EURASIP Journal on Image and Video Processing
a SpringerOpen Journal

RESEARCH Open Access

Full-reference video quality metric assisted the development of no-reference bitstream video quality metrics for real-time network monitoring

Iñigo Sedano[1*], Kjell Brunnström[2,4], Maria Kihl[3] and Andreas Aurelius[2,3]

Abstract

High-quality video is being increasingly delivered over Internet Protocol networks, which means that network operators and service providers need methods to measure the quality of experience (QoE) of the video services. In this paper, we propose a method to speed up the development of no-reference bitstream objective metrics for estimating QoE. This method uses full-reference objective metrics, which makes the process significantly faster and more convenient than using subjective tests. In this process, we have evaluated six publicly available full-reference objective metrics in three different databases, the EPFL-PoliMI database, the HDTV database, and the Live Video Wireless database, all containing transmission distortions in H.264 coded video. The objective metrics could be used to speed up the development process of no-reference real-time video QoE monitoring methods that are receiving great interest from the research community. We show statistically that the full-reference metric Video Quality Metric (VQM) performs best considering all the databases. In the EPFL-PoliMI database, SPATIAL MOVIE performed best and TEMPORAL MOVIE performed worst. When transmission distortions are evaluated, using the compressed video as the reference provides greater accuracy than using the uncompressed original video as the reference, at least for the studied metrics. Further, we use VQM to train a lightweight no-reference bitstream model, which uses the packet loss rate and the interval between instantaneous decoder refresh frames, both easily accessible in a video quality monitoring system.

Keywords: H.264; Full reference; Objective metrics; Video quality; Transmission distortions

1. Introduction

Streaming high-quality digital video over Internet Protocol (IP)-based networks is increasing in popularity both among users and operators. Two examples of these applications are *IPTV* and *Over The Top (OTT) Video* [1]. IPTV systems are managed by one operator, from video head-end to the user, and are based on ordinary broadcast television, using IP multicast. OTT Video is used to describe the delivery of TV over the public Internet, using unicast.

In order to ensure a high Quality of Experience (QoE), the network operators and service providers need methods to monitor the quality of the video services [2]. The monitoring and prediction should be performed in real-time and in different parts of the network. Since users' experienced quality is not easily understood and depends on many aspects [3], subjective assessments involving a panel of observers constitutes the most accurate method to measure the video QoE. However, in a monitoring situation, subjective assessments are very hard to perform and therefore objective measurement methods are desirable. Even for the development of these measurement methods subjective data is usually required, which may be cumbersome and time consuming to obtain when developing real-time monitoring systems. Furthermore, for subjective testing to be accurate, it requires careful planning, preparation and involvement of a number of viewers. This makes it costly to conduct.

Instead, objective metrics, which accurately characterize the video quality and predict viewer quality of experience, have evolved for some time now, but there is still a long way to go before they, in general, can accurately predict the results of subjective measurements [4]. The objective

* Correspondence: inigo.sedano@tecnalia.com
[1]TECNALIA, ICT - European Software Institute, Parque Tecnológico de Bizkaia, Edificio 202, Zamudio E-48170, Spain
Full list of author information is available at the end of the article

metrics can be classified as no reference, reduced reference, and full reference [5]. Traditionally, in the full-reference scenario, an original undistorted high-quality video is compared to a degraded version of the same video, for example, pixel by pixel or block based. Reduced-reference methods require partial or parameterized information about the original video sequence. No-reference methods rely only on the degraded video. Here, we generalize the concept of FR, RR, and NR by also including packet header models, bitstream models, and hybrid models together with the pure video-based models based on the amount of reference information used by the models, as suggested in Barkowsky et al. [6].

Objective video quality metrics are usually argued to be useful because subjective quality assessment is expensive and time consuming to perform. However, in the development of the objective metrics, subjective data is essential to train, optimize, and evaluate these metrics. Therefore, it would be advantageous and shorten the development time, if it was possible to use well performing objective quality metrics in the development process of new video quality metrics. The purpose of this paper is to show a cost- and time-effective development strategy for computationally efficient light weight no-reference bitstream video quality metrics. Here, the scope of model, meaning the area in which it is valid, is an additional important parameter, for example, see [6]. This is especially true for NR models where it is harder to develop good performing models with broad scopes. Therefore, by limiting the scope of a NR video quality metric, it is possible to achieve high prediction performance using a limited number of parameters. The drawback is that the usage should be within the scope that it was designed for. However, the strategy is then to redesign the model using the same method again to tailor it for the new application area. Specifically, the proposed method is to first find a full-reference metric that performs well for the types of distortions we are interested in, and then we use this metric to develop a no-reference metric. This could be summarized in a four step procedure as shown in the Figure 1 in a bit more detail.

1. Definition of scope
2. Find FR model for the scope
3. Train NR mode using FR model
4. Evaluate the performance of NR model

To illustrate the methodology, we have selected a concrete example, where we perform all the necessary steps for the procedure described above. This does not mean that we claim that this is the first time FR models are evaluated for packet loss, but to our knowledge, it has not been done for this particular scope, i.e., packet loss combined with coded reference. Still, it is presented for illustrating the methodology. In case it is already known which FR model that is best for a particular scope, then this step can be omitted. There may also be better NR models also for this scope, but it should be taken into account the relatively high performance combined with its simplicity of model and most importantly the low development effort.

As an example in this paper, see Figure 2, we first evaluate and select the best full-reference metric for transmission distortions in the case of compressed reference, which is a very specific scope. Then, we create a training database and we execute on it the selected full-reference metric. Finally, we model and validate the no-reference bitstream model. The created model will therefore be valid only in the scope of evaluating only transmission distortions. However, it can be redesigned following the same procedure in order to take into account other types of distortions, such as compression distortions. Here, we show the necessary steps in order to extend the model that we present in this paper. It would be necessary first to find the most suitable already existing objective metric to measure compression distortions, create a new training

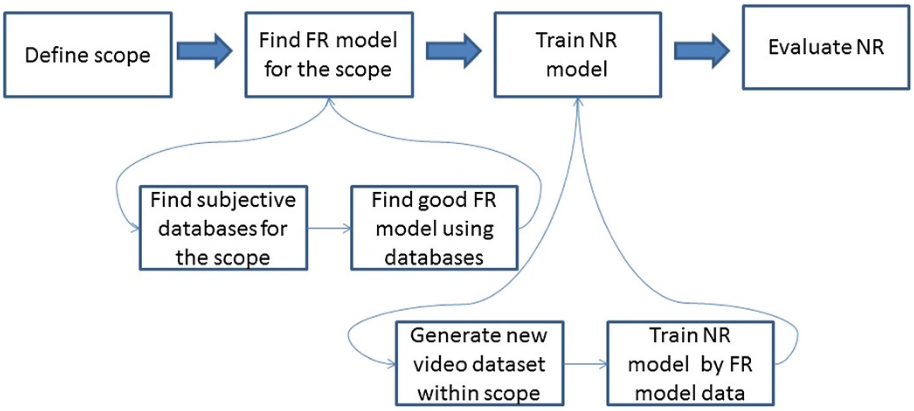
Define scope
Find FR model for the scope
Train NR model
Evaluate NR
Find subjective databases for the scope
Find good FR model using databases
Generate new video dataset within scope
Train NR model by FR model data

Figure 1 Proposed methodology to develop no-reference models.

