LCU-LEVEL BIT ALLOCATION FOR RATE CONTROL IN HIGH EFFICIENCY VIDEO CODING

Heng Sun, Can Zhang and Shaoshuai Gao

School of Electric, Electrical and Communication Engineering
University of Chinese Academy of Sciences

ABSTRACT

This paper presents an adaptive bit allocation algorithm for rate control in High Efficiency Video Coding (HEVC). Video texture complexity reflects the substantial of video content and it is a valuable tool for bit prediction. We discover that, there is always a quasilinear relationship between the coding bit rate and the texture complexity at the LCU level. This linearity leads to a novel and unified rate model for different types of source data. Meanwhile, a simple but accurate model is proposed to estimate the texture complexity by considering the spatial-temporal correlations. Experiments have been carried out to compare the coding performance of our proposed algorithm with existing method. Results demonstrate that the proposed algorithm achieves better rate-distortion performance with a much lower and steadier buffer level.

***Index Terms*—** Bit allocation, LCU level, Texture complexity, Rate control, HEVC

1. INTRODUCTION

The latest video coding standard, called High Efficiency Video Coding (HEVC) [1], aims at achieving the coding efficiency improvement of about 50% or more compared to H.264/AVC [2]. Many new coding tools are adopted by HEVC to improve the coding efficiency, such as a more adaptive quadtree structure.

In real-time video communication applications, video data is transmitted over a bandwidth-limited channel. Rate control plays a key role for ensuring effective channel adaption during video delivery. Its goal is to regulate the coded bit stream to satisfy the available bandwidth while achieving the best video quality. A typical rate control scheme consists of two basic operations: 1) bit allocation, i.e., how to efficiently distribute the available bandwidth among coding units, and 2) bit-rate control, i.e., how to properly adjust the encoder parameters to encode each basic unit with the allocated bits.

Performing a good rate control requires both an effective bit allocation and an accurate bit-rate control. Algorithms used to perform bit allocation can be roughly classified into buffer-status-based approaches [3] [4] and texture-complexity-based approaches [5] [6]. These two approaches can also be jointly applied to perform bit allocation [7]. With the available bandwidth, many rate control algorithms [8] [9] employ the mean absolute difference (MAD) as an indication of coding complexity to perform bit allocation. Since the residual signal is not available until the rate-distortion optimization (RDO) process, the MAD of each basic unit in the current frame is predicted by using the MAD of the collocated basic unit in the previous frame. In general, the prediction of MAD is carried out using a linear model that is updated after the encoding of each basic unit through linear regression. The linear model attempts to approximate the temporal variation of coding complexity and the prediction performance is weakened by the poor use of the spatial-temporal correlations. Liu *et al*. [10] proposed a switched MAD prediction model to achieve more accurate MAD predictions at both the frame and macroblock layer. Dong and Ling pointed out in [11] that MAD and model parameters estimated dynamically are often inaccurate and inaccurately estimated model parameters lead to large oscillations of future estimated MAD and model parameters. In [11], two contexts were constructed: one for the current basic unit and the other for the collocated basic unit in the previous frame. However, for high-motion videos, the temporal correlations among the basic units within these two contexts are not strong and the context for the collocated basic unit should be replaced by the context for motion-predicted basic unit in the previous frame. Besides, since each basic unit used during linear regression can be coded with different modes, predicting MAD can be both unwidely and inaccurate. Other works [12] [13] use the sum of absolute transform differences (SATD) instead of MAD because SATD has better performance in rate control.

This paper proposes a new LCU-level bit allocation algorithm based on texture complexity. In our scheme, we propose a simple but accurate algorithm to estimate texture complexity by taking account the spatial-temporal correlations. For each basic unit, the number of coding bits is modeled as a linear function of texture complexity. We implement the proposed scheme to verify its performance, and it is shown that the proposed algorithm based rate control scheme obtains better rate-distortion (RD) performance and a steadier and lower

This research was partially supported by the State Key Program of National Natural Science Foundation of China (Grant No. 61032006).

buffer level than that of HM12.1 which is the state-of-art rate control scheme.

2. RATE CONTROL OVERVIEW

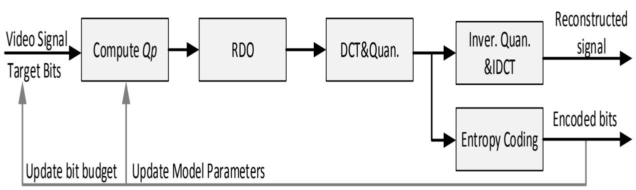


Fig. 1. Structure of the rate control scheme based on the proposed bit allocation algorithm.

