INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG04
MPEG VIDEO CODING

ISO/IEC JTC1/SC29/WG04 MPEG/M56335
April 2021, Online

Source	Poznań University of Technology, Poznań, Poland Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea
Status	Input document
Title	Rectangular blocks in encoder-derived features for decoder-side depth estimation
Author	Błażej Szydełko, Dawid Mieloch, Adrian Dziembowski (PUT), Gwangsoon Lee, Jun Young Jeong (ETRI)

Abstract

This document presents a description of the experiments on enabling the rectangular division of blocks in encoder-derived features used in DSDE. The results show improvement of the proposal over both the G17 anchor and use of square-only division. The recommendation of the document is to establish a new CE on the potential improvements of the geometry absent profile.

1 Proposal

1.1 Deriving of features from depth maps

The first step is to split the frame into main blocks of maxBlockSize x maxBlockSize. During division, Zmin and Zmax are found in each block. Then, for each main block, the condition (1) is checked if the difference of Zmax and Zmin is greater than the set partition threshold (paritionThresh), if the condition is true, the main block can be divided into subblocks.

$$\text{If}(Z_{max} - Z_{min} > \text{partitionThresh})\ doParition \qquad (1)$$

The first method of division is to divide main block into square subblocks only – the size of the subblock is equal to half the length of the currently divided block (Fig. 1).

In the second method of division, in addition to square blocks, it is considered to divide into rectangular sub-blocks in various variants (Fig. 2 and Fig. 3):

A. Vertical symmetrical
B. Horizontal symmetrical
C. Vertical asymmetrical in the ratio 25:75
D. Vertical asymmetrical in the ratio 75:25
E. Horizontal asymmetrical in the ratio 25:75
F. Horizontal asymmetrical in the ratio 75:25

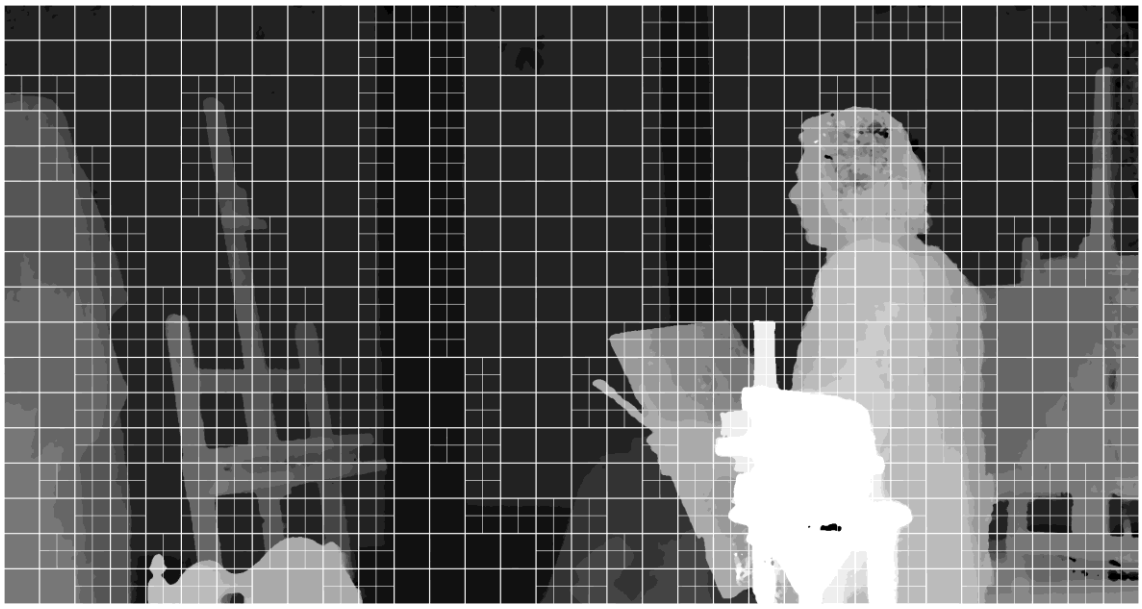

Fig. 1. Blocks of encoder-derived features: square division only.

