

# When Disparity meets Distance: HEVC Compression of Double-Faced Depth Map

Yu-Hsun Lin\*, Ja-Ling Wu\*, Toshihiko Yamasaki† and Kiyoharu Aizawa†

\*Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan, R.O.C.

E-mail: lymanblue@cmlab.csie.ntu.edu.tw, wjl@cmlab.csie.ntu.edu.tw

†Department of Information and Communication Engineering,

Graduate School of Information Science and Technology, University of Tokyo, Tokyo, Japan

E-mail: yamasaki@hal.t.u-tokyo.ac.jp, aizawa@hal.t.u-tokyo.ac.jp

**Abstract**—A depth map is inherently double-faced, one single depth map can provide two closely related 3D-vision parameters (i.e. the disparity  $s$  and the distance  $z$ ). Most existing depth map compression techniques considered only one of them at a time, rather than addressing the effect of compression for the two parameters simultaneously. In order to remedy this shortage, a new distortion function is proposed and integrated into HEVC 3D compression framework to examine the effectiveness with respect to  $s$  and  $z$  at the same time. Experimental results show that, with the aid of the proposed distortion function, the rate-distortion performance of the distance quality will be significantly enhanced (from 0.5 dB to 8 dB), while keeping the corresponding quality of disparity almost the same. In addition to the coding performance improvements, we also discovered some interesting phenomena that never occurred in traditional compression framework. We expect this work can inspire more interesting research works on depth map compression.

## I. INTRODUCTION

The booming of low-cost depth cameras (e.g. Kinect) enabled numerous novel vision-based applications and made great impact on the computer vision research area. On the other hand, the advances of 3D display technologies and the great success of 3D movies inspired the developments of new 3D video data formats (e.g. multi-view plus depth [1]) and the HEVC-based 3D video compression standard. For a given depth map, we can utilize the physical distance,  $z$ , between an object and the depth camera for developing various vision-based applications. We can also incorporate the same depth map to compute the binocular disparity,  $s$ , for providing 3D content viewing. That is, one single depth map possesses two related but different physical quantities,  $s$  and  $z$ , which can be applied to different application domains.

The depth map compression is one of the core research topics in 3D video compression. The depth map has been applied to image based rendering process for virtual view synthesis [2], [3]. The analyses of rendering processing with depth information were conducted in [4], [5]. The research works [6]–[10] took the rendering results into account during the depth map compression.

The studies [11]–[13] of joint color/depth compression showed there is certain coding gain by utilizing the relations between the color image and the depth map. The depth map can also reflect the motion activities which was utilized for fast mode selection [14]. A depth map has more smooth regions

than that of the corresponding color image, where [15] defined them as the plate-based bases for conducting the depth map compression. The study [16] proposed a reconstruction filter for depth map coding by using an up/down sampling approach. The view scalable issue is a new dimension for scalability which was addressed in [17].

The existing compression techniques only addressed one aspect (e.g. disparity) of a depth map. However, there are two meaningful physical parameters in a depth map,  $s$  and  $z$ , which are inversely proportional to each other (i.e.  $s \propto \frac{1}{z}$ ). Figure 1 shows the relation between the disparity  $s$  and the distance  $z$ , where  $s_{\max}$  and  $s_{\min}$  correspond to  $z_{\min}$  and  $z_{\max}$ , respectively.

Due to the above mentioned double-faced characteristic, the compression of a depth map needs a new view point on the integrity of a data with multiple physical meanings. In order to address the qualities of disparity and distance simultaneously, we have to take both of them into consideration during the rate-distortion optimization procedure for conducting the depth map compression.

The rest of this paper is organized as follows. Section II introduces the computation of the disparity  $s$  and the distance  $z$  in a double-faced depth map. Section III describes the proposed method for depth map compression. Section IV presents the experimental results and Section V concludes this work and discusses the future works.

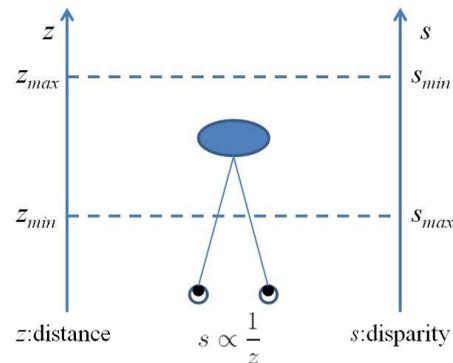


Fig. 1. The inversely proportional relation between the disparity  $s$  and the distance  $z$  of a depth map.

