Impact of Video Content on Video Quality for Video 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.
e-mail: (asiya.khan, l.sun, e.ifeachor)@plymouth.ac.uk

***Abstract*—Video streaming is a promising multimedia application and is gaining popularity over wireless/mobile communications. The quality of the video depends heavily on the type of content. The aim of the paper is threefold. First, video sequences are classified into groups representing different content types using cluster analysis based on the spatial (edges) and temporal (movement) feature extraction. Second, we conducted experiments to investigate the impact of packet loss on video contents and hence find the threshold in terms of upper, medium and lower quality boundary at which users' perception of service quality is acceptable. Finally, to identify the minimum send bitrate to meet Quality of Serive (QoS) requirements (e.g. to reach communication quality with Mean Opinion Score (MOS) greater than 3.5) for the different content types over wireless networks. We tested 12 different video clips reflecting different content types. We chose Peak-Signal-to-Noise-Ratio (PSNR) and decodable frame rate (Q) as end-to-end video quality metrics and MPEG4 as the video codec. The work should help optimizing bandwidth allocation for specific content in content delivery networks.**

Keywords-MPEG4; 802.11b; NS-2; PER; Video quality evaluation

I. INTRODUCTION

Multimedia services are becoming commonplace across different transmission platforms such as Wi-Max, 802.11 standards, 3G mobile, etc. The current trends in the development and convergence of wireless internet IEEE802.11 applications and mobile systems are seen as the next step in mobile/wireless broadband evolution. Users' demand of the quality of streaming service is very much content dependent. Streaming video quality is dependent on the intrinsic attribute of the content. For example, users request high video quality for fast moving contents like sports, movies, etc. compared to slow moving like news broadcasts, etc. where to understand the content is of more importance. The future internet architecture will need to support various applications with different QoS (Quality of service) requirements [1]. QoS of multimedia communication is affected both by the network level and application level parameters [2]. In the application level QoS is driven by factors such as resolution, frame rate, colour, video codec type, audio codec type, etc. The network level introduces impairments such as delay, cumulative inter-frame jitter, burstiness, latency, packet loss, etc.

Recent work has focused on the wireless network (IEEE 802.11) performance of multimedia applications [3,4,5]. In [6,7,8] the authors have looked at the impact of transmission errors and packet loss on video quality. In [9] 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, but their work is limited to videophone. Similarly, in [10] authors have taken into consideration a combination of content and network adaptation techniques to propose a fuzzy-based video transmission approach. In [11] the authors have proposed content based perceptual quality metrics for different content types, whereas, in [12] video content is divided into several groups using cluster analysis [13]. However, very little work has been done on the impact of different types of content on end-to-end video quality e.g. from slow moving (head and shoulders) to fast moving (sports) for streaming video applications under similar network conditions considering both network level and application level parameters. We have looked at the two main research questions in the network level and application level as:

(1) What is the acceptable packet error rate for all content types for streaming MPEG4 video and hence, find the threshold in terms of upper, medium and lower quality boundary at which the users' perception of quality is acceptable?

(2) What is the minimum send bitrate for all content types to meet communication quality for acceptable QoS (PSNR >27 dB) as it translates to a MOS of greater than 3.5 [14]?

To address these two questions, we first classified the video contents based on the spatial and temporal feature extraction into similar groups using cluster analysis [13]. We then carried out experiments to investigate the impact of Packet Error Rate (PER) and hence, find the threshold in terms of upper, medium and lower quality boundary at which the users' perception of quality is acceptable and identified the minimum acceptable Send Bitrate (SBR) for the content types. We chose Peak-Signal-to-Noise-Ratio (PSNR) and decodable frame rate (Q) [8] as end-to-end video quality metrics and MPEG4 as the video codec. In the presence of packet loss video quality becomes highly time-variant [15,16]. One of the significant problems that video streaming face is the unpredictable nature of the internet in terms of the send bitrate, and packet loss. We further investigated the impact of video quality over the entire duration of the sequence and hence observe the type of errors using objective video quality metrics such as PSNR. These could help in resource optimization and the development of QoS control mechanisms in the future. Our focus ranges from low resolution and low send bitrate video streaming for 3G applications to higher video send bitrate for WLAN applications depending on type of content and network conditions The proposed test bed is based on simulated network scenarios using a network simulator (NS2) [17] 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 organized as follows. Section 2 classifies the contents. In section 3 the experimental set-up is given. Section 4 presents the experiments conducted and analysis of results. Conclusions and areas of future work are given in section 5.

