Signal Processing 112 (2015) 189–198

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

View synthesis distortion model based frame level rate control optimization for multiview depth video coding

Xu Wang [a,b], Sam Kwong [a,b,*], Hui Yuan [a,c], Yun Zhang [a,d], Zhaoqing Pan [a,e]

[a] Department of Computer Science, City University of Hong Kong, Hong Kong
[b] City University of Hong Kong Shenzhen Research Institute, Shenzhen 518057, China
[c] Key Laboratory of Wireless Sensor Network & Communication, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China
[d] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[e] School of Computer Science and Engineering, Hebei University of Technology, Tianjin 300401, China

ARTICLE INFO

Article history:
Received 28 February 2014
Received in revised form
5 June 2014
Accepted 27 June 2014
Available online 24 July 2014

Keywords:
Virtural view quality
Rate control
Nash bargaining solution
Multiview depth video coding

ABSTRACT

Quality of virtual view image is one of the important issues in multiview depth video coding (MDVC). When people enjoy three dimensional (3D) applications, they may have an unpleasant perceptual experience if the virtual view has low image quality. Thus, high virtual view image quality is mostly expected in the rate control optimization for MDVC with limited channel bandwidth and buffer resources. In this paper, a view synthesis distortion model is proposed first to indicate the importance of each frame in the depth video. Second, to achieve a balance between virtual view image quality and buffer constraint, the proposed view synthesis distortion model is incorporated in the bargain game theoretic model to handle the frame level bit allocation problem for Hierarchical B-picture (HBP) structure. Experimental results demonstrate that the proposed rate control algorithm achieves 0.18 dB BDPSNR gain in average compared to the benchmark. The average mismatch of the proposed algorithm is only 0.71%.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

With the rapid developments on content generation and 3D display technology [1], 3D applications and services are recently very common in nearly every facets of people's daily life [2,3]. For example, people can go to 3D movies for immersive viewing experiences, play 3D video games for attractive scenarios. With the help of 3D display, doctors can accomplish diagnose and surgeries. Compared to traditional single view video, 3D video contains higher dimensional visual information, which can improve the immersion experience [4] of user such as depth perception [5,6].

Currently, there are many types of video format to represent the 3D visual content [7,8]. The most popular one is multiview video plus depth (MVD) based representation [9]. The MVD based 3D video format contains two or more captured views with associated color and depth videos. The virtual view can be synthesized [10] through the depth based image rendering (DIBR) technique to support the multiview auto-stereoscopic display environment or the interactive free view access function. On the other hand, the MVD format is back-compatible with the existing display environment. However, MVD based 3D video requires huge volume of bandwidth resources for

* Corresponding author at: Department of Computer Science, City University of Hong Kong, Hong Kong. Tel.: +852 3442 2907; fax: +852 3442 0503.
E-mail addresses: xuwang6-c@my.cityu.edu.hk (X. Wang), cssamk@cityu.edu.hk (S. Kwong), yuanhui0325@gmail.com (H. Yuan), zhangyun_8851@163.com (Y. Zhang), zqpan3-c@my.cityu.edu.hk (Z. Pan).

http://dx.doi.org/10.1016/j.sigpro.2014.06.025
0165-1684/© 2014 Elsevier B.V. All rights reserved.

storage and transmission [11]. To compress large volume of 3D video data, multiview color video (MCV) is encoded by the mulitview video coding (MVC) [12] directly. Since the multiview depth video (MDV) can be treated as the illumination component of the color video, the MDVC is developed based on the MVC [13].

Rate control (RC) aims to achieve a balance between image quality, quality smoothness and buffer smoothness under the bandwidth constraint, which is indispensable for MDVC with respect to bitstream storage and video streaming usage. A general RC scheme contains two basic components including the bit allocation and quantization parameter (QP) determination. For MDVC, the bit allocation problems exist in hierarchical coding units. For example, the bit allocation problem can be decomposed as view level, frame level, macroblock (MB) level, which makes the RC optimization in MDVC more complicated.

Currently, existing works mainly focus on the view level bit allocation. For example, Liu et al. [14] proposed a multi-pass algorithm to maximize the image quality of synthesized virtual view. To reduce the computation complexity, Yuan et al. [15] proposed a linear distortion model of virtual view based on theoretical analysis. Hu et al. [16] exploited the rate–distortion (R–D) properties among views to maximize the sum of the quality of color videos and virtual view videos. These algorithms mentioned above try to seek a optimal combination of bit allocation among views and determine a initial QP. However, encode each view by a fixed QP may not help for achieving buffer smoothness and reducing bit rates mismatch. Liu et al. [17] proposed a joint RC algorithm that the frame level rate controller was implemented when the target bit rates for each view is determined. Shao et al. [18] extended the frame level RC scheme in [19] to support MVC. However, existing works on frame level RC tries to minimize the total sum of depth distortion, instead of the distortion of virtual view images. Thus, the overall performance improvement is limited.

In this paper, a view synthesis distortion (VSD) model is firstly proposed as the optimization object in frame level bit allocation. To achieve a balance between virtual view image quality and buffer constraint, the proposed VSD model is incorporated in the bargaining game theoretic model to handle the frame level bit allocation problem of HBP structure. Third, an overall algorithm is developed to accomplish the bit rate control.

The rest of this paper is organized as follows. In Section 2, the VSD model is described. Section 3 describes the VSD based frame level bit allocation game. Then, the overall RC optimization algorithm is provided in Section 4. In Section 5, the experimental results are provided to demonstrate the efficiency of the proposed rate control algorithm. Finally, the summary is provided in Section 6.

2. Proposed view synthesis distortion model for multiview depth video coding

Fig. 1 shows a typical 3D video system. In the system, the multiview color and depth videos are encoded at server and the bitstreams are transmitted to the client for decoding. The MCV and MDV are encoded separately. The rate controller adjusts the coding parameters of encoder to balance the image quality and buffer smoothness under the channel constraints. Then, the virtual views are synthesized by the reconstructed color and depth videos [10] to support the multiview auto-stereoscopic display environment or the interactive free view access function. In this work, we focus on the frame level RC problem of MDVC.

