

V-SLAM-AIDED PHOTOGRAMMETRY TO PROCESS FISHEYE MULTI-CAMERA SYSTEMS SEQUENCES

A. Elalailiyi ^{1,3}, L. Perfetti ², F. Fassi ¹, F. Remondino ³

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

² Università degli Studi di Brescia, Italy - luca.perfetti@unibs.it

³ 3D Optical Metrology (3DOM) Unit, Bruno Kessler Foundation (FBK), Trento, Italy - aelalailiyi@fbk.eu, remondino@fbk.eu

Commission II

KEYWORDS: Multi-camera Systems, V-SLAM, photogrammetry, 3D mapping.

ABSTRACT:

The advent of mobile mapping systems (MMSs) and computer vision algorithms has enriched a wide range of navigation and mapping tasks such as localisation, 3D motion estimation and 3D mapping. This study focuses on Visual Simultaneous Localisation and Mapping (V-SLAM) in the context of two in-house MMSs: Ant3D, a patented five-fisheye multi-camera rig and GeoRizon, a high-resolution stereo fisheye rig. The aim is to leverage V-SLAM to enhance the systems performance in near-real-time and non-real-time 3D reconstruction applications. The research investigates both Monocular and Stereo V-SLAM applied to both MMSs and tackles the challenge of combining the V-SLAM estimated trajectory of one or a pair of cameras with known multi-camera relative orientation. We propose a state-of-the-art code that serves as a flexible and extensible platform for MMSs image acquisition and processing, along with an adapted version of the well-established ORB-SLAM3.0. Evaluation is performed in a cultural heritage challenging setup: the Minguzzi spiral staircase in the Duomo di Milano Cathedral. Performed tests highlight that introducing V-SLAM trajectories as well as pre-calibrated interior orientation and multi-camera constraints improve speed, applicability and accuracy of 3D surveys.

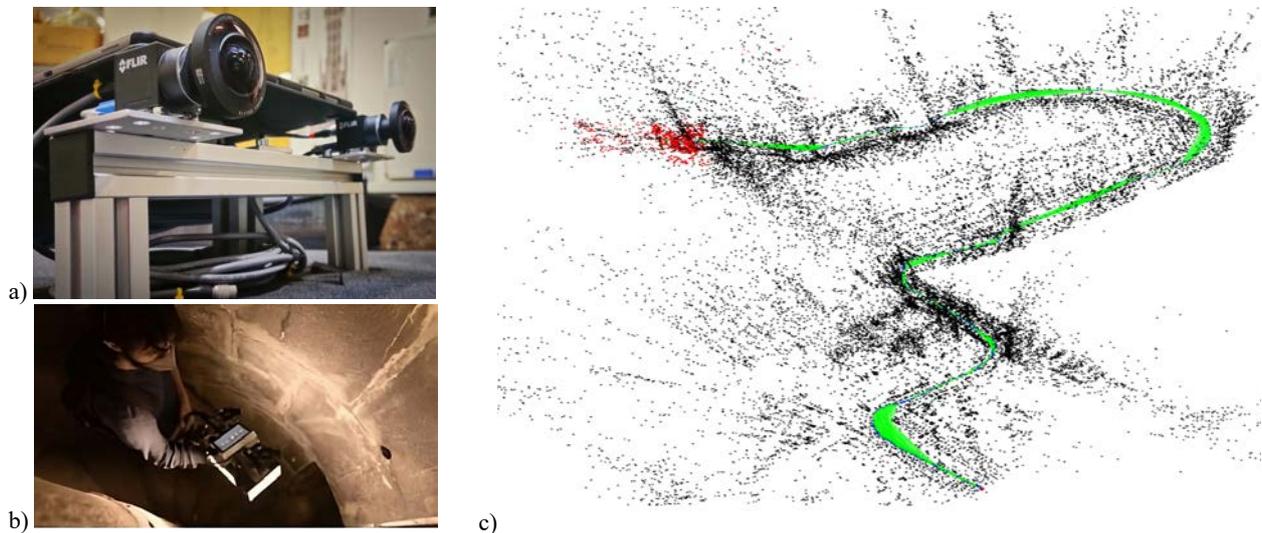


Figure 1: The in-house handheld GeoRizon (a) and Ant3D (b) systems used in this work. An application on Duomo di Milano Cathedral rooftop, with the recovered camera trajectory estimation (green trajectory) and 3D sparse map point cloud (c).

