



HAL
open science

Wavelet Decomposition Pre-processing for Spatial Scalability Video Compression Scheme

Glenn Herrou, W. Hamidouche, Luce Morin

► **To cite this version:**

Glenn Herrou, W. Hamidouche, Luce Morin. Wavelet Decomposition Pre-processing for Spatial Scalability Video Compression Scheme. 2018 Picture Coding Symposium (PCS), Jun 2018, San Francisco, United States. 10.1109/PCS.2018.8456307 . hal-01874660

HAL Id: hal-01874660

<https://hal.science/hal-01874660>

Submitted on 1 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Wavelet Decomposition Pre-processing for Spatial Scalability Video Compression Scheme

Glenn Herrou^{*}, Wassim Hamidouche^{*,†} and Luce Morin^{*,†}

^{*}IRT b<>com, Cesson-Sevigne, France

[†]IETR/INSA Rennes, Rennes, France

Email: glenn.herrou@b-com.com, {whamidou,lmorin}@insa-rennes.fr

Abstract—Scalable video coding enables to compress the video at different formats within a single layered bitstream. SHVC, the scalable extension of the High Efficiency Video Coding (HEVC) standard, enables x2 spatial scalability, among other additional features. The closed-loop architecture of the SHVC codec is based on the use of multiple instances of the HEVC codec to encode the video layers, which considerably increases the encoding complexity. With the arrival of new immersive video formats, like 4K, 8K, High Frame Rate (HFR) and 360° videos, the quantity of data to compress is exploding, making the use of high-complexity coding algorithms unsuitable. In this paper, we propose a low-complexity scalable coding scheme based on the use of a single HEVC codec instance and a wavelet-based decomposition as preprocessing. The pre-encoding image decomposition relies on well-known simple Discrete Wavelet Transform (DWT) kernels, such as Haar or Le Gall 5/3. Compared to SHVC, the proposed architecture achieves a similar rate distortion performance with a coding complexity reduction of 50%.

I. INTRODUCTION

In the past few years, several new video formats, such as 4K-8K, HDR, HFR and 360° video, have emerged in the video industry motivated by the will to provide more realistic and immersive experiences to consumers. These new formats present the characteristic of requiring a very large amount of data to be encoded before transmission to the end-user. In addition, the heterogeneity of the user's requirements, in terms of available bandwidth, display, computing and energy capabilities, drastically increases the storage and bandwidth resources required to encode and deliver the video. High Efficiency Video Coding (HEVC) [1] and its scalable extension Scalable High efficiency Video Coding (SHVC) [2] have been developed by the Joint Collaborative Team on Video Coding (JCT-VC) and released in January 2013 and October 2014, respectively. SHVC enables to encode the video in several formats, depending on the chosen type of scalability (spatial, quality, bit depth, color-gamut or codec), within a single layered bitstream. However, these state-of-the-art codecs have been designed and optimized for the encoding of HD and UHD videos in SDR format. With the new immersive video formats and their increased resolutions, SHVC becomes unsuitable due to its high complexity coding algorithm [3], [4]. Indeed, it relies on the use of one HEVC encoder instance per layer and inter-layer predictions to produce a scalable bitstream.

As part of the Joint Video Exploration Team (JVET) effort, E. Thomas et al. proposed a new scalable scheme [5], based on a polyphase sub-sampling performed prior to encoding,

achieving x2 spatial scalability with a single HEVC encoder instance, thus greatly reducing the coding complexity compared to SHVC. In this paper, we propose several improvements to the polyphase sub-sampling pre-processing technique. The first consists in the correction of the phase difference between the chroma planes of the resulting sub-resolution images by introducing a simple chroma filtering process. In order to improve the visual quality of the output video layers and avoid the potential aliasing introduced by the polyphase sub-sampling, we also propose a different decomposition step based on the use of well-known wavelet kernels modified to fit in the scalable coding chain. The proposed solution achieves similar rate distortion performance as SHVC with a 50% coding complexity reduction.

