J. Vis. Commun. Image R. 27 (2015) 7–27

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Packet losses and objective video quality metrics in H.264 video streaming ☆

Franco Tommasi *, Valerio De Luca, Catiuscia Melle

Department of Engineering for Innovation, University of Salento, Lecce 73100, Italy

ARTICLE INFO

Article history:
Received 19 October 2013
Accepted 5 December 2014
Available online 16 December 2014

Keywords:
Video quality metric
Video over IP
Video streaming
Video encoding
Quality of experience
Packet loss
Video quality optimization
Video quality impairments

ABSTRACT

Video quality is a key factor in modern video streaming systems. Video compression artifacts affect it but also delay, jitter and packet loss may compromise it. Objective metrics have been proposed to emulate the human visual system: several experimental works have evaluated their adherence to opinions expressed by real users. Instead of dealing with subjective tests, we focused on the effects of network packet losses on objective metrics: since quality metrics are generally computationally intensive, a convenient approach could consist in inferring information about transmission quality from network impairment statistics. We did experimental tests on two computer-animated videos, whose high color contrasts allow a fair comparison between content dependent and content independent metrics. We studied the encoding parameters that minimize/maximize the values of some metrics (*PSNR*, *BI-PSNR*, *SSIM*, *3SSIM*, *MSSSIM*, *VQM*) for several packet loss percentages. We analyzed also the Empirical Cumulative Density Function of the degradations of quality metrics.

1. Introduction

Video quality evaluation is very important in both the design and the maintenance of VoD and real-time streaming systems. It aims at offering the best quality of experience (QoE) to end-users. In a design phase it is a critical factor in system requirements assessment, in the encoder selection and in the tuning of an encoding algorithm. In the maintenance phase of an existing system, video quality determines when a switch from one technique to another is needed to adapt to parameter changes. When a system is in operation, video quality estimation should be not only accurate, but also fast in order to allow quick corrective or recovery actions especially in presence of strict real-time constraints.

From a network point of view, video quality is indeed affected by bandwidth constraints, packet losses, delays and jitters, which are relatively fast and straightforward to measure or estimate. These parameters, which characterize the quality of service (QoS), are often inadequate to reliably estimate the video quality perceived by final users, which is known as quality of experience (QoE). The overall QoE is in part subjective and can be modelled by means of various parameters [1], which are difficult to quantify yet. The impact of network impairments on video quality also depends on some specific properties of the original video sequence, such as spatial and temporal resolution, amount of texture, content and especially content motion. A packet loss would produce significant quality degradation in animated video regions and a negligible effect in mostly still scenes [2].

Some metrics concerning motion properties of video sequences are described in [3]. A first measure of the temporal variation of a video is the frame difference: it is often normalized by the contrast to prevent high contrast sequences being attributed larger variations than similar less contrasted ones. The contrast is defined as the average standard deviation of the pixel values in each frame. The motion feature is defined as the mean of the top 10% motion vector magnitudes. Just as frame difference, also motion vector is often normalized by contrast.

Both subjective and objective methods were standardized for digital video quality assessment [4]. A typical example of subjective metric is the mean opinion score (*MOS*) [5,6], which is the mean of subjective ratings expressed by real users within a scale of values ranging from 1 (very bad) to 5 (very good). However, subjective methods are highly time consuming: they require a significant number of users watching the video and expressing a mark about the perceived quality. For this reason, researchers have tried to model the quality perceived by the human visual system, introducing plenty of objective metrics. Moreover, some empirical methods were proposed to predict*MOS* from objective measurements. The

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.
* Corresponding author.
E-mail addresses: franco.tommasi@unisalento.it (F. Tommasi), valerio.deluca@unisalento.it (V. De Luca), catiuscia.melle@unisalento.it (C. Melle).

http://dx.doi.org/10.1016/j.jvcir.2014.12.003
1047-3203/

Video Quality Expert Group (VQEG) was founded with the aim of evaluating objective quality metrics [7].

Objective quality metrics can be classified into full-reference, no-reference and reduced-reference metrics depending on the amount of reference information they require. Full-reference (FR) metrics perform a frame-by-frame comparison between the original video, which is required to be available in its uncompressed form, and the impaired test video. For this reason, they require a spatio-temporal alignment between the two videos. No-reference (NR) metrics analyse only the test video, trying to separate distortions from content: it is not an easy task, since it needs specific assumptions about the video content and the distortions of interest. Reduced-reference (RR) metrics perform a comparison between the original and the test video based on a limited number of features such as motion information or spatial detail. They have alignment requirements just as FR metrics, although in a less stringent form. While FR metrics are generally suitable for offline video quality evaluation, NR and RR metrics are better suited for a real-time monitoring of a running video system.

