Multiview Video Coding: A Comparative Study Between MVC and MV-HEVC

Seif Allah El Mesloul Nasri, Abdul Hamid Sadka, Noureddine Doghmane, and Khaled Khelil

1 Introduction

During the last few years, the demand for higher-resolution video has witnessed a steady increase as much as the demand for interactive and three-dimensional (3D) visual content. It is predicted that the video traffic on the Internet will occupy 82% of all transmitted data by 2021 [1], and the 3D video content with its different formats will indeed be part of this traffic. 3D video content is today not only used for entertainment and leisure but it is also applied in several critical domains such as education, surveillance, cultural heritage and medicine [2]. The multiview video format offers a 3D experience to the end user, through depth sensation in addition to motion parallax. At least two cameras capture the multiview video from slightly different view angles. There exists a considerable amount of inherent redundancy between the viewpoints of the multiview video. Consequently, inter-view coding has been proposed taking into account the resemblance between the recorded views. Recent video coding standards such as H.264 [3] and H.265 [4] provide extended profiles that take advantage of the inter-view resemblances for better compression

---

S. A. El. M. Nasri
Department of Electronic and Computer Engineering, Brunel University, London, UK
e-mail: seif.nasri@brunel.ac.uk

Department of Electronics, University Badji Mokhtar, Annaba, Algeria

A. H. Sadka (✉)
Department of Electronic and Computer Engineering, Brunel University, London, UK
e-mail: abdul.sadka@brunel.ac.uk

N. Doghmane
Department of Electronics, University Badji Mokhtar, Annaba, Algeria

K. Khelil
LEER Lab, University Mohamed-Cherif Messaadia, Souk Ahras, Algeria

J. M. Alja'am et al. (eds.), *Recent Trends in Computer Applications*,
https://doi.org/10.1007/978-3-319-89914-5_7

efficiency. Based on the exploitation of both temporal and inter-view prediction, Merkle et al. proposed an approach that ensures a good trade-off between the bit rate and the video quality [5]. It was adopted and implemented by the Joint Video Team of ISO/IEC, Moving Picture Experts Group and ITU-T, and Video Coding Experts Group in a reference model named the Joint Multiview Video Model (JMVM) [6] or simply MVC, which is the extended profile of AVC/H.264. High compression efficiency is still the main requirement for multiview video coding in addition to other specific requirements such as low-delay random temporal and view access. Many research efforts [7–11] based on the MVC standards have been made with view of improving the coding capability with regard to the multiview video requirements list [12].

The first edition of the High Efficiency Video Coding standard (H.265) was finalised in 2013 by the Joint Collaborative Team on Video Coding (JCT-VC). The H.265 standard can achieve 50% bit rate saving for an equal perceptual video quality compared to H.264 [4]. Back in July 2012, the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) was established by the ISO/IEC MPEG and ITU-T Video Coding Experts Group (VCEG) in order to develop the next generation of the 3D video coding standards. As a result, the second edition with scalability extension (SHVC) [13] and multiview extension (MV-HEVC) [14] was completed in 2014 and published in early 2015.

In this manuscript, an evaluation of the MV-HEVC coding is presented in terms of compression efficiency relative to MVC. The assessment was conducted using multiple multiview video sequences with different contents and qualities.

The remainder of this chapter is organised as follows. Section 2 describes the multiview video and its coding principles. MV-HEVC technical concepts and features are presented in Sect. 3. Section 4 reports the compression performance of the two compared extensions. Finally, the conclusion is given in Sect. 5.

2 MVC Background

Multiview video can be produced when a set of synchronised cameras capture the same scene. The cameras record different angles of the same scene with an enriched overlapped content. Thus, 3D information of the scene is generated based on the cameras' similar content, offering an enhanced visual experience through depth feeling and motion parallax.

Multiview video visualisation is possible through a flat panel screen employing either parallax barrier or lenticular sheet technology (Fig. 1) to perceive a 3D image when both of the viewer's eyes are anywhere within the viewing zone. This so-called autostereoscopic display can support multiple viewers, each seeing 3D from his or her point of view. Looking around objects in the scene simply needs moving the viewer's head. More types of autostereoscopic displays are detailed in [15].

Typically, the conventional two-dimensional (2D) video is formed of continuous groups of pictures with a frequency of (25, 30 or 60 . . . etc.) frames per second.

