Feature-Driven Decoder-Side Depth Estimation in MPEG Immersive Video

Patrick Garus, Felix Henry, Joel Jung,
Thomas Maugey, Christine Guillemot

Summer School on Video Coding and Processing – SVCP 2020
Berlin, July 2021

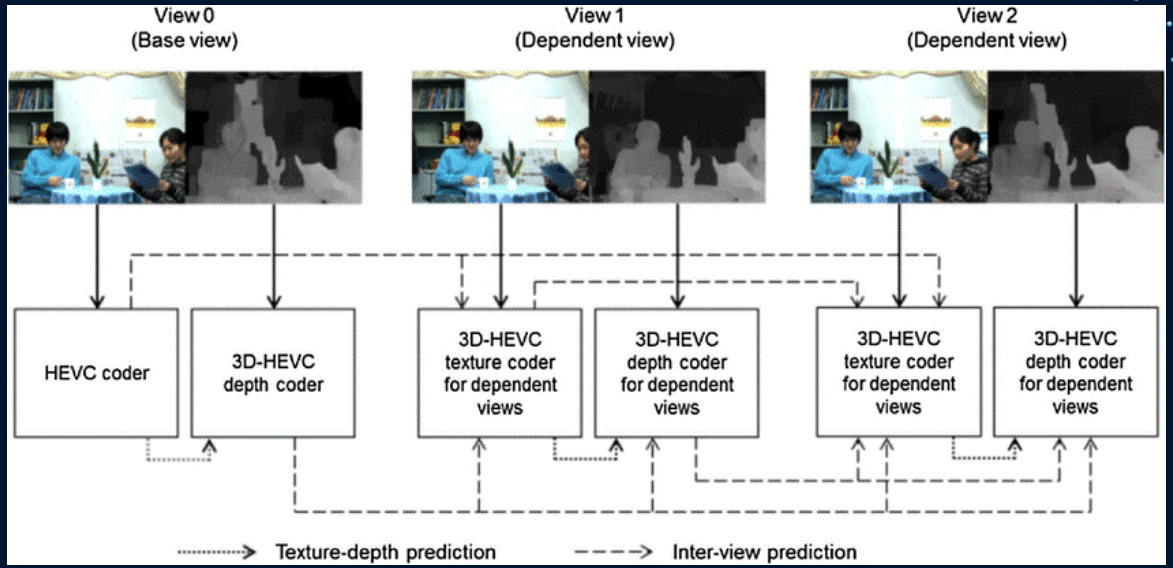
Outline

- Introducing Decoder Side Depth Estimation
- Feature-Driven Decoder Side Depth Estimation
- Decoder Side Depth Estimation in MPEG Immersive Video

Outline

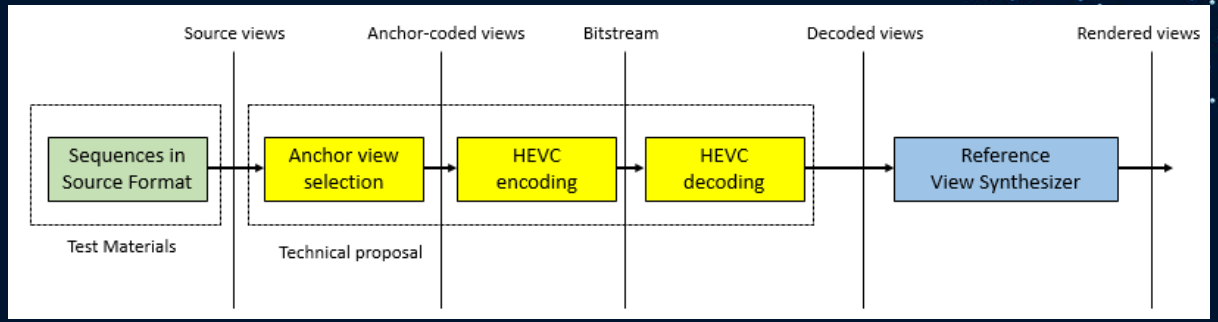
- **Introducing Decoder Side Depth Estimation**
- Feature-Driven Decoder Side Depth Estimation
- Decoder Side Depth Estimation in MPEG Immersive Video

3D-HEVC



- **Concept of 3D-HEVC**
  - Inter-View & Inter-Component Prediction
  - Dedicated depth coding tools

