on the maximum quality level has a significant impact on human perception. In [15], the impact of maintenance of low quality values is considered, and the authors show that this impact grows exponentially with maintenance

Manuscript received May 30, 2016.
Manuscript revised September 10, 2016.

[†]The authors are with the University of Aizu, Aizuwakamatsu-shi, 965-8580 Japan.
[††]The author is with Hanoi University of Science and Technology, 1 Dai Co Viet, Hanoi, Vietnam.
[†††]The author is with Gachon University, 1342 Seongnam-daero, Seongnam-si, Gyeonggi-do, 461-701 South Korea.
a) E-mail: thang@u-aizu.ac.jp

DOI: 10.1587/transfun.E100.A.555

time.

In addition, some qualitative observations have been presented in recent literature regarding the impacts of quality switching on human perception. In [18], the authors give some findings about the impacts of the amplitude and the frequency of quality switches. In particular, for user satisfaction, both the switching amplitude and the frequency of quality switches should be kept as low as possible. In [19], the authors also investigate the impact of quality switching on human perception, including the impacts of the direction (up/down), the amplitude (smooth/abrupt), and the number of switches. Specifically, for up-switches, the perceptual difference between smooth and abrupt switches is negligible. In contrast to up-switches, the higher the down-switching amplitude is, the more negative its impact becomes. This observation is the same as the conclusion in [13]. Regarding the number of switches, the authors in [19] show that the impact of switching frequency on perceptual quality is negligible, which agrees with a finding in [13].

Recency is also one of factors affecting human perception. However, its impact actually has not been clear yet. In recent literature, recency effect is formulated by different time-weighted functions [14], [15], but these functions have not been clearly proved by subjective tests with video contents. Based on the analysis of subjective test results, the authors in [19] observe that recency effect caused by quality values at the end of a session is significant. In contrast, the authors in [13] observe that recency effect can be neglected if more than two switches occur.

The authors in [10] considered initial buffering time, quality of segments and interruptions as metrics that directly impact a session's overall quality. Each segment quality value is computed as a linear function of QP. The overall quality is then predicted as the accumulation of instant quality values, which are weighted by human memory effects. The work in [16] investigated the impacts of QP and interruptions on the overall quality which are based on a methodology called Pseudo-Subjective Quality Assessment. However, the work in [10] and [16] did not consider the impact of quality variations on the overall quality.

The work in [14] proposed a model based on segment quality values in the temporal dimension and the number of switches. Each segment quality value is derived based on bitrate and motion parameter of video content. It is claimed that the quality values at the beginning and the end of a session have higher impacts on user impression. So, the weight of each segment is decided by the impression to user based on memory effects. In [17], the authors presented a QoE model, where the quality amplitude of a session is represented by the average of the segment quality values. Meanwhile, the factor of quality variations is represented by the frequency, types, and temporal locations of quality switching.

In [12], two quality models are proposed. In the first model, quality variations in a session are decomposed into frequency components. The overall quality is then predicted from the quality levels of the composing frequency components. It is observed that the frequency component with the worst quality among all frequency components has the biggest impact on the overall quality. The second model is based on the median and the minimum of segment quality values. Here, the minimum quality value can be considered as a measure of quality variations. Though being very simple, the second model is found to have better performances than the first one. The authors in [11], [20] proposed a model which considered four quality metrics, namely PSNR/SSIM, bitrate, version level, and segment quality in the mean opinion score (MOS). The finding is that the segment-quality based model provides the best performance. Specifically, the overall quality is predicted from the average, the standard deviation, and the switching frequency of segment quality values. In other words, the average of segment quality values is considered as the session's quality amplitude, and the standard deviation and the switching frequency of segment quality values are used to represent quality variations.

The work in [15] is one of the first studies that combines multiple influence factors into a QoE model, namely initial delay, stalling and quality variations. In this work, segment quality values are obtained as VQM metric and the quality variations are modeled by a heuristic function of low quality values and switch amplitudes.

As mentioned, most of the previous quality models use the average, the median, the maximum, the minimum, the standard deviation of segment quality values, as well as the number of switches, to predict the overall quality of a session. However, the impacts of segment quality values and switch amplitudes on overall quality are not fully understood yet. In this paper, this problem is quantitatively tackled by employing the segment quality histogram and the quality gradient histogram of a session. Our key contributions in this paper are as follows.

- First, we present a new quality model with very high prediction performance for different videos.
- Second, we quantify of the weights of different segment quality values and highlight the importance of high quality values in a session.
- Third, we show that switch-up events have a weight of zero, but switch-down events (especially large switches) have significant impacts on the overall quality.
- Fourth, the dependence of quality models on the content is investigated for the first time in this work.
- Finally, based on the findings, various suggestions to improve the quality of streaming service are provided.

It should be noted that the quality model proposed in this paper can be extended with factors of the initial delay and stalling in the same manner as [15]. Moreover, the recency effect can be incorporated in our model as in [10].

3. Proposed Quality Model

In HTTP Adaptive Streaming, we assume that each segment is represented by a quality value. Typically, QP, resolution, and frame rate are factors affecting the segment quality val-

ues. In this paper, we currently focus chiefly on the impact of QP on the segment quality. However, because our proposed model is essentially based on segment quality values (irrespective of QP, frame rate, or resolution), it could be applied to the cases of different resolutions and frame rates. In future work, we will consider applying our model to data sets with multiple quality dimensions (i.e. resolution, frame rate and QP). In this current work, each segment quality value is modeled as a function of the segment's average QP. Then the distribution of segment quality values and the distribution of switch amplitudes are used to predict the overall quality in a session. Note that QP values can be easily obtained from the headers of groups of pictures (GoP) and pictures/slices of a segment.

