On Measuring the Perceptual Quality of Video Streams over Lossy Wireless Networks

Ilias Politis, Michail Tsagkaropoulos, Tasos Dagiuklas, and Lampros Dounis

Department of Telecommunications Systems and Networks,
Technological Educational Institute of Messolonghi, Nafpaktos 30300, Greece
{ipolitis,mtsagaro}@ece.upatras.gr, ntan@teimes.gr,
dounis@gmail.com

Abstract. This paper studies the perceptual quality of video streams over lossy wireless networks. The focus is on investigating the impact on the perceived video quality of both physical error impairments and packet losses due to network congestion, by using objective and subjective evaluation methods. Extensive video quality assessments have shown that packet losses due to congestion are more severe than packet losses due to the physical error on the objective video quality. Furthermore, the comparison of MOS among different spatial resolution video sequences of the same bit rate indicates that a better perceived video quality can be achieved for lower resolution when the network is characterized by both high BER and network load.

Keywords: Simultaneous Double Stimulus for Continuous Evaluation method (SDSCE), Mean Opinion Score (MOS), Peak Signal to Noise Ratio (PSNR), video quality estimation (VQE), video streaming, wireless networks.

1 Introduction

Due to rapid growth of wireless communications, multimedia applications such as video conferencing, digital video broadcasting (DVB), streaming video, and audio over networks are becoming increasingly popular. In the recent years, this progress has also been aided by the proliferation of technologies such as IEEE 802.11X, 3G, LTE and WiMAX, and the trend has been to allocate these services more and more on mobile users. Mobile video delivery across heterogeneous wireless networks poses many challenges, including the issue of coping with losses due to both physical impairments and network congestion, as well as, maintenance of QoS and session continuity.

Wireless channels are prone to errors due to fading and interference effects resulting in error bursts. The bursty error characteristics of the wireless channel can be modeled by Gilbert-Elliot model using a discrete two state Markov Chain. The original two-state Markov model considers one “Good” and one “Bad” state where no errors occur in the “Good” State. The above model has been enhanced where errors occur in “Good” State. There are many authors that have studied the impact of physical error characteristics on the video quality [1] and [2], however most of these

L. Atzori, J. Delgado, and D. Giusto (Eds.): MOBIMEDIA 2011, LNICST 79, pp. 46–54, 2012.

research works focused on the monitoring of the distortion due to network impairments based on objective quality evaluation methods.

The deployment of real-time multimedia applications over wireless networks proved that the simple network parameters like bandwidth, loss, delay, jitter, etc. are inadequate to assess accurately the perceived by the human viewer, quality of service. As mobile users expect high perceptual quality that depends not only on technical parameters, but also on user experience. The operators need to control resources and at the same time maintain user satisfaction. To this end, several objective and subjective video quality assessment (VQA) methodologies have been used to evaluate video quality that can be introduced at any stage in the end-to-end video delivery system, including coding distortion, network impairments (congestion/packet loss, physical impairments) and decoding process (i.e. error concealment). Objective VQA methods to calculate video distortion in terms of parameters such as MSE, PSNR and SSIM have been studied extensively by [3]. On the other hand, subjective QoS measurements evaluate video as perceived by users, i.e., what is their opinion on the quality of particular audio/ video sequences, have been extensively studied by [4], [5] and described in ITU-T recommendation BT 500-11 [6].

As opposed to studies that consider the impact of physical layer and network layer impairments on the video quality, separately, this paper focus is on examining through extensive VQA tests (40 subjects-evaluators both experts in video QoS and non-experts, three different video sequences) how packet losses due to network congestion and physical channel errors, affect the perceived video quality using both objective and subjective evaluation methodologies. This methodology has been applied for different spatial resolution video sequences.

The rest of the paper is organized as follows. Section 2 outlines the wireless model that is used during the simulations. In Section 3 the simulation setup is described, while Section 4 includes an analysis of the selected objective and subjective methods used during the VQA procedure. The quality assessment scores are presented and discussed in Section 5 and Section 6 concludes the paper.

