Towards End-to-End Autonomous Wall Defect Inspection for Interior Finishing

Alessio Lovato[1,*], Edoardo Romiti[1], Luca Muratore[1] and Nikolaos Tsagarakis[1]

***Abstract*— Shallow defects on putty-coated walls are difficult to observe under conventional illumination, yet their precise localization is essential for targeted sanding or putty reapplication. We present an autonomous mobile-manipulation system that extends controlled photometric-stereo inspection from fixed-view sensing to wall regions and registers individual defects in a common site reference frame. Given operator-selected regions, the robot autonomously plans inspection viewpoints, navigates, aligns a multi-illumination sensing tool with the wall, and registers instance-level detections into a global defect map. In a separate five-run localization study, the largest location-wise mean planar error was approximately 5 cm. During a nine-viewpoint inspection, the system obtained 0.3504 m² of depth-confirmed coverage from 0.3708 m² of planned coverage and successfully produced a globally registered defect map. The experiments demonstrate an end-to-end transition from standalone defect perception to autonomous quality inspection, while identifying mobile-base localization and open-loop arm tracking as the main factors limiting viewpoint accuracy and robustness.**

A video demonstration can be found at: https://bit.ly/4xo1hEz

Fig. 1: The CONCERT mobile manipulator autonomously inspecting a putty-coated test wall with the controlled-illumination RGB-D sensing tool.

I. INTRODUCTION

Construction robots are progressing from isolated operations toward integrated workflows for bricklaying, surface preparation, and painting [1], [2], [3]. As these processes become automated, quality assessment should become part of the robotic workflow rather than remain a separate manual operation. Existing robotic systems use laser scanning to assess wall alignment and evenness [4], [5], or combine visual inspection with spatial mapping and corrective action [6], [7]. However, they address structural defects with strong appearance or depth cues, defects in fresh plaster, or quality decisions over coarse wall patches.

Interior finishing presents a different problem. After putty application and before painting, sparse irregularities must be localized so that rework can target individual defects rather than complete wall regions and these defects have weak texture and depth contrast under conventional lighting. Human workers reveal them using grazing illumination, motivating controlled acquisition under several light directions. Our previous work encoded photometric-stereo estimates of mean curvature, relative height, and albedo in a three-channel representation and demonstrated instance segmentation of dried-putty defects from fixed-view acquisitions [8]. The remaining construction-robotics challenge is to reproduce those acquisition conditions autonomously over multiple wall regions and to express detections in a common site reference frame.

This work integrates that perception method into the CONCERT mobile manipulator [9]. From wall regions selected in a site point cloud, the system autonomously plans inspection viewpoints, navigates between walls, resolves arm reachability through mobile-base repositioning when required, aligns the custom sensing tool, collects multi-illumination RGB-D observations, and registers detected defects in a global voxel map. The contribution is therefore the integration of the ecosystem into a single inspection workflow intended to support targeted downstream rework.

II. AUTONOMOUS INSPECTION SYSTEM

A. Platform, Mapping, and Navigation

CONCERT is a modular mobile-manipulation platform equipped with front and rear Velodyne VLP-16 LiDARs and an interchangeable arm and end-effector. The custom sensing tool integrates a RealSense D435f RGB-D camera and four independently controlled lights around its aperture (Fig. 2).

[1]Humanoids and Human Centered Mechatronics, Istituto Italiano di Tecnologia, Via S. Quirico, 19D, 16163 Genoa, Italy

*Corresponding author alessio.lovato@iit.it

Fig. 2: The custom sensing tool.

The lights are activated sequentially while the tool remains fixed and the camera operates approximately 0.40 m from the wall, a distance selected to balance image resolution, sensing footprint, depth availability, and collision clearance.

The system requires a point-cloud model of the site for localization, wall selection, navigation, and collision checking. When an as-built model is unavailable, the operator teleoperates the platform while RTAB-Map constructs the site map from LiDAR measurements [10]. Alternatively, a registered CAD- or BIM-derived point cloud can be used. In both cases, the origin of the reference point cloud defines the common map/world frame used in the pipeline. HDL [11] is used to localize the robot with scan matching of the live 3D LiDAR data against this reference model; Nav2 is then employed to perform obstacle-aware navigation using the known environment and live LiDAR observations [12].

