SINGLE-LOOP SNR SCALABILITY USING BINARY RESIDUAL REFINEMENT CODING

Christian Feldmann *Fabian Jäger* *Juliana Hsu* *Mathias Wien*

Institut für Nachrichtentechnik, RWTH Aachen University

ABSTRACT

In scalable video coding, a stream contains multiple layers of different temporal resolution, spatial resolution, or quality. Each enhancement layer can utilize the information from the lower layers in order to improve the overall coding efficiency. In the draft scalable extension of HEVC the enhancement layer uses the lower layer reconstructed pixel values for prediction. While this dual-loop approach is efficient in terms of rate-distortion performance, it doubles the decoder complexity. In this paper a method is presented that allows for single-loop decoding using a new coefficient refinement scheme for the enhancement layer. The scheme conceptually allows for binary mapping between arbitrary quantizer step sizes in base and enhancement layer. Simulation results reveal a promising rate-distortion performance, especially for small quantizer deltas.

***Index Terms*—** Scalable Video Coding, SNR Scalability, Single-loop Decoding, Residual Refinement, Quantization Mapping

1. INTRODUCTION

Currently, a scalable extension for the new video coding standard High Efficiency Video Coding (HEVC) [1] is under joint development by the Joint Video Team on Video Coding (JCT-VC). Similar to the scalable extension for H.264/AVC (SVC) the new scalable extension is planned to support temporal, spatial and quality scalability (signal to noise ratio (SNR) scalability). However, following the Call for Proposals (CfP) on Scalable High Efficiency Video Coding (SHVC) all responses proposed a multi-loop decoder design that utilizes the reconstructed pixels of the lower layer for prediction [2]. This implies that full decoding of all layers has to be performed on the decoder side. While in the case of spatial scalability the complexity increase for multi-loop decoding may be limited, multi-loop SNR scalable decoding is expected to induce a significant increase of the overall decoder complexity compared to single-loop decoding. In this paper, we use a single-loop SNR scalability concept with a prediction structure similar to SVC. The proposed method inherits the coding tree, transform tree, motion information, and residual information from the lower layer but does not require reconstruction. For residual coding a new coding mode is introduced that performs residual refinement in the transform domain. The scheme employs a binary refinement step in successive SNR layers and allows for re-writing of the multi-layer residual signal to a single-layer residual. The results were also presented to the JCT-VC and lead to the creation of an ad-hoc group for further studies [3].

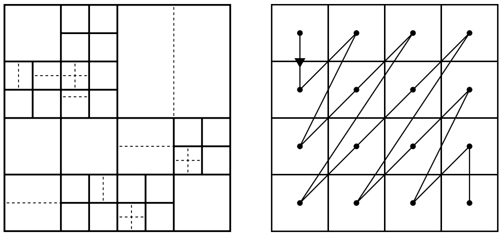
The paper is organized as follows. In Section 2 a short overview of HEVC and the current SHVC test model is given. The single-loop coding approach and the proposed binary residual refinement coding are detailed in Sections 3 and 4, respectively. Finally in Sections 5 and 6 the performance of the approach is evaluated and conclusions are drawn.

2. HEVC AND SHVC

In HEVC, each frame is subdivided into Coding Tree Units (CTUs) and each CTU is subdivided into Coding Units (CUs) of different size. Each CU is then split into one or four square prediction units (PUs) or into two non-square PUs, which can also be asymmetric (asymmetric motion partitions (AMP)) (see Fig. 1a). Each prediction unit is then predicted from the current frame using already decoded neighboring pixel values (Intra prediction) or from previously coded frames using motion compensation (Inter prediction).

After prediction, every CU is split into Transform Units (TUs) of different size. Each TU is transformed and the transform coefficients are quantized. For coding, the quantized transform coefficients are processed in a diagonal scan (see Fig. 1b). At first, the (x, y) position of the last significant (non zero) coefficient is coded. In the subsequent significance scan, all coefficients are processed in a reversed diagonal scan from the position of the last significant coefficient and a significance flag is coded for every coefficient. For all significant coefficients, subsequent flags indicate if each coefficient level is greater than one and greater than two. In case the level is greater than two, the remaining level difference is coded. In the last step, the signs of all significant coefficients in the TU are coded. All flags are coded using Context Adaptive Binary Arithmetic Coding (CABAC) [4]. The remaining level and the sign information is coded in CABAC bypass mode.

