SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2017, Vol. 60 042404:1–042404:14
doi: 10.1007/s11432-015-1044-3

HAS QoE prediction based on dynamic video features with data mining in LTE network

Fei WANG, Zesong FEI*, Jing WANG, Yifan LIU & Zhikun WU

School of Information Science and Engineering, Beijing Institute of Technology, Beijing 100081*, China*

Received April 19, 2016; accepted May 23, 2016; published online November 16, 2016

Abstract Evaluation of HTTP adaptive streaming (HAS) quality of experience (QoE) over LTE network is a challenging topic because of multi-segment and multi-rate features of dynamic video sequences. Different from the traditional QoE evaluation methods based on network parameters, this paper proposes the HAS QoE prediction methods based on its dynamic video segment features with data mining. Considering the application requirement of the trade-off between accuracy and complexity, two sets of methodologies are designed to evaluate the HAS QoE including regression and classification. In regression method, we propose the evolved PSNR (ePSNR) model using differential peak signal to noise ratio (dPSNR) statistics as the segment features to evaluate HAS QoE. In classification method, we propose the improved weighted k-nearest neighbors (WkNN) by using dynamic weighted mapping according to the position of video chunk to meet the dynamic segment and rate features of HAS. In order to train and test these methods, we build a real-time HAS video-on-demand (VOD) system in LTE network and do subjective test in different video scenes. With the mean opinion score (MOS), the regression and classification methods are trained to predict the HAS QoE. The validated results show that the proposed ePSNR and WkNN methods outperform other evaluation methods.

Keywords HTTP adaptive streaming (HAS), quality of experience (QoE), regression, classification, data mining, video-on-demand (VOD), long term evolution (LTE)

Citation Wang F, Fei Z S, Wang J, et al. HAS QoE prediction based on dynamic video features with data mining in LTE network. Sci China Inf Sci, 2017, 60(4): 042404, doi: 10.1007/s11432-015-1044-3

1 Introduction

The mobile video tsunami has flown into everyday life in different kinds of services and applications. The study and forecast from [1] show that the mobile video accounts for 55% of global mobile data traffic by the end of 2014 and it will increase to 72% in 2019. The mobile video traffic is still in infancy compared with its potential market. The bottleneck of the explosive mobile video is the contradiction of high user experience and the limited available bandwidth. To solve this problem, the HTTP adaptive streaming (HAS) delivery method is proposed as a flagship on the ocean of mobile video service.

As a video transmission method, HAS encodes the source video into different bit rates and cuts the whole video into many small segments. The client probes the current bandwidth and requests the sequential video segments with a proper bit rate. The HAS as a key technique can transmit video

* Corresponding author (email: feizesong@bit.edu.cn)

© Science China Press and Springer-Verlag Berlin Heidelberg 2017

segments in different video rates under the rapidly changing network. As a solution of the multi-rate video transmission [2–5], HAS was first proposed by the company Move Networks in 2007. Many other corporations followed the step. In 2008, Microsoft launched its own HAS named Microsoft Smooth Streaming (MSS) [6]. The Apple Inc. developed the HTTP Live Streaming (HLS) for its iPhone and iPad products in 2009 [7]. In 2010 the Adobe System Inc. proposed the Adobe HTTP dynamic streaming (HDS) to guard his media throne in the Internet [8]. On the other hand, the standardization work of HAS is also promoted by the 3rd Generation Partnership Project (3GPP). The latest 3GPP TS 26.247 (v13.1.0) [9] introduces the 3GP-Dynamic Adaptive Streaming over HTTP (3GP-DASH) structure in detail. Meanwhile, Dynamic Adaptive Streaming over HTTP (DASH) was launched as a standard of Moving Picture Experts Group (MPEG) belonging to ISO/IEC. The MPEG-DASH is taken as a standard of HAS. Although different names are called for the HAS, their key technologies and system constructions are similar. We call HAS as a universal name in the following content.

The burgeoning HAS technology proposes great demand for HAS quality evaluation. Especially in the mobile network, the end user's subjective feeling is the most important factor with network service provider's (NSP) concern. Because of multi-segment and multi-rate features of dynamic video sequences, the HAS quality of experience (QoE) evaluation is a challenging research aspect attracting the academia and industry. In order to assess the accuracy of HAS QoE evaluation model, we use pearson correlation coefficients (PCC) as a metric recommended in ITU-T video quality experts group (VQEG) report [10]. When PCC is larger than 0.9, it means that QoE model is accurate. For the universality, the outlier ratio (OR) is the metric recommended in VQEG report [10]. The lower OR means the model has a higher consistency of prediction. In the following, we will introduce the existing HAS QoE models we investigate. In [11], authors use the network-level packet characteristics and streaming events to evaluate the QoE. The PCC is only 0.85 and there is no result about the OR. In [12], authors mention the concept of HAS and give some metrics from 3GP-DASH to evaluate the HAS QoE in different aspects. There is no normal formulation to evaluate the HAS QoE. The work in [13] presents the parameters influencing the HAS QoE including the fluency, startup bit rate, bit rate switching and bit rate distribution. The analyses only show the influence of the parameters separately. It doesn't show the expression between the parameters and QoE. Ref. [14] presents the HAS QoE model concerning three parameters: the re-buffering event frequency (RER), re-buffering event average duration (RED) and representation quality switching rate (RQSR). The author shows the model in polynomial function with rational exponent power format. The results show that in high bandwidth about 8 Mbps, the predicted QoE is overlapped with the real subjective mean opinion score (MOS). The PCC and OR are not mentioned. In 2013, the next generation mobile networks (NGMN) alliance did some work about HAS quality evaluation and presented results in [15, 16]. The most exciting part in [15, 16] is that the NGMN presents a HAS QoE model named linear MOS predictor (LMP). The PCC is 0.863 and the OR is 11%. It is a lightning QoE evaluation method. The quality level used in the LMP is the peak signal to noise ratio (PSNR) value of the chunks [16]. The NGMN LMP is taken as a benchmark with great performance and we will compare our methods with it in Section 5. To sum up, in traditional network centric HAS QoE evaluation method, researchers only focus on the network performance but neglect the end users' perceive quality of the service. In terms of the video service, especially the HAS, the main parameter influencing the user is the dynamic video feature. The video bit rate distribution and multi-segment with different video rates reflect the direct feelings from end users. We use classification and regression method with data mining to fit the dynamic video feature and MOS. As the survey we present, the HAS QoE model from the network is not enough to evaluate the subjective feeling accurately [17, 18]. In this paper, we focus on the dynamic video features and show a further step on the way of the HAS QoE evaluation with high accuracy and the methods can be implemented easily in different conditions.

