A Local-Adapted Disparity Vector Derivation Scheme for 3D-AVS

Qi Mao #1, Shanshe Wang #1, Jing Su *1, Xiang Zhang #1, Xinfeng Zhang †2, Siwei Ma #*1

Institute of Digital Media, Peking University, Beijing, China

* Peking University Shenzhen Graduate School, Shenzhen, China

† Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore

1 {qimao, sswang, sujing_jy, x_zhang, swma}@pku.edu.cn 2 xfzhang@ntu.edu.sg

***Abstract*—In the 3D extension of Audio Video Coding Standard (AVS), i.e. 3D-AVS, the Global Disparity Vector (GDV) derivation technique has been proposed to provide an estimation for Disparity Vector (DV) in inter-view prediction, where the GDV is generated by averaging all Disparity Vectors (DVs) in the latest previously coded frame. The prediction accuracy of GDV may be however limited by the lack of local adaptivity. In this paper, we introduce a novel Local Disparity Vector (LDV) derivation scheme. Specifically, the DV of the current block is calculated from the DVs within a neighboring region, whose size can be adaptively expanded to increase the robustness and accuracy. Experimental results show that the proposed LDV derivation method can provide around 2.12% and 1.37% bitrate reductions for compressed views and synthesized views compared with the GDV scheme, respectively.**

***Index Terms*—3D video coding, local disparity vector, 3D-AVS.**

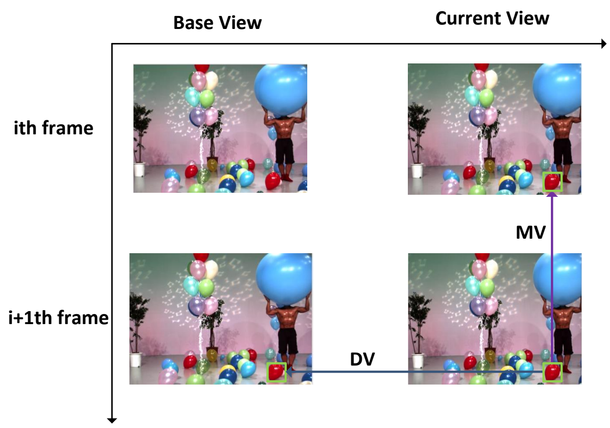

Fig. 1. Illustration of Disparity Vector (DV) and Motion Vector (MV).

I. Introduction

With the development of three-dimensional (3D) video techniques, 3D video services such as 3D television [1] and Free Viewpoint Television [2] (FTV) become more and more popular. To better support various 3D video applications, the working group of China Audio Video Coding Standard (AVS) [3][4] begins to standardize 3D video coding and develop a 3D extension of AVS called 3D-AVS. Since all cameras capture the same scene simultaneously from different views, 3D video also contains inter-view redundancy besides spatial redundancy and temporal redundancy. Therefore, 3D-AVS has included all efficient coding tools in 2D video and adds some new tools for 3D video coding to further improve the coding efficiency.

In 3D-AVS, the base view can be encoded independently by traditional 2D video coding standard, and other views, i.e. dependent view, can be efficiently coded by referring to the base view due to the strong correlation between video views. To best utilize the inter-view correlation, extensive 3D coding tools have been developed such as the Disparity Compensated Prediction (DCP) and the Inter-View Motion Prediction (IVMP) [5]. In Motion Compensated Prediction (MCP), the Motion Vector (MV) is to find the optimal reference between video frames. Analogously, DCP exploits the similarity between video views by the Disparity Vector (DV), as shown in Fig. 1. The IVMP can directly derive the motion information from the DV-pointed Prediction Unit (PU) in the base view. It can be seen that the DV plays an essential role in the DCP, IVMP and other inter-view techniques. As the accuracy of the derived DV can greatly affect the inter-view prediction, how to efficiently derive an effective DV is a critical issue in 3D video coding.

