Content-Based Video Quality Prediction for MPEG4 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.
Email: 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 addition, video content has an impact on video quality under same network conditions. The main aim of this paper is the prediction of video quality combining the application and network level parameters for all content types. Firstly, video sequences are classified into groups representing different content types using cluster analysis. The classification of contents is based on the temporal (movement) and spatial (edges, brightness) feature extraction. Second, to study and analyze the behaviour of video quality for wide range variations of a set of selected parameters. Finally, to develop two learning models based on – (1) ANFIS to estimate the visual perceptual quality in terms of the Mean Opinion Score (MOS) and decodable frame rate (Q value) and (2) regression modeling to estimate the visual perceptual quality in terms of the MOS. We trained three ANFIS-based ANNs and regression based-models for the three distinct content types using a combination of network and application level parameters and tested the two models using unseen dataset. We confirmed that the video quality is more sensitive to network level compared to application level parameters. Preliminary results show that a good prediction accuracy was obtained from both models. However, the regression based model performed better in terms of the correlation coefficient and the root mean squared error. 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.**

Index Terms— **ANFIS, neural networks, Content clustering, MOS, MPEG4, video quality evaluation.**

I. INTRODUCTION

Streaming video services are becoming commonplace with the recent progress of broadband access lines. Video content will be the main contributor to the future traffic in multimedia applications. Perceived quality of the streaming videos is likely to be the major determining factor in the success of the new multimedia applications. It is therefore important to choose both the application level i.e. the compression parameters as well as network setting so that they maximize end-user quality.

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 as they are more suitable for on-line quality prediction/control.

In this paper we aim to recognize the most significant content types, classify them using cluster analysis [6] based on the temporal (movement) and spatial (edges, brightness) feature extraction and 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 [7],[8] 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 [9],[10] authors have proposed an opinion and parametric model for estimating the quality of interactive multimodal and videophone services that can be used for application and/ or network planning and monitoring. However, in these work content types are not considered. Whereas, in [11] a theoretical framework is proposed based on both application and network level parameters to predict video quality. 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 [13] we have proposed an ANFIS-based prediction model that considers both application and network level parameters with subjective content classification. Recent work has also shown the importance of video content in predicting video quality. In [14],[15][16][17] video content is classified based on the spatial (edges, colours, etc) and temporal (movement, direction, etc) feature extraction which were then 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. Video content is classified in [18],[19] based on content characteristics obtained from users' subjective evaluation using cluster [6] and Principal Component Analysis (PCA) [20]. In [21],[22] authors have used a combination of PCA [20] and feature extraction to classify video contents.

This motivated us to look into ways of classifying video content based on feature extraction using cluster analysis [6]. Furthermore, based on the content types we are looking for an objective measure of video quality simple enough to be calculated in real time at the receiver side. We present two new reference-free approaches for quality estimation for all content types[13],[23].

The contributions of the paper are three-fold:
(1) Most frequent content types are classified into three main groups by extracting temporal (movement) and spatial (edges, brightness) feature using a well known tool called cluster analysis [6]. (2) Second, 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. (3)Third, we develop two models for video quality estimation as (a) a hybrid video quality prediction model based on an Adaptive Neural Fuzzy Inference System (ANFIS), as it combines the advantages of a neural network and fuzzy system [24] for the three content types [13]. (b) a regression based model for the three content types [23]. 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 ANFIS-based 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. For the regression based model we used frame rate and send bitrate as application level and packet error rate as network level parameters and estimate video quality in terms of the MOS only. 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 [25] with an integrated tool Evalvid [26]. It gives a lot of flexibility for evaluating different topologies and parameter settings used in this study.

The paper is organized as follows. In section II we introduce the test sequences, simulation set-up and platform and the variable test parameters. Section III classifies the contents using cluster analysis. In section IV we discuss the impact of parameters on video quality. Section V briefly describes the ANFIS neural network structure and training methods whereas, section VI evaluates the performance of the proposed artificial neural network. Section VII outlines the regression based video quality model and compares the two models and with existing results. Section VIII concludes the paper and outlines the future directions of our research.

II. Evaluation set-up

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

A. Simulation set-up and platform

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. 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.
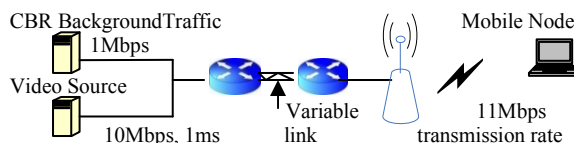

Figure. 1 Simulation setup

All the experiments in this paper were conducted with an open source framework Evalvid [26] and network simulator tool NS2 [25]. Video quality is measured by taking the average PSNR over all the decoded frames.

Further the decodable frame rate (Q) [3] was also obtained for the same testing combinations.

PSNR given by (1) computes the maximum possible signal energy to noise energy. PSNR measures the difference between the reconstructed video file and the original video trace file.

$$PSNR(s,d)]_{db} = 20 \log \frac{Vpeak}{MSE(s,d)} \qquad (1)$$

Mean Square Error (MSE) is the cumulative square between compressed and the original image.

Decodable frame rate (Q) [3] is defined as the number of decodable frames over the total number of frames sent by a video source. Therefore, the larger the Q value, the better the video quality perceived by the end user. MOS scores are calculated based on the PSNR to MOS conversion from Evalvid [26] given in Table I below.

