IMPROVED DISPARITY VECTOR DERIVATION IN 3D-HEVC

Na Zhang[1], Yi-Wen Chen[2], Jian-Liang Lin[2], Xiaopeng Fan[1], Siwei Ma[3], Debin Zhao[1], Wen Gao[3], Fellow IEEE

[1]Dept. of Computer Science and technology, Harbin Institute of Technology, Harbin, China
[2]Mediatek Inc., Taiwan
[3]Peking University, Beijing, China
{hitnzhang, fxp, dbzhao}@hit.edu.cn, {yiwen.chen, jl.lin}@mediatek.com, {swma, wgao}@jdl.ac.cn

ABSTRACT

In the High Efficiency Video Coding (HEVC) based 3D video coding, 3D-HEVC, the disparity vector (DV) derivation is critical for inter-view motion prediction, inter-view residual prediction, disparity-compensated prediction (DCP) or any other tools exploiting inter-view correlation. In HTM-5.0.1, the DV is derived from some spatial and temporal neighbors to locate a corresponding block in another view. This paper advocates modifications of the DV derivation to reduce the complexity and to achieve slightly better coding efficiency. Firstly, we propose to remove the additional temporal block, unify the DV searching order for all views, and to impose restrictions on the temporal blocks for memory access bandwidth and complexity reduction in DV derivation. We also propose an improved DV searching order for slightly better coding performance. These proposed methods were adopted into the 3D-HEVC standard in the 3rd JCT-3V meeting in Jan. 2013.

Index Terms— 3D video coding, DV derivation, inter-view prediction, 3D-HEVC.

1. INTRODUCTION

As an extension of High Efficiency Video Coding (HEVC), 3D-HEVC is an ongoing multi-view data compression standard which is being developed by the JCT-3V, the joint working group of MPEG and ITU-T VCEG. Since all cameras capture the same scene simultaneously just from different viewpoints, multi-view video data contains a large amount of inter-view redundancy. Besides inter-view texture prediction (namely disparity compensated prediction, DCP), inter-view motion prediction and inter-view residual prediction [1] have also been integrated to 3D-HEVC as two major coding tools to exploit the inter-view redundancy.

In 3D-HEVC, a disparity vector (DV) is derived to specify the location of a corresponding block in the neighboring view for the inter-view motion prediction, inter-view residual prediction, DCP, or any other tools which need to indicate the correspondence between inter-view pictures. Since the accuracy of the derived DV affects the prediction efficiency of inter-view motion parameter prediction and residual prediction, how to derive an efficient DV with low complexity has become a critical issue in 3D-HEVC.

In the first version of 3D-HEVC test model, HTM 1.0, the DV was derived from an estimated depth map. The depth map is estimated and updated based on an already coded depth map or coded DVs and motion vectors (MVs) [1]. The DV was then derived by a depth-to-DV conversion using the depth samples within the associated depth block in the estimated depth map [1] [2]. However, deriving the DV from an estimated depth map has several drawbacks: 1) extra buffers are required to store the estimated depth maps; 2) depth mapping, hole-filling, and conversion between disparity vector and depth value introduces additional computational complexity; 3) high memory access bandwidth is required to update the estimated depth maps through the motion-compensated process. To tackle the above issues, a scheme of neighboring blocks disparity vector (NBDV) was proposed to derive the DV from several spatial and temporal neighboring blocks [3].

Based on the NBDV scheme, several spatial and temporal neighboring blocks are searched in a given order, and the DV is derived as the first available one. As shown in Fig. 1(a), the spatial neighboring blocks are the same as those used to derive the spatial motion vector predictor (SMVP) candidate in HEVC. However, to increase the probability to find a DV from the temporal neighboring blocks, all temporal reference pictures are searched for the temporal DV derivation. And for each temporal reference picture, at least 18 temporal neighboring blocks are searched resulting in a large quantity of temporal data accesses.

