Quadruped Robot Control Using Immersive Interfaces in Emergency Contexts

Anabel Díaz-Labrador, Ángel Delgado, Héctor J. Pérez-Iglesias, Oscar Fontenla-Romero, and Jose Luis Calvo-Rolle

Universidade da Coruña, CTC, CITIC, Department of Industrial Engineering, Ferrol, 15471 A Coruña, Spain
Universidade da Coruña, LIDIA, CITIC, Faculty of Computer Science, 15471 A Coruña, Spain
Fundación Instituto Tecnológico de Galicia, A Coruña, Spain
Correspondence: anabel.dlabrador@udc.es

DOI: *https://doi.org/10.17979/spu.23.c21*

Abstract: This study presents an experimental system for the immersive teleoperation of a quadruped robot in emergency scenarios. Based on the Unitree Go1 robot and integrates the Meta Quest 3 headset for immersive visual feedback, along with its built-in joystick controllers for robot navigation. The user can control the robot's movement intuitively via joystick inputs while receiving real-time video from the robot's onboard cameras directly into the headset. Preliminary tests in controlled environments show that this approach improves user situational awareness and responsiveness during remote monitoring tasks. Designed as a modular and low-cost prototype, the system aims to explore new interaction paradigms in immersive human-robot collaboration.

1 Introduction

Throughout history, humans have survived by adapting to changing environments. This has been made possible by the constant development of tools and strategies Boyd et al. (2011). Although many risks to life have been reduced by advances in technology, today's society still faces various risks. These events can be any kind of disaster, from natural (like floods or storms) to human-made (like accidents involving technology) Shi (2019). In these situations, advanced technologies, standard procedures and specific equipment are needed to protect the public and emergency response teams Zhu and Li (2021).

Over time, different tools and procedures have been introduced to reduce the risks to people without affecting the performance Zhu and Li (2021). In recent decades, robotics has become another tool for emergency response tasks. The way robots are designed depends on their shape as unmanned aerial vehicles, underwater or caterpillar robots. Lately, quadrupedal robots have gained attention because of their ability to move around complex surfaces more easily and with more stability than other types of robots. Tong et al. (2024).

But even with these advances, there are still limitations in how robots are controlled and how they interact with the operator. Conventional teleoperation interfaces, such as joysticks and monitors, limit perception and make telepresence less effective Schwenk and Smith (2025). Immersive interfaces, such as virtual reality (VR), offer more natural and intuitive control schemes, helping operators better understand the environment and act with greater accuracy Wan et al. (2024). In emergency situations, the portability of control equipment is very important. Traditional systems that use external screens need extra devices to be carried, a flat surface to be available, and electrical or network connections to be managed, which makes it hard for the operator to move around and slows down deployment Zhu and Li (2021). A virtual reality device like the Meta Quest 3 combines visualisation and interaction into one device, making it more portable and faster to use. A feature that can make a real difference in the effectiveness of the intervention.

This research work is motivated by the need to combine well-established technologies: quadrupedal robotics and immersive teleoperation. The aim is to improve and facilitate telepresence in emergency contexts in remote locations where screens may not be available. The main contribution of this article is the presentation of an integrated framework for the immersive teleoperation of a quadrupedal robot, applied to emergency contexts.

This research work does not aim to solve the problem of immersive teleoperation as a whole, but rather to demonstrate that it can be implemented in a portable and economical way, serving as a basis for future applications.

The remainder of this paper is structured as follows. Section 2 describes the robot and XR hardware and the software used. Section 3 describes the proposed approach theoretically. Section 4 shows how the experiments were configured for the present study. Section 5 presents the experimental evaluation. Finally, Section 6 discusses the conclusions and outlines future research directions.

2 Case Study

For this study, two different systems were integrated: a VR interface and a quadrupedal robot.

In this case, we used the Unitree Go1 Edu quadruped robot, which has four legs and 12 Degrees of Freedom (DoF). It reaches a maximum speed of 5 m/s and has an autonomy of 1-2 hours. It is equipped with five cameras: one at the front, jaw, belly and one on each side of the robot. All of which are wide-angle, as well as an IMU sensor.