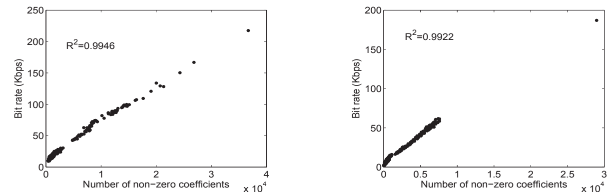
The basic structure of our bit allocation algorithm based rate control scheme is shown in Fig. 1. The first step of the rate control is bit allocation with the available bandwidth. The target bits for the current coding frame are a weighted combination of the frame target bits estimated from buffer status and those from the bits remaining for encoding the video sequence taking into the consideration of the frame complexity. The complexity of each frame is estimated as the average complexity per basic unit for the frame. For each basic unit, the computation of texture complexity is carried out using the average gradient per pixel factoring into the consideration of the spatial-temporal correlations. Based on the texture complexity, the current frame target bits are allocated to each basic unit. The basic idea for bit allocation is to allocate more bits to the high complexity basic units and less bits to the low complexity ones in order to achieve a good visual quality with small fluctuation. Once the target number of bits is determined, the next step is to determine the appropriate quantization parameter denoted as Qp to meet the target bit allocation. Then, Qp can be applied to perform motion estimation with RDO for the current basic unit. After encoding each basic unit, the model parameters are updated using a linear regression method.

3. THE PROPOSED BIT ALLOCATION ALGORITHM

It has been proved in [4] that the rate function in the ρ-domain is approximately a linear function that has the following expression:

$$R = \alpha \cdot (1-\rho) = \beta \cdot N_{\text{non\_zero}} \tag{1}$$

where R and ρ are the bit rate of the coding frame and the percentage of zeros among the quantized transform coefficients, respectively. α and β are two constants and their relation can be represented as $\alpha = \beta \cdot N_{\text{pixel}}$. $N_{\text{non\_zero}}$ and N_{pixel} are the number of non-zero quantized transform coefficients and the number of pixels in a frame. As Fig. 2 shows, there are two



(a) RaceHorsesC (b) BQMall

Fig. 2. The linear relation between R and $N_{\text{non\_zero}}$.

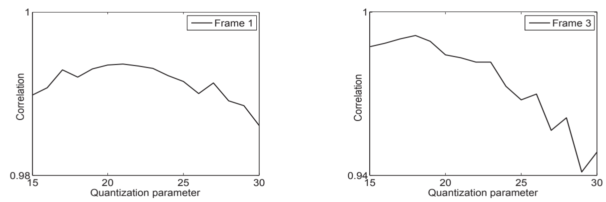


Fig. 3. The correlation coefficient of each LCU between R_{LCU} and $N_{\text{non\_zero,LCU}}$.

different sequences in the figure, and every dot in the figure represents a frame from a sequence in HEVC. Every dot denotes the relation between R and $N_{\text{non\_zero}}$ for that frame. Obviously, although they are distributed in different areas in the figure, the dots form a quasilinear relation.

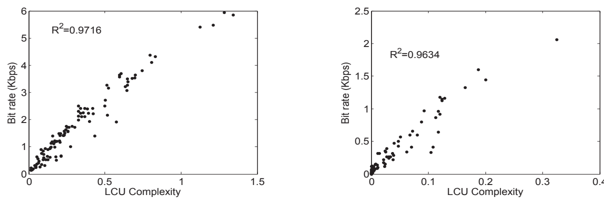
We run the HEVC codec on the *BQMall* sequence at different quantization parameters and generate several points $\{N_{\text{non\_zero,LCU}}, R_{\text{LCU}}\}$. $N_{\text{non\_zero,LCU}}$ and R_{LCU} are the number of non-zero quantized transform coefficients and the coding bit rate for the coding LCU. Let $\mathcal{C}(N_{\text{non\_zero,LCU}}, R_{\text{LCU}})$ be the correlation between $N_{\text{non\_zero,LCU}}$ and R_{LCU}. In Fig. 3, we plot the average value of $\mathcal{C}(N_{\text{non\_zero,LCU}}, R_{\text{LCU}})$ for all LCUs in a frame. It can be seen that $\mathcal{C}(N_{\text{non\_zero,LCU}}, R_{\text{LCU}})$ is very close to 1, which implies that there is also a linear relationship between $N_{\text{non\_zero,LCU}}$ and R_{LCU} at the LCU level.

