

**INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 4
MPEG VIDEO CODING**

ISO/IEC JTC 1/SC 29/WG 4 m 55751
October 2020, Online

Title: Proposal of IVDE 3.0
Source: Dawid Mieloch, Adrian Dziembowski
(Poznań University of Technology)

Abstract

The document describes a proposal of new version of IVDE. The proposal includes new features added to IVDE required to perform exploration experiments (EE2 and EE5), the automatic calculation of depth range required by the group in decoder-side depth estimation, and other smaller fixes of recently discovered issues.

1 Introduction and description of the proposal

In this document, new functionalities proposed for the IVDE 2.0 are described. Moreover, fixes of errors discovered during the works are also listed and explained.

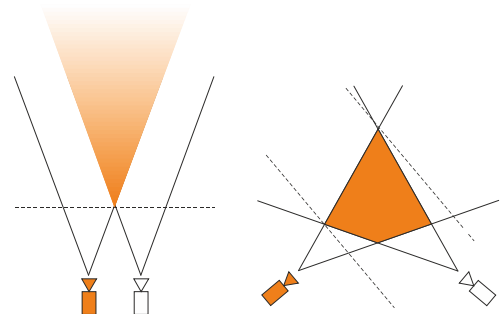
1.1 New functionalities

- *Automatic depth range calculation*

In ERP content, z_{near} is equal to the largest distance between views used in inter-view matching. z_{far} was calculated to give minimum of 1 degree shift per pixel when the most distant views are matched.

In perspective content, the calculation of z_{far} and z_{near} are different for linear camera arrangement and for other arrangements.

For the orange camera, in the linear arrangement (left), the z_{near} should be no closer than the z for which a point on the right side of the image becomes visible on the left side in the white camera. z_{far} on the other hand, can be estimated as z for which the disparity for the point in the middle of the image is no smaller than 1 pixel.



For other arrangements, the z_{near} should be no closer than the z for which a point on the left side of the image becomes visible on the left side in the white camera, while z_{far} should be no further than the z for which a point on the left side of the image becomes visible on the left side in the white camera.

To calculate the range for all possible arrangement, all abovementioned conditions are checked and, in the end, the closest calculated z becomes z_{near} , while the most distant z becomes z_{far} .

To reduce the calculated range to typical real-world cases, the conditions are checked with set threshold (minimum 10 pixel disparity, width/10 for matching of pixels on the edges of images). In the end, calculated z_{near} and z_{far} values are cropped to the range [0.3, 1000] (30 cm to 1 km).

- *Handling of encoder-derived features*

In order to perform the exploration experiment on the use of the encoder-derived features in the depth estimation, the required functionalities were added to the IVDE 3.0. The new version handles z_{near} and z_{far} values and skip flag send from the encoder in order to increase the quality of decoder-side estimated depth maps and decrease the time of the estimation.

1.2 Bugfixes

- *Moved calculation of projection matrices for ERP content*

In IVDE 2.0, projection matrices for ERP content were calculated for each segment in each view and each depth label (e.g., for SB it resulted in $100000 * 24 * 256$ computations of these matrices in the first frame). In IVDE 3.0 it is done once per frame.

- *Fixed choosing of the central camera*

During the optimization, the depth is calculated as a distance from the central camera of the system. In IVDE 1.0, the central camera was chosen as camera $\text{NumOfCameras}/2$, while in IVDE 2.0, was chosen as the camera with the smallest summarized distance to all other cameras. Unfortunately, the change was not made in all functions, causing the problems discovered in the SR sequence.

- *Removed redundant calculation of overall energy in GraphCut estimation*

The estimation of depth in IVDE is based on a cost function minimization. The process is performed by *maxflow()* function, which after the optimization returns a new value of the optimized function (which is also known as energy value).

For testing purposes, the energy value was checked if it is lower than in the previous optimization (for the previous depth label). If it was lower, then the new depth level was added to the depth map. For estimation of the first frame, this condition was always true. Unfortunately, when temporal enhancement is used, therefore, not all segments are used

in estimation, the new energy was not always lower, because the number of active segments was not the same.

The error was discovered during the development of EE2 because the number of active segments varies significantly for different labels.

- Parameters of "Center" and "Posetrace" cameras are skipped

This change was made to have the possibility to use CTC jsons with camera parameters to estimate depth maps for input views more easily (encoder-side depth estimation).

- Increased precision of π

Increased precision of estimation for ERP content.

