A Journey in the Challenges of Immersive Streaming

Lucile Sassatelli

Université Côte d'Azur, CNRS, I3S

Institut Universitaire de France

Coresa Tutorial, November 3rd, 2021

About me

UNIVERSITÉ CÔTE D'AZUR

- Research activities: Network Streaming of Immersive Media with Machine Learning and User-centric approaches
  - Joint control of streaming and VR experience
  - Modeling of human attention in VR

- Research projects:
  - **Europe H2020 AI4Media**, PI for 3IA-UCA (2020-2024, 30 partners): A European Excellence Centre for Media, Society and Democracy
  - **ANR PRC TRACTIVE** (2022-2026, 6 partners): Towards a computational multimodal analysis of film discursive aesthetics
  - ReVEGO (2021-2024, 3 partners): VR pour l'Egalité de Genre dans l'Orientation

- Co-responsable du Centre XR^2C^2 de UCA (13 laboratoires) : centre multi-disciplinaire de recherche et création sur la XR
- TPC chair of ACM Multimedia Systems 2021

Sources

- F. Chiariotti, ‘‘A survey on 360-degree video: Coding, quality of experience and streaming,’’ Comput. Commun., vol. 177, pp. 133–155, Sep. 2021.
- Christian Timmerer and Ali C. Begen. 2019. A Journey Towards Fully Immersive Media Access. In ACM Multimedia 2019.

Outline

UNIVERSITÉ CÔTE D'AZUR

- Streaming 360° videos: a multi-disciplinary problem

1. How to project, how to compress spherical content?
2. How to stream high-quality spherical content?
3. How to predict human head motion?
4. How to define Quality of Experience for spherical content?

- New interdisciplinary levers for VR streaming
- Conclusions

Immersive contents: societal potential

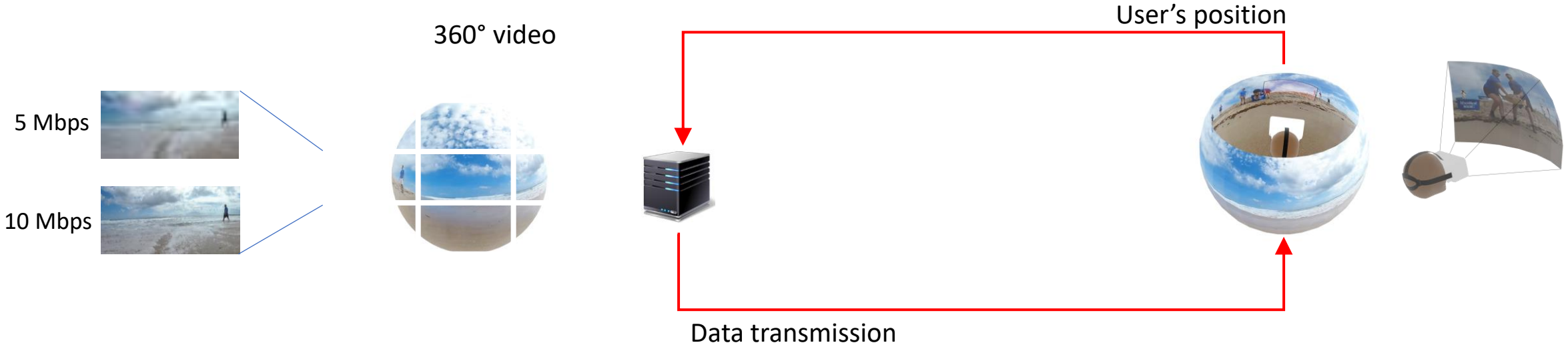
UNIVERSITÉ CÔTE D'AZUR

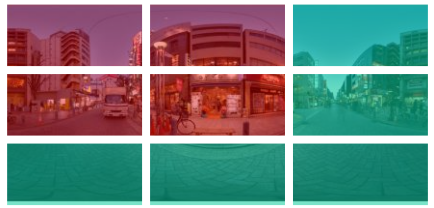
- Promising applications:
  - Journalism, fiction
  - « *The empathy machine* »: change the perception
  - Remote education

- But the development of immersive contents is still hindered by:
  - Headset technology
  - Design and production
  - Distribution

    → How to stream over the Internet?

Problem: streaming 360° videos





- Requires very high data rates
  - 5.2Gbps
- Adaptation of resolution based on possible field of view

- To match the fovea's sensitivity: 200 pixels/degree

→ 360°*200px/°*180°*200px/°*36b/px*60fps*2stereo/600 = **18.7Gbps**

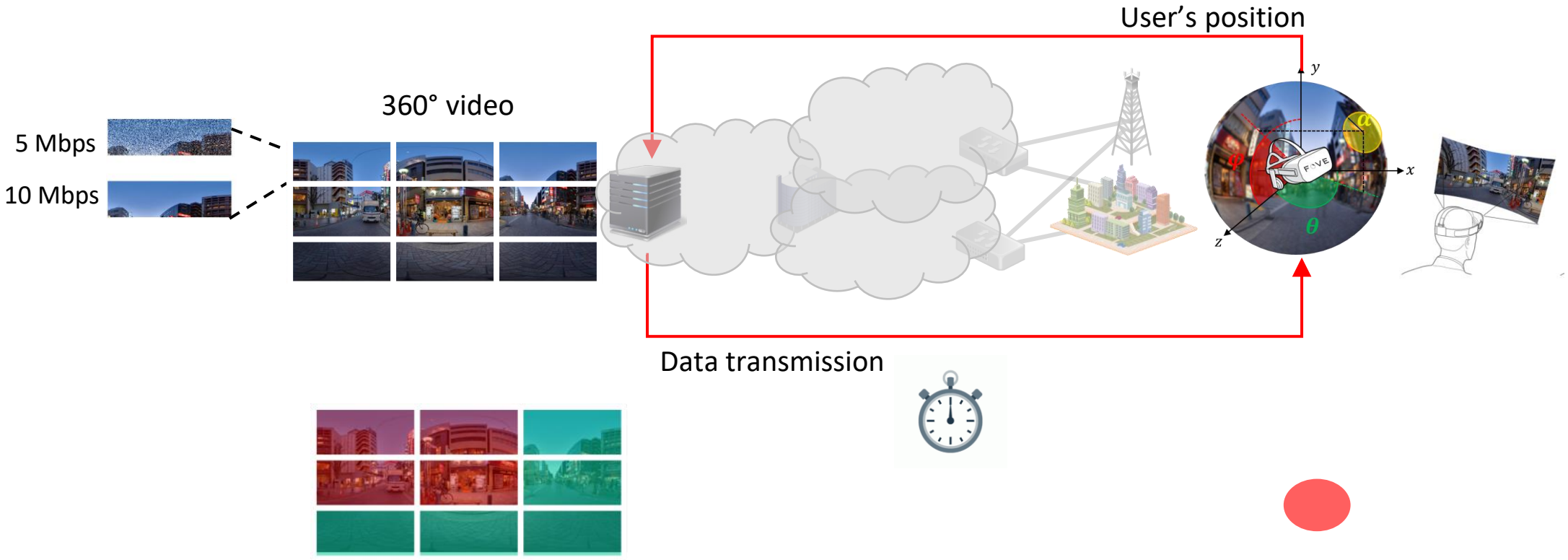
→ For a FoV of 150°x120°: **5.2Gbps**

- Acceptable resolution of 4K per FoV => 24K res, up to 100 fps

→ **500Mbps**

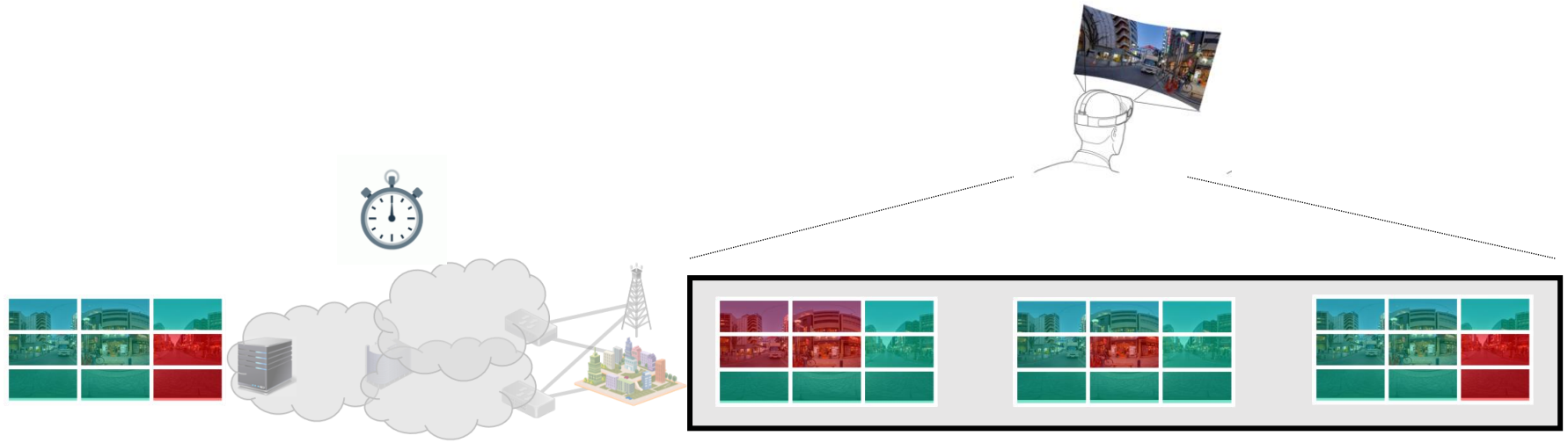
[1] E. Bastug et al.. Toward Interconnected Virtual Reality: Opportunities, Challenges, and Enablers. Jun. 2017. IEEE Communiations Mag..

Problem: streaming 360° videos



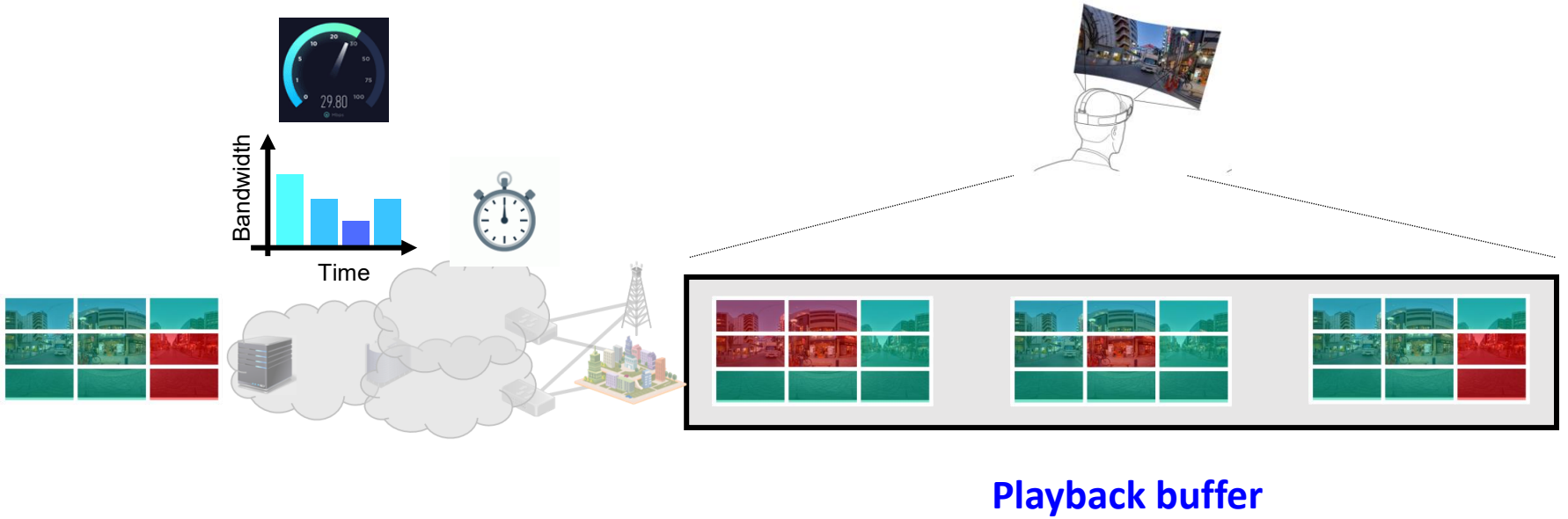
- Network latency requires to know user's motion in advance.

