Simulation Platform for Fast Video Quality Evaluation of H.264/AVC and SVC Transmission on Error-Prone Channels

Robert Skupin[1], Cornelius Hellge[1,2], Karsten Grüneberg[1,2], Thomas Schierl[1,2], and Thomas Wiegand[1,2]

[1] Fraunhofer HHI, Image Processing Department, Germany
[2] Technische Universität Berlin, Department of Telecommunications Systems, Germany
{robert.skupin,cornelius.hellge,karsten.grueneberg, thomas.schierl,thomas.wiegand}@hhi.fraunhofer.de

Abstract. Video transmission systems for IPTV, mobile TV or conversational video communication services are constantly subject to investigations regarding service reliability, quality of service, quality of experience, or bandwidth efficiency. The multitude of influencing parameters in terms of channel settings, error protection or media coding turns optimal setup into a challenge. Simulating video transmission is an important instrument to assess new developments and optimize settings in the first place, but existing solutions demand high computational capacity to provide application-level evaluations, such as peak signal-to-noise ratio. The required video decoding consumes a substantial part of the overall simulation runtime. In this paper, we present a simulation platform that utilizes the architecture of H.264/AVC and SVC to provide a fast and accurate video quality evaluation on packet level with application-level metrics. The presented approach significantly reduces overall runtime in the evaluation phase of large simulation sets. The underlying mechanisms are explained and their benefit in terms of time savings and accuracy is analyzed based on an exemplary simulation of a mobile broadcast scenario.

Keywords: network simulation, video quality evaluation, SVC, mobile TV, IPTV, conversational video communication.

1 Introduction

Due to the continuous progress of their components, video transmission systems for IPTV, mobile TV, conversational video communication services and others are constantly subject to investigations regarding service reliability, quality of service, quality of experience, or bandwidth efficiency. The amount of influencing parameters turns optimal setup into a challenge. Various channel parameters have an influence on transmission reliability, e.g. hierarchical modulation schemes increase the robustness of definable signal components and interleaving techniques protect data against burst errors by distributing transmission errors over non consecutive data bits. In terms of

J. Rodriguez, R. Tafazolli, C. Verikoukis (Eds.): MOBIMEDIA 2010, LNICST 77, pp. 634–647, 2012.

media coding, the state of the art video codec standard H.264/AVC and its scalable video coding (SVC) extension [1] are used for a variety of video transmission systems nowadays. Both offer numerous tools to optimize coding for specific transmission scenarios with respect to coding delay, error resilience, rate distortion performance and others. Transmission on error-prone channels generally relies on forward error correction (FEC) schemes to protect data against transmission errors. H.264/AVC and especially SVC coded video with its layered nature endorse stronger protection of more important parts of video data, which is referred to as unequal error protection (UEP). Advanced FEC schemes such as Layer-Aware FEC [2] utilize this fact by generating connected repair symbols across layers, exploiting the layered nature of SVC further.

Best possible channel setup, error protection and media coding is vital to achieve optimal user experience. Simulation is an important instrument to evaluate effects of advancements in single components on the overall system, but the multitude of parameters and the need to gather statistically consistent results require vast simulation sets to be conducted. Furthermore, computational complexity and therefore simulation runtime usually corresponds to video resolution and bitrate, making simulations with high-definition or 3-dimensional video data particularly time-consuming. Different approaches have been made to provide a realistic and adequate simulation platform for video transmission. The EvalVid framework [3] and its numerous extensions [4][5] 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 perceived quality by end users, especially for SVC coded video. However, EvalVid currently still lacks SVC support and the conventional approach to individually decode each transmission result for performance evaluation makes simulations time-consuming and computationally complex. Different models have been proposed in order to estimate the additional distortion of coded video data on packet level without decoding video. These models are beneficial where a conventional performance evaluation approach is unfeasible due to the limited computational power or missing reference data, but still have individual weaknesses regarding accuracy or significance. Models and metrics for the MPEG-2 and MPEG-4 coding standards in different transmission systems were introduced in [6] and [7] and the effects of burst-error propagation were examined in [8].

