EECE5512
Networked XR Systems

Lecture Outline for Today

- 360-Degree VR Video Streaming
  - Viewport
  - Viewport prediction
  - Viewport culling
  - Viewport adaptation

- Homework3

360-Degree or Panoramic Video

3-DoF on angular motion

360-Degree or Panoramic Video

- Central to many immersive applications (e.g., VR/AR)

Image credit: Oculus

Immersive Experience

$ Billion Market

http://blog.dsky.co/tag/head-tracking/

360-Degree or Panoramic Video

- Motion-to-photon latency – a unique metric of importance compared to regular videos
  - “*the lag between a user making a movement and the movement being displayed within the display*”
  - Should be in the order of a few 10s of milliseconds (<20ms)
- Other metrics that we discussed in case of regular videos still apply here (e.g., quality, stalls., etc.)

360-Degree or Panoramic Video

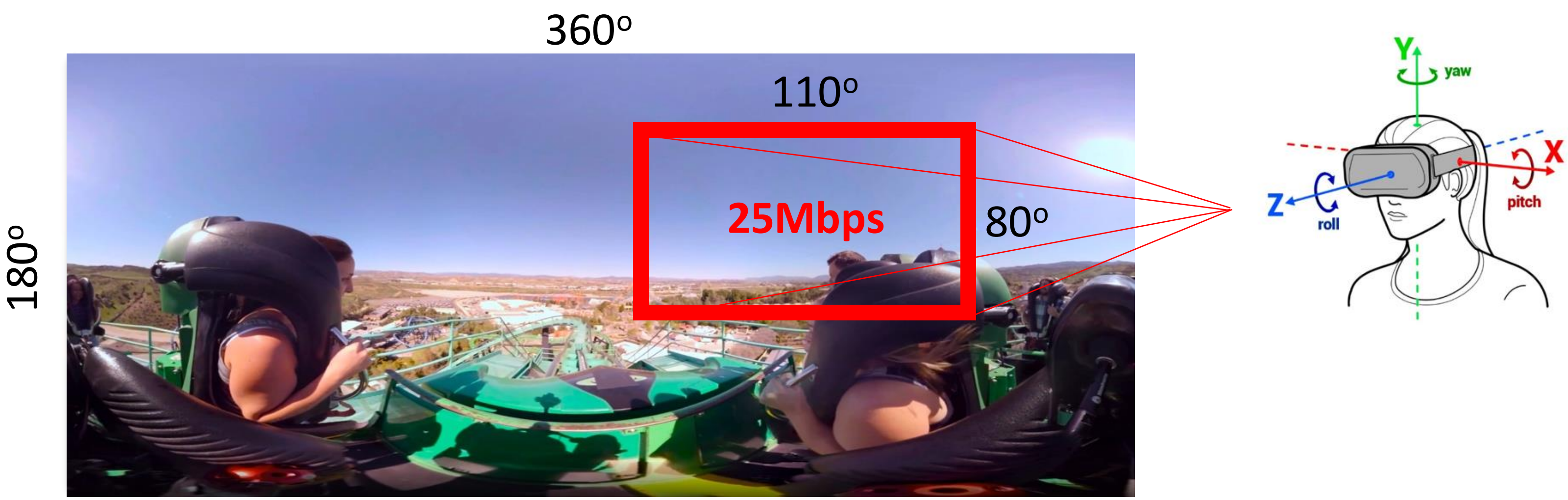
- How to compress 360-degree videos?
  - Often projected to a 2D equirectangular video and then adopt standard video codecs

360-Degree or Panoramic Video

- How streaming a 360-degree video is different from regular video?
  - Stream it like a regular video - problem?

Streaming Challenge

How streaming a 360-degree video is different from regular video?



200Mbps

Image from Rollercoaster video

360° videos require more bandwidth compared to regular videos for the same perceived quality

