Hosking, B., Agrafiotis, D., Bull, D., & Easton, N. (2016). An adaptive resolution rate control method for intra coding in HEVC. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016): Proceedings of a meeting held 20-25 March 2016, Shanghai, China. (pp. 1486-1490). Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/ICASSP.2016.7471924

AN ADAPTIVE RESOLUTION RATE CONTROL METHOD FOR INTRA CODING IN HEVC

Brett Hosking[1], *Dimitris Agrafiotis*[1], *David Bull*[1] *and Nick Easton*[2]

[1] Visual Information Lab, University of Bristol, Bristol, UK
[2] BAE Systems, Chelmsford, UK

ABSTRACT

Previous work has shown that spatial resampling can improve rate-distortion performance by providing a higher and more consistent level of video quality at low bitrates. Rate control aims to regulate the video bitrate in accordance to the bit budget. While this is a well studied problem in the single resolution case, very little progress has been made on the adaptive resolution case. In this paper we present an enhanced method of rate control for intra coding that allows the algorithm to learn from previously coded frames and make more accurate predictions, resulting in a lower average mismatch ratio. Our main contribution, however, lies in our adaptive resolution approach where the best scale factor is selected after prediction of the best Quantisation Parameter (QP). We show that our method closely conforms to the bit budget and provides a more stable bitrate. The likelihood of frame skipping is therefore reduced and a more consistent level of video quality is provided compared to standard single resolution methods.

***Index Terms*—** Adaptive, Resolution, Rate, Intra, HEVC

1. INTRODUCTION

It has been shown that spatial resampling of video sequences, or even single images, can provide better compression performance than simply coding at the original High Resolution (HR) [1, 2, 3, 4, 5, 6]. In [2], Bruckstein et al. show that at low bitrates, an image downsampled, compressed using JPEG and later interpolated, produces better results than an image compressed at the original HR. Wu et al. demonstrate that coding oversampled frames is not only a waste of resources but can also be counterproductive to image quality given a tight bit budget [3]. Dong at al. proposed resampling the entire video sequence after determining the optimal scale factor by minimising the overall distortion caused by downsampling and coding [4, 5]. In [6], Nguyen et al. proposed a method of adapting the scale factor and the quantisation step size according to spatial content. In our previous work [1], we demonstrated the performance of our Spatial Resampling of IDR Frames (SRIF) method using the High Efficiency Video Coding (HEVC) standard. We concentrated on resampling the intra coded Instantaneous Decoding Refresh (IDR) frames only as inter-coded frames are already well compressed, especially in HEVC. We show that SRIF coding can improve rate-distortion performance by providing a higher and more consistent level of video quality at low bitrates. However, the results are not necessarily optimal; a fixed scaling factor is used that provides the best overall performance for the entire sequence at any given bitrate, similar to the work in [4, 5]. For a more practical system, the best scale factor, QP and resampling technique would need be determined for each IDR frame independently.

This paper investigates the problem of assigning the best QP values to intra coded frames for adaptive resolution coding. Predicting the QP that will minimise distortion while conforming to a certain bit budget is a well studied rate control problem in the single resolution case [7, 8, 9, 10, 11, 12] but less so for adaptive resolution, where the best scale factor has to be selected alongside the best QP. This is a problem that in some respect arises in the cases of Scalable Video Coding (SVC) [13] and Scalable HEVC (SHVC) [14] which are extensions of H.264/AVC and H.265/HEVC, respectively, and solutions have been suggested in [15, 16]. Scalable coding provides improved bitstream adaptability by enabling the decoder to extract and decode sub-streams from the full high quality bitstream. Standard coding is applied at the base layer with enhancement layers providing increased temporal, spatial and/or quality gains. Unlike SVC and SHVC, adaptive resolution coding does not deliver multiple resolutions but rather has to decide upon the optimal one given certain bitrate restrictions. In that sense our work is more related to the work of [4, 5]. Our contribution lies in describing a method for estimating the optimal QP and resolution for intra coding on a frame by frame basis without the need for multiple coding passes at each resolution, which would render a real-time frame-based adaptation very difficult.

The remainder of this paper is formatted as follows: Section 2 provides background on rate control along with details of existing work. Next, our proposed method and the corresponding results are described in Sections 3 and 4, respectively. Finally, we present our conclusions in Section 5.

