

# Synthesized Views Distortion Model Based Rate Control in 3D-HEVC

Songchao Tan<sup>1</sup>(✉), Siwei Ma<sup>2</sup>, Shanshe Wang<sup>2</sup>, and Wen Gao<sup>2</sup>

<sup>1</sup> School of Computer Science and Technology,  
Dalian University of Technology, Dalian 116024, China  
sctan@jdl.ac.cn

<sup>2</sup> Institute of Digital Media, Peking University, Beijing 100871, China

**Abstract.** In this paper, we propose a synthesized views distortion model based rate control algorithm for the high efficiency video coding (HEVC) based 3D video compression standard. The major contributions of the paper include the following two aspects. Firstly, we investigate the distortion dependency between the synthesized views and the coded views including texture video and depth maps. Then we propose a synthesized views distortion model for 3D-HEVC, and based on the distortion model an efficient joint bit allocation scheme is proposed. Experimental results show that the proposed rate control algorithm achieves better performance on both the coded texture views and synthesized views. The maximum overall (including all coded texture views and all synthesized views) performance improvement can be up to 14.4 % and the average BD-rate gain is 6.9 %. Moreover, it can accurately control the bitrate to satisfy the total bitrate constraint.

**Keywords:** Rate control · Bit allocation · 3D video coding

## 1 Introduction

The 3D extension of High Efficiency Video Coding [1] (3D-HEVC) was developed by the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3 V) led by ISO/IEC MPEG and ITU-T VCEG. 3DV video coding aims at coding the visual information of a 3D scene that usually contains multi-view texture data and its corresponding depth information [2]. In 3D-HEVC, one view is selected as a base view which is coded independently of the other views to provide backward compatibility to HEVC decoders. Other views (termed as dependent views) are coded with inter-view prediction using the base view or other reference views to reduce the redundancy between views.

As an important module of video encoder, rate control (RC) is employed to regulate the bit rate meanwhile to guarantee a good video quality. For each video coding standard, rate control is always a hot research topic and many different rate control schemes for different video coding standards have been proposed, such as quadratic model for H.264 [3] and URQ model [4], R-lambda [5] and rate-GOP [6] for HEVC. However for 3D-HEVC, rate control becomes more complicated. Different from the traditional 2D video coding standards, 3D-HEVC utilizes the depth map to generate the

synthesized virtual views. The coding quality of synthesized views depends on the quality of texture video and depth maps [7]. As such, it is important to balance the bit allocation for texture video and depth maps to get better quality of synthesized views. And many other techniques are adopted to improve the coding performance. All these techniques bring great challenges to establish an accurate rate distortion (R-D) model and bit allocation scheme in rate control for 3D-HEVC.

In the literature, several rate control schemes for 3D-HEVC are proposed. Two representative rate control methods, URQ and R-lambda have been proposed for 3D-HEVC in [8,9], which are the extension of the methods in HEVC. In order to get better performance, depth maps-based inter-views MAD prediction is proposed to improve the prediction accuracy of the to-be-generated bits for the current unit. However, there is still large room for R-D performance improvement. In our previous work [10], we proposed an adaptive rate control scheme for 3D-HEVC. The algorithm performance including bit rate mismatch and R-D performance is significantly improved compared to the above two algorithms.

In this paper, we further propose a novel rate control scheme based on the synthesized views distortion model in 3D-HEVC. Firstly, we investigate the distortion dependency between the synthesized views and the input texture video and depth maps, and formulate a distortion model for synthesized views. Secondly, based on this model, the bit allocation scheme for texture video and depth maps is formulated as an optimization problem.

The rest of this paper is organized as follows. In Sect. 2, the R-D characteristics of both the coded views and synthesized views are investigated. In Sect. 2.1, the R-D model for coded texture views is proposed. A view synthesis distortion model to characterize the distortion dependency of the texture video and the depth maps on the synthesized virtual views is investigated in Sect. 2.2. In Sect. 2.3, an effective joint bit allocation based rate control scheme is designed for 3D-HEVC. In Sect. 3, the experimental results are given to demonstrate the efficiency of the proposed RC algorithm. Finally, Sect. 4 concludes this paper.

