



**HAL**  
open science

## Low-complexity Scalable Encoder based on Local Adaptation of the Spatial Resolution

Glenn Herrou, Wassim Hamidouche, Luce Morin

► **To cite this version:**

Glenn Herrou, Wassim Hamidouche, Luce Morin. Low-complexity Scalable Encoder based on Local Adaptation of the Spatial Resolution. IEEE International Conference on Image Processing (ICIP 2019), Sep 2019, Taipei, Taiwan. hal-02118078

**HAL Id: hal-02118078**

**<https://hal.science/hal-02118078v1>**

Submitted on 2 May 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.

# LOW-COMPLEXITY SCALABLE ENCODER BASED ON LOCAL ADAPTATION OF THE SPATIAL RESOLUTION

Glenn Herrou<sup>\*†</sup>, Wassim Hamidouche<sup>\*†</sup> and Luce Morin<sup>\*†</sup>

<sup>\*</sup>IRT b<>com, 35510 Cesson-Sevigne, France

<sup>†</sup>Univ Rennes, INSA Rennes, CNRS, IETR - UMR 6164, 35000 Rennes, France

## ABSTRACT

A two-layer low-complexity scalable encoding scheme based on local Adaptive Spatial Resolution (ASR) is proposed. This scheme relies on a block-level spatial resolution adaptation in the enhancement layer encoder. For each block, the optimal resolution is either obtained via a rate-distortion optimization followed by a decision refinement process or by a prediction via motion compensation exploiting the base layer motion vectors. The proposed architecture has been integrated over of the High Efficiency Video Coding (HEVC) reference software (HM16.12) which is used as a base layer encoder. Compared to SHVC, the scalable extension of HEVC, experimental results show bitrate savings of 0.76 % as well as encoding complexity reductions of 47 % for the whole scalable encoder and 96 % for the enhancement layer encoder.

*Index Terms*— Video compression, spatial scalability, low-complexity, adaptive spatial resolution, SHVC.

## 1. INTRODUCTION

The deployment of the latest Ultra High Definition TV (UHDTV) [1] systems aims to increase the user's Quality of Experience (QoE) by introducing to the existing High Definition TV (HDTV) system [2] new features such as higher spatial resolution, High Dynamic Range (HDR), wider color gamut and High Frame-Rate (HFR) [3, 4]. The introduction of these new features is driven by the will to provide a more realistic and immersive experience to the consumer. These new formats require 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.

The scalable extension of the High Efficiency Video Coding (HEVC) standard, SHVC [5], enables to encode the video in several formats, depending on the chosen type of scalability, within a single layered bitstream. However, with its architecture comprising one HEVC encoder instance per layer, SHVC becomes unsuitable for the coding of new immersive video formats and their increased resolutions due to its high complexity [6].

Despite the current trend of increasing the capture and display devices spatial resolutions further and further, such resolutions are not necessarily required to represent the critical information of the signal. A well-known strategy consists in downsampling the video prior to encoding and upsampling back to the original resolution

after decoding, in an effort to reduce the bitrate and/or the coding complexity for a similar reconstructed output quality [7–15]. Several studies have considered applying the down/up-sampling at an image level, showing that coding at a lower resolution performs better, especially at low bitrate [7, 8]. Since the spatial frequency components are highly variable within different areas comprised in a single natural image, several investigations have been carried out on coding based on local Adaptive Spatial Resolution (ASR), where the down/upsampling is performed at a block level [9, 10], showing promising results.

Based on the recent advances in the active field of single image Super Resolution (SR) [16, 17], several studies have used SR up-sampling for ASR-based coding showing increased gains compared to common upsampling filter-banks [11–15]. However, these gains come at the expense of a significant increase in processing complexity, making SR unsuitable for low-complexity encoding/decoding.

Inspired by the above mentioned previous works, we propose, in this paper, a two-layer low-complexity scalable encoding scheme, which uses a standard Base Layer (BL) encoder and an Enhancement Layer (EL) encoder based on local adaptation of the spatial resolution.