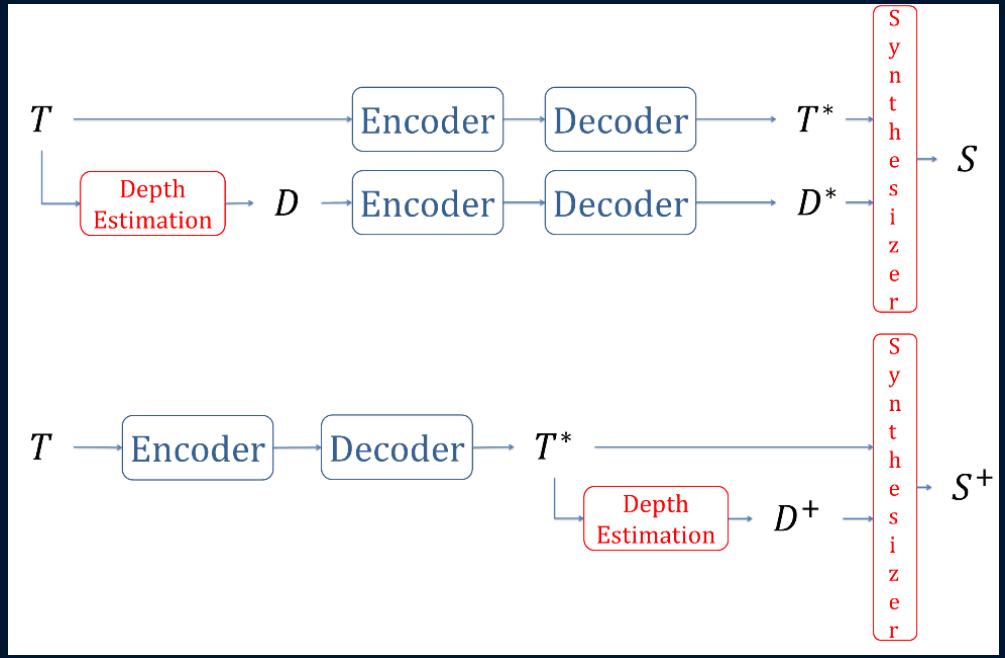
2019: CfP for Future Immersive Video (now: MIV)



- **Coding based on Legacy 2D codecs**
  - Depth are encoded with texture 2D texture codecs.
    - Compression Artifacts in depth reduce synthesis performance
  - „Pre-Processing“ to remove redundancy & define MIV metadata
    - Reduced BD-Rate performance in order to meet Pixel Rate constaint.

Our proposal: Decoder Side Depth Estimation

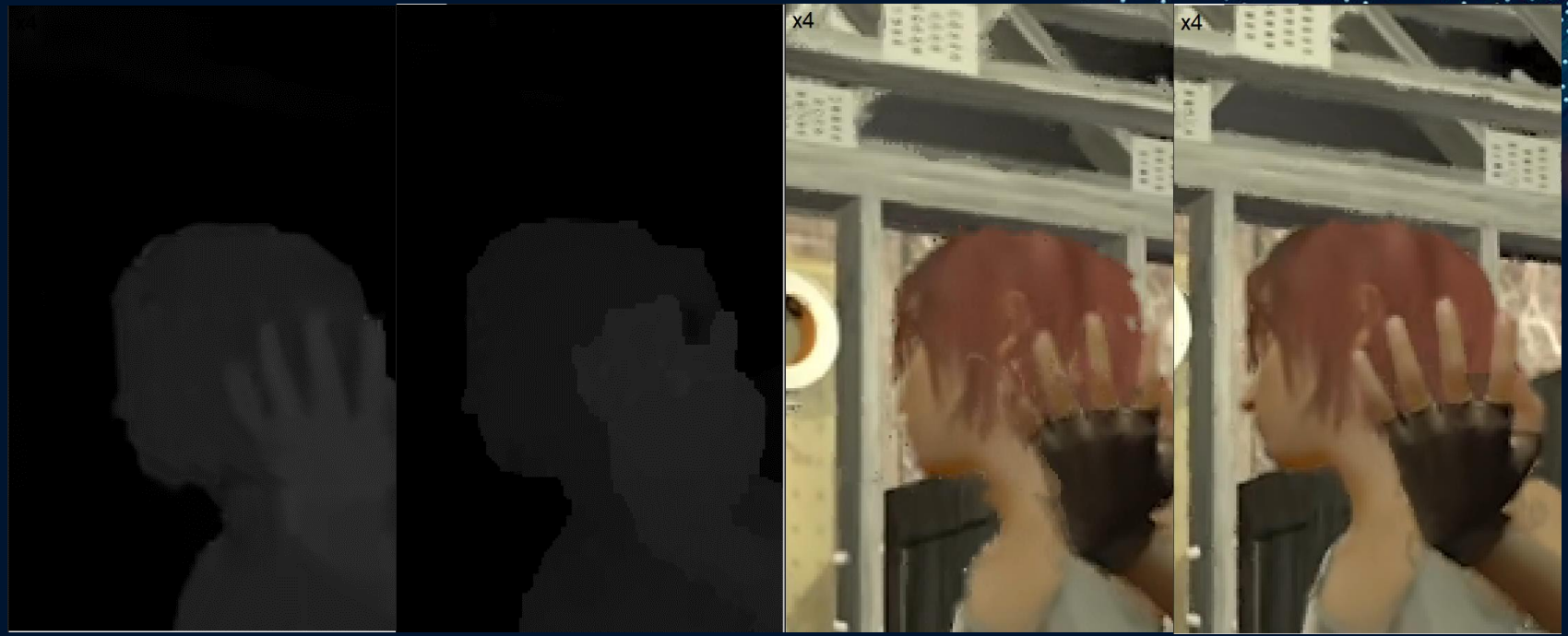
- A lot of problems of MIV are related to the depth transmission
- Instead of transmitting depth in a harmful & expensive way, we propose DSDE:



Encoder Side Depth Estimation – ESDE
e.g. 3D-HEVC, MIV, …

Decoder Side Depth Estimation – DSDE
e.g. „MIV Geometry Absent“ profile

Example – ESDE vs. DSDE (high bitrate)



ESDE-Depth DSDE-Depth ESDE-Synthesis DSDE-Synthesis

Comparative Study: ESDE vs DSDE

- **Depth Compression (MV-HEVC) is more harmful than Decoder Side Depth Estimation.**

TABLE I: BD-Rate and MS-SSIM per test sequence.

