INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2019/M54407
June 2020, Online

Source	**Poznań University of Technology (PUT), Institute of Multimedia Telecommunication, Poznań, Poland**
Status	**Input**
Title	**[VCM] Stereoscopic and multiview video coding for machines**
Author	Marek Domański, Jarosław Samelak, Sławomir Różek, Tomasz Grajek, Sławomir Maćkowiak, Olgierd Stankiewicz

1 Introduction

In the document, we call to consider stereoscopic and multiview video coding in the use cases for video coding for machines. The document comprises a study on stereoscopic and multiview video coding for machines using Screen Content tools.

2 Stresocopic and multiview video for machines

In this document, we suggest to include stereoscopic and multiview video coding into Use Cases and Requirements for Video Coding for Machines in general, or at least for Intelligent Transportation and Intelligent Industry usecases.

The autonomous vehicles are related to an important application of video coding for machines [1]. Intelligent transportation is considered as one of the major use cases. Some of the preliminary experiments have been conducted in the context of autonomous driving [2]. Evaluation framework [3] for VCM includes datasets dedicated to autonomous driving research eg. BD100K [4], CityScapes [5].

However, the expected sub-tasks listed in [1], like object detection, segmentation, tracking, etc. are investigated using only one input view. Even though modern sensors are capable of collecting multiview content, the redundancy of data between the views has not been considered yet in the pipeline for VCM.

Applications and scenarios related to fully connected cars are emerging more and more. We already have infrastructure to vehicle (I2V) and vehicle to infrastructure (V2I) communication to achieve infotainment, on-line navigation, remote diagnostics, safety & security, and communications. The assumption of next generation systems will be communication between vehicles (V2V). These applications will aim to warn drivers and improve traffic efficiency as basic sets of application. Processing of multiview video and depth estimation can support systems like Radar and LiDAR or supersede them in the case of malfunction. Moreover, LiDAR has the disadvantage of high cost, relatively short perception range (~100 m), and sparse information (32, 64 lines comparing to 720p image resolution and even more) [6]. Connected vehicles can share video data and features to improve navigation performance and driving safety. The driverless cars require compression algorithms allowing computer vision based on compressed data [7]. The depth information can be predicted from single view by semantic properties in scenes, object size, etc. However, the inferred depth cannot guarantee the accuracy, especially for unseen scenes. Key image features allowing vision recognition should be carefully preserved.

In the use case of vehicle-to-vehicle video communication, the low delay is also an important requirement. Effective stereo and multiview encoding for inter-vehicle video transmission applications will be crucial.

There are also several use cases in [1] involving AR/VR and Video Game Goggles, Sports Game animation, and Smart Glasses, where efficient multiview video coding seems to be unavoidable. Another example mentioned in [1] is UAV, where multiview video processing and depth estimation can improve mobility in a complex environment. A driver assistance system (in robot navigation, autonomous vehicles) based on the immersive video can inform the driver, and yet increasingly the machine to be prepared for unexpected situations, such as keeping relative distance between preceding cars and the driving vehicle. Scenarios that focus on the use of technology of immersive video for machines should therefore also be considered.

Steroscopic and multiview video also plays crucial role when providing data for depth estimation for mobile robots and robot arm control.

Regarding that, we propose to **add stereoscopic and multiview video coding** in the requirements list, especially in the context of intelligent transportation use case. We also call to modify the proposed architecture for VCM.

3 Stereoscopic and multiview video coding using Screen Content Coding

Multiview video compression can be performed using different approaches. One of them is to use dedicated multi-layer encoder, such as MV-HEVC [8]. This method provides the best compression efficiency, however multi-layer encoders are not popular due to their complexity and limited number of applications. Another solution is simulcast encoding, which means all views that compose multiview video are encoded separately with single-layer encoder, such as HEVC [9, 10]. This technique is much simpler, but it does not exploit similarities between the views.