The main contributions of this paper include:

- the design of a low-complexity block-level ASR-based EL encoder that determines the optimal spatial resolution among a list of several sampling-rates through a Rate Distortion Optimization (RDO) stage and a decision refinement process.
- the prediction of the block-level EL spatial resolution thanks to a derivation process based on the re-utilization of BL encoder Motion Vector(s) (MV).
- the implementation of this low-complexity EL encoder over the HEVC reference software (HM16.12) as base layer.

The remainder of this paper is organized as follows. Section 2 describes the proposed low-complexity scalable architectures, first the ASR-based encoder and then its extension with resolution prediction via motion compensation. Section 3 presents the test methodology and provides an analysis of the experimental results. Section 4 concludes this paper.

## 2. PROPOSED METHOD

### 2.1. Scalable coding scheme

The proposed scalable architecture, based on the one used in SHVC [5], is depicted in **Fig. 1**. On one hand, the input Ultra High Definition (UHD) (3840x2160 pixels) video is first downsampled to High Definition (HD) (1920x1080 pixels) resolution with the filter bank detailed in [18] before being fed to the base-layer encoder, a standard HEVC encoder. On the other hand, the input UHD

The authors acknowledge funding from the French government through the National Research Agency (ANR) investment ANR-A0-AIRT-07. E-mail: glenn.herrou@b-com.com, wassim.hamidouche@insa-rennes.fr and luce.morin@insa-rennes.fr

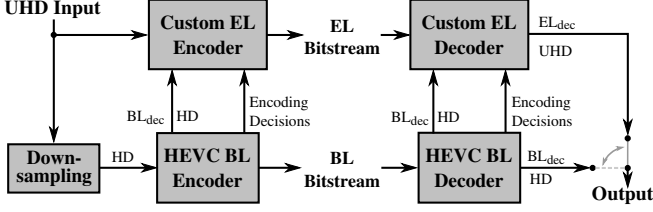


Fig. 1: Block diagram of the proposed scalable coding chain.

video is fed to the enhancement layer encoder whose architecture is described in the next section.

Both bitstreams are then transmitted and processed by the decoder of their respective layer, the EL bitstream only being decoded if the desired output is the higher resolution video, e.g. UHD in the chosen use-case.

To facilitate the processing of the UHD input video, inter-layer communications are enabled in this scalable coding scheme, with several encoding decisions made by the BL encoder (partitioning, prediction mode etc.) transmitted to the EL encoder, in addition to the decoded BL video.

## 2.2. ASR-based enhancement layer encoder

The local adaptation of the spatial resolution is at the heart of the proposed low-complexity EL encoder. Indeed, since natural images/videos comprise locally variant spatial frequency components, it can be beneficial to adjust the sampling rate at a block level to perform a down/up-sampling based encoding [9, 10]. The proposed EL encoder allows for different combinations of horizontal and vertical sampling rates of 1 and 1/2, achieved by successive 1D down-samplings of the EL input block. This leads to the set of possible resolution modes  $L = \{2N \times 2N, N \times 2N, 2N \times N, N \times N\}$ , with  $N$  the width of the square block in BL resolution, and  $W \times H$  referring to an EL block coding spatial resolution of width  $W$  and height  $H$ .

Fig. 2 depicts the proposed EL encoder architecture for a selection of the spatial resolution based on Rate-Distortion (R-D) optimization. For each EL block, whose size has been directly derived from the decision made by the BL encoder for its co-located BL block, the spatial resolution selection process is as follows.

The block is first successively encoded using each possible resolution mode to obtain the different R-D costs, computed as:

$$J(m) = D(m) + \lambda \cdot R(m), \quad (1)$$

with  $J$  the R-D cost,  $D$  the distortion and  $R$  the bit cost associated to resolution mode  $m$ . The derivation of Lagrangian weighting factor  $\lambda$  from the Quantization Parameter (QP) has been carefully optimized for the resolution mode selection problem following the method presented in [19]. Since the selection of the resolution with the lowest R-D cost would result in a highly variable coding spatial resolution, even in areas with homogeneous spatial features, a resolution refinement process has been designed. Indeed, excessively frequent resolution transitions, especially at higher QP values, would tend to lower the visual quality of the reconstructed output.

