Patch Decoder-Side Depth Estimation in MPEG Immersive Video

Marta Milovanović, Félix Henry, Marco Cagnazzo, Joël Jung

Orange Labs

LTCI, Télécom Paris, Institut Polytechnique de Paris

Tencent Media Lab

Summer School on Video Coding and Processing - SVCP2020
Berlin, July 2021

Outline

- MPEG Immersive Video
- Test Model for Immersive Video
- Proposed Method
- Results
- Discussion and Summary

Outline

- **MPEG Immersive Video**
- Test Model for Immersive Video
- Proposed Method
- Results
- Discussion and Summary

MPEG Immersive Video (MIV)

- ISO/IEC 23090-12 MPEG Immersive Video (MIV) specification
- Standard for **streaming** and **storage** of immersive content
- Aims to provide the 6 DoF (degrees of freedom) user experience
- MIV does not utilize specialized depth coding tools
- **MIV constraints:** bitrate, pixel rate, number of simultaneous 2D decoders
- Variety of use cases and devices: head mounted displays, light field displays, tablets, laptops…

https://mpeg-miv.org/
https://venturebeat.com/2019/05/05/how-virtual-reality-positional-tracking-works/

MPEG Immersive Video (MIV)

- Input format: Multiview video plus depth (MVD)
- 3D scene is captured by multiple real or virtual cameras
- Depth maps can be computer generated or natural
- Computer generated depth maps are obtained using mathematical models of a 3D scene
- Natural depth maps are obtained by a sensor or some depth estimation algorithm (not perfect)
- Output format: texture and depth atlases
- These video data streams are compressed by 2D video codecs

https://mpeg-miv.org/index.php/content-database-2/

Outline

MPEG Immersive Video

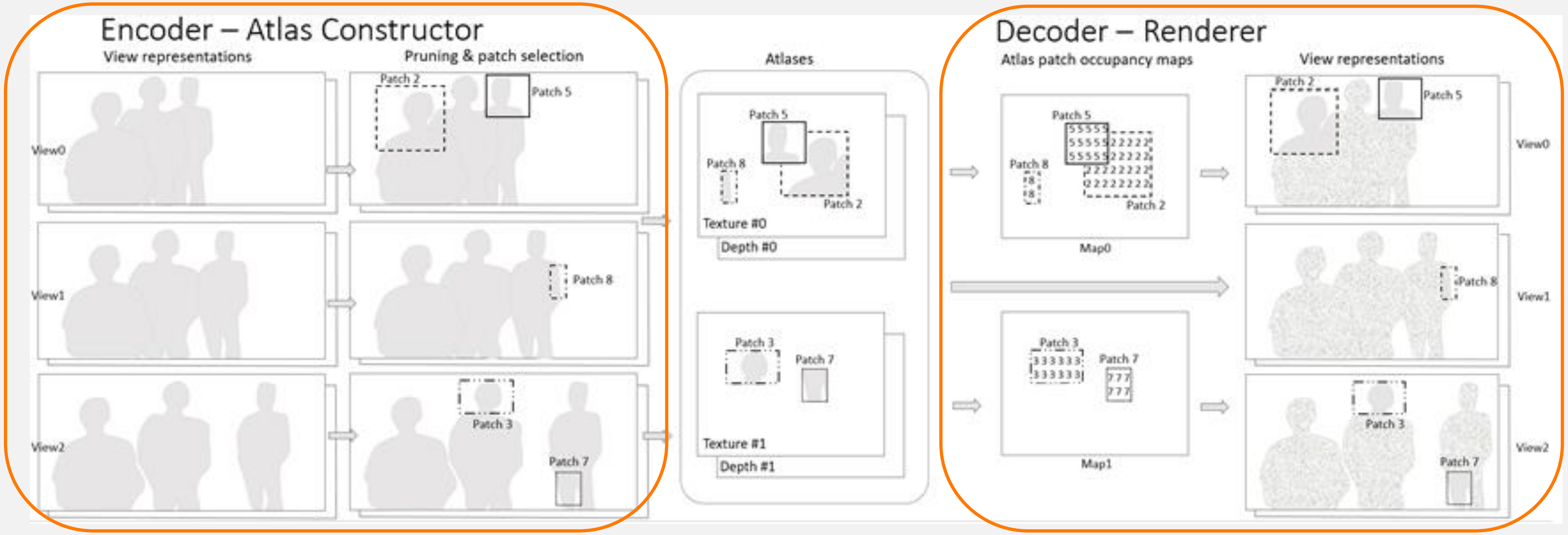
Test Model for Immersive Video

Proposed Method

Results

Discussion and Summary

Test Model for Immersive Video (TMIV)