B. Coverage and Mobile Manipulation

For each selected wall region, the coverage planner projects the surface into a wall-aligned two-dimensional raster. A target mask describes the requested surface, whereas a valid-center mask restricts tool-center placement. The latter is produced by eroding the selected region according to the physical tool footprint and a safety margin; detected openings are excluded from both domains. The projected camera footprint is computed from the camera model, image crop, camera-to-tool transform, and requested standoff. Candidate viewpoints follow a boustrophedon pattern and are selected according to their marginal contribution to uncovered cells [13], [14]. Optional route optimization changes their visiting order without altering achieved coverage.

MoveIt 2 first attempts each viewpoint with the arm-only planning group. If the target is unreachable, a whole-body planning group searches for a mobile-base configuration from which arm-only planning can succeed. If succeeds, the arm is homed, Nav2 moves the base along the extracted waypoints, and the target is replanned with the arm-only group. Before acquisition, a task-space controller estimates the wall plane from the depth image and applies bounded corrections that align the tool with the measured normal via CartesIO [15]. Overlap between adjacent sensing footprints partly accommodates the translational error, whereas correct orientation is important for consistent projected footprints and controlled illumination.

C. Perception and Global Registration

At each reached viewpoint, the system records aligned depth and the camera pose in the map frame, then acquires four RGB images under sequential illumination. Automatic exposure was disabled because adaptation to the inspection lights required almost 10 s; the lights were therefore operated at maximum brightness and the RGB gain and exposure were manually set to 4 and 15, respectively. Once the fourth image is accepted, the acquisition is queued for asynchronous processing so that the robot can proceed to the next viewpoint.

Photometric stereo estimates a dense surface-normal field from the four images [16]. Mean curvature, relative height, and albedo are assembled into the three-channel representation used by the Mask R-CNN instance-segmentation model [17], [8]. For each accepted mask, valid aligned-depth pixels are back-projected to 3D and transformed into the common map/world frame defined by the reference site model. A defect record stores its class, confidence, and occupied voxel indices, allowing the resulting global map to be queried or visualized by class and position.

D. Mission Coordination

The behavior tree in Fig. 3 coordinates startup, wall selection, navigation, manipulation, acquisition, and failure handling [18]. A reactive safety guard checks joint-state and map-to-odometry availability throughout execution. Startup either loads a reference point cloud and triggers relocalization or uses the map produced in SLAM mode; execution begins only after the required transforms and Nav2 are available.

The operator selects and approves wall regions and coverage plans before deployment using the Graphical User Interface (GUI) displayed in Fig. 4. Autonomous execution then iterates over walls and their precomputed viewpoints. The arm-only/whole-body recovery sequence is retried at most twice for each viewpoint, and an optional approval barrier can be enabled before arm motion. After alignment and acquisition, perception runs asynchronously; therefore, the mission waits for queued jobs before reporting completion.

III. EXPERIMENTAL DEMONSTRATION

The system was evaluated in an environment containing two reconstructed walls coated with putty. The site was first mapped with RTAB-Map and then inspected through autonomous missions. The localization experiment described below was conducted separately from the nine-viewpoint run used to evaluate coverage, alignment, and defect-map generation. A summary of the results is reported in Table I.

