Human-robot interaction for robotic inspections based on mixed reality

S. Sarida, J. Brokmeier, & K. Reguieg
Roboverse Reply, Munich, Germany

J. Stührenberg & K. Smarsly
Hamburg University of Technology, Hamburg, Germany

ABSTRACT: Mobile robots are increasingly being adopted for automated construction and industrial inspection to enhance operational efficiency and reliability. In particular, quadruped robots are deployed because of advanced locomotion techniques, providing enhanced maneuverability and adaptability, as compared to wheeled robots. In dynamic environments, inspection tasks must be adapted manually. Until now, teaching of inspection tasks has been a laborious process requiring skilled operators. In this study, a human-robot interaction (HRI) framework based on mixed reality (MR) is presented to support teaching of inspection tasks, thereby easing the recording of inspection missions. The HRI framework enables operators wearing an MR headset to teleoperate robots intuitively and hands-free, while interacting with the environment. The HRI framework allows to teach inspection tasks user-friendly and effective and to adapt inspection missions to dynamic environments. The HRI framework is implemented on the Microsoft HoloLens 2 MR headset and the Spot quadruped robot. Validation tests conducted in an indoor office environment highlight the intuitive and hands-free control capabilities of the HRI framework and the user-friendly and effective teaching of inspection tasks.

1 INTRODUCTION

Civil and industrial infrastructure requires routine inspections to maintain safety standards. Manual inspections can be time-consuming, labor-intensive, costly, potentially hazardous, and may suffer from inconsistencies, owing to repetitive tasks, declining productivity, and subjective judgment (Halder & Afsari 2023). Moreover, the construction industry faces labor shortage, escalating labor costs, and low productivity levels (Barbosa et al. 2017), which further hinder the frequency and efficiency of manual inspections. To enhance the efficiency and reliability of construction and industrial inspection processes, the adoption of automated inspections using robots is a promising direction (Smarsly & Dragos 2024). Specifically, quadruped robots are increasingly being deployed for automated inspections, due to advanced locomotion techniques offering improved maneuverability and adaptability to different environments (Smarsly et al. 2023).

Typically, robotic inspection processes involve operators remotely controlling robots to navigate through dynamic infrastructure and to collect sensor data (Delmerico et al. 2022). The sensor data is analyzed by trained inspectors. However, remote robot control is resource-intensive and dependent on the skills and availability of specialized operators. By contrast, automated inspection processes include specialized operators teaching robots to perform inspection tasks, enabling the robots to repeat the inspection tasks. Automated inspection processes are efficient in static or less dynamic environments, but dynamic and constantly changing environments, such as construction sites, are challenging for automated inspection processes. In dynamic environments, mixed reality (MR) solutions may bridge the gap by accommodating new environments and enable operators, even less experienced operators, to control the robots (Holly et al. 2022).

Several studies have investigated controlling quadruped robots using MR. In Cruz Ulloa et al. (2023), for example, an MR system to assist rescuers during search and rescue operations is presented. The MR system allows operators to control quadruped robots and manage visual sensor data in post-disaster settings, alerting the operators upon victim detection. In Delmerico et al. (2022), Azure Spatial Anchors (ASA) are utilized to design an inspection mission using the Spot quadruped robot. ASA enable the attachment of digital content, stored in the cloud, to specific physical world points in MR. Human operators, equipped with an MR headset, place holographic markers for Spot to follow during the inspection missions. In Quesada & Demiris (2022), ASA are used to align the coordinate frames of Spot and the MR headset, allowing to converse between the frames. However, ASA require cloud connectivity and a pre-

mapped environment. Cloud connectivity may not be guaranteed in shielded and remote environments, and pre-mapping may be cumbersome and not feasible in dynamic environments.

To bridge the gap between adapting to changing environments and enabling operators with less expertise to control robots, this study presents an MR-based human-robot interaction (HRI) framework for construction and industrial inspection. The MR-based HRI framework is designed for intuitive and hands-free teleoperation of quadruped robots. Operators interact in real-time with virtual models of quadruped robots, augmented in the real world, using an MR headset to provide goal poses and to command further actions to the robots. The HRI framework allows to adapt inspection tasks effectively and user-friendly to changing environments and provides operators with useful information about the robots and inspection results displayed on MR headsets.

