

Fast Application-level Video Quality Evaluation for Extensive Error-Prone Channel Simulations

Robert Skupin¹, *Student Member, IEEE*, Cornelius Hellge^{1,2}, *Member, IEEE*,
Thomas Schierl^{1,2}, *Member, IEEE*, and Thomas Wiegand^{1,2}, *Member, IEEE*

¹Fraunhofer HHI, Image Processing Department, Germany

²Technische Universität Berlin, Image Communication Group, Germany

{robert.skupin, cornelius.hellge, thomas.schierl, thomas.wiegand}@hhi.fraunhofer.de

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

Index Terms— Network Simulation, Video Quality Evaluation, H.264/AVC, SVC, Mobile TV

I. INTRODUCTION

Video transmission over error prone channels, such as typical for Mobile TV or IPTV systems, is constantly subject to research. This is due to the progress of involved components, changing requirements, or new user demands. A wide range of parameters influences the overall system performance. In the physical layer, modulation techniques, interleaving schemes, or physical layer forward error correction (FEC) settings allow adaptation to the characteristics of different transmission systems. In the application layer, the state of the art video codec standards H.264/AVC and its scalable video coding (SVC) extension [1] offer numerous tools to adjust video coding to the specific requirements of an application or service. Furthermore, application layer FEC schemes, transmission scheduling, or transport protocols have an effect on the performance. H.264/AVC and especially SVC coded video provide the opportunity to increase protection of more important parts of data, which is addressed by FEC schemes such as unequal error protection (UEP) or Layer-Aware FEC [2]. Fine-tuning of channel parameters, media coding and error protection is vital to achieve the best system performance and an optimal user experience.

Simulation of video transmission systems is an important instrument to evaluate the overall system performance in the first place, but the multitude of parameters often results in vast

simulation sets. Different approaches have been made to provide a realistic and adequate simulation platform for video transmission. The EvalVid framework [3] and its extensions allow the evaluation of H.264/AVC video transmission with application-level metrics such as peak signal-to-noise ratio (PSNR), instead of relying on network-level metrics such as packet error rate that are inadequate to evaluate system performance, especially for transmission of SVC coded video. However, EvalVid currently lacks SVC support and the conventional approach to decode each transmission result for video quality evaluation (VQE) makes simulations particularly time-consuming. Models to estimate the additional distortion from packet losses without decoding are beneficial when limited processing power makes the conventional approach unfeasible, but they still have individual weaknesses regarding accuracy or significance.

In this paper, we present a mechanism for fast evaluation of extensive error-prone H.264/AVC and SVC video transmission on packet level. Our approach significantly reduces the overall simulation time by eliminating redundancy in the evaluation phase. In order to allow fast application-level VQE of transmission results, a VQE database is constituted in a preprocessing phase by combining and reducing redundant calculations that are usually carried out in each simulation cycle. Exploitation of the prediction structure of H.264/AVC and SVC coded video leads to a further reduction and a significant speed up. The presented mechanism has already been successfully used in the simulations within the context of SVC for mobile satellite transmission [4] and in investigations of different FEC schemes [2].

The remainder of this paper is organized as follows: first, Section II provides a system overview and details on media coding, transmission simulation and VQE. Section III describes the proposed approach for fast application-level VQE. A validation of the proposed approach is presented in Section IV, followed by the conclusion in Section V.

II. SYSTEM OVERVIEW

The structure of the simulation platform used to implement the presented approach closely resembles the different tasks that come along with video transmission, namely encoding,

transmission and evaluation, as can be seen in Fig. 1.

