Journal of Computers Vol. 30 No. 5, 2019, pp. 172-184
doi:10.3966/199115992019103005013

# Network Video Quality Assessment Method Based on Artificial Learning Method

Zhiming Shi[1,2*], Chengti Huang[1,2]

[1] College of Engineering, Huaqiao University, Quanzhou Fujian, China

[2] Fujian Provincial Academic Engineering Research Centre in Industrial Intellectual Techniques and Systems, Fujian, China
szm@hqu.edu.cn

Received 24 December 2017; Revised 2 April 2018; Accepted 2 June 2018

Abstract. Now, people watch network video in various ways, especially mobile terminals. The service of network video will become the busiest network business. However the quality of network video will be impaired by various factors. So it is very important to monitor the video quality in real time and ensure the service. Many existing objective methods are designed for the specific video distortion, which don't have extensive applicability. Other disadvantage is that the spatial and temporal parameters are not considered simultaneously. Objective methods still need a lot of research to do. In this paper, the artificial learning method is used in the network video quality assessment. It can adjust the objective assessment model, according to the actual network environment. Network and video parameters are comprehensively considered in this method. They are media delivery index (*MDI*), Noise standard deviation ($N_{sd}$), Blur degree ($B_d$), and Block effect ($B_e$). Every parameter has relationship with the video quality. The contributions of this paper are: (1) The *M5*' model tree is used to train the network parameters and video parameters. It is an innovation in this domain. (2) Spatial and temporal parameters are considered simultaneously. All the extracted parameters have relationship with the visual perception. (3) The proposed method can improve the accuracy of objective score. In order to validate the proposed method, six videos of different bit-rate are tested under different experimental environment. Firstly 23 people are arranged to watch the network video and give the subjective score. Meanwhile the network and video parameters are extracted from the video. Secondly the *M5*' model tree is used to model the objective method and train the parameters. Next the objective scores are given and the similarity between the subjective and objective can be computed. Other existing methods are compared with the proposed method. All experimental results show that the proposed method has higher similarity with the subjective assessment. It improves the accuracy of objective score.

Keywords: artificial learning method, *M5*' model tree, *MDI*, similarity, video parameter, video quality

## 1 Introduction

Nowadays, wireless systems are replacing wire-line systems rapidly. New-generation encoders with tremendously improved compression efficiency are being standardized. In this environment, the service of network video is increasing rapidly [1-2]. People watch network video from different terminal, especially the mobile phone. But the quality of video quality may be impaired by various factors, such as network parameters, encoded parameters and video content [3]. Some factors that impair the quality of experience (*QoE*) are listed below [4-5]: (1) Every person has individual interest, such as favorite program. It determines the level and focus of attention. (2) The performance of display terminal, such as size, resolution. (3) Network environment of video. (4) The properties of network video, such as blur,

* Corresponding Author

block. As the wide variety and subjectivity of these factors, the measurement of network video quality is a complex work.

Low video quality may reduce the *QoE* [6-7]. In order to ensure the *QoE*, it needs to predict and monitor the video quality. Unfortunately, the quality of network video is a rather ill-defined concept. International telecommunications union telecommunication standardization sector (*ITU-T)* initially defines quality of service (*QoS*) as “user satisfaction and the service performance of the integrated effect”. This definition is consistent with *QoE*. But *QoS* only measures the network parameters, which can’t reflect the network video quality [8-9]. With the development of network video, people want to survey mark to reflect the network video quality [10]. So the *QoE* is proposed to define the user experience. Now a lot of research institution and scholars have done research on this field [11-12]. *VQEG* (Video Quality Expert Group) is a professional institute which does research on video quality assessment. It has tested many assessment methods, but there are many problems to be solved.

Commonly, the video quality assessment methods are classified into subjective and objective assessment method [13-14]. Subjective video quality assessment methods measure the video quality by the Human Visual System (*HVS*). They are crucial for evaluating the performance of objective quality assessment method. The mean opinion score (*MOS*) or differential mean opinion score (*DMOS*) are obtained while people watch the network video. *VQEG* and *ITU-T* provides the detailed test plans of subjective assessment method. But subjective methods have several drawbacks, which make them impractical for real-time use. So people develop many objective assessment methods. The objective video quality assessment can predict the video quality automatically. According to the related reference information, objective methods are divided into three classes. Full-reference (*FR*) methods need the original videos to be compared with the impaired videos. Reduced-reference (*RR*) methods only need partially related videos. No-reference (*NR*) methods assess the video quality without any related videos. Many researchers have focused on *FR* and *RR* methods, and related methods have been proposed in the past. The traditional objective methods have peak-signal-to-noise ratio (*PSNR*) and structural similarity index (*SSIM*). Other methods have video quality metric (*VQM*) and motion-based video integrity evaluation (*MOVIE*). Although these methods may have accurate objective score, it is difficult to obtain reference video [15-16]. So in this condition, the *NR* methods are needed. *NR* methods are more difficult than *FR* and *RR* methods. *NR* methods still need to do a lot of work. Now *NR* methods tend to consider the effects of distortions caused by the encoder or network impairments. A discrete cosine transform (DCT)-based no-reference video quality prediction model is proposed that measures artifacts and analyzes the statistics of compressed natural videos [17-18] proposes a novel estimation method of the quantization in H.264/AVC videos without bit-stream access, which can also be used for peak signal to noise ration estimation. But videos are complicated and have more dimensional information. The video quality may be impaired by various spatial and temporal artifacts at the same time. So many *NR* methods extract more spatiotemporal features of videos to improve the accuracy. [19] proposes video quality assessment method based on fuzzy interface system, it considers different features of video. [20] considers both spatial and temporal information of video to model the objective method.

