

A Survey on Multi-view and 3D High Efficiency Video Coding on Real Time Streaming

Pang Yik Siang¹ and Tew Yiqi²

Department of Computer Science and Embedded Systems,
Faculty of Computing and Information Technology,
Tunku Abdul Rahman University College
¹pangys@tarc.edu.my, ²yiqi@tarc.edu.my

ABSTRACT. In the past few years, the demand of multi-view was increased rapidly and there was a lot of research works to improve the technique and fulfil its needs. High Efficiency Video Coding (HEVC) compression standard has been implemented in this work. HEVC is a compression standard designed to reduce bitrate and remain the same quality compared to the previous compression standard Advance Video Coding (H.264). It will provide a better compression to higher resolution video such as Ultra High Definition (UHD). In this paper present a preliminary study on Multi-view with depth by using HEVC compression on a real-time streaming protocol. The study of proposed work may help the industry to enhance the viewing experience by multiple camera capture and also resolve the data traffic issue to transmit UHD video.

KEYWORDS: High Efficiency Video Coding, 3D-HEVC, Multi-view, real-time streaming, Ultra High Definition

1 INTRODUCTION

In 21st century, the video has become our necessities which are not to escape from our living life. In our daily life, the video can be use in our mobile device as a recreational purpose, even applicable in the medical and manufacturing industry as a scanning process. CISCO global Internet Protocol (IP) traffic forecast shows that the IP video will be in the range of 80% to 90% of total IP traffic in the rest of the time (CISCO 2019). The statistic shown the UHD video consume among all the different resolution video will be increasing from 3% to 22% in the years from 2017 to 2022. In future it is predicted that the new revolution of industrial is strictly rely on the speed of the process and tend to develop all data transmission processes are near to real-time. The latest 5G cellular network emerged may help to drive the technology to another level as its average data transmission speed in-

creased estimated up to 100 times faster than previous version cellular network 4G.

On the other hand, the video resolution was up to 8K resolution also known as UHD Video. In apart from wireless network transmission speed, 8K resolution video significantly affected the sustainability of hardware storage because of video size is increased hugely. Besides, the number of live stream video cameras are installed in the industry increase year by year, the storage for supporting the huge number video files must take into account.

However, video technology was also growing rapidly to resolve the storage and the bitrate transfer issue. The beginning of video compression as we know is Motion Picture Expert Group (MPEG) standard commonly used in Video Compact Disc (VCD) launched in 1993. After two years, MPEG-2 was established and utilised in digital video disc (DVD) and also satellite TV. Besides, Advance Video Coding (AVC) codec which published in 2003 by Video Coding Experts Group (VCEG) was widely used in current industry such as video gaming, blue-ray video disc, IP Video On Demand (VOD) and online video streaming. In January 2013, VCEG collaborated with MPEG and formed a team called Joint Collaborative Team on Video Coding (JCT-VC) finalised a new video project named HEVC (H.265 or MPEG-H Part 2) (Sullivan et al. 2012). HEVC enables to resolve the issue of the higher bitstream from UHD video by reduced 50% of the bitrate with same video quality in respect to AVC (Sullivan et al. 2012). The second edition of HEVC standard such as MV-HEVC and 3D-HEVC investigated by the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) by the ITU-T VCEG and ISO/IEC MPEG finalised in 2014 (Tech et al. 2016). MV-HEVC extension supporting stereo and multi-view capture by without changing

low-level syntax. Moreover, the 3D-HEVC extension is designed for better compression to multi-view with depth video. In the same year, JCT-VC and JCT-3V have developed Scalability extension of HEVC called SHVC to support spatial, quality and colour gamut scalability. SHVC is the successor of Scalable Video Coding (SVC) which is scalable extension of H.264/AVC standard.

The rest of this work will be discussing the review of other relevant works and theory behind the HEVC extension. The relevant validation method will be introduced as well.

2 Technical Background

The main concern on this survey is the real-time multiple video streaming by using a higher efficiency coding extension such as MV-HEVC and 3D-HEVC. Therefore, several video streaming protocols in real time and multi-view and 3D concepts based on HEVC extension standard are discussed in this section.