The encoding phase consists of a rate-controlled video encoder that features chunk-wise encoding of a continuous test sequence. Chunks that match the simulation criteria in terms of bitrate or quality can subsequently be concatenated into a continuous bitstream. The MP4 file format [5] is utilized for RTP packetizing and extraction of a so-called packet trace file, which is a textual description of the RTP packets and the packetized video data. The packet trace serves as input for a trace-driven transmission system simulation that leads to a possibly erroneous packet-trace with missing lines, resembling transmission errors of certain packets. The evaluation step is referred to as Virtual Video Decoder (VVD). It consists of a preprocessing phase, in which original source and compressed video are analyzed to acquire a VQE database. This database allows a packet-level trace evaluation of error-prone transmission results later on. The conventional approach of evaluation including bitstream reconstruction, decoding and VQE measurement is optional. Further details on the rate-controlled encoding mechanism can be found in [4].

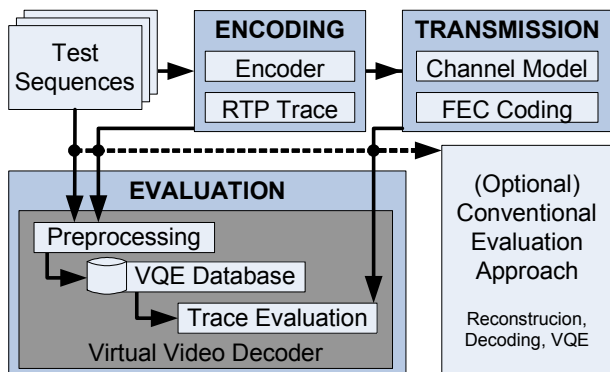


Fig. 1: Schematic illustration of a video transmission simulation platform.

A. Media Coding

The applied media codecs, H.264/AVC and SVC, are state of the art block oriented motion compensation based codec standards used for a variety of video transmission systems today. First introduced, 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 and quantized transform coding. Coded VCL frame data and additional information are further processed in the NAL by encapsulation in so-called VCL-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 such as the MP4 file format.

The SVC extension of H.264/AVC allows further structuring the bitstream and extracting different video representations of one 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. The scalability functionalities of SVC have great potential to achieve a more efficient and flexible provisioning of Mobile TV services. Compared to using a simulcast approach, where the same content is delivered multiple times at different video resolutions, SVC provides efficient means to cope with heterogeneous receiver capabilities (screen size and processing power) and extending existing services in a backwards compatible way.

Fundamental details of H.264/AVC and SVC for the presented approach are hierarchical prediction and the group of pictures (GOP) structure [6], as illustrated in Fig. 2. Hierarchical prediction refers to the concept of providing temporal scalability with the use of hierarchical B frames that predict from temporal preceding and succeeding frames with lower temporal level. A set of frames between two successive video frames of the lowest temporal layer with the succeeding lowest temporal layer picture constitutes a GOP structure. SVC coded video extends the GOP with further representations of video frames.

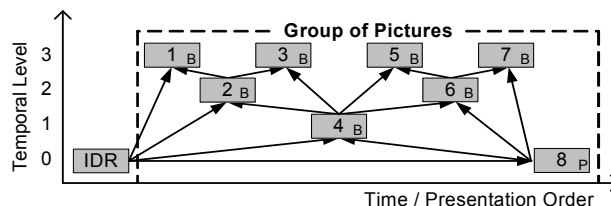


Fig. 2: Illustration of hierarchical prediction within a H.264/AVC group of pictures (GOP) with a single slice per frame.

B. External Transmission Simulation

In order to simulate a specific transmission system, appropriate channel models have to be chosen. For instance, a service provided via ADSL has to cope with channel characteristics that are different from those of a mobile broadcast channel such as DVB-SH. Different 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 not. A classic link/application layer model is the Gilbert Elliot model that consists of a varying binary symmetric channel with crossover probabilities determined by a binary-state Markov process. Otherwise, binary erasure channel (BEC) models that simulate channel behavior down to the physical layer have to be considered. Typical physical layer models are the additive white Gaussian noise (AWGN) model, or the Typical Urban 6-tap (TU6) channel, which features six paths with different attenuation and delay. The latter was found to be representative for typical mobile transmission scenarios [7] and is used in the validation section for exemplary simulations of a DVB-H mobile broadcast scenario.

