information

Article

HEVC Based Frame Interleaved Coding Technique for Stereo and Multi-View Videos

Bruhanth Mallik [1], Akbar Sheikh-Akbari [1,*], Pooneh Bagheri Zadeh [1] and Salah Al-Majeed [2]

[1] School of Built Environment, Engineering and Computing, Headingley Campus, Leeds Beckett University, Leeds LS6 3QR, UK
[2] School of Computer Science, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK
* Correspondence: a.sheikh-akbari@leedsbeckett.ac.uk; Tel.: +44-113-81-21767

Abstract: The standard HEVC codec and its extension for coding multiview videos, known as MV-HEVC, have proven to deliver improved visual quality compared to its predecessor, H.264/MPEG-4 AVC's multiview extension, H.264-MVC, for the same frame resolution with up to 50% bitrate savings. MV-HEVC's framework is similar to that of H.264-MVC, which uses a multi-layer coding approach. Hence, MV-HEVC would require all frames from other reference layers decoded prior to decoding a new layer. Thus, the multi-layer coding architecture would be a bottleneck when it comes to quicker frame streaming across different views. In this paper, an HEVC-based Frame Interleaved Stereo/Multiview Video Codec (HEVC-FISMVC) that uses a single layer encoding approach to encode stereo and multiview video sequences is presented. The frames of stereo or multiview video sequences are interleaved in such a way that encoding the resulting monoscopic video stream would maximize the exploitation of temporal, inter-view, and cross-view correlations and thus improving the overall coding efficiency. The coding performance of the proposed HEVC-FISMVC codec is assessed and compared with that of the standard MV-HEVC's performance for three standard multi-view video sequences, namely: "Poznan_Street", "Kendo" and "Newspaper1". Experimental results show that the proposed codec provides more substantial coding gains than the anchor MV-HEVC for coding both stereo and multi-view video sequences.

Keywords: texture 3D videos; multiview video coding; stereo video codec; HEVC; MV-HEVC; frame-interleaved video coding

check for updates

Citation: Mallik, B.; Sheikh-Akbari, A.; Bagheri Zadeh, P.; Al-Majeed, S. HEVC Based Frame Interleaved Coding Technique for Stereo and Multi-View Videos. *Information* **2022**, *13*, 554. https://doi.org/10.3390/info13120554

Academic Editor: Zoran H. Perić

Received: 6 September 2022
Accepted: 17 November 2022
Published: 25 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years 3D video entertainment market has grown enormously, however, the application of 3D videos is not limited to multimedia purposes alone. Three-dimensional videos are employed in immersive video conferencing, e-learning, cloud-based multimedia services, real-time surveillance, automation, robotics, and machine vision [1]. Multi-view videos are generated by using geometrically aligned and synchronized multiple cameras, which capture the same scene simultaneously. The vast amount of visual information contained in multi-view videos largely constitutes their demand for huge storage space, higher transmission bandwidth over a communication channel, and greater computational power for coding [2]. The bitrate for encoding the views of the multi-view videos individually, using monoscopic codecs, increases approximately linearly with the number of views, hence efficient compression techniques are necessary for such applications [3]. Unlike single-view video codecs, texture-based multi-view video codecs employ scene geometry implicitly, through disparity prediction and compensation across views, to efficiently compress stereo/multi-view videos. Stereo/multi-view video codecs extensively use disparity prediction/compensation (DPC) and motion prediction/compensation (MPC) techniques, which are designed to exploit the inter-view and temporal correlations, respectively [3,4]. In addition to DPC and MPC techniques, stereo and multi-view video codecs are provided

with advanced coding tools, such as hierarchical B picture (HBP) prediction structure, variable block-size motion estimation (ME), and disparity estimation (DE), to improve the coding efficiency. Standard 3D video codecs use the principle of combining temporal and inter-view prediction techniques to improve the coding performance. Numerous coding standards such as: H.264/AVC, MPEG 3DAV, H.264/MVC, and MV-HEVC, have been developed over the years to efficiently compress multi-view videos [5–8]. In [9], authors explored features and challenges with designing 3D virtual conferencing tools considering user experience, where virtual reality is employed for interacting with screen-based 3D conference environments. The mixed resolution multi-view video transmission was considered one of the solutions to overcome 3D video transmission bandwidth limitation challenges. An approach for the synchronized mixing of real-time audio/video streams from multiple peers while minimizing latency was presented in [10]. This method allows online live conversation system implementation of mixed live conversation streams from many peers and then again rebroadcast the mixed stream to many audiences. In [11], a Fuzzy based adaptive deblocking filter-based method for low-bitrate HEVC video transmission was presented. Authors demonstrated that transmission of complex videos at low bitrates induces visible artifacts to the decoded video and considerably degrades the picture quality. They introduced a four-step fuzzy-based adaptive deblocking filter selection technique to remove the quantization noise, blocking artifacts and corner outliers efficiently from HEVC decode videos. They demonstrated the effectiveness of their method using simulation. Their results show that their method's videos demonstrate higher objective quality in terms of PSNR and subjective quality. In [12] a model with two phases network-related settings (NRS) and video-related settings (VRS) for video transmission over mobile networks was introduced. The limitation of the mobile transmission link and the constraints of real-time video transmission were studied. The author used five distinct transcoding algorithms, including MPEG-4, H.264, H.265, VP8, and VP9, and assessed the contribution of the transmission links and transcoding techniques for optimum video transmission in terms of QoS and QoE. The presented simulation results show that the LTE network users frequently achieve a greater data rate, PSNR, SSIM, and lowest loss rate, regardless of the type of their used video compression methods. A view-dependent video encapsulation method was reported in [13]. The proposed method creates videos with different resolutions from the input video, generating a mixed-resolution MP4 with multiple tracks, which forms an encapsulated mixed-resolution file according to the user's viewport. The authors demonstrated that the proposed method could provide full-resolution video within the field of the view at significantly lower bandwidth without much noticeable quality impacts. In [14] a structure and characteristics of the bandwidth-efficient stereoscopic 3D broadcasting system, which uses the video resolution asymmetry between the left and right eye views, known as mixed-resolution, to save the bitrate, was presented. The authors demonstrated that the proposed system's video has a satisfactory objective and subjective quality when the proper bitrate is allocated. In [15], a view synthesis quality mapping for depth-based super-resolution on mixed resolution 3D video method was reported. The proposed method uses a mixed-resolution architecture to code the videos, where the center view is coded at its full resolution, and its neighboring views and depth map of all views are coded at a lower resolution. The authors showed that their method achieves superior results to that of the depth-based super resolution up sampling method. A combination of spatial mix-resolution frame interleaving for stereo and multi-view video compression has been further investigated in [16–18]. Although the standard multiview video codecs are designed to efficiently code various 3D video applications, they still fail to fulfill the requirement for multi-view video transmission at lesser bandwidth and computation costs. Whereas today's wireless and low-cost digital data transmission channels often operate at a much lower bitrate and most of the applications use high definition (HD) and ultra-high definition (UHD) videos. The objective of the research reported in this paper is to develop a less complex texture-based stereo/multi-view video codec by analyzing the coding process and prediction structures of standard HEVC codec.

