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

ACM Multimedia Systems Conference 23rd Packet Video Workshop
Alireza Zare et. al.

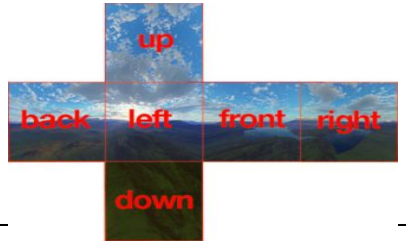
Presented by Hsuan-Chi Kuo

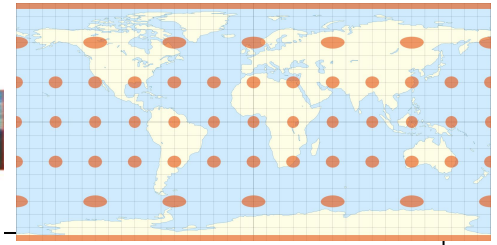
Introduction

- VR is getting more and more popular
- Head-mounted displays (HMD) can support a wider field of view (180 degree) and a higher resolution (6~8K).

Photo credit:https://uploadvr.com/new-john-carmack-software-doubles-oculus-home-resolution-gear-vr/

Background





CMP and ERP

- Omnidirectional MediA Format(OMAF) :
  - Standard that regulate the delivery of 360° content
  - Only supports equirectangular projection (ERP) and cubemap projection (CMP) and their region-wise packing.
    - Region-wise packing(RWP): rectangular regions of the projected frame may be resampled, rotaxted or mirrored.
  - HEVC-based viewport-dependent OMAF video profile supports picture size of 8,912,896 pixels, corresponding to **4K**.

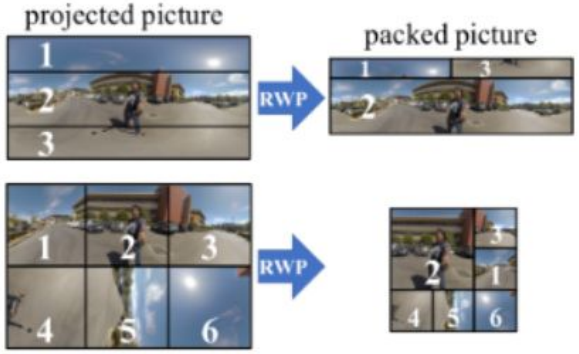


Figure 1: Two RWP examples based on the ERP format.

Figure credit:Skupin, R., Sanchez, Y., Wang, Y. K., Hannuksela, M. M., Boyce, J., & Wien, M. (2017, December). Standardization status of 360 degree video coding and delivery. In 2017 IEEE Visual Communications and Image Processing (VCIP) (pp. 1-4). IEEE.

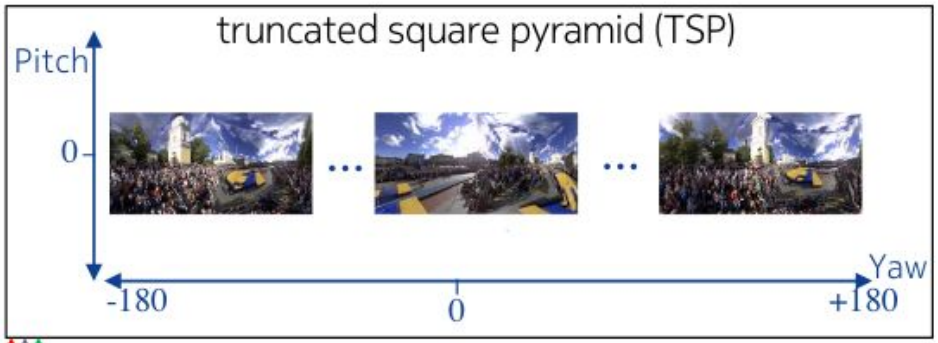
Background

- Viewport-adaptive streaming (VAS)
  - Theoretically, we can only transmit video contents corresponding to the current viewport. The entire 360° video is transmitted due to the limitation of the VR system.
  - Higher resolution for current viewport and lower resolution for the non-viewport of the video.
  - Viewport-dependent projection
  - Tile-based streaming

VAS: viewport-dependent projection

In each viewport stream:

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



Input 360° content

Viewport-based projection/mapping

Encoding and packaging

Orientation feedback

Network

Selected viewport representation

Standard HEVC decoder

Player and HMD

VR viewport-dependent unicast streaming system

Slide credit:https://2018.packet.video/wp-content/uploads/sites/6/2017/12/3001.pdf

VAS: Tile-based streaming

- Divide the 360° video into segments(tiles) with different resolutions.
- Combine tiles with different resolutions to generate the viewport.

