DECODER COMPLEXITY REDUCTION FOR THE SCALABLE EXTENSION OF HEVC

Christian Feldmann, Fabian Jäger, Mathias Wien

Institut für Nachrichtentechnik, RWTH Aachen University

ABSTRACT

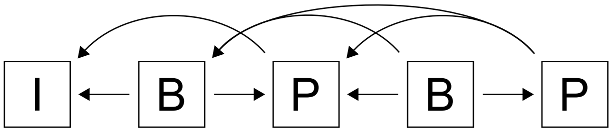
In the current standardization process of the scalable extension to High Efficiency Video Coding (SHVC) a high level syntax multi-loop approach is close to completion. On the one hand this multi-loop approach offers a reasonable rate-distortion performance while only minimal modifications to the encoder and decoder in both layers are required. On the other hand this approach requires full reconstruction of all pictures of all layers at the decoder side which, in the case of quality scalability with two layers, doubles the decoder complexity.

In this paper high layer modifications to the prediction structure similar to the scalable extension of H.264—AVC are implemented in SHVC and studied. These modifications allow for an enhancement layer decoder implementation to skip a significant amount of motion compensation and deblocking operations in the base layer. It is shown that the decoder complexity can hereby be reduced up to 55% for the random access configuration and up to 64% for the low delay configuration compared to SHVC. An overall coding performance increase of 1.2% when decoding the enhancement layer is reported while when only decoding the base layer a drift can be observed between -0.16 dB for random access and -0.39 dB for low delay.

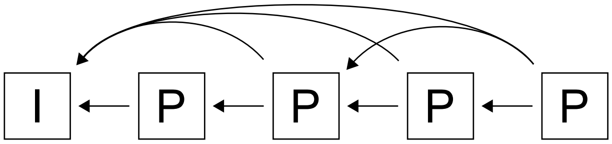
***Index Terms*—** Scalable Video Coding, SNR Scalability, HEVC, Complexity reduction

1. INTRODUCTION

Parallel to the standardization of the HEVC standard the development of a scalable extension to this standard in the Joint Video Team on Video Coding (JCT-VC) has been started [1]. This Scalable High Efficiency Video Coding (SHVC) standard is intended to cover the cases of temporal, spatial and Signal to Noise Ratio (SNR) scalability. The first version of SHVC is planned to be approved by mid 2014 and is focused on a high level implementation that requires no low layer changes to the HEVC coding process. While this approach yields a reasonable performance and can be easily realized using existing HEVC inplementations, it significantly increases the decoder complexity since for every layer full decoding of all pictures is required. This increase may be perfectly acceptable for spatial scalability. However, in case of SNR scalablity the decoder complexity directly scales with the number of layers in the stream which can be unsuitable for some applications.



(a) Random Access (RA)



(b) Low Delay (LD)

Fig. 1. Example HEVC prediction structure for a GOP size of 2.

In this paper a high level modification to SHVC is implemented alongside some encoder restrictions. These modifications are similar to the coding scheme used in Scalable Video Coding (SVC) and allow for a significant decoder complexity reduction [2]. By applying minor changes to the prediction structure and some restrictions for coding the lower layers we can enable the Enhancement Layer (EL) decoder to reconstruct a frame without the necessity of the full lower layer reconstruction. This permits the decoder to skip the computational complex motion compensation and filtering operations for the lower layers.

These concepts were also proposed to JCT-VC and were further studied in an Ad Hoc Group (AHG) and in a Core Experiment [3] [4]. While in [5] a related modification has been used for a new residual mapping technique, this paper focuses on a detailed performance and complexity analysis of the altered coding scheme.

The paper is organized as follows. In section 2 a short overview of HEVC and the current SHVC test model is given. In section 3, a description of the applied modifications to the SHVC coding scheme is provided. Section 4 details the reduction of complexity at the decoder. Finally experimental results of the performance and the complexity reduction are presented in chapter 5 and a conclusion is drawn in section 6.

Since the complexity increase is largest for SNR scalabil-

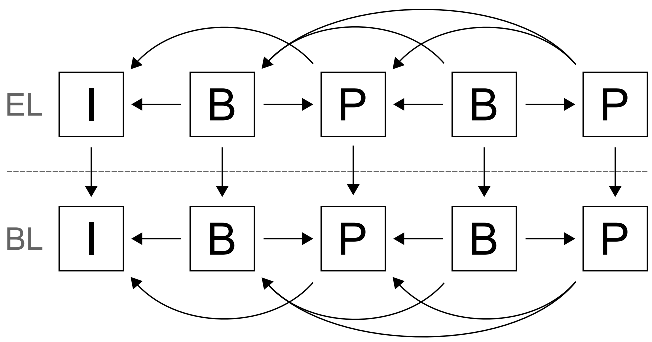


Fig. 2. Scalable prediction scheme of SHVC for the RA configuration and a GOP size of 2.

