Fault-tolerance based Block-level Bit Allocation and Adaptive RDO for Depth Video Coding

Shengwei Wang and Li Yu
School of Electron. Inf. & Commun.
Huazhong Univ. of Sci. & Tech.
Wuhan, Hubei, China 430074
Email: {kadinwang, hustlyu}@hust.edu.cn

Sen Xiang
School of Inf. Sci. & Eng.
Wuhan Univ. of Sci. & Tech.
Wuhan, Hubei, China 430081
Email: xiangsen@wust.edu.cn

Zixiang Xiong
Dept. of Elect. & Comput. Eng.
Texas A&M University and Monash University
College Station, TX, USA 77843
Email: zx@ece.tamu.edu

***Abstract*— Depth videos affect the visual quality of virtual view greatly, while conventional encoders are not adapted to encode depth videos. To solve this problem, we propose, in the paper, a fault-tolerance based joint block-level bit allocation scheme for depth video coding. The scheme classifies depth blocks into two classes by the fault-tolerance, and constructs virtual view perceptual quality based rate-distortion (R-D) model for each class. Based on the model, a joint block-level bit allocation scheme is proposed to obtain the optimized quantization parameter (QP) to encode each class of blocks. Further, an adaptive RDO algorithm is designed to determine λ_{MODE}, in order to achieve better visual quality of virtual views. Experimental results demonstrate that, compared with 3D-HEVC, the proposed method improves the visual quality of virtual views, especially at preserving details and boundaries of objects.**

I. INTRODUCTION

Multiview video plus depth (MVD) [1] consists of texture and depth videos from multiple views. As the widely recognized 3D video representation [2], MVD has been applied in 3D applications, such as 3D television (3DTV) [3] and free-viewpoint television (FTV) [4]. Moreover, the standard of the 3D video coding, which is known as 3D Video Coding Extension of High Efficiency Video Coding (3D-HEVC) [5][6][7], was finalized in 2015. Numerous proposals aiming at the part of 3D video coding were proposed. Meanwhile, great efforts have been made on depth video coding. Different from texture videos, depth videos have sharp boundaries and large smooth regions. Depth videos are also not for display, but used to render virtual view videos by depth-image-based rendering (DIBR) [3]. Conventional coding standard, namely High Efficiency Video Coding (HEVC), are designed for coding texture videos, not considering the above properties of depth videos.

Since sharp boundaries of depth maps highly affect the quality of rendering, some researchers optimized depth video coding methods, especially at protecting boundary information. Daribo *et al.* [8] divided depth blocks by sharp edges, and proposed an arbitrarily shaped motion prediction for depth video coding, considering that neighboring pixels of similar depth have similar motion. Nguyen *et al.* [9] proposed to encode a low spatial resolution of depth maps at the encoder, so that the decoded depth maps can be up-sampled and interpolated using weighted mode filtering (WMF) in order to maintain the sharp boundaries. Extending the work in [10], Zhu *et al.* [11] considered that the conventional video coding may fail to predict the sharp edges of depth blocks, and proposed a new texture-aware depth inter-prediction method. The method incorporates pixel-oriented weighting in the bi-prediction process by exploiting motion and structure similarities between texture and depth videos. Lei *et al.* [12] proposed a depth-texture cooperative clustering-based prediction method in the intra-coding, which exploits the structure similarity for the current coding block and its neighboring pixels. After intra prediction, a detection and rectification approach is incorporated to induce the depth-texture misalignment. Above methods improve the coding efficiency and quality of depth videos by maintaining the sharp boundaries of compressed depth blocks. And yet, the relationship between depth videos and virtual view videos is out of consideration.