TABLE I
PSNR TO MOS CONVERSION

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

B. Test sequences and variable test parameters

For the tests we selected nine different video sequences of qcif resolution (176x144) and encoded in MPEG4 format with an open source ffmpeg [27] encoder/decoder with a Group of Pictures (GOP) pattern of IBBPBBPBB. Each GOP encodes three types of frames - Intra (I) frames are encoded independently of any other type of frames, Predicted (P) frames are encoded using predictions from preceding I or P frames and Bi-directionally (B) frames are encoded using predictions from the preceding and succeeding I or P frames

A GOP pattern is characterized by two parameters, GOP(N,M) – where N is the I-to-I frame distance and M is the I-to-P frame distance. For example, as shown in Fig.1, G(9,3) means that the GOP includes one I frame two P frames, and six B frames. The second I frame marks the beginning of the next GOP. Also the arrows in Fig. 1 indicate that the B frames and P frames decoded are dependent on the preceding or succeeding I or P frames [28].
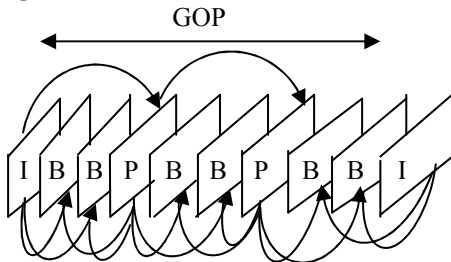

Figure. 2 A sample of MPEG4 GOP (N=9, M=3)

The chosen video sequences ranged from very little

movement, i.e. small moving region of interest on static background to fast moving sports clips. Each of the testbed sequences represent typical content offered by network providers.

For quality evaluation we used a combination of application and network level parameters as Frame Rate (FR), Send Bitrate (SBR), Link Bandwidth (LBW) and Packet Error Rate (PER). The video sequences along with the combination parameters chosen are given in Table II. In total, there were 1500 encoded test sequences.

In the application level we considered: (1) The frame rate - the number of frames per second. It takes one of three values as 10, 15 and 30fps. (2) 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.

In the network level we considered: (1) 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'. (2) Packet Error Rate: the simulator (NS-2) [25] drops packet at regular intervals using the random uniform error model, taking five values as 0.01, 0.05, 0.1, 0.15 and 0.2. It is widely accepted that a loss rate higher than 0.2 (20%) will drastically reduce the video quality.

TABLE II
TESTBED COMBINATIONS

Video sequences	Frame Rate (fps)	SBR (kb/s)	Link BW (kb/s)	PER
Akiyo, Suzie, Grandma	10, 15, 30	18	32	0.01, 0.05, 0.1, 0.15, 0.2
	10, 15, 30	44	64	
	10, 15, 30	80	128	
Carphone, Foreman	10, 15, 30	44	256	0.01, 0.05, 0.1, 0.15, 0.2
	10, 15, 30	80	256	
Rugby, Stefan, Table Tennis, Football	10, 15, 30	128	384	0.01, 0.05, 0.1, 0.15, 0.2
	10, 15, 30	104	512	
	10, 15, 30	384	768	
	10, 15, 30	512	1000	

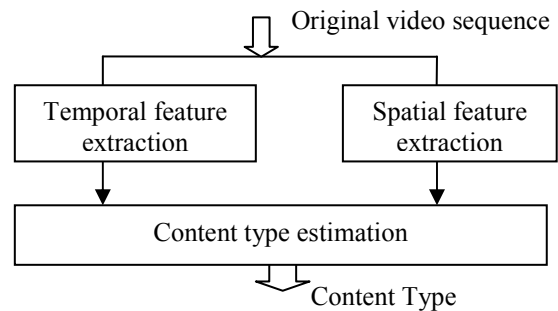## III. CONTENT CLASSIFICATION BASED ON CLUSTER ANALYSIS

Figure. 3 Content classification design

Video content is classified using a well known multivariate statistical analysis called cluster analysis [6]. This technique is used as it groups samples that have various characteristics into similar groups. Cluster

analysis is carried out on the nine video sequences given in Table II based on the temporal and spatial feature extraction. The design of our content classification method is given in Fig. 3.

A. Temporal feature extraction

The movement in a video clip given by the SAD value (Sum of Absolute Difference). The SAD values are computed as the pixel wise sum of the absolute differences between the two frames being compared and is given by (2):

$$SAD_{n,m} = \sum_{i=1}^{N} \sum_{j=1}^{M} |B_n(i,j) - B_m(i,j)| \qquad (2)$$

Where B_n and B_m are the two frames of size $N \times M$, and i and j denote pixel coordinates.

B. Spatial feature extraction

The spatial features extracted were the edge blocks, blurriness and the brightness between current and previous frames. Brightness (B_r) is calculated as the modulus of difference between average brightness values of previous and current frames.

$$Br_n = \sum_{i=1}^{N} \sum_{j=1}^{M} |Br_{av(n)}(i,j) - Br_{av(n-1)}(i,j)| \qquad (3)$$

Where $Br_{av(n)}$ is the average brightness of n-th frame of size $N \times M$, and i and j denote pixel coordinates.

C. Cluster analysis

For our data we calculate Euclidean distances in 10-dimensional space between the SAD, edge block, brightness and blurriness measurements and conduct hierarchical cluster analysis. Fig. 4 shows the obtained dendrogram (tree diagram) where the video sequences are grouped together on the basis of their mutual distances (nearest Euclid distance).
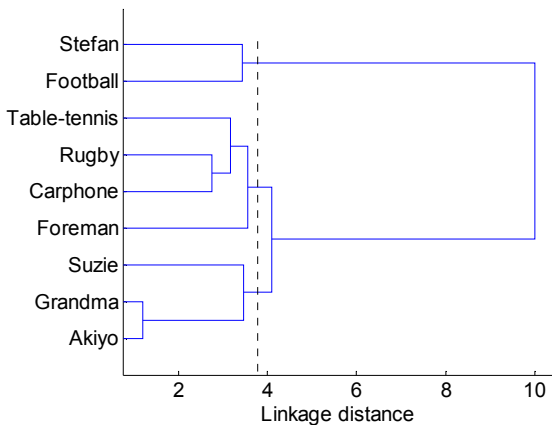

Figure. 4 Tree diagram based on cluster analysis

In this paper, we divided the test sequences at 38% from the maximum of Euclid distance into three groups as the data contains a clear 'structure' in terms of clusters that are similar to each other at that point (see the dotted line on Fig. 4). Group 1 (sequences Grandma, Suzie and Akiyo) are classified as 'Slight Movement', Group 2

(sequences Carphone, Foreman, Table-tennis and Rugby) are classified as 'Gentle Walking' and Group3 (sequences Stefan and Football) are classified as 'Rapid Movement'. See Table III. Future work will concentrate on reducing the maximum Euclid distance and hence increase the content groups.

