

# Assessing the QoE in Video Services Over Lossy Networks

Francisco J. Suárez · Andrés García ·  
Juan C. Granda · Daniel F. García · Pelayo Nuño

Received: 2 January 2014 / Revised: 27 January 2015 / Accepted: 6 February 2015 /  
Published online: 12 February 2015  
© Springer Science+Business Media New York 2015

**Abstract** Quality of experience is of critical importance in streaming video services, because the traditional quality of service cannot represent the quality perceived by viewers. This work evaluates several objective quality metrics under realistic bursty packet loss conditions in the network, with the support of a packet loss model. Alignment of reference and streamed video sequences (with different levels of spatial-temporal information) are also explored as a technique to prevent inaccurate computation of objective metrics due to frame loss. Finally, the correlation between subjective and objective metrics for each motion level and the computing time of metrics are analysed. The most suitable objective metrics to characterize the real degradation in the quality perceived by viewers, for both off-line and real-time assessment, are proposed. The integration of motion, bursty packet loss, sequence alignment after frame loss and computing time of metrics are the main contributions of this research work.

**Keywords** Video streaming · Packet loss · Subjective and objective video quality metrics · Quality of experience

---

F. J. Suárez (✉) · J. C. Granda · D. F. García · P. Nuño  
Department of Computer Science and Engineering, University of Oviedo, 33203 Gijón, Spain  
e-mail: fjsuarez@uniovi.es

J. C. Granda  
e-mail: jcgranda@uniovi.es

D. F. García  
e-mail: dfgarcia@uniovi.es

P. Nuño  
e-mail: nunopelayo@uniovi.es

A. García  
A.G. Ingeniera y Sistemas S.L. (labSens), Avenida de Asturias 12, 5<sup>o</sup>A, 33600 Mieres, Spain  
e-mail: agmangas@gmail.com

### 1 Introduction

Quality is an important subjective concept. Like everything related to human perceptions, it can be said to be related to the level of satisfaction with an experience. In the case of streaming video services, if the end user is not satisfied by his experience, the utility of the service decreases. Thus, prediction and measurement of the quality perceived by viewers must be studied in depth. Providing streaming video services at different quality levels is one of the most important business models in current technology. Lossy encoding techniques are necessary to keep the network requirements of these services within specified limits. Lossy methods introduce degradation and distortions depending on the selected encoding parameters. Thus, user satisfaction must be correlated with encoding parameters to provide a satisfactory level of quality without using more resources than strictly necessary.

In addition to encoding parameters, impairments introduced by the network (such as packet loss) can lead to quality degradation. Previous work such as [1] has shown that packet loss patterns tend to be bursty; that is, the probability of packet loss is not randomly distributed but depends on previous losses. This can be represented with finite-state stochastic models adapted to the parameters and characteristics observed in traffic traces obtained under realistic conditions [2, 3].

Traditional quality of service (QoS) is based on network metrics such as jitter, throughput, packet loss, corrupted packets, and latency. Two approaches can be considered: the reservation of network resources by applications, and the differentiation between services carried out by network hardware such as routers. The main problem with classic QoS is the heterogeneous nature of multimedia services in the Internet, which complicates the identification of the most suitable metric for each case. In addition to network features, other aspects such as codecs, compression, video source, or the nature of the content must be taken into account. In [4] it is concluded that the classical approach is no longer valid when considering services oriented to the end user. Although network metrics are still important, they can no longer provide a faithful representation of the quality perceived by the user. As a result, a new vision should be introduced: quality of experience (QoE).

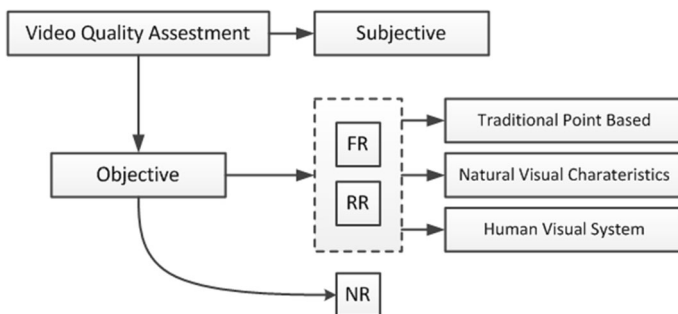


Fig. 1 Video quality assessment classification

QoE is meaningful only within the mind of the user; it is qualitative rather than quantitative, built from different factors and human behaviours that together provide a final level of perceived satisfaction. A measure of the quality perceived by viewers can be obtained through a wide variety of assessment methods. These methods are classified as subjective or objective. The former involve the participation of viewers and require time and resources. The latter can be implemented on computing devices and automated, although they are not as reliable as their subjective counterparts. There is some consensus in the classification of existing methodologies to assess video quality, with minor differences between the most prominent authors. In this section, we follow the classification presented in [5] and shown in Fig. 1.

Subjective methods are based on the selection of a small representative group of people exposed to a variety of video qualities in a controlled situation. Their opinions are registered on a scale that measures relative degradation or absolute quality. Closely related is the concept of mean opinion score (MOS), which is the most common level of satisfaction for a given resource as interpreted by an average user on a numerical scale.

Objective methods represent human perception to predict the MOS in an automated way. There are three main classes:

- Full reference (FR) methods insert reference signals alongside the tested signals to improve the estimation. They are usually more precise, but need more resources to process the extra workload.
- Reduced reference (RR) methods insert features extracted from the reference signal instead of the reference itself. They introduce a reasonable overhead and are an intermediate state between FR and NR methods.
- No reference (NR) methods only take into account the degraded signal, and are therefore adequate when the reference, or information about it, is not available.

FR and RR objective methods are further classified as follows:

- Traditional point-based metrics (FR only) are traditional mathematical metrics (PSNR, MSE) calculated comparing the reference and the noisy/degraded sequences in the luminance or colour components.
- Natural visual characteristic models extract statistical and visual features from the sequences. Some of these models use traditional point-based metrics as their foundation.
- A third group of metrics model the human visual system (HVS) (e.g., modelling the visual masking effects that occur in the eye) operating in the pixel or frequency domains of the video sequences (e.g., through DCT transform).

The objective of our work is to correlate subjective and objective metrics when video sequences are received by viewers under realistic packet loss conditions in the network. As a result, the objective metrics most suited to characterize the real degradation observed by viewers will be fixed.

The remainder of this paper is organized as follows. Section 2 is an overview of the related work. Section 3 describes the packet loss model that supports the emulation of realistic bursty packet loss conditions in the network. Section 4 introduces the technique of sequence alignment used to prevent inaccurate computation of objective metrics due to frame loss. Section 5 details the experimental framework for the quality assessment. Section 6 presents and analyses the results of the assessment. Finally, Sect. 7 presents our conclusions and outlines future work.

## 2 Related Work

Several works have been published that measure the influence of packet loss on video quality. The effects of the main network impairments (packet loss and jitter) on the perceptual quality of the video are analysed by emulation in [6]. The author concludes that the effects are similar. In fact, jitter could be mostly translated to extra packet loss when analysing the effects of network impairments. The impact of packet loss on MPEG streams transmitted over the Internet is measured in [7]. In [8] the results of an exhaustive experiment to detect the possible artifacts that may appear in an MPEG-2 video due to packet loss is presented. The results relate the number of frames affected by packet loss to the number of frames in the final degraded sequence that present artifacts, highlighting the nature and quantity of these artifacts. The effects of packet loss on video-chat applications [9] and HDTV [10] are also analysed.

The literature also includes exhaustive analyses of video traffic over Internet in streaming systems [11], conferencing systems [12], or IPTV systems [13] where packet loss rates and patterns are detailed. These patterns can be used to emulate packet loss in further experiments under realistic conditions with the support of a suitable packet loss model.

In recent years, much work has been focused on designing new and efficient approaches to quality assessment, to obtain reliable, simple, global metrics to represent perceived quality of multimedia services. There are also studies that compare and classify these proposed metrics. The studies developed by the video quality experts' group (VQEG) [14] are among the most relevant and recognized, included as recommendations of the International Telecommunication Union (ITU). In [5], a more recent extensive classification, comparison, and review of FR and RR objective video quality assessment methods are presented. A comparison of objective quality metrics for video scalability can also be found in [15].

