

Rate control of hierarchical B prediction structure for multi-view video coding

Jianjun Lei · Kun Feng · Meimin Wu · Shuai Li ·
Chunping Hou

Published online: 2 March 2013

© Springer Science+Business Media New York 2013

Abstract Rate control plays an important role in regulating bit streams in video coding. In order to obtain good coding performance, the hierarchical B prediction structure has been adopted in Multi-view Video Coding (MVC). However, the conventional rate control scheme is not efficient in the hierarchical B prediction structure. In this paper, we propose a rate control algorithm to address this problem. First, the accurate estimation of Mean Absolute Distortion (MAD) of the current frame is desired for both quantization parameter (QP) selection and Rate Distortion Optimization (RDO). Considering the hierarchical B structure, a bi-directional MAD prediction model is proposed to predict the MAD of the current frame by using the actual MADs of the encoded frames in the lower Temporal Layers (TLs). Second, the number of header bits has a close relationship with the TLs in the hierarchical B prediction structure. Therefore, we propose an enhanced prediction method in which a proportional relationship of the header bits is introduced if the frames are located in different TLs. Experimental results show that our proposed algorithm can achieve both accurate target bit rate and good coding performance.

Keywords MVC · Rate control · Mean Absolute Distortion (MAD) · Header bits · Hierarchical B prediction

1 Introduction

In order to present the real and natural video scenes, Multi-view Video (MVV) with three-Dimensional (3D) visual effect has been receiving more and more attention [24]. MVV is essential for the success of many types of 3D video applications, such as 3D Television (3DTV), Free Viewpoint Video (FVV) as well as Free Viewpoint Television (FTV).

J. Lei (✉) · K. Feng · M. Wu · S. Li · C. Hou
School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China
e-mail: jjleitg@gmail.com

J. Lei
Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA

Audiences can not only experience the 3D depth perception of a scene, but also select the viewpoint interactively within certain ranges [9]. In order to store and transmit the huge MVV data for practical application, an efficient compression technique is indispensable [30]. MVV contains a large amount of inter-view statistical dependencies, since all the cameras capture the same scene from different viewpoints. These dependencies can be exploited by the combined temporal/inter-view prediction, where images are not only predicted from the temporally neighboring images but also from corresponding images in the adjacent views, known as Multi-view Video Coding (MVC). The main structure of MVC [17] is depicted in Fig. 1. The encoder receives MVV streams and generates one bit-stream [27]. The decoder decodes and outputs MVV signals for various applications.

Based on both temporal and inter-view redundancy of MVV, researchers have proposed many different prediction structures [6]. Figure 2 depicts the temporal and inter-view hybrid hierarchical B prediction structure proposed by the Heinrith-Hertz-Institute [17], which is adopted by JVT MVC [18]. In July 2006, JVT released the Joint Multi-view Video Model (JMVM) [10]. Consequently, the standard of MVC was published as Annex H of the H.264/AVC in March 2009 [7]. MVC has brought great convenience for the transmission and storage of MVV. However, there are still many problems. Rate control is one of them considering the transmission and application [22]. The main purpose of rate control is to regulate the bit stream according to the available bandwidth to ensure the highest video quality.

There are many existing rate control schemes for video coding, such as TM5 for MPEG-2, VM8 for MPEG-4, TMN8 for H.263 and JVT-G012 for H.264/AVC [2, 8, 12]. MVC is more complex than the existing coding methods due to its temporal and inter-view hierarchical structure and thus leads to new challenges in bit allocation and rate control. Furthermore, the MVC reference software [1] does not provide an effective rate control mechanism, and the number of target bits for each Temporal Layer (TL) is achieved by coding it with a fixed quantization parameter (QP), which

