Received July 31, 2018, accepted September 4, 2018, date of publication September 19, 2018, date of current version October 12, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2870183

HEVC Based Mixed-Resolution Stereo Video Codec

BRUHANTH MALLIK, AKBAR SHEIKH-AKBARI, AND AH-LIAN KOR

School of Computing, Creative Technology & Engineering, Leeds Beckett University, Leeds LS6 3QS, U.K.

Corresponding author: Bruhanth Mallik (b.mallik6347@student.leedsbeckett.ac.uk)

ABSTRACT This paper presents a high efficiency video codec-based spatial mixed-resolution stereo video codec. The proposed codec applies a frame interleaving algorithm to reorder the stereo video frames into a single stream of monoscopic video. The challenge for mixed-resolution video coding is to enable the codec to encode frames with different frame resolutions. This issue is addressed by superimposing a low-resolution replica of the decoded I-frame on its respective decoded picture, where remaining space of the frame is set to zero. This significantly reduces the computation cost for finding the best match. The proposed codec's reference frames structure is designed to efficiently exploit both temporal and inter-view correlations. The performance of the proposed codec is assessed using five standard multiview video data sets and benchmarked against that of the anchor and the state-of-the-art techniques. Results show that the proposed codec yields significantly higher coding performance compared with the anchor and state-of-the-art techniques.

INDEX TERMS HEVC, mixed-resolution video codec, stereo video coding, low bitrate transmission.

I. INTRODUCTION

Over the last decade, numerous 3D video technologies, such as: holography, stereo video systems, video plus depth and multiview video codecs, have been developed. Stereo and multiview video codecs have evolved over time and are used in a range of applications such as: 3D surveillance, remote vehicle navigation, robotics and automation, e-learning systems and 3D machine-vision applications for object localization, identification and measurement [1], [2]. With the growing popularity of 3D video applications, ranging from low resolution to ultra-high definition (UHD) video content, demand for their efficient compression has drastically increased. According to Cisco's Visual Networking Index (VNI) forecast for 2016-2021, it is anticipated that by 2021, 73% of the overall IP traffic will be used by both business and domestic users would be from non-PC devices. It is also anticipated that, 82% of the total IP traffic would be used for visual content communication [3]. In order to meet the growing video coding demands in various fields of applications, many standard as well as non-standard stereo and Multiview Video Coding (MVC) techniques, e.g. H.264/MPEG-4 AVC, MVC and MV-HEVC codecs, have been released over the years [4]–[6]. Video coding standardization organizations, namely the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG), jointly published High Efficiency Video Coding (HEVC) standard in 2013 to address diverse video transmission applications' requirements [7]. A standardized extension of HEVC for multiview view videos, called MV-HEVC, was published to encode texture based stereo and multiview videos in 2014 [6]. In addition to exploiting spatial and temporal correlations within video sequences, stereo and multiview video codecs extensively utilize inter-view correlations to efficiently compress the huge amounts of visual information within views [8], [9]. These coding techniques primarily use advanced disparity or motion prediction and compensation techniques in a multi-layered architecture to exploit correlations between temporal frames and neighboring views. However, the amount of buffer memory that MVC requires for additional decoded pictures due to its multi-layered coding architecture, restricts the number of views that can be used to perform efficient disparity prediction and compensation [10]. In addition, it is obligatory for MVC based coding techniques to use larger quantization steps for encoding stereo videos in order to meet the requirements of low bitrate transmission. This results in distortions and blocking artefacts in its decoded frames, due to the loss of high frequency information. An insight into the low bitrate HEVC and MV-HEVC's bitstreams has revealed that slice data overhead, due to the large frame size, consumes more bits than required to send the actual visual information of the video frames [6], [7], [11]. In order to address these shortcomings of the current stereo video codecs, a single layered asymmetric spatial resolution video codec for stereo videos based on HEVC codec is presented in this paper. The remainder of this paper is organized as follows: Section II gives an overview of asymmetric spatial resolution video codecs and a review on the state-of-the-art stereo video coding techniques for low bitrate transmission. In section III, the mixed-resolution stereo videos frame interleaving algorithm, up- and down-sampling methods for the intermediate frames and the design of the proposed codec are presented. Experimental results are analyzed in Section IV and finally, the paper is concluded in Section V.

II. RELATED WORK

A mixed-resolution based stereo video coding method was first reported in [8], which reduces the number of bits required to represent the stereo video sequences. Mixed-resolution based stereo video coding methods, which are sometimes referred to as asymmetric spatial resolution video codecs, were reported in [8] and [12]. These codecs deliver a subjectively acceptable 3D viewing experience, because the human visual system is less sensitive to high frequency components of the video frames [8], [9]. Subjective quality assessment of decoded frames of these codecs has confirmed that the asymmetric spatial stereo video codecs' video frames do not exhibit any statistical change with regard to eye dominance of the subject [13], [14]. In another subjective evaluations, related to the quality of both spatially and temporally down-sampled stereo video frames, have shown that the spatially down-sampled stereoscopic videos are more favored for low bitrate transmission [15], [16]. Many algorithms for asymmetric mixed-resolution stereoscopic video coding have been developed in recent years [17]–[22], [24]–[28]. It was shown in [17] and [18] that a combination of frame down-sampling and inter-view prediction, when employed in asymmetric stereo video codecs, produces higher coding performance than that of the symmetric codecs for low bitrates transmission. Disparity or motion prediction and compensation of a down-sampled frame from a full resolution frame has been investigated in [19] and [20], and the findings of these investigations show that this prediction results in a higher coding performance. An adaptive spatial mixed-resolution coding technique using MV-HEVC was proposed in [21]. This codec selects a down sampling factor that yields highest coding performance by comparing the frequency power spectrums of adjacent views using various quality metrics. Authors reported 3.38% bitrate reduction over MV-HEVC codec for coding five different camera-captured standard stereo video sequences.

A perceptually driven Non-uniform Asymmetric Stereoscopic Video Codec (NASVC) was reported in [22]. This coding technique splits the input videos into a set of significant and manageable video segments, which have similar contents, using shot boundary detection algorithm [23]. The right view frames are encoded in full resolution and are used as the reference frames. It filters and down-samples left view frames using a blur filter. The blur-filter is a circular kernel filter and its parameters for each segment are calculated by taking into account disparity map of that segment's key-frame and perceptual weights. Authors reported an average of 1.13 dB higher coding performance by the NASVC method compared to MV-HEVC codec. However, no significant coding performance at low bitrates was reported. A mixed-resolution stereoscopic video coding technique using HEVC, called Symmetric Mixed-Resolution Stereoscopic Video Codec (SMRSVC), was introduced in [24], to reduce the interview frame quality-imbalances to less than 2dB. The frame prediction structure of the SMRSVC codec shows that the left view frames are coded independent of the right view frames, with the first frame in each Group of Picture (GOP) of each view coded at its full resolution. Though this study shows that the proposed codec reduces the quality-imbalance to less than 2 dB, the inter-view frames correlations are not fully exploited, as the first frames of both views are encoded in full resolution. The full capability of the mixed-resolution stereo video coding technique is not efficiently utilized. The coding performance of SMRSVC codec shows comparable results to that of the MV-HEVC codec, when coding stereo videos.

