A Software Platform Design for Objective Video QoE Assessment

Yitong Liu, Hao Liu, Yuchen Li, Yun Shen, Jianwei Wu, Dacheng Yang
Beijing University of Posts and Telecommunications, Beijing, China
Email: liuyitong@bupt.edu.cn

***Abstract*—Along with the rapid development of 3G and 4G technologies, mobile video services have gained its popularity among users around the world. Consequently, Content Providers (CPs), Service Providers (SPs), and especially, the operators are paying increasing attentions to the quality of experience (QoE) of the video services which could be easily affected by the quality of network. In this paper, a novel real-time objective video QoE assessment method is proposed and a software assessment system is built to test the video service quality in the real network. Firstly, in the test terminals, the QoE measurement of the entire video services is conducted by collecting all of the customers' experience in full-reference method, and then the QoE scores are evaluated through an accurate mathematic model. Secondly, the artifact of compression caused by video encoding should also be taken into account. Model in this part adopts no-reference method in consideration of the varied screen sizes in different terminals. What's more, the platform also evaluates the error of network in the part of video transmission by associating no-reference PSNR with network delay, jitter, and packet loss ratio. The results of Mean Opinion Score (MOS) tests show that the proposed models estimate QoE with high quality estimation accuracy respectively. We develop a software toolkit using the test methodologies above, which can help the operators to make measurements for its network. This software toolkit is useful as a QoE monitoring tool on video streaming services and can be deployed on real network conveniently.**

***Index Terms*—QoE, video service, objective assessment, software toolkit**

I. Introduction

The advances in video encoder technologies and broad IP networks lead to the popularity of video streaming services. Furthermore, with the development of 3G/4G technologies, the number of customers attracted by mobile video streaming services is growing rapidly. In recent years, mobile network operators in China have launched a variety of video services, including VOD, video telephone, and etc. Besides, the increase in the amount of Internet services is making the video services much more bustling.

As shown in Fig.1 [1], Web service, video, IPTV and P2P contribute to the major part of the total Internet traffic since 2011. Among these services, Video, IPTV and P2P are relative to video distribution.

In China, a statistic analysis report from CNNIC shows that the mobile terminals became the NO.1 internet access device in China by the year of 2013 [2] , used by 75% of users.

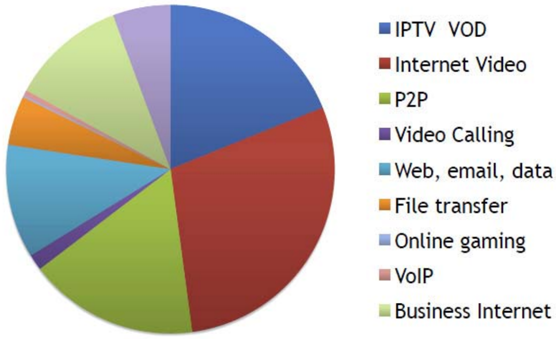


Figure 1. 2011 Internet Traffic distribution

A survey from iReseach revealed that the video traffic is turning to mobile market. They took 3 most popular videos appeared in one year as examples. The mobile share of traffic doubles nearly every 6 months [3], shown in Fig.2.

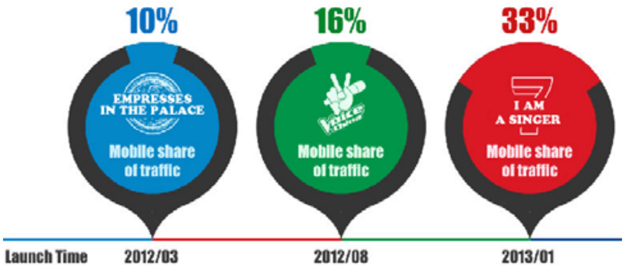


Figure 2. The mobile traffic of 3 most popular videos in 2012

The quality of video service is a reflection to the quality of network. The services' perceived quality draws the most attention form operators since it is closely related to customers' personal feelings. The better the quality is, the more customers could be attracted. Nevertheless, traditional indicators to evaluate the performance of services, in terms of Quality of Service (QoS) for the network, are not accurate enough to reflect the customers' degree of satisfaction.

Then came the question. In our paper, we discuss how to evaluate the customers' feeling, experience, or

doi:10.4304/jsw.9.11.2955-2965

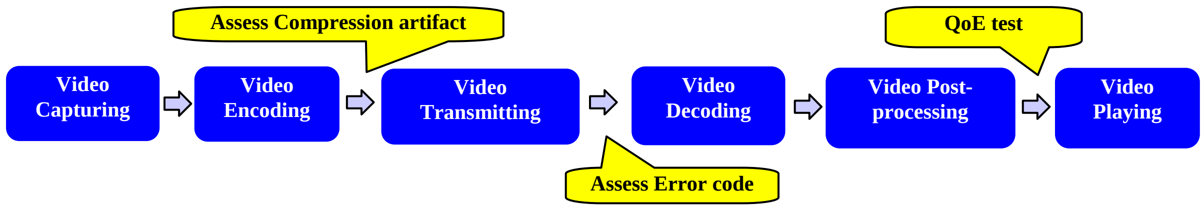


Figure 3. The test purpose of proposed QoE-based evaluation toolkit in a video service

reception on the video service objectively, and how to use this kind of evaluation result as the baseline for the provision of network quality for this kind of service.

In[4],the Quality of Experience (QoE) is proposed. In recent years, many proposals are made in order to evaluate, measure, and improve QoE of the video services.

