2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

3DGS-Holo-Inspector: A Mixed Reality UAV Controller with 3D Gaussian Splatting Localization for Infrastructure Inspection

Syed Muhammad Raza Rizvi, Huaiyuan Weng, and Chul Min Yeum

Video: https://youtu.be/jKusI9VYvzU

***Abstract*— Unmanned aerial vehicles (UAVs) are increasingly used for infrastructure inspection, but conventional joystick and first-person-view (FPV) controllers remain unintuitive, error-prone, and cognitively demanding, particularly in cluttered or safety-critical environments. We present 3DGS-Holo-Inspector, a Mixed Reality (MR) UAV controller that combines holographic goal-setting with autonomous UAV navigation. Using natural hand gestures, operators can define and preview navigation goals directly in MR before flight, ensuring precise and safe data capture at inspection viewpoints. The system complements existing inspection pipelines by leveraging pre-built 3D maps (e.g., photogrammetry or LiDAR reconstructions) to enable refinement of Regions of Interest (ROIs) where coverage is incomplete or the detail is insufficient. Robust headset–UAV alignment is achieved through a LiDAR–RGB 3D Gaussian Splatting (3DGS) localization backbone, which provides dense, markerless, and persistent spatial registration in both indoor and outdoor settings. Once goals are placed, the UAV autonomously navigates to the specified pose, with real-time telemetry and live video overlaid in MR to enhance situational awareness. Experimental validation using a ModalAI Starling UAV and Microsoft HoloLens 2 demonstrated accurate UAV-goal alignment, achieving a positional Root Mean Square Error (RMSE) of 0.090 m (median = 0.084 m) indoors and 0.119 m (median = 0.118 m) outdoors, with orientation (yaw) RMSEs of 1.491° (median = 1.400°) and 2.233° (median = 2.268°), respectively. These results confirm that 3DGS-Holo-Inspector provides reliable MR-based UAV control, augmenting inspection workflows by enabling safe, intuitive, and high-precision UAV operations in real-world environments.**

***Index Terms*— UAVs, Mixed Reality, Gaussian Splatting, Autonomous Navigation, Localization, Infrastructure Inspection**

I. Introduction

Vision-based inspection has become a cornerstone of Structural Health Monitoring (SHM), as deterioration in civil infrastructure often manifests through visual cues such as cracks, corrosion, and deformation [1, 2]. Advancements in computer vision have enabled the automated detection and quantitative assessment of such defects from perception data with high accuracy [3, 2]. Coupled with unmanned aerial vehicles (UAVs), vision-based methods enable automated defect detections on bridges, wind turbines, pipelines, and power lines, reducing both risk and cost compared to manual surveys [4, 5]. Studies report that UAV-based inspections can be up to three times faster than traditional methods while eliminating the need for scaffolding, lifts, or rope access [6, 7]. These capabilities have accelerated adoption in safety-critical domains, further supported by initiatives such as the U.S. Drone Infrastructure Inspection Grant Act [8] and expanded Federal Aviation Administration (FAA) Beyond Visual Line of Sight (BVLOS) exemptions [9].

Despite these benefits, UAV-based inspection remains bottle-necked by limitations of existing piloting interfaces. Traditional remote controllers and first-person-view (FPV) systems require extensive training and mostly provide only 2D visual feedback [10]. The lack of depth perception complicates distance estimation, while frequent orientation changes of the UAV confuse the operator about the UAV's current state, often resulting in crashes or inconsistent coverage [11, 12]. These challenges are particularly acute in cluttered or confined environments, where precise navigation is essential. As UAV adoption scales, there is a growing need for intuitive, spatially aware control interfaces that ensure safety and reliable data capture.

Mixed Reality (MR) offers a compelling solution by merging digital overlays with the physical environment. MR has been applied in construction, manufacturing, and inspection for tasks such as collaborative design, sensor data visualization, and real-time 3D annotations [13, 14]. For UAVs, MR interfaces allow pilots to place navigation goals holographically in 3D space, preview planned paths, and monitor telemetry in situ. Prior systems such as PinpointFly [10], FlyAR [15], and Drone-Augmented Vision [16] have demonstrated the feasibility of holographic goal placement. However, these solutions often depend on external markers or additional infrastructure, limiting robustness and scalability.

In this work, we introduce **3DGS-Holo-Inspector**, a MR UAV control system that bridges holographic goal-setting with autonomous navigation. The system complements existing inspection workflows that generate 3D reconstructions of infrastructure from photogrammetry or LiDAR for inspection purposes. Such reconstructions often lack sufficient coverage in areas requiring high detail, such as joints, cracks, or edges. Our system leverages these pre-built maps to enable focused data collection in such Regions of Interest (ROIs). Operators define and preview holographic goals in MR, which are then transformed into UAV control commands through a 3D Gaussian Splatting (3DGS) [17] localization backbone that provides dense, markerless, and persistent Spatial Alignment (SA) between the headset and the UAV. The UAV autonomously executes these goals, while real-time telemetry and live video streams are overlaid in MR to enhance situational awareness.

We acknowledge the support from Rogers Communications and the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors are with the Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. (Emails: smr4rizv@uwaterloo.ca, huaiyuan.weng@uwaterloo.ca, cmyeum@uwaterloo.ca.

Experimental validation was conducted using a ModalAI Starling UAV [18] and Microsoft HoloLens 2 (HL2) [19] in both indoor and outdoor environments. Augmented Reality University of Cordoba (ArUco) marker based pose estimation [20] was used as ground truth for evaluation. Across several trials, the system achieved a **Euclidean Root Mean Square Error (RMSE) of** 0.09 **m (median =** 0.084 **m) and an orientation RMSE of** $1.49°$ **(median =** $1.40°$**) indoors**, and a **Euclidean RMSE of** 0.119 **m (median =** 0.118 **m) and orientation RMSE of** $2.233°$ **(median =** $2.268°$**) outdoors** between the holographic goal and the UAV's actual pose.