In our research work, we predict the HAS QoE to meet the multi-segment and multi-rate features of HAS with data mining. Considering the trade-off between accuracy and complexity, two sets of methodologies are proposed: regression and classification. In regression method, we explore an objective factor or variable that can affect the human's subjective feeling of the HAS video. The evolved PSNR (ePSNR) is a HAS QoE evaluation model we proposed formed by differential PSNR (dPSNR) statistics. The ePSNR method only contains the application parameters to calculate the objective MOS. In Subsection 4.2, the detail of ePSNR is presented. In classification method, the traditional k-nearest neighbors (kNN) and mean k-nearest neighbors (MkNN) are launched firstly. Unlike the kNN and MkNN method, we give the different weights in the improved weighted kNN (WkNN) based on the video chunks positions of the sequences. The results present that both the ePSNR and WkNN have higher PCC than the NGMN LMP method with PSNR.

By means of data mining methods of both regression and classification, we build the HAS QoE evaluation schemes based on perception, psychics and feeling of end user. Different from the other models using the network parameters to "reflect" the MOS, we take the video quality level dPSNR or video bit rates level itself to "present" the MOS. For one thing, these parameters are in dynamic form and represent adequately the multi-segment and multi-rate features of HAS. For another, when calculating the distance between HAS video samples in WkNN method, the dynamic weights are also catering to the dynamic segment characteristic of HAS. Because of these reasons, the ePSNR and WkNN are more direct and accurate methods to predict the HAS QoE.

To complete the research, we first give the QoE and video feature relationship from a new perspective. Although Refs. [19,20] give the analysis relationship between the video quality and video rate, the authors only conclude them in a statistical way and do not show the nature of this relationship. We analyze the video feature and QoE in a new perspective from psychophysical way. This gives a reasonable explanation of research about the MOS and objective video features. We bulid a real-time HAS QoE testbed with different radio access network (RAN) parameters over LTE network. The testbed is different from the traditional one only for emulation. Our testbed transmits the HAS segments in real-time with different RAN parameters and it creates the real LTE network environment of video play. The users can mark the video sequences simultaneously. It saves much time and cost of subjective test. What's more, the video segments, stored in the database, can be used for playback and future data mining. The HAS QoE evaluation testbed can launch the live or on-demand video transmission. The models we proposed can be trained and tested in testbed with subjective MOS. It is also a universal system and users can mark the video in iOS, Android and Windows system.

The remainder of this paper is organized as follows. In order to follow the analysis process, Section 2 analyzes the relationship between the QoE and video feature in a new perspective. In Section 3, the real-time HAS QoE testbed over LTE network is introduced. The subjective test work like the video sequences choice, marking MOS process and the data analysis are aslo given in Section 3. In Section 4, the proposed regression and classification methodologies are given. The performance comparisons of different methods in different metrics are discussed in Section 5. Finally, Section 6 provides the conclusion and future work.

2 QoE and video feature relationship from psychophysical perspective

Traditional QoE and video feature researches only give the statistical data, while the nature of relationship is not stated clearly [19–22]. In this section, we will analyze the QoE and video feature relationship from a different perspective. The relationship is given from psychophysical way and the human visual system inspired by Weber's law. We try to find objective parameters of video feature as the shadow role of the subjective MOS.

2.1 Inspiration from Weber Fechner law

In psychophysics, the Weber's law is proposed by Emst Hernrich Weber. It states that the ratio of the increment threshold to the background intensity is constant. It can be shown as follows:

$$\frac{\Delta I}{I} = K, \tag{1}$$

where I represents the original stimulus intensity and ΔI is the increment threshold. The Gustav Fechner proposes the different equation of this relationship,

$$dP = k \cdot \frac{dS}{S}, \tag{2}$$

$$P = k \cdot \ln \frac{S}{S_0}, \tag{3}$$

where dP is the differential change in perception and dS is the differential change of stimulus. The S_0 is the initial threshold of the stimulus which means people cannot perceive any subjective feeling. The combination of (1), (2) and (3) is called Weber Fechner Law (WFL). It is a basic theory that depicts the human response to the physical stimulus in a quantitative way.

2.2 Relationship between QoE and video feature with WFL

The mean opinion score (MOS) definition is given in [23]. It is the values on a predefined scale that subjects assign to their opinion of the performance of the telephone transmission system. And in [24] the MOS conception is extended to the video aspects. To rate the video quality, five level scale has been used: Excellent (5), Good (4), Fair (3), Poor (2) and Bad (1). These scores represent the viewer's perceptual opinion to video quality. It is also called the MOS.

In the video quality evaluation aspect, the subjective and objective methods are proposed to assess the video quality. The subjective evaluation method is a complex work with much cost and time. In [25], some subjective test method are proposed such as absolute category rating (ACR), absolute category rating with hidden reference (ACR-HR), degradation category rating (DCR) and pair comparison (PC) methods. These methods has a common score scheme. The objective evaluation method uses the characteristic parameters in single or multi-dimension from the video source to the client, but the objective video quality (the quality scale achieved from the objective assessment method) sometimes cannot reflect the people's real watch feeling to the video. The concept of video QoE is proposed to show an evaluation of combining user perception, experience, and expectations to the video service with technical and non-technical factors [26]. Apart from the subjective opinion, the abbreviation MOS is also used as the scores originated from the objective method [23]. Because the human's subjective perception to the video quality is the most important part of QoE, as a general rule, an objective method to generate the MOS score is called as QoE calculation method.

Although Refs. [19, 20] give the analysis relationship between the video quality and video rate, the authors only conclude them in a statistical way and do not show the nature of this relationship. Based on our introduction, WFL can be used in the sound, vision, numerical cognition and pharmacology. It shows the good behavior in a wide application scenario. As we said earlier, the QoE is also a metric to evaluate the human's subjective perception to the video feature. Here, the feature parameters can be taken as the stimulus and the QoE is the human response. The correlation between video feature and QoE can be written as WFL (2) and (3). It can be given as

$$dQ = k \cdot \frac{dF}{F}, \tag{4}$$

where Q represents QoE and F represents the video features.

