

# HEVC based Multi-View Video Codec using Frame Interleaving technique

Bruhanth Mallik

School of Computing, Creative Technology and  
Engineering  
Leeds Beckett University, UK  
b.mallik6347@student.leedsbeckett.ac.uk

Akbar Sheikh Akbari

School of Computing, Creative Technology and  
Engineering  
Leeds Beckett University, UK  
a.sheikh-akbari@leedsbeckett.ac.uk

**Abstract**—this paper presents a HEVC based multi-view video codec. The frames of the multi-view videos are interleaved to generate a monoscopic video sequence. The interleaving is conducted in a way to increase the exploitation of the temporal and inter-views correlations. The MV-HEVC standard codec is configured to work as a single layered codec, which functions as a monoscopic HEVC codec with AVC capabilities, and used to encode interleaved multi-view video frames. The performance of the codec is compared with the anchor standard MV-HEVC codec by coding the three standard multi-view video sequences: “Balloon”, “Kendo” and “Newspaper1”. Experimental results show the proposed codec out performs the anchor standard MV-HEVC codec in term of bitrate and PSNR.

**Keywords**—HEVC; MV-HEVC; Multi-view video coding; video compression; texture coding.

## I. INTRODUCTION

Recent developments in display technology, storage, computing, wireless mobile communication and emergence of high speed broadband internet services have contributed to the current advancement in 3D video coding technology and its standardization [1]. Over the last decade digital video services have moved from Standard Definition (SD) to High Definition (HD) and Ultra-High Definition (UHD) resolution videos. HD videos have become the new quality standard for media services. It is now considered that upgradation in the coming decade will be in 3D video production and delivery. The current popular 3D video contents are stereoscopic, multi-view, video plus depth data and software rendered computer aided 3D videos. 3D video applications are not limited to entertainment market, they are employed in automation, robotics, machine-vision, e-Learning, autonomous navigation and surveillance systems.

The multi-view version of 3D video uses multiple camera views to capture the same scene simultaneously by geometrically aligned and synchronized cameras. Multi-view video coding can be classified into texture based, texture plus depth based and model based categories, according to their scene geometry usage to represent depth information in the encoded videos [1]. The texture based approaches employ scene geometry implicitly through disparity prediction and compensation across views. Though the texture plus depth coding schemes are more popular than the texture based codecs, they require much higher computational power for

their coding scheme. Texture plus depth based approaches use depth information, which interpret 3D scene by assigning same depth value to pixels located in the same distance from the camera, to render virtual views. This increases the complexity of the decoder and the rendered views may also suffer from artefacts due to inaccuracy of the depth estimation and view synthesis algorithms. The model-based approaches define the scene geometry explicitly by using 3D mesh models of the scene besides the texture mapping [2]. Model-based approaches facilitate view random access for the scene, but they are too complex and can only be used for a scene with one or very limited objects. The research objective in this paper is driven towards developing a reduced complexity texture based multi-view video codec by analyzing the coding process and prediction structures of multi-view video codecs.

Multi-view videos contain huge amount of visual information and their transmission over a bandwidth limited communication channel needs high level of efficient compression. The bitrate of the encoded multi-view videos increase approximately linearly to the number of views used to create the 3D content, hence efficient compression techniques are necessary in realizing such application [3]. To efficiently compress multi-view videos, the video codecs extensively use inter-view prediction and multi-frame referencing tools, which are designed to explore and exploit the inter-view and temporal correlations simultaneously [4].

The state of the art High Efficiency Video Coding (HEVC), developed by Joint Collaborative Team on Video Coding (JCT-VC) experts from the ITU-T Visual Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) was approved as an ITU-T/ISO standard in early 2013 [5]. HEVC standard and its extension to 3D videos, with the ability to deliver up to twice the bitrate savings compared to previously popular H.265/AVC, has opened new avenues for innovation and research in this field. Currently the research focus is driven towards HEVC’s extension to efficiently compress different class of 3D videos.

