Hindawi
Advances in Multimedia
Volume 2017, Article ID 8317590, 19 pages
https://doi.org/10.1155/2017/8317590

Research Article

A No-Reference Modular Video Quality Prediction Model for H.265/HEVC and VP9 Codecs on a Mobile Device

Debajyoti Pal and Vajirasak Vanijja

IP Communications Laboratory, School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

Correspondence should be addressed to Debajyoti Pal; debajyoti.pal@gmail.com

Received 14 September 2017; Revised 25 October 2017; Accepted 6 November 2017; Published 23 November 2017

Academic Editor: Constantine Kotropoulos

We propose a modular no-reference video quality prediction model for videos that are encoded with H.265/HEVC and VP9 codecs and viewed on mobile devices. The impairments which can affect video transmission are classified into two broad types depending upon which layer of the TCP/IP model they originated from. Impairments from the network layer are called the network QoS factors, while those from the application layer are called the application/payload QoS factors. Initially we treat the network and application QoS factors separately and find out the 1 : 1 relationship between the respective QoS factors and the corresponding perceived video quality or QoE. The mapping from the QoS to the QoE domain is based upon a decision variable that gives an optimal performance. Next, across each group we choose multiple QoS factors and find out the QoE for such multifactor impaired videos by using an additive, multiplicative, and regressive approach. We refer to these as the integrated network and application QoE, respectively. At the end, we use a multiple regression approach to combine the network and application QoE for building the final model. We also use an Artificial Neural Network approach for building the model and compare its performance with the regressive approach.

1. Introduction

There has been a rapid advance in various video services and their applications, like video telephony, High Definition (HD) and Ultra High Definition (UHD) television, Internet protocol television (IPTV), and mobile multimedia streaming in recent years. Thus, quality assessment of videos that are being streamed and watched online has become an area of active research. As per a report published in [1], video streaming over the Internet is becoming increasingly popular on devices having small form factor screens (4 inches to 6 inches). Most of the mobile phones as of today have a screen resolution of at least 1280 × 720 pixels (HD) or 1920 × 1080 pixels (Full HD). Some phones even have a higher screen resolution of 2556 × 1440 (2K resolution) pixels. Price of the mobile phones has also fallen to a great extent which makes them a perfect candidate for watching videos on the go. Advancements in mobile hardware coupled with decreasing costs have resulted in a greater demand for high resolution video content that could be watched anytime. A report published in [2] confirms this fact stating that presently video traffic constitutes more than 55% of the overall Internet traffic. In order to mitigate the problem of increased load on the existing network infrastructure, sophisticated video compression techniques have been developed that provide an excellent viewing quality without consuming a large network bandwidth during the streaming sessions. H.265/HEVC (High Efficiency Video Coding) developed by the International Telecommunication Union's (ITU) Video Coding Expert Group (VCEG) and VP9 by Google Inc. are prime examples of such modern codecs. Both of them provide an excellent quality to compression ratio.

Quality of Service (QoS) has been defined by ITU as *"a characteristic of a telecommunications service that bear on its ability to satisfy stated and implied needs of the users of the service"* [3]. On the contrary, the concept of Quality of Experience (QoE) is multidimensional that is influenced by a number of systems, users, and other contextual factors

[4]. For the purpose of successful QoE management by the Internet service providers (ISPs), it is extremely important to understand the relationships between QoE and the underlying network and application-layer QoS parameters. In fact, QoS parameters are the most important business relevant parameters for the ISPs [5]. Therefore, in order to measure the user satisfaction, there is a need for mapping from the QoS to the QoE domain.

In this paper, a video quality prediction model has been presented for videos encoded with H.265/HEVC and VP9 codecs and viewed on mobile devices that are connected to a Wireless Local Area Network (WLAN-802.11 ac standard). We have considered only the infrastructure mode. A total of seven QoS parameters are considered (four representing the network QoS and the remaining three used for application QoS). Large scale subjective tests are carried out for the purpose of model building. Packet loss, jitter, throughput, and auto resolution scaling are the network QoS factors taken into consideration, while bit rate, frame rate, and video resolution are the application QoS factors taken into account. In order to introduce a modularity concept, the video model has been built in three stages. In stages one and two the video quality model for the network and application QoS factors is built separately one after another independent of each other. In stage three, the video models from stage one and stage two are combined together to obtain the final comprehensive model. This modular approach provides more flexibility since it treats the network and application video models independently. This modular feature should be particularly useful to the ISPs because if any change or modification in the model is required afterwards; then they can work on only those specific factors without having to disturb the remaining ones. Detailed methodology has been provided in a later section.

Rest of the paper is organized as follows. In Section 2, related literature review is done. Section 3 presents the detailed methodology for building the video model. Section 4 illustrates the subjective tests that have been conducted along with the relevant data analysis. Sections 5 and 6 present the video model for network and application QoS factors, respectively, while Section 7 presents the final integrated video model. Section 8 introduces the Artificial Neural Network (ANN) approach for model building along with the relevant statistics. Finally, in Section 9, we provide the conclusion and the scope for future work.

2. Literature Review

In this section, we present a brief review of all the relevant works.

2.1. Video Service Quality Estimation Methods. There are two main techniques for video quality assessment: subjective and objective methods. A concise report on both techniques is provided in the following paragraphs.

Till date, subjective tests are considered to be the most accurate video quality assessment method. Typically, in a subjective test, users are gathered together in a room to view some video samples. Then they are asked to rate those samples (typically on a scale of 1 to 5), where 1 denotes the worst and 5 the best quality. The rating which is given by the users is commonly referred to as the Mean Opinion Score (MOS), also known as the QoE. ITU has several recommendations where the procedure for conducting subjective tests has been laid down in detail. Different techniques are available for conducting the subjective tests depending upon the application requirement. Absolute Category Rating (ACR), Absolute Category Rating with Hidden Reference (ACR-HR), Degradation Category Rating (DCR), and Pair Comparison (PC) are some of the most frequently employed techniques. Both ACR and ACR-HR are examples of single stimulus method where only one video sequence is shown at a time. DCR and PC are examples of double stimulus methods where both the original and the degraded video sequences are shown to the users in pair. Further detail about these techniques can be obtained from relevant ITU recommendations [6–8]. The recommendations suggest using video sequences having duration of at least 10 seconds. However, the effect of using videos lesser or greater than 10 seconds on the subjective quality assessment has not been accounted for [9]. Papers [10, 11] provide a detailed comparison among the different subjective techniques. Although subjective methods are very accurate, they are time consuming and very expensive to be carried out. Hence, there is a necessity of objective approach.

Objective techniques are based upon certain algorithms or mathematical formulae that try to predict the perceived video quality by a human observer. An objective approach can be of intrusive or nonintrusive type. Intrusive methods are also known as Full Reference (FR) techniques, because the evaluation process requires both the original and degraded video sequences to be present. Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Video Quality Metric (VQM) are examples of such a scheme [12–15]. There is another variation to the intrusive method where only a subset of the original video sequence is presented to the user for the purpose of quality evaluation. This is referred to as the Reduced Reference (RR) method [16]. Nonintrusive methods do not require the presence of any original video sequence; hence, they are also called no-reference (NR) methods. The models presented in [17–19] represent such a technique. For a video streaming scenario, a NR model is more preferable since it involves minimum overhead as it does not require the presence of the original video sequence. Further detail about these methods can be obtained from [20].

There is a third technique that is increasingly becoming popular in recent years. It involves a combination of the subjective and objective approach as mentioned above. The very recently published ITU-T Recommendation P.1203 is an example of such an approach [21].

2.2. Studies of Correlation between QoS and QoE. Due to the high cost of the subjective tests and relatively low accuracy of the objective algorithms, researchers have tried to estimate the QoE from various QoS factors. ITU-T Recommendation P.861 estimates the listening quality from various voice transmission factors and establishes a nonlinear relationship between the two [22]. Similar work has been done by authors in [23], where they discuss in detail how the human satisfaction of HTTP service (web browsing) is affected by

two network QoS parameters, namely, network bandwidth and latency. A nonlinear relationship between the QoS and QoE for web user satisfaction has been proposed by the authors. However, both the works consider only the network QoS parameters and it needs to be seen if the same type of relationship hold true for video traffic also.

A generic formula in which QoE and QoS parameters are connected through an exponential relationship called the IQX hypothesis is presented by the authors in [24]. The IQX hypothesis is tested for two different services: voice over IP (VoIP) and web browsing for different values of packet loss, jitter, and reordering conditions. However, the validity of IQX hypothesis for other interactive and streaming applications like video is a matter of further investigation. Similarly, the authors in [25] explain the relationship between QoE and QoS in terms of the Weber-Fechner Law (WFL) that is an important principle in psychophysics. The testing environment is limited to a VoIP system and a mobile broadband application scenario involving web browsing, e-mails, and downloads only. Both IQX hypothesis and WFL have been tested for VoIP application and web services only and they are found to be just the inverse of each other. With respect to video streaming, other factors like video resolution, type of the codec used for compression, nature of the video content, and so on are very important towards determining the video QoE. However, they cannot be taken into account by the IQX hypothesis or WFL since they explain only the network QoS factors.

An adaptive QoE measurement scheme for IPTV services has been presented in paper [26]. The authors propose a Video Intelligent Probe (VIP) that integrates the analysis of video processing and network parameters together. The assessment is based upon the quality of images contained in the video signal, packet loss, and the packet arrival time. Similar work has been done by authors in [27] for an IPTV service where the effects of delay, jitter, packet loss rate, error rate, bandwidth, and signal success rate are considered. Both these works primarily take into consideration network QoS factors and they are targeted towards watching videos on a big screen like television. However, there is a considerable difference in the viewing experience on a television and small form factor mobile devices [28, 29].

Authors in [30] evaluate the video quality on a mobile platform considering the impact of spatial resolution, temporal resolution, and the quantization step size. All the videos that are considered have a resolution of 4 CIF (704×576) only. QoE modeling for VP9 and H.265 encoded videos on mobile devices have been investigated by the authors in [31]. Although application QoS factors like bit rate and video resolution have been taken into consideration, they do not include any network QoS factors. Also, the effect of frame rate has not been taken into account. A content based video quality prediction model over wireless local area networks that combine both application and network level parameters has been proposed in [32]. Bit rate and frame rate are the application QoS factors considered, whereas only the effect of packet loss has been taken into account as the network QoS factor across a variety of video content. Similar models have also been proposed in [33–36].

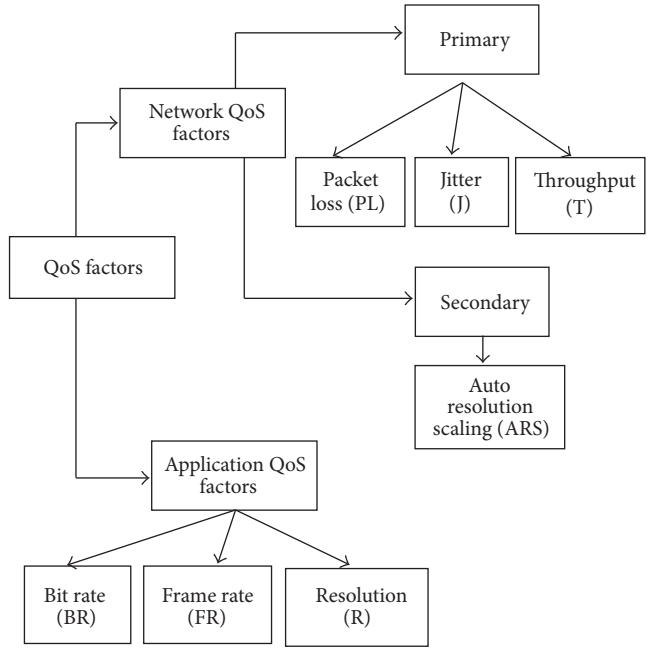

Figure 1: Classification of QoS factors.