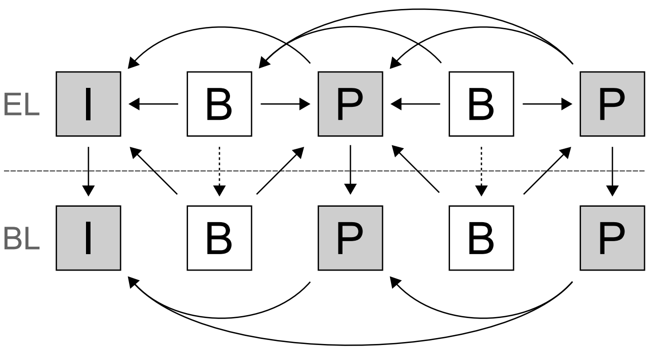


Fig. 3. Modified scalable prediction scheme. Key pictures are marked in grey.

ity we only consider a SNR scalability scenario in this paper. Also the number of layers is set to two, although an extension to more layers is possible.

2. HEVC AND SCALABLE HEVC

Similar to previous standards, the coding order and the display order in HEVC are decoupled. A coding structure can be defined in which the sequence is split into Groups of Pictures (GOP) with a certain coding order and a specific reference structure within each group of pictures (GOP). This enables a very flexible coding structure in HEVC that allows to utilize frames that are in the temporal future for bidirectional (bi-) prediction. Two configurations for particular applications are commonly studied. The Random Access (RA) configuration yields a high coding performance and frequent decoding entry points where the Low Delay (LD) configuration offers a very low structual delay by aligning the display and coding order. Fig. 1 shows an example prediction structure for a GOP size of 2 for the RA and LD configurations [6].

In the current draft of the scalable extension to HEVC the EL reference picture list is extended by the up-sampled corresponding reconstructed picture of the lower layer. The EL can then use conventional Inter prediction to utilize this lower layer reference. In order to limit the complexity of this inter layer prediction approach, the motion vector for this prediction is forced to zero motion so that only a copy operation is performed. This high level approach allows an implementation without any changes to the low layer coding process [7]. Fig. 2 shows a toy example for two layers and a GOP size of 2.

The high level approach yields some advantages compared to a more complex implementation as it was taken for the SVC standard [2]. The amount of modifications which are required is rather low so existing hardware and software implementations can be easily extended to support scalable coding. This also applies to the number of changes in the standardization text.

However, a drawback of this approach is the increased computational complexity at the decoder side. For SVC a single-loop decoding approach was chosen which allowed reconstruction of the EL pictures while skipping the lower layer motion compensation and loop filter operations, thus limiting the increase of computational complexity compared to single layer coding. In SHVC however, a multi-loop approach was chosen where reconstruction, loop filtering and buffering of all pictures in all layers is required. Particularly for SNR scalability this practically multiplies the decoder computational complexity by the number of layers used.

3. PREDICTION STRUCTURE MODIFICATIONS

In order to reduce the decoder complexity we modify the prediction structure and apply some restrictions at the encoder side. Firstly we introduce the concept of key pictures which is also used in SVC [2]. Here the pictures bordering each GOP are marked as key pictures (See Fig. 3). For the non key pictures in the Base Layer (BL) the prediction is changed so that they predict from previously coded EL pictures rather then the corresponding BL pictures. While this increases the overall performance at the EL quality it also raises a problem when decoding only the BL. In this case the EL references are not available for inter prediction and have to be replaced with the corresponding BL pictures which introduces drift at the decoder side. In order to limit the drift in the BL key pictures may only predict from other BL key pictures. Also BL non key pictures are only allowed to predict from EL pictures within the same GOP so the drift is concealed within each GOP. The prediction structure in the EL remains unchanged.

In a second step the inter layer prediction process is modified. For key pictures the inter layer process remains unchanged. The reconstructed and filtered picture from the BL is used in the EL as an additional reference picture. For non key pictures however, the unfiltered lower layer reconstruction is used for prediction in the EL (dotted references in Fig. 3).

The final modification applies to the non key pictures in the BL. Here constrained intra prediction is used in order to make intra predicted areas independently decodable from inter predicted information. This allows for the decoder to skip inter prediction operations when decoding the EL (see the following section).

4. DECODER COMPLEXITY IMPACT

When decoding only the BL the decoder complexity is unchanged and corresponds to the complexity of a single layer decoder. As for SHVC the BL stream is compatible with HEVC and can be decoded by a compliant HEVC decoder.

For the key pictures there is no complexity reduction compared to SHVC. All key pictures in the BL have to be fully reconstructed and filtered in order to reconstruct the other BL key pictures and the corresponding EL key pictures. However, the modifications described in Section 3 allow for the decoder complexity to be significantly reduced when decoding the non key pictures in the EL.