An asymmetric stereo video coding model based on Binocular Just-Noticeable-Difference (BJND), which harnesses the human binocular vision's properties, to determine the minimum distortion in one view that induce binocularly visible differences was introduced in [25]. In this BJND model, the perceptible distortion threshold of stereoscopic images are measured for binocular vision to create the luminance mask, which is subsequently integrated with the contrast mask. A Just-Noticeable-Difference (JND) based asymmetric stereo video codec was proposed in [26], which discards chrominance channels, while controlling the luminance quality of the right view frames by applying thresholds at just above the noticeable distortion level. A binocular-combination-oriented (BCO), perceptual based, stereoscopic 3D rate distortion optimization (BCO-stereo-RDO) coding technique was reported in [27], which uses difference of Gaussian model in frequency domain of video frames to quantitatively characterize binocular 3D perceptual distortion between the stereo views. The BCO-stereo-RDO video coding technique mainly focuses on saving bitrate by optimizing the perceived stereoscopic 3D video quality. Hence, this codec is unable to achieve significant bitrate saving compared to the standard codec. An asymmetric stereo video coding technique, using a disparity JND model, was proposed in [28]. This codec combines disparity JND model with spatio-temporal JND, in addition to luminance and texture masking. In this technique, the disparity JND model is obtained from disparity masks of the human visual system (HVS). The experimental results demonstrate that their coding technique effectively removes the perceptual redundancy within the stereo videos. Although their experimental results show that the codec outperforms the MV-HEVC codec at higher bitrates, its coding performance deteriorates as bitrate decreases.

Although, BJND and JND based asymmetric stereo video coding methods can save 6.6% to 34% of the bitrate, compared to the standard codecs, a subjective evaluation of the coded videos at low bitrates show that their videos do not fulfil the minimum quality requirement [29], [30]. This is due to the use of larger quantization parameters by BJND and JND based asymmetric stereo video coding methods. On the other hand, stereo video coding techniques based on the standard multiview extensions (in the likes of MVC or MV-HEVC), such as [21], [22] and [24], extensively rely on inter-view prediction/compensation for asymmetric stereo video coding. Mallik *et al.* [31] previous study on coding frame interleaved monoscopic stereo videos, has shown a reduced decoding complexity and signaling overhead, compared to multi-layered stereo video coding approaches (MVC and MV-HEVC). As the stereo video frames are temporally interleaved, the proposed codec's reference frame structure allows cross-frame (also called as lateral frame) referencing [5], which is neither supported by the standard MV-HEVC codec nor by the state-of-the-art codecs [21], [22], [24].

In this paper, a HEVC texture based mixed spatial resolution codec, which was initially reported in [32], that has been extended to code stereo videos is presented. The proposed codec applies an interleaving algorithm and resolution downsampling technique on the selected video frames to generate a single stream video sequence of the same size [33]. Each resulting low resolution frame is superimposed on the top left quadrant of a zero pixel value frame as shown in Figure 3. The standard HEVC codec is amended and configured to code frame interleaved mixed spatial resolution stereo video frames. Experimental results using five benchmark multiview video datasets show that the proposed codec generates significantly higher coding performance than the anchor MV-HEVC codec as well as other state-of-the-art mixed-resolution stereo video codecs. A comprehensive description of the proposed HEVC based Mixed-Resolution Stereo Video codec, detailing modifications to the standard HEVC codec, is presented in Section III.

III. PROPOSED HEVC BASED MIXED-RESOLUTION STEREO VIDEO CODEC

Stereo videos involves the use of two identical cameras with parallel or converging axes, where video pairs are simultaneously acquired. Stereo video codecs facilitate efficient video compression for transmission and saves storage space. The proposed HEVC based Mixed-Resolution Stereo Video Coding (HEVC-MRSVC) scheme is developed within the standard HEVC codec's software framework [11]. The stereo video codec proposed in this research aims to attain high compression ratio while maintaining high stereo video quality for low bitrate transmission. The frame architecture of the proposed HEVC-MRSVC codec is shown in Figure 1. This figure depicts the encoding of the selected key frames at

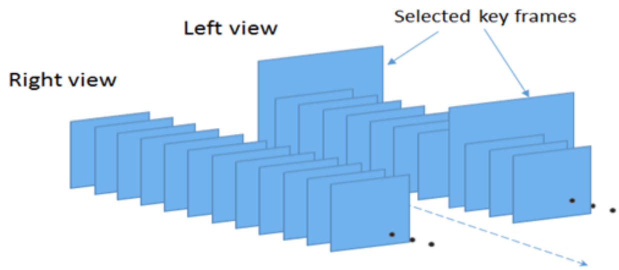


FIGURE 1. Frame architecture of the proposed mixed-resolution stereo video codec.

TABLE 1. Blackman 2D FIR filter coefficients.

0.0381	0.1051	0.0381
0.1051	0.4273	0.1051
0.0381	0.1051	0.0381

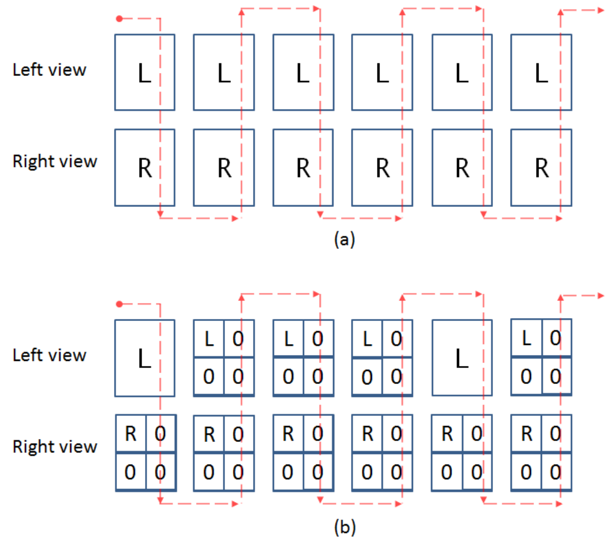


FIGURE 2. Frame interleaving algorithm's contour to reorder stereo frames (a) full resolution input video frames and (b) their respective mix-resolution frames.

their original resolution, whereas the intermediate (non-key) frames are filtered and down-sampled to reduce their resolution and then encoded. From Figure 1, it can also be noted that the key frames are always selected from the left view frames. The proposed HEVC-MRSVC codec encodes the key frames at their full resolution as I-frames, whereas the low resolution intermediate frames are encoded as P- or B-frames. The frame interleaving algorithm, for stereo videos [31], enables the proposed codec to rearrange the mixed-resolution stereo video frames of each GOP, such that two temporal consecutive frames of each view are always adjacent to each other, as shown in Figure2. The resulting single stream video sequence, is represented by the dotted arrows in Figure 2(a) and Figure 2(b) shows the intermediate frames low resolution that are superimposed onto zero values frames. The HEVC standard video coding scheme, JCT-VC HEVC

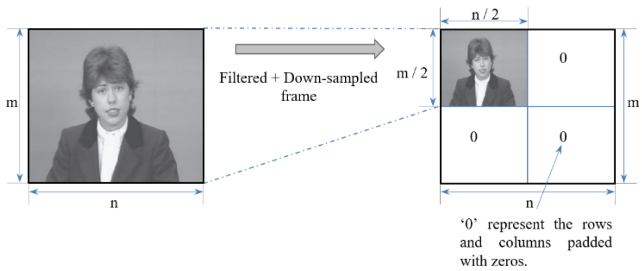


FIGURE 3. Intermediate input B- or P-frame and its superimposed low resolution frame.

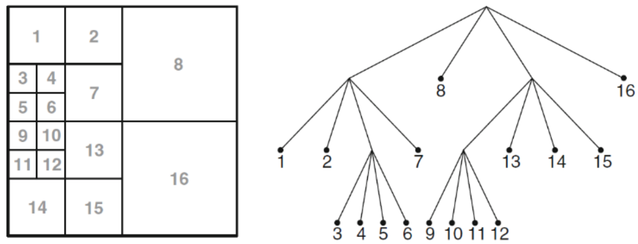


FIGURE 4. An example of HEVC's 64X64 luma CTU partitioned into CUs of 8x8 to 32X32 luma samples, with the numbers representing the coding order of the CUs [7].

