

Depth Intra Skip Prediction for 3D Video Coding

Kwan-Jung Oh, Jaejoon Lee, and Du-Sik Park

Samsung Advanced Institute of Technology, Samsung Electronics, Korea

E-mail: kwanjung.oh@samsung.com

Abstract— Depth image compression play a key role in the 3D video system. It can be used as supplementary data for rendering. In this paper, we present a depth intra skip prediction for 3D video coding. The depth intra skip prediction is designed based on the intra 16x16 mode. It exploits the estimated prediction direction which is derived from the adjacent neighboring pixels and does not encode any residual data. The proposed depth intra skip prediction is allowed for I slice as well as P and B slices. The usage of the depth intra skip prediction is signaled by a newly defined macroblock level flag. From the experiments, we confirm that the proposed depth intra skip prediction reduces the depth bit rate by up to 18.69% while preserving same synthesis quality for the virtual views.

I. INTRODUCTION

The recent successive big hits of the 3D movie and releasing of various 3D displays lead to the rapid growth of the 3D market and bring us much closer to the 3D age [1]. To keep pace with this trend, the moving picture experts group (MPEG) has prepared an international standard on 3D video coding. The MPEG 3D video coding (3DVC) standard aims to compression of a huge amount of 3D video data which consists of multiview color video and its corresponding depth data. In 98th MPEG meeting, various H.264/AVC (advanced video coding) compatible and HEVC (high efficiency video coding) compatible responses of call for proposal (CfP) were evaluated and several coding techniques have been studied in core experiments (CE).

Depth image is spatially high correlated, i.e., typically consist of almost flat regions separated by object boundaries. Moreover, preserving edges in depth images often yields better rendering quality. Thus, various edge-preserving intra coding methods were proposed for depth image compression. Edge-adaptive wavelet transforms have been proposed for depth image coding [2, 3], but these are not easily amenable to block based video coding standards. Alternatively, platelet [4] has been proposed for efficient representation of piecewise planar images, and these ideas were extended to depth map coding in [5]. Geometry-based block partitioning [6] and edge-aware intra prediction [7] have been proposed to represent arbitrary edge shapes. However, the previous intra coding methods for depth images focuses only high quality edge-preserving coding. Thus, they are good for complex depth images and high bit rate conditions, however not good for simple depth images and low bit rate conditions.

H.264/AVC has a high priority for inter coding than intra coding when both are available such as P or B picture coding. However, the above rule should be changed for depth video coding. Currently, most of depth video data are acquired by

mathematical stereo matching algorithms and remains are obtained by depth cameras or computer graphics tools. Even though the accuracy and temporal correlation of depth video depends on the acquisition type, the depth images are spatially high correlated compared to its corresponding texture images. The inter coding based on the motion estimation and motion compensation processes is useful for the temporally consistent and textured video data, but the depth images do not have enough textures and its pixel intensity value is varied when the object moves back and forth. That is, depth images do not have enough textures to guarantee accurate motion estimation and is temporally unstable.

Based on above observations, we propose a depth intra skip prediction (DISP). The DISP follows coding mechanism of the intra 16x16 mode but the prediction direction is estimated and residual data are not encoded. The prediction direction is derived by comparing the complexity of the adjacent left and above lines. The proposed DISP reduces the bit-rate for the flat regions and predictable simple edge regions. The DISP mode is applied to I slice as well as P and B slices.

II. BACKGROUNDS

Depth image represents the distance between a camera and an object. The depth image has totally different characteristics from texture image. It is not affected by illumination change, chrominance difference, and object shape so that the same object likely has the same depth values as shown in Fig. 1. Therefore, depth map consists of multiple planes and distinct edges, where the former represents the objects and the latter represents the object boundaries. In terms of frequency domain analysis, depth map has strong low-frequency energy for flat regions, and also strong mid-frequency energy for strong edges.

