

LOW-COMPLEXITY RATE CONTROL FOR EFFICIENT H.263 TO H.264/AVC VIDEO TRANSCODING

Chi-Wang Ho[†] Oscar C. Au[‡] S.-H. Gary Chan[†] Shu-Kei Yip[‡] Hoi-Ming Wong[‡]

[†]Dept. of Computer Science [‡]Dept. of Electrical and Electronic Engineering

The Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, Hong Kong, China.

Email: {jodyho, eeau, gchan, sukuyip, hoimingw}@ust.hk

ABSTRACT

Rate control is a complicated problem in the H.264/AVC coding standard, extra computation is usually needed for the existing rate control schemes to estimate the complexity of frames or macroblocks (MBs). However, during transcoding, information from precoded video could be used to simplify the rate control. In this paper, we propose a low-complexity rate control scheme for transcoding from H.263 to H.264/AVC. The relationship between the rate of the precoded video and both the rate and distortion of the transcoded video are studied. By using only the rate information from the precoded video, we introduce a row-layer bit allocation and perform average rate shaping across a row of MBs. Estimation error diffusion is also introduced. The proposed scheme has sufficiently lower computational complexity than other methods as there is no explicit complexity measurement of MBs and complicated parameters updating. Experimental results show the effectiveness of the proposed scheme.

Index Terms— Video coding, video signal processing

1. INTRODUCTION

H.264/AVC is the most current international video coding standard jointly developed by ITU-T and ISO/IEC [1]. With various new tools to enhance the ability to predict the picture content, it outperforms all of the existing standards, such as H.263+ and MPEG-4, in terms of coding efficiency. Therefore, H.264/AVC is suitable for video application in all areas and intended to replace the existing standards.

Video transcoding facilitates the deployment of H.264/AVC by providing an efficient way to convert existing video content to and from H.264/AVC. During the transcoding process, it is important to regulate the bitrate to meet the rate constraints of the target application. An adaptive quadratic rate-distortion model has been proposed for H.264/AVC [2], and some well-known schemes for other existing coding standards, such as H.263+ TMN-8 [3], etc, can also be used as well. However, due to the rate-distortion optimization method for motion estimation and mode decision in H.264/AVC, the complexity of MBs cannot be easily determined for rate control algorithm. Either predictive or two-pass method is usually used to estimate the complexity of MBs. This introduces extra computation and may result in inefficient rate control because of the error of the estimated complexity of MBs.

Some studies have developed simplified and efficient rate control schemes for H.264/AVC transcoding by re-using the information

from the coded bitstream [4,5]. In [4], adaptive method for calculating QP for I and B frame is proposed, and in [5], the method applies on frame-level bit allocation and QP selection only. In this paper, we present a simple rate control scheme for H.263 to H.264/AVC transcoding, which aims at reducing computational complexity and maintaining high quality. We have studied the relationship between the rate of the precoded video and both the rate and distortion of the transcoded video. Based on this relationship, a MB-level rate control is developed by using the bit information from the precoded video to perform average rate shaping across a row of MBs. In addition, a row-layer rate control is introduced to allocate the target number of bits for each row and also distribute the estimation error by diffusion.

The rest of the paper is organized as follows. Section 2 discusses the observations we have and the relationship between the precoded video and the transcoded video in terms of both the rate and distortion. In Section 3, the proposed rate control algorithm is presented in details. The performance of the proposed rate control scheme is evaluated by comparing with adaptive quadratic rate-distortion model proposed for H.264/AVC [2], and end with concluding remarks in Section 5.

2. RATE AND DISTORTION OF TRANSCODED VIDEO

We are going to discuss our observations which lead to the proposed rate control. Then, statistical analysis will be given on the relationship between the rate of the precoded video and both the rate and distortion of the transcoded video.

2.1. Relationship between Precoded Video and Transcoded Video

The rate of the transcoded video usually depends on several factors, they are the video content, the quality of the precoded video and the QP for re-encoding. The basic idea of the proposed rate control is based on the following two observations: We first assume that the QP for re-coding is fixed. Therefore, the rate of the transcoded video, in general, is determined by both the content and quality of the precoded video. However, these two factors are closely related to the rate of the precoded video. In other words, the higher output rate will give a better quality of the output video, and more bits is usually needed for encoding a complex scene than a simpler one in order to achieve same level of quality in terms of PSNR. Therefore, it may be possible to establish a relationship between the rate of the precoded video and the rate of the transcoded video.

