30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'20), June 2020.

Low-latency Cloud-based Volumetric Video Streaming Using Head Motion Prediction

Serhan Gül [1], Dimitri Podborski [1], Thomas Buchholz [2], Thomas Schierl [1], Cornelius Hellge [1]

[1] Fraunhofer Heinrich Hertz Insitute

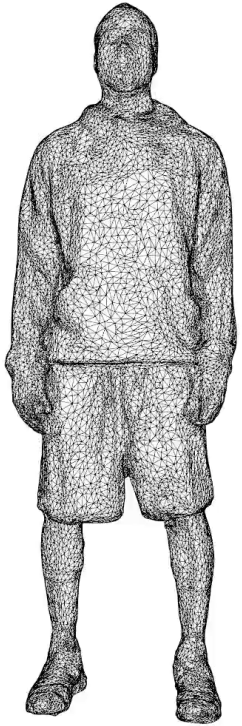
[2] Deutsche Telekom AG

Presenter: Serhan Gül

Mesh-based Volumetric video

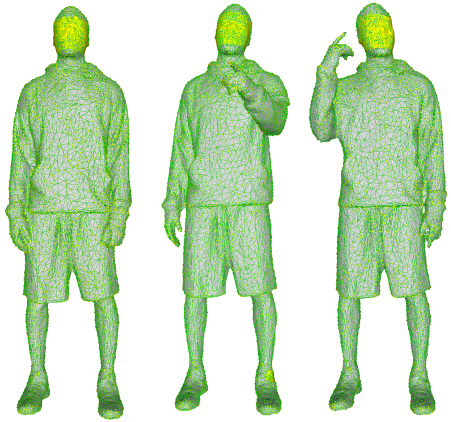
Rendering result

Meshes



Texture

Volumetric video: Content preparation



Video encoder

Mesh encoder

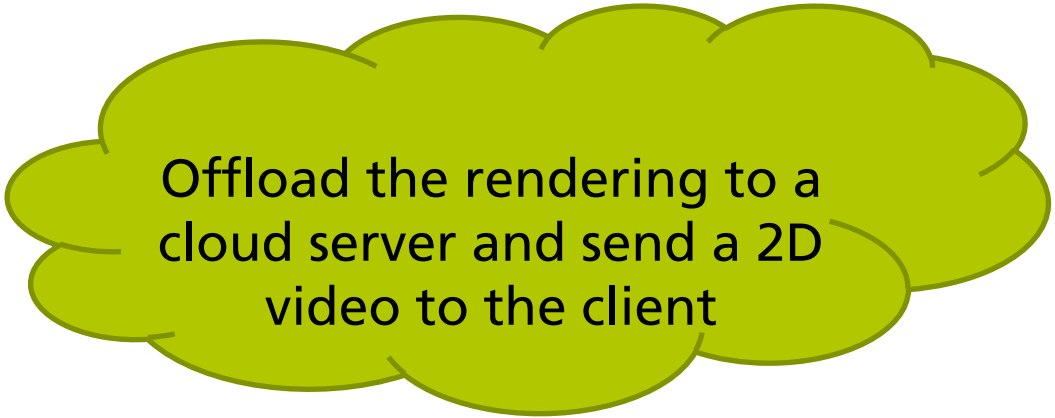
Multiplexer

MP4 file

Processing and Streaming Challenges

- Multiple volumetric objects
  - Not enough processing power on mobile devices
- No hardware decoders for point clouds/meshes available
- Bitrates are still too high

- Solution:



Cloud-based rendering

AR Client

Control data

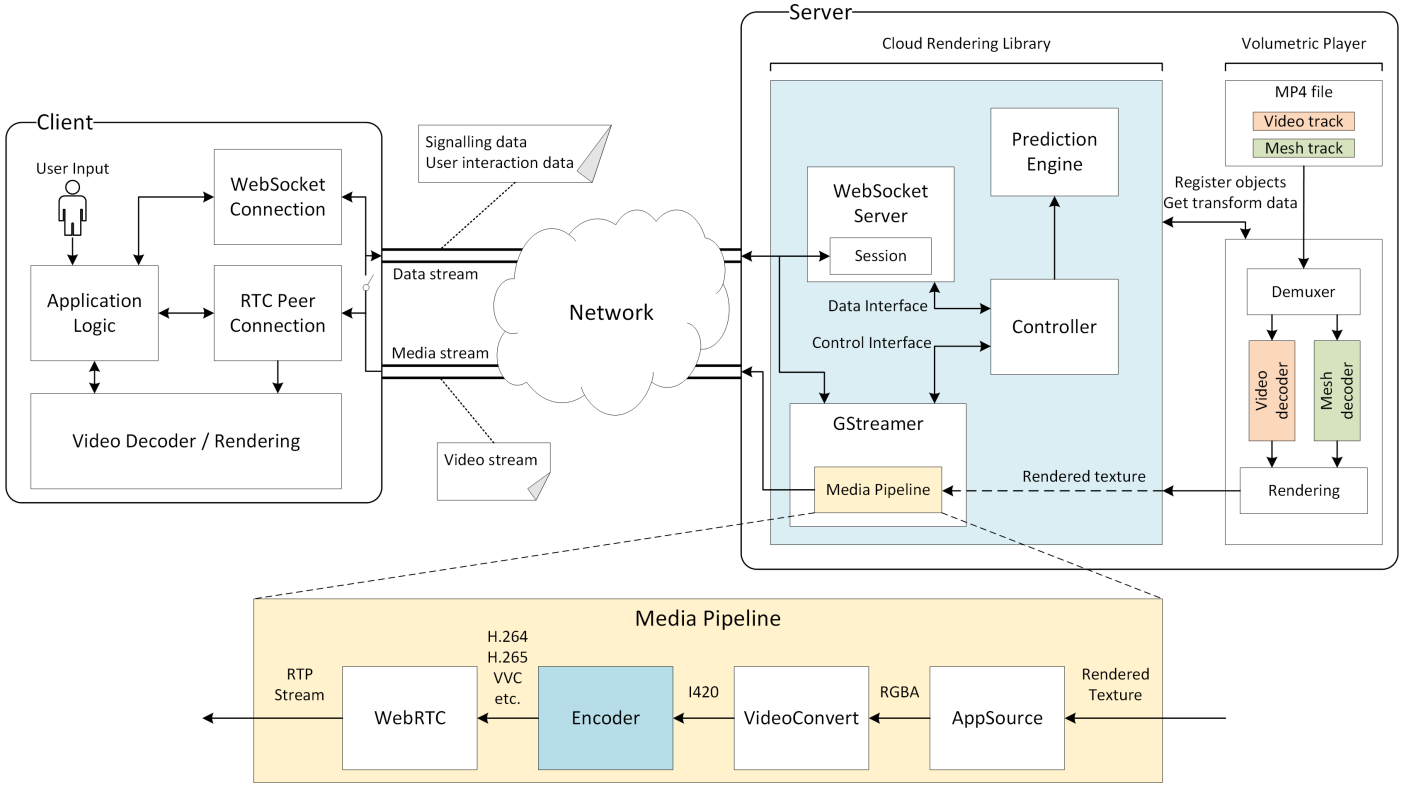
- Change viewport
- Move objects
- Control playback