The contributions of this work are fourfold:

1) A novel MR interface for holographic goal definition and pre-flight visualization, addressing the limitations of joystick and FPV control while ensuring consistent inspection data capture.
2) Developed a LiDAR–RGB 3DGS localization server that enables robust, markerless, and persistent headset–UAV alignment across diverse environments.
3) Integrated holographic goal-setting with autonomous UAV execution, reducing operator workload, improving safety, and supporting ROI refinement in underrepresented areas of initial inspection maps.
4) Validated the system in both indoor and outdoor environments using a ModalAI Starling UAV and HL2, demonstrating feasibility, scalability, and practical value for infrastructure inspection.

II. RELATED WORK

A. UAV Control Methods

Research on UAV control interfaces has explored both egocentric and exocentric approaches: Egocentric control, commonly realized through FPV systems, provides an immersive perspective from the UAV's onboard camera. This approach can be implemented through head-mounted displays (HMDs), or even through eye-gaze–based control schemes where the operator's gaze direction influences UAV motion [21]. While egocentric methods can enhance immersion, restricted situational awareness, and difficulty estimating the UAV's orientation and depth often result in piloting errors or crashes, particularly in cluttered environments [11, 12].

Exocentric control, in contrast, emphasizes third-person or line-of-sight perspectives to improve situational awareness. Traditional radio controllers represent the most common form, but recent work has introduced gesture-based control [22, 23], external camera frameworks [16], and marker-assisted systems [24]. While exocentric methods improve situational awareness, they typically require additional infrastructure such as external cameras or fiducial markers, limiting scalability in unstructured or large-scale environments.

MR-based control extends exocentric paradigms by overlaying digital elements onto the physical scene, allowing operators to interact with spatially registered holograms. MR enables intuitive 3D goal-setting, pre-flight path validation, and autonomous navigation, thereby ensuring safe UAV control. Representative systems include PinpointFly [10] and FlyCam [15], which allow goal placement through mobile AR; Drone-Augmented Human Vision, which uses exocentric overlays to visualize occluded areas [16]; and FlyAR, which enhances navigation and spatial understanding with AR overlays [15]. Despite these advances, most MR–UAV systems rely on markers and external cameras, constraining their robustness and limiting their scalability.

In contrast, 3DGS-Holo-Inspector provides a markerless MR interface for pre-flight goal visualization and holographic placement while removing the reliance on external infrastructure.

B. Localization Techniques

Accurate localization is essential for aligning MR interfaces with UAV coordinate frames. Existing methods can be categorized into markerless vision-based, learning-based, and hybrid approaches: Markerless vision-based methods exploit natural image features such as edges, textures, and keypoints, avoiding reliance on fiducials. Classical approaches [25, 26] laid the foundation for visual localization using feature descriptors. More recent methods, such as SuperPoint [27] and SuperGlue [28], extract stable keypoints and perform deep feature matching that is robust to lighting variation, occlusion, and viewpoint changes.

Learning-based methods leverage deep neural networks for end-to-end pose estimation. PoseNet [29] demonstrated that camera poses can be regressed directly from monocular images, while MapNet [30] improved accuracy by incorporating geometric constraints. These methods are robust in challenging conditions but often require large training datasets and incur significant computational overhead.

Hybrid approaches combine classical feature-based methods with learning-based models to improve scalability and accuracy [31]. While these methods improve overall accuracy, they remain sensitive to environmental dynamics, such as lighting variations, partial occlusions, and feature-sparse regions.

To address these limitations, we introduce a 3DGS localization backbone. 3DGS represents scenes as dense, photorealistic point-based splats, providing a continuous and differentiable representation that supports accurate alignment. Recent advances demonstrate the potential of 3DGS for localization: GSLoc [32] backpropagates pose gradients over the rendering pipeline, and HGSLoc [33] and GS-CPR [34] introduce a lightweight plug-and-play refinement framework. GSplatLoc [35] combines dense keypoint matching and photometric refinement in a two-stage process for improved localization.

In this work, by integrating LiDAR and RGB sensing, 3DGS-Holo-Inspector leverages 3DGS map representation for image relocalization to achieve dense, markerless, and persistent registration between the MR headset and UAV. This enables reliable headset–UAV alignment even in outdoor environments where traditional visual localization methods struggle, ensuring scalable operation across inspection contexts.

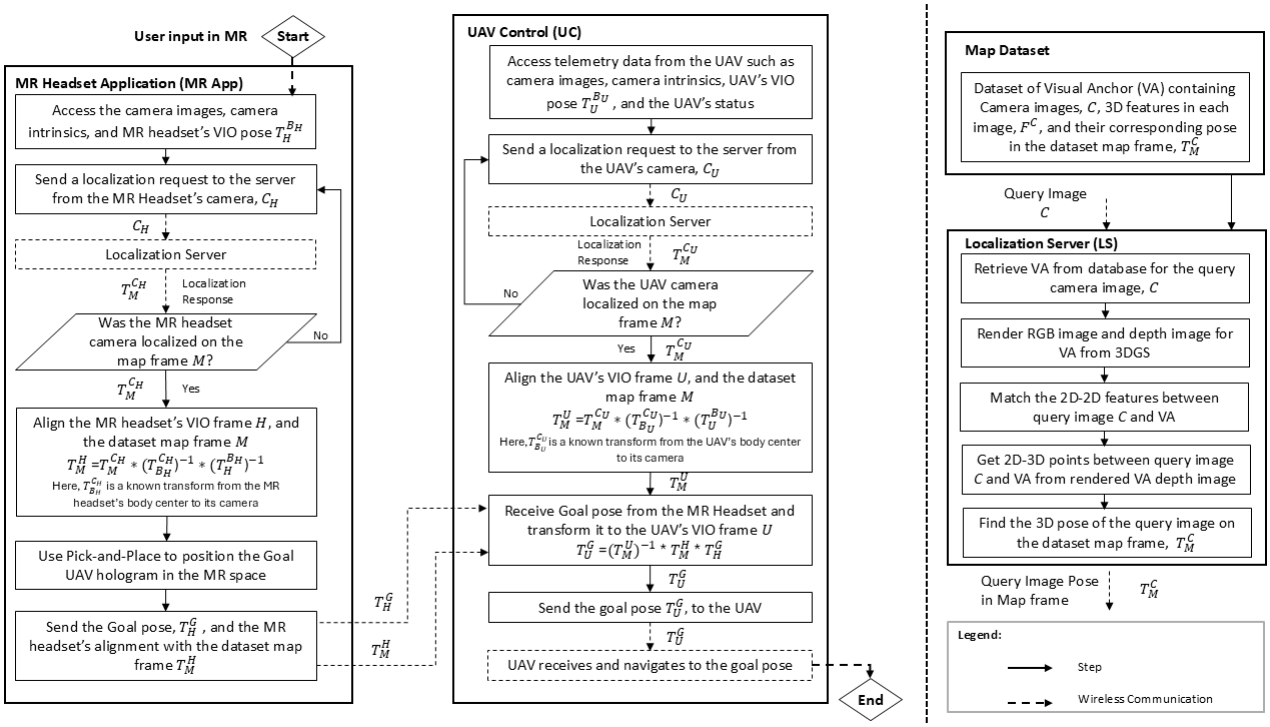

Fig. 1. System overview of 3DGS-Holo-Inspector, consisting of three interconnected subsystems: Localization Server (LS), MR Headset Application (MR App), and UAV Control (UC).