Second, we believe that the distortion of the transcoded video is somehow correlated with the rate of the precoded video. For example, in the precoded video, if more bits are used to code MB i than MB j , this means MB i contains more details or residual energy than

This work has been supported in part by the Innovation and Technology Commission (projects no. ITS/122/03 and GHP/033/05) and the Research Grant Council (DAG04/05.EG34) of the Hong Kong Special Administrative Region, China.

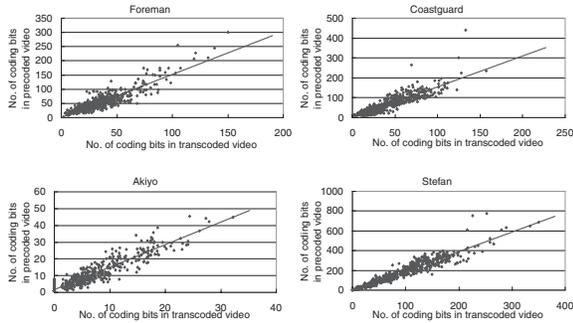


Fig. 1. Scatter plots of the rate of the precoded video against the rate of the transcoded video

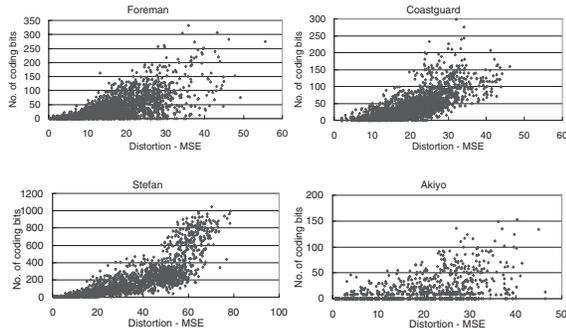


Fig. 2. Scatter plots of the rate of the precoded video against the distortion of the transcoded video

MB j . We expect that MB i is also relatively harder to predict than MB j in the re-encoding process. Therefore, MB i provides more room for introducing distortion than MB j .

Based on the above two observations, the rate of the precoded video will potentially give us some valuable information about the behavior of both the rate and distortion of the transcoded video which should be useful in developing a simple rate control algorithm.

2.2. Statistical Analysis

To study the relationship between the rate of the precoded video and both the rate and distortion of the transcoded video, four test sequences, including foreman, coastguard, stefan and akiyo, are coded into H.263 using fixed QP 8 at QCIF resolution and then transcoded into H.264/AVC using fixed QP 32 by their reference codec [6, 7]. Samples are taken at every fifth frames.

The relationship between the rate per frame of the precoded video and that of the transcoded video has been studied in [5] and frame-layer rate control is proposed based on their results. However, the frame-level relationship is not able to capture the variation of content within a frame. In order to understand the variation of content, we further study the relationship in a per-row basis. The reason of choosing per-row instead of per-MB is that different prediction modes supported in H.264/AVC introduce a large variation of the rate at MB-level. It is hard to approximate the relationship between the rate of the input and the output MB. Figure 1 shows four scatter plots between the average number of bits per row of the MBs in the precoded and transcoded video. The results reveal that they are likely to have a linear relationship. With the linear relationship, the

number of bits per row can be estimated from input video and the variation across the frame can be known. Due to space limitation, the graph for different QPs is not shown here, however, they show similar linear relationship with different slopes.

Similarly, Figure 2 shows some scatter plots between the rate of the precoded video and the distortion of the transcoded video. Both the rate and distortion are measured in per-MB basis. Two observations can be made from these figures. First, we observe that the MB which used more coding bits in the precoded video tends to have higher distortion in the transcoded video. This verifies that our second observation in previous sub-section. Our second observation is that there are some samples with very low coding bits in the precoded video, but large distortion is resulted after transcoding. This can be explained by investigating the skipped mode in both H.263 and H.264/AVC. When skipped mode is used in H.263 or H.264/AVC, the MB is copied from the previous frame so that the coding bits are effectively zero, except the bit used in header indicating the MB is being not coded in H.263. However, skipped mode still introduces distortion. The distortion depends on how good the MB copied from the previous frame, in other words, it depends on the distortion of the copied MB. It breaks the relationship between the rate and the distortion. Therefore, a large distortion is possible even the rate of the MB in the precoded video is very small.

