Research on Video Services with QoE Perception Over Future Wireless Networks

Qinghua Zhu[1(✉)], Jinglu Zhang[1], Gongsheng Zhu[1], and Xiaokui Chen[2]

[1] School of Telecommunication Engineering, Beijing Polytechnic, Beijing 100176, China
3170312109@QQ.com
[2] Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract. Research on video services with QoE (quality of experience) perception, through the analysis of the characteristics of the video service as well as its service quality influence and combined with the practical application of the scene, this paper presents a QoE perception model based on the characteristics of video frame. Based on MPEG compression coding and the independence between the video frames are proposed by this paper, and discussed the frame structure and the effect of FLR for QoE of video services. Simulation results show that QoE perception model proposed can effectively evaluate the influence of FLR of QoE of video services, and has very strong practicability of this scene in the real-time video transmission. The results show that in order to ensure the QoE, the packet loss of video frames must control in the process of transfer.

Keywords: Wireless video services · QoE perception
Future wireless network

1 Introduction

With the rapid development and wide application of in 5G wireless network, video streaming has taken up most of today's network traffic, and the digital video services have brought new challenges to the future wireless network [1, 2], which mainly reflect in: the evaluation on service quality is experiencing the transformation from QoS to QoE [3]. In the conventional mobile communication system, the band width, time delay, jitter, packet loss rate and other QoS parameters are typically considered as the indicator of measuring service quality. For the mobile Internet business especially the video services, the corresponding evaluation of the service quality is not only related to the transmission parameters of the wireless network physical layer, but also related to service characteristics and users' experience environment, and therefore QoE is generally used for evaluating the user quality. QoE is an application and service-level protection mechanism. The service quality evaluated from the perspective of users is not only based on the relevant QoS parameters, but also combined with the service characteristics, which can be better able to reflect the users' subjective feelings. QoS/QoE requirements are varied for different types of services [4–7]. For example, the files downloading, web pages browsing and other data services do not have the

K. Long et al. (Eds.): 5GWN 2017, LNICST 211, pp. 38–51, 2018.
https://doi.org/10.1007/978-3-319-72823-0_5

real-time requirements, but they have the clear requirements for the accuracy of the transmission data; for example, the real-time video services are more sensitive to time delay, so it's essential to make sure that data can be transmitted timely.

The video service QoE perception algorithm plays a significant role in video call, video conferencing, video surveillance, and video on demand and other fields, which is the feedback measurement of users towards the video service quality, but even an important tool to guarantee video quality [8, 9].

The video quality is originally reflected through people's subjective visual feeling towards the video, but because people's subjective evaluation will require a lot of time and efforts, and the differences between individuals have a larger impact on the evaluation results, people are in an urgent demand for some objective, quantitative mathematical model to express their subjective feeling towards the wireless video. This requires to starting from the basic characteristics of the video services of the wireless network to build an video QoE perception system that can penetrate into the wireless network and meet different demands, so as to improve the accuracy of video quality assessment and increase the utilization of wireless network resources as much as possible, safeguarding the service quality of future wireless network.

Currently, the QoE subjective and QoE objective evaluation methods are mainly used as the video service QoE assessment method. Among them, the subjective QoE perception method is to let the subjects to continuously watch the test sequence in a controlled environment, which will be lasted for about 10–30 min, and then is to conduct the subjective ratings for the watched video sequences, thus to finally calculate MOS (Mean Opinion Score). Wherein the controlled environment includes: The selection of the test environment, the viewing distance, the test material (video sequences), and the intervals of time display of the test sequence, etc. For subjective quality assessment method, currently the widely used method is mainly from ITU organization, including DSCQS (Double Stimulus Continuous Quality Scale), SSCQE (Single Stimulus Continuous Quality Scale), DSIS (Double stimulus Impairment Scale), ACR (Absolute Category Rating) and PC (Pair Comparison), etc.

