An HEVC Extension for Spatial and Quality Scalable Video Coding

Tobias Hinz[1a], Philipp Helle[a], Haricharan Lakshman[a], Mischa Siekmann[a], Jan Stegemann[a], Heiko Schwarz[a], Detlev Marpe[a], and Thomas Wiegand[a,b]
[a]Fraunhofer Institute of Telecommunication – Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany; [b]Berlin Institute of Technology, Einsteinufer 17, 10587 Berlin, Germany

ABSTRACT

This paper describes an extension of the upcoming High Efficiency Video Coding (HEVC) standard for supporting spatial and quality scalable video coding. Besides scalable coding tools known from scalable profiles of prior video coding standards such as H.262/MPEG-2 Video and H.264/MPEG-4 AVC, the proposed scalable HEVC extension includes new coding tools that further improve the coding efficiency of the enhancement layer. In particular, new coding modes by which base and enhancement layer signals are combined for forming an improved enhancement layer prediction signal have been added. All scalable coding tools have been integrated in a way that the low-level syntax and decoding process of HEVC remain unchanged to a large extent. Simulation results for typical application scenarios demonstrate the effectiveness of the proposed design. For spatial and quality scalable coding with two layers, bit-rate savings of about 20-30% have been measured relative to simulcasting the layers, which corresponds to a bit-rate overhead of about 5-15% relative to single-layer coding of the enhancement layer.

Keywords: HEVC, scalable video coding, spatial scalability, quality scalability

1. INTRODUCTION

The increased efficiency of video coding technology, the continuous improvement of network infrastructure and the rapid development of computing and communication devices brought digital video into more and more areas of our daily life. Today's video applications range from video telephony and video conferencing, over wireless and Internet video streaming to home entertainment, digital cinema and video surveillance. In particular, video transmission over the Internet and mobile networks is continuously increasing. People regularly use mobile devices such as smartphones, tablet computers or notebooks for retrieving videos as information sources or for entertainment. Today, already more than half of the traffic in the consumer Internet are video data. In contrast to classical TV broadcasting, the receiving devices are characterized by widely varying properties. Smartphones have typically a lower screen resolution as well as lower computing capabilities and battery power than tablet computers and notebooks. Furthermore, modern video applications use the Internet and mobile networks, which are characterized by adaptive resource sharing resulting in unreliable connection qualities. In order to provide each user with a video quality that fits to the capabilities of its receiving device and its network connection, multiple coded versions of the same source content with different picture sizes and bit rates have to be generated. For coping with varying throughput of the used network connection, mechanisms for seamless switching between different coded versions of a source are highly desirable.

Scalable video coding is an attractive solution to the challenges of modern video applications that are characterized by receiving devices with heterogeneous properties and unreliable network connections. Scalable video coding refers to a design by which the source content is encoded in a format that is easily adaptable to the capabilities of multiple receiving devices or network conditions. In this paper, a video bitstream is called scalable if it allows the removal of parts of the bitstream in a way that the resulting bitstream can be decoded and provides a lower resolution or quality of the source content. If a receiving device does not have the capabilities to decode the maximum spatio-temporal resolution or maximum bit rate contained in a scalable bitstream, the non-required parts of the bitstream can be removed and a lower spatio-temporal resolution or quality can be decoded. The concept of data discarding can also be used for responding to changing network conditions or low battery power. Furthermore, scalability can improve error resilience in certain application areas. Since different parts of the bitstream have different importance, the most important parts, which represent a base quality of the input video, can be protected stronger than the less important parts. By using such unequal

[1] tobias.hinz@hhi.fraunhofer.de; phone +49 30 31002 605; fax +49 30 31002 190; hhi.fraunhofer.de

Visual Information Processing and Communication IV, edited by Amir Said, Onur G. Guleryuz, Robert L. Stevenson,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8666, 866605 · © 2013 SPIE-IS&T
CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2009393

error protection mechanisms, the end user receives, with a high probability, at least a base quality of the transmitted content and, thus, does not observe a disturbing interruption of the video playback. Another potential use of scalable video coding is the backward-compatible introduction of extended video formats in video broadcasting. Ultra high definition television (UHDTV) formats could be introduced as enhancement layer of today's high-definition formats, so that legacy devices remain capable of decoding the HDTV resolution, while new devices can decode both layers and the content can be displayed in the improved format.

Similarly as for conventional non-scalable video coding, interoperability between sending and receiving devices is an important aspect. It has to be ensured that a bitstream created by a particular encoding product of one manufacturer can be decoded by receiving devices of other manufacturers. Interoperability is usually achieved by defining international video coding standards, which specify the bitstream syntax and the decoding process. In the context of scalable video coding, it has to be further ensured that scalable and non-scalable devices can interoperate with each other. On the one hand, a scalable decoder has to be capable of decoding non-scalable bitstreams. On the other hand, a scalable bitstream has to include a sub-bitstream (which is also referred to as base layer) that conforms to a non-scalable profile of a widely-used international video coding standard. This basic concept of scalable video coding with a legacy base layer is illustrated in Figure 1.

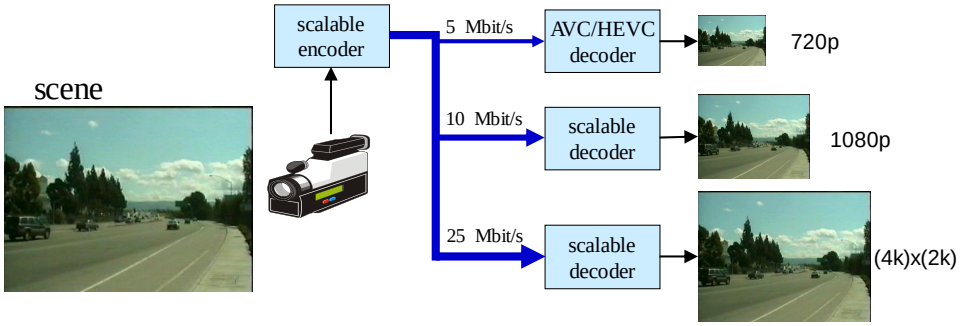

Figure 1: Basic concept of scalable video coding. By discarding data from a scalable video bitstream, versions with lower spatio-temporal resolution or quality can be decoded. The bitstream corresponding to the lowest resolution and quality point is conforming to a non-scalable profile of a base coding standard.

Scalable video coding has been an active research area for several decades. 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] already specify scalable profiles. The topic of scalable video coding has also been actively investigated in the research community and, in particular, several scalable coding techniques based on 3-d wavelet transforms have been developed [5][6][7][8]. Although the wavelet transform based approaches have not been included into international video coding standards, their investigation provided various insights into scalable coding and influenced the design of scalable coding standards. While the scalable profiles of the older standards H.262 | MPEG-2 Video, H.263, and MPEG-4 Visual are rarely used in practice, the scalable extension of H.264 | MPEG-4 AVC is the first scalable standard that found its way into some application areas. In particular, several video conferencing applications successfully use the advantages of the scalable profiles of H.264 | MPEG-4 AVC.

One reason for the rare use of the scalable profiles of older standards can be seen in the fact that the main application area for digital video in the times when these standards were approved was digital television and optical storage for home entertainment, where the scalable features do not provide any advantage over conventional single-layer coding. Another aspect is the loss of coding efficiency compared to non-scalable coding. An alternative to scalable coding is the transmission of multiple single-layer bistreams using the method of simulcast, which basically provides similar functionalities as a scalable bitstream. The adaptation of a single-layer bitstream to the capabilities of a receiving device can also be achieved by transcoding, i.e., a combination of decoding, a potential format conversion and re-encoding. Since scalable video coding comes at additional implementation costs and an increased decoding complexity, it has to provide clear advantages in terms of coding efficiency compared to the alternatives of simulcast and transcoding. In addition, for keeping the implementation costs low, it is desirable that the decoding complexity of scalable video coding design is only slightly increased in comparison to single-layer coding and that most parts of the underlying coding standard remain unmodified.

The state-of-the-art in video coding is the High Efficiency Video Coding (HEVC) standard [9][10] that is currently developed by 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) and will be technically finalized in January 2013. For the main targeted application area of high and ultra-high definition video coding, the HEVC design is capable of providing approximately 40% to 50% bit rate reduction [11] compared to its predecessor H.264 | MPEG-4 AVC at the same reproduction quality, while the decoding complexity is not significantly changed relative to H.264 | MPEG-4 AVC [12]. In order to address the potential needs of future video applications, the JCT-VC issued a Call for Proposals (CfP) on Scalable Video Coding Extensions for High Efficiency Video Coding [13]. If the coding efficiency of a scalable extension of HEVC is close to that of single-layer coding for typical scenarios, scalable video coding can play a greater role in future video applications than it did in the past.