2.1 Real time video streaming protocol

Various research works were doing about real-time streaming prototyping set up with a different kind of approach. In recent year, most of the proposed work was focusing on single view of real-time video coding (Tech et al. 2016, Parois et al. 2016, Viitamäki et al. 2018, Le et al. 2018). HEVC Scalability Extension (SHVC) is the most common video coding extension has been applied because it provides network-friendly scalability when a bitstream level is low due to bandwidth availability and also variety decoder in different devices such as smartphone, laptop and television (TV). Based on the theory (Boyce et al. 2016), SHVC extension provides a technique for coding video in multiple layers, every layer represents a diverse quality of the same video scene. The base layer (BL) represents the lowest quality. One or more Enhancement Layers (ELs) may be coded by referencing lower layers and provide improved video quality. Decoding a subset of layers of a scalable coded video bitstream results in video with a lower quality but still acceptable compare to full bitstream were decoded. This allows a more graceful degradation compared with non-scalable video bitstreams, where a reduction in bitrate typically causes more severe drops in video quality, often becoming of unacceptable quality for viewing. Compared with non-scalable video coding, scalable video coding usually required more bits to achieve the

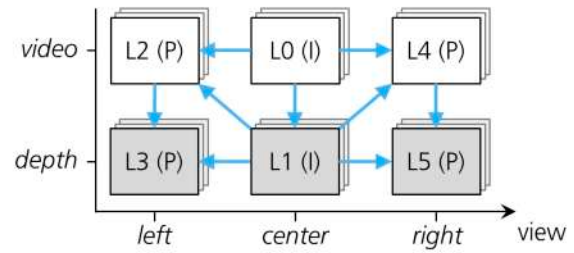


Figure 1: inter-layer prediction structure for 3D-HEVC with video and depth of three views (HHI 2019)

same video quality. SHVC is able to process a few resolution videos into one encoded bitstream, user is able to select different resolution decoded video to show. Therefore, SHVC is capable of solving the unstable of bandwidth was regularly happened in our real-life especially transmit the live HD/UHD video to play on our online social application, most of the research works will test it as real-time video streams coding.

2.2 MV & 3D-HEVC Extension

The concept of Multi-view Video Coding (MVC) extension was initially developing for H.264/AVC in 2009, and now it was also implemented in MV-HEVC and 3D-HEVC. MVC extension without modifying the basic encoding process like motion-compensated prediction (MCP), it only improving the high-level syntax (HLS), with the feature of disparity-compensated prediction (DCP) to reducing inter-view redundancies created higher coding efficiency (Tech et al. 2016). With the high demanding of the 3D video market, multi-view Video plus Depth (MVD) has been applied to develop a new HEVC extension standard named 3D-HEVC. 3D video coding will be more complicated of encoding and decoding process compared with 2D video coding. The 3D-HEVC video coding is supporting multiple views and associated depth data suitable for delivering 3D videos development. In Fig 1 shows inter-layer prediction structure for 3D-HEVC coded with MVD, the base-layer L(0) has inter-layer independent prediction same as the concept of inter-frame prediction in HEVC standard (HHI 2019). Layer (L1 to L5) are predicted dependencies on inter-component and inter-view between both video and depth. MVD can be reduced up to 70% bitrate compared to simulcast video coding which supported by few types of inter-layer prediction which are technique of inter-view prediction between different view or

Table 1: Functional Comparison Table on various HEVC research works

Authors	Codec					Real-time	Result Validation
	AVC	SHVC	HEVC	3D	MV		
Fezza et al. (2011)	✓				✓		PSNR/Bit-rate
Zhouye Gu et al. (2013)			✓	✓			Time and BD-rate
Gopalakrishna et al. (2013)			✓	✓			Time and PSNR/Bit-rate
Zhouye Gu et al. (2014)			✓	✓			FPGA Synthesis (frequency)
Springer et al. (2014)			✓			✓	N/A
Lee et al. (2014)		✓	✓		✓		BD-rate
Diaz-Honrubia et al. (2015)	✓	✓					Time, BD-rate 2 & 3views
Parois et al. (2016)		✓				✓	DEMO
Chan et al. (2016)			✓	✓	✓		BD-rate, Depth Bit-rate
Viitamäki et al. (2018)		✓				✓	DEMO
Le et al. (2018)		✓				✓	Time and Throughput (Mbps)

depth from MVC; inter-component prediction between the video and depth component of the same view; and combined inter-view or inter-component prediction between video and depth of different views.

3 State-of-the-Art 3D HEVC

In the earlier stage of multi-view video coding was used in AVC standard, (Fezza et al. 2011) was introduced the comparison of MVC prediction structure promoted by HHI with simulcast scheme and IPP structure. The simulcast scheme classified as the most straight forward method by setting up individual video stream and temporally encode independently to each other. The IPP structure proposes the prediction of middle view remain the same (base view) like simulcast temporal prediction structure, other view using inter-view prediction and the video sequence of individual view unable to process independent because they were sharing reference pictures. The results in this experiment to prove that (Fezza et al. 2011) HHI proposed MVC is more efficient because the inter-view prediction is applied to every second view, but it will be more complex.