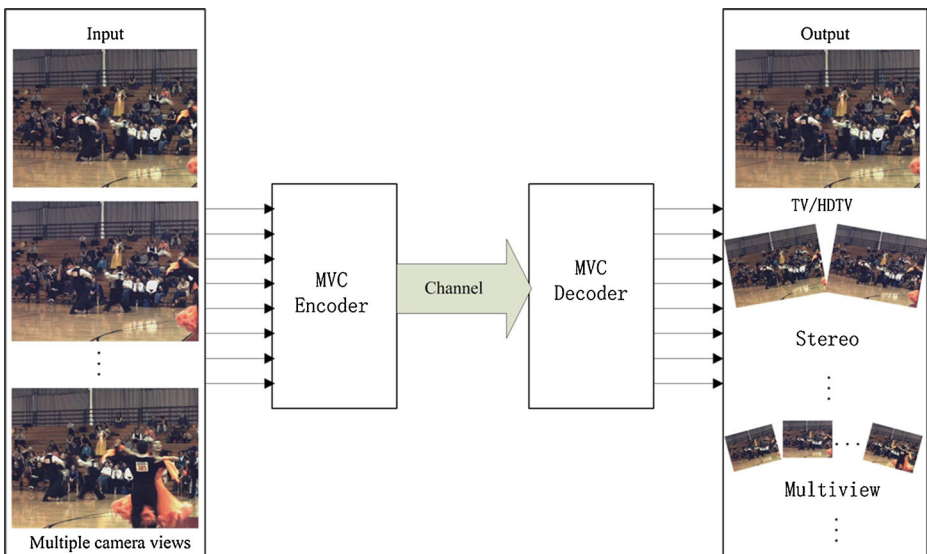


Fig. 1 The structure of the MVC system

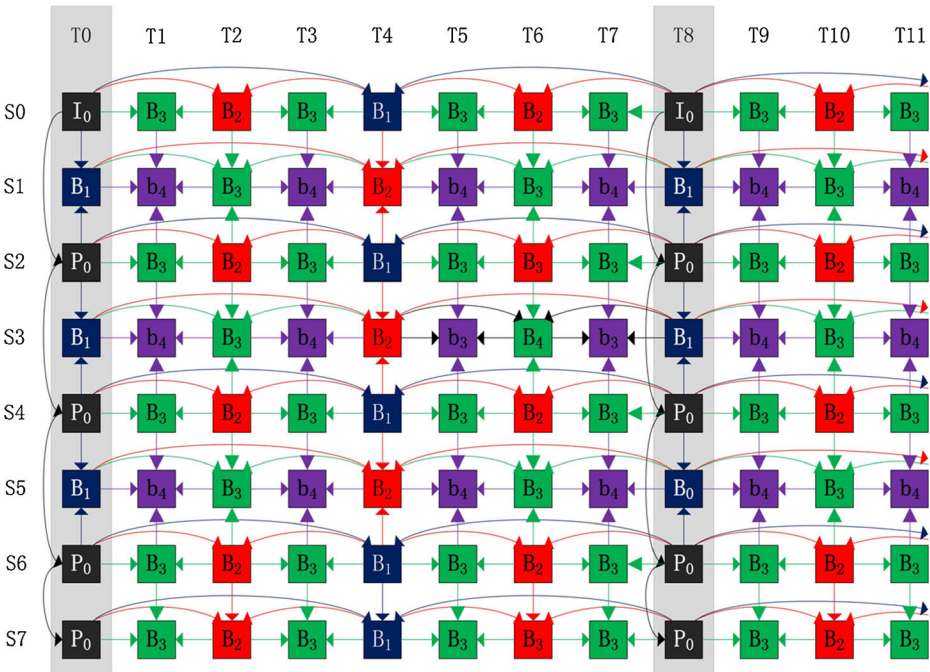


Fig. 2 The prediction structure of MVC

is determined by a logarithmic search. Therefore, many researchers have explored the way of bit allocation between views as well as TLs and rate control scheme for MVC. Natio and Matsumoto [19] proposed an algorithm which is tested by TM5, where the two channels of bit-stream share a same buffer. While the bit allocation method of TM5 is relatively simple, and the result turns worse with the increase of the coding modes [28]. Based on the Rate Distortion (RD) theory, Woo et al. proposed a basic idea of optimal bit allocation and related algorithms [25]. Lim revised the quadratic RD model for 3D multi-view sequences on the basis of the frame types [13]. Although these rate control algorithms can yield good performance, the computing cost is too high, to be practical in many applications [15]. Park et al. [20] proposed a rate control algorithm for MVC, but it did not give an efficient rate control scheme in the frame layer. Cho et al. proposed a GOP-based dependent distortion model [3] and a Q-selection tree solution [4] for bit allocation among different layers. Hu et al. developed an optimized rate control scheme [5] and Seo et al. introduced an efficient bit allocation in temporal layers [21]. Xu et al. proposed a bit allocation method for the hierarchical B prediction structure [26], which introduces some weighting factors for each TL with some heuristically predefined values. These weighting factors are then used to determine the target bits budget of the frame layer. However, conventional rate control schemes do not consider efficiently the characteristic of hierarchical B prediction structure.

In this paper, we propose a rate control algorithm employing the features of hierarchical B prediction structure for MVC. First, a bi-directional Mean Absolute Distortion (MAD) prediction method is introduced to predict the MAD by using the encoded frames in lower TLs. Furthermore, an enhanced header bits prediction method is developed to improve the

accuracy of the predicted header bits, in which the header bits of first frame in each TL are predicted by the ratio of the encoded frame header bits in the lower TLs. With the accurate MAD and header bits prediction model, the proposed rate control algorithm can achieve both accurate target bit rate and good coding performance compared to conventional rate control schemes.

The rest of this paper is organized as follows. The background and related work are presented in section 2. The proposed rate control algorithm with bi-directional MAD prediction model and enhanced header bits prediction method is described in section 3. Section 4 presents our experimental results. Section 5 concludes the paper.

2 Background and related work

2.1 Conventional rate control schemes

Rate control is employed as an important component of video coding to provide the best reconstructed video quality at the destination under certain network conditions and channel bandwidth. The process of conventional rate control is as follows. First, the number of target bits for the current frame is allocated. Then, the MAD of the current frame is predicted. The R-D model together with the target bits are used to calculate the QP, which is finally adopted to encode the current frame.

2.1.1 The MAD prediction

H.264/AVC utilizes RDO to select the optimal coding mode [14]. With the RDO introduced in H.264/AVC, QP affects both the Motion Estimation/Motion Compensation (ME/MC) and the spatial quantization steps to improve the coding performance. On the other hand, this feature makes the H.264/AVC rate control more complicated. For example, QP is determined by MAD before the RDO process. However, the MAD of the current frame is available only after RDO-based coding is completed. JVT-G012 [12] gives a linear prediction model to solve the QP-MAD dilemma caused by the RDO in H.264/AVC. As shown in Fig. 3, the MAD of the current frame (F_c)



Fig. 3 The prediction of MAD model in JVT-G012

can be predicted from the previous encoded frame (F_p). The linear prediction model can be depicted as follows.

$$MAD_c = c_1 \cdot MAD_p + c_2 \quad (1)$$

where MAD_c is the predicted MAD of the current frame, MAD_p is the actual MAD of the previous frame [23], c_1 and c_2 are two coefficients of the prediction model. The initial value of c_1 and c_2 are set to 1 and 0, respectively. And they are updated after coding each frame using the linear regression model [12].