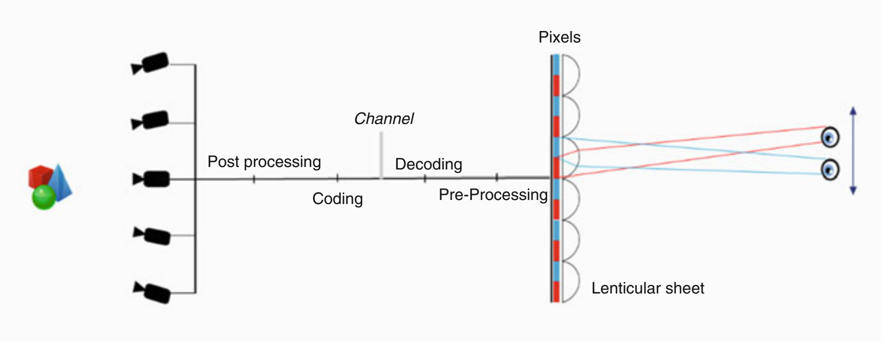


Fig. 1 Multiview video system

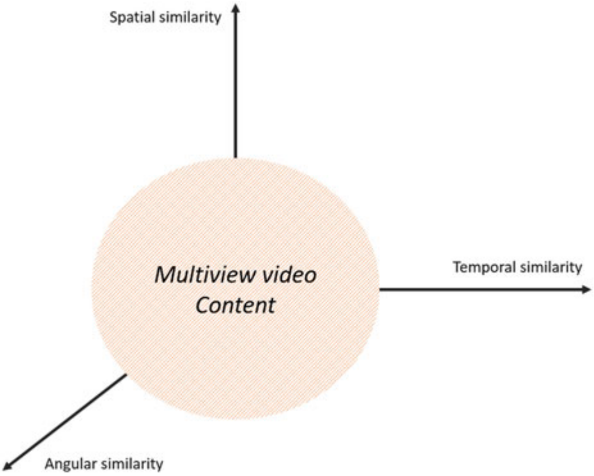


Fig. 2 Multiview video content similarities diagram

Successive frames have a certain degree of similarity, for example, an action video will have less similarity between its successive frames compared to an official speech clip with a fixed background. The same remark is noted for the temporal level between the successive frames [16]. This is associated with the inter-view correlation that exists among the views of the multiview video. The three types of similarity represent, in fact, a redundant information to be exploited to improve the compression efficiency. Figure 2 depicts the 3D similarity that exists within a multiview video content.

The synchronised sequences of the multiview video are coded jointly and simultaneously by only one video codec which is the MVC, as shown in Fig. 3. The MVC has to employ algorithms and techniques to reduce the amount of redundant data during the compression process.

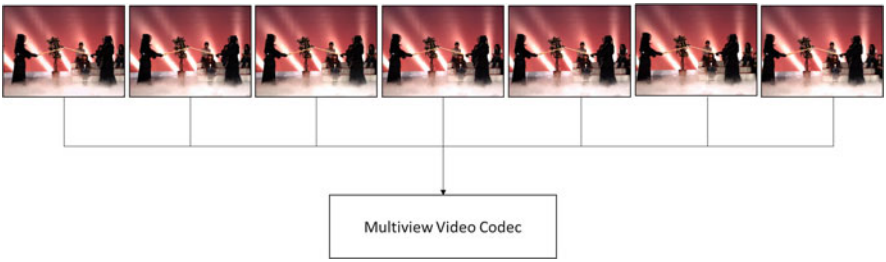


Fig. 3 Multiview video codec input

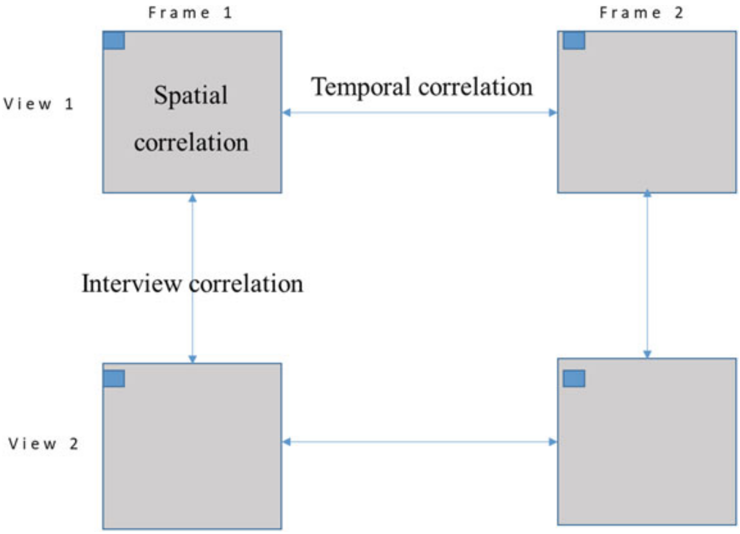


Fig. 4 Redundant information exploitation in the multiview video content

Effective motion-compensated and disparity prediction algorithms are used to eliminate the redundant information between the successive frames and adjacent views, respectively (Fig. 4).

