J Internet Serv Appl (2011) 2:129–138
DOI 10.1007/s13174-011-0025-z

SI: FUTURE NET SERVICE MODELS & DESIGNS

Packet level video quality evaluation of extensive H.264/AVC and SVC transmission simulation

Robert Skupin · Cornelius Hellge · Thomas Schierl · Thomas Wiegand

Received: 10 February 2011 / Accepted: 18 May 2011 / Published online: 19 June 2011

Abstract Video transmission over error prone channels as present in most of today's communication channels, such as Mobile TV or some IPTV systems, is constantly subject to research. Simulation is an important instrument to evaluate performance of the overall video transmission system, but the multitude of parameters often requires large and time-consuming simulation sets. In this paper, we present a packet level mechanism for fast evaluation of error-prone H.264/AVC and SVC video transmission with application layer video quality metrics, such as PSNR. Our approach significantly reduces the overall simulation time by eliminating redundancy in the evaluation phase and utilizing the prediction structure of the video codec. The benefit of the presented packet level video quality evaluation is evaluated with an exemplary simulation setup of an IPTV service with link congestion.

Keywords Network simulation · Video quality evaluation · H.264/AVC · SVC · IPTV

R. Skupin (✉) · C. Hellge · T. Schierl
Image Processing Department, Multimedia Communications Group, Fraunhofer Institute for Telecommunications—Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany
e-mail: robert.skupin@hhi.fraunhofer.de

C. Hellge
e-mail: cornelius.hellge@hhi.fraunhofer.de

T. Schierl
e-mail: thomas.schierl@hhi.fraunhofer.de

C. Hellge · T. Wiegand
Department of Telecommunication Systems, Image Communication Group, Berlin Institute of Technology, Einsteinufer 17, 10587 Berlin, Germany

T. Wiegand
e-mail: thomas.wiegand@tu-berlin.de

T. Wiegand
Image Processing Department, Fraunhofer Institute for Telecommunications—Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany

1 Introduction

Video transmission over error prone channels as present in most of today's communication channels, such as Mobile TV or some IPTV systems, is a topic of ongoing research encouraged by progress of involved components, new system requirements or user demands. A wide range of parameters influences the overall system performance. Various error control techniques like Automatic Repeat Request or Forward Error Correction (FEC), varying network conditions and congestion states, or prioritization schemes of data influence the performance of an IPTV video service. Furthermore, today's state of the art video codec standard H.264/AVC [1] and its Scalable Video Coding (SVC) extension [2] offer numerous tools to adjust the video coding setup to the specific requirements of a service. Fine-tuning of channel parameters, media coding, and error control is vital to achieve flawless operation and an optimal user experience within given system constraints.

Simulation is an important instrument to evaluate a certain parameter setup in the first place, but the multitude of parameters often results in very large simulation sets. Network level statistics, such as packet or block error rate are easily gathered but inadequate for a concluding evaluation from a users perspective, as users judgment highly depends on perceived visual quality. Numerous applica-

tion layer metrics have been proposed to allow algorithmic Video Quality Evaluation (VQE). When simulating video transmission, there are basically two approaches of gathering measurements with these VQE metrics, i.e. the conventional VQE approach using bitstream reconstruction, video decoding, and measurement or estimation of video quality degradation. Most general simulation frameworks, such as EvalVid [3] and its numerous extensions [4–7] or simulators for specific transmission systems such as presented in [8] for WiMAX or in [9] for DVB-T2, take the conventional VQE approach that includes video decoding in each simulation cycle. As availability of an error resilient decoder implementation is not always given, some of the above frameworks introduce simplifications to the decoding process that challenge the results validity. Considering possibly large simulation sets, the conventional VQE approach can be very time-consuming as video decoding is still a computational challenging task. In general, a sufficiently large number of simulation cycles includes (at least partially) identical video output, thus redundant operations are carried out. Models to estimate the additional video signal distortion from packet losses without decoding are beneficial when limited processing power makes the conventional VQE approach unfeasible, access to an undistorted reference is not given or live quality monitoring is targeted [10, 11], but these models still have individual weaknesses such as a limited accuracy [12, 13] or are restricted to video streams with specific coding parameters [14].

This work presents and analyzes a mechanism for fast evaluation of extensive error-prone video transmission scenarios with application layer metrics on packet level, referred to as Packet Level Video Quality Evaluation (PLVQE) introduced in [15]. Our approach allows fast application layer VQE and significantly reduces the simulation time by combining and reducing redundant calculations that are usually carried out in each simulation cycle. We utilize knowledge of the video prediction structure used in the coding process to define a set of relevant transmission errors and take corresponding VQE measurements offline to constitute a VQE database. During simulation time, the VQE database allows for fast evaluation of video quality on packet level without further video decoding. We analyze the presented approach in terms of accuracy and achievable runtime savings with an exemplary simulation of an IPTV service over a congested link. The presented mechanism has successfully been used in simulations within the context of SVC for mobile satellite transmission [16] and in investigations of different FEC schemes [17].

The remainder of this paper is organized as follows. In Sect. 2, we give a brief overview of the H.264/AVC and SVC video coding standard, prediction structures used for motion compensation, and video quality evaluation metrics. Section 3 explains the proposed PLVQE in detail, and Sect. 4 presents the simulator used for implementation of PLVQE. A performance analysis is given in Sect. 5, and we conclude with Sect. 6.