In this paper, we present a simulation platform that takes a different approach to provide video quality evaluation based on actual measurements of erroneous video sequences instead of estimating additional video distortion. The presented approach reduces redundancy and evaluation phase runtime by preprocessing and exploiting prediction structures within H.264/AVC and SVC coded video. Thereby, the overall simulation time is reduced while providing a fast and accurate application-level video quality evaluation of transmission results on packet level. On encoding side, the architecture of H.264/AVC and SVC is utilized to provide coded video at variable (VBR) or constant bitrates (CBR) in a flexible way by chunk-wise offline encoding. The presented simulation platform has already been put to use successfully in context of research projects on SVC in mobile satellite based hybrid networks [9] and further research on FEC schemes [2]. The focus of this publication is on the media coding section and the designed quality evaluation mechanism. Hence the transmission simulation and the underlying channel model are not stressed intensely, as they can be adjusted according to the simulation objectives, making the presented approach universally applicable. The remainder of this paper is organized as follows: First of all, Section 2 provides a system overview, including details on video coding, channel simulation and quality evaluation. Next, the implemented video encoding method is explained in Section 3. Section 4 describes the proposed video decoding mechanism for fast application-level video quality evaluation. A validation of the proposed platform is presented in Section 5, followed by a summary in Section 6.

2 System Overview

The main objective of the proposed simulation platform is a significant speed up of overall simulation time by providing a fast and accurate video quality evaluation. Its design closely resembles the different tasks that come along with video transmission, as depicted in Fig. 1. It also distinguishes clearly between static and dynamic parts, i.e. modules that perform offline calculations and those that operate dynamically during simulations. The offline media encoding phase, in which uncompressed test video sequences are encoded using a simple rate control, is referred to as Virtual Video Encoder (VVE). Rate control is achieved by chunk-wise encoding and selection in regard of simulation objectives. The encoding phase leads to coded video data and a textual description of the packetized coded video in form of a so-called packet trace file. During the preprocessing phase of the Virtual Video Decoder

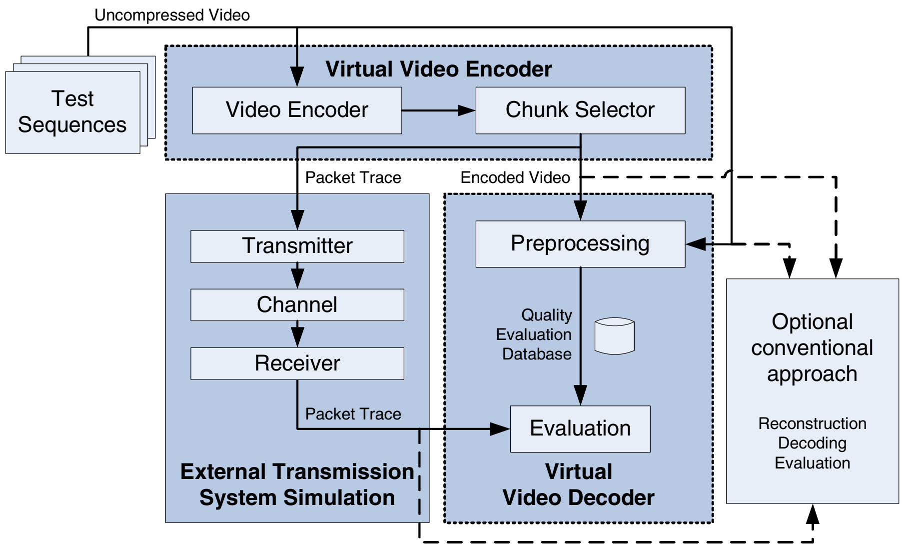

Fig. 1. Schematic illustration of the proposed simulation platform

(VVD), coded video data is used to acquire a database for quality evaluation in a way that exploits the nature of H.264/AVC and SVC and omits redundant operations. The textual description serves as basis for an external transmission system simulation that operates only on packet traces. The resulting packet trace files can subsequently be analyzed and evaluated on packet level with application-level metrics in the evaluation phase of the VVD via the previously created database. Optionally, simulated packet traces can be evaluated with the conventional approach of trace-to-bitstream reconstruction, video decoding and quality evaluation, as implemented in [3].