In HTM-5.0.1, to reduce the overhead of accessing temporal neighboring blocks while maintaining competitive coding performance, a modified disparity vector derivation algorithm is utilized to enlarge the source of disparity information by including disparity vector motion compensated prediction (DVMCP) blocks [4][5]. The DVMCP block is a block whose motion information is derived from a corresponding block in an inter-view reference picture and the location of the corresponding block is located by a disparity vector. The DVs of spatial DCP blocks and spatial DVMCP blocks are both used for the DV derivation. As for the temporal DV derivation, at

most two reference pictures are searched. One is the same as the one used for temporal motion vector predictor (TMVP) derivation and the other one is implicitly derived at both encoder and decoder side. The second temporal picture is chosen in a way that disparity vectors can have more chance to be present in the picture.

In this paper, several complexity issues of DV derivation in HTM-5.0.1 are revealed. We then propose a simplification of the DV derivation. We also propose an improvement of the DV searching order of DV derivation to slightly increase coding efficiency. The rest of the paper is organized as follows. Section 2 describes the details of the proposed method. In Section 3, the experimental results are presented. Finally, Section 4 concludes this paper.

2. PROPOSED METHOD

In the current 3D-HEVC test model version 5, HTM-5.0.1 [6], the DV is derived from several spatial and temporal neighboring blocks. The spatial neighboring blocks include A_1, B_1, B_0, A_0, B_2 as shown in Fig. 1(a) and three temporal neighboring blocks, located in the temporal collocated pictures as shown in Fig. 1(b) are used for the DV derivation. The DV is derived as the first available DV from searching the spatial and temporal neighboring blocks in a predefined order. The searching order is: spatial disparity-compensated prediction (DCP) blocks (A_1->B_1->B_0->A_0->B_2), temporal DCP blocks, and then spatial DVMCP blocks (A_0->A_1->B_0->B_1->B_2).

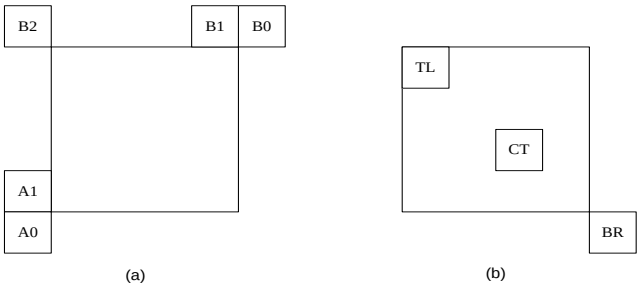

Fig.1. Candidate blocks for DV derivation: (a) spatial neighboring blocks; (b) temporal neighboring blocks in temporal collocated pictures.

In this paper, we first propose three simplifications to reduce the complexity for DV derivation. These three simplifications include: the removal of the additional temporal block, the unification of the temporal searching order, and the restriction on the temporal block. We also propose a modified DV searching order to improve coding performance.

2.1. Removal of the third temporal block

The temporal neighboring blocks include the top left corner 4x4 block within current prediction unit (PU), TL, the center 4x4 block within current PU, CT, and the bottom right corner 4x4 block of the current PU, BR (please refer to Fig. 1(b)). The temporal neighboring blocks, BR and CT, are the same blocks used in the derivation of temporal motion vector predictor (TMVP) in HEVC. However, an additional temporal neighboring block, TL, is required to be searched when the BR temporal block is outside the image boundary as shown in Fig. 2. In order to reduce the complexity of DV derivation and reduce the memory access bandwidth, we propose to remove the TL temporal block as shown in Fig. 3 so that the temporal blocks used for DV derivation could be aligned with those used for TMVP in HEVC.