http://blog.dsky.co/tag/head-tracking/

Streaming Challenge

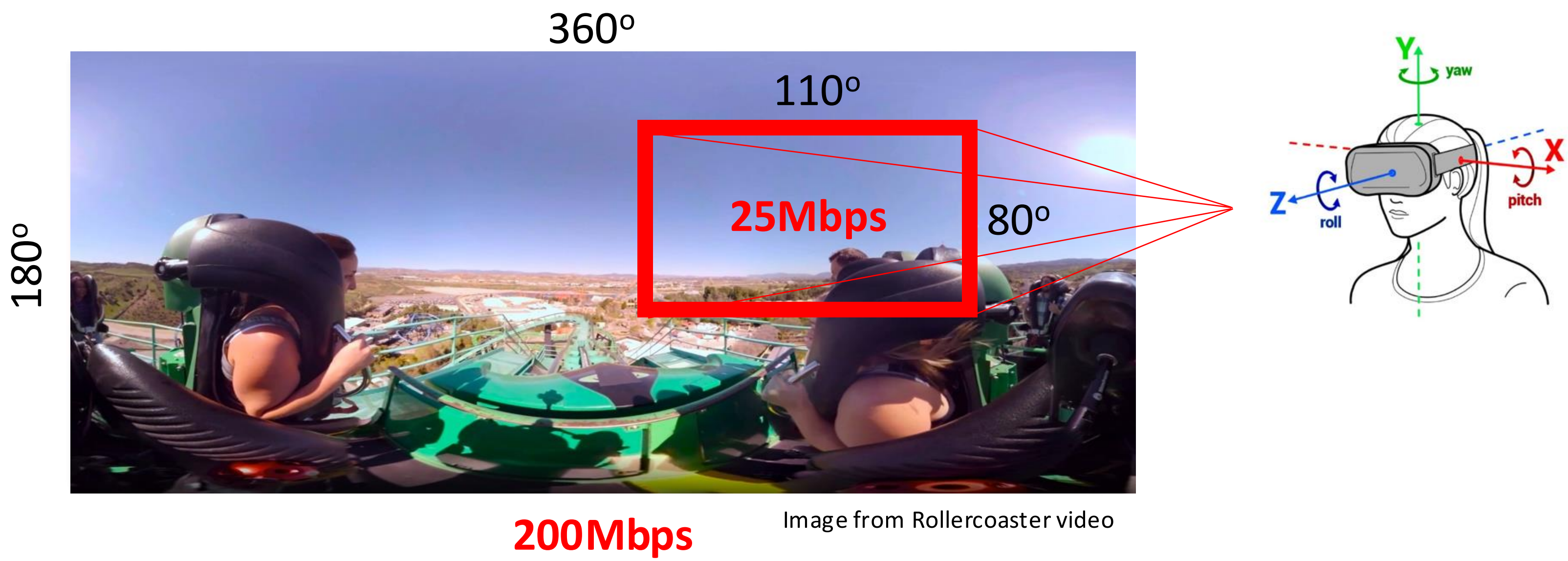


Image from Rollercoaster video

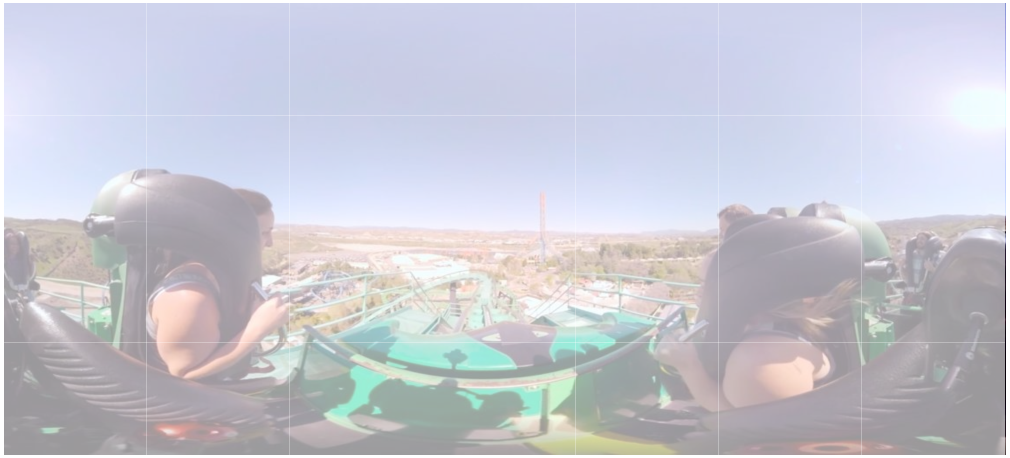
Can we just stream the viewport?