In the following, a scalable video coding extension of HEVC is described. Section 2 briefly reviews the design of the underlying HEVC coding standard. The extensions of HEVC for supporting spatial and quality scalable coding are described in Section 3, where particular differences to the SVC extension of H.264 | MPEG-4 AVC are highlighted. Simulation results comparing the proposed coding scheme with simulcast and single-layer coding for typical application scenarios are presented in Section 4. Section 5 concludes the paper.

2. HEVC OVERVIEW

Similar to all other major video coding standards, the design of HEVC follows the so-called block-based hybrid video coding approach. In the following, the basic concepts of the HEVC design are briefly described with a focus on aspects that have been modified or extended in the proposed scalable video coding approach. For further details on HEVC, the reader is referred to the draft standard text [9] and the overview paper in [10]. For illustration, Figure 2 shows a simplified block diagram of an HEVC encoder, where the gray-shaded box represents the embedded decoder.

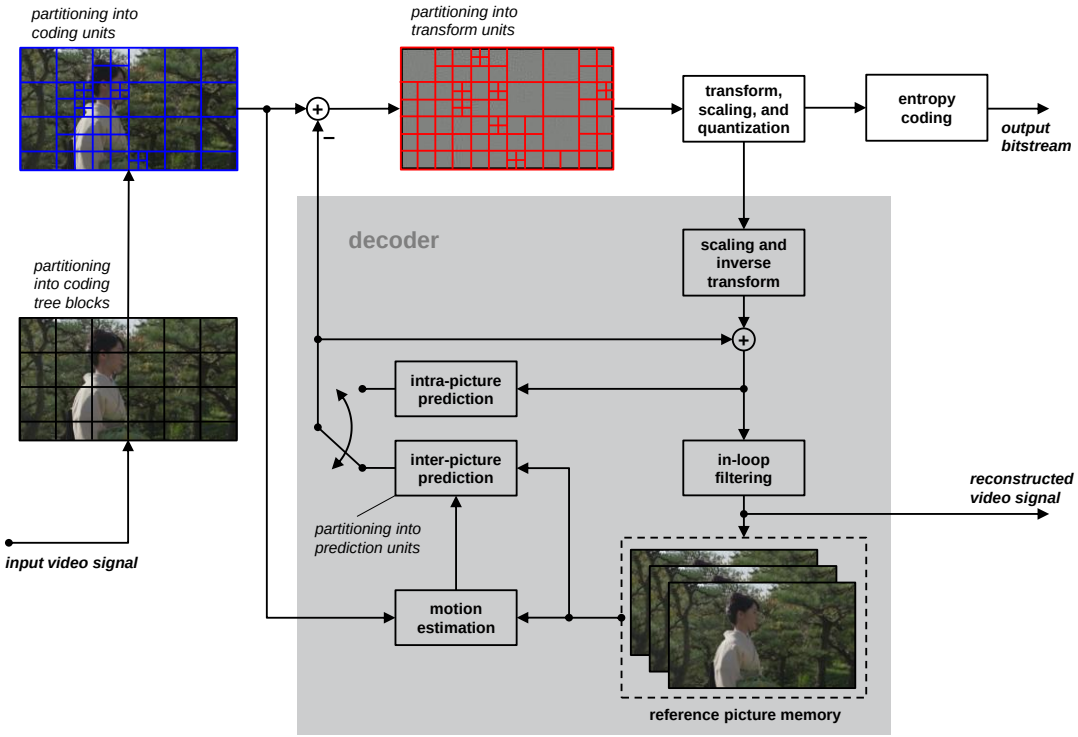

Figure 2: Simplified block diagram of an HEVC encoder. The gray shaded box represents the integrated decoder.

A picture is partitioned into so-called coding tree blocks (CTBs). The size of the CTBs can be chosen by the encoder. A luma CTB covers a square picture area of $N \times N$ samples of the luma component and, in the 4:2:0 chroma format, the

associated chroma CTBs cover each a square picture area of (N/2)×(N/2) samples of each of the two chroma components. The value of N is signaled in the bitstream and can be 16, 32, or 64. The luma CTB and the two associated chroma CTBs, together with the associated syntax, form a coding tree unit (CTU). The CTU is the basic processing unit in HEVC and can be conceptually compared to a macroblock (MB) in prior standards.

The luma and chroma CTBs can be further partitioned into multiple coding blocks (CBs). The CTU contains a quadtree syntax that allows further splitting into square blocks of variable size. The size of a CB can range from the same size as the CTB to a minimum size (8×8 luma samples or larger) that is specified by a syntax element transmitted to the decoder. The luma CB and the chroma CBs, together with the associated syntax, form a coding unit (CU). For each CU, the prediction mode is signaled to the decoder. It can be either intra or inter.

If a CU is coded in an intra prediction mode, one of 35 spatial intra prediction modes can be selected for predicting the luma CB using reconstructed samples of already coded neighboring blocks. Of these intra prediction modes, 33 represent angular prediction modes, by which the samples inside the luma CB are predicted from neighboring samples using a particular prediction direction. In addition, a DC prediction mode, which uses an average of neighboring samples for the prediction, and a planar prediction mode, for which the prediction signal is formed by average values of two linear predictions using the corner samples, are supported. If the luma CB has the smallest allowable size, it can also be split into four square blocks of the same size, in which case an intra prediction mode is signaled for each of these four sub-blocks. For both chroma CBs, a single intra prediction mode is selected, which specifies horizontal, vertical, left-downward diagonal, planar, or DC prediction, or the usage of the same intra prediction modes as used for luma. Depending on the direction and block size, a low pass filter is applied to the input samples to further enhance the prediction signal.

For inter-coded CUs, the luma and chroma CBs can be partitioned into one, two, or four prediction blocks (PBs), as illustrated in Figure 3. The luma and chroma PBs, together with the associated syntax, form a prediction unit (PU). Each PU is associated with a set of motion parameters. The inter prediction signal of a PU is either formed by uni-directional prediction or bi-prediction, where the latter is only supported in B slices. When uni-directional prediction is selected, the prediction signal is formed by displacing a block of a previous coded reference picture. For bi-prediction, the prediction signal is formed by a weighted average of two displaced blocks of previously coded reference pictures. The employed reference pictures are signaled using indices into reference picture lists. The reference picture lists include a subset of previously coded reference pictures; their construction is signaled in the slice header. The motion vectors have quarter luma sample precision. The luma prediction signal for all fractional sample locations is generated by separable 8-tap or 7-tap interpolation filters; for chroma, 4-tap filters are applied.

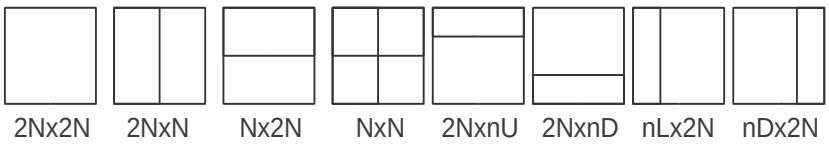

Figure 3: Supported partitioning types of a coding unit into prediction units.

For an efficient signaling of motion parameters for large areas that have the same translational motion, HEVC supports a so-called merge mode [14], which replaces the skip mode of prior video coding standards. Instead of transmitting motion vectors and reference indices, a candidate list of motion parameters is derived for the corresponding PU and only an index into the list is transmitted. In general, the candidate list includes the motion parameters of spatially neighboring blocks as well as temporally predicted motion parameters that are derived based on the motion data of the co-located block in a reference picture. By using the merge concept, it can be very efficiently signaled that a PU has the same motion parameters, i.e., the same number of hypotheses, reference picture indices, and motion vectors, as a neighboring PU. For PUs that are not coded using the merge mode, the number of motion hypothesis, the reference indices, and motion vector differences are transmitted. The prediction of motion vectors is done using advanced motion vector prediction (AMVP), in which a candidate list of motion vectors is constructed for each motion vector to be transmitted. The candidate list includes the motion vectors of neighboring blocks with the same reference index as well as a temporally predicted motion vector. An index into the candidate list specifying the chosen motion vector predictor is coded together with the corresponding difference vector.

