INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 2
MPEG TECHNICAL REQUIREMENTS

ISO/IEC JTC 1/SC 29/WG 2 M55108
Online – October 2020

Title: **[VCM] Test material for stereoscopic and multiview video coding for machines**
Source: **WG 2 MPEG Technical requirements**
Author(s): **Marek Domański, Tomasz Grajek, Sławomir Maćkowiak, Dawid Mieloch, Sławomir Różek, Olgierd Stankiewicz, Jakub Stankowski**
Poznań University of Technology, Poznań, Poland
Status: **Input**

Abstract
This document is on video test sequences for the use in experiments related to multiview and stereoscopic video coding for machines. It describes sequence parameters and shows sample views taken from the sequences. Exemplary segmentation results have been also presented.

1. Introduction

Prospective technology of Video Coding for Machines is expected to be used within a wide range of use cases described in [1]. For many of them (i.e. surveillance, intelligent transportation, smart city, intelligent industry) stereoscopic and multiview video processing was proposed [2] as one of the expected sub-tasks.

Evaluation Framework for VCM document [3] describes many datasets for VCM experiments. However, there is still a lack of high definition, multiview video sequences with accurate depth maps. Moreover, some of them cannot be utilized for standardization due to licensing issues.

Therefore, we propose to provide new test materials and include it into the recommended datasets in the Evaluation Framework for Video Coding for Machines. Proposed data contains high resolution multiview video sequences with camera parameters (intrinsic and extrinsic) and depth maps. Videos were recorded by well-synchronized HD cameras and contain mostly outdoor, surveillance scenes.

2. Group A – Stereoscopic test video sequences

The specifications of used cameras are as follows:

Basler Aviator – avA1900-50gm/gc	
Resolution Size (H x V pixels)	1920 x 1080
Sensor	Kodak KAI-02150
Sensor Size (optical)	2/3"
Sensor Technology	Progressive Scan CCD, Global Shutter
Pixel Size	5.5 x 5.5 µm
Frame Rate	51 fps
Mono/Color	Mono/Color
Video Output Format	Mono 8, Mono 16, Mono 12 packed, YUV 4:2:2, Raw 8, Bayer BG8, Bayer BG16, Bayer BG12 Packed
Interface	Gigabit Ethernet (1000 Mbit/s)
Synchronization	Via external trigger, via the Ethernet connection, or free-run
Exposure Control	Freely programmable
Housing Size (L x W x H)	40.7 mm x 62 mm x 62 mm
Housing Temperature	Up to 50°C
Lens Mount	C-mount
Digital I/O	2 opto-isolated input / 4 opto-isolated output
Power Requirements	12 VDC (+/- 10%) via 12-pin Hirose connector
Power Consumption (typical)	5.8 W
Weight (typical)	< 300 g

2.1. PoznanHighway test sequence

- PoznanHighway (Fig. 2.1) – A2 Highway near Poznan,
  250 frames, outdoor sequence

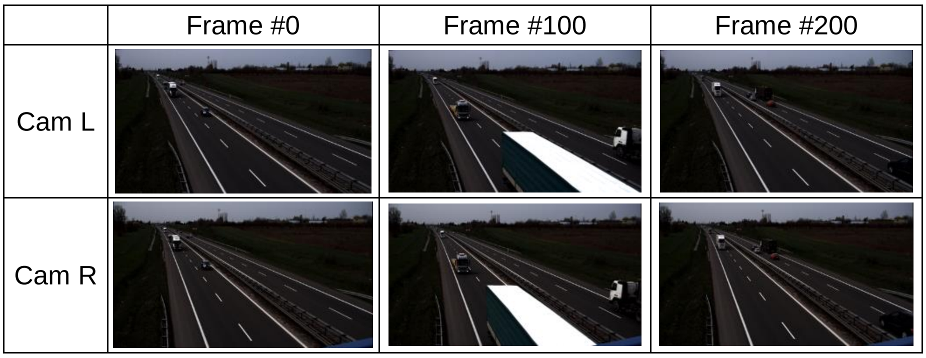


Fig. 2.1. Selected frames of PoznanHighway sequence.