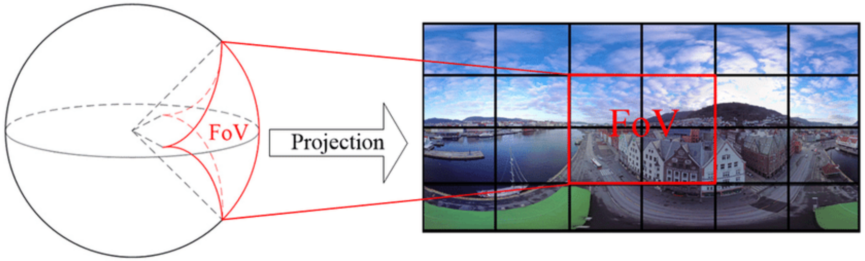
Viewport-Adaptive Streaming

- How to stream a viewport?
  - There can be many viewports i.e., different users may look different parts of the scene at different times during the video
  - Viewport is continuous stream of pixels – hard to identify viewport pixels and stream viewport directly

Viewport-Adaptive 360-Degree Streaming

- Tiled Streaming
  - Divide video into tiles (e.g., 192x192 pixels)

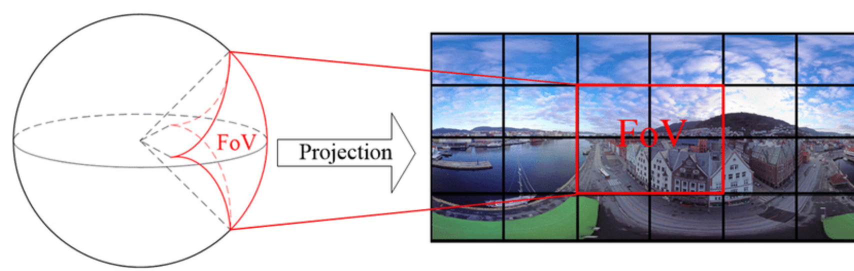




Flare [MobiCom'18], Rubiks [MobiSys'18], MOSAIC [IFIP Networking'19]
PANO [SIGCOMM'19], ClusTile [INFOCOM'19]

Viewport-Adaptive 360-Degree Streaming

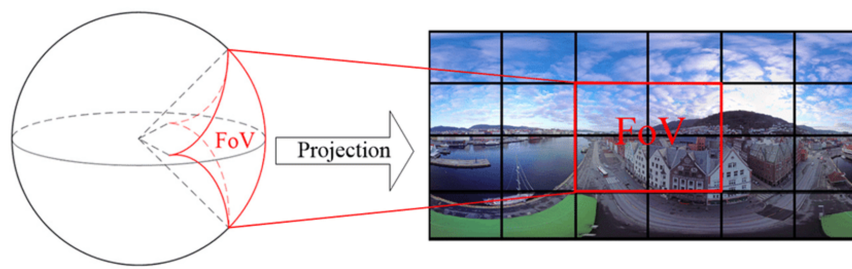
- Tiled streaming
  - Divide video into tiles (e.g., 192x192 pixels)
  - Predict viewport tiles based on head tracking and video saliency analysis
  - Stream only viewport specific tiles



Flare [MobiCom'18], Rubiks [MobiSys'18], MOSAIC [IFIP Networking'19]
PANO [SIGCOMM'19], ClusTile [INFOCOM'19]

Viewport-Adaptive 360-Degree Streaming

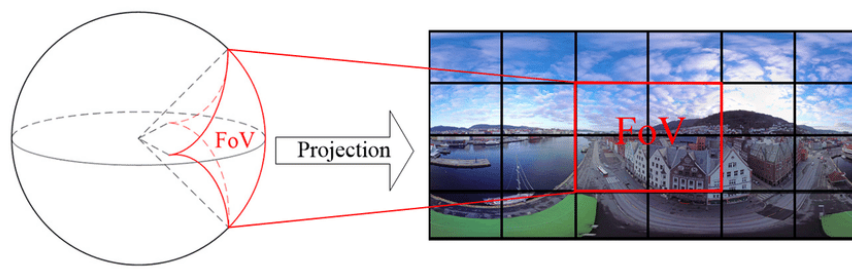
- Tiled streaming
  - Divide video into tiles (e.g., 192x192 pixels)
  - Predict viewport tiles based on head tracking and video saliency analysis
  - Stream only viewport specific tiles



Which tiles and how many tiles to stream?

Viewport-Adaptive 360-Degree Streaming

- Viewport prediction
  - Use video features and users' past history (i.e., head motion data) to predict where the user will look at in the near future
  - Need prediction models



Which tiles and how many tiles to stream?

Viewport-Adaptive 360-Degree Streaming

- Viewport prediction
  - Use video features and users' past history (i.e., head motion data) to predict where the user will look at in the near future
  - Need prediction models

Which tiles and how many tiles to stream?

