To appear in *IEEE Inter. Conf. Image Processing*, Sept., 2002, Rochester, NY.

SEQUENCE-BASED RATE CONTROL FOR CONSTANT QUALITY VIDEO

Bo Xie and Wenjun Zeng

PacketVideo Corporation
4820 Eastgate Mall, San Diego, CA 92121
{xie, wzeng}@packetvideo.com

ABSTRACT

Most model based rate control solutions have the generally questionable assumption that video sequence is stationary, and also suffer from the fundamental problems of model parameter mis-estimation. In this paper, we propose a sequence based bit allocation solution with the capability of tracking the non-stationary characteristics in the video source without look-ahead encoding and thus stationarity assumption is no longer needed. In addition, a new model parameter estimation approach is provided to solve the problems in the existing model parameter estimation schemes. Moreover, a general concept of bit allocation guarantee is presented to achieve the allocated bits in a deterministic way. The proposed rate control solution can achieve constant quality video with less quality flicker and motion jerkiness.

1. INTRODUCTION

Rate control is a central piece for standard video codecs (e.g., MPEG-1, MPEG-2, MPEG-4 and H.26x) to achieve consistent good quality for the whole sequence under the channel bandwidth and buffer constraints. In general, rate control includes two parts, one is bit allocation and the other is bit allocation achievement, i.e., QP (Quantization Parameter in standard video codecs) determination for achieving the allocated bits for the current frame or macroblock accurately.

For the bit allocation part, MPEG-2 TM5 [1] is a benchmark, i.e. GOP (Group of Pictures) based bit allocation in which the GOP size is fixed, and constant bit allocation among GOPs is used. Basically, the assumption of this bit allocation is that video sequence is stationary, which is usually questionable. Therefore there are some problems in this GOP based bit allocation, and people have tried to improve it from the following three directions. First, most work focuses on the modifications or improvement in bit allocation within a GOP. Some bit allocation schemes [3] allocate more bits on high complexity frames and less bits on low complexity frames. Others do R-D optimized bit allocation [5]-[6]. Second, since a fixed-size GOP doesn't match the scene structure in a multiple-scene sequence, people have tried dynamic GOP [7] to determine GOP structure and frame characteristics using look-ahead techniques. Dynamic GOP solves the scene change problem, but does not touch the problem of bit allocation among different GOPs (fixed size or dynamic). Third, GOP based bit allocation, if not done wisely, is hard to achieve constant quality for the whole sequence. Scene complexity was taken into account in [3]-[4] in an effort to solve the problem of bit allocation among scenes (GOPs) in a sequence. But, they used previous data to estimate the model parameters of the current scene or frame, which may cause some problems because the model parameters estimated based on the previous data may not truly reflect the statistics of the current scene.

For bit allocation achievement part, i.e. QP determination, most previous work focuses on developing all kinds of rate-quantization (R-Q) models, for example, logarithmic [9], power [8], spline [10], polynomial (including linear and quadratic) [1]-[2], [11] models etc. All R-Q model based work has no guarantee for bit allocation achievement because: (1) all R-Q models are some kind of approximation to the reality; (2) in most work, model parameter estimation is based on the previous data, and per se a kind of linear estimation (for example, MSE estimation etc.). That may not track the statistics of non-stationary parameters. As a result, more unexpected frame dropping would happen. An exception is Wei Ding's work [8] where re-encoding to guarantee the achievement of bit allocation was considered. Other work, for example, operational R-D optimized rate control [5]-[6] has the guarantee of bit allocation achievement, but needs huge complexity.

Our goal is thus to develop a generic rate control solution that can achieve constant quality (less quality flicker and motion jerkiness) across the entire sequence for a large variety of sequences with different characteristics. Basically, we need to provide a sequence based bit allocation with the capability of tracking the variability of non-stationary characteristics in the video source without resorting to two-pass or look-ahead encoding. We also provide a solid mechanism to achieve the bit allocation once determined. It should be noted that this solution is a frame level rate control solution and is derived under the assumption that bitrate, framerate and buffer size are given. For comparison, we choose MPEG-4 Annex L [2] frame level rate control as the main reference.

This paper is organized as follows. In Section 2, we present our new sequence based bit allocation solution. In Section 3, we address bit allocation achievement issue, and a general concept of bit allocation guarantee is proposed. We provide a new model parameter estimation method to improve the accuracy of estimation in Section 4. Finally we compare our solution with MPEG-4 Annex L rate control and show the simulation results in Section 5.