This work was co-funded by the German federal state North Rhine Westphalia (NRW) and the European Union (European Regional Development Fund: Investing In Your Future).



(a) Partitioning of a CTU into CUs and PUs (dashed).

(b) Diagonal scan of a 4×4 transform coefficient block.

Fig. 1: HEVC coding tree structure and diagonal scan.

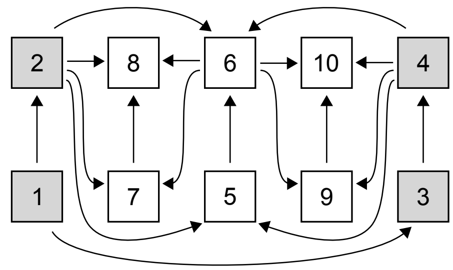


Fig. 2: Hierarchical key picture prediction structure for 4 frames. Numbers denote the coding order of the frames. Key pictures are marked in grey.

As additional tools, Rate-Distortion Optimal Quantization (RDOQ) and Sign Data Hiding are used. These techniques can alter the coded levels in order to achieve a better rate-distortion (RD) performance for the coded coefficients. For details see [5].

After the CfP on SHVC, a Common Scalable Model under Consideration (SMuC) was established which uses only basic inter layer prediction techniques [2]. In this minimal scalability approach, a new Intra BL prediction mode is introduced. This mode is signalled for every CU and indicates that the base layer reconstruction is to be used as prediction. The residual signal is then coded using the conventional HEVC coding scheme. This implies that motion compensation and loop filtering has to be performed in both layers when decoding the enhancement layer.

3. SINGLE-LOOP CODING SCHEME

In the following section, a scalable single-loop coding approach is presented that does not utilize the reconstruction of the lower layer. With this constraint, the enhancement layer can be decoded without performing motion compensation and in-loop filtering in the base layer which significantly reduces the complexity of the decoder.

Similar to the scalable extension of H.264/AVC (SVC), the key picture concept is used for prediction [6]. In this prediction technique, the base layer pictures use the enhancement layer picture reconstruction for inter prediction (see Fig. 2). On the one hand, this allows for a more efficient prediction in the base layer because a reconstruction of higher quality is used. On the other hand, the approach introduces drift when only the base layer is decoded as the enhancement layer reconstruction is not available for prediction in this situation. Key pictures are used to limit this drift by only using other key pictures for prediction and are thus reconstructable without drift.

In the enhancement layer, the CU tree from the base layer is inherited and no CU tree needs to be coded. For every CU in the enhancement layer, conventional HEVC inter prediction from enhancement layer picture references as well as conventional intra prediction from neighboring enhancement layer pixels can be performed. A new mode further allows for utilization of the lower layer information. In this mode, three cases have to be considered:

- If the base layer has no residual signal, the base layer prediction signal is used and a conventionally coded residual signal is added in the enhancement layer.
- A skip mode can be signaled that copies the base layer prediction and residual signal and adds no residual in the enhancement layer.
- A refinement mode can be signaled that copies the base layer prediction signal and refines the base layer residual. This residual refinement process is explained in the next section.

4. RESIDUAL REFINEMENT

In previous work on Fine Grain Scalability (FGS) various methods for refining existing residual data in the transform domain were presented, see [7] [8]. Here, a binary mapping scheme is presented that allows for refinement of the quantizer step size in configurable steps while preserving the option of rewritability to a single-layer residual.

4.1. Quantizer Mapping

In residual refinement mode, the transform tree as well as the prediction signal is inherited from the base layer. Thereby, the base layer and the enhancement layer perform the same transform on the identical residual signal. This results in equal transform coefficients in the two layers. The only difference between the two layers is now that they use different quantization parameters (QP) for the transform coefficients.