TABLE III
VIDEO CONTENT CLASSIFICATION

Content type	Content features	Video Clip
Slight Movement	A newscaster sitting in front of the screen reading news only by moving her lips and eyes	Grandma
		Suzie
		Akiyo
Gentle Walking	with a contiguous change of scene at the end – 'typical for video call'	Table-tennis
		Carphone
		Rugby
		Foreman
Rapid Movement	A professional wide angle sequence where the entire sequence is moving uniformly	Football
		Stefan

We found that the 'news' type of video clips were clustered in one group, however, the sports clips were put in two different categories i.e. clips of 'stefan' and 'football' were clustered together, whereas, 'rugby' and table-tennis' were clustered along with 'foreman' and 'carphone' which are both wide angle clips in which both the content and background are moving.

To further verify the content classification from the tree diagram obtained (Fig. 4) we carried out K-means cluster analysis in which the data (video clips) is partitioned into k mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. K-means computes cluster centroids differently for each measured distance, to minimize the sum with respect to the specified measure. We specified k to be three to define three distinct clusters. In Fig. 5 K-means cluster analysis is used to partition the data for the nine content types. The result set of three clusters are as compact and well-separated as possible giving very different means for each cluster. Cluster 1 in Fig. 5 is very compact for three video clips instead of four. The fourth clip of table-tennis can be within its own cluster and will be looked in much detail in future work. All results were obtained using MATLAB™ 2008 functions.
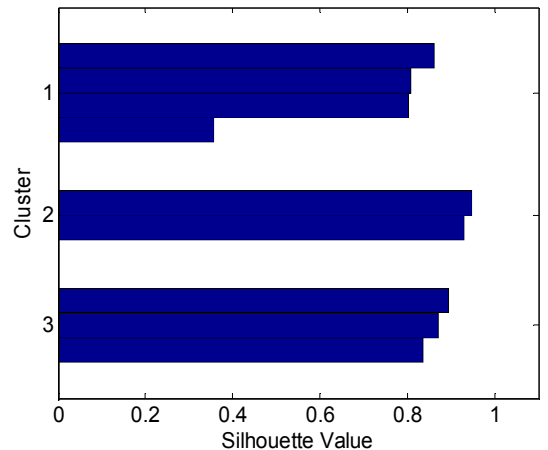

Figure. 5 K-means Cluster analysis

The cophenetic correlation coefficient, c, is used to measure the distortion of classification of data given by cluster analysis. It indicates how readily the data fits into the structure suggested by the classification. The value of c for our classification was 88.1% indicating a good classification result. The magnitude of c should be very close to 100% for a high-quality solution.

IV. 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 scores is computed as a function of the values of all four parameters.

A. Mos vs Send Bitrate vs Packet Error Rate
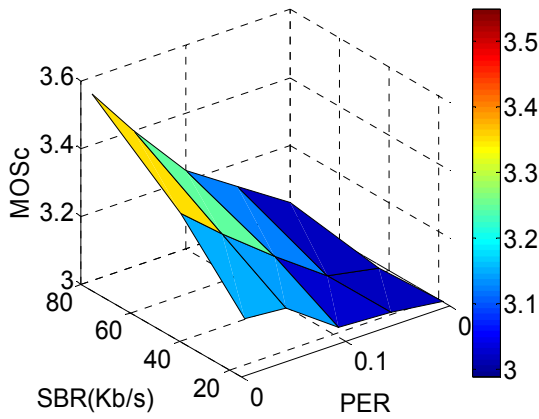
Figure. 6 MOS vs SBR vs PER for 'Slight movement'

Fig. 6 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.
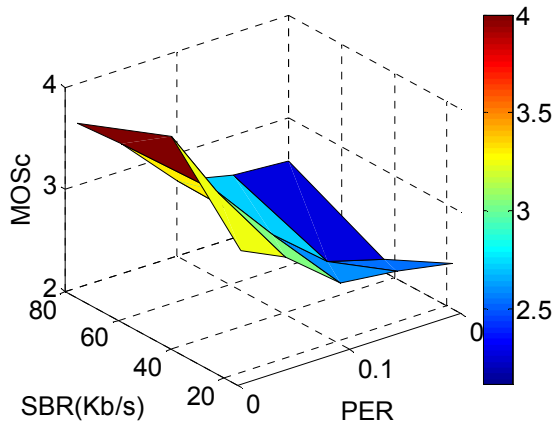

Figure. 7 MOS vs SBR vs PER for 'Gentle walking'

Fig. 7 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 (MOS > 3.5), however, the quality fades rapidly with increasing packet loss.
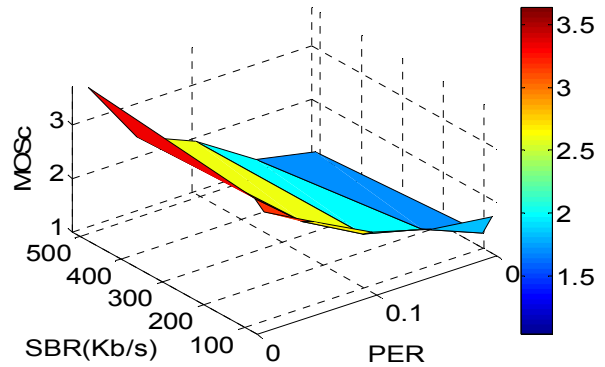

Figure. 8 MOS vs SBR vs PER for 'Rapid movement'

Fig. 8 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.