As our work is a comparison of FR and RR objective video quality metrics, this background review focuses mainly on these metrics. The most simple metrics in this category are those originally developed for analysis of still images. Although these metrics only consider spatial quality degradation in video frames, they can be complemented with temporal aspects for video quality prediction. Peak signal to noise ratio (PSNR) is one of the most popular metrics because of its simplicity and direct mathematical interpretation. The original video sequence is the reference, and the streamed (degraded) video sequence is considered as noise, the metric representing the

level of degradation between the two. This metric is currently adopted as the reference metric for performance comparisons in video quality assessment (VQA). Structural similarity (SSIM), presented in [16], is based on the comparison of the luminance, contrast and structural components of both the reference and degraded sequences. The multi-scale SSIM (MS-SSIM) metric [17] is based on the previous SSIM metric. The authors argue that the perceptual evaluation of a given image varies with the parameters that depend on the signal, the visualization environment, and the particular characteristics of the observer. Thus, a single-scale method seems to be appropriate only for a limited set of specific settings. A multi-scale method is proposed as a more convenient way to consider these parameters. The noise quality measure (NQM) [18] measures the effect of additive noise on the human visual system. A degraded image is modelled as an original image subjected to linear frequency distortion and noise injection. A complementary metric, distortion measure (DM), is also defined. The NQM takes the following into account: the variation in contrast sensitivity with distance, image dimensions, and spatial frequency; variation in the local luminance mean; contrast interaction between spatial frequencies and contrast masking effects. The [19] universal quality index (UQI) is a mathematically defined metric that models any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. Visual information fidelity (VIF) [20] approaches the image quality assessment problem as an information fidelity problem. It is an image information measurement that quantifies the information that is present in the reference image and how much of this reference information can be extracted from the distorted image.

When the image quality metrics described above are applied to the VQA, the metrics are computed for each frame and finally aggregated into a single value. The aggregation or pooling strategy [21] assigns weights to the frames in order to correlate the quality scores of viewers. The average score of all the frames is the basic pooling strategy [22]. An example of this kind of video quality metrics is VSSIM [23], the metric based on SSIM specifically designed for video. It leads to a quality measurement that aggregates SSIM metric measurement from all sampling windows in all frames. Lower weights are given to frames with dark regions, because they attract fewer fixations, and to high motion frames, where distortion perception competes with the motion perception. Other VQA works based on image quality metrics are VIF for video [24] and PSNR for 3D video [25].

A second group of FR and RR objective quality metrics are those specifically designed for video quality assessment, taking into account aspects of both spatial and temporal domains. The discrete cosine transform-video quality metric (DCT-VQM) [26] (based on Watson's proposal in [27]) operates in the frequency domain of the sequences through a DCT. It takes into account the decrease in human visual sensitivity at high spatial and temporal frequencies. The VQM [28] is a different metric supported by the General Model. The model utilizes reduced-reference parameters extracted from spatial-temporal regions of the video sequence. It includes associated calibration techniques that comprise a complete automated objective video quality measurement system. The calibration of the original and processed video streams includes spatial alignment, valid region estimation, gain and level offset calculation, and temporal alignment. The General Model contains

seven independent parameters, four based on features extracted from spatial gradients of the luminance component, two based on features extracted from the chrominance components, and one based on the product of features that measure contrast and motion, both of which are extracted from the luminance component. The metric was included in two International Telecommunication Union (ITU) recommendations. A recent highly competitive metric is motion based video integrity evaluation index (MOVIE) [29]. Quality is evaluated in space and time using motion information from the reference video sequence, and the spatial and temporal quality scores are pooled to obtain an overall video integrity index score. MOVIE is also related to SSIM and VIF metrics, but is a very complex and computational intensive quality metric. Perceptual evaluation of video quality (PEVQ) [30] is a recent, very precise, patented algorithm, divided into four blocks. The first block, pre-processing, is responsible for the spatial and temporal alignment of the reference and the impaired signal. The second block calculates the perceptual difference of the aligned signals. The third block classifies the previously calculated indicators and detects certain types of distortions. Finally, in the fourth block, all the appropriate indicators according to the detected distortions are aggregated, forming the final result—the mean opinion score (MOS). A family of RR VQA algorithms, that vary in the amount of reference information required for quality computation, is presented in [31]. Finally, recent works deal with VQA under packet loss influence, e.g., the algorithm with low computational cost in [32], the packet layer model for VQA in [33] and the perceptual model in [34].

Although not directly related to this work, there are also several recent works dealing with NR metrics, more suitable for real time monitoring of video quality [15, 35–42]. They propose QoE models and metrics considering system parameters like the packet loss rate and application level parameters like sender bitrate, frame rate, and content type.

In the context of the packet loss impact on video quality assessment, [43] addresses the problem of frame desynchronization due to packet loss in video sequences evaluated with objective metrics. Since FR and RR objective metrics are calculated by comparing corresponding frames in the original and degraded sequences, this desynchronization generates incorrect metric scores. The authors highlight the fact that almost no previous work considers this problem and proposes a frame-matching algorithm, extending the peak signal to noise ratio (PSNR) metric to include the effects of packet loss. The new metric is called MPSNR and is validated through subjective assessment. Another work that considers frame synchronization is [44]. Once the frames are synchronized, several pooling strategies can be adopted to weigh the video frames when computing the quality metric [21, 22].

Many of the most widely referenced and adopted methods regarding subjective assessment are proposed by the International Telecommunication Union (ITU), e.g., [45]. Taking into account the related work and to the best of the authors' knowledge, there is further need to compare the quality prediction of FR and RR objective metrics with the opinion of the end users. The comparison should consider different motions in video content and realistic bursty packet loss conditions. With the study we would achieve a better understanding of the degradation in the QoE due to packet and frame loss, and how objective metrics reflect that degradation.

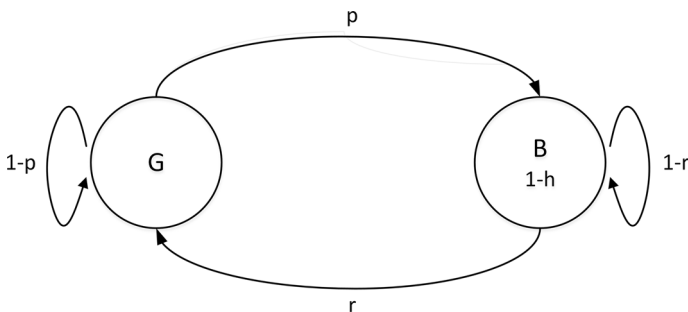
### 3 Packet Loss Model

The simplest packet loss model is the Bernoulli model, which is only able to represent uncorrelated losses and whose only parameter is loss probability. Because of its simplicity, it is not suitable to represent packet loss in real networks.

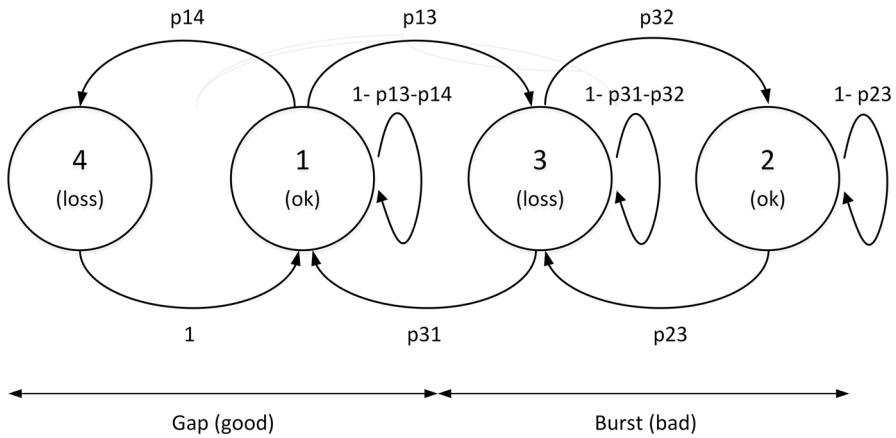
One of the first and most referenced models to represent bursty losses was proposed by Gilbert in [46]. The Gilbert Model is based on a Markov chain and has two states, a zero loss state G (Good) and a lossy state B (Bad). No errors occur in the G state, and the loss probability in the B state (loss density within the burst) is  $1-h$ . Two other parameters,  $p$  and  $r$ , represent the transition probability between states G and B. A graphical representation of the Gilbert model is shown in Fig. 2.

The Gilbert model was extended in [47] by Elliot. The Gilbert-Elliot model introduces a new parameter  $(1-k)$  that represents the loss probability when the system is in G state. Thus, it is possible to have loss events in both states G and B (note that the Gilbert-Elliot model is equivalent to the Gilbert Model when  $k=1$ ). This model is able to represent systems that have one state in which the loss rate is low (packet losses in this state can be considered independent) and another state where loss probability is relatively high (bursty loss).

