LOW COMPLEXITY VIDEO CODING BASED ON SPATIAL RESOLUTION ADAPTATION

Mariana Afonso, Fan Zhang, Angeliki Katsenou, Dimitris Agrafiotis and David Bull

Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK
{Mariana.Afonso, Fan.Zhang, Angeliki.Katsenou, Dimitris.Agrafiotis, Dave.Bull}@bristol.ac.uk

ABSTRACT

In this paper, a novel spatial resolution adaptation approach for video compression is proposed. Its ability to dynamically apply downsampling to frames exhibiting low spatial detail delivers improved rate distortion performance, together with a reduction in computational complexity of the encoding process. This method is based on an experimental investigation of the dependence between the QP threshold, which determines when to encode lower resolution frames, and the distortion obtained after downsampling/upsampling. The proposed approach is integrated with the High Efficiency Video Coding (HEVC) reference codec for intra coding, and evaluated on 15 high-resolution test sequences with varying levels of spatial detail. The results show a promising average bitrate savings of approximately 4% (B-D measurements), and significant complexity reduction (29% on average).

Index Terms— Video compression, spatial resolution adaptation, image scaling, HEVC

1. INTRODUCTION

Due to the significant increased demands on bandwidth from immersive video formats, the challenge remains for video coding to achieve high compression efficiency while maintaining a relatively low complexity. In this context, numerous different approaches have been developed to improve rate-distortion performance. One such method, referred to resolution adaptation, attempts to trade off the relationship between spatial detail and quantisation through encoding a video at a lower resolution and reconstructing the full resolution version after decoding.

Resolution adaptation algorithms can be applied at various stages in video compression; for example, as a new mode at the macroblock or Coding Tree Unit (CTU) level [1–3], or as a pre-processing step prior to encoding for each frame [4], for each group of pictures (GOP) [5], or for a whole sequence [6,7]. In all these methods, the downsampling and upsampling approaches are crucial for reconstructing high quality full resolution video content. Simple interpolation filters [4, 5], such as bicubic, are commonly used due to their high efficiency, while super-resolution techniques [3,4,7] provide improved performance in terms of visual quality at the expense of increased complexity.

Several previous works on this topic assumed that resolution adaptation was only applicable at lower bitrates [4, 7]. This assumption neglects the content dependent nature of these adaptation approaches, and restricts their application in video compression. To overcome this problem, adaptive methods were proposed to determine the optimal spatial resolution at a target bitrate by decomposing the overall distortion into a downsampling distortion and a coding distortion, assuming both distributions to be independent [5, 6]. However modelling the rate-distortion (R-D) performance is difficult and normally requires pre-encoding, which may further increase computational complexity. Finally, adaptive resolution methods have also applied in the context of the Scalable Video Coding (SVC) [8] and rate control [9].

This paper proposes an efficient adaptive resolution coding scheme, where given a frame and a quantization parameter (QP), a decision is performed on whether to encode a lower resolution frame and upsample at the decoder, or to encode the original resolution frame. Additionally, it also determines the best QP value to use when encoding at a lower resolution. It has been integrated into the High Efficiency Video Coding (HEVC) [10] reference codec, and shows promising (up to 7.6%) overall bitrate savings using the Bjøntegaard delta rate metric (BD-rate) [11]. In addition, complexity is also significantly reduced (up to 54%) compared to encoding full resolution frames.

The rest of the paper is organized as follows. Section 2 describes the adaptive downsampling coding scheme proposed, including the two key modules, image scaling (Section 2.1), and the resolution-quantization optimization (Section 2.2). Next, Section 3 describes the experimental design employed and the results obtained. Finally, Section 4 discusses the conclusions of this work and explores ideas for future research.

2. ADAPTIVE RESOLUTION FRAMEWORK

The proposed adaptive resolution coding scheme is illustrated diagrammatically in Figure 1. First, each frame of the input original video is downsampled to a lower resolution (I_d) and

This work was supported by the Marie Skodowska-Curie Actions - FP7 EU programme, project PROVISION ITN.