These methods may consider network parameters or video parameters, and ignore other impair factors. They are not comprehensive. All these methods have fixed model based on specific scenarios. The application of them is not universally. So there are many problems in the research of *NR* methods. In order to change this situation, this paper proposes an objective assessment model based on artificial learning method. As shown in Fig. 1, it considers both network parameters and spatiotemporal features of videos to model the assessment method. There are five quality parameters [Media delivery index (*MDI*): Delay factor (*DF*), Media loss rate (*MLR*), Noise standard deviation ($N_{sd}$), Blur degree ($B_d$), and Block effect ($B_e$)]. Later, *M5*’ model tree are selected as the artificial learning method. All the parameters are trained by it to model the objective assessment method. In order to validate the proposed method, six videos are tested under different network bandwidth. Other existing methods are compared with the proposed method. The experimental results show that this method can improve the accuracy of objective method.

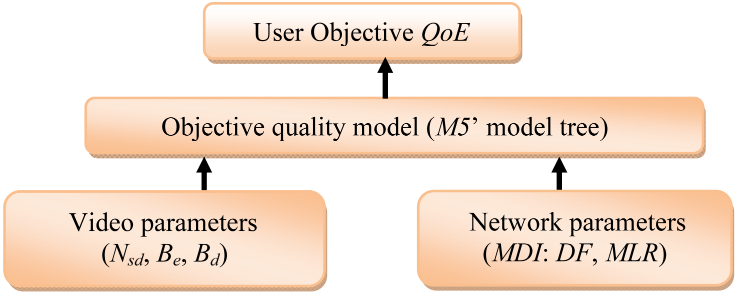


**Fig. 1.** The objective quality method

As shown in Table 1, the proposed method is different from other methods and has many advantages: (1) Comprehensive parameters are considered, including network and video parameters. All parameters have relationship with the visual perception. (2) In order to innovate the objective method, the *M5*' model tree is used. The proposed method can be adjusted, according to the actual experimental results. It is a flexible method 3. This method improves the similarity between the subjective and objective score.

**Table 1.** The advantage of method

| Method | Advantage 1 | Advantage 2 |
|---|---|---|
| Proposed method | Comprehensive | Flexible |
| Other methods | Limited | Fixed |

The contributions of this paper are as follows: (1) A flexible method is proposed to change the fixed model. (2) The *M5*' model tree is used to assess the video quality. (3) Many experiments are presented to explain the effectiveness of this method. The remainder of the paper is organized as follows: In section II, the parameters are introduced. The experiments are shown in section III. The proposed method based on *M5*' model tree is introduced in section IV. In section V, the experimental results are given and the similarity between objective and subjective is analyzed. Other methods are compared with the proposed method. This paper is concluded in section VI.

## 2 The Quality Parameters

Feature selection is the key of artificial intelligence learning. It relates to the video quality mostly and improves the performance of objective method. Many intelligence learning methods may be trained to estimate *MOS* based on the selected features. In this paper, the network parameters and video parameters are comprehensively considered. The network parameter is *MDI*. *MDI* is proposed by *CISCO* and *IneoQuest*. It is one of important index to assessment the video quality. Some packet loss rate and delay distortions have bad impact on the perceived quality. The video parameters have $N_{sd}$, $B_d$, $B_e$. They can impact the visual perception. Especially block, noise and blur are the most annoying distortions in the network video. All the parameters will be trained to model the objective method.

### 2.1 MDI

*MDI* is consisted of delay factor (*DF*) and media loss rate (*MLR*). The value of *DF* reflects the delay and jitter of the network video. *DF* can convert video stream jitter to the need of video transmission and decode buffer. The larger jitter of measured video stream mainly has bigger *DF* value. When the buffer of network device and decoder is less than the *DF*, the quality of video will reduce. Because network nodes need to allocate buffers no less than *DF* to smooth video stream jitter. So the max value of *DF* is the minimum delay of video content passing through the network nodes. The detail function of *DF* is (1)-(3).

$$VB(i, pre) = sum(S_j) - MR \times T_i \; j = 1 \ldots i-1 \; . \quad \textbf{(1)}$$

$$VB(i, post) = VB(i, pre) + S_i \; . \quad \textbf{(2)}$$

$$DF = [VB(Max) - VB(Min)] / MR \ . \quad \textbf{(3)}$$

*VB* represents the virtual buffer. When a packet arrives, there are two *VB* values: *VB(i,pre)* and *VB(i,post)*. In a measurement interval, there exists 2i+1 *VB*. Then the max *VB* minus the min VB and divided by media rate (*MR*) to get *DF*.

*MLR* represents the packet loss rate of network video in a time interval. The lost packet will cause the video quality impairment, including visual distortion and anomalies. The detail equation of *MLR* is (4). The expected packet minus the received packet and divided by the sample interval to get *MLR*.

$$MLR = \frac{P\_expected \text{ - } P\_received}{sample\ interval} \ . \quad \textbf{(4)}$$

### 2.2 Video Parameters

Table 2 lists the main video feature parameters. In this paper, blur, block and noise are chosen as the video parameters. According to the previous paper, the three parameters have greatly relationship with the *MOS*.

**Table 2.** Video feature parameters

| No | Feature |
|---|---|
| 1 | Blur |
| 2 | Blockiness |
| 3 | Noise |
| 4 | Contrast |
| 5 | Ringing |
| 6 | Edge activity |
| 7 | Z score |
| 8 | Gradient activity |

Blur degree can reflect the details of image changes. But videos are constituted of many images. So the temporal and spatial variation of blur is considered as the video quality parameter. The detail equation of blur is given in [21]. The vertical difference between *y(f,w,h+1)* and *y(f,w,h)* is calculated as follows:

$$d_h(f,w,h) = y(f,w,h+1) - y(f,w,h) \ . \quad \textbf{(5)}$$

$y(f,w,h)$ is the luminance value at position (*w,h*) in the *fth* frame. $b_h(f,w,h)$ is defined as follows.

$$b_h(f,w,h) = \begin{cases} 1, if\ d_h(f,w,h+1) \cdot d_h(f,w,h) < 0 \\ 0, else \end{cases} \ . \quad \textbf{(6)}$$

$Z_h$ is calculated for average vertical pixels. In the same way, the average horizontal pixels is calculated to get $Z_w$

$$Z_h = \frac{1}{N \cdot W(H-2)} \sum_{f=1}^{N} \sum_{w=1}^{W} \sum_{h=1}^{H-2} b_h(f,w,h) \ . \quad \textbf{(7)}$$

$$Z = \frac{Z_h + Z_w}{2} \ . \quad \textbf{(8)}$$

*N* is the frame of video, *W* is the number of vertical pixels and *H* is the number of horizontal pixels.

