A Proposed Measurement for Video Quality of Experience

Rana Fareed Ghani [1] and Amal Sufiuh Ajrash [2*]
[1] Department of Computer Science, University of Technology, Baghdad-Iraq.
[2] Department of Computer Science, Collage of Science for Women, University of Baghdad, Baghdad-Iraq.
[*] Corresponding author: amalsa_comp@csw.uobaghdad.edu.iq

Abstract

Technological development in the last years leads to increase the access speed in the networks that allow a huge number of users watching videos online. The Quality of Experience (QoE) Knowledge of services that provide from the network is a very critical matter to have a strong design of multimedia streaming networks. This paper provides a video streaming QoE prediction metric that does not require any information on the reference video. The proposed system extract numbers of features from videos that used to train the neural network and finally prediction the QoE value. Verify models prediction using 10-fold cross-validation that in a regular way split dataset (training set and test set) with multiple percentages. The proposed system verifies the best result.
[DOI: 10.22401/ANJS.22.3.10]

Keywords: Cross Validation, Feature Extraction, Mean Opinion Score (MOS), Neural Network, Quality of Experience (QoE).

1. Introduction

In general, there isn't a common definition for the term Quality of Experience (QoE). The QoE measures the performance as the user can understand subjectively [1], so that the QoE is an expansion to the Quality of Service (QoS) [2]. There are many applications and services such as Internet video and mobile broadcasting interested in video Quality of Experience assessment. The methods used to measure video quality can be classified into two groups: Subjective and Objective metrics [2,3].

The subjective metric is used to check the accuracy of objective scores so can be used in video QoE prediction, where the more familiar subjective measurement is the Mean Opinion Score (MOS) where viewers look videos in real-time and classify such as following: 1-worst, 2-poor, 3-fair, 4-good, and 5-excellent. The objective metrics depend on a practical mathematical approach to apply without needed of viewers [4].

The QoE measurement has great importance in video processing applications [1], in case of measure video QoE must follow one of the three methodologies: Full Reference (FR), Reduce Reference (RR) and No Reference (NR). In FR type, the original video must be available as reference and used it in the assessment process, RR and NR are same because they have no original video, but the RR used information about the original video during assessment progress. The NR has no information about the original video, it will calculate the QoE in real-time by take the QoS features or pixel-based features or both them. For that reason, it one of the practical QoE measurement is adopted in real-time video stream [1,5].

One of the trends of multimedia applications, like video streaming and audio conferencing, is the QoE direction, which is the goal of many studies due to it gives the ability to determine factors that affect the QoE and control them. As a result, QoE can react to variations in network performance [2].

2. Related work

The increased transmission of streaming video over the Internet leads to the need for quantitative assessment for video or audio quality transmission. One of the most adopted methods is to use a group of experienced persons to perform the assessment on actual test sequences representative of the conditions studied. Using such type of subjective tests are costly and not highly accurate. Instead, there are much research on using automatic measurement assessment [4].

Rubino, G., et al. in 2006 [6] developed new technology uses the Pseudo-Subjective Quality Assessment (PSQA) that have the ability to measure the quality of a video or audio communication over a packet network, as perceived by the user. The PSQA technique results are accurate, where it correlates with the experienced opinion with high efficiency because it uses video in real time. This system used Random Neural Network (RNN) for training to learn the system how observers quantify the quality.

Zheng, K., et al. in 2015 [7] proposed a new quality assessment method to video streaming based on neural networks (NN), the system used a number of objective factors to design and train the NN. This matric has the ability to estimate the Mean Opinion Score (MOS) value that depends on a Long Term Evolution (LTE) quality metrics values like (delay, delay-jitter, and packet loss rate). A comparison is done with the proposed system that used the freezing features that happened due to packet loss or late arrival and this feature has a major role in calculating best QoE value. So, after training the system on a large number of MOS scores in the future, the operators will need no human-based training.

Calyam, P., et al. in 2012 [8] proposed Multilayer Artificial Neural Network (ANN) that use the QoS parameters as input to the input layer, while the MOS, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) and Video Quality Metric (VQM) as an output resulted from output layer, and finally, it gets on QoE value. The authors depend at research on three features (jitter, loss frame, bitrate) with different video resolution to predict the QoE, the proposed system used pixel-based features and QoS features to get a more accurate result.

