Pose Graph Data Fusion for Visual- and LiDAR-based Low-Cost Portable Mapping Systems

Ahmad Elalailyi [1,2], Paweł Trybała [2], Luca Morelli [2,3], Francesco Fassi [1], Fabio Remondino [2], Luigi Fregonese [4]

[1] Politecnico di Milano, Dept. of Architecture, Built environment and Construction engineering (ABC), Italy
Email: ahmad.elalailyi@polimi.it, francesco.fassi@polimi.it

[2] 3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation (FBK), Trento, Italy
Email: aelalailyi@fbk.eu, lmorelli@fbk.eu, ptrybala@fbk.eu, remondino@fbk.eu

[3] Dept. of Civil, Environmental and Mechanical Engineering (DICAM), University of Trento, Italy

[4] Politecnico di Milano, Dept. of Architecture, Built environment and Construction engineering (DABC), Campus Mantova, Italy
Email: luigi.fregonese@polimi.it

Keywords: Multi-camera systems, LiDAR SLAM, Visual SLAM, 3D Point Clouds, Data Fusion, Accuracy, Pose Graph.

Abstract
Traditional 3D surveying methods often fall short in complex spaces due to lack of mobility, time constraints and high risk. For this reason there is an actual demand for 3D data acquisition tools and methods, particularly suitable for complex and narrow environments, due to their capacity for efficiently capturing detailed and accurate spatial information, maybe also automatically. This study presents a novel approach for fusing 3D spatial data collected by two separate and independent mobile mapping systems: (1) ATOM-ANT3D and (2) MandEye. We propose an innovative fusion technique that combines visual and LiDAR data from asynchronous acquisitions, reducing the need for strict temporal and spatial synchronizations between the two systems. We compare the outputs of both systems before and after fusion, studying the individual limitations and highlighting the complementary benefits achieved by the proposed fusion framework. Results demonstrate improved accuracy of global alignment and spatial completeness of the final point clouds, proving the efficiency and flexibility of the proposed approach.

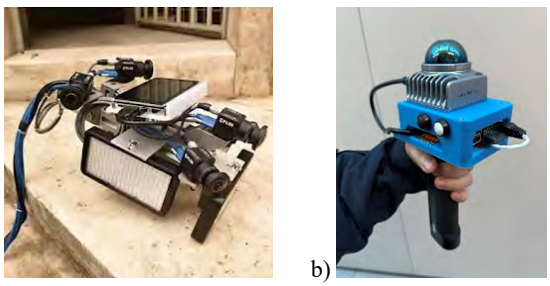
1. INTRODUCTION

With the increasing demand for capturing 3D spatial data, there is a growing need for cost-effective and efficient methods for data acquisition. Traditional methods, such as photogrammetry and static Terrestrial Laser Scanning (TLS), provided highly accurate 3D spatial data for many years in heritage and architectural scenarios (El-Hakim et al., 2007; El-Hakim et al., 2008; Fassi et al., 2011; Remondino, 2011). Recently, Mobile Mapping Systems (MMS), either with LiDAR sensors, cameras (visual) or an integration of both sensors, offer a promising alternative combining affordability, flexibility and reduced acquisition times, making them versatile and suitable for a wide range of applications, even in complex scenarios (Nocerino et al., 2017; Di Stefano et al., 2021; Elhashash et al., 2022). Coupled with real-time data processing capabilities (generally called SLAM – Simultaneous Localization and Mapping), data collection becomes more effective and post-processing requirements are generally reduced, therefore speeding up surveying processes and enhancing the quality of generated 3D data (Torresani et al., 2021; Będkowski, 2024; Perfetti et al., 2024).
While existing literature have often extensively compared the characteristics of LiDAR- and visual-based point clouds (Kadobayashi et al., 2004; Bayram et al., 2015; Dietmaier et al., 2019; Trybała et al., 2023a), an advantageous and underexplored opportunity lies in the fusion of these data modalities at raw data level. By integrating LiDAR and photogrammetric data, the resulting hybrid approach can potentially benefit from the best qualities of both methods. However, since such systems usually rely on the relative positioning methods, not using the global reference, they are prone to long-term drift of the trajectory. When registering the 3D data coming from independent SLAM-based acquisitions, this issue often prevents achieving an accurate global alignment of two-point clouds, even despite their locally high quality and compliance.

1.1 Paper's Aim

This study focuses on the fusion and quality evaluation of 3D spatial data generated by two portable and hand-held Mobile Mapping Systems (MMS) based on different sensing modalities:

- ATOM-ANT3D (Elalailyi et al. 2024), the new version of the ANT3D (Perfetti et al, 2024) multi-camera mobile mapping system based on multi-view (5) fish-eye cameras (Figure 1a);
- MandEye system (Figure 1b), a LiDAR-based portable SLAM system (Będkowski, 2024; Trybała et al., 2023b).



a) b)

Figure 1. The two portable handheld surveying systems: ATOM-ANT3D (a) and MandEye (b).

With the goal of employing data fusion techniques to portable mobile mapping systems, the paper aims to identify and mitigate sensor weaknesses. Therefore, we pursue a more reliable hybrid approach that can improve the overall 3D data quality in terms of reciprocal compliance of the two datasets, global accuracy and completeness of the final 3D reconstruction. No ground control points (GCPs) are used for constraining or optimizing any solution. The 3D reconstruction's scale relied solely on the known rigid baselines between ATOM-ANT3D cameras and the metric range measurements of LiDAR in MandEye. The assessment is carried out through comparisons with a high-quality ground truth TLS point cloud.