In Fig. 3 the quantization used in HEVC for the QP values 36 and 30 ($\Delta\mathrm{QP} = 6$) is shown. The bold vertical lines denote the decision thresholds while the reconstruction values are marked with a cross. As we can see for $\Delta\mathrm{QP} = 6$, the number of reconstruction values is exactly doubled for the enhancement layer quantizer. For the new residual refinement

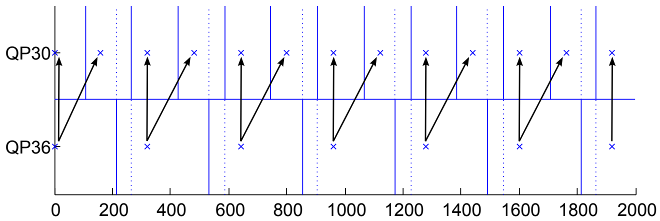


Fig. 3: Example for mapping reconstruction values from QP $= 36$ in base layer to QP $= 30$ in the enhancement layer with shift of quantization thresholds.

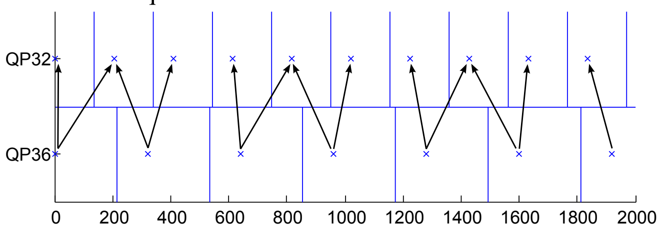


Fig. 4: Possible mapping from base (QP $= 36$) to enhancement (QP $= 32$) layer reconstruction values.

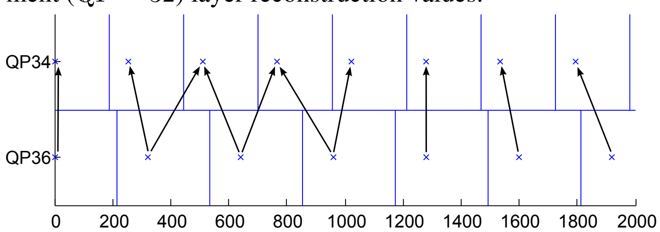


Fig. 5: Possible mapping from base (QP $= 36$) to enhancement (QP $= 34$) layer reconstruction values.

coding, only a binary mapping (represented by the arrows in Fig. 3) from base layer reconstruction values to enhancement layer reconstruction values is coded.

The mapping has an impact on the quantization process and by that on the RD performance. When the initial quantization is done in the base layer, the refinement step, in which they are mapped to the enhancement layer reconstruction values, can be interpreted as a quantization in the enhancement layer with shifted quantization thresholds. These are indicated by dotted vertical lines in Fig. 3. For example the value 1200 in Fig. 3 is quantized to level 4 in the base- and to level 7 in the enhancement layer. However, we only allow a mapping from base layer level 4 to enhancement layer levels 8 and 9. So effectively, the quantization threshold in the enhancement layer is shifted. This may lower the quantization performance in the enhancement layer. Likewise, if the initial quantization is performed in the enhancement layer, the quantization thresholds in the base layer are shifted and the base layer quantization performance might be reduced.

The mapping can be freely configured and is not limited to specific ΔQP values. Example mappings for $\Delta\text{QP} = 4$ and $\Delta\text{QP} = 2$ are shown in Figures 4 and 5. As can be seen from Fig. 4, some reconstruction values in the enhancement layer can be reached from two values in the base layer. Furthermore, reconstruction values in the base layer can map to exactly one reconstrucion value in the enhancement layer so that coding of a refinement flag can be omitted. (See Fig. 5). If the QP difference is greater than 6, the number of reconstruction values in the enhancement layer is more than doubled compared to the the base layer quantizer. In this case, a manifold mapping, which requires more than a single flag for the mapping, might be desirable. As an alternative approach, a stacked mapping can be used, where a binary mapping to an intermediate quantizer is performed before performing a second binary mapping to the final reconstruction values.

4.2. Refinement Coding

For every enhancement layer CU, the mode (Intra, Inter or Base Layer) is coded. For Intra and Inter, the conventional HEVC coding is performed. When residual refinement coding is used, the TU tree is inherited from the base layer and residual coding starts at the transform coefficient level. The coefficient coding for the refinement has been changed in the following way.

