FULL PARALLAX SUPER MULTI-VIEW VIDEO CODING

A. Dricot [1,2], *J. Jung* [1]

M. Cagnazzo [2], *B. Pesquet* [2], *F. Dufaux* [2]

[1] Orange Labs

[2] Institut Mines-Télécom;
Télécom ParisTech; CNRS LTCI

ABSTRACT

Super Multi-View (SMV) video is a key enabler for future 3D video services that allows a glasses-free visualization and eliminates many causes of discomfort existing in current available 3D video technologies. SMV video content is composed of tens or hundreds of views, that can be aligned in horizontal only or both horizontal and vertical directions, providing respectively horizontal parallax or full parallax. This paper compares several coding schemes and coding orders, and proposes a coding structure that exploits inter-view correlations in the two directions, providing BD-rate gains up to 29.1% when compared to a basic anchor structure. Additionally, Neighboring Block Disparity Vector (NBDV) and Inter-View Motion Prediction (IVMP) coding tools are further improved to efficiently exploit coding structures in two dimensions, with BD-rate gains up to 4.2% reported over the reference 3D-HEVC encoder.

***Index Terms*—** 3D Video compression, Multi-view, Full Motion Parallax

1. INTRODUCTION

3D video aims at enabling an immersive viewing experience. Current 3D video technologies available on the consumer market present however several limitations. In stereoscopy, the lack of comfort due to the use of glasses is combined with annoying perception stimuli like vergence-accommodation conflict that causes eyestrain and headaches. In existing glasses-free auto-stereoscopic systems, the small number of views induces artifacts, lack of smooth motion parallax (i.e. the visualization is not continuous when moving in front of the display) and a restricted viewing zone that particularly alter the quality and comfort of visualization [1].

A study of Super Multi-View video (SMV) has been initiated during the October 2013 MPEG FTV meeting [2]. SMV uses tens or hundreds of views to create a lightfield representation of a scene. This representation tends to eliminate many of current 3D technologies artifacts, especially the vergence-accomodation issue. It allows a glasses-free realistic visualization with a smooth motion parallax, which is a key cue in the perception of depth, in horizontal and potentially vertical directions. Several companies have already shown interest in SMV by working on ligthfield display systems. A typical target use case would be the live 3D broadcast of sport events, like 2020 Olympics in Japan, that could be shot by SMV camera arrays and projected on large SMV display systems at several public viewing facilities in major cities around the world [3]. Therefore a need exists for new efficient coding technologies that can handle the large amount of data required for SMV [4].

Multi-view standard encoder extensions are adequate to encode SMV content with horizontal only parallax. Modifications of these encoders have been proposed in the literature to encode content with full parallax. State of the art methods present however limitations in the use of the two dimensions for inter-view predictions. Here we propose an efficient inter-view prediction scheme to exploit horizontal and vertical dimensions at the coding structure level. Then we propose improvements of inter-view coding tools to exploit the two dimensional structure also at the coding unit level.

This paper is organized as follows. Section 2 describes state of the art methods for full parallax SMV encoding. In section 3, we describe the proposed inter-view reference picture scheme and show experimental results against state of the art schemes. Improved interview coding tools adapted to full parallax are proposed in section 4 including experimental results. Section 5 finally concludes this paper.

2. STATE OF THE ART

2.1. Multi-view video coding standards

SMV defines 3D multi-view video content with tens or hundreds of views, with either horizontal only or full motion parallax. The massive number of views increases the amount of data to process compared to current 3D video technologies. The amount of inter-view correlation available is also increased. Current multi-view standard encoders have been designed for horizontal parallax content with limited number of views. MVC and MV-HEVC are the multi-view extensions of respectively H.264/AVC and HEVC standards [5]. These extensions provide additional high level syntax that allows the inter-view prediction. 3D-HEVC extension provides depth related tools and new tools at Coding Unit level (CUs in HEVC replace H.264/AVC macroblocks) for side views.