However, it is difficult to map ρ to quantization parameters. The number of non-zero quantized transform coefficients indicates that for a coding unit, the higher the texture complexity is, the more bits will be generated. In other words, the coding bit rate is proportional to the texture complexity. As we know, gradient can be used to estimate texture complexity in image processing. In order to measure the texture complexity of each basic unit, the average gradient G_{LCU} per pixel for each basic unit is introduced, which is calculated as follows:

$$G_{\text{LCU}} = G_{\text{LCU, H}} + G_{\text{LCU,V}} \tag{2}$$

where $G_{\text{LCU,H}}$ and $G_{\text{LCU,V}}$ are the average horizontal gradient and the average vertical gradient per pixel for each basic unit. $G_{\text{LCU,H}}$ and $G_{\text{LCU,V}}$ can be determined as follows:

$$G_{\text{LCU,H}} = \frac{1}{W \cdot H} \sum_{w=1}^{W-1} \sum_{h=1}^{H} |I(w,h) - I(w+1,h)| \tag{3}$$



(a) Frame 1 (b) Frame 5

Fig. 4. The linear relation between the texture complexity and the coding bit rate for sequence *BQMall*.

and

$$G_{\text{LCU,V}} = \frac{1}{W \cdot H} \sum_{w=1}^{W} \sum_{h=1}^{H-1} |I(w,h) - I(w,h+1)| \tag{4}$$

where W and H are the width and height of the current block, respectively. $I(w,h)$ is either the (w,h)th luminance value in the current LCU considering the spatial correlation or the (w,h)th value in the residue block of the current block and the reference block considering the temporal correlation. The reference block is selected from the collocated blocks in the reference frames, which is able to provide the biggest matching degree MD measured by

$$MD = \frac{1}{N_{\text{pixel,LCU}}} \sum_{i=1}^{N_{\text{hist}}} \min\{H_{\text{curr}}(i), H_{\text{col}}(i)\} \tag{5}$$

where H_{curr} and H_{col} are the luminance histograms of the current block and the collocated blocks in the reference frames. N_{hist} is the number of bins of the histogram and $N_{\text{pixel,LCU}}$ is the number of pixels in the current block [14].

Therefore, by considering the spatial-temporal correlations, the texture complexity C_{LCU} of the current LCU can be calculated by:

$$C_{\text{LCU}} = \min\{G_{\text{LCU,spatial}}, G_{\text{LCU,temporal}}\} \tag{6}$$

where $G_{\text{LCU,spatial}}$ and $G_{\text{LCU,temporal}}$ are calculated by (2) considering the spatial and temporal correlations, respectively.

With the texture complexity measurement in (6), from Fig. 4, it can be observed that for each basic unit, there is a quasilinear relation between the texture complexity and the coding bit rate. In other words, our texture complexity measurement algorithm is quite effective and it is able to achieve an accurate predicted R_{LCU}. Therefore, R_{LCU} can expressed as follows:

$$R_{\text{LCU}} = \sigma \cdot C_{\text{LCU}} \tag{7}$$

where σ is a constant.

As we mentioned above, when allocating bits from a fixed bit budget to each basic unit in a video frame, the relative complexity of each basic unit has to be considered for good coding performance. Based on (7), the number of target bits R^i_{LCU} for the ith basic unit is given by

$$R^i_{\text{LCU}} = R_{\text{left}} \cdot \frac{C^i_{\text{LCU}}}{\sum_{k=i}^{N} C^k_{\text{LCU}}} \tag{8}$$

where R_{left} is the remaining number of bits in the frame bit budget left for coding the ith and other uncoded basic units in the frame. C^i_{LCU} is the texture complexity of the ith basic unit. N is the total number of basic units in the frame. It can also be observed that no much computation is introduced by the proposed algorithm. The most complex part in the algorithm is the computation on $G_{\text{LCU,spatial}}$ and $G_{\text{LCU,temporal}}$ which total cost $W_f \cdot H_f \cdot N_{\text{ref}} + 2 \cdot \{(W_f - 1) \cdot H_f + W_f \cdot (H_f - 1)\}$ additions and subtractions for one single frame where W_f and H_f are the frame width and height, N_{ref} is the reference frame number for each frame. Simulations on Intel Core 2 2.93G PC show that, on average, the proposed algorithm takes less than 0.01% coding time which can be negligible in real-time applications.