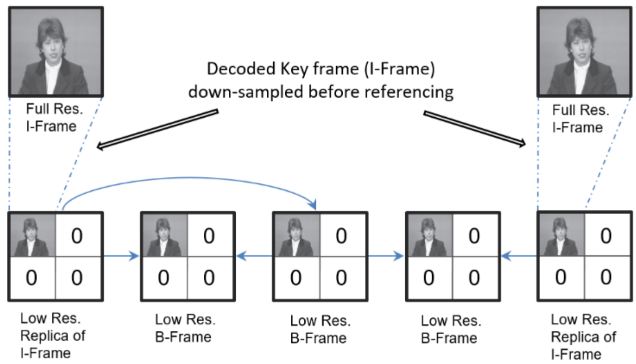


FIGURE 5. Frame referencing architecture of the proposed codec.

version HM16.12 software, has been amended and configured to code the frame interleaved mixed-resolution stereo video frames. The full resolution original size intermediate P- and B-frames are first smoothened using a Blackman 2D FIR low-pass filter to diminish aliasing artefacts and phase shift due to down-sampling [34]. The Blackman filter coefficients are tabulated in Table 1. The Blackman 2D FIR filter has shown to deliver perceptually higher visual quality, when it is used to generate low resolution images compared to other smoothing methods (e.g. Hanning or Hamming windows). A filter size of 3×3 was found to be perceptually effective while not inducing significant computation cost to the codec. The filtered P- and B-frames are both horizontally and vertically down-sampled by a factor of two, to maintain the aspect ratio of the full resolution frames [33]. The resulting low resolution P- and B-frames are overlaid on the top left quadrant of their full resolution frames, where their remaining pixels are set to zeros, as illustrated in Figure 3, [33]. Since HEVC codec encodes spatial uniform areas with larger Coding Tree Units (CTU), the three quadrants' zero value pixels are encoded with minimum signaling bits. Therefore, in the proposed HEVC-MRSVC code's design, the intermediate and the key frames have the same frame size, though their frame information are at different resolutions. However, this poses another challenge, which is the inherent challenge in designing any mixed-resolution video codecs, where motion or disparity estimation and compensation must be performed between frames at different resolutions. When implementing the codec in HEVC's framework, this challenge becomes even more complex as the codec splits each frame into its coding tree units (CTU). Each CTU represents one luma and two chroma coding tree blocks (CTB), with a syntax associated with the CTBs. The CTUs are further partitioned into coding unit (CU), where the size of the CUs have frame data dependency [7], [11], and the CU subdivisions are specified in the CTU through their respective quad tree syntax. Similar to CTUs, CUs also contain luma and chroma details with their related syntax. The CUs in a CTU are coded in a z-scan order, as shown in Figure 4 [7]. To address this issue, the proposed codec saves a copy of the decoded I-frame locally, in its full resolution, along with its status flag and CTU details (from the slice header) outside the Decoded Picture Buffer (DPB). In HEVC motion prediction, intra-frame prediction and in-loop deblocking filtering are all based on CTB. The HEVC reconstruction function, which is available to deblocking filter, is used in conjunction with DPB in the proposed codec to generate a low resolution replica of the I-frame, as shown in Figure 3. The resulting low resolution replica of the I-frame contains the superimposed low resolution I-frame at the top left quadrant of the zero value frame. In HEVC, group of CTUs of a picture are stuck together to form one or more slices. Encoded slice segments contains their specific slice segment headers, followed by the slice segment data. All the control and signaling information of the CTUs within a slice segment of the codec are kept in the header of the slice segment. The slice header and the first slice segment header of the codec contain the Picture Order Count (POC), slice type and picture output flag details along with other control information. Therefore, in the proposed HEVC-MRSVC codec the coding tree algorithm splits the resulting low resolution replica of each I-frame into CTUs and updates its I-Slice header, accordingly. In addition, the full resolution frame I-frame will be replaced by its low resolution replica I-frame in DPB memory buffer. The resulting frame referencing architecture of the proposed HEVC-MRSVC coding scheme is illustrated in Figure 5. From Figure 5, it can be noted that all video frames including the I-frames, have the same frame size, which enables the standard HEVC codec's algorithm to perform motion or disparity estimation and compensation. A flow chart of the algorithm for generating a low resolution replica of the I-frame is shown in Figure 6. The encoded frame is reconstructed and saved in the DPB, as long as it is required, for referencing. In the proposed

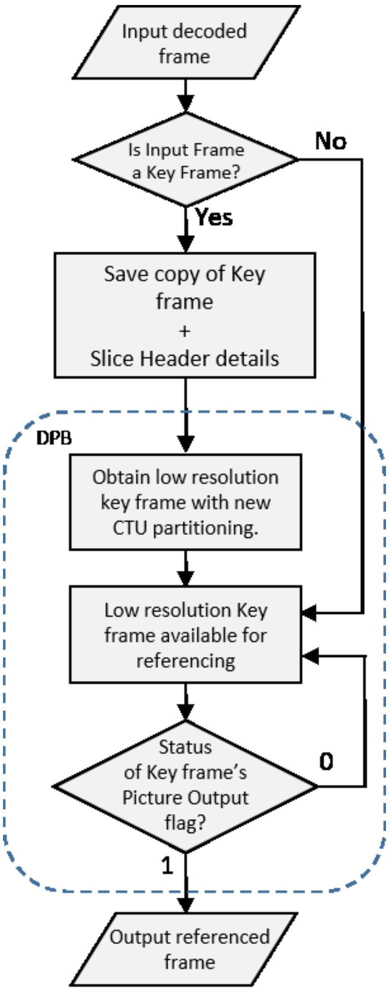


FIGURE 6. Flow chart of Key frame down-sampling function.

codec, the header of the reconstructed frame to be saved in the DPB is checked for key frame (which is encoded as an I-frame), if it is an I-frame, it saves a copy of the I-frame and its slice header details in a local memory outside the DPB. It then generates a low resolution replica of this I-frame and saves it in its respective location in the DPB and makes it available for motion or disparity estimation and compensation. The low resolution replica of the I-frame will remain in the DPB until it is needed for referencing (this is determined by the reference frame architecture). As the stereo video frames are interleaved, the proposed HEVC-MRSVC codec's reference frame structure allows cross-frame (also called as lateral frame) referencing [5], in addition to temporal and inter-view referencing, as demonstrated in Figure 7. The GOP size and intra frame period for the presented results are set to 16 and 48 frames, respectively. On the decoder end, the decoded low resolution video frames are enlarged using the Bi-cubic interpolation method to recover the original video frame size, as displayed in Figure 8. The decoded intermediate low resolution frames are enlarged and de-interleaved to generate the decoded full resolution stereo video sequences [33].

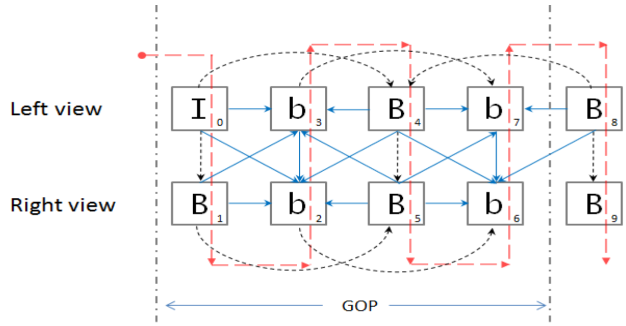


FIGURE 7. Reference frame structure of the proposed mixed resolution stereo video codec.

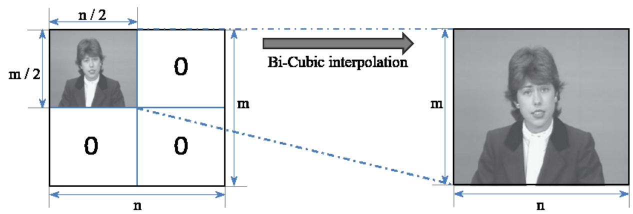


FIGURE 8. Enlarged decoded intermediate B- or P-frame.