1. INTRODUCTION

The demand for efficient and precise methods to capture and transform reality into 3D digital data is intensifying. From 3D reconstruction to autonomous navigation, augmented reality, and other applications, the need to sense reality and generate 3D digital data, such as point clouds and meshes, to perform and enhance multiple tasks has increased. Over the years, several well-established techniques, such as photogrammetry, 3D scanning, and LiDAR, have been developed and applied extensively across various domains including architecture, heritage conservation, construction, civil engineering, virtual reality, entertainment, and many others. Nevertheless, digitisation, which is often an intricate and time-consuming endeavour, necessitates innovative approaches to accelerate and streamline the workflow to satisfy the increasing demand. The

present investigation focuses on image-based digitisation techniques, exploring the latest advances in acquisition and processing to speed up and enhance image-based reconstruction. Structure from Motion (SfM) and Multi-View Stereo (MVS) are nowadays the most adopted image-based techniques for solving image orientation and mapping the 3D structure of the environment using only a set of overlapped images. This process, made available to non-experts by the various commercial photogrammetric software, involves taking static images and later processing the data offline. This task has proven challenging for large projects regarding acquisition time, processing time, and computational power needed (Leduc et al., 2019). On the other hand, portable handheld mobile mapping systems (MMSs) are a digitalisation trend nowadays, mainly because they offer higher acquisition speeds and flexibility. Range-based MMSs equipped with LiDAR sensors and Simultaneous Localisation And

Mapping (SLAM) algorithms have recently dominated research and industrial domains (Trybala et al., 2023). Similarly, portable multi-camera MMSs solutions (Torresani et al., 2021; Menna et al., 2023; Perfetti et al., 2022a) provide time-effective field operations and added benefits and opportunities such as cost-effectiveness and machine learning prospects (Padkan et al., 2023). Several earlier studies have explored and compared handheld and portable MMSs due to their multiple advantages (Nocerino et al., 2017; Perfetti et al., 2017; Campos et al., 2018; Blaser et al., 2019; Ortiz-Coder & Sánchez-Ríos, 2019; Perfetti, et al., 2022b). MMSs also promise to process the data in real-time on the move. Still, in reality, accurate data processing is mainly performed in post-acquisition in an offline context, and the delivery of products is not immediate and may require several days (Chang et al., 2021).

1.1 Visual SLAM (V-SLAM) integration

SLAM originated in the robotics community to localise robots while simultaneously mapping an unknown environment (Chen et al., 2022;). It is mainly used in autonomous driving and GNSS-denied environments. V-SLAM has been developed using computer vision and image-based techniques (Macario Barros et al., 2022). By analysing, tracking, and matching distinctive features in sequential images, we can estimate the relative positions of the camera, compute the trajectory, and construct a map of the environment. Sharing the same input, V-SLAM and MMS can be used to enhance real-time 3D mapping opportunities (Figure 1c) in terms of speed, flexibility, efficiency, and accuracy of work (Ortiz-Coder & Sánchez-Ríos, 2019; Kuo et al., 2020; Menna et al., 2022; Kaveti et al., 2023).

1.2 Paper objectives

This study is part of a broader ongoing research to enhance the traditional photogrammetric static acquisition and offline processing into dynamic acquisition and near-real-time processing able to provide a poses and trajectory as well as a sparse 3D reconstruction. This initial output can be refined in post-processing.

This paper aims to demonstrate the integration of multi-camera rigs and Visual SLAM methods and evaluate the performance of a complete processing pipeline from acquisition and real-time processing to offline SfM refinement. The focus is mainly on the processing time while surveying challenging narrow spaces and the achievable accuracy. Section 2 introduces the two multi-camera platforms that were used in this study with their respective setup and configuration and describes the handling of the image acquisition and the integration of the ORB-SLAM3.0 V-SLAM algorithm (Campos et al., 2021). Section 3 describes the case study where the experimental setup and testing were conducted. Section 4 presents the field survey conducted to acquire two image datasets with the two multi-camera and the multiple processing methods tested, including two V-SLAM configurations (Mono and Stereo SLAM). Sections 5 and 6 present and discuss the findings of this study, including conclusions and future works.

2. HANDHELD MMS AND V-SLAM INTEGRATION

2.1 GeoRizon – fisheye stereo system

The first multi-camera system tested is a stereo system, called GeoRizon (Figure 1a). It is composed of a handheld device and a backpack unit that houses the processing computer. The handheld device integrates two ultra-wide fisheye cameras (Table 1)

pointing forward and assembled on a rigid bar separated by a 24 cm horizontal baseline. The cameras have a 4096×3000 pixels resolution and use a global shutter readout method, ensuring robustness against fast movements. The lenses (Table 2) are a 4k Large Format Fisheye Lens that can support images with resolutions of up to 20 megapixels, ensuring high-quality, detailed images and providing a 195° wide field of view. The system combines the benefits of wide-angle coverage for tracking, mapping, and high-resolution imaging for detailed data capture, making it suitable for applications such as photogrammetry, industrial inspections, and scientific imaging.