2.1 Media Coding

The applied media coding, H.264/AVC and SVC, are state of the art block oriented motion compensation based codec standards used for a variety of broadcast video transmission systems today. First introduced H.264/AVC achieves significant improvements in coding efficiency compared to prior 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 like Real-time Transport Protocol (RTP) internet services and MPEG-2 transport streams or storage in containers such as the MP4 file format [10]. The SVC extension of H.264/AVC allows further structuring the bitstream and extracting different video representations of one single bitstream, where the different substreams are referred to as layers. The base layer of SVC provides the lowest quality level and is a 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 present a 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), distributing different service qualities within one scalable stream 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 structure [11], 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 proceeding and succeeding frames with lower temporal level. A set of frames between two successive video frames of the temporal base layer with the succeeding base layer picture constitutes the so-called group of pictures (GOP). SVC coded video extents the GOP with further representations of video frames.

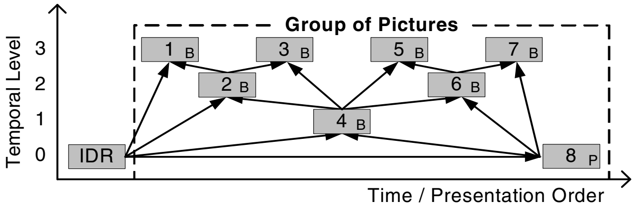

Fig. 2. Illustration of hierarchical predication within a H.264/AVC group of pictures (GOP)

2.2 External Transmission System Simulation

In order to simulate a specific transmission system, appropriate channel models have to be chosen according to the channel characteristics and parameters under test. For instance, a service provided via ADSL has to cope with different channel characteristics than a service using a mobile broadcast channel such as DVB-SH. Different phenomena influence the channel, e.g. path loss or fading for wireless, and attenuation or congestion for wired connections. The parameters under test determine the required simulation depth, i.e. whether the use of packet erasure channel (PEC) models is sufficient or not. A classic link/application layer emulating 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 [12] and is used in the exemplary simulations in the validation section.

Data transmission can be error-prone due to various reasons. Channel coding addresses this issue 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. Several FEC codes have been found to offer beneficial protection of data 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 [13]. The highly structured data of H.264/AVC and SVC coded video allows for stronger protection of data with higher priority, which is generally referred to as UEP. This is utilized to a high degree by priority aware FEC schemes such as Layer-Aware FEC that creates connected repair symbols across different SVC layers. Depending on the simulation depth, further techniques such as interleaving, modulation, or multiplexing have to be taken into account by the used transmission system simulation. Various generalized tools such as Network Simulator 2 or simulators for specific transmission systems such as DVB-H/SH [14][9] can be used for trace-driven channel simulation of different environments, i.e. internet streaming, peer-to-peer applications or mobile television services.