3. PROPOSED RATE CONTROL

Having illustrated the relationship between the rate of the precoded video and both the rate and distortion of the transcoded video, we propose a simple and yet efficient rate control scheme for fast transcoding based on the number of bits spent to code the MBs of the precoded video. The proposed rate control scheme attempts to regulate the bitrate of the transcoded video such that the average rate variation across a row and a frame is similar to the precoded video. In other words, more bits will be allocated to the MB which uses more bits in the precoded video. The details will be presented in section 3.3. In addition to the frame-layer bit allocation and the MB-layer rate control, the row-layer bit allocation is introduced. It selects the target number of bits for each row based on the relationship between the precoded rate and the target rate. Also, after encoding all the MBs in a row, we handle the estimation error between the estimated and the actual encoded number of bits by evenly distributing the error to the remaining MBs in the frame. This can avoid sudden change in QP due to unexpected high or low bits used for some MBs.

We first define N_{row} and N_{col} be the number of row and column of MBs in a frame respectively. And, $b_{i,j}$ and $\tilde{b}_{i,j}$ is the number of bits used for coding the MB at row i and column j in the precoded video and the transcoded video respectively.

3.1. Frame-Layer Bit Allocation

Many papers suggest that the number of bits allocated for a given frame should depend on its complexity. Therefore, the available number of bits can use more effectively over a video sequence. However, these methods allow a certain degree of bit fluctuation over a short period of time and tolerate for longer delay. In the proposed rate control scheme, a simple frame-layer bit allocation used in TMN-8 is adopted to allocate the target number of bits B for each frame depending on the buffer fullness W [3].

$$B = \frac{C}{F} - \Delta, \quad (1)$$

where C and F are the channel and frame rate respectively and Δ is defined as follows:

$$\Delta = \begin{cases} W/F & W > ZM, \\ W - ZM & \text{otherwise.} \end{cases} \quad (2)$$

By default, $Z = 0.1$ and $M = C/F$. This feedback mechanism maintains steady buffer occupancy.

3.2. Row-Layer Weighted Bit Allocation and Estimation Error Diffusion

The proposed row-layer rate control consists of two parts, they are the weighted bit allocation and the estimation error diffusion. The weighted bit allocation performs before coding the first MB of each row. With the linear relationship of the bits per row of MBs between the precoded video and the transcoded video, it assigns the target number of bits for each row according to the variation of bits per row of the precoded video. Then, the initial target average number of bits for each row is computed. After finished coding of each row, the estimation error diffusion is used to distribute the estimation error to the remaining non-coded MBs in the current frame and also updates the target average number of bits for each of the remaining row.

3.2.1. Weighted Bit Allocation

The initial target number of bits for i^{th} row, B^i , is calculated as follows

$$B^i = \alpha \times \sum_{l=1}^{N_{col}} b_{i,l} \quad (3)$$

where $\alpha = \frac{B}{\sum_{k=1}^{N_{row}} \sum_{l=1}^{N_{col}} b_{k,l}}$ is the ratio of the target number of bits of the current frame to the total number of bits its spent in the precoded video. Then, the initial target average number of bits for i^{th} row, \bar{B}^i , is calculated by

$$\bar{B}^i = \frac{B^i}{N_{col}} \quad (4)$$

For a given row, this value is updated after encoding each MB on that row. In the MB-layer rate control, the QP is determined in order to keep this value close to the one calculated based on the precoded video as shown in Section 3.3.

3.2.2. Estimation Error Diffusion

The error between the estimated and achieved number of bits, e , is usually controlled by involving the available number of bits for remaining MBs in QP calculation. The QP will be adjusted for the subsequent MBs if the number of bits spent by the current MB is unexpectedly large or small. The change of QP introduces some overhead and affects the quality of the immediately subsequent MBs. The estimation error diffusion will diffuse or distribute e to the remaining MBs of the current frame equally at the end of each row. This has an advantage over the traditional method that the impacts on error will be shared by all the remaining MBs. Let $N_{non-coded}$ be the number of non-coded MBs of the current frame. The error per non-coded MB is given as

$$\bar{e} = \frac{e}{N_{non-coded}} \quad (5)$$

And, the initial target average number of bits for the remaining row will be updated.

$$\bar{B}^i = \bar{B}^i + \bar{e} \quad (6)$$

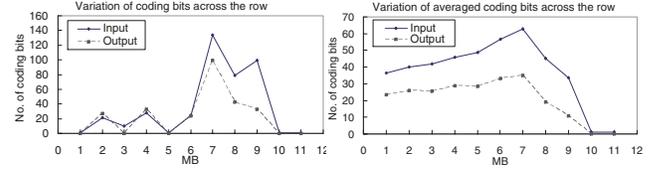


Fig. 3. The variation of coding bits (left) and the one of averaged coding bits (right) across a row