A Dynamic Adaptive Streaming over HTTP (DASH) based multiview video streaming system that can minimize the view-switching delay by employing proper buffer control, parallel streaming, and server push schemes has been presented by authors in [37]. Similar HTTP based video streaming for long-term evolution (LTE) cellular networks has been proposed in [38]. Authors in [39–41] try to predict the video QoE for a DASH based video streaming scenario. Papers [42–44] provide an excellent survey on the QoE estimation techniques in place for a video streaming scenario in general.

From this section, we conclude that a lot of research has been done on the quality estimation of videos. The factors taken into consideration belong either to the application layer or the network layer of the TCP/IP protocol suite. Some authors who have considered an effect of both use low resolution videos encoded with older generation codecs like MPEG-2 and H.264/AVC. Also, very little work has been done with respect to video quality estimation on mobile devices. Considering the tremendous and ever increasing popularity of online video streaming services, it is the need of the hour to develop a comprehensive quality prediction model which takes into account factors from both the network and application layers for videos that have been encoded with the current generation H.265/HEVC and VP9 codecs on a mobile device.

3. Methodology

3.1. Problem Statement. In a video streaming service, there are several factors that affect the visual quality as perceived by the end users. These QoS factors can be grouped under the category of network and application QoS factors. Figure 1 gives a detailed classification of the factors that we have

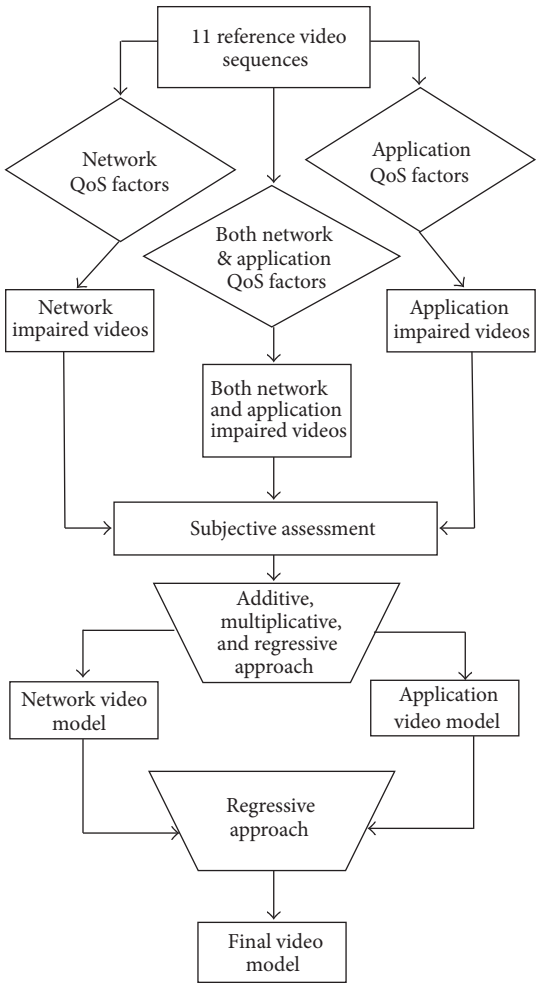

Figure 2: Methodology for building the video model.

considered in this paper. Therefore, for our case, the network layer perceptual QoS/video model will be a function of

$$\frac{\text{Perceptual QoS}}{\text{QoE}_{\text{Network}}} = f\,(\text{PL}, \text{J}, \text{T}, \text{ARS})\,. \tag{1}$$

Similarly, in the application layer, it will be a function of

$$\frac{\text{Perceptual QoS}}{\text{QoE}_{\text{Application}}} = f\,(\text{BR}, \text{FR}, \text{R})\,. \tag{2}$$

As both $\text{QoE}_{\text{Network}}$ and $\text{QoE}_{\text{Application}}$ have the same scale (equivalent to the MOS scale of 1 to 5), hence, the overall/final video model can be expressed as

$$\text{QoE}_{\text{Overall}} = f\left(\text{QoE}_{\text{Network}}, \text{QoE}_{\text{Application}}\right). \tag{3}$$

Equations (1) and (2) are absolutely independent of each other, while (3) integrates (1) and (2) together. This is the reason that our proposed model is modular in nature. Depending upon the requirement either $\text{QoE}_{\text{Network}}$, $\text{QoE}_{\text{Application}}$, or $\text{QoE}_{\text{Overall}}$ can be obtained. Figure 2 depicts the overall methodology that is followed for building the video model.

11 reference videos are chosen from the SVT High Definition Multiformat Test Set database maintained by the Video Quality Experts Group (VQEG) [45]. These reference videos are then subjected to various types of network and application level impairments. The degraded video sequences are then shown to the users who rate them on a scale of 1 to 5. The results from the subjective test are used for creating the objective model.

We begin by mapping the individual network and application QoS factors to their corresponding QoE. Various types of fitting functions are considered, but we choose the optimal one based upon a decision variable (DV) that is discussed in a later part of the paper. After this, we find the perceived video quality due to multiple impairment factors, that is, multiple network and application QoS factors, and refer to them as $\text{QoE}_{\text{Network}}$ and $\text{QoE}_{\text{Application}}$, respectively. This is done in three steps as discussed below:

(1) In step 1, we use a weighted sum (additive) approach to find the network and application QoE by using the Analytic Hierarchy Process (AHP) algorithm and refer to them as $\text{QoE}_{\text{Add}_{(\text{Network})}}$ and $\text{QoE}_{\text{Add}_{(\text{Application})}}$, respectively. However, due to a drawback in this approach next we use a multiplicative technique.

(2) In step 2, we use a multiplicative approach to find the network and application QoE and refer to them as $\text{QoE}_{\text{Mul}_{(\text{Network})}}$ and $\text{QoE}_{\text{Mul}_{(\text{Application})}}$, respectively.

(3) In step 3, we take into account the interaction of the additive and multiplicative approaches to find the final network and application QoE denoted by $\text{QoE}_{\text{Network}}$ and $\text{QoE}_{\text{Application}}$, respectively.

The final video model $\text{QoE}_{\text{Video-Model}}$ is found out from $\text{QoE}_{\text{Network}}$ and $\text{QoE}_{\text{Application}}$ by using a multiple regression approach. Due to the widespread use of different machine learning algorithms we also find $\text{QoE}_{\text{Video-Model}}$ using an Artificial Neural Network (ANN) approach and compare the results across the two methods.

Next we discuss in detail the various QoS factors that have been considered in this paper.

3.2. Network QoS Factors. Here we provide the detail of the considered network QoS factors.

(1) Packet Loss (PL). IP packets may be discarded during their transit over the network or dropped at any intermediate nodes due to network congestion or buffer overflow. Here, we consider a random packet loss pattern as it has a significant detrimental effect on the video stream quality as compared to other types of packet losses [46]. The different packet loss levels that we have considered have been taken from the recommended range of values as suggested by ITU-T via their recommendation ITU-T Y.1541 [47] and presented in Table 4.

(2) Jitter (J). It is defined as the variable delay in receiving packets at the receiver end. It can occur due to network congestion, improper queuing, or several other factors.

(3) Throughput (T). It refers to the amount of data that is successfully transferred from one place to another in a given

time period. Its influence towards the video QoE has been well accepted by the research community.

(4) Auto Resolution Scaling (ARS). In an adaptive video streaming scenario, the videos are encoded at multiple discreet bitrates, that is, at different resolutions. For example, the most commonly used video resolution by YouTube is at 144p, 240p, 360p, 480p, 720p, or 1080p. Depending upon the available network bandwidth and other factors, a particular bitrate stream is broken into multiple segments or chunks, with each segment lasting between 2 and 10 seconds. For the sake of this research, the resolution combinations that we choose are (360p + 480p), (720p + 360p), (720p + 480p), (360p + 1080p), and (1080p + 720p). The duration of the video sequences that we use in our experiment are 10 seconds each. Considering the fact that the duration of each fragmented segment should be between 2 and 10 seconds in case of a resolution switch, we take into account only two resolution switches for a particular video playback. Higher number of resolution switches have not been considered keeping in mind the total length of the original video sequences. For the purpose of data analysis, we express the ARS factor as the ratio of a particular resolution combination to the minimum resolution combination of the videos that is used. For example, the ARS factor for (720p + 360p) combination is $(1280 \times 720 + 640 \times 360)/(640 \times 360 + 854 \times 480) = 1.8$. Similarly, for (360p + 480p), (720p + 480p), (360p + 1080p), and (1080p + 720p), the ARS factors are 1, 2.1, 3.6, and 4.7, respectively.

Now we explain how the secondary ARS factor is related to the primary ones. Auto resolution scaling is a type of adaptive bitrate streaming technique that is used by the video content providers with an aim to improve the viewing QoE. The video content provider stores the same video contents in multiple resolutions and then depending on various network factors like the available network bandwidth, extent of jitter, and the overall network loading conditions select a particular resolution for showing them to the users. Automatic switching to lower or higher resolutions than what is currently being played happens depending upon the network conditions and factors like amount of playout buffer left, video rendering capability of the viewer's device, and so on. Hence, the ARS factor that we have considered is a consequence of the primary ones.

3.3. Application QoS Factors. Bitrate, frame rate, and resolution of the source videos are the application QoS factors that are considered. The videos that are used in the experiment vary over a wide range of video content. The bitrate factor is different from the throughput one (although they are measured using the same units). Bitrate is a codec related feature, while throughput is a network property that refers to the available bandwidth at any point of time.

The perceived video quality depends on the type of video content to a great extent which has been discussed by authors in [32, 48–50]. To define the different types of video contents we have considered the Spatial Information (SI) and Temporal Information (TI) of the source videos. SI gives an indication of the amount of spatial details that each frame has and it has a higher value for more spatially complex scenes. The SI value for every video frame has been calculated by filtering each one of them using the Sobel filter followed by computing the standard deviation. The maximum value in the frame represents the SI content of the scene. Similarly, TI values give an indication of the amount of temporal changes in a particular video sequence. It has a higher value for sequences having greater amount of motion. Equations (4) and (5) show the calculation of the SI and TI values, respectively,

$$\mathrm{SI} = \max_{\mathrm{time}} \left\{ \mathrm{std}_{\mathrm{space}} \left[\mathrm{Sobel} \left(F_n \right) \right] \right\}, \tag{4}$$

$$\mathrm{TI} = \max_{\mathrm{time}} \left\{ \mathrm{std}_{\mathrm{space}} \left[F_n \left(i, j \right) - F_{n-1} \left(i, j \right) \right] \right\}, \tag{5}$$

where F_n is the video frame at time n, $\mathrm{std}_{\mathrm{space}}$ is the standard deviation across all the pixels for each filtered frame, and $\max_{\mathrm{time}}$ is the corresponding maximum value in the considered time interval. The SI and TI values are multiplied in order to arrive at the overall content complexity of any video sequence.

The Sobel filter is implemented by convolving two 3×3 kernels over the video frame and taking the square root of the sum of the squares of the results of these convolutions. For $y = \mathrm{Sobel}(x)$, if $x(i, j)$ denotes the pixels of the input image at the ith row and jth column, then the result of the first convolution denoted by $Gv(i, j)$ is given by

$$\begin{aligned} Gv\left(i, j\right) = {} & -1 \times x\left(i-1, j-1\right) - 2 \times x\left(i-1, j\right) - 1 \\ & \times x\left(i-1, j+1\right) + 0 \times x\left(i, j-1\right) + 0 \\ & \times x\left(i, j\right) + 0 \times x\left(i, j+1\right) + 1 \\ & \times x\left(i+1, j-1\right) + 2 \times x\left(i+1, j\right) + 1 \\ & \times x\left(i+1, j+1\right). \end{aligned} \tag{6}$$

Similarly, $Gh(i, j)$ which is the result of the second convolution is given by

$$\begin{aligned} Gh\left(i, j\right) = {} & -1 \times x\left(i-1, j-1\right) + 0 \times x\left(i-1, j\right) + 1 \\ & \times x\left(i-1, j+1\right) - 2 \times x\left(i, j-1\right) + 0 \\ & \times x\left(i, j\right) + 2 \times x\left(i, j+1\right) - 1 \\ & \times x\left(i+1, j-1\right) + 0 \times x\left(i+1, j\right) + 1 \\ & \times x\left(i+1, j+1\right). \end{aligned} \tag{7}$$

Hence, the output from the Sobel filter image at the ith row and jth column is given by

$$y\left(i, j\right) = \sqrt{\left[Gv\left(i, j\right)\right]^2 + \left[Gh\left(i, j\right)\right]^2}. \tag{8}$$

The calculations are performed for all $2 \le i \le N - 1$ and $2 \le j \le M - 1$, where N denotes the number of rows and M the number of columns.