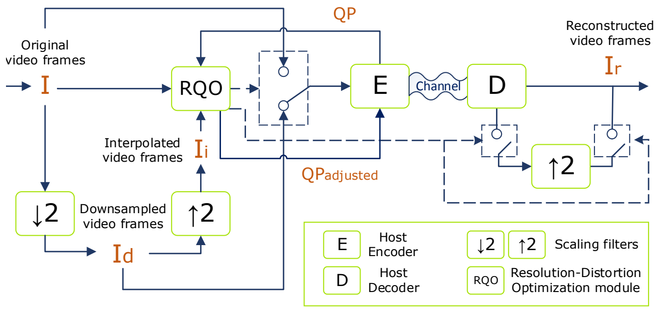


Fig. 1: Diagram of the proposed adaptive resolution framework.

then upsampled to its original resolution (I_i). For simplicity and due to the observation of superior performance compared to others [12], a fixed scaling factor of 2 is applied. The distortion of the upsampled frame with respect to its original version, alongside the input quantisation parameter (QP) used by in the host encoder, are used by the Resolution-Quantization Optimization (RQO) module to determine at which video resolution to encode. If the lower resolution is selected, the downsampled frame is sent to the host encoder (E), using an adjusted QP. Otherwise, the original frame will be compressed. This adaptation process requires one additional bit per frame to signal the downsampling status at the decoder (D) and output the reconstructed frame (I_r).

The performance of this coding scheme depends on two key modules: (i) Image Scaling and (ii) Resolution-Quantisation Optimisation. These are described below.

2.1. Image Scaling

Image scaling refers to the resizing of a digital image, which can be realised using various sampling filters or through super-resolution algorithms. The latter generally produces more accurate reconstruction at the expense of high computationally complexity. In this paper, two popular filter kernels for image scaling are tested: bicubic and Lanczos3 [13], alongside a state-of-the-art super-resolution (SR) algorithm [14]. It is noted that the Lanczos filter with order 3 is used in this work given that it has been shown to be the best compromise between reconstruction accuracy and avoiding ringing artifacts [15]. Additionally the window size of the bicubic and Lanczos3 filters are 4×4 and 6×6 pixels respectively. Figure 2 illustrates the performance of the tested methods by comparing the visual quality and PSNR of the interpolated first frame of the test sequence *BQTerrace* (see Section 3), which had been previously downsampled using a scale of two.

All three image scaling methods were evaluated on twelve HD (1920×1080) video sequences listed in Table 2, and only the first frame of each sequence was tested. For the SR algorithm and bicubic filter, the original frames are first down-

(a) Original

(b) Bicubic [PSNR = 25.4 dB]

(c) Lanczos3 [PSNR = 26.4 dB]

(d) SR [14] [PSNR = 29.1 dB]

Fig. 2: Visual quality comparison of a patch of the first frame of the BQTerrace sequence interpolated by three different methods.

sampled using the bicubic filter and then upsampled using the tested approach. For the Lanczos3 filter, the same method was used for both the downsampling and upsampling processes.

Table 1: Average PSNR and relative execution time of the three interpolation methods tested for 12 HD test sequences (see Table 2).

	Bicubic	**Lanczos3**	**SR**	**HEVC Intra Coding**
PSNR (dB)	33.5	34.4	36.0	-
Relative time	1	1.24	66701	1025

The performance of tested approaches, in terms of the reconstructed accuracy and computational complexity, is summarised in Table 1. Here PSNR is used to assess the quality of the upsampled frames. The complexity figures of all test algorithms are estimated based on their average execution time, obtained on an Intel Core i7-4790 CPU@3.60GHz PC Platform. The encoding time of the HEVC reference codec (HM 16.2, All Intra configuration [16]) is also presented here for benchmarking, and this figure is obtained by averaging the execution time of cases using QP values 22, 27, 32, 37 and 42.

It can be observed from Table 1 that the SR method offers the best reconstruction accuracy among all three approaches, with an average PSNR value of 36 dB. The Lanczos3 filter is the second best performer, providing approximately 1 dB PSNR gain over Bicubic. However, the SR method requires much higher computational complexity, several orders of magnitude relative to that of Bicubic, and 65 times that of HEVC intra coding. On the other hand, the complexity of the Lanczos3 filter is only 24% greater than Bicubic.

Based on the previous analysis, the method selected for this work is the Lanczos3 filter, since it offers the best trade-off between accurate reconstruction and complexity.

2.2. Resolution-Quantization Optimization

As discussed earlier, the Resolution-Quantization Optimization module is responsible for determining which frames to downsample based on an input or target QP value ($\text{QP}_{\text{target}}$). Previous works [2, 7] have reported that there is generally a bitrate/QP value at which the R-D performance of encoding lower resolution frames and upsampling at the decoder just outperforms encoding full-size frames. This value is referred as the QP threshold. In order to understand how the QP threshold varies with the video content, encoding at the original and lower resolution was performed using the Lanczos3 scaling filter (see Section 2.1). For these experiments, a subset of 50 low resolution (256×256) sequences from the HomTex dataset [17] were used. The decision for selecting this dataset was justified in terms of time constraints, the availability of a large number of sequences and the large variety of content. Ten equally sampled frames from each sequence were selected giving a total of 500 frames. Again, the host codec used was HEVC test model HM 16.2, All Intra configuration (main profile), with a QP range from 20 to 50 and a step size of one. Finally, the R-D curves of the original and lower resolution frames were plotted and the QP thresholds extracted.