Thus, a list  $L_{best}$  of resolution modes having an associated R-D cost close to the optimal one,  $J_{min}$ , is first computed by Equation (2).

$$m \in L_{best} \text{ if } J(m) \in [J_{min}, J_{min} + \alpha \cdot W_B], \quad (2)$$

where  $W_B$  is the width of the block  $B$  under consideration in input EL resolution,  $\alpha$  is a parameter defining the strength of the resolution refinement process.

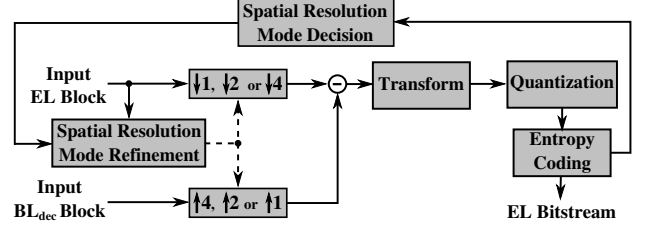


Fig. 2: Proposed EL encoder for R-D optimized resolution selection. Upsampling factors 4, 2 and 1 and downsampling factors 1, 2 and 4 correspond to resolution  $2N \times 2N$ ,  $N \times 2N$  or  $2N \times N$  and  $N \times N$ , respectively.

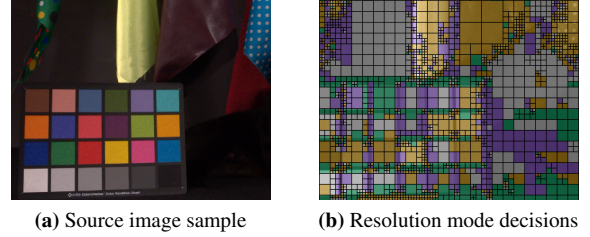


Fig. 3: Example of resolution mode decisions for rate-distortion optimized resolution selection. (yellow:  $2N \times 2N$ , green:  $N \times 2N$ , purple:  $2N \times N$ , white:  $N \times N$ )

Then, two different tests are carried out depending on the optimal resolution mode candidates comprised in the  $L_{best}$  list. First, if one of the candidates is a rectangular resolution, i.e. either  $2N \times N$  or  $N \times 2N$ , the input EL block is analysed to assess the presence of strong directional edges. Sobel-based gradients are thus computed along vertical and horizontal directions, and the rectangular resolution under consideration is selected as the final coding spatial resolution if the inequality expressed in Equation (3) is verified.

$$\overline{S_{B,d_{2N}}} > T_G \cdot \overline{S_{B,d_N}}, \quad (3)$$

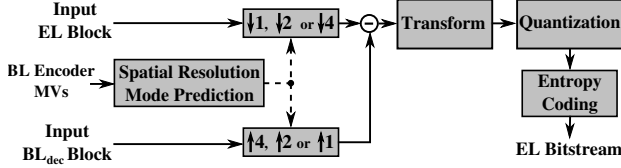
with  $\overline{S_{B,d_{2N}}}$  and  $\overline{S_{B,d_N}}$  the average Sobel-based gradient, along axes with higher and lower resolution, respectively, for the image block  $B$ .  $T_G$  is a threshold used to adjust the edge detection process.

Secondly, if the previous condition does not yield to the final resolution choice, the top and left neighboring blocks are subsequently analysed and compared to the current block under consideration. If the neighboring block coding spatial resolution is comprised in  $L_{best}$ , a list  $Q$  of similarity metrics is computed from the Sobel-based gradients of the current EL block and its considered neighbor. Then, if the criterion defined in Equation (4) is satisfied, the selected final coding resolution is the resolution of the neighboring block.

$$\forall q \in Q : q \in \left[ \frac{1}{T_C}, T_C \right], \quad (4)$$

with  $Q = \left\{ \frac{W_B}{W_H}, \frac{\overline{S_{B,h}}}{\overline{S_{H,h}}}, \frac{\overline{S_{B,v}}}{\overline{S_{H,v}}}, \frac{\overline{S_{B,h} \cdot S_{H,v}}}{\overline{S_{B,v} \cdot S_{H,h}}} \right\}$ ,  $B$  the current block,  $H$  the neighboring block under test,  $W_i$  the width of block  $i$  and  $S_{i,d}$  the average Sobel-based gradient of block  $i$  along direction  $d$ .  $T_C$  is a parameter defining the similarity threshold between two considered blocks.  $T_G$  and  $T_C$  have been optimized to give the best performance results on a subset of the test sequences used in Section 3.

