Video Quality Assessment Metric based on Network Packet Loss, Jitter and the Bit Rate

Chao Feng[1] Lian-Fen Huang[2*] Feng-Lian Gao[3]

School of Information Science and Engineering, Xiamen University

Xiamen, Fujian, PROC

[1]chaof @xmu.edu.cn, [2]lfhuang@xmu.edu.cn, [3]flgao@xmu.edu.cn

[*] Corresponding author

Received 18 August 2015; Revised 19 December 2015; Accepted 14 January 2016

Abstract. Video packets are discarded during transmission when the network signal attenuation is bad. Intensive study has been performed on the delay and jitter phenomena. Jittering is caused by delaying. It is extracted to establish assessment model with video bit rate. At last, packet loss and jitter are studied together. Several video sequences are used and WAN emulator is used to generate loss and jitter. The experiment of subjective assessment is done strictly. For packet loss, this paper analyzes the packet loss rate, packet loss type and the length which affect the video quality. The linear neural network is used to establish the evaluation metric. The establishment of a transport network simulation platform simulates network packet loss, jitter and bit rate simultaneously for the video transmission quality evaluation model. Finally, for jitter, mainly with the jitter buffer size for the normalization process, and then together with the packet loss and video bit rate as the parameters, the use of BP neural network evaluation is the metric.

Keywords: video quality assessment, network packet loss, neural network

1 Introduction

Unlike circuit-switched networks, the packet-switched network is not a physically point to point circuit. Packet data is transmitted from the sender to the receiver along the distribution router. Routed packets are generally queued using FIFO. However, multi-path fading in a wireless network will produce signal attenuation issues, obstructions caused by packet loss, damage to the package is rejected by, or defective network hardware, drivers and other circumstances can cause packet loss. Packet loss produces decoded message frames, resulting in video generation blur, mosaic and corruption of the entire frame. With enough corrupted frame the video image is jammed, preventing video playback. Thus, video transmission quality loss research has important practical value.

2 Packet Level Impact on VQA

The video signal quality will suffer distortion because the analog image signal must be encoded into a digital signal. In the network transmission distribution, the network system design parameters lead to code stream generation packet loss, delay and jitter, especially through a wireless network stream. This chapter will analyze the transmission process stream factors causing the channel loss phenomenon. Chapter three will discuss mathematical statistics, the neural network model parameters and video quality.

2.1 The Jitter

The desired delay time is defined as the time the data is transmitted from one end of the network to the other end. There are four types of delay: Send delay, propagation delay, processing delay and queuing delay. Jitter is defined as the change in the data set delay [1]. The first set of data sent by express time represents the first group data reception time. The delay and jitter formulas are as follows：

$$Delay_i = T_{2,i} - T_{1,i} \tag{1}$$

$$Delay_{avg} = \frac{\sum_{i=1}^{n} Delay_i}{n} \quad (2)$$

$$Jitter_i = Delay_i - Delay_{avg} \quad (3)$$

Jitter impacts cache related video quality and playback. When the players are set at a higher cache value, the cache data in the next stream buffer reduces the impact of the delay. Here network delay jitter and the Player caches impact video quality are simulated.

2.2 The Packet Loss

The actual network channel is usually a memory channel. A Markov Chain [2] is used to capture the temporary loss correlation established based on the 6 error pattern Gilbert model [3]. Network packet loss program simulation is shown in Figure 1. The program defines class Parameters for the source stream, error patterns and transport stream file name, dropping the data store file name. A Packet class is used to analyze the input RTP header information and slice RTP header information with data stream format. After initialization, the Simulator is run to obtain the source stream, bit error rate and packet loss starting position. The error pattern is selected according to BER. The stream is read at a predetermined position to judge the error pattern. If a packet is lost the packet is dropped and stored in a data file for later use. Returns read the data step and cycles until the end of the stream is read.