Another model (*4 state Markov*) is implemented in the patch available for the Linux kernel [48]. It is based on the previous models and is the combination of two 2-state Markov models that represent two periods: a burst period (similar to the previous B state) and a gap period (similar to the G state). The resulting four states, shown in Fig. 3, have a specific physical meaning: packet received successfully (1), packet received within a burst (2), packet lost within a burst (3), and isolated packet lost within a gap (4). This model uses transition probabilities between states as parameters to characterize the packet loss process. These parameters are not closely related to quantities that have an understandable meaning for an end user, so a set of five equivalent and more intuitive parameters are defined for the model: loss probability, mean burst length, loss density within the burst, isolated loss probability, and mean good burst length. An interesting feature of this model is that through manipulations and assumptions (e.g., no correlation between losses occurred during burst periods), the initial set of five input parameters can be reduced to represent the Gilbert-Elliot, Gilbert and Bernoulli models.



**Fig. 2** Gilbert model



**Fig. 3** Four state Markov model

Gilbert models have been widely used for emulation of packet loss during transmission in IP networks. When used together with real traffic traces, the results show that these simple models fit the observed loss patterns, as stated in recent works [2, 3].

To achieve valid results and generate packet loss traces that resemble real networks, the input parameters for the model must be realistic and obtained from real traces in previous work. Research conducted to obtain data from recent studies of traces to feed a five parameter model gave poor results. Thus, for this work, a compromise between simplicity and accuracy was chosen, opting for a simplification of the original *4-State Markov* model with two parameters. This simplified model is equivalent to a Gilbert model with  $h = 0$  (*Simple Gilbert*). The two-parameter *4-State Markov* is identified by the transition probabilities  $p$  and  $r$ , or by the loss parameters  $P_{Loss}$  and  $E(B)$ , where  $P_{Loss}$  is the packet loss probability and  $E(B)$  is the average loss burst length. Equation 1 defines the correspondence between the two pairs of parameters.

$$p = \frac{P_{Loss}}{E(B)(1 - P_{Loss})} \quad \text{and} \quad r = \frac{1}{E(B)} \tag{1}$$

The simplicity of the selected model has the disadvantage of somewhat lower accuracy, but the advantage of managing only two very intuitive parameters. This makes it much more feasible to find studies of real network traces to obtain data to feed the model.

### 4 Sequence Alignment

Video sequences streamed through lossy transport networks tend to lose frames in the process, as explained in [43]. This generates a critical problem when calculating FR and RR metrics. Metrics usually compare sequences of reference and streamed video frame to frame, but if the degraded sequence has lost frames, both sequences



became miss-aligned, causing inaccuracies in quality metrics calculations. Both high and low quality degraded sequences tend to produce low quality scores that can seriously influence the correlation with the perceptual quality assessed by viewers. Thus, the degraded sequences must be aligned with the original sequence.

VQA studies based on metrics that were originally developed for image quality assessment do not usually consider sequence alignment. Even recent top ranked metrics specifically developed for VQA, like VQM, PEVQ or MOVIE, do not explicitly handle frame loss despite their complexity. Thus, in this work we address the sequence alignment problem explicitly.

Figure 4 compares the computing of metrics with and without alignment. Frames in the reference sequence are consecutively numbered and the frames in the streamed sequence are identified with their original numbers. If there is no alignment, frame loss is not analysed. In consequence, the frames of the sequences are incorrectly compared. Metrics are calculated by comparing the first frame of the

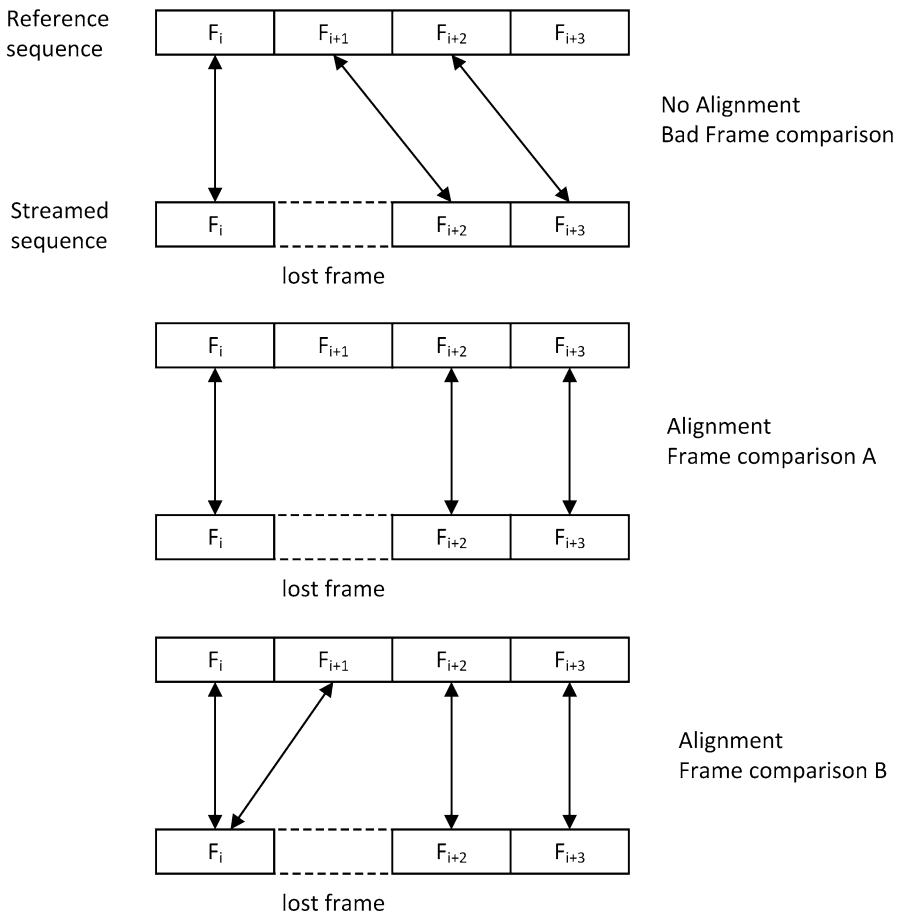


Fig. 4 Sequence alignment

streamed video with the first frame of the reference video, and then comparing the second frames of the streamed and the reference videos, and so on. Thus, the frame in the streamed sequence received after the lost frame ( $F_{(i+2)}$ ) is compared with the frame ( $F_{(i+1)}$ ) in the reference sequence. The same gap is introduced in the comparative of subsequent frames. Additional frame loss, both consecutive and isolated, increases the gap in the comparatives, making metrics' calculation less accurate.

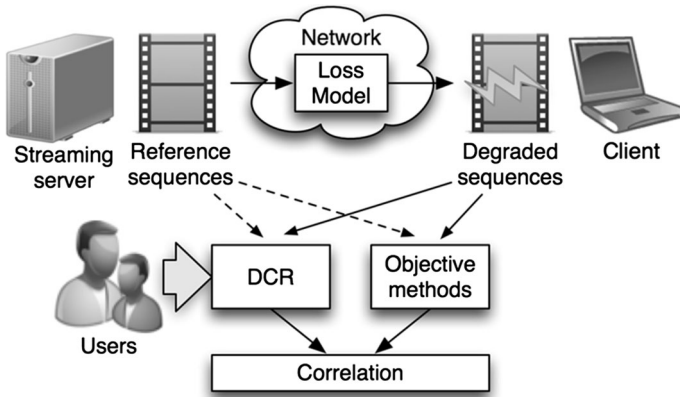
Sequence alignment requires a previous matching of the frames in the streamed sequence with the frames in the reference sequence. Thus, the matching process helps to locate the correct frame for comparison. The International Telecommunication Union (ITU-T) has recently proposed an improved PSNR calculation algorithm to address the problem of constant delays in a processed video [49]. Although it did not consider the problem of frame loss in the processed video, its approach to finding the corresponding frame could be used. In this work, we use the matching process proposed in [43], where the similarity between the streamed video and the reference video is used to find the correct match. The proposal assumes that the sum of PSNR of all frames in a streamed video is at its maximum when all the frames are correctly matched with the corresponding frames in the reference video. The dynamic programming algorithm shown in Eq. 2 is used to obtain the maximum PSNR value of the streamed video sequences and also finds the optimum match of the frames in the streamed video to the frames in the reference video. The time complexity of the matching algorithm is  $O(gn)$ , where  $n$  is the number of frames in the streamed video and  $g$  is the total number of frames lost. This time complexity can be reduced to values close to the time complexity  $O(n)$  of the traditional PSNR through the heuristic algorithm detailed in [43].

$$OPT(i, j) = \max[psnr(i, j) + OPT(i - 1, j - 1), OPT(i - 1, j)] \quad (2)$$

Once the sequences are aligned, we propose two alternatives for frame comparison. The first and simplest compares the matched frames directly, ignoring the corresponding lost frames in the reference sequence. The second alternative takes into account the common way of playing a streamed video, where the frame that precedes a lost frame is “frozen” until the arrival of the next frame. This implies comparison of a frame in the streamed sequence ( $F_i$ ) not only with the matched frame in the reference sequence ( $F_i$ ), but also with the frame or frames in the reference sequence that were lost just after ( $F_{i+1}$  in this example).