2. RELATED WORKS

Sensor integration and data fusion has long been studied, especially in SLAM and mobile mapping systems (MMS) to enhance the accuracy of spatial data collection. LiDAR multi-sensor fusion have witnessed developments on both data and approaches levels. For example, data coupling methods occur both on the front-end (raw data fusion) and back-end (error coupling and optimization) (Xu et al., 2022). Multi-sensor integration has become essential for reliable state estimation in SLAM applications. By leveraging the complementary strengths of different sensors, these frameworks improve the robustness and accuracy of SLAM systems, offering enhanced performance across a variety of conditions and addressing the challenges of drift, scale and environmental variability effectively. Zhu et al. (2024) demonstrated the different combinations of sensors fusion between visual, LiDAR and inertial measurement unit (IMU), analysing each approach and the methods used for fusing the data in each category. Zhao et al. (2023) demonstrated point clouds registration of image-based and LiDAR techniques through Robust Point Matching using Learned Features (RPM-Net) (Yew & Lee, 2020) for improving the 3D reconstruction. Cao et al. (2018) used thermal consistency in a multi-modal approach to fuse information from depth and thermal images to improve the 3D thermal reconstruction. Wei et al. (2024) fused point clouds generated by LiDAR-inertial SLAM with RGB and thermal cameras through post-processing to produce enriched 3D maps. Kumar et al. (2020) integrates data between aligned LiDAR and camera sensors at a low-level to improve vehicle to object depth estimations for navigation and obstacle avoidance.

Although such methods can leverage the advantages of different sensors, using them jointly to improve the positioning quality, they rely on acquiring the data simultaneously, providing time synchronization and pre-calibrated fixed sensor rigs. Such constrains make them less flexible compared to more common scenarios, where the different devices acquire data independently and not exactly at the same time, additionally following slightly varying trajectories. A data fusion method free of these constrains, despite being more challenging, could potentially allow to complementarily use even different MMSs or achieve better alignment of the spatial data acquired periodically, e.g., for monitoring purposes.

Sammartano and Spanò (2018) investigated the repeatability of 3D surveys performed with the GeoSLAM Zeb LiDAR-based SLAM device. They demonstrated the variability between the data acquisitions performed in two directions and evaluated the quality of the internally optimized trajectories. Moreover, the work also tackled the problem of fusing such data with point clouds of other source, e.g., TLS or drone-based surveys. In the case of rigidly registering such point clouds in larger area, the authors encountered a visible influence of the drift, degrading the alignment of the SLAM-based 3D reconstruction to other point clouds. Despite they show that the issues can be addressed at the local level, no solution was proposed for obtaining a correct global alignment of the full datasets. Medici et al. (2024) explored possibilities of fusing TLS and photogrammetry at different levels, i.e., raw data, intermediate processing steps or results. Although the proposed methods are relevant to the case studies in this work, the aforementioned work used high-quality data, from a comprehensive survey, which greatly reduced the non-rigid discrepancies between LiDAR and vision-based 3D reconstructions, in contrast to drift-prone low-cost SLAM systems.

3. PROPOSED FUSION APPROACH

The idea here is to introduce a novel approach for fusing data generated by the two different MMSs (LiDAR- and image-based), mitigating the requirement of temporal and spatial synchronization of both systems during the acquisition. The method adds flexibility as it allows a more practical and adaptable approach for real-world applications, without requiring a fixed rigid setup between different systems/sensors. The overview of the proposed method is presented in Figure 2.

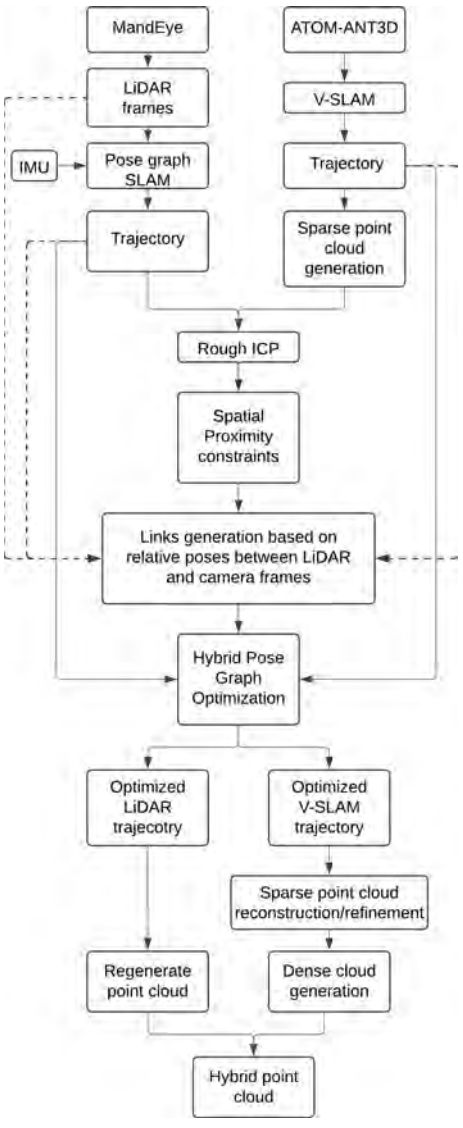


Figure 2. Diagram outlining the scheme of the proposed framework.

Given the initial 3D sparse point cloud of the complete survey site from both systems, a co-registration of both is performed using iterative closest point (ICP). As we demonstrate later in Section 5.3, such a global alignment does not yield good quality results, as both point clouds have varying local densities, completeness, noise and are not evenly influenced by a natural drift of a SLAM method.