As an example, in the test environment [25], we take the experiment to find the relationship between the MOS and video rate. A logarithmic function relationship is show in Figure 1. The equation is

$$y = a \times \ln(x) + b, \tag{5}$$

where $a = 1.7787$ and $b = 1.4829$ are calculated from our test data.

The video features are not limited in the rate, all the parameters representing video can be the features. Inspired by the WFL relationship between the QoE and video feature and the PSNR formula, we try to find out objective parameters to play as the shadow role of the subjective MOS. It must be from

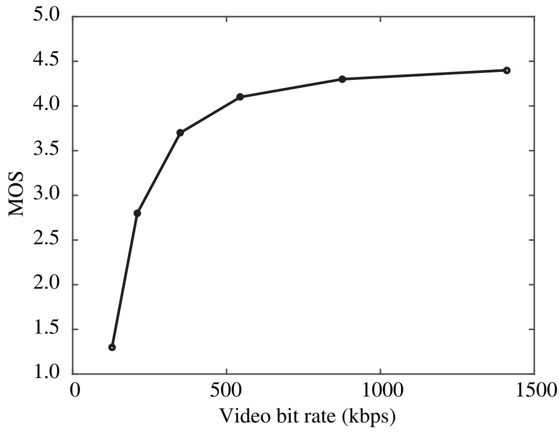

Figure 1 The MOS and the video rate logarithmic relationship.

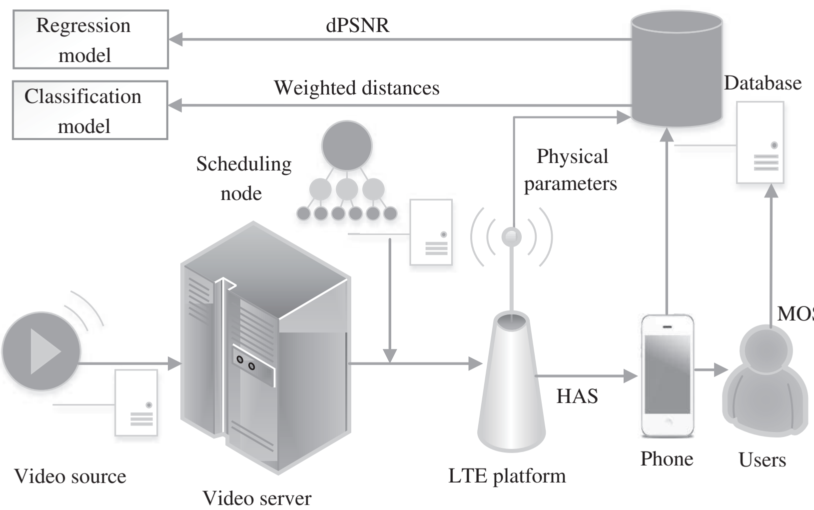

Figure 2 The LTE HAS QoE evaluation system.

the trait of the HAS itself and catering to the multi-rate and multi-chunk characteristic of HAS. In regression method, we design a parameter dPSNR as key indicator of QoE and generate the mapping model ePSNR. In classification method, we use the relationship of video rates and video quality to find a dynamic weighted distance to predict the QoE. Before showing the details of them in Section 4, we introduce the testbed first.

3 Real-time HAS QoE testbed over LTE network

In this section, we will introduce our real-time QoE HAS testbed. Different from traditional testbeds just for emulation [27, 28], our testbed gives a real-time HAS transmission with different RAN parameters in LTE network. The LTE platform in testbed fits the 3GPP specifications. The volunteers can do subjective test at the same time with different video play scenes in our real-time testbed.

3.1 Architecture and feature of real-time testbed

In our work, the framework of HAS QoE evaluation testbed is shown in Figure 2. The real-time testbed transmits videos in HAS with different RAN parameters for subjective test. Not like the traditional testbed just for emulation [27–29], the real-time HAS QoE evaluation testbed creates real video play in LTE network environment. The users can mark the video sequences simultaneously. It saves much time of subjective test. Furthermore, the video segments, stored in the database, can be used for playback and

Table 1 LTE simulation parameters

Parameters	Value
Scenario	UMa: 2 G CF, 500 m ISD, 10 M BW, speed 0.3 km/h
Cellular layout	Hexagonal grid, 19 sites, 3 cells per site, wrap around
Pico layout	1 Picos per cell
Load	Average 10 UE per cell
UE distribution	Users dropped uniformly in entire cell
Total eNB TX power	46 dBm
Total Pico Tx power	30 dBm
BS antenna gain plus cable loss	14 dBi
UE antenna gain	0 dBi
Noise figure at Pic	5 dB
Noise figure at UE	9 dB
Noise power of spectral density of UE	−174 dBm/Hz
Distance-dependent path loss for macro to UE	$L = 128.1 + 37.6 \cdot \lg(R)$ (R in km)
Distance-dependent path loss for macro to Pic	$L = 124.5 + 37.6 \cdot \lg(R)$ (R in km)
Distance-dependent path loss for Pico to UE	$L = \mathrm{Prob}(R) \cdot PL_{\mathrm{LOS}}(R) + [1 - \mathrm{Prob}(R)] \cdot PL_{\mathrm{NLOS}}(R)$, $PL_{\mathrm{LOS}}(R) = 103.8 + 20.9 \cdot \lg(R)$, $PL_{\mathrm{NLOS}}(R) = 145.4 + 37.5 \cdot \lg(R)$, $\mathrm{Prob}(R) = 0.5 - \min(0.5, 3 \cdot \exp(-0.3/R))$ $+\min(0.5, 3 \cdot \exp(-R/0.095))$, ($R$ in km)

future data mining. The LTE parameters are shown in Table 1. The LTE platform follows the 3GPP specifications [30–34]. The video source can be the IP camera for live event or the video files for video-on-demand (VOD). Because the HAS is based on HTTP, the video server can use the traditional HTTP server like Apache [35]. In order to analyze the different objective parameters and the subjective MOS. We set a scheduling node to choose the different video scenes. The physical parameters are collected from the LTE platform [36, 37]. The dPSNR is the value of the QoE metric of the video sequences collected from the terminal equipment in regression model and the dynamic weighted distances are calculated from the existing video scenes in classification method in the database. The database aggregates the subjective MOS and the video features and formulates the QoE evaluation model.

