



Predictive Video Streaming for Immersive 6G Teleoperation: From Stream Metrics to Task Performance

Kevin-Ismet Šabanović
kevin.sabanovic@tu-dortmund.de
TU Dortmund University
Dortmund, Germany

Hendrik Schippers
hendrik.schippers@tu-dortmund.de
TU Dortmund University
Dortmund, Germany

Niklas A. Wagner
niklas.wagner@tu-dortmund.de
TU Dortmund University
Dortmund, Germany

Simon Schippers
simon.schippers@tu-dortmund.de
TU Dortmund University
Dortmund, Germany

Christian Wietfeld
christian.wietfeld@tu-dortmund.de
TU Dortmund University
Dortmund, Germany

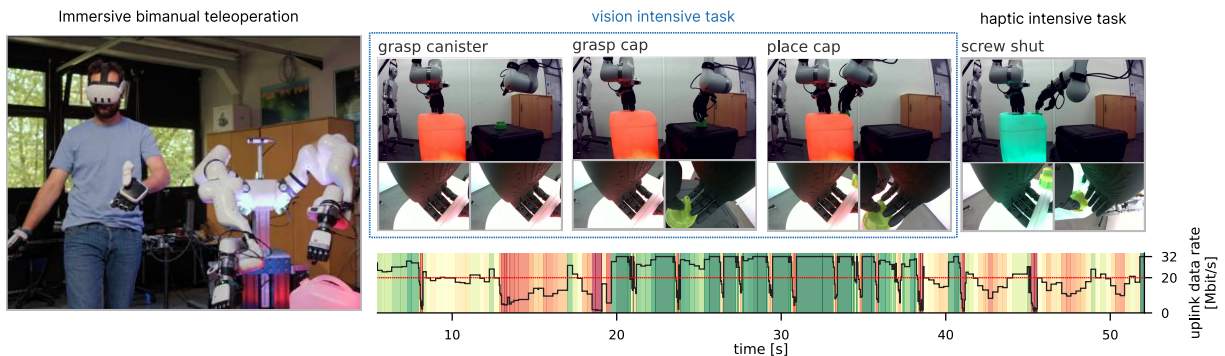


Figure 1: Immersive bimanual teleoperation of the capping task. The first three phases are guided by vision, the screwing phase by touch. The trace below gives the uplink data rate the three camera streams have to fit into.

Abstract

Two-handed robotic systems are needed for complex tasks like sealing a container or operating a valve. Immersive teleoperation in combination with haptic feedback is seen as a core service of 6G due to low latency and reliability requirements. While video requires tens of Mbit/s, tactile feedback costs only kbit/s. We used an emulated link driven by a capacity trace recorded from an urban channel emulation (1 to 32 Mbit/s) and ran 19 trials with one operator across three

rate-control schemes: fixed rate, predicted capacity, prediction with queue feedback. We report a reduction of frame loss from 14 to 1 %, a p_{95} queueing delay reduction from 251 to 68 ms and a blind time reduction from 19 to 1 %. At task level the 6.4 s difference stays within the run-to-run spread.

CCS Concepts

• **Networks** → **Network measurement**; *Network experimentation*; • **Human-centered computing** → Human computer interaction (HCI).

Keywords

teleoperation, predictive video streaming, network emulation, task performance, haptic feedback

ACM Reference Format:

Kevin-Ismet Šabanović, Hendrik Schippers, Niklas A. Wagner, Simon Schippers, and Christian Wietfeld. 2026. Predictive Video



This work is licensed under a Creative Commons Attribution 4.0 International License.

CoNEXT Companion '26, Utrecht, Netherlands

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-3006-1/2026/12

<https://doi.org/10.1145/3844665.3859468>

Streaming for Immersive 6G Teleoperation: From Stream Metrics to Task Performance. In *The 22nd International Conference on emerging Networking EXperiments and Technologies (CoNEXT Companion '26)*, December 07–10, 2026, Utrecht, Netherlands. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3844665.3859468>

1 Introduction

Robotic systems are increasingly used to perform tasks in cases where humans could be in danger [6]. For example, sealing a hazardous container or operating a valve requires two hands to grasp, align and fit one part to another (Fig. 1). Hence, immersive teleoperation is often the only viable option. It is one of the key use cases driving the requirements of the 6th Generation of Mobile Communication Networks (6G). While 6G research proposes mechanisms such as multi-link transmission and proactive video streaming to stabilize such links [7], their benefit is commonly quantified at stream level. Immersive visual and tactile feedback both need widely different data rates. The force feedback requires around 60 kbit/s for pose and force values [5]. In comparison, video transmission requires data rates around 15–20 Mbit/s, while both share the same 6G uplink. Thus, on a challenging uplink the video stream is mainly the target of insufficient data rates, while the operator is aligning a thread.