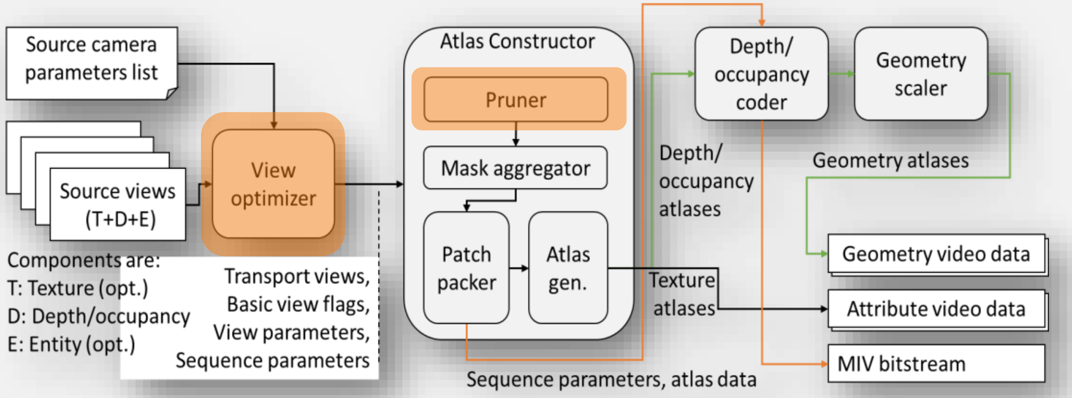
Representing source views using patch atlases

Pruned view reconstruction

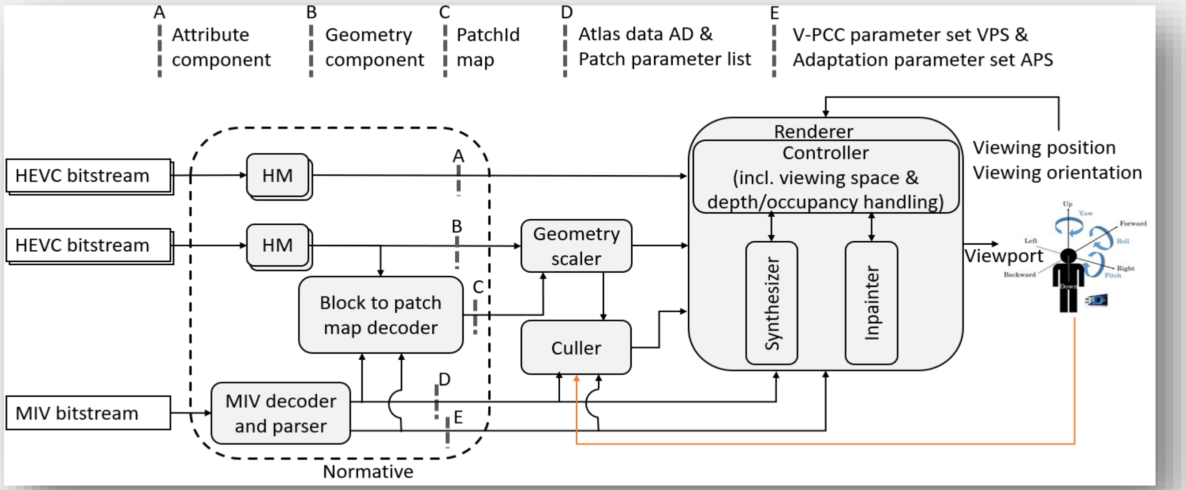
Basel Salahieh, Bart Kroon, Joel Jung, and Marek Domanski,
"Test Model 4 for Immersive Video," ISO/IEC JTC 1/SC29/WG 11 N19002, Feb. 2020.