3.1 Histogram of Segment Quality Values

Previous studies have identified and quantified the impact of different factors on the perceptual quality, such as QP [2], [3], resolution [5], frame rate [6], PSNR [7]. In this paper, we assume that video versions have the same frame rate and the same resolution. We adopt the model in [2] to determine the segment quality based on the average QP. When the versions have different resolutions and frame rates, segment quality values can be predicted in a similar manner as shown in [2].

In H.264/AVC, every increase of 6 in QP is equal to a double of quantization step size (QS) [9]. Specifically, QS and QP are related by

$$QS = 2^{\frac{QP-4}{6}}. \tag{1}$$

The video quality Q_{QS} of a segment given quantization step size QS is defined by [2]:

$$Q_{QS} = Q_{min} \times \frac{1 - e^{-\sigma\left(\frac{QS_{min}}{QS}\right)}}{1 - e^{-\sigma}}, \tag{2}$$

where Q_{min} is the video quality given the lowest quantization step size QS_{min} and σ is a model parameter.

In our method, segment quality values of each session are split into N bins $\{B_{Q_n}\}$ where $n \in \{1, 2, \ldots, N\}$. Each bin B_{Q_n} corresponds to an interval I_{Q_n} of segment quality values, which is defined by

$$I_{Q_n} = [n - \gamma_{L_n}, n + \gamma_{U_n}), \tag{3}$$

where γ_{U_n} and γ_{L_n} are parameters to define the width of the interval I_{Q_n}.

The histogram of segment quality values, which represents the distribution of different bins, is defined based on the frequency of segment quality values within the corresponding intervals. In the current work, the quality is based on the mean opinion score (MOS), which varies in the range [1, 5]. So we split segment quality values into N=5 bins with $\gamma_{U_n} = \gamma_{L_n} = 0.5$ $(1 \le n \le 5)$.

An example of the segment quality histogram of a streaming session is shown in Fig. 1, where the frequency values are normalized to the range [0, 1]. We can see that

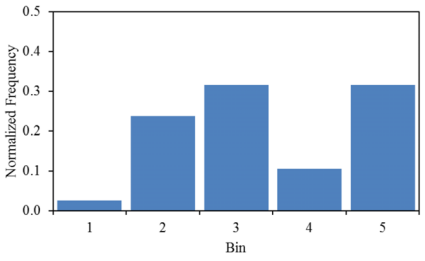

Fig. 1 An example of the segment quality histogram.

the segment quality values vary from 1 to 5 (MOS), and about 25% of segment quality values in this session belong to bin 1 and bin 2 of the histogram.

3.2 Histogram of Segment Quality Gradients

As mentioned above, quality variations are considered as the prominent feature (and also challenge) of HTTP Adaptive Streaming. In this study, we use the concept of "quality gradient" to represent quality variations. The instant gradient of segment quality values is given by

$$\nabla Q = \frac{\partial Q}{\partial t}, \tag{4}$$

where ∂Q is the difference between segment quality values. Currently, we use the quality changes between two consecutive segments to represent the instant gradients of a session. A positive (negative) gradient represents a switch-up (switch-down).

As the quality is mostly affected by switch-down events, we mainly focus on the negative gradients. Specifically, the negative gradients are split into M groups, corresponding to different switch-down amplitudes. So the histogram of instant gradient values is composed of M+2 bins $\{B_{\nabla Q_m}\}$ where $m \in \{-M, -M+1, \ldots, -1, 0, 1\}$. Each bin $B_{\nabla Q_m}$ corresponds to an interval $I_{\nabla Q_m}$ of instant gradient values, which is defined by

$$I_{\nabla Q_m} = [m - \delta_{L_m}, m + \delta_{U_m}), \tag{5}$$

where $\delta_{U_m}, \delta_{L_m}$ are parameters to define the width of the interval $I_{\nabla Q_m}$.

Currently, M is set to 4 and the histogram of instant gradients has 6 bins. The bin $B_{\nabla Q_1}$ corresponds to the interval of the positive instant gradient values, which is defined by $I_{\nabla Q_1} = [0.5, 4.5)$. The other bins correspond to the intervals with $\delta_{L_m} = \delta_{U_m} = 0.5$ $(-4 \le m \le 0)$.

It can be seen that the bin $B_{\nabla Q_1}$ represents the quality increases, the bin $B_{\nabla Q_0}$ represents the quality maintenance (or unchange), and the bins $\{B_{\nabla Q_m}\}$ with $(-4 \le m < 0)$ represents the quality decreases.

The instant gradient histogram, which belongs to the same session of Fig. 1, is shown in Fig. 2, where the frequency values are also normalized to [0, 1]. This histogram shows that there are a lot of switch-up events (for all switch amplitudes) in this session. Meanwhile, the number of

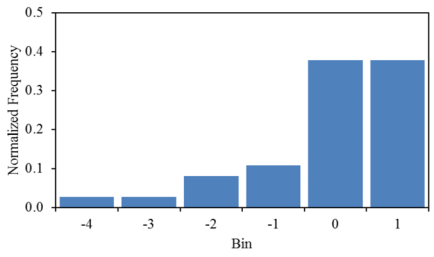

Fig. 2 An example of the instant gradient histogram.

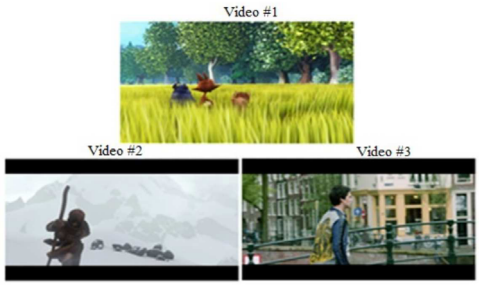

Fig. 3 Snapshots of three test videos.

switch-down events is very small, especially with switch amplitudes of −3 or −4 (MOS).