Sequence	Medium BD-Rate [%]		Low BD-Rate [%]		MS-SSIM	
	video	synth	video	synth	ESDE	DSDE
TPainter	-36.0	-31.7	-39.2	-35.1	0.9487	0.9448
UUnicornA	-4.7	6.0	-5.4	-6.7	0.9744	0.9746
UUnicornB	-5.6	-3.6	-6.7	-13.8	0.9743	0.9745
OShaman	-32.9	-55.2	-39.7	-50.0	0.9217	0.9212
OKitchen	-18.2	-39.2	-22.3	-36.0	0.9444	0.9452
ODancing	-5.3	-72.4	-8.0	-51.4	0.9749	0.9738
EChef2	-27.8	-58.4	-31.2	-54.3	0.9456	0.9478
IFrog	-11.4	-57.8	-15.0	-42.8	0.8968	0.9036
PFencing	-48.5	-23.2	-53.6	-39.9	0.9276	0.9248
Average	**-21.2**	**-37.3**	**-24.6**	**-36.7**	**0.9454**	**0.9456**

- **Improved synthesis performance with 0 bitrate spent for depth.**
- **Perceptual quality preserved.**
- **Reduction of pixel rate by 50%.**
- **No QP_D definition required.**

P. Garus, J. Jung, T. Maugey and C. Guillemot, ``Bypassing Depth Maps Transmission For Immersive Video Coding,'' 2019 Picture Coding Symposium (PCS), Ningbo, China, 2019

Outline

- Introducing Decoder Side Depth Estimation
- **Feature-Driven Decoder Side Depth Estimation**
- Decoder Side Depth Estimation in MPEG Immersive Video

FD-DSDE - Concept

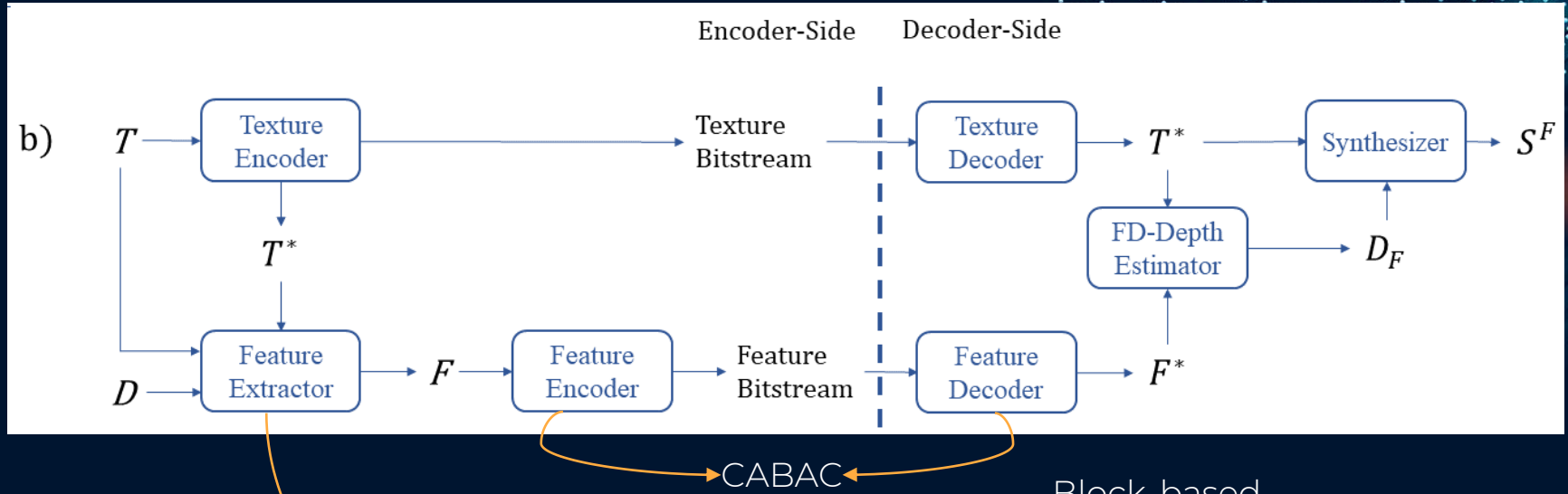
- Divide views into blocks.
- For each block, transmit a few parameters (features) to help depth estimation of the decoder side.
- Optimize parameters so the block-based depth is as close as possible to the GT depth.

Transmitted features

1. **Partition flag**
2. **Skip flag**
3. **Zmin/Zmax**
4. **Depth Estimation Parameters**

P. Garus, F. Henry, J. Jung, T. Maugey and C. Guillemot, ``Immersive Video Coding: Should Geometry Information be Transmitted as Depth Maps?,'' Transaction on Circuits and Systems for Video Technology (TCSVT), 2021, *accepted*.

