Passage-Aware Structural Mapping for RGB-D Visual SLAM

Ali Tourani[1], Miguel Fernandez-Cortizas[1], Saad Ejaz[1], David Pérez Saura[2], Asier Bikandi-Noya[1], Jose Luis Sanchez-Lopez[1], and Holger Voos[1,3]

***Abstract*—Doorways and passages are critical structural elements for indoor robot navigation, yet they remain underexplored in modern Visual SLAM (VSLAM) frameworks. This paper presents a passage-aware structural mapping approach for RGB-D VSLAM that detects doors and traversable openings by jointly fusing geometric, semantic, and topological cues. Doors are modeled as planar entities embedded within walls and classified as traversable or non-traversable based on their coplanarity with the supporting wall. Passages are inferred through two complementary strategies: traversal evidence accumulated from camera–wall interactions across consecutive keyframes, and geometric opening validation based on discontinuities in the mapped wall geometry. The proposed method is integrated into vS-Graphs as a proof of concept, enriching its scene graph with passage-level abstractions and improving room connectivity modeling. Qualitative evaluations on indoor office sequences demonstrate reliable doorway detection, and the framework lays the foundation for exploiting these elements in BIM-informed VSLAM. The source code is publicly available at https://github.com/snt-arg/visual_sgraphs/tree/doorway_integration.**

***Index Terms*—Visual SLAM, Semantic SLAM, Prior Maps, Situational Awareness**

I. Introduction & Related Works

Simultaneous Localization and Mapping (SLAM) has emerged as a fundamental capability of modern autonomous robots, allowing them to estimate their pose while incrementally reconstructing the surrounding environment [1]. Such a capability is crucial for robotic situational awareness [2] and encompasses a broad range of tasks, including navigation, exploration, and inspection. Among the available sensing modalities in SLAM, vision sensors provide a cost-effective means of capturing rich visual and structural data, leading to the emergence of Visual SLAM (VSLAM) [3]. With the considerable progress in VSLAM, retrieving geometric information (*e.g.,* 3D points) is often insufficient for understanding the environment, despite its impact on localization and basic mapping. Accordingly, modern VSLAM approaches also incorporate the **semantic** and **structural information** of mapped entities, such as walls, chairs, and tables [4], [5].

[1] Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg, Luxembourg. Holger Voos is also affiliated with the Faculty of Science, Technology, and Medicine at the University of Luxembourg, Luxembourg. {ali.tourani, miguel.fernandez, saad.ejaz, asier.bikandi, joseluis.sanchezlopez, holger.voos}@uni.lu

[2] Universidad Politécnica de Madrid, Spain. david.perez.saura@upm.es

[3] Faculty of Science, Technology, and Medicine at the University of Luxembourg, Luxembourg.

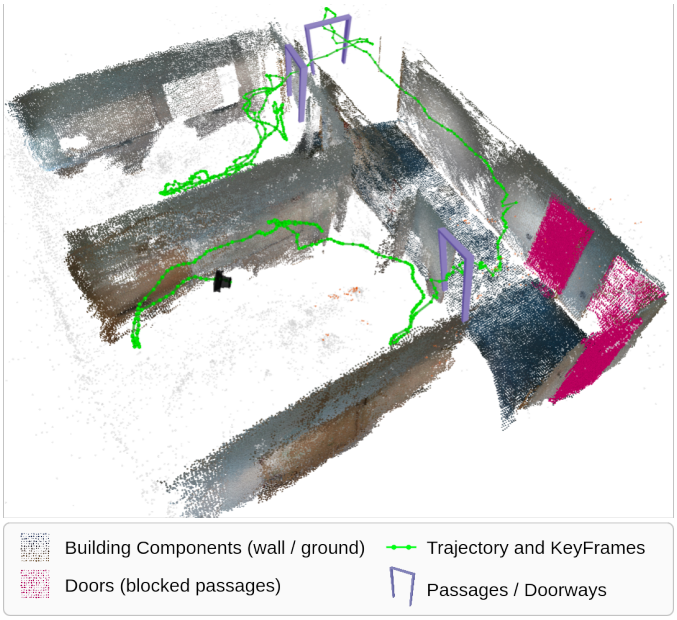


Fig. 1: Overview of door and passage mapping within vS-Graphs [9], where semantic and geometric cues jointly localize traversable openings, enabling robust downstream tasks such as navigation and path planning.