To solve this, we leverage spatial proximity-based subsampling of camera and LiDAR frames to establish links between them. First, the initial LiDAR trajectory is subsampled, taking 1 LiDAR frame (oriented according to the initial SLAM output) out of every 500 frames. For these poses, closest camera positions are

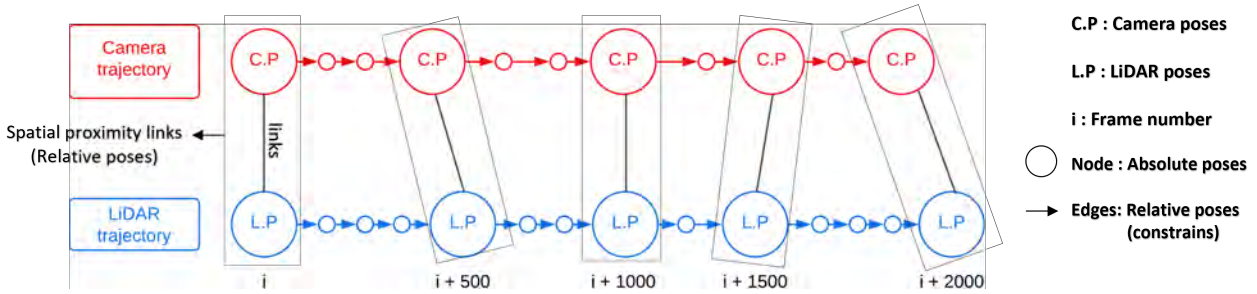


Figure 3. A graph demonstrating the spatial proximity links created between the two systems trajectories every 500th frame.

selected, based on the spatial proximity derived from the initially co-registered point clouds coming from the two systems. These frames are assigned as links between the two 3D reconstructions to generate constraints in the pose graph adjustment. Additionally, if a passage was traversed in two directions, i.e., the nearest camera poses to a selected LiDAR frame belonging to two distinct parts of the trajectory, two links are created to maintain the local consistency of the reconstruction. As this part of the methodology is use-case specific for a challenging setting of a narrow staircase and the use of a multi-camera system, it may need to be adapted for other applications / scenarios. The constraint between LiDAR and camera frames are generated as relative poses, calculated as a result of both the initial alignment of whole point clouds, and a fine-tuning through ICP matching of the LiDAR frame to a sparse photogrammetric reconstruction. This directly enforces the LiDAR point cloud to be locally consistent with the photogrammetric reconstruction, which is assumed to be created mostly by the cameras in the vicinity of the specific LiDAR pose.

Then, the initial trajectories in the form of relative poses are input into a Pose Graph Optimization (PGO), jointly with the aforementioned LiDAR-camera constrains (Dellaert et al., 2022). A simplified structure of the pose graph is presented in Figure 3. Additionally, the consistency of both SLAM-based reconstructions is ensured through constraining the relative poses between the start and ending points of the acquisition for both systems separately, using the original estimates of the SLAM solutions. Finally, the pose graph is optimized with the Gauss-Newton algorithm, producing adjusted trajectories of both the LiDAR and visual mapping systems, already in a common reference frame.

Since such pose graph adjustment lacks the information of the actual 3D data being mapped (i.e., 3D features, tie points), an additional step of local refinement of cameras is carried out. The pose graph optimized trajectory of the cameras is input into photogrammetric adjustment, using centimeter-level covariances for the optimized camera positions and higher covariances for their orientations for a further refinement and for the creation of the final dense cloud.

Unlike single-source mapping, which might have inconsistencies due to sensor specific limitations (e.g., LiDAR's SLAM sensitivity to sparse features or visual SLAM's sensitivity to lighting), the combination of the two systems can exploit each sensor's alignment strengths and compensate for weaknesses.

By linking spatial proximate frames, we fuse both 3D poses without the need for strict temporal synchronization and allow for a flexible correction of the sensors' trajectories, thus in the end essentially applying a non-rigid adjustment of the point clouds, which ensures their high coherence.

3.1 Evaluation approach

The overall alignment and 3D reconstruction quality of the original (separated) point clouds and the one produced with the proposed fusion approach are evaluated with respect to some reference data acquired with a Terrestrial Laser Scanning (TLS) survey (Leica C10). The assessment investigates accuracy and completeness of the dense clouds (before and after applying the fusion method). For accuracy (sometimes called also precision), we evaluate the ratio of points aligned to the ground truth data at different distance thresholds constructing the accuracy curve. On the other hand, the completeness (i.e., recall) curve is determined through thresholding the closest distances from the dense clouds to the ground truth data. This allows us to assess the compliance and coverage of the 3D reconstruction with respect to the ground truth as ratios of the total point count, satisfying the criteria at different metric distance thresholds (Knapitich et al., 2017; Nocerino et al., 2020). Additionally, the global alignment (and thus, global accuracy) is investigated in more detail, calculating signed distances from the evaluated point cloud to the reference data, using M3C2 method (Lague et al., 2013). The results are visualized in 3D, which properly highlights all problematic regions, in turn enabling identification of the causes of the lower accuracy (Trybała et al., 2023a).

Figure 4. TLS data of the staircase highlighting the interesting architecture of the study case (left) and some images taken during the acquisition (right).

