MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

# Depth Reconstruction Filter and Down/Up Sampling for Depth Coding in 3D Video

Kwan-Jung Oh, Sehoon Yea, Anthony Vetro, Yo-Sung Ho

TR2009-030 July 2009

**Abstract**

Depth images represent three-dimensional (3D) scene information and are commonly used for depth image based rendering (DIBR) to support 3D video and free-viewpoint video applications. The virtual view is generally rendered by DIBR technique and its quality highly depends on the quality of depth data. Thus, efficient depth coding is a core technique to realize 3D video systems. In this letter, we propose a depth reconstruction filter and depth down/up sampling techniques. Experimental results demonstrate that the proposed methods reduce the depth coding bit rate and achieve better rendering quality.

*IEEE Signal Processing Letters*

# Depth Reconstruction Filter and Down/Up Sampling for Depth Coding in 3D Video

Kwan-Jung Oh, *Student Member*, Sehoon Yea, *Member*, Anthony Vetro, *Senior Member*, and Yo-Sung Ho, *Senior Member*, *IEEE*

***Abstract*— Depth images represent three-dimensional (3D) scene information and are commonly used for depth image based rendering (DIBR) to support 3D video and free-viewpoint video applications. The virtual view is generally rendered by DIBR technique and its quality highly depends on the quality of depth data. Thus, efficient depth coding is a core technique to realize 3D video system. In this letter, we propose a depth reconstruction filter and depth down/up sampling techniques. Experimental results demonstrate that the proposed methods reduce the depth coding bit rate and achieve better rendering quality.**

***Index Terms*— Depth reconstruction filter, Depth down/up sampling, 3D Video, Free viewpoint television (FTV)**

## I. INTRODUCTION

DEPTH images represent a relative distance from a camera to an object in 3D space, and are widely used in computer vision and computer graphics to represent 3D scene information. Most image-based rendering (IBR) methods [1] utilize depth images in combination with stereo or multi-view video to realize 3D and free-viewpoint video applications [2]-[7]. Efficient compression of both the depth images and multi-view videos and rendering techniques are key issues for such applications and various coding schemes have been proposed.

The recently finalized multi-view video coding (MVC) [8] standard supports inter-view prediction, but it does not include any particular provisions for coding of depth. Only very recently are depth related issues such as depth estimation, rendering, and depth coding being investigated in the context of an end-to-end system that considers acquisition, transmission, and display in MPEG 3DV group.

The main objective of depth coding is how to synthesize the high quality virtual view with the low bit rate depth. That is, whereas texture coding focuses on the improvement coding performance itself, the depth coding should consider rendering quality than depth quality. In this letter, we propose a depth reconstruction filter and down/up sampling techniques. The proposed methods mainly focus on the object boundary which is sensitive to both coding and rendering.

The proposed depth reconstruction filter is designed as an in-loop filter to recover the object boundary and consists of newly designed frequent-low-high filter and bilateral filter. The proposed down/up sampling employs non-linear scaling and an up-scaled image is post-filtered by combination of median filter and depth reconstruction filter. The efficiency of the proposed algorithms is evaluated through the depth coding rate, depth quality, and rendering quality. The experimental results show better performance compared with the original depth coding.

Manuscript received February 15, 2009. This work was supported in part by ITRC through RBRC at GIST (IITA-2009-C1090-0902-0017) and in part by MOE through the BK21 project. The associate editor coordinating the review of this manuscript and approving it for publication was.

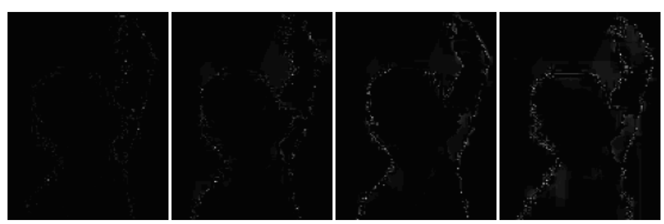
K.-J. Oh and Y.-S. Ho are with the School of Information and Mechatronics, Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea (e-mail: kjoh81@gist.ac.kr, hoyo@gist.ac.kr)

S. Yea and A. Vetro are with the Mitsubishi Electric Research Laboratories (MERL), MA 02139, USA. (e-mail: yea@merl.com, avetro@merl.com).

