Hybrid video-quality-estimation model for IPTV services

Kazuhisa Yamagishi, Taichi Kawano, and Takanori Hayashi
NTT Service Integration Laboratories, NTT Corporation,
3-9-11 Midori-cho, Musashino-shi, Tokyo 180-8585, Japan
E-mail: yamagishi.kazuhisa@lab.ntt.co.jp

Abstract— **We propose a no reference hybrid video-quality-estimation model for estimating video quality by using quality features derived from received packet headers and video signals. Our model is useful as a quality monitoring tool for estimating the video quality during use of an Internet protocol television service. It takes into account video quality dependence on video content and can estimate video quality per content, which our previously developed packet-layer model cannot do. We conducted subjective quality assessments to develop the model and validated its quality-estimation accuracy. The quality-estimation results showed that the Pearson-correlation coefficients were larger than 0.9 and the quality-estimation errors were equivalent to the statistical uncertainty of subjective quality.**

I. Introduction

Internet protocol television (IPTV) services have become popular due to advances in broadband IP networks and encoders and decoders (codecs). The quality of experience (QoE) of IPTV services is influenced by the following five factors: what kind of features (e.g., fast/slow motion) the source video content has; how the source video content is encoded and packetized before transmission; how the IP packets are carried over networks; how the IP packets are recovered by forward error correction (FEC) or automatic repeat-request (ARQ); and how packets are decoded and displayed at the client terminal (e.g., set top box (STB)). It is therefore important for service providers and network providers to monitor the QoE affected by these factors when providing an IPTV service.

To monitor end-user QoE, no reference (NR) video-quality-estimation models are essential because quality information is obtained from only a client terminal. Therefore, packet-layer models [1], [2], [3], which take transmitted packet headers as input, bitstream-layer models [4], [5], which take transmitted packet headers and payloads as input, and media-layer models [6], [7], [8], which take video signals as input, have been studied not only for the full reference (FR) mode but also the NR and reduced reference (RR) modes.

Packet-layer models can be used to estimate the average video quality over assumed sets of typical video content, rather than video quality per video content, by using packet headers (e.g., IP, user datagram protocol (UDP), real-time transport protocol (RTP), transport stream (TS), and packetized elementary stream (PES) headers). Packet headers do not include information about codec type (e.g., MPEG2 and H.264), coding parameters (e.g., frame rate and video format), and video content, so video quality dependence on the video codec and content cannot be taken into account by the model. Therefore, these models must make some assumptions with respect to video codec and content.

Bitstream-layer models can estimate the video quality per video content by using transmitted packet headers and payloads. However, transmitted packet payloads are often encrypted to protect the copyright of the video content. Obtaining bitstream information at the client terminal is thus difficult in these cases.

Media-layer models can estimate the video quality per video content by using video signals. However, according to test results of the video quality experts group (VQEG) [9], the quality-estimation accuracy of NR media-layer models was lower than that of peak signal-to-noise ratio (PSNR)-based model. Therefore, these models need to be improved in terms of quality-estimation accuracy.

We developed a packet-layer model [1] that can be used to estimate the average video quality with a low computational load because such a function usually needs to be implemented in client terminals such as home gateways (HGWs) and STBs. However, our model cannot be used to estimate the video quality per content because, by definition, our model does not have access to the video-related bitstream information and video signals.

For taking into account video quality dependence on video content, we propose a hybrid video-quality-estimation model that estimates the video quality degraded by video compression and packet loss by using the average video quality estimated by the packet-layer model and quality features derived from video signals. We conducted subjective quality assessments for developing the hybrid video-quality-estimation model and verified the quality estimation performance. A hybrid video-quality-estimation model for packet loss is for further study.

The remainder of this paper is structured as follows. The concept of a hybrid video-quality-estimation model for IPTV services is described in Section II, and a method of subjective quality assessment is described in Section III, and experimental results are shown in Section IV. We propose a hybrid video-quality-estimation model for IPTV services in Section V, and in Section VI, we discuss our model's validity after applying our model to unknown data sets that were different from the training data sets. Finally, in Section VII, we summarize our findings and suggest possible directions for future studies.