III. METHODOLOGY

The proposed 3DGS-Holo-Inspector system integrates three tightly coupled subsystems: (i) a *Localization Server (LS)* that establishes a precise SA between the MR headset and UAV using a 3DGS backbone, (ii) an *MR Headset Application (MR App)* that provides the operator with a holographic interface for goal definition, pre-flight visualization, and situational awareness, and (iii) a *UAV Control (UC)* module that transforms holographic goals into UAV control commands and ensures autonomous execution through the onboard autopilot. Fig. 1 shows the overall system architecture.

The workflow unfolds in four stages. First, the LS localizes both the MR headset and UAV within a shared database map frame, establishing SA between the headset and the UAV. Second, the operator defines a holographic goal in MR space using natural hand gestures via the MR App. Third, the UC module transforms the goal pose from the headset's visual inertial odometry (VIO) frame into the UAV's local VIO frame using the established SA. Finally, the UAV autonomously navigates to the goal while real-time telemetry and the live onboard video feed are streamed into the MR view, enhancing operator awareness.

A. 3DGS-Rendering Localization Server (LS)

The LS establishes a shared reference frame between the MR headset and UAV using markerless vision-based localization, enabling precise SA between the two devices. Our implementation builds on top of the framework introduced in [13], and extends it with 3DGS–based rendering to improve robustness under challenging inspection conditions.

A dataset map is first created by capturing high-resolution images with a calibrated RGB camera, augmented with depth scans from LiDAR or RGB-D sensors. These inputs are processed using a Simultaneous Localization and Mapping (SLAM) pipeline to estimate the pose of each image relative to the dataset map frame M. Each image is stored as a visual anchor (VA), consisting of the original camera image C, extracted 3D features F^C such as edges, corners, and key points, and their coordinates T_M^C in the map frame. In parallel, the posed images are used to train a 3DGS model, which generates a dense and photorealistic point-based representation of the scene.

During operation, an incoming query image I_Q from either the MR headset or the UAV is localized against the dataset map. The system first retrieves the closest VA using NetVLAD descriptors [36]. From the selected VA pose, the 3DGS model renders a synthetic RGB image I_R together with a depth map I_D, as shown in Fig. 2, effectively simulating the expected view from that location. A matcher implemented using Mast3R [37] is employed to establish dense pixel correspondences between I_Q and I_R, maintaining robustness under viewpoint and illumination variations. Using the rendered depth image I_D, the established 2D–2D correspondences between I_Q and I_R are lifted into 2D–3D correspondences in the dataset map frame.

These correspondences are then processed by a Perspective-n-Point (PnP) solver with P3P RANSAC to reject outliers and estimate the query camera pose in

the dataset map frame, $T_M^{C_Q}$. The transformation from the dataset map frame to the query device's VIO frame is subsequently computed as:

$$T_M^Q = T_M^{C_Q} \cdot (T_{B_Q}^{C_Q})^{-1} \cdot (T_Q^{B_Q})^{-1}$$

where:

- T_M^Q: transformation from map frame to the query device's VIO frame,
- $T_M^{C_Q}$: localized pose of the query device's camera in the map frame,
- $T_{B_Q}^{C_Q}$: known transform from the device body center to its camera,
- $T_Q^{B_Q}$: onboard VIO pose of the device body frame.

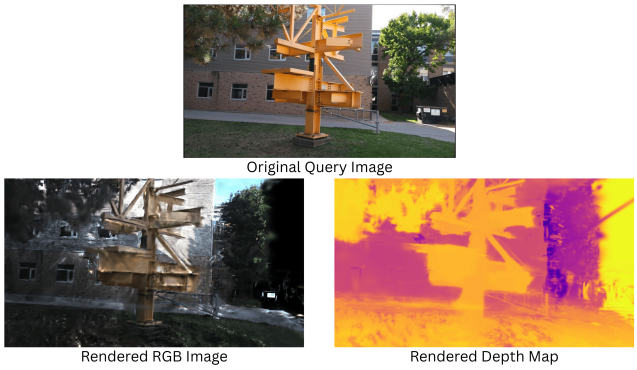

Fig. 2. Synthetically Rendered RGB and Depth Map by the 3DGS-Rendering Localization Server

This spatial alignment, illustrated in Fig. 3, allows holographic goal coordinates defined in the headset's VIO frame to be accurately transformed into the UAV's VIO frame.

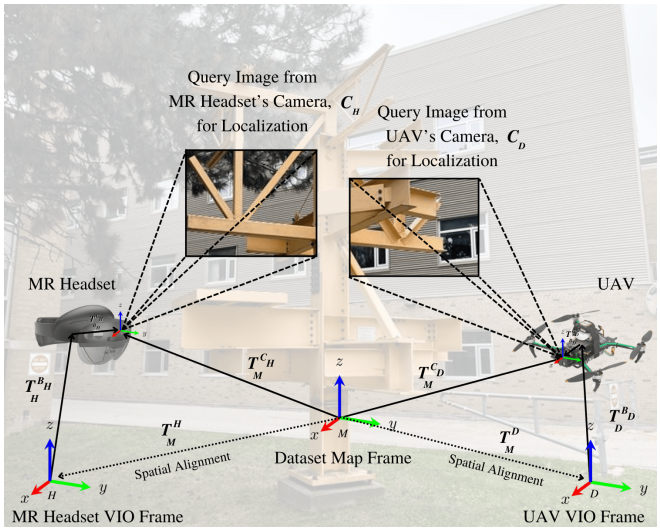

Fig. 3. Spatial alignment between the dataset map frame and the query device (MR headset and UAV) VIO frame.