User head motion ->

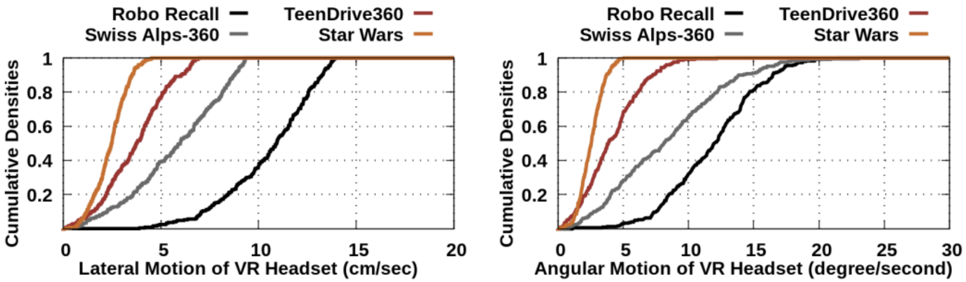


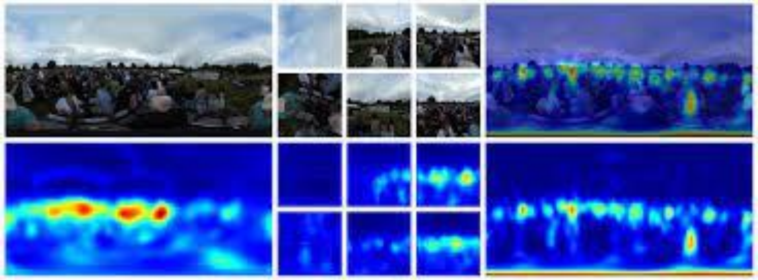
Figure 3: CDFs of VRH linear and angular speeds for VR applications.

Viewport-Adaptive 360-Degree Streaming

- Viewport prediction
  - Use video features and users' past history (i.e., head motion data) to predict where the user will look at in the near future
  - Need prediction models

Which tiles and how many tiles to stream?

Saliency features ->



Viewport-Adaptive 360-Degree Streaming

- Viewport prediction
  - Use video features and users' past history (i.e., head motion data) to predict where the user will look at in the near future
  - Need prediction models

Which tiles and how many tiles to stream?

Attention features ->

Viewport-Adaptive 360-Degree Streaming

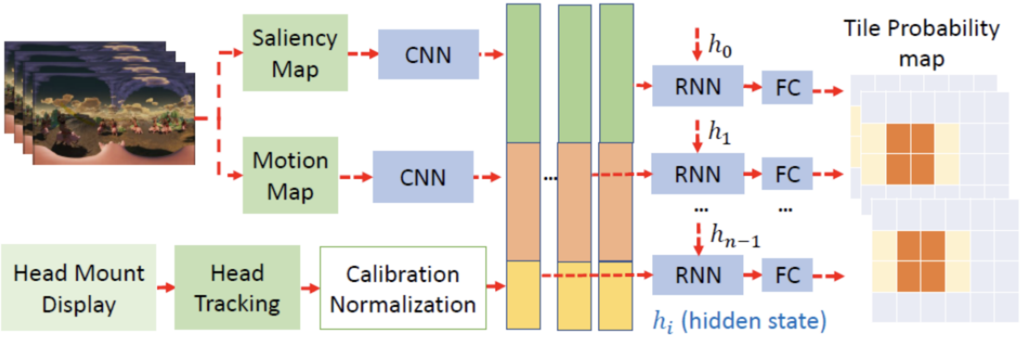
- Viewport prediction models
  - Simple ML Models (e.g., SVM)
    - Faster, less accurate
  - Neural Networks
    - Slower, slightly more accurate

Which tiles and how many tiles to stream?