2 Modeling Wireless Error

Apart from losses that are due to network congestion, we are interested in studying the impact of physical impairments on the perceived video quality. The classical two-state Gilbert–Elliott model for bursty noisy channels [7],[8] has been extensively studied by many researchers [9]-[12]. In [9], a finite-state Markov channel is presented for packet transmission where the received instantaneous Signal-to-Noise Ratio (SNR) is partitioned into K disjoint intervals. The channel is in state k when the SNR takes a value within the k^{th} interval. Clearly, each state is characterized by a different BER in PHY layer. In this paper, the Rayleigh fading channel is reflected by a two-state Markov model. A low mobility scenario is assumed (5 Km/h) [13] and the SNR threshold has been used in order to determine the steady state probabilities, the average BERs within each state and the transitional probabilities between the two states as it is shown in the following Table 1. In the rest of the paper bad, medium and good channel qualities will be referred to as bad, medium and low physical channel BER, accordingly.

Table 1. Physical channel simulation parameters

	Bad Channel Quality	Medium Channel Quality	Good Channel Quality
SNR (Threshold)	25	30	35
BER^B	$1.29\ 10^{-2}$	1.2910^{-2}	$1.25\ 10^{-2}$
BER^G	$4.1\ 10^{-12}$	$1.3\ 10^{-13}$	$4.13\ 10^{-14}$
$P_{G\text{->}B}$	0.013	0.007	0.004
$P_{B\text{->}G}$	0.198	0.360	0.664
$P_{B\rightarrow B}$	0.802	0.64	0.336
$P_{G\rightarrow G}$	0.987	0.993	0.996

In Table 1, BER^B and BER^G refer to the bit error rate at the BAD and the GOOD states of the two-state GE model, accordingly. Obviously, BER^B is always larger than BER^G. In order to simulate three different wireless channel qualities (BAD, MEDIUM, GOOD), the values of BER^B and BER^G have been selected in such a way that when the channel is at a GOOD state the BER^B and BER^G are assigned their lowest values. Moreover, Table 1 includes the transition probabilities between the two states of the GE model.

3 Simulation Setup

In the process of video quality evaluation (VQE), three high quality uncompressed video sequences have been used named “highway”, “deadline” and “paris”, which are freely available by PictureTel at [15]. Both YUV 4:2:0 color CIF and QCIF spatial resolution of the three video sequences were used at a frame rate of 30fps. All video sequences have been compressed by the H.264/AVC reference encoder (JM12) available at [16]. The encoding configuration parameters include a GOP size of 12 frames (GOP has the form of IPP...I) a number of 5 reference frames and different QP values for every encoded video sequence in order to achieve the same (or almost the same) average encoded video bit rate among the CIF and QCIF resolutions of each video sequence.Table 2 summarizes the video sequences characteristics.

Table 2. Video sequence characteristics

Video Sequence	CIF (352×288)	QCIF(176×144)
Highway	700kbps@30Hz, QP=12	700kbps@30Hz, QP=6
Deadline	800kbps@30Hz, QP=12	800kbps@30Hz, QP=4
Paris	1.1Mbps@30Hz, QP=12	900kbps@30Hz, QP=2

A unicast H.264 video transmission (one video server and one video client) is simulated and a single NAL unit packetization scheme (one RTP packet – one NAL unit) is adapted with an RTP packet size of 1024 bytes (payload). The generated video packets are delivered through the simulated wireless network.

A NS-2 based simulation environment with the appropriate extensions for simulating 802.11b WLANs is adopted [17]. Additionally to the video server, a second server generates background traffic at Constant Bit Rate (CBR) over UDP in order to overload the simulated 802.11b network. The background CBR traffic is transmitted at three different transmission rates 2.5Mbps, 3.75Mbps and 4.5Mbps that correspond to 50%, 75% and 90% network load respectively.

