QUANTIZATION ERROR REDUCTION IN DEPTH MAPS

Ku-Chu Wei, Yung-Lin Huang, and Shao-Yi Chien

Media IC & System Lab
Graduate Institute of Electronics Engineering and Department of Electrical Engineering
National Taiwan University
1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan
{kcwei, cary}@media.ee.ntu.edu.tw, sychien@cc.ee.ntu.edu.tw

ABSTRACT

Since most depth maps are quantized to 8-bit numbers in current 3D video systems, the induced cardboard effects can disturb human perception. Moreover, depth maps with larger resolution suffer more from the quantization error. Therefore, this paper proposes an optimization approach to reduce the depth quantization error with well-preserved structure of the depth maps. The experimental results demonstrate that the proposed approach can successfully recover the structure characteristics from the quantized depth maps. Evaluation in mean square error (MSE) and mean structural similarity index (MSSIM) also strongly support our theory and algorithm. Through enhancing the quality of the depth maps from the very beginning, this work can benefit most 3D processing applications, such as 3D modeling, shape registration, and view synthesis.

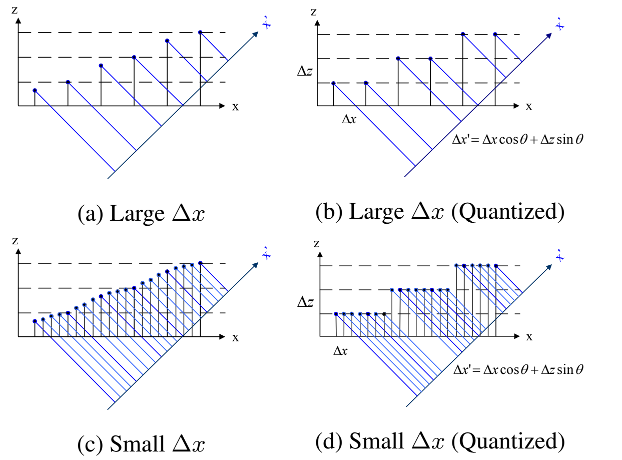
Index Terms— depth maps, 3D processing, quantization error, optimization

1. INTRODUCTION

With the improvement of depth sensors and stereo video systems, videos with depth maps enable various applications such as 3D reconstruction [1] and 3D TV [2]. Graphics models can be derived from multiple depth maps in 3D reconstruction [1]. However, depth maps often accompany sensor noise when capturing. Deriving accurate graphics models while reducing the negative effect of sensor noise attracts lots of researchers for a long time.

On the other hand, 3D TV provides realistic stereoscopic feeling using extra information from depth maps. The possibility of altering the viewpoint when playing back the video introduces amazing visual effects with stereo and autostereoscopic displays. View synthesis is essential for such an application, and depth-image-based rendering (DIBR) [3] is the most popular and well-grown technique for view synthesis, where new viewpoints can be synthesized via depth images. However, poor quality of depth maps leads to bad results for DIBR [4]. Unreasonable structure appears when watching such a video, resulting uncomfortable feeling.

Since these applications are sensitive to the quality of depth maps, lots of researches in the last decade focused on the depth refinement to derive good quality of depth maps. By imposing statistical methods with constraints, the quality of depth maps was upgraded [5][6], with little visible artifacts. However, all the techniques operate on the depth *images* only. That is, all the depth signals are recorded as depth maps, where the amplitudes are often represented by 8-bit numbers in current 3D video formats [2].



(a) Large Δx

(b) Large Δx (Quantized)

(c) Small Δx

(d) Small Δx (Quantized)

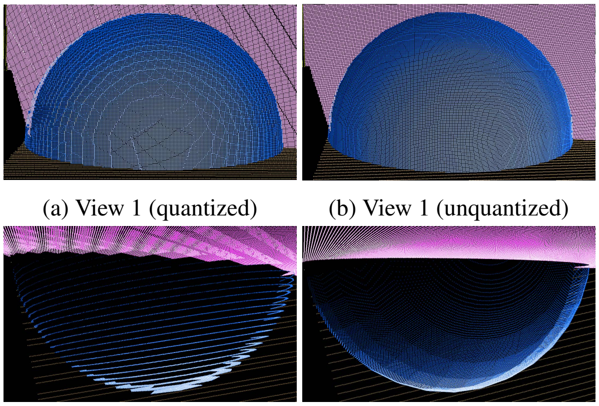
Fig. 1: Illustration of the effect from coarse quantization step Δz. (c)(d) With small Δx, artifacts due to coarse Δz becomes visible as the viewpoint altered.