II. Concept of hybrid video-quality-estimation model

We propose a hybrid video-quality-estimation model (Fig. 1). As described in section I, our proposed model translates the average video quality estimated by the packet-layer model [1] into video quality per content with quality features derived from video signals.

In general, video quality depends on video content. However, the qualitative tendency of video quality does not depend on video content. For example, video quality increases with increasing bit rate, and video quality decreases with increasing packet loss. Therefore, the average video quality (Vq_{ave}), which is averaged over an assumed set of typical video content, is expressed by a generalized mathematical equation (e.g., logistic equation and exponential equation) [1], while the model's coefficients differ for each video content. Video quality also depends on codec type (e.g., MPEG2 and H.264), codec implementation (e.g., rate distortion and motion detection algorithms), and coding parameters (e.g., frame rate, group of picture (GoP), and video format). These types of information are not included in transmitted packet headers. However, information about codec type, codec implementation, and coding parameters can be provided, for example, by an IPTV service provider because it must know such information. Therefore, a hybrid video-quality-estimation model (i.e., the model's coefficients) can be optimized for each assumed service condition with such a priori information.

It is possible to estimate the video quality per content (Vq) if the differential video quality ($dVq = Vq - Vq_{ave}$) between the Vq and the Vq_{ave} can be estimated with content-based information (e.g., spatial and temporal features) derived from video signals. Therefore, we tried to develop a hybrid video-quality-estimation model that translates the average video quality estimated by the packet-layer model into video quality per content with video signals.

Our proposed model works as follows. First, the parameter-calculation module calculates parameters (e.g., bit rate, packet-loss information, and content-based information) derived from packet headers and video signals. Second, the estimation module for compression takes the bit rate and content-based information as input and outputs the video quality affected only by video compression. Third, the estimation module for packet loss takes packet-loss information and the video quality degraded by video compression as input and outputs the video quality degraded by video compression and packet loss. The database of the model's coefficient tables stores coefficients that are optimized for an assumed service condition (e.g., video codec and coding parameters) and inputs their coefficients to the estimation modules.

III. Subjective quality assessment

We built a viewing system for deriving video-quality characteristics necessary for developing the hybrid video-quality estimation model.

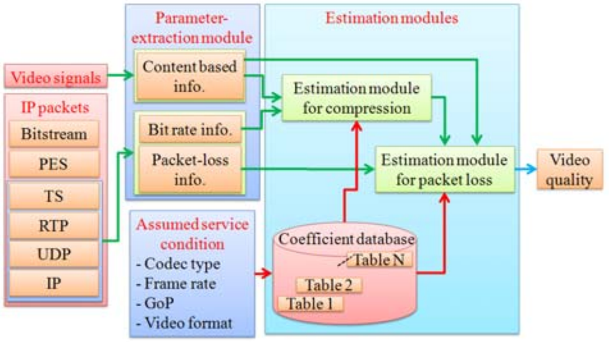

Fig. 1. Hybrid video-quality-estimation model.

TABLE I
Video Contents for Each Group

(a) Group A

No.	Title	Criticality [bit/pixel]
1	Streetcar	0.2
2	Opening ceremony	1.0
3	Crowded crosswalk	0.3
4	Boy and toys	0.2
5	Buildings along the canal	0.3
6	Baseball	0.3
7	Summertime tanning	0.3
8	Flamingos	0.4

(b) Group B

No.	Title	Criticality [bit/pixel]
1	European market	0.3
2	Harbour scene	0.4
3	Whale show	0.7
4	Soccer action	0.8
5	Green leaves	1.1
6	Japanese room	0.2
7	Ice hockey	0.2
8	Weather report	0.1

We used 16 different types of video content (10 seconds each) defined by ITU-R Rec. BT.1210.3 (Table I). Video contents were classified into two groups so that the ranges of criticality (see Fig. 5 of ITU-R Rec. BT.1210.3) would be almost the same. The video contents of group A were used as training data for the model (Experiment 1), and the video contents of group B were used as unknown data for the model (Experiment 2). The experimental parameter was bit rate (BR), as listed in Table II. The experiments had 20 test conditions.

