

Optical Engineering

SPIDigitalLibrary.org/oe

View synthesis distortion model optimization for bit allocation in three-dimensional video coding

Feng Shao
Gang-Yi Jiang
Mei Yu
Fu-Cui Li

View synthesis distortion model optimization for bit allocation in three-dimensional video coding

Feng Shao, Gang-Yi Jiang, Mei Yu, and Fu-Cui Li
 Ningbo University, Faculty of Information Science and Engineering, Ningbo, 315211, China

Abstract. In this letter, a novel optimized view synthesis distortion model is proposed for bit allocation in three-dimensional video coding. The proposed model separates the view synthesis distortion into two independent terms, and the two terms are modeled respectively by quadratic distortion models. Finally, the optimal quantization parameters for texture and depth can be determined by minimizing the view synthesis distortion under the total bitrate constraint. Experimental results show that compared with a fixed 5:1 method, the proposed method can obtain higher view synthesis rate-distortion performance. © 2011 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.3665436]

Subject terms: view synthesis distortion; quadratic distortion model; bit allocation; 3D video coding.

Paper 110835LRR received Jul. 17, 2011; revised manuscript received Nov. 1, 2011; accepted for publication Nov. 14, 2011; published online Dec. 8, 2011.

1 Introduction

The three-dimensional (3D) video has drawn increasing attention as a new multimedia technique for capturing real-world scenes, which is expected as the ensuing evolution of the two-dimensional video.¹ The 3D video captures photorealistic texture as well as complex geometric shape of real-world scenes. Two types of 3D video applications are envisioned. Three-dimensional television (3DTV) aims to provide viewer depth perception of the scene by simultaneously rendering multiple images from different viewing angles.² Instead, free-viewpoint television (FTV) application allows for interactive selection of the viewpoint and direction in the scene within a certain operating range.³ In order to promote the 3DTV and FTV applications, multiview acquisition, multiview video coding (MVC), and virtual view rendering, etc., are being developed as key technologies.

However, the quality of the rendered virtual views can be affected by the compression of texture videos and depth maps.⁴ Given a maximum bitrate budget to represent the 3D scene, how to optimally distribute the bitrate between texture and depth, such that the rendering distortion is minimized, is still an open-ended problem. Results from the European ATTEST project had shown that the bitrate of depth map was fixed at a percentage of 20% of the texture bitrate, but it cannot guarantee that the bit allocation is optimal to practical multi-view video plus depth (MVD) data. Morvan et al. used

a full-search method to exhaustively search the optimal bitrate trade-off between texture and depth.⁵ Liu et al. proposed a view synthesis distortion model to seek the optimal bitrate trade-off between texture and depth.⁶ Kim et al. proposed a new distortion metric to quantify the effect of depth coding on synthesized view quality.⁷ Yuan et al. proposed a planar distortion model to solve the bit allocation problem by the Lagrangian multiplier method.⁸ However, the accurate relationship between the view synthesis distortion and the corresponding quantization needs to be found for selecting the optimal quantization parameters (QPs) for texture videos and depth maps.

In order to solve the bit allocation problem in 3D video coding with high accuracy, a novel optimized view synthesis distortion model is proposed in this letter. The proposed model separates the distortion into two independent terms: the distortion from the compression of texture videos and the distortion from the compression of depth maps. Then, the two terms of distortions are modeled, respectively, by quadratic distortion models. Finally, the bit allocation problem is solved by determining the optimal QPs for texture and depth in the view synthesis distortion equation.

2 View Synthesis Distortion Model Optimization

An accurate bit allocation model plays an important role in 3D video coding due to its efficiency in rendering quality and low complexity. The optimization is to find the associated quantization levels for texture videos and depth maps, such that the view synthesis distortion is minimized under the total bitrate constraint R_c

$$\begin{aligned} & \arg \min_{(R_t, R_d) \in Q} D_v \\ & s.t. R_d + R_t \leq R_c, \end{aligned} \quad (1)$$

where R_t and R_d are the coding bitrates of texture videos and depth maps, respectively, D_v is the view synthesis distortion, and Q is the candidate set of the bitrate pair.

