We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Rate Control for Low Delay Video Communication of H.264 Standard

Chou-Chen Wang and Chi-Wei Tung
I-Shou University
Taiwan

1. Introduction

The demand for multimedia services is increasing rapidly in the last years. Therefore, efficient video compression has become a very important research for the multimedia communication. H.264 is the state-of-the-art digital video coding recommendation, which is also known as H.264/MPEG-4 Advanced Video Coding (H.264/AVC) (ITU-T, 2003). The standard is a joint collaborative effort between the ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG). The team responsible for the development and evolution of the standard is known as the Joint Video Team (JVT) and officially the standard is known as H.264 by the ITU-T and MPEG-4 Part 10 by ISO/IEC. The H.264 standard can achieve much higher coding efficiency than the previous standards such as MPEG-1/2/4 (LeGall, 1991; ISO/IEC, 1994 & ISO/IEC, 1999) and H.261/H.263 (CCITT, 1990 and ITU-T, 2003). In addition to coding efficiency, the rate control also plays a key role in a video encoder for multimedia services, especially for real-time communication such as video streaming, video conference and video surveillance. The number of bits required for encoding a video sequence varies with time to provide consistent visual quality because complexity of each frame generally differs from the other frames in the input sequence. Therefore, a rate control scheme which meets a constrained channel rate by controlling the number of generated bits is necessary in an encoder. Nowadays, real-time video streaming scenarios requiring very low end-to-end delay are getting more and more popular. However, it is very difficult to adjust the encoding parameters directly to obtain fixed bits for every encoded frame in the constant bit rate channel. Therefore, it is necessary that the buffer to regulate the bit stream before transmission. With a good rate control technique, it should adjust the output rate to prevent the buffer from overflow and underflow. If the buffer suffers from overflow and underflow, it will cause frames skipping and wastage of channel resource, respectively. Furthermore, the size of buffer is usually very small to achieve low end-to-end delay requirement for real-time communication. It causes the buffer overflowing and underflow easier. So, the low delay video communication requires more accurate bit allocation and encoder parameter adjustment to achieve a suitable rate control.

There are two parts that should be considered when designing a rate control scheme. One is about the bit allocation for each basic unit according to its complexity. The other is the adjustment of the encoder parameter, *i.e.*, quantization parameter (QP) to encode each basic unit to match target bits. Rate control scheme have been widely studied in video standards, such as TM5 for MPEG-2 (ISO/IEC, 1993), TMN8 for H.263 (ITU-T, 1997), and VM-8 for

MPEG-4 (ISO/IEC, 1997). Figure 1 shows the rate control scheme for MPEG-2, H.263 and MPEG-4 using rate-distortion (R-D) model. The amount of encoding bits of the current basic units (macroblock: MB) is predicted from the recent encoded basic units. The encoder shown in Fig. 1 can obtain the motion vectors (MV) using motion estimator (ME) and calculate the statistical data of the residual frame with actual mean absolute difference (MAD) after motion compensation (MC). And then, the rate controller can adjust the quantization parameter (QP) according to the rate-quantization (R-Q) model.
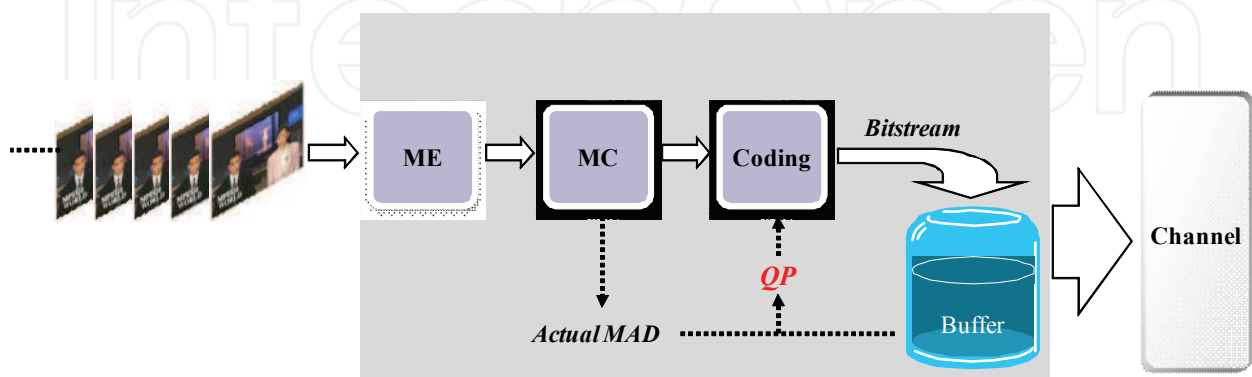

Fig. 1. Rate control scheme for MPEG-2, H.263 and MPEG-4

Compared with these previous standards, there is an additional problem for rate control in H.264 as shown in Fig. 2. The problem is due to that the H.264 encoder determines motion information by using the rate-distortion optimization (RDO) calculations. Before performing RDO for each MB, the quantization parameter should be defined by using MAD of MB. However, the statistical MAD of MB is only available after performing RDO. This is typical chicken and egg dilemma. Therefore, the rate control scheme is more difficult in H.264 (Li, et al., 2003).
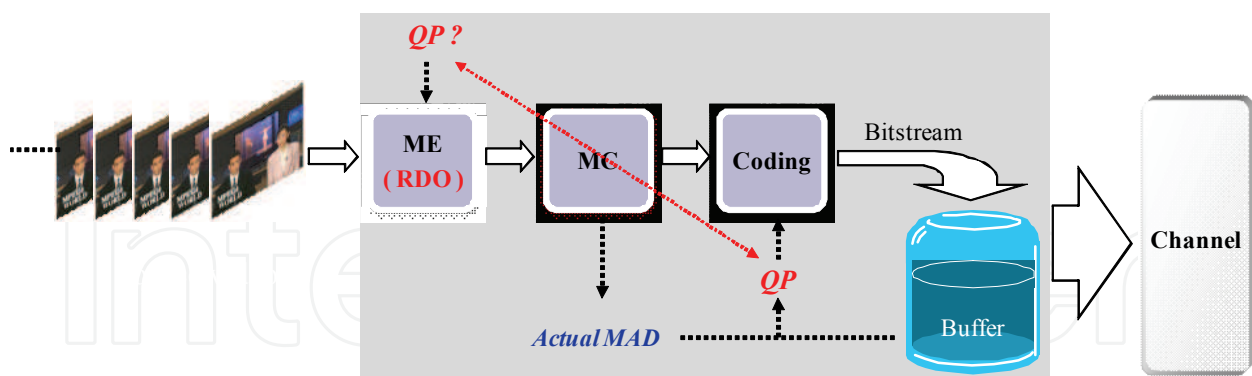

Fig. 2. The problem of QP dilemma for rate control scheme in H.264

In order to solve the QP dilemma problem caused by RDO, one rate control scheme was proposed in JVT-G012 (Li, et al., 2003) and was adopted by the JVT in the H.264 reference model JM 12.1 (JVT). JVT-G012 utilizes the method of temporal MAD prediction and R-Q quadratic model to achieve rate control. However, there are three problems existing in this scheme.

1. *Inaccurate initial quantization parameter*: For real time video applications, an improper initial QP maybe lead to the buffer overflowing and underflow seriously the front frames. It will affect continuity, quality and the demand of low delay directly.

2. *Inaccurate overhead bit prediction*: Due to a much more complex motion compensation strategy adopted by inter/intra coding mode in the H.264, the bits of overhead may highly fluctuate at the differential modes. Therefore, it is inaccuracy to predict overhead bit rate by using an average overhead bits in the JVT-G012.

3. *Inaccurate MAD prediction*: If the high motion or scene changes in the video sequences, the temporal correlation is reduced in the presence of such sudden changes. Therefore, only using the temporal MAD prediction model in the JVT-G012 is poorly in this status.

In this chapter, we will introduce a new rate control scheme for low delay H.264 video communication. A fast and best selection of initial QP is first proposed in the GOP layer rate control (Armstrong, et al. 2006 and Wang, 2008). Then, an improved MAD prediction model and overhead bits prediction method is adopted in the MB layer rate control. The simulation results show that the proposed scheme gives an average PSNRY gain of about 0.55 dB and 0.58 dB when compared with JVT-G012 and the method proposed in (Jiang & Lin, 2006), respectively. In addition, the proposed scheme improves the number of frame skipped and reduces the quality deviations of the initial frames by choosing the best initial QP.

2. Improved rate control in H.264

2.1 Efficient selection of initial QP

In JVT-G012, the I frame and the first P frame of the GOP are coded by initial QP. Therefore, the initial QP has a great effect in the front frames. The bad selection of initial QP may highly exceed buffer budgets so that the encoder attempting to salvage the over-spend bits in later frames. It leads to decrease rapidly the quality in later frames and even frame skipped. Therefore, according to channel bandwidth and complexity of encoded sequence to choose the best quantization parameter is a very important issue to be overcome. To increase the accuracy of selection, (Armstrong et al. 2006) proposed a selection of initial QP based on binary search scheme. As we know that the optimal QP setting can be obtained by full search for all QP indexes. To reduce the cost time of the full search, they use the binary search algorithm (BSA) to obtain the initial QP. For an example, assuming the bit-rate at QP=1 is 1,040 kbps and QP=51 is 20 kbps, the following example shown in Fig. 3 demonstrates the binary search for a desired target rate of 128 kbps on the first frame of a sequence. The QP=47 is determined as an initial QP to meet the channel bandwidth. The BSA can achieve the selection of QP with 6 processes, while a full search would require 51 processes. By using BSA, we can decrease the huge processing time rapidly.
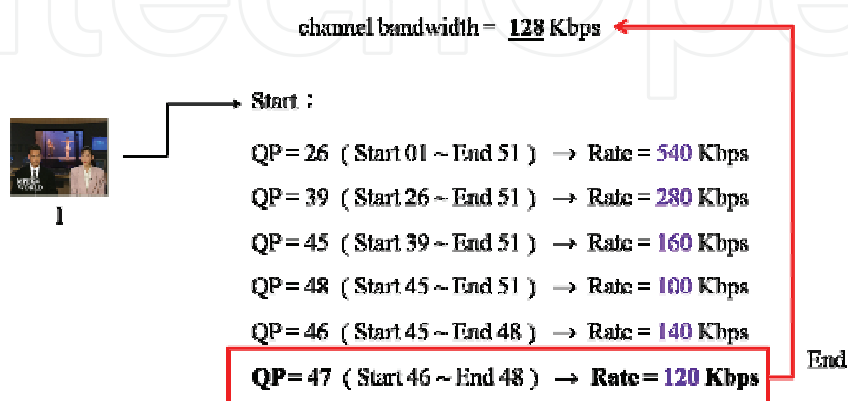

Fig. 3. Example of BSA for a desired target rate of 128 kbps.