2.2 ANT3D – fisheye multi-camera system

The second multi-camera system is called Ant3D (Figure 1b). As GeoRizon, it is composed of a handheld device and a backpack unit. The handheld multi-camera houses five cameras with a resolution of 5 Megapixels (2448×2048) equipped with fisheye lenses with a 190° field of view. The system has been developed mainly for digitising narrow spaces such as staircases and underground tunnels, owing to its portability and ability to manoeuvre in such environments (Perfetti et al., 2022a,b). The system setup ensures a quasi- 360° view at all acquisition times (excluding the area occupied by the operator), which enables better feature tracking and 3D reconstruction. The motivation behind the setup is to ensure the maximum coverage of the 3D scene and sufficient overlap between the cameras in narrow environments, which results in robust image acquisitions in such areas.

Cameras				
	Brand	Resolution (pixels)	Pixel Size (μm)	Shutter
GeoRizon	FLIR	4096*3000	3.45	Global
Ant3D	FLIR	2448*2048	3.45	Global

Table 1. Cameras' specifications.

Lenses			
	Brand	Focal length (mm)	Field of View (deg)
GeoRizon	Sunex	3.3	195°
Ant3D	Sunex	2.7	190°

Table 2. Lenses' specifications.

2.3 Image acquisition with V-SLAM integration

Handling the image acquisition demands a custom code framework consisting of two units:

- (1) one responsible for camera initialisation and operation: it is designed to be flexible and extensible to suit both single and multi-camera capture. It is responsible for controlling the camera exposure parameter, executing the capture trigger, and passing the captured images to be processed by the second unit;
- (2) one responsible of the implementation of ORBSLAM3.0 adapted to be efficient in our processing framework: this unit operates in parallel to the capture unit, awaits for images, perform the V-SLAM processing and store the solution.

The linkage between the two units is established by exploiting ROS (Robot Operating System), which serves as a middleware. From unit 1, the images acquired are published as ROS topics, which are taken over by Unit 2. The code employs parallel directives for parallelising specific computational tasks, using multicore architectures to optimise the performance.

The integration of ORBSLAM3.0 (Campos et al., 2021) introduces a powerful V-SLAM component for real-time mobile mapping activities. It provides real-time image processing for localisation and 3D sparse mapping. The current integrated version of ORBSLAM3.0 has been modified to work seamlessly with ROS and has been adapted to work efficiently with the high-resolution fisheye images acquire at high frame rates, up to 30 fps for each camera.

Finally, to ensure the integration of our approach into a broader data processing workflow, the outputs of the image acquisition can be transformed into results compatible with renowned software solutions, such as Agisoft Metashape and OpenCV. This expands the usability of our solution, enabling a broader spectrum of applications in photogrammetry and computer vision tasks.

3. CASE STUDY

The chosen case study is part of an extensive survey of the Duomo di Milano Cathedral aimed at building a detailed, accurate 3D model supporting on-site construction in an online smart manner (Achille et al., 2020). In particular, the work focuses on the Minguzzi spiral staircase (Figures 2), located at the front right corner of the cathedral and extended from the floor level to the rooftop at a height of 25m. The inner dimensions of the staircase consist of a very narrow passage of about 70cm in width and extreme curvature.

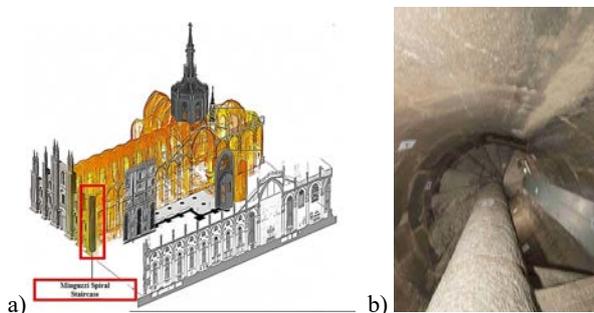


Figure 2: The Minguzzi spiral staircase within the Duomo di Milano cathedral (a) and the illuminated narrow staircase (b).

Artificial and natural lights are dim. Such an environment is a challenging case study for most surveying techniques. Terrestrial laser scanning would suffer the highly confined spaces and a limited line of sight leading to occluded areas. Therefore, it is necessary to perform multiple setups for comprehensive coverage, thus, time- and effort-consuming surveys. LiDAR-based mobile scanners might be considered a better flexible option, thinking of the lack of illumination and mobility requirements. However, due to the fast dynamic change of geometric features along the acquisition track, the drift error would exponentially increase, especially without the ability to perform efficient loop closures. Hence, visual mobile mapping systems are chosen due to their flexibility, low cost and ease of use. The mapping systems are equipped with ultra-wide fisheye lenses and supplementary lighting, to support better feature tracking and mapping. Real-time image processing is also envisioned to monitor the acquisitions and 3D results at every step and take fast decisions based on the outputs.