Figure 3 shows the SI and TI values for the eleven video sequences that have been used in this paper.

Table 1: Video sequences used.

Seq. number	Seq. name	Frame rate	Resolution	Chroma format	Content complexity
(1)	Harbor	60 fps	1920 × 1080	4.2.0	1014
(2)	Ice	60 fps	1920 × 1080	4.2.0	756
(3)	DucksTakeOff	50 fps	1920 × 1080	4.2.0	2728
(4)	ParkJoy	50 fps	1920 × 1080	4.2.0	2450
(5)	Crew	60 fps	1920 × 1080	4.2.0	1053
(6)	CrowdRun	50 fps	1920 × 1080	4.2.0	2688
(7)	Akiyo	30 fps	1920 × 1080	4.2.0	255
(8)	Soccer	60 fps	1920 × 1080	4.2.0	2704
(9)	Foreman	30 fps	1920 × 1080	4.2.0	1140
(10)	Football	30 fps	1920 × 1080	4.2.0	2760
(11)	News	30 fps	1920 × 1080	4.2.0	1470

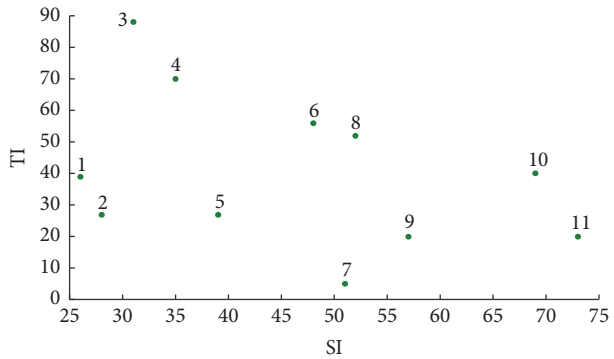

Figure 3: Calculated SI and TI values of chosen video sequences.

Table 2: Encoder configuration for H.265/HEVC codec.

Parameter	Details
Encoder version	HM 16.6
Profile	Main
Reference frames	4
R/D optimization	Enabled
GOP	8
Coding unit size/depth	64/4
Fast encoding	Enabled
Rate control	Disabled
Internal bit depth	8

For each video sequence, we have taken four different resolutions (VGA, qHD, HD, and Full HD). The resolution factor R that is considered is totally different from the ARS factor discussed previously. R refers solely to the resolution of the videos that have not been subjected to any sort of network impairments. However, the ARS factor has been introduced as a network QoS factor in order to take into consideration the effects of adaptive bitrate streaming. For the sake of data analysis, we express the resolution of a particular video sample denoted by R_{Video} in a ratio format given by

$$R_{\text{Video}} = \frac{R_{\text{Original}}}{R_{\text{Minimum}}}, \tag{9}$$

where R_{Original} refers to the actual resolution of the video under consideration and R_{Minimum} refers to its corresponding minimum resolution. For example, the resolution value for any Full HD content will be $(1920\times1080)/(640\times480) = 6.75$. Similarly, the resolution value for any VGA content will be $(640 \times 480)/(640 \times 480) = 1$. Thus, a video having a higher R_{Video} value will be at a higher resolution. Next, we discuss the experiment that has been carried out in detail and the subsequent data analysis.

4. Experiment Details

First, we present the video sequences that have been used in this research.

4.1. Video Selection. The publicly available video database of VQEG has been used for selecting our reference videos. A total of 11 sequences are taken; the details of which are shown in Table 1. All the sequences are roughly of 10-second duration each and in native YUV 4.2.0 format. The raw videos are encoded using the H.265 and VP9 codecs. Tables 2 and 3 show the encoder configuration used for both the codecs, respectively.

Figures 4(a)–4(k) show the snapshot of the videos that are considered.

4.2. Simulation Test-Bed. The simulation test-bed has been shown in Figure 5. We have created 2 sending nodes, namely, a constant bitrate (CBR) background traffic source node and a streaming server that contains all the video sequences we use encoded with the H.265 and VP9 codecs. The bitrate of the CBR has been fixed at 2 Mbps in order to simulate a realistic scenario. Both these sending nodes are connected to router A over the Internet across a 20 Mbps link. Router A is in turn connected to router B over a variable link. Router B is again connected to a wireless access point at 20 Mbps which further transmits this traffic to a mobile node at transmission speeds of up to 600 Mbps typically found in 802.11ac WLANs. No packets are dropped in the wired portion of the video delivery path. The maximum transmitted packet size is 1024 bytes. We use a random pattern for packet loss that takes six values at (0.1, 0.5, 1, 3, 5, and 10%). The effect of jitter

Table 3: Encoder configuration for VP9 codec.

Parameter	Details
Encoder version	Ffmpeg 3.1.3
Encoding quality	Best
Number of passes	2
Bit rate control mode	Variable bit rate (VBR defined by target bit rate)
Constrained quality (CQ) level	Same as quantization Parameter QP
Initial, optimal, and maximum buffer level	4000 ms, 5000 ms, 6000 ms
GOP size	Auto
GOP length (intraperiod)	320
Internal bit depth	8

Table 4: Simulation parameters.

Parameter	Value
Packet loss (%)	0.1, 0.5, 1, 3, 5, 10
Jitter (ms)	1, 2, 3, 4, 5
Throughput (Kbps)	500, 1000, 2000, 3000, 5000
Autoresolution scaling	1, 1.8, 2.1, 3.6, 4.7
Bitrate (Kbps)	500, 1000, 2000, 4000, 8000
Frame rate (fps)	10, 15, 25, 30, 50/60
Resolution	1, 1.69, 3, 6.75

Table 5: Viewing conditions.

Parameter	Setting
Viewing distance from screen	76 cm
Peak luminance of the screen	890 nits
Background room illumination	180 nits
Ratio of brightness of background to peak brightness of screen	0.20

has been added by introducing a fixed delay of 100 ms plus five variable delays corresponding to (1, 2, 3, 4, and 5 ms). The network throughput is varied by changing the bandwidth of the variable link between routers A and B and has been fixed at (500, 1000, 2000, 4000, and 8000 Kbps). As previously mentioned, range of all the values considered is based upon the ITU and ETSI recommendations provided in [47, 51, 52]. Values of all the parameters used in the experiment are provided in Table 4. For videos that have been impaired by a single ARS factor only or any particular application QoS factor, the simulation test-bed has not been used. In order to simulate the ARS factor, a particular video is segmented, with each segment being played back at two different resolutions. For example, in case of a video having 300 frames in total, the first 150 frames are played back at a particular resolution and the remaining 150 frames are played back at a different resolution.

The experiment has been conducted with Evalvid framework [53] and network simulator toolkit NS2 [54]. Integrating NS2 with the Evalvid platform gives us a lot of flexibility in choosing the parameters.

Next, the subjective evaluation process has been described in detail.

4.3. Subjective Evaluation. 59 participants are involved in the subjective test and they are mixed in gender. Figure 6 shows the percentage breakdown of the participants' age. Before recruiting the participants, an Ishihara color vision test has been conducted on them in order to ensure that none of them suffer from color blindness [55]. The test has been conducted in a controlled laboratory environment. It took 16 weeks to complete the entire subjective test. Table 5 gives the details of the viewing conditions under which the test has been performed.

The subjects had to evaluate 462 video sequences that have been impaired by exactly 1 network QoS factor. The total number of network impairment conditions is 21 (6 for PL + 5 for J + 5 for T and 5 for ARS). Considering the 11 video sequences across 2 codecs (21 impaired conditions × 11 video sequences × 2 codecs), we arrive at the number 462. In order to assess the quality of videos impaired by multiple network QoS factors, we limit the number of test sequences to 176. Since carrying out a subjective test consumes considerable amount of time and effort; hence, it was not feasible to include all possible values of the different impairment combinations while presenting the test video sequences. Instead, we choose 176 specific combinations, the details of which have been shown in Table 6.

32 video sequences are impaired by all the network QoS factors, while for all the other remaining conditions, we use 16 sequences for each one. For both the single and multifactor impaired videos, exactly half the number of samples is used for model building and the rest for validation.

Similarly, for creating the application video model, we have 308 video sequences impaired by exactly 1 application QoS factor. Five different BR and FR levels, respectively, along with 4 different resolution values across 2 codecs and 11 sequences give a total of 308 combinations. For creating the multifactor impaired videos, as explained previously we have used a subset of the total possible combination. In particular, 140 sequences are used, the details of which are provided in Table 7. As before, the 140 sequences are split evenly for the purpose of model creation and validation.

The final model is created by combining all the network and application QoS factors together. Thus, we have a total of 7 different factors. Since it is not possible to let the users

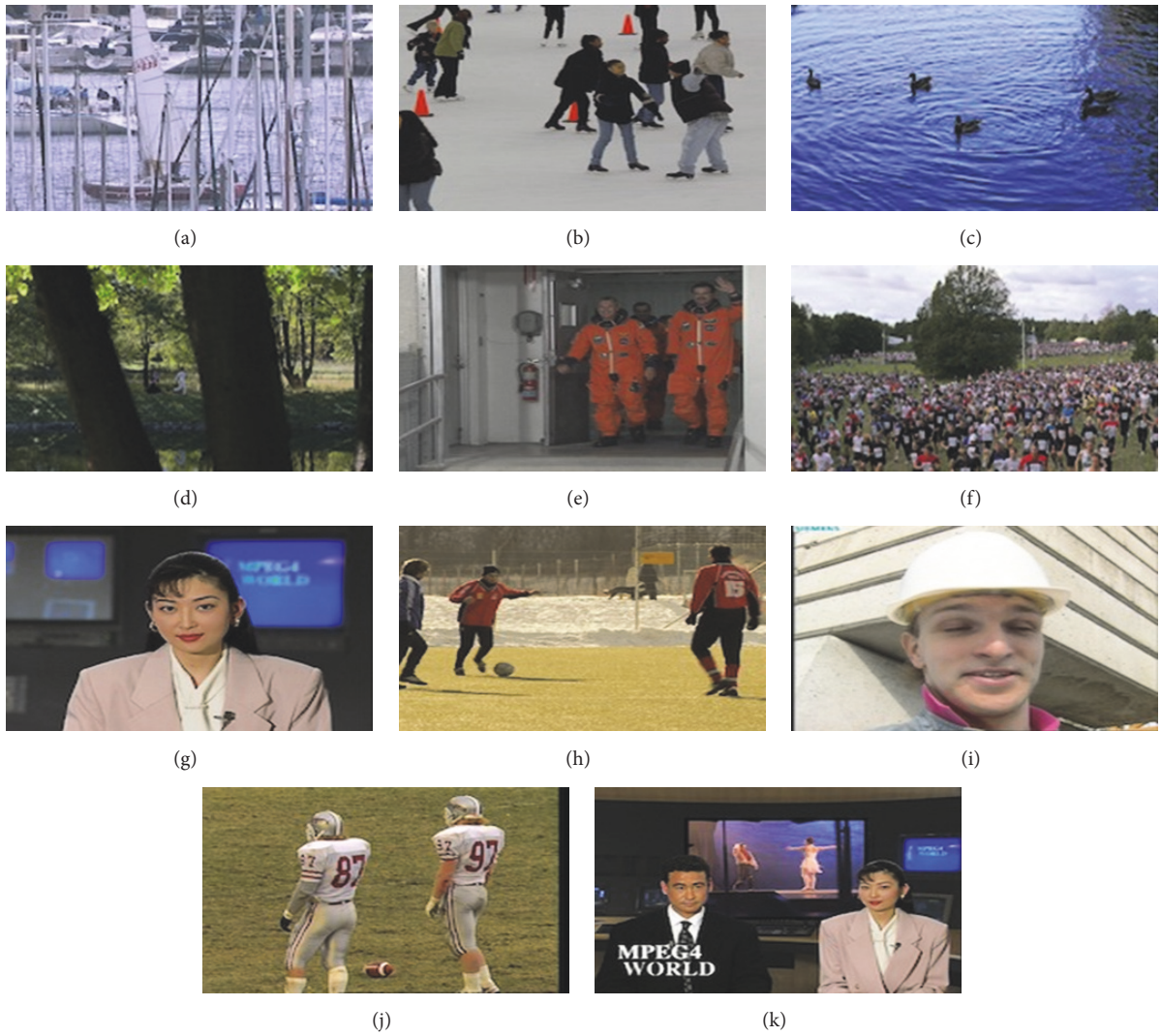

FIGURE 4: Snapshot of used video sequences: (a) Harbor, (b) Ice, (c) DucksTakeOff, (d) ParkJoy, (e) Crew, (f) CrowdRun, (g) Akiyo, (h) Soccer, (i) Foreman, (j) Football, and (k) News.

