An ANFIS-based Hybrid Video Quality Prediction Model for Video Streaming over Wireless Networks

Asiya Khan, Lingfen Sun and Emmanuel Ifeachor
Centre for Signal Processing and Multimedia Communication
School of Computing, Communications and Electronics
University of Plymouth, Plymouth PL4 8AA, UK.
(asiya.khan, l.sun, e.ifeachor)@plymouth.ac.uk

Abstract

There are many parameters that affect video quality but their combined effect is not well identified and understood when video is transmitted over mobile/ wireless networks. In this paper our aim is twofold. First, to study and analyze the behaviour of video quality for wide range variations of a set of selected parameters. Second, to develop a learning model based on ANFIS to estimate the visual perceptual quality in terms of the Mean Opinion Score (MOS) and decodable frame rate (Q value). We trained three ANFIS-based ANNs for the three distinct content types using a combination of network level and application level parameters such as frame rate, send bitrate, link bandwidth and packet error rate and tested the ANN models using unseen dataset. We found that the video quality is more sensitive to network level parameters compared to application level parameters. Preliminary results show that a good prediction accuracy was obtained from the ANFIS-based ANN model. The work should help in the development of a reference-free video prediction model and Quality of Service (QoS) control methods for video over wireless/mobile networks.

Keywords. *ANFIS, neural networks, MOS, MPEG4, video quality evaluation.*

1. Introduction

It is of great importance to develop efficient and reliable mechanisms to assess the quality of video streams for the success of multimedia communications over wireless networks. Video quality can be evaluated either subjectively or based on objective parameters. Subjective quality is the users' perception of service quality (ITU-T P.800) [1]. The most widely used metric is the mean opinion score (MOS). While subjective quality is the most reliable method, it is time consuming and expensive and hence, the need for an objective method that produces results comparable with those of subjective testing. Objective measurements can be performed in an intrusive or non-intrusive way. Intrusive measurements require access to the source then compares the original and impaired videos. Full reference and reduced reference video quality measurements are both intrusive [2]. Quality metrics such as Peak-Signal-to-Noise-Ratio (PSNR), more recently the Q value [3], VQM [4] and PEVQ [5] are full reference metrics. VQM and PEVQ are commercially used and are not publicly available. Non-intrusive methods (reference-free), on the other hand do not require access to the source video. Non-intrusive methods are either signal or parameter based. Non-intrusive methods are preferred to intrusive analysis and they are more suitable for on-line quality prediction/control.

In this paper we estimate the perceptual video quality through a reference-free parameter based learning model. There are many parameters that affect video quality and their combined effect is unclear, and their relationships are thought to be non-linear. Artificial Neural Networks (ANNs) can be used to learn this non-linear relationship which mimics human perception of video quality. ANN has been widely used in assessing the video quality from both network and application based parameters. In [6,7] the authors have developed neural-network models to predict video quality based on application and network parameters. The work was based on video subjective tests to form training and testing datasets. Further, different video contents have not been considered in developing neural network models and their work is only limited in fixed IP networks. Similarly, in [8] authors have proposed a parametric model for estimating the quality of videophone services that can be used for application and/or network planning and monitoring. Recent work has also shown the importance of video content in predicting video quality. In [9,10,11] features representing video content were used to predict video quality together with other application-level parameters such as send bitrate and frame rate. However, this work did not consider any network-level parameters in video quality prediction. Work in [12] is only based on network parameters. (e.g. network bandwidth, delay, jitter and loss) to predict video quality with no consideration of application-level parameters.

In this paper, we aim to investigate the combined effects of network and application parameters on end-to-end perceived video quality over wireless networks for three distinct content types and further develop a hybrid video quality prediction model based on an Adaptive Neural-Fuzzy Inference System (ANFIS) [13] as it combines the advantages of a neural network and fuzzy system. We use ANFIS to train the three neural networks for three distinct content types to predict the video quality based on a set of objective parameters. The ANN is validated with three different contents in the corresponding categories. We predict video quality (in terms of MOS score and Q-value[3]) from both network and application parameters for video streaming over wireless network application. We used frame rate and send bitrate as application level and packet error rate and link bandwidth as network level parameters. Our focus ranges from low resolution and send bitrate video streaming for 3G applications to higher video send bitrate for WLAN applications depending on the type of the content and network conditions. Our proposed test bed is based on simulated network scenarios using a network simulator NS-2 with an integrated tool Evalvid [14]. It gives a lot of flexibility for evaluating different topologies and parameter settings used in this study.

