A Scalable Video Coding Extension of HEVC

Philipp Helle, Haricharan Lakshman, Mischa Siekmann, Jan Stegemann, Tobias Hinz, Heiko Schwarz, Detlev Marpe, and Thomas Wiegand

Fraunhofer Institute for Telecommunications – Heinrich Hertz Institute, Berlin, Germany

Abstract

The paper describes a scalable video coding extension of the upcoming HEVC video coding standard for spatial and quality scalable coding. Besides coding tools known from scalable profiles of prior video coding standards, it includes new coding tools that further improve the enhancement layer coding efficiency. The effectiveness of the proposed scalable HEVC extension is demonstrated by comparing the coding efficiency to simulcast and single-layer coding for several test sequences and coding conditions.

1. Introduction

Due to the increased efficiency of video coding technology and the improvements of network infrastructure, storage capacity, and computing power, digital video is used in more and more application areas, ranging from multimedia messaging, video telephony and video conferencing over mobile TV, wireless and Internet video streaming to standard- and high-definition TV broadcasting. On the one hand, the rising demand for video streaming to mobile devices such as smartphones, tablet computers, or notebooks and their broad variety of screen sizes and computing capabilities motivate the need for a scalable extension. On the other hand, modern video transmission systems using the Internet and mobile networks are typically characterized by a wide range of connection qualities, which are a result of the used adaptive resource sharing mechanisms. In such heterogeneous environments with varying connection qualities and different receiving devices, a flexible adaptation of once-encoded content is desirable.

Scalable video coding is an attractive solution to the challenges posed by the characteristics of modern video applications. In this context, scalability refers to the property of a video bitstream that allows removing parts of the bitstream in order to adapt it to the needs of end users as well as to the capabilities of the receiving device or the network conditions, where the resulting bitstream remains conforming to the used video coding standard. It should, however, be noted that two or more single layer bitstreams can also be transmitted using the method of simulcast, which provides similar functionalities as a scalable bitstream. Moreover, the adaptation of a single layer bitstream can be achieved by transcoding. Scalable video coding has to compete against these alternatives. In particular, scalable coding is only useful if it provides a higher coding efficiency than simulcast. Hence, the objective of a scalable extension for a video coding standard is to enable the creation of a video bitstream that contains one or more sub-bitstreams that can themselves be decoded with a complexity and reconstruction quality comparable to that achieved using single-layer coding with the same quantity of data as that in the sub-bitstream.

The international video coding standards H.262 | MPEG-2 Video [1], H.263 [2], MPEG-4 Visual [3], and H.264 | MPEG-4 AVC [4] include several tools by which scalable video coding can be supported. While the scalable profiles of H.262 | MPEG-2 Video, H.263, and

1068-0314/13 $26.00 © 2013 IEEE
DOI 10.1109/DCC.2013.28

MPEG-4 Visual are rarely used in practice, the scalable extension of H.264 | MPEG-4 AVC [5][6] is successfully used in a number of video conferencing applications.

The Joint Collaborative Team on Video Coding (JCT-VC) of experts from the ITU-T Visual Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) is currently developing a new video coding standard with the name High Efficiency Video Coding (HEVC) [7][8], for which the first version will be finalized in January 2013. For the targeted application area of high and ultra-high definition video coding, HEVC is capable of providing approximately 40% to 50% bit rate reduction compared to H.264 | MPEG-4 AVC at the same reproduction quality [9].

In this paper, we describe a scalable video coding extension for HEVC and present simulation results, which compare the efficiency of the scalable video coding extension to simulcast and single-layer coding. The proposed coding scheme can be used with an HEVC conforming base layer or an H.264 | MPEG-4 AVC conforming base layer. Potential application areas are video conferencing, for which the scalable extension of H.264 | MPEG-4 AVC is already successfully used, video streaming to mobile devices, or the backwards-compatible introduction of ultra-high definition broadcast formats.

