# ATOM-ANT3D in Action: 3D Surveying from Confined Spaces to Urban Environments

Ahmad El-Alailyi [1,2], Luca Perfetti [3], Fabio Remondino [2], Francesco Fassi [1]

[1] 3D Survey Group, ABC Department, Politecnico di Milano, Milano, Italy - (ahmad.elalailyi, francesco.fassi)@polimi.it
[2] 3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation (FBK), Trento, Italy - (aelalailyi, remondino)@fbk.eu
[3] Department of Civil, Architectural, Environmental Engineering and Mathematics (DICATAM), Università degli Studi di Brescia, Brescia, Italy - luca.perfetti@unibs.it

**Keywords:** Multi-camera 3D Mobile Mapping, Multi-camera V-SLAM, Optimization, 3D Reconstruction, Complex Environments.

**Abstract**

3D digitization has advanced considerably in both hardware and algorithmic development, driven by the growing demand across a wide range of applications. In parallel, increasing attention has been given to methods that accelerate the digitization workflow by reducing computational time while preserving reconstruction accuracy. In this work, we evaluate the in-house handheld multi-camera mobile mapping system (MMS), ATOM-ANT3D, equipped with a V-SLAM-based 3D reconstruction pipeline, in a range of complex real-world environments. The processing pipeline relies on a multi-instance V-SLAM framework that processes synchronized image streams in real time to generate sparse 3D reconstructions and camera poses (i.e., trajectories), which are subsequently refined through the developed near-real-time optimization techniques to improve geometric accuracy of the final outputs. The proposed workflow is assessed across diverse case studies to test its generalizability, including a historical tower composed of rooms and narrow staircases, complex underground tunnels, inner-corridor urban city mapping, and confined infrastructure. The results demonstrate centimeter-level accuracy comparable to photogrammetric reconstruction, while significantly reducing computational time up to ca. 93%. These findings highlight the potential of the proposed system as an efficient and reliable solution for rapid and accurate 3D reconstruction in complex environments.

## 1. Introduction

3D surveying and mapping have advanced rapidly due to improvements in sensing technologies, algorithms, and computing power, enabling expanded applications across heritage, infrastructure, industry, and beyond. Among visual 3D reconstruction methods, photogrammetry (Luhmann et al., 2007) remains the method of choice for its proven geometric accuracy and rich texture information. However, static images acquisition is often time-consuming, requiring sufficient image overlap, controlled image capture, multiple setups, and the offline processing is compute- and memory-demanding, with runtime growing with image count, resolution, and scene complexity (Leduc et al., 2019; Vincke et al., 2019; Berra and Peppa, 2020). These factors can constrain throughput in large, complex, or hard-to-access environments.

Terrestrial Laser Scanning (TLS) provides an alternative sensing modality that delivers metrically scaled, highly accurate point clouds and remains robust in low-texture or low-light conditions (Muralikrishnan, 2021). However, tripod-based operation requires multiple scan stations, careful line-of-sight planning, target placement or targetless registration, and frequent instrument relocation. In large or complex sites, the number of setups, occlusions, and access constraints increases acquisition time and logistics. Post-processing likewise grows with project scale; for example, registering and optimizing large scan networks, de-ghosting moving objects, and data management become computationally and labor-intensive (Rüther et al., 2012; Waqar et al., 2023).

Mobile mapping systems (MMSs), being handheld, back-packed or on vehicles and leveraging simultaneous localization and mapping (SLAM), whether visual, LiDAR, or multi-sensor, have gained traction for their acquisition speed and operational flexibility, enabling continuous, dynamic 3D capture with real-time positioning and quality feedback (Elhashash et al., 2022). MMSs address key limitations of TLS by enabling flexible deployment in inaccessible, hazardous, or large-scale environments where tripod-based instruments are difficult to operate. In the case of visual MMS, the high frame rate and continuous motion of the platform allow rapid image acquisition, substantially accelerating field work. Visual SLAM (V-SLAM) exploits image sequences to jointly estimate a sensor’s pose and the surrounding scene structure (Davison et al., 2007).

Handheld and portable V-SLAM-based mobile mapping systems have recently become an attractive solution for integrated 3D acquisition, localization, and reconstruction, with particular interest in configurations based on multi-camera setups (i.e., systems with two or more cameras) (Ortiz-Coder and Sánchez-Ríos, 2019; Kuo et al., 2020; Torresani et al., 2021; Elalailyi et al., 2024; Perfetti et al., 2024). Despite their versatility, MMSs outputs can exhibit lower raw density and accuracy than static methods and typically require extensive post-processing and heterogeneous sensor integration (Szrek et al., 2024; Zhu et al., 2024).

### 1.1 Paper’s Aim

This paper recaps the system capabilities of the developed ATOM-ANT3D handheld multi-camera MMS, with its data acquisition and processing strategies. The 3D surveying capabilities are showcased across a range of challenging environments, from complex indoor environments to urban city mapping. The processing pipeline performs real-time multi-instance V-SLAM processing, supporting both multi-stereo and monocular trajectory estimations, and applies near-real-time optimizations (El-Alailyi et al., 2025) to enhance the camera's pose accuracy. Performances are assessed across four real-world case studies: (1) a vertical tower with circular rooms connected through narrow staircases, (2) an underground tunnel with restricted mobility, (3) urban inner-city corridors, and (4) a confined infrastructure pipe. Geometric accuracy is evaluated against a ground truth 3D point cloud, comparing both accuracy and end-to-end processing time with standard multi-view stereo photogrammetry, reporting the resulting runtime improvements.

