Rate-Distortion Optimized Rate Control for Depth Map-Based 3-D Video Coding

Sudeng Hu, Sam Kwong, *Senior Member, IEEE*, Yun Zhang, and C.-C. Jay Kuo, *Fellow, IEEE*

***Abstract*— In this paper, a novel rate control scheme with optimized bits allocation for the 3-D video coding is proposed. First, we investigate the R-D characteristics of the texture and depth map of the coded view, as well as the quality dependency between the virtual view and the coded view. Second, an optimal bit allocation scheme is developed to allocate target bits for both the texture and depth maps of different views. Meanwhile, a simplified model parameter estimation scheme is adopted to speed up the coding process. Finally, the experimental results on various 3-D video sequences demonstrate that the proposed algorithm achieves excellent R-D efficiency and bit rate accuracy compared to benchmark algorithms.**

***Index Terms*— 3-D video coding, depth map, rate control, view synthesis.**

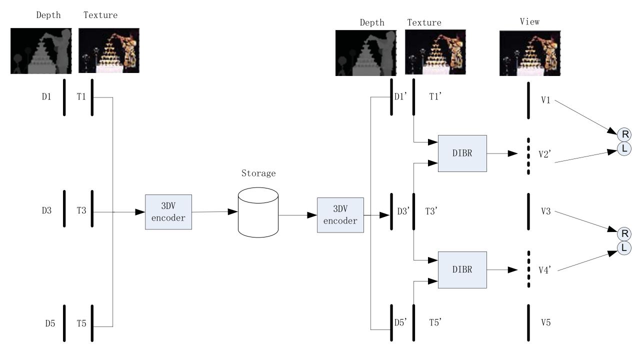


Fig. 1. Overview of 3-D video coding system. *m* views are encoded at the sender side and decoded at the receiver side. The virtual views are synthesized with DIBR by referencing the decoded views at the receiver side.

I. Introduction

THREE-DIMENSIONAL VIDEO (3DV) has gained increasing interests recently. The typical 3DV is stereo-view video which provides each eye with one video separately at the same time. The small differences between these two videos cause the illusion of depth perception for human. In addition to the stereoscopic 3D video, the emerging autostereoscopic display [1]–[4] which emits a number of views enable autostereoscopic 3D video. Comparing with stereoscopic viewing, it involves a more general case of *n*-view multiview video. In this scenario, the viewpoint can be interactively changed by selecting different stereo pairs of views from *n*-view.

Delivering or storing *n*-view video requires tremendous bits that beyonds current transmission or storage capacity. Multiview Video Coding (MVC) [5] is developed to encode the multiview videos, where both the temporal redundancy within each view and inter-view redundancy among the neighbouring views are exploited [6]. Although the MVC encoder performs excellent coding efficiency, it is still not efficient enough to store or to transmit large numbers of views. Multiview plus depth format (MVD) [7], [8] presents a promising solution for the efficient delivery of 3DV. As shown in Fig. 1, only a subset *m* of *n* views are coded and transmitted, along with additional supplementary information such as per-pixel depth map which provides scene geometry information. At receiver side, these *m* coded views provide references for generating the rest views, which are synthesized as the virtual views via Depth-Image-based-Rendering (DIBR) [9], [10]. MVD reduces the number of the views to be transmitted but it can still reconstruct all the required views at the receiver side.

Rate Control (RC) is employed in video coding to regulate the bit rate meanwhile guarantee good video quality. As for 3DV, it becomes more complicated because multiple views are involved in coding and within each view there are two kinds of video sequences (*i.e.* texture and depth map). One of challenge problems is the bit allocation between the texture and depth map. Since the quality of virtual view is affected by the quality of both the texture and depth map, the bits should be allocated to balance their quality. In [11], bits are allocated to minimize the total distortion of the texture and depth map. However since the depth map is not presented for viewing, the minimum total distortion does not guarantee the optimal quality in the virtual views. In [12] the optimal bit allocation between the texture map and depth map is exhaustively searched by a hierarchical search method. In [13], the distortion of the virtual view is modeled and the optimal bit allocation between the texture and depth map is searched based on this distortion model. In [14], a similar distortion model is derived for virtual view and to achieve optimal virtual view quality, bits are allocated between the texture and depth map based on the derived distortion model. In [15], a joint

Manuscript received March 1, 2012; revised September 3, 2012; accepted September 4, 2012. Date of publication September 20, 2012; date of current version January 10, 2013. This work was supported by the Hong Kong Research Grants Council General Research Fund, under Project 9041495 (CityU 115109) and the National Nature Science Foundation of China under Grant 61102088. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. James E. Fowler.

S. Hu and C.-C. J. Kuo are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: sudenghu@usc.edu; cckuo@sipi.usc.edu).

S. Kwong and Y. Zhang are with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (e-mail: cssamk@cityu.edu.hk; yunzhang@cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2012.2219549

RC scheme is proposed where inter-view bit allocation are performed according to the sequence complexity of each view, while the bits are allocated at fixed ratio between texture and depth map within each view. However in these algorithms, inter-view bit allocation is rarely considered or only simply allocated according to the sequence complexity. In 3DV, more general case involves m views coding, thus bits allocation among different views is highly desired.

In this paper, the RC algorithm is proposed aiming at improving the overall quality in 3DV, where both the qualities of the coded views and the virtual views are considered. This is more reasonable, since both the virtual view and the coded view would be presented for viewing at the receiver side. On the other hand, the virtual view is synthesized by referencing nearby coded views, thus its quality depends on the coded references' quality. As shown in Fig. 1, different coded views are referenced by different number of virtual views. Intuitively, the coded view with more dependants should have better quality, as it would benefit more virtual views. In order to achieve the optimal R-D performance in 3DV, we first investigate the R-D characteristics of the texture and depth map. Then the quality dependency between the virtual view and the coded view is studied in the texture and depth map respectively. Based on the R-D characteristics of both the coded view and the virtual view, a bit allocation scheme is proposed for both the texture and depth map of all coded views. In this paper, a simple case of multiview 3DV is discussed, where only three views are coded and two views are synthesized, but the bit allocation scheme and the conclusions derived in this paper can be easily extend to n views cases.

The rest of this paper is organized as follows. In Section II, the R-D characteristics of both the coded view and the virtual view are investigated. In Section III, a RC scheme is proposed with optimal bit allocation. In Section IV, the experimental results are given to demonstrate the efficiency of the proposed RC algorithm. Finally, Section V concludes this paper.

II. Rate and Quality Analysis in 3-DV

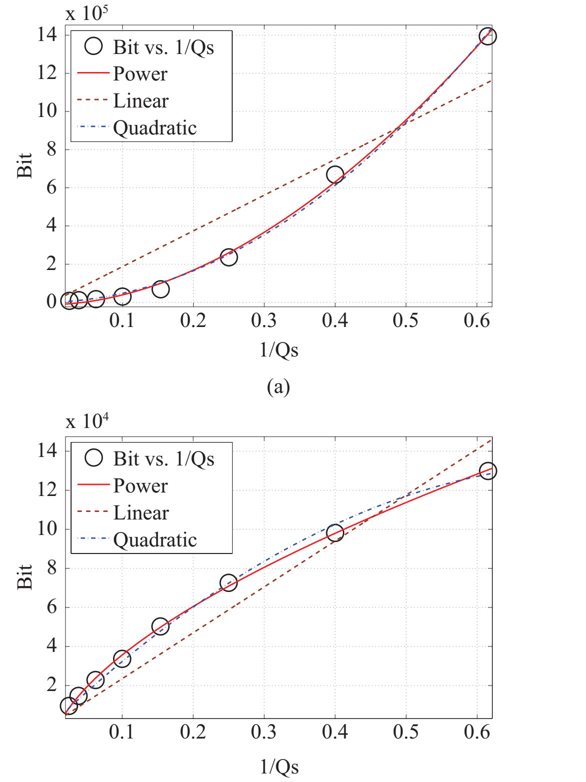
A. R-D Model for Coded Views

The tradeoff between the output bit rate (R) and the quality (D) of compressed video is determined by quantization step size (Q_s), which is indexed by quantization parameter (Q). The R-Q_s and D-Q_s model have been studied extensively for previous video coding standards such as MPEG-2 and H.264/AVC.