It has been used because it has a good performance-cost ratio, it is versatile in irregular environments, in addition to having a good number of cameras to make a more immersive VR experience.

The VR interface is the Meta Quest 3, featuring a standalone HMD with passthrough, and 2064x2208 px per eye. Controllers with integrated joysticks and Wi-Fi 6 connectivity for real-time video streaming.

Godot Engine, an open-source video game engine that offers VR integration, was used as middleware. Its open and free approach makes it an accessible and easily replicable option, as well as facilitating collaboration and customisation without relying on commercial licenses. It is also lighter, which consumes fewer resources in VR.

3 Methods

The proposed system enables immersive teleoperation of the Unitree Go1 quadruped robot by combining real-time video streaming with joystick-based motion control inside a virtual reality (VR) environment. The architecture integrates three main components: the robot, a middleware layer, and the Meta Quest 3 headset running a custom XR application developed in the Godot 4.2 engine.

To provide the operator with a situationally aware view, the front and side cameras of the Unitree Go1 were used to create a panoramic feed. The robot transmits an H.264-encoded video stream using WebRTC. On the receiving side, the XR application decodes the stream, extracts still frames in real time, and maps them as textures within the Godot rendering pipeline. These textures are then displayed directly inside the VR headset, enabling an immersive first-person perspective.

Robot locomotion is controlled via the joysticks of the Meta Quest 3. Godot captures the input events and translates them into motion commands, which are sent to the middleware through a WebSocket channel. The left joystick controls linear translation (forward, backward, lateral), while the right joystick controls yaw rotation. The middleware forwards these commands to the Unitree Go1 through its native API, ensuring low-latency bidirectional communication.

To support remote operation without requiring local infrastructure, the middleware was made accessible through a secure network channel. A VPN service provided by the Instituto Tecnológico de Galicia (ITG) was used to bridge the operator in Ferrol and the robot deployed in A Coruña. The VPN employs asymmetric key-based authentication to ensure data confidentiality and integrity during teleoperation.

Given the critical role of responsiveness in teleoperation, the system incorporates latency monitoring at multiple stages. Control channel latency was measured using a ping-pong mechanism over WebSockets, recording timestamps for command issuance (T_0) and reception by the robot (T_1). Video latency was assessed by analyzing WebRTC internal metrics W3C (2025), decomposing the total one-way delay — from camera capture to display - into encoding, network, and receiver-side processing components Bray et al. (2024).

4 Experiment Setup

The experimental evaluation focused on validating the proposed teleoperation system in a realistic remote inspection scenario, inspired by the potential use of quadruped robots to support security services in risky environments.

Figure 1 illustrates the overall setup, structured into three layers: user, network, and robot. The operator was physically located in Ferrol, while the Unitree Go1 was deployed at the Advanced Services Centre (CSA) of the Instituto Tecnológico de Galicia (ITG) in A Coruña Cidade das TIC (2025). The environment included small static obstacles to test basic locomotion and remote situational awareness.

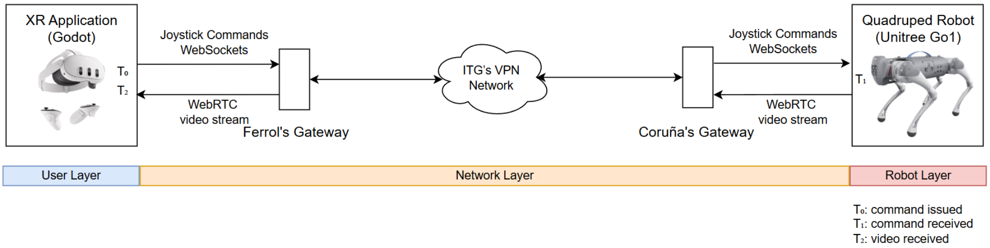

Figure 1: Experimental setup of the XR-based teleoperation system

The operator used a Meta Quest 3 headset running the XR application described in Section 3. The system communicated through the middleware over a secure VPN provided by ITG, which ensured encrypted and authenticated connectivity between both locations without additional on-site infrastructure.

System responsiveness was evaluated by collecting latency, jitter, and packet-loss metrics under controlled but realistic network conditions.

