AN IMPROVED RATE-QUANTIZATION MODEL FOR RATE CONTROL IN REAL-TIME VIDEO ENCODING

Bo Xie

PacketVideo Corporation
San Diego, CA 92121
xie@packetvideo.com

Wenjun Zeng

Dept. of Computer Science
Univ. of Missouri-Columbia
zengw@missouri.edu

ABSTRACT

There is a constant need to improve the rate-quantization (*R-Q*) model to achieve accurate rate control for real-time video encoding. In the literature, considerable efforts have been focused on developing all kinds of *R-Q* models. In this paper, we extend our previous work in [1] and propose a new algorithm for *R-Q* model parameter estimation, as well as a new approach to address the model failure issue by thresholding the quantization parameter (QP) determined by the R-Q model without introducing additional complexity. Experimental results show that the new *R-Q* model parameter estimation algorithm improves over the one we previously proposed in [1], and the QP thresholding approach can address the model failure issue well. With both contributions, the proposed QP determination solution can achieve very close performance to the benchmark solution discussed in [1] that achieves the best performance by using an operational re-quantization approach.

1. INTRODUCTION

Quantization parameter (QP) determination is one of the basic components in rate control for standard video codecs. In [1], we discussed the important concept of *bit allocation guarantee*, which means the achievement of the allocated target bit budget should be enforced. Without bit allocation guarantee, a good bit allocation scheme would not produce expected performance, and buffer overflow and underflow in buffer-constrained applications would not be prevented effectively. Therefore, QP determination for bit allocation guarantee serves as a solid foundation for rate control.

In real-time video encoding, given a target bit budget allocated for a frame, typically an *R-Q* model is used to determine an appropriate QP for such bit budget. In the literature, there are basically three kinds of work for *R-Q* model development and improvement. First, most work focuses on developing different kinds of *R-Q* models [2]- [6]. He and Mitra [7] proposed ρ-domain source modeling where ρ is defined as the percentage of zero quantized DCT (Discrete Cosine Transform) coefficients. A more accurate *R-Q* model is built based on a more accurate linear R-ρ relationship and ρ to Q mapping. Second, in [1] we identified a fundamental issue that affects the *R-Q* model accuracy, i.e., model parameter estimation, and showed that the typical linear regression approach (e.g., the Least Mean Square Error (MSE) approach used in MPEG-4 Annex L rate control [9]) may not work well for non-stationary parameter estimation. We then proposed a *non-linear* search algorithm based on statistical similarity to improve the *R-Q* model parameter estimation over the Least MSE approach used in MPEG-4 Annex L. Third, in [1]-[2], when necessary, an operational approach based on re-quantization to select the best QP for the target bit budget is proposed. Since this approach is based on the actual data, it can provide the best performance in QP selection with the introduction of some extra complexity (about 30% extra quantization on average). Especially, it can handle the model failure issue which the first two kinds of approaches mentioned above cannot address.

In summary, the above three kinds of work basically suggests three directions in improving the *R-Q* model accuracy. The first is to develop a more accurate *R-Q* model, the second is to further improve the model parameter estimation, and the third is to provide an effective mechanism to address the model failure issue. In this paper, we follow the second and third direction to improve the *R-Q* model accuracy. We extend our previous work [1] and propose a new algorithm to improve model parameter estimation. In addition, we also provide a new approach to address the model failure issue without using re-quantization. We use the same simple quadratic *R-Q* model $R = \frac{X}{Q^2}$ (X is the sole parameter) as the one used in [1] as an example to show the idea. It should be noted that our work focuses on the frame-level QP determination.

The paper is organized as follows. In Section 2, we present the new *R-Q* model parameter estimation algorithm. In Section 3, we propose a new approach to address *R-Q* model failure issue. Finally we compare our approach with some existing work in Section 4.