4. DATA ACQUISITION USING MULTIPLE CONFIGURATIONS

Both mobile mapping systems (Section 2) are equipped with artificial LED lights. The V-SLAM method is operated in real-

time, allowing the operator to monitor the acquisition at every instant through an attached screen (Figure 3).

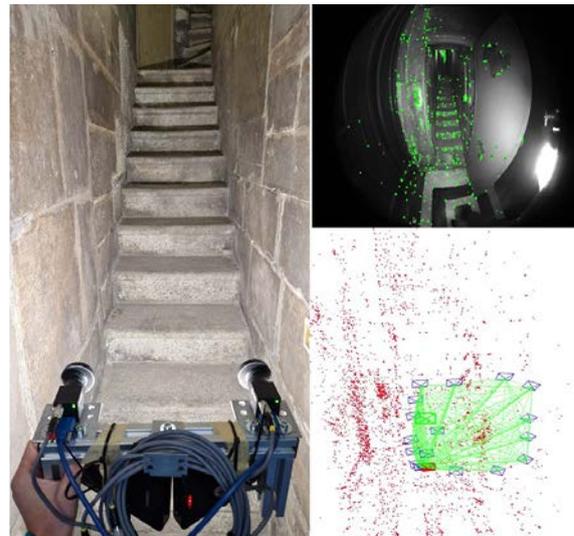


Figure 3: On-site acquisition with the GeoRizon MMS (left). Real-time keypoint detection and simultaneous estimation of camera poses and 3D points (right).

The staircase dimensions and curvature required thorough understanding for the acquisition to run smoothly without losing track at any instant and to ensure a fully oriented real-time acquisition with acceptable initial camera trajectory and 3D point cloud reconstruction. The spiral nature and extreme curvature provided limited viewpoints to maintain a sufficient number of distinct features for reliable real-time tracking. The strong curvature obstructs the line of sight, even between instant consecutive images, therefore obstructs the already tracked features after taking as little as three steps along the passage. This can also affect the stereo-matching process due to the limited matches between current and previously tracked frames, resulting in a higher drift error. Hence, the use of fisheye cameras in such an environment is strategic as their wider field of view provides larger overlapping regions between stereo images and distinct features in view for longer instants of time, which aids against loss of tracking. For both handheld systems, the image acquisition phase was done in one continuous and upward spiral motion without any stops and took approximately 8 minutes.

The survey was initialised at the floor entrance of the tower and finalised at the rooftop exit (Figure 4a). For a better evaluation of the processing methods and a better comparison of the tests discussed later, a single dataset per each system was used to test the different V-SLAM setup configurations by re-processing the data offline. 31 natural reference points (Figure 4b) were identified along the passage, starting at the main entrance, going up the main body of the staircase, and finishing around the exit area of the tower at the rooftop level. The reference points, also used for scaling and georeferencing, are primarily used to check the accuracy of both systems in the final 3D reconstruction by measuring the drift error accumulated from the bottom to the top. This allows an accurate comparison and understanding of each multi-camera system's performance and the processing method. As mentioned, a single image dataset for each handheld system was re-processed using an offline implementation of ORBSLAM3.0, testing both the mono and the stereo setup. After each process, offline refinements for achieving higher accuracy are performed within Agisoft Metashape. The performed tests allow us to analyse the accuracy of the V-SLAM-aided processing pipeline compared to a classical photogrammetric method. The comparison is twofold:

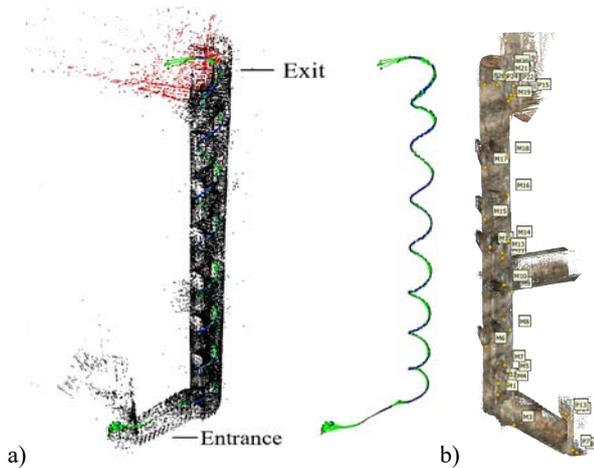


Figure 4: Minguzzi staircase: trajectory and sparse point cloud recovered with the V-SLAM approach (a) and marker distribution along the staircase (b).

- (i) overall accuracy, estimated by computing the drift error along the path, where similar results are to be expected;
- (ii) overall processing time, where we expect the V-SLAM aided approach to significantly reduce the offline computation requirements.

The overall comparison process is shown in Figure 5.

