# Ant3D: Evolution of a Multi-Camera Fisheye System for 3D Mapping in Confined Spaces

Luca Perfetti [1], Ahmad El-Alailyi [2,3], Francesco Fassi [2], Fabio Remondino [3], Matteo Sgrenzaroli [4], Giorgio Paolo Maria Vassena [1,4]

[1] Dept. of Civil Engineering, Architecture, Territory, Environment and Mathematics (DICATAM), Università degli Studi di Brescia, Brescia, Italy, (luca.perfetti; giorgio.vassena)@unibs.it
[2] Dept. of Architecture, Built environment and Construction engineering (ABC), Politecnico di Milano, Milano, Italy, (ahmad.elalailyi; francesco.fassi)@polimi.it
[3] 3D Optical Metrology Unit (3DOM), Fondazione Bruno Kessler (FbK), Trento, Italy, (aelalailyi; remondino)@fbk.eu
[4] Gexcel srl, via Branze 45, Brescia, Italy, (matteo.sgrenzaroli; giorgio.vassena)@gexcel.it

**Keywords:** Multi-camera, Photogrammetry, SLAM, Visual SLAM, Mobile Mapping, Architecture

**Abstract**

Ant3D is a fisheye multi-camera system for 3D mapping in confined environments, designed and refined over a seven-year period. This paper has two aims: to present the latest tests reflecting recent system developments and to summarize the design journey and milestones. We compare three successive surveys tests conducted in the same challenging test site: meandering masonry interiors and rock-cut passages within San Vigilio Castle (Bergamo, Italy). The 2018 test used a low-cost GoPro multi-camera rig; the 2020 test used Ant3D v1, a custom prototype improved based on the previous experiences; the 2025 test employed the first industrialized Atom-Ant3D, integrating SLAM-based pose estimation, pose-graph optimization, and SLAM-aided SfM to accelerate processing. The comparison considers reconstruction drift against a TLS ground-truth point cloud and processing times. Results show Ant3D v1 achieves high accuracy (max deviation ~3cm) but required extensive processing times (~20h), Atom-Ant3D offers major time savings (~4.5h) with moderate accuracy trade-offs, and the GoPro rig performs worst. The discussion outlines design and processing guidelines for reliable, fast mapping of narrow spaces.

## 1. Introduction

Fisheye, multi-camera, and spherical-camera photogrammetry have been thoroughly investigated and explored in the past years. The strongest drive in the development and adoption of methodologies and novel instrumentation that exploits these techniques is the 3D reconstruction of confined environments with limited accessibility. Such conditions are not rare instances in both cultural heritage and archaeological digitization projects. Consequently, researchers and practitioners have started experimenting with fisheye lenses and action cameras, including them in the photogrammetric pipeline in search of more compact and agile instruments and wider field of views. The boom in the use and experimentation of compact fisheye cameras started with the implementation of the fisheye camera model into popular Structure from Motion (SfM) software approximately ten years ago. The focus of the research has been on fast mapping and mobile acquisition utilizing low-cost sensors. In recent years, the use of off-the-shelf action cameras, such as the GoPro, has declined in favor of spherical cameras and ad-hoc solutions, mainly multi-camera systems targeted at solving specific survey challenges. A parallel line of research is the development of real-time or quasi-real-time processing aimed at reaching the ambitious goal of obtaining the final 3D reconstruction directly on field at the end of the acquisition operation. A popular investigation topic is the integration of visual SLAM (simultaneous localization and mapping) solutions that, if not the full reconstruction yet, allow for estimating the device movement path in real-time, significantly speeding up the post-processing operations by providing approximate values for camera poses.
This study has two aims: (1) to present the latest achievements in the development of a 3D measuring system aimed at the reconstruction of confined spaces named Ant3D (Perfetti et al., 2024a); and (2) to present a synthesis of the 7 years long development journey that led to the current solution, presenting a comparison with previous approaches and versions, offering comments on the main milestones.

The opportunity for comparison is offered by the case study of choice, that is, the confined spaces of the San Vigilio Castle (Città Alta, Bergamo, Italy), which was the subject of a first test in 2018 using an arrangement of GoPro action cameras; a second test in 2020 using our own patented multi-camera system prototype; and a recent third test in 2025 utilizing the first industrialized prototype of the same system augmented with multi-camera visual SLAM processing (Figure 1).

Figure 1. Latest version of the Ant3D multi-camera system. The picture shows the first production prototype developed by Gexcel srl with the control backpack.

## 2. Background

Early experimentation with novel photogrammetric solutions – featuring fisheye lenses, compact hardware, and multi-camera rigs – focused on streamlining and accelerating survey operations. These initial efforts often leveraged readily available low-cost devices such as action cameras (Teo, 2015; Koehl et al., 2016): Holdener et al. (2017) investigated multi-camera configurations consisting of five GitUp cameras and conceptualizing a possible image-based mobile mapping system.