The objective QoE perception method of video quality is to use a mathematical model to analyze the test sequence, thus to finally obtain the perception results of the video quality. The objective perception method can be divided into the following three types according to whether there's the original video for reference: FR (Full-Reference), RR (Reduce-Reference) and NR (No-Reference). The corresponding full reference evaluation method requires a full video reference to compare its quality with the distorted test sequence, thereby getting the score. Reduced reference evaluation method is mainly to discover the distortion through comparison of the extracted reduced reference, so these reference characteristics should be able to better reflect the distortion of coding and channel transmission. The vast majority of the algorithm is to extract the characteristic of space-time domain, as well as to extract the video image edges, but there's also the algorithm of extracting the values calculated based on the pixel values, such as the mean, standard deviation, etc., in addition, there's also the algorithm of extracting video spatial or temporal activity [10] (video activity describes the strong extent of the visual perception motion of video sequences to users [11]). Saviotti proposed evaluation method based on digital watermark, adding watermark in a video image to predict the video distortion at the receiving end through measurement of the watermark image distortion [12]. No-reference evaluation method does not require any reference information, and it derives the video distortion directly through the evaluation of the distortion video. The method has a very great significance for future wireless video services, and it can reduce the transmission content compared with the two evaluation methods of full reference and reduced reference, thereby reducing the band width. However, since no-reference algorithm is too complicated, there's limited research results currently.

The subjective evaluation method is not widely used due to its low real-time, cost of manpower and resources, while the objective method is to use the mathematical models, there's no need to consume too much manpower. In the objective evaluation method, the results of full reference are closer to the results of human eye perception; and the results of the reduced reference are worse than the results of the full reference and it requires the additional band width to transmit the reference signal, therefore, it has little research significance; no-reference method is with higher practical value, because it can use the video features captured currently within a short time to directly obtain the video QoE indicators, which is with the characteristics of simple calculation and high real-time, and therefore it is also the research focus.

Through the study of QoE perception method based on pixel-domain and QoE method based on frame characteristics, this thesis uses mutual independence between MPEG-based coding-decoding characteristics and video frames to propose the transmission strategy with a combination of modulation and coding strategy, video packet re-transmission, and service quality, and also the study of the impact of frame structure and frame loss rate on QoE, and builds the objective evaluation model of the video service transmission.

2 Service Quality Evaluation Method Based on Video Frame Characteristics

The wireless video quality is closely related to its consistency of time domain, this feature determines the real-time of video services, and even does not allow the video to have any Caton phenomenon. In addition, the video quality is also greatly related to the audio corresponding to videos, and therefore it needs the received video and audio information to be strictly synchronous, there's even more strict requirement for the synchronization between voice and image of the video on demand, video conferencing, if the voice is not exactly match with the image, it will seriously affect the viewing effects of the audience. Since the real-time video feature, the corresponding evaluation also needs to have real-time feature. Thus, although the full-reference video quality evaluation method have better performance, it can't meet the requirements, therefore, this thesis adopts the no-reference video quality evaluation method considering the frame loss rate based on video frame characteristics.

In MPEG coding and decoding, the image group includes three categories of frame, namely I-frame, B-frame and P-frame, the loss of different categories of frame will have different impacts on decoding quality of the receiving end [13]. The method proposed in this thesis is to use the independence between MPEG-based coding-decoding characteristics and video frames to study the impact of the frame structure and frame loss rate on QoE.

2.1 MPEG Image Group

The MPEG (Moving Pictures Experts Group) is a video compression coding standard, which mainly uses inter-frame compression coding technology with motion compensation to reduce the temporal redundancy; using DCT (Discrete Cosine Transform) technology to reduce the spatial redundancy of the image, the use of the entropy coding is to reduce the statistical redundancy in terms of information expression. In order to save network band width, video data will be transmitted after being encoded. There are three types of frames in MPEG standard video coding sequence: I-frame (Intra-coded Picture), P-frame (Predictive-coded Picture), B-frame (Bidirectional Predicted Picture). Wherein I-frame is to code by using its own information without any reference to other frames; P-frame refers to the preceding I-frame or P-frame, and it is to code based on the motion differences; B-frame is to make the bidirectional coding by using P-frame before and after it, so it is with the higher compression ratio. When decoding, if B-frame loss will not cause too much impact, P-frame loss will not affect the I-frame, but it will affect P-frame and B- frame associated with it, I-frame loss will make P-frame and B-frame associated with it can't be decoded, so the loss of different frames will have varied impacts on QoE [14].