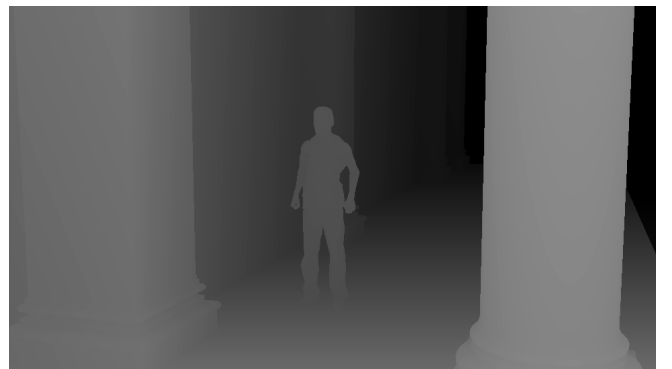


Fig. 1 Macroblock partitions for depth image of “Dancer” sequence.

The H.264/AVC exploits various macroblock coding modes and they are ordered by occurrence frequency. Each mode has its own code to represent mode information and the high priority mode has a shorter code length compared to the low priority mode. Table 1 shows the priority of macroblock modes for the B slice in H.264/AVC when context-based adaptive binary coding (CABAC) is used. The B slice is the most frequently used in MVC and the skip mode is the highest priority mode in H.264/AVC.

As shown in Table 1, the length of bin string of the intra 16x16 mode is longer than any other modes. The bins for the I macroblock in the B slice consists of 6 bins for a prefix and remaining bins for a suffix. According to the Table 1, to be selected the intra 16x16 mode as the best mode, 12 or 13 length of bin string is additionally coded compared to the skip mode. That is, the intra 16x16 mode has a big penalty in depth video coding based on the H.264/AVC even though it is more efficient than other modes if coding bits for the mode information are not considered. Thus, we propose a high priority depth intra coding method in this paper.

TABLE I
PRIORITY OF MACROBLOCK MODES FOR B SLICE AND ITS BIN STRING

Priority	MB mode	Length of bin string
0	B SKIP	-
1	B Direct	1
2	B 16x16	3 or 6
3	B 16x8 & B 8x16	6 or 7
4	B 8x8	6
5	I NxN	7
6	I 16x16	12 or 13
7	I PCM	8

III. PROPOSED DEPTH INTRA SKIP PREDICTION

The conventional intra 16x16 mode in H.264/AVC selects the best prediction direction in terms of rate and distortion optimization and sends it as mb_type with cbp information. In addition, residual data are also encoded according to cbp. However, if we can design a new intra mode which does not encode any data like a skip mode, it will be useful for depth video coding. Thus, we propose the intra skip mode based on the intra 16x16 mode. The proposed DISP estimates the prediction direction from adjacent neighboring blocks at both encoder and decoder sides. Moreover, the residual data are not sent. Fig. 2 shows the prediction direction used in DISP. The intra 16x16 supports four directional modes but plane mode is excluded for DISP.

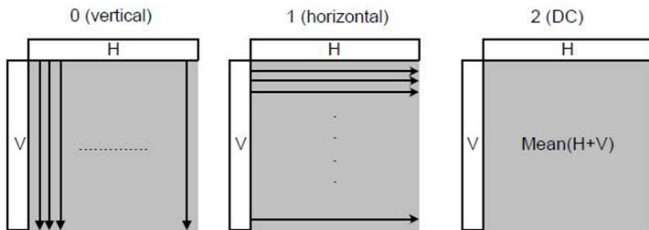


Fig. 2 Prediction directions for intra skip mode.

A. Estimation of Prediction Direction

The proposed intra skip mode needs direction information to employ prediction in the intra 16x16 mode. To use direction information without any additional coding bits, the same direction can be derivable both at an encoder and a decoder. Thus, we propose the estimation method for the prediction direction using adjacent neighboring pixels. The proposed method supports three prediction directions except for plane; vertical, horizontal, and DC. We decide a stronger direction by comparing adjacent left and above lines which belong to left and above macroblocks as shown in Fig. 3.

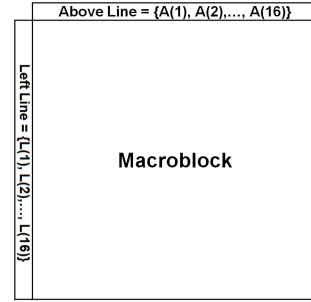


Fig. 3 Macroblock and its adjacent neighboring lines.