3.2 Evaluation scenes and configuration

In order to minimize the single scene's random impact on subjective test, we choose two typical video clips and present 90 different video evaluation scenes. The Sintel [38] is an excellent open movie film by the Blender Institute, part of the Blender Foundation. The NGMN also uses this film for the test [16]. We take two video clips from the whole movie. The clip1 from 0:00 to 1:40 represents the static to movement scene and the clip2 from 3:52 to 5:42 represents the movement to static scene shown in Figure 3. We first encode the original video into the H.264 format and the audio into the AAC format. The video resolution is 988×420 and bit rates are in 6 levels (128 kbps, 210 kbps, 350 kbps, 545 kbps, 876 kbps, 1410 kbps). The audio is static 64 kbps. The segment duration is 5 s and 22 segments are contained in each clip.

When the client plays the video, the first two segments are requested in 210 kbps (getting a subminimum value in order to decrease the initial buffering time and guarantee a relative good QoE), and then the client posts the requirement of the next video segment according to the current bandwidth.

In the lab environment, we set 90 different scenes of clip1 and clip2. The 90 video scenes are chosen typically in different network fluctuation conditions, the normal network is stable and the change times of segment video rate mostly distribute from 1 to 5 times shown as in Figure 4. The process and video play are in real-time and the experiment is designed by ACR method. The method specifies that after each presentation the subjects are asked to evaluate the quality of the sequence shown in [25]. Each of 90 volunteers is invited to mark 6 scenes in random. We guarantee each scene is marked with 6 different subjective MOS. 90% of all the data are collected randomly in training set and 10% are in the test set.

Figure 3 Two typical clips from the Sintel [38]. (a) Clip1 scene: fighting on the snow; (b) clip2 scene: chasing in the market.

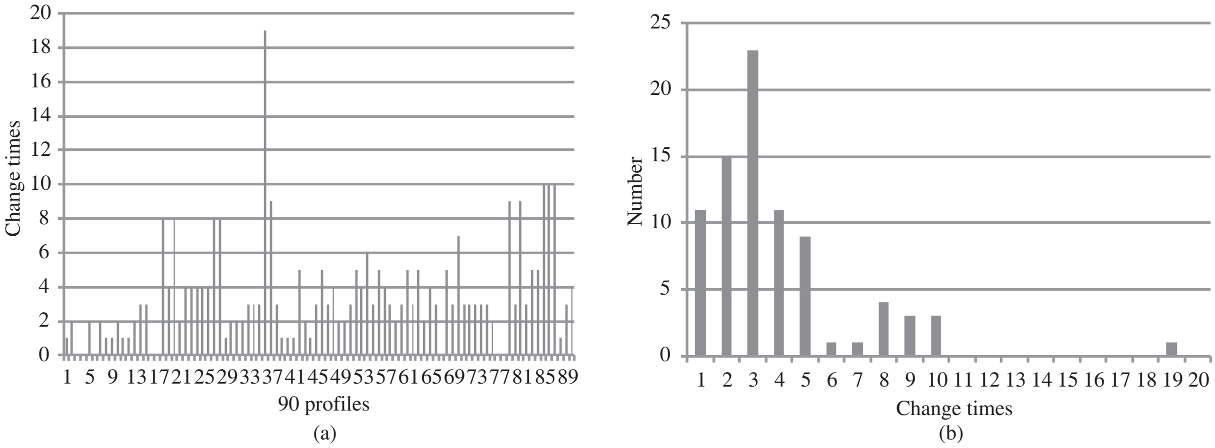

Figure 4 The video rate change times and distribution of different scenes. (a) Video bit rate change times of 90 profiles; (b) distribution of video bit rate change times.

4 Proposed methodologies with data mining

Different from the traditional QoE evaluation methods using network parameters, we proposes the HAS QoE evaluation methods based on its dynamic change of video segment features with data mining methods. In the following parts, the motivation of our work is given and the regression and classification methods fitting the multi-segment and multi-rate features of HAS are introduced in detail.

4.1 Motivation

The QoE and video features can be simply formulated as (4) in psychophysical perspective. If we want to derive the accurate relationship, more specific representations of video feature must be found. A more general formulation between HAS QoE and video features can be expressed as

$$Q_{\mathrm{HAS}} = f(v_1, v_2, v_3, \ldots, v_n), \tag{6}$$

the $v_1, v_2, v_3, \ldots, v_n$ are the representations of HAS video feature. More video features can reflect QoE more clearly, but it will induce computational complexity and overfitting. For the HAS video transmission, end users get the different video segments with different video bit rates. The bit rates change between consecutive segments. The multi-segment and multi-rate features of HAS are distinct from the traditional video streaming. These are the features we can use to derive the HAS QoE. Considering the application requirement of the trade-off between accuracy and complexity, two sets of methodologies are designed to evaluate the HAS QoE including regression and classification. In order to represent dynamic video features, we get the ePSNR as a formula containing the average, maximum, minimum and standard deviation of dPSNR. What's more, the distance between the different segments with different video rates can be the metric of similarity between two different play scenes. The details of two methods are given in next two parts.

4.2 Regression: dPSNR and ePSNR

As to the HAS transmission process we analyze before, the terminal requests the video segments at the client side. In our test, after the whole video play, 22 segments and their information are the basic factors influencing the QoE. In the traditional video quality or QoE evaluation, the PSNR is an objective method in the image evaluation. It is calculated as follows:

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[f(i,j) - f'(i,j)\right]^2, \tag{7}$$

$$\mathrm{PSNR} = 10\lg\frac{I^2}{\mathrm{MSE}}, \tag{8}$$

where $f(i,j)$ is the pixel value in ith row and jth column of original image and the $f'(i,j)$ is the pixel value of the the degraded image. I is the maximum possible pixel value of the image. For color images with three RGB values per pixel, the definition of PSNR is same except the MSE is the sum over all squared value differences divided by image size. In the calculation of PSNR, we usually calculate average value of RGB three dimensions. In video quality assessment, we calculate the mean value of all the frames' PSNR values. The video consists of continuous image feature. But when we evaluate the quality of video, as we introduce in Section 2, the comparison between the high bit rate chunk we receive and the current chunk can reflect the difference of video feature. The high rate chunk is taken as the reference value.