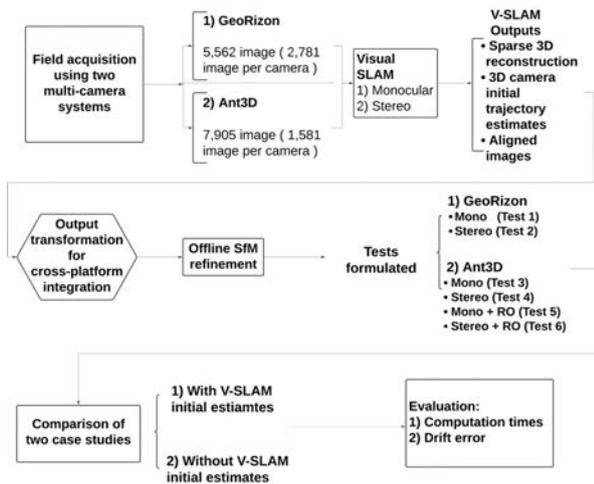


Figure 5: Followed approach.

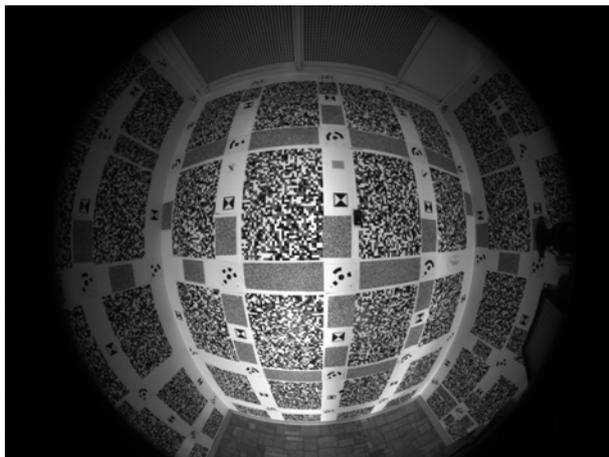


Figure 6: Fish-eye image of the employed 3D calibration wall used to calibrate both handheld mobile mapping systems.

Both multi-camera systems were pre-calibrated using a 3D calibration setup (Figure 6) equipped with 40 frame-filling coded reference points of known 3D coordinates distributed uniformly across the field. The wall was populated with rich texture in the form of random speckled texture patterns printed at different resolutions to provide texture-rich scenes with clear, well-defined, and trackable corresponding features. The available 3D points were evenly divided into Ground Control Points (GCPs) and Check Points (CPs) for RMSE estimations (Table 3).

	RMSE (mm) on CPs			
	X	Y	Z	XYZ
ANT3D	0.30	0.20	0.16	0.37
GeoRizon	0.16	0.10	0.10	0.21

Table 3. CPs' RMSEs of the MMS calibration process.

4.1 GeoRizon

The full images dataset acquired by GeoRizon consisted of 5,562 image (2,781 image per camera). A total of 847 images for a single camera were processed using the Monocular V-SLAM, ensuring full coverage of the Minguzzi staircase, while a total of 4,872 images (2,436 images per camera) were processed in the stereo V-SLAM configuration. No relative orientation of the second camera was done with respect to the monocular case since stereo mode replaces that task in a two-camera-only system.

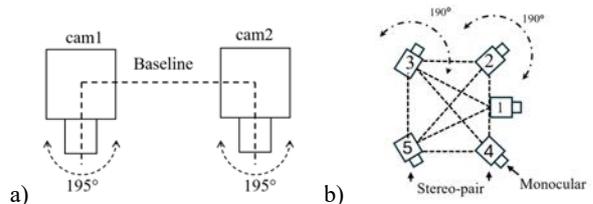


Figure 7: GeoRizon (a) and Ant3D (b) camera configurations

4.2 Ant3D

The image acquisition was simultaneously performed for all five cameras. A total of 7,905 images by the five cameras (1,581 images per camera) were acquired. However, the V-SLAM processing operated solely using the monocular and stereo ORBSLAM3.0, as later described. The monocular V-SLAM estimated EO for the images referring to only one camera, and the V-SLAM stereo setup estimated the EO for the stereo pair. Images acquired with the other cameras entered the processing during the offline SfM refinement, where they were constrained to the raw-oriented images based on the pre-calibrated multi-camera Relative Orientations (ROs) parameters.

4.2.1 V-SLAM Mono

The front right camera of Ant3D (Figure 7b - camera 4) was used in the monocular V-SLAM configuration. A total of 552 images out of the 1,581 images taken by that camera were oriented by the V-SLAM and were enough to cover the whole Minguzzi staircase. During the offline SfM refinement, the Mono dataset was split in two, for Test 3 (Figure 5), the post-process refinement was operated on the pre-oriented 552 images of camera 4 only; while for Test 5 (Figure 5), a total number of 552 x 5 = 2,760 images were processed and multi-camera relative-orientation (RO) constraints were imposed to link the pre-processed images with estimated raw EO with the others.

