The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W5-2025
13th International Conference on Mobile Mapping Technology (MMT 2025)
"Mobile Mapping for Autonomous Systems and Spatial Intelligence", 20–22 June 2025, Xiamen, China

STELVIO: Exploring Factor Graphs for a Robust Stereo-Visual-LiDAR-Inertial Odometry

Paweł Trybała[1], Luca Morelli[1], Samuele Facenda[1], Armando Vittorio Razzino[1,2], Fabio Remondino[1]

[1] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Web: http://3dom.fbk.eu – Email: (ptrybala, lmorelli, sfacenda, arazzino, remondino)@fbk.eu
[2] Dept. Mathematics, Computer Science and Physics, University of Udine, Italy

Commission II

KEY WORDS: data fusion, multi-sensor systems, odometry, mobile mapping, SLAM

ABSTRACT:

Accurate and robust odometry is critical for mobile mapping and autonomous navigation, particularly in complex environments where single-sensor approaches struggle. While LiDAR and visual odometry each provide valuable motion estimation, they are susceptible to failures in conditions unfavorable for the specific odometry type. Fusing multiple modalities enhances robustness, yet effective integration remains challenging due to differences in heterogenous sensor data representation. This study presents STELVIO, a flexible factor graph-based framework for Stereo LiDAR-Visual-Inertial Odometry. By combining LiDAR-inertial odometry with stereo visual odometry, STELVIO improves trajectory estimation by leveraging the strengths of each modality. The system introduces adaptive fusion strategies, ranging from loose, odometry-only pose graph coupling to an extensive factor graph approach, utilizing visual features and LiDAR-derived range factors. This modular structure allows for balancing computational efficiency with robustness, making it suitable for real-time applications and accuracy-oriented mapping applications. Evaluation is conducted using an in-house mobile mapping system in a challenging indoor environment. Initial results highlight the effectiveness of the fusion approach in reducing drift and improving localization consistency compared to single-sensor methods. The findings demonstrate the potential of multi-sensor integration for robust and scalable mobile mapping solutions.

1. INTRODUCTION

In the last decade, lightweight mobile mapping systems have gained popularity in the geomatics community, driven by the democratization of hardware access and continuous improvements in the quality of generated 3D data. These advancements are progressively approaching survey-grade accuracy achieved by well-established methods: static photogrammetry or terrestrial laser scanning. In many fields of applications, such as construction monitoring and Scan-to-BIM (Roman et al., 2023, Vassena et al., 2023), open-pit and underground mining (Wajs et al., 2021; Trybała et al., 2024), forestry (Muhojoki et al., 2024) and vehicle positioning in autonomous navigation (Dai et al., 2023), modern mobile mapping solutions often meet the required accuracy standards and provide useful and trustworthy 3D data.

Alongside commercially available solutions, such as handheld scanners, backpack-mounted systems, and drone-based mapping kits developed by surveying equipment manufacturers, the open-source community has contributed significantly to the evolution of Simultaneous Localization and Mapping (SLAM) algorithms. In particular, both visual- and LiDAR-centric odometry approaches have seen continuous refinements, improving the accuracy of trajectory estimation in real-world conditions. Since mapping quality is highly dependent on hardware-related sensor characteristics (e.g., resolution, physical pixel size, range measurement precision, sampling frequency), as well as environmental conditions (e.g., distance to scanned objects, achieved ground sampling distance (GSD), surface albedo, and texture richness), significant research efforts have been dedicated to enhancing sensor trajectory estimation robustness and accuracy. In this aspect, numerous studies indicated the strengths and weaknesses of different sensor modalities employed in SLAM, including LiDAR, radar, cameras, and inertial measurement units (IMUs). A clear consensus in the scientific community is that sensor fusion can bring substantial benefits to mobile mapping frameworks, increasing their robustness, accuracy and richness of the generated 3D data. For instance, comparisons of LiDAR-only versus LiDAR-inertial SLAM consistently demonstrate the benefits of multi-sensor integration (Fasiolo et al., 2023).

While coupling inertial measurements with a primary sensor can enhance motion estimation, it does not significantly improve environmental perception, as only one sensor actively captures information about the surroundings. In contrast, fusing visual and LiDAR data offers greater potential for enhancing SLAM performance, while enhancing data for the environment reconstruction, both in terms of geometry and colorimetry. Yet, this integration introduces notable challenges due to the fundamentally different nature of the data each sensor provides (2D vs 3D), necessitating advanced processing techniques for effective fusion.

1.1 Aim of the study

In this work, we propose a fusion framework for flexible coupling of LiDAR-inertial (or pure LiDAR) and stereo visual odometry solutions, built on top of existing odometry approaches. Due to the distinct list of supported input data types, we call it STELVIO: **Ste**reo-**L**iDAR-**V**isual-**I**nertial **O**dometry. It easily adapts to different components, especially utilizing various visual local feature extraction and matching strategies (Jin et al., 2021). The goal of our method is to enhance the long-term odometry precision and reliability in cases where one of the odometry modalities would struggle through inclusion of stable, high-quality landmarks.