The standardized multi-view video extension of HEVC referred to as MV-HEVC, was finalized in late 2013. In order to extend HEVC to encode multi-view videos, HEVC’s high level syntax has been revised to handle signaling between views for multi-view video data and prediction structure is enabled to have a flexible multi-frame reference picture management capability. In addition to modifications to HEVC

standard for coding multi-view videos, block-level modules have been improved to exploit the correlation in motion, texture characteristics and residual data between scene objects projected to different viewpoints. MV-HEVC is provided with disparity compensation for inter-view prediction which uniquely derives disparity motion vectors from neighboring blocks without additional bits being used for signaling [7].

Despite MV-HEVC’s toolsets to reduce the overhead of signaling motion information, the layered architecture followed by the codec for representing dependent views of multi-view video diminishes the compression efficiency of the encoder. An inter-view motion vector prediction method was proposed in [8] to improve coding efficiency of the dependent views by using previously encoded motion information of the reference view using temporal motion vector prediction. This method calculates a global disparity value by accessing the look up table used for disparity conversion from previously encoded frames. Then global disparity value is used to modify the motion field of inter-view reference pictures. A multi-view video compression scheme using HEVC’s single view coding tools was proposed in [9] in which the prediction structure closely matched H.264/AVC’s multi-view video coding with minimized prediction signaling. A less complex but improved prediction for motion data in inter-views of multi-view videos coded by HEVC codec was proposed in [10], which uses vector scaling for target prediction units, additional decision choices for deriving prediction candidate from co-located units in the reference frame and to track the neighboring unit vector to identify unit vector that can be used as a source of prediction (nested prediction). Though the method proposed in [10] was a complex way of improving prediction in multi-view videos, the modifications did not produce significant coding gain than the standard MV-HEVC.

Another less complex means of delivering 3D video content is by using frame-compatible formats, which packs the frames from different views into single video, through temporal multiplexing. Frame-compatible formats have received considerable amount of attention from the broadcasting industry due to feasibility of encoding and transmitting 3D contents through existing infrastructure [11]. Setbacks such as the lack of signaling for a frame-compatible format and deficient benefit from inter-view redundancies after frame interleaving through temporal multiplexing, have limited their application to stereo videos [12]. In a study conducted to develop a less complex HEVC based stereo video codec led to using frame interleaving scheme which encodes the reordered stereo video frames using a modified MV-HEVC [13]. The results for this stereo video codec showed significant coding gains compared to the standard MV-HEVC’s stereo video coding performance.

From analyzing the existing methods for coding multi-view videos with reduced complexity and improved coding performance, it is found that only modifications to reduce the codec’s complexity would not deliver significant compression efficiency, on the other hand frame interleaving method has the potential to deliver superior results for stereo videos. Therefore in this paper a novel HEVC based multi-view video Codec using Frame Interleaving technique is proposed. The coding

performance of the proposed codec is compared with the performance of standard MV-HEVC for a three views scenario.

The rest of this paper is organized as follows: Section 2 outlines the framework of the proposed multi-view video coding technique by introducing the frame interleaving algorithm and the codec’s design parameter for encoding the interleaved videos. Section 3 presents the experimental results of the proposed multi-view video coding scheme and finally, the paper is concluded in section 4.

## II. FRAMEWORK OF THE PROPOSED CODEC

The proposed HEVC based multi-view video coding scheme interleaves the frames of the multi-view videos, generating a single stream video sequence, as shown in Fig. 1. In this figure the dotted arrows represent the frame reordering and interleaving route traversed by the multi-view frame interleaving algorithm. The interleaving algorithm developed for the multi-view videos is designed to start and complete the frame reordering at the center view. Hence the minimum size of the Group of Pictures (GoP) in a 3-view multi-view video scenario would be of 12 frames. This ensures the I-frame always to be a center view frame. The frame interleaving structure has been designed to maximize the coding efficiency of the codec by exploiting spatial, temporal and inter-views correlations.