## 2 Rate and Distortion Analysis in 3D-HEVC

As illustrated in Fig. 1, the texture videos are captured by synchronizing the multiple camera arrays. The associated depth maps are also generated for virtual view synthesis. At the encoder, texture video and depth maps are encoded using 3D-HEVC. At the client side, the arbitrary virtual views are synthesized from the decoded texture video and depth maps. Then the decoded texture video and synthesized views would be presented for viewing at receiver side. Therefore, the quality of coded texture views and the virtual synthesized views needs to be optimized as follows

$$\begin{aligned} & \min(D_v + D_c), \\ & s.t. R_d + R_t \leq R_c, \end{aligned} \quad (1)$$

where  $R_t$  and  $R_d$  are the bit rate of texture video and depth maps, respectively.  $D_c$  and  $D_v$  are the distortion of texture video and synthesized views respectively.

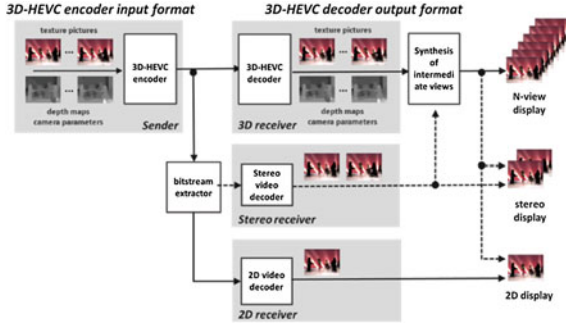


Fig. 1. General framework of the 3D-HEVC [1] system

In order to model the expression in (1), we need to investigate the rate and distortion (R-D) relationship for coded texture views and synthesized views, respectively.

### 2.1 R-D Model for the Coded Texture Views

To obtain the R-D characteristics of the coded views, we encode the original texture video with 4 quantization parameters (QP) (25, 30, 35, and 40). As an example, the R-D curve of the texture distortion and the bit rate of test sequence ‘*Newspaper\_CC*’ are illustrated in Fig. 2.

It can be observed that power functions can be used to fit the R-D points of texture video well.

$$D_t(R_t) \cong \alpha_c R_t^{-\beta_c}, \tag{2}$$

where  $D_t$  is the distortion of the coded texture views.  $R_t$  is the bit rate of texture views.  $\alpha_c$  and  $\beta_c$  are model parameters.

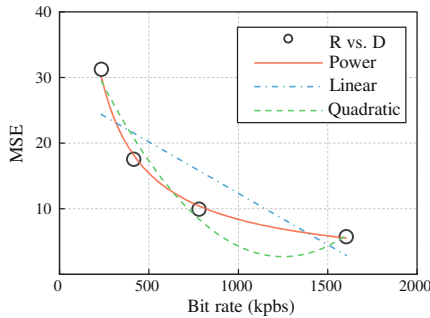
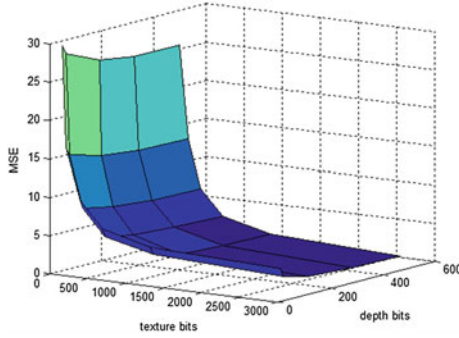


Fig. 2. The relationship between distortion and bit rate of coded texture views, ‘*Newspaper\_CC*’.

## 2.2 R-D Model for Synthesized Views

To find the best bit budget between the texture and depth, we also need to establish the relationship of bit rate and the synthesized views distortion. We investigate the synthesized views quality influence on the bit rate of texture video ( $R_t$ ) and the bit rate of depth map ( $R_d$ ).

In Fig. 3, the quality influence of texture and depth on synthesized views is investigated by changing the texture quantization parameter  $Q^T$  from 5 to 45 meanwhile fixing the depth quantization parameter  $Q^D$  at 24, 34, 39, 44 and 49 respectively. The quality of synthesized views ( $D_s$ ) is measured in term of MSE. As shown in Fig. 3, once  $R_t/R_d$  is determined, the  $D_s - R_t/D_s - R_d$  relationship can be approximated as power expression.



**Fig. 3.** The R-D surface of synthesized views distortion and bit rate. ‘kendo’.

The  $D_s - R_t$  relationship as

$$D_s(R_t) = \alpha_t R_t^{-\beta_t}. \quad (3)$$

And the  $D_s - R_d$  relationship as

$$D_s(R_d) = \alpha_d R_d^{-\beta_d}. \quad (4)$$

Therefore from (3) to (4), we get the distortion model for the synthesized views as follows,

$$D_s = \alpha_t R_t^{-\beta_t} + \alpha_d R_d^{-\beta_d}, \quad (5)$$

where  $D_s$  is the distortion of synthesized views.  $R_t$  and  $R_d$  are the bits for texture and depth.  $\alpha_t$ ,  $\beta_t$ ,  $\alpha_d$  and  $\beta_d$  are the model parameters.

