



CREATING THE LIVING NETWORK™

JVET-G0024

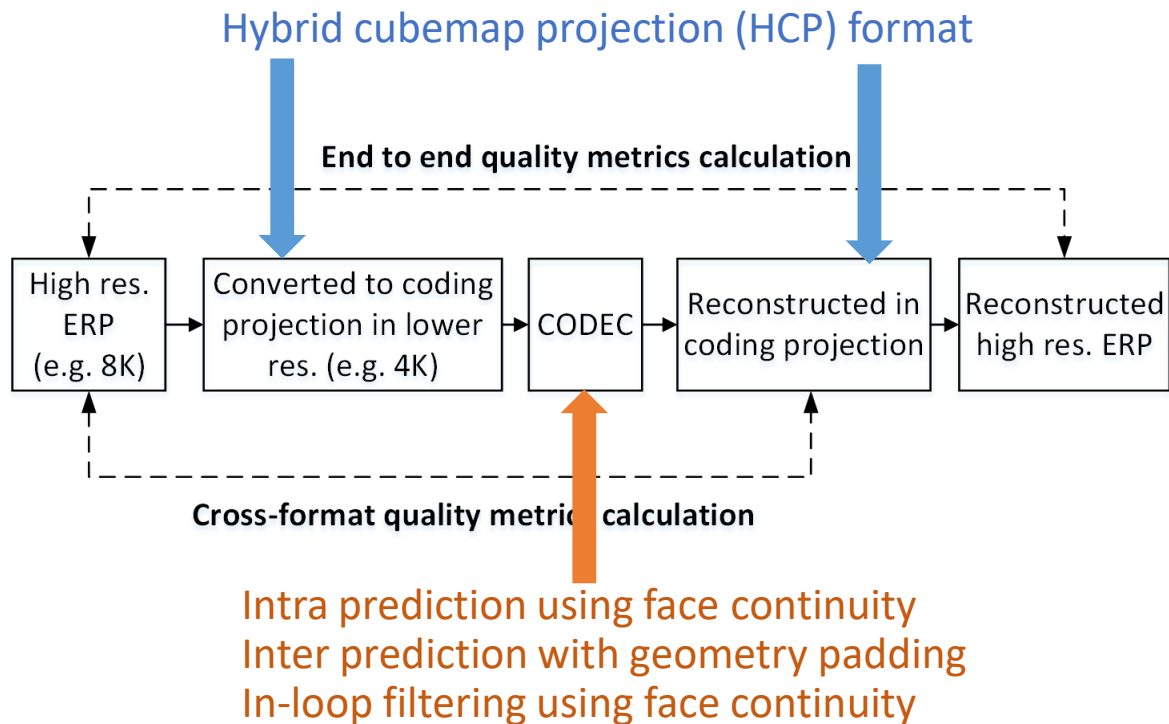
InterDigital's Response to the 360° Video
Category in Joint Call for Evidence on Video
Compression with Capability beyond HEVC

Philippe Hanhart, Xiaoyu Xiu, Fanyi Duanmu, Yuwen He, Yan Ye
(InterDigital)



Introduction

InterDigital's response includes the following coding technologies



Hybrid cubemap projection format (1)

- The HCP format is proposed to generalize CMP and other CMP-like projections using the mapping functions

- Map (x, y) in CMP to (x', y') in HCP

$$\begin{cases} x' = f_X(x) = \text{sgn}(x) \cdot (a \cdot x^2 + (1 - a) \cdot |x|) \\ y' = f_Y(y) = \text{sgn}(y) \cdot (b \cdot y^2 + (1 - b) \cdot |y|) \end{cases}$$