2D Video stream

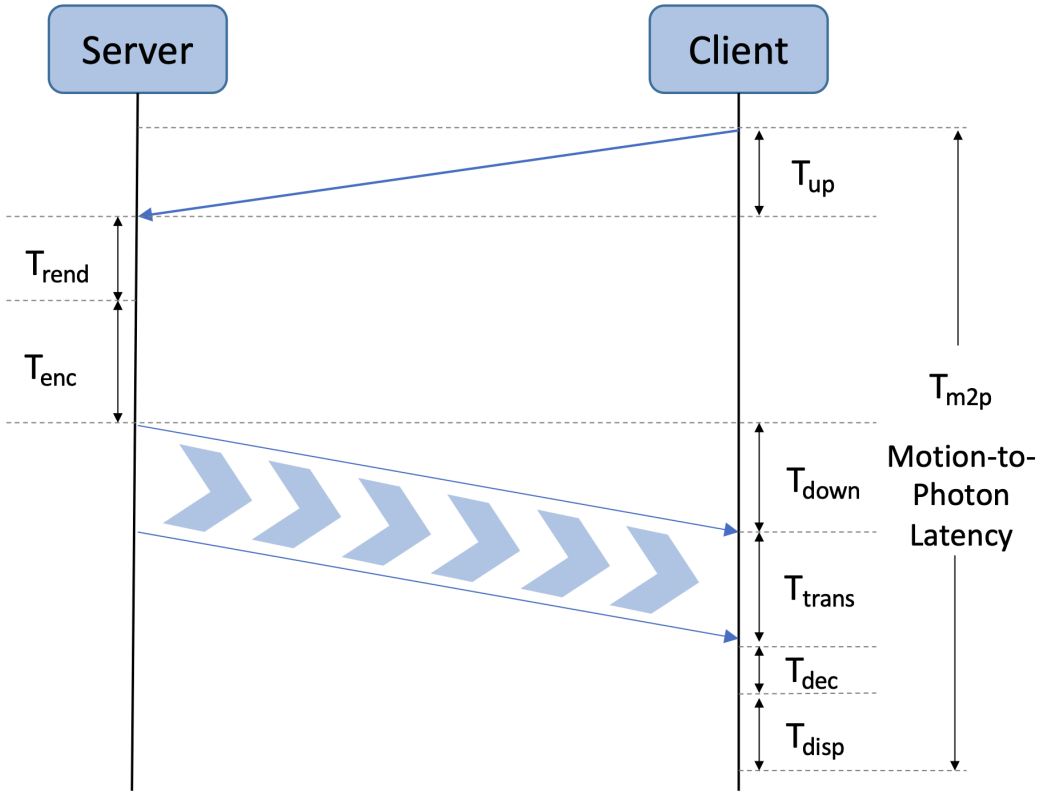
- Volumetric video processing
  - demux / decode / render
  - multiple objects
- Rendering based on user input
- Encode a 2D rendered view

Cloud server

Architecture overview



Motion-to-Photon Latency



Our good friends

- Hardware video encoding (NVIDIA Encoding API)
- WebRTC (ultra low latency streaming)
- 5G / Edge computing
- Head motion prediction
- Post-rendering correction

Autoregression-based predictor

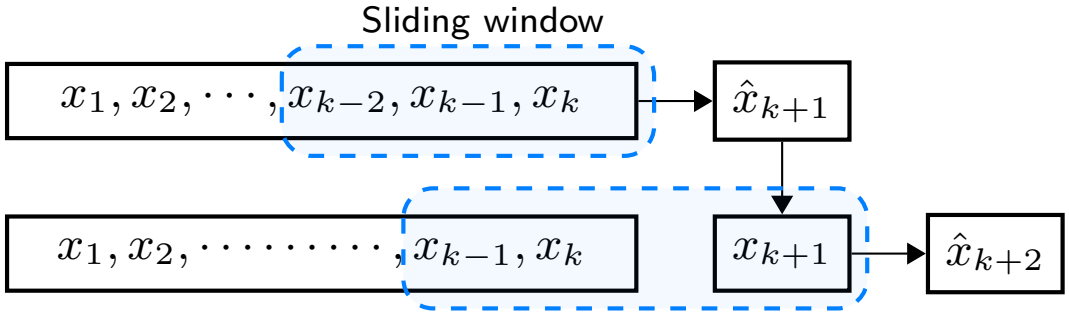
- We collected 5 user traces while interacting with a virtual object overlaid on real world
- Used one of the collected traces as training data and created two models (translational and rotational)
- Predict the next sample using a history window of the past 32 samples (=160 ms, sampling time= 5ms).

