Interview Rate Distortion Analysis-Based Coarse to Fine Bit Allocation Algorithm for 3-D Video Coding

Hui Yuan, *Member, IEEE*, Sam Kwong, *Fellow, IEEE*, Chuan Ge, Xu Wang, *Student, IEEE Member, IEEE*, and Yun Zhang, *Member, IEEE*

***Abstract*—In three dimensional video coding (3-DVC), it is reasonable to allocate bits for texture videos and depth maps differently as the characteristic of texture videos is different with that of depth maps. In order to improve the accuracy of bit allocation performance for 3-DVC, a more accurate distortion model for virtual view and rate models for texture videos and depth maps are proposed based on interview rate distortion analysis. Based on the proposed models, the bit allocation optimization problem is resolved by a coarse to fine strategy. The proposed bit allocation algorithm is implemented in the High Efficiency Video Coding-based 3-DVC platform version 6.0 (3-D-HTM 6.0). Experimental results demonstrate that the proposed distortion and rate models are accurate; meanwhile, the absolute difference (denoted as "rate inaccuracy") between the actual coding bit rate (R_{ABR}) and the target coding bit rate (R_{TBR}) of the proposed method is only 2.67% on average; while the rate inaccuracy of the existing planar model-based method is 17.71%. Besides, there is an average 29.68% complexity increment when comparing with the planar model-based method.**

***Index Terms*—Bit allocation, rate distortion analysis, depth maps, 3-D video coding.**

I. Introduction

With the rapid growth of social media applications, video information is taking up more and more network bandwidth and terminal storage space. Compared with the traditional monoscopic videos, three dimensional (3D) videos (3DVs) are more appealing to audiences for its impressive visual experiences. However, the bandwidth/storage constrain of network/terminal become intensified. In order to reduce the huge data volume of 3DV, Joint Collaborative Team for 3DV (JCT-3V), established by Moving Pictures Experts Group (MPEG) of International Organization for Standardization (ISO) and Video Coding Experts Group (VCEG) of International Telecommunication Union (ITU), is developing 3DVC standards [1], [2] based on the High Efficiency Video Coding (HEVC) Standard [3].

The widely used 3DV format consists of two kinds of information, i.e. n-view texture videos and the corresponding n-view depth maps. In the current 3D video coding test model [4], n is set as 2 or 3 for reducing data volume as much as possible. When users want to enjoy more views (in the n-view range) which are not coded and transmitted, the display terminal will render a virtual view based on user specified view position. The process is named as virtual view synthesis which is based on Depth-Image Based Rendering (DIBR) [5] technique, as shown in Fig. 1.

From Fig. 1, it is observed that the two kinds of information of a 3DV system, i.e. texture videos and depth maps, are different. Texture videos are adopted to describe the image content of the scene; however, depth maps are utilized to present the transformed physical distance between camera centers and objects in the scene. Compared with texture videos, depth maps are sparser and smoother. Besides, depth maps are not for viewing, but only for virtual views rendering. Accordingly, the influences of depth maps and texture videos on the quality of a virtual views as well as the 3DV system should be different. Thus, for achieving high coding performance in a bandwidth/storage limited 3DV system, bit allocation between texture videos and depth maps becomes an important issue in 3DV system.

In [5], fixed ratio (5:1) bit allocation is used, but the quality of virtual views and the 3DV system are not guaranteed.

Manuscript received April 23, 2014; revised September 1, 2014; accepted September 10, 2014. Date of publication October 30, 2014; date of current version December 5, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 61201211, Grant 61272289, and Grant 61471348, in part by the Hong Kong Scholar Project under Grant XJ2012029, in part by the City University of Hong Kong Applied Research Grant 9667094, in part by the City University of Hong Kong Shenzhen Research Institute, Shenzhen, China, 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 China Post-Doctoral Science Foundation Funded Project under Grant 2013M530320, in part by the China Post-Doctoral Special Foundation Funded Project under Grant 2012T50629, in part by the Shandong Post-Doctoral Innovation Project Fund under Grant 201203053, and in part by the Open Research Fund of Key Laboratory of Wireless Sensor Network and Communication under Grant 2013002.

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

S. Kwong and X. Wang are 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; xuwang6-c@my.cityu.edu.hk).

C. Ge is with the School of Information Science and Engineering, Shandong University, Jinan 250100, China (e-mail: gavinkeh.chuan@gmail.com).

Y. Zhang is with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: yunzhang@cityu.edu.hk).

Digital Object Identifier 10.1109/TBC.2014.2361964

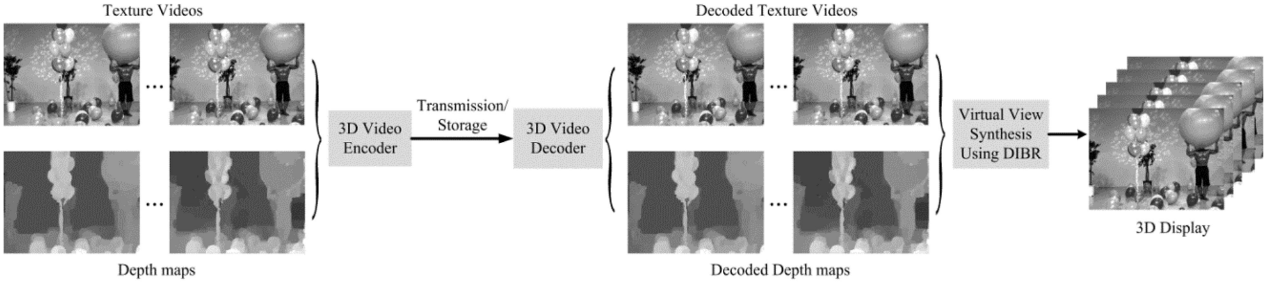


Fig. 1. Example of a typical 3-D video system.

Daribo *et al.* [6] made use of the standard deviations of depth maps and texture videos for the allocation of bits. However, the influences of depth maps degradation on the quality of virtual views are not analyzed; it also could not guarantee the optimal quality in the virtual views. A full search algorithm was proposed in [7] to find the optimal encoding parameters, i.e. quantization parameters (QP), for texture videos and depth maps. However, the complexity is too high to be applied for many practical applications. A distortion estimation model for virtual views is proposed in [8]. Based on the estimated distortion, the bit allocation problem is also solved by a search algorithm and the complexity of this algorithm is still so high that it cannot be applied easily. In order to reduce the computation complexity, an analytical distortion model was proposed for virtual views in our previous work [9]. Then, the bit allocation problem was formulated as a constrained optimization problem, and solved by using the Lagrangian multiplier method directly. Then, Hu *et al.* [10] proposed an improved distortion model for virtual views, and use the total distortion of texture videos and virtual views as an objective function for the bit allocation optimization problem. In [11], a region-based view synthesis distortion estimation approach is investigated with respect to the distortion of texture videos and depth maps, and a corresponding bit allocation scheme is also designed. Besides, in [12], an asymmetric coding method of multi-view video plus depth (MVD) based 3D video is proposed on purpose of providing high-quality view rendering. In [13] and [14], during the bit allocation of texture videos and depth maps, inter-view bit allocations are considered and performed according to a predefined ratio. In [15], a scalable bit allocation method for 3DV streaming is proposed based on a varied rate constraint. In this paper, based on the coding structure of HEVC based 3DVC Test Platform, 3D-HTM [16], a coarse to fine bit allocation algorithms for 3DV is proposed based on inter-view rate distortion analysis so as to improve the accuracy of the bit allocation problem.

The remainder of the paper is organized as follows: In Section II, inter-view rate distortion analysis is performed so as to derive an exact distortion model for virtual views and rate models for both texture videos and depth maps. Based on the models, a coarse to fine bit allocation algorithms for 3DV is proposed in Section III. Experiments and conclusions are presented in Sections IV and V respectively.

II. Interview Rate Distortion Analysis

The inter-view bit allocation problem of 3DVs can be regarded as an optimization problem, i.e. to achieve the minimum distortion of virtual view under a certain bit rate constraint. Once the optimization problem is solved, an optimal set of encoder parameters will be determined for 3DVs. In a 3DV encoder, inter-view prediction [17], [18] has been adopted as an important and effective coding tool. Therefore, it is necessary to derive a more precise rate and distortion model for 3DVC based on the inter-view dependency relationship. Currently, the test platform 3D-HTM contains 3 different QPs to control the output bit rates, i.e. QP of base view (BV) texture video (QP_t), QP of BV depth map (QP_d), and the QP difference (Δ_{QP}) between BV and non-base view (NBV). Therefore, QP_t (the corresponding quantization step, Q_t), QP_d (the corresponding quantization step, Q_d), and Δ_{QP} should be considered in the rate and distortion analysis.

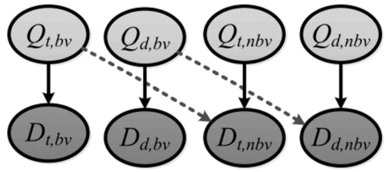


Fig. 2. Example of interview error propagation, $Q_{t,\mathrm{bv}}$ could affect both $D_{t,\mathrm{bv}}$ and $D_{t,\mathrm{nbv}}$; while $Q_{d,\mathrm{bv}}$ could affect both $D_{d,\mathrm{bv}}$ and $D_{d,\mathrm{nbv}}$.

A. Distortion Analysis

During the analysis, the mean squared error (MSE) between the original image and the corresponding reconstructed image is used as distortion criterion, which could be calculated as,

$$D = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \left[I\left(i,j\right) - I'\left(i,j\right) \right]^2, \tag{1}$$

where W and H are the width and height of the image, $I(i,\ j)$ and $I'(i,j)$ are the original image and the reconstructed image respectively. For the synthesized virtual view, $I(i,\ j)$ represents the virtual view synthesized from uncompressed texture and depth images; while $I'(i,j)$ represent the virtual view synthesized from those compressed texture and depth images.

For a 3DV encoder, there exists an inter-view propagation error caused by inter-view prediction, i.e. the quantization step of BV ($Q_{t,bv}$) could also affect the distortion of NBV ($D_{t,nbv}$), as shown in Fig. 2.