2. BACKGROUND

Rate control schemes aim to regulate the video bitrate in order to prevent overabundant or excessively compressed coded bitstreams. While in most cases the former will result in frame skipping, the latter can cause unnecessary degradation of video quality. On a frame level, rate control first assigns a *bit target* to the current frame given the buffer status and the available bandwidth. For best results, we wish to select the lowest QP that minimises distortion while preventing overflow of the channel buffer. Compression efficiency is determined from a combination of video content and the effectiveness of the encoder, so without coding there is no way of determining the resulting number of coded bits that each QP will generate with absolute certainty. However, for most applications it is only desirable to apply a single-pass rate control method where the QP is determined prior to coding and not subsequently after exhaustively coding with a range of QP values, as is the case with multi-pass methods. Multi-pass methods provides a means of deducing the optimal QP with a higher degree of certainty but at the cost of increased computational complexity. Selecting the optimal QP that meets these requirements without any prior coding is a challenging problem and one that has encouraged a great deal of research, as is described in [17].

2.1. Complexity Measures

In [18], Kim et al. proposed a fast bit allocation method for still image coding using an image *complexity* measure. Image complexity refers to frequency content; the more high frequencies an image contains, the more bits required to code the image given a fixed amount of quantisation. It therefore stands to reason that an image with a higher complexity will be subjected to more distortion given a low bit budget. Using an effective complexity measure it is possible to estimate the number of bits required to code each frame for a selected QP. This removes the need for an exhaustive search and provides a basis for real-time applications. Kim et al. [18] also provide comparisons between various complexity measures and it was found that the average gradient per pixel provided the most statistically accurate result which led to later work on rate control for intra coding adopting the same approach [9, 7, 8, 10]. In practice, to reduce computation while preserving statistical accuracy, it is more desirable to use the simplified formula as provided in [7] and given as follows:

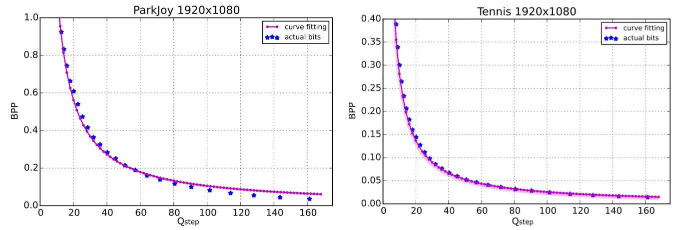
$$G = \frac{\sum_{i=1}^{N_w-1}\sum_{j=1}^{N_h-1}(|I_{i,j} - I_{i+1,j}| + |I_{i,j} - I_{i,j+1}|)}{N_w \times N_h} \quad (1)$$

where G is the average gradient per pixel, N_w and N_h are the number of pixels in the horizontal and vertical dimensions, respectively, and I is the luminance image.

2.2. Single-Resolution R-Q Model

It was found in [8] that the bitrate for an intra coded sequence can be predicted using the following formula:

$$R_{pred}(Q_{step}) = G.\alpha.Q_{step}^{\beta} \quad (2)$$



(a) $\alpha = 0.96$, $\beta = -1.04$ (b) $\alpha = 0.49$, $\beta = -1.04$

Fig. 1. Accurate curve fitting using fixed value of β. Average bits per pixel over Q_{step} size for sequences (a) *ParkJoy* and (b) *Tennis*

where R_{pred} is the predicted bitrate normalised to the average Bits per Pixel (BPP), $\alpha > 0$ and $\beta < 0$ are parameters that depend on content and Q_{step} is the quantisation step size which has the following relationship:

$$Q_{step} = 2^{\frac{QP-4}{6}} \quad (3)$$

For any given frame, the optimal values for parameters α and β can be found by solving:

$$[\alpha_{opt}, \beta_{opt}] = argmin \sum_{i=0}^{D-1}(R_{actual,i} - R_{pred,i})^2 \quad (4)$$

where D is the total number of data points and R_{actual} contains all actual coded data points.

After solving (4) for a variety of sequences we found that in each case the value of β is fairly consistent and only α is subject to change. The parameter β can therefore be fixed and only α need be adjusted to produce a minimum close to the result produced in equation (4), as shown in Fig. 1.

