Signal Processing: *Image Communication* 28 (2013) 903–916

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Assessing quality of experience for high definition video streaming under diverse packet loss patterns

Mikołaj Leszczuk*, Lucjan Janowski, Piotr Romaniak, Zdzisław Papir

AGH University of Science and Technology, Kraków, PL-30059, Poland

ARTICLE INFO

Available online 23 October 2012

Keywords:
Objective evaluation techniques
Subjective evaluation techniques
RSVP
HDTV
QoS
QoE

ABSTRACT

This paper presents an approach for deriving a QoE model of High Definition video streaming in the presence of different patterns of packet losses. The goal was achieved by using the SSIM video quality metric, temporal pooling techniques and content characteristics. Subjective tests were performed in order to verify the proposed models. The impact of several network loss patterns on diverse video content was analysed. The paper also deals with the encountered difficulties and presents intermediate steps to give a better understanding of the final result. The research aims to evaluate the perceived performance of IPTV and video surveillance systems. The model has been evaluated in the Quality of Experience (QoE) domain. The final model is generic and shows high correlation with the subjective results.

1. Introduction

IP transmission of video streams can be affected by packet losses even if a resource reservation algorithm is used. Therefore, the effect of packet loss on the perceived quality needs to be considered carefully.

An important example of a streaming service is IPTV implemented by many network operators. The recent premiere of High Definition IPTV brought new requirements in terms of bit-rate and quality of service assurance. The problem of network losses remains vivid, and affects mainly the "last mile" of the delivery path. Competition on the markets is fierce and service providers desperately seek video quality monitoring and assurance solutions in order to satisfy growing numbers of quality-aware customers. The impact of network losses on the perceived video quality remains a challenging task because (among others) "not all packets are equal" as claimed in [1].

Evaluation of packet loss effect on video content has been extensively analysed over recent years. Several models have been proposed for low bit-rate videos (for mobile use) and Standard Definition (SD) resolution. The majority of the proposed solutions are known as parametric models, operating on the network and transport layers. Verscheure in [2] explains the problem of quality prediction and control of an MPEG-2 video stream transmitted through a lossy network. The MPEG-2 video standard is analysed and the impact on the visual quality of packet loss is discussed. In [3] the authors presented a quality metric based on two network parameters related to packet loss. An application of customer-oriented measurements for H.264/AVC video is presented in [4]. Another model accounting for effects of burst losses and correlation between error frames was detailed in [5]. It is dedicated to low bit-rate H.264/AVC video. In contrast to the parametric approaches, a simple model for network impairments based on image analysis was proposed by Dosselmann in [6].

The *structural similarity index Metric* (SSIM) [7,8] is a top-down approach using a functional model of the Human Visual System (HVS). More detailed description of HVS can be found in [9]. Suppose x and y the reference and the distorted image signals respectively. The overall

* Corresponding author. Tel.: +48607720398; fax: +48125342372.
E-mail address: leszczuk@agh.edu.pl (M. Leszczuk).
URL: http://qoe.kt.agh.edu.pl/ (M. Leszczuk).

0923-5965/$ - see front matter © 2012 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.image.2012.09.006

similarity metric $S(x,y)$ is combined of three components: local luminance $l(x,y)$, local contrast $c(x,y)$ and structure $s(x,y)$ comparison between the original and the distorted images. Video quality assessment using SSIM is performed in three levels: the local region level, the frame level, and the video sequence level. First, random 8×8 pixels blocks are extracted from the original and the distorted video sequences. In this level, SSIM index is calculated for each block for *Y*, *Cb* and *Cr* components separately. In the second step, local quality values are combined into a frame-level quality. Quality of local regions is weighted according to the mean luminance level (dark regions less sensitive to quality degradation). In the last step, quality for entire video sequence is computed. Frame-level quality is weighted using frame motion vectors, since some types of distortion (e.g. blur) does not affect perceived quality for scenes, where large motion occurs [7].

The *MOtion-based Video Integrity Evaluation* (*MOVIE*) *Index* for video quality assessment utilizes a general, spatio-spectrally localized multi-scale framework for evaluating dynamic video fidelity that integrates both spatial and temporal (and spatio-temporal) aspects of distortion assessment. Video quality is evaluated not only in space and time, but also in space–time, by evaluating motion quality along computed motion trajectories. The MOVIE index delivers quality scores that correlate quite closely with human subjective judgement, using the Video Quality Expert Group (VQEG) FRTV Phase 1 database and the LIVE Video Quality Database. Indeed, the MOVIE index is found to be quite competitive with, and even outperform, algorithms developed and submitted to the VQEG FRTV Phase 1 study, as well as more recent VQA algorithms tested on both databases [10].