The key contribution of this work is the use of a factor graph-based fusion approach for LiDAR and visual odometry, supported by deep learning-based image tie points and LiDAR-based metric depth measurements. This implementation allows to adjust the computational effort needed for data processing, so that the time constraints of specific application can be met, or, alternatively, greater robustness can be pursued in a more extensive fusion approach. The modularity of the method enables inclusion of different odometry frameworks, enhancing its reusability with the future developments in the field.

We perform evaluation of the proposed method on an indoor sequence, acquired with our in-house built mobile mapping device – the hardware component of STELVIO. We compare the results of multiple STELVIO configurations to single-mode solutions (i.e., stereo visual odometry, LiDAR-inertial odometry) and a state-of-the-art fusion-based baseline.

2. RELATED WORKS

The topic of a multi-modal LiDAR-Inertial-Visual Odometry (LIVO) has been explored by various researchers. FAST-LIVO2 (Zheng et al., 2025) fuses LiDAR, Visual and Inertial Odometry via an Error-State Iterated Kalman Filter (ESIKF), combining scan registration, image alignment and IMU readings for performing the state update. The method also maintains a tightly-coupled 3D voxel-based map structure to fuse the visual and LiDAR map points. However, FAST-LIVO2 utilizes only data coming from a monocular camera and requires a precise synchronization of all sensors involved.

LVI-SAM (Shan et al., 2021) adopts a factor graph approach for a tightly coupled SLAM framework. The Visual Inertial Odometry (VIO) couples LiDAR Inertial Odometry (LIO). The former performs visual feature tracking, while the latter extracts feature depth from the LiDAR data. Furthermore, the LiDAR scan matching is used to refine loop closures identified by the visual submodule. Both FAST-LIVO2 and LVI-SAM can retain pose estimation during a temporal outage of one of the sensors.

R3LIVE++ (Lin and Zhang, 2024) uses the LIO to build the geometry of the local map in a manner similar to FAST-LIO2 (Xu et al., 2022). The VIO module perform standard feature tracking, but instead of minimizing the photometric error, as most methods, it minimizes the radiance differences of the projected landmarks. The two systems exploit the IMU data jointly with visual and LiDAR in the state estimation. The framework constructs a single 3D point cloud, named radiance map, which stores the 3D reconstruction of the environment based on points obtained from both LiDAR and visual modules. R3LIVE++ operates at individual pixel level, relies on an accurate initial state estimation and employs a dense direct method in the VIO, resulting in reduced robustness and increased computational load.

CLIC (Lv et al., 2023) proposes a continuous-time fixed lag smoothing approach. It enables using both LiDAR-Inertial (LI) and LiDAR-Inertial-Camera (LIC) systems. CLIC maintains temporal and key-frame constant-size sliding windows by fusing asynchronous sensor measurements and marginalizing older states and features in a factor-graph optimization.

Coco-LIC (Lang et al., 2023) is an extension of CLIC. It provides a tightly-coupled solution instead of a loosely-coupled one and is based on non-uniform B-splines for trajectory representation, instead of the uniform B-splines used by CLIC. The LIC data are fused without interpolation and the tight coupling is achieved by formulating frame-to-map reprojection errors for the current frame, exploiting the LI-reconstructed point cloud and the optical-flow tracking of pixels.

SR-LIVO (Yuan et al., 2024) employs a sweep reconstruction module that aligns the timestamp of a captured image with the reconstructed sweep from the LiDAR and greatly improve frame synchronization between the sensors. The LIO component estimates the state of the system and performs environment reconstruction in real time. However, the vision module is not used in pose estimation; it only optimizes the camera parameters to improve the texturing of the LiDAR-based 3D reconstruction.

In summary, regarding the most influential LIVO methods, the solutions tend to rely on direct or semi-direct algorithms for the visual data processing (Yuan et al., 2023; Fan et al., 2025), with feature-based visual odometry backbones being investigated only recently (Zhou et al., 2025).

3. METHODOLOGY

3.1 STELVIO mobile mapping system

To test the proposed approach, the open-source and open-hardware mobile mapping device MandEye (Będkowski, 2024) has been extended. Originally, the solution enables data acquisition with a Livox Mid-360 LiDAR and its internal 6 Degrees-of-Freedom IMU. To this set, a pair of 2MP ELP cameras with global shutters is added (Figure 1). We employ a custom kernel patch on a Linux-based system of the controller device for minimizing image acquisition delays. Nevertheless, the speed of the stereo image acquisition is limited by the hardware components of the data storage device (Raspberry Pi 4). The images are acquired synchronously by both cameras and although these low-cost sensors do not allow performing strict hardware synchronization, our tests show that the software-side synchronization allows us to acquire images within the timespan of a few milliseconds of the nearest point cloud timestamp. To address this minor discrepancy, we incorporated the so-called sweep reconstruction method, proposed by Yuan et al. (2024) in SR-LIVO, in turn synchronizing the LiDAR point clouds with the actual timestamps of the images. Jointly with the proposed novel data processing method, the developed hardware creates a STELVIO mobile mapping system.