Viewport-Adaptive 360-Degree Streaming

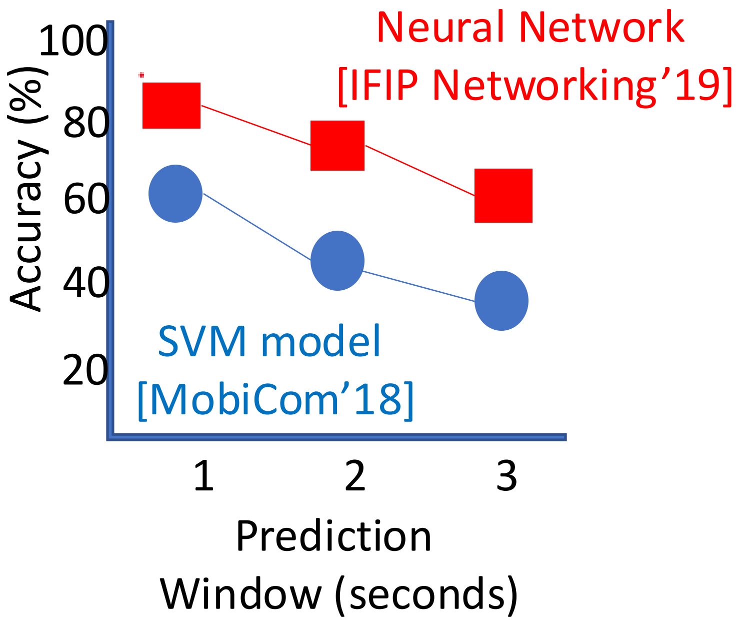
- Viewport prediction models
  - Simple ML Models (e.g., SVM)
    - Faster, less accurate
  - Neural Networks
    - Slower, slightly more accurate

Which tiles and how many tiles to stream?



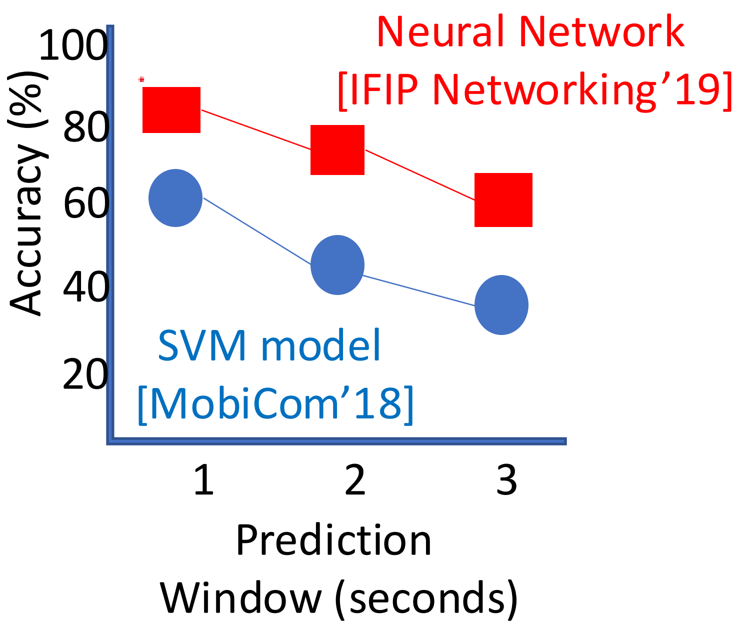
Viewport-Adaptive 360-Degree Streaming

- ❑Viewport prediction models
  - Predicting user head movement is hard
  - The accuracy drops significantly as we predict a longer horizon user motion



