COST-EFFECTIVE FRAME-LAYER H.264 RATE CONTROL FOR LOW BIT RATE VIDEO

Changhyun Lee[1], *Seongjoo Lee*[2], *Yunje Oh*[3], *and Jaeseok Kim*[1]

[1]School of EEE, Yonsei University, 134 Sinchon-Dong, Seodaemun-Gu, Seoul, Korea 120749
[2]School of ICE, Sejong University, 98 Gunja-Dong, Gwangjin-Gu, Seoul, Korea 143747
[3]TN R&D Center, Samsung Electronics Co., Ltd., 416 Maetan-Dong, Paldal-Gu, Suwon, Korea 442742

ABSTRACT

The rate control is important to allocate bits efficiently for getting better performance, such as high quality, low fluctuation of quality and low mismatch between a target bit rate and an encoded bit rate. In this paper, we present an improved frame-layer H.264/AVC rate control scheme using enhanced complexity measure, scene-change detection, and quantization parameter adjustment for low bit rate video. Especially, our complexity measure contains not only the residual information but also the motion information. We use the motion vector difference encoded previously as motion information for low computation-load encoder. Experimental results show that Y-PSNR and bit-rate mismatch are remarkably improved and the fluctuation of Y-PSNR is similar as compared with the conventional H.264/AVC rate control scheme (JVT-G012).

1. INTRODUCTION

As the video streams are transmitted to the wired or wireless network, the rate control is necessary to adjust the encoded bit-streaming to certain channel bandwidth and buffer constraints. In general, the coding process related to the rate control of H.264 is to the followings: estimation of the number of target bits (target-bit), computation of quantization parameter (QP), execution of rate distortion optimization (RDO), and then obtainment of mean absolute difference (MAD). When the QP is computed, it needs the MAD for the current frame. However, we cannot obtain the MAD of current frame before the RDO processes which require the QP. This is a typical chicken and egg dilemma. To solve the problem, it has been proposed how to estimate the MAD for the current frame [1]. However, because it is difficult to estimate the current MAD exactly and a good performance cannot be expected with just simple MAD estimation, some approaches tend to measure the content complexity for applying it to the calculation of QP [2]-[4]. Even though the content complexity has to be measured by considering both residual and motion information, these schemes use only residual information. The pre-analysis like [5], which requires another motion estimation (ME) process for each macroblock (MB), may be the unique method to get the real motion information of current frame prior to the computation of QP. However, the pre-analysis is not suitable for the encoder requiring low computational load because the ME module is still the most time consuming part. Generally, the ME process takes more than about 20% of the total encoding time at one reference frame and 32-pixel search range [6]. Therefore, we present a complexity measure considering both the residual and motion information without increasing computational load significantly. We also show a method of detecting scene changes to prevent the buffer from overflowing. By using both the complexity measure and the detection method, we propose a cost-effective frame-layer H.264 rate control for low bit rate video.

2. PROPOSED FRAME COMPLEXITY MEASURE

In order to get better performance, the content complexity is measured in [2] as follows

$$CM_{conv,i} = \frac{PMAD_i}{\left(\frac{1}{i-1}\right)\sum_{j=1}^{i-1} CMAD_j} \qquad (1)$$

where $PMAD_k$ and $CMAD_k$ denote the predicted and the actual MAD computed in the kth P frame, respectively. The denominator of (1) indicates an average CMAD over the previous P frames. Because the motion information is not considered in (1), we use the number of motion vector difference bits encoded previously (MVD-bit) for avoiding significant computational loads. In MVD-bit the number of reference-frame bits is also included. MVD-bit is closely connected with motion complexity and also the proportion of MVD-bit to the total number of generated bits (generated-bit) for encoding each frame is high. Table 1 shows the proportion of MVD-bit to the generated-bit. We can see from Table 1 that the lower bit rate is, the higher the proportion of MVD-bit is. Therefore, the MVD-bit has to be considered in the complexity measure for low bit rate, and our complexity measure is given by

$$CM_{prop,i} = \varepsilon \times RatioMVD_i + (1-\varepsilon) \times CM_{conv,i} \qquad (2)$$

where ε is a weighting factor having 0.5.

