Media-layer Objective Video Quality Assessment Technology for Video Communication Services (ITU-T J.247)

Jun Okamoto†, *Keishiro Watanabe, and Akira Takahashi*

Abstract

We summarize the media-layer objective video quality assessment technology standardized as ITU-T (International Telecommunication Union, Telecommunication Standardization Sector) Recommendation J.247. This technology objectively estimates with good precision the user's quality of experience (QoE) of video distorted by encoding and packet loss in IP-based video communication services (IP: Internet protocol).

1. Background

Broadband services and video communication services for personal computer and cell-phone users have been expanding rapidly. Providing such services to customers at an appropriate level of quality requires quality assessment technology that can accurately measure the quality of experience (QoE) of the video communication services. To efficiently design and manage services taking service quality into consideration, we need a video quality assessment technology that enables automatic assessment of this quality.

The video quality assessment technology for assessing the coding distortion of MPEG-2 (MPEG: Motion Picture Experts Group) encoding on standard television (SDTV) signals has been standardized as ITU-T (International Telecommunication Union, Telecommunication Standardization Sector) Recommendation J.144 [1]. This conventional technology, however, cannot assess the effects of the diverse coding systems and bit rates used in video communication services and cannot assess video degraded by packet loss in IP (Internet protocol) networks. To solve this problem, the Video Quality Experts Group (VQEG), an international study group in ITU consisting of video quality researchers, conducted a technical examination [2]. As a result, four systems, including an NTT method, were adopted as an international standard called ITU-T Recommendation J.247 in August 2008 [3].

In this article, first we explain the technical features, target applications, and some application examples of the J.247 standardized technology. Then, we describe the J.247 international standardization algorithm (NTT method) and its quality estimation accuracy. Finally, we mention future developments.

2. Summary of J.247

2.1 Overview

Recommendation J.247 describes a media-layer objective video quality assessment technology that estimates the video quality of video watched by customers from pixel information. Specifically, it quantifies quality by comparing the pixel information of reference and degraded videos. As such, it is a full-reference-type objective quality assessment technology (**Fig. 1**) .

The flow of a video delivery service (a video communication service as an example) is shown in Fig. 1 from left to right. First, the video content to be delivered is encoded to compress the amount of information to be transmitted. Second, the compressed video data is delivered over the network from the delivery server. The customer receives the data at his or her

† NTT Service Integration Laboratories
Musashino-shi, 180-8585 Japan

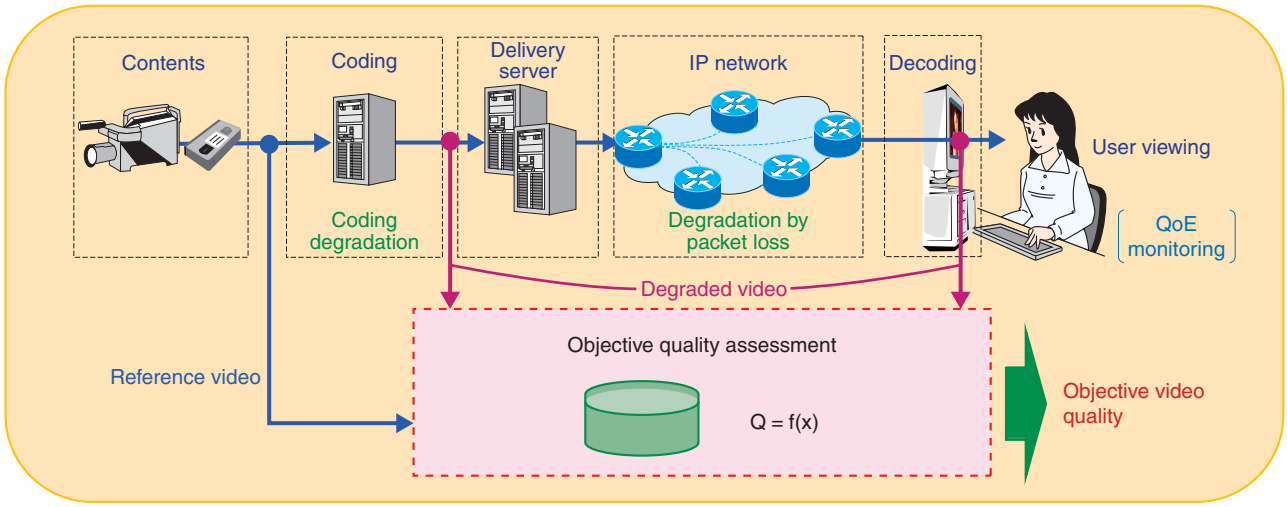

Fig. 1. Full-reference-type objective quality assessment technology.