For coding the inter or intra prediction residual, a luma CB is either represented as a single transform block (TB) or it is split recursively into four equal size transform blocks. The same splitting applies also to chroma CBs, with the exception that 4×4 chroma CBs are not further split. The splitting concept is also referred to as residual quadtree (RQT) and the luma and chroma TBs together with the associated syntax form a transform unit (TU). Each TB is transformed using a separable 2-d transform. The transforms are integer approximations of a discrete cosine transform (DCT). The inverse transforms are specified by exact integer operations. For intra-predicted luma CBs of a block size of 4×4, an alternative transform representing an integer approximation of a discrete sine transform (DST) is used. The maximum and minimum transform sizes are selected by the encoder; the standard supports transform sizes of 4×4, 8×8, 16×16, and 32×32. The transform coefficients are represented using a quantizer with uniformly spaced reconstruction levels.

All slice data syntax elements including the selected coding mode, the intra prediction modes, reference indices, motion prediction indices, motion vector differences, and transform coefficient levels are entropy-coded using context-adaptive binary arithmetic coding (CABAC).

The HEVC design supports two in-loop filters. The first one is a de-blocking filter, which adaptively removes block artifacts introduced by block-based motion compensation and transform coding. The second tool is called sample adaptive offset (SAO). It classifies the reconstructed samples into different categories, e.g., depending on edge orientation, and reduces the distortion by adding a separate offset for each class of samples.

3. SCALABLE HEVC EXTENSION

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. Temporal scalability is already supported by the flexible reference picture handling in HEVC. Changes of the HEVC design are only required for supporting spatial and quality scalability.

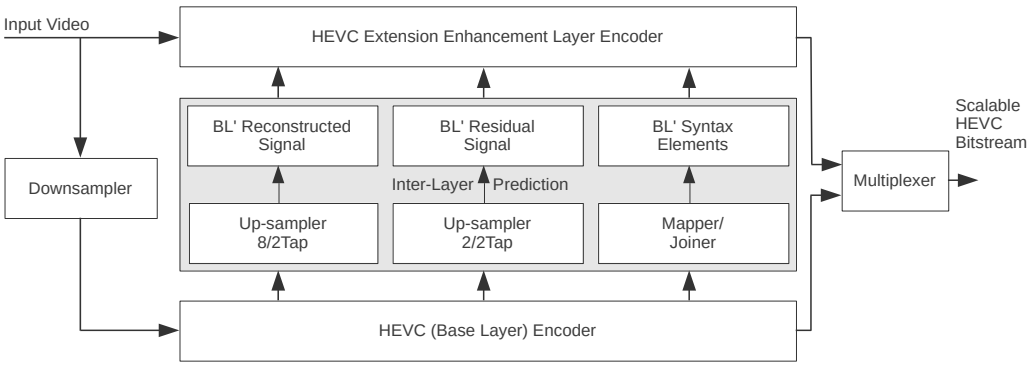

Figure 4: Simplified block diagram of the encoder for the proposed scalable extension of HEVC.

A simplified block diagram of the proposed HEVC extension for a configuration with two layers is depicted in Figure 4. For spatial scalable coding, the input video for the base layer encoder is downsampled to the base layer resolution, while the same input video as for the enhancement layer is used in quality scalable coding. The base layer coder conforms to a single layer standard, so that backward compatibility with single-layer coding is achieved. In the proposed design HEVC as well as H.264 | MPEG-4 AVC are considered as base layer codecs. The enhancement layer encoder represents an extended version of HEVC. The outputs of the encoders for base and enhancement layers are multiplexed to form the scalable bitstream. The multiplexing is done on an access unit basis. An access unit basically represents the data for a particular time instant. For each access unit, first all network abstraction layer (NAL) units (NAL units represent the basic data packets in HEVC or H.264 | MPEG-4 AVC) for the base layer are transmitted. Thereafter, the NAL units of the first enhancement layer are transmitted, etc. Then, the next access unit again starts with the base layer NAL units. Each NAL unit contains an identifier inside the NAL unit header, which specifies to which layer the NAL unit belongs

to. Hence, for extracting a sub-bitstream with a reduced spatial resolution or quality, data packets that belong to a particular layer can be easily identified and forwarded or discarded.

For coding a spatial or quality enhancement layer, basically the same concepts as for single layer HEVC are used. A picture is partitioned into coding units, for which a prediction mode can be chosen. Inter-coded CUs are further partitioned into prediction units that are associated with a set of motion parameters. The residual for a CU is coded using a partitioning in transform units. However, to improve the enhancement layer coding efficiency relative to simulcasting the layers, additional inter-layer prediction methods are incorporated.

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 prediction tools was included. In these standards, it can be signaled that the reconstructed and, for spatial scalability, upsampled base layer is used as prediction signal for a macroblock. The scalable extension of H.264 | MPEG-4 AVC [15] provides basically three methods of inter-layer prediction. A first tool enables the re-usage of the base layers coding mode and inter-prediction partitioning. The syntax includes a flag at the beginning of the macroblock syntax, which specifies whether the coding mode and inter-prediction partitioning is re-used for the enhancement layer coding. If this flag is equal to 1 and the co-located area in the base layer is coded using an intra prediction mode, the prediction signal for the enhancement layer is formed by the reconstructed and, for spatial scalability, upsampled base layer signal, similar to H.262 | MPEG-2 Video, H.263, and MPEG-4 Visual. If the flag is equal to 1 and the co-located area is not coded using an intra prediction mode, the enhancement layer block is coded using motion-compensated prediction mode and the partitioning used for motion compensation as well as all motion parameters are completely inferred from the co-located area in the base layer. As a second tool, the motion vector predictor can be adaptively selected between the conventional spatial motion vector predictor and an inter-layer motion vector predictor, which represents the motion vector of the co-located base layer block, scaled according to the resolution ratio between base and enhancement layers. As a third tool, the design includes the concept of residual prediction for inter-coded macroblock. It can be adaptively selected whether the reconstructed and, for spatial scalability, upsampled residual signal of the co-located base layer area is used as prediction for the residual signal of the enhancement layer macroblock.

As a particular feature, the SVC extension of H.264 | MPEG-4 AVC was designed in a way that each spatial resolution or quality layer could be decoded with a single motion compensation loop. The intention of this single-loop decoding feature was to reduce the decoder complexity by eliminating the reconstruction path of the base layer. However, applications in error prone environments usually use multi-loop decoding for improving the error resilience capabilities. Furthermore, with single-loop decoding not all blocks in the enhancement layer can employ inter-layer intra prediction and the intra blocks in the base layer must not depend on inter-coded blocks, i.e., constrained intra prediction must be used. Both aspects have a negative impact on coding efficiency. The proposed scalable extension of HEVC focuses on multi-loop decoding, which imposes no restrictions on the base-layer and offers new opportunities for combining prediction signals from base and enhancement layer for further improving the enhancement layer coding efficiency.

3.1 Intra prediction using reconstructed base layer samples

The first scalable coding tool is the so-called IntraBL prediction mode, in which the enhancement layer prediction signal is formed by copying or up-sampling the reconstructed samples of the co-located area in the base layer, as illustrated in Figure 5. This coding mode is already known from the scalable profiles of H.262 | MPEG-2 Video, H.263, and MPEG-4 Visual. It is also similar to the inter-layer intra mode in the scalable extension of H.264 | MPEG-4 AVC. However, in H.264 | MPEG-4 AVC, the usage of this mode is restricted to macroblocks for which the co-located base layer area is coded using an intra prediction mode (for enabling single-loop decoding).

In the proposed HEVC extension, this mode can be selected at a CU level. This mode is also used implicitly for the InterBL mode, which is described in sec. 3.6, when a CU completely or partially covers an intra block in the base layer. The residual signal is transmitted by transform coding using the syntax for inter-predicted CUs. At the decoder side, the final reconstruction signal is obtained by adding the transmitted residual signal to the inter-layer intra prediction signal.

In order to use base layer signals for prediction in the enhancement layer, up-sampling filters are required if the enhancement layer uses a larger spatial resolution than the base layer. While the scalable extension of H.264 | MPEG-4 AVC uses 4-tap FIR filters for upsampling of the luma signal [16], 8-tap filters are applied in the proposed HEVC extension. For chroma, bi-linear filters are used. The filters are 2-d separable, i.e., 1-d filters operate horizontally and vertically. Similar to H.264 | MPEG-4 AVC, the filters are provided with approximately 1/16th sample phase offsets. For supporting arbitrary resolution ratios, for each enhancement layer sample position, the used filter is selected based on the required phase shift [16].