The rest of this paper is organized as follows. Section II provides a brief overview of the scalable extension of HEVC, as well as the related work on image decomposition and wavelet-based image coding. The proposed low-complexity scalable method is detailed in Section III. Section IV presents a performance evaluation and provides an analysis of the results. Finally, Section V concludes this paper.

II. RELATED WORK

A. SHVC extension

The SHVC extension defines high level syntax elements mostly at the level of Video Parameter Set (VPS) header. These syntax elements provide information on the video layers such as the number of layers, and for each layer, resolution, bit depth as well as the inter-layer dependencies. The SHVC encoder architecture consists of L HEVC encoders, one to encode each layer with L the number of layers: one Base Layer (BL) and $L - 1$ Enhancement Layers (EL).

In the case of SHVC spatial scalability, the BL HEVC encoder encodes a down-sampled version of the original video and feeds the first EL encoder with the decoded picture and its Motion Vectors (MVs). The enhancement layer encoder l ($l = 2, \dots, L$) encodes a higher resolution video with the decoded picture from lower layer used as an inter-layer reference picture (included in the reference picture lists). The inter-layer reference picture is up-sampled and its MVs up-scaled to match with the resolution of the layer being decoded. The up-sampling operation is performed by a 8-tap interpolation filter for luma samples and 4-tap interpolation filter for chroma samples. The outputs from the BL and EL

encoders are multiplexed to form one bitstream that conforms to SHVC. The SHVC encoder architecture improves the rate distortion performance by up to 30% under the Common Test Conditions (CTC) [6] compared to the independent coding of each layer (simulcast) with an HEVC encoder.

The HEVC standard version 2 defines two SHVC profiles: Scalable Main and Scalable Main 10 [2]. The Scalable Main enables a base layer that conforms with the Main HEVC profile, while the Scalable Main 10 profile allows a base layer that conforms with the Main 10 HEVC profile. The 4th HEVC version defines four more scalable profiles for base layer in monochrome format with 8, 12 and 16 bit depth (Scalable Monochrome, Scalable Monochrome 12, Scalable Monochrome 16) and one Scalable Main 4:4:4 profile that conforms to the Main 4:4:4 HEVC profile.

B. Polyphase Sub-sampling for Spatial Scalability

As part of the JVET activities to develop the future video coding standard, E. Thomas et al. proposed a new scalable coding scheme relying on the polyphase sub-sampling of the input video prior to encoding [5]. This new architecture enables x2 spatial scalability with a single HEVC encoder instance. The input video is first decomposed into four sub-resolution components, each component having one over four pixels of the original signal. The resulting sub-resolution images are then packed sequentially to form a new video at a quarter of the original spatial resolution but with four times the number of frames. The polyphase sub-sampled video is then fed to an HEVC encoder. Spatial scalability is achieved at the decoder side by choosing to decode only one resolution component or the entire bitstream. If the four resolution components are decoded, an additional reconstruction step is performed to produce the full resolution decoded video.

In [7] and [8], authors evaluate the polyphase scalable coding chain in terms of rate distortion compared to simulcast coding with the Joint Exploration Test Model (JEM) software [9]. Results showed highly variable performance of the polyphase sub-sampling technique, depending on the test video sequence content and resolution. However, several questions have been raised in [8] concerning the test procedure, especially on the Group Of Pictures (GOP) structure, intra period alignment and aliasing introduced by the sub-sampling. The quality of the sub-resolution images has been investigated in [10], showing that aliasing is strongly perceptible for resolutions up to HD, while it is hardly visible for higher resolutions. Thus, the polyphase sub-sampling technique is an interesting solution for the coding of the new immersive video formats and their resolutions of 4K and greater.