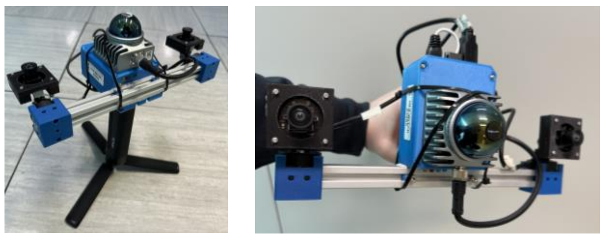

Figure 1: The hardware of the STELVIO mobile mapping system, an extension of MandEye (Będkowski, 2024).

3.2 Software system components

3.2.1 LiDAR-inertial odometry

As the core solution utilizing point cloud and inertial data, we use FAST-LIO2 (Xu et al., 2022), a LiDAR–inertial odometry framework that achieves real-time performance and high accuracy in various conditions (Trybała et at., 2023). The method uses a direct registration approach for associating raw LiDAR points to an incrementally built local map, managed dynamically through incremental k-d tree (ikd-Tree) structure. Point cloud motion distortion is compensated using IMU-driven back-propagation. The main component of FAST-LIO2 for high-accuracy state estimation is a tightly-coupled Iterated Kalman Filters on Manifolds (IKFoM) (Xu and Zhang, 2021), which fuses the point cloud matching and preintegrated IMU information into a state estimate. Nonetheless, any frame-based, direct or indirect, LiDAR odometry approach can be used in a plug-and-play manner with STELVIO.

3.2.1 Stereo Visual Odometry

For the visual odometry component, we employed COLMAP-SLAM (Morelli et al., 2023), a software framework that integrates feature-based odometry with windowed bundle adjustment, incorporating deep learning-based algorithms for extracting and matching local features in images. COLMAP-SLAM supports multi-camera systems in any configuration and is compatible with both pre-calibrated and non-calibrated camera setups. In this study, we utilized ALIKED (Zhao et al., 2023) for local feature extraction, which leverages convolutional neural networks optimized for real-time applications, along with

LightGlue (Lindenberger et al., 2023), a matcher based on graph neural networks that dynamically adjusts matching speed based on the perceived complexity of the image.
Any feature-based visual odometry method could be directly used in STELVIO in place of COLMAP-SLAM. Although a direct visual odometry method could be utilized as a source of pose estimation in our method, due to the additional role of the visual features, an additional layer of feature extraction would be required.

3.2.2 LiDAR-Visual Fusion Framework: STELVIO

The core idea of STELVIO is the seamless fusion of odometry data from two independent systems, supported by the selective tracking of distinct high-quality 3D landmarks, both visible in an image sequence and identifiable in the LiDAR point clouds. In this way, we locally strengthen the trajectory estimation via factor graph optimization, which acts as a local bundle adjustment, while retaining lightness of the solution by means of discarding less informative data for pose estimation in the short and medium time horizon.
An additional benefit of our approach is the complementariness of the odometry solutions. As STELVIO is built on top of independent odometry estimators with equal importance, it can retain pose estimation when any of the two solution fails. This is a key advantage in cases of performing mapping based on only relative positioning techniques (i.e., dead reckoning), which is common in environments without access to GNSS signals.
The overview of the framework is shown in Figure 2. First, synchronized image timestamps are used to recut the point cloud data stream, producing LiDAR point cloud frames at epochs matching the images. Then, the LiDAR-inertial and stereo images are processed separately by FAST-LIO2 and COLMAP-SLAM odometry methods, respectively, generating metric relative pose factors. Alongside the change in stereo camera poses, the latter outputs the 2D image feature locations, their 3D triangulations and characteristics. The features are tracked in time; if a tie point is seen in a sufficient number or frames, it is deemed stable enough to be used as a landmark in the factor graph. To do so, the image projection factors are input into the graph independently for each camera. The approximation of the 3D position of such landmark is estimated with COLMAP-based triangulation and refined later in the graph optimization.
For the images in which such a salient feature is present, the LiDAR 3D points are transformed to the camera reference frame and projected into the image. If such point lies within a pixel distance smaller than a set threshold (to accommodate for the sparsity of the 3D data), its depth is additionally used as a range factor, from the camera to the 3D landmark.
Finally, as the relative pose, tie point projection and landmark distance factors are input into the factor graph, the final estimate of the system’s pose is calculated. A priori covariance values are used for the odometry values and robust kernels (e.g., Cauchy) are used for visual features and range factors to limit the influence of outliers. The optimization problem is solved with Powell's dog leg method.

