FAST MULTI-VIEW VIDEO PLUS DEPTH CODING WITH HIERARCHICAL BI-PREDICTION

Brian W. Micallef [1], Carl J. Debono[2] and Reuben A. Farrugia[3]

Department of Communications and Computer Engineering, University of Malta, Msida, Malta
{[1] brian.micallef, [2] c.debono}@ieee.org, [3] reuben.farrugia@um.edu.mt

ABSTRACT

The Multi-view Video Coding (MVC) standard was developed for efficient encoding of multi-view videos. Part of it requires the calculation of both disparity and motion estimations using a bi-prediction structure. These estimations involve an exhaustive search for the optimal compensation vectors from multiple forward and backward reference frames which, while being very efficient in terms of compression, results in high computational costs. This paper proposes a solution that utilizes the multi-view geometry along with the available depth data, to calculate more accurate predictors for both motion and disparity estimations, and for both directions of the prediction structure. Simulation results demonstrate that this technique is reliable enough to allow a substantial reduction in the search areas in all the reference frames. This in turn results in a significant speed-up gain of 3.2 times with a negligible influence on the coding efficiency, while encoding both the color and the depth MVVs.

Index Terms— 3DTV, fast estimations, Hierarchical Bi-Prediction, Multi-view Video plus Depth coding, MVC.

1. INTRODUCTION

Multi-View Videos (MVVs) can provide the users with a sense of complete scene perception by transmitting several viewpoint videos to the receivers simultaneously. This provides information about the scene structure such that appropriate displays can provide the 3D perception with an arbitrary viewing angle selection, to all the end users. With the advances in 3D Television (3DTV) and Free-Viewpoint Television (FTV) technologies, MVV transmission is attracting more and more attention. However, for efficient enabling of both technologies, MVVs with their associated depth data are required at the receiver [1, 2]; thus, they form the new Multi-view Video plus Depth (MVD) format as a 3D video [3]. These are suitable because they allow view synthesis/rendering [1-4] for arbitrary viewpoint generation.

Efficient MVV Coding (MVC) techniques that exploit the spatial, temporal and inter-view redundancies are required for transmission. The first two redundancies are removed using the normal Intra and Motion Estimation (ME) techniques of the H.264/AVC. However, to remove inter-view redundancies, Disparity Estimation (DE) is required. For efficient MVC, disparity estimation with an I-B-P bi-prediction structure and ME with a Hierarchical Bi-Prediction (HBP) structure within a Group of Pictures (GOP) [5] are required. This imposes a high computational burden on the MVC encoder. Yet, this technique was found efficient to compress both the color [5] and depth MVVs [6], so faster MVC techniques are desirable for MVD coding.

This paper presents an efficient MVC technique that exploits the full geometrical scene information available from the depth data in the MVD, together with the multi-view geometry, to calculate more accurate Motion Vector (MV) and Disparity Vector (DV) predictors for Hierarchical Bi-Prediction MVC. These can respectively indicate optimal search areas for ME and DE in both the ForWard (FW) and BackWard (BW) reference frames within the structure. The performance of this technique is investigated for MVC of both color and depth MVVs. Results show that the proposed MVC reduces encoding duration by about 3.2 times for both MVV types, with an insignificant loss in coding efficiency.

The paper is structured as follow: Section 2 presents the estimation techniques and the HBP structure used for efficient MVC. Section 3 proposes a method that exploits the available depth data to obtain faster estimations in both directions of the HBP MVC structure. Section 4 provides a simulation overview, while section 5 gives the simulation results obtained. Finally section 6 concludes this paper.

2. MULTI-VIEW VIDEO CODING WITH HIERARCHICAL BI-PREDICTION

The MVC scheme utilizes the built-in block-based ME of the H.264/AVC [7] standard to remove the temporal redundancies inside a single viewpoint video. This is the most computational intensive component of the encoder. MVC extends this technique to perform also DE between the viewpoint frames to suppress the inter-view redundancy. Consequently, ME and DE become the most computational intensive parts of the system. Furthermore, the most efficient ME in the MVC scheme [5] was found to use the HBP structure which further increases the computational requirement. This is also called the HHI MVC scheme and is illustrated in Fig. 1. Therefore, to reduce this massive computational cost and for its practical application, efficient ME and DE are required.

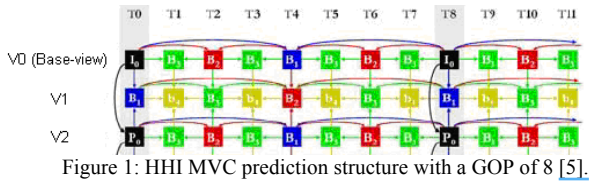

Figure 1: HHI MVC prediction structure with a GOP of 8 [5].