TMIV Encoder & Decoder

Encoder



Decoder

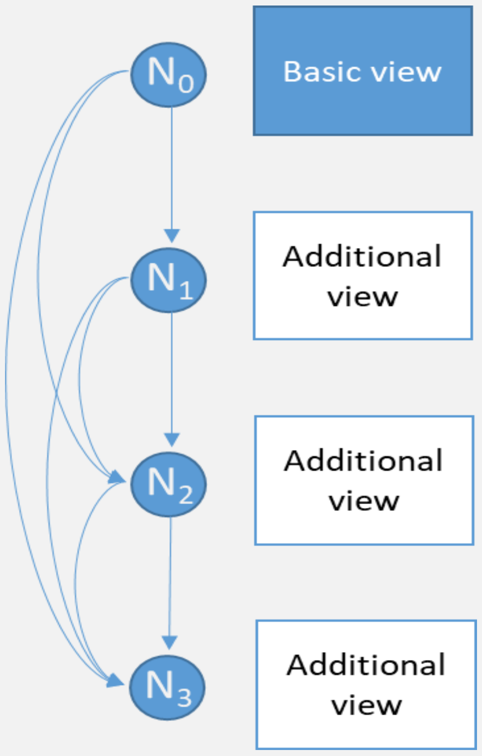


Basel Salahieh, Bart Kroon, Joel Jung, and Marek Domanski,
“Test Model 4 for Immersive Video,”
ISO/IEC JTC 1/SC29/WG 11 N19002,
Feb. 2020.

