



ISSN: 0067-2904

Quality of Experience Measurement for Video Streaming Based On Adaptive Neural Fuzzy Inference System

Rana Fareed Ghani¹, Amal Sufiuh Ajrash^{*2}

¹Department of Computer Science, University of Technology, Baghdad, Iraq

²Department of Computer Science, Collage of Science for Women, University of Baghdad, Baghdad, Iraq

Abstract

Technological development in recent years leads to increase the access speed in the networks that allow a huge number of users watching videos online. Video streaming is one of the most popular applications in networking systems. Quality of Experience (QoE) measurement for transmitted video streaming may deal with data transmission problems such as packet loss and delay. This may affect video quality and leads to time consuming. We have developed an objective video quality measurement algorithm that uses different features, which affect video quality. The proposed algorithm has been estimated the subjective video quality with suitable accuracy. In this work, a video QoE estimation metric for video streaming services is presented where the proposed metric does not require information on the original video. This work predicts QoE of videos by extracting features. Two types of features have been used, pixel-based features and network-based features. These features have been used to train an Adaptive Neural Fuzzy Inference System (ANFIS) to estimate the video QoE.

Keywords: Quality of Experience (QoE), Main Opinion Score (MOS), Back Propagation of neural network.

مقياس جودة التجربة للفيديو المتسلسل على أساس نظام الاستدلال العصبي المضرب

رنا فريد غني¹، امال سفيح عجرش^{*2}

¹قسم علوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق

²قسم علوم الحاسوب، كلية العلوم للبنات، جامعة بغداد، بغداد، العراق

الخلاصة

التطور التكنولوجي في السنوات الأخيرة يؤدي إلى زيادة سرعة الوصول في الشبكات التي تسمح لعدد كبير من المستخدمين بمشاهدة مقاطع الفيديو عبر الإنترنت. يعد بث الفيديو أحد التطبيقات الأكثر شيوعاً في أنظمة الشبكات. قد يتعامل مقياس جودة التجربة (QoE) لبث الفيديو المرسل مع مشكلة نقل البيانات مثل فقدان الرزم والتأخير. قد يؤثر ذلك على جودة الفيديو ويؤدي إلى استهلاك الوقت. لقد طورنا خوارزمية موضوعية لقياس جودة الفيديو تستخدم ميزات مختلفة تؤثر على جودة الفيديو. استطاعت الخوارزمية المقترحة تقدير جودة الفيديو بدقة مناسبة تقابل تقدير الخبير. في هذا العمل، يتم تقديم مقياس تقدير جودة الفيديو QoE لخدمات بث الفيديو حيث لا يتطلب المقياس المقترح معلومات حول الفيديو الأصلي. هذا العمل يتنبأ بالجودة التجريبية من مقاطع الفيديو عن طريق استخراج عدد من المعلومات من الفيديو إضافة إلى معلومات تؤخذ من

*Email: amalsa_comp@csw.uobaghdad.edu.iq

الشبكة. وقد استخدمت هذه المعلومات لتدريب نظام الاستدلال العصبي المضطرب القابل للتكيف (ANFIS) لتقدير جودة الفيديو.

1. INTRODUCTION

The increasing of access speed on the internet led to a large amount of online video on the internet and an increasing number of Consumers have own multimedia devices that allow users to watch online video anytime, anywhere and it tries to introduce these services to the person who actually uses a particular product with specific resolutions. The experience and customer satisfaction during the transport of video streams over network infrastructure is happening due to the success and acceptability of these new digital service offerings, but at the same time, the fact of being impaired in the service like delay, packet loss and jitter can severely damage the audiovisual quality as perceived by the end users.

The provider must ensure that customers have received appropriate quality at whole times. The variety of methods for measuring video quality aims to understand and ensuring a good user experience:

The research in the last years, Quality of Experience (QoE) term appeared and act as representative to user satisfaction about the quality of the displayed content and matching the computed quality to people's opinion like (bad, poor, good, very good and Excellent), it depends on the virtual experience of the human virtual system (HVS) with the video content [1].

It is important to differentiate between Quality of services (QoS) term and Quality of Experience (QoE) term because the first one refers to the performance of IP-based networks and services and the other depend on the degree of the end user delight to the application, service, or system [1,2].