During the view synthesis stage, each virtual view is rendered by two coded views with their associated color and depth videos as shown in Fig. 2. Therefore, the distortion of referenced color or depth video will affect the distortion of numbers of virtual view. To optimize the virtual view image quality of the 3D video system under the bit rate constraints, a lot of theoretical virtual view distortion model has been proposed. For example, view level distortion models such as linear model [14–16] and cubic model [20] were employed to solve the view level bit allocation problem. To optimize the R–D performance of MB-level bit allocation, regional-level distortion model [13,21] and MB-level distortion model [22] were proposed for depth video coding. For the mathematical simplicity and fitting accuracy, the linear model [15] between the distortion (in terms of MSE) of virtual view and its reference color and depth videos is employed in Eq. (1). For better observation, an example is provided in Fig. 3(a) where the distortion surface is close to a planar.

$$D_v^l = \sum_{n \in \mathbf{A}_l} (\xi_{n,l} D_d^n + \eta_{n,l} D_t^n) + \rho_l. \tag{1}$$

The subscript l stands the virtual view index; $\mathbf{A}_l$ is the set of index of coded views which is used to synthesize virtual view l, e.g., $\mathbf{A}_2 = \{1, 3\}$ and $\mathbf{A}_4 = \{3, 5\}$; D_v^l is the distortion of virtual view l; D_d^n and D_t^n are the distortion of depth video and color video for coded view n, respectively; $\xi_{n,l}$ and $\eta_{n,l}$ are the model parameters.

After determining the target bit rates for each view, traditional frame level rate controller tried to maximize the R–D performance of depth video where the D is in terms of the distortion of depth video. However, the depth video is for virtual view image rendering. The optimizing criteria for 3D system should be replaced by the virtual view image quality. On the other hand, for the frames belonging to the same view of depth video, the contributions of the frame on the virtual view quality may be different. Fig. 3(b) shows the model parameters a of virtual views in Eq. (1) respects to the frame index, where the model parameters are varied along the frame index. Therefore, in order to optimize the overall visual quality of virtual view, the frame level bit allocation of MDVC should consider the contributions difference among the frames.

For the coded view n of depth video, the VSD $P_{n,j}$ of frame j is defined as

$$P_{j,n} = \alpha_{j,n} \cdot D_d^{j,n}, \tag{2}$$

where

$$\alpha_{j,n} = \frac{1}{|\mathbf{V}_n|} \sum_{l \in \mathbf{V}_n} \xi_{n,l}^j. \tag{3}$$

$\mathbf{V}_n$ is the set of index of virtual view that refer to the coded view n, e.g. $\mathbf{V}_1 = \{2\}$, $\mathbf{V}_3 = \{2, 4\}$, and $\mathbf{V}_5 = \{4\}$. The

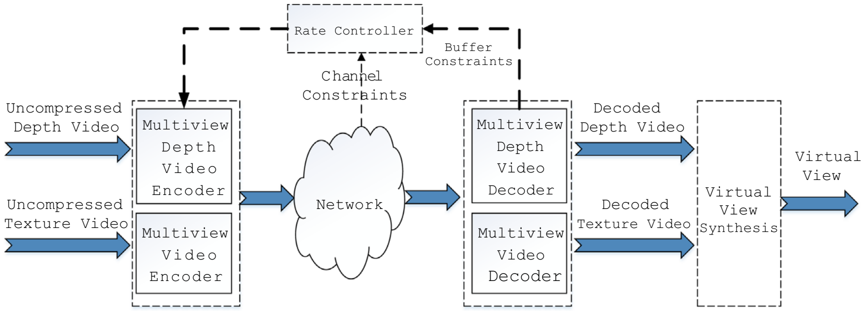

Fig. 1. Example of typical 3D video system. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

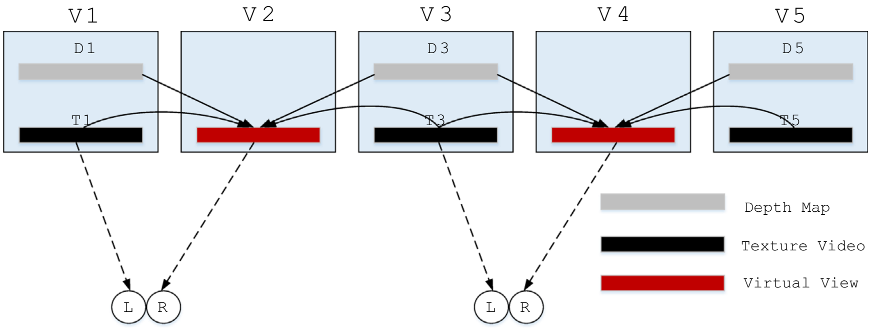

Fig. 2. Illustration of the rendering reference relationship of the virtual views and the coded views [16]. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

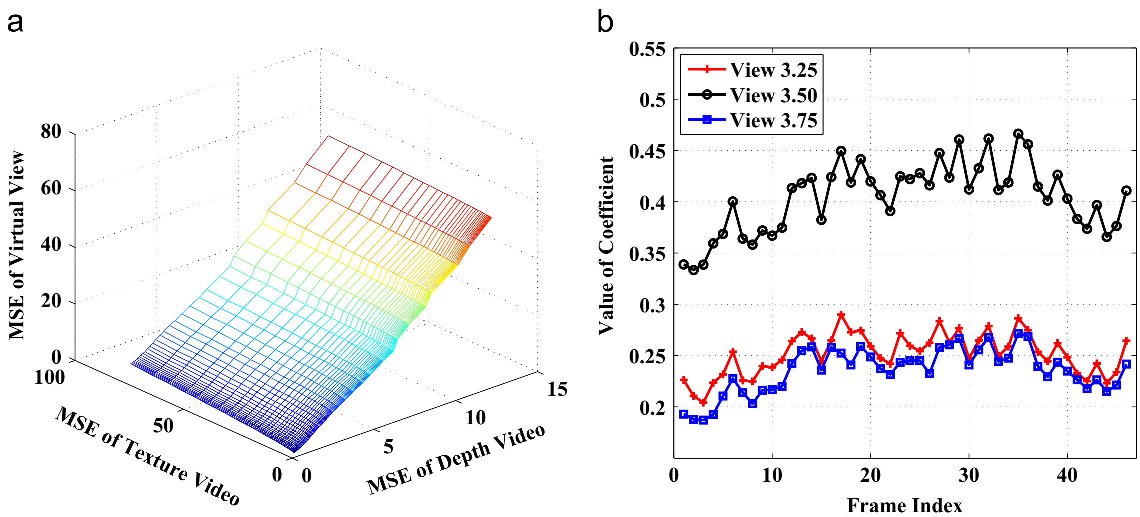

Fig. 3. Illustration of the virtual view distortion for 3D sequence “PoznanStreet”. (a) Effects of color and depth video distortion on the virtual view distortion. (b) Model parameter of virtual views against the frame index. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

expression |.| is the cardinality of a set, i.e., the number of elements of the set.