Therefore, the relationship between $D_{t,bv}$ and $D_{t,nbv}$ should also be studied in order to derive an accurate distortion model for virtual view. Let $I_{bv}(i)$ denotes the i^{th} pixel value in the BV

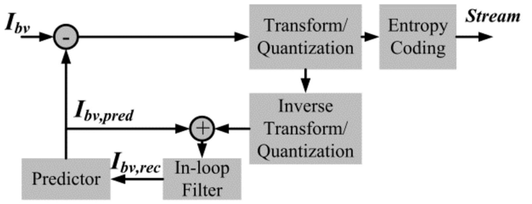


Fig. 3. Relationship among $I_{\rm bv}$, $I_{\rm bv,\,pred}$, and $I_{\rm bv,\,rec}$ for a typical encoder.

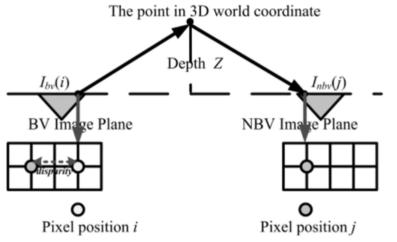


Fig. 4. Interview pixel matching, $I_{\rm bv}(i)$ is the matching pixel of $I_{\rm nbv}(j)$.

texture image, $I_{bv,pred}(i)$ denotes the predicted pixel value, and $I_{bv,rec}(i)$ denotes the reconstructed value of the pixel, as shown in Fig. 3. Accordingly, $I_{nbv}(j)$, $I_{nbv,pred}(j)$, and $I_{nbv,rec}(j)$ are used to represent the $j^{\rm th}$ pixel (the matching one of the $i^{\rm th}$ pixel of BV, as shown in Fig. 4) value, its predicted pixel value, and its reconstructed pixel value in the NBV texture image.

Thus, the coding error of the pixel i in the BV and that of the pixel j in the NBV could be written as

$$e_{bv}(i) = I_{bv}(i) - I_{bv,rec}(i), \tag{2}$$

$$e_{nbv}(j) = I_{nbv}(j) - I_{nbv,rec}(j). \tag{3}$$

Taking the inter-view prediction into account, since the i^{th} pixel in the base view and the j^{th} pixel in the non-base view are the matching pixels; the pixel difference between them should be zero. Accordingly, during the inter-view prediction of the encoder, the i^{th} pixel in the base view will be chosen to predict the j^{th} pixel in the non-base view. Thus, $I_{nbv,pred}(j)$ approximates to $I_{bv,rec}(i)$ because of the noise. As a result, e_{nbv} could also be written as

$$\begin{aligned} e_{nbv}(j) &= I_{nbv}(j) - I_{nbv,rec}(j) \\ &= I_{nbv}(j) - I_{nbv,pred}(j) + I_{nbv,pred}(j) - I_{nbv,rec}(j) \\ &\approx I_{nbv}(j) - I_{bv,rec}(i) + I_{bv,rec}(i) - I_{nbv,rec}(j) \\ &= e_{bv}(j) + I_{bv,rec}(i) - I_{nbv,rec}(j). \end{aligned} \tag{4}$$

Therefore, the mean squared error ($D_{t,nbv}$) of all the pixels in the NBV texture image could be written as

$$\begin{aligned} D_{t,nbv} &= \sum_{j}^{N} e_{nbv}(j)^2 \approx E\left\{e_{nbv}(j)^2\right\} \\ &= E\left\{\left[e_{bv}(j) + I_{bv,rec}(i) - I_{nbv,rec}(j)\right]^2\right\} \\ &= E\left\{e_{bv}^2\right\} + E\left\{\left[I_{bv,rec}(i) - I_{nbv,rec}(j)\right]^2\right\} \\ &\quad + 2 \cdot E\left\{e_{bv}\left[I_{bv,rec}(i) - I_{nbv,rec}(j)\right]\right\}, \end{aligned} \tag{5}$$

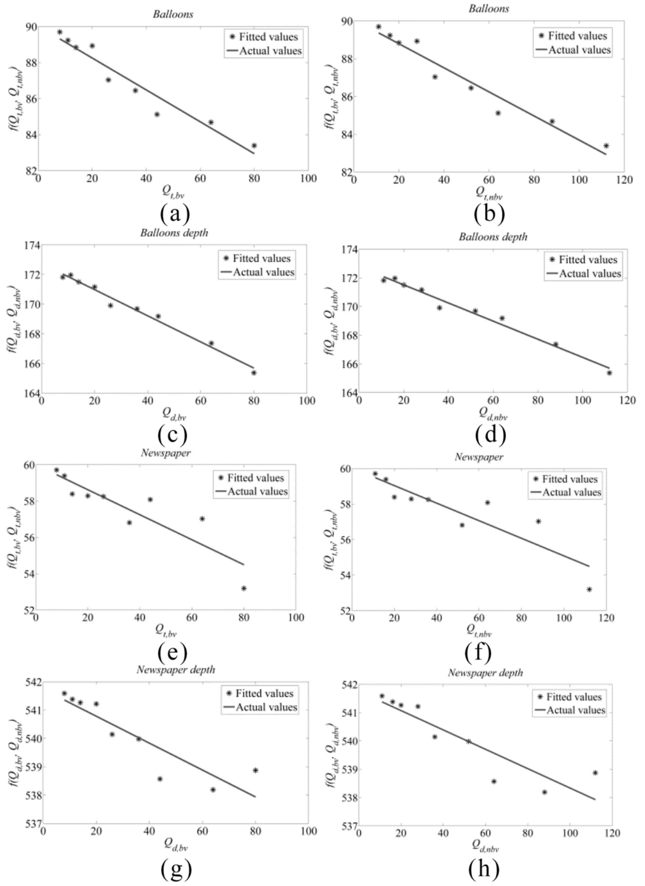


Fig. 5. Verification of (7). (a) and (b) *Balloons*. (c) and (d) *Balloons depth*. (e) and (f) *Newspaper*. (g) and (h) *Newspaper depth*.

where $E\{\cdot\}$ denotes the expectation value of a certain $e_{nbv}(j)^2$, and because of the Law of Large Numbers [19], the average value of $e_{nbv}(j)^2$ approximates to its expectation value.

Since e_{bv} could be proved to be independent with $I_{bv,rec}(i)$-$I_{nbv,rec}(j)$ (the proof is provide in Appendix I), (5) could be rewritten as

$$\begin{aligned} avg\left\{e_{nbv}^2\right\} &= E\left\{e_{bv}^2\right\} + E\left\{\left[I_{bv,rec}(i) - I_{nbv,rec}(j)\right]^2\right\} \\ &= D_{t,bv} + f\left(Q_{t,bv}, Q_{t,nbv}\right), \end{aligned} \tag{6}$$

were $f(Q_{t,bv},\ Q_{t,nbv}) = E\{[I_{bv,rec}(i)$-$I_{nbv,rec}(j)]^2\}$ which could be expanded via Taylor Series [20], i.e.

$$\begin{aligned} & f\left(Q_{t,bv}, Q_{t,nbv}\right) \\ &\quad = f(0,0) + \left(Q_{t,bv}\frac{\partial}{\partial Q_{t,bv}} + Q_{t,nbv}\frac{\partial}{\partial Q_{t,nbv}}\right) f(0,0) \\ &\quad + \frac{1}{2!}\left(Q_{t,bv}\frac{\partial}{\partial Q_{t,bv}} + Q_{t,nbv}\frac{\partial}{\partial Q_{t,nbv}}\right)^2 f(0,0) + \cdots \\ &\quad \approx f(0,0) + \left(Q_{t,bv}\frac{\partial}{\partial Q_{t,bv}} + Q_{t,nbv}\frac{\partial}{\partial Q_{t,nbv}}\right) f(0,0) \\ &\quad = \mu Q_{t,bv} + \nu Q_{t,nbv} + c_0, \end{aligned} \tag{7}$$

where $c_0 = f(0,0)$, which represents the inherent discrepancy between the $i^{\rm th}$ pixel in the BV and the $j^{\rm th}$ pixel in NBV, μ and ν are the two parameters. The reason why $f(Q_{t,bv},\ Q_{t,nbv})$ could be expanded via Taylor Series is proved in Appendix II. The accuracy of (6) is verified in Fig. 5, from which we can observe

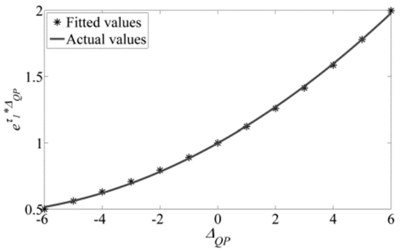


Fig. 6. Relationship between $e^{\tau 1^* \Delta \mathrm{QP}}$ and Δ_{QP}, where τ_1 = 0.11535.

that the $f(Q_{t,bv}, Q_{t,nbv})$ (or $f(Q_{d,bv}, Q_{d,nbv})$) varies linearly with $Q_{t,bv}$ (or $Q_{d,bv}$) and $Q_{t,nbv}$ (or $Q_{d,nbv}$). Besides, for the two sequences *Balloons* and *Newspaper* shown in Fig. 5, the squared correlation coefficients between $f(Q_{t,bv}, Q_{t,nbv})$ and $Q_{t,bv}$ are 0.9379 and 0.9035 respectively; those between $f(Q_{t,bv}, Q_{t,nbv})$ and $Q_{t,nbv}$ are 0.9467 and 0.9122 respectively; those between $f(Q_{d,bv}, Q_{d,nbv})$ and $Q_{d,bv}$ are 0.9824 and 0.9784 respectively; those between $f(Q_{d,bv}, Q_{d,nbv})$ and $Q_{d,nbv}$ are 0.9649 and 0.9742 respectively, i.e. (7) is accurate.