For the R-Q_s model, the classic quadratic model is developed in [16], [17] and a linear model is widely studied and applied for its simple form [18]–[20]. In 3DV, the R-D characteristics is different from that in previous coding standards. First, the depth map is a grey image, which has no chrominance (UV) components for YUV color space. Second, the inter-view prediction is employed to reduce the redundancy among the different views. We employ the power R-Q_s model [21], [22] for the depth and texture map as

$$R = \rho Q_s^{\tau} + c \quad (1)$$

where model parameter ρ and τ depend on the video content and the sequence types (*i.e.* texture or depth map); c represents



(a)

(b)

Fig. 2. R-Q relationship in the texture and depth map for the sequence *Kendo*. (a) Texture map. (b) Depth map.

the bit to code the header information. At high bit rate, header bits usually take a small part of the total output bits, therefore we simplify (1) by ignoring the header bits as

$$R \approx \rho Q_s^{\tau} \quad (2)$$

In Fig. 2, video sequence Kendo is coded with Q from 8 to 36. We can see that both the power model and the quadratic model fit the actual data well. For its simple form, the power model is adopted in our work. In Fig. 2, the texture and depth map exhibit different R-D characteristics. For example, parameter τ is quite different in the texture map and depth map.

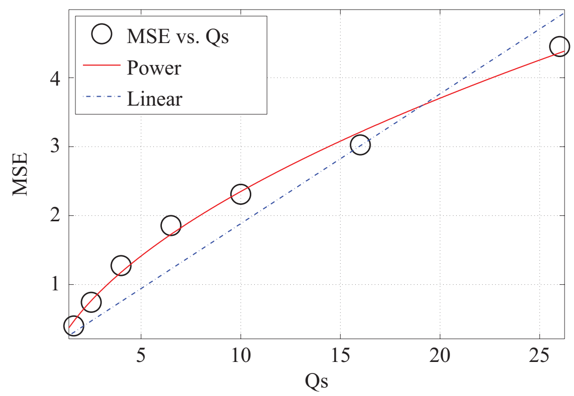
In H.264/AVC and its MVC extension, Q_s and Q have non-linear relationship, *i.e.*, Q_s double in size for every increment of 6 in Q [23]. This relationship can be approximated as

$$Q_s \approx e^{c_1 Q + c_2} \quad (3)$$

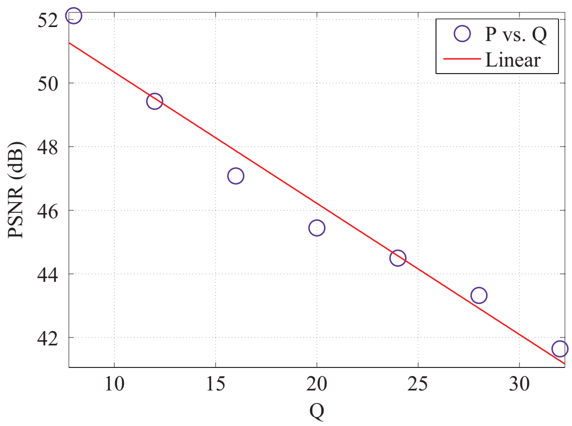
where c_1 and c_2 are constants that $c_1 = \frac{1}{6}ln2$ and $c_2 = -\frac{2}{3}ln2$. Therefore R-Q relationship can be derived by substituting (3) into (2) as

$$R = \rho \cdot e^{\tau(c_1 Q + c_2)} \quad (4)$$

As for the D-Q model, we investigate the relationship between Q and the quality of texture map. Since the depth map will not be presented for viewing, Q of depth map will have no direct effect on view quality, but it will have indirect influence on the quality of virtual view. Such influence will



(a)



(b)

Fig. 3. *D*-*Q* relationship in the texture map for the sequence *Balloons*. (a) MSE-Q_s relationship. (b) $PSNR$-Q relationship.

be discussed in Section II-B. For the texture map, the *D*-*Q* model is adopted as

$$MSE = \chi {Q_s}^{\varphi} \tag{5}$$

where MSE is mean square error indicating the quality of reconstructed pictures; χ and φ are model parameters related to the video content. The MSE-Q_s relationship is illustrated in Fig. 3(a), where Q varies from 6 to 32. We can see that the actual relationship can be precisely depicted by (5). MSE and peak signal noise ratio (PSNR) have following relationship

$$P = 10log_{10}(\frac{255^2}{MSE}) \tag{6}$$

where P refers to PSNR. By substituting (3) and (6), we obtain the P-Q relation as

$$P = \alpha Q + \eta \tag{7}$$

where $\alpha = -\frac{10}{ln10}c_1\varphi$ and $\eta = \frac{10}{ln10}(ln\frac{255^2}{\chi} - \varphi c_2)$. The P-Q relationship in (7) is illustrated in Fig. 3(b). As we can see, the PSNR decreases almost linearly with increase of Q value.

B. Quality Analysis for Virtual Views

At the receiver side, the virtual views can be synthesized from nearby coded views with DIBR as shown in Fig. 1. In this paper, we investigate in the multiview video captured by parallel camera array with small intervals. When DIBR is applied at receiver side, distortion will be introduced due to the compression error in the texture and depth map. In [24], analysis on the effect of geometry distortions caused by depth coding artifacts is presented. In [25], the bound of synthesis error is derived for various configurations such as depth errors. Even without compression, the distortion would be introduced by the DIBR tools. Various DIBR algorithms have been proposed to reduce the synthesis error [26], [27], however it cannot be avoided. In this paper, we are only interested in issues on compression that causes distortion. We directly investigate the relationship between the quality of virtual view and Q of texture map (Q^T) or depth map (Q^D).

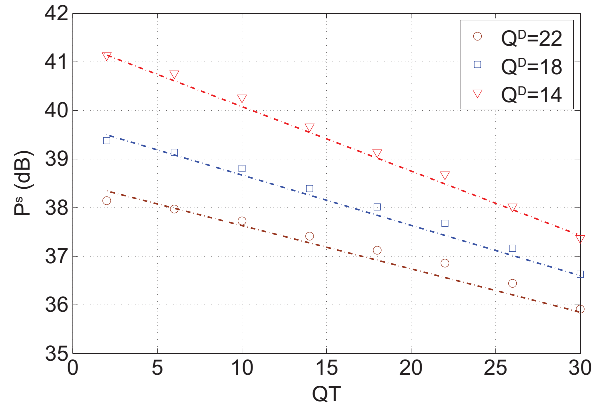


Fig. 4. Linear relationship between the quality of virtual view and Q^T on the sequence *Champagne_tower*. View 37 is coded while view 38 is synthesized with DIBR. Q^T is changed from 2 to 30 when Q^D is fixed at 14, 18, and 22, respectively.