4.2.2 V-SLAM Stereo

For the stereo configuration, the right pair was used (Figure 7b - cameras 4 and 5), and 1,549 images per camera were oriented. During the offline SfM refinement, as for the mono setup, the

Stereo dataset was split in two: in Test 4 (Figure 5) refinement was run for the pre-oriented stereo images only. In Test 6 (Figure 5), the remaining three cameras were used with the multi-camera RO constraint, with respect to the already oriented stereo pair, and the total number of offline processed images was $1,549 \times 5 = 7,750$.

5. RESULTS

Table 4 presents the results obtained by the six configurations (Tests 1 to 6) studied in this work. Two cases were compared:

- (1) the V-SLAM aided method, using the V-SLAM output as initial EO estimates for the offline refinement in Metashape: this case has been further divided into two configurations for GeoRizon, and into four configurations for Ant3D (Tests 1-6 shown in Figure 5);
- (2) the classical offline photogrammetric process, without any initial estimates for the EO, on the same six configurations.

To assess the drift error in the different tests, 7 markers located at the base of the staircase were used as GCPs, and the remaining 24 markers were set as CPs to compute RMSEs (Table 4).

Regarding processing times, the V-SLAM-aided case performed much better than regular SfM. For the GeoRizon tests, we observed a decrease in the offline processing time with respect to the regular SfM method of ca 36% and ca 55% for Tests 1 and 2, respectively. On the same note, for the Ant3D tests, we observed a decrease of ca 50%, ca 58%, ca 60% and ca 73% for Tests 3 to 6.

Regarding the accuracy evaluation, the GeoRizon Mono setup (Test 1) resulted in an RMSE of 12 cm, while the GeoRizon Stereo (Test 2) resulted in an RMSE of 6.7 cm. The drift error is effectively reduced by processing more images and employing a more robust image network. A nearly identical result is obtained for Test 2 without initial EO. On the other hand, Test 1 performs

better with V-SLAM pre-processing, suggesting that a good V-SLAM pre-orientation of the images can effectively constrain and improve a weak image dataset.

As for Ant3D, RMSEs obtained from the V-SLAM aided cases are very similar to that obtained from regular offline SfM, registering an improvement of 5% and 25% for Test 3 and Test 4, that is, when only Mono or Stereo images are processed (weaker image network). No improvement is visible for Test 5 and Test 6 using the full multi-camera dataset (more robust image network).

Looking only at the V-SLAM aided results, the Ant3D Mono setup (Test 3) performed very poorly with an RMSE of 71 cm, which can be explained by the image quality of the images acquired by camera 4 that shows non-ideal illumination.

Ant3D Stereo (Test 4) performs much better than the monocular case (Test 3) with an RMS of 12 cm on par with GeoRizon Stereo.

Finally, Test 5 and Test 6, with constrained RO of all five cameras, performed slightly better with an RMSE of around 10 and 11 cm, respectively. Test 5 and Test 6 performed very similarly but were not identical as expected. Also, the improvement in processing all multi-camera images from processing only the stereo pair is marginal, suggesting that the V-SLAM Stereo estimated EO for Ant3D is not as accurate as those of GeoRizon. That again can be explained by the lower image quality of the Ant3D Stereo setup compared to the GeoRizon Stereo setup in terms of ideal illumination, which is due to the GeoRizon stereo pair being front forward while the Ant3D one points towards the wall very up-close. The difference in terms of the overall resolution of the camera may also play a role in the better results obtained by GeoRizon Stereo.

Visual results of the image orientation results are reported in Figure 8. Mesh models derived by handheld systems are presented in Figure 9.

		Total numb. of images	V-SLAM as initial estimates					Without initial estimates				
			Processing time (min)	RMSE (cm)				Processing time (min)	RMSE (cm)			
				X	Y	Z	XYZ		X	Y	Z	XYZ
GeoRizon (4096 × 3000)	Mono (Test 1)	847	7	2.31	2.85	11.51	12.08	11	3.35	4.95	29.53	30.13
	Stereo (Test 2)	4,872	80	2.19	3.87	4.96	6.66	180	1.87	3.81	4.91	6.50
Ant3D (2448 × 2048)	Mono (Test 3)	552	4	4.85	5.86	70.95	71.35	8	4.97	6.21	73.91	74.34
	Stereo (Test 4)	3,098	25	6.46	9.35	3.86	12.00	60	9.26	3.63	12.23	15.76
	Mono + RO (Test 5)	2,760	20	1.65	3.03	9.32	9.94	50	2.00	3.00	9.33	10.00
	Stereo + RO (Test 6)	7,750	55	4.64	2.79	9.46	10.90	206	4.02	3.30	9.51	10.84

Table 4. Comparisons between photogrammetric results achieved with/without initial estimates of the exterior orientation parameters (EO) from the V-SLAM solution.