Nicolas Staelens proposed a novel subjective quality assessment methodology based on full-length movies [5]. Their subjects took DVD together with a questionnaire enclosed in a sealed envelope home to watch it in real-life environments. Ozgur Oyman reviewed the recently standardized QoE metrics and reporting framework in 3GPP and presented an end-to-end QoE evaluation study conducted over 3GPP LTE networks [6].Ricky K.P. Mok investigated the relationship between network QoS, application QoS, and QoE and then proposed their QoE measurement [7]. They evaluated QoE of Flash video perceived by users and quantified how the QoE is influenced by the application QoS.Hyun Jong Kim developed a QoS/QoE correlation model, which could evaluate the QoE using QoS parameters in offered network environment [8].Karan Mitra proposed a novel approach [9] to estimate and predict QoE in Heterogeneous Access Networks (HAN). This system is based on Hidden Markov Models (HMM) and Multi-homed Mobility Management Protocol (M-MIP), which improved the accuracy of QoE estimation in many network conditions. All of these researches about the QoE of video services are quite constructive. However, the limits of QoE assessment can be not ignored. Above all, the subjective test consumes lots of manpower and material resources, which is unacceptable for both of the operators and the customers. Therefore, the objective and quantitative QoE test is introduced in this paper.[10][11]

Besides, QoE assessment is posteriori so that the evaluation result could only show the QoE level of the whole service no matter how many factors and parts there are. In consequence, it is very difficult to monitor what caused the QoE down.

So our video QoE assessment system is divided into 3 platforms to describe the performance of the entire service, impairment from the video encoding, and the impairment in delivering.

The paper is organized as follows. In section II, the methodology of QoE-based assessment is introduced. The structure and deployment of our software toolkit is described. Section III describes the method and the software platform we proposed to assess the quality of entire video service. A reference sequence is chosen, and the service is recorded while playing. Full-reference method is adopted. KPIs in the video playing process, which are directly felt by customers, are collected to map to QoE score. In Section IV, we introduce the software platform which evaluates the compression artifact. Because there are various encoded videos in different sizes and the source video cannot be acquired in most of time, the no-reference test method is adopted. Through this platform, the behaviors of different providers and encoders, as well as the quality of videos with different contents and sizes are assessed and presented for the operators to monitor and control. Then in section V, assessing error code software is introduced to test and evaluate the impacts of network environment on the video quality. The structure and modules are introduced in this part. Finally, paper is ended with conclusion and some future works in section VI.

II. QoE-Based Evaluation Software Platform Of Video Quality

Fig.3 shows a complete processing procedure for a video service, including video capturing, video encoding, video transmitting, video decoding, video post-processing, and video playing. Obviously, the QoE that worked out from the terminal should be the final result of the entire video service. And there are only 2 processes, video encoding and video transmission, that could be monitored and controlled by the operators. It is significant for the operators to find out which part causes the degradation of service quality, and then to quantize and compare the QoE loss.

Video encoding is a kind of loss compression coding due to the limitation of storage and bandwidth. Thus, video encoding is one of the main sections causing quality degradation. Take H.264 encoding for example, the quantization of conversion coefficients, which is controlled by a quantization parameter (QP), degrades the image quality via increasing the QP value.[12] This degradation can be reflected by blocking artifact, blur, and Peak Signal to Noise Ratio (PSNR).

The quality of encoded video will degrade again while transmitted in the network. As a result of the network delay, jitter, and packet loss, frame skipping and frame frozen will appear in the received video services.

In this paper, we build a video QoE evaluation platform, which could provide QoE information to the operators, in order to help them in optimizing their services and to define the responsibilities clearly. The designed QoE evaluation framework functions in three parts.

The first part is placed in the test terminals. The target of this part of assessment is to give the QoE result of the whole service. This platform will be described in detail in the section III. The online video service is involved and the full-reference method is adopted. To begin with, we establish the standard original videos and upload them to the web servers. In the terminals, the service is requested and at the same time the course of the service is recorded. Three groups of KPIs is extracted from the comparison between the recorded video and the original video, including KPIs in the connecting period, KPIs of the image and quality of the voice. We use an accurate mapping model to obtain the QoE score from network response delay, KPIs of the images and the sound QoE scores from the Perceptual Evaluation of Speech Quality (PESQ) [13], an international standard to evaluate the QoE on the voice service and also a prevailing algorithm inbuilt by many network optimizing instruments. The QoE score is calculated out and presented in the assessment windows and the detail parameters are illustrated as graphs and tables to compare with each other. This part of this assessment system is named the online video QoE assessment and this assessment can be performed in anywhere we aim to acquire the quality of the network and service.

In the first part, the operators have obtained the value of the service quality. However, it is difficult to judge whether the damage of the service comes from the quality of the network provided by Themself. The second part is to assess compression artifact of video services. This platform is installed in the video center of the operators, and is implemented to test the encoded videos in various contents and sizes uploaded by the CPs. It is of great importance to distinguish the quality degradation owing to the coding compression from the loss because of the bandwidth limitation. The video content can also be an important factor for video coding.[14] Generally, videos in the same content would be encoded into different sizes and different qualities because of the demand of various mobile terminals. The performance of the different CPs and different commercial video encoders are monitored in 3 video types, the rapid movement, the slow movement and the colorful scenario.

The compression artifact assessment is performed after the CPs upload all of the encoded video chips. The encoded videos are in different sizes and the source videos cannot be acquired in most of the time. Therefore, we use the no-reference method to estimate the KPIs of the image, including PSNR, blur, block and motion. It is the most significant and difficult issue in the assessment. Our toolkit records these KPIs and the final QoE results in case of subsequent statistic analysis.

The third part mainly focuses on assessing error code. This part is distributed to the net nodes of individual province companies through Content Delivery Network (CDN) as the video of demand for customers. This part of assessment is to test the quality of transmission. The operators can monitor the behaviors of the individual province companies. We extracts the KPIs of the networks and the qualities of the video service. Packet loss, delay, jitter and so on are associated with the no-reference PSNR and other parameters. The assessment toolkit shows and records the real-time parameters in every transmission node.