To achieve high coding performance, MDVC employs the HBP structure as the basic coding structure. Fig. 4 shows an example of MDVC with two views. The R–D characteristics of different TLs in HBP structure for depth video are quite different as shown in Fig. 5. The number of bits generated by the frame of lower TLs is generally larger than that of higher TLs with the same quantization step size Q. The slopes of the $D-Q$ lines are also significantly different. Therefore, a single R–D model is not accurate enough to describe the R–D relationship for all the frames in different TLs. For the mathematical simplicity and fitting accuracy, the linear R–D model [23] based on the Laplacian

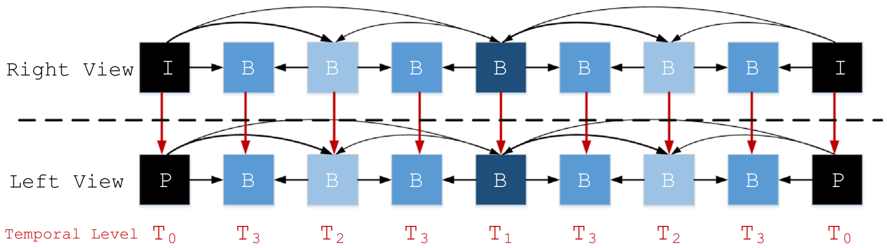

Fig. 4. Illustration of coding structure for MDVC with two views.

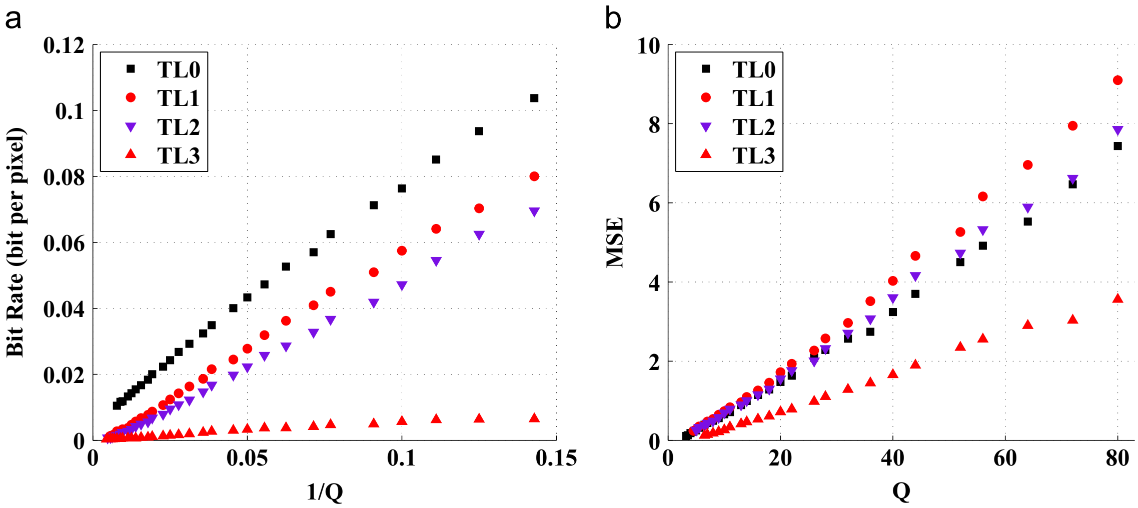

Fig. 5. R–D relationship for the 16th picture in TL 0, the eighth picture in TL 1, the fourth picture in TL2 and the second picture in TL 3 within the same GOP (the GOP size is 15). (a) Relationship between R and $1/Q_{step}$ of the first GOP in depth video for 3D sequence "Newspaper"; (b) Relationship between D and Q_{step} of the first GOP in depth video for 3D sequence "Newspaper".

distribution assumption of transform coefficients [24] is adopted to approximately represent the relationship of frames located in TL i as

$$R_{i,n}(Q_{i,n}) = \frac{k_{i,n} \cdot m_{i,n}}{Q_{i,n}} + C_{i,n}, \tag{4}$$

and

$$D_d^{i,n}(Q_{i,n}) = \gamma_{i,n} \cdot Q_{i,n}, \tag{5}$$

where $m_{i,n}$ indicates the MAD of the residual between the original and reconstructed frame, $k_{i,n}$ and $\gamma_{i,n}$ are the model parameters, and $C_{i,n}$ is the number of header bits.

3. Proposed VSD based frame level bit allocation game

For HBP structure in MDVC, the rate and distortion characteristics are more complex than those of traditional coding structures, which makes the frame level RC of MDVC more difficult than that of traditional coding structures [25]. Based on the reference dependency relationship, the frames of different TLs have different contributions to the R–D performance of depth video. In addition, the frames of depth video have different contributions to the virtual view image quality as discussed in Section 2. Therefore, to maximize the virtual view image quality, the frame level RC of MDVC should focus on how to optimal allocate the bandwidth among frames by considering the contributions difference among the frames. In previous RC works, the bits are allocated according to the weighting factor where the lower TLs will have greater chance to obtain more bits to minimize the average distortion [23,26]. However, these optimization methods will be unfair to the frames in the higher TL that obtain few bits due to inaccurate weighting factor.

To tackle the above-mentioned problem, a bargaining game theoretical model was proposed in [25], in which the bit allocation problem among frames that belong to different TLs was modeled as a bargaining game. To derive the theoretical model, each frames in the same GOP was regarded as a player and linear R–D model is employed to define the utility function of player. The proposed bit allocation game tries to seek the bargaining interactions among players achieve a good tradeoff between image quality and buffer constraint. However, the algorithm in [25] is not designed for the MDVC, thus extend the algorithm directly to support frame level RC of MDVC can not guarantee the optimal of virtual view image quality. In this section, the proposed VSD model is imported into the bit allocation bargaining game to update the utility function of player. The updated utility function is expected to improve the R–D performance by optimizing the virtual view image quality.

3.1. Frame level bit allocation problem formulation

Once the target bit rates of views are determined, the frame level bit allocation of HBP structures for each view can be regarded as the bargaining problem. The total number of view is denoted as L as shown in Fig. 4. For the view n, the basic components of bargaining game are described as follows.

Players: For convenience, we assume that current GOP contains S_n non-encoded frames. Each frame level rate controller is regarded as a player in the bit allocation game. They cooperate to divide the remaining bit rates of the current GOP R_c^n.