FD-DSDE – System Overview

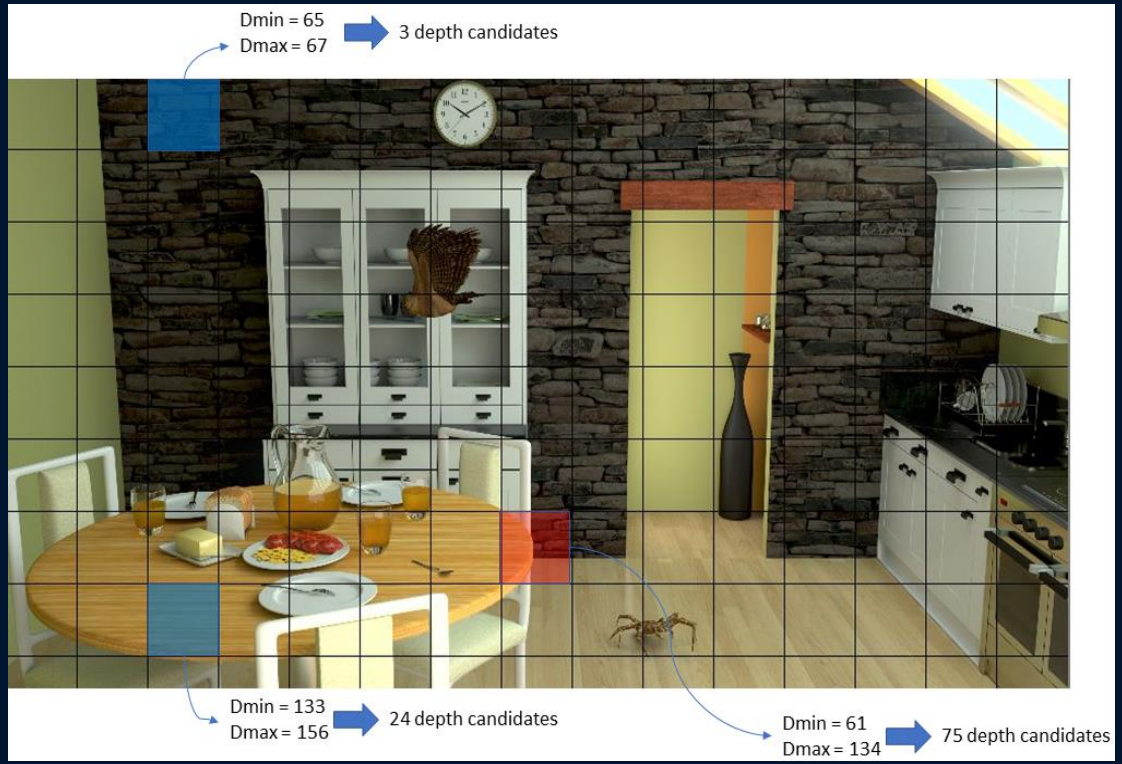


Block-based
Encoder-Side depth estimator*
(feature extractor)

Block-based
Decoder-Side depth estimator*
(Feature driven)

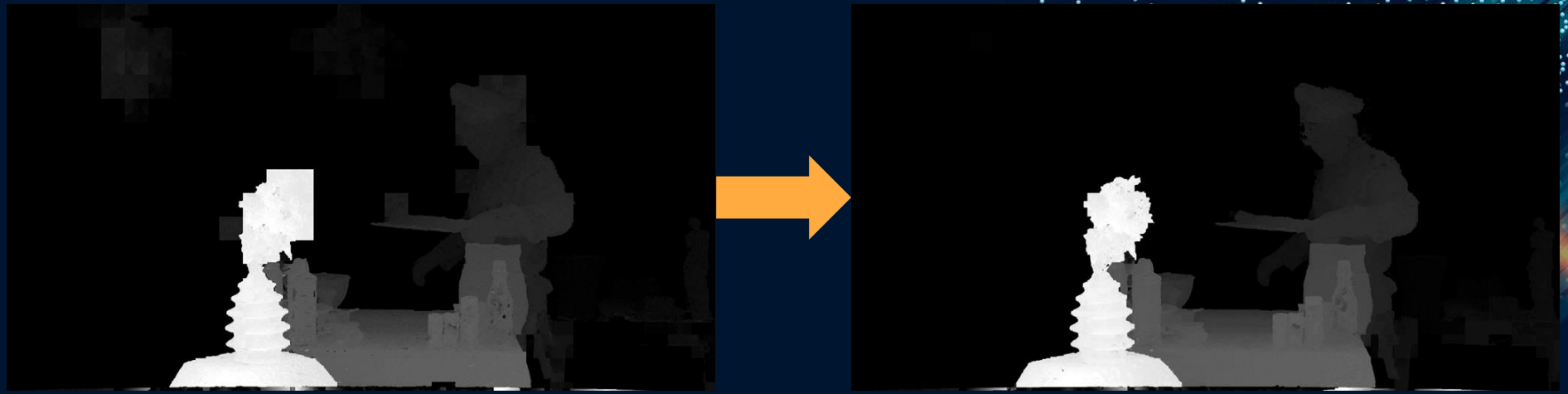
*based on DERS8

Feature #1 : Z_{min} & Z_{max}(d_{max}& d_{min})



- **Reduced complexity** by minimized Cost Volume.
- **Increased quality** by simplified cost aggregation.

