https://doi.org/10.2352/EI.2022.34.10.IPAS-365
© 2022, Society for Imaging Science and Technology

A frame level rate allocation algorithm based on temporal dependency model for AV1

Cheng Chen, Jingning Han, Paul Wilkins, and Yaowu Xu
WebM Codec Team, Google Inc.
1600 Amphitheatre Parkway, Mountain View, CA 94043
Emails: {chengchen, jingning, paulwilkins, yaowu}@google.com

***Abstract*—Rate control is an essential module in video coding. Rate control strategies strive to deliver a stable playback experience as well as achieving high compression efficiency for modern video applications, constrained by restricted bandwidth and buffer limits. The difficulty of rate control often lies in the adaptation ability of the underlying algorithm to capture the variability of content and temporal correlation across frames. In this paper, we present a rate allocation algorithm to model the distortion propagation in the hierarchical coding structure premised on the temporal dependency model at frame level. Our experiments show that with the information collected from the temporal dependency model, the proposed rate allocation algorithm significantly improves the coding efficiency over the AV1 baseline on a set of variable user generated video clips.**

***Index Terms*—rate control, temporal dependency model, AV1**

I. Introduction

Rate control is an essential module in video coding. It strives to deliver a stable playback experience for modern video applications, constrained by restricted bandwidth and buffer limits. The objective of rate control is often to maximize the usage of available bandwidth as well as delivering videos of the highest quality. Rate control is also a crucial research topic for practical encoders to improve compression efficiency. In practice, rate control strategies are designed to match the use case of video applications. For example, for video on demand (VOD), where out-of-order coding is allowed and a hierarchical coding structure is adopted, rate control strategies take advantage of the fact that frames at low hierarchical layers are typically referenced more often and encodes them with higher quality accordingly. However, how to quantitatively distribute bit rates to achieve the optimal compression efficiency is challenging due to the variability of content, and temporal correlations across video frames.

Given a target bit rate, a practical rate controller may first determines rate allocation, followed by a rate-distortion modelling, which aims to select an appropriate quantization parameter (QP) to approximate the allocated bits. Rate-distortion modelling has been well studied in literature. The R-Q model builds a quadratic relation between the rate R and QP [1], based on the assumption that transform coefficients follow a Laplacian distribution. Linear regression analysis is used to estimate model parameters. The R-λ model premises on a hyperbolic relation between rate and distortion and achieves a high rate control accuracy in HEVC [2]. A number of research works of rate control have emerged based on the R-λ model, where the rate allocation problem draws the most interest. Both picture-level and basic-unit-level bit allocation algorithms [3] are proposed as the extension of [2]. A frame level bit allocation for HEVC low delay is presented in [4]. A bit allocation algorithm at the largest coding unit level is proposed with recursive Taylor expansion in [5].

Among these bit allocation algorithms, the ability to precisely model the inter-frame dependency is a major differentiator. For video applications that allows the encoder to collect look-ahead information, a two-pass encoding scheme empowers richer information and a better approximation of temporal correlation. [6] refines the R-λ model by pre-encoding with 16x16 coding units to estimate the characteristics of the largest coding unit. [7] improves rate control of x265 using information from quarter-resolution motion estimation.

In this work, we propose a frame level rate allocation method based on a temporal dependency model [8]. We provide a rate-distortion analysis based on the R-λ model and approximate the inter-frame correlations with stats collected from the temporal dependency model. Experiments show that the proposed method achieves a significant coding gain against the baseline rate allocation of AV1.

II. Algorithm

A. Rate Control Problem

The rate control problem in video coding can be formed as a constrained optimization problem

$$\min_{\pi} \ D, \quad s.t. \ R \leq R_T, \tag{1}$$

where D, R, R_T represents the total distortion, the total bit rate, and the target bit rate, respectively. π represents the control policy and parameters that minimize the target function. A Lagrangian multiplier λ is introduced to convert it to a unconstrained optimization problem

$$\min_{\pi} \ D + \lambda R. \tag{2}$$

The optimal λ is the negative slope of the tangent line to the R-D curve,

$$\lambda = -\frac{\partial D}{\partial R}, \tag{3}$$

assuming the R-D curve is convex and differentiable [9].

In practice, the rate control problem often breaks down to two major parts: rate allocation and QP selection (Fig. 1).