In 3D-AVS, the Global Disparity Vector (GDV) [6] is adopted as the DV derivation method. It is calculated by averaging all the disparity vectors of 16×16 block in the latest previous coded frame. In the 3D extension of High Efficiency Video Coding (3D-HEVC), a Neighboring Blocks Disparity Vector (NBDV) [7] scheme is proposed, where the DV can be derived from spatial and temporal neighboring blocks. Motivated by NBDV, in this paper, we propose a novel Local Disparity Vector (LDV) derivation scheme to replace the GDV in 3D-AVS, where a Local-Adapted Neighboring Region (LANR) is defined and all the DVs in this region are averaged to generate the LDV.

The rest of this paper is organized as follows. Section II describes the motivation and details of the proposed LDV derivation method. Experimental results are presented in Section III. Finally, Section IV concludes this work.

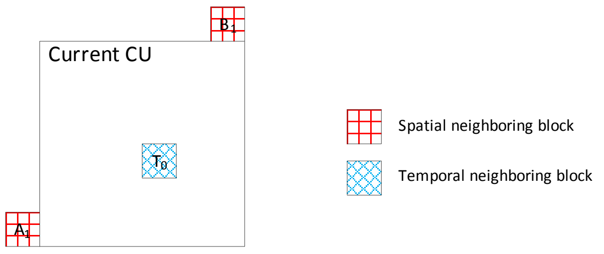

Fig. 2. Neighboring Blocks Disparity Vector (NBDV) scheme [10].

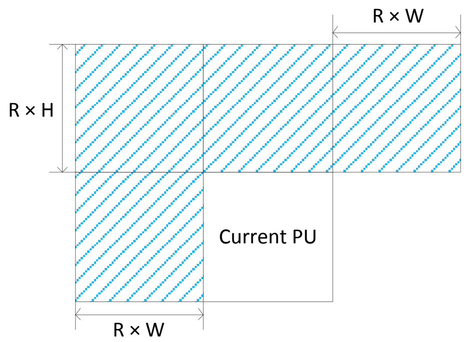

Fig. 3. Local-Adapted Neighboring Region (LANR).

II. Proposed Method

A. Motivation

In 3D-HEVC, the NBDV scheme is applied to derive a local DV of current Coding Unit (CU) from its spatial and temporal neighboring blocks, if the neighbouring block is disparity compensated or by Disparity Vector Motion Compensated Prediction (DVMCP) [8], [9], as shown in Fig. 2. In DVMCP, the DV utilized to specify an inter-view block for motion information derivation is considerd.

Generally, NBDV is more flexible and accurate comparing with the GDV method in current 3D-AVS due to its local adaptivity. Nevertheless, the efficiency of NBDV may be still limited when the neighboring blocks are coded by neither DCP nor DVMCP, in this case a default zero DV is used as substitution, which may be inefficient for inter-view prediction.

B. LDV Derivation Based on LANR

To solve this problem, an improved method is proposed in this work, where the Local-Adapted Neighboring Region (LANR) is defined to avoid invalid DV derivation.

As shown in Fig. 3, the LANR of current Prediction Unit (PU) is defined as its left, upper left, upper and upper right blocks, which are already compressed and their coding information can be utilized to derive the DV for current PU. In the figure, W and H indicate the width and height of current PU. The parameter R is utilized to adaptively adjust the range of LANR.

Subsequently, the Local DV $\boldsymbol{LDV_R}$ can be calculated by averaging all the DVs of 4×4 blocks in the LANR as follows,

$$\boldsymbol{LDV_R} = \frac{1}{N_R} \sum_{\boldsymbol{DV_i} \in \boldsymbol{\Omega_R}} \boldsymbol{DV_i}, \tag{1}$$

where $\boldsymbol{\Omega_R}$ indicates the set of all DVs in the LANR given R, and N_R is size of $\boldsymbol{\Omega_R}$.