The remainder of the paper is organized as follows. In the next section, Section 2, the methodology and implementation of the HRI framework is described. In Section 3, tests in an office environment are conducted to validate the HRI framework and results are given and discussed. Section 4 summarizes and concludes the paper.

2 A HUMAN-ROBOT INTERACTION FRAMEWORK BASED ON MIXED REALITY

This section provides a detailed overview of the HRI framework, proposed to facilitate intuitive and hands-free control of quadruped robots and user-friendly and effective teaching of inspection missions using MR headsets. In the first subsection, the coordinate alignment of MR headsets and quadruped robots is described. Subsequently, in the second subsection, robot actions devised to record and perform inspection missions are covered. The third subsection closes with information on the implementation of the HRI framework.

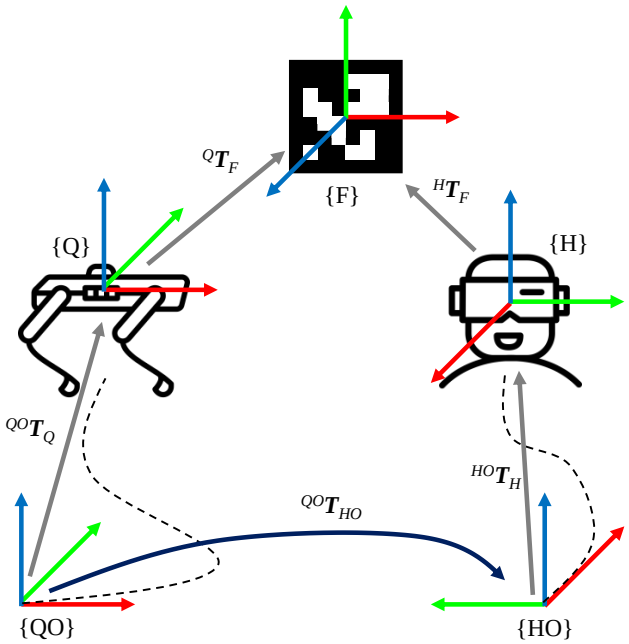
2.1 *Coordinate alignment of mixed reality headsets and quadruped robots*

Controlling robots in MR requires shared reference frames between robots and MR headsets. The shared reference frames are established through colocalization. The MR-based framework presented in this study employs fiducial markers for colocalization. By observing a fiducial marker with cameras embedded in both an MR headset and a robot, a coordinate transform from the MR headset frame to the robot frame is obtained. The coordinate transform allows translating goal poses commanded in the MR headset frame to the quadruped robot frame, and translating spatial information obtained from the quadruped robot to the MR headset frame.

The coordinate alignment process is illustrated in Figure 1. The quadruped robot frame {Q} is fixed to the quadruped robot, and the MR headset frame {H} is fixed to the MR headset. At startup, both the quadruped robot and the MR headset establish static coordinate frames at their starting locations, denoted as {QO} and {HO}, respectively. While moving, both devices consult sensor data to compute the odometry, i.e., to estimate the current pose in the respective static coordinate frame. The current pose of the quadruped robot, with respect to its origin frame, is denoted as the dynamic transformation ${}^{QO}\boldsymbol{T}_{Q}$. The current pose of the MR headset with respect to its origin frame is denoted as the dynamic transformation ${}^{HO}\boldsymbol{T}_{H}$. If both cameras (embedded in the quadruped robot and the MR headset) are facing and detecting the same fiducial tag aligned with the fiducial tag frame {F}, a static transformation between the coordinate frames can be computed using Equation 1,

$$ {}^{QO}\boldsymbol{T}_{HO} = {}^{QO}\boldsymbol{T}_{Q}\ {}^{Q}\boldsymbol{T}_{F}\ {}^{H}\boldsymbol{T}_{F}^{-1}\ {}^{HO}\boldsymbol{T}_{H}^{-1}, \tag{1} $$

where ${}^{QO}\boldsymbol{T}_{HO}$ denotes the static transformation from the origin of the quadruped robot frame to the origin of the MR headset frame, ${}^{H}\boldsymbol{T}_{F}$ denotes the dynamic transformation from the MR headset frame to the fiducial tag frame, and ${}^{HO}\boldsymbol{T}_{H}$ denotes the dynamic transformation from the origin of the MR headset frame to the MR headset frame. The static transformation ${}^{QO}\boldsymbol{T}_{HO}$ allows translating coordinates and poses between the quadruped robot and the MR headset.