Because the relationship between $D_{t,bv}$ and $Q_{t,bv}$ could be represented as (8) [21],

$$D_{t,bv} = \varphi_t Q_{t,bv}, \tag{8}$$

$D_{t,nbv}$ could be represented as

$$\begin{aligned} D_{t,nbv} &= D_{t,bv} + f\left(Q_{t,bv}, Q_{t,nbv}\right) \\ &= \varphi_t Q_{t,bv} + \mu_t Q_{t,bv} + \nu_t Q_{t,nbv} + c_{0,t} \\ &= m_t \cdot Q_{t,bv} + n_t \cdot Q_{t,nbv} + c_{0,t}, \end{aligned} \tag{9}$$

where m_t, n_t, and $c_{0,t}$ are model parameters. Since the relationship between Q and QP could be represented as (10) [10] by fitting,

$$Q \approx e^{\tau_1 QP + \tau_2}, \tag{10}$$

where τ_1 = 0.11535, τ_2 = -0.4593. Thus, the QP difference (Δ_{QP}) between $QP_{t,bv}$ and $QP_{t,nbv}$ could be derived as

$$Q_{t,nbv} = e^{\tau_1 \Delta_{QP}} Q_{t,bv}, \tag{11}$$

by solving the following equations

$$\begin{cases} Q_{t,nbv} \approx e^{\tau_1 QP_{t,nbv} + \tau_2} \\ Q_{t,bv} \approx e^{\tau_1 QP_{t,bv} + \tau_2} \\ \Delta_{QP} = QP_{t,nbv} - QP_{t,bv}. \end{cases} \tag{12}$$

Because, $e^{\tau_1 \cdot \Delta_{QP}}$ could be expanded by power series, i.e.

$$\begin{aligned} e^{\tau_1 \cdot \Delta_{QP}} &= 1 + \left(\tau_1 \Delta_{QP}\right) + \left(\tau_1 \Delta_{QP}\right)^2 / 2! + \cdots \\ &\approx p_2 \cdot \Delta_{QP}^2 + p_1 \Delta_{QP} + p_0, \end{aligned} \tag{13}$$

which is verified in Fig. 6, $D_{t,nbv}$ could be further represented as

$$\begin{aligned} D_{t,nbv} &= m_t Q_{t,bv} + n_t Q_{t,nbv} + c_{0,t} \\ &= m_t Q_{t,bv} + n_t Q_{t,bv} \left(p_2 \Delta_{QP}^2 + p_1 \Delta_{QP} + p_0\right) + c_{0,t} \\ &= \left(n_t p_2 \Delta_{QP}^2 + n_t p_1 \Delta_{QP} + m_t + n_t p_0\right) Q_{t,bv} + c_{0,t}, \end{aligned} \tag{14}$$

which means that there exists a linear relationship between $D_{t,nbv}$ and $Q_{t,bv}$, and the slope between $D_{t,nbv}$ and $Q_{t,bv}$ depends on Δ_{QP}. Similarly, $D_{d,nbv}$ could also be derived as,

$$D_{d,nbv} = \left(n_d p_2 \Delta_{QP}^2 + n_d p_1 \Delta_{QP} + m_d + n_d p_0\right) Q_{d,bv} + c_{0,d}. \tag{15}$$

Consequently, by taking the planar model [9] into consideration, a more precise distortion model of virtual view could be written as,

$$\begin{aligned} D_v &= A\left(D_{t,bv} + D_{t,nbv}\right) + B\left(D_{d,bv} + D_{d,nbv}\right) \\ &= A\left(n_t p_2 \Delta_{QP}^2 + n_t p_1 \Delta_{QP} + m_t + n_t p_0 + \varphi_t\right) Q_{t,bv} \\ &\quad + B\left(n_d p_2 \Delta_{QP}^2 + n_d p_1 \Delta_{QP} + m_d + n_d p_0 + \varphi_d\right) Q_{d,bv} \\ &\quad + \left(Ac_{0,t} + Ac_{0,d}\right). \end{aligned} \tag{16}$$

For simplification, (16) is rewritten as

$$\begin{aligned} D_v = &\left(g_2 \Delta_{QP}^2 + g_1 \Delta_{QP} + g_0\right) Q_{t,bv} \\ &+ \left(h_2 \Delta_{QP}^2 + h_1 \Delta_{QP} + h_0\right) Q_{d,bv} + C, \end{aligned} \tag{17}$$

where g_2, g_1, g_0, h_2, h_1, h_0, and C are the model parameters.

B. Rate Analysis

In a 3DV encoder, the output bits are composed of four parts, i.e. the coding bits of BV texture videos, NBV texture videos, BV depth maps, and NBV depth maps. For each part, the coding bit rate model could be represented as (18) [22],

$$R = \alpha Q^{-1}, \tag{18}$$

where α is a model parameter, Q is the quantization step. Therefore, take texture video as an example, the coding bit rate of BV and NBV texture video could be written as

$$\begin{aligned} R_t &= R_{t,bv} + R_{t,nbv} = \alpha_{t,bv} Q_{t,bv}^{-1} + \alpha_{t,nbv} Q_{t,nbv}^{-1} \\ &= \alpha_{t,bv} Q_{t,bv}^{-1} + \alpha_{t,nbv} e^{-\tau_1 \cdot \Delta_{QP}} Q_{t,bv}^{-1}. \end{aligned} \tag{19}$$

Since $e^{-\tau_1 \cdot \Delta_{QP}}$ could also be expanded via power series, i.e.

$$\begin{aligned} (e)^{-\tau_1 \cdot \Delta_{QP}} &= 1 - \left(\tau_1 \cdot \Delta_{QP}\right) + \left(\tau_1 \cdot \Delta_{QP}\right)^2 / 2! + \cdots \\ &= q_2 \cdot \Delta_{QP}^2 + q_1 \cdot \Delta_{QP} + q_0, \end{aligned} \tag{20}$$

which is verified in Fig. 7, therefore, (19) could be further rewritten as

$$\begin{aligned} R_t &= \alpha_{t,bv} Q_t^{-1} + \alpha_{t,nbv} e^{-\tau_1 \cdot \Delta_{QP}} Q_{t,bv}^{-1} \\ &= \alpha_{t,bv} Q_t^{-1} + \alpha_{t,nbv} \left(q_2 \Delta_{QP}^2 + q_1 \Delta_{QP} + q_0\right) Q_{t,bv}^{-1} \\ &= \left(\alpha_{t,nbv} q_2 \Delta_{QP}^2 + \alpha_{t,nbv} q_1 \Delta_{QP} + \alpha_{t,nbv} q_0 + \alpha_{t,bv}\right) Q_{t,bv}^{-1}. \end{aligned} \tag{21}$$

Similarly, the coding bit rate of depth maps could be represented as

$$\begin{aligned} R_d = &\left(\alpha_{d,nbv} q_2 \Delta_{QP}^2 + \alpha_{d,nbv} q_1 \Delta_{QP}\right. \\ &\left. + \alpha_{d,nbv} q_0 + \alpha_{d,bv}\right) Q_{d,bv}^{-1}. \end{aligned} \tag{22}$$

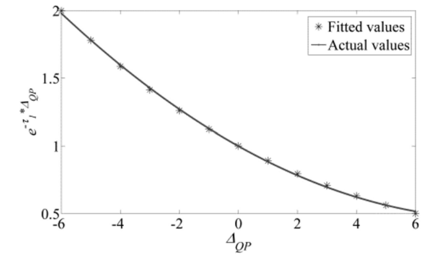


Fig. 7. Relationship between $e^{-\tau 1^* \Delta \mathrm{QP}}$ and Δ_{QP}, where $\tau_1 = 0.11535$.

Thereby, the total target bit rate could be represented as

$$\begin{aligned} R_{TBR} &= R_t + R_d \\ &= \Big(\alpha_{t,nbv} q_2 \Delta_{QP}^2 + \alpha_{t,nbv} q_1 \Delta_{QP} \\ &\quad + \alpha_{t,nbv} q_0 + \alpha_{t,bv}\Big) Q_{t,bv}^{-1} \\ &\quad + \Big(\alpha_{d,nbv} q_2 \Delta_{QP}^2 + \alpha_{d,nbv} q_1 \Delta_{QP} \\ &\quad + \alpha_{d,nbv} q_0 + \alpha_{d,bv}\Big) Q_{d,bv}^{-1}. \end{aligned} \tag{23}$$

For easy representation, (23) could be rewritten as

$$\begin{aligned} R_{TBR} = &\left(\alpha_2 \Delta_{QP}^2 + \alpha_1 \Delta_{QP} + \alpha_0\right) Q_{t,bv}^{-1} \\ &+ \left(\beta_2 \Delta_{QP}^2 + \beta_1 \Delta_{QP} + \beta_0\right) Q_{d,bv}^{-1}. \end{aligned} \tag{24}$$

III. Proposed Coarse to Fine Bit Allocation Algorithm

Based on the distortion and rate analysis in Section II, the 3DV bit allocation problem could be resolved by solving the optimization problem

$$\begin{aligned} \min \ &\left(g_2 \Delta_{QP}^2 + g_1 \Delta_{QP} + g_0\right) Q_{t,bv} \\ &+ \left(h_2 \Delta_{QP}^2 + h_1 \Delta_{QP} + h_0\right) Q_{d,bv} + C, \\ s.t. \ &\left(\alpha_2 \Delta_{QP}^2 + \alpha_1 \Delta_{QP} + \alpha_0\right) Q_{t,bv}^{-1} \\ &+ \left(\beta_2 \Delta_{QP}^2 + \beta_1 \Delta_{QP} + \beta_0\right) Q_{d,bv}^{-1} = R_{TBR}, \end{aligned} \tag{25}$$

where R_{TBR} represents the target bit rate. However, both the objective and the constraint function of (25) are not convex, which makes the solving procedure difficult. Therefore, in this work, Δ_{QP} is used to fine tune the 3D video bit allocation procedure so as to meet the target bit rate requirements. Firstly, we can set a default Δ_{QP} value (denoted as Δ_{QP1}) to calculate Q_t and Q_d using the method in [9]. Then in the second step, we calculate the optimal Δ_{QP} (denoted as Δ_{QP2}) under the given Q_t (the corresponding QP is QP_t) and Q_d (the corresponding QP is QP_d) to make the R_{ABR} close to the R_{TBR}.

