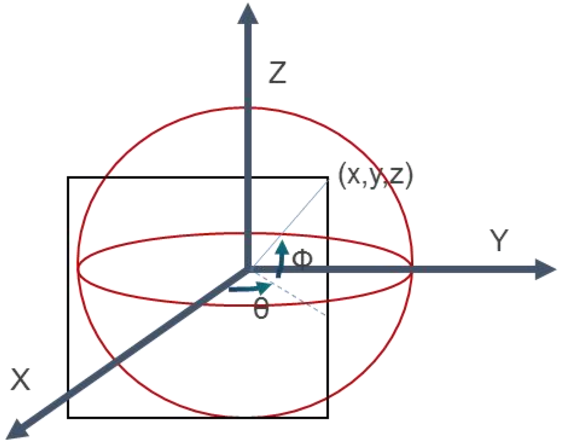
AHG8: TSP Evaluation With Viewport-Aware Quality Metric For 360 Video (JVET-E0070)

Geert Van der Auwera, Muhammed Coban, Hendry, Marta Karczewicz

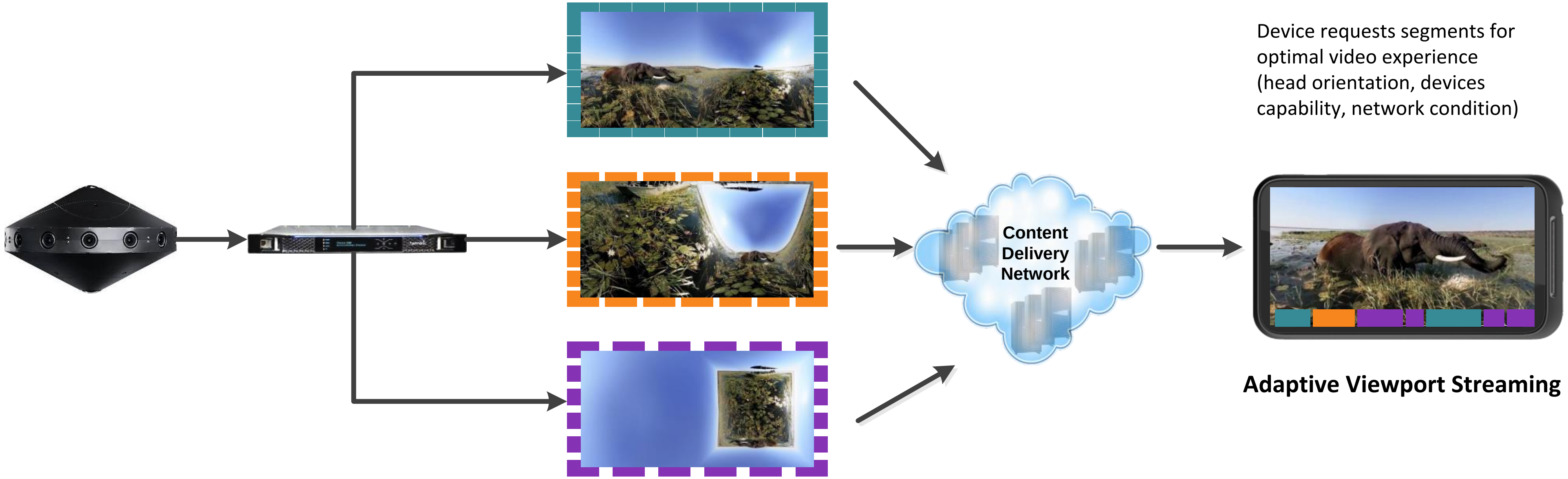
Qualcomm

Introduction



- Virtual reality and 360 degree video content require high quality to provide an immersive experience to the viewer
- HVS can distinguish up to 60 pixels per degree of field of view (FOV)
- Field-of-view (FOV) $\geq 90^0$
- “VR-Retina” requirement: 21K (21600 x 10800) 360^0 video
- Frame rate: 60fps and higher
- High bandwidth and decoding requirement