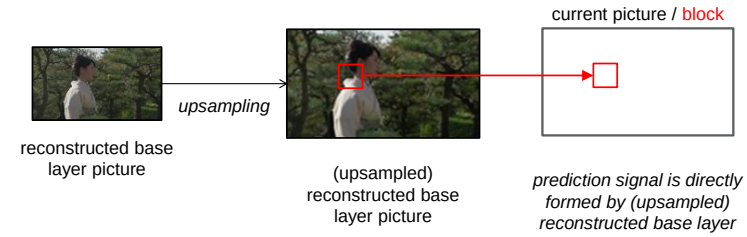

Figure 5: IntraBL prediction mode: The (upsampled) base layer samples are used to predict an enhancement layer CU.

The upsampling filters used for 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, another inter-layer intra coding mode, referred to as InterBLFilt mode, is introduced. This mode is used in the same way as the InterBL mode. The only difference is that, for generating the enhancement layer prediction signal, a smoothing filter with coefficients [1 2 1] / 4 is applied horizontally and vertically after upsampling or copying the reconstructed base layer samples.

If the IntraBL or IntraBLFilt mode is selected, the intra deblocking filter strength as specified in HEVC can be too high, since the base layer signal that is used as prediction has already been deblocked. This is taken into account by adapting the de-blocking filter strength derivation in the enhancement layer.

3.2 Intra prediction using a difference signal

In the IntraBL and IntraBLFilt modes only the reconstructed base layer samples are used for generating the prediction signal for an enhancement layer. As a consequence, the prediction signal has a systematic error. In spatial scalable coding, the generated prediction signal mainly contains low-frequency components; the high-frequency components which represent the difference between the base and enhancement layer resolution are missing in the enhancement layer prediction signal. In quality scalable coding, the quantization step size used for the base layer is larger than the quantization step size for the enhancement layer. Hence, the inter-layer intra prediction signal contains the larger quantization noise of the base layer. To some extent, this effect also appears in spatial scalable coding. For reducing these systematic errors, two further inter-layer intra prediction modes have been introduced, for which the final enhancement layer prediction signal is obtained by superimposing two intermediate prediction signals. The first intermediate prediction signal is generated by using the reconstructed base layer signal of the co-located area in the base layer and the second intermediate prediction signal is generated by using enhancement layer data.

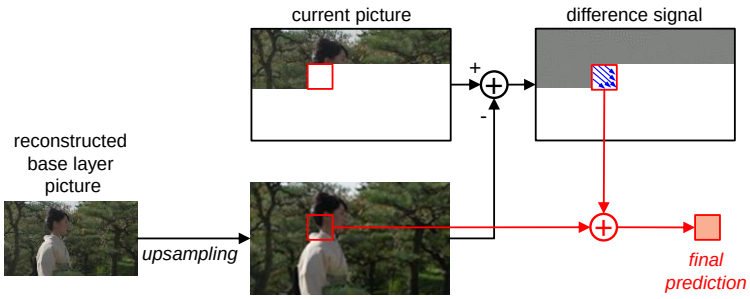

Figure 6: Spatial intra prediction using a difference signal. The (upsampled) base layer signal is added to the intra-predicted difference signal to form the final intra prediction signal.

The first combined base and enhancement layer intra prediction mode, which is also referred to as DiffIntra mode, uses a difference signal between already reconstructed enhancement layer samples and (upsampled) base layer samples [17]. The generation of the prediction signal for an enhancement layer block is illustrated in Figure 6. The first component of the enhancement layer prediction signal is derived by copying or, for spatial scalable coding, upsampling the reconstructed base layer samples of the co-located area in the base layer. For upsampling the base layer signal, the same

interpolation filters as for the IntraBL mode described in Sec. 3.1 are used. The second component of the prediction signal is derived by spatial intra prediction using a difference signal for neighboring samples of already reconstructed blocks. This difference signal represents the sample-wise difference between the reconstructed enhancement layer signal and the upsampled or copied reconstructed base layer signal of the co-located area. For spatial intra prediction, the same spatial intra prediction modes as for single-layer HEVC are used. The selected intra prediction modes are transmitted using the conventional HEVC syntax. Basically, the only difference to the spatial intra prediction as specified in HEVC is that difference samples instead of reconstructed enhancement layer samples are used. The final enhancement layer prediction signal is obtained by adding up the (upsampled) base layer signal and the intra-predicted difference signal.

Similar to the IntraBL and IntraBLFilt modes, the DiffIntra mode can be selected at a CU level. The residual signal is transmitted via transform coding using the syntax for intra-predicted CUs.

3.3 Weighted intra prediction using base and enhancement layer samples

Another intra prediction mode in which a reconstructed base layer signal is combined with an enhancement layer prediction signal is called WeightIntra. Similarly to the DiffIntra mode, the reconstructed and, for spatial scalable coding, upsampled base layer signal of the co-located area in the base layer constitutes one component of the prediction signal. For upsampling, the same interpolation filters as for the IntraBL mode are used. The second component is obtained by conventional spatial intra prediction using neighboring enhancement layer samples of already reconstructed blocks. The selected spatial intra prediction modes are transmitted using the regular HEVC syntax. The final enhancement layer prediction signal is obtained by low-pass filtering the base layer component, high-pass filtering the enhancement layer component, and adding up the resulting signals.

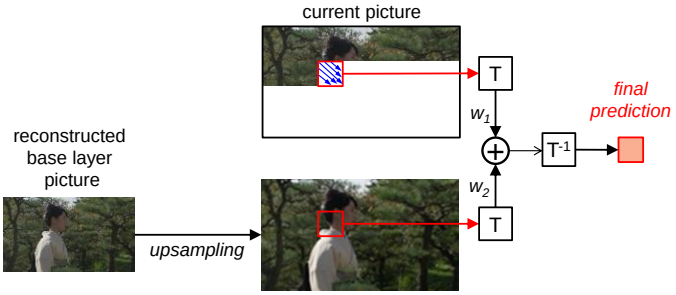

Figure 7: Weighted intra prediction. The (upsampled) base layer signal is low-pass filtered and combined with a high-pass filtered version of a spatially intra-predicted enhancement layer signal for forming the final intra prediction signal.

In our implementation, the low- and high-pass filtering is done in the transform domain as illustrated in Figure 7. Therefore, both intermediate prediction signals are first transformed using an approximation of a DCT of the corresponding block size. Then, the resulting transform coefficients are weighted according to the spatial frequencies, where 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 final enhancement layer prediction signal is obtained by summing up the weighted base and enhancement layer coefficients and applying an approximation of an inverse DCT.

The used inverse transforms are the same as the ones specified in the HEVC decoding process. The used forward transforms are the inverses of these transforms. The weighting matrices have been obtained by least-squares optimization using a set of test sequences. Different weighting matrices are used for different block sizes. Furthermore, one out of three sets of weighting matrices is selected based on the ratio of enhancement and base layer resolution. The sets of weighting matrices have been optimized for resolution ratios of 1.0, 1.5, and 2.0.

The WeightIntra mode can be selected on a CU level and the residual signal is transmitted via transform coding using the syntax for intra-predicted CUs.

3.4 Inter coding with inter-layer residual prediction

Similar to the scalable extension of H.264 | MPEG-4 AVC, inter-layer residual prediction is supported in combination with conventional motion-compensated prediction. The general idea behind residual prediction between layers is that by adding the residual signal of the base layer to the motion-compensated prediction signal in the enhancement layer,

systematic prediction errors that are present in the enhancement layer can be reduced. Such systematic errors could be for example a result of occlusions or lighting changes. In contrast to the scalable extension of H.264 | MPEG-4 AVC where the usage of residual prediction is signaled on a macroblock level, in the HEVC extension, the residual prediction mode, also referred to as InterResPred mode, can be selected on each CU level. In InterResPred mode, the same partitioning types for motion-compensated prediction as for the conventional inter coding mode are supported. The motion parameters are coded in the same way as in single-layer HEVC.

The enhancement layer prediction signal is constructed by adding the reconstructed and, for spatial scalable coding, upsampled base layer residual samples to the motion-compensated prediction signal. Hereby, the base layer residual samples are the samples that are obtained by scaling and inverse transformation of the transmitted transform coefficient levels. For upsampling, bi-linear interpolation is used for both the luma and chroma components. Similar to the scalable extension of H.264 | MPEG-4 AVC and the upsampling in IntraBL mode, interpolation filters with 1/16th sample phase offsets are provided for supporting arbitrary resolution ratios (see Sec. 3.1). The enhancement layer residual signal is coded by transform coding using the syntax for inter-predicted CUs.

