HEVC based Stereo Video codec

B Mallik*, A Sheikh Akbari*, P Bagheri Zadeh [†]
**School of Computing, Creative Technology & Engineering, Faculty of Arts, Environment & Technology, Leeds Beckett University, U.K.*
b.mallik6347@student.leedsbeckett.ac.uk, a.sheikh-akbari@leedsbeckett.ac.uk
[†]*School of Computer Science and Informatics, De Montfort University, U.K., pooneh.bagherizadeh@dmu.ac.uk*

Keywords: MV-HEVC, HEVC, stereo video compression, texture coding.

Abstract

Development of stereo video codecs in latest multi-view extension of HEVC (MV-HEVC) with higher compression efficiency has been an active area of research. In this paper, a frame interleaved stereo video coding scheme based on MV-HEVC standard codec is proposed. The proposed codec applies a reduced layer approach to encode the frame interleaved stereo sequences. A frame interleaving algorithm is developed to reorder the stereo video frames into a monocular video, such that the proposed codec can gain advantage from inter-views and temporal correlations to improve its coding performance. To evaluate the performance of the proposed codec; three standard multi-view test video sequences, named “Poznan_Street”, “Kendo” and “Newspaper1”, were selected and coded using the proposed codec and the standard MV-HEVC codec at different QPs and bitrates. Experimental results show that the proposed codec gives a significantly higher coding performance to that of the standard MV-HEVC codec at all bitrates.

1 Introduction

Three dimensional (3D) video enhances the ability to perceive the relative depth information of real world scenes. Over the last decade, stereo video coding has evolved as a viable option to produce 3D video contents [1]. Stereo videos have found specific roles in various applications, such as industrial automation, automatic surveillance, remotely operated vehicle navigation, robotic systems, e-Learning systems and in 3D machine-vision applications for object location, identification and measurements. In recent years 3D video entertainment market has grown enormously and the increasing popularity of 3D-TVs and movies has led to development of some efficient stereo video codecs. The simplest and a cost-effective way to produce stereoscopic videos is by using video pairs acquired simultaneously through two parallel axes geometrically aligned, identical cameras [2].

Compressing video content effectively is the elementary role of a video codec. The generic aim of video compression is to meet the requirements of compact storage in memory spaces and/or to minimize bandwidth requirement for rapid transmission through a communication channel. A single view (monocular) video codec uses DCT and DPCM techniques along with quantisation methods to compress the video data by removing spatial and temporal redundancies existing within the video sequences [2, 4]. A straight forward approach to encode stereo videos is using the standard video codecs. Another approach is to alter the stereo videos to meet the video codec’s input requirements such that the visual quality of the decoded videos is not lost. Alternatively, suitably modified video codes are used to encode stereo videos. Some coding techniques use modified codec along with altered stereo videos to gain maximum coding performance. Depending on how the video pairs are encoded, the coding techniques are classified as simulcast, SEI, multi-view based, scalable video, mixed resolution, and Video + Depth based coding [3]. Standard 3D video codecs use the principle of combining temporal and inter-view prediction techniques to improve the coding performance. Various coding standards such as: H.264/AVC, MPEG 3DAV, H.264/MVC, H.264/MVD and 3D-HEVC, have been developed over the years to efficiently compress the 3D videos [3, 4, 5, 20]. With the release of H.264 and its extension H.264/AVC standard video codec that support multiple frames referencing, many innovative research have been carried out to adopt this monocular video codec to code stereo/multi-view videos.

The technique proposed by Gunatilake et al. [6] introduced the concept of cross image or wordline correlation between the left and the right views to compress the stereo videos. The compression scheme in this technique preselects frames that need high bandwidth and uses intra-coding to encode them, then it uses a modified motion estimation and compensation technique for the remaining frames. Li et al. [7] found that the coding performance of their three inter-view prediction schemes for stereoscopic videos are better than simulcast coding techniques, in terms of objective quality of the video content. The technique discussed in [7] along with added capabilities to predict motion vectors, which is based on disparity and worldline correlation to optimise the performance, was proposed by Adikari et al. [8]. This technique has gained superior coding efficiency by utilising the toolsets featured in H.264 standard and feeding the combined stereo view streams through a multiplexer to the encoder. By doing so the technique has successfully encoded stereo videos by adapting to a monocular H.264 standard video codec. These methods made modifications to standard codecs to develop a robust stereo video codec, by introducing additional disparity predictions between the two views to encode the multiplexed stereo video stream [8, 9].