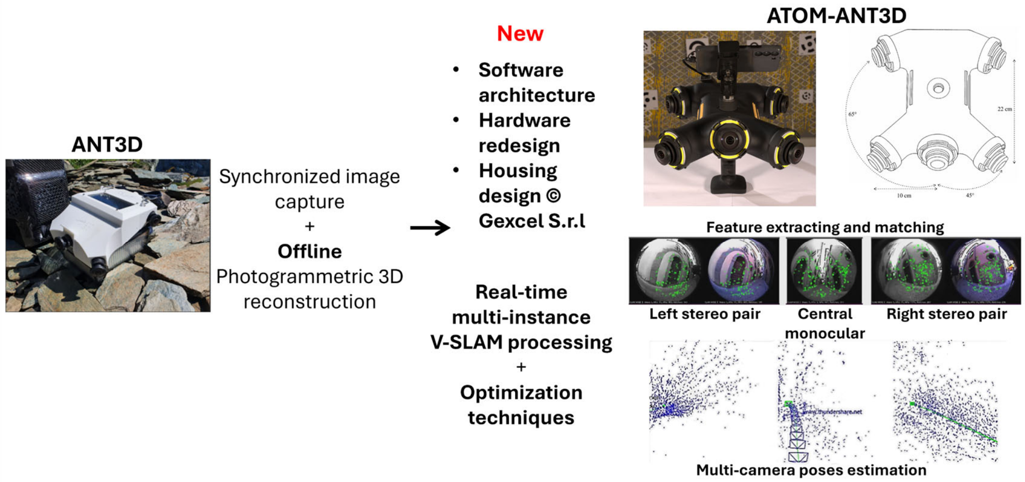


Figure 1. ATOM-ANT3D, version 2.0 of ANT3D handheld multi-camera fisheye MMS.

## 2. Methodology

### 2.1 ATOM-ANT3D

ATOM-ANT3D (Elalailyi et al., 2024; Perfetti et al., 2026) is the version 2.0 of the ANT3D handheld multi-camera fisheye MMS (Perfetti et al., 2024) (Figure 1). It features a redesigned architecture, including new software for data capture and management, redesigned hardware for processing and data handling, and a new industrial housing developed by Gexcel s.r.l. (Gexcel, 2026) In addition, its processing strategy (Section 2.3) differs from the standard photogrammetric reconstruction of ANT3D by incorporating a multi-instance V-SLAM framework for real-time image processing and near-real-time optimization techniques for camera poses refinement.

### 2.2 System Calibration

The ATOM-ANT3D is calibrated using our 3D calibration testfield designed with rich textures and feature patterns at varying colors and spatial resolutions (Figure 2). Some 176 coded, frame-filling targets with known 3D coordinates are uniformly distributed and divided equally into control points and check points. Two calibrated Brunson scale bars of length 1.099 m are included in the scene, each with a manufacturer-specified length uncertainty of 0.0015 mm. Both intrinsic and extrinsic camera parameters are estimated from the synchronized multi-camera images within a bundle-block adjustment. The process achieved a sub-millimeter root mean square error (RMSE) on check points, as reported in Table 1.

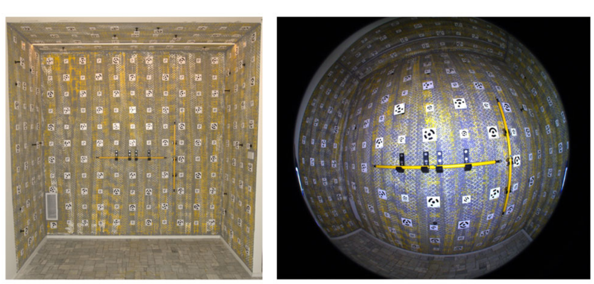

Figure 2. The 3D calibration wall (left) and a fisheye image acquired by the ATOM-ANT3D (right).

| | Check Points RMSE (mm) | | | |
|---|---|---|---|---|
| | X | Y | Z | XYZ |
| **ATOM-ANT3D (Rig configuration)** | 0.18 | 0.16 | 0.21 | 0.3 |

Table 1. The RMSE on check points for the ATOM-ANT3D 3D wall calibration.

### 2.3 Data Acquisition and Processing

Synchronized image streams from the field acquisitions were recorded at 3 fps and processed using the multi-instance V-SLAM (El-Alailyi et al., 2025).
For all datasets reported in Section 3, multi-instance V-SLAM was run on the full five-camera rig, with stereo mode used for the left and right side camera pairs (i.e., four cameras) with the largest baselines and higher image overlap, and monocular mode used for the fifth forward-facing central camera. Each V-SLAM instance produces the estimated camera poses (i.e., trajectory). The initial trajectory estimates have their own drift profile and outliers. Given the scale ambiguity of monocular trajectory estimation (Figure 3), a monocular scale recovery approach is applied to estimate a scale factor converting the trajectory from arbitrary to metric units, and additionally accounting for the local scale drift. To recover the metric scale, the consistency of the monocular camera pose estimates with the stereo pairs is exploited, whose known fixed baselines provides metrically scaled motion.

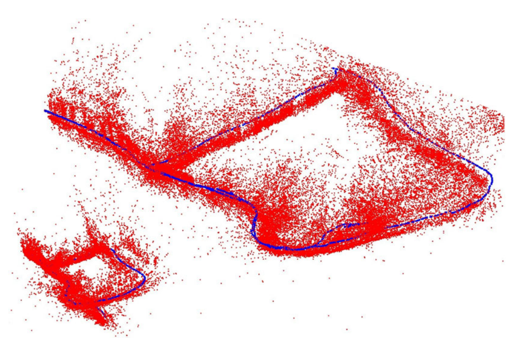

Figure 3. The V-SLAM trajectory estimation (blue trajectory) and the 3D sparse point cloud highlighting the ambiguity in scale between monocular (left) and stereo (right) configurations.