Figure 3 illustrates an example of two R-D curves obtained by the experimental setting described for a frame of the *flag* test sequence. It can be seen that the QP threshold is approximately 27. This is what the proposed model aims to predict for every input frame. By analysing the data collected from the experiments, it was found that there is a high correlation between QP thresholds and the distortion caused by scaling the frame to a lower resolution and back up again, i.e., the downsampling distortion. In practice, the downsampling quality (which follows an inverse correlation with the QP threshold) was estimated by computing the PSNR between the original and the interpolated frames. Figure 4 shows a scatter plot of this correlation where every data point represents a different frame. The Pearson Correlation Coefficient was calculated as 0.86 (absolute value). Following the shape of the data points, an exponential fitting was applied (green curve in Figure 4). The fitted curve (in green) is represented by the following equation:

$$\text{QP}_{\text{thres}}(q) = 10^{\alpha+\beta q} + K \tag{1}$$

Where QP_{thres} is the QP threshold, q is the downsampling quality or PSNR, expressed in dB, $\alpha = 1.92$ and $\beta = -0.01$ are the fitted parameters and K is an optional constant. The goodness of fit, evaluated by the Mean Absolute Error is 2.708 QP steps. In this work, in order to obtain a more conservative prediction of the QP threshold, $K = 2$ was used (red curve in Figure 4). This offset is able to reduce by 30%

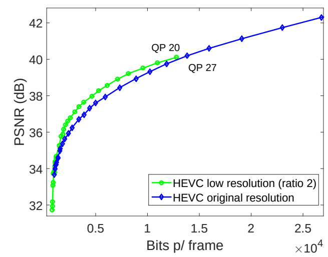


Fig. 3: R-D curves obtained by coding the original and lower resolution of the 89th frame of *flag*.

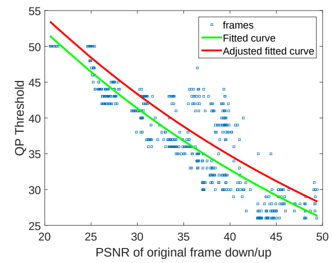


Fig. 4: Relationship between the PSNR of the downsampled frame and the QP threshold.

the number of frames where the predicted QP threshold is lower than the actual value, while maintaining a relatively accurate model.

Based on this model and under the assumption that it generalizes for full size frames, it is possible to obtain a prediction of the QP_{thres} for a given frame based only on the value of q. However, once the value of QP_{thres} is exceeded, the QP value needs to be adjusted for the low resolution frame. Generally, for the same bitrate, the QP for the low resolution frame, QP_{low}, is smaller than QP_{thres}. This is illustrated in Figure 3, where QP_{low} is a lower value than the QP_{thres} in the intersection of the two curves. Using the training data mentioned before, these two values were compared and the result is that there is a linear relationship between them in the form of $\text{QP}_{\text{low}} = \text{QP}_{\text{thres}} - K'$ with K' between 6 and 7. Therefore, a linear model was fitted to this data and used for the prediction model. Finally, for an input QP, QP_{thres}, the QP to use for encoding, QP_{new} is given as follows:

$$\begin{aligned} \text{QP}_{\text{diff}} &= \text{QP}_{\text{target}} - \text{QP}_{\text{thres}} \\ \text{QP}_{\text{new}} &= \begin{cases} \text{QP}_{\text{target}}, & \text{if } \text{QP}_{\text{target}} < \text{QP}_{\text{thres}} \\ \text{QP}_{\text{diff}} + \text{QP}_{\text{low}}, & \text{if } \text{QP}_{\text{target}} \geq \text{QP}_{\text{thres}}. \end{cases} \end{aligned} \tag{2}$$

3. EXPERIMENTAL RESULTS

The proposed resolution adaptation approach has been integrated into the HEVC reference codec (HM 16.2), and was fully tested using the All Intra configuration (main profile) [16] on the first 300 frames (when available) of 15 high resolution test sequences. The QP values used ranged from 22 to 42 with a step of 5.