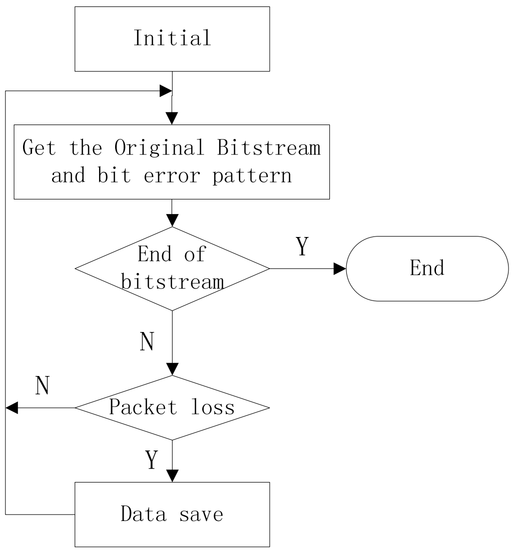

Fig. 1. Packet loss simulation with bit error pattern

During transmission the receiver can detect whether a packet is lost based on the RTP continuity packet number and calculate the packet loss rate. When doing the video encoding each video frame can be divided into several slices and each slice is packaged into a RTP packet [4]. As shown in Figure 2 a slice contains one or more macro blocks. A macro block is divided into smaller sub- macro blocks with thin yellow and purple segmented lines. With visible, eyes, teeth, collar and other complex regional picture, the macro block is divided more finely into sub macro blocks. Typically, a longer packet length contains more information.

Fig. 2. The slice and Macroblock

Packet loss rate and packet loss length are important parameters, but not important enough to be used to judge video quality. Packet loss location is also critical information. Even if the packet loss ratio is the same, dropping the same video frame in different positions may create different effects. More obvious changes in the image or data stream that the human eye is more sensitive to can cause more serious deterioration phenomenon. Therefore, the position of the lost frame also has a greater impact on video quality.

3 The Simulation of Packet Level Impact on VQA

The Sirannon software package and send H.264 files into the WAN emulator (WANem). The VLC receives and decodes the stream. The WANem simulates a network transport that generates the delay or jitter process. Due to different amounts of delay and jitter the VLC cache can be set into different sizes. The whole network transmission platform is shown in Figure 3.

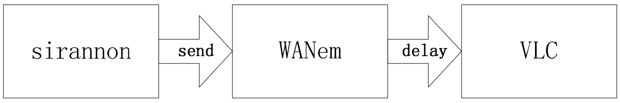

Fig. 3. the Simulation Platform

3.1 Sirannon

Sirannon is the latest standard tool chain (HRC Automation Toolchain) developed by VQEG [5]. The tool chain provides a plurality of modules established from video file encoding, packaging, sending, transmitting, receiving, decoding the set of processes. Sirannon software can be used to encapsulate H.264 stream file sent to a specified IP. This stream supports multiple video formats and streaming protocols. It provides software interfaces for the Linux environment. Element avc-reader read a H.264 pure stream and encapsulates a NAL unit into the media kit. The PES-packetizer-audio/video separately labeled PES packet transmission to TS-multiplexer to multiplex audio and video serial lined up using a certain percentage. The frame-scheduler belongs to the same frame package that smoothly transfers to the RTP-transmitter according to the RFC3984 RTP header added to the stream sent to the sink. It terminates the program when it receives the last packet. It specifies the receiving end IP. The RTP-transmitter element ensures package delivery to the right end point.

3.2 WANem

The WAN emulator (WANem) is a Linux-based open source simulator that can simulate WAN or LAN. The simulation includes bandwidth, packet loss, delay and jitter. It supports full graphical interface to configure easy set up. You need to adjust the rules to apply the routing table settings. The Packet Limit sets the packet queue size. If the queue exceeds this value the excess will be discarded. This article does not consider the packet queue buffer. Symmetrical network means when elected yes symmetrical network, this study only transmits direction, not the two options effect. Choose BW to set the bandwidth and the paper analog wireless network transmission environment, TD-SCDMA standard reference sets the value at 28000Kbps [6]. The first column is labeled block diagram delay column, Delay time simulation network delay time, Jitter simulation network jitter time units are milliseconds.

Correlation represents the correlation, the delay time after setting a packet with the previous packet time correlation. Distribution represents the delay time distribution function, including normal, Pareto distribution. Since the normal distribution network transmission performance is the most common this paper selects the normal study law [3]. The block diagram is marked as a loss second column. The Loss column indicates the packet loss rate, Correlation indicates that the packet loss probability and a packet of relevance.