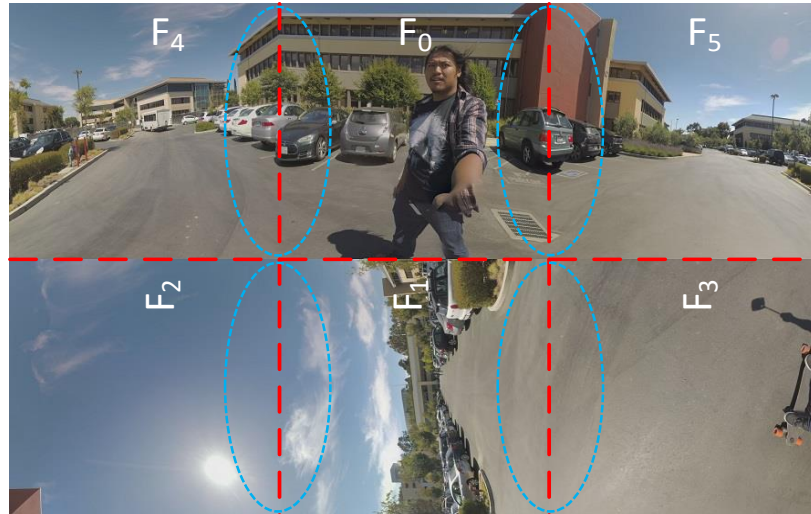
- **a** and **b** are the HCP horizontal and vertical mapping function parameters defined in the range of $(-1, 0]$;
 - If the parameter **a** or **b** is 0, HCP becomes the same as CMP in that corresponding direction

Hybrid cubemap projection format (2)

- 3x2 layout is used for HCP frame packing
- The mapping functions for the i-th face are defined by two parameters (a_i , b_i) for horizontal and vertical directions
- Constraint is applied to keep the vertical continuity of neighboring faces in the same face row

$$b_4 = b_0 = b_5$$

$$b_2 = b_1 = b_3$$



Hybrid cubemap projection format (3)

- The HCP parameter derivation process is done row-by-row:
 - deriving parameters for the first face row (F_4, F_0, F_5)
 - deriving parameters for the second face row (F_2, F_1, F_3)
- Iterative parameter searching algorithm for horizontal and vertical mapping functions is applied
 - Usually converges in 3-4 steps
- End-to-end conversion-only weighted error is minimized in the search process for the first picture in an IRAP

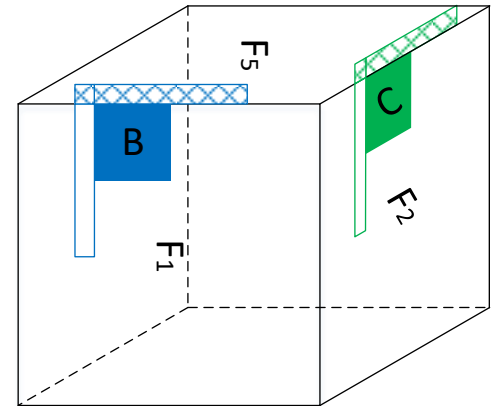
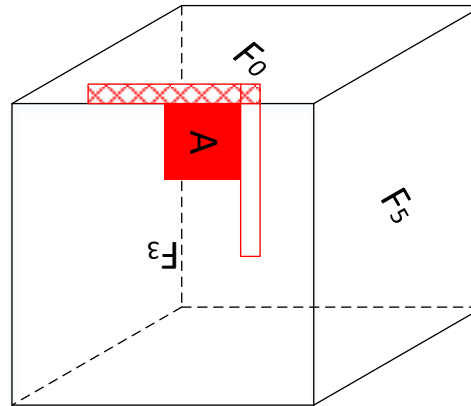
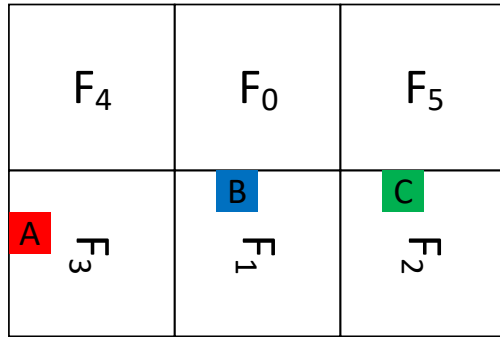
$$Dist = \sum_{F_i \in \text{face_row}} SSE(F_i(ERP), F_i(ERP'))$$