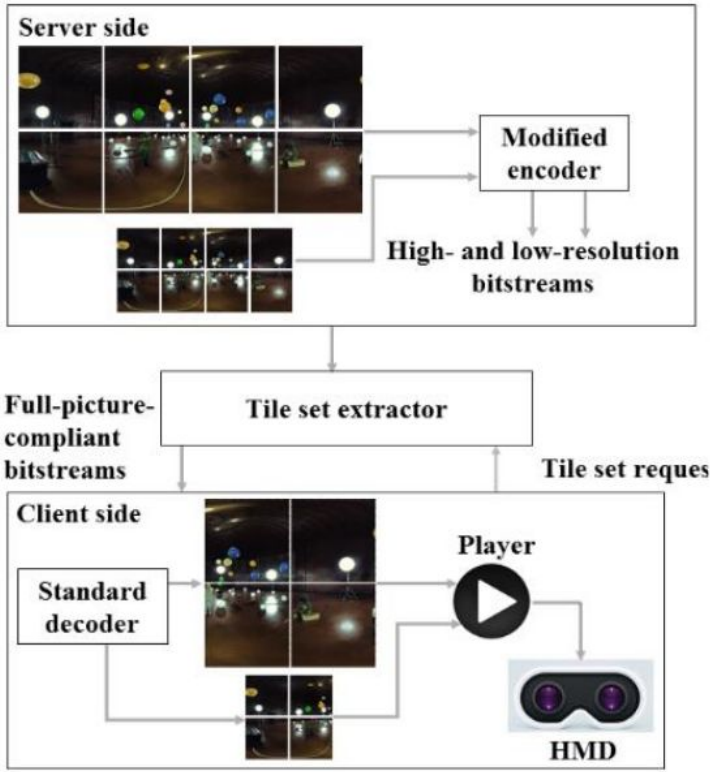
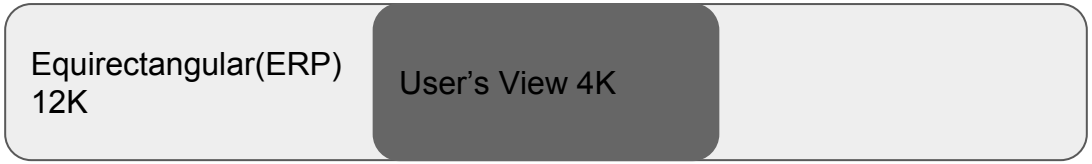


Figure 1. System overview (with 8-grid tiling)

Photo credit:Zare, A., Aminlou, A., Hannuksela, M. M., & Gabbouj, M. (2016, October). HEVC-compliant tile-based streaming of panoramic video for virtual reality applications. In Proceedings of the 2016 ACM on Multimedia Conference (pp. 601-605). ACM.

Problem

- If users want to see a 4K view, the video content should be 12K.