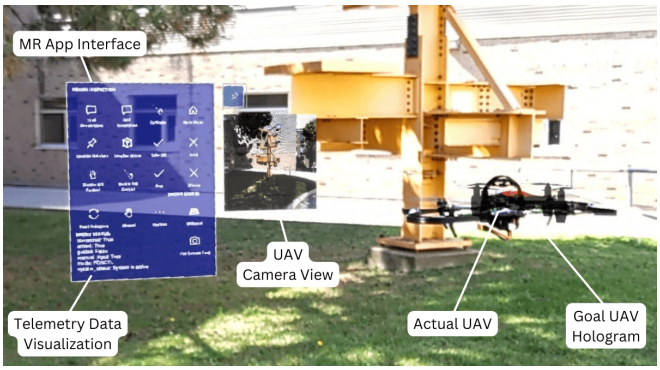

Fig. 4. MR App view: MR App interface for sending localization queries and visualizing telemetry data (blue canvas), UAV camera view, along with a goal UAV hologram and the actual UAV

B. MR Headset Application (MR App)

The MR App, as shown in Fig. 4, provides the operator-facing interface for triggering SA, defining holographic goals, previewing them before execution, and monitoring UAV telemetry in real time. Within the MR environment, a holographic UAV model is rendered that the operator manipulates using natural hand gestures to specify inspection goals. Each goal pose is expressed in the VIO frame of the MR headset, T_H^G, and transmitted to the UC module, where it is transformed into the UAV control frame for execution. Before dispatching a goal command, the operator can preview and validate the holographic goal directly in MR, confirming both the intended position and orientation of the UAV relative to the structure. This pre-flight visualization reduces the likelihood of collisions and ensures consistent data capture by allowing potential errors to be identified before the UAV begins navigation.

To support situational awareness and remove the need for the operator to shift attention to an external monitor, the MR App overlays all the important information, such as live telemetry, UAV pose, flight mode, and battery status, within the operator's field of view. The live camera feed from the UAV is also streamed directly into the headset, providing additional egocentric awareness. Holographic control buttons enable direct triggering of functions such as starting or stopping data capture, changing drone modes, takeoff, and land, streamlining all the necessary interactions. Finally, the UAV's current pose estimated in the dataset map frame, $T_M^{C_U}$, is also transformed into the MR headset frame, $T_H^{C_U}$, and visualized as a secondary hologram following the actual UAV. This allows the operator to continuously monitor localization accuracy and detect drift or misalignment in real time, further improving the reliability and safety of the inspection process.

C. UAV Control (UC)

The UC module serves as the execution backbone of 3DGS-Holo-Inspector, running on the ground control station (GCS) and interfacing directly with the UAV's autopilot. Its primary role is to transform holographically defined goals from the MR headset into control commands that can be

executed autonomously by the UAV. When a goal pose T_H^G is received from the headset, the UC module uses the SA estimated by the LS, namely the transformation from the dataset map frame to the headset's VIO frame T_M^H, and the transformation from the dataset map frame to the UAV's VIO frame T_M^U, to compute the corresponding goal in the UAV's VIO frame T_U^G. The transformation is expressed as:

$$T_U^G = (T_M^U)^{-1} \cdot (T_M^H) \cdot (T_H^G)$$

This ensures that the holographic goal is consistently represented within the UAV's local coordinate system. Once the transformation is complete, the goal pose T_U^G is transmitted to the UAV's onboard autopilot as a local waypoint to reach. The autopilot continuously evaluates the error between the current UAV VIO pose $T_U^{B_U}$ and the desired goal pose T_U^G, and minimizes this error iteratively through a closed-loop feedback position controller (e.g., PX4's model predictive position controller) which generates the required velocity/acceleration and attitude commands. This process enables the UAV to navigate precisely to the defined goal without requiring incremental joystick input.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

Experiments were conducted in both indoor and outdoor environments at the university campus. Indoor tests were carried out in a research lab, which provided a controlled setting for accuracy evaluation. Outdoor trials were performed around a steel tree structure on campus, shown in Fig. 5, a site commonly used for testing inspection applications, where natural illumination variations, background clutter, and weather effects such as wind introduced additional challenges.

Hardware: The MR App was deployed on the HL2 [19]. The UAV platform was the ModalAI Starling [18], equipped with a VOXL2 companion computer running the VOXL software development kit (SDK) [38] and a PX4 autopilot [39]. The GCS was an MSI Cyborg 15 laptop with an NVIDIA GeForce RTX 4090 GPU. The LS was hosted on a Lambda Vector PC with dual NVIDIA GeForce RTX 4090 GPUs.

Software: Robot Operating System (ROS) Noetic provided inter-process communication, with MAVROS [40] serving as the ROS–MAVLink bridge to the PX4 autopilot. The MR App was developed in Unity 2020.3.24 with the Mixed Reality Toolkit (MRTK) [41]. Communication between the HL2 and the GCS was established through a TCP link via the ROS–TCP–Connector [42].

Localization Server: For the LS dataset preparation, RGB frames, depth maps, IMU data, and odometry were first collected using a custom LiDAR–RGB scanner consisting of a Livox AVIA LiDAR (with integrated IMU) and a Teledyne FLIR BFS-PGE-16S2C-CS GigE camera at 1,440 × 1,080 resolution.

A state-of-the-art LiDAR–Inertial–Visual SLAM algorithm, R3Live [43], was employed to generate a globally consistent 3D point cloud together with accurate camera poses in the dataset map frame. We then trained a 3DGS model on top of this reconstruction following [17], for 30,000 iterations. The training also included a depth-regularization loss to improve geometric fidelity, resulting in a dense, photorealistic, and differentiable scene representation for relocalization. In terms of map construction cost, 3DGS training takes approximately 35 minutes on ∼340 images using a workstation equipped with an NVIDIA RTX A5000 GPU and an AMD Ryzen Threadripper PRO 5955WX 16-core processor.