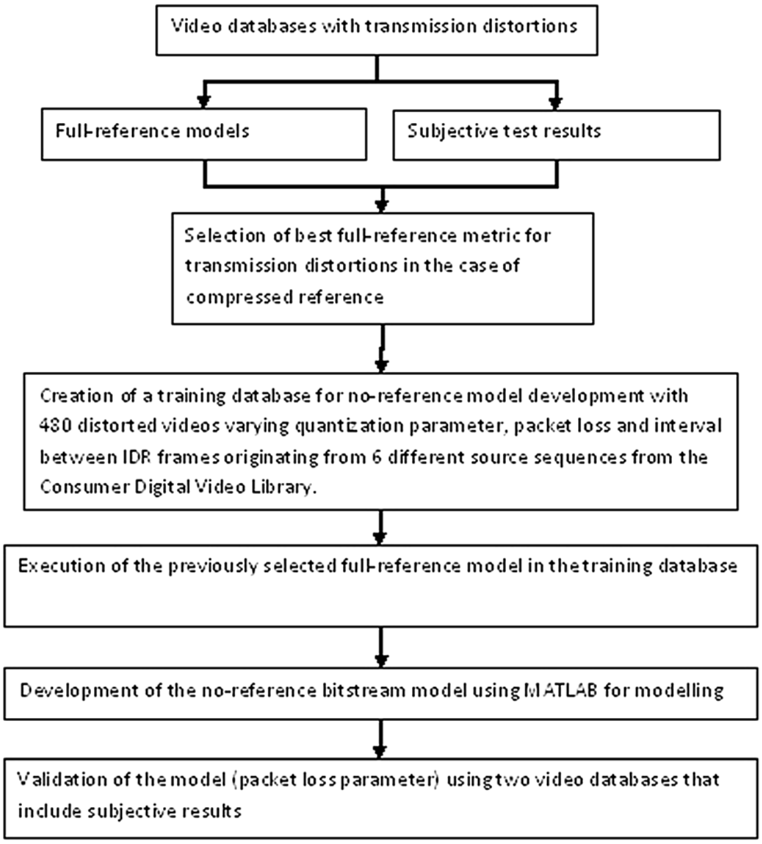
Video databases with transmission distortions
Full-reference models
Subjective test results
Selection of best full-reference metric for transmission distortions in the case of compressed reference
Creation of a training database for no-reference model development with 480 distorted videos varying quantization parameter, packet loss and interval between IDR frames originating from 6 different source sequences from the Consumer Digital Video Library.
Execution of the previously selected full-reference model in the training database
Development of the no-reference bitstream model using MATLAB for modelling
Validation of the model (packet loss parameter) using two video databases that include subjective results

Figure 2 **Proposed methodology to develop no-reference models considering packet loss.**

database varying all the parameters of the extended model (model with compression and transmission distortions), execute the selected full-reference metric for compression distortions on the training database, and finally redesign the model incorporating the compression dependent parameters.

2. Background

Traditionally, video quality metrics make predictions on computations on the video data itself. Nowadays, there are emerging models also utilizing network information either by itself, i.e., bitstream models or in combination with the video data, i.e., hybrid models. A review of no-reference video quality estimators can be found in [7].

Most of the objective metrics have been developed and tested to estimate the perceived quality when the video is only compression degraded, for example, see [8,9]. However, video delivered over Internet will be degraded by transmission distortions, for example, packet loss. Several studies have shown that even a low packet loss rate can and most often will affect the video quality, for example, see [10].

Also, for objective quality monitoring, there are two other aspects of performance apart from prediction accuracy that are important: computational requirements and run time [11]. To be used in a real-time monitoring and prediction system, the objective quality methods must be lightweight and cannot require the original video reference [12,13]. Independent evaluations are scarce with the notable exception of the work by the Video Quality Experts Group (VQEG) [14]. One of the problems a developer or a tester must face is the unavailability of video databases, especially if they contain videos subject to packet losses. Also, in real network deployments, the uncompressed original sequence is usually not available. Therefore, we believe that it is important to evaluate the performance of the metrics when a compressed reference is used instead. For example, the video quality degradation introduced by a network node could be evaluated applying a full-reference metric such as VQM comparing the compressed reference that is available before the video enters the network node and the degraded video due to transmission distortions that is available after the video exits the network node.

The paper starts by describing the publicly available video databases that have been used for the development work. The details about the video sequences and how they have been compressed and distorted are described.

Also, the subjective tests that have been performed with the aforementioned video sequences are described in terms of number of viewers, viewing conditions, etc.

In the following section, the objective full-reference assessment algorithms are reviewed, and the scenarios in which they are used are outlined.

The following section contains the results of the evaluation of the objective full-reference assessment algorithms against the databases with subjective test data. The means of evaluation are the Spearman Rank Order Correlation Coefficient (SROCC), the Pearson correlation coefficient, the Root Mean Square Error (RMSE), and the Outlier Ratio (OR).

In the last section of the paper, we show how a no-reference objective model can be developed by training it against a full-reference objective metric. Naturally, we choose the metric with the best performance, as evaluated in the previous section.

3. Video databases

In the paper, we evaluate some full-reference objective metrics against three different publicly available databases with subjective quality ratings, based on H.264 coded videos that also contained transmission distortions. The transmission in the databases was handled by RTP over UDP (Real-time Transport Protocol over User Datagram Protocol), and the distortions were in the form of packet losses. This means that the work in this paper is applicable to cases with these conditions, such as, e.g., IPTV cases. This is an extention of the study done in [15]. The databases considered are EPFL-PoliMI (Ecole Polytechnique Fédérale de Lausanne and Politecnico di Milano) video quality assessment database [16-18], an HDTV video database made available by IRCCyN [19] and the LIVE Wireless Video Quality Assessment database [20]. In addition to the following descriptions, Table 1 summarizes the parameters corresponding to the three databases.

3.1 EPFL-PoliMI video database

3.1.1 Description

The freely available EPFL-PoliMI (Ecole Polytechnique Fédérale de Lausanne and Politecnico di Milano) video quality assessment database [16-18] was specifically designed for the evaluation of transmission distortions. The database contains 78 video sequences at 4CIF spatial resolution (704 × 576 pixels). The distorted videos were created from five 10-s-long and one 8-s-long uncompressed video sequences in planar I420 raw progressive format [21].

The reference videos were lightly compressed to ensure high video quality in the absence of packet losses. Therefore, a fixed Quantization Parameter between 28 and 32 was selected for each sequence. The Quantization Parameter regulates how much spatial detail is saved. The sequences were encoded and decoded in H.264/AVC [22] High Profile in the H.264/AVC reference software. B-pictures and Context-adaptive binary arithmetic coding (CABAC) were enabled for coding efficiency. Each frame

Table 1 Summary of conditions of all databases

	EPFL-PoliMI	HDTV video database	LIVE Wireless video database
Number of sequences	Total 78, 6 different source sequences	Total 45, 9 different source sequences in compressed and uncompressed formats	Total 170, 10 different source sequences
Resolution	4CIF (704 × 576)	1,920 × 1,080	768 × 480
Duration	10 and 8 s	10 s	10 s
Reference	Compressed (with high quality)	Compressed (with high quality) and uncompressed	Compressed (with high quality)
Compression parameter	Fixed QP between 28 and 32	Fixed QP value 26	Reference video: Fixed QP value 18
			Degraded videos: bitrates 500 kbps, 1 Mbps, 1.5 Mbps and 2 Mbps
I-frame period	Not available	24 frames	Reference video: 14 frames
			Degraded videos: 96
Frame rate	25 and 30 fps	59,94 (interlaced) fps	30 fps
Transmission distortions	PLR 0.1%, 0.4%, 1%, 3%, 5%, 10%. Two different channel realizations for each PLR	PLR 0.7% (from 42% to 56% of the way), 4.2% (from 21% to 64% of the way), 4.2% (from 42% to 56% of the way).	PLR 0.5%, 2%, 5% and 17%
Encoder/decoder	H.264/AVC JM reference software	Not available	H.264/AVC JM reference software
Number of subjects	21 at PoliMI lab and 19 at EPFL lab	24	31
Processing of subjective scores	Difference scores → Z-scores (with outliers detection) → re-scaling to range [0,5] → DMOS	Difference scores → Z-scores (with outliers detection) → re-scaling to range [0,5] → DMOS both for compressed and uncompressed reference	Difference scores → DMOS → re-scaling to range [0,5]

was divided into a fixed number of slices, where each slice consists of a full row of macroblocks.

The compressed videos in the absence of packet losses were used as the reference for the computation of the DMOS (Differential Mean Opinion Score) values. Three of the reference videos have a frame rate of 25 frames per second (fps). This was accomplished by cropping HD resolution video sequences down to 4CIF resolution and reducing the frame rate from 50 to 25 fps. The other three videos have a frame rate of 30 fps.

The transmission distortions were simulated at different packet loss rates (PLR) (0.1%, 0.4%, 1%, 3%, 5%, 10%). The packet loss was generated using a two-state Gilbert's model with an average burst length of three packets and two different channel realizations were selected for each PLR.

Forty naive subjects took part in the subjective tests. The subjective evaluation was done using the ITU continuous scale in the range [0–5] [23]. Twenty-one subjects participated in the evaluation at the PoliMI lab and 19 at the EPFL lab. More details about the subjective evaluation can be found in [16-18].