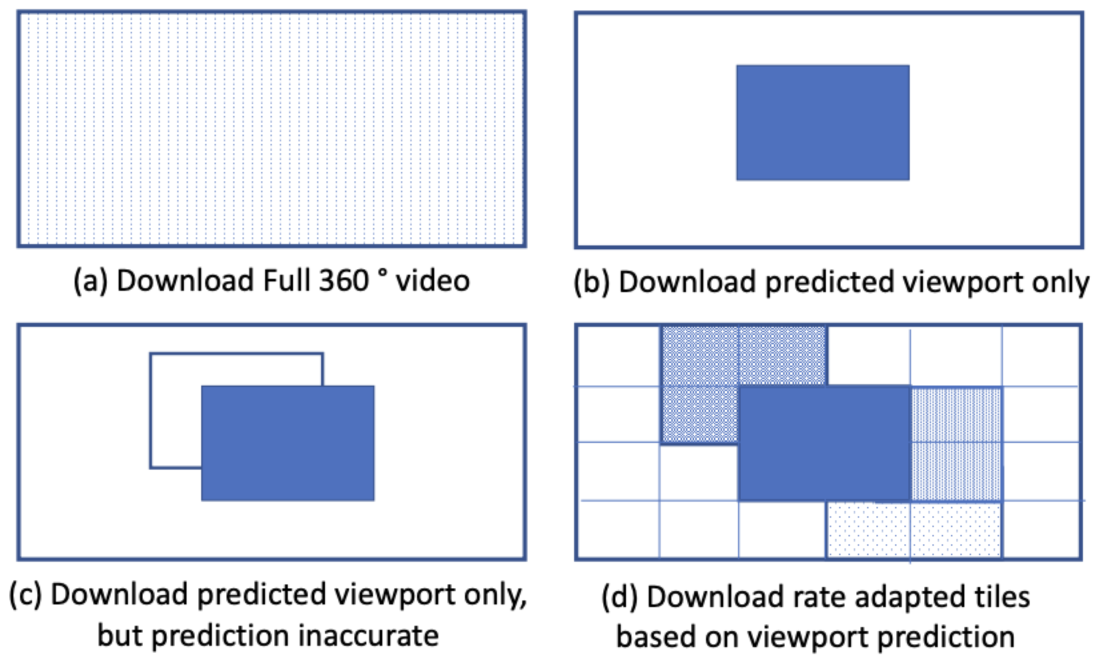
Viewport-Adaptive 360-Degree Streaming

- ❑ Viewport prediction models
  - Predicting user head movement is hard
  - **Solution**:
    - Fetch more tiles to avoid the tile misses
    - Fetching more tiles competes for bandwidth and reduces video quality