Otherwise, the final choice is set to the resolution mode resulting in the lowest R-D cost. Fig. 3 shows the resolution mode decisions after refinement for a sample of the *CatRobot* sequence.



**Fig. 4:** EL encoder architecture for resolution mode derivation via motion compensation.

Once the final spatial resolution has been selected, the EL block (resp. corresponding decoded BL block) is downsampled (resp. upsampled) to the coding resolution, using the same filter banks as the SHVC reference software [18], to generate the prediction residuals. Then, these residuals are transformed, quantized and entropy coded using the same tools as HEVC [20]. The only difference is the addition of two context-coded bins to the Context-Adaptive Binary Arithmetic Coding (CABAC) engine in order to signal, in the bitstream, the chosen resolution among the four possible ones.

### 2.3. Resolution derivation via motion compensation

To further leverage the decisions made by the base layer encoder, the BL motion vectors, when available, are used to derive the resolution mode from previously encoded EL frames. **Fig. 4** depicts the EL encoder architecture for this case of resolution mode prediction through motion compensation.

The RDO and resolution refinement steps are thus replaced by a resolution prediction stage for EL blocks with a corresponding inter-coded block in the BL. For such an EL block, the MVs and reference frame are first extracted from the base layer. Then, after a proper scaling of the MVs to match the input EL resolution, the EL reference block is obtained to perform the resolution mode prediction. Since it is possible for the reference block to overlap several coding blocks of the reference frame, thus comprising pixels that have been encoded using different resolutions, a derivation process has been designed as follows:

$$r = \begin{cases} 2N \times 2N & \text{if } \mathbb{P}(m = 2N \times 2N) > p_{Th} \\ \arg \max_{s \in L} \mathbb{P}(m = s) & \text{otherwise} \end{cases} \quad (5)$$

with  $r$  the derived resolution from the set of possible resolutions  $L = \{2N \times 2N, N \times 2N, 2N \times N, N \times N\}$  and  $\mathbb{P}(m)$  the probability corresponding to the relative frequency of pixels in the EL reference block in resolution mode  $m$ , with  $m \in L$ . The probability threshold  $p_{Th}$  for the selection of the  $2N \times 2N$  resolution has been introduced to avoid losing important details when the majority of the reference block has not been coded in the EL input resolution. The value of 0.3 has been selected to give the best performance results on a subset of the test sequences used in Section 3. In the case of bi-predictions, the derivation process remains the same, with the probabilities computed across both reference blocks. It should be noted that to reuse BL decisions during the EL encoding process, the EL encoder has to follow the same Group of Pictures (GOP) structure and keep the same reference frame buffer management as the BL encoder.

With this resolution mode prediction process, the signalling of the chosen spatial resolution is not required, for each block, in the generated EL bitstream since the prediction can also be performed at the decoder side. Thus, in addition to the added temporal coherence brought by the motion compensated resolution prediction, the signalling cost of the ASR technique is considerably reduced when MVs are available in the base layer.

**Table 1:** BD-Rate results (%) for proposed ASR-based scalable encoder compared to SHVC. *BL in AI config, EL in P<sub>BL</sub> config.*

Sequence	PSNR-Y	PSNR-U	PSNR-V	PSNR-YUV
S1. DaylightRoad	4.43	2.11	-1.49	3.67
S2. CatRobot	3.16	9.71	12.98	5.4
S3. Drums	5.15	9.33	7.67	6.03
S4. Tango	3.63	3.67	4.55	4.22
S5. RollerCoaster	4.0	10.65	7.81	5.38
S6. FoodMarket2	5.08	7.92	8.26	6.06
<b>Average</b>	<b>4.24</b>	<b>7.23</b>	<b>6.63</b>	<b>5.13</b>
<b>Without resolution signalling cost</b>	<b>0.67</b>	<b>3.8</b>	<b>3.26</b>	<b>1.62</b>