C. Wavelet-based scalable image coding

The wavelet transform representation [11] allows for the decomposition of an input image into sub-bands, one representing a low-pass filtered version of the input and the other containing the details along different spatial orientations. In [12], Le Gall et al. design a short kernel symmetric analysis/synthesis filter bank. This wavelet kernel provides

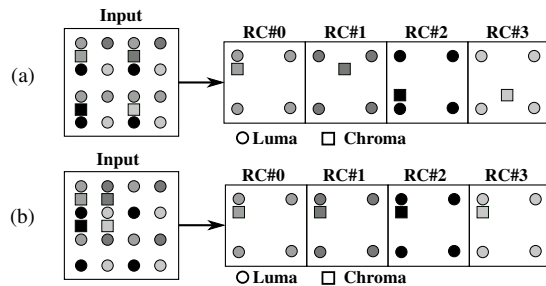


Fig. 1: Pixel positions in 4:2:0 format for (a) original polyphase sub-sampling and (b) polyphase sub-sampling with chroma pixel alignment.

an integer-to-integer mapping with the advantage of being computationally efficient (only a limited number of shifts and additions) while offering a perfect reconstruction of the input signal and a visually pleasant low-pass sub-band.

In [13], Sweldens et al. introduce the lifting scheme, a method to construct a bi-orthogonal wavelet with a fast implementation. The lifted wavelet construction relies on a simple relationship: the high pass sub-band is obtained by applying a high-pass filter on the odd samples whereas the low-pass sub-band is a linear combination of the filtered even sampled and the scaled high-pass sub-band. The lifting implementation of the 1D Le Gall 5/3 wavelet is presented in (1) and (2) where n is the sample index, x the input signal and $\lfloor \cdot \rfloor$ the floor rounding operator. y is the output signal with its even samples representing the low-pass sub-band and its odd samples the high-pass sub-band.

$$y(2n+1) = x(2n+1) - \left\lfloor \frac{x(2n) + x(2n+2)}{2} \right\rfloor \quad (1)$$

$$y(2n) = x(2n) + \left\lfloor \frac{y(2n-1) + y(2n+1) + 2}{4} \right\rfloor \quad (2)$$

For image and video compression, wavelet transforms offer a native x2 spatial scalability. Indeed, the full resolution output image can be obtained by applying the inverse wavelet transform on the four sub-bands, while the sub-resolution output can be obtained by directly taking the low-pass sub-band which corresponds to a low-pass filtered and down-sampled version of the input image. Lifted wavelet transforms have notably been used in the lossless profile of the JPEG2000 still image coding standard [14].

III. PROPOSED SCALABLE SOLUTIONS

The polyphase sub-sampling technique presents several issues: a chroma pixel misalignment and the potential aliasing introduced by the sub-sampling process. This section describes the different solutions proposed to address each of these issues.

A. Modified Polyphase Sub-sampling

The polyphase sub-sampling technique takes one over four pixels, in each 2x2 block of the input image, to create the four resolution components. This decomposition, as described in [5], is performed on both the luma and chroma planes of the

input images. However, for a video in 4:2:0 format, the most commonly used chroma sub-sampling format, the polyphase decomposition introduces a misalignment between the chroma pixels of the different sub-resolution images, as depicted in Figure 1.a. During the motion compensation step, an encoder usually derives the MVs of the chroma planes from the MVs of the luma plane. For example, in HEVC, the luma MVs are scaled by a factor of two in both directions to compute the chroma MVs. For a polyphase sub-sampled video, when a frame from a different resolution component is used as reference for an inter-prediction, the derived chroma motion vector is inherently wrong due to the misalignment of the chroma pixels between resolution components. The resulting chroma plane prediction is sub-optimal thus increasing the amount of residuals to encode. To overcome this issue we can either change the chroma motion vector derivation process depending on the resolution components or realign the chroma planes before the polyphase sub-sampling. To avoid changing the core encoding process and keep the solution entirely as a pre-processing step, we propose to realign the chroma planes of the different resolution components, as shown in Figure 1.b, by using a simple mean filter before the polyphase sub-sampling. The mean filter is applied on each 2x2 chroma block of the input full resolution image, as described in (3). The inverse operation is performed after the reconstruction of the output full resolution image from the four decoded resolution components.

$$\begin{cases} I_{align}(2n, 2m) &= I_{src}(2n, 2m) \\ I_{align}(2n+1, 2m) &= \frac{1}{2} \cdot (I_{src}(2n, 2m) \\ &\quad + I_{src}(2n+1, 2m)) \\ I_{align}(2n, 2m+1) &= \frac{1}{2} \cdot (I_{src}(2n, 2m) \\ &\quad + I_{src}(2n, 2m+1)) \\ I_{align}(2n+1, 2m+1) &= \frac{1}{2} \cdot (I_{src}(2n, 2m) \\ &\quad + I_{src}(2n+1, 2m+1)) \end{cases} \quad (3)$$

with I_{src} the input picture, I_{align} the output chroma aligned picture, n and m the row and column indexes, respectively.