Adaptive viewport streaming

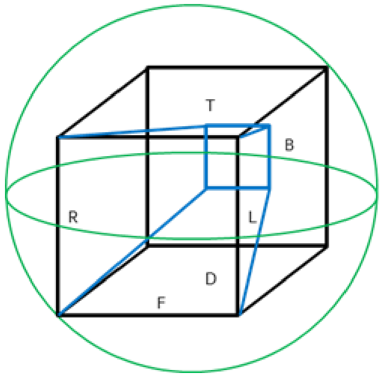


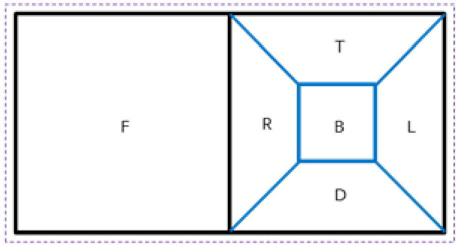
- Time aligned IRAP pictures (e.g., 1 sec. period)
- During switching, the viewer will continue viewing the previous viewport stream until the arrival of the current (last requested track) viewport
- Viewport switch duration: IRAP period and playback buffering duration

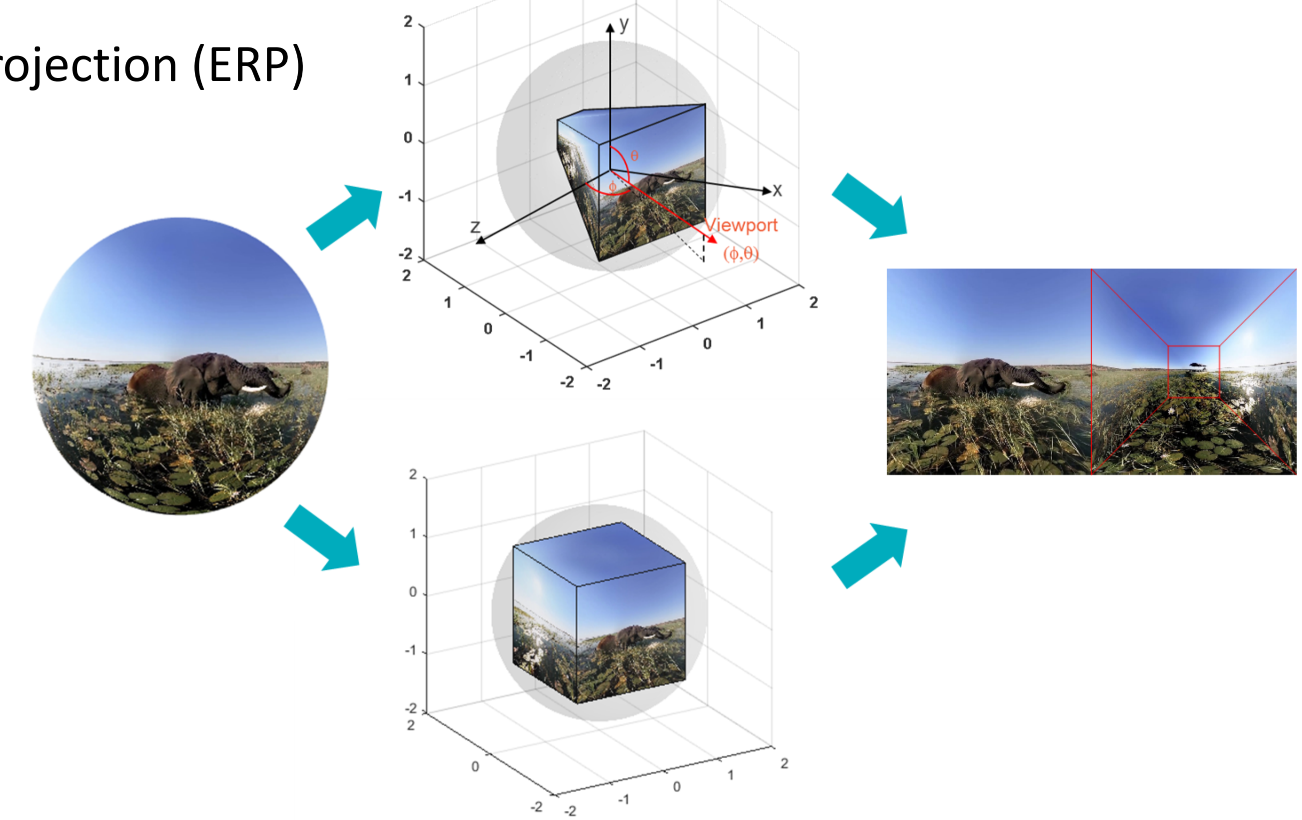
Truncated Square Pyramid Projection

- Viewport-based schemes
  - FOV coded at high resolution/quality
  - Rest of the 360 view represented with lower resolution/quality