2.2.PoznanFootbridge test sequence

- PoznanFootbridge (Fig. 2.2) – Footbridge over Baraniaka street,
  250 frames, outdoor sequence

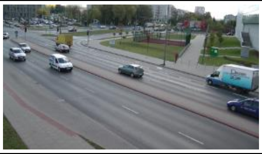
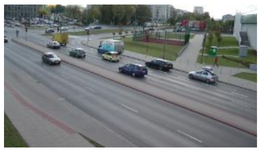
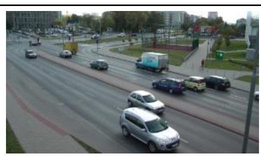
	Frame #0	Frame #100	Frame #200
Cam L			
Cam R			

Fig. 2.2. Selected frames of PoznanFootbridge sequence.

3. Group B – Multiview test video sequences

The proposed test sequences were already accepted by MPEG as test sequences for 3D-HEVC-related experiments [5] and recently are part of Common Test Conditions for Immersive Video [6].

The specifications of used cameras are as follows:

- 3-chip 1/3” CCD sensors,
- progressive scan RGB camera,
- HDTV resolution (1920 × 1088),
- 25 frames / second,
- Frame and shutter synchronization by TTL trigger,
- Parameter control via LANC interface,
- Frame timestamp output,
- SDI signal output,
- Dimensions (W × H × L) 163 × 189 × 350 mm.

The proposed depth maps that are available for all views were calculated using the Immersive Video Depth Estimation method [7], based on [8]. The depth maps were also enhanced using the PDR software [9].

3.1. PoznanStreet test sequence

- PoznanStreet (Fig. 3.1 and 3.2) – Polanka Street near Poznan University of Technology building, 50 frames, outdoor sequence

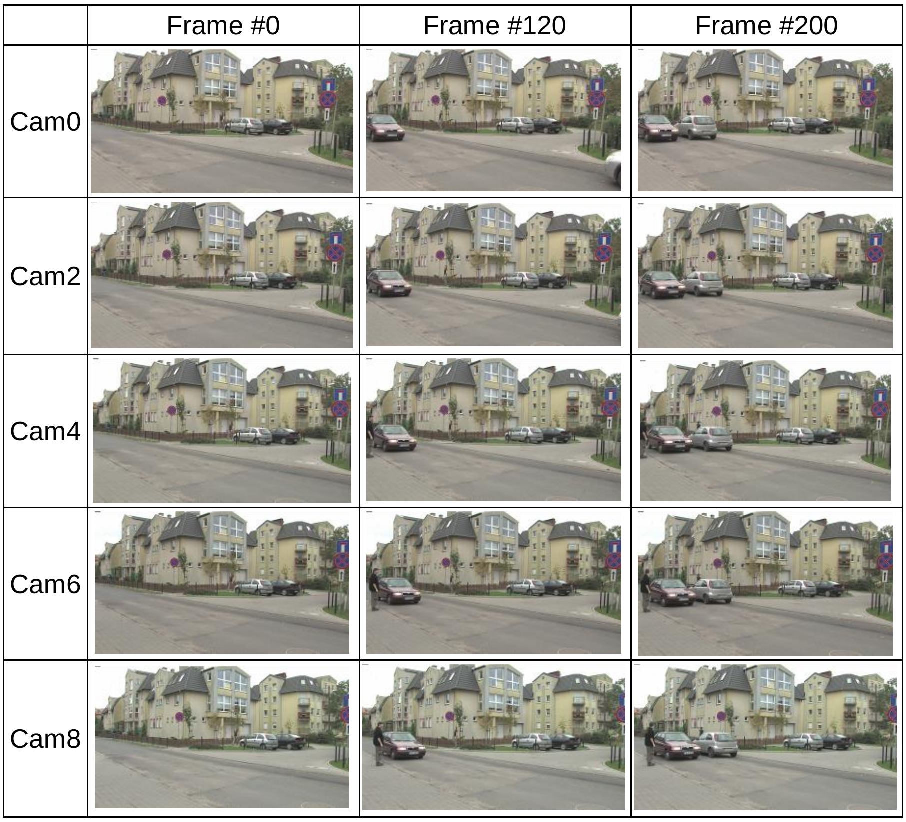


Fig. 3.1. Selected views and frames of PoznanStreet sequence.