2 Video coding

Most of today's video transmission systems such as specified within 3GPP, DVB, and ATSC, or Internet video services such as YouTube or Vimeo make use of H.264/AVC, which is a state of the art hybrid video coding standard featuring block oriented motion compensation and transform coding. H.264/AVC achieves significant improvements in coding efficiency compared to previous standards and provides a network-friendly video representation of the coded data. Its design consists of the Video Coding Layer (VCL) and the Network Abstraction Layer (NAL). The VCL constitutes a hybrid of block-based prediction, quantized transform coding, and entropy coding. Coded VCL frame data and additional information are further processed in the NAL by encapsulation in so-called NAL units with additional header information. The concept of NAL units strongly simplifies transportation of VCL data in systems such as Real-time Transport Protocol (RTP) Internet services and MPEG-2 transport streams or storage in containers, e.g. the MP4 file format.

The extension for Scalable Video Coding (SVC) in H.264/AVC allows further structuring the bitstream and extracting different video representations of a single bitstream, referred to as layers. The base layer of SVC provides the lowest quality level and is an H.264/AVC compliant bitstream to ensure backward-compatibility with existing receivers. Each additional enhancement layer improves the video quality in a certain dimension. SVC allows up to three different scalability dimensions within one bitstream: temporal, spatial, and quality scalability, which yields great potential to achieve a more efficient and flexible provisioning of video services. Compared to using a simulcast approach, i.e. providing several versions of the same content in multiple H.264/AVC streams, SVC provides efficient means to cope with heterogeneous receiver capabilities (screen size and processing power) and extending existing services in a backwards compatible way. In contrast to relatively large stream switching delays introduced with H.264/AVC simulcast, SVC allows a simple implementation of graceful quality switching.

2.1 Prediction structures

Motion compensation based video codes utilize inter-frame (i.e. temporal) prediction to reduce redundancy of video data. H.264/AVC coded video frames can be classified according to a set of frame types: Intra-coded (I) frames,

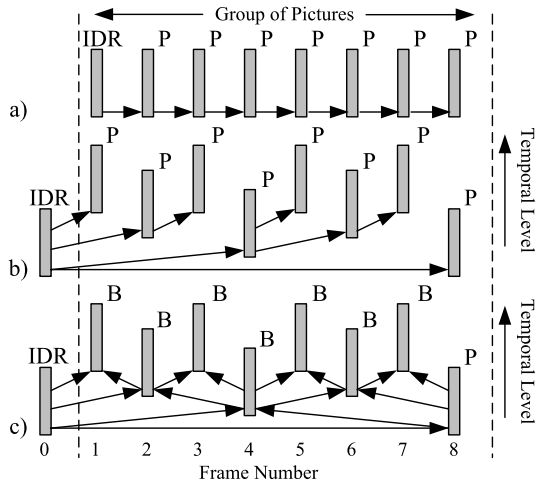

Fig. 1 Prediction structures with (**a**) P frames, (**b**) hierarchical P frames, and (**c**) hierarchical B frames

Instantaneous Decoder Refresh (IDR) frames, Predictive-coded (P) frames and Bi-predictive-coded (B) frames. I frames do not predict data from surrounding frames and can therefore be decoded independently. IDR frames are intra-coded frames that refresh the decoding picture buffer. Thus, all frames following in coding order do not have access to frames prior to the IDR frame for prediction. P frames use temporally preceding frames for prediction, whereas B frames use both temporally preceding and following frames for prediction. In general, prediction is done according to a defined prediction structure build up of the aforementioned frame types. By decoupling decoding and presentation order of video frames, H.264/AVC and SVC allow for using arbitrary prediction structures.

Figure 1 illustrates three possible prediction structures within a H.264/AVC video sequence that consists of 8 frames. The frames are numbered in presentation order and the arrows represent the dependencies between individual frames that arise from prediction using reference frames. The structure depicted in Fig. 1(a) is referred to as IPPP coding and uses only P frames except for an IDR frame in the beginning and periodic I frames that serve as random access points. This prediction structure allows for very low coding delay as there is no difference between the coding and presentation order of frames. On the downside, despite being the common coding structure in H.264/AVC due to its simplicity, its coding efficiency is not optimal and decoding errors may propagate until the following I frame.

Hierarchical prediction structures, such as shown in Fig. 1(b) and (c), utilize temporal levels for hierarchical prediction, which are indicated in Fig. 1. Such prediction structures have been found to be advantageous in terms of coding efficiency and additionally offer temporal scalability as a benefit. Unlike the base layer, i.e. frames of the lowest temporal level, that may have prediction dependencies on frames of the base layer, frames of higher temporal levels are typically restricted to prediction from lower temporal levels. After a mandatory intra-coded IDR frame at the beginning, SVC coded video typically uses prediction structures with hierarchical B frames [18] as shown in Fig. 1(c). Hierarchical P frame structures are beneficial when a low coding delay is necessary, e.g. as in low latency video conversation applications, but lead to a lower coding efficiency [19].

A set of frames between two successive video frames of the lowest temporal level with the succeeding lowest temporal level frame constitutes a Group of Pictures (GOP) structure. When there is only a single temporal level available, as in Fig. 1(a), we define the GOP size as the distance between intra-coded frames. SVC coded video adds an additional level of frames to the GOP structure depicted in Fig. 1, which does not serve for prediction by the lower layers. The proposed PLVQE has been implemented for a prediction structure with hierarchical B frames, but can be extended to any given prediction structure.

2.2 Video quality evaluation

As video coding and transmission may introduce distortion into the processed video, the non-trivial task of VQE is an important instrument to evaluate compression efficiency or transmission system performance. A large test population is necessary to gather statistically relevant results with subjective tests, which is rather costly and time consuming. Thus, this approach is not feasible for large simulation sets and objective algorithmic metrics are advantageous.