2.3 Video Quality Evaluation

As video coding and transmission may introduce a distortion to the processed video, the non-trivial task of measuring video quality is an important instrument to evaluate compression efficiency or transmission performance. Statistically relevant and meaningful results require subjective tests with a large test population. This approach is rather costly and time consuming, thus it is not adequate for a large amount of simulations. Several approaches have been made to find algorithmic quality evaluation metrics that correspond to the characteristics of human perception, as can be found in ITU-T recommendation J.247 [15]. 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 original and coded frame. The clear physical meaning and its simple calculation made PSNR the most commonly used metric to score the quality of coded video as of today, although it is only 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 evaluate video quality by utilizing known characteristics of human perception to a higher degree. Metrics such as Structural Similarity Index (SSIM) or Perceptual Evaluation of Video Quality (PEVQ), along with a variety of others, extract image features in form of structures, blocking or image activity, and consider the movement in a given video sequence, most likely 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 (per layer in case of SVC) is simple to compute and a meaningful indicator that allows calculating an Erroneous Seconds Ratio (ESR). The simulation presented in this paper use the well established combination of PSNR and decodable frame counts as metrics for quality evaluation. However, the proposed mechanisms are not limited to the PSNR metric and can easily be extended to offer other metrics as well.

3 Virtual Video Encoder

The main objective of the implemented encoding mechanism is to provide coded video at variable (VBR) or constant (CBR) bitrates by individually encoding video chunks with different quantization parameters. A compliant bitstream that meets the simulations criteria can be created by concatenating chunks after the encoding process. The proposed mechanism ensures that the coded video matches the target bitrate or video quality. For this purpose the continuous uncompressed source video sequence is divided into n chunks containing a smaller number of frames as depicted in Fig. 3. Since each video chunk is encoded individually, its coded representation starts with an IDR-frame. IDR-frames are independently decodable regardless of prior data and therefore serve as random access point (RAP). The chunk size controls the rate of RAPs in the concatenated bitstream. The IDR-frame is followed by an arbitrary number of GOP structures, thus the chosen chunk size should preserve an integer number of GOPs. Each chunk is encoded m times according to the number of selected parameters in terms of quantization or others to achieve an array of *n* x *m* encoded video chunks with different bitrate and quality. To form a valid H.264/AVC or SVC bitstream, coded chunks are selected and concatenated according to the simulation scenario, as illustrated in Fig. 4. Possible criteria for selection are continuous bitrate adaptation for each chunk in order to optimize the bitstream for a statistical multiplex scenario to maximize the number of available services within a given channel bandwidth or a constant video bitrate to not exceed a given service bandwidth [9]. Numerous bitstreams with different characteristics can be created from the array of encoded chunks without further encoding.

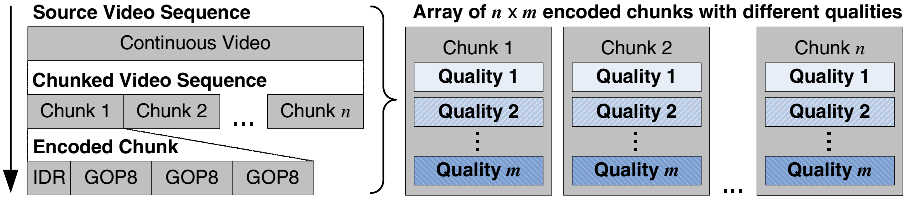

Fig. 3. Illustration of VVE mechanism and the resulting array of *n* x *m* encoded chunks

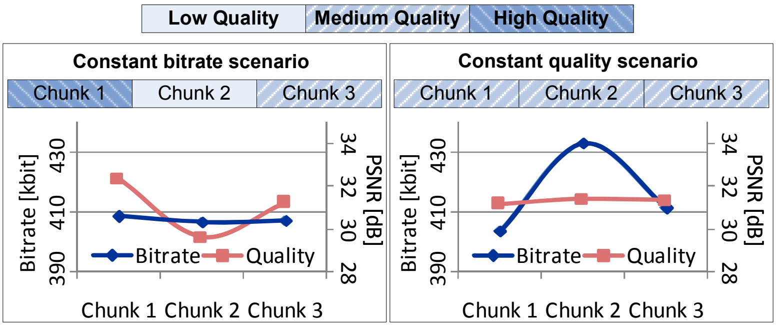

Fig. 4. Illustration of chunk selection in 'constant bitrate' and 'constant quality' scenarios