All test sequences are from public video databases, namely, the BVI-HFR database [18], the HEVC recommended test database [16], the Tampere 4K test dataset [19], and the Derf's Test Media Collection [20]. Table 2 lists all the test sequences used. The sequences contain different levels of spatial detail, ranging from very textured to more homogeneous or blurry content. Based on the amount of spatial detail, they were divided into three groups: low (Group I), medium (Group II) and high spatial detail (Group III). The metric used

Table 2: Results of the proposed adaptive resolution coding scheme against HEVC using several test sequences divided into three groups (G). BVI-H: Bristol Vision Institute High Frame Rate dataset [18], HEVC: test sequences from the Common Test Conditions [16], T4K: Tampere 4K test sequences dataset [19], Xiph: Derf's Test Media Collection [20], Res.:Spatial resolution, HFE: High frequency energy density (J/Hz), T: Total execution time difference, R: BD-Rate.

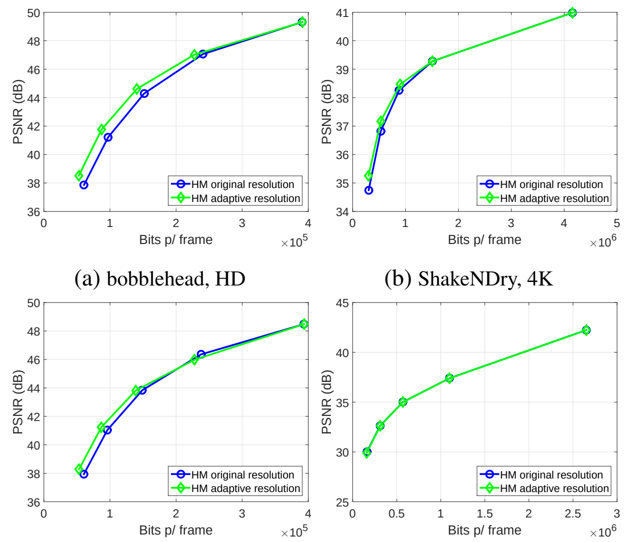
G	Sequence	Dataset	Res.	HFE	T (%)	R (%)
I	bobblehead	BVI-H	HD	0.2	-56.0	-12.5
	bouncyball	BVI-H	HD	0.4	-57.4	-7.1
	catch_track	BVI-H	HD	0.2	-45.5	-3.7
	guitar_focus	BVI-H	HD	0.2	-55.3	-7.2
	pour	BVI-H	HD	0.3	-57.2	-7.8
	Average			**0.3**	**-54.3**	**-7.6**
II	ParkScene	HEVC	HD	5.4	-17.2	-0.5
	Kimono1	HEVC	HD	3.1	-32.6	-2.4
	Bosphorus	T4K	4K	2.7	-31.0	-5.3
	HoneyBee	T4K	4K	8.1	-30.4	-5.3
	ShakeNDry	T4K	4K	8.1	-28.8	-6.9
	Average			**6.0**	**-28.0**	**-4.1**
III	BQTerrace	HEVC	HD	16.1	-8.2	0.1
	Cactus	HEVC	HD	10.1	-10.4	0.0
	CrowdRun	Xiph	HD	20.1	-3.9	0.1
	DucksTakeOff	Xiph	HD	22.8	-7.2	-0.3
	ParkJoy	Xiph	HD	33.5	-0.3	0.1
	Average			**20.5**	**-6.0**	**-0.03**
	Overall average				**-29.4**	**-3.9**

was the high frequency energy density which was obtained by integrating the Power Spectral Density (PSD) [21] of the frequencies that would ideally by cutoff by the downsampling filter (using a factor of two). These values along with the groupings can also be found in Table 2.

The compression performance of the proposed method for HEVC intra coding is compared with the corresponding encoding of full resolution content. The results are based on BD-rate measurements [11] over all frames are shown in the last column of Table 2. It can be observed that the proposed method always provides better or nearly equivalent performance against full resolution compression for all test sequences, with average bitrate savings of 3.9%. This improvement is more significant on sequences with lower spatial detail (Group I and II) than on those with higher spatial detail.

Moreover, Table 2 also shows that the proposed coding scheme significantly reduces the computational complexity compared to original HEVC compression on full resolution content, by 29% on average. This is due to the fact that selected frames are encoded at a lower resolution, which decreases the overall execution time in exchange for a very low added complexity (by applying scaling filters).