3.2.3 Experimental setup

A test survey has been carried out to validate the improvement brought by the proposed method. A set of indoor corridors of the office building was traversed in an eight-shaped loop while holding the STELVIO system by the operator. The cameras were pointing forward, thus the LiDAR field of view was mostly limited to the side walls of the narrow corridors. Such a scenario constitutes a challenging sequence for both the LiDAR-based odometry, due to the presence of constant geometry of the surrounding, as well as to the visual component, thanks to the commonly featureless wall textures. The data collection was initialized and finished at the same spot (a ground control point), so that the drift accumulated through the full, approximately 200 m sequence could be assessed. The dataset consisted of synchronized stereo image sequences at 2 Hz, LiDAR point clouds at approximately 10 Hz and inertial readings at 200 Hz. Internal parameters of the cameras, as well as relative orientation of the LiDAR, IMU and both cameras, were determined with kalibr multi-camera and camera-IMU calibration procedures (Furgale et al., 2013; Rehder et al., 2016) and fixed throughout all the processing. A pinhole camera model with radial-tangential distortion was used for both cameras.

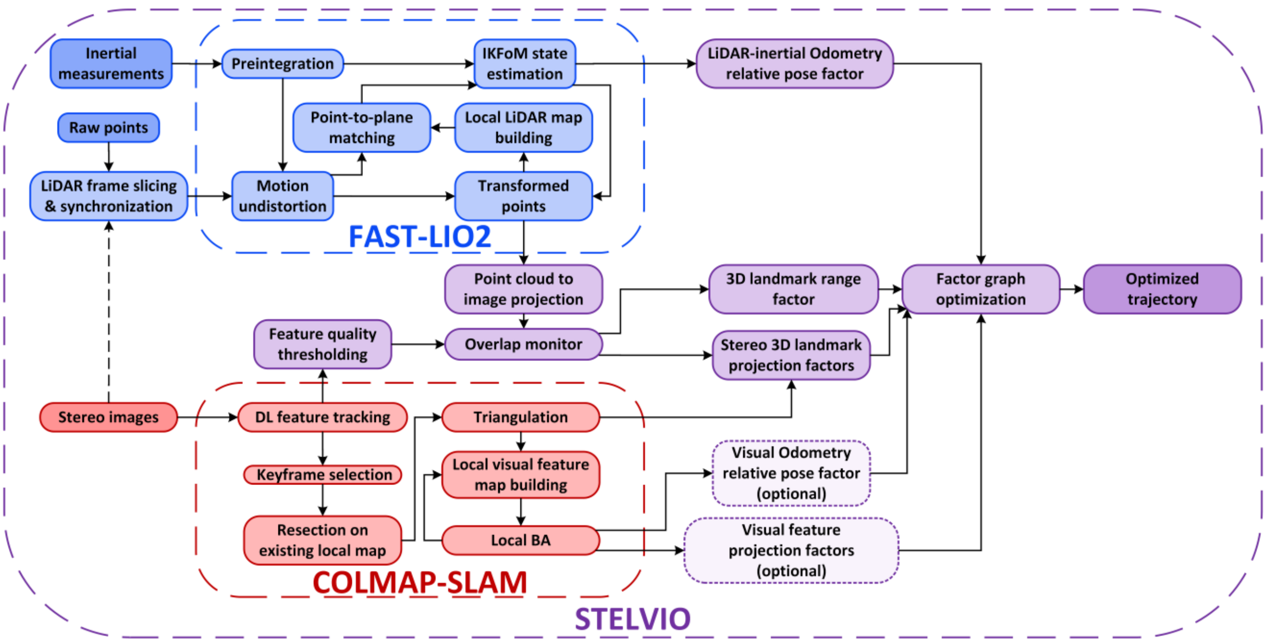

Figure 2: Overview of the STELVIO framework.

In this work, we investigated different complexity levels of graph-based approaches in STELVIO for fusing LiDAR and visual data. Such adjustment can balance computation complexity to the requirements related to the time constrains, like real-time applications, or increase robustness achieved through greater data redundancy. Thus, factor graph-based fusion was performed at different levels:

- Pose graph – including only two independent odometry estimates
- Simplified Factor Graph – two independent odometry estimates aided by sparse high-quality 3D visual-LiDAR and visual-only landmarks
- Full Factor Graph – two odometry trajectories aided by numerous visual-LiDAR landmarks and visual-only landmarks.

This way, we compare how the quality and quantity of landmarks used in the optimization affects the results. As a proxy of the factor quality, we selected the track length of tie points used by the stereo visual odometry frontend.

4. RESULTS AND DISCUSSION

First, the test sequence was processed with separate components of STELVIO, i.e., LiDAR-inertial and stereo-visual odometry methods, as well as fusion methods: STELVIO and FAST-LIVO2. All obtained trajectories are presented in Figure 3.

Both odometry methods produced lacking results. Namely, the former solution struggled with insufficient distinct geometry of the plain corridors, multiple times wrongly estimating the rotation at turns and distances at the straight parts of the corridor (Figure 3). The latter performed better, but suffered from a significant vertical drift. As indicated in the Table 1, these resulted in the total position error at the end of the sequence equal to 6.44 m and 1.90 m, respectively. Relative errors, i.e., the position residual divided by the traversed distance, were equal to 3.6% and 1.1%. The baseline fusion method, FAST-LIVO2, failed to recover the useful trajectory. Despite performing reasonably well in some parts of the sequence, the solution drifted substantially at several locations.

