Spatio-Temporal Activity based Tiling for Panorama Streaming

Y. Sanchez, R. Skupin, C. Hellge, and T. Schierl
Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI)

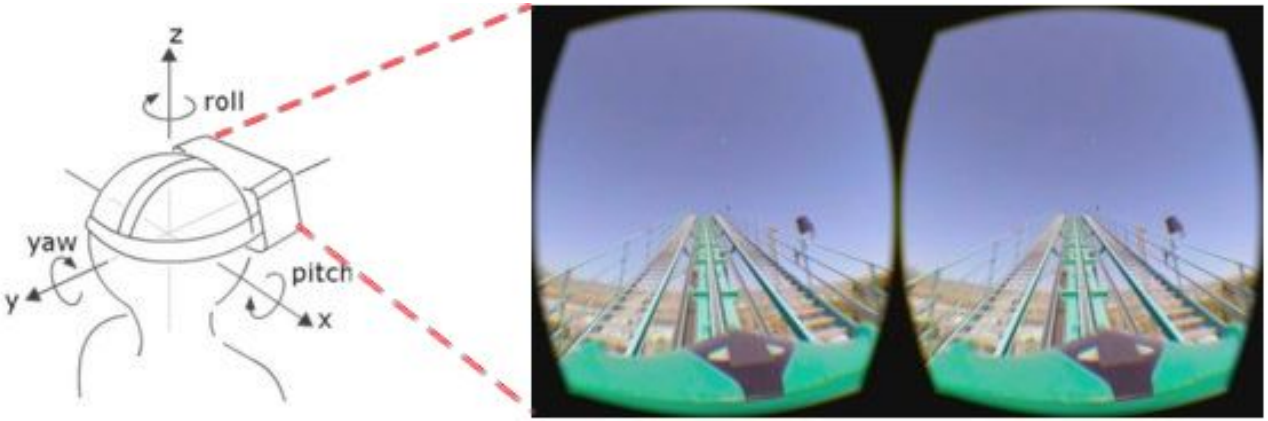
In Proc. of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'17), Taipei, Taiwan, June, 2017, pp. 61–66.

Outline

- Motivation
- Problems to be solved
- Related work
- Spatial-temporal activity metrics
- Validations
- Conclusion

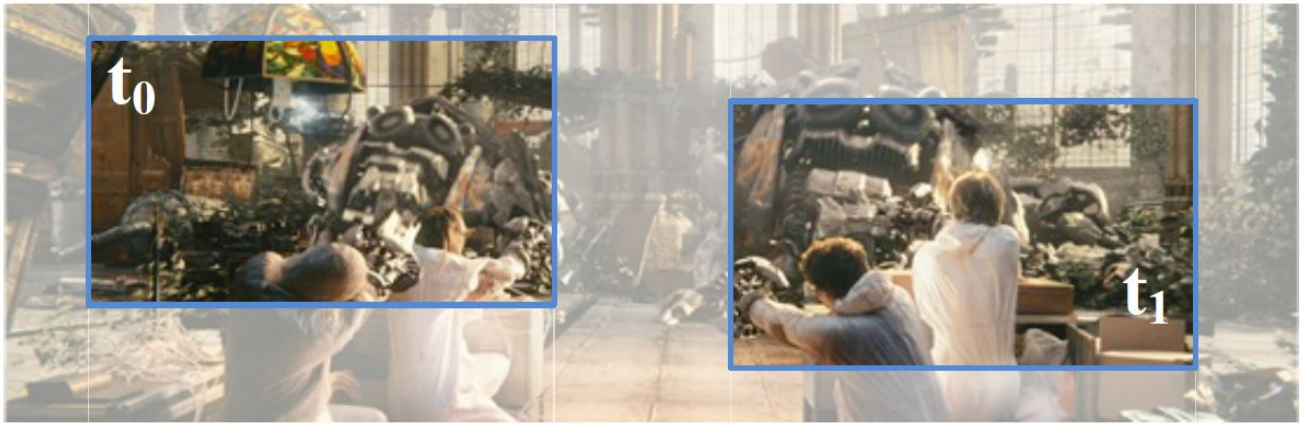
Motivation

- In panorama streaming, users can navigate the high-resolution videos with an arbitrary Region-of-Interest (RoI)
- Transmitting the whole video is unfeasible
- Tile-based panoramic streaming overcomes the mentioned drawbacks



To be solved...