After the HEVC standard released, (Springer et al. 2014) was developed rapid prototyping to let all interested researcher understand HEVC encoding process by providing a visualisation of encoder decisions. The presented work is proposing real-time video streams support by encoder and decoder to monitor encoder behaviour, e.g. Rate-Distortion Optimisation (RDO) analysis. It was also providing an option for chosen either using HM 14 or x265 encoder. HM and x265 are the H.265 / HEVC video encoder application library. In the com-

parison of HM 14 between x265, we found that x265 is delivered faster encoding speed due to reducing the complexity of the feature that used to increase the encoding efficiency.

Besides that, HTM code is the source code of MV and 3D-HEVC standard which had been widely studied experimentally and numerically by many researchers. They tested different depth prediction model that newly appear in second edition HEVC standard (HTM code) such as Depth Mode Modelling (DMM) (Zhouye Gu et al. 2013, Sanchez et al. 2014), Backward View Synthesis Prediction (B-VSP) and Flexible Coding Order (FCO) (Gopalakrishna et al. 2013). Among the 3D-HEVC researched work, DMM is commonly used due to its better representation of edges in depth maps compared to four new intra prediction modes for depth coding (Tech et al. 2016). The proposed method of (Zhouye Gu et al. 2013) to skip needless DMM full-RD cost calculation based on the selection results of Most Probable Mode (MPM). The proposed work achieves an average of 27.8% encoding time saving with a 0.31% average bitrate loss in coded view (Zhouye Gu et al. 2013). Besides, (Gopalakrishna et al. 2013) also done a similar work with Bi-partition mode exempted since most coding units (CUs) of depth map were flat and smooth, which 34.4% encoding time saving with 0.3% bitrate increasing. In the experiment of (Gopalakrishna et al. 2013), they turned off several depth coding tools such as DMM, Motion Parameter Inheritance (MPI), View Synthesis Optimization (VSO) and Depth Quadtree Prediction, then made a comparison of anchor Neighboring Block Disparity Vector (NBDV) method with FCO and Backward View Synthesis Prediction (B-VSP) separately. The results showed that FCO scheme is ben-

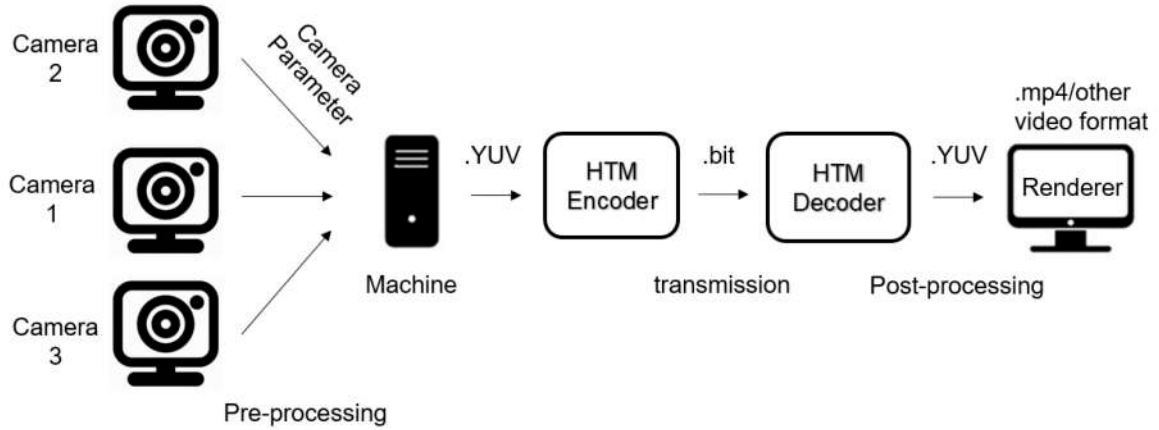


Figure 2: Real Time Multi-view with Depth Video Streaming Approach

eficial for coding dependent textures with low complexity which is better than NBDV method. Moreover, the B-VSP method was implemented on top of depth-based disparity vector derivation and reduced the complexity of decoding by over 8%.

Earlier work by (Sanchez et al. 2014) was doing similar case tested on hardware performance to achieve real-time streaming with 3D-HEVC extension at processing speed 30 frames per second (fps). In this work, they proposed five views of multi-view video streaming with depth. The researches have been focusing on logistic architecture design of DMM which is a insertion of 3D-HEVC extension. They utilised Altera Stratix V FPGA to enhance the performance of the encoding and decoding process, architecture has been described in VHDL and synthesised using Quartus II software. The result shows that proposed architecture able to process more than 6 views for HD 1080P video, more than 3 views for FHD video and more than 18 views for HD 720P.