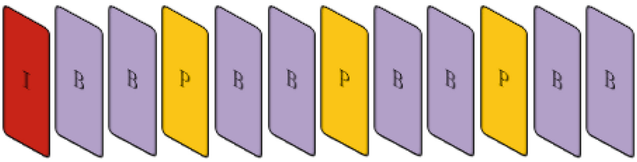

Fig. 1. Open GOP structure diagram

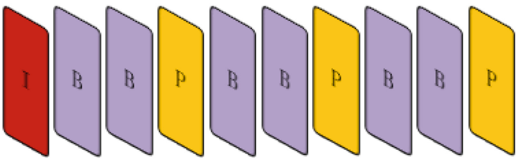

Fig. 2. Closed GOP structure diagram

According to the distance between I-frames in a video sequence, all of the frames are divided into GOP (Group of Picture), so there's only one I-frame in each GOP. A common GOP structure is generally described by using (N, M), wherein N is the distance between one I-frame and the next I-frame, M is the distance between

I-frame and the next P-frame. For example, one GOP may be such a composition: IBBPBBPBBPBB, which is referred to as (12, 3), where 12 represent the length of GOP, 3 is the distance between I-frame and P-frame. For such a frame, the last B-frame is coded with reference to I-frame of the next GOP, and such GOP is called as the open GOP, as shown in Fig. 1. The other is such as GOP of (10, 3): IBBPBBPBBP, the last B-frame is with the reference to P-frame in GOP, so such GOP is called as the closed GOP as shown in Fig. 2. Suppose N represents the total number of I-frame, P-frame and B-frame in one GOP, and N_I, N_P and N_B respectively represents the number of I-frame, P-frame and B-frame, thus it may be inferred that for an open GOP, the number of P-frame is $N/M - 1$, and for a closed GOP, the number of P-frame is $(N-1)/M$, the number of B-frame is $N - N_P - 1$. Therefore, regarding to any kind of GOP, there will be the following formula:

$$\begin{aligned} N_I &= 1 \\ N_P &= [(N-1)/M] \\ N_B &= N - [(N-1)/M] - 1. \end{aligned} \tag{1}$$

In the formula (1), $[a]$ represents the rounding of a. In an open GOP, $N_B = (N_P + 1) \times (M-1)$, and in a closed GOP, $N_B = N_P \times (M-1)$, this thesis uses a variable φ to represents the type of GOP, when $\varphi = 1$, GOP is an open GOP, when $\varphi = 0$, indicating that GOP is a closed GOP. Therefore, φ value can be calculated directly from GOP result, as shown in formula (2):

$$\varphi = \frac{N_B}{M-1} - N_P = \frac{N - 1 - \frac{N-1}{M}}{M-1} - \frac{N-1}{M}. \tag{2}$$

2.2 QoE Perception Model Based on Frame Characteristics

Decodable frame rate refers to the part of the video frames that can be completely and correctly seen by users after they are decoded by the Player, so it is directly related to the user perception and only related to the video signal received by users. The literature [15] introduces the decodable frame rate, supposing that the loss of each frame is independent, and the analytical model of the decodable frame rate depends on the probability of a frame loss, such analytical model has been used in some research on wireless transmission network. In some other studies [16, 17], they compare PSNR and decodable frame rate under the same scene and frame parameters, the conclusion is that decodable frame rate is consistent with the video quality evaluation result based on PSNR, both can reflect the visual effect quality shown by the video to users, so MOS video quality can be estimated by decodable frame rate at the reasonable estimation accuracy.

Decodable frame rate can evaluate the video quality by spending less time on the receiving end to assess the video quality. Decodable frame rate is also known as Q, in this thesis Q will be used as the assessment measurement of video services in the wireless network.

Prior to the video transmission, the first is to divide each frame of the images of video data into certain number of data packets for transmission, real-time transmission of video services is conducted according to the corresponding needs, since there's large amount of data and demanding real-time, UDP connection is generally adopted for transmission of video data in order to avoid the network collapse. However, the characteristics of best efforts of UDP show that the non-reliability of UDP is bound to make the packet loss become a problem must be considered, the missing of a packet in one frame will make the whole video impossible to be decoded, and this will certainly affect the video QoE, and it is also a major factor affecting QoE in wireless network video services.

