Optimize X265 Rate Control

An Exploration of Lookahead in Frame Bit Allocation and Slice Type Decision

Zhenyu Liu[†], Libo Wang[‡] and Xiaobo Li[‡]

[†]Tsinghua University and [‡]Alibaba Co. Ltd

Contact Information:
4-305, FIT Building,
Tsinghua University
Beijing 100084, China

Phone: +86-13552444129
Email: liuzhenyu73@tsinghua.edu.cn

Abstract

x265 is an open source commodity software encoder. The problem of original x265 rate control stems from the long-term complexity based frame-level base quantization parameter, which is intrinsically not sensitive to the complexity variations of pictures being coded. In addition, the applied threshold in slice type decision, which is compared with the frame coding costs estimated by lookahead, does not consider the impacts of quantization. Because of the aforementioned irrational elements, the rate accuracy and the picture quality of original x265 were both deteriorated. In this paper, we explore the lookahead information to derive I-/P-frame base quantization parameter values, which considered not only the coding history but also the current and the near future picture complexities. Moreover, we introduced the quantization scale as a critical factor in the threshold for slice type decision, with which more pictures could be identified as B-type properly. The proposed algorithms were implemented in x265 version 2.4. Experiments revealed that, under the default preset (–preset medium), an averaged 0.59dB and up to 1.72dB BDPNSR coding quality gains were achieved, while saving the encoding time by 1.52% and improving the rate accuracy by 4.2% on average.

Introduction

As the critcal technique in video coding, rate control plays the vital role in the band-limited practical communication systems, which are subject to the discrepancy between the timing-variant complexities of pictures being coded and the constant band limitation. If we apply an constant quantization scale (Q), the required rate bandwidth might overflow or underflow the limitation seriously. The function of rate control is to find the Q for each coding unit (CU) that minimizes the rate-distortion-cost under the specified rate constrains. The rate control algorithm can be formulated as

$$\min_{Q_i} \sum_i (D_i + \lambda R_i) \text{ s.t. } \sum_i R_i = R_C,$$

The traditional rate control algorithms endeavored to improve the accuracy of rate model parameterized by Q, but neglected data dependency. In practice, the optimal bit allocation in a spatial-temporal dependent coding environment is more efficient to coding quality improvement. The CU-tree of x265 analyze CU grain information propagation in consecutive pictures by lookahead module, and then estimates the overall distortions to alter the Q of one CU. In the original algorithm, the frame-level base QP value merely depends on the complexity sum of the previously encoded frames, which cannot respond to the scene change sensitively. To overcome this hindrance, we exploit the lookahead results to estimate the overheads in the full resolution coding and the degree of information dependency in the future pictures, and then improve the performance of frame-level base QP definition. Additionally, we refine the slice type decision to improve the coding quality by considering the impacts of quantization in lookahead procedure. Experimental results revealed that an averaged 0.617dB quality gain was obtained by our proposals.

I-Frame base Q Estimation

R_T denote the desired bit-rate of one frame. R_I and R_{BP} represent the bit-rate of I-frame and the averaged bit-rate of the following B-/P-frames, respectively. The ratio between R_I and R_{BP} is expressed as $\alpha = \frac{R_I}{R_{BP}}$. If we want to achieve the averaged R_T in $k+1$ frames, we have

$$R_I = \frac{R_T\,(k+1)\,\alpha}{k+\alpha}.$$

Given the low-resolution costs $\dot{C}_I$, $\dot{C}_P$, and $\dot{C}_B$, and the target rate R_T, the key parameter α and I-frame QP estimation are derived by the regression algorithm 1. $\dot{C}_P$ and $\dot{C}_B$ are the averaged costs of P-frames and B-frames in the pre-motion-estimation, respectively. $\gamma_I = \frac{R_I}{\dot{R}_I}$, where the original I-frame rate and the corresponding quarter-I-frame rate are written as R_I and $\dot{R}_I$. For the resolution larger than HD720p, $s = 0.53$; Otherwise, $s = 1$. $\Delta\overline{QP}_{\text{CUTree}}$ is the averaged CU-Tree based QP offset of the I-frame. ΔQP_{BP} denotes the QP offset between B-frame and P-frame.

Algorithm 1: I-Frame QP Derivation

```
Q́P_I ← 30; Q́P_P ← Q́P_I + 1.5; Q́P_B ← Q́P_I + 3
Ṙ_I ← Ċ_I/QP2Q(Q́P_I); Ṙ_P ← Ċ_P/QP2Q(Q́P_P);
  Ṙ_B ← Ċ_B/QP2Q(Q́P_B)
while |QP_I − Q́P_I| > 0.1 do
    α ← Ṙ_I/(a_0(Q́P_I)Ṙ_P + a_1(Q́P_I)Ṙ_B)
    R_I ← αR_T (k + 1)/(k + α)
    Ṙ_I ← R_I/γ_I(Q́P_I)
    QP_I ← Q́P_I
    Q̃ ← sĊ_I/Ṙ_I
    Q́P_I ← Q́P_I + 0.5 (Q2QP(Q̃) − Q́P_I)
    Q́P_P ← Q́P_I − ΔQP̄_CUTree
    Q́P_B ← Q́P_P + ΔQP_BP
    Ṙ_I ← Ċ_I/QP2Q(Q́P_I)
    Ṙ_P ← Ċ_P/QP2Q(Q́P_P)
    Ṙ_B ← Ċ_B/QP2Q(Q́P_B)
end
QP_I ← Q́P_I − ΔQP̄_CUTree
```

