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/M54151
June 2020, Online

Source **Intel, PUT**
Status **Input**
Title **[MIV] Grouping and anchor study on MIV content**
Author Basel Salahieh, Adrian Dziembowski, Jill Boyce

1 Introduction

This document is an informative contribution presenting:

- comparison of TMIV performance for different number of groups at similar pixel rate,
- comparison of MIV vs. MIV view anchor results for different number of frames.

2 Overview of the experiment

The experiment was performed using TMIV5.0 [N19213]. Results presented in the following sections were obtained for 3 anchors and two additional configurations:

- A97 (MIV anchor, ‘ff’ configuration),
- A17 (MIV anchor, ‘rf’ configuration),
- V17 (MIV view anchor, ‘rf’ configuration),
- A97_3G (MIV, ‘ff’ configuration, 3 groups),
- V97 (MIV view, ‘ff’ configuration).

All the experiments followed Common Test Conditions [N19214].

3 Experimental results

3.1 3 groups vs. 2 groups on the perspective content

In order to generate atlases for 3 groups while meeting all CTC constraints (other than #decoder instantiation), 3 parameters in .json configuration file for TMIV encoder were changed (for all perspective sequences).

Table 1. Modified parameters in TMIV encoder .json file.

Parameter	Configuration	
	A97	A97_3G
numGroups	2	3
maxAtlases	2	3
maxLumaPictureSize	8912896	6300000