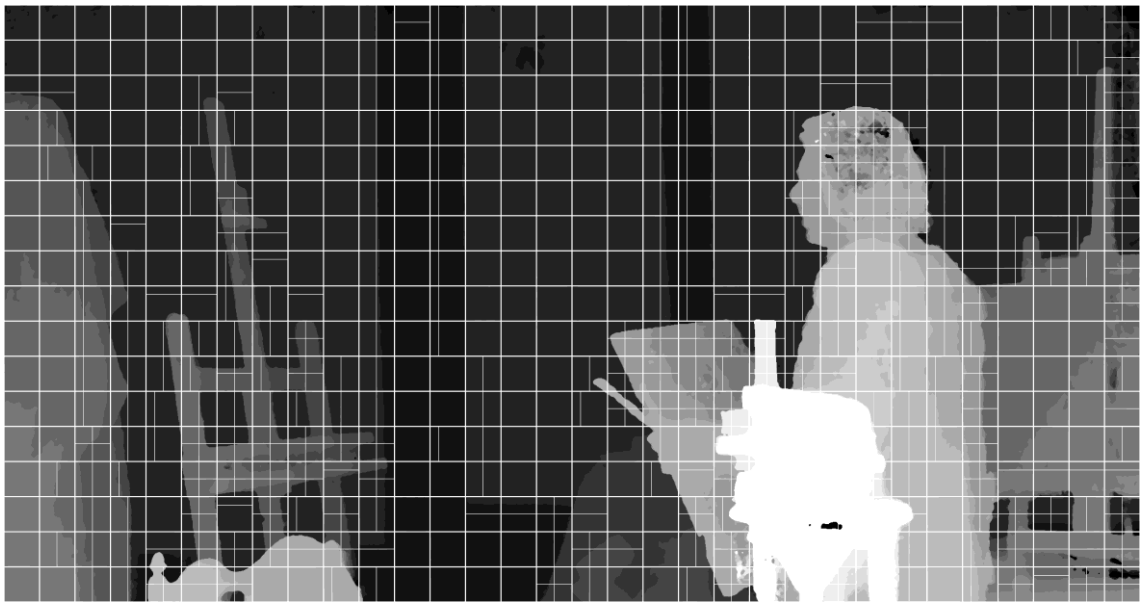

Fig. 2. Blocks of encoder-derived features: rectangular division enabled.

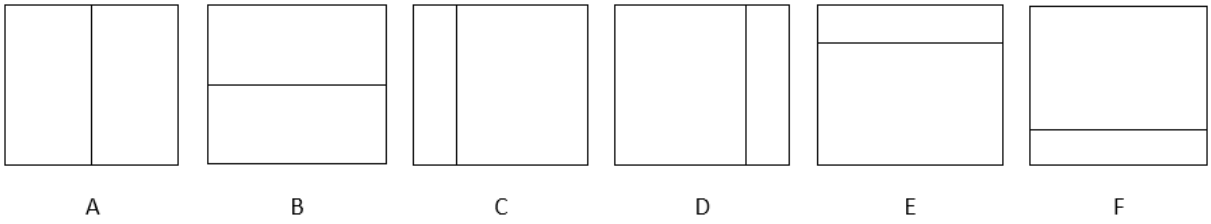

Fig. 3. Possible rectangular divisions of the main block.

After generating the possible divisions, it is checked if the sub-blocks from a given variant meet the condition (1). If subblocks do not meet the condition, the tested variant is rejected from the set and is not taken into account in the following steps. If all variants have been rejected - the block is divided into square subblocks.

Having several split variants to choose, the sum of the costVolume of the subblocks is calculated:

$$costVolume = blkSize.x * blkSize.y *(Zmax - Zmin + 1) \quad (2)$$

The split variant with the lowest costVolume is selected.