Problem: streaming 360° videos



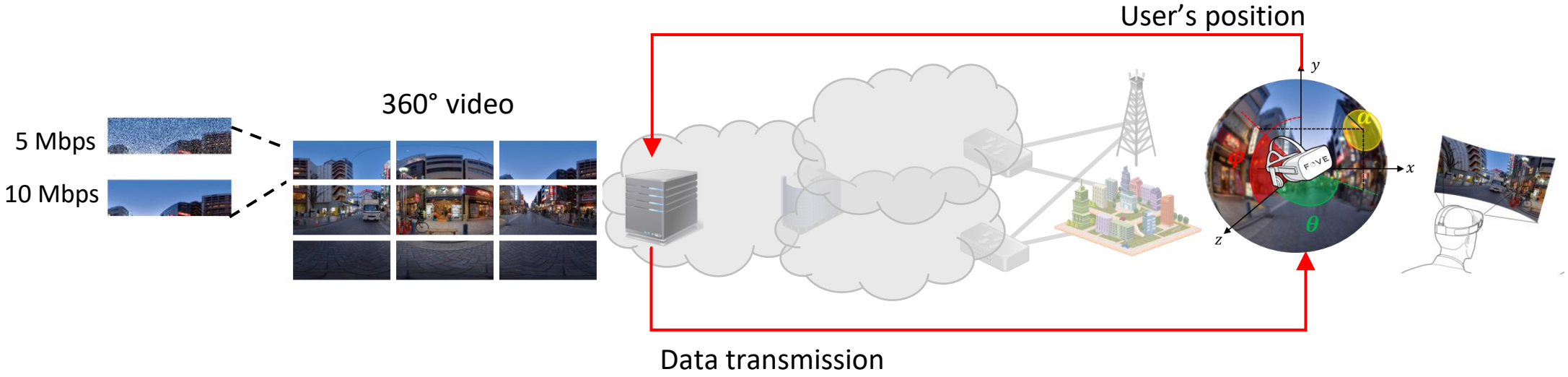
- A playback buffer is crucial to absorb bandwidth variations, but increases end-to-end delay.

Problem: streaming 360° videos



- A playback buffer is crucial to absorb network latency and bandwidth variations.

Problem: streaming 360° videos

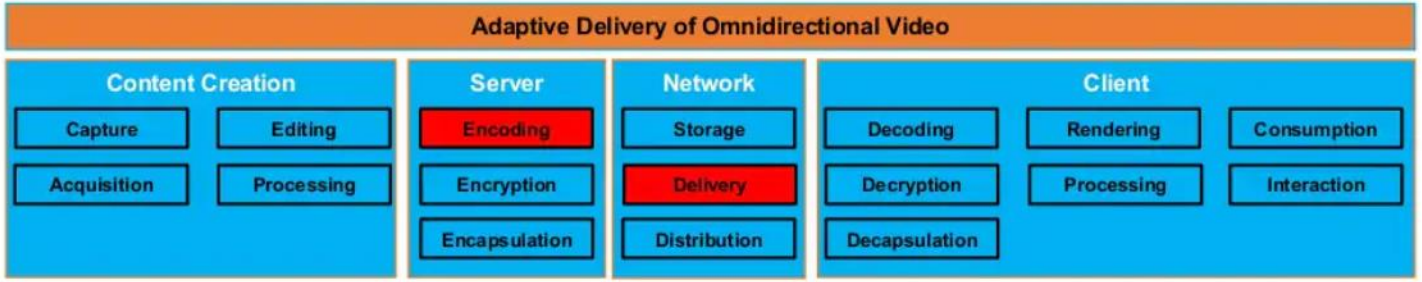


→ Visual quality and consumed rate get dependent on human motion

→ Harder challenges, Solid groundwork for interdisciplinarity

Functional architecture

Capturing Devices



Consumer Devices

[1] Christian Timmerer and Ali C. Begen. 2019. A Journey Towards Fully Immersive Media Access. ACM Multimedia 2019.

Outline

UNIVERSITÉ CÔTE D'AZUR

- Streaming 360° videos: a multi-disciplinary problem

1. How to project, how to compress spherical content?
2. How to stream high-quality spherical content?
3. How to predict human head motion?
4. How to define Quality of Experience for spherical content?

- New interdisciplinary levers for VR streaming
- Conclusions

Acquisition

- Omnidirectional images and videos are usually stitched from multiple cameras → several types of issues at the edges

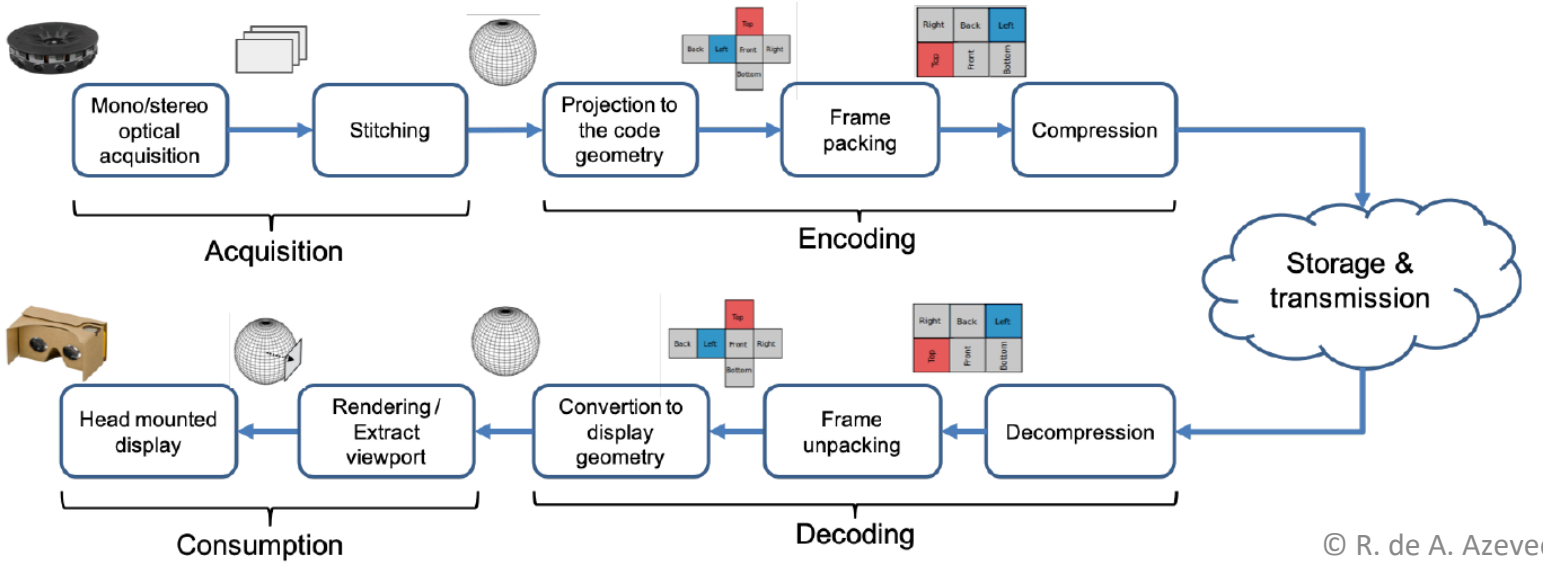
(a) (b) (c) (d) (e)

Fig. 3: Examples of *stitching* artifacts: (a) broken edges and (b) missing information; *blending* artifacts: (c) ghosting and (d) exposure; and *warping* artifacts: (e) geometrical distortions / object deformations.

R. G. d. A. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli and P. Frossard, "Visual Distortions in 360° Videos," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 8, pp. 2524-2537, Aug. 2020, doi: 10.1109/TCSVT.2019.2927344.

Coding, compression and distortion

- Most of the omnidirectional video systems re-use the same encoding tools as classical video solutions (H.264, H.265, VP9, or AV1).
- Since filters and coding tools are made for 2D images, the spherical content needs to be projected to a flat surface to be processed and encoded.

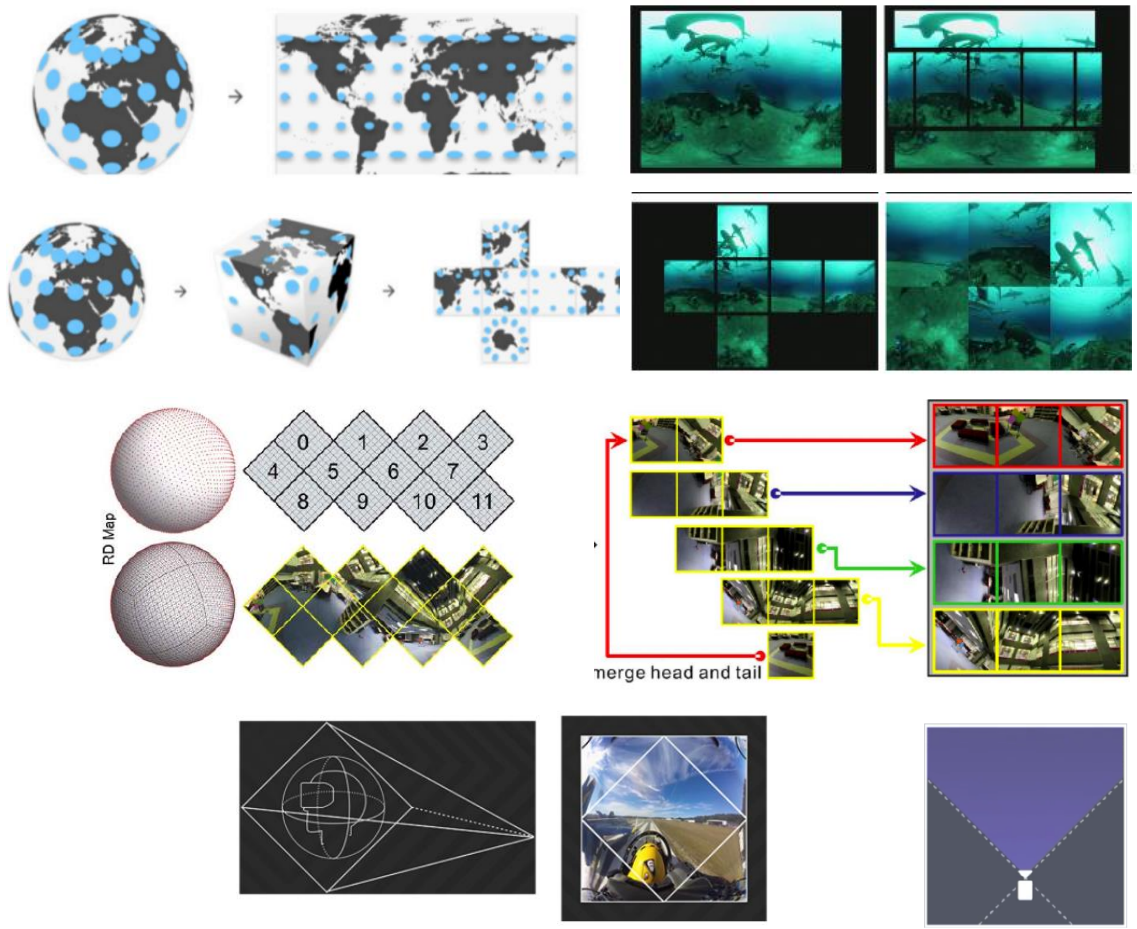


© R. de A. Azevedo

R. G. d. A. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli and P. Frossard, "Visual Distortions in 360° Videos," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 8, pp. 2524-2537, Aug. 2020.

Encoding

- Converting the 360° image to a planar representation with:
  - Equirectangular Projection (ERP): the poles get more pixels than the equator
  - CubeMap Projection (CMP):
  - Rhombic Dodecahedron Map:
  - Offset projections: views in one direction have a higher pixel density than in other directions.



Tarek El-Ganainy, Mohamed Hefeeda: Streaming Virtual Reality Content. CoRR abs/1612.08350 (2016)

Projection artifacts

- Projection spherical ←→ planar involves resampling and interpolation
  - May result in aliasing, blurring, and ringing distortions
  - visible poles due to oversampling at the poles, may appear when using the ERP representation; and visible seams in the discontinuity regions.
- Solution: graph-based techniques like [2]
- However: current 360° systems exclusively rely on sampling and interpolation techniques in the classical rectangular geometry.

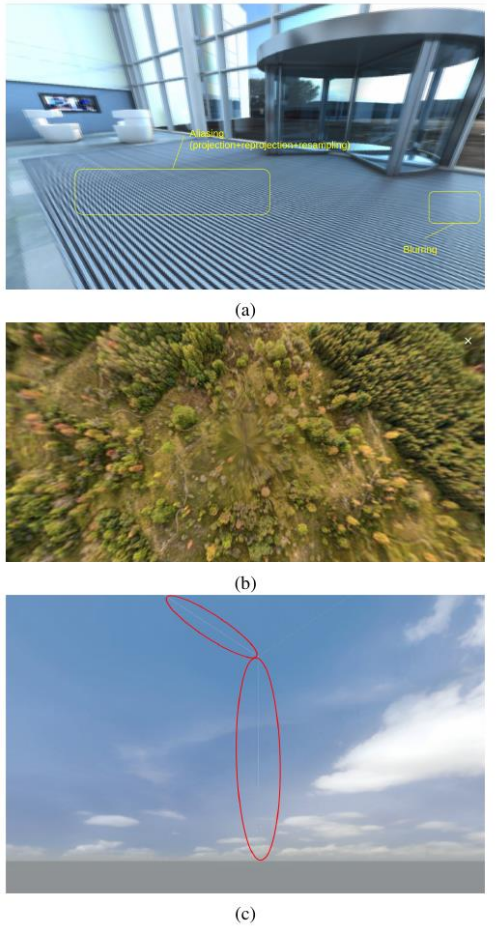


Fig. 6: Examples of (a) aliasing and blurring, (b) visible poles; and (c) visible seams due to projection and resampling.

[1] R. G. d. A. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli and P. Frossard, "Visual Distortions in 360° Videos," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 8, pp. 2524-2537, Aug. 2020, doi: 10.1109/TCSVT.2019.2927344.
[2] L. Bagnato, P. Frossard, and P. Vandergheynst, “Plenoptic spherical sampling,” in IEEE ICIP 2021.