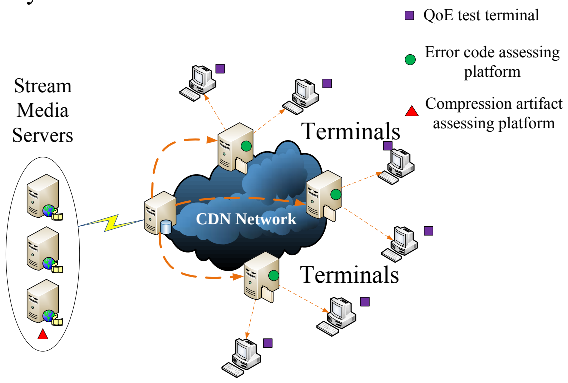


Figure 4. The network framework of the QoE evaluation platform

As previously mentioned, Fig.4 illustrates the network framework of the QoE-based evaluation platform. The online QoE test terminal is used to obtain the QoE performance of the video service in any point. The compression artifact assessment is deployed to administrate the encoded videos required by a variety of screens. Besides, the error code assessment is monitor of the transmission quality in the CDN and even in the wireless connections.

III. Online Video QoE Evaluation Platform In Terminals

A. *Procedures of the Video QoE Test*

We propose an online test module structure in this section. The progressive streaming scenario is considered. The full-reference method, which compares the degraded video against the original video to get results, is adopted. The standard source video samples are divided into 3 groups of different durations: 30s, 1mins and 3mins. Besides, different contents are involved. The procedures of the assessment are as follows:

1. Build a network service server and upload all of the original standard videos.

2. Choose a video sequence in the test terminals to test.

3. Request a video service, and record the playing process automatically. The KPIs during the service request, such as *successful access ratio* and *response delay*, are counted.

After recording the test videos online,the degraded sequences are obtained.

Figure 5. The comparison between the source videos and the recorded videos

4. Build a new project to test this video service. Input the original video and the degraded video for comparison and start the QoE testing automatically. The QoE score of the whole service is calculated. This score is based on image quality, audio quality and service access delay.

5. Detailed parameters and KPIs can also be viewed. Besides, the original video and replay of the recorded video service can be watched, as shown in Fig.5.

Some Graphical User Interfaces (GUIs) of this part of QoE test platform are demonstrated in Fig.6. The main interface gives the assessment progress and the basic parameters of the video in Fig.6.

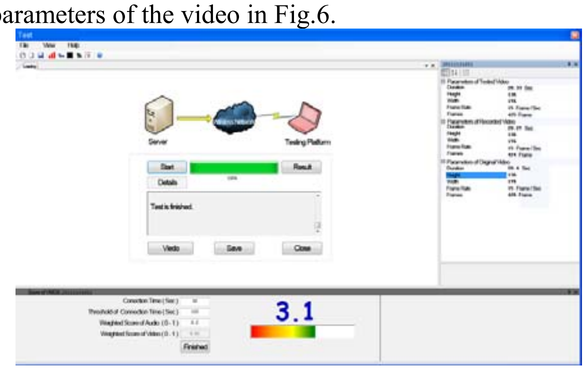


Figure 6. The main interface of the online video QoE evaluation platform

The final QoE score is presented in the end of the assessment. After the test finished, the detail KPIs, such as delay, PSNR, frame skipping and frame frozen, can be analyzed in graphs and can be output to a file named after the test time. The analysis window is demonstrated in Fig.7.

Figure 7. The detailed KPIs analysis interface

B. KPIs Extraction Fliter in Video QoE Test

We defined several KPIs that influence the customers' experience. These KPIs is obtained in the toolkit and presented for the operators.

1.KPIs during the period of connecting server, including *successful access ratio* and *service access delay*.

Concretely, *service access delay* is defined as T :

$$T = t_1 + t_2 \tag{1}$$

Where t_1 is *network access time* and the t_2 is *response delay*, described in the Fig.8. The *network access time* equals to the time length from the time when customer demands the video to the time when video starts to buffer. It is used to estimate the network response time, related to the net environment. And *response delay*, depending on the predetermined strategy, records the time lag from the time that video starts to buffer to the time that video starts to play. Therefore, *service access delay* equals to the waiting time after the costumers request the service.

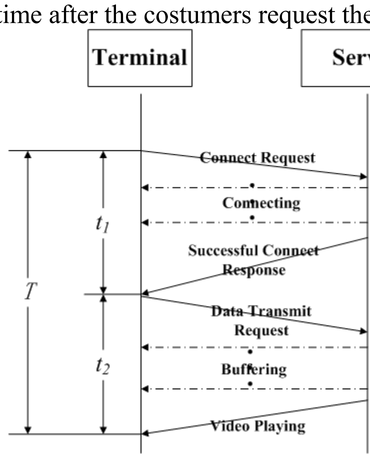


Figure 8. The definitions of the service access delay, the network access time and the response delay

2.KPIs of video images influence the fluency and resolution, including:

--*Image activity ratio*

--*Activity in time domain* which means image variation degree in time domain

--*Spatial complexity* which depends on video types

--*Luminance*

--*PSNR*

--*Frame skipping* caused by frame loss

--*Frame frozen* caused by frame repetition

--*Block*

--*Blur*

--*Delay distribution while playing*

PSNR is the most frequently used indicator for video quality. It can be calculated from the luminance values of the source signal *p(x, y)* and the degraded signal *q(x, y)* as follows.

$$MSE = \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (q(x,y) - p(x,y))^2 \tag{2}$$

$$PSNR = 10\lg\left(\frac{(2^Q - 1)^2}{MSE}\right) \tag{3}$$

MES is the mean squared error and *Q* is the bit of an intensity value. *X,Y* is the frame width and height separately. The classic *PSNR* algorithm provides three types of values: $PSNR_Y$, $PSNR_{Cr}$ *and* $PSNR_{Cb}$.

The blur parameter is calculated based on an extreme luminance value in a frame. We use the zero-crossing rate to extract blur. The image definition reflects the degree of changes in image details. Namely, the higher the definition is, the better the image presents. In consequence, gray scale can be more sensitive to the changes in location and variations in image details could also be high, resulting in a good degree of recognition

[15][16]. The value range of blur is (0,1], where higher value means higher ambiguity.

Generally speaking, blocking artifacts are caused by low bit rate encoding in the flat area of image and border of moving objects. The key of Blockiness algorithm is to extract block border areas and non-block border areas. The block border area consists of pixels that are adjacent to a block border or pixels that include a block border, which can be detected by canny-edge detector. And non-block border area consists of the rest pixels. The block is calculated by reference to the computing method proposed by [17].

