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 m56680
Online – April 2021

Title: **[VCM] Depth estimation from compressed video in VCM-related applications**
Source: **WG 2 MPEG Technical requirements**

Author(s):
Olgierd Stankiewicz, Marek Domański, Tomasz Grajek,
Sławomir Różek, Dominik Cywiński, Jakub Szekiełda
Poznań University of Technology, Poznań, Poland

Abstract

In this contribution, we consider a use case for video coding for machines, related to autonomous driving, where depth estimation is performed from remotely acquired video. We present the results of experiment, in which in video is encoded with various QP values/bitrate/quality, and then is used for depth estimation. We show the relation between the video quality and the accuracy of estimated depth for various objects: trees, cars and people.

1. Introduction

The application scenario proposed for VCM [1] during 131st MPEG Meeting was related to multiview and stereoscopic video. In the context of such applications, an important problem is the dependency of the accuracy of depth estimation from compressed video. The influence of compression on depth estimation was studied in several papers in the context of fidelity of virtual view synthesis rather than the accuracy of the distance itself. Therefore, the relevant study is included into this work.

2. Experiment configuration

Sequences:

The experiment has been performed with use of the following sequences:

- Poznan Street, Poznan Carpark (MPEG test multiview test sequences) [2].

Objects :

Depth has been estimated for objects selected from compressed video (manual labelling):

- 7 objects have been selected from Poznan Street sequence (Fig. 1),
- 12 objects have been selected for Poznan Carpark sequence (Fig. 2).

Codec:

The HEVC codec was configured according the relevant MPEG configuration files [4].

Ground truth depth:

Ground truth was estimated from the measurements (from the scene and uncompressed video) as shown in Table 1.

Depth estimation software:

Software is included (on www.ieee.org) as appendix to the paper [3].

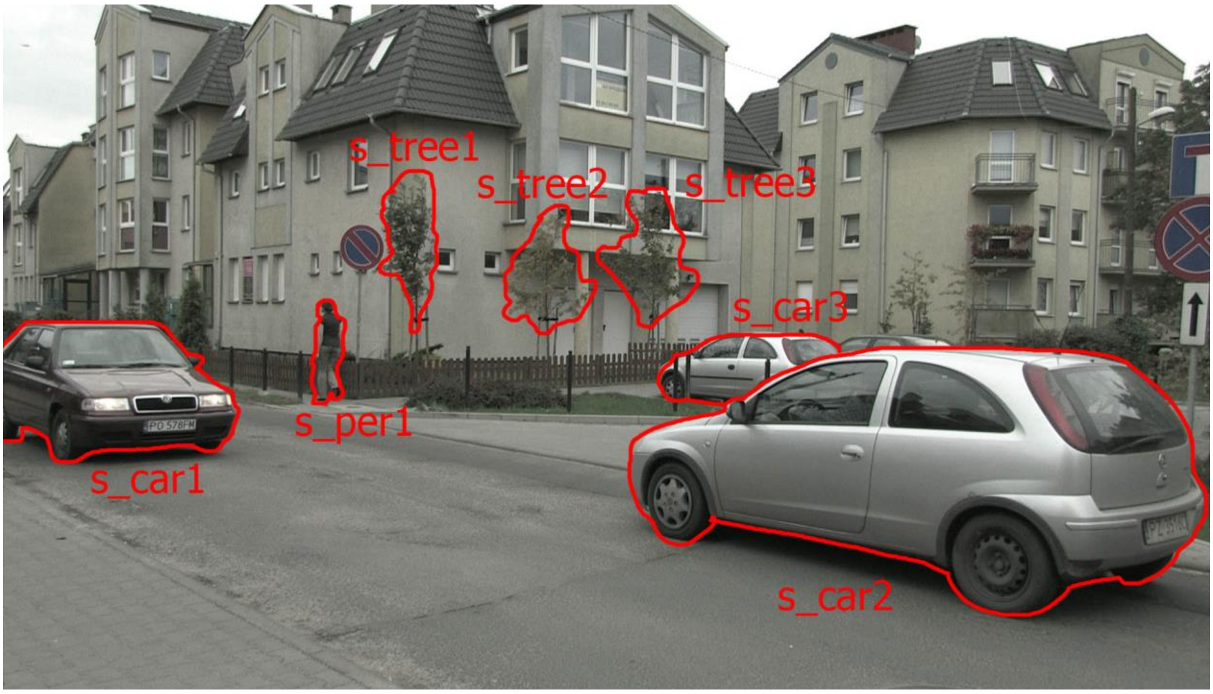


Fig 1. Objects selected for evaluation in Poznan Street sequence.