In more recent developments, researchers have introduced custom-built devices incorporating specialized hardware components, including global shutter sensors and precise synchronization: Torresani et al. (2021) presented Gupho, a general-purpose modular stereo system for guided photogrammetric mapping featuring stereo visual SLAM for trajectory estimation. Ortiz-Coder and Sánchez-Ríos (2019) also proposed a general-purpose system similar to Gupho, leveraging two cameras. In their solution, one camera is dedicated to visual SLAM tracking (monocular visual SLAM), while the second camera captures images for the photogrammetric process.
Most recently – testifying for the field interest in hand-held image-based survey platforms and ad-hoc guided photogrammetric solutions – novel products are being commercialized, such as: the Stonex XVS Scanner (Stonex srl Development Team), device based on the Ortiz-Coder and Sánchez-Ríos's system and the Looq AI qCam (Looq AI Development Team), a hand-held system comprising 4 cameras, mainly marketed as the imaging component of the Looq platform, but that can be applied to challenging 3D survey scenarios (Clark et al., 2025; Klingspon et al., 2025).
Our system was instead developed specifically to be optimized for narrow space mapping (Perfetti, 2020; Perfetti et al., 2024a), and in the initial iterations, visual SLAM capabilities were not included. Elalailyi et al. (2024) later developed a module implementing Multi-Instance visual SLAM to be included in the current version.

Figure 2. Examples of recent multi-camera portable visual SLAM systems from research: Gupho (top left, Torresani et al., 2021); Looq qCam (top right, Looq AI); system from Ortiz-Coder and Sánchez-Ríos, 2019 (bottom left); Stonex XVS vSLAM scanner (bottom right).

## 3. Case study

The San Vigilio Castle, specifically the confined and narrow passages within the historic Castagneta Tower (Figure 3), provide an ideal environment for testing the performance of systems aimed at efficiently reconstructing meandering narrow spaces. The tower comprises two interconnected environments, each presenting distinct challenging features.
The first environment is an underground tunnel approximately 80 meters in length, excavated in natural rock. It features a narrow, linear, and meandering geometry with a width of approximately 0.70 to 0.80 metres and a ceiling height averaging around 1.5 metres. The tunnel is characterized by a muddy floor, high humidity, and no lighting, necessitating the use of artificial illumination during data acquisition. The second environment is the interior of the tower, which includes two circular rooms connected by a narrow internal staircase.
These spaces are characterized by refined masonry surfaces, planar walls, sharp edges, and relatively uniform textures composed of squared stone blocks.
Together, these environments offer a diverse range of surface types – from rough, irregular rock to smooth, structured masonry – and spatial configurations that include tight corridors, vertical connections, and curved geometries. The combination of poor accessibility, confined space, and complex architectural features made this site an ideal testbed during the development of the Ant3D multi-camera system.
The case study was first chosen for the FINE benchmark (Fisheye Indoor Narrow spaces Evaluation, Perfetti et al. (2024a)), a study aimed at evaluating strategies for achieving the 3D reconstruction of tunnels employing low-cost cameras, that is, a multi-camera system comprising GoPro cameras. The FINE benchmark provided raw data of multiple acquisitions performed in the Castle environment, including the video footage from the GoPro camera rig and a ground truth data derived from a highly redundant laser scanning survey (Figure 4). This experience highlighted the limitations of the low-cost action camera approach and hinted for improvements. A subsequent acquisition was conducted in 2020, repeating the survey of the area with the first functional prototype of Ant3D. The results as well as the design of the first version of the system are described in Perfetti et al. (2024). More recently, following further improvements to the systems regarding primarily the illumination setup, capturing frame rate, data management, and implementation of multi-stereo visual SLAM (Elalailyi et al., 2024; El-Alailyi et al., 2025), a new acquisition was carried out in 2025 employing the first industrialized prototype of the system developed by Gexcel srl.

Figure 3. Images of the Castagneta Tower's confined spaces.

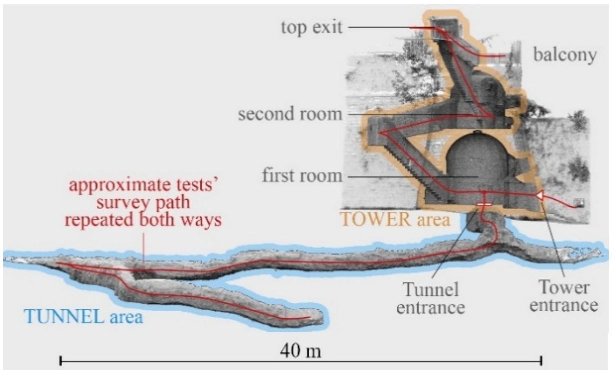


Figure 4. Elevation view of the ground truth: laser scanning point cloud, with annotations.

