# MIXED-RESOLUTION HEVC BASED MULTIVIEW VIDEO CODEC

*Bruhanth Mallik, Akbar Sheikh Akbari, Ah-Lian Kor*

School of Computing, Creative Technology & Engineering,
Faculty of Arts, Environment & Technology, Leeds Beckett University, U.K.

## ABSTRACT

Studies have shown that mixed resolution based video codecs, also known as asymmetric spatial inter/intra view video codecs are successful in efficiently coding videos for low bitrate transmission. In this paper a HEVC based spatial resolution scaling type of mixed resolution coding model for frame interleaved multiview videos is presented. The proposed codec is designed such that the information in intermediate frames of the center and neighboring views are down-sampled, while the frames still retaining the original size. The codec's reference frames structure is designed to efficiently encode frame interleaved multiview videos using a HEVC based mixed resolution codec. The multiview test video sequences were coded using the proposed codec and the standard MV-HEVC. Results show that the proposed codec gives significantly higher coding performance over the MV- HEVC codec at low bitrates.

***Index Terms*** **—** HEVC, multiview video, mixed-resolution, video compression, low bitrate

## 1. INTRODUCTION

It is anticipated that, by 2020, 71% of the overall IP traffic used by both business and consumer users would be from non-PC devices, of which 82% would be/being used for visual content communication [1]. With the growing popularity of video-on-demand (VoD) and introduction of ultra-high-definition (UHD) video streaming, the IP VoD service is expected to constitute 21% of global VoD traffic by 2020.This means while the quality of network connectivity and bandwidth speed are fast improving, the demand for high definition (HD) and UHD quality monoscopic and 3D visual communication keeps growing at the same time. The multiview version of 3D videos uses multiple views (more than 2 views) to capture the same scene concurrently by geometrically aligned and synchronized cameras [2]. Due to huge amount of visual information in multiview videos, the bitrate needed to encode these videos increases approximately linearly with the number of views used to create the 3D content [3]. For this purpose multiview video coding (MVC) techniques have been extensively studied and standard texture based multiview point-to-point codecs, based on H.264/MPEG-4 AVC [4], [5] and high efficiency video coding MV-HEVC [6] have been developed. To efficiently compress huge amount of visual information in multiview videos, MVCs resort on extensively exploiting temporal and inter-view correlations using their multi-frame referencing features [5]. Despite the standard multiview extension of HEVC's (MV-HEVC) toolsets for reducing the overhead of signalling motion information, the codec's layer architecture for representing dependent views of multiview video stagnates the compression efficiency of the codec [6]. Hence, multiview video transmission over a bandwidth limited communication channel demands high level of compression efficiency. However, achieving high compression ratios through large quantization values leads to distortions and blocking artefacts, due to loss of high frequency information within the videos. To overcome these setbacks encountered by 3D video codecs for low bitrate transmission, mixed-resolution based video codecs, also known as asymmetric spatial inter/intra view frames resolution video codecs, were proposed in [7] and [8]. Further studies have shown that mixed-resolution based 3D video coding in case of stereoscopic videos for low bitrate transmission offer significant bandwidth savings while retaining adequate video quality [9],[10]. Studies conducted in [11] and [12] provide insight into improving compression efficiency of mixed-resolution 3D video codecs by using inter-view prediction. An inter-view motion vector prediction technique for MV-HEVC was proposed in [13], 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 referring to a look up table for disparity conversion from previously encoded frames and used this value to modify the motion field of inter-view reference pictures.