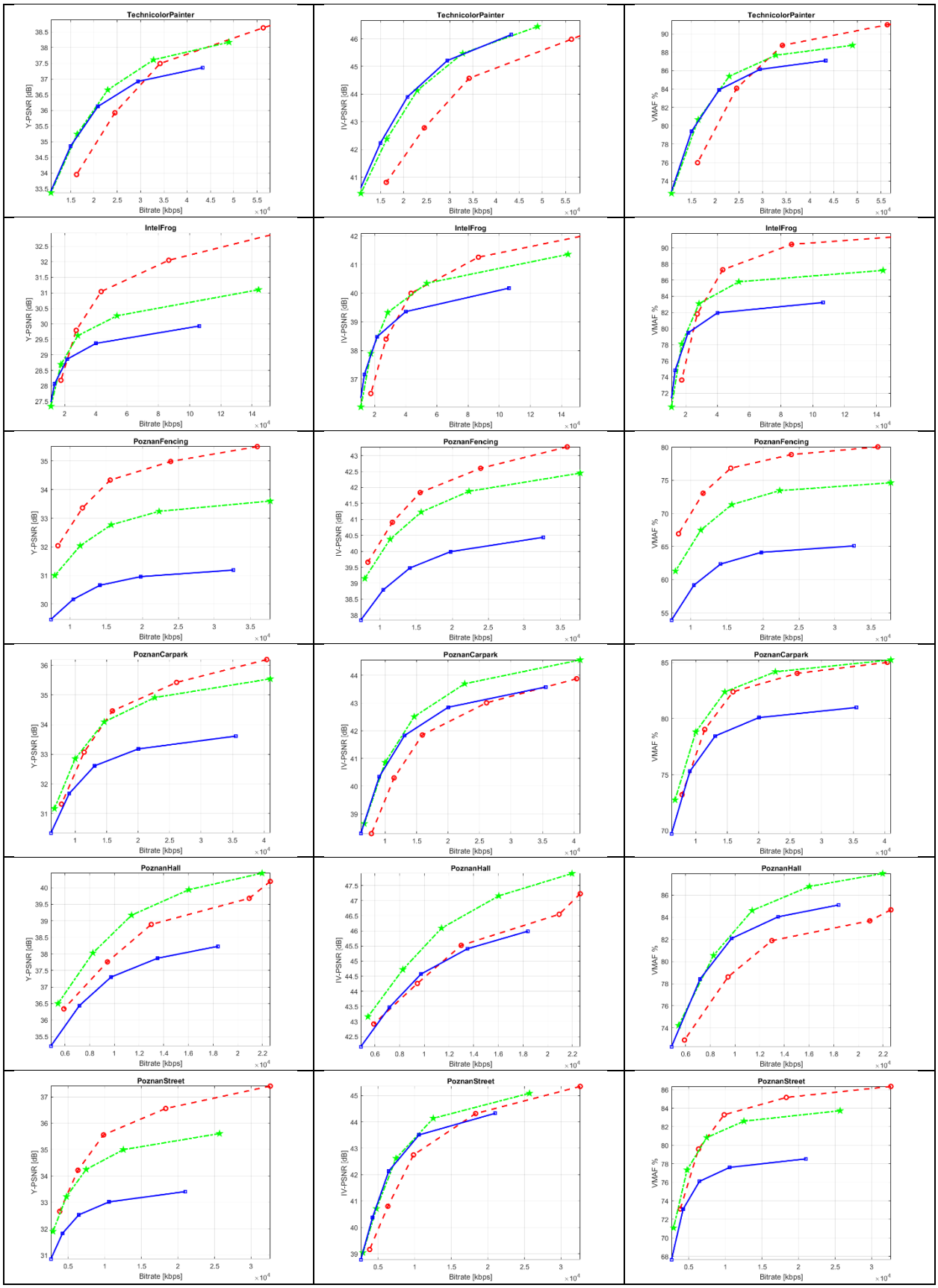

Fig. 1. RD-curves for V97 (red), A97 (blue) and A97_3G (green); natural content.

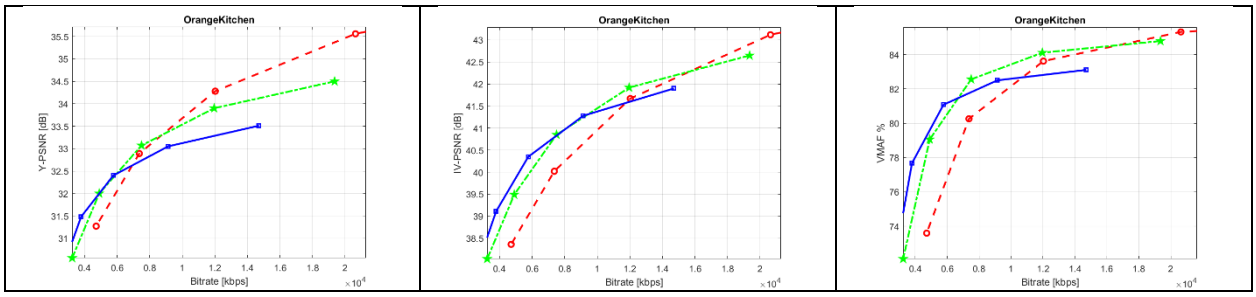

Fig. 2. RD-curves for V97 (red), A97 (blue) and A97_3G (green); CG content.

Also with regards to pixel rate, we present them in Table 2 for all three cases.

Table 2. Comparision of pixel rate in GP/S between V97, A97, A97_3G

	Orange Kitchen SJ	Technicolor Painter SD	Intel Frog SE	Poznan Fencing SL	Poznan Carpark SP	Poznan Hall ST	Poznan Street SU
A97	0.67	0.67	0.67	0.56	0.56	0.56	0.56
A97_3G	0.71	0.71	0.71	0.59	0.59	0.59	0.59
V97	1.12	1.07	0.87	0.52	0.52	0.52	0.52

Numerically, comparing A97 (left columns) vs A97_3G (right columns) reveals the following summary results shared in table 3.