  - Lower bandwidth, reduced pixel count
- Truncated square pyramid
  - Pixel count is 25% of Equirectangular Projection (ERP)
  - Preserves continuity across face edges

Video Frame Resolution (Decode)	
ERP	**TSP**
3840x1920 (4K)	1920×960
5760x2880 (6K)	2880×1440
7680x3840 (8K)	3840×1920

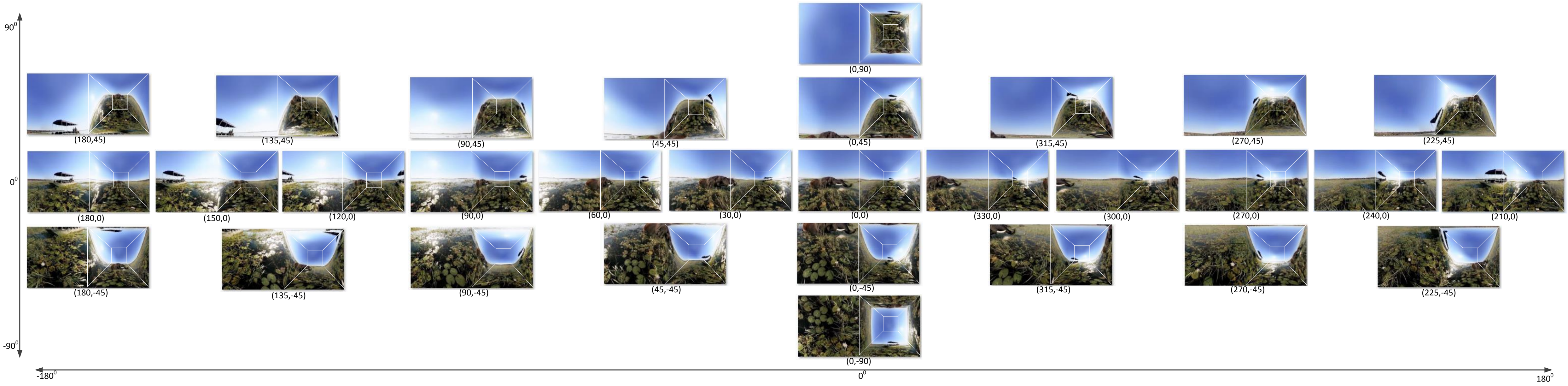






Viewport Streaming

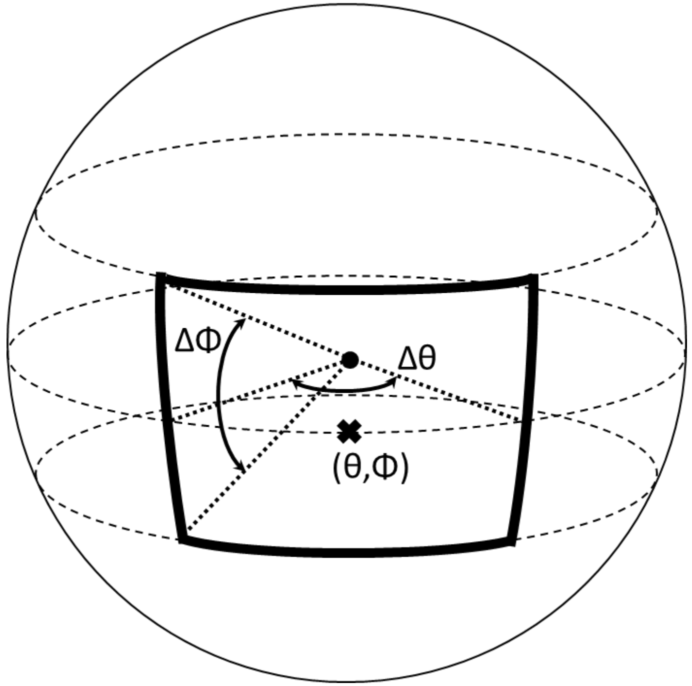
- Viewport aware 360 video streaming schemes target delivery of high resolution (>4K) 360-degree video
- Entire sphere is covered by multiple overlapping viewports