In the subjective quality assessment, video quality was evaluated using an absolute category rating (ACR) method [10]. The quality descriptions on the rating scale were given in Japanese. Twenty four subjects aged 20–39 participated in each experiment. They were non-experts who were not directly concerned with video quality as part of their work and, therefore, not experienced assessors. The subjects viewed each video sequence at a distance of 3H (about 110 cm), where H indicates the ratio of viewing distance to picture height.

Subjective video quality (Vqs) was represented as a mean opinion score (MOS) averaged over the 24 subjects.

TABLE II

EXPERIMENTAL SETTINGS

Parameter	Value	Unit
Codec	H.264 high profile level 4	-
Video format	1440 × 1080	pixel
Frame rate	30	fps
Group of picture	M = 3, N = 15	-
Bit rate	18.0, 16.0, 15.0, 13.4, 11.0, 13.0, 9.6, 7.9, 6.9, 6.2, 5.7, 5.4, 5.0, 4.7, 4.3, 4.0, 3.7, 3.4, 3.0, 2.0	Mbps

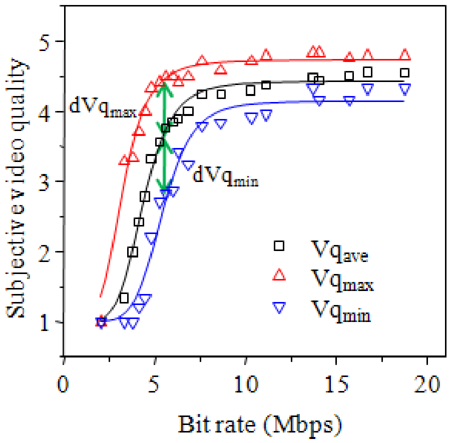

Fig. 2. Subjective video quality characteristics.

IV. EXPERIMENTAL RESULTS

A. Subjective video quality characteristics

These are the video quality characteristics for the video contents of group A. As the BR increased, the average video quality (Vq_{ave} [1]), the maximum video quality (Vq_{max} [2]), and the minimum video quality (Vq_{min} [3]) increased and saturated, as shown in Fig. 2. These curves were formulated by logistic equations.

As BR increased, the difference between Vq_{max} and Vq_{ave} ($dVq_{max} = Vq_{max} - Vq_{ave}$) increased, and then dVq_{max} decreased. On the other hand, as BR increased, the difference between Vq_{min} and Vq_{ave} ($dVq_{min} = Vq_{min} - Vq_{ave}$) decreased, and then dVq_{min} increased. dVq_{max} and dVq_{min} were approximated by a convex equation, as shown in Fig. 3.

B. Video signal characteristics

We now discuss the video signal characteristics for the video contents of group A. In general, video coding difficulty depends on spatial information (SI) and temporal information (TI). Therefore, we first investigate the relationship between SI and TI [10].

The SI is based on the Sobel filter. Each video frame (luminance plane) at time n (Fn) is first filtered with the Sobel filter [$Sobel(Fn)$]. The standard deviation over the pixels (std_{space}) in each Sobel-filtered frame is then computed. This operation is repeated for each frame in the video content and results in a time series of SI of the scene. The average value in the time series is averaged over all SI of the scene. This process can be represented as

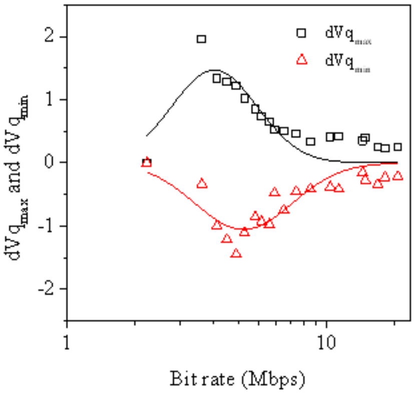

Fig. 3. Differential video quality characteristics.

$$SI = \frac{1}{300}\sum_{n=1}^{n=300} std_{space}(Sobel[F_n(i,j)]), \tag{1}$$

where $F_n(i,j)$ is the pixel at the ith row and jth column of the nth frame at a specific time.