Table 1. Application domain of J.247.

Video resolution		QCIF (176 x 144), CIF (352 x 288), VGA (640 x 480)
Coding distortion	Codecs	H.264/AVC (MPEG-4 Part 10), VC-1, Windows Media 9, RealVideo (RV10), MPEG-4 Part 2 (Cinepak, DivX, H.261, H.263, H.263+, JPEG2000, -MPEG MPEG-2, Sorenson, H.264 SVC,Theora)
	Bit rates	QCIF: 16 to 320 kbit/s CIF: 64 kbit/s to 2 Mbit/s VGA: 128 kbit/s to 4 Mbit/s
	Frame rates	5–30 fps
Other types of distortion		Transmission error caused by packet loss (visually represented by block distortion and freezing)

location, decodes it, and finally watches the video. Here, the reference video is the content before encoding, and the degraded video is the video either just after encoding or just after decoding. The video quality assessment method compares the pixel information between the reference and degraded videos, and it estimates with good accuracy the user's QoE taking into consideration human visual characteristics.

This technology lets us assess the effects of the diverse coding systems and bit rates used in video communication services. We can assess the quality of video distorted by packet loss in IP networks.

2.2 Target applications

The target applications of this technology are video delivery services for personal computers, smartphones, and other devices and videophone and videoconferencing services. Details of the application domain are given in **Table 1**. The target video resolutions are QCIF (176 × 144 pixels), CIF (352 × 288 pixels), and VGA (640 × 480 pixels). The target video codecs are almost all the main video codecs used in actual video delivery services, such as H.264/AVC, MPEG-4, Windows Media, and RealVideo. The types of video distortion caused by packet loss and affected by the various codec types, bit rates, and frame rates were selected taking into consideration their variations in actual services.

The applications of this technology include in-service quality monitoring at the head end, remote destination quality monitoring when a copy of the source is available, quality verification of archived video, and codec performance comparisons. If we monitor the coding quality in real time at the time of encoding, we can quickly see any problems in the encoding process. When this technology is applied to these

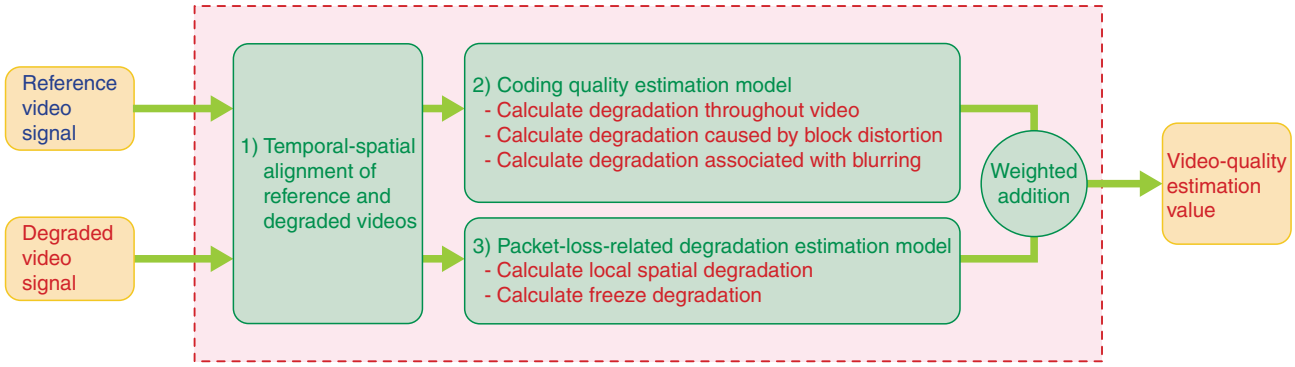

Fig. 2. Video quality objective assessment model.

applications, it provides the following benefits.