The objective of the research reported in this paper is to develop a texture HEVC for multiview videos based on spatial resolution scaling type of mixed resolution coding. In our previous study to develop a less complex HEVC based video codec for stereo and multiview videos, frame interleaving technique was used to encode the reordered stereo and multiview video frames using a modified MV-HEVC [14] [15]. The results of this study show that significant coding gain can be achieved compared to the standard MV-HEVC's stereo and multiview video codecs. The proposed coding model applies mixed spatial resolution coding to multiview video frames that are reordered using a frame interleaving algorithm developed for multiview videos in [14]. Experimental results for the standard video sequences show that the proposed video codec generates significantly higher coding performance than anchor MV-HEVC codec. The remainder of this paper is organized as follows: Section 2 presents the framework of the proposed technique by introducing the structure of the mixed resolution multiview videos, the proposed intermediate frame down-sampling method, multi-view video frame interleaving algorithm and the design of the codec's. Section 3 presents the experimental results and finally, the paper is concluded in Section 4.

## 2. PROPOSED MIXED-RESOLUTION HEVC BASED MULTIVIE VIDEO CODEC

The proposed mixed resolution HEVC based multiview video codec (MRHEVC-MVC) has been developed using a standard HEVC codec. The standard codec has been modified and configured to encode frames with different resolution, as shown in Figure 1, for low bitrate multiview video transmission. The MRHEVC-MVC's design allows I-frames to retain their full resolution whereas the intermediate P- and B-frames are spatially down-sampled by factor of 2. In addition to encode video frames of different resolutions, the proposed

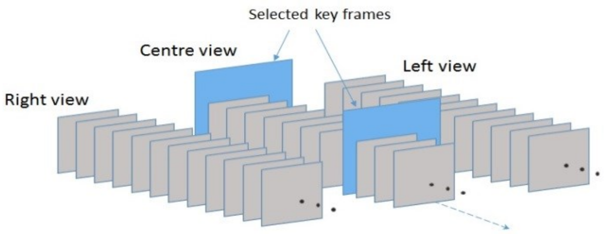


Figure 1. Frame structure of the proposed mixed resolution multiview video codec.

| | | |
|---|---|---|
| 0.0381 | 0.1051 | 0.0381 |
| 0.1051 | 0.4273 | 0.1051 |
| 0.0381 | 0.1051 | 0.0381 |

Table 1. Blackman 2D FIR filter coefficients.

MRHEVC- MVC is designed to be able to use different resolution decoded video frames for inter-frame and inter-view motion/disparity estimation/compensation. Further to simplify frame referencing in mixed resolution videos, the low resolution intermediate frames can use the I-frames for inter-frame referencing after reducing the reference picture size to the same size as the frame to be encoded. But the low resolution intermediate frames are not allowed by the MRHEVC-MVC's design to be used as reference frames. Thus, the size of the video frame to be encoded is never larger than its reference frame. A Blackman 2D FIR low-pass filter (filter coefficients are given in Table 1) is used to mitigate aliasing artefacts due to down-sampling [16]. The filtered intermediate frames are both horizontally and vertically down-sampled to maintain the same aspect ratio as the full resolution key-frames (I-frames). The development of the proposed MRHEVC-MVC's algorithm is summarized in the following order:

- First the standard HEVC is modified to encode mixed resolution video frames.
- The multiview video frames are temporally reordered into a monoscopic video using the frame interleaving algorithm
- Reference frame structure for the mixed resolution multiview video codec designed.
- The proposed mixed resolution multiview video codec is configured to encode frame interleaved multiview videos.
- Lastly the encoded/reconstructed intermediate frames are up-sampled to their original frame resolution (as shown in Figure 4).

One of the primary challenges for a mixed resolution video codec is to be able to encode different spatial resolution frames in the same sequence. To address this problem, in the proposed codecs the low resolution intermediate frames are designed to have the same resolution as the full resolution I-frames. The filtered and down-sampled low resolution intermediate frames are superimposed on top left corner of a full resolution sized frame with every pixel value set to zero, as illustrated in Figure 3. As a result, the intermediate frames and the key frames of the mixed-resolution video sequence have the same frame size. Reordering the mixed resolution multiview video frames to a monocular video facilitates the mixed resolution HEVC video codec to exploit interview correlation and the temporal disparity of the multiple views, through a single view. The frame interleaving algorithm rearranges the multiview frames (as shown in Figure 5) such that at least two frames of a view are as close as possible such that they are easily available for inter-frame referencing and thus improving compression efficiency. The reference frame structure of the proposed MRHEVC-MVC is designed by taking into consideration that the mixed resolution multiview video frames have been reordered into a monocular sequence, starting