Channel coding addresses transmission issues at different layers on transmitter and receiver side. FEC codes protect data against transmission errors by adding redundancy, which enables the receiver to detect and correct transmission errors. Different FEC codes offer beneficial protection for transmission in error-prone channels, e.g. low-density-parity-check (LDPC) codes in DVB-T2/S2 and turbo codes in DVB-SH on physical layer or Reed Solomon and Raptor codes on link or application layer [8]. The highly structured data of H.264/AVC and SVC coded video allows stronger protection of data with higher priority. This can be utilized by priority aware FEC schemes such as Layer-Aware FEC that create connected repair symbols across different SVC layers. Depending on the simulation depth, further techniques such as interleaving, modulation, or multiplexing have to be considered. Various generalized tools, e.g. Network Simulator 2 or simulators for specific systems can be used for trace-driven simulations of IPTV, peer-to-peer applications or DVB-H/SH mobile broadcast services [4][9].

C. Video Quality Evaluation

As coding and transmission may introduce a distortion into the processed video, the non-trivial task of VQE is an important instrument to evaluate compression efficiency or transmission system performance. Statistically relevant results require subjective tests with a large test population, which is rather costly, time consuming and not adequate for large simulation sets. Numerous metrics offer VQE that corresponds to the characteristics of human perception, as can be found in ITU-T recommendation J.247 [10]. PSNR is the ratio of maximum pixel value within a frame to the corrupting noise that affects its coded representation. The corrupting noise is derived from the mean square error of pixel values between the original and coded frame. Its clear physical meaning and simple calculation made PSNR the commonly used VQE metric, although it can only be seen as an approximation of the human visual perceptions behavior and therefore fails to match results of subjective tests in certain respects.

Apart from calculating sheer pixel differences among original and coded video frames, other metrics utilize known characteristics of human perception to a higher degree. Structural Similarity Index (SSIM) or Perceptual Evaluation of Video Quality (PEVQ), 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, often with a significant increase in computational complexity compared to PSNR, but still failing to match the human visual perception exactly. The amount of erroneous and decoded frames is simple to compute and is a meaningful indicator that can be used to calculate the Erroneous Seconds Ratio (ESR). The simulations presented in this paper use the well established combination of PSNR and decodable frame counts.

III. VIRTUAL VIDEO DECODER

The proposed Virtual Video Decoder (VVD) provides an evaluation of transmitted video without the need to decode each simulation cycle result. The main idea is to allow

application-level quality evaluation on packet level by precalculating a VQE database that covers quality measurements for all possible video outputs. This database is created during a preprocessing phase, in which decoding and evaluation operations that are usually conducted in each simulation cycle are combined to omit redundant calculations. Considering the prediction structure of H.264/AVC and SVC coded video allows to eliminate unnecessary further calculations. In the evaluation phase, transmitted video sequences can be evaluated using the pre-calculated VQE database. Packet losses are analyzed and mapped to the corresponding VQE values in the database. Thus, after preprocessing, vast simulations can be evaluated in a very short time without any decoding operation.

The proposed approach requires decoder implementation and media coding to fulfill certain constraints to reduce complexity and processing time. An error resilient decoder implementation, which is compatible with the H.264/AVC and SVC standard, is used during creation of the VQE database. Basic error concealment techniques include base layer upsampling for loss of SVC enhancement layer data and the insertion of freeze frames in case of frame loss to keep video output in sync [11]. Further constraints concerning the video coding are a known prediction structure and limitation to a small number of slices per frame to reduce necessary calculation and achieve a reasonable complexity.

