6K Effective Resolution with 4K HEVC Decoding Capability for OMAF-compliant 360° Video Streaming

ACM Multimedia Systems Conference
23rd Packet Video Workshop
Tuesday, July 12 2018

Alireza Zare, Alireza Aminlou, Miska M. Hannuksela
Nokia Technologies, Tampere, Finalnd

NOKIA

Problem statement

Head-mounted displays (HMDs) FOV is getting larger.

Hardware decoders have restrictions on frame resolutions, which is up to 4K

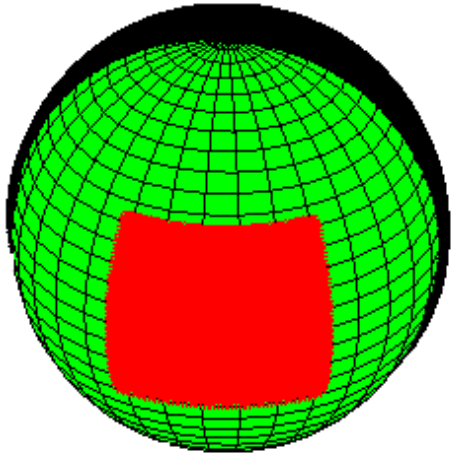
Omnidirectional MediA Format (OMAF) has 4K-decoding constraint.

Streaming of 4K ERP content allows only a limited viewport resolution.

4K viewport requires video content at 12K

The solution

In order to take the full advantage of HMDs with wide FOV, a specific mixed-resolution packing of 6K (6144x3072) ERP content is proposed, while sticking with the 4K-decoding constraint.



A limited part of content is displayed on HMDs at a time

The propose packing guarantees that content for the viewport originates from ERP panorama of 6K

Omnidirectional MediA Format (OMAF)

A standard that specifies delivery of 360° content

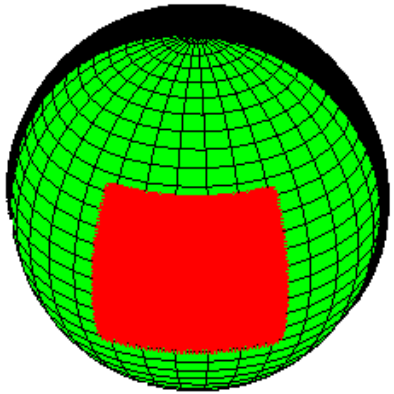
The Current OMAF supports only equirectangular (ERP) and cubemap projections and their region-wise packing.

OMAF 4K-decoding constraint:

HEVC-based viewport-dependent OMAF video profile supports picture size of 8 912 896 luma samples, corresponding e.g. to 4096x2176

Viewport-adaptive streaming (VAS)

- Ideally, only the current viewport content could be transmitted.
- Because of delay in the VR system: a full representation of the 360° content is transmitted.
- VAS: higher resolution or fidelity content for viewport and a lower fidelity or resolution content for the remaining non-viewport area
  - mixed-quality VAS
  - mixed-resolution VAS
- Two practical methods for VAS:
  - Viewport-dependent projection
  - Tile-based streaming

