Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Sensor-Centric Survey of SLAM and Odometry for GPS-Denied Environments

Aristeidis Geladaris
Hellenic Mediterranean University
Panagiotis Polygerinos
polygerinos@uniwa.gr
University of West Attica

Systematic Review

Keywords: SLAM, odometry, sensor fusion, multi-sensor SLAM, GPS-denied, survey

Posted Date: September 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10409860/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Sensor-Centric Survey of SLAM and Odometry for GPS-Denied Environments

Aristeidis Geladaris[1] and Panagiotis Polygerinos[1,2*]

[1]Control Systems and Robotics Laboratory, Department of Mechanical Engineering, School of Engineering, Hellenic Mediterranean University, Heraklion, 71004, Greece.

[2]Department of Industrial Design and Production Engineering, School of Engineering, University of West Attica, Athens, 12241, Greece.

*Corresponding author(s). E-mail(s): polygerinos@uniwa.gr;
Contributing authors: ddk177@edu.hmu.gr;

Abstract

Simultaneous localization and mapping (SLAM) and odometry are the principal means by which autonomous platforms localize in environments where satellite positioning is unavailable, such as subterranean, indoor, underwater, and obstructed spaces. Because localization performance in such environments is fundamentally constrained by the available sensing modalities, this review organizes the literature by sensing configuration rather than algorithmic technique. We survey six categories: visual, visual-inertial, LiDAR, LiDAR-inertial, LiDAR-visual-inertial, and a heterogeneous group comprising sonar, radar, and thermal sensing. Within each modality, methods are grouped into four maturity tiers: foundational, established, recent, and frontier. To ensure that the field-defining and very recent works are weighed on comparable terms, these tiers are selected via maturity-appropriate criteria, with temporal boundaries calibrated per modality rather than fixed to common calendar years. Furthermore, the review compares datasets, evaluation metrics, and the accuracy-robustness-resource trade-offs that govern modality choice. Three key movements recur across all configurations: progressive multi-sensor fusion, the integration of learning within otherwise geometric pipelines, and a migration toward neural and differentiable map representations. We conclude that no configuration is universally optimal. Robust GPS-denied localization depends on combining sensing modalities with complementary failure characteristics while employing estimators capable of quantifying and adapting to their own uncertainty. Finally, we identify fragmented, degradation-agnostic benchmarking as the principal obstacle to rigorous cross-modal evaluation and sustained progress in the field.

Keywords: SLAM, odometry, sensor fusion, multi-sensor SLAM, GPS-denied, survey

1 Introduction

Simultaneous localization and mapping (SLAM) is the problem of estimating a platform's trajectory while concurrently building a map of an unknown environment. It is a foundational capability for autonomous systems, underpinning navigation, planning, and interaction wherever no prior map is available. In many of the most demanding of these settings, a global positioning reference is absent or unreliable: subterranean mines and tunnels, building interiors, underwater, dense forest canopy, and planetary surfaces all lie beyond the reach of the Global Navigation Satellite System (GNSS), as do disaster sites and adversarial environments where signals are obstructed or jammed. In such GPS-denied[1] conditions the burden of localization falls entirely on onboard sensing, and SLAM becomes not a convenience but the principal means by which a platform can determine where it is and what surrounds it. This survey reviews SLAM and odometry methods through the lens of GPS-denied operation, organizing them by the sensing configuration on which they rely.

Two closely related problems must be distinguished at the outset, as the distinction defines the scope of this review. Odometry estimates a platform's incremental ego-motion and, in most modern formulations, a local map by registering successive sensor measurements. Because errors compound, odometry estimates accumulate unbounded drift along the trajectory. SLAM augments this front-end with mechanisms for recognizing previously visited places (loop closure) and for jointly optimizing the trajectory and map, thereby bounding drift and yielding a globally consistent reconstruction. Although the term SLAM properly denotes only the latter, the boundary is porous in practice: many widely deployed and influential systems are, strictly, odometry pipelines, or expose loop closure only as an optional module, and they occupy a continuum rather than discrete classes. This blurring is especially consequential under GPS-denied operation, where no external fix is available to reset accumulated error and the accuracy of the odometry estimate is frequently the dominant determinant of overall localization performance. We therefore adopt a deliberately inclusive scope, surveying both full SLAM systems and odometry-only methods under the umbrella term SLAM, and we note explicitly, where it bears on accuracy and evaluation, which systems provide global consistency and which estimate motion incrementally.

The performance of any such system is inseparable from the sensors that feed it, and this coupling is most acute precisely in the degraded conditions characteristic of GPS-denied environments, where each sensing modality fails in its own way. Cameras are compact and information-rich but degrade under low illumination, weak texture, motion blur, and abrupt changes in dynamic range. LiDAR yields accurate geometry but struggles where structure is self-similar or absent, e.g. long corridors or featureless tunnels, and is corrupted by airborne particulates, such as dust, fog, and smoke.