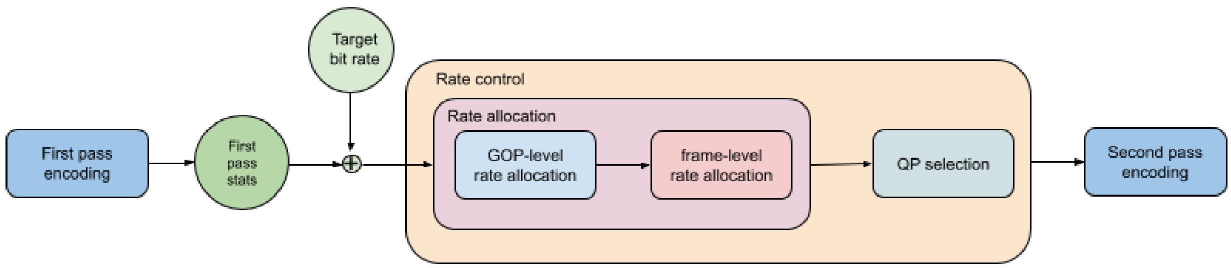

Fig. 1. The two-pass encoding workflow of AV1. The rate control module contains rate allocation and QP selection. The rate allocation consists of GOP level and frame level rate allocation.

Given the target bit rate, a practical rate controller first allocates bits to different frames, or blocks. A QP determination mechanism then selects the appropriate QP to approximate the allocated bits for each frame, or block. In this paper, we focus on the rate allocation.

B. AV1's Two-pass Rate Allocation Algorithm

AV1's rate allocation algorithm is based on two-pass encoding. In the first pass, a simplified and fast encoding run is executed with fixed partition, transform sizes, and limited coding tools. A set of parameters are collected per frame to aid the rate control decisions in the second pass.

As shown in Fig. 1, the rate allocation process is divided to two stages: group of picture (GOP) level and frame level rate allocation. First pass stats are used to determine the rate allocation in each stage. The complexity measure of a frame, denoted as M_i, $i = 1, 2, ..., N$, is the aggregation of block errors for frame i, where the block error is the minimum of motion compensated error and intra prediction error. N is the number of frames.

The allocated bit rate of the m-th GOP, R_{G_m}, is linearly proportional to the accumulated complexity measure of each frame with respect to the total complexity of the video clip,

$$R_{G_m} = \frac{\sum_{i \in \Omega_m} M_i}{\sum_{i=1}^{N} M_i} \cdot R_T, \tag{4}$$

where R_T is the target bit rate; Ω_m is the set of frames in the m-th GOP.

Within each GOP, the rate allocation of each frame is based on a heuristic measure. Taking into the consideration that frames at low hierarchy are referenced more often than those at high hierarchy, a boost score S_i of frame i is calculated to reflect its importance in the hierarchical structure. And the bit budget of frame i is then determined as

$$R_i = \frac{S_i}{\sum_{i \in \Omega_m} S_i} \cdot R_{G_m}. \tag{5}$$

The boost score S_i is the accumulated ratio of intra prediction error (P) over motion compensated error (Q),

$$S_i = \sum_{j=i-\Delta}^{i+\Delta} \frac{P_j}{Q_j}, \tag{6}$$

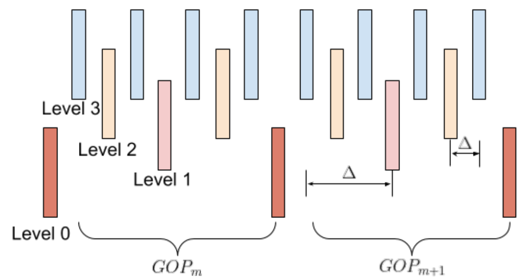

Fig. 2. An example of the hierarchical structure and GOPs.

in its neighbour frames $[i - \Delta, i + \Delta]$, where Δ is the size of the neighbourhood, which aims to approximate how far the current frame has influence on other frames. For example, as shown in Fig. 2, Δ is set as the distance from the current frame to the neighbour frame at the previous level minus one. Δ of frames at the top level is 0.

C. A New Frame-level Rate Allocation Algorithm

Within a GOP G_m, the optimization problem (Eqn. (2)) can be written as

$$\min_{\pi} \sum_{i \in \Omega_m} D_i + \lambda \sum_{i \in \Omega_m} R_i, \tag{7}$$

where λ is the global Lagrangian multiplier. To find the optimal rate allocation for each frame, we take derivative of Eqn (7),

$$\frac{\partial(\sum_{i \in \Omega_m} D_i + \lambda \sum_{i \in \Omega_m} R_i)}{\partial R_j} = 0, \tag{8}$$

for any $j \in \Omega_m$. It can then be further derived as

$$\frac{\partial \sum_{i \in \Omega_m} D_i}{\partial R_j} + \lambda \frac{\partial \sum_{i \in \Omega_m} R_i}{\partial R_j} = 0, \tag{9}$$

