Rate Distortion Optimized Inter-View Frame Level Bit Allocation Method for MV-HEVC

Hui Yuan, *Member, IEEE*, Sam Kwong, *Fellow, IEEE*, Xu Wang, Wei Gao, and Yun Zhang, *Member, IEEE*

***Abstract*—In multi-view video coding, since inter-view prediction has been adopted as an important coding tool which could improve coding efficiency greatly, inter-view dependency is inevitable, i.e., the distortion of the reference view (RV) picture could be propagated to the non-reference view (NRV) pictures. Therefore, in order to achieve higher coding efficiency, the inter-view dependency must be taken into account for inter-view bit allocation. In this paper, the inter-view dependency is analyzed in detail, and a rate-distortion (RD) model for NRVs is derived by taking the distortion of RV into account. Based on the derived RD model, the inter-view bit allocation is represented as a mathematical problem with an analytic form, and is solved by a convex optimization (Lagrangian Multiplier) method. Experimental results demonstrate that the RD performance and the inter-view quality consistency of the proposed method is better than existing methods, while the complexity of the proposed method is comparable with the existing methods.**

***Index Terms*—Bit allocation, inter-view dependency, MV-HEVC, rate distortion optimization.**

I. Introduction

DUE TO THE rapid development of 3D multimedia technologies, e.g. IMAX movie theaters[1] and low cost personal 3D displays, 3D video (3DV) applications is becoming more and more popular for both movie industry and home entertainment. In 3DV applications, 2 pictures, i.e. the left and the right view pictures, which correspond to the left and the right eyes of an observer, are projected onto a special-coated screen simultaneously. The observer can use 3D glasses to separate the 2 pictures so as to let them be projected into his corresponding eyes. Then, the human brain will fuse the 2 pictures together and enjoy a depth perception. To support the free view function of 3D applications, a set of view pictures (at least 2 view pictures) should be captured/generated, compressed, and transmitted to the 3DV application terminal. Thus, compared with the traditional 2D videos, more storage space and network bandwidth are needed for 3DVs. Due to the limited storage capacity and network bandwidth, efficient compression of 3DVs becomes critical and important for 3DV applications.

In order to compress 3DVs efficiently, Multi-view Video Coding (MVC) [1] was developed as an important extension of H.264/AVC [2]. Unlike the traditional method, the MVC exploits the inter-view correlation [3] and encodes multiview pictures simultaneously. Recently, a Joint Collaborative Team for 3DV (JCT-3 V), which was established by Moving Pictures Experts Group (MPEG) of International Organization for Standardization (ISO) and Video Coding Experts Group (VCEG) of International Telecommunication Union (ITU), developed a new MVC encoder based on the state of the art High Efficiency Video Coding (HEVC) Standard [4] in order to further improve the coding efficiency. Accordingly, the new MVC standard is named as MV-HEVC [5]. In the current MV-HEVC encoder common test condition [6], only 2 and 3 view cases are considered. Based on the MV-HEVC codec, a full 3D-HEVC standard [7] was also developed to generate more views, in which depth maps have been taken into considerations.

In 3DV coding, bit allocation is very important to improve the rate distortion (RD) performance. The bit allocation problem of 3DVs can be divided as 2 steps, *i.e.*, bit allocation for texture videos and depth maps (*e.g. Wang's* algorithm [8], *Cheung's* algorithm [9], and our previous work [10]), and bit allocation for multiview (only texture videos are considered) pictures. In this paper, the bit allocation for multiview pictures is discussed based on MV-HEVC.

In MV-HEVC, inter-view prediction [11] based on disparity compensation is adopted as an important coding tool. Recently, lots of researchers focused on developing novel inter-view prediction methods [12]–[14] so as to improve the coding efficiency. Besides, for the same rate constraint, different inter-view bit allocation methods may give different average qualities of reconstructed view pictures. *Yao et al.* [15] proposed a joint bit allocation scheme for multiple stereoscopic 3DVs. In this method, the reference view (RV) picture is encoded

Manuscript received December 28, 2014; revised May 24, 2015 and July 25, 2015; accepted September 02, 2015. Date of publication September 10, 2015; date of current version November 13, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61201211, Grant 61571274, Grant 61272289, and Grant 61501299, in part by the Young Scholars Program of Shandong University under Grant 2015WLJH39, in part by the City University of Hong Kong Applied Research Grant 9667094, in part by the Ph.D. Programs Foundation, Ministry of Education of China under Grant 20120131120032, in part by the Excellent Youth Scientist Award Foundation of Shandong Province under Grant BS2012DX021, in part by the Key Laboratory of Wireless Sensor Network and Communication, Chinese Academy of Sciences under Grant 2013002, and in part by the City University of Hong Kong Shenzhen Research Institute. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Adrian Munteanu.

H. Yuan is with the School of Information Science and Engineering, Shandong University, Ji'nan 250100, China, with the Key Laboratory of Wireless Sensor Network & Communication, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China, and with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: yuanhui0325@gmail.com; huiyuan@sdu.edu.cn).

S. Kwong is with the Department of Computer Science, City University of Hong Kong, Hong Kong, and also with the City University of Hong Kong Shenzhen Research Institute, Shenzhen 5180057, China (e-mail: cssamk@cityu.edu.hk).

X. Wang is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: wangxu@szu.edu.cn).

W. Gao is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: gaowei262@126.com).

Y. Zhang is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (e-mail: yun.zhang@siat.ac.cn).

Digital Object Identifier 10.1109/TMM.2015.2477682

[1]"The IMAX experience," [Online]. Available: http://www.imax.com/about/experience/3d/

independently which is the same with MV-HEVC encoder, but the non-reference view (NRV) pictures are encoded as an enhancement layer which is compatible with scalable video codec. Then, a fast recursive golden-section search method is employed to allocate bits for different view pictures. Although a sub-optimal result could be achieved, multi-pass coding which is time-consuming must be used so as to generate enough search points. *Chang et al.* [16] proposed a frame level joint rate control scheme for MVC. In this method, the bit allocation between the left and the right view is converted as an optimization problem. However, the inter-view dependency is not considered in the optimization model. Moreover, in [17]–[20], a predefined fixed bit ratio is used for inter-view bit allocation; while for different 3DVs, the predefined fixed bit ratios are different. In the current MV-HEVC codec, an empirically optimal bit ratio is also adopted for inter-view bit allocation. For 2 view case, the empirically optimal bit ratio is 0.8 for RV picture, and 0.2 for NRV picture; while for 3 view case, the empirically optimal bit ratio is 0.66 for RV picture, and 0.17 for each of the 2 NRV pictures [21]. Besides, it should be noted that only sequence level bit allocation method were proposed in [17]–[21]. Motivated by the above analysis, a frame level inter-view bit allocation method based on inter-view dependency analysis is proposed to improve the coding efficiency for MV-HEVC.

In order to design a RD optimal bit allocation method for multiview pictures, accurate RD model must be obtained first. In the early studies, the RD model was derived based on the assumption that the transformed coefficients have Gaussian distributions [22]. Later, more accurate RD models were derived by considering that the transformed coefficients have Laplacian [23], [24], Generalized Gaussian [25], and Cauchy [26] distributions. Among them, the Cauchy distribution based RD model is with the best accuracy. Therefore Cauchy distribution based RD model is used in this paper. Exactly speaking, there are mainly 3 contributions, i.e.:

a) a simplified Cauchy density [26] based RD model is proposed to describe the RD characteristic for RV picture;
b) a novel RD model for NRV pictures is proposed based on inter-view dependency analysis; and
c) an optimal analytic solution and the corresponding one-pass coding scheme for inter-view bit allocation.

The remainder of this paper is organized as follows. In Section II, a simplified Cauchy density based RD model is derived to describe the RD characteristic for RV pictures; besides, based on the inter-view dependency analysis, the RD model for NRV pictures is derived as well. In Section III, the inter-view bit allocation problem is modeled as a convex optimization problem by taking different temporal levels (TLs) into account, and is solved by Lagrangian Multiplier Method. Experimental results and conclusions are given in Sections IV and V, respectively.

II. Rate Distortion Analysis

In H.264/MVC or MV-HEVC, the coding tools of RV pictures are the same with those of single view video coding; while for NRV pictures, inter-view prediction is used as an additional coding tool by taking the RV picture as a reference picture so as to achieve higher coding efficiency. In order to model the inter-view bit allocation problem as an analytical form, the rate distortion characteristics of RV and NRV pictures are analyzed as follows.

A. Simplified Cauchy Model Derivation

The aim of video coding and transmission is to provide the best reconstruction quality under a certain network bandwidth. Therefore, it is no doubt that seeking the best tradeoff between rate and distortion is very important. In order to find the best tradeoff, the relationship between rate and distortion, i.e., rate-distortion (RD) model must be confirmed first.

In the Cauchy distribution based RD model, the relationship between distortion (D, measured by mean squared error, i.e., MSE) and rate (R, measured by average coding bits per pixel) could be written as

$$D(R) = \alpha \cdot R^{-\beta} \tag{1}$$

where α, $\beta > 0$ are model parameters. However, the Cauchy based RD model cannot be applied easily because the power β is an uncertainty parameter. For easier application, the model (1) can be simplified by using the Taylor Series[2] expansion at $R = 0$. Therefore, we have

$$\begin{aligned}\frac{1}{D} = \frac{1}{\alpha}R^{\beta} &= \sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}(R-0)^n \\ &= \frac{1}{\alpha}\cdot\left[\beta\cdot R + \frac{1}{2!}\beta(\beta-1)R^2 + \cdots\right] \\ &= \frac{\beta}{\alpha}R + \frac{1}{2!}\frac{\beta(\beta-1)}{\alpha}R^2 + \cdots.\end{aligned} \tag{2}$$

Usually, the average coding bits of all pixels is far less than 1, therefore, (2) could be simplified as

$$\frac{1}{D} = \frac{\beta}{\alpha}R + O(R) \approx \frac{\beta}{\alpha}R + \varepsilon \tag{3}$$

where $O(R)$ represents the high order infinitesimals of R, which is approximated as a constant value ε. Thus, for easy illustration, (3) could be written as

$$D = (a\cdot R + b)^{-1}. \tag{4}$$

Where a (corresponds to β/α) and b (corresponds to ε) are model parameters.

Therefore, the RD model of RV pictures could be represented as

$$D_{RV} = (a_{RV}\cdot R_{RV} + b_{RV})^{-1} \tag{5}$$

in which D_{RV} and R_{RV} represent the distortion and average coding bits of RV pictures, a_{RV} and b_{RV} are parameters. Fig. 1 shows the relationship between $1/D_{RV}$ and R_{RV}, and the curves fitted by the proposed model and Laplacian based model [25]. Besides, Table I shows the detailed quantitative comparison between the proposed model and the Laplacian based model. From Fig. 1 and Table I, we can observe that the propose model is more accurate than the Laplacian based model.