## 3. EXPERIMENTAL RESULTS

### 3.1. Test methodology

In this section, the proposed scalable encoder architecture is evaluated and compared to SHVC both in terms of coding performance and encoding complexity. Two different studies are carried out, one comparing the proposed ASR-based encoder depicted in **Fig. 2** with SHVC, and the other comparing the proposed scalable encoder with resolution prediction via motion compensation, as shown in **Fig. 4**, with SHVC.

Encodings have been performed with the SHVC reference software SHM9.0 [21]. Both the proposed encoder and the reference SHVC encoder use the same base layer, respectively encoded in All Intra (AI) or Random Access (RA) configuration for the first and second studies. The considered RA configuration uses a GOP size of 16 images with an intra-period of approximately one second.

Considering the EL encoders, for both studies, the SHM software has been limited to use either intra prediction modes or inter-prediction modes with the upsampled decoded base layer as reference ( $P_{BL}$  only) for fair comparison purposes. Fixed-QP configuration has been used for every encodings carried out in both studies. The parameters used to define the strength of the resolution refinement process have been set to  $\alpha = 2.56$ ,  $T_G = 2$  and  $T_C = 1.5$ . These values have been carefully optimized to offer the best results in terms of R-D performance and resolution choice homogeneity. Coding performance has been assessed in terms of rate-distortion using the Bjøntegaard delta metric (BD-Rate) [22], representing the average bit-rate gain/loss (-/+ 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 37 with a step of one have been used for each of the six UHD video test sequences.

### 3.2. Results and analysis

The first study focuses on the performance of the proposed low-complexity ASR-based scalable encoder with only the R-D optimized resolution selection process. BD-rate results are summarized in **Table 1**, positive values corresponding to a bitrate overhead for the proposed scalable encoder compared to SHVC for the same EL output PSNR value.

The average bitrate overhead of the proposed scheme is 5.13 % compared to SHVC. The per-sequence results are fairly homogeneous, around 4 ~ 6 % bitrate overhead for equal PSNR-YUV values, the variations mainly depending on the content of the video, i.e. the ability of the resolution selection process to locally reduce

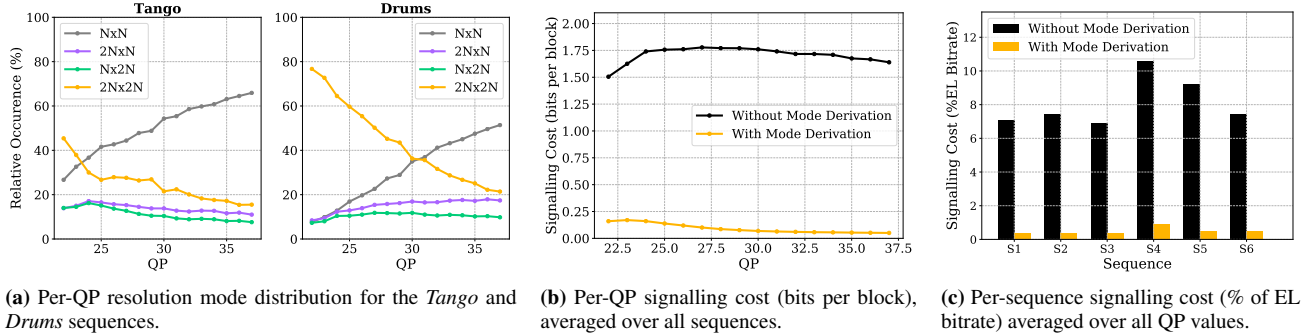


Fig. 5: Example of resolution mode distribution and signalling costs with and without motion compensated resolution prediction.

Table 2: BD-Rate results (%) for proposed scalable encoder with mode derivation compared to SHVC. *BL* in *RA* config, *EL* in *P<sub>BL</sub>* config.