We assume that more complex line has a stronger direction. The NOC (number of change) is used as a complexity metric and NOC(L), NOC of left line, is calculated as follow:

$$\text{NOC}(L) = \sum_{i=1}^{15} C_L(i), C_L(i) = \begin{cases} 0, & L(i) = L(i+1) \\ 1, & L(i) \neq L(i+1) \end{cases} \quad (1)$$

The NOC(A) for above line is obtained in the same way. $L(i)$ represents the intensity value of i th position in the left line. We also consider the mode type and prediction directions for neighboring macroblocks in I slice. Fig. 4 shows the flow diagram of estimation of prediction direction.

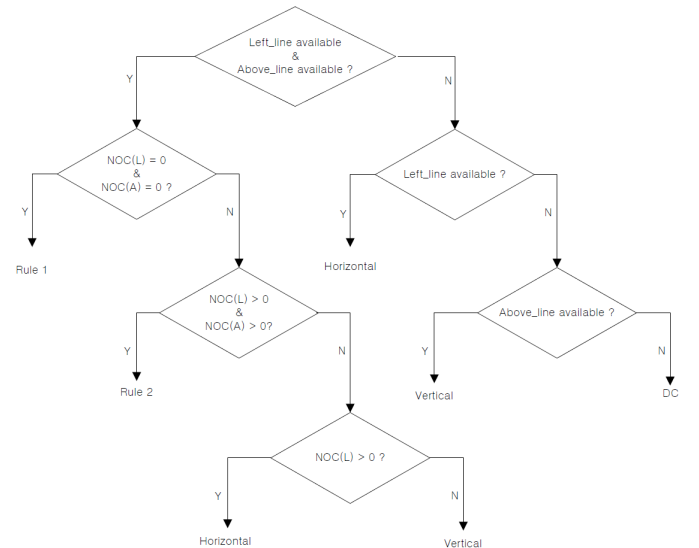


Fig. 4 Flow diagram of estimation of prediction direction.

Rule 1 and Rule 2 in Fig. 4 are as follows.

<Rule 1>

```

if (left MB = ver I16x16 mode & above MB = ver I16x16 mode)
: vertical direction
else if (left MB = hor I16x16 mode & above MB = hor I16x16 mode)
: horizontal direction
else if (left[0] < up[0])
: vertical direction
else if (left[0] > up[0])
: horizontal direction
else
: DC direction

```

<Rule 2>

```

if (left MB = ver I16x16 mode & above MB = ver I16x16 mode)
: vertical direction
else if (left MB = hor I16x16 mode & above MB = hor I16x16 mode)
: horizontal direction
else if (left MB ≠ hor I16x16 mode & above MB = ver I16x16 mode)
: vertical direction
else if (left MB = hor I16x16 mode & above MB ≠ ver I16x16 mode)
: horizontal direction
else if (NOC(L) < NOC(A))
: vertical direction
else if (NOC(L) > NOC(A))
: horizontal direction
else if (abs(above[0] - above[15]) > abs(left[0] - left[15]))
: vertical direction
else if (abs(left[0] - left[15]) > abs(above[0] - above[15]))
: horizontal direction
else
: DC direction

```

B. Signaling of DISP

The DISP mode is a kind of skip mode. Thus, we newly define a `mb_disp_flag` as Fig. 5. It is located between a inter skip and other modes. `mb_disp_flag` is encoded similar to `mb_skip_flag`.

slice_data() {	C	Descriptor
...		
mb_skip_flag	2	u(1) ae(v)
if (!mb_skip_flag && DepthFlag) {		
mb_disp_flag	2	u(1) ae(v)
...		
}		
...		

Fig. 5 Syntax of `mb_disp_flag` for DISP mode.

C. Deblocking Filter for DISP

The intra coded block has strong boundary strength such as 3 or 4 for deblocking filter because the intra block is predicted from the neighboring pixels which are not deblocking filtered. The boundary strength of the skipped block equals to 0. It is because of that the skipped block does not have any residual data. The DISP mode is intra block, but at the same time, skipped block be. Thus the deblocking filter for the proposed DISP is designed based on both. The boundary strength of DISP block equals to 2 for macroblock boundary and equals to 0 for others. Fig. 6 shows a bS (boundary strength) decision rule of deblocking filter when DISP is used. The macroblock type, macroblock boundary, cbp, and etc are considered for bS decision.