The blockiness can reflect the discontinuity of two adjacent blocks. [22] provides the equation of the block. Firstly, the difference of horizontal matrix is defined by $D_h = \{d_{i,0}, d_{i,1}, \ldots d_{i,N-1}\}$ .

$$d_{ij} = \left| f(i,j) - f(i,j+1) \right| \ . \quad \textbf{(9)}$$

*f(i,j)* represents the luminance of pixel at *ith* row and *jth* column. *N* is the width of image. Commonly, 8×8 block is used in video coding. So a one-dimension vector $A=[a_0,a_1...a_7]^T$ is defined. Because the periodic of block artifacts, $a_k$ is calculated by $a_k=\sum_{x=0}^{M-1}\sum_{y=0}^{N/8-1} d_{x,8y+k}C_{x,8y+k}$ . *M* is the height of image.

$$C_{x,8y+k}=C_{x,8y+k}{}^{(1)}\ \&\ C_{x,8y+k}{}^{(2)}\ \&\ C_{x,8y+k}{}^{(3)}\ . \quad \textbf{(10)}$$

$C_{x,8y+k}$ is a weight value. If $d_{x,8y+k}$ meets the three constrains which are denoted by $C_{x,8y+k}{}^{(1)},C_{x,8y+k}{}^{(2)}$ *and* $C_{x,8y+k}{}^{(3)}$, the $C_{x,8y+k}$ is set to 1, otherwise is set to 0.

$$C_{x,8y+k}{}^{(1)}=\{\begin{matrix}1,S\geq T_2\\0,others\end{matrix}\ . \quad \textbf{(11)}$$

$$S=\sum_{j=-2}^{2}\phi_j,\ \ \phi_j=\{\begin{matrix}1,d_{x,8y+k}\leq T_1\\0,others\end{matrix}\ . \quad \textbf{(12)}$$

*S* can be calculated by examining local content activities of the related pixel.

$$C_{x,8y+k}{}^{(2)}=\{\begin{matrix}1,d_{x,8y+k}\leq \max(d_{x,8y+k},d_{x,8y+k-1})\\0,others\end{matrix}\ . \quad \textbf{(13)}$$

$$C_{x,8y+k}{}^{(3)}=\{\begin{matrix}1,d_{x,8y+k}\leq T_3\\0,others\end{matrix}\ . \quad \textbf{(14)}$$

$T_1$, $T_2$ and $T_3$ are the threshold. The horizontal blocking artifacts $M_h$ can be calculated.

$$M_h=\frac{m_h}{\mu_h+\sigma_h}\ . \quad \textbf{(15)}$$

$m_h$ is the max element of *A*, $\mu_h$ is the mean of elements except $m_h$ of *A*, $\sigma_h$ is the standard derivation of elements except $m_h$ of *A*. The vertical blocking artifacts $M_v$ can be similarly defined. Finally the blocking artifact is given. Commonly the larger Metric, the image has heavier blocking artifacts. In this paper, the threshold values of $T_1$=5, $T_2$=4, $T_3$=30, and *a*=1 is used. All the blocking artifacts of video are averaged.

$$Metric=aM_h+(1-a)M_v\ . \quad \textbf{(16)}$$

The noise can impair the video quality [23]. It is a key index. Firstly the noise is filtered respectively from the horizontal and vertical.

$$y_v=\frac{1}{\sqrt{2}}(y(m+1,n)-y(m,n))\ .$$
$$y_h=\frac{1}{\sqrt{2}}(y_1(m,n+1)-y_1(m,n))\ . \quad \textbf{(17)}$$

$y(m,n)$ is the gray value at position (*m, n*) in a frame. Then the image is divided into 8×8 block, and computed the average standard deviation of every block.

$$SD=\sqrt{\frac{\sum_{k=1}^{M}\sum_{l=1}^{N}D}{M\times N}}\ . \quad \textbf{(18)}$$

*D* represents the average standard deviation of every block. $M\times N$ represents the number of block. At last the noise of every image is averaged.

## 3 Experimental Environment

In order to validate the proposed method, six videos of different bit-rate are tested under different experimental environment. Firstly the experimental environment is shown in Fig. 2.The video server connects the switch through the Ethernet port. The switch can control the network bandwidth and transmit the video to client personal computer (*PC*) through the wireless access point (AP) in wireless way. Meanwhile the *MDI* is programmed and the frame grabbing software is installed. So the client *PC* can monitor the *MDI* and analyze the video features.

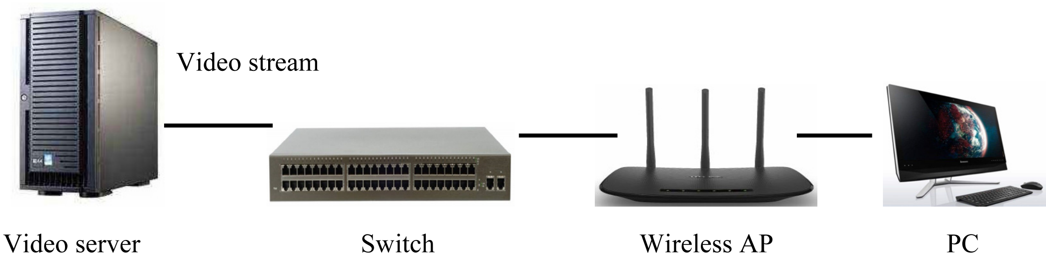


**Fig. 2.** The experimental environment

Six different video sequences with different resolution and bit rate are tested under different network bandwidth. In the Fig. 3, rocket represents standard video, concert represents high definition video, else represents the ultra high definition video. They all have different scene and the detail information is shown in Table 3.

