

Chapter 23

Real-Time Quality Monitoring for Networked H.264/AVC Video Streaming

Jiarun Song and Fuzheng Yang

Abstract In this paper a packet-layer model for video quality assessment is proposed to monitor the service quality of networked H.264/AVC video streaming in real-time. Using the information extracted from packet header, the frame type is first predicted, based on which the structure of the group of picture (GOP) is estimated since quality degradation caused by packet loss is significantly affected by the loss position. Then the number of impaired frames due to packet loss is calculated as an input of the proposed model. Moreover, since the video quality significantly relies on the motion characteristic of the video content, an algorithm to measure the temporal complexity is designed and incorporated into the proposed model. Eventually, the quality of each GOP can be calculated and the correspondingly video quality is derived using the temporal pooling strategy. Experimental results show that the proposed model remarkably outperforms the model recommended as ITU-T G.1070.

Keywords Video quality · Packet-ayer model · Networked video streaming · Packet loss · Temporal complexity

23.1 Introduction

With the explosive development of technologies in networks and H.264/AVC video compression, the networked video application has become popular in our daily life. However, the quality of these applications cannot be guaranteed in an IP network due to its best-effort delivery strategy. The networked videos are often subject to packet loss, delay, and jitter when transmitted over the network,

J. Song (✉) · F. Yang

State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, China
e-mail: jrsong@stu.xidian.edu.cn

therefore, it is essential to employ quality evaluation of the network video for quality of service (QoS) planning or control in the involved applications [1].

Packet loss is one of the major factors which influence the perceived quality for video streaming over the IP network. A variety of objective quality assessment methods have been reported to evaluate quality degradation caused by packet loss. In [2] and [3] for example, the packet loss rate is employed to estimate the video quality. Considering burst packet loss, more accurate estimations of the related distortion, resorting to both the packet loss rate and the average burst length, have been introduced in [4] and [5]. However, due to the fact that the motion compensation scheme has been commonly used in the video codec, packet loss impacts not only the current video frame but also its successive frames by error propagation. Therefore, since only statistical parameters are considered, the detailed influence of lost packets on a specific video cannot be well captured by these models. To find a more accurate mapping between packet loss and its impact on quality degradation, the temporal complexity, representing the moving character of a video, is taken into account in [6] and [7] by using quantization parameter (QP) and motion vector (MV), respectively. However, these parameters cannot be accessed if the video streaming is encrypted and correspondingly it is difficult to obtain the temporal complexity in this situation. Therefore, how to efficiently and accurately evaluate the video quality affected by the packet loss is still left open.

Since the packet-layer model only utilizes information from packet headers, it is very efficient in quality monitoring due to its low complexity, especially suitable for real-time quality monitoring at network internodes [8]. In this paper, a novel packet-layer assessment model is proposed to evaluate the quality of networked video considering packet loss. The main contributions of this paper can be summarized as follows: (1) though limited information is obtained from the packet header, the motion characteristic of the video content is considered and an efficient estimated temporal complexity is introduced to evaluate the video quality affected by packet loss, (2) knowing the error position caused by packet loss, the quality of each GOP is accurately evaluated and the video quality is assessed by temporal pooling strategy. Specially, in this work it is assumed that direct packet loss always leads to frame loss, which holds true when a corrupted frame is to be discarded by the error control strategy, and is always valid for low bit-rate transmissions where a packet typically contains an entire frame.

The remainder of this paper is organized as follows: [Sect. 23.2](#) presents a detailed description of the frame type detection. The quality assessment model is addressed in [Sect. 23.3](#). Performance evaluation and conclusions are given in [Sects. 23.4](#) and [23.5](#), respectively.

23.2 Frame Type Detection

For a packet-layer model, the coding type of each frame is not directly available since the payload information cannot be accessed. However, due to the fact that different frames have different sensitivity to packet loss, the knowledge of the

23.3 Estimation of Distortion Caused by Packet Loss

When a packet gets lost, not only the current video frame is subject to errors but also the subsequent frames will be affected due to error propagation. The severity of video quality degradation caused by packet loss also depends on the GOP structure and different manners of errors propagate. In this section, key factors that greatly affect the quality of the video will be discussed, based on which the quality of each GOP is estimated and then further integrated to obtain the overall video quality.