2. Overview

The main types of scalability are temporal, spatial, and quality scalability. Spatial scalability and temporal scalability describe cases in which a sub-bitstream represents the source content with a reduced picture size (or spatial resolution) and frame rate (or temporal resolution), respectively. With quality scalability, which is also referred to as signal-to-noise ratio (SNR) scalability or fidelity scalability, the sub-bitstream provides the same spatial and temporal resolution as the complete bitstream, but with a lower reproduction quality and, thus, a lower bit rate. Similar to H.264 | MPEG-4 AVC, temporal scalability is already supported by the flexible reference picture memory control of HEVC.

Figure 1 depicts a simplified block diagram for spatial and quality coding with 2 layers. At a first glance the scalable encoder consists of two encoders, one for each layer. In spatial scalable coding, the input video is downsampled and fed into the base layer encoder, while the input video of the original size represents the input of the enhancement layer encoder. In quality scalable coding, both encoders use the same input signal. The base layer encoder conforms to a single-layer video coding standard, so that backwards compatibility with single-layer coding is achieved; the enhancement layer encoder usually includes additional coding features. The outputs of both encoders are multiplexed to form the scalable bitstream.

If both encoders are operated independently, the bitstream for the different spatial resolutions or quality levels are coded using the method of simulcast. In order to improve the coding efficiency, the data of the base layer need to be employed for an efficient enhancement layer coding by so-called inter-layer prediction methods. In the scalable profiles of the older video coding standards H.262 | MPEG-2 Video, H.263, and MPEG-4 Visual, a single inter-layer coding tool was included by which the reconstructed and, for spatial scalable coding, upsampled base layer picture of the same time instant could be used as an additional reference for predicting blocks of the current enhancement layer picture. In the scalable extension of H.264 | MPEG 4 AVC, basically three methods of inter-layer prediction are supported [5][6]:

- *Inter-layer intra prediction*: A block of the enhancement layer is predicted using the reconstructed (and upsampled) base layer signal.
- *Inter-layer motion prediction*: The motion data of a block are completely inferred using the (scaled) motion data of the co-located base layer blocks, or the (scaled) motion data of the base layer are used as an additional predictor for coding the enhancement layer motion.
- *Inter-layer residual prediction*: The reconstructed (and upsampled) residual signal of the co-located base layer area is used for predicting the residual signal of an inter-picture coded block in the enhancement layer, while the motion compensation is applied using enhancement layer reference pictures.

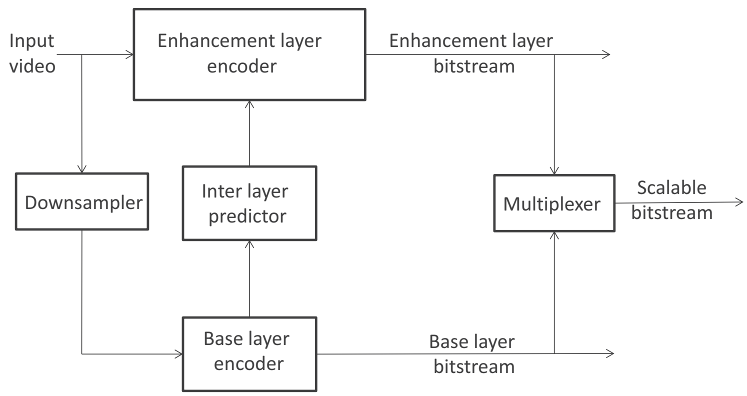

Figure 1: Simplified block diagram of a scalable encoder with 2 layers.

As a particular feature, the scalable extension of H.264 | MPEG-4 AVC was designed in a way that all layers could be coded with a single motion-compensation loop. This is achieved by restricting the inter-layer intra prediction to regions that are coded using an intra-coding mode in the base layer and only allowing constrained intra prediction in the base layer, i.e., samples from inter-predicted regions may not be used for inter-layer intra prediction. Although this restriction decreases the coding efficiency, it was considered a good trade-off between coding efficiency and decoder complexity during the development of the scalable profiles for H.264 | MPEG-4 AVC.