B. Inter-View Distortion Dependency Analysis

In multi-view video coding, a NRV picture of the current frame could be predicted (called as disparity compensation prediction, DCP [27]) from the corresponding RV picture, as shown

[2][Online]. Available: http://en.wikipedia.org/wiki/Taylor_Series

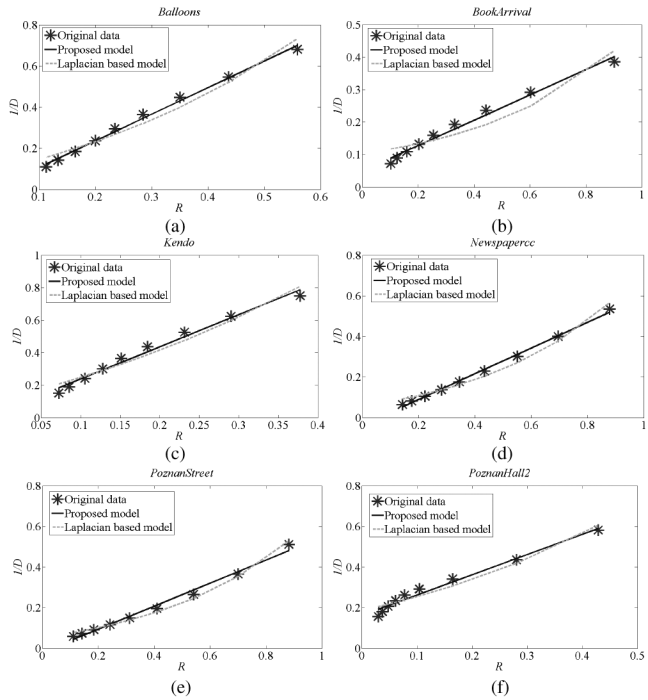


Fig. 1. Comparison between the proposed model and the Laplacian-based R-D model.

TABLE I
ROOT OF MEAN SQUARED ERROR (RMSE) AND CORRELATION COEFFICIENTS COMPARISONS (CC)

Sequence	Proposed Model		Laplacian Based Model	
	RMSE	**CC**	RMSE	CC
Balloons	**0.0129**	**0.9975**	0.0361	0.9803
BookArrival	**0.0115**	**0.9930**	0.0325	0.9430
Kendo	**0.0245**	**0.9918**	0.0411	0.9769
Newspapercc	**0.0085**	**0.9984**	0.0227	0.9886
PoznanStreet	**0.0161**	**0.9937**	0.0143	0.9950
PoznanHall2	**0.0189**	**0.9892**	0.0282	0.9757

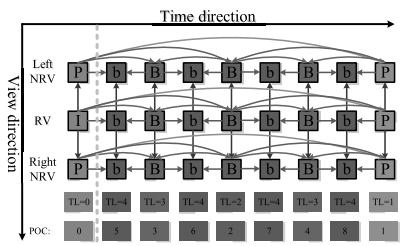


Fig. 2. Coding structure example of MV-HEVC.

in Fig. 2. During the DCP procedure, when *interview SKIP mode* is selected, the matching Prediction Units (PUs) in the RV picture will be directly copied to the current PU in the NRV picture. Thus, the distortion of the matching PU in the RV picture will also be propagated to the current PU in the NRV picture directly. Accordingly, for those PUs with interview SKIP mode, the distortion could be written as

$$D_{NRV,SKIP} = \eta\left(D_{RV} + \delta\right) \tag{6}$$

where D_{RV} is the distortion of the RV picture, η is the percentage of interview SKIP mode which is predicted from RV

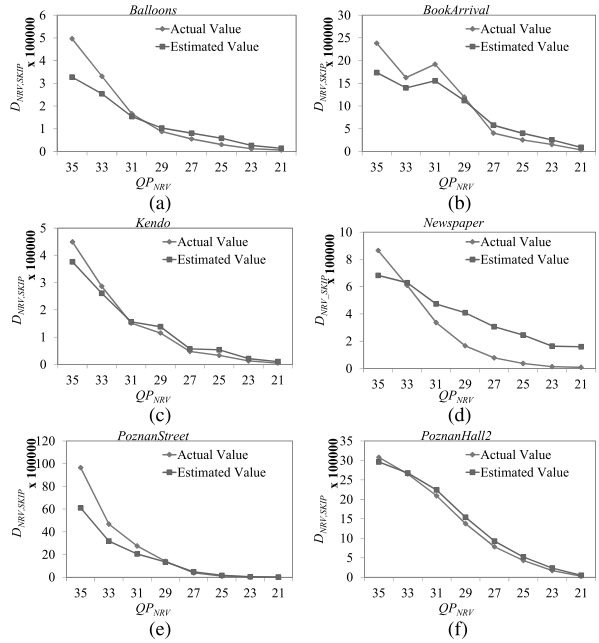


Fig. 3. Relationship between the actual D_{NRV_SKIP} and the estimated one.

TABLE II
CC AND RMSE BETWEEN THE ACTUAL $D_{NRV,SKIP}$ AND THE ESTIMATED ONE

Sequence	CC	RMSE	Sequence	CC	RMSE
Balloons	0.986	0.086	*Newspaper*	0.9773	0.226
BookArrival	0.985	0.292	*PoznanStreet*	0.995	0.662
Kendo	0.995	0.038	*PoznanHall2*	0.996	0.053

picture, δ is determined by wrong estimated disparity, non-verified block-based correlation assumption, and the inherent luminance discrepancy between the RV and NRV pictures, etc.

The relationships between the actual D_{NRV_SKIP} and the estimated D_{NRV_SKIP} by using (6) are shown in Fig. 3, in which D_{RV} is obtained by encoding the RV picture with a quantization parameter (QP) of 28 (denoted as QP_{RV}); D_{NRV_SKIP} is obtained by encoding the NRV picture with variable QPs (i.e. 21, 23, 25, 27, 29, 31, 33, and 35, denoted as QP_{NRV}); both D_{RV} and D_{NRV_SKIP} are normalized by the total number of pixels in a picture (this is why the distortions are small). For easy statistics, only the first frame was encoded. From Fig. 3, we can observe that the actual D_{NRV_SKIP} and the estimated one are similar for all the test video sequences. The estimation errors are caused by δ which is usually non-zero and could also be affected by the compression procedure. For example, since the interview discrepancy of the sequence *Newspapercc* is larger than other sequences, the estimation of D_{NRV_SKIP} is not accurate enough compared with other sequences. Besides, Table II shows the correlation coefficients (CC) and the root of mean squared estimation errors (RMSE) between the actual D_{NRV_SKIP} and the estimated one, from which it could be concluded that (6) is accurate to some extent.

When other (non-SKIP) prediction modes are selected, the matching PU in the RV picture is only used as a prediction of the current PU, and the residual signal is then obtained by subtracting it from the original signal of the current PU. Let

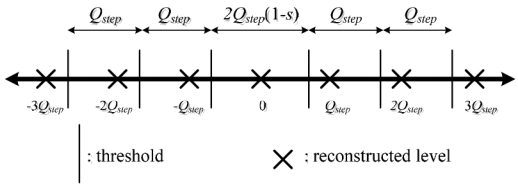


Fig. 4. Quantization example.

$\mathbf{L_{NRV,orig}}$ denotes a transform unit (TU) of the current PU, $\mathbf{L_{NRV,pred}}$ denotes the prediction (with a non-SKIP mode) of the TU. Then, the transform coefficients could be written as

$$\mathbf{W} = \mathbf{C}\left[\mathbf{L}_{\mathrm{NRV,orig}} - \mathbf{L}_{\mathrm{NRV,pred}}\right]\mathbf{C}^{\mathrm{T}} \tag{7}$$

where $\mathbf{W}$ is the transform coefficients matrix, $\mathbf{C}$ is the transform basis matrix. During the quantization procedure, a transformed coefficient w (an element of $\mathbf{W}$) could be quantized as

$$z = \lfloor |w| / Q_{step} + s \rfloor \cdot sgn(w) \tag{8}$$

where z is the quantization level of w, Q_{step} is the quantization step size (determined by QP), s denotes rounding offset [28], $\lfloor \cdot \rfloor$ is the floor operation, and $sgn(\cdot)$ is a sign function which returns the sign of the input signal. From (7) and (8), we can conclude that when the distortion of the RV picture is changed, e.g., the QP of the RV picture is changed, $\mathbf{L_{NRV,pred}}$ will also be changed; and then, the quantization level z will be changed as well. At the decoder, the quantization level z could be reconstructed to w_{rec} by the inverse quantization operation as shown in (9) and Fig. 4

$$w_{rec} = z \cdot Q_{step}. \tag{9}$$

It can be concluded from (9) and Fig. 4 that the w_{rec} may be changed when z is changed. However, no matter what the change is, the difference between w and w_{rec} will fall within the range of $[-(1-s)Q_{step}, (1-s)Q_{step}]$ which depends on Q_{step}. Furthermore, consider N level uniform quantizer, the quantized distortion could be calculated as

$$\begin{aligned} D = &\int_{-(1-s)Q_{step}}^{(1-s)Q_{step}} (w)^2 f(w)\, d_w \\ &+ \sum_{n=1}^{N} \int_{(n-s)Q_{step}}^{(n-s+1)Q_{step}} (w - nQ_{step})^2 f(w) d_w \end{aligned} \tag{10}$$

from which we can observe that the quantized distortion depends on Q_{step} and the distribution function of w, i.e. $f(w)$. For a uniform quantizer, it is well known that the quantized distortion could be approximated as [29]

$$D = {Q_{step}}^2/12 \tag{11}$$

no matter what $f(w)$ is. This means that the distortion of RV picture will not affect the distortion of blocks with non-SKIP modes in NRV pictures. D_{NRV,non_SKIP} is mainly determined by the QP of the NRV pictures. In order to verify this point, Fig. 5 shows the relationship between D_{RV} and the distortions of non-SKIP PUs in the NRV pictures (denoted as D_{NRV,non_SKIP}). In Fig. 5, D_{RV} is obtained by encoding the RV picture with QP = 20, 22, 24, 26, 28, 30, 32, 34; while D_{NRV,non_SKIP} is obtained by encoding the NRV picture with

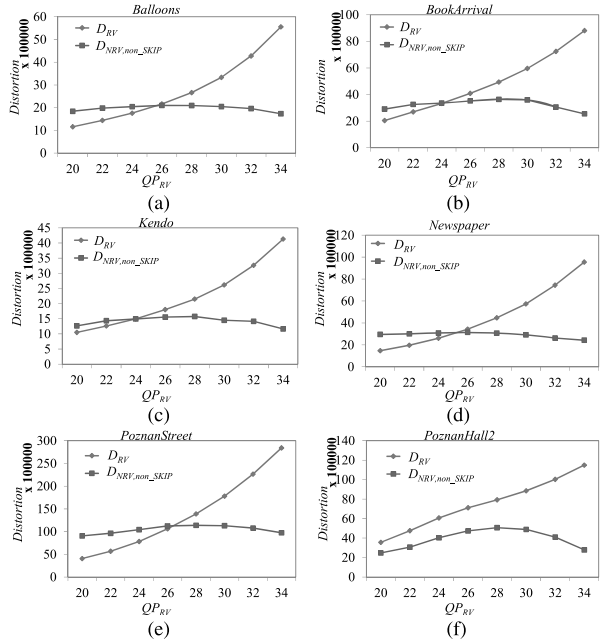


Fig. 5. Relationship between D_{RV} and D_{NRV,non_SKIP}.