Furthermore, we try to address the problem of how to adaptively change the range of LANR for maximum information utilization. The algorithm of the proposed method is shown in Alg. 1. Firstly we examine whether a non-zero $\boldsymbol{LDV_R}$ is derived by initializing $R = 1$. If not, R would be continually increased by a step of 1 to expand the LANR until a non-zero $\boldsymbol{LDV_R}$ is acquired or the R reaches the maximum fixed R_M. Finally, if the derived $\boldsymbol{LDV_R}$ is non-zero, it will replace the $\boldsymbol{GDV}$ in 3D-AVS.

Algorithm 1: Algorithm of Local Disparity Vector (LDV) Derivation Based on LANR

Input:
- R_M : Given Maximum R
- $\boldsymbol{GDV}$: Global Disparity Vector

Output:
- $\boldsymbol{LDV_R}$: Local Disparity Vector

Initialization:
- $\boldsymbol{LDV_R} = \mathbf{0}$
- $\boldsymbol{LDV_S} = \mathbf{0}$
- $R = 1$

```
while (R ≤ R_M) and (LDV_R == 0) do
    Find Ω_R in LANR;
    if Ω_R exists then
        for DV_i ∈ Ω_R do
            LDV_S = LDV_S + DV_i;
        LDV_R = LDV_S / N_R;
    R = R + 1;
if LDV_R == 0 then
    LDV_R = GDV;
return LDV_R;
```

III. Experimental Result

To verify the performance of the proposed method, it has been implemented on the 3D-AVS reference software RFD-4.0 [11] and simulated strictly in accordance with the common test conditions of 3D-AVS [12], where the low-delay P (LDP) configuration is used for simulation. Note that the GDV has already involved in the RFD-4.0 platform, which is set as the anchor. The information of all the five test sequences in 3D-AVS is listed in Table I.

The commonly used BD-rate index [13] is utilized for comparing the performance of two codecs. A negative value of the BD-rate indicates coding gains over the anchor. In Table II and Table III, The first and second column represent the BD-rate performance considering Y-PSNR of view 1 and 2 (dependent view). The third and fourth column represent the

TABLE I
TESTING SEQUENCES

Sequence Name	Resolution	Frame Rate	Input Views
Balloons	1024x768	30	3-1-5
Kendo	1024x768	30	3-1-5
Newspaper	1024x768	30	4-2-6
PoznanHall	1920x1088	25	6-5-7
PoznanStreet	1920x1088	25	5-1-9

TABLE IV
ENCODING TIME COMPARISON BETWEEN ANCHOR AND THE PROPOSED LDV WITH DIFFERENT R_M

Methods	Anchor	$R_M = 1$	$R_M = 4$
Encoding Time	100%	101.21%	101.90%

BD-rate performance considering Y-PSNR of the coded texture views over the bitrates of texture data and over the bitrates of texture data and depth data. The last column represents the BD-rate performance considering Y-PSNR of the synthesized texture views over the bitrates of texture data and depth data.

A. Results of LDV Derivation Scheme

Firstly, the overall performance is evaluated, where Table II shows the results of dependent views, compressed views and synthesized views when $R_M = 1$, and Table III shows the results when $R_M = 4$. From the results, It can be observed that the proposed LDV scheme can reach 1.61% and 0.95% BD-rate saving when $R_M = 1$, and 2.12% and 1.37% BD-rate saving when $R_M = 4$ for the compressed views and synthesized views, respectively. Table IV demonstrates that not much complexity increase is introduced by the improved scheme.

B. Analysis of LDV Derivation Scheme

To evaluate the performance of LDV, we separately count the usage proportion of GDV and LDV when $R_M = 1$ and $R_M = 4$. It can be confirmed that the coding efficiency improvements in Table II and Table III are all brought by LDV scheme. As shown in Fig. 4 and Fig. 5, with the increase of R_M, we can get more non-zero LDVs.

Furthermore, the parameter impact of R_M on the coding performance is investigated. Fig. 6 illustrates the BD-rate gain for the value of R_M ranging from 1 to 5. It can be clearly explained that the larger the LANR size used, the better the performance improvements obtained. However, these improvements taper off as the LANR becomes larger.