II. CONTENT CLASSIFICATION

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 test sequences represent typical content offered by network providers. The content classification was done based on the temporal and spatial feature extraction using well known tool called cluster analysis [13].

The design of our content classification method is given in Fig. 1

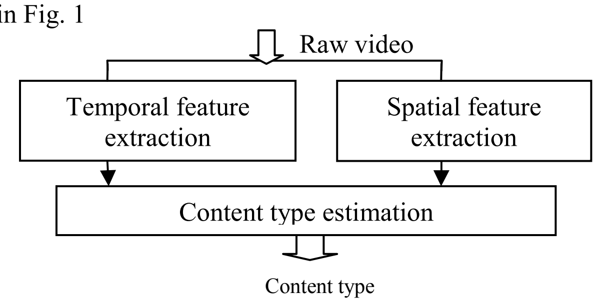

Figure 1. Content classification design

A. TemporalFfeature 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:

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

Where B_n and B_m are the two frames of size N X 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 is calculated as the modulus of difference between average brightness values of previous and current frames.

C. Cluster Analysis

For our data we calculate Euclidean distances in 13-dimensional space between the SAD, edge block, brightness and blurriness measurements and conduct hierarchical cluster analysis. Fig. 2 shows the obtained dendrogram (tree diagram) where the video sequences are grouped together on the basis of their mutual distances (nearest Euclid distance).

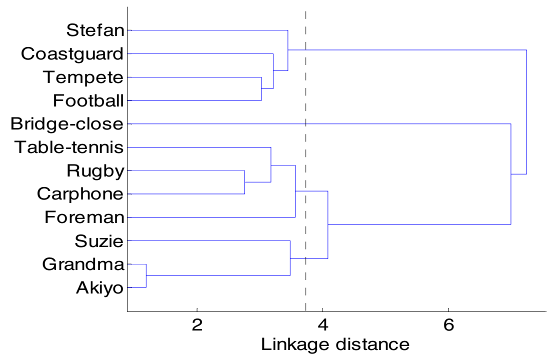

Figure 2. Tree diagram based on cluster analysis

According to Sturge's rule (k = 1 + 3.3logN), which for our data will be 5 groups. However because of the problems identified with this rule [18] we split the data (test sequences) at 38% from the maximum Euclid distance into three groups. (see the dotted line on Fig. 2) as the data contains a clear 'structure' in terms of clusters that are similar to each other at that point. 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'. 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. Also 'bridge-close' can be classified on its own creating four groups instead of three. But as it is closely linked with the first group of SM we decided to put it in SM. In future, we will create more groups and compare it to our existing classification.

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 79.6% indicating a good classification result. The magnitude of c should be very close to 100% for a high-quality solution.

The three content types are defined for the most frequent contents for mobile video streaming as follows:

1. Content type 1 – Slight Movement (SM): includes sequences with a small moving region of interest (face) on a static background. See Fig. 3.

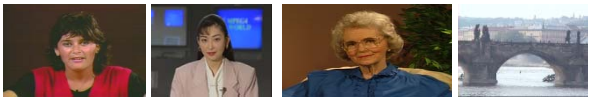

Figure 3. Snapshots of typical 'SM' content

2. Content type 2 – Gentle Walking (GW): includes sequences with a contiguous scene change at the end. They are typical of a video call scenario. See Fig. 4.