## II. PROPOSED DEPTH CODING METHODS

### A. Depth Reconstruction Filter

Unlike natural images, depth images have a well-defined object boundary, and it is very important to maintain the object boundary for high-quality rendering. However, since this is where the most significant information resides, the object boundaries are also prone to coding errors, which increase with lower bit rates as shown in Fig. 1.



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

Fig. 1. Depth coding errors: (a) QP22, (b) QP27, (c) QP32, (d) QP37.

In this letter, we propose a depth reconstruction filter to compensate for depth coding errors and to recover the object boundary. The proposed depth reconstruction filter consists of the following two filters: 1) frequent-low-high filter to recover object boundaries, 2) bilateral filter to eliminate the remaining errors. The frequent-low-high filter is a kind of non-linear filter and is designed as follows:

Firstly, we defined a window *W(i, j)* which is a set of pixels in a $m \times n$ window. *m* and *n* should be odd. And then the pixels in *W* are sorted by its intensity value as in (1) and the median value for *W* is defined as in (2).

$$sort(W(i,j)) = \{I_0, I_1, I_2 \cdots I_{m\times n-1}\} \tag{1}$$

$$median(W) = I_{(m \times n+1)/2} \quad (2)$$

The pixels in $W$ are categorized into low and high group by $median(W)$ as depicted in (3).

$$W(i,j) \in \begin{cases} S_{low}, & if \;\; W(i,j) < median(W) \\ S_{high}, & otherwise \end{cases} \quad (3)$$

For the pixel set $S_{low}(k)$, we count the number of occurrence ($C_v$) for each intensity value ($v$) belonging to $S_{low}$ with (4).

$$C_v = \sum_{k=0}^{(m \times n-1)/2} \delta[v, S_{low}(k)] \quad \text{with} \quad \delta[a,b] = \begin{cases} 1, & if \;\; a=b \\ 0, & otherwise \end{cases} \quad (4)$$

Among $v$ values, we chose the intensity having the maximum number of occurrence and we define it as $v_{low}$ as depicted in (5). That is, $v_{low}$ is the pixel intensity value of the most frequently occurred intensity value among the set of pixel values which are smaller than the median value of $W$.

$$C_{v_{low}} = \max\{C_{v_0}, C_{v_1}, \cdots C_{v_k}\} \quad (5)$$

In the same manner, $v_{high}$ is defined. Finally, we decide the representative value for a given pixel by (6).

$$W(x,y) = \begin{cases} v_{low}, & if \;\; |W(x,y) - v_{low}| < |W(x,y) - v_{high}| \\ v_{high}, & otherwise \end{cases} \quad (6)$$

where $W(x, y)$ is a center point in $W$. $W(x, y)$ is replaced by the closer of the two representative values $v_{low}$ and $v_{high}$ in the proposed frequent-low-high filter.

The proposed frequent-low-high filter has following advantages over other linear filters: 1) it is more robust against outliers; a single pixel that does not have a similar intensity as those of neighboring pixels will not affect the frequent-low-high value significantly, and 2) since the frequent-low-high value must actually be the value of one of the pixels in the neighborhood, the frequent-low-high does not create new unrealistic pixel values when the filter straddles an edge.

However, some errors may still remain and they look like Gaussian noise. To eliminate the remaining errors, we apply the bilateral filter [9]. The bilateral filter is non-linearly designed and is a representative edge-preserving filter. The bilateral filter extends the concept of Gaussian smoothing by weighting the filter coefficients with their corresponding relative pixel intensities. Pixels that are very different in intensity from the central pixel are weighted less even though they may be in close proximity to the central pixel. This is effectively a convolution with a non-linear Gaussian filter, with weights based on pixel intensities as in (7).

$$BF[I]_p = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_s}(\| p-q \|) \; G_{\sigma_r}(| I_p - I_q |) I_q \quad (7)$$

where $I_p$ is a center pixel to be bilateral filtered and $I_q$ is a $I_p$'s neighboring pixel in the window $W_p$. The filter parameters, space sigma ($\sigma_s$) and range sigma ($\sigma_r$), which define the spatial extent of the kernel and threshold for which color differences in the image, respectively. The depth reconstruction filter is implemented as an in-loop filter and it is located after the deblocking filter in H.264/AVC compatible codecs.