Mocanu, D. et al. in 2015 [9] displayed a new metric that measures the user dissatisfaction, which not always refers to averaged scores. This is done by using deep learning framework / deep belief networks and two modelled the average scores and user dissatisfaction levels.

In most of QoE research, the researchers depended in works on QoS features that related to network and their effects on video streaming quality. The proposed system found the QoS features are not enough to measure the QoE value, as a result, has been used the pixel-based features to get more powerful value from the subjective QoE.

3. The Proposed System

The main objective of the proposed system is designing an efficient objective video QoE system that can predict the MOS of a video stream.

The proposed No Reference QoE (NRQoE) system is a multi-level system. At first, is extracting different features from a video sequence. Then, the features extracted are using to train the neural network system and predict QoE values.

The main steps of the proposed NRQoE system shown in Fig.(1). The system consists of four steps. These steps are: data set selection process, the second step is feature extraction process, predicting phase, and testing system accuracy.

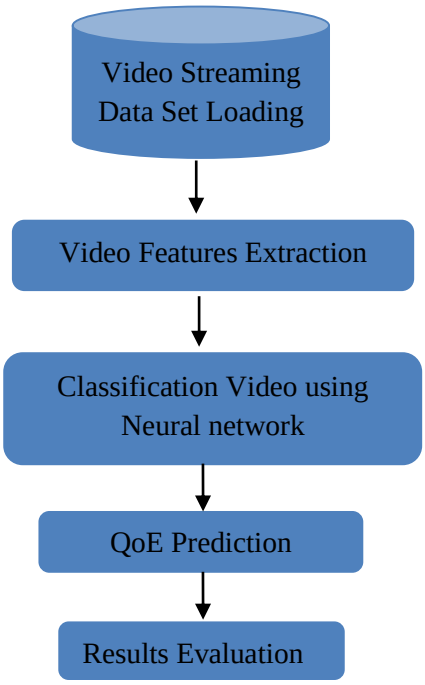

Fig.(1): Proposed system diagram.

3.1 Video Streaming Dataset

In this work, we used the Waterloo Streaming QoE Database III of video sequences with a good distribution of spatial and temporal properties to get the best result of the proposed system. The database contains 20 sources of videos and 450 simulated streaming videos in an average duration of 13 seconds at different contents. The streaming sessions

generated by 6 adaptive streaming algorithms under 13 wide-ranging bandwidth conditions are recorded and evaluated by 34 subjects. Subjects score the quality of each video sequence according to the 0-100 numerical quality. Fig.(2) shows the frames of the original video of the used dataset [10,11].

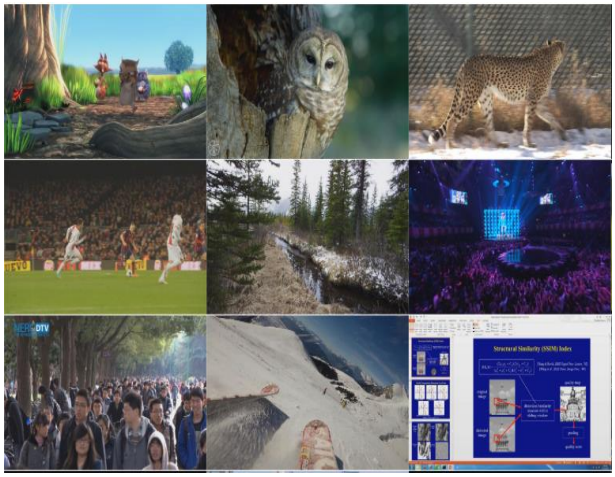

Fig.(2): Sample images of the source videos contents from Waterloo Video Quality dataset [11].

3.2 Feature Extractions

Feature extraction phase is responsible for extracting the information from huge data, and use this information in the prediction process.