Figure 4. Snapshots of typical 'GW' content

3. Content type 3 – Rapid Movement (RM): includes a professional wide angled sequence where the entire picture is moving uniformly e.g sports type. See Fig. 5.

Figure 5. Snapshots of typical 'RM' content

III. Experimental Set-Up

For the tests we selected twelve different video sequences of qcif resolution (176x144) and encoded in MPEG4 format with an open source ffmpeg [19] encoder/decoder with a Group of Pictures (GOP) pattern of IBBPBBPBB. The frame rate was fixed at 10fps. 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.

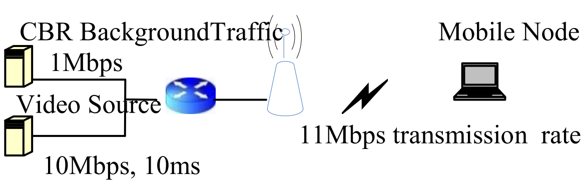

Figure 6. Simulation setup

The experimental set up is given in Fig 6. 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 maximum transmission packet size is 1024 bytes. The video packets are delivered with the random uniform error model. The CBR rate is fixed to 1Mbps to give a more realistic scenario. 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.

IV. Experiment And Analysis Of Results

We considered both network level and application level factors and used performance metrics to evaluate video quality affected by both factors. The performance metrics used were average PSNR and decodable frame rate [8]. 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.

$$PSNR(s,d)]_{db} = 20 \log \frac{Vpeak}{MSE(s,d)} \qquad (2)$$

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

Decodable frame rate (Q) [8] 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.

A. Experiment 1 – Average PSNR Vs PER

Video quality is measured by taking the average PSNR over all the decoded frames across network PER from 0.01 to 0.2 (20%). All videos were encoded at a send bitrate of 256kb/s. This experiment is conducted to answer the first research question: What is the acceptable PER for maintaining the minimum QoS requirement of 27dB for the different content types ?

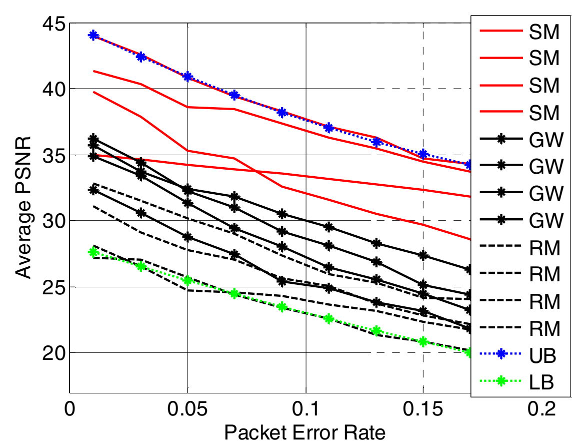

Figure 7. Packet Error Rate vs Average PSNR

Fig. 7 show the average PSNR vs the PER for all 12 video clips. It shows that the average PSNR is better for slight movement compared to gentle walking which in turn is better than rapid movement which shows the dependence on content type. From our results, we found that for slight movement the video quality stays above the threshold of

PSNR > 27dB (MOS >3.5) for upto 20% packet loss. However, for gentle walking and rapid movement that value drops to 10% and 6% respectively.

Further, we derive an upper, medium and lower boundary for PSNR as a function of PER for the three content types of SM, GW and RM and hence know the threshold for acceptable quality in terms of the PSNR for the three content types with 95% confidence level and goodness of fit of 99.71% and Root Mean squared Error (RMSE) of 0.3235 is given by equations (3), (4) and (5):

SM: $PSNR= 122.3(PER)^2-88.36(PER)+42.6$; $PER \leq 20\%$ (3)

GW: $PSNR= 64.9(PER)^2-73.75(PER)+34.43$; $PER \leq 10\%$ (4)

RM: $PSNR= 76.8(PER)^2-68.87(PER)+31.43$; $PER \leq 6\%$ (5)

B. Experiment 2 – Q Vs PER

The experimental set up is the same as in A but we measured Q value [8] instead of PSNR vs PER and addressed the above research question in terms of Q [8] instead of PSNR.