Delay parameters are divided into *maximum delay, minimum delay* and *average delay*, which come from the full reference algorithm. And these detailed parameters are listed on the interface separately.

Some of the parameters mentioned above work as intermediate variables to calculate parameters afterwards. Others are used to build the eventual QoE model.

3.Quality of sound in the video

For the sound in the video,we choose the PESQ test, which is introduced in the ITU P.862. PESQ is a most prevailing voice service QoE assessment method and has been inbuilt by many network optimizing instruments. However, our former research have drawn a conclusion that it is not convinced enough to use PESQ to assess Chinese voice service directly [18]. So the English voice is preferred in the test video sequences.

C. Structure of Image Quality Assessment

As it shows in Fig.9, the image quality assessment module is composed of pre-process and analysis module, parameter extraction module and quality estimation module. First, the pre-process and analysis module takes source video signal and degraded video signal as inputs, and extract the Region of Interest (ROI) and some other information from spatial domain and temporal domain. Then, the parameter-extraction module derives delay parameters, PSNR, block and blur parameters. Finally, the quality-estimation module estimates video quality using these parameters.

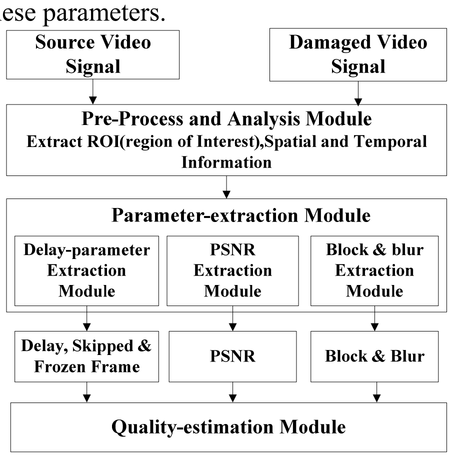


Figure 9. Full-reference QoE online test module structure

D. QoE Mapping Modeling

The QoE assessment model is established to map *PSNR, delay, frameskipped, framefrozen, blockiness, blur* and frame rate $Rate_{fps}$ to QoE score. The modeling method and result have been introduced in our previous research paper [19].

$$Score_{PSNR} = a * PSNR_Y + b * PSNR_{Cb} + c * PSNR_{Cb} \quad (4)$$

$$Score_{frame} = \ln(framefrozen) + d * frameskipped \quad (5)$$

Here， $PSNR, framefrozen, frameskipped \in [1,5]$ ·

$$Score_{Block} = (e * \ln(Block + 1) + f) * (Score_{frame} + Score_{psnr}) + g \quad (6)$$

$$Score_{Blur} = \begin{cases} Score_{Block} + 10 / Blur - h & 10 / Blur <= 5 \\ (\frac{i}{Blur} + j) * Score_{Block} & 10 / Blur > 5 \end{cases} \quad (7)$$

$$S' = Rate_{fps} * (k * Score_{Blur} + l) \quad (8)$$

$$Rate_{fps} = \min(1, m * fps^n + o) \quad (9)$$

The final score is

$$S_{QoE} = \begin{cases} 1 & \text{if } S' < 1 \\ 5 & \text{else if } S' > 5 \\ S' & \text{else} \end{cases} \quad (10)$$

Where *a, b, c, d, e, f, g, h, j, k, l, m,* and *o* are coefficients. These coefficients are optimized using least-square method to minimize the difference between subjective video quality and estimated video quality.

E. Subjective Test and Data Acquisiton

According to [20], 30 seconds standard videos are accepted in the subjective test. In this section, 29 sequences with different damages are produced as counterparts. No less than 10 viewers vote every counterpart.

Before beginning the subjective test, every viewer accepts a simple training using 5 examples. And each viewer should rate the counterpart using the integral MOS scale of 1, very bad, to 5, excellent.

After cancelling the invalid test samples, according to [21], 315 votes are accepted and the coefficients in (4)-(10) are determined, listed in Table I.

TABLE I.
COEFFICIENTS OF QOE MODEL FOR ONLINE VIDEO TEST

Coeff.	Value	Coeff.	Value	Coeff.	Value
a	0.3	***f***	1	***k***	0.9654
b	0.1	***g***	1.3	***l***	-0.0421
c	0.1	***h***	5	***m***	-10.12
d	0.2	***i***	1.348	***n***	-0.03116
e	-0.279	***j***	-0.069	***o***	10.45

The Pearson Correlation Coefficient of above QoE model is calculated as high as 0.925, as shown in Fig 10.

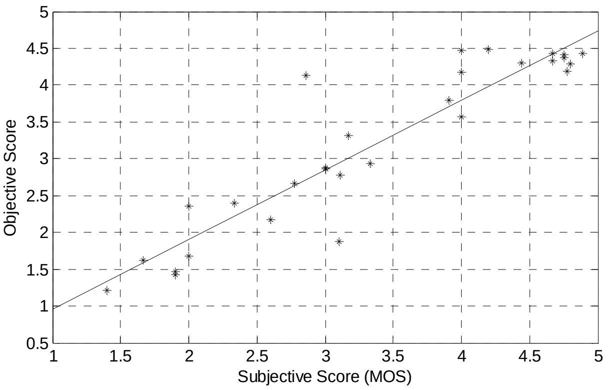


Figure 10. Scatter diagram between MOS and QoE score

IV. COMPRESSION ARTIFACT ASSESSING PLATFORM

The mobile Internet users has been dramatically increasing and the usage scenarios are diverse. Internet access is common to various mobile devices, such as smart phones, tablet PCs, laptops, TVs and so on. All kinds of video content are provided by a sight of CPs, whereas the videos are aimed to be applied to different terminals. Therefore, even videos of the same content should have different sizes and qualities.

As known, different encoding types lead to different compression artifacts, which cause varying degrees damage of quality. Even the same encoding type can make different degrees of artifacts by using different encoders.