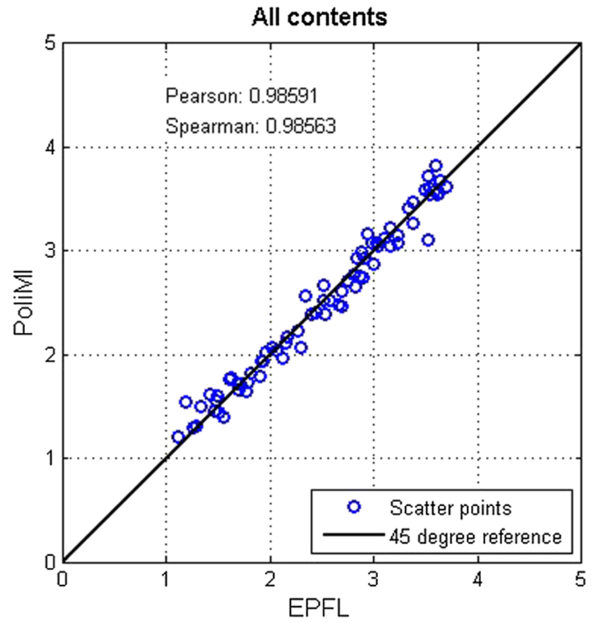

Figure 3 **Scatter plot showing the correlation between the DMOS values obtained in the EPFL lab and in the PoliMI lab.**

3.1.2 Processing of subjective scores

Although the raw subjective scores were already processed in the EPFL-PoliMI database, we processed them in a different way in order to merge the data from the two labs.

A Student *T* test considering the overall mean and standard deviation of the raw MOS individual scores of each lab showed that at 95% confidence level the data from the two labs were not significantly different, and therefore, we decided that they could be merged. As an additional verification, the DMOS and confidence interval values (in this case after normalization, screening, and re-scaling) were calculated for each content and distortion type and compared between the two labs, confirming that the data from the two labs were sufficiently similar to be merged. Seventy-two PVS were checked corresponding to six different packet loss rate, two different channel realizations for each PLR, for each of the six source sequences. In the scatter plot in Figure 3, it can be seen that the linear correlation between the DMOS values obtained in the PoliMI lab and the EPFL lab is high (0.986).

First of all, we calculated the difference scores by subtracting the scores of the degraded videos to the score of the reference videos. The difference scores for the reference videos were set to 0 and were removed. Accordingly, a lower difference score indicates a higher quality.

Each subject may have used the rating scale differently and with different offset. In order to account for this, the Z-scores were computed for each subject separately by means of the Matlab zscore function. The Z-scores transform the original distribution to one in which the mean becomes zero and the standard deviation becomes one. Indeed, this normalization procedure reduces the gain and offset between the subjects. Subsequently, the outliers were detected according to the guidelines described in ITU-T Rec 910 Annex 2 Section 2.3.1 [23] and removed.

Next, the Z-scores were re-scaled to the range [0,5]. The Z-scores are assumed to be distributed as a standard Gaussian. Consequently, 99% of the scores will be in the range [−3,3]. In our study, 100% of the scores were placed in that range. All the data was in fact in the range [−3,3] so no clipping was done. The re-scaling was performed by linearly mapping the data range [−3,3] to the range [0,5] using the following formula:

$$z' = \frac{5.(z+3)}{6}$$

Finally, the Difference Mean Opinion Score (DMOS) of each video was computed as the mean of the re-scaled Z-scores from the 36 subjects that remained after rejection. Additionally, the confidence intervals were also computed. The methodology for the processing of the scores shown in this paper has been applied by many authors. For example, see [24].

3.2 HDTV video database

The HDTV video database was made freely available by Barkowsky et al. [19]. The video database contains nine different source video sequences, and we selected three different conditions corresponding only to transmission distortions. In [19], these are referred to as the Hypothetical Reference Circuit (HRC) 5, 6, and 7. HRCs 5 to 7 are coded with high quality (QP26) and contain simulated transmission errors, mainly blurriness and motion artifacts. The errors were inserted in the middle of the video sequence. In HRC 5, from 42% to 56% of the way through the 14-s sequence's bitstream (before removing the beginning and end of the sequence), 0.7% of packets were randomly lost. HRC 6 contained 4.2% of packets randomly lost from 21% to 64% of the way through the bitsream. HRC 7 contained 4.2% of packets randomly lost from 42% to 56% of the way through the bitstream.

The encoder always used two interlaced slice groups of two macroblock lines. For error recovery, an intra image was forced every 24 frames and the ratio of intra macroblock refresh was 5%. The video resolution was 1,920 × 1,080 pixels at 59.94 fields-per-second in interlaced format. The sequences have a duration of 10 s. In total, 24 naive observers viewed the content. The Absolute Category Rating with Hidden Reference (ACR-HR) conforming to ITU-T P.910 with a five-point rating scale was used. The subjects viewed the content at a distance of 1.5 m corresponding to three times the picture height. More details about the subjective experiment can be found in [19].

The processing of the subjective scores was performed in the same way as for the EPFL-PoliMI video database. The DMOS values were calculated both for the scenario with compressed reference (QP26, HRC1) and with uncompressed reference (HRC0). Two outliers were found in the case of compressed reference and no outliers in the case of uncompressed reference.

3.3 LIVE Wireless video database

Moorthy et al. [20] evaluated publicly available full-reference video quality assessment algorithms on the LIVE Wireless Video Quality Assessment database. The LIVE Wireless video database contains ten source sequences, each 10 s long at a rate of 30 frames per second. The source videos are in RAW uncompressed progressive scan YUV420 format with a resolution of 768 × 480. However, the videos used as reference were already compressed with high quality (average PSNR > 45 dB). For the reference sequences, the Quantization Parameter was set to 18 and the I-frame period to 14. One-hundred sixty distorted videos were created (4 bitrates × 4 packet loss rates = 16 distorted videos per reference sequence). The simulated wireless transmission errors were inserted to the H.264 compressed videos, which were generated with the JM reference software (Version 13.1). The source videos were encoded using different bitrates: 500 kbps, 1 Mbps, 1.5 Mbps, and 2 Mbps with three different slice groups and an I-frame period of 96. The RD Optimization was enabled, and the baseline profile was used for encoding and hence did not include B-frames. The packet size was set to between 100 and 300 bytes. The Flexible Macroblock Ordering (FMO) mode was set as 'dispersed'.

Packet loss rates of 0%, 5%, 2%, 5%, and 17% were simulated using bit-error patterns captured from different real or emulated mobile radio channels. The JM reference software was used to decode the compressed video stream.

For the subjective test, the Single Stimulus Continuous Quality Evaluation with hidden reference was used. A total of thirty-one subjects participated in the study. The difference scores were calculated by subtracting the score that the subject assigned to the distorted sequence to the score that the subject assigned to the reference sequence. One subject was rejected. The scores from the remaining subjects were then averaged to form a Differential Mean Opinion Score (DMOS) for each sequence. No Z-scores were used. Finally, we re-scaled the DMOS values to the range [0–5]. More details on the subjective study can be found on [20]. The LIVE Wireless video database is no longer publicly available because of the uniformity and simplicity of the content. However, we use this database because our study involves various video databases.

4. Objective assessment algorithms

The video quality metrics that were evaluated are the following well-known publicly available algorithms: Peak Signal-to-Noise Ratio (PSNR) [4], Structural SIMilarity (SSIM) index [25], Multi-scale SSIM (MS-SSIM) [26], Video Quality Metric (VQM) [27], Visual Signal to Noise Ratio (VSNR) [28], and MOtion-based Video Integrity Evaluation (MOVIE) [29]. The performance of the objective models is evaluated using the Spearman Rank Order Correlation Coefficient, the Pearson Linear Correlation Coefficient, the Root-Mean-Square Error (RMSE) and the Outlier Ratio. A non-linear regression was done using a monotonic function. The performance of the different metrics was compared by means of a statistical significance analysis based on the Pearson, RMSE, and Outlier Ratio coefficients.

4.1 Scenarios

The typical full-reference scenario is shown in Figure 4. The original uncompressed video is compared to the uncompressed video that contains the compression and transmission distortions.

In this paper, we also consider the scenario shown in Figure 5 that corresponds to compressed reference. The reference videos are lightly compressed to ensure high video quality in the absence of packet losses. The references are thus similar in quality to the uncompressed original. Therefore, in the compressed reference scenario, the video is first compressed before being used in the evaluation.