The Full Search Estimation (FSE) in MVC exhaustively searches for an optimal displacement vector from all the search points within a search area. Meanwhile, it estimates each Macro-Block (MB) from both temporal and viewpoint reference frames to form appropriate MVs and DVs, respectively. Only the winning vectors together with the optimal mode which minimize the Lagrangian Rate-Distortion Optimization (RDO) [8] function are transmitted. The central position of this search area is indicated by the predictor that is the median of the neighborhood vectors. This exhaustive search obtains the optimal RDO vectors but it is the most computational intensive way to achieve them. Thus, FAst Search Estimation (FASE) such as the Diamond Search [9] has been proposed in H.264/AVC to reduce the computations and these can still be applicable for MVC. These sub-optimal techniques reduce the number of search points while maintaining almost optimal R-D performances.

To obtain fast encoding, the real-time MVC structure is generally used since it allows only one temporal reference frame. Our work in [10]-[12] demonstrated that, for real-time MVC, the multi-view geometry can be used with the available depth data to increase the speed of the estimations by obtaining accurate predictors that allow a reduced search area. However, for optimal MVC efficiency, the HBP MVC is required [5], which allows multiple hierarchical levels of reference frames from both the FW and BW directions, to obtain optimal estimations. This increase in reference frames drastically increases further the number of search points and its associated computations. Generally, this structure is used for broadcasting applications where fast encoding is highly desirable but efficiency is more important. Thus, efficient techniques that identify the optimal MV and DV predictors and respectively reduce the ME and DE search areas in both directions of the HBP structure need also investigation, to speed-up efficient MVD coding with HBP MVC.

3. PROPOSED FAST MULTI-VIEW VIDEO CODING USING HIERARCHICAL B FRAMES

Generally, the optimal DVs are found in the region that represents the same area in the viewpoint reference frames. This region can be easily calculated using the multi-view geometry together with the depth [10]. These DVs lie also around the epipolar line, [13] which is identified by the epipolar geometry. Thus, as shown in our previous work [11], searching around these two areas can drastically reduce the DE search areas in both of the FW and BW directions of the I-B-P inter-view prediction structure. These techniques are still applied to speed-up DE within MVC.

However, the complexity of HBP MVC results also from the fact that for each MB, ME should also be performed repetitively for each reference frame. Our work in [12] demonstrated that for real-time MVC, the encoded MVs between successive frames can be obtained from the corresponding area in the Base view and can be projected in the temporal reference frame of the current frame in a non-Base view. This is used to obtain a better MV predictor that allows a reduced search area for faster ME.

However, with the HBP structure, a MB is compensated from multiple FW and BW temporal reference frames from an upper hierarchal level within the GOP that are not necessarily sequential. Moreover, during it's encoding, only the optimal MVs of the optimal mode, from the optimal temporal frame, which minimize the global RDO cost function are encoded. Thus, not all the required MVs, for each partition of the modes and for each reference frame are easily accessible and available for projection. Nonetheless, these optimal MVs were identified by the ME process to be tested for RDO. Therefore, during Base view coding, these MVs are available and should be temporary stored successively according to the order of their reference frame in this level of the hierarchical structure, to ease their accessibility. To provide a good MV resolution, the MVs of the 8×8 mode's ME are selected for storage, and then they are averaged for the other larger partitions. These MVs are initially obtained from the Base-view, since it has the proper MV calculated by ME with a full search area. However, later, these can be obtained from any other encoded view since appropriate MV predictors were estimated for its ME.

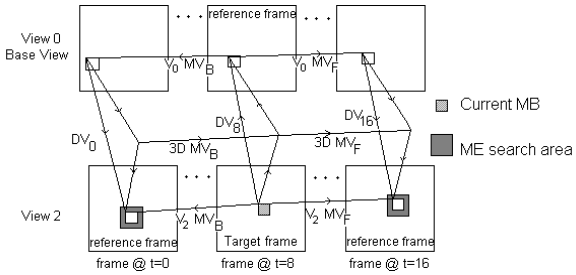

Figure 2: MV prediction for both directions of HBP MVC with GOP of 16.