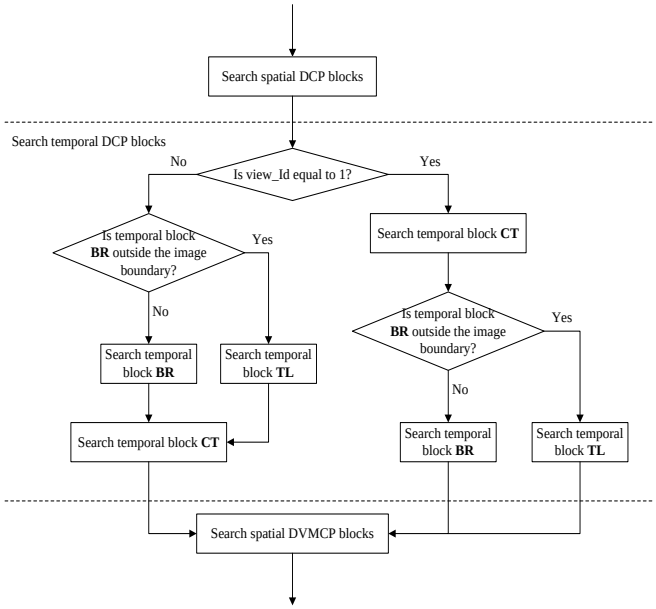

Fig. 2. The flowchart of DV derivation in HTM-5.0.1.

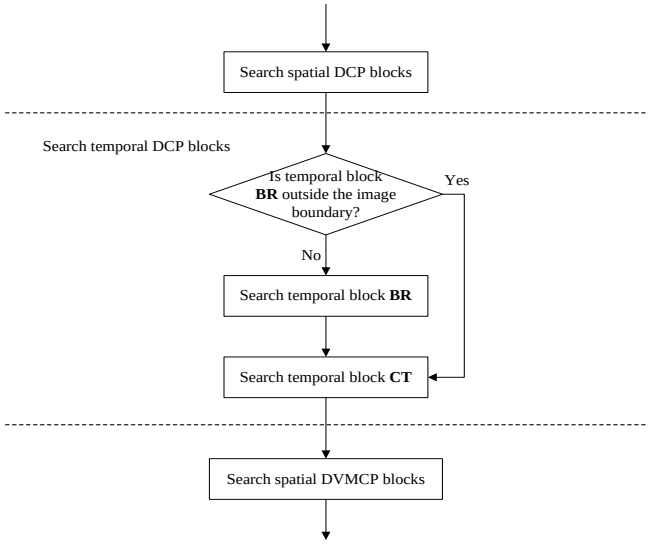

Fig. 3. The flowchart of the proposed scheme for DV derivation.

2.2. Unification of searching order

In current 3D-HEVC, when searching a DV among the temporal neighboring blocks, it utilizes different searching orders for different dependent views as shown in Fig. 2. For view 1 (view_Id=1), the searching order CT->BR (please refer to Fig. 1(b) for notations) is utilized. However, for the other dependent views, the searching order BR->CT is utilized. This requires additional operations to check whether the view_Id of current view is equal to 1.

In order to reduce the complexity of DV derivation, we propose to unify the searching order of the temporal neighboring blocks for all dependent views when deriving DV. We propose to use the searching order BR->CT for all dependent views as shown in Fig. 3.

2.3. Restriction on the BR temporal block

In current design, it permits to access the temporal block BR which resides in the lower coding tree unit (CTU) row. The major drawback of such design is that the collocated motion data fetch area is not aligned with the current CTU and thus causes higher memory bandwidth cost. In order to reduce the overhead of data fetch for the DV derivation and also to align with the derivation of TMVP, we prohibit the access of the BR temporal block residing at the lower CTU row as shown in Fig. 4.

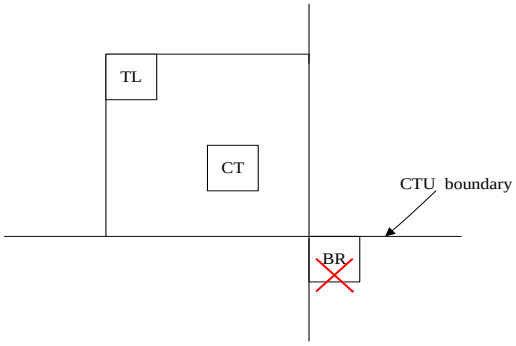

Fig. 4. BR temporal block at the lower CTU row is not used in the proposed scheme.

With the proposed simplifications, we can simplify DV derivation by using only one searching order for the temporal neighboring blocks and can reduce memory access bandwidth in DV derivation by unifying the temporal neighboring blocks used for TMVP and those for DV derivation.

