Joint Bit Allocation for 3D Video with Nonlinear Depth Distortion—An SSIM-Based Approach

Y. Harshalatha and Prabir Kumar Biswas

Abstract Perceptual quality improvement approach for 3D video through bit allocation is presented in this paper. Bit allocation between texture video and depth map plays an important role in deciding quality of synthesized views at the decoder end. To have better visual quality, structural similarity (SSIM) index is used as a distortion metric in rate distortion optimization (RDO) of the 3D video. In this paper, we used the nonlinear relationship of depth distortion with synthesis distortion in computing rate distortion cost resulting in better mode decision. Using the same depth map RDO in bit allocation algorithm, more accurate results are obtained when compared to the linear relation of depth distortion with synthesis distortion.

Keywords 3D video · Perceptual quality · Bit allocation

1 Introduction

The 3D video is a motion picture format that gives the real depth perception and thus gained huge popularity in many application fields. The depth perception is possible with two or more videos. With two views, the stereoscopic display produces a 3D vision and the viewer needs to wear special glasses. Multiview acquisition, coding, and transmission are necessary for an autostereoscopic display to provide 3D perception to the viewers without special glasses. Instead of multiple views, representation formats are used and only a few views along with the depth maps are coded and transmitted. Virtual view synthesis [2] is used to render intermediate views. These views are distorted and thus affect the end display.

Normally, rate distortion optimization (RDO) computes the rate distortion (RD) cost for a macroblock (MB). If distortion is measured using metrics like structural

Y. Harshalatha (✉) · P. K. Biswas
Indian Institute of Technology Kharagpur, Kharagpur, India
e-mail: harshalatha.y@ece.iitkgp.ernet.in

P. K. Biswas
e-mail: pkb@ece.iitkgp.ernet.in

B. B. Chaudhuri et al. (eds.), *Proceedings of 3rd International Conference on Computer Vision and Image Processing*, Advances in Intelligent Systems and Computing 1024, https://doi.org/10.1007/978-981-32-9291-8_12

similarity (SSIM) index, the quality of the reconstructed block matches the human vision. In [5], the authors extended the perceptual RDO concept to 3D video by considering the linear relationship between depth distortion and synthesis distortion. However, depth map distortion is nonlinearly related to view synthesis distortion, i.e., depth distortion remains the same whereas synthesis distortion varies according to the details in the texture video. In this paper, a suitable Lagrange multiplier is determined for RDO using nonlinear depth distortion. We used this framework of RDO in SSIM-based bit allocation algorithm [6].

Section 2 gives a brief explanation of view synthesis distortion and its relation with texture and depth distortion. In Sect. 3, SSIM index is discussed briefly. In Sect. 4, we discuss the RDO process with linear and nonlinear depth distortion. Bit allocation criteria for 3D video is explained in Sect. 5. Performance evaluation of the proposed method is discussed in Sect. 6 and concluding remarks are given in Sect. 7.

2 View Synthesis Distortion

3D video system shown in Fig. 1 has multiple videos as inputs and uses multiview plus depth (MVD) representation format that is more economical compared to other representation formats. View synthesis process generates a new intermediate view (target view) from the available texture views (reference views) and depth maps. It consists of two steps: warping and blending [10]. Warping is a process to convert the reference viewpoint to 3D point and then to target viewpoint. This pixel mapping is not one-to-one mapping, and thus holes are created. These holes are filled by the blending process. Warping uses depth data in converting a reference viewpoint to target viewpoint and accuracy of conversion depends on depth data. As lossy compression method is used in encoding depth map, it affects the warping process and causes distortion in the synthesized view.

