ACM Multimedia Conference
12-16 October 2020, Seattle, United States (virtual)

Kalman Filter-based Head Motion Prediction for Cloud-based Mixed Reality

Serhan Gül, Sebastian Bosse, Dimitri Podborski, Thomas Schierl, Cornelius Hellge

Fraunhofer Heinrich Hertz Institute

Video Coding and Analytics Department

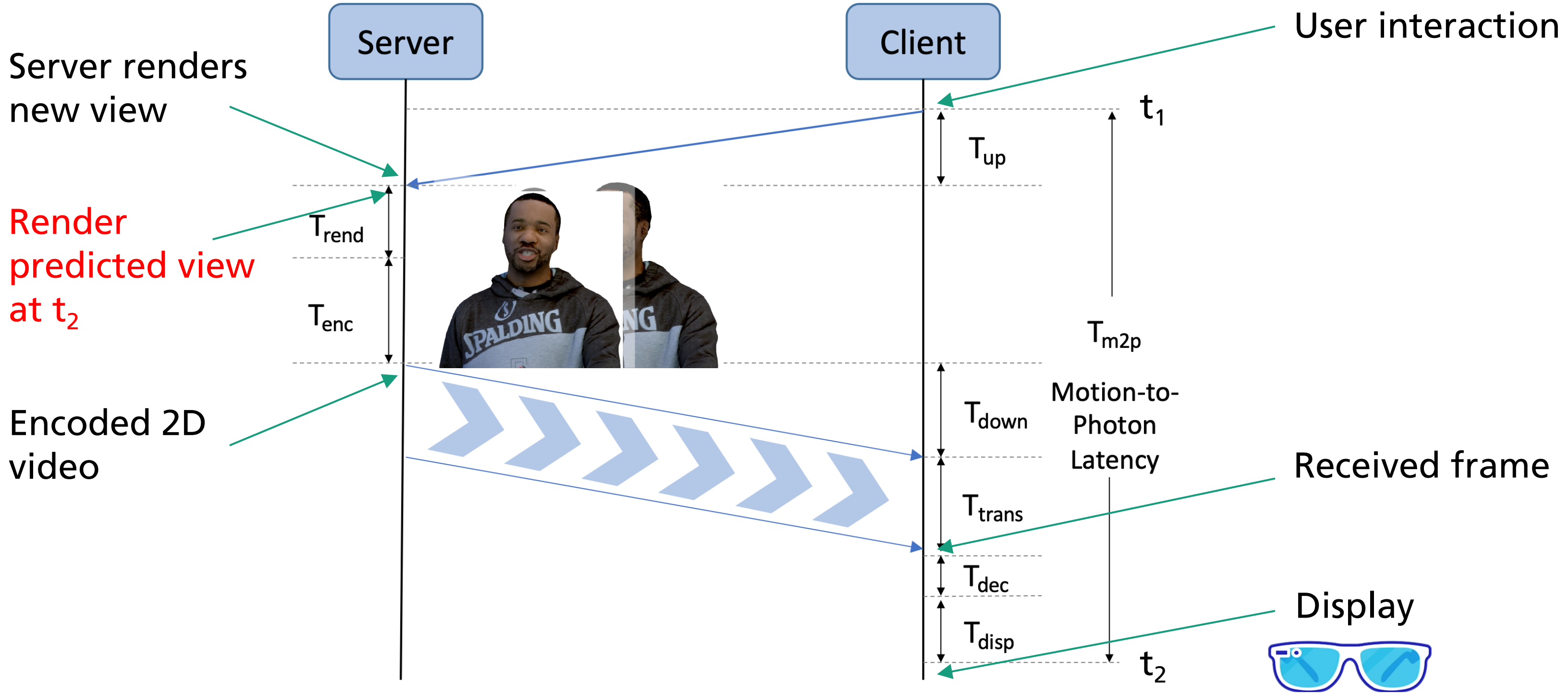
Volumetric Video Streaming for Mixed Reality

- Problem:
  - Complex rendering
- Solution:
  - Offload rendering to cloud server
  - Transmit 2D video stream based on user’s head pose
- Drawback:
  - Increased interaction (motion-to-photon) latency

Serhan Gül

Fraunhofer HHI

Interaction Latency of a Remote Rendering System



Head motion prediction: State of the Art

Viewport Prediction for VR

- Predict user's future viewport for optimized 360° video streaming
- Different approaches
  - Sensor-based [1]
  - Content-based [2]
  - Data-driven [3]
- 3DoF prediction for VR