According to the different content, we divide videos into 3 groups: slight movement, rapid movement and colorful scenario. All of the 3 video content types are considered. We choose 6 original video samples with 1.5Mbps and D1 (720*480) resolution. Rapid movement group contains *Car Racing* and *Tennis*. Slight movement group contains *News* and *Football*. The Colorful group has *Movie* and *Natural scenario*.

5 CPs are invited to participate in the assessment. 5 different commercial encoders are involved to produce compression artifact counterparts. In each encoder, every original video is encoded into 3 versions with different bitrates and resolutions: D1 for high bitrate, CIF for middle bitrate and QCIF for low bitrate. Finally, 90 degraded samples are collected.

Same as that in section III, no less than 10 viewers are recruited to rate each counterpart. 2700 votes are accepted in the end, half of which is used to establish the QoE model and some other is used to verify the performance of the model. In addition, the invalid votes are filtered out as specification in [21].

The objective of assessing compression artifact is to get optimal encoders for different type of videos. In the video encoding section, the full-reference method is infeasible. It is difficult to compare the degraded video and the original video because the encoded video chips have different sizes from the original samples or even the original video samples may not be acquired. Therefore, the proposed model adopts no-reference method to assess compression artifact by using h.264 encoded video instead. The model structure can be seen in Fig.11.

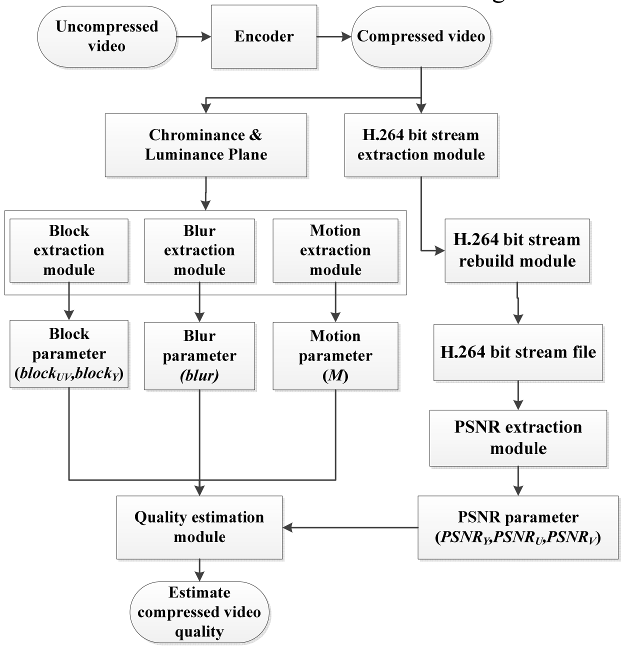


Figure 11. Structure of assess compression artifact platform

The complete algorithm is composed of two modules, Chrominance and Luminance Plane and H.264 Bit Stream Rebuilt Module. There are totally 4 parameters that can be extracted by the model.

In the first part, the signal is sent to Block Module, Blur Module and Motion Module respectively. In these three modules, video is processed to obtain its block, blur and motion value. In our platform, the algorithm of calculating these three parameters are a no-reference method, which means that there is no need to acquire the source video. Many studies provided solution of this aspect, here we adopted to the algorithm from [22].

In the second pare, the PSNR can only be extracted from bit stream of the whole video signal. Usually, the traditional PSNR algorithm is a full-reference method which needs the source signal in (2) and (3). However, it's hard to calculate PSNR when lacking of source signal, i.e., uncompressed video. Thus, we adopt a no-reference method to estimate the value of PSNR, according to [23]. This algorithm uses the coded transform coefficients to estimate the PSNR in a statistical manner. And the Pearson Correlation Coefficient between full-reference PSNR and this no-reference PSNR is as high as 0.96. By the simulation results in [23][24], this algorithm is confirmed to suit different content types and resolutions.

Based on the above 4 KPIs, in our previous research paper [24], we have given the QoE evaluation formula and verified a good performance of the model in estimation the subjective experience. In the Quality Estimation Module of the software platform, the QoE model is adopted to acquire the QoE score.

For the slow movement and rapid movement videos, the QoE evaluation formula is:

$$Score_{PSNR} = a_1 * PSNR_Y + a_2 * PSNR_U + a_3 * PSNR_V + a_4 \quad (11)$$

$$Score_{Block} = (b_1 * block_Y + b_2 * block_{UV} + b_3) * Score_{PSNR} \quad (12)$$

$$S_{QoE} = \begin{cases} c_1 * Score_{Block} + c_2 * Blur + c_3 & 176 \le W < 352 \\ c_4 * Score_{Block} + c_5 * Blur - c_6 & 352 \le W < 720 \\ c_7 * Score_{Block} + c_8 * Blur + c_9 & W \ge 720 \end{cases} \quad (13)$$

Where $PSNR_Y$, $PSNR_U$, and $PSNR_V$ are the no-reference PSNR values on chrominance and luminance plane. W is the width in pixels.

The formula reflects the map relationship between PSNR value and QoE score. The coefficients *a1-c9* depend on the specific content type of video, which are shown in TABLE II. The Pearson Correlation Coefficient is approaching to 0.97 for the Rapid movement video and 0.84 for the slow movement.

TABLE II.
COEFFICIENTS OF QOE MODEL FOR RAPID&SLOW MOVEMENT VIDEO

Coeff.	Rapid movement	Slow Movement	Coeff.	Rapid movement	Slow movement
a_1	0.9104	1.1590	c_2	0.1394	-0.7480
a_2	-2.2971	-3.0432	c_3	0.3378	0.5694
a_3	1.6015	1.9305	c_4	0.2137	-9.9208
a_4	-0.9782	2.9416	c_5	0.322	8.8242
b_1	0.7421	0.8251	c_6	-0.1591	0.2134
b_2	-0.5819	-0.0822	c_7	0.316	-1.7434
b_3	0.8908	1.2075	c_8	0.5762	1.6814
c_1	0.0688	1.1194	c_9	0.0954	0.6651

For the colorful scenario, another QoE evaluation model is more suitable:

$$Score_{PSNR} = a_1 * PSNR_Y + a_2 * PSNR_U + a_3 * PSNR_V + a_4 \quad (14)$$

$$Score_{Block} = b_1 \cdot Score_{PSNR} + b_2 \cdot \ln(block_Y + 1) + b_3 \cdot \ln(block_{UV} + 1) + b_4 \quad (15)$$

$$S_{QoE} = c_1 \cdot Score_{block} + c_2 \cdot blur + c_3 \quad (16)$$

With the huge subjective test results, the coefficients in (14) (15) (16) are listed in TABLE III. The Pearson Correlation Coefficient for the colorful scenario is as high as 0.95.