Various stereo/multi-view video coding techniques have been developed to address the standard multi-view video codec's shortcomings. An analysis of different combinations of temporal and inter-view prediction techniques was conducted by Merkle et al. [19], for multi-view video compression technique based on the standard H.264/AVC video codec. Their results revealed that the efficiency of inter-view/temporal prediction combinations strongly depends on the properties of the multi-view video sequences and adding inter-view reference pictures for disparity prediction/compensation could increase the achieved coding gain. Over the last decade, the introduction of H.264/AVC and its multi-view extension, H.264-MVC, has attracted the interest of many researchers towards further developing advanced stereo/multi-view video codecs; however, the challenging aspects of these techniques have been to deal with inherent computational complexity and high bandwidth requirement, due to the nature of the multi-view videos. Many coding techniques have been proposed based on motion vector quantization, flexible group of pictures (GoP) structures that can adapt to different characteristics of multi-view 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 prediction vectors [20,21]. The results from MVC-based stereo/multi-view video coding techniques have shown that increasing the number of inter-view predictions effectively reduces the required bitrates [21,22]. Another way of coding stereo video is using asymmetric resolution coding techniques, where the video quality of the additional views is reduced by scaling down the resolution spatially or temporally. Asymmetric video coding techniques benefit from the 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 resolutions for the stereo videos was studied in [23,24], the coding performance of these techniques was found to be close to that of the standard multi-view video coding technique, while they were able to deliver higher subjective qualities. A subjective study on the visual quality of the decoded video frames of asymmetric and symmetric stereo videos was conducted in [25] using H.264/MVC codec. Their results showed that asymmetric video frames exhibit superior visual quality to those of symmetric videos at high bitrates, with compression efficiency close to that of H.264/MVC codec.

An inter-view motion vector prediction method was proposed in [26] to improve the coding efficiency of the dependent views by using previously encoded motion information of the reference views using temporal motion vector prediction. This method calculates a global disparity vector by accessing a look-up table generated from disparity vectors of the previously encoded frames. Then the global disparity vector is used to adjust the motion field of inter-view reference pictures. A multi-view video compression scheme using HEVC monoscopic codec was proposed in [27], the prediction structure of this technique closely matches that of H.264/AVC-based multi-view video codec with minimized prediction signaling. A less complex but improved motion and inter-view prediction multi-view video codec using HEVC was proposed in [28], which uses vector scaling for the targeted prediction units. In addition, it uses decision choices for selecting prediction candidates from co-located units in the reference frame to track the neighboring unit motion/disparity vector and identify a unit vector that can be used as the source of prediction (nested prediction). Although this method adds complexity for finding the best match in multi-view motion/disparity prediction, it does not produce significant coding gain to that of the standard MV-HEVC.

From the literature, it is evident that researchers have mainly investigated techniques that adapt the standard monoscopic and multi-view video codecs to encode stereo/multi-view videos. So far, the modification mostly encompasses changing the resolution of video frames and improving the motion/disparity estimation/compensation unit of the codec. However, an investigation into techniques, which enables the exploitation of cross-frame (also called as lateral frame) along with inter-view and temporal correlations for stereo and multi-view video coding, have been less reported in the literature. In this paper, an HEVC-based frame interleaved stereo/multi-view video codec is presented. The proposed

codec applies a novel frame interleaving technique on the stereo/multi-view video frames to increase the exploitation of cross-frame, temporal and inter-view correlations. The coding performance of the proposed codec is assessed and compared with that of the standard MV-HEVC using three standard multi-view video datasets. Experimental results show the merit of the proposed technique. The rest of this paper is organized as follows: Section 2 outlines the framework of the proposed stereo/multi-view video codec. Section 3 presents the experimental results and finally, the paper is concluded in Section 4.

2. HEVC Based Frame Interleaved Stereo and Multiview Video Coding Technique

The extension of the HEVC for code multi-view videos, known as MV-HEVC, uses a framework similar to the multi-view extension of H.264/ MPEG-4 AVC, namely H.264-MVC. In contrast to H.264-MVC, which uses a single loop coding approach, MV-HEVC uses a multi-loop decoding technique. Hence, MV-HEVC requires decoding all the encoded reference layers, prior to decoding a new layer. It also uses a layered representation for encoding multi-view videos, this layer encoding dependency significantly increases the decoding complexity of the MV-HEVC codec [29,30]. In this section, an HEVC-based Frame Interleaved Stereo/Multi-view video codec (HEVC-FISMVC), which uses a single layer encoding approach, is presented. The proposed frame interleaving algorithm reorders video frames in such a fashion that the resulting monoscopic video sequence has two consequent frames of each view next to each other, as shown in Figure 1. The proposed HEVC-FISMVC codec's frame interleaving algorithm's contour is shown by red dotted lines in Figure 1, for two-view (stereo) and three-view scenario multiview videos. From Figure 1a,b, the interleaving technique starts and completes the frame reordering at the left- and center-view for stereo and multiview video frames, respectively. This is to ensure that the I-frames are always from the left-view frame (L) for stereo videos and the center view frame (M) for multiview video frames. Figure 1a shows the frame interleaving algorithm's design for stereo video frames, where the algorithm's contour starts from the left view frame and selects the next frame from the right view, and from then two consecutive frames of each view are selected into the reordered sequence. In this investigation, it was found that a direct extension of two view scenario frame interleaving algorithm, as shown in Figure 1a, to three-view scenario multiview videos would not yield significant coding gains. This is because the frame referencing and signaling are not simplified by a large margin, as it would in the case of stereo video, therefore wanting a similar number of bits to represent them as it would be in MV-HEVC. Hence, a more resolute frame interleaving design for multiview videos with more than two views is presented in Figure 1b. From Figure 1b it can be seen that the frame interleaving algorithm's contour for three view scenarios in the proposed HEVC-FISMVC codec starts from the center view frame and then selects the next frame from the left view into the reordered sequence. The multiview view scenarios' frame interleaving algorithm in the proposed codec is designed to select two consecutive center view frames and a pair of two consecutive (which makes it quadruple) left view or right view frames into the reordered monoscopic sequence. This kind of frame interleaving in a multiview scenario will allow the reference frame architecture to use lesser bits to represent the different frames as they become referenced from a center view frame, which in turn refers to I-frame in the center view. The resulting monocular video sequence is then encoded using the MV-HEVC codec, which is configured to code a single-layered video. The proposed codec's reference frame architecture for frame-interleaved stereo and multiview videos is explained in the following paragraph. It has been shown in [21,22] that exploiting inter-view correlations, DPC, delivers higher coding efficiency than the temporal correlations, hence the standard codecs use intelligent combinations of disparity and motion compensation to increase the coding efficiency, which adds to the computation cost of the codec [19]. To further improve the coding efficiency of the codec, codecs exploit inter-view and temporal redundancies by taking up a group of advanced coding tools, such as hierarchical B picture (HBP) prediction structure, variable block-size motion estimation (ME) and disparity estimation (DE). The reference frame architecture

in MV-HEVC has been combined with signaling for prediction dependencies between different views. The MV-HEVC uses a multi-loop encoding design to encode frames from other views, it also uses a layered representation for encoding multi-view videos, this increases the decoding complexity of the codec since the multi-layered decoding process is an essential need for prediction prior to encoding a new layer. As a consequence, the video sequences of individual views cannot be processed independently since they share reference pictures with other views. The proposed HEVC-FISMVC codec benefits from coding the reordered multi-view monoscopic video frames. The reference frames architectures for the frame interleaved stereo and multi-view video codecs are shown in Figures 2 and 3, respectively. From these figures, it can be seen that in addition to adjustment inter-view and temporal frames referencing, the proposed reference frame architectures allow cross-frame referencing (also called as lateral frame referencing), if not, this would be computationally complex when implemented within the AVC framework (using the standard HEVC).