However, the time of the search processes is still expended too much for each intra frame of a sequence (or on the first intra frame of each sequence in a group). To further speed up the best decision of initial QP, we propose a fast intra mode decision (FIMD) method to combine the BSA. In our previous research (Wang et al, 2006), we have proposed a fast intra mode selection algorithm for H.264 standard that take advantage of the correlation between MBs/blocks. Since H.264 is a block-based coding scheme, the frame is encoded block by block in a raster scan order, *i.e.*, from the left to right and top to bottom. For a luma MB in an I-slice, RDO exhaustively searches the combinations of the predefined 13 intra modes to produce the best mode for this MB. Figure 4 shows part of the RDO intra-mode map of an I-frame (News video sequence) conducted by the JM 12.1 (JVT) with RDO procedure. Each point (location) in the two intra-mode maps corresponds to a luma Intra_16×16 MB and a luma Intra_4×4 block, respectively. Since MBs/blocks are highly correlated, many MBs/blocks in I-frame correspond to the same modes. In other words, many points in the map have same mode, as shown in Fig. 4.

2	1	1	2	1	1	1	1	1	1	1
0	3	0	2	2	1	3	1	2	1	1
0	0	0	2	3	2	1	1	0	2	1
0	0	0	2	1	2	2	2	3	0	3
0	1	0	3	2	1	2	0	0	0	0
0	2	0	3	3	1	1	1	0	2	3
0	0	2	1	2	3	3	1	1	3	3
0	2	2	1	1	1	3	3	2	3	3
2	2	3	2	0	2	0	0	3	1	0

(a)

2	1	1	1	1	1	1	1	2	1	2	1	8	1	1	1	1	1	1	1	1	1	1	1
0	0	4	5	6	6	1	0	0	8	0	7	4	4	4	0	0	1	1	1	1	1	1	0
0	4	4	4	4	4	6	0	0	0	0	0	4	4	4	5	0	1	1	1	5	0	1	2
0	0	2	4	6	6	8	0	0	7	0	0	6	5	4	4	0	0	0	0	1	1	1	1
2	2	2	2	2	2	2	0	0	0	0	0	0	5	7	4	0	0	1	1	1	1	1	1
2	2	2	2	2	2	0	0	0	0	0	0	0	0	5	3	0	6	1	1	1	1	1	1
2	2	2	2	2	2	1	0	0	0	0	3	1	0	0	0	0	0	0	0	0	1	1	1
2	2	2	2	2	2	1	0	0	0	0	0	0	3	0	0	0	5	1	4	0	0	0	0
0	2	2	2	2	2	2	2	0	0	0	0	4	0	1	0	0	0	7	4	0	1	1	1
0	0	1	1	2	2	2	2	0	0	0	0	6	6	6	0	0	0	0	5	5	1	1	1
0	1	1	1	2	2	2	2	0	0	0	0	1	1	6	1	0	0	0	1	3	4	1	0
0	0	0	0	2	2	2	2	0	0	0	0	1	1	1	0	0	3	4	0	5	1	1	1
2	2	2	2	2	2	2	2	0	0	0	7	1	1	0	0	0	0	6	2	1	1	6	3
2	2	2	2	2	2	2	2	0	0	0	0	6	1	1	0	0	0	0	0	0	0	8	4
2	2	2	2	2	2	2	2	0	0	0	2	1	1	1	8	0	0	3	8	0	7	0	0
2	2	2	2	2	2	2	2	0	0	0	0	0	0	8	0	0	3	0	4	0	0	0	0
0	0	0	0	1	1	1	0	0	0	0	0	0	8	8	8	0	5	1	4	0	1	1	0
0	1	0	1	2	1	1	0	0	0	0	8	8	8	8	8	0	0	8	7	8	1	1	8
0	0	0	0	1	1	8	8	0	0	0	3	3	3	7	0	0	0	0	0	1	1	1	1
2	1	8	8	3	8	8	8	0	5	0	0	0	0	0	0	0	3	8	7	1	1	1	1

(b)

Fig. 4. RDO intra-mode map for an I-frame (a) luma Intra_16×16 (b) luma Intra_4×4

According to the observation of intra prediction modes (including luma Intra_4×4 and luma Intra_16×16) of any MB and those of its four neighboring blocks from different real video sequences, we find that there are a high mode correlation exists in intra mode map of H.264. We exploit the interblock correlation in the intra mode domain to early terminate the RDO calculations. Four modes of neighboring coded macroblocks/blocks shown in Fig. 5 are

Fig. 5. Four causal neighboring modes of the current block (a) 16×16 MB (b) 4 ×4 block

considered as the good candidate intra modes of the current block in I frame, and we can use the threshold combination to achieve early detection. A detailed description of the method is available in (Wang et al, 2006). In order to evaluate the performance, we use the News in QCIF (176×144) format and 15 Hz frame rate as testing sequence. We compare the results using different methods in terms of PSNR and bitrate. Figure 6 demonstrates the performance of the proposed decision of initial QP is almost the same as BSA under 24 kbps for channel rate. In addition, from the Table 1, we can find that the decision time of initial QP using the proposed method is obviously less than the BSA. The simulation results show that the proposed method achieves an average 62% computational saving compared to BSA method.

News QCIF 15Hz

(a) PSNRY vs. Frame

News QCIF 15Hz
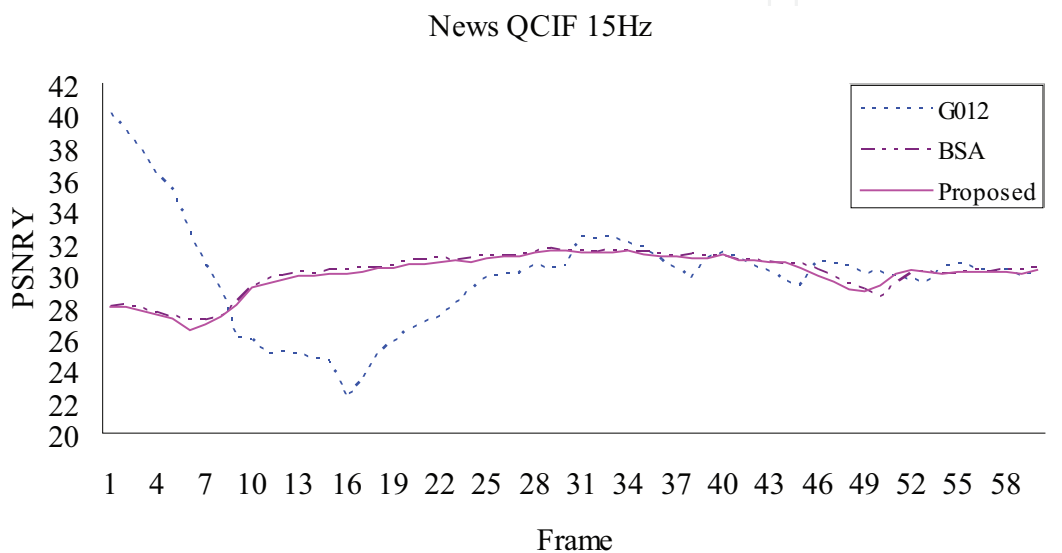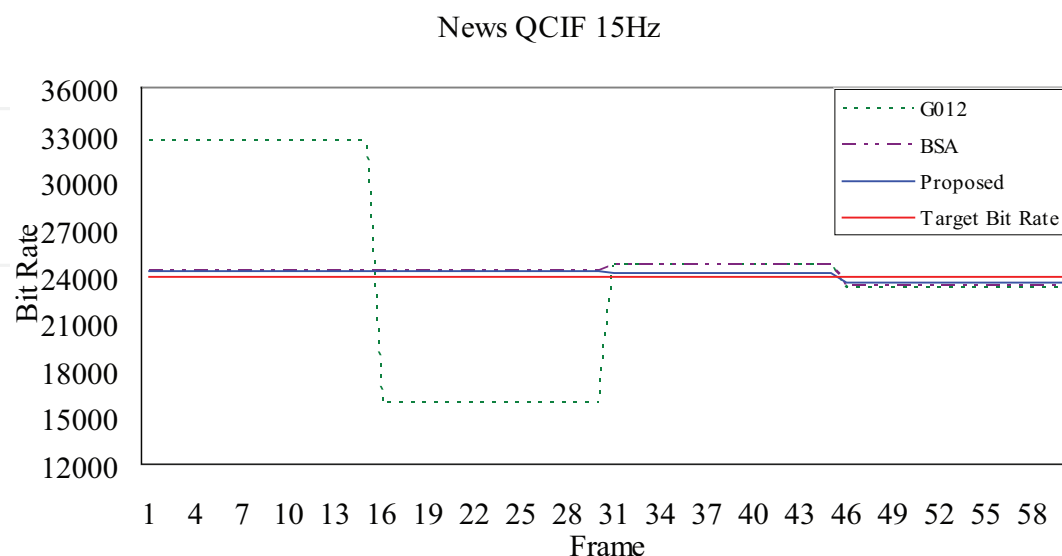
(b) Bitrate vs. Frame

Fig. 6. Comparisons of the PSNRY(dB) and bitrate for initial QP by different methods

Method \ Sequence	News	Forman	Bus
Full Search	44.3	42.2	49.7
BSA	3.3	3.3	3.4
Proposed	1.2	1.1	1.2

Table 1. Comparisons of processing times (seconds) for initial QP using various sequences

2.2 Improved overhead bit rate prediction

The overall bitstream generated by the source code is mainly comprised of texture bits and overhead bits. And texture bits are used to recode MB's residua after motion compensation. The overhead bits are used for differential coding and the overhead bits are used to record importance information such as the MB mode, the motion vector and the QP. The overhead bit prediction of rate control schemes in previous standards (including MPEG-1/2/4 and H.261/H.263) usually adopt the average overhead bits to predict, and update it with the average overhead bit rate after coding a MB or a frame. As we know that a more complex motion compensation adopted by the H.264 will lead to the overhead bits fluctuate at the differential mode.Unfortunately, the JVT-G012 rate control scheme follows that way and it produces large prediction error.