Predictive encoders are choosing a bit rate depending on the current networking conditions and try to remain under the maximum available bandwidth [4]. They measure video encoder performance indicators such as the queue delay, frame loss or image quality. This is typically done with prerecorded video that is retransmitted, and not with an immersive teleoperation task in the loop. Remote driving over commercial 4G and 5G networks is evaluated in the same way, but without a manipulation task [2, 3].

On the other side, Quality of Experience (QoE) is measured as well [1], but again without a contact task. Subjective evaluations of impaired teleoperation video have been reported for a comparable setup [8], but they stop at the perception of the operator and do not cover the completion of a contact task. Stream-level gains are therefore well studied, but whether they suffice for a demanding manipulation mission to succeed, such as screwing a cap onto a canister, remains open. In this work we demonstrate a measurement chain that evaluates immersive haptic teleoperation at stream, image and task level under realistic cellular emulation.

2 Experimental Methodology

Instead of simulating network condition, we use a hierarchical token bucket (HTB) shaper to replay a capacity trace recorded from an urban channel emulation. HTB follows the trace between 1 and 32 Mbit/s over a 59 s period. The encoder and the media server run in separate network namespaces. Queued packets stay charged to the socket and

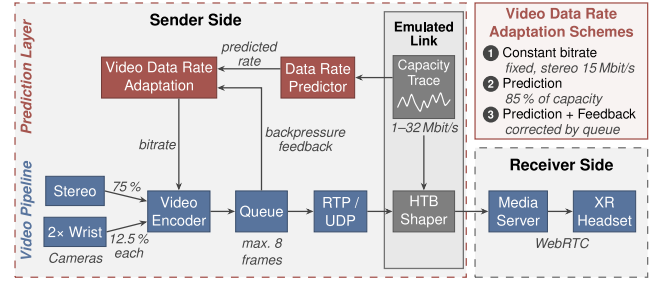


Figure 2: Measurement setup and the three video data rate adaptation schemes. The encoder pipeline is drawn once and stands for all three camera streams.

the send call blocks. This propagates back pressure into the encoder pipeline, simulating the behaviour a real link which packet-dropping qdisc does not. The three rate-control schemes shown in Fig. 2 reproduce the single-link configurations of the reference system [4]. The constant scheme encodes a fixed 15 Mbit/s for the stereo pair (1), the prediction scheme targets 85 % of the predicted capacity (2), and the feedback scheme corrects that target by the fill level of the send queue (3). The available rate is split by pixel rate, 75 % to the stereo pair and 12.5 % to each wrist camera. They run at 30 fps with a queue depth of 8 frames on a single link, against 60 fps, 15 frames and multiple links in the reference system. Both the predictor and the shaper are driven by the same capacity trace, so the controller knows the channel without prediction error.

The teleoperation task consists of four consecutive phases: grasping the canister, grasping the cap, placing the cap, and screwing it shut. For each video frame, the sender records the enqueue and dequeue timestamps using a common clock. The corresponding decoded-frame timestamps are obtained from the recorded video. A video stall is defined as an interval exceeding 100 ms without a newly decoded frame. A trial is considered successful if the cap is fully closed within the 120 s time limit. After a short training period 19 runs were performed by one operator. The haptic loop is closed locally and its traffic never crosses the emulated link, so the screwing phase cannot be affected by the channel. The task-level comparison therefore uses the visually guided time, that is the mission time without the screwing phase.

3 Results

The stereo encoder reached median rates of 15.0, 16.8 and 14.9 Mbit/s under the three schemes of Fig. 2. They are within a few percent of each other, so the regulated schemes do not simply send less data, but redistribute out of capacity dips. Across the three rate-control schemes, the frame loss decreases from 14 to 1 % and the p_{95} queuing delay decreases

