Perceptual Video Quality Assessment in HTTP Adaptive Streaming

Jan Lievens[1], Adrian Munteanu[1], *Member, IEEE*, Danny De Vleeschauwer[2], Werner Van Leekwijck[2]
[1]Department of Electronics and Informatics, Vrije Universiteit Brussel – iMinds, Brussels, Belgium
[2]Bell Labs – Alcatel Lucent, Antwerp, Belgium

***Abstract*--In this paper, we propose a quality metric for video transmitted over HTTP adaptive streaming (HAS). Perceived video quality in HAS strongly depends on the employed rate adaptation mechanisms, transmission conditions and buffer filling levels. Predicting perceived quality in HAS is an open problem, due to the lack of appropriateness of existing metrics. To overcome this problem, a quality metric based on extensive user evaluations is proposed in this paper. A thorough investigation is carried out, accounting for all possible types of video quality impairments occurring in HAS applications. The proposed metric correlates well with the perceived quality, allowing for optimizing the HAS descriptions and of the switching mechanism between them. It is also shown that the classical intrinsic quality metrics (PSNR, SSIM, VQM) fail to predict user-perceived video quality due to quality fluctuations inherent to HAS systems.**

I. Introduction

Internet video delivery platforms are increasingly moving towards HTTP based technologies. HTTP adaptive streaming systems [1] (e.g. MPEG-DASH, Apple's HTTP Live Streaming, Microsoft's Smooth Streaming, Adobe's Dynamic Streaming ...) can accommodate a large array of end-users with different terminal and channel characteristics. All HAS systems rely for their video data transmission on client driven HTTP downloads. On the HAS server, video streams are partitioned into independently decodable segments. The HAS connection starts with a client downloading a media presentation description that will enable the client to select which from the available video segments to download based on both client terminal capabilities as well as current channel characteristics. The plethora of descriptions available on the server allows clients with widely different capabilities to connect. However, there is a trade-off between the amount of storage the server has to provide and the diversity of available video parameters. Each description constitutes an additional copy of the same source material. Typical parameters that can change between descriptions are spatial resolution, temporal resolution, and the amount of encoder quantization present in the coded video. These parameters each provide a different type of bandwidth adaptability to the system. Optimally selecting the parameters based on the perceived quality of the resulting video would help balance the trade-off. In this paper we evaluate the end-user perception of video where the content parameters are changed during playback, as would be the case in a HAS system.

II. Video Quality in HAS

Initial research on perceived quality of video has reported dependencies on spatial resolution, temporal resolution and quantization artifacts [2]. Compared to classical video broadcast systems where the video parameters (spatio-temporal resolution and image fidelity) are constant during playback, HAS video delivery systems will exhibit (frequently) varying parameters and thus affecting the perceived quality [3] [4]. The temporal effects on perceived quality are highly non-linear due to presences of persistence effects [5]. Existing research has mainly focused on low spatial resolutions, and low frame rate video source material. Our investigation focuses on current standards for high video fidelity, i.e. high definition spatial resolutions (1920x1080), frame rates beyond 30Hz and high signal to noise ratios. In this paper we take into consideration non-constant spatial resolution, temporal resolution and quantization. Additionally, we assess the effect on perceived end-user quality caused by the occurrence of frame freezes linked to buffer depletion.

III. Subjective Testing

A. Experiments

We have organized several rounds of subjective evaluations to measure the user-perceived quality linked to a variable video description during playback. Table 1 presents an overview of the experiments. In a first series of tests, the test panel was asked to score video sequences where all segments share the same description. These Mean Opinion Scores (MOS) correspond to the quality of a non-adaptive system where the channel characteristics are fixed. In a second series of tests, the test panel evaluated the quality of video where the segments' description changed once during playback. These tests were performed using short pieces of video material in which one distinct change occurred. This allows us to quantify the effect of individual parameter changes.

TABLE 1
EXPERIMENT DESCRIPTION

Test Series	Sequence Description(s)	Sequence duration	Segment duration
I	Single, constant	10s	10s
II	Two with one change	8s	2s
III	Multiple with multiple changes	50s	10s

A short pre-test session was held amongst a different test panel to establish the relationship between Quantization Parameter (QP) and MOS for each sequence. Based on this information, we have selected four QPs so that the entire MOS scale is evenly covered for each sequence. For a third series of experiments, we have exposed the panel to a set of sequences in which the video parameters changed multiple times. We have used these results to build a model that accounts for the time-dependent effects on the perceived quality.

978-1-4799-7543-3/15/$31.00 ©2015 IEEE

B. Experimental Set-Up

The subjective tests were held at our lab in Brussels, which has controlled lighting conditions featuring a color-calibrated screen specifically for the purpose of subjective evaluations of image and video data. The video material was coded using H.264/AVC. The material consisted of 2 sets of 3 1080p video sequences. The first set was used in experiments I and III. The other set was used for experiment series II. The experiments followed the Single Stimulus (SS) methodology, where the reference was shown only once at the beginning. The video description with the highest spatio-temporal resolution and the finest quantization served as reference during these tests. The test panels were drafted randomly amongst university students and employees. Groups of 17, 18 and 18 people participated in experiment series I, II and III respectively.