WANem implement network packet control substance is called directly under Linux code. Linux 2.2 kernel network stack from the start to achieve support for quality of service, this function module is called Traffic Control (TC). The basic principles of Linux traffic control shown in Figure 4

Fig. 4. the simulation of jitter

A packet is received from an input interface from the input de-multiplexer judge selection. If the received packet is for this host, the packet is sent to upper layer processing; otherwise it is forwarded to the received packet forwarding module. The forwarding module also receives data packets generated by the machine top. The forwarding module looks at the routing table and determines the next hop for the packet. The packages are sorted in order to transfer them to the output interface. General Linux limits card packets sent and cannot restrict the card received packets. Therefore, you can change the order of Linux sent, delay and other parameters to control the flow of data packets. Linux flow control occurs mainly through the output interface package sorting before implementation. Packet sequencing uses the Linux kernel library libnl in a core data structure Qdisc (queueing discipline). Qdisc is located between the network protocol stack and network device interface. Data packets sent by any kernel will enter the Qdisc module queue and are then sent to the network device, to achieve data packet QoS control [7]. WANem TC based on Linux controls the network. It can be based on user settings. Qdisc queue parameters can be modified to control the network latency, packet loss, jitter and other functions.

3.3 VLC

The VLC player is a good open source player that can play MPEG-1, MPEG-2, MPEG-4, DivX, DVD, digital satellite channels and in many systems platform through broadband IPv4, IPv6 network players for online movies. "Open Network Stream” and enter in the menu item “media" drop-down menu and input the URL of the sender to receive the designated network streaming.

3.4 The subject assessment

Twenty testers were invited to participate in subjective video quality assessment. All testers possessed normal vision. Each test involved playing and evaluating viewing material. The testers sat in front of a screen in the middle position. The display distance was from 0.5m. Laboratory lighting was maintained consistent.

Separate subjective evaluation test videos were used for the tests. No reference video was used for comparison. The players watched the original video files to determine which video quality is the best without considering the impact of sources. Each video was viewed for 5 minutes. In the actual test the original playback sequence was scored [8]. The packet loss rate and jitter impact on video quality are shown in Figure 5. In container and foreman video, for example, presented greater dropout packet loss rate, with poor video quality. The bigger the jitter, the more the video quality is degraded. A combination of the two resulted in a rapid decline in video quality. The greater the change in video image content (i.e. multi-scene motion or scene cut picture) the more rapidly the video quality decreased [9].

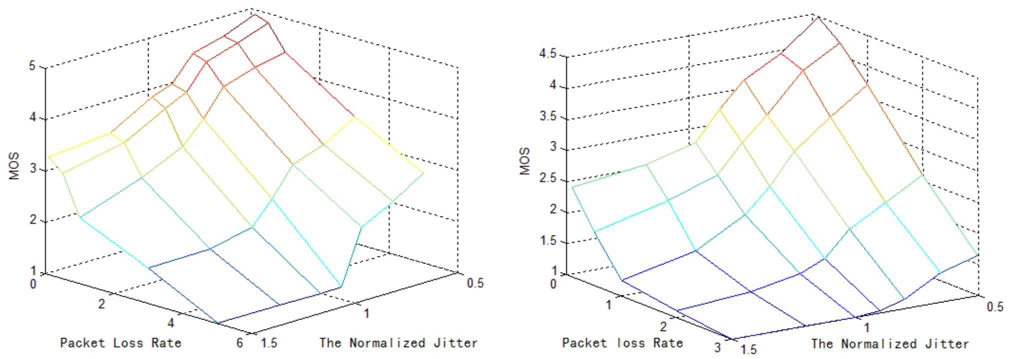

Fig. 5. The Packet loss rate and jitter impact on MOS (Left: container 352*288; Right: foreman 352*288)

4 The VQA Metric based on Neural Network

4.1 The Principle

The BP neural network principle in this paper is shown in Figure 6. This training set input layer has two nodes, with three nodes in the hidden layer and output layer [10]. The input layer of the two nodes, the jitter normalized values for the rate, and other connection weights with the input layer to the hidden layer. The hidden layer has three nodes and there is a node i.e.

x_i in the output layer. It means the jitter and packet loss rate. The $W_{k,i}$ represents the weight between the input layer and the hidden layer. The z_k represents the weight of the node. The ϕ means the activation function of the hidden layer. We used the non-linear function tansig as the hidden layer activation function [11]. The V_k represents the k-th node to node output layer connection weights ; and b represents the output threshold layer; ψ presents the output layer activation function , this takes a linear function purloin; y represents the output value of the output layer.