IV. CONCLUSION

In this paper, we propose to derive a novel Local Disparity Vector (LDV) using Local-Adapted Neighboring Region (LANR), and it will replace GDV if it is not zero vector. Experimental results show that the proposed LDV method provides 1.61% and 0.95% BD-rate saving for the compressed views and synthesized views when $R_M = 1$, while provides 2.12% and 1.37% BD-rate saving when $R_M = 4$ without much complexity increase. It can be concluded that LDV performs better than GDV in 3D-AVS. With the expand of the LANR, more non-zero LDV will be derived and can achieve more coding efficiency .

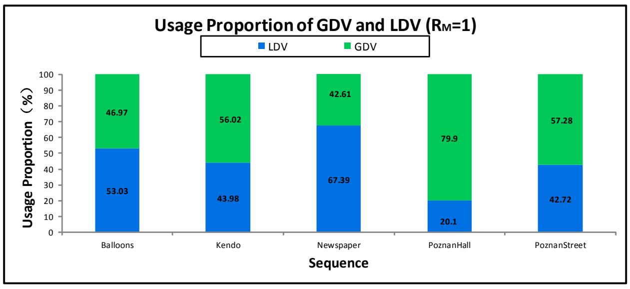

Fig. 4. Usage proportion of GDV and LDV ($R_M = 1$).

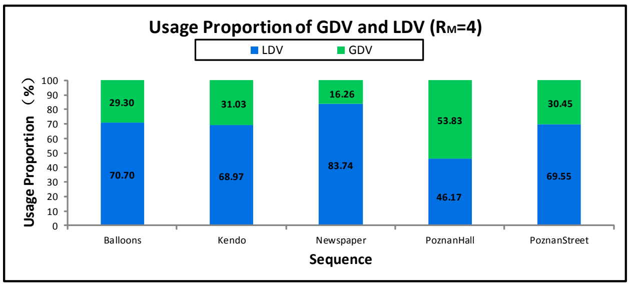

Fig. 5. Usage proportion of GDV and LDV ($R_M = 4$).

ACKNOWLEDGMENT

This work was supported in part by the National High-tech R&D Program of China (863 Program, 2015AA015903), National Natural Science Foundation of China (61322106, 61571017, 61421062), Shenzhen Peacock Plan, which are gratefully acknowledged.

REFERENCES

[1] C. Fehn, E. Cooke, O. Schreer, and P. Kauff, "3D analysis and image-based rendering for immersive TV applications," *Signal Processing: Image Communication*, vol. 17, no. 9, pp. 705–715, 2002.

[2] J. Carranza, C. Theobalt, M. A. Magnor, and H.-P. Seidel, "Free-viewpoint video of human actors," *ACM transactions on graphics (TOG)*, vol. 22, no. 3, pp. 569–577, 2003.

[3] S. Ma, S. Wang, and W. Gao, "Overview of IEEE 1857 video coding standard," in *IEEE International Conference on Image Processing*, 2013, pp. 1500–1504.

[4] L. Fan, S. Ma, and F. Wu, "Overview of AVS video standard," in *IEEE International Conference on multimedia*, vol. 1, 2004, pp. 423–426.

[5] J. Ma, N. Zhang, X. Fan, R. Xiong, and D. Zhao, "Quad-tree based Inter-view Motion Prediction," in *Visual Communications and Image Processing (VCIP)*, 2015.

[6] J. Ma, Y. Zhang, N. Zhang, X. Fan, and D. Zhao, "An effective method for obtaining disparity vector," *in Document M3582, 51st AVS Meeting, Fuzhou, China*, 2014.

[7] L. Zhang, Y. Chen, and M. Karczewicz, "3D-CE5.h related: Disparity vector derivation for multiview video and 3DV," *ISO/IEC JTC1/SC29/WG11 MPEG2012/m24937*, 2012.

[8] J. Sung, M. Koo, and S. Yea, "3D-CE5.h: Simplification of disparity vector derivation for HEVC-based 3D video coding," *ITU-T SG*, 2012.