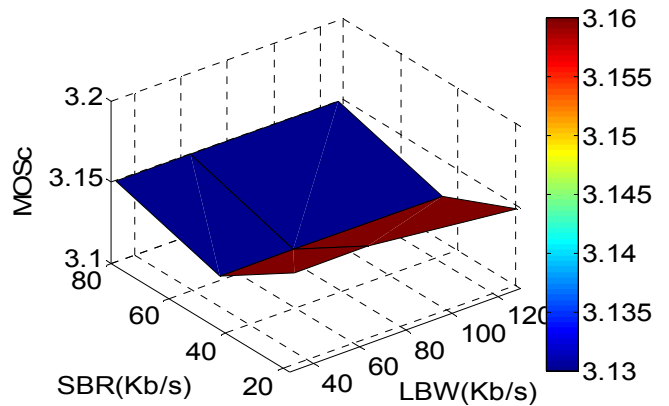*B. MOS vs Send Bitrate vs Link Bandwdth*

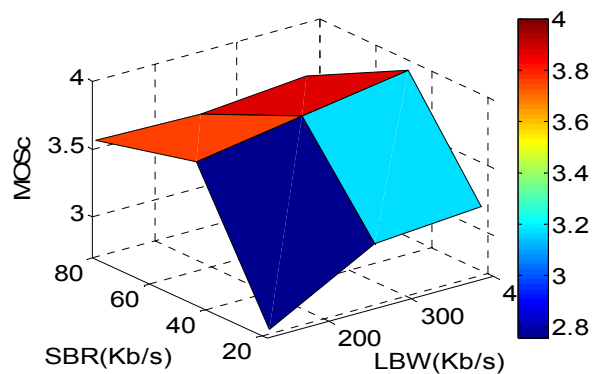Figure. 9a MOS vs SBR vs LBW for 'Slight movement'
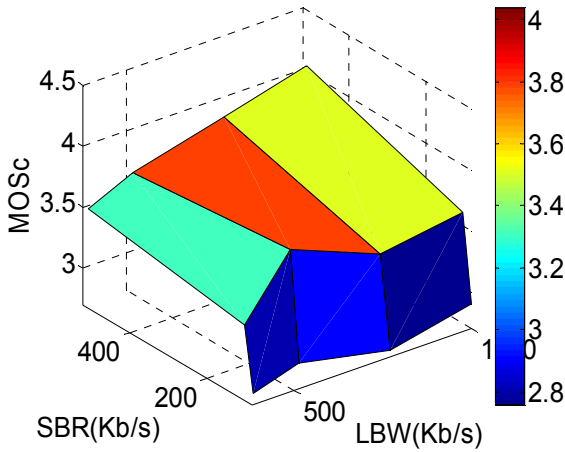

Figure. 9b MOS vs SBR vs LBW for 'Gentle walking'

Figure. 9c MOS vs SBR vs LBW for 'Rapid movement'

In Figs 9a, 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. If the send bitrate is greater than the link bandwidth then video quality worsens due to network congestion.
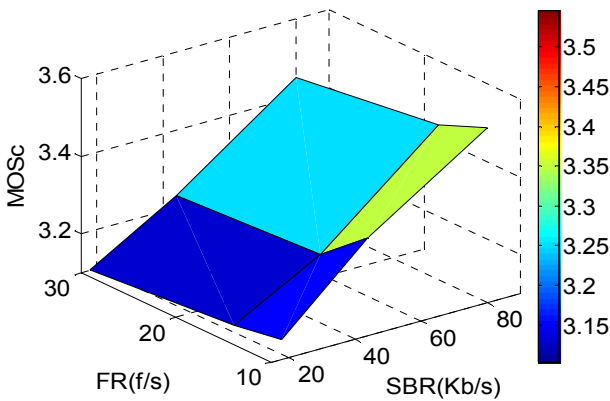
C. MOS vs Frame Rate vs Send Bitrate

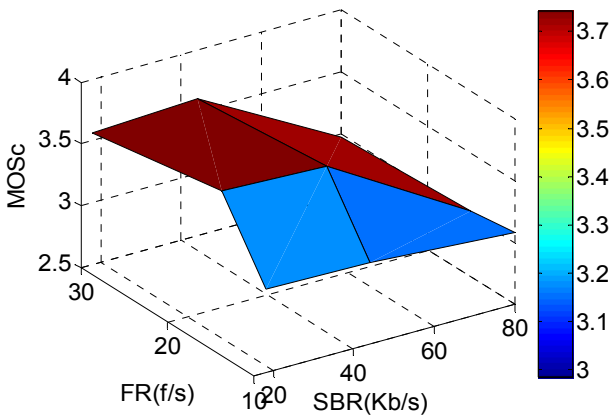Figure. 10a MOS vs FR vs SBR for 'Slight Movement'

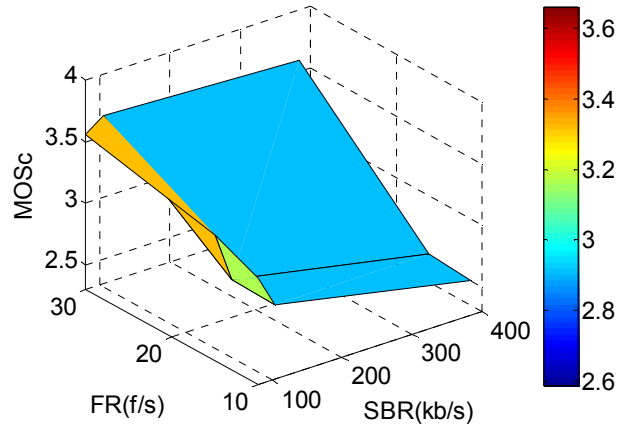Figure. 10b MOS vs FR vs SBR for 'Gentle Walking'

Figure. 10c MOS vs FR vs SBR for 'Rapid Movement'

From Figs. 10a,b&c we found that frame rate is not as significant as send bitrate. Improvement in video quality is only achived upto frame rates of 15fps. This confirms that for low send bitrate videos low frame rates give better quality (e.g. frame rate $\leq$ 10fps). However, for higher send bitrate higher frame rate will not reduce video quality.

