

Research Article

Nonintrusive Method Based on Neural Networks for Video Quality of Experience Assessment

Diego José Luis Botia Valderrama and Natalia Gaviria Gómez

Engineering Department, Universidad de Antioquia, Medellín, Colombia

Correspondence should be addressed to Diego José Luis Botia Valderrama; diego.botia@gmail.com

Received 11 August 2015; Revised 4 December 2015; Accepted 17 December 2015

Academic Editor: Stefanos Kollias

Copyright © 2016 D. J. L. Botia Valderrama and N. Gaviria Gómez. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The measurement and evaluation of the QoE (Quality of Experience) have become one of the main focuses in the telecommunications to provide services with the expected quality for their users. However, factors like the network parameters and codification can affect the quality of video, limiting the correlation between the objective and subjective metrics. The above increases the complexity to evaluate the real quality of video perceived by users. In this paper, a model based on artificial neural networks such as BPNNs (Backpropagation Neural Networks) and the RNNs (Random Neural Networks) is applied to evaluate the subjective quality metrics MOS (Mean Opinion Score) and the PSNR (Peak Signal Noise Ratio), SSIM (Structural Similarity Index Metric), VQM (Video Quality Metric), and QIBF (Quality Index Based Frame). The proposed model allows establishing the QoS (Quality of Service) based in the strategy *Diffserv*. The metrics were analyzed through Pearson's and Spearman's correlation coefficients, RMSE (Root Mean Square Error), and outliers rate. Correlation values greater than 90% were obtained for all the evaluated metrics.

1. Introduction

The assessment of quality in digital video systems is a topic of great interest to the telecommunications companies that hope to increase the quality to their users. The QoE (Quality of Experience) is the degree of user's satisfaction with any kind of multimedia service. This concept has been defined in different ways by various authors. Liu et al. [1] suggested that QoE involves two aspects:

- (i) The monitoring of the user's experience online.
- (ii) The service control to ensure that QoS (Quality of Service) can satisfy the user's requirements.

QoE is an extension of QoS, since the former provides information about the services delivery from point of view of the end users. QoE refers to personal preferences of users and so seeks to assess the subjective perception of the received service [2, 3]. However, this perception is influenced, by the network performance in terms of QoS and video encoding parameters. Different methodologies proposed in the literature aim at estimating subjective Quality of Experience,

through the assessment of different metrics of video quality generally using objective methods.

In the implementation of digital television platforms (e.g., IPTV and DVB), some important restrictions can affect the management and proper operation of the network. Some of these are

- (i) the large amount of bandwidth the user should contract,
- (ii) the limitation of internal buffers in routers and STB (Set Top Box), which can generate problems such as packet loss that are critical on video or audio transmission,
- (iii) the type of video compression format, which will reduce the channel use without affecting the quality,
- (iv) other items that should be installed and properly configured, such as the last mile link used and the admission control class used.

Different researches have proposed quality assessment models for video streams and measurement strategies in

order to identify the optimal values for each metric guaranteeing the experience of the viewer.

According to Winkler [4], there are projects for QoE assessment: VQEG (Video Quality Experts Group), QoS (Quality of Service Metrics) from ATIS IPTV Interoperability Forum, and specific metrics such as those oriented to packets, a bitstream, hybrids, and images metrics [5–8]; but these are more complex, and the correlation methods have not yet been applied for real time services assessment. Other works propose a new relationship between KPIs (Key Performance Indexes) with QoE assessment on mobile environments, but it can be applied in other scenarios as fixed broadband networks focused particularly on telecommunications providers [9].

One of the main problems on the estimation of the video subjective quality is the lack of proper estimation or correlation models. These must guarantee results with accuracy but in many cases heavily depend on objective metrics [10–12]. Also, the correlation models in QoE metrics are not accurate and reliable.

According to [13], three strategies were developed to perform the estimation of video quality. The first one is to apply a subjective assessment with a selected group of people; the main drawback of the evaluations is the cost and time. The second one is the objective quality metrics assessment for video, in which the principal disadvantage is the low correlation with regard to subjective quality metrics [14]; in addition, such metrics obviate the network and content parameters. The last one is to use machine learning methods to analyse the objective and subjective assessment but the major drawback is the difficult for configuring and testing; furthermore, some methods may fail whether a suitable design and optimal parameter selection are not performed.

According to Kuipers et al. [15], the minimum threshold of accepted quality is a MOS (Mean Opinion Score) value of 3.5. The subjective tests (such as MOS) are widely useful in assessing the users satisfaction due to its accuracy; however, the application of them is still very complex due to the high consumption of time and money; therefore, they are impractical for tasks of testing in network devices, and a controlled environment is required (in some cases, its implementation are complex).

Also if we want to develop traffic management techniques in real time, it is necessary to find a relationship among them and the objective metrics, measurable by the network equipment.

Objective methods are based on algorithms for the assessment of video quality, making them less complex and, furthermore, can be performed on controlled simulation environments. The objective metrics are supported in mathematical models that approximate themselves to the Human Vision System (HVS) behavior and, therefore, try to estimate as accurately as possible the true QoE. However, the perception of each viewer is highly influenced by the quality of the data network, expressed by the QoS parameters [26]. A lot of proposals for assessing QoE metrics have been created to define the user's experience. ITU-T has carried out standards for some of them [27, 28].

Due to complex factors such as the HVS, different kinds of solutions such as the implementation of machine learning

techniques have been proposed. Some of the most used methods are Artificial Neural Network (ANN), fuzzy logic based on rules, neural-fuzzy networks (e.g., ANFIS), support vector machines (SVM), Gaussian processes, and genetic algorithm, among others. Nonintrusive QoE estimation methods for video are mainly based on application layer and network parameters.

Models as artificial neural networks have been little studied for estimation and prediction of video quality. For the evaluation of nonintrusive methods, we decided to employ two methods of machine learning: BPNN feedforward and RNN (Random Neural Network). We assess the output of each system, estimated MOS versus expected MOS. In each case, the fit of the correlation, determined by Pearson and Spearman rank order correlation coefficients, RMSE (Root Mean Square Error), and Outliers Ratio were calculated. The expected MOS was calculated from the VQM (Video Quality Metric), SSIM (Structural Similarity Index Metric), PSNR (Peak Signal Noise Ratio), and QIBF (Quality Index Based Frame) metrics.

This paper presents the key objective and subjective quality metrics and introduces a nonintrusive QoE assessment methodology based on machine learning techniques, showing its main features and functionality. Afterwards, the developed testbed is explained; both results obtained as their analyses are presented. Finally the main conclusions and further works will be shown.

2. Related Works

2.1. Methods for Quality of Experience Assessment. In spite of all proposals for objective metrics, they are always not close to the human perception, due to the fact that the perception is highly influenced by the performance of the network, defined in terms of QoS parameters. According to [29, 30], the objective metrics are computational models that predict the image quality perceived by any person and can be classified as intrusive and nonintrusive methods, as shown in Figure 1.

The Subjective Pseudo Quality Assessment model or PSQA (Pseudo Subjective Quality Assessment) is an example of nonintrusive method (NR). This model uses an RNN (Random Neural Network) [31–33] to learn and recognize the relationship between video and the characteristics of the network with the quality perceived by users.

Initially, for the training process of the RNN, a database is required which contains different sequences to assess the distortions generated by several QoS and coding parameters. Afterwards, the training of the RNN with any video sequence is evaluated in order to validate the MOS measure.

2.2. QoE Metrics

2.2.1. PSNR and MSE Metrics. The PSNR (Peak Signal Noise Ratio) and MSE (Mean Square Error) metrics are the most frequently applied ones [11]. They assess the quality of the received video sequence and, thus, can be mapped on a subjective scale PSNR aiming to compare pixel-by-pixel and frame-by-frame the quality of the received image with the

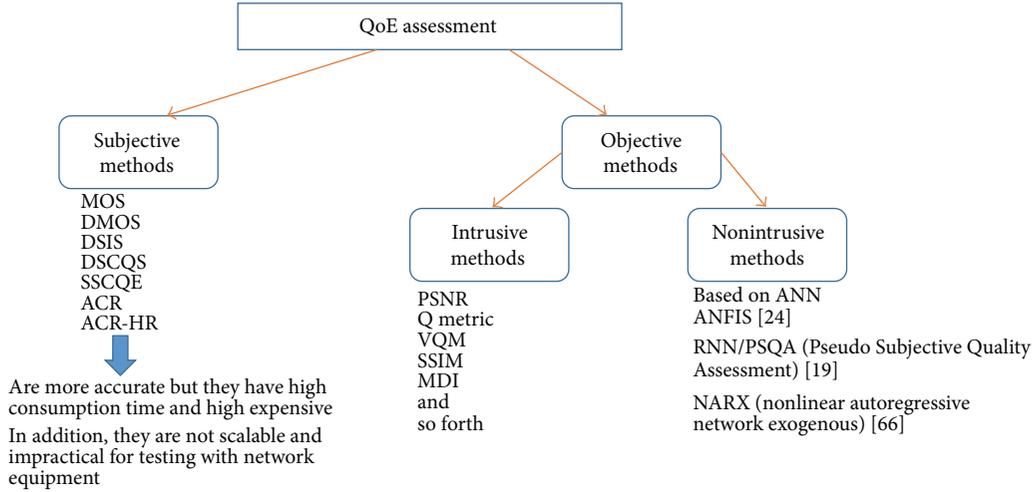


FIGURE 1: Methods for Quality of Experience assessment.

source image. It is the most known FR (Full Reference) metric. If we consider frames with a size of $M \times N$ pixels and 8 bits/sample, the PSNR can be calculated using (1) according to [34, 35] as follows:

$$\text{PSNR} = 20$$

$$\cdot \log_{10} \left(\frac{255}{\sqrt{(1/(M \times N)) \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \|Y_s(i, j) - Y_d(i, j)\|^2}} \right), \quad (1)$$

where $Y_s(i, j)$ denotes a pixel in the (i, j) position of the original frame and $Y_d(i, j)$ refers to the pixel located at (i, j) position of the frame reconstructed in the receiver side. Around 255 elements are the maximum value that the pixel can take (255 for 8-bit images). The denominator is known as MSE or Mean Square Error, which is the mean square of the differences among the grey level values of the pixels into the pictures or sequences Y_s and Y_d .