In practice, the scalable extension of H.264 | MPEG-4 AVC is mainly used in video conferencing applications, which are characterized by varying network conditions. The decoders used in this application area typically implement multi-loop decoding for improving the error resilience capabilities and do not make any use of the single-loop decoding feature. Due to this observation, we did not include the single-loop decoding feature in the proposed HEVC extension. Instead, an enhancement layer is generally decoded using multiple motion compensation loops. This already increases the coding efficiency for the set of inter-layer prediction tools known from the scalable extension of H.264 | MPEG-4 AVC, but also offers new possibilities to combine prediction signals from base and enhancement layer, which can further increase the coding efficiency relative to simulcast coding.

3. Description of the Scalable Coding Tools

The basic processing unit for each picture is the Coding Tree Unit (CTU) as defined in HEVC. The CTU contains a quad-tree syntax that allows for splitting a region into multiple

Coding Units (CUs) based on local characteristics. For each CU, a prediction mode is signaled, which can be intra, inter or inter-layer mode. Intra and inter prediction processes are the same as in HEVC. In order to reduce the bitrate required for signaling intra or inter parameters, methods are provided to predict the parameters from available base layer data. Additionally, inter-layer prediction modes are defined that provide the possibility to use base layer samples or residuals to form the prediction signal for an enhancement layer block. Inter-layer prediction can be classified as inter-layer intra modes that only access base or enhancement layer pictures corresponding to a particular time instant and inter-layer inter modes that additionally access previously coded pictures. A syntax element ilpred_type is included to identify the inter-layer prediction type for an enhancement layer CU. Besides temporal scalability, which is already supported in HEVC, the proposed scalable extension supports spatial scalability with arbitrary resolution factors and quality scalability (as a special case of spatial scalability with a resolution ratio of one).

3.1. IntraBL mode

For an enhancement layer CU, when ilpred_type indicates the IntraBL mode, the prediction signal is formed by copying or, for spatial scalable coding, upsampling the co-located base layer reconstructed samples. Since the final reconstructed samples from the base layer are used, multi-loop decoding architecture is necessary. For upsampling the luma component, one-dimensional 8-tap FIR filters are applied horizontally and vertically. The chroma components are upsampled using bilinear filters. For both luma and chroma, a set of 15 filters having approximately 1/16th phase shifts [6] are supported for handling arbitrary spatial scalability. The operation is similar to the inter-layer intra prediction in the scalable extension of H.264 | MPEG-4 AVC, except that it is possible to use the samples of both intra and inter predicted blocks from the base layer.

3.2. IntraBLFilt mode

The upsampling filters used in the IntraBL mode are designed to provide a good coding efficiency over a wide variety of base and enhancement layer signals. However, even within each picture, video signals may show a high degree of non-stationarity. Additionally, quantization errors and noise may show varying characteristics in different parts of a picture. Hence, to adapt the upsampling filter to local signal characteristics, a smoothing filter with coefficients [1 2 1]/4 is defined that can be turned on after upsampling to modify the overall frequency response. This happens when ilpred_type indicates IntraBLFilt mode, where the one-dimensional smoothing filter is applied horizontally and vertically after upsampling.

3.3. Weighted Intra prediction

In this mode, the (upsampled) base layer reconstructed signal constitutes one component for prediction. Another component is obtained by regular spatial intra prediction as in HEVC, using the samples from the causal neighborhood of the current enhancement layer block. The base layer component is low pass filtered and the enhancement layer component is high pass filtered and the results are summed to form the prediction. In our implementation, both low pass and high pass filtering happen in the DCT domain, as depicted in Figure 2. First, the DCTs of the base and enhancement layer prediction signals are computed and the resulting coefficients are weighted according to spatial frequencies. The weights for the base layer signal are set such that the low frequency components are retained and the high frequency components are suppressed and the weights for the enhancement layer signal are set vice versa. The weighted base and enhancement layer coefficients are summed and an inverse DCT is computed to obtain the final prediction.