V. 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. 11. For the tests we selected three different video sequences from the three content types classified in section III (See Table III) of qcif resolution (176x144) and encoded in MPEG4 format with an open source ffmpeg [27] encoder/decoder. The three video clips are send over WLAN (IEEE 802.11 standard) using NS2 [25] and Evalvid [26] as shown in Section II, sub-section A. The application level parameters considered are Content Type (CT), Frame Rate (FR) and Send Bit Rate (SBR). The network parameters are Packet Error Rate (PER) and Link Bandwidth (LBW).
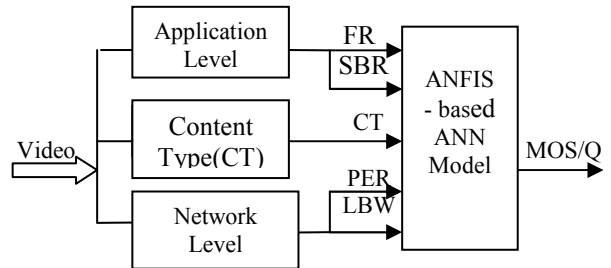

Figure. 11 Functional block of proposed ANFIS-based model

A. ANFIS architecture

The corresponding equivalent ANFIS architecture [24] is shown in Fig. 12.

The entire system architecture is made of five layers, consisting of - a fuzzy layer, a product layer, a normalized layer, a defuzzy layer and a total output layer.

Inputs x and y are frame rate, send bitrate, link bandwidth and packet error rate. Output f is the MOS score and Q value.
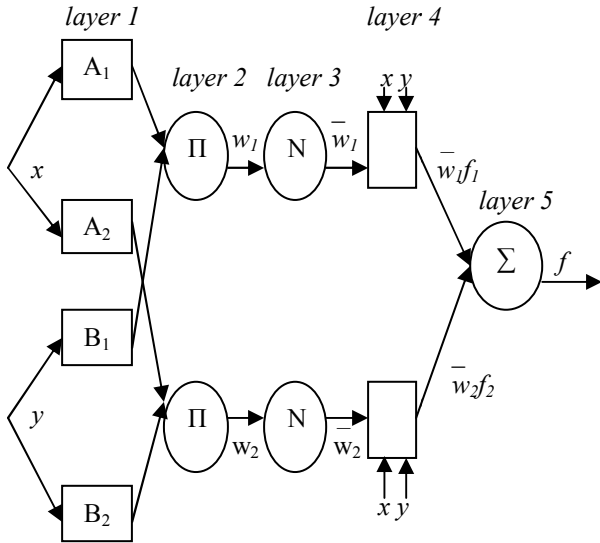

Figure. 12 ANFIS architecture [24]

B. Training and validating 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. From the three content types classified in section III, we chose one video clip from each content type (see Table III) for training purposes and a different video clip from the same content type for validation purposes. The three ANFIS-based ANN models were trained with the three distinct content types of 'Akiyo', 'Foreman' and 'Stefan' (see Table III) from the three content types of 'slight movement', 'gentle walking' and 'rapid movement for training and validated by three different content types of 'Suzie', 'Carphone' and Football' in the corresponding content categories. Snapshost of the three video clips in the three content types used for validation are given in Fig. 13 below.

Figure. 13 Snapshots of three content types

The data selected for validation was one third that of testing with different parameter values to that given in Table II. In total there were 135 encoded test sequences for the first two content categories and 180 encoded test

sequences for the third content category.

To summarize, the ANFIS-based ANN gives good prediction accuracy for both MOS and Q prediction. We feel that the choice of parameters are crucial in achieving good prediction accuracy. Parameters such as Link Bandwidth in real systems are measured in terms of packet loss and delay. However, in a simulation system it was interesting to capture the impact of Link Bandwidth. Also, in the application level the Send Bitrate has a bigger impact than Frame rate. Finally, to predict video quality content type is very important. Contents with less movement require low Send Bitrate and Link Bandwidth compared to that of higher movement. In future, we are looking at one model for all content types.

VI. 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 [28]. For the three content types we obtained results in terms of the MOS score and decodable frame rate Q [3].
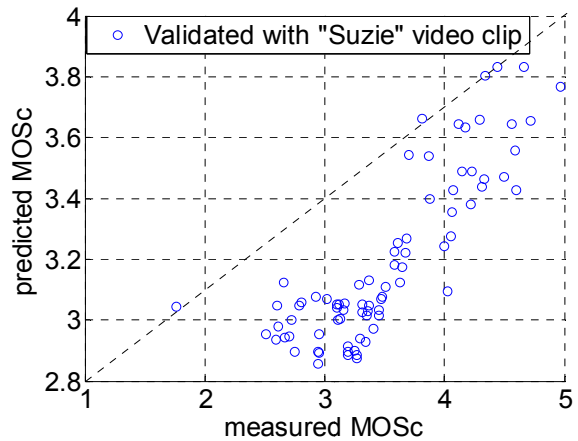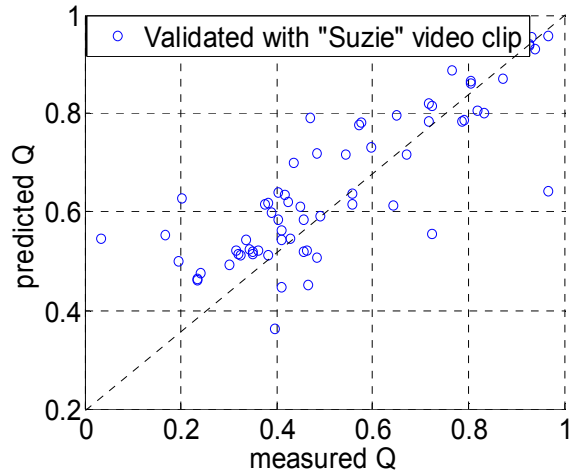

Figure. 14a ANN mapping of MOS for 'Slight movement'