### 2.3 A Joint Bit Allocation Based RC Scheme for 3D-HEVC

Rate control for 3D-HEVC needs to solve the bit allocation on texture/depth level, view level and frame level. The optimum bit allocation problem is to effectively distribute the bit budget between texture and depth so that the minimum views synthesis and coded views distortion are achieved. Based on the proposed coded texture views and synthesized views R-D model (5), we formulate the overall quality based optimum bit allocation as

$$\begin{aligned} (R_t^{opt}, R_d^{opt}) &= \arg \min \left( \alpha_t R_t^{-\beta_t} + \alpha_c R_t^{-\beta_c} + \alpha_d R_d^{-\beta_d} \right), \\ &s.t. R_t + R_d \leq R_c. \end{aligned} \quad (6)$$

$\zeta$  is used to represent the proportional relationship between  $R_t$  and  $R_d$ , defined as

$$\zeta = \frac{R_t^{opt}}{R_d^{opt}}. \quad (7)$$

Therefore, from (6) and (7), we get the objective optimization function with only one variable  $\zeta$ , as shown below,

$$(\zeta) = \arg \min \left( \alpha_t \left( \frac{\zeta}{1+\zeta} R_c \right)^{-\beta_t} + \alpha_c \left( \frac{\zeta}{1+\zeta} R_c \right)^{-\beta_c} + \alpha_d \left( \frac{1}{1+\zeta} R_c \right)^{-\beta_d} \right). \quad (8)$$

Many optimization methods can be used to find the optimal solution of (8). In this paper, Newton iterative method is used to get the approximate optimal value. The target bit rate for the texture and depth can be expressed as follows

$$R_t = R_c \cdot \frac{\zeta}{1+\zeta}, \quad (9)$$

$$R_d = R_c \cdot \frac{1}{1+\zeta}. \quad (10)$$

In order to estimate these parameters, we first encode the frames in the first GOP. Then the model parameters are calculated by the least square error method.

Based on the optimal target bit rate for the texture and depth, the bit rate ratio between the different views can be further determined by the statistical analysis. In this paper, we use anchor's bits ratio between the base view and the dependent views to allocate the bits for different views.

After allocating the target bit rate for texture/depth level and view level, the target bit rate needs to be allocated for the different frames. The frame level bit allocation is proposed in our previous work [7] as follows

$$R_{n,i} = \begin{cases} R_n^{remain} \cdot \phi & I \text{ frame} \\ R_n^{remain} \cdot w_i \cdot (1 - \phi) & \text{others} \end{cases}, \quad (11)$$

$$R_n^{remain} = \frac{\text{bit rate}}{\text{framerate}} \cdot N_n + (R_{n-1}^{remain} - R_{n-1}^{actual}) / N_{rest}^G, \quad (12)$$

where  $R_{n,i}$  is the target bits for  $i^{\text{th}}$  frame in  $n^{\text{th}}$  GOP.  $R_n^{remain}$  and  $R_n^{actual}$  are the target and actual bits in  $n^{\text{th}}$  GOP.  $N_n$  is the numbers of  $n$  th GOP's frames.  $N_{rest}^G$  is the number of the rest GOP which is not coded.  $\phi$  is a proportion of the I frame in a GOP which is recommended to be 0.4 and 0.25 respectively for the first and the rest GOPs based on experiments.  $w_i$  is the weight of the frames in RA hierarchical structure getting from experience.

$$w_i = \begin{cases} 0.07 & \text{if}(POC\%8 == 0) \\ 0.056 & \text{if}(POC\%8 == 4) \\ 0.0454 & \text{if}(POC\%4 == 2) \\ 0.035 & \text{else} \end{cases}, \quad (13)$$

where  $POC$  denotes Picture Order Count and represents an output order of the pictures in the video stream.

When overflow or underflow occurs, the difference between the target bits and the actual bits in a GOP will be distributed to the rest GOPs averagely.