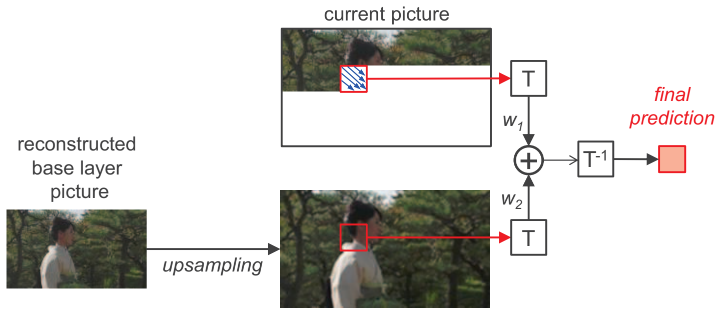

Figure 2: Weighted intra prediction mode. The (upsampled) base layer reconstructed samples are combined with the spatially predicted enhancement layer samples to predict an enhancement layer CU to be coded.

3.4. Difference prediction modes

The principle in difference prediction modes is to reduce the systematic error when using the (upsampled) base layer reconstructed signal for prediction. It is achieved by reusing the previously corrected prediction errors available to both encoder and decoder. To this end, a new signal, denoted as the difference signal, is derived using the difference between already reconstructed enhancement layer samples and (upsampled) base layer samples. The final prediction is formed by adding a component from the (upsampled) base layer reconstructed signal and a component from the difference signal [10].This mode can be used for inter as well as intra prediction cases.

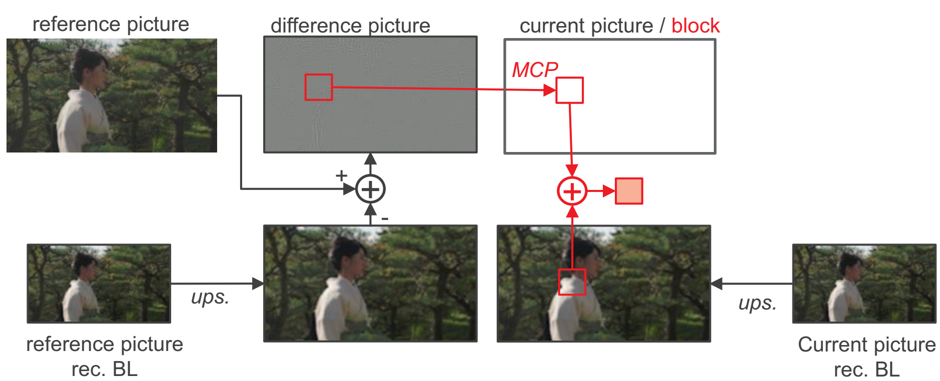

Figure 3: Inter difference prediction mode. The (upsampled) base layer reconstructed signal is combined with the motion compensated difference signal from a reference picture to predict the enhancement layer CU to be coded.

In inter difference prediction, the (upsampled) base layer reconstructed signal is added to a motion-compensated enhancement layer difference signal corresponding to a reference picture to obtain the final prediction for the current enhancement layer block, as depicted in Figure 3. For the enhancement layer motion compensation, the same inter prediction technique as in single-layer HEVC is used, but with a bilinear interpolation filter. The motion vectors indicated in the bitstream point to areas of enhancement layer difference signals in this mode.

In intra difference prediction, the (upsampled) base layer reconstructed signal constitutes one component for the prediction. Another component is derived by spatial intra prediction using the difference signal from the causal neighborhood of the current enhancement layer block. The intra prediction modes that are used for spatial intra prediction of the difference signal are coded using the regular HEVC syntax. The final prediction signal is formed by adding the (upsampled) base layer reconstructed signal and the spatially predicted difference signal, as shown in Figure 4.

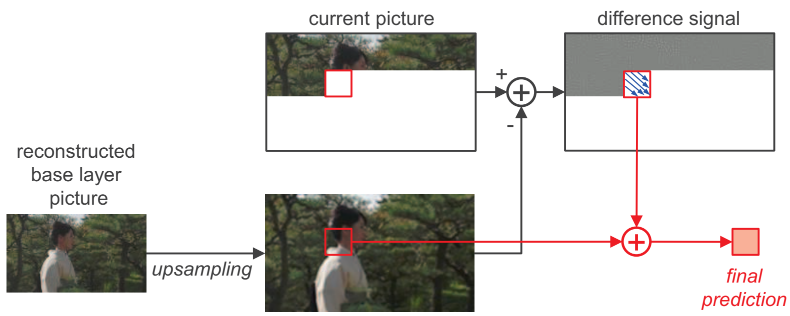

Figure 4: Intra difference prediction mode. The (upsampled) base layer reconstructed signal is combined with the intra predicted difference signal to predict the enhancement layer block to be coded.