The network provider viewpoint must measure video quality along the path that video delivery, so it easy to detect any fault happened. To measurement the points in the network we were using QoS metric parameters, such as packet loss, jitter, loss ratio, and arrival time.

QoS-based metrics models have the ability to distribute at any point in the network, but It does not have an idea on the user impression of the quality of the video being played. Subsequently, QoS models have a limitation to predict the video quality as perceived by the end-user. two types to measure video Quality of experience:

(1) Subjective video quality experience that has different scales for video sequence likes (excellent, good, fair, poor and bad). These linguistic variables will be reflected people's instant impression of a video sequence.

(2) Objective video quality metrics used to typically place the decoded video as seen by the end-user.

The objective video quality metrics can be divided into three types of (1) Full Reference (FR) metrics measure the video quality by comparisons the video with the quality of the original video. (2) Reduced Reference (RR) metrics are like FR metrics because RR has information about the original video, but it can be near as NR metrics because it has no information about the original video. (3) No Reference (NR) metrics measure the perceived quality using only the decoded video. The NR metrics are really difficult to design can be used on any video stream because it requires perfect models of human vision, object recognition, and quality judgment. There are researches on this field and this paper one of them [3, 4].

2. Related work

No-reference video quality of experience estimation is an important and challenging task which must be considered an essential part of future interactive multimedia communication applications such as IPTV or online games [5]. Researchers have also tried to develop some relationship between the quality of experience and Media -layer parameters. Also, for estimation the video quality, different measures/metrics can be used [5].

In [6] Bao and authors a new method of QoE estimation based on fuzzy clustering heuristic algorithm which is focused on service score at the server side. The server side is responsible for collect network QoS parameter and other information. Then save this information in a big database and use it in a heuristic rule model to predict user score, this process called fuzzy clustering analysis and then they will generate service QoE score that will feedback to clients.

In [7] Mocanua and the authors display a new metric that measures the user dissatisfaction which not always refer to averaged scores. Do that by using deep learning framework / deep belief networks and two modelled the average scores and user dissatisfaction levels.

In [8] Kawano and the authors have proposed a no reference QoE measurement system comprised of two training and test parts. In the first part training phase, they calculate the sensitivity of the low quality coded videos from some features such as blockiness, blurriness, edge, and continuity, etc. then rank these features using the Principal Component Analysis (PCA) method.

3. The Proposal System

The main objective is to design an efficient objective metric to measure the video quality of experience where it can predictive the MOS to video streaming. The proposed NR-QoE measurement system consists of two stages at the first the system is extracting different features (network-based, pixel-based) from streaming video, and in the second, training the system where we used neural network and fuzzy system.

Multi-level for the video QoE system: the proposed system consists of four steps the first step: data set selection, second step: feature extraction, third step: training phase, and fourth step: testing phase.

3.1 Data Set Selection

This work used numbers of datasets of video sequences with a good distribution of spatial and temporal properties to get the best result of the proposed system. The video quality of experience databases for testing our models :

1. ReTRiEVED Video Quality Database

The database contains eight source videos of different content by using Video LAN (NETEM) and Network Emulator that adjusting network parameter jitter, packet loss rate, throughput, and delay, so it generated from reference videos 184 test videos. The reference video is being used for the evaluation where Mean Opinion Score (MOS) and Difference Mean Opinion Score (DMOS) used as quality assessment metrics [9,10,11]. Figure-1 shows frames of the original video of the database used.



Figure 1-Sample images of source videos content of ReTRiEVED Video Quality database.

2. The LIVE-Netflix Video Quality of Experience database

It has 14 reference videos with 1080p as a resolution that they used to generate 112 distorted videos with 24, 25 and 30 fps evaluated by over 55 human subjects on a mobile device, get the Mean Opinion Scores (MOS) from the subjective evaluations [12,13]. Figure-2 show frames of the original videos of the database used.



Figure 2-Sample images of source videos content of LIVE-Netflix database.

3.2 Feature Extractions

Feature extraction stage is the most important phase in the prediction system that deals with extracting features from the raw data of information and use the relevant for classification purposes. In the extraction features process, there are many numbers of features extracted, but not all give as a

good result so must apply the dimensionality reduction of data by pruning the bad one. This is an important step because of the technical limits in memory and achieving the time.