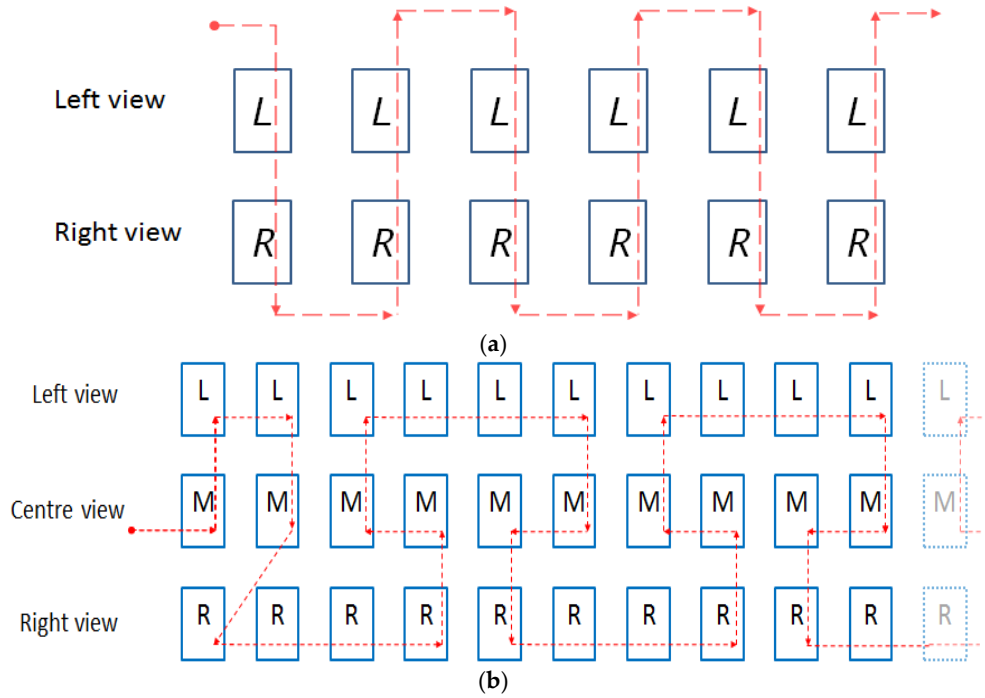


Figure 1. Stereo and multiview frame interleaving block diagram: (**a**) Algorithm's contour to interleave stereo frames and (**b**) Interleave representation of the stereo frames.

The goal of motion/disparity estimation/compensation is to reduce the energy of the difference block [31]. This is achieved by finding the same scene in either neighboring-view or the previous frame [32]. In the case of neighboring views, the scene location is a function of the distance of the cameras from the scene and their inter-camera angles. A study on the impact of camera separation on the performance of the multi-view video codecs has shown that, as the angle between the optical reference lines of the cameras, increases the inter-view motion correlation decreases [33,34]. For the standard test videos, used in this study, the motion vector search range was set to 96 pixels to mitigate the effect of the inter-camera angles and camera distances from the scene. The motion/disparity search area for the proposed codec is set to 96 pixels, so that the HEVC's advanced and temporal motion vector prediction toolsets work efficiently for frame interleaved multi-view monoscopic videos. The standard MV-HEVC video codec's software is configured to provide a single-layered HEVC codec operating with AVC capabilities and reduced transmission overhead. HTM-16.0-MV-Draft 5 software version [35] was used to implement the proposed HEVC-FISMVC codec. The values assigned to the parameters of the standard MV-HEVC codec in order to implement the proposed codec are tabulated in Table 1. The parameter "NumberOfLayer" is set to the value 1 to configure the single layer mode

of operation of the standard MV-HEVC codec. The parameters "Number OfViewId", "OutputLayerSetIdx" and "LayerIdsInAddOutputLayerSet_0" were assigned minimum values to run the standard codec with the least number of signaling 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 [35], but the minimum GoP size for the proposed codec's design for 2-view and 3-view multi-view video scenarios are set to 8, and 12 frames respectively.

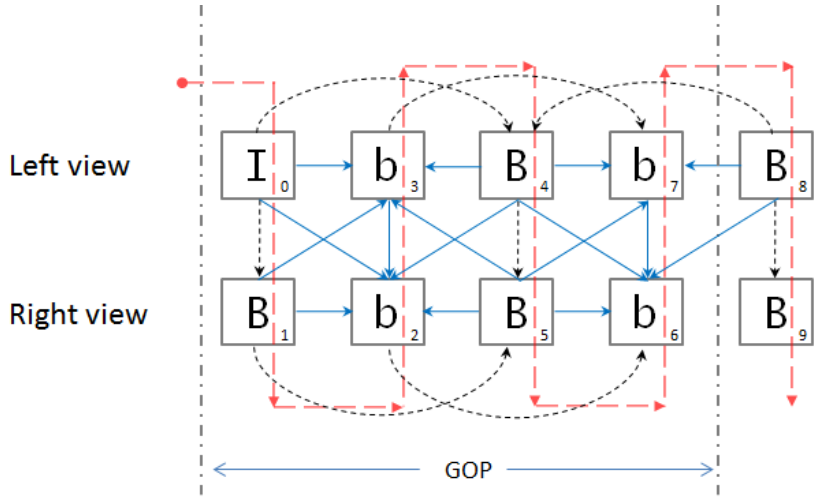


Figure 2. Reference frame architecture of the proposed HEVC-FISMVC for stereo videos.

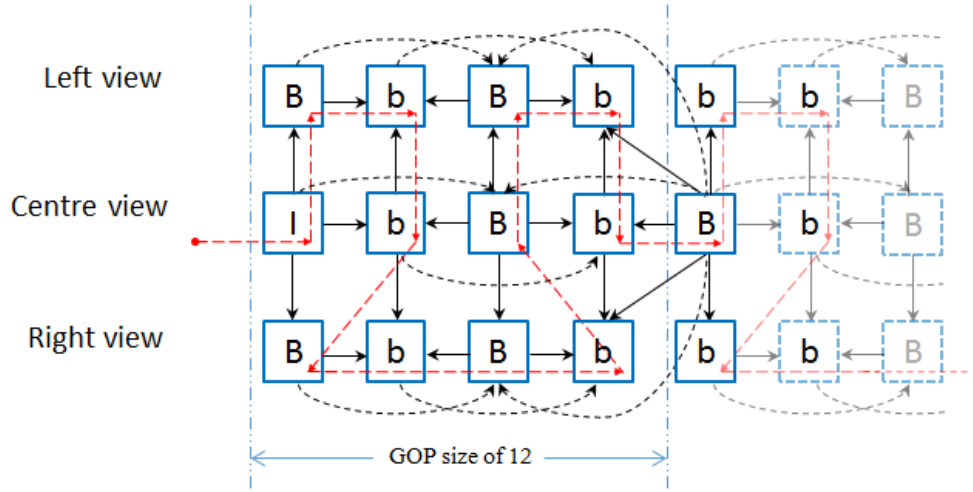


Figure 3. Reference frame architecture of the proposed HEVC-FISMVC for multiview videos.

Table 1. MV-HEVC design parameter for the proposed codec.

Parameters	Value
NumberOfLayers	1
NumberOfViewId	1
VpsNumLayerSets	1
OutputLayerSetIdx	0
LayerIdsInAddOutputLayerSet_0	0
GOP Size: 2-view scenario	8
GOP Size: 3-view scenario	12
Intra Period	24
QP	25, 30, 35, 40

3. Experimental Results and Analysis

The primary goal of the study presented in this paper is to investigate the merits of a frame interleaved video coding technique over the layered multi-loop approach of MV-HEVC while coding stereo and multiview videos. The coding performance of the proposed HEVC-based frame interleaved coding technique is compared with the anchor MV-HEVC codec, as presented in the JCT3V-G1100 document [34], for 2-view and 3-view scenarios. To evaluate the coding performance of the proposed HEVC-FISMVC codec, in a 2- view scenario, views 4-3, 1-3, and 2-4 of "Poznan Street", "Kendo" and "Newspaper1" standard multiview video sequences were chosen respectively and coded using the pro- posed codec. In the case of a 3-view scenario, views 5-4-3, 1-3-5, and 2-4-6 of "Poznan Street", "Kendo" and "Newspaper1" standard multiview video sequences were chosen, respectively, and coded using the proposed HEVC-FISMVC codec. The test video sequences illustrate different entertainment and interactive applications, with varying scene characteristics, illumination, and camera distances. The Peak Signal to Noise Ratio (PSNR) measure is then used to assess the quality of the decoded frames. The combined PSNR (YUV-PSNR) weighted sum of the average PSNR per video frames of the individual components (Y PSNR, U PSNR, and V PSNR) of the decoded stereo and multiview videos, as defined in Equation (1) [35], were calculated and compared with that of the anchor MV-HEVC codec.

$$YUVPSNR = \frac{(6.YPSNR + UPSNR + VPSNR)}{8} \quad (1)$$

The PSNR of each video frame component (with 8-bit pixel resolution), is calculated as shown in Equation (2).

$$YUVPSNR = -10.\log_{10}\left[\frac{1}{255^2.W.H}\sum_i\sum_j\left(I_{ref}(i,j) - I_{dec}(i,j)\right)^2\right] \quad (2)$$

where, $I_{ref}(i,j)$ and $I_{dec}(i,j)$ represent the corresponding pixel values of the reference Ire f and decoded Idec video video frame, respectively, while W and H represent width and height of the video frames, respectively. The experimental results presented in this paper for anchor MV-HEVC codec were taken from JCT3V-G1100 CTC documentation [34]. Figures 4–6 show the resulting YUV-PSNRs of the proposed and anchor codec for coding "Poznan Street", "Kendo" and "Newspaper1" test videos, for two view and three view scenarios, at Quantization Parameters (QP) 25, 30, 35 and 40, respectively.