3.3. Macroblock-Layer Rate Control

As shown in section 2.2, the distortion after transcoding tends to be larger if the input MB spent more bits to code. So, it would be better to preserve the shape of the rate usage or rate variation across a row, even across a frame. In Figure 3, an example of coding bits for each MB across a row in the stefan sequence is shown on the left. It can be noted that the fluctuation is pretty large as there are some MBs used zero bits. However, the plot on the right shows another representation of the same row with a smoothed curve. The value of y-axis for MB j is the average number of coding bits from MB j to the end of the row. The proposed MB-layer rate control is based on this average number of coding bits and attempted to shape the average number of bits of the output video such that it is similar to the one of the precoded video across a row. The average number of coding bits from the j^{th} MB on row i to the end of the row of the precoded video is formally defined as

$$\hat{b}^{i,j} = \frac{\sum_{k=j}^{N_{col}} b_{i,k}}{N_{col} - j + 1} \quad (7)$$

A simple QP selection scheme is based on the QP of previous MB, QP_{prev} , the target average number of bits for j^{th} MB on row i , $\bar{B}^{i,j}$, and the possible error correction factor, δ . We define

$$\bar{B}^{i,j} = \begin{cases} \bar{B}^i & j = 1, \\ \frac{\bar{B}^{i,j-1} \times (N_{col} - j + 1) - \hat{b}_{i,j-1}}{(N_{col} - j)} & \text{otherwise,} \end{cases} \quad (8)$$

The above equation updates the target average number of bits after encoding a MB on row i . The QP for $j + 1^{th}$ MB on row i is chosen according to the following conditions

$$QP_{j+1} = \begin{cases} QP_j + 1 & \bar{B}^{i,j+1} < \alpha \hat{b}^{i,j+1} - \delta, \\ QP_j - 1 & \bar{B}^{i,j+1} > \alpha \hat{b}^{i,j+1} + \delta, \\ QP_j & \text{otherwise,} \end{cases} \quad (9)$$

where $\delta = \arg \max_{k \geq j} \alpha \hat{b}^{i,k} \times 0.5$. The starting QP for the first MB is simply specified by the user or it can be determined by the method proposed in [5]

In the above equations, the QP increases (decreases) when the target average number of bits for the remaining MBs is less (greater) than the value obtained from the precoded video by Eq. (7) minus (plus) δ . The δ is the half of the maximum average number of bits for the remaining MBs of the row i of the precoded video, in other words, this is the maximum allowable error which is expected to be able to compensate in the remaining MBs of this row. By selecting the QP using Eq. (9), QP will adjust gradually and match the average number of bits of row i in the transcoded video with the precoded video. This can avoid a noticeable different quality between adjacent MBs in a row. Additional constraints are added in frame-level and row-level for the same reason. The difference between the QP of

Table 1. Comparison of performance achieved by JVT-G012r1 and the proposed scheme for QCIF sequences to 64kbps and 96kbps

Sequence	Achieved bitrate (kbps)		PSNR (dB)		Δ PSNR
	JVT	Proposed	JVT	Proposed	
Foreman	64.04	64.01	32.45	32.67	+0.22
QP=18	96.00	96.05	34.55	34.72	+0.17
Foreman	64.04	64.06	31.22	31.51	+0.29
TMN8-384kbps	96.00	96.02	33.30	33.51	+0.21
Akiyo	64.07	62.73	45.34	46.35	+1.01
QP=18	96.09	95.22	48.86	49.59	+0.73
Akiyo	64.12	63.00	40.38	40.98	+0.60
TMN8-384kbps	96.13	95.59	42.34	42.86	+0.52
Children	64.14	63.80	27.11	27.23	+0.12
QP=18	96.24	95.77	29.59	29.87	+0.28
Children	64.14	63.45	26.03	26.14	+0.11
TMN8-384kbps	96.20	95.42	28.75	29.04	+0.29
Stefan	64.03	64.36	24.94	25.00	+0.06
QP=18	96.05	96.34	26.59	26.64	+0.05
Stefan	64.07	64.30	24.17	24.30	+0.13
TMN8-384kbps	96.06	96.24	25.96	26.17	+0.21

Table 2. Comparison of performance achieved by JVT-G012r1 and the proposed scheme for CIF sequences to 256kbps and 384kbps