In this proposed system we extracted from the sequence video frames two types of different features:

1. Network-based features that refer to it as QoS parameters. Most of the video characteristics can be taken from application layers such as bitrate, frame bitrate, pixel bitrate and others. While some features result from the network or channel performance such that Packet Loss Rate (PLR), Frame bitrate, Pixel bitrate, packet structure, video codec, and video content type and we also can calculate average PLR, Average of Frame Rate, Average of Bit Rate [5]. Sample of extracted features shows in Table-I.

Table I-NETWORK-BASED FEATURES

Video name	No. frame	Bitrate	Frame bitrate	Pixel bitrate	PLR
V1	25	6954610	278184	0.68602	0.1
V2	25	7129481	285179	0.70327	0.4
V3	25	7195273	287811	0.70976	1
V4	25	6794059	271762	0.67018	3
V5	25	6894681	275787	0.68011	5
V6	25	6447505	257900	0.636	8
V7	25	6356907	254276	0.62706	10
V8	25	3139928	125597	0.30973	0.1
V9	25	3126372	123593	0.30479	0.4
V10	25	3126366	125055	0.30839	1
V11	25	3101483	124059	0.30594	3
V12	25	2759193	110368	0.27217	5
V13	25	2663657	106546	0.26275	8
V14	25	2812350	112494	0.27742	10

2. Pixel-based features applied the features extraction process on video database and extract features like blocking, an average of blocking, blurring, an average of blurring, the blur ratio, Average of luminance, average noise, the noise ratio. Pixels-based features have a high correlation to the Human virtual System (HVS) and consequently impact the end user of video stream service. Eventually improved the accuracy performance of the prediction model [7,12,13]. Pixels-based features have a high correlation to the Human virtual System (HVS) and consequently impact the end user of video stream service. Eventually, improve the accuracy performance of the prediction model.

From the distorted videos, we extract the Spatial perceptual Information (SI) and Temporal perceptual Information (TI). These two features give us the complexity information in the videos that depend on the amount of distortion.

This type of information will be useful in the proposed system. We calculate the SI and TI values for each frame individually and then calculate the average of the features over the frames, equation 1 used to calculate SI and equations 2,3 used to calculate TI [14].

$$SI = \text{Max} \{sd_{i,j}[\text{Sobel}(\text{Fr}_n)]\} \quad (1)$$

$$TI = \text{Max}_{\text{time}} \{sd_{i,j}[M_n(i,j)]\} \quad (2)$$

$$M_n(i,j) = \text{Fr}_n(i,j) - \text{Fr}_{n-1}(i,j) \quad (3)$$

SI represents the Edge block that calculates by using Sobole filter and standard deviation (sd) of the frame pixels. TI value refers to the biggest video motion [15], that counted by found the maximum standard deviation video sequence, M_n refers to the pixel difference in successive frames at the same place.

Luminescence and average of luminescence deference are used to measure the impair quality of perception. equation 4 used to calculate the luminance of the single pixel [16], equation 5 used to calculate the average of luminescence deference [17, 18]. Table II shows a sample of extracted features.

$$\text{Pixel-lum} = (0.2989 R) + (0.5866 \times G) + (0.1145 \times B) \quad (4)$$

$$\text{Diff-lum} = \sum_{i=1}^x \sum_{j=1}^y |\overline{\text{Br}}_x(i,j) - \overline{\text{Br}}_{x-1}(i,j)| \quad (5)$$

Table II-PIXEL-BASED FEATURES

Video name	Lum	LumDiff	SI	TI
V1	8.539079203	85.10788385	17.06461465	65.16397593
V2	8.660247853	85.02203828	17.48668441	62.39615669
V3	8.519748052	85.0938272	16.92623517	66.11277563
V4	9.470517631	85.16416771	18.70986766	52.48608693
V5	9.162661799	84.9810674	19.9247935	71.78290429
V6	8.519707131	85.09014278	16.81514695	66.32441138
V7	8.534096553	85.10999489	16.60241191	52.01307474
V8	8.519707131	85.09014278	16.81514695	66.32441138
V9	8.399239124	85.13286587	16.70960969	69.70004949
V10	9.034755497	86.63403089	16.7811327	46.08300495
V11	8.552029608	85.22597284	16.11221547	73.83637765
V12	8.678576209	85.13621386	16.93871519	72.85598519
V13	8.533077537	85.21487672	16.11221547	73.83637765
V14	8.539079203	85.08599379	16.09368001	72.67646641

3.3 Classification and Clustering

Classification and clustering are the most important parts of the prediction process, in addition to feature extraction.