2.4. Improved DV searching order

It is observed that the DVs derived from temporal neighboring blocks are more accurate than those derived from spatial neighboring blocks. This can be evident from the experiments of excluding spatial or temporal neighboring blocks from DV derivation [7]. The results show that excluding temporal neighboring blocks from DV derivation results in 1.7% BD-rate increase while excluding spatial neighboring blocks from DV derivation only results in 0.1% BD-rate increase.

In the current NBDV derivation process, the process is terminated once a DV is found. Although a DV derived from spatial neighboring block would be found earlier than any other candidates using current searching order, the DV derived from a spatial neighboring block may not serve as well as the one derived from a temporal neighboring block. Therefore, in this paper, we propose to search the DV among the temporal neighboring blocks first. The proposed order is: temporal DCP blocks, spatial DCP blocks (A1->B1->B0->A0->B2), and spatial DV-MCP blocks (A0->A1->B0->B1->B2).

3. EXPERIMENTAL RESULTS

To verify the performance of the proposed improvements, all of them have been implemented on HTM-5.0.1 [6] and simulated strictly in accordance with the common test conditions under the JCT-3V configurations as specified in [8]. The testing sequences with associated depth data and corresponding input views are listed in Table 1. The average PSNR and bit-rate of the coded texture and depth views are measured. Besides, the average PSNR of the synthesized views between two coded views are also measured. Experimental results of the proposed schemes for simplified DV derivation and improved DV searching order are demonstrated separately as well as jointly.

Table 1. Testing sequences and corresponding input views

Test Sequence	Resolution	Input views
Balloons	1024x768	1-3-5
Kendo	1024x768	1-3-5
Newspapercc	1024x768	2-4-6
GhostTownFly	1920x1088	9-5-1
PoznanHall2	1920x1088	7-6-5
PoznanStreet	1920x1088	5-4-3
UndoDancer	1920x1088	1-5-9

3.1. Results of simplified temporal DV derivation

The Bjontegaard deltas calculated from average PSNR and bit-rate of the proposed schemes as described in sections 2.1, 2.2, and 2.3 for simplified DV derivation are given in Table 2 and Table 3. The anchor is HTM-5.0.1. As can be seen in Table 3, negligible BD-rate changes are observed for overall coded and synthesized results while the searching orders for all dependent views are unified and the temporal blocks used in the DV derivation and those used for the TMVP derivation are also aligned. With the proposed three simplifications, the DV derivation process is simplified and the memory access bandwidth is also reduced, which can be reflected by the slightly reduced encoding and decoding run time as shown in Table 4.

Table 2. The BD-rate performance of the proposed simplified scheme for DV derivation compared to HTM-5.0.1 for all input views

Sequence	video0[1]	video1[2]	video2[2]
Balloons	0.0%	0.2%	0.6%
Kendo	0.0%	0.2%	0.3%
Newspapercc	0.0%	0.1%	0.1%
GhostTownFly	0.0%	0.2%	0.2%
PoznanHall2	0.0%	0.0%	0.2%
PoznanStreet	0.0%	0.0%	0.2%
UndoDancer	0.0%	0.1%	0.0%
1024x768	0.0%	0.2%	0.3%
1920x1088	0.0%	0.1%	0.1%
average	**0.0%**	**0.1%**	**0.2%**

[1]**video 0:** The BD-rate performance considering Y-PSNR of view 0 (base view)
[2]**video 1 &2:** The BD-rate performance considering Y-PSNR of view 1 and 2 (dependent views)

Table 3. The BD-rate performance of the proposed simplified scheme for DV derivation compared to HTM-5.0.1 for overall coded and synthesized views

Sequence	video only[3]	synthesized only[4]	coded & synthesized[5]
Balloons	0.1%	0.1%	0.1%
Kendo	0.1%	0.1%	0.1%
Newspapercc	0.0%	0.0%	0.0%
GhostTownFly	0.0%	0.0%	0.0%
PoznanHall2	0.1%	0.1%	0.1%
PoznanStreet	0.0%	0.0%	0.0%
UndoDancer	0.0%	-0.1%	-0.1%
1024x768	0.1%	0.1%	0.1%
1920x1088	0.0%	0.0%	0.0%
average	**0.1%**	**0.0%**	**0.0%**