In order to model the expression in Eq. (1), let S_v denote the virtual image synthesized by the original texture images and the original depth maps, \bar{S}_v denote the virtual image synthesized by the original texture images and the compressed depth maps, \tilde{S}_v denotes the virtual image synthesized by the compressed texture images and the original depth maps, and \hat{S}_v denotes the virtual image synthesized by the compressed texture images and the compressed depth maps. Our goal is to obtain a mathematical expression for the view synthesis distortion model as

$$\begin{aligned} D_v &= E\{[(S_v - \bar{S}_v) + (\bar{S}_v - \hat{S}_v)]^2\} \\ &= E[(S_v - \bar{S}_v)^2] + E[(\bar{S}_v - \hat{S}_v)^2] \\ &\quad + 2E[(S_v - \bar{S}_v)(\bar{S}_v - \hat{S}_v)]. \end{aligned} \quad (2)$$

Two hypotheses are assumed: 1. The term $E[(S_v - \bar{S}_v)(\bar{S}_v - \hat{S}_v)]$ is supposed to be negligible. 2. The term $E[(\bar{S}_v - \hat{S}_v)^2]$ can be approximated by $E[(S_v - \tilde{S}_v)^2]$. The two hypotheses are validated by experiments. Through experiments we find that the term in the first hypothesis is close to zero and the two terms of distortions in the second hypothesis are very similar. Thus, the error propagation problem between texture and depth can be omitted. Finally,

the following expression is obtained

$$D_v = \underbrace{E[(S_v - \bar{S}_v)^2]}_{D_d} + \underbrace{E[(S_v - \tilde{S}_v)^2]}_{D_t}. \quad (3)$$

The view synthesis distortion is separated into two independent terms: the distortion D_t from the term related to the distortion of the synthesized virtual view induced by texture compression, and the distortion D_d from the term related to the distortion of the synthesized virtual view induced by depth compression. Let TQ_{step} denote the quantization step of texture image, and DQ_{step} denotes the quantization step of depth map, the motivation of the proposed view synthesis distortion model is to find the mathematical relationship between distortion and quantization for texture and depth. Through experiments we find that when the texture quality is fixed, the relationship between the distortion D_d induced by depth compression and the DQ_{step} is approximately quadratic. Similarly, the relationship between the D_t induced by texture compression and the TQ_{step} can also be approximated by a quadratic model. Thus, the two terms in Eq. (3) are described by quadratic distortion models

$$D_d(DQ_{step}) \cong \alpha_d(DQ_{step})^2 + \beta_d(DQ_{step}) + \gamma_d, \quad (4)$$

$$D_t(TQ_{step}) \cong \alpha_t(TQ_{step})^2 + \beta_t(TQ_{step}) + \gamma_t, \quad (5)$$

where $\alpha_d, \beta_d, \gamma_d$ and $\alpha_t, \beta_t, \gamma_t$ are model parameters for depth and texture, respectively.

In order to characterize the rate-quantization (R-Q) relationship, a lot of R-Q models have been designed, such as linear R-Q model and quadratic R-Q model. Through experiment we find that the quadratic R-Q model can fit well to the actual R-Q curve.⁹ Thus, R_d and R_t can be expressed as

$$R_d(DQ_{step}) \cong \mu_d(1/DQ_{step})^2 + \nu_d(1/DQ_{step}) + \gamma_d, \quad (6)$$

$$R_t(TQ_{step}) \cong \mu_t(1/TQ_{step})^2 + \nu_t(1/TQ_{step}) + \gamma_t, \quad (7)$$

where μ_d, ν_d, γ_d and μ_t, ν_t, γ_t are model parameters for depth and texture, respectively.

In the experiments, we encode the original depth maps and texture videos with a small number of quantization steps, and fit the quadratic distortion models and quadratic R-Q models in the above formulation by using linear regression. Given the above formulation, the optimal view synthesis distortion problem can be converted to the optimization problem of TQ_{step} and DQ_{step} . Thus, Eq. (1) can be represented as

$$\begin{aligned} \arg \min_{(DQ_{step}, TQ_{step}) \in Q} & D_d(DQ_{step}) + D_t(TQ_{step}) \\ \text{s.t.} & R_d(DQ_{step}) + R_t(TQ_{step}) \leq R_c. \end{aligned} \quad (8)$$

To solve Eq. (8), we restrict the ranges of quantization step and perform simple searches to find the minimum view synthesis distortion cost. The basic idea is to construct a subset of all possible solutions, and try to find the approximate optimal solution. Thereby, the optimal quantization steps TQ_{step} and DQ_{step} can be calculated by solving the above equation, and the optimal QP_t for texture video coding and QP_d for depth map coding can be determined based on the relationship between QP and quantization step Q_{step} , $Q_{step} = 2^{(QP - 4)/6}$.