4. CASE STUDY AND DATA ACQUISITION

The case study is the complex staircase of Sant'Andrea church in Mantua (Italy), which constitutes a intricated and narrow architectural environment for evaluating the capabilities of both

surveying systems in terms of spatial accuracy and coverage (Figure 4).
The surveyed area of the staircase structure has a width of ca 1 m and extends for ca 24 m length from ground till rooftop through 8 levels/floors. It is a double U-shaped staircase type where the separated up and down flights run mirror each other and interconnect between levels 4 and 5, enabling loop closure during the data acquisitions. For both acquisition systems, the survey started and finished at the entrance point with separate acquisition timings. The taken trajectory ensured forward and backward passes along the staircase, although certain areas were surveyed only once due to the complicated nature of the environment and the presence of obstructions (Figure 5). Both devices followed similar, but not identical trajectories.

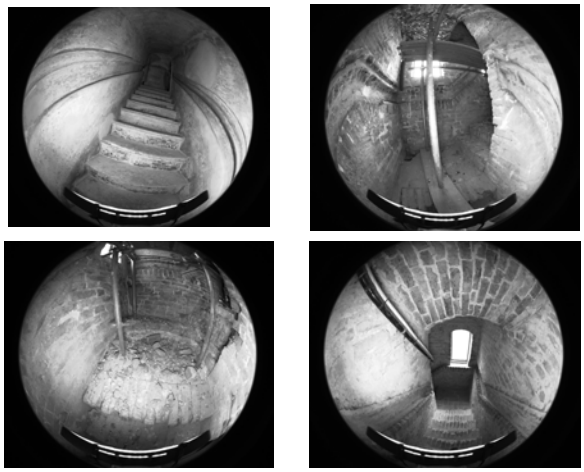


Figure 5. Examples of fisheye images acquired with the ATOM-ANT3D showing the narrow and complex site.

5. DATA PROCESSING

5.1 ATOM-ANT3D: V-SLAM + photogrammetry

The ATOM-ANT3D visual MMS captures data through five fisheye cameras, with multi-instance V-SLAM built on ORBSLAM3.0 (Campos et al., 2021), and produces (1) oriented images, (2) trajectory paths and (3) global map / sparse point cloud of the survey scene (Figure 6). The survey lasted ca 15 minutes starting and ending at the entrance point.
V-SLAM algorithms subsample image streams into keyframes based generally on the rate of change in the viewed environment. The front view, particularly for this narrow and complex study case, is exposed to larger depth with respect to the lateral cameras and therefore a smaller view rate of change while navigating the staircase compared to the side stereo pairs, resulting in lesser number of keyframes required.
The camera trajectories are refined using a pose graph optimization (PGO) approach available in GTSAM (Dellaert et al., 2022), leveraging both frame-to-frame and keyframe-based constraints to enhance spatial accuracy. Starting with initial pose estimates, a Nonlinear Factor Graph is constructed in which each frame is connected through relative transformations, while keyframes anchor the trajectory with higher-certainty constraints. The first frame is set as a fixed reference and subsequent frames are linked using Between Factors, capturing the relative transformations between consecutive frames. This pose graph structure minimizes cumulative error on unoptimized frames, producing a richer refined trajectory. After the optimization process a dense cloud can be generated thru MVS photogrammetry (Figure 7).

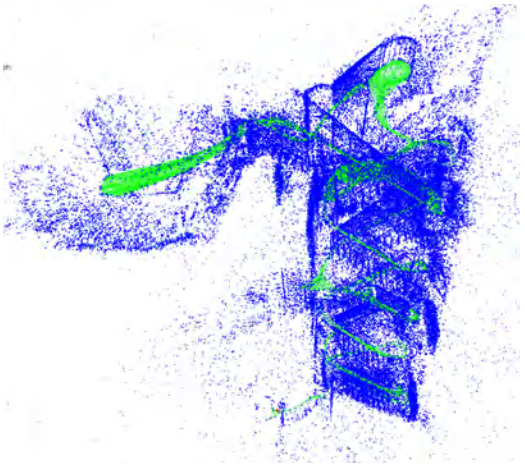


Figure 6. ATOM-ANT3D sparse point cloud (blue dots) and camera trajectories (green lines).

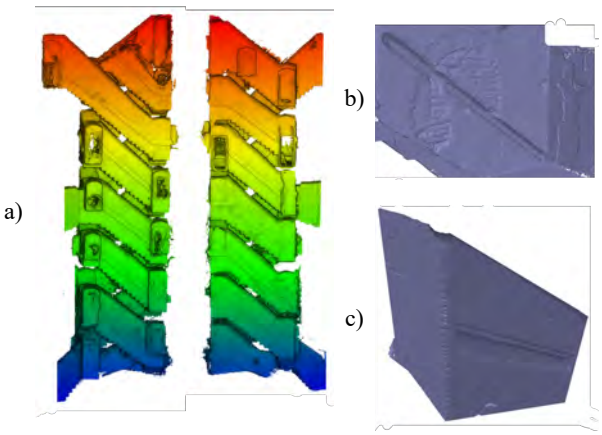


Figure 7. Point cloud of the staircase produced by the multi-camera system and V-SLAM aided photogrammetry (a) (coloured by relative elevation) and some views of dense clouds of inner parts of the staircase (b, c).