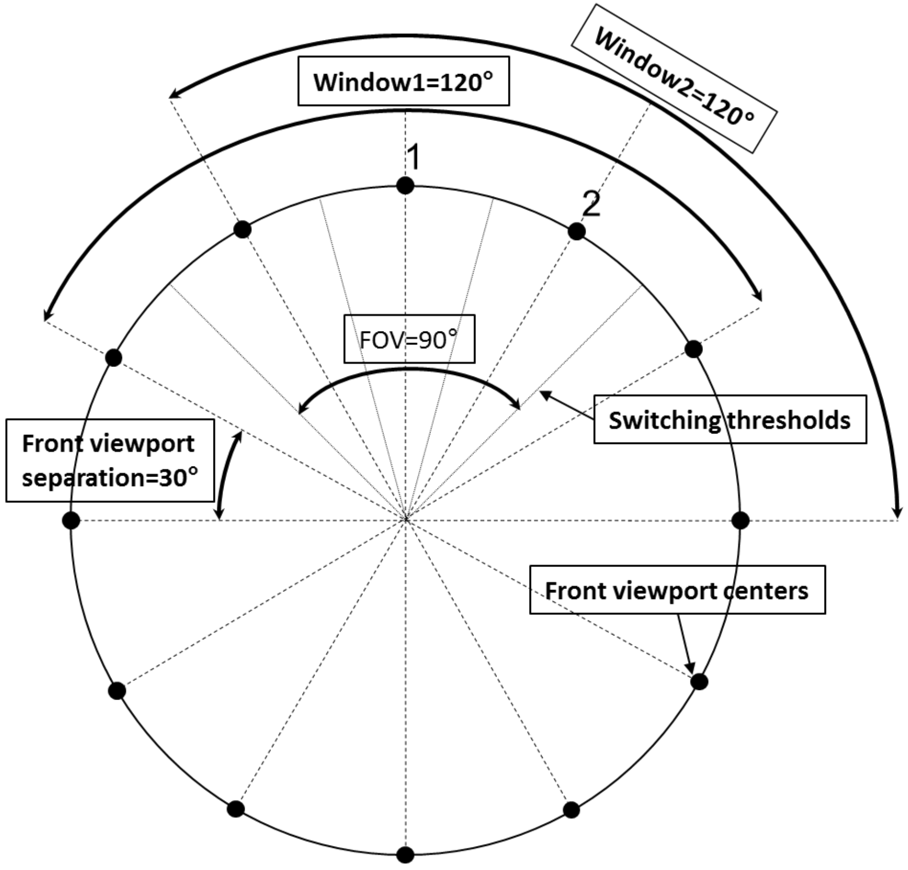
30 viewports

Viewport-Aware Quality Metric

- Windowed S-PSNR
  - Window with size (Δθ, ΔΦ) is centered on front viewport of the representations.
    - [Φ – ΔΦ/2, Φ + ΔΦ/2] and [θ – Δθ/2, θ + Δθ/2]
  - Window sizes can range from 0° to 180° for pitch Φ and 0° to 360° for yaw θ.