3.3 Overall Quality Model

The overall quality of a session is modeled by a pooling strategy of the above statistics of segment quality values and quality gradients. Let F_{Q_n} denote the frequency of segment quality values in bin $B_{Q_n}(1 \le n \le N)$, and $F_{\nabla Q_m}$ denote the frequency of quality gradients in bin $B_{\nabla Q_m}(-M \le m \le 1)$. The predicted overall quality of a session Q_{pred} is given by

$$Q_{pred} = \sum_{n=1}^{N} \alpha_n F_{Q_n} + \sum_{m=-M}^{1} \beta_m F_{\nabla Q_m}, \quad (6)$$

where α_n and β_m are the weights of the corresponding frequencies F_{Q_n} and $F_{\nabla Q_m}$ in the quality model.

As mentioned, the total time on each quality level and quality switching are factors affecting on human perceptual quality. In Eq. (6), the time on each quality level is represented by the component of segment quality values, and quality switching is represented by the component of quality gradients. In particular, the time on each quality level in a session is represented by the frequency of segment quality values in a corresponding bin. Regarding quality switching, each switching type determined by the switching direction and the switching amplitude is represented by quality gradient value, and the number of quality switches of each type in a session is represented by the frequency of the corresponding quality gradient values. Based on the values of α_n and β_m, the impacts of the time on quality levels and quality switching (including the direction, the amplitude, and the number of switches) on the human perception could be quantified. More discussions regarding the impacts of these factors will be provided in Sect. 5. In the following sections, content-specific and content-generic quality models will be obtained and validated by subjective data.

4. Model Evaluation and Analysis

4.1 Experiment Settings

To create versions for HTTP Adaptive Streaming, there are two kinds of video encoding modes, namely Constant Bitrate (CBR) (e.g. [11], [14]) and Variable Bitrate (VBR) (e.g. [12], [28]). With CBR mode, it is not easy to obtain segment quality values because the quality actually varies during a version or even a segment. Therefore, obtaining the segment quality values from the constant bitrate or version index of each version like in [11], [14] is not always correct. In contrast to CBR mode, although VBR mode has variable bitrate, it has consistent quality during each version. Therefore, the segment quality values can be accurately obtained from encoding parameters. In this study, a dataset in VBR mode is used to build and to evaluate our proposed model. Note that our model can be directly employed in VBR streaming [28]. In future work, we will investigate our model in the context of CBR streaming.

In this experiment, we use three videos of 74 seconds (1776 frames) from public short movies [21], namely Big Buck Bunny (denoted by BBB), Sintel (denoted by ST), and Tears of Steel (denoted by ToS) with starting timestamps of 00:05:00, 00:00:20, and 00:04:30, respectively. Fig. 3 shows the snapshots of the video contents. Features of the videos are presented in Table 1. The videos are encoded by using H.264/AVC (libx264) with a frame rate of 24 fps and a resolution of 1280x720. For each video, 9 versions are generated with corresponding QP of 20, 24, 28, 32, 36, 40, 44, 48, and 52. The duration of each segment is 2 seconds. A GoP structure of "IBBP" with a GoP length of 24 is used for all videos. Our data set has 49 streaming sessions. The first 44 sessions are generated using two adaptation methods of [22], [23] and 34 bandwidth traces (extracted from [24]) with different types of variations. The average bandwidth of each trace varies between 1Mbps and 4 Mbps. Figure 4 shows some bandwidth traces of the experiment. Additionally, we generate 5 sessions with fixed QPs, which are 24, 32, 36, 40 and 48. We can see that the first 44 sessions are of the variable-quality type and the last 5 sessions are of the constant-quality type.

Before doing actual subjective tests, the subjects are trained to get accustomed to the rating procedure and the range of video quality. During the tests, the test sequences of each video are randomly presented. The sequences are displayed on a 14-inch screen with a resolution of 1366×768 and a black background. There are totally 25 subjects taking part in this experiment.

The Absolute Category Rating (ACR) method is used in our experiments [25]. The viewers give a rating score at the end of each test sequence with the score ranging from 1

Table 1 Features of source videos.

Video	Content	Type	Motion activity	Spatial complexity
Video #1 (BBB)	Slow movements of characters	Animated video	Low	Complex (Forest scene)
Video #2 (Sintel)	A fight between 2 characters	Animated video	High	Simple (Snow mountain)
Video #3 (ToS)	Conversations of characters	Natural video	Low	Complex (Street scene)

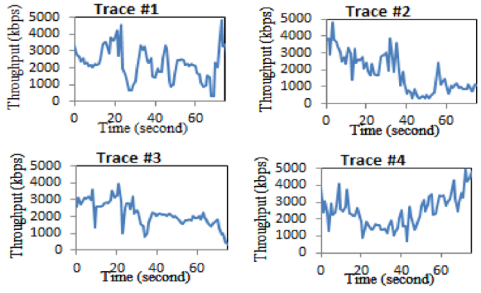

Fig. 4 Examples of bandwidth traces.

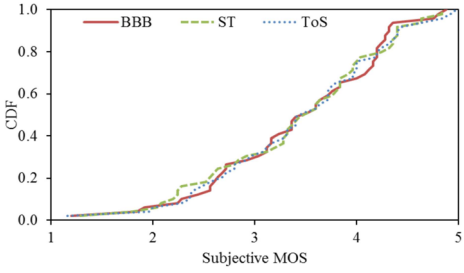

Fig. 5 CDF of MOS values corresponding to 49 streaming sessions.

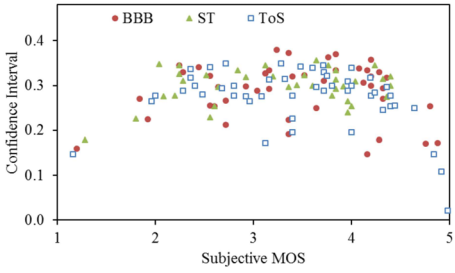

Fig. 6 Confidence intervals of the subjective MOSs of each video.

(worst) to 5 (best). Every 20 minutes, there is a break for the subjects. The mean opinion score (MOS) is determined as the average of viewers' scores.

