

Efficient Quantization Step Selection Scheme For I-Frame In Rate-Constrained Video Coding

ShuMin Zhou^{1,2} JinTao Li¹ YongDong Zhang¹

¹Digital Technology Lab, Institute of Computing Technology, Chinese Academy of Sciences
P. O. Box 2704, Beijing, 100080, P. R. China, email: smzhou@ict.ac.cn

²Graduate School of the Chinese Academy of Sciences

ABSTRACT

Quantization step (QS) selection for I-frame has an obvious impact on rate-distortion performance in rate-constrained video coding. An efficient QS selection scheme for I-frame is presented which is based on a balanced bits allocation scheme between I and P frames and a new experimental rate-quantization step (R-Q) model. In the balanced bits allocation scheme, firstly a model on the ratio of bits allocation between I and P frame is derived under the constraint of certain target bit rate; then based on the model, a feasible bits allocation algorithm is proposed by experimental analysis among different target bit rates. The analysis and final simulation results show that compared to the existing methods, the proposed algorithm improves PSNR performance significantly (up to 1.2dB).

1. INTRODUCTION

Rate-constrained video coding techniques have been widely used in streaming media processing and bandwidth-constrained video communication. However, most of them focus on the rate control scheme for inter frames (i.e. P and B frame) rather than intra frames. In fact, bits allocation for I-frame in a GOP also has an obvious impact on rate-distortion performance, which can be illustrated in Fig. 1. It can be seen from Fig.1 that: 1) the optimal quantization parameter (QP) for I-frame does exist and it can outperform the bad one up to 2.0 dB in PSNR; 2) the optimal QP for I-frame varies from sequences and different target bit rate. As to the second point, F.Pan[1] and Z.G.Li[2] have presented some methods. In Z.G. Li's scheme, only target bit rate was considered. F. Pan has proposed an empirical model in which bits allocation for I-frame was determined by its spatial complexity and target bit rate. However, it is well-

known that the optimal bits allocation for I-frame not only depends on target bit rate and I-frame itself, but also the interdependency between I and P frames. Here the interdependency can be defined as follows:

$$D_{inter} = \sum_{i=1}^N \Delta D_i | \Delta D_i \quad (1)$$

$$R_{inter} = \sum_{i=1}^N \Delta R_i | \Delta R_i$$

where $\Delta R_i, \Delta R_i, \Delta D_i, \Delta D_i$ stand for rate and distortion change for I frame and frame i (P frames) respectively in a GOP.

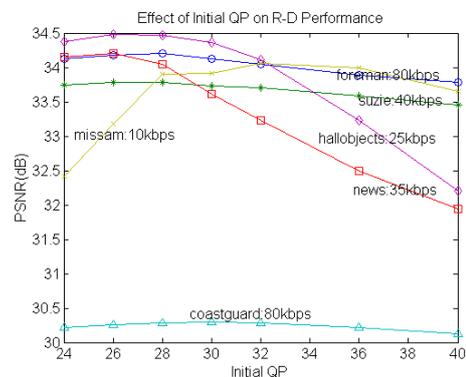


Fig.1. Effect of initial QP on R-D performance, all sequences here are QCIF, 30Hz

When target bit rate is fixed, we observe that the distortion and rate interdependency between I frame and successive P frames in a GOP are nearly constant. Based on the observation, firstly a model on the ratio of bits allocation between I and P frames is derived under certain target bit rate. Next, based on experimental analysis, bits allocation for I-frame is obtained among different target bit rate. Then, a new experimental rate-quantization step (R-Q) model for I-frame is derived by experimental analysis. Finally, with the R-Q model and bits allocation scheme, a new QS (or QP) selection scheme is proposed. The final simulation results on JM76 show that compared

to the existing methods, the proposed scheme has an obvious gain in PSNR.

The rest of the paper is organized as follows. Section 2 gives the balanced bits allocation scheme and the new experimental R-Q model for I-frame. In section 3, with the techniques proposed, an easy but efficient quantization step selection scheme is presented. The experimental results and discussions are given in section 4. Finally, our summary and conclusion are presented in section 5.

2. BALANCED BITS ALLOCATION AND NEW R-Q MODEL

2.1 Balanced Bits Allocation between I and P Frames

In this section, first, we allocate bits under fixed target bit rate. Second, an adaptive bits allocation for I-frame among different target bit rate can be made out based on the derived result and experimental analysis.

Prior to encoding a GOP under certain target bit rate, we assume that all P frames use the same bits so as to get an analytical model for balanced bits allocation between I and P frames. Besides, to simplify the deduction, we also assume that I and P frames has the same analytical R-Q format.