Based on the fact that compressed depth videos lead to the distortion of virtual views in DIBR, many methods have been developed to explore the relationship between depth videos and synthesized views. Liu *et al* [13] utilized the above fact, and constructed the rate-distortion (R-D) model between depth video bit-rate and distortion of virtual view. Based on the model, a frame-level bit allocation scheme is proposed. Shao *et al.* [14] studied the R-D models between depth bit-rate and distortion of virtual views, and extended the frame level rate control scheme in [15] to support MVD. Wang *et al.* [16] constructed a view synthesis distortion model to indicate the importance of each frame in the depth video. In his method, the view synthesis distortion model is incorporated in the bargain game theoretic model to handle the frame level bit allocation problem for Hierarchical B-picture (HBP) structure. Hu *et al.* [17] exploited the R-D properties among views to maximize the sum of the quality of reference texture videos and virtual view videos. The algorithm seeks an optimal combination of bit allocation among reference views, and allocates bits at frame level. Above proposed methods construct R-D models of the virtual view distortion, and develop a variety of methods to allocate bits at view or frame level. A block-level bit allocation scheme can be expected to achieve better performances.

978-1-5090-3649-3/17/$31.00©2017 IEEE

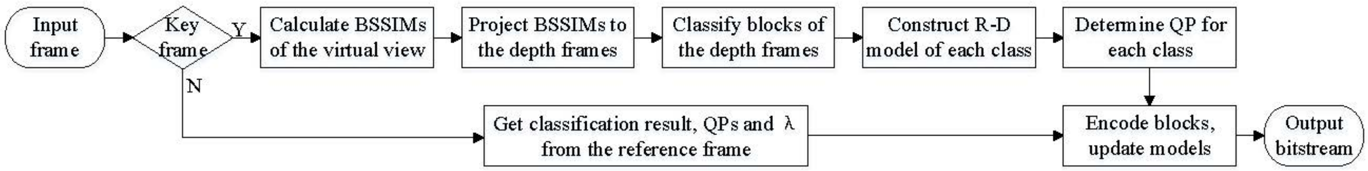


Fig. 1. Flowchart of the proposed bit allocation scheme.

In order to improve the coding quality of virtual views, we proposes, in this paper, a fault-tolerance based block-level bit allocation scheme. The scheme adopts the block distortion of virtual view to evaluate the fault-tolerance of depth blocks, by considering that same distortion in depth blocks can cause different quality loss in virtual view. In order to allocate the bits preciously, the depth blocks with different fault-tolerance are divided into two classes by an adaptive threshold. One is with high fault-tolerance, the other one is with low-tolerance. For each class of blocks, a synthesized view quality based R-D model is constructed. Based on the models, a joint block-level bit allocation scheme is designed to determine the optimized quantization parameter (QP) for each class of blocks. In order to further improve the quality of encoded virtual views, an adaptive rate distortion optimization (RDO) parameter λ_{MODE} is designed. Moreover, we have studied various visual-based image quality assessment methods [18][19][20], in order to ensure that the quality of virtual view coincides with the perceptual experiences. Finally, the structural similarity index metric (SSIM) is adopted to evaluate the quality of virtual view, for its high consistence with the subjective quality [18]. Experimental results demonstrate that the proposed method improves the quality of rendered virtual view, especially at preserving details and boundaries.

The paper is organized as follows: Section II briefly introduces the concept of block-based SSIM. Section III explains the proposed block-level bit allocation scheme in detail. The adaptive RDO is presented in Section IV. Section V shows the experimental results. Finally, the conclusion is drawn out in Section VI.

II. Concept of Block-based SSIM

SSIM consists of three parts including luminance, contrast, and structure. The combination of these parts makes SSIM adapt to human-visual system well. Thus, we adopts SSIM to evaluate the quality of the virtual view. In the paper, SSIM is used to calculate the quality of block of virtual view. The block-based SSIM (BSSIM) is as follows:

$$S(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \cdot \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (1)$$

where C_1, C_2, and C_3 are small constant numbers. x and y respectively denote the distorted and reference blocks at the same position. μ_x, μ_y represent the average values of all the pixels from x and y respectively. σ_x, σ_y are the autocorrelation coefficients, and σ_{xy} is the cross-correlation coefficient. BSSIM of x is denoted as $S(x,y)$.