The main purpose of disparity or motion prediction and compensation is to lessen the energy of the difference block by finding the block contents in the frame of its neighboring view or the previous frame of the same camera. Investigation has shown that disparity estimation and compensation search area is dependent on inter-cameras angle and the distance of the camera from the scene [35]. In this research, the search area is set to 96 for motion or disparity estimation and compensation, to mitigate the effect of multiview imaging system on coding performance of the codec.

IV. EXPERIMENTAL RESULTS

To assess the coding performance of the proposed HEVC based mixed-resolution stereo video codec (HEVC-MRSVC) [33], five standard multiview video datasets (of 4:2:0 format with 8 bit pixel resolution) called: "Poznan_Street", "Undo_Dancer", "Newspaper1", "Balloons" and "Kendo" were selected and tested [36]. These datasets represent scenes illuminated by various light sources with different scene features, covering a range of entertainment and interactive applications captured by different multiview imaging systems. To provide the reader an overview of these datasets, the first frame of the left view of each of these videos are shown in Figure 9. These datasets include scenes with both stationary and non-stationary backgrounds exposed to different levels of illumination. The datasets "Kendo", "Balloons" and "Newspaper1" are recorded at 1024×768 resolutions with 30 frames per second (30 Hz), while "Poznan_Street" and "Undo_Dancer" have 1920×1088 resolutions with 25 frames per second (25 Hz). To assess and compare the coding competency of the proposed HEVC-MRSVC codec against the anchor MV-HEVC

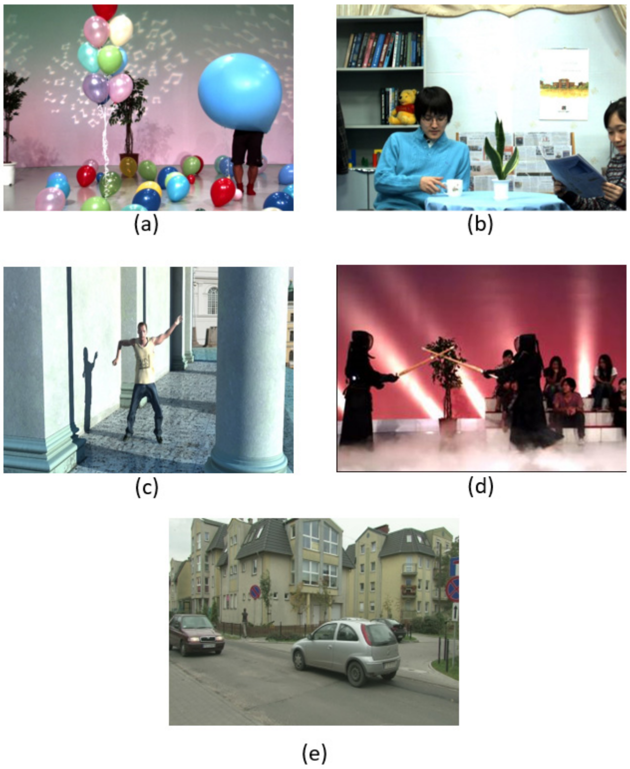


FIGURE 9. First frame of view 1 of standard multiview dataset: (a) "Balloons", (b) "Newspaper1", (c) "Undo_Dancer", (d) "Kendo", (e) "Poznan_Street".

codec, views 5-4, 1-5, 2-4, 1-3, and 1-3, of the "Poznan_Street", "Undo_Dancer", "Newspaper1", "Balloons" and "Kendo, benchmark multiview videos were respectively chosen to create five pairs of stereo video datasets and used in the analysis. The standard MV-HEVC codec's performance results, which are used as the anchor in the analysis, for the chosen test stereo video views, are freely accessible in JCT3V-G1100 Common Test Condition (CTC) documentation [36]. Therefore, same views of these test stereo videos were used in this research for a generalized evaluation across the community.

A. CODING PERFORMANCE ANALYSIS

The effectiveness of the proposed HEVC-MRSVC codec is compared with that of the anchor MV-HEVC for encoding each of the selected stereo video dataset at different Quantization Parameters (QP). Combined peak signal to noise ratios of Y-, U- and V-components (YUV-PSNR), which is a weighted sum of the average PSNR per video frames of the individual components (Y PSNR, U PSNR, and V PSNR) of the decoded and enlarged left and right view videos' frames together, as defined in (1) [37], is employed to assess the objective quality of the codec.

$$YUV\ PSNR = \frac{(6.Y\ PSNR + U\ PSNR + V\ PSNR)}{8} \tag{1}$$

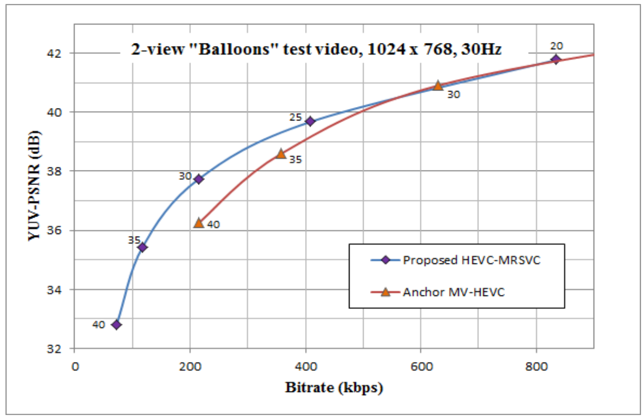


FIGURE 10. YUV-PSNR for MV-HEVC anchor and the proposed HEVC-MRSVC codecs for coding "Balloons" stereo test videos at different bitrates.

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

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

where, $I_{ref}(i,j)$ and $I_{dec}(i,j)$ represent the corresponding pixel values of the reference (I_{ref}) and decoded (I_{dec}) 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 [36]. Figures 10–14 show the resulting YUV-PSNRs at QPs of 20, 25, 30, 35, 40 and 25, 30, 35, 40 for the proposed HEVC-MRSVC and the anchor MV-HEVC codecs, respectively. Figure 10 and 11 illustrate the resulting YUV-PSNR for "Balloons" and "Kendo" benchmark stereo videos, respectively. The "Balloons" and "Kendo" datasets represent indoor scenes lit by multiple controlled illuminants. These videos contain progressive background changes with a few quick moving entities in the foreground. It is evident from Figure 10 that the proposed codec's videos exhibit higher objective quality (up to 1.8dB) than the anchor MV-HEVC codec's videos, for "Balloons" dataset. The proposed HEVC-MRSVC codec's videos exhibit 2.33 to 1 dB greater PSNR than that of the anchor codec's videos between 214 and 450 kbps, respectively. Since the scene characteristics of these stereo videos contain progressive background with moving entities in the foreground (as shown in Figure 9.a), when coding at lower QPs, it requires extra bits to represent difference blocks. The competency of the proposed HEVC-MRSVC codec gradually merges with that of the anchor codec from 550 kbps. Therefore, the competency of the proposed codec drops and matches the anchor codec. From Figure 11, it is evident that the proposed codec attains improved coding performance of 0.8 dB compared to the anchor codec at 400 kbps. The coding competency of the proposed codec

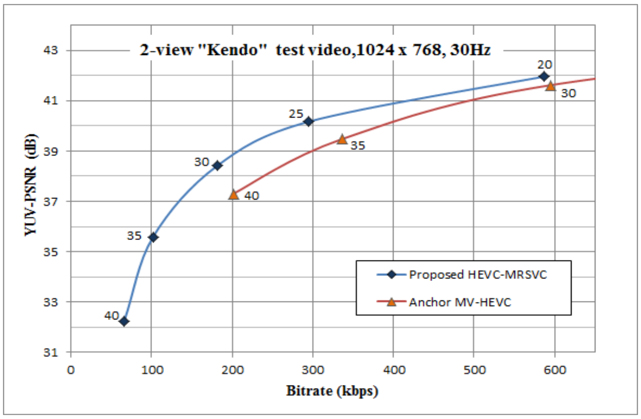


FIGURE 11. YUV-PSNR for MV-HEVC anchor and the proposed HEVC-MRSVC codecs for coding "Kendo" stereo test videos at different bitrates.