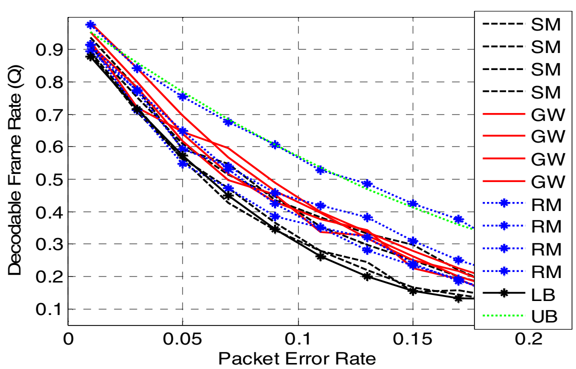

Figure 8. PER vs Q for all content types

Fig. 8 shows the decodable frame rate (Q) of all 12 contents and shows that Q is higher when the PSNR is higher for all the video clips. In comparison to Fig 3 the decodable frame rate does not directly compare to the PSNR. However, from our results we found higher values for the average PSNR for 'slight movement' and it did not correspond to a higher value of Q. This is because the Q value is derived from the number of decodable frames over the total number of frames sent by a video source [8] i.e. it is sensitive to the number of frames and packets lost. Therefore, as the content becomes more complex we would expect the video quality to degrade more for less I-frames lost compared to that of simpler contents. Hence, we conclude that for slight movement 20%, for gentle walking 10% and for rapid movement 6% packet loss is acceptable.

Further, we derive an upper, medium and lower boundary for Q value as a function of PER for the three content types of SM, GW and RM and hence know the threshold for acceptable quality in terms of the Q value for the three content types with 95% confidence level and goodness of fit of 99.71% and RMSE of 0.0117 is given by the equations (6), (7) and (8):

SM: $Q = 19.89(PER)^2 -8.03(PER) + 0.967$; $PER \leq 20\%$ (6)

GW: $Q= 18.09(PER)^2-7.88(PER)+1.02$; $PER \leq 10\%$ (7)

RM: $Q= 13.84(PER)^2 - 6.5(PER) + 0.975$; $PER \leq 6\%$ (8)

C. Experiment 3 – Average PSNR Vs PER Vs SBR

The experimental set up is the same as in 3.1 but we changed the video send bitrate to achieve the minimum send bitrate for QoS requirements and to address the research question: What is the minimum SBR for the different video content types with time variant quality acceptable for communication quality (>27dB)?

The send bitrates ranged from 18kb/s to 384kb/s. We chose one video clip from each category. We suggest a minimum send bitrate for all three categories that achieve an average PSNR values of higher than 27dB for the video content types as it translates to a MOS of greater than 3.5 [14] which is an acceptable score for the telecommunication industry.

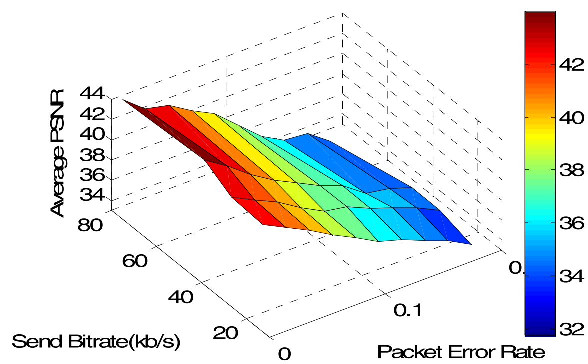

Figure 9. Average PSNR Vs PER and SBR for 'SM'

Fig. 9 shows the average PSNR over the video send bitrates of 18kb/s, 32kb/s, 44kb/s and 80kb/s. We found that for slow movement low bitrate of 18kb/s is acceptable as it yields an average PSNR of 30dB without any packet loss. As the send bit rate is increased to 80kb/s, average PSNR is greater than 40dB indicating that the bandwidth should be re-allocated to optimize it.