Over sliding temporal windows, the path length travelled according to the stereo (in meters) and the corresponding monocular (in arbitrary units) configurations are compared; their ratio yields a local scale estimate. Because these ratios can be corrupted by segments with unreliable stereo depth or monocular tracking failures, we aggregate all valid window ratios across the stereo cameras and estimate a robust global scale using median statistics with interquartile range (IQR) outlier rejection. This global factor is sufficient when scale is approximately constant; however, the monocular scale can drift over long sequences (e.g., low-parallax motion or abrupt rotations). To compensate for this, a local scale profile is built assigning each window-based ratio to the window's center timestamp and taking a per-timestamp median across available samples. The estimates are then smoothed by grouping frames into fixed-length bins and applying a piecewise-constant scale per bin. A per-bin translation offset is computed to enforce continuity at bin boundaries, preventing jumps in the scaled trajectory. The result is a continuous metric monocular trajectory that remains locally and globally consistent with the stereo reference. By utilizing redundancy in camera position estimations, the post-processing multi-camera pose graph optimization (El-Alailyi et al., 2025) combines available trajectory estimates into a single, globally consistent solution. It starts by initializing a nonlinear factor graph, where nodes are entire keyframe poses, and edges are relative pose constraints between successive keyframes. To improve resilience against outliers, Cauchy noise model combined with a loss function is used to account for pose uncertainty. A Gauss-Newton optimization is then applied, eliminating residuals to produce a globally optimized trajectory (Figure 4).

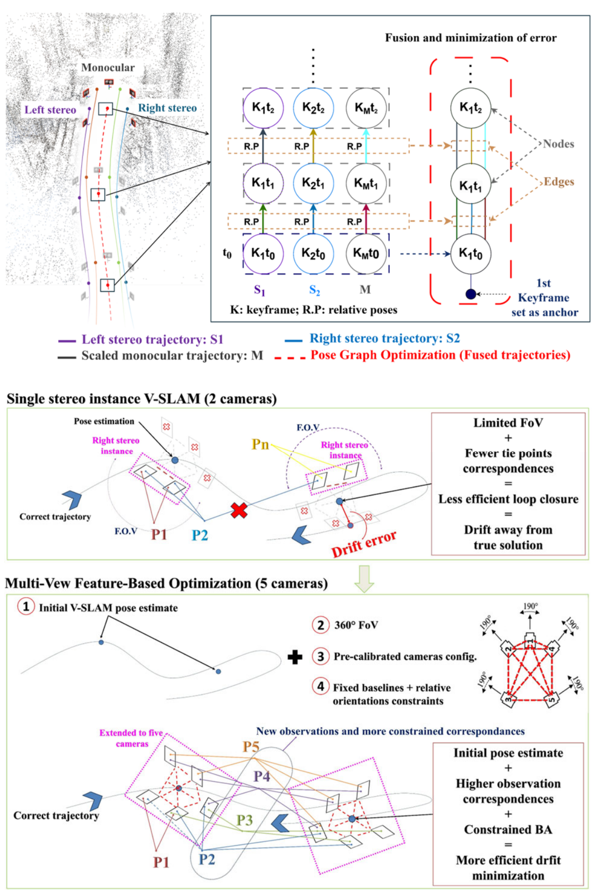


Figure 4. The multi-camera pose graph optimization (top) and the multi-view feature-based optimization (bottom) schemes.

The optimized trajectory is then refined using a multi-view feature-based optimization (Figure 4) (El-Alailyi et al., 2025). The initial trajectory supports loop-closure detection and matching between spatially adjacent cameras. The multi-camera rig pre-calibrated rigid relative orientation (i.e., baselines and orientations) is used in a constrained bundle adjustment (BA) to further refine the sparse 3D tie points cloud produced by multi-view triangulation. To ensure geometric consistency and precise camera position estimation, 3D points with large reprojection errors are successively filtered, and optimization is iterated until the overall average projection RMSE converges below 1 px threshold. Figure 5 demonstrates the full pipeline, from V-SLAM processing to subsequent optimizations. The final dense 3D reconstruction is then obtained through standard multi-view stereo photogrammetric depth maps and point cloud generation.

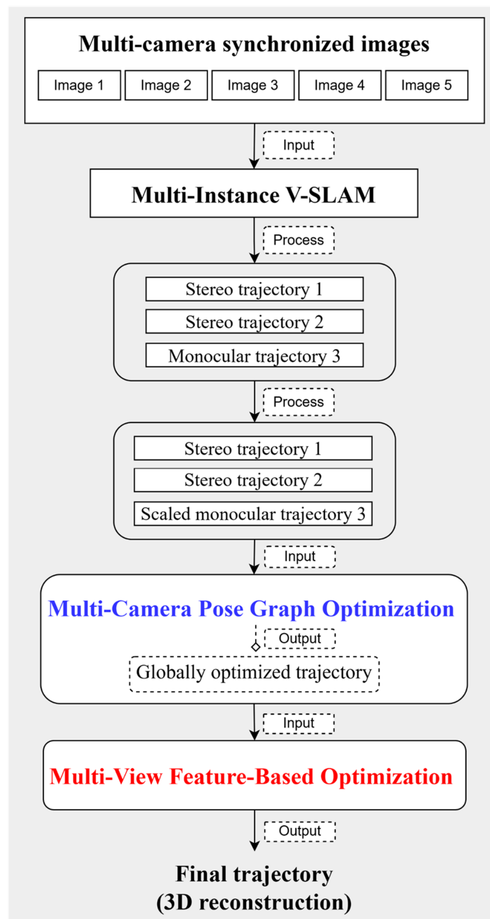


Figure 5. Scheme of the proposed multi-camera V-SLAM-based processing pipeline.