For convenience, the whole proposed bit allocation based rate control scheme is summarized in Algorithm 1.

Algorithm 1 Encode one Fame with the proposed algorithm

Calculate the frame target bits R as the way of HM12.1.
Estimate the texture complexity of each LCU in the current frame by (2), (3), (4), (5) and (6).
Estimate the frame complexity C as the average complexity per LCU for the current frame.
Modify the frame target bits as $R' = R + (R_{\text{prev}} - R_{\text{prev,actual}}) \cdot \frac{C}{C_{\text{prev}}}$, where R_{prev} and $R_{\text{prev,actual}}$ are the target bits and the actual coding bits for the previous frame. C_{prev} is the texture complexity of the previous frame.
Use R' as the target bits for the current frame. Let $i = 1$.
while $i \le N$ **do**
    Calculate the target bits of the ith LCU by (8).
    Calculate the quantization parameter Qp as the way of HM12.1.
    Encode the current LCU with Qp.
    Update the remaining target bits R_{left} by subtracting the total encoded bits of LCU i from it.
    Let $i = i + 1$
end while
Finish the encoding process of the current frame.

4. EXPERIMENTAL RESULTS

To evaluate the performance of our proposed algorithm, we carry out a number of experiments using HM12.1. In all the experiments, the performance comparisons are carried out between the proposed algorithm based rate control and the rate

Table 1. SIMULATION RESULTS OF OUR PROPOSED ALGORITHM IN TERMS OF PSNR (dB) AND BIT RATES (kb/s)

Sequences	Target Bit Rates R_T (kbps)	Frame	Average PSNR (dB)			Bit Rates R_A (kbps)		$\frac{\lvert R_T - R_A \rvert}{R_T} \cdot 100\%$	
			HM12.1	Ours	Gain	HM12.1	Ours	HM12.1	Ours
BasketballDrill	512	500	32.38	33.03	+0.65	512.01	512.04	0.00%	0.01%
BQMall	512	600	32.52	33.17	+0.65	513.89	512.12	0.37%	0.02%
PartyScene	384	500	26.47	27.37	+0.90	387.55	384.09	0.92%	0.02%
RaceHorsesC	384	300	29.66	29.88	+0.22	383.98	384.01	0.01%	0.00%
BQSquare	256	600	30.69	32.00	+1.31	258.54	256.10	0.99%	0.04%
BlowingBubbles	256	500	30.21	31.03	+0.82	258.87	255.83	1.12%	0.07%
RaceHorses	256	300	31.77	32.00	+0.23	255.51	255.38	0.19%	0.24%
ChinaSpeed	512	500	30.78	31.77	+0.99	513.15	511.41	0.22%	0.12%

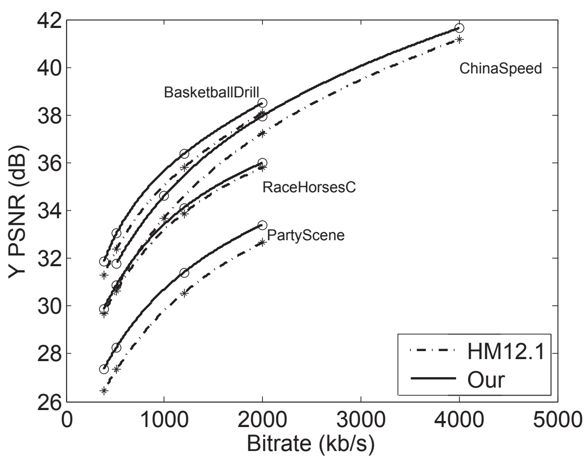


Fig. 5. R-D curves for *BasketballDrill*, *RaceHorsesC*, *ChinaSpeed* and *PartyScene*.