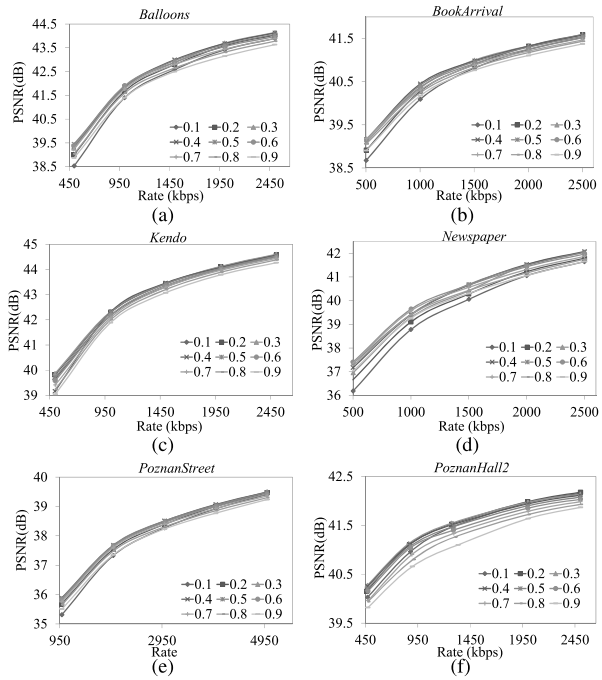


Fig. 6. RD performances under different initial bit ratio for 2 view case.

QP=27, from which we can observe that D_{NRV,non_SKIP} remain stable when D_{RV} is changed. Actually, the standard deviations of D_{NRV,non_SKIP} are only 0.163, 0.480, 0.179, 0.317, 0.425, and 0.484 for Fig. 5(a), (b), (c), (d), (e), and (f) respectively. Therefore, it could be concluded that the influence of D_{RV} on D_{NRV,non_SKIP} is small.

Since rate and distortion are one to one correspondence, the coding bit rate of the PU with non-SKIP modes is also mainly determined by the quantization process. Furthermore, since the residual coding bits of SKIP mode is 0, the coding bits of the NRV picture depends only on those non-SKIP PUs. Based on the simplified Cauchy rate distortion model shown in (5), the

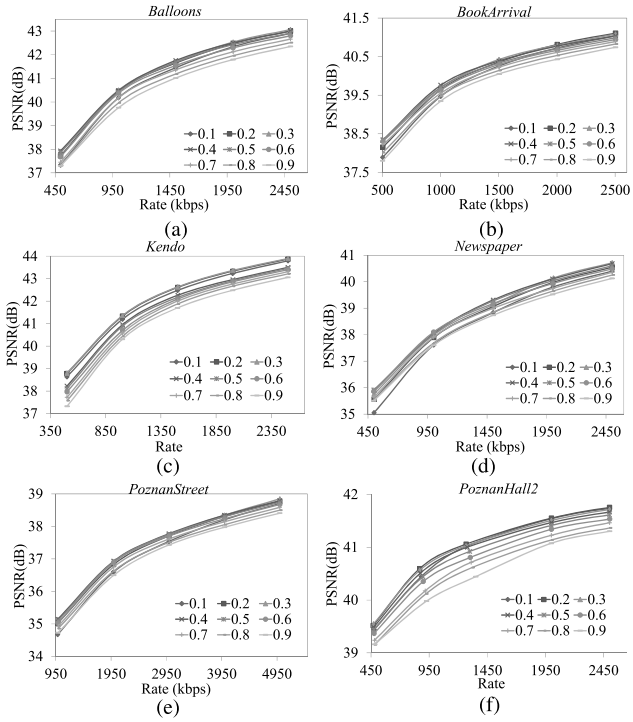


Fig. 7. RD performances under different initial bit ratio for 3 view case.

TABLE III
CODING CONFIGURATIONS

Encoder Parameters	Values	Encoder Parameters	Values
Unit Definition		**Motion Search**	
MaxCUWidth	64	FastSearch	1
MaxCUHeight	64	SearchRange	64
MaxPartitionDepth	4	BipredSearchRange	1
QuadtreeTULog2MaxSize	5	HadamardME	1
QuadtreeTULog2MinSize	3	FEN	1
QuadtreeTUMaxDepthInter	3	FDM	1
QuadtreeTUMaxDepthIntra	3		
Coding Structure		**Other Parameters**	
IntraPeriod	-1	WaveFrontSynchro	0
DecodingRefreshType	1	UniformSpacingIdc	0
GOPSize	8	PCMEnabledFlag	0
		SliceMode	0
Quantization		**Deblock Filter**	
MaxDeltaQP	0	DeblockingFilterControlPresent	1
MaxCuDQPDepth	0	LoopFilterOffsetInPPS	0
DeltaQpRD	0	LoopFilterDisable	0
RDOQ	1	LoopFilterBetaOffset_div2	0
RDOQTS	1	LoopFilterTcOffset_div2	0
ScalingList	0	DeblockingFilterMetric	0
Coding Tools		**Multiview Coding Tools**	
SAO	1	IvMvPred	1
AMP	1	AdvMultiviewResPred	1
TransformSkip	1	IlluCompEnable	1
TransformSkipFast	1	ViewSynthesisPred	1
SAOLcuBoundary	0	DepthRefinement	1
Rate Control			
RateControl	1		
KeepHierarchicalBit	1		
LCULevelRateControl	1	**NA**	
RCLCUSeparateModel	1		
InitialQP	0		
RCForceIntraQP	0		

relationship between the distortion of non-SKIP PUs and the coding bits could be written as

$$D_{NRV,non\_SKIP} = \left(\alpha_{NRV} R_{NRV} + \beta_{NRV}\right)^{-1} \tag{12}$$

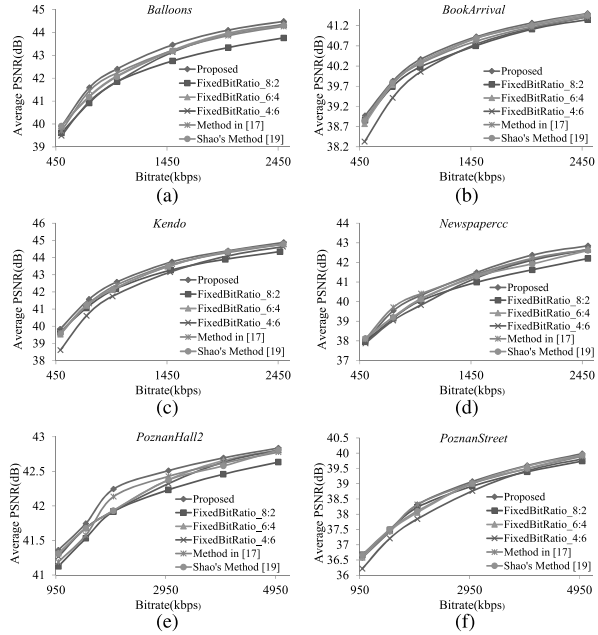


Fig. 8. RD curves comparison for 2 view case.

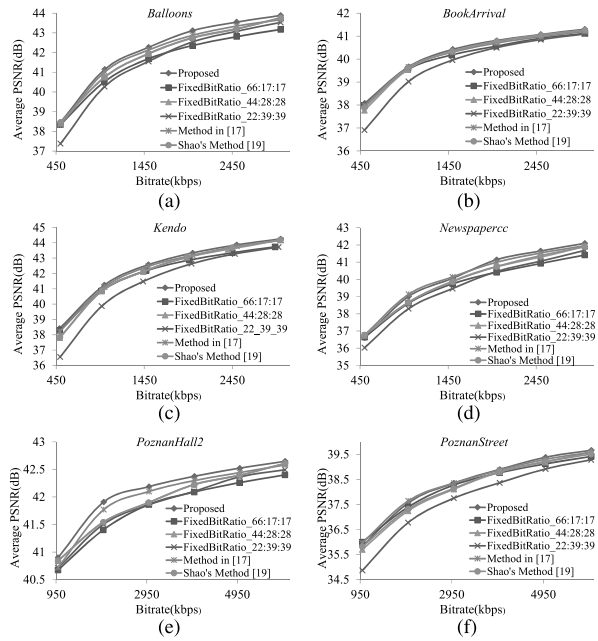


Fig. 9. RD curves comparison for 3 view case.

where R_{NRV} denotes the coding bits of the NRV picture, α_{NRV} and β_{NRV} are model parameters. Accordingly, the distortion of NRV picture could be written as

$$\begin{aligned} D_{NRV} &= D_{NRV,SKIP} + D_{NRV,non\_SKIP} \\ &= \eta\left(D_{RV} + \delta\right) + \left(a_{NRV} R_{NRV} + b_{NRV}\right)^{-1} \\ &= \eta\left(a_{RV} R_{RV} + b_{RV}\right)^{-1} \\ &\quad + \left(a_{NRV} R_{NRV} + b_{NRV}\right)^{-1} + \eta\delta. \end{aligned} \tag{13}$$

Moreover, for 3 view case of MV-HEVC, since the middle view picture is coded as the RV picture, both the left and the right view pictures are coded as NRV pictures, the distortion dependency between RV and NRV pictures is much larger than