3 Experimental Results

In the experiments, we used four typical 3D video sequences of “Altmoabit,” “Bookarrival,” “Doorflower,” and “Leavelaptop” with size of 1024×768 . The 8th and 10th views in “Altmoabit” and “Doorflower” are used to synthesize the 9th virtual view. The 9th and 11th views in “Bookarrival” and “Leavelaptop” are used to synthesize the 10th virtual view. For all experiments, we used H.264 JM 17.2 video codec to independently encode texture videos and depth maps, and used VSRS 3.0 software to synthesize the virtual view.

To demonstrate the accuracy of the proposed model, we demonstrate the relationship between view synthesis distortion and quantization step in Fig. 1. It is obvious that the model for texture can well fit the actual changes because the distortions in texture videos can be directly propagated to the virtual views. The accuracy of the model for depth will be dropped slightly because distortions in depth maps may induce geometry changes in the virtual views, but the correlation coefficients between the actual distortions and the fitted values are all larger than 0.98.

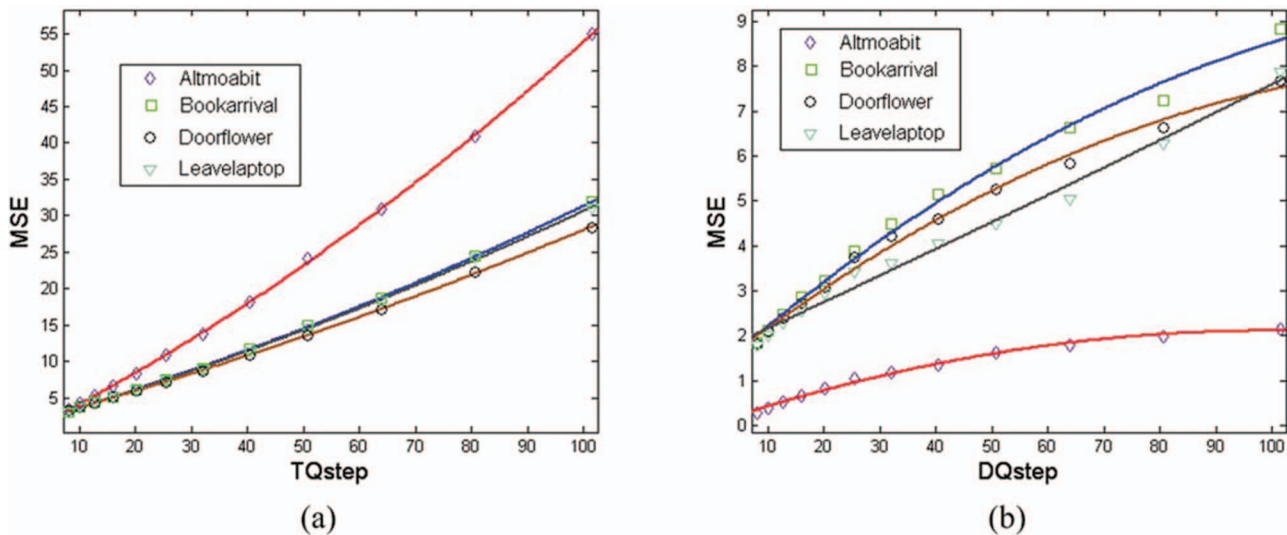


Fig. 1 The relationship between view synthesis distortion and quantization step. (a) texture; (b) depth.