To obtain reasonable bits balance between I and P frames over a GOP, we introduce a new variable L which represents the ratio of bits assignment between I and P frame:

$$L = \frac{R_0}{R_p} \quad (2)$$

Based on Lagrange relaxation rate-distortion scheme, the R-D cost function of a GOP (format is IPPP) can be expressed as:

$$J = \sum_{i=0}^M D_i + \lambda (\sum_{i=0}^M R_i - B) \quad (3)$$

where D_i , R_i stand for distortion and bits for frame i , M is the length of GOP, B is target bits.

D-Q model can be provided as follows [3]:

$$D_i(Q_i) = \frac{Q_i^2}{12} \quad (4)$$

Here Q_i is quantizer step size of frame i . Note that frame 0 is I-frame in a GOP.

The analytical R-Q model can be approximately formulated as follows [3]:

$$R_i = \frac{e}{\ln 2} \frac{\delta_i^2}{Q_i^2}, 0 \leq i \leq M-1 \quad (5)$$

Here δ_i^2 is the variance of source data.

According to the above assumption, here we assume that all P frames in a GOP use the same bits approximately, which can be expressed as (6):

$$\frac{\delta_i^2}{Q_i^2} = \frac{\delta_{i+1}^2}{Q_{i+1}^2}, 1 \leq i \leq M-1 \quad (6)$$

Then, plugging (2)(4)(5)(6) into (3) provides the result:

$$J(Q_0^2, L) = \frac{Q_0^2}{12} + \frac{LQ_0^2}{12\delta_0^2} \sum_{i=1}^M \delta_i^2 + \lambda \frac{e}{\ln 2} \left(\frac{M}{L} + 1\right) \frac{\delta_0^2}{Q_0^2} - \lambda B$$

The minimum of J is obtained by setting its partial derivatives to 0, i.e:

$$\begin{cases} \frac{\partial J}{\partial Q_0^2} = \frac{1}{12} + \frac{LM\delta_\mu^2}{12\delta_0^2} - \lambda \frac{e}{\ln 2} \left(\frac{M}{L} + 1\right) \frac{\delta_0^2}{(Q_0^2)^2} = 0 \\ \frac{\partial J}{\partial L} = \frac{M\delta_\mu^2 Q_0^2}{12\delta_0^2} - \lambda \frac{e}{\ln 2} \frac{M\delta_0^2}{L^2 Q_0^2} = 0 \end{cases} \quad (7)$$

where $\delta_\mu^2 = \frac{1}{M} \sum_{i=1}^M \delta_i^2$ is the mean variance of P frames in a

GOP.

From (6), (7) can be deduced easily.

$$L = \frac{\delta_0}{\delta_\mu} \quad (8)$$

here δ_μ is the average standard deviation of P frames in a GOP.

The relationship in (8) is an ideal model of L under certain target bit rate, but it reflects the linear correlation between optimal L and RSD (ratio of standard deviation: $RSD = \frac{\delta_0}{\delta_\mu}$) at least. To confirm this, exhaustive

experiments have been done, in which foreman, akiyo, mother&daughter, coastguard, container, hallobjects, miss_am, suzie and news are tested under 25kbps, 50kbps, 70kbps, 100kbps, 150kbps, 300kbps respectively. For each bit-rate of each sequence, an experimental optimal L is made out. All of the sequences used are QCIF(30Hz). The optimal L is obtained by exhaustive search in [1,100]. As to the average standard deviation of P frames, since it is mainly determined by local motion strength in a sequence when target bit rate is fixed, we replace it with a statistic feature which can represent the local motion efficiently and be proportional to δ_μ . The statistic feature known as block variance difference will be detailed later.

Fig. 2 shows the relationship between optimal L and RSD under certain target bit rate, from which we can see: 1) the optimal L is approximately linear with RSD when target bit rate is fixed, which verifies the validness of (8), whose experimental model is given in (9); 2) among different target bit rates, the gradient and offset of each line, i.e parameter A and B in (9), decline as the target bit rate increases. It should be noted that RSD is fixed for certain a sequence, i.e, RSD only varies from different sequences. To make out the relationship between the parameters in (9) and target bit rate, further experiments have been done, whose results are shown in Fig. 3. It can be seen clearly in Fig.3 that both A and B are approximately linear with

target bit rate except that there is an abrupt change in gradient and offset when the target bit rate is up to certain value.

Thus, A and B can be expressed as the function of target bit rate, which is shown in (9) also.

$$L = A * RSD + B, \begin{cases} A \\ B \end{cases} = \begin{cases} a1 & a2 \\ b1 & b2 \end{cases} \begin{cases} TBR \\ 1 \end{cases} \quad (9)$$

where TBR is the abbreviation of target bit rate, $a1, a2, b1, b2$ are statistic constant, which can be obtained as follows by linear regression in the paper:

$$(a1, a2) = \begin{cases} (-0.0014, 0.1688), TBR < T_{ibr,a} \\ (-0.0001, 0.0724), TBR \geq T_{ibr,a} \end{cases}$$

$$(b1, b2) = \begin{cases} (-0.0922, 17.9151), TBR \leq T_{ibr,b} \\ (-0.0165, 8.7518), TBR > T_{ibr,b} \end{cases}$$

where $T_{ibr,a}$ and $T_{ibr,b}$ are the target bit rate threshold of abrupt change in gradient and offset respectively. In the paper, both of them are set to be 100kbps through experiments.