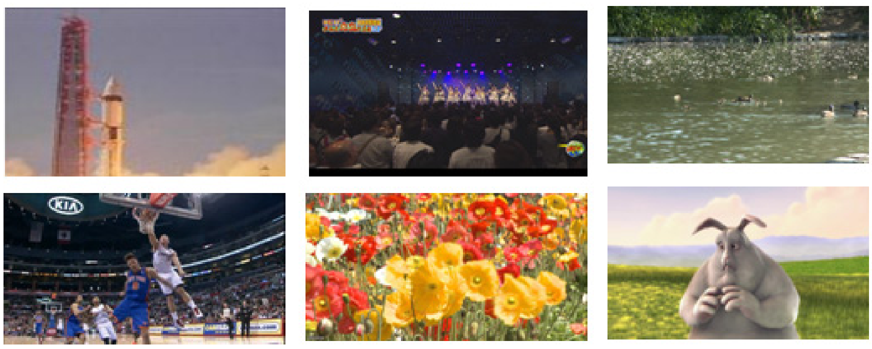

**Fig. 3.** The videos

**Table 3.** The information of test videos

| No | Video | Resolution | Average bit rate(Mbps) |
|---|---|---|---|
| 1 | Rocket | 640×480 | 3.75 |
| 2 | Concert | 1280×720 | 13.8 |
| 3 | Duck | 1920×1080 | 20 |
| 4 | Basketball | 1920×1080 | 20 |
| 5 | Flower | 1920×1080 | 20 |
| 6 | Tale | 1920×1080 | 30 |

In the next stage, the switch is configured to adjust the bandwidth, simulating the shortage of network bandwidth. Every video is tested 10 times under different bandwidth. The network bandwidth of every video is set in Table 4.

**Table 4.** The results under different network bandwidth

| Rocket | | | Concert | | | Duck | | |
|---|---|---|---|---|---|---|---|---|
| Bandwidth Mbps | Damage of video | MOS | Bandwidth Mbps | Damage of video | MOS | Bandwidth Mbps | Damage of video | MOS |
| 4 | Clear and fluent | 5 | 14 | Clear and fluent | 5 | 20 | Clear and fluent | 5 |
| 3.75 | Clear and fluent | 5 | 13.5 | Clear and fluent | 5 | 19 | Clear and fluent | 4.79 |
| 3.5 | Occasional pause | 4.5 | 13 | Occasional pause | 4.6 | 18 | Occasional pause | 4.42 |
| 3.25 | 2-4 pause | 3.14 | 12.5 | 1-3 pause | 3.34 | 17 | 3 pause | 3.01 |
| 3.0 | Mosaic, 3-7 pause | 2.78 | 12 | Mosaic, more than 5 pause | 3.01 | 16 | Mosaic, more than 5 pause | 2.66 |
| 2.75 | Blur image, more than 10 pause | 2.51 | 11.5 | Blur image, more than 10 pause | 2.83 | 15 | Blur image, more than 10 pause | 2.41 |
| 2.5 | Blur image, 10-20 pause | 2.24 | 11 | Blur image, 15 pause | 2.43 | 14 | Blur image, more than 10 pause | 2.0 |
| 2.25 | Influence viewing | 2.0 | 10.5 | Serious blur | 2.1 | 13 | Serious blur | 1.81 |
| 2 | Influence viewing | 1.5 | 10 | Influence viewing | 1.8 | 12 | Influence viewing | 1.0 |
| 1.75 | Unable to watch | 1.0 | 9.5 | Unable to watch | 1.0 | 11 | Unable to watch | 1.0 |
| Basketball | | | Flower | | | Tale | | |
| Bandwidth Mbps | Damage of video | MOS | Bandwidth Mbps | Damage of video | MOS | Bandwidth Mbps | Damage of video | MOS |
| 20 | Clear and fluent | 5 | 20 | Clear and fluent | 5 | 30 | Clear and fluent | 5 |
| 19 | Clear and fluent | 4.6 | 19 | Clear and fluent | 5 | 28 | Clear and fluent | 4.8 |
| 18 | Occasional pause | 4.0 | 18 | Few pause time | 4.8 | 26 | Few pause time | 4.3 |
| 17 | More than 3 pause | 2.8 | 17 | 1-3 pause | 4.0 | 24 | Occasional pause | 3.7 |
| 16 | Mosaic, more than 5 pause | 2.43 | 16 | Mosaic, 1-5 pause | 3.6 | 22 | Mosaic, 2-5 pause | 3.3 |
| 15 | Blur image, more than 10 pause | 2. 1 | 15 | Blur image, more than 10 pause | 3.0 | 20 | Blur image, 5-10 pause | 2.6 |
| 14 | Blur image, more than 20 pause | 1.5 | 14 | Blur image, more than 10 pause | 2.6 | 18 | Blur image, more than 10 pause | 2.0 |
| 13 | Serious blur | 1.2 | 13 | Serious blur | 2.2 | 16 | Serious blur | 1.4 |
| 12 | Influence viewing | 1.0 | 12 | Influence viewing | 1.8 | 14 | Influence viewing | 1.0 |
| 11 | Unable to watch | 1.0 | 11 | Unable to watch | 1.0 | 12 | Unable to watch | 1.0 |

According to *ITU-R BT*.1788, 23 non professional viewers take part in the experiment and give the subjective assessment score. As shown in Fig. 4, the given score is from 1 to 5. All the raw should be normalized before analysis. At last the average score is *MOS*. With the decrease of bandwidth, the quality of network video is impaired. The mosaic phenomenon appears in the video. The image of video becomes blurred and the number of pause increases. When the bandwidth maintains about 80% average bit rate (average bit rate×0.8), the video quality is better. When the bandwidth decreases to 80%-50% average bit rate, the video quality reduces significantly. When the bandwidth decreases to 50% average bit rate, the video can’t watch. On the other hand, the higher resolution video has stronger anti damage ability. If the video has more violent picture, the video impairs much, such as “Basketball”.