2. AN IMPROVED *R-Q* MODEL PARAMETER ESTIMATION

In [1], we proposed a non-linear *R-Q* model parameter estimation to improve over the Least MSE based linear parameter estimation. In particular, the proposed non-linear approach searches within a set of previous frames and finds one particular frame that has the largest statistical similarity to the current frame, and then uses the *R-Q* data of that frame to estimate the sole parameter X of the simple quadratic *R-Q* model. We extend this non-linear approach and further improve the *R-Q* model parameter estimation in this section.

2.1 Basics of the new parameter estimation algorithm

In [1], we take the following specific form of the simple quadratic *R-Q* model, i.e.,

$$\frac{R}{mad} = \frac{X1}{Q^2} \iff R = \frac{X}{Q^2} \quad (X = X1 \cdot mad) \qquad (1)$$

where *mad* is the mean absolute difference of the frame. For I frames or Intra macroblocks (MB), we use the absolute value of each original pixel, instead of that of each residue pixel in calculating *mad*.

In addition, we choose *mad* as the measure of the statistical similarity to perform the search for parameter estimation, i.e., search one particular frame in a candidate set of previous frames, such that its *mad* is most similar to the *mad* of the current frame. Therefore, the goodness of *mad* as a statistical similarity measure essentially determines the accuracy of the *R-Q* model parameter estimation and the model itself.

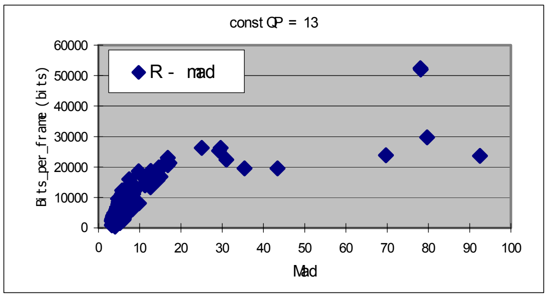

Fig. 1: R-mad data for the movie sequence "Hangingup" with constant QP. "Hangingup": 320x224, 300 frames, lots of scene change and high motion

Fig. 1 shows the actual *R-mad* data in the constant *QP* encoding scenario. We can see that *mads* for I frames or high complexity frames are usually sparse in a large range (i.e., $mad > 20$). For example, for one previous I frame that is most similar to the current I frame, due to the sparseness of *mad* at the high end in the range, the *mad* difference between the two frames can still be very large, which may cause some approximation error. In addition, it is seen that there are multiple *Rs* that correspond to the same *mad*, and *R* is not quite converged given the same *mad* for the same *QP*. This could also potentially introduce some approximation error. Therefore, *mad* may not be sufficiently good for use as a statistical similarity measure in *R-Q* parameter estimation.

We thus propose to use *J* as an alternative for the statistical similarity measure, where *J* is defined as follows.

$$J = mdev + \lambda \cdot \overline{R_{motion}} \,,$$

$$mdev = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{A}\sum_{j=1}^{A}\left|x_{j,i} - \overline{x_i}\right|\right), \quad \overline{R_{motion}} = \frac{1}{M} R_{motion} \qquad (2)$$

where *mdev* is referred to as the mean deviation, and $x_{j,i}$, $\overline{x_i}$, A, M and R_{motion} are the intensity of the *j*-th residue pixel of the *i*-th MB, the average residue intensity of the *i*-th MB, the number of pixels for each MB, the number of MBs in each frame and the total coding bits spent on the motion vectors of a frame, respectively; λ is a scaling factor that controls the relative contribution between the motion information and the residue.