Frames:

{QO}	Origin of quadruped robot	{HO}	Origin of MR headset
{Q}	Quadruped robot	{H}	MR headset
{F}	Fiducial tag		

Figure 1. Coordinate alignment of the quadruped robot and the MR headset.

2.2 *Robot actions*

For intuitive, hands-free teleoperation of robots, substituting the traditional remote control, and for user-friendly and effective recording of inspection missions, four actions, described in the following paragraphs, are employed in the HRI framework, namely

i. *walk to goal*,
ii. *follow MR headset*,
iii. *record mission*, and
iv. *replay mission*.

The *walk to goal* action is used to command single goal poses intuitively to robots. Operators wearing the MR headset interact with a virtual model of the robot and place it to a desired location. Upon confirmation, the real robot moves to match the goal pose of the virtual robot.

The process of commanding goal poses to a quadruped robot in the *walk to goal* action is visualized in Figure 2. Operators wearing the MR headset interact with the virtual model of the quadruped robot, attached with the goal pose frame {G}, in multiple ways. Specifically, operators may use far interaction pointers to select locations, and grab and position the virtual robot model and adjust the model with both hands as desired. The virtual model of the quadruped robot is referenced in the origin of the MR headset frame {HO}, denoted by the transformation ${}^{HO}\boldsymbol{T}_G$. Upon confirming the goal pose visualized by the virtual model of the robot, the goal pose is transformed to the origin of quadruped robot frame {QO}, denoted by ${}^{QO}\boldsymbol{T}_G$, using Equation 2:

$$ {}^{QO}\boldsymbol{T}_G = {}^{QO}\boldsymbol{T}_{HO}\ {}^{HO}\boldsymbol{T}_G \tag{2} $$

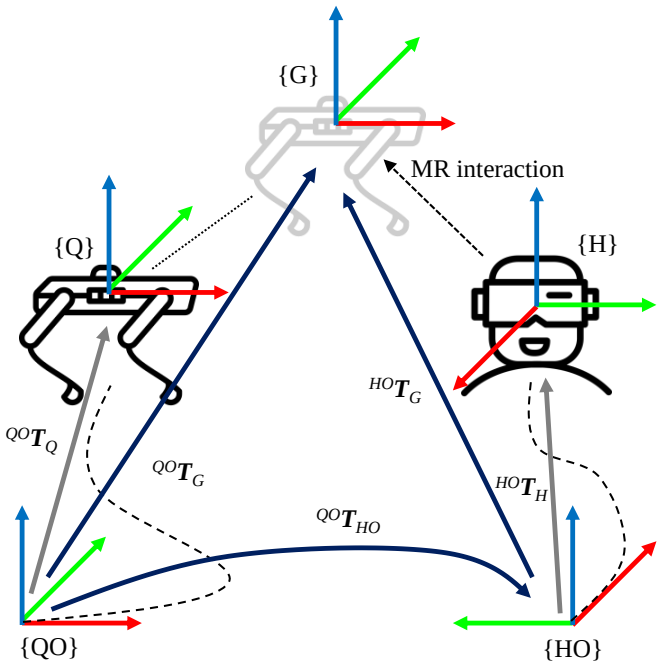
With the goal pose being defined in the origin of the quadruped robot frame, the quadruped robot plans a collision-free path from its current pose to the goal pose and follows the path.

In the *follow MR headset* action, robots follow the operator wearing the MR headset while keeping a determined distance. The *follow MR headset* action offers hands-free control of the quadruped robot and requires nothing but “start” and “stop” inputs from the operator. To follow the operator, the quadruped robot computes the transformation from the origin of quadruped robot frame {QO} to the MR headset frame {H}, denoted as ${}^{QO}\boldsymbol{T}_H$, by leveraging the coordinate alignment ${}^{QO}\boldsymbol{T}_{HO}$:

$$ {}^{QO}\boldsymbol{T}_H = {}^{QO}\boldsymbol{T}_{HO}\ {}^{HO}\boldsymbol{T}_H \tag{3} $$

The quadruped robot plans its movement accordingly and maintains a fixed distance between its position and the position of the MR headset, both defined in the origin of quadruped robot frame.