- Tile based panoramic streaming allows users to receive a set of tiles that match their RoI instead of the whole video
- How to derive from the video content the optimal tile size in a low complexity manner



Related work

- Brute force approach [1]
- High computational complexity
- Pixel overhead & bitrate per pixel (bpp)

$$BR(t_w, t_h) = RoI_{size} * \eta(t_w, t_h) * \varphi(t_w, t_h) \qquad (1)$$

with η being:

$$\eta(t_w, t_h) = \frac{(r_w + t_w - 1) * (r_h + t_h - 1)}{RoI_{size}} \qquad (2)$$

[1] A. Mavlankar, B. Girod, “Spatial-Random-Access-Enabled Video Coding for Interactive Virtual Pan/Tilt/Zoom Functionality”, IEEE Transactions on Circuits and Systems for Video Technology (CSVT), Vol.21, No. 5, May 2011.

Spatial-temporal activity metrics

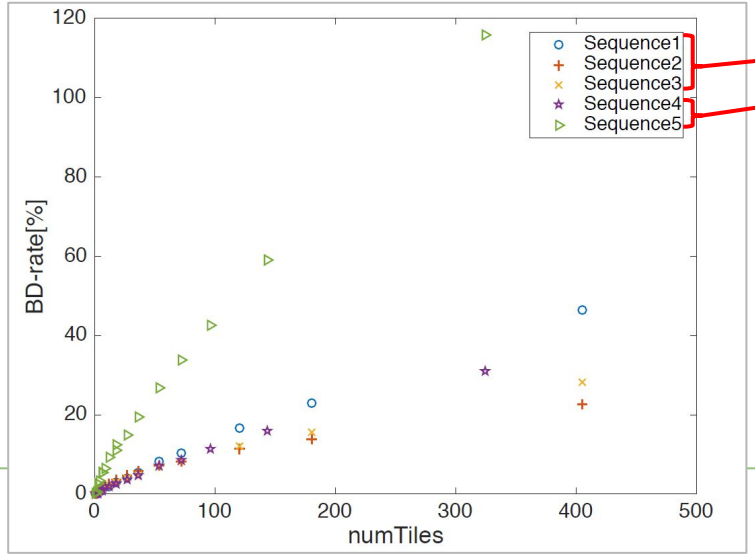
- Based on aforementioned method,

$$\left(t_w^{opt}, t_h^{opt}\right) = \arg \min_{(t_w, t_h)} \eta(t_w, t_h) * \left(1 + BD(t_w, t_h)\right) \quad (5)$$

- Pixel overhead is easy to be computed and does not depend on the content
- Bjøntegaard-Delta bitrate (BD-rate) measurement method
  - For the same PSNR, and
  - Negative values tell how much lower the bitrate is reduced (coding efficiency is increased), and positive values tell how much the bitrate is increased (coding efficiency is reduced)

Spatial-temporal activity metrics (cont.)

- For a high number of tiles the BD-rate can vary from around 20% to around 120%
- The test conditions used by JCT-VC during standardization



Num. tiles	Tile sizes for 6912x1920	Tile sizes for 6912x1532
2	3456x1920	3456x1536
4	3456x960	3456x768
3	2304x1920	2304x1536
6	2304x960	2304x768
9	2304x640	2304x512
6	1152x1920	1152x1536
12	1152x960	1152x768
18	1152x640	1152x512
18	768x960	768x768
27	768x640	768x512
54	768x320	768x256
36	576x640	576x512
72	576x320	576x256
120/96*	576x192	576x192
180/144*	384x192	384x192
405/314*	256x128	256x128

*different number of tiles for the different resolution videos

Spatial-temporal activity metrics (cont.)