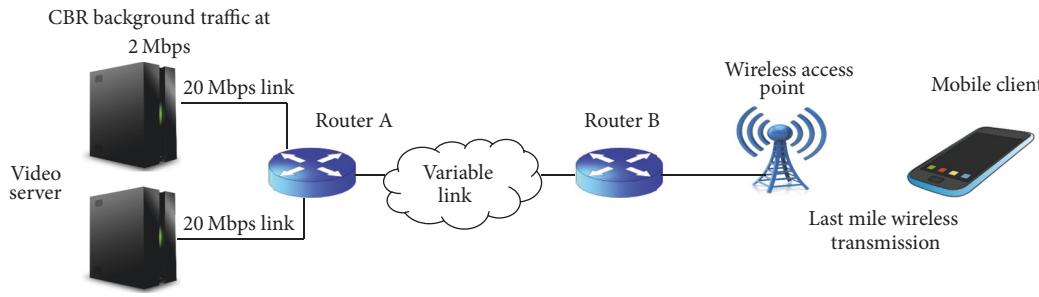

FIGURE 5: Simulation test-bed.

watch such a huge number of videos, we limit the number of impaired videos to 156. Table 8 shows the relevant details. For this case, while creating the video sequences, care has been taken to include the effect of both the network and application QoS factors for every condition. 78 sequences are being used for model creation and the rest for validation.

All the videos are presented on a Samsung Galaxy Note 5 for the purpose of evaluation. We chose this device as it has hardware level decoding capability for the H.265 and VP9 codecs. Hardware level decoding has certain advantages over software decoding. Sometimes software decoding results in a jittery/distorted playback for certain format of videos

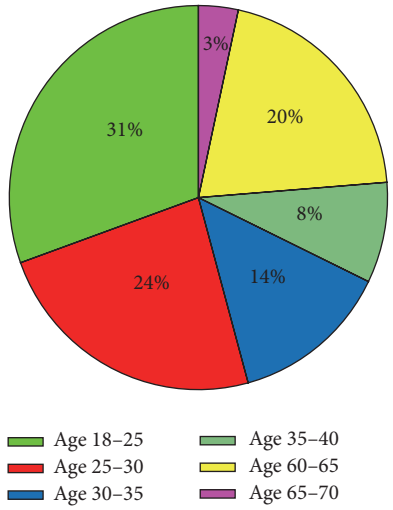

Figure 6: Breakdown of participants' ages.

Table 6: Impairment choice for 176 video sequences (network QoS).

Impairment combination	Total number of impairments	Number of impaired video sequences
(PL + J)	2	16
(J + T)	2	16
(T + ARS)	2	16
(PL + T)	2	16
(PL + ARS)	2	16
(J + ARS)	2	16
(PL + J + T)	3	16
(PL + J + ARS)	3	16
(J + T + ARS)	3	16
(PL + J + T + ARS)	4	32

Table 7: Impairment choice for 140 video sequences (application QoS).

Impairment combination	Total number of impairments	Number of impaired video sequences
(BR + FR)	2	20
(BR + R)	2	20
(FR + R)	2	20
(BR + FR + R)	3	80

Table 8: Sequence detail for creating final video model.

Impairment combination	Total number of impairments	Number of impaired video sequences
(PL + J + BR + FR)	4	28
(J + T + BR + FR + R)	5	30
(PL + J + T+ ARS + BR + FR)	6	20

encoded specially with newer codecs. Hardware acceleration is very useful for such cases. In case of hardware acceleration manufacturers specifically implement multimedia chipsets as a part of the motherboard to assist with the video decoding process, whereas software decoders only use the CPU to play videos. Hence, the choice is between something specific (hardware decoders) versus something generic (software decoders). This is the exact reason why we choose Samsung Galaxy Note 5 as it has a dedicated decoder chip for H.265 and VP9 codecs.

Single stimulus ACR technique as outlined in ITU-T Recommendation P.910 has been used for designing the experiment. The total number of test videos that the participants have to watch is quiet large (1242 sequences). Approximately, each subject needs about 4 hours of time in order to complete the entire assessment procedure. We divided the entire test duration into 9 different subsessions. Five sessions were completed on the first day and the remaining 4 on the next day for each subject. Each session lasted for about 30 minutes followed by a 15-minute break in order to remove any sort of tiredness and fatigue that may arise due to the extended viewing period. The videos were presented to the subjects in a random order.

Next, we discuss the data processing method used.

4.4. Outlier Detection and Score Estimation. In case of the subjects whose scores deviate to a certain extent from the mean score, outlier detection has to be carried out in order to remove the subject bias. Roughly, the score distribution should be normal which we find out using β_2 test (by calculating the kurtosis coefficient of the function, i.e., the ratio of the fourth-order moment to the square of the second-order moment). For a particular test video sequence k, we calculate the mean ($\overline{x}_k$), standard deviation (S_k), and the kurtosis coefficient (β_{2k}) as

$$\beta_{2k} = \frac{m_4}{m_2^2}, \quad \text{with } m_x = \frac{\sum_{i=1}^{N} (x_{ik} - \overline{x}_k)^x}{N}, \tag{10}$$

where N is total number of subjects and x_{ik} is score given by ith user for kth test video

For each observer i we find P_i and Q_i as given below.

If $2 \leq \beta_{2k} \leq 4$, then:

$$\text{if } (x_{ik} \geq \overline{x}_k + 2s_k),\ P_i = P_i + 1$$
$$\text{if } (x_{ik} \leq \overline{x}_k - 2s_k),\ Q_i = Q_i + 1$$

Else:

$$\text{if } (x_{ik} \geq \overline{x}_k + \sqrt{20}s_k),\ P_i = P_i + 1$$
$$\text{if } (x_{ik} \leq \overline{x}_k - \sqrt{20}s_k),\ Q_i = Q_i + 1.$$

Following the above procedure, any subject i will be removed from the analysis if $(P_i + Q_i)/N > 0.05$ and $(P_i - Q_i)/(P_i + Q_i) < 0.3$. Consequently, the ratings from 4 subjects for the packet loss factor, 5 subjects for the jitter factor, 7 subjects for the ARS factor, and 3 subjects for the frame rate

factor have been removed from further analysis. Based upon the user rating the QoE or MOS is calculated as

$$\frac{\text{QoE}_k}{\text{MOS}_k} = \frac{1}{N}\sum_{i=1}^{N} x_{ik}, \tag{11}$$

where N is number of subjects and x_{ik} is score given by the ith user for kth video

Next, we present the network video model.

5. Network Video Model

To build the network video model first we consider the effect of single network QoS factors on the user QoE. Thereafter, we find the joint effect of all the network QoS factors considered.

5.1. Mapping for Individual QoS Factors to User QoE. We do a nonlinear curve-fitting on the subjective dataset to arrive at the relationships between the QoS factors and their corresponding QoE. An optimal fitting is chosen based upon a decision variable (DV) that is introduced here. The overall goodness-of-fit statistics is generally expressed in terms of the sum of squared error (SSE), root mean square error (RMSE), R^2 change, or the adjusted $-R^2$ change values. For SSE and RMSE, values closer to 0 indicate that the model has a smaller random error component and that the fit will be more useful for prediction. Similarly, R^2 and adjusted $-R^2$ values close to 1 indicate that a greater proportion of variance is accounted for by the model. R^2 and adjusted $-R^2$ are given as

$$\begin{aligned} R^2 &= \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}, \\ \text{Adjusted}R^2 &= 1 - \left(\frac{n-1}{n-p}\right)\frac{\text{SSE}}{\text{SST}}, \end{aligned} \tag{12}$$

where SST is sum of squared total, n is number of observations, and p is number of regression coefficients including the intercept. Based upon the above discussion, we propose the variable DV as

$$\text{DV} = \frac{\left(R^2 \times \text{Adjusted}R^2\right)}{(\text{SSE} \times \text{RMSE})}. \tag{13}$$

Equation (13) suggests that a higher value of DV is always desirable. We considered various types of fitting models and choose the one which is optimized to get the highest value of DV possible. The goodness-of-fit statistics for each individual mapping has been shown in Table 9.

Equations (14)–(17) give the mapping from QoS to QoE domain for packet loss, jitter, throughput, and auto resolution scaling factor, respectively.

$$\text{QoE}_{\text{PL}} = a \times \exp^{(b\times\text{PL})} + c \times \exp^{(d\times\text{PL})}, \tag{14}$$

$$\text{QoE}_{\text{J}} = a \times \exp^{(b\times\text{J})} + c \times \exp^{(d\times\text{J})}, \tag{15}$$

$$\text{QoE}_{\text{T}} = a \times \log(\text{T}) + b, \tag{16}$$

$$\text{QoE}_{\text{ARS}} = a \times \exp^{(-b\times\text{ARS})} + c, \tag{17}$$

Table 9: Model fitting statistics for network factors.

Parameter	Codec	SSE	RMSE	R^2	AdjustedR^2	DV
PL	H.265	0.011	0.076	0.998	0.996	1132.51
J	H.265	0.002	0.049	0.999	0.997	8328.25
T	H.265	0.091	0.175	0.985	0.980	60.13
ARS	H.265	0.016	0.04	0.985	0.981	1509.82
P	VP9	0.006	0.059	0.999	0.997	2407.75
J	VP9	0.013	0.022	0.999	0.996	3479.03
T	VP9	0.161	0.282	0.962	0.960	20.34
ARS	VP9	0.092	0.097	0.958	0.936	100.48

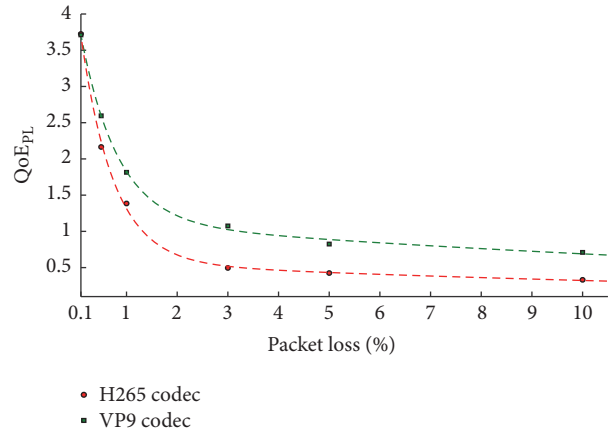

Figure 7: Relationship between PL (%) and QoE_{PL} (from subjective test).

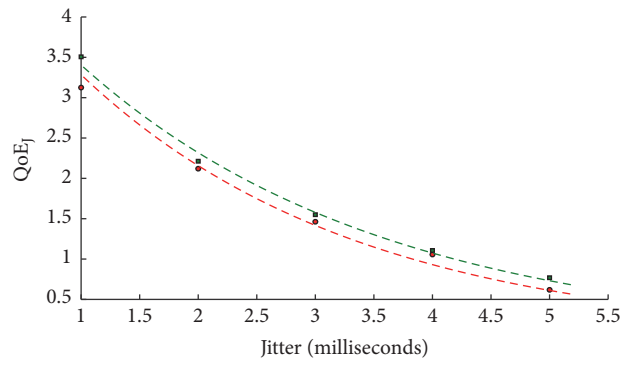

Figure 8: Relationship between jitter (ms) and QoE_{J} (from subjective test).

where a, b, c, and d are the coefficients that are found out from curve-fitting and presented in Table 10.