and

$$\frac{\partial \sum_{i \in \Omega_m} D_i}{\partial R_j} + \lambda = 0, \tag{10}$$

where we can safely assume R_j and R_k ($j \neq k$) are independent of each other. By multiplying $\frac{\partial R_j}{\partial D_j} = -\frac{1}{\lambda_j}$ on both sides, we get

$$\frac{\partial \sum_{i \in \Omega_m} D_i}{\partial R_j} \cdot \frac{\partial R_j}{\partial D_j} + \lambda \cdot \frac{\partial R_j}{\partial D_j} = 0, \tag{11}$$

$$\frac{\partial \sum_{i\in\Omega_m} D_i}{\partial D_j} - \frac{\lambda}{\lambda_j} = 0. \tag{12}$$

By denoting $\frac{\partial \sum_{i\in\Omega_m} D_i}{\partial D_j}$ as θ_j, we obtain

$$\lambda_j = \frac{\lambda}{\frac{\partial \sum_{i\in\Omega_m} D_i}{\partial D_j}} = \frac{\lambda}{\theta_j}. \tag{13}$$

θ_j measures the distortion propagation along with the reference relations in the hierarchical structure.

As shown in literature [2], [10], [11], the rate-distortion relation can be modelled as

$$D = CR^{-K}, \tag{14}$$

where C and K are parameters that relate to the characteristics of the video clip. The corresponding Lagrangian multiplier can be written as

$$\lambda = -\frac{\partial D}{\partial R} = C \cdot K \cdot R^{-(K+1)}. \tag{15}$$

Combining Eqn. (13) and (15) we get

$$R_j = \theta_j^{\frac{1}{K+1}} R. \tag{16}$$

Bit rates of all frames within the GOP G_m sum up to the target rate budget R_{G_m},

$$\sum_{j\in\Omega_m} R_j = R_{G_m}. \tag{17}$$

Finally, combining Eqn. (16) and (17), the rate allocation for individual frame is

$$R_j = \frac{\theta_j^{\frac{1}{K+1}}}{\sum_{j\in\Omega_m} \theta_j^{\frac{1}{K+1}}} R_{G_m}. \tag{18}$$

In this model, frame level rate allocation depends on the distortion propagation factor θ and parameter K. To verify that the relation between λ and R matches the hyperbolic model (Eqn. (15)) in AV1, we fit R-λ curves, as shown in Fig. 3. It captures the relation extremely well with correlation coefficient $r^2 > 0.995$. We can also notice that K varies significantly for different video clips. However, it is hard to get an accurate estimation without encoding the video. For this study, we universally set $K = 0.5$. The approximation of θ is based on the temporal dependency model as illustrated below.

D. Temporal Dependency Model

Inspired by prior research work of inter-frame dependency analysis [12], [13], the temporal dependency model in VP9 and AV1 [8], referred to as TPL hereinafter, captures the temporal correlation across frames by tracing block-based motion trajectories along the reference structure.

Similar to the first pass encoding, TPL runs a simplified encoding process for a GOP, conducting motion search and collecting information such as intra prediction error and motion compensated error. Nevertheless, the major difference is that TPL builds upon the full frame referencing system (Fig. 2) and fully reconstructs frames for referencing, while the first pass encoding only uses the last source frame as reference without transform, quantization, or reconstruction. As demonstrated in [8], the use of reconstructed frames as well as accounting for the quantization effect successfully captures temporal dependency information.

Algorithm 1: Frame-level rate allocation algorithm.

Input: target bit rate R_T, GOP set Ω, complexity measure M, K
Output: bit rate budget R_j for each frame j
Method:
foreach $m \in \Omega$ **do**
  Compute rate budget R_{G_m} according to Eqn. (4).
  foreach $j \in \Omega_m$ **do**
    Run TPL and collect stats: $intra\_cost_j$, $inter\_cost_j$, and $propagation\_cost_j$.
  foreach $j \in \Omega_m$ **do**
    Compute θ_j according to Eqn. (19).
  foreach $j \in \Omega_m$ **do**
    Compute R_j according to Eqn. (18).

In this work, we estimate the distortion propagation factor θ with TPL stats. Specifically, we calculate the intra prediction cost, inter prediction cost, and the propagation cost, which represents how much information the current frame carries for this GOP, denoted as $intra\_cost$, $inter\_cost$ and $propagation\_cost$ in [8], respectively. The estimation of the distortion propagation θ_j is formulated as

$$\theta_j = \frac{inter\_cost_j * propagation\_cost_j}{intra\_cost_j}. \tag{19}$$

The complete rate allocation algorithm is summarized in Algorithm. 1.