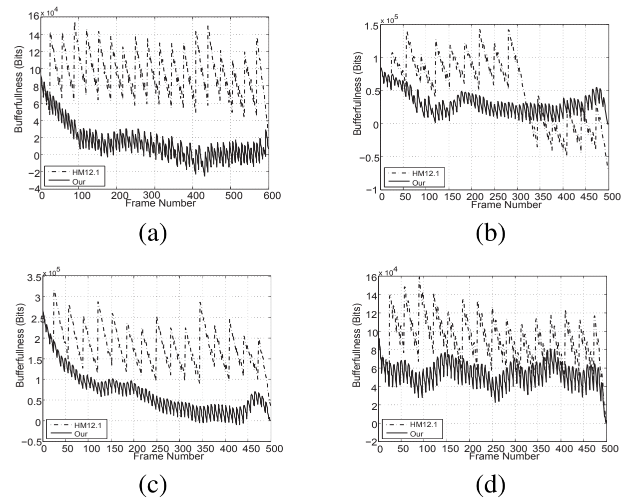


Fig. 6. Buffer occupancy. (a) *BQSquare*, (b) *BlowingBubbles*, (c) *PartyScene* and (d) *BasketballDrill*.

control scheme in HM12.1 [9] which is the state-of-art rate control scheme. The motion search range is 64 and the GOP size is set to 8. The RDO is enabled and hierarchical B picture structures are used. The LCU width and height are both set to 64. Table 1 summaries the simulation results. On average, 0.72 dB gain in PSNR and less bit-rate mismatching are achieved by the proposed algorithm. From the R-D curves shown in Fig. 5, it can be observed that the proposed adaptive bit allocation algorithm can achieve better R-D performance. Fig. 6 shows the buffer level variations. It is found that the actual buffer level using our scheme is kept steadier and closer to 0. This implies that our scheme produces a steadier buffer status. It is also noted that underflow occurs when HM12.1 used, as in Fig. 6(b). Although underflow does not affect motion continuity, it wastes channel bandwidth.

5. CONCLUSION

In this paper, an adaptive bit allocation based rate control scheme is presented. By considering the spatial-temporal correlations, accurate texture complexity and rate models are first developed. Then an adaptive bit allocation scheme is derived from these models. Comprehensive simulations verify that, compared with the state-of-art rate control scheme, the proposed bit allocation scheme based rate control shows a much better coding performance with a smaller mismatching between the target bits and the actual coding bits.

6. REFERENCES

[1] G.J. Sullivan, J. Ohm, Woo-Jin Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[2] J. Ohm, G.J. Sullivan, H. Schwarz, Thiow Keng Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards–Including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst.*

Video Technol., vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[3] S. Ma, W. Gao, and Y. Lu, “Rate-Distortion Analysis for H.264/AVC Video Coding and its Application to Rate Control,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[4] Z. He and S.K. Mitra, “Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρ-Domain Source Modeling,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840–849, Oct. 2002.

[5] J. Ribas-Corbera and Shaw-Min Lei, “A Frame-Layer Bit Allocation for H.263+,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 7, pp. 1154–1158, Oct. 2000.

[6] Y. H. Tan, C. Yeo, and Z. Li, “Single-Pass Rate Control With Texture and Non-Texture Rate-Distortion Models,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 8, pp. 1236–1245, Aug. 2012.

[7] M. Jiang and Nam Ling, “On Lagrange Multiplier and Quantizer Adjustment for H.264 Frame-Layer Video Rate Control,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 5, pp. 663–669, May 2006.

[8] H. Choi, J. Nam, J. Yoo, D. Sim, and IV Bajić, “Rate control based on unified RQ model for HEVC,” in *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 8th Meeting: San José, CA, USA*, 2012, pp. 1–10.

[9] B. Li, “Rate control by R-lambda model for HEVC,” in *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T 1/SC 29/WG11 11th Meeting: Shanghai, CN*, 2012, pp. 10–19.

[10] Y. Liu, Z.G. Li, and Y. C. Soh, “A Novel Rate Control Scheme for Low Delay Video Communication of H.264/AVC Standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, pp. 68–78, Jan. 2007.

[11] J. Dong and N. Ling, “A Context-Adaptive Prediction Scheme for Parameter Estimation in H.264/AVC Macroblock Layer Rate Control,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp. 1108–1117, Aug. 2009.

[12] Do-Kyoung Kwon, Mei-Yin Shen, and C.-C.J. Kuo, “Rate Control for H.264 Video With Enhanced Rate and Distortion Models,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 5, pp. 517–529, May 2007.

[13] S. Ma, J. Si, and S. Wang, “A study on the rate distortion modeling for high efficiency video coding,” in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 181–184.

[14] M. J. Swain and D. H Ballard, “Color indexing,” *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.