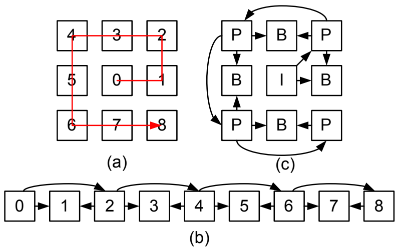

Fig. 1. State of the art method [9] for 9 views (0...8)
(a) spiral scan, (b) IBP structure for inter-view prediction,
(c) equivalent IBP scheme in 2 dimensions

In the reference software version used in our experiments (HTM7.0), the following applies: Neighboring Block Disparity Vector (NBDV) [6] and Inter-View Motion Prediction (IVMP) [7] are specific 3D-HEVC coding tools designed for standard horizontal multi-view encoding. For the current CU, NBDV searches for a disparity vector (DV) through already coded temporal and spatial neighboring CUs. The DV derived by NBDV is used by IVMP to create the Inter-view Motion Predicted Candidate (IMPC). IMPC corresponds to the motion parameters (motion vectors and temporal references) of the CU pointed by the DV in the reference view. IPMC is introduced at the first place in the merge [8] candidate list. Finally the DV itself is inserted in the merge list as Disparity Motion Vector candidate (DMV).

2.2. Improvement for full parallax configuration

The first approach considered to encode full parallax SMV content is the use of a multi-view standard encoder with an adaptation at the inter-view references structure level. In [9], the views are first scanned in spiral as illustrated in Figure 1 (a) and realigned horizontally. Then the horizontal arrangement is coded using a IBP prediction structure (b) by an MVC encoder. Figure 1 (c) shows the resulting scheme of equivalent IBP structure with the views represented in two dimensions. The main drawback of this approach is the introduction of unsuitable and ineffective predictions.

In [10], it is proposed to apply horizontal IPP or IBP structures (Fig. 4(e) and (f)) to each line of the views array, and to add vertical inter-view prediction only for the first or central column of views as illustrated in Figure 2 (a),(b) and (c). The number of available vertical inter-view predictions is very limited in such structures.

In [11], [12] and [13], another structure is proposed as illustrated in Figure 2 (d). Each line of views uses an horizontal IBP structure and additional vertical inter-view predictions are introduced, giving views of types B1 with two horizontal or vertical only references, B2 with one horizontal and two vertical references, and B3 with two references in both directions. The main limitation is the small number of views that use both horizontal and vertical references (less than half of the views are of types B2 or B3) and the distance between the coding views and reference views.

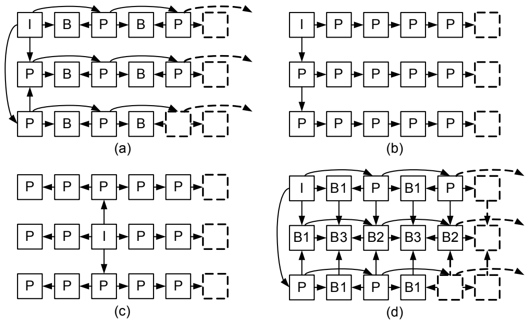

Fig. 2. State of the art structures
(a),(b) and (c) proposed in [10],
and (d) proposed in [11][12][13].

A second approach at the coding unit level is considered in [14] and in [11], [12] and [13]. Similar methods are proposed based on the prediction of a DV for the current coding view by interpolation of DVs from neighboring views.

3. PROPOSITION FOR A NEW INTER-VIEW REFERENCE PICTURES CONFIGURATION

3.1. Reference and proposed schemes

We propose a two dimension inter-view reference picture structure, namely *Central2D*, that can exploit efficiently a two dimensional view alignment as illustrated in Figure 3 (b). For a $N \times M$ views configuration, *Central2D* scheme is built as follows. The central view is coded first and cannot use inter-view references. The $N-1$ (respectively $M-1$) views that are in the same horizontal (resp. vertical) axis as the central view are then coded using only one inter-view reference, being the nearest view in the central direction. All the other views are coded using one horizontal and one vertical inter-view references being the nearest views in the central direction, hence it allows the use of an horizontal and a vertical inter-view reference picture for a large number of views (only $M+N-1$ views do not have horizontal and vertical reference pictures). Moreover this method minimizes the distance between the coding views and their inter-view

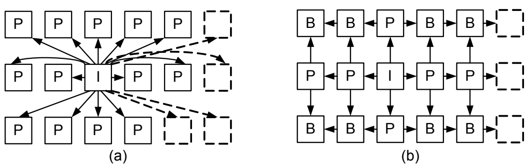

Fig. 3. (a) basic anchor, (b) proposed *Central2D*

reference pictures and does not use diagonal references.