## II. THE DOUBLE-FACED DEPTH MAP

The distance  $z$  and the disparity  $s$  refer to the same current visible objects. The relation between the disparity  $s$  and the distance  $z$  is

$$s = \frac{f \cdot \Delta}{z}, \quad (1)$$

where  $f$  denotes the focal length and  $\Delta$  stands for the inter-eye distance.

From Eq. (1), the disparity  $s$  and the distance  $z$  are inversely proportional to each other. This relation brings challenges for double-faced depth map compression, i.e., compressing one expanding another.

For a given depth map  $I_d$ , the definition of the intensity values will depend on the application domain. For the newly developed 3D format [1], the disparity  $s$  is proportional to the intensity values of  $I_d$ :

$$\begin{aligned} s(p) &= f \cdot \Delta \cdot \frac{1}{255} \left( \frac{1}{z_{\min}} - \frac{1}{z_{\max}} \right) \cdot I_d(p) + \frac{1}{z_{\max}} \\ &= k \cdot I_d(p) + c, \end{aligned} \quad (2)$$

where  $p$  is the position for a given pixel,  $k$  and  $c$  are the constants that are defined by the maximum distance  $z_{\max}$ , the minimum distance  $z_{\min}$ , and the camera parameters (e.g. the focal length  $f$ ).

In order to address the qualities of  $s$  and  $z$ , independently of the camera parameters, we define the normalized disparity map  $I_s$  and the normalized distance map  $I_z$ , where the corresponding values of  $s$  and  $z$  are normalized into the range from 0 to 255.

### A. The Normalized Disparity Map

The normalized disparity map  $I_s$  is defined as

$$I_s(p) \stackrel{\text{def}}{=} I_d(p) \in [0, \dots, 255]. \quad (3)$$

The PSNR of the disparity  $s$ , obtained from the depth map, is denoted as S-PSNR and is computed as

$$\text{S-PSNR}(I_d, I'_d) = 10 \log_2 \left( \frac{255^2}{\text{MSE}(I_s, I'_s)} \right), \quad (4)$$

where  $I'_d$  is the distorted depth map and  $I'_s$  is the corresponding normalized disparity map of  $I'_d$ ,  $\text{MSE}(I_s, I'_s)$  computes the mean squared error between  $I_s$  and  $I'_s$ .

### B. The Normalized Distance Map

Eqs. (1) and (2) show the value of distance  $z$  is inversely proportional to the intensity of depth map  $I_d$ . Therefore, we define the normalized distance map  $I_z$  as

$$I_z(p) \stackrel{\text{def}}{=} \left\lfloor \frac{255}{I_d(p) + 1} \right\rfloor \in [0, \dots, 255], \quad (5)$$

where  $\lfloor \cdot \rfloor$  denotes the floor rounding operation.

Let Z-PSNR represent the PSNR of the distance  $z$ , obtained from the depth map, which is computed as

$$\text{Z-PSNR}(I_d, I'_d) = 10 \log_2 \left( \frac{255^2}{\text{MSE}(I_z, I'_z)} \right), \quad (6)$$

where  $I'_z$  is the corresponding normalized distance map of the distorted depth map  $I'_d$ .

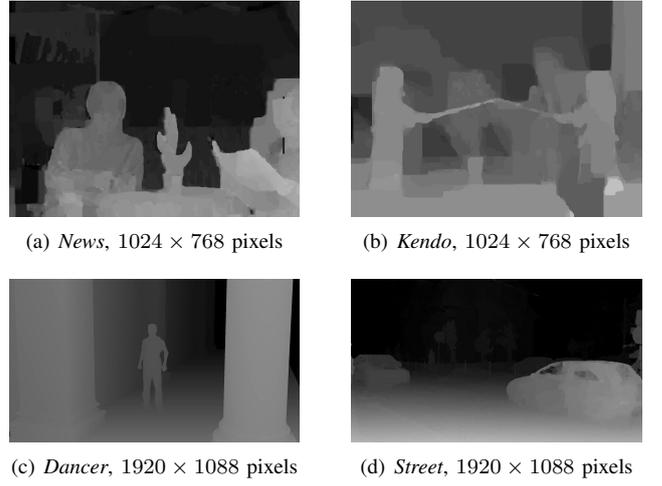


Fig. 2. The left-eye depth maps of the test data. The 2 view configurations [left-view-cam, right-view-cam] for the test sequences are (a) *News* [3,5], (b) *Kendo* [3,5], (c) *Dancer* [2,5], and (d) *Street* [4,3].

## III. PROPOSED METHOD

In order to address S-PSNR and Z-PSNR simultaneously, a new distortion function is proposed and integrated into the HEVC 3D compression framework.