Augmenting VSLAM with semantic information enables more interpretable and structurally meaningful map reconstruction. In this direction, Tao *et al.* [6] integrate semantic perception and mapping to generate trajectories that maximize information gain with uncertainty reduction. Despite the rich map reconstructions, the method introduces considerable system complexity and limited scalability. SemanticFusion [7] produces temporally consistent semantic maps by probabilistically fusing predictions from multiple viewpoints at the expense of high computational cost. Khronos [8] employs semantic segmentation to identify short-term and long-term scene dynamics; however, its primary focus remains on dynamic objects rather than structural scene understanding.

Other approaches emphasize semantic representations of the environment layout, aiming to construct richer, more interpretable maps that include architectural entities such as walls and rooms. vS-Graphs [9] proposes a tightly coupled framework that jointly optimizes VSLAM and 3D scene graph generation within a unified factor graph. Incorporating *build-*

ing components (*i.e.,* walls and ground surfaces), *structural elements* (*i.e.,* rooms and floors), and their spatial relations produces geometrically grounded and semantically coherent environment representations. However, the framework does not model *doorways*, largely due to the absence of dense point cloud information. In [10], doorway detection and mapping are achieved using fiducial markers placed on door frames. However, the approach relies on prior environmental instrumentation, which limits its practicality and scalability.

It should be noted that passages are critical components of indoor spaces and have a special SLAM signature, as they define **traversable openings** between enclosed regions (*e.g.,* rooms) and establish their inter-connectivity. From a robotic perspective, passages are topologically meaningful landmarks that indicate transitions between navigable spaces. Their detection extends room modeling beyond wall-only enclosures toward a richer structural-topological abstraction, supporting downstream tasks such as situationally aware navigation [14]. Furthermore, incorporating prior knowledge from **Building Information Modeling (BIM)** can provide complementary structural constraints, enabling more reliable validation of passages. Such priors enhance robustness in partially observed environments by aligning reconstructed geometry with known architectural layouts. To address this gap, this paper makes the following contributions:

- Formulating passage detection as traversable openings in walls by jointly fusing geometric, semantic, and topological cues.
- Integrating passages into a VSLAM system (vS-Graphs [9] as a proof of concept, extendable to other backbones), enabling improved room connectivity modeling and environment understanding.
- Publicly available source code to facilitate reproducibility and further research in the field.

II. Proposed Method

A. Formalism

Given an RGB-D point cloud at the VSLAM KeyFrame level, a panoptic segmentation method such as YOSO [11] can be used to extract semantically meaningful planar entities, including walls and doors [9]. Each KeyFrame is defined as $K_t = \{\mathbf{P}_t, L_t, \varepsilon_t\}$, where $\mathbf{P}_t = \{\mathbf{p}_i \in \mathbb{R}^3\}$ denotes the 3D point cloud, L_t the RGB image, and ε_t the associated mapping metadata (e.g., camera pose). Applying panoptic segmentation to L_t yields pixel-wise semantic labels and instance-level masks, which are projected onto $\mathbf{P}_t$ to obtain semantically segmented point subsets $\mathbf{P}_\Psi^{K_t} = S(\mathbf{P}_t, L_t) = \{\mathbf{P}_{wall}^{K_t}, \mathbf{P}_{door}^{K_t}, \dots\}$, where each $\mathbf{P}_\psi^{K_t} \subset \mathbf{P}_t$ corresponds to a semantic class ψ. To reduce redundancy and sensor noise, each segmented subset is refined through downsampling and range-based filtering, and then processed with RANSAC plane fitting [13] to estimate semantically validated planar entities. For each subset $\mathbf{P}_\psi^{K_t}$, a plane $\pi = (\mathbf{n}, d)$ is estimated, where $\mathbf{n} \in \mathbb{R}^3$ is the normal and $d \in \mathbb{R}$ the plane offset, such that a point $\mathbf{p}_r \in \mathbf{P}_\psi^{K_t}$ is considered an inlier if $\mathrm{dist}(\mathbf{p}_r, \pi) \leq \epsilon$, with ϵ denoting the inlier threshold. The validated planar entities detected in KeyFrame K_t are subsequently inserted into the map and used for continuous structural reconstruction. Among these semantic entities, locating **doors** and **passages** (*e.g.,* doorways, archways, *etc.*) is particularly valuable for downstream robotic tasks such as navigation and exploration.