Table 3. A97_3G (right) objective results with respect to A97 (left) for perspective content

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	Pixel rate ratio
OrangeKitchen	SJ	-12.9%	3.7%	14.85	-4.0%	12.8%	2.8%	11.9%	0.66
TechnicolorPainter	SD	-9.5%	-2.0%	8.00	-7.1%	-0.2%	3.7%	5.3%	0.66
IntelFrog	SE	-46.8%	-10.7%	10.82	-35.7%	-5.0%	-26.1%	-1.3%	0.66
PoznanFencing	SL	0.0%	0.0%	13.91	0.0%	-44.4%	-63.4%	-41.4%	0.55
PoznanCarpark	SP	-49.8%	-24.9%	11.20	-46.0%	-20.8%	-14.3%	-3.5%	0.55
PoznanHall	ST	-49.6%	-31.0%	12.26	-14.1%	-3.4%	-28.8%	-19.3%	0.55
PoznanStreet	SU	-71.7%	-44.0%	10.21	-63.0%	-29.1%	-4.4%	1.4%	0.55

It maybe thought that the results presented in Table3 are biased since 3 groups have an extra basic view being coded hence artifically boosting the synthesis results. Thus, we share as well in table4 the results by averaging only the non-coded views (i.e. excluding synthesis results of basic views). This still show improvments for A97_3G and those improvments are actually bigger for contents with more challenging depth maps.

Table 4. A97_3G objective results with respect to A97 for perspective conten (non-coded only)

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	Pixel rate ratio
OrangeKitchen	SJ	-10.1%	6.9%	5.92	3.1%	16.5%	8.9%	15.7%	0.66
TechnicolorPainter	SD	-7.7%	-1.0%	5.20	-4.0%	1.4%	7.3%	7.3%	0.66
IntelFrog	SE	-43.4%	-13.5%	4.75	-37.1%	-5.1%	-28.2%	-2.8%	0.66
PoznanFencing	SL	0.0%	0.0%	5.18	0.0%	-48.0%	0.0%	-52.0%	0.55
PoznanCarpark	SP	0.0%	-32.8%	4.98	-41.6%	-18.4%	-12.7%	-2.8%	0.55
PoznanHall	ST	0.0%	-42.2%	5.07	-4.6%	3.2%	-37.3%	-23.1%	0.55
PoznanStreet	SU	0.0%	-68.2%	3.36	-65.1%	-28.7%	1.5%	4.3%	0.55

The results reported here prove, that TMIV in 3-group configuration provides much better results than for 2 groups and stands better against MIV View. Also, we note that the optional poznan fencing content have been actually quite improved and in fact their results are even better than some of the mandatory contents (i.e. PoznanFencing), hence we recommend considering them as mandatory content going forward.

Subjective results are available on the MPEG content server at /MPEG-I/Intel/m54151

3.2 MIV vs. MIV view for 'ff' and 'rf' configurations

In the second experiment, results of MIV and MIV view anchor were compared in two configurations: 'ff' and 'rf'.

Results presented in Figs. 3-4 and Tables 5-6 show that despite the shape of the curves for 'rf' vs 'ff' are similar for both anchors, the gab between them is large and the bitrate where the intersection of both anchor types are different in the two configurations. We also note that the 'ff' results are always better than 'rf' ones which means that the 17 frame selected for the 'rf' study are more challenging than the other frames in 'ff' on average.