## 4. Materials and Methods

In this study, we present a comparison of three subsequent surveys conducted in the Castagneta tower, revisiting all datasets to compare both the accuracy of the 3D reconstruction and the required processing time.

### 4.1 Multi-camera solutions

The three multi-camera solutions are composed as follows:

- **GoPro rig** (Figure 5, top),
  The hardware was composed of five GoPro Hero3 cameras with a resolution of 1920 x 1080 pixels and one GoPro Hero4 camera with a resolution of 3840 x 2160 pixels. The cameras were mounted on an aluminium frame, forming two horizontal camera pairs and one vertical pair. The multi-camera rig was mostly forward oriented with the cameras only slightly tilted towards the inside and outside. Image acquisition was controlled independently for each camera by initializing a full resolution video recording from which images were later extracted at 1 fps. Sub-optimal multi-camera synchronization was achieved by synchronizing the video recordings based on the audio tracks. Within initial testing, three main limitations of the system were observed: (i) lack of accurate frame synchronization, (ii) rolling shutter sensors, and (iii) rig geometry not ideal for confined environments.
- **Ant3D v1** (Figure 5, centre),
  The hardware was composed of five FLIR USB3 cameras with a resolution of 2448 x 2048 pixels, equipped with fisheye lenses. Addressing the main limitation encountered with the GoPro rig, Ant3D features a rig geometry designed to better suit confined environments, global shutter cameras, and hardware frame synchronization. Images were acquired at 1 fps.
- **Atom-Ant3D** (Figure 5, bottom),
  First industrialized prototype of the system, designed with hardware similar to the previous version, reducing weight and improving handling, robustness and usability. Regarding image acquisition, options are available to acquire the images at a specified frame rates and to initialize Multi-Instance visual SLAM on-field. For the purpose of this test, images were saved at 3 fps.

### 4.2 Acquisitions and dataset preparation

The three acquisitions covered both the tower and the tunnel areas (Figure 4). In the 2018 survey the two environments were acquired separately but with overlapping rooms, while in the 2020 and 2025 surveys, both environments were covered by one unified acquisition performed with loop closure. By combining all images from each survey, three image datasets were obtained:

1. **D1**, GoPro rig, 1637 x 6 images
2. **D2**, Ant3D v1, 2565 x 5 images
3. **D3**, Atom-Ant3D, 1001 x 5 images

For each dataset, the initial self-calibration of the cameras' intrinsics was carried out making use of associated image sets of specifically prepared calibration environments. For the 2018 and 2020 datasets, the calibration environments were set up on-site in the Castagneta tower, while for the 2025 survey, a higher accuracy calibration was achieved using a laboratory test field.
For all datasets, images pre-processing was performed to balance the scene high contrast due to the dark environments being illuminated solely by the multi-camera LED illuminators. The images pre-processing was carried out in Adobe Lightroom by adjusting highlight and shadows levels. Figure 6 shows a comparison between original and processed images from D3.

Figure 5. Milestones in the development of Ant3D. Multi-camera GoPro rig from 2018 FINE Benchmark (top), Ant3D v1 (centre), and first industrialized prototype Atom-Ant3D (bottom).

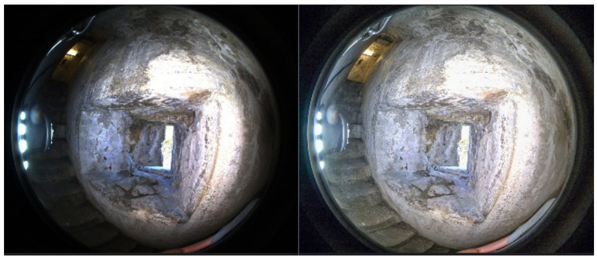

Figure 6. Example of the images pre-processing applied to D3. Original (left column) vs processed (right column) images.

Figure 7. Images of the survey activities from the three tests.

### 4.3 Data processing

Two distinct processing pipelines were implemented:

- Datasets D1 and D2 were processed using a standard photogrammetric SfM pipeline with Relative Orientation (RO) constraints.
- Dataset D3 was processed by leveraging a visual SLAM approach to estimate initial image poses. Subsequently, the initial poses were refined through a visual SLAM aided SfM.

Figure 8 summarizes the two processing pipelines.

#### 4.3.1 D1 and D2 processing

Datasets D1 and D2, previously compared in Perfetti et al. (2024a), were completely reprocessed using the approach that proved most successful in the comparative analysis conducted in Perfetti et al. (2024b), which includes multi-camera RO constraints and initial values for camera intrinsics that are then refined during the Bundle Adjustment (BA). For dataset D1, acquired with the GoPro rig, the rolling shutter compensation was also estimated. SfM processing was conducted in Agisoft Metashape (latest version available as of Dec 2025, v2.2.2).
After SfM, if required, manual macro adjustments were performed to fix images misalignment. This was necessary only for D1 where the image orientation of small groups of images had to be recomputed to achieve a satisfactory result.
The last step of the pipeline was refinement, which was performed through tie points filtering and additional BA.