The TI is based on the motion difference feature, $M_n(i,j)$, which is the difference between the pixel values of the luminance plane at the same location in space in successive frames. $M_n(i,j)$ is defined as

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j). \tag{2}$$

The measure of TI is computed as the average of all frames of the standard deviation over space (std_{space}) of $M_n(i,j)$ over all i and j.

$$TI = \frac{1}{299}\sum_{n=2}^{n=300} std_{space}(M_n(i,j)). \tag{3}$$

The correlation between SI and TI is shown in Fig. 4. This figure shows that almost all of the scattered points correlate with each other. From this result, we try to estimate the dVq using TI rather than SI because the computational load of SI is higher than that of TI.

The relationship between BR and TI_{ave} is shown in Fig. 5, where TI_{ave} is averaged over eight video contents at each BR. As the BR increased, the TI_{ave} increased and saturated. This curve is formulated by an exponential equation.

When the video compression difficulty of a video content is high, TI is large. For example, when the difference between the TI of an estimated video content and TI_{ave} ($dTI = TI - TI_{ave}$) is large, the Vq is lower than Vq_{ave}. That is, the smaller dTI becomes, the larger dVq becomes. As described

[1] Vq_{ave} is averaged over eight video contents at each BR

[2] Vq_{max} represents the maximum video quality in eight video contents at each BR

[3] Vq_{min} represents the minimum video quality in eight video contents at each BR

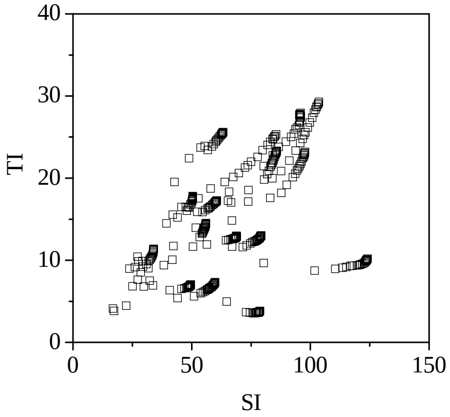

Fig. 4. Relationship between SI and TI.

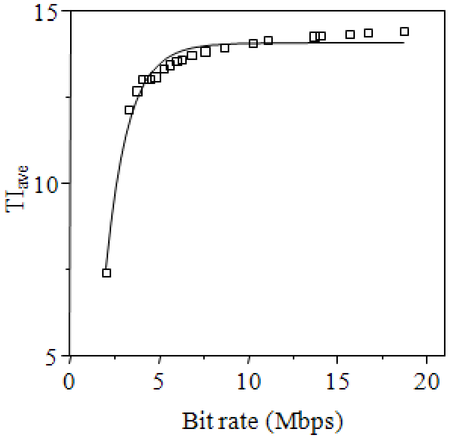

Fig. 5. Relationship between BR and TI_{ave}.

in section IV-A, the characteristics of dVq_{max} and dVq_{min} were approximated by a convex equation. From the results, the relationship between dVq and X could be formulated by a linear equation, as shown in Fig. 6, where X is defined as the following equations.

$$X = |dTI| \cdot dVq_{max} \quad TI < TI_{ave}, \tag{4}$$

$$X = |dTI| \cdot dVq_{min} \quad TI \geq TI_{ave}. \tag{5}$$

V. Hybrid video-quality-estimation model

Using the experimental results, we developed a model for estimating the video quality degraded by video compression. As described in section IV-A, Vq_{ave}, Vq_{max}, and Vq_{min} could be approximated by using the following logistic equations:

$$Vq_{ave} = 1 + v_1 - \frac{v_1}{1 + (BR/v_2)^{v_3}}, \tag{6}$$

$$Vq_{max} = 1 + v_4 - \frac{v_4}{1 + (BR/v_5)^{v_6}}, \tag{7}$$

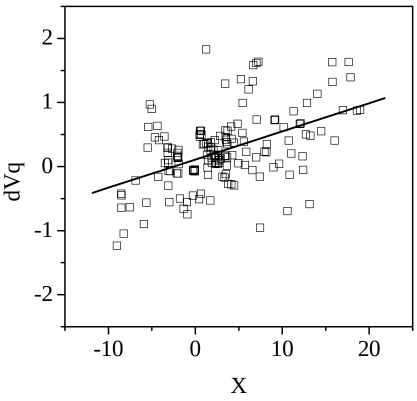

Fig. 6. Differential video quality characteristics.