B. Wavelet-based Decomposition

The main flaw of the polyphase sub-sampling technique is the potential aliasing introduced in the four sub-resolution images obtained after decomposition. Indeed, aliasing creates high frequencies which are costly to encode, for usual video coding algorithms, due to the magnitude and position of the corresponding coefficients in the transformed residual blocks. This is especially a problem for the resolution component which serves as base layer, since it will be encoded with a lower Quantization Parameter (QP) than the sub-images of the enhancement layer, if an HEVC encoding with a traditional hierarchical GOP is considered. To avoid this issue, we propose to use a wavelet-based decomposition instead of the polyphase sub-sampling.

The proposed wavelet-based decomposition process is depicted in Figure 2. First, a Discrete Wavelet Transform (DWT)

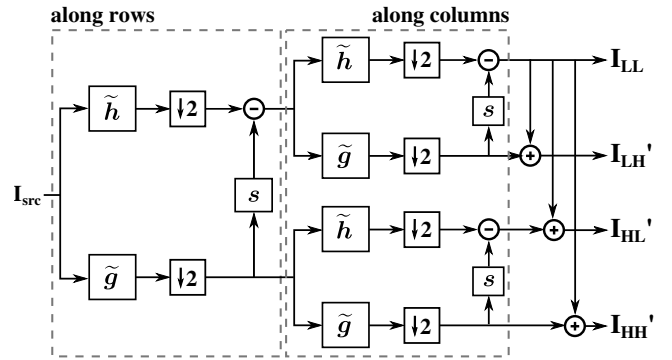


Fig. 2: Modified wavelet lifting scheme.

is applied on the input signal to produce four different sub-bands. The LL sub-band, corresponding to a lowpass filtered version of the input signal, is used as base layer in the scalable coding chain. The LH , HL and HH sub-bands, which correspond to horizontal, vertical and diagonal high-frequencies, respectively, form the enhancement layer. To make the decomposed signal suitable for a standard HEVC encoder, we add the LL sub-band to each of the other three high-frequency sub-bands as shown in Figure 2, thus enabling temporal prediction between the base and enhancement layers. Indeed, the residuals of a temporal prediction between a LL block and a LH' block with a (near-)zero motion vector would be equivalent to the corresponding original LH block. The temporal packing of the sub-bands follows the same process as for the polyphase sub-sampling, each sub-band corresponding to a different resolution component.

In our experiments, we chose the Haar or Le Gall 5/3 wavelets [12] for their low-complexity kernels and integer-to-integer lifting implementation with perfect reconstruction. Since the Le Gall 5/3 wavelet is bi-orthogonal, energy is not preserved throughout the transform process resulting in a sub-optimal bit-allocation with usual Rate-Distortion Optimization (RDO) algorithms such as in the HEVC reference software. Usevitch proposed in [15] an optimal bit-allocation for bi-orthogonal wavelet based on a weighting of the quantization error depending on the sub-band filter coefficients. We transposed this weighting into the RDO stage of the HEVC reference software using the following cost function:

$$J = D + \frac{\lambda}{\omega_{SB}} \cdot R \quad (4)$$

with J the cost function to minimize, D the distortion, R the rate in number of bits required to signal the coding parameters, λ the Lagrangian multiplier derived from the prediction mode and QP, ω_{SB} the weight applied to sub-band SB .