In order to generate packet traces of the encoded video, the concatenated bitstream is stored using the MP4 file format standard which has recently been extended to support SVC. Protocol encapsulation for further processing depends on the available interface to the transmission system simulation, which is RTP packet-oriented in our case. RTP encapsulation is done by analyzing the bitstream and adding so-called server hint tracks to the MP4 file. They contain all necessary information for media encapsulation according to the appropriate RTP payload format [16]. Utilizing hint tracks, a media agnostic server is able to encapsulate the media correctly. A textual description of RTP hint tracks containing packetization type, timestamps, packet and NAL unit size, NAL unit type and other parameters of the packetized coded video is extracted to a packet trace file. Transmission system simulation is based upon packet traces instead of the media data itself, which significantly speeds up the overall simulation process.

4 Virtual Video Decoder

The proposed VVD provides quality evaluation of transmitted video sequences without the need to decode results individually. The main idea is to allow application-level quality evaluation on packet level by pre-calculating a database that covers quality measurements for all possible video outputs. This database is created during a preprocessing phase, in which time usually needed for decoding and evaluation of single simulation results is combined to omit redundant calculations. Considering the structure of H.264/AVC and SVC coded video drastically eliminates unnecessary calculations further. In the evaluation phase, transmitted video sequences can be evaluated using the pre-calculated database. Transmission losses are analyzed on packet level and mapped to the corresponding quality metric values in the database. Thus, after preprocessing, vast simulations can be evaluated in very short time without any decoding operation. The proposed solution requires decoder implementation and media coding to fulfill certain constraints to reduce complexity and processing time. An error resilient decoder implementation, which is compatible to the H.264/AVC and SVC standard, is used to create the database of quality measurements for erroneous sequences. Basic error concealment techniques include base layer upsampling for loss of SVC enhancement layer data and insertion of freeze frames in case of frame loss to keep video output in sync [17]. Further constraints concerning the coding structure are hierarchical prediction and limitation to a single slice per frame, which reduces necessary calculation to a reasonable amount.

4.1 Relevant Error Pattern

Fig. 5 gives an exemplary error distribution within a single-layer H.264/AVC GOP structure. The frames are numbered in presentation order and vertically sorted regarding 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. The amount of combinations of erroneous frames within a GOP equals two raised to the power of n, 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. 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 erroneous 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. 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 in terms of other frames. Frame 1 and frame 3 are not decodable due to partially missing reference data in the form of 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 effected directly by transmission errors. Since the resulting video output is identical for error combinations that consist of the same initial errors, considering initial error combinations only is sufficient to cover all transmission errors. Processing these relevant error patterns (REP) reduces the number of necessary decoding operations significantly, as can be seen from Table 1.

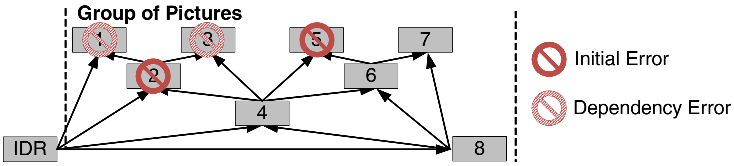

Fig. 5. Illustration of erroneous frame categories within a GOP structure

Table 1. Overview of error combinations, Relevant Error Pattern and calculation savings for different video codings and GOP sizes. BL = Base Layer, EL = Enhancement Layer

Video Coding	GOP Size BL	GOP Size EL	Number of Error Combinations	Number of REP	Savings
H.264/AVC	8	-	$2^8 = 256$	27	89.5%
H.264/AVC	16	-	$2^{16} = 65536$	678	99.0%
SVC Spatial or Quality Scalability	8	8	$2^{16} = 65536$	278	99.6%
SVC Spatial or Quality Scalability	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%

4.2 Database Generation

To generate a database of PSNR measurements, the preprocessing of each given coded video sequence utilizes the previously described error patterns. Each REP is mapped to the corresponding NAL units within all GOPs of the video sequence to create erroneous bitstreams. NAL units unaffected by initial or dependency errors are extracted, concatenated and processed with an error resilient video decoder. In conjunction with a unique REP identifier, a frame-wise PSNR measurement of the resulting video output is averaged for each GOP and stored in a database, as well as PSNR measurements for IDR-frames.