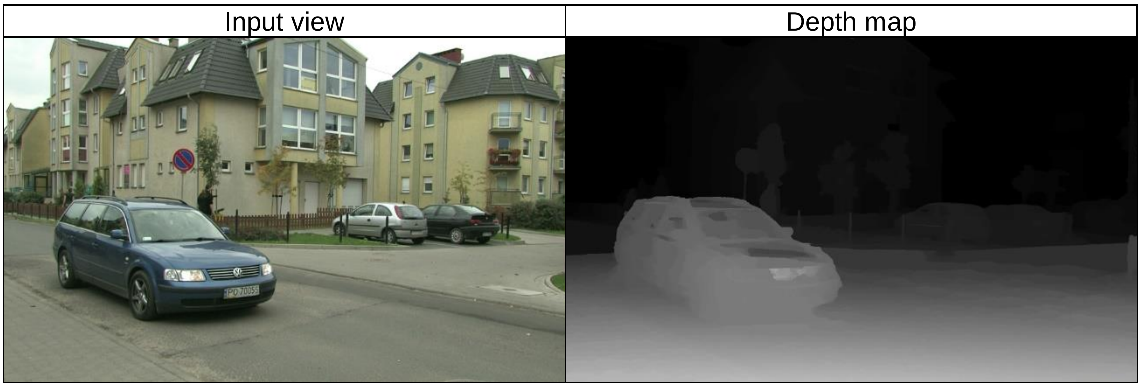


Fig. 3.2. Selected view and depth map of proposed sequence.

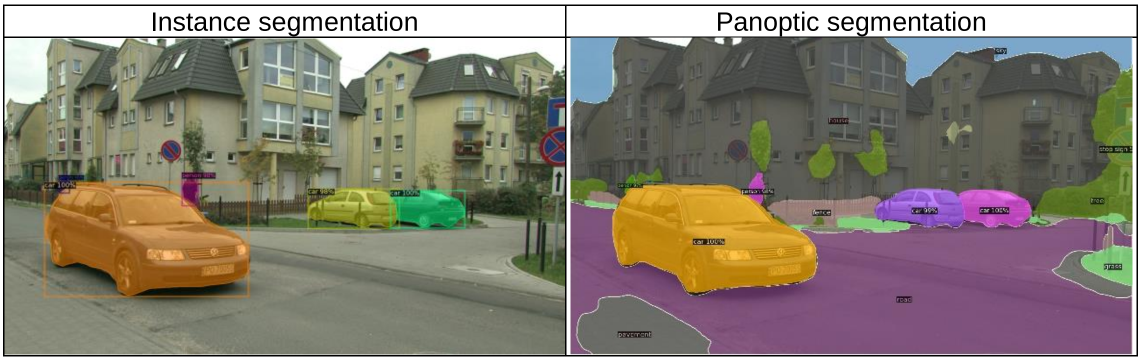


Fig. 3.3. Exemplary segmentation done onto selected view of proposed sequence

3.2. PoznanHall test sequence

- PoznanHall (Fig. 3.4 and 3.5) – Poznan University of Technology Hall
  500 frames, indoor, moving camera