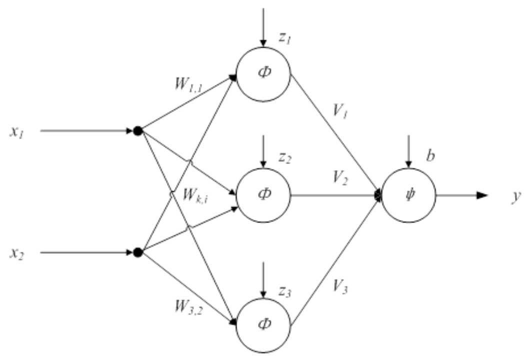

Fig. 6. the principle of two input of BP Neural Networks

(1) the forward signal propagation process

Enter the k-th hidden layer nodes net_k :

$$net_k = \sum_{i=1}^{2} W_{k,i} x_i + z_k \tag{4}$$

k-th hidden layer output node o_k :

$$o_k = \phi(net_k) = \tan sig(\sum_{i=1}^{2} W_{k,i} x_i + z_k) \tag{5}$$

Input and output layer node net :

$$net = \sum_{k=1}^{3} V_k o_k + b = \sum_{k=1}^{3} V_k \tan sig\left(\sum_{i=1}^{2} W_{k,i} x_i + z_k\right) + b \tag{6}$$

The output value of the output layer y :

$$y = \psi\left(net\right) = \psi(\sum_{k=1}^{3} V_k \left[\tan sig\left(\sum_{i=1}^{2} W_{k,i} x_i + z_k\right)\right] + b) = purelin(\sum_{k=1}^{3} V_k \tan sig\left(\sum_{i=1}^{2} W_{k,i} x_i + z_k\right) + b) \tag{7}$$

(2) Error back- propagation process

The training used a total of 52 samples. Let T_p t p is the true value of the output samples (i.e. subjective MOS min)， y_p The model output y is the corresponding value of, Quadratic error criterion function for each sample E_p is：

$$E_p = \frac{1}{2}\left(T_p - y_p\right)^2 \tag{8}$$

The total error criterion included 52 training samples:

$$E = \frac{1}{2} \times \frac{\sum_{p=1}^{52} \left(T_p - y_p\right)^2}{52} \tag{9}$$

BP neural network learning rule using supervised learning, error correction gradient descent algorithm sequentially output layer weights correction ΔV_k , Output layer threshold Δb , Correction amount hidden layer weights $\Delta W_{i,k}$, Correction amount hidden layer threshold Δz_k , as follows:

$$\Delta V_k = -\eta \frac{\partial E}{\partial V_k} \text{ , } \Delta b = -\eta \frac{\partial E}{\partial b} \text{ , } \Delta W_{k,i} = -\eta \frac{\partial E}{\partial W_{k,i}} \text{ , } \Delta z_k = -\eta \frac{\partial E}{\partial z_k} \tag{10}$$

This correction amount of publicity a few adjustments to the output layer weights Case:

$$\Delta V_k = -\eta \frac{\partial E}{\partial V_k} = -\eta \frac{\partial E}{\partial net} \frac{\partial net}{\partial V_k} = -\eta \frac{\partial E}{\partial y} \frac{\partial y}{\partial net} \frac{\partial net}{\partial V_k} \tag{11}$$