23.3.1 Effect of the Number of Impaired Frames

The impaired frame is defined as the frame which is contaminated by the packet loss or by error propagation. When a packet gets lost, the number of impaired frames is not specific since it may occur at different positions. Figure 23.2 shows the distribution of the impaired frames with different positions of packet loss. It is clear that when packet loss occurs in the front of a GOP as shown in Fig. 23.2a, the impaired frames are more than those in Fig. 23.2b which the packet loss occurs closely to the end of the GOP. Consequently, the distortion in these two situations differs.

To evaluate the impact of impaired frames number, the sequences “Soccer”, “Carphone”, “City”, and “Grandma” with a lost packet in the second GOP were examined and the loss position is chosen to lead to different numbers of impaired frames. Then the second GOP is extracted for subjective test, which is described in detail in the next section. The subjective quality of the corresponding video without packet loss was also evaluated, denoted as coding quality. Figure 23.3 shows the relationship between the quality degradation and the number of impaired frames in the second GOP of each sequence at the bit-rate of 150 kbps with the GOP size of 2 s. It can be found that the quality degradation increases approximately linearly with the growth in the number of impaired frames at first, and tends to steady when the number reaching a certain value. In addition, the slope of the linear piece is different for different video sequences. Compared with a low temporal complexity sequence, e.g., “Grandma”, the quality degradation of the sequence with a higher temporal complexity, e.g., the sequence “Soccer”, is more acute.

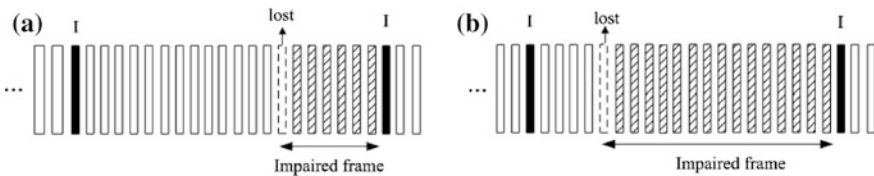


Fig. 23.2 Distribution of impaired frames **a** packet loss occurs in the front of the GOP; **b** packet loss occurs close to the end of the GOP

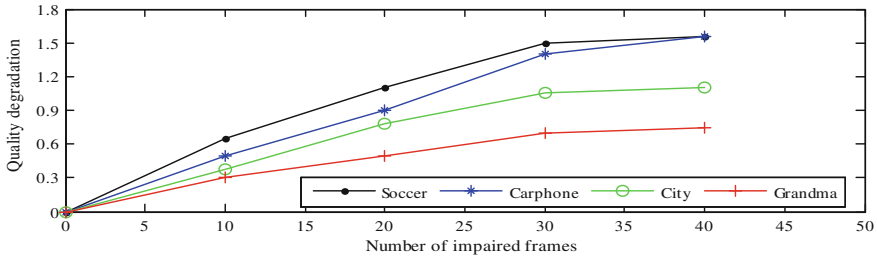


Fig. 23.3 Relationship between quality degradation and impaired frames number for different sequences

From the Fig. 23.3, it is clear that for different movement characteristics of the sequence, the degree of quality degradation due to the packet loss is not identical. The temporal complexity, which generally means the acuteness of changes in the temporal domain, is of great significance to evaluate the quality of a specific video. In [10], the magnitude of motion vectors is used to interpret the temporal complexity. It can also be evaluated based upon the difference between the pixel values (of the luminance plane) at the same spatial location but at successive times or frames as recommended by ITU-T P.910 [11]. However, for a packet layer model, neither the motion vectors nor the residuals are available for quality assessment. Therefore, in our earlier work [12], the temporal complexity δ_T is estimated by the ratio of the frame size for coding I-frames and P-frames as follows:

$$\delta_T = \frac{a_1 \cdot \ln(BR) + a_2}{\ln(B_I) - \ln(B_P)} \quad (23.2)$$

where BR is the average bit-rate which can be calculated by the information extracted from packet header, B_I is the frame size of the I-frame, B_P is the average size of P-frames, and a_1, a_2 are constants empirically obtained by experiments.

23.3.2 Quality Degradation Caused by Coding

Apart from the number of impaired frames and temporal complexity, the degree of quality degradation caused by packet loss is also varying with the video quality related to coding. To examine the impact of the video coding quality, the sequence ‘‘Soccer’’ was encoded at the different target bit-rates with a fixed number of impaired frames. Then the second GOP is extracted for subjective tests to measure its quality degradation. It should be noticed that the video coding quality Q_c can be obtained when the number of the impaired frame is set to zero.