We propose the dPSNR which is calculated as follows:

$$\mathrm{dPSNR} = \mathrm{PSNR} - \mathrm{PSNR_{ref}} = 10\lg\frac{I^2}{\mathrm{MSE}} - 10\lg\frac{I^2}{\mathrm{MSE_{ref}}} = 10\lg\frac{\mathrm{MSE_{ref}}}{\mathrm{MSE}}. \tag{9}$$

Here the dPSNR is the basic video feature parameter representing the QoE of each video chunk. As the pooling method shown in the NGMN technical paper [16], we take four statistical magnitudes of dPSNR to fit the subjective MOS. The mean value, minimum, maximum and standard deviation are the components to evaluate the quality of HAS QoE. In the following descriptions, we take the matrix $\boldsymbol{Q}$ representing the dPSNR. $\boldsymbol{Q} = \{q_{ij}\}_{M\times N}$ and the q_{ij} indicates the dPSNR value of ith video scene and jth video segment, here, $i = 1,2,\ldots,k, j = 1,2,\ldots,n$ and in our test $k = 90$, $n = 22$. The statistical value of $\boldsymbol{Q}$ is $\tilde{\boldsymbol{Q}}$ which can be simplified to

$$\tilde{\boldsymbol{Q}} = \left(\operatorname*{mean}_j(q_{ij})\ \operatorname*{max}_j(q_{ij})\ \operatorname*{min}_j(q_{ij})\ \operatorname*{std}_j(q_{ij})\right)^{\mathrm{T}}, \tag{10}$$

where the $\operatorname*{mean}_j(q_{ij})$, $\operatorname*{min}_j(q_{ij})$, $\operatorname*{max}_j(q_{ij})$ and $\operatorname*{std}_j(q_{ij})$ are calculated in

$$\operatorname*{mean}_j(q_{ij}) = \frac{1}{n}\sum_{j=1}^{n} q_{ij}, \tag{11}$$

$$\operatorname*{min}_j(q_{ij}) = \min(q_{i1}, q_{i2}, \ldots, q_{in}), \tag{12}$$

$$\operatorname*{max}_j(q_{ij}) = \max(q_{i1}, q_{i2}, \ldots, q_{in}), \tag{13}$$

$$\operatorname*{std}_j(q_{ij}) = \sqrt{\frac{1}{n}\left(\sum_{j=1}^{n}(q_{ij} - \operatorname*{mean}_j(q_{ij}))\right)}. \tag{14}$$

We then propose the concept ePSNR as the predicted MOS of the QoE model. It is shown that

$$\mathrm{ePSNR} = (a,b,c,d)\tilde{\boldsymbol{Q}} + \boldsymbol{e}. \tag{15}$$

Our purpose is using the multiple linear regression to fit the MOS with best a, b, c, d and e. In Section 5, we will give the detail of comparison of ePSNR and NGMN LMP method with PSNR.

4.3 Classification: improved WkNN

As the lightning and simple method of the classification, the kNN aims at finding the k closest training examples of the object in the feature space. In the HAS video, the basic feature of each segment is the video bit rate and segment duration. In the real application, the segment durations are the same, so we focus on the video bit rate. We take each chunk's bit rate attribute tagged as six level number from 1 to 6. The video scene is taken as the example representing as x. Its feature vector is written as $(a_1(x), a_2(x), \ldots, a_n(x))$. Here $a_r(x)$ is the rth attribute of example x ($r = 1, 2, \ldots, n$, here, n is 22 in our test) and the value is from 1 to 6. The distance between two examples x_i and x_j is $d(x_i, x_j)$. It can be expressed as

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2}. \tag{16}$$

The set $V = (v_1, v_2, \ldots, v_s)$ is the MOS set marked by te subjective users. The target function is $f : R^n \to V$. If the $x_q \in R^n$ is the unclassified variable, we assume that $\widehat{f}(x_q)$ is the estimation of $f(x_q)$. The $\widehat{f}(x_q)$ is decided by

$$\widehat{f}(x_q) \leftarrow \underset{v\in V}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(v, f(x_i)), \tag{17}$$

$$\delta(v, f(x_i)) = \begin{cases} 1, \ v = f(x_i), \\ 0, \ v \neq f(x_i). \end{cases} \tag{18}$$

While in the application of kNN in video QoE evaluation, the subjective MOS from volunteers are processed after test. Thus the value of final MOS is continuous. The target function is the continuous. The $\widehat{f}(x_q)$ is rewritten as

$$\widehat{f}(x_q) \leftarrow \frac{\sum_{i=1}^{k} f(x_i)}{k}. \tag{19}$$

Because this method takes the average of k nearest $f(x_i)$, we call it MkNN. While, it is obvious that these two methods take the weight of the $f(x_i)$ as the same value. The improved kNN method considers the different weight of different $f(x_i)$. The Eqs. (17) and (19) are rewritten as follows:

$$\widehat{f}(x_q) \leftarrow \underset{v\in V}{\operatorname{argmax}} \sum_{i=1}^{k} w_i \delta(v, f(x_i)), \tag{20}$$

$$\widehat{f}(x_q) \leftarrow \frac{\sum_{i=1}^{k} f(x_i)}{\sum_{i}^{k} w_i}, \tag{21}$$

the w_i is calculated by

$$w_i = \frac{1}{d(x_q, x_i)^2}. \tag{22}$$

As an improvement and extension, we focus on the dynamic weighted distances of HAS video QoE evaluation. The improved distance $\widetilde{d}(x_i, x_j)$ is rewritten as

$$\widetilde{d}(x_i, x_j) = \sqrt{\sum_{r=1}^{n} \lambda_r (a_r(x_i) - a_r(x_j))^2}, \tag{23}$$

where $\widetilde{d}(x_i, x_j)$ represents the whole difference between the different video scenes. The human visual system is sensitive to difference of video feature. But for the whole video scenes, the different position of video segment has different influencing weight on the human visual system [39]. We add the λ_r as the

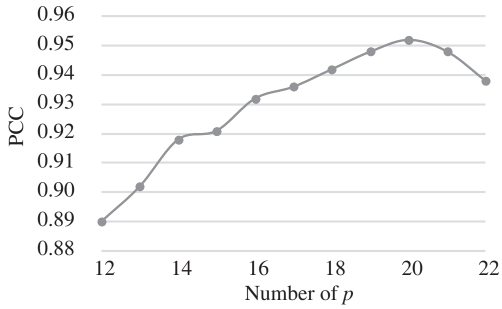

Figure 5 The PCC performance under different number of p.

weight factor of the video segment position. So the distance between video scene x_i and x_j cannot be written as simple way in (16). The distance of last segment should give the larger weight. Where λ_r is the weight and is trained in the test, when r is larger than the pth video segment $\lambda_r > 1$ or else $\lambda_r = 1$. In our HAS video test, we set the p as a variable with $\lambda_r = 2$. The PCC results is shown in Figure 5. We observe that when p is 20, the PCC is the maximum. It means that the last two segments are more important than other segments from end user's view.