From (Yuan, et al., 2006), we can observe that the overhead bit rate is usually close among spatially and temporally adjacent MBs. Figure 7 shows the correlation of overhead bits count for two successive P frames by the Forman sequence in QCIF format, 15 Hz and initial QP=32. A further experiment is conducted to justify the temporal correlation and the typical experiment results are shown in Fig. 8. The experiment is simulated by searching for minimum difference of overhead bits count between the current MB and all of MBs in the previous P frame with a distance to the same position of the current MB not exceeding the searching radius.

47	10	10	22	14	3	49	20	8	13	44
28	9	11	9	3	3	2	56	2	3	2
1	2	4	10	**25**	38	11	44	45	2	8
0	1	10	30	31	12	27	36	82	8	8
9	10	2	15	42	19	37	77	19	18	8
5	12	2	31	12	19	29	71	15	3	43
2	2	8	24	10	9	6	58	15	13	18
1	**0**	4	4	14	8	11	7	3	**1**	11
1	0	14	21	28	10	23	27	8	14	12

(*j-1*)th frame

42	24	4	10	20	7	12	77	3	24	5
6	22	3	22	10	3	2	75	6	29	3
13	17	2	14	**26**	24	18	32	19	65	7
11	12	3	56	70	13	15	23	61	20	2
15	19	2	31	62	31	49	21	39	18	3
6	3	9	21	20	12	12	23	19	20	2
1	1	2	27	89	30	6	8	32	16	14
1	**0**	1	12	93	9	37	24	4	**2**	14
0	1	15	17	41	12	27	67	48	15	9

*j*th frame

Fig. 7. Temporal correlation of overhead bits count for two successive P frames

From the Fig. 8, we can observe that the relation between the MB percentage and the difference of overhead bits count in the different searching radius. When the searching radius is increasing, the probability of the similar overhead bits count and the searching complexity are also increasing at the same time. Furthermore, if the searching radius is other

than 0, we will get confused with the choice among more than one MB. In order to simplify the problem, (Yuan, et al., 2006) predict the overhead bits count of not yet coded MB directly by that at the co-located position in the previous P frame. However the accuracy of prediction is limited by using their method. In addition, we can also find when the radius more than two, the cumulated probability distribution (CDF) of the overhead bit rate correlation is almost the same. So, we select the searching radius equal to 2 (R=2) to achieve a trade-off between the accuracy of prediction and the searching complexity. In order to overcome the confusion with the choice among more than one MB, we will make use of the rough motion compensation information of MB to determine the overhead bit rate prediction in the searching radius. The rough MAD will be explained in the subsection 2.3 explicitly.
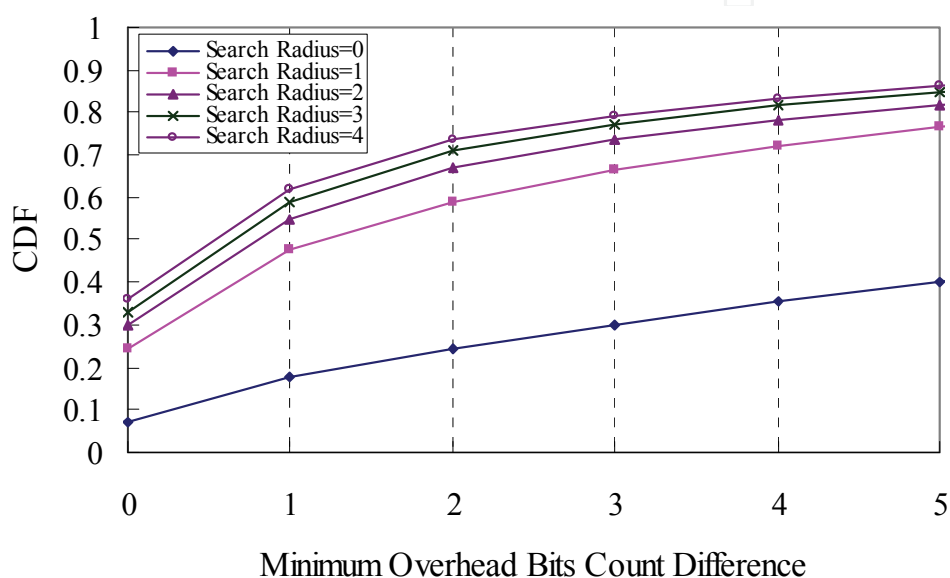
Fig. 8. The CDF of temporal correlation of overhead bits count

2.3 Improved MAD prediction

Because RDO in H.264 needs QP parameter before encoding, bit rate controller can't get MAD information of the current MB. To solve the QP dilemma caused by RDO in H.264, a simple linear model is proposed in the JVT-G012 to predict the MAD of not yet coded MB. The prediction model using the temporal information is then given by

$$MAD_{cur,temp}[i] = a_1 \times MAD_{pre}[i] + a_2 \tag{1}$$

where $MAD_{cur,temp}[i]$ denotes the temporal predicted MAD of the current ith MB, $MAD_{pre}[i]$ denotes the actual MAD of the co-located MB in the previous frame, a_1 and a_2 are two coefficients of prediction model. The initial value of a_1 and a_2 are set to 1 and 0, respectively. They would be updated after coding each MB.

The accuracy of prediction model using the previous temporal information is poorly when MAD changes abruptly due to high motion or scene changes. Thus, the prediction model is unable to predict the current changes, and is less sensitive to input data fluctuations. This situation will lead to error of prediction and error propagation by linear regression model. Therefore, it is desirable to collect more information that is helpful to predict MAD before RDO. In the procedure of H.264 encoding, the RDO module select the best mode by

calculating the rate-distortion cost (RDcost) for 9 modes including INTRA16×16, INTRA 4×4, INTER16×16, INTER 16×8, INTER 8×16, INTER 8×8, INTER 8×4, INTER 4×8 and INTER 4×4. Let MAD_{rough} be a rough measure for evaluating the difference between the current original frame and the previous reconstructed frame. To get the additional information, Liu *et al.* build the MAD_{rough} after rough motion determination only with the INTRA16×16 and INTER 16×16 modes. (Liu, et al., 2007) has shown that the MAD_{rough} is highly related to MAD_{actual}. A detailed description of the experimental results regarding the relationship between MAD_{rough} and MAD_{actual} is available in (Liu, et al., 2007). Therefore, the prediction model is helpful to improve the accuracy of prediction, especially for abrupt changes. The spatial linear prediction model is determined by

$$MAD_{cur,spat}[i] = b_1 \times MAD_{rough}[i] + b_2 \tag{2}$$

where $MAD_{cur,spat}[i]$ denotes the spatial MAD of the current MB, $MAD_{rough}[i]$ represents the rough MAD of the current MB, b_1 and b_2 are two coefficients of prediction model. They would be updated after coding each MB.

To combine the temporal prediction model with the spatial prediction model so that we can obtain more accurate MAD prediction. Therefore, we introduce two similarity measures as indicators to switch temporal and spatial prediction models

$$E_{temp}[i] = \sum_{n=i-S}^{i} \left| MAD_{cur,temp}[n] - MAD_{actual}[n] \right| \tag{3}$$

$$E_{spat}[i] = \sum_{n=i-S}^{i} \left| MAD_{cur,spat}[n] - MAD_{actual}[n] \right| \tag{4}$$

where S is the number of MAD samples used to measure E. When the measure E_{spat} is greater than E_{temp}, the MAD of current MB could be predicted by the temporal prediction model. On the other hand, it could be predicted by the spatial prediction model if E_{temp} is greater than E_{spat}. Therefore, in addition to the temporal prediction model, we can also use the spatial prediction model to predict current MB. Due to the fluctuation of MAD_{rough} roughly reflects the fluctuation of MAD_{actual} (Liu, et al., 2007), we adopt the characteristics to determine the optimal position (v_x, v_y) of MB in the searching radius $(R = 2)$ for the temporal and spatial prediction model. Therefore, the optimal position is obtained when the difference of MAD between MAD_{rough} and MAD_{actual} around the searching radius is the minimum value.

Figure 9 shows the proposed searching procedure to find the optimal position of predicted MB. We further use the MB in the position (v_x, v_y) to substitute for the co-located MB adopted in JVT-G012 to predict MAD of the current MB. Therefore, the equations (1) and (2) are rewritten as the following:

$$MAD_{cur,temp}[i] = a_1 \times MAD_{pre,temp,\min}(v_x, v_y) + a_2 \tag{5}$$

$$MAD_{cur,spat}[i] = b_1 \times MAD_{pre,spat,\min}(v_x, v_y) + b_2 \tag{6}$$

where $MAD_{pre,\min}(v_x, v_y)$ denotes the actual MAD of the position (v_x, v_y) in the temporal or spatial frame. In addition, the same technique as MAD prediction is also employed to the prediction of overhead bit rate. They are formulated as the following:

$$H_{i,temp} = MAD_{rough}[i] - MAD_{actual}[i-1]_{-2 \leq R \leq 2, \min} \qquad (7)$$

$$H_{i,spat} = MAD_{pred,spat}[i] - MAD_{actual}[i]_{-2 \leq R \leq 2, \min} \qquad (8)$$

where $H_{i,temp}$ denotes the temporal prediction and the spatial prediction of overhead bit rate of the ith MB in the previous P frame, and $H_{i,spat}$ denotes the spatial prediction of overhead bit rate of the MB in the current P frame.
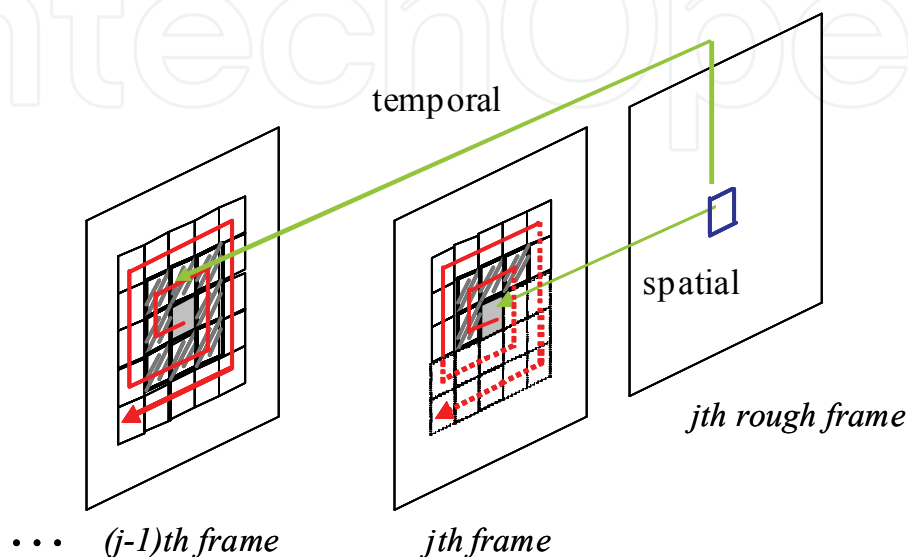
Fig. 9. The proposed searching procedure