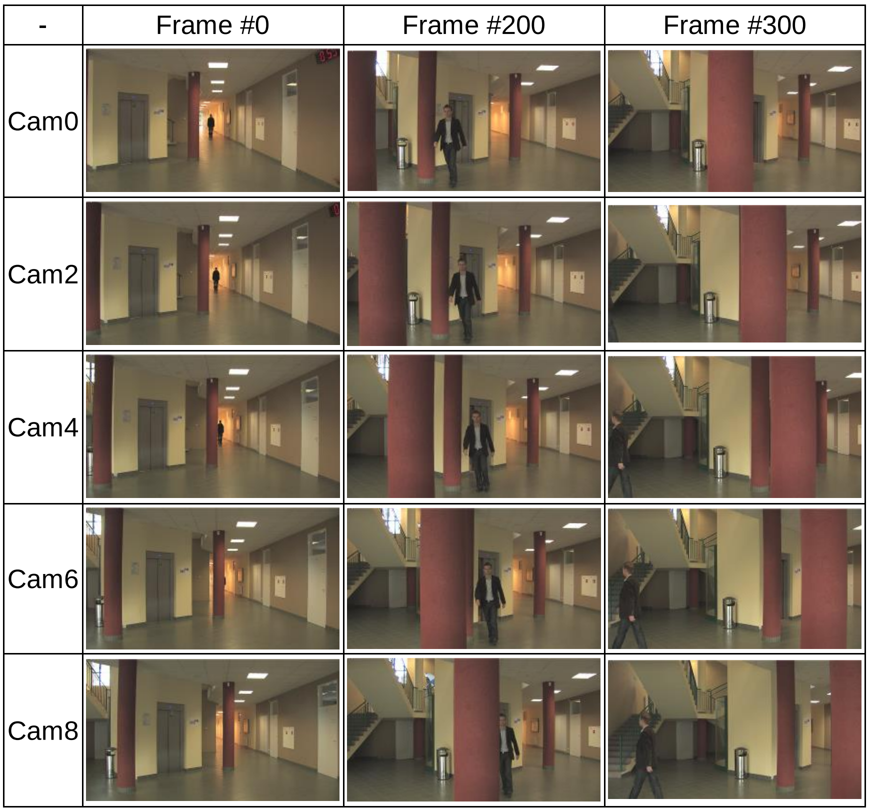


Fig. 3.4. Selected views and frames of PoznanHall sequence.

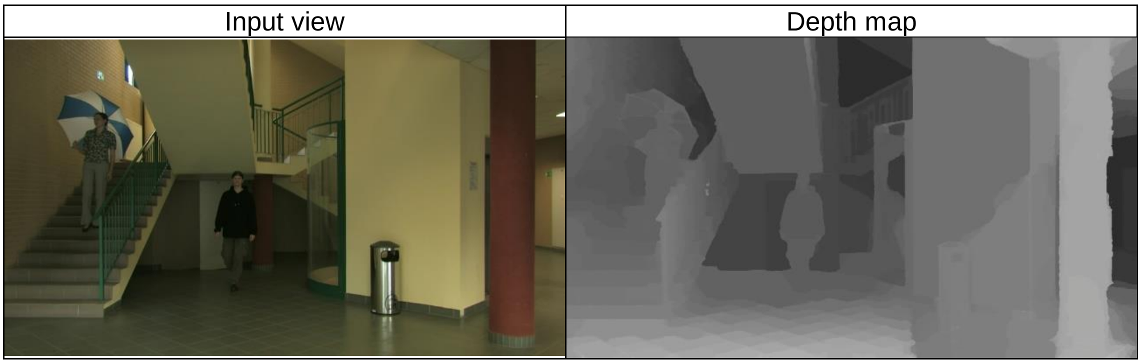


Fig. 3.5. Selected view and depth map of proposed sequence.