In the feature extraction phase, the dimensionality reduction of the data is necessary to decrease the memory and time lose [3,4]. The proposed system adopted the features extraction process on video frames. It extracted 10 features, which are:

• **Freezing feature** is a major type of packet loss. It happend when the frames dropped and that frame was still displayed until the following right frame is received. Therefor, this phenomenon made freeze in a video sequence [12], Which affected on the video quality. The frozen frames are discovered by computing the difference between sequence frames; if the difference is zero then it means that there is a frozen frame.

• **Blockiness feature** analyz the frame components to vertical and horizontal edge enhancement filtering and considered the high gradient information in eight directions by using Kirsch Masks and used this information for the distortion measurement[13]. Fig.(3) show the kirsch masks for detected in eight directional edges information (N, NE, E, SE, S, SW, W, NW). The edge direction is defined by the mask that produces the maximum magnitude.

$$G_1=\begin{bmatrix}-3 & -3 & 5\\ -3 & 0 & 5\\ -3 & -3 & 5\end{bmatrix}\ G_2=\begin{bmatrix}-3 & 5 & 5\\ -3 & 0 & 5\\ -3 & -3 & -3\end{bmatrix}\ G_3=\begin{bmatrix}5 & 5 & 5\\ -3 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}\ G_4=\begin{bmatrix}5 & 5 & -3\\ 5 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}$$

$$G_5=\begin{bmatrix}5 & -3 & -3\\ 5 & 0 & -3\\ 5 & -3 & -3\end{bmatrix}\ G_6=\begin{bmatrix}-3 & -3 & -3\\ 5 & 0 & -3\\ 5 & 5 & -3\end{bmatrix}\ G_7=\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & -3\\ 5 & 5 & 5\end{bmatrix}\ G_8=\begin{bmatrix}-3 & -3 & 5\\ -3 & 0 & 5\\ -3 & 5 & 5\end{bmatrix}$$

Fig.(3): Eight directional Kirsch Masks.

• **Blurring feature** is one of the most important NR features, where the blur effect on frames which leads to loss of the necessary details required for the scene interpretation [13]. In this system, blurring feature is extracted by applying Laplacian operator. The Laplacian highlights regions of an image containing rapid intensity changes. The Laplacian is often used for edge detection. It convolves the frame with the 3x3 Laplacian operator shown in Fig.(4) for each frame, then it found the blur average for video frames and returns the variance.

0	1	0
1	-4	1
0	1	0

Fig.(4): Laplacian Kernel.

• **Natural Sences Statistics (NSS) feature** has a high correlation to the Human Visual System (HVS) and consequently impact the end user of video stream service. The NSS extracted features are:- (N_Shape, N_Variance, N_H Shape, N_H Variance, N_V Shape, N_V Variance), by using the Blind/ Referenceless Image Spatial QUality Evaluator (BRISQUE) which is considered as an NSS model framework of locally normalized luminance coefficients and quantifies naturalness using the parameters of the model [14]. The motivation to use this feature is each distortion afflicts natural frames in a different way. Therefore, a different set of features are important for every distortion. We use a generic feature set to represent the

'naturalness' of images. These features still capture the characteristics of diverse distortions making it possible to train a classifier which can map the model parameters to the type of distortion the frame is afflicted with.

Eventually, all these features leads to improve the accuracy performance of the prediction model as shown in Pseudocode (1). Samples of the extracted features from samples video listed in Table (1).

Pseudocode 1: Feature Extraction Model.

Input: Set of NR video streaming;
Output: list of extracted features;
1 Detect video stream;
2 For each vide ∈ to data input stream buffer Do.
3 Extract video frame per second;
4 For each frame from *N* sequence Do
5 Create 4 threads to analysis pixel per
6 frame using multithreading:
7 Thread one calculate blurring;
8 Thread two calculated the blockiness;
9 Thread three freezing;
10 Thread four calculate naturalness;
11 Execute the multi-threads and the average for each process extracted and save to array;
12 Feature extracted value saved to file;
13 File sent to the supervised trained model;
14 Capture N sequence, go to step 2.

Table (1)
Samples of extracted features.