Figure 5 shows the distribution of the MOS scores corresponding to 49 streaming sessions. We can see that the MOS values span across the range [1, 5]. The 95% confidence intervals of the 49 streaming sessions of all videos are shown in Fig. 6. The confidence intervals are generally smaller at the two ends of the score range. This is because the subjects are more confident in rating sessions of very high (or low) quality scores.

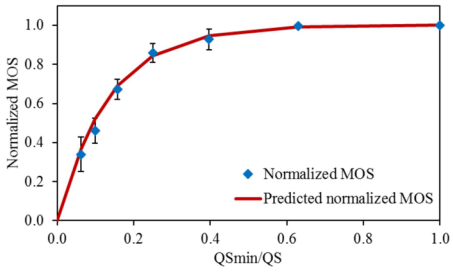

Fig. 7 Normalized MOS and predicted normalized MOS.

Table 2 Quality model.

	Video1 (BBB)	Video2 (ST)	Video3 (ToS)
σ	7.4	6.1	7.4
PCC	0.998	0.996	0.999
RMSE	0.14	0.13	0.05
ACI	0.22	0.26	0.18

4.2 QP-MOS Relationship

As mentioned, the quality of a segment can be represented by different measures, e.g. PSNR, subjective MOS, objective MOS, bitrate, etc. As our goal is to obtain a quality model for real-time session monitoring, we use the objective MOS predicted from QP (or QS).

To obtain segment quality values based on QP, we employ 7 test sequences of with the corresponding QP of 20, 24, 28, 32, 36, 40, and 44 for each of the three videos described in Sect. 4.1. Figure 7 shows the normalized MOS versus QS and the 95% confidence intervals for the BBB video. For closed-form relationship, the function of Eq. (2) is fitted to the corresponding data.

Table 2 summarizes the values of model parameter σ, the Pearson Correlation Coefficient (PCC) values, the Root Mean Squared Error (RMSE) values, and the average 95% confidence intervals (ACI) of the experiment. We can see that the model well matches the quality for each video. In practice, parameter σ can also be obtained by machine learning [25]. Based on the corresponding value of σ and content features, the training videos are clustered into groups. Then, classification is performed with any test video to derive the corresponding value of σ.

4.3 Prediction Performance

The segment quality values in each session are determined by the corresponding QP values of the segments. Figure 8 shows an example of segment quality variations given a bandwidth trace.

Similar to other studies (e.g. [4], [11], [16]), the generated sessions in this experiment are divided into two sets, namely a training set of 29 sessions and a test set of 20 sessions. Parameters $\{\alpha_n, \beta_m\}$ in Eq. (6) for content-specific

Table 3 Parameters of the proposed quality model.

Parameter	Video1 (BBB)	Video2 (ST)	Video3 (ToS)	All videos
α_1	1.2	1.3	1.2	1.2
α_2	1.6	2.0	1.8	1.8
α_3	2.6	3.4	2.6	2.8
α_4	4.2	3.9	4.1	4.1
α_5	4.6	4.8	4.6	4.7
β_1	0.0	0.0	0.0	0.0
β_0	0.0	0.0	0.0	0.0
β_{-1}	-1.7	-1.3	-1.6	-1.5
β_{-2}	-1.7	-5.5	-2.7	-3.2
β_{-3}	-15.4	-9.7	-13.9	-11.1
β_{-4}	-15.4	-13.1	-13.9	-11.1

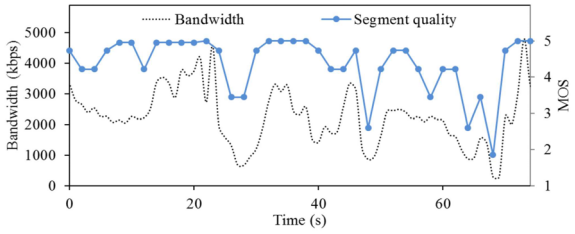

Fig. 8 An example of segment quality variations in a session.

Table 4 Performance of the proposed model.

Set	Metrics	Video #1	Video #2	Video #3	All videos
Training	PCC	0.95	0.97	0.95	0.95
	RMSE	0.26	0.20	0.26	0.27
Test	PCC	0.96	0.94	0.95	0.94
	RMSE	0.28	0.25	0.24	0.28

models (i.e. for each test video) and a content-generic model (i.e. for all test videos) are determined by curve-fitting to the MOS data of the corresponding training data. Then, the prediction performance of a model is evaluated by the corresponding test set.

The parameters of our content-specific and content-generic quality models are shown in Table 3. It can be seen that the lower the segment quality value is, the smaller the corresponding weight in the model is. In addition, the weight of positive gradients is zero. This means that it is unnecessary to quantify the weights of different positive gradient values (like negative gradients). However, when the quality decreases, the lower the quality gradient value is, the higher the impact on the overall quality becomes. More discussion about the impacts of different factors in our quality model will be provided in the next section.

The relationship between the predicted MOS and the subjective MOS in the training set and the test set for each video and all videos are shown in Fig. 9. The accuracy for the training set and the test set (i.e. PCC and RMSE) is shown in Table 4. For the test sets, our model achieves very high PCC values (0.94~0.96) and low RMSE values (0.24~0.28).

In this part, we also compare our proposed model with two reference methods of [12] and [11], using the same training sets and test sets. Note that segment quality values in these models are also MOS. The best quality model in

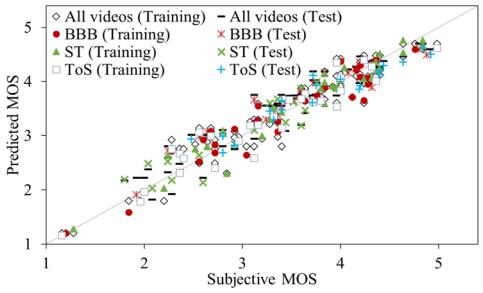

Fig. 9 Scatter diagram of subjective MOS and predicted MOS.