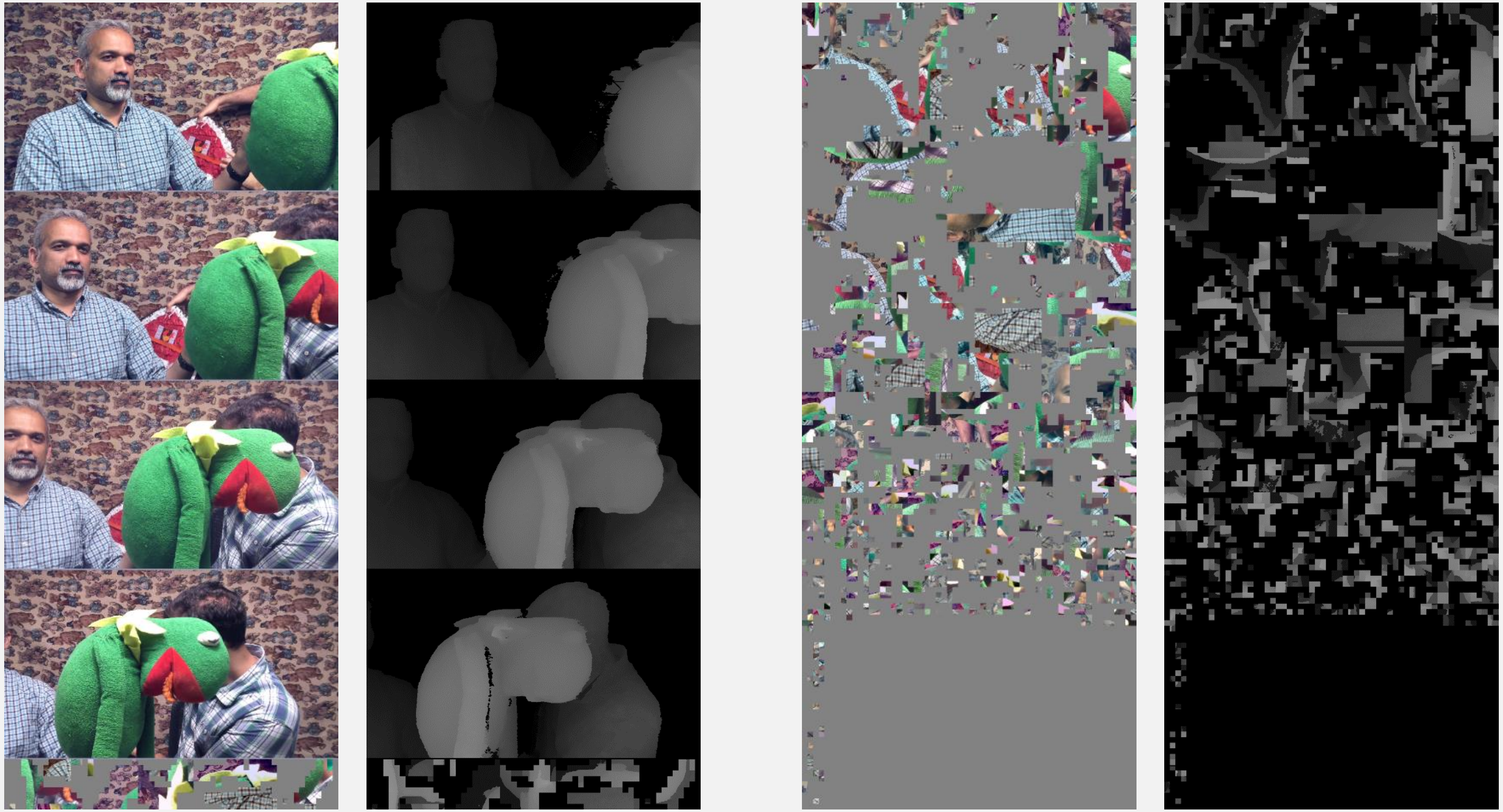
TMIV Pruning

- The pruning process **removes the inter-view redundancy**
- The pruner uses three criteria to determine if a pixel may be pruned:
  - The pixel should be synthesized from the views higher up in the hierarchy
  - The difference between synthesized and source geometry should be less than a threshold
  - The minimum difference between luma of a synthesized pixel and luma of all pixels within a collocated source 3×3 block should be less than a pruning luma threshold
- **Second-pass pruning:** identifying the pixels that are not to be pruned among the pixels that were initially determined to be pruned (global color component differences)
- **Temporal consistency:** the pruning masks are aggregated frame-by-frame and reset at the beginning of each intra period

The pruning graph is created:



TMIV Atlases Demo



Outline

MPEG Immersive Video

Test Model for Immersive Video

Proposed Method

Results

Discussion and Summary

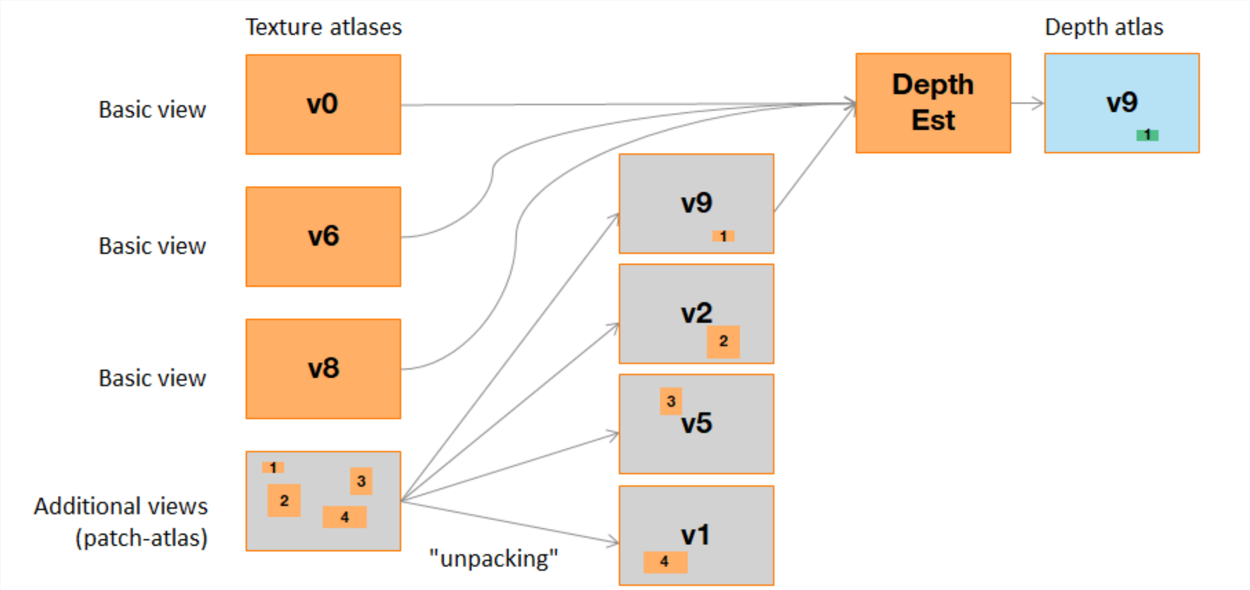
Proposed Method

- **Goal:**
  - Reduce the transmission of the depth data in the context of TMIV
- **Motivation:**
  - MIV constraints (bitrate, pixel rate, number of 2D decoders)
  - Decoder-Side Depth Estimation: transmission of the depth maps is not needed
- **The main idea:**
  - Omit the depth component of some patches that belong to pruned views