Based on (21) and (22), for a 3DV sequence, when the quantization steps Q_t and Q_d are confirmed, the distortion and rate model could be rewritten as

$$D_v = G_2 \cdot \Delta_{QP}^2 + G_1 \cdot \Delta_{QP} + G_0, \tag{26}$$

$$R_{TBR} = \mathrm{K}_2 \cdot \Delta_{QP}^2 + \mathrm{K}_1 \cdot \Delta_{QP} + \mathrm{K}_0, \tag{27}$$

where, G_2, G_1, G_0, K_2, K_1, K_0 are six parameters that depends on Q_t and Q_d, and they can be calculated by pre-code the NBV texture video and depth maps under the given Q_t and Q_d. Then, the bit allocation problem could be represented as (28),

$$\begin{aligned} \min \quad & G_2 \cdot \Delta_{QP}^2 + G_1 \cdot \Delta_{QP} + G_0 \\ s.t. \quad & \mathrm{K}_2 \cdot \Delta_{QP}^2 + \mathrm{K}_1 \cdot \Delta_{QP} + \mathrm{K}_0 = R_{TBR}. \end{aligned} \tag{28}$$

By solving the quadratically constrained quadratic program (QCQP) [23] problem in (28), the optimal Δ_{QP} (denoted as Δ_{QP2}) could be written as

$$\Delta_{QP2} = \begin{cases} \left(-K_1 - \sqrt{K_1^2 - 4K_2\left(K_0 - R_{TBR}\right)}\right) / \left(2\mathrm{K}_2\right) \\ if\ K_1^2 - 4K_2\left(K_0 - R_{TBR}\right) \geq 0 \\ -K_1 / \left(2K_2\right) \ else \end{cases} \tag{29}$$

Since $R_{ABR,1}$ could be represented as

$$R_{ABR,1} = K_2 \cdot \Delta_{QP1}^2 + K_1 \cdot \Delta_{QP1} + K_0, \tag{30}$$

Subsequently,

$$\begin{aligned} &\Delta_{QP2} - \Delta_{QP1} \\ &= \frac{\sqrt{K_1^2 - 4K_2\left(K_0 - R_{ABR,1}\right)} - \sqrt{K_1^2 - 4K_2\left(K_0 - R_{TBR}\right)}}{2K_2}. \end{aligned} \tag{31}$$

From (31), we can conclude that when $R_{ABR},_1$ is larger than R_{TBR}, Δ_{QP} should be increased ($\Delta_{\mathrm{QP2}} > \Delta_{\mathrm{QP1}}$); and when the $R_{ABR},_1$ is smaller than R_{TBR}, Δ_{QP} should be decreased ($\Delta_{\mathrm{QP2}} < \Delta_{\mathrm{QP1}}$). In summary, the proposed coarse to fine bit allocation strategy could be described as the following,

Step 1: Set Target Bit Rate (R_{TBR});
Step 2: Set Initial Δ_{QP}=Δ_{QP1} (a default value);
Step 3: Calculate Q_t and Q_d by using the method in [9];
Step 4: Calculate model parameters K_2, K_1, and K_0 by precoding NBV texture video and depth maps for two times with different Δ_{QP} (specially, Δ_{QP}) respectively.
Step 5: Calculate Δ_{QP2} by using (29);
Step 6: Encode the video sequences with Q_t, Q_d, and Δ_{QP2}.

IV. Experimental Results

In the experiments, we first check the accuracy of the proposed distortion model and rate models. Then, the performance of the proposed coarse to fine bit allocation algorithm is verified. Six 3DV sequences [24] which are adopted by JCT-3V, i.e., *Balloons* (view 3 is encoded as BV, view 1 is encoded as NBV), *BookArrival* (view 8 is encoded as BV, view 10 is encoded as NBV), *Kendo* (view 3 is encoded as BV, view 1 is encoded as NBV), *Newspapercc* (view 4 is encoded as BV, view 2 is encoded as NBV), *GhostTownFly* (view 5 is encoded as BV, view 9 is encoded as NBV), and *UndoDancer* (view 5 is encoded as BV, view 1 is encoded as NBV), are used in the experiments. The 3DVC platform, 3D-HTM version 6.0 [16] is employed to encode those sequences, while the rendering software [16] integrated in 3D-HTM is used to synthesize virtual views. The encoder configuration files are set as the same with that of common test condition [25] of 3D-HTM.

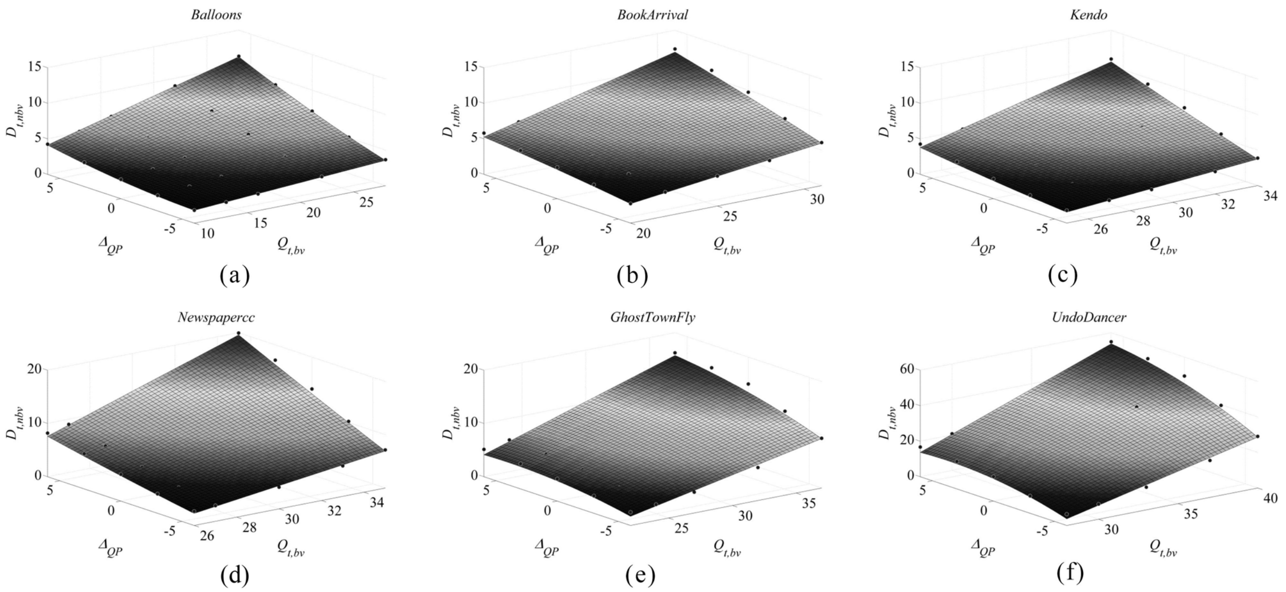


Fig. 8. Relationship between $D_{t,\mathrm{nbv}}$ and Δ_{QP}, each line corresponds to a fixed $Q_{t,\mathrm{bv}}$. (a) *Balloons*. (b) *BookArrival*. (c) *Kendo*. (d) *Newspapercc*. (e) *GhostTownFly*. (f) *UndoDancer*.

TABLE I
SQUARED CORRELATION COEFFICIENTS (r^2) BETWEEN THE ACTUAL $D_{t,\mathrm{nbv}}$ AND THE VALUES FITTED BY (14)

Sequences	r^2	Sequences	r^2
Balloons	0.9987	*Newspapercc*	0.9898
BookArrival	0.9802	*GhostTownFly*	0.9698
Kendo	0.9857	*UndoDancer*	0.9829

A. Distortion Model Verification

In this part, the accuracy of the distortion model shown in (17) is verified. However, (17) is a model with 3 parameters, which means that it cannot be shown by figures easily. From the derivation of (17), we can observe that (17) is obtained by adding (14) with (15). Therefore, we could verify (14) and (15) so as to validate the accuracy of (17).

From (14), we can observe that there is a quadratic relationship between $D_{t,nbv}$ and Δ_{QP}, and a linear relationship between $D_{t,nbv}$ and $Q_{t,bv}$. Fig. 8 shows the relationship among $D_{t,nbv}$, $Q_{t,bv}$, and Δ_{QP}. Besides, the squared correlation coefficients (r^2) between the actual $D_{t,nbv}$ and the fitted values are shown in Table I. From the table, we can observe that the r^2 are all larger than 0.96, which implies that (14) is accurate. Similarly, from (15), it can be observed that there is a quadratic relationship between $D_{d,nbv}$ and Δ_{QP}, and a linear relationship between $D_{d,nbv}$ and $Q_{d,bv}$, as shown in Fig. 9. Moreover, Table II, the r^2 for all the different video sequences are also larger than 0.99, i.e. (15) is accurate.

Besides, the accuracy of the proposed model in (17) is also compared with the planar model in [9]. The squared correlation coefficients (denoted as ${r_1}^2$) and the root of mean squared error ($RMSE_1$) between the actual D_v and the one fitted from (17) and those (denoted as ${r_2}^2$ and $RMSE_2$) between the actual D_v and the one fitted from the planar model in [9] are shown in Table III, from which we can observe that ${r_1}^2$ is larger than ${r_2}^2$, while $RMSE_1$ is smaller than $RMSE_2$. Thus, it implies that the proposed model is more accurate than the planar model in [9].

TABLE II
SQUARED CORRELATION COEFFICIENTS (r^2) BETWEEN THE ACTUAL $D_{d,\mathrm{nbv}}$ AND THE VALUES FITTED BY (15)

Sequences	r^2	Sequences	r^2
Balloons	0.9962	*Newspapercc*	0.9964
BookArrival	0.9982	*GhostTownFly*	0.9941
Kendo	0.9953	*UndoDancer*	0.9981

TABLE III
SQUARED CORRELATION COEFFICIENTS AND ROOT OF MEAN SQUARED FITTING ERROR BETWEEN THE ACTUAL D_v AND THE FITTED VALUES

Sequences	${r_1}^2$	${r_2}^2$	$RMSE_1$	$RMSE_2$
Balloons	0.9955	*0.9914*	*1.0827*	*1.4231*
BookArrival	0.9982	*0.9449*	*0.8651*	*3.2052*
Kendo	0.9971	*0.9946*	*0.7363*	*0.926*
Newspapercc	0.9974	0.9334	1.3483	3.8705
GhostTownFly	0.9944	0.9583	1.2572	2.2058
UndoDancer	0.9949	0.9595	2.2286	5.9160

B. Rate Model Verification

In this part, in order to verify the accuracy of the rate model shown in (24) which could not be shown as figures easily, the rate models (21) and (22) are verified. From (21), when Δ_{QP} is confirmed, there will be a linear relationship between R_t and ${Q_{t,bv}}^{-1}$; when $Q_{t,bv}$ is confirmed, there will be a quadratic relationship between R_t and Δ_{QP}. The relationship among R_t, ${Q_{t,bv}}^{-1}$, and Δ_{QP} is shown in Fig. 10 and Table IV.