Figures 7–10 show the relationship between the QoS and the corresponding QoE/MOS.

Generally we observe that, for all the factors, videos encoded with VP9 codec have a slightly better QoE as compared to the H.265 codec.

The PCC (Pearson Correlation Coefficient) has also been calculated for the set of equations obtained above for the individual factors. This has been shown in Table 11. Results

Table 10: Coefficient values for network factors.

Parameter	a H.265/VP9 (95% CI)	b H.265/VP9 (95% CI)	c H.265/VP9 (95% CI)	d H.265/VP9 (95% CI)
PL	3.66/2.96	−1.56/−1.38	0.57/1.13	−0.06/−0.05
J	4.51/11.62	−0.37/−3.39	-2×10^{-16}/4.4	6.73/−0.35
T	−1.39/−1.65	−7.44/−9.40	—	—
ARS	3.47/3.38	$-4.4 \times 10^{-8}/-3.7 \times 10^{-7}$	8.6×10^{-16}/0.69	—

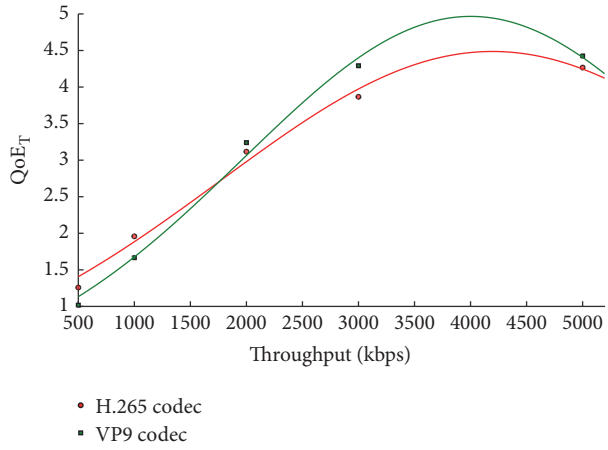

Figure 9: Relationship between throughput (Kbps) and QoE_T (from subjective test).

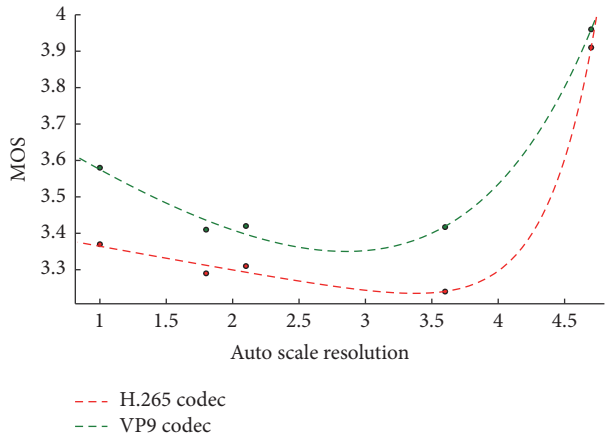

Figure 10: Relationship between auto resolution scaling and QoE_{ARS} (from subjective test).

Table 11: Correlational analysis of the different network QoS factors.

Parameter	PCC
PL	0.952
J	0.978
T	0.874
ARS	0.924

show that the QoE values which the equations predict have a high degree of correlation with the actual subjective scores.

Next we present the integrated QoE measurement technique from the individual QoS factors.

5.2. Integrated QoE Measurement for Network Factors. An additive and multiplicative approach is used for finding out the integrated QoE. The final network video model is obtained by carrying out a regression across both the approaches.

In an additive form, the QoE is generally represented as

$$\mathrm{QoE}(x_1, x_2) = w_1 \mathrm{QoE}_{x_1} + w_2 \mathrm{QoE}_{x_2}, \quad (18)$$

where w_1 and w_2 are the weights that need to be found out for QoS factors x_1 and x_2, respectively. Not all the network QoS factors considered here have the same impact on the perceived video quality. The factor which affects more should be given a higher weight as compared to the factor which is lesser important. Before going for the additive approach, in order to explicitly find out the effect of the different network QoS parameters on the QoE, we perform ANOVA (Analysis of Variance) on the subjective dataset that has the score collected from the 176 video sequences which have been impaired by all the network factors. Table 12 shows the result from the ANOVA analysis.

Small p value ($p \leq 0.01$) suggests that all the parameters that are considered are significant. Based upon the magnitude of the p values we can make further claim that jitter impacts the MOS results the most followed by packet loss and auto resolution scaling. Throughput has the least influence. This observation is extremely important in assigning proper weights to the different factors in the additive approach.

For assigning the weights, Analytic Hierarchy Process (AHP) algorithm has been used [56, 57]. It is a well-known structured technique that is often used in multicriteria decision making systems. As the first step we obtain the criteria comparison matrix that has been shown in Table 13.

The next step is to build the normalized matrix from which we can get the weight of every factor considered. This normalized matrix has been shown in Table 14.

Thus, for the case of the network QoE in additive form (18) reduces to

$$\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Network})}} = 0.26\mathrm{PL} + 0.55\mathrm{J} + 0.07\mathrm{T} + 0.12\mathrm{ARS}. \quad (19)$$

From (19), it is evident that the weight associated with the jitter factor is maximum, while it is minimum for the throughput factor. The QoE which is calculated by the additive method has a disadvantage that is now explained.

A video that has been distorted by two QoS metrics should not have a better QoE than the video which has been distorted by only one of the two QoS metrics. For example, we refer to Table 15 that shows a sample calculation.

Table 12: Four-way ANOVA on MOS collected from 176 video sequences.

Parameter	Sum of squares	Degrees of freedom	Mean squares	F-statistic	p value
PL	15.74	5	3.15	37.38	1.91×10^{-4}
J	8.50	4	2.12	114.56	4.2×10^{-5}
T	15.69	4	3.92	109.57	4.7×10^{-3}
ARS	1.43	4	0.36	59.70	2.08×10^{-4}

Table 13: Criteria comparison matrix for network factors.

Factors	PL	J	T	ARS
PL	1	0.333	5	3
J	3	1	7	5
T	0.2	0.143	1	0.333
ARS	0.333	0.2	3	1
Intermediate loading	4.533	1.676	16	9.333

Table 14: Normalized matrix for weight calculation.

Factors	PL	J	T	ARS
PL	0.22	0.19	0.31	0.32
J	0.66	0.59	0.43	0.53
T	0.04	0.08	0.06	0.03
ARS	0.07	0.12	0.18	0.10
Weight contribution	0.26	0.55	0.07	0.12

Table 15: Sample calculation of $\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Network})}}$.

Network factor	QoS value	QoE	Additive QoE
PL	1%	1.31	0.35
J	5 ms	0.62	0.34
T	2000 Kbps	3.10	0.22
ARS	1	4.07	0.49
$\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Network})}}$			1.4

The QoE value for each network factor is calculated from the individual QoS to QoE mapping functions that we presented previously in (14)–(17). The additive contribution of each QoE factor is calculated by multiplying each individual network QoS factor by its weight. Finally, $\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Network})}}$ is obtained by adding the contribution of the corresponding impairment terms. For this particular case, the range of the QoE values for the individual factors varies from 0.62 to 4.07. The additive QoE value obtained is 1.4, which is within this range. However, it contradicts the fact that the QoE should not be greater than 0.62 (which is the minimum QoE obtained). Thus, clearly there is an anomaly while calculating the QoE using the additive approach.

Hence, we consider an alternative multiplicative approach. As the subjects give their opinion on a scale of 1 to 5, we present the QoE equation in multiplicative form by

$$\mathrm{QoE}_{\mathrm{Mul}_{(\mathrm{Network})}} = 5 \times \left(\frac{\mathrm{PL}}{5}\right) \times \left(\frac{\mathrm{J}}{5}\right) \times \left(\frac{\mathrm{T}}{5}\right) \times \left(\frac{\mathrm{ARS}}{5}\right). \quad (20)$$

Each individual QoS factor has been weighed on a scale of 5, while evaluating its contribution towards the final multiplicative QoE. Table 16 shows a sample calculation using (20). For the purpose of illustration, same set of QoS values have been taken for both the approaches. The $\mathrm{QoE}_{\mathrm{Mul}_{(\mathrm{Network})}}$ value obtained is 0.08 (lesser than the minimum QoE value of 0.62 corresponding to jitter).

Table 16: Sample calculation of $\mathrm{QoE}_{\mathrm{Mul}_{(\mathrm{Network})}}$.

Network factor	QoS value	QoE	Multiplicative QoE
PL	1%	1.31	0.26
J	5 ms	0.62	0.12
T	2000 Kbps	3.10	0.62
ARS	1	4.07	0.81
$\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Network})}}$			0.08

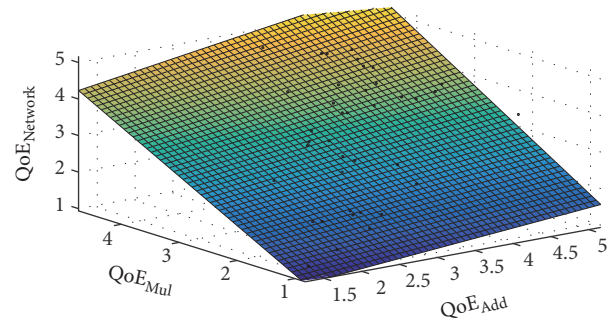

Figure 11: Network video model.

Comparison of the QoE values obtained from both the approaches for the same set of network QoS conditions reveal that the additive approach tends to overpredict the actual viewing quality, while the multiplicative approach tends to underpredict the same. Hence, for building the final network video model $\mathrm{QoE}_{\mathrm{Network}}$, we use a regression based approach that combines the additive and multiplicative techniques just presented.

The regressive model is built based upon (19) and (20) along with the results of the subjective dataset that have 88 video sequences impaired by multiple network QoS factors. Equation (21) represents the network video model which is further shown in Figure 11.

$$\begin{aligned}\mathrm{QoE}_{\mathrm{Network}} &= 0.14\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Network})}} + 0.81\mathrm{QoE}_{\mathrm{Mul}_{(\mathrm{Network})}} \\ &\quad + 0.04\left(\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Network})}} \times \mathrm{QoE}_{\mathrm{Mul}_{(\mathrm{Network})}}\right).\end{aligned} \quad (21)$$

Table 17: Modeling accuracy of each stage.

Model stages	R^2	AdjustedR^2
Additive	0.654	0.649
Multiplicative	0.889	0.888
Regression	0.913	0.912

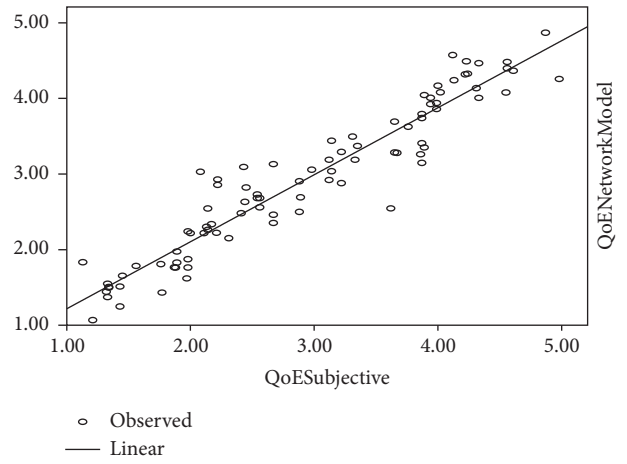

Figure 12: Accuracy of network video model.