The paper is organised as follows. In section 2 we introduce the simulation set-up and platform, the test sequences and variable test parameters. In section 3 we briefly describe the ANFIS neural network structure and training methods. Section 4 discusses the impact of parameters on video quality, whereas, in section 5 we evaluate the performance of the proposed artificial neural network and compare with existing results. Section 6 concludes the paper and outlines the future directions of our research.

2. Evaluation Set-up

This section describes the simulation set-up and platform, defines the test sequences and variable test parameters.

2.1 Simulation set-up

The experimental set up is given in Fig 1. There are two sender nodes as CBR background traffic and MPEG4 video source. Both the links pass traffic at 10Mbps, 1ms over the internet which in turn passes the traffic to another router over a variable link. The second router is connected to a wireless access point at 10Mbps, 1ms and further transmits this traffic to a mobile node at a transmission rate of 11Mbps 802.11b WLAN. We assume that no packet loss occurs in the wired segment of the video delivered path. The max transmission packet size is 1024 bytes. The video packets are delivered with the random uniform error model. The CBR rate is fixed to 1Mbps. The packet error rate is set in the range of 0.01 to 0.2 with 0.05 intervals. To account for different packet loss patterns, 10 different initial seeds for random number generation were chosen for each packet error rate. All results generated in the paper were obtained by averaging over these 10 runs.

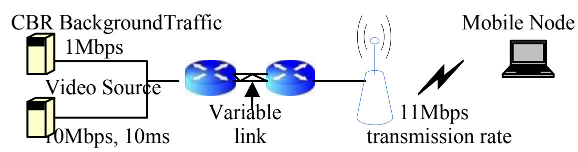

Fig. 1 Simulation setup

2.2 Simulation platform

All the experiments in this paper were conducted with an open source framework Evalvid [14] and network simulator tool NS2 [15]. Video quality is measured by taking the average PSNR over all the decoded frames. MOS scores are calculated based on the PSNR to MOS conversion from Evalvid [14] given in table 1 below. Further the decodable frame rate (Q) [3] was also obtained for the same testing combinations.

Table 1. PSNR to MOS Conversion

PSNR (dB)	MOS
>37	5
31 – 36.9	4
25 – 30.9	3
20 – 24.9	2
< 19.9	1

2.3 Test sequences

The test sequences used were divided in three categories as slight movement, gentle walking and rapid movement depending on the type of content subjectively. We selected one video representing each category. We selected the video sequence 'akiyo' in the first category in which a female moderator is reading the news only by moving her lips and eyes. It represents the news scenario. In the second category we chose the video sequence 'foreman'. Foreman sequence contains a monologue of a man moving his head dynamically and at the end of the sequence there is a rapid scene change. We chose 'stefan' as the video sequence in the third category. It is a wide angle sequence in which two players are playing tennis.

Table 2. Testbed combinations

Video sequence	Frame Rate	SBR kb/s	Link BW (kb/s)	PER
Slight Movement	10, 15, 30	18	32, 64, 128	0.01, 0.05, 0.1, 0.15, 0.2
	10, 15, 30	44		
	10, 15, 30	80		
Gentle Walking	10, 15, 30	18	128, 256, 384	0.01, 0.05, 0.1, 0.15, 0.2
	10, 15, 30	44		
	10, 15, 30	80		
Rapid Movement	10, 15, 30	80	384, 512, 768, 1000	0.01, 0.05, 0.1, 0.15, 0.2
	10, 15, 30	104		
	10, 15, 30	512		

All sequences were encoded with MPEG4 video codec available from [16]. We used the combination of Frame Rate (FR), video Send Bitrate (SBR), Link Bandwidth (LBW) and Packet Error Rate (PER) as shown in Table 2. In total there were 135 encoded test sequences for the first two content categories and 180 encoded test sequences for the third content category.