Similarly, the relationship among R_d, ${Q_{d,bv}}^{-1}$, and Δ_{QP} is shown in Fig. 11 and Table V. From Tables IV and V, we can observe that the r^2 between the actual R_t and the value fitted by (21) are all larger than 0.88, and the r^2 between the actual R_d and the value fitted by (22) are all larger than 0.94,

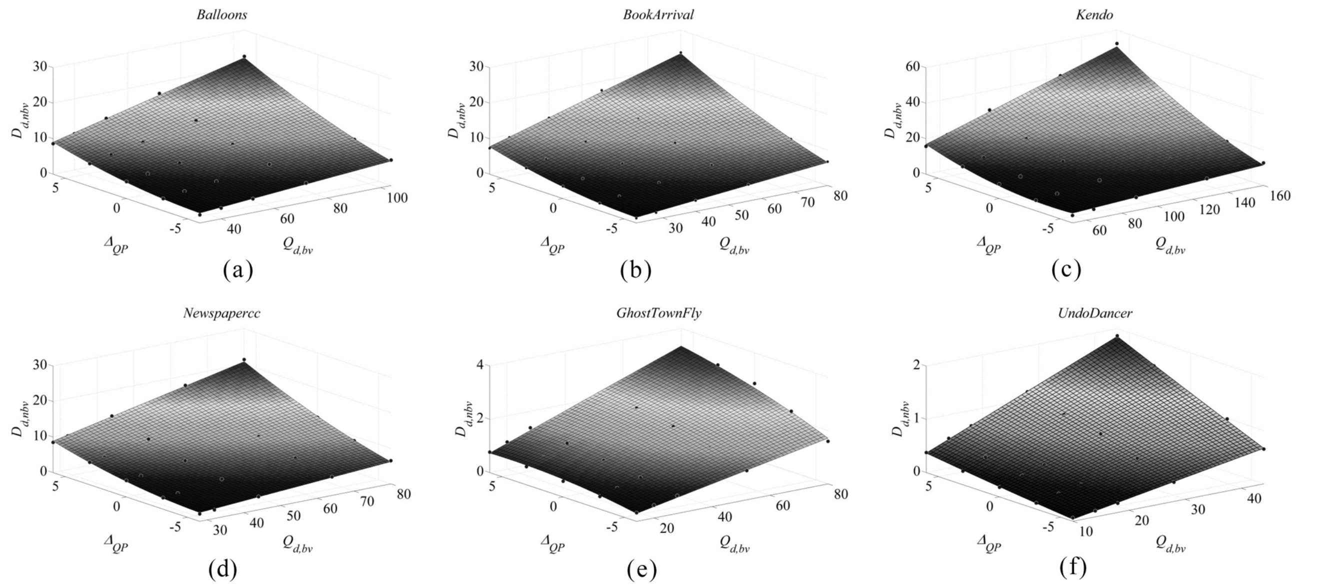


Fig. 9. Relationship between $D_{d,\mathrm{nbv}}$ and Δ_{QP}, each line corresponds to a fixed $Q_{d,\mathrm{bv}}$. (a) *Balloons*. (b) *BookArrival*. (c) *Kendo*. (d) *Newspapercc*. (e) *GhostTownFly*. (f) *UndoDancer*.

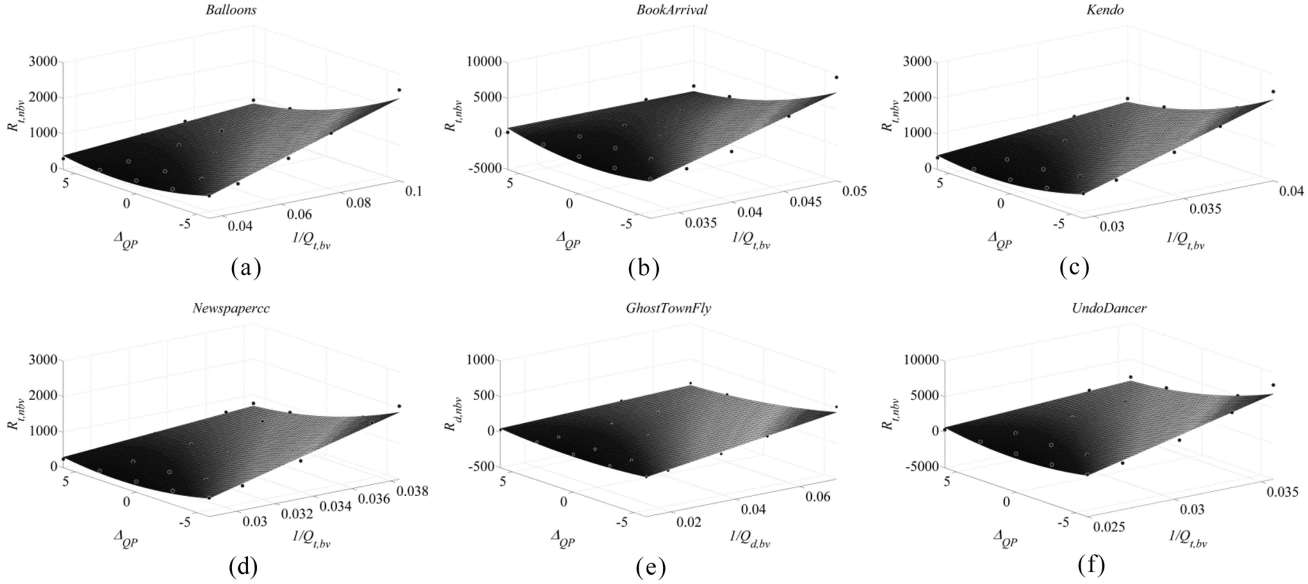


Fig. 10. Relationship among $R_{t,\mathrm{nbv}}$, $Q_{t,\mathrm{bv}}$, and Δ_{QP}. (a) *Balloons*. (b) *BookArrival*. (c) *Kendo*. (d) *Newspapercc*. (e) *GhostTownFly*. (f) *UndoDancer*.

TABLE IV
SQUARED CORRELATION COEFFICIENTS (r^2) BETWEEN THE ACTUAL $R_{t,\mathrm{nbv}}$ AND THE VALUES FITTED BY (21)

Sequences	r^2	Sequences	r^2
Balloons	0.9756	*Newspapercc*	0.9787
BookArrival	0.8829	*GhostTownFly*	0.9182
Kendo	0.9781	*UndoDancer*	0.9448

TABLE V
SQUARED CORRELATION COEFFICIENTS (r^2) BETWEEN THE ACTUAL $R_{d,\mathrm{nbv}}$ AND THE VALUES FITTED BY (22)

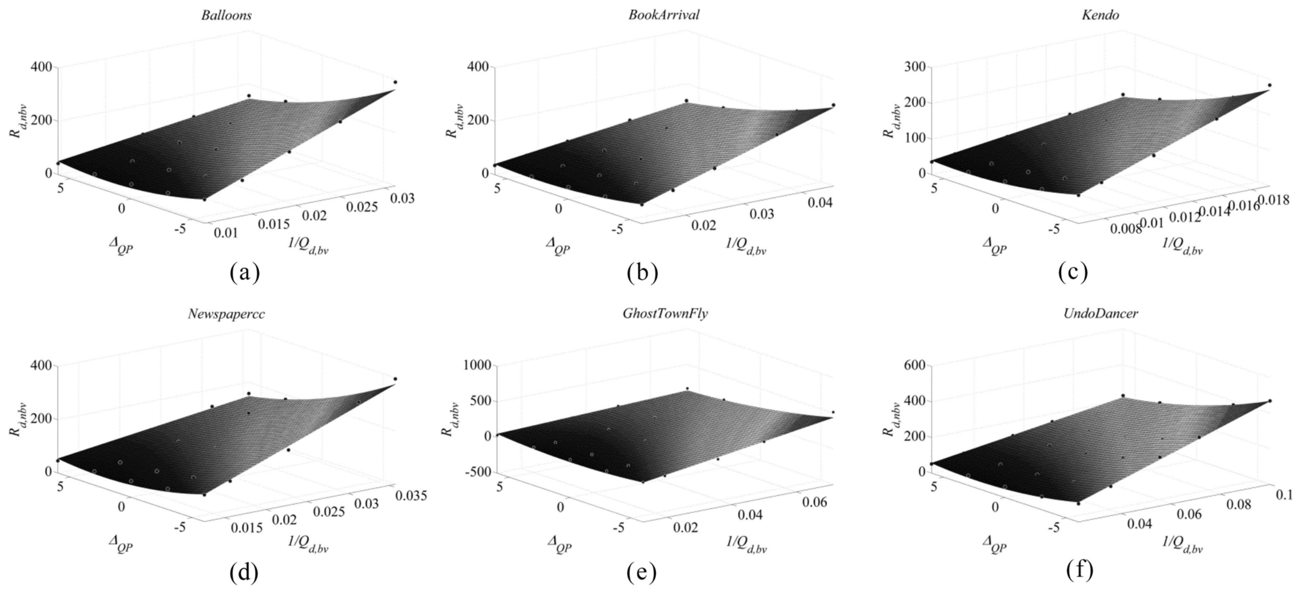
Sequences	r^2	Sequences	r^2
Balloons	0.9877	*Newspapercc*	0.9888
BookArrival	0.9949	*GhostTownFly*	0.9484
Kendo	0.9940	*UndoDancer*	0.9960

which mean that (21) and (22) are accurate. Therefore, the rate model shown in (24) is accurate.