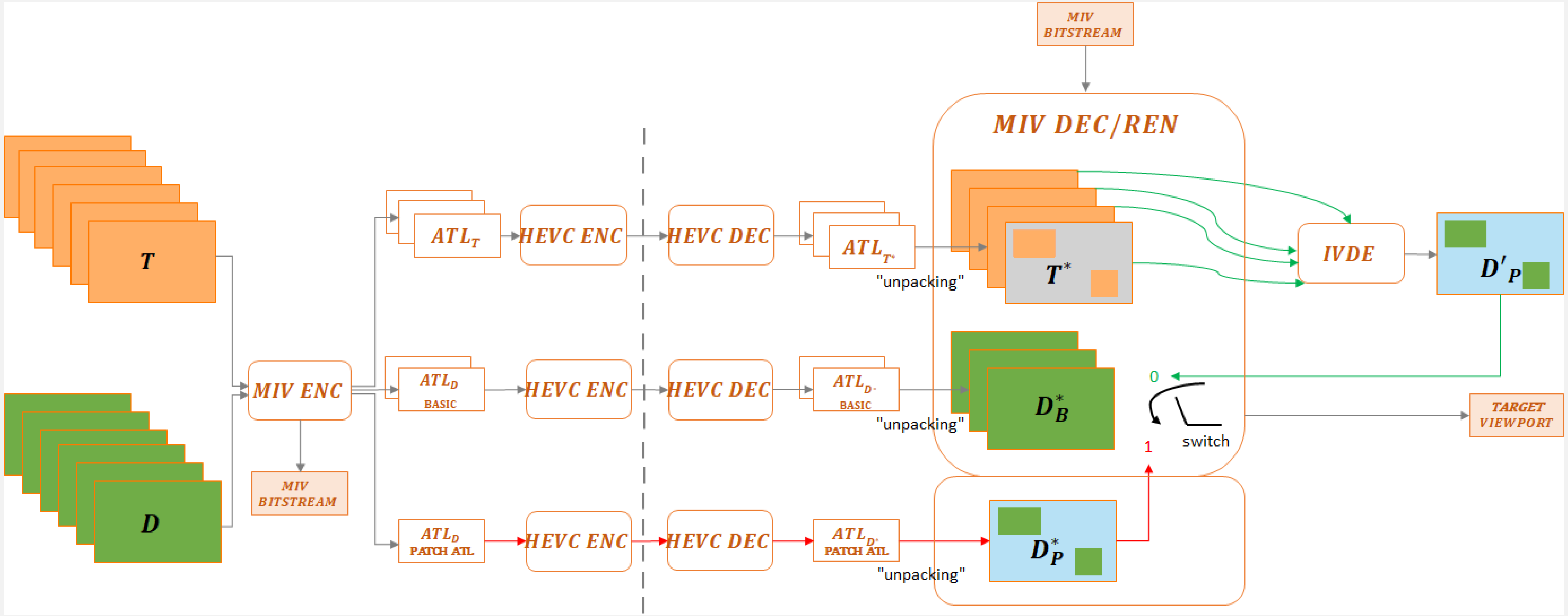
Patrick Garus, Joel Jung, Thomas Maugey, and Christine Guillemot, "Bypassing Depth Maps Transmission For Immersive Video Coding," in *2019 Picture Coding Symposium (PCS)*, Ningbo, China, Nov. 2019, pp. 1–5, IEEE.

Proposed Method

- Use the information available at the decoder side:
- Recover the pruned views: take the patches from atlases and put them to the correct positions in corresponding views using the metadata information
- Estimate the patch-depth using all textures from available basic views and the corresponding pruned view



Proposed Method



Process diagram for anchor (switch = 1) and proposed method (switch = 0) in TMIV framework

Outline

MPEG Immersive Video

Test Model for Immersive Video

Proposed Method

Results

Discussion and Summary

Synthesis Results

Sequence	CTC - High bitrate	CTC - Medium bitrate	Low bitrate
Shaman (CG)	26.34	8.99	0.73
Kitchen (CG)	66.95	30.33	10.72
Painter (NC)	2.75	-7.81	-12.86
Frog (NC)	3.57	-3.49	-8.01
Fencing (NC)	-12.33	-16.02	-18.35
Carpark (NC)	0.26	-8.33	-12.63
Street (NC)	-6.10	-8.67	-10.65
Hall (NC)	-8.53	-9.48	-10.17
Average (all)	**9.11**	**-1.81**	**-7.65**
Average (NC)	**-3.40**	**-8.97**	**-12.11**