the portion of source ERP
corresponding to face F_i in HCP

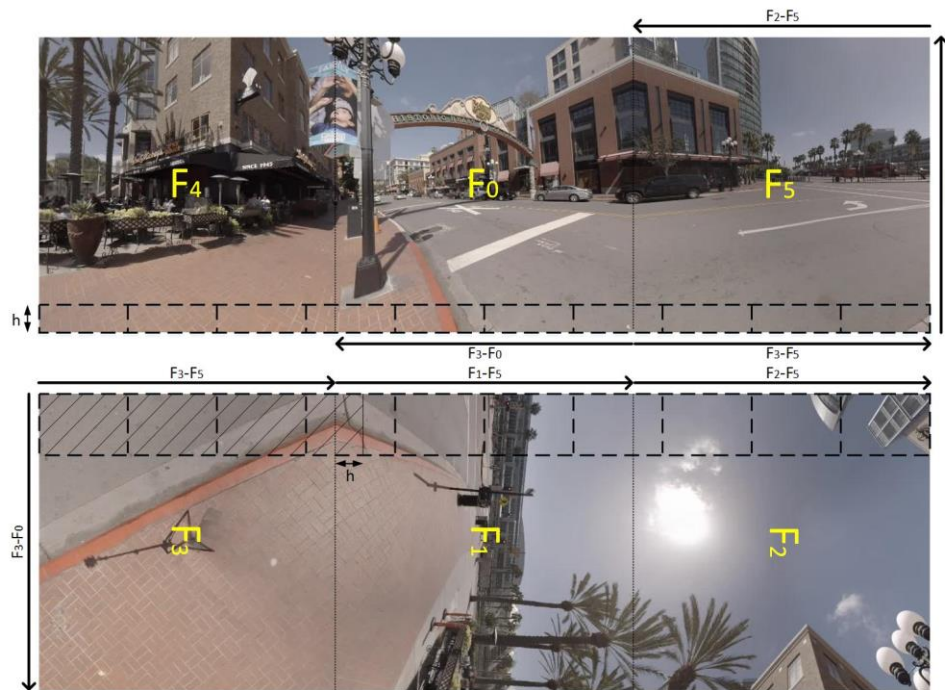
the portion of ERP after conversion
corresponding to face F_i in HCP

Intra prediction using face continuity (1)

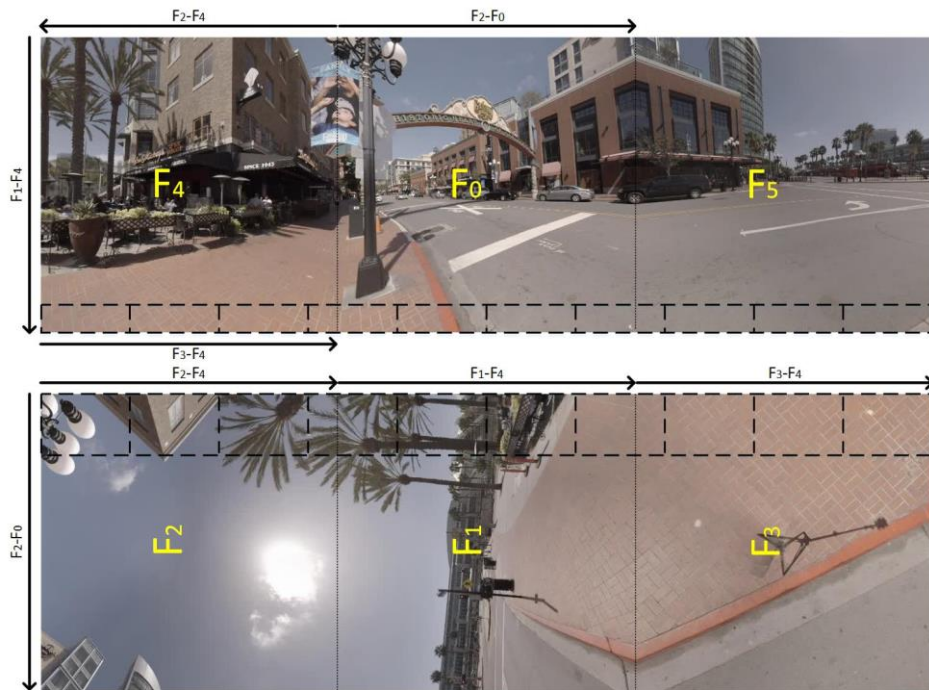
- Two kinds of spatial neighbors
 - Frame packed neighbor: neighbor in frame packed picture (this is the conventional neighbor)
 - Spherical neighbor: neighbor in 3D geometry
- Spherical neighbor may be different from frame packed neighbor
- For intra prediction, if reference sample is located outside of a face, then spherical neighbor is used