Different types of frame loss will make the number of decodable frame loss to be varied, herein the decodable frame rate is defined as Q to represent the quality of the video sequence, and Q represents the ratio of the number of decodable frame and the total number of received frame. Supposing, and respectively represents the loss probability of I-frame, P-frame and B-frame, and there's the mutual independence between frames, then for an open GOP structure that can be expressed as (N, M), the decodable frame rate Q of its video sequence can be derived as follows:

(1) The mathematical expectation for calculating the number of decodable I-frame is N_{DI}. In a GOP, I-frame is to code by using its own information without any reference to other frames, therefore $N_{DI} = (1 - FLR_I)$;
(2) The mathematical expectation for calculating the number of decodable P-frame is N_{DP}. In a GOP, only when the preceding I-frame and P-frame can be decoded, can P-frame be decoded accordingly. $S(P_n)$ represents probability for n^th P-frame to be decoded:

$$S(P_1) = (1 - FLR_I) * (1 - FLR_P)$$
$$S(P_2) = (1 - FLR_I) * (1 - FLR_P)^2$$
$$\ldots$$
$$S(P_{N_P}) = (1 - FLR_I) * (1 - FLR_P)^{N_P}$$

Thus, there's:

$$N_{DP} = \sum\nolimits_{i=1}^{N_P} S(P_i) = (1 - FLR_I) \sum\nolimits_{i=1}^{N_P} (1 - FLR_P)^{N_P}. \tag{3}$$

(3) The mathematical expectation for calculating the number of decodable P-frame is N_{DB}. In a GOP, only when the preceding or the following I-frame and P-frame can be decoded, can P-frame be decoded accordingly. Especially the last B-frame is based on the preceding P-frame coding and following I-frame coding, thus the last B-frame is affected by two I-frames. $S(B_n)$ represents the probability for n^th B-frame can be decoded:

$$S(B_1) = (1 - FLR_I) * (1 - FLR_P) * (1 - FLR_B)$$
$$S(B_2) = (1 - FLR_I) * (1 - FLR_P)^2 * (1 - FLR_B)$$
$$\ldots$$

$$S(B_{\frac{N}{M}-1}) = (1 - FLR_I) * (1 - FLR_P)^{\frac{N}{M}-1} * (1 - FLR_B)$$

$$S(B_{\frac{N}{M}}) = (1 - FLR_I)^2 * (1 - FLR_P)^{\frac{N}{M}-1} * (1 - FLR_B)$$

Therefore, the formula (4) is concluded:

$$N_{DB} = (M-1)\sum_{i=1}^{\frac{N}{M}} S(B_i) = (M-1)(1 - FLR_I)(1 - FLR_B) \left[\sum_{j=1}^{\frac{N}{M}-1} (1 - FLR_P)^j (1 - FLR_B) + (1 - FLR_I)(1 - FLR_P)^{\frac{N}{M}-1}\right]. \quad (4)$$

(4) Calculate the ratio of decodable frame Q, as shown in formula (5).

$$Q = \frac{N_{DI} + N_{DP} + N_{DB}}{N}. \quad (5)$$

Formula (5) can be further expressed as:

$$Q = \frac{(1 - FLR_I) + (1 - FLR_I)\sum_{i=1}^{N_P}(1 - FLR_P)^i}{N} + \frac{(M-1)(1 - FLR_I)(1 - FLR_B)\left[\sum_{i=1}^{N_P}(1 - FLR_P)^i + (1 - FLR_I)(1 - FLR_P)^{N_P}\right]}{N}. \quad (6)$$

The last part $(1 - FLR_I)(1 - FLR_P)^{N_P}$ of formula (6) is only related to the last B-frame that needs to take a reference of the decoding I-frame in the following GOP, so decodable frame rate Q of the video sequence for any GOP structure can be expressed as:

$$Q = \frac{(1 - FLR_I) + (1 - FLR_I)\sum_{i=1}^{N_P}(1 - FLR_P)^i}{N} + \frac{(M-1)(1 - FLR_I)(1 - FLR_B)\left[\sum_{i=1}^{N_P}(1 - FLR_P)^i + \varphi(1 - FLR_I)(1 - FLR_P)^{N_P}\right]}{N}. \quad (7)$$

Q is a measurement of the video quality of the receiving end, and we have finally established the method for video QoE assessment based on frame loss rate.