Since several studies have used this mapping, PSNR is limited by the image content and it is not able to identify artifacts due to packet loss. In addition, the measure does not always correlate with the real user's perception due to the fact that a pixel-by-pixel comparison is carried out without performing an analysis of the image structural elements (e.g., contours or specific distortions introduced by either encoders or transmitting devices on the network or spatial and temporal artifacts). Therefore, some metrics are proposed to generate the extraction and analysis of the features and artifacts into video sequences such as SSIM metric [36].

2.2.2. SSIM Metric. Assuming that the human visual perception is highly adapted for extracting structures from a scene, SSIM (Structural Similarity Index) calculates the mean, variance, and covariance between the transmitted and received frames [37]. To apply SSIM, three components (luminance, contrast, and structural similarities) are measured and combined into one value called SSIM index ranging between -1 and 1 , where a negative or 0 value indicates zero

correlation with the original image and 1 means that it is the same image [35].

According to Wang et al. [38], SSIM gives a good approximation of the image distortion due to changes in the measurement of structural information. On the video sequences, this metric considers a wide range of scenes complexity, in terms of movement, spatial details, and color. According to Wang et al. [37], this metric uses a structural distortion measure instead of the error. The above is focused on the human vision system to extract structural information from the visual field and ignored the extraction of errors. Therefore, a calculation of the structural distortion should give a better correlation with subjective metrics. In [39, 40], a simple and effective algorithm was proposed to calculate the SSIM index.

Let x be the original signal, $x = \{x_i \mid i = 1, 2, \dots, N\}$, and let y be the distorted signal, $y = \{y_i \mid i = 1, 2, \dots, N\}$; the similarity structure index is given by

$$\text{SSIM} = \frac{(2\bar{x}\bar{y} + C1)(2\sigma_{x,y} + C2)}{[(\bar{x})^2 + (\bar{y})^2 + C1](\sigma_x^2 + \sigma_y^2 + C2)}, \quad (2)$$

where \bar{x} is the mean of x , \bar{y} is the mean of y , σ_x^2 and σ_y^2 are the variances of x and y , $\sigma_{x,y}$ is the covariance of x and y , and $C1$ and $C2$ are constant values. The SSIM value can be defined as

$$\text{SSIM}(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma, \quad (3)$$

where $l(x, y)$, $c(x, y)$, and $s(x, y)$ are comparison functions of the luminance, contrast, and structure components, and the parameters α , β , and $\gamma > 0$ are constants. $\text{SSIM}(x, y)$ is the quality assessment index. For more details see [39].

2.2.3. VQM Metric. The NTIA VQM metric (Video Quality Metric) [39] considers two image inputs, the original and processed video, in which the quality levels are verified through the human vision system and some subjective aspects. This metric divides the image into sequences of space-temporal

TABLE 1: Comparative table of the main QoE metrics.

Metric	Standard	Class	Estimate	Type
PSNR		Objective	Pixel-by-pixel comparison between reference image and compressed image.	FR
VQM (Video Quality Metric)	NTIA	Objective	Sequences are divided into temporal-space blocks. Calculate the orientation of each block using spatial luminance gradient. Score tends to 0 (best). Assess blurring, global noise, block distortion, and color distortion	RR
SSIM (Structural Similarity Index)		Objective	Find the mean, variance, and covariance within a frame and combine them in a distortion map. Use luminance, contrast, and structure similarity. Use decimal value from 0 to 1 (high correlation with original picture).	FR
MOS	ITU-T P.800	Subjective	Subjective scale from 1 (poor) to 5 (good).	NA

blocks, measuring elements like blurring, general noise, block distortion, color distortion, and mix among them into a single metric. The score closer to 0 is considered the best possible value. According to Wang [41], this metric shows a good correlation with subjective methods and has also been adopted by ANSI as a standard for the assessment of video quality.

In Table 1, the comparative table of the main QoE metrics is summed up, which is used in this work.

2.2.4. New Mapping QoE Metrics. Zinner et al. [42] proposed a framework for the assessment of the QoE by using streaming video systems. On the other hand, Botia et al. [16] proposed a new mapping among the PSNR, SSIM, VQM, and MOS metrics, shown in Table 2. In our simulations, we calculated the average of each MOS for all video sequences with the value of each FR metric from Table 2 and then we obtained the expected MOS.

2.2.5. QIBF Metric. Serral-Garcia et al. [43] propose a framework called PBQAF (Profile Based QoE Assessment Framework). This framework defines three states for frames (correct, altered, and lost) through the analysis of the payload. Moreover, a mapping is performed to generate and associate the quality index. This metric is calculated from payload of the received packets, associated with a PLR (Packet Loss Rate) in particular. Equation (4) shows the quality function to generate the mapping function M :

$$Q(f) = M(\text{PLR}(f)), \quad (4)$$

with $\text{PLR}(f)$ being the packet loss rate of the frame f . The mapping function is given by

$$M(x) = 1 - x, \quad (5)$$

where x shows the ratio of lost frames; therefore, when a high packet loss rate is measured, the quality index will be lower and tends to be 0. Thus, the final quality of the video stream into a set of quality values $q(f_x)$ for particular frames is given by

$$Q_f = \{q(f_1), \dots, q(f_n)\}, \quad (6)$$

where $f_{1,\dots,n}$ is the set of all frames in the video stream. From Botia et al. [44], a mapping between QIBF (Quality Index

TABLE 2: Mapping among QoE metrics (SSIM, PSNR, VQM, and MOS) [16].

MOS	PSNR (dB)	SSIM	VQM
5 (excellent)	≥ 37	≥ 0.93	< 1.1
4 (good)	$\geq 31 - < 37$	$\geq 0.85 - < 0.93$	$\geq 1.1 - < 3.9$
3 (fair)	$\geq 25 - < 31$	$\geq 0.75 - < 0.85$	$\geq 3.9 - < 6.5$
2 (poor)	$\geq 20 - < 25$	$\geq 0.55 - < 0.75$	$\geq 6.5 - < 9$
1 (bad)	< 20	< 0.55	≥ 9

TABLE 3: QIBF versus MOS mapping [16].

MOS	QIBF value
5 (excellent)	≥ 0.85
4 (good)	$\geq 0.65 - < 0.85$
3 (fair)	$\geq 0.45 - < 0.65$
2 (poor)	$\geq 0.25 - < 0.45$
1 (bad)	< 0.25

Based Frame) metric and MOS metric is proposed for each frame class $I/P/B$; that is, each frame is equivalent to one class. In (7), the QIBF is given by

$$\text{QIBF}(\text{seq}, \text{nt}) = \sum_{f=1}^F \frac{1 - \text{NFL}(f)}{3} - \alpha, \quad (7)$$

where F is the maximum number of frames, seq is the sequence of actual video to be evaluated, nt refers to the class network applied over the transmitted sequence (*BestEffort* or *DiffServ*), $\text{NFL}(f)$ is the number of frames ($I/P/B$) lost, determined by the set of MPEG images, and α is an adjustment factor set to 0.05, obtained from several developed tests.

The factor $1/3$ in (7) is defined through 3 classes of frames used in the tests. As $\text{NFL}(f) \in [0, 100]$, the value of NFL is divided by 100 in order to define $\text{NFL}(f) \in [0, 1]$ and, therefore, to calculate $\text{QIBF}(\text{seq}, \text{nt}) \in [0, 1]$. The demonstration is shown in Appendix A.

Table 3 shows the proposed mapping between quality index and MOS metric, where $\text{QIBF}(f) \in [0, \dots, 1]$ and, as observed in Table 3, for an index value $\text{QIBF}(f) > 0.65$, a good value for the MOS metric is obtained, indicating a video sequence with a minimum of artifacts.



FIGURE 2: Screenshots from video sequences assessed.

3. Simulation Testbed

For our test, a testbed is built up to use a simulation software tool *NS-2* and the framework *Evalvid* for assessing a set of video sequences. The results obtained from FR/RR metrics are evaluated to find the possible correlation with subjective metrics using machine learning techniques.

In the simulation, we used a selection of different video raw uncompressed sequences in format *YUV* with color mode or sampling $4:2:0$, encoded with the *ffmpeg* and main concept software tools, to adapt them to different bitrates and GOP (Group of Pictures) lengths. Four video sequences were initially assessed with different levels of movement, encoded in the MPEG-4 format, which were adapted to be transmitted by a simulated IP network.

Figure 2 illustrates some screenshots of the evaluated videos (News of the Spanish Public TV, Mass, Highway and Winter Tree), converted and encoded at a resolution of 720×480 pixels (standard definition) under the NTSC standard, with frame rate of 30 fps. For each video stream, several parameters were combined as length of the GOP (10, 15, and 30), the bit rates recommended by DSL Forum [45] (1.5, 2, 2.5, and 3 Mbps), and packet loss rate for both networks (*BestEffort* and *Diffserv* using the congestion control algorithm, *WRED*), which produced 385 different video sequences for testing [16].

The generated video traces were adapted to be sent to the data network through the encapsulation of each packet with an MTU (Maximum Transfer Unit) of 1024 bits, using the RTP protocol (Real Time Transport Protocol) with *MP4trace* software tool. Considering the simulation tool *NS-2* and *Evalvid* framework [46], the sender and receiver trace files that were created, to calculate the sent and received lost frames and packets, delays, and jitters. The above facilitates the analysis of each video sequence for both implemented scenarios (*BestEffort* and *Diffserv* data networks).

The Evalvid framework also supports PSNR and MOS metrics and has a modular structure, making it easily adaptable to any simulation environment. MSU VQMT software tool [47] allows getting the Y-PSNR, SSIM, and VQM metrics values through the original reference video and received video with distortion.