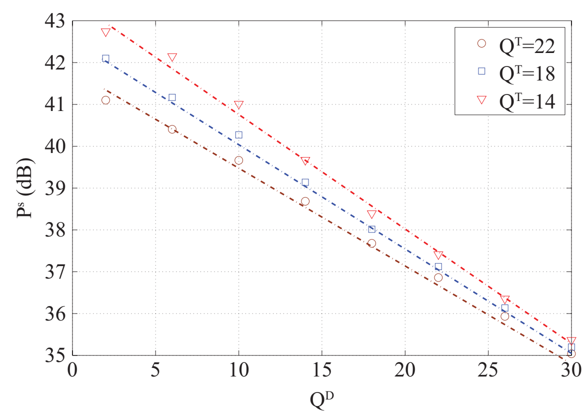
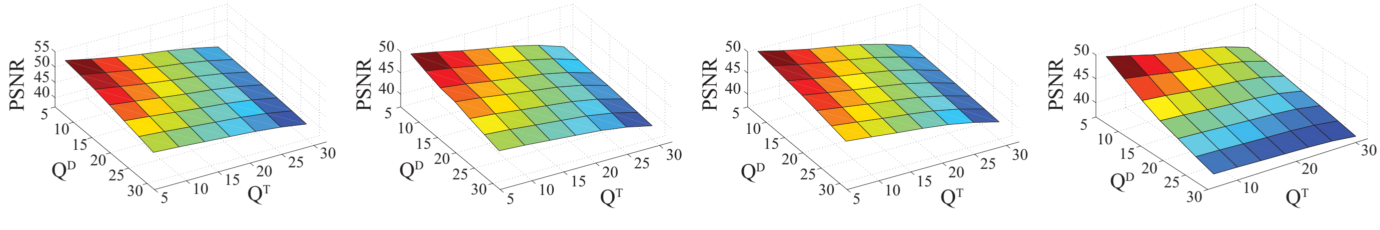


Fig. 5. Linear relationship between the quality of virtual view and Q^D on the sequence *Champagne_tower*. View 37 is coded while view 38 is synthesized with DIBR. Q^D is changed from 2 to 30 when Q^T is fixed at 14, 18, and 22, respectively.

Since the virtual view is projected from pixel value in the texture map, its quality will be affected by the quality of decoded texture map. In Fig. 4, the quality influence of texture map on virtual view is investigated by changing Q^T from 2 to 30 meanwhile fixing Q^D at 14, 18 and 22 respectively. The quality of virtual view (P^s) is measured in term of PSNR. As shown in Fig. 4, once Q^D is determined, the P^s-Q^T relationship can be approximated as linear.



(a) (b) (c) (d)

Fig. 6. Joint relationship between quality of synthesized view and Q^T and Q^D. (a) *Pantomime.* (b) *Balloons.* (c) *Kendo.* (d) *Champagne_tower.*

Similarly in Fig. 5, Q^D is changed from 2 to 30, while Q^T is fixed at 14, 18, 22 respectively. We can see that linearity also can be observed between P^s and Q^D. Therefore we have the P^s-Q^T relationship as

$$\frac{\partial P^s(Q^T, Q^D)}{\partial Q^T} = \beta(Q^D) \tag{8}$$

and the P^s-Q^D relationship as

$$\frac{\partial P^s(Q^T, Q^D)}{\partial Q^D} = \gamma(Q^T). \tag{9}$$

Moreover, we can observe that the values of $\beta(Q^D)$ or $\gamma(Q^T)$ change slowly with Q^D or Q^T. Table I shows the slopes of linear P^s-Q^T relation and linear P^s-Q^D relation in Fig. 4 and Fig. 5, when the corresponding Q^D or Q^T are fixed at 14, 18, 22. In Table I, the derivatives of $\beta(Q^D)$ and $\gamma(Q^T)$ ($\Delta\beta(Q^D)/\Delta Q^D$ and $\Delta\gamma(Q^T)/\Delta Q^T$), which indicate the change rate with Q^D or Q^T, are very small. Therefore for simplicity, we approximate $\beta(Q^D)$ and $\gamma(Q^T)$ as constant and approximate (8) as

$$\frac{\partial P^s(Q^T, Q^D)}{\partial Q^T} = \beta \tag{10}$$

and (9) as

$$\frac{\partial P^s(Q^T, Q^D)}{\partial Q^D} = \gamma \tag{11}$$

where β and γ are considered as constants and their values depend on the video content.

In Fig. 6, the joint P^s-Q^T-Q^D relationship is illustrated by carrying out extensive experiments, where depth and texture maps are coded with Q from 6 to 30 respectively. The relation in Fig. 6 can be approximated as 2D plane which indicates linear and decoupled relation between the P^s-Q^T and P^s-Q^D. For the completely decoupled linear relations, ideally the derivatives of $\beta(Q^D)$ and $\gamma(Q^T)$ should be 0. As shown in Table I, the actual derivatives are very close to 0, which indicates the approximation is close to the ideal cases, thus the caused approximation error is neglected in the rest of the paper.

III. Rate Control Scheme

A. RDO Bit Allocation at Sequence Level

The reference relation between the coded views and the virtual views is illustrated in Fig. 7, where V1, V3 and V5 are coded views, and V2 and V4 are virtual views synthesized with the texture and depth maps of V1, V3 and V3, V5 respectively. Since both the coded views and virtual views will be presented for human, their qualities are equally important. Due to the limitation of the transmission capacity or the storage space, the problem is how to allocate bit reasonably to optimize the overall quality performance. For convenience, in the rest of this paper the superscripts T and D are used to indicate the texture map and depth map; the subscripts n, m and i, j stand for the view index. The optimization problem is formulated as

$$Max\left(\sum_{n\in C} P_n(Q_n^T) + \sum_{m\in S} P_m(\vec{Q}^T, \vec{Q}^D)\right) \tag{12}$$

where C is the set of the coded view index, e.g. $C = \{1, 3, 5\}$; S is the set of virtual view index, e.g. $S = \{2, 4\}$; $P_n(Q_n^T)$ and $P_m(\vec{Q}^T, \vec{Q}^D)$ are the quality of nth view and mth view. Since the quality of coded view is determined by the corresponding Q^T, P_n is the function of Q_n^T. For virtual view, its quality is determined by both the Q^T and Q^D of the nearby coded views. For example, P_2 is determined by Q_1^T, Q_3^T and Q_1^D, Q_3^D as illustrated in Fig. 7. In general, P_m is function of $\vec{Q}^T$ and $\vec{Q}^D$, where $\vec{Q}^T = [Q_1^T, Q_3^T, \ldots]$ and $\vec{Q}^D = [Q_1^D, Q_3^D, \ldots]$.

TABLE I

Slope of P^s-Q^T Relation and P^s-Q^D Relation

		Q = 14	Q = 18	Q = 22	Der
Champagne	γ	−0.3076	−0.2853	−0.2636	0.0055
	β	−0.1214	−0.1088	−0.0942	0.0034
Kendo	γ	−0.0835	−0.0756	−0.0694	0.0018
	β	−0.2789	−0.2604	−0.2557	0.0029
Pantomime	γ	−0.1215	−0.1103	−0.0977	0.0030
	β	−0.3056	−0.2762	−0.2533	0.0065
Balloons	γ	−0.1293	−0.1215	−0.1095	0.0025
	β	−0.2319	−0.2180	−0.1847	0.0059

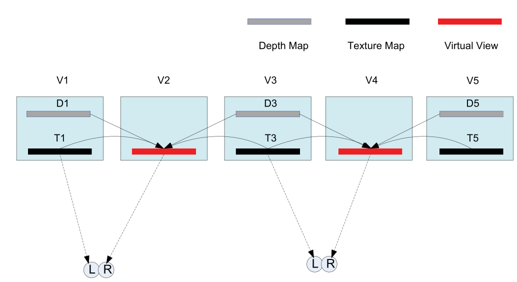


Fig. 7. Illustration of the reference relationship of the virtual view and the coded view.