6. CONCLUSIONS

The study examined the performances of the two in-house handheld mobile mapping systems named GeoRizon and Ant3D. Two main scenarios were compared: one utilising V-SLAM processing output as initial EO estimates and the other without any initial EO estimates. Regarding computational efficiency, the study's findings highlight a significant reduction in processing time when utilising initial estimates for camera alignment. This efficiency gain is particularly significant in scenarios with a higher number of images, where the advantages of leveraging initial image orientation estimates from V-SLAM become more pronounced. Regarding accuracy, performing SfM with initial EO estimates shows an increase in accuracy in all tests with

various degrees of effectiveness, from no improvement for robust image blocks to significant improvements for weak image blocks. For both systems, the mono setup (Test 1 and Test 3) proved not optimal as should be expected. These image datasets do not provide complete coverage of the environment or scale information. The drift error from these tests gives an indication of the robustness of the image dataset processed for the V-SLAM estimation. It is evident that the front-facing camera used for V-SLAM Mono tracking in the stereo system provides a better image dataset than the side-facing camera from Ant3D.

This comparison suggests that it is worth investigating selecting a different camera or stereo pair from the Ant3D multi-camera to achieve optimal V-SLAM tracking.

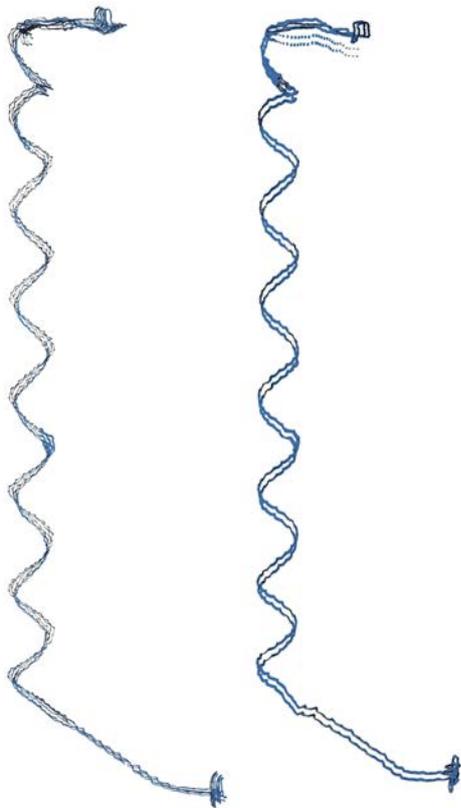


Figure 8: Recovered camera trajectories for Ant3D (left), and GeoRizon (right).

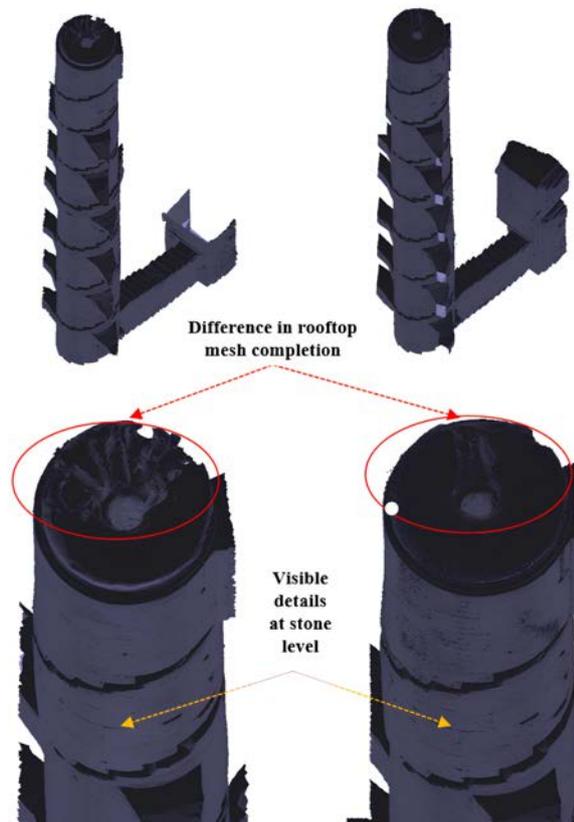


Figure 9: 3D mesh models: GeoRizon (left), Ant3D (right).

The V-SLAM Stereo setup produced better results for both systems compared to the mono setup, again highlighting the stronger image dataset acquired. On the other hand, with the ORBSLAM Stereo setup, many more images can be acquired, clearly resulting in elongated processing time but not always improved accuracy. Indeed, among Ant3D tests, Test 5, where the full multi-camera has been constrained to the EO from V-SLAM Mono, resulted in a lower drift error with a lower count of images and processing times compared to Test 6. The results also suggest that image resolution and illumination quality play a role in determining the quality of the outcome. Indeed, GeoRizon was able to achieve the lower drift error, and the images acquired are both at a higher resolution with respect to Ant3Ds and have better illumination.