Preferences: During the bargaining game, each player will try to request for more bits to reflect its preference, which is measured by utility function. For player j_n, the utility is denoted as $U_{j,n}$.

Strategies: To achieve the utility $U_{j,n}$, the number of allocated bits consumed by player j_n is denoted as $R_{j,n}$. Given a combination of strategies carried out by all non-encoded frames $R_n=(R_{1,n}, R_{2,n}, \ldots, R_{S_n,n})$, then $U_n=(U_{1,n}\ (R_{1,n}), U_{2,n}(R_{2,n}), \ldots, U_{S_n,n}(R_{S_n,n}))$ is the corresponding utility of the game.

Initial Utility: To guarantee the basic requirement, the player j_n claims a initial utility $d_{j,n}$ that can be derived from the bit rate $R^0_{j,n}$. $\mathbf{d_n}=(d_{1,n}, d_{2,n}, \ldots, d_{S_n,n})$ is called disagreement point.

Assume the feasible set $\mathbf{U_n}=\{U_n\}\subset\mathbb{R}^{S_n}$ is nonempty, convex, closed and bounded, then the pair $(\mathbf{U_n}, \mathbf{d_n})$ defines the bargaining problem for the view n.

3.2. Definition of utility function

Based on the proposed distortion model in Eq. (2), the utility function of the player j_n with TL index t is defined as

$$U_{j,n}=\frac{1}{P_{j,n}}=\frac{R_{j,n}-C_{t,n}}{\beta_{t,n}}, \tag{6}$$

where $\beta_{j,n}=\alpha_{j,n}\gamma_{t,n}k_{t,n}m_{t,n}$. $R_{j,n}$ is the allocated bits for the player j_n involved in the bit allocation game of coded view n. It is easy to prove that the feasible utility set $\mathbf{U_n}$ is convex. Detailed proof can be referred in [25].

3.3. Nash bargaining solution

In the bargaining problem, the properties of feasible utility set will determine the selection criteria of bargaining solution. Since the utility set $\mathbf{U_n}$ is convex and bounded, the Nash bargaining solution (NBS) is adopted to obtain the optimal solution for the frame level bit allocation bargaining game.

Based on the definition, the solution R_n^* is an NBS of the view n, if and only if $\prod_{j=1}^{S_n}(U_{j,n}(R^*_{j,n})-d_{j,n})\geq\prod_{j=1}^{S_n}(U_{j,n}(R_{j,n})-d_{j,n})$ for all $U_n(R_n)\in\mathbf{U_n}$. Therefore, we can solve the following optimization problem (7) to obtain the NBS:

$$\begin{aligned}&\max_{R_n}\ \prod_{j=1}^{S_n}(U_{j,n}(R_{j,n})-d_{j,n})\\ &\text{s.t.}\ \sum_{j=1}^{S_n}R_{j,n}\leq R_c^n,\\ &j=1,\ldots,S_n.\end{aligned} \tag{7}$$

The problem can be transformed as

$$\begin{aligned}&\max_{R_n}\ \sum_{j=1}^{S_n}\ln(U_{j,n}(R_{j,n})-d_{j,n})\\ &\text{s.t.}\ \sum_{j=1}^{S_n}R_{j,n}\leq R_c^n,\\ &j=1,\ldots,S_n.\end{aligned} \tag{8}$$

Since the transformed problem is convex, the optimal solution can be obtained by solving the Karush–Kuhn–Tucker (KKT) conditions. Suppose the λ_n are the Lagrange multipliers, then we have the following Lagrangian function $L(R_{j,n}, \lambda_n)$:

$$\begin{aligned}L=&\sum_{j=1}^{S_n}\ln(U_{j,n}(R_{j,n})-d_{j,n})\\ &+\lambda_n\cdot\left(R_c^n-\sum_{j=1}^{N_s}R_{j,n}\right)\end{aligned} \tag{9}$$

The KKT conditions of (9) are written as follows:

$$\begin{cases}\dfrac{\partial L}{\partial R_{j,n}}=\dfrac{\partial\ln(U_{j,n}-d_{j,n})}{\partial R_{j,n}}-\lambda_n=0\\ \dfrac{\partial L}{\partial\lambda_n}=R_c^n-\sum_{j=1}^{S_n}R_{j,n}\geq 0\\ \lambda_n\left(R_c^n-\sum_{j=1}^{S_n}R_{j,n}\right)=0\end{cases} \tag{10}$$

where $j\in 1,\ldots,S_n$.

Based on the above KKT conditions (10), we have

$$\begin{aligned}&\frac{\partial L}{\partial R_{j,n}}=\frac{\partial\ln(U_{j,n}-d_{j,n})}{\partial R_{j,n}}-\lambda_n\\ &\Rightarrow R_{j,n}=\frac{1}{\lambda_n}+\beta_{j,n}d_{j,n}+C_{t,n}.\end{aligned} \tag{11}$$

and then

$$\begin{aligned}&R_c^n-\sum_{j=1}^{S_n}R_{j,n}=0\\ &\Rightarrow\frac{1}{\lambda_n}=R_c^n-\sum_{j=1}^{S_n}\left(\beta_{j,n}d_{j,n}+C_{t,n}\right).\end{aligned} \tag{12}$$

Therefore,

$$R_{j,n}=\frac{1}{\lambda_n}+\beta_{j,n}d_{j,n}+C_{t,n}. \tag{13}$$

To guarantee the buffer smoothness, the target bit rates should be clipped as follows:

$$R_{j,n}=\max(R^{\min}_{j,n}, \min(R_{j,n}, R^{\max}_{j,n})), \tag{14}$$

where $R^{\min}_{j,n}$ and $R^{\max}_{j,n}$ is determined by the lower bound and upper bound of the buffer. Finally, the optimal quantization step size for the player j_t is determined by

$$Q_{j,n}=\frac{\kappa_{t,n}m_{t,n}}{R_{j,n}-C_{t,n}}. \tag{15}$$

4. Proposed overall rate control algorithm

4.1. Parameter determination

To obtain a NBS of the frame level bit allocation bargaining game, the model parameters should be determined first.