The "Poznan Street" dataset is an outdoor video sequences captured under natural lighting condition. Its videos contain multiple moving objects with a stationary background, which were recorded by stationary cameras. From Figure 4a, it can be noted that, for 2-view "Poznan street" scenario, the proposed codec's video frames on average exhibit 0.74 dB greater YUV-PSNR than the anchor codec's video frames between 550 kbps to 3000 kbps bitrates. From Figure 4b, which shows results for coding the 3-view multiview video sequences, it is obvious that the proposed code outperforms the standard codec by an average of 0.52 dB, between 300 kbps and 3400 kbps. The "Kendo" dataset is an indoor multiview video captured under multiple controlled lighting sources. These videos contain progressive background changes with a number of fast-moving objects in the foreground. From Figure 5a, it is clear that the proposed codec gives higher coding performance than that of the anchor codec for coding 2-view "Kendo" with an average YUV-PSNR of 0.45 dB, at bitrates 280 kbps to 970 kbps. For coding 3-view "Kendo" multiview test sequences, the YUV components of the proposed codec's video frames have an average of 0.51 dB greater PSNR than that of anchor MV-HEVC codec's video frames, between 250 kbps and 1250 kbps, as shown in Figure 5b. The "Newspaper1" multiview video is an indoor video dataset that represents a scene with stationary background and moving objects close to stationary cameras, where the scene is moderately illuminated by artificial illuminants. The results for coding the "Newspaper1" 2-view and 3-view multiview video sequences are shown in Figure 6a,b, respectively. From these figures, the proposed HEVC-FISMVC

codec outperforms the anchor MV-HEVC codec by a factor of 0.32 dB average YUV-PSNR, between 180 kbps and 950 kbps, and 0.45 dB average YUV-PSNR, between 270 kbps and 1580 kbps, for coding 2-view and 3-view video scenario sequences, respectively. To validate and compare the archived objective quality of the proposed single-layered approach coding schemes for stereo and multiview videos with that of the anchor MV-HEVC codec, Bjøntegaard delta-PSNR (BD-PSNR) and Bjøntegaard delta-bitrates (BD-rate) of the decoded 'Poznan Street', ' Kendo' and 'Newspaper1' stereo and multiview videos are used. For these calculations, a piece-wise cubic interpolation, introduced in [36,37] for a five data points-based interpolation polynomial, as recommended in the JCTVC-B055 document [38], is used, and the resulting BD-PSNR and BD-Rate are tabulated in Tables 2 and 3 and Tables 4 and 5, respectively. From Tables 2 and 3 it is evident that the proposed codec delivers superior coding performance in terms of BD-PSNR than the anchor codec for all Y-, U-, and V-components of the stereo and multiview videos. The average BD-PSNR Y, BD-PSNR U, and BD-PSNR V performance of the proposed codec for stereo videos is at 0.225208 dB, 0.048408 dB, and 0.024790 dB respectively higher than that of the anchor MV-HEVC. Similarly, from Table 3 it can be noted that the proposed codec's average BD-PSNR Y, BD-PSNR U, and BD-PSNR V performance for three view scenario multiview videos with respect to the anchor MV-HEVC is 0.248668 dB 0.041242 dB and 0.164715 dB higher, respectively.

Table 2. BD-PSNR of the proposed single-layered stereo video codec with respect to the anchor MV-HEVC codec.

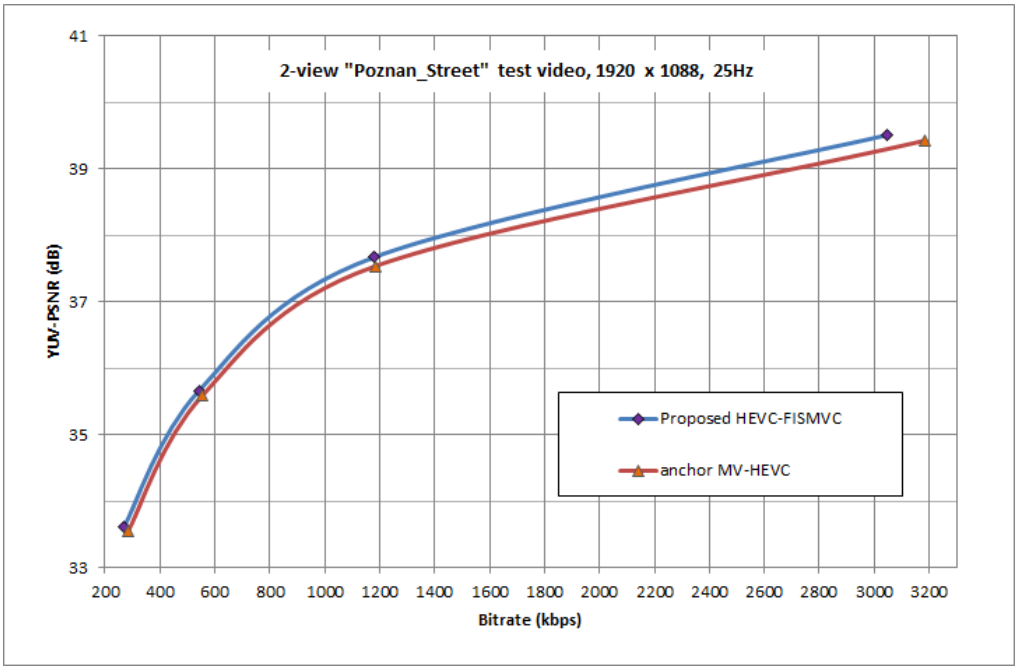
	BD-PSNR (dB)		
Sequence	**BD-PSNR Y**	**BD-PSNR U**	**BD-PSNR V**
Poznan_Street	0.293725	0.092543	0.035271
Kendo	0.165743	0.015274	0.021561
Newspaper1	0.216158	0.037409	0.017538
Average	0.225208	0.048408	0.024790

Table 3. BD-PSNR of the proposed single layered multiview video codec with respect to the anchor MV_HEVC codec.