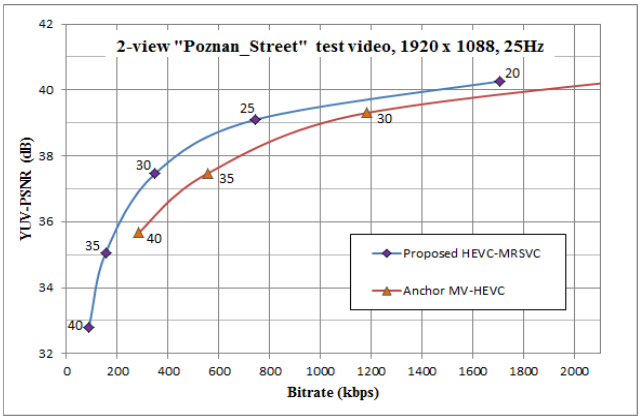


FIGURE 12. YUV-PSNR for MV-HEVC anchor and the proposed HEVC-MRSVC codecs for coding "Poznan_Street" stereo test videos at different bitrates.

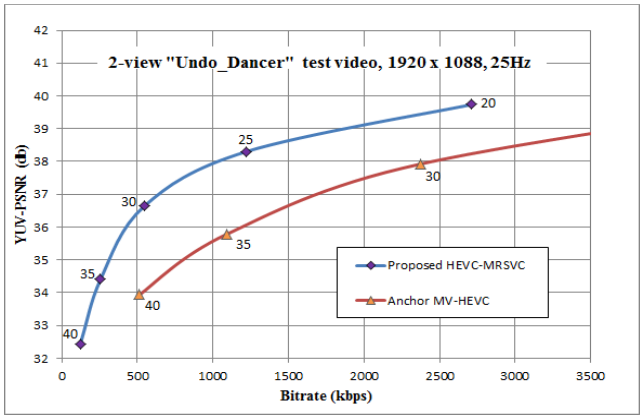


FIGURE 13. YUV-PSNR for MV-HEVC anchor and the proposed HEVC-MRSVC codecs for coding "Undo_Dancer" stereo test videos at different bitrates.

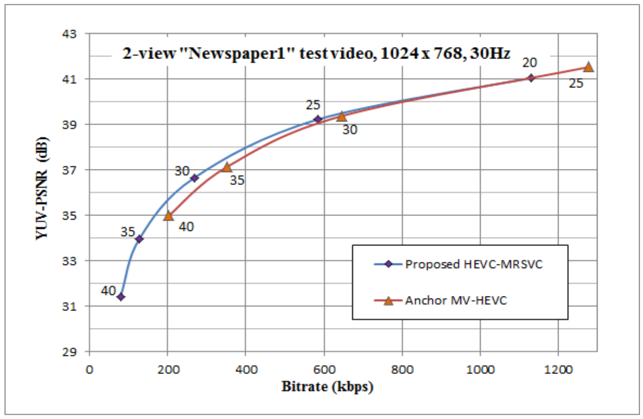


FIGURE 14. YUV-PSNR for MV-HEVC anchor and the proposed HEVC-MRSVC codecs for coding "Newspaper1" stereo test videos at different bitrates.

surges as the QP increases and it touches 1.8 dB at 250 kbps. By observing video contents (as shown in Figure 9.d) and experimental results, it can be established that the proposed codec yields greater coding competency when videos contain moving entities in either foreground or background. The resulting YUV-PSNR for coding "Poznan_Street" video sequences by the proposed HEVC- MRSVC and the anchor codecs are shown in Figure 12. The "Poznan_Street" dataset represents an outdoor naturally lit scene. The scene, which is recorded by stationary cameras, contains multiple moving objects in a stationary background. From Figure 12, it can be noted that the proposed codec's videos exhibit 0.8 to 1.9 dB greater PSNR than the anchor codec's videos between 1300 kbps and 284 kbps, respectively. By looking at the "Poznan_Street" videos (as shown in Figure 9.e), it can be seen that many regions of the video frames contain large stationary background, while moving objects cover small regions of the frames. Since the stationary areas of the scenes are coded by larger coding tree units, which require fewer bits to be coded, the coding competency of the proposed codec is capped by the moving entities' area. Experimental results for coding "Undo_Dancer" stereo video dataset are shown in Figure 13. This dataset contains computer graphic animation videos with dynamic changing scenes, which represent both camera and moving objects, with uniform level of illumination across the scene (as shown in Figure 9.c). From Figure 13, it is obvious that the proposed codec yields a greater coding competency than the anchor codec (about 1.7 dB at 2700kbps, it increases as the bitrate decreases and reaches 2.4 dB at 500kbps). The characteristics of video content of "Undo_Dancer" dataset is compared to that of "Poznan_Street" dataset. From this comparison, it is evident that the former has relatively higher camera panning than the moving objects, while the latter has a large stationary background with small moving objects. Hence, it can be concluded that the proposed codec's performance is higher when coding videos with scenes containing larger moving areas (an overall average gain of 2.05 dB in this case). Experimental results for coding "Newspaper1" stereo videos by the proposed HEVC-MRSVC and the anchor codecs are illustrated in Figure 14. This indoor stereo video dataset represents a scene with stationary background and moving objects close to cameras, moderately illuminated by artificial illuminants. From Figure 14, it can be inferred that the proposed codec's videos exhibits higher YUV-PSNR than that of the anchor codec (up to 0.47 dB). The proposed codec's videos demonstrate 0.2 dB higher YUV-PSNR than the anchor codec's videos at 600 kbps and this rises as the bitrate declines, ultimately reaches 0.75dB at 203 kbps. Upon examining characteristics of the video scene (from Figure 9.b), it is obvious that the stereo videos contain a stationary background, with foreground entities primarily in uniform motion. Therefore, few blocks are needed to be motion/disparity compensated and coded when coding P-/B-frames. I-frames in the proposed and the anchor codecs are coded in full resolution and these videos have very limited motion/disparity information in their P-/B-frames to be encoded. Therefore, the performance of the proposed codec does not show significant gain than that of the anchor codec. Upon examining the visual contents of the stereo video datasets with the coding competency of the codec, it can be concluded that the proposed codec yields considerably greater coding competency for videos containing larger moving areas. From Figures 10–14, it is evident that the proposed codec's performance grows as the total bitrate declines compared to the anchor codec.

TABLE 2. BD-PSNR of the proposed codec with respect to anchor codec.

	BD-PSNR (dB)		
Sequence	**BD-PSNR Y**	**BD-PSNR U**	**BD-PSNR V**
Balloons	1.327523	-0.002741	-0.046235
Kendo	1.214322	0.132543	1.726943
Poznan_Street	1.317625	-0.007961	0.120743
Undo_Dancer	2.839443	-0.002154	0.544325
Newspaper1	1.274581	-0.007451	-0.074563
Average	**1.594698**	**0.022447**	**0.454242**

A compact way of assessing and comparing the objective quality of the decoded HEVC-SVC and the anchor MV-HEVC's videos is using Bjøntegaard delta -bitrates (BD-rate) and -PSNR (BD-PSNR), which are computed using piece-wise cubic interpolation, as in [38] and [39]. A five data points based interpolation polynomial, as recommended in JCTVC-B055 document [40], has been used to generate BD-PSNR and BD-Rate for the proposed video codec with respect to the anchor codec's objective results. The resulting BD-PSNR and BD-Rate are tabulated in Table 2 and Table 3, respectively. From Table 2, it is evident that the proposed codec generally delivers greater performance compared to that of the standard MV-HEVC anchor codec in all video components (Y, U and V). The average BD-PSNR Y of the proposed codec for coding Balloons, Kendo, Poznan_Street and Newspaper1 videos is 1.283512 dB higher than that of the anchor codec. Whereas for the computer graphic animated dataset (Undo_Dancer), the proposed codec's videos exhibit 2.839443 dB higher BD-PSNR Y than that of the anchor codec. Overall, the proposed codec's videos have an average of 1.594698 dB

TABLE 3. BD-Rate of the proposed codec with respect to anchor codec.