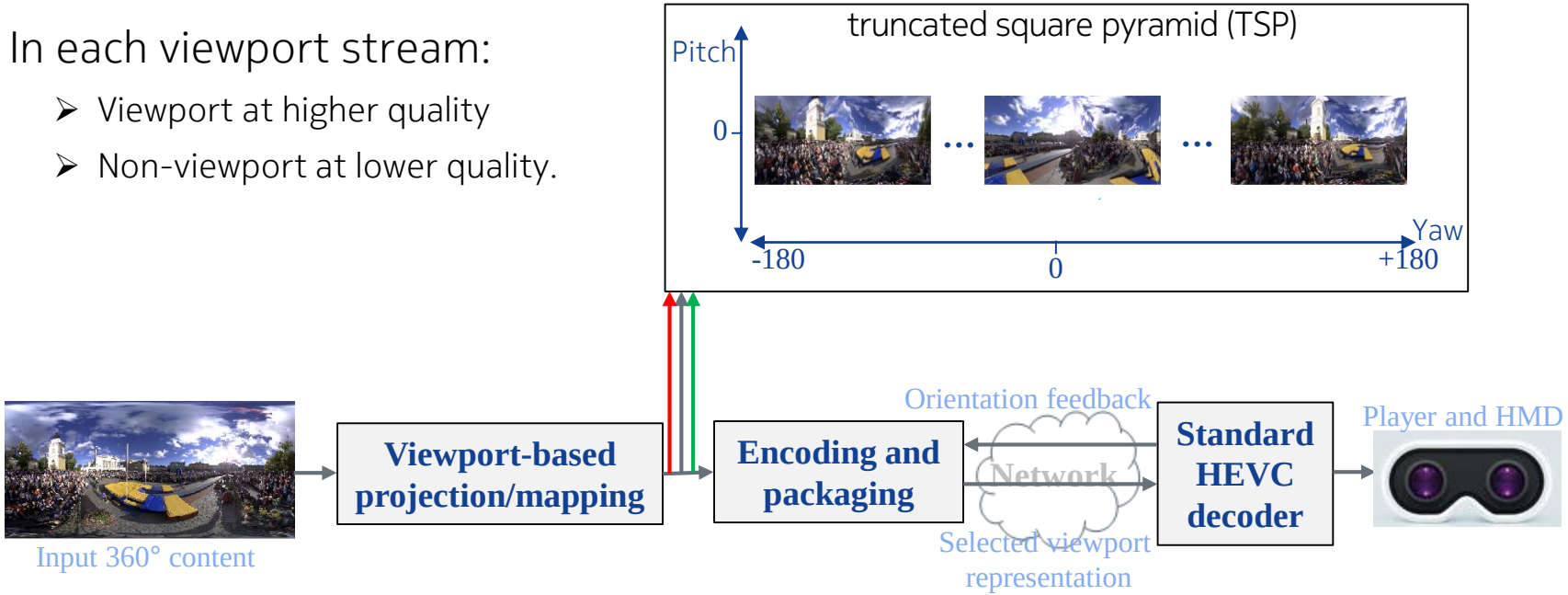


NOKIA

Viewport-dependent projection

In each viewport stream:

- Viewport at higher quality
- Non-viewport at lower quality.

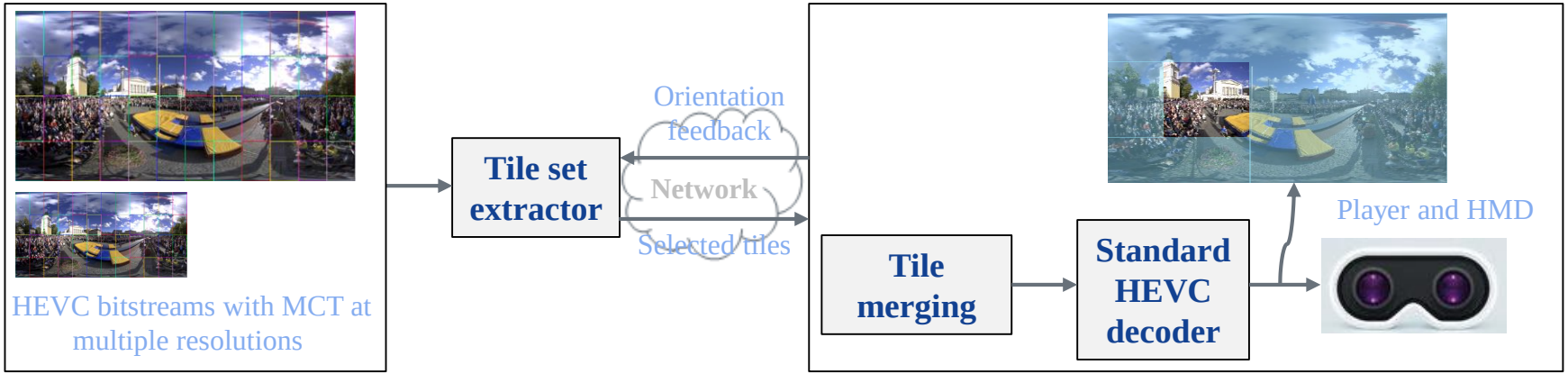


VR viewport-dependent unicast streaming system

NOKIA

Tile-based streaming

- Dividing 360° video to several tiles coded independently in varying quality
- Combining tiles with varying quality to generate a viewport representation