An analysis of different combinations of temporal and inter-view prediction techniques was conducted by Merkle et al. [10], for multi-view video compression technique, based on H.264/AVC standard video coding. Results revealed that efficiency of the mixed inter-view/temporal prediction modes strongly depends on properties of the multi-view video sequence and coding gains can be achieved by additional inter-view reference pictures for disparity-compensated prediction. Over the last decade MVC based approaches have attracted stereo and 3D video codec development; however the challenging aspect in this technique has been to deal with inherent computational complexity and the high bandwidth requirements incurred due to multiple views. Coding algorithms have been proposed based on motion vector quantisation, flexible GOP structures that can adapt to different characteristics of multi-views videos, estimating motion homogeneity by calculating the difference in horizontal and vertical motion vectors for complex motions and an adaptive search window range algorithm by calculating differences between the predictor vectors [11, 12]. The results from MVC based stereo and 3D video coding techniques have shown that increasing the number of inter-view prediction effectively saves the encoded video bitrates [12, 13].

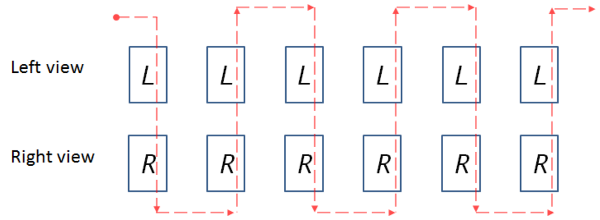
Another way of coding stereo videos is by using asymmetric resolution coding techniques, where video quality of the additional views are reduced by scaling down the resolution spatially or temporally. Asymmetric video coding techniques benefit from human visual system's tolerance to suppressed high frequency components and reduced resolution in one of the views. Coding efficiency for different scaling levels and resolution for the stereo views was studied by Hewage et al. [14] and Gürler et al. [15], the coding performance of their techniques was found to be close to the standard multi-view video coding technique, whereas it was able to deliver higher subjective qualities. The subjective study on coding performance of asymmetric and symmetric stereo video coding techniques conducted by Saygili and Gürler [16], based on H.264/MVC codec, showed that asymmetric coding out performs symmetric coding at high bitrates, with compression efficiency close to that of H.264/MVC codec. An adaptive spatial resolution down sampling technique was proposed by Aflaki et al. [17], wherein a frequency domain analysis is used to estimate the spatial resolution of both views of the stereo video streams. To achieve the best objective compression performance they have defined down sampling thresholds for a non-linear filter, after analysing high frequency components of the first frame of the video sequence. Majority of these techniques, propose modifications to the codec to encode available videos. Video frame resolution sampling, frame packing, and video sequence altering techniques proposed by few researchers assemble the video sequence to be encoded without modifications to the video codec. Setbacks in the current coding techniques necessitate a newer approach to encode stereo videos.

The objective of the present study is to find a computationally less complex way to encode stereo videos, by bringing in alteration to existing standard video codec and yet be able to deliver the immersive 3D video experience. The stereo video coding schemes proposed by other researcher in this context have been analysed. A novel texture based coding technique for coding stereo videos based on the latest standard multi-view extension of HEVC referred to as MV-HEVC codec is proposed in this paper. The idea behind the proposed technique is to have the input video sequence and the video codec synchronised with each other in order to get maximum coding performance. The frames of the stereo views are interleaved before they are encoded by the codec using a frame interleaving algorithm which rearranges the two views into a single view video. Then the resulting monocular video is coded using MV-HEVC codec which has been modified to encode the single view video. The proposed codec has the flexibility of accessing the I-frame from the next Group of Pictures (GoP), has a higher number of B-Frames within GoPs and uses only one layer to encode the stereo views. The coding performance of the designed codec is compared against the performance of standard MV-HEVC for two views scenario. The remainder of this paper is organized as follows: Section 2 presents the framework of the proposed technique by introducing the frame interleaving algorithm and the codec design parameter to encode interleaved videos. Section 3 presents the experimental results of the proposed framework and finally, the paper is concluded in section 4.

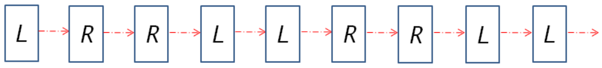
2 MV-HEVC based stereo video codec framework

The proposed MV-HEVC based stereo video coding scheme first interleaves the frames from stereo video pair to generate a monocular video stream, the interleaving algorithm arranges stereo video frames in such a fashion that two consequent frames of each view are always next to each other, as in Figure 1. The resulting monocular video sequence is then coded, as a single layered video, by the modified MV-HEVC codec. This enables the proposed codec exploit both temporal and inter-view correlations more efficiently.