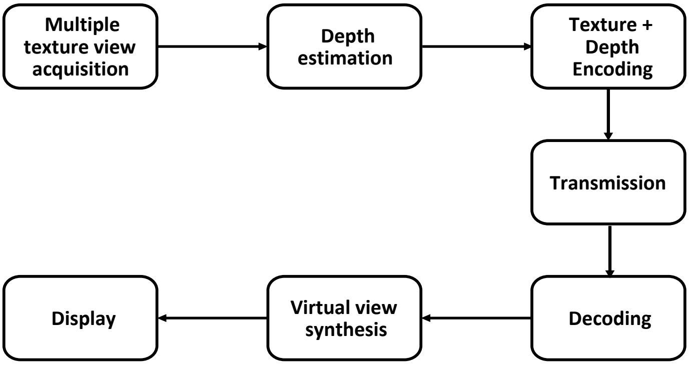

Fig. 1 Block diagram of 3D video system

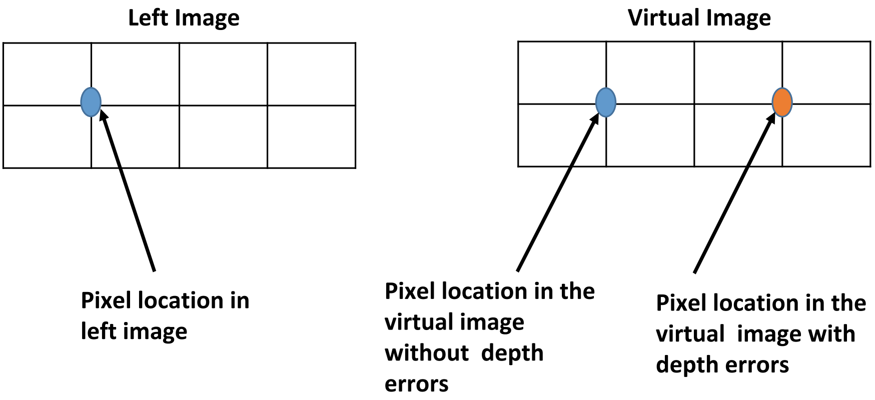
Left Image
Virtual Image
Pixel location in left image
Pixel location in the virtual image without depth errors
Pixel location in the virtual image with depth errors

Fig. 2 Geometric error due to inaccurate depth data

Distortion in synthesized views is mainly due to texture distortion and depth inaccuracy. Distortion model derived in [14] as well as in [8] assumed that texture and depth distortion (D_t and D_d) are linearly related to synthesis distortion (D_v) as in Eq. 1.

$$D_v = AD_t + BD_d + C \tag{1}$$

During the synthesis process, geometric errors will be minimum if the depth data is accurate. Inaccurate depth data leads to change in pixel position as shown in Fig. 2. This position error is linearly proportional to depth error as given in Eq. 2 [7].

$$\Delta P = \frac{f \cdot L}{255}\left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right) \tag{2}$$

where f represents camera focal length, L is the baseline distance, and Z_{near} and Z_{far} are nearest and farthest depth values, respectively. However, synthesis distortion depends on details in the texture video and is not the same in all the regions with same position error. This implies, even with the same amount of geometric error, degradation of synthesized view will be different for textured and textureless areas. Texture area with edge information will have more error compared to smooth regions. Considering these factors, synthesis distortion caused by depth distortion is formulated as in Eq. 3 [13].

$$D_{d\to v} = \Delta P \cdot D_d \cdot \left[D_{t_{(x-1)}} + D_{t_{(x+1)}}\right] \tag{3}$$

where $D_{t_{(x-1)}}$ and $D_{t_{(x+1)}}$ are the horizontal gradients computed between collocated texture blocks.

3 SSIM Index

Traditional methods for image quality measurement use objective evaluations and most of the metrics do not match with human visual characteristics. Human vision is sensitive to structural information in the scene and the quality metric must measure the structural degradation. Wang et al. [12] proposed structural similarity (SSIM) that measures structural degradation and thus evaluates according to human vision. For measuring SSIM, two images are required and measurement is done at the block level. For each block x and y, three different components, namely, luminance $(l(x, y))$, contrast $(c(x, y))$, and structure $(s(x, y))$ are measured and therefore SSIM is expressed as in Eq. 4.

$$SSIM(x, y) = f(l(x, y), c(x, y), s(x, y)) \tag{4}$$

Instead of similarity measure, we need distortion based on SSIM in RDO. Therefore, dSSIM is used and is defined as $dSSIM = \frac{1}{SSIM}$.

4 SSIM-Based RDO with Nonlinear Depth Distortion

Rate distortion optimization helps to reduce the distortion of reconstructed video with minimum rate while increasing the computation complexity. SSIM is used instead of sum of squared error (SSE) in mode decision and motion estimation to improve the visual quality. In our previous work [5], the linearity of texture and depth distortions with view synthesis and suitable Lagrange multiplier is determined as in Eq. 5.

$$\lambda_{new} = \frac{2\sigma_{x_i}^2 + C_2}{S_f \left(\exp(\frac{1}{M} \sum_{j=1}^{M} \log(2\sigma_{x_j}^2 + C_2)) \right)} \lambda_{SSE} \tag{5}$$

where S_f is the scaling factor, $\sigma_{x_i}^2$ is the variance of ith macroblock, M is the total number of macroblocks, and C_2 is constant to limit the range of SSIM.