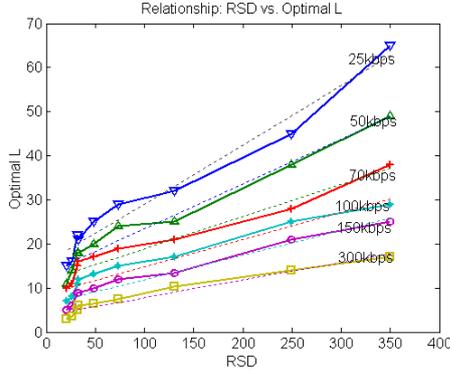


Fig. 2 The linear model of optimal L

With the L in (9), balanced bit allocation for I-frame can be calculated as follows:

$$R_0 = M \frac{bit_rate}{frame_rate} \frac{L}{L + M} \quad (10)$$

where bit_rate is target bit rate, $frame_rate$ is frame frequency.

2.2 Derived R-Q model for I-frame

According to R-Q model in [4], relationship between R and Q can be expressed as follows in high bit rate:

$$R_0(Q_0) = \log_2(2e^2 \frac{\delta_0^2}{Q_0^2})$$

To confirm the accuracy of the relationship above, foreman, akiyo and news are used to generate statistical data. All of them are encoded in I-frame only. 100 frames are used for each sequence and QP set is [24,26,28,30,32, 35,38,40]. Fig.4 shows the results that R (bpp) is quadratic function of $\ln(Q)$ rather than linear. Thus, a more accurate R-Q function for I-frame is presented as follows:

$$R_0(Q_0) = a \ln^2(\frac{\delta_0}{Q_0}) + b \ln(\frac{\delta_0}{Q_0}) + c \quad (11)$$

where a, b and c are statistic parameters, which can be obtained by linear regression: $a = 0.2346, b = 0.5657, c = 0.6206$.

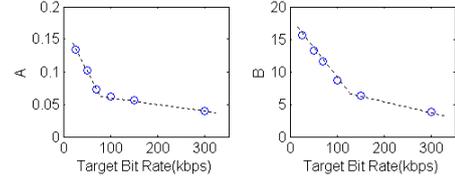


Fig.3. Relationship between A,B and target bit rate

The relationship between QP and QS in H.264 can be approximately expressed as $Q = 2^{\frac{QP}{6}}$. Therefore, QP_0 can be easily obtained once $\ln(\frac{\delta_0}{Q_0})$ is made out from (11) at certain target bits R_0 .

3. QUANTIZATION STEP PREDICTION SCHEME

In the paper, as is mentioned above, block variance difference [5] is used to replace the average standard deviation of P frames, which can be formulated as follows:

$$\delta_\mu = \frac{1}{N_f - 1} \sum_{k=1}^{N_f} BV(f_{k-1}, f_k)$$

$$BV(f_n, f_m) = \frac{1}{N_{MB}} \sum_{i=1}^N |\text{var}_n(i) - \text{var}_m(i)|$$

where BV is the abbreviation of block variance difference, N_{MB} is number of MB in a picture. $\text{var}_n(i)$ indicates the block variance of MB_i in frame n . N_f is the frames used to predict δ_μ .

If the current GOP is the first GOP, in order to reduce the delay, N_f is set to 2, i.e, 2 frames should be buffered at the beginning. With the two frames, we can make out δ_0 and estimate δ_μ .

If the current GOP is not the first GOP, only the first frame in current GOP needs to be buffered so as to make out δ_0 . δ_μ is estimated by using the average block variance in last GOP.

With δ_0 and δ_μ , RSD can be obtained. Then, according to the current target bit rate, the gradient A and offset B in (9) can be made out. With A and B , the optimal L can be estimated according to (9) and accordingly, bits allocation for I-frame can be derived by (10). Finally, the initial QP of current GOP can be made out as follows:

$$\theta = \ln\left(\frac{\delta_0}{Q_0}\right), QP_0 = \frac{6}{\ln 2} Q_0 \Rightarrow$$

$$\theta = \begin{cases} \frac{\sqrt{b^2 - 4a(c - R_0)} - b}{2a}, & b^2 - 4ac \geq 0 \\ -\frac{b}{2a}, & \text{others} \end{cases}$$

where a,b,c is the parameters in (11).