The previous phases should be designed and set aiming to succeed in the classification phase. The classifier depends on his decision on the feature vector value that extracted from the video sequence to be classified, as input, and then places it (MOS) to the suitable one [19].

The clustering operation and classification have similar concepts with each other. The main job of the clustering process is to group the feature vectors to clusters in which the correspondence of the patterns is stronger than between the clusters [20].

Training phase

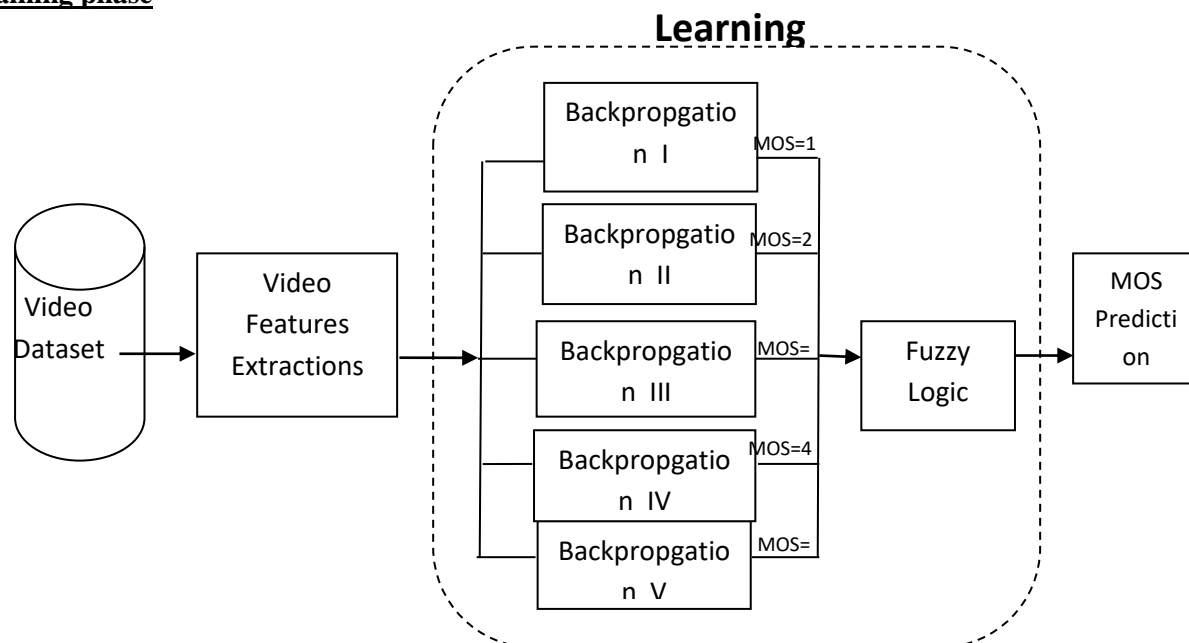


Figure 3-The Proposed System of Learning Model

1- The first level: Train the system with Neural Network algorithm to build a cluster module to detect the video has the lowest value of MOS from other videos.

The proposed system main idea depends on using five Neural Network of Back Propagation each one responsible on calculating the score of MOS that has value (1=worst, 2=poor, 3=fair, 4=good, and 5=excellent) and this value represent the input to the Fuzzy Inference system.

Each one of the neural networks has three layers (input, hidden, output) layers. The back propagation algorithm has local gradient value used for updating the weights. In the first for each output neuron calculating the local gradient, updated weights value that backtracking from the neurons in the hidden layer to the neurons in input and calculated the local gradients at the neurons in the hidden layer. The repetitions procedure responsible to update the weights in two path from the output neurons to the hidden neurons and from the hidden neurons to the input neurons.