2.4. Variable test parameters

We considered the following quality affecting parameters both in the application level and the network level as follows:

Application Level parameters: The frame rate: the number of frames per second. It takes one of three values as 10, 15 and 30fps. The send bitrate: the rate of the encoders output. It is chosen to take 18, 44, 80kb/s for slight movement and gentle walking whereas, 80, 104 and 512kb/s for rapid movement.

Network Level Parameters: The link bandwidth: the variable bandwidth link between the routers (Fig. 1). It takes the values of 32, 64 and 128kb/s for 'slight movement', 128, 256, and 384kb/s for 'gentle walking' and 384, 512, 768 and 1000kb/s for 'rapid movement'. Packet Error Rate: the simulator (NS-2) [15] drops packet at regular intervals using a random uniform error model. It takes five values as 1, 5, 10, 15 and 20%. It is widely accepted that a loss rate higher than 20% will drastically reduce the video quality.

3. ANFIS-based ANN learning model

The aim is to develop three ANFIS-based learning models to predict video quality for three distinct content types from both network and application parameters for video streaming over wireless networks application as shown in Fig. 2. The application level parameters considered are Frame Rate (FR) and Send Bit Rate (SBR). The network parameters are Packet Error Rate (PER) and Link Bandwidth (LBW).

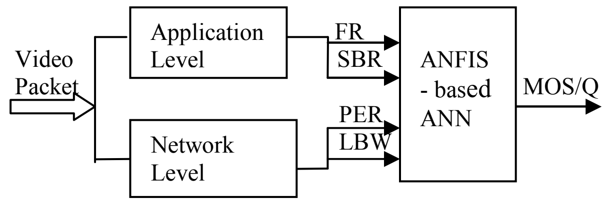

Fig. 2 Functional block of proposed ANFIS-based model

3.1 ANFIS architecture

The corresponding equivalent ANFIS architecture [13] is shown in Fig. 3.

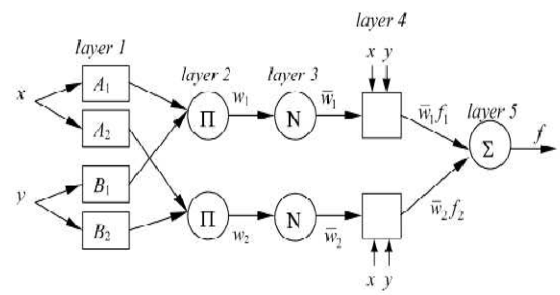

Fig. 3 ANFIS architecture [13]

The entire system architecture consists of five layers, namely, a fuzzy layer, a product layer, a normalized layer, a defuzzy layer and a total output layer. The inputs x and y in our paper are frame rate, send bitrate, link bandwidth and packet error rate. The output *f* is the MOS score and Q value.

3.2 Training and testing of ANFIS-based ANN

For ANNs, it is not a challenge to predict patterns existing on a sequence with which they were trained.

The real challenge is to predict sequences that the network did not use for training. However, the part of the video sequence to be used for training should be 'rich enough' to equip the network with enough power to extrapolate patterns that may exist in other sequences. For this reason, the three ANFIS-based

ANN models were trained with the three distinct content types of ‘akiyo’, ‘foreman’ and ‘stefan’ (see Table 2) and validated by three different content types of ‘suzie’, ‘carphone’ and ‘rugby’ in the corresponding content categories. The data selected for validation was one third that of testing with different parameter values to that given in Table 2.

4. Impact of Parameters on Video Quality

In this section we study the effects of the four parameters on video quality. We chose three-dimensional figures in which two parameters were varied while keeping the other two fixed. The MOS value is computed as a function of the values of all four parameters.

4.1 Mos vs Send Bitrate vs Packet Error Rate

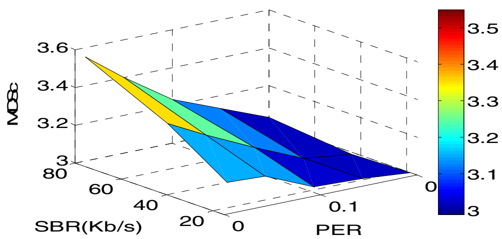

Fig. 4 MOSc vs SBR vs PER for ‘Slight movement’

Fig. 4 shows the MOS scores for ‘slight movement’. The frame rate was kept fixed at 10fps and the link bandwidth was fixed at 128kb/s. We observed that the MOSc dropped to 3 when the packet loss was 20% which is an acceptable value for communication quality. This shows that when the there is very little activity in content the video quality is still acceptable at low send bitrates and with high packet loss.