5.2 MandEye: LiDAR-inertial SLAM

The MandEye system embeds a Livox Mid-360 laser scanner and a low-cost PC to store the raw data (Będkowski, 2024). It was employed solely for the purpose of data acquisition and then a LiDAR-inertial SLAM (Trybała et al., 2023b) was used to process continuous streams of LiDAR and inertial data. This led to the sensor's poses estimation and the generation of a dense point cloud of the surveyed scene (Figure 8). Although SLAM was performed in an offline mode, we simulated a real-time mapping process through streaming the data from a rosbag file.
The employed SLAM method utilizes a pose graph approach to optimize the sensor's trajectory, given the relative poses from LiDAR-inertial odometry. The survey of the area of interest started and ended at the exact same point (a survey mark), forcing a loop closure to reduce the long-term drift. All other constrains in the graph are generated automatically by the system.

5.3 LiDAR-Visual data fusion

4131 images/camera were used for the fusion approach. Only poses of the master camera of the right stereo instance were utilized in the graph. For the other cameras, their poses were calculated later using a rigid transformation, obtained during a rigorous calibration performed before the measurements. On the other hand, a total of 8173 frames from the LiDAR-based system

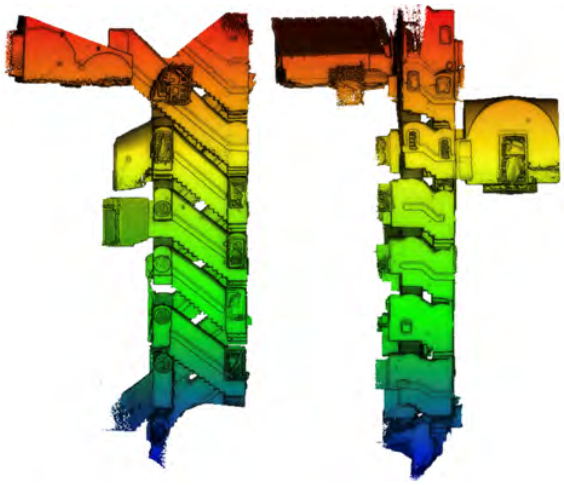
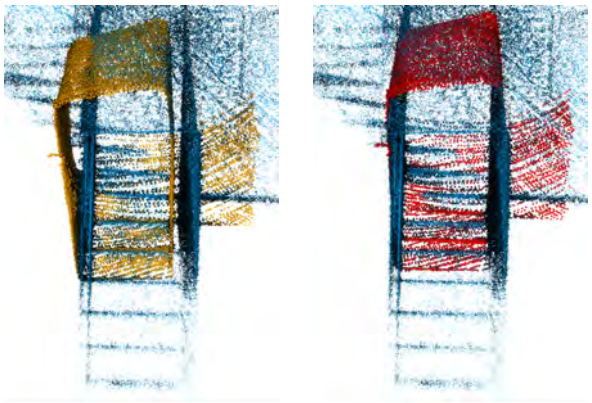


Figure 8. Point cloud of the staircase produced with the portable LiDAR-based system (coloured by relative elevation).

were used. Both trajectories were input into the pose graph optimization with a total of 20 links through local ICP (i.e. LiDAR - cameras relative pose constraints) established every 500th LiDAR frame. An example of creating such a link is shown in Figure 9. The solution is constrained at the entrance for both trajectories (relative poses separately generated by the two SLAM solutions) in the pose graph representing the loop closure at the starting and ending point. Figure 9 demonstrates the effect of the global ICP between the sparse V-SLAM point cloud and a LiDAR frame. The resulting optimized poses for the two trajectories were used to compute the final optimized dense clouds of the LiDAR and visual techniques.



a) b)

Figure 9. Correcting the point cloud alignment through ICP: a single LiDAR frame oriented according to global ICP between the point clouds (a; orange) and according to local ICP refinement (b; red) to the sparse V-SLAM point cloud (blue).

The fusion of the two-point clouds, apart from elevating the quality of the joint trajectory estimation, brings different benefits. Figure 10 presents part of the complementary role provided by the fusion for improving the completeness of the point clouds. In the LiDAR point cloud, one part of the staircase was not measured due to the lack of coverage during the survey. After the fusion, vision-based points filled this gap, greatly increasing the completeness of the 3D reconstruction.

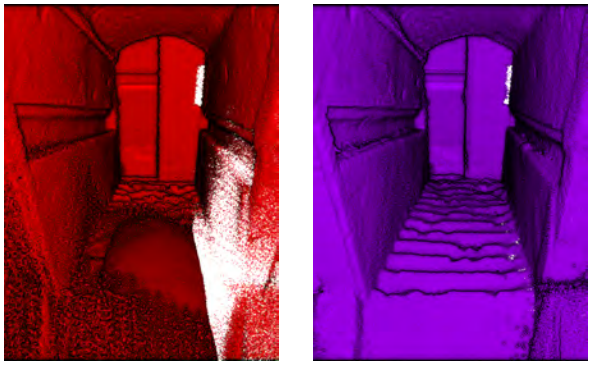


Figure 10. Gaps in in the LiDAR-based point cloud (left) are minimized by integrating the vision-based point cloud (right).

But, beside increasing completeness, there is a potential accumulated noise in the fused point cloud, therefore a filtering procedure is necessary for each original data. The visual point cloud was filtered based on the confidence level of depth maps supporting each point used. On the other hand, the LiDAR point cloud was clean based on (i) a range threshold (e.g. points beyond 5 meters distance) and (ii) a noise filter considering the nearest neighbor's approach (20 nearest neighbor) and a standard deviation threshold (1-sigma). Figure 11 shows the result of the filtering process which, while marginally decreasing the completeness of the 3D reconstruction, is more consistent, with most of the noisy points being removed.