The *record mission* action can be combined with the *walk to goal* and *follow MR headset* actions. During mission recording, robots record a map of the environment and log multiple goal poses and inspection tasks, such as collecting sensor data or taking photos.



Frames:

{QO}	Origin of quadruped robot	{HO}	Origin of MR headset
{Q}	Quadruped robot	{H}	MR headset
{G}	Goal pose		

Figure 2. Commanding goal poses to the quadruped robot in the *walk to goal* action.

Finally, the *replay mission* action allows to replay missions previously recorded using the *record mission* action. The *replay mission* action facilitates performing automated robotic inspections.

2.3 *Implementation*

The HRI framework is implemented on the Microsoft HoloLens 2 (HL2) MR headset (Microsoft 2024) and the quadruped robot Spot from Boston Dynamics (Boston Dynamics 2020). The Robot Operating System (ROS) (Quigley et al. 2009) is used to handle communication between the HL2 and Spot. The ROS tf2 library is employed to calculate the transformations.

A user interface (UI) application for controlling Spot with the HL2 and displaying information relevant to the operator is implemented and designed in the Unity game engine, using the Microsoft Mixed Reality Toolkit (MRTK) (MixedRealityToolkit 2024). After starting the application, the operator is prompted to enter an IP address to connect to a Spot. Upon successfully establishing connection between the HL2 and Spot the operator is prompted to execute the coordinate alignment process.

The coordinate alignment utilizes ArUco marker for fiducial tags (Garrido-Jurado et al. 2014). It is to be noted that Unity employs left-handed y-up z-forward coordinate systems, while ROS employs right-

handed z-up x-forward coordinate systems. Therefore, the pose of ArUco markers detected from the HL2 are transformed to a right-handed coordinate system for further computations. To guide the operator in the coordinate alignment process, visual confirmations of the ArUco marker detected by Spot and the HL2 are displayed to the operator, as illustrated in Figure 3. The video stream of the camera of Spot is displayed to the operator including a coordinate system, if an ArUco marker is detected. The coordinate system corresponds to the transformation ${}^{Q}\boldsymbol{T}_{F}$ and should lie in the middle of the ArUco marker. If the ArUco marker is detected from the HL2, a virtual marker is overlaid to match the real ArUco marker, corresponding to the transformation ${}^{H}\boldsymbol{T}_{F}$. The operator confirms the coordinate alignment, if both markers are detected accurately, resulting in computing ${}^{QO}\boldsymbol{T}_{HO}$ according to Equation 1. Upon confirming the coordinate alignment, the operator is displayed with the main menu of the UI application.

The main menu of the UI application shows useful information of Spot, such as connectivity status, current battery charge percentage, and estimated battery runtime. Furthermore, the main menu contains buttons to execute the four robot actions described in Section 2.2.

The actions are implemented on Spot using the Spot SDK (Boston Dynamics 2024). The Spot SDK facilitates the teleoperation, localization, mapping, navigation, and autonomy functionalities for Spot. On the MR headset, each action opens a UI panel with short instructions, options, and visual feedback to the operator. An example of the *walk to goal* action is given in Figure 4, where the virtual model of Spot, i.e. the goal, stands next to the real Spot. The *follow MR headset* action UI panel includes two buttons, one to start the MR headset following, and one to stop the MR headset following. The *record mission* action UI panel includes three buttons. The first button starts a new mission recording, including recording a map, tracking actions and locations of the actions. The second button ends the mission recording and saves the map, actions, and locations. The third button opens a menu of all possible actions during the mission recording. Possible actions are the *walk to goal* action, the *follow MR headset* action, collecting sensor data for inspection purposes, and inbuilt functions of Spot, such as to sit and stand. The *replay mission* action offers two buttons, one to load missions previously recorded using the *record mission* action, and one to replay the mission loaded by the first button.