Table 5 Model parameters of the reference methods.

Quality models	Parameter	Video1 (BBB)	Video2 (ST)	Video3 (ToS)	All videos
$Q_{\text{ref}-1}$ [12]	α	0.6	0.7	0.5	0.6
	β	0.4	0.4	0.5	0.4
$Q_{\text{ref}-2}$ [11]	α	1.0	1.0	1.0	1.0
	β	0.69	0.6	1.0	0.7
	γ	0.0	0.0	0.0	0.0

[12] predicts the overall quality from the median and the minimum of the segment quality values as follows.

$$Q_{ref-1} = \alpha Q_{median} + \beta Q_{min}, \tag{7}$$

with α and β being the model parameters.

In [11], the overall quality is predicted from the average, the standard deviation, and the switching frequency of the segment quality values as follows.

$$Q_{ref-2} = \alpha Q_{aver} - \beta Q_{std} - \gamma Q_{SwFreq}, \tag{8}$$

with α, β and γ being the model parameters.

The parameters of content-specific and content-generic models using the reference methods are shown in Table 5. The PCC and RMSE of the two reference models and our proposed model, together with average 95% confidence intervals (ACI) of the experiment, are shown in Table 6.

It can be seen that, compared to the two reference models, our model achieves higher PCC values and lower RMSE values. The model of [11], which uses average quality and quality standard deviation, performs better than that of [12]. Interestingly, the weight of switching frequency in Eq. (8) is zero, implying that this factor has no contribution in the overall quality, which is inline with the qualitative finding in [13]. In addition, among the three considered models, RMSE of our proposed model not only is the lowest but also is lower than the corresponding ACI for all test videos.

The above results show that, using the proposed model, we can not only quantify the impacts of segment quality values and quality gradients on the overall quality, but also predict the overall quality accurately. In the next section, we will discuss different influence factors in the quality models.

Table 6 Comparison of the quality models.

Quality models	Metrics	Video1 (BBB)	Video2 (ST)	Video3 (ToS)	All videos
Q_{ref-1}[12]	PCC	0.91	0.84	0.92	0.88
	RMSE	0.41	0.44	0.32	0.43
Q_{ref-2}[11]	PCC	0.96	0.91	0.93	0.93
	RMSE	0.35	0.32	0.28	0.34
Q_{pred}	PCC	0.96	0.94	0.95	0.94
	RMSE	0.28	0.25	0.24	0.28
Average confidence interval (ACI)		0.29	0.29	0.28	0.29

5. Discussions

5.1 Impacts of the Number of Quality Bins

In this part, we investigate the accuracy of the proposed model with respect to the number of the segment quality bins N, where N is 3, 5, 7. The intervals of segment quality values for N of 3 and 7 are shown in Table 7. For each N, we randomly select 20 training sets, each set includes 29 streaming sessions among 49 sessions of each video. In each selection, the 20 remaining sessions is used for the test set. The performance of our model is averaged over the 20 sets. Note that the number of quality gradient bins M is changed according to the value of N. In particular, the number of quality gradient bins corresponding to N = 3 and N = 7 are M = 2 and M = 6, respectively.

Figure 10 shows the average PCC and the average RMSE of all test sets versus the number of the segment quality bins for each video. We can see that, in general, PCC increases and RMSE decreases as N is increased from 3 to 7. However, the performance difference between N=5 and N=7 is very small, especially for BBB and ST videos. In addition, PCC values are equal to or higher than 0.94, and RMSE values are equal to or lower than 0.29 for all test sets when N is equal to or higher than 5. Therefore, to achieve good performance and low complexity for the proposed model, a reasonable number of segment quality bins is 5. Besides, it is interesting that the performance with N=3 is not bad. This could be because the users cannot differentiate many close quality levels. In future work, we will consider customizing both the histogram of segment quality and the histogram of quality gradient.

5.2 Impact of Training Set Size

In this part, we investigate the accuracy of the proposed model with respect to the training set size. Among 49 sessions, we select S sessions for the training set, where S is 20, 24, 26, 28, 29, 30, 32 and 34 for each video. The (49-S) remaining sessions of each video are used for the test set. For each value of S, 20 different training sets of size S are randomly selected. The performance of our model is averaged over the 20 sets.

Figure 11a shows the average PCC corresponding to the training set and the test set versus the training set size

Table 7 Interval values of 3 and 7 segment quality bins.

N	Intervals I_{Q_n}						
	I_{Q_1}	I_{Q_2}	I_{Q_3}	I_{Q_4}	I_{Q_5}	I_{Q_6}	I_{Q_7}
3	[1.0;2.3)	[2.3;3.6)	[3.6;5.0]	NA			
7	[1.0;1.6)	[1.6;2.2)	[2.2;2.8)	[2.8;3.4)	[3.4;4.0)	[4.0;4.6)	[4.6;5.0]

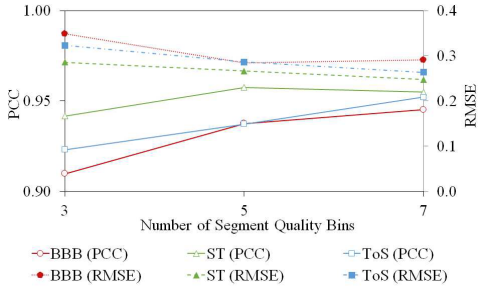

Fig. 10 PCC and RMSE vs. the number of segment quality bins.