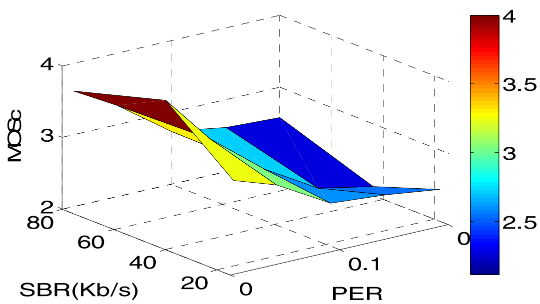

Fig. 5 MOSc vs SBR vs PER for ‘Gentle walking’

Fig. 5 show the MOS scores for ‘gentle walking’. The frame rate is fixed at 10fps and the link bandwidth at 384kb/s. We observe that with higher send bitrate of 80kb/s the video quality is very good (MOSc > 3.5), however, the quality fades rapidly with increasing packet loss.

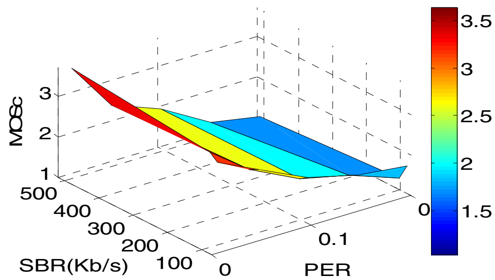

Fig. 6 MOSc vs SBR vs PER for ‘Rapid movement’

Fig. 6 show the MOS scores for ‘rapid movement’. The frame rate was kept fixed at 10fps and the link bandwidth was fixed at 512kb/s. Again, the video quality is very good for higher send bitrate of 512kb/s, but fades very rapidly with increasing packet loss.

4.2 MOSc vs Send Bitrate vs Link Bandwdth

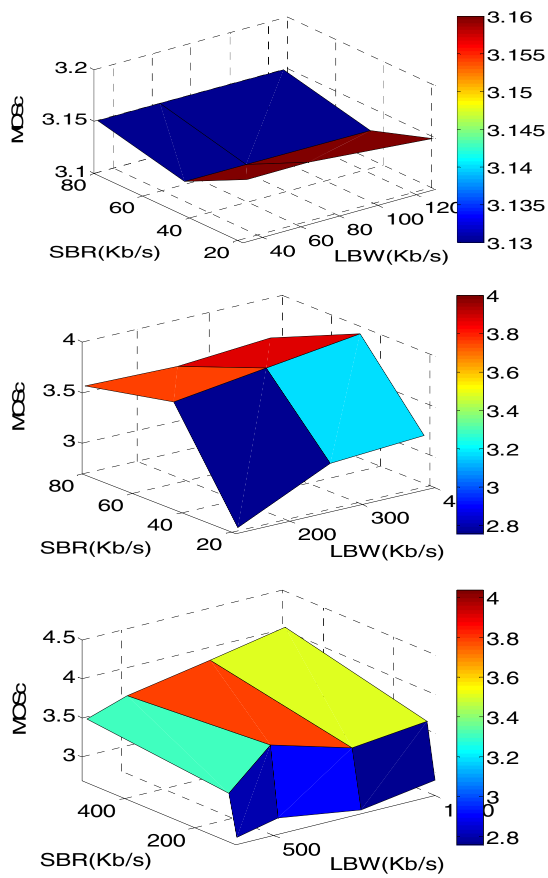

Fig. 7a, b & c MOSc vs SBR vs LBW for ‘Slight movement’, ‘Gentle walking’ & ‘Rapid movement’

In Figs 7a, b and c the frame rate is fixed at 10fps without packet loss, for three content types, increasing the link bandwidth only improves the MOS score if the video is encoded at a bitrate less than the LBW. However, we observe a worsening in video quality if the send bitrate is more than the link bandwidth.