IV. PERFORMANCE EVALUATION

A. Test Methodology

The coding chain under evaluation is shown in Figure 3. The decomposition is first applied on the input video clips before encoding with a legacy HEVC encoder. Then, the obtained bitstreams are decoded and reconstructed to obtain the output

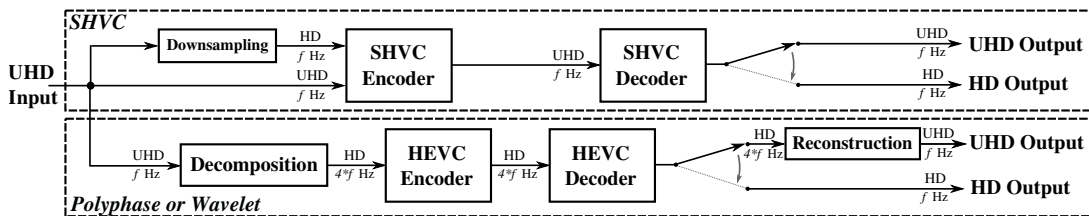


Fig. 3: Scalable coding chains under evaluation.

TABLE I: BD-Rate results for chroma aligned polyphase compared to original scheme.

| Sequence | PSNR-Y | PSNR-U | PSNR-V | PSNR-YUV |
|-----------------|-------------|--------------|--------------|-------------|
| DaylightRoad | -0.16 | -12.5 | -11.2 | -2.96 |
| CatRobot | -3.55 | -28.0 | -31.8 | -8.98 |
| Drums | -1.94 | -15.0 | -10.9 | -3.51 |
| Tango | -0.10 | -7.87 | -14.9 | -3.66 |
| RollerCoaster | -0.52 | -14.5 | -7.64 | -2.64 |
| ToddlerFountain | -0.03 | -4.47 | -4.39 | -0.11 |
| Average | -1.1 | -13.7 | -13.5 | -3.6 |

video which will be compared to a SHVC encoded video in terms of rate-distortion performance. The low-complexity scalable coding chain has been evaluated with the two proposed decompositions using a dataset composed of six UHD test sequences submitted to JVET for the evaluation of their future video coding standard.

Encodings have been performed with the HEVC reference software (HM16.12) [16] and SHVC reference software (SHM9.0) [17]. For the Le Gall 5/3 decomposition, the RDO stage of the HM software has been modified using the Usevitch weighting described in (4). Random Access (RA) with fixed QP configuration has been used for all encodings, with a hierarchical GOP of size 16 and an intra-period of approximately one second for both the reference SHVC coding chain and the proposed one. The performance has been assessed in terms of rate-distortion using the Bjøntegaard delta metric (BD-Rate) [18], representing the average bit-rate savings for equal Peak Signal-to-Noise Ratio (PSNR) values. In order to evaluate the proposed scalable coding chain over a wide range of bit-rates, QPs ranging from 22 to 40 with a step of two have been used for each test sequence.

B. Results and Analysis

Table I shows the results of the realignment of chroma pixels of the polyphase sub-sampled signal. Negative BD-Rate values represent bit-rate savings for the proposed decomposition compared to the original polyphase sub-sampling. We can observe that, the proposed chroma realignment improves the performance for both chroma channels (U and V) with an average gain of 13.6% compared to the original polyphase sub-sampling. Thus, the losses due the rounding error introduced by the proposed filtering are more than compensated by the improved prediction step due to the correct chroma motion vector derivation. Table I also shows minor gains for the luma channel, which can be explained by the overall decrease in terms of number of bits necessary to achieve the same quality.

The proposed wavelet-based decomposition, with Haar and Le Gall 5/3 kernels, and the polyphase sub-sampling with chroma realignment are compared to SHVC in Table II. For equal PSNR-YUV, an average bit-rate overhead of 6.6%, 1.9% and 7.1% can be observed for polyphase with chroma realignment, Haar and Le Gall 5/3, respectively. For the polyphase scheme, we can observe that the per-sequence results are highly variable, ranging from a 45% overhead to a 13.3% bit-rate reduction. For most test sequences, the temporal prediction between the enhancement and base layers allows for the recovery of the spatial details of the full resolution input video, resulting in a bit-rate reduction compared to SHVC. For the *DaylightRoad* and *CatRobot* test sequences, the large losses can be explained by a noisy input, especially for *Daylightroad*, the presence of very sharp details, mostly text. The spatial high frequencies of these sharp details are partially or totally lost in the base layer during the decomposition step and are thus too costly to recover due to the high residual energy in the enhancement layer frames. The same behaviour is observed in the chroma channels for all sequences, where the smaller resolution of the chroma planes amplifies this effect, which results in high positive chroma BD-Rate values.