Table 1. The proportion of motion vector difference header-bit to the generated-bit over 100 frames

Sequence	76.8kbps	38.4kbps	19.2kbps
Foreman	34.9%	43.0%	46.1%
Salesman	19.9%	25.1%	25.9%
Trevor	29.0%	34.3%	38.8%

$$RatioMVD_i = \frac{AMVD_{i-1}}{\left(\frac{1}{i-1}\right)\sum_{j=1}^{i-1} AMVD_j} \tag{3}$$

RatioMVD$_i$ is shown in (3) where *AMVD$_i$* denotes the average MVD-bit over all MBs of the ith frame.

3. PROPOSED DETECTION METHOD FOR SCENE CHANGE

When a scene is changed, the number of intra-coded MBs is increased and then the buffer is filled up instantaneously. In order to avoid buffer's overflow, both detection of scene change and appropriate rate control are necessary at scene change. In [3], the PSNR difference ratio is used in order to improve the video quality at scene change as follows

$$Diff_{ratio,i} = \frac{Diff(i,i-1)}{Diff(i-1,i-2)} \tag{4}$$

where *Diff(i,i-1)* denotes the PSNR difference between the ith and the (i-1)th frame. Unfortunately this method may make a wrong decision because $Diff_{ratio,i}$ becomes undesirably high if *Diff(i-1,i-2)* is very low in (4).

We present a new detection method for scene change solving the defect of (4). The scene change is decided by

$$RatioPSNR_i = \frac{PPSNR_{i,i-1}}{\left(\frac{1}{i-1}\right)\sum_{j=1}^{i-1} CPSNR_{j,j}} \tag{5}$$

where $PPSNR_{m,n}$ and $CPSNR_{m,n}$ denote the predicted and the actual PSNR computed for the mth frame with the nth reconstructed frame. We can see from Figure 1 that the values of *RatioPSNR* at scene changes are less than 0.4. The threshold of *RatioPSNR* to recognize a scene change is set to 0.5 with some margin. In addition, we use this detection method in determining QP for the frame where a scene is changed.

4. PROPOSED RATE CONTROL

The ultimate task of rate control is the obtainment of QP to get high quality, low fluctuation of quality, and low mismatch of bit rate. We select the frame-layer rate control to get a good trade-off between computational load and obtainment of precise QP and assume without loss of generality that the GOP structure is an IPPP…P where I and P denote an intra-coded picture and a forward predicted picture, respectively.

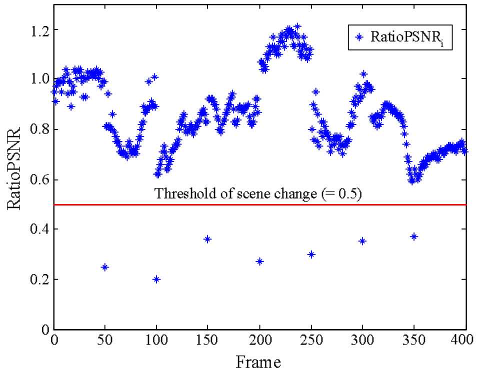

Figure 1. The *RatioPSNR* for the cascaded sequence where a scene is changed every 50 frames.

In order to obtain a fine QP for the current frame, the following three steps are needed: estimation of frame target-bit, computation of a QP, and adjustment of the QP. The estimation of frame target-bit in the proposed rate control scheme is based on the conventional rate control scheme [1]. In the computation and adjustment QP, we consider both the proposed complexity measure in Section 2 and detection method for scene change in Section 3.