C. Bit Allocation Performance Comparison

In this part, the Planar Model Based Method in [9] is used as a reference to evaluate the performance of the proposed algorithm (denoted as "Proposed Method"). For the sequences *Balloons*, *Bookarrival*, *Kendo*, and *Newspapercc*, the target coding bit rates are set as 1024kbps, 768kbps, 512kbps, and 384kbps; while for the sequences *GhostTownFly* and *UndoDancer*, the target coding bit rates are set as 3072kbps, 2048kbps, 1024kbps, and 512kbps.

In the experiments, $Q_{t,bv}$ and $Q_{d,bv}$ are first confirmed by pre-coding all the BV and NBV texture videos and depth



(a) (b) (c)

(d) (e) (f)

Fig. 11. Relationship among $R_{d,\text{nbv}}$, $Q_{d,\text{bv}}$, and Δ_{QP}. (a) *Balloons*. (b) *BookArrival*. (c) *Kendo*. (d) *Newspapercc*. (e) *GhostTownFly*. (f) *UndoDancer*.

TABLE VI
CONFIRMED $Q_{t,\text{bv}}$, $Q_{d,\text{bv}}$, AND Δ_{QP} OF THE PROPOSED METHOD AND THE PLANAR MODEL-BASED METHOD

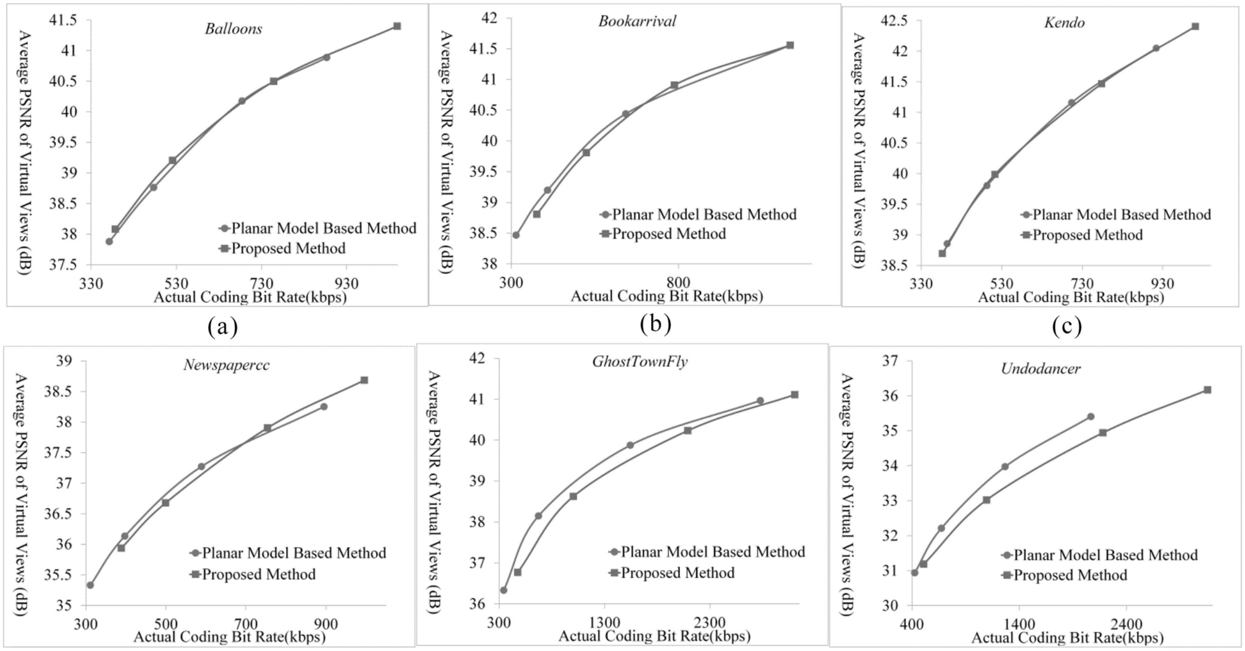
Sequences	RTBR(kbps)	Planar Model Based Method					Proposed Method				
		***Q** t,bv*	***Q** d,bv*	***QP** t,bv*	***QP** d,bv*	**Δ*QP***	***Q** t,bv*	***Q** d,bv*	***QP** t,bv*	***QP** d,bv*	**Δ*QP***
Balloons	1024	12.33	41.40	26	36	**3**	12.33	41.40	26	36	**0**
	768	15.97	53.63	28	38	**3**	15.97	53.63	28	38	**1**
	512	22.67	76.10	31	41	**3**	22.67	76.10	31	41	**1**
	384	28.69	96.10	33	44	**3**	28.69	96.10	33	44	**2**
BookArrival	1024	8.25	29.29	22	33	**3**	8.25	29.29	22	33	**3**
	768	11.01	39.06	25	36	**3**	11.01	39.06	25	36	**0**
	512	16.51	58.58	28	39	**3**	16.51	58.58	28	39	**-1**
	384	22.01	78.11	31	42	**3**	22.01	78.11	31	42	**-2**
Kendo	1024	14.06	66.14	27	40	**3**	14.06	66.14	27	40	**1**
	768	18.35	86.33	29	43	**3**	18.35	86.33	29	43	**1**
	512	26.41	124.27	32	46	**3**	26.41	124.27	32	46	**2**
	384	33.83	159.84	34	48	**3**	33.83	159.84	34	48	**4**
Newspapercc	1024	15.01	33.29	27	34	**3**	15.01	33.29	27	34	**1**
	768	19.54	43.35	30	37	**3**	19.54	43.35	30	37	**-1**
	512	27.99	62.11	33	40	**3**	27.99	62.11	33	40	**-1**
	384	35.72	79.26	35	42	**3**	35.72	79.26	35	42	**-1**
GhostTownFly	3072	10.04	19.38	24	30	**3**	10.04	19.38	24	30	**1**
	2048	14.49	27.97	27	33	**3**	14.49	27.97	27	33	**-1**
	1024	26.02	50.25	32	38	**3**	26.02	50.25	32	38	**-2**
	512	43.24	83.49	37	42	**3**	43.24	83.49	37	42	**-2**
UndoDancer	3072	19.75	13.46	30	27	**3**	19.75	13.46	30	27	**-2**
	2048	27.27	18.58	33	29	**3**	27.27	18.58	33	29	**-3**
	1024	44.07	30.03	37	33	**3**	44.07	30.03	37	33	**-3**
	512	63.67	43.38	40	37	**3**	63.67	43.38	40	37	**-1**

maps for two times. Then based on $Q_{t,bv}$ and $Q_{d,bv}$, NBV texture video and depth map are pre-coded for two times so as to confirm the Δ_{QP}. Table VI shows the confirmed $Q_{t,bv}$, $Q_{d,bv}$ and Δ_{QP}. Furthermore, the coding results based on those confirmed $Q_{t,bv}$, $Q_{d,bv}$ and Δ_{QP} values are shown in Table VII. From the table, we can observe that the absolute difference (denoted as "Rate inaccuracy") between the actual coding bit rate (R_{ABR}) and the target coding bit rate (R_{TBR}) of the Proposed Method is only 2.67%, while that of the Planar Model Based Method is 17.71%, i.e. the Proposed Method is more accurate than the Planar Model Based Method.

Besides, the rate distortion performance of the Proposed Method and the Planar Model Based Method are also compared. The BD PSNR [26] and rate distortion curves are used to denote the rate distortion performance difference between the Proposed Method and the Planar Model Based Method. Since the proposed bit allocation algorithm is a coarse to fine solution of (25), the rate distortion performance of the proposed method is not the best. The R_{ABR}-*PSNR* performance of the Proposed Method and the Planar Model Based Method are shown in Fig. 12 and Table VII, from which we can observe that the rate distortion performance of the Proposed Method

TABLE VII
BD PSNR BETWEEN THE PROPOSED METHOD AND THE PLANAR MODEL-BASED METHOD

Sequences	Planar Model Based Method				Propsoed Method				BD PSNR (dB) (Using R_{ABR})
	R_{TBR}(kbps)	R_{ABR}(kbps)	Rate inaccuracy (%)	Virtual View PSNR(dB)	R_{TBR}(kbps)	R_{ABR}(kbps)	Rate inaccuracy (%)	Virtual View PSNR(dB)	
Balloons	**1024.00**	**883.03**	**13.77**	40.89	**1024.00**	**1048.52**	**2.39**	41.40	**0.050**
	768.00	**684.38**	**10.89**	40.18	**768.00**	**758.61**	**1.22**	40.50	
	512.00	**476.90**	**6.86**	38.77	**512.00**	**521.31**	**1.82**	39.21	
	384.00	**372.62**	**2.96**	37.88	**384.00**	**386.85**	**0.74**	38.08	
Kendo	**1024.00**	**1134.62**	**10.80**	41.56	**1024.00**	**1134.62**	**10.80**	41.56	**-0.059**
	768.00	**642.53**	**16.34**	40.44	**768.00**	**788.09**	**2.62**	40.91	
	512.00	**408.52**	**20.21**	39.20	**512.00**	**524.69**	**2.48**	39.81	
	384.00	**314.36**	**18.14**	38.47	**384.00**	**376.23**	**2.02**	38.81	
BookArrival	**1024.00**	**914.09**	**10.73**	42.05	**1024.00**	**1011.29**	**1.24**	42.41	**-0.007**
	768.00	**703.57**	**8.39**	41.16	**768.00**	**778.09**	**1.31**	41.47	
	512.00	**493.05**	**3.70**	39.80	**512.00**	**512.85**	**0.17**	39.99	
	384.00	**394.14**	**2.64**	38.86	**384.00**	**382.07**	**0.50**	38.70	
Newspapercc	**1024.00**	**894.76**	**12.62**	38.25	**1024.00**	**995.54**	**2.78**	38.68	**-0.064**
	768.00	**588.81**	**23.33**	37.27	**768.00**	**753.82**	**1.85**	37.90	
	512.00	**397.56**	**22.35**	36.13	**512.00**	**499.35**	**2.47**	36.68	
	384.00	**311.29**	**18.93**	35.33	**384.00**	**388.79**	**1.25**	35.94	
GhostTownFly	**3072.00**	**2779.69**	**9.52**	40.97	**3072.00**	**3106.51**	**1.12**	41.11	**-0.351**
	2048.00	**1546.43**	**24.49**	39.87	**2048.00**	**2091.51**	**2.12**	40.23	
	1024.00	**674.98**	**34.08**	38.15	**1024.00**	**1005.65**	**1.79**	38.62	
	512.00	**344.64**	**32.69**	36.33	**512.00**	**477.46**	**6.75**	36.77	
UnderdoDancer	**3072.00**	**2067.29**	**32.71**	35.40	**3072.00**	**3154.87**	**2.70**	36.17	**-0.492**
	2048.00	**1267.98**	**38.09**	33.97	**2048.00**	**2179.69**	**6.43**	34.94	
	1024.00	**674.76**	**34.11**	32.21	**1024.00**	**1095.73**	**7.01**	33.02	
	512.00	**426.12**	**16.77**	30.94	**512.00**	**509.87**	**0.42**	31.19	
Average			**17.71**	**Average**			**2.67**	**Average**	**-0.154**