**Fig. 4.** Rating scale used to assess subjective quality

Meanwhile the *MDI* and video parameters are measured under different experimental environment. The *MDI* is extracted by preset program. The video parameters are extracted by the above calculations. These parameters have been averaged under different network bandwidth. The *MDI* is shown in Fig. 5.

When the bandwidth decreases, the *MDI* increases. The video quality has also changed. Fig. 6 shows the video parameters. It can be seen that the parameters increase when network bandwidth decreases. The increase of parameters will impair the video quality. There exists a relationship between these parameters and *MOS*. Next the *M5*' model tree is used to model the relationship between the six parameters and *MOS*.

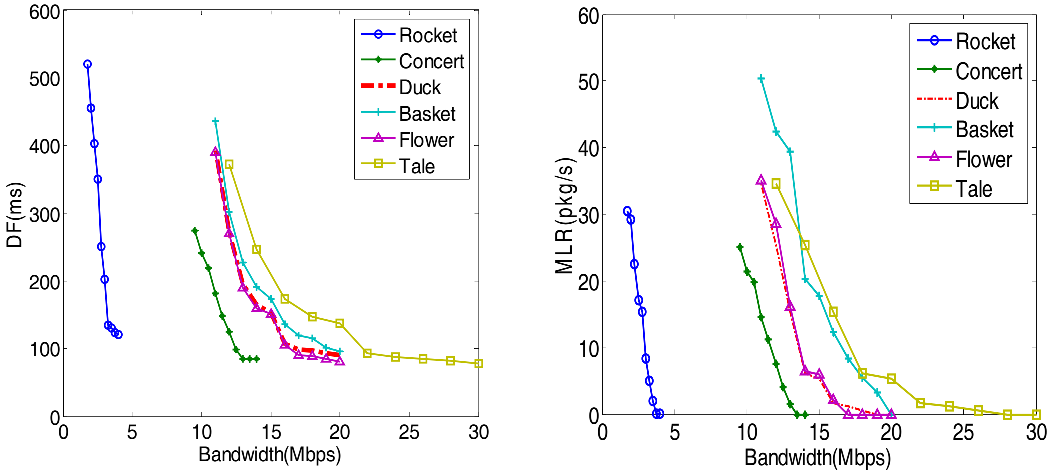


**Fig. 5.** The MDI

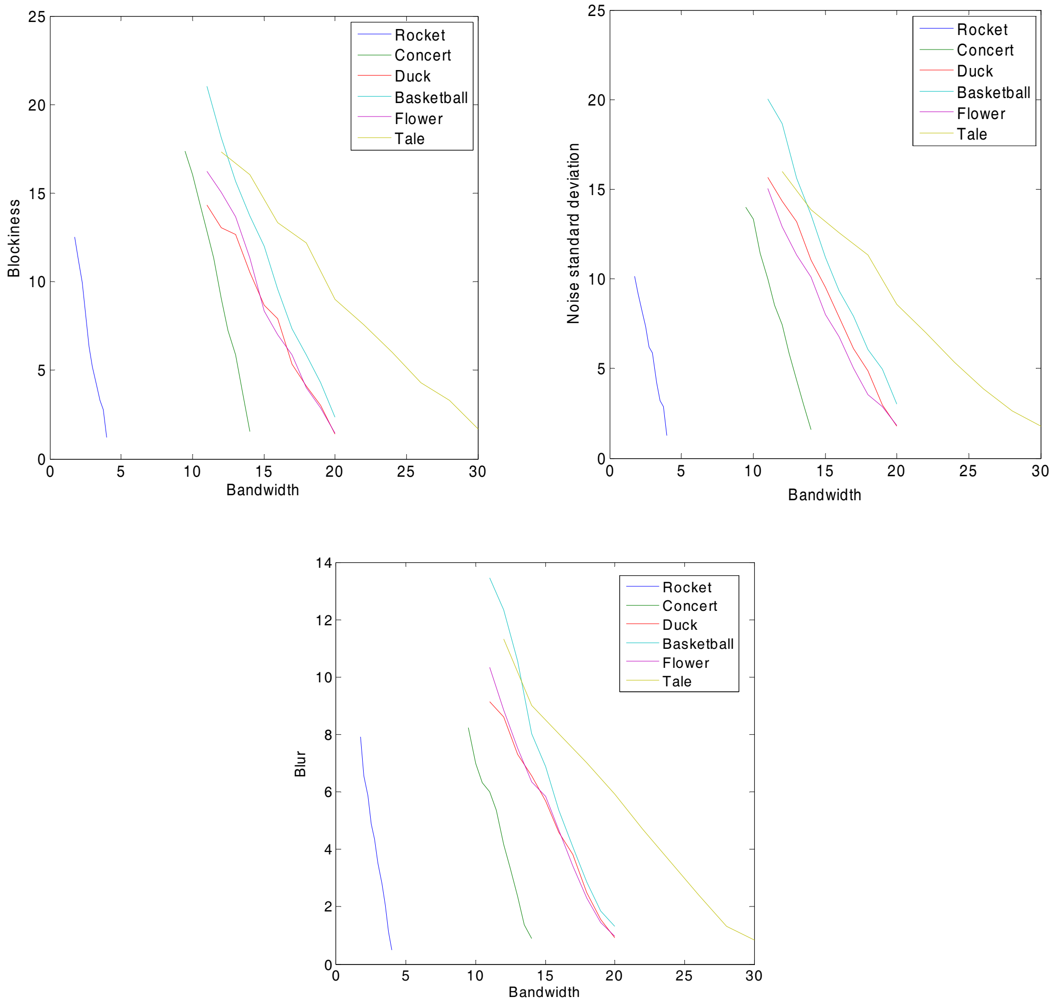


**Fig. 6.** The video parameters

## 4 The Proposed Method

*M5*' model tree is an artificial learning method. It can divide the complex problem into a number of simple tasks and combine the tasks. As shown in Fig. 7. It has two steps. The first step is to train the samples and build the decision tree. The second step is the pruning of decision tree [24-25]. In this paper, the *M5'* model tree is used to model the objective assessment method. The detail steps are as follows.