2.1.2 QP Calculation

The relation between target bits and QP fits the quadratic R-D model and can be depicted as follows [8].

$$\frac{R_c - H_c}{MAD_c} = \frac{x_1}{Qstep_c} + \frac{x_2}{Qstep_c^2} \quad (2)$$

where R_c is the total number of the bits used for encoding the current frame, H_c is the predicted header bits of the current frame and it mainly includes header information and overheads like motion vectors. MAD_c is the predicted MAD for the current frame. $Qstep_c$ is the quantization step for the current frame. The relation between $Qstep_c$ and QP is shown as follows.

$$Qstep_c = 2^{(QP-4)/6} \quad (3)$$

x_1, x_2 are the first and second-order coefficients, the default value of x_1 is the bit rate, and x_2 is 0. x_1, x_2 can be obtained based on the least square method and linear regression.

2.2 Hierarchical B prediction structure

It is generally known that the hierarchical B prediction structure is developed for H.264/AVC at the very beginning. Due to its efficiency and good R-D performances, the hierarchical B prediction structure is also adopted in SVC and MVC [16, 32]. Figure 4 shows a typical hierarchical reference frame structure for temporal prediction. The frames located at the top of the structure are called key frame, and it is encoded in regular intervals (usually a multiple of 4). A key frame and all frames that are temporally located between two key frames construct a Group of Pictures (GOP), as illustrated in Fig. 4. Generally the GOP length is 4, 8, or 16. Hierarchical B frames with a larger TL Identifier (ID) refer to higher TL, while a smaller TL ID refers to lower TL [11]. The key frames are either intra-coded or inter-coded using the previous encoded key frames as references. The B frames in a GOP are hierarchical predicted and coded using the frames of higher temporal as references. Unlike the traditional IPPP and IBBP structure, the frames of lower layers are referenced by the frames of higher layers, resulting in complex quality dependency among Tls.

2.3 Related bit allocation scheme in MVC Hierarchical B prediction structure

The temporal and inter-view hybrid hierarchical B prediction structure is adopted in MVC. One of the problems arisen from the MVC hierarchical B prediction structure is how to allocate source bits efficiently among different view and temporal layers.

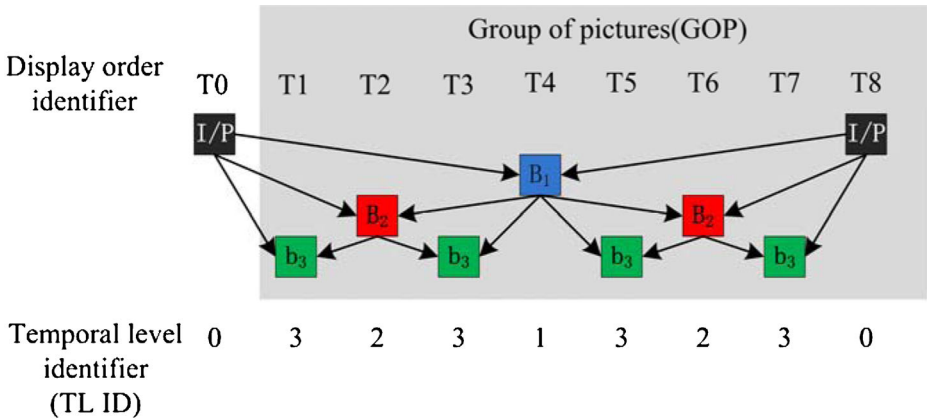


Fig. 4 The hierarchical reference picture structure for temporal prediction

2.3.1 The view layer bit allocation

In the view layer, the target bits for each view can be computed with the number of views and the weighting factor of each view. In MVC, each view has different priorities. The weighting factors can be set according to the users, depending on the application or priority of the view. Furthermore, the bit allocation between views also depends on whether the current view belongs to the basic views or enhancement views and more bits can be allocated to the basic views than the enhancement ones. The target bits allocated to each view can be derived from [31]:

$$T_v = w_v \cdot T_{tot} \tag{4}$$

where T_{tot} is the total bits allocated to all the views, T_v is the target bits allocated to the current view. w_v is a weighting factor, and its value depends on the features of the sequence and the importance of the current view.

2.3.2 The frame layer bit allocation

In the frame layer, as shown in Fig. 4, the B frames of such a typical hierarchical structure are predicted by using the two nearest pictures of the lower TLs as references. Xu et al. [26] took this feature into consideration in the bit allocation. They allocated more target bits for the frames which would be used as references to the following B-frames. The principle of this method is that the B-frames of different TLs carry varying importance to the entire coding quality, and the frame level target bits can be allocated according to the weighting factors in the hierarchical B prediction structure. The weighting factors w_I , w_P and w_k can be depicted as [26]:

$$\begin{cases} X_{tot} = N_I X_I + N_P X_P + \sum_k N_B(k) X_B(k) \\ w_I = \frac{X_I}{X_{tot}} \\ w_P = \frac{X_P}{X_{tot}} \\ w_k = \frac{X_B(k)}{X_{tot}} \end{cases} \tag{5}$$

where k ranges from 0 to D and denotes the hierarchy level from the lowest TL to the highest TL, D is the maximum TL ID, X_I , X_P , and X_B represent the complexities of the I-frame, P-frame

and B-frame respectively, X_{tot} is the total complexity of the frame in the current GOP. The complexity is defined as the product of bits and quantization step. N_I and N_P represent the numbers of the I-frame and P-frame in the current GOP, and $N_B(k)$ is the number of B frames in the k -th TL. With the weighting factors, the target bits allocated to each frame can be calculated as:

$$\begin{cases} T_{l,i}^B = \frac{w_l B_l(i)}{\sum_{k=l}^D w_k N_B^i(k)} \\ T_{I/P} = \frac{w_{I/P} B_{tot}}{w_{I/P} + \sum_{k=1}^D w_k N_B(k)} \end{cases} \quad (6)$$

where T is the allocated bits for the current frame, l is the number of remaining TLs that have not been encoded and ranges from 0 to D . $N_B^i(k)$ is the number of the frames that have not been encoded in the k -th TL, $B_l(i)$ is the remaining bits when coding the i -th frame of the l -th hierarchical level in the current GOP, and B_{tot} is the total bits allocated for the current GOP.

