Hybrid Method for Patch-Depth Transmission in MPEG Immersive Video

Marta Milovanović

Supervisors: Félix Henry, Marco Cagnazzo

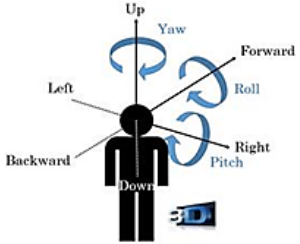
Orange Labs

LTCI, Télécom Paris, Institut Polytechnique de Paris

4th July 2022

Ilmenau, Germany

MPEG Immersive Video (MIV)

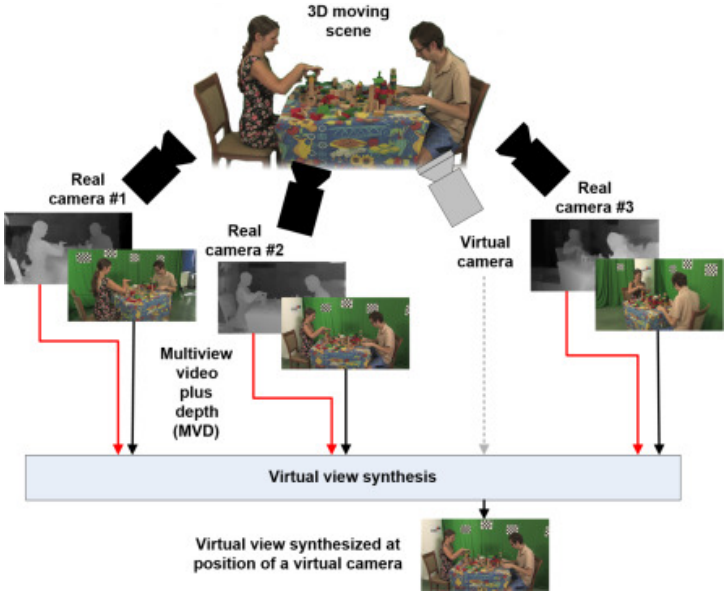


- Multiview (plus Depth) representation

- **Constraints**:
  - Bitrate
  - Pixel rate
  - Number of simultaneous 2D decoders

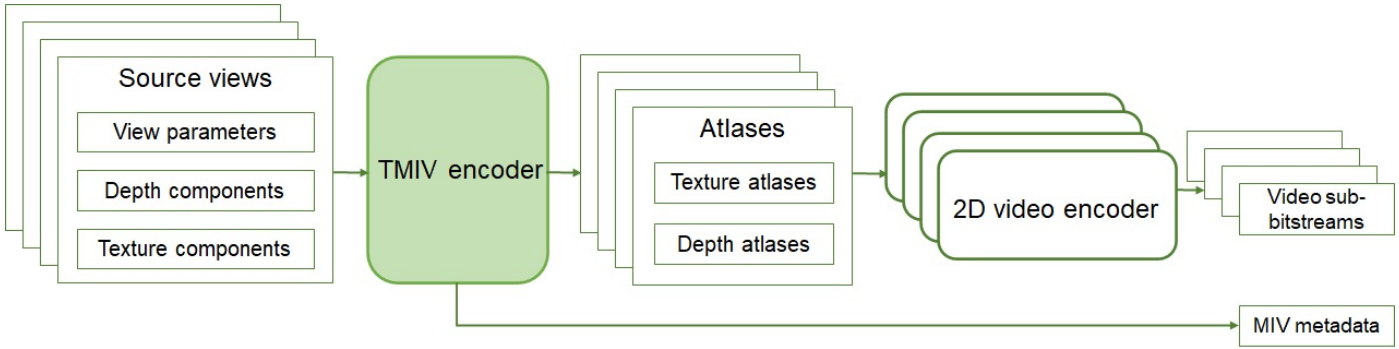
- **MIV solution**:
  - Send a few full views + the parts of other views that are essential for a good rendering (patches)
  - Re-use existing 2D video codecs