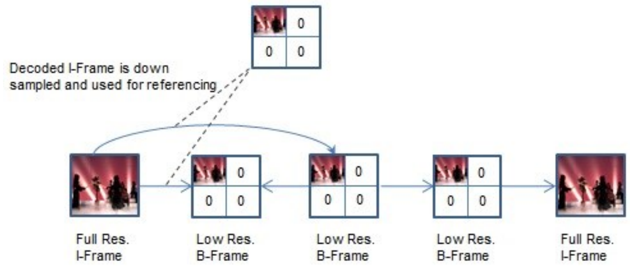


Figure 2. Inter-frame referencing of the proposed mixed resolution multiview video codec.

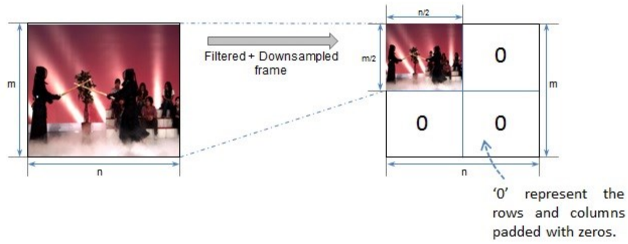


Figure 3. Intermediate frame resolution downscaling.

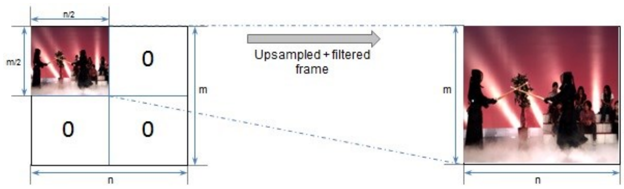


Figure 4. Intermediate frame resolution up-sampling.

from the middle view. The reference frame structure of the proposed codec for 3 view multiview scenario is shown in Figure 6. A mixed resolution based video codec is designed to be able to use video frames with different resolution as reference frame. The proposed MRHEVC-MVC's inter-frame referencing is designed to allow low-resolution intermediate frames to use low resolution intermediate frames and also down-sampled replica of the I-frames. However, the proposed MRHEVC-MVC codec's reference frame structure does not allow full resolution frames to use low resolution intermediate frames as an inter-frame reference. The copy of the reconstructed full resolution I-frames saved in the decoded picture buffer (DPB) goes through the down-sampling process, shown in Figure 2. Thus, by making these design modifications to the standard HEVC, the codec is equipped to encode mixed-resolution multiview video frames.

To implement the proposed MRHEVC-MVC video codec, JCT-VC HEVC software version HM16.12 was used. In addition to the modifications to the standard HEVC, the configuration parameters of the codec were set to encode frame interleaved mixed-resolution multiview video frames with the reference frame structure shown in Figure 6. Since the frames from the 3 view of the multiview videos are temporally reordered using the proposed frame interleaving algorithm, the group of pictures (GoP) size is set to 12 frames and the intra-frame period is set to 48. The goal of motion/disparity estimation/compensation is to reduce the energy of the difference block in order to reduce the number of bits required to code the coefficients in the block. This is achieved by finding the same scene in either neighboring view or previous frame. In case of neighboring views, the scene location is a function of distance of the camera from the scene and inter-camera angles [17]. For the standard test videos, used in this study, motion vector search range was set to 96 to mitigate the effect of the inter-camera angle and camera distance from the scene for disparity estimation/ compensation.