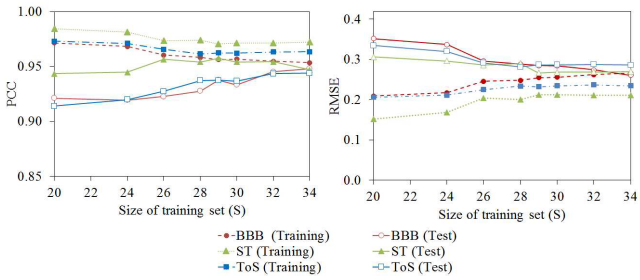

Fig. 11 Performance vs. size of training set for each video: (a): left, PCC; (b): right, RMSE.

for each video. We can see that, for the training set, when S increases, PCC decreases. On the other hand, for the test set, PCC is improved as S increases. When S is equal to or higher than 28, PCC values of the test set are higher than 0.92 for all videos.

Similar results of the average RMSE are shown Fig. 11b. For the training set, when S increases, RMSE increases as well. On the other hand, for the test set, RMSE decreases as S increases. In other words, when S increases, the MOS values of sessions in the test set are more accurately predicted. We can see that, when S is equal to or higher than 28, RMSE values of both the training set and the test set are stable.

Therefore, to achieve stable and good performance for the proposed model, at least 28 sessions should be used for the training set of each video. Based on the dataset we have, the RMSE values are in the range of 0.19~0.27 for training sets and are in the range of 0.26~0.29 for test sets.

5.3 Impacts of Model Parameters and Their Implications

As mentioned, the overall quality is affected by quality amplitudes and quality variations. In previous models, the first component is represented by the average [11], [17] or the median [12] of the segment quality values. Instead of using

the average quality, the study in [13] focuses on the impacts of different quality levels (in the context of 2-3 levels). It is observed that the time (or frequency) on the highest video quality level has a significant contribution to the overall quality. Meanwhile, the second component is represented by the standard deviation of quality values [11] and switching frequency [11], [14], or simply by the minimum quality value [12].

Our proposed model can quantify the weight of each segment quality bin of the histogram, which is the contribution in the overall quality. From Table 3, we can see that the lower the segment quality value is, the smaller the corresponding weight in the model is. One of the implications of this finding is that the highest quality bin has the biggest contribution to the overall quality, which is similar to the finding in [13]. Note that, only two quality levels are considered in [13] while our solution supports any kinds of quality value distributions.

In addition, our quality model also quantifies impacts of variations by the weight of each quality gradient bin. Regarding quality decreases, Table 3 shows that the higher the switch amplitude (or the lower the bin of negative quality gradient) is, the larger the absolute value of the weight (i.e. contribution) in the overall quality is. Especially, the weight of bin -3 is equal to the weight of bin -4 for video #1 and video #3 and content-generic case. This implies that these types of quality decreases have the same negative impact on users and should be avoided. This supports the motivations of proposals to avoid large quality decreases (e.g. [26], [27]).

On the other hand, when the quality increases, the weight of the positive instant gradients in the model is zero. That means, while the impact of switch-down events are significant, the impacts of switch-up events themselves are negligible. This could be explained that, when the quality is increased, the important factor to the overall quality is the time on higher quality values, not the switches themselves. So, a streaming client should focus on switching up to, and then maintaining, a high quality level, rather than gradual switching-up. To better understand the impacts of the average, the median, the minimum, etc., Fig. 12 illustrates some cases of segment quality variations. We can see that these cases have the same statistics (namely the median, the minimum, the average, and the standard deviation of quality values). However, the switch amplitudes of these cases are very different, and so their overall quality measures are also different. Thus, the median, the minimum, the average, and the standard deviation are not able to fully represent quality variations in a streaming session. Similarly, using switching frequency (or number of switches), all types of switch amplitudes could not be differentiated. Moreover, as seen in the evaluation of the model of [11] (Sect. 4.3), the weight of switching frequency turned out to be zero. So, the switching frequency should not be used in a quality model.

From the above results and discussion, it can be seen that using the histograms of segment quality values and quality gradients is more flexible and comprehensive than existing models, which use the medium, the median, the standard deviation, or the minimum, etc., of the segment quality values. Also, our model provides the insights into the contributions of different quality values and gradients, which are not identified or quantified in existing quality models.

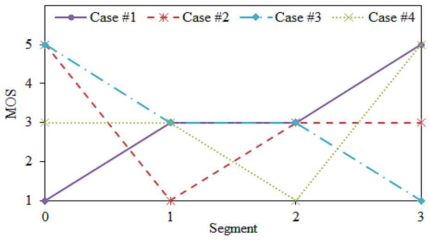

Fig. 12 Examples of the quality variations with the same statistics.

5.4 Impact of Content Characteristics

One of the goals in this study is to investigate the dependence of overall quality on the content. It should be noted that this issue has not been considered in previous studies of adaptive streaming. From Table 3, it can be seen that the behaviors of the model parameter sets for all test videos are consistent and actually dependent on the content. For all test videos, the impacts of quality gradients in bin -3 and bin -4 on the overall quality are the most significant. However, video #2 has high motion activity, so the weight of bin -2 is much higher than in the other videos.

The performance of a quality model is also dependent on the content. For quality models of [12] and [11], the performance in term of PCC is the lowest with video #2. In addition, from Table 4, we can see that generic model for all videos always have lower performance (for both PCC value and RMSE value) than content-specific models for the test sets. It is because that the impacts of the segment quality values and the switch amplitudes on the overall quality depend on the content characteristics (e.g. low or high motion activity). In practice, specific quality models could be generated for different content types using machine learning approaches (e.g. [25]).

It should be noted that, in HAS, the metadata of each video content is sent to the client before a streaming session. So, the above finding suggests that content-specific models could be obtained in advance and then included in the metadata for the purpose of QoS monitoring and adaptation.

5.5 Investigation with a Test Set of Different Resolutions

In this part, an evaluation of the above model given a test set of different resolutions is conducted. Now, the inputs to the model are not the QP values, but segment quality values obtained from a subjective test.