[3]**video only:** The BD-rate performance considering Y-PSNR of the coded texture views over the bitrates of texture data
[4]**synthesized only:** The BD-rate performance considering Y-PSNR of the synthesized texture views over the bitrates of texture data and depth data
[5]**coded & synthesized:** The BD-rate performance considering Y-PSNR of the coded texture views and synthesized texture views over the bitrates of texture data and depth data

Table 4. Run time ratio of the proposed simplified scheme for DV derivation over HTM-5.0.1

Sequence	encoding time	decoding time	synthesis time
Balloons	98.5%	100.7%	99.8%
Kendo	97.6%	99.2%	100.2%
Newspapercc	98.4%	103.6%	98.3%
GhostTownFly	99.2%	100.5%	99.8%
PoznanHall2	99.1%	100.7%	99.8%
PoznanStreet	99.3%	94.4%	100.0%
UndoDancer	98.2%	98.0%	99.9%
1024x768	98.2%	101.2%	99.4%
1920x1088	99.0%	98.4%	99.9%
average	**98.6%**	**99.5%**	**99.7%**

3.2. Results of improved DV searching order

The BD-rate performance of the proposed scheme as described in section 2.4 for improved DV searching order is illustrated in Table 5 and Table 6. From Table 6, we can see that the proposed scheme can bring 0.1% overall BD-rate gain compared to HTM-5.0.1. Moreover, no complexity increase is introduced by the improved scheme compared to HTM-5.0.1. Table 7 presents the run time ratio of the proposed scheme and the anchor. As can be seen, the proposed scheme could reduce the overall encoding time and decoding time.

Table 5. The BD-rate performance of the proposed improved DV searching order compared to HTM-5.0.1 for all input views

Sequence	video0	video1	video2
Balloons	0.0%	-0.2%	-0.5%
Kendo	0.0%	-0.7%	-0.6%
Newspapercc	0.0%	-0.3%	-0.2%
GhostTownFly	0.0%	-0.1%	-0.1%
PoznanHall2	0.0%	0.2%	0.2%
PoznanStreet	0.0%	-0.4%	0.0%
UndoDancer	0.0%	-0.3%	0.3%
1024x768	0.0%	-0.4%	-0.4%
1920x1088	0.0%	-0.1%	0.1%
average	**0.0%**	**-0.3%**	**-0.1%**

Table 6. The BD-rate performance of the proposed improved DV searching order compared to HTM-5.0.1 for overall coded and synthesized views

Sequence	video only	synthesized only	coded & synthesized
Balloons	-0.2%	-0.1%	-0.1%
Kendo	-0.3%	-0.2%	-0.2%
Newspapercc	-0.1%	-0.1%	-0.1%
GhostTownFly	0.0%	0.0%	0.0%
PoznanHall2	0.1%	0.1%	0.1%
PoznanStreet	-0.1%	-0.1%	-0.1%
UndoDancer	0.0%	-0.1%	-0.1%
1024x768	-0.2%	-0.1%	-0.1%
1920x1088	0.0%	0.0%	0.0%
average	**-0.1%**	**-0.1%**	**-0.1%**

Table 7. Run time ratio of the proposed simplified scheme for improved DV searching order over HTM-5.0.1

Sequence	encoding time	decoding time	synthesis time
Balloons	98.4%	98.5%	99.8%
Kendo	97.8%	95.9%	100.8%
Newspapercc	98.3%	98.6%	99.9%
GhostTownFly	98.6%	103.9%	98.9%
PoznanHall2	98.4%	101.7%	98.7%
PoznanStreet	99.5%	95.5%	100.4%
UndoDancer	97.9%	97.4%	98.9%
1024x768	98.2%	97.6%	100.2%
1920x1088	98.6%	99.5%	99.2%
average	**98.4%**	**98.7%**	**99.6%**

3.3. Results of combination of the simplified scheme and improved DV searching order

Table 8 and Table 9 provide the experimental results of combining the simplified DV derivation and the improved

DV searching order. Compared to the proposed scheme for simplified DV derivation, the combining scheme achieves 0.2% and 0.1% BD-rate reduction for video1 and video2 respectively.