- Control latency was measured using the ping-pong mechanism introduced in Section 3, recording timestamps for command issuance (T_0) and reception at the robot (T_1).
- (T_2) was derived from WebRTC internal statistics, allowing decomposition of total one-way delay into encoding, network, and rendering components.

Each test run consisted of 200 samples, with the first 10 discarded as warm-up. Three independent runs were performed under similar network load. These measurements were carried out separately from the user evaluation sessions to avoid bias.

To complement the quantitative measurements, the system was tested by five external participants with no prior involvement in the project. Each session lasted about 15 minutes and focused on basic teleoperation and situational awareness tasks. Afterward, participants provided subjective feedback on perceived latency, control responsiveness, and overall sense of presence and usefulness.

5 Results

This section presents the experimental results obtained with the setup described in Section 4. Two key aspects were evaluated: (i) the end-to-end latency of the control channel, and (ii) the performance of the video streaming pipeline. Together, these metrics allow estimating the overall delay perceived by the operator during teleoperation.

5.1 End-to-end Control-Channel Latency

For each of the three test runs, 200 round-trip time (RTT) samples were collected over the WebSocket control channel (the first 10 samples of each run were discarded as warm-up). Figure 2 shows the RTT distributions per run, while Figure 3 illustrates the time evolution. Table 1 summarizes the main statistics.

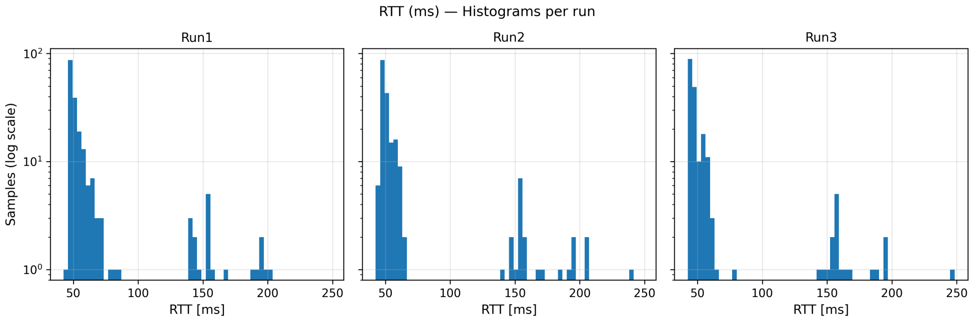

Figure 2: Histograms of RTT measured over the control channel (WebSockets) on the three test runs.

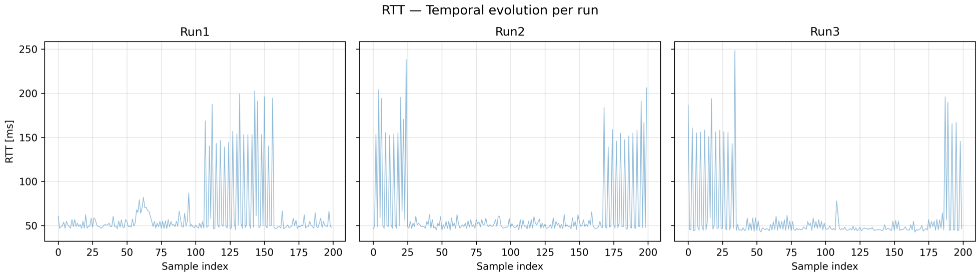

Figure 3: Time Series of net RTT measured over the control channel (WebSockets)

The average RTT ranged between 58-64 ms, with 95th percentiles around 153-158 ms and rare peaks up to 194-204 ms. The maximum observed RTT during the experiments was about 203-248 ms. As shown in Table 2, the majority of packets (50–70%, depending on the run) stayed below 50 ms, indicating generally low control latency, but with occasional jitter spikes above 150 ms.