3. Rate control for low delay video communication

The proposed rate control scheme is composed of three layers: group of picture (GOP) layer rate control, frame layer rate control and MB layer rate control. In the GOP layer, we have to select the best initial QP in each GOP and compute the total number of remaining bits for all non-coded frames. The frame layer rate control determines the target bits for each P frame. The MB layer rate control mainly determines the QP for each MB so that the sum of MB bits count is close to the target bits of frame. Figure 10 shows the flowchart of our proposed rate control scheme.

3.1 GOP layer rate control

In low delay applications, the typical format of a GOP used is IPPP…P. In this layer, we first compute the initial QP using our proposed method described in Section 2.1, and then compute the total number of remaining bits for all non-coded frames in the GOP as follows:

$$T_{r,GOP}[j] = \begin{cases} \dfrac{u}{F} \times N_{GOP} & j = 1 \\ T_{r,GOP}[j-1] - A[j-1] & 2 \leq j \leq N_{GOP} \end{cases} \qquad (9)$$

where $T_{r,GOP}[j]$ denotes the remaining bits of the jth frame that not yet coded in the GOP, u denotes the channel bandwidth, F denotes the frame rate, N_{GOP} denotes the total number of frames in the GOP, and the $A[j-1]$ denotes the actually generated bits in the (*j-1*)th frame.
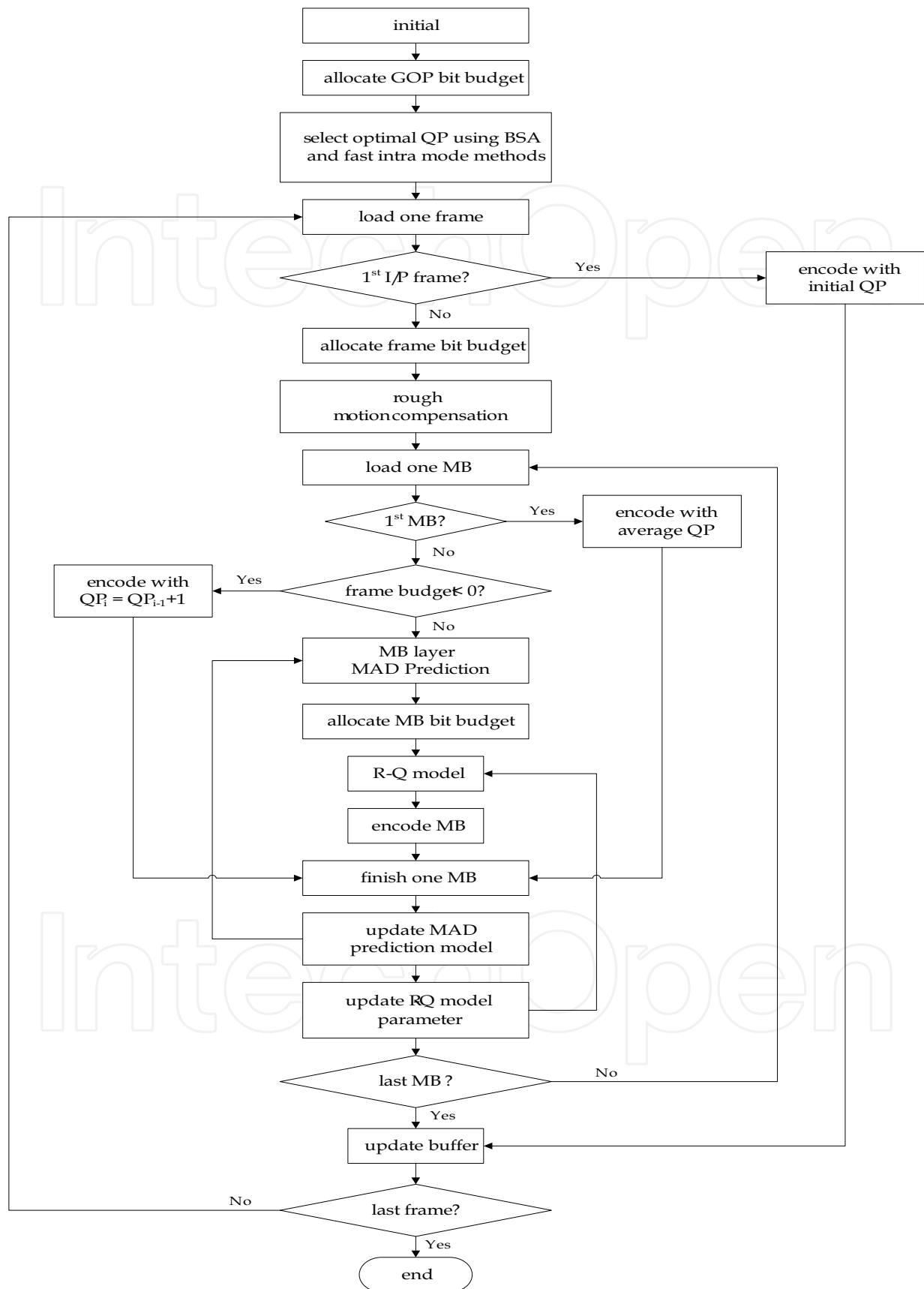
Fig. 10. The flowchart of proposed rate control scheme

3.2 Frame layer rate control

The layer mainly allocates the target bits for each frame. It can be calculated from

$$T_{frame}[j] = \beta \times f_1[j] + [1 - \beta] \times f_2[j] \tag{10}$$

where $T_{frame}[j]$ is the target sum bits to encode the jth frame, β is a constant and is set 0.75 in JVT-G012. In addition, the two parameters of f_1 and f_2 are expressed as follows:

$$f_1[j] = \frac{T_{r,GOP}[j]}{N_{r,p}} \tag{11}$$

$$f_2[j] = \left(\frac{u}{F}\right) + \gamma \times (Tbl[j] - B_c[j]) \tag{12}$$

where f_1 represents the number of remaining bits for remaining frames. f_2 is estimated from the previous actual buffer occupancy B_c, target buffer level Tbl, frame rate and the channel bandwidth. The details of these parameters are referred in JVT-G012.

Finally, the target bits of frame $T_{frame}[i]$ is used to avoid overflow and underflow with the hypothetical reference decoder (HRD) which has been defined in H.264. So, the target bit of frame is bounded as follows:

$$\begin{cases} T_{frame}[i] = \max\{ L_{bound} , T_{frame}[i]\} \\ T_{frame}[i] = \min\{ U_{bound} , T_{frame}[i]\} \end{cases} \tag{13}$$

where L_{bound} and U_{bound} denote the lower and upper boundary, respectively.

3.3 MB layer rate control

The MB layer rate control is the detail section to achieve more accurate output rate. If the basic unit is set as one frame, the MB layer rate control is skipped and using the frame layer predicted QP to encode all MBs within the frame. In other words, if the basic unit is set as one MB, the MB layer is to compute QP for each MB. For low delay applications with a small buffer size, the MB layer rate control is demanded generally to avoid buffer overflow and underflow accurately. The MB layer rate control can be described as follows.

Step 1. Check whether ith MB is the first MB in the current frame. If it is true, calculate the average value of QP for all MBs in the previous frame, then go to step 8. Otherwise go to step 2.

Step 2. Update the budget of the current frame as follows:

$$T_{budget}[i] = \begin{cases} T_{frame}[i] & i = 0 \\ T_{frame}[i-1] - R_{MB}[i-1] & otherwise \end{cases} \tag{14}$$

where $T_{budget}[i]$ denotes the number of remaining bits in the current frame after coding $(i-1)th$ MB, and $R_{MB}[i]$ denotes the actual bits after coding.

Step 3. Check whether T_{budget} is not enough. If T_{budget} is smaller than zero, go to step 7. Otherwise go to step 4.

Step 4. Predict the MAD of current MB using the two improved prediction model described in Section 2.3.

Step 5. Predict the overhead bits of current MB using our method described in Section 2.2 and Section 2.3, and compute the texture bits for the current MB as follows:

$$R_{text} = \frac{f_{rb}}{N_{rb}} - R_{header} \tag{15}$$

where R_{text} denotes the texture bits, R_{header} denotes overhead bits that is predicted by ours method. f_{rb} and N_{rb} denote the number of remaining bits for all non-coded MB in the current frame and the number of non-coded MB, respectively.

Step 6. Determine the quantization step of current MB using the quadratic Rate-Quantization
(R-Q) model. Then go to step 8.

$$R_{text} = X_1 \times \frac{MAD_{cur}}{Q_{step}} + X_2 \times \frac{MAD_{cur}}{Q_{step}^2} \tag{16}$$

where X_1 and X_2 are coefficients of R-Q model, MAD_{cur} denotes $MAD_{pre,temp}$ or $MAD_{pre,spat}$ described in Section 2.3.