Since:

$$\frac{\partial E}{\partial y} = -\frac{\sum_{p=1}^{52} \left(T_p - y_p\right)}{52} \tag{12}$$

$$\frac{\partial net}{\partial V_k} = o_k \text{, } \frac{\partial net}{\partial b} = 1 \text{, } \frac{\partial net_k}{\partial W_{k,i}} = x_i \text{, } \frac{\partial net_k}{\partial z_k} = 1 \tag{13}$$

$$\frac{\partial E}{\partial o_k} = -\sum_{p=1}^{52} \left(T_p - y_p\right) \cdot \psi'\left(net\right) \cdot V_k \tag{14}$$

$$\frac{\partial o_k}{\partial net_k} = \phi'\left(net_k\right) \tag{15}$$

$$\frac{\partial y}{\partial net} = \psi'\left(net\right) \tag{16}$$

Finally, the correction amount obtained formula:

$$\Delta V_k = \eta \sum_{p=1}^{52} \left(T_p - y_p\right) \cdot \psi'\left(net\right) \cdot o_k \tag{17}$$

$$\Delta b = \eta \sum_{p=1}^{52} \left(T_p - y_p\right) \cdot \psi'\left(net\right) \tag{18}$$

$$\Delta W_{k,i} = \eta \sum_{p=1}^{52} \left(T_p - y_p\right) \cdot \psi'\left(net\right) \cdot V_k \cdot \phi'\left(net_k\right) \cdot x_i \tag{19}$$

$$\Delta z_k = \eta \sum_{p=1}^{52} \left(T_p - y_p\right) \cdot \psi'\left(net\right) \cdot V_k \cdot \phi'\left(net_k\right) \tag{20}$$

4.2 The Metric based the packet loss and jitter

When the network channel transmission package, due to system design and other issues, packet loss and delay exist is what often happens [12]. Packet loss can cause loss of information in the video sequence. The delay caused by jitter and packet loss is generated indirectly. A joint study was performed on the channel impact on the video quality with combined loss rate and jitter values. Since WANem can simulate network delay, jitter network packet loss can be simulated, so this section continues to use the 4.2 used in the simulation platform [13]. The loss principle also uses the Linux TC control packet sequencing random packet loss probability Poisson distribution. The metric based on BP neural network is shown in Table 1. The performance of this metric is shown in the left half of Figure 7.

Table 1. The metric based on the packet loss and jitter

The parameters	Value	The parameters	Value
$W_{1,1}$	3.3218	V_1	-0.9013
$W_{1,2}$	0.0058	V_2	1.1218
$W_{2,1}$	-1.2359	V_3	1.8642
$W_{2,2}$	-0.0139	Z_1	-3.1801
$W_{3,1}$	26.3911	Z_2	3.2664
$W_{3,2}$	-4.4362	Z_3	-2.4096
		b	4.8713

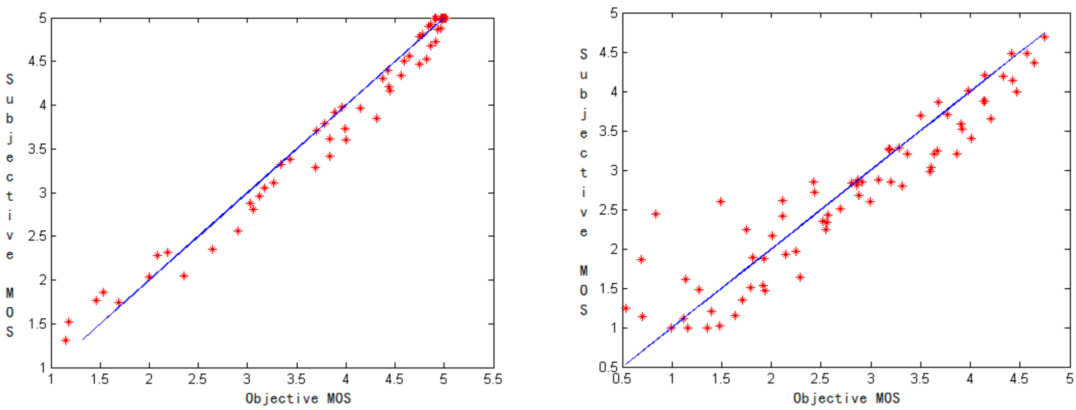

Fig. 7. Left: the performance of BP metric based on packet loss and jitter (mse=0.0595, corr= 0.9673)

Right: the performance of BP metric based on packet loss, jitter and bit rate (mse= 0.1824, corr=0.9304)

4.3 The Model based the packet loss, jitter and bit rate

The channel parameter fraction on the video quality is not linear, it is still selected nonlinear BP neural network modeling principles, but an increase in the input packet loss ratio this parameter is the jitter normalized values for the packet loss rate for the rate [14]. A corresponding increase is shown in the five connection values. The trained neural network parameters are shown in Table 3. The BP neural network performance based on packet loss, jitter and bit rate is shown in the right side of Figure 7.