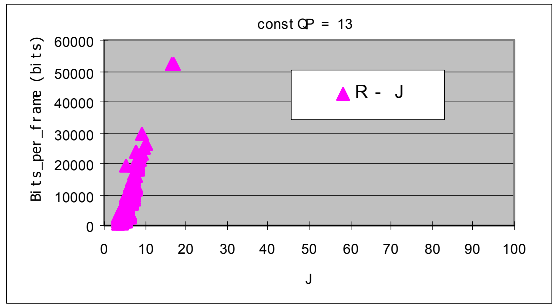

Fig. 2: *R-J* modeling results for the movie sequence "Hangingup" with constant QP ($\lambda = 2.3 \cdot QP$)

Fig. 2 shows the *R-J* curve for the same sequence, where the scaling factor λ is chosen as $\lambda = k \cdot QP$, which is similar to that used for rate-distortion based motion estimation in [8]. We can see that the range of *J* becomes much smaller than that of *mad*, and the distribution of *J* is denser than that of *mad* within the corresponding range, though there are a couple of *Js* at the very high end. In addition, the *R* data is much more converged for the same QP given the same *J* when compared with the case with the same *mad*. Therefore, *J* appears to be better than *mad* in determining the similarity. One of the reasons for this is that *J* reflects both residue and motion information.

By using the *J* measure, we change the simple quadratic *R-Q* model into the following form.

$$\frac{R}{J} = \frac{X2}{Q^2} \iff R = \frac{X}{Q^2} \quad (X = X2 \cdot J) \qquad (3)$$

Thus we choose *J* instead of *mad* as the measure of the statistical similarity, i.e., search for one particular frame in a candidate set of previous frames, such that its *J* is most similar to the *J* of the current frame.

2.2 Details of the new parameter estimation algorithm

A. New design of the candidate set

For the candidate set of previous frames, we choose a new way to collect historical data, to improve over the temporal sliding window approach adopted in [1]. The problem of the sliding window approach is that it is difficult to collect enough data of high complexity or I- frames within the limited window (e.g. 1sec window). To address this issue, we adopt a complexity-based classification. In particular, a fixed number of groups ranging from low complexity to high complexity are defined and each group is represented in a fixed-size sliding window that holds the latest frames with similar complexity. When a new frame is encoded, the group that has similar complexity as this frame is identified and then updated by the new frame data. We use *mad* to do the classification, as suggested in [1] that *mad* is a good measure of coding complexity.

$$Group\ i:\ low[i] < \frac{mad_n}{\overline{mad}_{n-1}} \le high[i], \quad i = 1, 2, \cdots, 7,$$

$$\{low[i]\} = \{0, 0.5, 1, 2, 3, 4, 5\}, \{high[i]\} = \{0.5, 1, 2, 3, 4, 5, +\infty\} \quad (4)$$

where mad_n is the *mad* of the current frame, and $\overline{mad}_{n-1}$ is the average *mad* from the starting frame to the previous frame.

Therefore, based on $\frac{mad_n}{\overline{mad}_{n-1}}$, each incoming frame will be categorized into the corresponding group.

B. Statistical similarity based search with the *J* measure

Next, the search for one specific frame with the most similar J as the current frame is performed within the available historical data set $\boldsymbol{X} = \{(A(i,j), QP(i,j), mad(i,j), J(i,j))\}$, where $i = 1, 2 \cdots, K$, $j = n - n_{n_1}^{(i)}$, $n - n_{n_2}^{(i)}, \cdots, n - n_{n_N}^{(i)}$, $0 < n_{n_1}^{(i)} < n_{n_2}^{(i)} < \cdots < n_{n_N}^{(i)}$; $A(i, j)$, $QP(i, j)$, $mad(i, j)$ and $J(i, j)$ are the actual bit count, QP, *mad* and J for the j-th frame of the i-th group in the past, respectively; K and N are the number of groups in $\boldsymbol{X}$ and the number of frames in each group; n represents the current frame.

Then search the whole data set $\boldsymbol{X}$ to get

$$(i^*, j^*) = \arg \min_{\substack{1 \le i \le K \\ n - n_{n_1}^{(i)} \le j \le n - n_{n_N}^{(i)}}} (\,|J(i,j) - J(n)|\,). \tag{5}$$

Note that the whole data set instead of the corresponding single group of data is used for the search because $\overline{mad}_{n-1}$ shown in (4) is changing all the time during the encoding, and it is likely that frames with identical or very similar *mad* can be categorized into different groups, which could potentially affect the search if only single group is chosen.