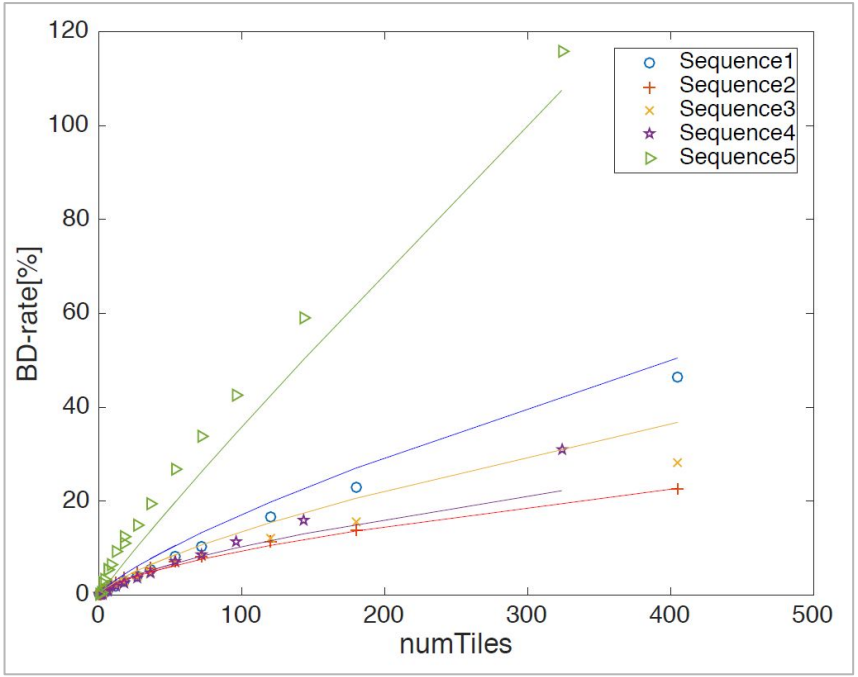
- Based on [2] and [3], it can derive the equation below,

$$BD_{model}(t_w, t_h) = \frac{1}{0.82+1.5*SA^{-0.05}*TA^{-0.04}} * (N_{tiles}(t_w, t_h) - 1)^{\frac{1}{0.32*SA^{0.33}*TA^{0.11}}} \quad (10)$$

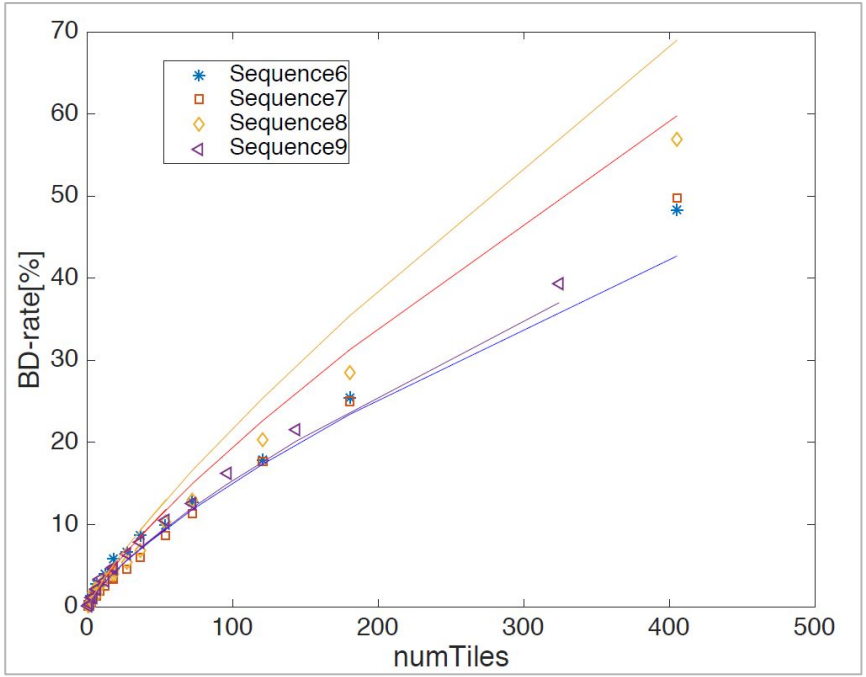
$$BD_{model}(t_w, t_h) = \quad (10)$$
$$= \frac{1}{0.82+1.5*SA^{-0.05}*TA^{-0.04}} * (N_{tiles}(t_w, t_h) - 1)^{\frac{1}{0.32*SA^{0.33}*TA^{0.11}}}$$

Spatial-temporal activity metrics (cont.)