3 The rate control algorithm with bi-directional MAD prediction model and enhanced header bits prediction model

The traditional rate control algorithms are not suitable for the hierarchical B prediction structure. Considering the quadratic R-D model shown in (2), the efficiency of rate control can be improved by accurately predicting MAD and header bits. While conventional schemes for MAD and header bits prediction do not consider the dependency relationship of TLs in hierarchical B prediction structure. To address this shortcoming, we propose a rate control algorithm with bi-directional MAD prediction model and improved header bits prediction model based on the properties of hierarchical video coding structure.

3.1 The bi-directional MAD prediction model

The residual component MAD can be a good indicator of the encoding complexity [29]. The accuracy of MAD is essential to the accuracy of rate control. As shown in Fig. 5, we propose a MAD prediction model which makes use of the relationship



Fig. 5 The prediction of MAD model for the proposed rate control algorithm

between different TLs. The MAD of the current frame (F_c) is predicted from the former frame (F_{p1}) and the latter one (F_{p2}) in the nearest lower TLs. Then the MAD prediction model can be proposed as:

$$MAD_c = C_1 \cdot MAD_{p1} + C_2 \cdot MAD_{p2} + C_3 \tag{7}$$

where MAD_{p1} and MAD_{p2} are the actual MADs of F_{p1} and F_{p2} . C_1 , C_2 , and C_3 are the model coefficients. They are set as 0.5, 0.5, and 0 initially, and are updated using the regression model. Define matrices M_p , C , and M_c as:

$$M_p = \begin{bmatrix} MAD_{p1}(1) & MAD_{p2}(1) & 1 \\ MAD_{p1}(2) & MAD_{p2}(2) & 1 \\ \vdots & \vdots & \vdots \\ MAD_{p1}(n) & MAD_{p2}(n) & 1 \\ \vdots & \vdots & \vdots \\ MAD_{p1}(N) & MAD_{p2}(N) & 1 \end{bmatrix} \tag{8}$$

$$C = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} \tag{9}$$

$$M_c = \begin{bmatrix} MAD_c(1) \\ MAD_c(2) \\ \vdots \\ MAD_c(n) \\ \vdots \\ MAD_c(N) \end{bmatrix} \tag{10}$$

where $MAD_{p1}(n)$, $MAD_{p2}(n)$, and $MAD_c(n)$ are the actual MADs of the encoded frames at the location n in the sliding windows. n ranges from 1 to N , N is the number of selected data samples [8]. Then the model coefficients can be estimated by

$$C = (M_p^T M_p)^{-1} M_p^T M_c. \tag{11}$$

3.2 The bit allocation and header bits prediction

In the view layer, the target bits allocated to each view can be obtained from (4). Considering the quality dependency among the view layers, we set the QP of key frame by increasing from lower view layers to higher layers. The QP of the key frame in the GOP is depicted as follows.

$$QP_\alpha = bQP + \alpha \tag{12}$$

where QP_α is the QP of the key frame. α indicates the type of key frames in different view layers. $\alpha = 0, 1, 2$ represent the key frame being I-frame, P-frame, and B-frame respectively. bQP is the basis QP that has been predefined.

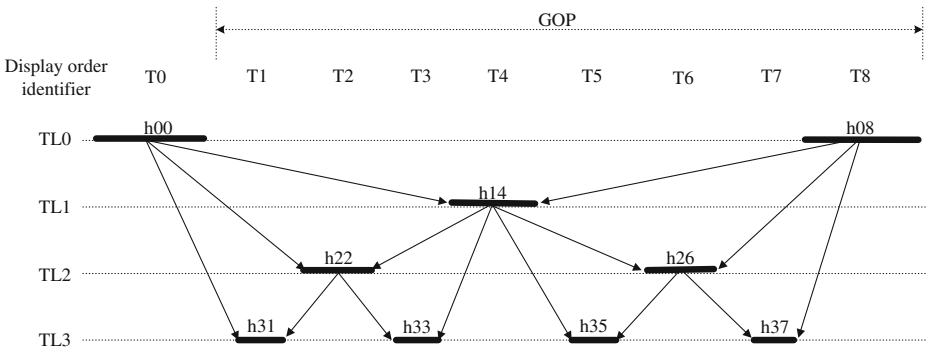


Fig. 6 The relation of header bits in different TLs of hierarchical B prediction structure

In the frame layer, we allocate the bits for the current frame in the same way as (5), and take the fixed QP to encode the key frame and the B-frame of first TL. The relation between them can be depicted as follows.

$$QP_{B_1} = QP_a + \Delta QP \tag{13}$$

where QP_{B_1} is the QP of the B-frame in the first TL, ΔQP is a constant which is related with the length of GOP.

The target bits for the rest of B-frames can be obtained from (6). We utilize the fixed QP scheme to encode the first GOP and compute the weighting factors according to the results of the first GOP. However, the weighting factors derived from this method is not so accurate, especially when the scene changes frequently. To maximize the coding efficiency, it is desirable to set the weighting factors adaptively to different video content [14], which will be considered in our next step work.