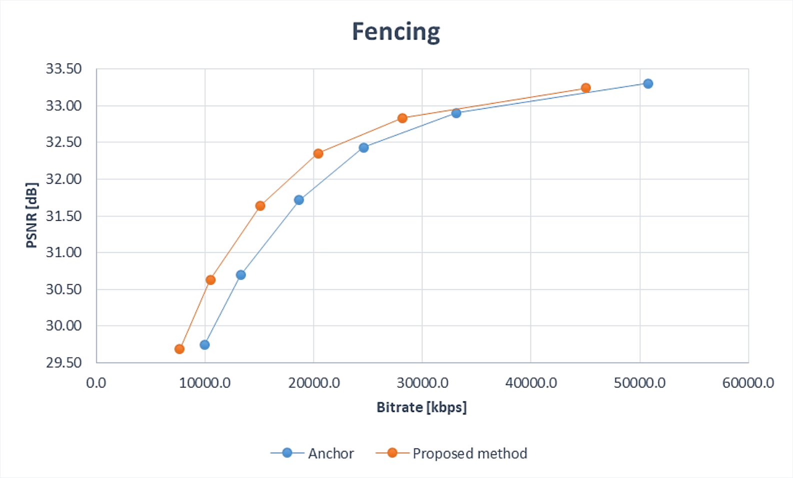
Table 1. BD-rate results per test sequence, in terms of Y-PSNR of synthesized texture [%]. Negative values indicate gains.

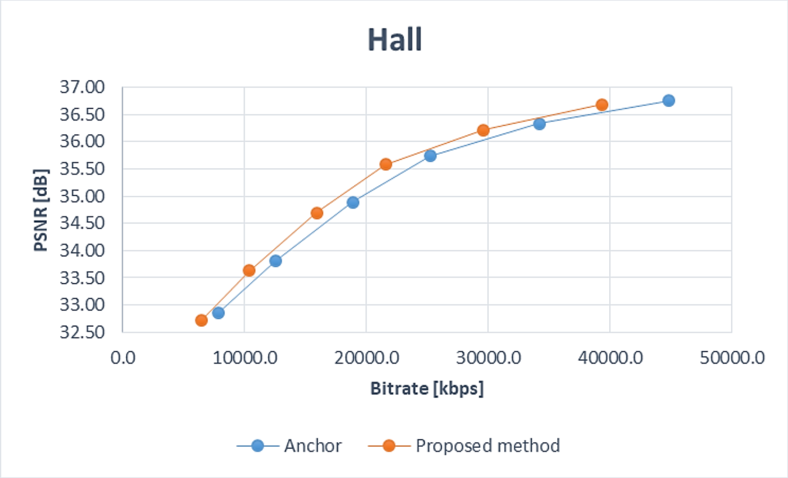
Sequence	VMAF			MS-SSIM		
	High	Med	Low	High	Med	Low
Shaman (CG)	3.44	-6.69	-11.53	20.60	1.54	-6.33
Kitchen (CG)	46.96	12.77	-0.28	20.14	5.08	-3.38
Painter (NC)	-13.04	-18.59	-21.31	-4.21	-13.96	-18.22
Frog (NC)	-3.57	-8.57	-11.13	6.19	-5.49	-10.07
Fencing (NC)	-12.91	-16.87	-19.60	-3.74	-13.29	-17.03
Carpark (NC)	-5.53	-13.14	-16.52	-1.70	-12.27	-16.15
Street (NC)	-7.00	-9.52	-11.32	-6.54	-9.31	-11.29
Hall (NC)	-10.84	-13.35	-15.09	3.93	-5.73	-10.22
Average (all)	**-0.31**	**-9.25**	**-13.35**	**4.33**	**-6.68**	**-11.59**
Average (NC)	**-8.82**	**-13.34**	**-15.83**	**-1.01**	**-10.01**	**-13.83**