As compared to the baseline, TPL builds upon the full reference structure for the current GOP, and well captures the information flow and distortion propagation. While the baseline algorithm uses heuristic terms, for example Δ, to indirectly measure the information propagation.

III. Experimental results

The proposed frame level rate allocation algorithm was implemented in libaom [14], the reference software of AV1 [15]. We evaluated the compression performance of the proposed method against the baseline, which depends on the first pass stats, as described in section II-B.

In this work, we focus on the rate allocation. Other rate control constraints, for example the QP selection constraint as shown in Fig. 1, are relaxed in the experiment. Specifically, we allow the encoder to encode a frame multiple times in order to select the QP that best approximates the allocated rate for each frame.

The evaluation dataset is collected from YouTube's user generated content (UGC) [16], a large scale dataset intended for video compression and quality assessment research. Forty

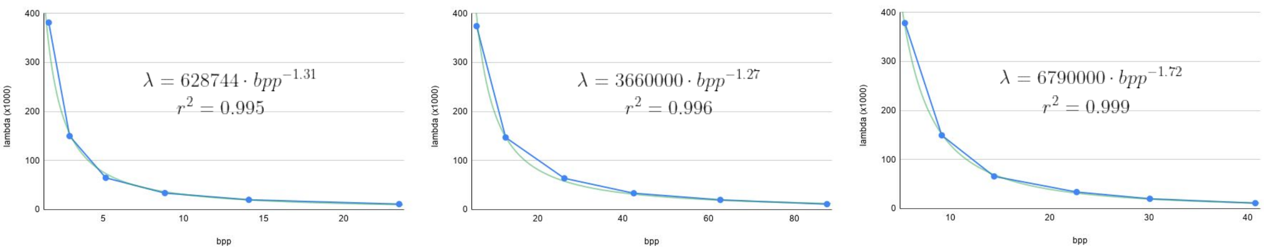

Fig. 3. Some examples of the hyperbolic model of λ and R (bit per pixel). Correlation coefficient r^2 indicates the model captures the correlation between λ and R very well.

360p videos are encoded at various target bit rates (from 50 kbps to 1000 kbps), which covers the typical PSNR range from 30 dB to 40 dB. These videos contain various types of content, from animation and lyrics video, to challenging content with camera motion, and scene changes, such as vlog, music video, etc.

The first 150 frames of each video was encoded at speed setting 2 (*–cpu-used=2*), using the variable bit rate mode (*–end-usage=vbr*). The compression efficiency improvement was measured in terms of BD rate reduction in PSNR and VMAF. A negative value indicates better coding performance.

The result is summarized in Table I. The averaged gain over the baseline is 6.15% and 6.54%, for PSNR and VMAF respectively. Whilst some clips or classes of content exhibited much larger gains, including animation, gaming, music videos and vlogs, in natural videos with slow motion, such as TV 1-5, the baseline rate allocation in AV1 works well. For animation videos and some gaming videos, even those that are relatively static, the temporal relationship between frames differs from natural videos and the proposed method is better to model the correlation. For clips with frequent scene changes (music video) and large camera motion (vlogs), the proposed method also captures temporal correlation better and leads to a better rate allocation.

It is worth noting that TPL has already been used in libaom to adjust QPs at superblock level [8]. The proposed algorithm takes advantage of TPL stats without introducing extra computation to the encoder. The encoding time difference with respect to the baseline is negligible for VOD applications.

TABLE I
CODING PERFORMANCE GAINS OF OUR METHOD OVER THE BASELINE.

Clip	PSNR (%)	VMAF (%)
Animation_1	-17.40	-12.62
Animation_2	-17.86	-19.42
Animation_3	-4.46	-10.52
Gameplay_1	-3.49	-4.99
Gameplay_2	-2.25	-2.06
Gameplay_3	-0.28	-2.12
Gameplay_4	-15.23	-17.79
Lecture_1	-4.64	-5.47
Lecture_2	4.15	3.83
Lecture_3	0.60	2.89
Lecture_4	-2.82	-1.72
Lecture_5	-0.47	-1.33
Lecture_6	0.14	-0.04
LiveMusic_1	-15.65	-15.03
LiveMusic_2	-3.65	-5.87
LiveMusic_3	-17.65	-15.35
Lyrics_1	2.65	2.92
Lyrics_2	-2.73	-4.56
Music_1	-3.15	-5.23
Music_2	-2.28	-2.60
Music_3	-1.67	-2.60
Music_4	-8.27	-13.06
Music_5	-29.03	-21.80
Music_6	-11.60	-13.31
Music_7	-9.76	-2.12
News_1	-3.26	-3.35
News_2	-4.23	-5.38
News_3	-8.76	-12.33
News_4	0.15	-0.04
TV_1	-3.40	-2.46
TV_2	-0.29	-2.34
TV_3	-1.77	-4.38
TV_4	-0.68	-0.56
TV_5	3.30	0.66
TV_6	-18.54	-21.02
UgcVert_1	2.66	1.47
UgcVert_2	-18.29	-17.12
UgcVert_3	-2.87	-3.42
Vlog_1	-21.37	-18.42
Vlog_2	-2.09	-2.84
OVERALL	**-6.15**	**-6.54**