- Important non normative stages:
  - Depth estimation
  - View synthesis

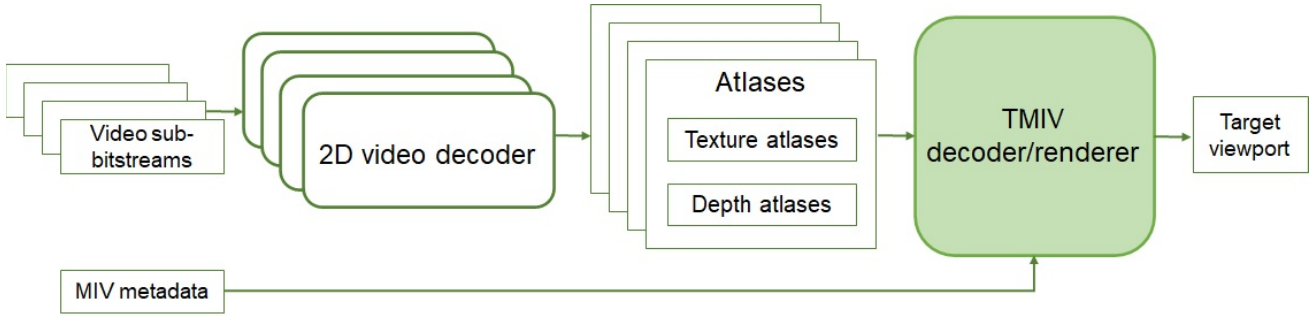


Test Model for Immersive Video (TMIV)