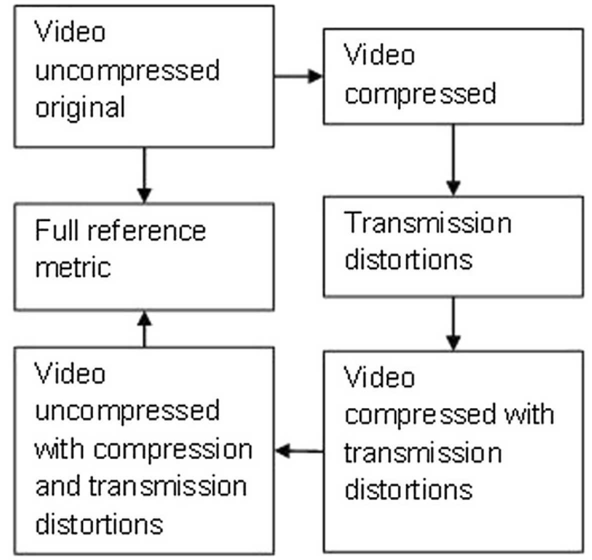
Video uncompressed original
Video compressed
Full reference metric
Transmission distortions
Video uncompressed with compression and transmission distortions
Video compressed with transmission distortions

Figure 4 Full-reference scenario with original uncompressed reference.

The decompressed video with compression distortions is compared to the decompressed video with compression and transmission distortions.

4.2 Video quality algorithms

We have evaluated and compared several well-known objective video quality algorithms using the videos and subjective results in the three databases. The objective algorithms are described below. The default values of the metrics were used for all the metrics. No registration problems, i.e., a misalignment between the reference and degraded videos due to the loss of entire frames, occurred in the dataset.

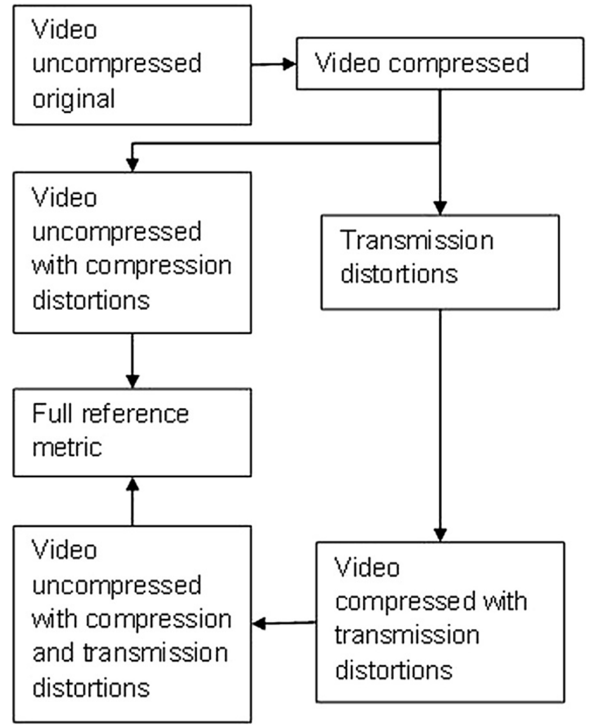
Video uncompressed original
Video compressed
Video uncompressed with compression distortions
Transmission distortions
Full reference metric
Video uncompressed with compression and transmission distortions
Video compressed with transmission distortions

Figure 5 Full-reference scenario with compressed reference.

4.2.1 Peak Signal-to-Noise Ratio

PSNR is computed using the mean of the MSE vector (contains the Mean Square Error of each frame). The MSE is computed per frame. The implementation used is based on the 'PSNR of YUV videos' program (yuvpsnr.m) by Dima Pröfrock available in the MATLAB Central file repository [30]. Only the luminance values were considered.

4.2.2 Structural SIMilarity

SSIM [25] is computed for each frame. After that an average value is produced. The implementation used is an improved version of the original version [25] in which the scale parameter of SSIM is estimated. The implementation, named ssim.m, can be downloaded in the author's implementation home page [31]. Only the luminance values were considered.

4.2.3 Multi-scale SSIM

MS-SSIM [26] is computed for each frame. Afterwards, an average value is produced. The implementation used was downloaded from the Laboratory for Image & Video Engineering (LIVE) at the University Of Texas at Austin [32]. Only the luminance values were considered.

4.2.4 Video Quality Metric

For VQM, we used the software version 2.2 for Linux that was downloaded from the author's implementation home page [27]. We used the following parameters: parsing type none, spatial, valid, gain and temporal calibration automatic, temporal algorithm sequence, temporal valid uncertainty false, alignment uncertainty 15, calibration frequency 15, and video model general model. The files were converted from planar 4:2:0 to the format required by VQM (Big-YUV file format, 4:2:2) using ffmpeg.

4.2.5 Visual Signal-to-Noise Ratio

VSNR [28] is computed using the total signal and noise values of the sequence. We modified the authors' implementation available at [33] to extract the signal and noise values in order to sum them separately. Only the luminance values were considered. The VSNR was obtained dividing the total amount of signal by the total amount of noise.

Table 2 EPFL-POLIMI video database

	Pearson	Spearman	RMSE	Outlier ratio
PSNR	0.958	0.961	0.219	0.625
SSIM	0.959	0.969	0.217	0.597
MS-SSIM	0.964	0.978	0.204	0.597
VSNR	0.974	0.973	0.173	0.472
VQM	0.961	0.960	0.210	0.541
MOVIE	0.965	0.962	0.202	0.625
SPATIAL MOVIE	0.981	0.978	0.148	0.458
TEMPORAL MOVIE	0.924	0.914	0.294	0.611

4.2.6 MOtion-based Video Integrity Evaluation

MOVIE [29] includes three different versions: the Spatial MOVIE index, the Temporal MOVIE index and the MOVIE index. The MOVIE Index version 1.0 for Linux was used and can be downloaded from [32]. The optional parameters framestart, frameend, or frameint were not used. Only EPFL-PoliMI was analyzed with MOVIE.

4.3 Statistical analysis

In order to test the performance of the objective algorithms, we computed the Spearman Rank Order Correlation Coefficient (SROCC), the Pearson correlation coefficient, the RMSE, and the Outlier Ratio (OR) [34]. The Spearman coefficient assesses how well the relationship between two variables can be described using a monotonic function. The Pearson coefficient measures the linear relationship between a model's performance and the subjective data. The RMSE provides a measure of the prediction accuracy. Finally, the consistency attribute of the objective metric is evaluated by the Outlier Ratio.

The Pearson, RMSE, and Outlier Ratio were computed after a non-linear regression. In the analysis of the EPFL-PoliMI video database, the regression was performed using a monotonic cubic polynomial function with four parameters. The function is constrained to be monotonic:

$$\mathrm{DMOSp} = a{\cdot}x^3 + b{\cdot}x^2 + c{\cdot}x + d.$$

In the above equation, the DMOSp is the predicted value. The four parameters were obtained using the MATLAB function ‘nlinfit’.

In the analysis of the other two databases, a monotonic logistic function with four parameters was used instead:

$$\mathrm{DMOSp} = \frac{\beta_1 - \beta_2}{1 + \exp\left(-\frac{x-\beta_3}{|\beta_4|}\right)} + \beta_2$$

In each of the databases, we used the function providing the best fitting. The performance of the metrics is compared by means of a statistical significance analysis

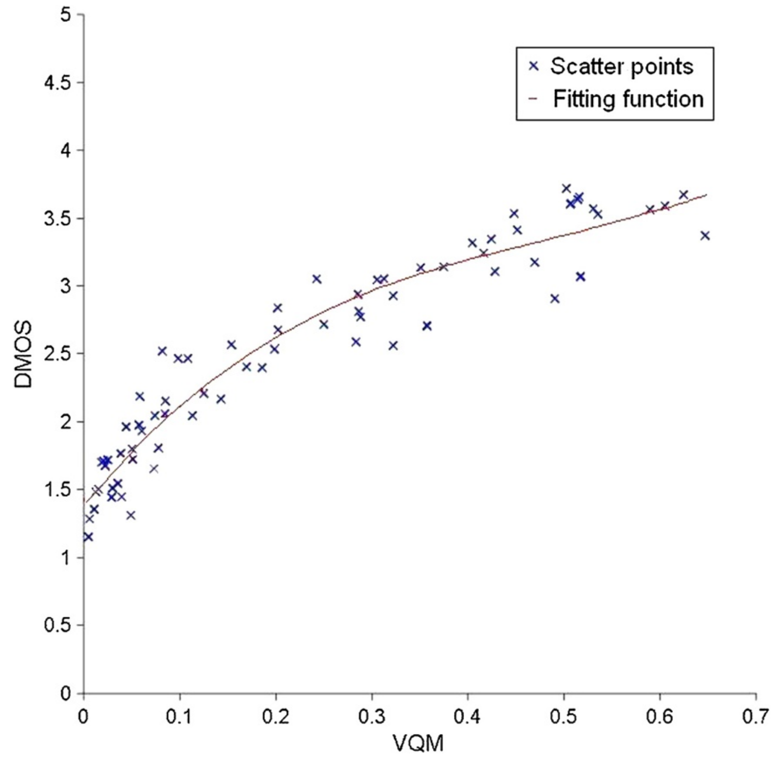

Figure 6 Scatter plot VQM for EPFL-PoliMI.