3.5. Inter layer motion vector prediction

Our scalable video extension of HEVC employs various methods to improve the coding of enhancement layer motion information by exploiting the availability of base layer motion information. First, we introduced a new mode called InterCopy which allows to completely infer motion information for an enhancement layer block from the base layer. This mode is similar to a macroblock mode employed in the scalable video extension of H.264 | MPEG-4 AVC [4][5] by signaling the *base mode flag* equal to 1. When the base mode flag is equal to 0 in the H.264 | MPEG-4 AVC extension, it allows by signaling an additional flag for each coded motion vector, whether a co-located base-layer motion vector is used as a predictor or not. While H.264 | MPEG-4 AVC only employs a single motion vector predictor, the motion prediction used in HEVC employs a list of competing predictors. We incorporate base layer motion information into this motion predictor competition scheme by an extension of the predictor list.

Motion information inference

In InterCopy mode, all motion information required for a CU in the enhancement layer is inferred from the base layer. Each 8×8 sub-block within an enhancement layer CU copies the motion information from the base layer block containing the position of the upper-left sample of the enhancement layer block. This way, the inference of motion parameters is defined for arbitrary spatial scaling ratios. In case the corresponding base layer block is intra predicted, i.e., no base layer motion is available, the enhancement layer block uses the IntraBL mode for reconstruction.

For transform coding the prediction residual, transform block boundaries have to be defined. On the one hand, the signaling of transform block boundaries could probably benefit from taking into account the particular prediction block structure in the base layer. On the other hand, this would introduce unwanted parsing dependencies. We avoid those by using the same residual quad tree (RQT) syntax as otherwise used for a CU in HEVC.

Predictor list extension

In the motion prediction scheme employed in single layer HEVC, the motion vectors of a block can be either coded differentially by advanced motion vector prediction (AMVP), or they can be inferred by merging the block with one of its neighboring blocks, i.e., copying the motion information from a particular neighboring block [11]. Either way, this scheme uses a list of motion parameters to be used as candidate predictors. We exploit base layer motion information by including additional candidates obtained from the base layer as described below. Note however, that the syntax for signaling the motion parameters has not been changed compared to the single-layer HEVC specification.

The advanced motion vector prediction (AMVP) in HEVC uses two spatial and one temporal motion vector predictor as possible candidates for motion vector prediction [11]. In the presented scheme, the candidate list is extended by an additional inter-layer motion vector predictor, which is obtained by mapping the center position of the current block into the base layer. The motion vector of the corresponding base layer block is scaled according to the resolution ratio and inserted at the first position of the candidate list.

We similarly extend the candidate list used for the HEVC block merging [11]. It may also use an inter-layer motion parameter predictor which is obtained from the same center position of the current block. This new candidate is inserted to the top of the merge candidate list. This inter-layer predictor is only inserted if the merge flag of the co-located block in the base layer is equal to 0. Otherwise, it is not considered as a candidate to avoid redundant motion information among the candidates. For spatial scalability, the base layer motion vectors need to be scaled according to the resolution ratio. After the candidate list construction is completed, a re-ordering process may be invoked. This re-ordering depends on the co-located base layer prediction block that covers the top-left sample position inside the current prediction block. If the merge flag of this block equals 1, the origin of the base layer motion predictor is retrieved, which can be any of the candidates depicted in Figure 5. If the current list contains a candidate with the same origin, this candidate is put to the first position, shifting the other candidates and thus increasing their index by one. As the first index tends to have the shortest codeword, the reordering step implicitly considers the selection of the same predictor origin as in the base layer as the most likely event.