The simplest method for coding a multiview video is the simulcast method, which performs the compression by exploiting only the spatiotemporal redundancies and coding each view independently using a conventional video codec. By making use of H.264/AVC and the hierarchical B pictures, video compression has been efficiently improved in comparison to the traditional simulcast coding structures [17]. Figure 5 depicts the hierarchical B pictures structure where the number of frames in the group of pictures (GOP) is equal to 8. The first picture is independently coded as an instantaneous decoder refresh (IDR) picture, and the

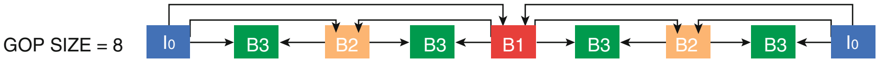


Fig. 5 Hierarchical B pictures structure

so-called anchor or key pictures are coded within regular intervals. The B pictures, located between two I pictures and known as non-key frames, are hierarchically predicted using the concept of hierarchical B pictures.

Despite the fast random access provided by the simulcast coding method, its coding efficiency is not optimal as it neglects the inter-view dependencies during the compression process. Simulcast method is typically employed as a reference model for coding performance comparisons between different MVC schemes.

In fact, research on multiview video coding has been active for more than 30 years since the emergence of the disparity compensated concept in 1986 [18], followed by other propositions in 1989 [19] and 1992 [20]. The first official standardisation of the MVC was in 1996 [21] and consisted of extending H262/MPEG-2 [22] capabilities to support the multiview video content. However, at that time, the ultimate challenge was to upgrade video services from the standard analogue definition to the digital high definition. This fact prevented the multiview extension of H.262/MPEG-2 from being applied and developed.

Following the progress in video compression technologies and multimedia services, MPEG launched a call for proposal on MVC in July 2005. Based on the AVC/H.264 coding standards, some proposed responses introduced different forms of inter-view prediction structures [23–25]. Compared to the simulcast coding where each view is coded independently, the inter-view coding methods offer significant gains in terms of bit rate saving. Merkle et al.'s [5] approach was adopted and implemented by the Joint Video Team in a reference model named the Joint Multiview Video Model (JMVM) [6].

Figure 6 presents the inter-view prediction structure used as the default structure of JMVM. Eight views (cameras) are employed in this scheme where S_n indicates the different cameras, while T_n represents the time location of the frames. Moreover, in this case, each group of groups of pictures (GGOP) is composed of eight views and eight pictures per GOP.

The IBP structure employs three types of views: I-view as one base view per Group of groups of pictures, P-views which are predicted from a unique direction and B-views involving bi-directional inter-view prediction for coding its set of frames. Much research work based on MVC/H.264 has been undertaken to improve the outcomes of the coding process in terms of view random access [26, 27]. Meanwhile, research on MVC/AVC is still underway while HEVC codec implementation in the market is going at slow pace due to its loyalty cost and complexity.