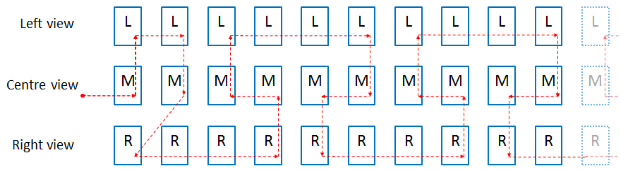


Figure 5. Frame interleaving algorithm's contour to reorder multiview video frames.

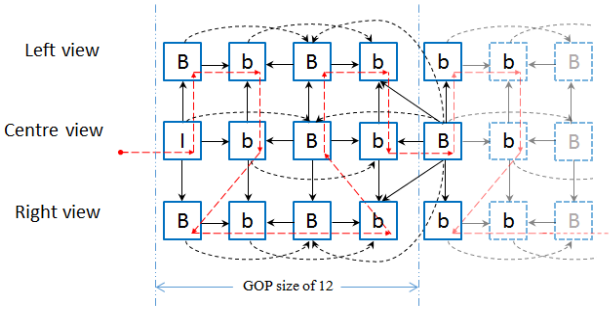


Figure 6. Reference frame structure of the proposed MRHEVC-MVC codec.

## 3. EXPERIMENTAL RESULTS

To evaluate the coding performance of the proposed mixed resolution HEVC based multiview video codec (MRHEVC-MVC), three views 1-3-5, 2-4-6 and 5-4-3, of "Balloon", "Newspaper1" and ""Poznan_Street" standard multiview video sequences respectively, were selected and coded using the proposed codec. The experimental results were compared with the anchor MV-HEVC's coding performance for the same 3 view scenario multiview datasets, which are available in JCT3V-G1100 common test condition documentation [18]. Figures 7–9 show the resulting average PSNR for the Y-frames at Quantization Parameters (QP) of 20, 25, 30, 35 and 40, respectively. The selected standard test video sequences cover both static and dynamic backgrounds at different level of illuminations. The "Balloons" data set is recorded under multiple artificial lighting conditions, where the background is progressively changing with a number of fast moving objects in the foreground. Figure 7, shows the resulting average Y-PSNR for "Balloons" test videos and it can be seen that the proposed codec, on average has higher luminance quality metric, i.e. Y-PSNR of up to 1.3dB greater than the anchor MV-HEVC codec. The proposed MRHEVC-MVC codec gives 1.7dB higher PSNR than the anchor codec at 508.88kbps and about 2dB higher PSNR between 450kbps to 260kbps. The coding performance of the proposed codec shows an increasing trend as the QP increases and the difference with anchor codec reaches to 2.1231dBs at 265 kbps. Figure 8 shows the comparison of the resulting coding performance of the proposed codec with the anchor MV-HEVC for "Poznan_Street" test video. The "Poznan_Street" data set is an outdoor-recorded video sequence captured under natural lighting condition. It contains multiple moving objects with a stationary background and fixed camera position. From figure 8, it can be noted that the proposed codec's Y-frames have an average 0.7dB greater PSNR than the anchor codec's frames at 350kbps to 870kbps. Since this test videos have huge still background with small moving areas, the still areas are coded by larger CTUs, which represented by smaller number of bits. Hence the coding performance of the proposed codec is limited by the areas' of the small moving objects and the results do not exhibit significant gains. The "Newspaper1" test video is an indoor recording with moderate illumination using an artificial light source. The scene in this test video have a constant background without that many moving objects in the foreground. The resulting Y-frames' average PSNR for coding "Newspaper1" test video sequence using the proposed MRHEVC-MVC and the anchor codecs is shown in Figure 9. From Figure 9, it can be seen that the proposed codec on average exhibits higher luminance quality metric, i.e. Y-PSNR of up to 1.17dB.The proposed MRHEVC-MVC codec gives 1.2dB higher PSNR greater than the anchor codec at 720 kbps and the difference increases to 1.8dB at 252 kbps.