The optimization problem is under the constraint that

$$\sum_{n\in C}\left(R_n^D(Q_n^T)+R_n^D(Q_n^D)\right) < R_{tot} \tag{13}$$

where R_n^T and R_n^D are the bits to code the nth texture and depth map respectively; R_{tot} is total target bits. It is obvious that the optimization problem is generally convex, since P of the coded view and virtual view have linear relationship with Q^T and Q^D as discussed in section II. By applying the method of Lagrangian multiplier, we have

$$J=\left(\sum_{n\in C}P_n(Q_n^T)+\sum_{m\in S}P_m(\vec{Q^T},\vec{Q^D})\right) +\lambda\left(\sum_{n\in C}\left(R_n^T(Q_n^T)+R_n^D(Q_n^D)\right)-R_{tot}\right) \tag{14}$$

where λ is Lagrangian multiplier. The Q values of texture and depth map have different effects on the total quality. Q^D only affects the quality of virtual view while Q^T influences the quality of both coded views and virtual views. Therefore depending on the type of Q value (*i.e.*, Q^T or Q^D), we have different differential equations for (14). According to (7) and (10), the partial derivative of the first term of right side of (14) with respect to Q^T is derived as

$$\frac{\partial\left(\sum_{n\in C}P_n+\sum_{m\in S}P_m\right)}{\partial Q_i^T}=\alpha_i+\textstyle\sum_{j\in K_i}\beta_{ij} \tag{15}$$

where α_i is the slope of linear P_i-Q_i^T relation in (7); β_{ij} is the slope of linear P_j-Q_i^T relation in (10); K_i is the set of virtual views whose qualities depend on ith coded view. For example in Fig. 7, $K_1=\{2\}$ since the 1st view only affects the 2nd virtual view, while $K_3=\{2,4\}$ since the 3rd view affects the 2nd and 4th virtual views. Therefore Q^T in different positions have different effects on the total quality, and thus the right side of (15) varies for different views.

Meanwhile, by taking the partial derivative of the second term of right side of (14) with respective to Q^T, and combining with (4) we obtain

$$\frac{\partial R_i^T}{\partial Q_i^T}=\tau_i\cdot c_1\cdot R_i^T \tag{16}$$

where R_i^T is bits for texture map. Finally, for the optimal solution, by differentiating (14) on both sides and replacing with (15) and (16), we get

$$0=k_i^T+\lambda\cdot\tau_i^T\cdot c_1\cdot R_i^{T*} \tag{17}$$

where R_i^{T*} is optimal bit for the texture map of ith view; k_i^T is a parameter of the texture map related to the view position that

$$k_i^T=\alpha_i+\sum_{j\in K_i}\beta_{ij}. \tag{18}$$

Similarly the partial derivative of the first term of right side of (14) with respective to Q_n^D is derived from (11) as

$$\frac{\partial\left(\sum_{n\in C}P_n+\sum_{m\in S}P_m\right)}{\partial Q_i^D}=\textstyle\sum_{j\in K_i}\gamma_{ij} \tag{19}$$

where γ_{ij} is the slope of linear P_j-Q_i^D relation in (11). By taking the partial derivative of (14) with respect to Q_i^D and replacing with (19), we have

$$0=k_i^D+\lambda\cdot\tau_i^D\cdot c_1\cdot R_i^{D*} \tag{20}$$

where R_i^T is the optimal bits for ith texture map; k_i^D is model parameter of the depth map in ith view that

$$k_i^D=\sum_{j\in K_i}\gamma_{ij}. \tag{21}$$

Therefore from (17) and (20) we can get the optimal bit allocation for both depth or texture map of ith view as

$$R_i^*=\frac{k_i/\tau_i}{\sum_{n\in C}(k_n^T/\tau_n^T+k_n^D/\tau_n^D)}R_{tot} \tag{22}$$

where τ_i and k_i can be the parameters for either texture map or depth map. Therefore the bit allocation scheme in (22) can be applied both for the texture and depth map. In this way, optimal bit allocation among different views and among the texture and depth map are automatically achieved.

B. Model Parameter Estimation

In order to allocate bits according to (22), we have to access model parameter α_i, β_{ij}, γ_{ij} and τ_i before coding. Therefore the texture map is precoded at Q_A^T and Q_B^T and the depth map is precoded at Q_A^D and Q_B^D. Based on (7), (10) and (11), these parameters can be estimated as

$$\alpha_i=\frac{P_{iA}-P_{iB}}{Q_{iA}^T-Q_{iB}^T} \tag{23}$$

$$\beta_{ij}=\frac{\hat{P}_{jA}-\hat{P}_{jB}}{Q_{iA}^T-Q_{iB}^T} \tag{24}$$

$$\gamma_{ij}=\frac{\check{P}_{jA}-\check{P}_{jB}}{Q_{iA}^D-Q_{iB}^D} \tag{25}$$

where P_{iA} and P_{iB} are the PSNR of ith coded view precoded at Q_A^T and Q_B^T respectively; $\hat{P}_{jA}$ and $\hat{P}_{jB}$ are the PSNR of jth virtual view synthesized with the ith texture map precoded at Q_A^T and Q_B^T respectively; $\check{P}_{jA}$ and $\check{P}_{jB}$ are the PSNR of jth virtual view synthesized with the ith depth map precoded at Q_A^D and Q_B^D respectively.

In order to estimate these parameters, each view has to be precoded twice, which would involve heavy computation, especially when the view number is large. Usually the video contents of different views are highly similar. Thus we assume the R-D characteristics are similar in the same type of videos. To reduce the computational complexity, instead of precoding m views, only the texture and the depth map of the first view are precoded. Then the model parameters α, β, γ are