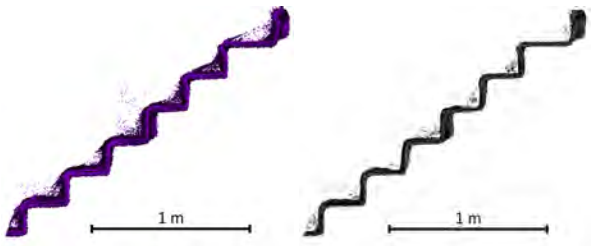


Figure 11. Cross-section of the fused point cloud representing part of the stairs: before (left) and after (right) filtering.

5.4 Quality assessment for the fused 3D data

Figure 12 illustrates the accuracy and completeness curves of the dense points downsampled at 5 mm level against the ground truth (TLS). The x-axis represents the distance threshold in meters whereas the y-axis reports the percentage of points within those thresholds. For the accuracy assessment (Figure 12top), the photogrammetric point cloud (blue curve) achieves ca 70% of points falling within the 6 cm accuracy range. LiDAR SLAM curve (orange curve) improves upon it with 80% of points falling with 6 cm accuracy. Both techniques plateau at ca 95% of points falling within 14 cm distance thresholds.
In contrast, photogrammetry with PGO adjustment (green) achieves over 90% alignment within 5 cm, demonstrating a substantial improvement from the standalone photogrammetry (V-SLAM aided) method. Similarly, the LiDAR with PGO adjustment (red) reaches about 95% alignment at 5 cm, highlighting the impact of optimization in further enhancing LiDAR accuracy. Both photogrammetry (green) and LiDAR (red) with optimization demonstrate a notable improvement with approximately 100% of points falling within the 6 cm accuracy threshold. The point cloud fused with the proposed approach (violet) and aided with the filtering process (grey) also converge at 100% of points falling within 6 cm accuracy, whereas the filtered point cloud and LiDAR SLAM with PGO optimization have both equivalent results.

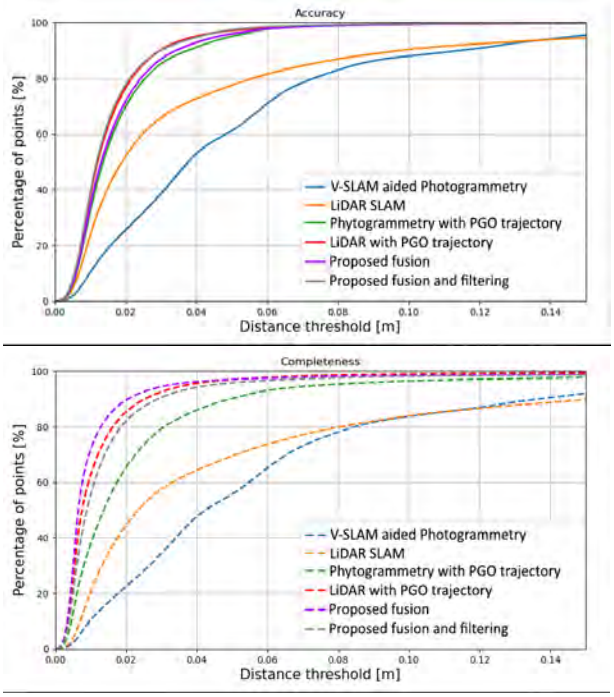


Figure 12. Accuracy (top) and completeness (bottom) assessment results.

The completeness results (Figure 12bottom) show that the photogrammetric and LiDAR SLAM clouds feature some 80% of points falling within the 8 cm distance threshold, with the later having better completeness at 2 and 4 cm distance thresholds. Both techniques plateau at approximately 90% of points with distance thresholds exceeding 14 cm level.

The PGO applied to photogrammetry data yields a significant improvement, achieving over 90% completeness at a 5 cm threshold whereas for LiDAR SLAM data it further increases the coverage reaching approximately 95% within a 5 cm threshold. The point cloud fused with the proposed approach performs nearly optimally and marginally better than the other approaches. Compared to the separately processed photogrammetry and LiDAR SLAM, the optimized techniques are able to overcome the plateau and reach near 100% completeness levels compared to the TLS.

Figure 13 presents the global accuracy evaluation results of each 3D surveying method based on the comparisons of the ground truth with the point clouds generated with 6 studied methods. The V-SLAM aided photogrammetric method (Figure 13a) demonstrates error values higher than 10 cm along the connecting flights of each level (blue and red color regions). The LiDAR SLAM (Figure 13b) shows similar behavior at the entrance and top level of the staircases. Considerable improvement can be seen for both methods enriched with the PGO (Figure 13c-d) and after the proposed fusion (Figure 13e-f).

Finally, Table 1 reports means and standard deviations of the M3C2 signed distances. The mean shift of the photogrammetric solution is probably related to the complexity of the staircase in terms of lighting, overexposure of images at narrow spaces, repetitive turns within narrow spaces limiting the field of views. All other proposed solutions reached close to zero mean shifts, with a high reduction over the standard deviations of the global alignment with respect to the ground truth data. Both fusion methods (with/without filtering) achieved standard deviations close to the PGO LiDAR solution, demonstrating only a tiny compromise in terms of accuracy after fusing with the other point cloud.

Method	M3C2 distance [mm]	
	μ	σ
V-SLAM aided Photogrammetry	-24	74
LiDAR SLAM	-3	60
Photogrammetry with PGO trajectory	-4	27
LiDAR with PGO trajectory	-1	20
Proposed fusion	-2	22
Proposed fusion and filtering	-1	21

Table 1. Mean and standard deviations of the global alignment assessment using M3C2 signed distances (see also Figure 13).