(1) Reduces personnel expenses incurred by service providers by automating the pre-delivery content quality check that is currently performed visually.

(2) Raises customer satisfaction through speedy troubleshooting and responses to customer complaints.

(3) Reduces the extent of quality degradation by monitoring and managing the quality experienced by customers in terms of customer sensations.

3. NTT algorithm in J.247

The J.247 international standardization algorithm (NTT method) is shown in **Fig. 2**. Specifically, this method assesses subjective quality influenced by video distortion through the following steps.

Step 1: Temporal/spatial alignment process between the reference and distorted videos

This step matches the pixels and frames of the reference and degraded videos so they can be compared appropriately. Unless all pairs of pixels in the reference and degraded videos are aligned correctly, a pixel-wise full-reference objective video assessment method cannot properly estimate subjective video quality in the following estimation process.

First, macro-alignment is performed. This process consists of temporal/spatial alignment, noise removal, and gain/offset alignment. Temporal/spatial alignment is performed once per pair of video clips, i.e., the reference and degraded videos, to align all the pixels in the spatial and temporal directions. Noise removal removes the influence of high-frequency noise in the degraded video that is imperceptible to humans. Gain/offset alignment matches the pixel value distribution of the reference video with that of the degraded video. This degradation is due to the color arrangement in a decoder or a player (including a video board) that receives the video.

Second, micro-alignment is performed to match the frames between the reference and degraded videos taking into consideration the influence of video frame skipping and freezing.

Step 2: Coding quality estimation model

This step derives three characteristic parameters related to encoding distortion [4].

(1) Overall distortion that occurs throughout all the frames is derived by calculating the luminance difference between the reference and degraded videos.

(2) Distortion in the form of block distortion is derived by calculating the ratio between horizontal and vertical edges and other edges.

(3) Distortion in the form of motion blur is derived by calculating the frame-to-frame luminance differences expressed for each 8 × 8-pixel block between the reference and degraded videos.

Step 3: Packet-loss-related degradation estimation model

This step derives two additional parameters related to video distortion caused by packet loss [5].

(4) Local block distortion that occurs locally in specific frames is derived by calculating the degree of temporal variation of the local block distortions in all the frames when frame-to-frame luminance differences between the reference and degraded videos are large.

(5) Distortion in the form of freeze distortion and variance of the frame rate is derived by calculating the weighted duration of time in which the same image is displayed while the reference image changes.

Step 4: Overall quality estimation

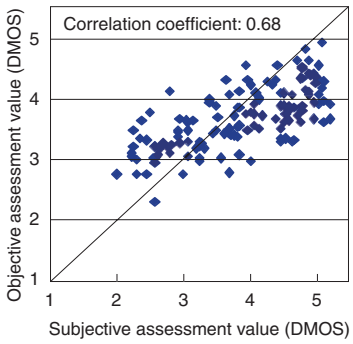

Fig. 3. Results estimated by conventional method (peak signal noise ratio).

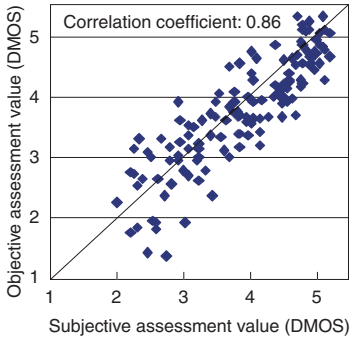

Fig. 4. Results estimated byJ.247 (NTT method).

This step estimates the total effect of quality degradation on subjective quality by calculating the weighted sum of the five characteristic parameters derived in steps 2 and 3.