When decoding the non key pictures in the EL a full reconstruction and filtering of the BL pictures is no longer necessary. Only the area using constrained intra prediction in the BL has to be reconstructed. This information can then be utilized for inter layer prediction in the higher layer. If the EL uses inter layer information that uses Inter prediction in the lower layer the reconstruction of this area can be directly performed in the EL. This is possible because both layers use the same EL pictures as references for inter prediction and the motion information between the layers is forced to zero. Only the residual signal of the lower layer has to be reconstructed for the inter layer prediction.

It is noted that while these techniques can be used to reduce the decoder complexity such a decoder implementation is not mandatory. A multi-loop decoder can also be used if simultaneous decoding of all layers is required. In this case the EL decoder only requires access to the unfiltered lower layer reconstruction.

5. RESULTS

Performance results for the proposed coding scheme as well as an analysis of the reduced decoder complexity are presented. The common SHVC test conditions were followed [8]. These conditions consist of a set of 1080p and 4k sequences as well as various quantization parameter (QP) settings for the two layers and configurations for low delay and random access. The GOP size is set to 8 for the RA and 4 for the LD configuration. Key pictures are set at the boundary pictures of each GOP so at every 8th picture for RA and ever 4th picture for LD. For comparison the reference Test model for SHVC (SHM) 3.0.1 software using the common test conditions was used. The complete results that were reported to the AHG as well as the software can be retrieved from [9].

Although the common test condition define only a two layer test case the proposed method can be extended to more than two layers. However, the evaluation of this scenario is outside of the scope of this paper.

		Y	U	V
RA	Class A	-1.2%	-8.6%	-8.5%
	Class B	-1.2%	-6.9%	-8.7%
	Average	-1.2%	-7.4%	-8.6%
LD	Class A	-1.1%	-6.8%	-7.1%
	Class B	-1.1%	-5.9%	-7.8%
	Average	-1.1%	-6.2%	-7.6%

Table 1. Average BD-Rate difference of the presented scheme compared to the SHM reference when decoding EL quality.

		Y	U	V
RA	Class A	-0.21	-0.05	-0.05
	Class B	-0.13	-0.04	-0.06
	Average	-0.16	-0.04	-0.05
LD	Class A	-0.50	-0.06	-0.03
	Class B	-0.35	-0.35	-0.05
	Average	-0.39	-0.39	-0.04

Table 2. Average BD-PSNR difference of the presented scheme compared to the SHM reference when decoding only the BL.

5.1. Coding performance

In Table 1 the Bjøntegaard Delta (BD-rate difference) of the proposed coding scheme compared to conventional SHM is presented [10]. It can be observed that the overall coding performance for both layers increases by about 1.2% for the random access as well as 1.1% for the low delay configuration. This increase can be explained by the prediction from EL pictures in the BL which are of higher quality and thus yield a better prediction signal (see section 3).

While an overall performance increase can be observed when decoding both layers, there is a drift due to the modified prediction structure when decoding only the BL. Table 2 shows the BD-PSNR difference of the decoded BL with drift compared to the SHM BL without drift. The BD-rate difference is not shown here because the drift only impacts the reconstruction quality compared to the drift free SHM approach. On average one can observe a BD-PSNR reduction of 0.16 dB for the RA configuration and 0.39 dB for the LD configuration.

In a first viewing of the drift prone BL reconstruction no obvious artifacts could be observed. However, a subjective viewing comparing to the drift free SHM reconstruction is still outstanding. Also note that an optimized encoder implementation could determine and control the drift by replicating the decoding process with drift.

5.2. Decoder complexity

In order to measure the required motion compensation and deblocking operations the functionality of counting the number of motion compensation and deblocking operations required for decoding is added. The operations are weighted by the number of pixels that are affected by them in order to get an estimate of the required computational complexity. HEVC allows for half and quarter pixel accurate motion compensation. Since half and quarter precision motion compensation require interpolation with differing complexity the operations are counted separately. This functionality was added to the proposed coding scheme, to SHM and also to the single layer test model for HEVC (HM). This way a three way comparison between the approaches can be performed. This complexity measurement was agreed by the participating experts in the AHG [3].

Since the coding process of the BL in SHM is unchanged compared to single layer coding in HM the measured complexity values are identical. For the proposed scheme, the decoding process of the BL itself is not modified so the measured results show no change in the complexity values as well.