4.1.1. Parameters determination of distortion model

For the view n, all the model parameters are unavailable. To obtain the parameter $\alpha_{j,n}$ of frame j, the depth videos are pre-encoded at Q^A and Q^B, then the parameter $\xi^j_{n,l}$ can be estimated as

$$\xi^j_{n,l} = \frac{D_v^{j,l,A} - D_v^{j,l,B}}{D_d^{j,n,A} - D_d^{j,n,B}}, \tag{16}$$

where $D_v^{j,l,A}$ and $D_v^{j,l,B}$ are the MSE of the jth frame in lth virtual view image synthesized with the corresponding original color videos and compressed depth videos pre-encoded at Q^A and Q^B, respectively; $D_d^{j,l,A}$ and $D_d^{j,l,B}$ are the MSE of the jth frame in nth depth video pre-encoded at Q^A and Q^B, respectively. In order to obtain the parameters, each depth video has to be pre-encoded twice, which can be implemented during the view level bit allocation.

The same method in [25] is employed to initialize the parameters of TL t. Once the frame located in TL t is encoded, the R–D model parameters such as $\kappa_{t,n}$, $m_{t,n}$ and $\gamma_{t,n}$ are updated by using the linear regression technique.

4.1.2. Parameters determination of bargaining game

Before the optimal allocation result from Eq. (14), the parameters $d_{j,n}$ and $\beta_{j,n}$ for the view n can be computed. The prediction of $d_{j,n}$ is based on the history information of previously encoded frames that located in the same TL. The estimation method is the same as in [25].

4.1.3. Buffer management and parameters updating

Before solving the optimization problem from Eq. (7), the remaining target bits R_c^n for the view n of current GOP is determined based on the buff status and target bandwidth. $R_{j,n}^{\min}$ and $R_{j,n}^{\max}$ is determined by the lower bound and upper bound of the buffer. Detailed buffer management and parameters updating methods can be referred in [27].

4.2. Overall algorithm

Suppose the depth video contains L views, the target bit rate for each view is already allocated and fixed. Then the proposed overall RC algorithm for depth video coding is described as follows.

Step 1: For each view, encode the depth video with two different QPs. Render the virtual view by the reconstructed depth videos. Initialize the parameters of models as described in Section 4.1.

Step 2: For the current view n, obtain the target bit rates according to the system setting or view level bit allocation algorithm.

Step 3: Obtain the remaining target bits of the current GOP k according to the buffer status.

Step 4: For the current frame $j_{t,n}$ in the GOP k, determine the target bits by (14). Compute quantization step $Q_{j,n}$ for the current frame $j_{t,n}$ by (15).

Step 5: Encode the current frame $j_{t,n}$ of depth video with the determined $Q_{j,n}$ of Step 4.

Step 6: Update the buffer status and model parameters. If the last frame of current GOP is not reached, go to Step 4.

Step 7: If all the views end, terminates procedure; else if the current view ends, goto Step 2 for next view, else goto Step 3.

5. Experimental results

To assess the performance of the proposed overall RC algorithm for depth video coding, we implement the algorithm in the reference software JMVC 8.5 [28]. Seven standard 3D video sequences with different resolution and motion properties are tested for the coding and rendering process. Two depth videos of each sequence were encoded as reference views to render three intermediated views. The reference virtual views rendered by the uncompressed color and depth videos are used as reference in measuring the virtual view image quality. Details of these test sequences and rendered views are listed in Table 1. The virtual views are rendered by the synthesis algorithm [29] included in the software 3D-HTM [30]. Table 2 provides the key parameters for the reference software JMVC. Each 3D sequences are firstly encoded by five different QPs as described in Table 2. Then the output bit rates are set as the target bit rates for the views of depth video. In the experiment, the uncompressed color videos and reconstructed depth videos are used to render the views.

Table 1
Summary of simulation parameters for depth video coding.

Sequences	Resolution	View index	
		Reference	Rendered
GhostTownFly	1920 × 1088	5–9	6, 7, 8
PoznanHall2	1920 × 1088	6–7	6.25, 6.50, 6.75
PoznanStreet	1920 × 1088	4–5	4.25, 4.50, 4.75
UndoDancer	1920 × 1088	5–1	2, 3, 4
Balloons	1024 × 768	3–1	1.5, 2, 2.5
Kendo	1024 × 768	3–1	1.5, 2, 2.5
Newspaper	1024 × 768	4–2	2.5, 3, 3.5

Table 2
Summary of simulation parameters for depth video coding.

Parameter	Value
View number	2
Intra period	8
Reference number	1
GOP size	8
GOP number	20
Symbol mode	CABAC
Search mode	4(FastSearch)
QP	22-27-32-37-42
Search range	64

To make a fair comparison, two RC schemes are implemented, including the Wang's bargaining game scheme [25] (denoted as BG) and the proposed scheme. The original JMVC (denoted as FixedQP) is employed as the benchmark. RC scheme should guarantee a small gap between the target bit rate and the actual coded bit rate, and then achieve the average PSNR improvement. Meanwhile, the picture quality smoothness and buffer smoothness are also very important in evaluating the efficiency of a rate control algorithm [31]. To evaluate the performance of our proposed algorithm, bit rate accuracy, and R–D performance (in terms of BDPSNR gain and BDBR saving) are compared with the benchmarks. For the virtual view image quality assessment, average PSNR value between the reference and reconstructed virtual views is employed. In addition, the buffer status is also evaluated. Detailed results are discussed in the following section.

5.1. Accuracy of bit rate achievement

The accuracy of bit rate achievement is one of important criteria to evaluate the performance of rate control algorithms. In this section, we measure the mismatch error between the target BR and the actual output BR by using the following equation:

$$E = \frac{|R_t - R_o|}{R_t} \times 100\%, \tag{17}$$

where R_t and R_o are the target bit rate (BR) and actual output BR of depth video, respectively. The results of BR mismatch E (%) for the proposed overall RC algorithms are provided in Table 3. Compared to the BG scheme, the mismatch value of the proposed algorithm is increased. The average mismatch of the proposed algorithm is only 0.71%, which is acceptable in practical applications.

Table 3
Summary of R–D performance on 3D sequences.