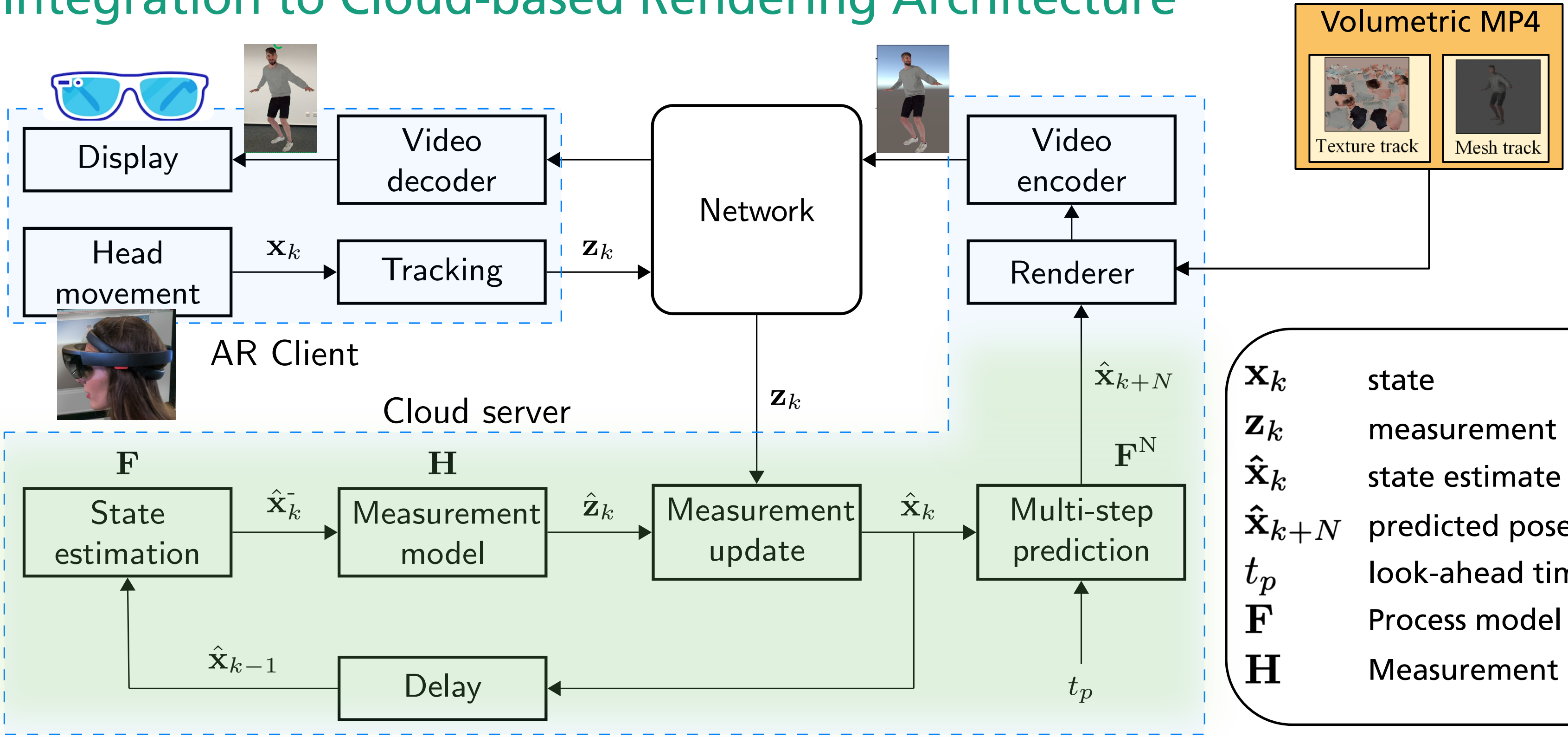
[1] Bao, Yanan, et al. "Shooting a moving target: Motion-prediction-based transmission for 360-degree videos." 2016 IEEE International Conference on Big Data (Big Data). IEEE, 2016.

[2] Ozcinar, Cagri, Julian Cabrera, and Aljosa Smolic. "Visual attention-aware omnidirectional video streaming using optimal tiles for virtual reality." *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 9.1 (2019): 217-230.