In order to reduce the amount of information transmitted for individual blocks in successive depth map frames, the depth information of the block in the previous frame may be used. Having the block from the previous frame and the block from the current frame, the size and position of which (upper left corner of the block) are identical, the check of how the depth range in the currently processed block has changed from the previous one is performed. The basic condition that does not allow the re-use of information from the previous block is the extension of the depth range in the current block:

$$\text{if}(currentBlock.Zmin < previousBlock.Zmin \,||\, currentBlock.Zmax > previousBlock.Zmax)\ \textit{noSkip} \quad (3)$$

Another condition that determines the possibility of skipping is the value of the skipThresh threshold, expressed as a percentage of the possible maximum depth value – MAX_DEPTH_VALUE (for 10 bit 1023). This threshold is used to check how the current block differs from the previous one. For this purpose, the percentage of the difference Zmin and Zmax of the current and previous block with respect to MAX_DEPTH_VALUE and the sum of the absolute difference (SAD) of the depth value between the current and previous block (per pixel), also as a percentage. When all specified values are greater than or equal to skipThresh, the depth information cannot be omitted.

$$\text{if}(ZminCurrPrevPerc >= skipThresh \,||\, ZmaxCurrPrevPerc >= skipThresh \,||\, SADPerc >= skipThresh)\ \textit{noSkip} \quad (4)$$

Otherwise, you can use the depth information from the previous block.

1.2 Compression of features

The quantization of z values that are send as Zmin or Zmax values was set to 32 (number of possible levels divided by 32). The Zmin and Zmax values are sent as the difference from the previously coded block as 16 bit values, 6 flags (skip, division, orientation, symmetry, ratio, square/rectangular) are represented by 1 bit each. The resulting binary file was compressed using the default options of 7Zip file encoder.

1.3 Features example

Fragment of decoded features used by IVDE:

```
f02 y0192 x1664 skip=0 partition=0 zmin=1664 zmax=3456
f02 y0192 x1728 skip=0 partition=1 divideintosquares=1 [ skip=0,zmin=1728,zmax=14528 ] [ skip=0,zmin=1664,zmax=16064 ] [ skip=0,zmin=1984,zmax=22976 ] [ skip=0,zmin=1984,zmax=23616 ]
f02 y0192 x1792 skip=0 partition=1 divideintosquares=1 [ skip=0,zmin=2080,zmax=17792 ] [ skip=0,zmin=15968,zmax=18432 ] [ skip=0,zmin=15840,zmax=24032 ] [ skip=0,zmin=21568,zmax=24512 ]
f02 y0192 x1856 skip=0 partition=1 divideintosquares=0 orientation=0 symmetry=0 ratio=0 [ skip=0,zmin=16800,zmax=18208 ] [ skip=0,zmin=16800,zmax=24576 ]
f02 y0256 x0256 skip=0 partition=1 divideintosquares=0 orientation=1 symmetry=1 [ skip=0,zmin=16800,zmax=18208 ] [ skip=0,zmin=16800,zmax=24576 ]
f02 y0256 x0000 skip=1
```

2 Experimental results

Table 1. G17 vs. G17 with encoder-derived features (rectangular blocks)

Mandatory content - Proposal vs. Low/High-bitrate Anchors

Sequence	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	Max delta Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR
ClassroomVideo	---	-76.4%	5.71	---	-71.8%	-42.3%	-34.4%
Museum	---	---	13.36	-68.9%	-44.9%	-37.7%	-24.8%
Fan	20.6%	22.8%	10.92	49.8%	42.7%	15.4%	21.9%
Kitchen	-14.8%	-3.9%	12.66	-7.3%	1.7%	-5.8%	2.7%
Painter	-21.9%	32.8%	6.58	29.9%	79.1%	-20.7%	29.7%
Frog	28.5%	57.4%	7.01	32.4%	62.1%	38.8%	66.3%
Carpark	-29.3%	-4.5%	10.05	-29.3%	-4.6%	-33.1%	-8.7%
Chess	---	---	24.93	143.4%	79.9%	---	---
Group	---	---	13.29	-72.3%	-48.2%	---	---
MIV	**---**	**---**	**11.61**	**---**	**10.7%**	**---**	**---**