## 3. Datasets

### 3.1 San Vigilio Castle (Castagneta Tower and Tunnel)

The first case study focuses on the Castagneta tower at the San Vigilio Castle in Bergamo, Italy. It is composed of two environments: (1) the interior of the tower and (2) the tunnel (Figure 6). The tower consists of two circular rooms and extends to a height of ca. 11 m. The rooms are connected by narrow staircases, ca. 1 m wide, that extend diagonally with predominantly planar surfaces, sharp transitions between floor levels, and adequate surface texture. In contrast, the tunnel is characterized by narrow tubular passages with irregular surfaces, variable cross-sections, and constricted bends, in the absence of natural light. The floor surface and overall atmosphere are highly humid and uneven, requiring cautious navigation, especially near steep areas. The tunnel extends for ca. 80 m in one direction.

Figure 6. Representative fisheye images from the San Vigilio Castle case study acquisition, showing the interior spaces of the Castagneta tower (top) and the tunnel environment (bottom).

### 3.2 Città Alta Inner-City Corridors

The second case study focuses on a portion of the inner-city corridors of Bergamo’s Città Alta, Italy, representing an urban-style surveying and mapping scenario. The inner-city corridors (Figure 7) are characterized by a dense historic urban fabric composed of narrow streets and stone-paved alleys with irregular geometries and rich surface textures. The environment includes dynamic elements, such as pedestrians and vehicular traffic. The corridor width ranges from ca. 2.5 to 6 m, with some open spaces reaching up to ca. 16 m (façade-to-façade). The total traveled scanning distance on foot is ca. 500 m. The building façade heights vary between ca. 3.5 and 16 m.

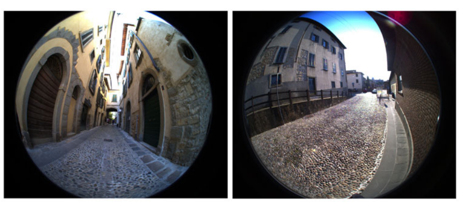

Figure 7. Representative fisheye images acquired by ATOM-ANT3D for the Città Alta inner-city corridors.

### 3.3 Infrastructure Pipe

The third case study consists of a confined, narrow, tubular conduit exhibiting a uniform circular cross section with an estimated inner diameter of ca. 1 m, resulting in highly constrained, predominantly single-directional movement.

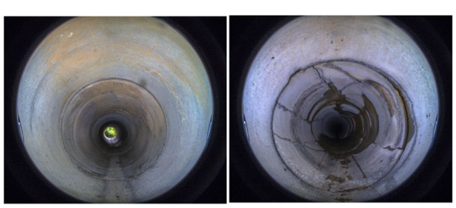

Figure 8. Representative fisheye images from the pipe acquisitions, illustrating the narrow and low texture variability of the conduit.

The interior surface is constructed of composite concrete segments with a smooth, very low-texture surface, making the environment particularly challenging for visual tracking. The environment lacks natural light and presents a nearly monotonous and highly symmetric structure extending ca. 30 m in length (Figure 8).

### 3.4 Evaluation Methodology

For the first two case studies, highly accurate laser scanning dense clouds are used as ground truth references. These reference datasets provide metrically reliable and high-density 3D representations suitable for quantitative geometric validation (Figure 9). In contrast, for the third case study, no external ground truth is available due to the geometric complexity, limited accessibility, and operational constraints of the environment. Therefore, a cross-discrepancy analysis is performed between the dense point cloud generated by the proposed V-SLAM-based and 3D reconstruction pipeline and the dense cloud reconstructed using a conventional photogrammetric workflow processed using Agisoft Metashape software (Agisoft Metashape, v2.2.2).

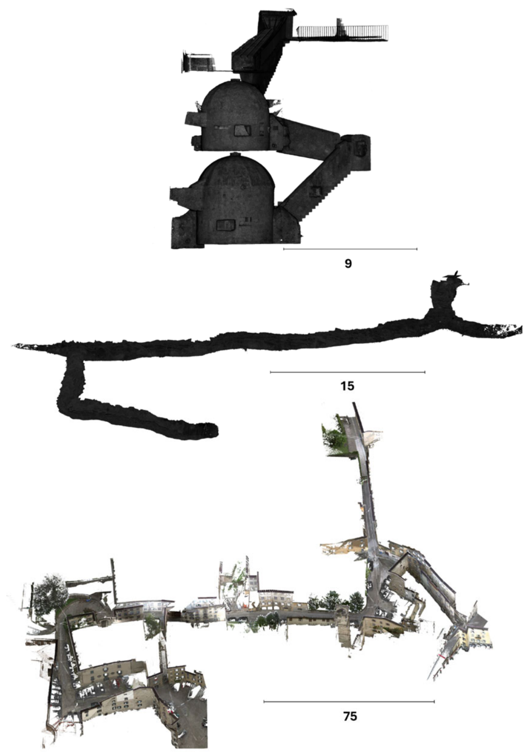


Figure 9. Ground truth reference dense point clouds acquired through high-accuracy laser scanning for the San Vigilio Castle (Castagneta tower and tunnel) (side views) and the Città Alta city corridor (top view) case studies. Scale in m.

For all case studies, evaluation is conducted on dense point clouds generated by (i) the proposed V-SLAM-based 3D reconstruction pipeline and (ii) a standard multi-view stereo photogrammetric workflow. The number of images acquired and processed by each method is reported in Table 2. In the proposed V-SLAM pipeline, all acquired images are processed in real-time; however, only a subset of selected keyframes by the V-SLAM is retained for subsequent near-real-time optimizations and dense reconstruction. In contrast, the photogrammetric workflow uses the full set of acquired images for reconstruction.

For the first two datasets, both reconstructed dense clouds are independently compared against the laser scanning ground truth using the Multiscale Model-to-Model Cloud Comparison (M3C2) algorithm implemented in CloudCompare (CloudCompare v2.14) for a robust assessment of geometric discrepancies. In addition to geometric accuracy, computational efficiency is evaluated by measuring the end-to-end processing time required by each workflow. This enabled quantification of the runtime reductions achieved by the proposed pipeline relative to the photogrammetric benchmark.

| | **Multi-Instance V-SLAM** | **Photogrammetry** |
|---|---|---|
| **Tower** | 606 img/cam | 2,070 img/cam |
| **Tunnel** | 393 img/cam | 1,813 img/cam |
| **Città Alta** | 917 img/cam | 2,679 img/cam |
| **Pipe** | 162 img/cam | 310 img/cam |

Table 2. Total number of images used by each of the approaches per each case study.