IV. MOS Results

A. Spatial Resolution and Quantization

When all segments belong to the same HAS description, the MOS increases with larger spatial resolution or higher bit rate. However, for each tested sequence we have found that a bit rate exists below which a lower spatial resolution (up-sampled at the receiver) is preferred over a higher spatial resolution. For each considered bit rate there is a combination of spatial resolution and quantization that yields the highest MOS. These points should be selected as descriptions.

When different descriptions are used during playback, we noticed that a visual fidelity decrease has an immediate negative effect on the MOS. However, an increase in fidelity translates into an increased MOS only after a certain delay. We consider that a persistence effect plays here, penalizing poor visual quality.

B. Temporal Resolution

We have tested frame rates that are integral divisions the base frame rate 50, namely 25, 16.7 and 12.5. The difference in observed quality for 50Hz and 25Hz is very small for static conditions. The test panel scored the lower two frame rates poorly compared to the original 50Hz. By alternating the frame rates between the above-mentioned options, the quality perception was lower than in the static case. Members of the test panel mentioned experiencing nausea in some of the more extreme changes between frame rates.

C. Frame Freezes

Frame freezes are always dynamic occurrences. The negative effect on perceived quality of playback freezes is highly non-linear and depends on the duration of the individual freeze as well as the frequency of occurrence.

V. Quality Metric

Based on our test results, the proposed MOS predictor *PQM* is given by (1), where T is the period over which the MOS is predicted, $T_{freezes}$ is the total time of the frame freezes, γ, ε, and α are constants and τ, β and δ are functions that reflect the changes in fidelity, amount of freezes and frame rate

$$PQM(T)=\frac{1}{T+\gamma T_{freezes}}\sum_{T} Q\left[fidelity\left(t-\tau\left(\frac{\partial fidelity(t)}{\partial t}\right)\right)\right] -\varepsilon\alpha^{\beta\left(\frac{\partial freezes(t)}{\partial t}\right)}-\delta\left(\frac{\partial framerate(t)}{\partial t}\right) \quad (1)$$

respectively, and Q is the encoder-side MOS associated with a given *fidelity*. Compared to traditional metrics for video quality our proposed metric outperforms the other metrics significantly in terms of correlation, as shown in Fig. 1.

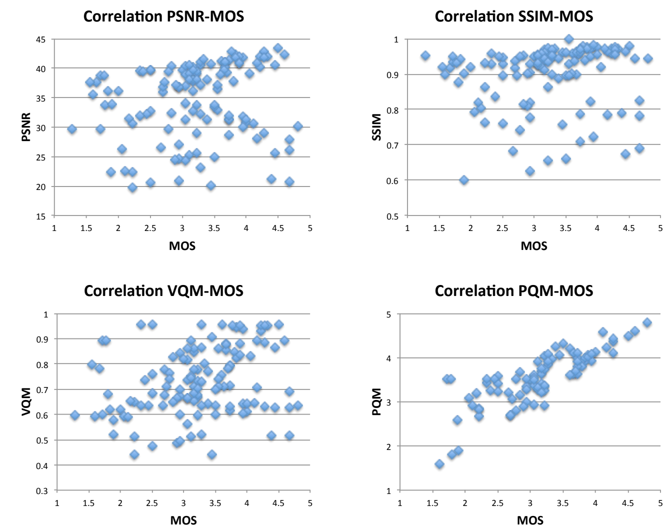

Fig. 1. Correlation between Mean Observer Score and different quality metrics. Top left: PSNR, top right: SSIM, bottom left: VQM, bottom right: the proposed PQM metric.

VI. Conclusion

In this paper, we have presented results for a series of subjective evaluations pertaining the variability in video quality in HAS. We have shown that the end-user evaluation of the visual impairments associated with HAS playback are poorly predicted by classical video quality metrics (PSNR, SSIM, VQM). Additionally, we have proposed a new quality metric that correlates well with the perceived visual quality.

References

[1] T. Stockhammer, “Dynamic adaptive streaming over HTTP – design principles and standards”, *ACM Conf. Multimedia Syst.*, Feb 2011.

[2] J.-S. Lee, F. De Simone, N. Ramzan, Z. Zhao, E. Kurutepe, T. Sikora, J. Ostermann, E. Izquierdo, and T. Ebrahimi, “Subjective evaluation of scalable video coding for content distribution”, *Proc. ACM Multimedia*, pp. 65–72, Oct. 2010.

[3] J. De Vriendt, D. De Vleeschauwer, D. Robinson, “Model for estimating QoE of video delivered using HTTP adaptive streaming”, *IFIP/IEEE Int. Workshop on QoE Centric Management*, pp. 1288-1293, May 2013.

[4] C.,Alberti, D. Renzi, C. Timmerer, C. Mueller, S. Lederer, S. Battista, M. Mattavelli, "Automated QoE evaluation of Dynamic Adaptive Streaming over HTTP", *IEEE Int. Workshop on Quality of Multimedia Experience*, pp. 58-63, July 2013.

[5] S. Tavakoli, K. Brunnström, K. Wang, B. Andrén, M. Shahid, et al., "Subjective quality assessment of an adaptive video streaming model", *Proc. SPIE 9016*, Jan. 2014.

[6] A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. de Veciana, "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies", *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, pp. 652-671, Oct. 2012