2. SEQUENCE BASED BIT ALLOCATION

The basic idea is to allocate more bits for scene change frames (which can usually be regarded as I frames), or high complexity frames, and less bits for low complexity frames in an effort to achieve constant quality in the sequence level. We use MAD (Mean SAD, Sum of Absolute Difference; for I frames or Intra macroblocks, we take the absolute value of each original pixel instead of that of each residue pixel) as an index to quantify the coding complexity of different kinds of frames. The coding complexity refers to the number of bits needed for encoding such frames. Specifically, we choose the following quadratic model.

$$R = K * \sqrt{mad - \alpha} \qquad (1)$$

where K, α are two parameters. K controls the slope of the R-mad curve, while α is the offset of *mad* due to quantization effect (when mad is small, the quantized DCT coefficients may become zero).

In fact, this model is based on the following observation. In general, the goal of this bit allocation is to achieve constant quality, i.e. constant distortion. And we know constant QP for the whole sequence is some sort of constant distortion solution. So we look at actual R-mad data in constant QP cases and try to model the R-mad relationship.

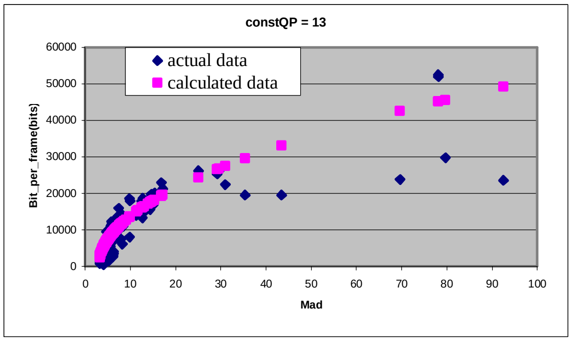

Fig.1 R-mad modeling results for movie sequence Hangingup with constant QP. Hangingup: 320x224, 300 frames, lots of scene change and high motion.

The actual R-mad data in Fig.1 shows that the linear model does not work in bit allocation. So we choose a quadratic model. We can see that this model with K=5200 basically follows the trend of the actual curve. But for high mad frames (which can be regarded as I frames), there is apparent discrepancy between the actual bit count and calculated bit count. Considering I frames are generally more important than P frames, we need to give I frames more priority and allocate more bits to them. So this model should work fine in those high mad cases. In fact, we tried other complexity measures such as σ^2 (variance of the residue) and found MAD is often better. More details can be found in [12].

In addition, to match the actual data well, we introduced the shifting factor α in the model. Ideally, when mad is zero, the coded bits for the residue should be zero. Thus this shifting factor should be very small. So in the implementation, we use the following model to do bit allocation.

$$R = K * \sqrt{mad} \qquad (2)$$

Based on this model, we can calculate the allocated bits for the current frame. Basically, what we have is previous information (average bit count per frame, average *mad*) and current *mad*. Note that the previous average bit count per frame should be close to C=bitrate/framerate. Therefore, we can get the following equation for bit allocation.

$$T_n = C * \sqrt{\frac{mad_n}{\overline{mad}_{n-1}}} \qquad (3)$$

where $C = \frac{Bitrate}{Framerate}$, i.e., VBV (Video Buffering Verifer) buffer output rate; T_n is the target bit allocated for the current frame, mad_n is the mad of the current frame, $\overline{mad}_{n-1}$ is the average mad from the starting frame to the previous frame.

In this target bit calculation, we do not need stationarity assumption any more and use *mad* to track the complexity variation in the video sequences effectively. Therefore, it automatically takes care of scene change and determines the bit allocation for all I, P and B frames in the sequence in a unified way. In addition, we use current data and average previous data which are highly reliable. The only assumption we make is the quadratic model shown in Eq.(2), supported by our R-mad modeling results.

After this sequence based bit allocation, there is one minor adjustment for the current frame.

$$T_n = T_n - \Delta_n\,,\ \Delta_n = \sum_{i<n}(R_i - T_i) \qquad (4)$$

where R_i and T_i are the actual bit count and target bit count in the past. Actually, Δ_n represents the cumulative difference between actual bit count and target bit count so far.

Next, we have to meet the buffer constraint. We need to check the buffer to avoid buffer overflow and underflow. In addition, this would guarantee that the final average bitrate converges to the given bitrate.

```
If ($VBV\_fullness + T_n - C > B_s$)
        $T_n = B_s - VBV\_fullness + C$;
Else if ($VBV\_fullness + T_n - C < 0$)                         (5)
        $T_n = C - VBV\_fullness$;
```

where *VBV_fullness* shows the buffer usage, B_s is buffer size, and $C = \frac{bitrate}{framerate}$ is VBV buffer output rate.