	BD-Rate (kbps)		
Sequence	**BD-Rate Y**	**BD-Rate U**	**BD-Rate V**
Balloons	-35.8758	2.743165	-3.853461
Kendo	-25.7485	-1.841752	-13.854723
Poznan_Street	-42.6367	-3.782347	-2.764591
Undo_Dancer	-64.8762	-5.172641	-24.79438
Newspaper1	-28.7587	-2.834725	-12.64723
Average	**-39.5792**	**-2.17766**	**-11.582877**

BD-PSNR Y with respect to the anchor codec. The proposed codec's 'U' video components demonstrate slightly higher BD-PSNR U than that of the anchor codec, which is an average of 0.022447 dB. While, the proposed HEVC-MRSVC codec's 'V' components have an average improvement of 0.454242 dB over the anchor codec's videos. From Table 2, it can be noted that the proposed codec's BD-PSNR U for "Poznan_Street", "Undo_Dancer", "Newspaper1" and also its BD-PSNR V of "Newspaper1" video is marginally lower than that of anchor codec's videos. Meanwhile, the BD-PSNR U and BD-PSNRV of the proposed codec's "Balloons" videos show a curtailed performance than the anchor codec. Similarly for "Newspaper1" videos, the BD-PSNR U and BD-PSNRV of the proposed codec's videos are faintly lower than the anchor codec's. On the contrary, for the "Kendo" dataset, which has similar indoor lighting conditions as that of "Balloons", the proposed codec achieves the highest BD-PSNR U and BD-PSNR V amongst all other datasets.

The bitrate difference between the proposed and the anchor codec for encoding different stereo video datasets, in terms of Bjøntegaard delta -bitrates (BD-rate), are calculated and tabulated in Table 3. The negative values in this table imply that the proposed codec requires less number of bits than the anchor codec, when it delivers similar objective quality (PSNR). From this table, it can be noted that the proposed codec's BD-Rate Y and BD-Rate V are on an average 39.5792 kbps and 11.582877 kbps less than the anchor codec's bitrates, respectively. Whereas, BD-Rate U of the proposed codec is on an average 2.17766 kbps less than the anchor codec. The total average bitrate savings of the proposed codec compared to the anchor codec is 53.339717 kbps in terms of BD-Rate. Although the average BD-Rate U of the proposed codec is less than the anchor codec by a small margin, the total average BD-Rate of the proposed codec is considerably less than the anchor codec. From Table 2 and Table 3, it is evident that the proposed codec's videos exhibit considerably greater objective qualities in terms of BD PSNR at lower bitrates in terms of total BD-Rate.

From these analyses, it is evident that the proposed HEVC-MRSVC codec needs fewer frame signaling (e.g. motion and disparity vectors and CTUs) to represent low resolution video frames, since fewer bits are used to represent the video frames. In addition, the proposed codec benefits from cross-frame referencing the neighboring view frames via frame interleaving approach, which reduces its signaling overhead.

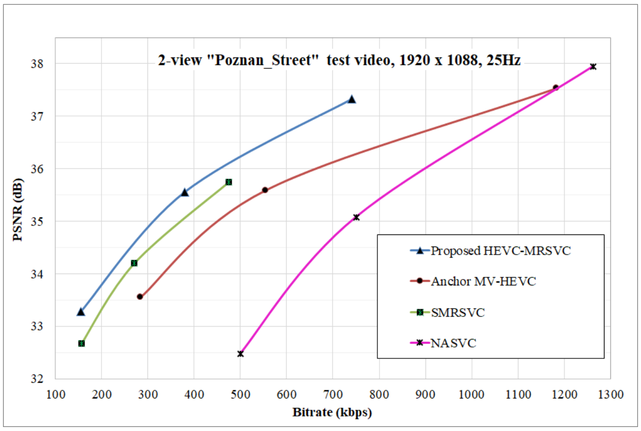


FIGURE 15. PSNR versus bitrate for coding "Poznan_street" stereo videos using proposed HEVC-MRSVC, anchor MV-HEVC, SMRSVC and NASVC codecs.

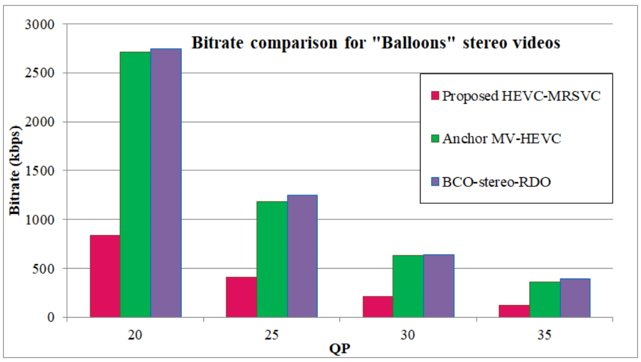
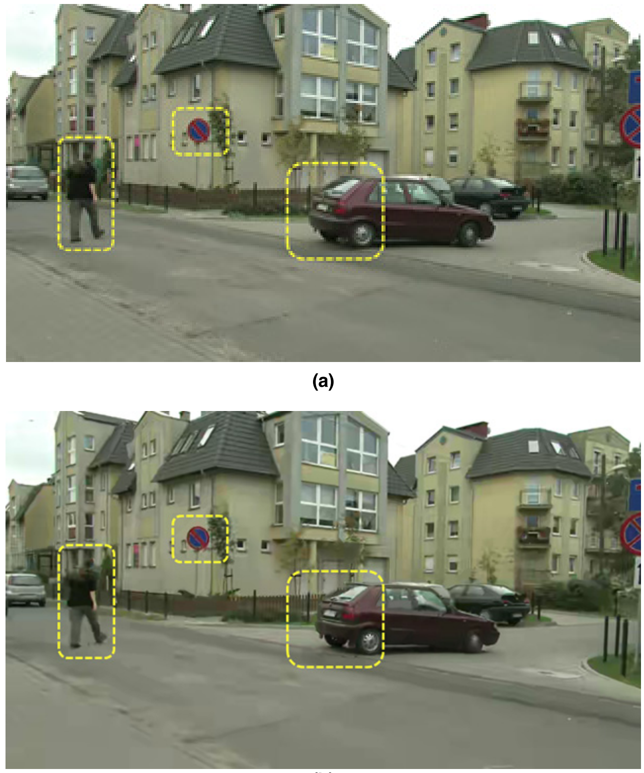


FIGURE 16. Total bitrate for coding "Balloons" stereo test videos using proposed HEVC-MRSVC, MV-HEVC anchor and BCO-stereo-RDO codecs.

Figure 15 shows the coding competency of the proposed HEVC-MRSVC codec, anchor MV-HEVC codec, symmetric mixed-resolution stereoscopic video codec (SMRSVC) [24] and perceptually driven non-uniform asymmetric stereoscopic video codec (NASVC) [22] for coding "Poznan_street" stereo videos in terms of PSNR for low bitrate transmission. From Figure 15, it is evident that the proposed codec has the highest coding competency in comparison with other three codecs. The proposed codec achieves 0.3, 0.9 and 3.7 dBs greater coding competency than SMRSVC, anchor MV-HEVC and NASVC codecs at 500 kbps, respectively. The superiority of the proposed codec can also be observed at low bitrates compared to the other three codecs as its coding gain further increases as the bitrate decreases. The coding competency of the proposed codec is also compared with the BJND based asymmetric stereo video coding model (BCO-stereo-RDO), proposed in [27] for coding "Balloons" videos. Figure 16 shows the total bitrates for coding "Balloons" stereo test videos at QP: 20, 25, 30 and 35, for the proposed HEVC-MRSVC, anchor and BCO-stereo-RDO and codec. From Figure 16, it is evident that the proposed codec requires significantly fewer bits to represent the encoded videos compared to the anchor and BCO-stereo-RDO codecs.