In this paper we will describe a RTP/UDP streaming experiment we made to study the effects of packet losses on some full-reference quality metrics. We will analyze *PSNR*, Brightness Independent *PSNR* (*BI-PSNR*), *SSIM*, *3SSIM*, *MSSSIM* and *VQM* metrics, which we will describe in Section 3. We used NetEm [8] to emulate 15 different packet loss percentages and EvalVid framework [9] and MSU VQMT [10] to compute quality metrics. The goals of the experiment we made were two:

- we tried to detect the encoding parameters (bitrate and keyframe frequency) that may optimize the average video quality of the whole transmission in presence of different packet loss percentages; since quality impairments may vary for videos encoded at various bitrates and keyframe frequencies, the goal was to achieve the best trade-off between compression artifacts and packet loss impairments;
- we evaluated how various quality metrics (*PSNR*, *BI-PSNR*, *SSIM*, *3SSIM*, *MSSSIM* and *VQM*) may be affected by packet losses by analyzing and comparing the ECDFs (Empirical Cumulative Density Functions) of their degradations.

We point out that the goal of this work is not that of evaluating the correspondence between objective metrics' predictions and the human perception, since other authors [11] have already performed several subjective tests with plenty of human beings to study that correlation. We rather want to evaluate how objective quality metrics react to packet losses. We think that such analysis may offer useful hints for both the fine-tuning of video source encoding parameters and the design of rate adaptive streaming systems, which constantly monitor network conditions and consequently adapt video streaming rate. Nowadays the adoption of rate adaptive streaming is increasing together with the concern in the Internet community for congestion collapses. For this reason, the IETF introduced the Datagram Congestion Control Protocol (DCCP) [12], a real-time transport protocol that supports the deployment of rate adaptive codecs to face queuing delay and packet loss issues. DCCP still uses UDP non-reliable packet flow, but it also establishes a session in which context the exchange of some parameters is allowed in a connection-oriented fashion. However, some packet losses are independent from congestion: they occur unpredictably and are often caused by noise on the network links. We want to analyze the effects of this category of loss events, which we will call "isolated packet losses", on video quality. Some papers [13,14] proposed FEC [15] and retransmission as techniques for recovering from solitary loss events: unfortunately, not only congestive losses but also many single packet losses within a short time interval may prevent FEC-based recovery. For this reason, we want to experimentally analyze whether lower bitrates may significantly reduce the distortion in video frames caused by isolated packet losses.

After quickly mentioning some related works in Section 2 (and specifically focusing on works dealing with subjective tests in Section 2.1), we will briefly describe some important quality metrics in Section 3. Then we will describe our experimental testbed in Section 4 and we will analyze the obtained results in Section 5. In Section 6 conclusions will be drawn.

2. Related works

Several algorithms have been proposed until recently to estimate in real-time QoS parameters such as bandwidth, packet loss, delay and jitter. Although QoS parameters give useful insights about network status, which is the key to success of many multimedia applications (especially real-time streaming), they are not obviously related to the quality perceived by end-users (QoE). On the other hand, especially full-reference quality metrics' computation is often very complex and requires a heavy workload, which is not tolerable in real-time systems. For these reasons, many authors have tried to detect, in both analytical and experimental ways, some kind of relation between QoE and QoS, whose parameters are certainly easier to compute. On the contrary, other authors tried to characterize distortions by only processing the impaired video (that is proposing no-reference methods).

In [16] a generic dependence between QoE and QoS was formulated by means of partial differential equations, which correlate an infinitesimal decrement ∂QoE in the user satisfaction with an infinitesimal increment ∂QoS of the level of disturbance. In [17] a more specific model, based again on partial differential equations, was presented: QoE represents a score in the *MOS* scale, while QoS refers to the packet losses. Both models show that QoE suffers more degradations when the current QoE level is high, even in presence of low QoS disturbance.

Some authors proposed a utility function [18] that predicts IPTV QoE on the basis of packet losses. The same authors in their subsequent work [19] expressed QoE as utility on the basis of end-to-end frame delay and planned to merge loss rate and delay into a comprehensive solution for QoE prediction.

In [20,21] a PSQA [22] technique was used to combine the loss rate of video frames (LR) and the mean size of lost bursts (MLBS) into perceived quality functions: in particular MLBS describes the way losses are distributed within the received stream.

In a previous work [23] we developed a library for building application layer multicast overlays for video streaming. We included also a module for QoS monitoring that tries to prevent QoE degradations by keeping under control QoS variations in terms of loss ratio and jitter. We computed autocorrelation for loss ratio and jitter to consider their historical evolution, assigning some weights also to the past values. In this way we aimed at partially ignoring transient degradations, which may not motivate a swap towards another video source (in real-time streaming systems any overlay refinement action has to be taken only in presence of persistent issues, since it may introduce a significant delay).

Some works [24,25] tried to experimentally detect mapping functions between QoE and QoS metrics. In [24] packet loss, jitter and delay are mapped into the subjective *MOS* score, derived from users' point of view, and the objective *VQM* metric. In [25] EvalVid integrated into NS-2 simulator is used to estimate QoS parameters, which are compared to perceived video quality, represented by *SSIM* index.

Ref. [26] evaluated the effects of a noise impulse on a MPEG-2 video. Furthermore, several databases [27–34] collected plenty of subject rated images: they show the artifacts deriving by