Metric	Run1	Run2	Run3
Number of samples	200	200	200
Mean RTT	62.87	63.91	58.93
Standard deviation	34.23	38.38	35.82
Median RTT	49.85	49.98	46.41
95th percentile RTT	153.27	158.10	156.48
99th percentile RTT	196.06	204.19	193.91
Maximum observed	202.85	238.41	248.37

Table 1: Latency statistics summary for the three test runs

RTT Range	Run1	Run2	Run3
$<$ 50 ms	101 (50.5%)	101 (50.5%)	141 (70.5%)
50–100 ms	80 (40.0%)	77 (38.5%)	41 (20.5%)
100–150 ms	6 (3.0%)	3 (1.5%)	2 (1.0%)
150–200 ms	12 (6.0%)	16 (8.0%)	15 (7.5%)
$>$ 200 ms	1 (0.5%)	3 (1.5%)	1 (0.5%)

Table 2: Latency distribution by ranges

5.2 End-to-end video latency

Video delay was estimated from WebRTC internal statistics (jitter buffer occupancy, decoder time, and network RTT) combined with typical hardware encoder and display refresh values. Table 3 lists the parameters used to compute the one-way video latency.

Metric	Value	Source
JitterBufferDelay / EmittedCount	40-60ms	inbound-rtp
TotalDecodeTime / FramesDecoded	0.5-1.5ms	inbound-rtp
currentRoundTripTime	40-60ms	candidate-pair
VSync (display refresh)	16ms	assumed (60 Hz)
Encode (camera)	20ms	value H.264 HW

Table 3: Values used to calculate the total video latency

Using these values in Eq. 12.1, the typical one-way video delay was approximately 110 ms, dominated by encoding and network propagation. Under normal conditions, jitter buffering remained between 40-60 ms, but occasionally increased to 150 ms, leading to worst-case delays of about 200 ms.

$$L_{\text{video}} = L_{\text{encode}} + \frac{\text{RTT}}{2} + L_{\text{jitterbuffer}} + L_{\text{decode}} + L_{\text{vsync}} \tag{12.1}$$

5.3 Teleoperation round-trip latency

The overall delay perceived by the operator is the sum of the control RTT and the one-way video latency (as shown in Eq. 12.2).

$$L_{\text{op}} = RTT_{\text{control}} + L_{\text{video}} \tag{12.2}$$

Using the measured mean RTT ($\sim$ 60 ms) and video latency ($\sim$ 110 ms), the typical operator-perceived delay is about 170 ms. Under degraded network conditions (RTT up to 160 ms, video up to 200 ms), the delay can increase to 350−360 ms.

These values suggest that the system remains usable for exploratory teleoperation, though high jitter can affect precision maneuvers.

5.4 User Experience

Five external participants tested the system in 15-minute sessions. All users reported perceivable but manageable latency during remote operation (Ferrol–A Coruña over VPN), sufficient for navigation and exploration but challenging for fine alignment or passing through narrow spaces. In local tests (same LAN), participants perceived roughly one-third less latency, describing a smoother and more responsive experience. These observations align with the quantitative results: mean control RTT $\approx$ 60 ms and video latency $\approx$ 110 ms yield an end-to-end delay near 170 ms under typical conditions.

6 Conclusion and Future Work

This work presented a modular and affordable prototype for the immersive teleoperation of a quadruped robot using a standalone VR headset, with Godot as the integration layer and a differentiated communication stack for control and video. The proposed architecture proved viable for emergency telepresence and surveillance scenarios, maintaining control latencies in the tens of milliseconds and video delays around 110 ms. These values enabled basic remote navigation and situational awareness, fulfilling the initial goal of demonstrating that immersive teleoperation can be implemented in a portable and economical way. User evaluation supported these findings: in the remote setup, participants perceived noticeable latency that made fine maneuvers challenging but still allowed effective exploration; under local conditions, the perceived delay decreased by about two-thirds, resulting in a noticeably smoother and more responsive experience.

Nevertheless, the study also highlights several limitations. The performance of the system depends strongly on the quality and stability of wide-area network links, a factor that can be highly variable in real emergency deployments. The absence of advanced visual aids, such as multi-camera fusion, depth visualization, or obstacle overlays, limits situational awareness in complex environments. Moreover, the small sample size and the non-professional profile of participants restrict the generalization of the results to demanding operational contexts, underlining the need for broader and more controlled validation campaigns.