Formula (7) can be further simplified, it is considered that there's little change of the frame loss rate of a video sequence, therefore, it is assumed that each type of frames has the same frame loss rate, if the frame loss rate of I-frame, P-frame and B-frame is the same when there's the transmission of the video services, namely, $FLR = FLR_I = FLR_P = FLR_B$, then

$$Q = \frac{(1 - FLR) + (1 - FLR)\sum_{i=1}^{N_P}(1 - FLR)^i}{N} + \frac{(M - 1)(1 - FLR)^2\left[\sum_{i=1}^{N_P}(1 - FLR)^i + \varphi(1 - FLR)^{N_P + 1}\right]}{N}. \tag{8}$$

3 Simulation Results and Performance Analysis

3.1 Experimental Network and Its Encoding Parameter Setting

Figure 3 is the architecture of the experimental network, the entire network environment is built based on NS2 [18] environment. Network has two emission sources, one is to generate the background stream of CBR (Constants Bit Rate), which is to transmit the CBR data packet at a rate of 1 Mbps; and the other is to generate MPEG4 video stream, which is to transmit the data packet at a rate of 10 Mbps. The router is to connect the wireless network at a rate of 10 Mbps, the connection frequency is 1 ms. The wireless network is to transmit the data to the mobile endpoint via WLAN at a rate of 11 Mbps.

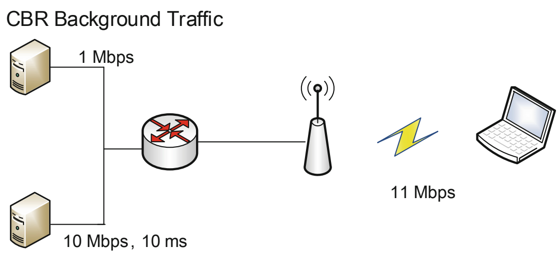

Fig. 3. Experimental network architecture

In the simulation, this thesis assumes that: (1) when connected to the network, there is no loss of the video packet; (2) the loss of a video packet will make it impossible to decode the entire video frame.

In this thesis, the video can be divided into three types according to the time characteristics and space characteristics: SM (Slight Movement), GW (Gentle Walking) and RM (Rapid Movement). Parameter settings based on application layer and the transport layer are as shown in Table 1.

Table 1. Parameter values of application layer and the transport layer:

Video sequences	Frame rate (fbps)	Transmission rate (kb/s)	Link rate (kb/s)	FLR
SM	10, 15, 30	18	32, 64, 128	0.01, 0.05, 0.1, 0.15, 0.2
		44		
		80		
GW	10, 15, 30	18	128, 256, 384	0.01, 0.05, 0.1, 0.15, 0.2
		44		
		80		
RM	10, 15, 30	80	384, 512, 768	0.01, 0.05, 0.1, 0.15, 0.2
		104		
		512		

3.2 Comparative Analysis of Q and PSNR

The experiment of this thesis is to use PSNR and Q indicators in Evalvid [19] to study the effectiveness of video service quality evaluation. Three different types of videos are used respectively in the simulation, namely, foreman, highway and coastguard. Figure 4 is the Q and PSNR comparative figure of different types of videos.

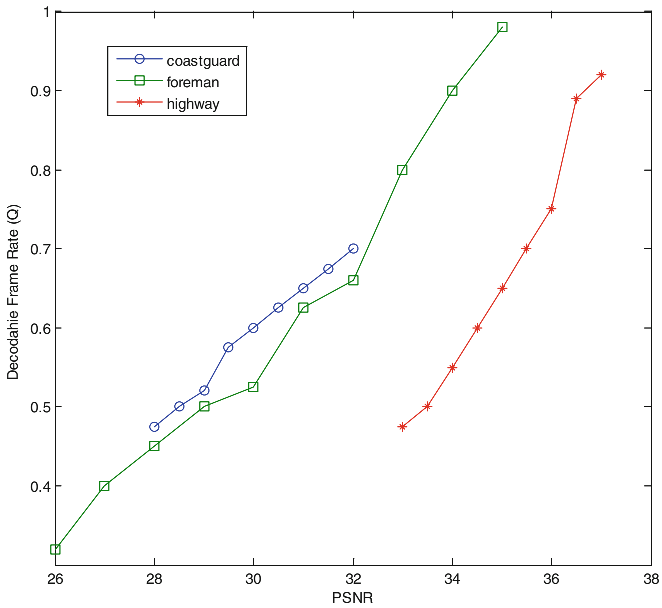

Fig. 4. Comparison of PSNR and Q in different types of video

It can be known from the simulation results that when Q is increased, PSNR is also increased accordingly. Based on the relationship between Q and PSNR, it can be considered that Q indicator is the same to PSNR to reflect the service quality of the video.