4 Video Quality Evaluation

The aim of this study is to measure the perceived video quality of video streams over error prone wireless channels, using both objective and subjective video quality assessment (VQA) methods. In particularly, emphasis has been given on estimating the impact of both physical errors due to physical impairments and packet losses due to congestion, on the perceived video quality. Moreover, focus is given also on comparing the quality assessments of CIF and QCIF spatial resolution, under the same network conditions. The aim is to provide evidence that although the perceived video QoS under a lossless transmission environment is significantly better when the video sequence resolution is CIF instead of QCIF, in case of severe network conditions with high BER and network loads, the perceived QoS of a QCIF video sequence can be significantly better compared with the perceived QoS of the same video sequence at CIF resolution. To this end all the test video sequence have been encoded at the same (or similar) bit rate, thus the impact of background traffic will be almost the same in every case.

4.1 Objective Evaluation

As objective video quality evaluation method the Peak Signal to Noise Ratio (PSNR) is selected, since it is the most common, widely used by the research community and simple objective VQE scheme. In short PSNR is the ratio of the maximum (peak) power of the signal over the power of the signal's noise. In order to calculate the PSNR of a video sequence at the receiver, first Mean Square Error (MSE) between the original frame $\mathrm{F(i,j)}$ and the distorted frame $\mathrm{F'(i,j)}$needs to be defined as:

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|F(i,j) - F'(i,j)|^2$$

where, every video frame consists of $M \times N$ pixels. PSNR is then defined as the logarithm of the ratio of the peak signal value over the MSE due to noise. This also implies that PSNR calculation requires full video reference (i.e. reconstructed frame at the receiver), hence its use is limited in real time applications. Without loss of generality, in this study the distortion introduced to the video at the receiver is considered to be due to physical errors occurred at the wireless channel and packet losses due to congestion at the transport layer.

4.2 Subjective Evaluation

In the case of the subjective assessment of the perceived video quality, the tests have been carried out according to the ITU-T BT.500 recommendations for laboratory environments [6]. The simultaneous double stimulus for continuous evaluation method (SDSCE) has been preferred over the single stimulus schemes, since it is more appropriate for evaluating time varying degradations on the fidelity of visual information. According to SDSCE method, the original reference sequence and the test sequence are displayed simultaneously side by side. The subject is informed about the reference video (Stimulus A) and the distorted video (Stimulus B) and is allowed to evaluate continuously the test material in a scale from 0 (Bad) to 100 (Excellent) during the testing session. The votes from the voting bar are sampled every 0.5 sec, as described from the SDSCE in ITU BT-500.

In order to produce reliable and repeatable results, the tests have been conducted in a controlled testing environment provided by the Converged Networks and Services Group (CONES) of the TEI of Mesolonghi[1], Department of Telecommunication Systems and Networks, Greece. The facilities include high quality LCD displays, controlled light conditions and mid-gray background using appropriate curtains. During the video quality evaluation test 40 subjects were asked to evaluate the test videos. These subjects-evaluators include academic staff and students of the department. According to the subjective video assessment recommendations, the pool of the evaluators consists of a small number of experts on video quality, and the rest of them have no expertise in video evaluation. In accordance to the SDSCE specifications the test included three phases:

- a training and a demonstration phase – where subjects get familiar to the testing procedure and understand how to recognize artifacts.
- a pseudo-test phase – where selected represented conditions are shown with a different video sequence than the ones used for the test.

Moreover, to avoid subject's fatigue and decreasing level of attention, the test sessions lasted less than 30 minutes, including the training, demonstration and pseudo test phases. Since the entire set of test material presented as a single test session exceeds 30 minutes, multiple sessions were scheduled so that each subject could perform all sessions and rate all the test material.

Finally, the resulting scores need to be statistically processed before presented as final results. There is the need to remove subjects whose scores deviate from the scores of the other subjects, thus the technique of outliers detection was performed. The outlier detection refers to the detection and removal of scores in cases where the difference between mean subject vote and the mean vote for this test case from all other subjects exceeds 15%. This is a general rule that has been also used in other research works [14].

5 Processing of Results

In this section both objective and subjective evaluation results are illustrated and discussed. Since the aim is to identify the impact of physical errors and packet losses,

---

[1] `http://www.teimes.tesyd.gr/cones`

all video perceived QoS measurements for all six testing video sequences are compared against different bit error rates and network load conditions.