TABLE II

MODEL PARAMETER VALUE AND THE CORRESPONDING ESTIMATION

		Newspaper			Champagne_tower			Balloon			Mobile		
		Actual	Est	E(%)	Actual	Est.	E(%)	Actual	Est.	E(%)	Actual	Est.	E(%)
α	α_2	−0.515	−0.512	0.6	−0.406	−0.405	0.1	−0.413	−0.434	5.3	−0.739	−0.736	0.4
	α_3	−0.535	−0.512	4.3	−0.420	−0.405	3.5	−0.405	−0.434	7.2	−0.746	−0.736	1.3
β	β_{32}	−0.240	−0.242	0.8	−0.112	−0.104	7.0	−0.255	−0.258	1.4	−0.525	−0.514	2.1
	β_{34}	−0.250	−0.242	3.4	−0.098	−0.104	6.3	−0.252	−0.258	2.6	−0.519	−0.514	1.0
	β_{54}	−0.263	−0.242	8.0	−0.105	−0.104	1.1	−0.256	−0.258	0.9	−0.541	−0.514	5.0
γ	γ_{32}	−0.167	−0.151	9.5	−0.251	−0.256	2.1	−0.102	−0.107	5.4	−0.143	−0.155	8.5
	γ_{34}	−0.144	−0.151	4.9	−0.278	−0.256	7.8	−0.096	−0.107	11.5	−0.146	−0.155	5.6
	γ_{54}	−0.167	−0.151	9.7	−0.281	−0.256	9.0	−0.101	−0.107	6.9	−0.132	−0.155	17.1
τ^T	τ_2^T	−1.126	−1.073	4.7	−1.370	−1.314	4.1	−1.103	−1.045	5.3	−1.137	−1.093	3.9
	τ_3^T	−1.164	−1.073	7.8	−1.396	−1.314	5.9	−1.113	−1.045	6.2	−1.187	−1.093	7.9
τ^D	τ_2^D	−1.048	−1.010	3.6	−0.832	−0.922	10.8	−0.956	−0.993	3.8	−0.727	−0.721	0.9
	τ_3^D	−0.970	−1.010	4.1	−0.904	−0.922	1.9	−1.027	−0.993	3.3	−0.629	−0.721	14.6
ρ	ρ_2^T	19435	21041	8.3	50493	48792	3.4	23864	23637	1.0	6438	6618	2.8
	ρ_3^T	22032	21991	0.2	60480	57388	5.1	25290	25786	2.0	8040	7596	5.5
Average				**5.0**			**4.9**			**4.5**			**5.5**

TABLE III

RESULT SUMMARY OF DIFFERENT RC ALGORITHMS ON THE SEQUENCE *Balloons*

		Liu2011				Yuan2011					Liu2009					Proposed+CQ					Proposed+FL				
Target (Mbps)	**View**	**Rate** (kbps) **T**	**D**	**PSNR** (dB)	**E** (%)	**Rate** (kbps) **T**	**D**	**PSNR** (dB)	**E** (%)	**ΔP** (dB)	**Rate** (kbps) **T**	**D**	**PSNR** (dB)	**E** (%)	**ΔP** (dB)	**Rate** (kbps) **T**	**D**	**PSNR** (dB)	**E** (%)	**ΔP** (dB)	**Rate** (kbps) T	D	**PSNR** (dB)	**E** (%)	**ΔP** (dB)
	V1	832	211	41.13		718	265	41.31			642	412	40.79			725	129	41.36			759	127	41.16		
	V2			40.71				41.09					41.06					40.96					40.75		
3 (Mbps)	**V3**	811	203	41.05	**3.1**	692	244	41.24	**2.0**	**0.29**	619	374	40.74	**5.5**	**−0.03**	1006	166	42.51	**0.2**	**0.62**	1065	174	42.60	**0.3**	**0.36**
	V4			40.58				41.06					41.02					40.82					40.59		
	V5	828	210	40.82		712	309	41.05			635	484	40.54			865	103	41.74			759	125	40.96		
	V1	1126	281	42.27		908	353	42.30			913	472	42.33			1015	173	42.57			1013	167	42.46		
	V2			41.74				42.06					42.34					41.88					41.71		
4 (Mbps)	**V3**	1106	269	42.12	**5.1**	880	321	42.19	**5.6**	**0.16**	885	426	42.22	**4.0**	**0.29**	1311	232	43.33	**0.7**	**0.48**	1418	231	43.36	**0.2**	**0.43**
	V4			41.62				42.00					42.28					41.72					41.57		
	V5	1143	278	41.98		903	413	41.98			909	557	42.02			1149	148	42.63			1011	166	42.80		
	V1	1381	346	43.09		1164	450	43.09			1059	708	42.81			1178	226	43.16			1265	207	43.19		
	V2			42.49				42.81					43.04					42.56					41.62		
5 (Mbps)	**V3**	1344	336	42.89	**2.5**	1136	406	42.93	**2.9**	**0.18**	1029	633	42.67	**6.7**	**0.11**	1814	288	43.92	**1.1**	**0.21**	1778	288	43.90	**0.6**	**−0.03**
	V4			42.31				42.76					42.95					42.33					41.71		
	V5	1371	349	42.62		1169	530	42.71			1060	844	42.45			1366	182	42.45			1266	166	42.80		
Average:					**3.6**				**3.5**	**0.21**				**5.4**	**0.12**				**0.7**	**0.43**				**0.4**	**0.25**

TABLE IV

RESULT SUMMARY OF DIFFERENT RC ALGORITHMS ON THE SEQUENCE *Newspaper*

		Liu2011				Yuan2011					Liu2009					Proposed+CQ					Proposed+FL				
Target (Mbps)	**View**	**Rate** (kbps) **T**	**D**	**PSNR** (dB)	**E** (%)	**Rate** (kbps) **T**	**D**	**PSNR** (dB)	**E** (%)	**ΔP** (dB)	**Rate** (kbps) **T**	**D**	**PSNR** (dB)	**E** (%)	**ΔP** (dB)	**Rate** (kbps) **T**	**D**	**PSNR** (dB)	**E** (%)	**ΔP** (dB)	**Rate** (kbps) T	D	**PSNR** (dB)	**E** (%)	**ΔP** (dB)
	V2	885	227	39.96		577	377	39.31			670	329	39.93			720	215	40.22			700	201	40.01		
	V3			37.64				37.92					38.11					38.12					37.54		
3 (Mbps)	**V4**	798	209	39.56	**1.0**	540	400	38.95	**6.7**	**−0.22**	634	347	39.57	**2.1**	**0.26**	831	342	40.59	**2.3**	**0.50**	918	281	40.91	**0.1**	**0.47**
	V5			38.46				38.67					38.99					38.96					39.01		
	V6	722	191	38.96		550	354	38.64			648	310	39.29			623	201	39.16			702	201	39.45		
	V2	1152	304	40.90		741	538	40.37			748	569	40.42			963	309	41.30			934	267	41.13		
	V3			38.30				38.95					39.00					38.92					38.68		
4 (Mbps)	**V4**	1088	279	40.54	**1.6**	702	575	40.00	**5.7**	**−0.02**	710	610	40.06	**2.7**	**0.04**	1094	488	41.57	**0.7**	**0.61**	1221	374	41.76	**0.1**	**0.48**
	V5			39.26				39.78					39.88					39.95					39.47		
	V6	987	253	39.94		722	495	39.74			730	523	39.79			846	270	40.28			934	267	40.33		
	V2	1498	373	41.75		970	631	41.34			882	830	41.04			1134	375	41.94			1167	334	41.97		
	V3			38.82				39.61					39.92					39.51					39.14		
5 (Mbps)	**V4**	1347	347	41.32	**1.7**	932	680	40.98	**4.8**	**0.19**	844	905	40.67	**1.8**	**0.08**	1487	599	42.47	**0.8**	**0.69**	1527	466	42.48	**1.4**	**0.43**
	V5			39.89				40.64					40.69					40.71					40.57		
	V6	1208	313	40.58		968	579	40.73			875	757	40.42			1099	345	41.18			1170	267	40.33		
Average					**1.4**				**5.7**	**−0.02**				**2.2**	**0.13**				**1.3**	**0.60**				**0.5**	**0.46**

calculated according to (23), (24), (25) and they are used to predict the other similar parameters as

$$\alpha_i = \alpha, \quad \beta_{ij} = \beta, \quad \gamma_{ij} = \gamma \tag{26}$$