Therefore, taking Fig. 2 as reference, for the current sub-MB being estimated in a non-Base view, the multi-view geometry together with the depth data is used to identify the DV predictor that points to the corresponding sub-MB area in the Base view reference frame (DV_8) [10]. The stored 8×8 MVs which lie within this partition's area are averaged for each of its potential temporal reference frame (V_0 MV_B and V_0 MV_F), and are used to locate its equivalent compensation sub-MBs. These identified locations are re-projected in 3D space and located in the temporal reference frames of the current view (DV_0 and DV_{16}), to identify good potential sub-MB replacements for the current sub-MB. These positions are used to obtain more reliable MV predictors for the current sub-MB, for each temporal reference frame (V_2 MV_B and V_2 MV_F). These allow a

reduced search area in all the temporal reference frames in both of the prediction's directions and this is equivalent to obtaining appropriate 3DMVs for each MV. Finally, the obtained optimal MVs are still transmitted as residual vectors from the original median MV, to obtain H.264/MVC compatible bit-streams.

A new ME that has its search range adaptive to motion, can also be applied for MVC. This is because a good indication of the potential temporal sub-MB's motion can be obtained from the Base view such that its ME search area can be adapted accordingly. Therefore, if the magnitude of the MVs in the corresponding area in the Base view is encoded with a magnitude smaller than ±2 pixel elements (pels) (including also the MVs of the SKIP mode), a very small ME search area of ±2 pels can be allowed. The majority of the picture is usually static or with slow motion. Therefore, they are generally compensated with a small MV or the SKIP mode, and thus allowing a further small search area will further reduce the ME's computations. Then, if the average of the MVs is larger than ±2 pels, this indicates a more dynamic area, so the reduced search area around the new MV predictor described above, must be used.

4. SIMULATION OVERVIEW

The proposed algorithms were implemented within the Joint Multi-view Video Coding model (JMVC ver 6.0) [14] to calculate the resulting computational reduction. Two calibrated MVD sequences known as the *Breakdancers* and the *Ballet* sequences, were tested. Their color MVV (YUV 4:2:0, 1024×768, 15Hz) was captured by eight cameras arranged on an arc with precise camera parameters [15]. Their per-pixel depth MVV was estimated using the method described in [4]. The first three views of both the color and depth MVVs were encoded. Since optimal efficiency is required, the HBP structure and the CABAC as the main entropy encoder were selected. Both exhaustive FSE and diamond FASE [9] were used to determine the optimal compensation MVs and DVs for color and depth type MVC. The MVC configuration parameters [16] are given in Table 1.

TABLE I. MVC SIMULATION PARAMETERS

Multi-view HIGH Profile
Temporal HBP structure with a GOP of 16
I – B – P prediction between view-points
Max number of 4 FW and 4 BW reference frames
CABAC as main entropy encoding
Original search area of ± 64 pels
Proposed search area of ± 10 pels
Estimation resolution of ¼ pel
Fixed Quantization Parameters (QPs) of 24, 28, 32, and 36

All the simulations were carried out on a PC with an Intel® Core™ i7 CPU @ 3.2 GHz, with 6GB of RAM and running Microsoft Windows® 7 Ultimate x64. The reduction in computational cost was measured as a speed-up gain obtained in the time required to encode the whole MVV using the encoder with the proposed methods, with respect to the original encoder. Finally, the original MVC decoder was used to decode the formed bit-streams and objective evaluation was performed.

5. RESULTS AND ANALYSIS

Table II and Table III give the simulation results obtained after encoding and decoding the first three views of the color and the depth MVVs from the *Ballet* MVD, respectively. These include both the performance of the FSE and the FASE methods to determine the optimal DVs. The comparison of the MVC performances is presented as the distortion in the Luminance Peak-Signal-to-Noise-Ratio (PSNR), the percentage increase in the total MVV bit-rate and the speed-up gain obtained in their encoding time, with respect to the original encoder with FSE; since this provides the optimal quality with the highest encoding computational cost. Then, Fig. 3 and Fig. 4 illustrate the R-D performance curves obtained for the color MVC and the depth MVC of the *Breakdancers* MVD sequence, respectively.