Compression

- Most omnidirectional streaming systems reuse the 2D coding pipelines [1].
- Distortion from projection does not impact QoE only: it also impact coding efficiency.
- The interaction between the geometrical distortions and the lossy compression processing may result in visible artifacts.

(a)

(b)

© R. de A. Azevedo

[1] A. Zare, A. Aminlou, M. M. Hannuksela, M. Gabbouj, HEVC-compliant tile-based streaming of panoramic video for virtual reality applications, in ACM Multimedia 2016.

Compression artifacts

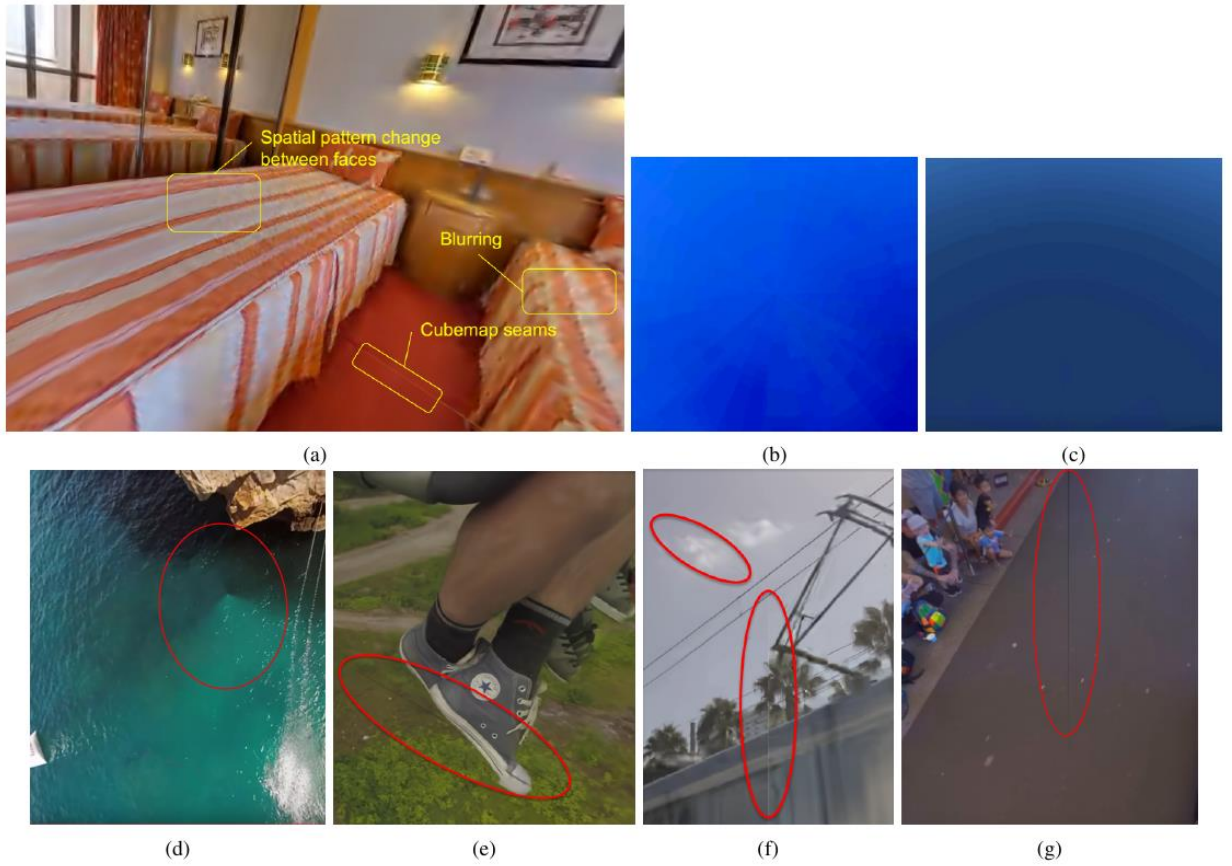


Fig. 7: Examples of compression artifacts on 360-degree content: (a) blurring, cubemap seams, and spatial pattern changes (due to jpeg2000 compression); (b) radial blocking; (c) radial banding; (d) visible pole; (e)-(f) cubemap seams artifacts; and (g) equirectangular seam artifact.

Optimize coding for a certain projection

UNIVERSITÉ CÔTE D'AZUR

- Adapt QPs: using the Weighted to Spherically Uniform PSNR (WS-PSNR) weights:
  - regions less important in the metric encoded with a rougher compression [1]
  - same can be performed with other metrics (S-PSNR) [2]
  - more advanced: combine the geometric information with the saliency [3]
- Adapt filters:
  - A correction to the standard HEVC deblocking filter can reduce the CMP edge distortion by aligning the face edges with the filter edges [4].
- Adapt motion estimation and temporal coding:
  - standard video coders only allow for block translations in the critical tool of motion compensated prediction → adapt motion models to use rotation [5]

[1] Y. Li, J. Xu, Z. Chen, Spherical domain rate-distortion optimization for 360-degree video coding, in: IEEE ICME 2017.
[2] Y. Liu, L. Yang, M. Xu, Z. Wang, Rate control schemes for panoramic video coding, Journal of Visual Communication and Image Representation 53 (2018).
[3] G. Luz, J. Ascenso, C. Brites, F. Pereira, Saliency-driven omnidirectional imaging adaptive coding: Modeling and assessment, in IEEE MMSP 2017.
[4] J. Sauer, M. Wien, J. Schneider, M. Bläser, Geometry-corrected deblocking filter for 360 video coding using cube representation, in IEEE Picture Coding Symposium (PCS), 2018
[5] B. Vishwanath, T. Nanjundaswamy, K. Rose, Rotational motion model for temporal prediction in 360 video coding, in IEEE MMSP 2017.

Outline

UNIVERSITÉ CÔTE D'AZUR

- Streaming 360° videos: a multi-disciplinary problem

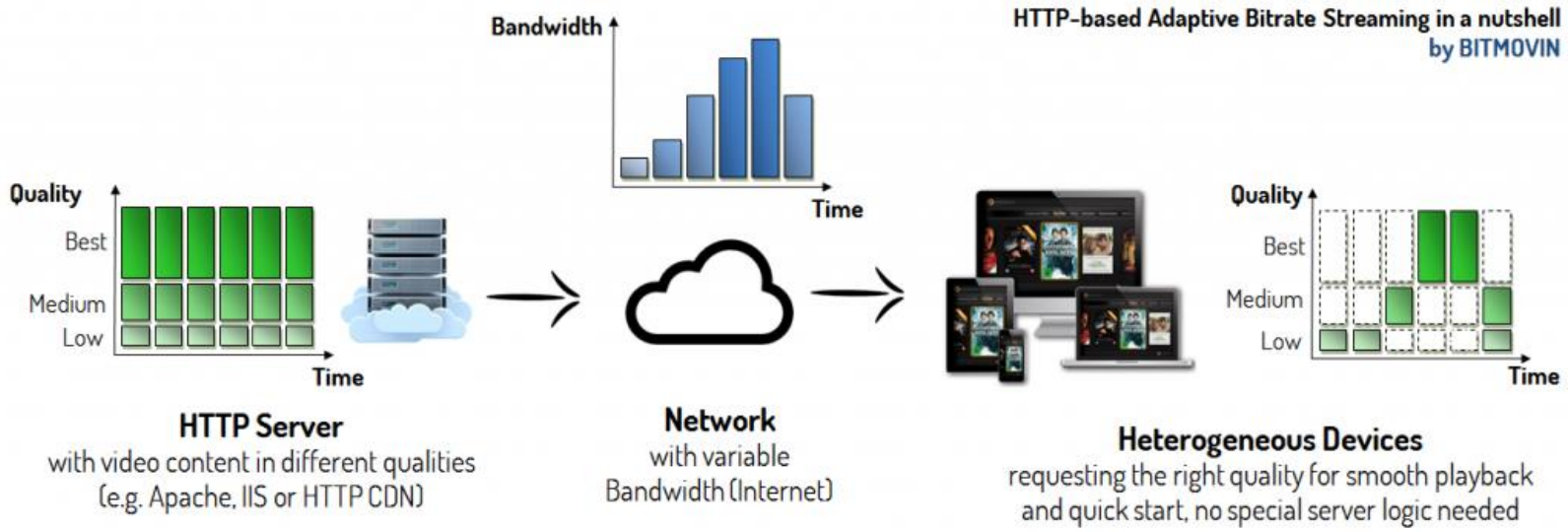
1. How to project, how to compress spherical content?
2. How to stream high-quality spherical content?
3. How to predict human head motion?
4. How to define Quality of Experience for spherical content?

- New interdisciplinary levers for VR streaming
- Conclusions

Regular video streaming: MPEG-DASH Dynamic Adaptive Streaming over HTTP

- Adapt the encoding rate to the available bandwidth

→ A DASH adaptation logic must be network-adaptive



© Bitmovin

360° video streaming strategies

UNIVERSITÉ CÔTE D'AZUR

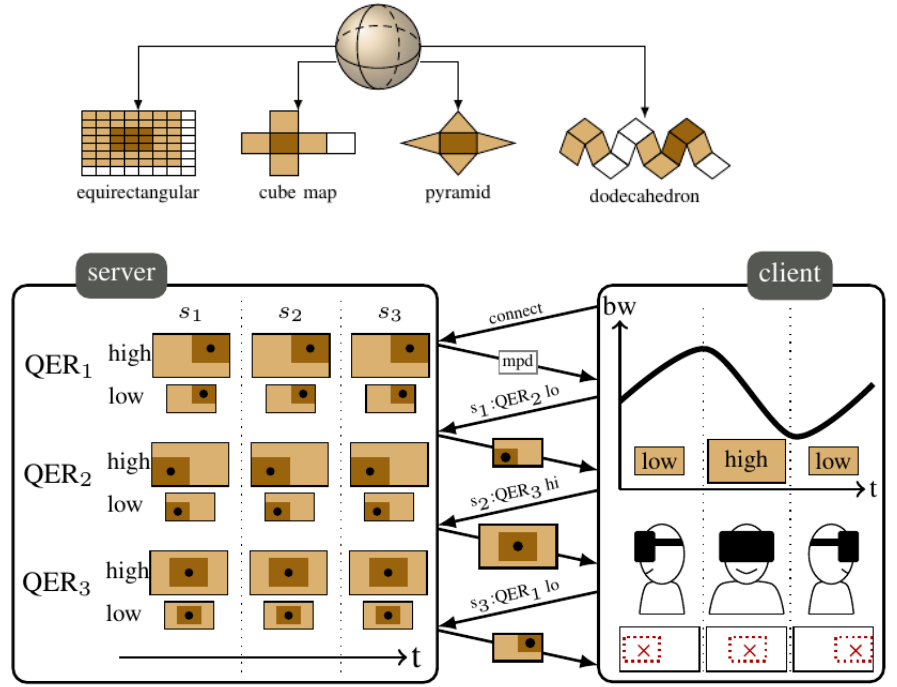
Viewport-agnostic streaming:

- Perform bitrate adaptation as for regular video, once the 360° projection and encoding has been made
  - Homogeneous quality, low FoV quality, bandwidth waste [1]

Viewport-dependent streaming:

- with monolithic frames from Quality Emphasized Representation [2]:
  - Multiple version for pre-defined viewports
  - Increased storage costs (at CDN), limited flexibility, good coding efficiency
  - Optimization problem: **for every segment:**

given past bandwidth measurements and FoV locations, **choose** which FoV-emphasized representation in which encoding rate to download



© X. Corbillon

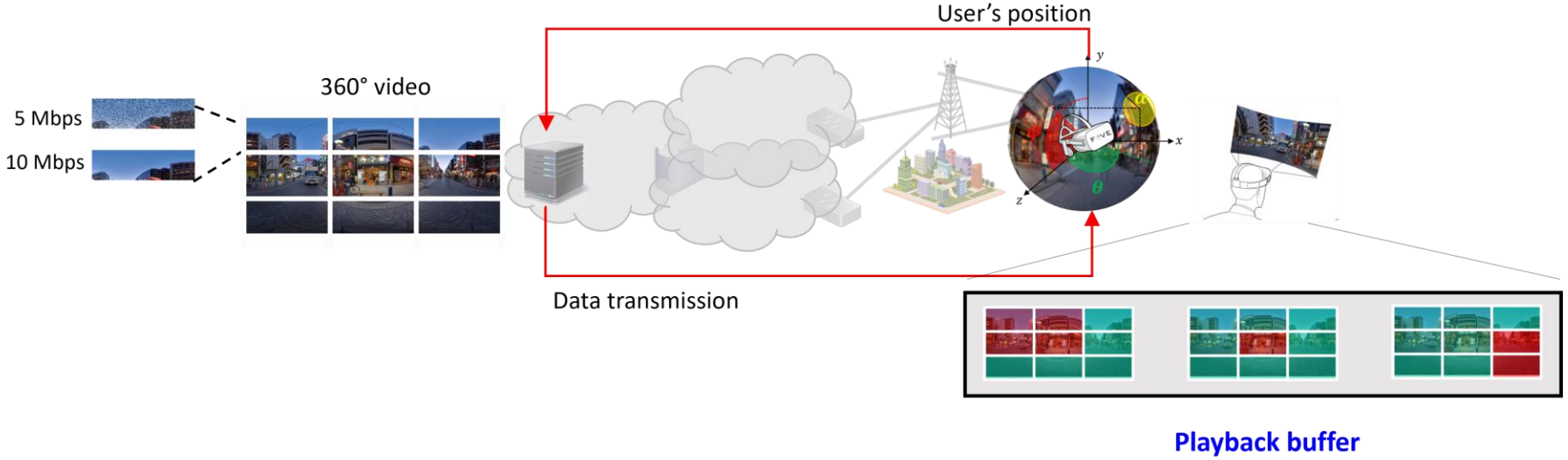
[1] Xing Liu, Bo Han, Feng Qian, and Matteo Varvello. 2019. LIME: understanding commercial 360° live video streaming services. ACM MMSys 2019.
[2] Xavier Corbillon, Gwendal Simon, Alisa Devlic, Jacob Chakareski. Viewport-Adaptive Navigable 360-Degree Video Delivery. IEEE ICC 2017.