Meanwhile from (4), we can estimate τ_i for the texture map or the depth map as

$$\tau_i = \frac{ln(R_{iA}) - ln(R_{iB})}{ln(Q_{siA}) - ln(Q_{siB})} \tag{27}$$

TABLE V

RESULT SUMMARY OF DIFFERENT RC ALGORITHMS ON THE SEQUENCE *Champagne_Tower*

		Liu2011				Yuan2011					Liu2009					Proposed+CQ					Proposed+FL				
Target	**View**	**Rate**		**PSNR**	**E**	**Rate**		**PSNR**	**E**	**ΔP**	**Rate**		**PSNR**	**E**	**ΔP**	**Rate**		**PSNR**	**E**	**ΔP**	**Rate**		**PSNR**	**E**	**ΔP**
(Mbps)		(kbps)		(dB)	(%)	(kbps)		(dB)	(%)	(dB)	(kbps)		(dB)	(%)	(dB)	(kbps)		(dB)	(%)	(dB)	(kbps)		(dB)	(%)	(dB)
		T	**D**			**T**	**D**				**T**	**D**				**T**	**D**								
	V38	891	240	41.48		567	603	40.29			521	720	39.96			628	331	40.78			617	292	40.60		
	V39			39.05				40.35					39.79					39.90					39.47		
3	**V40**	640	171	41.20	**0.6**	518	471	40.39	**5.6**	**−0.12**	474	388	40.05	**2.8**	**−0.55**	761	479	41.92	**5.4**	**0.33**	769	410	41.82	**3.9**	**0.13**
(Mbps)	**V41**			38.60				39.82					39.31					39.87					39.44		
	V42	823	217	40.99		544	464	39.85			525	455	39.48			645	318	40.48			620	410	40.66		
	V38	1239	315	42.27		713	1035	41.22			646	1051	40.93			819	433	41.67			823	390	41.68		
	V39			39.59				41.38					41.11					40.84					40.46		
4	**V40**	896	233	41.83	**2.2**	649	571	41.37	**7.4**	**0.31**	590	534	41.09	**3.2**	**0.00**	1003	630	42.69	**4.5**	**0.69**	1020	545	42.73	**0.1**	**0.52**
(Mbps)	**V41**			39.12				41.04					40.54					41.08					40.53		
	V42	1121	285	41.47		731	595	40.84			658	647	40.58			856	438	41.43			837	389	41.49		
	V38	1494	397	42.98		832	1292	41.77			852	1182	41.89			936	547	42.16			1035	487	42.33		
	V39			40.10				42.13					41.83					41.44					41.14		
5	**V40**	1107	292	42.52	**0.6**	759	636	41.93	**3.3**	**0.35**	777	590	42.06	**0.1**	**0.31**	1178	707	43.13	**1.7**	**0.61**	1303	681	43.30	**2.4**	**0.50**
(Mbps)	**V41**			39.56				41.84					41.55					41.73					41.06		
	V42	1383	356	42.21		862	784	41.46			884	721	41.61			992	557	41.97			1033	583	42.03		
Average					**1.1**				**5.4**	**0.18**				**2.0**	**-0.08**				**3.8**	**0.54**				**2.1**	**0.38**

where R_{iA} and R_{iB} refer to the output bits of texture map or depth map that are precoded at the corresponding Q_{siA} (Q_{iA}) and Q_{siB} (Q_{iB}). τ^T of the texture and τ^D of the depth map in the first view are used as estimation for those of other coded views.

On the other hand, the sequence complexity related parameter ρ_i needs to be estimated for the texture and depth map of each view. For the first view, ρ_1 can be estimated according to precoding result as

$$\rho_1 = \frac{R_{1A} - R_{1B}}{e^{\tau_1(c_1 Q_{1A}+c_2)} - e^{\tau_1(c_1 Q_{1A}+c_2)}} \tag{28}$$

where ρ_1 refers to model parameters for either the texture map or the depth map. Since R_A and R_B of other views are unavailable, a limited number of frames are encoded for the texture and depth map of each view, and the output bit rate is recorded as sample complexity r_i. ρ_i of other views is estimated as

$$\rho_i = \frac{r_i}{r_1}\rho_1 \tag{29}$$

In this way, we can estimate the model parameters with reduced computational complexity.

C. Frame Level Bit Regulation

Given the total bit rate constraint, the optimal target bit rate (R_i^*) for each texture or depth map of each view can be calculated according to (22). To achieve the target bit of the each sequence, two RC schemes can be applied. One is to apply RC at frame level (FL) to adjust Q dynamically along the sequence to achieve the target bit rate. The other is to adopt a constant Q (CQ) to code the entire sequence. Since the fluctuation in Q usually degrade the R-D performance, the CQ usually has better R-D performance than FL. On the other hand, FL has more accurate target bit rate achievement due to the adaptive adjustment of Q value.

In this paper, we adopt both the FL and the CQ schemes to achieve the target bit. For the FL scheme, bit allocation algorithm in [28] is used to allocate target bit (R_t) at frame level and the corresponding Q_s for the coding frame is calculated based on (2) as

$$Q_s = \left(\frac{R_t}{\rho}\right)^{1/\tau}. \tag{30}$$

Then the corresponding Q can be attained with the Q-Q_s relation.

For the CQ, with the target bit of the sequence, the Q can be calculated based on (4) as

$$Q = \frac{ln(R^*/\rho) - c_2\tau}{c_1\tau} \tag{31}$$

where R^* is the target bit; ρ and τ are the estimated model parameters. Since we have already accessed the R-D characteristics of the sequence and estimated these model parameters, the calculated Q lead to the achievement of the target bit.

IV. EXPERIMENTAL RESULTS

The experiments are conducted under 5-view scenario as shown in Fig. 1, where 3 views are coded, and 2 virtual views are synthesized. The testing sequences include *Champagne_tower* (1280×960), *Balloons* (1024×768) provided by Nagoya University [29], and *Newspaper* (1024×768) provided by Gwangju Institute of Science and Technology (GIST) [30]. The texture and depth map of three views are separately encoded with MVC encoder [31] as I-view, P-view and P-view respectively. For P-view, the interview prediction is only applied for key frames. View Synthesis Reference Software (VSRS) [32] is used to synthesize the virtual view. 201 frames are encoded for each view.

A. Verification of Parameter Estimation Scheme

In this section, we verify the effectiveness of the proposed parameter estimation scheme in Section III-B. In the experiments, α, β, γ are estimated according to (26), and ρ is