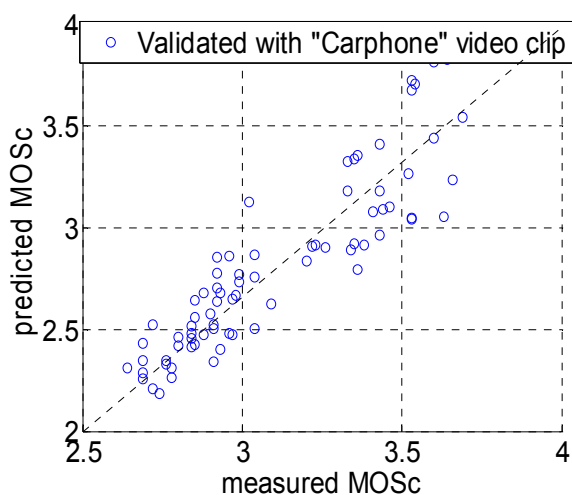Figure. 14b ANN mapping of predicted Q for 'Slight movement'

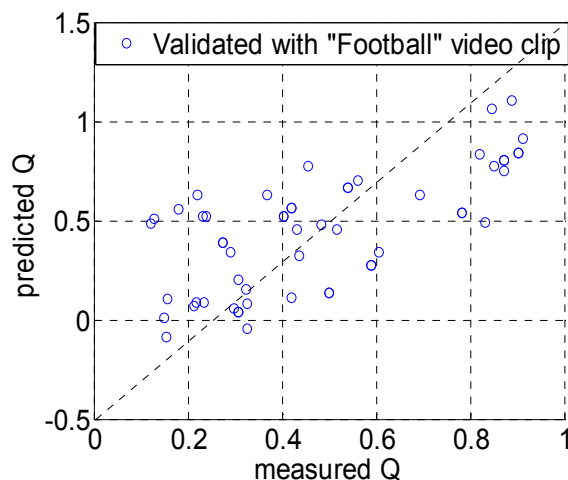Figure.15a ANN mapping of predicted MOSc 'Gentle walking
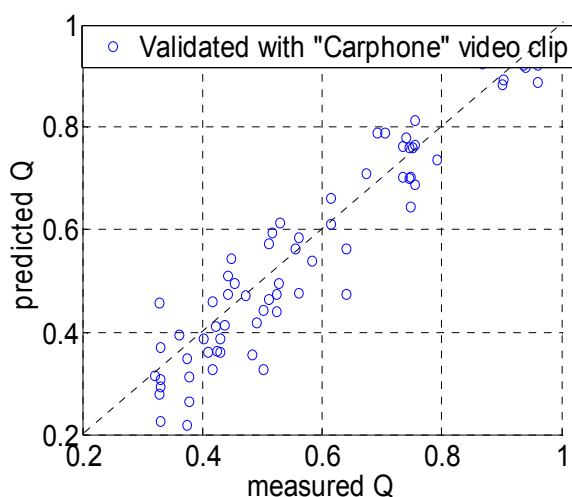

Figure.15b ANN mapping of predicted Q for 'Gentle walking

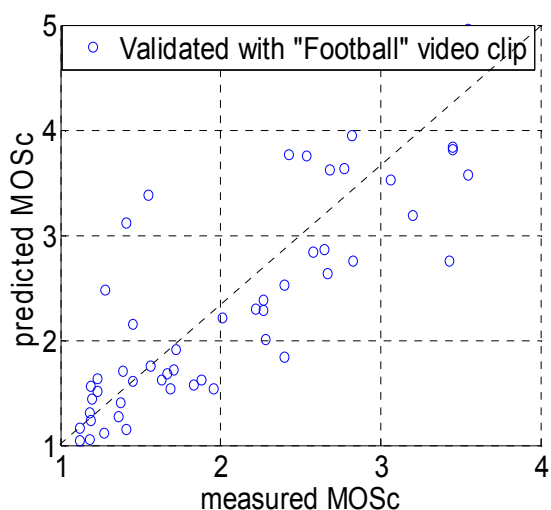Figure.16a ANN mapping of predicted MOSc for 'Rapid movement'

Figure.16b ANN mapping of predicted 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. 14a&b, 15a&b and 16a&b). However, more realistically the relationship between the measured MOS/Q (x) and the predicted MOS/Q (y) is represented as $y = a_1 x + a_2$.

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 IV:

TABLE IV
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	1.178/0.8268	-0.05656/0.0528

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 V below.

TABLE V.
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	0.754/0.6911	0.5659/0.2181

We achieved better correlation for 'gentle walking' compared to 'rapid movement' and 'slight movement'. Additionally, 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.

VII. REGRESSION-BASED VIDEO QUALITY PREDICTION

This section describes the regression-based video quality prediction model based on the application level parameters of Send Bitrate and Frame Rate and network level parameter of Packet Error Rate.

A. PCA Analysis

Principal Component Analysis (PCA) [20] reduces the dimensionality of the data while retaining as much information as possible. For this reason, PCA was carried out to determine the relationship between MOS and the objective video parameters of send bitrate, frame rate, packet error rate and link bandwidth. PCA involves calculating eigenvalues and their corresponding eigenvectors of the covariance or correlation matrix. Covariance matrix is used where the same data has the same set of variables and correlation matrix is used in the case where data has a different set of variables. In this paper, we used a covariance matrix because of the same data set. The PCA was carried out to verify the applicability of the objective parameters of SBR, FR PER and LBW for metric design. The PCA was performed for the three content types of SM, GW and RM separately. The variance of the data for the three content types is given in Table IV.