4.1. Estimation of frame target-bit

To estimate frame target-bit, the number of remaining bits is needed and both the buffer fullness and the target buffer level are considered for avoiding overflow or underflow. According to [1], the target-bit, $T_{b,i}$, can be estimated before encoding the ith frame as follows

$$T_{b,i} = \beta \frac{R_{b,i}}{N_{\Pr,i}} + (1-\beta)\left[\frac{b_r}{f_r} - \Gamma(CBF_{i-1} - TBL_i)\right] \tag{6}$$

where $R_{b,i}$ and $N_{Pr,i}$ denote the total number of remaining bits for all non-coded P frames and the number of non-coded P frames before encoding the ith frame respectively, b_r and f_r denote the predefined bit rate and frame rate respectively, CBF_i denotes the current buffer fullness after encoding the ith frame, and TBL_i denotes the target buffer level before encoding the ith frame, and both β and Γ are constants and their typical values are 0.5 and 0.75, respectively.

4.2. Computation and Adjustment of QP

After estimating the target-bit, we compute and adjust QP with considering two cases.

4.2.1. Negative target-bit ($T_{b,i} < 0$)

We can see from (6) that the negative target-bit results from a big CBF_{i-1} in comparison with TBL_i, and a big CBF_{i-1} comes from a big generated-bit. Once the target-bit falls negative, it is urgent to make the target-bit to be positive for getting the better visual quality. Since the more generated-bit is expected in the higher frame complexity, we adjust the current QP based on the proposed complexity measure as shown in (7) where Q_i is the QP for encoding the ith frame.

$$Q_i = \begin{cases} Q_{i-1}+2 & CM_{prop,i} < 0.8 \\ Q_{i-1}+3 & 0.8 \le CM_{prop,i} < 1.4 \\ Q_{i-1}+4 & CM_{prop,i} \ge 1.4 \end{cases} \tag{7}$$

4.2.2. Positive target-bit ($T_{b,i} \geq 0$)

As the target-bit has the positive value, the QP can be computed by using the quadratic R-D model [7] as follows

$$\frac{T_{tb,i}}{PMAD_i} = \frac{x_1}{Q_{c,i}} + \frac{x_2}{Q_{c,i}^2} \tag{8}$$

where $T_{tb,i}$ is the estimated texture-bit which is the difference between the estimated target-bit and the number of header bits (header-bit) encoded previously for the ith frame, $Q_{c,i}$ is the computed QP for the ith frame, x_1 and x_2 are the first and the second-order coefficients, respectively. The estimated texture-bit may be below zero when the previously encoded header-bit is comparable to the estimated target-bit in the low bit rate. In that case, we limit the estimated texture-bit to one bit.

Since the computed QP (8) may oscillate noticeably for difficult sequences having rapid changes in content complexity, we limit changes in QP to no more than ±2 units between pictures as follows.

$$Q_{lm,i} = MAX\{Q_{i-1}-2, MIN\{Q_{i-1}+2, Q_{c,i}\}\} \tag{9}$$

The limited QP is then adjusted depending on whether a scene is changed or not as follows.

$$Q_i = \begin{cases} Q_{sc} & RatioPSNR_i \le 0.5 \\ Q_{nor} & RatioPSNR_i > 0.5 \end{cases} \tag{10}$$

When scene change is detected by using the method presented in Section 3, the QP is adjusted as given by

$$Q_{sc} = \begin{cases} MAX\{Q_{lm,i}+4, Q_0\} & (b_r/f_r) < 400 \\ MIN\{Q_0, Q_{lm,i}+4\} & elsewhere \end{cases} \tag{11}$$

where Q_0 is the predefined QP for the first frame in the GOP.

In the normal picture without scene change, the QP is adjusted differently with considering both buffer status and proposed complexity measure as shown in (12). When both the buffer status and the proposed complexity measure are low, the QP is decreased by one to improve the visual quality because it is expected that the generated-bit after encoding the current frame will be small. Also when both the buffer status and the proposed complexity measure are high, the QP is increased by one because it is expected that the target-bit will be negative for the next frame.

After encoding a frame, a linear regression method like [7] is used in order to update the parameters of linear prediction model for MAD, as well as x_1 and x_2 of quadratic R-D model (8) for the next frame. The generated-bit is also added to the current buffer fullness.