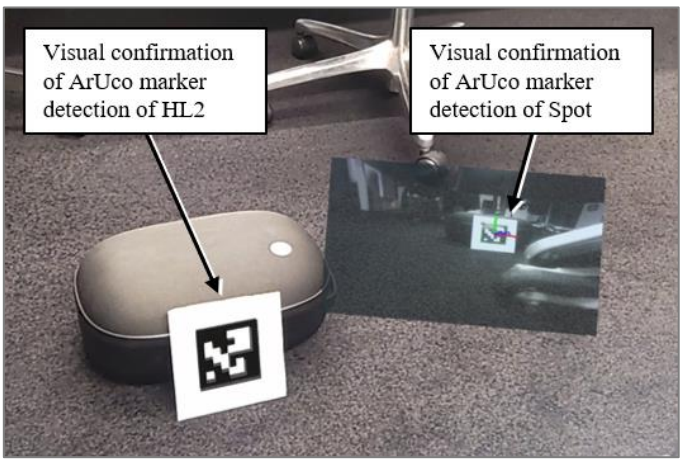


Figure 3. ArUco marker detection for coordinate alignment.

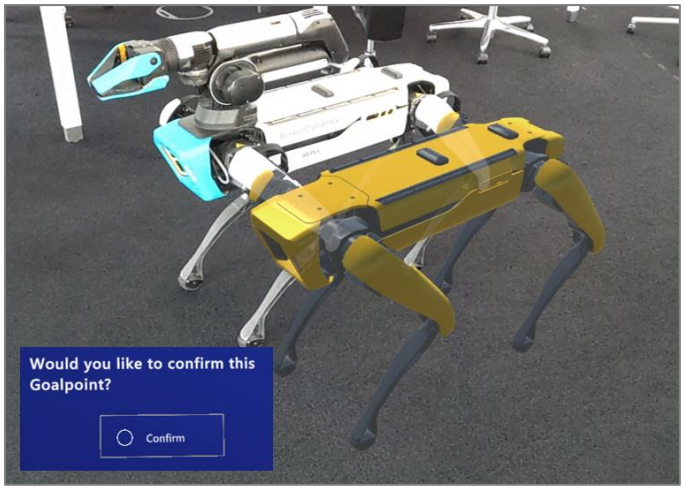


Figure 4. Virtual Spot (front right) representing a goal pose for the real Spot (back left) in the *walk to goal* action.

3 VALIDATON AND RESULTS

This section describes the validation tests and the test results, to assess the accuracy of the HRI framework. The accuracy is influenced by the coordinate alignment step described in Section 2.1. Furthermore, the intuitiveness and effectiveness of the HRI framework is discussed.

The validation tests are conducted in a dynamic indoor office environment with multiple people walking through the corridors and offices. First, an operator wearing the HL2 conducts the coordinate alignment step. Then, the operator starts the *record mission* action. During the action, Spot is commanded to navigate to three goal poses using the *walk to goal* action. Ground truth of the goal poses is provided by cardboard boxes cut to the dimensions of the outline of Spot, devised to fix the goal poses in the real world. Next, the operator places the virtual Spot accurately to align with the edges of the cardboard boxes. Then, offsets of the real Spot assuming the goal poses are measured with a tape measure from the central edge of the cardbox to the center of the head of the real Spot. In between commanding the goal poses, Spot is commanded to follow the operator wearing the HL2 using the *follow MR headset* action. After teaching the three goal poses, the mission recording is stopped, saving the goal poses and the map of the environment in a mission. The mission is subsequently replayed by Spot six times using the *replay mission* action. During mission replay, the accuracy of the goal poses is

evaluated by extracting coordinates from the odometry of Spot at the goal poses.

The map recorded during the *record mission* action including the three goal poses {G1}, {G2}, and {G3} is shown in Figure 5. Figure 6 illustrates the alignment of the real Spot and the virtual Spot at {G1}. The offsets measured are presented in Table 1, given in the respective {G} coordinate frames.

Table 1. Offsets of the real Spot to the virtual Spot at the goal poses during mission recording.