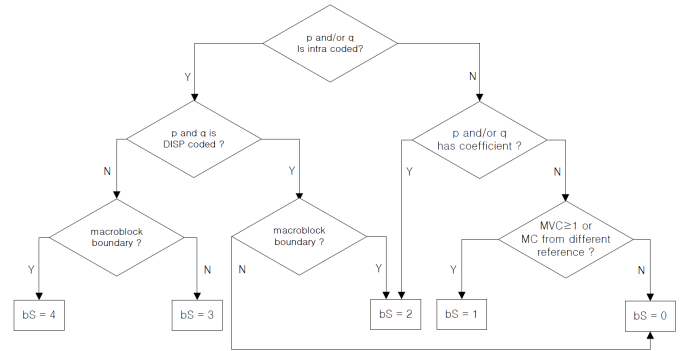


Fig. 6 bS decision for deblocking filter when DISP is used.

IV. EXPERIMENTAL RESULTS

To evaluate the efficiency of the proposed depth intra skip prediction, DISP has been implemented in 3DV-ATM version 0.3 [8]. The simulation setup mostly follows the common test condition [9]. All seven test sequences are tested with 3-view configuration and other depth tools are disabled. The full size texture and quarter size depth are coded with mixed resolution. The coding results are summarized in Table 2 using BDBR (Bjontegaard delta bit rate) metric [10] in terms of the total bit rate and PSNR of decoded texture views as well as the total rate and PSNR of synthesized views and depth rate and PSNR of synthesized views. The virtual views are synthesized by VSRS 1-D fast software [11]. Fig. 7 shows mode distribution for DISP.

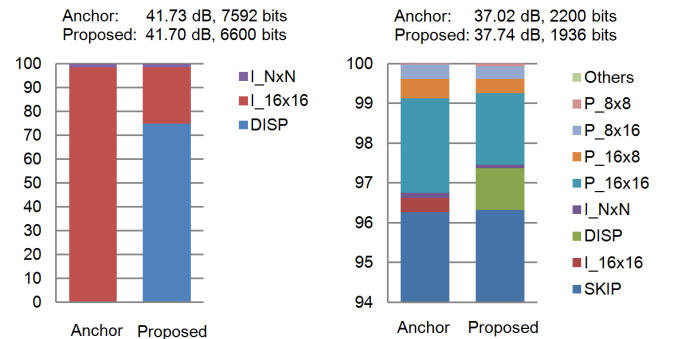
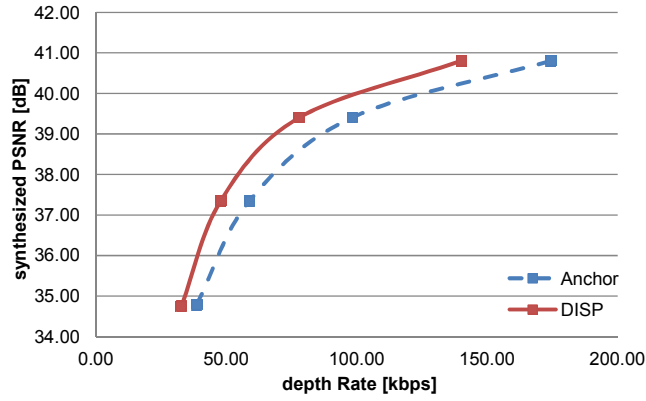


Fig. 7 Mode distribution of I slice and B slice for "Kendo" sequence.

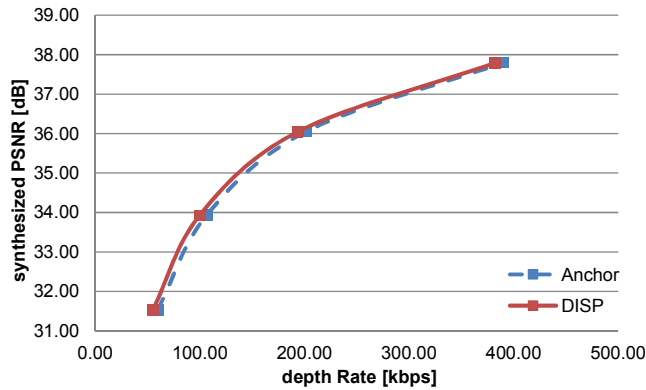
TABLE 2
EXPERIMENTAL RESULTS FOR DISP [BD-RATE(%)]