The HEVC standard’s extension to multi-view video coding, which is known as the MV-HEVC, has a framework similar to MVC extension of H.264/ MPEG-4 AVC. The referencing structure in MV-HEVC is supplemented with signaling for prediction dependencies between different views for multi-view video coding. The MV-HEVC uses a multi-loop encoding design to encode frames from other views. However, MV-HEVC also uses a layered representation for encoding multi-view videos, which increases the decoding complexity due to the much needed multi-layer decoding for prediction prior to encoding a new layer. Hence, a single layer encoding approach with reduced complexity and bitstream overhead of the standard MV-HEVC codec, which makes a monoscopic HEVC, called HEVC and Frame Interleaving based Multi-view Video (HEVC-FIMV) codec, is designed to code the resulting inter-leaved multi-view video frames. Fig. 2 outlines the reference frame structure strategized for the HEVC-FIMV codec.

HTM-16.0-MV- Draft 5 software version [14] was used to implement the proposed HEVC-FIMV codec. The modifications to the standard MV-HEVC video codec’s software were intended to provide a single layered HEVC codec operating with AVC capabilities and reduced transmission overhead.

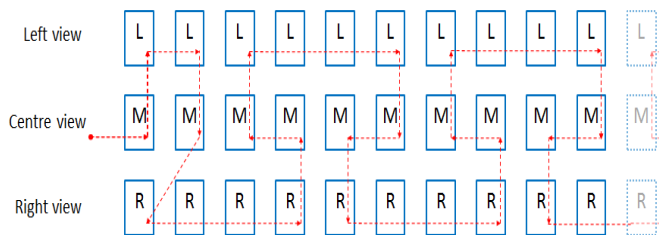


Fig. 1. Multi-view video frames interleaving block diagram.

The values assigned to the parameters of the standard MV-HEVC codec in order to implement the proposed codec are tabulated in Table I. The parameter “Number of Layer” is set to the value 1 to configure single layer mode of operation of the standard MV-HEVC codec. The parameters “Number of ViewId”, “OutputLayerSetIdx” and “LayerIdsInAddOutputLayerSet\_0” were assigned minimum values to run modified codec with least number of bits in the transmission overhead. The intra period is set to 24 frames as restricted by the specifications as per the common test condition document JCT3V-G1100 [14], but the minimum GoP size for the proposed codec’s design for a 3-view multi-view video scenario has to be set to 12 frames.

Multi-view 3D video contents are acquired through geometrically aligned parallel axised or convergent camera setup. A study on the impact of camera separation on performance of video codecs has shown that, as the angle between the optical reference lines of the cameras increases the inter-view motion correlation decreases [15]. The proposed HEVC-FIMV codec’s design is based on the standard MV-HEVC codec, which uses Advanced Motion Vector Prediction (AMVP) and Temporal Motion Vector Prediction (TMVP). Therefore, for AMVP and TMVP to work efficiently for a frame interleaved multi-view monoscopic video, the motion vectors search region for the proposed codec needs to be extended from common test condition’s 64 to 96.

### III. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed HEVC-FIMV codec, views 1-3-5, 1-3-5 and 2-4-6 of “Balloon”, “Kendo” and “Newspaper1” standard multi-view video sequences respectively, were selected and coded using the proposed HEVC-FIMV codec. These video sequences cover both static and dynamic backgrounds at different levels of illuminations. The coding performance of the proposed codec was then compared with the anchor MV-HEVC codec, as presented in JCT3V-G1100 document [14].

The results of PSNR and consumed bitrate for the standard multi-view videos test sequences “Balloons”, “Kendo” and “Newspaper1” in comparison to the anchor MV-HEVC codec’s performance, at Quantization Parameters (QP) of 25, 30, 35 and 40, are tabulated in Table II – IV, respectively. From Table II, it can be seen that the proposed codec gives

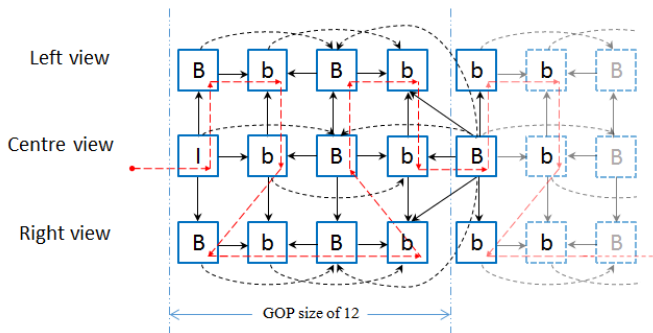


Fig. 2. Reference frame structure of the proposed HEVC-FIMV codec.