Intra prediction using face continuity (2)



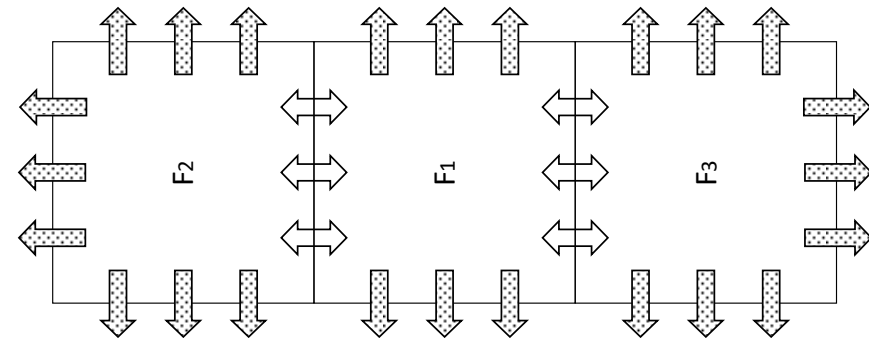
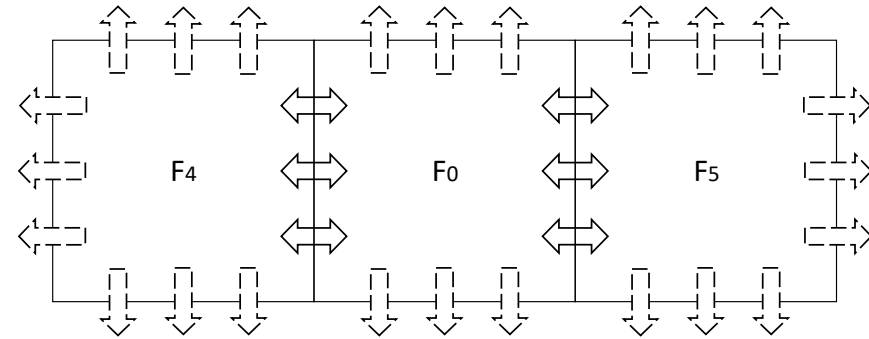
Default 3x2 packing layout in 360Lib



proposed 3x2 packing layout

Intra prediction using face continuity (3)