TABLE II
BD-RATE PERFORMANCE OF THE PROPOSED LDV DERIVATION ($R_M = 1$) COMPARED TO RFD-4.0 FOR ALL CODED AND SYNTHESIZED VIEWS

Sequence	video1	video2	video PSNR \video bitrate	video PSNR \total bitrate	synth PSNR \total bitrate
Balloons	-4.84%	-2.02%	-1.68%	-1.42%	-1.10%
Kendo	-3.35%	-2.04%	-1.37%	-1.17%	-0.87%
Newspaper	-4.85%	-1.23%	-1.64%	-1.45%	-0.86%
PoznanHall	-3.03%	-1.79%	-1.32%	-1.15%	-0.67%
PoznanStreet	-2.81%	-5.44%	-1.90%	-1.84%	-1.23%
1024x768	-4.35%	-1.77%	-1.56%	-1.35%	-0.94%
1920x1088	-2.92%	-3.62%	-1.61%	-1.49%	-0.95%
average	**-3.78%**	**-2.51%**	**-1.58%**	**-1.61%**	**-0.95%**

TABLE III
BD-RATE PERFORMANCE OF THE PROPOSED LDV DERIVATION ($R_M = 4$) COMPARED TO RFD-4.0 FOR ALL CODED AND SYNTHESIZED VIEWS

Sequence	video1	video2	video PSNR \video bitrate	video PSNR \total bitrate	synth PSNR \total bitrate
Balloons	-6.08%	-2.80%	-2.28%	-2.03%	-1.53%
Kendo	-3.97%	-2.77%	-1.77%	-1.58%	-1.12%
Newspaper	-6.35%	-2.14%	-2.38%	-2.16%	-1.21%
PoznanHall	-4.84%	-3.25%	-2.36%	-2.16%	-1.15%
PoznanStreet	-3.96%	-7.31%	-2.78%	-2.68%	-1.84%
1024x768	-5.47%	-2.57%	-2.14%	-1.92%	-1.29%
1920x1088	-4.40%	-5.28%	-2.57%	-2.42%	-1.50%
average	**-5.04%**	**-3.66%**	**-2.32%**	**-2.12%**	**-1.37%**

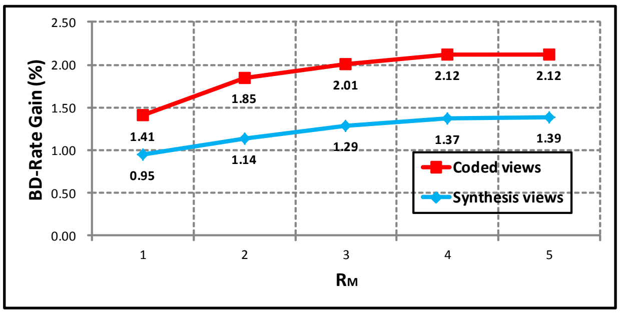

Fig. 6. Parameter R_M impact on the coding gains of coded views and synthesis views, respectively.

[9] J. Kang, Y. Chen, L. Zhang, and M. Karczewicz, "3D-CE5.h related: Improvements for disparity vector derivation," in *Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) document JCT3V-B0047, 2nd Meeting, Shanghai, CN*, 2012.

[10] W. K. Y. S. Chen Y, Tech G, "3D-HEVC test model 11," in *Joint collaborative team on 3D video coding extensions (JCT-3V) document JCT3V- K1003, 11th Meeting, Geneva, CH*, 2015.

[11] S. Wang, "AVS2-P2-3D: Describe of the RFD 4.0 platform," in *Document N2246, 55th AVS Meeting, Beijing, China*, 2015.

[12] ——, "AVS2-P2-3D: RFD Common Test Condition and Software reference configuration," in *Document N2247, 55th AVS Meeting, Beijing, China*, 2015.

[13] G. Bjontegaard, "Calcuation of average psnr differences between rd-curves," *Doc. VCEG-M33 ITU-T Q6/16, Austin, TX, USA*, 2001.