I-frame Bits Amortization Algorithm

As the prediction residue energy of I-frame is larger than that of the P- and B-frame couterpart, and on the other hand, the Q of I-frame is always smaller than the following frames, the ration between R_I and R_{BP}, i.e., $\alpha = R_I/R_{BP}$, is much larger than 1. If no special treatment is carried out, when encoding the first P-frame next to the I-frame, the quality of this P-frame will be deteriorated seriously. x265 amortizes 77%, which is denominated as the amortize fraction (a_F), of the I-frame bits R_I over the next $k = 68$ frames. The amortize cost (C_{amort}) is written as

$$C_{amort} = \frac{a_F \cdot C_I}{k}.$$

In our work, a_F is rectified as

$$a_F = \frac{R_I - R_T}{R_I} = \frac{k\,(\alpha - 1)}{(k+1)\,\alpha},$$

P-frame Base Q Estimation

In x265, the frame level Q is derived from the long-term complexity. The main drawback is that, as the increase of encoded frame number, the frame level Q becomes more insensitive to the current frame complexity change. In our algorithm, the base quantization Q considers both effects of the long-term complexity history and the low resolution motion estimation costs in the current MiniGOP.

Let N_B denote the number of B-frames in the current MiniGOP. If the prediction residue complexities of the P frame and B frames in the current MiniGOP are C_P and C_B, respectively, the short-term P-frame base quantization step Q is expressed as

$$Q = \frac{(0.04/fd)^{1-q_c}}{(N_B+1)R_T}\left[C_P 2^{\frac{-\Delta\overline{QP}_{\text{CUTree}}}{6}} - \tau C_P(2^{\frac{-\Delta\overline{QP}_{\text{CUTree}}}{3}} - 1) + N_B C_B 2^{\frac{-\Delta QP_{\text{BP}}}{6}}\right]$$

The term $\tau C_P(2^{\frac{-\Delta\overline{QP}_{\text{CUTree}}}{3}} - 1)$ denotes how much we can borrow in the future mini-GOP. The higher dependency, more we can borrow. τ is a parameter derived by regression.

Quantization Strength based Slice Type Decision

If one of the following three conditions is satisfied, it is claimed that the next P frame or frames are found:

- First, $\mathrm{PME}(-1,1,1)$ is carried out. If $N_{Intra}(-1,1,1) > 0.5N_{CU}$, f_0 and f_1 in **F** are decided as P-frames.
- Second, if $\dot{C}(-1,0,0) + \dot{C}(0,1,1) < \dot{C}(-1,0,1) + \dot{C}(-1,1,1)$, f_0 is determined as one P-frame.
- Finally, we define the low threshold as $T_L = \max(300 - 50\imath_C, 30)$ and the high threshold as $T_H = 66Q'_P$, where Q'_P is the last P-frame base Q. The applied threshold is

$$T = \dot{C}(-1,0,0) < 150N_{CU} \&\& \dot{C}(0,1,1) < 150N_{CU} || T_H < 300\ ?\ T_L : T_H.$$

We process $\mathrm{PME}(-1,\imath_C,\imath_C)$ for each $\imath_C > 0$ in sequence. Once $\dot{C}(-1,\imath_C,\imath_C) > N_{CU} \cdot T$, the frame $f_{\imath_C}$ is defined as the P-frame and its leading undetermined frames are defined as B-type. The middle position B-frame is the reference B-frame. By introducing Q in the threshold T_H, with the increase of quantization, more frames could be identified as B-frames properly.

Results

The proposed methods were conducted on x265 version 2.4[1]. The coding performance results of the overall rate control optimizations are shown in Table 1. We applied preset medium in these tests. On average, +0.5904dB quality gain, i.e., 17.67% rate reduction, was achieved by our methods. The coding time was reduced by 1.52%.

Table 1: Coding Quality of Whole Rate Control Algorithm

Class	Sequence	BP[dB]	BR[%]	ΔT[%]
A	PeopleOnStreet	0.1029	-02.46	+1.07
	Traffic	0.6232	-21.28	+1.59
B	BasketballDrive	0.1524	-05.38	+3.36
	BQTerrace	0.5113	-29.52	+2.33
	Cactus	0.4367	-15.62	+0.23
	Kimono	0.1358	-04.61	-3.38
	ParkScene	0.6103	-19.61	+2.50
C	BasketballDrill	0.6759	-17.91	+1.25
	BQMall	0.5025	-12.51	+1.14
	Flowervase	0.3813	-12.56	+0.77
	PartyScene	0.6773	-20.21	+0.84
	RacehorsesC	0.1576	-04.90	+1.65
D	BasketballPass	0.5123	-10.62	+5.27
	BlowingBubbles	0.8624	-21.64	+5.60
	BQSquare	1.7205	-41.37	+2.29
	Flowervase	0.5058	-14.13	+0.41
	RaceHorses	0.2321	-05.37	+1.75
E	Vidyo1	0.5723	-22.76	+1.34
	Vidyo3	0.6529	-20.71	+2.53
	Vidyo4	0.6304	-22.52	+2.45
	KristenAndSara	0.7858	-28.52	-0.47
	Johnny	0.8728	-46.77	+3.23
	FourPeople	0.9236	-28.77	+2.89
F	SlideEditing	0.1211	-00.48	+2.15
	ChinaSpeed	0.3779	-10.84	-2.51
	SlideShow	1.4936	-17.29	+1.59
	BasketballDrilltext	0.7094	-18.60	-0.87
	Average	0.5904	-17.67	+1.52

Figure 1: Subject Comparisons

Conclusions

- Explore data dependency to optimize frame-level bit allocation.
- Introduce quantization strength in slice type decision.
- BDPSNR was improved by 0.59dB on average.

References

[1] Z. Liu. x265 ratecontrol demo. https://sites.google.com/site/liuzhenyu73/home/publications/x265_ratecontrol_demo.rar.