[3] Petrangeli, Stefano, Gwendal Simon, and Viswanathan Swaminathan. "Trajectory-based viewport prediction for 360-degree virtual reality videos." 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). IEEE, 2018.

Head Motion Prediction with Kalman Filter

Integration to Cloud-based Rendering Architecture

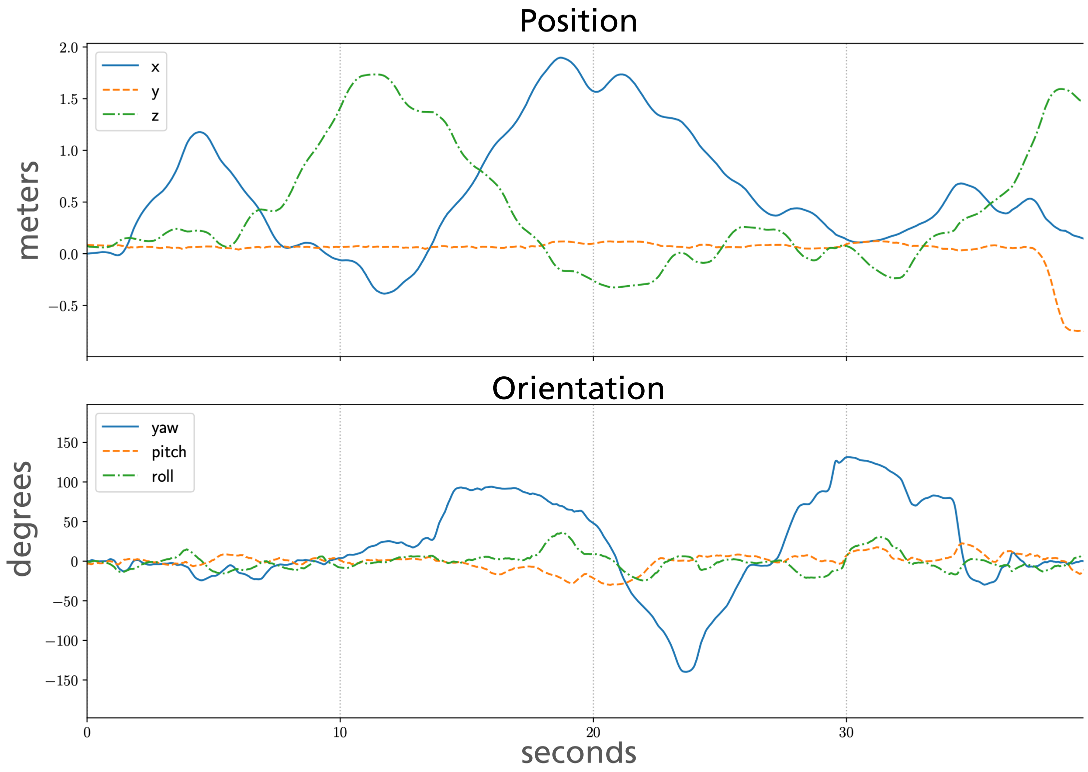


$\mathbf{x}_k$	state
$\mathbf{z}_k$	measurement
$\hat{\mathbf{x}}_k$	state estimate
$\hat{\mathbf{x}}_{k+N}$	predicted pose (N steps)
t_p	look-ahead time
$\mathbf{F}$	Process model
$\mathbf{H}$	Measurement model