#### 4.3.2 D3 processing pipeline

Building on the pose-graph and feature-based optimization framework for multi-camera mobile mapping systems (El-Alailyi et al., 2025) and the visual SLAM-aided photogrammetric processing of fisheye multi-camera sequences (Elalailyi et al., 2024), initial image poses were estimated trough the following steps:

1. **Multi-Instance visual SLAM**
   This step estimates multiple trajectories from stereo pairs and a monocular stream of the multi-camera system. In this study, visual SLAM was operated using stereo mode for the side camera pairs and monocular mode for the central camera. Each visual SLAM instance produces estimated camera poses (i.e., trajectories) and their corresponding sparse point clouds. Each initial output trajectory has its own drift profile and outliers.
2. **Monocular stereo scale estimation**
   This step involves bringing the previously estimated monocular trajectory to metric scale using the simultaneously estimated stereo SLAM trajectories. The scale factor is computed from synchronized time motion consistency between monocular and stereo tracks over overlapping time intervals and robustly aggregated across the available stereo pairs to yield a single global scale. This scale is then applied to the monocular translations while keeping the orientations unchanged.
3. **Multi-camera Pose Graph Optimization (PGO)**
   This step estimates a single optimized and globally consistent trajectory from the previously estimated trajectories (Figure 9) by exploiting the redundancy in multi-camera observations and independently estimated motion. The procedure begins with the initialization of a nonlinear factor graph, where each node represents the full pose of a camera and each edge encodes a relative-pose constraint between consecutive frames. Robustness to outliers is increased by adopting an uncertainty-aware noise model with a robust loss function to down-weight inconsistent constraints. The resulting nonlinear problem is solved using a Gauss-Newton optimizer, which iteratively minimizes the residual errors to obtain a globally consistent trajectory (El-Alailyi et al., 2025).

The resulting initial estimate of the image poses is then refined using the visual SLAM-aided photogrammetric adjustment in order to obtain the final 3D reconstruction.
As for D1 and D2 processing, the SfM adjustment was performed using Agisoft Metashape (v2.2.2). The spatial priors provided by the optimised PGO trajectory were used to (i) identify candidate loop closures and (ii) restrict feature matching to cameras that were spatially proximate. The visual SLAM-informed feature matching significantly accelerates the regular photogrammetric processing by reducing unnecessary exhaustive pairwise comparisons. Subsequently, the images orientation was performed using the optimized PGO trajectory as external orientation priors to accelerate the photogrammetric computation. As for D1 and D2 processing, the last step was the refinement through tie points filtering and final BA.

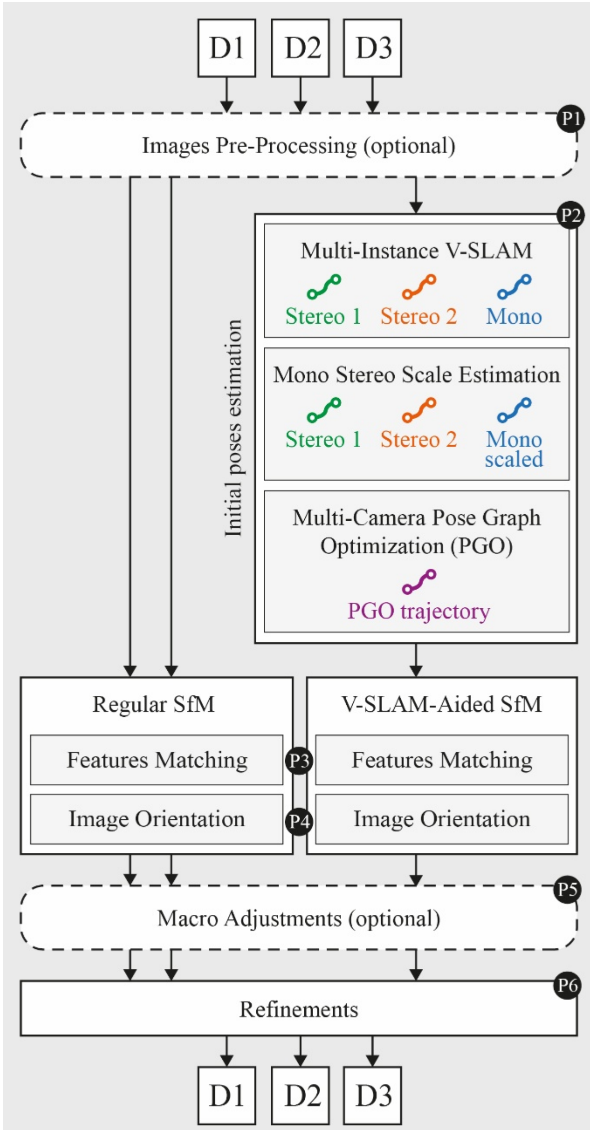


Figure 8. Scheme of processing phases. D1 and D2 were processed with a regular SfM pipeline plus manual adjustment (D1 only) and refinements. D3 was processed in two steps: (i) initial image poses estimation and (ii) visual SLAM aided accelerated SfM, plus refinements. For each processing step (P1-P6), Table 1 provides provides the processing time.