Sequences	Index	Target rate (kbps)	FixedQP) rate (kbps)	FixedQP) PSNR (dB)	BG [25] Rate (kbps)	BG [25] PSNR (dB)	BG [25] Error (%)	BG [25] BDPSNR (dB)	BG [25] BDBR (%)	Proposed Rate (kbps)	Proposed PSNR (dB)	Proposed Error (%)	Proposed BDPSNR (dB)	Proposed BDBR (%)
GhostTownFly	R1	2003	2003	48.22	2000	48.10	0.14	−0.19	7.61	1992	48.20	0.53	−0.06	2.15
	R2	1583	1583	47.58	1581	47.36	0.13			1564	47.48	1.18		
	R3	1125	1125	46.70	1123	46.52	0.21			1122	46.61	0.29		
	R4	908	908	46.14	911	45.93	0.29			910	46.08	0.22		
	R5	670	671	45.24	678	45.18	1.12			688	45.37	2.68		
PoznanHall2	R1	1042	1042	52.51	1043	52.36	0.11	−0.35	12.22	1043	52.32	0.14	−0.20	5.09
	R2	536	535	50.02	538	49.58	0.46			540	49.68	0.71		
	R3	235	235	47.87	236	47.52	0.59			236	47.59	0.33		
	R4	124	124	46.00	125	45.73	0.79			125	46.10	1.15		
	R5	77	77	43.89	78	43.80	0.71			78	44.01	1.27		
PoznanStreet	R1	1770	1759	48.93	1771	48.91	0.07	−0.15	6.18	1779	49.11	0.53	0.01	−0.32
	R2	850	846	47.54	849	47.53	0.13			858	47.64	0.97		
	R3	430	427	46.29	429	46.11	0.29			433	46.28	0.78		
	R4	235	230	44.88	235	44.70	0.13			236	44.90	0.28		
	R5	135	131	43.39	132	42.85	2.30			137	43.22	1.33		
UndoDancer	R1	1080	1076	44.27	1070	43.91	0.89	−0.30	6.54	1072	44.39	0.79	0.05	−0.97
	R2	650	640	41.85	640	41.33	1.50			641	41.86	1.43		
	R3	410	408	39.53	405	39.20	1.26			405	39.27	1.33		
	R4	275	275	37.59	272	37.49	0.92			272	37.70	1.19		
	R5	190	189	36.07	188	36.00	0.83			188	36.19	0.88		
Balloons	R1	1700	1692	49.86	1698	49.73	0.12	−0.20	7.37	1697	49.72	0.15	−0.03	1.27
	R2	950	915	48.03	948	47.92	0.21			949	47.98	0.06		
	R3	490	489	46.33	489	46.08	0.20			490	46.30	0.02		
	R4	260	262	44.75	260	44.54	0.00			260	44.80	0.08		
	R5	140	139	42.93	140	42.87	0.00			140	43.09	0.24		
Kendo	R1	1680	1670	53.35	1680	53.37	0.00	−0.18	5.09	1676	53.27	0.24	−0.31	8.70
	R2	1000	990	51.25	1002	51.22	0.20			998	51.12	0.21		
	R3	580	577	49.30	582	49.24	0.34			579	49.10	0.24		
	R4	330	333	47.50	331	47.20	0.30			330	47.03	0.02		
	R5	170	167	45.28	170	45.00	0.00			169	44.82	0.47		
Newspaper	R1	1810	1806	45.12	1816	44.94	0.33	−0.16	6.76	1808	45.43	0.11	0.23	−8.46
	R2	970	973	43.32	973	43.14	0.31			963	43.70	0.69		
	R3	520	515	41.82	522	41.75	0.38			514	41.98	1.08		
	R4	290	286	40.45	292	40.34	0.69			287	40.57	0.95		
	R5	160	158	38.83	161	38.78	0.63			163	38.87	2.15		
Average							**0.47**	**−0.22**	**7.40**			**0.71**	**−0.04**	**1.06**

5.2. *R–D performance*

The R–D performances in terms of the output BR of depth video and average PSNR values of virtual view are provided in Table 3. For each sequence, five different target BRs are tested. Fig. 6 shows the R–D performance comparison among the three schemes for the seven 3D sequences. The *y*-axis is the average PSNR value of virtual views, and the *x*-axis is the total bit rates of the output depth video stream. From the experimental results, we can observe that the proposed algorithm can obtain similar R–D performance compared to the FixedQP scheme. For the sequence “Newspaper”, the proposed algorithm superiors to the FixedQP scheme. This is because the model parameters of virtual view distortion model are varied dramatically among the frames, thus the proposed algorithm can achieve a better bit allocation strategy. For most cases, the proposed algorithm superiors to the BG scheme, except for the 3D sequences “Kendo”. It may be caused by the inaccuracy of depth information.

To measure the overall improvement of the proposed scheme, the Binotegaard Delta Bit Rate (BDBR) and Biontegaard Delta PSNR (BDPSNR) [32] are employed in the experiments for fair comparison. The FixedQP is set as the benchmark, then the performances of BG and proposed algorithm are compared with algorithm FixedQP in terms of BDPSNR and BDBR. Positive BDPSNR indicates quality improvement and negative BDBR indicates the bit rate reduction compared to the FixedQP scheme. The results are summarized in Tables 3. We observe that the BG scheme is inferior to the FixedQP for all the sequences. This is because the RC algorithm often makes sacrifices on the R–D performance to achieve the buffer smoothness under the bandwidth constraints. The average BDPSNR loss is 0.22 dB. For the proposed scheme, the average BDPSNR loss is only 0.04 dB compare to the FixedQP