Correspondingly, BSSIM distortion of x is $DS(x,y)$, which is defined as follows:

$$DS(x,y) = 1 - S(x,y) \quad (2)$$

III. Virtual View SSIM-based Bit Allocation

The whole process of the proposed bit allocation scheme is as Fig. 1 shows. For an input frame, the scheme first judges whether it is a key frame. A key frame is firstly pre-encoded and used to render the virtual view by DIBR. Then BSSIM of each block in the virtual view is calculated. Next, BSSIMs of the virtual view are inverse 3D-warped to the two reference depth frames to evaluate the fault-tolerance of the corresponding depth blocks. Then, these measured depth blocks are classified into two categories by an adaptive threshold. The blocks with low fault-tolerance are considered as complex blocks, while others are classified into simple ones. Next, the R-D models are constructed for the two classes of blocks respectively, and a joint block-level bit allocation algorithm is designed to determine the QPs for the two classes. Finally, the scheme encodes the key frame with the optimized QPs, and updates two R-D models.

While the input frame is not the key frame, the classification result and optimized encoding parameters from the last encoded key frame are directly utilized to encode the current frame. In detail, the non-key frame is set as P-frame, and the QP for the current block can be obtained from the encoded reference block in previous frames via the process of the motion estimation. With the motion estimation, the optimized QP can be passed from one frame to the next in P frames, until next key frame appears. After encoding the current frame, the parameters of two R-D models are also updated.

A. Adaptive Classification

In the proposed scheme, the size of group of pictures (GOP) is set as 8, the first frame in every GOP is the I-frame and the key frame. For every key frame, pre-encoding generates compressed depth images. The compressed depth and uncompressed texture images from the reference view V_1 and V_2 synthesize the distorted virtual view image in the virtual view V_0, as Fig. 2 shows. After pre-encoding, the BSSIMs of the virtual view image are calculated as shown in I_0. Darker blocks have smaller BSSIMs with larger BSSIM distortion DS, while lighter blocks have smaller DS.

Since depth images are used to render virtual views, the quality of reference depth images is measured with BSSIMs of the synthesized view. Specifically, BSSIMs in I_0 with the distortion DS are projected to the reference depth images in

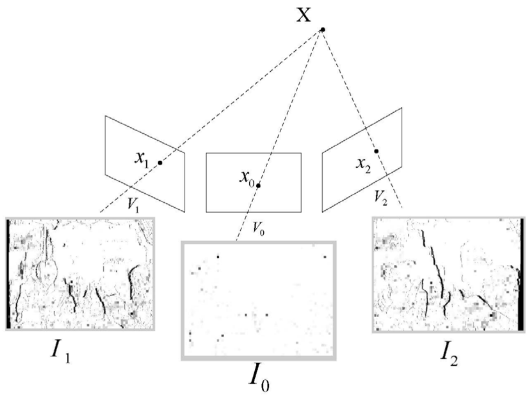
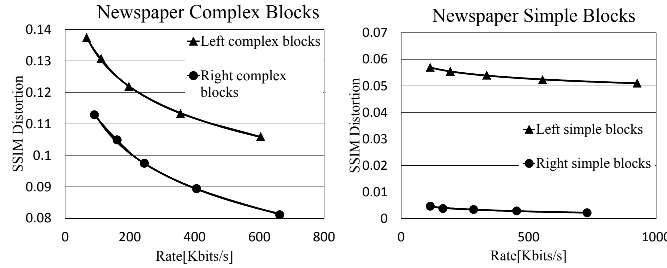


Fig. 2. BSSIMs of the synthesized view image is projected to the reference depth images by reverse 3D-warping



(a) Complex blocks (b) Simple blocks

Fig. 3. R-D curves of complex and simple blocks in sequence 'Newspaper'

V_1 and V_2, in order to evaluate the quality of corresponding depth blocks. The results of evaluation are shown as I_1 and I_2. Same as I_0, darker blocks have larger DS, the distortion of BSSIM, while lighter blocks are opposite. The process of projection is exactly as x_0 in V_0, which is the projection of X of the world coordinate, is respectively projected to x_1, x_2 in V_1 and V_2 by reverse 3D-warping.

All of the depth blocks are encoded with the same QP in the process of pre-encoding, however, the blocks lead to different quality loss in the virtual view. The blocks, which has low fault-tolerance, cause server distortion in the virtual view, while blocks with high fault-tolerance only lead to slight quality loss. Thus, the depth blocks with large DS can be treated as blocks with low fault-tolerance, and the blocks with small DS have high fault-tolerance.