At first, the position of the last significant coefficient is coded. Since the position in the base layer is known, only the position difference needs to be coded. In the subsequent significance scan, the significant coefficients from the base layer are skipped and only coefficients that were not significant in the base layer are scanned. Thus, the sign information only needs to be coded for the coefficients that got significant in the enhancement layer significance scan. In the final step, all coefficients that are significant in the base layer are refined using a binary flag that codes the respective mapping decision between the base- and enhancement layer quantizers. This flag is encoded in one CABAC context which is initialized with the initial probability set to equal probability.

5. RESULTS

In order to evaluate the performance of the design, similar conditions to the CfP on SHVC were chosen [9]. The SMuC software version 0.1.1 is used for comparison to the multi-loop approach. The proposed single-loop method has been implemented based on the software of the HEVC test model (HM) version 6.1 [10]. For generating single layer and simulcast anchors, the HM software in version 6.1 is used. Due to limitiations in the current implementation of the presented method, some changes to the coding conditions had to be applied in order to allow for a fair comparison. For the current implementation of the proposal as well as for both references, RDOQ, AMP and Sign Data Hiding are disabled. Following the configuration setting of the CfP, the 1080p sequences *Kimono*, *ParkScene*, *Cactus*, *BasketballDrive* and *BQTerrace* were tested with the base layer QP values of 26, 30, 34, 38 and enhancement layer values of

Table 1: Average luminance BD-Rate savings over simulcast per ΔQP of the SMuC reference and the proposed codec in Random Access configuration.

	SMuC	**Proposed Codec**
$\Delta\mathrm{QP} = 2$	-30.68%	-33.12%
$\Delta\mathrm{QP} = 4$	-20.95%	-14.21%
$\Delta\mathrm{QP} = 6$	-14.14%	-6.10%
Average	-21.92%	-17.81%

Table 2: Average luminance BD-Rate savings over simulcast per ΔQP of the SMuC reference and the proposed codec in All Intra configuration.

	SMuC	**Proposed Codec**
$\Delta\mathrm{QP} = 2$	-38.88%	-41.70%
$\Delta\mathrm{QP} = 4$	-30.52%	-28.18%
$\Delta\mathrm{QP} = 6$	-23.97%	-21.39%
Average	-31.13%	-30.42%

$\mathrm{QP}_{\mathrm{enh}} = \{\mathrm{QP}_{\mathrm{base}} - 2, \mathrm{QP}_{\mathrm{base}} - 4, \mathrm{QP}_{\mathrm{base}} - 6\}$. The results were then averaged over the sequences for each of the three ΔQP values and over all sequences at all ΔQP values for the overall average results.

Please note that in the current software implementation, the mapping function is not yet optimized for different ΔQP values. Also, the current implementation optimizes the CU and TU tree as well as the mode decisions for the base layer. This diminishes the performance of the enhancement layer and could be improved by an inter-layer optimization.

In Table 1, the average luminance BD-Rate savings of the SMuC reference and the proposed codec compared to simulcast are shown for the CfP Random Access configuration [11]. On average, the proposed approach can gain 17.81% of BD-Rate over simulcast where the SMuC approach gains 21.92%. Table 2 shows the results for the CfP All Intra configuration where the difference in BD-Rate savings over simulcast is smaller than for the Random Access configuration (SMuC 31.13%, Proposed codec 30.42%). This can be explained by the fact that in the All Intra case more transform coefficients are encoded. Furthermore, it can be observed that the proposed scheme performs better for small QP differences where it is even able to surpass the SMuC reference.