Figure 9. PGO optimization of the visual SLAM trajectories implemented in D3. Tower area (top) and tunnel area (bottom).

### 4.4 Evaluation

The evaluation of the three datasets and related approaches is based on both the accuracy of the 3D reconstruction and on the comparison of the processing time required to achieve such results. The accuracy of the 3D reconstruction is evaluated against the original laser scanning ground truth from the FINE benchmark (Figure 4) in two ways: (i) by quantitatively measuring the deviation between the sparse tie points clouds and ground truth; and (ii) by qualitatively comparing cross-sections.

#### 4.4.1 Accuracy evaluation – M3C2 distance

The deviations from the ground truth are measured as the M3C2 signed distance (Lague et al., 2013) computed along the normals of the ground truth. Due to the time span between the three acquisitions, no fixed points could be reliably identified. Therefore, to determine a common coordinate system, an Iterative Closest Point (ICP) registration of the resulting point clouds was performed with the ground truth. The authors would like to note that a scale adjustment was also computed during the ICP registration. This was done to adjust for scale-error effects due to sub-optimal RO calibration for D1 and D2. Because the comparison was performed using the sparse tie points clouds, deviations within 2 cm were considered negligible.

#### 4.4.2 Accuracy evaluation – cross-sections

Multiple cross-sections have been traced along the tower and tunnel environments comparing the laser scanning ground truth with the three photogrammetric results. Visual comparison of cross-sections allows for the evaluation of the accuracy of the results, precision, density, and noise.

#### 4.4.3 Processing times

For each of the main processing steps (see Figure 8), the processing times are recorded, comparing D2 and D3 specifically to evaluate the benefit of the visual SLAM implementation and related processes developed to accelerate and simplify SfM. All tests ran on a single workstation: CPU Intel Core i9-8950HK @2.90GHz, RAM 32GB, GPU NVIDIA Quadro P3200.

## 5. Results and Discussion

The comparative results for accuracy and efficiency across the three datasets (D1 - GoPro rig; D2 - Ant3D v1; D3 - Atom-Ant3D with Multi-Instance visual SLAM) are synthesized in Figure 10 (M3C2 distance maps and statistics), Figure 11 (cross-sections), and Table 1 (processing times per step P1-P6 as in Figure 8).

Overall, D2 achieved the highest geometric fidelity, D1 the lowest, and D3 ranked intermediate. After ICP registration to the ground truth and subsampling of the tie-point clouds to a uniform 5 cm resolution, the maximum deviations (excluding outliers/noisy areas) were approximately: 35 cm for D1, 3 cm for D2, and 11 cm for D3. The percentage of points falling within the ±2 cm / ±5 cm ranges were respectively: 30 % / 50 % for D1, 83 % / 91 % for D2, and 42 % / 72 % for D3.

Errors in D1 concentrate strongly in the tunnel (local maxima >30 cm), while the tower performs comparatively better (local maxima ~10 cm), reflecting the limitations of the action camera rig's geometric configuration applied to confined, meandering passages. D2 exhibits consistently small residuals across both environments. D3 shows low deviations in the tower; however, localized deviations increase in the tunnel, owing to the reduced image redundancy in that setting.

The cross-sections confirm the M3C2 analysis. D1 reveals clear mis-orientations and shape distortions in the tunnel, with improved, though still imperfect, adherence in the tower area. D2 closely matches the ground truth, showing good adherence and low noise levels. D3 generally mirrors D2 in profile sharpness and adherence, except for specific tunnel segments where deviations enlarge and density drops, consistent with the reduced number of retained images.

Regarding processing times, D1 required ~26 h in total, including 13 h for feature matching, 11 h for image orientation, and ~1 h of manual macro adjustment, reflecting the fragility of the rig configuration. D2 converged without manual intervention but still required ~21 h 30 min, dominated by the feature matching process (~14 h) despite general image-pairs pre-selection strategies that were used. By contrast, D3 completed in ~4 h 30 min, fully unattended, with 40 min required for the visual SLAM and PGO process, ~40-60 min for image orientation, and noticeably reduced feature matching times (~ 20 min per sub-block). Moreover, the ~1 h image pre-processing could be substituted via live histogram balancing during acquisition, and when Multi-Instance visual SLAM runs live, the total processing time can drop around 3 h.

The outcomes of the GoPro rig, showing the weakest performance and longest processing time, confirm the previous findings, highlighting issues stemming from un-ideal design, that is: (i) imperfect multi-camera synchronization, (ii) rolling-shutter sensors, and (iii) a rig geometry sub-optimal for narrow environments.