Trade-off between the output bit rate ( $R$ ) and the quality ( $D$ ) of the compressed video are determined by the quantization step size ( $Qs$ ), which is indexed by quantization parameter ( $Q$ ). The R-Qs and D-Qs model have been studied extensively for the previous video coding standards such as H.264/AVC and HEVC. Here we use a linear model which is proposed in our previous work [7] as follows

$$R = \alpha \times X / QP, \quad (14)$$

where  $\alpha$  is the model parameter.  $R$  is the coding rate.  $QP$  is the quantization parameter.  $X$  is the complexity estimation for the current picture which is computed as following.

$$X = \left( \sum_{i=0}^n (w_i \times SAD_i) / \sum_{i=0}^{n-1} (w_i \times SAD_i) \right)^{1-\lambda} \times R_{n-1} \times QP_{n-1}, \quad (15)$$

where  $n$  is the current frame number.  $QP_{n-1}$  is the quantization parameter of the  $(n-1)$  th frame.  $R_{n-1}$  is the actual bits of the  $(n-1)$  th frame.  $w_i$  is defined as:

$$w_i = 0.5^{n-i} / \sum_{i=0}^n 0.5^{n-i} \quad (16)$$

### 3 Experimental Results

To evaluate the proposed 3D-HEVC rate control algorithm, the proposed algorithm is integrated into the reference software HTM10.0. In order to evaluate the performance of the proposed RC algorithm and R-lambda algorithm is utilized for comparison. We have tested our algorithm on all of eight sequences defined in the CTCs ( $1024 \times 768$  and  $1920 \times 1088$ ). Each sequence is composed of three views: the left, the center (coded first) and the right view. After coding, six synthesized views were rendered.

#### 3.1 Control Accuracy

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

$$Error = \frac{|R_{actual} - R_{target}|}{R_{target}} \times 100\%, \quad (17)$$

where *Error* is the bits error.  $R_{target}$  and  $R_{actual}$  are the number of target bits and the actual output bits, respectively.

As illustrated in Table 1, it can be seen that the proposed RC algorithm achieves smaller mismatch between target bits and actual output bits. That is because the frame level bit allocation proposed in our previous work [10] is designed more suitable for I-SLICE instead of relying on overflow/underflow handling strategy.

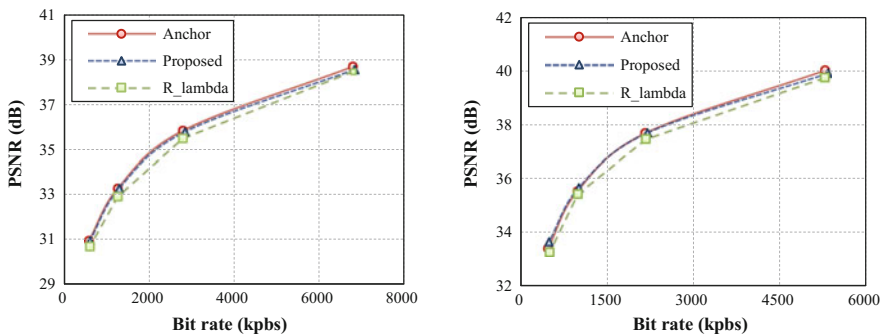
**Table 1.** Proposed Algorithm R-D Performance Compared With Anchor And R-lambda

	Anchor VS Proposed (BD-Rate)			R-lambda VS Proposed (BD-Rate)			Bits Error (%)	
	Texture PSNR / Texture bits	Syn views PSNR / Total bits	All views PSNR / Total bits	Texture PSNR / Texture bits	Syn views PSNR / Total bits	All views PSNR / Total bits	R-lambda	Proposed
Balloons	2.2 %	2.1 %	2.0 %	-3.8 %	-5.3 %	-5.2 %	1.47 %	0.28 %
Kendo	2.5 %	2.3 %	2.2 %	-4.3 %	-5.7 %	-5.6 %	0.58 %	0.37 %
Newspaper_CC	3.8 %	3.0 %	3.0 %	-2.6 %	-3.4 %	-3.3 %	2.60 %	0.54 %
GT_Fly	0.1 %	-0.8 %	-0.5 %	-8.4 %	-9.5 %	-9.8 %	1.45 %	1.46 %
Pozan_Hall2	4.7 %	3.3 %	3.5 %	-12.0 %	-14.0 %	-14.4 %	2.88 %	0.62 %
<b>Poznan_Street</b>	5.0 %	3.3 %	3.6 %	-4.1 %	-4.6 %	-4.8 %	6.01 %	3.88 %
<b>Undo_Dancer</b>	4.9 %	3.4 %	3.6 %	-6.4 %	-7.7 %	-7.8 %	0.98 %	0.96 %
<b>Shark</b>	3.5 %	2.1 %	2.5 %	-3.7 %	-4.7 %	-4.6 %	1.43 %	1.63 %
<b>1024*768</b>	2.8 %	2.5 %	2.4 %	-3.6 %	-4.8 %	-4.7 %	1.55 %	0.40 %
<b>1920*1088</b>	3.6 %	2.3 %	2.5 %	-6.9 %	-8.1 %	-8.3 %	2.55 %	1.71 %
<b>average</b>	<b>3.3 %</b>	<b>2.4 %</b>	<b>2.5 %</b>	<b>-5.7 %</b>	<b>-6.9 %</b>	<b>-6.9 %</b>	<b>2.18 %</b>	<b>1.22 %</b>