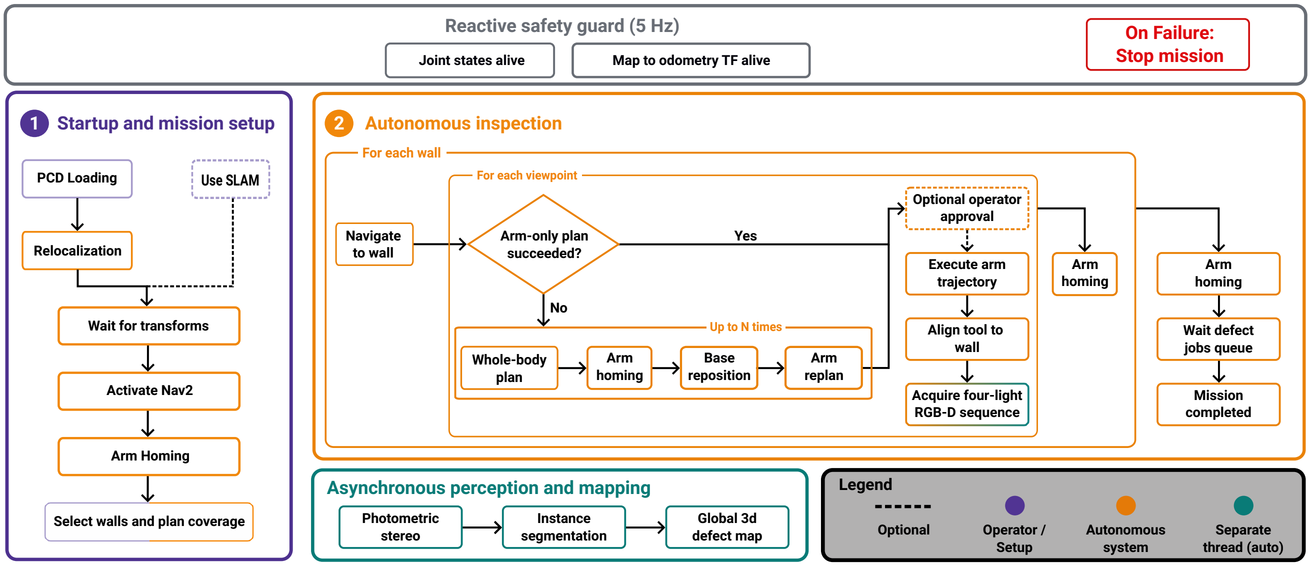
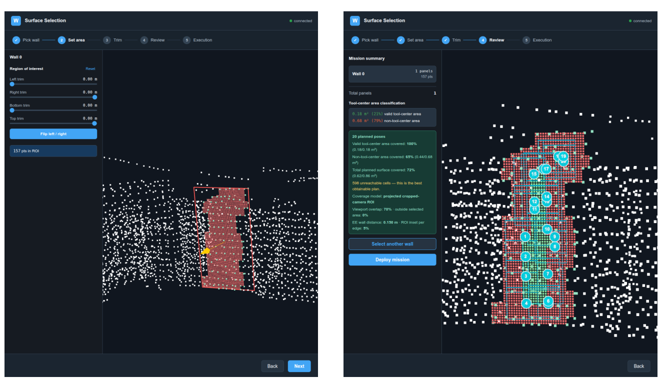


Fig. 3: Behavior-tree-based mission logic. A reactive guard monitors robot-state and localization availability. After startup and wall selection, the robot iterates over walls and viewpoints, attempting arm-only planning before whole-body base repositioning. Each reached viewpoint triggers tool alignment and four-light RGB-D acquisition, while perception and global registration run asynchronously.



(a) Wall selection (b) Planned viewpoints

Fig. 4: Graphical User Interface for selecting wall regions and reviewing the planned inspection poses.

A. Localization Study

A Leica Nova MS60 MultiStation provided an external position reference through a prism rigidly mounted on the mobile base. Because one prism does not determine the relative yaw, two calibration positions along the robot's local x-axis were combined with the onboard Vectornav VN100 yaw's measurements to estimate the MS60-to-map orientation. Five runs were performed, each comprising the origin and five Nav2 waypoints, for 30 location visits in total. Reference poses reconstructed from MS60 and IMU data were compared with the HDL transform chain at each location.

Localization deviations increased while the robot was moving and stabilized after it stopped, although yaw and vertical estimates remained noisy even during stationary periods. This observation supports evaluating reachability and planning arm trajectories only after navigation completes. Across the six locations, the largest mean planar error, computed over the five runs, was approximately 5 cm and should be noted that this value is subject to uncertainty from the IMU yaw and the initial MS60-to-map orientation used to reconstruct the reference pose.