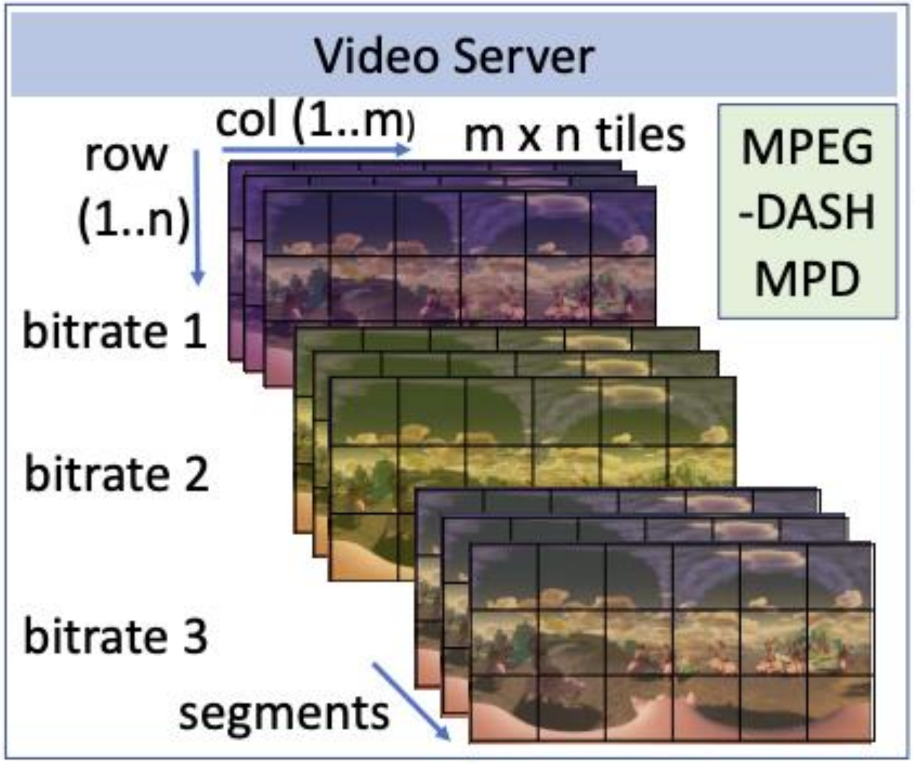
Viewport-Adaptive 360-Degree Streaming

- Tile selection strategies



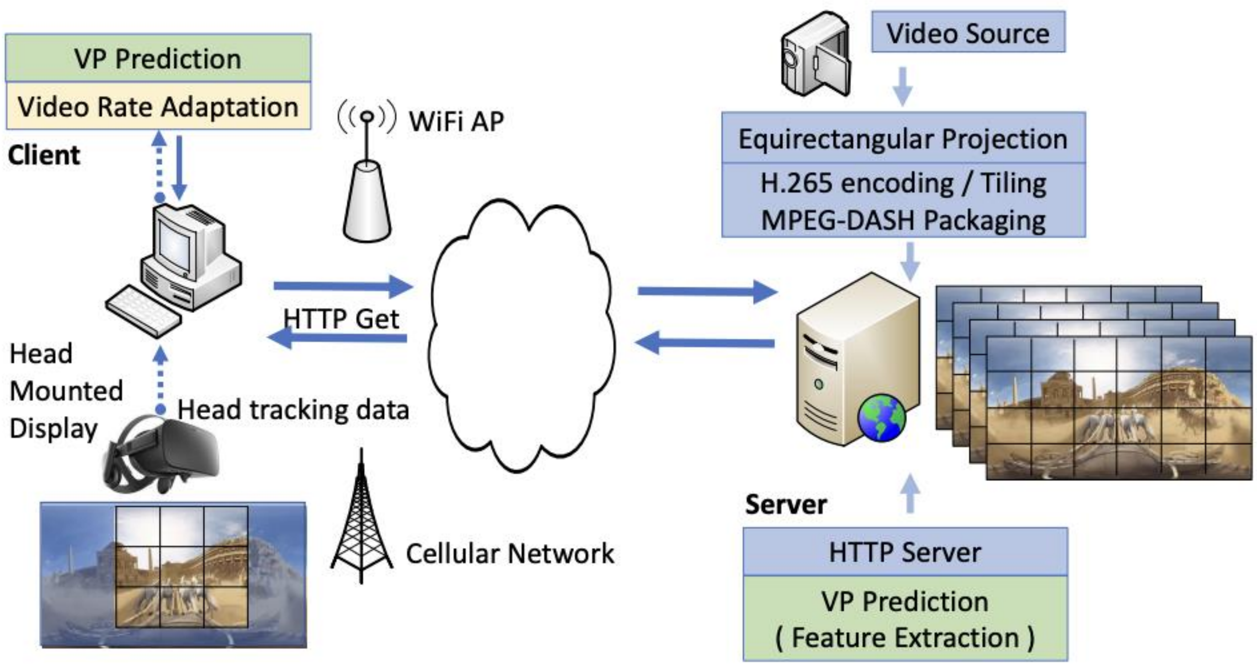
Viewport-Adaptive 360-Degree Streaming

- Tiles stored at the server in different qualities



Viewport-Adaptive 360-Degree Streaming

- End-to-end streaming pipeline



Viewport-Adaptive 360-Degree Streaming

Equation: Tile Selection Optimization

Let T be the set of all tiles in a video frame, and V_p be the set of tiles within the predicted viewport. Each tile t has an associated quality level q_t and required bitrate $b_t(q_t)$. The optimization problem aims to maximize the overall quality of the viewport under the total available bandwidth B.

Objective:

$$\max_{\{q_t\}} \sum_{t \in V_p} w(t) \cdot q_t$$

Subject to:

$$\sum_{t \in T} b_t(q_t) \leq B$$

where $w(t)$ is the weight (importance) of tile t based on its position within the predicted viewport V_p, reflecting the user's likely focus area.

Viewport-Adaptive 360-Degree Streaming

We need throughput estimation similar to the case of regular videos

Equation: Throughput Estimation

Let T_i be the throughput estimate after downloading the i-th video segment, S_i be the size of the i-th segment (in bits), and D_i be the download duration (in seconds). The throughput estimate can be updated as:

$$T_{i+1} = \alpha \cdot T_i + (1 - \alpha) \cdot \left(\frac{S_i}{D_i} \right)$$

where $0 < \alpha < 1$ is a smoothing factor that controls the impact of past throughput measurements on the current estimate.

Viewport-Adaptive 360-Degree Streaming

- Tile Miss Ratio
  - measures how often a user's viewport contains tiles that were not pre-fetched or delivered at the desired quality.

The Tile Miss Ratio is calculated as:

$$\text{TMR} = \frac{\sum(\text{missed tiles in viewport})}{\sum(\text{total tiles in viewport})}$$

Where:

- **Missed tiles** refer to tiles that are either delivered at a lower-than-desired quality or not delivered in time to be displayed when the user shifts their viewport.
- **Total tiles in viewport** refer to the total number of tiles that should have been displayed at high quality.

Viewport-Adaptive 360-Degree Streaming

You are given a 360-degree video streaming scenario where the video sphere is divided into **100 tiles**. The system uses a **viewport-adaptive streaming** mechanism to stream only the tiles within a user's **90-degree** horizontal field of view.