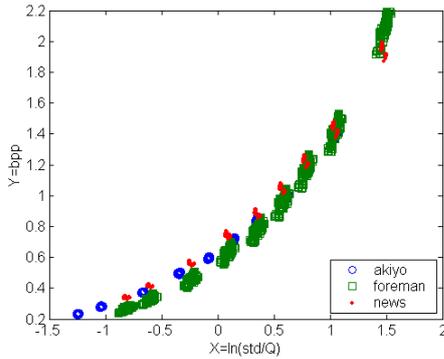


Fig.4. Relationship between bits per pixel (bpp) and $\ln(\text{std}/Q)$ for I-frame

4. EXPERIMENTS AND ANALYSIS

Numerous experiments have been conducted to evaluate the performance of the proposed algorithm. The test sequences used are in QCIF format, with I-frame appearing every 50 frames. Note that no frame skip is concerned here. The total frames are 150 for each sequence (3 GOPs). All the experiments are carried out on JM76 (reference = 1).

Table 1 shows the comparison in PSNR between the proposed scheme and Z.G.Li's (ZGL) (default for JM76) and F.Pan's (FP) methods. It can be seen that, for the proposed scheme, 1) obvious gains in PSNR have been obtained over both of the existing methods (0.33dB and 0.18dB on average relative to ZGL and FP respectively); 2) FP scheme works better than ZGL; 3) generally, gains in low bit rate are higher than in high bit rate.

5. CONCLUSION

An easy but efficient QP prediction scheme for I-frame is presented in the paper. Compared to the existing techniques, the proposed algorithm has two main distinguished features:

- 1)an new balanced bits allocation scheme is adopted to assign bits for I-frame, which is more comprehensive than the existing techniques.
- 2)an accurate experimental R-Q model was proposed to obtain the QP for I-frame, which is coherent to the relationship between QS and QP in H.264.

Although the scheme proposed is not optimal one, it performs better than the existing algorithms, which is due to the fact that more factors are considered when allocating bits for I-frame. The optimal QP (or QS) selection is a very complex problem (NP complexity), therefore it is useful to look for an alternative to approach it as closely as possible. However, many statistic parameters are used in the proposed scheme, we'll further improve them in adaptive way in the future.

Test Name	Z.G.Li PSNR	Proposed PSNR	F.Pan PSNR	Gain(dB)	
				vs.ZGL	vs.FP
F-60	32.51	32.71	32.59	+0.20	+0.12
F-75	33.37	33.64	33.43	+0.27	+0.21
F-100	34.77	34.80	34.78	+0.03	+0.02
F-150	36.43	36.44	36.44	+0.01	+0.00
S-45	34.71	34.84	34.80	+0.13	+0.04
S-65	36.16	36.30	36.20	+0.14	+0.10
S-95	37.86	37.89	37.82	+0.03	+0.07
S-135	39.28	39.32	39.28	+0.04	+0.04
N-25	30.16	30.55	30.35	+0.39	+0.20
N-60	34.63	35.76	34.98	+1.13	+0.78
N-90	38.00	38.25	38.02	+0.25	+0.23
N-130	40.06	40.27	40.14	+0.21	+0.13
H-25	32.25	32.69	32.46	+0.44	+0.23
H-45	35.35	36.26	35.98	+0.91	+0.28
H-70	37.27	38.56	38.01	+1.29	+0.55
H-100	39.99	40.01	40.02	+0.02	-0.01

Table 1. Comparison PSNR(dB) among three schemes, here test name F-n, S-n, N-n and H-n stand for Foreman, Suzie, News and Hallobjects respectively at n-Kbps.

Acknowledgement: This work is supported by National Nature Science Foundation of China under grant number 60302028.

6. REFERENCES

- [1]F. Pan, Z. Li. X. Lin et al, "An intra-frame quantization optimization algorithm for very low bit rate video coding". Inst. For Infocomm Research, Singapore, Tech. Rep. June 2002.
- [2]Z.G Li, F. Pan, K. P. Lim, G. Feng, X. Lin and S. Rahardja, "Adaptive Basic Unit Layer Rate Control for JVT", ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6-G012, Pattaya II ,Thailand, 7-14 March, 2003.
- [3]Christos Grecos and Jianmin Jiang. "On-line improvement of the rate-distortion performance in MPEG-2 rate control", IEEE Trans. On CSVT, Vol . 13, No. 6, June 2003, pp.519~528.
- [4]Jordi Ribas-Corbera and Shawmin Lei. "Rate control in DCT video coding for low-delay communications". IEEE Trans. On CSVT. Vol.9.No.1.Feb 1999, p172-185.
- [5]Jungwoo Lee and Bradley W. Dickinson, "Temporally Adaptive Motion Interpolation Exploiting Temporal Masking in Visual Perception". IEEE Trans. On Image Processing, Vol 3. No.5. September 1994, p513-526.