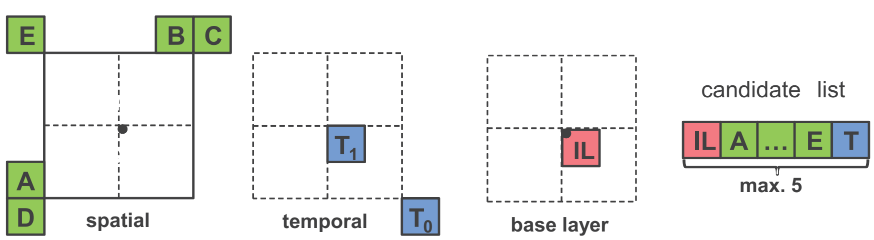

Figure 5: Spatial, temporal and inter-layer sample positions used to obtain the candidate predictor list for block merging.

3.6. Residual prediction mode

Our HEVC extension employs an additional CU mode InterResi that uses residual samples from the base layer to improve the inter-picture prediction of enhancement layer samples. This mode is similar to the macroblock mode referred to as inter-layer residual prediction employed in the scalable video extension of H.264 | MPEG-4 AVC. The InterResi mode derives the final prediction signal for a CU by adding to an enhancement layer inter prediction signal the co-located (upsampled) reconstructed residual samples. When this mode is signaled in a spatial enhancement layer, the co-located residual signal is first upsampled by bilinear interpolation. In a quality enhancement layer, the reconstructed residual samples are directly added to the prediction signal without any interpolation.

3.7. Transform and entropy coding

Conceptually, coding of the enhancement layer transform coefficients is done in the same way as the coding of the base layer transform coefficients, and thus, follows the coding scheme of single-layer HEVC. Only the coding of the significant coefficient positions within a transform unit (TU) is extended, whereas the coding of the remaining level and sign information is unchanged. The coding of significant positions utilizes scan patterns to map the positions within a TU into a one-dimensional sequence.

While in single layer HEVC a static scheme is used for selecting a particular scan pattern for each TU, an adaptive scan selection is additionally introduced for enhancement layer TUs with certain criteria. For luminance signal components with TU sizes of 16×16 or 32×32, it is checked whether the last significant coefficient position lies outside the first 4×4 sub-block. If this is the case, the encoder can select a diagonal, vertical, or horizontal scan pattern by minimizing the rate-distortion cost. The selection is signaled in the bitstream.

4. Simulation Results

For evaluating the efficiency of the proposed scalable HEVC extension, we compared the coding efficiency of the scalable approach with two layers to that of simulcast and single layer coding. We followed the test conditions specified in the Joint Call for Proposals on Scalable Video Coding Extensions of HEVC [12]. All layers have been coded using hierarchical B pictures with a GOP size of 8 pictures. Intra pictures for enabling random access have been inserted about every 1.1 seconds. For both scalable coding and simulcast, the same base layers are used. For spatial scalable coding, the base layer intra QP was set equal to 34, 30, 26, and 22. For quality scalability, the base layer intra QP was set to 38, 34, 30, and 26. For spatial scalable coding, the enhancement layer QPs have been set to BQP + 4, BQP + 2, BQP, BQP – 2. For quality scalable coding, the enhancement layer QPs BQP – 2, BQP – 4, BQP – 6, and BQP – 8 have been used.

Figure 6 exemplifies the coding efficiency of the presented scalable extension by showing RD-curves of different tool subsets together with the coding efficiency of simulcast and single-layer coding for a selected configuration. The "IntraBL only" subset contains the IntraBL tool as the only inter-layer prediction tool. The "SVC tools" subset contains the tools similar to those found in the scalable extension of H.264 | MPEG-4 AVC, i.e. IntraBL, the motion prediction tools, and the residual prediction mode, but without the restrictions for single-loop decoding. "All tools" denotes the complete tool set as proposed in this paper. In terms of RD-performance, this scalable extension is one of the top ranking of all proposals submitted to the Call for Proposals on scalable extensions of HEVC [13], when excluding the additional single-layer coding tools for enhancement layers present in some proposals.