## 5 Experimental Framework

Figure 5 resumes the experimental framework for computing and evaluating video quality metrics. The videoLan client (VLC) application [50] was used as a streaming server and video streaming client. The streaming server was configured to use the real time streaming protocol (RTSP) when transmitting the test video streams. Finally, the packet loss model was implemented in a network emulator, the



**Fig. 5** Experimental framework

NetemCLG patch for the Linux Kernel [48]. In the following subsections, each part of the framework is detailed.

### 5.1 Objective Quality Metrics

The selection of FR objective metrics for comparison in our study is based on the non patented character of the metrics and the public availability of tools and software for computing them. Finally, eight metrics have been selected, six originally designed for image quality assessment: PSNR, SSIM, MS-SSIM, NQM, UQI and VIF; and two specifically designed for video quality assessment: DCT-VQM and VQM. The metrics originally developed for still image analysis are computed using the Python Visual Quality Metric Package (PyMetrikz) [51]. The Moscow State University Video Quality Measurement Tool (MSU VQMT) [52] allows us to compute the DCT-VQM metric. Finally, the matlab based visual quality metric (VQM) software of the Institute of Telecommunication Sciences (ITS) in the US National Telecommunication and Information Administration (NTIA) [53] was configured to compute the VQM metric with our test video sequences.

### 5.2 Parameters of the Packet Loss Model

A representative packet loss model is implemented in our test environment to faithfully represent realistic network conditions. These conditions have been extracted from the related work in the video traffic analyses [11–13] which were previously commented on above. From the results of these works and our first tests with the streaming environment, we reach the following conclusions.

- Values of  $P_{Loss} > 5\%$  completely degrade the video sequence and produce the lowest perceptual quality possible with many visible artifacts. The streams often block suddenly in the middle of the transmission.

- $E(B)$  is approximately within the range of 1 – 2 for most of the lost packets, increasing with the packet loss rate.

Taking these conclusions into account, a total of 36 values, 6  $P_{Loss}$  and 6 for  $E(B)$ , were selected for the parameters of the packet loss model, as shown in Table 1. Most of the values are in the region of  $P_{Loss} \leq 1\%$  because the decrease in perceptual quality caused by an increase in packet loss is more pronounced when packet loss is low. In addition, the central point of  $E(B)$  increases with  $P_{Loss}$ . As detailed later, a more compact set of values, shown in Table 2, will also be used in order to reduce the number of experiments.

### 5.3 Video Sequences

The video sequences selected are cartoons, where the complexity of the video objects is low and viewers assessment is clearly linked to the video artefacts caused by network impairments.

Because the quality degradation caused by packet loss increases with the motion level, spatio-temporal information was also used to select the video sequences. Repositories of standard video sequences like *ITS Standard Video Sequences* and *Video Trace Library* were analysed in order to select the sequences, but only one cartoon with few motion changes was found. Finally, the sequences were recorded from TV.

Three different sequences—with high, medium and low motion levels—were chosen, selecting sequences that show scenes with similar conditions of color, luminance, etc. Thus, the parameter that differs in each of the sequences is the spatio-temporal information level. The three sequences can be downloaded from [54] and the spatio-temporal information of the sequences is measured with the metrics proposed in [45]:

*Spatial information (SI)*: Level of detail of the sequence, that is, the complexity of its elements.

$$SI_n = std_{space}[Sobel(F_n)] \tag{3}$$

$$SI_{scene} = max_{time}(SI_n) \tag{4}$$

Equations 3 and 4 are used to calculate the spatial information of sequences.  $F_n$  is the  $n$ th frame of the sequence. The global  $SI$  level of a sequence ( $SI_{scene}$ ) is the

**Table 1** Full set of values for parameters of the packet loss model

$P_{Loss}$	0.1 %	0.25 %	0.5 %	1 %	2.5 %	5 %
$E(B)$	1.0	1.0	1.0	1.0	1.1	1.2
	1.1	1.1	1.1	1.1	1.2	1.3
	1.2	1.2	1.2	1.2	1.3	1.4
	1.3	1.3	1.3	1.3	1.4	1.5
	1.4	1.4	1.4	1.4	1.5	1.6
	1.5	1.5	1.5	1.5	1.6	1.7

**Table 2** Reduced set of values for parameters of the packet loss model

$P_{Loss}$	0.25 %	1 %	2.5 %	5 %
$E(B)$	1.0	1.0	1.2	1.3
	1.3	1.3	1.4	1.5
	1.5	1.5	1.6	1.7

maximum of the *SI* levels of its frames. *Sobel* refers to a Sobel filter for edge detection.

*Temporal information (TI)*: Motion level of a sequence over time, that is, the dynamism of its elements.

$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j) \tag{5}$$

$$TI_n = std_{space}[M_n(i, j)] \tag{6}$$

$$TI_{scene} = max_{time}(TI_n) \tag{7}$$

Equations 5, 6 and 7 are used to calculate temporal information about sequences. The *TI* level of a frame  $TI_n$  is the standard deviation of  $M_n$ , where  $M_n$  is the pixel plane that comes from removing the luminance component of frame  $F_{n-1}$  from the luminance component of frame  $F_n$ . The *SI* and *TI* of the selected sequences are shown in Table 3 ( $\mu$  and  $\sigma^2$  are the mean and variance of the *SI* and *TI* for all frames of the sequence). The final specifications for the video sequences are listed in Table 4.

The length of the sequences is determined from the number of packets needed for the loss model to converge, that is, to show the loss behaviour expected from the input parameters. Salsano et al. [48], *Sqngen* is a loss sequence generator, developed with the network emulator, to evaluate the behaviour of the various loss models before implementing them. It provides statistical information (mean, variance, and standard deviation) after generating a series of sequences with certain input parameters, helping to check the convergence of the loss model with the number of packets. The *Sqngen* tests (with an increasing number of packets) allowed us to find a threshold of 4000 packets for the model to converge. With this number of packets, the average value of all 95 % confidence intervals is 0.013 for packet loss

**Table 3** Spatio-temporal information of sequences

	High	Medium	Low
$TI_{scene}$	46.5410	54.5922	17.8479
$\mu_{TI_n}$	17.6871	2.8905	1.8155
$\sigma^2_{TI_n}$	75.4723	26.2917	6.0946
$SI_{scene}$	122.5812	117.2555	89.7910
$\mu_{SI_n}$	85.9009	76.4562	87.6395
$\sigma^2_{SI_n}$	125.3617	206.0620	0.6432

**Table 4** Video sequences specification

Video codec	H.264/MPEG-4 AVC
Bitrate	800 Kbps
Frames per second	25 fps
Resolution	HVGA (480 × 320)
Container format	AVI
Length	~ 1 min

probability and 0.011 for average loss burst length among all the video sequences generated in the experimental plan detailed below.

The packet size corresponds to the maximum transmission unit of transmission control protocol (TCP MTU, 1500 bytes) and the video bitrate is 800 Kbps, so the time to send 4000 packets can be calculated. Taking into account that not all the data in the packets corresponds to video information, the length of time of the video sequences which guaranties more than 4000 packets is about a minute.

#### 5.4 Subjective Assessment Methodology

To obtain subjective data, the *degradation category rating* (DCR) method (introduced in [45]) was used. Due to the high number of different sequences (many different quality levels), we used the extended degradation scale (from 1 to 9) as proposed in appendix V of the recommendation. The viewers must evaluate the perceived quality in the degraded sequence when compared to the reference sequence. A score of 1 means the highest degradation (minimum quality) and 9 means no perceived degradation at all (maximum quality). Ten viewers participated in the experiment (a number within the range suggested by the recommendation): 6 men and 4 women, 20–29 years old, most of them with university degrees and a relationship with the media world. The sequences were presented to them in pairs (the reference along with the degraded) in sessions of 30 min, with 15 min of break time between each session. The results of the subjective experiment captured the perception of viewers, and once correlated with the objective metrics, allowed us to estimate the QoE with network impairments.

The subjective scores given by the viewers for the video sequences in the training set ranged from 1 to 9, covering the full scale. This shows that we chose degraded sequences covering a wide range of quality. The average value of all 95 % confidence intervals for the scores among all the sequences in the training set was 0.45 on the 1–9 scale and hardly varied with the motion level. This small confidence interval indicates a very good agreement among the viewers.