For each of the three videos of Sect. 4.1, an adaptation set of five versions having resolutions of 1280x720, 854x480, 640x360, 426x240, and 256x144 is generated. Each version is encoded with the QP of 24 and the frame rate of 24 fps. For each video, a test set of 20 streaming sessions is generated

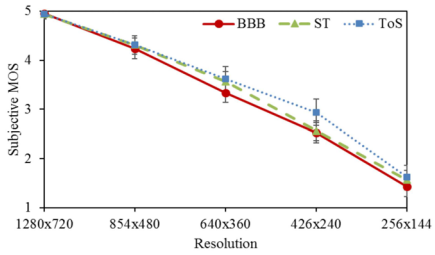

Fig. 13 Subjective experiment results with fixed resolution.

Table 8 Performance of our model for the test set of different resolutions.

Metrics	Video1 (BBB)	Video2 (ST)	Video3 (ToS)	All videos
PCC	0.95	0.92	0.93	0.91
RMSE	0.27	0.29	0.29	0.26

using the adaptation method of [22] and 20 bandwidth traces (extracted from [24]).

To validate our proposed model for the test set of different resolutions, a subjective test is conducted with participations of 25 subjects. This subjective test includes two parts for each video. The first part, which includes 5 test sequences with 5 corresponding fixed resolutions, is used for determining the versions' MOS scores, which are shown in Fig. 13. The quality value of each segment is then the MOS score of the corresponding version. The second part is used for validating the obtained model for the test set of 20 streaming sessions above. The distribution of segment quality values and the distribution of quality gradients are used to predict the overall quality of a session.

Table 8 shows the average PCC and the average RMSE corresponding to the test set of different resolutions for each video. We can see that, our model achieves high PCC values (0.91~0.95) and low RMSE values (0.26~0.29). Therefore, our proposed model, which is obtained by a training set of varying-QP sequences, can also be applied to sequences with varying-resolutions. More investigations with data sets of different quality dimensions will be carried out in our future work.

6. Conclusions

In this paper, we have presented a quality model for HTTP adaptive streaming. The model took into account the segment quality histogram and the quality gradient histogram of a session. It was shown that the proposed quality model had very high prediction performance for different videos. Especially, we showed that switching-up events had a weight of zero, but switching-down events (especially large switches) had significant impacts to the overall quality. It was also found that model parameters were dependent on content characteristics. Based on these findings, various suggestions to improve the quality of streaming service were provided. For future work, we will employ our quality model to evaluate in real-time the performance of different adaptation strategies for adaptive streaming.

Acknowledgment

The authors are grateful to the subjects who participate in the tests and to Prof. C. Timmerer of Klagenfurt University for the bandwidth traces used in this study.

References

[1] T.C. Thang, Q.D. Ho, J.W. Kang, and A.T. Pham, "Adaptive streaming of audiovisual content using MPEG DASH," IEEE Trans. Consumer Electron., vol.58, no.1, pp.78–85, Feb. 2012.
[2] Y. Ou, Y. Xue, and Y. Wang, "Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions," IEEE Trans. Image Process., vol.23, no.6, pp.2473–2486, June 2014.
[3] M. Barkowsky, M. Pinson, R. Pépion, and P.L. Callet, "Analysis of freely available dataset for HDTV including coding and transmission distortions," Workshop on Video Processing and Quality Metrics, 2010.
[4] Y. Liu, Z.G. Li, and Y.C. Soh, "A novel rate control scheme for low delay video communication of H.264/AVC standard," IEEE Trans. Circuits Syst. Video Technol., vol.17, no.1, pp.68–78, Jan. 2007.
[5] A.M. Demirtas, A.R. Reibman, and H. Jafarkhani, "Full reference video quality estimation for videos with different spatial resolutions," in IEEE Conference on Image Processing, pp.1997–2001, Oct. 2014.
[6] Y. Peng and E. Steinbach, "A novel full-reference video quality metric and its application to wireless video transmission," in IEEE Conference on Image Processing, pp.2517–2520, Sept. 2011.
[7] Y. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," IEEE Trans. Circuits Syst. Video Technol., vol.21, no.3, pp.286–298, March 2011.
[8] H. Sohn, H. Yoo, W. De Neve, C.S. Kim, and Y.M. Ro, "Full-reference video quality metric for fully scalable and mobile SVC content," IEEE Trans. Broadcast., vol.56, no.3, pp.269–280, Sept. 2010.
[9] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol.13, no.7, pp.560–576, July 2003.
[10] X. Jingteng, Z. Dong-Qing , H. Yu, and C.W. Chen, "Assessing quality of experience for adaptive HTTP video streaming," in IEEE International Conference on Multimedia and Expo Workshops, pp.1–6, July 2014.
[11] J.D. Vriendt, D.D. Vleeschauwer, and D. Robinson, "Model for estimating QoE of video delivered using HTTP adaptive streaming," IFIP/IEEE International Symposium on Integrated Network Management, pp.1288–1293, May 2013.
[12] Z. Guo, Y. Wang, and X. Zhu, "Assessing the visual effect of non-periodic temporal variation of quantization stepsize in compressed video," IEEE International Conference on Image Processing, pp.3121–3125, Sept. 2015.
[13] T. Hossfeld, M. Seufert, C. Sieber, and T. Zinner, "Assessing effect sizes of influence factors towards a QoE model for HTTP adaptive streaming," in QoMEX, pp.111–116, Sept. 2014.
[14] Y. Shen, Y. Liu, Q. Liu, and D. Yang, "A method of QoE evaluation for adaptive streaming based on bitrate distribution," in IEEE International Conference on Communications Workshops, pp.551–556, 2014.
[15] Y. Liu, S. Dey, F. Ulupinar, M. Luby, and Y. Mao, "Deriving and validating user experience model for DASH video streaming," IEEE Trans. Broadcast., vol.61, no.4, pp.651–665, Dec. 2015.
[16] K.D. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC," in Conference on Consumer Communications and Networking, pp.127–131, Jan. 2012.