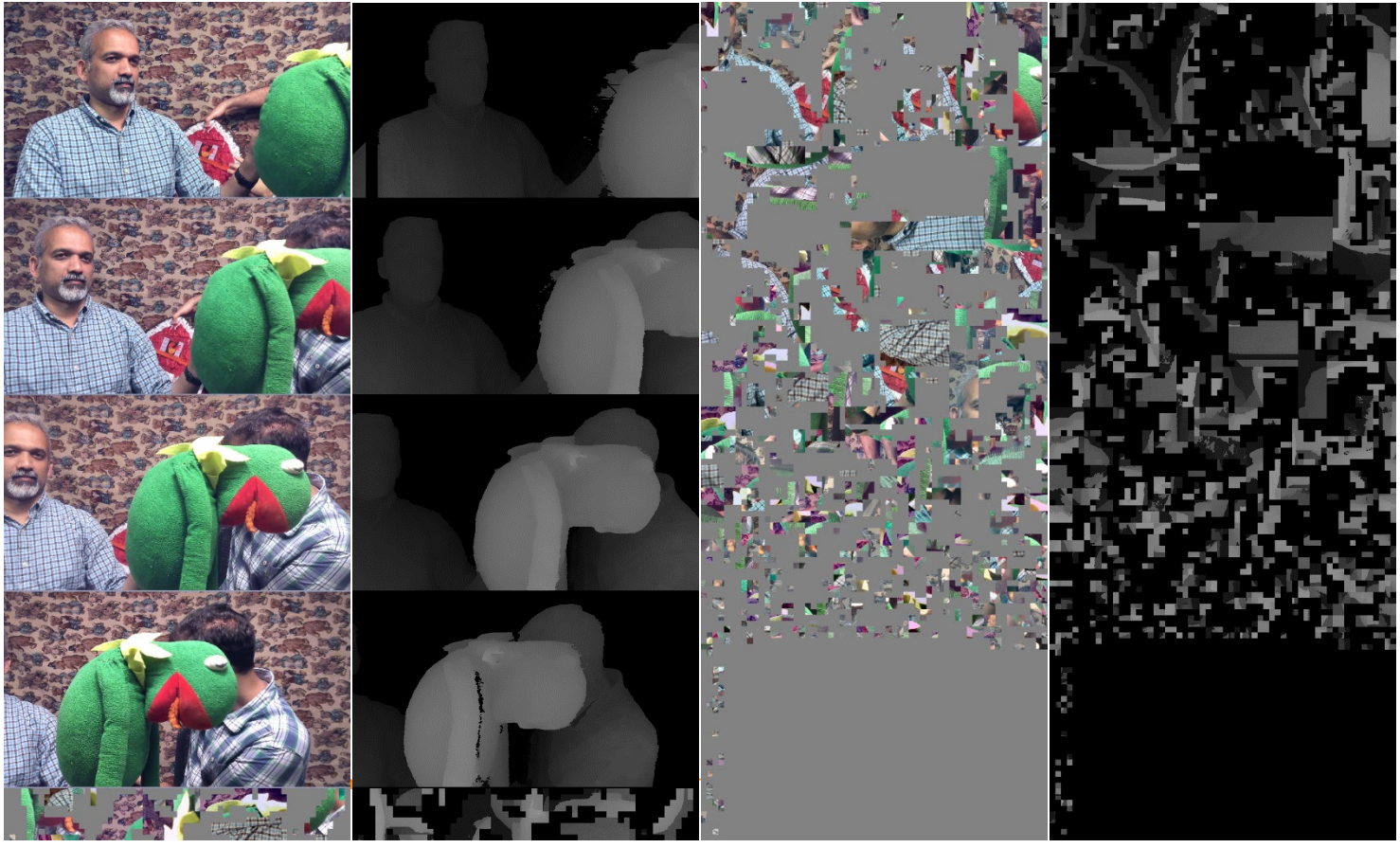
Encoder



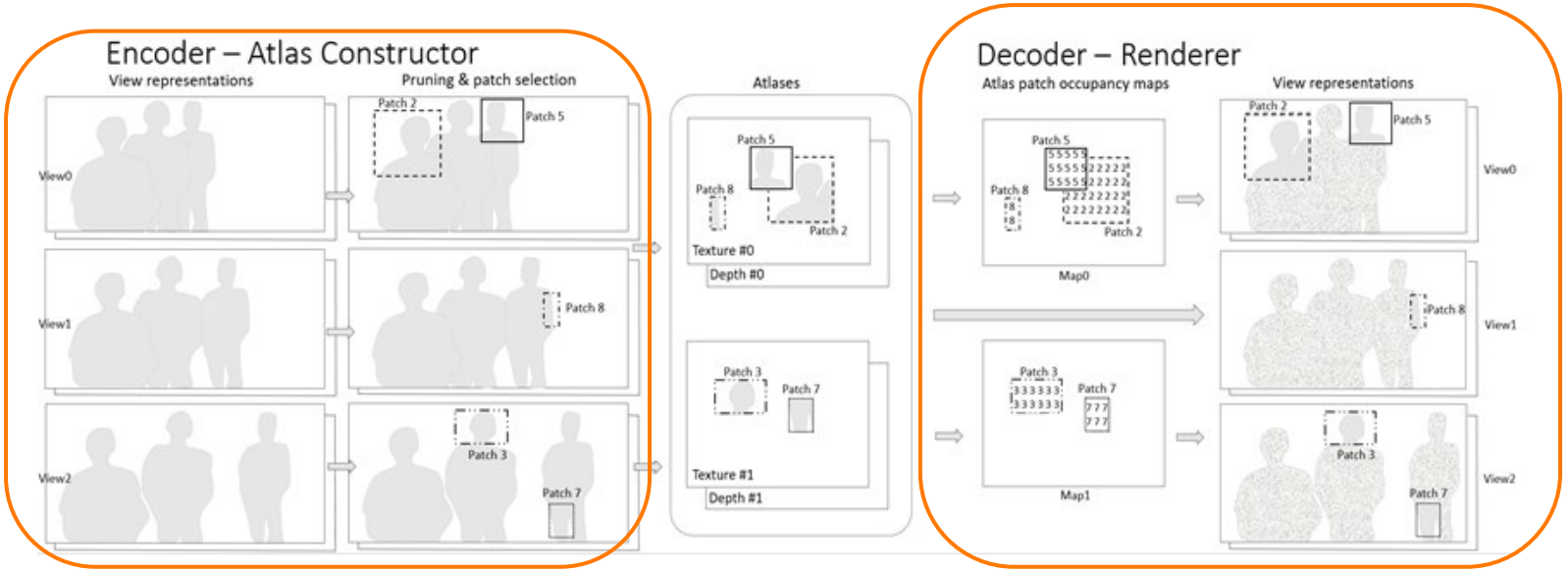
Decoder



TMIV Atlases

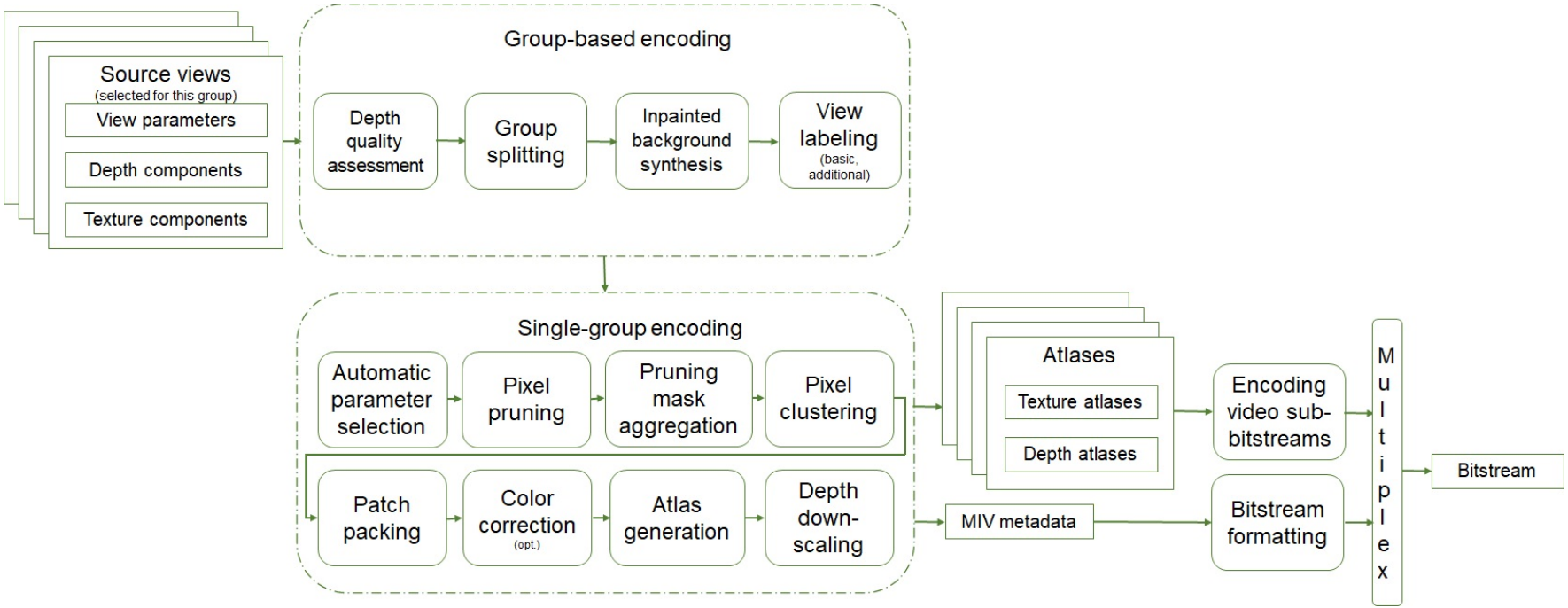


TMIV Key Components



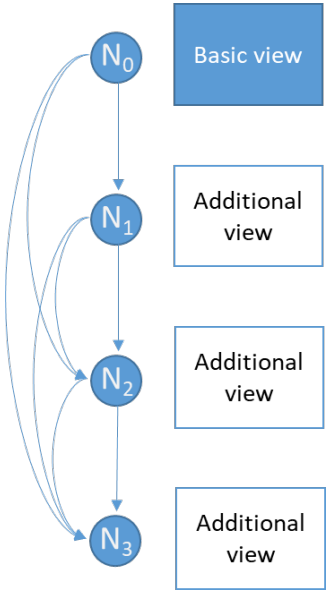
- Representing source views using patch atlases
- Pruned view reconstruction

TMIV Encoder (Detailed)



TMIV Pruning

- Remove the inter-view redundancy
- 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/luma should be less than a threshold
- Second-pass pruning: global color component differences
- Temporal consistency: the pruning masks are aggregated frame-by-frame (intra period reset)