TABLE IV
DETAILED RD VALUES COMPARISON FOR 2 VIEW CASE

Sequences	Proposed Method		Bit Ratio 8:2		BD PSNR (L/H)	Bit Ratio 6:4		BD PSNR (L/H)	Bit Ratio 4:6		BD PSNR (L/H)	Method in [17]		BD PSNR (L/H)	Shao's Method [19]		BD PSNR (L/H)
	Total Rate	Avg. PSNR	Total Rate	Avg. PSNR		Total Rate	Avg. PSNR		Total Rate	Avg. PSNR		Total Rate	Avg. PSNR		Total Rate	Avg. PSNR	
Balloons	**499.9**	**39.9**	499.6	39.6		500.3	39.9		500.0	39.5		500.0	39.7		500.1	39.9	
	752.5	**41.6**	749.4	40.9		749.2	41.2		750.1	41.0		752.3	41.4		749.2	41.2	
	1002.3	**42.4**	1001.5	41.9	**0.578 /**	999.3	42.1	**0.280 /**	999.1	41.8	**0.525 /**	1002.1	42.2	**0.181 /**	999.0	42.1	**0.303 /**
	1500.4	**43.5**	1500.2	42.8	**0.695**	1500.3	43.2	**0.216**	1499.6	43.1	**0.309**	1500.3	43.2	**0.233**	1500.1	43.2	**0.253**
	1999.8	**44.1**	2000.6	43.3		2000.5	44.0		2001.9	43.9		1999.7	43.9		2000.6	43.9	
	2500.0	**44.5**	2500.1	43.8		2500.8	44.3		2500.9	44.3		2500.1	44.3		2500.9	44.3	
BookArrival	**499.1**	**39.0**	497.7	38.9		494.3	38.8		494.2	38.3		499.7	38.9		493.9	38.8	
	748.8	**39.8**	744.6	39.7		743.0	39.7		746.9	39.4		749.2	39.8		743.5	39.8	
	997.8	**40.4**	994.0	40.2	**0.160 /**	998.6	40.3	**0.074 /**	998.5	40.1	**0.372 /**	998.0	40.3	**0.078 /**	991.9	40.3	**0.063 /**
	1499.1	**40.9**	1493.1	40.7	**0.196**	1498.4	40.9	**0.031**	1498.4	40.7	**0.182**	1499.3	40.8	**0.103**	1499.0	40.9	**0.051**
	1998.9	**41.3**	1999.0	41.1		1998.7	41.2		1998.9	41.1		1999.2	41.2		1999.1	41.2	
	2499.5	**41.5**	2499.5	41.3		2499.8	41.5		2499.0	41.4		2499.4	41.4		2499.5	41.5	
Kendo	**488.0**	**39.8**	489.3	39.7		491.3	39.5		491.0	38.6		488.5	39.6		491.8	39.6	
	748.4	**41.6**	725.1	41.1		726.0	41.2		725.9	40.6		749.1	41.3		725.3	41.2	
	998.7	**42.6**	990.1	42.2	**0.363 /**	960.5	42.1	**0.280 /**	959.8	41.7	**0.806 /**	998.8	42.4	**0.199 /**	961.9	42.1	**0.260 /**
	1495.4	**43.8**	1482.0	43.3	**0.454**	1456.0	43.5	**0.167**	1478.2	43.2	**0.497**	1495.5	43.6	**0.132**	1460.4	43.5	**0.172**
	2000.3	**44.4**	1974.7	43.9		1997.0	44.3		1995.5	44.1		2000.2	44.3		1996.2	44.3	
	2500.4	**44.9**	2466.7	44.3		2497.8	44.8		2493.5	44.6		2500.5	44.8		2497.5	44.8	
Newspapercc	499.9	37.9	499.9	38.0		499.6	38.1		500.3	37.9		**499.9**	**38.1**		500.1	38.1	
	752.0	39.5	750.4	39.1		749.9	39.2		750.0	39.0		**751.8**	**39.7**		750.0	39.2	
	1001.3	40.3	1000.7	40.1	**0.301 /**	1000.4	40.1	**0.168 /**	999.7	39.8	**0.400 /**	**1000.9**	**40.4**	**-0.10 /**	999.8	40.1	**0.193 /**
	1501.0	41.5	1500.5	41.0	**0.522**	1500.3	41.3	**0.167**	1501.1	41.2	**0.29**	**1500.8**	**41.4**	**0.088**	1500.7	41.3	**0.243**
	2000.6	**42.4**	2000.6	41.6		2001.0	42.2		2001.7	42.1		2000.5	42.2		2000.2	41.9	
	2501.4	**42.8**	2500.5	42.2		2501.3	42.7		2501.7	42.6		2500.4	42.6		2500.8	42.6	
PoznanHall2	**994.3**	**41.4**	995.1	41.1		996.1	41.3		995.8	41.3		995.4	41.2		995.6	41.3	
	1494.2	**41.7**	1495.2	41.5		1494.7	41.7		1494.9	41.7		1494.8	41.6		1493.8	41.7	
	2000.7	**42.2**	1998.2	41.9	**0.265 /**	1995.1	41.9	**0.160 /**	1995.7	41.9	**0.177 /**	2001.0	42.1	**0.136 /**	1995.2	41.9	**0.162 /**
	3001.6	**42.5**	2998.0	42.2	**0.268**	2999.9	42.4	**0.129**	2998.7	42.3	**0.167**	3001.6	42.4	**0.082**	3000.0	42.4	**0.145**
	4001.7	**42.7**	3999.0	42.5		4001.9	42.7		4001.9	42.6		4002.1	42.6		4000.2	42.6	
	5002.5	**42.8**	5000.1	42.6		5002.0	42.8		5001.9	42.8		5001.8	42.8		5002.1	42.8	
PoznanStreet	997.9	36.6	997.7	36.7		996.6	36.6		998.1	36.2		**998.0**	**36.6**		997.9	36.7	
	1497.9	37.4	1496.7	37.5		1498.3	37.5		1496.6	37.2		**1497.7**	**37.5**		1498.3	37.5	
	2000.2	38.3	1998.4	38.2	**0.016 /**	1997.5	38.1	**0.091 /**	1998.4	37.8	**0.342 /**	**2000.3**	**38.3**	**-0.017**	1997.3	38.1	**0.045 /**
	3000.3	39.1	2997.8	38.9	**0.162**	2998.4	39.0	**0.081**	2999.0	38.8	**0.292**	**3000.3**	**39.0**	**/ 0.059**	2997.8	39.0	**0.087**
	4000.3	**39.6**	3997.4	39.4		4000.4	39.6		4000.7	39.4		4000.4	39.5		3999.3	39.5	
	5000.5	**40.0**	4998.9	39.7		5000.8	39.9		5000.9	39.8		5000.9	39.9		5000.9	39.9	
Average					**0.280 / 0.383**			**0.175 / 0.132**			**0.437 / 0.289**			**0.079 / 0.116**			**0.171 / 0.158**

that between the left and the right NRV pictures. Accordingly, in this paper, only the dependency between RV and NRV picture is considered for simplicity. Thus the RD model of the left and right NRV pictures could be written as

$$\begin{aligned} D_{NRV\_\phi} &= D_{NRV\_\phi,SKIP} + D_{NRV\_\phi,non\_SKIP} \\ &= \eta_\phi \left(D_{RV} + \delta_\phi\right) + \left(a_{NRV\_\phi} R_{NRV\_\phi} + b_{NRV\_\phi}\right)^{-1} \\ &= \eta_\phi \left(a_{RV} R_{RV} + b_{RV}\right)^{-1} \\ &\quad + \left(a_{NRV\_\phi} R_{NRV\_\phi} + b_{NRV\_\phi}\right)^{-1} + \eta_\phi \delta_\phi \end{aligned} \quad (14)$$

where ϕ *denotes "l" and "r" for the left and the right NRV picture respectively*, and thus, D_{NRV_l} and D_{NRV_r} represent the distortions of the left and right NRV pictures, R_{NRV_l} and R_{NRV_r} are the average coding bits of the left and right NRV pictures, η_l and η_r denote the percentage of PUs with SKIP mode (predicted from RV picture, i.e.the middle view picture) of the left and the right NRV picture, δ_l and δ_r are the average squared luminance difference between the original RV and the left NRV pictures and that between the original RV and the right NRV pictures, respectively.

III. PROPOSED INTER-VIEW BIT ALLOCATION METHOD

A. Theoretical Solution

Mathematically, the inter-view bit allocation problem could be written as

$$\begin{aligned} \min \quad & D_{RV} + D_{NRV} \\ s.t. \quad & R_{RV} + R_{NRV} \le R_c \end{aligned} \quad (15)$$

where R_c represents the rate constraint of current frame (including RV and NRV pictures). Based on the analysis of Section II, for 2 view case of MV-HEVC, (15) could be written as

$$\begin{aligned} \min \quad & \left\{ \begin{array}{c} \left(a_{RV} R_{RV} + b_{RV}\right)^{-1} + \eta \left(a_{RV} R_{RV} + b_{RV}\right)^{-1} \\ + \left(a_{NRV} R_{NRV} + b_{NRV}\right)^{-1} + \eta\delta \end{array} \right\} \\ s.t. \quad & R_{RV} + R_{NRV} \le R_c \end{aligned} \quad (16)$$

by taking (5) and (13) into (15). Since the objective function is composed of 2 power functions, i.e., $\left(a_{RV} R_{RV} + b_{RV}\right)^{-1}$

TABLE V
DETAILED RD VALUES COMPARISON FOR 3 VIEW CASE

Sequences	Proposed Method		Bit Ratio 66:17:17		BD PSNR (L/H)	Bit Ratio 44:28:28		BD PSNR (L/H)	Bit Ratio 22:39:39		BD PSNR (L/H)	Method in [17]		BD PSNR (L/H)	Shao's Method [19]		BD PSNR (L/H)
	Total Rate	Avg. PSNR	Total Rate	Avg. PSNR		Total Rate	Avg. PSNR		Total Rate	Avg. PSNR		Total Rate	Avg. PSNR		Total Rate	Avg. PSNR	
Balloons	**500.3**	**38.5**	500.3	38.3		500.6	38.4		500.6	37.4		500.3	38.3		500.6	38.5	
	1003.1	**41.1**	998.5	40.5		999.6	40.8		999.7	40.3		1003.3	41.0		999.3	40.7	
	1501.6	**42.3**	1500.4	41.7	**0.558 /**	1498.3	41.9	**0.324 /**	1499.8	41.6	**0.855 /**	1501.9	42.1	**0.146 /**	1499.4	41.9	**0.326 /**
	1999.7	**43.1**	2000.4	42.3	**0.723**	2000.6	42.8	**0.313**	1999.8	42.5	**0.54**	2000.0	42.9	**0.211**	2000.2	42.7	**0.333**
	2499.4	**43.5**	2500.5	42.8		2498.6	43.2		2499.8	43.1		2499.6	43.3		2499.7	43.2	
	2999.9	**43.9**	2999.9	43.2		3000.8	43.8		2999.1	43.5		2999.7	43.7		3000.9	43.8	
BookArrival	**499.3**	**38.1**	494.7	38.0		496.0	37.8		496.4	36.9		499.8	38.0		495.3	37.9	
	998.8	**39.7**	990.6	39.5		988.6	39.5		997.0	39.0		999.3	39.7		989.2	39.6	
	1499.2	**40.4**	1487.6	40.2	**0.121 /**	1498.0	40.4	**0.120 /**	1494.5	40.0	**0.670 /**	1499.3	40.3	**0.045 /**	1492.8	40.3	**0.061 /**
	1999.3	**40.8**	1987.1	40.6	**0.216**	1998.6	40.8	**0.043**	1997.9	40.5	**0.299**	1999.2	40.7	**0.106**	1998.8	40.8	**0.047**
	2499.1	**41.1**	2499.5	40.9		2499.0	41.1		2498.6	40.9		2499.3	41.0		2499.2	41.0	
	2999.1	**41.3**	2999.7	41.1		2999.7	41.3		2998.3	41.1		2998.9	41.2		2999.6	41.3	
Kendo	**489.4**	**38.4**	492.0	38.3		492.3	37.8		495.0	36.6		490.1	38.2		492.8	37.8	
	998.0	**41.3**	965.7	40.9		968.7	40.9		967.4	39.9		999.1	41.1		968.4	40.9	
	1497.8	**42.6**	1472.5	42.2	**0.280 /**	1439.0	42.1	**0.333 /**	1437.8	41.5	**1.269 /**	1497.6	42.5	**0.152 /**	1440.6	42.1	**0.308 /**
	2000.0	**43.3**	1963.4	42.9	**0.428**	1931.0	43.1	**0.195**	1986.1	42.6	**0.673**	1999.9	43.2	**0.12**	1932.0	43.1	**0.170**
	2500.2	**43.9**	2450.6	43.4		2400.6	43.6		2479.5	43.3		2500.4	43.8		2449.5	43.7	
	3000.4	**44.3**	2936.2	43.7		2996.9	44.2		2975.1	43.7		3000.6	44.2		2996.8	44.2	
Newspapercc	500.1	36.6	500.2	36.6		500.2	36.7		500.5	36.0		**500.3**	**36.8**		500.2	36.8	
	1001.7	39.0	1000.1	38.6		1000.3	38.7		1000.0	38.3		**1001.7**	**39.1**		999.9	38.7	
	1500.7	40.1	1501.5	39.7	**0.382 /**	1499.8	39.9	**0.264 /**	1499.6	39.5	**0.686 /**	**1501.2**	**40.1**	**-0.083**	1500.2	39.9	**0.267 /**
	2000.4	41.1	2000.4	40.4	**0.665**	2000.1	40.7	**0.329**	2001.2	40.5	**0.618**	**2000.5**	**41.0**	**/ 0.108**	2000.7	40.7	**0.315**
	2500.4	**41.6**	2500.2	40.9		2499.3	41.3		2500.8	41.1		2500.7	41.5		2501.0	41.4	
	3001.1	**42.1**	2999.9	41.4		3001.9	41.9		3000.8	41.7		3001.3	42.0		3002.0	41.9	
PoznanHall2	**993.7**	**40.9**	994.5	40.7		995.6	40.8		995.6	40.7		994.8	40.7		995.3	40.8	
	2000.6	**41.9**	1989.9	41.4		1990.7	41.5		1993.4	41.5		2000.7	41.8		1992.2	41.5	
	3001.7	**42.2**	2995.7	41.9	**0.421 /**	2993.6	41.9	**0.279 /**	2995.2	41.9	**0.346 /**	3001.7	42.1	**0.136 /**	2992.4	41.9	**0.277 /**
	4002.5	**42.4**	3997.2	42.1	**0.281**	3998.6	42.2	**0.149**	3995.6	42.1	**0.24**	4002.0	42.3	**0.079**	3999.4	42.2	**0.151**
	5002.4	**42.5**	4997.5	42.3		4999.6	42.4		4999.2	42.4		5002.7	42.4		4999.6	42.4	
	6003.2	**42.6**	5999.6	42.4		6002.5	42.6		5998.4	42.5		6002.8	42.6		6003.2	42.6	
PoznanStreet	995.3	35.8	994.7	36.0		995.2	35.7		998.5	34.9		**994.6**	**36.0**		995.2	35.8	
	2000.1	37.6	1995.0	37.4		1996.4	37.2		1996.7	36.8		**2000.0**	**37.6**		1997.0	37.3	
	3000.4	38.4	2995.4	38.3	**0.104 /**	2995.8	38.1	**0.246 /**	2996.1	37.8	**0.772 /**	**3000.3**	**38.4**	**-0.074**	2994.9	38.1	**0.177 /**
	4000.5	38.9	3997.3	38.8	**0.169**	3997.4	38.8	**0.138**	3997.5	38.4	**0.509**	**4000.4**	**38.9**	**/ 0.046**	3995.7	38.8	**0.128**
	5000.2	**39.4**	4996.2	39.1		4996.2	39.2		4999.8	38.9		4999.9	39.3		4995.8	39.2	
	6000.0	**39.7**	5997.6	39.4		6000.9	39.6		5999.1	39.3		6000.2	39.6		5995.0	39.5	
Average					**0.311 / 0.413**			**0.261 / 0.194**			**0.766 / 0.48**			**0.054/ 0.112**			**0.236 / 0.191**