TABLE I. MV-HEVC DESIGN PARAMETERS FOR THE PROPOSED CODEC

Parameters	Value
NumberOfLayers	1
ViewId	1
VpsNumLayerSets	1
OutputLayerSetIdx	0
LayerIdsInAddOutputLayerSet_0	0
GOP Size	12
Intra Period	24
QP	25, 30, 35, 40

superior coding performance to that of the anchor codec for coding “Balloons” sequence, in terms of the required bitrates and the average PSNR of the decoded frames, about 6.27% of bitrates and up to 0.73 dBs for Y-colour component, with the best performance at QP 30. From Table III, which shows results for “Kendo” test sequence, it is clear that the proposed codec’s video sequences requires 16.65% less bitrates than the anchor MV-HEVC codec’s bitrate, while it offers up to 0.88 dBs higher visual quality in term of PSNR.

From Table IV, which tabulated the experimental results for “Newspaper1” standard multi-view test sequences, it can be found that the proposed codec has an average bitrate reduction of 9.04% in comparison to that of anchor codec in addition to the higher average luminance quality metric Y-PSNR of up to 0.79dB than the anchor MV-HEVC. Although the chrominance quality metrics U-PSNR and V-PSNR, of the proposed codec for the tested standard experimental multi-view videos is marginally lower (about 0.45dB) and the average Y-PSNR is 0.8dB greater than that of anchor codec, the proposed codec requires about 5% less bandwidth to transmit the videos. This implies that the proposed HEVC-FIMV codec has an improved compression/coding efficiency in comparison to that of MV-HEVC.

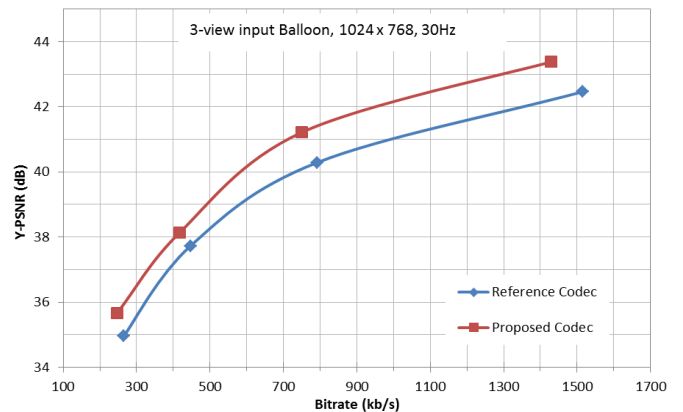


Fig. 3. PSNR vs bitrate for MV-HEVC and the proposed codec for coding “Balloons”.

To help better understand the coding performance of the proposed codec the Y-PSNR results of the proposed HEVC-FIMV codec and the anchor standard MV-HEVC codec for coding “Balloons”, “Kendo” and “Newspaper1” standard multi-view test sequences plotted with respect to bitrate are shown in Fig. 3-5, respectively. From the graph in Fig. 3, it can be seen that the proposed codec outperforms the standard MV-HEVC at all bitrates for coding the “Balloons” multi-view video sequences. At lower bitrates the proposed codec has 0.8dB gain and at higher bitrates the proposed codec continues to maintain the gain at 1.2dB over the standard MV-HEVC codec. From Fig. 4, it is clear that the proposed gives higher coding performance to that of anchor codec for coding “Kendo” multi-view sequences. It is something between 1.4dB to 1.8dB gain. Similarly in Fig. 5, a constant performance gain of about 1.1dB can be seen, by proposed codec’s experimental results for the standard multi-view test sequence “Newspaper1” over standard MV-HEVC codec. From these figures, it is clear that the proposed codec generates significantly higher coding performance to that of the standard MV-HEVC at all bitrates (up to 1.25 dBs).