These factors led to unstable orientation necessitating manual macro fixes and amplified deviations, particularly in the tunnel environment. While a low-cost option, solutions such as the GoPro rig are advisable only for relatively simple indoor environments, and with Ground Control Points (GCPs) constraints, such solutions can deliver at best 1:100-1:200 mapping scales, yet nowadays they are difficult to justify against other more practical low-cost alternatives.

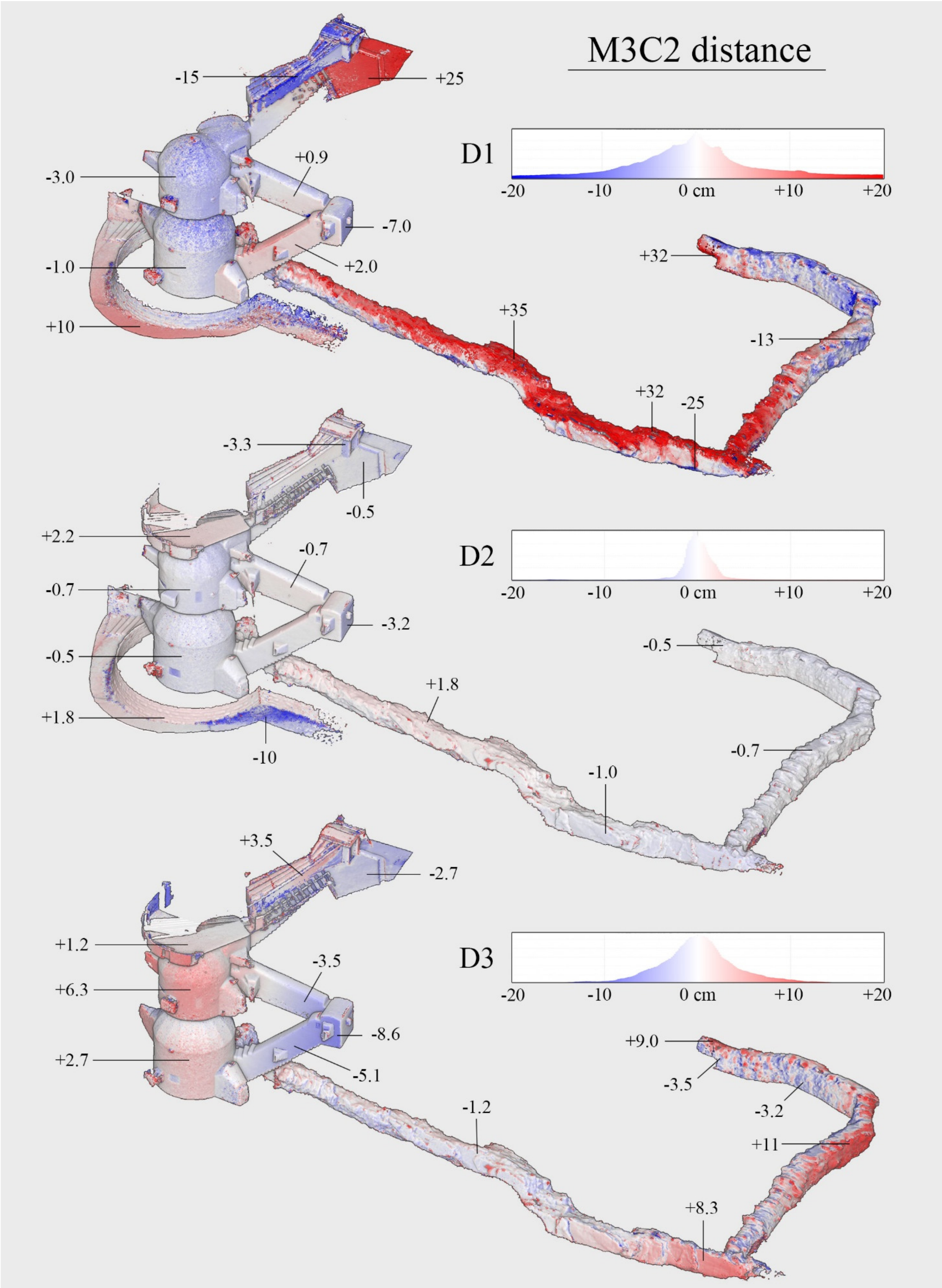


Figure 10. Evaluation of the 3D reconstructions' deviation. Comparison of M3C2 signed distances to the ground truth with plotted distributions, deviations ≤ 2 cm are considered negligible.