Feature #2 : Depth Estimation Parameters

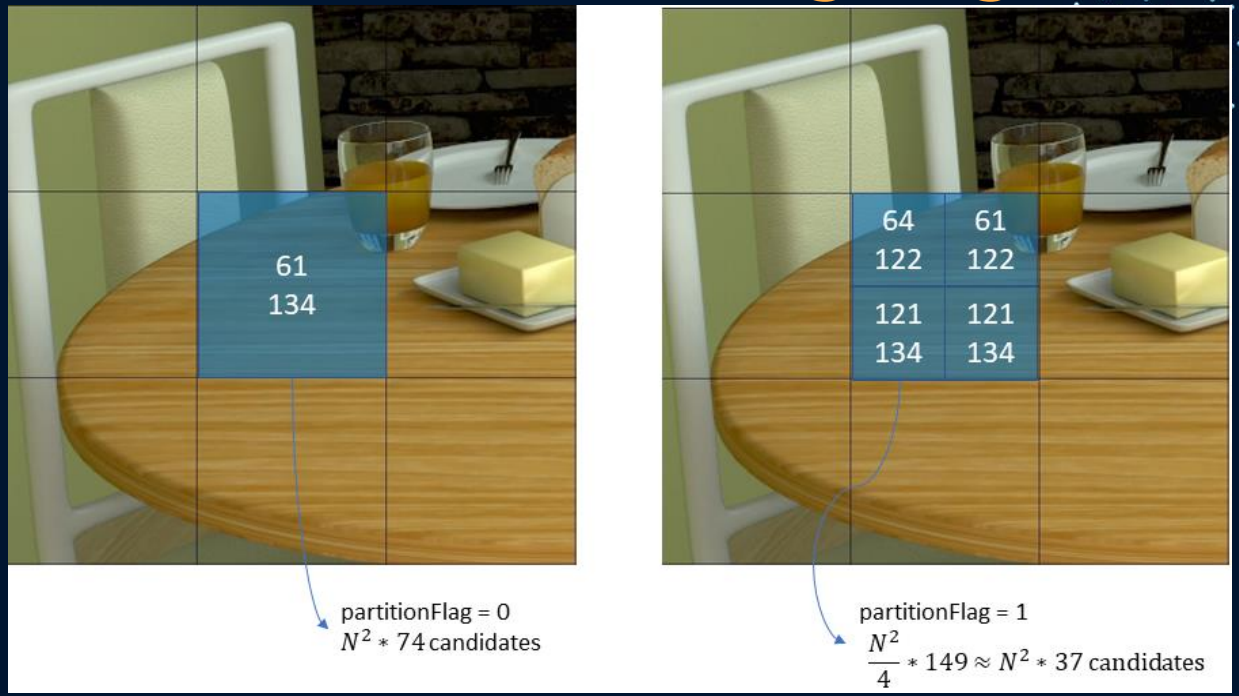


- Optimized smoothing coefficient per block
- Coefficient selected to minimize L2-distance to a Ground Truth depth map.
- Mitigates downside of “block-based” approach & estimation from decoded textures.

Feature #3 : Skip Flag

- Entirely skip depth estimation if no depth change detected at the encoder.
- Further reduces complexity for future frames, provides temporal stability & reduces bitrate for features.

Feature #4 : Partitioning Flag



- Enables partitioning of blocks.
- Further minimizes complexity or improves performance.

FD-DSDE: Complexity reduction

- Maintained pixel rate reduction of -50% for all sequences.
- Cost volume reduced by 88% on average (up to 98%)
- Runtime reduced by a factor of 18 compared to the DERS8 reference

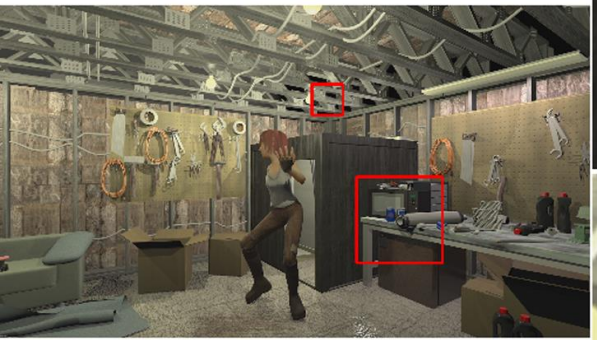
TABLE IV: Average runtime per frame and view using the unmodified DERS8 and FD-DSDE.