Finally, STELVIO was applied to the same dataset in several configurations. In its best one, it reached the position residuals of only 89 cm (0.5% of the trajectory length). Moreover, all XYZ components of the final drift of STELVIO were lower than their counterparts obtained using FAST-LIO2 and COLMAP-SLAM – the input odometry to STELVIO. This underlines the improvement of the trajectory estimation quality brought by the proposed fusion method.

The ablation study was carried out to examine the effect of visual feature quality on the STELVIO performance. The dataset was processed limiting the visual features to those with minimal track lengths of: 27 (a single, longest-tracked landmark), 25, 20, 15 and 10. Additionally, only stereo visual and LiDAR-inertial odometry outputs, without any tie point projection factors, were fused together. The results are shown in Table 1 and the trajectories are plotted in Figure 4. The lowest drift was achieved by the setup with a single landmark, i.e., the most rigorous feature filtering. Then, using more features slowly, but increasingly degraded the solution. The results of setting the threshold at the minimal track length of 10 resulted in using actually less features, than for the previous threshold, since more projections were filtered out as outliers. Consequently, this run of STELVIO resulted in the drift minimally worse than the odometry-only configuration. Notably, all of the results achieved with STELVIO surpassed the loop closure error of other baseline methods, with the COLMAP-SLAM stereo solution being the closest one and reaching 213% of the best STELVIO run error value and 132% error of the worst result of STELVIO.

The 3D reconstructions of the site, generated through combining the LiDAR scans with different sources of the mobile mapping system trajectory, are displayed in Figure 5. For clarity, the LiDAR range was limited to 5 meters. The coregistered point cloud of STELVIO shows visibly greater coherence than the results of FAST-LIO2. Although the COLMAP-SLAM-based solution demonstrates remarkably good quality, on the parts with trajectory overlaps, as well as at the start and end area, STELVIO outperformed vision-only approach in terms of the local point cloud coherence.

Thanks to the synchronization of the sensors of STELVIO and its precalibration, the adjusted system trajectory can be used for coloring the point cloud. A simple direct method was tested: projecting the RGB colors from the images oriented in the global reference frame to the coregistered point cloud. For the points visible in multiple images, only the closest projection was kept. An example part of the point cloud of the examined indoor environment is shown in Figure 6. Although the texturing matches the general shapes visible in the geometry, some artifacts, most likely related to the sensor scanning pattern, are visible. Thus, further work on the optimal point cloud coloring approach, incorporating multiple projections per point, is needed.

Method	Minimal visual feature track	No . of visual feature factors	No. of distance factors	Loop closure error [m]			
				X	Y	Z	Linear
COLMAP-SLAM	-	-	-	0.16	1.52	1.68	1.90
FAST-LIO2	-	-	-	6.23	1.56	0.39	6.44
FAST-LIVO2	-	-	-	2.59	45.42	13.17	47.36
STELVIO	-	0	0	0.67	1.13	0.11	1.32
	27	27	5	0.05	0.88	0.15	**0.89**
	25	77	5	0.04	1.00	0.12	1.00
	20	249	29	0.29	1.06	0.23	1.12
	15	744	99	0.04	1.15	0.40	1.22
	10	426	59	0.67	1.26	0.00	1.43

Table 1. Results of all tested approaches on the closed-loop indoor sequence.