Most researchers think quantization (sampling) in X-Y domains dominates the perception, since human visual system (HVS) has poor discriminability in depth than that in spatial domain. Consequently, the resolution (width×height) of depth maps becomes larger with the improving acquisition techniques, whereas depth values remain being recorded into 8-bit numbers. The gap between quantization step in spatial domain (sampling interval) and that in intensity domain for depth maps becomes much larger nowadays.

The big gap would result in perceivable artifacts, as illustrated in Fig. 1. Since the y-axis behaves similar to the x-axis does, we only illustrate and explain the signal $z(x)$ rather than $z(x, y)$ here. As shown in Fig. 1, view synthesis or altering viewpoints is equivalent to projecting the signal from x-axis to x'-axis. Assuming the spatial sampling interval is Δx and the quantization step is Δz, after projection, the distance of adjacent samples in x'-axis becomes

$$\Delta x' = \Delta x \cos\theta + \Delta z \sin\theta, \qquad (1)$$

where θ is the angle between x-axis and x'-axis, as shown in Figs. 1a and 1b. When the spatial resolution becomes larger, the spatial sampling interval Δx is getting smaller, whereas Δz remains the same. As θ is approaching to 90 degrees, altering a novel viewpoint, the value of $\Delta x'$ becomes dominated by Δz rather than Δx. That is, the coarse quantization step Δz will lead to low resolution or large sampling interval $\Delta x'$, as shown in Figs. 1c and 1d. Although HVS



(a) View 1 (quantized) (b) View 1 (unquantized)

(c) View 2 (quantized) (d) View 2 (unquantized)

Fig. 2: Visualization of the cardboard effect due to quantization. As altering viewpoint from view 1 to view 2, the noticeable cardboard effect appears.

may not be so sensitive in z-direction, the quantization step along z-direction now contributes to x'-direction as large viewpoint change occurs. The quantization effect in z-direction becomes perceptible to HVS, and the cardboard effect [4] is introduced accordingly. It is a new problem for the field of 3D video processing when the spatial resolution continues increasing. Another example is shown in Fig. 2. Figs. 2a and 2c are rendered with a quantized depth map, while Figs. 2b and 2d are rendered with an unquantized depth map. The cardboard effect is apparently perceivable in Fig. 2c.

Dithering [7] is a mature technique in audio and video processing that randomizes quantization error by applying noise intentionally, using the fact that random noise is less perceivable or objectionable than the harmonic tones. However, in current 3D video coding systems, only quantized depth maps can be derived in the application sides, while dithering can only deal with quantization or re-quantization given the original data. Reducing quantization error without knowing the original data becomes an ill-posed problem, which is the focus of this paper.

2. PROPOSED ALGORITHM

The stair-like quantized signal with high-frequency noise comes from the original signal, as shown in Fig. 3. Normally, low-pass filters can be applied to eliminate the stair-like shape. All high-frequency components, including the quantization error and the edge information, are removed in the meanwhile. The 3D perception can be severely influenced with the incomplete edge information. On the other hand, though the bilateral filters remove high-frequency noise while preserving edges, determining the proper size of the filter kernel is crucial. Filtering with a small-size kernel is not effective, whereas artifacts such as halo effects appears when filtering with a large-size kernel. Therefore, an optimization framework is proposed in this paper to conquer the quantization error.

Given the original signal I and the quantization step q, we assume round-off is used in quantization, and the quantized signal X can be accordingly modeled as:

$$-\frac{q}{2} \le I - X \le \frac{q}{2}. \tag{2}$$

As mentioned before, generally the quantization step (sampling interval) in spatial domain is much smaller than that in intensity domain for depth signals; that is, the signals should be smooth in spatial domain. The preferred estimated signal I' can be approximated by diagonal line segments ($\left\|\nabla^2 I'\right\|_1$) and horizontal line segments ($\|\nabla I'\|_1$). Meanwhile, we impose (2) to constrain the estimated signal I' being similar to I. In summary, the following energy function is utilized to recover signals from the quantized signals:

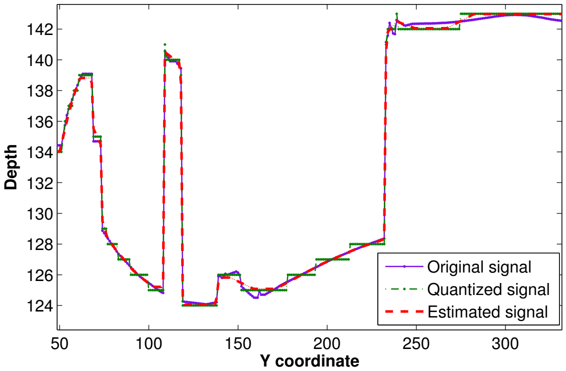


Fig. 3: An example 1D depth signal to verify (3). Compared to the quantized signal, the estimated signal is much similar to the original one.