In [11], the authors have proposed a novel approach for efficient compression of stereoscopic video, using Screen Content Coding (SCC) extension of HEVC [12]. The proposal uses Intra Block Copy (IBC) tool [13] to search for inter-view similarities in a frame-compatible video, simulating the Disparity Compensated Prediction from MV-HEVC.

Application of SCC to multiview video compression requires preparing frame-compatible video. For the case of 3 views, the views are joined side by side in order: middle, left, right, as presented on Fig. 2. This allows to follow the coding order from MV-HEVC Common Test Conditions [14].

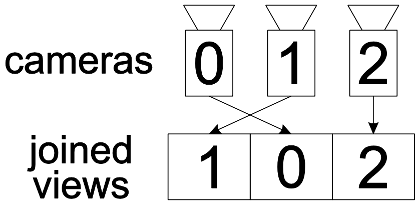

Fig. 2. Preparation of frame-compatible video in multiview system composed of 3 cameras

In order to increase the compression efficiency of SCC encoder for multiview video, a number of changes were introduced into SCC configuration, compared to default configuration from Common Test Conditions [15]:

- Tile encoding – each tile corresponds to a single view. This change allows IBC to use whole previously encoded view as a reference for prediction of remaining views.
- Intra Boundary Filter enabled [12].
- Hash-Based Motion Estimation disabled [12].
- Palette Mode disabled [12, 16].
- Colour Transform disabled [12].

Next section reports experimental results of compression of stereoscopic and multiview video using simulcast HEVC and frame-compatible SCC encoder.

4 Experimental results

The goal of the experiment was to investigate if the proposed SCC encoder would be more efficient in compression of stereoscopic and multiview video for machines, compared to simulcast HEVC. The experiment was conducted on publicly available version of SCC HEVC test model (HM-16.9+SCM-8.0 [17]), using eight commonly used multiview sequences [18-20]. Tests were performed in All Intra and Random Access coding scenarios. Encoders were configured according to Common Test Conditions [15, 21], except for the changes described in Secion 3. The results are presented in Table 1 as an average bitrate reduction, calculated as Bjøntegaard metric for luma PSNR [22].

Table 1. Percentage reduction of bit stream produced by SCC encoder, compared to HEVC simulcast encoding

Sequence	**All Intra**		**Random Access**	
	2 views	**3 views**	**2 views**	**3 views**
Poznan Hall 2	-16.79	-21.08	-11.30	-13.82
Poznan Street	-20.66	-27.53	-13.88	-19.29
Kendo	-20.25	-26.33	-10.30	-13.80
Balloons	-21.49	-30.08	-13.26	-18.44
Newspaper	-18.32	-25.51	-16.00	-20.46
Undo Dancer	-35.14	-47.16	-25.82	-34.32
GT Fly	-38.56	-50.71	-21.74	-31.67
Shark	-35.39	-49.78	-29.50	-40.39
Average	**-26.28**	**-34.56**	**-16.61**	**-22.58**

The results show that using Screen Content Coding for compression of stereoscopic and multiview video provides significant gain, compared to simulcast HEVC. Moreover, the more views are encoded, the greater is the percentage bitrate reduction. Other our results show that Screen Content Coding for HEVC can be easily upgraded by sub-pixel vectore resolution. In that case, its applications to Multiview video coding provides virtually the same results as MV-HEVC [23].

5 Conclusions

Stereoscopic and multiview video coding should be considered as an important sub-task for Video Coding for Machines.

The available technology for stereoscopic and multiview video coding is either MV-HEVC and 3D-HEVC or Screen Content Coding of HEVC/VVC used for multiview frames.

The latter technology does not need multi-layer codecs and is compatible with some other contributions that propose to use Screen Content Coding for Video Coding for Machines. Therefore such technology should be taken for further considerations.

6 Acknowledgment

This work was supported by Project 0314/050.

7 References

[1] "Use cases and draft requirements for Video Coding for Machines", ISO/IEC JTC1/SC29/WG11 Doc. w19365, Alpbach, Austria, October 2020.