Sequence	PSNR-Y	PSNR-U	PSNR-V	PSNR-YUV
S1. DaylightRoad	-0.31	-15.3	-16.41	-3.98
S2. CatRobot	-1.48	-4.42	3.71	-0.85
S3. Drums	4.16	0.86	0.49	3.81
S4. Tango	-2.58	-5.96	-7.63	-3.7
S5. RollerCoaster	-0.36	-0.03	-1.26	-0.2
S6. FoodMarket2	0.22	0.29	0.12	0.39
<b>Average</b>	<b>-0.1</b>	<b>-4.09</b>	<b>-3.50</b>	<b>-0.76</b>

the resolution whenever possible. For example, Fig. 5a depicts the per-QP resolution mode distribution for the *Tango* and *Drums* sequences. As can be observed, the ASR algorithm decisions lead to a majority of blocks coded at reduced resolutions even at low QP values for the *Tango* sequence, while the majority resolution transition from  $2N \times 2N$  to  $N \times N$  happens at a much higher QP for the *Drums* sequence. This correlates with the lower (resp. higher) than average bitrate overhead for the *Tango* (resp. *Drums*) sequence. The measured performance gain if the resolution mode signalling cost is not taken into account reaches, on average, 3.5 % of the total bitrate. Hence the interesting coding potential of the adaptive spatial resolution scheme if the signalling cost can be reduced, which is the purpose of the following experiment.

Table 2 summarizes the BD-rate results for the proposed encoder compared to SHVC. Overall, the proposed ASR-based encoder with resolution mode derivation enabled achieves average bitrate savings of 0.76 % compared to SHVC. The performance of both U and V chroma channels are better than the results obtained for the luma channel, with respectively  $-4.1$  %,  $-3.5$  % and  $-0.1$  % average BD-rate values. As can be observed, the proposed encoder performs better than the RDO-based version of the ASR algorithm on sequences that benefit from the changes in spatial resolution, and shows equivalent losses on the other sequences with high spatial details. In particular, performance of chroma components is highly variable depending on the sequence, with BD-rate values ranging from 16.4 % bitrate savings to 3.7 % bitrate overhead for *DaylightRoad* and *CatRobot* sequences, respectively.

Fig. 5b shows the mean bit per block signalling cost, averaged over all sequences, with each tested QP value for the proposed encoder with and without the resolution mode derivation enabled. As can be expected from the BD-rate results, the signalling cost is at least divided by six for every QP when the derivation is enabled. The cost is gradually decreasing when the quantization parameter in-

Table 3: Encoding time (% of SHVC) of the proposed scalable encoder compared to SHVC.

	Enhancement Layer	Overall (EL + BL)
Without inter-predicted BL	12 %	28 %
With inter-predicted BL	4 %	53 %

creases due to the overall lower amount of intra coded blocks in the base layer, at lower bitrates. Fig. 5c depicts the per-sequence portion of the EL bitrate taken by the resolution mode signalling, averaged over all QP values. The cost reduction factor is nearly constant for all sequences, with an average value of 17.5.

Concerning the codec complexity, the proposed ASR-based scalable coding scheme achieves a 72 % encoding time reduction compared to the SHVC reference software, as reported in Table 3. If only the EL layer is considered, the proposed architecture is 8.5 times faster than SHM at the encoder side. This complexity gain is mostly due to the low computational demand of the few modes to test during the RDO stage and the derived EL partitioning. For the proposed encoder with motion compensated resolution prediction, the EL encoder only represents 4 % of the EL SHM encoding time. The lower computation time compared to the previously presented version of the ASR-based encoder is due to the skip of the RDO stage when MVs are available. Considering the overall complexity, the reduction reaches 47 % compared to the SHM. This high difference between the EL-only and overall results can be explained by the increased complexity related to the inter-picture predictions in the BL encoder.