	{G1}		{G2}		{G3}	
	x [m]	y [m]	x [m]	y [m]	x [m]	y [m]
Offset	0.027	0.042	0.035	0.040	0.025	0.037

The results of the six *replay mission* actions are presented in Table 2 and visualized in Figure 7. The coordinates at the goal poses are given in the origin of quadruped robot frame {QO} and are extracted from the odometry of Spot ${}^{QO}\boldsymbol{T}_{Q}$. The validation test results in a mean distance offset of 0.407 m from the ground truth with a standard deviation of 0.196 m.

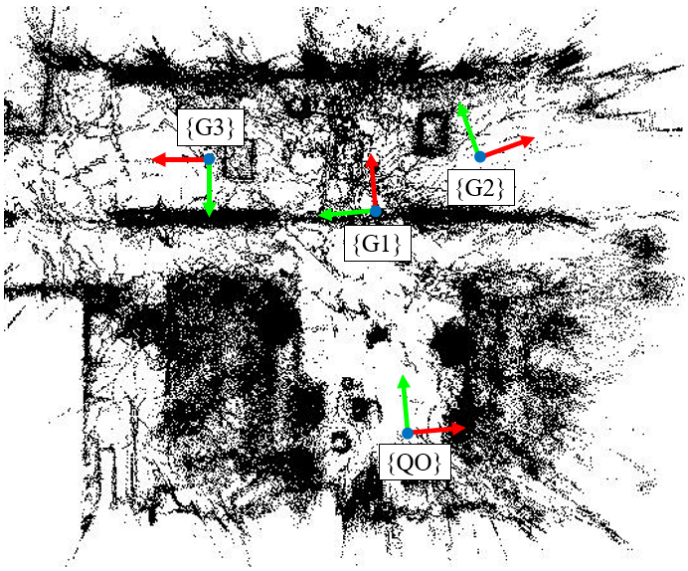


Figure 5. Map with goal poses recorded during the *record mission* action.

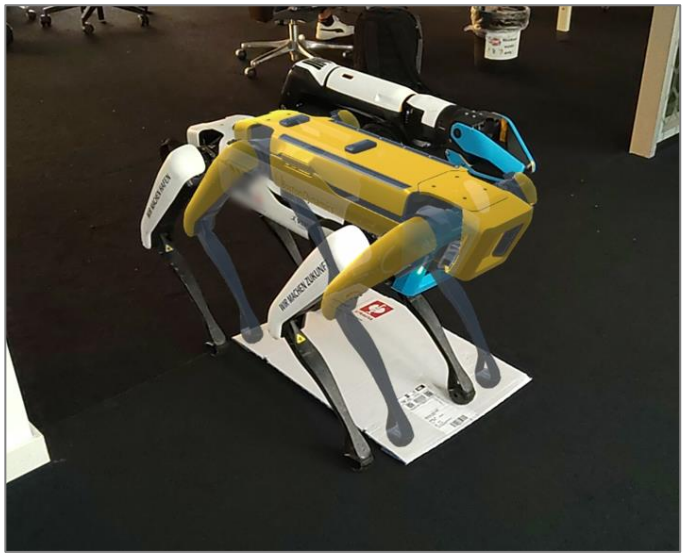


Figure 6. Alignment of the virtual and the real Spot at {G1} during mission recording.

Table 2. Coordinates of Spot at the goal poses in the mission replays.

	{G1}		{G2}		{G3}	
	x [m]	y [m]	x [m]	y [m]	x [m]	y [m]
Truth	0.015	4.755	1.590	4.913	-2.218	5.527
Replay 1	-0.034	4.935	1.579	5.064	-2.255	5.576
Replay 2	-0.130	5.080	1.598	5.153	-2.181	5.645
Replay 3	-0.329	5.150	1.759	5.298	-2.168	5.790
Replay 4	0.122	5.266	1.739	5.401	-2.076	5.836
Replay 5	-0.049	5.358	1.656	5.576	-2.148	5.970
Replay 6	0.008	5.431	1.661	5.617	-2.166	6.039

In summary, the validation test results show offsets in the goal poses commanded by the operator wearing the HL2 and the goal poses assumed by Spot. The offsets derive from multiple factors, described in the following two paragraphs.