Table 8. The BD-rate performance of the combination compared to the simplified scheme for all input views

Sequence	video 0	video 1	video 2
Balloons	0.0%	-0.4%	-0.4%
Kendo	0.0%	-0.7%	-0.1%
Newspapercc	0.0%	-0.2%	-0.2%
GhostTownFly	0.0%	-0.2%	-0.1%
PoznanHall2	0.0%	0.1%	0.2%
PoznanStreet	0.0%	-0.2%	-0.1%
UndoDancer	0.0%	0.0%	0.2%
1024x768	0.0%	-0.4%	-0.2%
1920x1088	0.0%	-0.1%	0.1%
average	0.0%	-0.2%	-0.1%

Table 9. The BD-rate performance of the combination compared to the simplified scheme for overall coded and synthesized views

Sequence	video only	synthesized only	coded & synthesized
Balloons	-0.2%	0.0%	-0.1%
Kendo	-0.2%	-0.1%	-0.1%
Newspapercc	-0.1%	-0.1%	-0.1%
GhostTownFly	0.0%	0.0%	0.0%
PoznanHall2	0.0%	0.1%	0.0%
PoznanStreet	0.0%	0.0%	0.0%
UndoDancer	0.0%	0.1%	0.1%
1024x768	-0.1%	-0.1%	-0.1%
1920x1088	0.0%	0.0%	0.0%
average	**-0.1%**	**0.0%**	**0.0%**

4. CONCLUSION

In this paper, we propose to unify the searching order of temporal blocks for all dependent views. We also propose to align the temporal blocks used in the DV derivation and those used for the TMVP derivation. With the proposed simplification, the DV derivation process is simplified and the memory access is also reduced. Experimental results show that the proposed simplification introduces no BD-rate change for overall coding performance. Besides, we also propose a new searching order for the DV derivation and the results show that the proposed order can bring 0.1% overall coding performance gain compared to HTM-5.0.1 without complexity increase.

ACKNOWLEDGMENT

This work was supported in part by the Major State Basic Research Development Program of China (973 Program 2009CB320905), the 863 program (2012AA011505), the National Science Foundation of China (61103088, 61121002), and the Program for New Century Excellent Talents in University (NCET) of China under grants NCET-11-0797.

5. REFERENCES

[1] H. Schwarz, C. Bartnik, S. Bosse et al., "Description of 3D Video Technology Proposal by Fraunhofer HHI (HEVC compatible; configuration A)," M22570, Geneva, Nov. 2011.

[2] H. Schwarz, K. Wegner, "Test Model under Consideration for HEVC based 3D video coding v3.0," N12744, Geneva, Apr. 2012.

[3] L. Zhang, Y. Chen, M. Karczewicz, "3D-CE5.h related: Disparity vector derivation for multiview video and 3DV," m24937, Geneva, Apr. 2012.

[4] J. Sung, M. Koo, S. Yea, "3D-CE5.h: Simplification of disparity vector derivation for HEVC-based 3D video coding," Document of Joint Collaborative Team on 3D Video Coding Extension Development, JCT2-A0126, Stockholm, Jul. 2012.

[5] J. Kang, Y. Chen, L. Zhang, M. Karczewicz, "3D-CE5.h related: Improvements for disparity vector derivation," Document of Joint Collaborative Team on 3D Video Coding Extension Development, JCT3V-B0047, Shanghai, Oct. 2012.

[6] https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-5.0.1/

[7] J. Lee, H. Wey, D. Park, "3D-CE2.h related results on disparity vector derivation," Document of Joint Collaborative Team on 3D Video Coding Extension Development, JCT3V-C0097, Geneva, Jan. 2013.

[8] D. Rusanovskyy, K. Müller, A. Vetro, "Common test conditions of 3DV Core Experiments," Document of Joint Collaborative Team on 3D Video Coding Extension Development, JCT3V-B1100, Shanghai, Oct. 2012.