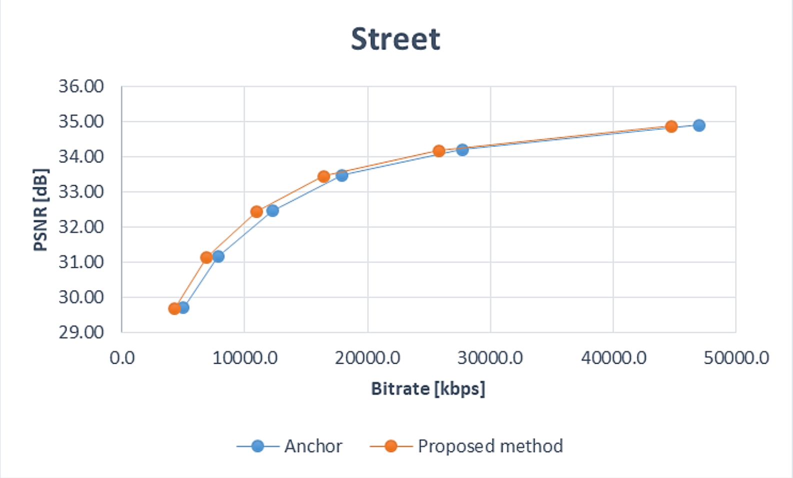
Table 2. BD-rate results per test sequence, in terms of VMAF and MS-SSIM metrics [%]. Negative values indicate gains.

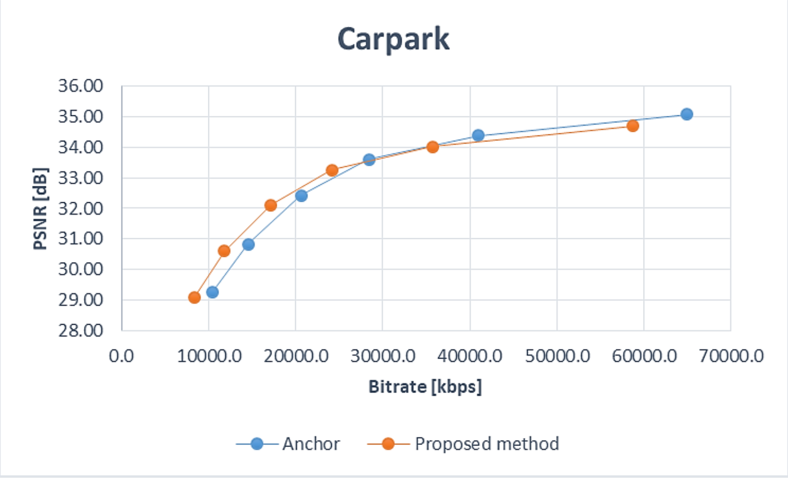
+ pixel rate reduction of 8.3% per sequence

Synthesis Results









Synthesis Results

Anchor

Our proposal

Source

Fencing ~ 10 000 kbps

Marta Milovanović, Félix Henry, Marco Cagnazzo and Joël Jung, "Patch Decoder-Side Depth Estimation in MPEG Immersive Video," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, June 2021, pp. 1945-1949, IEEE.

Outline

MPEG Immersive Video → Test Model for Immersive Video → Proposed Method → Results → **Discussion and Summary**

Discussion and Summary

- **We demonstrate:**
  - Gains measured with objective metrics for natural content:
  - 3.4%, 9.0%, 12.1% Y-PSNR BD-rate gains on high, medium, and low bitrate, respectively
  - Preserved perceptual quality as measured with MS-SSIM and VMAF
  - Pixel rate reduction of 8.3% per sequence
- For CG content, comparison is done with the Blender ground-truth depths!
- **Limitations:**
  - Pruning strategy
  - Local depth estimation on very small patch areas
  - Depth estimation from compressed textures

Discussion and Summary

- **Future work:**
  - Adapt the **pruning strategy** to ensure a reliable patch depth estimation at the decoder side
  - Current tests involve patch level decisions based on depth estimation quality
  - This approach gives new possibilities, *e.g.* sending more textures, instead of depths

Thank you for your attention !

Patch Decoder-Side Depth Estimation in MPEG Immersive Video

?

Marta Milovanović, Félix Henry, Marco Cagnazzo, Joël Jung

Orange Labs

LTCI, Télécom Paris, Institut Polytechnique de Paris

Tencent Media Lab

Summer School on Video Coding and Processing - SVCP2020
Berlin, July 2021