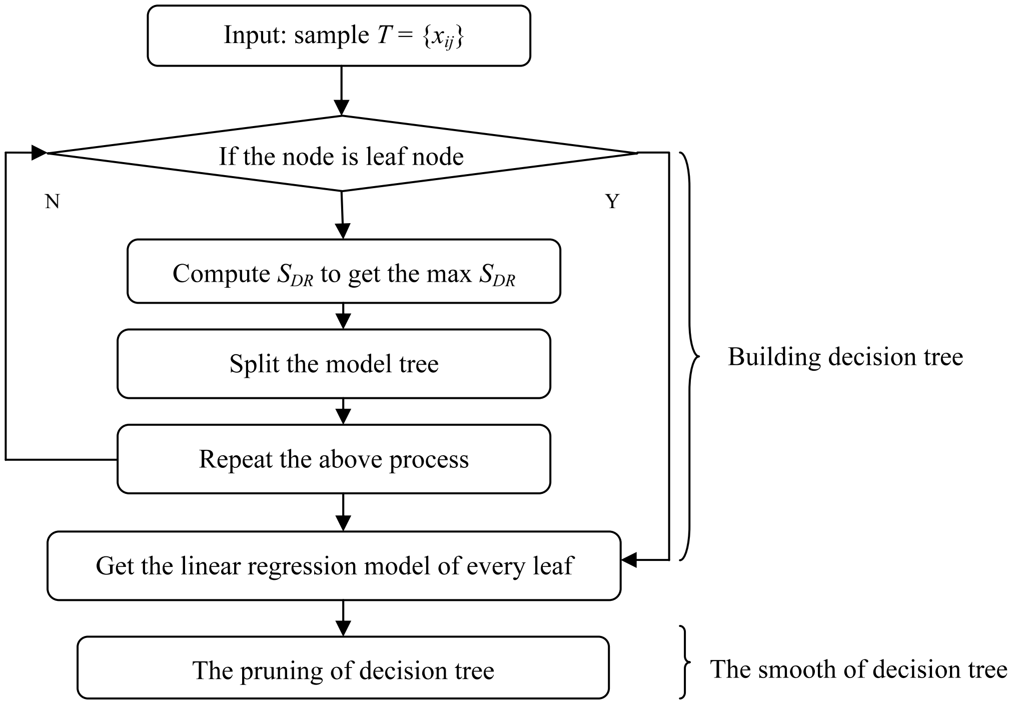


**Fig. 7.** The process of *M5'* model tree

The trained samples are the actual test parameters, including *MDI* and video parameters. The set *T* is constituted by them.

$$T = \begin{array}{c} \begin{array}{ccccc} T_1 & T_2 & T_3 & T_4 & T_5 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{array} \\ \begin{bmatrix} DF & MLR & N_{sd} & B_d & B_e \\ a_1 & b_1 & c_1 & d_1 & e_1 \\ . & . & . & . & . \\ a_{10} & b_{10} & c_{10} & d_{10} & e_{10} \end{bmatrix} \end{array}. \quad \textbf{(19)}$$

The input samples are divided into a number of subspaces or regions based on the $S_{DR}$. The formula of $S_{DR}$ is:

$$S_{DR} = sd(T) - \sum_i \frac{|T_i|}{|T|} sd(T_i) \ . \quad \textbf{(20)}$$

*T* represents the set of samples; $T_i$ is the *ith* subset of sample; *sd* is the standard deviation of sample. After examining all $S_{DR}$, *M5*' chooses the max one as the split node. When the class values of all instances reach a leaf node, the *M5*' split ceases.

However, this step may produce needless structures. So it is need to be pruned back. The second step is the pruning of decision tree. It is performed to compensate for the sharp discontinuities that will inevitably occur between adjacent linear models at the leaves of the pruned tree, particularly for some models constructed from a smaller number of training examples. In smoothing, the adjacent linear equations are updated in such a way that the predicted outputs for the neighbouring input vectors corresponding to the different equations are becoming close in value. As shown in Fig. 8, the model tree is built finally. Next the leaf node will have a linear regression to give the objective assessment model.

$$objective\ score = \alpha MLR + \beta DF + \gamma N_{sd} + \xi B_e + \psi B_d \,. \tag{21}$$

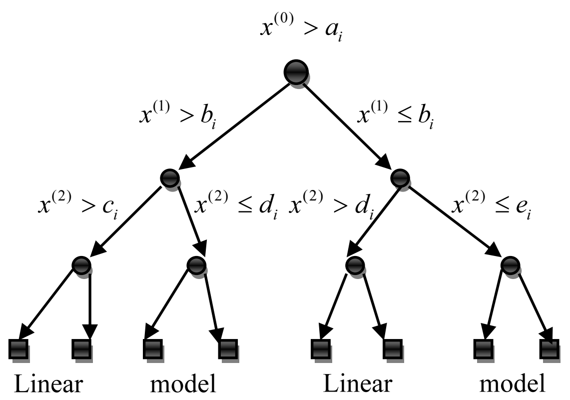


**Fig. 8.** The decision tree

## 5 Result Analyses

The steps of simulation are as follows: (1) According to the above process, six videos are tested to extract the parameters. (2) Then the *M5'* model tree is used to model the objective assessment method and give the objective score. During the training, the *M5'* is required to produce a tree with different leaves covering no less than 10 instances, corresponding to the different network bandwidth. (3) So every leaf node will have a linear regression to model the objective method. The objective score is consisted of five parameters. Meanwhile other related methods are used to compare with this method. They are PSNR, video signal to noise ratio (VSNR), visual information fidelity (VIF), SSIM and multi scale structural similarity index metric (MSSIM).