2 Experimental results

2.1 IVDE 2.0 vs. IVDE 3.0 (without automatic depth range)

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

Sequence		High-BR BD rate	Low-BR BD rate	Max delta	High-BR BD rate	Low-BR BD rate	High-BR BD rate	Low-BR BD rate	TMIV encoding	HM encoding	TMIV decoding
		Y-PSNR	Y-PSNR	Y-PSNR	VMAF	VMAF	IV-PSNR	IV-PSNR			
ClassroomVideo	SA	-4.8%	-5.0%	5.52	-1.0%	-0.4%	2.3%	1.7%	100.0%	100.0%	52.4%
Museum	SB	78.8%	-2.1%	15.50	-0.7%	-0.7%	-1.0%	-0.8%	100.0%	100.0%	28.4%
Fan	SO	-0.1%	-0.1%	10.22	-0.0%	0.0%	0.1%	0.1%	100.0%	100.0%	94.4%
Kitchen	SJ	1.0%	1.1%	13.50	-0.1%	-0.0%	0.8%	0.9%	100.0%	100.0%	95.7%
Painter	SD	2.3%	1.4%	9.61	2.4%	0.6%	4.5%	2.3%	100.0%	100.0%	98.7%
Frog	SE	0.1%	0.0%	6.82	0.0%	0.0%	0.1%	0.1%	100.0%	100.0%	106.3%
Carpark	SP	-3.3%	-2.5%	9.48	-1.1%	-0.8%	-1.6%	-1.9%	100.0%	100.0%	92.3%
Chess	SN	-10.9%	-26.8%	24.61	-5.0%	-5.5%	-68.0%	678.1%	100.0%	100.0%	22.5%
Group	SR	---	---	13.20	-81.4%	-64.6%	---	---	100.0%	100.0%	95.4%
MIV		---	---	12.05	-9.6%	-7.9%	---	---	100.0%	100.0%	76.3%

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

Fencing	SL	-4.3%	3.9%	12.75	-1.0%	0.9%	0.8%	1.4%	100.0%	100.0%	96.0%
Hall	SU	1.2%	0.7%	7.94	0.3%	0.1%	0.7%	0.5%	100.0%	100.0%	94.1%
Street	ST	36.4%	-98.2%	17.93	32.8%	12.9%	18.7%	#####	100.0%	100.0%	92.3%
ChessPieces	SQ	-97.2%	-46.9%	28.93	-7.5%	-4.2%	5.7%	-1.6%	100.0%	100.0%	23.5%
Hijack	SC	115.4%	41.1%	22.75	10.0%	12.3%	70.6%	25.9%	100.0%	100.0%	35.5%
MIV		10.3%	-19.9%	18.06	6.9%	4.4%	19.3%	#####	100.0%	100.0%	68.3%

The table above shows the comparison of the G17 anchor with the proposal in the same configuration. As it can be seen, the time of depth estimation (included in the TMIV decoding time) was significantly decreased for ERP content.

Below, the tables with the comparison of WS-PSNR can be find. These tables show just the quality of the synthesized views, not a BD-rate (as the bitrate for the anchor and the proposal is of course the same). The quality was mainly improved for ERP content, but the largest improvement can be seen in SR, for which the quality is now in the same range as for other sequences.

Mandatory

SA		
WS-PSNR		
Anchor SA	Proposal SA	Delta
31.57	31.61	0.04
31.50	31.51	0.01
31.33	31.34	0.01
31.21	31.24	0.02
30.93	30.94	0.01

SB		
WS-PSNR		
Anchor SB	Proposal SB	Delta
26.91	26.90	-0.01
26.57	26.62	0.05
26.82	26.85	0.03
26.59	26.60	0.02
26.06	26.06	0.00

SO		
WS-PSNR		
Anchor SO	Proposal SO	Delta
32.17	32.17	0.00
31.39	31.40	0.01
30.47	30.47	0.00
29.34	29.34	0.00
27.76	27.76	0.00

SJ		
WS-PSNR		
Anchor SJ	Proposal SJ	Delta
35.70	35.70	-0.01
34.64	34.62	-0.02
33.79	33.77	-0.02
32.62	32.61	-0.01
31.57	31.53	-0.04

SD		
WS-PSNR		
Anchor SD	Proposal SD	Delta
37.68	37.68	0.00
37.19	37.10	-0.10
36.43	36.43	0.00
35.20	35.16	-0.04
33.55	33.53	-0.02

SE		
WS-PSNR		
Anchor SE	Proposal SE	Delta
31.11	31.11	0.00
30.16	30.15	0.00
28.74	28.73	0.00
27.26	27.26	0.00
25.54	25.54	0.00

SP		
WS-PSNR		
Anchor SP	Proposal SP	Delta
35.59	35.59	0.00
35.38	35.39	0.01
34.95	34.99	0.04
34.43	34.45	0.02
33.53	33.56	0.03