Equation (21) suggests that for the network video model the contribution of the multiplicative part is more as compared to the additive one. Accuracy of the network model is shown in Figure 12, while Table 17 reports the accuracy of each stage in the model building phase. While creating Figure 12, we have used the unseen subjective data that has not been used for the purpose of model building. We note that at each stage there is a gradual increase in the modeling accuracy.

Next, we present the application video model.

6. Application Video Model

A similar approach like the network video model is followed to build the application video model. First, the effects of the individual application QoS factors to the viewing quality are examined followed by an integrated application QoE estimation using the same three techniques previously presented.

6.1. Mapping for Individual Application QoS Factors to User QoE. Equations (22)–(24) show the variation of QoE with respect to bitrate, frame rate, and resolution of the impaired videos, respectively. All the mappings have been done with respect to the decision variable that has already been introduced in the previous section of the paper.

$$\mathrm{QoE}_{\mathrm{BR}} = a \times \log(\mathrm{BR}) + b, \tag{22}$$

$$\mathrm{QoE}_{\mathrm{FR}} = a \times \exp(b \times \mathrm{FR}) + c \times \exp(d \times \mathrm{FR}), \tag{23}$$

$$\mathrm{QoE}_{\mathrm{R}} = a \times \exp\left(-\left(\frac{R-b}{c}\right)^2\right). \tag{24}$$

Table 18: Coefficient values for application factors.

Parameter	a	b	c	d
BR	0.36/0.34	0.86/1.05	—	—
FR	10.27/7.1	−0.01/−0.02	−8.40/−5.1	−0.03/−0.02
R	3.47/3.51	5.15/5.98	8.64/12.2	—

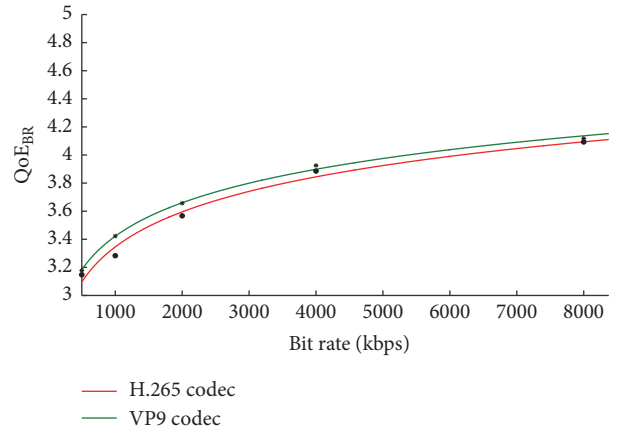

Figure 13: Relationship between bit rate (Kbps) and $\mathrm{QoE}_{\mathrm{BR}}$ (from subjective test).

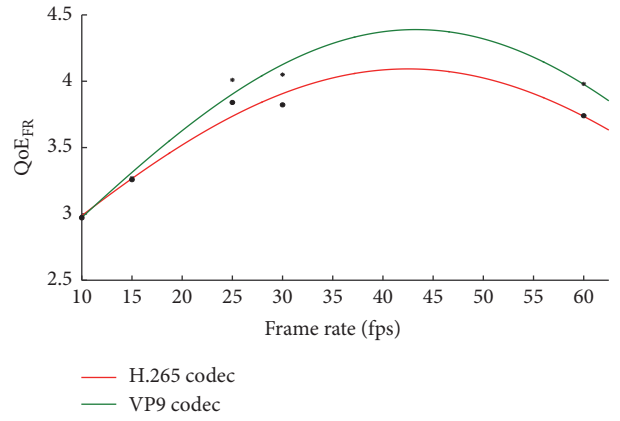

Figure 14: Relationship between frame rate (fps) and $\mathrm{QoE}_{\mathrm{FR}}$ (from subjective test).

The relevant coefficients are found out from the experiment and presented in Table 18. The corresponding graphs have been shown in Figures 13–15. For all the factors VP9 codec offers a better viewing experience. Figure 14 suggests that for every video sequence there is an optimal frame rate beyond which the viewing quality does not improve and enters a saturation stage. Similarly, the effect of resolution on the perceived quality follows a Gaussian distribution as evident from (24) and Figure 15. We attribute this observation to the limitations of the human visual system and the size of the screen on which the video is being watched. In case of our experiment, the videos are viewed on a mobile device. The results clearly indicate that, for small sized screens, there will not be any substantial improvement in the viewing quality by increasing the resolution of the videos.

Table 19: Model fitting statistics for application factors.

Parameter	Codec	SSE	RMSE	R^2	AdjustedR^2	DV
BR	H.265	0.009	0.055	0.985	0.981	1952.01
FR	H.265	0.018	0.095	0.972	0.964	547.96
R	H.265	0.002	0.014	0.991	0.989	35003.5
BR	VP9	0.001	0.020	0.998	0.997	49750.3
FR	VP9	0.019	0.097	0.981	0.962	512.06
R	VP9	0.00004	0.006	0.999	0.992	4.1×10^6

Table 20: Correlational analysis of the different application QoS factors.

Parameter	PCC
BR	0.916
FR	0.941
R	0.987

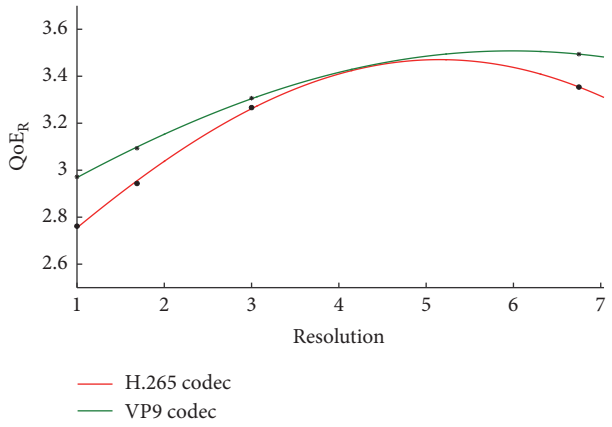

Figure 15: Relationship between resolution and QoE_R (from subjective test).

The model fitting statistics for the individual application factors are shown in Table 19. PCC coefficients presented in Table 20 show a relatively high correlation between the subjective scores and the calculated MOS.

Next, we present the integrated approach towards finding the application level QoE.

6.2. Integrated QoE Measurement for Application Factors. The application video model is also built in three stages comprising the additive, multiplicative, and the regressive approach, respectively. As before, an ANOVA analysis is carried out in the beginning over the subjective dataset containing 140 videos that have been impaired by all the application QoS factors. The result of this analysis is used to choose the relative importance of the factors and assign proper weights to them based upon the AHP algorithm. The ANOVA report has been presented in Table 21. The viewing quality is most impacted by frame rate followed by bitrate and resolution, respectively.

The additive form for the application factors has been shown in (25). Intermediate criteria comparison matrix and

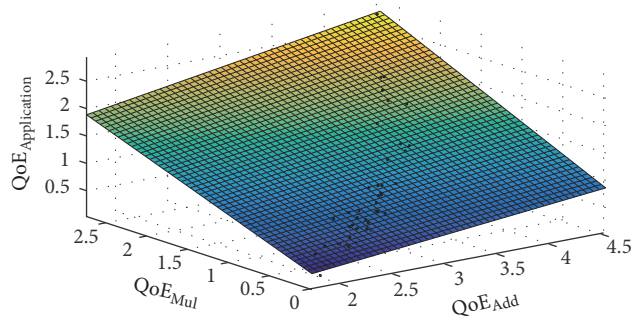

Figure 16: The application video model.

the final normalized weight matrix that are obtained from the AHP algorithm are presented in Tables 22 and 23, respectively.

$$\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Application})}} = 0.26\mathrm{BR} + 0.63\mathrm{FR} + 0.11\mathrm{R}. \quad (25)$$

The additive approach suffers from the same type of anomaly that has already been discussed in the previous section. Hence, we present the multiplicative form in

$$\mathrm{QoE}_{\mathrm{Mul}_{(\mathrm{Application})}} = 5 \times \left(\frac{\mathrm{BR}}{5}\right) \times \left(\frac{\mathrm{FR}}{5}\right) \times \left(\frac{\mathrm{R}}{5}\right). \quad (26)$$

As before, the additive approach tends to overpredict the viewing quality, whereas the multiplicative approach tends to underpredict the same. Therefore, a regression based model is presented in (27) that integrates both the approaches for finding the final video quality due to the application factors. The regression model is build based on (25) and (26) along with the subjective score obtained from the 70 video sequences that have been impaired by all the concerned application QoS factors.

$$\begin{aligned}\mathrm{QoE}_{\mathrm{Application}} &= 0.19\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Application})}} + 0.49\mathrm{QoE}_{\mathrm{Mul}_{(\mathrm{Application})}} \\ &\quad + 0.042\left(\mathrm{QoE}_{\mathrm{Add}_{(\mathrm{Application})}} \times \mathrm{QoE}_{\mathrm{Mul}_{(\mathrm{Application})}}\right).\end{aligned} \quad (27)$$

The application video model and its accuracy are shown in Figures 16 and 17, respectively. Table 24 presents the modeling accuracy across all the three stages.

Next, we find the final integrated video model by combining the network and application video models just presented.

Table 21: Three-way ANOVA on MOS collected from 140 video sequences.

Parameter	Sum of squares	Degrees of freedom	Mean squares	F-statistic	p value
BR	112.458	4	28.114	44.661	1.4×10^{-4}
FR	277.677	5	55.535	118.872	5.2×10^{-5}
R	14.039	3	4.680	6.448	0.003

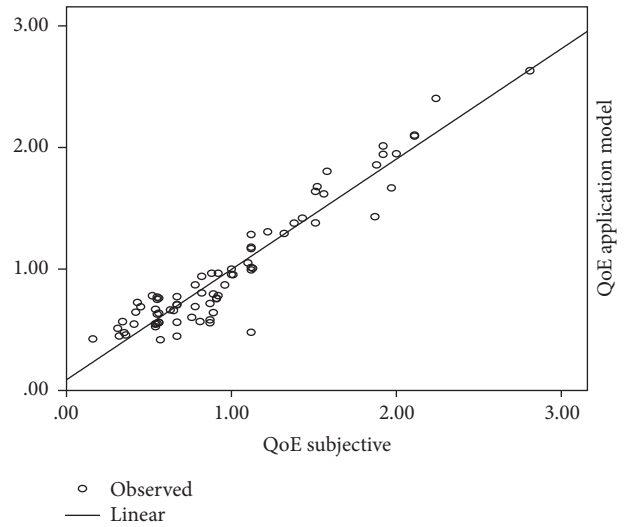

Figure 17: Accuracy of application video model.

Table 22: Criteria comparison matrix for application factors.

Factors	BR	FR	R
BR	1	0.33	3
FR	3	1	5
R	0.33	0.20	1
Intermediate loading	4.33	1.53	9

Table 23: Normalized weight matrix for application factors.

Factors	BR	FR	R
BR	0.23	0.22	0.33
FR	0.69	0.65	0.55
R	0.08	0.13	0.11
Weight contribution	0.26	0.63	0.11

Table 24: Modeling accuracy of each stage.

Model stages	R^2	AdjustedR^2
Additive	0.848	0.846
Multiplicative	0.904	0.903
Regression	0.912	0.910

7. Final Integrated Video Model

Till now, separate models have been built for the network and application QoS factors. With an aim to build a cross-layer model, we now combine these two models into one single entity.

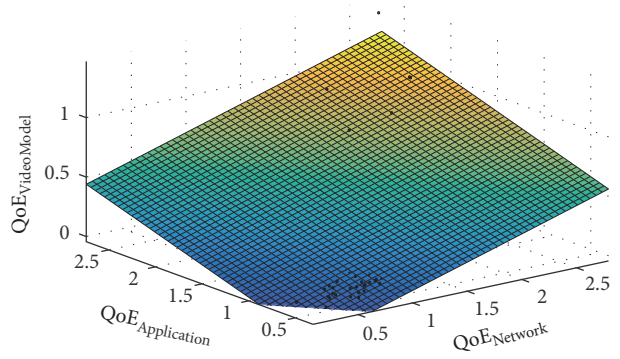

Figure 18: Final integrated video model.