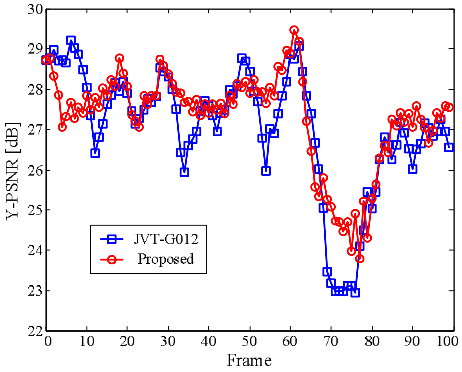

Figure 2. PSNR curve of ‘Foreman’ sequence with high motion, encoded at 19.2kbps/30fps.

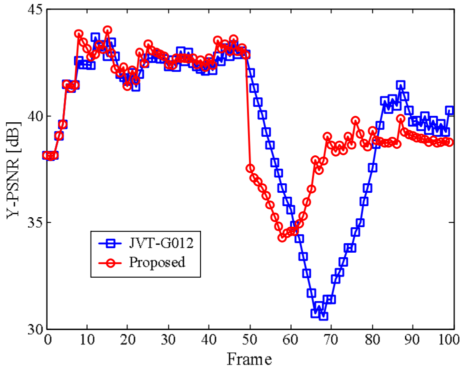

Figure 3. PSNR curve of ‘Akiyo_Container’ sequence with a scene change at the 50th frame, encoded at 76.8kbps/30fps.

5. EXPERIMENTAL RESULTS

We conduct numerous experiments on 23 test sequences (16 non-cascaded normal sequences and 7 cascaded sequences). The cascaded sequences by two normal sequences have a scene change which happens at the 50th frame. All the used test sequences are in QCIF 4:2:0 formats. We employ the H.264/AVC reference software version JM6.1 as the test platform and compare our algorithm with JVT-G012 by using the software [8]. For the tests, we set parameters to the followings: RDO is enabled, search range for ME is 16, the number of reference frames is one, and entropy coding method is CABAC. All other parameters are carefully selected for both algorithms to be equivalent. The initial QP for the first I frame and P frame is appropriately selected to avoid buffer’s overflow due to the first I frame.

Figure 2 and Figure 3 show the PSNR curve for sequences with high motion and a scene change, respectively. In Figure 3, ‘Akiyo_Container’ is made by concatenating ‘Akiyo’ and ‘Container’ for the scene change effect. We can see from these figures that the proposed

$$Q_{nor} = \begin{cases} Q_{lm,i}-1 & (CBF_{i-1}-TBL_i) < b_r/(f_r \times \Gamma)\, and (CM_{prop,i} < 0.8) \\ Q_{lm,i}+1 & (CBF_{i-1}-TBL_i) > b_r/(f_r \times \Gamma)\, and (CM_{prop,i} > 1.4) \end{cases} \tag{12}$$

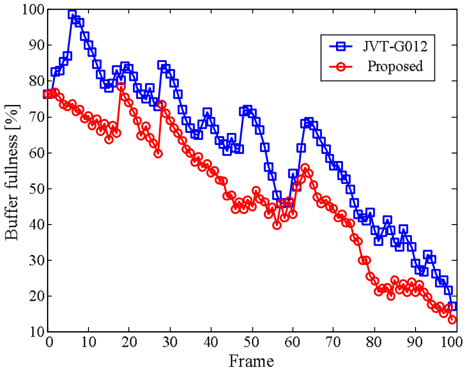

Figure 4. Buffer fullness of ‘Foreman’ sequence encoded at 19.2kbps/ 30fps.