$$\begin{aligned} I'^* = \underset{I'}{\operatorname{argmin}} \|X - I'\|_2 + \lambda_1 \left\|\nabla^2 I'\right\|_1 + \lambda_2 \|\nabla I'\|_1 \\ s.t. -\tfrac{q}{2} \le I' - X \le \tfrac{q}{2} \end{aligned} \tag{3}$$

A 1D signal is employed to demonstrate the performance of the proposed approach, as shown in Fig. 3. Since we impose the quantization constraint (2) during optimization, the proposed algorithm would maintain the critical high-frequency edge information.

3. EXPERIMENTS

Although (3) can be directly minimized via convex optimization [8], the high computation and memory requirements make that infeasible. In practice, a row-column decomposition is employed to reduce the high computation requirements by separating a 2D problem into multiple 1D problems. In the following experiments, we set $\lambda_1 = 1$ and $\lambda_2 = 0.1$, since better results can be derived by using diagonal line segments than using horizontal ones.

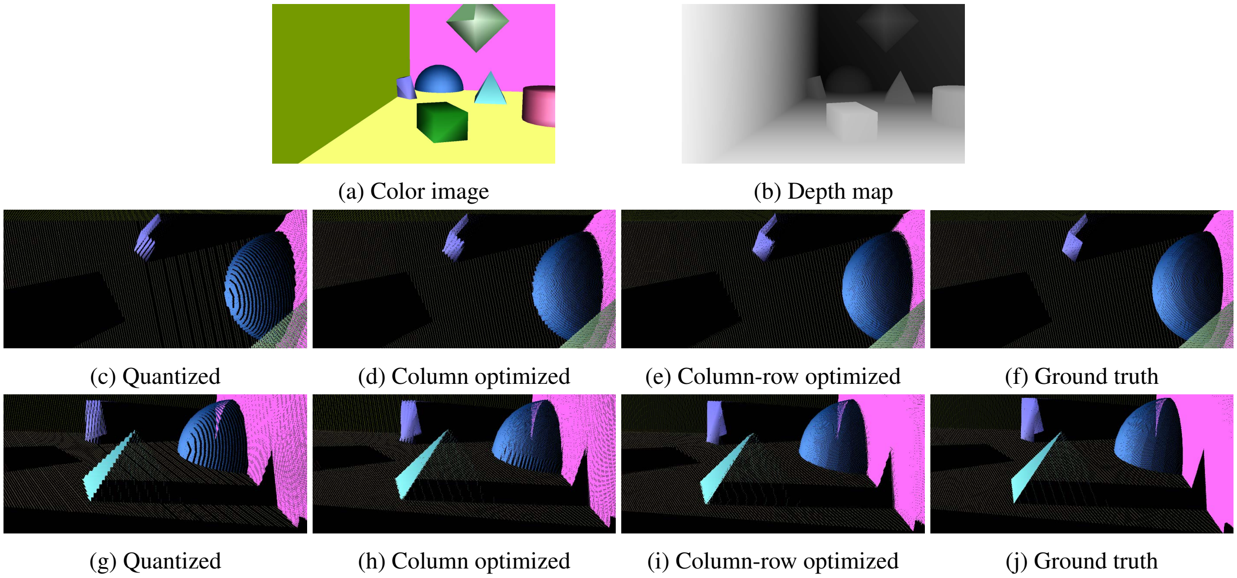
For the convenience of analysis, we use computer graphics rendering to derive floating-point depth maps z as the ground truth. The depth maps z are then normalized using

$$z_{normalized} = 255 - \frac{255}{z_{far} - z_{near}} \times (z - z_{near}), \tag{4}$$

and rounded off to 8-bit integers (0 to 255), generating quantized 8-bit depth maps.