360° video streaming strategies

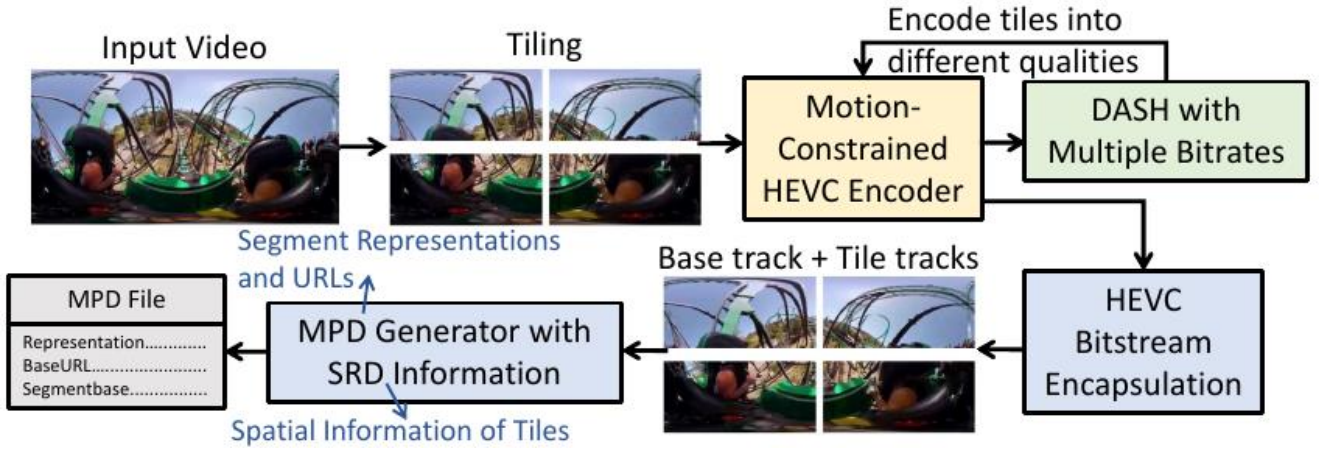
UNIVERSITÉ CÔTE D'AZUR

Viewport-dependent streaming:

- with monolithic frames from Quality Emphasized Representation
- with tiling:
  - Divides the content into independent video tiles
  - standards: MPEG DASH-SRD, MPEG OMAF, using tiling capabilities of modern video codecs (AVC and HEVC)



Pre-process to tile a 360° video

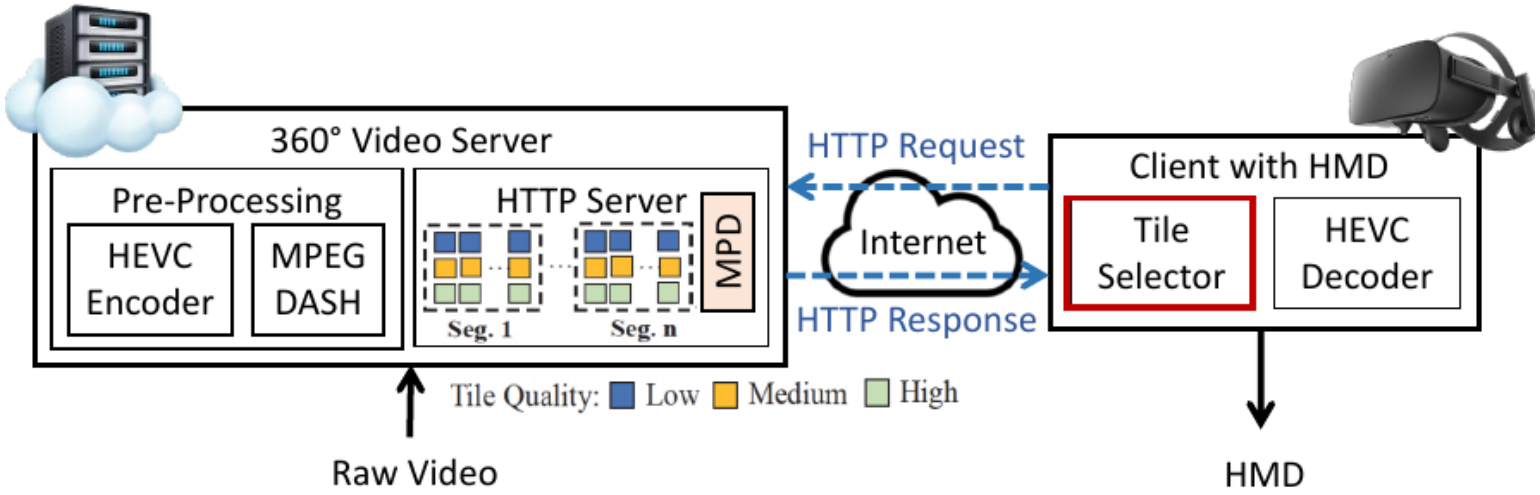


© W. Lo

- Split the video into tiles of sub-videos
- Encode the tiles using motion-constrained HEVC encoder with different bitrates (qualities)
- Encapsulate tiles into single HEVC bitstream
- Integrate with DASH for spatial index generation (MPD and SRD)

W. Lo, C. Fan, S. Yen and C. Hsu, "Performance measurements of 360° video streaming to head-mounted displays over live 4G cellular networks," 2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS), 2017, pp. 205-210, doi: 10.1109/APNOMS.2017.8094203.

Tile-based streaming platform



- Optimization problem: **for every segment, for every tile:**

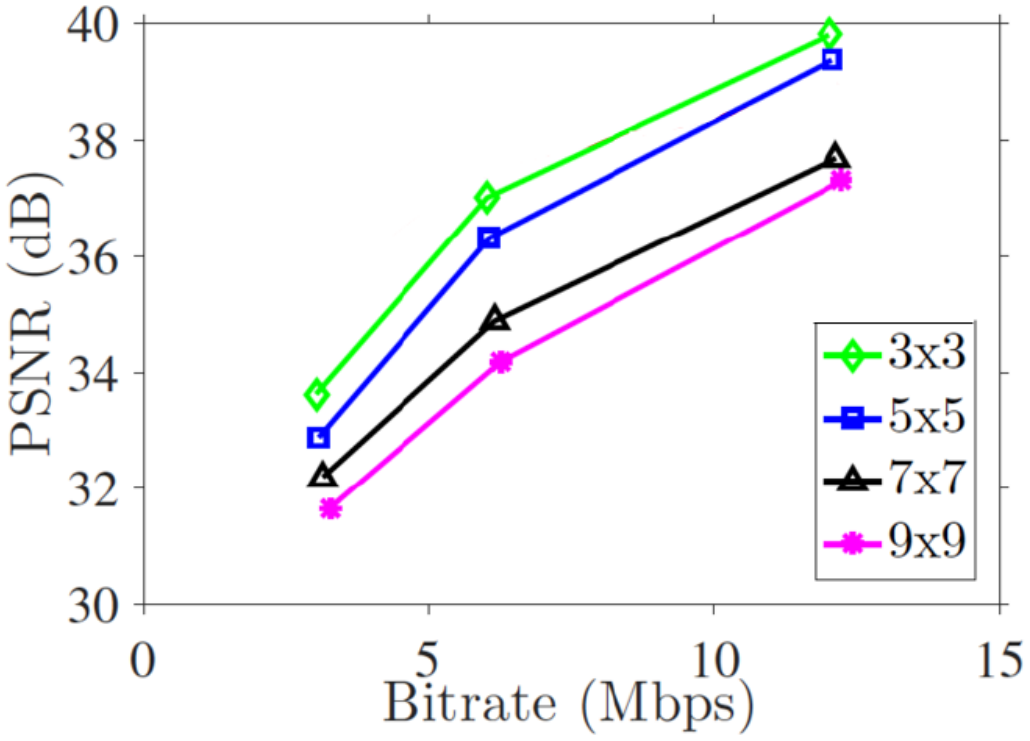
  given past bandwidth measurements and FoV locations,
  choose which encoding rate to download

- Full flexibility compared with monolithic encoding, but the bandwidth savings and QoE gains depend on:
  1. The compression overhead due to tiling
  2. The accuracy of the FoV prediction

Compression overhead due to tiling

UNIVERSITÉ CÔTE D'AZUR

- Higher number of tiles → Lower coding efficiency

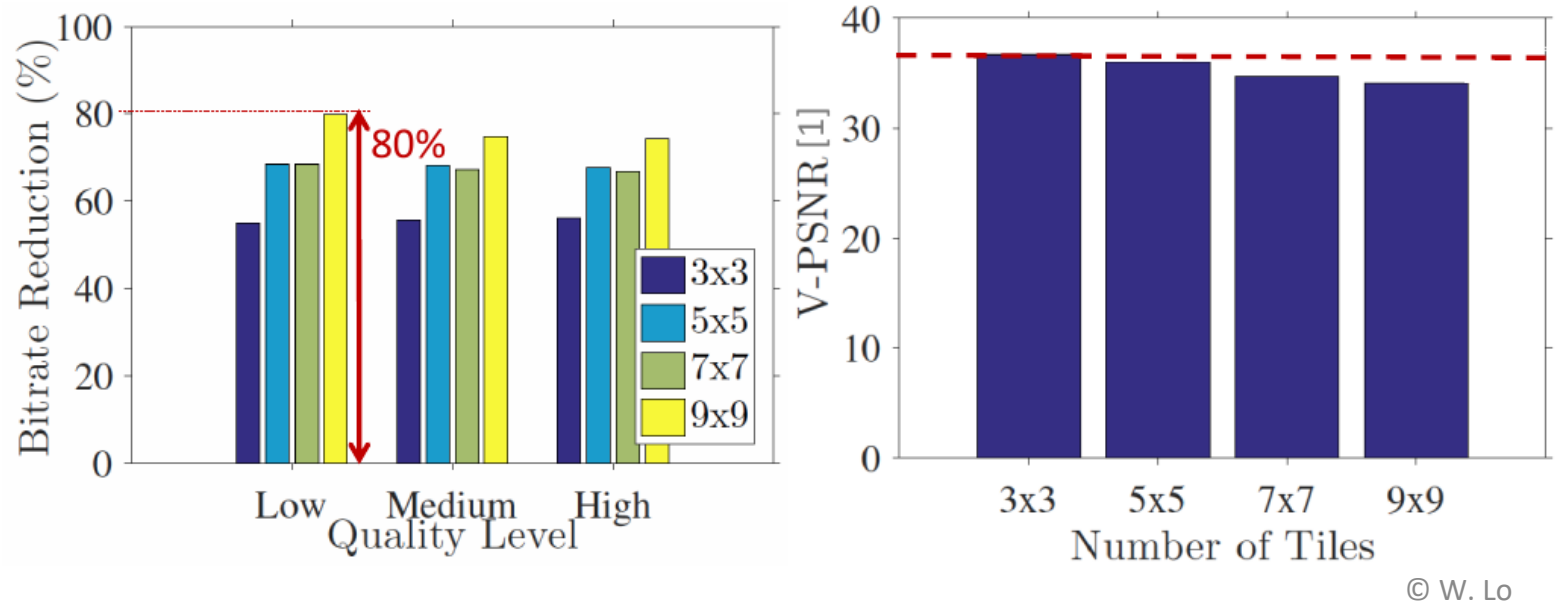


© W. Lo

W. Lo, C. Fan, S. Yen and C. Hsu, "Performance measurements of 360° video streaming to head-mounted displays over live 4G cellular networks," 2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS), 2017, pp. 205-210, doi: 10.1109/APNOMS.2017.8094203.

Can we still gain with tiling?

- Yes! Considering protocol overhead and lower compression rate due to tiling: tile-based streaming saves bandwidth and preserves quality



© W. Lo

→ Trade-off between quality in FoV, consumed bandwidth and storage

W. Lo, C. Fan, S. Yen and C. Hsu, "Performance measurements of 360° video streaming to head-mounted displays over live 4G cellular networks," 2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS), 2017, pp. 205-210, doi: 10.1109/APNOMS.2017.8094203.