and $(a_{NRV}R_{NRV}+b_{NRV})^{-1}$ which are convex, (16) could be converted to an unconstraint optimization problem

$$\begin{aligned}\min F = &(1+\eta)\left(a_{RV}R_{RV}+b_{RV}\right)^{-1}\\ &+\left(a_{NRV}R_{NRV}+b_{NRV}\right)^{-1}\\ &+\lambda\left(R_{RV}+R_{NRV}-R_c\right)+\eta\delta\end{aligned} \tag{17}$$

via Lagrangian Multiplier Method [30]. Subsequently, by solving

$$\begin{cases}\frac{\partial F}{\partial R_{RV}} = -a_{RV}\left(1+\eta\right)\left(a_{RV}R_{RV}+b_{RV}\right)^{-2}+\lambda = 0\\ \frac{\partial F}{\partial R_{NRV}} = -a_{NRV}\left(a_{NRV}R_{NRV}+b_{NRV}\right)^{-2}+\lambda = 0\\ R_{RV}+R_{NRV}=R_c\end{cases} \tag{18}$$

the solution of (17) could be obtained as

$$R_{RV} = \frac{R_c+\frac{b_{NRV}}{a_{NRV}}-\frac{b_{RV}}{\sqrt{a_{RV}a_{NRV}(1+\eta)}}}{1+\sqrt{\frac{a_{RV}}{a_{NRV}(1+\eta)}}} \tag{19}$$

$$\begin{aligned}R_{NRV} = &\sqrt{\frac{a_{RV}}{a_{NRV}\left(1+\eta\right)}}R_{RV}\\ &+\frac{b_{RV}}{\sqrt{a_{NRV}a_{RV}\left(1+\eta\right)}}-\frac{b_{NRV}}{a_{NRV}}.\end{aligned} \tag{20}$$

Similarly, for 3 view case, the inter-view bit allocation problem could be formulated as

$$\begin{aligned}\min &\left\{\begin{array}{l}\left(1+\eta_l+\eta_r\right)\left(a_{RV}R_{RV}+b_{RV}\right)^{-1}\\ \quad+\left(a_{NRV\_l}R_{NRV\_l}+b_{NRV\_l}\right)^{-1}\\ \quad+\left(a_{NRV\_r}R_{NRV\_r}+b_{NRV\_r}\right)^{-1}\end{array}\right\}+\eta_l\delta_l+\eta_r\delta_r\\ s.t.\quad &R_{RV}+R_{NRV\_l}+R_{NRV\_r}\le R_c,\end{aligned} \tag{21}$$

and could be solved as

$$\begin{aligned}R_{RV} = &\begin{pmatrix}R_c-\sqrt{\frac{{b_{RV}}^2}{a_{NRV\_l}a_{RV}(1+\eta_l+\eta_r)}}\\ -\sqrt{\frac{{b_{RV}}^2}{a_{NRV\_r}a_{RV}(1+\eta_l+\eta_r)}}+\frac{b_{NRV\_l}}{a_{NRV\_l}}+\frac{b_{NRV\_r}}{a_{NRV\_r}}\end{pmatrix}\\ &\Big/\left(1+\sqrt{\frac{a_{RV}}{a_{NRV\_l}\left(1+\eta_l+\eta_r\right)}}\right.\\ &\left.+\sqrt{\frac{a_{RV}}{a_{NRV\_r}\left(1+\eta_l+\eta_r\right)}}\right)\end{aligned} \tag{22}$$

$$\begin{aligned}R_{NRV\_l} = &\sqrt{\frac{a_{RV}}{a_{NRV\_l}\left(1+\eta_l+\eta_r\right)}}\cdot R_{RV}\\ &+\sqrt{\frac{{b_{RV}}^2}{a_{NRV\_l}a_{RV}\left(1+\eta_l+\eta_r\right)}}-\frac{b_{NRV\_l}}{a_{NRV\_l}}\end{aligned} \tag{23}$$

TABLE VI
INTER-VIEW QUALITY CONSISTENCY COMPARISON

Sequences	*IVQC2view*						*IVQC3view*					
	Proposed Method	Bit Ratio 8:2	Bit Ratio 6:4	Bit Ratio 4:6	Method in [17]	Shao's Method [19]	Proposed Method	Bit Ratio 66:17:17	Bit Ratio 44:28:28	Bit Ratio 22:39:39	Method in [17]	Shao's Method [19]
Balloons	1.65	3.71	1.29	**1.18**	2.90	1.67	2.41	3.80	**1.40**	2.77	3.42	1.59
	0.91	3.50	1.23	0.86	2.48	1.65	1.51	3.56	1.48	**1.38**	2.68	1.69
	0.70	3.54	1.15	0.75	2.23	1.50	1.19	3.49	1.41	**1.02**	2.24	1.56
	0.64	3.19	1.25	0.96	1.67	1.56	**1.08**	3.24	1.58	1.20	1.82	1.74
	0.39	2.90	0.76	0.59	1.19	1.05	**0.99**	3.03	1.50	1.08	1.60	1.63
	0.28	2.70	0.68	0.56	1.00	0.96	**0.90**	2.89	1.10	0.96	1.44	1.22
BookArrival	0.59	1.69	**0.26**	1.34	1.17	0.92	0.90	1.74	**0.29**	2.45	1.32	0.72
	0.44	1.70	0.45	1.03	1.10	1.04	0.80	1.74	**0.54**	1.59	1.22	0.89
	0.51	1.66	**0.45**	0.71	1.01	1.05	0.64	1.67	**0.53**	1.20	1.00	0.86
	0.29	1.63	0.54	0.40	0.78	0.97	**0.51**	1.62	0.55	0.89	0.81	0.80
	0.27	1.49	0.57	0.26	0.68	0.99	**0.44**	1.46	0.56	0.69	0.73	0.78
	0.26	1.49	0.61	0.23	0.66	1.01	**0.43**	1.45	0.57	0.58	0.76	0.80
Kendo	1.41	2.33	**0.27**	3.63	2.03	0.08	1.94	2.35	0.52	5.38	2.31	**0.34**
	0.83	2.37	**0.19**	2.43	1.88	0.51	1.41	2.58	**0.52**	3.61	2.16	0.63
	0.52	2.59	**0.25**	1.84	1.29	0.55	0.95	2.86	**0.69**	2.43	1.54	0.79
	0.36	2.57	0.70	1.82	0.94	0.95	**0.88**	2.82	1.08	2.15	1.28	1.18
	0.21	2.45	0.42	1.25	0.71	0.64	**0.77**	2.73	1.16	1.83	1.16	1.06
	0.13	2.39	0.42	1.05	0.60	0.63	**0.71**	2.68	0.84	1.70	1.06	0.91
Newspapercc	**1.07**	3.05	1.20	0.59	2.34	1.69	1.86	3.49	1.66	**1.06**	3.07	1.90
	0.63	3.29	1.14	**0.62**	2.43	1.68	1.36	3.63	1.58	**0.66**	2.91	1.83
	0.53	3.60	1.29	0.67	2.32	1.82	**1.21**	3.88	1.71	0.75	2.59	1.96
	0.64	3.49	1.57	0.86	1.91	2.09	1.41	3.80	1.98	**1.09**	2.45	2.24
	0.57	3.47	1.14	**0.52**	1.89	2.03	1.28	3.74	1.92	**1.15**	2.27	2.03
	0.48	3.60	1.09	0.56	2.05	1.62	1.20	3.85	1.53	**0.97**	2.09	1.77
PoznanHall2	0.37	1.17	**0.32**	0.46	0.71	0.39	0.61	1.25	**0.45**	0.70	0.96	0.46
	0.19	1.03	0.24	0.40	0.54	0.31	0.38	1.06	0.36	0.53	0.59	**0.35**
	0.07	1.12	0.24	0.43	0.40	0.29	**0.27**	1.15	0.32	0.52	0.51	0.32
	0.02	1.05	0.36	0.55	0.30	0.43	**0.23**	1.13	0.49	0.48	0.45	0.49
	0.00	1.06	0.21	0.31	0.25	0.40	**0.27**	1.12	0.46	0.59	0.47	0.47
	0.04	1.09	0.24	0.34	0.28	0.29	**0.26**	1.15	**0.35**	0.58	0.44	0.35
PoznanStreet	0.49	1.42	**0.29**	1.18	1.02	0.73	0.81	1.47	**0.32**	2.27	1.35	0.75
	0.43	1.52	**0.36**	0.95	0.87	0.80	0.42	1.61	0.53	1.29	0.94	0.88
	0.04	1.71	0.39	0.92	0.86	0.87	**0.36**	1.80	0.53	1.16	0.87	0.92
	0.05	1.78	0.59	1.02	0.78	1.04	**0.34**	1.82	0.70	1.15	0.81	1.07
	0.14	1.82	0.46	0.69	0.72	1.05	**0.49**	1.84	0.71	1.28	0.87	1.07
	0.21	1.85	0.50	0.63	0.71	0.93	**0.46**	1.85	0.55	1.23	0.82	1.07
Average	**0.41**	2.17	0.61	0.90	1.15	0.97	**0.82**	2.29	0.87	1.36	1.38	1.05

$$R_{NRV\_r} = \sqrt{\frac{a_{RV}}{a_{NRV\_r}\left(1+\eta_l+\eta_r\right)}}R_{RV} + \sqrt{\frac{{b_{RV}}^2}{a_{NRV\_r}a_{RV}\left(1+\eta_l+\eta_r\right)}} - \frac{b_{NRV\_r}}{a_{NRV\_r}}. \quad (24)$$

B. Technical Implementation

Based on the theoretical solution, in order to allocate bits for RV and NRV pictures optimally, parameters a_{RV}, b_{RV}, a_{NRV}, b_{NRV}, and η should be obtained first for 2 view case; while for 3 view case, a_{RV}, b_{RV}, a_{NRV_l}, b_{NRV_l}, a_{NRV_r}, b_{NRV_r}, η_l and η_r should be obtained first similarly. In order to implement the proposed bit allocation method by one pass coding, those parameters are calculated based on the statistics of previous encoded frames.