Future work will address these challenges to evolve the current proof-of-concept into an operational tool for emergency response teams. Planned improvements include enhancing the communication and video pipeline through adaptive encoding, latency-oriented congestion control, and dynamic buffering to reduce jitter and improve frame pacing. In parallel, XR perception will be enriched with multi-camera composition, depth cues, and obstacle highlighting to provide more reliable situational awareness. The integration of intelligent assistance — such as shared control, reactive obstacle avoidance, waypoint navigation, and viewpoint prediction — is expected to mitigate the impact of latency and reduce the operator's low-level workload.

Additionally, the system will be evaluated with professional users in scenarios representative of security and emergency response operations. This will include metrics such as cognitive load, presence, and task performance, as well as robustness tests in outdoor environments with redundant communication links (e.g., 5G/LTE combined with Wi-Fi), fail-safe policies, and cybersecurity considerations. Finally, ergonomic aspects such as interface comfort, HUD configuration, and lightweight haptic feedback will be explored to further improve immersion and control precision.

These developments aim to transform the presented prototype into a safer, more efficient, and operationally ready immersive teleoperation system, capable of reducing risk and enhancing situational awareness for emergency and security personnel in real deployments.

Acknowledgements

This project is funded by the European Union (Scalable Platform for Innovations on Real-time Immersive Telepresence (SPIRIT), 101070672). This work has also received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI) and the UK Research and Innovation (UKRI).

Anabel Díaz Labrador's research was supported by the Xunta de Galicia (Regional Government of Galicia in Spain), through grants to industrial PhD (*http://gain.xunta.gal/*), under the "Doutoramento Industrial 2024" grant with reference: *02_IN606D_2024_3100897*.

CITIC, as a Research Center of the University System of Galicia, is funded by Consellería de Educación, Universidade e Formación Profesional of the Xunta de Galicia through the European Regional Development Fund (ERDF) and the Secretaría Xeral de Universidades (Ref. ED431G 2019/01).

Xunta de Galicia. Grants for the consolidation and structuring of competitive research units, GPC (ED431B 2023/49).

Bibliography

R. Boyd, P. J. Richerson, and J. Henrich. The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, 108(supplement_2): 10918–10925, 2011. URL *https://www.pnas.org/doi/abs/10.1073/pnas.1100290108*.

N. Bray, M. Boeding, M. Hempel, H. Sharif, T. Heikkilä, M. Suomalainen, and T. Seppälä. A latency composition analysis for telerobotic performance insights across various network scenarios. *Future Internet*, 16(12), 2024. URL *https://www.mdpi.com/1999-5903/16/12/457*.

Cidade das TIC. Advanced services center (csa) cidade das tic. *https://www.cidadedastic.gal/centro-de-servicios-avanzados/*, 2025. Accessed: September 21, 2025.

R. Schwenk and S. Smith. An immersive spatially consistent multi-modal augmented virtuality human-machine interface for telerobotic systems. *Computers in Industry*, 167:104260, 2025. URL *https://www.sciencedirect.com/science/article/pii/S0166361525000259*.

P. Shi. Hazards, disasters, and risks. In *Disaster Risk Science*, pages 1–48. Beijing Normal University Press and Springer Nature Singapore Pte Ltd, Singapore, 2019.

K. Tong, Y. Hu, B. Dikic, S. Solmaz, F. Fraundorfer, and D. Watzenig. Robots saving lives: A literature review about search and rescue (sar) in harsh environments. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 953–960, 2024.

W3C. Identifiers for webrtc's statistics api. *https://www.w3.org/TR/webrtc-stats/*, 2025. Accessed: 2025-09-26.

K. Wan, C. Li, F.-S. Lo, and P. Zheng. A virtual reality-based immersive teleoperation system for remote human-robot collaborative manufacturing. *Manufacturing Letters*, 41:43–50, 2024. URL *https://www.sciencedirect.com/science/article/pii/S2213846324000658*. 52nd SME North American Manufacturing Research Conference (NAMRC 52).

Y. Zhu and N. Li. Virtual and augmented reality technologies for emergency management in the built environments: A state-of-the-art review. *Journal of Safety Science and Resilience*, 2(1): 1–10, 2021. URL *https://www.sciencedirect.com/science/article/pii/S266644962030030X*.