Table 3 HDTV video database compressed reference

	Pearson	Spearman	RMSE	Outlier ratio
PSNR	0.817	0.804	0.346	0.296
SSIM	0.871	0.856	0.295	0.370
MS-SSIM	0.891	0.884	0.273	0.296
VSNR	0.837	0.774	0.328	0.444
VQM	0.887	0.860	0.277	0.333

based on the Pearson, RMSE, and Outlier Ratio coefficients [34].

5. Evaluation of full-reference objective metrics

In this section, we present the results of the statistical analysis. Also, in several figures, the scatter plots of the VQM objective metric scores vs. DMOS for the different databases are shown. We show the plots of the VQM objective metric because the VQM metric performs very well in all the video databases. The fitting function is also plotted.

5.1 EPFL-PoliMI

In Table 2, the values of the coefficients corresponding to all the metrics for the EPFL-PoliMI video database are shown. The meaning of each coefficient was explained in the previous section. The values for the Pearson correlation coefficient ranged from 0.92 (for TEMPORAL MOVIE) to 0.98 (for SPATIAL MOVIE). The values for the Spearman rank order correlation coefficient were confined within 0.91 (TEMPORAL MOVIE) and 0.98 (SPATIAL MOVIE). Looking also at the RMSE, we can see that the TEMPORAL MOVIE performed significantly worse than the other methods. In general, the magnitude of the coefficients was high and the differences between them were small. The statistical significance analysis based on Pearson and RMSE confirms that at 95% confidence level MS-SSIM, VSNR, VQM, MOVIE, and SPATIAL MOVIE performed better than TEMPORAL MOVIE, being SPATIAL MOVIE the best performing metric.

Further, in Figure 6, the scatter plot of VQM is shown including the fitting function. The horizontal axis corresponds to the values of the VQM metric. The vertical axis corresponds to the DMOS values. A lower DMOS means higher video quality. The fitting function is plotted in circles (one circle per VQM value). In the scatter plot, we can see that the correlation between VQM and DMOS is not linear and that the correlation is very high.

5.2 HDTV video database

In Table 3, the values of the coefficients corresponding to all the metrics for the HDTV video database when the reference is lightly compressed can be observed. It can be seen in the table that the values for the Pearson correlation coefficient were distributed within 0.82 (for PSNR) and 0.89 (for MS-SSIM). The values for the

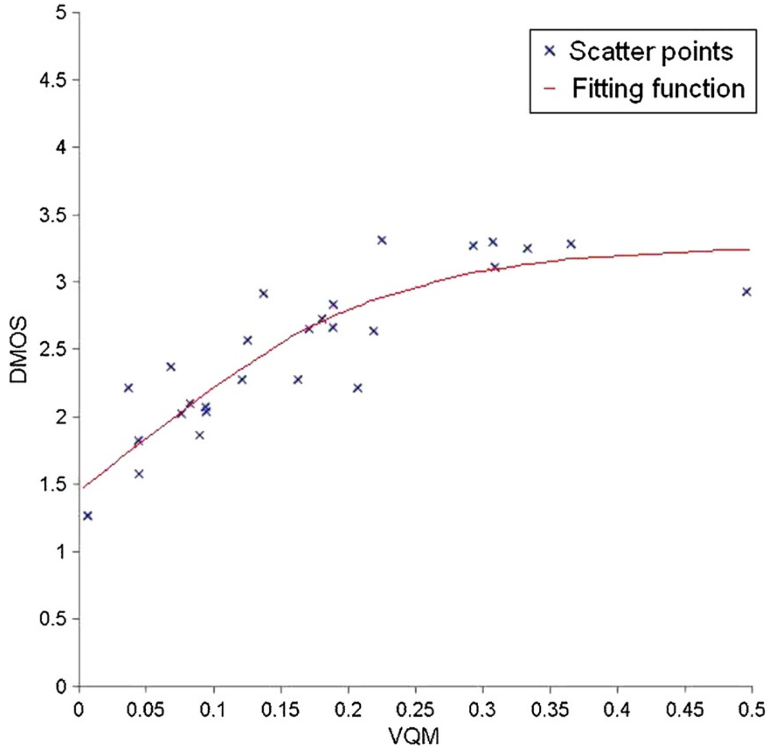

Figure 7 Scatter plot VQM for HDTV video database compressed reference.

Table 4 HDTV video database uncompressed reference

	Pearson	Spearman	RMSE	Outlier ratio
PSNR	0.661	0.600	0.422	0.555
SSIM	0.720	0.653	0.391	0.518
MS-SSIM	0.727	0.664	0.386	0.518
VSNR	0.629	0.511	0.438	0.592
VQM	0.840	0.782	0.305	0.370

Spearman rank order correlation coefficient were confined within 0.80 (for PSNR) and 0.88 (for MS-SSIM). The general magnitude of the coefficients was high. The statistical significance analysis based on Pearson and RMSE shows that at 95% confidence level, there were no significant differences between the studied metrics.

Further, in Figure 7, the scatter plot of VQM using compressed reference is shown including the fitting function. In the scatter plot, we can see that the correlation between VQM and DMOS is not linear and that the correlation is high.

In Table 4, the values of the coefficients corresponding to all the metrics for the HDTV video database when the reference is uncompressed are presented. The values for the Pearson correlation coefficient ranged from 0.63 (for VSNR) to 0.84 (for VQM). The values for the Spearman rank order correlation coefficient had the lowest value at 0.51 (VSNR) and the highest at 0.78 (VQM). The general magnitude of the coefficients was low. The statistical significance analysis based on RMSE shows that at 95% confidence level, VQM performed better than VSNR.

Further, in Figure 8, the scatter plot of VQM using uncompressed reference is shown including the fitting function. The correlation between VQM and DMOS is high and not linear.

5.3 Live Wireless database

The coefficients corresponding to the LIVE Wireless database are shown in Table 5. The values for the Pearson correlation coefficient are distributed within 0.93 (for VSNR) and 0.97 (for VQM). The values for the Spearman rank order correlation coefficient are confined within 0.95 (VSNR) and 0.97 (VQM). The general magnitude of the coefficients is very high and the differences between them are small. The statistical significance analysis based on Pearson and RMSE shows that at 95% confidence level VQM performed better than all the other metrics.

Further, in Figure 9, the scatter plot of VQM is shown including the fitting function. In this case, the correlation between VQM and DMOS is approximately linear and very high.

5.4 Discussion

Our results show that VQM has a very good performance in all the databases, being the best metric among

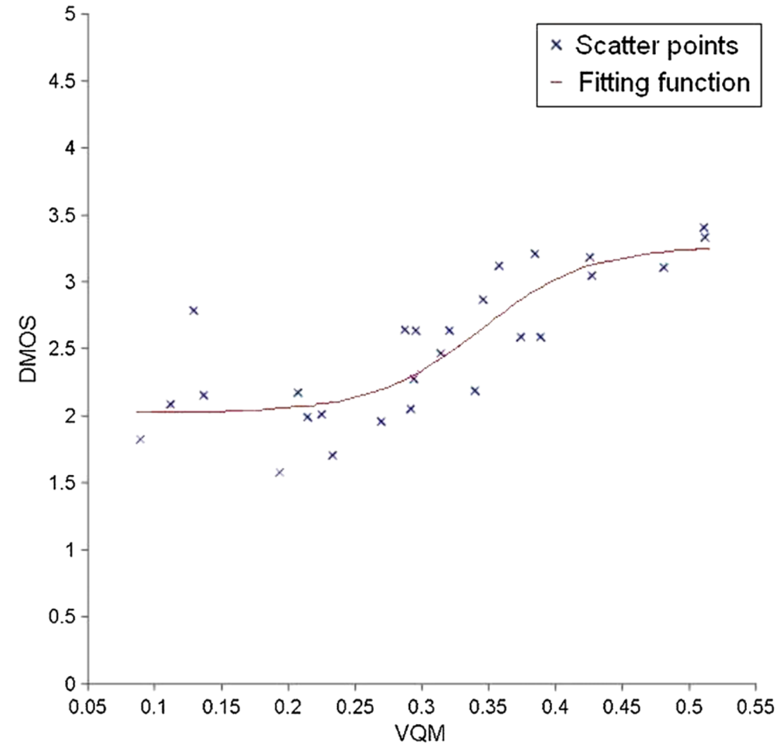

Figure 8 Scatter plot VQM for HDTV video database uncompressed reference.