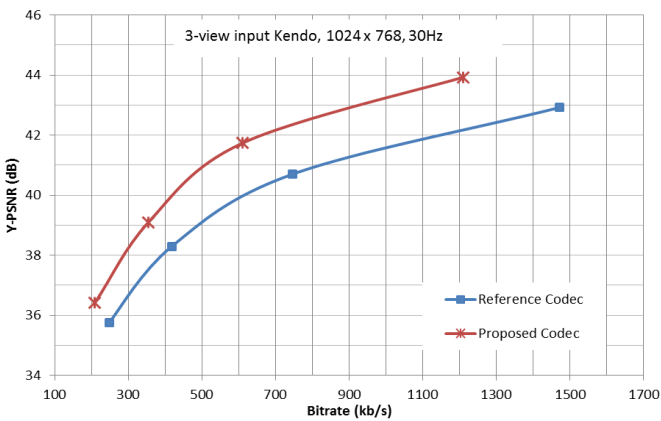


Fig. 4. PSNR vs bitrate for MV-HEVC and the proposed codec for coding “Kendo”.

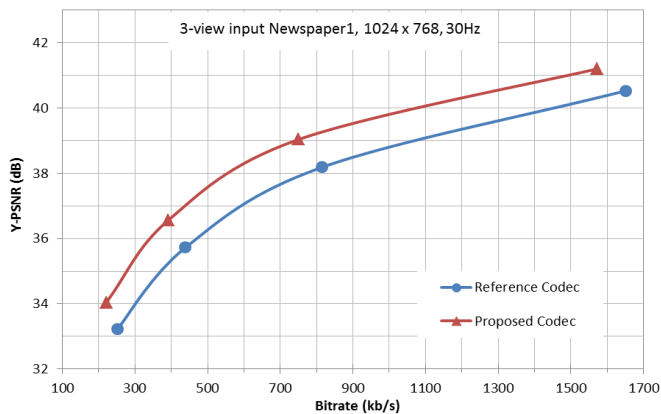


Fig. 5. PSNR vs bitrate for MV-HEVC and the proposed codec for coding “Newspaper1”.

#### IV. CONCLUSIONS

In this paper a HEVC based Frame Interleaved video coding technique, for multi-view videos (HEVC-FIMV) that uses a reduced layer approach to encode frame interleaved multi-view videos is proposed. The coding performance of the proposed codec was compared with the standard MV-HEVC video codec using three standard multi-view video sequences. Experimental results show that substantial amount of bitrate savings can be achieved through the proposed coding scheme compared to the standard MV-HEVC codec. Further, the proposed HEVC-FIMV codec delivers superior video quality in comparison to the standard MV-HEVC codec at different QPs and bitrates.