$$Vq_{min} = 1 + v_7 - \frac{v_7}{1 + (BR/v_8)^{v_9}}, \tag{8}$$

where v_1, ..., v_9 are constants calculated from the subjective data for each assumed service condition.

As described in section IV-B, TI_{ave} could be approximated by using the following exponential equation:

$$TI_{ave} = t_1 + t_2 \exp(-BR/t_3), \tag{9}$$

where t_1, ..., t_3 are constants calculated from the video signals for each assumed service condition.

As described in section IV-B, dVq could be approximated by using the following linear equation:

$$dVq = Vq - Vq_{ave} = d_1 + d_2 \cdot X, \tag{10}$$

where d_1 and d_2 are constants calculated from the subjective data and the video signals for each assumed service condition.

By using the above equations, Vq could be expressed as follows:

$$Vq = Vq_{ave} + dVq. \tag{11}$$

VI. Performance Evaluation of Proposed Model

A. Performance requirements

We used the Pearson correlation (R) $\geq$ 0.84, the root mean square error (RMSE) $\leq$ 0.52, and the outliers ratio (OR) $\leq$ 0.46 as the performance requirements to determine if the quality-estimation accuracy of the model was sufficient. The calculation methods of R, RMSE, and OR are defined in ITU-T Rec. J.247 Appendix 2. Because the above values of R, RMSE, and OR are the same as those in the J.247 model (FR media-layer model), which are averaged over all experiments, it is reasonable to use these values as criteria.

B. Quality-estimation accuracy of packet-layer model

In this section, we show the quality-estimation accuracy of the packet-layer model [1]. Equation 6 was used as the packet-layer model. The results of the R, RMSE, and OR of the model for the training data and unknown data sets are shown in Table

TABLE III

QUALITY-ESTIMATION ACCURACY OF PACKET-LAYER MODEL

	Experiment 1	Experiment 2
R	0.89	0.88
RMSE	0.54	0.64
OR	0.48	0.53

TABLE IV

COEFFICIENTS OF HYBRID VIDEO-QUALITY-ESTIMATION MODEL

v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
3.3	4.7	4.7	3.6	3.6	13	2.8	6.1	6.8

t_1	t_2	t_3	d_1	d_2
19	-32	1.9	0.018	0.060

III. The RMSE and OR values of the model did not satisfy the performance requirements because it cannot use video-related information.

C. *Quality-estimation accuracy of proposed model*

We optimized the hybrid video-quality-estimation model for the subjective data in Experiment 1. Then, we estimated the subjective video quality. The relationship between subjective video quality and estimated video quality is shown in Fig. 7. The R, RMSE, and OR are also shown. In this experiment, they satisfy the performance requirements.

To verify the validity of our model, we used subjective data sets in Experiment 2. The model's coefficients were trained by the subjective data sets in Experiment 1. The quality-estimation accuracy of our model is shown in Fig. 7. The R, RMSE, and OR are also shown. These values satisfy the performance requirements, as described in section VI-A. Therefore, we concluded that our model can be applied to the quality estimation of video quality degraded by video compression and that our model with optimized coefficients for the training data sets of group A was also valid for the unknown data sets of group B.

From these results, we found that the quality-estimation accuracy of the proposed model for the training data and unknown data sets was better than that of the packet-layer model. These improvements were due to the proposed model taking into account the motion difference feature (i.e., TI_{ave}) of the video content.