In the following section, we will compare the kNN, MkNN and WkNN. When applying the regression and classification method, 90% of the subjective MOS are taken as the training examples and the other data are used to test model.

5 Performance analysis

5.1 Five metrics to assess the performance of methods

In the data analysis, the prediction model can be evaluated by many metrics. Here we take mean absolute percentage error (MAPE), root mean square error (RMSE) and PCC, spearman's rank correlation coefficient (SROCC) and OR [10]. Assuming that the total number of the video sequence i is N, the S_i represents the subjective score marked by the testee and the P_i is the prediction score (the objective score) calculated by the QoE prediction model. Following are the details of five metrics and their realistic meaning in the QoE evaluation.

5.1.1 *MAPE*

MAPE is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses accuracy as a percentage and is defined by the formula,

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{S_i - P_i}{S_i}\right|. \tag{24}$$

Always, the MAPE value is multiplied by 100 to make it to be a percentage error. The small MAPE value indicates that the model is efficient.

5.1.2 *RMSE*

The RMSE is an indictor used to measure the accuracy of the predicted model. It represents the sample standard deviation between the observed and predicted value. It is calculated by

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - P_i)^2}. \tag{25}$$

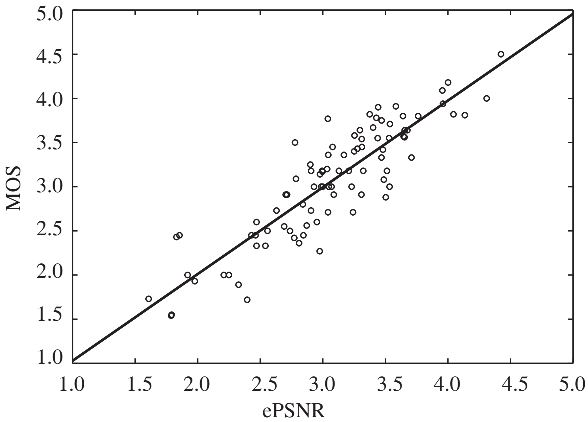

Figure 6 The mapping between subjective MOS and the ePSNR in different video scenes.

5.1.3 *PCC*

In statistics, the PCC is a measure of linear correlation between two variables. Here we use the PCC to measure the different QoE predicted model or method, the bigger PCC indicates the better model. When calculating the PCC, the division is formed covariance and the product of standard deviation S and P. The formula is

$$R = \frac{\sum_{i=1}^{N}\left(S_i - \bar{S}\right)\left(P_i - \bar{P}\right)}{\sqrt{\sum_{i=1}^{N}\left(S_i - \bar{S}\right)^2}\sqrt{\sum_{i=1}^{N}\left(P_i - \bar{P}\right)^2}}, \tag{26}$$

where R represents the value of PCC, $\bar{S}$ is the average value of the actual score and $\bar{P}$ is the average value of the predicted MOS.

5.1.4 *SROCC*

The SROCC is defined as the PCC between ranked variables. The SROCC value ρ can be computed by

$$\rho = 1 - \frac{6\sum_{i=1}^{N} {d_i}^2}{N\left(N^2 - 1\right)}, \tag{27}$$

where $d_i = s_i - p_i$ is the difference between ranks. The s_i is the rank of the value S_i and p_i is the rank of the value P_i. The Spearman correlation increases in magnitude as two variables become closely perfect monotone functions of each other.

5.1.5 *OR*

The OR reflects the ratio of the "false" objective score to the total number of scores. The P_i is the "false" score when $|P_i - S_i| > 2\sigma_i$, where σ_i is the standard deviation of the S_i. The OR is calculated from

$$\mathrm{OR} = \frac{n_{\mathrm{OR}}}{N}, \tag{28}$$

where n_{OR} is the total number of the "false" objective score. The metric OR can assess the stationarity of the QoE evaluation model. If the OR is 0, the objective method will be stable to predict the QoE.

5.2 Results and discussion

In this section, we compare the method in regression and classification aspects with five assessment metrics. Figure 6 shows the mapping between the subjective MOS and the ePSNR. The ePSNR's PCC value is 0.891 in linear relationship and the WkNN's PCC is 0.952. In Figure 7, the MAPE, RMSE, PCC and SROCC metrics are comparison in the LMP model with PSNR, ePSNR, kNN, MkNN and WkNN. In the view of MAPE and RMSE metric, the classification methods are better than the regression methods with the smaller error. The ePSNR method we proposed is more accurate than the LMP method with

