ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Low-Latency VR Telepresence for Remote Inspection in Fence-Free Collaborative Manufacturing

Stanislav Svediroh

***Abstract*—Fence-free collaborative manufacturing, where humans and industrial machines share workspace without physical barriers, requires reliable safety monitoring. Mobile inspection robots can patrol autonomously, but when anomalies are detected—such as unauthorized personnel or safety zone violations—a human operator must rapidly assess the situation through immersive remote control. We present a fully open-source VR framework enabling low-latency stereoscopic video streaming from a mobile robot to a standalone Meta Quest headset. The system supports stereo and mono video modes with hardware-accelerated encoding (H.264/H.265) on NVIDIA Jetson and hardware-accelerated decoding on the headset. Head-coupled camera control maps the operator's gaze to the robot's camera, providing intuitive situational awareness during remote inspection. A key contribution is built-in end-to-end latency instrumentation: per-frame timestamps embedded in RTP header extensions enable continuous monitoring of each pipeline stage from camera capture to photon emission. Measured glass-to-glass latency is approximately $120\,\mathrm{ms}$ over $5\,\mathrm{GHz}$ WiFi. The robot-agnostic architecture requires only a thin adapter layer for integration with any platform. The framework, validated on Boston Dynamics Spot, is publicly available as open source.**

I. Introduction

Fence-free collaborative manufacturing lets workers and machines share space without physical barriers, increasing flexibility but introducing safety challenges. Autonomous safety monitoring systems—multi-camera tracking, geofencing—can detect anomalies such as unauthorized access or unexpected worker behavior, but cannot resolve every situation alone. When an alert is raised, a human operator must visually assess the scene.

Existing telepresence systems for robotic inspection are typically proprietary, tightly coupled to specific hardware, and offer no visibility into their latency pipeline [1]. Operators cannot diagnose whether degraded performance originates from the camera, encoder, network, or decoder—making it difficult to adapt to varying conditions or integrate with new robot platforms.

Our open-source framework addresses these gaps by deploying a mobile inspection robot controlled through an immersive VR headset. The operator sees through the robot's cameras with low-latency head-coupled video, navigates to the scene, and assesses the situation remotely—closing the loop between autonomous detection and human decision-making. The key contributions are: (1) a robot-agnostic telepresence architecture with runtime-configurable streaming, (2) per-frame end-to-end latency instrumentation embedded in the RTP stream, and (3) a fully open-source implementation validated on a real robot platform.

II. System Architecture

The system consists of two sides connected over a standard IP network. Fig. 1 shows the GStreamer pipelines on both sides with the latency probe points used for end-to-end instrumentation.

Robot side (NVIDIA Jetson): GStreamer pipelines capture camera frames from MIPI CSI-2 sensors, apply hardware-accelerated encoding (H.264/H.265), and transmit them as RTP/UDP streams. A REST API (Flask/OpenAPI) enables live reconfiguration of codec, bitrate, resolution, and frame rate without restarting the application. A Python robot controller receives head pose and movement commands via UDP and routes them to the 2-DOF camera pan-tilt and mobility platform through an abstract translator interface—adding support for a new robot requires implementing a single adapter class.

Operator side (Meta Quest): A standalone native C++17 OpenXR application hardware-decodes and renders stereo video at $90\,\mathrm{Hz}$. The operator's head orientation drives the robot-mounted camera in real time, with pose prediction (up to $100\,\mathrm{ms}$) compensating for transmission delay. An in-VR GUI provides live streaming controls, connection status, and per-stage latency readout. The system supports stereo and mono video modes, selectable at runtime without restarting the stream.

III. Latency Instrumentation

A distinguishing feature of the framework is built-in, continuous end-to-end latency measurement. Every video frame carries its own timing metadata through the entire pipeline.

On the robot side, GStreamer identity probes record per-stage durations (camera capture, video conversion, encoding, RTP payloading) and embed them as 16-bit values in an RTP header extension (RFC 5285). The VR client extracts these on arrival, appends decoding and presentation measurements using NTP-synchronized clocks ($<1\,\mathrm{ms}$ accuracy), and displays the full pipeline breakdown in the in-VR GUI overlay and an optional InfluxDB/Grafana dashboard for long-term monitoring and data export.

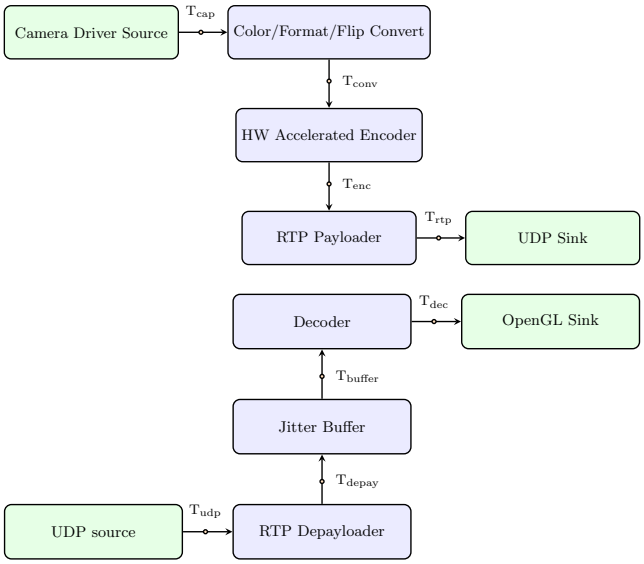

Fig. 1. Sender (top) and receiver (bottom) GStreamer pipelines with latency probe points. Per-stage durations are embedded into the RTP header extension on the sender side and extracted by the VR client for continuous end-to-end monitoring.