A. Relevant Error Patterns

Fig. 3 shows an exemplary error distribution within a single-layer H.264/AVC GOP structure. The frames are numbered in presentation order and vertically sorted according to their temporal level. The arrows represent the dependencies between individual frames that arise from the hierarchical structure used for temporal prediction from surrounding frames. SVC introduces additional dependencies across layers. Solid and striped symbols illustrate non-decodable frames due to transmission errors. There are 2^n possible combinations of erroneous frames within a GOP, where n is the number of frame representations within the GOP for SVC coded video or the GOP size in case of H.264/AVC. The depicted GOP structure allows $2^8 = 256$ error combinations. Taking inter-frame (and inter-layer in case of SVC) dependencies into account significantly reduces 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 depicted GOP structure belong to this category and are referred to as initial errors.

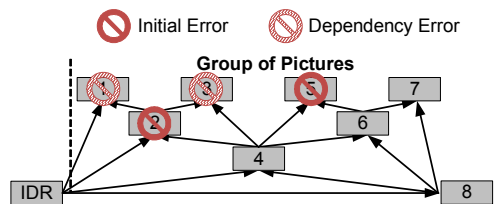


Fig. 3: Illustration of erroneous frames within a GOP structure.

Initial errors are always caused by transmission errors that directly affect the frames NAL units. The second category contains dependency errors, which are not decodable due to missing reference data from other frames. Frame 1 and frame 3 are not decodable due to partially missing reference data in frame 2. Even in case of correctly received corresponding NAL units, frame 1 and frame 3 are not decodable and therefore belong to the second category. Dependency errors can but do not need to be affected directly by transmission errors. Since the resulting video output is identical for error combinations that consist of the same initial errors, only considering initial error combinations is sufficient to cover all transmission errors. Processing these relevant error patterns (REP) reduces the number of necessary decoding

TABLE 1: OVERVIEW OF ERROR COMBINATIONS, REP AND CALCULATION SAVINGS FOR DIFFERENT VIDEO CODING SETUPS AND GOP SIZES. BL = BASE LAYER, EL = ENHANCEMENT LAYER.

Video Coding Setup	GOP size		Number of Error Combinations	Number of REP	Savings
	BL	EL			
H.264/AVC	8	-	$2^8 = 256$	27	89,5%
H.264/AVC	16	-	$2^{16} = 65536$	678	99,0%
SVC	8	8	$2^{16} = 65536$	278	99,8%
SVC	16	16	$2^{32} = 4294967296$	51318	99,9%
SVC Temporal Scalability	4	8	$2^{12} = 4096$	51	98,8%
SVC Temporal Scalability	8	16	$2^{24} = 16777216$	1763	99,9%

operations significantly, as can be seen in Table 1.

B. Preprocessing

To generate a VQE database of PSNR measurements, the preprocessing of a given coded video sequence utilizes the previously described error patterns. With information about timestamps, NAL types and sizes, each REP can be mapped to the corresponding NAL units within all GOPs of the video sequence in order to create an erroneous version of the video corresponding to the REP. NAL units unaffected by initial errors of the REP or dependency errors due to the use of prediction are extracted and concatenated to reconstruct an erroneous bitstream. This bitstream can subsequently be 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 additional PSNR measurements for IDR frames in the beginning of each chunk and a unique REP identifier.

Since the above way to calculate PSNR values of erroneous streams is GOP-based, its accuracy degrades when evaluating a complete loss of video signal that exceeds the duration of a GOP and leads to a long period of freeze frames. In this case another technique is used in parallel to extend the database. All frame representations are compared to the following original frames to achieve an evaluation of long-lasting freeze frames. As PSNR measurement of frames with different content gives to some extent arbitrary results, further calculations can be omitted by using a constant value. Our simulations indicated that meaningful values lie within a range

of 10dB to 15dB strongly depending on the video content.

C. Trace Evaluation

In order to evaluate a simulation cycle, packet traces of transmission results are analyzed on packet-level and erroneous packets are mapped to corresponding video data. A GOP-wise analysis of all occurred transmission errors with knowledge of the coded video structure allows identifying the initial errors. Information about initial errors is used to compose a unique REP identifier to query corresponding PSNR measurements from the database, which are subsequently averaged and combined with the count of erroneous and decoded frames.