B. Inspection Results

The nine-viewpoint run assessed coverage and tool alignment on one selected wall region. Complete observation was required only for the eroded legal tool-center area; the planner maximized coverage of the surrounding selected surface where feasible. The total executed surface coverage was 0.3504 m^2, compared with 0.3708 m^2 of planned coverage. The intersection between the planned and executed coverage masks was 0.3208 m^2, corresponding to 86.52% of the planned coverage area.

Wall alignment generally reduced the larger initial orientation errors. Across the nine viewpoints, absolute residual errors remained below 1.5° in roll, 2.8° in pitch, and 1.2° in yaw. Alignment also introduced predominantly vertical viewpoint offsets. These were attributed in part to open-loop joint tracking: the first gravity-loaded pitch joint exhibited a mean absolute tracking error of approximately 0.49°, which can produce centimeter-scale end-effector displacement over the arm length.

At least ten end-to-end missions were conducted with different wall selections and recovery conditions. In three missions, a requested tool-to-wall distance below 8 cm, localization error, and OctoMap padding led MoveIt 2 to report a virtual collision near the wall. A less frequent failure occurred when whole-body planning placed the base too close to the wall for Nav2 to execute the proposed path.

TABLE I: Summary of the reported experiments.

Metric	Value
Localization runs	5
Evaluated locations per run	6
Largest location-wise mean planar error	$\sim$0.05 m
Inspection viewpoints	9
Operator-selected region	0.5740 m^2
Legal tool-center area	0.1900 m^2
Planned coverage	0.3708 m^2
Depth-confirmed coverage	0.3504 m^2
Scan time, excluding planning/navigation	436 s

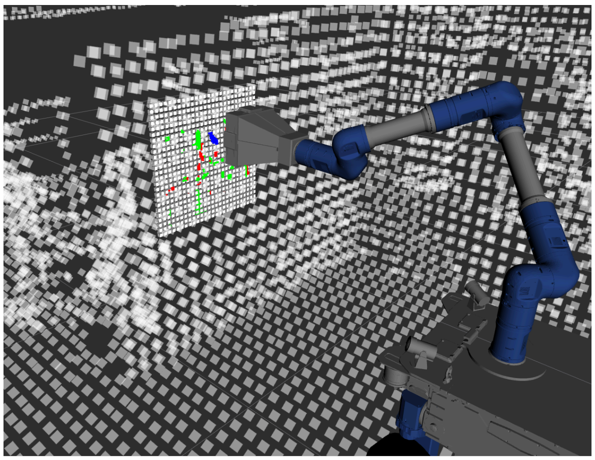


Fig. 5: Global defect map from the nine-viewpoint run, with registered detections overlaid on the rasterized selected wall region.

The reported nine-viewpoint scan required 436 s, excluding coverage planning and navigation, with arm velocity and acceleration limited to 5% of the motor limits and considering that each trajectory planning takes 5 s. Detection accuracy was not re-evaluated because it was assessed in our previous work [8].

IV. CONCLUSION

This work connects controlled photometric-stereo sensing with autonomous mobile manipulation to produce an instance-level wall-defect map in site coordinates. The integrated workflow covers site mapping, operator wall selection, coverage planning, navigation, reachability recovery, tool alignment, acquisition, and asynchronous registration. Its repeated operation in the two-wall test environment demonstrates a concrete step from standalone perception toward autonomous quality assessment for interior finishing and an inspection output suitable for targeted rework.

Mobile-base localization and open-loop arm tracking limit viewpoint accuracy after alignment, while the manually tuned camera exposure and gain limit the adaptability of the current sensing configuration.

Future work will therefore target more accurate localization, closed-loop arm control, and navigation-aware recovery for whole-body solutions. Larger construction-site trials should additionally quantify mission success, inspection throughput, and end-to-end defect-localization accuracy.

