Evaluating a Multi-Modal Robotic Approach for Deconstruction Inspection: A POP-Based Comparison with a Manual Approach

Heyaojing Huang, Vaishnavi Thumuganti, Martin Fischer

Department of Civil and Environmental Engineering
Stanford University, California, 94305

***Abstract*—Material reuse and waste reduction are gaining attention for their role in lowering carbon emissions and enabling net-zero buildings. Prior work has shown the potential of non-invasive deconstruction inspection using multi-modal sensing and machine learning to generate reuse recommendations. However, manual data collection remains operator-dependent, limiting consistency and repeatability. To address this, we integrated a quadruped robot for automated sensing and evaluated its feasibility using a product–organization–process (POP) framework. While the robotic workflow streamlines the organizational structure and simplifies the product setup, it still requires manual fine-tuning of viewpoints, as the robot lacks awareness of inspection targets and the ability to determine appropriate scanning behavior, and offers no real-time data quality feedback—relying instead on asynchronous review, which may lead to rework. Future work will focus on developing 3D object-level semantic perception, real-time quality assessment metrics, and mixed reality visualization to enhance human–robot collaboration, along with hardware adaptations and sensor integration to expand the robot's multi-modal capabilities.**

I. Introduction

As material scarcity and environmental concerns grow, material reuse and waste reduction have gained attention for their potential to reduce carbon emissions and enable net-zero buildings. Prior work combining multi-modal sensing including thermal, RGB and depth with machine learning-based data processing has shown a non-invasive method for deconstruction inspection, enabling the analysis of material conditions and reveals hidden geometries within existing buildings (Cabral et al. 2025). However, the previous pilot revealed that manual data collection involved multiple operators working together to mark distances, position the sensor, and trigger image capture, with the process repeated for each patch Cabral et al. [1]. To address this, we initiated efforts to integrate the multi-modal sensing system with a quadruped robot to enable automated data collection. Figure 1 illustrates the proposed end-to-end workflow, in which robot-collected multi-modal data will be processed by an AI system to generate material reuse suggestions. To assess the feasibility of the proposed robotic approach, we used the Construction Robot Evaluation Framework Markenson [2] to compare the manual and robotic workflows in terms of product, organization, and process.

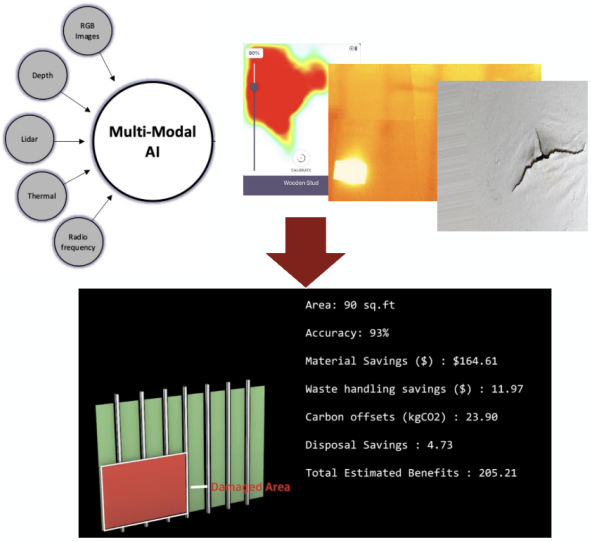

Fig. 1. Multi-modal AI-assisted deconstruction inspection workflow.

II. POP-based Evaluation

A. Product

The manual data capture setup consists of a multi-modal sensor rig mounted on a wheeled, height-adjustable tripod, allowing it to be repositioned throughout the site. The rig integrates an RGB-D camera and a thermal camera for simultaneous sensing. Operators manually measure the distance between the rig and the inspection target using a tape measure, and then trigger data capture via a laptop mounted on the rig.

For the robotic system, ANYmal was selected for automated multi-modal data collection and deployed to inspect structures slated for demolition, with the goal of identifying opportunities for material reuse and repurposing. This inherently requires the robot to perform multi-task operations, including multi-modal sensing and navigation.

For data collection, ANYmal is equipped with a 360° LiDAR, six RealSense RGB-D cameras positioned around its body (front, rear, left, and right), as well as wide-angle cameras mounted on the upper and lower front and rear. An embedded compute unit enables real-time data synchronization across

all sensors. The robot actively corrects for motion-induced noise to maintain data quality during movement. While it provides a full 360-degree horizontal field of view, its vertical sensing is limited by its maximum standing height (890 mm) and a tilt range of only ±15°, making it difficult to directly capture elevated targets. In such cases, angled views may degrade scan quality, suggesting that a telescoping sensor mount could be considered to improve vertical coverage. For navigation, the onboard sensors mentioned above also support autonomous navigation and obstacle avoidance. In terms of mobilty, ANYmal's four-legged, all-terrain mobility allows it to traverse both indoor and outdoor environments with uneven, dusty, or slippery surfaces, including grates, gravel, and wet floors. Its compact form factor (Length: 930 mm; Width: 530 mm; Height: 890 mm) and adjustable height (470–890 mm), along with its zero-turn capability, make it well-suited for navigating tight, cluttered, or obstructed deconstruction sites.