Optional content - Proposal vs. Low/High-bitrate Anchors

Fencing	-70.4%	-20.1%	13.03	-37.6%	-12.6%	-6.3%	8.2%
Hall	---	---	17.18	---	-53.6%	---	---
Street	34.8%	37.8%	8.25	8.7%	26.1%	8.2%	22.4%
Hijack	---	---	20.51	---	-70.9%	---	---
Mirror	---	293.4%	14.98	---	249.0%	---	439.5%
MIV	**---**	**---**	**14.79**	**---**	**27.6%**	**---**	**---**

Comments:

- The proposed feature extractor together with the proposed compression of metadata provides an increase of objective quality for most CTC sequences, especially for high-bitrate cases.
- Mixed results for SD and SE are the result of the high bitrate of metadata (see Table 3). Nevertheless, for high bitrates, the increase of the quality exceeds the bitrate increase due to metadata. The size of metadata can be decreased by stronger quantization of parameters.
- Loss of the quality in SI is due to the presence of mirrors in the scene – the encoder-derived features are based on the ground-truth depth maps, in which the depth of mirror shows the distance to the mirror, not to the reflected objects.

Table 3. The bitrate of features for the square blocks and rectangular blocks.

Sequence	Bitrate of features	
	Square	Rectangular
A	498.34	422.99
B	681.94	567.88
C	809.24	708.56
D	8236.32	6424.76
E	6155.42	4706.34
I	1624.55	1279.43
J	657.22	636.42
L	2574.69	1884.20
N	341.14	278.03
O	1599.32	1441.62
P	3101.71	2203.26
R	1179.67	908.96
T	3059.96	2259.22
U	2640.62	1902.00
Avg.	**2347.65**	**1824.75**

Table 2. G17 with encoder-derived features: square vs. rectangular blocks

Mandatory content - Proposal vs. Low/High-bitrate Anchors

Sequence	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	Max delta Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR
ClassroomVideo	-3.4%	0.3%	5.71	5.8%	2.0%	-0.6%	-1.2%
Museum	11.4%	5.2%	13.36	2.4%	0.4%	5.6%	1.8%
Fan	0.4%	-1.0%	10.92	-0.9%	-1.7%	0.1%	-1.2%
Kitchen	---	---	12.84	-26.0%	-14.6%	---	-75.0%
Painter	-33.4%	-23.5%	6.58	-35.6%	-23.9%	-22.2%	-17.8%
Frog	-5.9%	-9.6%	7.01	-7.0%	-10.3%	-6.0%	-10.0%
Carpark	38.4%	12.7%	10.05	9.0%	0.1%	30.3%	11.8%
Chess	172.8%	10.6%	24.93	-9.9%	-5.0%	1.2%	9.4%
Group	0.6%	-1.8%	######	-0.5%	-3.1%	-0.7%	-2.7%
MIV	**---**	**---**	**11.63**	**-7.0%**	**-6.2%**	**---**	**-9.4%**

Optional content - Proposal vs. Low/High-bitrate Anchors

Fencing	20.3%	0.2%	13.03	3.9%	-5.4%	0.5%	-4.1%
Hall	---	-57.2%	17.18	32.8%	1.2%	-54.1%	-44.1%
Street	0.4%	-5.8%	8.25	0.8%	-5.8%	1.2%	-5.5%
Hijack	---	-55.7%	######	6.6%	1.4%	---	-55.3%
Mirror	-2.3%	-3.6%	14.98	-3.6%	-4.4%	-0.2%	-2.7%
MIV	**---**	**-24.4%**	**14.79**	**8.1%**	**-2.6%**	**---**	**-22.3%**

Comments:

- Enabling rectangular blocks, on average, increases the final quality. The increase is mostly due to the smaller bitrate of encoder-derived features (Table 3), but for some sequences (SD, SJ, ST) the objective quality was also increased.

3 Recommendations

We recommend establishing a new CE on the potential improvements of the geometry absent profile.

4 Acknowledgement

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00207, Immersive Media Research Laboratory).