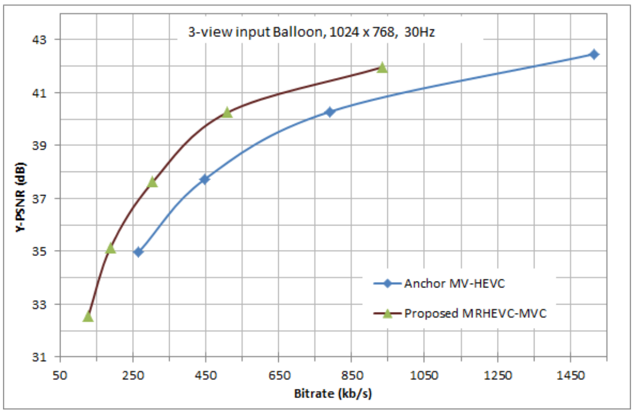


Figure 7. Y-frames avarage PSNR of the MV-HEVC anchor codec and the proposed MRHEVC-MVC codec for coding "Balloons" test videos at different bitrates.

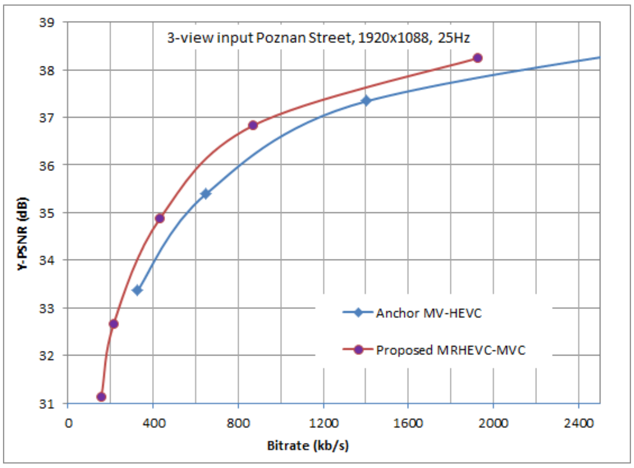


Figure 8. Y-frames average PSNR of the MV-HEVC anchor codec and the proposed MRHEVC-MVC codec for coding "Poznan Street" test video at different bitrates.

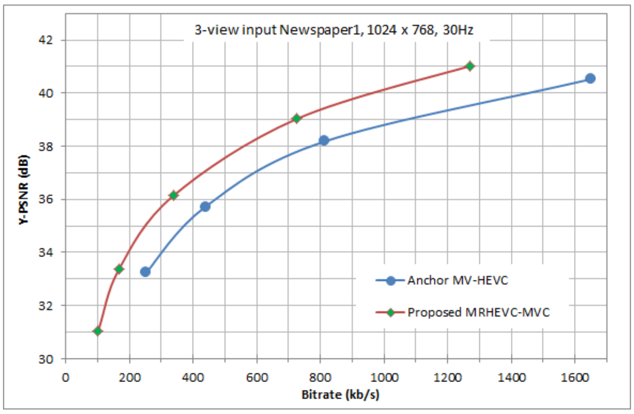


Figure 9. Y-frames average PSNR of the MV-HEVC anchor codec and the proposed MRHEVC-MVC codec for coding "Newspaper1" test video at different bitrates.

(a) Proposed MRHEVC-MVC

(b) Anchor MV-HEVC

Figure 10. Decoded intermediate center view frame 98 from "Balloons" test videos at 304.4 kb/s a) the proposed MRHEVC-MVC codec and b) the anchor MV-HEVC standard codec.

To give a sense of the visual quality of the decoded videos, decoded intermediate center view frame 98 of the proposed codec and the anchor codec from "Balloons" test video is shown in Figure 10. It can be seen that the proposed MRHEVC-MVC codec's image, shown in Figure 10.a, exhibits generally higher visual quality than the anchor MV-HEVC, shon in Figure 10.b.