### A. The Proposed Distortion Function

The newly proposed distortion function is defined as a linear combination of the distortions of the normalized disparity map and the normalized distance map, that is

$$D(I_d, I'_d) = w \cdot \text{SSE}(I_z, I'_z) + (1 - w) \cdot \text{SSE}(I_s, I'_s), \quad (7)$$

where  $D(I_d, I'_d)$  converges to the traditional distortion function when  $w = 0$ ,  $\text{SSE}(\cdot, \cdot)$  computes the sum of squared error between the original image and the distorted image.

### B. Rate-Distortion Optimization in HEVC

The rate-distortion optimization is addressed by introducing the Lagrange multiplier method which combines the distortion function  $D(I_d, I'_d)$  and the target bit rate  $R(I'_d)$  of the distorted depth map  $I'_d$  as

$$E = D(I_d, I'_d) + \lambda R(I'_d), \quad (8)$$

where  $\lambda$  is the Lagrange multiplier.

The rate-distortion optimization in HEVC compression will minimize the cost  $E$ , defined in Eq. (8), to find the optimal solutions in terms of distortion and bit rate considerations.

## IV. EXPERIMENTAL RESULTS

We integrate the proposed method in HEVC 3D compression reference software HTM 6.0. The settings of tested QP values are [25, 30, 35, 40, 45] for color image, and the corresponding QP values for the depth map are [34, 39, 42, 45, 48], as defined in the Common Test Conditions (CTC) of HEVC 3D compression experiments [18].

The 2-view configurations (i.e. left-eye, right-eye and the corresponding depth maps) are adopted in the experiments.

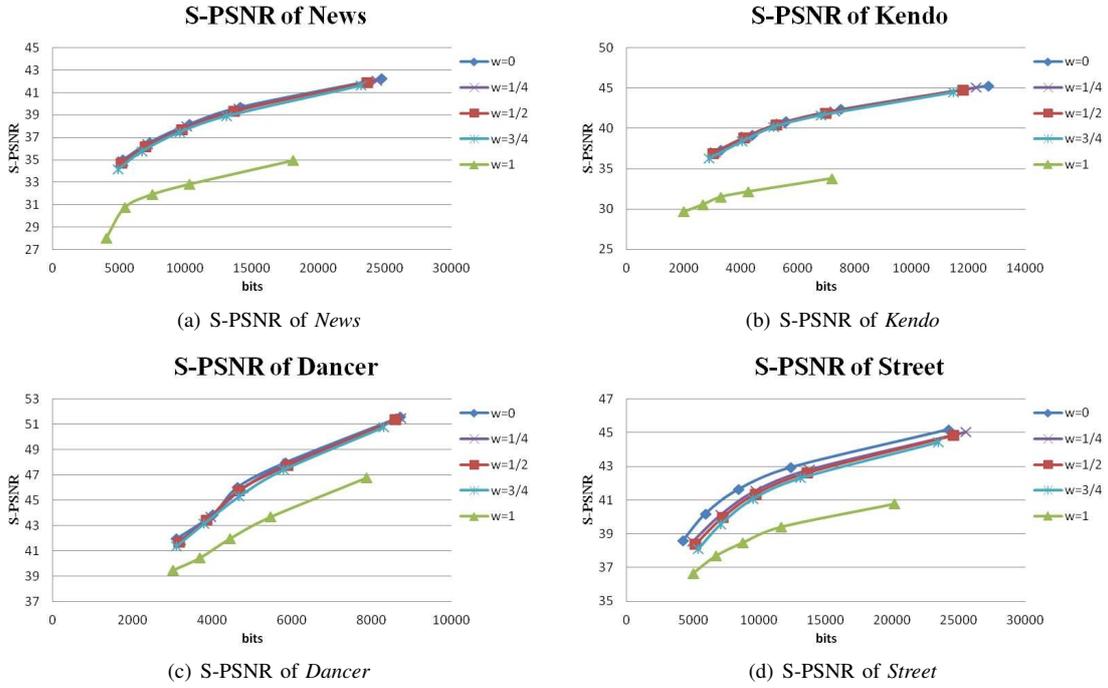


Fig. 3. The S-PSNR results of all test sequences, where  $w = 0$  corresponds to the traditional approach in HEVC 3D compression.

Since the depth map coding tools in HTM only address the quality of disparity value  $s$ , the depth map coding tools of HTM (e.g. wedgelets, simplified depth coding tree) are all disabled for evaluating the proposed distortion function during the experiments. We evaluate the performance of the proposed distortion function by setting the weights  $w$  in Eq. (7) to be  $\{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}$ , where  $w = 0$  corresponds to the case of traditional distortion function.