3.5 Inter prediction using difference pictures

In addition to the DiffIntra mode described in Sec. 3.2, the proposed scalable extension of HEVC also supports a difference prediction mode for motion-compensated prediction, in which the final enhancement layer prediction signal is built as a sum of a base layer signal and a difference prediction signal. This prediction mode is also referred to as DiffInter mode. The generation of the prediction signal for an enhancement layer block is illustrated in Figure 8. In the same way as for the DiffIntra mode, the first component of the enhancement layer prediction signal is derived by copying or, for spatial scalable coding, upsampling the reconstructed base layer samples of the co-located area in the base layer. For upsampling the base layer signal, the same interpolation filters as for the IntraBL mode described in Sec. 3.1 are used. The second component of the prediction signal is obtained by motion-compensated prediction using difference pictures. The difference pictures represent the difference between the enhancement layer reconstruction for already reconstructed pictures and the corresponding reconstructed and, for spatial scalable coding, upsampled base layer pictures. The upsampling of the base layer for generating the difference pictures is done using the same interpolation filters as for the InterBL mode described in Sec. 3.1. The final enhancement layer prediction signal is obtained by adding the motion-compensated difference signal to the base layer component.

For the motion-compensated prediction using difference pictures, the same partitioning types as for conventional inter modes are supported. The motion parameters are transmitted using the regular HEVC syntax. However, instead of using 8- and 7-tap interpolation filters for generating the motion-compensated prediction signal at fractional sample positions, a simple bi-linear interpolation with quarter sample precision is used.

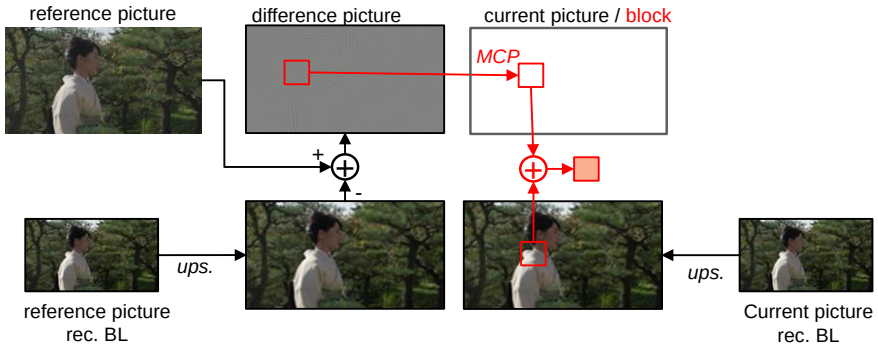

Figure 8: Motion-compensated prediction using difference pictures. The (upsampled) base layer signal is added to a motion-compensated difference signal. The difference pictures used for motion compensation represent the difference between the reconstructed enhancement layer and the (upsampled) reconstructed base layer.

3.6 Inter-layer prediction of prediction parameters

The proposed scalable extension of HEVC supports three methods for using the motion information of the base layer for efficient enhancement layer coding as well as a method for improving the coding of intra prediction modes by using the base layer intra prediction mode. Conceptually, two of the methods for motion parameter coding are similar to the inter-

layer motion prediction in the scalable extension of H.264 | MPEG-4 AVC, but they differ in several details. The methods for motion parameter prediction can be used in connection with all inter coding modes, i.e., they can be used with the conventional inter coding as well as in the InterResPred and DiffInter mode.

As a first method of motion data prediction, the scalable HEVC extension supports a mode in which all motion parameters and coding modes as well as the sub-partitioning into coding units and prediction units are inferred from the co-located region of the base layer. Therefore, if it is signaled that a coding unit is not split into smaller coding units, a syntax element is transmitted which indicates whether all prediction parameters are inferred from the co-located region of the base layer. If the syntax element indicates that the prediction parameters are not inferred, it is signaled whether the CU is coded in inter or intra prediction mode and the corresponding prediction parameters are transmitted. Otherwise, the prediction parameters are inferred from the base layer. Therefore, the CU is decomposed into 8×8 sub-blocks and for each of the sub-blocks a co-located block in the base layer is determined. If the co-located base layer block is coded in an intra prediction mode, the 8×8 enhancement layer sub-block is marked as intra-coded. Otherwise, the 8×8 sub-block is associated with the base layer motion parameters. The number of motion hypotheses and reference indices are copied from the base layer block, whereas the motion vectors are scaled according to the resolution ratio between base and enhancement layer. The partitioning of the considered enhancement layer CU into smaller CUs and PUs is obtained by recursively joining the 8×8 sub-blocks in a quadtree fashion. If all 4 sub-blocks of an 16×16 block have the same motion parameters (or are all marked as intra-coded), the sub-blocks are represented by a single 16×16 sub-block. In the next step, four neighboring 16×16 blocks can be represented by a single 32×32 block, etc. And if four neighboring blocks do not have the same motion parameters, but each horizontal or vertical pair of the sub-blocks has the same prediction parameters, the corresponding PU partitioning is selected. This joining of sub-blocks is done until no further joining is possible and a valid CU/PU partitioning for the considered CU is obtained. It should be noted that the derivation of the CU/PU partitioning and associated prediction parameters is the same for both an HEVC compliant and an H.264 | MPEG-4 AVC compliant base layer.

After deriving the CU/PU partitioning, the inter-coded blocks are predicted using the associated motion parameters. The blocks that are marked as intra-coded are predicted in the same way as CUs coded in IntraBL mode, i.e., the prediction signal is given by the reconstructed and, for spatial scalable coding, upsampled base layer signal of the co-located region (see sec. 3.1). For transform coding of the residual signal, transform block boundaries have to be determined. On the one hand, the signaling of transform block boundaries could benefit from taking into account the derived CU/PU structure, but on the other hand, this would introduce unwanted parsing dependencies. In our implementation, the parsing dependencies are avoided by using the same residual quadtree (RQT) syntax as for conventional CUs in HEVC. Hence, the residual signal is transmitted using the regular HEVC syntax for inter-coded CUs; the potential splitting into smaller CUs as a result of inferring base layer motion data is ignored for the purpose of residual coding.

If the motion parameters of an inter-coded CU are not completely inferred from the base layer as described above, they are basically coded as for conventional inter-coded CUs in HEVC. However, we have incorporated inter-layer motion parameter prediction concepts into these motion parameter coding methods as will be described in the following.

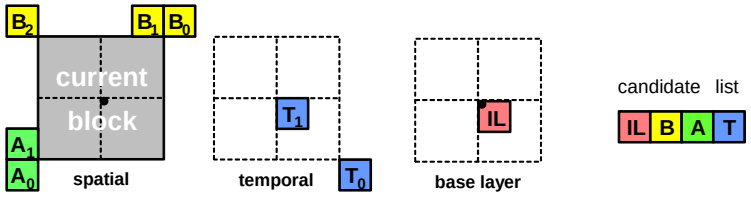

Figure 9: Derivation of the motion vector prediction candidate list using spatial, temporal, and inter-layer predictors. For the candidates A, B, and T, the blocks are evaluated in the order indicated by the subscripts and the first available candidate for the considered reference list and reference index is selected.

In HEVC there are two methods of transmitting motion data for a PU. In the first method, the number of motion hypotheses specifying uni-directional prediction or bi-prediction as well as a reference picture index and a motion vector difference for each motion hypothesis are transmitted. The motion vector predictor is selected by transmitting an index into the so-called advanced motion vector prediction (AMVP) list. The AMVP list in HEVC includes two spatial and one temporal motion vector predictor as potential candidates for motion vector prediction. In the scalable extension, an

additional inter-layer motion vector predictor extending the candidate list by one item is included in the AMVP list as illustrated in Figure 9. The inter-layer candidate is added if the co-located base layer prediction block contains motion information for the considered reference picture list and reference picture index. In this case, the corresponding motion vector is scaled according to the resolution ratio and inserted at the first position in the candidate list of the current prediction block in the enhancement layer. The co-located base layer block is the block that covers the sample location of the center sample of the current enhancement layer PU. The syntax for coding the motion parameters in the AMVP mode is not modified.

The second method for coding motion parameters in HEVC is the so-called merge mode, which replaces the skip mode used in prior standards. Similar to AMVP, a candidate list of motion parameters is constructed. But in contrast to AMVP, the candidate motion parameters include the number of motion hypotheses, the reference indices, and the motion vectors. In single-layer HEVC, this candidate list typically includes a temporally predicted set of motion parameters and the motion parameters of up to 4 neighboring blocks. The motion parameters are signaled by transmitting only an index into the candidate list.