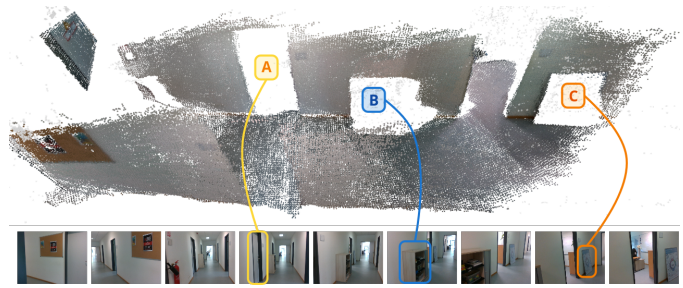


Fig. 2: Examples of geometric openings detected as gaps in semantic point clouds generated by vS-Graphs [9] across sequential frames, arising from (A) a door, (B) a storage cabinet, and (C) a poster on the ground, highlighting the ambiguity of purely geometric cues and the need for contextual validation.

1) Doors: A door $\mathbf{P}_{door}^{K_t} \subset \mathbf{P}_\Psi^{K_t}$ is modeled as a *physical planar structural object* extracted from RGB-D observations. Since doors are typically embedded within walls, their *traversability* can be inferred from their geometric relation to the associated supporting wall. Let $\pi_d = (\mathbf{n}_d, d_d)$ and $\pi_w = (\mathbf{n}_w, d_w)$ be the estimated door and wall planes, respectively. The door is "*closed/non-traversable*" if it is approximately coplanar with its supporting wall:

$$\cos^{-1}\left(|\mathbf{n}_d^\top \mathbf{n}_w|\right) < \tau_\theta \quad \text{and} \quad |d_d - d_w| < \tau_d. \tag{1}$$

where τ_θ and τ_d are predefined thresholds for angular alignment and coplanarity, respectively. Satisfying both conditions indicates that the detected door lies approximately on the same plane as its supporting wall.

2) Passages: A passage is defined as $Q_i = \{\mathbf{P}_{wall}, \rho_i, \eta_i, m\}$, representing an opening (potentially blocked) embedded within a wall. Here, $\rho_i \in \mathbb{R}^3$ denotes the passage centroid in the global reference frame, $\eta_i \in \mathbb{R}^2$ encodes its geometric extent (*e.g.,* width and height), and m stores its semantic variant (*e.g.,* `doorway`, `archway`, or `unknown`). When a closed door instance $\mathbf{P}_{door}$ is detected, a corresponding passage is instantiated at the same location and classified as a `doorway`, with its dimensions inferred from the associated door geometry. However, not all passages correspond to doors, and detecting generic traversable openings remains essential for tasks such as navigation and planning. Accordingly, passage candidates are identified using two complementary strategies, as described below:

I. Traversal evidence-based approach: passages are inferred from geometric and observational evidence across multiple KeyFrames by analyzing the interaction between the *camera trajectory* and the *mapped wall geometry*. Having $\mathbf{c}_t$ as the

camera center at KeyFrame K_t and $\pi_w = (\mathbf{n}_w, d_w)$ as the supporting wall plane, the signed distance of $\mathbf{c}_t$ to the wall is $s_t = \mathbf{n}_w^\top \mathbf{c}_t + d_w$. To ensure locality, only KeyFrames within a bounded distance to the wall ($d_{\max}$) are considered, *i.e.,* $|s_t| < d_{\max}$. A *wall traversal event* is detected when two consecutive KeyFrames K_t and K_{t+1} exhibit a sign change in their signed distances, as $s_t \cdot s_{t+1} < 0$, indicating that the camera trajectory crosses the wall plane. Such traversal events provide strong evidence of a *local geometric opening* in the wall, as a true wall surface would otherwise prevent visibility and motion across it. To improve robustness, traversal events are evaluated within a sliding temporal window over ordered KeyFrames, ensuring *locally consistent* detections. For each valid traversal event, the crossing point is estimated by linear interpolation between $\mathbf{c}_t$ and $\mathbf{c}_{t+1}$, yielding a candidate passage with centroid ρ_i. If a door instance $\mathbf{P}_{door}$ is detected in proximity to a candidate passage centered at ρ_i, the passage is classified as a `doorway`, associated with the detected door, and its dimensions are inferred directly from the door geometry. Otherwise, the passage remains `unknown`, and a default extent is assigned (*e.g.,* $1.5\,\mathrm{m} \times 2\,\mathrm{m}$).