Video name	No. frame	Freeze	Blur.	Bloc.	NSS					
					N_ Shape	N_ Var.	N_H Shape	N_H Var.	N_ V Shape	N_V Var.
V1	30	28.5	16	162809	1.4387	0.0809	0.5224	0.0102	0.5198	0.0105
V2	30	4.7	24	179468	1.4809	0.099**5**	0.5413	0.0149	0.5377	0.0155
V3	30	6.2	46	192568	1.5010	0.1229	0.5438	0.0252	0.5517	0.0236
V4	30	2.3	36	187993	1.5023	0.11**40**	0.5458	0.0202	0.5494	0.0202
V5	30	1.6	37	188854	1.5139	0.116**5**	0.5506	0.0210	0.5544	0.0209
V6	30	1.3	43	192521	1.4900	0.1180	0.5428	0.0231	0.5487	0.0219
V7	30	1.6	49	195197	1.5012	0.125**9**	0.5450	0.0261	0.5534	0.0246
V8	30	1.3	62	195670	1.4869	0.130**6**	0.5400	0.0288	0.5482	0.0273
V9	30	10.4	24	183855	1.47**60**	0.0967	0.5408	0.0141	0.5349	0.0147
V10	30	10.4	30	192114	1.5039	0.1079	0.5514	0.0173	0.5489	0.0717
V11	30	43.2	58	202899	1.5319	0.1300	0.5599	0.0259	0.5619	0.0258
V12	30	10.1	35	198568	1.5245	0.1143	0.5625	0.0189	0.5573	0.0196
V13	30	15.2	38	201045	1.5112	0.1098	0.5629	0.0174	0.5484	0.0186
V14	30	17.8	39	201280	1.5276	0.1125	0.5699	0.0179	0.5551	0.0191

3.3 Classification

Classification is the most critical phase in the proposed system. The main goal of the classification process is to convert the quantitative input data to qualitative output information. The classification phase used a vector of extracted features from the input data that need to be classified, and then assign it's using Prediction QoE into a class that is the most appropriate one using Artificial Neural Network (ANN) algorithm.

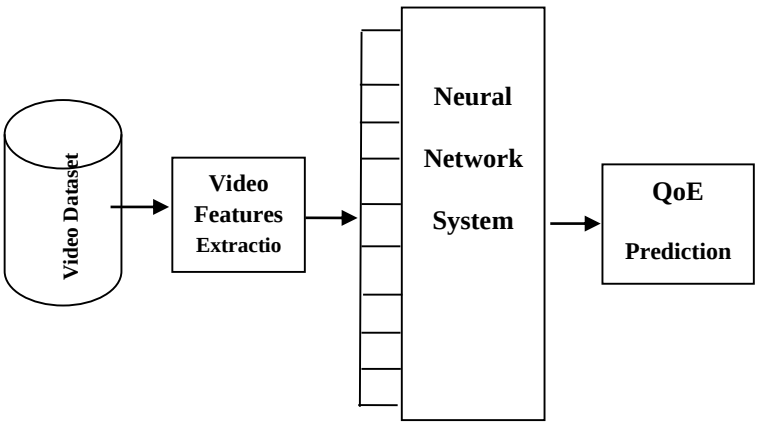

Fig.(5):Proposed system of the learning model.

The system is trained with a Multi-Layer Perceptron network (MLP), its process dealing with samples one by one, comparing the prediction with actual known data in learning phase to each recorded sample. MLP constructed from a neuron, which formed from multiple input values, combined according to its weight. In addition, it is a feed-forward used standard backpropagation algorithm for training. It considered a supervised algorithm that required the desired solution.

The MLP algorithm work as classifier used to build the proposed model to predicate real-time automated end-user QoE which represent the MOS values (Bad, Poor, Fair, Good, Excellent). The MLP network constructed from 10 units as input each one represent an extracted feature, 10 neurons in the hidden layer and one neuron as output that represents the predicates MOS, Fig.(6) shows the structure of the used MLP neural network.

The training algorithm used here is the Levenberg-Marquardt that applied with mean squared error as loss function and number of iteration range is (0-1000). The local gradient in each output neuron is calculate and update the weights between the neurons of the hidden layer and input neurons, and then the local gradients of the neurons in the hidden layer are calculated. The same process used to update the weights between the output neurons and hidden neurons as used to update the weights between the hidden neurons and input neurons.