The simulation scenario is composed of a video sender (server for video on demand) and 9 cross-traffic sources, which consist of *CBR* and *On-Off* traffic sources. The network is based on dumbbell topology [48] (see Figure 3). In our test, we send several video packets over a network with congestion and will be allowed to test the defined QoS scheme (*Diffserv with WRED*). The MPEG-4 video flow is complete with background *On-Off* traffic flows, which has an exponential distribution with an average packet size of 1500 bytes, burst time of 50 ms, idle time of 0.01 ms, and sending rate of 1 Mbps [16, 44]. The access network is represented by a video receiver (simulating a last mile with ADSL2), with a bandwidth link of 10 Mbps and several receiving nodes (sink) for cross traffic with a bandwidth of 10 Mbps for each one. Transmission distortions were simulated at different PLR (Packet Loss Rate). The traffic behavior and QoE metrics were tested with several error rates over a link established between the core and edge routers, using a loss model with uniform distribution at rates of 0%, 1%, 5%, and 10% and delay of 5 ms. Tables 4 and 5 show the main parameters used in the simulation and encoding.

4. Implementation of Nonintrusive Methods to Estimate Video Quality by Objective and Subjective Metrics

For the evaluation of nonintrusive methods, we decided to use two machine learning methods (BPNN feedforward and Random Neural Network). These methods have been used in different environments described in the next section. To

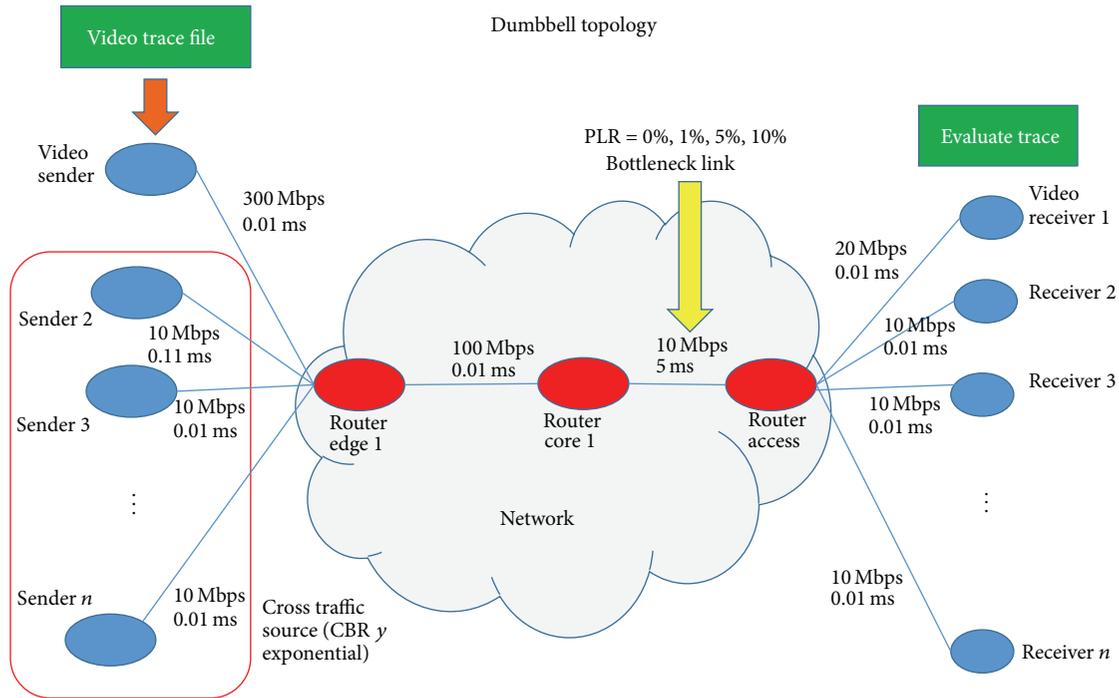


FIGURE 3: Simulation scenery with NS-2 and Evalvid.

TABLE 4: Encoder parameters.

GOP length	10, 15, and 30 frames
Frame rate	30 fps
Bitrate	1.5, 2, 2.5, and 3 Mbps
GOP sequence	IPBBPBBPBBP...
Resolution	NTSC 720 × 480 p
Video color mode	YUV (4 : 2 : 0)

TABLE 5: Simulation parameters.

Bandwidth link	10/100 Mbps
Link delay	0.01 ms–5 ms
Buffer size	50
Tail behavior	DropTail/WRED random early detection
Packet size	1052 bytes (8 bytes UDP and 20 bytes IP)
Max fragment size	1024 bytes
PLR, loss model with uniform distribution	0%, 1%, 5%, and 10% and delay of 5 ms

assess the output of each system, we considered the following: estimates MOS versus expected MOS, where the latter is computed through the mapping of PSNR, VQM, SSIM, and QIBF metrics (see Tables 1 and 2).

In each case, the correlation adjustment was calculated, determined by the Pearson correlation coefficient and RMSE to estimate the error. The general methodology is presented in Figure 4 [26, 49].

Initially, the video sequences in RAW-format were obtained and they were codified in MPEG-4/AVC-format. Each sequence was sent through an IP-simulated network. The above is based on “BestEffort” and “Diffserv” where each FR/RR metric is calculated and their respective values are mapped. In this manner, average MOS value is obtained by using Tables 2 and 3. Considering the codification values, the kind of QoS network and the obtained FR/RR metrics allow building the input database for training two neural networks.

MATLAB software suite was used for BPNN, using the Neural Networks Toolbox. For RNNs case, we used the QoE-NNR software tool [50]. From 385 different sequences at different configured parameters, around 70% were used as training data and the remaining 30% as test and validation data. Finally, obtaining the estimated MOS value, the correlation is calculated with respect to the average MOS value. The given results of machines learning will be presented and discussed.

Table 7 shows a brief state of the art where several papers assess QoE using neural networks as PNN (Probabilistic Neural Network), BPNN (Backpropagation Neural Network), RNN (Random Neural Network), and ANFIS (Adaptive Neurofuzzy Inference System). Works, using a few video sequences, do not show resolutions or codec information and apply low resolutions (for mobile devices). Codec formats more used are H.264 and MPEG-4 AVC. We found that PSNR, SSIM, and VQM metrics are frequently applied and few works use 2 metrics or more. In our case, we were using 5 different metrics and 4 video sequences with motion levels in each scene. Finally, the machine learning based on neural networks (RNN, BPNN, and ANFIs) is used in these works. The results show a good performance of the MOS estimation.

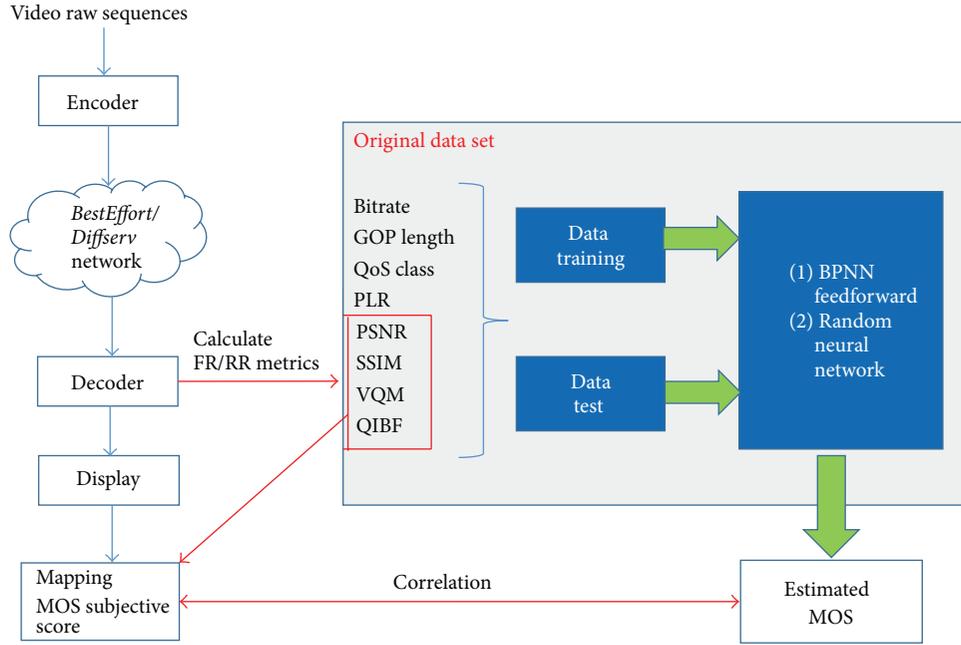


FIGURE 4: Proposed methodology to estimate the MOS metric using machine learning techniques.

However, they are defined by few input parameters, only assessing 1 or 2 video sequences in low resolution(s). They do not compare with other kinds of ANNs and, in most cases, the Pearson correlation coefficient was <0.90 .

4.1. Case 1. Implementation of a Feedforward Artificial Neural Network with Backpropagation. The ANNs are a paradigm for processing information inspired by the human neural system. Usually, ANNs are composed of a large number of highly interconnected processing elements called neurons, which work together to solve problems [13]. The base is the creation of a neural network also called MLP (Multilayer Perceptron), usually divided into 3 or N -layers.

The first layer contains neurons connected to the input vector data; the second layer is called the hidden layer and includes a set of synapses and a number of weights W_{ij} and some activating functions defined for exciting or inhibiting each neuron, generating a response. According to Rubino et al. [51], if the number of hidden neurons is low, it can have large training and generalization errors due to the underfitting. Otherwise, if there are many neurons in the hidden layer, low training errors may occur, but high generalization errors may appear, causing the undesired effect of overtraining (overfitting) and high variance. The third layer is the output layer directly connected to the hidden layer where data vectors of each estimated output will be obtained. Multiple layers of neurons with nonlinear transfer functions (such as tangential-sigmoid) allow the ANNs learning the linear and nonlinear relationships between the inputs (PLR, GOP, bitrate, QoS class, QIBF, PSNR, SSIM, and VQM) and the desired output vectors (MOS). The structure of Backpropagation Neural Network is shown in Figure 5.

ANNs type feedforward are the most commonly used ones to perform estimation of subjective metrics. Several

works propose different models based on ANN [52–58]. These approaches are generally applied on mobile systems or with low resolutions (QCIF or CIF); furthermore, these works obviate the network parameters or video objective metrics. In most cases, one or two metrics as input to the network are used, and the results from Pearson's autocorrelations almost always are below 0.90.