#### 5.5 Evaluation Metrics

The Pearson correlation coefficient (PCC), the linear correlation coefficient (LCC), and the outlier ratio (OR) were the metrics used to evaluate and compare the selected objective quality metrics. The *PCC* coefficient represents the strength of the linear relationships between two variables. The *PCC* between variables  $X$  and  $Y$

can be calculated with Eq. 8, where an absolute value of 1 means a perfect relationship, and 0 means no relationship at all.

$$PCC = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (8)$$

In our study, *PCC* represents the correlation or level of linear relationship between the objective and the subjective data. It is important to have a strong linear relationship to derive subjective information from objective data. Using a linear regression (least squares), we modelled the subjective data with a third degree polynomial. The regression function provided a formal way to predict subjective scores from objective scores. The *LCC* represents the PCC between the original subjective data and the output of non-linear cubic polynomial regression. High values of this coefficient mean that the approximation fits the observed subjective data very well; that is, that the predictions obtained from the regression function are reliable. Finally, we obtained the *OR* that is defined as the percentage of the number of predictions outside the range  $\pm 2\sigma_{dcr}$ , where  $\sigma_{dcr}$  is the standard deviation of the subjective data.

## 5.6 Experimental Plan

Taking in account the number of selected metrics (8), sequence alignment alternatives (3), motion sequences (3), and values for packet loss probability (6), and the average loss burst length (6), the total number of experiments to carry out can be obtained. If the influence of the sequence alignment is also explored, the total number of experiments is 5184. This number of experiments is too high to be practical, because some metrics are computationally intensive and the test video sequences are long. For this reason, the experiments were carried out in two phases.

In the first phase only three PSNR-based metrics were evaluated: a traditional PSNR without sequence alignment; a “modified” PSNR with previous sequence alignment and frame comparison of type A (M-PSNR-A); and a modified PSNR with previous sequence alignment and frame comparison of type B (M-PSNR-B). The objective in this phase was to fully understand the influence of sequence alignment, video motion, and packet loss on PSNR, commonly used as a reference metric for comparison with other metrics. For each test video sequence (high, medium, and low motion), each metric was computed for the full combination of parameters (36) of the packet loss model, resulting in 324 experiments in total. The pooling strategy adopted was the simple mean of the frame scores, since this is the most suitable for long test videos, as stated in [22]. Finally, the PCC coefficients for each metric and motion were calculated. Phase I ends with the selection for each motion level, of the PSNR-based metric with the best PCC. These selected metrics will be the reference metrics in the next phase.

In the second phase, the comparison of all the selected metrics was carried with and without sequence alignment. Thus, the influence of sequence alignment on the performance of the metrics could be analysed. To reduce the number of experiments

in this phase, the combinations of parameters of the packet loss model were reduced from 36 to 12, as shown in Table 2. This reduced set of values was selected ensuring very similar values of the PCC for the PSNR-based metrics to those obtained for the full set. The number of experiments in this phase was 504, making 828 experiments in total for both phases instead of the original 5184.

## 6 Results

The evaluation of the selected video quality metrics was carried out following the experimental plan in the two phases detailed above. In each phase, the results were analysed independently for each motion level in video sequences and also for the aggregation of the video sequences. Aggregation helps to understand the behaviour of the metrics for any kind of video. Correlation for the aggregation of video sequences is always lower than correlation for a specific motion level.

### 6.1 Phase I: Initial Evaluation of PSNR-Based Metrics

Table 5 shows the PCC values and corresponding 95 % confidence intervals for the three reference PSNR-based metrics, computed from the full set of values of the packet loss model. As indicated above, the results are presented independently for each level of motion in video sequences and also for the aggregation of the video sequences. Correlations clearly increased with the level of motion in video. For high motion, both alternatives of the modified PSNR had correlations much higher than the traditional PSNR, highlighting the importance of sequence alignment. For low motion, the modified PSNRs had correlations slightly above the traditional PSNR. For medium motion, correlations of modified and traditional PSNRs were also very close but, surprisingly, the correlation of traditional PSNR was above. Finally, for aggregation of sequences, the modified PSNRs had correlations slightly above traditional PSNRs. In summary, the sequence alignment seems to be critical only for

**Table 5** Correlation of PSNR-based metrics for motion scenarios

Motion	Metric	<i>PCC</i>	<i>CI</i> (95%)
High	PSNR	0.6849	0.4591–0.8270
	M-PSNR-A	0.9198	0.8468–0.9585
	M-PSNR-B	0.9201	0.8474–0.9587
Medium	PSNR	0.9057	0.8211–0.9510
	M-PSNR-A	0.8834	0.7811–0.9391
	M-PSNR-B	0.8832	0.7808–0.9390
Low	PSNR	0.8478	0.7191–0.9199
	M-PSNR-A	0.8562	0.7335–0.9244
	M-PSNR-B	0.8553	0.7320–0.9239
Aggregated	PSNR	0.7516	0.6551–0.8232
	M-PSNR-A	0.7825	0.6959–0.8460
	M-PSNR-B	0.7799	0.6925–0.8441



high motion, and much less important for low and medium motion levels. In all cases, the two alternatives of the modified PSNR have very similar correlations, and so the type of frame comparison does not seem to be relevant for the level of frame loss in the test video sequences. Finally, we selected the traditional PSNR and the M-PSNR-A (with the simplest frame comparison) to be the definitive reference metrics in the next evaluation phase.

### 6.2 Phase II: Comparison of all the Selected Metrics

In this phase a comparison of all the selected metrics was carried out with and without sequence alignment. Figure 6 shows the PCC and LCC values of the eight selected metrics for high motion sequences. The PCC values are represented for both the traditional metrics without sequence alignment, and the modified metrics with sequence alignment (noted as “mod” in the figure), while the LCC values are represented only for modified metrics. Traditional PSNR and M-PSNR-A, as indicated in the first phase, are the reference metrics. The strong linear relationships between all the metrics is because network impairments introduced in high motion sequences were magnified and easier to detect both for viewers and for the metrics. For a better comparison of metrics, the curves were built in decreasing order of the PCC values of the modified metrics. All the modified metrics, with sequence alignment, have higher correlations than the corresponding traditional metrics without sequence alignment. This is consistent with the results in the first evaluation phase. The slope of both PCC curves are almost opposite, so the behaviour of metrics drastically changed with sequence alignment for this level of motion. The correlation increased with sequence alignment 55.7 % for the best metric and 28.3 % on average for all the metrics. Modified versions of NQM, PSNR, and VQM are the metrics with the best correlations for this level of motion, with similar high

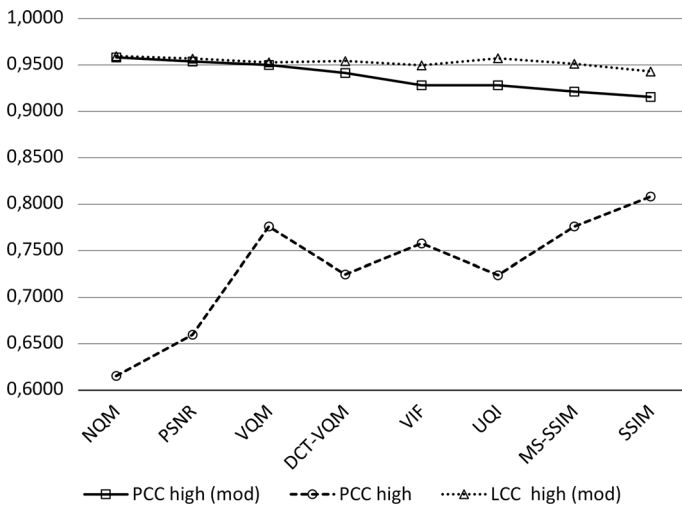


Fig. 6 Correlation of objective quality metrics for high motion

values above 0.95. All the other metrics have correlation values above 0.91. The LCC values are between 0.95 and 0.96 for seven of the eight metrics.

Figure 7 shows the PCC and LCC values of the metrics for low motion video sequences. These values are quite low in general if compared with the high motion scenario. The low dynamism of the scenes in low motion sequences helps the codec to predict video frames, even under frame loss, so degradation is lower. The viewers are able to detect the slight degradation differences, but the objective metrics are not. This is reflected in the low correlation of the metrics. Only two metrics, the modified versions of UQI and VIF, have correlation values above 0.91, the rest are below 0.87. It is interesting to note the low correlation values of the two “native” video quality metrics, DCT-VQM and VQM. The correlation of metrics increases with sequence alignment for three metrics, decreases for four metrics, and does not change for one metric; on average the differences are negligible. The LCC values are also at their maximum for the best correlated metrics, with values around 0.92.

Figure 8 shows the PCC and LCC values of the metrics for medium motion video sequences. In general, correlation values have intermediate values between those corresponding to high and low motion. The curves were built in decreasing order of PCC values for the traditional version of the metrics. With regard to the influence of sequence alignment on correlation, the slightly worse correlation of a modified PSNR in the first evaluation phase is confirmed here for the modified versions of all other metrics, with a decrease of 3.1 % on average. The curves are quite plain in this scenario, with low differences between correlations. The traditional DCT-VQM has the highest correlation, with a value close to 0.93, but is quite remarkably the lowest correlation of VQM. The LCC values, around 0.94 for most of the metrics, correspond to traditional metrics.