The compression artifact assessing platform can be located at the video center servers in which the CPs upload the encoded videos. This software tool tests the encoded video quality provided by content providers, and furthermore, evaluates the behaviors of different CPs and different encoders, as well as the quality of videos with different contents and different sizes.

TABLE III.
COEFFICIENTS OF QOE MODEL FOR COLORFUL SCENARIO

Coeff.	Resolution(pixel*pixel)		
	QCIF (176x144)	CIF (352x288)	D1 (720x480)
b_1	0.1943	-0.4637	-0.1549
b_2	-7.842	39.1546	2.1186
b_3	7.5542	-38.5645	-1.7643
b_4	1.315	7.7655	4.7507
a_1	0.6903		
a_2	0.0041		
a_3	-0.6606		
a_4	3.6908		
c1			1.0218
c2			0.0214
c3			-0.3951

Fig.12 shows the GUI of compression artifact assessing toolkit. After CPs uploaded the encoded videos to the center server of the operators, the software tool loads all the video sequences automatically and tests them successively. The user chooses the file folder and the videos are loaded in the left bar. Through the QoE test, the quality score, PSNR, Blockiness, Blur, Motion, Luminance, Skipped Frame, Frozen Parameter and other KPIs are calculated out in the right side of the window. The video information is listed below the QoE results. The results can be compared in different KPIs, CPs, contents, or other dimensions. All the results are stored for further analysis and research.

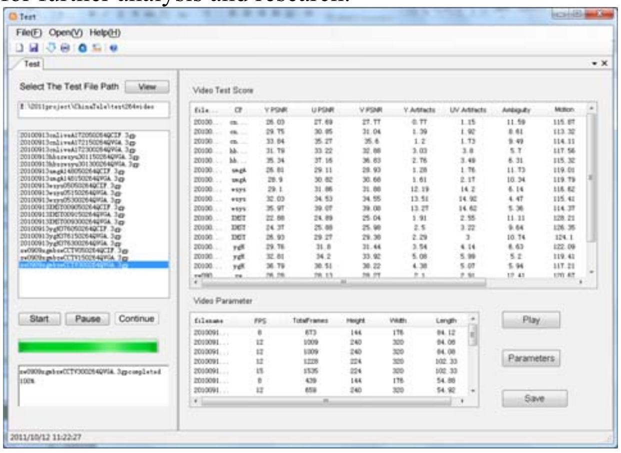
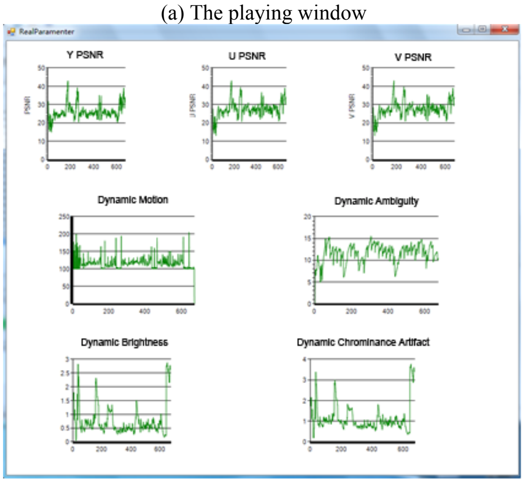


Figure 12. Compression artifact assessing platform main GUI—the QoE and KPIs results

In the Fig.13 (a), different versions of the test videos can also be played on this platform. The users could decide which video to play on the basis of their test results. In addition, the results can be further analyzed on the platform. Real time parameters are provided for detailed observation. For example, brightness, chrominance and other parameters of every frame are calculated in real time and shown in line chart, which are demonstrated in Fig.13 (b).

(a) The playing window



(b) Real-time results analysis GUI

Figure 13. Assess compression artifact GUIs—the playing window and real-time parameters analysis in graphs

V. Error Code Assessing Platform

When the compressed video signal transmits through the network to the receiver following the Real Time Streaming Protocol (RTSP), different kinds of error code would appear under the influence of real network defects, such as delay, jitter, and packet loss. The target of assessing error code is to analyze the influence to video quality caused by network error code.

A. Procedure of the Error Code Assessment

The structures of error code assessing platform are as follows:

1. A RTSP video service connecting should be built in the first place. After opening the platform, the customer should input the server IP, home IP and URL address, and choose the recording equipment in the computer.

2. Press the Play button, and begin to view the RTSP video on demand. Meanwhile, the platform records the screencast of the video automatically. The blur, blockness, movement and PSNR are calculated in the recorded video.

3. Capture all the RTP packets to rebuild a bit stream file to extract the PSNR parameter when the video are playing. By analyzing the RTP packets, the network KPIs are calculated, such as delay, jitter and packet loss ratio.

4. Evaluate the video quality based on the KPIs from the RTP packet capture and the KPIs of the images. The final QoE score and the KPIs are presented in the GUIs.

5. After the QoE test, the KPIs and real time parameters can be reviewed in another windows.

B. KPIs in the Error Code assessment

KPIs that influence the customers' experience are chosen. Both the average result and the real-time parameters are presented on the toolkit.

The network level KPIs: *delay, jitter* and packet loss ratio. The definition of this KPIs is specified in [25].

The video information: *video size, duration of the video, frame rate, file name, provider's name* and so on.

The KPIs of the image quality: *luminance, chrominance, PSNR, the frame frozen, blur, blockness* and so on.