In the following section, the proposed scheme is compared to a basic anchor (see Figure 3 (a)) with only the central view as inter-view reference picture for all the other views, in order to assess the benefit of inter-view prediction in two directions and of a small distance between the coding and the reference views. State of the art structures are also tested in our experiments: [10] and [11] correspond to the schemes illustrated in Figure 2 (c) and (d). [9] corresponds to the spiral scan with IBP structure (see Figure 1). For comparison purpose, we also propose to extend method [9] by varying the scan order and the structure as illustrated in Figure 4.

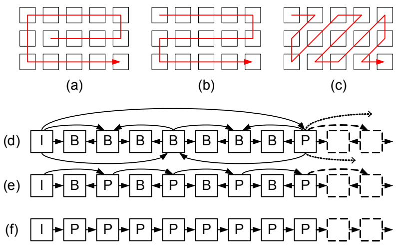

Fig. 4. Scan order: (a) spiral, (b) perpendicular, (c) diagonal and Horizontal inter-view reference picture structures: (d) hierarchical, (e) IBP, (f) IPP

3.2. Experimental results

In this section, we test the state of the art and proposed schemes within MV-HEVC. The temporal prediction structure is as described in the Common Test Conditions (CTC) [15]. Experiments are performed under MV-HEVC reference software version 7.0 (HTM7.0 with QC_MVHEVC macro). Two sequences are tested: *CoastalGuard* (50 frames, computer generated, resolution 768×384) and *Akko&Kayo* (290 frames, captured, resolution 640×480). Configurations of 3×3 views and 11×5 views are tested. Results are measured using the Bjøntegaard Delta (BD) rate [16] on the QPs range 22-27-32-37. The reference is the basic anchor scheme (Figure 3 (a)). Negative values represent improvement over the anchor.

Table 1 shows that for both sequences with a 3×3 views configuration, the *Central2D* scheme, method [10] and IPP structure with perpendicular and spiral scan outperform the other methods. These schemes do not use diagonal inter-view reference pictures and minimize the distance between the coding views and the inter-view reference pictures. The extra gain for *Central2D* is due to the use of both horizontal and vertical inter-view reference pictures. Table 2 shows that *Central2D* remains the most coherent and efficient configuration with a larger number of views.

The final BD-rate gain for the proposed structure *Central2D* against the basic anchor is up to 8.2% and 29.1% in the 3×3 and 11×5 views configuration respectively.

Coast 3 × 3			
	spiral	perpendicular	diagonal
IPP	-1.2%	-2.2%	5.1%
IBP	9.1%	7.1%	11.4%
Hierarchical	3.0%	4.4%	8.4%
Method [11]	2.1%		
Method [10]	-6.8%		
CENTRAL2D	-7.1%		
Akko 3 × 3			
	spiral	perpendicular	diagonal
IPP	-4.9%	-5.5%	8.8%
IBP	2.7%	-4.0%	-1.9%
Hierarchical	1.9%	2.4%	4.0%
Method [11]	7.8%		
Method [10]	-7.7%		
CENTRAL2D	-8.2%		

Table 1. BD-rate variations for state of the art and proposed structures compared to basic anchor - *with* 3×3 *views*

Coast 11 × 5			
	spiral	perpendicular	diagonal
IPP	-20.5%	-19.6%	16.1%
IBP	-15.9%	-14.9%	-13.9%
Hierarchical	-8.4%	-9.3%	-13.0%
Method [11]	-19.5%		
Method [10]	-24.4%		
CENTRAL2D	-29.1%		
Akko 11 × 5			
	spiral	perpendicular	diagonal
IPP	-22.9%	-24.8%	-6.5%
IBP	-20.0%	-23.4%	-2.4%
Hierarchical	-14.9%	-20.2%	-3.7%
Method [11]	-24.2%		
Method [10]	-25.9%		
CENTRAL2D	-27.6%		

Table 2. BD-rate variations for state of the art and proposed structures compared to basic anchor - *with* 11×5 *views*

4. ADAPTATION AND IMPROVEMENT OF INTER-VIEW CODING TOOLS

4.1. Merge candidate list improvement

In this section we propose a normative modification of the NBDV and IVMP coding tools. NBDV and IVMP are specific coding tools implemented to work in the Common Test Conditions [15], i.e with only one horizontal inter-view reference picture, which is the central baseview (with view index 0). We adapt these tools by allowing the use of several inter-view reference pictures with a view index different from 0, and possibly horizontal or vertical.