SN		
WS-PSNR		
Anchor SN	Proposal SN	Delta
27.83	27.98	0.15
27.97	28.10	0.13
27.91	27.96	0.05
27.61	27.67	0.06
27.84	27.78	-0.06

SR		
WS-PSNR		
Anchor SR	Proposal SR	Delta
20.29	30.62	10.33
20.41	29.61	9.21
20.12	28.68	8.57
20.80	27.73	6.93
20.10	26.61	6.51

Optional

SL		
WS-PSNR		
Anchor SL	Proposal SL	Delta
34.15	34.14	0.00
33.98	33.99	0.01
33.90	33.92	0.02
33.74	33.70	-0.04
33.24	33.25	0.01

SU		
WS-PSNR		
Anchor SU	Proposal SU	Delta
36.65	36.65	0.00
36.36	36.35	-0.01
36.10	36.09	-0.01
35.50	35.49	-0.01
34.42	34.42	0.00

ST		
WS-PSNR		
Anchor ST	Proposal ST	Delta
35.44	35.41	-0.03
35.35	35.30	-0.05
35.27	35.22	-0.05
35.31	35.28	-0.03
35.27	35.22	-0.04

SC		
WS-PSNR		
Anchor SC	Proposal SC	Delta
29.65	29.57	-0.08
29.54	29.49	-0.05
29.53	29.41	-0.12
29.41	29.37	-0.04
29.20	29.03	-0.17

SQ		
WS-PSNR		
Anchor SQ	Proposal SQ	Delta
27.56	27.61	0.05
27.71	27.77	0.06
27.81	27.78	-0.03
27.56	27.67	0.10
27.50	27.59	0.09

2.2 Automatic depth range

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

Sequence		High-BR BD rate	Low-BR BD rate	Max delta	High-BR BD rate	Low-BR BD rate	High-BR BD rate	Low-BR BD rate	TMIV encoding	HM encoding	TMIV decoding
		Y-PSNR	Y-PSNR	Y-PSNR	VMAF	VMAF	IV-PSNR	IV-PSNR			
ClassroomVideo	SA	-28.8%	-17.2%	5.77	-48.4%	-24.3%	28.1%	24.4%	100.0%	100.0%	70.8%
Museum	SB	#####	63.7%	15.31	62.7%	29.2%	-26.7%	-12.8%	100.0%	100.0%	71.7%
Fan	SO	-5.9%	-0.6%	10.29	-15.7%	-8.9%	-7.8%	-4.0%	100.0%	100.0%	94.3%
Kitchen	SJ	-45.4%	-37.5%	12.68	-36.0%	-24.1%	-37.1%	-33.1%	100.0%	100.0%	67.3%
Painter	SD	-24.2%	-12.7%	8.46	-14.2%	-7.0%	-22.0%	-11.1%	100.0%	100.0%	69.6%
Frog	SE	-0.6%	-0.2%	6.86	1.5%	1.3%	-14.3%	-8.8%	100.0%	100.0%	89.6%
Carpark	SP	65.4%	40.3%	10.24	35.1%	23.2%	120.3%	72.5%	100.0%	100.0%	67.2%
Chess	SN	---	---	25.77	---	-81.0%	---	-79.5%	100.0%	100.0%	89.4%
Group	SR	---	---	21.20	184.0%	87.6%	---	---	100.0%	100.0%	91.3%
MIV		---	---	12.95	---	-0.4%	---	---	100.0%	100.0%	79.0%

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

Fencing	SL	256.1%	71.5%	13.39	45.4%	14.4%	13.3%	5.9%	100.0%	100.0%	70.8%
Hall	SU	9.3%	8.6%	8.12	6.4%	5.0%	21.5%	13.1%	100.0%	100.0%	64.5%
Street	ST	---	---	18.67	---	284.2%	---	---	100.0%	100.0%	78.4%
ChessPieces	SQ	---	---	30.39	16.8%	-5.2%	---	---	100.0%	100.0%	86.9%
Hijack	SC	---	---	21.60	-24.0%	-19.4%	---	---	100.0%	100.0%	74.1%
MIV		---	---	18.43	---	55.8%	---	---	100.0%	100.0%	75.0%

The table above shows the comparison of the the IVDE 3.0 with CTC depth ranges and IVDE 3.0 with automatic depth ranges. Here, shorter time of depth estimation can be seen in almost all sequences.