Viewport-Aware Quality Metric



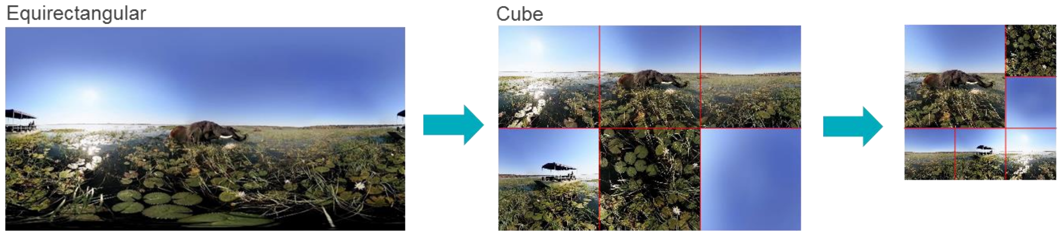
- Zero switching delay (ideal case)
  - Window size = window_zero_delay = FOV + representation_separation
  - First proposal:
    - Compute windowed S-PSNR (window_zero_delay) for each representation and average all bit rates and S-PSNRs
- Nonzero switching delay (real-world)
  - Switching delay (~1s) depends on I period, frame rate, buffering
  - Window size = window_delay(head_rotational_speed) =

  FOV + representation_separation + head_rotational_speed x switching_delay x 2
  - Second proposal:
    - Compute windowed S-PSNR for multiple window sizes, each size corresponding with a head rotational speed
    - Per window size, compute average of all bit rates and S-PSNRs
    - Probability may be assigned to each window size and weighted average computed

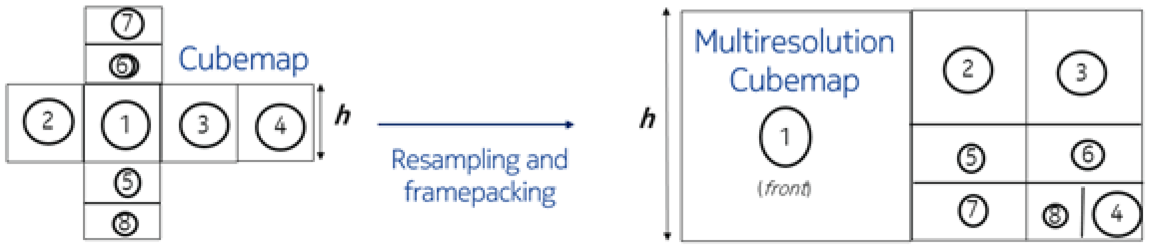
Experimental Results

- Compare viewport based switching schemes against ERP:
  - TSP
  - Downsampled cubemap (DCP) and alternative DCP-2 (JVET-D0078)

DCP



DCP-2



- Test conditions:
  - Viewport-aware metric
  - HM 16.11, 360tools
  - 10-bit conversion
  - Random access configuration, 1 sec duration
  - 7 GoPro and 3 Nokia sequences
  - 30 viewports

ERP	TSP	DCP	DCP-2
3840×1920	1920×960	1440×1440	1920×960
4096×2048	2048×1024	1536×1536	2048×1024
4320×2160	2160×1080	1632×1632	2160×1080