TABLE II. R-D VALUES FOR THE COLOR MVC OF THE *BALLET* MVD

QP	FSE	Change	Prop. FSE	FASE	Prop. FASE
24	42.17 dB	Δ PSNR (dB)	-0.005	-0.008	-0.006
	1697.97 kbps	Δ Bit-rate (%)	+1.98	+0.51	+1.20
	432.45 hrs	Gain in Speed	+3.28	+33.57	+100.72
28	41.16 dB	Δ PSNR (dB)	-0.011	-0.011	-0.020
	934.11 kbps	Δ Bit-rate (%)	+1.76	+0.76	+1.29
	430.81 hrs	Gain in Speed	+3.29	+36.08	+105.19
32	39.69 dB	Δ PSNR (dB)	-0.047	-0.014	-0.064
	589.57 kbps	Δ Bit-rate (%)	+2.33	+0.32	+0.41
	430.13 hrs	Gain in Speed	+3.29	+39.13	+112.59
36	37.89 dB	Δ PSNR (dB)	-0.100	-0.032	-0.131
	386.22 kbps	Δ Bit-rate (%)	+2.13	-0.20	-0.22
	425.80 hrs	Gain in Speed	+3.31	+41.99	+115.70

TABLE III. R-D VALUES FOR THE DEPTH MVC FROM THE *BALLET* MVD

QP	FSE	Change	Prop. FSE	FASE	Prop. FASE
24	48.77 dB	Δ PSNR (dB)	-0.058	-0.078	-0.217
	2222.69 kbps	Δ Bit-rate (%)	+2.22	+0.094	+3.31
	393.77 hrs	Gain in Speed	+3.27	+23.18	+65.16
28	46.13 dB	Δ PSNR (dB)	-0.033	-0.094	-0.218
	1502.77 kbps	Δ Bit-rate (%)	+2.41	+1.44	+4.11
	393.89 hrs	Gain in Speed	+3.29	+26.26	+72.70
32	43.21 dB	Δ PSNR (dB)	-0.041	-0.165	-0.203
	995.94 kbps	Δ Bit-rate (%)	+2.52	+1.85	+3.73
	394.42 hrs	Gain in Speed	+3.27	+30.31	+83.26
36	40.33 dB	Δ PSNR (dB)	-0.059	-0.141	-0.176
	626.62 kbps	Δ Bit-rate (%)	+2.27	+2.00	+2.59
	392.99 hrs	Gain in Speed	+3.29	+34.66	+91.77

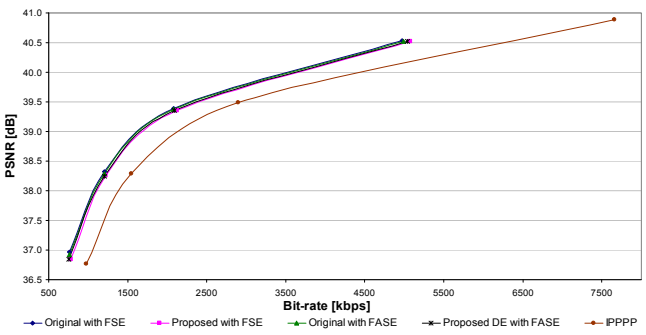

Figure 3: R-D curves for the color MVC of the *Breakdancers* MVD.

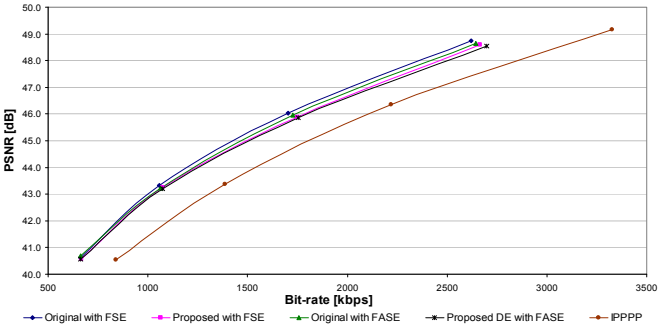

Figure 4: R-D curves for the depth MVC of the *Breakdancers* MVD.

Comparing the results obtained shows that the proposed fast MVC technique provides a significant decrease in coding time of about 3.2 for FSE and 2.7 for FASE, when averaged over all the encoded MVVs. Moreover, the proposed technique registered only a small average loss in the original quality of about 0.09 dB in Bjøntegaard Delta (BD)-PSNR [17] for the color MVVs and 0.2 dB in BD-PSNR for the depth MVVs. These represented the speed-up gains obtained for the whole MVC process. However, an individual speed-up gain of about 72.8 for FSE and 6.4 for FASE were obtained while encoding an inter-view predicted viewpoint utilizing the proposed techniques for fast encoding. Thus, the limitation of the overall speed-up gains lies when encoding the base view since the original encoder has to be used, significantly taking up the majority of the overall MVC time. Thus overall speed-up gains are expected to increase as more inter-view predicted views are encoded within the MVC system.