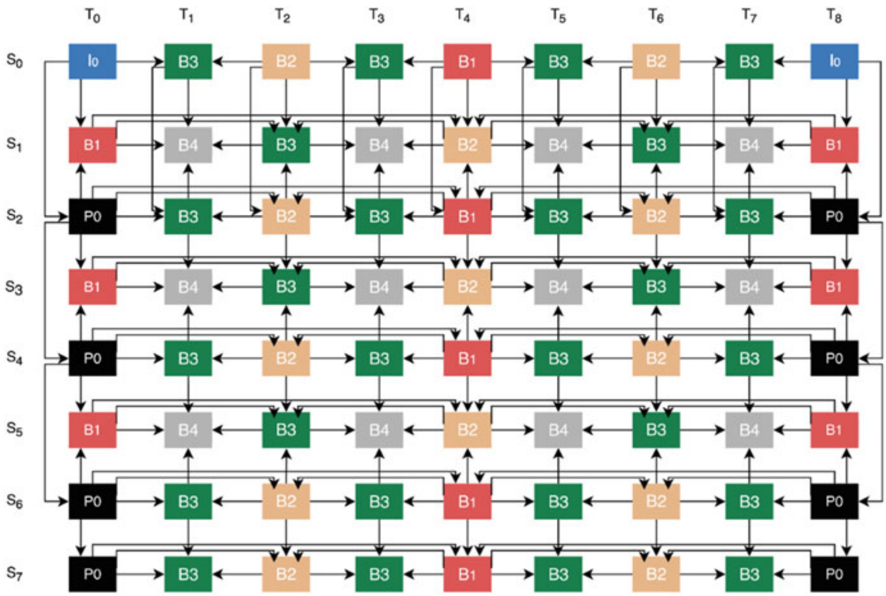


Fig. 6 IBP prediction structure

3 MV-HEVC

Results of subjective evaluation [28] show that HEVC/H.265 standard can reach the same quality levels as H.264/AVC whilst generating approximately 50% lower bit rate on average. HEVC standard adopts innovative tools which contribute to achieving this gain, such as accurate intra-/inter-predictions, in-loop sample adaptive offset filter and quadtree-based block partitioning [4].

HEVC benefits from using variable pattern comparison and difference-coding areas starting from blocks of 16 × 16 to 64 × 64 pixels. The concept behind this is based on partitioning the frame into coding tree units (CTUs), which replace the macroblocks used in H.264. Each CTU contains two chroma and one luma coding tree blocks CTBs. CTB size can be 16 × 16, 32 × 32 or 64 × 64, where larger pixel block size increases the compression efficiency. The CTBs are then divided into one or more coding units (CUs) as shown in Fig. 7. The CU is split into prediction units (PUs), a basic entity for intra- and inter-predictions, variable in size from 64 × 64 to 4 × 4 pixels. Variable partition scenarios have been defined in the design of the HEVC encoder considering a certain attention to complexity. For instance, to deal with critical case memory bandwidth in the decoding process, PUs coded using temporal inter-prediction are restricted to the minimum size of 8 × 8 if they are bi-predicted from two references, or 8 × 4 or 4 × 8 if they are predicted from a single reference [4].

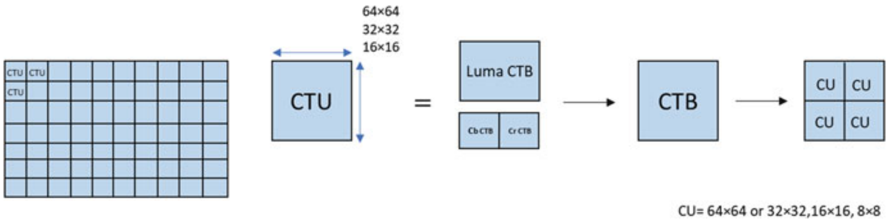


Fig. 7 Block partitioning in HEVC

Compared to AVC which includes only eight directional modes for the intra-picture prediction, HEVC employs 33 intra-picture prediction modes in addition to planar (Mode 0, surface fitting) and DC (Mode 1, flat) prediction modes. Due to the increased number of modes (35), efficient coding of intra-prediction mode is achieved by using a list-based approach. For each prediction unit, the most probable three modes are determined and a Most Probable Mode (MPM) list is constructed from these modes.

The HEVC bitstreams include an elementary unit called a network abstraction layer (NAL) unit, composed of payload and a header. The NAL header consists of a 5-bit NAL unit type, 6-bit layer identifier called nuh_layer_id, and a 3- bit temporal sub-layer identifier. A new video parameter set (VPS) structure has been included in HEVC as metadata representation to allow the extension compatibility of the standard including dependences between temporal sub-layers. It also contains essential data that can be shared with the decoding process.

The multiview extension of HEVC uses the same fundamental coding tools of the HEVC main profile in addition to some specific features mainly related to the stereoscopic and multiview representations. MV-HEVC provides bit rate saving compared to the standard HEVC simulcast by enabling the exploitation of the inter-view references within the motion-compensated prediction. It is also noted that MV-HEVC utilises the same coding design principle (IBP) as the multiview extension of H.264.

However, the concept of inter-view has been replaced in MV-HEVC by the inter-layer prediction design. The multi-layer approach is employed in all multi-layer extensions [29], including MV-HEVC, 3D-HEVC, as well as the scalable extension of HEVC (SHVC). A layer can represent a depth, texture or other auxiliary information related to a particular camera view. All layers of the same camera perspective are marked as a view; while layers representing the same type of information are denoted as components in 3D video (Fig. 8).

MV-HEVC includes high-level syntax (HLS) additions [14] and can be implemented using existing 2D single-layer decoding cores. Moreover, MV-HEVC shares the same HLS with all HEVC multilayer extensions. HLS enables the extraction of a single texture base view from MV-HEVC bitstream which is decodable by the main profile HEVC decoder.