TABLE VI
VARIANCE OF THE FIRST TWO COMPONENTS FOR ALL CONTENT TYPES

Sequence	Var. of PC1(%)	Var. of PC2(%)
Slight Movement	58	33
Gentle Walking	63	31
Rapid Movement	74	20

The first two components account for more than 90% of the variance and hence are sufficient for the modeling of the data. Therefore, the PCA suggests which parameters in our data set are important and which ones of little consequence. The parameters of SBR and FR had the largest impact followed by PER. LBW was not considered for metric design as it had the least impact. The PCA results are shown in Fig. 17.
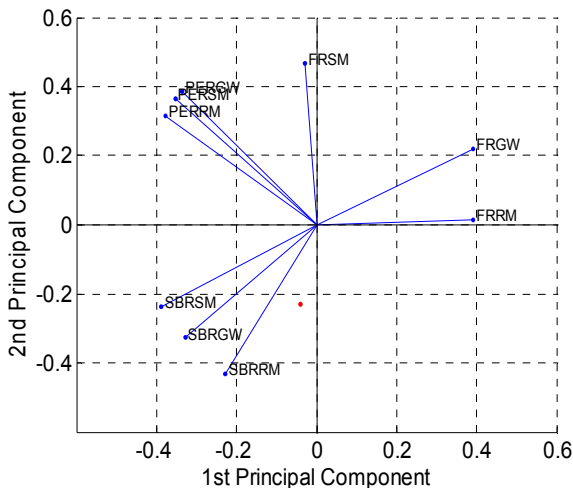

Figure. 17 PCA results for all content types

The PCA results from Fig. 17 show the influence of the chosen parameters (SBR, FR and PER) on our data set for the three content types of SM, GW and RM. In Fig. 17 the horizontal axis represents the first principal component (PC1) and the vertical axis represents the second principal component (PC2). Each of the objective parameters (e.g. FRGW, etc) are represented by a vector.

B. Proposed Model

The final step in this paper is to predict video quality based on the objective parameters of send bitrate, frame rate and packet error rate (see Table I) for the three content types of 'Slight movement', 'Gentle walking' and 'Rapid movement'. From the three content types classified in the previous section, we chose one video clip from each content type (see Table III) for testing purposes and a different video clip from the same content type for validation purposes. We chose video clips of 'Akiyo', 'Foreman' and 'Stefan' from the three content types. For validation purposes we chose 'Suzie', 'Carphone' and 'Football' as previously.
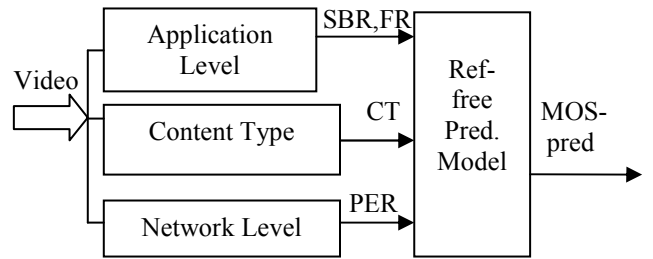

Figure. 18 Video quality metric prediction design

Video quality prediction is carried out for three distinct content types from both network and application parameters for video streaming over wireless networks application as shown in Fig. 18. The application level parameters considered are Content Type (CT), Frame Rate (FR) and Send BitRate (SBR). The network parameters are Packet Error Rate (PER).

C. MOS Prediction

The proposed low complexity metric is based on three objective parameters (SBR, FR and PER) for each content type as given by (4):

$$MOS = f(SBR, FR, \text{Content type}, PER) \qquad (4)$$

We propose one common model for all content types given by (4). The prediction model for video quality evaluation in terms of the Mean Opinion Score (MOS_v) is given by a rational model with a logarithmic function (See (5)).

$$MOS_v = \frac{a_1 + a_2 FR + a_3 \ln(SBR)}{1 + a_4 PER + a_5 (PER)^2} \qquad (5)$$

The metric coefficients were obtained by a linear regression of the proposed model with our training set (MOS values obtained by objective evaluation given in

Table I). The coefficients for all three content types are given in Table VII.

TABLE VII
COEFFICIENTS OF METRIC MODELS FOR ALL CONTENT TYPES

Coeff	SM	GW	RM
a1	4.5796	3.4757	3.0946
a2	-0.0065	0.0022	-0.0065
a3	0.0573	0.0407	0.1464
a4	2.2073	2.4984	10.0437
a5	7.1773	-3.7433	0.6865

The proposed metric has different coefficient values for the three different content types because spatial and temporal sequence characteristics of the sequences are significantly different. The model's prediction performance is given in terms of the correlation coefficient R^2 (indicates the goodness of fit) and the RMSE (Root Mean Squared Error) and is summarized in Table VIII.

TABLE VIII
METRIC PREDICTION PERFORMANCE BY CORRELATION COEFFICIENT AND RMSE

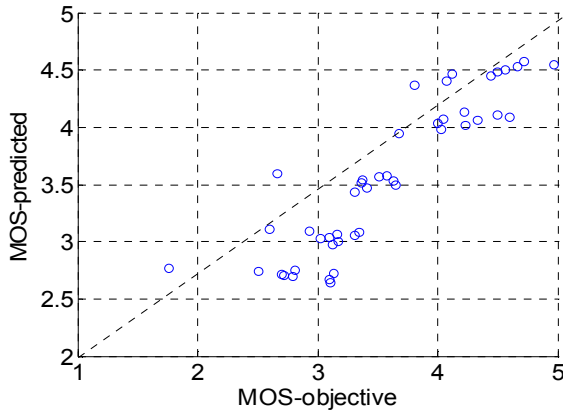| Content type	SM	GW	RM
Corr coef	79.9%	93.36%	91.7
RMSE	0.2919	0.08146	0.2332

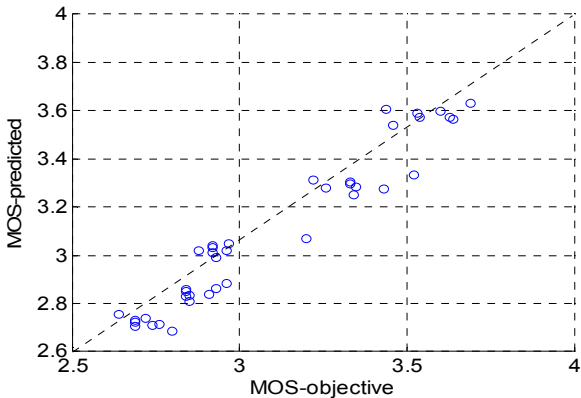Figure. 19a Predicted vs. objective MOS results for 'SM'