During operation, device cameras from either the HL2 or the Starling submits query images to the LS deployed on a Lambda Vector PC. ROS middleware manages the communication, where queries are published as ROS topics and the LS responds with the estimated camera pose in the dataset map frame. This enables persistent and accurate SA between the headset and UAV VIO frames.

MR Application: The MR App provided the operator-facing interface for holographic goal definition, pre-flight visualization, and real-time telemetry. As shown in Fig. 5, the pilot defined and previewed goals in 3D MR space using natural hand gestures. The TCP link allowed the HL2 to publish goal poses, issue localization queries, and receive telemetry data from the drone through ROS topics.

UAV Control: The UC module running on the GCS transformed goal poses received from the HL2 VIO frame onto the Starling VIO frame using the SA. Transformed goals were sent to the PX4 autopilot via MAVROS, running onboard the UAV, as local waypoints. PX4's closed-loop model predictive position controller tracked these waypoints and generated the required motion commands, enabling the UAV to autonomously navigate to the commanded pose without any manual interventions.

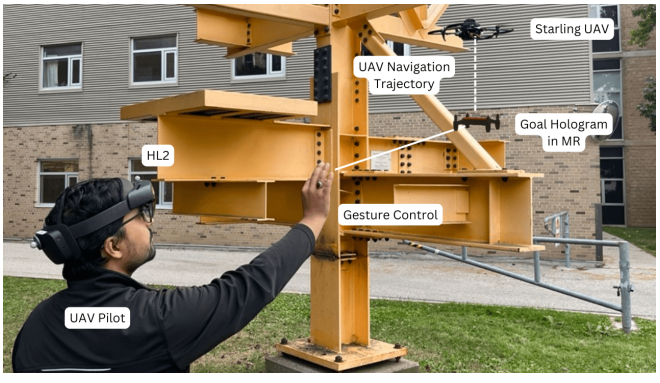

Fig. 5. UAV pilot using gesture control to set a goal pose in the MR App. The UAV autonomously navigates to the specified goal.

B. VIO Accuracy of Starling and HL2

System performance depends on the accuracy of the VIO running onboard the MR headset and UAV. The Starling's VIO was evaluated by comparing its estimated camera pose

in its VIO frame, $T_U^{C_U}$, against ground-truth poses from ArUco markers [20] as shown in Fig. 6, $T_A^{C_U}$. The VIO-based camera pose was derived from the body pose $T_U^{B_U}$ given by the VOXL SDK [38] and the known body–camera transform $T_{B_U}^{C_U}$. Euclidean distances between two positions were computed in both frames, and the difference served as the error metric. During four indoor experiments with 10 data points in each, Starling's VIO achieved a positional RMSE of 0.052 m (median = 0.0245 m.

The HL2's VIO was evaluated in the same manner by comparing $T_H^{C_H}$ against $T_A^{C_H}$. MRTK provided the HL2 camera pose directly. Four indoor experiments with 10 data points in each, yielded a positional RMSE of 0.064 m (median = 0.035 m).

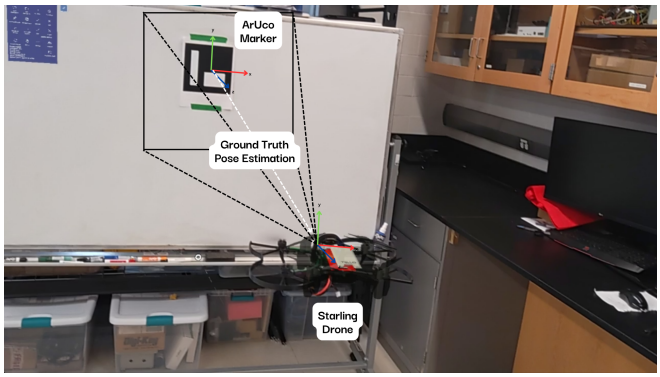

Fig. 6. UAV facing ArUco marker for ground truth pose estimation (Indoor)

C. Goal Alignment Accuracy

Goal alignment accuracy quantifies the difference between the holographic goal set in the MR App and the UAV's final pose after navigation. To obtain reliable ground truth for this evaluation, ArUco marker–based pose estimation was employed, consistent with the procedure used for VIO accuracy assessment (Fig. 6). The HL2 goal pose T_H^G was transformed into the ArUco frame as

$$T_A^G = T_A^{C_H} \cdot (T_H^{C_H})^{-1} \cdot T_H^G,$$

where $T_A^{C_H}$ is the HL2 camera pose in the ArUco frame and $T_H^{C_H}$ its VIO pose. The Starling's body pose in the ArUco frame was obtained as

$$T_A^{B_U} = T_A^{C_U} \cdot (T_{B_U}^{C_U})^{-1},$$

where $T_A^{C_U}$ is the camera pose and $T_{B_U}^{C_U}$ the body–camera transform. Errors between T_A^G and $T_A^{B_U}$ were computed in positional x, y, z, and orientation yaw.

1) Indoor Experiments: For the indoor evaluation, the dataset map was generated several days before testing, during which the lab environment had changed slightly due to equipment movement, introducing discrepancies between the stored map and the actual scene. Despite these variations, 3DGS-Holo-Inspector consistently achieved centimeter-level positional accuracy, and maintained orientation errors below $2°$. This validates both the precision of holographic goal alignment and the system's robustness to moderate environmental dynamics often encountered in real-world inspection tasks.

Across 10 trials, the system achieved:

- **Positional RMSE:** $x = 0.056$ m, $y = 0.046$ m, $z = 0.054$ m
- **Positional Median:** $x = 0.056$ m, $y = 0.041$ m, $z = 0.048$ m
- **Euclidean RMSE:** 0.090 m
- **Euclidean Median:** 0.084 m
- **Orientation RMSE (yaw):** $1.491°$
- **Orientation Median (yaw):** $1.400°$

2) Outdoor Experiments: For the outdoor evaluation, experiments were conducted near the steel tree structure, introducing challenges such as variable lighting and natural environmental dynamics. Despite these factors, 3DGS-Holo-Inspector maintained reliable holographic goal alignment and autonomous navigation, with MR higher than the indoor tests. This demonstrates the system's robustness under real-world conditions and confirms its practicality for infrastructure inspection tasks.