Step 7. Due to the budget of frame is overspent early, the QP value is increased by 1 to reduce the output rate, that is

$$QP_i = QP_{i-1} + 1 \tag{17}$$

where QP_i denotes the QP of the ith MB.

Step 8. Actual encoding of the current MB. The QP is applied to perform RDO for the current MB.

Step 9. Update the R-D model and two MAD prediction models. After encoding each MB, the encoder should update the parameters of R-Q model and MAD prediction model using the linear regression method.

Step 10. Check whether ith MB is the last MB in current frame. If it is true, go to step 12. Otherwise go to step 11.

Step 11. Set $i = i+1$, then repeat step 1 to step 10.

Step 12. Finish the MB layer rate control.

4. Experimental results

4.1 Simulation setup

In this section, we discuss the experimental model used to simulate the proposed low delay H.264 video communication scheme, the performance metrics to evaluate performance of different methods, and the parameters and methods to encode the video for comparison. We evaluate eight video sequences including Foreman, Bus, Highway, Coastguard, News, Stefan, Mobile and Akiyo. Each sequence consists of 100 frames at QCIF format. For the real-time applications, all the sequences are intra-coded for the first frame (I-frame) and the remaining frames (P-frames) are inter-coded. In addition, the initial QP is set as 38 for JVT-G012 and the rate control presented by (Jiang & Lin, 2006), the channel bandwidth is set as 24 kbps, the symbol mode is CAVLC for low-delay, and RDO is enabled. To evaluate and

compare the performance of the different methods, we have implemented our proposed rate control scheme with JVT reference software JM 12.1 (JVT) serving as a test benchmark. For a fair comparison, the JVT-G012 and (Jiang & Lin, 2006) are also implemented based on JM 12.1 (JVT), respectively.

4.2 Performance metrics
To analyze the performance of the decoded video sequences, we use the average peak signal-to-noise ratio for luma (PSNRY) of all frames over all realizations to evaluate the objective video quality, because it is the most widely used objective video quality metric. PSNRY is defined by

$$PSNRY = 10\log\frac{255^2}{MSE} \quad \text{dB} \tag{18}$$

where *MSE* is the mean square error between the original pixel and the decoded pixel.

The video quality is evaluated in terms of PSNRY, buffer fullness and skipped frames. In this work, the size of buffer is set as $u/F{\times}1.25$ for low delay applications. The buffer overflow threshold is set as 80% of the buffer size, which it is u/F. If the current buffer fullness exceeds 80% of the encoder buffer size, the encoder will skip encoding the next frame until the buffer fullness is lower than 80% of the encoder buffer size. When frame skipping occurs, the decoder displays the previous encode frame in place of the skipped one. Therefore, the previous frame is used in the PSNRY calculation.

4.3 Performance evaluation
To evaluate the performance of the proposed rate control scheme for low delay video communication, we compare the JVT-G012 and the coding scheme in (Jiang & Lin, 2006). All other parameters are selected the same among these schemes.

The comparisons of PSNRYs and buffer fullness by adopting the proposed scheme, JVT-G012 and (Jiang & Lin, 2006) are shown in Fig. 11 and Fig. 12, respectively. Figs. 11(a)-(h) shows frame by frame PSNRYs for various sequences, and Figs. 12(a)-(h) shows the number of bits in the buffer at each frame. From Fig. 11, we can find that our proposed scheme can improve the PSNRY significantly for most frames in these sequences. In addition, our proposed scheme can achieve much fewer skipped frames for sequences with high motion than the other two schemes. The improvements in PSNRY and skipped frames are very high for scene change video such as "Mobile". Despite the first frame has higher PSNRY quality for JVT-G012, the buffer fullness above the buffer threshold at the same time as shown in Fig. 12. Therefore, the number of frame skipped is increased for initial frames. The proposed scheme improves the number of frame skipped and reduces the quality deviations of the initial frames by choosing the best initial QP. Thus, it can supply user with a stable and constant viewing experience. In addition, the proposed scheme also improves the accuracy of prediction method to increase the reconstructed quality. From the experimental results, the best initial QP can improve the buffer fullness of initial frames efficiently. It is found from Figs. 12(a)-(h) that our proposed scheme achieves much steadier buffer fullness when compared to that of JVT-G012 and (Jiang & Lin, 2006). This implies that our proposed scheme produces stable buffer delay so that it is suit for real-time video communication. In Figs. 12(a)-(h), if the curve of buffer fullness falls below zero, it yields a buffer underflow problem. In such case, stuffing bits should be inserted into bit stream. Although underflow does not affect motion continuity, it wastes channel bandwidth.

On the other hand, since the bit allocation of each basic unit is not further considered in our rate control scheme, we still use the method of JVT-G012 to allocate the target bits for frame layer and MB layer. Therefore, the improvement of buffer control is limited when compared with JVT-G012. From the analysis of the experimental results, the proposed rate control scheme indeed can achieve better improvements than those of the JVT-G012 and the (Jiang & Lin, 2006) in low delay video communication.
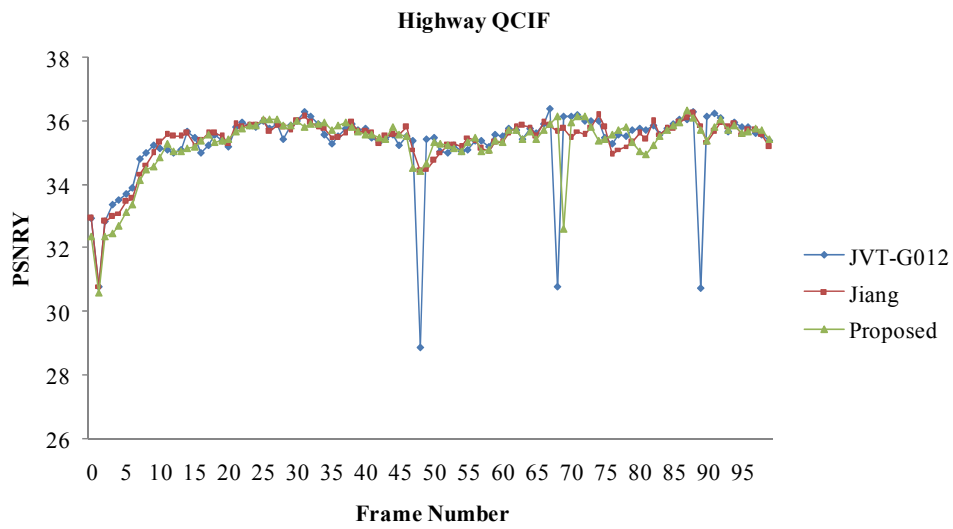
The experimental results are further reported in Table 2. and Table 3. From the two tables we can find that the proposed scheme gives an average PSNRY gain of about 0.55 dB and 0.58 dB when compared with JVT-G012 and (Jiang & Lin, 2006), respectively. In addition, the proposed scheme improves the number of frame skipped and reduces the quality deviations of the initial frames. To compare with all test sequences, the proposed rate control scheme achieves more accurate rate control, especially for high motion sequences. For all sequences, the proposed method can reduce the number of skipped frames with the best reconstructed video quality.

(a)

(b)

Highway QCIF

(c)

Coastguard QCIF

(d)

News QCIF

(e)

Stefan QCIF

(f)

Mobile QCIF

(g)