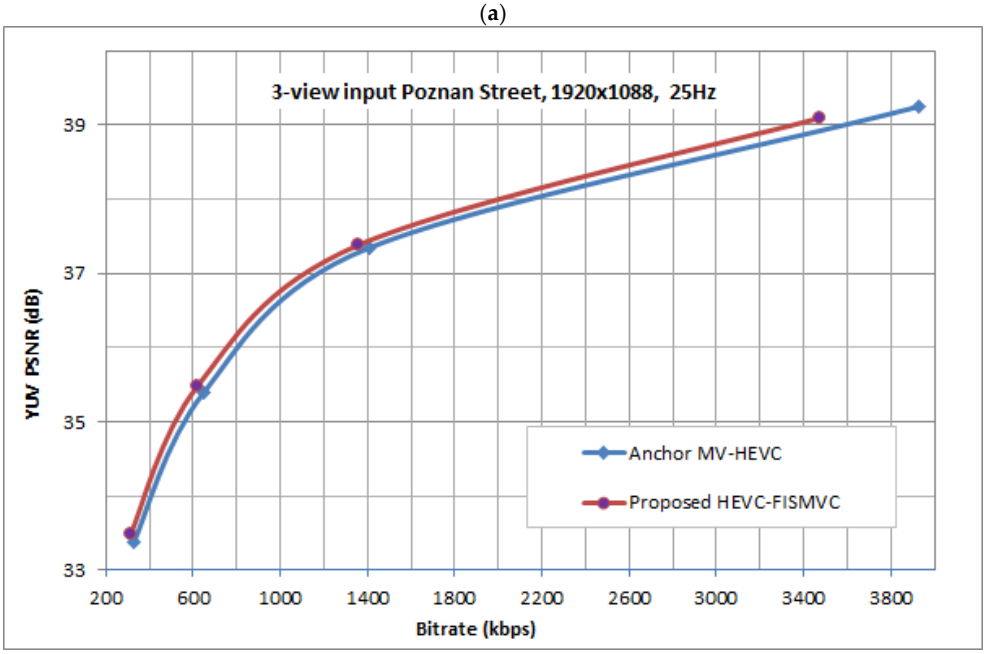
	BD-PSNR (dB)		
Sequence	**BD-PSNR Y**	**BD-PSNR U**	**BD-PSNR V**
Poznan_Street	0.317251	0.071045	0.032501
Kendo	0.154372	0.015274	0.074198
Newspaper1	0.274381	0.037409	0.058016
Average	0.248668	0.041242	0.164715

Table 4. BD-Rate of the proposed single-layered stereo video codec with respect to the anchor MV-HEVC codec.

	BD-Rate (kbps)		
Sequence	**BD-Rate Y**	**BD-Rate U**	**BD-Rate V**
Poznan_Street	−3.72356	−0.76472	−0.43817
Kendo	−2.17652	−0.84175	−1.15233
Newspaper1	−3.65218	−0.85279	−0.97857
Average	−3.184086	−0.819732	−0.856492



(**a**)

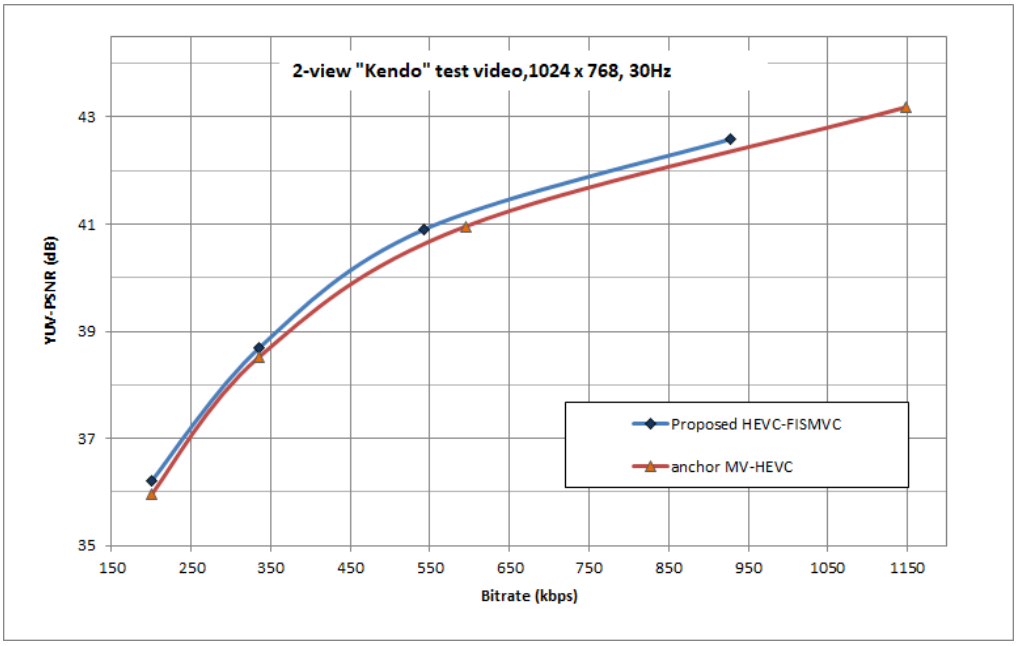


(**b**)

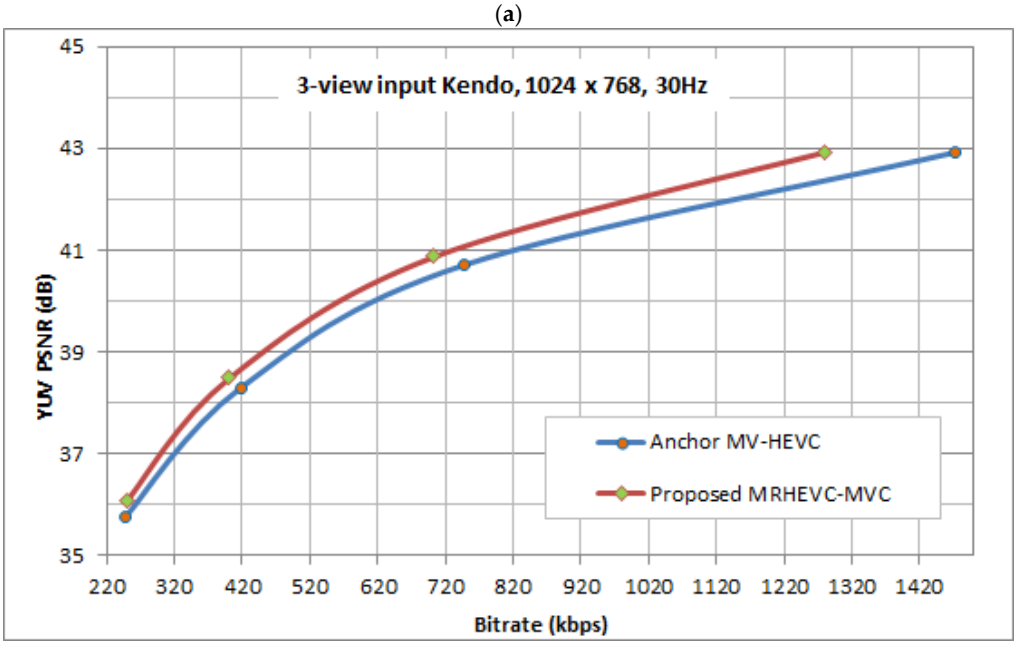
Figure 4. Average YUV-PSNR vs bitrate of the anchor MV-HEVC and the proposed HEVC-FISMVC codecs for coding: (**a**) 2-view "Poznan_street" and (**b**) 3-view "Poznan_street" videos.

Table 5. BD-Rate of the proposed single-layered multiview video codec with respect to the anchor MV-HEVC codec.