When encoding depth blocks, bits should be allocated to blocks by the fault-tolerance in order to ensure the high quality of virtual views. The blocks with low fault-tolerance should consume more bits to achieve high quality of virtual views, while the blocks with high fault-tolerance can be allocated less bits with negligible quality loss. Thus, a threshold is set to classify the depth blocks into two classes, which makes it convenient to allocate bits by the fault-tolerance of blocks. The blocks with higher DS than the threshold are regarded as complex blocks, which have low fault-tolerance. The others are treated as simple blocks with high fault-tolerance. In order to make the threshold adapt to diverse sequences and QPs well, the threshold is adaptively set as the average value of DS of all the blocks in the virtual view.

B. R-D Model of Blocks

After classification, in order to construct R-D models for two classes of depth blocks, R-D curves for both two classes are studied firstly. Within each class, the distortion is the average projected DS of all the blocks in the class, and the bit-rate is the total bit-rate needed to encode all the blocks in the class. As an example, Fig. 3(a) and Fig. 3(b) illustrate the R-D curves of complex and simple blocks respectively. As the aforementioned analysis, the curve of simple blocks with high fault-tolerance keeps stable, while complex blocks with low fault-tolerance are more sensitive to bit-rate changes.

By studying various sequences, we find that the R-D curves of two classes of blocks in different sequences are similar to Fig. 3, and the curves can be fitted with logarithmic functions. For complex and simple blocks, the fitting equations are as follows respectively:

$$DS_c = A_c \ln R_c + B_c \tag{3}$$

$$DS_s = A_s \ln R_s + B_s \tag{4}$$

R denotes the total bit-rate. A, B are coefficients. Subscript c and s denote complex blocks and simple blocks respectively.

C. Block-based Joint Bit Allocation

For the complex and simple blocks, the statistical results from various sequences show the following R-Q relationships:

$$R_c = k_c e^{l_c Q_c} \tag{5}$$

$$R_s = k_s e^{l_s Q_s} \tag{6}$$

Q denotes the QP which is used to encode blocks. k, l are the coefficients which are determined by the sequence.

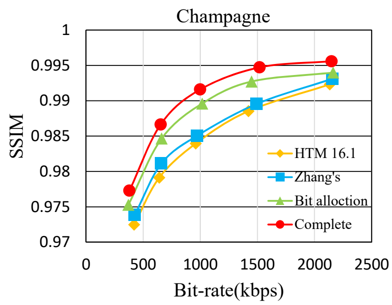
After obtaining R-D and R-Q models for each class of blocks, the joint block-level bit allocation can be transformed to the following optimization problem:

$$\begin{cases} \{Q_c^*, Q_s^*\} = \arg\min(DS_c + DS_s) \\ R_c + R_s \leq R_T \end{cases} \tag{7}$$

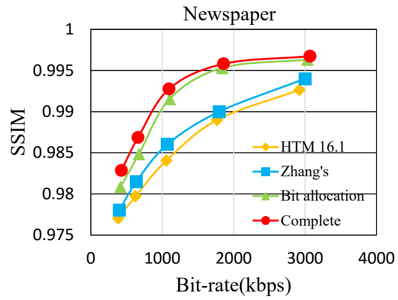
Q_c^* and Q_s^* are the optimized QPs to encode the complex and simple blocks respectively. R_T is the target bit-rate.

Lagrangian multiplier is used to solve Eq. (7) [21]. To minimize the R-D cost J, Eq.(7) is expressed as:

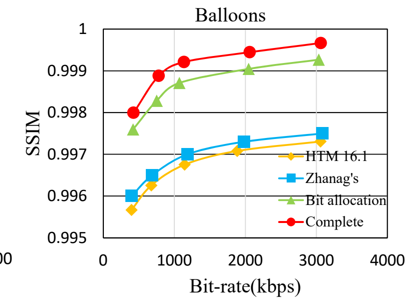
$$\begin{aligned} \{Q_c^*, Q_s^*, \lambda^*\} &= \arg\min J \\ &= \arg\min(DS_c + DS_s \\ &\quad + \lambda(R_c + R_s - R_T)) \end{aligned} \tag{8}$$