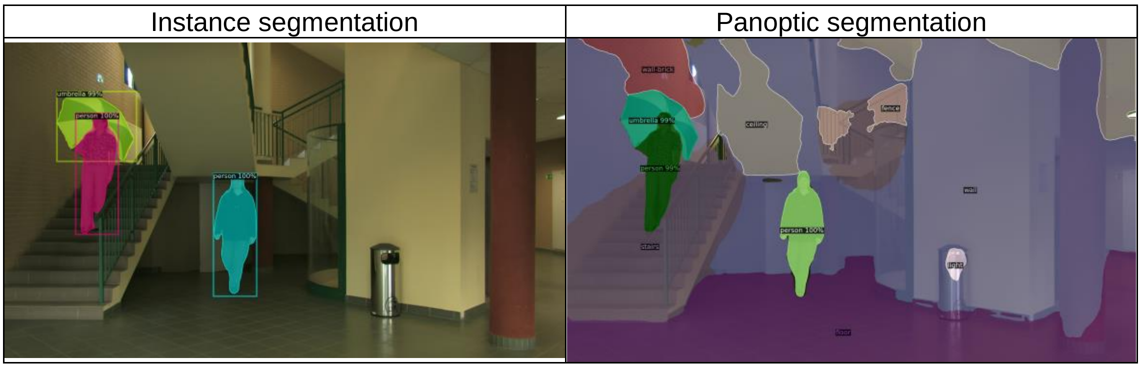


Fig. 3.6. Exemplary segmentation performed onto selected view of proposed sequence.

3.3.PoznanCarpark test sequence

- PoznanCarpark (Fig. 3.7 and 3.8) – Poznan University of Technology Car Park
  250 frames, outdoor

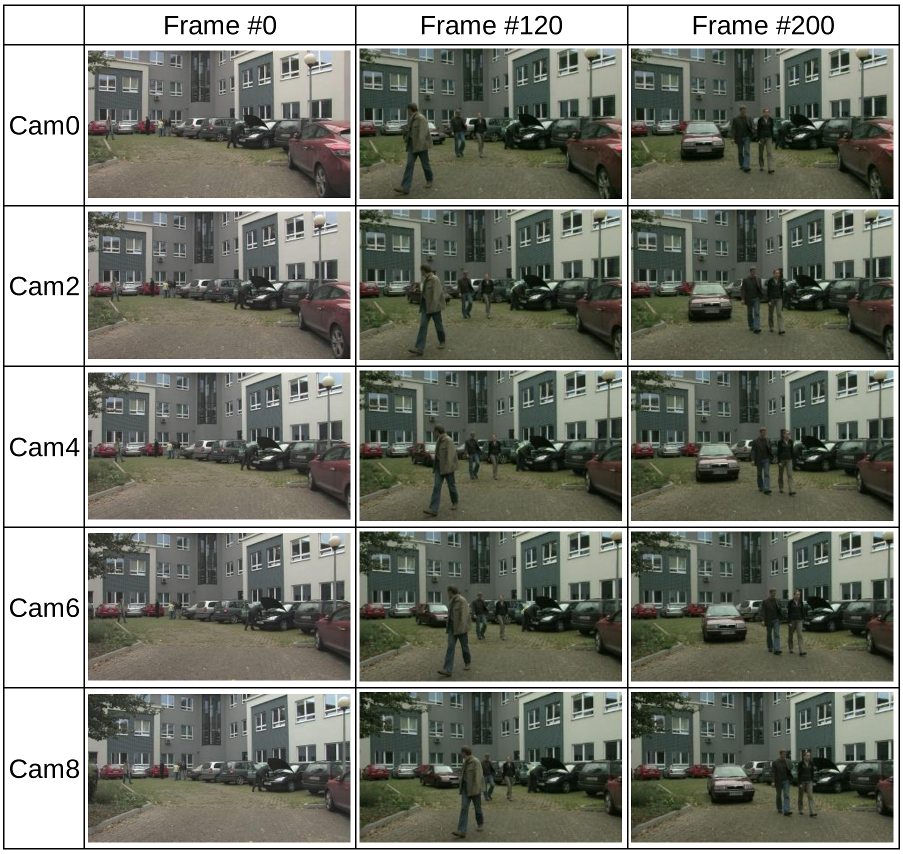


Fig. 3.7. Selected views and frames of PoznanCarpark sequence.

Fig. 3.8. Selected view and depth map of proposed sequence.