	BD-Rate (kbps)		
Sequence	**BD-Rate Y**	**BD-Rate U**	**BD-Rate V**
Poznan_Street	−3.013856	−0.943276	−0.06723
Kendo	−1.854732	−0.087431	−0.075832
Newspaper1	−2.714531	−0.274352	−0.004951
Average	−2.527706	−0.435019	−0.049337



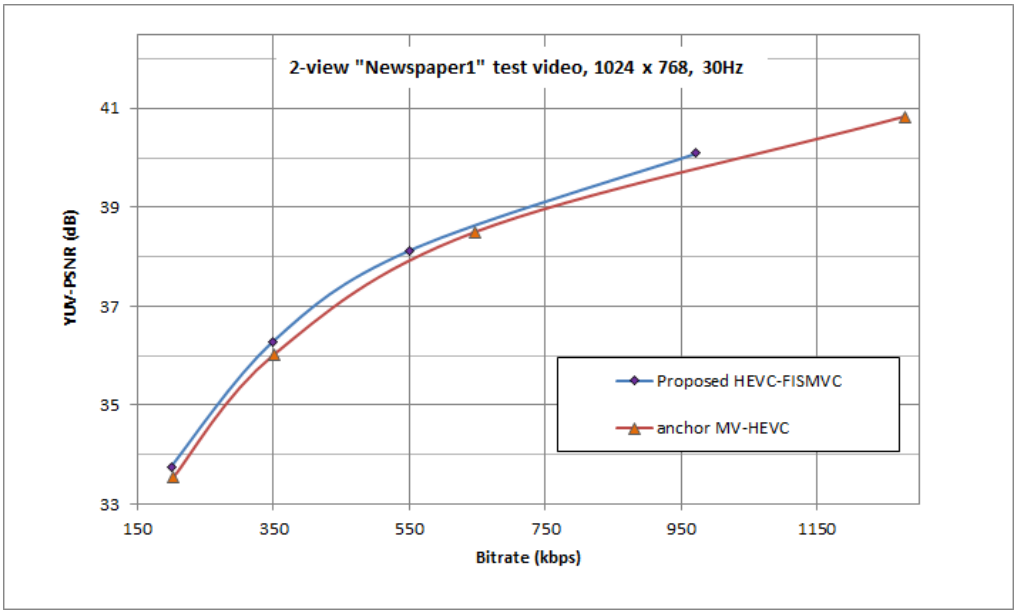
(**a**)



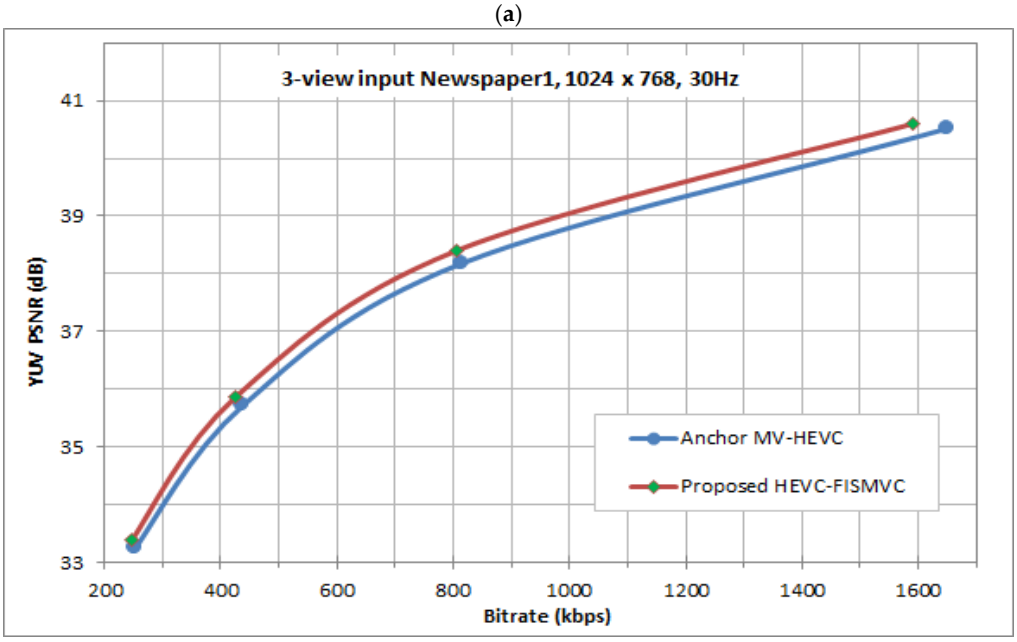
(**b**)

Figure 5. Average YUV-PSNR vs bitrate of the anchor MV-HEVC and the proposed HEVC-FISMVC codecs for coding: (**a**) 2-view "Kendo" multiview videos and (**b**) 3-view "Kendo" multiview videos.

Tables 4 and 5, shows the achieved Bjøntegaard delta-bitrates (BD-rate) for 'Poznan Street', 'Kendo' and 'Newspaper1', stereo and multiview videos of the proposed codec with respect to that of the anchor codec, respectively. From Table 4 it can be seen that the average BD-Rate Y, BD-Rate U, and BD-Rate V of the proposed codec for stereo videos is at 3.184086 kbps, 0.819732 kbps, and 0.856492 kbps lesser than the anchor MV-HEVC. Table 5 shows the BD-Rate Y, BD-Rate U and BD-Rate V of the proposed codec for multiview videos, which are on an average 2.527706 kbps, 0.435019 kbps and 0.049337 kbps lesser than that of anchor MV-HEVC for the same three view scenario multiview videos.



(**a**)

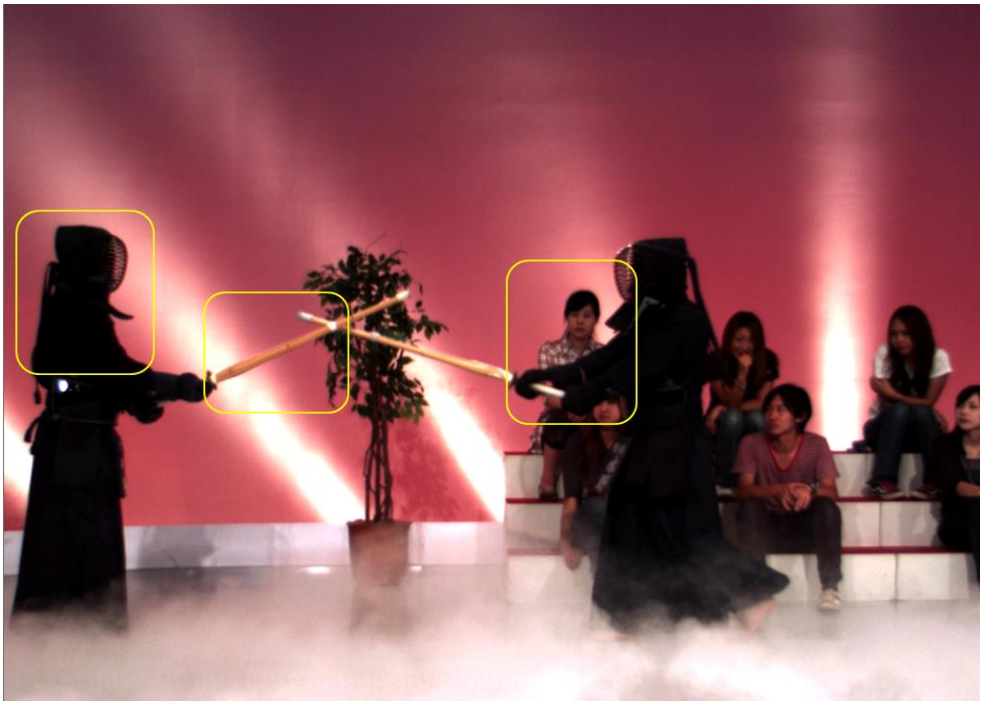


(**b**)