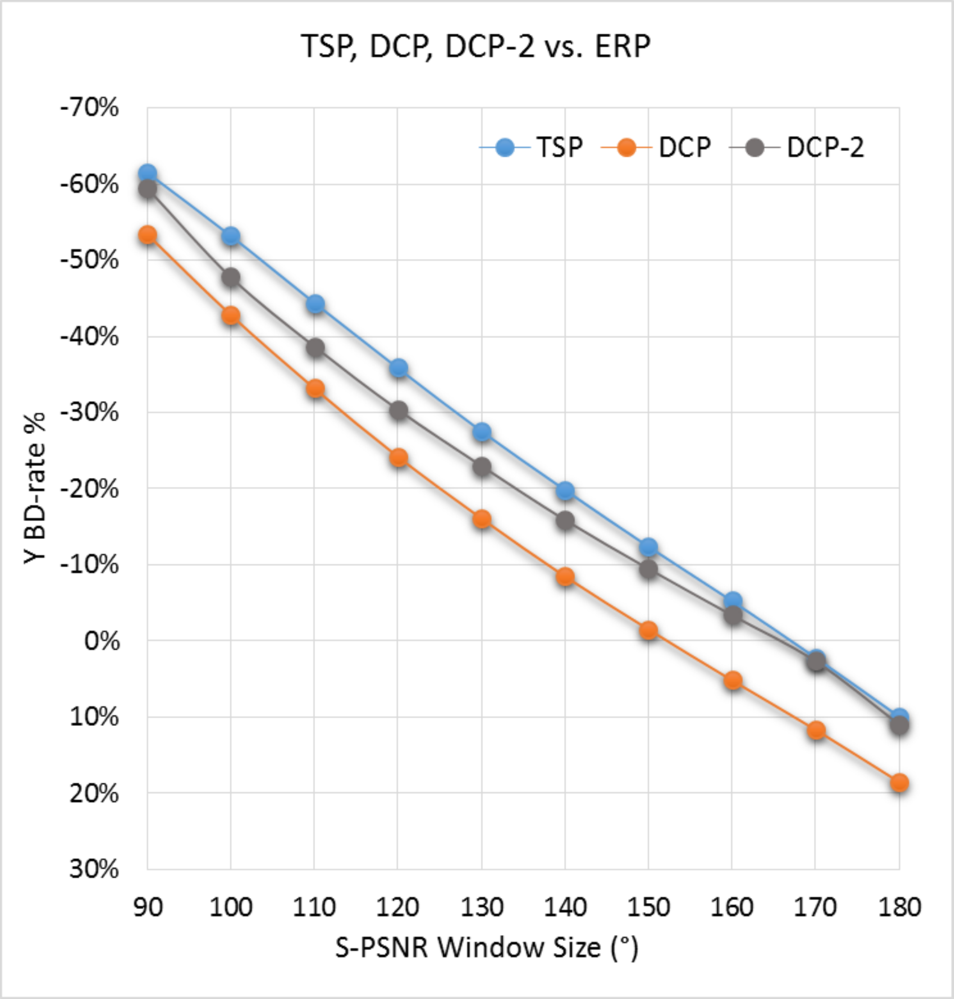
DCP 12.5% more pixels than TSP
DCP-2 same no pixels as TSP

Experimental Results

BD-rates for TSP, DCP, DCP-2 comparison with ERP (negative BD-rate is gain)

ERP (Anchor)	TSP BD-rate (Test)			DCP BD-rate (Test)			DCP-2 BD-rate (Test)		
S-PSNR Win (°)	Y	U	V	Y	U	V	Y	U	V
90	-61.54%	-64.67%	-65.40%	-53.47%	-58.63%	-60.19%	-59.53%	-59.53%	-59.53%
100	-53.25%	-58.73%	-59.68%	-42.88%	-50.30%	-52.12%	-47.85%	-47.85%	-47.85%
110	-44.45%	-52.88%	-54.10%	-33.22%	-42.76%	-44.91%	-38.74%	-38.74%	-38.74%
120	-35.89%	-46.95%	-48.45%	-24.25%	-35.52%	-38.21%	-30.45%	-30.45%	-30.45%
130	-27.59%	-40.81%	-42.94%	-16.14%	-28.40%	-31.82%	-23.06%	-23.06%	-23.06%
140	-19.89%	-34.70%	-37.16%	-8.53%	-21.60%	-25.48%	-15.96%	-15.96%	-15.96%
150	-12.44%	-28.73%	-31.17%	-1.57%	-15.23%	-19.15%	-9.56%	-9.56%	-9.56%
160	-5.30%	-22.78%	-24.81%	5.11%	-8.97%	-12.47%	-3.46%	-3.46%	-3.46%
170	2.14%	-16.18%	-17.89%	11.61%	-2.35%	-5.69%	2.58%	2.58%	2.58%
180	9.89%	-9.40%	-10.93%	18.47%	4.61%	1.27%	10.87%	10.87%	10.87%