3.1. Results

We use point-cloud models to visualize the results properly by projecting the depth maps into 3D space. Figs. 4 and 5 show the results of two sequences, Room and Sponza, respectively. Although theoretically the row-column decomposition might degrade the results slightly, the processed depth maps are much better than the quantized ones. Sequence Room is employed to show the optimization



(a) Color image

(b) Depth map

(c) Quantized

(d) Column optimized

(e) Column-row optimized

(f) Ground truth

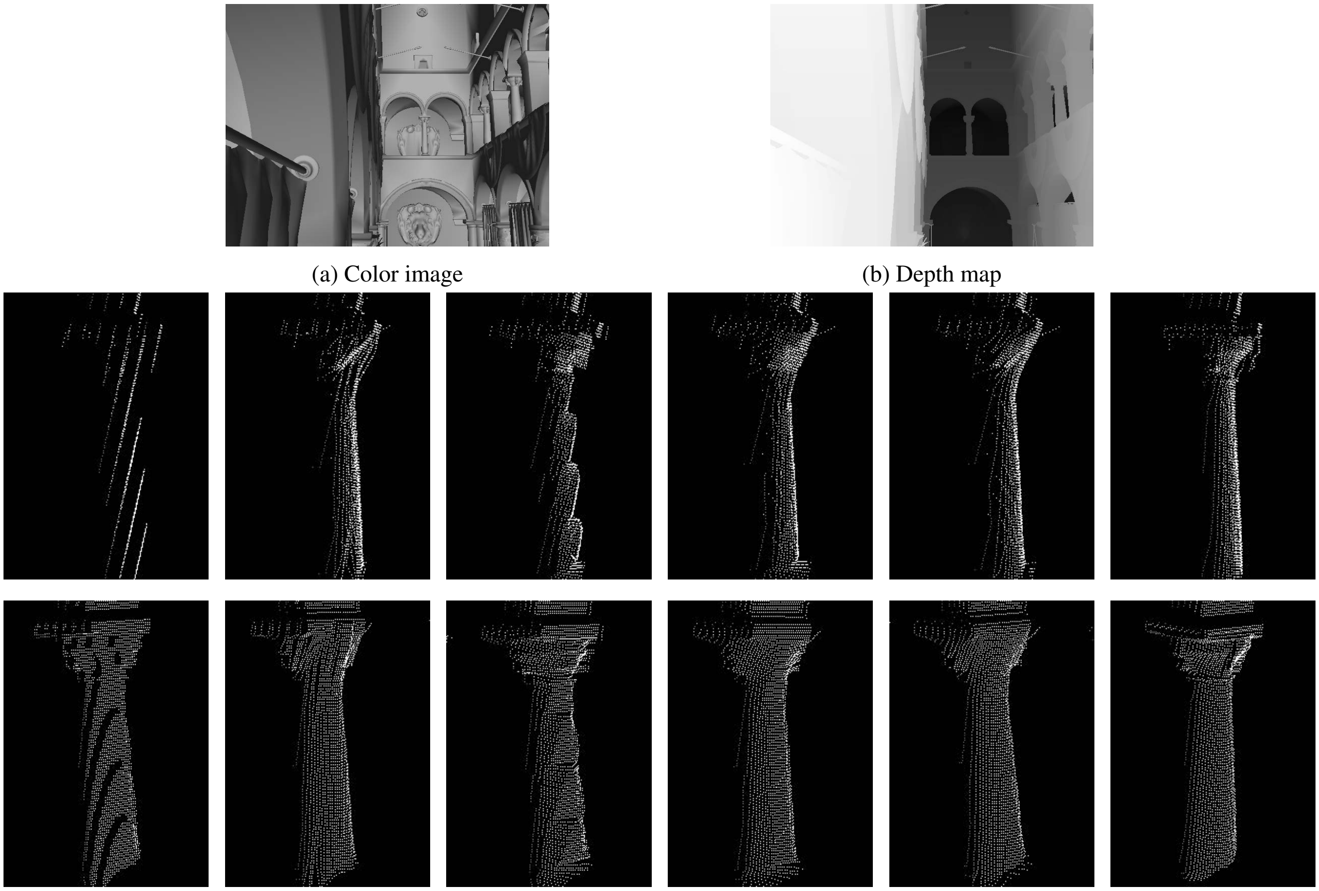
(g) Quantized

(h) Column optimized

(i) Column-row optimized

(j) Ground truth

Fig. 4: Comparison between the reconstructed models and the ground truth for sequence Room. The second row and the third row show images from two selected views. The optimization reduces the cardboard effect effectively.



(a) Color image

(b) Depth map

Fig. 5: Comparison between the reconstructed models and the ground truth for sequence Sponza. The second row and the third row show images from two selected views. From left to right: quantized, column optimized, row optimized, column-row optimized, row-column optimized, ground truth.