Figure 9 shows the PCC and LCC values of the metrics for the aggregated sequences. All the metrics (except VQM) have a higher correlation with sequence

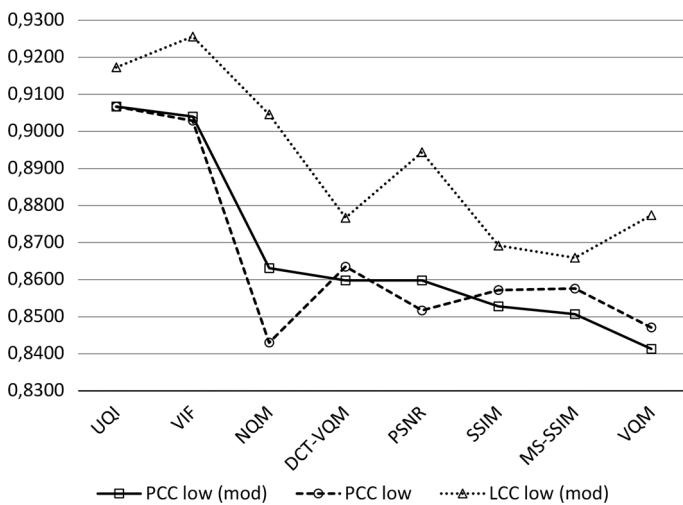
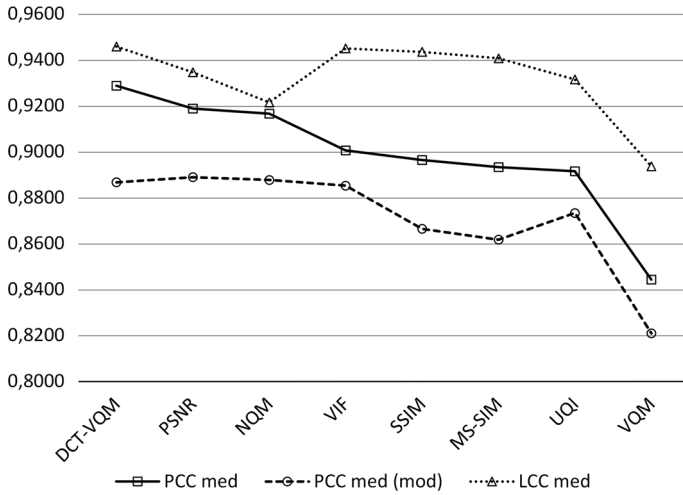
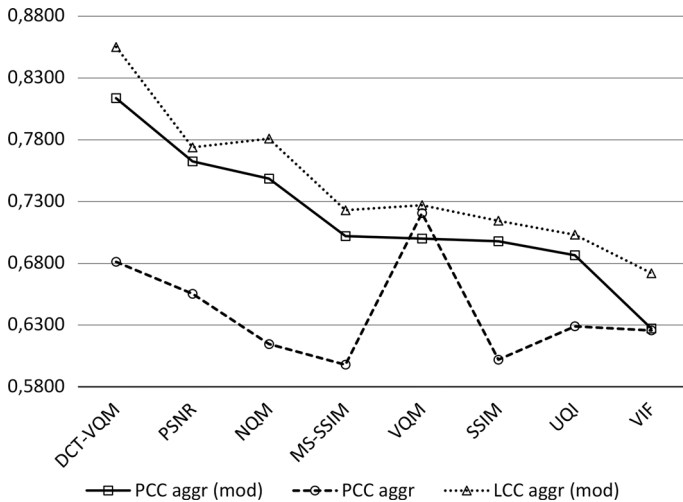


Fig. 7 Correlation of objective quality metrics for low motion



**Fig. 8** Correlation of objective quality metrics for medium motion



**Fig. 9** Correlation of objective quality metrics for aggregated sequences

alignment, with an increase of 19.5 % for the best metric and 11.9 % on average for all the metrics. The best PCC value, the only one above 0.8, and the best LCC value above 0.85, also correspond to the modified version DCT-VQM, followed by modified versions of PSNR and NQM.

Figure 10 shows the outlier ratios corresponding to the best correlated metrics for each motion and the aggregated sequences.

Finally, Fig. 11 shows the best PCC values of the metrics for each motion level in video sequences and also for the aggregated video sequences. An additional

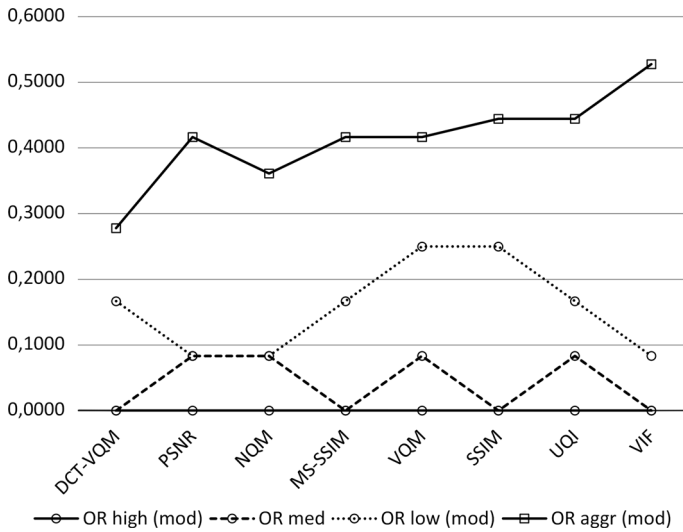


Fig. 10 Outlier ratios of polynomial predictions

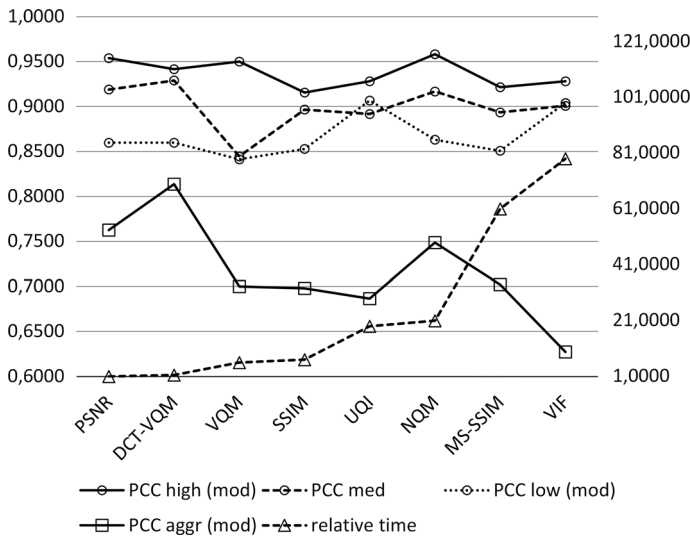


Fig. 11 Correlation and computing time of the objective quality metrics

curve, with an independent scale on the right, represents the approximate relative computing times for the metrics. The time of the modified version of the PSNR, with value 1, is the reference time. The curves were built in increasing order of computing time. The metrics with the best correlation values, the modified versions of the DCT-VQM and PSNR, are also the fastest to compute. Table 6 summarizes our proposal of metrics for off-line and real-time assessment. In the table, the modified versions of metrics are denoted with a *M*- prefix.

**Table 6** Proposed metrics for off-line and real-time assessment

Motion	Computing	Metric	<i>PCC</i>	<i>LCC</i>	<i>OR</i>
High	Off-line	M-NQM	0.9582	0.9597	0
	Real-time	M-PSNR	0.9538	0.9570	0
Medium	Off-line	DCT-VQM	0.9290	0.9460	0
	Real-time	DCT-VQM	0.9290	0.9460	0
Low	Off-line	M-UQI	0.9067	0.9173	0.1667
	Real-time	M-PSNR	0.8598	0.8944	0.0833
Aggregated	Off-line	M-DCT-VQM	0.8137	0.8551	0.2778
	Real-time	M-DCT-VQM	0.8137	0.8551	0.2778

## 7 Conclusions and Future Work

In this paper, we have explored the influence of important aspects that affect the correlation between objective and subjective video quality metrics in a QoE assessment: motion level in video sequences with bursty packet loss and sequence alignment after frame loss. As expected, the results showed higher correlations with the level of motion in video sequences. Furthermore, sequence alignment strongly increased correlations for high motion videos, making it a mandatory technique for this common kind of video. On the other hand, the technique did not change correlations for low motion videos and slightly decreases correlations for medium motion videos. Finally, taking into account the relative computing times of the metrics, the most suitable objective metrics to characterize the quality scores from viewers, both for off-line and real-time assessments, were proposed. Future work is focused on assessing new objective metrics and exploring the potential of standard video sequences in publicly available databases for packet loss analysis.