C. Modules of the KPI extractions

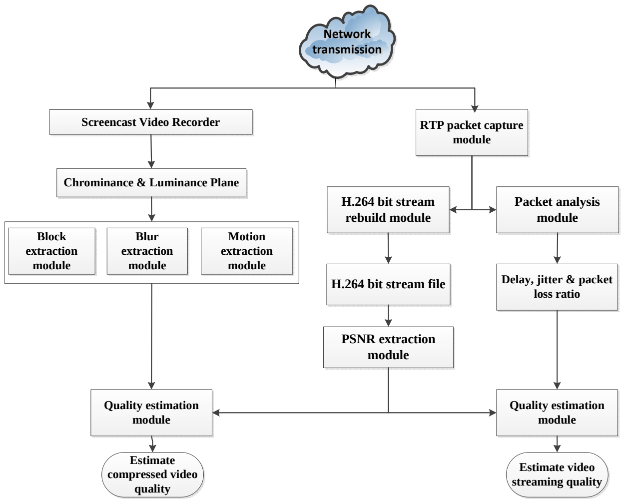


Figure 14. Structure of assessing error code

As Fig.14 shows, the assessment is deployed in the network transmission. This module can also be divided into two sections. One is RTP packet capture module. In this part, the H.264 bit stream file is rebuilt from the RTP packets on the one hand. The model calculating no-reference PSNR is the same with the one in section IV. On the other hand, the packets are analyzed to calculate the delay, jitter and packet loss ratio[25].

Another sections bases on the recorded videos. The no-reference method is also accepted. It is important to acquire the frozen frames caused by the delay or packet loss during the transmission. Besides, blur, blockness and movement are obtained with the same algorithms as former sections.

D. GUIs in the Error Code Assessment

In the Fig.15, Fig.16 and Fig.17, the GUIs of the error code assessing platform are revealed and the assessment process is specified.

Fig. 15 gives the main interface of the error code assessing platform. To start with, users should fill the server's IP and the maximum waiting time on the left blanks followed by choosing voice record devices on the right. Demanded videos will be played in the bottom while its assessment results is shown in the lower right region.

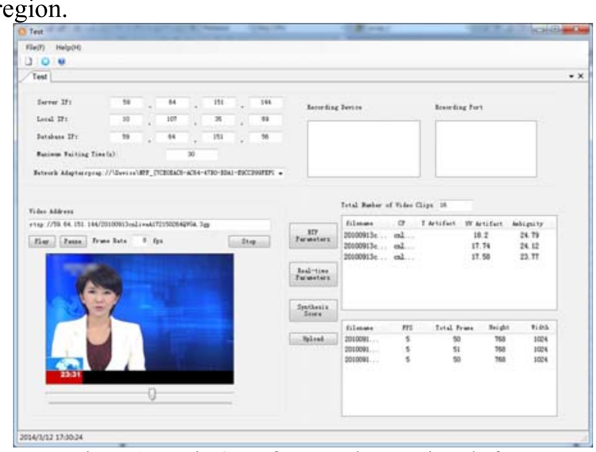
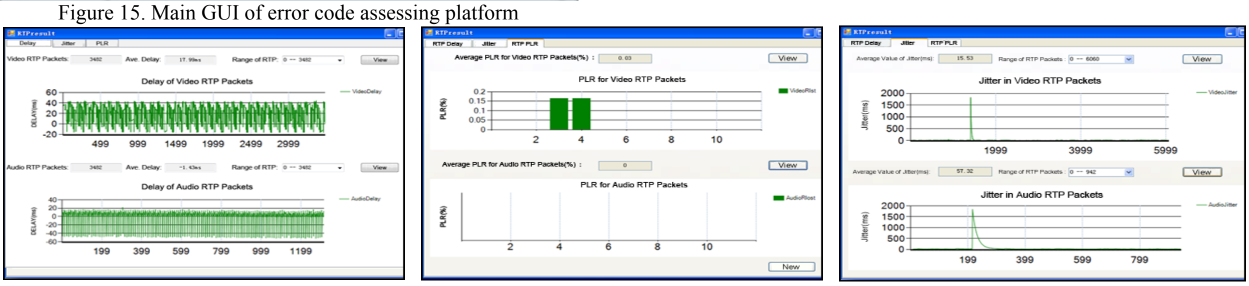


Figure 15. Main GUI of error code assessing platform



(a) Network delay (b) Packet loss ratio (c) Jitter

Figure 16. The real-time KPI analysis GUIs in the error code assess software platform

The users can also review the real-time parameters after pressing the “Real-time parameters” button as shown in Fig.16. The RTP packet arrived delay and the jitter can be recorded in chronological order.

The QoE evaluation result and information of current video service is given in the QoE score windows, illustrated in Fig.17.

Figure 17. QoE score and video information GUI

VI. Conclusion

In this paper, an objective QoE assessment methodology for the video service is proposed. This evaluation can be used as the baseline for provision of network quality for video streaming services.

The KPIs that can influence the feelings of the customers are extracted, which are based on our former and others’ constructive research achievements. The QoE mapping models are built and huge subjective tests are implemented. The result demonstrates our objective QoE assessment models are accurate.

Based on this assessment methodology, a software platform is built for the Operators to evaluate the quality of video streaming service and the performance of 3G/4G network. The software platform has 3 parts. The first part, video QoE Evaluation platform in terminals, can assess the QoE of the entire video service. Moreover, the second part, compression artifact assessing platform can quantify the QoE loss in the video encoding process. This part of the software platform can be deployed in the video service’s center server to test the behavior of the CPs. Through evaluating the quality of videos with different content and different sizes, the performances of different video encoders are ascertained. It is helpful for the operators to distinguish the QoE loss in the video encoding process from the loss in the transmission process. The third part is error code assessing platform, which monitors the network parameters and map them to the QoE score. The output of this part is the QoE loss between any two points in the network. The error code assessing platform is deployed in the CDN.

The GUIs of these 3 software platforms are shown. The assess procedure is automatic and the real-time parameters are recorded and can be remote reviewed.