The state of the art MV-HEVC design is based on MVC extension of H.264/ MPEG-4 AVC framework. The high level syntax of the MV-HEVC codec is an extension of the H.264/ MPEG-4 AVC codec. The monoscopic HEVC has been extended to multi-view video coding by including signalling for prediction dependencies between different views. The reference picture list in HEVC has been modified to improve inter-view prediction process between different views. Hence, other views' decoded frames can be used for the prediction of the current frame. MV-HEVC follows a layer representation for the additional views of multi-view videos to enable inter-view motion and texture parameter predictions. One of the views that is encoded, in full resolution, is named base layer. The additional views are



a) Algorithms contour to interleave stereo frames.



b) Interleave representation of the stereo frames.

Figure 1: Stereo frame interleaving block diagram: a) Algorithms contour to interleave stereo frames and b) Interleave representation of the stereo frames

treated as enhancement layers and are encoded either in lower resolution or same resolution as the base layer.

The MV extension of the HEVC, which is known as the MV-HEVC, uses a multi-loop decoding design, which is in contrast to the single loop decoding design of the H.264/AVC. Hence, MV-HEVC requires decoding all the encoded reference layers representing encoded views, prior to decoding a new layer. This layer encoding dependency significantly increases the decoding complexity of the MV-HEVC codec [18, 19]. Therefore, a reduced layer encoding approach could reduce decoding complexity of the MV-HEVC codec.

The proposed MV-HEVC based stereo video codec uses a reduced layer approach to reduce the decoding complexity of the codec. The flexibility of accessing I-frame from the next Group of Pictures (GoP) and supporting higher number of B-Frames within GoPs have made the MV-HEVC codec a suitable candidate for implementing the proposed single layer stereo video coding. The HTM-14.0-MV-draft 3 MV-HEVC software platform [20] was modified to implement single layer coding approach using the reference frame structure illustrated in Figure 2.

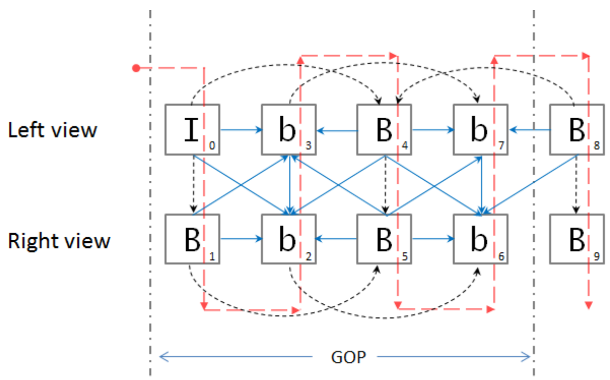


Figure 2: Reference frame structure of the proposed MV-HEVC based stereo video codec.

Parameter	***Value***
GOP size	8
Intra period	24
NumberOfLayers	1
VpsNumLayerSets	1
Number of ViewId	1
OutputLayerSetIdx	0
LayerIdsInAddOutputLayerSet_0	0
QP	25, 30, 35, 40

Table 1: MV-HEVC parameter settings for the proposed codec.

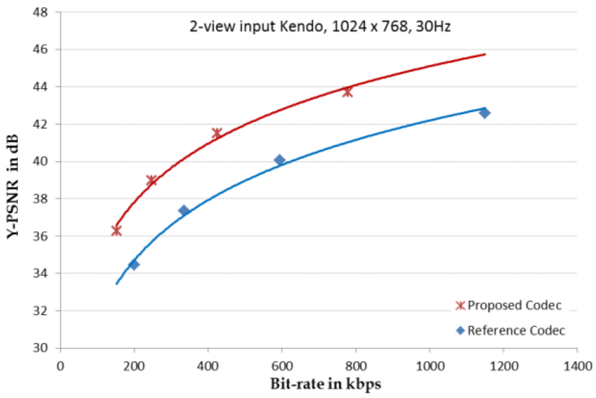
The parameter modifications that were applied to the standard MV-HEVC codec to implement the proposed codec are tabulated in Table 1. As it can be seen from Table 1, "NumberOfLayers" and "ViewId" parameters in the configuration file of the MV-HEVC are set to value one to enforce the codec to operate in the single layer mode. This will facilitate the stereo video coding, where stereo video frames have been interleaved. In addition, "OutputLayerSetIdx" parameter is set to value zero to indicate single layer decoding method both in encoder and decoder side.