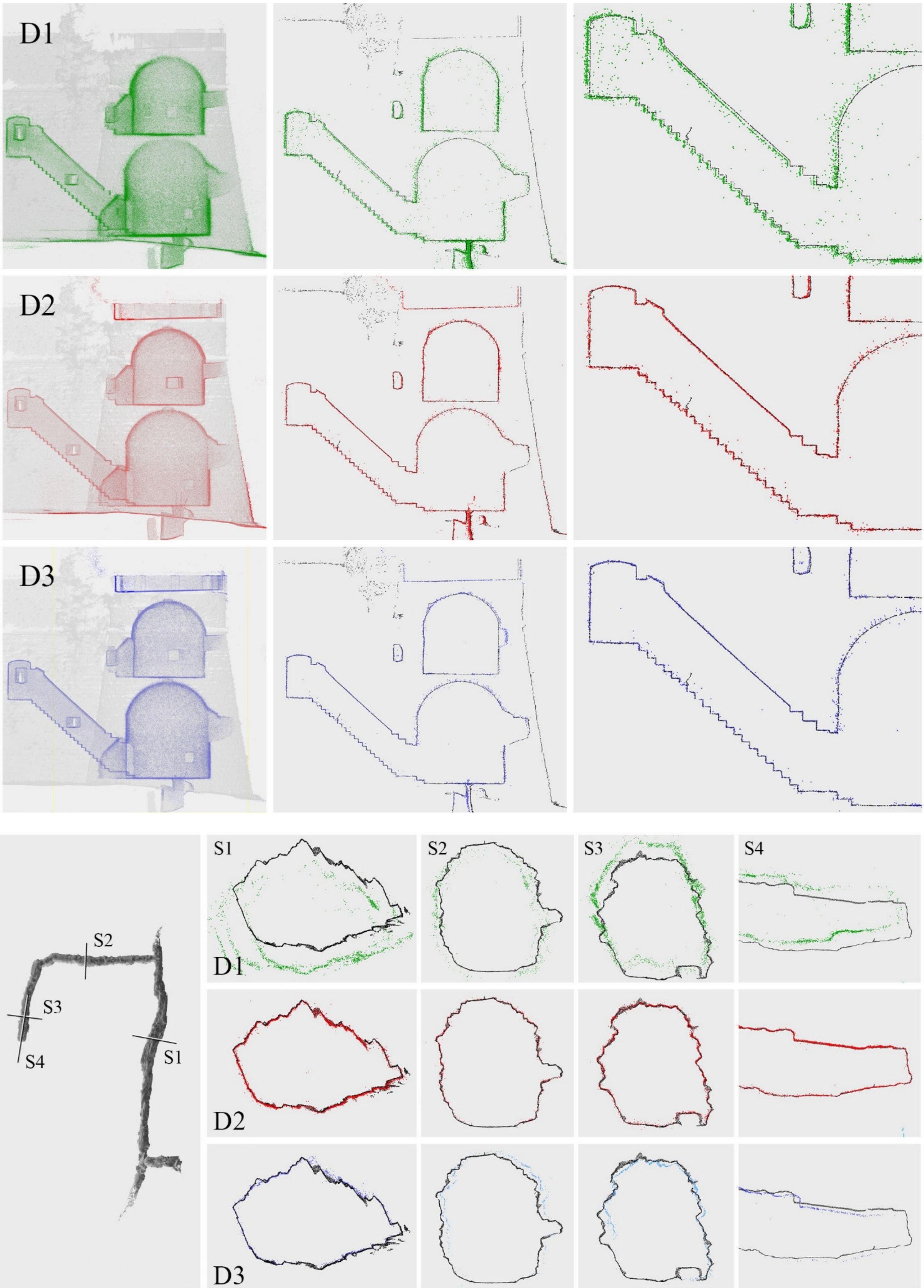


Figure 11. Evaluation of the 3D reconstructions' deviation. Cross-sections of D1 (green), D2 (red) and D3 (blue) overlaid with the ground truth (black) for both the tower (top) and tunnel (bottom) areas.

| | # of poses | Survey duration | Img Pre-Processing | V-SLAM + PGO | Features' Matching | Image Orientation | Macro Adjustments | Refinements + Filtering | Tot. processing |
|---|---|---|---|---|---|---|---|---|---|
| | | | P1 | P2 | P3 | P4 | P5 | P6 | P1-P6 |
| D1 | 1637 | 27 min | ~ 01 h | - | 13 h | 11 h | ~ 01 h | ~ 01 h | 26 h |
| D2 | 2565 | 40 min | ~ 01 h | - | 14 h | 06 h | - | ~ 30 min | 21 h 30 min |
| D3 | 608 | 25 min | ~ 01 h | 40 min | 20 min | 01 h | - | ~ 30 min | 04 h 30 min |
| | 393 | | | | 20 min | 40 min | - | | |

Table 1. Per-step processing times and totals for the three datasets (D1-D3), including poses counts and survey durations.