## References

- Altman, E., Avrachenkov, K., Barakat, C.: TCP in presence of bursty losses. *Perform. Eval.* **42**(2–3), 129–147 (2000). doi:[10.1016/S0166-5316\(00\)00027-4](https://doi.org/10.1016/S0166-5316(00)00027-4)
- Hasslinger, G., Hohlfeld, O.: The Gilbert–Elliott model for packet loss in real time services on the internet. In: Proceedings of the 14th GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB 2008), pp. 269–283. Dortmund, Germany (2008)
- Ellis, M., Pezaros, D., Kypraios, T., Perkins, C.: Modelling packet loss in RTP-based streaming video for residential users. In: Proceedings of the 37th IEEE Conference on Local Computer Networks (LCN 2012), pp. 220–223. Clearwater, USA (2012). doi:[10.1109/LCN.2012.6423613](https://doi.org/10.1109/LCN.2012.6423613)
- Van Moorsel, A.: Metrics for the internet age: quality of experience and quality of business. In: Proceedings of the 5th Performability Workshop, pp. 26–31. Erlangen, Germany (2001)
- Chikkerur, S., Sundaram, V., Reisslein, M., Karam, : Objective video quality assessment methods: a classification, review, and performance comparison. *IEEE Trans. Broadcast.* **57**(2), 165–182 (2011). doi:[10.1109/TBC.2011.2104671](https://doi.org/10.1109/TBC.2011.2104671)
- Claypool, M., Tanner, J.: The effects of jitter on the peceptual quality of video. In: Proceedings of the 7th ACM International Conference on Multimedia (ACM Multimedia'99) (part 2), pp. 115–118. Orlando, USA (1999). doi:[10.1145/319878.319909](https://doi.org/10.1145/319878.319909)

7. Boyce, J., Gaglianello, R.: Packet loss effects on MPEG video sent over the public internet. In: Proceedings of the 6th ACM International Conference on Multimedia (ACM Multimedia'98), pp. 181–190. Bristol, England (1998). doi:[10.1145/290747.290770](https://doi.org/10.1145/290747.290770)
8. Greengrass, J., Evans, J., Begen, A.: Not all packets are equal, part 2: the impact of network packet loss on video quality. *IEEE Internet Comput.* **13**(2), 74–82 (2009). doi:[10.1109/MIC.2009.40](https://doi.org/10.1109/MIC.2009.40)
9. Boyaci, O., Forte, A., Schulzrinne, H.: Performance of video-chat applications under congestion. In: Proceedings of the 11th IEEE International Symposium on Multimedia (ISM'09), pp. 213–218. San Diego, USA (2009). doi:[10.1109/ISM.2009.45](https://doi.org/10.1109/ISM.2009.45)
10. Pinson, M., Wolf, S., Cermak, G.: HDTV subjective quality of H.264 vs. MPEG-2, with and without packet loss. *IEEE Trans. Broadcast.* **56**(1), 86–91 (2010). doi:[10.1109/TBC.2009.2034511](https://doi.org/10.1109/TBC.2009.2034511)
11. Loguinov, D., Radha, H.: Large-scale experimental study of Internet performance using video traffic. *Comput. Commun. Rev.* **32**(1), 7–19 (2002). doi:[10.1145/510726.510727](https://doi.org/10.1145/510726.510727)
12. Calyam, P., Lee, C.G.: Characterizing voice and video traffic behavior over the Internet. In: Proceedings of the 20th International Symposium on Computer and Information Sciences (ISCIS 2005). Istanbul, Turkey (2005)
13. Ellis, M., Perkins, C.: Packet loss characteristics of IPTV-like traffic on residential links. In: Proceedings of the 7th Annual IEEE Consumer Communications and Networking Conference (CCNC 2010), pp. 1–5. Las Vegas, USA (2010). doi:[10.1109/CCNC.2010.5421735](https://doi.org/10.1109/CCNC.2010.5421735)
14. Final reports from the video quality experts group on the validation of objective models of video and multimedia quality assessment. Technical report, Video Quality Experts Group (VQEG). “<http://www.its.bldrdoc.gov/vqeg/reports.aspx>”. Accessed 23 Jan 2015
15. Besson, A., De Simone, F., Ebrahimi, T.: Objective quality metrics for video scalability. In: Proceedings of the 20th IEEE International Conference on Image Processing (ICIP), pp. 59–63. Melbourne, Australia (2013). doi: [10.1109/ICIP.2013.6738013](https://doi.org/10.1109/ICIP.2013.6738013)
16. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004). doi:[10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861)
17. Wang, Z., Simoncelli, E., Bovik, A.: Multi-scale structural similarity for image quality assessment. In: Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers, pp. 1398–1402. Piscataway, USA (2003). doi:[10.1109/ACSSC.2003.1292216](https://doi.org/10.1109/ACSSC.2003.1292216)
18. Damera-Venkata, N., Kite, T., Geisler, W., Evans, B., Bovik, A.: Image quality assessment based on a degradation model. *IEEE Trans. Image Process.* **9**(4), 636–650 (2000). doi:[10.1109/83.841940](https://doi.org/10.1109/83.841940)
19. Wang, Z., Bovik, A.: A universal image quality index. *IEEE Signal Process. Lett.* **9**(3), 81–84 (2002). doi:[10.1109/97.995823](https://doi.org/10.1109/97.995823)
20. Sheikh, H., Bovik, A.: Image information and visual quality. *IEEE Trans. Image Process.* **15**(2), 430–444 (2006). doi:[10.1109/TIP.2005.859378](https://doi.org/10.1109/TIP.2005.859378)
21. Rimac-Drlje, S., Vranjes, M., Zagar, D.: Influence of temporal pooling method on the objective video quality evaluation. In: Proceedings of the 2nd IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '09), pp. 1–5. Bilbao, Spain (2009). doi:[10.1109/ISBMSB.2009.5133781](https://doi.org/10.1109/ISBMSB.2009.5133781)
22. Seufert, M., Slanina, M., Egger, S., Kottkamp, M.: “To pool or not to pool”: A comparison of temporal pooling methods for HTTP adaptive video streaming. In: Proceedings of the 5th International Workshop on Quality of Multimedia Experience (QoMEX 2013), pp. 52–57. Klagenfurt, Austria (2013). doi:[10.1109/QoMEX.2013.6603210](https://doi.org/10.1109/QoMEX.2013.6603210)
23. Wang, Z., Lu, L., Bovik, A.: Video quality assessment based on structural distortion measurement. *Signal Process. Image* **19**(2), 121–132 (2004). doi:[10.1016/S0923-5965\(03\)00076-6](https://doi.org/10.1016/S0923-5965(03)00076-6)
24. Sheikh, H., Bovik, A.: A visual information fidelity approach to video quality assessment. In: Proceedings of the 1st International Workshop on Video Processing and Quality Metrics for Consumer Electronics, pp. 23–25. Scottsdale, USA (2005)
25. Politis, I., Dounis, L., Tselios, C., Kordelas, A., Dagiuklas, T., Papadakis, A.: A model of network related QoE for 3D video. In: Proceedings of the IEEE Global Communications Conference (GLOBECOM 2012), pp. 1335–1340. Anaheim, USA (2012). doi: [10.1109/GLOCOMW.2012.6477776](https://doi.org/10.1109/GLOCOMW.2012.6477776)
26. Xiao, F.: DCT-based video quality evaluation. Technical Report, MSU Graphics and Media Lab (Video Group) (2000). “[http://compression.ru/video/quality\\_measure/vqm](http://compression.ru/video/quality_measure/vqm)”. Accessed 23 Jan 2015
27. Watson, A., Hu, J., McGowan III, J.: DVQ: a digital video quality metric based on human vision. *J. Electron. Imaging* **10**, 20–29 (2001). doi:[10.1117/1.1329896](https://doi.org/10.1117/1.1329896)