By applying Eq. (3), we get the current QP,

$$QP(n) = QP(i^*, j^*) * \sqrt{\frac{A(i^*, j^*)/J(i^*, j^*)}{T(n)/J(n)}}\,. \tag{6}$$

where $T(n)$ and $J(n)$ are the target bit count and J for the current frame.

3. HANDLING THE *R-Q* MODEL FAILURE CASES

One challenge for improving the *R-Q* model accuracy is to handle the model failure cases. When an *R-Q* model fails, usually the actual bit count based on the determined QP will overshoot or undershoot too much compared to the target bit count. This potentially increases the chances of buffer overflow or underflow. [1]-[2] introduces an operational approach based on re-quantization to control the QP and the actual bit count to hit the target in a deterministic way. This approach completely addresses this issue, but incurs additional complexity, which may be a concern for some real-time applications.

We thus propose a new approach to address this issue without introducing extra complexity. The basic idea is to come up with the lower or upper thresholds for the determined QP to avoid it being too small or too large, which is usually the case when a *R-Q* model fails. Note that we put more focus on the lower thresholds, and leave the upper thresholds as an open issue for the future work.

For high complexity frames ($mad_n >= \overline{mad}_{n-1}$), we believe a reasonable lower threshold for the current QP is the average QP over all previously encoded frames, considering that using constant QP for the entire sequence is the asymptotically optimal solution for coding independent sources at high bit rates.

For low complexity frames ($mad_n < \overline{mad}_{n-1}$), we realize that when the model fails, the model parameter estimation based on the data of one particular frame does not work. In such cases, we feel that using the average data might be a good alternative to come up with the threshold for the QP. Therefore, we use the previous average data to do the parameter estimation for the *R-Q* model shown in Eq. (3) and get the lower threshold.

$$QP(n) = \overline{QP}_{n-1} * \sqrt{\frac{C/\overline{J}_{n-1}}{T(n)/J(n)}}$$

$$\overset{T(n)\le C}{\ge} \overline{QP}_{n-1} * \sqrt{\frac{C/\overline{J}_{n-1}}{C/J(n)}} = \overline{QP}_{n-1} * \sqrt{\frac{J(n)}{\overline{J}_{n-1}}} \tag{7}$$

where C=*bitrate/framerate*, $\overline{QP}_{n-1}$ and $\overline{J}_{n-1}$ are the average QP and J over all previously encoded frames. Note that for the previous average bit count, we use C instead of the actual data because the previous average bit count should converge asymptotically to C. In addition, for low complexity frames, the allocated bit count T_n is usually smaller than the target bit count C in order to achieve good consistent quality.

Therefore we get the lower thresholds for the QP determined by Eq. (6).

$$\textit{If}\ (mad_n >= \overline{mad}_{n-1})\ \ QP_{low\_threshold,\,n} = \overline{QP}_{n-1},$$

$$\textit{else}\ \ QP_{low\_threshold,\,n} = \overline{QP}_{n-1} * \sqrt{\frac{J(n)}{\overline{J}_{n-1}}}\,. \tag{8}$$

Therefore, if $QP(n) < QP_{low\_threshold,\,n}$, then $QP(n) = QP_{low\_threshold,\,n}$.

In the implementation, considering thresholding QPs could potentially reduce the actual bitrate, we add a constraint, i.e., only when the current actual bitrate is higher than the target bitrate, we do QP thresholding as illustrated in (8).