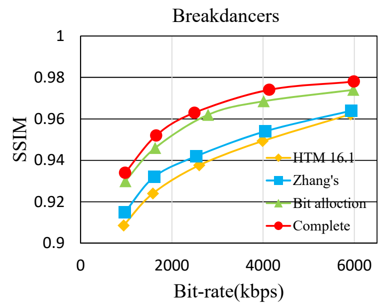
(a) Champagne sequence



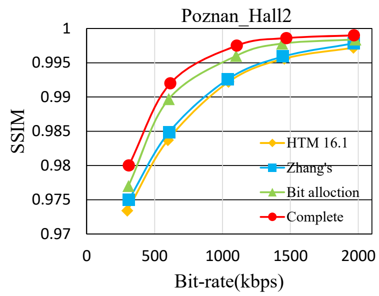
(b) Newspaper sequence



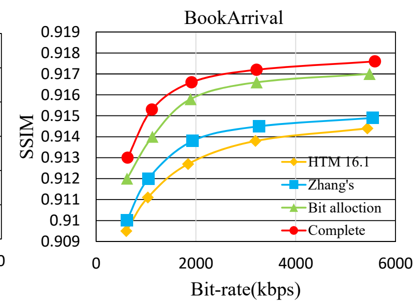
(c) Balloons sequence



(d) Breakdancers sequence



(e) Poznan_Hall2 sequence



(f) BookArrival sequence

Fig. 4. R-D curves comparing the proposed complete method with adaptive RDO, the proposed bit allocation scheme, Zhang's method and HTM 16.1. Notes: x-axis is the total bit-rate for two reference depth videos; y-axis is SSIM of the synthesized view.

Set the partial derivatives of J to Q_c, Q_s, and λ as zero. The optimized Q_c^* and Q_s^* for the complex and simple blocks are derived. The results are as follows:

$$\begin{cases} \lambda = -\frac{A_c+A_s}{R_T} \\ Q_c^* = \frac{1}{l_c}\ln(-\frac{A_c}{\lambda k_c}) \\ Q_s^* = \frac{1}{l_s}\ln(-\frac{A_s}{\lambda k_s}) \end{cases} \tag{9}$$

From Eq. (9), it is obviously to see that Q_c^* is smaller than Q_s^*, which means that the complex blocks need more bits to encode for their bad fault-tolerance, and less bits can meet the demand of the simple blocks.

IV. RDO FOR DEPTH VIDEO CODING

Based on the proposed R-D models, an adaptive RDO algorithm is designed for each class of blocks. Adjusting λ_{MODE} via RDO makes depth blocks choose the best prediction mode to achieve high SSIM and maintain the detailed texture information of virtual views. The proposed joint bit allocation scheme provides block-level R-Q and R-D modes for the two classes of depth blocks. By utilizing constructed models, the proposed RDO scheme is able to make λ_{MODE} adaptively fit various depth videos with different QPs.

To be specifically, for each class of blocks, according to Eq. (3)-(6), the following equation can be derived [22], and the subscript is omitted for simplification:

$$\begin{aligned} \frac{1}{\lambda_{MODE}} &= -\frac{dR}{dDS} \\ &= -\frac{dR}{dQ}\cdot\frac{dQ}{dDS} \\ &= -\frac{ke^{lQ}}{A} \end{aligned} \tag{10}$$

So the novel adaptive λ_{MODE} is as follows:

$$\lambda_{MODE} = -\frac{A}{ke^{lQ}} \tag{11}$$

In Eq. (11), it obviously shows that the property and QP of each class of depth blocks adaptively determine the optimized λ_{MODE}. In practical, the λ_{MODE} can refresh with the update of R-Q and R-D modes.

V. EXPERIMENTAL RESULTS

The experiments are performed on six 3D video sequences, and two views are selected as the reference views to synthesize one virtual view. The detailed information of sequences and views is listed in Table I. In the experiments, GOP is set as 8 (IPPPPPPP), and QP of 3D-HEVC is set as 26, 31, 36, 41, 46 respectively.