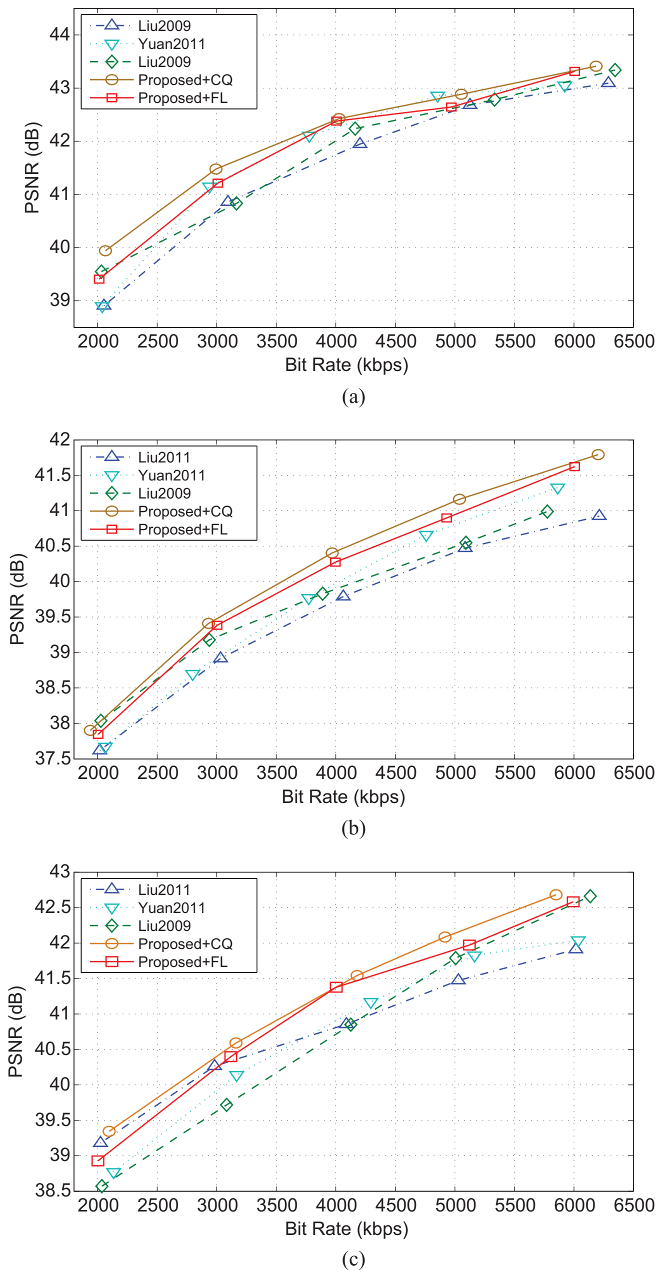


Fig. 8. R-D curves. The autostereoscopic 3-D video is set to 5-view scenario, where 3 views are coded views and 2 views are virtual views. The three-coded views are coded with MVC codec as I-view, P-view, and P-view, respectively. Search range is set to 96 with GOP size 4. Target bits are set at 2.0, 3.0, 4.0, 5.0, and 6.0 Mb/s and the corresponding R-D points are depicted for each algorithm. (a) *Balloons*. (b) *Newspaper*. (c) *Champagne_tower*.

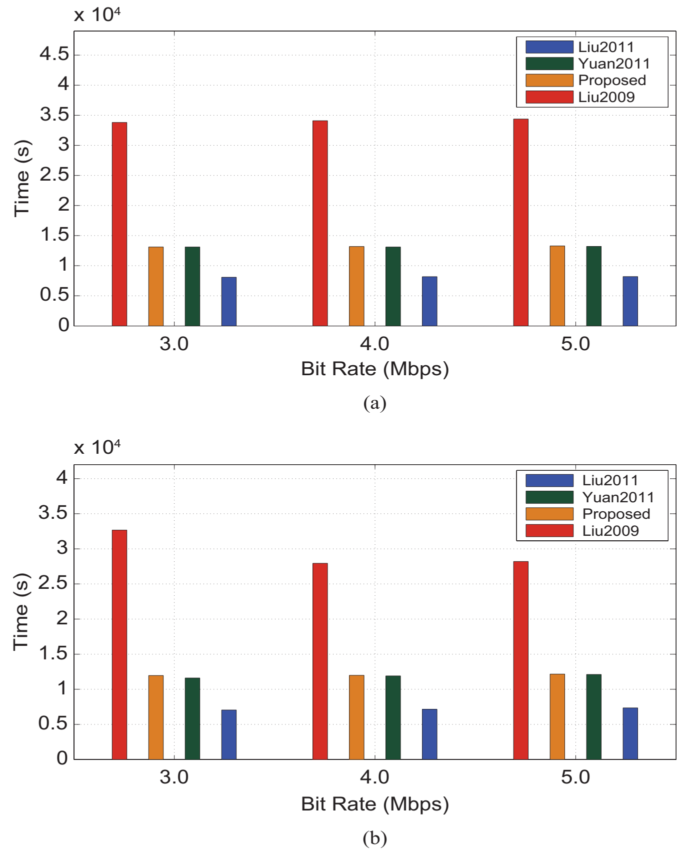


Fig. 9. Consumed coding time. The autostereoscopic 3-D video is set to 5-view scenario, where 3 views are coded views and 2 views are virtual views. The three-coded views are coded with MVC codec as I-view, P-view, and P-view, respectively. Search range is set to 96 with GOP size 4. Target bits are set at 3.0, 4.0, and 5.0 Mb/s for each algorithm. (a) *Balloons*. (b) *Newspaper*.

estimated according to (29), and τ^T and τ^D are estimated based on (27).

The results and the estimation errors are presented in Table II. We can see that the estimation is accurate enough that the mismatch is less than 5.5% on overage, which indicates the proposed scheme can achieve accurate model parameter estimation.

B. R-D Performance and Rate Accuracy

In order to evaluate the performance of the proposed RC algorithm, Liu2011 [15], Yuan2011 [14] and Liu2009 [13] are utilized for comparison. For the proposed algorithm, both the FL (proposed+FL) and the CQ (proposed+CQ) are employed to achieve the target bit for each sequence. Table III, IV and V summarize the output bits of the coded texture and the depth map of each view. Since the virtual views are generated with DIBR, for V2 and V4 in Table III and V3 and V5 in Table IV, V38 and V40 in Table V, there are no output bits for the texture and depth map. The ratio of output bits of the texture and depth map is fixed close to 4:1 for Liu2011.

To evaluate the accuracy of the bit rate achievement, the following measurement is adopted

$$E = \frac{|R_{all} - R_{target}|}{R_{target}} \times 100\% \tag{32}$$

where R_{all} is the total bits used to encode the depth map and texture map of three views; R_{target} is the target bit rate. Table III, IV and V present the rate achievement accuracy of different algorithms. We can see the proposed+FL generally has the best performance that its mismatch is 0.4%, 0.5%, 2.1% on average for different sequences. The proposed+CQ also achieves acceptable accuracy, i.e. 0.7%, 1.3% and 3.8% on average.

The PSNR of both coded and virtual views are recorded for each view. The average PSNR is used to evaluate the overall quality performance of five views for different algorithms. The average PSNR of Liu2011 is set as the benchmark and the performances of other algorithms are measured as

$$\Delta P = P_i - \hat{P} \tag{33}$$

where $\hat{P}$ refers to the average PSNR of Liu2011 and P_i refers to the average PSNR of the rest algorithms. The results are presented in Table III, IV, and V, where we can see the proposed+CQ achieves the best performance that the average PSNR gains are 0.43 dB, 0.60 dB and 0.54 respectively, while the proposed+FL has little degradation in R-D performance achieving 0.25 dB, 0.46 dB and 0.38 dB gain respectively.

For further illustration, typical R-D curves for different algorithms are shown in Fig. 8, where we can see the proposed+CQ demonstrates the best R-D efficiency among the different algorithms.

The computational complexity is compared for different algorithms. The computation complexity mainly comes from view coding and view synthesis process. For Liu2011, each view including the texture and depth map is coded with single pass, while for the proposed algorithm and Yuan2011, two additional iterations are required for the first view. For Liu2009, the texture maps of each view have to be coded for M times and the depth maps have to be code for three times, where M is the number of proper Q values which generate bit rate falling into the range $[1/2R_t, R_t]$. As for view synthesis, Liu2011 has to synthesize two virtual views, while for the proposed method and Yuan2011 three more virtual views need to be synthesized to assist model parameter calculation. For the Liu2009, $3M$ more virtual view synthesis are required to calculate the model parameters. Therefore, Liu2011 consumes the lowest computation complexity. Although the proposed algorithm requires additional computation in the precoding stage, with the increase of view number, the portion of the additional complexity in the total complexity will decrease. The actual consuming time for each algorithm is presented in Fig. 9, where we can see the proposed algorithm takes less time than Liu2009 and is comparable with Yuan2011.