(a)

(b)

FIGURE 17. Decoded intermediate left view frame number 72 from "Poznan_Street" stereo videos at 545kbps bitrate of a) the proposed MRHEVC-MVC codec and b) the anchor MV-HEVC standard codec.

To give an idea of the attained visual quality for using the proposed codec and facilitate reader to compare it with that of anchor codec's videos, decoded and enlarged intermediate left view frame number 72 for coding "Poznan_Street" stereo videos by the proposed codec and the anchor codec are shown in Figure 17. Through the comparison of these two images, it is noticeable that the proposed codec's image has considerably greater visual quality than the anchor codec, as it contains much lesser blocking artefacts. To further distinguish and compare the achieved visual quality of the two methods' video frames, snippets of the highlighted areas are included in Figure 18(a)-18(c). From this figure, it is evident that the anchor codec's video frames have considerable levels of blocking artefacts with noticeable blurs on edges. In contrast, the video frames of the proposed codec have preserved more details in the background, along with edges of the objects. For example, the man's clothes and head edges, traffic sign, the signpost board's edges and the number plate of the car, which is a relatively faster moving object with respect to the man crossing the road. In summary,

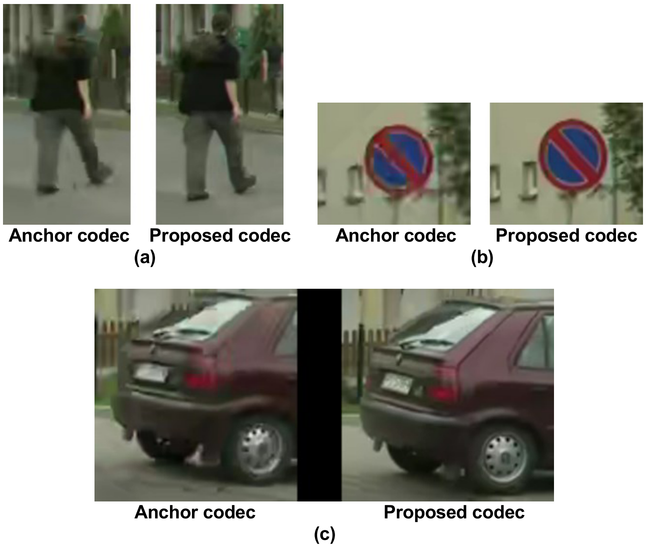


FIGURE 18. Snippets of the highlighted areas of the anchor and the proposed technique from Figure 17.

the proposed codec's videos exhibit lesser blocking artefacts, while retaining details, at the same bitrate.

B. COMPLEXITY ANALYSIS

Execution time is one of the metrics that is used to empirically assess the complexity of the codecs. This has been incorporated in the standard HEVC codec and coding complexities are presented in terms of encoding and decoding times. The JCT3V-G1100 CTC document, provides calculations for encoding time ratio (t_E) of geometric mean of encoding/decoding time at each QP for a test codec to that of the anchor codec [36]. The percentage change in computation time (ΔT), which is commonly used for direct comparison, is calculated using (3),

$$\Delta T = \frac{t_p - t_a}{t_a} \times 100\% \tag{3}$$

where t_p and t_a are the respective encoding times of the proposed and anchor codecs. In order to compare the coding complexity of the proposed codec with anchor codec, similar coding parameter, as suggested in JCT3V-G1100 CTC documentation (coding parameters, GOP size and I-frame period), were factored into analyzing the execution time. The proposed and anchor MV-HEVC codecs for two view scenarios were run on the same Microsoft Windows 7 based personal computer, running on a 6th generation core i5 microprocessor, with 8GB of random access memory and 250 GB hard-disk drive, without any dedicated graphic processing unit (no other applications, updates or background programs were running during the simulation). The resulting ratio of encoder execution time (t_E) of the proposed codec with respect to that of the anchor codec for encoding "Balloons", "Kendo", "Poznan_Street", "Undo_Dancer" and "Newspaper1" stereo videos are shown in Table 4. From Table 4, it can be derived that the proposed codec's geometric mean is between 32.58 to 40.13% less than that of the anchor codec. Furthermore, encoding percentage change, ΔT, is between 54.35% and 64.59% less than that of the anchor codec for coding the above named five standard test stereo videos. This can be clarified by the fact that the proposed HEVC-MRSVC's intermediate frames have quarter of size of the anchor codec. Hence, the proposed codecs requires less time to encode P- and B-intermediate frames than the anchor codec, which encodes intermediate frames in their full resolution.

TABLE 4. Encoder execution time.

Sequence	Encoding Time Ratio (t_E)	Encoding Time Percentage Change (ΔT)
Balloons	34.23%	-63.79%
Kendo	38.11%	-60.26%
Poznan_Street	43.83%	-54.35%
Undo_Dancer	32.58%	-64.59%
Newspaper1	40.13%	-57.87%

V. CONCLUSION

In this paper, a HEVC based Mixed-Resolution Stereo Video Codec (HEVC-MRSVC) has been introduced. The proposed codec applied a frame interleaving algorithm, along with a frame down-sampling and overlaying technique on the input stereo videos frames, to generate a monoscopic mixed spatial resolution video stream. The standard HEVC coding platform has been amended and configured to code videos of different spatial resolutions. Experimental results were generated using stereo views of five benchmark multiview videos. Results show the merit of the proposed HEVC-MRSVC codec compared to the anchor codec and state-of-the-art techniques for low bitrate transmission both objectively and subjectively.

REFERENCES

[1] P. Merkle, H. Brust, K. Dix, A. Smolic, and T. Wiegand, "Stereo video compression for mobile 3D services," in *Proc. 3DTV Conf.*, Potsdam, Germany, 2009, pp. 1–4.
[2] A. Smolic, K. Mueller, P. Merkle, P. Kauff, and T. Wiegand, "An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution," in *Proc. PCS*, Chicago, IL, USA, 2009, pp. 1–4.
[3] Cisco Visual Networking Index, "The Zettabyte era: Trends and analysis," Cisco, San Jose, CA, USA, White Paper 1465272001663118, Jun. 2017.
[4] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003, doi: 10.1109/TCSVT.2003.815165.
[5] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007, doi: 10.1109/TCSVT.2007.903665.
[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 J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1001–1016, Dec. 2013, doi: 10.1109/JSTSP.2013.2283657.

[7] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012, doi: 10.1109/TCSVT.2012.2221191.

[8] M. G. Perkins, "Data compression of stereopairs," *IEEE Trans. Commun.*, vol. 40, no. 4, pp. 684–696, Apr. 1992, doi: 10.1109/26.141424.

[9] I. Dinstein, G. Guy, J. Rabany, J. Tzelgov, and A. Henik, "On stereo image coding," in *Proc. IEEE 9th Int. Conf. Pattern Recognit.*, Rome, Italy, Nov. 1988, pp. 357–359, doi: 10.1109/ICPR.1988.28242.

[10] 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," *Proc. IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011, doi: 10.1109/JPROC.2010.2098830.

[11] *High Efficiency Video Coding*, document ISO/IEC 23008-2, ITU-T Recommendation H.265, Dec. 2016.

[12] P. D. Gunatilake, M. W. Siegel, and A. G. Jordan, "Compression of stereo video streams," in *Int. Signal Process. HDTV Workshop*, 1994, pp. 173–185.

[13] L. B. Stelmach and W. J. Tam, "Stereoscopic image coding: Effect of disparate image-quality in left- and right-eye views," *Signal Process. Image Commun.*, vol. 14, nos. 1–2, pp. 111–117, Nov. 1998, doi: 10.1016/S0923-5965(98)00031-9.