Depth map RDO is performed by computing RD cost as in Eq. 6.

$$J_d = \Delta P \cdot D_{t_G} \cdot D_d + \lambda_{SSE} R_d \tag{6}$$

where R_d is the depth map rate and $D_{t_G} = D_{t_{(x-1)}} + D_{t_{(x+1)}}$ is horizontal gradient computed from texture video. For depth map RDO, Eq. 6 is minimized and Lagrange multiplier is derived as in Eq. 7.

$$\lambda_{i(d)} = \frac{\lambda_{new}}{S_f \cdot \kappa} \lambda_{SSE} \tag{7}$$

where $\kappa = \Delta P \cdot D_{tG}$.

5 Bit Allocation Algorithm

In 3D video, bit rate must be set properly between the views to improve the virtual view quality. In the literature [3, 9, 13–16], many joint bit allocation methods are proposed, and all these methods improve the PSNR of synthesized views. Visual quality enhancement can be achieved by using dSSIM as distortion metric ($dSSIM_v$) and bit allocation to improve SSIM is given by Eq. 8.

$$\min_{(R_t, R_d)} dSSIM_v \qquad s.t. \;\; R_t + R_d \le R_c \tag{8}$$

In terms of $dSSIM$, a planar model for synthesis distortion is determined as in Eq. 9.

$$dSSIM_v = a \cdot dSSIM_t + b \cdot dSSIM_d + c \tag{9}$$

Using SSIM-MSE relation, the distortion model in Eq. 9 is converted into Eq. 10.

$$dSSIM_v = \frac{a}{2\sigma^2{}_{x_t} + C_2} D_t + \frac{b}{2\sigma^2{}_{x_d} + C_2} D_d + z \tag{10}$$

$$dSSIM_v = p_1 D_t + p_2 D_d + c \tag{11}$$

where $p_1 = \frac{a}{2\sigma^2{}_{x_t} + C_2}$ and $p_2 = \frac{b}{2\sigma^2{}_{x_d} + C_2}$. Q_t (Eq. 13a) and Q_d (Eq. 13b), quantization steps of texture video and depth map determined by minimizing distortion-quantization model (Eq. 12).

$$\min \quad (p_1 D_t + p_2 D_d) \qquad s.t. \;\; (a_t Q_t^{-1} + b_t + a_d Q_d^{-1} + b_d) \le R_c \tag{12}$$

$$Q_t = \frac{a_t + \sqrt{\frac{K_1 a_t a_d}{K_2}}}{R_c - b_t - b_d} \tag{13a}$$

$$Q_d = \sqrt{\frac{K_2 a_d}{K_1 a_t}} Q_t \tag{13b}$$

where $K_1 = p_1 \alpha_t$ and $K_2 = p_2 \alpha_d$.