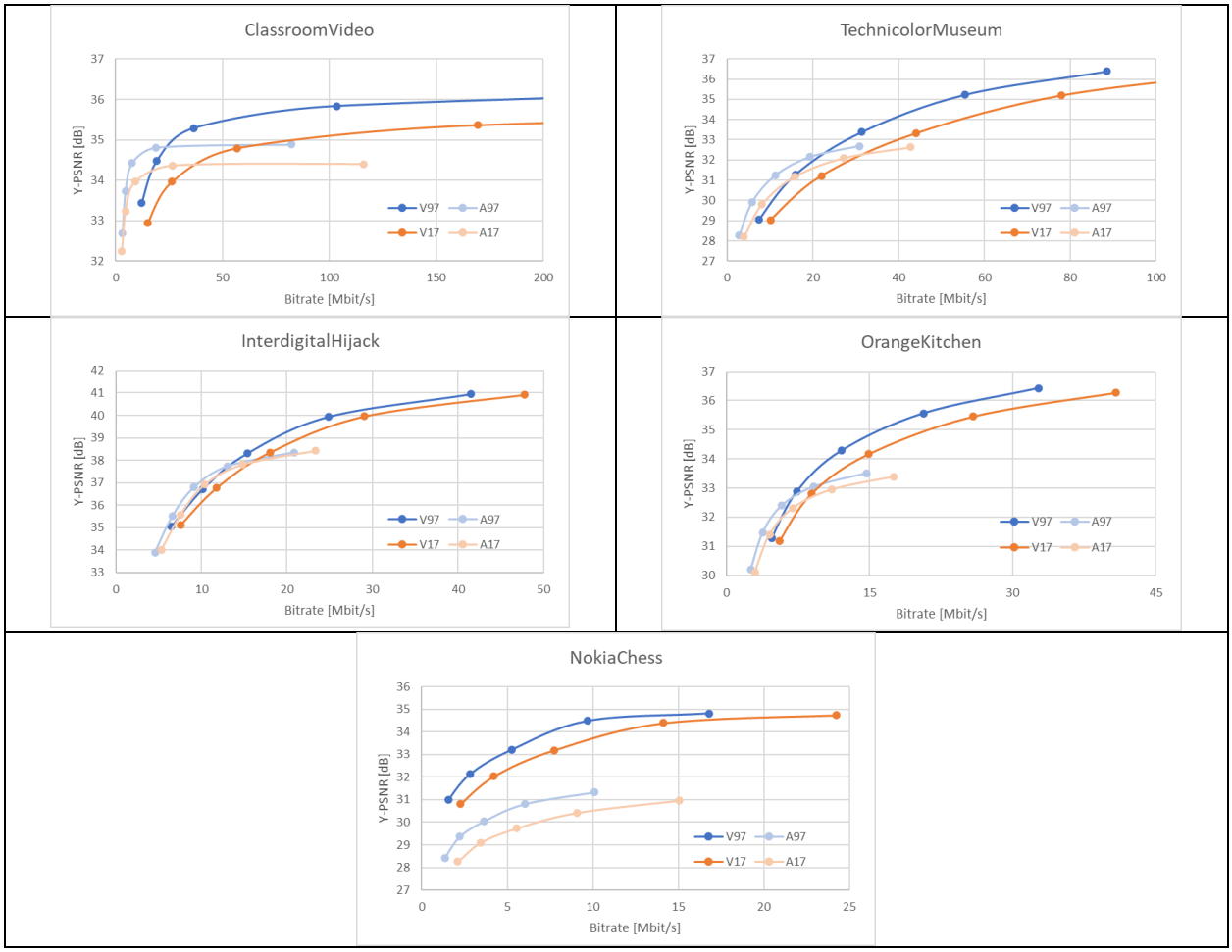

Fig. 3. RD-curves for 'ff' (blue) and 'rf' (orange) configuration,
for MIV anchor (light) and MIV view (dark) anchor; CG content.