According to Ding et al. [52], the neural network can be used to obtain mapping functions between objective quality assessment and subjective quality assessment indexes. This affirmation allows the understanding of the usefulness of the ANN to analyse the estimated MOS and the proposed model presented in this research.

In that case, the network training was carried out through several parameters, which will turn into input variables. As stated, the objective parameters may affect the video quality. After training, an evaluation with a set of test data (96 sequences) will be performed in order to reach the corresponding network validation. The idea is to reach the lowest error and to be able to correlate the estimated MOS by ANN versus the average MOS computed from the objective metrics defined by Tables 1 and 2.

For the case of study, we want to build the estimated MOS function, defined by

$$\text{MOS_Est} = f(x_1, x_2, \dots, x_n), \quad (8)$$

where $\{x_1, x_2, \dots, x_n\}$ are the input parameters established by bitrate, packet loss rate, GOP length, QoS class (1 for *BestEffort* and 2 for *Diffserv*), SSIM, PSNR, VQM, and QIBF.

Like a human being, the ANN system needs a learning phase and another one for validation and testing to establish when the neural network generalized its learning for any data set. For this process, the network is trained through a training

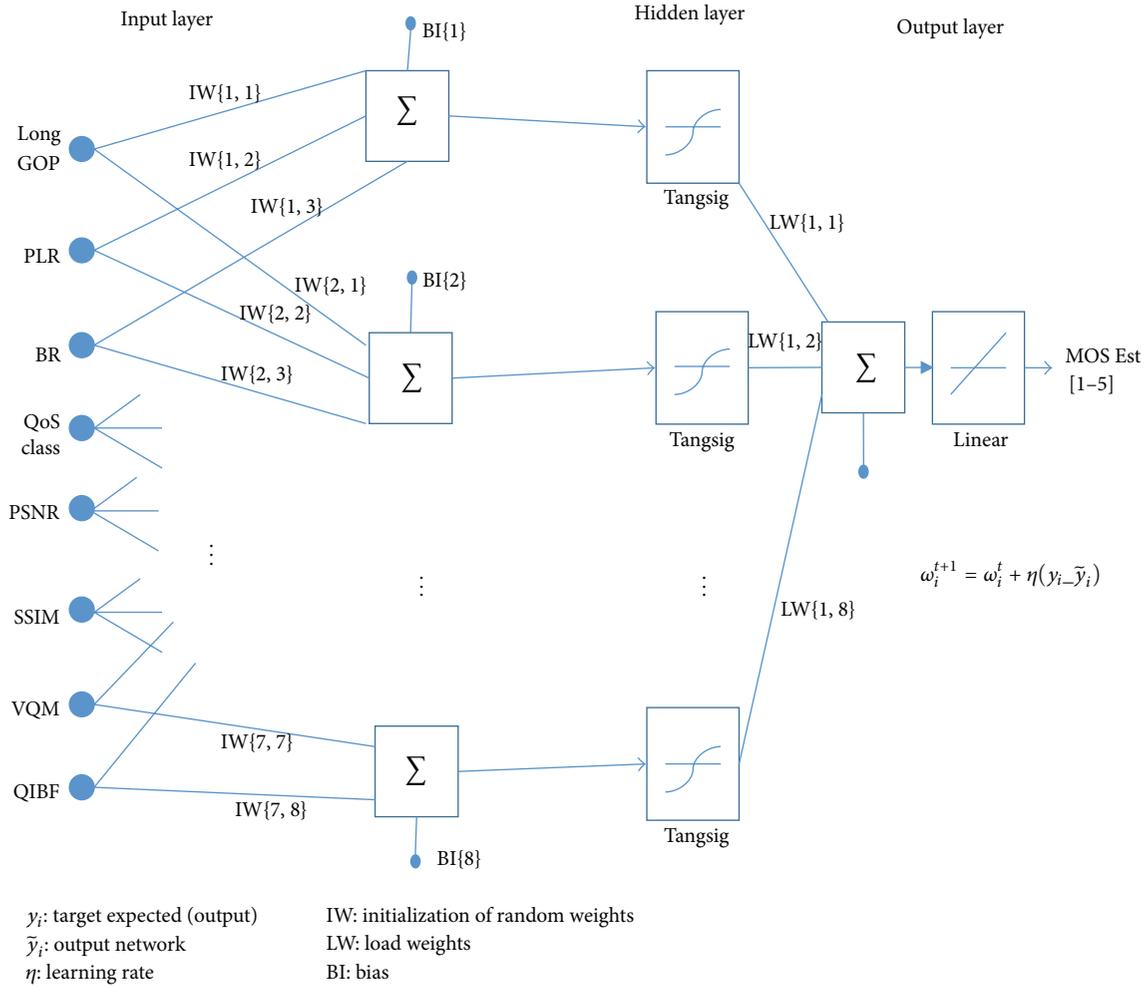


FIGURE 5: Proposed architecture for estimating MOS through BPNN feedforward.

algorithm and the lowest possible error is computed through the cost function MSE (Mean Square Error). In that case, we used an iterative gradient descent algorithm or other learning algorithms to achieve convergence to a target value (target), that is, to calculate the minimum training error of the network.

According to Ries et al. [59], in the multilayer network MLP, with a wide variety of nonlinear continuous activation functions on the hidden layers, one of these layers, which contain a largely arbitrary number of neurons, is sufficient to satisfy the property called universal approach. This allowed defining the neural network with a single hidden layer (with 20 neurons) satisfying the outputs (estimated MOS and desired MOS), calculated from the mapping of objective metrics. One of the major drawbacks is to find the right number of hidden neurons due to factors such as the quantity of input/output neurons, the number of training cases, the amount of noise in the output and input, the used architecture, the learning algorithms, and the kind of activation functions in the neurons on the hidden layer.

An empirical methodology was performed starting with 16 neurons in the hidden layer equal to twice of input neurons and a new neuron to reach the objective function with the

training algorithm Levenberg-Marquardt was added in each training and testing cycle. This is a widely used algorithm to solve the problem of least squares.

The proposed BPNN feedforward architecture is shown in Figure 6. In the input and output layer, the neurons have a linear activation function (*purelin*). In the hidden layer after various tests, the lowest training error was obtained with 20 neurons with function tangent-sigmoid activation (*tansig*).

After performing several iterations and resetting the weights in the training stage, the best performance is obtained as shown in Figure 7.

In Figure 8, the validation stage is illustrated. As shown, a good fit between the output data of the network and the desired MOS is obtained. An analysis of the linear regression between them is performed. The relationship is established between the estimated value by the BPNN (y variable) and desired data (x variable). The representation is given by the following classical linear equation:

$$y = mx + b. \quad (9)$$

According to Pearson correlation parameter between the estimated MOS by BPNN and the expected MOS, a linear fit of 96.72% and a RMSE of 0.1977 were obtained. Figure 9

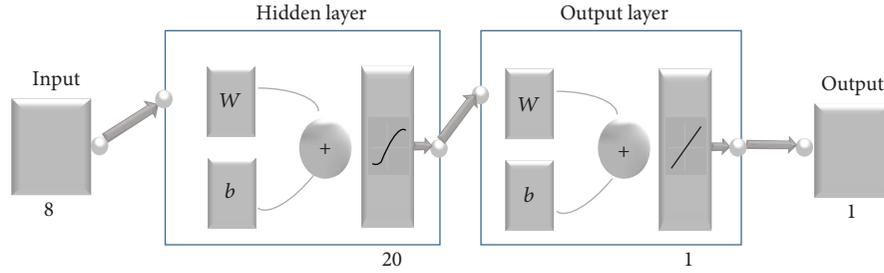


FIGURE 6: Neural Network feedforward 3-layer architecture.

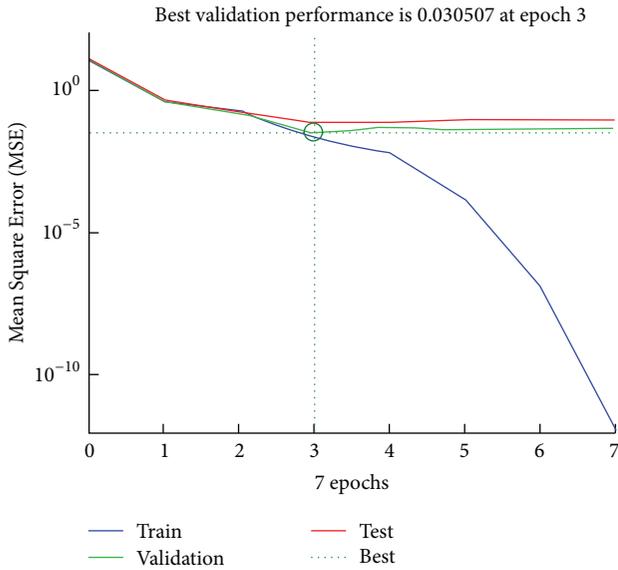


FIGURE 7: MSE obtained in the training stage.

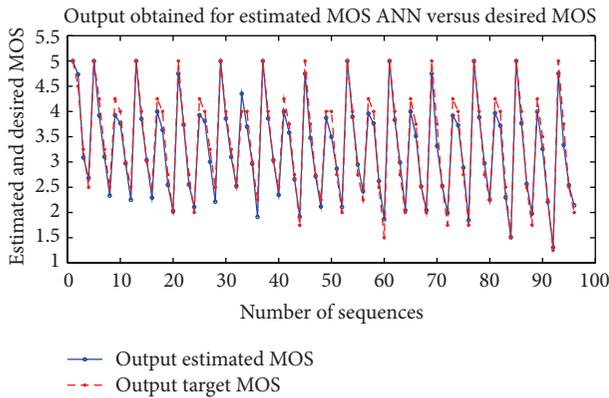


FIGURE 8: Results of the output, estimated MOS versus desired MOS.

shows the output of the correlation obtained. The results establish that the feedforward neural network allowed a good generalization with data used for validation and full linear relationship.