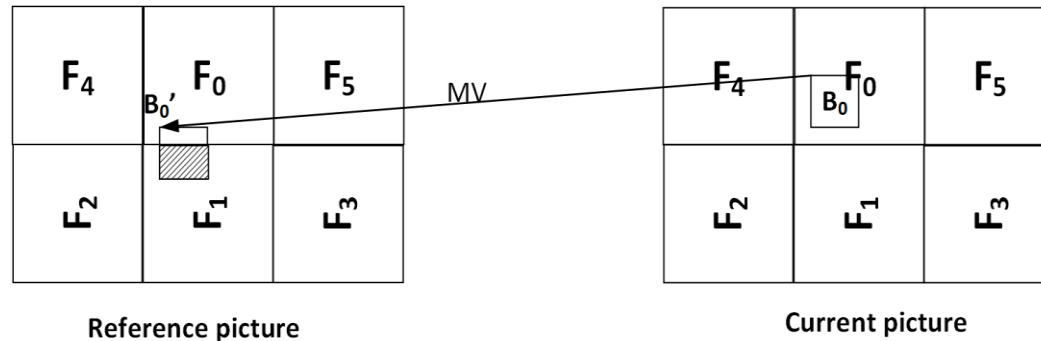
- Spherical neighbor derivation
 - Nearest neighbor in 2D plane is used to avoid interpolation
- Same approach is used in cross-component linear model prediction



- ←--- Not available
- ← JEM default reference sample derivation method
- ←···· Reference sample derivation based on spherical neighbors

Inter prediction with geometry padding (1)

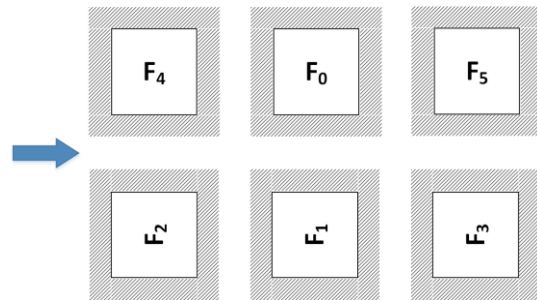
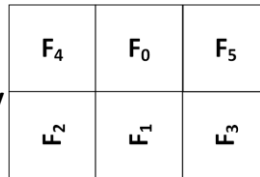
- Geometry padding is applied in the motion compensated prediction when the reference sample is outside of the face that current block belongs to
 - Those samples in the shaded region are not directly derived from samples in face F_1 (back face), but derived from face F_2 (bottom face) with geometry padding



Inter prediction with geometry padding (2)

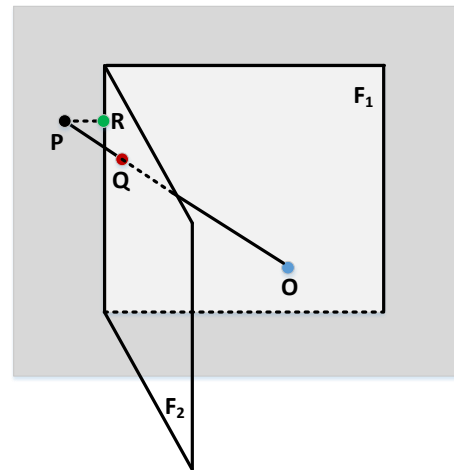
- Geometry padding process for HCP

- The reference picture of HCP is frame packed
- Each face of the HCP is padded with geometry padding



- Derivation of padded samples based on 3D geometry

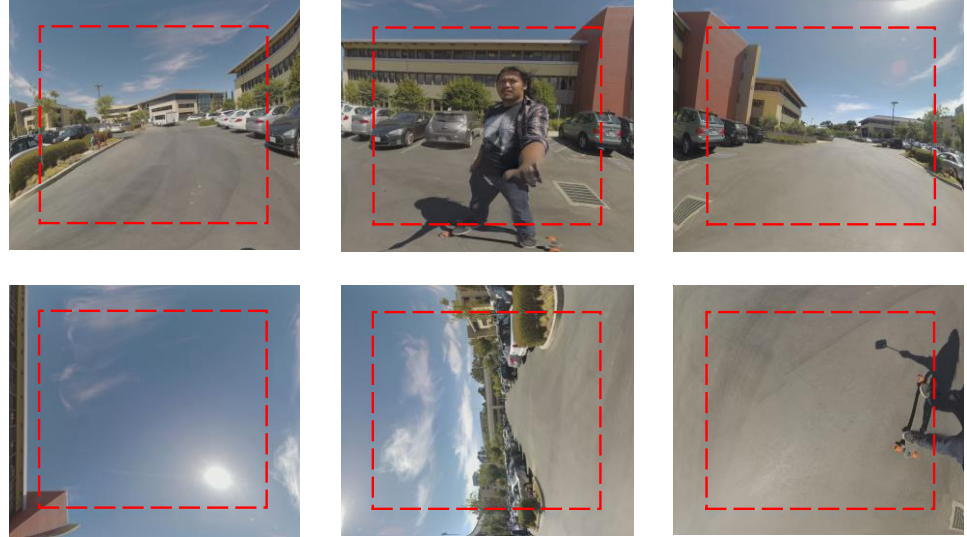
- Point P is outside of face F₁ and needs to be padded
- Point Q is the projected point on face F₂ of point P from the center of the sphere point O
- The sample value at point P will be filled with the sample value at point Q on face F₂