Network- and User-adaptive quality control

UNIVERSITÉ CÔTE D'AZUR

- Optimization problem: **for every segment, for every tile:**

  given past bandwidth measurements and FoV locations,
  choose which encoding rate to download
  So that QoE is maximized

$$\max_{\{x_{ijl}\}} \sum_{i=1}^{M} \sum_{j=j_t}^{j_t+K} q_l p_i(j) x_{ijl}, \quad \text{s.t.} \tag{1a}$$

$$x_{ijl} = 0, \quad \forall i,j,l : buf_i(t) \geq B_{max} \text{ or } j < j_{t_i} \tag{1b}$$

$$buf_i(t) - \Delta_{Dl} + \sum_j \sum_l x_{ijl} \Delta_{Dl} \geq B_{min}, \quad \forall i \in \mathcal{M} \tag{1c}$$

$$\sum_{i,j,l} x_{ijl} s_{ijl} \leq C_t \Delta_{Dl}, \quad \sum_l x_{ijl} \leq 1, \quad \forall i \in \mathcal{M}, \forall j \in [j_t, j_t + K] \tag{1d}$$

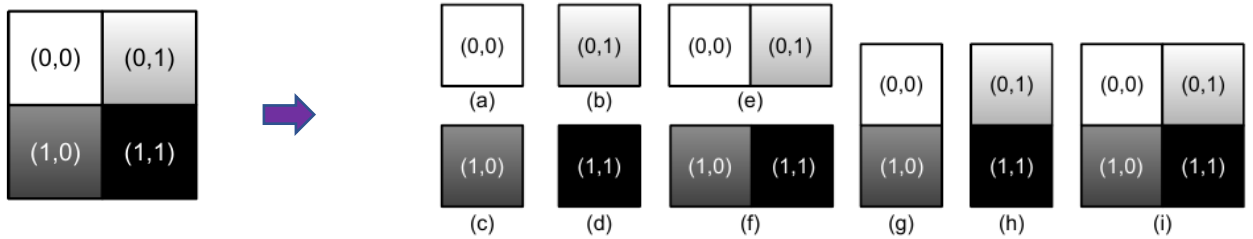
$$\sum_l x_{ijl} \leq \sum_l x_{i(j-1)l}, \quad \forall i \in \mathcal{M}, \forall j \in [j_t + 1, j_t + K] \tag{1e}$$

- The gains depend on:
  1. The compression overhead due to tiling
  2. The accuracy of the FoV prediction

- Approaches to optimize tile-based streaming:
  - Modulate the tiling pattern based on viewing statistics and QoE
  - FoV prediction as input:
    - Simple prediction over short horizon = short buffer (vulnerable to stalls)
    - Longer-term prediction and tile-viewing probability
  - End-to-end FoV prediction and quality control with Reinforcement Learning
  - Enabling quality corrections with replacements

Modulation of the tiling pattern

- Choose which subtiles to use:



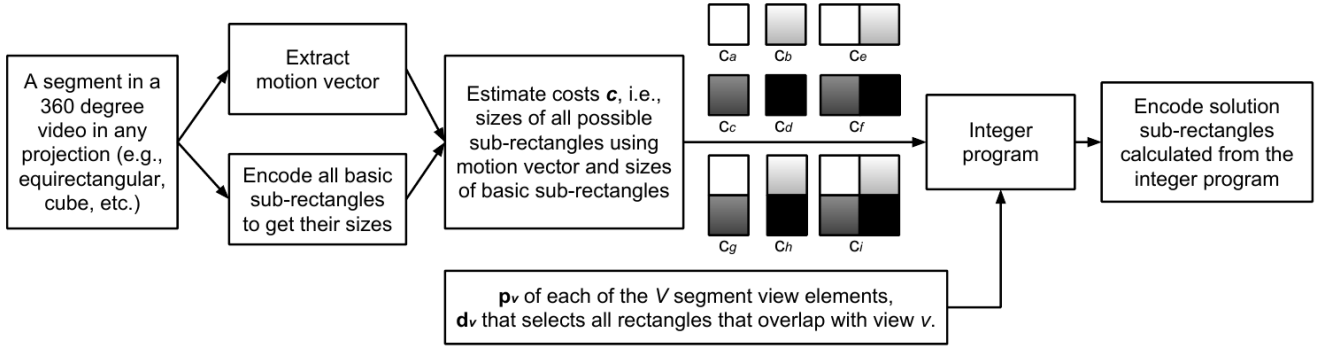
- So as to minimize storage cost and downloading cost:

$$\begin{aligned} \text{maximize:} \quad & (-c^{(stor)} - \alpha c^{(view)})^{\top} x \\ \text{subject to:} \quad & Ax = \mathbf{1} \\ & x_i \in \{0,1\} \quad \forall i \end{aligned}$$

A=

Position	a	b	c	d	e	f	g	h	i
(0,0)	1	0	0	0	1	0	1	0	1
(0,1)	0	1	0	0	1	0	0	1	1
(1,0)	0	0	1	0	0	1	1	0	1
(1,1)	0	0	0	1	0	1	0	1	1

- Estimate costs:



- Savings in downloaded data: 44% wrt homogeneous tiling

Enabling corrections of tile quality with replacements

UNIVERSITÉ CÔTE D'AZUR

- To tackle the high variability of visual region prediction and the unpredictable network fluctuations, employ 2 features of HTTP/2: *stream termination* and *stream priority*.
- Schedule tiles of K segments in specific order to download
- Given the head position and bandwidth at the previous segment, generate the optimal transmission sequence and:
  - Update existing transmission sequence by adding, removing, and changing order of the tile transmissions.
  - If the measured bandwidth has significantly deviated from the predicted value, terminate all active tile transmissions and prepare for rescheduling.

M. Xiao, C. Zhou, V. Swaminathan, Y. Liu and S. Chen, "BAS-360°: Exploring Spatial and Temporal Adaptability in 360-degree Videos over HTTP/2," IEEE INFOCOM 2018.

Joint FoV prediction and quality control with Reinforcement Learning

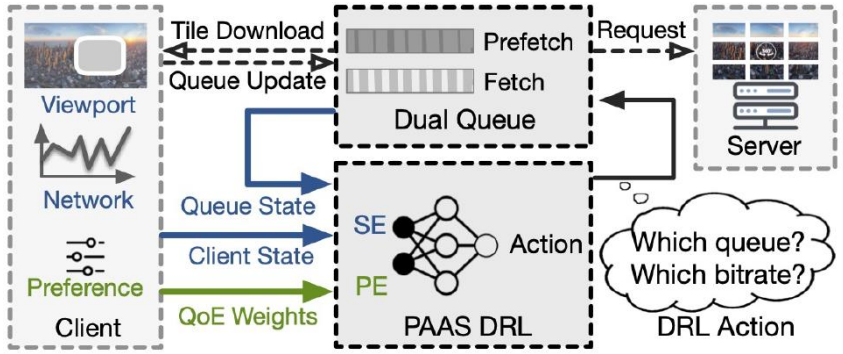


Figure 1: General Overview of PAAS

© C. Wu

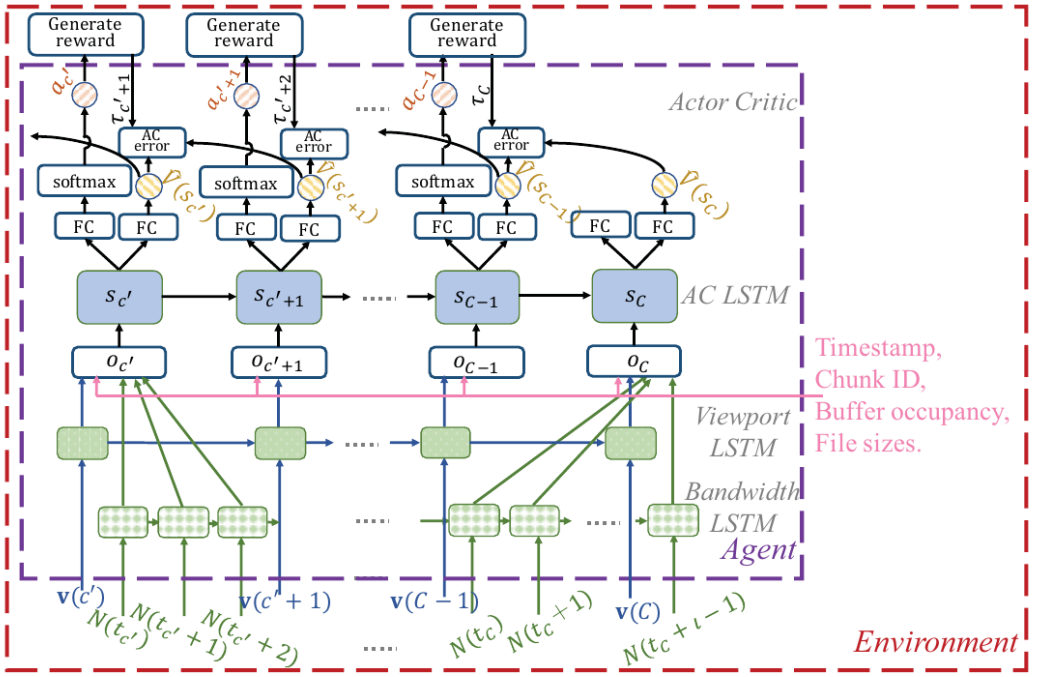


Fig. 2: The structure of the DRL-based model in the DRL360 streaming system.

© Y. Zhang

Y. Zhang, P. Zhao, K. Bian, Y. Liu, L. Song and X. Li, "DRL360: 360-degree Video Streaming with Deep Reinforcement Learning," IEEE INFOCOM 2019.
Chenglei Wu, Zhi Wang, and Lifeng Sun. 2021. PAAS: a preference-aware deep reinforcement learning approach for 360° video streaming. ACM NOSSDAV 2021.

Approaches to optimize tile-based streaming

- Modulate the tiling pattern based on viewing statistics and QoE
- Enabling quality corrections with replacements
- End-to-end FoV prediction and quality control with Reinforcement Learning
- FoV prediction as input:
  - Simple prediction over short horizon = short buffer (vulnerable to stalls)
  - Longer-term prediction and tile-viewing probability

Outline

UNIVERSITÉ CÔTE D'AZUR

- Streaming 360° videos: a multi-disciplinary problem

1. How to project, how to compress spherical content?
2. How to stream high-quality spherical content?
3. How to predict human head motion?
4. How to define Quality of Experience for spherical content?

- New interdisciplinary levers for VR streaming
- Conclusions

Saliency estimation and FoV prediction

- Saliency: quality that makes part of an image or video stand out and capture viewers' attention
- Saliency estimation and FoV prediction at the core of immersive QoE modeling and streaming optimization
- 360° saliency estimation can inform projection, compression, QoE estimation
- FoV prediction of the current user can be informed by past trajectories (of current and other users) and by the estimated saliency (content)
- Machine Learning is key in the models, particularly for information fusion
- Examples of 360° saliency models:
  - SalGAN360 [2], V-BMS360 [3], PanoSalNet [1]

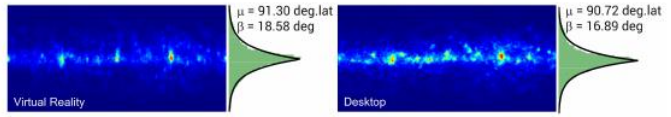


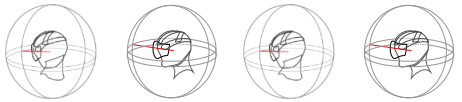
Fig. 3. Average saliency maps computed with all the scenes for both the *VR* (left) and the *desktop* (right) conditions. These average maps demonstrate an "equator bias" that is well-described by a Laplacian fit modeling the probability of a user fixating on an object at a specific latitude.

©V. Sitzmann

[1] A. Nguyen, Z. Yan, K. Nahrstedt, Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction, in ACM Multimedia 2018.
[2] F. Chao, L. Zhang, W. Hamidouche, O. Deforges, SalGAN360: Visual saliency prediction on 360 degree images with Generative Adversarial Networks, in ICME Workshops 2018.
[3] P. Lebreton, S. Fremerey, A. Raake, V-BMS360: A video extention to the BMS360 image saliency model, in ICME Workshops 2018.