5. Evaluation of the ANN

We trained three ANFIS-based learning models for the three distinct content types and validated them with three different video test sequences in the corresponding content categories. The accuracy of the ANN can be determined by the correlation coefficient and the RMSE of the validation results [17]. For the three content types we obtained results in terms of the MOS score and decodable frame rate Q [3].

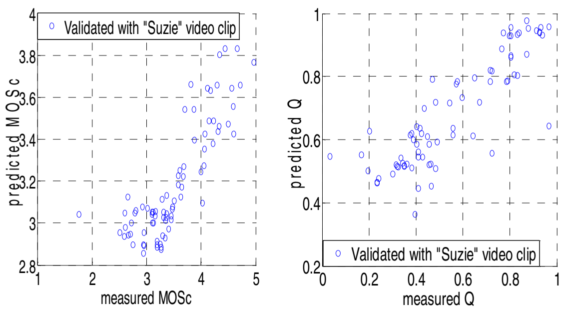

Fig. 8 ANN mapping of MOSc and Q for 'Slight movement'

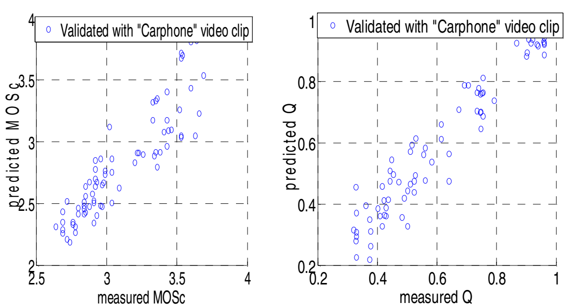

Fig.9 ANN mapping of MOSc and Q for 'Gentle walking'

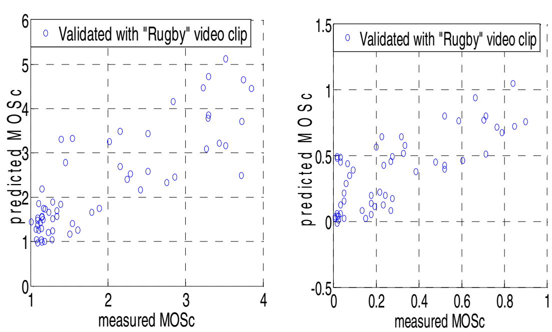

Fig.10 ANN mapping of MOSc and Q for 'Rapid movement'

We carried out a linear regression analysis between the predicted and measured MOS scores and Q value to aim to achieve y = x (see Figs. 8, 9 and 10). However, more realistically the relationship between the measured MOS/Q (x) and the predicted MOS/Q (y) is represented as $y = a_1x + a_2$.

We aim to achieve a_1 as close to 1 as possible and a_2 close to 0. For the three content types we obtained the following results given in table 3:

Table 3. Coefficients for the linear model

Content	a_1(MOS/Q)	a_2 (MOS/Q)
Slight Movement	0.3696/0.6241	1.8999/0.3359
Gentle Walking	1.222/1.032	-0.9855/-0.03716
Rapid Movement	1.014/0.7898	0.3247/0.1476

The validation results of the proposed ANFIS-based ANN in terms of the correlation factor and the root mean squared error (RMSE) between the predicted and measured MOS/Q for all three content types is given in Table 4 below.

Table 4. Validation results of ANFIS-based ANN by correlation coefficient and RMSE

Content type	Correlation coef (MOS/Q)	RMSE (MOS/Q)
Slight Movement	0.7007/0.7384	0.1545/0.08813
Gentle Walking	0.8056/0.9229	0.1846/0.06234
Rapid Movement	0.7034/0.5845	0.6193/0.1816

We achieved better correlation for 'gentle walking' compared to 'rapid movement' and 'slight movement'. Also the ANFIS-based ANN gave better results for Q value compared to MOS for 'gentle walking'. We also observed that video clips in 'rapid movement' are very sensitive to packet loss. The quality degrades rapidly compared to the other two categories as packet loss is introduced. In future, we are looking at classifying the three content types objectively.