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. For 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 accurate evaluation of long-lasting freeze frames. This does not replace the GOP-based procedure described before as it fails to reproduce the behavior of an error resilient decoder. As PSNR measurement of two frames with different content gives to some extent arbitrary results, further calculations can be omitted by using a constant value which will speed up preprocessing further. The conducted simulations indicated that meaningful values lie within a range of 10dB to 15dB strongly depending on the video content regarding scene changes, movement and luminance.

4.3 Evaluation of Transmission Results

In order to evaluate simulations, resulting packet traces of channel simulations are analyzed 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 detects initial errors. Information of initial errors is used to compose a unique REP identifier to query corresponding video quality measurements from the database, which are subsequently averaged and combined with the count of erroneous and decoded frames per layer. Analyzing erroneous packet-traces is significantly faster than video decoding.

5 Validation

A simulation speed-up in conjunction with accurate results is the main objective of the proposed simulation platform and its benefit strongly depends on both. These goals are examined by comparing the VVD with the conventional approach of bitstream reconstruction, video decoding and evaluation. This is done based on an exemplary simulation of a mobile broadcast scenario that was conducted during investigations of intra-burst LA-FEC for SVC delivery in DVB-H [2]. In order to illustrate achieved time savings of the VVD, the overall evaluation time of the conventional approach is compared to VVD. Its accuracy is examined with a comparison of resulting measurements from VVD and the conventional approach.

5.1 Simulation Settings

The context of the simulations is a broadcast scenario with QVGA and VGA devices served by a single DVB-H service with SVC or simultaneous broadcast (simulcast) of H.264/AVC as shown in Fig. 6. Different FEC schemes for layered media (LA-FEC vs. conventional FEC) have been evaluated using different code rate distributions. Video encoding was done using the VVE to achieve an approximately constant video

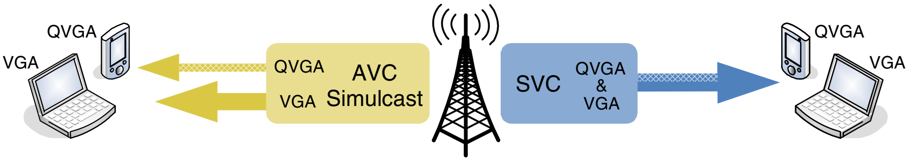

Fig. 6. H.264/AVC Simulcast vs. SVC layered transmission.

bitrate. The test sequence "soccer" with duration of 30 seconds was encoded without CABAC and 8x8 transform according to a restricted version of the scalable high profile. The SVC stream offers QVGA resolution at 12.5 frames per second (fps) as H.264/AVC compatible base layer and increased quality with a VGA enhancement layer at 25 fps. The single layer stream providing VGA resolution at 25 fps was encoded on a slightly lower quality in terms of PSNR. The additional SVC QVGA service leads to an overhead of roughly 7.5% compared to the single layer VGA stream. The GOP size is 8 frames and video chunks consists of one IDR frame plus three GOP structures, which results in a random access point rate of 25 frames or 1 second. Table 2 gives an overview of the SVC and simulcast encoding results.

Transmission system simulation was conducted with a TU6 channel using a DVB-H System-Level Simulator [14]. In detail the simulations included different Doppler frequencies (i.e. user velocity), a CNR range resembling correlated shadowing, several FEC schemes, FEC code rate distributions, and different transmission scheduling variants of the base layer. Each combination of parameters underwent several iterations to simulate a sufficient length of video for statistically consistent results.