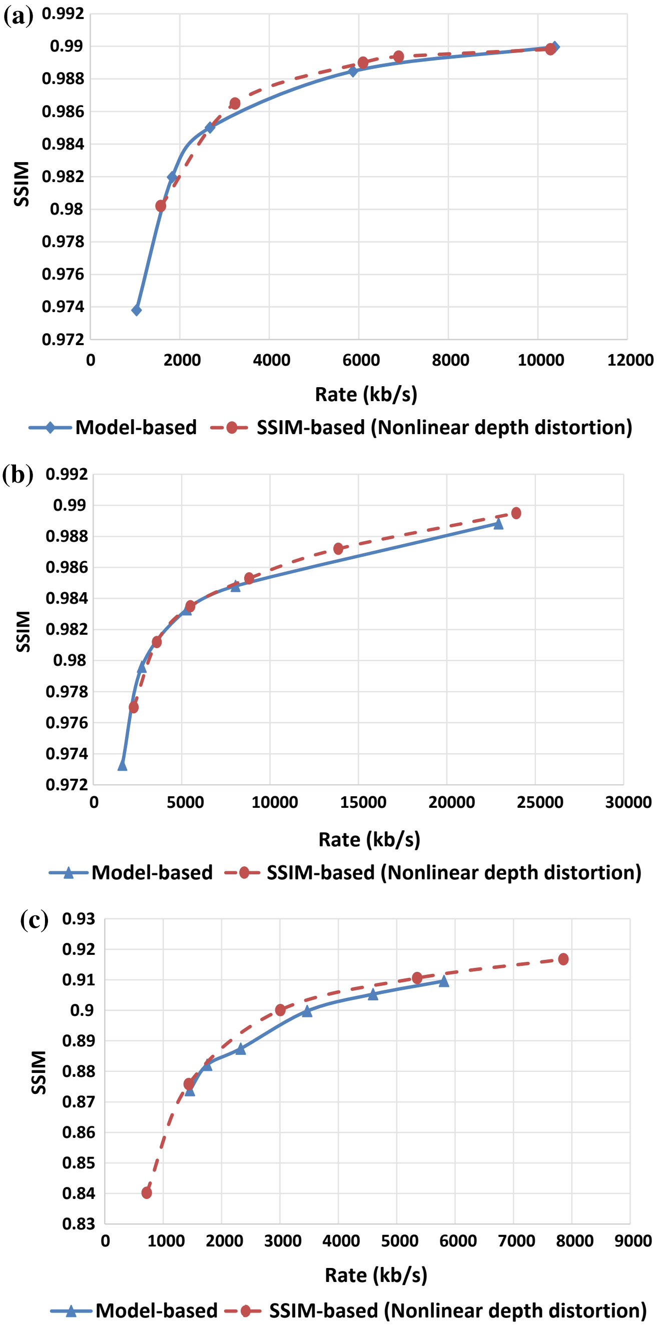

Fig. 3 SSIM-based bit allocation with nonlinear depth distortion **a** Kendo sequence, **b** Balloons sequence, and **c** Breakdancer sequence

Table 1 BD-rate comparison

Sequence	Proposed VS model-based bit allocation [14]		Proposed VS SSIM-based bit allocation [6]	
	ΔSSIM	ΔRate	ΔSSIM	ΔRate
Kendo	0.0003	−6.9196	0.0002	−3.8245
Balloons	0.0002	−2.8413	0.00004	−1.0327
Breakdancer	0.0034	−12.6183	0.0019	−7.1566

6 Results

We conducted experiments to check the performance of the joint bit allocation algorithm with nonlinear depth distortion. Encoding is done using 3DV-ATM reference software [1] and VSRS 3.0 [11] reference software is used for virtual synthesis. The test sequences used are Kendo, Balloons [4], and Breakdancer [17] with a frame rate of 30 frames/s. Kendo and Breakdancer sequences have 100 frames whereas Balloons sequence has 300 frames.

Experiments were conducted to evaluate SSIM-based bit allocation where nonlinear depth distortion is implemented in RDO. SSIM is computed between synthesized views and original views. For comparison, we utilized model-based algorithm of Yuan et al. [14] and Harshalatha and Biswas's algorithm [6]. RD curves in Fig. 3 give a comparison between our proposed algorithm and bit allocation with model parameters. Bjontegaard distortion-rate (BD-rate) calculations are done and tabulated in Table 1.

Further, bit allocation algorithm with linear and nonlinear effect of depth distortion RDO is compared as in Fig. 4 and also using BD-rate (Table 1). Nonlinear effect of depth distortion considered in RDO gives more accurate bit allocation results.

7 Conclusions

Rate distortion optimization improves the efficiency of an encoder and we proposed depth map RDO for 3D video by considering nonlinear relation of depth map distortion with view synthesis distortion. To improve the visual quality of synthesized views, dSSIM is used as distortion metric. Bit allocation algorithm is verified by using nonlinear depth distortion RDO and gives better performance over linear depth distortion RDO.