In the experiments, the proposed method are compared with 3D-HEVC and Zhang's method [23]. The reference software of 3D-HEVC is HTM 16.1. In Zhang's method, a full reference synthesized video quality metric (SVQM) was proposed to measure the perceptual quality of the synthesized video. Based on the SVQM, an RDO algorithm is developed to minimize

TABLE I
EXPERIMENTAL SEQUENCES

Sequences	Left View	Right View	Virtual View
Champagne	View 37	View 39	View 38
Newspaper	View 4	View 6	View 5
Balloons	View 1	View 3	View 2
Breakdancers	View 3	View 5	View 4
Poznan_Hall2	View 5	View 7	View 6
BookArrival	View 6	View 10	View 8

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

Fig. 5. Synthesized images of 3rd frame of sequence 'Newspaper' (QP=36). Note: (a) is encoded with HTM 16.1, (b) is encoded with Zhang's method, (c) is encoded with the proposed joint bit allocation scheme, (d) is encoded with the proposed complete method with RDO

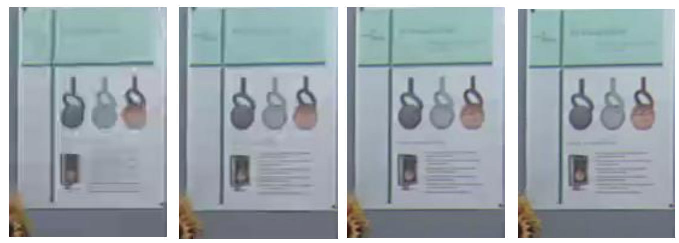
the perceptual distortion of synthesized view at a given bit-rate. The proposed method in the paper is set as two scheme. One scheme only includes the proposed joint block-level bit allocation scheme. The other one is the complete method consisted of the bit allocation scheme and the adaptive RDO algorithm. The R-D curves of six test sequences are shown in Fig. 4. SSIM is calculated by comparing the luminance component of the virtual view. Bit-rate denotes the total bit-rate used to encode two reference depth videos from left and right views. The curve 'HTM 16.1' shows the performance of the reference software of 3D-HEVC. The curve 'Zhang's' represents the performance of the proposed method in [23] by Zhang *et al.*. Moreover, the curve 'Bit allocation' denotes the performance of the proposed joint block-level bit allocation scheme without RDO. The complete method with the adaptive RDO algorithm is shown in the curve 'Complete'.

The figure shows that the proposed method has better SSIM performance than the other two methods, especially with the adaptive RDO algorithm. The proposed bit allocation scheme classifies the depth blocks into two classes by the fault-tolerance, and constructs R-D models, which connects the depth bit-rate and the perceptual quality of synthesized views. The model based bit allocation scheme ensures that the bits are preciously allocated by the fault-tolerance at block level. Thus, the proposed scheme achieves better performance. By adding the adaptive RDO algorithm, the proposed complete method chooses the best predicting mode to maintain the detailed edge information of virtual views, further improving the perceptual quality of virtual views.

The subjective results coincide with the R-D curves in Fig. 4, as Fig. 5-7 shows. The proposed method preserves boundaries well and obtains better viewing experiences, especially the letters on books in Fig. 5, the handrail in Fig. 6, and the words on the poster in Fig. 7. The experimental results demonstrate that the proposed method can truly improve the subjective quality of virtual views, and preserve the detailed texture information of objects, providing better experiences to viewers.

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

Fig. 6. Synthesized images of 14th frame of sequence 'Poznan_Hall2' (QP=31). Note: (a) is encoded with HTM 16.1, (b) is encoded with Zhang's method, (c) is encoded with the proposed joint bit allocation scheme, (d) is encoded with the proposed complete method with RDO



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

Fig. 7. Synthesized images of 8th frame of sequence 'BookArrival' (QP=41). Note: (a) is encoded with HTM 16.1, (b) is encoded with Zhang's method, (c) is encoded with the proposed joint bit allocation scheme, (d) is encoded with the proposed complete method with RDO

VI. CONCLUSION

In the paper, we proposed a fault-tolerance based joint block-level bit allocation scheme for depth video coding. The scheme applied the perceptual distortion of virtual view to describe the fault-tolerance of depth blocks. Then, the depth blocks were classified into two classes by an adaptive threshold, and R-D model of each class was constructed. Based on the models, the joint block-level bit allocation scheme was designed to optimize QPs for each class of blocks. Further, an adaptive λ_{MODE} was designed via RDO, and the proposed λ_{MODE} could change with various depth blocks and QPs. Experimental results demonstrated that the proposed method achieved the best perceptual quality of synthesized views, and subjective results showed that the proposed method provided better viewing experiences, especially at preserving object boundaries and details.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (NSFC) (No. 61231010), National High Technology Research and Development Program (No. 2015AA015903), Research Program of Hubei Provincial Department of Education (Q20171106).