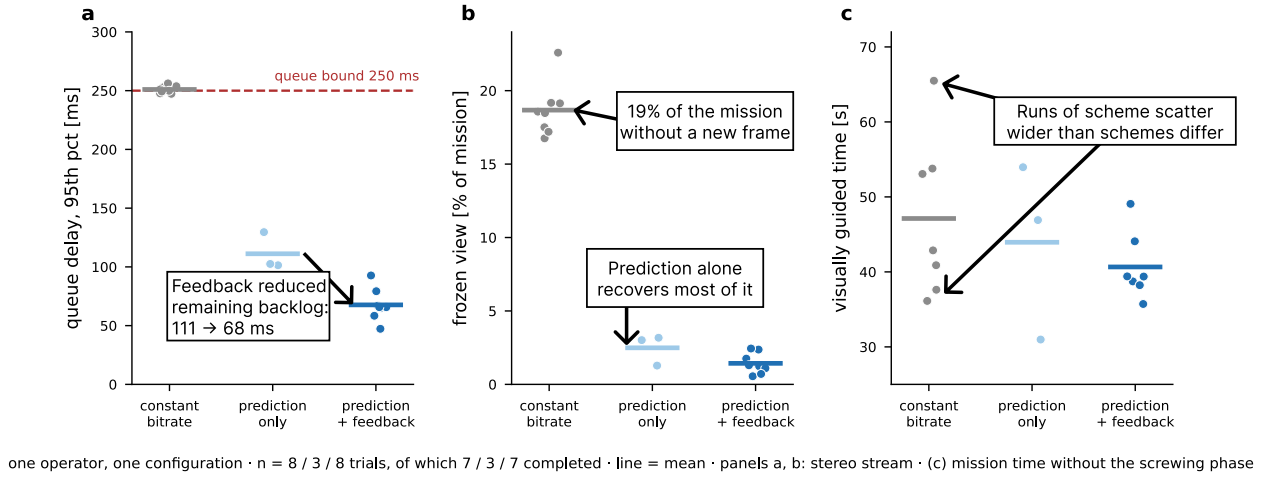


Figure 3: Stream, image and task level under the three rate-control schemes. Each point is one trial, the line is the mean; (a) and (b) show the stereo stream, (c) is the mission time without the screwing phase.

from 251 to 68 ms. The time the operator is effectively blind is reduced from 19 to 1 % of the mission, while the effective frame rate increases from 24.6 to 29.9 fps. As Fig. 3a and b show, the scatter within a scheme stays far below this separation. Prediction alone is sufficient to substantially reduce frame loss, but remains limited by the accuracy of the predicted channel capacity. Queue feedback does not further reduce frame loss; instead, it corrects overly optimistic rate estimates and reduces the p_{95} queueing delay from 111 to 68 ms. The schemes averaged visually guided times of 47.1 s under constant bitrate and 40.7 s with prediction and queue feedback over seven completed runs each (Fig. 3c); the corresponding mission times are 76.5 and 69.9 s. The difference of 6.4 s is smaller than the variation of about 11 s between runs of the constant scheme, while the feedback runs scatter less. Since both schemes were used at similar positions during the session, the run order does not explain it either. Seven runs per scheme resolve only differences of about 13 s. In total, 17 of 19 trial runs were completed successfully, with both failures occurring during the screwing phase. Predictive streaming removes the stream-level failure entirely, while the task-level gain of 6.4 s stays within the run-to-run spread. Sizing that benefit therefore requires measurement at task level rather than inference from stream metrics.

4 Future Work

In the future more runs with multiple operators are planned as well as a harsher channel to test whether the effect grows beyond the spread, and receiver-side timestamps so that late frames count as well as missing ones.

Acknowledgments

This work has been funded by the German Federal Ministry of Research, Technology and Space (BMFTR) in the course of the *6GEM+ Transfer Hub* and *GEM-X* under grant numbers 16KIS2412 and 16KISS005.

References

- [1] Nabajeet Barman and Maria G. Martini. 2019. QoE Modeling for HTTP Adaptive Video Streaming—A Survey and Open Challenges. *IEEE Access* 7 (2019), 30831–30859. doi:10.1109/ACCESS.2019.2901778
- [2] Markus Hofbauer et al. 2022. Traffic-Aware Multi-View Video Stream Adaptation for Teleoperated Driving. In *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*. 1–7. doi:10.1109/VTC2022-Spring54318.2022.9860513
- [3] Andreas Schimpe et al. 2021. Adaptive Video Configuration and Bitrate Allocation for Teleoperated Vehicles. In *2021 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*. 148–153. doi:10.1109/IVWorkshops54471.2021.9669258
- [4] Hendrik Schippers et al. 2025. RISE: Multi-Link Proactive Low-Latency Video Streaming for Teleoperation in Fading Channels. In *IEEE 101st Vehicular Technology Conference (VTC-Spring)*. Oslo, Norway.
- [5] Eckehard Steinbach et al. 2019. Haptic Codecs for the Tactile Internet. *Proc. IEEE* 107, 2 (2019), 447–470. doi:10.1109/JPROC.2018.2867835
- [6] Ioannis Tsitsimpelis et al. 2019. A review of ground-based robotic systems for the characterization of nuclear environments. *Progress in Nuclear Energy* 111 (2019), 109–124. doi:10.1016/j.pnucene.2018.10.023
- [7] Cheng-Xiang Wang et al. 2023. On the Road to 6G: Visions, Requirements, Key Technologies, and Testbeds. *IEEE Communications Surveys & Tutorials* 25, 2 (2023), 905–974. doi:10.1109/COMST.2023.3249835
- [8] Kevin-Ismet Šabanović et al. 2026. QoE Benchmarking of Resilient 6G Teleoperation: VR Digital Twin vs. Video Streaming. In *IEEE International Conference on Communications (ICC)*.