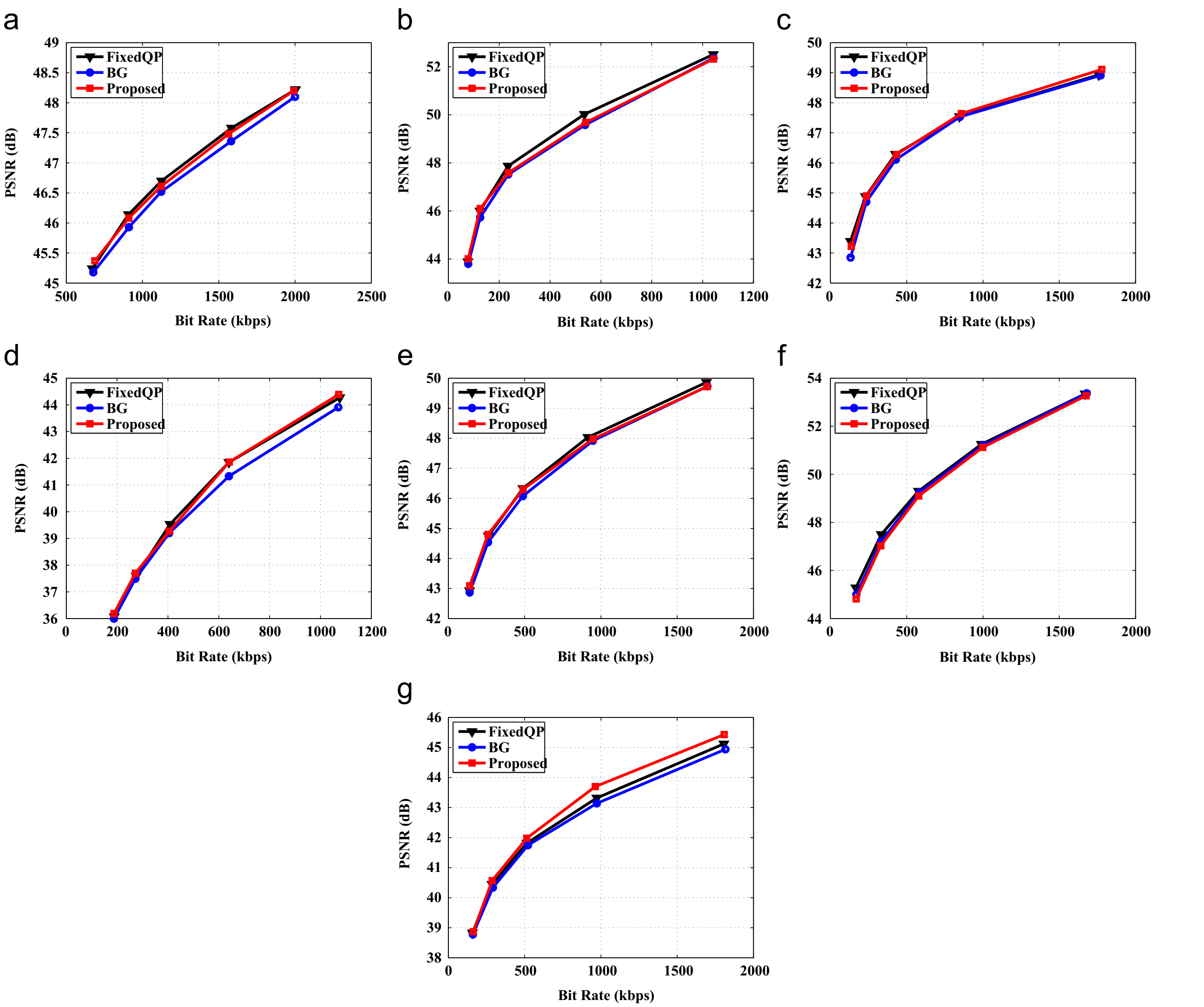

Fig. 6. R–D curves. (a) “GhostTownFly”. (b) “PoznanHall2”. (c) “Poznanstreet”. (d) “Undodancer”. (e) “Balloons”. (f) “Kendo”. (g) “Newspaper”.

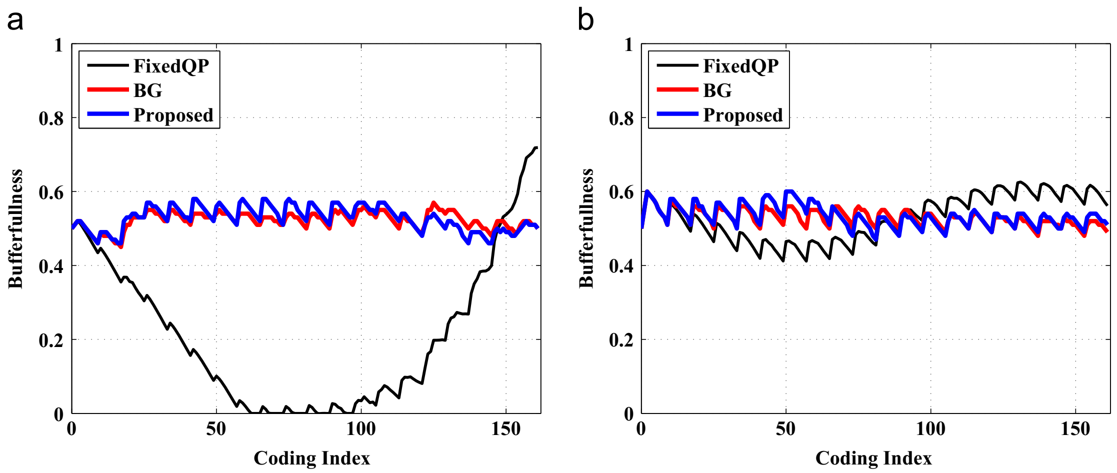

Fig. 7. Buffer regulation performance comparison. (a) "GhostTownFly" (View index: 5, Target bit rate: 508 kbps). (b) "Balloons" (View index: 3, Target bit rate: 70 kbps).

scheme. In other words, the proposed algorithm can improve the image quality of virtual views with the same amount of bit rates compared to the BG scheme. The average BDPSNR gain is 0.18 dB. In other words, our proposed algorithm achieves a significant gain.

5.3. Buffer regulation

In the experiments, the buffer delay is set as 2.56 s. The initial buffer fullness and target buffer fullness are also set as 0.5 for all the views. The lower bound and upper bound of buffer are set as 0.1 and 0.9, respectively. Fig. 7 shows the buffer regulation performance comparison among the FixedQP scheme, BG scheme and the proposed algorithm for two selected sequences. The y-axis is the buffer status of the RC schemes, and the x-axis is the coding index of frames in the depth video. Form Fig. 7, it could be observed that the buffer status of the proposed overall algorithm is more stable than that of FixedQP algorithm. The buffer status of the proposed overall algorithm can converge to the target buffer fullness as the number of encoded frames increases, which indicates the effectiveness of the proposed algorithm.

6. Conclusion

In this paper, we proposed a frame level rate control algorithm for MDVC to optimize the overall virtual view quality for 3D video. For depth video, the VSD model is proposed to indicate the importance of frames in the same view. Then, the proposed VSD model is incorporated in the bargain game theoretic based RC model to achieve the balance between virtual view quality and buffer constraint. Experimental results show that the proposed RC algorithm can achieve 0.18 dB BDPSNR gain in average compared to existing RC scheme.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under Grants 61272289, 61102088, City University of Hong Kong Applied Research Grant (ARG) 9667094, Shenzhen Emerging Industries of the Strategic Basic Research Project under Grant JCYJ20120617151719115, the Guangdong Nature Science Foundation under Grant S2012010008457, and the open research fund of Key Laboratory of Wireless Senor Network & Communication under Grant 2013002, and was supported in part by a grant from the Shenzhen Research Institute, City University of Hong Kong.