In feedforward phase, the input layer has the unit ($X_1 \dots X_{10}$) represent the feature value that receives from the previous phase and distribute it's to the hidden layer that has ($Z_1, \dots Z_{15}$) units. After that, the hidden units compute the activation function and sends the result to the output layer. In the training cycle, each output unit compares its computed activation with its target value to determine the associated error for that pattern with that unit. During the backpropagation phase of learning, signals are sent in the reverse direction [21, 22].

The proposed system used Back Propagation (BP) algorithm because of its accuracy where it allows itself to learn and improve itself. The network in this system contains 10 units in the input layer that refer to extracted features, 15 unities in the output layer and one unit in the output layer that represents the predicates Main Opinion Score figure (4) shows the structure of the used neural network. Numbers of iterations increase within the recommended limits until they can reach the appropriate parameter values. ANFIS model training can be completed at the end of the iteration. After this step, the model can control the problem [18]. ANFIS is an off-line learning model and it is completely used in modeling and control of dynamic systems by making a set of fuzzy if-then rules with suitable MFs [20, 23].

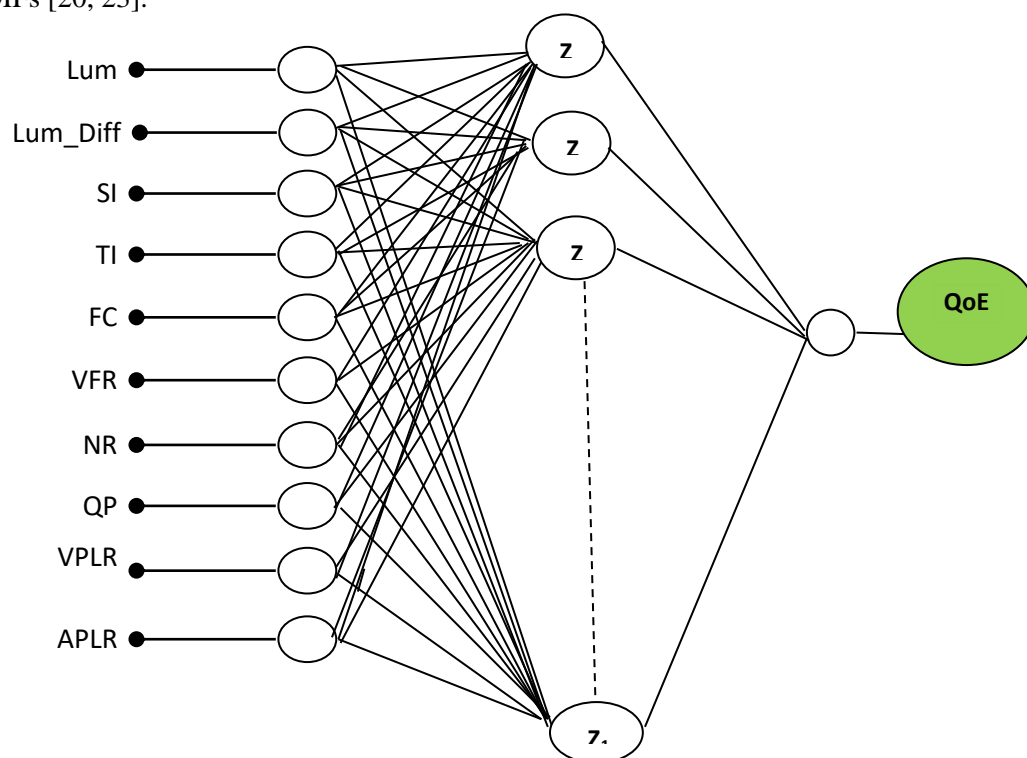


Figure 4-Artificial Neural Network Architecture

2- The second level: Train with fuzzy algorithm to build classifier module to detect the prediction value to Mean Opinion Score by framing the output of neural networks in rule and used them to prediction the final MOS [18]. The fuzzy logic role it to decided which output will take and display as

result to MOS, it do that by making numbers of rule to help it in classification process. Example of fuzzy rules:

if (net5 > net1 && net5 > net2 && net5 > net3 && net5 > net4) QoE = 5

if (net1 > net2 && net1 > net3 && net1 > net4 && net1 > net5) QoE = 1

Table III display the result of MOS before training and after training, Table IV show the result of each neural net. In figure 5 display the correlation between the original QoE (red line) and QoE assessment from the system (blue line).