3. BIT ALLOCATION GUARANTEE

After target bit calculation, quantization level will be determined. This will be discussed in the next section. In this section, we present a concept of bit allocation guarantee. That means for any bit allocation scheme, when the target bits are allocated, the achievement of this target bit count should be guaranteed, i.e. the difference between the actual bit and the target bit should be controlled in a deterministic way (the ideal case is the actual bit count after encoding is as close as possible to the target bit count). Without this guarantee, any bit allocation scheme would not be well realized. This is particularly important when model-based quantization level determination is used. To achieve this bit allocation guarantee, we need a mechanism to address those cases where the model fails.

We do multi-pass quantization and entropy coding to achieve bit allocation. In particular, the multi-pass quantization framework is as follows:

Step 1. Go through the normal rate control (bit allocation, initial QP calculation) and do the quantization and entropy coding to get the actual number of bits for the current frame.

Step 2. Model mismatch detection and QP re-adjustment: If |Actual bit – Target bit| / Target bit > Threshold, then do QP re-adjustment; otherwise, go to remaining encoding modules, i.e. de-quantization and IDCT etc..

Step 3. Use updated QP to re-do quantization and entropy coding, then go to step 2.

This framework is similar to Wei Ding's work[8]. The difference, however, lies in the initial QP. In Wei Ding's work, two initial QPs are required, and the first QP is randomly chosen. After two-pass encoding with these two QPs, the actual data will be obtained to fit his R-Q model. So at least three-pass re-encoding is needed and this is not efficient. In our multi-pass framework, any model based initial QP calculation could be used, and if the model we use is accurate enough, then only one-pass is needed.

4. QUANTIZATION LEVEL DETERMINATION

Based on the multi-pass quantization framework outlined in the previous section, QP determination includes the initial QP determination and QP re-adjustment.

4.1. Initial QP determination

Due to multi-pass processing in each frame encoding, there may be several coding cases. So the basic idea is to pick up a case for some frame in the past (restricted by a window), which has the most similar statistical characteristics to the current frame, i.e., pick up a case in the past that most likely resembles the current frame in the statistical sense. Then we use the data for that case to estimate the parameter of the R-Q model used for calculating the current QP. We use *Actual bit/mad* as the metric for describing the statistical similarity between two frames.

Specifically, suppose we have the historical data set $\{(A(i,j), Q(i,j), mad(i))\}_{i=n-1,\, j=1}^{i=n-M,\, j=K_i}$, where $A(i, j)$ is the actual number of bit for the *j*th case of the *i*th frame in the past; $Q(i, j)$ is QP for the *j*th case of the *i*th frame in the past; $mad(i)$ is mad for the *i*th frame in the past; M is number of frames observed in the past; K_i is the number of cases of the *i*th frame; n represents the current frame. For the current frame, the data we can obtain before quantization and entropy coding are target bits $T(n)$ and $mad(n)$.

Search the above historical data set to get

$$QP(p,q)^* = \arg \min_{\substack{n-1\le i\le n-M \\ 1\le j\le K_i}} \left(\left| \frac{A(i,j)}{mad(i)} - \frac{T(n)}{mad(n)} \right| \right), \qquad (6)$$
$$where \quad p < n,\ q \le K_p$$

Then by applying the quadratic model $\frac{R}{mad} = \frac{X}{Q^2}$ similar to the model in MPEG-4 Annex L[2], we can get the current QP,

$$QP(n) = QP(p,q)^* * \sqrt{\frac{A(p,q)/mad(p,q)}{T(n)/mad(n)}} \qquad (7)$$

4.2. QP re-adjustment

We can extend the above idea to the re-quantization process of the current frame. During the iteration, we get actual number of bits for the QP of the previous pass. We then use quadratic model to calculate the updated QP for the new pass. In particular,

$$QP\_new = QP\_prev * \sqrt{\frac{Actual\ bit\_prev}{T\,arg\,et\ bit}} \qquad (8)$$

After final QP determination, we will do the actual encoding and update the parameters in our rate control. In particular, we need to check if frame dropping is needed to avoid buffer overflow after updating the current buffer status.

```
If(VBV_fullness >= Bs)  /* Bs is the VBV buffer size*/
        Drop currently encoded frame;
            /* VBV_fullness is the buffer status */                (9)
Else if(VBV_fullness > Bs*95%)
```

Drop next frame;

In MPEG-4 Annex L, when the buffer usage is over 80% of VBV buffer size, the next frame will be dropped. In our solution, since we have bit allocation guarantee, we can push this limit to a higher level.

5. SIMULATION RESULTS

We applied the above solution to MPEG-4 simple profile encoder and compared this with Annex L rate control. The basic coding settings are, combined mode with no error resilience, 10s I frame refresh time, four motion vectors in motion estimation with 16x16 motion search range and 0.85s VBV buffer size. We chose QCIF Foreman (400 frames), Stephan (300frames) and Glasgow (750frames) as the testing sequences with the framerate of 10f/s. The results are shown in Table1 and Fig.2.