4 Discussion

Based on our studies, we simplified and grouped the critical information of similar works by several researchers in Table 1. The majority proposed works were tested 3D-HEVC extension to reduce the bitrate and encoding time by modification of different approach. The bit-rate reduction value usually calculated by using Bjøntegaard delta bit rate (BD-rate) method, (Zhouye Gu et al. 2013, Lee et al. 2014, Diaz-Honrubia et al. 2015, Chan et al. 2016) were comparing the BD-rate percentage to determine the proposed model that is significantly reduced with the bit rate. In the-

oretically, the BD-rate method (Ohm et al. 2012) was used to calculate the physical differences between rate-distortion curves, that are the average differences in bit-rate between two curve which figured in percentage. In the used method, the rate-distortion curves of the combined luma and chroma components are used. The average peak signal-to-noise ratio ($PSNR_{YUV}$) is calculated as the weighted sum of the PSNR per picture of the individual components ($PSNR_Y, PSNR_U, and PSNR_V$), the calculation as shown at Eq. 1. The PSNR unit is in (dB).

$$PSNR_{YUV} = (6 \times PSNR_Y + PSNR_U + PSNR_V) / 8 \quad (1)$$

The bit-rate and combined PNSR as the input to calculate Bjøntegaard measurement method. In the results, average difference bit-rate shows more precisely because of consideration of involving the trade-offs between luma and chroma component. Sometime the researcher (Fezza et al. 2011) will use the graph of PSNR against bit-rate to make a comparison of the bit-rate reduction.

In addition, the time reduction also often used to compare the speed of encoding and decoding process in between anchor model (also known as a simulcast in multiple view tasks) and proposed model such as (Zhouye Gu et al. 2013, Gopalakrishna et al. 2013, Diaz-Honrubia et al. 2015, Le et al. 2018). For HEVC, the encoding time is highly increased compared to the previous codec due to its complexity. The reduction of encoding time and remaining the similar quality of the video is also a novelty project. If the project is decided to compare the time reduction between the anchor model and proposed

model, they have to fix the several parameters as constant for example Frame Per Second (FPS), Quantization Parameter (QP), required bit-rate and also same computational power.

However, there is no previous work deals with real-time streaming by using MV-HEVC and 3D-HEVC. (Fezza et al. 2011, Zhouye Gu et al. 2013, Gopalakrishna et al. 2013, Zhouye Gu et al. 2014, Lee et al. 2014, Chan et al. 2016) are only comparing their proposed model with pre-recorded video encoding result. In contrast, most of the project deal with the real-time mode is using single view codec like SHVC and HEVC (Springer et al. 2014, Parois et al. 2016, Viitamäki et al. 2018, Le et al. 2018). Most of them used SHVC as real-time research because of this extension is scalable. When the internet transmission speed is low, the CODEC itself still have another option to complete the transmission which is lower down the video resolution. Unfortunately, SHVC extension unable to encoding in multiple views in a single time, it must set up separately as simulcast video encoding in numerous view.

As a result of survey, the tested real-time video streaming prototype only applies to single viewer HEVC. Multi-view high efficiency video coding such as MV-HEVC and 3D-HEVC yet to be real-time on video streaming without FPGA. Therefore, we proposed an architecture to test on real-time HEVC video streaming, as shown in Fig 2 for further studies. Three cameras with depth vision will be utilized to provide camera parameters and views to be encoded in HEVC, and decoded for multi-view rendering purpose. Then, the proposed architecture will be validated based on the studied method using BD-rate and PSNR in our future work. The proposed architecture provides instant multiple video viewing compared to previous research work which only tested on recorded video.

5 CONCLUSION

In this paper, we have discussed the similar works were doing real-time streaming and multiple views with depth video coding and compared their pros and cons reflect on our proposed work. MV & 3D-HEVC codec was also proposing to encode and decode in near real-time video streaming by setting up a three different view of cameras with depth prototyping on higher performance CPU. Few validation methods like BD-rate, time, QP and PSNR will be considered to used to make a comparison with simulcast real-time multiple views architecture in difference HEVC extension.

Acknowledgment

The authors would like to acknowledge the Computational Intelligence Research Laboratory, Faculty of Computing and Information Technology, Tunku Abdul Rahman University College and Ministry of Education Malaysia (MOE) for the financial support of this research. This research is supported MOE under the Fundamental Research Grant Scheme (FRGS) with project code: FRGS/1/2018/ICT01/TARUC/02/1.