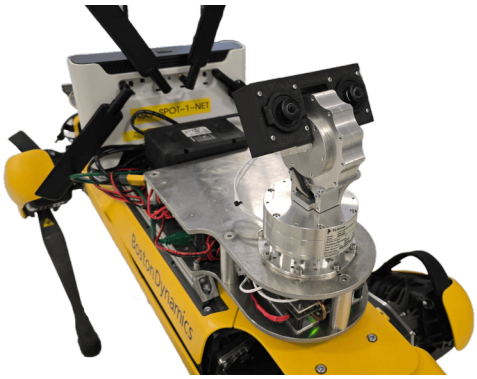

Fig. 2. The telepresence payload mounted on a Boston Dynamics Spot: Jetson Orin NX on a custom carrier board, 2-DOF pan-tilt manipulator with stereo 4K@60FPS cameras, and a 5G-capable router.

IV. Experimental Validation

The framework was validated on a Boston Dynamics Spot equipped with a custom 2-DOF pan-tilt stereo camera payload and an NVIDIA Jetson Orin NX, connected over 5 GHz single-hop WiFi (Fig. 2). The robot was teleoperated across a testbed factory floor by a remote operator wearing a Meta Quest Pro headset. The built-in per-frame latency instrumentation recorded the full pipeline breakdown continuously throughout the experiment.

V. Results

The measured end-to-end pipeline latency was 50 ms to 80 ms and the glass-to-glass latency approximately 120 ms (Fig. 3). The dominant contributors were camera capture ($\sim$17 ms at 60 FPS) and network transmission ($\sim$10 ms to

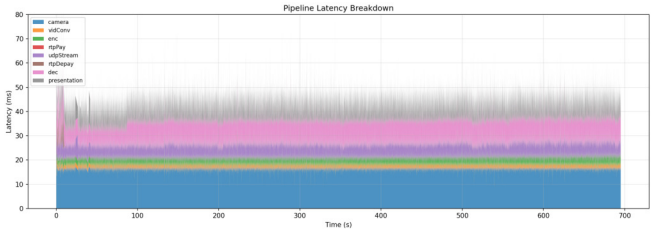

Fig. 3. Stacked per-stage latency breakdown during teleoperation over 5 GHz WiFi, showing camera capture, encoding, network transmission, decoding, and presentation components.

20 ms including jitter). Encoding and decoding each contributed under 10 ms thanks to hardware acceleration on both the Jetson (NVENC) and the Quest (Qualcomm MediaCodec).

OpenXR pose prediction (up to 100 ms lookahead) reduces the operator's perceived latency by anticipating head motion, making head-coupled camera tracking feel responsive despite the physical delay. The per-stage breakdown identifies bottlenecks in real time, enabling the operator to manually reconfigure streaming parameters (codec, bitrate, frame rate) from the in-VR GUI to maintain optimal experience under varying network conditions.

VI. Conclusion

We presented a fully open-source VR telepresence framework for remote robot inspection in collaborative manufacturing environments. The system achieves glass-to-glass latency of approximately 120 ms over 5 GHz WiFi, with built-in per-frame latency instrumentation that provides continuous diagnostic visibility into every pipeline stage. The robot-agnostic architecture enables rapid integration with new platforms through a single adapter class. All source code and build instructions will be made publicly available.

Acknowledgment

The work has been performed in the project Cynergy4MIE: Leverage synergy by cyber-physical systems for the convergence of the eco systems mobility, infrastructure and energy in the circular economy for the Society 5.0 No 101140226/9A24003. The work was co-funded by grants of Ministry of Education, Youth and Sports of the Czech Republic and Chips Joint Undertaking. The work was supported by the infrastructure of RICAIP that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857306 and from Ministry of Education, Youth and Sports under OP RDE grant agreement No CZ.02.1.01/0.0/0.0/17_043/0010085. The completion of this paper was made possible by the grant No. FEKT-S-26-8988 – "Advanced Methods in Cybernetics, Robotics, and Artificial Intelligence" financially supported by the Internal science fund of Brno University of Technology.

References

[1] [Stanislav Svediroh], "A fully open-source VR framework for low-latency stereoscopic video streaming with end-to-end latency instrumentation," *Virtual Reality*, submitted, 2025.