In the quadratic RD model shown in (2), the number of header bits is also essential for calculating the QP accurately. In JVT-G012, the number of header bits of the current frame is estimated from the average header bits of the previous encoded pictures. Due to the hierarchical structure in MVC, the header bits of the frames from different TLs are different from each other. The header bits prediction method in JVT-G012 is thus inaccurate when it is directly applied to the rate control of MVC. Figure 6 shows the relation of header bits in different TLs of hierarchical B prediction structure, in which four TLs are denoted by TL0, TL1, TL2 and TL3. In the hierarchical video coding structure, the frames in the same TL exhibit similar characteristics in motion vectors and coding mode, especially in the constant or slow-changing scene, and the numbers of header bits are closely related within the same TL. For example, in TL2, the header bits h26 is similar with h22. The frames in different TLs show different characteristics, and the header bits can be estimated using a proportional relationship of the header bits considering the hierarchical structure. According to this basic guideline, we use the following method to predict the header bits of the current frame, in which the number of header bits of the first frame in each TL is predicted by the weighted sum of the encoded frame header bits in the lower TLs, and the header bits of other frames are predicted by the average header bits of previous encoded frames in the same TL. For example, h22 in TL2 has different characteristics with frames in TL0 and TL1, and h22 can be predicted by

the weighted sum of h00, h08 in TL0 and h14 in TL1. In our proposed model, key frames and the B-frame of first TL are not taken into consideration when using fixed QP. The header bits of first frame in the rest of Tls, H_c , can be predicted as follows.

$$H_c = \frac{\sum_{k_e=0}^{k_l} (a_{k_e} \cdot SH_{k_e})}{\sum_{k_e=0}^{k_l} s_{k_e}} \tag{14}$$

where k_e is the TL ID for the encoded frame, k_l is the maximum TL ID of the encoded Tls, and s_{k_e} is the number of the encoded frames in the k -th TL.

SH_{k_e} is the sum of header bits of the encoded frame in the k -th TL and is given by:

$$SH_{k_e} = \sum h_{k_e,p} \tag{15}$$

where $h_{k_e,p}$ is the number of header bits of the encoded frame with the display order identifier of p in the k -th TL. For example, in TL2, SH_2 is the sum of h22 and h26.

The weighting factor of the header bits, a_{k_e} , can be obtained by:

$$a_{k_e} = \prod_{m=0}^{k_e} r_m \tag{16}$$

where r_m is the ratio of average header bits between $(m+1)$ -th TL and m -th TL and is calculated with the header bits obtained from the result of the first GOP. For example, a_1 is the product of r_0 and r_1 , r_0 is the ratio of SH_1 and $SH_0/2$, r_1 is the ratio of $SH_2/2$ and SH_1 .

4 Experimental results

To evaluate the performance of our proposed rate control algorithm, several experiments have been conducted. The sequences used in our experiments are of different properties, which include “ballroom” and “exit”. All the test sequences are in 4:2:0 formats. The features of the sequences are shown in Table 1. The simulation is implemented on the reference software JMVC8.5 and the target bit rate is generated by coding sequences with a fixed QP. The value of ΔQP is set to 3, which is the same as the default value in JMVC. The test conditions are shown in Table 2.

In order to test the performances of our proposed MAD and header bits prediction method, the benchmark method is developed by using 1) the view layer bit allocation method of (4) derived from [31], 2) the frame layer bit allocation method of (6) proposed in [26], combined with 3) the exiting linear MAD and header bits predicted model in JVT-

Table 1 The features of the testing sequences

Sequences	Picture resolution	Camera distance	views
Ballroom	640*480	20 cm	0–7
Exit	640*480	20 cm	0–7

Table 2 The testing conditions

Frame rate	Search range	Symbol model	Frames to be encoded	GOP size	BiPredlter
25	64	CABAC	113	8	4

Table 3 The results of the proposed MAD prediction model

Sequences	QP	JMVC	JVT-G012	PSNR	Proposed-MAD		ΔR_{RCE}
		Target bit rate	Actual bit rate		Actual bit rate	PSNR	
Exit	22	853.147	923.724	39.732	928.233	39.771	-0.529
	24	595.307	761.129	39.212	636.684	39.168	20.904
	26	425.221	476.887	38.530	459.467	38.532	4.097
	28	314.771	370.998	37.856	332.2947	37.838	12.296
Ballroom	22	1555.629	1703.542	38.724	1605.668	38.784	6.292
	24	1177.202	1310.843	37.852	1208.765	37.929	8.671
	26	890.354	1021.609	36.974	921.679	37.013	11.224
	28	676.741	736.983	35.954	714.864	36.001	3.268

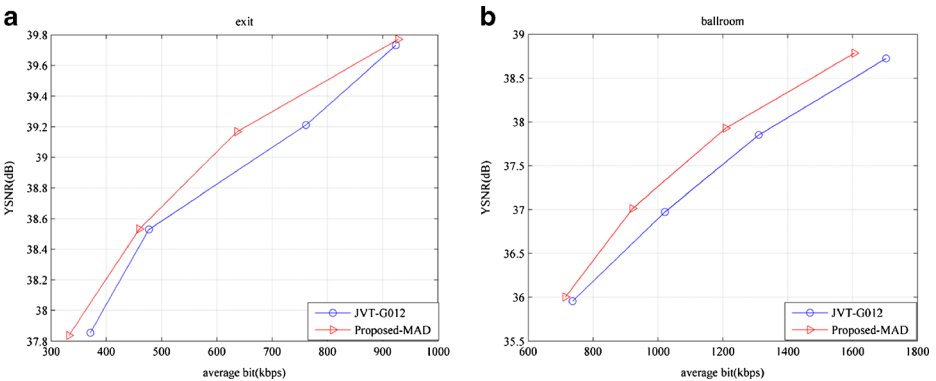


Fig. 7 The RD results for sequences **a** “exit” encoded at QP=22, 24, 26, 28, **b** “ballroom” encoded at QP=22, 24, 26, 28