Table. 2. The test video bit rate

Video Sequence (352*288)	Bit rate (Kb/s)
container	14.61664
highway	19.31616
hall	40.856
coastguard	97.436
foreman	103.8432
mobile	135.9553

Table. 3. The metric based on the packet loss, jitter and the bit rate

Parameters	Value	Parameters	Value
$W_{1,1}$	-7.7113	V_1	-0.3888
$W_{1,2}$	-0.4078	V_2	2.5778
$W_{1,3}$	0.9933	V_3	-0.3983
$W_{2,1}$	-1.2573	Z_1	-3.7634
$W_{2,2}$	-0.3021	Z_2	2.0181
$W_{2,3}$	-0.0091	Z_3	43.0883
$W_{3,1}$	88.8684	b	3.3971
$W_{3,2}$	-20.256		
$W_{3,3}$	-1.3589		

5 Acknowledgement

The work presented in this paper was partially supported by 2015 National Young Natural Science Foundation of China (Grant number 61401381), 2014 National Natural Science Foundation of China (Grant number 61371081), 2012 Natural Science Foundation of Fujian (Grant number 2012J01424) and Foundation by Comba Telecom.

References

[1] C. Keimel et al., "No-reference video quality metric for HDTV based on H. 264/AVC bitstream features," in *Proceedings of 18th IEEE International Conference on Image Processing (ICIP)*, pp. 3325-3328, 2011.

[2] K. Brunnstrom, D. Hands, "VQEG validation and ITU standardization of objective perceptual video quality metrices," *IEEE Signal Processing Magazine*, Vol. 26, No. 3, pp. 96-101, 2009.

[3] N. Staelens et al., "Standardized toolchain and model development for video quality assessment—The mission of the joint effort group in VQEG," in *Proceedings of 2011 Third International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 61-66, 2011.

[4] H. K. Kalitay, M. K. Nambiarz, "Designing WANem: A wide area network emulator tool," in *Proceedings of 2011 Third International Conference on Communication Systems and Networks (COMSNETS)*, pp. 1-4, 2011.

[5] M. H Pinson, S. Wolf, "A new standardized method for objectively," *IEEE Transactions on Broadcasting*, Vol. 50, No. 3, pp. 312-322, 2004.

[6] C. Feng, "Image quality assessment-oriented frame capture of video phone over 3G system," in *Proceedings of 2010 International Conference on Anti-Counterfeiting Security and Identification in Communication (ASID)*, pp. 359-362, 2010.

[7] X.-B. Gao, *Quality Assessment Methods for Visual Information*, Xi'an, XiDian University Publishing House, 2010.

[8] ITU-T, *J.247 Objective Perceptual Multimedia Video Quality Measurement in The Presence of a Full Reference*, 2008

[9] C. Keimel et al., "Video is a cube: Multidimensional analysis and video quality metrics," *IEEE Signal Processing Magazine*, Vol. 28, No. 6, pp. 41-49, 2011.

[10] Z. Wang, A.C. Bovik, "Mean squared error: love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, Vol. 26, No. 1, pp. 98-117, 2009.

[11] M. H. Kayvanrad et al., "Reduced reference watermark-based image transmission quality metric," in *Proceedings of 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP 2008)*, pp. 526-531, 2008.

[12] C. Keimel, T. Oelbaum, K. Diepold, "No-reference video quality evaluation for high-definition video," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1145-1148, 2009.

[13] C. Feng, L.F. Huang, "A method about video quality assessment in video phone service over 3G network," in *Proceedings of 2014 International Conference on Mechatronics Engineering and Electrical Engineering*, pp. 43-47, 2014.

[14] Z. Shi, P. Chen, C. Feng, L. Huang, W. Xu, "Research on quality assessment metric based on H.264/AVC bitstream," in *Proceedings of the International Conference on Anti-Counterfeiting, Security and Identification*, pp. 1-5, 2012.