Sequence	CTC depth range	Estimated depth range
SA	[0.8, 1000]	[0.3, 135.495]
SB	[0.5, 25]	[0.611215, 398.45]
SO	[0.35, 12.5]	[1, 18.8]
SJ	[2.24, 7.17]	[2, 69.9]
SD	[1.80881, 5.40248]	[0.8, 13.7]
SE	[0.3, 1.62]	[0.5, 16.4]
SP	[3.45064, 276.051]	[1.4, 75.3]
SN	[0.1, 500]	[0.6, 391.139]
SR	[1.5, 25]	[1.9, 4.8]
SL	[3, 7]	[1.5, 12.3]
SU	[3.45064, 276.051]	[1.4, 75.3]
ST	[1.85064, 276.051]	[1.4, 75.3]
SQ	[0.747777, 195.163]	[0.6, 391.139]
SC	[0.5, 25]	[0.5706, 743.947]

Below, the tables with the comparison of WS-PSNR can be find. For most of the mandatory sequences a gain in the quality of synthesized views. Value of z_{far} larger than in CTC seems to not cause any degradation to the quality of estimated depth. However, in the case of z_{far} closer than in CTC the quality is decreased (SP, SR, SU and ST).

Mandatory

SA		
WS-PSNR		
Anchor SA	Proposal SA	Delta
31.61	31.75	0.14
31.51	31.54	0.03
31.34	31.47	0.13
31.24	31.27	0.04
30.94	30.98	0.05

SB		
WS-PSNR		
Anchor SB	Proposal SB	Delta
26.90	26.71	-0.20
26.62	26.63	0.01
26.85	26.48	-0.37
26.60	26.36	-0.24
26.06	26.18	0.12

SO		
WS-PSNR		
Anchor SO	Proposal SO	Delta
32.17	32.41	0.24
31.40	31.56	0.16
30.47	30.54	0.07
29.34	29.30	-0.05
27.76	27.71	-0.05

SJ		
WS-PSNR		
Anchor SJ	Proposal SJ	Delta
35.70	36.29	0.59
34.62	35.58	0.96
33.77	34.69	0.92
32.61	33.61	1.00
31.53	32.12	0.59

SD		
WS-PSNR		
Anchor SD	Proposal SD	Delta
37.68	38.37	0.70
37.10	37.76	0.67
36.43	36.84	0.41
35.16	35.43	0.27
33.53	33.70	0.17

SE		
WS-PSNR		
Anchor SE	Proposal SE	Delta
31.11	31.11	0.01
30.15	30.19	0.03
28.73	28.73	0.00
27.26	27.26	0.00
25.54	25.54	0.01

SP		
WS-PSNR		
Anchor SP	Proposal SP	Delta
35.59	35.23	-0.36
35.39	35.07	-0.31
34.99	34.62	-0.37
34.45	34.06	-0.39
33.56	33.28	-0.28

SN		
WS-PSNR		
Anchor SN	Proposal SN	Delta
27.98	29.35	1.37
28.10	29.15	1.05
27.96	29.07	1.11
27.67	29.01	1.35
27.78	28.47	0.70

SR		
WS-PSNR		
Anchor SR	Proposal SR	Delta
30.62	22.91	-7.70
29.61	22.77	-6.84
28.68	22.63	-6.05
27.73	22.29	-5.44
26.61	21.96	-4.65

Optional

SL		
WS-PSNR		
Anchor SL	Proposal SL	Delta
34.14	33.75	-0.39
33.99	33.69	-0.30
33.92	33.61	-0.31
33.70	33.46	-0.24
33.25	33.04	-0.21

SU		
WS-PSNR		
Anchor SU	Proposal SU	Delta
36.65	36.56	-0.09
36.35	36.31	-0.04
36.09	36.03	-0.06
35.49	35.37	-0.12
34.42	34.34	-0.09

ST		
WS-PSNR		
Anchor ST	Proposal ST	Delta
35.41	34.78	-0.63
35.30	34.48	-0.82
35.22	34.48	-0.74
35.28	34.41	-0.87
35.22	34.41	-0.82

SQ		
WS-PSNR		
Anchor SQ	Proposal SQ	Delta
27.61	27.21	-0.40
27.77	27.13	-0.64
27.78	27.26	-0.52
27.67	27.05	-0.62
27.59	27.09	-0.50

SC		
WS-PSNR		
Anchor SC	Proposal SC	Delta
29.57	29.80	0.24
29.49	29.83	0.34
29.41	29.82	0.41
29.37	29.86	0.49
29.03	29.74	0.71

3 Recommendations

We recommend to:

- accept the proposal as a new version of IVDE and issue new manual for the software as an output document,
- discuss the use of the proposed automatic depth range in decoder-side depth estimation MIV anchor.

Acknowledgement

The research was supported by the Ministry of Science and Higher Education of Republic of Poland.