In [7, 8, 10] an update procedure is performed to account for the changes in video content – QP selection is dependent on α, as indicated in (2), which is determined from a weighting of values calculated from previous frames. The update procedure is given as:

$$\alpha_{k+1} = \lambda.\alpha_k + (1 - \lambda).\frac{R_{actual,k}}{G_k.Q_{step}{}_k^{\beta}} \quad (5)$$

where λ is a forgetting factor which is commonly stated to have a typical value of 0.5 [8, 10] and k is the frame index.

When calculating α for the next frame, the forgetting factor determines how much weight is given to the value generated from the current frame over the value generated from the previous frame. A factor of 1 results in infinite memory and therefore the parameter is never updated – the initial value of α is applied to all future frames. Alternatively, a factor of 0 applies the value calculated from the current frame only when estimating the optimal value for the preceding Q_{step} size.

3. PROPOSED ADAPTIVE RESOLUTION R-Q MODEL

Contrary to previous works on rate control for intra coding, which provide a single resolution solution, our method adapts both the QP and spatial resolution to produce a coded bitstream that improves rate-distortion performance and also provides better matching of the target bitrate. For the multi-resolution rate control problem, each frame has multiple solutions; a frame may be encoded at the original HR or downsampled to a lower spatial resolution prior to encoding – frames coded at a lower spatial resolution are then upsampled after coding. Our method has two stages: the lowest QP possible at each resolution is first predicted for a given bit budget and then the combination that minimises distortion is determined.

3.1. Initial QP Selection

To determine the QP for the first frame we use prior calculations to generate a set of linear models. We know from [18] that the correlation between the average gradient per pixel and the actual number of coded bits is high. Using data from a range of different video sequences, a set of linear models for each QP can be produced to provide an initial prediction of the optimal QP given the target BPP and the measured complexity of the frame calculated in (1). Note that these models are only generated once and then applied to all future coded sequences.

3.2. Modified Updating Procedure

The performance of the update procedure given in (5) relies heavily on the correlation between successive frames. Better results can be achieved by applying a weighting of parameters calculated from frames with similar complexities. The rate control algorithm can learn from previously coded frames and make more accurate predictions. We therefore propose a modified updating procedure:

$$\alpha_{k+1} = (\lambda.\alpha_k + (1-\lambda).\frac{R_{actual,k}}{G_k.Q_{step_k}^{\ \beta}}).(1-\tau) + \tau.\alpha_G \quad (6)$$

where α_G is a weighted value determined from the complexities of all previous frames (7) that satisfy certain conditions and τ is selected based on the availability of frames with similar complexities stored.

We use a normal distribution to determine the weights of previously calculated parameters:

$$\alpha_G = \frac{1}{\omega}\sum_{i=0}^{K}\mathcal{N}(|G_{k+1} - G_i|).\alpha_i \quad (7)$$

where ω is the normalising factor equal to the sum of the weights and K is the number of useful previously stored parameters.

Complexity and α parameter pairs are only stored for future calculation if the result from the corresponding coded frame satisfies the conditions: $R_{actual} \leq R_{target}$ and $M \leq \gamma$ where R_{target} is the target number of BPP, M is the mismatch ratio given in (8) and γ is a threshold.

$$M = \frac{|R_{target} - R_{actual}|}{R_{target}} \times 100\% \quad (8)$$

3.3. Adaptive Resolution

Predicting the optimal QP for each spatial variation of the frame can be achieved in much the same way as the single resolution method as described in Section 2.2. We also apply our proposed modifications as given in Sections 3.1 and 3.2. Excluding the first frame, Q_{step} can be calculated by:

$$Q_{step} = e^{\frac{ln(\frac{R_{target}}{G.\alpha})}{\beta}} \quad (9)$$

When selecting the best resolution and QP combination the combined resampling and coding distortions need to be considered. Similar work on predicting the optimal scale factor has been carried out by Dong et al. [4, 5] and they show that the combined Mean Squared Error (MSE) distortion can be estimated from the addition of both the resampling and coding distortions. It is also suggested that the resampling distortion can be estimated by applying a simple box filter in the frequency domain. Estimating the resampling distortion alleviates some computational complexity, however, coding distortion is still calculated from actual results as the rate control algorithm is required to code at each resolution. After calculating the combined distortions, the combination that produces the best result while satisfying the condition $R_{actual} \leq R_{target}$ is selected. If none of the combinations meet this requirement, we select the resolution that provides the smallest mismatch ratio.