Given the original frame I and its coded representation K, both of size $m \times n$, a simple metric to measure the difference between I and K is the Mean Square Error (MSE) as shown in (1).

$$\mathrm{MSE} = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \big[I(i,j) - K(i,j) \big]^2. \tag{1}$$

Today's de facto standard metric in the video coding community is the Peak-Signal-to-Noise-Ratio (PSNR), which is the logarithmic ratio of the maximum pixel intensity of image I to the square root of the MSE according to (2).

$$\mathrm{PSNR} = 10 \cdot \log_{10} \left(\frac{\mathrm{MAX}_I}{\sqrt{\mathrm{MSE}}} \right). \tag{2}$$

PSNR measurements are typically taken for the luminance component of a video frame and averaged over the video sequence. Its clear physical meaning and simple calculation made PSNR the commonly used VQE metric. However, error sensitivity of the human visual system and masking effects in spatial and temporal frequency domain heavily influence the perceived video quality. While a human

observer pays less attention to homogeneous regions than to image details, i.e. edges and textures, simple pixel- and frame-based metrics such as MSE or PSNR lack a corresponding semantic capability. Therefore, MSE and PSNR do not correlate well to results of subjective tests in certain respects and can only be seen as an approximation, especially in case of error-prone video transmission [20, 21].

Ongoing research addresses the development of new algorithmic VQE metrics that correspond to the characteristics of human visual perception to a higher degree and go beyond calculating sheer pixel differences among original and coded video frames [22–25]. Perceptual Evaluation of Video Quality (PEVQ) [26] or Structural Similarity Index (SSIM) [27], along with a variety of others, extract image features in the form of structures or image activity, and consider the movement in a video sequence to weight frame-wise measurements. This is often coupled to a significant increase in computational complexity but still none of the above correlates exactly with subjective test results or is as widely used as PSNR. Considering the complexity of human visual perception and that people's idea of state-of-the-art video quality changes over time, it is clear that development of video quality metrics is a challenging task.

When evaluating the quality of erroneous video, analyzing the playout behavior in addition to the video quality metrics described above can be beneficial. A simple metric for measuring the robustness of playout is the Errored Second Ratio (ESR) [28, 29]. It is defined as the ratio of seconds that contain errors, i.e. at least one non-decodable frame in the context of video, to the overall length in seconds. The proposed PLVQE has been implemented for the PSNR and ESR metric. However, the presented concepts are applicable to any frame-based metric such as MSE or PSNR, whereas metrics that incorporate temporal aspects, e.g. movement within a scene, would require adjustments to cover the additional layer of complexity.

3 Packet level video quality evaluation

The proposed Packet Level Video Quality Evaluation (PLVQE) provides an application layer quality evaluation of transmitted video on packet level without the need to decode the result of each simulation cycle individually. The general idea is to constitute a database of VQE measurements offline during a preprocessing phase that covers every possible erroneous video output. Decoding and evaluation operations that are carried out within each simulation cycle in the conventional VQE approach are thereby combined to omit redundant calculations. Considering the prediction structure of H.264/AVC and SVC coded video allows to reduce the amount of required calculations while maintaining coverage of all possible video outputs, which will be discussed in detail in the following Sect. 3.1. In the evaluation phase, transmitted video sequences can be evaluated online using the VQE database. Packet losses are analyzed and mapped to the corresponding VQE values in the database. Thus, after preprocessing, large simulation sets can be evaluated in a very short time without executing any video decoding operation. The results of PLVQE thereby closely represent the behavior of the underlying video decoder.

The measurements used to constitute the VQE database are gathered with an error resilient decoder implementation that is compatible with the H.264/AVC and SVC standard. It supports basic error concealment techniques, such as base layer upsampling (for SVC spatial scalability) in case of SVC enhancement layer data loss, the insertion of freeze frames in case of SVC base layer data loss to keep video output in sync and further advanced techniques such as enhancement layer utilization for base layer error concealment [30]. Additionally, the proposed PLVQE requires the video coding to fulfill certain constraints. First, the prediction structure has to be known, which is crucial for PLVQE as the number of decoding operations is reduced according to the prediction structure. Second, a limitation to a small number of slices per frame is necessary to reduces complexity and processing time to a reasonable degree. The implementation used to analyze the approach in Sect. 5 supports prediction structures with hierarchical B frames with various GOP sizes and a single slice per frame.

3.1 Relevant error pattern

The space of all possible (frame) error combinations within a video sequence is very large. With the given constraints and a number of n video frames, the number of all error combinations within an H.264/AVC video is 2^n and $2^{m \cdot n}$ for SVC coded video with m layers. The amount of possible error combinations is very large even for short video sequences with a length of a few seconds. The first step to reduce the amount of error combinations is to focus on the level of GOPs instead of the whole video sequence. It is assumed that the video quality of a decoded GOP largely relates to errors within the GOP. The previous GOPs affect the video quality of the current GOP only to a minor degree through error propagation.

There is only one case, i.e. an complete loss of the lowest temporal level preceding the GOP, in which the decoding result solely depends on the last decoded frames as none of the frames within the GOP can be decoded due to missing reference data. A way to cope with such severe errors is to repeat the last decoded frame until the decoder receives a decodable NAL unit. This is referred to as freeze frame error concealment, which we will have to address with an additional technique that is described in a subsequent subsection. All other errors are covered with the GOP-based approach that is described in the following.