## 4. Results and Discussion

Figure 10 demonstrates the results of the multi-instance V-SLAM trajectories and the multi-camera pose graph optimization for each case study. The scaled monocular trajectories (blue trajectory) are globally consistent with the stereo-based (green and orange trajectories) estimates across all case studies, since their metric scale is derived from the stereo configurations. This suggests that the monocular solution, once properly scaled, is robust enough to preserve the overall path geometry. Moreover, the monocular trajectory appears less susceptible to tracking failure, likely due to its lower computational burden and the use of a single forward-looking view (central forward camera), which provides improved feature observability and a more favorable visual configuration for tracking. The multi-instance V-SLAM estimated stereo trajectories, right (green trajectory) and left (orange trajectory), demonstrate a case study-dependent behaviour.

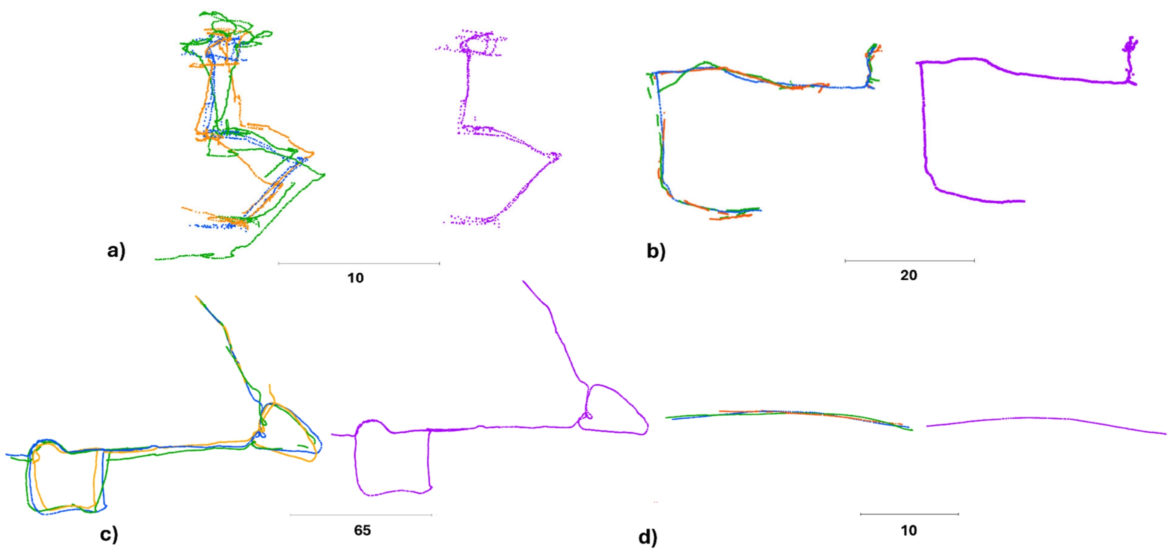


Figure 10. Trajectory estimates obtained from the multi-instance V-SLAM (left) and the corresponding multi-camera pose graph optimization (right) results for the four case studies: (a) Castagneta tower, (b) Castagneta tunnel, (c) Città Alta inner-city corridor, and (d) infrastructure pipe. Right stereo pair (green), left stereo pair (orange), scaled monocular (blue) trajectories, and the optimized global trajectory post multi-camera pose graph optimization (purple). Scale in m.

In the Castagneta tower, composed of rooms and narrow staircases, local tracking failures and drift are observed due to the challenging spatial configuration, limited visual continuity, and scarcity of features in certain areas. This behavior is even more evident in the tunnel case study, where tracking loss becomes more pronounced due to adverse conditions, including narrow tubular geometry, irregular surfaces, constricted bends, and the absence of natural light. Moreover, the humid and uneven ground, together with steep sections, imposes cautious yet sometimes unstable motion, which can generate abrupt viewpoint changes, motion blur, and reduced feature detectability. Under these conditions, the side stereo pairs are especially challenged in preserving continuous tracking. However, the multi-camera system ensures continuity of tracking and mapping, since temporary failure in one instance can be compensated for by the others, allowing the platform to remain localized within the global solution.

In the Città Alta inner-city corridor case study, tracking loss is generally less evident due to the smoother acquisition motion, presence of sufficiently rich visual features, and wider field of view offered by the urban corridor geometry, which together support more stable tracking and pose estimation. However, localized deviations can still be observed, mainly associated with the open nature of the environment, where moving pedestrians/cars introduce dynamic features, and with portions of the scene characterized by large sensor-to-object distances. In such areas, the combination of stereo baseline limitations and fisheye imaging effects can reduce the strength and reliability of depth estimation, leading to local trajectory deviations.

Conversely, in the confined infrastructure pipe case study, the stereo trajectory estimates differ. The narrowness and uniformity of the pipe strongly limit stereo depth and field of view, while the reduced availability of reliable visual features in some sections increases the sensitivity of the estimation and outliers. As a result, pose estimation in these areas becomes less stable, leading to larger discrepancies between the reconstructed trajectories.