The robot is powered by a lithium-ion battery with an operational runtime of 1–2 hours. Control is available via a tablet-based Workforce app or a PC-based ROS GUI. During operation, continuous Wi-Fi connectivity is required to maintain communication between the robot and the control interface.

To evaluate the robot's performance under realistic conditions, we selected the an academic building as the initial deployment site. The facility includes part of the ground floor, a multi-story staircase with a high ceiling, and a 3,000-square-foot basement area. In this study, as shown in Figure 2, three testing areas were selected: (1) a construction area, (2) a stairwell, and (3) a hallway with minimal interference from glass walls. This selection enables testing of the robot's capabilities across flat indoor surfaces, vertical movement, and active construction environments.

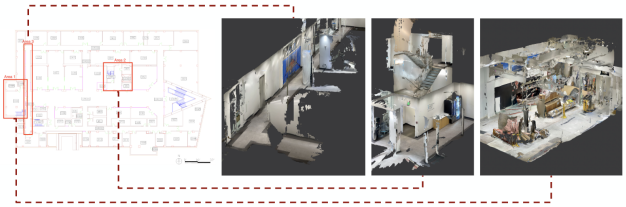

Fig. 2. Floor plan of the academic building and 3D scans of the testing areas, captured using the 3D Scanner app.

B. Organization

From an organizational perspective, the manual data capture workflow requires a team of three operators: one to measure distances and heights, another to adjust the tripod and align the sensor rig, and a third to operate the multi-modal sensor for data collection.

In contrast, the robotic workflow with ANYmal typically requires only one or two people to supervise the robot in the area of interest and manage ROS (Robot Operating System) topic recording on a PC. With proper planning, the entire process can be handled by a single operator, especially when autonomous navigation is enabled via BIM or a pre-built map. The robot autonomously handles both motion and data packaging, generating complete ROS bag files that include multi-modal sensor data, precise positioning, sensor intrinsics and extrinsics, and synchronized timestamps. This significantly reduces manual input and coordination efforts, and streamlines team structure.

1) Process: The manual data collection process required three operators and a sensor box tripod for simultaneous RGB-D and thermal data capture. Specifically, one operator measured the sensor height and distance based on the desired field of view (FOV), another adjusted the tripod to match the measured position, and the third operated the sensor to capture the image. Each patch cycle began with manually measuring and marking distances from the wall using paper tape to establish different resolution levels. The selected resolution setting directly influenced the number of patches required. Higher resolutions demanded multiple captures at closer distances, while lower resolutions required fewer patches taken from farther away. For instance, capturing data for a 6.79 × 8 ft drywall section required nine patches at high resolution, two at mid-resolution, and one at low resolution (Cabral et al. 2025). Once the distances were marked, the team positioned the sensor at the designated point and triggered image capture—completing that patch. The cycle then repeated for the next patch, starting again with distance measurement. Throughout the data collection process, operators had to control and verify the scanning angle and motion, as both significantly impacted accuracy. The previous case study showed that vertical motion generally produced better results than horizontal scanning, especially for blank walls with minimal visual features. Additionally, a straight-on view of the object yielded more accurate results than angled views.

Robotic data collection using ANYmal follows a different process structure. Initial deployment in a new environment typically requires manual mapping by a human operator using the tablet-based Workforce app. If a BIM model is available, it can be directly imported, eliminating the need for manual mapping. The manual mapping process employs SLAM (Simultaneous Localization and Mapping) to build a map of the surroundings, with a typical resolution of 10 to 15 cm. However, the system may face challenges with reflective surfaces such as glass, and in dynamic environments—such as active deconstruction sites—significant layout changes may require re-scanning to ensure accurate navigation. If pre-mapping is not performed, the robot must be manually controlled through the Workforce app for navigation throughout the operation. Once mapping is complete, users can define points of interest (POIs) via a PC-based GUI.

ANYmal then computes an optimized path and navigates between POIs semi-autonomously. In its default autonomous mode, the robot follows a stop-and-go behavior when encountering obstacles—pausing until the obstruction is removed before resuming its path. If alternate POIs are predefined, ANYmal can reroute and revisit previously inaccessible locations once conditions improve. However, its navigation

performance is generally more reliable in open obstacle -free spaces. In complex or highly obstructed environments, human teleoperation can enhance robustness by allowing manual intervention when needed.