The offsets are affected by the accuracy of the ArUco marker detection, which in turn is affected by lighting, the size of the marker, the distance to the marker, and movement. Head movement cannot be fully prevented when confirming the coordinate alignment, which may affect the ArUco marker detection accuracy from the HL2. Furthermore, the odometry of the HL2 and Spot exhibit small drift over time. The effect of drift is amplified by computing the odometry in two different frames, {QO} for Spot and {HO} for the HL2, which are connected by the

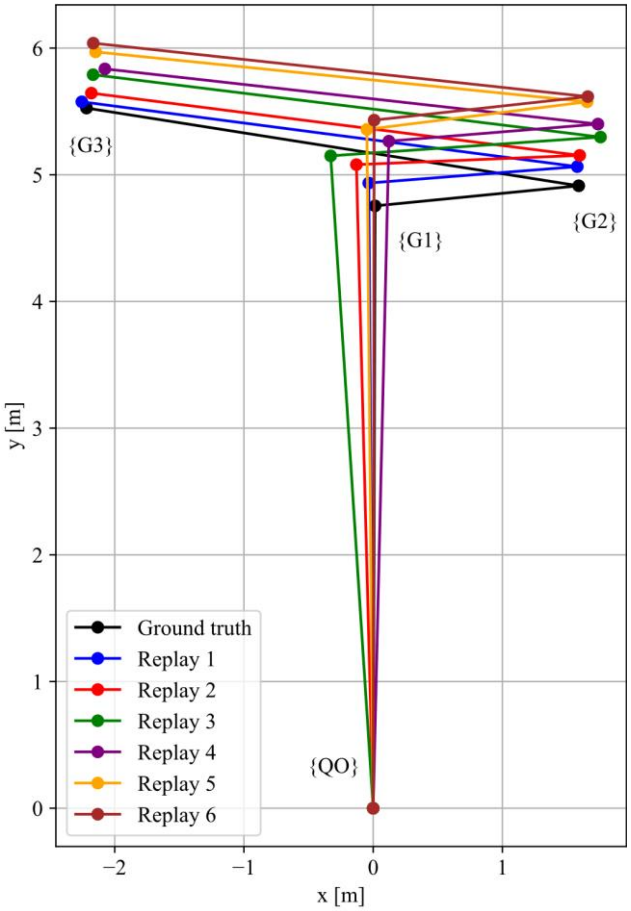


Figure 7. Results of the replay mission validation tests.

transformation ${}^{QO}\boldsymbol{T}_{HO}$ computed at the location of the ArUco marker. Errors in orientation of ${}^{QO}\boldsymbol{T}_{HO}$ result in offsets accumulating with increasing distance to the ArUco marker.

The offsets in the mission replays may result from the map recorded shown in Figure 5, which exhibits significant noise. The noise partly originates from dynamic objects, such as persons moving around during map recording. Furthermore, the map is created using camera data, which can be challenging in office environments with featureless white walls and glass walls, respectively. Facilitating an additional light detection and ranging (LiDAR) sensor may improve the map recording process and map quality. The map quality directly impacts the localization and navigation accuracy of Spot.

Regarding the goals of implementing intuitive and user-friendly capabilities, from the subjective experience of the operator during the mission recording, the HRI framework offers intuitive and hands-free control of Spot. The HRI framework allows recording inspection missions user-friendly and effective. Pointing with fingers or dragging and dropping the virtual model to goal poses provides a more intuitive control method compared to traditional remote controls. The *follow MR headset* action allows the operator hands-free teleoperation. Visual confirmations in the UI application inform the operator about the success or failure of the coordinate alignment and the execution of actions.

4 SUMMARY AND CONCLUSIONS

To maintain safety standards, civil and industrial infrastructure requires periodic inspections. Quadruped robots are increasingly adopted to perform automated inspections to enhance the efficiency and reliability of infrastructure. Dynamic environments present challenges to automated robotic inspections, requiring skilled operators to modify the inspection missions.

In this study, an MR-based HRI framework has been presented, easing the teleoperation of quadruped robots and the recording of inspection missions. Operators wearing an MR headset command goal poses to quadruped robots by interacting with a virtual model of the robot, record and replay inspection missions, and receive information from the robot displayed in an UI application. Commanding goal poses is enabled by a coordinate alignment step, where the detection of a fiducial tag by the quadruped robot and the MR headset allows computing a transformation between the two devices. Validation tests have been devised, highlighting the intuitive and hands-free control of quadruped robots and user-friendly and effective recording of inspection missions.