3 Experimental results

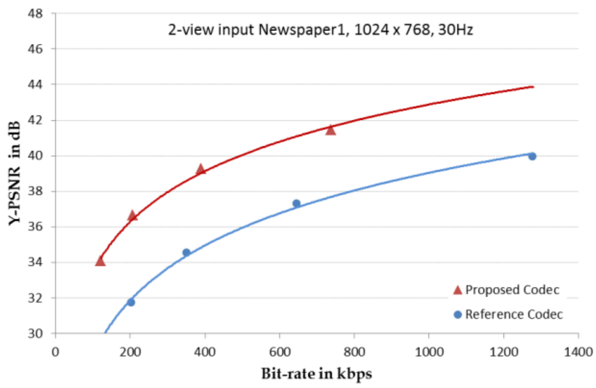
The compression efficiency of the designed codec was compared with the standard MV-HEVC codec. To achieve this, views 5-4, 1-3 and 2-4 of "Poznan_Street", "Kendo" and "Newspaper1" standard multi-view video sequences were chosen respectively and coded using the proposed MV-HEVC based stereo video coding scheme. These video sequences cover both the indoor and outdoor scenes with 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 [21]. Tables 2a-c show the resulting PSNR and consumed bitrate for coding "Poznan_Street", "Kendo" and "Newspaper1" stereo videos at QP 25, 30, 35 and 40. The proposed codec outperforms the MV-HEVC codec in terms of the PSNR of the decoded frames (up to 1dBs) while it significantly reduces the bandwidth requirements. From Table 2a, it can be seen that the proposed codec exhibits almost the same coding performance to that of MV-HEVC in terms of PSNR at QP of 35 when coding Poznan_Street sequences. However, the proposed codec requires about 8% lower bandwidth to transmit the videos, which implies improvement in coding in comparison to that of MV-HEVC. The proposed stereo video codec in general provides an average bitrate savings of 25% relative to the reference standard MV-HEVC codec.

To help better understand the performance of the proposed codec, the resulting Y-PSNR of the proposed codec and the MV-HEVC codec for coding "Poznan_Street", "Kendo" and "Newspaper1" test videos with respect to the bitrate are shown in Figures 3a-c. From these figures, it is clear that the proposed codec generates significantly higher coding performance to that of MV-HEVC at all bitrates (up to 1.2 dBs).

a) Poznan_Street



b) Kendo



c) Newspaper1

Figure 3: PSNR vs bitrate for MV-HEVC and the proposed codec for coding a) "Poznan_Street", b) "Kendo" and c) "Newspaper1" sequences.

(a) Proposed codec

(b) MV-HEVC

Figure 4: Decoded frame number 16 from Kendo videos of a) the proposed codec and b) the MV-HEVC standard codec.

To give a sense on the visual quality of the decoded videos, decoded frame number 16 of the proposed codec and MV-HEVC codec from Kendo video for the same view are shown in Figure 4. As it can be seen from these figures, the proposed codec's frame exhibits generally higher visual quality to that of MV-HEVC.

4 Conclusions

A MV-HEVC based stereo video codec that uses a reduced layer approach to encode frame interleaved stereo videos is proposed. The coding performance of the proposed codec was compared with the standard MV-HEVC stereo video codec using three standard stereo video sequences at different QPs and bitrates. 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 stereo video codec delivers superior video quality in comparison to the standard MV-HEVC codec at different QPs and bitrates.

Acknowledgements

This work was supported by School of Computing, Creative Technology & Engineering; Faculty of Arts, Environment & Technology, Leeds Beckett University, U.K

References

[1] P. Merkle, H. Brust, K. Dix, A. Smolic, T. Wiegand. "Stereo video compression for mobile 3D services", *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pp.1-4, (2009).

[2] M. G. Perkins. "Data compression of stereopairs", *IEEE Transactions on Communications*, **40**(4), pp. 684-696, (1992).

[3] 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), pp. 626-642, (2011).

[4] T. Wiegand, G. J. Sullivan., G. Bjøntegaard, A. Luthra. "Overview of the H. 264/AVC video coding standard", *IEEE Transactions on Circuits and Systems for Video Technology*, **13**(7), pp. 560-576, (2003).

[5] Information Technology-Coding of Audio-Visual Objects-Part 10: Advanced Video Coding, Amendment 1: Constrained Baseline Profile, Stereo High Profile and Frame Packing Arrangement SEI Message, document N10707, ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Jul. 2007.

[6] P. D. Gunatilake, M. W. Siegel, A. G. Jordan. "Compression of stereo video streams", *Signal Processing of HDTV International Workshop on HDTV'93*, pp. 173-185, (1994).