Table III- MOS RESULT BEFORE TRAINING AND AFTER TRAINING

MOS from system	4	2	2	4	3	4	2	2	5	3	5
MOS from Dataset	3.1897	1.25	2	3.8	3.2083	3.9483	1.15	1.7321	4.1034	2.8148	4.103448

Table IV-THE RESULT TO EACH NEURAL NET

Network 1	100%
Network 2	99.38%
Network 3	90.60%
Network 4	86.88%
Network 5	96.88%

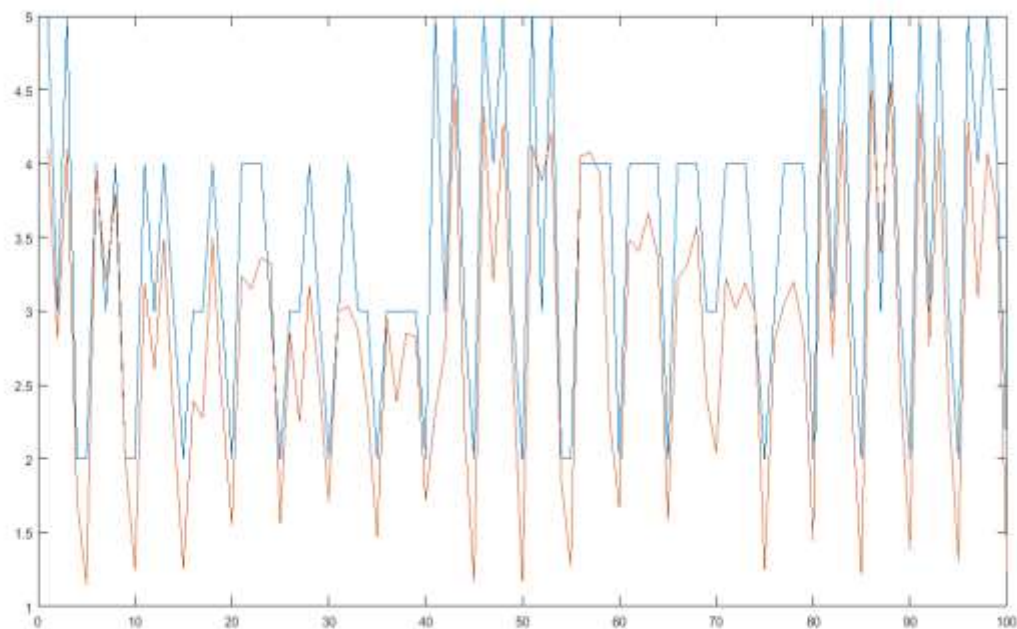


Figure 5- Changing between training data and the ANFIS output.

4. CONCLUSIONS

In this work, the quality of experience of streaming video is predicted using the ANFIS model. This model can be easily used video parameters and control on the output.

It is showed that the ANFIS model is compatible with the experimental results. The model, which has the learning ability and the parallel computation of five training back propagation algorithm which is a supervised learning method and fuzzy inference system, is a hybrid artificial intelligence method. It has the advantages of both ANN and FL methods and gives an opportunity for the solution of critic and complex problems.