ACKNOWLEDGMENT

This work was supported by the European Union's Horizon Europe research and innovation programme (grant number 101120731 MAGICIAN).

During the preparation of this work the author(s) used OpenAI's Chatgpt in order to improve language and readability. After using this tool/service, the author(s) ensured that the content was not changed and assume(s) full responsibility of the publication.

REFERENCES

[1] P. Štibinger, G. Broughton, F. Majer, Z. Rozsypálek, A. Wang, K. Jindal, A. Zhou, D. Thakur, G. Loianno, T. Krajník, *et al.*, "Mobile manipulator for autonomous localization, grasping and precise placement of construction material in a semi-structured environment," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2595–2602, 2021.

[2] X. Jiang and X. Li, "Robotized interior finishing operations with visual feedback," *Industrial Robot: the international journal of robotics research and application*, vol. 49, no. 1, pp. 141–149, 2022.

[3] E. Asadi, B. Li, and I.-M. Chen, "Pictobot: A cooperative painting robot for interior finishing of industrial developments," *IEEE Robotics & Automation Magazine*, vol. 25, no. 2, pp. 82–94, 2018.

[4] R.-J. Yan, E. Kayacan, I.-M. Chen, L. K. Tiong, and J. Wu, "Quicabot: Quality inspection and assessment robot," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 506–517, 2019.

[5] H. Wu, M. Ma, Y. Yang, L. Han, and S. Wu, "On-site measuring robot technology for post-construction quality assessment of building projects," *Buildings*, vol. 14, no. 10, p. 3085, 2024.

[6] L. Yang, B. Li, J. Feng, G. Yang, Y. Chang, B. Jiang, and J. Xiao, "Automated wall-climbing robot for concrete construction inspection," *Journal of Field Robotics*, vol. 40, no. 1, pp. 110–129, 2023.

[7] C. Zhao, H. Bai, L. Li, K. Yang, Z. Yang, W. Chen, and C. Liu, "Dual-arm cooperative construction robots for wall repair with adaptive spraying and troweling," *Automation in Construction*, vol. 181, 2026.

[8] A. Lovato, E. Romiti, L. Muratore, and N. Tsagarakis, "Shedding light on walls: Photometric stereo for autonomous defect detection," in *2026 IEEE 22nd International Conference on Automation Science and Engineering (CASE)*, 2026.

[9] L. Rossini, E. Romiti, A. Laurenzi, F. Ruscelli, M. Ruzzon, L. Covizzi, L. Baccelliere, S. Carrozzo, M. Terzer, M. Magri, *et al.*, "Concert: A modular reconfigurable robot for construction," *Journal of Field Robotics*, 2025.

[10] M. Labbé and F. Michaud, "Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of field robotics*, vol. 36, no. 2, pp. 416–446, 2019.

[11] K. Koide, J. Miura, and E. Menegatti, "A portable three-dimensional lidar-based system for long-term and wide-area people behavior measurement," *International Journal of Advanced Robotic Systems*, vol. 16, no. 2, 2019.

[12] S. Macenski, F. Martín, R. White, and J. Ginés Clavero, "The marathon 2: A navigation system," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[13] H. Choset, "Coverage of known spaces: The boustrophedon cellular decomposition," *Autonomous Robots*, vol. 9, no. 3, pp. 247–253, 2000.

[14] E. Galceran and M. Carreras, "A survey on coverage path planning for robotics," *Robotics and Autonomous systems*, vol. 61, no. 12, pp. 1258–1276, 2013.

[15] A. Laurenzi, E. M. Hoffman, L. Muratore, and N. G. Tsagarakis, "Cartesi/o: A ros based real-time capable cartesian control framework," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 591–596.

[16] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical engineering*, vol. 19, no. 1, pp. 139–144, 1980.

[17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[18] M. Iovino, E. Scukins, J. Styrud, P. Ögren, and C. Smith, "A survey of behavior trees in robotics and ai," *Robotics and Autonomous Systems*, vol. 154, 2022.