https://doi.org/10.5194/isprs-archives-XLVIII-1-W5-2025-125-2025

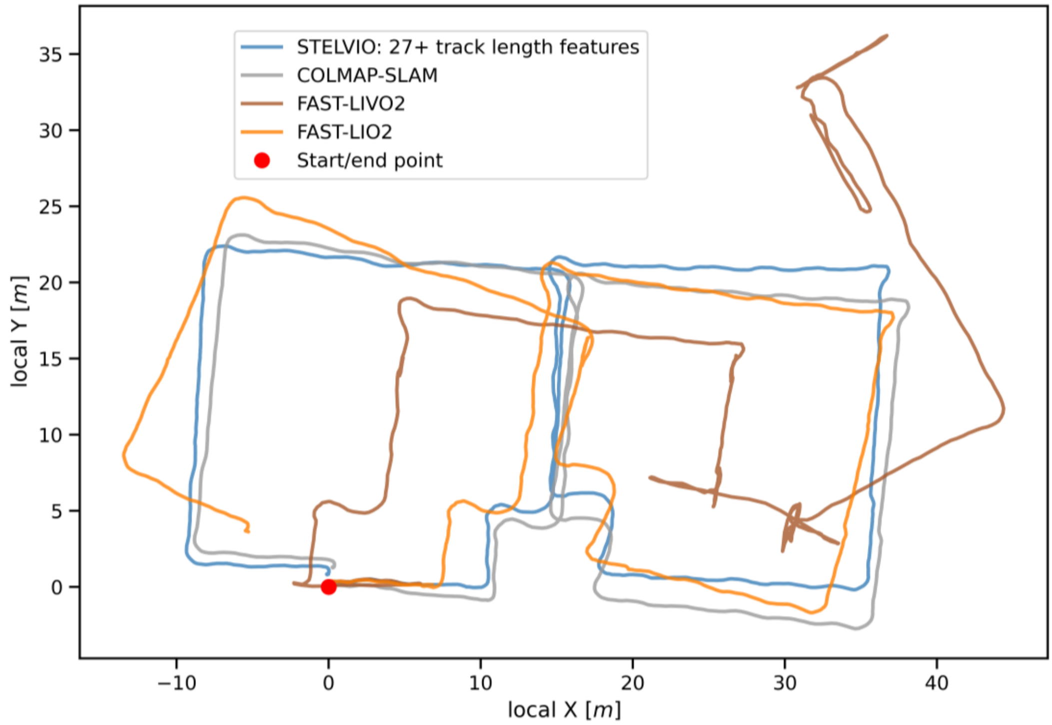

Figure 3: Estimated mobile mapping system trajectories: stereo COLMAP-SLAM (grey), LiDAR-inertial FAST-LIO2 (orange), LiDAR-inertial-visual FAST-LIVO2 (brown) and STELVIO (blue).

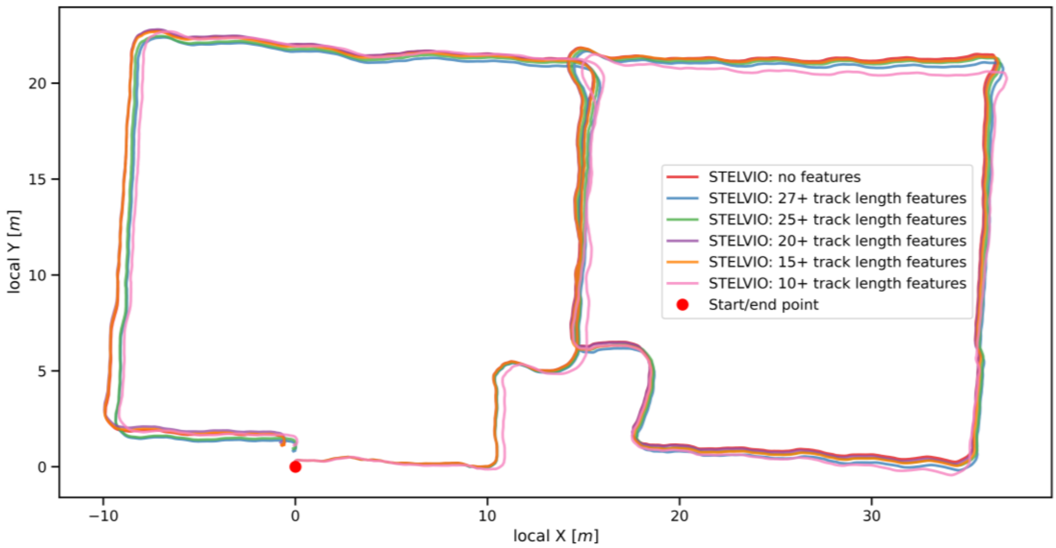

Figure 4: Estimated mobile mapping system trajectories with different STELVIO configurations.

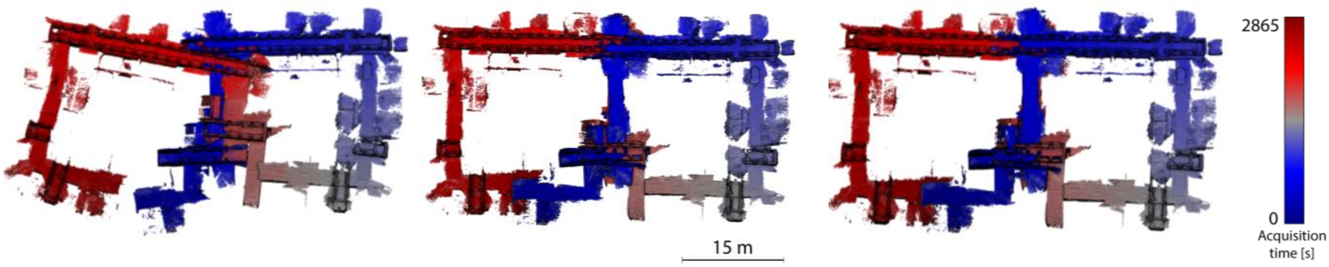

Figure 5: Orthographic views from the top of the 3D reconstructions using LiDAR scans and a trajectory obtained with: LiDAR-inertial odometry (left), stereo visual odometry (middle) and fused stereo visual-LiDAR-inertial data (right). Points colored by the time of acquisition, from blue to red.