4.2. Implementation of a Random Neural Network (RNN) through PSQA Methodology. This kind of network captures with great accuracy and robustness mapping functions,

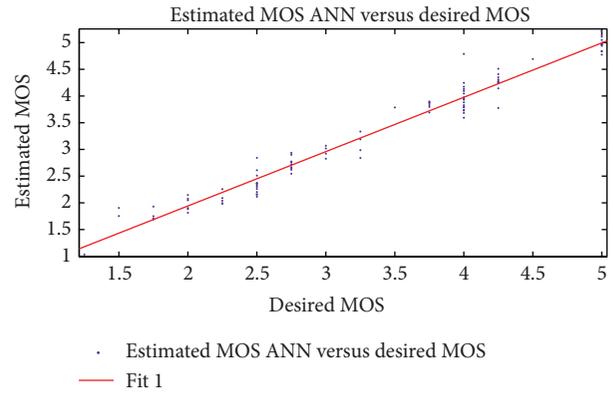


FIGURE 9: Estimated ANN MOS versus expected MOS.

where several parameters are involved. According to Casas et al. [60], such networks have been used on multiple engineering fields, highlighting and solving NP-complete optimization problems, generating textures in images problems, video and image compression algorithms, and classification problems for perceived quality of voice and video over IP, which make them ideal for its application in QoE assess. Appendix B explains in detail the RNNs and the main parameters used in these networks. The RNNs are a cross between neural networks and queuing networks. By definition, RNNs are sets of ANNs formed by a range of interconnected neurons. These neurons exchanged instantly signals which travel from one neuron to another and send signals from and to the environment. Each neuron is associated with an integer random variable associated with a potential. The potential of a neuron i at time t is defined by $q_{i(t)}$. If the potential of the neuron i is positive, the neuron is excited and randomly sends signals to other neurons or to the environment according to Poisson process of rate r_i . The signals can be positive (+) or negative (-). The RNN has a three-tier architecture. Thus, the set of neurons $R \in \{1, \dots, N\}$ is split into 3 subsets: the input neuron set, the set of hidden neurons, and the output neurons. An output $\vec{Y}^{(k)}$ is generated when the input is $\vec{X}^{(k)}$; therefore, $\vec{Y}^{(k)} = \text{MOSs}$, for $k = 1, \dots, K$, the set of weights for step k .

According to Casas et al. [60], it is necessary to apply a methodology to assess the Quality of Experience based on the use of network parameters (probability of packet loss, delay,

TABLE 6: General overview of the correlations obtained for the study cases.

Correlation	PCC	SROCC	RMSE	OR
Y-PSNR versus MOS BPNN	0.9303	0.9713	0.2882	0.5833
VQM versus MOS BPNN	0.9412	0.9707	0.2647	0.1979
SSIM versus MOS BPNN	0.9446	0.9733	0.257	0.1483
QIBF versus MOS BPNN	0.937	0.9711	0.2739	0.1562
Y-PSNR versus MOS _{psqa}	0.9698	0.9873	0.1796	0.5833
VQM versus MOS _{psqa}	0.9645	0.9900	0.1948	0.1666
SSIM versus MOS _{psqa}	0.9497	0.9889	0.2319	0.1354
QIBF versus MOS _{psqa}	0.9254	0.9853	0.2824	0.1354

Note: PCC (Pearson correlation coefficient) for prediction accuracy.
 SROCC (Spearman rank order correlation coefficient) for monotonicity.
 RMSE (Root Mean Square Error) for correlation quality assessment.
 OR (Outlier Ratio) for consistence.

jitter, etc.) and video parameters (encoding, bitrate, frame rate, GOP length, etc.), which will become input parameters. Based on these criteria, a mapping function between these parameters and subjective quality value defined by the MOS metric can be generated. To perform this task, the author proposes the methodology PSQA (Pseudo Subjective Quality Assessment), which uses the RNNs to learn the mapping between the parameters and perceived quality [51].

This methodology is characterized by its accuracy; it allows generating automatic evaluations in real time which is efficient and can be applied with many kinds of media codec and under different parameters and network conditions. Also, it can be extended for comparison with objective metrics, generating more accurate correlations [61].

The RNN is considered as a supervised learning machine, which uses multimedia and network characteristics and the expected MOS values. If in the training stage a relationship of the input parameters through the objective metrics and the expected output is found, it is possible to estimate and/or predict the subjective values with a higher level of accuracy. Due to the RNNs features, they are perfect to generate good assessments for a wide variation of all parameters that affect the quality [51]. Therefore, it is an accurate model, fast, and with low computational cost. To develop the proposed study case, RNN feedforward 3-layer architecture, proposed by Mohamed and Rubino [61], is implemented. QoE-RNN software tool was used to estimate the MOS value through the use of a RNN. This tool has LGPL license and was developed in the programming language-C [50].

The whole process is presented in Figure 10, where video sequences transmitted over the network are assessed by comparing the target average MOS. Thus, depending on the QoS strategy implemented, the MOS values associated to each objective metrics (PSNR, VQM, SSIM, and QIBF) are set and are defined by MOS_{obj} and so the average MOS (MOS_p) that is the target value of the RNN is calculated. MOS_p is expressed as shown in the following equation [26]:

$$MOS_p = \sum_{i=1}^N \frac{MOS_{obj}}{N}, \quad (10)$$

where MOS_{obj} refers to the MOS for each objective metric and N is the total of samples.

The input parameters are chosen and, with the MOS_{psqa} obtained from the network, the corresponding correlation analysis is performed.

The number of boundary iterations is 2000 and the network topology is defined by 9 input neurons, 10 neurons in the hidden layer, and one output neuron. In different conducted tests, the best fit is achieved with 10 neurons in the hidden layer.

Using MATLAB, an analysis of the linear correlation between the expected MOS and MOS_{psqa} is performed. A good linear fit is found by the Pearson coefficient R^2 at 0.9812 and RMSE at 0.1412. In Figure 11, this correlation is shown.

The general summary of all correlations obtained from each study case is presented in Table 6.

The performance of perceptual quality metrics depends on its correlation with the results of objective metrics. Thus, the accuracy in the estimation of subjective metrics with respect to issues such as prediction accuracy, monotonicity, and consistency is evaluated. This guarantees a high reliability in the subjective assessment of video quality over a range of video test sequences with different artifacts. The assessment methods are proposed by the Video Quality Experts Group (VQEG) [62]. Four statistical measurements are applied to evaluate the video quality metrics performance: Pearson correlation coefficient (PCC), Spearman's rank order correlation coefficient (SROCC), Mean Square Error (RMSE), and Outlier Ratio (OR) [63, 64].

As shown in the results of Table 6, the correlations between MOS_{psqa} metrics were certainly good with objective metrics, with high Spearman coefficient proving high linear trend in all cases. The generalization was very high and we noted the consistency, accuracy, and monotonicity results. In the simulations performed, we observed that the VQM, SSIM, and QIBF metrics are highly correlated and are close to the users' perception (established by the MOS metric).

For all cases, the correlation reached values higher than 90%, and the correlation between the objective and subjective metrics in each case presents an excellent linear behavior. Accordingly, the metrics as VQM, SSIM, and QIBF are largely related to the subjectivity established by MOS. Also

TABLE 7: Papers with machine learning methods for assessing QoE.

Paper	# video sequences	Resolution	Code c	SSIM	VQM	PSRN	MSE	BR	FR	PLR	Q value (decodificables frames rate)			Playout interruptions			Delay jitter	BW	MLBS	GOP	Machine learning	Assess MOS/DMOS?	Correlation performance			
											Q	QP	interruptions	PCC	SROCC	RMSE							OR			
He et al. [17], 2012	1	QCIF	MPEG4			X		X				X	X	X						Neural Network (PNIN)	Yes	0.9286	No	No	No	
Kipli et al. [18], 2012	N/D	No data	No data	X		X	X													BPNN	Yes	0.891	No	No	No	
Singh et al. [19], 2012	4	HD-720p	H.264					X	X											RNN	Yes	No	No	0.37	No	
Lia et al. [20], 2011	4	No data	No data																	BPNN	No	No	No	No	No	
Khan et al. [21], 2010	6	QCIF	H.264				X	X												ANFIS	Yes	0.8717	No	0.2812	No	
Du et al. [22], 2009	1	SD	No data			X			X	X		X	X	X						BPNN	Yes	No	No	No	No	
Piamrat et al. [23], 2009	1	No data	H.264			X				X										PSQA with RNN	Yes	No	No	No	No	
Khan et al. [24], 2008	3	CIF	MPEG4				X	X	X		X									ANFIS	Yes	Measure and R = 0.9229 for Q	No	RMSE = 0.1846 for MOS predicted versus MOS measure and RMSE = 0.06234 for Q predicted versus Q measure	No	No
Rubino et al. [25], 2004	Only VoIP	No data	No data							X										RNN	No	No	No	No	No	

BR: bitrate, FR: frame rate, PLR: Packet Loss Rate, QP: Quantization Parameter, BW: bandwidth, MLBS: Mean Loss Burst Size, GOP: Group of Pictures, PCC: Pearson correlation coefficient, SROCC: Spearman rank order correlation coefficient, RMSE: Root Mean Square Error, and OR: Outlier Ratio.

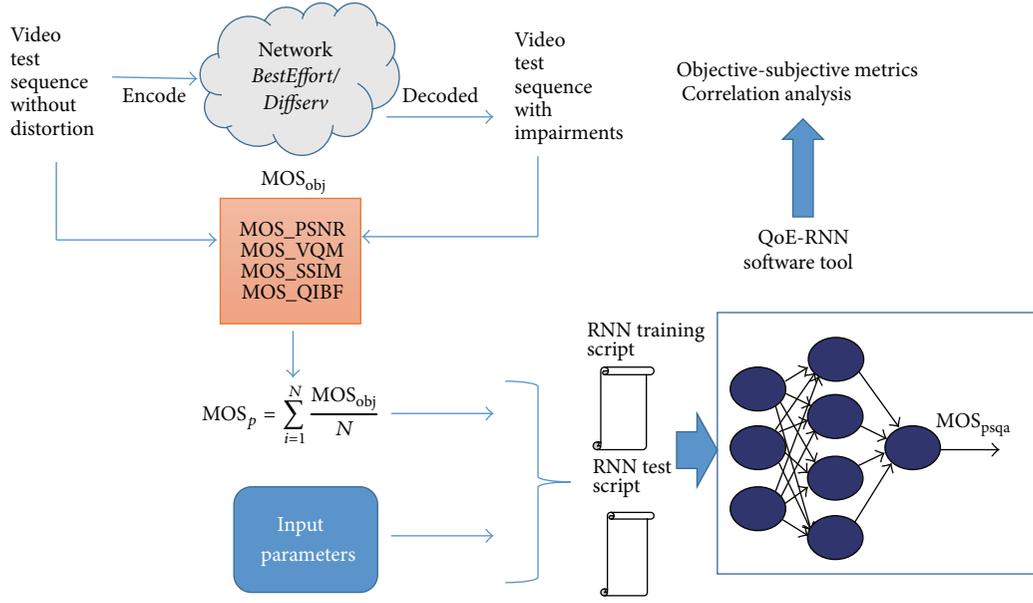


FIGURE 10: Employed methodology for the implementation of PSQA with the RNN.