The *Visual Information Fidelity* (*VIF*) is a measure for Image Quality Assessment criterion that quantifies the Shannon information that is shared between the reference and the distorted images relative to the information contained in the reference image itself. VIF uses Natural Scene Statistics (NSS) modelling in concert with an image degradation model and an HVS model. VIF is competitive with state of the art quality assessment methods, and outperforms them in some simulations [11].

The *Moving Picture Quality Metric* (MPQM) is an objective quality metric for moving pictures using a vision modelling approach [12]. The model accounts for spatio-temporal aspects of HVS, namely contrast sensitivity and masking. Based on assumption that HVS processes visual information in separated spatial and temporal channels, the original and coded video sequences are decomposed into perceptual channels segmented using uniform areas, textures, and contours classification block by block. Then, contrast sensitivity and masking are considered for each perceptual channel in detection threshold calculation. Afterwards, filtered error signal is divided by the detection threshold. In the last step, data from channels is gathered together in order to account for higher level quality and the overall distortion level is computed; this process is called pooling [12].

The *Perceptual Evaluation of Video Quality* (PEVQ) metric has been designed to estimate video quality degradation introduced within content networks. It is based on spatial and temporal artefacts measurement aided by replica of HVS [13]. PEVQ outputs Mean Opinion Score (MOS) [14] value ranging from 1 (bad) to 5 (excellent) as well as additional indicators for more detailed analysis of the perceptual level of distortion in the luminance, chrominance and temporal domain. The latest version of PEVQ, i.e. PEVQ v.2, is optimized and capable of a real-time video processing. It accepts the following input data: (i) AVI with RGB24, YUV video data, (ii) QCIF, CIF, VGA and Rec. 601 frame sizes, (iii) 6–20 s in length [13].

The NTIA General *Video Quality Metric* (VQM) measures the perceptual effects of video impairments including blurring, jerky/unnatural motion, global noise, block distortion and colour distortion, and combines them into one single metric [15,16]. VQM takes the original video and the processed video as input and is computed through the following steps: (i) Calibration: it estimates and corrects the spatial and temporal shift as well as the contrast and brightness offset of the processed video sequence with respect to the original one, (ii) Quality Features Extraction: a set of quality features is extracted that characterizes perceptual changes in the spatial, temporal, and chrominance properties from spatial–temporal sub-regions of video streams, (iii) Quality Parameters Calculation: this step computes a set of quality parameters that describe perceptual changes in video quality by comparing features extracted from the processed video with those extracted from the original video, and (iv) VQM Calculation: VQM is computed by using a linear combination of parameters calculated from previous steps [15,16].

The *V-Factor* is a particular MPQM implementation specifically designed for the Internet Protocol Television (IPTV), and leveraged MPQM research that several labs have developed over the last years [17]. Just like MPQM, V-Factor provides the video quality score but also some extra information needed for monitoring and diagnosing the root cause of problems.

Very few results are available on High Definition content. One of the first significant publications on this particular topic describes the performance of VQM in the High Definition TV (HDTV) video quality assessment task [18]. It is devoted mainly to compression artefacts for five different video encoders. Network losses are also considered, albeit with a lower stress. Another research paper published recently is dedicated exclusively to network losses [19]. Correlation of the three existing quality metrics was verified using the subjective results, namely PSNR, SSIM and VQM [15,16]. However, only one network loss pattern with a variable number of occurrences per video sequence was considered. Furthermore, subjective and objective scores were averaged over 2 min of video material consisting of 12 video sequences. This simplifies the quality assessment task because an important factor affecting the perceived quality is omitted this way. This factor is related to diverse content characteristics and may significantly affect the perceived quality of different types of content affected by the same (in terms of quantity) impairments [20,21]. As a result, the authors claim that even the PSNR metric can achieve extremely high correlation with the perceived quality, which is a very surprising result. Recent discussion on the performance of mean squared error metrics is presented by Wang and Bovik in [22].