[2] Yi Fan, "Task specific Compression for Lane Detection", ISO/IEC JTC1/SC29/WG11 MPEG2019/m51586, Brussells, Belgium, January 2020.

[3] "Evaluation Framework for Video Coding for Machines", ISO/IEC JTC1/SC29/WG11 Doc. w19366, Alpbach, Austria, October 2020.

[4] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, T. Darrell, "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2020.

[5] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2016.

[6] I. Ashraf, et al., "An investigation of interpolation techniques to generate 2D intensity image from LIDAR data", *IEEE Access*, vol.5, Apr. 2017, pp.8250-8260.

[7] A. Geiger, P. Lenz, R. Urtasun, "Are we ready for autonomous driving? The KITTI Vision benchmark suite", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2012, pp. 3354–3361.

[8] G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro and Y. K. Wang, "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding", in *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 26, no. 1, pp. 35-49, Jan. 2016.

[9] ISO/IEC Int. Standard 23008-2: 2015 “High efficiency coding and media delivery in heterogeneous environment – Part 2: High efficiency video coding” and ITU-T Rec. H.265 (V3) (2015), „High efficiency video coding”.

[10] G. J. Sullivan, J. Ohm, W. J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard”, in *IEEE Transactions on Circuits Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[11] J. Samelak, J. Stankowski, M. Domański, “Efficient frame-compatible stereoscopic video coding using HEVC Screen Content Coding”, *IEEE International Conference on Systems, Signals and Image Processing IWSSIP 2017*, Poznań, Poland, May 2017.

[12] J. Xu, R. Joshi, and R. A. Cohen, “Overview of the Emerging HEVC Screen Content Coding Extension”, in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 50-62, Jan. 2016.

[13] X. Xu et al., “Intra Block Copy in HEVC Screen Content Coding Extensions”, in *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 4, pp. 409-419, Dec. 2016.

[14] K. Müller, A. Vetro, “Common Test Conditions of 3DV Core Experiments”, JCT-3V of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 7th Meeting: Doc. JCT3V-G1100, San José, US, Jan. 2014.

[15] H. Yu, R. Cohen, K. Rapaka, J. Xu, “Common Test Conditions for Screen Content Coding”, JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 21st Meeting: Doc. JCTVC-U1015r2, Warsaw, PL, Jun. 2015.

[16] W. Pu et al., “Palette Mode Coding in HEVC Screen Content Coding Extension”, in *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 4, pp. 420-432, Dec. 2016.

[17] JCT-VC, HEVC Screen Content Coding reference software repository, https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftw are/tags/HM-16.9+SCM-8.0. Web. 15 Dec. 2016.

[18] Y.S. Ho, E.K. Lee, C. Lee, “Multiview video test sequence and camera parameters”, ISO/IEC JTC1/SC29/ WG11 MPEG Doc. M15419, Archamps, France, Apr. 2008.

[19] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, “Poznań multiview video test sequences and camera parameters”, ISO/IEC JTC1/SC29/WG11 MPEG Doc. M17050, Xian, China, Oct. 2009.

[20] M. Tanimoto, T. Fujii, N. Fukushima, “1D parallel test sequences for MPEG-FTV”, ISO/IEC JTC1/SC29/WG11, MPEG Doc. M15378, Archamps, France, Apr. 2008.

[21] F. Bossen, “Common Test Conditions and software reference configurations”, JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 12th Meeting: Doc: JCTVC-L1100, Geneva, CH, Jan. 2013.

[22] G. Bjøntegaard, “Calculation of Average PSNR Differences between RD-curves”, ITU-T SG16, Doc. VCEG-M33, Austin, USA, Apr. 2001.

[23] J. Samelak, M. Domański, Unified Screen Content and Multiview Video Coding - Experimental results, JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 12th Meeting: Doc: JCTVC- M0765, Marrakech, MA, 9–18 Jan. 2019, also ISO/IEC JTC1/SC29/ WG11 MPEG Doc. M46332.