During the encoding procedure, η (or η_l and η_r) could be obtained by using the percentage of SKIP PUs (predicted from RV picture) in the nearest previous frame which has the same TL of the current frame. For parameters a_{RV}, b_{RV}, a_{NRV}, and b_{NRV}, (or a_{NRV_l}, b_{NRV_l}, a_{NRV_r}, and b_{NRV_r}), a sliding window with a length of W is designed to store the rates and distortions of those W previous encoded frames with the same TL of the current frame. Then, for the RV picture of the current frame, a_{RV} and b_{RV} could be calculated by solving the linear regression

$$\left(D_{RV,0}^{-1}\cdots D_{RV,M}^{-1}\right)^T = a_{RV}(R_{RV,0}\cdots R_{RV,M})^T + b_{RV} \quad (25)$$

where $M \in [2, W]$ is the number of data points used in the regression. For the NRV pictures of the current frame, a_{NRV} and b_{NRV} could be calculated by solving (26)

$$\left(\left(D_{NRV,0}-\eta_0 D_{RV0}\right)^{-1}\cdots\left(D_{NRV,M}-\eta_M D_{RV,M}\right)^{-1}\right)^T = a_{NRV}\left(R_{NRV,0}\cdots R_{NRV,i}\cdots R_{NRV,M}\right)^T + b_{NRV} \quad (26)$$

when there are only 2 view pictures. When 3 view pictures should be encoded, parameters a_{NRV_l}, b_{NRV_l}, a_{NRV_r}, and b_{NRV_r} could be calculated by solving (27)

$$\left(\left(D_{NRV\_\phi,0}-\eta_{\phi,0}D_{RV\_\phi,0}\right)^{-1} \cdots\left(D_{NRV\_\phi,M}-\eta_{\phi,M}D_{RV\_\phi,M}\right)^{-1}\right)^T = a_{NRV\_\phi}\left(R_{NRV\_\phi,0}\cdots R_{NRV\_\phi,M}\right)^T + b_{NRV\_\phi} \quad (27)$$

where ϕ denotes "l" and "r" for the left and the right NRV picture respectively. In this work, W was set as 49 empirically.

C. Optimal Initial Bit Ratio

In order to allocate bits for different view pictures, model parameters should be obtained before encoding. Because the model parameters are updated frame by frame, they cannot be obtained for the first several frames. Therefore, we set fixed bit ratio for the first several frames to start the encoding procedure. Since the bit ratio of the first several frames could affect the overall rate-distortion performance of a video sequence, we implemented a statistical experiment so as to obtain the optimal bit ratio for the first several frames.

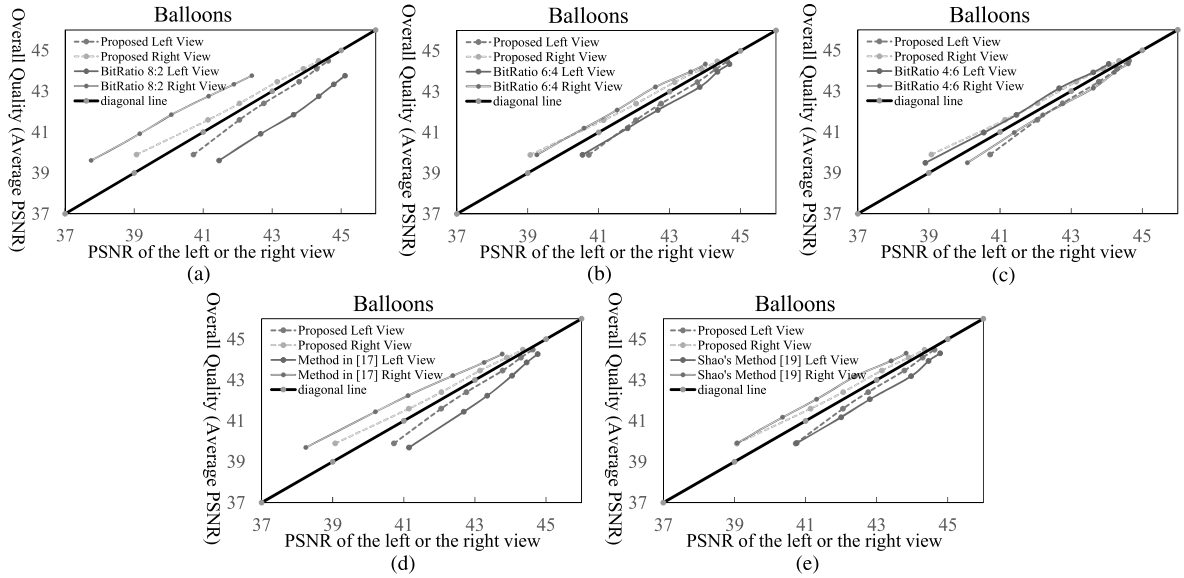


Fig. 10. Relationship between the overall quality and the distortion of each view of the *Balloon* sequence, 2 view case.

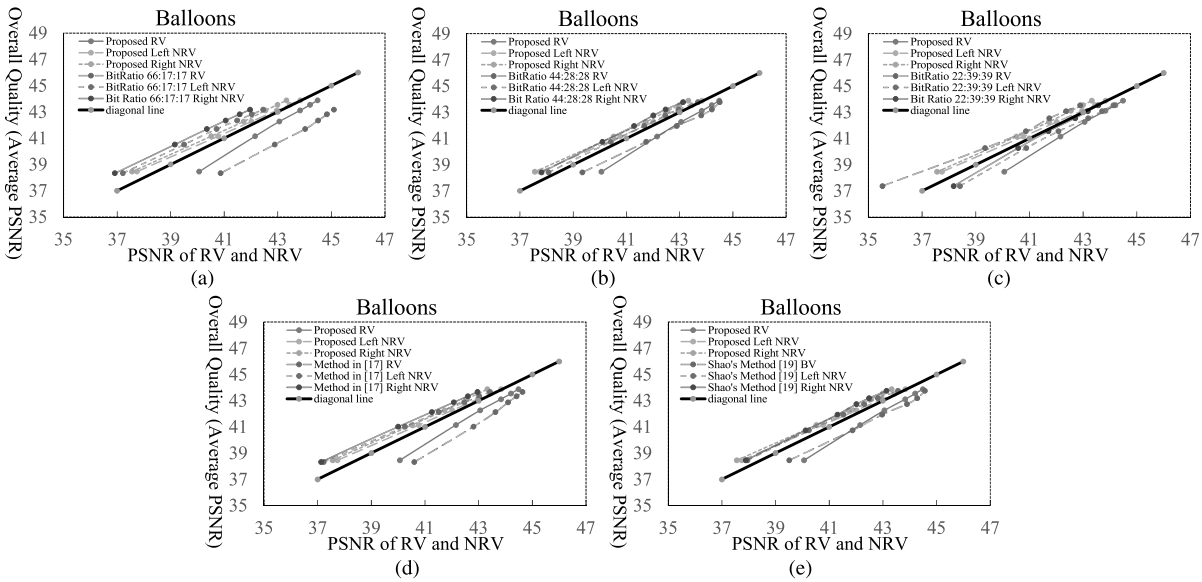


Fig. 11. Relationship between the overall quality and the distortion of each view of the *Balloon* sequence, 3 view case.

In the experiment, we vary initial bit ratio (i.e. 0.1:0.9, 0.2:0.8, …, and 0.9:0.1 for 2 view case, and 0.1:0.45:0.45, 0.2:0.4:0.4, …, and 0.9:0.05:0.05 for 3 view case) for the first 16 frames; while fixing the bit ratio for the remaining frames as 0.5:0.5 and 0.4:0.3:0.3 for 2 view and 3 view cases respectively. The total rate constraints are set as 500 kbps, 1000 kbps, 1500 kbps, 2000 kbps, and 2500 kbps. Then, the RD performances of different initial bit ratios are compared as shown in Fig. 6 and Fig. 7 (only the bit ratios of RV picture are labeled in the figures), from which we can observe that 0.4:0.6 and 30:35:35 are the quasi optimal initial bit ratios for 2 view and 3 view case respectively.

IV. Experimental Results

To evaluate the performance of the proposed method, experiments were divided into 2 parts, i.e. bit allocation performance comparison, and complexity comparison. 3DV sequences [31] which are adopted by JCT-3 V, i.e., *Balloons* (view 3 was encoded as RV, view 1 and 5 were encoded as NRV), *BookArrival* (view 8 was encoded as RV, view 10 and 6 were encoded as NRV), *Kendo* (view 3 was encoded as RV, view 1 and 5 were encoded as NRV), *Newspapercc* (view 4 was encoded as RV, view 2 and 6 were encoded as NRV), *PoznanHall2* (view 6 was encoded as RV, view 5 and 7 were encoded as NRV), and *PoznanStreet* (view 4 was encoded as RV, view 3 and 5 were encoded as NRV), were used. The MV-HEVC platform, 3D-HTM version 9.2 [32] was employed to encode those videos. The encoder configuration files were set as the same with that of common test condition [6] of MV-HEVC. Detailed encoder parameters are shown shown in Table III. Besides, the experiments were implemented on a computer with Intel Core i5-2300 CPU (2.8 GHz), and 4 GB memory size.