### 3.2 R-D Performance

In order to objectively evaluate the performance of the proposed RC algorithm, R-lambda algorithm proposed in [9] is utilized for comparison. In [9], the target bit rate of each texture video is set as corresponding bit rate in HTM anchor and depth maps are coded with fixed QP as the same as anchor. In the proposed algorithm, the target bit rate for all coded views' bit rate (including three texture videos and three depth maps) is set as anchor's total bit rate.

As illustrated in Table 1 and Fig. 4, we can see that the proposed algorithm shows much better R-D performance than R-lambda for both coded texture views and synthesized views. Based on the proposed synthesized views distortion model, the optimal bit allocation for the texture and depth is achieved. The maximum performance improvement for all views (including coded texture views and synthesized views) can be up to 14.4 % and the average BD-rate gain is 6.9 %.



**Fig. 4.** The R-D curves for all views of the proposed RC scheme compared with anchor and R-lambda ‘*Undo\_Dancer*’ and ‘*GT\_Fly*’

Furthermore, two R-D curves are shown in Fig. 4. It can be observed the proposed RC algorithm shows much better R-D performance than R-lambda model for both high bit rate and low bit rate.

## 4 Conclusions

This paper has presented a synthesized views distortion model based joint bit allocation and rate control method to achieve the best overall quality for 3D-HEVC. The distortion dependency is investigated between the coded views and the synthesized views. The proposed bit allocation method is classified into three levels, namely texture/depth level, view level and frame level. Experimental are conducted on different video sequences and the results show that the proposed method can achieve much better R-D performance than other algorithms for 3D-HEVC.



## References

1. Bross, B., Han, W.-J., Sullivan, G.J., Ohm, J.-R., Wiegand, T.: High efficiency video coding (HEVC) text specification draft 8. JCTVC-J1003, Stockholm, July 2012
2. Kauff, P., Atzpadin, N., Fehn, C., Müller, M., Schreer, O., Smolic, A., Tanger, R.: Depth Map Creation and Image Based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability. *Signal Processing: Image Communication, Special Issue on 3DTV*, pp. 217–234, February
3. Ma, S., Gao, W., Lu, Y.: Rate-distortion analysis for H.264/AVC video coding and its application to rate control. *IEEE Trans. Circ. Syst. Video Technol.* **15**(12), 1533–1544 (2005)
4. Choi, H., Nam, J., Yoo, J., Sim, D., Bajić, I.V.: Rate control based on unified RQ model for HEVC. JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCT-VC H0213 (m23088), San José, CA, USA, February 2012
5. Li, B., Li, H., Li, L., Zhang, J.: lambda Domain Rate Control Algorithm for High Efficiency Video Coding. *IEEE Trans. Image Process.* **23**(9), 3841–3854 (2014)
6. Wang, S., Ma, S., Wang, S., Zhao, D., Gao, W.: Rate-GOP based rate control for high efficiency video coding. *IEEE J. Sel. Top. Sign. Process.* **7**(6), 1101–1111 (2013)
7. Ma, S., Wang, S., Gao, W.: Low complexity adaptive view synthesis optimization in HEVC based 3D video coding. *IEEE Trans. Multimedia* **16**(1), 266–271 (2014)
8. Lim, W., Sim, D., Bajić, I.V.: JCT3 V – Improvement of the rate control for 3D multi-view video coding. ISO/IEC JTC1/SC29/WG11, JCT3 V-C0090, Geneva, Switzerland, January 2013
9. Lim, W., Sim, D., Bajić, I.V.: JCT3 V –The rate control schemes for 3D multi-view video coding. ISO/IEC JTC1/SC29/WG11, JCT3 V-D0111, Incheon, KR, April 2013
10. Tan, S., Si, J., Ma, S., Wang, S., Gao, W.: Adaptive Frame Level Rate Control in 3D-HEVC. *Visual Communication and Image Processing*, Malta, December 2014