V. Conclusion

In this paper, we proposed a RC scheme to achieve the best overall quality for 3DV. Based on power models for the R-Q_s and the MSE-Q_s relationship, we derived the exponential R-Q relationship and the linear $PSNR$-Q relationship. Furthermore, a linear model is approximated for the quality dependency between the virtual view and the coded view. Based on the above R-D characteristics of both the coded view and the virtual view, a R-D optimized RC algorithm is derived. Experiments are conducted on different video sequences and the results demonstrate the effectiveness of the proposed algorithm.

References

[1] A. Smolic, "3D video and free viewpoint video-from capture to display," *Pattern Recognit.*, vol. 44, no. 9, pp. 1958–1968, Sep. 2011.

[2] J. Konrad and M. Halle, "3-D displays and signal processing," *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 97–111, Nov. 2007.

[3] N. Dodgson, "Autostereoscopic 3-D displays," *IEEE Comput.*, vol. 38, no. 8, pp. 31–36, Aug. 2005.

[4] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. von Kopylow, "A survey of 3DTV displays: Techniques and technologies," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1647–1658, Nov. 2007.

[5] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proc. IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.

[6] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[7] K. Muller, P. Merkle, and T. Wiegand, "3-D video representation using depthmaps," *Proc. IEEE*, vol. 99, no. 4, pp. 643–656, Apr. 2011.

[8] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1. Oct. 2007, pp. 201–204.

[9] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Process., Image Commun.*, vol. 22, no. 2, pp. 217–234, Feb. 2007.

[10] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," *Proc. SPIE Stereosc. Displays Virtual Real. Syst. XI,* vol. 5291, pp. 93–104, Jan. 2004.

[11] I. Daribo, C. Tillier, and B. Pesquet-Popescu, "Motion vector sharing and bitrate allocation for 3D video-plus-depth coding," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–13, Jun. 2009.

[12] Y. Morvan, D. Farin, and P. H. N. de With, "Joint depth/texture bit allocation for multi-view video compression," in *Proc. 26th Picture Coding Symp.*, Nov. 2007, pp. 265–268.

[13] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "Joint video/depth rate allocation for 3-D video coding based on view synthesis distortion model," *Signal Process., Image Commun.*, vol. 24, no. 8, pp. 666–681, Sep. 2009.

[14] H. Yuan, Y. Chang, J. Huo, F. Yang, and Z. Lu, "Model-based joint bit allocation between texture videos and depth maps for 3-D video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 485–497, Apr. 2011.

[15] Y. Liu, Q. Huang, S. Ma, D. Zhao, W. Gao, S. Ci, and H. Tang, "A novel rate control technique for multiview video plus depth based 3D video coding," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 562–571, Jun. 2011.

[16] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246–250, Feb. 1997.

[17] Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT," in *Proc. 7th JVT Meeting*, Pattaya, Thailand, Mar. 2003, pp. 1–28.

[18] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[19] J. Dong and N. Ling, "A context-adaptive prediction scheme for parameter estimation in H.264/AVC macroblack layer rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp. 1108–1117, Aug. 2009.

[20] S. Hu, H. Wang, S. Kwong, and C.-C. J. Kuo, "Rate control optimization for temporal-layer scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, pp. 1152–1162, Aug. 2011.

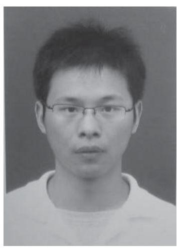
[21] N. Kamaci, Y. Altinbasak, and R. M. Mersereau, "Frame bit allocation for the H.264/AVC video coder via Cauchy density-based rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.

[22] S. Hu, H. Wang, S. Kwong, and C.-C. J. Kuo, "Novel rate-quantization model-based rate control with adaptive initialization for spatial scalable video coding," *IEEE Trans. Ind. Electron.*, vol. 59, no. 3, pp. 1673–1684, Mar. 2012.

[23] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia*. New York: Wiley, 2003.

[24] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Muller, P. H. N. de With, and T. Wiegand, "The effects of multiview depth video compression on multiview rendering," *Signal Process., Image Commun.*, vol. 24, pp. 73–88, Jan. 2009.

[25] H. T. Nguyen and M. N. Do, "Error analysis for image-based rendering with depth information," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 703–716, Apr. 2009.
[26] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View synthesis for advanced 3D video systems," *EURASIP J. Image Video Process.*, vol. 2008, pp. 1–11, 2008, Article ID 438148.
[27] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 600–608, Aug. 2004.
[28] A. Leontaris and A. M. Tourapis, "Rate control for the joint scalable 638 video model (JSVM)," Joint Video Team, San Jose, CA, Tech. Rep. JVT-W043, Apr. 2007.
[29] *MPEG-FTV Project*. (2011) [Online]. Available: http://www.tanimoto.nuee.nagoya-u.ac.jp/mpeg/mpeg_ftv.html
[30] *GIST* (2011, Jan.) [Online]. Available: ftp://203.253.128.142
[31] *Joint Multiview Video Coding JMVC 8.5 Software Package*, ISO/IEC Standard JTC1/SC29/WG11, Mar. 2011.
[32] *Report on Experimental Framework for 3-D Video Coding*, ISO/IEC Standard JTC1/SC29/WG11, Oct. 2010.



Sudeng Hu received the B.Eng. degree from Zhejiang University, Hangzhou, China, and the M.Phil. degree from the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, in 2007 and 2010, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, University of Southern California, Los Angeles.

He was a Research Associate with the Department of Computer Science, City University of Hong Kong, from 2010 to 2011. In 2012, he was an Intern with Mitsubishi Electric Research Laboratories, Cambridge, MA. His current research interests include image and video compression, rate control, scalable video coding, and 3-D video coding.

Sam Kwong (SM'04) received the B.S. and M.S. degrees in electrical engineering from the State University of New York at Buffalo, Buffalo, in 1983, and the University of Waterloo, Waterloo, ON, Canada, in 1985, respectively, and the Ph.D. degree from the University of Hagen, Hagen, Germany, in 1996.

He was a Diagnostic Engineer with Control Data Canada from 1985 to 1987. He joined Bell Northern Research, Ottawa, ON, as a Scientific Staff Member. In 1990, he became a Lecturer with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, where he is currently a Professor with the Department of Computer Science. His current research interests include video and image coding and evolutionary algorithms.

Yun Zhang received the B.S. and M.S. degrees in electrical engineering from Ningbo University, Ningbo, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, China, in 2010.

He was a Post-Doctoral Researcher with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, from 2009 to 2011. In 2010, he joined the Shenzhen Institute of Advanced Technology, CAS, as an Assistant Professor. His current research interests include multiview video coding, 3-D video processing, and visual perception.

C.-C. Jay Kuo (F'99) received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1980, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He is currently the Director of the Multimedia Communications Laboratory and a Professor of electrical engineering, computer science, and mathematics with the Ming-Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles. He has co-authored over 200 journal papers, 850 conference papers, and ten books. His current research interests include digital image and video analysis and modeling, multimedia data compression, communication and networking, and biological signal and image processing.

Dr. Kuo is a fellow of The American Association for the Advancement of Science and The International Society for Optical Engineers.