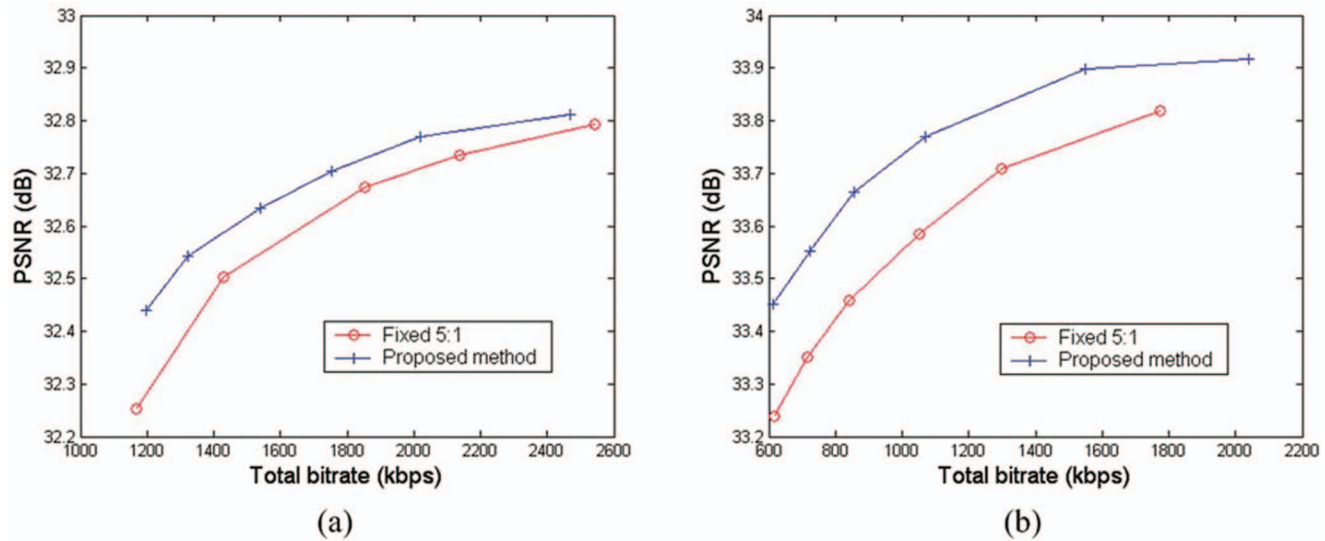


Fig. 2 View synthesis R-D performance comparison of different bit allocation methods. (a) Altmoabit; (b) Doorflower.

Then, by utilizing the obtained view synthesis distortion model, we can differentiate the view synthesis qualities at different texture and depth QPs without the actual encodings of texture videos and depth maps. We compared the view synthesis rate-distortion (R-D) performance of the proposed bit allocation method with a fixed ratio 5:1 bit allocation method. The R-D curves of different bit allocation methods for “Altmoabit” and “Doorflower” are shown in Fig. 2. As depicted by the results, compared with the fixed ratio 5:1 bit allocation method, more peak signal-to-noise ratio gains can be obtained by the proposed method.

4 Conclusions

In this letter, a novel optimized view synthesis distortion model is proposed. The proposed model separates the view synthesis distortion into two independent terms, and the two terms are modeled respectively by quadratic distortion models. The experimental results show the effectiveness of the proposed model. In the future probe, on the one hand, a more accurate model, such as cubic distortion model, is to represent the view synthesis distortion model. On the other hand, rate-distortion criterion for model decision of 3D video coding is to design by employing the proposed model.

Acknowledgments

This work was supported by the Natural Science Foundation of China (Grant Nos. 60902096, 60832003, and 61071120),

and the Specialized Research Fund for the Doctoral Program of Higher Education of China (20093305120002). It was also sponsored by K. C. Wong Magna Fund in Ningbo University.

References

1. A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang, “Three-dimensional video postproduction and processing,” *Proc. IEEE* **99**(4), 607–625 (2011).
2. C. Fehn, R. de la Barré, and S. Pastoor, “Interactive 3-DTV-concepts and key technologies,” *Proc. IEEE* **94**(3), 524–538 (2006).
3. M. Tanimoto, “Overview of free viewpoint television,” *Signal Process. Image Commun.* **21**(6), 454–461 (2006).
4. P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Muller, P. H. N. de With, and T. Wiegand, “The effects of multiview depth video compression on multiview rendering,” *Signal Process. Image Commun.* **24**(1–2), 73–88 (2009).
5. Y. Morvan, D. Farin, and P. H. N. de With, “Joint depth/texture bit-allocation for multi-view video compression,” in *Proceedings of Picture Coding Symposium*, Lisbon, Portugal (2007).
6. Y. W. Liu, Q. M. Huang, S. W. Ma, D. B. Zhao, and W. Gao, “Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model,” *Signal Process. Image Commun.* **24**(8), 666–681 (2009).
7. W. S. Kim, A. Ortega, P. L. Lai, D. Tian, and C. Gomila, “Depth map distortion analysis for view rendering and depth coding,” in *Proceedings of IEEE International Conference on Image Processing*, pp. 721–724, Cairo, Egypt (2009).
8. H. Yuan, Y. L. Chang, J. Y. Huo, F. Z. Yang, and Z. Y. Lu, “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).
9. H. J. Lee, T. Chiang, and Y. Q. Zhang, “Scalable rate control for MPEG-4 video,” *IEEE Trans. Circuits Syst. Video Technol.* **10**(6), 878–894 (2000).