Figure 9 shows an example of MV-HEVC bitstream with three texture views coded by the so-called IBP inter-view structure. The base layer (left view) is coded

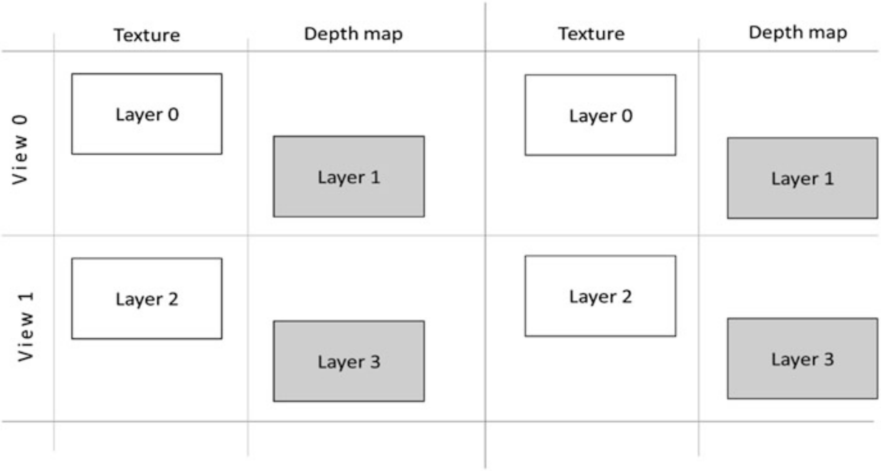


Fig. 8 Layers division in MV-HEVC

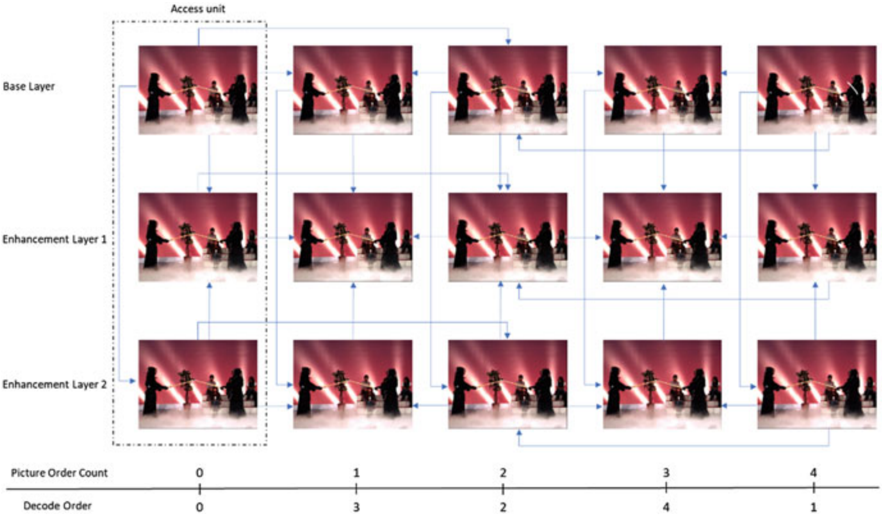


Fig. 9 MV-HEVC bitstream with three texture views using IBP inter-view prediction

independently of other views using HEVC main profile. The MV-HEVC profile is enabled to code the two enhancement layers (ELs). EL2 (right view) utilises inter-view prediction from the base layer, and EL1 (centre view) is predicted from both left and right views.

4 Experimental Results

In this section, the compression efficiency of MV-HEVC and MVC is compared and evaluated. Four different video sequences have been used in the experiments. Table 1 describes the used multiview video sequences and their parameters. Also, samples of the tested sequences are illustrated in Fig. 10.

Table 1 Multiview video sequences used for the compression efficiency evaluation

Database	Video sequences	Frame rate	Image resolution	Camera parameters
MERL	Vassar	25	640 × 480	8 cameras/20 cm spacing
MERL	Ballroom	25	640 × 480	8 cameras/20 cm spacing
Fujii Lab	Kendo	30	1024 × 768	7 cameras/5 cm spacing
Fujii Lab	Balloon	30	1024 × 768	7 cameras/5 cm spacing

Balloon

Ballroom

Kendo

Vassar

Fig. 10 First view picture of the used multiview video sequences

Table 2 Initial common encoding configuration

Frames to be encoded	250
GOP size	8
Intra period	8
Quantization Parameter	[25,30,35,40]
Search mode	Fast mode
Search range	64

The objective evaluation is shown using graphs of peak signal-to-noise ratio PSNR (dB) versus bit rate (kbit/s). The PSNR which expresses the video quality is given by:

$$\mathrm{PSNR} = 10 \times \log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right) \tag{1}$$

MSE represents the mean square error between the original and the compressed video signals. Conventionally, the objective measure of quality is applied to the luminance video signal regardless of the chrominance signals. Table 2 regroups the common primary conditions that have been used to obtain a fair comparison. The quantisation parameter (QP) controls the quality of the compressed video and the bit rate of the generated bitstream; the higher the value of the QP, the lower is the bit rate and the video quality. Four QP values are chosen according to the standardisation tests defined in [30].