Serhan Gül

Experimental Setup

Data Collection

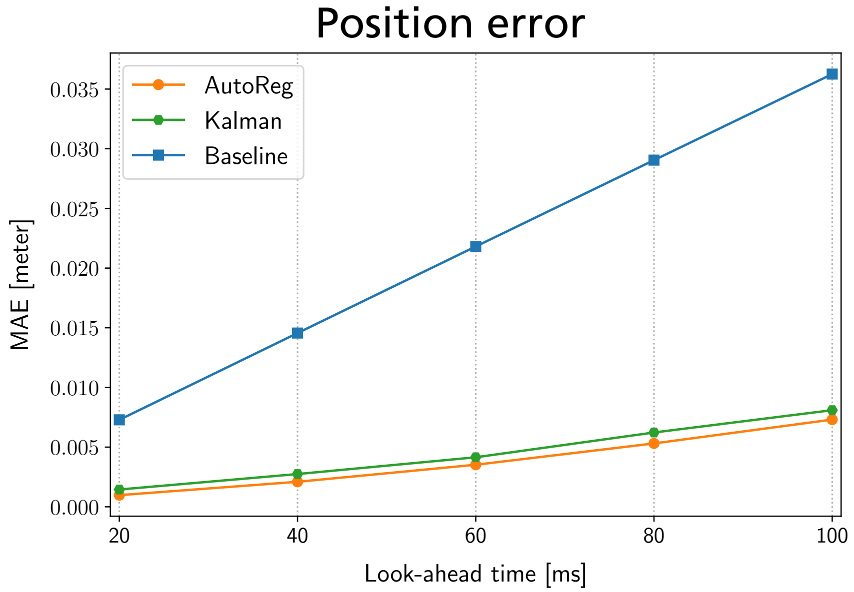


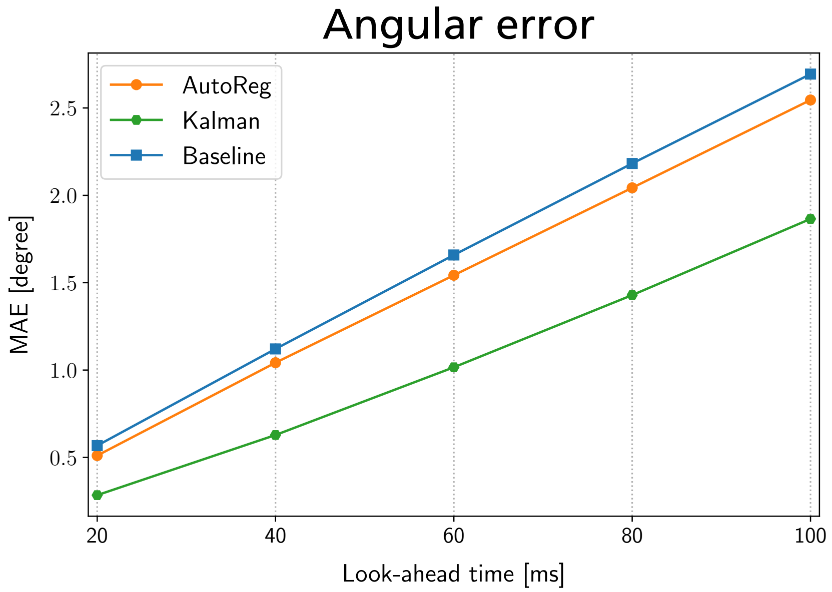
Serhan Gül

Fraunhofer HHI

Experimental Results

Different Prediction Models





MAE: Mean absolute error

AutoReg Gül, Serhan, et al. "Low-latency cloud-based volumetric video streaming using head motion prediction." *ACM NOSSDAV*. 2020.

Conclusion and Future Work

- Kalman filter-based head motion prediction at cloud server
- Predict orientation more accurately than AutoReg
- Robust to variations, needs no training
- Future work:
  - Use spherical distributions (e.g. von-Mises-Fisher) to predict orientation better
  - Subjective evaluation of latency and prediction errors

Serhan Gül

Fraunhofer HHI

Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, HHI

WE PUT SCIENCE INTO ACTION.

Kalman Filter-based Head Motion Prediction for Cloud-based Mixed Reality

Serhan Gül
Fraunhofer HHI
Berlin, Germany
serhan.guel@hhi.fraunhofer.de

Sebastian Bosse
Fraunhofer HHI
Berlin, Germany
sebastian.bosse@hhi.fraunhofer.de

Dimitri Podborski
Fraunhofer HHI
Berlin, Germany
dimitri.podborski@hhi.fraunhofer.de

Thomas Schierl
Fraunhofer HHI
Berlin, Germany
thomas.schierl@hhi.fraunhofer.de

Cornelius Hellge
Fraunhofer HHI
Berlin, Germany
cornelius.hellge@hhi.fraunhofer.de

ABSTRACT

Volumetric video allows viewers to experience highly-realistic 3D content with six degrees of freedom in mixed reality (MR) environments. Rendering complex volumetric videos can require a prohibitively high amount of computational power for mobile devices. A promising technique to reduce the computational burden on mobile devices is to perform the rendering at a cloud server. However, cloud-based rendering systems suffer from an increased interaction (motion-to-photon) latency that may cause registration errors in MR environments. One way of reducing the effective latency is to predict the viewer's head pose and render the corresponding view from the volumetric video in advance.