4. SIMULATION RESULTS

We take the frame-level bit allocation framework proposed in [1], and evaluate the performance of the proposed QP determination solution. Based on the same frame-level bit allocation scheme, we compare the following five QP determination solutions: (1) Baseline solution. This is the solution proposed in [1] but without using re-quantization, and is the basic solution we try to improve over. (2) The proposed simplified solution with the new *R-Q* model parameter estimation algorithm as described in Section 2 but without the QP thresholding described in Section 3. This solution serves as a reference to show how the new *R-Q* model parameter estimation algorithm and QP thresholding work independently. (3) The proposed complete solution with the new *R-Q* model parameter estimation algorithm AND QP thresholding. (4) Benchmark solution. This is the solution proposed in [1] based on re-quantization, which provides the best performance with additional complexity. (5) The linear *R-ρ* model proposed in [7] for frame-level QP determination. We applied the above solutions to MPEG-4 simple profile encoder with 0.5sec VBV (Video Buffering Verifier) buffer size. We chose QCIF (176x144) "Foreman" (400 frames, medium motion, one scene change), "Glasgow" (750 frames, lots of scene cuts) and "Charles Angels" (1000 frames, action movie, lots of scene changes and high motion) as the test sequences.

From Table 1, we can see that the proposed complete solution achieves an average of 0.55 dB gain over the baseline solution, while the benchmark solution has an average of 0.64 dB gain over the baseline solution. The proposed complete solution has a closer performance to the benchmark solution in complex sequences such as "Charles Angels" (0.05 dB loss on average) than the simple sequence "Foreman" (0.17 dB loss). In terms of frame dropping, the proposed complete solution only incurs about 2 more dropped frames than the benchmark solution. Our proposed solution achieves very close performance to the benchmark solution in terms of both PSNR and frame dropping.

Table 1 suggests that for the proposed simplified solution, the new *R-Q* model parameter estimation algorithm provides an average of 0.39 dB PSNR gain over the baseline solution. QP thresholding contributes an average of 0.16 dB additional gain, with more gain (0.34 dB) for "Charles Angels" at 112 kbps and 10 fps. In addition, it successfully reduces the frame dropping in all test cases while keeping the actual bitrate accurate, e.g., for "Charles Angels" at 112 kbps and 10 fps encoding, 11 dropped frames in the proposed simplified solution is reduced to 2, which translates into more consistent temporal visual quality.

Table 1 also shows that the proposed complete solution achieves close performance as the linear *R*- ρ model [7] for *frame-level QP* determination. We make a few notes here. First, in terms of complexity, the linear *R*- ρ model for frame-level *QP* determination needs at least two to three quantization processes for each frame [7]. It is actually more complex than the benchmark solution which needs only 30% re-quantization and entropy coding on average [1]. Second, since our *QP* determination approach can determine the *QP* prior to performing DCT, it can be combined with any fast encoding algorithms (e.g., [10]) where pre-transform decisions can be made to save a lot of the computation of DCT, quantization/inverse quantization and IDCT. On the other hand, the linear *R*- ρ model approach relies on the availability of the distribution of the DCT coefficients, therefore may not allow fast encoding desirable in the real-time encoding scenario. Third, it appears that, with slightly increased complexity, the linear *R*- ρ model is more robust and accurate than the quadratic *R-Q* model. However, our proposed non-linear model parameter estimation approach is generic. It can be applied to the linear *R*- ρ model as well to provide further improvement.

5. CONCLUSION

There are two contributions in this work. First, we propose a new non-linear *R-Q* model parameter estimation algorithm to achieve accurate rate control. Second, we propose a new approach to address the model failure issue by thresholding the QP determined by the *R-Q* model without introducing additional complexity. Experimental results show that both approaches help improve the *R-Q* model accuracy, and the combination can achieve very close PSNR performance as the benchmark solution in [1] that achieves the best performance by using the re-quantization approach. Requiring no re-quantization, the proposed solutions run as fast as the MPEG-4 Annex L rate control solution.

6. REFERENCES

[1] B. Xie and W. Zeng, "Sequence-based rate control for constant quality video," in *Proc ICIP*, Rochester, Sept. 2002.

[2] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. Cir. Sys. Video Tech.*, vol. 6, no. 1, pp. 12-20, Feb. 1996.