#### 4. CONCLUSION AND FUTURE WORK

This paper proposes a two-layer low-complexity scalable encoding scheme based on local adaptations of the spatial resolution in the enhancement layer encoder. This EL encoder, which heavily re-uses the BL encoder decisions, allocates the spatial resolution at a block-level, either by performing a RDO stage followed by a carefully designed decision refinement process to select the resolution among a list of several sampling rates or by predicting the spatial resolution via motion compensation. This framework has been implemented over the HM16.12 reference software used as BL encoder. Experimental results show that the proposed framework achieves 0.76 % bitrate savings compared to SHVC. Complexity-wise, encoding time reductions of 47 %, for the dual-layer architecture, and 96 %, if only the EL is taken into account, are observed. Future work will include the integration of inter-predictions in the ASR-based EL encoder and real-time implementation of the proposed solution.

## REFERENCES

- [1] ITU-R, "Recommendation BT.2020-1: Parameters Values of Ultra-High Definition Television Systems for Production and International Programme Exchange," .
- [2] ITU-R, "Recommendation BT.709-5: Parameter Values for the HDTV Standards for Production and International Programme Exchange," .
- [3] Mike Nilsson, "Ultra high definition video formats and standardisation," *BT Media and Broadcast Research Paper*, 2015.
- [4] Masayuki Sugawara and Kenichiro Masaoka, "UHDTV Image Format for Better Visual Experience," *Proceedings of the IEEE*, vol. 101, no. 1, pp. 8–17, January 2013.
- [5] Jill M Boyce, Yan Ye, Jianle Chen, and Adarsh 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.
- [6] Wassim Hamidouche, Mickael Raulet, and Olivier Déforges, "4k real-time and parallel software video decoder for multi-layer hevc extensions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 169–180, 2016.
- [7] Alfred M Bruckstein, Michael Elad, and Ron Kimmel, "Down-scaling for better transform compression," *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 1132–1144, 2003.
- [8] Mariana Afonso, Fan Zhang, Angeliki Katsenou, Dimitris Agrafiotis, and David Bull, "Low complexity video coding based on spatial resolution adaptation," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 3011–3015.
- [9] Weisi Lin and Li Dong, "Adaptive downsampling to improve image compression at low bit rates," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2513–2521, 2006.
- [10] Viet-Anh Nguyen, Yap-Peng Tan, and Weisi Lin, "Adaptive downsampling/upsampling for better video compression at low bit rate," in *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*. IEEE, 2008, pp. 1624–1627.
- [11] Mariana Afonso, Fan Zhang, and David R Bull, "Video compression based on spatio-temporal resolution adaptation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 275–280, 2019.
- [12] Minmin Shen, Ping Xue, and Ci Wang, "Down-sampling based video coding using super-resolution technique," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 6, pp. 755–765, 2011.
- [13] Ren-Jie Wang, Chih-Wei Huang, and Pao-Chi Chang, "Adaptive downsampling video coding with spatially scalable rate-distortion modeling," *IEEE transactions on circuits and systems for video technology*, vol. 24, no. 11, pp. 1957–1968, 2014.
- [14] Jie Dong and Yan Ye, "Adaptive downsampling for high-definition video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 3, pp. 480–488, 2014.
- [15] Yue Li, Dong Liu, Houqiang Li, Li Li, Feng Wu, Hong Zhang, and Haitao Yang, "Convolutional neural network-based block up-sampling for intra frame coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2316–2330, 2018.
- [16] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
- [17] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1637–1645.
- [18] J Chen, J Boyce, Y Ye, and MM Hannuksela, "Shvc test model 9 (shm 9) introduction and encoder description," *Document JCTVC-T1007. Geneva, Switzerland*, 2015.
- [19] Gary J Sullivan and Thomas Wiegand, "Rate-distortion optimization for video compression," *IEEE signal processing magazine*, vol. 15, no. 6, pp. 74–90, 1998.
- [20] Gary J Sullivan, J-R Ohm, Woo-Jin Han, and Thomas 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.
- [21] SHVC reference software version 9.0, ,  
[https://hevc.hhi.fraunhofer.de/svn/svn\\_SHVCSoftware/tags/SHM-9.0/](https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSoftware/tags/SHM-9.0/).
- [22] Gisle Bjøntegaard, "Improvements of the BD-PSNR model," *ITU-T SG16 Q*, vol. 6, pp. 35, 2008.