The robotic system improves the data collection workflow by enabling autonomous navigation, obstacle avoidance, and eliminating the need for manual distance measurement or physical repositioning of a sensor tripod. Its position can be continuously tracked through ROS topics, allowing for more precise and repeatable scanning. However, the current system still lacks 3D object-level understanding. Instead of recognizing specific inspection targets, the robot relies on operators to manually specify coordinate points through the GUI or Workforce app. Once it reaches the target location, operators often need to fine-tune the robot's viewing angle to properly frame the object of interest—similar to the manual workflow. Figure 3 shows ANYmal during one of the site tests. While this approach streamlines setup and navigation, it still requires human input for target identification and final alignment. In terms of quality control, both manual and robotic workflows face similar limitations. During data collection, operators can only perform rough visual checks to verify framing and coverage. However, actual quality assessment depends on downstream data processing by the AI/ML team, which happens asynchronously. In robotic workflow, data is typically transmitted via ROS bag files for post-processing, and if the quality is found insufficient, the entire capture must be repeated. Figure 4 shows a sample of processed multi-modal data collected from the construction site. As a result, despite improvements in automation and efficiency, the issue of real-time quality assurance remains unresolved.

Fig. 3. AnyMal Capturing Multi-Modal Data in the testing area.

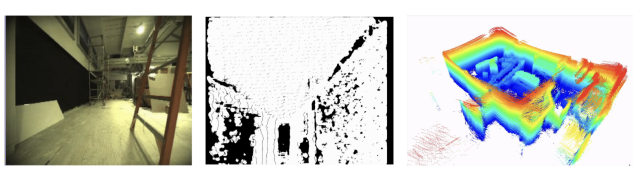

Fig. 4. Collected data from multi-modal sensors.

III. Conclusion

This evaluation was conducted in a realistic built environment as a preliminary pilot, with testing areas covering flat floor surfaces, stairwells, and active construction zones—conditions that reflect common scenarios in real deconstruction projects. These diverse settings support the generalizability of the robotic workflow beyond controlled lab environments.

The POP-based comparison with manual workflows shows that the robotic approach streamlines organizational structure and simplifies product setup. However, key limitations—specifically, the lack of object-level understanding and real-time data quality feedback—still hinder full autonomy and deployment efficiency, including fast execution, low rework rates, and reduced reliance on manual operation. These gaps are being addressed through ongoing research, including: (1) a 3D semantic perception module for recognizing and aligning with inspection targets; (2) real-time, object-level data quality assessment metrics to reduce rework; and (3) a mixed reality interface for live visualization and operator-robot alignment during scanning. We also plan hardware adaptations such as adjustable sensor mounts and the integration of thermal and RF sensors to expand multi-modal capabilities. These enhancements aim to make robotic data collection more deployable, robust, and generalizable across deconstruction inspection scenarios.

The conclusion goes here.

Acknowledgments

This research was supported by Autodesk and the Stanford Robotics Center.

References

[1] S. Cabral, M. Klimenka, F. Bademosi, D. Lau, S. Pender, L. Villaggi, J. Stoddart, J. Donnelly, P. Storey, and D. Benjamin. A contactless multi-modal sensing approach for material assessment and recovery in building deconstruction. *Sustainability*, 17:585, 2025. doi: 10.3390/su17020585. URL https://doi.org/10.3390/su17020585.

[2] C. B. Markenson. *A construction robot evaluation framework*. Doctoral dissertation, Stanford University, 2022. URL https://purl.stanford.edu/gj772cx6353. Stanford Digital Repository.

APPENDIX A: COMPARISON OF MANUAL AND ROBOTIC WORKFLOWS USING THE POP FRAMEWORK

This appendix provides a comprehensive comparison of the manual and robotic workflows using the Product–Organization–Process (POP) section from the Robot Evaluation Framework (REF).