5.1 Objective Scores

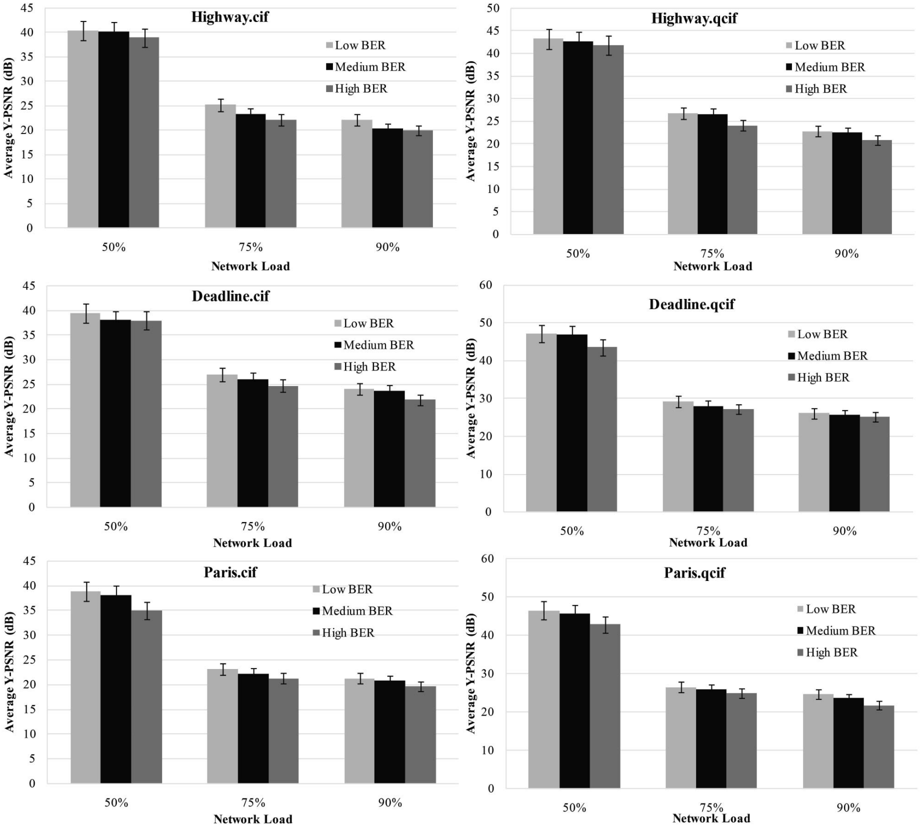

Fig. 1. PSNR measurements for CIF and QCIF spatial resolutions of the three test video sequences

The PSNR at the receiver for different network and channel conditions and a 5% error bar for each measurement, are illustrated in Figure 1. From the PSNR measurements, it can be seen that the increase in network load from 50% to 75% of the network capacity due to the background traffic, causes significant drop to the perceived video quality (e.g. from almost 40dB at load 50% to almost 25 dB at 75% load, in the case of “highway.cif”). A further increase of the network load from 75% to 90% of the network capacity results in a marginal drop of the average PSNR at the receiver (e.g. from 25dB at 75% load to 20dB at 90% in the case of “highway.cif”). Moreover, the impact of BER to the video quality at the receiver is limited mainly due to the fact the physical errors occur randomly and last only for short periods in time compared to packet losses from network congestion. In addition, the errors in the physical channel can be recovered using FEC mechanisms that are inherent at the physical and MAC layers. Similar conclusions can be derived from the QCIF measurements, as well. It must be mentioned that the average PSNR of the QCIF sequences measured at any physical and network conditions is higher than the corresponding average PSNR of the CIF sequences.