The results for both wavelet-based decompositions follow the same trend as the polyphase sub-sampling but with less bit-rate overhead. On average, it can be observed that the scheme based on the Haar wavelet performs better than the Le Gall 5/3 scheme. This is mainly due to the smaller energy present in the transformed coefficients in the high-pass sub-bands after the motion compensation stage. Figure 4 depicts the average energy in the 16x16 Transform Units (TU) obtained by encoding the sequence *CatRobot* with a QP of 22. It is shown that the Le Gall 5/3 scheme transformed residuals are less compact in the Discrete Cosine Transform (DCT) domain especially in the LH and HL sub-bands, thus greatly increasing the number of bits required to signal such transform blocks after the Rate-Distortion Optimized Quantization (RDOQ) stage [19]. For the HH sub-band, the residual energy is mainly contained in the DC coefficient for both wavelet schemes but is much more diffused in the spatial frequency coefficients for the Le Gall 5/3 scheme. Thus, the number of transmitted TUs is globally higher for the Haar wavelet scheme due to their limited bit cost, resulting in a higher reconstruction quality for the same bit-rate.

Encoding and decoding times are also summarized in Table II, showing an average encoding (resp. decoding) complexity reduction of 55%, 54% and 53% (resp. 42%, 50%

TABLE II: BD-Rate results (%) and complexity (%) for proposed schemes vs SHVC.

| Sequence | Polyphase with Chroma Realignment | | | | Haar | | | | Le Gall 5/3 | | | |
|----------------------|-----------------------------------|-------------|-------------|------------|-------------|-------------|-------------|------------|-------------|-------------|-------------|------------|
| | Y | U | V | YUV | Y | U | V | YUV | Y | U | V | YUV |
| DaylightRoad | 33.4 | 104 | 142 | 44.9 | 22.9 | 45.5 | 71.2 | 30.0 | 36.3 | 67.3 | 92.4 | 44.0 |
| CatRobot | 19.6 | 80.4 | 98.1 | 28.5 | 7.5 | 39.2 | 55.0 | 13.4 | 16.6 | 53.4 | 71.6 | 23.3 |
| Drums | -12.1 | 46.1 | 69.4 | -4.6 | -2.2 | 45.9 | 67.8 | 4.3 | 2.4 | 60.9 | 83.2 | 9.7 |
| Tango | -17.3 | 31.0 | 31.4 | -6.5 | -24.8 | 12.1 | 10.5 | -16.0 | -23.4 | 16.1 | 16.3 | -13.7 |
| RollerCoaster | -17.4 | 11.3 | 15.0 | -13.3 | -12.8 | 13.9 | 17.8 | -8.8 | -15.6 | 10.2 | 13.6 | -11.9 |
| ToddlerFountain | -11.8 | 50.9 | 26.9 | -9.5 | -12.6 | 12.9 | -4.1 | -11.5 | -11.4 | 56.4 | 25.9 | -8.7 |
| Average | -0.9 | 53.9 | 63.8 | 6.6 | -3.7 | 28.3 | 36.4 | 1.9 | 0.8 | 44.1 | 50.5 | 7.1 |
| Encoding time | 45 | | | | 46 | | | | 47 | | | |
| Decoding time | 58 | | | | 50 | | | | 60 | | | |

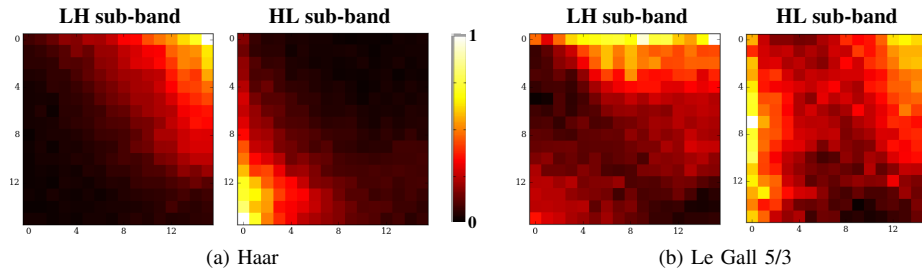


Fig. 4: Normalized average transformed coefficients of 16x16 HEVC TUs for (a) Haar and (b) Le Gall wavelet-based schemes.

and 40%) for polyphase with chroma realignment, Haar and Le Gall 5/3 schemes, respectively, compared to SHVC.