(a) (b) (c)

(d) (e) (f)

Fig. 12. R_{ABR}-PSNR curves comparison between the proposed method and the planar model-based method in [9]. (a) *Balloons*. (b) *BookArrival*. (c) *Kendo*. (d) *Newspapercc*. (e) *GhostTownFly*. (f) *UndoDancer*.

is similar with that of the Planar Model Based Method for sequences *Balloons*, *Bookarrival*, *Kendo*, and *Newspapercc*, and for sequences *GhostTownFly* and *UndoDancer*, the rate distortion performance of the Proposed Method is lower than that of the Planar Model Based Method. This is because that *GhostTownFly* and *UndoDancer* are generated from computer, while other sequences are captured from natural scenes, thus the rate distortion characteristics of the two video sequences are different from other sequences, i.e., a same bit rate increment may give lower virtual view quality increment compared with other sequences. As a result, the BD PSNRs of *GhostTownFly* and *UndoDancer* are much different from other sequences.

Since PSNR could not reflect the subjective quality effectively [27]–[30], we have given some subjective quality comparison as shown in Fig. 13. Since the proposed method

TABLE VIII
COMPLEXITY COMPARISON BETWEEN THE PROPOSED METHOD AND THE PLANAR MODEL-BASED METHOD

Sequences	RTBR(kbps)	Planar Model Based Method			Proposed Method			Time Increasing(%)
		Time for pre-coding (s)	Time for encoding (s)	Total Time (s)	Time for pre-coding (s)	Time for encoding (s)	Total Time (s)	
Balloons	1024	4144.82	2123.75	6268.57	6140.87	1915.25	8056.12	**28.52**
	768	4144.82	1967.04	6111.85	6155.36	1850.01	8005.37	**30.98**
	512	4144.82	1974.98	6119.80	5964.96	1785.22	7750.18	**26.64**
	384	4144.82	2019.47	6164.29	5874.05	1765.16	7639.21	**23.93**
BookArrival	1024	3243.38	2107.06	5350.45	5726.60	2048.24	7774.84	**45.31**
	768	3243.38	2111.54	5354.92	5397.07	1930.44	7327.51	**36.84**
	512	3243.38	2080.17	5323.55	5243.27	1886.15	7129.42	**33.92**
	384	3243.38	1926.00	5169.39	5132.36	1836.73	6969.09	**34.81**
Kendo	1024	4959.47	2810.85	7770.32	7527.59	2409.55	9937.14	**27.89**
	768	4959.47	2356.20	7315.67	7299.79	2271.69	9571.48	**30.84**
	512	4959.47	2392.00	7351.47	7218.65	2350.35	9569.00	**30.16**
	384	4959.47	2206.58	7166.05	7159.81	2072.76	9232.57	**28.84**
Newspapercc	1024	2921.34	1932.74	4854.08	4789.00	1855.36	6644.36	**36.88**
	768	2921.34	2086.68	5008.02	4703.64	1828.79	6532.43	**30.44**
	512	2921.34	1858.34	4779.67	4733.18	1824.66	6557.84	**37.20**
	384	2921.34	1783.91	4705.25	4472.74	1374.13	5846.87	**24.26**
GhostTownFly	3072	10842.15	7339.31	18181.46	16183.90	5174.34	21358.24	**17.47**
	2048	10842.15	6074.45	16916.60	15846.04	4821.97	20668.01	**22.18**
	1024	10842.15	5652.48	16494.63	15504.90	4651.94	20156.84	**22.20**
	512	10842.15	6143.89	16986.04	15271.63	4424.17	19695.80	**15.95**
UndoDancer	3072	11166.46	4964.43	16130.89	16098.98	5009.31	21108.29	**30.86**
	2048	11166.46	4824.53	15990.99	17548.95	4720.66	22269.61	**39.26**
	1024	11166.46	4755.70	15922.16	15646.17	4642.41	20288.58	**27.42**
	512	11166.46	4558.09	15724.55	15810.89	4541.76	20352.65	**29.43**
Average								**29.68**

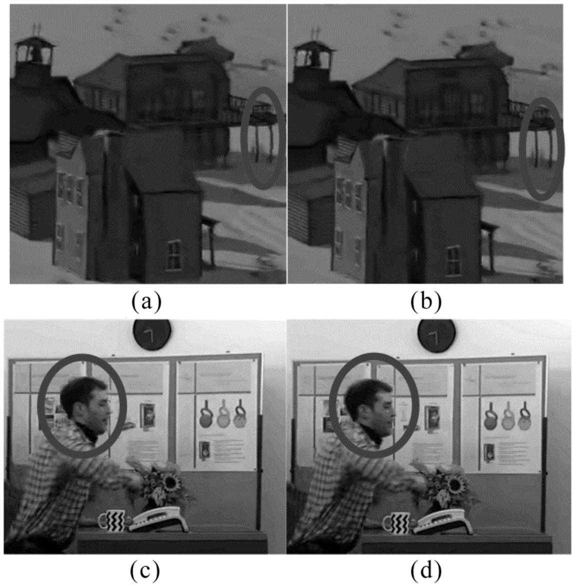


Fig. 13. Virtual view subjective quality comparison. Frame 24 of *GhostTownFly* by (a) planar model-based method [9] and (b) proposed method. Frame 24 of *BookArrival* by (c) planar model-based method [9] and (d) proposed method.

could utilize the given target bit rate effectively, from Fig. 13, it can be observed that the proposed method could give higher virtual view subjective quality than the planar model based method [9] at the same target bit rate.

D. Complexity Analysis

In the proposed method, model parameters should be calculated by pre-coding. For a certain 3D video sequence, we need first encode the texture videos (both the BV and the NBV) and depth maps (both the BV and the NBV) two times for determining $Q_{t,bv}$ and $Q_{d,bv}$ (the Planar Model Based Method could be completed at this step). Then, when the $Q_{t,bv}$ and $Q_{d,bv}$ are confirmed, the NBV texture video and depth map should be encoded for two times so as to confirm Δ_{QP}. Therefore, compared with the Planar Model Based Method, it is needed to additionally encode the NBV texture video and depth map for 2 times. Table VIII compares the encoding time of the Planar Model Based Method and the Proposed Method. From Table VIII, we can observe that compared with the Planar Model Based Method, an average 29.68% encoding time is increased by the Proposed Method.

V. CONCLUSION

In this paper, a coarse to fine bit allocation algorithm is proposed based on inter-view rate distortion analysis. Based on the analysis, a more accurate virtual view distortion model and rate model for texture videos and depth maps are proposed. Then, the 3DV bit allocation problem is converted as a constrained optimization problem, and solved by a coarse to fine strategy. Experimental results show that the proposed distortion and rate models are all accurate. Besides, the Rate inaccuracy of the Proposed Method is only 2.67%, while that of the Planar Model Based Method in [9] is 17.71%.

Moreover, the inter-view rate distortion analysis method could also be used in inter-frame rate distortion analysis so as to develop more accurate and efficient rate control algorithms for texture videos and depth maps, which is our future work.

APPENDIX I

Lemma 1: e_{bv} is independent with $\widetilde{I}_{bv}(i) - \widetilde{I}_{nbv}(j)$.

Proof: Since the i^{th} pixel in BV and the j^{th} pixel in the NBV are two matching pixels, $I_{bv}(i)$ approximates to $I_{nbv}(j)$. Accordingly, when $I_{bv}(i)$ and $I_{nbv}(j)$ are encoded with the same encoder parameters (quantization parameters, prediction modes, etc.), the reconstructed value $I_{bv,rec}(i)|_{Qbv,Pbv}$ is also closed to $I_{nbv,rec}(j)|_{Qbv,Pbv}$, where $|_{Qbv,Pbv}$ represents the pixel is encoded with quantization step Q_{bv} and prediction mode P_{bv}, i.e. $I_{bv,rec}(i)$ is the same with $I_{bv,rec}(i)|_{Qbv,Pbv}$. Similarly, $I_{nbv,rec}(j)$ could be represented as $I_{nbv,rec}(j)|_{Qbv,Pbv}$. Therefore, $I_{bv,rec}(i)$-$I_{nbv,rec}(j)$ could be represented as

$$\begin{aligned} & I_{bv,rec}(i) - I_{nbv,rec}(j) \\ & \quad = I_{nbv,rec}(j)\big|_{Q_{bv},P_{bv}} - I_{nbv,rec}(j)\big|_{Q_{nbv},P_{nbv}} \\ & \quad = I_{nbv,rec}(j)\big|_{Q_{bv},P_{bv}} - I_{nbv}(j) + I_{nbv}(j) \\ & \qquad\qquad - I_{nbv,rec}(j)\big|_{Q_{nbv},P_{nbv}} \\ & \quad = -e_{nbv}\big|_{Q_{bv},P_{bv}} + e_{nbv}\big|_{Q_{nbv},P_{nbv}}, \end{aligned} \tag{32}$$

where, $e_{nbv}|_{Qbv,Pbv}$ and $e_{nbv}|_{Qnbv,Pnbv}$ are the coding errors of the j^{th} pixel in the NBV at two different encoding parameters (Q_{bv}, P_{bv}) and (Q_{nbv}, P_{nbv}) respectively. Since an image coding error could be thought as a white noise [31], therefore e_{bv}, $e_{nbv}|_{Qbv,Pbv}$, and $e_{nbv}|_{Qnbv,Pnbv}$ are independent with each other. Thus e_{bv} is independent with $I_{bv,rec}(i)$-$I_{nbv,rec}(j)$.