IV. VALIDATION

The benefit of the presented approach relies both on the achievable time-savings and the accuracy of results. These goals are examined based on an exemplary simulation by comparing results of VVD with optional reconstruction and evaluation. The test sequence “soccer” with a length of 30 seconds was encoded according to a restricted version of the scalable high profile with an approximately constant bitrate and a single slice per frame. The SVC bitstream consists of an H.264/AVC compatible QVGA base layer at 12.5 fps with 34.7 dB PSNR and a VGA enhancement layer at 25 fps with 35.1 dB leading to an overhead of roughly 7.5% compared to the corresponding single layer VGA stream. One IDR frame plus three GOP structures of 8 frames per chunk result in a random access point rate of 1 sec. Channel simulation was conducted using a DVB-H System-Level Simulator [9]. The simulations included different Doppler frequencies, a carrier-to-noise-ratio (CNR) range resembling correlated shadowing, several FEC schemes, FEC code rate distributions and iterations to gather statistically consistent results.

A. Time Savings

The range of parameters leads to roughly 20000 simulation runs, equivalent to about 170 hours of video, which results in 170 hours of decoding with a real-time decoder plus additional 10 hours to reconstruct bitstreams from packet traces. As can be seen in Table 1, using the VVD a total of 278 erroneous versions, equivalent to only about 3 hours of video, have to be processed to constitute a VQE database. Approximately an hour is needed to create the corresponding erroneous bitstreams plus two hours to extend the database for long-lasting freeze frames. The following evaluation of transmission results can be done at comparatively high speeds of up to 6000 fps on standard PC hardware. Thus, evaluation of the simulation results takes less than 7 hours compared to about 180 hours required to evaluate reconstructed bitstreams with a real-time decoder. This leads to a significant speed-up of more than 90 percent of the overall evaluation process.

Based on the given simulation, Fig. 4 illustrates the estimated time of overall evaluation with VVD compared to the conventional approach for a range of simulation cycles and reference decoder speeds. Since the duration of the VQE

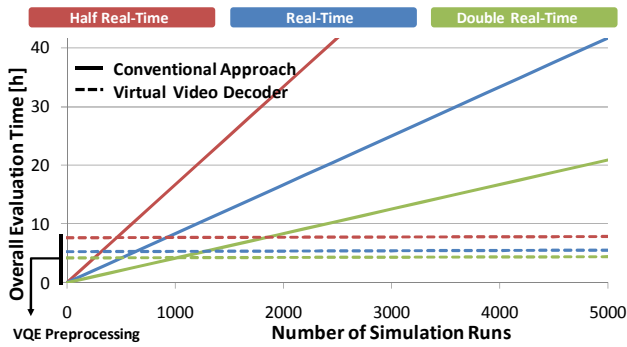


Fig. 4: Overall evaluation time of Virtual Video Decoder compared to evaluating each reconstructed transmission result with Real Video Decoder for different reference decoder speed and number of simulations.

preprocessing phase depends on video decoder speed and length of the video, overall VVD evaluation time only increases slightly whereas the benefit compared to reconstruction significantly grows with the amount of simulation cycles and decreasing reference decoder speed.

B. Accuracy

A comparison of VVD results with a conventional VQE approach serves as a basis for the evaluation of the accuracy of the presented approach. For reconstruction, erroneous transmitted NAL units are discarded from the original bitstream and the resulting erroneous bitstreams are decoded and evaluated. For simulation cycles with error-free data transmission, VVD results match the results of the conventional approach precisely. A loss of enhancement layer data in case of SVC, where VVD relies completely on the GOP-based approach, showed that there is practically no deviation between results. Table 2 gives an analysis on these particular simulations. For simulations with freeze frame error concealment due to loss of H.264/AVC frames or SVC base layer frames, the deviation of VVD results increases slightly but does not rise to a notable magnitude and is negligible. Considering the benefit in terms of time consumption compared to the conventional VQE approach, the VVD shows a more than satisfying accuracy.