Table 2. Encoding parameters of H.264/AVC and SVC stream for simulation

Encoding	Resolution	Bitrate	PSNR
H.264/AVC Base Layer	QVGA @ 12.5 fps	225 kbps	34.7 dB
SVC Enhancement Layer	VGA @ 25 fps	647 kbps	35.4 dB
H.264/AVC	QVGA @ 12.5 fps	225 kbps	34,7 dB
H.264/AVC	VGA @ 25 fps	811 kbps	35.3 dB

5.2 Time Savings

The range of parameters and iterations leads to roughly 20000 simulation cycles, equivalent of 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 and evaluate video quality. As can be seen from Table 1, using the proposed VVD a total of 278 erroneous versions, equivalent to less than 3 hours of video, have to be decoded to constitute a database for quality evaluation. Additional time of approximately an hour is needed for analyzing and creating erroneous bitstreams plus two hours to extend the database with measurements for long lasting freeze frames. The following evaluation of transmission simulations on packet level can be done in parallel at up to 6000 fps on standard PC hardware. Thus, evaluation of 20000 simulation cycles does not take longer than an hour. Complete VVD evaluation of the exemplary simulation takes less than 7 hours compared to about 180 hours of evaluating the reconstructed bitstreams of a real-time decoder, which leads to a significant speed-up of more than 90 percent of the overall evaluation process for the particular simulation.

Based on the given results, Fig. 6 illustrates estimated time of overall evaluation with VVD compared to evaluation of each reconstructed transmission result for a range of simulation cycles and reference decoder speeds. Since preprocessing phase duration depends on video coding and length, overall VVD evaluation time only slightly increases whereas benefit compared to reconstruction significantly grows with the amount of simulation cycles.

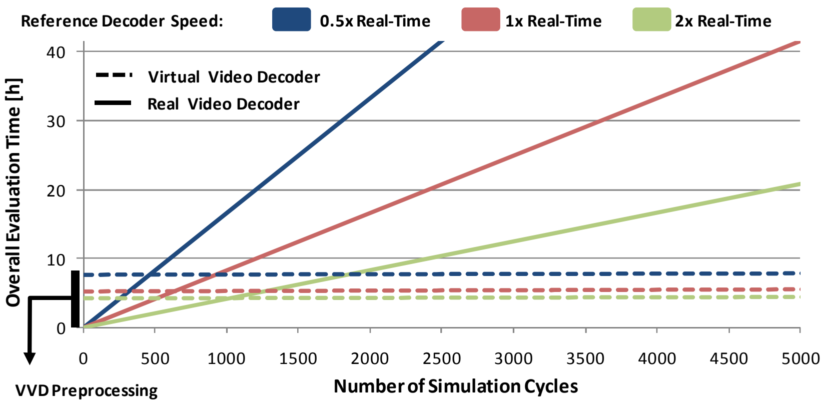

Fig. 7. Overall evaluation time of Virtual Video Decoder compared to the conventional approach of evaluating each reconstructed transmission result with Real Video Decoder for different reference decoder speed and number of simulation cycles

5.3 Accuracy

A comparison of VVD results with PSNR measurements of reconstructed transmission results from the exemplary simulations serves as a basis for evaluation of accuracy of the presented platform. For reconstruction, erroneous transmitted NAL units are discarded from the original bitstream and the resulting erroneous bitstreams are decoded and evaluated. The trivial case of error-free data transmission is evaluated 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 3 gives an analysis these particular simulations. Deviations for simulations that include frames losses slightly increase but do not rise to a notable magnitude and are negligible. The proposed VVD therefore meets the requirement of accuracy compared to time-consuming conventional approach.