Sequence	DERS8 [s]	FD-DSDE [s]	speedup
Painter	698.3	111.0	6.3
UnicornA	291.4	40.4	7.2
UnicornB	270.5	35.2	7.7
Shaman	3933.9	169.1	23.3
Kitchen	1051.5	22.6	46.5
Dancing	1195.0	30.2	39.6
Chef2	2565.1	228.8	11.2
Frog	1107.3	177.6	6.2
Average	**1389.1**	**101.85**	**18.5**

FD-DSDE: Objective Performance

Config	Sequence	synth PSNR [dB			LPIPS		
		ESDE	B-DSDE	FD-DSDE	ESDE	B-DSDE	FD-DSDE
Medium Bitrate	Painter	**34.38**	34.23	34.25	0.220	0.218	0.218
	UnicornA	29.58	29.38	**29.92**	0.059	0.063	**0.056**
	UnicornB	29.88	29.84	**30.42**	0.058	0.059	**0.055**
	Shaman	33.61	34.21	34.24	0.255	**0.243**	0.250
	Kitchen	30.55	31.05	**31.45**	0.181	0.176	**0.169**
	Dancing	28.15	29.97	**30.06**	0.218	**0.191**	**0.191**
	Chef2	31.50	31.91	31.70	0.260	0.260	**0.257**
	Frog	26.95	**27.59**	27.53	0.229	**0.208**	0.215
	Average	30.57	31.02	**31.20**	0.185	0.177	**0.176**
Low Bitrate	Painter	**32.96**	32.75	32.80	0.334	**0.331**	0.332
	UnicornA	28.21	28.20	**28.67**	0.123	0.122	**0.115**
	UnicornB	28.56	28.65	**29.12**	0.123	0.121	**0.118**
	Shaman	32.46	32.75	**32.93**	0.430	**0.418**	0.423
	Kitchen	29.42	29.79	**30.27**	0.324	0.317	**0.308**
	Dancing	27.02	28.11	**28.88**	0.386	0.374	**0.356**
	Chef2	30.82	31.29	31.13	0.323	0.322	**0.319**
	Frog	25.99	26.59	**26.60**	0.363	**0.337**	0.346
	Average	29.43	29.77	**30.05**	0.301	0.293	**0.290**

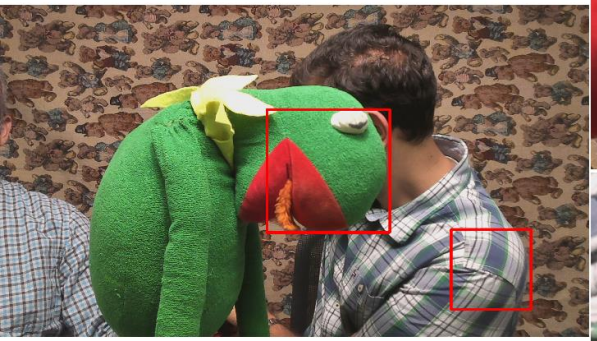
FD-DSDE: Subjective performance (1)



Dancing

Ground-Truth ESDE B-DSDE FD-DSDE

FD-DSDE: Subjective performance (2)



Frog

Ground-Truth ESDE B-DSDE FD-DSDE

FD-DSDE: Comparison with 3D-HEVC

TABLE VII: Synthesis PSNR BD-Rates using the 3-Vi configuration of the 3D-HEVC CTC. FD-DSDE average (*) underestimated due to missing overlaps for Frog and Dancin Instead, their RD-curves are shown in Fig. 10.

Config	Sequence	synth PSNR BD-Rate [%]		
		3D-HEVC	B-DSDE	FD-DSDE
Medium Bitrate	Painter	-24.3	**-36.4**	-27.7
	UnicornA	-58.3	-42.3	**-64.6**
	UnicornB	**-59.2**	-33.8	-44.1
	Shaman	**-59.5**	-46.8	-35.9
	Kitchen	**-66.4**	-61.1	-57.2
	Dancing	-62.2	-83.4	**no overlap**
	Chef2	-67.1	**-82.3**	-47.9
	Frog	-62.8	-79.9	**no overlap**
	Average	-57.5	-58.2	-46.0*
Low Bitrate	Painter	-25.0	**-36.5**	-27.0
	UnicornA	-39.3	-38.3	**-47.5**
	UnicornB	**-41.6**	-37.4	-40.0
	Shaman	**-50.6**	-25.5	-22.7
	Kitchen	**-57.4**	-50.2	-53.6
	Dancing	-32.8	-82.2	**no overlap**
	Chef2	-55.3	**-61.5**	-40.7
	Frog	-40.5	-58.6	**-77.8**
	Average	-42.8	-48.8	-44.2*