In this paper, we design a Kalman filter for head motion prediction in our cloud-based volumetric video streaming system. We analyze the performance of our approach using recorded head motion traces and compare its performance to an autoregression model for different prediction intervals (look-ahead times). Our results show that the Kalman filter can predict head orientations 0.5 degrees more accurately than the autoregression model for a look-ahead time of 60 ms.

KEYWORDS

volumetric video, augmented reality, mixed reality, cloud-based rendering, head motion prediction, Kalman filter, time series analysis

1 INTRODUCTION

With the advances in volumetric capture technologies, volumetric video has been gaining importance for the immersive representation of 3D scenes and objects for virtual reality (VR) and augmented reality (AR) applications [42]. Combined with highly accurate positional tracking technologies, volumetric video allows users to freely explore six degrees of freedom (6DoF) content and enables novel mixed reality (MR) applications where highly realistic virtual objects can be placed inside real environments and animated based on user interaction [14].

Geometry of volumetric objects is usually represented using meshes or point clouds. High-quality volumetric meshes typically contain thousands of polygons, and high-quality point clouds may contain millions to billions of points [41, 43]. Therefore, rendering complex volumetric content is still a very demanding task despite the remarkable computing power available in today's mobile devices [10]. Moreover, no efficient hardware implementations of mesh/point cloud decoders are available yet. Software-based decoding can be prohibitively expensive in terms of battery usage and may not be able to meet the real-time rendering requirements [37].

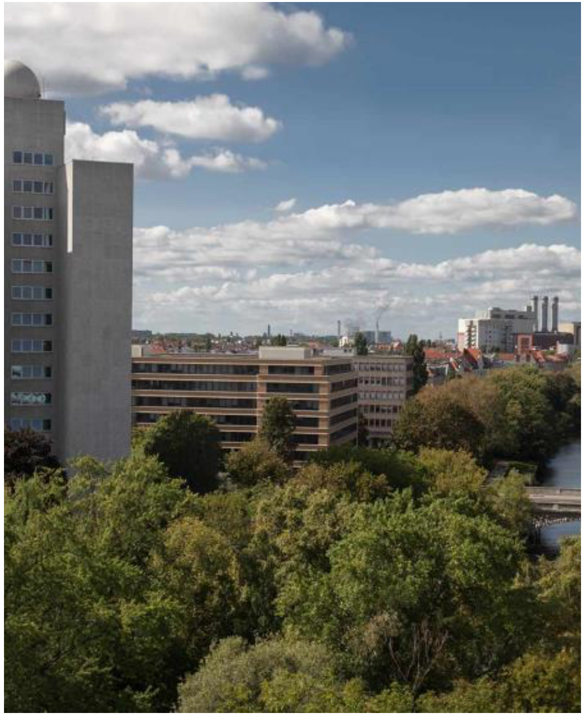
One way to avoid the complex rendering on mobile devices is to offload the processing to a powerful remote server which dynamically renders a 2D view from the volumetric video based on the user's actual head pose [46]. The server then compresses the rendered texture into a 2D video stream and transmits it over a network to the client. The client can then efficiently decode the video stream using its hardware decoder and display the dynamically updated content to the viewer. Moreover, the cloud-based rendering approach allows utilizing highly efficient 2D video coding techniques and thus can reduce the network bandwidth requirements by avoiding the transmission of the volumetric content [37].

However, one drawback of cloud-based rendering is the increased interaction latency, also known as the motion-to-photon (M2P) latency [47]. Due to the network round-trip time and the added processing delays, the M2P latency is higher than in a system that performs the rendering locally. Several studies show that an increased interaction latency may lead to a degraded user experience and motion sickness [8, 27, 30].

One way to reduce the latency is to predict the user's future head pose at the cloud server and render the corresponding view of the volumetric content in advance. Thereby, it is possible to significantly reduce or even eliminate the M2P latency, if the user pose is successfully predicted for a look-ahead time (LAT) equal to or larger than the M2P latency of the system. However, mispredictions of head motion may increase registration errors and degrade the user experience in AR environments [27]. Therefore, designing accurate head motion prediction algorithms is crucial for high-quality volumetric video streaming.

In this paper, we consider the problem of head motion prediction for cloud-based AR/MR applications. Our main contributions are as follows:

- We develop a Kalman filter-based predictor for head motion prediction in 6DoF space and analyze its performance compared to an autoregression model and a baseline (no prediction) model using recorded head motion traces.



Contact:

Serhan Gül
serhan.guel@hhi.fraunhofer.de

Einsteinufer 37
10587 Berlin