Table 3. Overview of VVD results derivation for different frame losses, EL = Enhancement Layer

Number of lost EL Frames	Deviation of VVD results			
	Max [dB]	Min [dB]	Average [dB]	σ [dB]
50	0.0003	-0.0045	-0.0021	0.0024
75	0.0035	0.0023	0.0032	0.0005
100	0.0042	-0.0049	0.0000	0.0024
125	0.0035	-0.0048	0.0011	0.0024
150	0.0047	-0.0028	0.0024	0.0019
175	0.0050	-0.0049	0.0002	0.0028
200	0.0046	-0.0050	-0.0001	0.0029
225	0.0038	-0.0046	-0.0004	0.0027

6 Summary

This work presents a simulation platform for H.264/AVC and SVC transmission on error-prone transmission channels, which offers a fast and accurate mechanism of video quality evaluation. Time savings arise from reduction of redundancy by combining decoding operations and exploiting hierarchical prediction structures within H.264/AVC and SVC coded video. A preprocessing of video data constitutes a database for quality evaluation 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 evaluation speed of more than 85 percent and an insignificant deviation of results compared to a time-consuming reconstruction of transmission results. 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 the reference decoder speed, making the presented approach favorably for large simulation sets and slowly decodable video data such as HDTV. Opportunities for further work include the extension of the simulation platform for 3-dimensional multi view coded (MVC) video, the support of a larger number of SVC scalability layers or the implementation of different quality measures.

Acknowledgement. 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 8 (including SVC extension) (Consented in July 2007)

2. Hellge, C., et al.: Intra-Burst Layer Aware FEC for Scalable Video Coding Delivery in DVB-H. In: IEEE ICME 2010, Hong Kong (September 2010)
3. Klaue, J., Rathke, B., Wolisz, A.: EvalVid - A Framework for Video Transmission and Quality Evaluation. In: 13th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation, Urbana, Illinois, USA, pp. 255–272 (September 2003)
4. Ke, C.H., et al.: A Novel Realistic Simulation Tool for Video Transmission over Wireless Network. In: IEEE SUTC 2006, Taiwan (2006)
5. Lie, A., Klaue, J.: Evalvid-RA: Trace Driven Simulation of Rate Adaptive MPEG-4 VBR Video. ACM/Springer Multimedia Systems Journal 14(1) (June 2008)
6. Kanumuri, S., Reibman, A.: Modeling Packet-Loss Visibility in MPEG-2 Video. IEEE Transactions on Multimedia 8(2) (April 2006)
7. Tao, S., Apostolopoulos, J., Guérin, R.: Real-Time Monitoring of Video Quality in IP Networks. IEEE/ACM Transactions on Networking 16(5) (October 2008)
8. Liang, Y., Apostolopoulos, J., Girod, B.: Analysis of Packet Loss for compressed Video: Does Burst-Length Matter? In: IEEE ICASSP 2003, Hong Kong (2003)
9. Liebl, G., et al.: Simulation Platform for Multimedia Broadcast over DVB-SH. In: SIMUTools 2010, Torremolinos, Malaga, Spain (March 2010)
10. 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)
11. Schwarz, H., Marpe, D., Wiegand, T.: Analysis of hierarchical B pictures and MCTF. In: IEEE ICME 2006, Toronto, Canada (July 2006)
12. Faria, G., et al.: DVB-H: Digital Broadcast Services to Handheld Devices. Proc. IEEE 94(1), 194–209 (2006)
13. DVB Bluebook A148, Upper Layer FEC in DVB (March 2010)
14. Gómez-Barquero, D., et al.: Development and Applications of a Dynamic DVB-H System-Level Simulator. To appear in IEEE Transactions on Broadcasting 56(3) (2010)
15. Objective perceptual multimedia video quality measurement in the presence of a full reference, ITU-T Rec. J.247 (August 2008)
16. Wenger, S., et al.: RTP Payload Format for SVC Video (April 2010), `http://tools.ietf.org/html/draft-ietf-avt-rtp-svc-21`
17. Uitto, M., Vehkaperä, J.: Spatial Enhancement Layer Utilisation for SVC in Base Layer Error Concealment. In: MobiMedia 2009, London, United Kingdom (June 2009)