Across 10 trials, the system achieved:

- **Positional RMSE:** $x = 0.079$ m, $y = 0.066$ m, $z = 0.076$ m
- **Positional Median:** $x = 0.079$ m, $y = 0.066$ m, $z = 0.075$ m
- **Euclidean RMSE:** 0.119 m
- **Euclidean Median:** 0.118 m
- **Orientation RMSE (yaw):** $2.233°$
- **Orientation Median (yaw):** $2.268°$

Detailed experimental data is reported in Tables I and II for indoor and outdoor experiments, respectively. Moreover, Figures 7, 8, 9, and 10 present boxplots of the positional as well as overall Euclidean and orientation errors. In these plots, the red line denotes the median error, while the blue dashed line marks the RMSE, which serves as our primary performance metric. The shaded box corresponds to the interquartile range (25th–75th percentiles), and the whiskers extend to the minimum and maximum error values, excluding statistical outliers.

TABLE I

POSITIONAL, EUCLIDEAN, AND ORIENTATION ERRORS (INDOOR)

	X (m)	Y (m)	Z (m)	Euclidean (m)	Orientation (°)
1	0.049	0.032	0.046	0.074	0.792
2	0.064	0.063	0.044	0.100	1.545
3	0.068	0.076	0.050	0.113	1.583
4	0.043	0.036	0.050	0.075	2.136
5	0.034	0.047	0.074	0.094	2.112
6	0.006	0.034	0.058	0.068	1.034
7	0.067	0.038	0.079	0.111	1.202
8	0.048	0.033	0.049	0.076	1.210
9	0.061	0.031	0.046	0.082	1.595
10	0.080	0.046	0.028	0.097	1.078
RMSE	0.056	0.046	0.054	0.090	1.491
Median	0.056	0.041	0.048	0.084	1.400

TABLE II

POSITIONAL, EUCLIDEAN, AND ORIENTATION ERRORS (OUTDOOR)

	X (m)	Y (m)	Z (m)	Euclidean (m)	Orientation (°)
1	0.079	0.066	0.073	0.120	2.241
2	0.073	0.062	0.069	0.113	1.984
3	0.087	0.055	0.082	0.123	2.453
4	0.065	0.071	0.077	0.116	1.873
5	0.083	0.068	0.072	0.120	2.295
6	0.070	0.074	0.070	0.116	2.011
7	0.086	0.060	0.080	0.121	2.376
8	0.073	0.066	0.071	0.115	2.189
9	0.078	0.059	0.077	0.116	2.322
10	0.090	0.072	0.082	0.127	2.501
RMSE	0.079	0.066	0.076	0.119	2.233
Median	0.079	0.066	0.075	0.118	2.268

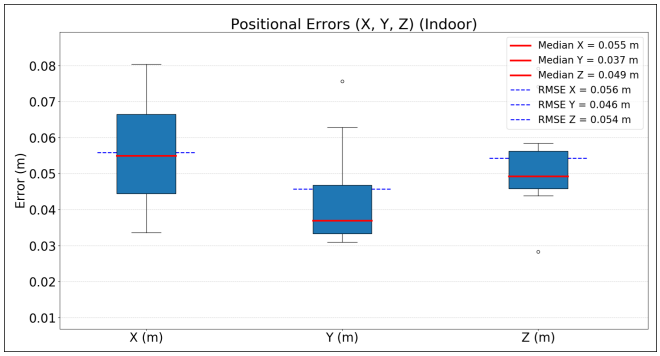

Fig. 7. Positional errors (x, y, z) (Indoor)

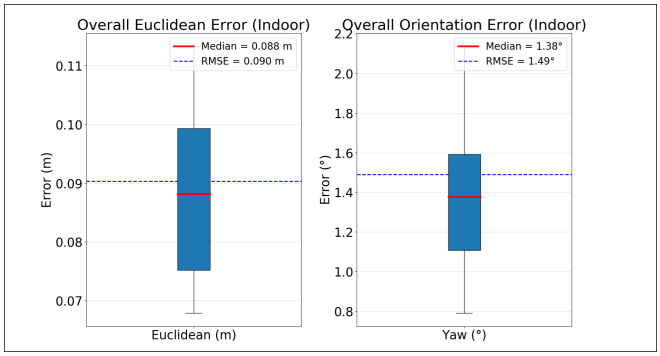

Fig. 8. Overall euclidean and orientation error (Indoor)

V. CONCLUSION AND FUTURE WORK

This paper presented 3DGS-Holo-Inspector, a MR UAV control system that unifies holographic goal definition with autonomous navigation to complement inspection workflows. By leveraging pre-built maps, the system enables operators to direct UAVs toward specific ROIs where coverage is incomplete or additional detail is needed, improving both efficiency and inspection quality. The system introduces an intuitive MR interface that allows operators to define and preview UAV navigation goals before flight, reducing piloting errors and ensuring consistent data capture. A LiDAR–RGB 3DGS LS provides dense, markerless, and persistent SA between the MR headset and UAV, enabling robust operation

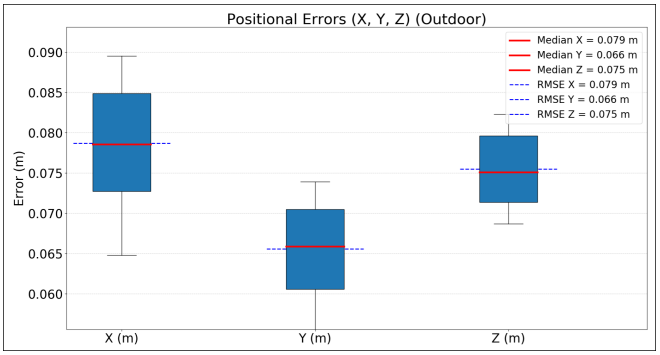

Fig. 9. Positional errors (x, y, z) (Outdoor)

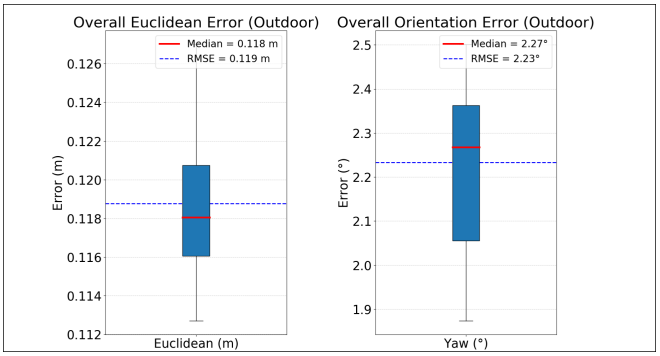

Fig. 10. Overall euclidean and orientation error (Outdoor)

in both indoor and outdoor settings. By tightly coupling holographic goal-setting with autonomous execution, 3DGS-Holo-Inspector improves safety and supports targeted ROI inspection without manual piloting.