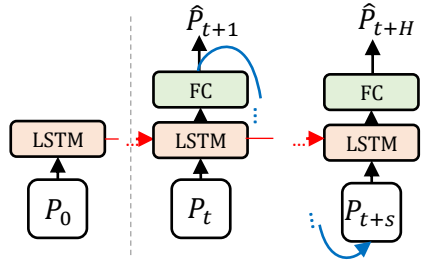
Predicting future FoV

- Offline (from saliency estimates) or online. In the online case, different possible assumptions:
- Predict only based on the current user's past:
  - Simple: linear prediction [1], dead reackoning
  - Deep Learning-based [2,3]
  - Generalizes over users and videos
- Predict based on the other trajectories of the current video:
  - k-NN and clustering approaches [4,5]
  - Does not generalize over videos, requires to record trajectories for every video



Past head positions trajectory

Deep Position-Only Baseline

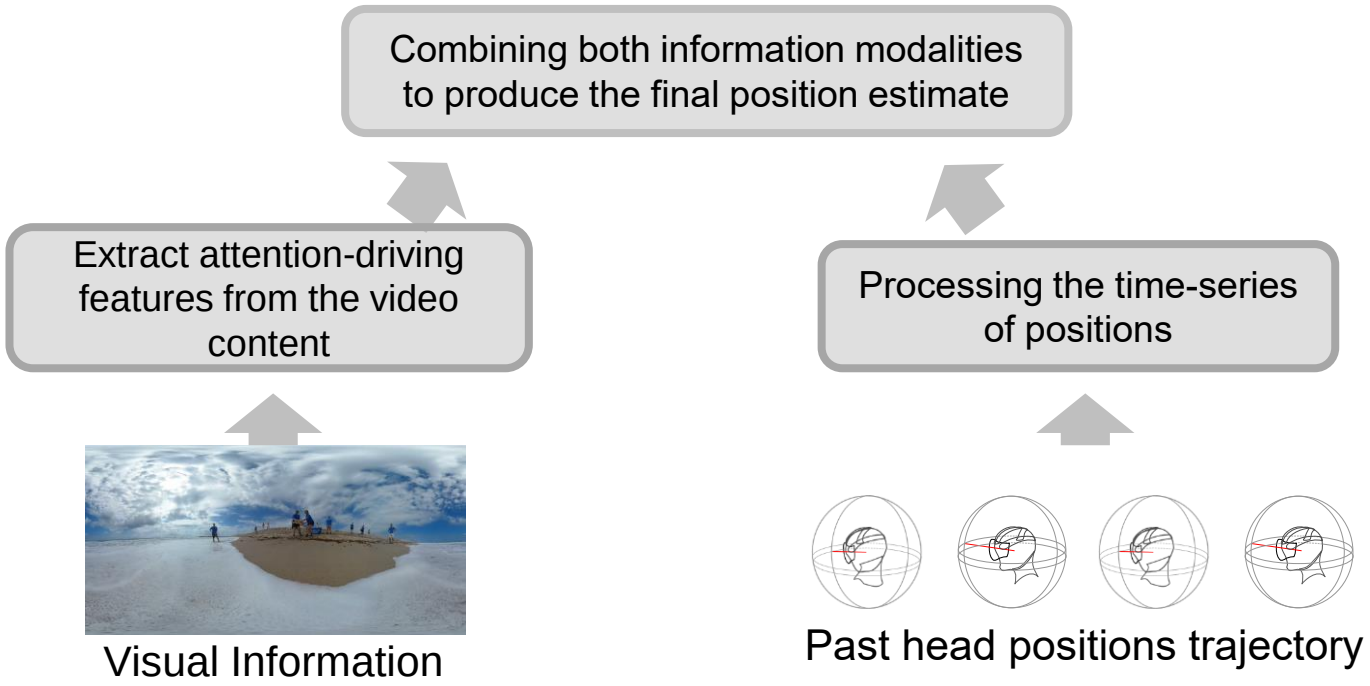


[1] Y. Bao, H. Wu, T. Zhang, A. A. Ramli, X. Liu, Shooting a moving target: Motion-prediction-based transmission for 360-degree videos, IEEE Big Data 2016.
[2] M. F. Romero Rondon, L. Sassatelli, R. Aparicio-Pardo and F. Precioso, "TRACK: A New Method from a Re-examination of Deep Architectures for Head Motion Prediction in 360-degree Videos," in IEEE TPAMI 2021.
[3] F.-Y. Chao, C. Ozcinar, and A. Smolic, "Transformer-based Long-Term Viewport Prediction in 360◦ Video: Scanpath is All You Need," IEEE MMSP 2021.
[4] Y. Ban, L. Xie, Z. Xu, X. Zhang,et al., "Cub360: Exploiting cross-users behaviors for viewport prediction in 360 video adaptive streaming," IEEE ICME 2018.
[5] S. Petrangeli, G. Simon, and V. Swaminathan, "Trajectory-based viewport prediction for 360◦ virtual reality videos," in IEEE AIVR 2018.

Predicting future FoV

- Predict based on current past and video content:
  - Generalizes over users and videos, does not require to collect per-video statistics

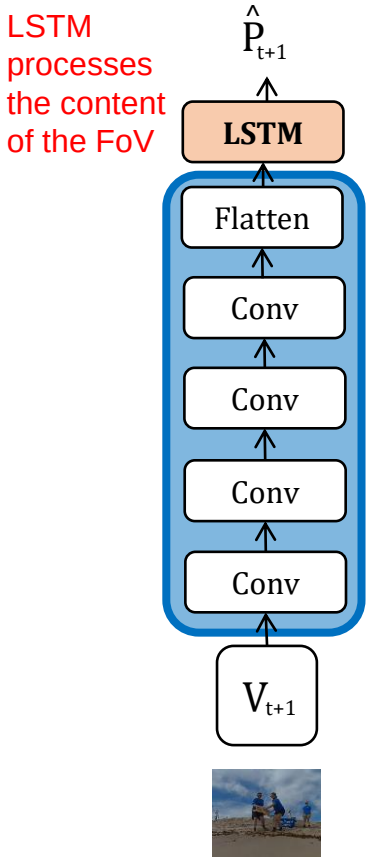
Reference	Prediction horizon
PAMI18	30 ms
CVPR18	1 s
MM18	2.5 s
ChinaCom18	1 s
NOSSDAV17	1 s



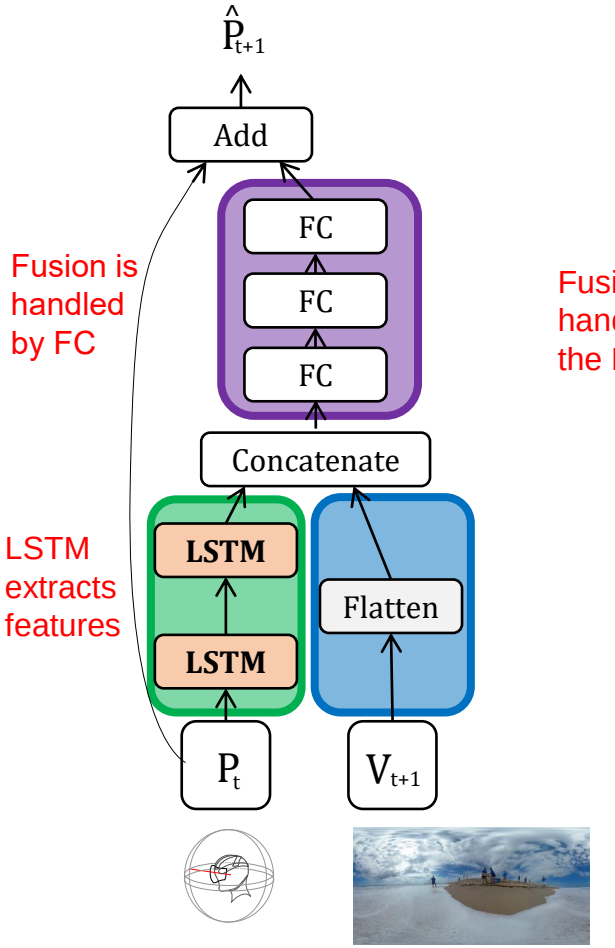
[M. Xu, et al., "Predicting head movement in panoramic video: A deep reinforcement learning approach," IEEE Trans. on PAMI, 2018.]
[Y. Xu, et al., "Gaze prediction in dynamic 360° immersive videos," in IEEE CVPR, 2018]
[A. Nguyen, et al., "Your attention is unique: Detecting 360° video saliency in head-mounted display for head movement prediction," in ACM Int. Conf. on Multimedia, 2018]
[Y. Li, et al., "Two-layer FoV prediction model for viewport dependent streaming of 360° videos," in EAI Int. Conf. on Communications and Networking (ChinaCom), 2018.]
[C.-L. Fan, et al., "Fixation prediction for 360 video streaming in head-mounted virtual reality," in ACM NOSSDAV, 2017]

Architectural choices for fusing modalities

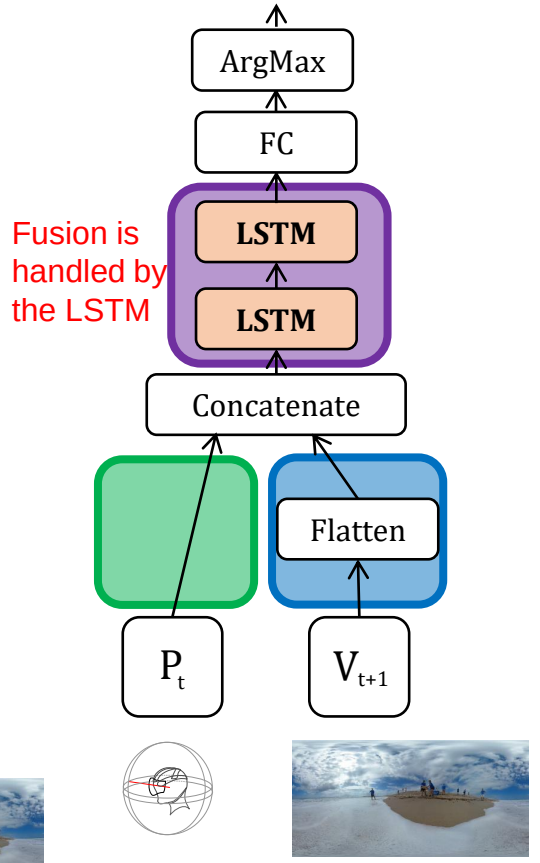
Reference	Prediction horizon	LSTM Performs Fusion?
PAMI18	30 ms	
CVPR18	1 s	Before Fusion
MM18	2.5 s	After Fusion
ChinaCom18	1 s	After Fusion
NOSSDAV17	1 s	After Fusion



PAMI18



CVPR18



MM18
ChinaCom18
NOSSDAV17

Taxonomy based on choice for fusion

UNIVERSITÉ CÔTE D'AZUR

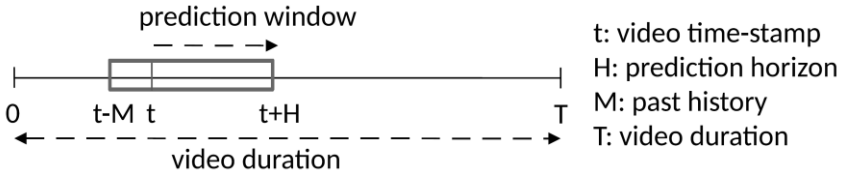
Reference	Objective	Prediction horizon	Dataset	Inputs	LSTM before/after fusion
PAMI18 [3]	Head coordinates	30 ms	76 videos, 58 users	Frame cropped to FoV	N/A no fusion
IC3D17 [4]	Head coordinates	2 s	16 videos, 61 users	Pre-trained sal. in FoV	N/A (no fusion, no LSTM)
ICME18 [5]	Tiles in FoV	6 s	18 videos, 48 users	Position history, users' distribution	N/A (no LSTM)
CVPR18 [6]	Gaze coordinates	1 s	208 videos, 30+ users	Video frame, position history as coordinates	Before
MM18 [7]	Tiles in FoV	2.5 s	NOSSDAV17's dataset with custom pre-processing	Pre-trained sal., mask of positions	After
ChinaCom18 [8]	Tiles in FoV	1 s	NOSSDAV17's dataset	Pre-trained sal., FoV tile history	After
NOSSDAV17 [9]	Tiles in FoV	1 s	10 videos, 25 users	Pre-trained sal., FoV position or tile history	After

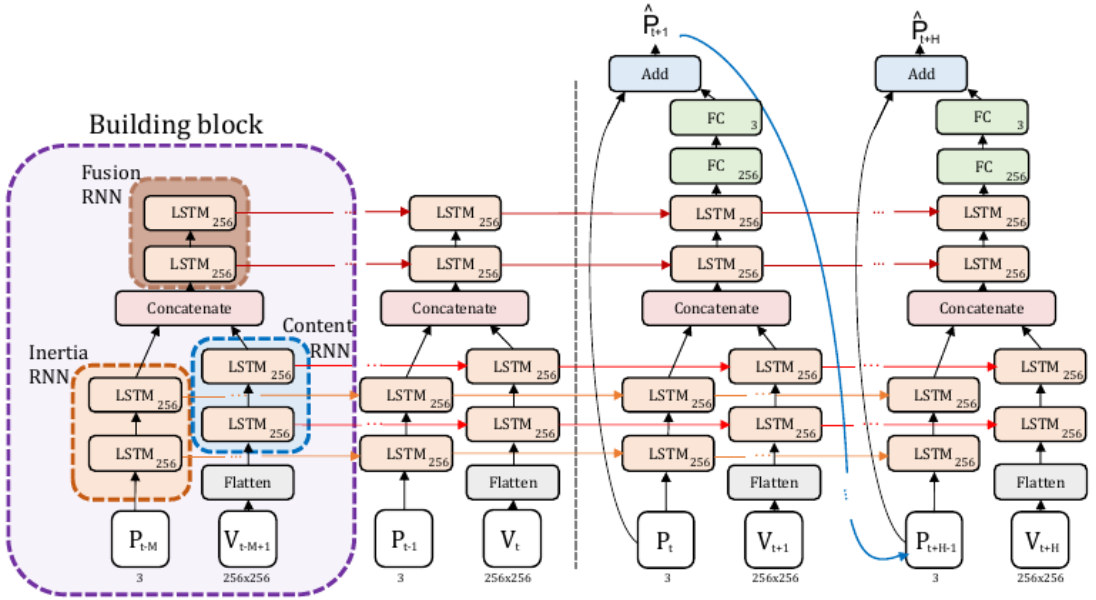
TRACK

- Predict trajectory $[\mathbf{P}_{t+1},\ldots\mathbf{P}_{t+H}]$

- Multi-modal inputs:
  - Past positions of the current user
  - 360° video content

- Design of TRACK [1] :
  - Information fusion based on Structural RNN
  - Best performance on all datasets, gains up to 20%





[1] M. Romero, L. Sassatelli, R. Aparicio-Pardo, and F. Precioso. TRACK : A New Method from a Re-examination of Deep Architectures for Head Motion Prediction in 360-degree Videos. IEEE Transactions on Pattern Analysis and Machine Intelligence, Mar. 2021.