It can be clearly inferred from Figs. 11 and 12 that the MV-HEVC exceeds the MVC in terms of bit rate saving and video quality. This outperformance ultimately covers all the conducted tests through the different datasets and conditions. The rate distortion (RD) curves of the high-definition multiview video sequences, shown in Fig. 11, prove that MV-HEVC codec improves the compression performance compared to MVC over the entire bit rate range. For instance, for QP = 25, the bit rate saving gain achieved by MV-HEVC exceeds 25% and 31% for Balloon and Kendo sequences, respectively. Furthermore, Fig. 12 shows that the MV-HEVC bit

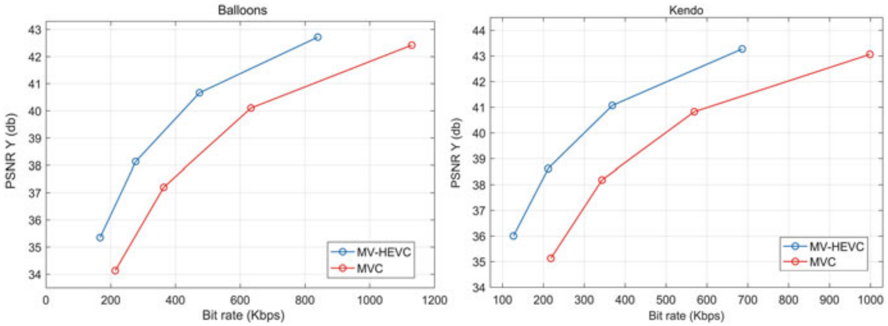


Fig. 11 Compression efficiency comparison through HD multiview video sequences

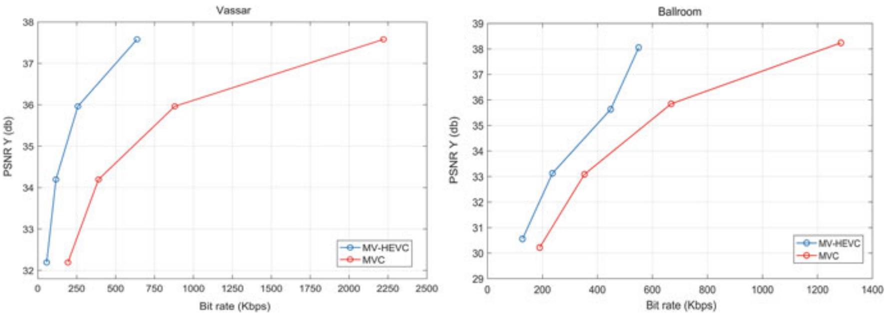
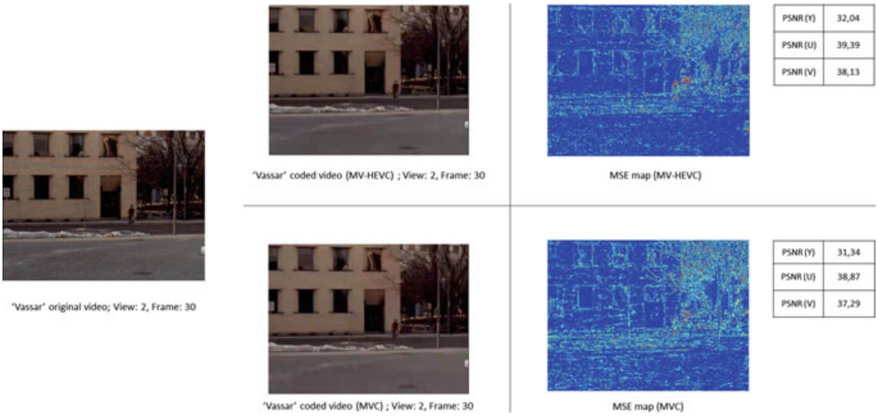