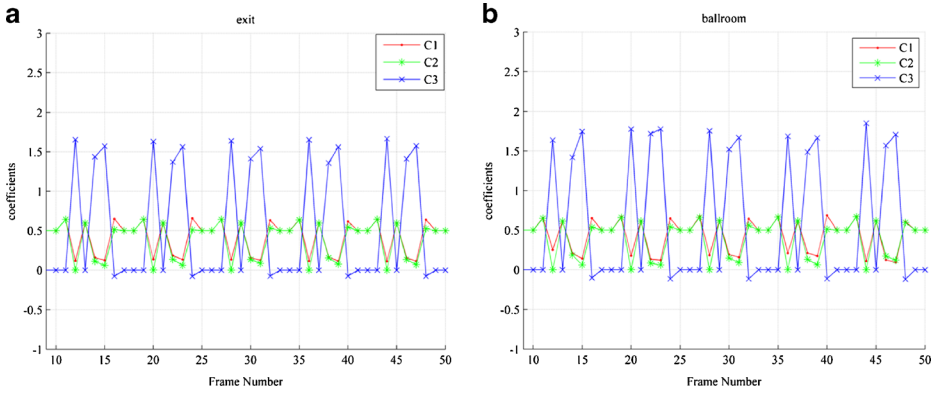


Fig. 8 The variation of model coefficients for base view of **a** “exit” encoded at QP=22, **b** “ballroom” encoded at QP=22

G012. The Rate Control Error (RCE) [28] is used to measure the accuracy of the bit rate estimation. The reduced RCE can be depicted as follows.

$$\Delta R_{RCE} = \frac{R1_{actual} - R2_{actual}}{R_{target}} \cdot 100\% \tag{17}$$

where R_{target} is the target bit rate, $R1_{actual}$ is the actual bit rate produced by the existing method and $R2_{actual}$ is the actual bit rate produced by the proposed methods.

First, we make some experiments to compare our proposed bi-directional MAD prediction model with the linear MAD prediction model in JVT-G012. In these experiments, the original header bits prediction model is adopted. The differences between actual bit rate and the target bit rate are showed in Table 3. The coding performances are illustrated in Fig. 7. The “Proposed-MAD” refers to our proposed MAD prediction model. The “JVT-G012” means the existing MAD prediction model in JVT-G012. ΔR_{RCE} is the reduced mismatch percentage of the actual bit rate and the target bit rate. As the information from the bi-directional reference is taken into consideration, we get more accurate bit rate. The average ΔR_{RCE} of our proposed

Table 4 The results of the proposed header bits prediction model

Sequences	QP	JMVC	JVT-G012	PSNR	Proposed-HB	PSNR	ΔR_{RCE}
		Target bit rate	Actual bit rate		Actual bit rate		
Exit	22	853.147	923.724	39.732	976.958	39.774	-6.240
	24	595.307	761.129	39.212	738.614	39.189	3.782
	26	425.221	476.887	38.530	453.838	38.509	5.420
	28	314.771	370.998	37.856	358.362	37.846	4.104
Ballroom	22	1555.629	1703.542	38.724	1643.031	38.690	3.890
	24	1177.202	1310.843	37.852	1336.156	37.870	-2.150
	26	890.354	1021.609	36.974	973.122	36.952	5.446
	28	676.741	736.983	35.954	701.165	35.936	5.293

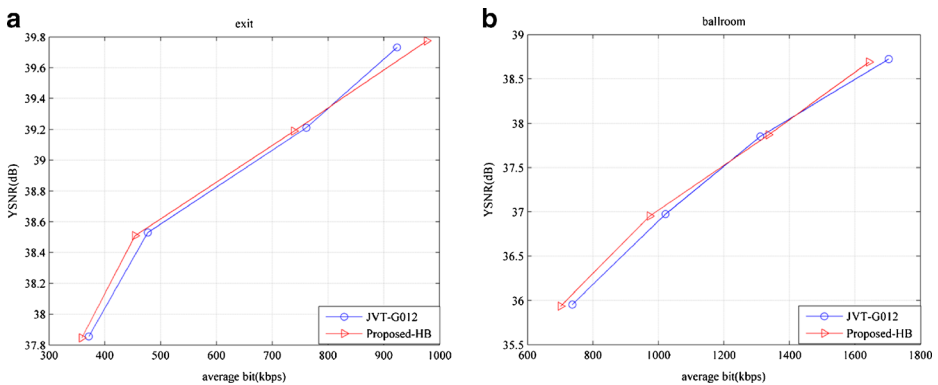


Fig. 9 The RD results for sequences **a** “exit” encoded at QP=22, 24, 26, 28, **b** “ballroom” encoded at QP=22, 24, 26, 28

MAD prediction model can reach up to 9.192 % for “exit” and 7.364 % for “ballroom”. In addition, we can also achieve better coding performance than the model in JVT-G012. To further investigate the effectiveness of the proposed MAD prediction model, we test the variation of model coefficients C_1 , C_2 , and C_3 in coding procedure. It is observed from Fig. 8 that both the forward and backward prediction information is utilized for predicting MAD, our proposed MAD prediction model is able to get better prediction.

Then, we implement some simulations to compare our proposed header bits prediction model with the existing header bits prediction method in JVT-G012. In these experiments, the original MAD model is adopted. Table 4 gives the differences between the actual bit rate and the target bit rate. Figure 9 shows the coding performances. The “Proposed-HB” indicates our proposed header bits prediction model. The “JVT-G012” means the existing header bits prediction method in JVT-G012. From the results, we can see that our proposed method achieve more accurate bit rate due to utilizing the relationship of the header bits between different TLs in hierarchical prediction structure. The average ΔR_{RCE} of our proposed header bits prediction model can reach up to 1.767 % for “exit” and 3.120 % for “ballroom”. Furthermore, we achieve more accurate bit rate as well as better coding performance than the model in JVT-G012.