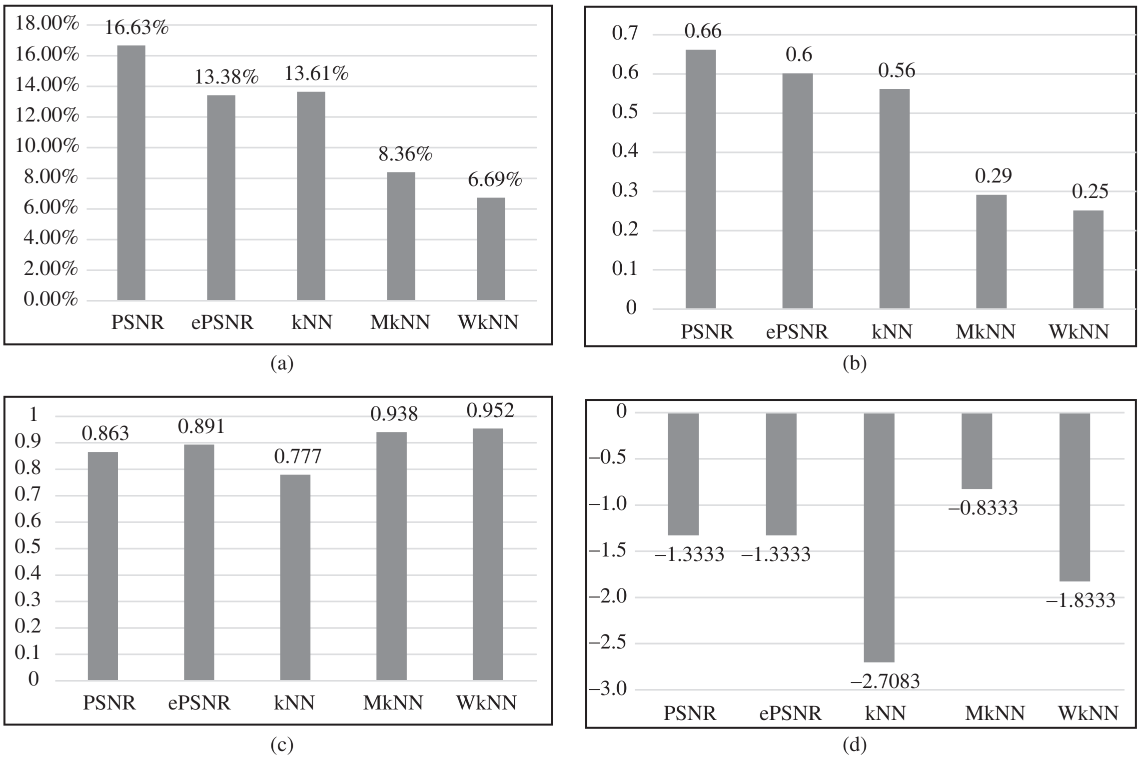

Figure 7 The metric comparison between different video QoE evaluation methods. (a) MAPE; (b) RMSE; (c) PCC; (d) SROCC.

Table 2 Time and space complexity of different methods

Model	Time complexity in training	Time complexity in test	Space complexity in training	Space complexity in test
PSNR	$\Theta(n)$	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$
ePSNR	$\Theta(n)$	$\Theta(1)$	$\Theta(1)$	$\Theta(1)$
kNN	$\Theta(1)$	$\Theta(n)$	$\Theta(1)$	$\Theta(1)$
MkNN	$\Theta(1)$	$\Theta(n)$	$\Theta(1)$	$\Theta(1)$
WkNN	$\Theta(1)$	$\Theta(n)$	$\Theta(1)$	$\Theta(1)$

PSNR. Because the WkNN method takes the dynamic weights as a factor when calculating the different distances, it shows a better performance than the kNN and MkNN.

In Figure 7(c) and (d), the kNN is the worst method considering the PCC value is only 0.777 and SROCC is −2.7083. It is lower than the other methods. The MkNN and WkNN present a better correlation. The OR value of the PSNR method is 11.11% and the other is 0%. It means that the PSNR method is too sensitive and cannot mark the QoE in a stable condition. Comprehensively considering, the WkNN is the most outstanding QoE evaluation method we investigated. It is accurate and has a high correlation with subjective MOS.

For the complexity analysis, we take time complexity and space complexity in training and test. As shown in Table 2, the regression method is $\Theta(n)$ and the classification kNN, WkNN and MkNN is $\Theta(1)$ in the time complexity of training. But when testing a new HAS QoE the time complexity is different, the regression is $\Theta(1)$ and the classification is $\Theta(n)$. So when the examples of QoE evaluation is large, the WkNN method will consume more time than the ePSNR. In space complexity part, the training data and test data size are same in all methods. The space complexity only contains the extra storage of the method. But there is no need for the extra space to store expect the calculate result, so all the space complexity is $\Theta(1)$. In the real application of QoE evaluation method, for example, we plant these methods in the real LTE system or the application software in smart phone, the balance of accuracy and the time consuming is the key point that should be considered. Normally, the ePSNR can be used in the terminal and the single equipment in the network. The WkNN can be used in the cloud system in parallel computing.

6 Conclusion and future work

In this paper, we predict the HAS QoE based on its multi-segment and multi-rate features with data mining methods. Considering the application requirement of trade-off between accuracy and time complexity, the regression and classification methodologies of HAS QoE evaluation are presented. The ePSNR and WkNN HAS QoE methods both can express the subjective MOS value in a certain way. The WkNN is the best method in the accuracy phase with PCC 0.952, while the ePSNR has the lower complexity. These two methods both present a higher PCC and smaller RMSE than the famous NGMN LMP method. Our work shows a further step on the way of the HAS QoE evaluation with high accuracy and our methods can be implemented easily in different conditions with good universality. The proposed QoE models can be used to predict the HAS video quality not only in individual, but also the whole cell and district. They also can be used in QoE monitor, management and the network optimization. The QoE is a metric or an indicator from the view of the consumer and the final goal is using these indicators to optimize the network and service and make further efforts to increase the satisfaction of consumer. In the future, the QoE-based resource management for HAS can be studied and our team will go on promoting the standardization work of HAS evaluation.

Acknowledgements This work was supported by National High Technology Research and Development Program of China (Grant No. 2014AA01A701) and Beijing Natural Science Foundation (Grant No. 4152047).