Appendix II

Based on the theory of Taylor Series, the constraint of equation (7) is that the function $f(Q_{t,bv}, Q_{t,nbv})$ should have partial derivatives at arbitrary orders.

To prove a function $f(x)$ is derivable, F_1 and F_2 (as shown in the following equations) must exist and equal.

$$\begin{cases} F_1 = \lim\limits_{\Delta x \to 0} \frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} \\ F_2 = \lim\limits_{\Delta x \to 0} \frac{f(x_0)-f(x_0+\Delta x)}{\Delta x} \\ F_1 = F_2 \end{cases} \tag{33}$$

Let x denotes $Q_{t,bv}$, y denotes $Q_{t,nbv}$. Therefore, for the function $f(Q_{t,bv}, Q_{t,nbv})$, F_1 could be written as,

$$\begin{aligned} F_1 &= \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \left\{ \begin{array}{l} \frac{E\left\{\left[I_{bv,rec}(i)|_{x_0+\Delta x} - I_{nbv,rec}(j)|_{x_0+\Delta x, y_0}\right]^2\right\}}{\Delta x} \\ -\frac{E\left\{\left[I_{bv,rec}(i)|_{x_0} - I_{nbv,rec}(j)|_{x_0,y_0}\right]^2\right\}}{\Delta x} \end{array} \right\}. \end{aligned} \tag{34}$$

Since the i^{th} pixel in the base view and the j^{th} pixel in the non-base view are the matching pixels, theoretically speaking, the pixel difference between them should be zero. Accordingly, statistically speaking, during the inter-view prediction of the encoder, the i^{th} pixel in the base view will be chosen to predict the j^{th} pixel in the non-base view. Thus, $I_{nbv,pred}(j)$ equals to $I_{bv,rec}(i)$. Accordingly, (34) could be written as,

$$\begin{aligned} F_1 &= \lim_{\Delta x \to 0} \left\{ \begin{array}{l} \frac{E\left\{\left[I_{bv,rec}(i)|_{x_0+\Delta x} - I_{bv,rec}(i)|_{x_0+\Delta x} - e_{y_0}\right]^2\right\}}{\Delta x} \\ -\frac{E\left\{\left[I_{bv,rec}(i)|_{x_0} - I_{bv,rec}(i)|_{x_0} - e_{y_0}\right]^2\right\}}{\Delta x} \end{array} \right\} \\ &= 0, \end{aligned} \tag{35}$$

where e_{y0} is the coding noise depends on $Q_{t,nbv}$.

Similarly, F_2 could also be proved to be zero. Furthermore, by using the same way, it could be proved that $f(Q_{t,bv}, Q_{t,nbv})$ have partial derivatives at arbitrary orders.

Acknowledgment

The authors would like to thank the editors and anonymous reviewers for their valuable comments. They would also like to thank Fraunhofer HHI, Berlin, Germany, GIST, Korea, ETRI, Daejeon, Korea, Nagoya University, Nagoya, Japan, and Poznan University, Poznan, Poland, for providing their 3-D video sequences and their valuable work on 3-DV.

References

[1] ISO/IEC JTC1/SC29/WG11, "Call for proposals on 3D video coding technology," in *Proc. 96th Meeting ISO/IEC JTC1/SC29/WG11*, Geneva, Switzerland, Mar. 2011, Art. ID N12036.

[2] K. Müller *et al.*, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3179–3194, Sep. 2013.

[3] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 9," in *Proc. 11nd Meeting ITU-T/ISO/IEC Joint Collaborative Team Video Coding (JCT-VC)*, Oct. 2012, Art. ID JCTVC-K1003.

[4] G. Tech, K. Wegner, Y. Chen, and S. Yea, "3D-HEVC test model 2," in *Proc. 2nd Meeting ITU-T/ISO/IEC Joint Collaborative Team 3D Video Coding (JCT-3V)*, Shanghai, China, Oct. 2012, Art. ID JCT3V-B1005_d0.

[5] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3DTV," in *Proc. Stereoscopic Displays Virtual Reality Syst., XI*, vol 93. San Jose, CA, USA, 2004, pp. 93–104.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] S. Hu, S. Kwong, Y. Zhang, and C.-C. J. Kuo, "Rate-distortion optimized rate control for depth map-based 3-D video coding," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 585–594, Feb. 2013.

[11] 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.

[12] F. Shao, G. Jiang, M. Yu, K. Chen, and Y.-S. Ho, "Asymmetric coding of multi-view video plus depth based 3D video for view rendering," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 157–167, Feb. 2012.

[13] Y. Liu *et al.*, "A novel rate control technique for multiview video plus depth based 3D video coding," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 562–571, Dec. 2011.

[14] 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, Nov. 2013.

[15] J. Xiao *et al.*, "Scalable bit allocation between texture and depth views for 3D video streaming over heterogeneous networks," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

[16] Joint Collaborative Team for 3DV. (2013, Mar.). *3D-HTM Software Platform* [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/

[17] 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.

[18] 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.

[19] *The Law of Larger Numbers*. (2013, Sep.) [Online]. Available: http://en.wikipedia.org/wiki/Law_of_large_numbers

[20] *Taylor Series*. (2013, Nov.) [Online]. Available: http://en.wikipedia.org/wiki/Taylor_Series

[21] H. Wang and S. Kwong, "Rate-distortion optimization of rate control for H.264 with adaptive initial quantization parameter determination," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 140–144, Jan. 2008.

[22] 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.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*, 6th ed. New York, NY, USA: Cambridge Univ. Press, 2008, pp. 152–153.

[24] Joint Collaborative Team for 3DV. (2013, Sep.). *3D Video Test Sequences* [Online]. Available: http://ftp.hhi.fraunhofer.de

[25] 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-3V)*, Geneva, Switzerland, Jan. 2013, Art. ID JCT3V-C1100.

[26] G. Bjontegaard, "Improvements of the BD-PSNR model," in *Proc. 35th Meeting ITU-T Video Coding Experts Group (VCEG)*, Berlin, Germany, Jul. 2008, Art. ID AI11.

[27] Y. Tong, F. A. Cheikh, H. Konik, and A. Tremeau, "Full reference image quality assessment based on saliency map analysis," *J. Imag. Sci.*, vol. 54, no. 3, pp. 30503-1–30503-14, May 2010.

[28] E. Ardizzone and A. Bruno, "Image quality assessment by saliency maps," in *Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, vol. 1, Feb. 2012, pp. 479–483.

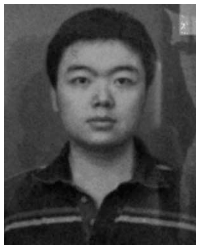
[29] Y. Tong, H. Konik, and A. Tremeau, "Color face-tuned salient detection for image quality assessment," in *Proc. 2nd Eur. Workshop Vis. Inf. Process. (EUVIP)*, Paris, France, Jul. 2010, pp. 253–260.

[30] Y. Tong, F. A. Cheikh, F. F. Guraya, H. Konik, and A. Tremeau, "A spatiotemporal saliency model for video surveillance," *Cogn. Comput.*, vol. 3, no. 1, pp. 241–263, Jan. 2011.

[31] L. Xiao, M. Johansson, H. Hindi, S. Boyd, and A. Goldsmith, "Joint optimization of communication rates and linear systems," *IEEE Trans. Autom. Control*, vol. 48, no. 1, pp. 148–153, Jan. 2003.

Hui Yuan (S'08–M'12) received the B.E. and Ph.D. degrees in telecommunication engineering from Xidian University, Xi'an, China, in 2006 and 2011, respectively. Since 2011, he has been a Lecturer with the School of Information Science and Engineering, Shandong University, Jinan, China. He is currently a Post-Doctoral Fellow with the Department of Computer Science, City University of Hong Kong, Hong Kong. His current research interests include video coding and multimedia communication.

Sam Kwong (SM'04–F'14) received the B.Sc. and M.Sc. degrees in electrical engineering from the State University of New York at Buffalo, Buffalo, NY, USA, and the University of Waterloo, Waterloo, ON, Canada, in 1983 and 1985, respectively, and the Ph.D. degree from the University of Hagen, Hagen, Germany, in 1996. From 1985 to 1987, he was a Diagnostic Engineer with the Control Data Canada, Mississauga, ON, Canada. He joined the Bell Northern Research Canada, Ottawa, ON, Canada, as a Scientific Staff Member and the City University of Hong Kong (CityU), Hong Kong, as a Lecturer with the Department of Electronic Engineering, in 1990. He is currently a Professor with the Department of Computer Science, CityU. His current research interests include video coding, pattern recognition, and evolutionary algorithms. He was an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS and the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.



Chuan Ge received the B.S. degree from Shandong Jianzhu University, Jinan, China, in 2007, and the M.S. degree from Software College, Shandong University, Jinan, in 2010, where he is currently pursuing the Ph.D. degree from the School of Information Science and Engineering. His research interests include 3-D video coding and image processing.

Xu Wang (S'11) received the B.S. degree from the South China Normal University, Guangzhou, China, the M.S. degree from Ningbo University, Ningbo, China, and the Ph.D. degree from the Department of Computer Science, City University of Hong Kong (CityU), Hong Kong, in 2007, 2010, and 2014, respectively. He is currently a Post-Doctoral Researcher with the Department of Computer Science, Shenzhen Research Institute, CityU. His research interests are video coding and stereoscopic image/video quality assessment.

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 Visiting Researcher and a Post-Doctoral Researcher with the Department of Computer Science, City University of Hong Kong, Hong Kong. In 2010, he joined Shenzhen Institutes of Advanced Technology, CAS. He has served as an Associate Professor since 2012. His research interests are 3D video coding, visual perception, and content-based video processing.