TABLE VII
COMPLEXITY COMPARISONS FOR 2 VIEW CASE

Sequences	Encoding Time (s)										
	Proposed Method	Bit Ratio 8:2	*CQ*	Bit Ratio 6:4	*CQ*	Bit Ratio 4:6	*CQ*	Method in [17]	*CQ*	Shao's Method [19]	*CQ*
Balloons	2002.08	1924.16	1.04	1969.92	1.02	1970.49	1.02	**1923.94**	1.04	2015.71	0.99
	2075.05	**2026.77**	1.02	2095.73	0.99	2113.11	0.98	2064.21	1.01	2115.20	0.98
	2178.14	**2117.10**	1.03	2178.85	1.00	2203.59	0.99	2156.70	1.01	2225.82	0.98
	2357.38	**2279.47**	1.03	2335.21	1.01	2365.58	1.00	2380.47	0.99	2405.38	0.98
	2452.89	**2417.28**	1.01	2501.20	0.98	2515.98	0.97	2484.59	0.99	2589.40	0.95
	2592.39	**2499.98**	1.04	2627.40	0.99	2654.25	0.98	2632.68	0.98	2703.44	0.96
BookArrival	1971.58	**1941.13**	1.02	1987.97	0.99	1981.46	1.00	1976.77	1.00	2027.54	0.97
	2120.42	**2068.47**	1.03	2101.64	1.01	2121.95	1.00	2136.77	0.99	2166.31	0.98
	2220.23	**2173.16**	1.02	2229.71	1.00	2225.52	1.00	2218.76	1.00	2294.62	0.97
	2395.34	**2334.13**	1.03	2402.85	1.00	2405.86	1.00	2452.01	0.98	2434.03	0.98
	2533.13	**2473.63**	1.02	2544.99	1.00	2541.61	1.00	2880.59	0.88	2538.41	1.00
	2653.73	**2592.84**	1.02	2658.15	1.00	2677.87	0.99	3029.57	0.88	2652.06	1.00
Kendo	2051.55	**2024.45**	1.01	2087.28	0.98	2085.68	0.98	2077.30	0.99	2077.38	0.99
	2220.43	**2140.14**	1.04	2209.42	1.00	2229.07	1.00	2176.53	1.02	2214.80	1.00
	2316.82	**2275.03**	1.02	2304.77	1.01	2318.29	1.00	2318.03	1.00	2305.93	1.00
	2519.26	**2415.84**	1.04	2473.92	1.02	2506.81	1.00	2502.36	1.01	2475.78	1.02
	2654.78	**2536.45**	1.05	2634.52	1.01	2661.70	1.00	2635.11	1.01	2632.07	1.01
	2750.97	**2629.00**	1.05	2742.11	1.00	2754.82	1.00	2793.81	0.98	2734.62	1.01
Newspapercc	1832.86	**1787.78**	1.03	1809.86	1.01	1824.21	1.00	1807.05	1.01	1810.14	1.01
	1950.50	**1848.03**	1.06	1897.64	1.03	1922.57	1.01	1890.22	1.03	1886.26	1.03
	2020.88	**1917.89**	1.05	1947.76	1.04	1972.46	1.02	1997.87	1.01	1943.32	1.04
	2120.82	**2035.25**	1.04	2072.84	1.02	2123.23	1.00	2117.11	1.00	2065.99	1.03
	2203.40	**2140.34**	1.03	2215.47	0.99	2232.51	0.99	2213.51	1.00	2174.06	1.01
	2315.13	**2210.03**	1.05	2323.75	1.00	2334.86	0.99	2323.60	1.00	2300.52	1.01
PoznanHall2	5242.84	5876.26	0.89	5327.87	0.98	5300.39	0.99	**5206.95**	1.01	5574.93	0.94
	5510.36	5418.81	1.02	5622.04	0.98	5579.81	0.99	5484.24	1.00	**5337.81**	1.03
	5779.06	**5653.94**	1.02	5782.52	1.00	5744.05	1.01	5813.25	0.99	6141.17	0.94
	6147.56	6139.68	1.00	6150.69	1.00	**6136.77**	1.00	6297.47	0.98	6477.61	0.95
	6467.66	6512.54	0.99	6489.42	1.00	**6372.89**	1.01	6839.15	0.95	6760.17	0.96
	6781.14	6891.66	0.98	6746.92	1.01	6675.75	1.02	7061.74	0.96	**6667.03**	1.02
PoznanStreet	4700.19	**4548.33**	1.03	4750.58	0.99	4906.83	0.96	4917.50	0.96	4660.71	1.01
	4923.93	**4746.98**	1.04	5045.45	0.98	5124.01	0.96	4913.96	1.00	4905.16	1.00
	5095.14	**4897.01**	1.04	5223.81	0.98	5248.89	0.97	5078.38	1.00	5037.34	1.01
	5374.66	**5258.08**	1.02	5453.12	0.99	5441.52	0.99	5373.74	1.00	5626.08	0.96
	5574.42	**5504.73**	1.01	5654.34	0.99	5664.28	0.98	5605.82	0.99	5975.04	0.93
	5801.68	**5793.24**	1.00	5906.02	0.98	5816.71	1.00	5822.67	1.00	6169.14	0.94
	Average		**1.02**		**1.00**		**0.99**		**0.99**		**0.99**

A. RD Performance Comparison

To verify the RD performance of the proposed method, we compared the RD curves of the proposed method with those of view-wise fixed bit ratio method [21] (in which we predefined several different bit ratios, i.e. bit ratio 8:2, 6:4, 4:6 for 2 view case and bit ratio 66:17:17, 44:28:28, 22:39:39 for 3 view case to cover all the possible bit ratios), the method in [16], and the step 2 (interview allocation for texture videos) of *Shao's* method [18]. In 2 view case, for sequences with 1024 × 768 resolution (*Balloons*, *BookArrival*, *Kendo*, and *Newspapercc*), the target bit rates were set as 500 kbps, 750 kbps, 1000 kbps, 1500 kbps, 2000 kbps, and 2500 kbps; while for sequences with 1920 × 1088 resolution (*PoznanHall2 and PoznanStreet*), the target bit rates were set as 1000 kbps, 1500 kbps, 2000 kbps, 3000 kbps, 4000 kbps, and 5000 kbps. For 3 view case, the target bit rates were set as 500 kbps, 1000 kbps, 1500 kbps, 2000 kbps, 2500 kbps, and 3000 kbps for sequences with 1024 × 768 resolution; and as 1000 kbps, 2000 kbps, 3000 kbps, 4000 kbps, 5000 kbps and 6000 kbps for sequences with 1920 × 1088 resolution. The RD curves for 2 view and 3 view cases are compared in Fig. 8 and Fig. 9 respectively, where we can observe that the proposed method is the best.

Besides, Table IV shows the detailed values of coding bit rates (kbps) and reconstructed qualities (evaluated by PSNR) for all of the tested methods for 2 view case. In Table IV, in order to compare RD performance, BD PSNR [33] is employed. The BD PSNR means the numerical average reduction (negative value) or increment (positive value) of the PSNR of proposed method compared with that of the reference method at the same coding bit rate. Actually, the BD PSNR is obtained by calculated the area of the gap between two Rate-PSNR curves [34]. We can observe that when compared with bit ratio 8:2, bit ratio 6:4, bit ratio 4:6 methods, the method in [16], and *Shao's* method [18], 0.383 dB, 0.132 dB, 0.289 dB, 0.116 dB, and 0.158 dB BD PSNR could be achieved for high bit rate (from 1000 kbps to 2500 kbps for 1024 × 768 resolution, from 2000 kbps to 5000 kbps for 1920 × 1088 resolution, denoted as **H** in Table IV), and 0.280 dB, 0.175 dB, 0.437 dB, 0.079 dB, and 0.171 dB BD PSNR could be achieved for low bit rate (from 500 kbps to 1500 kbps for 1024 × 768 resolution, from 1000 kbps to 3000 kbps for 1920 × 1088 resolution,, denoted as **L** in Table IV). For 3 view case, the detailed results are shown in Table V. It can be observed that when the bit rate is high (from 1500 kbps to 3000 kbps for 1024 × 768 resolution, from 3000 kbps to 6000 kbps for 1920 × 1088 resolution), the BD PSNR gains are 0.413 dB, 0.194 dB, 0.480 dB, 0.112 dB, and 0.191 dB by comparing the proposed method with bit ratio 66:17:17, bit ratio 44:28:28, bit ratio 33:39:39, the method in [16] and *Shao's* method [18]. When the bit rate is low (from 500 kbps to 2000 kbps for 1024 × 768 resolution, from 1000 kbps to 4000 kbps for 1920 × 1088 resolution), the BD PSNR gains are 0.311 dB, 0.261 dB, 0.766 dB, 0.054 dB, and 0.236 dB respectively.

Moreover, we also evaluated the inter-view quality consistency by comparing the proposed method with the other methods. For 2 view case, we use the difference of average reconstructed qualities between the 2 views to represent the inter-view quality consistency ($IVQC_{2view}$)

$$IVQC_{2view} = |PSNR_{RV} - PSNR_{NRV}|. \quad (28)$$

TABLE VIII
COMPLEXITY COMPARISONS FOR 3 VIEW CASE

Sequences	Encoding Time (s)										
	Proposed Method	Bit Ratio 66:17:17	*CQ*	Bit Ratio 44:28:28	*CQ*	Bit Ratio 22:39:39	*CQ*	Method in [17]	*CQ*	Shao's Method [19]	*CQ*
Balloons	2960.99	2937.18	1.01	2978.36	0.99	3014.11	0.98	**2899.39**	1.02	3423.23	0.86
	3242.51	**3225.29**	1.01	3282.86	0.99	3316.48	0.98	3249.39	1.00	3526.43	0.92
	3486.46	**3478.37**	1.00	3478.06	1.00	3527.29	0.99	3496.93	1.00	3571.04	0.98
	3615.41	3679.60	0.98	3638.00	0.99	3766.02	0.96	3649.70	0.99	3716.55	0.97
	3790.51	**3722.70**	1.02	3794.04	1.00	3910.52	0.97	3829.01	0.99	3868.10	0.98
	3925.15	**3803.59**	1.03	4016.96	0.98	4038.81	0.97	3966.88	0.99	4154.64	0.94
BookArrival	2982.29	**2962.66**	1.01	3000.00	0.99	3029.98	0.98	2976.67	1.00	3094.96	0.96
	3307.70	**3297.45**	1.00	3321.97	1.00	3379.31	0.98	3333.67	0.99	3440.63	0.96
	3564.14	**3497.44**	1.02	3572.97	1.00	3567.55	1.00	3604.74	0.99	3670.71	0.97
	3744.84	**3691.48**	1.01	3749.61	1.00	3765.82	0.99	3787.88	0.99	4005.09	0.94
	3977.11	**3861.67**	1.03	3909.29	1.02	3910.16	1.02	3942.97	1.01	3969.75	1.00
	4153.39	**4001.59**	1.04	4042.81	1.03	4064.75	1.02	4127.13	1.01	4124.39	1.01
Kendo	3145.23	3055.11	1.03	3087.13	1.02	3106.91	1.01	**3044.25**	1.03	3109.69	1.01
	3538.59	**3390.33**	1.04	3451.88	1.03	3482.94	1.02	3398.54	1.04	3515.06	1.01
	3702.11	**3630.04**	1.02	3722.44	0.99	3701.59	1.00	3656.51	1.01	3729.06	0.99
	3865.90	**3830.18**	1.01	3938.89	0.98	3887.26	0.99	3842.90	1.01	3874.24	1.00
	4026.69	**4005.01**	1.01	4070.27	0.99	4065.63	0.99	4039.52	1.00	4075.76	0.99
	4155.57	**4078.25**	1.02	4264.13	0.97	4166.55	1.00	4179.04	0.99	4237.81	0.98
Newspapercc	2773.30	**2740.35**	1.01	2858.87	0.97	2806.37	0.99	2761.01	1.00	2814.42	0.99
	3041.44	**2951.95**	1.03	3476.01	0.87	3014.70	1.01	2998.38	1.01	3015.12	1.01
	3216.71	**3092.22**	1.04	3486.27	0.92	3143.74	1.02	3179.10	1.01	3157.85	1.02
	3306.37	**3248.03**	1.02	3325.50	0.99	3373.11	0.98	3320.58	1.00	3337.56	0.99
	3462.96	**3365.83**	1.03	3460.12	1.00	3445.20	1.01	3485.40	0.99	3543.66	0.98
	3571.21	**3453.64**	1.03	3667.31	0.97	3549.51	1.01	3551.76	1.01	3716.47	0.96
PoznanHall2	8084.65	7972.16	1.01	8093.15	1.00	8244.51	0.98	8025.61	1.01	**7946.21**	1.02
	8749.43	**8666.61**	1.01	8931.09	0.98	8883.37	0.98	8687.21	1.01	8796.39	0.99
	9232.94	9122.84	1.01	9740.91	0.95	9270.78	1.00	**9119.29**	1.01	9782.02	0.94
	9652.64	**9382.23**	1.03	9976.00	0.97	9695.82	1.00	9436.95	1.02	10088.62	0.96
	10079.77	**9734.40**	1.04	9853.29	1.02	9961.59	1.01	9744.54	1.03	10633.97	0.95
	10548.35	**10055.86**	1.05	10229.98	1.03	10186.47	1.04	10114.58	1.04	10923.57	0.97
PoznanStreet	7381.88	7515.01	0.98	7732.35	0.95	**7240.38**	1.02	7453.42	0.99	7897.35	0.93
	7899.89	**7768.00**	1.02	8254.27	0.96	7833.64	1.01	7936.18	1.00	8379.31	0.94
	8316.25	8109.44	1.03	8556.62	0.97	**8038.63**	1.03	8374.50	0.99	8194.28	1.01
	8587.80	**8380.97**	1.02	**8517.89**	1.01	8619.30	1.00	8657.27	0.99	8442.26	1.02
	8946.48	8660.15	1.03	8734.04	1.02	8908.47	1.00	8889.92	1.01	**8571.22**	1.04
	9097.68	8869.35	1.03	9019.89	1.01	9069.82	1.00	9215.40	0.99	**8719.20**	1.04
Average			**1.02**		**0.99**		**1.00**		**1.00**		**0.98**

For 3 view case, the inter-view quality consistency ($IVQC_{3view}$) is defined as shown in (29), at the bottom of the page.