4. EXPERIMENTAL RESULTS

As in [8, 10], we partly evaluate the performance of each method by analysing the average mismatch ratio, given as:

$$\overline{M} = \frac{1}{N}\sum_{i}^{N}\frac{|R_{target} - R_{actual,i}|}{R_{target}} \times 100\% \quad (10)$$

where N is the total number of frames within the sequence

A total of 6 spatial resolutions were selected based on the conditions that the original aspect ratio remains the same and

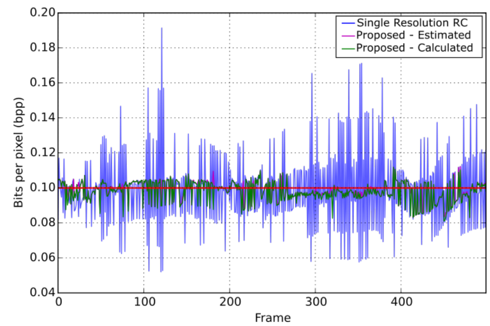
Video	Target kb/s	Single RC			Proposed - Calculated			Proposed - Estimated		
		$\overline{M}$-Ratio%	Rate	Y-PSNR	$\overline{M}$-Ratio%	Rate	Y-PSNR	$\overline{M}$-Ratio%	Rate	Y-PSNR
BlueSky	829.44	9.212	756.86	38.22	**5.074**	792.12	**38.52**	5.096	792.12	**38.52**
	414.72	10.792	371.17	33.82	**5.191**	396.06	**34.65**	5.797	393.98	34.63
	207.36	11.292	184.55	29.94	**5.673**	196.99	**30.98**	6.140	194.92	30.97
InToTree	1036.80	10.305	995.33	**35.89**	4.822	1032.65	35.73	**4.677**	1034.73	35.73
	622.08	8.289	603.42	**34.28**	3.899	624.15	34.12	**3.473**	626.23	34.09
	207.36	11.788	207.36	31.45	**4.722**	207.36	**31.47**	5.037	207.36	31.29
Station	207.36	8.225	201.14	**36.00**	6.122	196.99	35.95	**5.365**	199.06	35.94
	103.68	8.588	101.61	33.74	**6.191**	99.53	**33.76**	6.446	99.53	33.74
	20.74	13.56	20.74	29.53	**7.437**	20.74	**29.72**	9.098	18.66	29.02
ParkJoy	622.08	8.113	584.76	27.41	4.890	599.27	**28.88**	**4.869**	603.42	28.85
	207.36	16.825	205.29	24.03	**4.391**	205.29	**24.09**	4.788	203.21	24.08
	103.68	24.713	107.83	22.62	4.483	105.75	**22.67**	**4.448**	103.68	22.66

Table 1. Comparison of intra based Rate Control (RC) algorithms with bit-rates representing 1 frame per sec

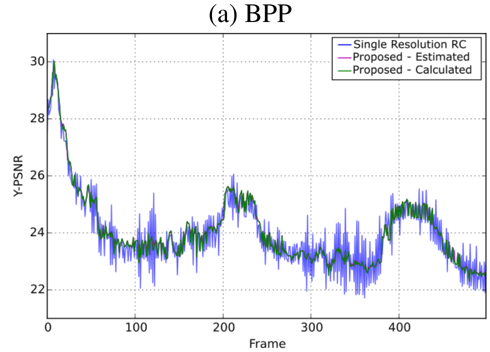
that the width and height are multiples of the smallest coding unit size in HEVC – which is 8×8. Given these criteria, the tested spatial resolutions were: 640×360, 768×432, 896×504, 1152×648, 1280×720, 1920×1080. All resampling was performed using Bicubic, although better methods can be applied to improve rate-distortion performance. The four tested video sequences were obtained from https://media.xiph.org/video/derf/.

For the updating procedure given in (5) and (7), λ was set to 0.1 so to give more weight to the value calculated from the current frame as this was shown to produce better results. For our modified procedure, γ was set to 10 so that any coded frames with a mismatch ratio less than 10% and $R_{actual} \leq R_{target}$ are considered to have a near optimal value of α.