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Cisco Systems Inc. Cisco visual networking index: global mobile data traffic forecast update 2014–2019 white paper. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html. 2015
2 Zhou Y Q, Liu H, Pan Z G, et al. Spectral and energy efficient two-stage cooperative multicast for LTE-advanced and beyond. IEEE Wirel Commun, 2014, 21: 34–41
3 Zhang Z S, Long K P, Vasilakos A V, et al. Full-duplex wireless communications: challenges, solutions and future research directions. Proc IEEE, 2016, 99: 1–41
4 Zhou Y Q, Liu H, Pan Z G, et al. Two-stage cooperative multicast transmission with optimized power consumption and guaranteed coverage. IEEE J Sel Areas Commun, 2014, 32: 274–284
5 Fei Z S, Luo C, Xing C W, et al. Joint linear processing design for distributed two-way amplify-and-forward MIMO relaying networks. China Commun, 2013, 10: 126–133
6 Zambelli A. Smooth streaming technical overview. http://www.iis.net/learn/media/on-demand-smooth-streaming/smooth-streaming-technical-overview. 2009
7 Apple Inc. HTTP live streaming overview. https://developer.apple.com/streaming. 2014
8 Adobe Systems Inc. HTTP dynamic streaming. http://www.adobe.com/products/hds-dynamicstreaming.html. 2015
9 3GPP. Transparent end-to-end packet-switched streaming service (PSS); progressive download and dynamic adaptive streaming over HTTP (3GP-DASH). TS 26.247 V13.1.0. http://www.3gpp.org/ftp/Specs/archive/26-series/26.247/26247-d10.zip. 2015
10 Video Quality Experts Group. VQEG FRTV phase I final report. http://www.its.bldrdoc.gov/vqeg/-projects/frtv-phasei/frtv-phase-i.aspx. 2000
11 Xue J T, Zhang D Q, Yu H, et al. Assessing quality of experience for adaptive HTTP video streaming. In: Proceedings of the IEEE International Conference on Multimedia and Expo Workshops, Chengdu, 2014. 1–6
12 Oyman O, Singh S. Quality of experience for HTTP adaptive streaming services. IEEE Commun Mag, 2012, 50: 20–27
13 Liu Y T, Shen Y, Mao Y N, et al. A study on quality of experience for adaptive streaming service. In: Proceedings of the IEEE International Conference on Communications Workshops, Budapest, 2013. 682–686
14 Alberti C, Renzi D, Timmerer C, et al. Automated QoE evaluation of dynamic adaptive streaming over HTTP. In: Proceedings of the 5th International Workshop on Quality of Multimeda Experiecne, Klagenfurt, 2013. 58–63
15 de Vriendt J, de Vleeschauwer D, Robinson D. Model for estimating QoE of video delivered using HTTP adaptive streaming. In: Proceedings of the IFIP/IEEE International Symposium on Inergrated Network Management, Ghent, 2013. 1288–1293

16 NGMN Alliance. Service quality definition and measurement. White Paper Version 1. http://www.ngmn.org/publications/all-downloads/article/technical-report-service-quality-definition-andmeasurement.html. 2013
17 Fei Z S, Xing C W, Li N. QoE-driven resource allocation for mobile IP services in wireless network. Sci China Inf Sci, 2015, 58: 012301
18 Zhang Z S, Chai X M, Long K P, et al. Full-duplex techniques for 5G networks: self-interference cancellation, protocol design and relay selection. IEEE Commun Mag, 2015, 53: 128–137
19 Germak G, Pinson M, Wolf S. The relationship among video quality, screen resolution, and bit rate. IEEE Trans Broadcast, 2011, 57: 258–262
20 Belmudez B, Moller S. An approach for modeling the effects of video resolution and size on the perceived visual quality. In: Proceedings of the IEEE International Symposium, Dana Point, 2011. 464–469
21 Xing C W, Ma Y, Zhou Y Q, et al. Transceiver optimization for multi-hop communications with per-antenna power constraints. IEEE Trans Signal Process, 2016, 64: 1519–1534
22 Fei Z S, Li N, Xing C W, et al. Energy-efficient transceiver design for multi-pair two-way relay systems. China Commun, 2015, 12: 133–140
23 ITU-T. Mean opinion score terminology. P.800.1. http://www.itu.int/rec/T-REC-P.800.1-201602-I. 2006
24 ITU-T. Mean opinion score interpretation and reporting. P.800.2. http://www.itu.int/rec/T-REC-P.800.2-201305-I. 2013
25 ITU-T. Subjective video quality assessment methods for multimedia applications. P.910. http://www.itu.int/rec/T-REC-P.910-200804-I. 2008
26 Fiedler M, Hossfeld T, Tran G P. A generic quantitative relationship between quality of experience and quality of service. IEEE Netw, 2010, 24: 36–41
27 Zhao S, Jiang H, Cai Q, et al. Hybrid framework for no-reference video quality indication over LTE networks. In: Proceedings of Wireless and Optical Communication Conference, Newark, 2014. 1–5
28 Ghalut T, Larijani H, Shahrabi A. Content-Based video quality prediction using random neural networks for video streaming over LTE networks. In: Proceedings of IEEE International Conference on Computer and Information Technology, Liverpool, 2015. 1626–1631
29 Zhang Z S, Long K P, Wang J P, et al. On swarm intelligence inspired self-organized networking: its bionic mechanisms, designing principles and optimization approaches. IEEE Commun Surv Tut, 2014, 16: 513–537
30 3GPP. Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; stage 2. TS 36.300. http://www.3gpp.org/DynaReport/36300.htm. 2014
31 3GPP. Evolved universal terrestrial radio access network (E-UTRAN); architecture description. TS 36.401. http://www.3gpp.org/DynaReport/36401.htm. 2014
32 3GPP. Evolved universal terrestrial radio access network (E-UTRAN); S1 general aspects and principles. TS 36.410. http://www.3gpp.org/DynaReport/36410.htm. 2014
33 Zhang Z S, Long K P, Wang J P. Self-organization paradigms and optimization approaches for cognitive radio technologies: a survey. IEEE Wirel Commun, 2013, 20: 36–42
34 Xing C W, Gao F F, Zhou Y Q. A framework for transceiver designs for multi-hop communications with covariance shaping constraints. IEEE Trans Signal Process, 2015, 63: 3930–3945
35 Apache Software Foundation. Apache HTTP Server. Version 2.2.21. http://httpd.apache.org/. 2015
36 Fei Z S, Ding H C, Xing C W, et al. Performance analysis for range expansion in heterogeneous networks. Sci China Inf Sci, 2014, 57: 082305
37 Xing C W, Ma S D, Fei Z S. A general robust linear transceiver design for multi-hop amplify-and-forward MIMO relaying systems. IEEE Trans Signal Process, 2013, 61: 1196–1209
38 Blender Foundation. Watch Sintel Online. https://durian.blender.org/download. 2015
39 Rimac D S, Vranjes M, Zagar D. Influence of temporal pooling method on the objective video quality evaluation. In: Proceedings of International Symposium on Broadband Multimedia Systems and Broadcasting, Bilbao, 2009. 1–5