V. CONCLUSION

In this paper we have proposed an improvement to the polyphase sub-sampling scalable coding chain and a new low-complexity wavelet-based scalable coding scheme using HEVC for 4K and greater videos. The modification to the polyphase sub-sampling technique is based on a chroma pixels re-alignment between the base and enhancement layers to obtain a correct chroma motion vector derivation. The experimental study shows a BD-Rate gain of 3.6% for the proposed scheme compared to the original polyphase sub-sampling technique. The proposed wavelet-based decomposition relies on a modified wavelet lifting scheme to enable x2 spatial scalability using a single HEVC encoder instance. The proposed scalable coding chain shows an average bit-rate overhead of 6.6%, 1.9% and 7.1% for the polyphase with chroma realignment, Haar and Le Gall 5/3 decompositions, respectively, compared to SHVC together with an overall encoding complexity reduction of approximately 50%.

Future work will focus on improving the prediction between the base and enhancement layers using a motion compensation filter optimized for the proposed decompositions. We will also extend the proposed scalable coding chain to other new immersive video formats such as 360° sequences.

REFERENCES

- [1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 22-12, pp. 1649–1668, 2012.
- [2] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramanian, "Overview of SHVC: scalable extensions of the High Efficiency Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, 2016.
- [3] W. Hamidouche, M. Raulet, and O. Deforges, "4k real-time and parallel software video decoder for multilayer HEVC extensions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 169–180, Jan. 2016.
- [4] R. Paroiss, W. Hamidouche, E.-G. Mora, and M. R. ad O. Deforges, "Efficient parallel architecture for a real-time UHD Scalable HEVC Encoder."
- [5] E. Thomas, "Polyphase subsampled signal for spatial scalability," *Doc. JVET-B0043*, San Diego, February 2016.
- [6] V. Seregin and Y. He, "Common SHM test conditions and software reference configurations," *Doc. JCTVC-Q1009*, Valencia, April 2014.
- [7] E. Thomas, "Experiments on polyphase subsampled sequence coding," *Doc. JVET-C0032*, Geneva, May 2016.
- [8] P. Philippe, T. Biatek, and V. Lorcy, "Cross-check of C0032 (polyphase subsampled sequence coding)," *Doc. JVET-C0078*, Geneva, May 2016.
- [9] J. Chen, E. Alshina, G. J. Sullivan, J.-R. Ohm, and J. Boyce, "Algorithm description of Joint Exploration test Model 1," *Doc. JVET-A1001*, Geneva, October 2015.
- [10] E. Thomas, "Subjective quality analysis of polyphase subsampled sequences," *Doc. JVET-D0034*, Chengdu, October 2016.
- [11] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [12] D. Le Gall and A. Tabatabai, "Sub-band coding of digital images using symmetric short kernel filters and arithmetic coding techniques," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP, International Conference on*. IEEE, 1988, pp. 761–764.
- [13] W. Sweldens, "The lifting scheme: a custom-design construction of biorthogonal wavelets," *Applied and computational harmonic analysis*, vol. 3, pp. 186–200, 1996.
- [14] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal processing magazine*, vol. 18, no. 5, pp. 36–58, 2001.
- [15] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Data Compression Conference, 1996. DCC'96. Proceedings*. IEEE, 1996, pp. 387–395.
- [16] "HEVC reference software version 16.12." [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.12/
- [17] "SHVC reference software version 9.0." [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSoftware/tags/SHM-9.0/
- [18] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," *VCEG-M33 ITU-T Q6/16, Austin, TX, USA, 2-4 April*, 2001.
- [19] M. Karczewicz, Y. Ye, and I. Chong, "Rate distortion optimized quantization," *Doc. VCEG-AH21*, Antalya, January 2008.