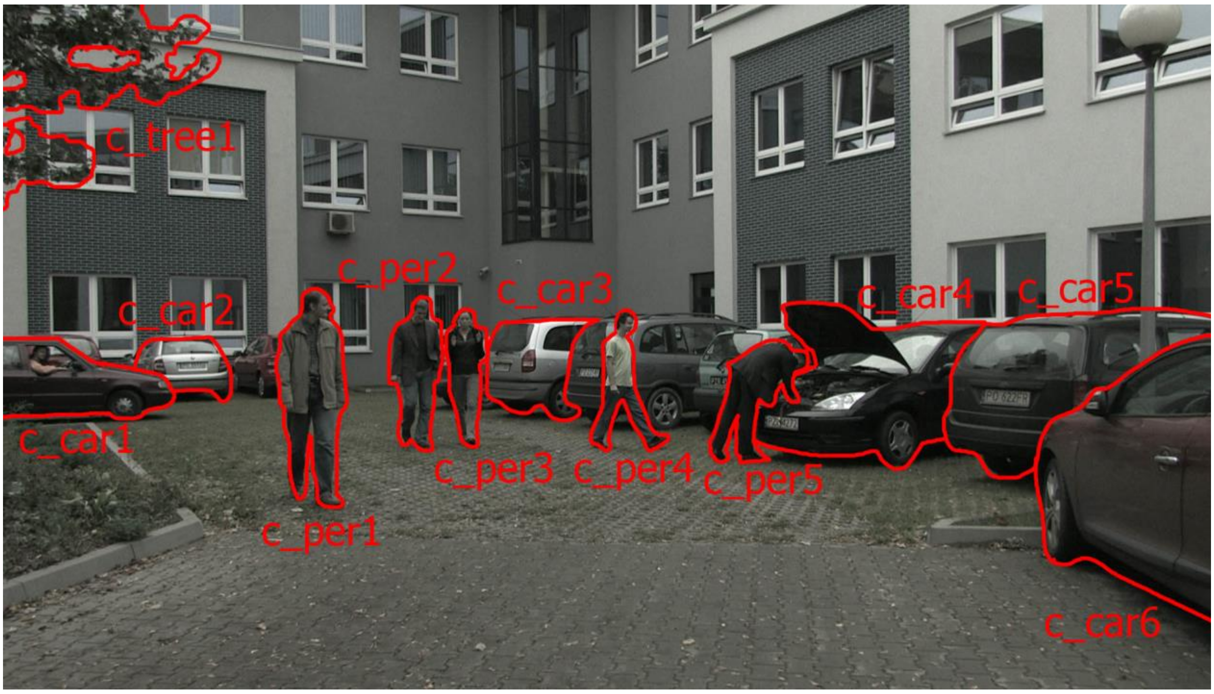


Fig 2. Objects selected for evaluation in Poznan Carpark sequence.

Table 1. Depth of the objects selected for evaluation.

Sequence	Object class	Object	*Z* distance [m]
Poznan Street	Person	s_per1	17.79
	Car	s_car1	11.24
		s_car2	7.46
		s_car3	20.71
	Tree	s_tree1	19.65
		s_tree2	21.58
		s_tree3	23.16
Poznan Carpark	Person	c_per1	8.78
		c_per2	12.98
		c_per3	13.10
		c_per4	12.55
		c_per5	11.25
	Car	c_car1	17.04
		c_car2	22.37
		c_car3	16.48
		c_car4	10.80
		c_car5	8.88
		c_car6	5.71
	Tree	c_tree1	5.72

3. The course of the experiment

Experiment has been performed as follows:

Reference:

- Depth (normalized disparity map) has been estimated for the original (uncompressed) sequences.
- Basing on the estimated disparity maps, depths (Z values) of particular objects (Table 1) have been extracted.
- This constitutes reference depth values for objects.