- The presented model for each sequence as well as the actual BD-rate values



Training sequences



Testing sequences

Validations

- Real -> the actual BD-rate values measured for tiled encodings
- Model -> the predict values measured by the proposed model

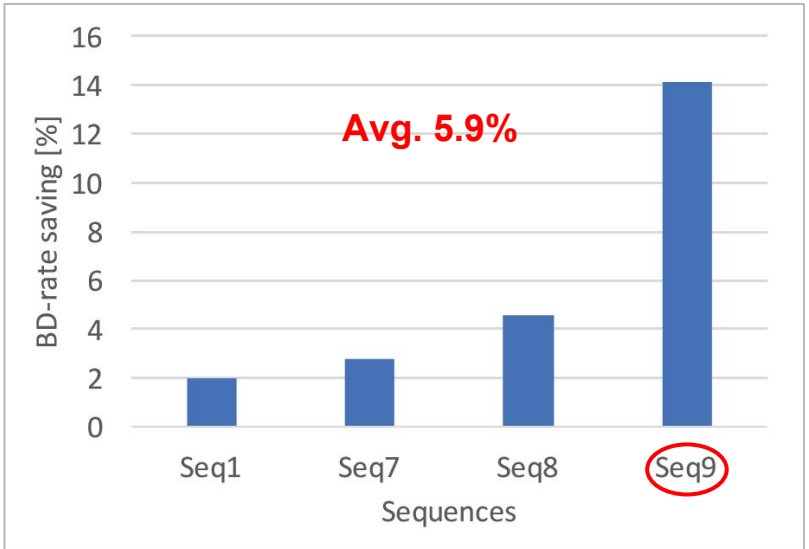
Table 4: Optimal size using the real BD-rate and the model in Eq. 10

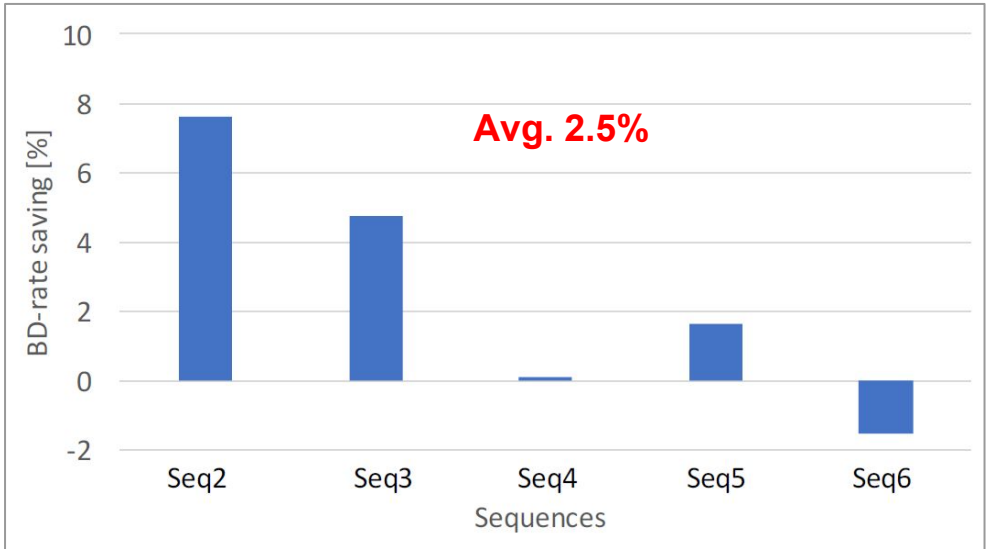
	Real	Model
Seq1	384x192	384x192
Seq2	256x128	256x128
Seq3	256x128	256x128
Seq4	256x128	256x128
Seq5	576x192	576x192
Seq6	384x192	256x128*
Seq7	384x192	384x192
Seq8	384x192	384x192
Seq9	256x128	256x128

*different result when using the model

Validations (cont.)

- The most occurring tile sizes among the 9 sequences is 384x192 and 256x128
- It computes the BD-rate savings in comparison to having a static configuration





Conclusion

- On average gains of 5.9% are achieved as tile size is 256x128
- On average gains of 2.5% are achieved as tile size is 384x192

- They proposed an optimization process, which has a low complexity compared to performing encodings for each tiling variant, in order to minimize the transmitted bitrate of the RoI content

Q & A