[3] B. Tao, H. A. Peterson and B.W. Dickinson, "A rate-quantization model for MPEG encoders," in *Proc. ICIP*, vol. I, pp. 338-341, Santa Barbara, CA, Oct. 1997.

[4] L.-J. Lin, A. Ortega and C.-C. J. Kuo, "Rate control using spline-interpolated R-D characteristics," in *Proc.1996 VCIP*, Orlando, FL, Mar. 1996, pp. 111-122.

[5] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 7, pp. 246-250, Feb. 1997.

[6] K. H. Yang, A. Jacquin and N. S. Jayant, "A normalized rate-distortion model for H.263-compatible codecs and its application to quantizer selection," in *Proc.ICIP*, vol. II, pp. 41-44, Santa Barbara, CA, Oct. 1997.

[7] Z. He and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Trans. Cir. Syst. Video Tech.*, vol. 11, no. 12, Dec. 2001.

[8] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," in *IEEE Sig. Proc. Mag.*, vol. 15, no. 6, pp. 74-90, Nov. 1998.

[9] Coding of Moving Pictures and Associated Audio MPEG-4 Committee Draft ISO/IEC 14496-2, Oct.1998.

[10] I-M. Pao and M.-T. Sun, "Modeling DCT coefficients for fast video encoding," *IEEE Trans. Cir. Syst. Video Tech..*, June, 1999.

Table 1: Comparisons of frame dropping and PSNR, among four QP determination solutions (1st line: baseline solution; 2nd line: proposed simplified solution without QP thresholding; 3rd line: proposed complete solution; 4th line: benchmark solution); 5th line: linear *R-ρ* model proposed in [7].

Sequence (bitrate, framerate)	Actual Bitrate (kb/s)	Actual framerate (f/s)	# of dropped frames (Percentage)	PSNR (db)	Gain in PSNR (db)
Foreman (64kb/s, 10f/s)	63.28	9.15	12(9.0%)	30.63	0
	63.73	9.75	4(3.0%)	31.29	+0.66
	63.43	9.90	2(1.5%)	31.37	+0.74
	63.39	10.00	0(0)	31.58	+0.95
	63.29	9.90	2(1.5%)	31.37	+0.74
Foreman (112kb/s, 10f/s)	111.22	9.38	9(6.8%)	33.31	0
	110.20	9.75	4(3.0%)	33.71	+0.40
	111.50	9.90	2(1.5%)	33.93	+0.62
	110.86	10.00	0(0)	34.06	+0.75
	110.54	9.97	1(0.75%)	33.96	+0.64
Glasgow (64kb/s, 10f/s)	62.39	9.44	14(5.6%)	28.01	0
	63.48	9.64	9(3.6%)	28.37	+0.36
	63.03	9.92	2(0.8%)	28.54	+0.53
	63.43	10.00	0(0)	28.56	+0.55
	63.41	9.88	3(1.2%)	28.39	+0.38
Glasgow (112kb/s, 10f/s)	110.18	9.44	14(5.6%)	30.71	0
	110.70	9.84	4(1.6%)	31.08	+0.37
	110.84	9.88	3(1.2%)	31.10	+0.39
	111.05	10.00	0(0)	31.20	+0.49
	111.07	9.92	2(0.8%)	31.09	+0.38
Charles Angels (64kb/s, 10f/s)	63.46	9.69	11(3.3%)	29.83	0
	63.49	9.75	9(2.7%)	30.13	+0.30
	63.07	9.87	5(1.5%)	30.27	+0.44
	64.02	9.93	3(0.9%)	30.31	+0.48
	63.91	9.90	4(1.2%)	30.34	+0.51
Charles Angels (112kb/s, 10f/s)	109.97	9.69	11(3.0%)	32.61	0
	111.25	9.69	11(3.3%)	32.83	+0.22
	111.29	9.96	2(0.6%)	33.17	+0.56
	111.91	10.00	0(0)	33.23	+0.62
	111.73	9.99	1(0.3%)	33.23	+0.62