Table 5 The results of the proposed RC algorithm

Sequences	QP	JMVC	JVT-G012	PSNR	Proposed-RC	PSNR	ΔR_{RCE}
		Target bit rate	Actual bit rate		Actual bit rate		
Exit	22	853.147	923.724	39.732	922.867	39.771	0.100
	24	595.307	761.129	39.212	656.449	39.167	17.584
	26	425.221	476.887	38.530	446.768	38.532	7.083
	28	314.771	370.998	37.856	330.832	37.843	12.760
Ballroom	22	1555.629	1703.542	38.724	1605.368	38.786	6.311
	24	1177.202	1310.843	37.852	1209.178	37.913	8.636
	26	890.354	1021.609	36.974	911.494	37.032	12.368
	28	676.741	736.983	35.954	693.402	36.037	6.440

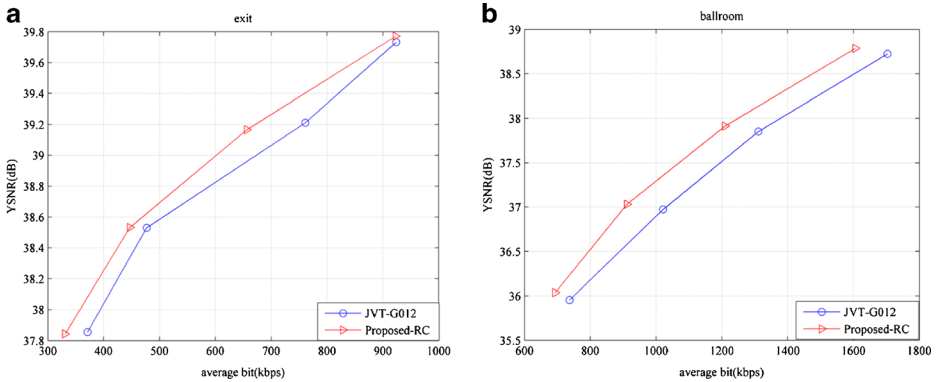


Fig. 10 The RD results for sequences **a** “exit” encoded at QP=22, 24, 26, 28, **b** “ballroom encoded at QP=22, 24, 26, 28

Finally, we compare the coding performance of using both of our proposed MAD prediction model and header bits method, with the linear MAD prediction model and the existing header bits method in JVT-G012. Table 5 shows the differences between the actual bit rate and the target bit rate. The “Proposed-RC” means combining both of our bi-directional MAD prediction model and enhanced header bits prediction model. The “JVT-G012” means the existing rate control algorithm in JVT-G012. From Table 5, the average ΔR_{RCE} of our proposed rate control algorithm are 9.382 % and 8.439 % respectively for the sequence “exit” and “ballroom”. Figure 10 gives the coding performances. It is observed that our proposed method is able to improve the coding performance. The experimental results demonstrate that our proposed rate control algorithm can achieve both accurate target bit rate and good coding performance.

5 Conclusion

In this paper, we propose an efficient rate control algorithm for the hierarchical B prediction structure which has been adopted in MVC. First, a bi-directional MAD prediction method based on the quadratic R-D model is introduced according to the prediction characteristics of the B frames. Second, an enhanced header bits prediction method is described for computing the header bits in the hierarchical B prediction structure according to the similar characteristics of previous encoded frames. The experimental results show that our proposed rate control algorithm outperforms the methods that directly apply the existing rate control algorithm to MVC. The proposed algorithm can achieve more accurate bit rate and better coding performance than the method used in JVT-G012. It is worthwhile to point out that the proposed MAD and header bits prediction method may be further adapted to the rate control of H.264/AVC and SVC with hierarchical B prediction structure, which will be our future work.

Acknowledgment The authors would like to thank the reviewers for their constructive and valuable comments on this paper. We would also like to thank Prof. Zhenhua Ling, Dr. Haoming Chen and Dr. Chen Zhao for comments and suggestions. This research was partially supported by the Natural Science Foundation of China (No.61271324, 60932007, 61002029, 61001178, 61202266), Natural Science Foundation of Tianjin (No.12JCYBJC10400, 12JCQNJC00500, 12JCQNJC00300), and Research Fund for the Doctoral Program of Higher Education of China (No.20110032120029).