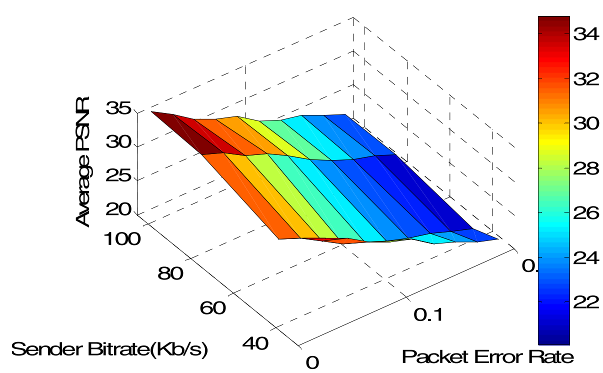

Figure 10. Average PSNR Vs PER and SBR for 'GW'

In Fig. 10 we chose send bitrates of 32kb/s, 44kb/s, 80kb/s and 104kb/s, as bitrates less than 18kb/s will give poor video quality rendering them meaningless. We suggest a send bitrate of 32kb/s for gentle walking as it gives an average PSNR value of approx. 29dB. However, with higher packet loss the quality falls below the acceptable level.

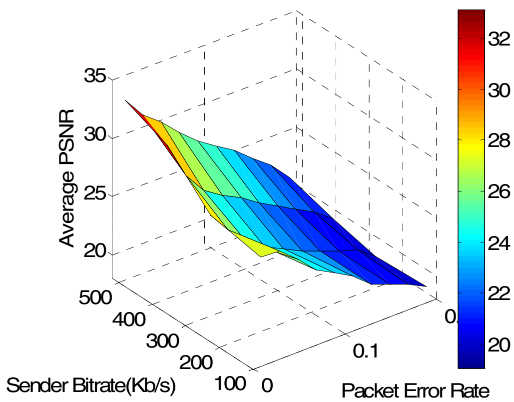

Figure 11. Average PSNR Vs PER and SBR for 'RM'

In Fig. 11 we chose bitrates of 80kb/s, 104kb/s, 256kb/s, 384kb/s and 512kb/s as bitrates less than 80kb/s will yield meaningless results. From our results we suggest a minimum send bitrate of 256kb/s as it yields a PSNR of 30dB. Increasing the send bit rate improves the quality with no packet loss. However, increasing the send bitrate does not compensate for the higher packet loss effect of streaming video quality for fast moving content due to network congestion issues.

Therefore, the quality of video in 'rapid movement' degrades much more rapidly with an increase in packet loss compared to that of 'slight movement' and 'gentle walking'.

D. Experiment4 – PSNR Vs Time

We further looked at the relationship between the PSNR over the entire duration of the sequence for all three content types.

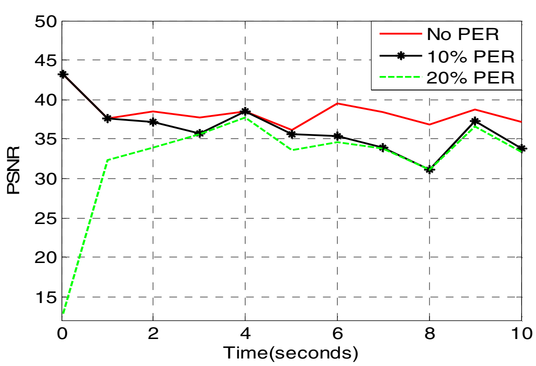

Figure 12. PER effects for SM for 32kb/s SBR

In Fig. 12 we investigate the source of effects caused by packet errors over the entire duration of the sequence. For 'slight movement' we compare the PSNR values for no transmission errors to 10% and 20% packet loss. The PSNR values are the same for a new I-frame over the duration of the sequence. The error occurs in the B-frames and propagates to the P-frames as expected. We observe two effects, the PSNR decreases over the entire duration and the second a more ragged response curve when packet errors of 10% and 20% are introduced. We also observe that for a send bitrate of 32kb/s the video quality is still acceptable for 20% packet loss.