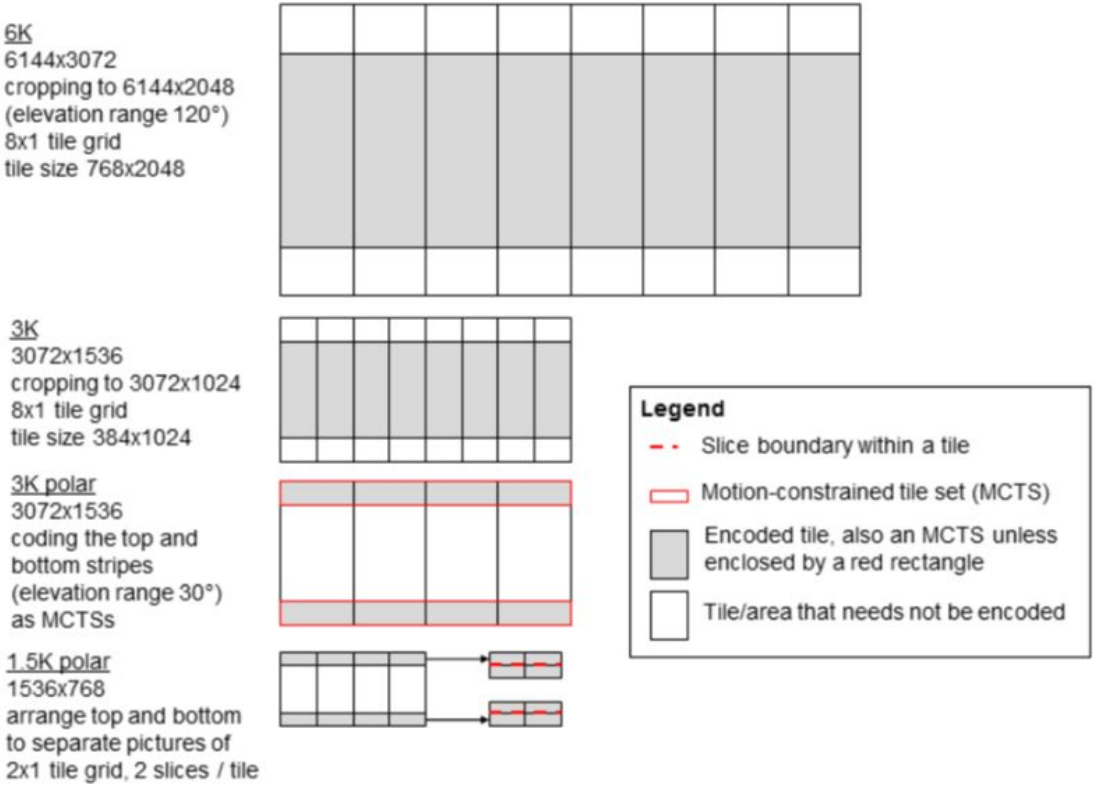
- The hardware decoders can only decode frames no bigger than 4K.
- OMAF also has 4K decoding constraint.
- It means that, If the ERP is 4K → user's view is lower than 4K.

Solution

- A way to pack **part of** a 6K video in to a 4K frame without breaking the 4k-decoding constraint

Extractor track
6x2 tile grid
tile widths 768,768,768,768,384,384
tile heights 2048,256
1 or 2 slices per tile
picture size 3840x2304

Solution


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
Legend
Slice boundary within a tile
Motion-constrained tile set (MCTS)
Encoded tile, also an MCTS unless enclosed by a red rectangle
Tile/area that needs not be encoded

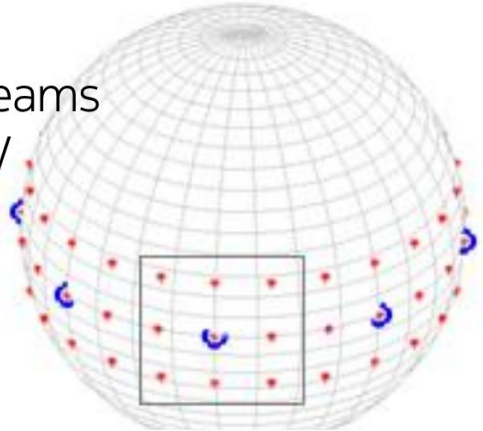
How to evaluate the QoE of a 360° video?

- The displayed content might partially rendered from the low-quality(resolution) content.
- Head movement
- is not predictable

Zonal-PSNR(Peak Signal to Noise Ratio)

- Measure the QoE over a set of quality assessment views. (QAVs)
- Render the video using the closet viewport stream
- More focus on the equator
  - Horizontal head movement is faster than the vertical one.