References

1. Chen Y, Pandit P, and Yea S (2009) WD 4 reference software for MVC (JMVC). ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-AD207
2. Chi MC, Chen MJ, Hsu CT (2004) Region-of interest video coding by fuzzy control for h.263+ standards. In: IEEE Int Symp and Circuits Syst (ISCAS), pp 93–96
3. Cho YJ, Kuo CCJ, Kwon DK (2007) Gop-based rate control for h.264/svc with hierarchical B-pictures. In: IEEE IntInf. Hiding Multimed Signal Processing (IHMSP), pp 387–390
4. Cho YJ, Liu JY, Kwon DK, and Kuo CCJ (2009) H.264/SVC temporal bit allocation with dependent distortion model. In: IEEE Int Conf Acoustics Speech Signal Processing (ICASSP), pp 641–644
5. Hu SD, Wang HL, Kwong S, Zhao TS, Kuo CCJ (2011) Rate control optimization for temporal-layer scalable video coding. IEEE Trans Circ Syst Video Technol 21(8):1152–1162
6. ISO/IEC JTC1/SC29/WG11 (2005) Survey of algorithms used for multi-view video coding (MVC). Doc. N6909
7. ITU-T Recommendation H.264 (2010) Advanced video coding for generic audio visual services
8. Lee HJ, Chiang TH, Zhang YQ (2000) Scalable rate control for MPEG-4 video. IEEE Trans Circ Syst Video Technol 10(6):878–894
9. Lee JY, Wey HC, Park DS (2011) A fast and efficient multi-view depth image coding method based on temporal and inter-view corrections of texture images. IEEE Trans Circ Syst Video Tech 21(12):1859–1868
10. Li ZG, An P, Yan Tao LF, Zhang ZY (2009) Macroblock layer rate control in multi-view video coding. J Appl Sci 27(5):502–507
11. Li M, Chang YL, Yang FZ, Wan S, Lin SX, Xiong LH (2009) Frame layer rate control for H.264/AVC with hierarchical B-frames. Signal Process: Image Commun 24:177–199
12. Li ZG, Pan F, Lim KP, Feng G, Lin X, Rahardja S (2003) Adaptive basic unit layer rate control for JVT. ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Doc. JVT-G012
13. Lim J, Kim J, Ngan KN (2003) Advanced rate control technologies for 3D-HDTV. IEEE Trans Consum Electron 49(4):1498–1507
14. Liu Y, Li ZG, Soh YC (2008) Rate control of H.264/AVC scalable extension. IEEE Trans Circ Syst Video Technol 18(1):116–121
15. Liu F, Xiong J, Fan JJ, Lin QJ (2011) Efficient rate control algorithm for multi-view video coding. China Commun:83–89
16. Merkle P, Müller K, Wiegand T (2010) 3D video: acquisition, coding, and display. IEEE Trans Consum Electron 56(2):946–950
17. Merkle P, Smolić A, Müller K, Wiegand T (2007) Efficient prediction structures for multiview video coding. IEEE Trans Circ Syst Video Tech 17(1):1461–1473
18. Müller K, Merkle P, Schwarz H (2006) Multiview video coding based on H.264/MPEG4-AVC using hierarchical B pictures. In: Picture Coding Symposium, pp 24–26
19. Natio S, Matsumoto S (1999) 34/35Mbps 3D-HDTV digital coding scheme using a modified motion compensation with disparity vectors. In: SPIE Vis Commun Image Process (VCIP), pp. 1082–1089
20. Park S, Sim DY (2009) An efficient rate-control algorithm for multi-view video coding. In: IEEE Int Symp Consum Electron (ISCE), pp 25–28
21. Seo CW, Kang JW, Han JK, Nguyen TQ (2010) Efficient bit allocation and rate control algorithms for hierarchical video coding. IEEE Trans Circ Syst Video Technol 20(9):1210–1223
22. Shen LQ, Liu Z, Zhang ZY (2011) A novel H.264 rate control algorithm with consideration of visual attention. Multimed Tools Appl. doi:10.1007/s11042-011-0893-z
23. Son NR, Shin YJ, Yoo JM, Lee GS (2006) A new macroblock-layer rate control for H.264/AVC using Quadratic R-D model. Lect Notes In Comput Sci 4319:822–831
24. Vetro A, Wiegand T, Sullivan GJ (2011) Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. Proc IEEE 99(4):626–642
25. Woo W, Ortega A (1999) Optimal blockwise dependent quantization for stereo image coding. IEEE Trans Circ Syst Video Technol 9(6):861–867
26. Xu L, Gao W, Ji XY, Zhao DB (2007) Rate control for hierarchical B-picture coding with scaling-factors. In: IEEE Int Symp and Circuits Syst (ISCAS), pp 27–30
27. Xu L, Kwong S, Zhao TS, Zhou Y (2011) Priority pyramid based bit allocation for multiview video coding. In: IEEE Vis Commun Image Processing (VCIP), pp 6–9
28. Yan T, An P, Shen LQ, Zhang Q, Zhang ZY (2009) Rate control algorithm for multi-view video coding based on correlation analysis. In: IEEE Symp On Photonic Opo electron (SOPO), pp 14–16
29. Yi XQ, Ling N (2006) Improved H.264 rate control by enhanced MAD-based frame complexity prediction. J Vis Commun Image Represent 17(2):407–424

30. Zeng HQ, Ma KK, Cai CH (2011) Fast mode decision for multiview video coding using mode correction. *IEEE Trans Circ Syst Video Tech* 21(11):1659–1666
31. Zhu ZJ, Liang F, Jiang GY, Yu M (2007) Bit allocation and rate control algorithm for stereo video coding. In: 3DTV conference, pp 7–9
32. Zhu C, Liu M (2009) Multiple description video coding based on hierarchical B pictures. *IEEE Trans Circ Syst Video Technol* 19(4):511–521



Jianjun Lei received the Ph.D. degree in signal and information processing from Beijing University of Posts and Telecommunications, Beijing, China, in 2007. He is currently an Associate Professor with the School of Electronic Information Engineering, Tianjin University, Tianjin, China. He is also a visiting researcher at the Department of Electrical Engineering, University of Washington, Seattle, WA, from August 2012. His research interests include 3D video processing, 3D display, and computer vision.



Kun Feng received the B.S. degree in telecommunication engineering from Tianjin University, Tianjin, China, in 2011. She is currently pursuing the M.S. degree at the School of Electronic Information Engineering, Tianjin University, Tianjin, China. Her research interests include multiview video coding and digital image processing.



Meimin Wu received the B.S. degree in communication engineering from Tianjin University, Tianjin, China, in 2011. She is currently a postgraduate majoring in information and communication engineering at the School of Electronic Information Engineering, Tianjin University, Tianjin, China. Her research interests include 3D imaging, video processing and multiview video coding.



Shuai Li received the B.S. degree in electronic engineering from Yantai University, Yantai, China, in 2011. He is currently pursuing the M.S. degree at the School of Electronic Information Engineering, Tianjin University, Tianjin, China. His research interests include multiview video coding and 3D imaging.



Chunping Hou received the M.Eng. and Ph.D. degrees, both in electronic engineering, from Tianjin University, Tianjin, China, in 1986 and 1998, respectively. Since 1986, she has been the faculty of the School of Electronic Information Engineering, Tianjin University, where she is currently a Full Professor and the Director of the Broadband Wireless Communications and 3D Imaging Institute. Her current research interests include 3D image processing, 3D display, wireless communication, and the design and applications of communication systems.