Fig. 12 Compression efficiency comparison through SD multiview video sequences



PSNR (Y)	32,04
PSNR (U)	39,39
PSNR (V)	38,13

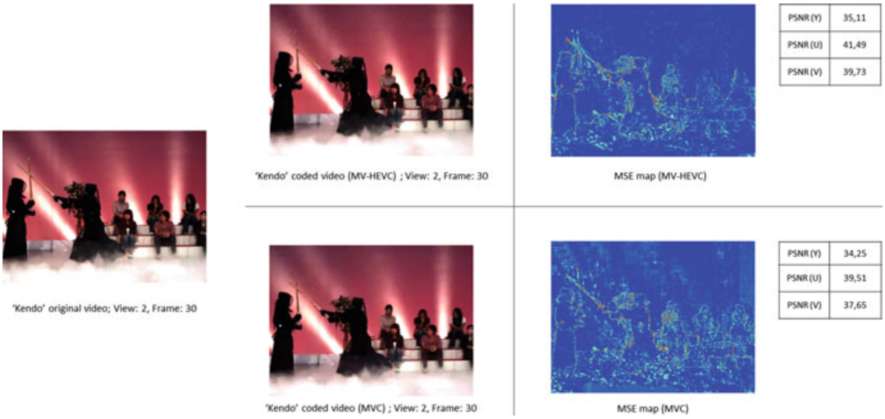
PSNR (Y)	31,34
PSNR (U)	38,87
PSNR (V)	37,29

Fig. 13 Image quality comparison between MV-HEVC and MVC using Vassar sequence

rate saving is further increased for the standard definition sequences, whereby a gain of 71% and 57% is achieved for Vassar and Ballroom sequences, respectively.

Figures 13 and 14 present a frame-based comparison between MV-HEVC and MVC codecs. Frame number 30 located in view 2 (camera 2) of the two chosen multiview video sequences is selected for this comparison. This frame, which comes after three successive groups of pictures, is coded using both temporal and inter-view predictions. Also, the quantization parameter QP = 40 has been selected for this comparison to evaluate the performance of the reported codecs at the lowest level of perceptual image quality.

Figure 13 shows the comparison using a standard resolution video (Vassar), the degradation can be seen in the compressed frame with MV-HEVC and MVC as well. However, the difference cannot be clearly perceived between the two compressed frames. The MSE maps slightly highlight the difference between the two compressed frames, where extra red regions are observed in the frame compressed by MVC codec, which indicates a larger number of mismatching errors.



PSNR (Y)	35,11
PSNR (U)	41,49
PSNR (V)	39,73

PSNR (Y)	34,25
PSNR (U)	39,51
PSNR (V)	37,65

Fig. 14 Image quality comparison between MV-HEVC and MVC using Kendo sequence

However, the blue regions, which represent the matching between the original and the compressed frames, are distinctly perceived in the frame compressed by MV-HEVC. The PSNR values confirm the MSE map results, where the PSNR (Y) gain of MV-HEVC is 0.7 dB, and the overall value is 0.69 dB.

Almost a similar perception can be obtained from Fig. 14 where HD multiview video sequences have been used with the same quantisation parameter value for both codecs. The results emphasize the same fact that MV-HEVC outperforms MVC in terms of image quality with a gain of 0.86 dB achieved for PNSR(Y) and 1.64 dB for the mean value which includes PSNR(Y), PSNR(U) and PSNR(V).

5 Conclusion

The chapter reviewed the multiview video coding theory and concepts, focusing on MVC and MV-HEVC coding standards. Both codecs use the same IBP design for the disparity compensation in addition to the hierarchical B algorithm for the temporal level. The MV-HEVC employs the powerful tools of HEVC such as the innovative block partitioning to improve the rate distortion capability. Both codecs have been implemented and evaluated through different datasets and common test conditions. The used test video sequences were multiple texture views without depth map of SD and HD resolutions. Test results have shown an increased compression efficiency of MV-HEVC compared to MVC. The significant bit rate saving gain starts from 24% for Balloon sequences and achieves 70% for Vassar sequences.