Example R-D curves comparing the proposed method with full resolution compression for four test sequences are shown in Figure 5. The selected sequences represent typical content from each test group, and are with various spatial resolutions. For the test sequence *bobblehead* (Group I), both approaches have equal performance for the first QP value,



(a) bobblehead, HD (b) ShakeNDry, 4K

(c) guitar_focus, HD (d) BQTerrace, HD

Fig. 5: R-D curves of the proposed coding scheme and the anchor, HEVC, for four sequences tested.

QP $= 22$. This means that this QP is lower than the predicted QP threshold and, therefore, the original resolution frames are encoded. However, for higher QP values, the adaptive resolution approach significantly outperforms HEVC. A similar pattern is shown in *ShakeNDry* (Group II), except that for this sequence, the QP threshold was predicted as a higher value. In the case of *guitar_focus* (Group I), the proposed adaptive resolution provides a slightly lower R-D performance compared to HEVC for the second QP value, QP $=$ 27, which reflects an inaccurate prediction of the QP threshold. Nonetheless, for the whole range of QP values, significant gains are still achieved (see Table 2). Lastly, for the sequence *BQTerrace* (Group III), the two R-D curves overlap for all QP values. This suggests that the prediction model is able to make conservative decisions at higher bitrates and sequences with high spatial detail.

4. CONCLUSIONS

This paper presents a resolution adaptation approach for HEVC intra coding, which adaptively enables frame downsampling before encoding and reconstructs to the original resolution after decoding. The determination of resolution adaptation is based on the observation of the relationship between the downsampling quality and the QP used for compression. When integrated into HEVC intra coding, the proposed method achieves promising BD-rate savings, and reduces the computational complexity significantly. The future work will focus on multiple sampling ratios and the application in inter coding.

5. REFERENCES

[1] S. Uslubas, E. Maani, and A. Katsaggelos, "A resolution adaptive video compression system," in *Intelligent Multimedia Communication: Techniques and Applications*, pp. 167–194. Springer, 2010.

[2] V. Nguyen, Y. Tan, and W. Lin, "Adaptive downsampling/upsampling for better video compression at low bit rate," in *IEEE International Symposium on Circuits and Systems*, 2008, pp. 1624–1627.

[3] Yue Li, Dong Liu, Houqiang Li, Li Li, and Feng Wu, "Convolutional neural network-based block up-sampling for intra frame coding," *arXiv preprint arXiv:1702.06728*, 2017.

[4] M. Shen, P. Xue, and C. 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.

[5] R. Wang, C. Huang, and P. 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.

[6] J. Dong and Y. 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.

[7] G. Georgis, G. Lentaris, and D. Reisis, "Reduced complexity superresolution for low-bitrate video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 2, pp. 332–345, 2016.

[8] B. Hosking, D. Agrafiotis, D. Bull, and N. Easton, "Enhancement of intra-coded pictures for greater coding efficiency," in *Picture Coding Symposium*, 2016.

[9] B. Hosking, D. Agrafiotis, D. Bull, and N. Eastern, "An adaptive resolution rate control method for intra coding in hevc," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 1486–1490.

[10] G. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[11] G. Bjontegaard, "Calcuation of average psnr differences between rd-curves," *Doc. VCEG-M33 ITU-T Q6/16, Austin, TX, USA*, 2001.

[12] R. Wang, M. Chien, and P. Chang, "Adaptive down-sampling video coding," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2010, pp. 75420P–75420P.

[13] Wilhelm Burger, Mark James Burge, Mark James Burge, and Mark James Burge, *Principles of digital image processing*, Springer, 2009.

[14] J. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5197–5206.

[15] J. Li, Y. Koudota, M. Barkowsky, H. Primon, and P. Le Callet, "Comparing upscaling algorithms from HD to ultra HD by evaluating preference of experience," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, pp. 208–213.

[16] F. Bossen, "Common test conditions and software reference configurations," *Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-L1100*, 2013.

[17] M. Afonso, A. Katsenou, F. Zhang, D. Agrafiotis, and D. Bull, "Video texture analysis based on hevc encoding statistics," in *Picture Coding Symposium*, 2016.

[18] A. Mackin, F. Zhang, and D. Bull, "A study of subjective video quality at various frame rates," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3407–3411.

[19] Tampere University Ultra Video Group, ," `http://ultravideo.cs.tut.fi/`, [Online; accessed 2017-01-23].

[20] Derf's Test Media Collection, ," `http://media.xiph.org/video/derf/`, [Online; accessed 2017-01-23].

[21] D. Bull, *Communicating pictures: A course in Image and Video Coding*, Academic Press, 2014.