Figure 23.4 shows the relationship between the quality degradation and the number of the impaired frames at different coding bit-rates 48, 80, 128, and 150 kbps, corresponding to different values of coding quality Q_{c1}, Q_{c2}, Q_{c3} and

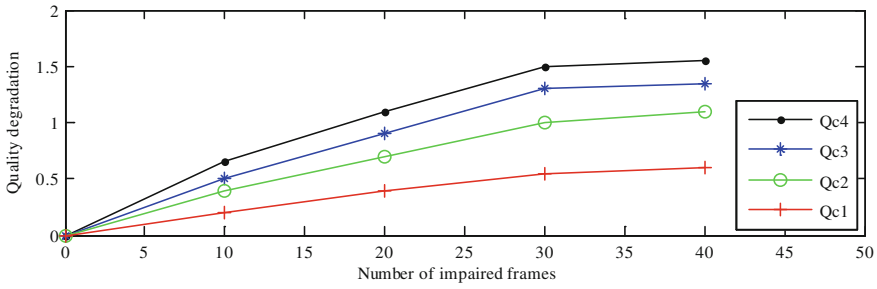


Fig. 23.4 Relationship between the quality degradation and the number of impaired frames at different bit-rates, e.g., “Soccer”

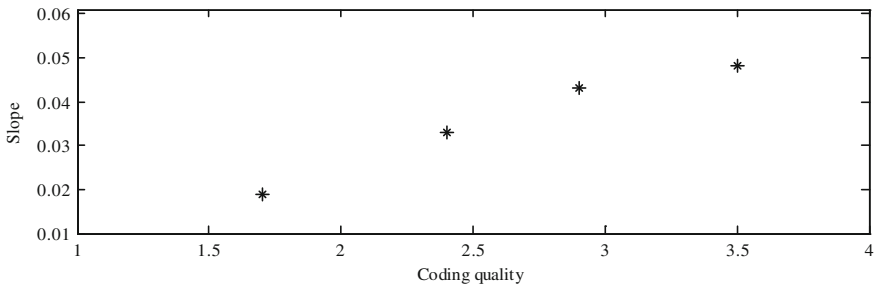


Fig. 23.5 Relationship between the slope of the linear piece in Fig. 23.4 and the coding quality

Q_{c4} . When a packet loss occurs at a high bit-rate, even a slight error due to packet loss will be rather visible, while it may not be very noticeable in a sequence of low quality due to quantization [6]. It can be seen that there is a linear piece in each curve in Fig. 23.4 when the number of impaired frames is less than 30. Figure 23.5 gives the relationship between the slope of the linear piece and the video coding quality.

Based on the analysis above, the quality degradation of the GOP introduced by packet loss, i.e., D_g , can be predicted by the number of impaired frames, the video coding quality, and the temporal complexity, as follows:

$$D_g = \begin{cases} a_3 \cdot \delta_T \cdot \frac{N_{err}}{30} \cdot Q_c N_{err} & N_{err} < 30 \\ a_3 \cdot \delta_T \cdot Q_c N_{err} & N_{err} \geq 30 \end{cases} \quad (23.3)$$

where N_{err} is the number of the impaired frames. Here a_3 is the parameter determined through training. Moreover, compared with the GOP which is not contaminated, the GOP with severe distortions affect the overall video quality much more. Thus, the sequence distortion D_s can be estimated by the following equation:

$$D_s = \left(\frac{1}{N} \cdot \sum_{k=1}^N (D_g(k))^P \right)^{\frac{1}{P}} \quad (23.4)$$

where N is the number of the GOP, $D_g(k)$ is the distortion of the k th GOP, and P equals to 2. Therefore, the video quality Q_s can be calculated as follows:

$$Q_s = Q_c - D_s \quad (23.5)$$

where Q_c is the coding quality which can be obtained as described in [12].

23.4 Experimental Results

All the experiments in this paper used the x264 encoder and the sequences are all QCIF format at 30 frames per second (fps). In simulations, the RTP/UDP/IP protocol stack is employed to packet the encoded data. The parameters defined in the metric were empirically obtained through the least square error fitting, where $r_1 = 0.81$, $r_2 = 0.02$, $a_1 = -0.38$, $a_2 = 1.43$, $a_3 = 0.32$.

In order to verify the efficiency and accuracy of the proposed model for video quality assessment, standard test sequences were employed including “Hall”, “Paris”, “Mother-Daughter”, “Highway”, and “Football”. The bit-rates were set at 64, 96, 128, and 150 kbps, respectively. In addition, the packet loss rates used in the experiments were set to 0.5, 1, 2, 3, 5, and 7 %, respectively, where a random packet loss model was employed to simulate packet loss distribution in IP networks.