5.1 Comparison with recent published work

A recent work that has estimated video quality based on ANNs is presented in [11]. Our results in terms of the correlation coefficients and mean squared error are comparable to theirs. However, they have not taken into account the effect of network parameters on video quality and also the video sequences we chose for validation are completely different to those for testing confirming the right choice of objective parameters and hence, a reliable tool for video quality prediction.

6. Conclusions

In this paper, we investigated the combined effects of application and network parameters on end-to-end perceived video quality and analyzed the behaviour of video quality for wide range variations of a set of selected parameters. We further, developed an ANFIS-based learning model from the suitable parameter range to predict video quality from both network and application parameters for video streaming over wireless network application.

We observed that network level parameters like link bandwidth and packet error rate have a much bigger impact on video quality compared to application level parameters such as frame rate and video send bitrate. We also found that if the video stream is encoded at a send bitrate greater than the link bandwidth then video quality is degraded.

Further, from the ANFIS-based ANN our results demonstrates that it is possible to predict the video quality if the appropriate parameters are chosen. The correlation coefficient and RMSE for MOS scores were generally better than decodable frame rate except in 'gentle walking' where Q results were better. Our results confirm that the proposed ANFIS-based ANN learning model is a suitable tool for video quality estimation for the most significant video streaming content types.

Our future work will focus on classifying the content objectively and to propose feedback mechanisms that could dynamically generate video streams that can adapt the send bitrate in the most efficient way taking into account network conditions to achieve optimum end-to-end video quality.

7. Acknowledgement

The work reported here is supported in part by the EU FP7 ADAMANTIUM project (contract No. 214751).

8. References

[1] ITU-T. Rec P.800, Methods for subjective determination of transmission quality, 1996.

[2] Video quality experts group, multimedia group test plan, Draft version 1.8, Dec 2005, www.vqeq.org.

[3] C. Lin, C. Ke, C. Shieh and N. Chilamkurti, "The packet loss effect on MPEG video transmission in wireless networks", *Proc. of the 20th Int. Conf. on Advanced Information Networking and Applications (AINA)*.Vol. 1, 2006, pp. 565-72.

[4] http://compression.ru/video/index.htm

[5] www.pevq.org

[6] 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* , Publisher, Vol. 12, No. 12. Dec. 2002, pp. 1071-83.

[7] P. Frank and J. Incer, "A neural network based test bed for evaluating the quality of video streams in IP networks", *Proceedings of the Electronics, Robotics and Automotive Mechanics Conference (CERMA),* Vol. 1, Sept. 2006, pp. 178-83.

[8] K. Yamagishi, T. Tominaga, T. Hayashi and A. Takashi, "Objective quality estimation model for videophone services", *NTT Technical Review*, Vol. 5, No. 6, June 2007.

[9] P. Gastaldo, S. Rovetta, and R. Zunino, 2001, "Objective assessment of MPEG-video quality: a neural network approach", *IEEE International Joint Conference on Neural Networks proceedings IJCNN,* Vol. 2, 2001, pp. 1432-37.

[10] M. Ries, O. Nemethova and M. Rupp, "Video quality estimation for mobile H.264/AVC video straming", *Journal of Communications*, Vol. 3, No.1, Jan. 2008, pp. 41-50.

[11] M. Ries, J. Kubanek, and M. Rupp, "Video quality estimation for mobile streaming applications with neuronal networks", *Published in the Proceedings of MESAQIN Conference,* Prague, Czech Republic, 5–6 June, 2006.

[12] P. Calyam, E. Ekicio, C. Lee, M. Haffner and N. Howes, "A gap-model based framework for online VVoIP QoE measurement", *Journal of Communications and Networks***,** **V**ol. 9, No.4, Dec. 2007, pp. 446-56.

[13] The application of an ANFIS and Grey system method in turning tool-failure detection, *Advanced Manufacturing Technology,* 2002.

[14] J. Klaue, B. Tathke, and A. Wolisz, "Evalvid – A framework for video transmission and quality evaluation", *In Proc. Of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation,* Urbana, Illinois, USA, 2003, pp. 255-272.

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

[16] Ffmpeg, http://sourceforge.net/projects/ffmpeg

[17] VQEG: "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment", 2000, available at http://www.vqeg.org/