The proposed video service assessment methodology solves the problem of the objective and automated QoE testing. And the QoE-based assessment software is useful as a QoE monitoring tool on video streaming services and can be flexibly deployed on real network.

REFERENCES

[1] Hui Zhang. "Internet Video: The 2011 Perspective," unpublished.
[2] CNNIC. 32nd Statistical Report on Internet Development in China. July, 2013.
[3] iResearch Inc. http://www.iresearch.com.cn/, 2013.3.
[4] GB923 TMF.v3.0-2004[J]. *Wireless service measurements handbook*, 2004.
[5] Staelens N, Moens S, Van den Broeck W, Marien IIse, et al. "Assessing quality of experience of IPTV and video on demand services in real-life environments," *Broadcasting, IEEE Transactions on* 56.4 (2010): 458-466.
[6] Ozgur Oyman, Sarabjot Singh. "Quality of experience for HTTP adaptive streaming services," *Communications Magazine, IEEE* 50.4 (2012): 20-27.
[7] Mok R K P, Chan E W W, Chang R K C. "Measuring the quality of experience of HTTP video streaming," *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*. IEEE, 2011.
[8] Hyun long Kim, t Seong Gon Choi. "A study on a QoS/QoE correlation model for QoE evaluation on IPTV service," *Advanced Communication Technology (ICACT), 2010 The 12th International Conference on*. Vol. 2. IEEE, 2010.
[9] Karan Mitra, Christer Ahlund, Arkady Zaslavsky. "QoE estimation and prediction using hidden Markov models in heterogeneous access networks," *Telecommunication Networks and Applications Conference (ATNAC), 2012 Australasian*. IEEE, 2012.
[10] Markus Fiedler, Tobias Hossfeld, Phuoc Tran-Gia. "A generic quantitative relationship between quality of experience and quality of service," *Network*, IEEE 24.2 (2010): 36-41.
[11] Brooks, Peter, Bjørn Hestnes. "User measures of quality of experience: why being objective and quantitative is important," *Network, IEEE* 24.2 (2010): 8-13.
[12] Jinbao Cai, Yalin Wu. An Adaptive Rate Control Initialization Method for H. 264 at Low Bit Rate[J]. Journal of Software (1796217X), 2013, 8(10).
[13] ITU-T Rec. P.862 "Perceptual Evaluation of Speech Quality," *International Telecommunication Union, Geneva, 2001*.
[14] Daxing Qian, Hongyu Wang, Wenzhu Sun and Kaiyan Zhu. Bit Stream Extraction Based on Video Content Method in the Scalable Extension of H. 264/AVC[J]. Journal of Software (1796217X), 2011, 6(10).
[15] Yan Tan, Chao Wang, "Fuzzy degree of image based on fuzzy mathematics," *Journal of Computer Aided Design and Computer Graphics*, vol. 14, No.8, August 2002.
[16] Jiegu Li, *Image Processing Technology*. Shanghai Jiaotong University Press, 1988.
[17] Xu Zheng, Bo Yang, Yawen Liu, Guangjian Shi. "Blockiness evaluation for reducing blocking artifacts in compressed images," *2009 Digest of Technical Papers International Conference on Consumer Electronics*. 2009: 1-2.
[18] Yitong Liu, Yun Shen, Qianhong Liu, Shi Ran, Yang Hongwen, Yang Dacheng, et al. "Performance Evaluation and Accuracy Upgrading of PESQ in Chinese Environment," *Vehicular Technology Conference (VTC Spring)*, 2013 IEEE 77th. IEEE, 2013: 1-5.
[19] Kui Tian, Yitong Liu, Yongyu Chang, Dacheng Yang, "A novel full-reference video quality assessment method." *Modern Science and Technology of Telecommunications*, vol.6, June 2011.
[20] ITU-R BT.500-11,"Methodology for the subjective assessment of the quality of television pictures," 2002.6
[21] ITU R BT.1788, "Methodology for the subjective assessment of video quality in multimedia applications," 2007.
[22] Taichi Kawano, Kazuhisa Yamagishi. "No reference video-quality-assessment model for video streaming services," *Packet Video Workshop (PV),* 2010 18th International. IEEE, 2010.
[23] Arnd Eden. "No-reference estimation of the coding PSNR for H. 264-coded sequences," *Consumer Electronics*, IEEE Transactions on 53.2 (2007): 667-674.
[24] Yun Shen, Yitong Liu, Nan Qiao, Lin Sang, Dacheng Yang. "QoE-based evaluation model on video streaming service quality," *Globecom Workshops (GC Wkshps),* 2012 IEEE.
[25] Network Working Group Request for Comments:3550, "RTP: A Transport Protocol for Real-Time Applications," *IETF RFC3550 (2003)*.

Yitong Liu received M.S degree in communication engineering from BUPT, China, in 2007. She is a Ph.D. student in School of Information and Communication Engineering in BUPT. Her research interests include QoE of streaming service, quality and performance for wireless network, and wireless applications.

Hao Liu received his B.E. degree in Communication Engineering from BUPT in 2013. Currently, he is a Master in WT&T Lab in BUPT. His research interests mainly focus on video coding, nature language processing(NLP) and Dynamic Adaptive Streaming of HTTP (DASH).

Yuchen Li is an M.S. candidate in Wireless Theories and Technologies Lab, Beijing University of Posts and Telecommunications (BUPT), China. He received her B.E. degree of Communication Engineering in BUPT in 2012. His research interest now is Quality of Experience.

Yun Shen received B.E. degree in communication engineering from BUPT, China, in 2010. He is currently a Ph.D. student in WT&T Lab. His research interests include source/channel coding, multimedia communication, and QoE study on application services.

Jianwei Wu is a senior engineer in BUPT. He received his M.S. degree in 2007. His research interests focus on the computer networks, switching and routing technologies, new generation networks theory and technologies.

Dcheng Yang received Ph.D. degree in communication engineering from BUPT in 1988, and is currently a professor in BUPT. He is also the director of the WT&T Lab at BUPT. Since 1988, he has engaged in studies on communication systems. His recent interests are in wireless transmission techniques and systems.