[7] S. Li, M. Yu, G. Jiang, T. Y. Choi, Y. D. Kim. "Approaches to H.264-based stereoscopic video coding", *IEEE 2004 IEEE First Symposium on Multi-Agent Security and Survivability*, pp. 365-368, (2004).

[8] A. B. B. Adikari, W. A. C. Fernando, H. K. Arachchi, K. K. Loo. "A H.264 compliant stereoscopic video codec", *IEEE 2005 Canadian Conference on Electrical and Computer Engineering*, pp. 1614-1617, (2005).

[9] P. Y. Yip, J. A. Malcolm, W. A. C. Fernando, K. K. Loo, H. K. Arachchi. "Joint source and channel coding for H.264 compliant stereoscopic video transmission", *IEEE 2005 Canadian Conference on Electrical and Computer Engineering*, pp. 188-191, (2005).

[10] P. Merkle, A. Smolić, K. Müller, T. Wiegand. "Efficient prediction structures for multiview video coding", *IEEE Transactions on Circuits and Systems for Video Technology*, **17**(11), pp. 1461-1473, (2007).

[11] Y. Kim, J. Kim, K. Sohn. "Fast disparity and motion estimation for multi-view video coding", *IEEE Transactions on Consumer Electronics*, **53**(2), pp. 712-719, (2007).

[12] L. Shen, Z. Liu, S. Liu, Z. Zhang, P. An. "Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding", *IEEE Transactions on Broadcasting*, **55**(4), pp. 761-766,(2009).

[13] S. Li, C. Hou, Y. Ying, X. Song, L. Yang. "Stereoscopic video compression based on H. 264 MVC." *2nd International Congress on Image and Signal Processing, IEEE CISP'09*, pp. 1-5, (2009).

[14] C. T. E. R. Hewage, H. A. Karim, S. Worrall, S. Dogan, A. M. Kondoz. "Comparison of stereo video coding support in MPEG-4 MAC, H.264/AVC and H. 264/SVC", *Proc. of IET Visual Information Engineering-VIE07*, (2007).

[15] C. G. Gürler, K. T. Bağci. "Adaptive stereoscopic 3D video streaming", *IEEE 2010 17th IEEE International Conference on Image Processing (ICIP)*, pp. 2409-2412, (2010).

[16] G. Saygili, C. G. Gürler. "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming", *IEEE Transactions on Broadcasting*, **57**(2), pp. 593-601, (2011).

[17] P. Aflaki, M. M. Hannuksela, M. Gabbouj. "Adaptive Spatial Resolution Selection for Stereoscopic Video Compression with MV-HEVC: A Frequency Based Approach", *2014 IEEE International Symposium on Multimedia (ISM)*, pp. 267-270, (2014).

[18] G. J. Sullivan, J. R. Ohm, W. J. Han, Wiegand, T. "Overview of the high efficiency video coding (HEVC) standard", *IEEE Transactions on Circuits and Systems for Video Technology*, **22**(12), pp. 1649-1668, (2012).

[19] 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).

[20] G. Tech. "MV-HEVC Software Draft 3", ITU−T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-K1009, (2015).

[21] 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).

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	3185.94	2609.142	39.4194	39.6869	46.68095	46.5996	45.4234	45.8668
30	1183.357	1035.892	37.53	37.6375	45.32565	45.1835	43.9465	44.3765
35	555.1744	498.0417	35.58665	35.5823	43.7643	43.3346	42.4607	42.6455
40	284.6808	260.9833	33.5462	33.549	42.64395	42.1984	41.46885	41.5371

a) Poznan_Street

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	1148.9408	777.9001	43.19105	43.7122	44.87895	44.7369	44.59105	44.6292
30	594.8096	424.19	40.95815	41.5079	44.039	43.8524	43.1651	43.1414
35	335.9528	246.32	38.51945	38.9898	43.06875	42.764	41.615	41.4274
40	201.1832	152.57	35.95415	36.3046	42.29845	41.8735	40.41975	40.06

b) Kendo

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	1278.1416	738.001	40.83645	41.462	43.5406	43.9861	43.6237	43.8154
30	646.5224	390.282	38.4995	39.2422	41.97955	42.3449	42.0138	42.213
35	351.3312	207.34	36.02505	36.6256	40.4611	40.4303	40.42235	40.2144
40	203.3856	120.92	33.5456	34.0556	39.3734	39.1183	39.2805	38.9349

c) Newspaper1

Table 2: PSNR comparison for MV-HEVC and the proposed codec for coding a) “Poznan_Street”, b) “Kendo” and c) “Newspaper1” sequences.