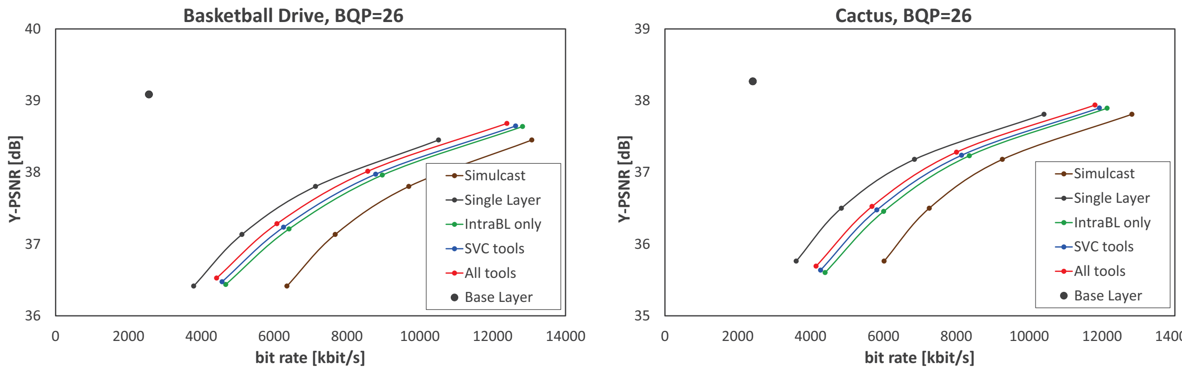

Figure 6: Rate-distortion curves for spatial scalability with a resolution ratio of 2 and BQP=26. The coding efficiency of the scalable HEVC extension is compared with that of simulcast, single-layer coding, and two versions with a reduced tool set.

Using the obtained bit rates and average PSNR values and the bit rates and PSNR values of the simulcast anchor provided by the JCT-VC, we calculated the bit-rate savings of the complete tool set relative to simulcast (SC), the bit-rate overhead relative to single-layer coding (SL), and a measure we call base layer usage (BLU). The base layer usage is given by the difference of the simulcast bit rate and the scalable bit rate divided by the base layer rate. It can be interpreted as the amount of the base layer rate that is re-used for the enhancement layer coding. A base layer usage of 0% and 100% represents a coding efficiency equal to that of simulcast and single-layer coding respectively. Table 1 summarizes the simulation results for various sequences for a fixed base layer QP of 26 in spatial and quality scalability tests, while Table 2 provides average results over the entire test set.

Table 1 – Simulation results for random access coding with selected base layer QPs**.**

	spatial 2.0 (base QP 26)			**spatial 1.5 (base QP 26)**			**quality (base QP 30)**		
	SC	SL	BLU	SC	SL	BLU	SC	SL	BLU
BQTerrace	12.0 %	8.3 %	75.3 %	19.4 %	9.7 %	79.8 %	22.1 %	-4.4 %	135.1 %
BasketballDrive	23.4 %	9.0 %	79.2 %	31.6 %	8.0 %	86.1 %	28.0 %	3.2 %	94.9 %
Cactus	21.0 %	13.3 %	70.2 %	30.2 %	12.6 %	79.3 %	26.2 %	6.1 %	89.1 %
Kimono	29.1 %	7.2 %	85.4%	36.0 %	6.5 %	89.9 %	29.1 %	6.6 %	87.1 %
ParkScene	17.4 %	15.4 %	60.6 %	28.4 %	14.6 %	74.2 %	23.8 %	11.9 %	73.8 %
PeopleOnStreet	27.5 %	9.9 %	80.9 %				31.1 %	2.5 %	94.9 %
Traffic	18.0 %	17.6 %	59.2%				23.3 %	11.9 %	73.3 %
Average	**21.2%**	**11.5%**	**72.9%**	**29.1%**	**10.3%**	**81.8%**	**26.2%**	**5.4%**	**92.6%**

Table 2 – Average simulation results for various tool configurations**.**

test sequence	**savings vs. simulcast**			**overhead vs. single-layer**			**base layer usage**		
	IntraBL	SVC tools	all tools	IntraBL	SVC tools	all tools	IntraBL	SVC tools	all tools
Spatial 2.0	16.3 %	18.3 %	21.2 %	18.8 %	15.8 %	11.6 %	55.2 %	62.4 %	73.7 %
Spatial 1.5	25.0 %	26.8 %	29.5 %	17.3 %	14.4 %	10.2 %	68.7 %	74.1 %	82.9 %
Quality	21.0 %	22.7 %	25.5 %	14.1 %	11.7 %	7.6 %	70.1 %	75.7 %	87.0 %