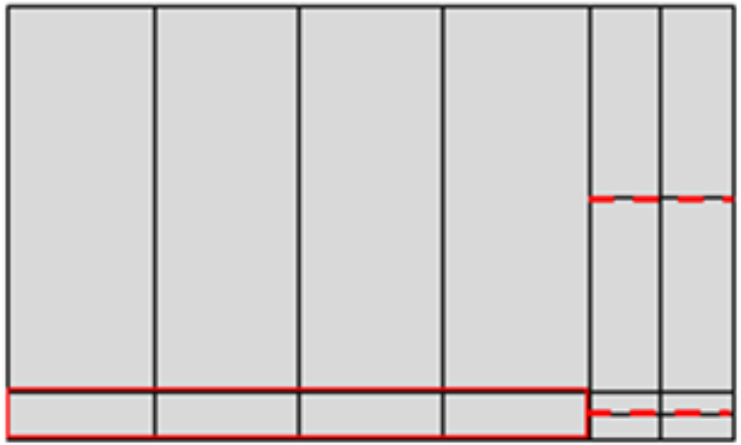
VR tile-based unicast streaming system

NOKIA

Proposed OMAF-complaint 6K effective resolution packing

- To benefit high-resolution capability of HMDs available in the market
- Meeting the OMAF 4K-decoding constraint
- The proposed packing is part of the OMAF Annex D

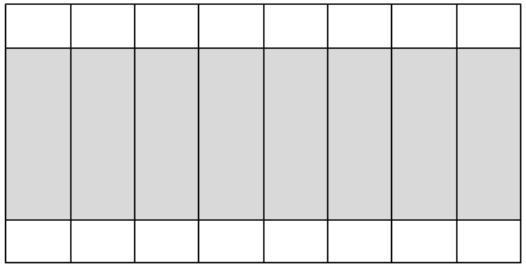
The stream layout for achieving effective 6K ERP resolution, picture size 3840x2304



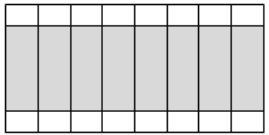
NOKIA

Proposed OMAF-complaint 6K effective resolution packing

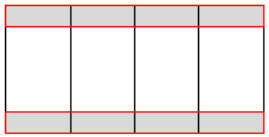
6K
6144x3072
cropping to 6144x2048
(elevation range 120°)
8x1 tile grid
tile size 768x2048



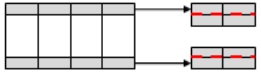
3K
3072x1536
cropping to 3072x1024
8x1 tile grid
tile size 384x1024

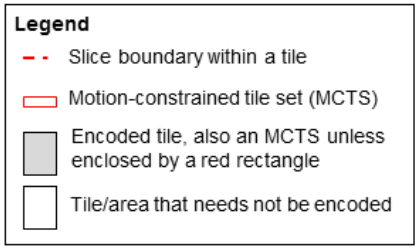


3K polar
3072x1536
coding the top and
bottom stripes
(elevation range 30°)
as MCTSs

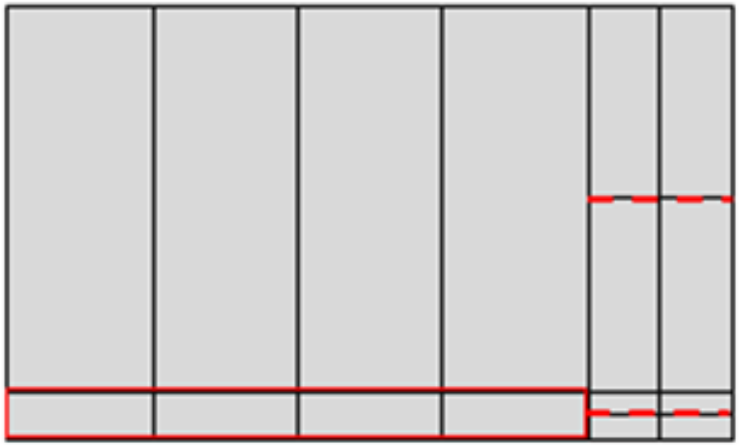


1.5K polar
1536x768
arrange top and bottom
to separate pictures of
2x1 tile grid, 2 slices / tile