[1] We use the entrenched term GPS-denied for what is, strictly, GNSS-denied operation; the definition is made precise in Section 3.1.

Inertial measurement units are immune to these visual and geometric failures but drift rapidly when used alone. Acoustic sensing is the natural recourse underwater, where optical attenuation and turbidity render cameras ineffective. Millimeter-wave radar offers all-weather robustness at the cost of sparse and noisy returns and thermal imaging extends perception into darkness and visual obscurants. Because the choice and combination of sensors is thus the primary design decision in GPS-denied SLAM, we organize the surveyed methods by sensing configuration into six categories: visual, visual-inertial, LiDAR, LiDAR-inertial, LiDAR-visual-inertial, and a residual category encompassing acoustic (sonar), radar, and thermal sensing.

Within each category the literature spans nearly two decades, from early formulations to methods published within the past year, and a survey that presents them as a flat ranking by citation count inevitably privileges older work, since citations accrue only after citing papers are themselves written and published. To present the evolution of each modality fairly, we instead group methods into four maturity tiers: foundational, established, recent, and frontier. Within each tier we select by the criterion appropriate to its maturity, ranging from community consensus for canonical methods to publication venue, open-source adoption, and benchmark standing for frontier work whose citation record does not yet exist. This organization allows field-defining and very recent methods to be weighed on comparable terms. Because subfields emerged and matured at different times, the tier boundaries are permitted to shift across modalities rather than being fixed to common calendar years. The criteria and their rationale are detailed in Section 3. To our knowledge, existing SLAM surveys either concentrate on a single modality or treat the field as a whole without a principled, maturity-aware account of method selection, and few foreground the GPS-denied setting across the full breadth of sensing configurations considered here, including the comparatively under-surveyed acoustic and radar modalities.

The contributions of this survey are as follows:

(i) a hardware-centric taxonomy of SLAM and odometry methods spanning six sensing configurations, unified under a GPS-denied operating assumption,

(ii) a four-tier temporal organization within each modality, with explicit, maturity-appropriate selection criteria that fairly accommodate both foundational and very recent work,

(iii) a cross-cutting analysis of datasets, evaluation metrics, and the comparative trade-offs that govern modality choice in GPS-denied environments, and

(iv) a synthesis of open challenges and future directions, with particular attention to robustness under sensor degradation, multi-sensor fusion, emerging map representations, and benchmarking gaps in the less-studied modalities.

The remainder of this paper is organized as follows. Section 2 reviews the conceptual foundations common to all SLAM systems - front-end estimation, back-end optimization, loop closure, and the evolution of map representations. Section 3 details the review scope, taxonomy, and selection methodology. Sections 4–9 survey each of the six sensing configurations in turn, tracing their development across the four maturity tiers and closing each with subfield-specific trends. Section 10 presents the cross-cutting analysis of datasets, metrics, and modality trade-offs, and Section 11 discusses open challenges and future directions. Section 12 concludes.

2 Background and Preliminaries

This section establishes the conceptual machinery common to the methods surveyed in the remainder of the paper. Modern SLAM and odometry systems share a common architecture and a common set of estimation problems. The primary distinguishing factors are the sensing configuration and the specific architectural choices made within a few well-defined modules. The sensor-specific front-end, which contains feature extraction and data association for cameras and scan registration for ranging sensors, is treated within each modality's section. Here we describe the shared back-end formulation, the mechanisms that bound drift, the role of inertial sensing, the representations used for the map, and the conditions under which estimation degrades. Establishing these once avoids repeating them across six sections and isolates what is genuinely modality-specific.