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_\rho x_{t-\rho} + \epsilon_t$$

Model coefficients

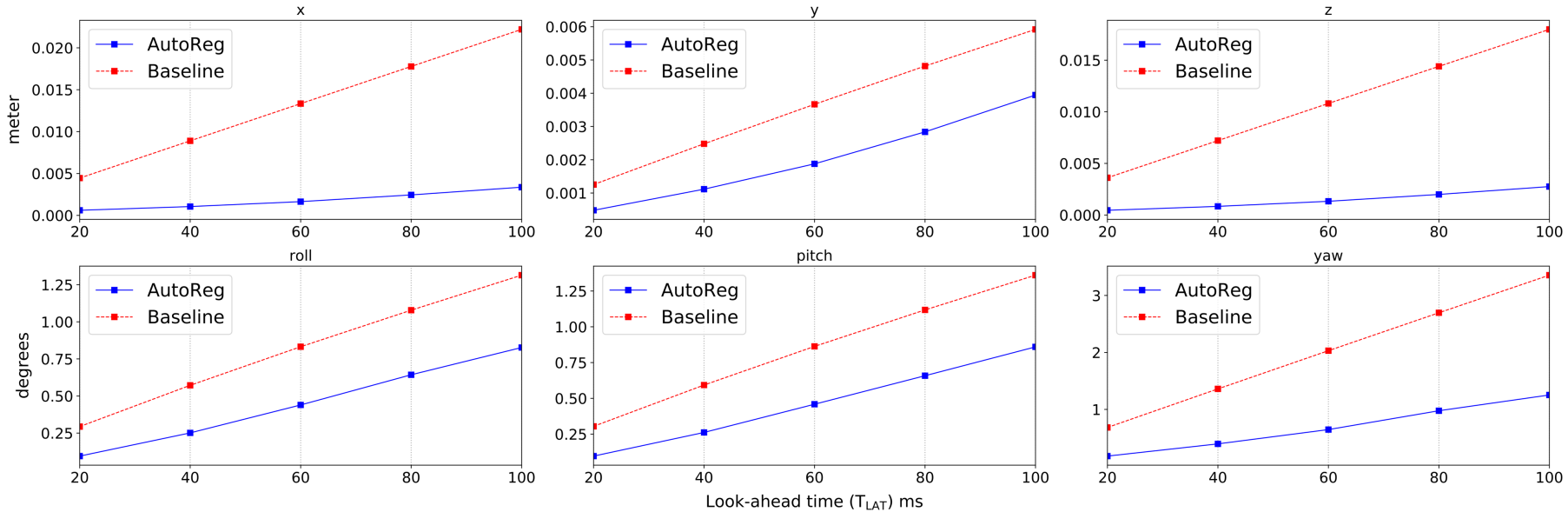
Lagged values

Multi-step prediction (example history window=3, prediction window=2)



Results: Autoregression-based predictor

- Prediction performed separately on x , y , z , q_w , q_x , q_y , q_z components.
- Quaternions converted to Euler angles for evaluation



Mean absolute error (MAE) for different translational and rotational components averaged over **5** users. Results are given for the look-ahead times in the range **20-100ms**.

Conclusion

- We developed a cloud-based based volumetric streaming system that reduces the rendering load on the client-side.
- Compensate the increased system latency using head motion prediction
- Our results show that the autoregression model reduces the rendering errors caused by the added latency
- Future work
  - analyze the effect of motion-to-photon latency on the user experience through subjective tests
  - develop more advanced prediction techniques e.g. based on Kalman filtering.

Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI

Thank you

Contact:

Serhan Gül
serhan.guel@hhi.Fraunhofer.de

Einsteinufer 37
10587 Berlin