Akiyo QCIF
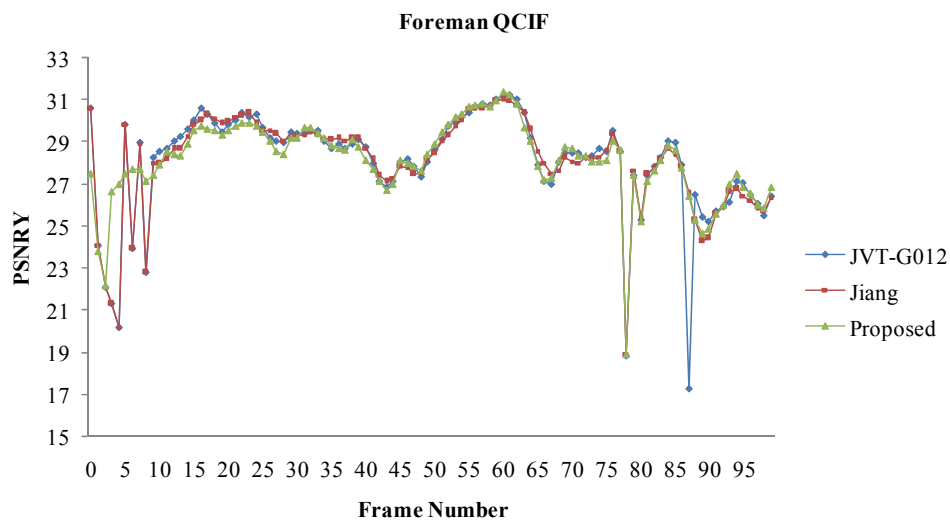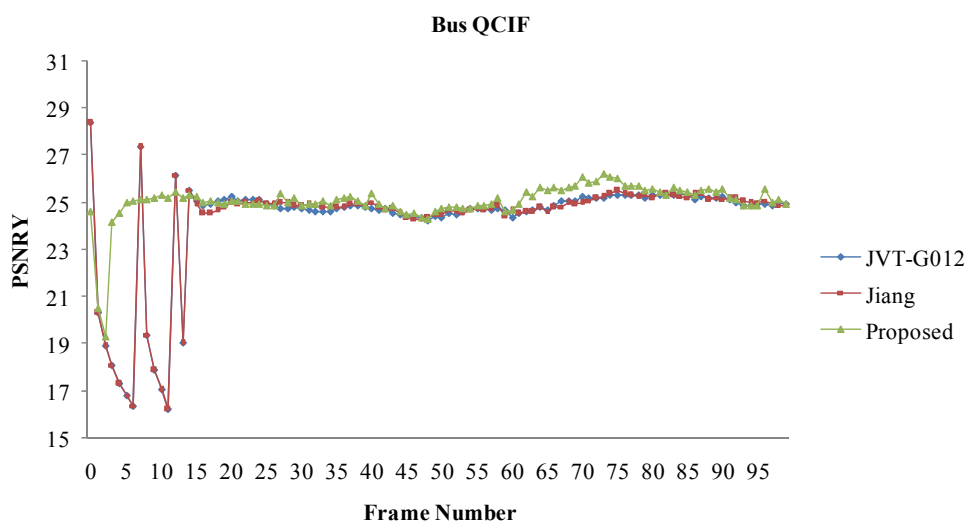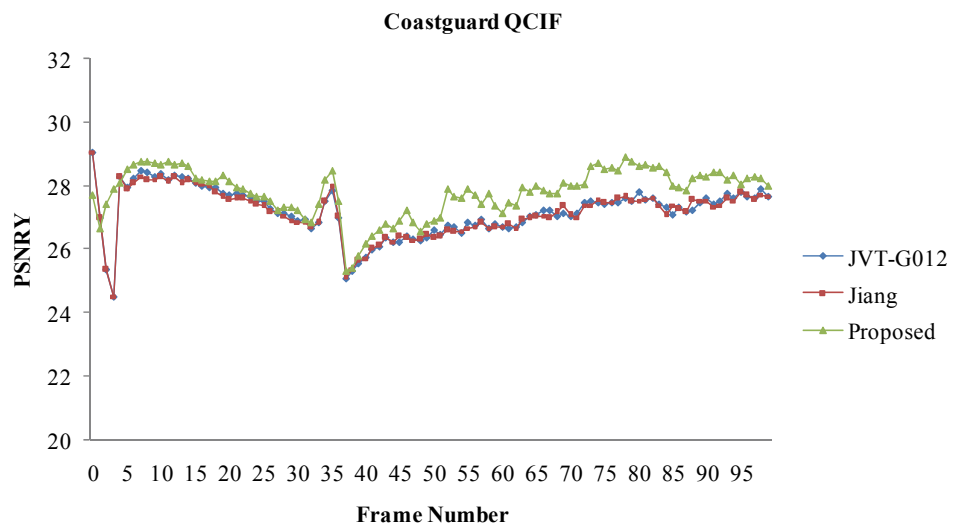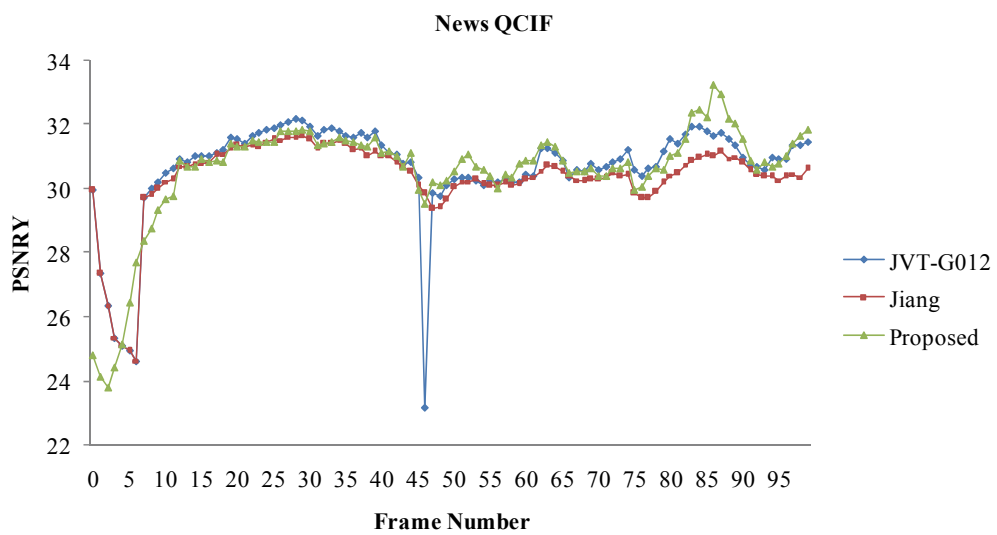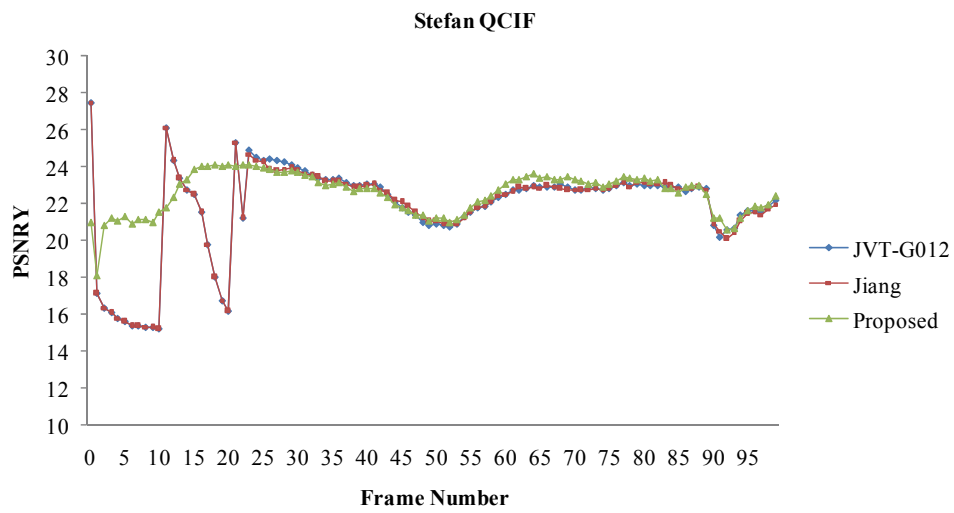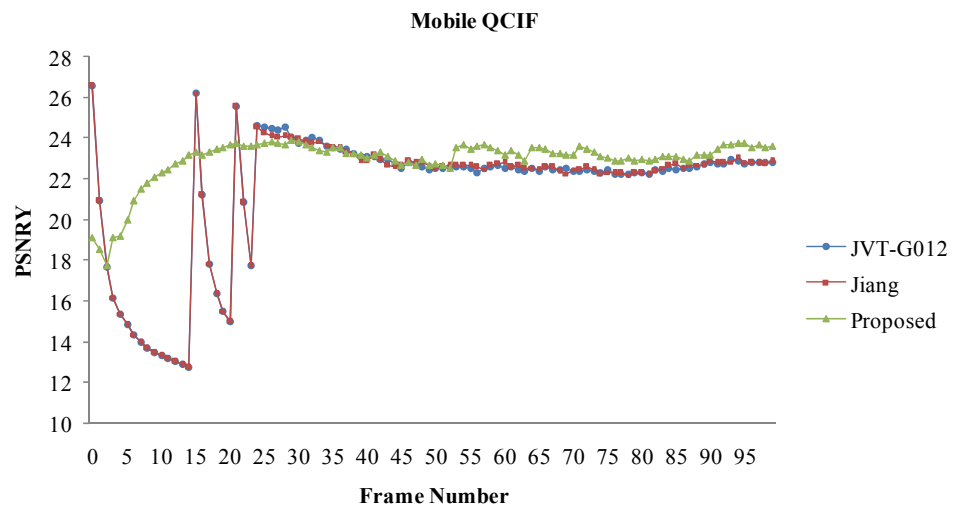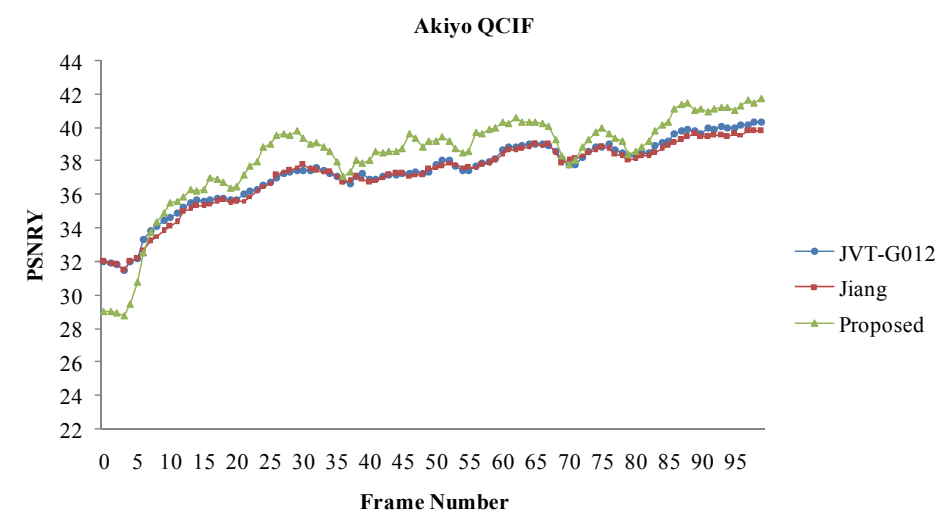
(h)

Fig. 11. Comparisons of PSNRYs for the proposed scheme, JVT-G012 and (Jiang & Lin, 2006)

Foreman QCIF

(a)

Coastguard QCIF

(b)

Stefan QCIF

(c)

News QCIF

(d)

Bus QCIF

(e)

Highway QCIF

(f)

(g)