POP	**Manual**	**Robot**
Product		
Single / Multi-task	Multi-task	Multi-task
Interior / Exterior	Interior (Partition Walls)	Interior, Exterior
Hardware	Tripod-mounted multi-modal sensor rig with external computer for data capture and processing	ANYmal Gen D robot with onboard LiDAR, RGB cameras, embedded compute unit, optional mounted sensors
Mobility	Full mobility – but slow, given the huge system movements	Full Mobility (four-legged all-terrain mobility, zero-turn radius, high agility)
Degree of mechanization / Autonomy	–	Between Level 3 and 4: semi-autonomous to high autonomy; human input needed for initial mapping or layout changes
Control interface	–	Controlled via tablet app or PC GUI
Software / sensors	RGB+Depth camera, Thermal Camera, LiDAR	Visual, Thermal cameras, LiDAR scanner, ultrasonic mic
Power and communications	Battery-powered with external charging; Wi-Fi connection	Li-ion battery, Wi-Fi for ROS and remote communication
Clearance	3m × 3m for 3 workers	Robot: 1100×700×1000mm; Operator: 1m × 1m
Site conditions	Must meet construction safety standards	Feasible in open, flat indoor space. May be limited by clutter, low ceilings, Wi-Fi dead zones, etc.
Reach (workspace)	Requires ladder to reach high areas	Max sensor-to-surface distance: 4m. Limited by sensor height and tilt angle.
Project types	Deconstruction	Deconstruction, Inspection, Safety Monitoring, Disaster Response, etc.
Units of work / zone	Depends on patch layout	105 ft² per section, or 54 ft² per fixed position
Organization		
Skills / experience	Manual tripod setup, marking, scanning	Visual monitoring, basic spatial awareness, touchscreen control, joystick navigation
Labor supply	3 people (height, camera, measurement)	1 operator, optionally 1 support technician
Organizations involved	General contractor, sensor supplier, facility owner	Robot provider, data team, GC, facility owner
Team experience	N/A	Basic interface familiarity, troubleshooting
Process		
Process changes	Allocate human resources; data assessment	Schedule robot time and charging; check image quality before downstream use
Handoffs of information	3 (capture → prepare → interpret)	3 (capture → prepare → interpret)
Data acquisition	Manual patch-by-patch RGB-D and thermal sensing; quality depends on operator	Continuous autonomous capture with more consistent results
Quality control (QC)	Rough on-site check; full QC deferred; rescans may be needed	Same, but smoother acquisition; improvement still needed

TABLE I
COMPARISON OF MANUAL AND ROBOTIC WORKFLOWS USING THE POP SECTION FROM THE ROBOT EVALUATION FRAMEWORK (REF).

APPENDIX B: ROBOTIC WORKFLOW DIAGRAM

This appendix presents a detailed flow diagram of the robotic workflow, illustrating the key steps, decision points, and points of information handoff throughout the process.

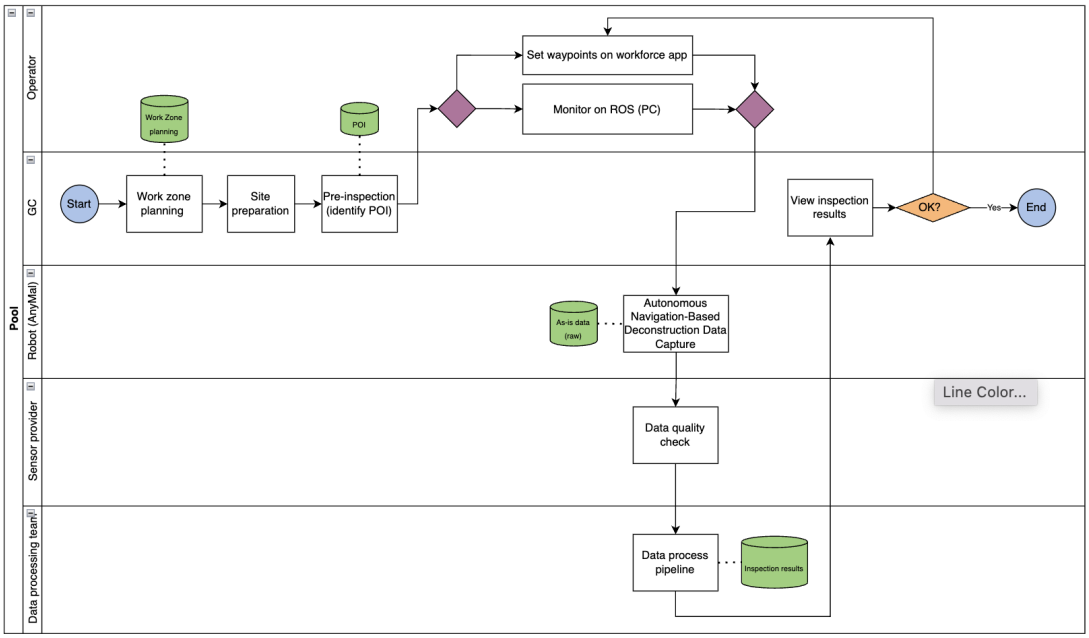

Fig. 5. Role-Based Diagram of the Robotic Deconstruction Workflow.