The multi-camera pose graph optimization (purple trajectory) results provide a global integration and refinement of the estimated trajectories while maintaining trajectory continuity, and are further optimized using the multi-view feature-based


https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-17-2026 

optimization, which enforces the multi-camera rig constraints and iteratively refines the camera poses and 3D sparse reconstruction (Figure 11). Finally, dense 3D point clouds are produced using standard multi-view stereo photogrammetry. Figure 12 presents the M3C2 distance results for the V-SLAM-based and photogrammetric 3D reconstructions for each case study, enabling both visual and quantitative comparison against the laser scanning ground truth. Overall, both methods exhibit good agreement with the reference, with discrepancies mainly confined to localized regions. In the Castagneta tower, both approaches achieve the same standard deviation of ca. 3 cm, with consistent agreement across both the rooms and staircase sections. In the tunnel, photogrammetry performs slightly better, yielding a standard deviation of 4.4 cm, compared with 5.7 cm for the proposed V-SLAM pipeline. Larger deviations are visible near the tunnel entrance in the V-SLAM result, likely because photogrammetry benefits there from denser image coverage and stronger geometric redundancy. In the Città Alta inner-city corridor, the two methods show comparable results, with standard deviations of 11 cm for the V-SLAM-based reconstruction and 12 cm for photogrammetry. In this case, a similar error propagation can be observed in both reconstructions, especially at ground level and, locally, on building façades.

These higher deviations are likely associated with weakly constrained surfaces, viewing geometry, occlusions, dynamic elements, and long object distances. Overall, the results confirm that the proposed pipeline is capable of producing metrically reliable 3D reconstructions comparable to those obtained through standard photogrammetry.

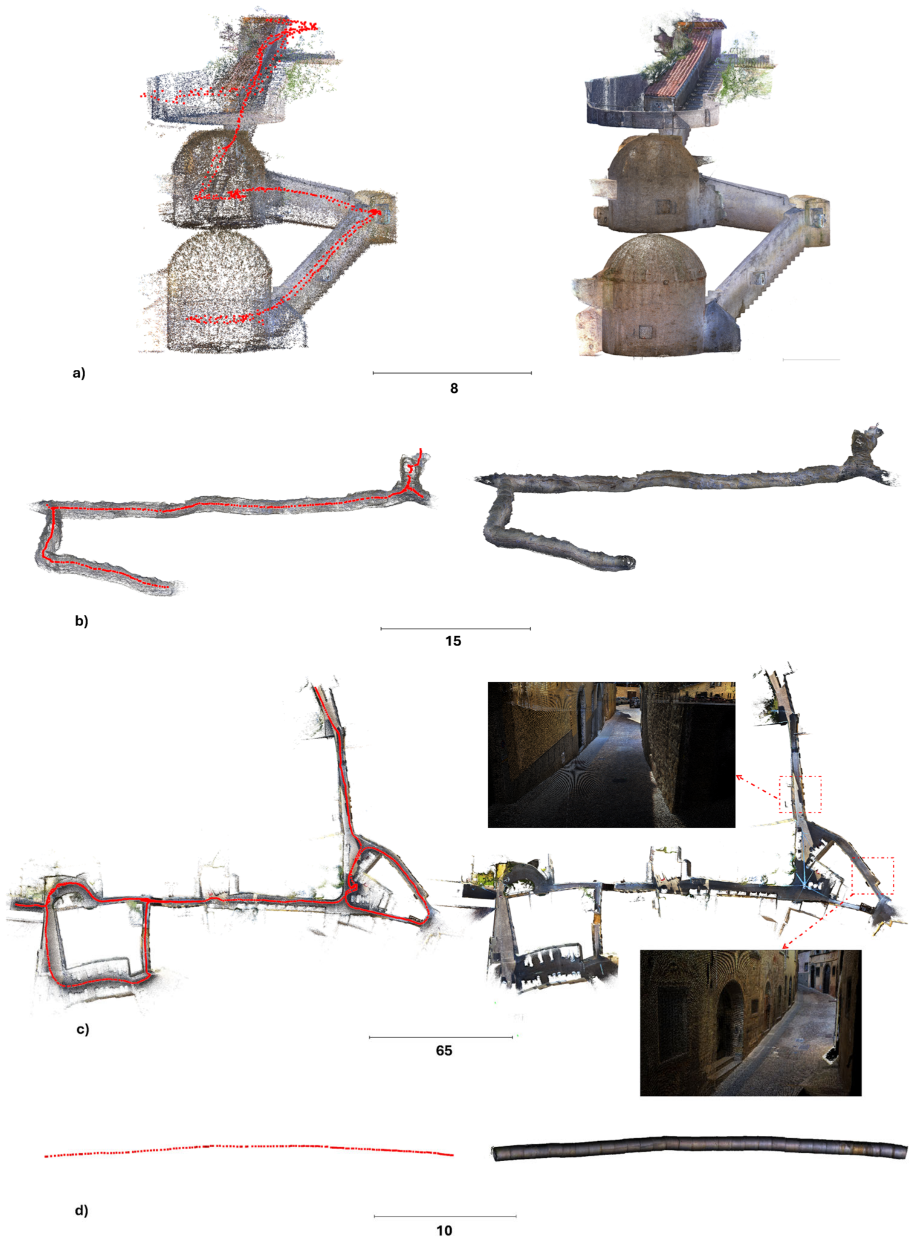


Figure 11. Results of the multi-view feature-based optimization for the four case studies: (a) Castagneta tower, (b) Castagneta tunnel, (c) Città Alta inner-city corridor, and (d) infrastructure pipe showing the optimized trajectories (red) overlaid on the corresponding sparse reconstructions, alongside the associated 3D dense point clouds. Scale in m.