#### A. Test Data

Figure 2 demonstrates the test data used in our experiments. The resolutions are ranged from  $1024 \times 768$  to  $1920 \times 1088$  pixels. The 1-st frame of each video in each sequence (4 videos in a sequence for a 2 view configuration) are selected for experiments.

#### B. RD Performance of Disparity (S-PSNR)

Figure 3 shows the S-PSNR for each one of the test sequence, where S-PSNR's of the proposed distortion function are almost the same with the traditional one (i.e.  $w = 0$ ) when the weight  $w \in \{\frac{1}{4}, \frac{2}{4}, \frac{3}{4}\}$ . The setting of  $w = 1$  will decrease the S-PSNR since the distortion function in Eq. (7) only optimizes the quality of the distance  $z$  regardless of the quality of disparity  $s$ .

#### C. RD Performance of Distance (Z-PSNR)

Figure 4 demonstrates the significant improvements on Z-PSNR (from 0.5 dB to 8 dB PSNR improvement, or from 10% to 62% bit rate saving) by incorporating the proposed distortion function into consideration. There are some interesting phenomena occurred in the experimental results, which are addressed as follows.

#### • Non-monotonic RD Curve Behavior

The traditional distortion function (i.e.  $w = 0$ ) will introduce non-monotonic RD curve of Z-PSNR. This non-monotonic behavior can be explained by Eq. (5) that the normalized distance map  $I_z$  is inversely proportional to the intensity values of the depth map  $I_d$ . For example, for a given noise  $\Delta n$ , the sign (i.e. positive or negative) of the noise  $\Delta n$  will not affect the results of the traditional distortion function (i.e.  $w = 0$ ), that is

$$\text{SSE}(I_d, I_d + \Delta n) = \text{SSE}(I_d, I_d - \Delta n). \quad (9)$$

However, the sign of the noise  $\Delta n$  will have obvious impact on the results of the  $\text{MSE}(I_z, I'_z)$  in Eq. (6). Therefore, the proposed distortion function can reduce the non-monotonic behaviors as shown in Figure 4.

#### • The Effect of $w = 1$

The setting of  $w = 1$  may not always improve the result of Z-PSNR. The test sequence *Kendo* has large intensity values in  $I_d$  as compared with other test sequences, which leads to small  $\text{SSE}(I_z, I'_z)$  in Eq. (8). Therefore, HEVC compression will reduce the bit rate size instead of maintaining the quality of Z-PSNR.

## V. CONCLUSIONS

Different to the color image, depth map is a special information which possesses 2 different meanings. This double-faced characteristic of a depth map provides a new view point for developing the depth map compression. The importance of each meaning (i.e. the distance and the disparity) will depend on the application domains. The S-PSNR is important for viewing applications and the Z-PSNR is important in the