6. CONCLUSIONS

The fusion of vision-based and LiDAR point clouds, through the proposed PGO optimization and filtering, significantly enhances both the completeness and accuracy of the reconstructed models. By improving the pose alignment of the visual approach through the link constraints with the LiDAR SLAM, and further filling gaps in the LiDAR data with photogrammetric points, this approach provides a more continuous and comprehensive representation of complex environments, as demonstrated in the multi-level staircase setting. Hybrid optimization allows the point clouds to achieve near-perfect alignment with the TLS ground truth, effectively overcoming the limitations of standalone methods. The hybrid point clouds achieve high levels of accuracy and reliability, with reduced alignment errors and consistent global alignment statistics for accuracy and completeness. This study highlights the complementary strengths of visual and LiDAR data, showcasing the potential of hybrid methods to produce high-fidelity 3D reconstructions that are robust, scalable, and adaptable to challenging survey environments.

Future works include further research for enhancing the pose graph optimization approach and using smarter and more sophisticated filtering techniques for the hybrid point clouds.

ACKNOWLEDGMENTS

Some of the activities reported in this paper were supported by the EIT-RM project VOT3D - Ventilation Optimizing Technology based on 3D scanning (https://vot-3d.com).

This research was partially funded by “Boostech Valorization Program 2022” funded by the Italian “Piano Nazionale di Ripresa e Resillienza—missione 1, Componente 2, investimento 6” funded by the European union’s—NextGenerationEU with the goal of industrializing the Ant 3D prototype, which is already the subject of the patent proposal n° 102021000000812. The patent was licensed on 24 January 2023.

REFERENCES

Bayram, B., Nemli, G., Özkan, T., Oflaz, O. E., Kankotan, B., Çetin, İ., 2015. Comparison of Laser Scanning and Photogrammetry and Their Use for Digital Recording of Cultural Monument Case Study: Byzantine Land Walls-Istanbul. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, II-5/W3, pp. 17–24.

Będkowski, J., 2024. Open source, open hardware hand-held mobile mapping system for large scale surveys. *SoftwareX*, 25, 101618.

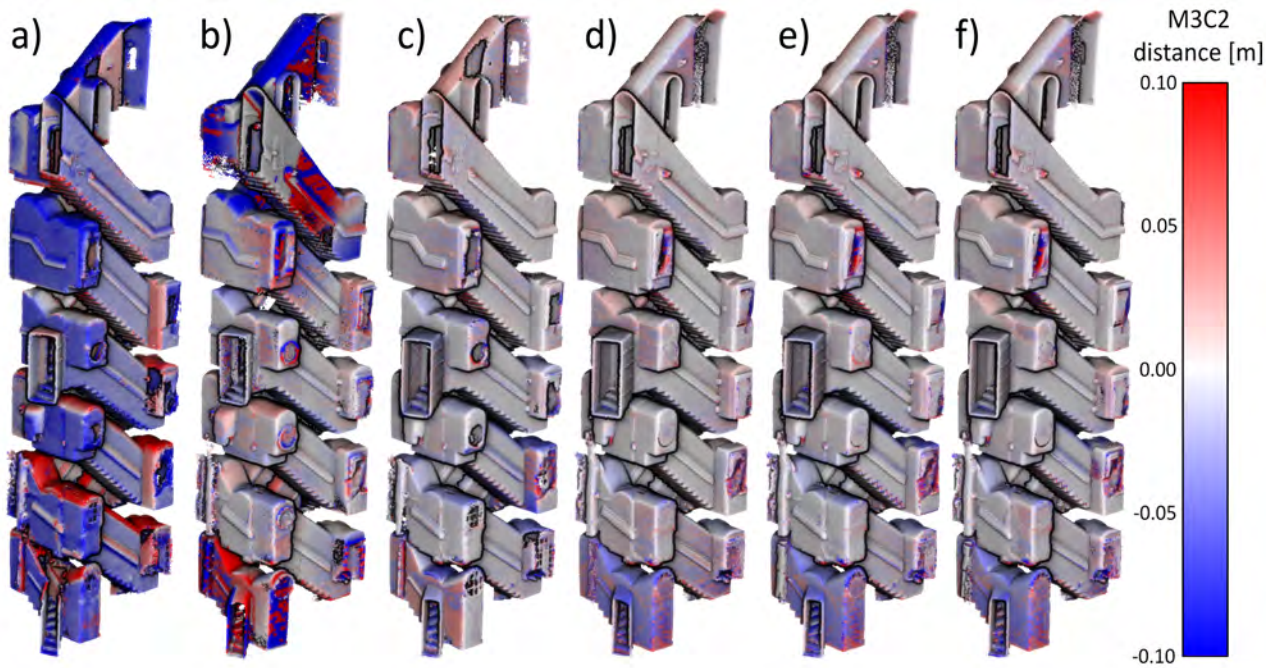


Figure 13. M3C2 signed distances computed against the ground truth for a) V-SLAM aided photogrammetry method, b) LiDAR SLAM, c) V-SLAM with PGO trajectory), d) LiDAR with PGO trajectory, e) fused point cloud, e) fused and filtered point cloud.

Campos, C., Elvira, R., Rodriguez, J. J. G., M. Montiel, J. M., D. Tardos, J., 2021. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*, 37(6), pp. 1874–1890.

Cao, Y., Xu, B., Ye, Z., Yang, J., Cao, Y., Tisse, C.-L., & Li, X., 2018. Depth and thermal sensor fusion to enhance 3D thermographic reconstruction. *Optics Express*, 26(7), 8179.