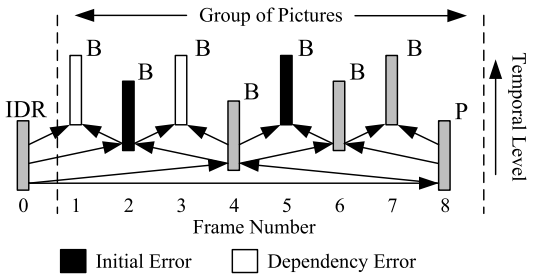

Fig. 2 Illustration of erroneous frames within a hierarchical prediction structure with B frames

Figure 2 illustrates an exemplary error distribution within a single-layer H.264/AVC hierarchical B frames GOP structure. The frames are numbered in presentation order and vertically sorted according to their temporal level. The arrows represent dependencies between individual frames that arise from the hierarchical structure used for prediction from surrounding frames. SVC introduces additional dependencies across layers. To illustrate the characteristics of decoding errors, exemplary errors are indicated by black and white blocks. Still, there are 2^n possible combinations to consider, but focusing on GOP level, n depends only on the GOP size, regardless of the video length. More precisely, n is the number of frame representations within the GOP for SVC coded video or simply the GOP size in case of H.264/AVC. The depicted hierarchical B frames GOP structure allows $2^8 = 256$ error combinations.

Taking inter-frame (and inter-layer in case of SVC) dependencies into account can significantly reduce the error combinations of interest. Erroneous frames can be divided into two categories. The first category is constituted by frames that are not decodable due to erroneously transmitted corresponding NAL units. Frame 2 and frame 5 within the GOP structure depicted in Fig. 2 belong to this category and are referred to as initial errors. Initial errors are always caused by transmission errors that directly affect the NAL units of the particular frame. The second category contains dependency errors, which are not decodable due to missing reference data of other frames. Frame 1 and frame 3 are not decodable due to (partially) missing reference data in form of frame 2. Regardless of the availability of NAL units belonging to frame 1 and frame 3, both frames fall into the second category, referred to as dependency errors. NAL units of frames that belong to this category are not necessarily affected directly by transmission errors. Since the resulting video output is identical for error combinations that consist of the same initial errors, considering only initial error combinations is sufficient to cover all error combinations. Identifying and processing these Relevant Error Patterns (REPs) reduces the number of necessary decoding operations significantly. The number of REPs highly depends on the prediction structure and GOP size. Figure 3 illustrates the numerically gathered amount of REPs for various video coding

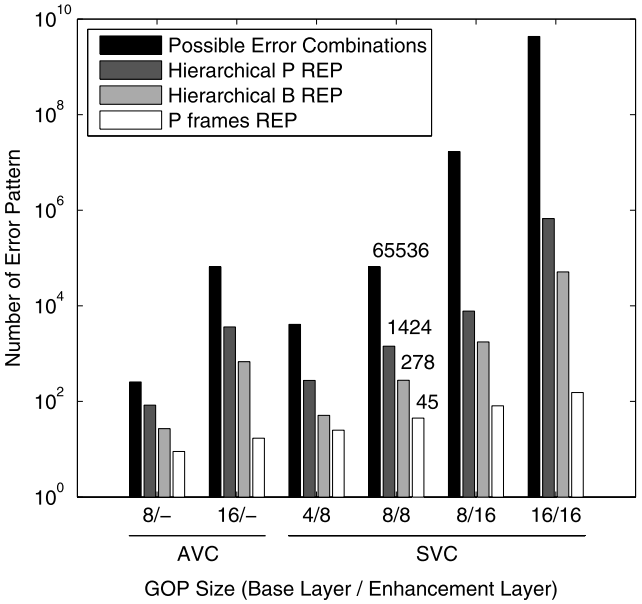

Fig. 3 Number of error combinations and REPs for various coding setups and GOP sizes of H.264/AVC and SVC

setups and states the number of REPs for SVC coded video with a GOP size of 8 frames in base and enhancement layer that is used in the subsequent exemplary simulations.

3.2 Offline preprocessing

To generate a VQE database of PSNR measurements, the preprocessing of a given coded video sequence utilizes the previously described REPs. With information about timestamps, frame types and layers, each REP is mapped to the corresponding NAL units within all GOPs of the coded original video sequence in order to create an erroneous version of the video corresponding to a specific REP. NAL units unaffected by initial errors of the REP or dependency errors determined by the prediction structure are extracted and concatenated to reconstruct an erroneous bitstream. This bitstream is subsequently processed with an error resilient video decoder. A frame-wise PSNR measurement of the resulting video output is averaged for each GOP and stored in the VQE database in conjunction with an unique REP identifier. Additional VQE measurements for possibly occurring IDR frames that do not belong to a GOP structure are taken and stored. Note that this procedure is very similar to the conventional VQE approach with erroneous bitstream reconstruction, decoding, and video quality evaluation. The main difference is that the bitstream errors are repeatedly produced according to a specific REP as opposed to the random transmission errors of a transmission channel.

As mentioned in Sect. 3.1, it is not possible to measure VQE with the GOP-based approach in case of video data loss that exceeds the duration of a GOP and leads to a long

period of freeze frames. Therefore, another technique is used in parallel to extend the VQE database. All frame representations (one per frame for H.264/AVC and one per frame and layer for SVC) are compared to the temporally following original frames to obtain VQE measurements for the case of long-lasting freeze frames. Note that image quality of a specific frame and layer within an erroneous transmitted video varies, as errors on preceding frames propagate until the next IDR frame and introduce small changes to the quality of the following frames video. Therefore, the extracted frames used for comparison in this process might not be accurate and lead to a deviation of VQE results.