Previous Work / Motivation

- Decoder-Side Depth Estimation for Immersive Video Coding [1]
  - Completely omit the transmission of depth maps (MV-HEVC+synthesis, full views)

- Decoder-Side Depth Estimation in MIV (Geometry Absent profile) [2]
  - Comparison of MIV profiles, Geometry Assistance SEI, Depth estimation benchmark (TMIV, full views)

- Patch Decoder-Side Depth Estimation in MIV [3]
  - Reduce the transmission of **patch depth data** (TMIV, partial views)

[1] P. Garus, J. Jung, T. Maugey, and C. Guillemot, “Bypassing Depth Maps Transmission For Immersive Video Coding,” in *2019 Picture Coding Symposium (PCS)*, Ningbo, China, Nov. 2019, pp. 1–5, IEEE.

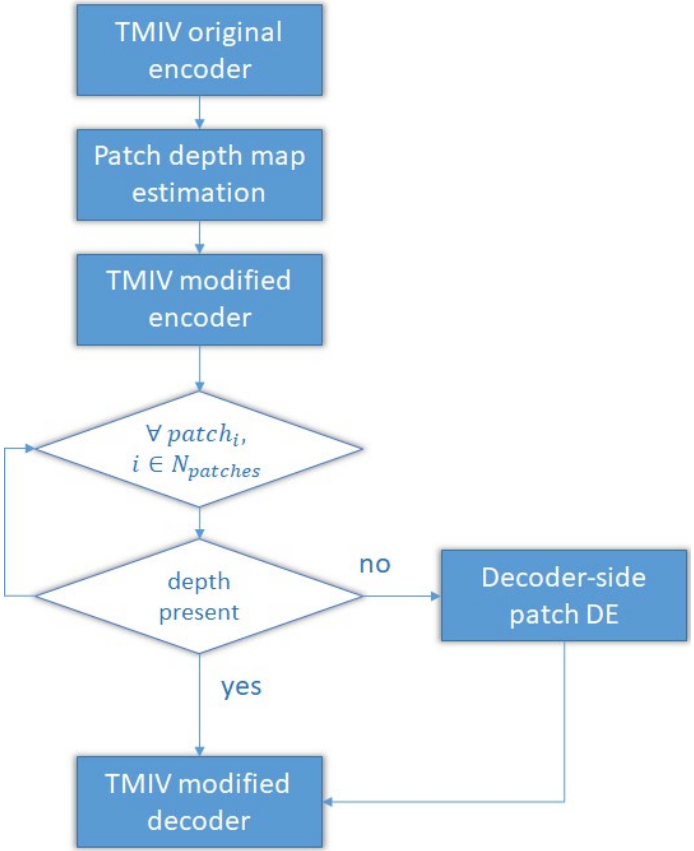
[2] D. Mieloch, P. Garus, **M. Milovanović**, J. Jung, J. Y. Jeong, S. L. Ravi, B. Salahieh, "Overview and Efficiency of Decoder-Side Depth Estimation in MPEG Immersive Video," in *IEEE Transactions on Circuits and Systems for Video Technology* (early access).

[3] **M. Milovanović**, F. Henry, M. Cagnazzo and J. Jung, "Patch Decoder-Side Depth Estimation In Mpeg Immersive Video," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1945-1949, IEEE.