3.3 Analysis of Factors Affecting Q

Figure 5 compares the effect of changes in B-frame on video quality. Figures 6 and 7 respectively compares the impact of the total length of the video frame on video quality under the condition of open and closed GOP. Figure 8 compares the impact of frame on video quality whether GOP is open.

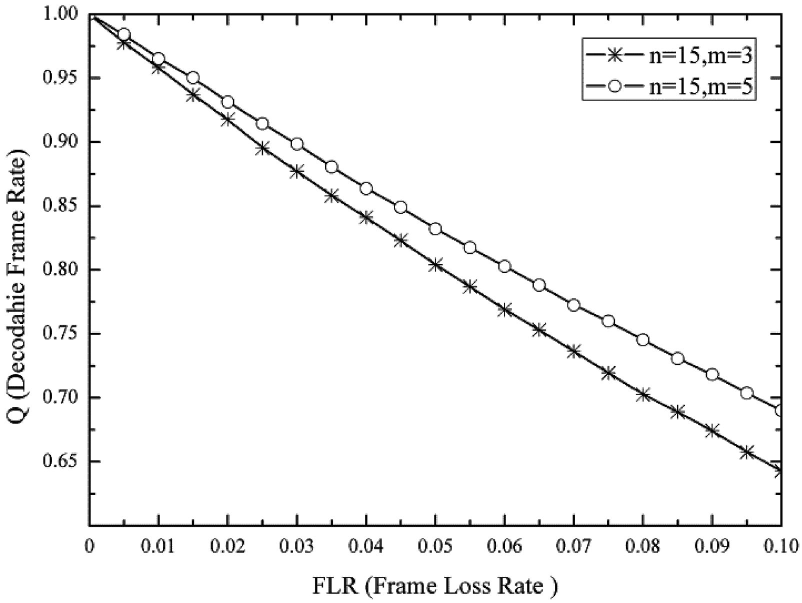

Fig. 5. Effects of changes in B frames on video quality

Through the analysis, it can be found that: (1) With the increase of video frame loss rate, the video quality at the receiving end shows the decreasing trend, and therefore in the wireless video transmission, it's essential to minimize the frame loss rate as much as possible in order to improve the quality of the service experience; (2) when the frame loss rate is smaller, the more B-frames in a sequence, the higher the video quality; (3) when the frame loss rate is smaller, the longer a GOP sequence, its video quality will be lower; (4) when the frame loss rate is smaller, there's little impact on the corresponding service quality whether a video frame is open.