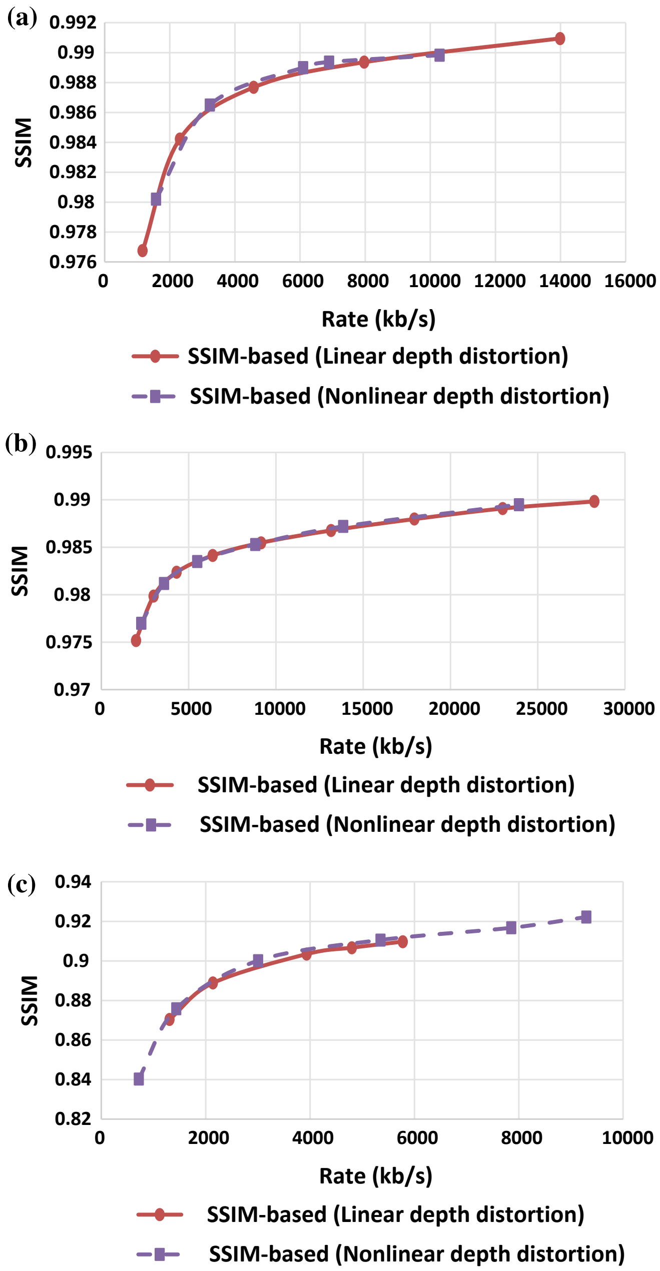

Fig. 4 SSIM-based bit allocation with nonlinear depth distortion compared with linear distortion **a** Kendo sequence, **b** Balloons sequence, and **c** Breakdancer sequence

References

1. 3DV-ATM Reference Software 3DV-ATMv5.lr2. http://mpeg3dv.nokiaresearch.com/svn/mpeg3dv/tags/3DV-ATMv5.1r2/. Accessed 29 July 2018
2. Fehn, C.: A 3D-TV approach using depth-image-based rendering (DIBR). In: Proceedings of VIIP, vol. 3 (2003)
3. Fehn, C.: Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In: Electronic Imaging 2004, pp. 93–104. International Society for Optics and Photonics (2004)
4. Fujii Laboratory, Nagoya University. http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/. Accessed 29 July 2018
5. Harshalatha, Y., Biswas, P.K.: Rate distortion optimization using SSIM for 3D video coding. In: 23rd International Conference on Pattern Recognition (ICPR), pp. 1261–1266. IEEE (2016)
6. Harshalatha, Y., Biswas, P.K.: SSIM-based joint-bit allocation for 3D video coding. Int. J. Multimedia Tools Appl. **77**(15), 19051–19069 (2018). https://doi.org/10.1007/s11042-017-5327-0
7. Kim, W.S., Ortega, A., Lai, P., Tian, D.: Depth Map Coding Optimization Using Rendered View Distortion for 3-D Video Coding (2015)
8. Shao, F., Jiang, G.Y., Yu, M., Li, F.C.: View synthesis distortion model optimization for bit allocation in three-dimensional video coding. Opt. Eng. **50**(12), 120502–120502 (2011)
9. Shao, F., Jiang, G., Lin, W., Yu, M., Dai, Q.: Joint bit allocation and rate control for coding multi-view video plus depth based 3D video. IEEE Trans. Multimed. **15**(8), 1843–1854 (2013)
10. Tian, D., Lai, P.L., Lopez, P., Gomila, C.: View synthesis techniques for 3D video. In: Applications of Digital Image Processing XXXII, Proceedings of the SPIE 7443, 74430T–74430T (2009)
11. View Synthesis Reference Software VSRS3.5. ftp://ftp.merl.com/pub/avetro/3dv-cfp/software/. Accessed 20 May 2018
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
13. Yang, C., An, P., Shen, L.: Adaptive bit allocation for 3D video coding. Circuits, Systems, and Signal Processing, pp. 1–23 (2016)
14. Yuan, H., Chang, Y., Huo, J., Yang, F., Lu, Z.: Model-based joint bit allocation between texture videos and depth maps for 3-D video coding. IEEE Trans. Circuits Syst. Video Technol. **21**(4), 485–497 (2011)
15. Yuan, H., Chang, Y., Li, M., Yang, F.: Model based bit allocation between texture images and depth maps. In: International Conference On Computer and Communication Technologies in Agriculture Engineering (CCTAE), vol. 3, pp. 380–383. IEEE (2010)
16. Zhu, G., Jiang, G., Yu, M., Li, F., Shao, F., Peng, Z.: Joint video/depth bit allocation for 3D video coding based on distortion of synthesized view. In: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1–6. IEEE (2012)
17. Zitnick, C.L., Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R.: High-quality video view interpolation using a layered representation. In: ACM Transactions on Graphics (TOG), vol. 23, pp. 600–608. ACM (2004)