### *B. Depth Down/Up Sampling*

Encoding a reduced resolution depth can reduce the bit rate substantially, but the loss of resolution also degrades the quality of the depth map, especially in high frequency regions such as the object boundary; the resulting image rendering artifacts could be very visible and disturbing. Conventional down/up samplers use a low-pass filter and an interpolation filter to reduce the quality degradation. However, since the depth video and image rendering results are quite sensitive to variations in space and time, especially on the object boundary, these traditional techniques are not sufficient.

Considering the above, we design a new down/up sampler for depth. In down sampling a 2D image, a representative value among the values in a certain window must be selected; we choose a median value as in (8).

$$img_{down}(x,y) = median(W_{s \times s}) \quad (8)$$

where $W_{sxs}$ represents a $s$x$s$ block and $s$ is a scaling factor for down sampling. That is, the down sampling reduces the image size by selecting the median value for each $W_{sxs}$.

The up sampling process consists of following steps: 1) image up-scaling, 2) 2D median filtering, 3) proposed depth reconstruction filtering. The up-scaling is an inverse process of the image down-sampling as depicted in (9). We use a same scaling factor used in down sampling process.

$$img_{up}(x,y) = img_{down}(\lfloor x/s \rfloor, \lfloor y/s \rfloor) \quad (9)$$

After the up-scaling, we apply 2D median filter to smooth the blocking artifact caused by image down sampling. Finally, the proposed depth reconstruction filtering is applied to reconstruct the object boundary distorted by down sampling.

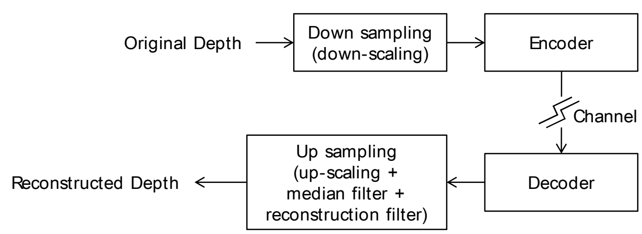



Fig. 2. Flow diagram of depth down/up sampling.

Fig. 2 shows a flow diagram of depth down/up sampling.

### C. *Evaluation of Depth Coding*

The coding efficiency of common texture images is measured in the encoding bit rate and a peak signal-to-noise ratio (PSNR) value as in (10).

$$PSNR = 10 \times \log_{10}\left(255^2 / MSE\right) \qquad (10)$$

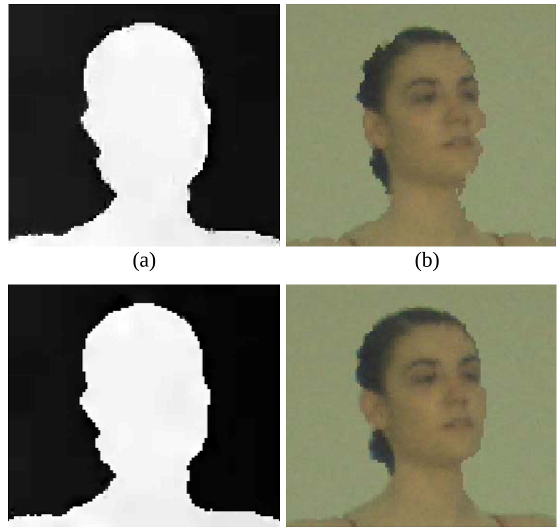
However, the depth image is 3D information to synthesize the virtual view, thus its quality should be evaluated in terms of rendering quality. In this letter, we measure the rendering PSNR by MSE between original image ($I_{org}$) and rendered image ($I_{ren}$) with the reconstructed depth image as in (11).

$$MSE_{ren} = \frac{1}{w \times h} \sum_{i=0}^{w-1} \sum_{j=0}^{h-1} \left\| I_{org}(i,j) - I_{ren}(i,j) \right\|^2 \qquad (11)$$

## III. Experimental Results and Analysis

We have tested the proposed algorithm on two test sequences, “Ballet” and “Breakdancers” [10]. Among the 8 views, view 3 and view 5 were selected as reference views and view 4 was set as the virtual view to be synthesized. Depth views were encoded using JMVC 3.0 [11] with QP 22, 25, 28, and 31. The delta QP, differential QP between the basis layer and sub-layer in hierarchical-B picture structure, was set as zero in all layers. We used own rendering method designed based on [12] and it was fairly used and the texture videos were not encoded.