Reference

1. Bentes, O., Camille, H., and De Errico, L. **2015**. A concise review of the quality of experience assessment for video streaming. *Computer Communication* Elsevier, **57**: 1–12.
2. Tsihrintzis, G., and Jain, L. **2008**. *Multimedia Services in Intelligent Environments: Advanced Tools and Methodologies*. Springer Science & Business Media, **120**(9).
3. Nicolas, S., Margaret, H., Philip, C., Filip, T., and Piet, D. **2015**. Measuring Video Quality In The Network: From Quality Of Service To User Experience. the Ninth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM) Chandler, AZ.
4. Arun, R. L., Dhananjay, K., Iswarya, H., S., and Srinivasan, A. **2017**. Adaptive video streaming over HTTP through 4G wireless networks based on buffer analysis. *EURASIP Journal on Image and Video Processing*, **41**(3).
5. Søggaard, J., Forchhammer, and S., Korhonen, J. **2015**. Video Quality Assessment and Machine Learning: Performance and Interpretability, Proceedings of IEEE QoMEX.
6. Bao, Y., Lei, W., Zhang, W. and Zhan, Y. **2016**. QoE collaborative evaluation method based on fuzzy clustering heuristic algorithm. *Springer Plus* ,**5**(1).
7. Mocanu, D.C., Pokhrelb, J., Garellac, J.P., Seppanen, Liotou, J.E. and Narwaria, M. **2015**. No-reference video quality measurement: The added value of Machine learning. *Journal of Electronic Image*, **24**(6).
8. Kawano, T., Yamagishi, K., Watanabe, K. and Okamoto, J. **2010**. No reference video-quality-assessment model for video streaming services. 18th International Packet Video Workshop, IEEE Explore Digital Library, pp.158–164.
9. Paudyal, P., Battisti, F. and Carli, M. **2014**. A Study on The Effects of Quality Of Service Parameters On Perceived Video Quality. Proceedings of 5th European Workshop on Visual Information Processing, France.
10. Battisti, F., Carli, M. and Paudyal, P. **2014**. QoS to QoE Mapping Model for Wired/Wireless Video Communication. Proceedings of European Med. Telecommunication Conference, Italy.
11. Paudyal, P., Battisti, F. and Carli, M. **2015**. Study of the Effects of Video Content on Quality of Experience. Proceedings of Image Quality and System Performance XII-IS&T/SPIE Electronic Imaging, United States.
12. Bampis, C.G., Li, Z., Moorthy, A.K., Katsavounidis, I., Aaron, A. and Bovik, A.C. **2017**. Study of Temporal Effects on Subjective Video Quality of Experience. *IEEE Trans. Image Proceedings* **26**(11):5217–5231.
13. Bampis, C.G., Li, Z., Moorthy, A.K., Katsavounidis, I., Aaron, A. and Bovik, A.C. **2016**. LIVE Netflix Video Quality of Experience Database. Online: http://live.ece.utexas.edu/research/LIVE_NFLXStudy/index.html.
14. Tsihrintzis, G. and Jain, L. **2008**. *Multimedia Services in Intelligent Environments: Advanced Tools and Methodologies*. Springer Science Bus. Media, **120**.
15. Mwela, J.S. **2010**. Impact of Packet Loss on the Quality of Video Stream Transmission. M.Sc. Thesis, Electrical Engineering department, School of Computing Blekinge, Sweden.
16. Feghali, R., Speranza, F., Wang, D., and Vincent, A. **2007**. Video Quality Metric for Bit Rate Control via Joint Adjustment of Quantization and Frame Rate. *IEEE Trans. Broadcast.*, **53**(1): 441-446.
17. Hameed, A., Dai, R. and Balas, B., **2016**. A Decision-Tree-Based Perceptual Video Quality Prediction Model and Its Application in FEC for Wireless Multimedia Communications. *IEEE Trans. Multimedia*, **18**(4):764–774.

18. Jang, JSR., **1993**. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE T System Man Cy.*, **23**: 665-685.
19. Takagi, H. and Hayashi, I., **1991**. NN-driven fuzzy reasoning. *International Journal Approximation Reason.*, **5**: 91-212.
20. Melin, P., and Castillo, O. **2005**. Intelligent control of a steeping motor drive using an adaptive neuro- fuzzy inference system. *Information Sciences*, **170**:133-151.
21. Simon, H. **1999**. Neural Network Comprehensive Foundation. 2nd edition, Pearson Education Inc., Third Chapter, Canada, pp. 142.
22. Achkar, R. and Nasr, C. **2010**. Real Time Application of an AMB using MLP: Study of Robustness. IEEE Second International Conference on Computational Intelligence Modeling and Simulation, Indonesia, pp. 9-14.
23. Mon, YJ. **2007**. Airbag controller designed by adaptive- network-based fuzzy inference system (ANFIS). *Fuzzy Sets System*, **158**: 2706-2714.