#### REFERENCES

- [1] A. Smolic, "3D video and free viewpoint video: From capture to display," *Pattern recognition*, vol. 44, no. 9, pp. 1958-1968, 2011.
- [2] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 519-528, 2006.
- [3] P. Merkle, M. Müller, T. Wiegand, "3D Video: Acquisition, Coding, and Display," *IEEE Transactions on Consumer Electronics*, 56 (2), 946 -950.
- [4] A. Vetro, T. Wiegand, G.J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard," *Proceedings of the IEEE*, 99(4), 626-642.
- [5] "High efficiency video coding," *ITU-T Recommendation H.265 and ISO/IEC 23008-2*, April 2013 (and subsequent editions).
- [6] G.J. Sullivan, J. Ohm, T.K. Tan, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, no. 12, pp. 1649-1668, December 2012.
- [7] G. J. Sullivan, J. M. Boyce, Y. Chen, J. R. Ohm, C. A. Segall, A. Vetro. "Standardized extensions of high efficiency video coding (HEVC)" *IEEE Journal of Selected Topics in Signal Processing*, 7(6), pp. 1001-1016, (2013). D.B. Sansli, K. Ugur, M.M. Hannuksela, M. Gabbouj, "Inter view motion vector prediction in multiview HEVC," *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1-4, 2014.
- [8] G. Van Wallendael, S. Van Leuven, J. De Cock, F. Bruls, and R. Van De Walle, "3D Video Compression Based on High Efficiency Video Coding," *IEEE Transactions on Consumer Electronics*, 58(1), pp. 137-145, March 2012.
- [9] J. Stankowski, M. Domanski, O. Stankiewicz, J. Konieczny, J. Siast and K. Wegner, "Extensions of the HEVC technology for efficient multiview video coding." In *2012 19th IEEE International Conference on Image Processing*, pp. 225-228, September 2012.
- [10] A. Vetro, "Frame compatible formats for 3D video distribution," In *2010 IEEE International Conference on Image Processing* (pp. 2405-2408, September 2010).
- [11] A. Vetro, T. Wiegand, and G.J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H. 264/MPEG-4 AVC standard," *Proceedings of the IEEE*, 99(4), pp. 626-642, April 2011.
- [12] Mallik, B., Akbari, A. S., & Zadeh, P. B, "HEVC based Stereo Video codec", *IET ISP*, 2015.
- [13] G. Tech, "MV-HEVC Software HTM-16.0-MV- Draft 5", *ITU-T SG16 WP3 and ISO/IEC JTC 1/SC 29/WG, JCT3* (2016)
- [14] K. Muller, A. Vetro. "Common Test Conditions of 3DV Core Experiments", *ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V G1100*, pp. 1-7, (2014).

TABLE II. PSNR COMPARISON FOR MV-HEVC AND THE PROPOSED HEVC-FIMV CODED FOR CODING "BALLOONS" SEQUENCE.

QP	kbps		Y-PSNR (dBs)		U-PSNR (dBs)		V-PSNR (dBs)	
	MV-HEVC	Proposed codec	MV-HEVC	Proposed codec	MV-HEVC	Proposed codec	MV-HEVC	Proposed codec
25	1516.275	1428.48	42.4662	43.3798	43.02393	42.7594	43.31677	43.1608
30	791.5968	749.375	40.28697	41.2112	41.81877	41.3412	41.78193	41.4232
35	446.3736	416.705	37.7224	38.1293	40.52787	39.9469	40.2097	39.7286
40	265.4616	246.025	34.96603	35.6514	39.5501	38.8975	39.06727	38.4474

TABLE III. PSNR COMPARISON FOR MV-HEVC AND THE PROPOSED HEVC-FIMV CODED FOR CODING "KENDO" SEQUENCE.

QP	kbps		Y-PSNR (dBs)		U-PSNR (dBs)		V-PSNR (dBs)	
	MV-HEVC	Proposed codec	MV-HEVC	Proposed codec	MV-HEVC	Proposed codec	MV-HEVC	Proposed codec
25	1472.535	1210.085	42.9236	43.9308	44.5902	44.3756	44.38583	44.1081
30	746.8848	611.375	40.70433	41.7435	43.7386	43.3352	42.99043	42.5611
35	418.0824	355.01	38.2951	39.1004	42.82693	42.2908	41.51783	40.9532
40	248.0648	209.54	35.75747	36.4223	42.10213	41.4779	40.35893	39.6628

TABLE IV. PSNR COMPARISON FOR MV-HEVC AND THE PROPOSED HEVC-FIMV CODED FOR CODING "NEWSPAPER1" SEQUENCE.

QP	kbps		Y-PSNR (dBs)		U-PSNR (dBs)		V-PSNR (dBs)	
	MV-HEVC	Proposed codec	MV-HEVC	Proposed codec	MV-HEVC	Proposed codec	MV-HEVC	Proposed codec
25	1650.587	1572.82	40.5277	41.2027	43.30997	43.881	43.32573	42.8121
30	814.9024	749.375	38.1854	39.0336	41.79477	41.2284	41.73343	41.1635
35	438.6056	391.315	35.7213	36.5581	40.35327	39.72	40.2007	39.4933
40	252.4216	220.58	33.23517	34.0369	39.3072	39.6486	39.1031	38.4274