[14] P. Seuntiens, L. Meesters, and W. Ijsselsteijn, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Trans. Appl. Perception*, vol. 3, no. 2, pp. 95–109, Apr. 2006, doi: 10.1145/1141897.1141899.

[15] G. Saygili, C. G. Gürler, and A. M. Tekalp, "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 593–601, Jun. 2011, doi: 10.1109/TBC.2011.2131450.

[16] A. Aksay *et al.*, "Temporal and spatial scaling for stereoscopic video compression," in *Proc. EUSIPCO*, Forence, Italy, 2006, pp. 1–5.

[17] H. Brust, A. Smolic, K. Mueller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices," in *Proc. 3DTV Conf.*, Potsdam, Germany, 2009, pp. 1–4.

[18] C. G. Gürler and K. T. Bağci, "Adaptive stereoscopic 3D video streaming," in *Proc. ICIP*, Hong Kong, 2010, pp. 2409–2412.

[19] 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, 2007, pp. 1–4.

[20] Y. Chen, Y. K. Wang, M. M. Hannuksela, and M. Gabbouj, "Picture-level adaptive filters for asymmetric stereoscopic video," in *Proc. ICIP*, San Diego, CA, USA, 2008, pp. 1944–1947.

[21] P. Aflaki, M. M. Hannuksela, and M. Gabbouj, "Adaptive spatial resolution selection for stereoscopic video compression with MVHEVC: A frequency based approach," in *Proc. Int. Symp. Multimedia*, Taichung, Taiwan, 2014, pp. 267–270.

[22] S. A. Fezza and M. C. Larabi, "Perceptually driven nonuniform asymmetric coding of stereoscopic 3D video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 27, pp. 2231–2245, Oct. 2017, doi: 10.1109/TCSVT.2016.2583618.

[23] I. Koprinska and S. Carrato, "Temporal video segmentation: A survey," *Signal Process., Image Commun.*, vol. 16, no. 5, pp. 477–500, 2001, doi: 10.1016/S0923-5965(00)00011-4.

[24] K. L. Chung and Y. H. Huang, "Efficient multiple-example based super-resolution for symmetric mixed resolution stereoscopic video coding," *J. Vis. Commun. Image Represent.*, vol. 39, pp. 65–81, Aug. 2016.

[25] Y. Zhao, Z. Chen, C. Zhu, Y. P. Tan, and L. Yu, "Binocular just-noticeable-difference model for stereoscopic images," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 19–22, Oct. 2010, doi: 10.1109/LSP.2010.2090041.

[26] F. Shao, G. Jiang, M. Yu, K. Chen, and Y. S. Ho, "JND-based asymmetric coding of stereoscopic video for mobile 3DTV applications," in *Proc. ICIP*, Shanghai, China, 2011, pp. 186–189.

[27] C. Chang, C. J. Chiang, W. N. Lie, and T. C. Wu, "Binocular JND model based stereo image coding," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2016, pp. 1–5, doi: 10.1109/APSIPA.2016.7820674.

[28] C. Jung, Q. Fu, and F. Xue, "Perceptual stereoscopic video coding using disparity just-noticeable-distortion model," *J. Vis. Commun. Image Represent.*, vol. 48, pp. 195–204, Jun. 2017, doi: 10.1016/j.jvcir.2017.06.007.

[29] F. Shao, W. Lin, S. Gu, G. Jiang, T. Srikanthan, and B. V. characteristics, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013, doi: 10.1109/TIP.2013.2240003.

[30] B. Du *et al.*, "Novel visibility threshold model for asymmetrically distorted stereoscopic images," in *Proc. VCIP*, Chengdu, China, 2016, pp. 1–4.

[31] B. Mallik, A. Sheikh-Akbari, and P. B. Zadeh, "HEVC based stereo video codec," in *Proc. Int. Conf. Intell. Signal Process.*, London, U.K., 2015, doi: 10.1049/cp.2015.1764.

[32] B. Mallik, A. Sheikh-Akbari, M. A. Hussain, and A. L. Kor, "A mixed resolution based high efficiency video codec (HEVC)," in *Proc. Int. Conf. Intell. Signal Process.*, London, U.K., 2017, doi: 10.1049/cp.2017.0364.

[33] B. Mallik, A. Sheikh-Akbari, and A. L. Kor, "HEVC based mixed-resolution stereo video coding for low bitrate transmission," in *Proc. Int. Conf. Imag. Syst. Techn.*, Beijing, China, 2017, pp. 1–5.

[34] A. Sheikh-Akbari and P. B. Zadeh, "Wavelet based image enlargement technique," in *Proc. Int. Conf. Global Secur. Saf. Sustainability*, London, U.K., 2015, pp. 182–188.

[35] S. Bouyagoub, A. Sheikh-Akbari, D. Bull, and N. Canagarajah, "Impact of camera separation on performance of H.264/AVC-based stereoscopic video codec," *Electron. Lett.*, vol. 46, no. 5, pp. 345–346, Mar. 2010.

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

[37] H. Schwarz, T. Schierl, and D. Marpe, "Block structure and parallelism," in *High Efficiency Video Coding (HEVC): Algorithms Architectures* (Integrated Circuits and Systems), vol. 5. New York, NY, USA: Springer, 2014, ch. 3, sec. 3.2, pp. 49–90.

[38] G. Bjontegaard, *Calculation of Average PSNR Differences Between RD Curves*, document ITU-T SG 16, VCEG-M33, Apr. 2001, pp. 1–4.

[39] G. Bjontegaard, *Improvements of the BD-PSNR Model*, document ITU-T SG 16, VCEG-AI11, Jul. 2008, pp. 1–2.

[40] K. Senzaki, *BD-PSNR/Rate Computation Tool for Five Data Points*, document ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCTVC-B055, 2010, pp. 1–3.

BRUHANTH MALLIK received the M.Sc. degree in electronic engineering from the University of Portsmouth. He is currently pursuing the Ph.D. degree with Leeds Beckett University. His research interests are in the direction of studying and developing multiview video coding techniques for bandwidth limited transmission, within HEVC's framework.

AKBAR SHEIKH-AKBARI received the B.Sc. degree in electronic and electrical engineering (Hons.), the M.Sc. degree (Hons.) in electronic and electrical engineering, and the Ph.D. degree in electronic and electrical engineering from the University of Strathclyde. After completing the Ph.D. degree, he joined Bristol University, where he was involved in the EPSRC Project in stereo/multiview video processing. He continued his career in industry, where he was involved in real-time embedded video analytics systems. He is currently a Senior Lecturer with the School of Computing, Creative Technologies & Engineering, Leeds Beckett University. His main research interests include signal processing, hyperspectral image processing, source camera identification, image/video forgery, image hashing, biometric identification techniques, assisted living technologies, compressive sensing, camera tracking using retro-reactive materials, and standard and non-standard image/video codecs, such as H.264 and HEVC, multiview image/video processing, color constancy techniques, resolution-enhancement methods, edge detection in low-SNR environments, and medical-image processing.

AH-LIAN KOR specializes in software development, web applications, data analytics, and AI. She is active in sustainable IT, intelligent systems, decision support systems, and data center research. She forges an industrial collaboration with Drax Power Station, which aims to develop and evaluate a virtual environment (in the form of an expert system) to support decision making in fault diagnosis for the centrifuge system. She is currently an Intelligent System Academic Expert for the Knowledge Transfer Partnership Project with Premier Farnell (a global electronics engineering company). She is with the School of Computing, Creative Technology & Engineering, Leeds Beckett University. She is a Project Team Committee Member for several EU Projects, including ALIOT, Green Co, and PERCCOM. She has chaired sessions in reputable international conferences including FTC 2016. She was also the Publicity Chair of Cyber SciTech 2017 Technical Programme Committee for journals and conferences, including ICAART, SEEDS, IARIA Intelligent Systems, and the IEEE Computing Conference.