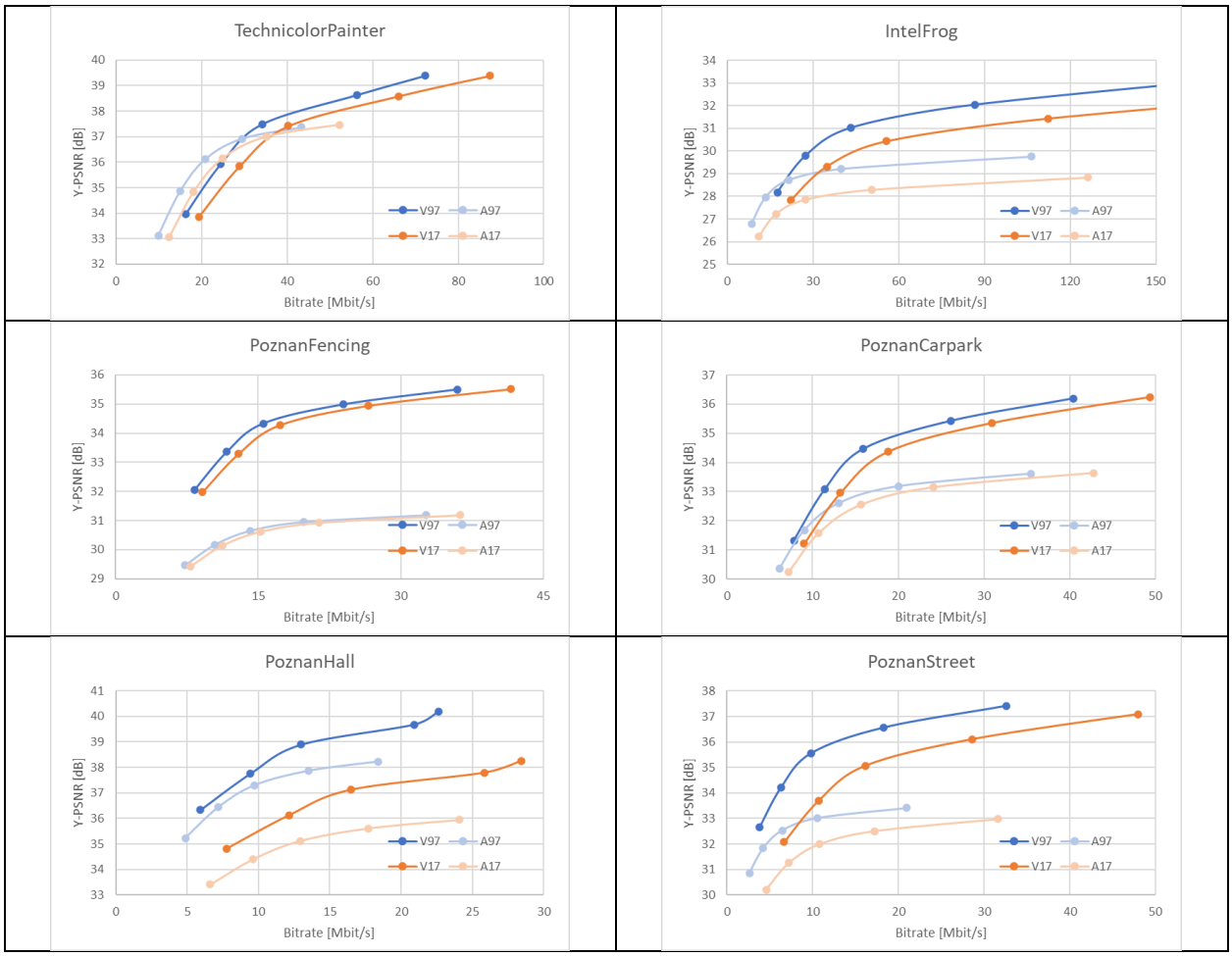

Fig. 2. RD-curves for ‘ff’ (blue) and ‘rf’ (orange) configuration, for MIV anchor (light) and MIV view (dark) anchor; natural content.

Table 5. BD-rates for ‘ff’ configuration

Mandatory content - V97 (left) vs. A97 (right)

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	Pixel rate ratio
ClassroomVideo	SA	-33.3%	-59.6%	3.38	-40.5%	-66.3%	-77.0%	-74.8%	0.63
TechnicolorMuseum	SB	-11.0%	-34.2%	14.24	-12.0%	-39.5%	-56.0%	-58.8%	0.63
InterdigitalHijack	SC	-0.2%	-9.6%	12.41	-15.3%	-18.2%	-10.2%	-16.1%	0.63
OrangeKitchen	SJ	30.2%	-11.9%	15.43	-16.1%	-33.0%	-17.6%	-29.8%	0.62
TechnicolorPainter	SD	-4.4%	-20.0%	7.01	5.6%	-17.3%	-29.8%	-30.7%	0.63
IntelFrog	SE	0.0%	11.1%	12.36	156.1%	-10.9%	22.1%	-24.2%	0.62
PoznanFencing	SL	0.0%	0.0%	15.15	0.0%	0.0%	0.0%	102.4%	0.52
MIV		**-2.7%**	**-17.8%**	**11.43**	**11.1%**	**-26.5%**	**-24.1%**	**-18.8%**	

Optional content - V97 (left) vs. A97 (right)