References

[1] K. Müller, H. Schwarz, D. Marpe *et al.*, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3366–3378, 2013.

[2] C. Fehn, R. De La Barre, and S. Pastoor, "Interactive 3DTV concepts and key technologies," *Proc. of the IEEE*, vol. 94, no. 3, pp. 524–538, 2006.

[3] C. Zhu, Y. Zhao, L. Yu *et al.*, *3D-TV system with depth-image-based rendering*. Springer, 2014.

[4] M. Tanimoto, "FTV: free-viewpoint television," *Signal Process.: Image Commun.*, vol. 27, no. 6, pp. 555–570, 2012.

[5] G. Tech, K. Wegner, Y. Chen *et al.*, "3D-HEVC draft text 1," in *Proc. 5th Meeting of Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V)*, 2013.

[6] Y. Chen, G. Tech, K. Wegner *et al.*, "Test model 9 of 3D-HEVC and MV-HEVC," *Joint Collaborative Team on 3D Video Coding Extension Develop. of ITU-T SG*, vol. 16, 2014.

[7] G. Tech, Y. Chen, K. Müller *et al.*, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, 2016.

[8] I. Daribo, D. Florencio, and G. Cheung, "Arbitrarily shaped motion prediction for depth video compression using arithmetic edge coding," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4696–4708, 2014.

[9] V.-A. Nguyen, D. Min, and M. N. Do, "Efficient techniques for depth video compression using weighted mode filtering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 189–202, 2013.

[10] S. Li, J. Lei, C. Zhu *et al.*, "Pixel-based inter prediction in coded texture assisted depth coding," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 74–78, 2014.

[11] C. Zhu, S. Li, J. Zheng *et al.*, "Texture-aware depth prediction in 3D video coding," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 482–486, 2016.

[12] J. Lei, S. Li, C. Zhu *et al.*, "Depth coding based on depth-texture motion and structure similarities," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 2, pp. 275–286, 2015.

[13] G. Zhu, G. Jiang, M. Yu *et al.*, "Joint video/depth bit allocation for 3D video coding based on distortion of synthesized view," in *Proc. 2012 IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, 2012, pp. 1–6.

[14] F. Shao, G. Jiang, W. Lin *et al.*, "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, 2013.

[15] Z. G. Li, F. Pan, K. P. Lim *et al.*, "Adaptive basic unit layer rate control for JVT," in *Proc. JVT-G012-r1, 7th Meeting, Pattaya II, Thailand*, vol. 14, 2003.

[16] X. Wang, S. Kwong, H. Yuan *et al.*, "View synthesis distortion model based frame level rate control optimization for multiview depth video coding," *Signal Process.*, vol. 112, pp. 189–198, 2015.

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

[18] Z. Wang, A. C. Bovik, H. R. Sheikh *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[19] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1527–1542, 2014.

[20] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 591–602, 2014.

[21] Y. Zhang, S. Kwong, L. Xu *et al.*, "Regional bit allocation and rate distortion optimization for multiview depth video coding with view synthesis distortion model," *IEEE Trans. on Image Process.*, vol. 22, no. 9, pp. 3497–3512, Sep. 2013.

[22] K. Takagi, Y. Takishima, and Y. Nakajima, "A study on rate distortion optimization scheme for JVT coder," in *Visual Commun. & Image Process. 2003*. Int. Soc. Optics and Photonics, 2003, pp. 914–923.

[23] Y. Zhang, X. Yang, X. Liu *et al.*, "High-efficiency 3D depth coding based on perceptual quality of synthesized video," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5877–5891, 2016.