https://doi.org/10.5194/isprs-archives-XLVIII-1-W5-2025-125-2025 |

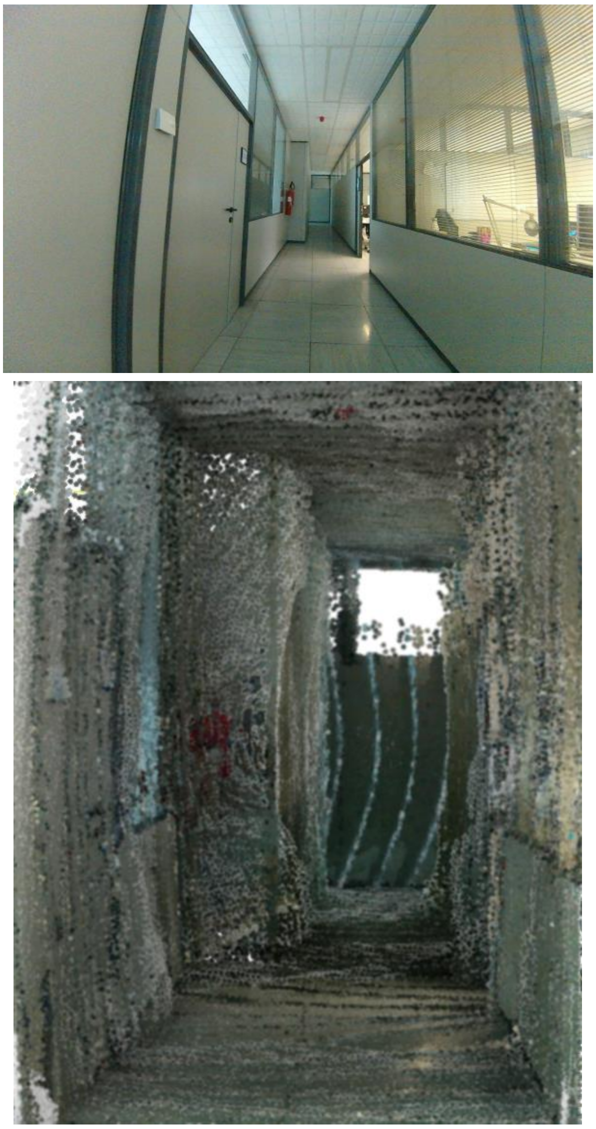

Figure 6: An image captured with one of the STELVIO cameras (top) and a zoom-in into the corresponding area of the colored point cloud obtained with STELVIO output (bottom).

5. CONCLUSIONS

This study presents STELVIO, a novel low-cost mobile mapping system which uses a set of sensors consisting of a stereo camera pair, a 360-degree LiDAR and its internal IMU. Along the dedicated in-house built hardware, the paper has introduced the software framework to perform joint trajectory adjustment, leveraging the strengths of the included active and passive sensors. Our approach, based on factor graphs, is designed as a plug-and-play framework for easier extendibility by future novel odometry approaches. In this study, we utilized some state-of-the-art LiDAR- and vision-oriented methods, FAST-LIO2 and COLMAP-SLAM, as backbones of STELVIO. The first field tests of mapping an indoor, office environment demonstrated promising capabilities of STELVIO. In the closed-loop test of a challenging indoor scenario, STELVIO outperformed all other tested odometry methods. Due to such good performance in the narrow, textureless corridors, we believe the system can generalize well to environments with similar characteristics, like underground mining tunnels or cave systems.

Tests of different configurations of our method were carried out, exploring different complexity levels of the definition of the factor graph. The results highlighted the importance of the reliable assessment of the quality of visual features, as it can greatly affect the benefit of incorporating more factors in the graph-based adjustment. To this end, we will explore a higher range of visual landmark characteristics for selecting the most salient observations.

Apart from the improvement in the trajectory estimation accuracy, applying our fusion framework allows to directly texture the 3D point clouds from the LiDAR with the image data. As in this study only an initial insight into the quality of the trajectory estimation of the method is investigated, future works will focus also on the time-wise performance of the odometry for real-time applications in longer sequences. Moreover we will extend STELVIO with loop closure capabilities to form a well-rounded, robust and universal fusion-based SLAM system. Additionally, enhancement in the quality of the point cloud texturing will be sought after.

ACKNOWLEDGEMENTS

This work has been partially supported by the EIT-RM project VOT3D - Ventilation Optimizing Technology based on 3D scanning (https://vot-3d.com).

REFERENCES

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

Dai, K., Sun, B., Wu, G., Zhao, S., Ma, F., Zhang, Y., & Wu, J. (2023). Lidar-based sensor fusion slam and localization for autonomous driving vehicles in complex scenarios. *Journal of Imaging*, 9(2), 52.

Fan, Z., Zhang, L., Wang, X., Shen, Y., Deng, F., 2025. LiDAR, IMU, and camera fusion for simultaneous localization and mapping: a systematic review. *Artificial Intelligence Review*, 58(6), 1-59.

Fasiolo, D. T., Scalera, L., Maset, E., 2023. Comparing LiDAR and IMU-based SLAM approaches for 3D robotic mapping. *Robotica*, 41(9), 2588-2604.

Furgale, P., Rehder, J., Siegwart, R., 2013. Unified temporal and spatial calibration for multi-sensor systems. Proc. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1280-1286.

Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K.M., Trulls, E., 2021. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, 129(2), pp. 517-547.

Lang, X., Chen, C., Tang, K., Ma, Y., Lv, J., Liu, Y & Zuo, X., 2023. Coco-LIC: Continuous-Time Tightly-Coupled LiDAR-Inertial-Camera Odometry Using Non-Uniform B-Spline. *IEEE Robotics and Automation Letters*, 8 (11), pp. 7074-7081.

Lin, J., Zhang, F., 2024. R^{3}3LIVE++: A Robust, Real-Time, Radiance Reconstruction Package With a Tightly-Coupled LiDAR-Inertial-Visual State Estimator. *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, 46 (12), pp. 1168-1185.

Lindenberger, P., Sarlin, P.E. and Pollefeys, M., 2023. Lightglue: Local feature matching at light speed. Proc. *ICCV*, pp. 17627-17638.

Lv, J., Lang, X., J, Xu, J., Wang, M., Liu, Y., Zuo, X., 2023. Continuous-Time Fixed-Lag Smoothing for LiDAR-Inertial-Camera SLAM. *IEEE/ASME Transactions on Mechatronics*, 28 (4), pp. 2259-2270.

Morelli, L., Ioli, F., Beber, R., Menna, F., Remondino, F. and Vitti, A., 2023. COLMAP-SLAM: A framework for visual odometry. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, pp. 317-324.

Muhojoki, J., Tavi, D., Hyyppä, E., Lehtomäki, M., Faitli, T., Kaartinen, H., Kukko, A., Hakala, T., Hyyppä, J., 2024. Benchmarking Under-and Above-Canopy Laser Scanning Solutions for Deriving Stem Curve and Volume in Easy and Difficult Boreal Forest Conditions. *Remote Sensing*, 16(10), 1721.

Rehder, J., Nikolic, J., Schneider, T., Hinzmann, T., Siegwart, R., 2016. Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4304-4311.

Roman, O., Farella, E. M., Rigon, S., Remondino, F., Ricciuti, S., Viesi, D., 2023. From 3D surveying data to BIM to BEM: The InCUBE dataset. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, pp. 175-182.

Shan, T., Englot, B., Ratti, C., Rus, D., 2021. Lvi-sam: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping. *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5692–5698.

Trybała, P., Rigon, S., Remondino, F., Banasiewicz, A., Wróblewski, A., Macek, A., Kujawa, P., Romańczukiewicz, K., Redondo, C., Espada, F., 2024. Optimizing Mining Ventilation Using 3D Technologies. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, pp. 427-434.

Trybała, P., Szrek, J., Remondino, F., Kujawa, P., Wodecki, J., Blachowski, J., & Zimroz, R. (2023). MIN3D Dataset: MultI-seNsor 3D Mapping with an Unmanned Ground Vehicle. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 91(6), pp. 425-442.

Vassena, G. P., Perfetti, L., Comai, S., Mastrolembo Ventura, S., Ciribini, A. L., 2023. Construction progress monitoring through the integration of 4D BIM and SLAM-based mapping devices. *Buildings*, 13(10), 2488.

Wajs, J., Trybała, P., Górniak-Zimroz, J., Krupa-Kurzynowska, J., Kasza, D., 2021. Modern solution for fast and accurate inventorization of open-pit mines by the active remote sensing technique—case study of mikoszów granite mine (lower silesia, sw poland). *Energies*, 14(20), 6853.

Xu, W., Cai, Y., He, D., Lin, J., Zhang, F., 2022. Fast-lio2: Fast direct lidar-inertial odometry. *IEEE Transactions on Robotics*, 38(4), pp. 2053-2073.

Xu, W., Zhang, F., 2021. Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter. *IEEE Robotics and Automation Letters*, 6(2), pp. 3317-3324.

Yuan, Z., Deng, J., Ming, R., Lang, F., Yang, X., 2024. Sr-livo: Lidar-inertial-visual odometry and mapping with sweep reconstruction. *IEEE Robotics and Automation Letters*, 9, 6, pp. 5110-5117,

Yuan, Z., Wang, Q., Cheng, K., Hao, T., Yang, X., 2023. SDV-LOAM: Semi-direct visual–LiDAR Odometry and mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9), pp. 11203-11220.

Zhao, X., Wu, X., Chen, W., Chen, P.C., Xu, Q. and Li, Z., 2023. Aliked: A lighter keypoint and descriptor extraction network via deformable transformation. *IEEE Transactions on Instrumentation and Measurement*, *72*, pp.1-16.

Zheng, C., Xu, W., Zou, Z., Hua, T., Yuan, C., He, D., ... & Zhang, F., 2025. Fast-livo2: Fast, direct lidar-inertial-visual odometry. *IEEE Transactions on Robotics*, 41, pp. 326-346.

Zhou, S., Wang, Z., Dai, X., Song, W., Gu, S., 2025. LIR-LIVO: A Lightweight, Robust LiDAR/Vision/Inertial Odometry with Illumination-Resilient Deep Features. *arXiv preprint* arXiv:2502.08676.