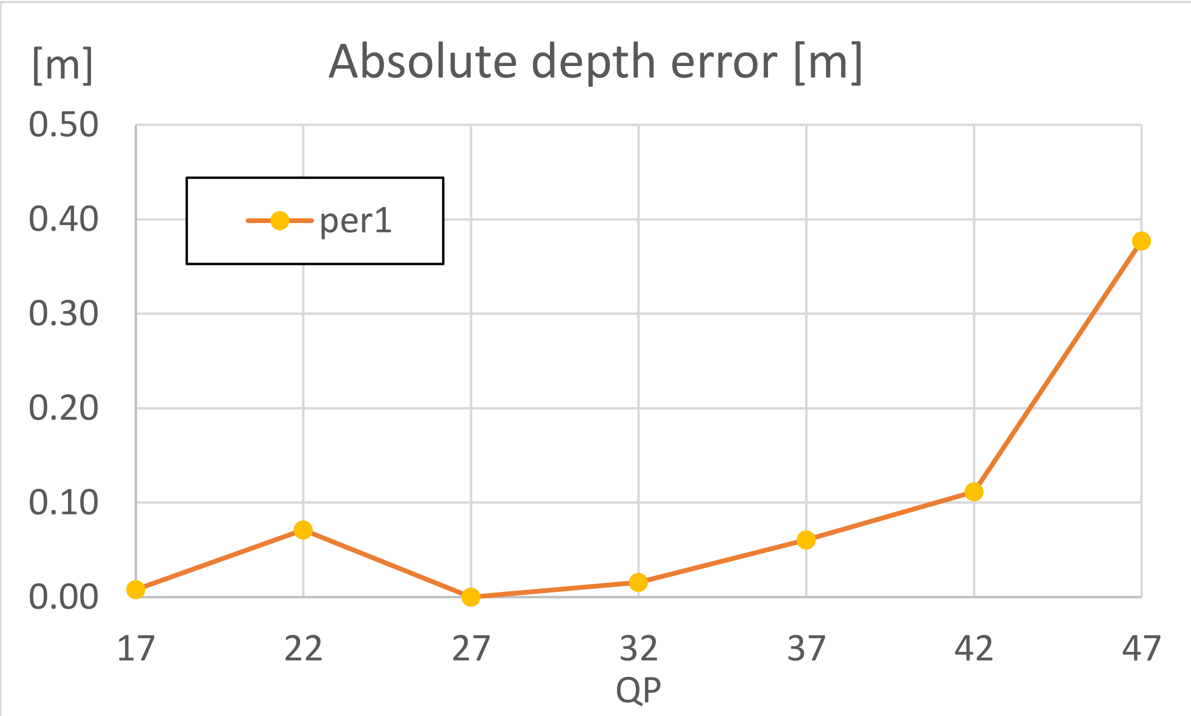
Test Cases:

- Images (texture) of the original sequences has been encoded with HEVC with given quantization parameter QP.
- Depth (normalized disparity map) has been estimated for the reconstructed sequences.
- Basing on the newly estimated disparity maps, depths (Z values) of particular objects have been extracted.
- Attained depth values have been compared with the reference values.

4. Results

4.1. Absolute depth errors of selected objects

Object s_per1 (a person):