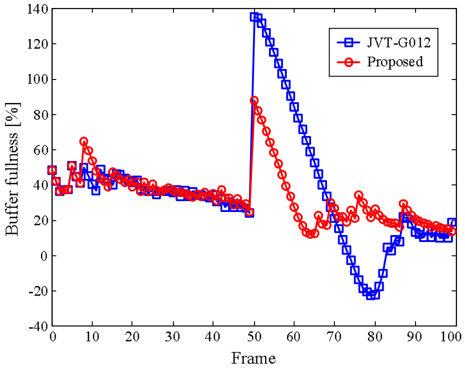

Figure 5. Buffer fullness of ‘Akiyo_Container’ sequence encoded at 76.8kbps/30fps.

frame-layer rate control can obtain better performance than that of JVT-G012. Figure 4 and Figure 5 show the buffer fullness at each frame. The buffer size is set to the half of target bit rate. According to these figures, the proposed rate control achieves steadier buffer fullness levels. In other words, the proposed algorithm is safer from the buffer’s overflow and underflow than JVT-G012. Table 2 shows the experimental results of both proposed algorithm and JVT-G012 for five out of 23 test sequences.

6. CONCLUSION

We proposed an improved frame-layer H.264/AVC rate control scheme. The improvement came from an enhanced content complexity measure, scene change detection, and quantization parameter adjustment as compared with the conventional H.264/AVC rate control scheme (JVT-G012). Especially, taking the computational loads into account, we used the previously encoded motion vector difference as motion information for obtaining complexity measure. In comparison with JVT-G012, we obtained the Y-PSNR improvement of 0.18dB and the mismatch improvement between target bit rate and encoded bit rate of 30.4% on average over 23 various test sequences, four bit rates and two frame rates. Also our standard deviation of Y-PSNR was almost similar to that of JVT-G012. The experimental results showed that the proposed rate control scheme was suitable for the real-time encoder supporting low bit rate video more efficiently because it had better performance than JVT-G012 without increasing hardware costs and computing powers significantly.

Acknowledgements: This research was supported by the University ITRC Project and partly by the TN R&D Center in Samsung Electronics Co., Ltd.

7. REFERENCES

[1] Z. G. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, “Adaptive basic unit layer rate control for JVT,” JVT-G012-r1, 7th Meeting, Pattaya II, Thailand, Mar. 2003.

[2] M. Jiang, X. Yi, and N. Ling, “Improved frame-layer rate control for H.264 using MAD ratio,” *IEEE International Symposium on Circuits and Systems*, vol. III, pp. 813-816, May 2004, Vancouver, Canada.

[3] X. Yi and N. Ling, “Rate control using enhanced frame complexity measure for H.264 video,” *IEEE Workshop on Signal Processing Systems*, pp. 263–268, 2004.

[4] M. Jiang and N. Ling, “On enhancing H.264/AVC video rate control by PSNR-based frame complexity estimation,” *IEEE Trans. Consumer Electronics*, vol. 51, pp. 281-286, Feb. 2005.

[5] L. Ping, X.K. Yang, and W.S. Lin, “Buffer-constrained R-D Model-Based Rate Control for H.264/AVC,” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 321-324, Mar. 2005.

[6] Z. B. Chen, P. Zhou, and Y. He, “Fast integer pel and fractional pel motion estimation for JVT,” JVT-F017, 6th Meeting, Awaji, Japan, Dec. 5–13, 2002.

[7] L. Hung-Ju, C. Tihao, and Z. Ya-Qin, “Scalable rate control for MPEG-4 video,” *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 10, pp. 878-894, Sept. 2000.

[8] http://ftp3.itu.ch/av-arch/jvt-site/2003_03_Pattaya/JVT-G012r1_software.zip

Table 2. Performance comparisons of the proposed algorithm with JVT-G012

Sequence	Target Bit rate /Frame rate	Average Y-PSNR			Standard Deviation of Y-PSNR		Encoded Bit rate	
		JVT-G012	Proposed	Gain	JVT-G012	Proposed	JVT-G012	Proposed
Carphone	9.6kbps/30fps	25.971	26.170	0.199	0.828	0.742	9.66	9.64
Foreman	19.2kbps/30fps	27.036	27.291	0.255	1.539	1.198	19.33	19.22
News	38.4kbps/15fps	36.786	37.182	0.396	1.702	1.825	38.52	38.63
Akiyo_Container	76.8kbps/30fps	39.568	40.037	0.469	3.631	2.706	77.51	76.89
MissAmerica_Carphone	19.2kbps/15fps	34.768	35.688	0.920	6.227	5.237	19.32	19.20