28. Pinson, M., Wolf, S.: A new standardized method for objectively measuring video quality. *IEEE Trans. Broadcast.* **50**(3), 312–322 (2004). doi:[10.1109/TBC.2004.834028](https://doi.org/10.1109/TBC.2004.834028)
29. Seshadrinathan, K., Bovik, A.: Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Trans. Image Process.* **19**(2), 335–350 (2010). doi:[10.1109/TIP.2009.2034992](https://doi.org/10.1109/TIP.2009.2034992)
30. MPEVQ, Advanced perceptual evaluation of video quality. Technical Report, OPTICOM GmbH (2005). “<http://www.opticom.de/download/PEVQ-WP-v07-A4>”. Accessed 23 Jan 2015
31. Soundararajan, R., Bovik, A.: Video quality assessment by reduced reference spatio-temporal entropic differencing. *IEEE Trans. Circuits Syst. Video Technol.* **23**(4), 684–694 (2013). doi:[10.1109/TCSVT.2012.2214933](https://doi.org/10.1109/TCSVT.2012.2214933)
32. Narwaria, M., Lin, W., Liu, A.: Low-complexity video quality assessment using temporal quality variations. *IEEE Trans. Multimed.* **14**(3–1), 525–535 (2012). doi:[10.1109/TMM.2012.2190589](https://doi.org/10.1109/TMM.2012.2190589)
33. Yang, F., Song, J., Wan, S., Wu, H.: Content-adaptive packet-layer model for quality assessment of networked video services. *IEEE J. Sel. Top. Signal Process.* **6**(6), 672–683 (2012). doi:[10.1109/JSTSP.2012.2207705](https://doi.org/10.1109/JSTSP.2012.2207705)
34. Joskowicz, J., Sotelo, R., López-Ardao, J.: Towards a general parametric model for perceptual video quality estimation. *IEEE Trans. Broadcast.* **59**(4), 569–579 (2013). doi:[10.1109/TBC.2013.2277951](https://doi.org/10.1109/TBC.2013.2277951)
35. Mohamed, S., Rubino, G.: A study of real-time packet video quality using random neural networks. *IEEE Trans. Circuits Syst. Video Technol.* **12**(12), 1071–1083 (2002). doi:[10.1109/TCSVT.2002.806808](https://doi.org/10.1109/TCSVT.2002.806808)
36. Khan, A., Sun, L., Ifeachor, E., Fajardo, J., Liberal, F.: Impact of RLC losses on quality prediction for H.264 video over UMTS networks. In: Proceedings of the 11th IEEE International Conference on Multimedia and Expo (ICME), pp. 702–707. Singapore (2010). doi: [10.1109/ICME.2010.5583033](https://doi.org/10.1109/ICME.2010.5583033)
37. Mok, R., Chan, E., Chang, R.: Measuring the quality of experience of HTTP video streaming. In: Proceedings of the 12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011), pp. 485–492. Dublin, Ireland (2011). doi: [10.1109/INM.2011.5990550](https://doi.org/10.1109/INM.2011.5990550)
38. Reibman, A., Vaishampayan, V., Sermadevi, Y.: Quality monitoring of video over a packet network. *IEEE Trans. Multimed.* **6**(2), 327–334 (2004). doi:[10.1109/TMM.2003.822785](https://doi.org/10.1109/TMM.2003.822785)
39. Babu, R., Bopardikar, A., Perkis, A., Hillestad, O.: No-reference metrics for video streaming applications. In: Proceedings of the 14th International Workshop on Packet Video (PV 2004). Irvine, USA (2004)
40. Venkataraman, M., Chatterjee, M., Chattopadhyay, S.: Evaluating quality of experience for streaming video in real time. In: Proceedings of the 24th IEEE Global Communications Conference (GLOBECOM 2009), pp. 1–6. Honolulu, USA (2009). doi:[10.1109/GLOCOM.2009.5425806](https://doi.org/10.1109/GLOCOM.2009.5425806)
41. Laghari, K., Pham, T., Nguyen, H., Crespi, N.: QoM: A new quality of experience framework for multimedia services. In: Proceedings of the 17th IEEE Symposium on Computers and Communications (ISCC 2012), pp. 851–856. Cappadocia, Turkey (2012). doi: [10.1109/ISCC.2012.6249408](https://doi.org/10.1109/ISCC.2012.6249408)
42. Staelens, N., Van Wallendael, G., Crombecq, K., Vercammen, N., De Cock, J., Vermeulen, B., Van de Walle, R., Dhaene, T., Demeester, P.: No-reference bitstream-based visual quality impairment detection for high definition H.264/AVC encoded video sequences. *IEEE Trans. Broadcast.* **58**(2), 187–199 (2012). doi:[10.1109/TBC.2012.2189334](https://doi.org/10.1109/TBC.2012.2189334)
43. Chan, A., Zeng, K., Mohapatra, P., Lee, S., Banerjee, S.: Metrics for evaluating video streaming quality in lossy IEEE 802.11 wireless networks. In: Proceedings of the 29th IEEE International Conference on Computer Communications (INFOCOM 2010), pp. 1613–1621. San Diego, USA (2010). doi:[10.1109/INFCOM.2010.5461979](https://doi.org/10.1109/INFCOM.2010.5461979)
44. McDonagh, P., Pande, A., Murphy, L., Mohapatra, P.: Toward deployable methods for assessment of quality for scalable IPTV services. *IEEE Trans. Broadcast.* **59**(2), 223–237 (2013). doi:[10.1109/TBC.2013.2255776](https://doi.org/10.1109/TBC.2013.2255776)
45. Subjective video quality assessment methods for multimedia applications (ITU-T P.910 Recommendation). Technical Report, International Telecommunication Union (ITU) (2009). “<http://www.itu.int/rec/T-REC-P.910-200804-1>”. Accessed 23 Jan 2015
46. Gilbert, E.: Capacity of a burst-noise channel. *Bell Syst. Tech. J.* **39**(5), 1253–1265 (1960). doi:[10.1002/j.1538-7305.1960.tb03959.x](https://doi.org/10.1002/j.1538-7305.1960.tb03959.x)
47. Elliott, E.: Estimates of error rates for codes on burst-noise channels. *Bell Syst. Tech. J.* **42**(5), 1977–1997 (1963). doi:[10.1002/j.1538-7305.1963.tb00955.x](https://doi.org/10.1002/j.1538-7305.1963.tb00955.x)
48. Salsano, S., Ludovici, F., Ordine, A., Giannuzzi, D.: Definition of a general and intuitive loss model for packet networks and its implementation in the Netem module in the Linux kernel. Technical Report, University of Rome (2012). “<http://netgroup.uniroma2.it/TR/TR-loss-netem>”. Accessed 23 Jan 2015

49. Reference algorithm for computing peak signal to noise ratio of a processed video sequence with compensation for constant spatial shifts, constant temporal shift, and constant luminance gain and offset (ITU-T J.340 Recommendation). Technical Report, International Telecommunication Union (ITU) (2010). "[https://www.itu.int/rec/dologin\\_pub.asp?lang=e&id=T-REC-J.340-201006-I!!PDF-E&type=items](https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-J.340-201006-I!!PDF-E&type=items)". Accessed 23 Jan 2015
50. VLC media player. "<http://www.videolan.org/vlc/>". VideoLAN Organization. Accessed 23 Jan 2015
51. García, P.: Python visual quality metric package (PyMetrikz). "<https://bitbucket.org/kuraiev/pymetrikz>". University of Brasilia. Accessed 23 Jan 2015
52. Moskow State University (MSU) Video quality measurement tool. "[http://compression.ru/video/quality\\_measure/video\\_measurement\\_tool\\_en.html](http://compression.ru/video/quality_measure/video_measurement_tool_en.html)". MSU Graphics and Media Lab (Video Group). Accessed 23 Jan 2015
53. Visual quality metric (VQM). "<http://www.its.bldrdoc.gov/resources/video-quality-research/software.aspx>". Institute of Telecommunication Sciences (ITS), US National Telecommunication and Information Administration (NTIA). Accessed 23 Jan 2015
54. Test video sequences. "<http://www.atc.uniovi.es/personal/fran/papers/SuarezFJ13-videos.rar>". Computer Engineering Group at University of Oviedo. Accessed 23 Jan 2015

**Francisco J. Suárez** is a professor in the Department of Computer Science and Engineering at the University of Oviedo, Spain, where he received his Ph.D. degree in 1998. His current research focuses on evaluation of networked multimedia systems and he has been working on several related projects as head of the Interactive Multimedia Services research team.

**Andrés García** is a software engineer and data analyst at labSens company in Spain. He received his M.Sc. degree in Computer Science from the University of Oviedo, Spain, in 2010. His current work focuses on the design and development of tools to analyze and improve the performance of networks and industrial processes.

**Juan C. Granda** is an associate professor in the Department of Computer Science and Engineering at the University of Oviedo, Spain, where he received his Ph.D. degree in Computer Science in 2008. He has been involved in the research activities of the Computer Engineering Area, focusing on performance analysis of synchronous e-learning systems.

**Daniel F. García** is a full professor in the Department of Computer Science and Engineering at the University of Oviedo, Spain, where he leads the Computer Engineering Area. His current research focuses on self-adaptive multimedia and real-time systems. He has led many research projects during the last 20 years and coauthored more than 50 articles in journals and more than 130 papers in conferences and workshops. He is a member of the IEEE Computer Society.

**Pelayo Nuño** received his Ph.D. degree in Computer Science from the University of Oviedo, Spain, in 2013. He has been involved in several research activities of the Computer Engineering Area. His research interests include multimedia networking and autonomic computing.