Table 1: Evaluation and execution time for sequence Room (1280x720)

Name	MSE	Reduction Ratio	MSSIM	Enhancement Ratio	Execution Time (sec.)
Original (8-bit)	0.0830	-	0.5118	-	-
Column optimized	0.0323	61.0843%	0.7693	1.5031	707.66
Row optimized	0.0253	69.5181%	0.9370	1.8308	609.79
Column-row optimized	0.0060	92.7711%	0.9888	1.9320	1318.0
Row-column optimized	0.0057	93.1325%	0.9903	1.9349	1326.1

Table 2: Evaluation and execution time for sequence Sponza (1280x960)

Name	MSE	Reduction Ratio	MSSIM	Enhancement Ratio	Execution Time (sec.)
Original (8-bit)	0.0822	-	0.2106	-	-
Column optimized	0.0532	35.2798%	0.4841	2.2987	836.63
Row optimized	0.0709	13.7470%	0.4076	1.9354	752.83
Column-row optimized	0.0585	28.8321%	0.6136	2.9136	1600.1
Row-column optimized	0.0564	31.3869%	0.6149	2.9196	1596.4

effect in each pass. Optimizing along column direction only refines the cardboard effect vertically as shown in Figs. 4d and 4h. By further optimizing along row direction, as shown in Figs. 4e and 4i, the cardboard effect is refined horizontally. The depth map after two-pass optimization has similar structure to the ground truth as shown in Figs. 4f and 4j. Similar results are also observed with sequence Sponza, as shown in Fig. 5, where the cardboard effects are successfully alleviated.

3.2. Objective Evaluation

Mean square error (MSE) is first employed to evaluate how well the signals are estimated. The evaluation results are listed in Tables 1 and 2, where great reduction is shown in terms of MSE.

However, the reduction in MSE does not reflect the signal properties completely since the structure of signals is not considered. Alternatively, we employ mean structure similarity index (MSSIM) [9] to measure the structure of signals. MSSIM is the average of SSIM calculated within each small window, and the SSIM measure between two signals x and y is:

$$SSIM(x,y) = \left(\frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2}\right)\left(\frac{2\sigma_{xy}}{\sigma_x^2+\sigma_y^2}\right), \qquad (5)$$

where μ_x and μ_y denote the means of x and y respectively, σ_{xy} is the covariance of x and y, and σ_x^2 and σ_y^2 are the variances of x and y respectively. SSIM would be 1 if two identical signals are measured. We use small windows (4×4) for SSIM to measure the local structures completely. Tables 1 and 2 also list the evaluation results. Obvious enhancement is achieved after optimizing the depth maps along both row and column directions. While only one-pass optimization of the depth maps may cause smaller MSE, MSSIM is always larger after two-pass optimization.

We use a PC to run these experiments, with the environment of i7-2600K with 4GB memory and Win7 64-bit OS. Tables 1 and 2 also show the execution time of each sequence. It can be seen that the execution time grows with the resolution of the depth map.

4. CONCLUSION

This paper first shows that the cardboard effect due to quantization in depth maps influences the applications of 3D reconstruction and view synthesis, especially when spatial resolution becomes larger and larger nowadays. We then model the quantization error and propose an optimization framework to reduce the error. The experimental results are visualized in 3D point clouds to demonstrate the effectiveness of our approach subjectively. Objective evaluation also strongly shows the effectiveness of the proposed algorithm. Applications requiring precise depth maps can be beneficial from this paper.

Even if the row-column decomposition is employed, it still takes minutes to minimize the energy function (3). The approximate methods for fast computing can be studied in the future. Furthermore, only the quantization error is considered in this paper. By joint optimizing the sensor noise, this framework can be extended to suit depth maps derived from various sensors.

5. ACKNOWLEDGEMENT

The authors would like to thank Chung-Te Li for providing useful advice. This work was supported by Himax Technologies, Inc.

6. REFERENCES

[1] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A.W. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking.," in *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, 2011, pp. 127–136.

[2] 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, April 2011.

[3] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Proc. SPIE 5291, Stereoscopic Displays and Virtual Reality Systems XI*, 2004.

[4] A. Boev, D. Hollosi, A. Gotchev, and K. Egiazarian, "Classification and simulation of stereoscopic artifacts in mobile 3DTV content," in *Proc. SPIE 7237, Stereoscopic Displays and Applications XX*, 2009.

[5] G. Zhang, J. Jia, T. T. Wong, and H. Bao, "Consistent depth maps recovery from a video sequence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 974–988, June 2009.

[6] C. H. Hung, L. Xu, and J. Jia, “Consistent binocular depth and scene flow with chained temporal profiles,” *International Journal of Computer Vision*, pp. 1–22, Augest 2012.

[7] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, “Quantization and dither: A theoretical survey,” *Journal of the Audio Engineering Society*, vol. 40, no. 5, pp. 355–375, May 1992.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.