Inter prediction with geometry padding (3)



Extended picture of HCP 3x2 with repetitive padding



Extended face pictures of HCP 3x2 with geometry padding

In-loop filtering using face continuity

- In 2D video coding, top, right, bottom, and left picture boundaries may not be filtered during in-loop filtering process
- In 360° video coding, top, right, bottom, and left boundaries of each face are connected to another face boundary and thus in-loop filtering could be applied across all face boundaries, including across picture boundaries
- In-loop filters applied by processing one face at a time
 - Using spherical neighbors instead of frame packed neighbors
 - Nearest neighbor in the 2D plane is used to avoid interpolation

Simulation setting

- The response software is built on JEM-6.0 and 360Lib-3.0
- HCP face set to 1184×1184, the same CMP and ACP in 360° video CTC
- The following information is signalled in the bitstream
 - SPS: projection format, # faces in horizontal and vertical directions, position and orientation of faces
 - PPS: the horizontal and vertical HCP parameters

Comparison with the HM-based 360° anchors

Sequence	E2E WS-PSNR Y	E2E WS-PSNR U	E2E WS-PSNR V
SkateboardInLot	-39.64%	-63.18%	-70.08%
Chairlift	-45.95%	-66.27%	-60.73%
KiteFlite	-23.02%	-56.60%	-63.16%
Harbor	-25.92%	-56.11%	-56.50%
Trolley	-21.56%	-44.27%	-50.52%
Average	-31.22%	-57.29%	-60.20%

Comparison with the JEM-based 360° anchors

Sequence	E2E WS-PSNR Y	E2E WS-PSNR U	E2E WS-PSNR V
SkateboardInLot	-15.38%	-25.27%	-24.29%
Chairlift	-22.06%	-17.77%	-17.60%
KiteFlite	-6.35%	-9.09%	-14.69%
Harbor	-7.41%	-12.71%	-12.91%
Trolley	-6.41%	-6.88%	-11.25%
Average	-11.52%	-14.35%	-16.15%

HCP's conversion-only performance

Sequence	E2E WS-PSNR Y	E2E WS-PSNR U	E2E WS-PSNR V
SkateboardInLot	50.14	61.48	61.10
Chairlift	50.06	59.55	59.61
KiteFlite	46.44	57.68	57.68
Harbor	49.22	59.14	58.09
Trolley	45.39	57.95	57.83
Average	48.25	59.16	58.86
Gain over ERP	3.01	0.91	0.87