5.2 Subjective Scores

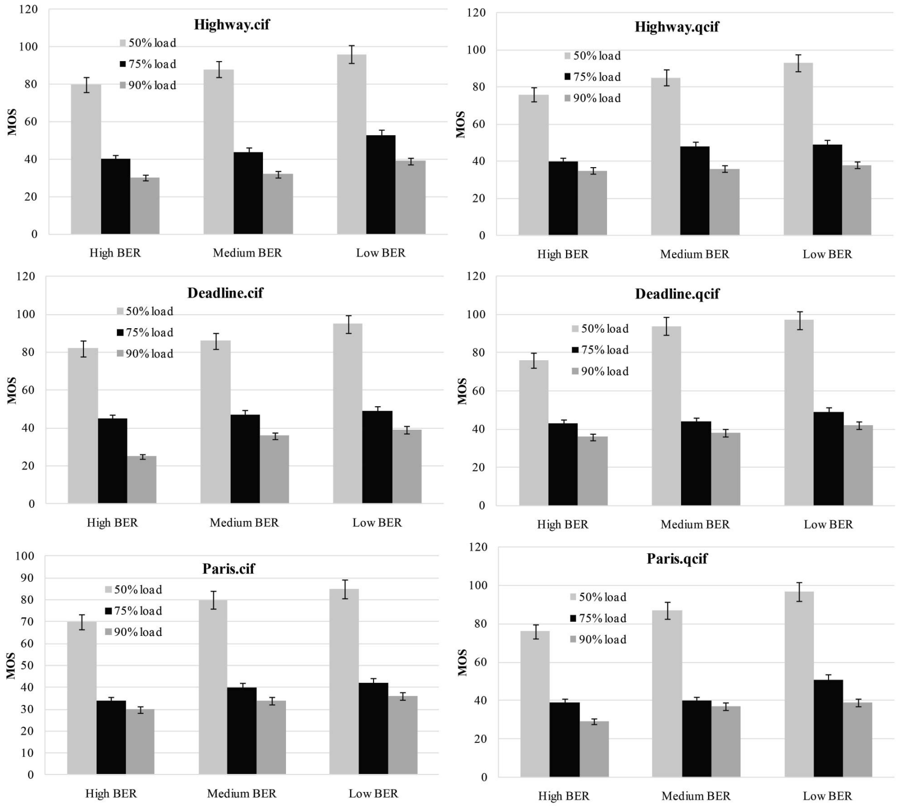

Fig. 2. MOS obtained with the SDSCE method for CIF and QCIF spatial resolutions of the three test video sequences

In Figure 2, the MOS obtained with the SDSCE video quality estimation method are shown. In particular, it is evident that the perceived video quality deteriorates fast as the network load increases, for the same channel conditions (BER). This MOS behavior is similar to the PSNR measurements, which means that the human viewer is more sensitive to the distortion that is introduced to the decoded video due to packet loss, rather than the distortion due to errors in the physical channel. Moreover, the comparison between MOS for CIF and QCIF, as shown in Figure 3, indicates that video streaming with lower spatial resolution may result in better-perceived QoS under high BER and increased network loads. However, this conclusion is not final as it also depends on the context of the video sequence, thus further investigation is required and more experiments to deeper understand the effect of visual context on perceived QoS are planned.