Table 5 Live Wireless video database

	Pearson	Spearman	RMSE
PSNR	0.959	0.960	0.365
SSIM	0.954	0.954	0.386
MS-SSIM	0.96	0.963	0.364
VSNR	0.949	0.946	0.409
VQM	0.974	0.974	0.294

the studied in the HDTV video database (uncompressed reference) and in the LIVE Wireless video database. In the EPFL-PoliMI video database, SPATIAL MOVIE performed better than the other metrics. On the other hand, the performance of TEMPORAL MOVIE was lower than the other metrics, at least for the EPFL-PoliMI video database.

The performance of MOVIE, SPATIAL MOVIE, and TEMPORAL MOVIE was not evaluated in HDTV video databases and in the LIVE Wireless video database because the execution of the metric requires a very significant amount of time (many days) in comparison with the other metrics. This fact decreases the usability of these metrics considerably. It may be argued that for development purposes, it is less important, but with computation times of several hours, this is a problem also for this usage.

In the results from the HDTV video database, we can appreciate that the accuracy in the prediction can be increased if the reference is compressed, compared to the case where the reference is uncompressed.

6. No-reference bitstream model development

In this section, we demonstrate how the full-reference objective metrics can be used to speed up the development process of no-reference bitstream real-time video QoE monitoring methods. In particular, we develop a no-reference bitstream model using the VQM full-reference metric, and we validate it using the subjective databases EPFL-PoliMI and LIVE Wireless Video Quality Assessment database.

We present a lightweight no-reference bitstream method that uses the packet loss rate and the interval between instantaneous decoder refresh frames (IDR frames) to estimate the video quality. IDR frames are 'delimiters' in the stream. After receiving an IDR frame, frames prior to the IDR frame can no longer be used for prediction. As well as this, IDR frames are also I-frames, so they do not reference data from any other frame. This means they can be used as seek points in a video. The no-reference bitstream model was fitted using several videos from the Consumer Digital Video Library (CDVL) database [35] and the VQM metric. Then, it was validated with the video databases EPFL-PoliMI and LIVE Wireless Video Quality Assessment database. The VQM metric has been used to train the no-reference bitstream

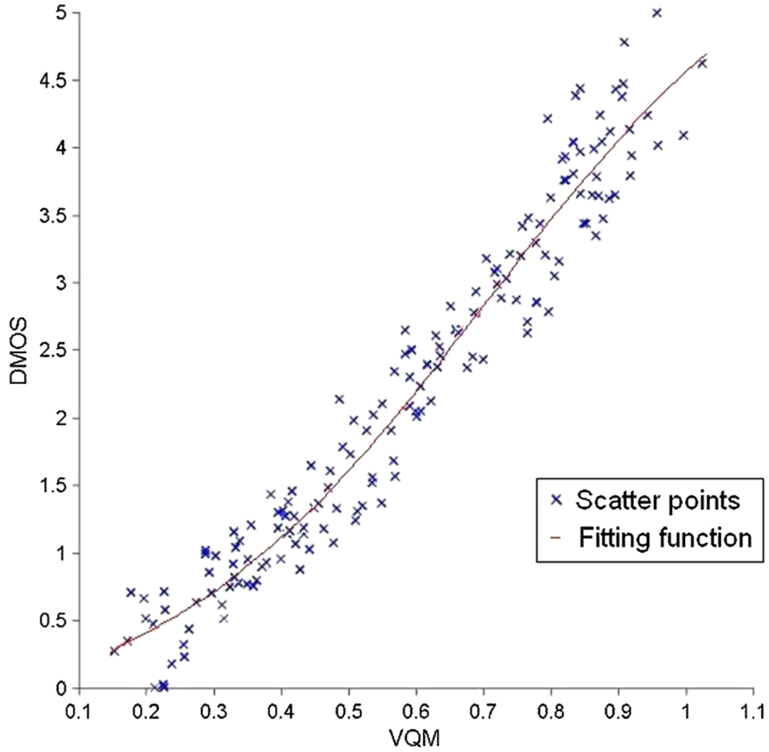

Figure 9 Scatter plot VQM for LIVE Wireless.

model regarding only the transmission distortions, and no compression distortions such as QP have been taken into consideration because it has been shown that VQM is very accurate when only transmission distortions are considered using a compressed reference. The case where VQM is used to measure a combination of compression and transmission distortions (for example, different QP and packet loss rate with uncompressed reference) is not evaluated in this paper.

6.1 Framework for model development

We selected the VQM metric to develop a no-reference bitstream model because of the very good performance shown in the previous section.

Six sequences with resolution 1,920 × 1,080 pixels were downloaded from the Consumer Digital Video Library (CDVL) database [35], with different characteristics. In five of the videos, the final part was removed to generate videos of a total length of 17 s at 30 fps. One of the sequences had a total length of 14 s at 25 fps. The SRC, listed in Table 6, were selected to spread a large variety of different content in Full-HD 1,920 × 1,080 format.

The videos were converted from YUV packed 4:2:2 to YUV planar 4:2:0. The videos were compressed with the Quantization Parameter set to 26, 32, 38 and 44. In order to make sure the no-reference model is valid for the different compression qualities the QP has been set to 26, 32, 38 and 44. However the performance of VQM in the case of compressed reference has been only tested in the case of compressed reference of high quality, which may not correspond to a QP value of 44. This causes a small degree of uncertainty in the obtained results because the scenario in which the compressed reference has low quality remains to be verified. The parameter keyint in the x264 encoder, corresponding to the interval between IDR frames, was set to 12, 36, 60 and 84. The maximum slice size was set to 1400 bytes. We consider that the keyint parameter is important since the distortion due to a packet loss propagates until the next IDR frame. Thus a higher value implies more error propagation and lower video quality. Finally the packet loss rate was set to 0.1%, 1%, 3%, 5%, and 10%. In total, 6 × 4 × 4 × 5 = 480 distorted videos were evaluated using the VQM metric.

Table 6 List of SRCs used in model development

SRC	Thumbnail	Description	Name of the sequence in CDVL
1		Woman smoking and people on a street, high contrast in the rock	NTIA outdoor mall with tulips (3e)
2		Kayaking, scene changes, fast moving water	NTIA Red Kayak
3		Trees, leaves, short and numerous movements in most of the image, scene changes	NTIA Aspen Trees in Fall Color, Slow Scene Cuts
4		Mountain with snow and moving fog in a sunny day, high brightness, scene changes	NTIA Snow Mountain
5		Global view of a city, buildings, scene changes, rather static	NTIA Denver Skyscrapers
6		Two people speaking in a table and showing an electronic device	NTIA Front End (Part of a Longer Talk)

The videos were encoded with the x264 encoder [36], random packet losses were inserted using a packet loss simulator [37] and the videos were decoded with the ffmpeg decoder. The ffmpeg decoder produces incomplete video files when random packet losses are inserted. To be able to apply the VQM metric, the videos were reconstructed so that they have the same length as the original. The reconstruction was done in two steps. First, the frame numbers were inserted into the luminance information of the uncompressed original sequence. After decoding the videos, the frame numbers were read and used to identify the missing frames and reconstruct the decoded video. The reconstruction method is explained in detail in [38].

The framework is described in Figure 10. As it can be seen in the figure, the VQM metric was applied (after conversion to packed 4:2:2 format) between the compressed reference (video compressed and uncompressed) and the reconstructed video. We used the same version of VQM than in the previous sections (described in Section 4.2). The same parameters as in Section 4.2 were used for VQM.