Table 5 and Table 6 give the correlation coefficient of different methods. It can be seen that the similarity of proposed method is better than other methods. The correlation coefficients are above 0.85 for every video sequence. Fig. 9 shows the subjective and objective *MOS*, the proposed method has a good linear relationship.

**Table 5.** The Spearman of every method

| Algorithm | Rocket | Concert | Duck | Basketball | Flower | Tale | Average |
|---|---|---|---|---|---|---|---|
| VSNR | 0.6421 | 0.6532 | 0.6678 | 0.6132 | 0.6235 | 0.6322 | 0.6387 |
| PSNR | 0.6132 | 0.5918 | 0.6231 | 0.5897 | 0.6214 | 0.6138 | 0.6088 |
| VIF | 0.6713 | 0.6214 | 0.6513 | 0.6028 | 0.6413 | 0.6415 | 0.6383 |
| SSIM | 0.7013 | 0.6987 | 0.7132 | 0.6883 | 0.7138 | 0.7242 | 0.7066 |
| MSSIM | 0.7921 | 0.7763 | 0.7814 | 0.7534 | 0.7832 | 0.7752 | 0.7769 |
| **M5' method** | 0.8953 | 0.8792 | 0.9131 | 0.8713 | 0.8821 | 0.8679 | 0.8848 |

**Table 6.** The Pearson of every method

| Algorithm | Rocket | Concert | Duck | Basketball | Flower | Tale | Average |
|---|---|---|---|---|---|---|---|
| VSNR | 0.6514 | 0.6432 | 0.6713 | 0.6089 | 0.6315 | 0.6389 | 0.6409 |
| PSNR | 0.6214 | 0.5897 | 0.6138 | 0.5918 | 0.6251 | 0.6205 | 0.6104 |
| VIF | 0.6821 | 0.6201 | 0.6618 | 0.6131 | 0.6216 | 0.6533 | 0.6420 |
| SSIM | 0.7156 | 0.6914 | 0.7103 | 0.6814 | 0.7214 | 0.7313 | 0.7086 |
| MSSIM | 0.8013 | 0.7815 | 0.7956 | 0.7713 | 0.7915 | 0.7959 | 0.7895 |
| M5' method | 0.9012 | 0.8915 | 0.9213 | 0.8823 | 0.8913 | 0.8724 | 0.8933 |

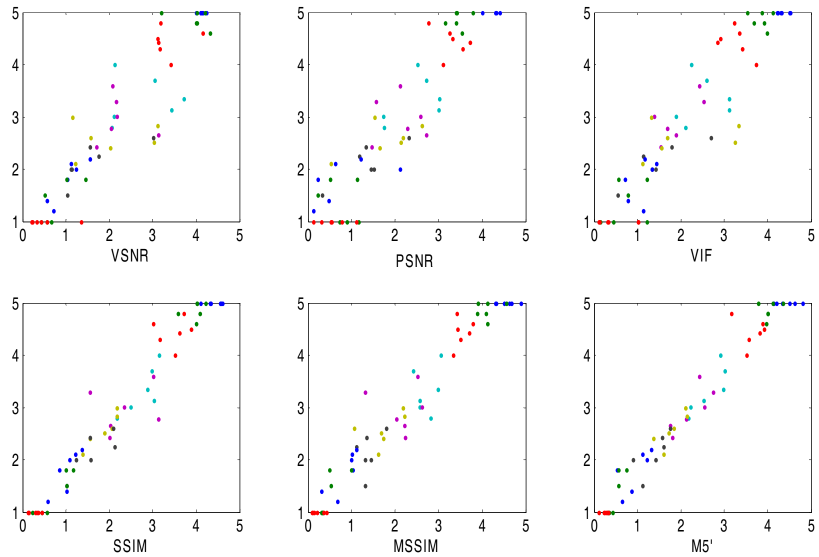


**Fig. 9.** The similarity between the subjective and objective *MOS*

Fig. 10 shows the subjective and objective scores of these videos. It can be seen that the objective score is close to the subjective value. The proposed method can improve the similarity between the subjective and objective assessment. It can give more accurate objective score than other methods. Because the proposed method takes into account different parameters and trains them to assess the objective score. This method can adjust the assessment model according to the real situation. The method is a widely applicable and practical method.

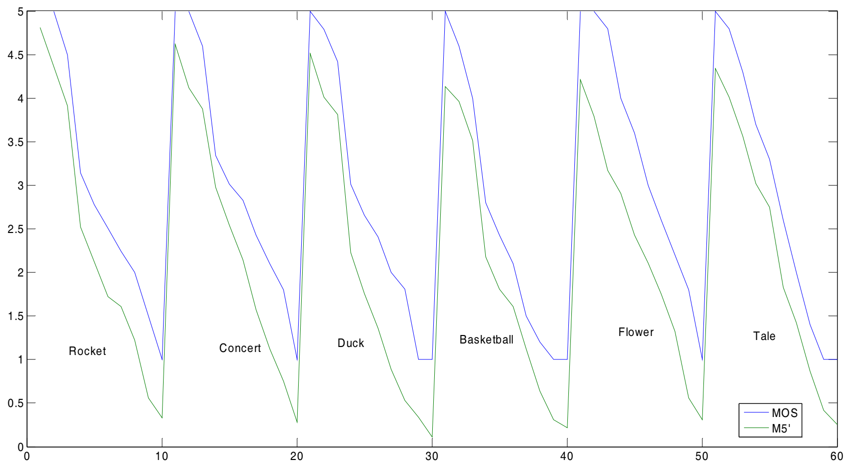


**Fig. 10.** *M5'* and *MOS*

## 6 Conclusion

This paper proposes an objective assessment model based on artificial learning method. The method considers both network parameters and spatiotemporal features of videos, including *MDI*, $N_{sd}$, $B_d$ and $B_e$. Next, *M5*' model tree is used to train the parameters and model the objective assessment method. Six video are tested under different network bandwidth to validate the proposed method. Meanwhile other existing methods are compared with the proposed method. The experimental results show that this method can improve the similarity between subjective and objective scores.