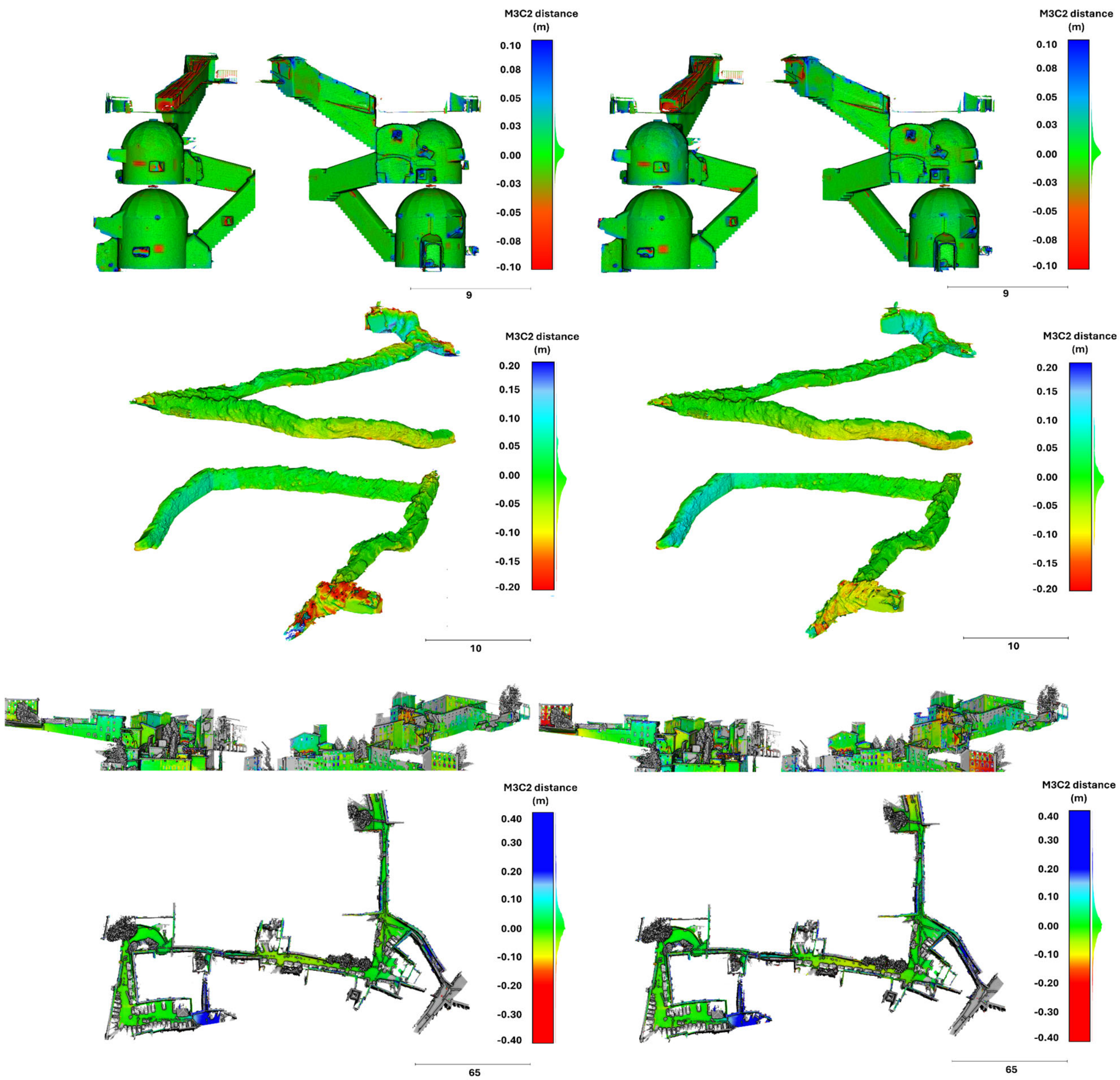


Figure 12. M3C2 distance results for the V-SLAM-based 3D reconstruction (left) and the photogrammetric 3D reconstruction (right), compared against the laser scanning ground truth for the Castagneta tower (top), Castagneta tunnel (middle), and Città Alta inner-city corridor (bottom) sites. Scale in m.

For the confined infrastructure pipe case study, no laser scanning ground-truth point cloud is available due to the extreme narrowness of the pipe and its challenging geometry. Therefore, the geometric evaluation is performed through a cross-comparison between the V-SLAM-based and photogrammetric 3D reconstructions by computing the M3C2 distances between the two dense point clouds (Figure 13). Deviation is observed to be higher at the boundaries, mainly because these areas are less constrained leading to registration or reconstruction errors. Nonetheless, the comparison indicates a generally reasonable agreement between the two models, with an overall reported standard deviation of 7 cm.

A comparison across the four case studies shows that the V-SLAM-based pipeline consistently required less processing time than photogrammetry (Table 3). For the Castagneta tower, processing time decreased from 13 hrs 50 mins hr to 1 hr 2 mins, while for the Castagneta tunnel, it decreased from 15 hrs 12 mins to 1 hr 5 mins. In the Città Alta inner-city corridors case, processing time is reduced from 17 hrs 26 mins to 2 hrs.

In contrast, the confined infrastructure pipe case showed the smallest difference, with processing time decreasing from 1 hrs 7 mins to 27 mins.

In relative terms, the Castagneta tunnel achieved a processing time reduction of ca. 93.0%, followed by the Castagneta tower (92.5%) and the Città Alta inner-city corridors (88.5%), whereas the confined infrastructure pipe showed a smaller but still notable reduction of ca. 59.5%.

The observed reduction in processing time is associated with the reduced number of images retained by the multi-instance V-SLAM pipeline relative to the photogrammetric workflow. This effect is further reinforced by the optimization procedures, which are executed in a near-real-time fashion by leveraging the preliminary solution generated directly from the real-time V-SLAM processing. As a result, the workflow focuses on refining an existing estimate instead of globally solving the 3D reconstruction problem from scratch, thereby contributing to the substantial time savings observed across the case studies.