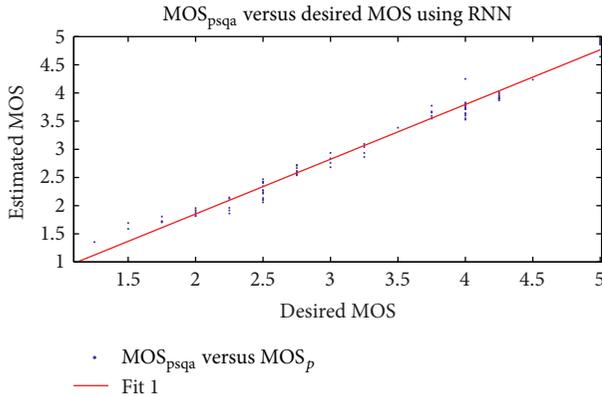


FIGURE 11: Correlation between MOS_{psqa} and MOS expected through the RNN.

it was observed that the results of the BPNN feedforward indicate a good generalization for estimated MOS against every assessed objective metrics, although it was slightly higher for applying the RNN. The RNN with PSQA provides a better correlation with the predicted values for MOS. The strategy generated by the nonintrusive model based on machine learning methods was shown to be accurate and highly flexible. It allows the estimation of subjective MOS values by relating them with FR (Full Reference) and RR (Reduced Reference) chosen for the research.

5. Conclusion

In this work, we obtained excellent correlation values between the objective and subjective QoE metrics through the use of nonintrusive methods. The accuracy, consistency,

and monotonicity were validated with the analysis of Pearson’s and Spearman’s correlations, outliers rate, and RMSE (Roots Mean Square Error). One of the main limitations of the objective and subjective methods is the lack of complete methodologies to analyze the accuracy of QoE. Therefore, this problem was addressed in order to propose a new methodology that allows finding new correlations between objective and subjective metrics. To improve the analysis, the machine learning techniques were proposed through back-propagation artificial neural networks and Random Neural Networks to enhance the approach of the estimation of human perception. In different performed simulations, we observed that VQM, SSIM, and QIBF metrics are highly correlated and are close to the user perception (determined by the MOS metric). Unlike previous works, we developed a general correlation model which uses network and coding parameters applied to several video sequences. Analyzing the results from learning machines, the BPNNs and RNNs generated high correlations with objective metrics, obtaining PCC values higher than 90% and low error rates. The consistency of correlations between the metrics through outliers was calculated with low values. We conclude that the application of nonintrusive methods allows us to generate more accurate approaches to human perception. In addition, telecommunications providers can use this methodology for estimating the QoE of users and improve their data network architectures and/or global settings on their platforms, optimize the QoS, and employ better encoding mechanisms. The development of new models for the assessment of QoE is a top research topic according to the state of the art. The topics that are being currently working include

- (i) development of new objective metrics, which are easy to apply and can be mapped accurately to the true perception of the viewer,

- (ii) the close relationship of all QoS factors which affect the QoE,
- (iii) new transmission strategies over highly congested data networks that can have a greater impact on transmission environments for streaming video over the internet which affect the user experience on the multimedia content over the next generation mobile devices,
- (iv) the application of new kinds of video codecs specifically aimed at *HD (H.265/MPEG-DASH, Dynamic Adaptive Streaming over HTTP)* [65],
- (v) the application of different methodologies based on machine learning techniques especially those related to artificial neural networks, neurofuzzy networks, support vector machines, and genetic algorithms.

Appendices

A. Proof of QIFB Metric

Let $QIBF(\text{seq}, \text{nt})$ be a QIBF metric $[0, 1]$; (7) fulfills the following axioms:

- (P.1) [Maximum] If $QIBF(\text{seq}, \text{nt}) = 1$ for all $f = \{1, 2, 3\}$, with 1 being equivalent to *I*, 2 equivalent to *P*, and 3 equivalent to *B*, the MOS index is excellent; $MOS = 5$.
- (P.2) [Minimum] If $QIBF(\text{seq}, \text{nt}) = 0$ for all $f = \{1, 2, 3\}$, with 1 being equivalent to *I*, 2 equivalent to *P*, and 3 equivalent to *B*, the MOS index is bad; $MOS = 1$.
- (P.3) [Resolution] $QIBF_1(\text{seq}, \text{nt}) < QIBF_2(\text{seq}, \text{nt})$ for all $f = \{1, 2, 3\}$, with 1 being equivalent to *I*, 2 equivalent to *P*, and 3 equivalent to *B*, if $NFL_1(1) > NFL_2(1)$; $NFL_1(2) > NFL_2(2)$; and $NFL_1(3) > NFL_2(3)$.
- (P.4) [Symmetry] $QIBF_1(\text{seq}, \text{nt}) = QIBF_2(\text{seq}, \text{nt})$ for all $f = \{1, 2, 3\}$, with 1 being equivalent to *I*, 2 equivalent to *P*, and 3 equivalent to *B*, if $NFL_1(1) = NFL_2(1) = NFL_1(2) = NFL_2(2) = NFL_1(3) = NFL_2(3)$.

Proof. (P.1) Considering $NFL(1) = NFL(2) = NFL(3) = 0$ for all $f = \{1, 2, 3\}$ and supposing that $\alpha = 0$, then $QIBF(\text{seq}, \text{nt}) = 1$. If $\alpha = 0.05$ as real adjustment, then $QIBF(\text{seq}, \text{nt}) = 0.95$ but this value can be approached at 1 in order to get a better evaluation of MOS. Therefore, as $QIBF(\text{seq}, \text{nt}) = 1$, the MOS score is around 5 as maximum value.

(P.2) If $NFL(1) = NFL(2) = NFL(3) = 1$ (maximum lost) for all $f = \{1, 2, 3\}$ and supposing that $\alpha = 0$, then $QIBF(\text{seq}, \text{nt}) = 0$. If $\alpha = 0.05$ as real adjustment, then $QIBF(\text{seq}, \text{nt}) = 0.05$ but this value can be approached at 0 in order to get a better evaluation of MOS. Therefore, as $QIBF(\text{seq}, \text{nt}) = 0$, the MOS score is around 1 as minimum value.

(P.3) Assuming $NFL_1(1) > NFL_1(2) > NFL_1(3)$ and $NFL_2(1) > NFL_2(2) > NFL_2(3)$, these relations are rewritten as

$$\begin{aligned} NFL_1(1) > NFL_2(1) > NFL_1(2) > NFL_2(2) \\ > NFL_1(3) > NFL_2(3). \end{aligned} \quad (\text{A.1})$$

Taking into account $QIBF_1(\text{seq}, \text{nt})$ and $QIBF_2(\text{seq}, \text{nt})$,

$$\begin{aligned} QIBF_1(\text{seq}, \text{nt}) &= \sum_{f=1}^F \frac{1 - NFL_1(f)}{3} - \alpha, \\ QIBF_2(\text{seq}, \text{nt}) &= \sum_{f=1}^F \frac{1 - NFL_2(f)}{3} - \alpha. \end{aligned} \quad (\text{A.2})$$

Then,

$$\begin{aligned} &QIBF_2(\text{seq}, \text{nt}) - QIBF_1(\text{seq}, \text{nt}) \\ &= \left[\sum_{f=1}^F \frac{1 - NFL_2(f)}{3} - \alpha \right] \\ &\quad - \left[\sum_{f=1}^F \frac{1 - NFL_1(f)}{3} - \alpha \right], \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} &QIBF_2(\text{seq}, \text{nt}) - QIBF_1(\text{seq}, \text{nt}) \\ &= \sum_{f=1}^F \frac{1 - NFL_2(f)}{3} - \sum_{f=1}^F \frac{1 - NFL_1(f)}{3}. \end{aligned}$$

If $NFL_1(1) > NFL_2(1) > NFL_1(2) > NFL_2(2) > NFL_1(3) > NFL_2(3)$, it is obtained that

$$\begin{aligned} &QIBF_2(\text{seq}, \text{nt}) > QIBF_1(\text{seq}, \text{nt}) \longleftrightarrow \\ &\sum_{f=1}^F \frac{1 - NFL_2(f)}{3} > \sum_{f=1}^F \frac{1 - NFL_1(f)}{3}. \end{aligned} \quad (\text{A.4})$$

Therefore, it is found that $QIBF_1(\text{seq}, \text{nt}) < QIBF_2(\text{seq}, \text{nt})$ in which if $NFL_2\{1, 2, 3\}$ is monotonically decreased, the loss is also decreased and $NFL_1\{1, 2, 3\}$ is monotonically decreased if the loss is also increased. Thus, $QIBF_1(\text{seq}, \text{nt}) < QIBF_2(\text{seq}, \text{nt})$ is a sufficient condition.

(P.4) If $NFL_1(1) = NFL_2(1) = NFL_1(2) = NFL_2(2) = NFL_1(3) = NFL_2(3)$, it is obvious that $QIBF_1(\text{seq}, \text{nt}) = QIBF_2(\text{seq}, \text{nt})$ where the quantity of loss is the same. \square

B. Random Neural Network

The RNNs are a cross between neural networks and queuing networks. By definition, RNNs are sets of ANNs comprising a series of interconnected neurons. These neurons exchange signals which instantly travel from one neuron to another and send signals from and to the environment. Each neuron is associated with an integer random variable and these are associated with a potential. The potential of a neuron i at time t is defined by $q_i(t)$. If the potential of the neuron i is positive, the neuron is excited and randomly it sends signals to other neurons or to the environment according to Poisson's process with rate r_i . The signals may be positive (+) or negative (-). Thus, the probability that the signal sent from neuron i to neuron j is positive is denoted by $P_{i,j}^+$, and the probability that the signal is negative is denoted by $P_{i,j}^-$.