For creating the final video model 78 video sequences are taken. All these video sequences are impaired by multiple network and application QoS factors considered here. Details about the video sequences are provided in Table 8. Based on the MOS scores obtained across these 78 sequences and (21) and (27); a regression approach is used to build the final video model. A stepwise method of variable entering scheme is used. During any step if we obtain a nonsignificant result, then the corresponding parameter is removed. Equation (28) represents the final video model and Figure 18 depicts the same. The coefficients of each of the contributing factors suggest that while calculating the overall video quality, the effect of the network QoE is more than the effect due to the

$$\mathrm{QoE}_{\mathrm{Video-Model}} = 0.38\mathrm{QoE}_{\mathrm{Network}} + 0.23\mathrm{QoE}_{\mathrm{Application}} - 0.19 \quad (28)$$

application QoE. R^2, adjustedR^2, and PCC values of 0.953, 0.952, and 0.976, respectively, are obtained for our final model. The modeling accuracy has been shown in Figure 19. For finding out the final model accuracy, we have used the remaining 78 sequences from the subjective dataset that have not been used for the purpose of model building.

Next, we present the same video model using an Artificial Neural Network (ANN) based approach.

8. ANN Based Video Model

Till now we have used a regression based technique for building the video quality prediction model. The model is able to predict the perceived video quality with reasonable accuracy. However, recently due to the widespread use of different machine learning techniques for analysis of data, we decided to use an Artificial Neural Network approach for building the same model limited to the same parameters

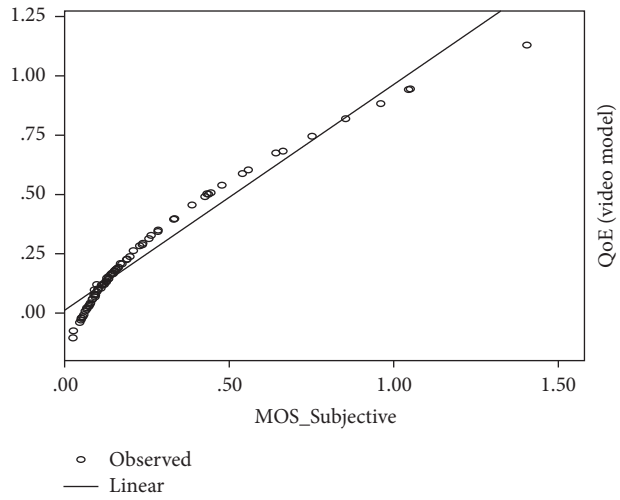

Figure 19: Accuracy of final video model.

that we have considered before and evaluate the performance of both. The same subjective data consisting of 78 impaired video sequences that we have used previously is taken in this case also.

Video quality estimation using different types of neural networks has been attempted by several researchers in the past. Probabilistic Neural Networks (PNN), Backpropagation Neural Network (BPNN), Adaptive Neurofuzzy Inference System (ANFIS), and Random Neural Networks (RNN) are some of the techniques commonly used. However, as pointed out in the literature review section, video quality assessment on mobile devices has been done with low resolution videos and using only the H.264 and MPEG-2/4 video codecs. In order to estimate the video quality from subjective metrics like MOS feedforward type ANNs are most commonly used [58–61]. This is the reason why we decided to use an ANN technique for this paper keeping in mind the current research gaps and trying to overcome those.

The ANN which we have used in our work is a multilayer perceptron model having one hidden layer. Considering the number of inputs that we have, that is, 7, going for more hidden layers would have increased the overall complexity of the system unnecessarily and also resulted in overfitting problems. Hence, we opted for the one hidden layer architecture. Training of the neural network has been done using the Levenberg-Marquardt (LM) algorithm by issuing the trainlm command in MATLAB. The trainlm command is a network training function that updates the weight and biases of the different nodes according to the LM optimization technique. It is considered to be one of the fastest back propagation algorithms and is highly recommended as a first-choice supervised algorithm, although it consumes more computer memory as compared to other algorithms. For the hidden layer, we have used a hyperbolic tangent sigmoid transfer function by issuing the tansig command. For the output layer, a pure linear transfer function is used by giving the purelin command. The neural network has the same 7 parameters that we have discussed previously as input, plus one extra factor for the type of codec used. As the output, we have the score that predicts the quality of the video.

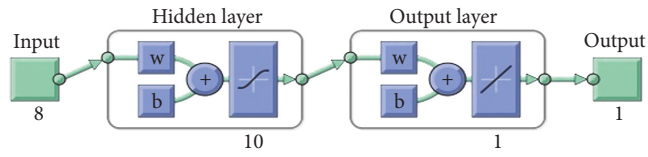

Figure 20: System architecture of the neural network.

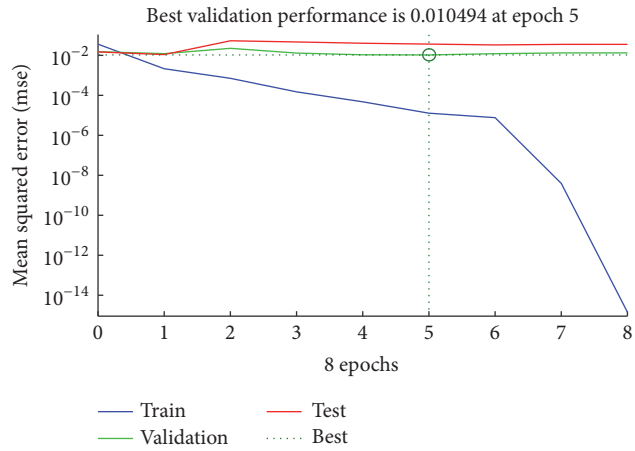

Figure 21: MSE variation across all sets.

Table 25: Weight and bias value for input layer.

Weight								Bias
PL	J	T	ARS	BR	FR	R	Codec	
0.769	0.073	0.644	−0.35	0.540	−0.20	−0.53	0.228	−2.18
0.099	0.342	1.006	0.182	0.933	0.682	1.657	0.101	1.44
1.085	0.183	0.256	0.383	−0.80	−0.10	−0.01	−0.28	−1.57
1.117	0.293	0.194	0.232	1.037	0.478	0.786	0.828	−1.13
−0.90	−0.31	0.238	0.900	−0.44	−0.24	0.847	−0.36	1.13
0.775	0.508	0.893	0.537	0.880	1.801	0.644	0.025	−0.23
0.106	0.351	0.694	−0.97	1.182	−0.53	0.566	0.002	−0.28
0.505	1.085	0.276	0.462	1.710	0.161	1.542	0.094	0.34
0.169	1.144	0.094	1.648	−0.59	1.476	0.571	−0.04	0.21
0.329	0.742	0.133	0.085	0.515	0.324	0.473	1.206	0.42

Table 26: Weight and bias value for hidden layer.

Weight										Bias
0.53	1.26	−0.3	−0.1	00.2	1.22	0.73	0.19	0.14	1.01	−0.8

We use a 70 : 30 split ratio for the input data as training, testing, and validation sets. To find the configuration of the network that achieves the best performance, several rounds of tests are conducted by varying the number of neurons in the hidden layer and observing the output. Since we have 8 inputs and 1 output, hence we varied the number of hidden neurons from 4 to 15. Optimal performance was observed with 10 hidden nodes. The system architecture showing the best configuration has been given in Figure 20. In the figure, the symbols w and b stand for the weight and bias factors for each node, respectively. The value of w and b for our configuration set for both input and hidden layer has been provided in Tables 25 and 26, respectively.

The performance of our model across the training, testing, and validation sets has been shown in Figure 21. The best

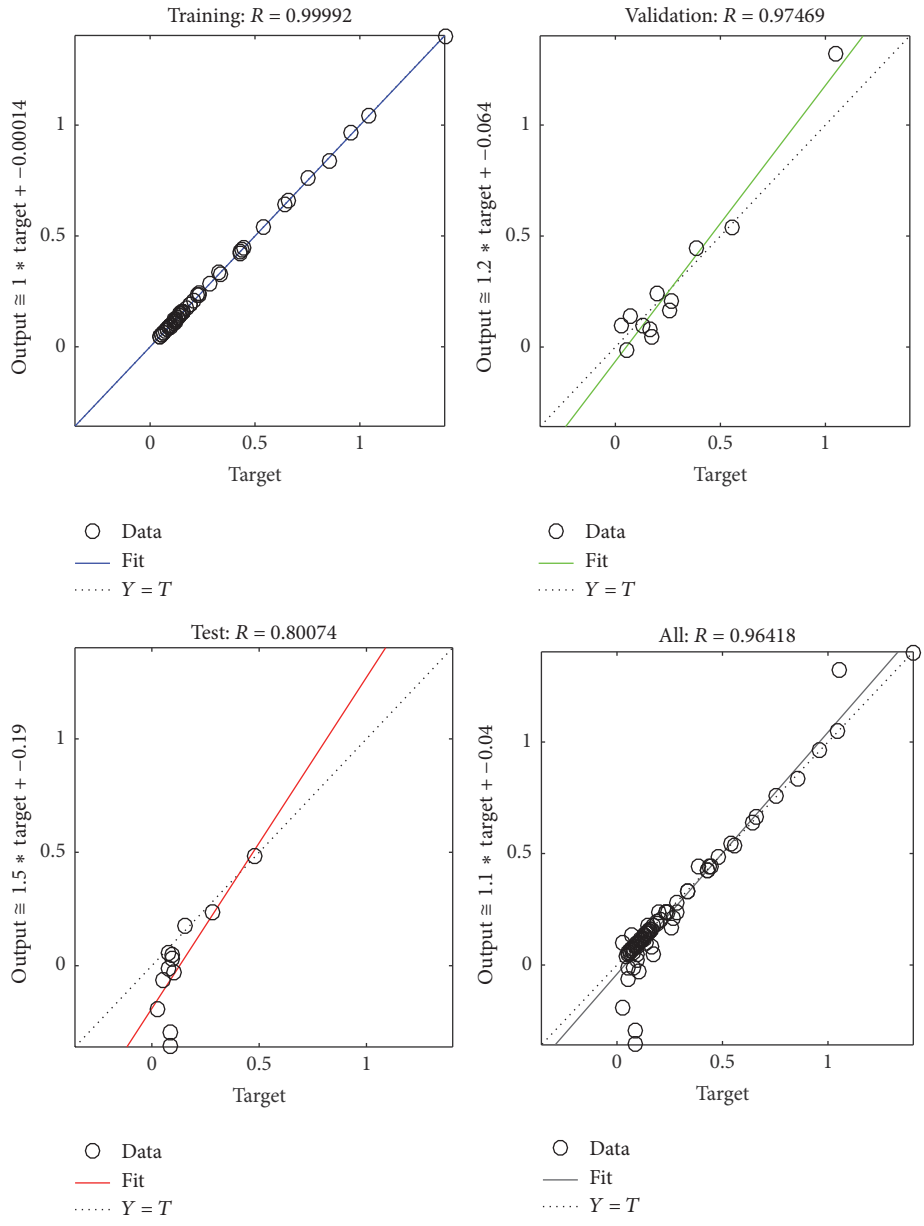

FIGURE 22: Regression plot across all sets.

validation performance is obtained at epoch 5 and marked in the figure. Also, we find that as the model learns during the training phase, the mean squared error (MSE) across all the three sets decreases and then becomes almost constant. Figure 22 shows the regression plot across all the 3 sets. The overall R^2 value for all the video sequences is 0.964 which is pretty high. The PCC value obtained is 0.984 at a significance level of less than 0.01. Compared to the regressive approach, the ANN model gives a slightly better performance.

9. Conclusion and Future Work

In this paper, we have proposed a no-reference video quality prediction model for relevant network and application QoS factors for a video streaming scenario. Our proposed model is a cross layered one as it takes into account factors from multiple layers of the TCP/IP protocol stack. At the same time, it has the unique characteristic of being a modular one. Depending upon the requirement, the model can be tuned for predicting the quality due to the network and application factors or a combination of both. At each stage, proper subjective tests have been done for the purpose of model building and validation. The ANN approach provides slightly better prediction accuracy as compared to the regression based approach.