REFERENCES

- Boyce, J. M., Ye, Y., Chen, J. & Ramasubramanian, A. K. (2016), 'Overview of shvc: Scalable extensions of the high efficiency video coding standard', *IEEE Transactions on Circuits and Systems for Video Technology* **26**(1), 20–34.
- Chan, C., Teng, C. & Chuang, J. (2016), New application of mv- and 3d-hevc for multi-intensity illuminated infrared video coding, in '2016 Visual Communications and Image Processing (VCIP)', pp. 1–4.
- CISCO (2019), 'Visual networking index (vni) forecast and trends, 2017–2022 white paper', <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html>. Last accessed 27 February 2019.
- Diaz-Honrubia, A. J., De Praeter, J., Van Leuven, S., De Cock, J., Martinez, J. L. & Cuenca, P. (2015), Using bayesian classifiers for low complexity multiview h.264/avc and hevc hybrid architecture, in '2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)', pp. 1–6.
- Fezza, S. A., Faraoun, K. M. & Ouddane, S. (2011), A comparison of prediction structures for multi-view video coding based on the h.264/avc standard, in 'International Workshop on Systems, Signal Processing and their Applications, WOSSPA', pp. 111–114.
- Gopalakrishna, S., Hannuksela, M. M. & Gabbouj, M. (2013), Flexible coding order for 3d video extension of h.265/hevc, in '2013 Picture Coding Symposium (PCS)', pp. 253–256.

- HHI, F. (2019), '3d video coding', <https://www.hhi.fraunhofer.de/en/departments/vca/research-groups/image-video-coding/research-topics/3d-video-coding.html>.
- Le, T., Van, D. N. & Ryu, E. (2018), Real-time 360-degree video streaming over millimeter wave communication, in '2018 International Conference on Information Networking (ICOIN)', pp. 857–862.
- Lee, H., Kang, J. W., Lee, J. H., Choi, J. S., Lee, J. Y. & Han, J. K. (2014), 'Scalable multiview video coding based on high efficiency video coding for flexible realistic video content services', <https://pdfs.semanticscholar.org/44c1/9a835827ed3b1e9918d416298cd78d0df375.pdf>.
- Ohm, J.-R., Sullivan, G. J., Schwarz, H., Keng Tan, T. & Wiegand, T. (2012), 'Comparison of the Coding Efficiency of Video Coding Standards-Including High Efficiency Video Coding (HEVC)', *IEEE Transactions on Circuits and Systems for Video Technology* **22**(12).
- Parois, R., Hamidouche, W., Mora, E. G., Raulet, M. & Deforges, O. (2016), Demo: Uhd live video streaming with a real-time scalable hevc encoder, in '2016 Conference on Design and Architectures for Signal and Image Processing (DASIP)', pp. 235–236.
- Sanchez, G., Zatt, B., Porto, M. & Agostini, L. (2014), A real-time 5-views hd 1080p architecture for 3d-hevc depth modeling mode 4, in '2014 27th Symposium on Integrated Circuits and Systems Design (SBCCI)', pp. 1–6.
- Springer, D., Schnurrer, W., Weinlich, A., Heindel, A., Seiler, J. & Kaup, A. (2014), Open source hevc analyzer for rapid prototyping (harp), in '2014 IEEE International Conference on Image Processing (ICIP)', pp. 2189–2191.
- Sullivan, G. J., Ohm, J., Han, W. & Wiegand, T. (2012), 'Overview of the high efficiency video coding (hevc) standard', *IEEE Transactions on Circuits and Systems for Video Technology* **22**(12), 1649–1668.
- Tech, G., Chen, Y., Müller, K., Ohm, J., Vetro, A. & Wang, Y. (2016), 'Overview of the multiview and 3d extensions of high efficiency video coding', *IEEE Transactions on Circuits and Systems for Video Technology* **26**(1), 35–49.
- Viitamäki, V., Sjövall, P., Vanne, J., Hämäläinen, T. D. & Kulmala, A. (2018), Live demonstration: 4k100p hevc intra encoder, in '2018 IEEE International Symposium on Circuits and Systems (ISCAS)', pp. 1–1.
- Zhouye Gu, Jianhua Zheng, Nam Ling & Zhang, P. (2013), Fast depth modeling mode selection for 3d hevc depth intra coding, in '2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)', pp. 1–4.
- Zhouye Gu, Jianhua Zheng, Nam Ling & Zhang, P. (2014), Fast bi-partition mode selection for 3d hevc depth intra coding, in '2014 IEEE International Conference on Multimedia and Expo (ICME)', pp. 1–6.