(h)
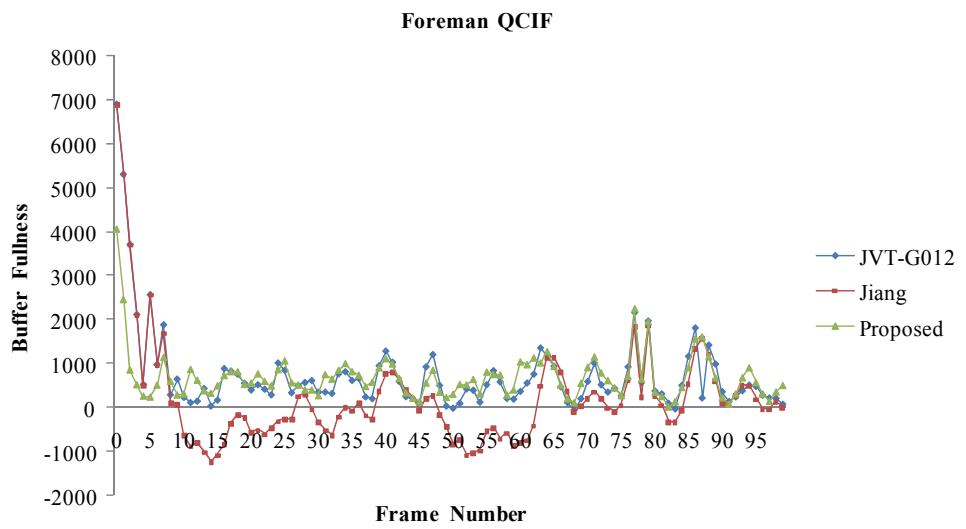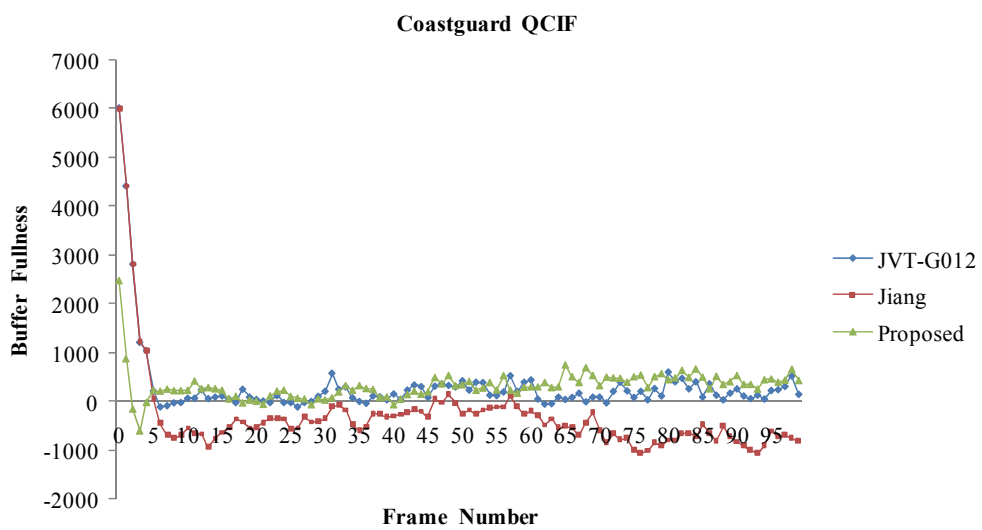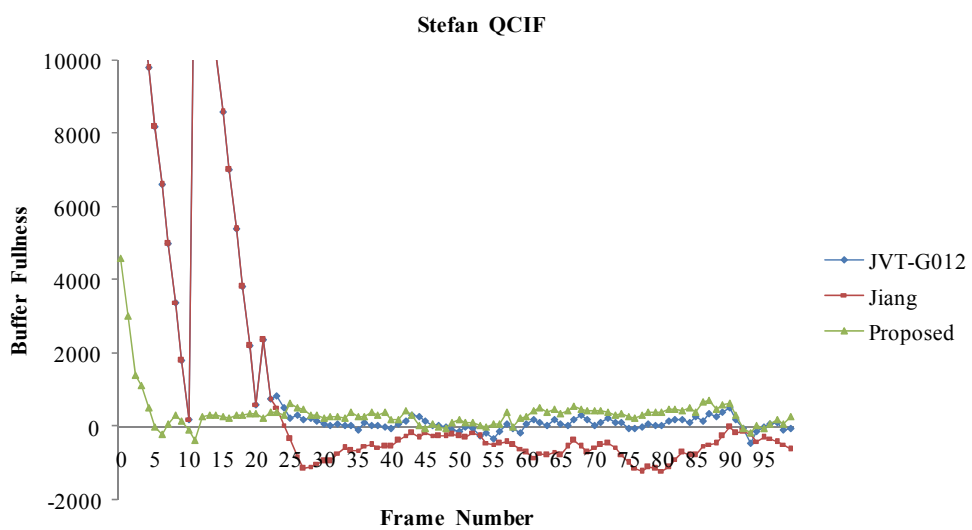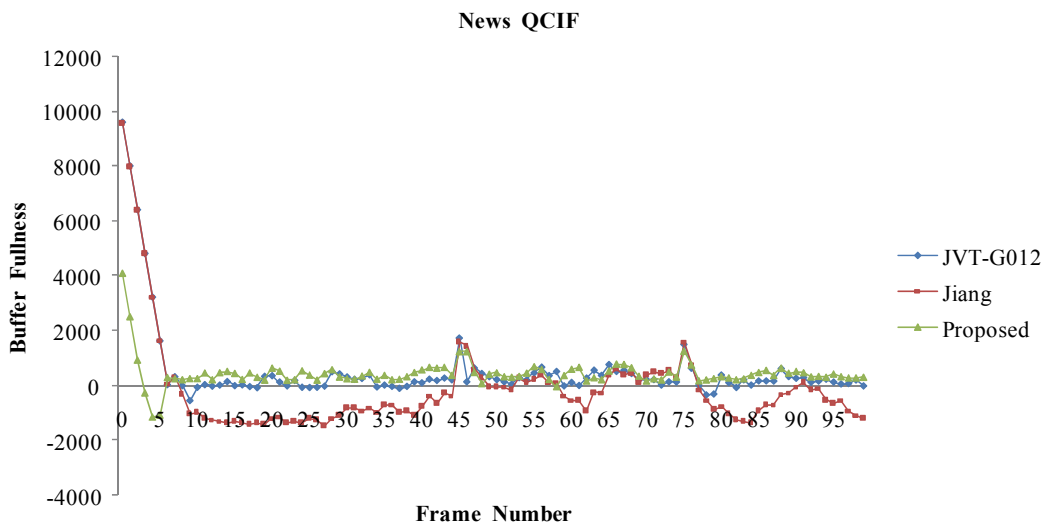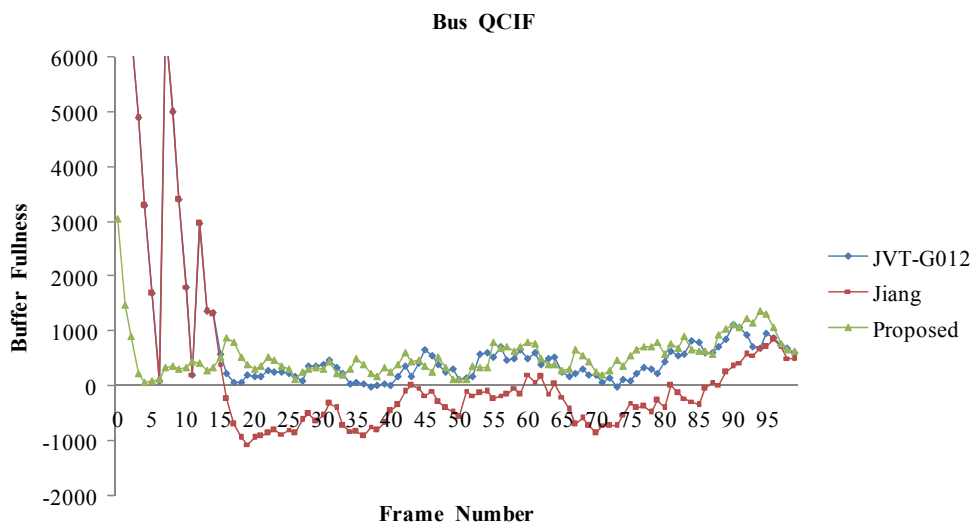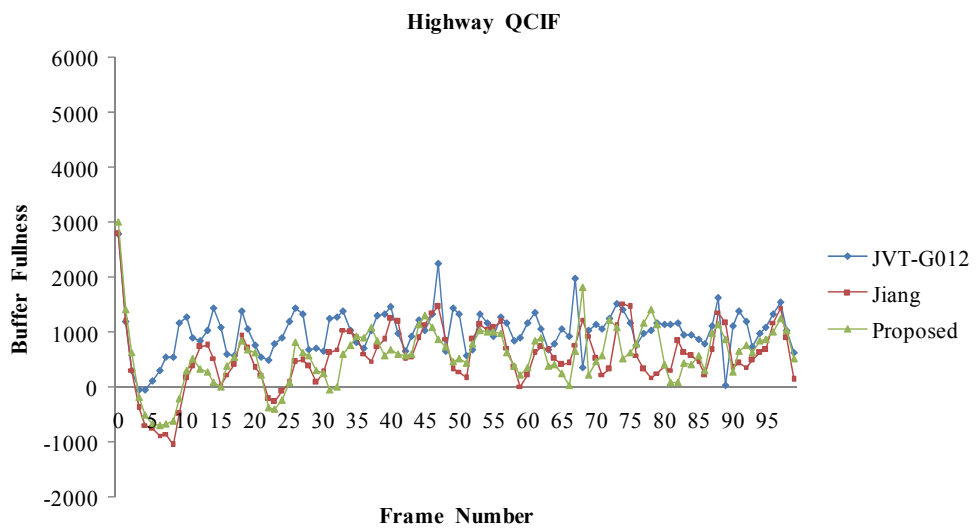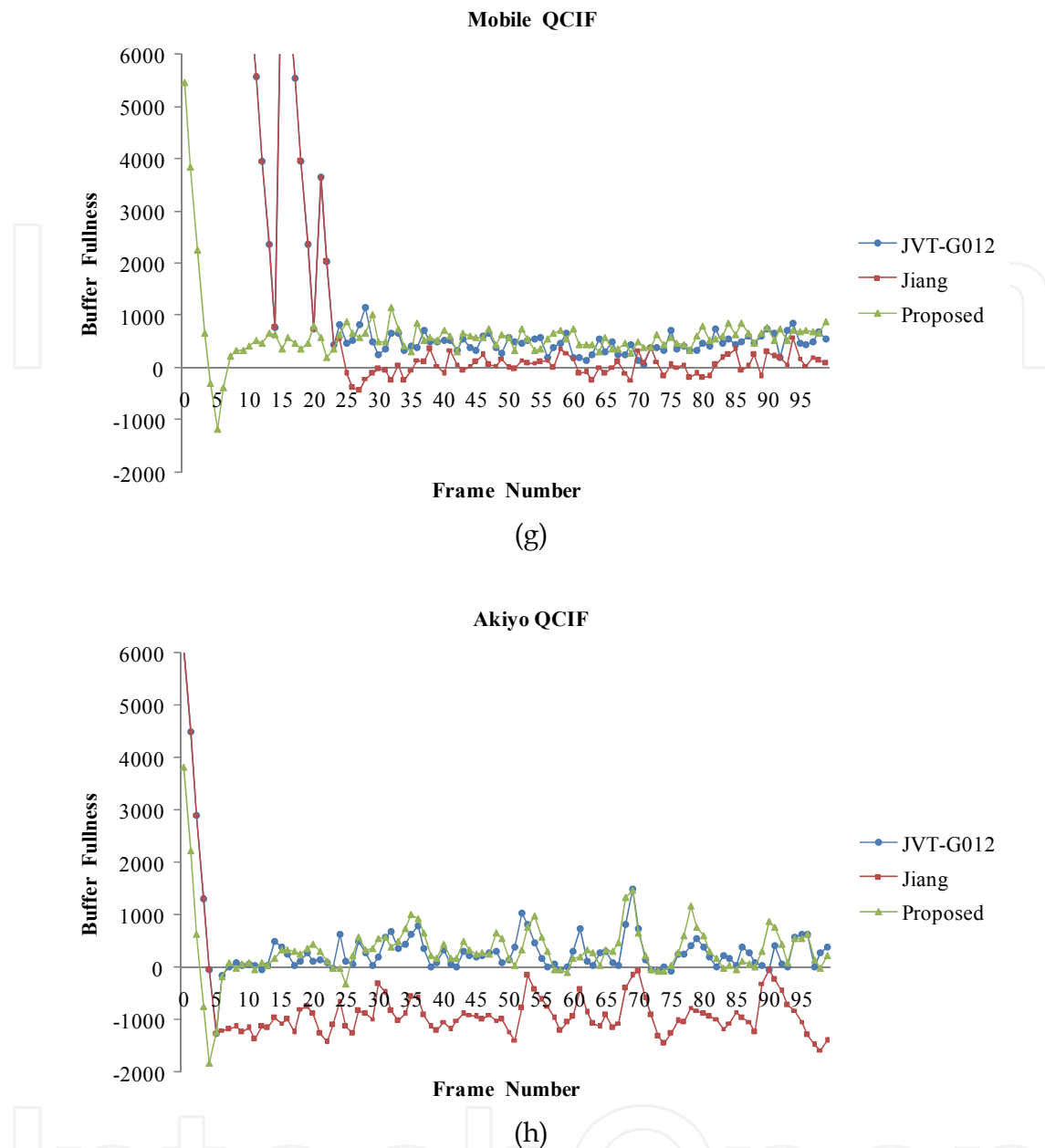
Fig. 12. Comparisons of buffer fullness for the proposed scheme, JVT-G012 and (Jiang & Lin, 2006)