3.3 Online evaluation

In order to evaluate a simulation cycle, transmission results are analyzed on packet level. Missing or erroneous packets are mapped to the corresponding NAL units, which can be associated with specific video frames and layers. A GOP-wise analysis of all transmission errors with knowledge of the video prediction structure allows identifying the initial errors among all transmission errors. Information on a specific combination of initial errors is used to compose the unique REP identifier and query the corresponding PSNR measurements from the VQE database. Finally, VQE measurements of all IDR frames and GOPs are averaged to gather the mean PSNR of the transmitted video sequence. Additionally, counts of erroneous and decoded frames are gathered for calculation of ESR.

4 Simulation environment

The simulation platform used to implement and evaluate the performance of PLVQE is divided into an offline and an online part. Its modular structure closely resembles the different tasks that come along with video transmission, i.e. encoding, transmission simulation and evaluation, as illustrated in Fig. 4. The encoding module features a simple offline rate-controlled mechanism to encode a continuous test sequence chunk-wise using SVC reference encoder JSVM [31]. Chunks that match the simulation criteria (e.g. a constant bitrate or quality scenario) are subsequently concatenated into a continuous bitstream. Video data is packetized according to the RTP payload format [32] and a packet trace file is extracted that contains a textual description of the relevant characteristics of the RTP packets and the packetized video data. Further details on the rate-controlled encoding mechanism, especially in the context of a statistical multiplex scenario with several video streams, can be found in [16].

The packet trace serves as input for a trace-driven transmission simulator. In order to simulate a specific transmission system, an appropriate channel model has to be chosen. For instance, a service provided via ADSL has to cope

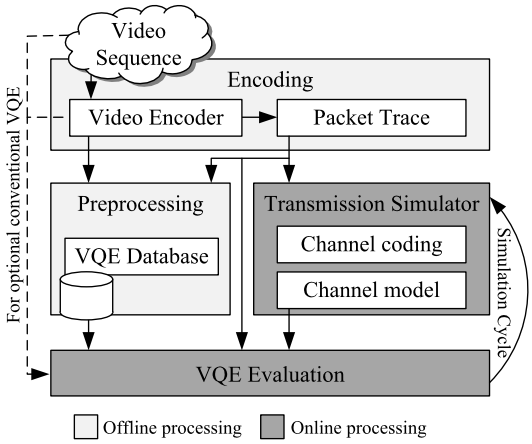

Fig. 4 Structure of simulation platform

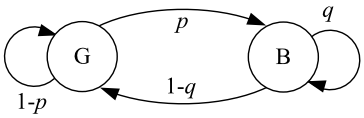

Fig. 5 Two state Markov process for the Gilbert Elliot model

with channel characteristics that are different from those of a mobile broadcast channel such as in DVB-SH [29]. Numerous effects influence the channel, e.g. path loss or fading for wireless, and attenuation or congestion for wired connections. The parameters under test determine whether the use of Packet Erasure Channel (PEC) models is sufficient or a physical layer binary erasure channel, such as the Additive White Gaussian Noise (AWGN) channel model, has to be used. The exemplary simulations conducted for the evaluation of PLVQE use a Gilbert Elliot model [33] that consists of a varying binary symmetric channel with crossover probabilities determined by a binary-state Markov process, as shown in Fig. 5. The average packet loss rate and the average loss burst length can be calculated with the crossover probabilities according to (3) and (4).

$$\textit{Average Packet Loss Rate} = \frac{1}{1+\frac{1-q}{p}}, \tag{3}$$

$$\textit{Average Burst Length} = \frac{1}{1-q}. \tag{4}$$

The modular structure of the simulation platform allows to apply various transmission simulators, e.g. Network Simulator 2 [34] or simulators for specific systems, that can be used for trace-driven simulations of IPTV, Mobile TV applications, or others. The only requirement for the transmission simulator is adaptability to the interfaces of the encoding and evaluation modules. The output of the transmission simulator, i.e. an erroneous packet trace with possibly missing lines resembling transmission errors of certain packets, serves as input for the evaluation module.

As described in Sect. 3.2, PLVQE includes an offline preprocessing phase, in which the original source and coded video are analyzed to acquire the VQE database. During simulation time, the VQE database allows for online VQE of the error-prone simulator traces on packet level. Optionally, the evaluation module provides conventional VQE that features bitstream reconstruction from packet trace, decoding and VQE measurement, which is done using the same error resilient video decoder implementation as for gathering the VQE database.

5 Results analysis

The main benefit of PLVQE are processing time savings while maintaining accuracy of results, which is proofed with an exemplary simulation of an IPTV service using SVC coded video with link congestion simulated according to the Gilbert Elliot channel. Results of PLVQE are compared with results of the conventional VQE approach featuring bitstream reconstruction from packet trace, video decoding and measurement of PSNR.

A concatenation of the four test sequences City, Crew, Habour, and Soccer with a total length of 34 seconds is encoded according to the scalable high profile using JSVM [31] with an approximately constant bitrate of 293 kbps and a single slice per frame. The quality scalable (CGS) SVC bitstream consists of an H.264/AVC compatible QVGA base layer at 15 fps with 29.58 dB PSNR and a QVGA enhancement layer at 15 fps with 34.41 dB PSNR. Both video layers have a GOP size of 8 frames and random access point intervals of approximately 0.5 s in the base layer and 2.1 s in the enhancement layer.