Experimental validation confirmed the feasibility and accuracy of the proposed approach. In indoor trials, the system achieved an Euclidean RMSE of 0.090 m and an orientation (yaw) RMSE of $1.491°$, demonstrating centimeter-level precision in goal alignment. Outdoor experiments near a steel tree structure subject to natural illumination and weather challenges showed slightly higher errors, with a Euclidean RMSE of $0.119°$ and orientation RMSE of 2.233 m.

Some limitations still remain. Currently, the UC relies on the UAV's autopilot for low-level waypoint tracking and control command generation. However, this command generation does not explicitly account for surrounding obstacles. In addition, the accuracy of goal alignment is constrained by the reliability of the onboard VIO systems on both UAV and the headset. Drift in VIO estimates can accumulate during long missions, particularly in feature-sparse outdoor environments, propagating into misalignment between holographic goals and UAV execution.

Future work will focus on addressing these challenges. First, we will introduce an intermediate autonomy layer between the UC and the autopilot to enable local, obstacle-aware planning that respects nearby structures and safety margins. Secondly, we will investigate improved onboard positioning for both the UAV and headset, with emphasis on pose-graph optimization techniques to improve robustness in

large-scale or dynamic environments. Third, view-based goal refinement algorithms will be developed to match the UAV's live camera feed with the rendered goal view, reducing residual misalignment after coarse navigation.

Finally, to encourage reproducibility and community adoption, the full implementation of 3DGS-Holo-Inspector is available at:
https://github.com/cviss-lab/3dgs-holoinspector.

REFERENCES

[1] Billie F. Spencer, Vedhus Hoskere, and Yasutaka Narazaki. "Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring". In: *Engineering* 5.2 (2019), pp. 199–222. DOI: 10.1016/j.eng.2018.11.030.

[2] C.-Z. Dong and F. N. Catbas. "A Review of Computer Vision–Based Structural Health Monitoring at Local and Global Levels". In: *Structural Health Monitoring* 20.2 (2021), pp. 692–743. DOI: 10.1177/1475921720935585.

[3] Chenyu Zhang, Zhaozheng Yin, and Ruwen Qin. "Attention-Enhanced Co-Interactive Fusion Network (AECIF-Net) for Automated Structural Condition Assessment in Visual Inspection". In: *Automation in Construction* 159 (2024), p. 105292. DOI: 10.1016/j.autcon.2024.105292.

[4] Szu-Pyng Kao, Yung-Chen Chang, and Feng-Liang Wang. "Combining the YOLOv4 Deep Learning Model with UAV Imagery Processing Technology in the Extraction and Quantization of Cracks in Bridges". In: *Sensors* 23.5 (2023), p. 2572. DOI: 10.3390/s23052572.

[5] Tommaso Panigati et al. "Drone-Based Bridge Inspections: Current Practices and Future Directions". In: *Automation in Construction* 173 (2025), p. 106101. DOI: 10.1016/j.autcon.2025.106101.

[6] Federal Highway Administration. *Use of Unmanned Aircraft Systems (UAS) to Enhance the Design, Construction, Inspection, and Maintenance of Transportation Infrastructure*. Tech. rep. FHWA-PL-23-007. Accessed: 2025-09-13. U.S. Department of Transportation, Federal Highway Administration, 2023. URL: https://international.fhwa.dot.gov/pubs/pl23007.pdf.

[7] Voliro AG. *Drones for Inspection: Advantages, Use Cases and Best Practices*. Accessed: 2025-09-01. 2024. URL: https://voliro.com/blog/best-drones-for-inspection/.

[8] U.S. Congress. *H.R. 3593 (118th Congress): Drone Infrastructure Inspection Grant Act*. Introduced May 22, 2023. 2023. URL: https://www.congress.gov/bill/118th-congress/house-bill/3593.

[9] U.S. Department of Transportation, Office of Inspector General. *FAA's Progress and Challenges in Granting Exemptions for Beyond Visual Line of Sight (BVLOS) Drone Operations*. Tech. rep. Accessed: 2025-09-13. U.S. Department of Transportation, June 2025. URL: https://www.oig.dot.gov/sites/default/files/library-items/FAA%20BVLOS%20Drone%20Operations%20Final%20Report_6.30.2025.pdf.

[10] Linfeng Chen et al. "PinpointFly: An Egocentric Position-control Drone Interface using Mobile AR". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445110.

[11] Julien Mellet et al. "Evaluation of Human-Robot Interfaces Based on 2D/3D Visual and Haptic Feedback for Aerial Manipulation". In: *Journal of Intelligent & Robotic Systems* 111 (2025), p. 104. DOI: 10.1007/s10846-025-02284-7.

[12] Bingjian Huang et al. "AeroHaptix: A Wearable Vibrotactile Feedback System for Enhancing Collision Avoidance in UAV Teleoperation". In: *IEEE Robotics and Automation Letters* 10.5 (2025), pp. 4260–4267. DOI: 10.1109/LRA.2025.3548866.

[13] Zaid Abbas Al-Sabbag, Chul Min Yeum, and Sriram Narasimhan. "Enabling Human–Machine Collaboration in Infrastructure Inspections through Mixed Reality". In: *Advanced Engineering Informatics* 53 (2022), p. 101709. DOI: 10.1016/j.aei.2022.101709.

[14] Yuze Dan et al. "HoloDesigner: A mixed reality tool for on-site design". In: *Automation in Construction* 129 (2021), p. 103808. DOI: 10.1016/j.autcon.2021.103808.

[15] Stefanie Zollmann et al. "FlyAR: Augmented Reality Supported Micro Aerial Vehicle Navigation". In: *IEEE Transactions on Visualization and Computer Graphics* 20.4 (2014), pp. 560–568. DOI: 10.1109/TVCG.2014.24.