BD-rate comparison between ERP (anchor) and TSP, DCP, DCP-2



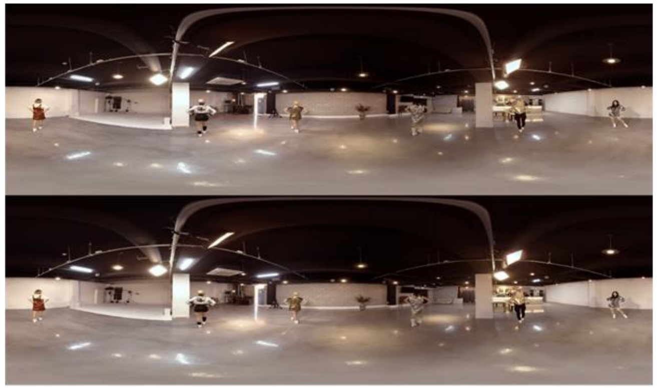
Thanks to Mediatek for cross-check (JVET-E0131)

Conclusions

- Viewport-aware quality metric was proposed based on windowed S-PSNR
- TSP is up to ~12% better in BD rate gain compared with ERP than for DCP according to the proposed evaluation scheme, and up to ~6% better than DCP-2
- TSP based viewport scheme provides lower bandwidth and reduced pixel count
- TSP demo available
    - Qualcomm Snapdragon VR820 kit
    - "Effective 6K" per eye stereo 60 fps
    - 4K 60 fps HEVC decoder

Backup Slides

Stereo TSP



3840x2160 Stereo Equirectangular Top-Bottom

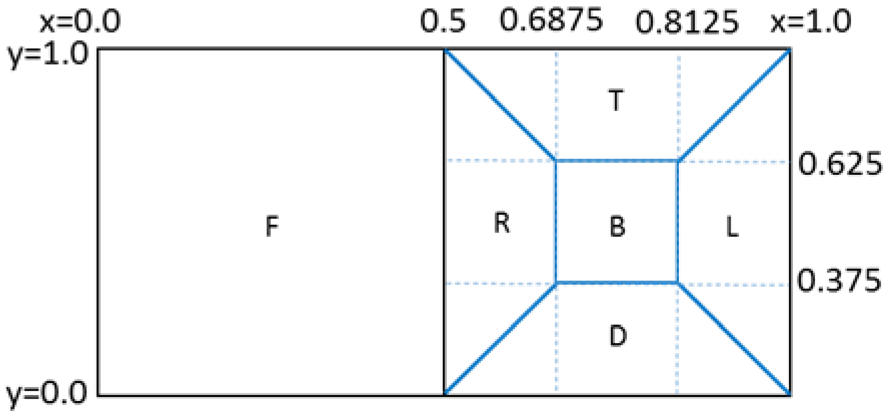
Half of 4K resolution

2880x2880 Stereo TSP Top-Bottom

"Effective" 6K resolution

“360tools”, “360Lib” software TSP Support

- Implementation



Forward equations (cube faces to TSP)	**Inverse equations (TSP to cube faces)**
Right TSP trapezoid from right cube face: x’ = (x – 0.5) / 0.1875 y’ = (y – 2.0x + 1.0) / (3.0 - 4.0x)	Right cube face from right TSP trapezoid: x = 0.1875x’ + 0.5 y = 0.375x’ – 0.75x’y’ + y’
Left TSP trapezoid from left cube face: x' = (x – 0.8125) / 0.1875 y' = (y + 2.0x – 2.0) / (4.0x – 3.0)	Left cube face from left TSP trapezoid: x = 0.1875x’ + 0.8125 y = 0.25y’ + 0.75x’y’ – 0.375x’ + 0.375
Bottom TSP trapezoid from bottom cube face: x' = (1.0 – x – 0.5y) / (0.5 – y) y’ = (0.375 – y) / 0.375	Bottom cube face from bottom TSP trapezoid: x = 0.1875y’ – 0.375x’y’ – 0.125x’ + 0.8125 y = 0.375 – 0.375y’
Top TSP trapezoid from top cube face: x' = (0.5 – x + 0.5y) / (y – 0.5) y’ = (1.0 – y) / 0.375	Top cube face from top TSP trapezoid: x = 1.0 – 0.1875y’ – 0.5x’ + 0.375x’y’ y = 1.0 – 0.375y’
Back TSP face from back cube face: x' = (x – 0.6875) / 0.125 y’ = (y – 0.375) / 0.25	Back cube face from back TSP face: x = 0.125x’ + 0.6875 y = 0.25y + 0.375