[17] Z. Demóstenes Rodríguez, Z. Wang, L.R. Rosa, and G. Bressan, "The impact of video-quality-level switching on user quality of experience in dynamic adaptive streaming over HTTP," EURASIP Journal on Wireless Communications and Networking, vol.2014, no.1, pp.1687–1499, Dec. 2014.
[18] P. Ni, R. Eg, A. Eichhorn, C. Griwodz, and P. Halvorsen, "Flicker effects in adaptive video streaming to handheld devices," Proc. 19th ACM International Conference on Multimedia, pp.463–472, New York, 2011.
[19] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, and N. García, "Perceptual quality of HTTP adaptive streaming strategies: Cross-experimental analysis of multi-laboratory and crowdsourced subjective studies," IEEE J. Sel. Areas. Commun., vol.34, no.8, pp.2141–2153, 2016.
[20] J.D. Vriendt, D.D. Vleeschauwer, and D.C. Robinson, , "QoE model for video delivered over an LTE network using HTTP adaptive streaming," Bell Labs Tech. J., vol.18, no.4, pp.45–62, March 2014.
[21] Xiph.org Test Media, https://media.xiph.org/
[22] T.C. Thang, H.T. Le, H.X. Nguyen, A.T. Pham, J.W. Kang, and Y.M. Ro, "Adaptive video streaming over HTTP with dynamic resource estimation," J. Commun. Netw., vol.15, no.6, pp.635–644, Dec. 2013.
[23] P. Juluri, V. Tamarapalli, and D. Medhi, "SARA: Segment aware rate adaptation algorithm for dynamic adaptive streaming over HTTP," in IEEE ICC, pp.1765–1770, 2015.
[24] C. Müller, S. Lederer, and C. Timmerer, "An evaluation of dynamic adaptive streaming over HTTP in vehicular environments," Proc. Fourth Annual ACM SIGMM Workshop on Mobile Video, pp.37–42, 2012.
[25] R.I.-T.P.913, Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment, 2014.
[26] T.C. Thang, H.T. Le, A.T. Pham, and Y.M. Ro, "An evaluation of bitrate adaptation methods for HTTP live streaming," IEEE J. Sel. Areas. Commun., vol.32, no.4, pp.693–705, April 2014.
[27] H.T. Le, H.N. Nguyen, N.P. Ngoc, A.T. Pham, H.L. Minh, and T.C. Thang, "Quality-driven bitrate adaptation method for HTTP live-streaming," Proc. IEEE ICC'15, QoE-FI WS, 2015.
[28] H.T. Le, H.N. Nguyen, N.P. Ngoc, A.T. Pham, and T.C. Thang, "A novel adaptation method for HTTP streaming of VBR videos over mobile networks," Mobile Information Systems, vol.2016, Article ID 2920850, 11 pages, 2016.

Huyen T. T. Tran received the B.E. degree from Hanoi University of Science and Technology in 2014. She is currently a graduate student and research assistant in the Computer Communications Lab., the University of Aizu. Her research interests include Quality of Experience (QoE), multimedia networking, and content adaptation. She is a recipient of Japanese government scholarship (MonbuKagaku-sho) for graduate study since 2015.

Nam Pham Ngoc received B.E. degree in Electronics and Telecom. from Hanoi University of Science and Technology (Vietnam) and M.Sc. degree in Artificial Intelligence from K.U. Leuven (Belgium) in 1997 and 1999, respectively. He was awarded a Ph.D. degree in Electrical Engineering from K.U. Leuven in 2004. From 2004 until now he has been working at Hanoi University of Science and Technology, Vietnam. His research interests include QoS management at end-systems for multimedia applications, reconfigurable embedded systems and low-power embedded system design.

Yong Ju Jung received the Ph.D. degree from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2005. From 2005 to 2010, he was a Principal Engineer with Samsung Advanced Institute of Technology, contributing to 3-D display processing for 3-D TV. From 2010 to 2014, he was a Research Associate Professor with the Department of Electrical Engineering, KAIST. From 2014 to 2015, he was a Principal Engineer with the System LSI division of Samsung Electronics, Gyeonggi-do, Korea. Since 2016, he has been an Assistant Professor with the Department of Software, Gachon University, Gyeonggi-do, Korea. His research interests include 3-D image/video processing, image quality assessment, human depth perception, and computer vision. He has co-organized special sessions on human 3-D perception and 3-D video assessments in DSP2011.

Anh T. Pham received the B.E. and M.E. degrees, both in Electronics Engineering from the Hanoi University of Technology, Vietnam in 1997 and 2000, respectively, and the Ph.D. degree in Information and Mathematical Sciences from Saitama University, Japan in 2005. From 1998 to 2002, he was with the NTT Corp. in Vietnam. Since April 2005, he has been on the faculty at the University of Aizu, where he is currently Professor and Head of Computer Communications Laboratory with the Division of Computer Engineering. Professor Pham's research interests are in the broad areas of communication theory and networking with a particular emphasis on modeling, design and performance evaluation of wired/wireless communication systems and networks. He has authored/co-authored more than 140 peer-reviewed papers, including 40+ journal articles, on these topics. Professor Pham is senior member of IEEE. He is also member of IEICE and OSA.

Truong Cong Thang received the B.E. degree from Hanoi University of Science and Technology, Vietnam, in 1997 and the Ph.D. degree from KAIST, Korea, in 2006. From 1997 to 2000, he worked as a network engineer in Vietnam Post & Telecom (VNPT). From 2007 to 2011, he was a Member of Research Staff at Electronics and Telecommunications Research Institute (ETRI), Korea. He was also an active member of Korean and Japanese delegations to standard meetings of ISO/IEC and ITU-T from 2002 to 2014. Since 2011, he has been an Associate Professor of University of Aizu, Japan. His research interests include multimedia networking, image/video processing, content adaptation, IPTV, and MPEG/ITU standards.