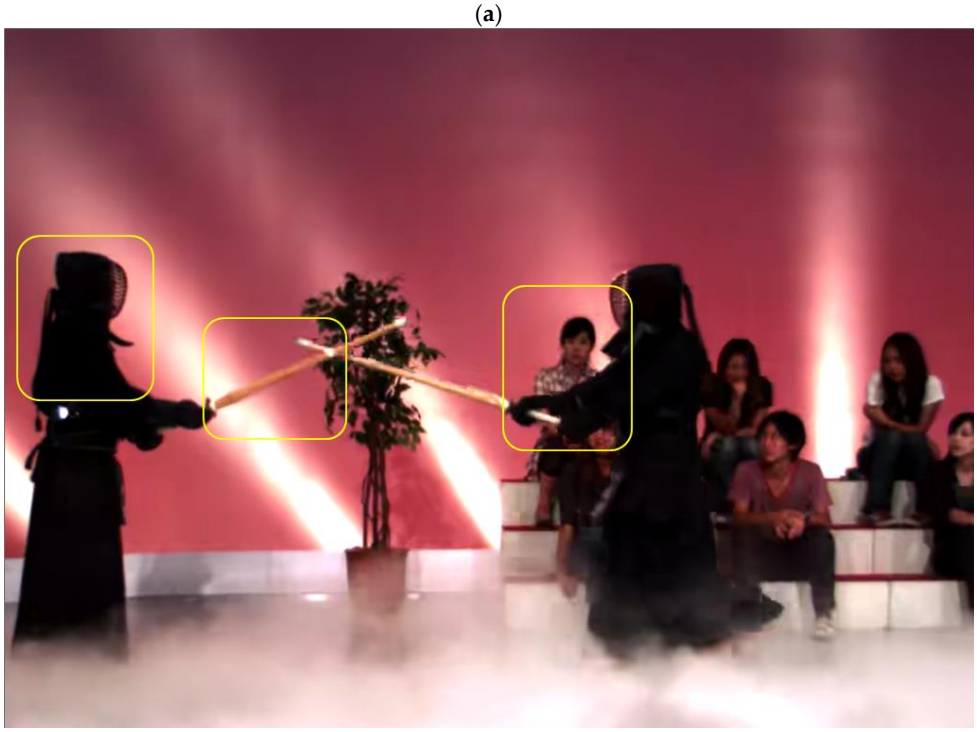
Figure 6. Average YUV-PSNR vs bitrate for the anchor MV-HEVC and the proposed HEVC-FISMVC codecs when coding: (**a**) 2-view "Newspaper1" and (**b**) 3-view "Newspaper1" multiview videos.

To enable readers to compare the achieved visual quality of the proposed codec with that of the anchor, frame 16 of the proposed and anchor MV-HEVC codec, from coding Kendo 3-view scenario multiview videos, were selected and are shown in Figure 7a,b, respectively. The snippets of the highlighted areas of the two methods' video frames are included in Figure 8a–c.

From Figure 7 it can be seen that the anchor codec's video frames exhibit noticeable levels of blocking artefacts and blurred edges in the background and moving objects, better-retained details are noticeable in kendo artist's mask and the mask's grill in Figure 8a, edges of the girl's head and facial details are more obvious in Figure 8b and better edge details in moving object, such as the bamboo kendo in Figure 8c. From these figures, it is evident that the proposed codec's decoded video frames exhibit generally higher visual quality with lesser blocking artefacts and well-retained details than that of MV-HEVC video frame, for the same bitrates.



(**a**)



(**b**)

Figure 7. Decoded frame number 16 from Kendo videos of (**a**) the proposed codec and (**b**) the MV-HEVC standard codec.

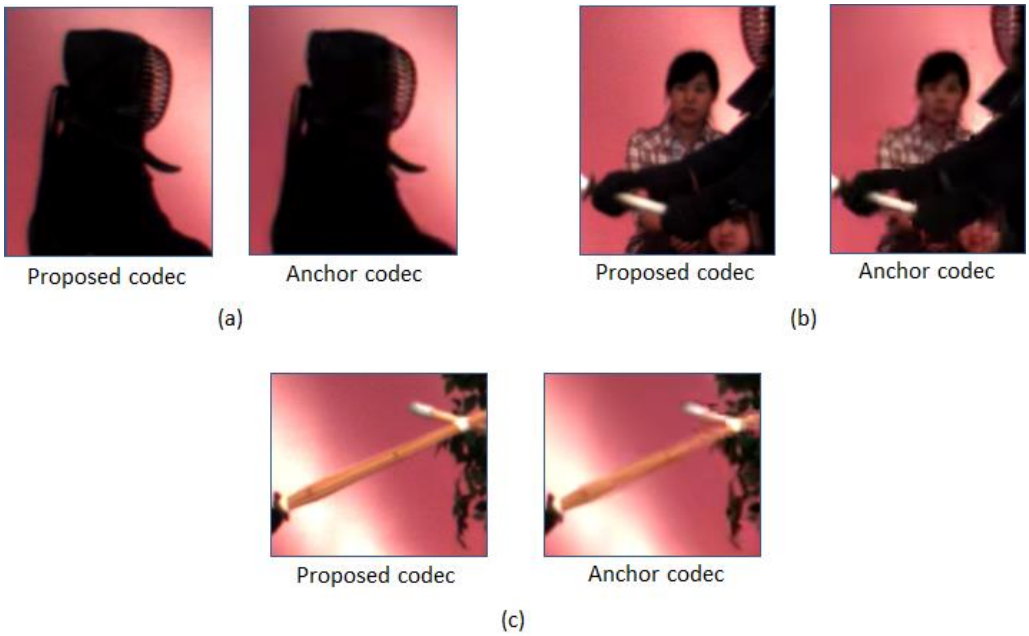


Figure 8. Snippets of the highlighted areas of the proposed and the anchor codecs from Figure 7a,b respectively, (**a**) kendo artist's mask's details (**b**) edges of the girl's head and facial details and (**c**) edge details in moving object.

4. Conclusions

In this paper an HEVC-based frame interleaved video coding technique, for coding stereo and multiview videos was presented. The proposed HEVC-based Frame Interleaved Stereo and Multiview video codec uses a single layered approach to encode frame interleaved multiview videos for two and three view scenarios. The proposed codec exploits correlations across views more efficiently than the anchor MV-HEVC codec, resulting in higher coding gains. The coding performance of the proposed codec was compared with that of the standard MV-HEVC video codec using three standard multiview video sequences at different QPs. The experimental results for both two view and three view scenarios have shown that the proposed codec outperforms the anchor MV-HEVC codec. In addition, the encoded single layer video bit stream mitigates the drawback of accessing multi-layer video frames, enabling fast video streaming across different views.

Author Contributions: Conceptualization, B.M. and A.S.-A.; methodology, B.M. and A.S.-A.; software, B.M.; validation, B.M. and A.S.-A.; formal analysis, B.M. and A.S.-A.; investigation, B.M.; resources, A.S.-A., P.B.Z. and S.A.-M.; data curation, B.M.; writing—original draft preparation, B.M. and A.S.-A.; writing—review and editing, B.M., A.S.-A. and S.A.-M.; visualization, B.M.; supervision, A.S.-A., P.B.Z. and S.A.-M.; project administration, A.S.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, Y.; Wang, Y.K.; Ugur, K.; Hannuksela, M.M.; Lainema, J.; Gabbouj, M. The emerging MVC standard for 3D video services. *EURASIP J. Adv. Signal Process.* **2008**, *2009*, 1–13. [CrossRef]
2. Urey, H.; Chellappan, K.V.; Erden, E.; Surman, P. State of the Art in Stereoscopic and Autostereoscopic Displays. *Proc. IEEE* **2011**, *99*, 540–555. [CrossRef]
3. Merkle, P.; Müller, K.; Wiegand, T. 3D video: Acquisition, coding, and display. *IEEE Trans. Consum. Electron.* **2010**, *56*, 946–950. [CrossRef]
4. Perkins, M.G. Data compression of stereopairs. *IEEE Trans. Commun.* **1992**, *40*, 684–696. [CrossRef]