In addition to this adaptation we improve NBDV as follows. When encoding one of the B views that use one horizontal and one vertical inter-view reference pictures, the modified NBDV searches for two DVs (one for each inter-view reference picture). The search of a second DV does not provide BD-rate gain in itself but will be used for IMPC and DMV. The new second DV is used to introduce a second IMPC at the second place of the merge candidate list. For the DMV merge candidate, the couple of DVs is used, allowing an inter-view bi-prediction in both directions at the same time.

4.2. Inter-view derivation of the second DV

We propose to increase the chances of finding a second DV with NBDV in order to improve the efficiency of modified IMPC and DMV candidates. The steps are illustrated in Figure 5. For the current coding view NBDV must find a first horizontal DV pointing a reference CU in an inter-view reference picture. If this horizontal reference picture uses itself a vertical inter-view reference picture and if the reference CU is coded by inter-view prediction, the vertical DV used for the prediction is inherited/derivated as a second DV for the current coding CU, and then used by IMPC and DMV as described in the previous section. We note that this method can be used for B views with one horizontal and one vertical references, which makes the *Central2D* structure the most adequate for this coding tools.

4.3. Experimental results

In this section, we test the proposed modifications on NBDV and IVMP coding tools. The experiments are performed under 3D-HEVC reference software version 7.0 (HTM7.0). The test conditions are the same as in Sec. 3.2, (i.e. allowing two dimensional structures configuration). Previously proposed *Central2D* structure is used in all following experiments. The reference is HTM7.0 without software modifications.

Table 3 shows that the adaptation of NBDV and IVMP to a two dimensions structure brings BD-rate gains up to 3.3%, confirming the impact of the use of horizontal and vertical dimensions at the inter-view references structure level. The insertion of a second IMPC in the merge candidate list and the bi-prediction for the DMV merge candidate separately increase the gains up to 2.4% for the 3×3 views configuration and 3.7% for the 11×5 configuration. The combination of both improvements provides a gain up to 2.5% and 3.9%

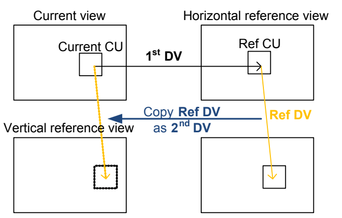

Fig. 5. Inter-view derivation of a second DV

Reference: 3D-HEVC (HTM7.0 without modifications)				
	3×3 views		11×5 views	
	Coast	Akko	Coast	Akko
Adaptation only	-1.1%	-2.3%	-2.4%	-3.3%
BiDMV	-1.2%	-2.4%	-2.7%	-3.7%
2 IMPC	-1.1%	-2.3%	-2.8%	-3.5%
Both	-1.3%	-2.5%	-3.1%	-3.9%

Table 3. BD-rate variations for improved NBDV and IVMP using one DV for each inter-view reference picture

Reference: 3D-HEVC (HTM7.0 without modifications)				
	3×3 views		11×5 views	
	Coast	Akko	Coast	Akko
BiDMV + derivation	-1.9%	-2.9%	-3.4%	-3.9%
2 IMPC + derivation	-1.3%	-2.4%	-2.8%	-3.5%
Both + derivation	-2.0%	-2.9%	-3.9%	-4.2%

Table 4. BD-rate variations for improved NBDV and IVMP with inter-view derivation of a second DV

respectively with 3×3 and with 11×5 views. The results for the combination of both tools are slightly higher than the sum of each taken separately because the bi-prediction allows NBDV to find more often a second DV, hence increasing the chances to have a relevant second IMPC candidate.

Table 4 shows that the proposed derivation for the second DV is efficient and increases the encoding performance of the complete proposed method (including the adaptation of NDBV and IVMP to a full parallax structure, the two IPMC, the DMV bi-prediction and the inter-view derivation of the second DV) up to 2.9% and 4.2% for the sequence *Akko&Kayo* respectively with 3×3 and with 11×5 views.