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



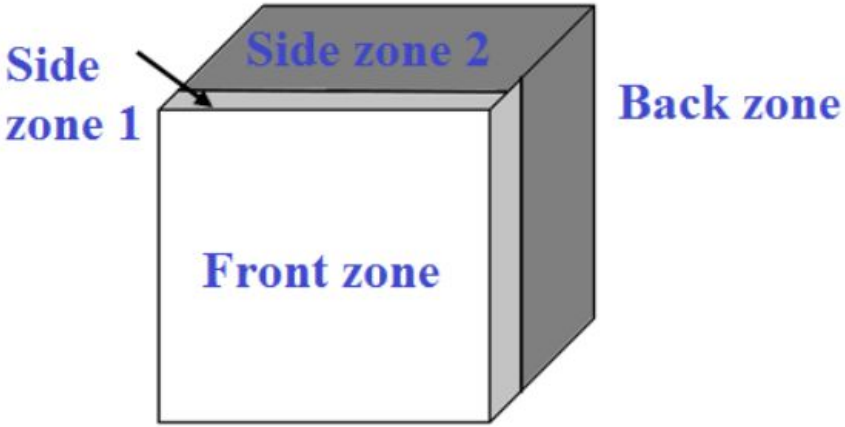


Photo credit:https://2018.packet.video/wp-content/uploads/sites/6/2017/12/3001.pdf

Quality assessment zones in ERP

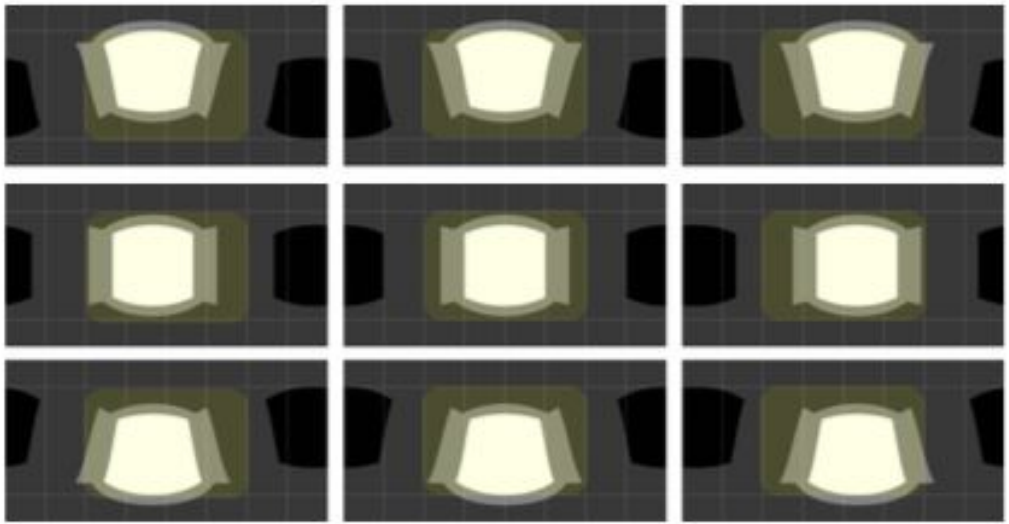


Figure 4: Different zones on ERP for (yaw, pitch) pairs of:

Top row: (15°, 15°), (0°, 15°), (-15°, 15°), Middle row: (15°, 0°), (0°, 0°), (-15°, 0°), Bottom row: (15°, -15°), (0°, -15°), (-15°, -15°)

Experiment Setup

- Comparing MCTS based mixed resolution VAS using the proposed packing (mix-resolution)
  - MCTS-based mixed-quality VAS (mix-quality)
  - Single-stream mixed-resolution or pixel-domain region-wise packing (RWP)
- A viewport of 90°x90° FOV is rendered for each QAV.
- 8 videos, 6k and 8k, downsampled to
  - 6K for mixed-resolution
  - 4K for mixed-quality
  - 3840x2304 for RWP

Mix-resolution vs mix-quality

- The PSNR(a metric to evaluate picture quality, the higher the better) values are averaged over the all QAVs
- BD-rate: negative values means that given the same PSNR, it needs less bits

Table 2: Streaming and storage performance of mixed-resolution VAS relative to mixed-quality VAS techniques in different zones (BD-rate (%))

	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**

Mix-resolution vs RWP

- Since RWP has no divided zones, only front-zone in mixed-resolution is compared.
- Other disadvantages of RWP:
  - Require many pre-processing efforts
  - Waste the encoding capacity

Table 3: Streaming and storage performance of mixed-resolution VAS relative to viewport-adaptive RWP (BD-rate (%))

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**

Takeaway

- A 6K-effective packing layout is proposed while complying to the current constraints.
- Outperform mixed-quality in the aspect of streaming and storage up to 32% and 20% respectively.
- A good candidate for future viewport-adaptive VR streaming.