[16] Okan Erat et al. "Drone-Augmented Human Vision: Exocentric Control for Drones Exploring Hidden Areas". In: *IEEE Transactions on Visualization and Computer Graphics* 24.4 (2018), pp. 1437–1446. DOI: 10.1109/TVCG.2018.2794058.

[17] Bernhard Kerbl et al. "3D Gaussian Splatting for Real-Time Radiance Field Rendering". In: *ACM Transactions on Graphics* 42.4 (July 2023). DOI: 10.1145/3592433.

[18] ModalAI. *Starling - ModalAI*. 2023. URL: https://www.modalai.com/products/starling-2.

[19] Microsoft Corporation. *Microsoft HoloLens Documentation*. Accessed: 2025-09-13. 2023. URL: https://learn.microsoft.com/en-us/hololens/.

[20] PAL Robotics and Contributors. *aruco_ros: ArUco Marker Detection for ROS*. Accessed: 2025-09-13. 2023. URL: https://github.com/pal-robotics/aruco_ros.

[21] Pavan Kumar B. N. et al. "GazeCamera: A Novel Gaze-Controlled UAV Camera". In: *Proceedings of the 2019 2nd International Conference on Control and Robot Technology*. Association for Computing Machinery, 2020, pp. 130–135. ISBN: 9781450372527. DOI: 10.1145/3387304.3387314.

[22] Dimitrios Sainidis et al. "Single-Handed Gesture UAV Control and Video Feed AR Visualization for First Responders". In: *Proceedings of the 18th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. 2021, pp. 835–848.

[23] Hao Kang et al. "FlyCam: Multitouch Gesture Controlled Drone Gimbal Photography". In: *IEEE Robotics and Automation Letters* 3.4 (2018), pp. 3717–3724. DOI: 10.1109/LRA.2018.2856271.

[24] Sourav Raxit, Simant Bahadur Singh, and Abdullah Al Redwan Newaz. "YoloTag: Vision-based Robust UAV Navigation with Fiducial Markers". In: *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. 2024, pp. 311–316. DOI: 10.1109/RO-MAN60168.2024.10731319.

[25] David G. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". In: *International Journal of Computer Vision* 60.2 (Nov. 2004), pp. 91–110. DOI: 10.1023/B:VISI.0000029664.99615.94.

[26] Ethan Rublee et al. "ORB: An efficient alternative to SIFT or SURF". In: *2011 International Conference on Computer Vision*. 2011, pp. 2564–2571. DOI: 10.1109/ICCV.2011.6126544.

[27] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. "SuperPoint: Self-Supervised Interest Point Detection and Description". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2018, pp. 337–33712. DOI: 10.1109/CVPRW.2018.00060.

[28] Paul-Edouard Sarlin et al. "SuperGlue: Learning Feature Matching With Graph Neural Networks". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 4937–4946. DOI: 10.1109/CVPR42600.2020.00499.

[29] Alex Kendall, Matthew Grimes, and Roberto Cipolla. "PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 2938–2946. DOI: 10.1109/ICCV.2015.336.

[30] Samarth Brahmbhatt et al. "Geometry-Aware Learning of Maps for Camera Localization". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018, pp. 2616–2625. DOI: 10.1109/CVPR.2018.00277.

[31] Torsten Sattler, Bastian Leibe, and Leif Kobbelt. "Improving Image-Based Localization by Active Correspondence Search". In: *Computer Vision – ECCV 2012*. Ed. by Andrew Fitzgibbon et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 752–765. ISBN: 978-3-642-33718-5.

[32] Kazii Botashev et al. "GSLoc: Visual Localization with 3D Gaussian Splatting". In: *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2024, pp. 5664–5671. DOI: 10.1109/IROS58592.2024.10801919.

[33] Zhongyan Niu et al. "HGSLoc: 3DGS-Based Heuristic Camera Pose Refinement". In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. 2025, pp. 1–7. DOI: 10.1109/ICRA55743.2025.11127431.

[34] Changkun Liu et al. "GS-CPR: Efficient Camera Pose Refinement via 3D Gaussian Splatting". In: *International Conference on Learning Representations*. 2025.

[35] Gennady Sidorov et al. "GSplatLoc: Grounding Keypoint Descriptors into 3D Gaussian Splatting for Improved Visual Localization". In: *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2025, pp. 12601–12607. DOI: 10.1109/IROS60139.2025.11246406.

[36] Relja Arandjelovic et al. "NetVLAD: CNN Architecture for Weakly Supervised Place Recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016. DOI: 10.1109/CVPR.2016.572.

[37] Vincent Leroy, Yohann Cabon, and Jerome Revaud. "Grounding Image Matching in 3D with MASt3R". In: *Computer Vision – ECCV 2024*. Springer Nature Switzerland, 2025, pp. 71–91. DOI: 10.1007/978-3-031-73220-1_5.

[38] ModalAI, Inc. *VOXL SDK Documentation*. Accessed: 2025-09-13. 2025. URL: https://docs.modalai.com/voxl-sdk/.

[39] PX4 Development Team. *PX4 Autopilot: Open Source Flight Control Software*. Accessed: 2025-09-13. 2023. URL: https://github.com/PX4/PX4-Autopilot.

[40] MAVROS Contributors. *MAVROS: MAVLink Extendable Communication Node for ROS*. Accessed: 2025-09-13. 2025. URL: https://wiki.ros.org/mavros.

[41] Microsoft Corporation and Contributors. *Mixed Reality Toolkit for Unity (MRTK)*. Accessed: 2025-09-13. 2023. URL: https://github.com/microsoft/MixedRealityToolkit-Unity.

[42] Unity Technologies and Contributors. *ROS-TCP Connector: Unity-ROS Communication Tool*. Accessed: 2025-09-13. 2023. URL: https://github.com/Unity-Technologies/ROS-TCP-Connector.

[43] Jiarong Lin and Fu Zhang. "R3LIVE: A Robust, Real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state Estimation and mapping package". In: *2022 International Conference on Robotics and Automation (ICRA)*. 2022, pp. 10672–10678. DOI: 10.1109/ICRA46639.2022.9811935.