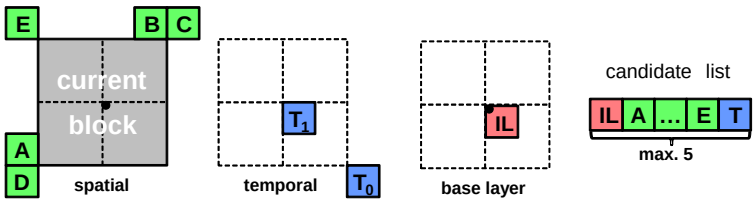

Figure 10: Derivation of the candidate list for the merge mode using spatial, temporal, and inter-layer predictors.

In the scalable HEVC extension, an inter-layer motion parameter candidate is included into the candidate list. Therefore, the center sample location is mapped to the base layer and the base layer block covering this sample is evaluated. If this base layer block is coded in an inter mode but not in merge mode, its motion parameters are inserted at the first position in the candidate list. Similarly as for AMVP, the motion vectors are scaled according to the resolution ratio. If the co-located base layer block is coded in an intra mode or its motion parameters are transmitted using the merge syntax, the inter-layer motion parameters are not inserted into the candidate list. After the potential insertion of the inter-layer candidate, the construction of the merge list continues as specified in HEVC. The total length of the merge candidate list is not extended and remains the same as in single-layer HEVC. After the candidate list construction is finished, a re-ordering process may be invoked depending on the co-located base layer prediction block that covers the top-left sample position inside the current PU. If this block is coded in merge mode, the origin of the motion vector predictor in the base layer is retrieved. The candidate list of the enhancement layer is checked if it includes a merge candidate that stems from the same origin. If such a candidate is found, it is moved to the top of the candidate list. The order of the remaining entries in the candidate list is retained. By this re-ordering process, it is taken into account that the motion description in base and enhancement layer should be similar. If a base layer block is merged with a particular neighboring block, it is likely that the co-located enhancement layer block is merged with the candidate at the same location.

As a further method for improving the coding of enhancement coding parameters by using base layer coding parameters, the prediction of spatial intra prediction modes in the enhancement layer has been slightly modified. In HEVC, a list of three most probable intra prediction modes is derived based on the intra prediction modes of the blocks that cover the sample to the left and the sample above the top-left sample of the current block. If the intra prediction mode of the current block is in the list of the most probable modes, it is very efficiently coded as an index into this list. Otherwise, the intra prediction mode is transmitted using a fixed-length code and requires more bits. In the proposed scalable extension of HEVC, the intra prediction mode derived from the above or left block is, under certain conditions, replaced by the intra prediction mode of the co-located base layer block, so that the base layer intra prediction mode is used for deriving the list of most probable modes. Therefore, it is checked if the co-located base layer is coded using an angular intra prediction mode. If this is not the case, the list of most probable modes is determined as specified in single-layer HEVC. Otherwise, if one of the intra prediction modes derived from the above and left block does not represent an angular intra prediction mode, the base layer intra prediction mode is used to replace this mode. In the case that both of the modes derived from the above and left block are non-angular prediction modes, only the mode derived from the left block is replaced by the base layer intra prediction mode. The remaining derivation and coding process for intra prediction modes is not modified relative to single-layer HEVC.

3.7 Entropy coding transform coefficients levels

Compared to HEVC, the coding of transformation coefficients in the enhancement layer is slightly modified. Vertical and horizontal scan patterns are introduced for 16×16 and 32×32 transformation blocks. These new scans follow a simple scheme for subdividing coefficient positions into subgroups. Starting from the DC position, each 16 consecutive positions in scan order form a subgroup, which leads to subgroups of the size 16×1 for horizontal scans, and 1×16 for vertical scans, respectively. Thus, in the case of 16×16 TUs, each row represents a subgroup in the horizontal scan, and each column represents a subgroup in the vertical scan, respectively. This concept is depicted in Figure 11.

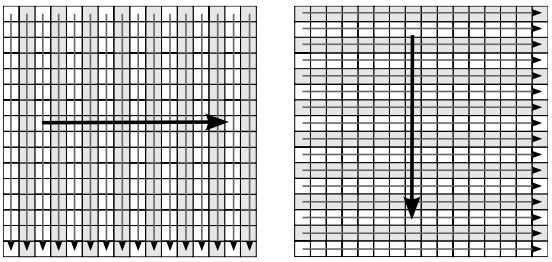

Figure 11: Vertical and horizontal scan for 16×16 transformation units.

The newly introduced scan patterns can be only selected for the luminance signal and for transform blocks with a last significant coefficient position outside the first 4×4 sub-block. For those cases, the scan pattern can be chosen between the conventional diagonal scan, the horizontal scan, and the vertical scan. The selected pattern is signaled inside the bitstream. In our implementation, the encoder decision is done by estimating the cost for each scan, using a scan dependent rate distortion optimized quantization, and selecting the scan pattern with minimum cost. The context model derivation for coding significant coefficient subgroups and the last significant position is slightly modified: A different set of context models is used for TUs for which the scan pattern is explicitly signaled inside the bitstream. Furthermore, in those cases, the coding of the vertical last significant position depends on the horizontal last position. If the horizontal last position exceeds three, a separate set of context models is used for coding the vertical position.

For signaling the selected scan pattern two context models are introduced. First, it is signaled whether the diagonal scan is selected. If the diagonal scan is not selected, it is signaled whether the more likely scan pattern is selected. If the last horizontal position exceeds the vertical position, horizontal scan is set as more likely; otherwise, the vertical scan is set to be more likely.

4. SIMULATION RESULTS

For evaluating the efficiency of the described scalable HEVC extension, the coding efficiency of the scalable approach with two scalable layers is compared to that of simulcast and single layer coding. In addition, the coding efficiency of two scalable versions with a reduced tool set is also evaluated. In the first version, which will be labeled as "IntraBL" in the following, the IntraBL mode described in Sec. 3.1 is the only scalability tool enabled. For the second version, only the tools that are similar to the ones supported in the scalable extension of H.264 | MPEG-4 AVC are enabled. This version will be labeled as "SVC tools" in the following and it includes the IntraBL mode, the concepts for motion parameter prediction, and the residual prediction mode.

For the simulations, the test conditions specified in the Joint Call for Proposals on Scalable Video Coding Extensions of HEVC [13] were followed. 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 each base layer QP, which will be also referred to as BQP, four different enhancement layer QPs have been tested. 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. Using the bit rates and average PSNR values obtained from our experiments and the bit rates and PSNR values of the simulcast anchor provided by the JCT-VC, we calculated the bit-rate savings relative to simulcast, the bit-rate overhead relative to single-layer coding, and a measure we call base layer usage. The base layer usage is given by the difference of the simulcast bit rate and the scalable bit rate (for the

same PSNR) divided by the base layer bit 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.

The anchors have been coded using only the 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 the tables below) to about 4k×2k (for the last 2 test sequences in Table 1 and Table 3). The scalable extension has been implemented in the reference software HM-6.1, which has also been used for producing the anchor bitstreams. The encoders for both scalable and single-layer coding have been operated using the same Lagrangian encoder control, which is described in [11] and implemented in the reference encoder implementation for HEVC. For all scalable encoders, a small change has been made to the derivation of the Lagrange parameters. The Lagrange parameter for the enhancement layer does not only depend on the enhancement layer quantization parameter, but also on the base layer quantization parameter.