During training, the output unit compares the computed activation sigmoid by activation algorithm with its target value to determine the associated error of that pattern with that unit. During the phase of learning, signals are sent in the reverse direction.

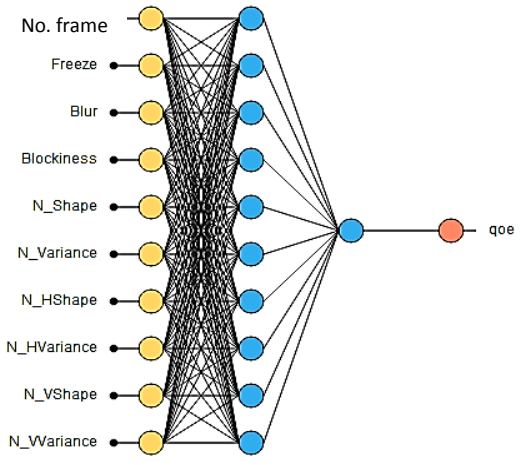

Fig.(6): Artificial neural network architecture .

Table (2) shows a sample result of QoE from the considered dataset and from the proposed system, QoE value between (0-100) that classified into subranges, (1-20) is worst, (21-40) is poor, (41-60) is fair, (61-80) is good, and (81-100) is excellent. The comparison indicates an improvement in the QoE results. The chart in Fig.(7) shows the changing between the reference data and the proposed system output.

Table (2)
QoE result before and after training.

Video name	QoE from Dataset	QoE from system
V1	31.377	45.373
V2	51.006	51.732
V3	48.948	55.894
V4	58.343	55.383
V5	62.817	63.067
V6	43.938	53.349
V7	64.339	64.121
V8	62.891	63.614
V9	45.61	46.929
V10	52.109	53.095
V11	46.137	45.373
V12	56.506	55.273
V13	62.346	50.766
V14	43.174	52.476
V15	52.513	50.079
V16	13.732	19.373
V17	49.42	48.746
V18	62.738	63.582
V19	67.517	64.929
V20	64.93	66.924

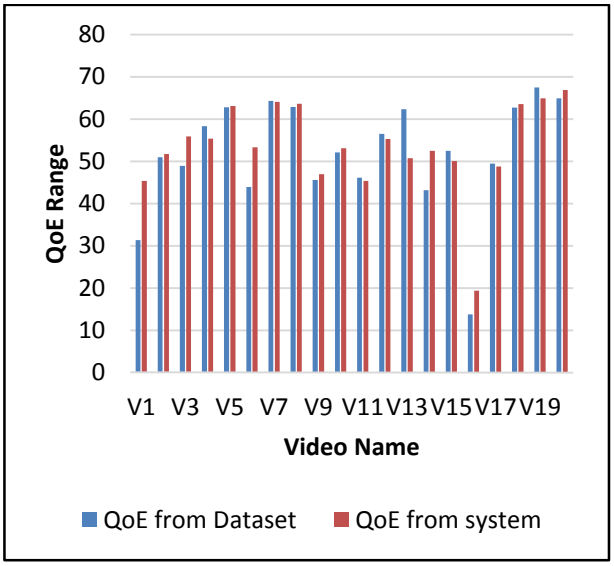

Fig.(7): Data change of the training and proposed system output.

3.4. Evaluation

The K-fold Cross Validation (CV) is used as an approach to investigate the prediction accuracy. Its randomly split the dataset into K samples; one sample for the test and K-1 samples for the training. This process repeated K times with changing K value as 80 % as a training set and 20% as a testing set. In the proposed system we used 10-fold CV as shown in Fig.(8).

```
Fold = 0: wrong = 10 correct = 35 error = 0.2222
Fold = 1: wrong = 0 correct = 45 error = 0.0000
Fold = 2: wrong = 0 correct = 45 error = 0.0000
Fold = 3: wrong = 2 correct = 43 error = 0.0444
Fold = 4: wrong = 2 correct = 43 error = 0.0444
Fold = 5: wrong = 2 correct = 43 error = 0.0444
Fold = 6: wrong = 0 correct = 45 error = 0.0000
Fold = 7: wrong = 3 correct = 42 error = 0.0667
Fold = 8: wrong = 1 correct = 44 error = 0.0222
Fold = 9: wrong = 2 correct = 43 error = 0.0444
Total Accumulated: wrong = 22 correct = 428 error = 0.0489
```

Fig.(8) The results of 10-Fold.