The probability that the signal goes to the environment is denoted by d_i . If N is the number of neurons, for all $i = \{1, \dots, N\}$, then d_i is expressed as shown in

$$d_i + \sum_{j=1}^N (p_{i,j}^+ + p_{i,j}^-) = 1. \quad (\text{B.1})$$

Therefore, when a neuron receives a positive signal from another neuron or from the environment, its potential is increased by 1. If a negative potential signal is received, it decreases by one. When a neuron sends a positive or negative signal, its potential decreases in a unit.

The flow of positive signals arriving from the environment to the neuron i is Poisson's process with a λ_i^+ or λ_i^- rate. It is thus possible to have $\lambda_i^+ = 0$ and $\lambda_i^- = 0$ for any neuron i . To have an active network, (B.2) is needed:

$$\sum_{i=1}^N (\lambda_i^+) > 0. \quad (\text{B.2})$$

If g_i is defined as the equilibrium probability for neuron i in excitation state then, (B.3) is considered:

$$g_i = \lim_{t \rightarrow \infty} P(q_i(t) > 0). \quad (\text{B.3})$$

In (B.3), if Poisson's process for the potential of the neurons $\overrightarrow{q}(t) = \{q_1(t), \dots, q_N(t)\}$ is ergodic, the network is defined as a stable and satisfies the conditions for a nonlinear system.

In RNNs, the purpose of the learning process is to obtain the values of R_i and probabilities $P_{i,j}^+$ and $P_{i,j}^-$. The above allows obtaining the weights of the connections between the neurons i, j , as shown in

$$\begin{aligned} \omega_{i,j}^+ &= R_i P_{i,j}^+, \\ \omega_{i,j}^- &= R_i P_{i,j}^-. \end{aligned} \quad (\text{B.4})$$

From (B.4), the set of weights on the network topology is initialized with arbitrary positive values and K iterations that are performed to modify the weights. For $k = 1, \dots, K$, the set of weights for the step k is calculated from the set of weights at step $k-1$. Let $R^{(k-1)}$ be the network obtained after step $k-1$ defined by weights $\omega_{i,j}^{+(k-1)}$ and $\omega_{i,j}^{-(k-1)}$; then, the set of inputs rates (positive external signals) on $R^{(k-1)}$ for $X_i^{(k)}$ will get a network that allows generating an output $\vec{Y}^{(k)}$ when the input is $\vec{X}^{(k)}$; therefore, $\vec{Y}^{(k)} = \text{MOSs}$.

The RNN has a three-tier architecture. Thus, the set of neurons $R \in \{1, \dots, N\}$ is split into three subsets: the set of input neuron, the set of hidden neurons, and the set of output neurons. The input neurons receive positive signals from the outside. For each node i , $d_i = 0$. For the output nodes $\lambda_i^+ = 0$, $d_i > 0$. The intermediate nodes are not directly connected to the environment, for any hidden neuron i , $\lambda_i^+ = \lambda_i^- = d_i = 0$.

There are several video sequences with different parameters for the case study, where a set of training data and another set of test data are selected. In (B.5), S denotes the set of

training sequences, where each sequence is defined by α_n and S' refers to the set of validation sequences. Moreover, P is the set of parameters λ that affects each sequence, where

$$\begin{aligned} S &= \{\alpha_1, \alpha_2, \dots, \alpha_S\}, \\ S' &= \{\alpha'_1, \alpha'_2, \dots, \alpha'_S\}, \\ P &= \{\lambda_1, \lambda_2, \dots, \lambda_n\}. \end{aligned} \quad (\text{B.5})$$

The value of the parameter λ_p in the sequence α_s is defined by V_{ps} where $V = (V_{ps})$, where s is a matrix $s = 1, 2, \dots, S$. Each sequence α_s receives a score MOSs $\in [1, \dots, 5]$. Moreover, for the sequences $\alpha'_s \in S'$, a function $f(V_{1S}, V_{2S}, \dots, V_{ps}) \approx \text{MOSs}$ is obtained and the training process is completed. Otherwise, we can try with more data or change some parameters of the RNN and proceed to build a new function f .

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was developed as part of the macroproject "System of Experimental Interactive Television" for the Research and Innovation Center (Regional Alliance of Applied ICT-Artica) with code 1115-470-22055 and project no. RC584 funded by Colciencias and MinTIC.

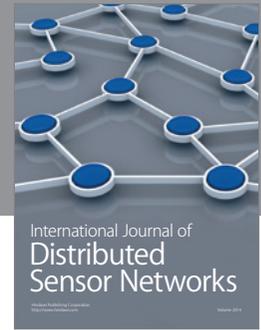
References

- [1] L.-Y. Liu, W.-A. Zhou, and J.-D. Song, "The research of quality of experience evaluation method in pervasive computing environment," in *Proceedings of the 1st International Symposium on Pervasive Computing and Applications*, pp. 178-182, IEEE, Urumqi, China, August 2006.
- [2] S. Mohseni, "Driving Quality of Experience in mobile content value chain," in *Proceedings of the 4th IEEE International Conference on Digital Ecosystems and Technologies (DEST '10)*, pp. 320-325, Dubai, United Arab Emirates, April 2010.
- [3] A. Devlic, P. Kamaraju, P. Lungaro, Z. Segall, and K. Tollmar, "Towards QoE-aware adaptive video streaming," in *Proceedings of the IEEE/ACM International Symposium on Quality of Service*, Portland, Ore, USA, February 2015, <http://people.kth.se/~devlic/publications/IWQoS15.pdf>.
- [4] S. Winkler, "Standardizing quality measurement for video services," *IEEE COMSOC MMTTC E-Letter*, vol. 4, no. 9, 2009.
- [5] S. Winkler, "Video quality measurement standards—current status and trends," in *Proceedings of the 7th International Conference on Information, Communications and Signal Processing (ICICS '09)*, Macau, China, December 2009.
- [6] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 335-350, 2010.
- [7] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430-444, 2006.

- [8] A. K. Moorthy and A. C. Bovik, "A motion compensated approach to video quality assessment," in *Proceedings of the 43rd Asilomar Conference on Signals, Systems and Computers*, pp. 872–875, Pacific Grove, Calif, USA, November 2009.
- [9] K. Radhakrishnan and L. Hadi, "A study on QoS of VoIP networks: a random neural network (RNN) approach," in *Proceedings of the Spring Simulation Multiconference (SpringSim '10)*, Society for Computer Simulation International, Orlando, Fla, USA, April 2010.
- [10] S. Winkler and P. Mohandas, "The evolution of video quality measurement: from psnr to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, 2008.
- [11] F. Boavida, E. Cerqueira, R. Choderek et al., "Benchmarking the quality of experience of video streaming and multimedia search services: the content network of excellence," in *Proceedings of the 23rd National Symposium of Telecommunications and Teleinformatics (KSTiT '08)*, Institute of Telecommunication and Electrical Technologies of University of Technology and Life Sciences, Bydgoszcz, Poland, September 2008.
- [12] P. Casas, A. Sackl, R. Schatz, L. Janowski, J. Turk, and R. Irmer, "On the quest for new KPIs in mobile networks: the impact of throughput fluctuations on QoE," in *Proceedings of the IEEE International Conference on Communication Workshop (ICCW '15)*, pp. 1705–1710, London, UK, June 2015.
- [13] M. Ries and J. R. Kubanek, "Video Quality Estimation for mobile streaming applications with neural networks," in *Proceedings of the Measurement of Speech and Audio Quality in Networks Workshop (MESAQIN '06)*, Prague, Czech Republic, June 2006.
- [14] O. Nemethova, M. Ries, E. Siffel, and M. Rupp, "Quality assessment for H.264 coded low rate and low resolution video sequences," in *Proceedings of the IASTED International Conference on Communications, Internet and Information Technology (CIIT '04)*, pp. 136–140, St. Thomas, Virgin Islands, November 2004.
- [15] D. Kuipers, R. Kooij, D. Vleeshauer, and K. Brunnstrom, "Techniques for measuring quality of experience," in *Wired/Wireless Internet Communications*, vol. 6074 of *Lecture Notes in Computer Science*, pp. 216–227, Springer, Berlin, Germany, 2010.
- [16] D. Botia, N. Gaviria, D. Jiménez, and J. M. Menéndez, "An approach to correlation of QoE metrics applied to VoD service on IPTV using a diffserv network," in *Proceedings of the 4th IEEE Latin-American Conference on Communications (IEEE LATINCOM '12)*, Cuenca, Ecuador, November 2012.
- [17] Y. He, C. Wang, H. Long, and K. Zheng, "PNN-based QoE measuring model for video applications over LTE system," in *Proceedings of the 7th International ICST Conference on Communications and Networking in China (CHINACOM '12)*, pp. 58–62, Kunming, China, August 2012.
- [18] K. Kipli, M. Muhammad, S. Masra, N. Zamhari, K. Lias, and D. Azra, "Performance of Levenberg-Marquardt backpropagation for full reference hybrid image quality metrics," in *Proceedings of International Conference of Multi-Conference of Engineers and Computer Scientists (IMECS '12)*, Hong Kong, March 2012.
- [19] K. D. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of experience estimation for adaptive Http/Tcp video streaming using H.264/AVC," in *Proceedings of the 9th Annual IEEE Consumer Communications and Networking Conf. Multimedia and Entertainment Networking and Services (CCNC '12)*, pp. 127–131, Las Vegas, Nev, USA, January 2012.
- [20] Q. Lia, J. Yangb, L. He, and S. Fan, "Reduced-reference video quality assessment based on bp neural network model for packet networks," *Energy Procedia*, vol. 13, pp. 8056–8062, 2011, Proceedings of the International Conference on Energy Systems and Electric Power (ESEP '11).
- [21] A. Khan, L. Sun, E. Ifeachor, J.-O. Fajardo, F. Liberal, and H. Koumaras, "Video quality prediction models based on video content dynamics for H.264 video over UMTS networks," *International Journal of Digital Multimedia Broadcasting*, vol. 2010, Article ID 608138, 17 pages, 2010.
- [22] H. Du, C. Guo, Y. Liu, and Y. Liu, "Research on relationship between QoE and QoS based on BP neural network," in *Proceedings of the IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC '09)*, pp. 312–315, Beijing, China, November 2009.
- [23] K. Piamrat, C. Viho, A. Ksentini, and J.-M. Bonnin, "Quality of experience measurements for video streaming over wireless networks," in *Proceedings of the 6th International Conference on Information Technology: New Generations (ITNG '09)*, pp. 1184–1189, IEEE, Las Vegas, Nev, USA, April 2009.
- [24] A. Khan, L. Sun, and E. Ifeachor, "An ANFIS-based hybrid video quality prediction model for video streaming over wireless networks," in *Proceedings of the 2nd International Conference on Next Generation Mobile Applications, Services, and Technologies (NGMAST '08)*, pp. 357–362, Cardiff, Wales, September 2008.
- [25] G. Rubino and M. Varela, "A new approach for the prediction of end-to-end performance of multimedia streams," in *Proceedings of the International Conference on Quantitative Evaluation of Systems (QUEST '04)*, pp. 110–119, IEEE CS Press, University of Twente, Enschede, The Netherlands, September 2004.
- [26] P. Goudarzi, "A no-reference low-complexity QoE measurement algorithm for H.264 video transmission systems," *Scientia Iranica Transactions: Computer Science & Engineering and Electrical Engineering*, vol. 20, no. 3, pp. 721–729, 2013.
- [27] ITU-T, "Subjective video quality assessment methods for multimedia applications," Recommendation P.910, International Telecommunication Union, Geneva, Switzerland, 1999.
- [28] ITU-T, *Recommendation, QoE Requirements in Consideration of Service Billing for IPTV Service*, UIT-T FG-IPTV, International Telecommunication Union, Geneva, Switzerland, 2006.
- [29] P. Casas, P. Belzarena, I. Irigaray, and D. Guerra, "A user perspective for end-to-end quality of service evaluation in multimedia networks," in *Proceedings of the 4th International IFIP/ACM Latin American Conference on Networking (LANC '07)*, San José, Costa Rica, 2007.
- [30] P. Casas, P. Belzarena, and S. Vatou, "End-2-end evaluation of IP multimedia services, a user perceived quality of service approach," in *Proceedings of the 18th ITC Specialist Seminar on Quality of Experience (ITEC '08)*, Karlskrona, Sweden, May 2008.
- [31] R. Immich, P. Borges, E. Cerqueira, and M. Curado, "QoE-driven video delivery improvement using packet loss prediction," *International Journal of Parallel, Emergent and Distributed Systems—Smart Communications in Network Technologies*, vol. 30, no. 6, pp. 478–493, 2015.
- [32] M. Kailash and D. Padma, "An overview of quality of service measurement and optimization for voice over internet protocol implementation," *International Journal of Advances in Engineering & Technology*, vol. 8, no. 1, pp. 2053–2063, 2015.
- [33] S. Timotheou, "The random neural network: a survey," *The Computer Journal*, vol. 53, no. 3, pp. 251–267, 2010.