The results of Ant3D v1 (D2), designed for confined-space photogrammetry (i.e., global shutter, synchronized frames, specifically designed rig), show high accuracy across both environments and indicate suitability for 1:100 to 1:50 surveys even without GCPs. This approach supports robust and reliable 1:50 surveys even in complex scenes when GCPs are introduced. The principal limitation is computational: near-exhaustive pairwise feature matching remains the dominant cost, despite successful automatic convergence of the reconstruction pipeline. The results of Atom-Ant3D with Multi-Instance visual SLAM (D3), although based on a multi-camera configuration matching D2, show slightly lower overall accuracy. These results can be motivated by the reduced image redundancy of D3 with respect to D2. From ~4250 x 5 images saved at 3 fps, the visual SLAM frame selection retained roughly one quarter (1001 x 5 images). While visual SLAM frame selection is generally beneficial for accelerating photogrammetric processing by preserving significant frames and discarding redundant ones, under the present highly challenging tunnel conditions, it appears to have reduced robustness in specific trajectory segments. This effect has been further enhanced by the fact that the real-time Multi-Instance visual SLAM run live on field had to be discarded and re-processed offline, since the abrupt motion that characterised the acquisition inside the tunnel (crawling) caused real-time failures. Offline post-processing succeeded for the forward trajectory but failed midway through the return trajectory, further reducing the number of usable selected frames in the tunnel. Despite these limitations, pronounced deviations occur mainly near the tunnel end; across most sections, D3 exhibits deviation profiles comparable to D2.

## 6. Conclusion

Purpose-built, multi-camera rigs (D2/D3) decisively outperform action camera setups (D1). The test conducted showed how visual SLAM aided processing (D3) offers significant speed gains over conventional SfM processing, while high geometric accuracy (suitable for 1:50 surveys) can be achieved, as demonstrated by D2. In highly challenging scenarios, consistent data coverage proved to be critical. The results indicate how over-pruning of frames during the visual SLAM processing and keyframe selection can reduce robustness and increase residuals. An adaptive frame-retention informed by motion patterns and scene complexity, would likely narrow D3's residual gap to D2 while preserving most of the computational speed advantages.

## Acknowledgements

The Atom-Ant3D projects were funded to Francesco Fassi (Politecnico di Milano) through the Boostech Valorization Program, via Proof of Concept (PoC) funding under Italy's National Recovery and Resilience Plan (PNRR), Mission 1, supporting the industrial property system and financed by the European Union – NextGenerationEU.

## References

Clark, L., Rissolo, D., McAvoy, S., Fortin, J., Nava Blank, A., Barba-Meinecke, H., and Kuester, F.: Integrated Multi-Camera Photogrammetry, 2025: Toward the simulation of human and faunal access to the pleistocene cave systems of Quintana Roo, Mexico, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-M-9-2025, 329–335.

Elalailyi, A., Perfetti, L., Fassi, F., Remondino, F., 2024: Visual SLAM-aided photogrammetry to process fisheye multi-camera systems sequences. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W4-2024, 189–195.

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.

Holdener, D., Nebiker, S., Blaser, S., 2017: Design and implementation of a novel portable 360° stereo camera system with low-cost action cameras. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2-W8, 105–110.

Klingspon, J., McAvoy, S., Chiabrando, F., Tanduo, B., Martino, A., Rissolo, D., and Kuester, F., 2025: A multi-modal evaluation of GPS assisted rapid multi-cam photogrammetry, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-M-9-2025, 741–747.

Koehl, M., Delacourt, T., and Boutry, C., 2016: Image capture with synchronized multiple-cameras for extraction of accurate geometries, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B1, 653–660.

Lague, D., Brodu, N., Leroux, J., 2013: Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 10-26.

Looq AI Development Team: Looq qCam, https://www.looq.ai, last accessed January 2026.

Ortiz-Coder, P., Sánchez-Ríos, A., 2019: A self-assembly portable mobile mapping system for archeological reconstruction based on VSLAM-photogrammetric algorithm. *Sensors*, 19(18), 3952.

Perfetti, L., 2020: Multi-camera rig for 3D reconstruction: concept, design and accuracy evaluation. in: Luhmann, T., Schumacher, C., Photogrammetrie - Laserscanning – Optische 3DMesstechnik, Beiträge der Oldenburger 3D-Tage, 124-131.

Perfetti, L., Fassi, F., Vassena, G., 2024a: Ant3D—a fisheye multi-camera system to survey narrow spaces. *Sensors*, 24(13), 4177.

Perfetti, L., Bruno, N., Roncella, R., 2024b: Multi-camera rig and spherical camera assessment for indoor surveys in complex spaces. *Remote Sensing*, 16(23), 4505.

Stonex srl Development Team: XVS vSLAM 3D scanner, https://www.stonex.it/product/xvs-vslam-scanner, last accessed January 2026.

Teo, T., 2015: Video-based point cloud generation using multiple action cameras, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XL-4/W5, 55–60.

Torresani, A., Menna, F., Battisti, R., Remondino, F., 2021: A visual SLAM guided and portable system for photogrammetric applications. *Remote Sensing*, 13(12), 2351.