5. Conclusion

In this chapter, we proposed a more efficient rate control scheme for low delay H.264 video coding. A fast and best selection of initial QP is first proposed in the GOP layer rate control. Then, an improved MAD prediction model and overhead bits prediction method is adopted in the MB layer rate control. For low-bandwidth transmission channel applications, the simulation results show that the proposed rate control scheme is more efficient than JVT-G012 and (Jiang & Lin, 2006) for low-delay applications. In addition, the proposed scheme improves the number of frame skipped and reduces the quality deviations of the initial frames by choosing the best initial QP.

Sequences (QCIF)	No. of skipped frames			Average PSNRY (dB)			Bitrate (kbps)	
	G012	Proposed	Gain	G012	Proposed	Gain	G012	Proposed
Foreman	9	4	5	28.07	28.16	+ 0.09	24.02	24.03
Bus	11	2	9	24.23	25.04	+ 0.81	24.11	24.11
Highway	4	2	2	35.28	35.29	+ 0.01	24.05	24.08
Coastguard	3	1	2	27.19	27.80	+ 0.61	24.02	24.03
News	7	2	5	30.51	30.57	+ 0.06	24.00	23.98
Stefan	20	1	19	21.80	22.56	+ 0.76	24.00	24.00
Mobile	21	2	19	21.45	22.95	+ 1.50	24.06	24.05
Akiyo	3	2	1	37.35	38.30	+ 0.95	24.00	24.00

Table 2. Comparisons of the number of skipped frames, average PSNRY and bitrate between the proposed scheme and JVT-G012

Sequences (QCIF)	No. of skipped frames			Average PSNRY (dB)			Bit rate (kbps)	
	Jiang	Proposed	Gain	Jiang	Proposed	Gain	G012	Proposed
Foreman	8	4	4	28.09	28.16	+ 0.07	24.01	24.03
Bus	11	2	9	24.21	25.04	+ 0.83	24.08	24.11
Highway	1	2	1	35.36	35.29	- 0.07	24.02	24.08
Coastguard	3	1	2	27.17	27.80	+ 0.63	24.00	24.03
News	6	2	1	30.32	30.57	+ 0.25	23.95	23.98
Stefan	20	1	19	21.77	22.56	+ 0.79	23.98	24.00
Mobile	21	2	19	21.46	22.95	+ 1.49	24.04	24.05
Akiyo	3	2	1	37.19	38.30	+ 1.11	24.00	24.00

Table 3. Comparisons of the number of skipped frames, average PSNRY and bitrate between the proposed scheme and (Jiang & Lin, 2006)

6. References

ITU-T (2003), Rec. H.264 and ISO/IEC 14496-10 AVC, JVT-G050, March 2003.

ITU-T (1997), Video Codec Test Model, ITU-T/SG15, TMN8, Portland, OR, June1997.

ISO/IEC (1993), Test Model Editing Committee, MPEG-2, Test Model 5, Doc. ISO/IEC, JTC1/SC29 WG11/93-225, April 1993.

ISO/IEC (1997), Coding of Moving Pictures and Associated Audio MPEG 97/W1796, Text of ISO/IEC 14496-2 MPEG-4 Video VM-Version 8.0, ISO/IEC JTC1/SC29/WG11, Video Group, Stockholm, Sweden, July 1997.

CCITT (1990), Rec. H.261 — Video codec for audiovisual services at p×64 Kbits/s, CCITT SG XV, COM XV-R37-E, 1990

ITU-T (1996), Rec. H.263 — *Video coding for low bit rate communication*, ITU-T Rec. H.263, March 1996

LeGall, D. (1991), MPEG: A video compression standard for multimedia applications, *Communications of ACM*, Vol. 34, No. 4, pp. 46-58, Apr. 1991

ISO/IEC (1994), Information technology - Generic coding of moving pictures and associated audio information - Part 2: Video, ISO/IEC FDIS 13818-2, MPEG-2 1994

ISO/IEC (1999), Information technology - Coding of Audio-Visual Objects - Part 2: Visual, ISO/IEC 14496-2, MPEG-4 1999

Jiang, M. & Ling, N. (2006). Low-delay rate control for real-time H.264/AVC video coding, *IEEE Transactions on Multimedia*, Vol. 8, No. 3, (June 2006), pp.467-477, March 2006.

Li, Z. G.; Pan, F.; Lim, K. P.; Feng, G. N.; Lin, X. & Rahardaj, S. (2003). Adaptive basic unit layer rate control for JVT, *Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6 Doc. JVT-G012*, Pattaya, Thailand, March 2003.

JVT H.264/AVC Reference Software version JM 12.1, http://iphome.hhi.de/suehring /tml/download/jm12.zip.

Armstrong, A.; Beesley, S. & Grecos, C. (2006). Selection of initial quantization parameter for rate controlled H.264 video coding, *Research in Microelectronics and Electronics*, pp. 249-252, June 2006.

Wang, C. C.; Chen, T. S. & Tung, W. C. (2006). Fast intra-mode decision in H.264 using interblock correlation, *IEEE International Conference on Image Processing* (ICIP 2006), pp.1345-1349, Atlanta, USA, October 2006.

Wang, H. & Kwong, S. (2008). Rate-distortion optimization of rate control for H.264 with adaptive initial quantization parameter determination, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 18, No. 1, pp.140-144, January 2008.

Yuan, W.; Lin, S.; Zhang, Y. & Luo, H. (2006). Optimum bit allocation and rate control for H.264/AVC, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 16, No. 6, pp.705-715, June 2006.

Gu, Y. & Song, B. C. (2009). An intra-frame rate control algorithm for ultralow delay H.264/AVC, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 19, No. 5, pp.747-752, May 2009.

Liu, Y.; Li, Z. G. & Soh, Y. C. (2008). Rate control of H.264/AVC scalable extension, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 18, No. 1, pp.116-121, January 2008.

Liu, Y.; Li, Z. G. & Soh, Y. C. (2007). A novel rate control scheme for low delay video communication of H.264/AVC standard, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 1, pp.68-78, January 2007.

Lee, H. J.; Chiang, T. & Zhang, Y. (2000). Scalable rate control for MPEG-4 video, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10, No. 6, pp.878-894, September 2000.

Recent Advances on Video Coding

Edited by Dr. Javier Del Ser Lorente

This book is intended to attract the attention of practitioners and researchers from industry and academia interested in challenging paradigms of multimedia video coding, with an emphasis on recent technical developments, cross-disciplinary tools and implementations. Given its instructional purpose, the book also overviews recently published video coding standards such as H.264/AVC and SVC from a simulational standpoint. Novel rate control schemes and cross-disciplinary tools for the optimization of diverse aspects related to video coding are also addressed in detail, along with implementation architectures specially tailored for video processing and encoding. The book concludes by exposing new advances in semantic video coding. In summary: this book serves as a technically sounding start point for early-stage researchers and developers willing to join leading-edge research on video coding, processing and multimedia transmission.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Chou-Chen Wang and Chi-Wei Tung (2011). Rate Control for Low Delay Video Communication of H.264 Standard, Recent Advances on Video Coding, Dr. Javier Del Ser Lorente (Ed.), ISBN: 978-953-307-181-7, InTech, Available from: http://www.intechopen.com/books/recent-advances-on-video-coding/rate-control-for-low-delay-video-communication-of-h-264-standard

INTECH
open science | open minds