- 4 different versions of the content are encoded (different resolutions and tilings)
- 180°×120° coverage of high-resolution content

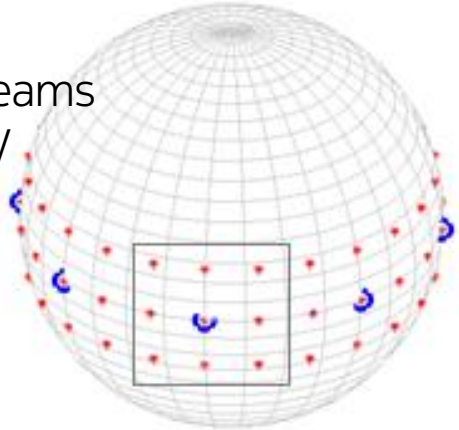


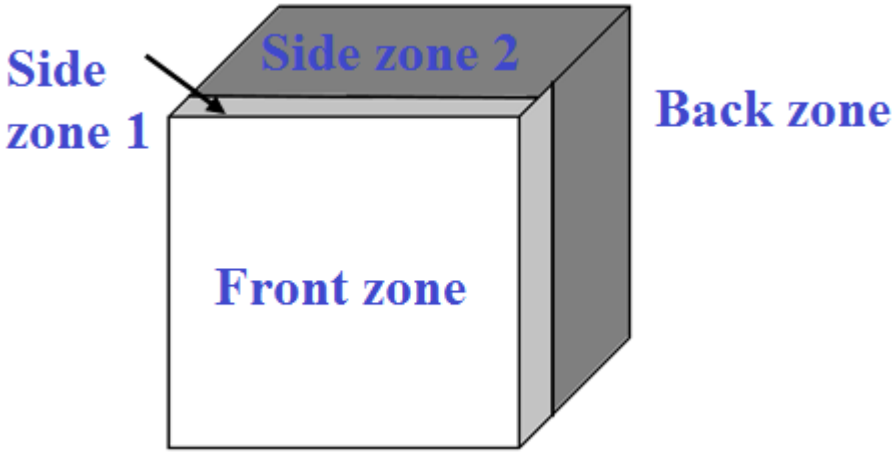
NOKIA

Zonal-PSNR quality assessment methodology

- Measuring the quality of experience over a set of discrete quality assessment views (QAVs).
- Rendering a cubemap using the closest viewport-based stream
- More focus on the equator, horizontal head movement rather than vertically and head movements are slow

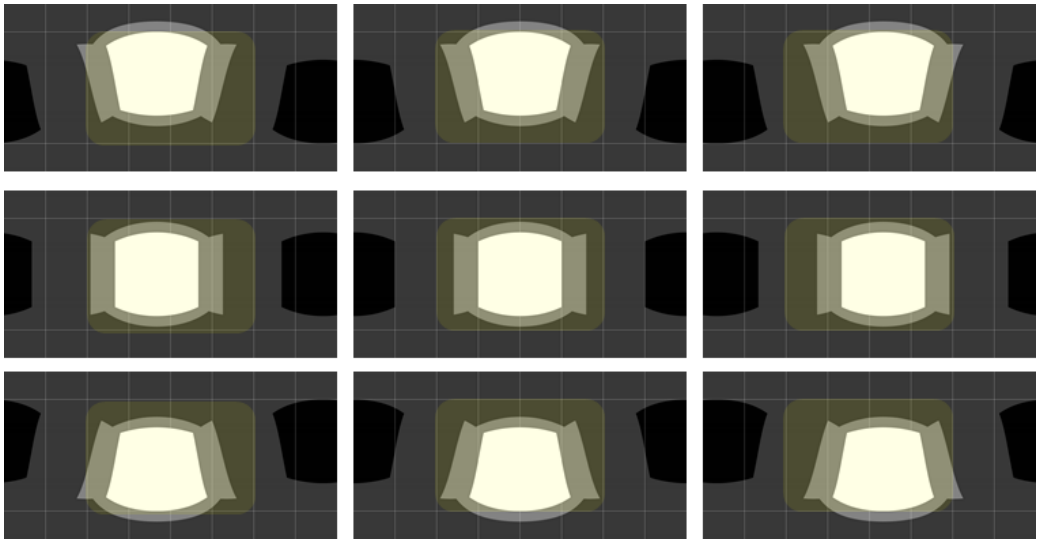
Blue marks: center of streams
Red marks: center of QAV