- [34] K. Kilkki, "Next generation internet and QoE," in *Presentation at EuroFGI IA.7.6 Workshop on Socio-Economic Issues of NGL*, Santander, Spain, June 2007.
- [35] ITU. FG-IPTV.DOC-0814, Quality of Experience Requirements for IPTV Services, 2007.
- [36] E. Cerqueira, L. Veloso, M. Curado, and E. Monteiro, *Quality Level Control for Multi-User Sessions in Future Generation Networks*, University of Coimbra INESC Porto, Coimbra, Portugal, 2008.
- [37] 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.
- [38] D. Wang, W. Ding, Y. Man, and L. Cui, "A joint image quality assessment method based on global phase coherence and structural similarity," in *Proceedings of the 3rd International Congress on Image and Signal Processing (CISP '10)*, vol. 5, pp. 2307–2311, IEEE, Yantai, China, October 2010.
- [39] M. Vranješ, S. Rimac-Drlje, and D. Žagar, "Objective video quality metrics," in *Proceedings of the 49th International Symposium ELMAR*, Faculty of Electrical Engineering, University of Osijek, Zadar, Croatia, September 2007.
- [40] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [41] Y. Wang, *Survey of Objective Video Quality Measurements*, EMC Corporation, Hopkinton, Mass, USA, 2004, <ftp://ftp.cs.wpi.edu/pub/techreports/pdf/06-02.pdf>.
- [42] T. Zinner, O. Abboud, O. Hohlfeld, T. Hossfeld, and P. Train-Gia, "Towards QoE management for scalable video streaming," in *Proceedings of the 21st ITC Specialist Seminar on Multimedia Applications Traffic, Performance and QoE*, Miyazaki, Japan, March 2010.
- [43] R. Serral-Garcia, Y. Lu, M. Yannuzzi, X. Masip-Bruin, and F. Kuipers, *Packet Loss Based Quality of Experience of Multimedia Video Flows, Crente de Recerca d'Arquitectures Avancades de Xarxes (Politencic University of Catalunya)*, Spain, Delft University of Technology, Delft, The Netherlands, 2009, http://personals.ac.upc.edu/rserral/research/techreports/psnr_mos.pdf.
- [44] D. Botia, N. Gaviria, J. Botia, D. Jiménez, and J. M. Menéndez, "Improved the quality of experience assessment with quality index based frames over IPTV network," in *Proceedings of the 2nd International Conference on Information and Communication Technologies and Applications (ICTA '12)*, Orlando, Fla, USA, 2012.
- [45] DSL Forum, "Triple play services quality of experience requirements and mechanism," Working Text WT-126 version 0.5, 2006.
- [46] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid—a framework for video transmission and quality evaluation," in *Proceedings of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pp. 255–272, Urbana, Ill, USA, September 2003.
- [47] D. Vatolin, "MSU Video Metric Quality Tool," MSU Graphics and media Lab, 2012, http://compression.ru/video/quality_measure/video_measurement_tool_en.html, Rusia.
- [48] P. Subramanya, K. Vinayaka, H. Gururaj, and B. Ramesh, "Performance evaluation of high speed TCP variants in dumbbell network," *IOSR Journal of Computer Engineering*, vol. 16, no. 2, pp. 49–53, 2014.
- [49] D. Botia, N. Gaviria, D. Jiménez, and J. M. Menéndez, "Strategies for improving the QoE assessment over iTV platforms based on QoS metrics," in *Proceedings of the IEEE International Symposium on Multimedia (ISM '12)*, pp. 483–484, Irvine, Calif, USA, December 2012.
- [50] QoE-RNN Tool, 2011, <http://code.google.com/p/qoe-rnn/>.
- [51] G. Rubino, M. Varela, and J.-M. Bonnin, "Controlling multimedia QoS in the future home network using the PSQA metric," *The Computer Journal*, vol. 49, no. 2, pp. 137–155, 2006.
- [52] W. Ding, Y. Tong, Q. Zhang, and D. Yang, "Image and video quality assessment using neural network and SVM," *Tsinghua Science and Technology*, vol. 13, no. 1, pp. 112–116, 2008.
- [53] A. Bouzerdoum, A. Havstad, and A. Beghdadi, "Image quality assessment using a neural network approach," in *Proceedings of the 4th IEEE International Symposium on Signal Processing and Information Technology*, pp. 330–333, Rome, Italy, December 2004.
- [54] J. Choe, K. Lee, and C. Lee, "No-reference video quality measurement using neural networks," in *Proceedings of the 16th International Conference on Digital Signal Processing (DSP '09)*, pp. 1–4, IEEE, Santorini-Hellas, Greece, July 2009.
- [55] 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.
- [56] C.-H. Kung, W.-S. Yang, C.-Y. Huan, and C.-M. Kung, "Investigation of the image quality assessment using neural networks and structure similarity," in *Proceedings of the International Symposium on Computer Science*, vol. 2, no. 1, p. 219, August 2010.
- [57] C. Wang, X. Jiang, F. Meng, and Y. Wang, "Quality assessment for MPEG-2 video streams using a neural network model," in *Proceedings of the IEEE 13th International Conference on Communication Technology (ICCT '11)*, pp. 868–872, IEEE, Jinan, China, September 2011.
- [58] 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.
- [59] M. Ries, J. Kubanek, and M. Rupp, "Video quality estimation for mobile streaming applications with neural networks," in *Proceedings of the Measurement of Speech and Audio Quality in Networks Workshop (MESAQIN '06)*, Prague, Czech Republic, June 2006.
- [60] P. Casas, D. Guerra, and I. Bayarres, "Perceived Quality of Service in Voice and Video Services over IP," School of Engineering, Universidad de la República, End of career project, Electrical Engineering—Plan 97, Telecommunications, Uruguay, 2005.
- [61] S. Mohamed and G. Rubino, "A study of real-time packet video quality using random neural networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 12, pp. 1071–1083, 2002.
- [62] VQEG, "Video Quality Experts Group, Final Report from the VQEG on the validation of Objective Models of Video Quality Assessment, Phase II," 2003, <http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx>.
- [63] S. Winkler, *Digital Video Quality—Vision, Models and Metrics*, John Wiley & Sons, 2005.
- [64] A. Bath, I. Richardson, and S. Kannangara, *A New Perceptual Quality Metric for Compressed Video*, The Robert Gordon University, Aberdeen, Scotland, 2007.

- [65] M. Alreshoodi, J. Woods, and I. K. Musa, "Optimising the delivery of Scalable H.264 Video stream by QoS/QoE correlation," in *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE '15)*, pp. 253–254, IEEE, Las Vegas, Nev, USA, January 2015.
- [66] A. Kuwadekar and K. Al-Begain, "User centric quality of experience testing for video on demand over IMS," in *Proceedings of the 1st International Conference on Computational Intelligence, Communication Systems and Networks (CICSYN '09)*, pp. 497–504, Indore, India, July 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