In conclusion, the HRI framework offers an intuitive, user-friendly, and effective solution for controlling robots and recording inspection missions, eliminating the dependency on hand-held remote controls. However, as has been unveiled in the validation tests, offsets between the goal poses commanded by the operator and the goal poses assumed by the quadruped robot are observed. The accuracy of assuming the goal poses by the quadruped robot is satisfactory in vicinity of the fiducial tag during the mission recording, but replaying missions show inaccuracies in the localization of the quadruped robot. Future work may address correcting the offsets of the commanded and assumed goal poses to increase the accuracy of the HRI framework. Employing a LiDAR sensor on the quadruped robot is believed to improve map and localization quality.

5 REFERENCES

Barbosa, F., Woetzel, L., Mischke, J., Ribeirinho, M.J., Sridhar, M., Parsons, M., Bertram, N., & Brown, S. 2017. Reinventing construction: A route to higher productivity. *McKinsey Global Institute report, February 2017*.

Boston Dynamics, 2020. Spot specifications. Available online: https://bostondynamics.com/wp-content/uploads/2020/10/spot-specifications.pdf (accessed on 22 March 2024).

Boston Dynamics, 2024. spot-sdk. Available online: https://github.com/boston-dynamics/spot-sdk (accessed on 18 March 2024).

Cruz Ulloa, C., Del Cerro, J., & Barrientos, A., 2023. Mixed-reality for quadruped-robotic guidance in SAR tasks. *Journal of Computational Design and Engineering* 10(4): 1479-1489.

Delmerico, J., Poranne, R., Bogo, F., Oleynikova, H., Vollenweider, E., Coros, S., Nieto, J., & Pollefeys, M. 2022. Spatial computing and intuitive interaction: Bringing mixed reality and robotics together. *IEEE Robotics & Automation Magazine* 29(1): 45-57.

Garrido-Jurado, S., Muñoz-Salinas, S., Madrid-Cuevas, F.J., Marín-Jiménez, M.J. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47(6): 2280-2292.

Halder, S. & Afsari, K. 2023. Robots in inspection and monitoring of buildings and infrastructure: A systematic review. *Applied Sciences* 13(4): 2304.

Holly, F., Zigart, T., Maurer, M., Wolfartsberger, J., Brunnhofer, M., Sorko, S.R., Moser, T., & Schlager, A. 2022. Gaining impact with mixed reality in industry – a sustainable approach. In *Proceedings of the 2022 8th International Conference on Computer Technology Applications, Vienna, Austria, 12-14 May 2022*.

Microsoft, 2024. HoloLens 2. Available online: https://www.microsoft.com/en-us/hololens/hardware (accessed on 22 March 2024).

MixedRealityToolkit, 2024. MixedRealityToolkit-Unity. Available online: https://github.com/MixedRealityToolkit/MixedRealityToolkit-Unity (accessed on 22 March 2024).

Moniruzzaman, M.D., Rassau, A., Chai, D., & Islam, S.M.S. 2022. Teleoperation methods and enhancement techniques for mobile robots: A comprehensive survey. *Robotics and Autonomous Systems* 150: 103973.

Quesada, R.C. & Demiris, Y. 2022. Holo-SpoK: Affordance-aware augmented reality control of legged manipulators. In *Proceedings of the 2022 IEEE/RSJ International Conference*

on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23-27 October 2022.

Quigley, M., Gerkey, B., Conley, K., Faust, J., Footey, T., Leibs, J., Berger, E., Wheeler, R., Ng, A. 2009. ROS: An open-source Robot Operating System. In *Proceedings of the International Conference on Robotics and Automation, Kobe, Japan, 12-17 May 2009.*

Smarsly, K., Dragos, K., Stührenberg, J., & Worm, M., 2023. Mobile structural health monitoring based on legged robots. *Infrastructures* 8(9): 136.

Smarsly, K., & Dragos, K., 2024. Advancing civil infrastructure assessment through robotic fleets. *Internet of Things and Cyber-Physical Systems* 4: 138-140.