4. Conclusions

In this work, the QoE of streaming video is predicted throw the MLP neural network model. The prediction of this model depends on different features extracted from videos content and other features from the network. All these features give an acceptable predictive of QoE value. The proposed system has shown that the system results are compatible with the experimental results.

Reference

[1] Mok, R. K.; Chan, E. W.; and Chang, R. K.; "Measuring the Quality of Experience of HTTP Video Streaming"; 12th IFIP/IEEE International Conference on Integrated Network Management; Dublin, Ireland; 2011.

[2] Mendi, E.; Milanova, M.; Zhou, Y.; and Talburt, J.; "Objective Video Quality Assessment for Tracking Moving Objects from Video Sequences"; International Conference on Signal Processing Robotics and Automation (ISPRA'10); Cambridge, UK; pp. 121-126; February 20-22, 2010.

[3] Zhengfang, D.; Kai, Z.; Kede, M.; Abdul Rehman; and Zhou, W.; "A Quality-of-Experience Index for Streaming Video"; IEEE Journal of Selected Topics in Signal Processing; 11(1); 2017.

[4] Alreshoodi, M.; "Prediction of Quality of Experience for Video Streaming Using Raw QoS Parameters"; Ph.D. thesis, Department of Computer Science and Electronic Engineering - University of Essex; 2016.

[5] Bao, Y.; Lei, W.; Zhang, W.; and Zhan, Y.; "QoE collaborative evaluation method based on fuzzy clustering heuristic algorithm"; Springer Plus; 5(1); 2016.

[6] Rubino, G.; Tirilly, P.; and Varela, M.; "Evaluating user's satisfaction in packet networks using random neural networks. Artificial Neural Networks"; Lecture notes in Computer science (ICANN 2006); PP.303–312; 2006.

[7] Zheng, K.; Zhang, X.; Zheng, Q.; Xiang, W.; and Hanzo, L.; "Quality-of-Experience Assessment and its Application to Video Services in LTE Networks"; IEEE Wireless Communications; 22(1); PP.70–78; 2015.

[8] Calyam, P.; Chandrasekaran, P.; Trueb, G.; Howes, N.; Ramnath, R.; Yu, D.; Ying, L.; Xiong, L.; and Yang, D.; "Multi-Resolution Multimedia QoE Models for IPTV Applications", International Journal of Digital Multimedia Broadcasting; 2012.

[9] Mocanu,D.; Pokhrelb, C.; Garellac, J. J.; Seppanen, J.; Liotou, E.; and Narwaria, M.; "No-reference video quality measurement: The added value of Machine learning"; Journal of Electronic Image; 24(6); 2015.

[10] Zhengfang, D.; Abdul Rehman; and Zhou, W.; "A Quality-of-Experience Database for Adaptive Video Streaming", IEEE Transactions on Broadcasting; 64(2); 2018.

[11] Bampis, C. G.; Li, Z.; Moorthy, A. K.; Katsavounidis, I.; Aaron, A.; and Bovik, A. C.; "Study of Temporal Effects on Subjective Video Quality of Experience", IEEE Trans. Image Process.; 26(11); pp. 5217–5231; 2017.

[12] Usman, M, A..; Shin, S., Y.; Shahid, M.; and Lövström, B.; "A No-Reference Video Quality Metric Based on Jerkiness Estimation Focusing on Multiple Frame Freezing in Video Streaming", IETE Technical Review;34(3); pp.309-320; 2017.

[13] Soltanian, N.; Karimi, N.; Karim, M.; and Samavi,, SH.; "Blind Image Quality Assessment Based on Natural Scene Statistics," IEEE 22nd Iranian Conference on Electrical Engineering (ICEE); Iran; 2014.

[14] Fieno, A.; "No-reference Video Quality Assessment Model Based on Artifact Metrics for Digital Transmission Applications"; PhD. Thesis; Department of Computer Science - University of Brasilia; Brasilia; 2010.