II. Geometric opening validation approach: in contrast to the local traversal-based strategy, passage candidates can also be inferred from a *global analysis* of the reconstructed mapped wall geometry by explicitly searching for *geometric discontinuities* or gaps. Given a wall plane π_w and its associated inlier point cloud $\mathbf{P}_{wall}$, the wall is projected onto a local 2D parameterization, where openings are identified as connected regions $\Omega_j \in \mathbb{R}^2$ with insufficient point support, *i.e.,* $\forall \mathbf{u} \in \Omega_j,\ \mathcal{N}(\mathbf{u}) < \tau_\rho$, where $\mathcal{N}(\cdots)$ denotes local point density. Each candidate region (*i.e.,* gap) is then back-projected to the 3D reference, with a centroid ρ_i and extent η_i. Since such gaps may correspond to various entities (*e.g.,* doors, windows, archways, or unmapped semantic objects near walls), as shown in Fig. 2, a validation step is required to further filter out non-passage candidates. This can be performed by evaluating *geometric consistency* (*e.g.,* size and aspect ratio of the gaps) and *contextual cues* such as proximity to a detected door instance $\mathbf{P}_{door}$. Only openings satisfying these constraints are retained as passages, while ambiguous regions are deferred until further evidence is accumulated across observations.

B. Implementation & Deployment

The proposed passage detection framework has been integrated into vS-Graphs [9], a real-time visual SLAM pipeline that enables online extraction of semantically verified structural elements. This extension enables the extraction of additional *geometric* and *semantic cues* beyond those originally modeled in vS-Graphs (i.e., walls and ground surfaces). In particular, it introduces higher-level abstractions that explicitly encode connectivity through passages within the environment. Such enriched representations are beneficial for downstream tasks (*e.g.,* navigation and planning) that rely on scene understanding (such as situationally aware path planning [14]), where reasoning about *traversability* is essential.

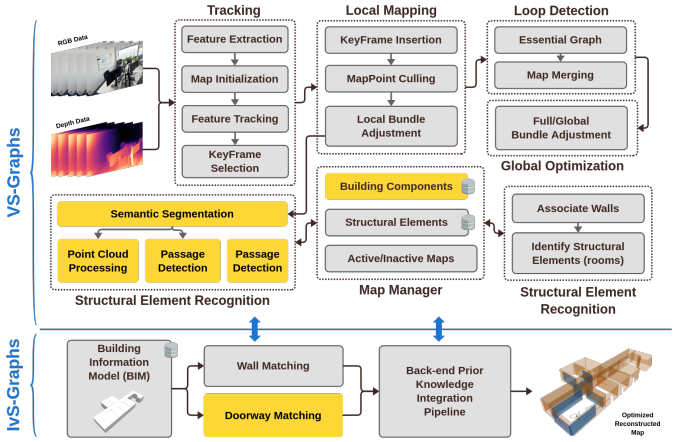
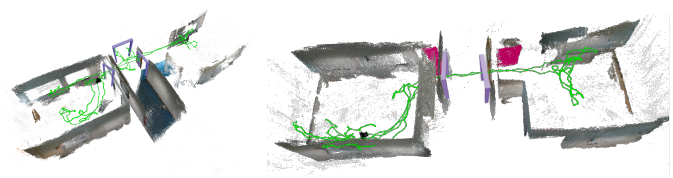


Fig. 3: System architecture showing the integration of passage detection within vS-Graphs [9] and its potential extension with structural priors in ivS-Graphs [12] for improved scene understanding. Bordered, light gray modules indicate baselines' components, while highlighted modules denote the contributions of this paper.



(a) SMapper - Seq. MR01 (b) SMapper - Seq. MR04

Fig. 4: Qualitative results of the proposed passage detection and integration approach in indoor office environments, with vS-Graphs [9] as the baseline.

Integration within vS-Graphs is depicted in Fig. 3. Accordingly, the *building component recognition* thread is extended to support door detection by adding the door class to the semantic space (wall and ground). This enables the identification of *closed doors* within their surrounding wall context while mapping them. Additionally, passage-related cues are detected in the *structural element recognition* thread by periodically checking for openings in mapped walls. This contains both *traversal evidence* and *geometric validation* approaches introduced in Section II-A2. It should be noted that, as these operations are performed at the KeyFrame level, the system preserves its real-time performance while enriching the scene graph with passage-level abstractions.