Hybrid Method for Patch-Depth Transmission in MIV

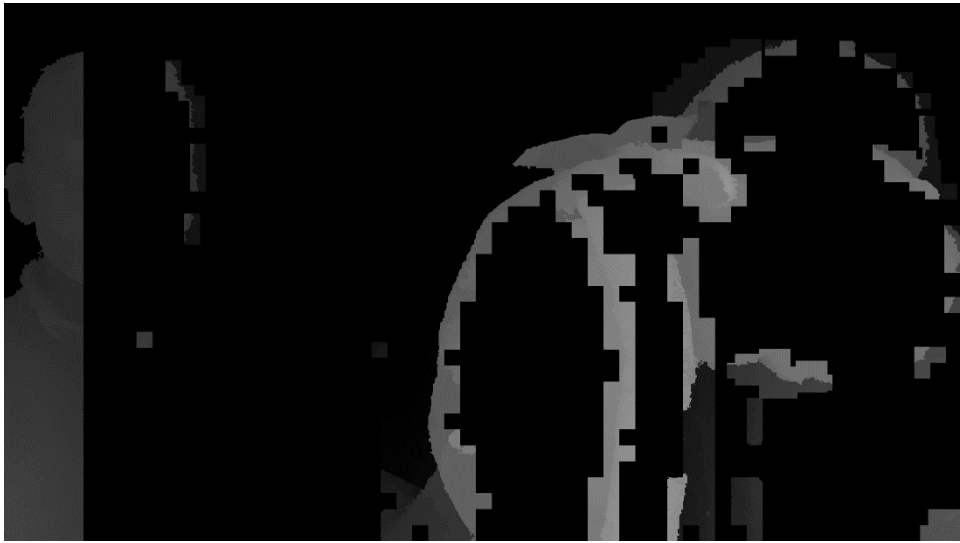
- Adapt the pruning strategy to ensure a reliable patch depth estimation at the decoder side
- Patch level decisions and patch depth selection based on the estimation quality

- “Anchor 1”: send everything
- “Anchor 2”: do not send the patch-depths at all
- Our “hybrid” proposal



Preliminary Synthesis Results

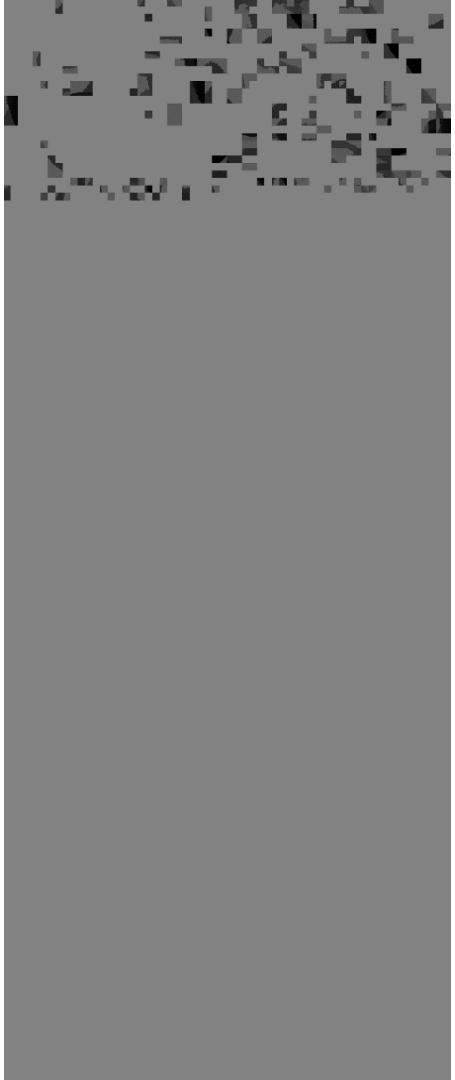
Pruned view depth map - anchor



Pruned view depth map - proposal

Preliminary Synthesis Results

Sequence	High BD-Rate Y-PSNR [%]	Low BD-Rate Y-PSNR [%]	High BD-Rate VMAF [%]	Low BD-Rate VMAF [%]
Frog	0.1	-1.4	-1.6	-2.3
Painter	-0.0	-0.0	-0.0	-0.0
Fencing	-2.8	-1.7	-1.3	-1.1
Street	-0.0	-0.4	-0.9	-0.9
Mirror	0.8	0.5	0.5	0.2



- We also observe small gains in terms of pixel rate saving (~0.2% – 1.0%)
- The amount of transmitted patch-depths vary per sequence
- Our method performs better on **low bitrate range** (same as for DSDE in general)

Conclusion

- **Summary:**
  - Smart selection of depth patches for transmission brings bitrate and pixel-rate savings while preserving the synthesized view quality

- **Possible improvements:**
  - Testing other selection criteria and finding the best one
  - Using a more convenient depth estimator
  - Enhancing the patch-depth atlas packing
  - New possibilities, e.g. sending more textures

Thank you for your attention!