Cloud Compare, 2024. GPL software, Version 2.13. www.cloudcompare.org/

Dellaert, F., & GTSAM Contributors, 2022. Georgia Tech Smoothing And Mapping, Version 4.2a8 [Computer software]. Georgia Tech Borg Lab. doi.org/10.5281/zenodo.5794541

Di Stefano, F., Torresani, A., Farella, E.M., Pierdicca, R., Menna, F., Remondino, F., 2021. 3D Surveying of Underground Built Heritage: Opportunities and Challenges of Mobile Technologies. *Sustainability*, Vol.13, 13289

Dietmaier, A., McDermid, G. J., Rahman, M. M., Linke, J., Ludwig, R., 2019. Comparison of LiDAR and Digital Aerial Photogrammetry for Characterizing Canopy Openings in the Boreal Forest of Northern Alberta. *Remote Sensing*, 11(16).

El-Hakim, S., Gonzo, L. Voltolini, F., Girardi, S., Rizzi, A., Remondino, F., Whiting, E., 2007: Detailed 3D modelling of castles. *Int. J. of Architectural Computing*, Vol.5(2), pp. 199-220.

El-Hakim, S., Beraldin, J.-A., Remondino, F., Picard, M., Cournoyer, L., Baltsavias, M., 2008: Using terrestrial laser scanning and digital images for the 3D modelling of the Erechteion, Acropolis of Athens. *Proc. DMACH*, pp. 3-16

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., 2002. A Review of Mobile Mapping Systems: From Sensors to Applications. *Sensors*, 22, 4262.

Fassi, F., Achille, C, Fregonese, L., 2011. Surveying and modelling the main spire of Milan Cathedral using multiple data sources. *The Photogrammetric Record*, 26(136), pp. 462-487.

Kadobayashi, R., Furukawa, R., Kochi, N., Otani, H., 2004. Comparison and evaluation of laser scanning and photogrammetry and their combined use for digital recording of cultural heritage. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*

Knapitsch, A., Park, J., Zhou, Q.Y., Koltun, V., 2017. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Trans. Graph. ToG*. 36, pp. 1–3.

Kumar, G. A., Lee, J. H., Hwang, J., Park, J., Youn, S. H., Kwon, S., 2020. LiDAR and Camera Fusion Approach for Object Distance Estimation in Self-Driving Vehicles. *Symmetry*, 12(2), 324.

Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ). *ISPRS journal of photogrammetry and remote sensing*, 82, pp. 10-26.

Medici, M., Perda, G., Sterpin, A., Farella, E. M., Settimo, S., Remondino, F., 2024. Separate and Integrated Data Processing for the 3D Reconstruction of a Complex Architecture. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 48, pp. 249-256.

Nocerino, E., Menna, F., Remondino, F., Toschi, I., Rodríguez-Gonzálvez, P., 2017. Investigation of indoor and outdoor performance of two portable mobile mapping systems. *Proc. SPIE,* Vol. 10332.

Nocerino, E., Stathopoulou, E. K., Rigon, S., Remondino, F., 2020. Surface reconstruction assessment in photogrammetric applications. *Sensors*, 20(20), 5863.

Perfetti, L., Fassi, F., Vassena, G., 2024. Ant3D—A Fisheye Multi-Camera System to Survey Narrow Spaces. *Sensors*, 24(13), 4177.

Remondino, F., 2011. Heritage Recording and 3D Modeling with Photogrammetry and 3D Scanning. *Remote Sensing*, 3(6), pp. 1104-1138.

Sammartano, G., Spanò, A., 2018. Point clouds by SLAM-based mobile mapping systems: accuracy and geometric content validation in multisensor survey and stand-alone acquisition. *Applied geomatics*, 10(4), pp. 317-339.

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

Trybała, P., Kasza, D., Wajs, J., Remondino, F., 2023a. Comparison of low-cost handheld lidar-based slam systems for mapping underground tunnels. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.,* 48, 517-524.

Trybała, P., Kujawa, P., Romańczukiewicz, K., Szrek, A., Remondino, F., 2023b. Designing and Evaluating a Portable LiDAR-Based SLAM System. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-1/W3-2023, pp. 191–198.

Wei, P., Fu, K., Villacres, J., Ke, T., Krachenfels, K., Stofer, C. R., Bayati, N., Gao, Q., Zhang, B., Vanacker, E., Kong, Z., 2024. A Compact Handheld Sensor Package with Sensor Fusion for Comprehensive and Robust 3D Mapping. *Sensors*, 24(8), 2494.

Xu, X., Zhang, L., Yang, J., Cao, C., Wang, W., Ran, Y., Tan, Z., Luo, M., 2022. A Review of Multi-Sensor Fusion SLAM Systems Based on 3D LIDAR. *Remote Sensing*, 14(12), 2835.

Yew, Z. J., Lee, G. H., 2020. RPM-Net: Robust Point Matching Using Learned Features. *Proc. CVPR*, pp. 11821–11830.

Zhao, L., Zhang, H., Mbachu, J., 2023. Multi-Sensor Data Fusion for 3D Reconstruction of Complex Structures: A Case Study on a Real High Formwork Project. *Remote Sensing*, 15(5), 1264.

Zhu, J., Li, H., Zhang, T., 2024. Camera, LiDAR, and IMU Based Multi-Sensor Fusion SLAM: A Survey. *Tsinghua Science and Technology*, 29(2), pp. 415–429.