The inter-view consistency comparisons for 2 and 3 view cases are shown in Table VI. We can observe that, the average $IVQC_{2view}$ and $IVQC_{3view}$ of the proposed method are only 0.41 dB and 0.82 dB which are considerably smaller than other methods.

Furthermore, we have also given the relationship between the overall quality and the distortion of each view of *Balloon* sequence so as to analyze the interview quality consistency, as shown in Fig. 10 and Fig. 11. From the two figures, it can be observed that both the qualities of the left and the right view for 2 view case (or RV and NRVs for 3 view case) are closer to the overall quality (closer to the diagonal line) compared with other methods in most cases.

B. Complexity Comparison

In the experiments, the encoding time (T) was used to evaluate complexity. The encoding time of the proposed method was used as benchmark. The ratio between the encoding times of other methods and that of the proposed method is employed to represent the complexity quotient (CQ)

$$CQ = T_{proposed}/T_{other\_method} \tag{30}$$

where, T_{other_method} denotes the encoding time of one of the reference methods, $T_{proposed}$ denotes the encoding time of the proposed method. If CQ is larger than 1, the complexity of the proposed method (benchmark) is larger than the compared method; if CQ is smaller than 1, the complexity of the proposed method is smaller than the compared method. The complexity comparisons are shown in Table VII and Table VIII, from which we can observe that the *CQs* are close to 1, which means that the complexity of the proposed method is comparable with the existing methods. The reason why the complexity of the proposed method is a little higher or lower than the other methods is because of the random fluctuation of CPUs. Despite of the random fluctuation of CPUs, from Table VII and Table VIII, it can also be observed that the complexity of the proposed method is only a little larger than that of the *Bit Ratio 8:2* and *Bit Ratio 66:17:17*

$$IVQC_{3view} = \frac{(|PSNR_{RV} - PSNR_{NRV\_l}| + |PSNR_{RV} - PSNR_{NRV\_r}|)}{2} \tag{29}$$

method for most of the test sequences and coding bit rates. Compared with other method, i.e., *Bit Ratio 6:4*, *Bit Ratio 44:28:28*, *Bit Ratio 4:6*, *Bit Ratio 22:39:39*, and *Shao's method*[18], the complexity of the proposed method is the least for almost all the test sequences and coding bit rates.

V. Conclusion

Inter-view dependency is analyzed in detail in this paper. Based on the analysis, the distortion propagation between RV picture and NRV pictures is mainly caused by the percentage of PUs with SKIP mode. Then, RD models for NRV pictures are derived. Subsequently, by employing the RD models, the inter-view bit allocation problem is formulated as a convex optimization problem, and was solved by Lagrangian Multiplier Method. Experimental results demonstrated that the RD performance of the proposed method is better than the existing method. Besides, the inter-view quality consistency of the proposed method is also better than the existing methods, and the complexity of the proposed method is comparable with other existing methods.

Acknowledgment

The authors would like to thank the editors and anonymous reviewers for their valuable comments. The authors would also like to thank the JCT-3 V Group for providing their 3D video sequences and their valuable work on 3DV.

References

[1] Y. Chen, Y.-K. Wang, K. Ugur, M. M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP J. Appl. Signal Process.*, vol. 2009, pp. 8:1–8:13, Jan. 2008.

[2] A. Puri, X. Chen, and A. Luthra, "Video coding using the H.264/MPEG-4 AVC compression standard," *Signal Process. Image Commun.*, vol. 19, no. 9, pp. 793–849, Jun. 2004.

[3] P. Merkle, A. Smolic, K. Müller, 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.

[4] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[5] G. J. Sullivan, J. M. Boyce, Y. Chen, J. R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of high efficiency video coding (HEVC)," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1001–1016, Dec. 2013.

[6] D. Rusanovskyy, K. Müller, and A. Vetro, "Common test conditions of 3DV core experiments," in *Proc. 3rd Meeting ITU-T/ISO/IEC Joint Collaborative Team 3D Video Coding (JCT-3 V)*, Jan. 2013, Doc. JCT3 V-C1100.

[7] K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3366–3378, Sep. 2013.

[8] Q. Wang, X. Ji, Q. Dai, and N. Zhang, "Free viewpoint video coding with rate-distortion analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 6, pp. 875–889, Jun. 2012.

[9] G. Cheung, V. Velisavljević, and A. Ortega, "On dependent bit allocation for multiview image coding with depth-image-based rendering," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3179–3194, Nov. 2011.

[10] 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, Nov. 2011.

[11] G. Tech, K. Wegner, Y. Chen, and M. Hannuksela, "MV-HEVC Draft Text 5," in *Proc. 5th Meeting ITU-T/ISO/IEC Joint Collaborative Team 3D Video Coding (JCT-3 V)*, Jul. 2013, Doc. JCT3 V-E1004.

[12] Y. Chen, L. Zhang, V. Seregin, and Y.-K. Wang, "Motion hooks for the multiview extension of HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 12, pp. 2090–2098, Dec. 2014.

[13] D. B. Sansli, K. Ugurt, M. M. Hannukselat, and M. Gabbouj, "Inter view motion vector prediction in multiview HEVC," in *Proc. 3DTV-Conf.: True Vis.—Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jul. 2014, pp. 1–4.

[14] H. Schwarz and T. Wiegand, "Inter view prediction of motion data in multiview video coding," in *Proc. Picture Coding Symp.*, May 2012, pp. 101–104.

[15] W. Yao, L. P. Chau, and S. Rahardja, "Joint rate allocation of stereoscopic 3D videos in next-generation broadcast applications," *IEEE Trans. Broadcast.*, vol. 59, no. 3, pp. 445–454, Sep. 2013.

[16] Y. Chang and M. Kim, "A joint rate control scheme in a hybrid stereoscopic video codec system for 3DTV broadcasting," *IEEE Trans. Broadcast.*, vol. 59, no. 2, pp. 265–280, Jun. 2013.

[17] 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. 562–572, Jun. 2011.

[18] F. Shao, G. Jiang, W. Lin, M. Yu, and Q. Dai, "Joint bit allocation and rate control for coding multi-view video plus depth based 3D video," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1843–1854, Dec. 2013.

[19] F. Shao, G. Jiang, and M. Yu *et al.*, "Asymmetric coding of multi-view video plus depth based 3D video for view rendering," *IEEE Trans. Multimedia*, vol. 14, no. 1, Feb. 2012.

[20] F. Shao *et al.*, "Depth map coding for view synthesis based on distortion analyses," *IEEE J. Emerging Sel. Topics Circuits Syst.*, vol. 4, no. 1, pp. 106–117, Mar. 2014.

[21] W. Lim, H. Jo, J. Yoo, D. Sim, and I. V. Bajić, "Inter-view MAD prediction for rate control of 3D multi-view video coding," in *Proc. 5th Meeting ITU-T/ISO/IEC Joint Collaborative Team 3D Video Coding (JCT-3 V)*, Aug. 2013, Doc. JCT3 V-E0227.

[22] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2001, vol. 3, pp. 542–545.

[23] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[24] H. J. Lee, T. Chiang, and Y. Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878–894, Sep. 2000.

[25] R. C. Reininger and J. D. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. Commun.*, vol. COM-31, no. 6, pp. 835–839, Jun. 1983.

[26] N. Kamaci, Y. Altunbasak, 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.

[27] M. Flierl, A. Mavlankar, and B. Girod, "Motion and disparity compensated coding for multiview video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1474–1484, Nov. 2007.

[28] Q. Xu, X. Lu, Y. Liu, and C. Gomila, "A fine rate control algorithm with adaptive rounding offsets (ARO)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 10, pp. 1424–1435, Nov. 2009.

[29] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*, 6th ed. Cambridge, U.K.: Cambridge Univ. Press, 2008, pp. 215–227.

[31] Joint Collaborative Team for 3DV, *3D Video Test Sequences* Oct. 2014 [Online]. Available: ftp.hhi.fraunhofer.de, Accessed on: Oct. 30, 2014

[32] Joint Collaborative Team for 3DV, , Oct. 2014, *3D-HTM Software Platform* [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/ Accessed on: Oct. 30, 2014

[33] G. Bjontegaard and Tandberg, "Improvements of the BD-PSNR model," in *Proc. 35th Meeting ITU-T Video Coding Experts Group*, Jul. 2008, Doc. AI11.

[34] *Calculation of Average PSNR Differences Between RD Curves*, ITU-T SG16/Q6 (VCEG), doc. VCEG-M33, Apr. 2001.

Hui Yuan (S'08–M'12) received the B.E. and Ph.D. degree in telecommunication engineering from Xidian University, Xi'an, China, in 2006 and 2011, respectively.

He was a Post-Doctoral Fellow with the Department of Computer Science, City University of Hong Kong, Hong Kong, from 2013 to 2014. He is currently an Associate Professor with the School of Information Science and Engineering, Shandong University, Jinan, China. His current research interests include video coding and multimedia communication.

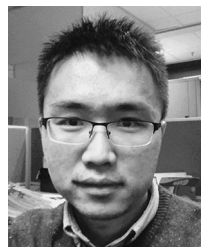
Sam Kwong (M'93–SM'04–F'13) received the B.S. degree from the State University of New York at Buffalo, Buffalo, NY, USA, in 1983, the M.S. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1985, and the Ph.D. degree from the University of Hagen, Hagen, Germany, in 1996.

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

Prof. Kwong serves as an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS and the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.

Xu Wang received the B.S. degree from South China Normal University, Guangzhou, China, in 2007, the M.S. degree from Ningbo University, Ningbo, China, in 2010, and the Ph.D. degree in computer science from The City University of Hong Kong, Hong Kong, in 2014.

He joined the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, in 2015, as an Assistant Professor. His research interests include video coding and stereoscopic image/video quality assessment.



Wei Gao received the M.S. degree in pattern recognition and intelligent systems from Huazhong University of Science and Technology, Wuhan, China, in 2012, and is currently working toward the Ph.D. degree in computer science at The City University of Hong Kong, Hong Kong.

His research interests include image and video processing, video coding and transmission, rate control, multimedia communication, visual perception, very-large-scale integration signal processing, and system-on-a-chip systems.

Yun Zhang (M'12) 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.

From 2009 to 2014, he was a Post-Doctoral Researcher with the Department of Computer Science, The City University of Hong Kong, Hong Kong. In 2010, he became an Assistant Professor with the Shenzhen Institutes of Advanced Technology, CAS, where he has been an Associate Professor since 2012. His research interests include video compression, 3D video processing, and visual perception.