NOKIA

Quality assessment zones in ERP



Different zones on ERP for (yaw, pitch) pairs of:
Top row: (15°, 15°), (0°, 15°), (-15°, 15°)
Middle row: (15°, 0°), (0°, 0°), (-15°, 0°)
Bottom row: (15°, -15°), (0°, -15°), (-15°, -15°)

NOKIA

Experimental conditions

- Comparing MCTS-based mixed-resolution VAS using the proposed packing
  - MCTS-based mixed-quality VAS
  - Single-stream mixed-resolution (pixel-domain region-wise packing)
- Standard HEVC encoder with motion-constrained tile set (MCTS)
- QPs for mixed-resolution VAS: [20, 23, 26, 29]
- QPs for mixed-quality VAS: [16, 19, 22, 25] for viewport and sequence-wise selection of QPs for non-viewport
- Eight Joint Video Exploration Team (JVET) video sequences at 6K and 8K
- For ERP to cubemap conversion JVET 360Lib tool was used.
- A viewport of 90°x90° FOV is rendered for each QAV.

Experimental results

Streaming and storage performance of mixed-resolution VAS relative to mixed-quality VAS techniques.

	Streaming				Storage
Sequences	FZ	SZ1	SZ2	BZ	
Balboa	3.03	3.88	-10.97	-29.08	4.22
Broadway	0.37	0.37	-2.06	-22.51	2.01
BrandCastle	-13.73	-15.21	-0.55	-13.44	-10.17
Landing	-28.41	-28.00	-20.05	-22.30	-27.09
Gaslamp	-19.27	-16.09	-6.14	-3.95	-17.39
Harbor	-31.52	-26.15	-8.33	-13.24	-30.23
KiteFlite	-9.78	-11.41	-7.88	-34.69	-7.16
Trolley	-31.66	-30.80	-3.66	-21.16	-29.80
Average	-16.37	-15.43	-7.46	-20.05	-14.45

- The PSNR values are averaged over the all QAVs.
- Storage and streaming bitrate saving in terms of Bjøntegaard Delta-rate
- Negative values indicate that mixed-resolution VAS method outperforms
- The proposed layout decreases the streaming bitrate up to 32%.

Experimental results

Streaming and storage performance of mixed-resolution VAS relative to viewport-adaptive RWP

Negative values indicate that mixed-resolution VAS method outperforms

Sequences	Streaming	Storage
Balboa	3.49	-87.06
Broadway	3.48	-87.06
BrandCastle	2.14	-87.23
Landing	2.37	-87.20
Gaslamp	1.79	-87.27
Harbor	1.45	-87.31
KiteFlite	0.76	-87.40
Trolley	0.71	-87.41
Average	2.03	-87.24

- RWP technique brings 2% higher streaming performance on average, however it requires 87% higher storage.
- Other drawbacks of RWP:
  - Not adapted to the characteristics of the HMDs
  - Significant pre-processing encoding overhead

NOKIA

Conclusion

An OMAF-compliant packing layout of 6K ERP was proposed.

The proposed packing provides 6K resolution on viewport

While complying with the standard HEVC 4K-decoding constraint

Comparing resolution adaptation against quality adaptation in MCTS-based VAS

The 6K effective packing is a suitable candidate for viewport-adaptive streaming of VR content.

Thank you for your attention!