5. Vetro, A.; Wiegand, T.; Sullivan, G.J. Overview of the stereo and multiview video coding extensions of the H. 264/MPEG-4 AVC standard. *Proc. IEEE* **2011**, *99*, 626–642. [CrossRef]
6. Wiegand, T.; Sullivan, G.J.; Bjontegaard, G.; Luthra, A. Overview of the H. 264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Technol.* **2003**, *13*, 560–576. [CrossRef]
7. 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). 2007. Available online: https://www.iso.org/standard/75400.html (accessed on 15 July 2020).
8. "High Efficiency Video Coding," ITU-T Recommendation H.265 and ISO/IEC 23008-2, April 2013 (and subsequent editions). Available online: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-H.265-201304-S!!PDF-E&type=items (accessed on 15 July 2020).
9. Hurst, W.; Withington, A.; Kolivand, H. Virtual conference design: Features and obstacles. *Multimed. Tools Appl.* **2022**, *81*, 16901–16919. [CrossRef] [PubMed]
10. Tang, D.; Zhang, L. Audio and Video Mixing Method to Enhance WebRTC. *IEEE Access* **2020**, *8*, 67228–67241. [CrossRef]
11. Gandam, A.; Sidhu, J.-S. Fuzzy Based Adaptive Deblocking Filters at Low-Bitrate HEVC Videos for Communication Networks. *J. Comput. Mater. Contin.* **2021**, *66*, 3045–3063. [CrossRef]
12. Malekzadeh, M. Perceptual service-level QoE and network-level QoS control model for mobile video transmission. *Telecommun. Syst.* **2021**, *77*, 523–541. [CrossRef]
13. Luo, Y.; Song, L.; Xie, R.; Luo, C. View-Dependent Omnidirectional Video Encapsulation Using Multiple Tracks. In Proceedings of the 2017 International Conference on Virtual Reality and Visualization (ICVRV), Zhengzhou, China, 21–22 October 2017; pp. 421–422.
14. Lee, J.; Kim, S.-H.; Jeong, S.Y.; Choi, J.S.; Kang, D.-W.; Jung, K.-H.; Kim, J. A Stereoscopic 3-D Broadcasting System Using Fixed and Mobile Hybrid Delivery and the Quality Assessment of the Mixed Resolution Stereoscopic Video. *IEEE Trans. Broadcast.* **2015**, *61*, 222–237. [CrossRef]
15. Joachimiak, M.; Hannuksela, M.; Gabbouj, M. View synthesis quality mapping for depth-based super resolution on mixed resolution 3D video. In Proceedings of the 2014 3DTV-Conference: The True Vision—Capture, Transmission and Display of 3D Video (3DTV-CON), Budapest, Hungary, 2–4 July 2014; pp. 1–4.
16. Mallik, B.; Sheikh-Akbari, A.; Kor, A.-L. HEVC Based Mixed-Resolution Stereo Video Codec. *IEEE Access* **2018**, *6*, 52691–52702. [CrossRef]
17. Mallik, B.; Sheikh-Akbari, A. HEVC Based Multi-view Video Codec Using Frame Interleaving Technique. In Proceedings of the 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, UK, 31 August–2 September 2016; pp. 181–185.
18. Mallik, B.; Sheikh-Akbari, A.; Bagheri-Zadeh, P. HEVC based stereo video codec. In Proceedings of the 2nd IET International Conference on Intelligent Signal Processing 2015 (ISP), London, UK, 1–2 December 2015; pp. 1–6.
19. Merkle, P.; Smolic, A.; Muller, K.; Wiegand, T. Efficient prediction structures for multiview video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2007**, *17*, 1461–1473. [CrossRef]
20. Kim, Y.; Kim, J.; Sohn, K. Fast Disparity and Motion Estimation for Multi-view Video Coding. *IEEE Trans. Consum. Electron.* **2007**, *53*, 712–719. [CrossRef]
21. Shen, L.; Liu, Z.; Liu, S.; Zhang, Z.; An, P. Selective Disparity Estimation and Variable Size Motion Estimation Based on Motion Homogeneity for Multi-View Coding. *IEEE Trans. Broadcast.* **2009**, *55*, 761–766. [CrossRef]
22. Li, S.; Hou, C.; Ying, Y.; Song, X.; Yang, L. Stereoscopic video compression based on H. 264 MVC. In Proceedings of the 2nd International IEEE Congress on Image and Signal Processing, CISP 2009, Tianjin, China, 17–19 October 2009; pp. 1–5.
23. Hewage, C.T.E.R.; Karim, H.A.; Worrall, S.; Dogan, S.; Kondoz, A.M. Comparison of stereo video coding support in MPEG-4 MAC, H.264/AVC and H. 264/SVC. In Proceedings of the IET Visual Information Engineering-VIE07, London, UK, 25–27 July 2007.
24. Gürler, C.G.; Bağci, K.T.; Tekalp, A.M. Adaptive stereoscopic 3D video streaming. In Proceedings of the 17th IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 2409–2412.
25. Saygili, G.; Gurler, C.G.; Tekalp, A.M. Evaluation of Asymmetric Stereo Video Coding and Rate Scaling for Adaptive 3D Video Streaming. *IEEE Trans. Broadcast.* **2011**, *57*, 593–601. [CrossRef]
26. Sansli, D.B.; Ugur, K.; Hannuksela, M.M.; Gabbouj, M. Interview motion vector prediction in multiview HEVC. In Proceedings of the 3DTV-Conference: The True Vision Capture, Transmission and Display of 3D Video, 2014. 3DTV-CON, Budapest, Hungary, 2–4 July 2014; pp. 1–4.
27. Van Wallendael, G.; van Leuven, S.; de Cock, J.; Bruls, F.; van de Walle, R. 3D video compression based on high efficiency video coding. *IEEE Trans. Consum. Electron.* **2012**, *58*, 137–145. [CrossRef]
28. Stankowski, J.; Domanski, M.; Stankiewicz, O.; Konieczny, J.; Siast, J.; Wegner, K. Extensions of the HEVC technology for efficient multiview video coding. In Proceedings of the 19th IEEE International Conference on Image Processing, ICIP 2012, Orlando, FL, USA, 3 September–3 October 2012; pp. 225–228.
29. Sullivan, G.J.; Ohm, J.R.; Han, W.J.; Wiegand, T. Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 1649–1668. [CrossRef]

30. Sullivan, G.J.; Boyce, J.M.; Chen, Y.; Ohm, J.-R.; Segall, C.A.; Vetro, A. Standardized Extensions of High Efficiency Video Coding (HEVC). *IEEE J. Sel. Top. Signal Process.* **2013**, *7*, 1001–1016. [CrossRef]
31. Paramkusam, A.V.; Reddy, V.S.K. An efficient multi-layer reference frame motion estimation for video coding. *J. Real-Time Image Process.* **2014**, *11*, 645–661. [CrossRef]
32. Bouyagoub, S.; Akbari, A.S.; Bull, D.; Canagarajah, N. Impact of camera separation on performance of H. 264/AVC-based stereoscopic video codec. *IET Electron. Lett.* **2010**, *46*, 345–346. [CrossRef]
33. Sheikh-Akbari, A.; Said, H.; Moniri, M. Effect of inter-camera angles on the performance of an H. 264/AVC based multi-view video codec. In Proceedings of the 2012 Picture Coding Symposium, Krakow, Poland, 7–9 May 2012; pp. 109–112.
34. Muller, K.; Vetro, A. "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. 2014, pp. 1–7. Available online: https://www.itu.int/wftp3/av-arch/jct3v-site/2014_03_H_Valencia/JCT3V-H_Notes_d7.doc (accessed on 10 June 2020).
35. Schwarz, H.; Schierl, T.; Marpe, D. Block Structure and Parallelism. In *High Efficiency Video Coding (HEVC): Algorithms and Architectures, Integrated Circuit and Systems*; Springer: New York, NY, USA, 2014; Chapter 3; pp. 49–90.
36. Bjontegaard, G. Calculation of Average PSNR Differences between RD Curves. In ITU-T SG 16, VCEG-M33. 2001, pp. 1–4. Available online: https://www.itu.int/wftp3/av-arch/video-site/0104_Aus/VCEG-M33.doc (accessed on 16 November 2022).
37. Bjontegaard, G. Improvements of the BD-PSNR Model. In ITUT SG 16, VCEG-AI11. 2008, pp. 1–2. Available online: https://www.itu.int/wftp3/av-arch/video-site/1707_Tor/VCEG-BD04-v1.doc (accessed on 16 November 2022).
38. Senzaki, K. BD-PSNR/Rate Computation Tool for Five Data Points. In ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCTVC-B055. 2010, pp. 1–3. Available online: https://www.itu.int/wftp3/av-arch/JCTVC-site/2010_07_B_Geneva/JCTVC-B055.doc (accessed on 16 November 2020).