Acknowledgement This publication was made possible by NPRP grant 9-181-1-036 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

1. Cisco, Visual Networking Index: Forecast and Methodology, 2016–2021 June 6, 2017.
2. A. Smolic et al., "3D video and free viewpoint video—technologies, applications and MPEG standards", IEEE Int. Conf. Multimedia and Expo, pp. 2161–2164, IEEE, Toronto, Ontario, Canada (2006).
3. J. Ostermann et al., "Video coding with H.264/AVC: tools, performance, and complexity", IEEE Circuits and Systems Magazine. 4(1), 7–28 (2004)
4. G. J. Sullivan et al., "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Circuits and Systems Magazine. 22(12), 1649–1668 (2012).
5. P. Merkle et al., "Efficient prediction structures for multiview video coding," IEEE Trans. Circuits Syst. Video Technol. 17(11), 1461–1473 (2007).
6. Y. Chen, P. Pandit, and S. Yea, "WD 4 reference software for MVC," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-AD207 (2009).
7. Y. Yang, Q. Dai, J. Jiang, and Y-S. Ho, "Coding order decision of B frames for rate-distortion performance improvement in single-view video and multiview video coding," IEEE Trans. Image Processing, vol. 19, no. 8, Aug. 2010.
8. S.H. Hany, M. El-Khamy and M. El-Sharkawy, "Blind configuration of multi-view video coder prediction structure", IEEE Trans. Consumer Electron, vol. 59, no. 1, pp. 191–199, 2013
9. A. B. Ibrahim, A. H. Sadka "Error resilience and concealment for Multiview video coding" IEEE Int. Symp. on Circuits and Systems 2014, pp. 1–5 (2014).
10. Pei-Jun Lee, Ho-Ju Lin, and Kuei-Ting Kuo, "Faster mode determination algorithm using mode correlation for multi-view video coding," IET Signal Process, vol. 8, no. 5, pp. 565–578, 2014.
11. S. Nasri et al., "Enhanced view random access ability for multiview video coding," J. Electron. Imaging 25(2), 023027 (2016).
12. "Requirements on multi-view video coding v.4," ISO/IEC JTC1/SC29/WG11, Doc. N7282, Poznan, Poland (2005).
13. J. M. Boyce et al., "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard", IEEE Trans. Circuits Syst. Video Technol. 26(1), 20–34 (2016).
14. G. Tech et al., "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding", IEEE Trans. Circuits Syst. Video Technol. 26(1), 35–49 (2016).
15. N.A. Dodgson, "Autostereoscopic 3D displays", Computer, 38(8), 31–36 (2005).
16. A. H. Sadka, Compressed Video Communications. Halsted Press, 2002.
17. L. Ma and F. Pan, "Efficient compression of multi-view video using hierarchical B pictures," in Int. Conf. on Multimedia and Ubiquitous Engineering, pp. 118–121 (2008)
18. M. E. Lukacs, "Predictive coding of multi-viewpoint image sets", IEEE Int. Conf. Acoust. Speech Signal Process, vol. 1, 521–524. Tokyo, Japan (1986).
19. I. Dinstein et al., "On the compression of stereo images: Preliminary results", Signal Process., Image Commun., 17(4), 373–382 (1989).
20. M. G. Perkins "Data compression of stereo pairs", IEEE Trans. Commun., 40(4), 684–696, (1992).
21. ITU-T and ISO/IEC JTC 1, Final draft amendment 3, Amendment 3 to ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video), ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. N1366 (1996).
22. ITU-T and ISO/IEC JTC 1, Generic coding of moving pictures and associated audio information Part 2: Video, ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video) (1994).
23. MPEG Video Sub-Group Chair (J.-R. Ohm), Submissions received in CfP on multiview video coding, Bangkok, Thailand, ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. M12969 (2006).
24. MPEG Video and Test Sub-Groups, Subjective test results for the CfP on multi-view video coding, Bangkok, Thailand, ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. N7799 (2006).
25. K. Muller et al., "Multiview coding using AVC", Bangkok, Thailand, ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. M12945 (2006).

26. A. Bekhouch et al., “Improving view random access via increasing hierarchical levels for multi-view video coding”, IEEE Transactions on Consumer Electronics, 62(4), 437–445 (2016).
27. S. Nasri et al., “Group of Pictures Effects on Proposed Multiview Video Coding Scheme”, 40th International Conference on Telecommunications and Signal Processing, pp. 548–554 (2017).
28. J.-R. Ohm et al., “Comparison of the coding efficiency of video coding standards— Including High Efficiency Video Coding (HEVC)”, IEEE Trans. Circuits Syst. Video Technol., 22(12), 1669–1684 (2012).
29. R. Sjoberg et al., “Overview of HEVC high-level syntax and reference picture management”. IEEE Trans. Circuits Syst. Video Technol., 22(12), 1858–1870 (2012).
30. V. Baroncini et al., “MV-HEVC Verification Test Report”, Joint Collaborative Team on 3D Video Coding Extensions of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Doc. JCT3V-N1001 (2016)