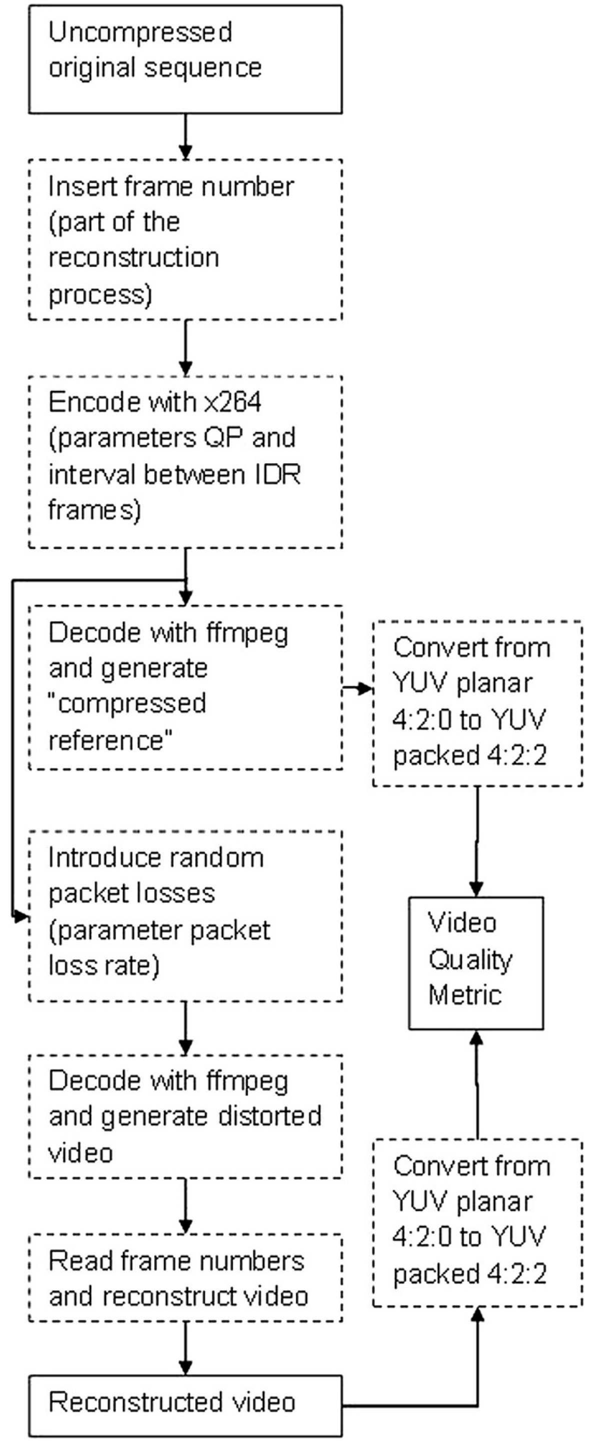
Uncompressed original sequence
Insert frame number (part of the reconstruction process)
Encode with x264 (parameters QP and interval between IDR frames)
Decode with ffmpeg and generate "compressed reference"
Convert from YUV planar 4:2:0 to YUV packed 4:2:2
Introduce random packet losses (parameter packet loss rate)
Video Quality Metric
Decode with ffmpeg and generate distorted video
Convert from YUV planar 4:2:0 to YUV packed 4:2:2
Read frame numbers and reconstruct video
Reconstructed video

Figure 10 Framework for model development.

6.2 Model development

In this case, our objective was to develop a lightweight model to predict the quality of the video as a function of two parameters: packet loss rate in percentage, denoted p, and interval between IDR frames in number of frames, denoted I. The MATLAB function nlinfit was used to calculate the coefficients of the following equation:

$$\begin{aligned}\text{VQM} = {} & b_0 + b_1 \cdot I^3 + b_2 \cdot I^2 + b_3 \cdot I + b_4 \cdot p^3 + b_5 \cdot p^2 \\ & + b_6 \cdot p.\end{aligned}$$

With the non-linear fit, we obtained the following no-reference bitstream model for the predicted quality, $f(I,p)$:

$$\begin{aligned}f(I,p) = {} & -0.16 – 0.0001 \cdot I^2 + 0.0064 \cdot I \\ & + 0.0003 \cdot p^3 – 0.0092 \cdot p^2 + 0.1106 \cdot p.\end{aligned}$$

The three-dimensional plot in Figure 11 shows the VQM values as a function of packet loss rate and interval between IDR frames together with the developed model (surface).

6.3 Validation of the model

To validate the no-reference bitstream model, we applied the model to the EPFL-PoliMI and LIVE Wireless Video Quality Assessment databases, and we calculated the linear correlation coefficient with the subjective values. The model was not checked on the HDTV database because the HDTV database was done applying a packet loss rate to a percentage of the way through the sequence. In order to apply our model, we expect a constant packet loss rate along all the sequence. As the interval between IDR frames is fixed in all the databases used, we are only able to verify the part of the equation related to the packet loss rate. For the EPFL-PoliMI, we

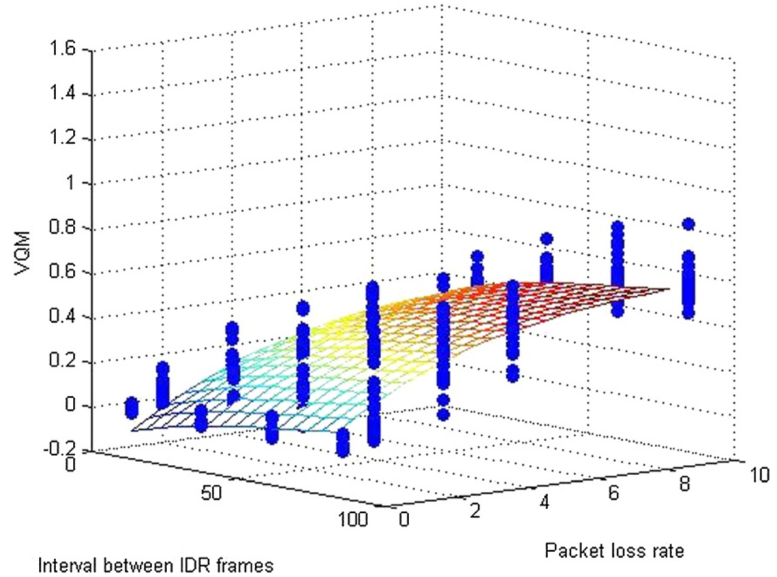

Figure 11 VQM as a function of packet loss rate and interval between IDR frames. The developed model is also plotted (surface).

obtained a linear correlation coefficient of 0.945, and for the LIVE Wireless Video Quality Assessment database, we obtained a linear correlation coefficient of 0.903. We believe that the model can be improved by adding new parameters and improving the fitting function used. The important fact is that these results validate the methodology followed in order to develop a no-reference bitstream model.

7. Conclusions

High-quality video streaming services over the Internet are increasing in popularity, and as people start to pay for the services, the quality must be guaranteed. Therefore, video quality monitoring and prediction become important in the development of Internet service management systems. Numerous objective assessment methods have been proposed; however, independent comparisons are scarce. Also, real-time monitoring requires lightweight no-reference bitstream models that perform accurately enough.

In this paper, we propose a strategy for developing new no-reference objective video quality metrics by using well performing full-reference video objective quality metrics to reduce the development time. The starting point is to define a relatively narrow scope. Find a FR model to create a big training database by varying the parameters that will be present. Train the NR model on this database. The NR model can then be validated using a smaller subjective test. In case there is a need for the use of the model outside the scope, the strategy is to retrain the model for the new scope.

This strategy is illustrated on the scope of transmission distortions in the case of compressed reference. As a first step, we have evaluated six publicly available full-reference metrics using three freely available video databases. The main objective of the evaluation was to compare the performance of the metrics when transmission distortions in the form of packet loss were introduced. The results show that VQM performs very well in all the video databases, being the best metric among the studied in the HDTV video database (uncompressed reference) and in the LIVE Wireless video database. In the EPFL-PoliMI database, SPATIAL MOVIE performed best and TEMPORAL MOVIE performed worst. When transmission distortions are evaluated, using the compressed video as the reference provides greater accuracy than using the uncompressed original video as the reference, at least for the studied metrics.

We believe that the correlation values obtained would be lower if registration problems occurred and different error concealment strategies were applied.