Fig. 2 shows a comparison of our proposed method, using both estimated and calculated distortions, alongside the standard single resolution rate control approach using the update procedure given in (5). As described in Section 3.3, the *estimated* approach uses a simple box filter in the frequency domain to approximate the resampling distortion which is then added to the calculated coding distortion. The *calculated* approach uses the actual measured distortions from the reconstructed HR versions of the frame. Table 1 includes additional comparisons between these three methods and it is shown that our proposed method significantly outperforms the standard single resolution approach by providing a much lower average mismatch ratio. It should also be noted that, due to the fluctuations in the video bitrate in the standard single resolution case, a high number of frames are likely to be skipped resulting in a much lower average Peak Signal-to-Noise Ratio (PSNR) than those stated in Table 1. This also causes variance of video quality over time, as indicated in Fig. 2 (b). Furthermore, results show that estimating the distortion is effective and is a viable solution if computational power is limited.



(a) BPP



(b) Y-PSNR

Fig. 2. Reduced variation of rate and quality using proposed method – *ParkJoy* coded using three different intra based rate control methods. Target: 0.1 BPP / 207.36 kb/s (1 fps)

5. CONCLUSIONS

In this paper we demonstrated that our proposed adaptive resolution rate control method outperforms the single resolution approach by generating a coded bitstream with a far better regulated bitrate. We also provide further evidence to support our previous claim [1] that spatial resampling of intra coded frames can provide a higher and more consistent level of video quality at low bitrates.

6. REFERENCES

[1] B. Hosking, D. Agrafiotis, D. Bull, and N. Easton, "Spatial resampling of IDR frames for low bitrate video coding with HEVC," *IS&T/SPIE Electronic Imaging*, Feb. 2015.

[2] A. M. Bruckstein, M. Elad, and R Kimmel, "Downscaling for better transform compression," *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 1132–1144, Sept. 2003.

[3] X. Wu, X. Zhang, and X. Wang, "Low bit-rate image compression via adaptive down-sampling and constrained least squares upconversion," *IEEE Transactions on Image Processing*, vol. 18, no. 3, pp. 552–561, Mar. 2009.

[4] J. Dong and Y. Ye, "Adaptive downsampling for high-definition video coding," *IEEE Transactions on Image Processing*, Sept. 2012.

[5] 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, Mar. 2014.

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

[7] X. Jing and L. Chau, "A novel intra-rate estimation method for H.264 rate control," *IEEE International Symposium on Circuits and Systems*, July 2006.

[8] X. Jing, L. Chau, and W. Siu, "Frame complexity-based rate-quantization model for H.264/AVC intraframe rate control," *IEEE Signal Processing Letters*, vol. 15, pp. 373–376, Mar. 2008.

[9] Y. Zhou, L. Tian, and X. Ning, "Intra frame constant rate control scheme for high efficiency video coding," *International Conference on Computing, Networking and Communications (ICNC)*, pp. 648–652, Jan. 2013.

[10] L. Sun, O. C. Au, W. Dai, Y. Guo, and R. Zou, "An adaptive frame complexity based rate quantization model for intra-frame rate control of High Efficiency Video Coding (HEVC)," *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1–6, Dec. 2012.

[11] L. Tian, Y. Zhou, and X. Cao, "A new rate-complexity-QP algorithm (RCQA) for HEVC intra-picture rate control," *International Conference on Computing, Networking and Communications (ICNC)*, pp. 375–380, Feb. 2014.

[12] S. Hu, H. Wang, S. Kwong, and T. Zhao, "Frame level rate control for H.264/AVC with novel rate-quantization model," *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 226–231, July 2010.

[13] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[14] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: scalable extensions of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, July 2015.

[15] X. Jing, J. Y. Tham, Y. Wang, K. H. Goh, and W. S. Lee, "Efficient rate-quantization model for frame level rate control in spatially scalable video coding," *IEEE International Conference on Networks (ICON)*, pp. 339–343, Dec. 2012.

[16] J. Liu, Y. Cho, Z. Guo, and C.J. Kuo, "Bit allocation for spatial scalability coding of H.264/SVC with dependent rate-distortion analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 7, pp. 967–981, July 2010.

[17] Z. Chen and K. N. Ngan, "Recent advances in rate control for video coding," *Signal Processing: Image Communication*, vol. 22, no. 1, pp. 19–38, Jan. 2007.

[18] W. J. Kim, J. W. Yi, and S. D. Kim, "A bit allocation method based on picture activity for still image coding.," *IEEE Trans. on Image Processing*, vol. 8, no. 7, pp. 974–977, July 1999.