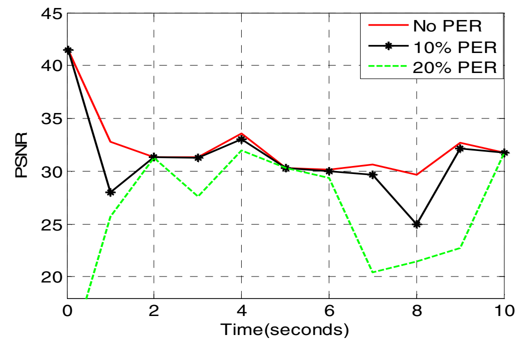

Figure 13. PER effects for GW for 80kb/s SBR

Fig. 13 shows the effects of no packet loss, 10% and 20% packet loss for 'Gentle walking' at a send bitrate of 80kb/s. Again as previously mentioned the video quality reduces over the time duration and we observe a much bigger loss in quality as the packet loss increases to 20%.

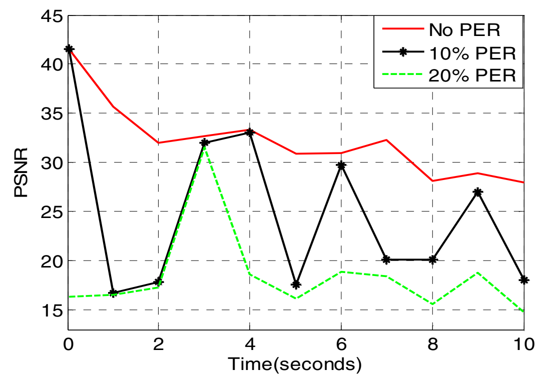

Figure 14. PER effects for GW for 80kb/s SBR

Whereas, from Fig. 14 in 'rapid movement' the video quality degrades fairly quickly with the increase of packet error rate i.e. for 20% packet loss the video quality is completely unacceptable.

While PSNR is not a good predictor of the visual quality, it can serve as a detector of clearly visible distortions. It can be observed, however that the perceived quality degradation increases in the duration of the sequence. Due to the auto-correlation of the time series (each sample is dependent on the previous and following sample) the values are not independent. We also observed that as the scene activity in the video sequence becomes more complicated e.g. for 'rapid movement' at 20% packet loss the quality is completely unacceptable deteriorating at a much faster speed. All degraded video clips can be found in [20].

V. Conclusions

Through this work we have classified the most significant content types and have established guidelines for the transmission of MPEG4 streaming video over wireless networks in terms of acceptable packet error and minimum send bitrate. The contents were first classified using cluster analysis into three groups with good prediction accuracy. The video quality is evaluated in terms of average PSNR and decodable frame rate. The acceptable PER was found to be 20%, 10% and 6% for the three content categories of SM, GW and RM respectively. We found that the PSNR was more sensitive to video content than decodable frame rate as the video quality of 'akiyo' in the category of 'SM' was overall best in terms of average PSNR compared to that of 'bridge-close' for the Q value.

We then derived the upper, medium and lower bounds for the three content types of SM, GW and RM to give the threshold for acceptable quality in terms of the PSNR and the Q value with 99.97% prediction accuracy. Similarly, we identified the minimum SBR for acceptable QoS for the three content types as 18, 32 and 256kb/s for SM, GW and RM respectively.

We further investigated the impact of video quality over the entire duration of the sequence and observed the PSNR variations by introducing a 10% and 20% packet loss confirming that the faster moving content is very sensitive to packet loss compared to slower moving content.