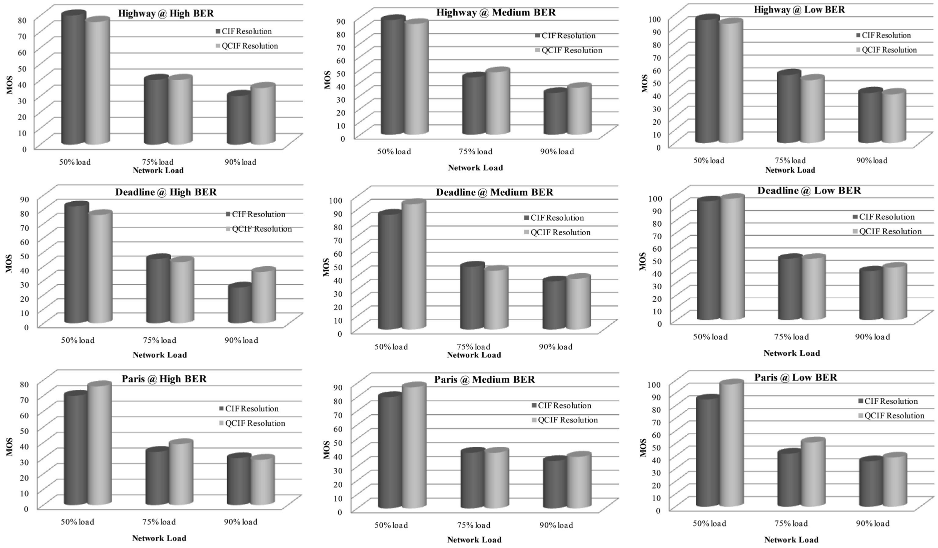

Fig. 3. Comparison of average MOS scores for CIF and QCIF sequences

6 Conclusions

This paper studies the impact of wireless physical channel impairments and packet losses due to network congestion on the perceived video quality of video streams with different spatial resolution. Extensive video quality assessments with objective measurements based on PSNR, as well as, subjective test according to the SDSCE method indicate that an increase of the BER has limited impact on the perceived video quality, as opposed to the impact on the QoS resulted by an increase of the network congestion. Moreover, the MOS comparisons among different network conditions indicate that better perceived video QoS can be achieved if lower spatial resolution video is transmitted over a network characterized by high BER and network load. Further experiments and real test-bed experiments are already undergoing, which will help to better understand the effect that specific visual context of the video sequence has to the video evaluator.

References

[1] Zhang, Q., Zhu, W., Zhang, Y.-Q.: End-to-End QoS for Video Delivery Over Wireless Internet. Proceedings of the IEEE 93(1), 123–134 (2005)

[2] Li, F., Liu, G.: Transmission Distortion Analysis for H.264 Video over Wireless Networks. In: 4th IEEE International Conference on Circuits and Systems for Communications, Shanghai, China, May 26-28, pp. 477–481 (2008)

[3] Chikkerur, S., Sundaram, V., Reisslein, M., Karam, L.J.: Objective Video Quality Assessment Methods:A Classification, Review, and Performance Comparison. IEEE Transactions onBroadcasting 57(2), 165–182 (2011)

[4] De Simone, F., Naccari, M., Tagliasacchi, M., Dufaux, F., Tubaro, S., Ebrahimi, T.: Subjective Quality Assessment of H.264/AVC Video Streaming with Packet Losses. EURASIP Journal on Image and Video Processing 2011, 1–12 (2011)

[5] De Simone, F., Goldmann, L., Lee, J.-S., Ebrahimi, T., Baroncini, V.: Subjective evaluation of next-generation video compression algorithms: a case study. In: Proc. of SPIE, San Diego, California (August 2010)

[6] ITU-R BT.500-11, Methodology for the Subjective Assessment (2002)

[7] Gilbert, E.N.: Capacity of a burst-noise channel. Bell Systems Technical Journal 39, 1253–1265 (1960)

[8] Elliot, E.: Estimates of error rates for codes on burst-noise channels. Bell Systems Technical Journal 42, 1977–1997 (1963)

[9] Wang, H.S., Moayeri, N.: Finite-state Markov Channel—A useful Model for Radio Communication Channels. IEEE Transactions on Vehicular Technology 44(1), 163–171 (1995)

[10] Zhang, Q., Kassam, S.: Finite-State Markov Model for Rayleigh Fading Channels. IEEE Transactions on Communications 47(11), 1688–1692 (1999)

[11] Sadeghi, P., Kennedy, R.A., Rapajic, P.B., Shams, R.: Finite-state Markov Modeling of Fading Channels: A Survey of Principles and Applications. IEEE Signal Processing Magazine 25(5), 57–80 (2008)

[12] Zhu, H., Karachontzitis, S., Toumpakaris, D.: Low Complexity Resource Allocation in Downlink Distributed Antenna Systems. IEEE Wireless Communications Magazine 17(3), 44–50 (2010)

[13] Aspelin, K.: (2005-05-25). Establishing Pedestrian Walking Speeds. Portland State University (retrieved August 24, 2009)

[14] Oelbaum, T., Schwarz, H., Wien, M., Wiegand, T.: Subjective performance evaluation of the SVC Extension of H.264/AVC. In: IEEE ICIP Conference, San Diego (October 2008)

[15] `http://media.xiph.org/video/derf/`

[16] H.264/AVC Software Coordination (2007), `http://iphome.hhi.de/suehring/tml/`

[17] The Network Simulator–NS-2, `http://www.isi.edu/nsnam/ns/`