4. Quality estimation accuracy

Here, we show one verification example of the video quality estimation accuracy of the NTT model. We assessed degraded videos encoded by two encoding methods by using eight different video scenes with VGA resolution that were not used in optimizing the model. The experimental parameters were bit rate, frame rate, and packet-loss ratio. We compared the subjective assessment values derived in the subjective experiment with the objective assessment values. Subjective assessment was performed using the 5-grade ACR-HR (absolute category rating with hidden reference) method* with 24 subjects [6]. The results estimated by the conventional method, which uses the peak signal noise ratio as an objective index of coding quality, and by the NTT method are shown in **Figs. 3** and **4**, respectively. The NTT method estimated subjective quality more accurately than the conventional method. Both Figs. 3 and 4 show the estimated quality of every video individually. In some cases, maximizing the average subjective quality of multiple videos is important, e.g., in optimizing the parameters of the codec. Therefore, we averaged the assessment values of the eight video scenes per experiment condition. The results are shown in **Fig. 5**. The correlation coefficient between the subjective and objective assessment values is 0.94.

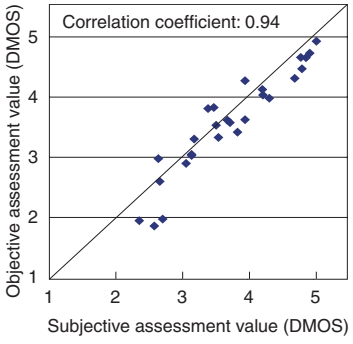

Fig. 5. Estimated results per condition for J.247 (NTT method).

5. Future development

NTT Service Integration Laboratories intends to expand the scope of the NTT method to high-definition television (HDTV) videos and get it standardized. In addition, we intend to contribute to the implementation of quality monitoring systems for video delivery services in the ubiquitous-broadband era as well as the implementation of these technologies in quality estimation and monitoring devices.

References

[1] ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," Mar. 2004.
[2] http://www.its.bldrdoc.gov/vqeg/projects/multimedia/
[3] ITU-T Recommendation J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," Aug.

* 5-grade ACR-HR method: After deriving the mean opinion score (MOS) by using a 5-grade quality scale (5: excellent, 4: good, 3: fair, 2: poor, 1: bad), this method removes the effect of degraded quality on the reference video.

2008.
[4] J. Okamoto, T. Hayashi, A. Takahashi, and T. Kurita, "Proposal for an Objective Video Quality Assessment Method that Takes Temporal and Spatial Information into Consideration," IEICE Trans. on Comm., Vol. J88-B, No. 4, pp. 813–823, 2005.
[5] K. Watanabe, J. Okamoto, and T. Kurita, "An Objective Video Quality Assessment Method for Freeze Distortion in Video Communication Services," IEICE Trans. on Comm., Vol. J90-B, No. 10, pp. 1036–1044, 2007.
[6] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Apr. 2008.

Jun Okamoto
Senior Research Engineer, Supervisor, Service Assessment Group, NTT Service Integration Laboratories.
He received the B.S. and M.S. degrees in electrical engineering from Tokyo University of Science, Tokyo, in 1994 and 1996, respectively. He joined NTT Laboratories in 1996 and has been engaged in quality assessment of visual communication services. He is currently studying objective video assessment methods and leading its standardization in ITU-T SG9 and VQEG. He is a member of the Image Media Quality committee in the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan.

Keishiro Watanabe
Research Engineer, Service Assessment Group, NTT Service Integration Laboratories.
He received the B.S. and M.S. degrees in computer science and communication engineering from Kyushu University, Fukuoka, in 2002 and 2004, respectively. He joined NTT Laboratories in 2004 and has been engaged in quality assessment of visual communication services. He is currently studying objective video quality assessment methods and working on their standardization in VQEG. He is a member of IEICE.

Akira Takahashi
Senior Research Engineer, Supervisor, Service Assessment Group, NTT Service Integration Laboratories.
He received the B.S. degree in mathematics from Hokkaido University, Hokkaido, the M.S. degree in electrical engineering from California Institute of Technology, USA, and the Ph.D. degree in engineering from the University of Tsukuba, Ibaraki, in 1988, 1993, and 2007, respectively. He joined NTT Laboratories in 1988 and has been engaged in the quality assessment of audio and visual communications. He is a Vice-chairman of ITU-T SG12. He has been a co-Rapporteur of ITU-T Question 13/12 on Multimedia QoE and its assessment since 2005. He received the Telecommunication Technology Committee Award in Japan in 2004, the ITU-AJ Award in Japan in 2005, the Best Tutorial Paper Award from IEICE Com. Soc., and the Telecommunication Advancement Foundation Award in Japan in 2008.