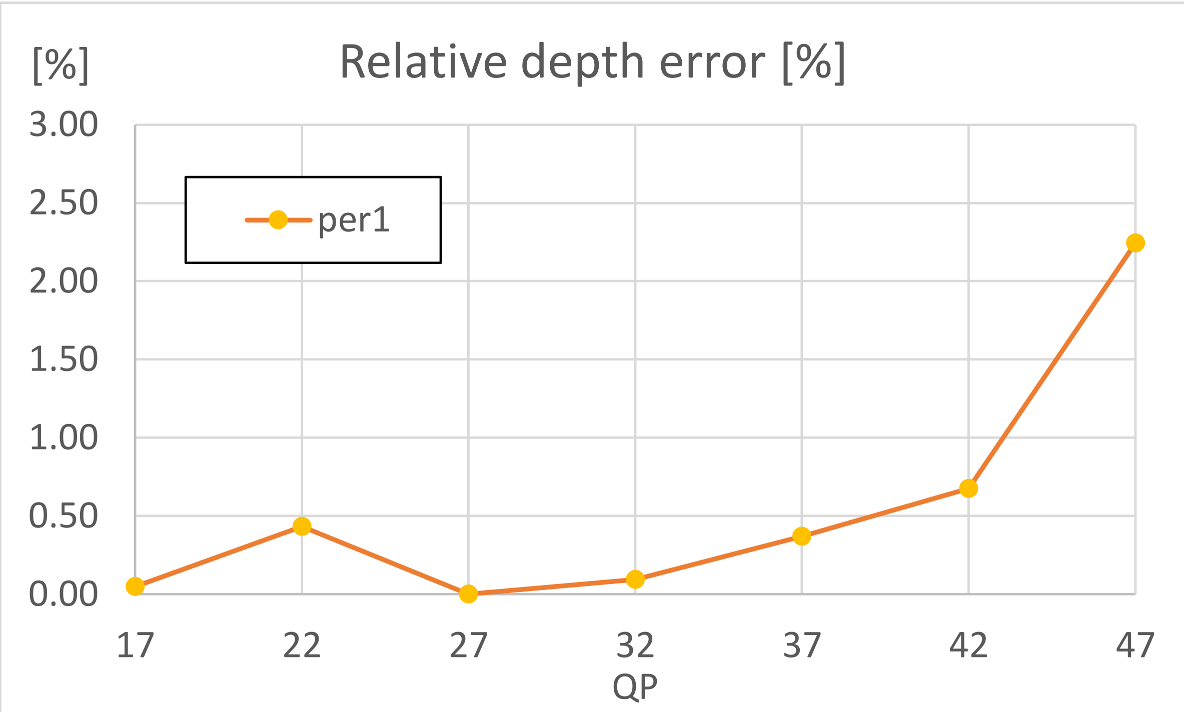
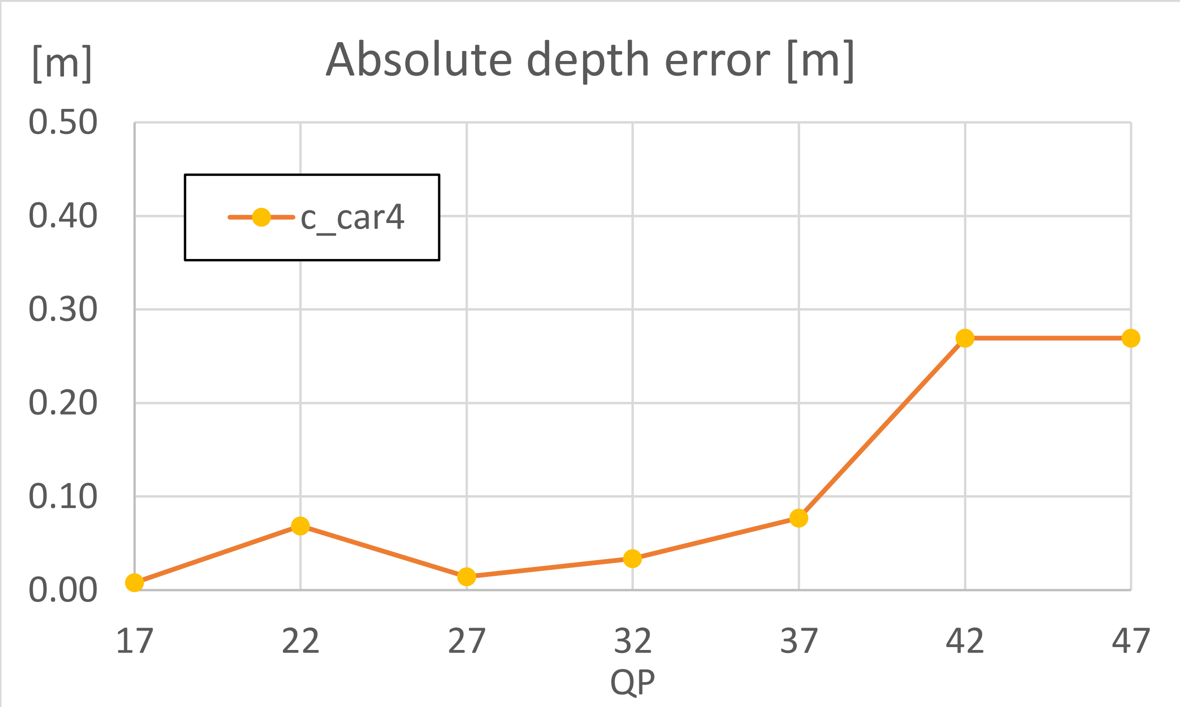


Fig. 3. Absolute and relative depth errors for the object s_per1.