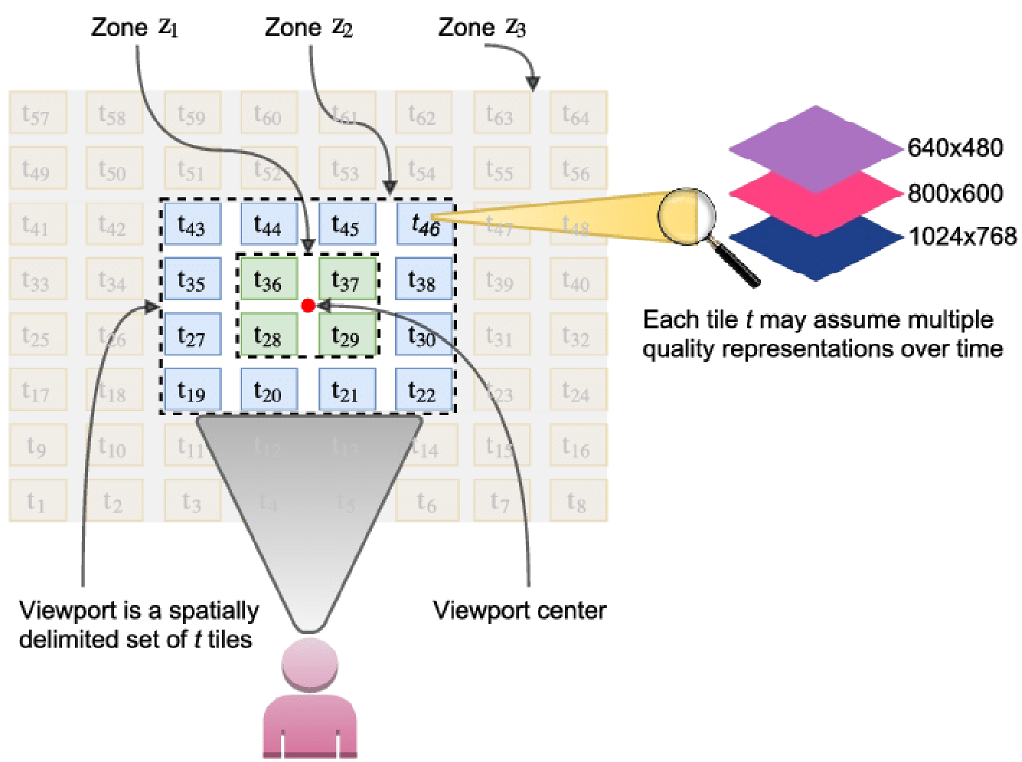
Assume the following:

- A user's viewport can cover **25 tiles** at a time.
- The system predicts the viewport for the next 1 second and fetches tiles accordingly.
- Due to unpredictable head movements, the user's actual viewport deviates from the prediction by **10% on average**.
- The available bandwidth allows for **20 tiles per second** to be streamed at high quality.
- If a tile is not delivered in high quality, it is considered a "missed tile" for that frame.

Calculate the Tile Miss Ratio based on the given data

Viewport-Adaptive 360-Degree Streaming

- Need to adapt spatial quality as well



Predicting_the_Performance_of_Virtual_Reality_Video_Streaming_in_Mobile_Networks, MMSys'2018

Viewport-Adaptive 360-Degree Streaming

- Compression overhead with tiling
  - Looses out on exploiting spatio-temporal redundancy that existing across the tiles
  - Why?

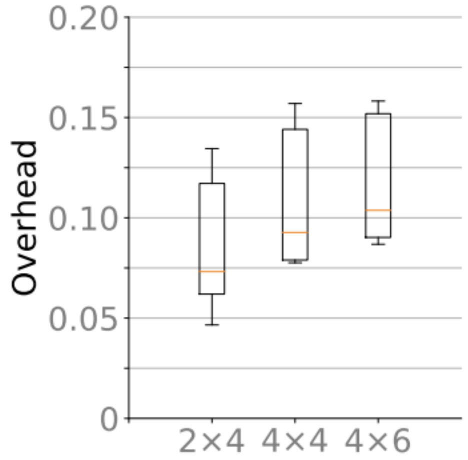
Viewport-Adaptive 360-Degree Streaming

- Compression overhead with tiling
  - Looses out on exploiting spatio-temporal redundancy that existing across the tiles
  - Why?
    - Tiles need to be encoded independently so that can be streamed independently

Viewport-Adaptive 360-Degree Streaming

- Compression overhead with tiling
  - Looses out on exploiting spatio-temporal redundancy that existing across the tiles

  - Why?
    - Tiles need to be encoded independently so that can be streamed independently



Summary of the Lecture

- 360-Degree Video Streaming
  - Viewport prediction
  - Viewport adaptation

Next up: Point cloud streaming