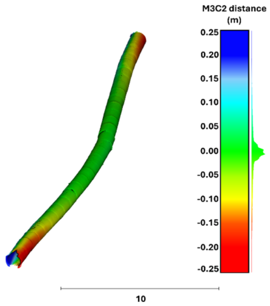


Figure 13. Infrastructure pipe M3C2 distance results for the cross-comparison between the V-SLAM-based and the photogrammetric 3D reconstructions. Scale in m.

| Processing Time | Multi-Instance V-SLAM + Optimizations | Photogrammetry |
|---|---|---|
| **Tower** | 1 hr 2 mins | 13 hrs 50 mins |
| **Tunnel** | 1 hr 5 mins | 15 hrs 12 mins |
| **Città Alta** | 2 hrs | 17 hrs 26 mins |
| **Pipe** | 27 mins | 1 hrs 7 mins |

Table 3. Processing time for the V-SLAM-based and the photogrammetric 3D reconstruction approaches.

## 5. Conclusions

The realized ATOM-ANT3D handheld multi-camera mobile mapping system, coupled with the proposed V-SLAM-based 3D reconstruction pipeline, proved to be a reliable solution across diverse real-world case studies with varying environmental conditions, geometries, and complexity. By processing images in real time and applying near-real-time optimization techniques, the proposed workflow achieved centimeter-level accuracy comparable to photogrammetric reconstruction, while significantly reducing computational time. While tracking losses are observed in the most challenging environments, the multi-camera configuration preserved the continuity of localization and mapping, thereby enabling the successful completion of the surveys. Overall, these results highlight the strong potential of the system (hardware + software) as an efficient and practical alternative for rapid, accurate, and scalable 3D reconstruction. Future work will focus on further enhancing the robustness of the V-SLAM pipeline, particularly under conditions of limited visual features, abrupt motion, tracking loss, and difficult geometric configurations.

## References

Agisoft Metashape, 2025. Agisoft: St. Petersburg, Russia, Version 2.2.2 www.agisoft.com/

Berra, E. F., & Peppa, M. V., 2020. Advances and Challenges of UAV SFM MVS Photogrammetry and Remote Sensing: Short Review. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., XLII-3/W12-2020, pp. 267–272.

CloudCompare, 2026. GPL software, Version 2.14. www.cloudcompare.org/

Davison, A. J., Reid, I. D., Molton, N. D., & Stasse, O., 2007. MonoSLAM: Real-time Single Camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(6), pp. 1052–1067.

El-Alailyi, A., Morelli, L., Trybała, P., Fassi, F., & Remondino, F., 2025. Optimizing Multi-Camera Mobile Mapping Systems with Pose Graph and Feature-Based Approaches. *Remote Sensing*, 17(16), 2810.

Elalailyi, A., Perfetti, L., Fassi, F., & Remondino, F., 2024. V-SLAM-aided Photogrammetry to Process Fisheye Multi-Camera Systems Sequences. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., XLVIII-2/W4-2024, pp. 189–195.

Elhashash, M., Albanwan, H., & Qin, R., 2022. A Review of Mobile Mapping Systems: From Sensors to Applications. *Sensors*, 22(11), 4262.

Gexcel. Available online: https://gexcel.it/it/ (accessed on 27 March 2026).

Kuo, J., Muglikar, M., Zhang, Z., & Scaramuzza, D., 2020. Redesigning SLAM for Arbitrary Multi-Camera Systems. *Proc. ICRA*, pp. 2116–2122.

Leduc, P., Peirce, S., & Ashmore, P., 2019. Short Communication: Challenges and Applications of Structure-from-Motion Photogrammetry in a Physical Model of a Braided River. *Earth Surf. Dynam*. 7, pp. 97–106.

Luhmann, T., Robson, S., Kyle, S., & Harley, I., 2007: Close Range Photogrammetry: Principles, Techniques and Applications. *Wiley*.

Muralikrishnan, B., 2021. Performance Evaluation of Terrestrial Laser Scanners - A Review. *Measurement Science and Technology*, *32*(7), 072001.

Ortiz-Coder, P., & Sánchez-Ríos, A., 2019. A Self-Assembly Portable Mobile Mapping System for Archaeological Reconstruction Based on VSLAM-Photogrammetric Algorithm. *Sensors*, 19(18), 3952.

Perfetti, L., El-Alailyi, A., Fassi, F., Remondino, F., Sgrenzaroli, M., and Vassena, G. P. M., 2026. Ant3D: Evolution of a Multi-Camera Fisheye System for 3D Mapping in Confined Spaces. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., XLVIII-2/W12-2026, pp. 375–382.

Perfetti, L., Fassi, F., Vassena, G., 2024. Ant3D—A Fisheye Multi-Camera System to Survey Narrow Spaces. *Sensors*, 24, 4177.

Rüther, H., Bhurtha, R., Held, C., Schröder, R., & Wessels, S., 2012. Laser Scanning in Heritage Documentation. *Photogramm eng remote sensing* 78, pp. 309–316.

Szrek, A., Romańczukiewicz, K., Kujawa, P., & Trybała, P. 2024. Comparison of TLS and SLAM Technologies for 3D Reconstruction of Objects with Different Geometries. *IOP Conference Series: Earth and Environmental Science,* 1295.

Torresani, A., Menna, F., Battisti, R., Remondino, F., 2021. A V-SLAM Guided and Portable System for Photogrammetric Applications. *Remote Sensing*, Vol.13(12), 2351.

Vincke, S., Bassier, M., & Vergauwen, M., 2019. Image Recording Challenges for Photogrammetric Construction Site Monitoring. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-2/W9, pp. 747–753.

Waqar, A., Othman, I., Saad, N., Qureshi, A.H., Azab, M., & Khan, A.M., 2023. Complexities for Adopting 3D Laser Scanners in the AEC Industry: Structural Equation Modeling. *Applications in Engineering Science,* 16, 100160.

Zhu, J., Li, H., & Tao, Z, 2024. Camera, LiDAR, and IMU Based Multi-Sensor Fusion SLAM: A Survey. *Tsinghua Science and Technology*.