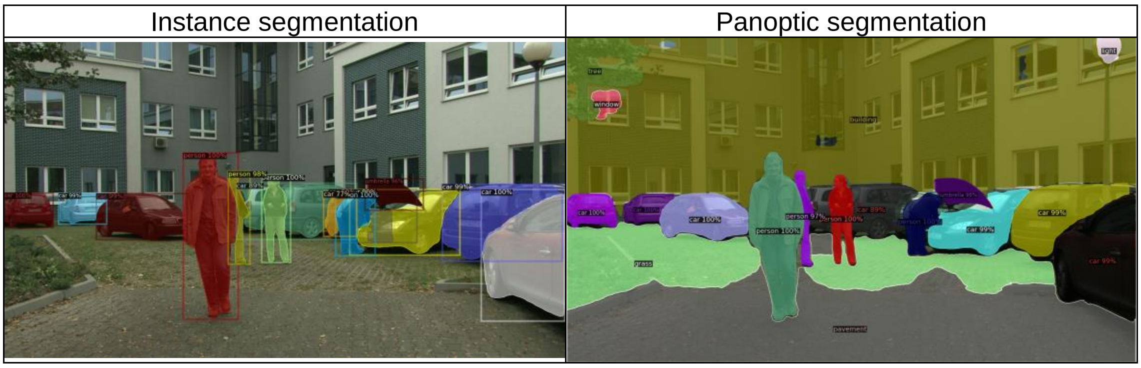


Fig. 3.9. Exemplary segmentation performed onto selected view of proposed sequence.

4. Conclusions

Proposed test sequences should be considered as a valuable dataset in the VCM evaluation framework. Presented materials complement the recommended datasets. The content match Surveillance and Smart City use cases. There are no licensing issues with using proposed videos, as Poznań University of Technology has released the sequences for standardization purposes. Depth maps for each view are available and accurate. Such data can be useful for development and testing multiview video compression path inside the VCM encoder.

5. Recommendations

We recommend including proposed sequences into recommended datasets in the Evaluation Framework for Video Coding for Machines.

6. Acknowledgment

This work was supported by Project 0314/050.

7. References

[1] "Use cases and requirements for Video Coding for Machines," Doc. ISO/IEC JTC1/SC29/WG11 N19506, June 2020.

[2] M.Domański, J. Samelak, S. Różek, T. Grajek, S. Maćkowiak, O. Stankiewicz,”[VCM] Stereoscopic and multiview video coding for machines”, Doc. ISO/IEC JTC1/SC29/WG11 M54407, June 2020.
[3] “Draft Evaluation Framework for Video Coding for Machines,” Doc. ISO/IEC JTC1/SC29/WG11 N19507, June 2020.
[4] M. Domański, T. Graje, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, “Poznań Multiview Video Test Sequences and Camera Parameters”, Doc. ISO/IEC JTC1/SC29/WG11, MPEG/M17050 M17050, Xian, China, 26-30 October 2009.
[5] K. Müller i A. Vetro, “Common Test Conditions of 3DV Core Experiments” JCT-3V (ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11) dok. JCT3V_G1100, San Jos-e, USA. Sty. 2014.
[6] “Common Test Conditions for MPEG Immersive Video”, ISO/IEC JTC1/SC29/WG11 MPEG2020/N19484, Online, July 2020.
[7] “Manual of Immersive Video Depth Estimation”, ISO/IEC JTC1/SC29/WG11 MPEG 2020/N19224, Online, May 2020.
[8] D. Mieloch, O. Stankiewicz and M. Domański, “Depth Map Estimation for Free-Viewpoint Television and Virtual Navigation,” IEEE Access, vol. 8, pp. 5760-5776, 2020.
[9] “Manual of depth refinement software PDR”, ISO/IEC JTC1/SC29/WG11 MPEG2019/N18708, Gothenburg, July 2019.
[10] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo and R. Girshick. Detectron2. 2019. https://github.com/facebookresearch/detectron2