The 7x7 frequent-low-high filter was used and the bilateral filter parameters, the color sigma and space sigma, were set as 10 and 1, respectively. The experimental results for depth reconstruction filter are given in Table I and Table II. The rate-distortion (RD) curves are shown in Fig. 3.



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

Fig. 4. Depth reconstruction filter (QP31, 33rd frame): (a) without depth reconstruction filter, (b) rendering result for (a), (c) with depth reconstruction filter, (d) rendering result for (c).

Fig. 4 illustrates example images of the depth reconstruction filter for “Ballet” sequence and the color of depth images was auto-adjusted for better visibility.

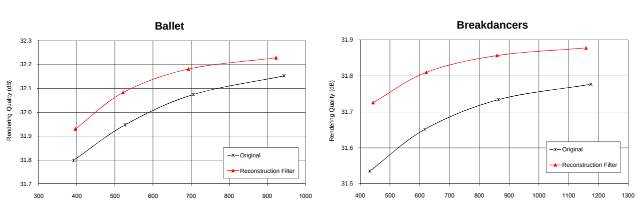
The experimental results for depth down/up sampling filter in depth rate and rendering quality are given in Table III and Table IV. The 5x5 median and 13x13 frequent-low-high filters were used for up sampling. The RD curves are illustrated in Fig. 5. Fig. 6 shows example images of the depth down/up sampling and they were auto-adjusted for better visibility.

TABLE I
EXPERIMENTAL RESULTS FOR RECONSTRUCTION FILTER (BALLET)

| QP | Depth rate (kbps) | | Rendering quality (dB) | |
|---|---|---|---|---|
| | original | proposed | original | proposed |
| 22 | 943.50 | 922.68 | 32.15 | 32.23 |
| 25 | 705.07 | 692.70 | 32.07 | 32.18 |
| 28 | 526.68 | 521.25 | 31.95 | 32.08 |
| 31 | 391.52 | 395.90 | 31.80 | 31.93 |

TABLE II
EXPERIMENTAL RESULTS FOR RECONSTRUCTION FILTER (BREAKDANCERS)

| QP | Depth rate (kbps) | | Rendering quality (dB) | |
|---|---|---|---|---|
| | original | proposed | original | proposed |
| 22 | 1174.53 | 1158.18 | 31.78 | 31.88 |
| 25 | 863.78 | 859.05 | 31.73 | 31.86 |
| 28 | 616.88 | 621.78 | 31.65 | 31.81 |
| 31 | 432.15 | 442.97 | 31.53 | 31.73 |




(a) (b)

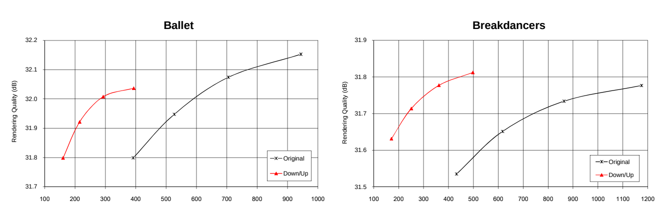
Fig. 3. RD curves for depth reconstruction filter: (a) Ballet, (b) Breakdancers.