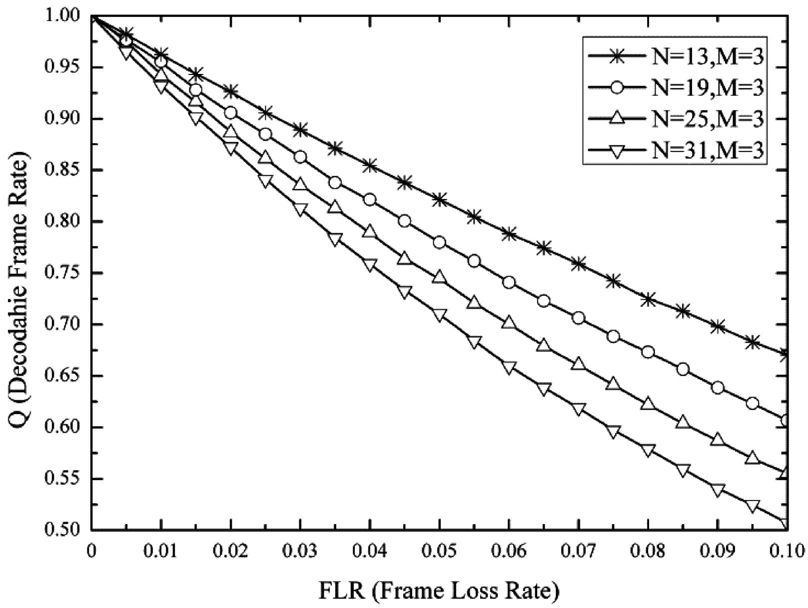

Fig. 6. Effects of open GOP video with different length frame on video quality

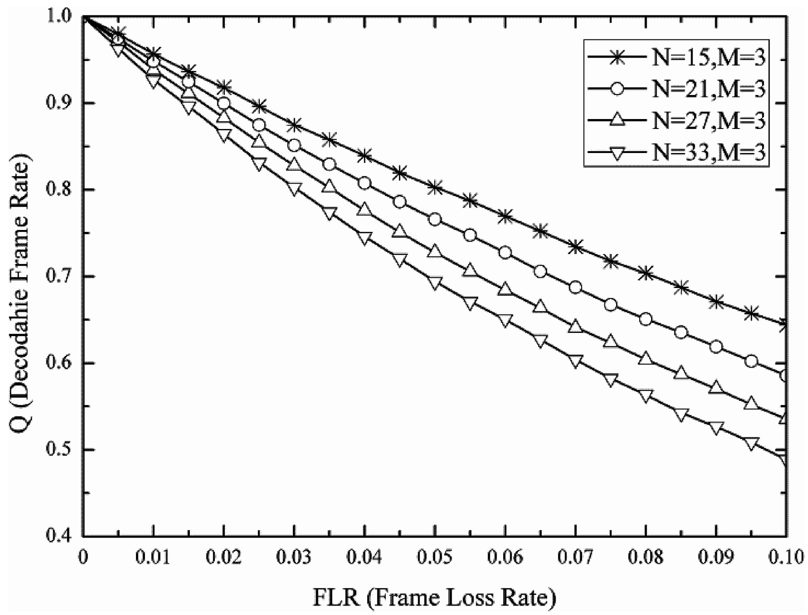

Fig. 7. Effects of closed GOP video with different length frame on video quality

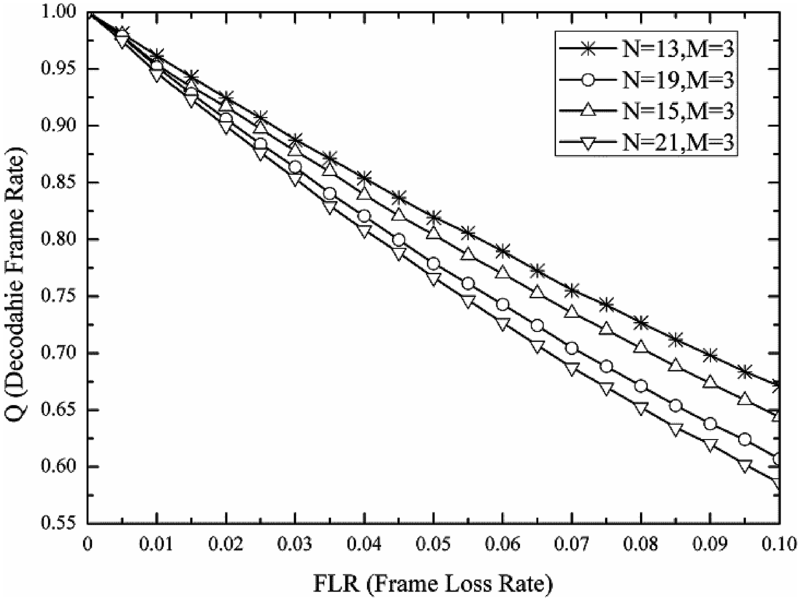

Fig. 8. The effect of video quality in open or closed GOP

4 Conclusion

The wireless network services are with the interaction, real-time, integration and other characteristics, thus the evaluation model is required to be a non-parameter evaluation model with good real-time. The impact of wireless network on the video is mainly for the packet loss resulting in frame loss, so this thesis is committed to finding a frame-level evaluation model. The perception method based on the pixel domain is simple and intuitive, and easy to calculate, but it only considers the energy of the signal, overlooking the impact of video content on the perception quality, which sometimes makes the results to be inconsistent with human eye perception, especially for the same original signal, two distortion signals with the same PSNR are not necessarily the same. In addition, PSNR needs to take a longer time to compare the corresponding frames of the original video and distortion video to obtain the evaluation parameters, and therefore it does not meet the requirements. This thesis presents a transmission strategy with a combination of modulation and coding strategy, video packet retransmission and service quality can meet the needs of future wireless network for the video services.