## 4. CONCLUSIONS

A mixed resolution HEVC based multiview video codec (MRHEVC-MVC) was presented in this paper. The proposed codec temporally reorders the multiview video frames into a monoscopic mixed resolution videos. The HEVC standard codec was modified and configured to encode frames with different resolutions in the interleaved mixed resolution multiview videos. The proposed MRHEVC-MVC encodes the I-frames in full resolution and the intermediate frames in low resolution. The low-resolution intermediate frames are superimposed on to full resolution zero frames. Thus, the information in the low resolution intermediate frames in the center and neighboring views are down-sampled, while the frames still retaining the original size. This reduces the complexity for implementing the proposed MRHEVC-MVC codec. Experimental results show that the proposed codec outperforms the anchor HEVC codecs for low bitrate transmission.

## 5. REFERENCES


[1] Visual Networking Index Cisco, "The Zettabyte Era: Trends and Analysis," Cisco white paper, Jun. 2016.

[2] A. Smolic, "3D video and freeviewpoint video: From capture to display," in Pattern recognition, vol. 44, no. 9, pp. 1958-1968, 2011.

[3] Y. Chen, Y.K. Wang, K. Ugur, M.M. Hannuksela, J. Lainema and M. Gabbouj, "The emerging MVC standard for 3D video services," in EURASIP Journal on Applied Signal Processing, pp. 8, 2009.

[4] ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audio visual service," in Document ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2010.

[5] 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," in Proc. of the IEEE, vol. 99, no. 4, pp. 626-642, Apr. 2011.

[6] G. J. Sullivan, J. M. Boyce, Y. Chen, J. R. Ohm, C. A. Segall and A. Vetro, "Standardized extensions of high efficiency video coding (HEVC)" IEEE Journal of Selected Topics in Signal Processing, vol. 7, pp. 1001-1016, 2013.

[7] M. G. Perkins, "Data compression of stereopairs," in IEEE Trans. on Communications, vol. 40, no. 4, pp. 684-696, 1992.

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

[9] L. B. Stelmach, W. J. Tam, D. Meegan and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," in IEEE Trans. Circuits and Systems for Video Technology, vol. 10, no. 2, pp. 188-193, 2000.

[10] A. Aksay, C. Bilen, E. Kurutepe, T. Ozcelebi, G.B. Akar, M. R. Civanlar and A. M. Tekalp, "Temporal and spatial scaling for stereoscopic video compression" in 14th European Signal Processing Conf., Proc. EUSIPCO, vol. 6, no. 8, pp. 1-5, 2006.

[11] C. Fehn, P. Kauff, S. Cho, H. Kwon, N. Hur and J. Kim " Asymmetric coding of stereoscopic video for transmission over T-DMB," in: Proc. 3DTV Conf., Kos Island, Greece, pp. 1–4, May 2007.

[12] H. Brust, A. Smolic, K. Mueller, G. Tech and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices" in 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, May 2009

[13] D.B. Sansli, K. Ugur, M.M. Hannuksela and 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.

[14] B. Mallik, A. Sheikh Akbari and P. Bagheri Zadeh, "HEVC based Stereo Video codec", in 2nd IET International Conference on Intelligent Signal Processing, 2015.

[15] B. Mallik and A. Sheikh Akbari, "HEVC based Multi-View Video Codec using Frame Interleaving technique", in 9th Int. Conf. Developments on eSystems Engineering, DeSE, UK, 2016.

[16] A. Sheikh Akbari and P. Bagheri Zadeh, "Wavelet Based Image Enlargement Technique," in Int. Conf. Global Security, Safety, and Sustainability. Springer International Publishing, pp. 182-188, 2015.

[17] S. Bouyagoub, A. Sheikh Akbari, D. Bull and N. Canagarajah, "Impact of camera separation on performance of H. 264/AVC-based stereoscopic video codec" IET Electronics letters, vol. 46, no. 5, pp. 345-346, 2010.

[18] K. Muller and A. Vetro, "Common Test Conditions of 3DV Core Experiments," in ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V G1100, pp. 1-7, 2014.