Object: c_car4:



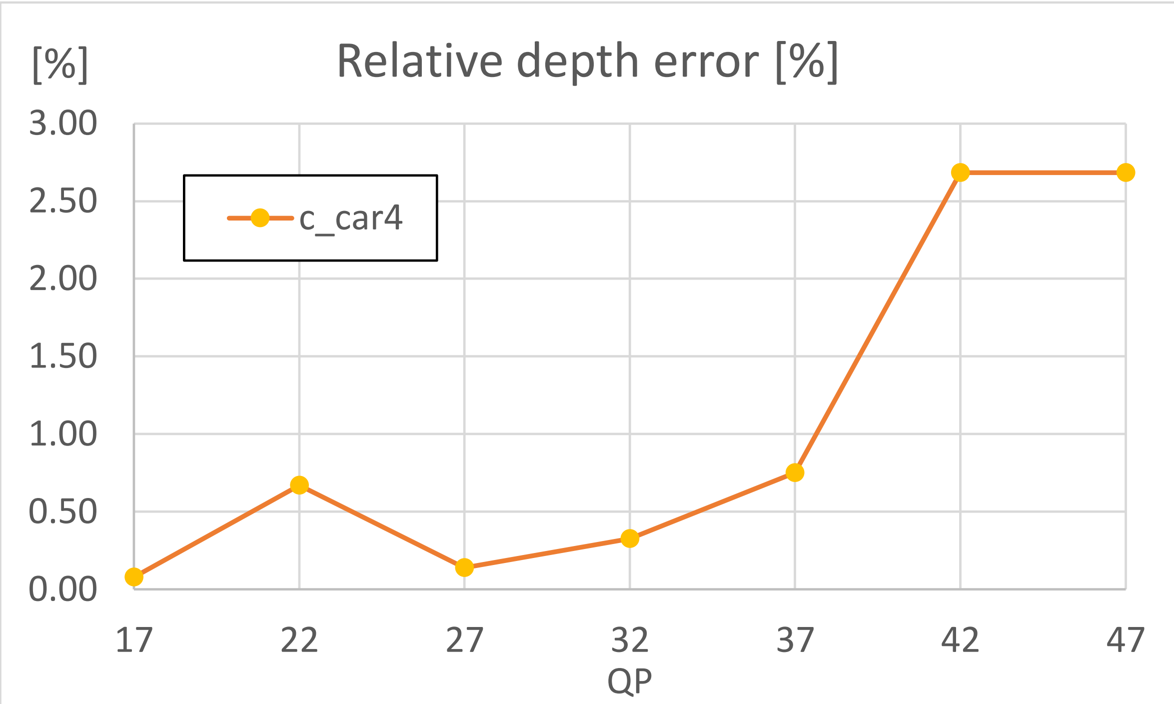
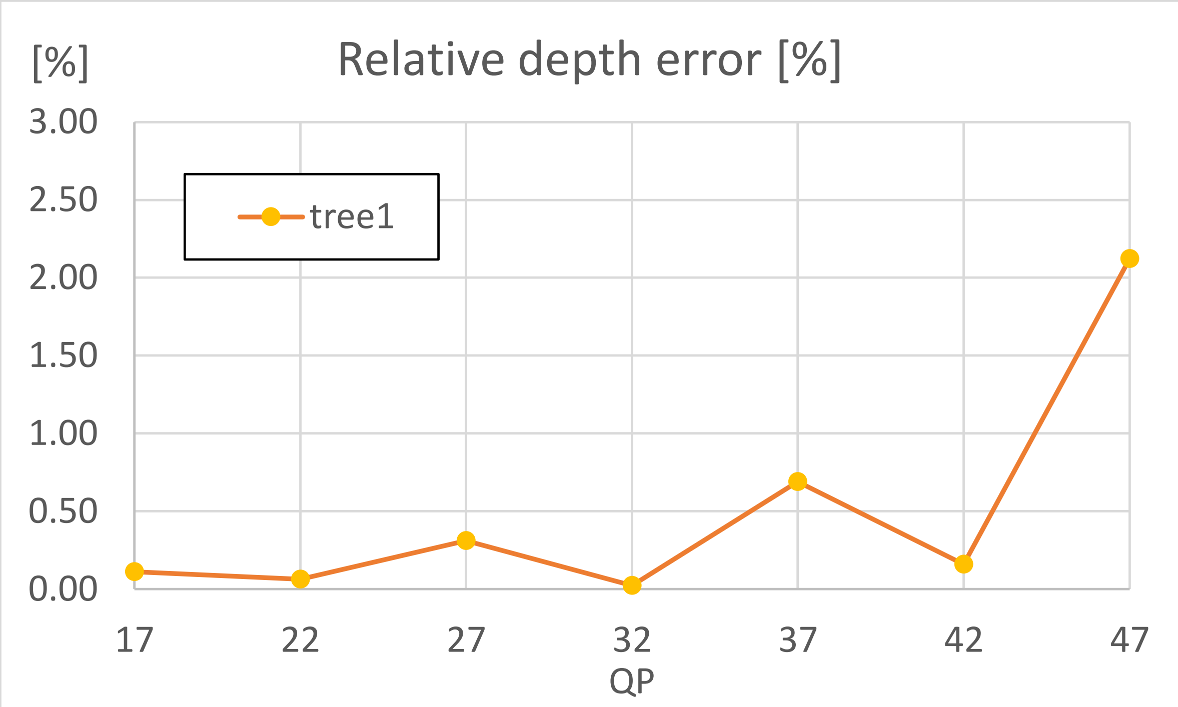