Sequences	depth rate vs. synthesized PSNR	total rate vs. synthesized PSNR	total rate vs. decoded PSNR
PoznanHall2	-18.69%	-2.04%	-2.15%
PoznanStreet	-6.39%	-0.49%	-0.67%
UndoDancer	-7.11%	-0.14%	-0.41%
GT_Fly	-12.96%	-1.16%	-1.19%
Kendo	-13.31%	-2.95%	-3.17%
Balloons	-7.37%	-0.74%	-0.98%
Newspaper	-5.19%	-0.74%	-0.90%
Average	-10.14%	-1.18%	-1.35%

The results in Table 2 indicate that DISP reduces the depth bit rate efficiently. The coding gains in terms of total bit rate versus decoded and synthesized PSNR are stable for all test sequences. Fig. 8 shows RD (rate-distortion) curves in terms of depth rate versus synthesized PSNR for the best case and worst case in Table 2. As seen in Fig. 8, DISP shows stable bit rate reduction at all bit rate ranges.



(a) RD curves for "PoznanHall2" sequence.



(b) RD curves for "Newspaper" sequence.

Fig. 8 Comparison of the RD curves between Anchor and DISP.

V. CONCLUSIONS

The intra prediction is useful to depth compression since depth image is spatially high-correlated. However, the current texture based video coding standards does not support any high priority intra prediction methods such as inter skip mode. Thus a new depth intra skip prediction has been proposed in this paper for depth video coding. The proposed DISP mode is designed based on the existing intra 16x16 mode. The DISP mode estimates the prediction direction both at the encoder and decoder sides and does not encode any residual data. The prediction direction is determined by comparing the line complexity of the adjacent neighboring pixels. The DISP is applied to all slice types and its use is signaled by the DISP flag which is designed like to convectional skip flag. The condition of the deblocking filter for DISP mode is modified by harmonizing of conditions of intra modes and skip mode. We implemented the proposed DISP method into 3DV-ATM 0.3 and confirmed that it achieves about 10.14% depth bit rate saving compared to anchor coding at the same rendering quality.

REFERENCES

- [1] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, and A. Koz, "Coding Algorithms for 3DTV - A Survey," *IEEE Trans. CSVT*, vol.17, no.11, 2007.
- [2] M. Maitre and M. N. Do, "Shape-adaptive wavelet encoding of depth maps," in *Proc. of PCS'09*, 2009.
- [3] A. Sanchez, G. Shen, and A. Ortega, "Edge-preserving depth map coding using graph-based wavelets," in *Proc. of Asilomar'09*, 2009.
- [4] R. Willett and R. Nowak, "Platelets: A multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no.3, pp. 332–350, March 2003.
- [5] Y. Morvan, P.H.N. de With, and D. Farin, "Platelet-based coding of depth maps for the transmission of multiview images," vol. 6055, SPIE, 2006.
- [6] M.K. Kang, J. Lee, J.Y. Lee, and Y.S. Ho, "Geometry-based block partitioning for efficient intra prediction in depth video coding," *SPIE Visual Information Processing and Communication*, vol. 7543, pp. 75430A-75430A-11, Jan. 2010.
- [7] G. Shen, W.-S. Kim, A. Ortega, J. Lee, and H. Wey, "Edge-aware intra prediction for depth-map coding", in *Proc. of IEEE Intl. Conf. Image Proc. (ICIP '10)*, Hong Kong, Sep. 2010.
- [8] MPEG 3DV AVC-based 3DV reference software version 0.3: <http://mpeg3dv.research.nokia.com/svn/mpeg3dv/>
- [9] ISO/IEC JTC1/SC29/WG11, "Common Test Conditions for 3DV experimentation," N12745, May 2012.
- [10] ITU-T SG16 Q.6, "An excel add-in for computing Bjøntegaard metric and its evolution," VCEG-AE07, Jan. 2007.
- [11] MPEG 3DV view synthesis reference software: VSRS 1-D Fast software.