Channel simulation uses a Gilbert Elliot model, as described in Sect. 4. Its parameters are chosen according to a recent investigation on interdomain IPTV performance with synthetic RTP traffic over UDP/IP on ADSL links [35] that indicated an average loss burst length of about 1.2 packets and an average packet loss rate of 0.39%. In order to analyze PLVQE performance on a wide operating range, average packet loss rates of up to 10% with 18 steps and the given average loss burst length of about 1.2 packets are simulated. For evaluating statistical relevance of results, 150 iterations are conducted leading to a total of $18 \cdot 150 = 2{,}700$ simulation cycles per simulation set. In order to analyze the interaction of the GOP-based approach and the freeze frame handling for severe SVC base layer losses, we simulate one set with equal packet losses in base and enhancement layer and an additional set with losses restricted to the enhancement layer. With two simulation sets, a total of 5,400 simulations have to be carried out.

All simulations are conducted on a Dell Precision T7400 with two Intel Xeon X5482 CPU at 3.2 GHz and 16 GB of memory. The simulation framework and the video decoder run in single-threaded mode without parallelization.

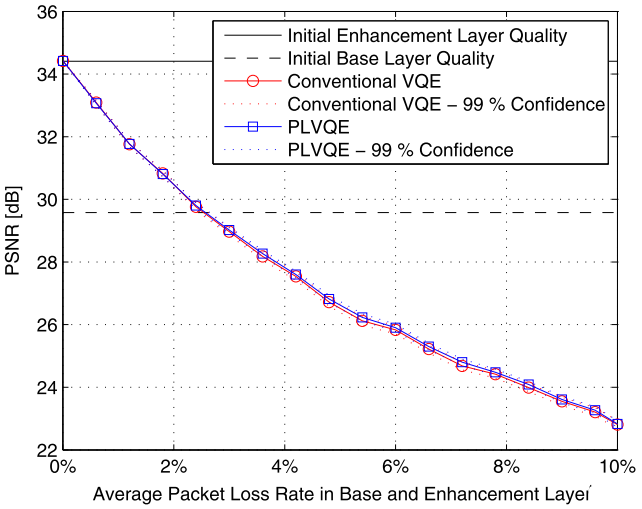

Fig. 6 PSNR results of the conventional VQE approach and PLVQE for packet losses in SVC base and enhancement layer with 99% confidence intervals

5.1 Accuracy

To analyze the accuracy of PLVQE, each simulation cycle is evaluated with PLVQE and conventional VQE. Figure 6 shows the results for both VQE approaches in terms of PSNR over average packet loss rate, where packets of both SVC layers are affected by packet loss. Therefore, this simulation set includes severe SVC base layer losses that have to be dealt with by the freeze frame handling of PLVQE. Within the analyzed operating area, the average deviation of PLVQE results compared to results of conventional VQE is 0.06 dB PSNR, with a maximum deviation of 0.15 dB. The measurements of conventional VQE and PLVQE for each setting of the Gilbert Elliot channel follow a normal distribution and the 99% confidence interval given in Fig. 6 and Fig. 7 is calculated accordingly.

Figure 7 shows simulation results with packet loss restricted to the SVC enhancement layer only. The given maximum packet loss rate leads to an almost complete loss of the enhancement layer, as can be seen from the PSNR of 30.12 dB. PLVQE entirely relies on the GOP-based approach to evaluate this simulation set. The average deviation of PLVQE results compared to the results of the conventional approach is 0.02 dB with a maximum deviation of 0.06 dB, which is significantly smaller than in the simulation set that includes packet losses of both SVC layers. However, in both exemplary simulations sets, deviation of PLVQE results does not rise to a notable magnitude and is marginal.

It can be seen that the resulting PLVQE deviations under severe SVC base layer losses, as explained in Sect. 3.2,

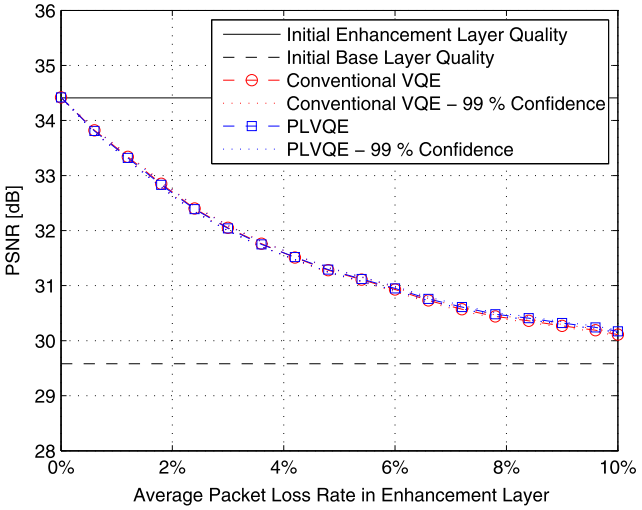

Fig. 7 PSNR results of the conventional VQE approach and PLVQE for packet losses in SVC enhancement layer only and 99% confidence intervals

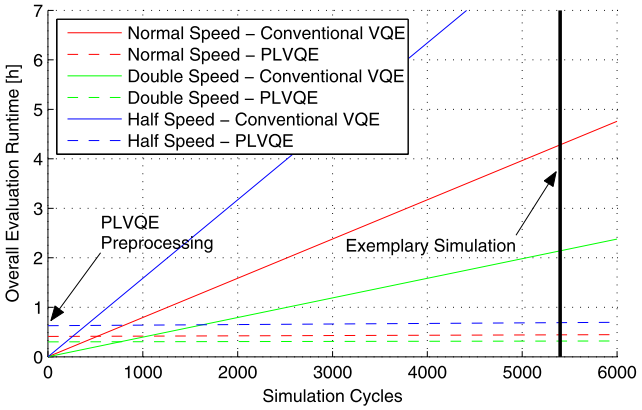

Fig. 8 Interpolation of overall evaluation runtime for different decoder speeds

are larger then the deviations when using the GOP-based approach only. The reason for deviations of the GOP-based approach is related to the assumption made in Sect. 3.1 concerning error propagation. The video quality of a GOP does not solely depend on itself, but also predicts from temporally preceding frames, which can lead to differences of video quality through propagating errors. During preprocessing phase, the GOP-based measurements of PLVQE are taken on bitstreams with a specific REP repeatedly mapped onto them. Therefore, the frames used for prediction at time of preprocessing slightly differ from the corresponding frames at simulation time that might be affected by random transmission errors.