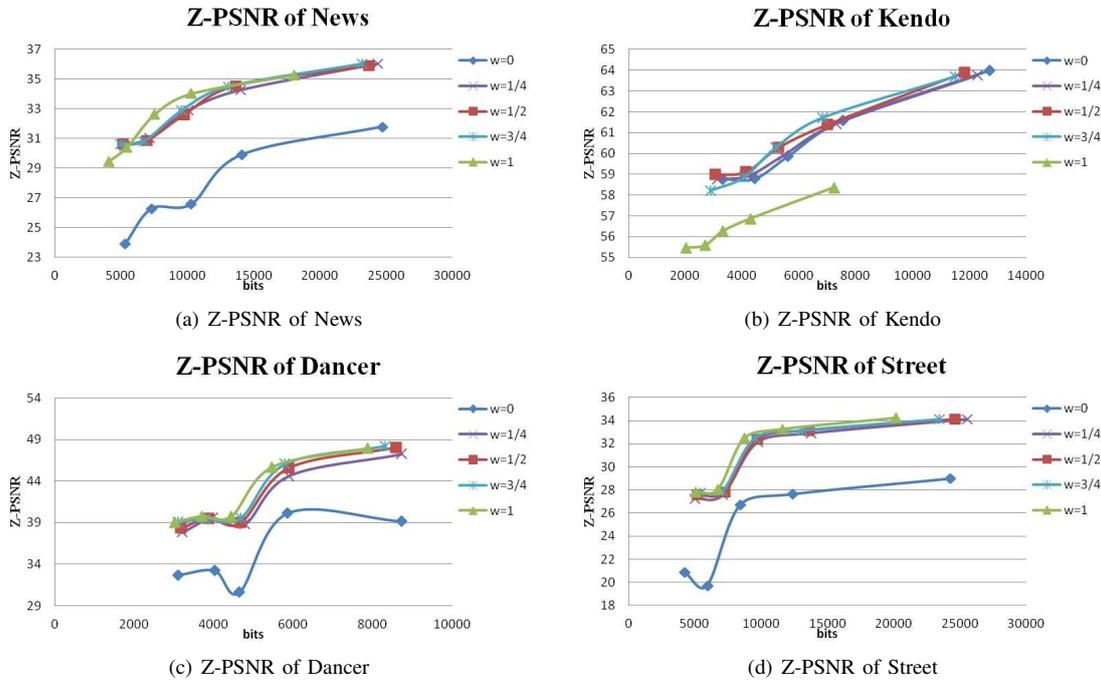


Fig. 4. The Z-PSNR results of all test sequences, where  $w = 0$  corresponds to the traditional approach in HEVC 3D compression.

computer vision related applications (e.g. action recognition). A new distortion function is proposed to address the qualities of the disparity  $s$  and the distance  $z$  simultaneously, in which S-PSNR remains almost the same as the traditional approach while the corresponding quality of Z-PSNR will be significantly improved (from 0.5 dB to 8 dB). Due to the inversely proportional relation between the disparity  $s$  and the distance  $z$ , there are some interesting discoveries (e.g. non-monotonic behavior in RD performance) of the double-faced depth map compression.

#### ACKNOWLEDGMENT

Yu-Hsun Lin is sponsored by the Graduate Students Study Abroad Program (102-2917-I-002-088) of NSC in Taiwan during his visiting in the University of Tokyo.

#### REFERENCES

- [1] K. Müller, P. Merkle, and T. Wiegand, "3-d video representation using depth maps," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 643–656, 2011.
- [2] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3d tv services providing interoperability and scalability," *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 217–234, 2007.
- [3] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3d tv," *IEEE Trans. on Broadcasting*, vol. 51, no. 2, pp. 191–199, 2005.
- [4] H. T. Nguyen and M. Do, "Error analysis for image-based rendering with depth information," *IEEE Trans. on Image Processing*, vol. 18, no. 4, pp. 703–716, 2009.
- [5] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, "Depth map distortion analysis for view rendering and depth coding," *IEEE ICIP*, pp. 721–724, 2009.
- [6] Y. Zhao, C. Zhu, Z. Chen, and L. Yu, "Depth no-synthesis-error model for view synthesis in 3-d video," *IEEE Trans. on Image Processing*, vol. 20, no. 8, pp. 2221–2228, 2011.
- [7] B. T. Oh, J. Lee, and D.-S. Park, "Depth map coding based on synthesized view distortion function," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1344–1352, 2011.
- [8] W. Hu, G. Cheung, X. Li, and O. Au, "Depth map compression using multi-resolution graph-based transform for depth-image-based rendering," *IEEE ICIP*, pp. 1297–1300, 2012.
- [9] Y.-H. Lin and J.-L. Wu, "Rendering lossless compression of depth image," *DCC*, p. 467, 2011.
- [10] G. Cheung, V. Velisavljevic, and A. Ortega, "On dependent bit allocation for multiview image coding with depth-image-based rendering," *IEEE Trans. on Image Processing*, vol. 20, no. 11, pp. 3179–3194, 2011.
- [11] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view video plus depth representation and coding," *IEEE ICIP*, vol. 1, pp. I–201–I–204, 2007.
- [12] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Müller, P. de With, and T. Wiegand, "The effects of multiview depth video compression on multiview rendering," *Signal Processing: Image Communication*, vol. 24, no. 1-2, pp. 73–88, 2009.
- [13] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "Joint video/depth rate allocation for 3d video coding based on view synthesis distortion model," *Signal Processing: Image Communication*, vol. 24, no. 8, pp. 666–681, 2009.
- [14] Y.-H. Lin and J.-L. Wu, "A depth information based fast mode decision algorithm for color plus depth-map 3d videos," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 542–550, 2011.
- [15] Y. Morvan, P. H. N. de With, and D. Farin, "Platelet-based coding of depth maps for the transmission of multiview images," *Proc. SPIE 6055, Stereoscopic Displays and Virtual Reality Systems XIII*, pp. 1–12, 2006.
- [16] K.-J. Oh, S. Yea, A. Vetro, and Y.-S. Ho, "Depth reconstruction filter and down/up sampling for depth coding in 3-d video," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 747–750, 2009.
- [17] S. Shimizu, M. Kitahara, H. Kimata, K. Kamikura, and Y. Yashima, "View scalable multiview video coding using 3-d warping with depth map," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1485–1495, 2007.
- [18] D. Rusanovskyy, K. Müller, and A. Vetro, "Jct3v-d1100: Common test conditions of 3dv core experiments," April 2013.