The anchors have been coded using only coding tools of the draft Main profile. For the scalable extension, only the described scalable coding tools have been enabled in addition to the Main profile tools. The spatial resolution of the test sequences ranges from 1080p (for the first 5 test sequences in Table 1) to about 4k×2k (for the last 2 test sequences in Table 1). The scalable extension has been implemented in the HEVC reference software HM-6.1, which has also been used for producing the anchor bit streams. The encoders for both scalable and single-layer coding have been operated using the same Lagrangian encoder control.

All bit rate savings, overheads, and base layer usages have been obtained by interpolating the obtained PSNR curves for a fixed base layer setting using cubic spline interpolation and numerically integration. As can be seen from the results, on average, the proposed scalable HEVC extension with all tools provided bit-rate savings between 20% and 30% relative to simulcast for the considered sequences and test cases. The overhead relative to single-layer coding is between 5% and 12%. And the base layer usage is in the range from 73% to 93%, which can be interpreted as the amount of the base layer rate could be re-used for the enhancement layer coding. The effectiveness of the proposed scalable HEVC extension generally improves with increasing base layer rate.

5. Conclusion

We presented a scalable extension of the upcoming video coding standard HEVC, which includes new scalable coding tools in addition to coding tools known from scalable profiles of prior coding standards. In contrast to the scalable extension of H.264 | MPEG-4 AVC, the proposed HEVC extension requires multi-loop decoding, which provides more freedom in designing improved scalable coding tools. The effectiveness of the described approach has been demonstrated by experimental results.

References

[1] ITU-T and ISO/IEC JTC 1, "Generic Coding of Moving Pictures and Associated Audio Information – Part 2: Video," ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2), version 1: 1994.

[2] ITU-T, *Video Coding for Low Bitrate Communication*, ITU-T Rec. H.263, version 1, 1995, version 2, 1998, version 3, 2000.

[3] ISO/IEC JTC 1, *Coding of Audio-Visual Objects – Part 2: Visual*, ISO/IEC 14496-2 (MPEG-4 Visual), version 1: 1999, version 2: 2000, version 3: 2004.

[4] ITU-T and ISO/IEC JTC 1, *Advanced Video Coding for generic audiovisual services*, ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC), vers. 1: 2003, vers. 2: 2004, vers. 3, 4: 2005, vers. 5, 6: 2006, vers. 7, 8: 2007, vers. 9, 10, 11: 2009, vers. 12, 13: 2010, vers. 14, 15: 2011, vers. 16: 2012.

[5] H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103-1120, September 2007.

[6] C. A. Segall, G. J. Sullivan, "Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1121-1135, September 2007.

[7] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 8," Joint Collaborative Team on Video Coding, doc. JCTVC-J1003, July, 2012.

[8] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, December 2012.

[9] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.22, no.12, pp.1669-1684, Dec. 2012.

[10] J. Boyce, D. Hong, W. Jang, A. Abbas, "Information for HEVC scalability extension," Joint Collaborative Team on Video Coding, doc. JCTVC-G078, Nov. 2011.

[11] P. Helle, S. Oudin, B. Bross, D. Marpe, M. Bici, K. Ugur, J. Jung, G. Clare, T. Wiegand, "Block Merging for Quadtree-based Partitioning in HEVC,"*IEEE Trans. Circuits Syst. Video Technol.*, Dec. 2012.

[12] ISO/IEC JTC 1/SC 29/WG 11 and ITU-T SG 16 WP 3, "Joint Call for Proposals on Scalable Video Coding Extensions for High Efficiency Video Coding (HEVC)," Doc. N12957, July 2012.

[13] P. Lai et al, "Summary of tools and performance of scalable video coding technology proposals," Joint Collaborative Team on Video Coding, Doc. JCTVC-K0346, October 2011.