References

[1] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, C. Von Kopylow, A survey of 3DTV displays: techniques and technologies, IEEE Trans. Circuits Syst. Video Technol. 17 (11) (2007) 1647–1657.
[2] Y. Gao, M. Wang, Z. Zha, Q. Tian, Q. Dai, N. Zhang, Less is more: efficient 3D object retrieval with query view selection, IEEE Trans. Multimed. 11 (5) (2011) 1007–1018.
[3] L. Onural, Signal processing and 3DTV, IEEE Signal Process. Mag. 27 (5) (2010) 141–144.
[4] Y. Gao, M. Wang, D. Tao, R. Ji, Q. Dai, 3D object retrieval and recognition with hypergraph analysis, IEEE Trans. Image Process. 21 (9) (2012) 4290–4303.
[5] Q. Liu, Y. Yang, R. Ji, Y. Gao, L. Yu, Cross-view down/up-sampling method for multiview depth video coding, IEEE Signal Process. Lett. 19 (5) (2012) 295–298.
[6] Q. Liu, H. Xiao, Semi-global depth estimation algorithm for mobile 3-D video applications, Tsinghua Sci. Technol. 17 (2) (2012) 128–135.
[7] A. Gelman, P.L. Dragotti, V. Velisavljevic, Multiview image coding using depth layers and an optimized bit allocation, IEEE Trans. Image Process. 21 (9) (2012) 4092–4105.
[8] Y. Gao, J. Tang, R. Hong, S. Yan, Q. Dai, N. Zhang, T.-S. Chua, Camera constraint-free view-based 3D object retrieval, IEEE Trans. Image Process. 21 (4) (2012) 2269–2281.
[9] K. Muller, P. Merkle, T. Wiegand, 3-D video representation using depth maps, Proc. IEEE 99 (4) (2011) 643–656.
[10] N. Plath, S. Knorr, L. Goldmann, T. Sikora, Adaptive image warping for hole prevention in 3D view synthesis, IEEE Trans. Image Process. 22 (9) (2013) 3420–3432.

[11] C.G. Gurler, B. Gorkemli, G. Saygili, A.M. Tekalp, Flexible transport of 3-D video over networks, Proc. IEEE 99 (4) (2011) 694–707.
[12] A. Vetro, T. Wiegand, G.J. Sullivan, Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard, Proc. IEEE 99 (4) (2011) 626–642.
[13] Y. Zhang, S. Kwong, L. Xu, S. Hu, G. Jiang, C.-C.J. Kuo, Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model, IEEE Trans. Image Process. 22 (9) (2013) 3497–3512.
[14] Y. Liu, Q. Huang, S. Ma, D. Zhao, W. Gao, Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model, Signal Process.: Image Commun. 24 (8) (2009) 666–681.
[15] H. Yuan, Y. Chang, J. Huo, F. Yang, Z. Lu, Model-based joint bit allocation between texture videos and depth maps for 3-D video coding, IEEE Trans. Circuits Syst. Video Technol. 21 (4) (2011) 485–497.
[16] S. Hu, S. Kwong, Y. Zhang, C.-C.J. Kuo, Rate-distortion optimized rate control for depth map-based 3-D video coding, IEEE Trans. Image Process. 22 (2) (2013) 585–594.
[17] Y. Liu, Q. Huang, S. Ma, D. Zhao, W. Gao, S. Ci, H. Tang, A novel rate control technique for multiview video plus depth based 3d video coding, IEEE Trans. Broadcast. 57 (2) (2011) 562–571.
[18] F. Shao, G. Jiang, W. Lin, M. Yu, Q. Dai, Joint bit allocation and rate control for coding multi-view video plus depth based 3d video, IEEE Trans. Multimed. 15 (8) (2013) 1843–1854.
[19] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, Doc. JVT-G012r1, Adaptive Basic Unit Layer Rate Control for JVT. Pattaya, Thailand, March 2003.
[20] G. Cheung, V. Velisavljevic, A. Ortega, On dependent bit allocation for multiview image coding with depth-image-based rendering, IEEE Trans. Image Process. 20 (11) (2011) 3179–3194.
[21] Q. Wang, X. Ji, Q. Dai, N. Zhang, Free viewpoint video coding with rate-distortion analysis, IEEE Trans. Circuits Syst. Video Technol. 22 (6) (2012) 875–889.
[22] H. Yuan, S. Kwong, J. Liu, J. Sun, A novel distortion model and lagrangian multiplier for depth maps coding, IEEE Trans. Circuits Syst. Video Technol. 24 (3) (2014) 443–451.
[23] S. Hu, H. Wang, S. Kwong, T. Zhao, C.-C.J. Kuo, Rate control optimization for temporal-layer scalable video coding, IEEE Trans. Circuits Syst. Video Technol. 21 (8) (2011) 1152–1162.
[24] H. Wang, S. Kwong, C.-W. Kok, Efficient prediction algorithm of integer DCT coefficients for H.264/AVC optimization, IEEE Trans. Circuits Syst. Video Technol. 16 (4) (2006) 547–552.
[25] X. Wang, S. Kwong, Y. Zhang, Applying game theory to rate control optimization for hierarchical b-pictures, IEEE Trans. Broadcast. 59 (4) (2013) 591–601.
[26] L. Xu, W. Gao, X. Ji, D. Zhao, Rate control for hierarchical B-picture coding with scaling-factors, in: Proceedings of IEEE ISCS, 2007, pp. 49–52.
[27] A. Leontaris, A.M. Tourapis, Rate Control for the Joint Scalable Video Model (JSVM), document JVT-W043, Video Team of ISO/IEC MPEG and ITU-T VCEG, San Jose, CA, April 2007.
[28] Joint Multiview Video Coding (JMVC) 8.5 Software Package, CVS Server for JMVC Software, Boston, MA, USA, 2011.
[29] C. Bartnik, S. Bosse, H. Brust, HEVC Extension for Multiview Video Coding and Multiview Video Plus Depth Coding, Doc. VCEG-AR13, 44th Video Coding Experts Group (VCEG) meeting, San Jose, CA, USA, February 2012.
[30] HEVC 3D extension software, 〈https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/trunk/〉.
[31] L. Xu, D. Zhao, X. Ji, L. Deng, S. Kwong, W. Gao, Window-level rate control for smooth picture quality and smooth buffer occupancy, IEEE Trans. Image Process. 20 (3) (2011) 723–734.
[32] G. Bjøntegaard, Calculation of Average PSNR differences Between RD-Curves, Doc. VCEG-M33, Austin, USA, April 2001.