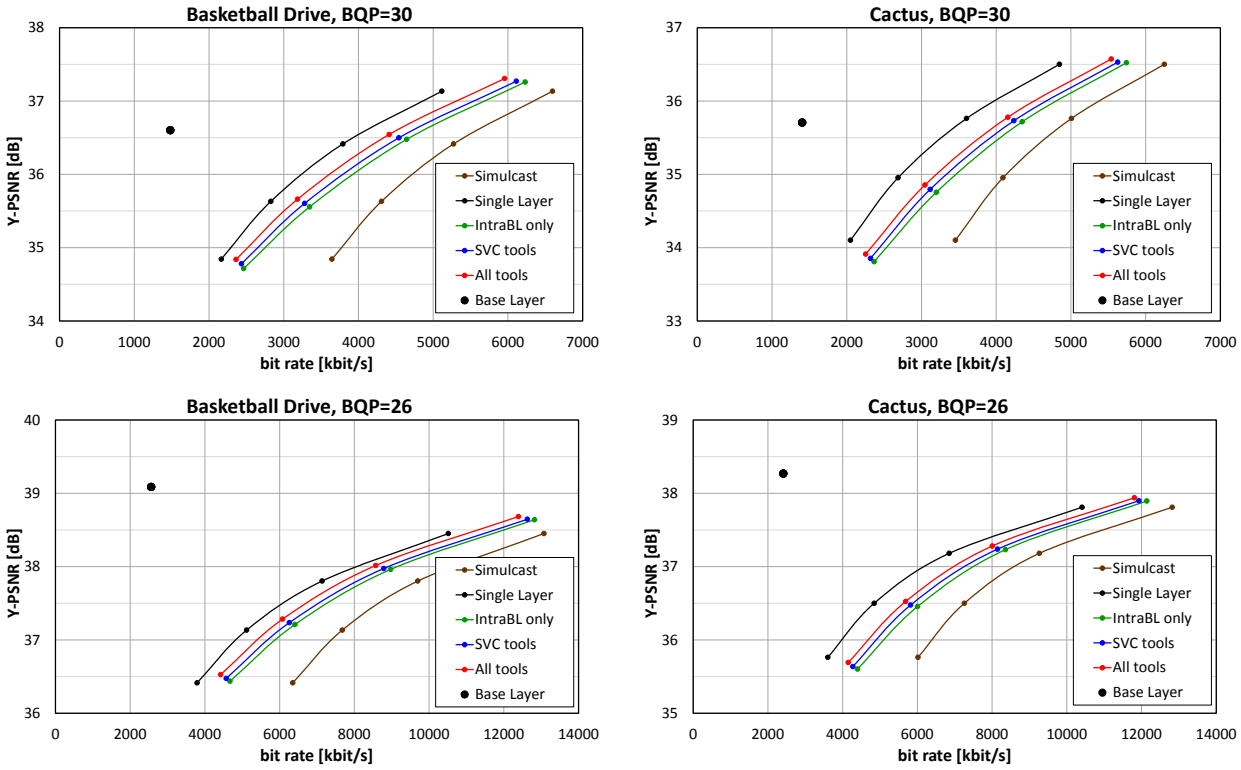

Figure 12: Selected rate-distortion curves for spatial scalability with a resolution ratio of 2. The coding efficiency of the scalable HEVC extension is compared with that of simulcast, single-layer coding, and two versions with a reduced tool set.

Table 1. Average simulation results for spatial scalable coding with a resolution ratio of 2.0.

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
BQTerrace	7.9 %	9.4 %	12.1 %	17.4 %	15.3 %	11.9 %	44.7 %	53.3 %	71.2 %
Basketball Drive	17.3 %	19.7 %	23.4 %	17.0 %	13.4 %	8.3 %	60.5 %	69.0 %	82.3 %
Cactus	14.9 %	17.5 %	20.8 %	20.6 %	16.9 %	12.2 %	52.6 %	61.5 %	74.2 %
Kimono	23.6 %	25.9 %	28.4 %	14.5 %	11.1 %	7.3 %	70.0 %	77.1 %	84.9 %
Park Scene	14.4 %	16.1 %	18.1 %	20.0 %	17.5 %	14.7 %	49.2 %	55.4 %	62.7 %
People on Street	22.0 %	23.5 %	26.9 %	17.9 %	15.5 %	10.3 %	64.7 %	69.3 %	79.8 %
Traffic	14.0 %	16.0 %	18.9 %	24.2 %	21.1 %	17.0 %	44.5 %	51.3 %	61.0 %
Average	16.3 %	18.3 %	21.2 %	18.8 %	15.8 %	11.6 %	55.2 %	62.4 %	73.7 %

Figure 12 shows selected rate-distortion curves for spatial scalable coding with a resolution ratio of 2 and Table 1 summarizes the average simulation results.. The 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 numerical integration. The results shown in Table 1 are averaged over the 4 tested base layer QPs. As can be seen from the results, on average, the proposed scalable HEVC extension provided bit-rate savings of about 21% relative to simulcast for the considered sequences and test cases. The overhead relative to single-layer coding is approximately 12%. And the base layer usage is approximately 74%, which can be interpreted in a way that on average 74% 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, as can be seen in Figure 12. Furthermore, the scalable extension with all tools enabled outperforms the two versions with a reduced tool set.

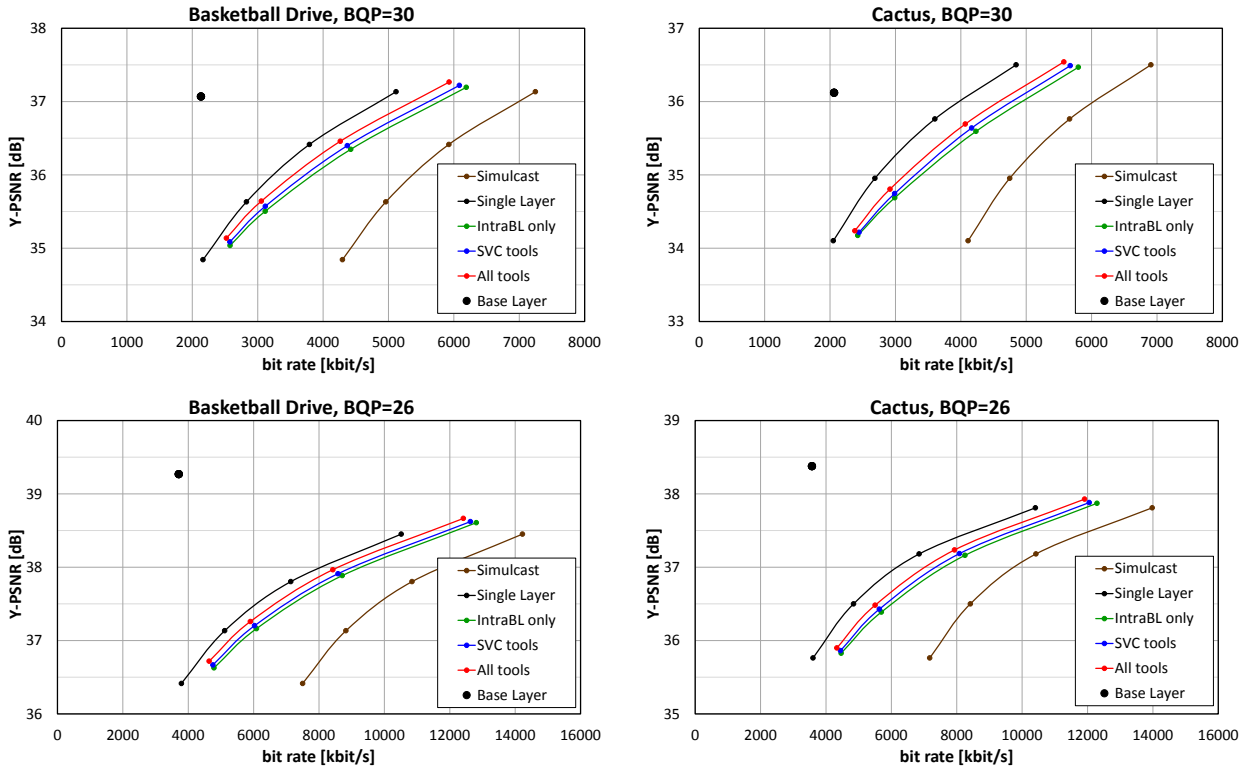

Figure 13: Selected rate-distortion curves for spatial scalability with a resolution ratio of 1.5. The coding efficiency of the scalable HEVC extension is compared with that of simulcast, single-layer coding, and two versions with a reduced tool set.

Table 2. Average simulation results for spatial scalable coding with a resolution ratio of 1.5.

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
BQTerrace	15.7 %	17.8 %	20.6 %	19.5 %	16.6 %	12.5 %	56.9 %	64.2 %	77.5 %
Basketball Drive	26.5 %	28.4 %	31.7 %	15.3 %	12.2 %	7.1 %	73.2 %	78.8 %	88.4 %
Cactus	24.9 %	26.8 %	30.2 %	19.8 %	16.7 %	11.4 %	67.1 %	72.6 %	82.5 %
Kimono	32.3 %	33.8 %	36.0 %	11.9 %	9.5 %	5.9 %	80.4 %	84.5 %	90.3 %
Park Scene	25.6 %	27.2 %	29.2 %	19.8 %	17.2 %	14.0 %	66.0 %	70.5 %	75.9 %
average	25.0 %	26.8 %	29.5 %	17.3 %	14.4 %	10.2 %	68.7 %	74.1 %	82.9 %

Selected rate-distortion curves for spatial scalable coding with a resolution ratio of 1.5 are shown in Figure 13. The average simulation results are summarized in Table 2. Here, the proposed scalable HEVC extension provided bit-rate savings of approximately 30% relative to simulcast. The overhead relative to single-layer coding is approximately 10%

and the base layer usage is approximately 83%. As for a resolution ratio of 2, the effectiveness of the proposed scalable HEVC extension generally improves with increasing base layer rate and the scalable extension with all tools enabled outperforms the two versions with a reduced tool set. Furthermore, it can be seen that the effectiveness of the scalable HEVC extension (which can be measured by the introduced base layer usage) for a resolution ratio of 1.5 is higher than that for a resolution ratio of 2.