Further, to demonstrate the suggested strategy of model development, we present a no-reference bitstream model trained and optimized using full-reference model evaluation. The objective of the model is to accurately enough predict the video quality when transmission distortions are introduced. We fit the model using videos from the Consumer Digital Video Library (CDVL) database and the VQM metric. Then, the model is validated using the video databases EPFL-PoliMI and LIVE Wireless Video Quality Assessment database with reasonable performance.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was developed inside the Future Internet project supported by the Basque Government within the ETORTEK Programme and the FuSeN project supported by the Spanish Ministerio de Ciencia e Innovación, under grant from Fundación Centros Tecnológicos - Iñaki Goenaga and Tecnalia Research & Innovation. The work was also partly financed by the CELTIC project IPNQ-SIS, and the national project EFRAIM, with the Swedish Governmental Agency for Innovation Systems (VINNOVA) supporting the Swedish contribution.

Author details

[1]TECNALIA, ICT - European Software Institute, Parque Tecnológico de Bizkaia, Edificio 202, Zamudio E-48170, Spain. [2]Acreo Swedish ICT AB, NETLAB: Visual Media Quality, Box 1070, Kista SE-164 25, Sweden. [3]Deptartment of Electrical and Information Technology, Lund University, Box 117, Lund SE-221 00, Sweden. [4]Department of Information Technology and Media, Mid Sweden, University, Holmgatan 10, Sundsvall SE-851 70, Sweden.

Received: 18 March 2013 Accepted: 23 December 2013
Published: 14 January 2014

References

1. K Ahmad, A Begen, IPTV and video networks in the 2015 timeframe: The evolution to medianets. IEEE. Commun. Mag. **47**, 68–74 (2009)
2. A Takahashi, D Hands, V Barriac, Standardization activities in the ITU for a QoE assessment of IPTV. IEEE. Commun. Mag. **46**, 78–84 (2008)
3. F Kuipers, R Kooij, DD Vleeschauwer, K Brunnström, *Techniques for measuring quality of experience*. Lecture notes in computer science 6074/2010 (Springer Berlin, Heidelberg, 2010)
4. S Winkler, P Mohandas, The evolution of video quality measurements: from PSNR to Hybrid metrics. IEEE. Trans. Broadcast. **54**, 660–668 (2008)
5. AR Reibman, VA Vaishampayan, Y Sermadevi, Quality monitoring of video over a packet network. IEEE. Trans. Multimed. **6**, 327–334 (2004)
6. M Barkowsky, I Sedano, K Brunnström, M Leszczuk, N Staelens, Hybrid video quality prediction: re-viewing video quality measurement for widening application scope. Multimed Tool Appl. in press
7. S Hemami, A Reibman, No-reference image and video quality estimation: applications and human-motivated design. Signal Process. Image Commun. Elsevier. **25**, 469–481 (2010)
8. M Pinson, S Wolf, A new standardized method for objectively measuring video quality. IEEE. Trans. Broadcast. **50**(3), 312–322 (2004)
9. C Lee, S Cho, J Choe, T Jeong, W Ahn, E Lee, Objective video quality assessment. Optic. Engineer. **45**(1), 017004 (2006)
10. J Greengrass, J Evans, A Begen, Not all packets are equal, part 2: the impact of network packet loss on video quality. IEEE. Int. Comput. **13**, 74–82 (2009)
11. K Brunnström, D Hands, F Speranza, A Webster, VQEG validation and ITU standardization of objective perceptual video quality metrics. IEEE. Signal. Process. Mag. **26**, 96–101 (2009)
12. M Naccari, M Tagliasacchi, S Tubaro, No-reference video quality monitoring for H.264/AVC coded video. IEEE. Trans. Multimed. **11**, 932–946 (2009)
13. M Garcia, A Raake, P List, Towards content-related features for parametric video quality prediction of IPTV services, in *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processes* (, Las Vegas, NV, 2008)
14. Video quality experts group page. http://www.its.bldrdoc.gov/vqeg. Accessed 07 Jan 2014
15. I Sedano, M Kihl, K Brunnstrom, A Aurelius, Evaluation of video quality metrics on transmission distortions in H.264 coded videos, in *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)* (Nuremberg, 2011)
16. F De Simone, M Naccari, M Tagliasacchi, F Dufaux, S Tubaro, T Ebrahimi, Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel, in *Proceedings of the International Workshop on Quality of Multimedia Experience (QoMEX)* (San Diego, CA, 2009)
17. F De Simone, M Tagliasacchi, M Naccari, S Tubaro, T Ebrahimi, A H264/AVC video database for the evaluation of quality metrics, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Dallas, TX, 2010)
18. EPFL-PoliMI video quality assessment database [Online]. Available: http://vqa.como.polimi.it. Accessed 07 Jan 2014
19. M Barkowsky, M Pinson, R Pépion, P Le Callet, Analysis of freely available subjective dataset for HDTV including coding and transmission distortions, in *Proceedings of the Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)* (Scottsdale, AZ, 2010)
20. AK Moorthy, K Seshadrinathan, R Soundararajan, AC Bovik, Wireless video quality assessment: a study of subjective scores and objective algorithms. IEEE. Transact. Circ. Syst. Video Technol. **20**(4), 513–516 (2010)
21. FOURCC, YUV formats [Online]. Available: http://www.fourcc.org/yuv.php. Accessed 07 Jan 2014
22. H.264/AVC reference software version JM14.2, Tech. Rep., Joint Video Team (JVT) [Online]. Available: http://iphome.hhi.de/suehring/tml/download/old_jm/. Accessed 07 Jan 2014
23. ITU-T, *Recommendation ITU-T P 910, September 1999, Subjective video quality assessment methods for multimedia applications* (ITU-T, Geneva)
24. K Seshadrinathan, R Soundararajan, AC Bovik, LK Cormack, Study of subjective and objective quality assessment of video. IEEE. Trans. Image Process. **19**, 1427–1441 (2010)
25. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE. Trans. Image Process. **13**(4), 600–612 (2004)
26. Z Wang, EP Simoncelli, AC Bovik, Multi-scale structural similarity for image quality assessment, in *Proceedings of the IEEE Asilomar Conference Signals, Systems and Computers* (Pacific Grove, CA, 2003)
27. Video Quality Metric (VQM) software [online]. Available: http://www.its.bldrdoc.gov/vqm/. Accessed 07 Jan 2014
28. DM Chandler, SS Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images. IEEE. Trans. Image Process. **16**(9), 2284–2298 (2007)
29. K Seshadrinathan, AC Bovik, Motion tuned spatio-temporal quality assessment of natural videos. IEEE. Trans. Image Process. **19**(2), 335–350 (2010)
30. MATLAB Central File Exchange [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/. Accessed 07 Jan 2014
31. The Structural SIMilarity (SSIM) index author's home page [Online]. Available: http://www.ece.uwaterloo.ca/~z70wang/research/ssim/. Accessed 07 Jan 2014
32. Laboratory for image & video engineering [online]. Available: http://live.ece.utexas.edu/research/Quality/index.htm. Accessed 07 Jan 2014
33. VSNR implementation from the authors [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html. Accessed 07 Jan 2014
34. Final report from the video quality experts group on the validation of objective models of multimedia quality assessment, phase I [Online]. Available: ftp://vqeg.its.bldrdoc.gov/Documents/VQEG_Approved_Final_Reports/VQEG_MM_Report_Final_v2.6.pdf. Accessed 07 Jan 2014
35. The consumer digital video library [online]. Available: http://www.cdvl.org/about/index.php. Accessed 07 Jan 2014
36. x264 software [Online]. Available: http://www.videolan.org/developers/x264.html. Accessed 07 Jan 2014
37. JVT-Q069 [Y. Guo, H. Li, Y.-K. Wang] SVC/AVC loss simulator [Online]. Available: http://wftp3.itu.int/av-arch/jvt-site/2005_10_Nice/. Accessed 07 Jan 2014
38. I Sedano, M Kihl, K Brunnstrom, A Aurelius, Reconstruction of incomplete decoded videos for use in objective quality metrics, in *Proceedings of the 19th Int. Conf. Syst. Signals Image Process (IWSSIP)* (s, Vienna, 2012), pp. 376–379

doi:10.1186/1687-5281-2014-4
Cite this article as: Sedano *et al.*: **Full-reference video quality metric assisted the development of no-reference bitstream video quality metrics for real-time network monitoring.** *EURASIP Journal on Image and Video Processing* 2014 **2014**:4.