Original



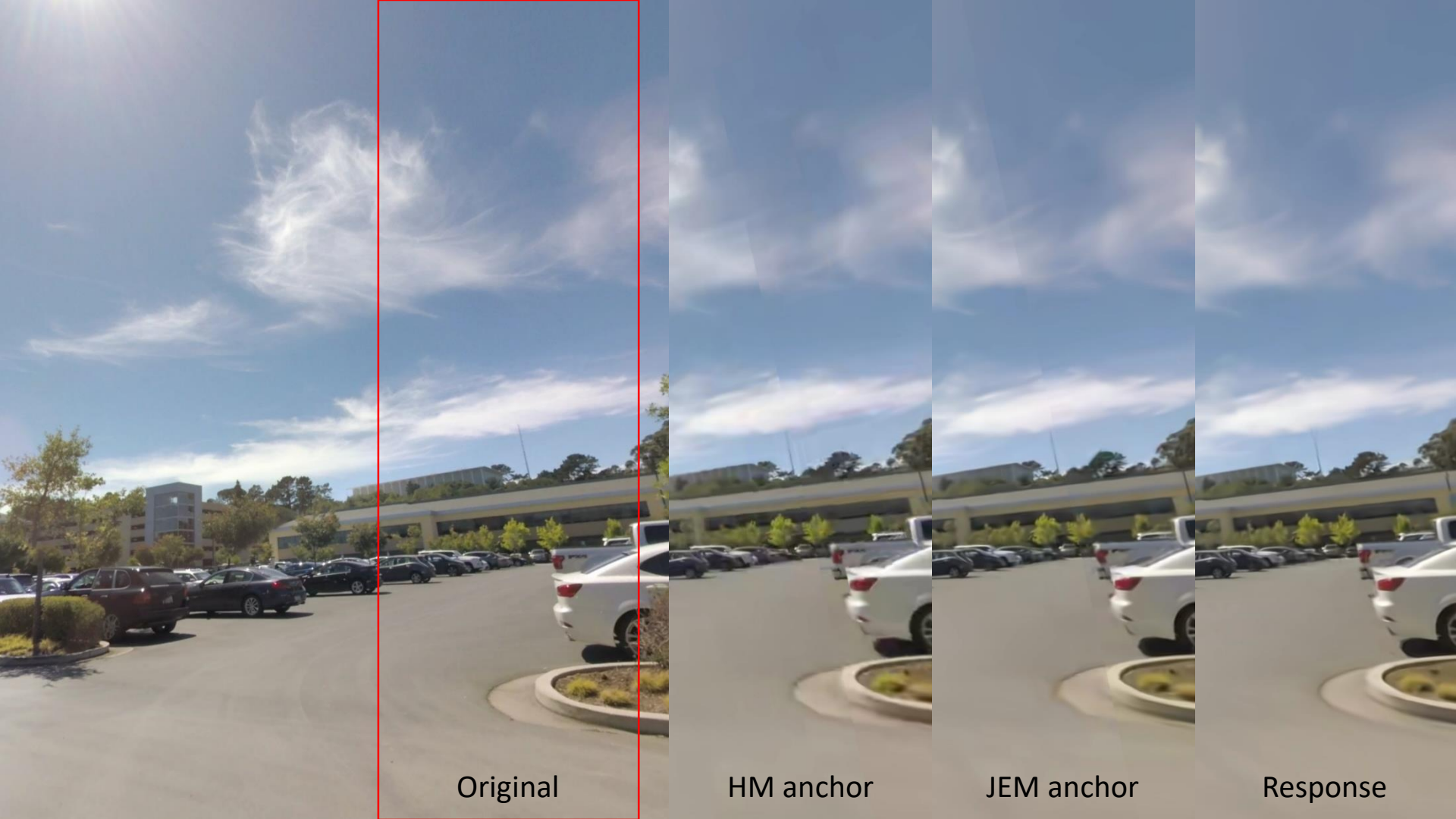
HM anchor



JEM anchor



Response



Original

HM anchor

JEM anchor

Response

Conclusion

- This response shows that it is feasible to significantly improve the compression performance of HEVC for 360° video
 - 31.22% average bitrate savings, with up to 45.95% bitrate reduction for *Chairlift*
- The response can effectively improve the coding performance of the current JEM for 360° video
 - Average BD-rate savings of 11.52%, maximum bit rate savings of 22.06% for *Chairlift*
- General improvement in subjective quality of the reconstructed video was observed
 - Less blocking artifacts, better texture and edges
- 360° video specific video quality problems were also reportedly fixed
 - In particular, no face seams were observed in the reconstructed video sequences encoded with this response

THANK YOU!