Although our proposed model satisfied the performance requirements, some scattered points of the "streetcar", "buildings along the canal", "flamingos", and "whale show" video contents were larger than the 95% confidence interval of subjective quality. The video content of "streetcar" and "buildings along the canal" have simple horizontal movement, so motion compensation worked well. On the other hand, motion compensation for "flamingos" and "whale show" did not work well because objects appeared and disappeared in these contents. For these reasons, the estimated qualities of "streetcar" and "buildings along the canal" were lower than subjective qualities of these contents, and the estimated qualities of "flamingos" and "whale show" were higher than the subjective qualities of these contents. That is, the quality-estimation accuracy of our model could be improved by taking into account the difficulty of motion compensation.

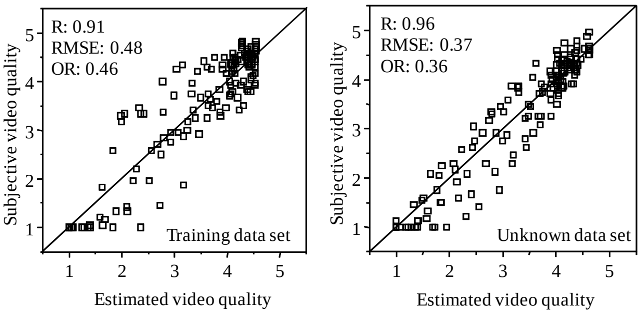

Fig. 7. Quality-estimation accuracy.

VII. CONCLUSION

We proposed a hybrid video-quality-estimation model that translates the average video quality estimated by the packet-layer model into video quality per content with quality features derived from video signals. We first conducted two subjective quality assessments to obtain video quality and video signal characteristics. We then developed a hybrid video-quality-estimation model that can be used to estimate the video quality affected by video compression. Finally, we verified the performance of our proposed model by using two sets of video content. Our proposed model is useful for QoE monitoring of IPTV services.

The following issues call for further study. For taking into account video quality dependence on motion compensation, we need to develop quality features. Our proposed model needs to be expanded to estimate video quality affected by packet loss.

REFERENCES

[1] K. Yamagishi and T. Hayashi, "Parametric Packet-Layer Model for Monitoring Video Quality of IPTV Services," IEEE ICC 2008, pp. 110–114, May 2008.

[2] J. Gustafsson, G. Heikkila, and M. Pettersson, "Measuring Multimedia Quality in Mobile Networks with an Objective Parametric Model," IEEE ICIP 2008, pp. 405–408, Oct. 2008.

[3] A. Raake, M. Garcia, J. Berger, F. Kling, P. List, J. Johann, and C. Heidemann, "T-V-Model: Parameter-based Prediction of IPTV Quality," IEEE ICASSP 2008, pp. 1149–1152, Mar. 2008.

[4] K. Watababe, K. Yamagishi, J. Okamoto, and A. Takahashi, "Proposal of New QoE Assessment Approach for Quality Management of IPTV Services," IEEE ICIP 2008, pp. 2060–2063, Oct. 2008.

[5] O. Verscheurei and X. Garcia, "User-oriented QoS in Packet Video Delivery," IEEE Network, vol. 12, no. 6, pp. 12–21, Nov. 1998.

[6] P. L. Callet, C. Viard-Gaudin, and D. Barba, "A Convolutional Neural Network Approach for Objective Video Quality Assessment," IEEE Transactions on Neural Networks, vol. 17, no. 5, pp. 1316–1327, Sept. 2006.

[7] Y. Fu-zheng, W. Xin-dai, C. Yi-linand, and W. Shuai, "No Reference Video Quality Assessment Method Based on Digital Watermark," IEEE International Symposium on Personal, Indoor and Mobile Radio Communication, vol. 3, pp. 2707–2710, Sept. 2003.

[8] M. Farias and S. Mitra, "No-reference Video Quality Metric Based on Artifact Measurements," ICIP 2005, vol. 3, pp. 141–144, Sept. 2005.

[9] Video Quality Experts Group website, http://www.vqeg.org/.

[10] ITU-T Recommendation P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," Sep. 1999.