IV. CONCLUSION

In this paper, we present a new frame level rate allocation method for AV1. Based on the R-λ model analysis and temporal dependency model, the proposed algorithm simulates the temporal correlation in the hierarchical coding structure and allocates bits accordingly at frame level. We present evidence that the proposed method captures the temporal correlation better than the baseline, especially for videos with large motion and frequent scene changes. The proposed rate allocation algorithm improves compression efficiency for the rate control system in AV1.

REFERENCES

[1] Tihao Chiang and Ya-Qin Zhang. A new rate control scheme using quadratic rate distortion model. *IEEE Transactions on circuits and systems for video technology*, 7(1):246–250, 1997.

[2] Bin Li, Houqiang Li, Li Li, and Jinlei Zhang. λ domain rate control algorithm for high efficiency video coding. *IEEE transactions on Image Processing*, 23(9):3841–3854, 2014.

[3] Li Li, Bin Li, Houqiang Li, and Chang Wen Chen. λ-domain optimal bit allocation algorithm for high efficiency video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(1):130–142, 2016.

[4] Hongwei Guo, Ce Zhu, Shengxi Li, and Yanbo Gao. Optimal bit allocation at frame level for rate control in HEVC. *IEEE Transactions on Broadcasting*, 65(2):270–281, 2018.

[5] Shengxi Li, Mai Xu, and Zulin Wang. A novel method on optimal bit allocation at LCU level for rate control in HEVC. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2015.

[6] Jiangtao Wen, Meiyuan Fang, Minhao Tang, and Kuang Wu. R-(λ) model based improved rate control for HEVC with pre-encoding. In *2015 Data Compression Conference*, pages 53–62. IEEE, 2015.

[7] Zhenyu Liu, Libo Wang, Xiaobo Li, and Xiangyang Ji. Optimize x265 rate control: An exploration of lookahead in frame bit allocation and slice type decision. *IEEE Transactions on Image Processing*, 28(5):2558–2573, 2018.

[8] Jingning Han, Paul Wilkins, Yaowu Xu, and James Bankoski. A temporal dependency model for rate-distortion optimization in video coding. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–4. IEEE, 2019.

[9] Yi-Hsin Huang, Tao-Sheng Ou, Po-Yen Su, and Homer H Chen. Perceptual rate-distortion optimization using structural similarity index as quality metric. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(11):1614–1624, 2010.

[10] Stéphane Mallat and Frédéric Falzon. Analysis of low bit rate image transform coding. *IEEE Transactions on Signal Processing*, 46(4):1027–1042, 1998.

[11] Min Dai, Dmitri Loguinov, and Hayder Radha. Rate-distortion modeling of scalable video coders. In *2004 International Conference on Image Processing, 2004. ICIP'04.*, volume 2, pages 1093–1096. IEEE, 2004.

[12] Jason Garrett-Glaser. A novel macroblock-tree algorithm for high-performance optimization of dependent video coding in H. 264/AVC. *Tech. Rep.*, 2009.

[13] Yanbo Gao, Ce Zhu, Shuai Li, and Tianwu Yang. Temporally dependent rate-distortion optimization for low-delay hierarchical video coding. *IEEE Transactions on Image Processing*, 26(9):4457–4470, 2017.

[14] Alliance for Open Media. https://aomedia.googlesource.com/aom/.

[15] Yue Chen, Debargha Murherjee, Jingning Han, Adrian Grange, Yaowu Xu, Zoe Liu, Sarah Parker, Cheng Chen, Hui Su, Urvang Joshi, et al. An overview of core coding tools in the AV1 video codec. In *2018 Picture Coding Symposium (PCS)*, pages 41–45. IEEE, 2018.

[16] YouTube Media Algorithms team. https://media.withyoutube.com/.