III. Benchmarking & Evaluations

The proposed approach has been evaluated through qualitative analyses of dataset instances recorded in indoor office environments with the SMapper device [15]. The collected data is used to validate the effectiveness of the proposed passage detection pipeline. Experimental results demonstrate that the system reliably identifies passages, primarily doorways, using both *traversal evidence* and *geometric validation* approaches,

as illustrated in Fig. 4. In the figure, light purple doorway markers indicate open passages, while crimson point clouds correspond to closed or obstructed doorways. The results further highlight the consistency of the proposed method across varying spatial layouts and levels of environmental complexity.

IV. Potentials & Discussions

The proposed approach can be augmented with a BIM-informed extension of vS-Graphs, as introduced in [12] (titled ivS-Graphs), thereby incorporating prior structural knowledge into the mapping process. Such an extension enables the establishment of correspondences between the reconstructed scene and the as-planned BIM model, allowing detected passages to be more robustly validated and refined. This integration improves scene understanding by reducing geometric ambiguity and increasing the structural consistency and completeness of the reconstructed environment. As depicted in Fig. 3, the underlying vS-Graphs framework is naturally coupled with BIM information, facilitating more robust reasoning about structural elements. In particular, the availability of prior knowledge simplifies the validation of doorway existence and enhances confidence in passage detection, as further depicted in Fig. 5. Accordingly, the current version of ivS-Graphs enables the detection of structural deviations by establishing correspondences between *as-built* walls in the SLAM reconstruction and their *as-planned* counterparts. This has been effectively demonstrated for wall alignment, where inconsistencies can be identified and corrected using BIM priors. Building on this capability, the proposed framework extends this principle to *passages*, enabling detected doorways and openings to be validated against the BIM model. This facilitates the identification of missing, misaligned, or falsely detected passages, further improving the structural reliability of the reconstructed environment.

Additionally, and from an application perspective, enriching reconstructed maps with passages significantly enhances their utility for downstream robotic tasks. In particular, explicitly modeling doorways and traversable passages provides critical information for path planning and navigation, enabling robots to reason about connectivity between spaces. Such representations improve situational awareness by allowing more informed decision-making, especially in complex indoor environments where accessibility and transitions between regions are essential.

V. Conclusions

This paper introduced a passage-aware structural mapping approach for RGB-D Visual SLAM, formulating doorway detection as the identification of traversable openings in walls through the joint use of geometric, semantic, and topological cues. The method has been integrated into vS-Graphs and qualitatively evaluated on indoor office sequences, where it reliably distinguishes between open passages and closed or obstructed doorways while preserving real-time operation at the KeyFrame level. Beyond the current proof of concept, the framework is designed to be naturally extended through BIM priors, as outlined in ivS-Graphs [12], to validate detected passages against as-planned models and improve structural consistency. Future work will focus on incorporating doors and doorways directly within the factor graph optimization, enabling tighter coupling between traversability reasoning and pose estimation, as well as quantitative benchmarking across more diverse indoor environments.

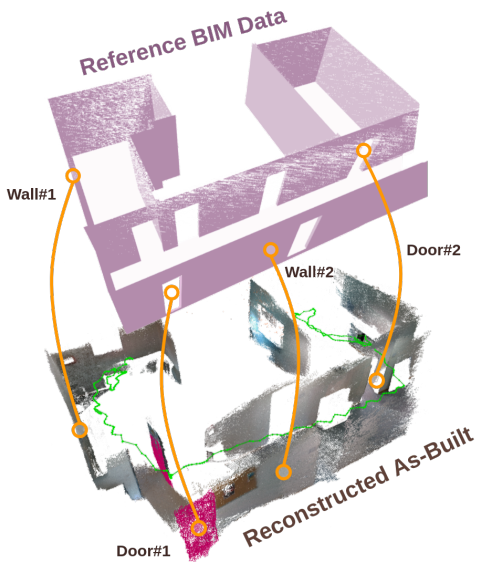


Fig. 5: Potential integration of the proposed passage detection methodology within the structural priors in ivS-Graphs [12] framework.