TABLE III
EXPERIMENTAL RESULTS FOR DOWN/UP SAMPLING (BALLET)

| QP | Depth rate (kbps) | | Rendering quality (dB) | |
|---|---|---|---|---|
| | original | proposed | original | proposed |
| 22 | 943.50 | 393.09 | 32.15 | 32.04 |
| 25 | 705.07 | 292.59 | 32.07 | 32.01 |
| 28 | 526.68 | 215.12 | 31.95 | 31.92 |
| 31 | 391.52 | 160.55 | 31.80 | 31.80 |

TABLE IV
EXPERIMENTAL RESULTS FOR DOWN/UP SAMPLING (BREAKDANCERS)

| QP | Depth rate (kbps) | | Rendering quality (dB) | |
|---|---|---|---|---|
| | original | proposed | original | proposed |
| 22 | 1174.53 | 497.67 | 31.78 | 31.81 |
| 25 | 863.78 | 361.36 | 31.73 | 31.78 |
| 28 | 616.88 | 250.52 | 31.65 | 31.71 |
| 31 | 432.15 | 170.50 | 31.53 | 31.63 |




(a) (b)

Fig. 5. RD curves for depth down/up sampling: (a) Ballet, (b) Breakdancers.

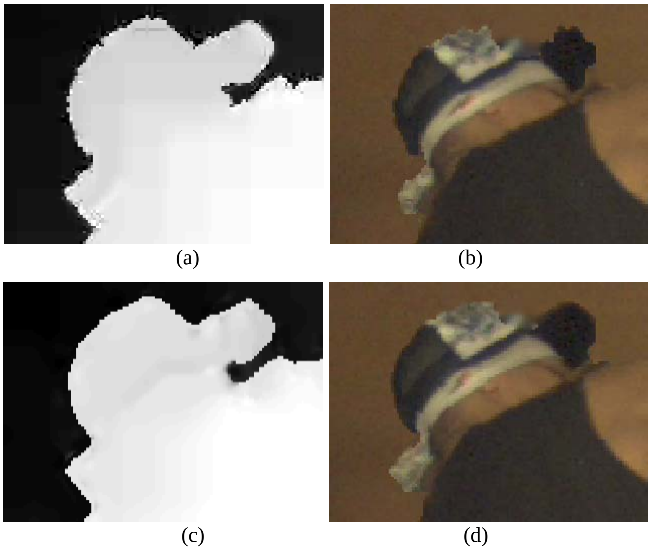



Fig. 6. Depth down/up sampling (QP31, 1st frame): (a) without depth down/up sampling, (b) rendering result for (a), (c) with depth down/up sampling, (d) rendering result for (c).

From the experiments, we confirmed that the proposed depth reconstruction filter and down/up sampling achieved better depth coding performance and better rendering quality.

Table V and Table VI show the experimental results for both the depth reconstruction filter and down/up sampling methods were used. Finally, we compared and analyzed the experimental result for three coding methods. As shown in Fig. 7 and Fig. 8, the combined scheme shows the best performance among three cases but is slightly better than the down/up sampling scheme. It is because of that the effect of the depth reconstruction filter was offset by the reconstruction filter in up sampling process.

TABLE V

EXPERIMENTAL RESULTS FOR COMBINED METHOD (BALLET)

| QP | Depth rate (kbps) | | Rendering quality (dB) | |
|---|---|---|---|---|
| | original | proposed | original | proposed |
| 22 | 943.50 | 384.39 | 32.15 | 32.04 |
| 25 | 705.07 | 287.29 | 32.07 | 32.01 |
| 28 | 526.68 | 212.83 | 31.95 | 31.92 |
| 31 | 391.52 | 159.07 | 31.80 | 31.80 |

TABLE VI

EXPERIMENTAL RESULTS FOR COMBINED METHOD (BREAKDANCERS)

| QP | Depth rate (kbps) | | Rendering quality (dB) | |
|---|---|---|---|---|
| | original | proposed | original | proposed |
| 22 | 1174.53 | 489.07 | 31.78 | 31.81 |
| 25 | 863.78 | 358.35 | 31.73 | 31.78 |
| 28 | 616.88 | 252.04 | 31.65 | 31.71 |
| 31 | 432.15 | 174.71 | 31.53 | 31.64 |

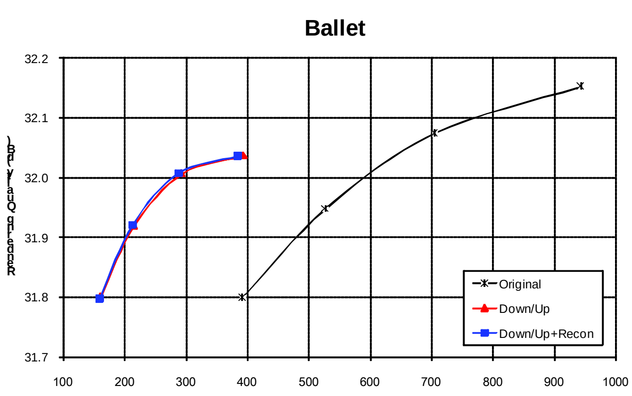



Fig. 7. Comparison of RD curves for three depth coding methods.