5.2 Runtime savings

Runtime savings of the conducted simulation set are evaluated by comparing the runtime behavior of the PLVQE with the runtime behavior of the conventional VQE approach. Our experiments indicated that, on average, conventional VQE takes more than 100 times the runtime of online PLVQE evaluation with the given hardware and simulation setup. For a complete comparison, the VQE database creation during the offline preprocessing phase has to be considered as well. The necessary operations for each REP in the GOP-based approach of PLVQE, i.e. bitstream reconstruction, video decoding, and PSNR measurement, are very similar to the conventional VQE procedure. Therefore, we assume an average preprocessing runtime per REP equal to the average runtime of conventional VQE per simulation cycle. As shown in Fig. 3, 278 REPs have to be evaluated for the given video coding setup in the offline PLVQE preprocessing. 277 of all REPs will be evaluated with the GOP-based approach. For the remaining single REP indicating a total loss of video frames within the GOP, the VQE database is extended with measurements for long lasting freeze frame, as described in Sect. 3.2.

The achievable time-savings depend on numerous factors. First, the number of simulation cycles is crucial as PLVQE is not beneficial when the number of simulation cycles is smaller than the number of REPs processed to acquire the VQE database. All simulations performed with the presented framework included a multitude of parameters, e.g. FEC schemes, FEC code rates, or channel parameters as in [17], making PLVQE very attractive for simulations with a large number of cycles or iterations. Second, the number of REPs required for a video coding setup determines the size of the VQE database, which linearly affects the necessary runtime for offline preprocessing. Third, the underlying decoder implementation largely influences the runtime, e.g. experiments with a set of decoder implementations led to large differences of evaluation runtime with the given video coding setup. At last, the video coding setup regarding bitrate, resolution, and framerate affects the evaluation runtime depending on decoder implementation and the given hardware. Figure 8 shows an extrapolation of overall runtime results based on the given simulation setup, referred to as normal speed. For the given simulation setup and hardware, PLVQE reduces evaluation runtime by 89.7% compared to the conventional approach. It can be seen that the achievable gain mainly depends on the amount of simulation cycles. Furthermore, a change of decoding speed has been considered, denoted as double and half speed. Both lines illustrate the evaluation runtime for varying decoder speeds due to a change of video coding (e.g. video resolution or data rate) or decoder implementation.

The speed of the underlying decoder influences the gradient of the conventional VQE runtime behavior and the size of the initial preprocessing offset of PLVQE, which is introduced by the VQE database creation. A change in coding

structure alters the initial offset only and, therefore, affects the number of simulation cycles at which PLVQE breaks even with the conventional approach, i.e. the intersection of similarly colored lines in Fig. 8.

6 Conclusion

This work presents and analyzes an approach for fast Packet Level Video Quality Evaluation (PLVQE) of error-prone H.264/AVC and SVC transmission with application layer metrics. Simulation time savings result from reduction of redundancy by combining decoding operations and exploiting prediction structures within H.264/AVC and SVC coded video. An offline preprocessing of video data constitutes a Video Quality Evaluation (VQE) database that allows online trace-driven packet level evaluation of simulation results with application layer video quality metrics such as PSNR. The conducted validation based on exemplary simulations proved enormous benefit with a reduction of the evaluation runtime of almost 90% and an only marginal deviation of results compared to the time-consuming conventional VQE approach that includes bitstream reconstruction, decoding, and VQE measurements of each simulation cycle. Moreover, the analysis showed that the time savings of the proposed platform in the overall evaluation scales with the amount of simulation cycles and the speed of the decoder implementation. This makes the presented approach favorable for large simulation sets and video data that demands relatively high computational power such as used in HDTV applications.

Acknowledgements The presented work has been supported by the European Commission under contract number FP7-ICT-248036, project COAST.