The application of *M5*' model tree in this domain is an innovation. There are many advantages in this method: 1. Many impair parameters are considered in this model, including network and video parameters. These parameters can impair the video quality directly. 2. The proposed method can adjust objective assessment model, according to the actual experimental results. In the future work, more parameters will be considered in this model to get more accuracy objective score.

## Acknowledgements

This work in supported by the High Level Talent Research Project in Huaqiao University (605-50Y14041) and Quanzhou City Science & Technology Program of China under Grant (No. 2018C106R).

## References

[1] Y.-J. Chen, K.-S. Wu, Q. Zhang, From QoS to QoE: a tutorial on video quality assessment, IEEE Communications Surveys & Tutorials 17(2)(2015) 1126-1165.

[2] W. Zhang, H.-T.Liu, Study of saliency in objective video quality assessment, IEEE Transactions on Image Processing 26(3)(2017) 1275-1288.

[3] L. Qian, Z.-X. Cheng, Z. Fang, L.-H. Ding, F. Yang, W. Huang, A QoE-driven encoder adaptation scheme for multi-user video streaming in wireless networks, IEEE Transactions on Broadcast 63(1)(2017) 20-31.

[4] T.-S. Zhao, Q. Liu, C.-W. Chen, QoE in video transmission: a user experience-driven strategy, IEEE Communication Surveys& Tutorials 19(1)(2016) 285-302.

[5] D. Ghadiyaram, J. Pan, A.C. Bovik, Learning a continuous-time streaming video QoE model, IEEE Transactions on Image Processing 27(5)(2018) 2257-2271.

[6] H. Jeong, J. Yang, M. Song, Video quality adaptation to limit energy usage in mobile systems, IEEE Transactions on Consumer Electronics 62(3)(2016) 301-309.

[7] C.-M. Lentisco, L. Bellido, J.C. Cuellar Q, E. Pastor, J. Arciniegas H, QoE based analysis of DASH streaming parameters over mobile broadcast networks, IEEE Access 5(2017) 20684-20694.

[8] J. Radun, M. Nuutinen, J.-N. Antons, S. Arndt, Did you notice it?- how can we predict the subjective detection of video quality changes from eye movements, IEEE Journal of Selected Topics in Signal Processing 11(1)(2017) 37-47.

[9] R.-D. Fréin, Source separation approach to video quality prediction in computer networks, IEEE Communications Letters 20(7)(2016) 1333-1336.

[10] F. Zhang, D.-R. Bull, A perception-based hybrid model for video quality assessment, IEEE Transactions on Circuits and Systems for Video Technology 26(6)(2016) 1017-1028.

[11] S.-D. Hu, L. Jin, H.-L. Wang, Y. Zhang, S. Kwong, C.-C. JayKuo, Objective video quality assessment based on perceptually weighted mean squared error, IEEE Transactions on Circuits and Systems for Video Technology 27(9)(2017) 1844-1855.

[12] M. Alreshoodi, E. Danish, J. Woods, A. Fernando, C.-D. Alwis, Prediction of perceptual quality for mobile video using fuzzy inference systems, IEEE Transactions on Consumer Electronics 61(4)(2015) 546-554.

[13] M.-T. Vega, D.-C. Mocanu, J. Famaey, S. Stavrou, A. Liotta, Deep learning for quality assessment in live video streaming, IEEE Signal Processing Letters 24(6)(2017) 736-740.

[14] A.-M. Demirtas, A.-R. Reibman, H. Jafarkhani, Full-reference video quality estimation for videos with different spatial resolutions, IEEE Transactions on Circuits and Systems for Video Technology 26(11)(2016) 1988-2000.

[15] M.K, S.-S. Channappayya, An optical flow-based full reference video quality assessment algorithm, IEEE Transactions on Image Processing 25(6)(2016) 2480-2492.

[16] A.-M. Demirtas, A.-R. Reibman, H. Jafarkhani, Full-reference video estimation for videos with different spatial resolutions, IEEE Transactions on Circuits and Systems for Video Technology 26(11)(2016) 1988-2000.

[17] K.-F. Zhu, C.-Q. Li, V. Asari, D. Saupe, No reference video quality assessment based on artifact measurement and statistical analysis, IEEE Transactions on Circuits and Systems for Video Technology 25(4)(2015) 533-546.

[18] J. Sogaard, S. Forchhammer, J. Korhonen, No reference video quality assessment using codec analysis, IEEE Transactions on Circuits and Systems for Video Technology 25(10)(2015) 1637-1650.

[19] A.-A. Naji, S.-H. Lee, J. Chahl, Quality index evaluation of videos based on fuzzy interface system, IET Image Processing 11(5)(2017) 292-300.

[20] X.-L. Li, Q. Guo, X.-Q. Lu, Spatiotemporal statistics for video quality assessment, IEEE Transactions on Image Processing 25(7)(2016) 3329-3342.

[21] T. Kawano, K. Yamagishi, K. Watanabe, J. Okamoto, No reference video quality assessment model for video streaming services, in: Proc. 2010 18th International Packet Video Workshop, 2010.

[22] Z.-B. Zhao, J.-S. Yuan, Q. Gao, Y.-H. Kong, Wavelet image de-noising method based on noise standard deviation estimation, in: Proc. 2007 International Conference on Wavelet Analysis and Pattern Recognition, 2007.

[23] X. Zheng, B. Yang, Y.-W. Liu, G.-J. Shi, Blockiness evaluation for reducing blocking artifacts in compressed Images, in: Proc. 2009 Digest of Technical Papers International Conference on Consumer Electronics, 2009.

[24] S. Zou, Y.-L. Tang, Y.-L. Sun, K. Su, An identification decision tree learning model for self-management in virtual radio access network: IDTLM, IEEE Access 6(2018) 504-518.

[25] E.F.D. Oliveira, M.E.D. Lima Tostes, C.A.O.D. Freitas, J.C. Leite, Voltage THD analysis using knowledge discovery in databases with a decision tree classifier, IEEE Access 6(2018) 1177-1188.