Outline

UNIVERSITÉ CÔTE D'AZUR

- Streaming 360° videos: a multi-disciplinary problem

1. How to project, how to compress spherical content?
2. How to stream high-quality spherical content?
3. How to predict human head motion?
4. How to define Quality of Experience for spherical content?

- New interdisciplinary levers for VR streaming
- Conclusions

What is immersive QoE

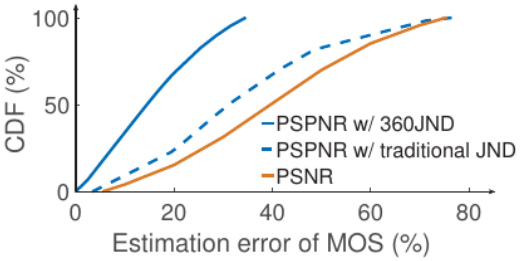
UNIVERSITÉ CÔTE D'AZUR

- Definition [1]:
  - Quality of Experience (QoE) is the degree of delight or annoyance of the user (persona) of an application or service. It results from the fulfillment of their expectations wrt the utility and/or enjoyment of the application or service in the light of user's personality and current state (context).
- But how is QoE defined from QoS or application-level metrics?
- For video streaming, important dimensions are: (ITU-T Rec. P.1203)
  - Visual quality, rebuffering time or frequency, temporal quality variations, startup delay
- For 360° video streaming, there are re-defined and new dimensions:
  - Visual quality, cybersickness, immersion, presence
  - And varies with content itself, user attention and experiment duration (more than with regular video)

[1] Kjell Brunnstrom, Sergio Ariel Beker, Katrien de Moor, Ann Dooms, Sebastian Egger, et al.. Qualinet White Paper on Definitions of Quality of Experience. 2013.

Measuring QoE: subjective methods

- Based Mean Opinion Score (MOS)
  - A new ACR methodology for 360° not requiring to take off the HMD [1]
- For 360° video, static image quality is not the only component impacting QoE:
  - [2] studies how perceived quality and simulator sickness are impacted by motion-to-high-quality latency:
    - With motion-to high-quality latency: MOS decreases, sickness increases
    - With session index: sickness increases
    - With camera motion: sickness increases
  - [3] builds on a new quality model for 360° videos by revisiting user's sensitivity to quality distortion relatively to the viewpoint-moving speed, the difference of depth-of-field (DoF) and the change of luminance.



© Y. Guan

[1] A. Singla, S. Fremerey, W. Robitza, P. Lebreton, A. Raake, Comparison of subjective quality evaluation for HEVC encoded omnidirectional videos at different bit-rates for UHD and FHD resolution. ACM Multimedia workshops 2017.
[2] A. Singla, S. Göring, A. Raake, et al., T. Buchholz, Subjective quality evaluation of tile-based streaming for omnidirectional videos. ACM MMSys 2019.
[3] Y. Guan, C. Zheng, X. Zhang, et al.. 2019. Pano: optimizing 360° video streaming with a better understanding of quality perception. ACM SIGCOMM 2019.

More dimensions to immersive QoE

UNIVERSITÉ CÔTE D'AZUR

- Impact of sound
- Multi-sensory environments: can include haptic feedback (wind), smells – this is mulsemedia
  - Can improve immersion and enrich experience [1]
  - multi-sensorial extension of the MPEG DASH: DASHMS
- Presence: psychological experience of 'being there' within a VR environment [2]
  - Presence is fundamentally a construct of the user rather than of the technology per se.
- Immersion: sensorial vividness of a virtual environment and its ability to replace real world stimuli with those from VR [2]
  - Immersion is closely tied to technological specifications.
- Creation of metrics based on objective physiological data
  - Electro-dermal resistance, heart rate and more [3]

[1] T. Bi, R. Lyons, G. Fox and G. -M. Muntean, "Improving Student Learning Satisfaction by Using an Innovative DASH-based Multiple Sensorial Media Delivery Solution," in IEEE Transactions on Multimedia, 2020.
[2] C. Jicol et al., "Effects of Emotion and Agency on Presence in Virtual Reality," in ACM CHI 2021.
[3] P. Arnau-González, T. Althobaiti, S. Katsigiannis, N. Ramzan, Perceptual video quality evaluation by means of physiological signals, IEEE QoMEX 2017.

Objective QoE metrics

UNIVERSITÉ CÔTE D'AZUR

- Easiest method to objectively measure the QoE of a 360° image is to directly use a classic 2D metric (PSNR, SSIM, etc.)
- But they need to be adapted to consider the geometric distortion sphere → plane:
  - S-PSNR, WS-PSNR, S-SSIM, WS-SSIM
- And they also should consider the FoV and content saliency:
  - Content Preference PSNR (CP-PSNR) and CP-SSIM [1]
  - Other more refined: PVQ (spatial resolution and QP, moments of luminance) [2]
- Machine Learning-based models:
  - e.g., Viewport-based CNN (V-CNN) [3]: a CNN predicts QoE for different FoVs, and a spherical CNN predicts possible FoVs and their weights in the expected QoE.

[1] M. Xu, C. Li, Z. Chen, Z. Wang, Z. Guan, Assessing visual quality of omnidirectional videos, IEEE Transactions on Circuits and Systems for Video Technology 2018.
[2] W. Zou, F. Yang, S. Wan, Perceptual video quality metric for compression artefacts: from two-dimensional to omnidirectional, IET Image Processing 2017.
[3] C. Li, M. Xu, L. Jiang, S. Zhang, X. Tao, Viewport proposal CNN for 360deg video quality assessment. IEEE CVPR 2019.

Comparison of the objective QoE metrics

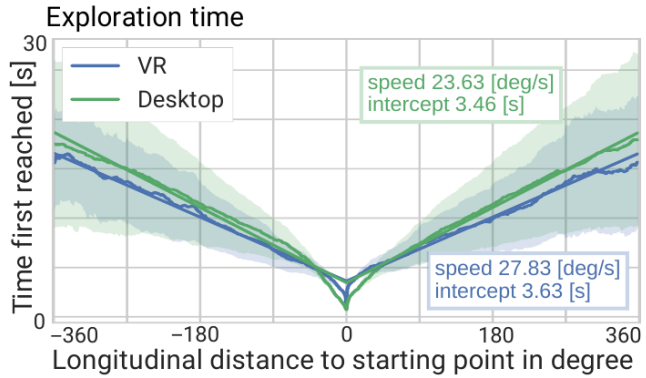
Table 5: Performance of the main presented objective QoE metrics. The table should be read horizontally: the metric in each row is compared to one for each column. Metrics whose rows have more green cells are more closely correlated with subjective MOS

	PSNR	SSIM	MS-SSIM	VIFP	WS-PSNR	S-PSNR	CPP-PSNR
PSNR		Worse	Worse	Worse	Worse	Worse	Worse
SSIM	Better		Similar	Worse	Better	Better	Slightly better
MS-SSIM	Better	Similar		Worse	Better	Better	Slightly better
VIFP	Better	Better	Better		Better	Better	Better
WS-PSNR	Better	Worse	Worse	Worse		Slightly worse	Slightly worse
S-PSNR	Better	Worse	Worse	Worse	Slightly better		Slightly worse
CPP-PSNR	Better	Slightly worse	Slightly worse	Worse	Slightly better	Slightly better	

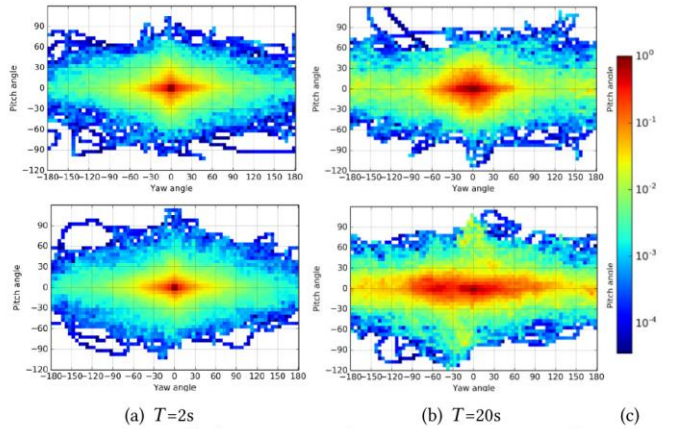
F. Chiariotti, ‘‘A survey on 360-degree video: Coding, quality of experience and streaming,’’ Comput. Commun., vol. 177, pp. 133–155, Sep. 2021.

Challenges in QoE assessment

- Conditions for testing QoE metrics in 360° videos specified [1,2]
  - Evil viewport problem: FoV with visible seams must be considered separately
  - Do not use too short videos [3]
  - Strong dependence of the correlation between objective metrics and MOS on the actual content of the images
- Need to consider how do people explore in VR:
  - Exploratory phase of about 20s [4]
  - Different scene types yield different users' behaviors [5]



©V. Sitzmann



(a) T=2s (b) T=20s (c)

Figure 1: Example heat-maps for two categories of movies: Rides (top) and Moving focus (bottom).

© M. Almquist

[1] E. Alshina, J. Boyce, A. Abbas, Y. Ye, JVET common test conditions and evaluation procedures for 360 degree video, Tech. Rep. G1030, JVET (Jul. 2017).
[2] P. Hanhart, Y. He, Y. Ye, J. Boyce, Z. Deng, L. Xu, 360-degree video quality evaluation, in: Picture Coding Symposium (PCS), IEEE, 2018, pp. 328–332.
[3] H. Huang, J. Chen, H. Xue, Y. Huang, T. Zhao, Time-variant visual attention in 360-degree video playback, IEEE HAVE 2018.
[4] V. Sitzmann, et al.. Saliency in VR: How Do People Explore Virtual Environments?. IEEE Trans. on Vis. and Comp. Graphics, April 2018.
[5] M. Almquist, et al.. The Prefetch Aggressiveness Tradeoff in 360 Video Streaming. In ACM MMSys 2018.

Outline

UNIVERSITÉ CÔTE D'AZUR

- Streaming 360° videos: a multi-disciplinary problem

1. How to project, how to compress spherical content?
2. How to stream high-quality spherical content?
3. How to predict human head motion?
4. How to define Quality of Experience for spherical content?

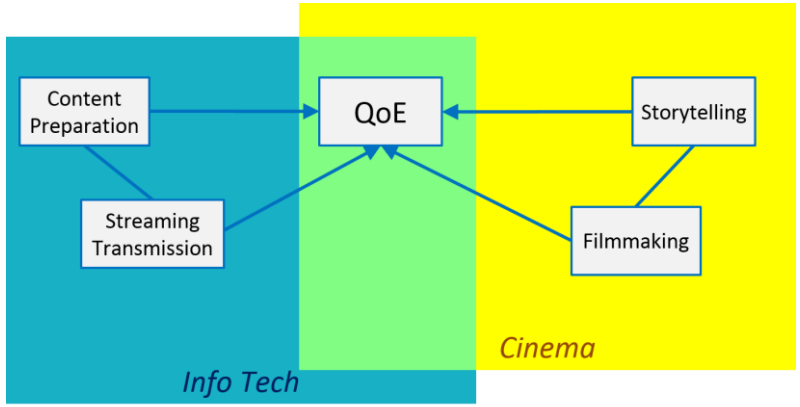
- New interdisciplinary levers for VR streaming
- Conclusions

Our general approach

UNIVERSITÉ CÔTE D'AZUR

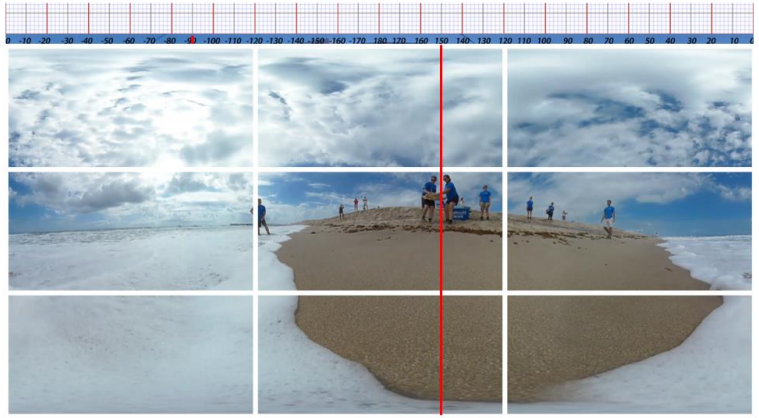
- **Research objective**:
  - Improve the quality of experience of 360° video streaming with new adaptation levers

- **Approach**:
  - We posit that the visual quality is **not** always the best dimension to impair the content to fit the available bandwidth.
  - Design a wider set of levers, beyond compression, to modulate the immersive content.
  - Build on interdisciplinarity: ex: HCI and filmmaking



Dynamic editing for 360° videos

- Snap-change to control field of view:
  - Re-position user in front of a pre-defined area, in a snap
  - Defined by the art director
  - Enables bandwidth saving **AND** serves the content's objective

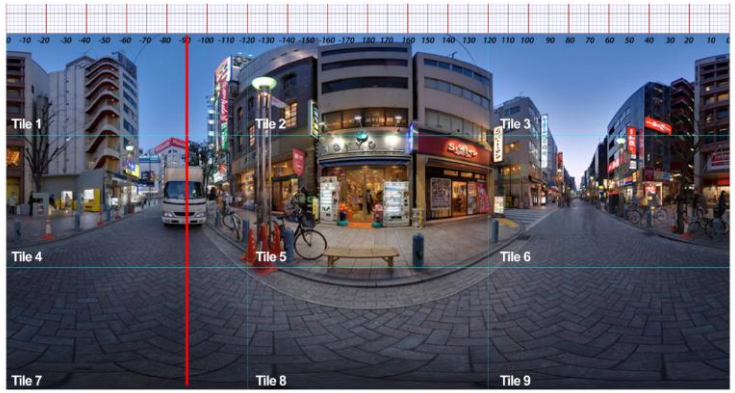


Identification of the Region of Interest: 140° at 6s

[13] B.E. Riecke, M. Von Der Heyde and H. Bulthoff. Visual Cues Can Be Sufficient for Triggering Automatic, Reflexlike Spatial Updating. ACM Trans. on Applied Perception 2005.