References

1. Wiegand T, Sullivan G, Bjontegaard G, Luthra A (2003) Overview of the H. 264/AVC video coding standard. IEEE Trans Circuits Syst Video Technol 13:560–576
2. Schwarz H, Marpe D, Wiegand T (2007) Overview of the scalable video coding extension of the H. 264/AVC standard. IEEE Trans Circuits Syst Video Technol 17:1103–1120
3. Klaue J, Rathke B, Wolisz A (2003) EvalVid—a framework for video transmission and quality evaluation. In: Computer performance. Lecture notes in computer science, vol 2794. Springer, Berlin, pp 255–272. doi:10.1007/978-3-540-45232-4_16
4. Ke C-H, Lin C-H, Shieh C-K, Hwang W-S (2006) A novel realistic simulation tool for video transmission over wireless network. In: International conference on sensor networks, ubiquitous, and trustworthy computing, vol 1, pp 275–283
5. Lie A, Klaue J (2008) EvalVid-RA: trace driven simulation of rate adaptive MPEG-4 VBR video. Multimed Syst 14:33–50. doi:10.1007/s00530-007-0110-0
6. Ke C, Shieh C, Hwang W, Ziviani A (2008) An evaluation framework for more realistic simulations of MPEG video transmission. J Inf Sci Eng 24:425–440
7. Le TA, Nguyen H, Zhang H (2010) EvalSVC—an evaluation platform for scalable video coding transmission. In: 14th international symposium on consumer electronics, Braunschweig, Germany, June 2010
8. Migliorini D, Mingozzi E, Vallati C (2010) QoE-oriented performance evaluation of video streaming over WiMAX. In: Wired/wireless Internet communications, pp 240–251
9. Kondrad L, Bouazizi I, Vadakital V, Hannuksela M, Gabbouj M (2009) Cross-layer optimized transmission of h.264/SVC streams over dvb-t2 broadcast system. In: IEEE international symposium on broadband multimedia systems and broadcasting. BMSB'09, May 2009, pp 1–5
10. Reibman A, Vaishampayan V, Sermadevi Y (2004) Quality monitoring of video over a packet network. IEEE Trans Multimed 6:327–334
11. Tao S, Apostolopoulos J, Guerin R (2008) Real-time monitoring of video quality in IP networks. IEEE/ACM Trans Netw, 16:1052–1065
12. Liang Y, Apostolopoulos J, Girod B (2003) Analysis of packet loss for compressed video: does burst-length matter. In: IEEE international conference on acoustics, speech, and signal processing. Proceedings (ICASSP'03), April. vol 5, pp V–684–7
13. Stuhlmüller K, Farber N, Link M, Girod B (2000) Analysis of video transmission over lossy channels. IEEE J Sel Areas Commun 18:1012–1032
14. Li Z, Chakareski J, Niu X, Zhang Y, Gu W (2009) Modeling and analysis of distortion caused by Markov-model burst packet losses in video transmission. IEEE Trans Circuits Syst Video Technol 19:917–931
15. Skupin R, Hellge C, Schierl T, Wiegand T (2010) Fast application-level video quality evaluation for extensive error-prone channel simulations. In: 15th IEEE international workshop on computer aided modeling, analysis and design of communication links and networks (CAMAD), Dec 2010, pp 6–10
16. Liebl G, Tappayuthpijarn K, Grüneberg K, Schierl T, Keip C, Stadali H (2010) Simulation platform for multimedia broadcast over DVB-sh. In: Proceedings of the 3rd international ICST conference on simulation tools and techniques, SIMUTools'10, ICST, Brussels, Belgium, pp 84:1–84:10.
17. Hellge C, Gómez-Barquero D, Schierl T, Wiegand T (2010) Intra-burst layer aware FEC for scalable video coding delivery in DVB-h. In: 2010 IEEE international conference on multimedia and expo (ICME), July 2010, pp 498–503
18. Schwarz H, Marpe D, Wiegand T (2006) Analysis of hierarchical b pictures and MCTF. In: IEEE international conference on multimedia and expo, July 2006, pp 1929–1932
19. Hong D, Horowitz M, Eleftheriadis A, Wiegand T (2010) H.264 hierarchical p coding in the context of ultra-low delay, low complexity applications. In: Picture coding symposium (PCS), Dec 2010, pp 146–149
20. Wang Z, Bovik A, Lu L (2002) Why is image quality assessment so difficult. In: IEEE international conference on acoustics speech and signal processing, vol 4. Springer, Berlin, pp 3313–3316. IEEE, New York 1999
21. Girod B (1993) What's wrong with mean-squared error. In: Digital images and human vision. MIT Press, Cambridge, pp 207–220
22. ITU-T (2008) International Telecommunication Union, Geneva, Switzerland, Recommendation J.247—objective perceptual multimedia video quality measurement in the presence of a full reference.
23. Brunnstrom K, Hands D, Speranza F, Webster A (2009) VQeg validation and ITU standardization of objective perceptual video quality metrics [Standards in a Nutshell]. IEEE Signal Process Mag 26:96–101
24. Winkler S (2010) Video quality measurement standards—Current status and trends. In: 7th international conference on information,

communications and signal processing. ICICS 2009. IEEE, New York, pp 1–5
25. Engelke U, Zepernick H (2007) Perceptual-based quality metrics for image and video services: a survey. In: 3rd EuroNGI conference on next generation Internet networks. IEEE, New York, pp 190–197
26. Opticom P (2005) Advanced perceptual evaluation of video quality.
27. Wang Z, Lu L, Bovik AC (2004) Video quality assessment based on structural distortion measurement. Signal Process Image Commun 19:121–132
28. ITU-T International Telecommunication Union, Geneva, Switzerland (2008) Recommendation G.826—End-to-end error performance parameters and objectives for international, constant bit rate digital paths and connections
29. DVB, Digital Video Broadcasting (2010) DVB-SH Implementation Guidelines Issue 2, DVB Document A120
30. Uitto M, Vehkaperä J (2009) Spatial enhancement layer utilisation for SVC in base layer error concealment. In: Proceedings of the 5th international ICST mobile multimedia communications conference, Mobimedia '09, ICST, Brussels, Belgium, pp 10:1–10:7.
31. JVT (2009) SVC reference software JSVM (joint scalable video model) 9.17.
32. Wenger S, Wang Y-K, Schierl T, Eleftheriadis A (2011) Rfc6190: Rtp payload format for SVC video. In: Internet engineering task force (IETF).
33. Mushkin M, Bar-David I (2002) Capacity and coding for the Gilbert-Elliot channels. IEEE Trans Inf Theory 35:1277–1290
34. Issariyakul T, Hossain E (2007) Introduction to Network Simulator 2 (NS2). pp. 1–18
35. Ellis M, Perkins C, Pezaros D (2011) End-to-end and network-internal measurements on real-time traffic to residential users. In: Proc of ACM multimedia systems