NokiaChess	SN	0.0%	0.0%	16.87	127.4%	74.7%	251.0%	199.1%	0.63
PoznanCarpark	SP	106.8%	21.8%	12.87	71.6%	12.6%	-16.4%	-19.9%	0.52
PoznanHall	ST	53.4%	20.5%	14.82	-29.3%	-23.2%	4.0%	-1.6%	0.52
PoznanStreet	SU	0.0%	112.9%	13.01	0.0%	34.1%	-19.0%	-23.3%	0.52
MIV		**40.0%**	**38.8%**	**14.39**	**42.4%**	**24.6%**	**54.9%**	**38.6%**	

Table 6. BD-rates for ‘rf’ configuration

Mandatory content - V17 (left) vs. A17 (right)

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	Pixel rate ratio
ClassroomVideo	SA	-51.4%	-67.7%	3.28	18.2%	-70.6%	-83.0%	-81.3%	0.63
TechnicolorMuseum	SB	-11.0%	-32.9%	13.86	-15.5%	-39.4%	-54.6%	-57.8%	0.63
InterdigitalHijack	SC	-6.1%	-14.3%	12.43	-20.3%	-22.2%	-14.4%	-19.7%	0.63
OrangeKitchen	SJ	28.8%	-12.9%	15.26	-19.5%	-35.0%	-17.4%	-30.2%	0.62
TechnicolorPainter	SD	-7.5%	-20.8%	6.89	-0.3%	-18.1%	-30.9%	-30.8%	0.63
IntelFrog	SE	0.0%	39.7%	13.10	0.0%	2.8%	73.2%	-7.4%	0.62
PoznanFencing	SL	0.0%	0.0%	15.15	0.0%	0.0%	0.0%	98.2%	0.52
MIV		**-6.8%**	**-15.5%**	**11.42**	**-5.3%**	**-26.1%**	**-18.2%**	**-18.4%**	

Optional content - V17 (left) vs. A17 (right)

NokiaChess	SN	0.0%	0.0%	15.28	145.9%	81.1%	0.0%	225.2%	0.63
PoznanCarpark	SP	104.2%	24.1%	12.86	69.7%	13.3%	-15.9%	-18.7%	0.52
PoznanHall	ST	0.0%	53.0%	16.04	-17.7%	-14.7%	54.6%	30.2%	0.52
PoznanStreet	SU	0.0%	95.5%	13.42	103.3%	12.3%	-12.8%	-19.5%	0.52
MIV		**26.0%**	**43.2%**	**14.40**	**75.3%**	**23.0%**	**6.5%**	**54.3%**	

Few significant differences between 'rf' and 'ff' configurations can be spotted:

1. high-bitrate VMAF BD-rate for SA: 18.2% (loss) for ‘rf’ and -40.5% (gain) for ‘ff’ (reason unknown),
2. big difference for all metrics for SE (probably because the 17 frame selected for ‘rf’ are the most challenging ones + artiacts in the associated depth maps).
3. huge difference for all metrics for ST (probably because of different content for ‘rf’ and ‘ff’ caused by moving camera rig).

We understand the motivation to progress faster with experiments on ‘rf’ only but these results suggest to do that with CAUTION! Also, it is recommended to use the ‘ff’ version when reporting results outside the MIV group.

4 Recommendations

We recommend to:

- Consider making at least two Poznań optional sequences as mandatory content (preferably SP and SU, which have better depth maps than ST),
- Revert A97/A17 anchors back to the 3groups case for the perspective content while finding an alternative way to meet the CTC constraints (i.e. # decoder instantiation) without sacrificing quality,
- Use ‘rf’ with CAUTION and only report ‘ff’ results to people outside the MIV group.

5 References

[N19213] Test Model 5 for Immersive Video,
ISO/IEC JTC1/SC29/WG11 MPEG/N19213, Alpbach, April 2020.

[N19214] Common Test Conditions for Immersive Video,
ISO/IEC JTC1/SC29/WG11 MPEG/N19214, Alpbach, April 2020.