References

[1] A. Macario Barros, M. Michel, Y. Moline, G. Corre, and F. Carrel. "A Comprehensive Survey of Visual SLAM Algorithms," Robotics, vol. 11, no. 1, p. 24, 2022. https://doi.org/10.3390/robotics11010024

[2] H. Bavle, J.L. Sanchez-Lopez, C. Cimarelli, A. Tourani, and H. Voos, "From SLAM to Situational Awareness: Challenges and Survey," Sensors, vol. 23, no. 10, p. 4849, 2023. https://doi.org/10.3390/s23104849

[3] A. Tourani, H. Bavle, J.L. Sanchez-Lopez, and H. Voos, "Visual SLAM: What are the Current Trends and What to Expect?" Sensors, vol. 22, no. 23, p. 9297, 2022. https://doi.org/10.3390/s22239297

[4] L. Qin, C. Wu, Z. Chen, X. Kong, Z. Lv, and Z. Zhao, "RSO-SLAM: A Robust Semantic Visual SLAM with Optical Flow in Complex Dynamic Environments," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 10, pp. 14669-14684, 2024. https://doi.org/10.1109/TITS.2024.3402241

[5] G. Li, J. Cai, C. Huang, H. Luo, and J. Yu, "PS-SLAM: A Visual SLAM for Semantic Mapping in Dynamic Outdoor Environment using Panoptic Segmentation," IEEE Access, vol. 13, pp. 46534-46545, 2025. https://doi.org/10.1109/ACCESS.2025.3547002

[6] Y. Tao, X. Liu, I. Spasojevic, S. Agarwal, and V. Kumar, "3D Active Metric-Semantic SLAM," IEEE Robotics and Automation Letters, vol. 9, no. 3, pp. 2989–2996, 2024, https://doi.org/10.1109/LRA.2024.3363542.

[7] J. McCormac, A. Handa, A. Davison, and S. Leutenegger, "Semanticfusion: Dense 3D Semantic Mapping with Convolutional Neural Networks," IEEE International Conference on Robotics and Automation, pp. 4628–4635, 2017. https://doi.org/10.1109/ICRA.2017.7989538

[8] L. Schmid, M. Abate, Y. Chang, and L. Carlone, "Khronos: A Unified Approach for Spatio-Temporal Metric-Semantic SLAM in Dynamic Environments," arXiv preprint arXiv:2402.13817, 2024. https://doi.org/10.48550/arXiv.2402.13817

[9] A. Tourani, S. Ejaz, H. Bavle, M. Fernandez-Cortizas, D. Morilla-Cabello, J.L. Sanchez-Lopez, and H. Voos, "vS-Graphs: Tightly Coupling Visual SLAM and 3D Scene Graphs Exploiting Hierarchical Scene Understanding," ArXiv preprint arXiv:2503.01783, 2025. https://doi.org/10.48550/arXiv.2503.01783

[10] A. Tourani, H. Bavle, D.I. Avşar, J.L. Sanchez-Lopez, R. Munoz-Salinas, and H. Voos, "Vision-based Situational Graphs Exploiting Fiducial Markers for the Integration of Semantic Entities," Robotics, vol. 13, no. 7, p. 106, 2024. https://doi.org/10.3390/robotics13070106

[11] J. Hu, L. Huang, T. Ren, S. Zhang, R. Ji, and L. Cao, "You Only Segment Once: Towards Real-time Panoptic Segmentation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 17819–17829, 2023. https://doi.org/10.1109/CVPR52729.2023.01709

[12] A. Bikandi-Noya, M. Fernandez-Cortizas, M. Shaheer, A. Tourani, J.L. Sanchez-Lopez, and H. Voos, "BIM Informed Visual SLAM for Construction Monitoring," ArXiv preprint arXiv:2509.13972, 2025. https://doi.org/10.48550/arXiv.2509.13972

[13] O. Chum, and J. Matas, "Optimal Randomized RANSAC," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 8, pp. 1472–1482, 2008. https://doi.org/10.1109/TPAMI.2007.70787

[14] S. Ejaz, M. Giberna, M. Shaheer, J.A. Millan-Romera, A. Tourani, P. Kremer, H. Voos, and J.L. Sanchez-Lopez, "Situationally-aware Path Planning Exploiting 3D Scene Graphs," IEEE Robotics and Automation Letters, vol. 11, no. 3, pp. 3358 - 3365, 2026. https://doi.org/10.1109/LRA.2026.3656775

[15] P.M. Bastos Soares, A. Tourani, M. Fernandez-Cortizas, A. Bikandi-Noya, H. Voos, and J.L. Sanchez-Lopez, "SMapper: A Multi-Modal Data Acquisition Platform for SLAM Benchmarking," Journal of Intelligent & Robotic Systems, vol. 112, no. 20, 2026. https://doi.org/10.1007/s10846-026-02351-7