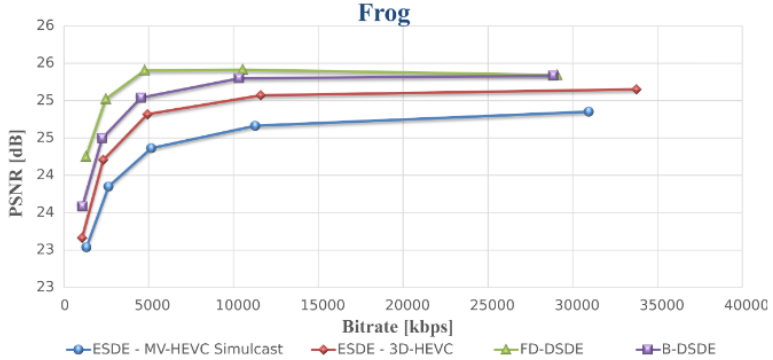


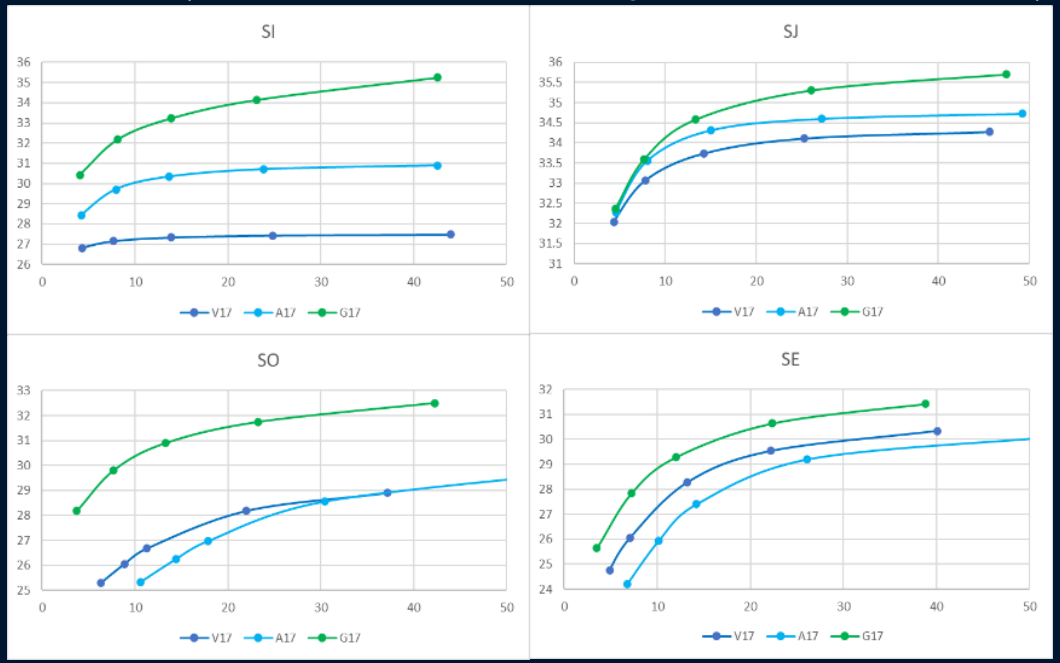
Fig. 10: RD-curves for the Dancing and Frog sequences.

Outline

- Introducing Decoder Side Depth Estimation
- Feature-Driven Decoder Side Depth Estimation
- **Decoder Side Depth Estimation in MPEG Immersive Video**

DSDE in MIV : Geometry Absent Profile

- DSDE Performance was reconfirmed using the MIV Standard involving a different depth estimator (IVDE) & different synthesizer (VWS / RVS)
- Adoption of a **MIV Geometry Absent Profile –** Depth Atlases are not transmitted.



- GA Profile outperforms other MIV anchors (GA=G17 – green curve).
- Especially for perspective content.

FD-DSDE in MIV : Geometry Assistance SEI (I)

- The FD-DSDE Features have been adopted into the MIV as **Geometry Assistance SEI**.
- With the exception of Depth Estimation Parameters
- **Depth Estimation & Synthesis are not part of the specification**

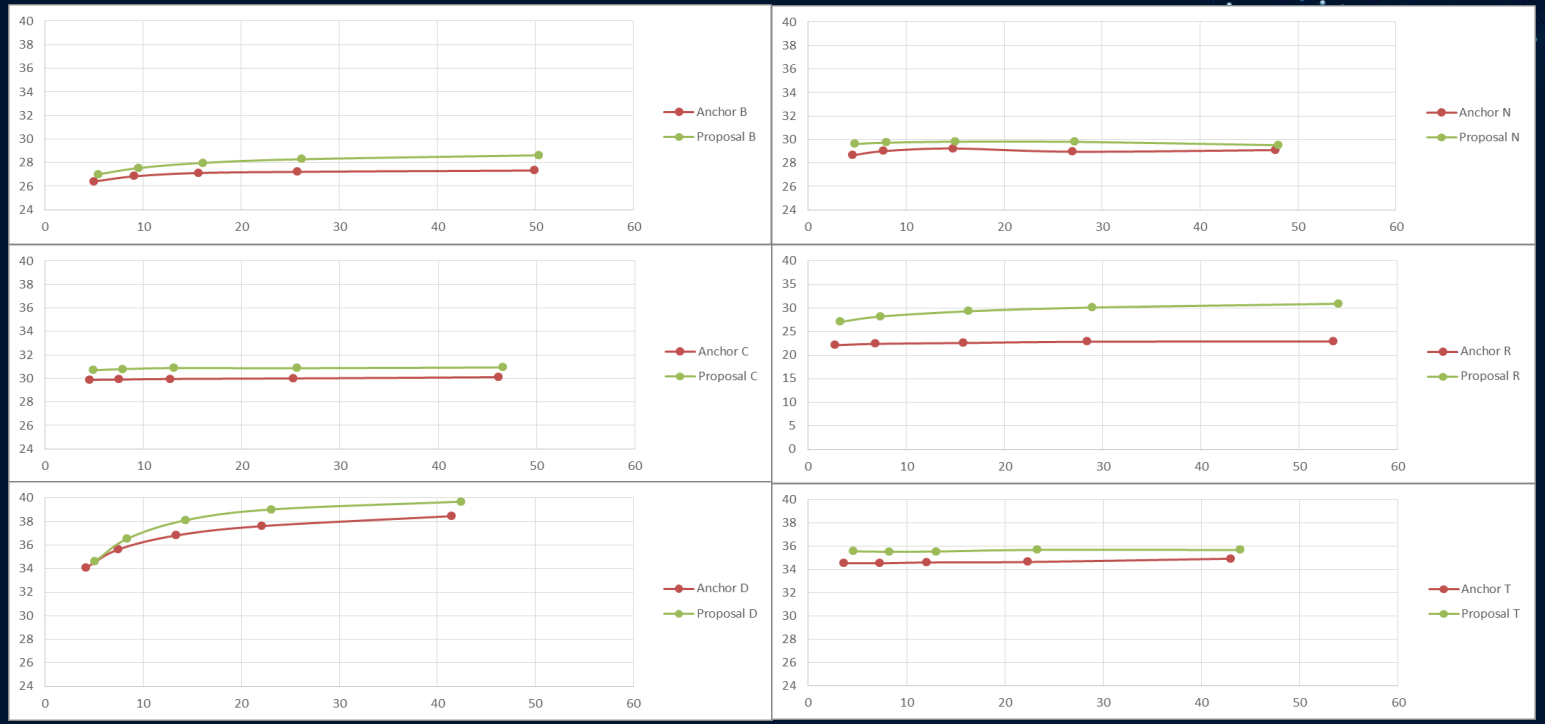
- Concept of Features supported in IVDE 3.0