TABLE 2: OVERVIEW OF VVD RESULTS DERIVATION FOR DIFFERENT FRAME LOSSES, EL = ENHANCEMENT LAYER.

Percentage of lost EL Frames	Deviation of VVD results			
	Max [dB]	Min [dB]	Average [dB]	σ [dB]
20%	0.0003	-0.0045	-0.0021	0.0024
30%	0.0035	0.0023	0.0032	0.0005
40%	0.0042	-0.0049	0.0000	0.0024
50%	0.0035	-0.0048	0.0011	0.0024
60%	0.0047	-0.0028	0.0024	0.0019
70%	0.0050	-0.0049	0.0002	0.0028
80%	0.0046	-0.0050	-0.0001	0.0029
90%	0.0038	-0.0046	-0.0004	0.0027

V. CONCLUSION

This work presents an approach for fast application-level video quality evaluation of error-prone H.264/AVC and SVC transmission. Time savings arise from reduction of redundancy by combining decoding operations and exploiting prediction structures within H.264/AVC and SVC coded video. A preprocessing of video data constitutes a video quality evaluation (VQE) database that allows trace-driven packet-level evaluation of simulation results with application-level metrics such as PSNR, decodable frame counts, and others.

The conducted validation based on exemplary simulations proved enormous benefit with a reduction of the evaluation runtime of more than 90 percent and an insignificant deviation of results compared to the time-consuming conventional VQE approach that includes bitstream reconstruction, decoding and VQE measurements of each transmission result. Moreover, the analysis showed that the benefit of the proposed platform in terms of time savings in the overall evaluation scales with the amount of simulations and decreasing reference decoder speed, making the presented approach favorable for large simulation sets and slowly decodable video data such as in HDTV applications. Opportunities for further work include the extension of the approach for 3-dimensional multi view coded (MVC) video or the support of a larger number of SVC layers.

ACKNOWLEDGEMENTS

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

REFERENCES

- [1] "Advanced Video Coding for Generic Audiovisual Services", ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, Version 11 (including SVC extension): Approved in March 2009.
- [2] C. Hellge, et al., "Inter Burst LA-FEC for Scalable Video Coding Delivery in DVB-H", ICME'10, Singapore, July 2010.
- [3] J. Klauke, B. Rathke, and A. Wolisz, "EvalVid - A Framework for Video Transmission and Quality Evaluation", 13th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation, pp. 255-272, Urbana, Illinois, USA, Sep.2003.
- [4] G. Liebl, et al., "Simulation Platform for Multimedia Broadcast over DVB-SH", SIMUTools'10, Torremolinos, Malaga, Spain, March 2010.
- [5] "ISO/IEC. Information technology – Coding of audio-visual objects – Part 15: Advanced Video Coding (AVC) File format AMENDMENT 2: File format support for Scalable Video Coding (SVC)", 14496-15:2004/Amd2, 2008.
- [6] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF", IEEE ICME'06, Toronto, Canada, July 2006.
- [7] G. Faria, et al., "DVB-H: Digital Broadcast Services to Handheld Devices," Proc. of the IEEE, vol. 94, no. 1, pp. 194-209, Jan. 2006.
- [8] DVB Bluebook A148, "Upper Layer FEC in DVB", March 2010.
- [9] D. Gómez-Barquero et al., "Development and Applications of a Dynamic DVB-H System-Level Simulator", to appear in IEEE Transactions on Broadcasting, vol. 56, issue 3, 2010
- [10] "Objective perceptual multimedia video quality measurement in the presence of a full reference", ITU-T Rec. J.247 (08/08).
- [11] M. Uitto, and J. Vehkaperä, "Spatial Enhancement Layer Utilisation for SVC in Base Layer Error Concealment", MobiMedia'09, London, United Kingdom, June 2009.