S. Dambra, G. Samela, L. Sassatelli, R. Pighetti, R. Aparicio-Pardo, and A.-M. Pinna-Déry. Film Editing: New Levers to Improve VR Streaming. ACM MMSys 2018.

Our 360° player: TOUCAN-VR



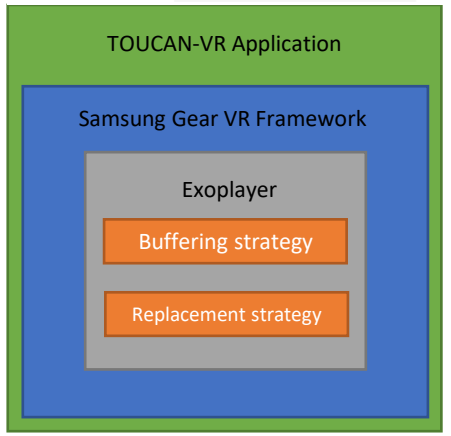
ACM reproducibility badge:
https://github.com/UCA4SVR/TOUCAN-VR

Identification of the region of interest

```
<?xml version="1.0"?>
<snapchange>
    <milliseconds>15000</milliseconds>
    <roiDegrees>-90</roiDegrees>
    <foVTile>1,2,4,5</foVTile>
</snapchange>
```

Description of a snap-cut :
xml file to download

App

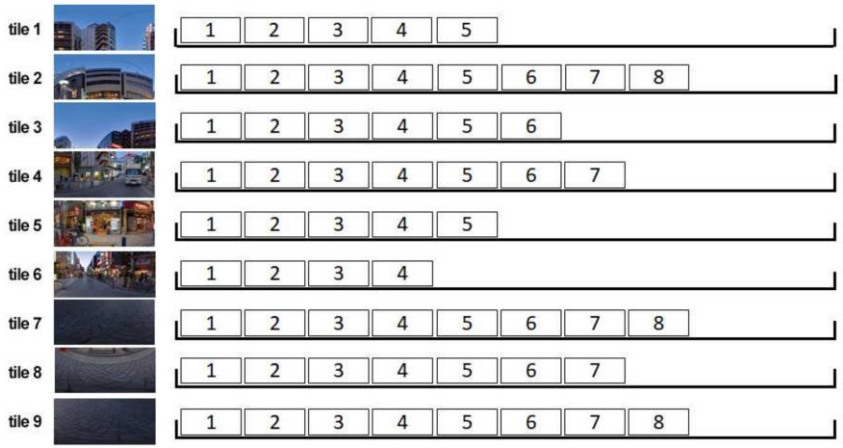


S. Dambra, G. Samela, L. Sassatelli, R. Pighetti, R. Aparicio-Pardo, and A.-M. Pinna-Déry. Film Editing: New Levers to Improve VR Streaming. ACM MMSys 2018.

Components of the streaming app

UNIVERSITÉ CÔTE D'AZUR

Buffering and quality selection



Qualities selected based on current FoV or next snap-change

→ Benefits from Dynamic editing

Replacements for responsiveness

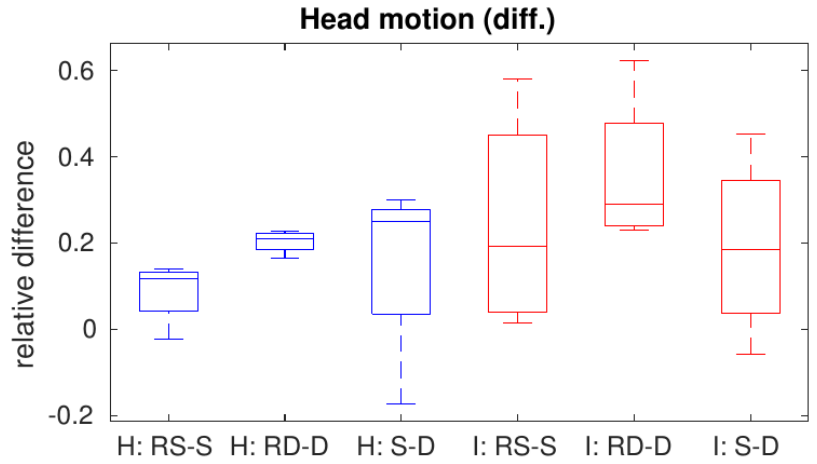
No replacements to make before a snap-change occurs

→ Benefits from Dynamic editing

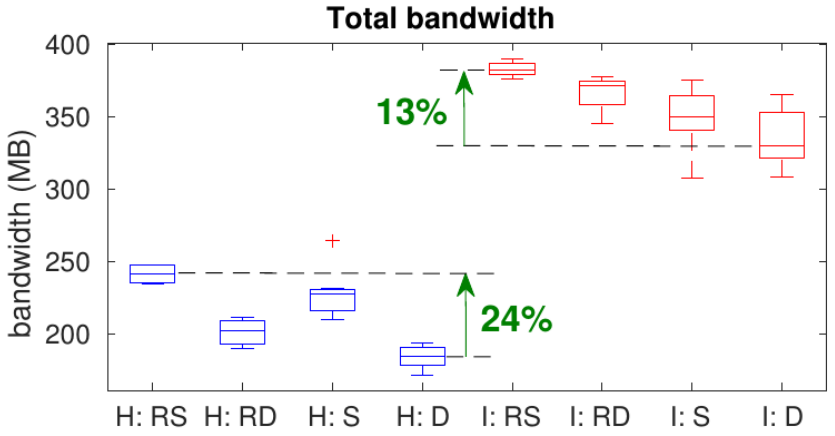
S. Dambra, G. Samela, L. Sassatelli, R. Pighetti, R. Aparicio-Pardo, and A.-M. Pinna-Déry. Film Editing: New Levers to Improve VR Streaming. ACM MMSys 2018.

Dynamic movie editing helps streaming

- Reduction of up to 30% in head motion speed with snap-changes



- Reduction of up to 24% in consumed data rate



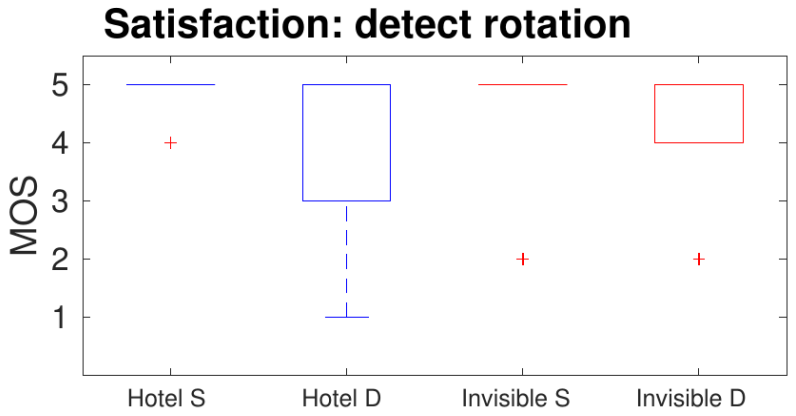
S. Dambra, G. Samela, L. Sassatelli, R. Pighetti, R. Aparicio-Pardo, and A.-M. Pinna-Déry. Film Editing: New Levers to Improve VR Streaming. ACM MMSys 2018.

AND dynamic editing improves the user's experience

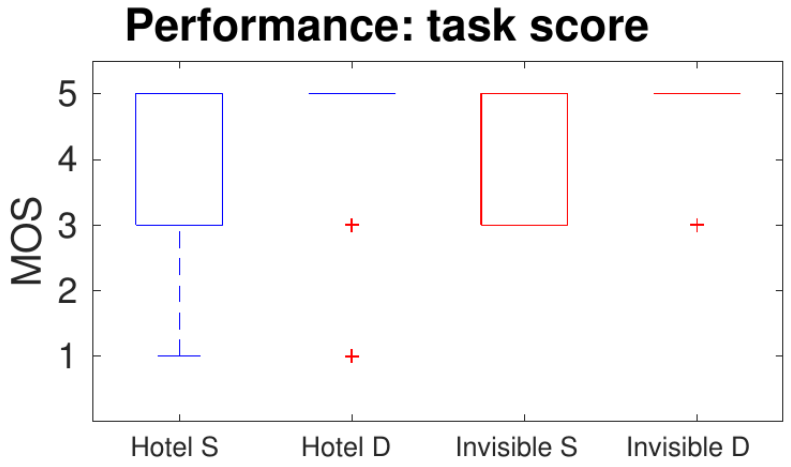
UNIVERSITÉ CÔTE D'AZUR

Awarded by the MPEG-DASH Industry Forum

- Snap-cuts go unnoticed



- The director can control what the user sees and remembers



S. Dambra, G. Samela, L. Sassatelli, R. Pighetti, R. Aparicio-Pardo, and A.-M. Pinna-Déry. Film Editing: New Levers to Improve VR Streaming. ACM MMSys 2018.

Design of Virtual Wall

- Preventing access to an angular sector
- Placed after exploration in Static focus and Rides scenes
- When the longitude of the user's position reaches the limit of the visible sector, the FoV refreshes in latitude only
- Do not affect latitude to keep balance

L. Sassatelli, M. Winckler, T. Fisichella, A. Dezarnaud, J. Lemaire, R. Aparicio and D. Trevisan. New Interactive Strategies for Virtual Reality Streaming in Degraded Context of Use. Elsevier Computers & Graphics 2020.

Outline

UNIVERSITÉ CÔTE D'AZUR

- Streaming 360° videos: a multi-disciplinary problem

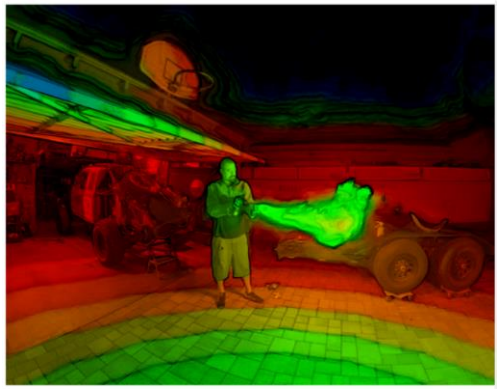
1. How to project, how to compress spherical content?
2. How to stream high-quality spherical content?
3. How to predict human head motion?
4. How to define Quality of Experience for spherical content?

- New interdisciplinary levers for VR streaming
- Conclusions

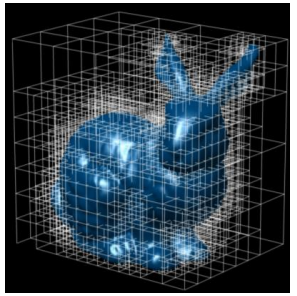
Conclusions

- Live and social VR [1]
  - Volumetric video formats: V-PCC, G-PCC

- Higher-resolution headsets (70 px/°, 115° FoV)
  - Foveated streaming
- Natural viewing, holographic videos: towards streamable light field [2,3]



- 6 DoF VR/AR content: PCC-DASH streaming [4,5]
  - 3D tiles, adapt based on viewpoint and depth
  - MPEG-I



[1] S. N.B. Gunkel, R. Hindriks, et al. VRComm: An end-to-end web system for real-time photorealistic social VR communication. ACM MMSys 2021.
[2] B. Wang, Q. Peng, Q., et al. User-dependent interactive light field video streaming system. Multimedia Tools and Applications 2021.
[3] M. Broxton, J. Flynn, et al. Immersive light field video with a layered mesh representation. ACM Trans. Graph. 2020.
[4] J. van der Hooft, T. Wauters, et al. Towards 6DoF HTTP Adaptive Streaming Through Point Cloud Compression. ACM Multimedia 2019.
[5] S. Subramanyam, I. Viola, et al.. User Centered Adaptive Streaming of Dynamic Point Clouds with Low Complexity Tiling. ACM Multimedia 2020.

Thank you!