References

1. Zhang, H., Dong, Y., Cheng, J., Hossain, M., Leung, V.C.M.: Fronthauling for 5G LTE-U ultra dense cloud small cell networks. IEEE Wirel. Commun. **23**(6), 48–53 (2016)
2. Zhang, H., Liu, N., Chu, X., Long, K., Aghvami, A., Leung, V.: Network slicing based 5G and future mobile networks: mobility, resource management, and challenges. IEEE Commun. Mag. **55**(8), 138–145 (2017)
3. Shao, H., Lu, L., Wen, X., Zhang, H., Chen, Y., Hong, Y.: Content-aware video QoE coverage analysis in heterogeneous wireless networks. Wirel. Pers. Commun. **90**(2), 487–502 (2016)
4. Shao, H., Wen, X., Jing, W., Lu, Z., Zhang, H., Chen, Y.: Joint optimization of quality of experience and power consumption in OFDMA multi-cell networks. IEEE Commun. Lett. **20**(2), 380–383 (2016)
5. Ma, W., Zheng, W., Zhang, H., Wen, X.: MOS-driven energy efficient power allocation for wireless video communications. In: Proceedings of IEEE Globecom (2012)
6. Ma, W., Zheng, W., Zhang, H., Wen, X., Lu, Z.: Distortion-guaranteed energy saving power allocation for downlink OFDMA multiuser video transmission. J. Inf. Comput. Sci. **9**, 3011–3019 (2012)
7. Gómez, G., Hortigüela, L., Pérez, Q., Lorca, J., García, R., Aguayo-Torres, M.C.: YouTube QoE evaluation tool for android wireless terminals. EURASIP J. Wirel. Commun. Netw. **2014**, 1–14 (2014)
8. He, S., Tang, A., Zhang, H.: High-performance implementation of OFDM-MIMO base-band in wireless video system. Inf. Technol. J. **13**, 1678–1685 (2014)
9. Shen, Y., Jiang, C., Quek, T., Zhang, H., Ren, Y.: Device-to-device cluster assisted downlink video sharing - a base station energy saving approach. In: Proceedings of IEEE GlobalSIP, Atlanta, Georgia, USA (2014)
10. Yang, C., Ro, C.W.: Resource utilization and QoS analysis for differentiated services in wireless cellular networks. In: 2012 8th International Conference on IEEE Computing and Networking Technology (ICCNT), pp. 264–267 (2012)
11. Huo, Y., Jing, T., Zhang, D., Feng, Y.: Error distribution feedback quantization scheme based on HVS. In: Signal Processing, vol. 25, no. 4, pp. 537–542 (2009)
12. Bucciol, P., et al.: Cross. In: Global Telecommunications Conference, GLOBECOM 2004, vol. 5, pp. 3027–3031. IEEE (2004)
13. Khan, A., Sun, L., Ifeachor, E.: Impact of video content on video quality for video over wireless networks. In: Fifth International Conference on Autonomic and Autonomous Systems, ICAS 2009, pp. 277–282. IEEE (2009)
14. Ma, X., Liu, J., Jiang, H.: Energy-efficient mobile data uploading from high-speed trains. Mob. Netw. Appl. **17**(1), 143–151 (2012)
15. Cheng, R.S., et al.: Improving performance of MPEG-based stream by SCTP multi-streaming mechanism. In: 2010 5th International ICST Conference on Communications and Networking in China (CHINACOM), pp. 1–6 (2010)
16. Ke, C.H., et al.: An evaluation framework for more realistic simulations of MPEG video transmission. J. Inf. Sci. Eng. **24**(2), 425–440 (2008)
17. Singh, K.D., Rubino, G.: Quality of experience estimation using frame loss pattern and video encoding characteristics in DVB-H networks. In: Packet Video Workshop International 2011, pp. 150–157 (2011)

18. NS2. http://www.isi.edu/nsnam/ns/
19. Klaue, J., Rathke, B., Wolisz, A.: EvalVid – a framework for video transmission and quality evaluation. In: Kemper, P., Sanders, William H. (eds.) TOOLS 2003. LNCS, vol. 2794, pp. 255–272. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45232-4_16