Concerning the generated 3D models, both systems provided detailed 3D reconstruction, but ANT3D resulted in a more complete reconstruction due to their larger field of view coverage between the five fisheye views.

Finally, these findings provide valuable insights for practitioners seeking optimal solutions in similar photogrammetric applications using multi-view visual MMS and V-SLAM techniques. Future works will further focus on improving accuracy and lowering computational times to reach near-real-time digitisation.

ACKNOWLEDGEMENTS

The authors would like to thank the VENERANDA FABBRICA DEL DUOMO DI MILANO for supporting us in conducting the tests at the Duomo di Milano Cathedral (Italy).

REFERENCES

- Achille, C., Fassi, F., Mandelli, A., Perfetti, L., Rechichi, F., Teruggi, S., 2020. From a traditional to a digital site: 2008–2019. The History of Milan Cathedral Surveys. *Research for Development*, pp. 331-341. Springer, Cham.
- Blaser, S., Nebiker, S., Wisler, D., 2019. Portable image-based high performance mobile mapping system in underground environments – system configuration and performance evaluation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-2/W5*, 255–262.
- 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), 1874–1890.
- Campos, M., Tommaselli, A., Honkavaara, E., Prol, F., Kaartinen, H., El Issaoui, A., Hakala, T., 2018. A backpack-mounted omnidirectional camera with off-the-shelf navigation sensors for mobile terrestrial mapping: development and forest application. *Sensors*, 18(3).
- Chang, A., Jung, J., Landivar, J., Landivar, J., Barker, B., Ghosh, R., 2021. Performance evaluation of parallel structure from motion (SfM) processing with public cloud computing and an on-premise cluster system for UAS images in agriculture. *ISPRS International Journal of Geo-Information*, 10(10), 677.

- Chen, W., Zhou, C., Shang, G., Wang, X., Li, Z., Xu, C., Hu, K., 2022. SLAM overview: from single sensor to heterogeneous fusion. *Remote Sensing*, 14(23), 6033.
- Elhashash, M., Albanwan, H., Qin, R., 2022. A review of mobile mapping systems: from sensors to applications. *Sensors*, 22, 4262.
- Kaveti, P., Vaidyanathan, S. N., Chelvan, A. T., Singh, H. 2023. Design and evaluation of a generic visual slam framework for multicamera systems. *IEEE Robotics and Automation Letters*, 8(11), 7368–7375. doi.org/10.1109/LRA.2023.3316609
- Kuo, J., Muglikar, M., Zhang, Z., Scaramuzza, D., 2020. Redesigning SLAM for arbitrary multicamera systems. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 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 Surface Dynamics*, 7(1), 97–106.
- Macario Barros, A., Michel, M., Moline, Y., Corre, G., Carrel, F., 2022. A comprehensive survey of Visual SLAM algorithms. *Robotics*, 11(1), 24.
- Menna, F., Battisti, R., Nocerino, E., Remondino, F., 2023.: FROG: a portable underwater mobile mapping system. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-1/W1-2023, 295–302.
- Menna, F., Torresani, A., Battisti, R., Nocerino, E., Remondino, F., 2022. A modular and low-cost portable VSLAM system for real-time 3D mapping: from indoor and outdoor spaces to underwater environments. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W1-2022, 153–162.
- Nocerino, E., Menna, F., Remondino, F., Toschi, I., Rodríguez-González, P., 2017. Investigation of indoor and outdoor performance of two portable mobile mapping systems. SPIE Proc. 10332, "Videometrics, Range Imaging, and Applications XIV", Munich, Germany.
- 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.
- Padkan, N., Battisti, R., Menna, F., Remondino, F., 2023. Deep learning to support 3D mapping capabilities of a portable VSLAM-based system. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-1/W1-2023, 363–370.
- Perfetti, L., Elalaily, A., Fassi, F., 2022a. Portable multicamera system: from fast tunnel mapping to semi-automatic space decomposition and cross-section extraction. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLIII-B2-2022, 259–266.
- Perfetti, L., Teruggi, S., Achille, C., & Fassi, F., 2022b. Rapid and low-cost photogrammetric survey of hazardous sites, from measurements to VR dissemination. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-2/W1-2022, 207–214.
- Perfetti, L., Polari, C., Fassi, F., 2017. Fisheye photogrammetry: Tests and methodologies for the survey of narrow spaces. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W3, 573–580.
- 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.
- Trybała, P., Kasza, D., Wajs, J., Remondino, F., 2023. Comparison of low-cost handheld LiDAR-based SLAM systems for mapping underground tunnels. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII-1/W1-2023, 517–524.