- Speedup of decoder runtime x2:

Mandatory content			
Sequence		**TMIV decoding (incl. IVDE)**	**Depth map estimation**
ClassroomVideo	A	30.8%	24.2%
Museum	B	63.1%	57.9%
Fan	O	46.2%	43.5%
Kitchen	J	67.4%	64.2%
Painter	D	49.0%	46.5%
Frog	E	23.3%	22.3%
Carpark	P	71.3%	69.1%
Chess	N	78.1%	73.7%
Group	R	56.5%	53.6%
MIV		54.0%	50.6%
Optional content			
Fencing	L	28.8%	26.2%
Hall	T	44.5%	42.9%
Street	U	49.3%	47.8%
ChessPieces	Q	83.8%	83.2%
Hijack	C	78.7%	63.8%
Mirror	I	59.3%	55.8%
MIV		57.4%	53.3%

FD-DSDE in MIV : Geometry Assistance SEI (2)

- > 1 dB synthesis PSNR gain for 6 sequences.

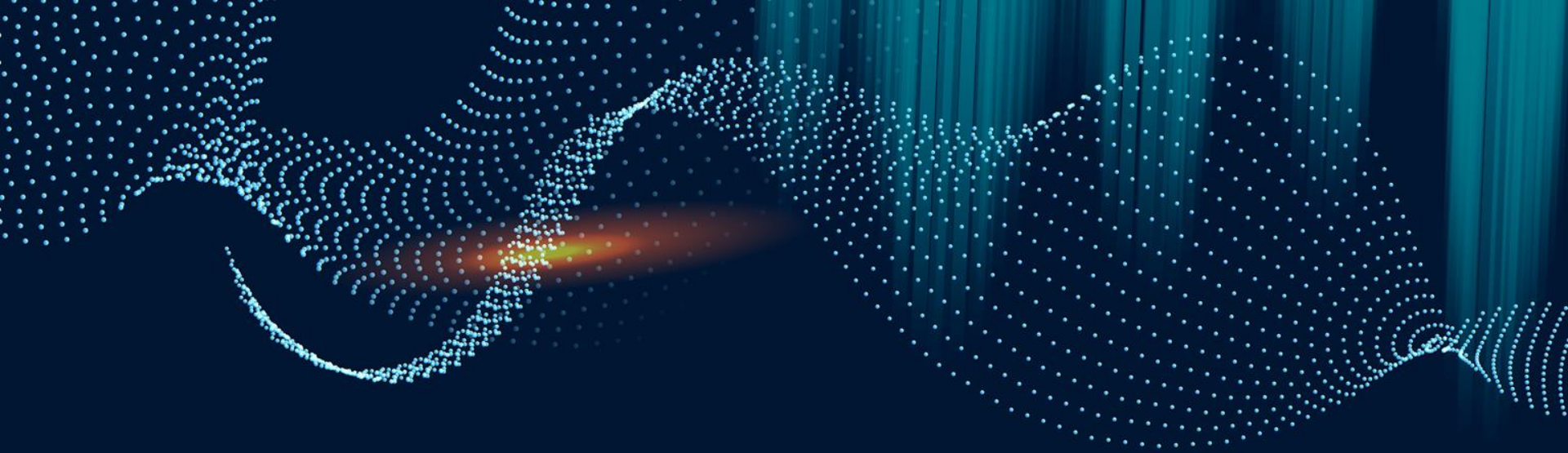


Summary

- In terms of research, DSDE is a „new“ system.
- It already outperforms ESDE-type systems easily.
- Complexity aspect has to be solved, but we are optimistic.

- DSDE is part of the new MIV standard as the **MIV Geometry Absent Profile.**
- **The Geometry Assistance SEI** message improves the complexity as well as the quality.

- Many more synergies can be found
- Merging **texture decoder + depth estimator**
- Merging **depth estimator + renderer**



Thank you for your attention

Questions?