Sequence	Achieved bitrate (kbps)		PSNR (dB)		Δ PSNR
	JVT	Proposed	JVT	Proposed	
Dancer	256.34	256.29	36.51	36.67	+0.16
QP=18	384.65	384.21	38.88	39.01	+0.13
Dancer	256.44	256.05	35.74	35.86	+0.12
TMN8-512kbps	384.55	384.06	37.92	38.11	+0.19
Flower	256.34	256.20	24.44	25	+0.56
QP=18	384.15	384.11	26.12	26.49	+0.37
Flower	256.33	256.36	24.85	25.33	+0.48
TMN8-512kbps	384.15	384.21	26.43	26.72	+0.29

the first MB in the adjacent rows is limited by ± 2 . And the QP of current frame is clipped by the average QP of the previous frame, $\bar{Q}P$, as follows

$$QP = \max(\min(QP, \bar{Q}P + 4), \bar{Q}P - 4) \quad (10)$$

4. EXPERIMENTAL RESULTS

In this section, the performance of proposed the rate control algorithm is evaluated and compared with the rate control scheme proposed for H.264/AVC in JVT-G012r1 [2]. A cascaded transcoder was implemented using H.263 and H.264/AVC reference codec [6, 7]. The input H.263 video bitstreams are encoded at 30fps either with fixed QP=18 or with TMN-8 enabled. This is used to verify the effectiveness of the proposed algorithm in both situations. The first frame is encoded as I frame with QP = 13. Then, the bitstream is converted into H.264/AVC format. The output are encoded with one reference frame, CAVLC, RDO disabled, search range ± 16 , all INTER mode enabled and FME enabled. Four QCIF sequences and two CIF sequences are tested.

Table 1 and 2 show the actual encoded bitrate and the PSNR achieved by the proposed and JVT-G012r1 methods. A gain in PSNR by the proposed algorithm over JVT-G012r1 method, up to 1.01 dB,

is observed. This is probably because the proposed algorithm takes the variation of complexity across a frame into account. With the knowledge of the variation of complexity, this avoids the bits used on the MB with low complexity. In terms of encoded bitrate, the proposed algorithm performs similarly and consistently as the JVT-G012r1 method. However, in Akiyo sequence, the achieved bitrate is a bit lower than the target bitrate because the average number of bits per row is too small from the precoded video. However, the proposed rate control scheme can achieve a higher PSNR than JVT-G012r1 although the bitrate is a bit lower than expected.

5. CONCLUSIONS

We have presented a low-complexity yet efficient rate control scheme for transcoding from H.263 to H.264/AVC using the rate information from precoded video. A rate control algorithm with additional row-layer rate control is proposed with no explicit MB's complexity measurement. The knowledge of the rate variation across the frame is utilized to assign the bits in a per-row basis. Also, the estimation error is shared by all the non-coded MBs with the estimation error diffusion. The experimental results can be used to verify the effectiveness of our scheme. In comparison with the rate control proposed in JVT-G012r1, the proposed rate control can achieve higher PSNR and require a significantly lower computational complexity.

Acknowledgment

This work has been supported in part by the Innovation and Technology Commission (projects no. ITS/122/03 and GHP/033/05) and the Research Grant Council (DAG04/05.EG34) of the Hong Kong Special Administrative Region, China.

6. REFERENCES

- [1] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC)*, May 2003, JVT-G050.
- [2] Zhengguo Li, Feng Pan, Keng-Pang Lim, Genan Feng and Xiao Lin, and Susanto Rahardja, *Adaptive Basic Unit Layer Rate Control for JVT*, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Mar. 2003, JVT-G012-r1.
- [3] J. Ribas-Corbera and Shawmin Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 172–185, Feb. 1999.
- [4] You-Neng Xiao, Hong Lu, Xiangyang Xue, Viet-Anh Nguyen, and Yap-Peng Tan, "Efficient rate control for MPEG-2 to H.264/AVC transcoding," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, May 2005, vol. 2, pp. 1238–1241.
- [5] Viet-Anh Nguyen and Yap-Peng Tan, "Efficient H.263 to H.264/AVC video transcoding using enhanced rate control," in *Proceedings of the IEEE International Conference on Image Processing*, Sept. 2005, vol. 3, pp. 652–655.
- [6] Image Processing Lab, University of British Columbia, "TMN (H.263+) encoder/decoder, version 3.2," Sept. 1997.
- [7] "H.264/AVC Reference Software JM8.6," Oct. 2004.