In Fig. 3 and Fig. 4, we illustrate the results obtained by the HBP structure with the CABAC as entropy encoder, with respect to the IPPP structure with the CAVLC as entropy encoder [13], for the *Breakdancers* MVD coding. These results demonstrate that the former is more efficient to encode both MVV types, giving an average gain of 22% in BD-bit-rate saving for the color MVC and 21% in BD-bit-rate saving for the depth MVC, and that this is vital for efficient encoding. Thus, they also demonstrate that the original coding efficiency of the HBP structure over the real-time structure can be almost preserved while still obtaining its fast encoding.

6. CONCLUSION

In this paper, we presented an efficient MVC technique that obtains faster motion and disparity estimations within the Hierarchal Bi-Prediction structure of MVC. This technique is so efficient to obtain the optimal search areas in both of the prediction's directions of the structure, that the estimations' search points can be reduced while encoding the inter-view predicted viewpoints. This highly speeds up the MVD coding while still maintaining the same high coding efficiency of this prediction structure. Simulation results demonstrated that a speed-up gain of up-to 3.2 was registered while encoding the first three-views of the MVVs under test, and that this gain is expected to increase as more inter-view predicted views are MVV encoded. The results also showed that there is little degradation in the Rate-Distortion performances of the original system.

7. ACKNOWLEDGEMENT

This research work is partially funded by STEPS-Malta and partially by the EU–ESF 1.25. We would like to thank the Interactive Media Group of Microsoft® Research (MSR) for providing the *Breakdancers* and *Ballet* Multi-View plus Depth sequences.

8. REFERENCES

[1] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth Map Creation and Image Based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability," *Signal Processing: Image Comm. Special Issue on 3DTV*, vol. 22, no. 2, pp. 217-234, February 2007.

[2] Y. Mori, N. Fukushima, T. Fujii, and M. Tanimoto, "View Generation with 3D Warping using Depth Information for FTV," *in Proc of 3DTV-CON 2008*, pp. 229-232, May 2008.

[3] ISO/IEC MPEG & ITU-T VCEG, "Multi-view Video Plus Depth (MVD) Format for Advanced 3D Video Systems," *Doc. JVT-W100*, April 2007.

[4] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winderm, and R. Szeliski, "High-Quality Video View Interpolation using a Layered Representation," *ACM SIGGRAPH and ACM trans. on Graphics*, pp. 600-608, August 2004.

[5] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient Prediction Structures for Multi-view Video Coding," *IEEE Trans. CSVT*, vol. 17, no. 11, pp. 1461–1473, November 2007.

[6] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Multi-view Video Plus Depth Representation and Coding," in *Proc. of ICIP 2007*, pp-201-204, September 2007.

[7] ISO/IEC IS 14496-10, Advanced Video Coding for Generic Audiovisual Services, ITU-T Rec. H.264, March 2009.

[8] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G.J. Sullivan, "Rate-Constrained Coder Control and Comparison of Video Coding Standards," *IEEE Trans. CSVT*, vol. 13, pp. 688-703, July 2003.

[9] S. Zhu, and K.–K. Ma, "A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation," *IEEE Trans. on Image Process.*, vol. 9, no. 2, pp. 387-392, February 2000.

[10] B.W. Micallef, C.J. Debono, and R.A. Farrugia, "Exploiting Depth Information for Fast Multi-view Video Coding," in *Proc. of PCS 2010*, pp. 38-41, December 2010.

[11] B.W. Micallef, C.J. Debono, and R.A. Farrugia, "Fast Disparity Estimation for Multi-view plus Depth Video Coding," in *Proc. of VCIP 2011,* November 2011.

[12] B.W. Micallef, C.J. Debono, and R.A. Farrugia, "Exploiting Depth Information for Fast Motion and Disparity Estimation in Multi-view Video Coding," in *Proc. of 3DTV-CON 2011,* May 2011.

[13] J. Lu, H. Cai, J.-G. Lou, and J. Li, "An Epipolar Geometry-Based Fast Disparity Estimation Algorithm for Multiview Image and Video Coding," *IEEE Trans. CSVT*, vol. 17, no. 6, pp. 737-750, June 2007.

[14] ISO/IEC MPEG & ITU-T VCEG, "Joint Multi-view Video Coding Model (JMVC 6.0)," JVT-AE207, September 2009.

[15] MSR Multi-view Sequences [Online]. Available: http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload.

[16] ISO/IEC MPEG, and ITU-T VCEG, "Common Test Conditions for Multiview Video Coding," Doc. JVT-U211, October 2006.

[17] G. Bjøntegaard, "Calculation of Average PSNR Differences Between RD-Curves," Doc. VCEG-M33, April 2001.

View publication stats