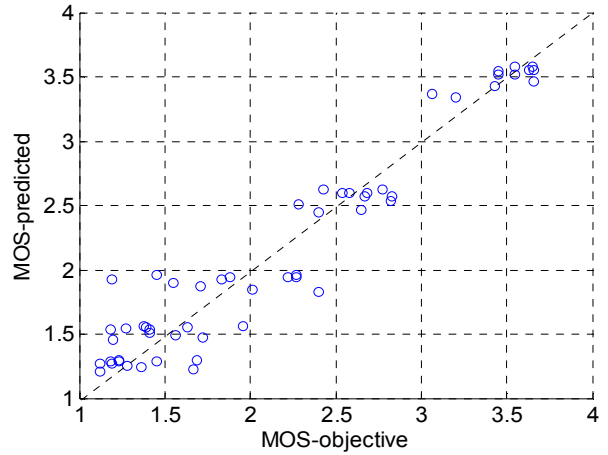Figure. 19b Predicted vs. objective MOS results for 'GW'

Figure. 19c Predicted vs. objective MOS results for 'RM'

The performance of the video quality prediction obtained by our metric compared to the video quality data obtained objectively using NS2[25] for the three content types of 'slight movement', 'gentle walking' and 'rapid movement' are shown in Fig. 19a,b&c. We achieved slightly better correlation for 'gentle walking' compared to the other two content types of 'slight movement' and 'rapid movement'. 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. Whereas, for 'slight movement' the video quality was still acceptable (MOS>3.5) for packet losses of up to 20%.

D. Comparison of the proposed models

The two models proposed in this paper are reference-free. The regression based model [23] has performed better compared to that of the ANFIS based [13] in terms of the prediction performance (See Table V & VIII). We feel that the video contents are fuzzy in nature and are looking at extending the three models to one for all content types. Furthermore, compared to a recent work that has estimated video quality based on ANNs is presented in [17]. 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. Furthermore, the video sequences we chose for validation are completely different to those for testing which confirm the right choice of objective parameters and hence, a reliable tool for video quality prediction.

VIII. CONCLUSIONS

In this paper, we have proposed two content-based perceptual quality reference-free metric for the most frequent content types for wireless MPEG4 video streaming applications. We have further 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 used cluster analysis to classify the most frequently used content into three specific content types of 'slow movement', 'gentle walking', and 'rapid movement' based on the combination of temporal and spatial feature extraction. The purpose of content clustering was to make new groups of video content with similar characteristics. The grouping can be used to apply priority control to content delivery, and hence optimize bandwidth allocation for specific content in content delivery networks. The automatic content classification enable video quality prediction within a content type.

We developed (1) ANFIS-based learning model and (2) Regression-based model for the three content types 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 as expected compared to application level parameters such as frame rate and video send bitrate. In real networks it is very difficult to measure link bandwidth. The effect of link bandwidth is usually measured in terms of delay and packet loss. However, in simulated scenario we wanted to directly measure link bandwidth and its impact on video quality. Also, if the video stream is encoded at a send bitrate greater than the link bandwidth then video quality is degraded due to network congestion.

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.

The regression-based model gave better prediction performance compared to that of ANFIS. Both the models were validated with video clips within the same content type with good prediction accuracy.

Our future work will focus on feedback mechanisms that could dynamically adapt the encoding parameters according to the content dynamics that satisfy a specific video quality level at a pre-encoding stage in the most efficient way taking into account network conditions to achieve optimum end-to-end video quality.

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. du Toit, A. Steyn and R. Stumpf, Cluster analysis", Handbook of graphical exploratory data analysis, ed. S.H.C. du Toit, pp.73-104, Springer-Verlag, New York, 1986.
[7] 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.
[8] 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.
[9] K. Yamagishi and T. Hayashi, "Opinion model using psychological factors for interactive multimodal services", IEICE Trans. Communication, Vol.E89-B, No. 2, Feb. 2006.
[10] 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.
[11] H. Koumaras, A. Kourtis, C. Lin and C. Shieh, "A theoretical framework for end-to-end video quality prediction of MPEG-based sequences", *Third international conference on Networking and Services*, 19-25 June 2007.
[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] A. Khan, L. Sun and E. Ifeachor, "An ANFIS-based hybrid video quality prediction model for video streaming over wireless networks", IEEE Computer Society Proceedings, NGMAST Conference, Cardiff, UK, 16-19 Sept., 2008.
[14] M. Ries, O. Nemethova and M. Rupp, "Video quality estimation for mobile H.264/AVC video streaming", *Journal of Communications*, Vol. 3, No.1, Jan. 2008, pp. 41-50.
[15] L. Yu-xin, K. Ragip, B. Udit, "Video classification for video quality prediction", *Journal of Zhejiang University Science A,* 2006 7(5), pp 919-926.
[16] 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.
[17] 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.
[18] Y. Kato, A. Honda and K. Hakozaki, "An analysis of relationship between video contents and subjective video quality for internet broadcasting", *Proc. Of the 19th Int. Conf. On Advanced Information Networking and Applications (AINA),* 2005.
[19] Y. Suda, K. Yamori and Y. Tanaka, "Content clustering based on users' subjective evaluation", *6th Asia-Pacific symposium on Information and Telecommunication Technologies ASPITT*, 2005, Volume , Issue , 09-10 Nov. 2005 pp. 177 – 182.
[20] W. J. Krzanowski, "Principles of multivariate analysis", Clarendon press, Oxford, 1988.
[21] J. Wei, Video content classification based on 3-d Eigen analysis, *"IEEE transactions on image processing"*, Vol. 14, No.5, May 2005.

[22] G. Gao, J, Jiang, J. Liang, S. Yang and Y. Qin, PCA-based approach for video scene change detection on compressed video, *Electronics letters*, Vol. 42, No.24, 23rd Nov. 2006.

[23] A. Khan, L. Sun and E. Ifeachor, "Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks", submitted to the IEEE ICC Conference, Dresden, Germany, 14-18 June 2009.

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

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

[26] 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.

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

[28] J. Mitchell and W. Pennebaker, "MPEG Video: Compression Standard, Chapman and Hall, 1996, ISBN 0412087715

[29] 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/