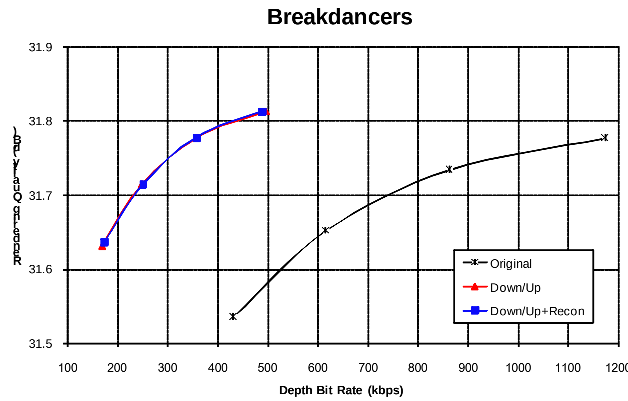



Fig. 8. Comparison of RD curves for three depth coding methods.

## IV. CONCLUSIONS

In this letter, we have proposed depth reconstruction filter and depth down/up sampling methods. The proposed depth reconstruction filter consisted of the frequent-low-high filter and bilateral filter. It was enough to recover the object boundary and gave advantages to both depth coding and rendering. In addition, the down/up sampling was designed by considering the characteristics of depth and enormously reduced the depth bit-rate without degradation of rendering quality. By experiments, we confirmed that the proposed schemes achieved better coding performance and rendering quality.

## REFERENCES

[1] H.Y. Shum, S.B. Kang, and S.C. Chan, "Survey of image-based representations and compression techniques," Proc. of IEEE Transactions on CSVT, vol. 13, no. 11, pp. 1020–1037, Nov. 2003.

[2] A. Smolic and D. McCutchen, "3DAV exploration of video-based rendering technology in MPEG," Proc. of IEEE Transactions on Circuit and System for Video Technology, vol. 14, no. 3, pp. 348–356, Mar. 2004.

[3] F.Isgro, E. Trucco, P. Kauff, and O. Schreer, "3D image processing in the future of immersive media," Proc. of IEEE Transactions on Circuits and Systems for Video Technology, vol.14, no. 3, pp. 288–303, 2004.

[4] A. Smolic and P. Kauff, "Interactive 3D video representation and coding technologies," Proc. of IEEE, Special Issue on Advances in Video Coding and Delivery. vol. 93, pp. 99–110, 2005.

[5] M. Tanimoto, "Overview of free viewpoint television," Proc. of Signal Processing: Image Communication, vol 21, pp. 454-465, July 2006.

[6] A. Smolic, H. Kimata, and A.Vetro, "Development of MPEG standards for 3D and free viewpoint video," Proc. of SPIE Optics East, Three-Dimensional TV, Video, and Display IV, Boston, MA, Oct. 2005.

[7] ISO/IEC JTC1/SC29/WG11, "Description of exploration experiments in 3-D video coding," Doc. N10173, Oct. 2008.

[8] ISO/IEC JTC1/SC29/WG11, "Survey of algorithms used for multi-view video coding (MVC)," Doc. N6909, Jan. 2005.

[9] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," Proc. of the IEEE International Conference on Computer Vision, Bombay, India, 1998.

[10] MSR 3D Video Sequences. Available at: http://www.research.microsoft.com/vision/ImageBasedRealities/3DVideoDownload/.

[11] ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, "Joint multiview video coding (JMVC) 3.0," Doc. JVT-AC207, Oct. 2008.

[12] ISO/IEC JTC1/SC29/WG11, “Reference softwares for depth estimation and view synthesis,” Doc. M15377, Apr. 2008.