All the videos that are used have Full HD resolution and encoded using the latest generation H.265/HEVC and VP9 codecs. Even though these codecs are capable of compressing videos at much higher resolutions up to 4K, we did not

consider them in this paper due to the limited availability of the 4K video contents. With more improved network speed and widespread availability of 4K video content, we plan to investigate the effect of higher resolutions in future work. Also, all the video sequences that were used had roughly 10 seconds duration. The effect of longer video sequences has not been investigated in this research, which would be considered in the future.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] *2013 Video Index-TV is no longer a single screen in your Living Room*, Ooyala Corp, USA, 2013.

[2] *Cisco Global Mobile Data Traffic Forecast Update Report 2014–2019*, Cisco Corp, USA, 2016.

[3] "Definitions of Terms related to Quality of Service," ITU-T Recommendation E.800, September, 2008.

[4] P. Le Callet, S. Möller, and A. Perkis, *Qualinet White Paper on Definitions of Quality of Experience- 2012*, Lausanne, Switzerland, 2012.

[5] T. Hoßfeld, R. Schatz, M. Varela, and C. Timmerer, "Challenges of QoE management for cloud applications," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 28–36, 2012.

[6] "Subjective Video Quality Assessment Methods for Multimedia Applications," ITU-T Recommendation P.910, June 200.

[7] "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU-T Recommendation BT.500, January 2012.

[8] "Methods for Subjective Determination of Transmission Quality," ITU-T Recommendation P.800, August 1996.

[9] F. M. Moss, K. Wang, F. Zhang, R. Baddeley, and D. R. Bull, "On the optimal presentation duration for subjective video quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, 2015.

[10] M. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," in *Proceedings of the Visual Communications and Image Processing 2003*, pp. 573–582, Switzerland, July 2003.

[11] D. M. Rouse, R. Pépion, P. Le Callet, and S. S. Hemami, "Tradeoffs in subjective testing methods for image and video quality assessment," in *Proceedings of the Human Vision and Electronic Imaging XV*, USA, January 2010.

[12] "Reference Algorithm for Computing Peak Signal to Noise Ratio of a Processed Video Sequence with Compensation for Constant Spatial Shifts, Constant Temporal Shift, and Constant Luminance Gain and Offset," TU-T Recommendation J.340, June 2010.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[14] "Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference," ITU-T Recommendation J.144, March 2001.

[15] "Objective Perceptual Video Quality Measurement Techniques for Standard Definition Digital Broadcast Television in the Presence of a Full Reference," ITU-R Recommendation BT.1683, June 2004.

[16] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. A. Vaishampayan, "Modeling Packet-Loss Visibility in MPEG-2 Video," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 341–355, 2006.

[17] J. Søgaard, S. Forchhammer, and J. Korhonen, "No-Reference Video Quality Assessment Using Codec Analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 10, pp. 1637–1650, 2015.

[18] M. Shahid, A. Rossholm, B. Lövström, and H.-J. Zepernick, "No-reference image and video quality assessment: a classification and review of recent approaches," *Eurasip Journal on Image and Video Processing*, vol. 2014, no. 1, article no. 40, 2014.

[19] "Opinion Model for Video Telephony Applications," *ITU-T Recommendation G.1070*, July 2012.

[20] M. Mu, P. Romaniak, A. Mauthe, M. Leszczuk, L. Janowski, and E. Cerqueira, "Framework for the integrated video quality assessment," *Multimedia Tools and Applications*, vol. 61, no. 3, pp. 787–817, 2012.

[21] "Parametric Bit-stream based Quality Assessment of Progressive Download and Adaptive Audio-visual Streaming Services over Reliable Transport," ITU-T Recommendation P.1203, January 2017, ITU-T Recommendation P.1203.

[22] "Objective Quality Measurement of Telephone Band (300-3400 Hz) Speech Codec," ITU-T Recommendation P.861, August 1996.

[23] S. Khirman and P. Henricksen, "Relationship between Quality of Service and Quality of Experience for Public Internet Services," in *Proceedings of the 3rd Workshop on Passive and Active Measurement*, pp. 23–28, March 2002.

[24] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, 2010.

[25] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo, "The logarithmic nature of QoE and the role of the Weber-Fechner law in QoE assessment," in *Proceedings of the IEEE International Conference on Communications (ICC '10)*, pp. 1–5, Cape Town, South Africa, May 2010.

[26] J. A. Lozano, A. Castro, B. Fuentes, J. M. González, and Á. Rodríguez, "Adaptive QoE measurement on Video streaming IP services," in *Proceedings of the 7th International Conference on Network and Service Management*, pp. 1–4, Paris, fra, 2011.

[27] H. J. Kim and S. G. Choi, "QoE assessment model for multimedia streaming services using QoS parameters," *Multimedia Tools and Applications*, pp. 1–13, 2013.

[28] W. Song, D. Tjondronegoro, and M. Docherty, "Exploration and optimization of user experience in viewing videos on a mobile phone," *International Journal of Software Engineering and Knowledge Engineering*, vol. 20, no. 8, pp. 1045–1075, 2010.

[29] H. Knoche, J. D. McCarthy, and M. A. Sasse, "Can small be beautiful? assessing image resolution requirements for mobile TV," in *Proceedings of the 13th ACM International Conference on Multimedia, MM 2005*, pp. 829–838, Singapore, November 2005.

[30] Y.-F. Ou, Y. Xue, Z. Ma, and Y. Wang, "A perceptual video quality model for mobile platform considering impact of spatial, temporal, and amplitude resolutions," in *Proceedings of the 2011 IEEE 10th IVMSP Workshop: Perception and Visual Signal Analysis, IVMSP 2011*, pp. 117–122, usa, June 2011.

[31] W. Song, Y. Xiao, D. Tjondronegoro, and A. Liotta, "QoE modelling for VP9 and H.265 videos on mobile devices," in *Proceedings of the 23rd ACM International Conference on Multimedia, MM 2015*, pp. 501–510, Australia, October 2015.

[32] A. Khan, L. Sun, and E. Ifeachor, "Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks," in *Proceedings of the 2009 IEEE International Conference on Communications, ICC 2009*, Germany, June 2009.

[33] H. Koumaras, A. Kourtis, C.-H. Lin, and C.-K. Shieh, "A theoretical framework for end-to-end video quality prediction of MPEG-based sequences," in *Proceedings of the 3rd International Conference on Networking and Services, ICNS 2007*, Greece, June 2007.

[34] Z. Duanmu, A. Rehman, K. Zeng, and Z. Wang, "Quality-of-experience prediction for streaming video," in *Proceedings of the 2016 IEEE International Conference on Multimedia and Expo, ICME 2016*, USA, July 2016.

[35] P. Calyam, E. Ekici, C.-G. Lee, M. Haffner, and N. Howes, "A "GAP-model" based framework for online VVoIP QoE measurement," *Journal of Communications and Networks*, vol. 9, no. 4, pp. 446–455, 2007.

[36] A. Khan, L. Sun, and E. Ifeachor, "Learning models for video quality prediction over wireless local area network and universal mobile telecommunication system networks," *IET Communications*, vol. 4, no. 12, pp. 1389–1403, 2010.

[37] D. Yun and K. Chung, "DASH-based Multi-view Video Streaming System," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1-1.

[38] S. Colonnese, F. Cuomo, T. Melodia, and I. Rubin, "A Cross-Layer Bandwidth Allocation Scheme for HTTP-Based Video Streaming in LTE Cellular Networks," *IEEE Communications Letters*, vol. 21, no. 2, pp. 386–389, 2017.

[39] K. Jia, Y. Guo, Y. Chen, and Y. Zhao, "Measuring and Predicting Quality of Experience of DASH-based Video Streaming over LTE," in *Proceedings of the 19th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 102–107, Shenzhen, China, 2016.

[40] T. Maki, M. Varela, and D. Ammar, "A Layered Model for Quality Estimation of HTTP Video from QoS Measurements," in *Proceedings of the 11th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2015*, pp. 591–598, tha, November 2015.

[41] J. Jiang, V. Sekar, and H. Zhang, "Improving fairness, efficiency, and stability in HTTP-based adaptive video streaming with festive," *IEEE/ACM Transactions on Networking*, vol. 22, no. 1, pp. 326–340, 2014.

[42] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A tutorial on video quality assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[43] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

[44] P. Juluri, T. Venkatesh, and D. Medhi, "Measurement of quality of experience of video-on-demand services: a survey," *IEEE Communications Surveys & Tutorials*, 2015.

[45] "VQEG Standard Database maintained," http://www.its.bldrdoc.gov/vqeg/downloads.aspx.

[46] J. Nightingale, Q. Wang, C. Grecos, and S. Goma, "The impact of network impairment on quality of experience (QoE) in H.265/HEVC video streaming," *IEEE Transactions on Consumer Electronics*, vol. 60, no. 2, pp. 242–250, 2014.

[47] "Network Performance Objectives for IP-based Services," TU-T Recommendation Y.1541, December 2011.

[48] K. Gu, J. Zhou, J.-F. Qiao, G. Zhai, W. Lin, and A. C. Bovik, "No-reference quality assessment of screen content pictures," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 4005–4018, 2017.

[49] H. Malekmohamadi, W. A. C. Fernando, and A. M. Kondoz, "Content-based subjective quality prediction in stereoscopic videos with machine learning," *IEEE Electronics Letters*, vol. 48, no. 21, pp. 1344–1346, 2012.

[50] T. Ghalut, H. Larijani, and A. Shahrabi, "Content-based video quality prediction using random neural networks for video streaming over LTE networks," in *Proceedings of the 15th IEEE International Conference on Computer and Information Technology, CIT 2015, 14th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2015, 13th IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC 2015 and 13th IEEE International Conference on Pervasive Intelligence and Computing, PICom 2015*, pp. 1626–1631, gbr, October 2015.

[51] "Framework and Methodologies for the Determination and Application of QoS Parameters," ITU-T Recommendation E.802, February 2007.

[52] "Speech and Multimedia Transmission Quality (STQ); End-to-End Jitter Transmission Planning Requirements for Real Time Services in an NGN Context," ETSI TR 103 210 v.1.1.1 (2013-10) Recommendation, 2013.

[53] J. Klaue, B. Rathke, and A. Wolisz, "Evalvid–a framework for video transmission and quality evaluation," in *Computer Performance Evaluation. Modelling Techniques and Tools*, vol. 2794 of *Lecture Notes in Computer Science*, pp. 255–272, Springer, New York, NY, USA, 2003.

[54] "NS2," http://www.isi.edu/nsnam/ns/.

[55] L. H. Hardy, G. Rand, and M. C. Rittler, "Tests for the Detection and Analysis of Color-Blindness III The Rabkin Test," *Journal of the Optical Society of America*, vol. 35, no. 7, p. 481, 1945.

[56] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, NY, USA, 1980.

[57] F. Zahedi, "The analytic hierarchy process-a survey of t he method and its applications," *Interfaces*, vol. 16, no. 4, pp. 96–108, 1986.

[58] X. Zhang, L. Wu, Y. Fang, and H. Jiang, "A study of FR video quality assessment of real time video stream," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 6, 2012.

[59] P. Reichl, S. Egger, S. Moller et al., "Towards a comprehensive framework for QOE and user behavior modelling," in *Proceedings of the 17th International Workshop on Quality of Multimedia Experience (QoMEX '15)*, pp. 1–6, IEEE, Pylos-Nestoras, Greece, May 2015.

[60] L. Pierucci and D. Micheli, "A Neural Network for Quality of Experience Estimation in Mobile Communications," *IEEE MultiMedia*, vol. 23, no. 4, pp. 42–49, 2016.

[61] E. Danish, M. Alreshoodi, A. Fernando, B. Alzahrani, and S. Alharthi, "Cross-layer QoE prediction for mobile video based on random neural networks," in *Proceedings of the IEEE International Conference on Consumer Electronics, ICCE 2016*, pp. 227-228, USA, January 2016.