Fig. 4. Absolute and relative depth errors for the object c_car1.

Object: c_tree1:



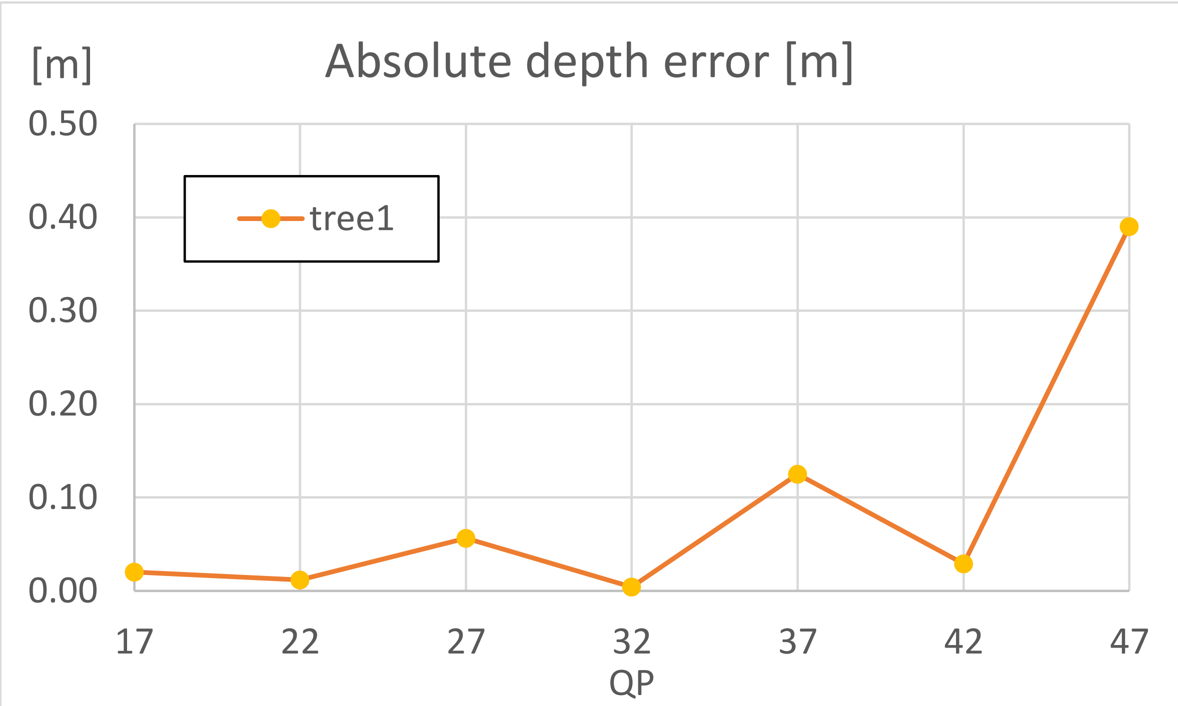
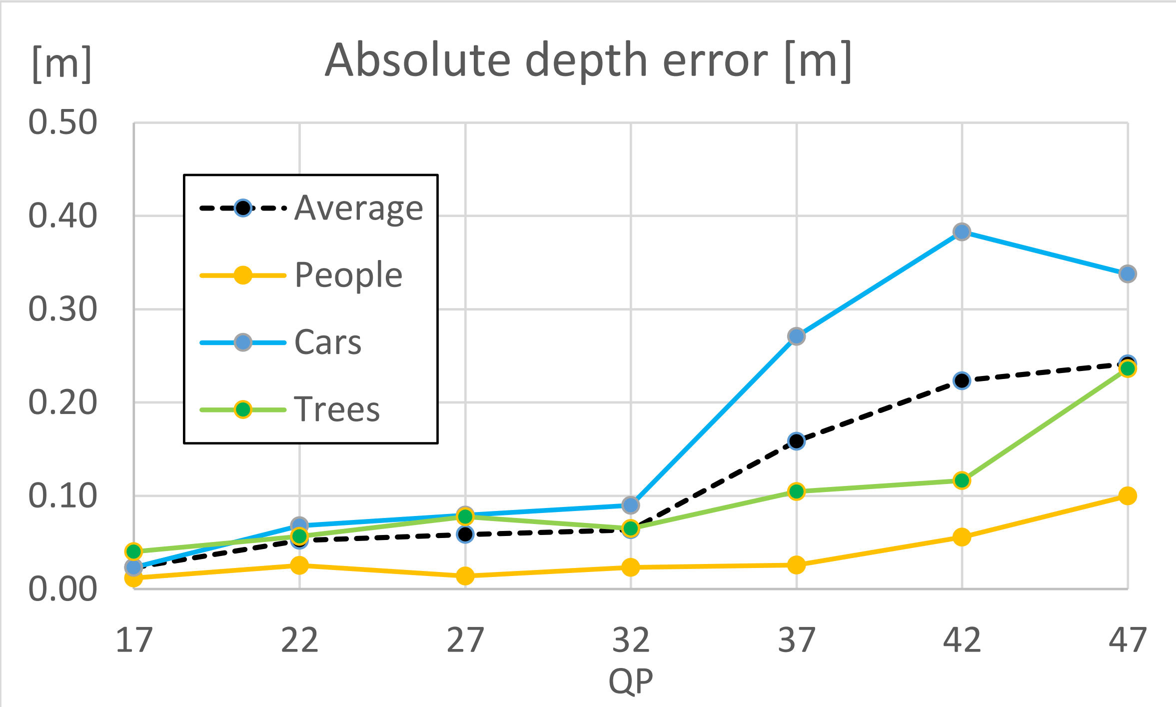


Fig. 5. Absolute and relative depth errors for the object c_tree1.

4.2. Aggregated results for all objects

Average errors for all objects in the 3 classes:



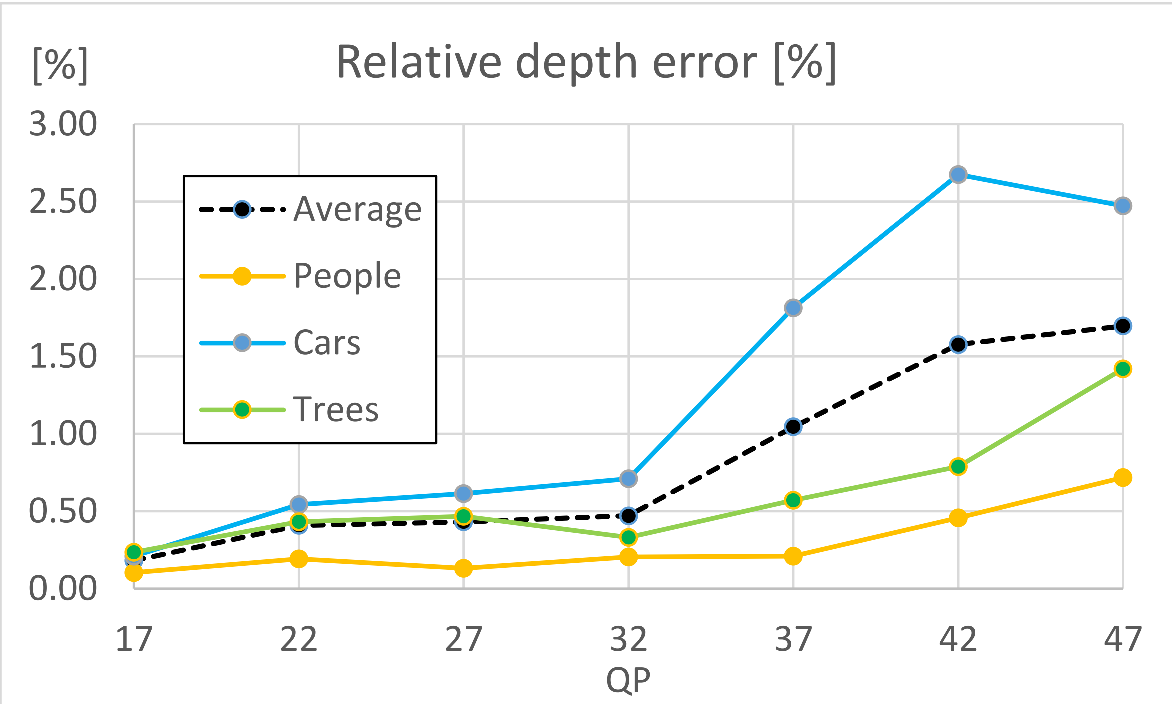


Fig. 6. Absolute and relative average depth errors for the all objects in the respective classes.

5. Conclusions

For the state-of-the-art depth estimation techniques, the accuracy of depth estimation is satisfactory for a wide range of moderate quantization parameter (QP) values. Even strong compression does not yield significant errors.

The measured depth errors depend on objects classes. The highest observed depth error have been observed for cars: approximately 0.4m (absolute error) and 2.5% (relative error). This phenomenon can be explained by the shiny surfaces of cars bodies, which increase difficulty of depth estimation. Significantly lower depth error have been observed for people and trees: approximately 0.1-0.2m (absolute error) and 0.5-1.5% (relative error).

Independently from object classes, depth error increases with QP parameter value. This slope of this increase is rather low below QP=32. Above QP=32, depth errors start to increase more rapidly, however are still limited to a level acceptable in the considered application (2.5% relative error).

Therefore, it may be concluded that in the considered use case scenario even strong compression may be used. Nevertheless, the issue requires further study.

An interesting direction of further research may use the proposals from MPEG-I where transmission of some metadata as side information is considered in order to reduce depth errors that result from estimation from decoded video.

Acknowledgment

This work was supported by Project 0314/050.

6. References

[1] Marek Domański, Tomasz Grajek, Sławomir Maćkowiak, Sławomir Różek, Olgierd Stankiewicz, Jakub Stankowski, "[VCM] Requirements for stereoscopic and multiview video in VCM environment", ISO/IEC JTC 1/SC 29/WG 2 M56105, Online – January 2021

[2] Marek Domański, Tomasz Grajek, Krzysztof Klimaszewski, Maciej Kurc, Olgierd Stankiewicz, Jakub Stankowski, Krzysztof Wegner, “Poznań Multiview Video Test Sequences and Camera Parameters”, ISO/IEC JTC1/SC29/WG11, MPEG/M17050, m17050, Xian, China, 26-30 October 2009.

[3] Dawid Mieloch, Olgierd Stankiewicz, Marek Domański, “Depth Map Estimation for Free-Viewpoint Television and Virtual Navigation,” IEEE Access, vol. 8, pp. 5760-5776, 2020.

[4] ISO/IEC-JCT1/SC29/WG11, Common Test Conditions and Software Reference Configurations, Geneva, Switzerland, 2012.

[5] Patrick Garus, Felix Henry, Gordon Clare, “Feature-driven Decoder Side Depth Estimation”, ISO/IEC JTC1/SC29/WG4, MPEG/M54994, October 2020, Online.