5. CONCLUSION

In this paper we propose an inter-view reference picture structure adapted to 3D lightfield video content with full motion parallax (horizontal and vertical view alignment). Its main features are the minimal distance between the coded and the reference views, and the use of both horizontal and vertical inter-view references. The proposed scheme outperforms a basic anchor by up to 29.1% in BD-rate, showing the impact of an efficient use of both horizontal and vertical directions in the inter-view reference picture scheme. We also propose to improve 3D-HEVC coding tools NBDV and IVMP in order to exploit both horizontal and vertical directions in a full parallax configuration, providing BD-rate gains up to 4.2%. The results of the proposed methods show that exploiting efficiently both horizontal and vertical dimensions of full parallax SMV content at the coding structure and coding tools level significantly improves the compression performance.

6. REFERENCES

[1] F. Dufaux, B. Pesquet-Popescu, and M. Cagnazzo, *Emerging technologies for 3D video: content creation, coding, transmission and rendering*, Wiley Eds, 2013.

[2] M. P. Tehrani, T. Senoh, M. Okui, K. Yamamoto, N. Inoue, and T. Fujii, "[m31095][FTV AHG] Use cases and application scenarios for super multiview video and free-navigation," in *International Organisation For Standardisation*. ISO/IEC JTC1/SC29/WG11, October 2013.

[3] M. P. Tehrani, T. Senoh, M. Okui, K. Yamamoto, N. Inoue, and T. Fujii, "[m31261][FTV AHG] Multiple aspects," in *International Organisation For Standardisation*. ISO/IEC JTC1/SC29/WG11, October 2013.

[4] M. P. Tehrani, T. Senoh, M. Okui, K. Yamamoto, N. Inoue, and T. Fujii, "[m31103][FTV AHG] Introduction of super multiview video systems for requirement discussion," in *International Organisation For Standardisation*. ISO/IEC JTC1/SC29/WG11, October 2013.

[5] J.-R. Ohm, "Overview of 3D video coding standardization," in *International Conference on 3D Systems and Applications*, Osaka, Japan, June 2013.

[6] L. Zhang, Y. Chen, and M.Karczewicz, "3d-ce5.h related: Disparity vector derivation for multiview video and 3dv," in *International Organisation For Standardisation*. ISO/IEC JTC1/SC29/WG11 MPEG2012/m24937, July 2012.

[7] G. Tech, K.Wegner, Y. Chen, and S. Yea, "3D-HEVC test model 2," in *International Organisation For Standardisation*. ITU-T SG 16 WP 3 ISO/IEC JTC1/SC29/WG11 JCT3V-B1005, October 2012.

[8] P. Helle, S. Oudin, B. Bross, D. Marpe, M. Bici, K. Ugur, J. Jung, G. Clare, and T. Wiegand, "Block merging for quadtree-based partitioning in HEVC," in *IEEE Transactions on Circuits and Systems for Video Technology*, December 2012, vol. 22, pp. 1720–1731.

[9] S. Shi, P. Gioia, and G. Madec, "Efficient compression method for integral images using multi-view video coding," in *IEEE 18th International Conference on Image Processing*, Brussels, Belgium, September 2011.

[10] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," in *IEEE Transactions on Circuits and Systems for Video Technology*, 2007, vol. 17, pp. 1461–1473.

[11] T. Chung, K. Song, and C.-S. Kim, "Compression of 2-d wide multi-view video sequences using view interpolation," in *IEEE 15th International Conference on Image Processing*, San Diego, CA, USA, October 2008.

[12] T. Chung, I.-L. Jung, K. Song, and C.-S. Kim, "Virtual view interpolation and prediction structure for full parallax multi-view video," in *Advances in Multimedia Information Processing - PCM*, 2009, vol. 5879, pp. 543–550.

[13] T. Chung, I.-L. Jung, K. Song, and C.-S. Kim, "Multiview video coding with view interpolation prediction for 2d camera arrays," in *Journal of Visual Communication and Image Representation*, July August 2010, vol. 21, pp. 474–486.

[14] A. Avci, J. De Cock, P. Lambert, R. Beernaert, J. De Smeta, L. Bogaert, Y. Meuret, H. Thienpont, and H. De Smeta, "Efficient disparity vector prediction schemes with modified p frame for 2d camera arrays," in *Journal of Visual Communication and Image Representation*, February 2012, vol. 23, pp. 287–292.

[15] D. Rusanovsky, K. Muller, and A. Vetro, "Common test conditions of 3dv core experiments," in *International Organisation For Standardisation*. ITU-T SG 16 WP 3 ISO/IEC JTC1/SC29/WG11 JCT3V-B11000, October 2012.

[16] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," in *VCEG Meeting*, Austin, USA, April 2001.