Then a subject test was carried out using the Single Stimulate Method (SSM) specified by the Video Quality Experts Group (VQEG). The Mean Opinion Scores (MOS) of reconstructed sequences is obtained by the Absolute Category Rating (ACR) with a 5-point scale from bad to excellent. There were 24 nonexpert observers that rated the quality. The G.1070 model was used for comparison purposes, and the proposed model outperforms G.1070 by getting an increment about 0.015 in PCC and a decrement about 0.049 and 0.007 in RMSE and OR, respectively, as shown in Table 23.1. The scatter plots of the objective scores versus the subjective scores are shown in Fig. 23.6, where the same conclusion can be drawn that the proposed model is very accurate in quality evaluation.

Table 23.1 Performance comparison of proposed model and G.1070

Assessment model	PCC	RMSE	OR
Proposed model	0.913	0.353	0.028
G.1070 model	0.898	0.402	0.035

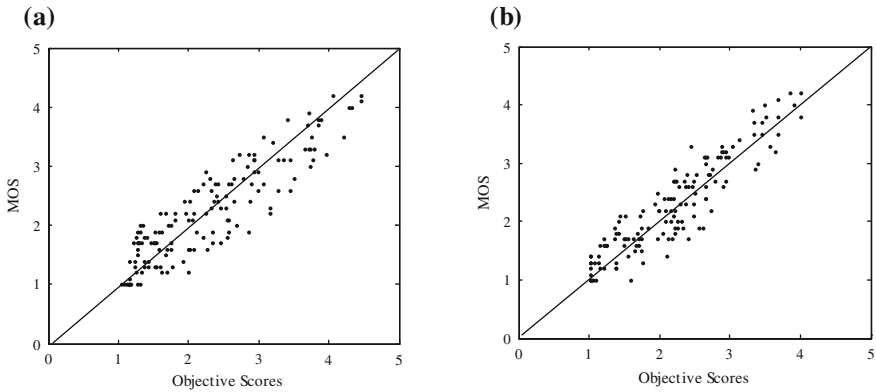


Fig. 23.6 Scatter plot of MOSs vs objective scores, **a** G.1070 Model; **b** Proposed Model

23.5 Conclusions

An accurate packet-layer model is proposed to evaluate the video quality. The impairment of packet loss has been evaluated, based on which the quality degradation of a specific video due to packet loss can be estimated using the number of impaired frames in each GOP. The proposed packet-layer model enables real-time and nonintrusive quality monitoring for networked video streaming.

References

1. Wu HR, Rao KR (2007) Digital video image quality and perceptual coding. CRC Press, London, pp 181–201
2. Verscheure O, Frossard P, Hamdi M (1999) User-oriented QoS analysis in MPEG-2 video delivery. *Real-Time Imaging* 5(5):305–314
3. Boyce JM, Gaglianella RD (1998) Packet loss effects on MPEG video sent over the public internet. In: *Proceedings of 6th ACM multimedia*, pp 181–190
4. You F, Zhang W, Xiao J (2009) Packet loss pattern and parametric video quality model for IPTV. In: *Proceedings of the 2009 8th IEEE/ACIS international conference on computer and information science*, pp 824–828
5. Frossard P, Verscheure O (2001) Joint source/FEC rate selection for quality-optimal MPEG-2 video delivery. *IEEE Trans Image Process* 10(12):1815–1825
6. Yang FZ, Wan S, Xie QP, Wu HR (2010) No-reference quality assessment for networked video via primary analysis of bit-stream. *IEEE Trans Circuits Syst Video Technol* 20(11):1544–1554
7. Wan S, Yang FZ, Xie ZQ (2010) Evaluation of video quality degradation due to packet loss. In: *international symposium on intelligent signal processing and communication systems*
8. Yamagishi K, Hayashi T (2008) Parametric packet-layer model for monitoring video quality of IPTV services. In: *IEEE international conference on communications ICC*

9. Yamada T, Yachida S, Senda Y (2010) Accurate video-quality estimation without video decoding. In: Proceedings of 2010 IEEE international conference on acoustics, speech and signal processing, pp 2426–2429
10. Feghali R, Speranza F, Wang D (2007) Video quality metric for bit rate control via joint adjustment of quantization and frame rate. *IEEE Trans Broadcast* 53(1):441–446
11. ITU-T Recommendation P.910 (2008) Subjective video quality assessment methods for multimedia applications
12. Su HL, Yang FZ, Song JR (2012) Packet-layer quality assessment for networked video. *Int J Comput Commun Control* 7(3):541–549