We believe that the results would help in optimizing resource allocation for specific content in content delivery networks and the development of QoS control methods for video over mobile/wireless networks. Future direction of our work is to further investigate the more perceptual-based quality metric and adapt the video send bitrate depending on network conditions over wireless networks.

Acknowledgment

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

References

[1] G. Ghinea and J. P. Thomas, "QoS impact on user perception and understanding of multimedia video clips", *Proc. Of ACM Multimedia'98*, Bristol, UK, pp. 49-54, 1998.

[2] A. Khan, Z. Li, L. Sun and E. Ifeachor, "Audiovisual quality assessment for 3G networks in support of E-healthcare", *Proc. of CIMED ,* Plymouth, UK, 25-27 July 2007.

[3]] P. Chondros, A. Prayati, C. Koulamas and G. Papadopoulos, "802.11 performance evaluation for multimedia streaming", *Fifth International Symposium on Communication Systems, Networks and Digital Signal Processing*, Patras, Greece, 19-21 July, 2006.

[4] B. Munir, N. K. Chilamkurti and B. Soh, "A comparative study of voice over wireless networks using NS-2 simulation with an integrated error model", *International Conf. on WiCOM,* 22-24 Sept. 2006.

[5] Y. Koucheryavy, D. Moltchanov and J. Harju, "Performance evaluation of live video streaming service in 802.11b WLAN environment under different load conditions", MIPS, Napoli, Italy, November 2003.

[6] Z. He, H. Xiong, "Transmission distortion analysis for real-time video encoding and streaming over wireless networks", IEEE transactions on Circuits and Systems for Video Technology, Vol. 16, No. 9, Sept. 2006.

[7] S. Kanumuri, P. C. Cosman, A. R. Reibman and V. A. Vaishampayan, "Modelling packet-loss visibility in MPEG2 video", *IEEE Transactions on Multimedia*, Vol. 8, No. 2, April 2006.

[8] Ke, C. Lin and C. Shieh, "Evaluation of streaming MPEG video over wireless channels", Journal of mobile multimedia, Vol. 3, No. 1, pp. 047-064, 2007

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

[10] V. Vassilou, P. Antoniou, I. Giannakou and A. Pitsillides " Delivering adaptive scalable video over wireless internet", *International Conference on Artificial Neural Networks (ICANN),* Athens, Greece, September 10-14, 2006.

[11] M. Ries, C. Crespi, O. Nemethova and M. Rupp, "Content based video quality estimation for H.264/AVC video streaming", *Proc. Of the IEEE Wiress Communication & Networking Conference*, Hong Kong, China, Mar. 2007.

[12] Y. Suda, K. Yamori and Y. Tanaka, "Content clustering based on users' subjective evaluation", Information and Telecommunication Technologies, 2005. APSITT 2005 Proceedings. 6th Asia-Pacific Symposium, Pages 177-182, Nov. 2005. W. J. Krzanowski, "Principles of Multivariate Analysis", Clarendon press, Oxford, 1998.

[13] W. J. Krzanowski, "Principles of Multivariate Analysis", Clarendon press, Oxford, 1998.

[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*, pp. 255-272, Urbana, Illinois, USA, September 2003.

[15] S. Winkler and F. Dufaux, "Video quality evaluation for mobile applications", *Proceedings of SPIE Visual Communications and Image Processing*, Lugano, Switzerland, 2003

[16] O. Nemethova, M. Ries, M. zavodsky and M. Rupp, "PSNR-based estimation of subjective time-variant video quality for mobiles", *Proc. of MESAQIN 2006*, Prag, Tschechien, June, 2006.

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

[18] R. J. Hyndman, "The problem with Sturges' rule for constructing histograms", Monash University 1995.

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

[20] www.tech.plymouth.ac.uk/spmc/staff/akhan/video_sequences/videoclips.html