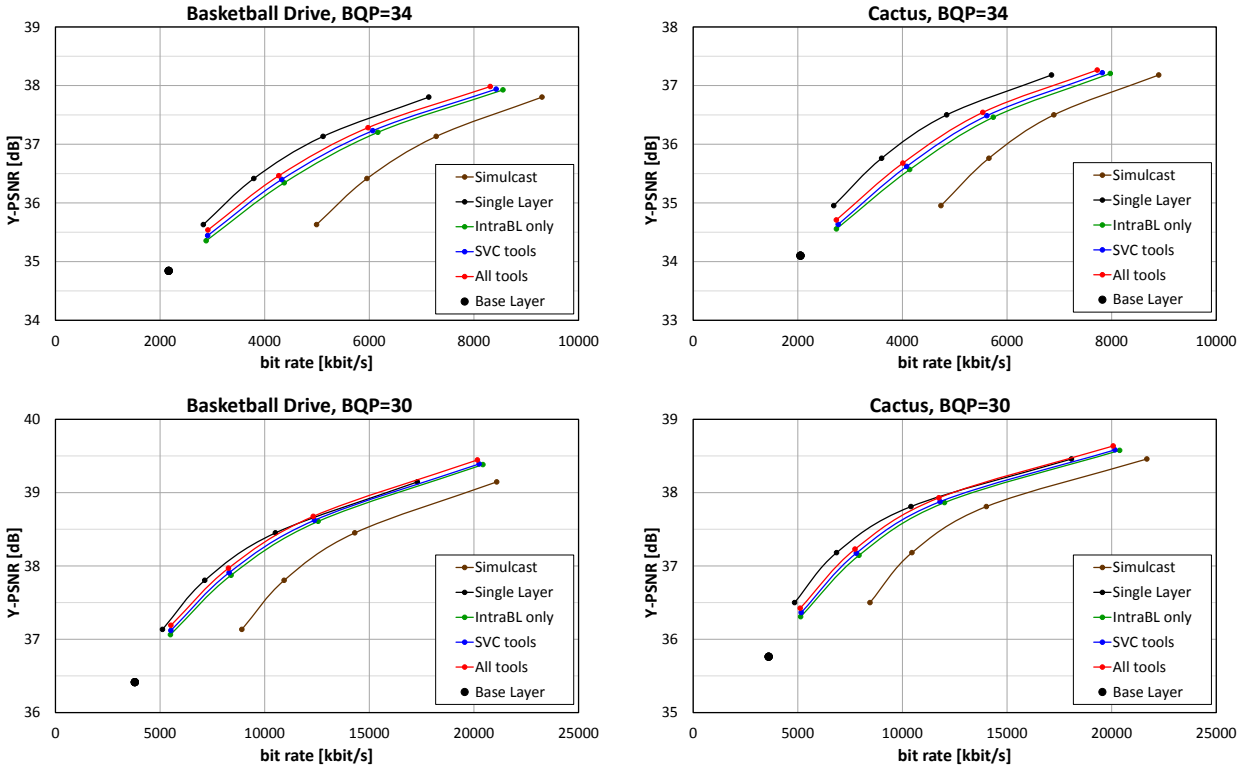

Figure 14: Selected rate-distortion curves for quality scalability. The coding efficiency of the scalable HEVC extension is compared with that of simulcast, single-layer coding, and two versions with a reduced tool set.

Table 3. Average simulation results for quality scalable coding.

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
BQTerrace	15.5 %	17.2 %	20.9 %	11.2 %	8.8 %	4.1 %	72.1 %	79.2 %	101.8 %
Basketball Drive	22.4 %	24.3 %	27.2 %	11.8 %	9.0 %	4.8 %	77.0 %	83.3 %	94.6 %
Cactus	19.6 %	21.6 %	24.9 %	15.3 %	12.4 %	7.8 %	69.2 %	75.9 %	88.9 %
Kimono	24.2 %	26.1 %	28.3 %	13.3 %	10.4 %	7.2 %	73.2 %	79.1 %	86.0 %
Park Scene	20.6 %	22.2 %	23.5 %	16.3 %	14.1 %	12.0 %	64.0 %	69.0 %	73.6 %
People on Street	25.5 %	26.3 %	30.6 %	12.1 %	10.8 %	4.5 %	75.8 %	78.4 %	91.5 %
Traffic	19.4 %	21.1 %	23.5 %	18.9 %	16.3 %	12.8 %	59.4 %	65.1 %	72.8 %
average	21.0 %	22.7 %	25.5 %	14.1 %	11.7 %	7.6 %	70.1 %	75.7 %	87.0 %

Figure 14 shows selected rate-distortion curves for quality scalable coding. The average simulation results are summarized in Table 3. The bit-rate savings relative to simulcast are approximately 25%. The overhead relative to single-layer coding is approximately 8% and the base layer usage is approximately 87%. The effectiveness of the proposed scalable HEVC extension generally improves with increasing base layer rate and the scalable extension with all tools enabled outperforms the two versions with a reduced tool set. Based on the measured base layer usage, it can be concluded that the scalable HEVC extension is more effective for quality scalability than for spatial scalable coding.

5. SUMMARY

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 video coding standards. In contrast to the scalable extension of H.264 | MPEG-4 AVC, the proposed HEVC extension has been designed for multi-loop decoding, which provides more freedom to develop improved scalable coding tools. In particular, the multi-loop design offered the possibility to include coding mode, in which a base and enhancement layer prediction signal are combined for forming an improved enhancement layer prediction signal. The scalable coding tools have been integrated in a way that the low-level syntax and decoding process of single-layer HEVC remains unchanged to a large extent. The effectiveness of the described approach has been demonstrated by experimental results for quality scalability and for different spatial scaling factors. It has been shown that the new coding tools provide an additional increase in coding efficiency relative to the tools known from the scalable extension of H.264 | MPEG-4 AVC.

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 Video), 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), version 1: 2003, version 2: 2004, version 3, 4: 2005, version 5, 6: 2006, version 7, 8: 2007, version 9, 10, 11: 2009, version 12, 13: 2010, version 14, 15: 2011, version 16: 2012.

[5] P. Choi, J. W. Woods, "Motion-compensated 3-d subband coding of video," IEEE Trans. Image Process., vol. 8, pp. 155-167, Feb. 1999.

[6] A. Secker, D. S. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," IEEE Trans. Image Process., vol. 12, pp. 1530-1542, Dec. 2003.

[7] R. Xiong, J. Xu, F. Wu, S. Li, "Barbell-lifting based 3-d wavelet coding scheme," IEEE Trans. Circuits Syst. Video Technol., vol. 17, pp. 1256-1269, Sep. 2007.

[8] R. Xiong, J. Xu, F. Wu, "In-Scale Motion Compensation for Spatially Scalable Video Coding," IEEE Trans. Circuits Syst. Video Technol., vol. 18, pp. 145-158, Feb. 2008.

[9] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 9," Joint Collaborative Team on Video Coding (JCT-VC), doc. JCTVC-K1003, Oct. 2012.

[10] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Trans. Circuits Syst. Video Technol., Dec. 2012.

[11] 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)," IEEE Trans. Circuits Syst. Video Technol., Dec. 2012.

[12] F. Bossen, B. Bross, K. Sühring, D. Flynn, "HEVC Complexity and Implementation Analysis," IEEE Trans. Circuits Syst. Video Technol., Dec. 2012.

[13] 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. MPEG/N12957, July 2012.

[14] 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.

[15] H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," IEEE Trans. Circuits Syst. Video Technol., vol. 17, pp. 1103-1120, Sep. 2007.

[16] C. A. Segall, G. J. Sullivan, "Spatial Scalability within the H.264/AVC Scalable Video Coding Extension," IEEE Trans. Circuits Syst. Video Technol., vol. 17, pp. 1121-1135, Sep. 2007.

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