Table 1: Comparison of the framerate, frame dropping and PSNR between the two solutions.

Sequence (bitrate)	Actual framerate(f/s)	Percentage of frame dropping	PSNR (db)	Gain in PSNR (db)
Foreman (48kb/s)	9.45 10.00	5.50% 0	29.97 30.39	+0.42
Foreman (64kb/s)	9.82 10.00	1.75% 0	31.42 31.67	+0.25
Stephan (64kb/s)	8.80 9.30	12.00% 7.00%	23.75 24.12	+0.37
Stephan (112kb/s)	9.50 10.00	5.00% 0	26.17 26.78	+0.61
Glasgow (64kb/s)	8.44 10.00	15.60% 0	27.68 28.60	+0.92
Glasgow (112kb/s)	9.36 10.00	6.40% 0	30.54 31.20	+0.66

Note: The first line data is for MPEG-4 Annex L, and the second line data is for the proposed solution. Here PSNR refers to the luminance PSNR. When frames were skipped, the respective previous encoded frames were used in the PSNR calculation.

Fig.2 shows the PSNR variation for the Foreman sequence.

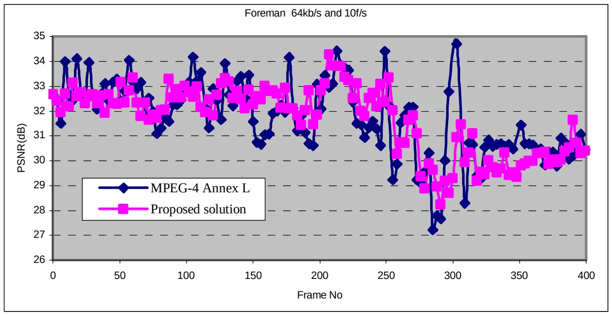

Fig.2 PSNR variation of the two solutions for the Foreman sequence. (Only the first and No.301 frame are I frames, and the rest are P frames.). The PSNR of the skipped frames are very low and not shown in this curve.

From the above table and figure, we can see the proposed solution not only achieves overall PSNR gain but also delivers more consistent visual quality across the sequence and temporally smoother video. Interestingly, the proposed solution achieves much better quality than MPEG-4 Annex L in the scene change frames of the sequence, and thus significantly reduces the annoying flickering effect. From the complexity point of view, we observe that on average, 25% frames need extra re-quantization.

6. CONCLUSION

There are two major contributions in this work. First, we have proposed a novel sequence based bit allocation model. Second, we proposed a solid mechanism to achieve the bit allocation, once determined. The proposed rate control solution can achieve constant quality video with less quality flicker and motion jerkiness.

7. REFERENCES

[1] ISO-IEV/JTC1/SC29/WG11, “Test Model 5”, Draft, Apr. 1993.

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

[3] N. Mohawnian, R. Rajagopalan and C. A. Gonzales, “Single -pass constant- and variable-bit-rate MPEG-2 video compression,”IBM J. Res. Dev.,vol.43, No. 4, pp489-509, Jul.1999.

[4] P. H. Westerink, R. Rajagopalan, and C. A. Gonzales, “Two-pass MPEG-2 variable-bit-rate encoding,” *IBM J. Res. Develop*., vol.43, No.4, pp. 471-488, Jul.1999.

[5] K. Ramchandran, A. Ortega and M. Vetterli, “Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders,” *IEEE Trans. Image Proc*., vol.3, no.5, pp. 533-544, Sep.1995.

[6] A. Ortega, K. Ramchandran and M. Vetterli, “Optimal trellis - based buffered compression and fast approximations,” *IEEE Trans. Image Proc*.,vol.3, no.1, pp.26-40, Jan.1994.

[7] J. Lee and B. W. Dickinson, “Temporally adaptive motion intgerpolation exploiting temporal masking in visual perception,” *IEEE Trans. Image Proc*., vol.3, no.5, Sep. 1994.

[8] 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.

[9] H.-M. Hang and J.-J. Chen, “Source model for transform video coder and its application – Part I: Fundamental theory, ” *IEEE Trans.Cir.Sys.Vid. Tech*.,vol.7, pp.287-298, Apr.1997.

[10] 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.

[11] J. Ribas-Corbera and S. Lei, “Rate control in DCT video coding for low- delay communications,” *IEEE Trans. Cir. Syst. Video Technol.*, vol.9, no.1, pp. 172-185, Feb. 1999.

[12] Bo Xie and Wenjun Zeng, “A sequence-based rate control framework for constant quality video”, submitted to *IEEE Trans. Circuits Syst. Video Technol.*