Table 3 shows the relative number of operations for full, half and quarter pixel accuracy for both RA and LD configuration when decoding the EL. When comparing the multi-loop SHM to single layer coding (HM), one can see that for SNR scalability there is a major complexity increase compared to single layer coding almost doubling the motion compensation complexity at the decoder. For the proposed coding scheme compared to HM one can observe that the complexity increase compared to HM is significantly reduced. Finally, comparing the two scalable approaches shows that a significant motion complexity reduction can be achieved by the proposed coding scheme. For RA a reduction to 64% and 65% compared to SHM can be observed for the computationally expensive half and quarter pixel precise motion compensation. For LD the complexity reduction is lower with 76% and 72% for half and quarter pixel precision respectively.

In Table 4 the number of pixels that are modified by the deblocking operation for the RA and LD configuration when decoding the EL is shown. Here a similar observation as for the motion compensation complexity can be made. SHM shows a major increase in complexity compared to single layer coding (HM). For the proposed coding scheme this increase is significantly reduced. Compared to SHM the deblocking complexity is reduced to 66.83% for the RA configuration and 69.12% for LD.

	MV	SHM vs HM	Prop vs HM	Prop vs SHM
RA	Full	238%	132%	55%
	Half	188%	121%	64%
	Quarter	187%	121%	65%
LD	Full	229%	146%	64%
	Half	164%	124%	76%
	Quarter	169%	122%	72%

Table 3. Decoder complexity for motion compensation. Comparison of single-loop coding (HM), multi-loop (SHM) and the proposed coding scheme.

	SHM vs HM	Prop vs HM	Prop vs SHM
RA	180%	120%	67%
LD	184%	127%	69%

Table 4. Comparison of deblocking complexity. Comparison of sinle-loop coding (HM), multi-loop (SHM) and the proposed coding scheme.

6. CONCLUSION

In the current standardization process of the scalable extension of HEVC a multi-loop approach is discussed which nearly doubles the decoder complexity in case of SNR scalability. In this paper an approach is presented to significantly reduce the decoder complexity. In order to achieve this a high level change to the coding structure is applied as well as some restrictions to the coding of base layer pictures. This way the base layer only has to be partially reconstructed in order to be able to obtain the enhancement layer reconstruction. The complexity measurements reveal a significant reduction of required operations for motion compensation as well as for the deblocking filter compared to the current multi-loop approach. The complexity reduction is combined with a slight increase in overall performance while some drift can be observed when only decoding the base layer.

Especially for SNR scalability the complexity increase of the current multi-loop approach of SHVC compared to single layer coding might be unacceptable for practical applications. The scheme that is evaluated in this paper offers an alternative approach which is similar to the approach that was taken in scalable video coding (SVC). This significantly reduces the decoder complexity increase that is required for scalable coding in SHVC.

7. REFERENCES

[1] ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, "Joint call for proposals on scalable video coding extensions of high efficiency video coding (HEVC)," Tech. Rep. VCEG-AS90, ITU-T SG16/Q6, July 2012.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *Circuits and Systems for Video Technology,*

IEEE Transactions on, vol. 17, no. 9, pp. 1103–1120, Sept 2007.

[3] M. Wien, M. Budagavi, K. Mishra, K. Ugur, and X. Xiu, “JCT-VC AHG report: Single-loop scalability (AHG16),” Tech. Rep. JCTVC-O0016, Geneva, CH, 15th meeting, Oct. 2013.

[4] M. Wien, K. Rapaka, and X. Xiu, “SCE2 summary report: Key picture concept and single loop decoding,” Tech. Rep. JCTVC-O0032, Geneva, CH, 15th meeting, Oct. 2013.

[5] C. Feldmann, F. Jäger, J. Hsu, and M. Wien, “Single-loop SNR scalability using binary residual refinement coding,” in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, Sept. 2013, pp. 1583–1587.

[6] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, and T. Wiegand, “High efficiency video coding (HEVC) text specification draft 10 (for FDIS & consent),” Tech. Rep. JCTVC-L1003, JCTVC, 12th Meeting, Geneva, Switzerland, Jan. 2013.

[7] J. Chen, J. Boyce, Y. Ye, M. M. Hannuksela, G. J. Sullivan, and Y.-K. Wang, “High efficiency video coding (HEVC) scalable extension draft 5,” Tech. Rep. JCTVC-P1008, JCTVC, 16th Meeting, San José, United States of America, Jan. 2014.

[8] X. Li, J. Boyce, P. Onno, and X. Xiu, “Common SHM test conditions and software reference configurations,” Tech. Rep. JCTVC-N1009, JCTVC, 14th Meeting, Vienna, Austria, July 2013.

[9] C. Feldmann and M. Wien, “[SCE]2: Key picture concept and single loop decoding,” Tech. Rep. JCTVC-O0145, Geneva, CH, 15th meeting, Oct. 2013.

[10] G. Bjøntegaard, “Calculation of average PSNR differences between rd curves,” Tech. Rep. VCEG-M33, ITU-T Q6/16, Austin TX, USA, Apr. 2001.