In Fig. 6 the RD plot for the sequence *Kimono* with $\Delta\mathrm{QP} = 6$ is shown. The leftmost curve shows the base layer of SMuC as well as the single layer reference. Both show almost identical results and overlap in the plot. The base layer of the proposed scheme shows some loss due to the drift that occurs when only decoding the base layer. Furthermore, it can be observed that the single-loop approach shows a reasonably small performance loss compared to SMuC. At the

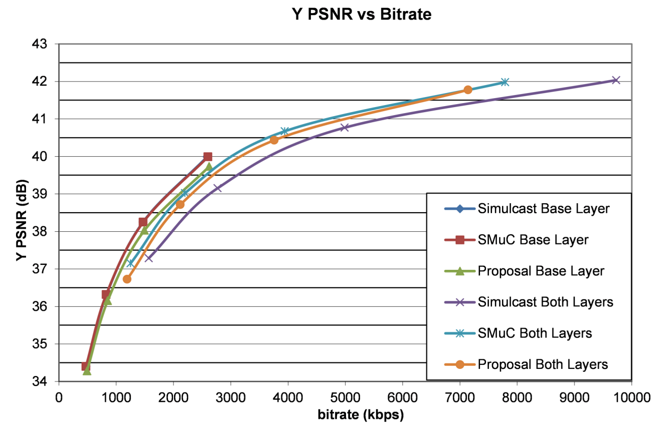


Fig. 6: Y-PSNR versus bitrate for the sequence *Kimono* and $\Delta\mathrm{QP} = 6$.

Table 3: Average enhancement layer decoding operations compared to the SMuC multi loop approach in random access configuration.

Intra prediction	79.56%
Inter prediction	57.20%
Inverse transform	75.39%
Deblocking	35.43%
Sample adaptive offset	88.07%

same time Fig. 3 shows a decrease in required decoding operations for the enhancement layer compared to the multi-loop approach. Especially the inter prediction operations, inverse transform operations and filtering operations are significantly reduced [12].

6. CONCLUSIONS

In this paper, a coefficient refinement coding method for single-loop SNR scalability is presented. The scheme uses binary mapping between different quantizer step sizes. Simulation results reveal promising RD performance in the context of the emerging scalable extension of HEVC. While not fully reaching the RD performance observed by the multi-loop SHVC test model, the scheme reduces the computational complexity of the decoder as multiple motion compensation loops, reference picture buffers and loop-filter operations are omitted.

The presented results suggest that single-loop SNR scalable decoding can be achieved at a reasonable RD trade-off compared to the increased complexity of the dual-loop approach investigated so far. In future work it is planned to extend the scheme by RDOQ, Sign Data Hiding and AMP. Furthermore an inter layer optimization of the shared CU- and TU trees, as well as the mode decision could further improve the overall RD performance.

7. REFERENCES

[1] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 9," Tech. Rep. JCTVC-K1003, JCT-VC, Shanghai, PRC, Oct. 2012.

[2] A. Segall, J. Boyce, J.-R. Ohm, and G. J. Sullivan, "Report of results of the joint call for proposals on scalable high efficiency video coding (SHVC)," Tech. Rep. JCTVC-K1008, JCT-VC, Shanghai, PRC, Oct. 2012.

[3] C. Feldmann, F. Jäger, J. Hsu, and M. Wien, "Single-loop SNR scalability using binary residual refinement coding," Tech. Rep. JCTVC-L0154, JCT-VC, Geneva, CH, Jan. 2013.

[4] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 620 – 636, July 2003.

[5] J. Sole, R. Joshi, N. Nguyen, T. Ji, M. Karczewicz, G. Clare, F. Henry, and A. Duenas, "Transform coefficient coding in HEVC," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1765 –1777, Dec. 2012.

[6] 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.

[7] I. Amonou, N. Cammas, S. Kervadec, and S. Pateux, "Improved SNR scalability for J-SVM," Tech. Rep. JVT-O060, JVT, Busan, KR, Apr. 2005.

[8] H. Radha, Y. Chen, K. Parthasarathy, and R. Cohen, "Scalable internet video using MPEG-4," *Signal Processing: Image Communication*, vol. 15, no. 12, pp. 95 – 126, 1999.

[9] 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.

[10] K. McCann, B. Bross, I.-K. Kim, K. Sugimoto, and W.-J. Han, "HM6: High efficiency video coding (HEVC) test model 6 encoder description," Tech. Rep. JCTVC-H1002, JCT-VC, San José CA, USA, Feb. 2012.

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

[12] Christian Feldmann, Fabian Jäger, and Mathias Wien, "[AHG16] analysis of single-loop SNR scalability using binary residual refinement coding," Tech. Rep. JCTVC-M0176, 13th Meeting, Incheon, Republic of Korea, Apr. 2013.