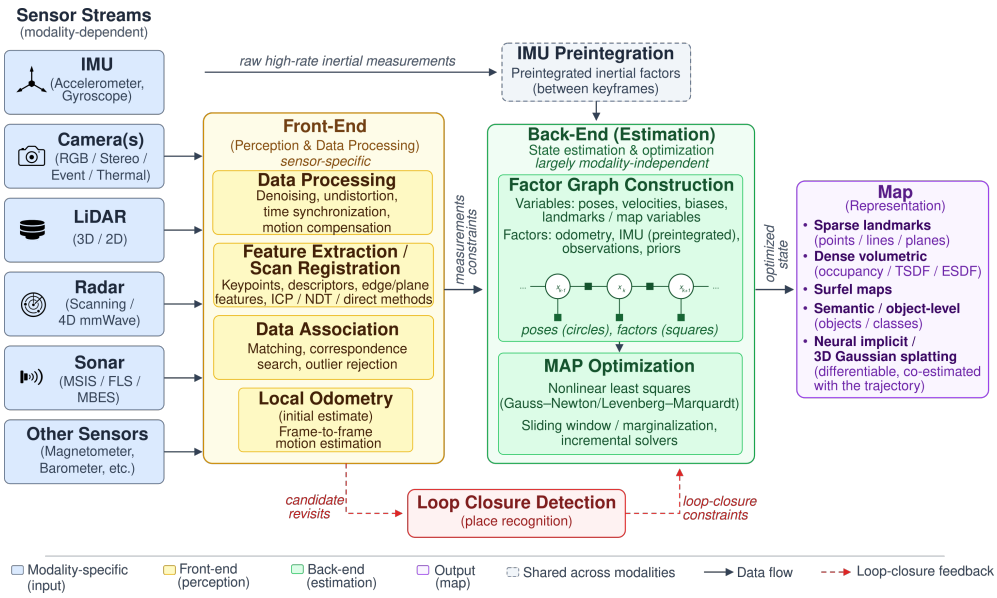


Fig. 1 Generic architecture of modern SLAM and odometry systems. Modality-dependent exteroceptive sensor streams feed a sensor-specific front-end that produces relative geometric constraints, while raw inertial measurements are summarized by IMU preintegration into relative-motion factors (Section 2.5). A largely modality-independent back-end constructs and solves the MAP estimation problem of Section 2.1. Loop-closure constraints from place recognition bound accumulated drift and yield a globally consistent map.

2.1 Problem formulation

Let the state to be estimated be denoted by $\mathbf{X} = \{\mathbf{x}_{0:t}, \mathbf{m}\}$, comprising the platform's trajectory, represented as a sequence of poses $\mathbf{x}_{0:t}$ with $\mathbf{x}_i \in SE(3)$, together with a map $\mathbf{m}$ of the environment, whose form depends on the chosen representation (Section 2.6). The system receives proprioceptive measurements from an inertial sensor and

exteroceptive measurements $\mathbf{z}_{1:k}$ from one or more environment-sensing modalities. Assuming independent, zero-mean Gaussian measurement noise, the MAP estimate can be written as the nonlinear least-squares problem:

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \sum_k ||h_k(\mathbf{X}) - \mathbf{z}_k||^2_{\Sigma_k} \tag{1}$$

where $\mathbf{X}^*$ denotes the MAP estimate of the complete state, h_k is the measurement model predicting the k^{th} observation, $\mathbf{z}_k$ the corresponding measurement, and Σ_k its covariance [1, 2]. The weighted norm is defined as $||\mathbf{e}||^2_\Sigma = \mathbf{e}^\top \Sigma^{-1} \mathbf{e}$, and the index k runs over all measurement factors, including inertial, exteroceptive, and prior constraints. This objective, together with the means of constructing and solving it efficiently, forms the conceptual core of nearly all methods reviewed in this survey.

Two problems sit on a continuum within this formulation. Odometry estimates incremental ego-motion by registering consecutive measurements and, in most modern formulations, maintains only a local window of recent state, marginalizing the past. Because each estimate is composed from the previous one, error accumulates without bound along the trajectory. SLAM additionally recognizes when the platform revisits a previously mapped region and introduces constraints that link the current state to that earlier one, then re-optimizes the trajectory and map jointly, bounding the accumulated error and producing a globally consistent reconstruction. Even though both estimate the same quantity, they differ in whether global consistency is enforced. This distinction is particularly consequential under GPS-denied operation: with no absolute position reference to reset error, the estimate is effectively dead-reckoned from the origin, and the magnitude of accumulated drift, together with whatever mechanisms exist to mitigate it, frequently dominates overall localization performance.

2.2 Front-end and back-end architecture

Most contemporary systems decompose into a front-end and a back-end [3] (Figure 1). The front-end converts raw sensor streams into relative geometric constraints. It extracts and matches features or registers successive scans, establishes data association, and produces short-term motion estimates and candidate correspondences. It is inherently sensor-dependent and is therefore the component that differs most across the modalities considered in this survey. The back-end consumes the resulting constraints and computes the MAP estimate of the trajectory and map by solving the optimization presented in Section 2.1. The back-end is, to a large degree, modality-independent. This separation is articulated in early keyframe-based visual systems that ran tracking and mapping as parallel processes [4] and underlies the organization of the present survey, in which back-end machinery is shared and front-ends are described per modality.

2.3 From filtering to optimization

Early SLAM systems were formulated as recursive Bayesian filters. The extended Kalman filter (EKF) maintained a joint Gaussian belief over the pose and all map landmarks, but the quadratic growth of the covariance matrix with map size limited

its scalability [5]. Rao-Blackwellized particle filters offered an alternative that factored the problem [6–8], but introduced sample-impoverishment concerns [9]. From the late 2000s the field shifted decisively toward smoothing, formulating SLAM as the full-MAP nonlinear least-squares problem above and solving it with Gauss–Newton or Levenberg–Marquardt iterations [10]. Bundle adjustment, long established in photogrammetry [11], became the back-end of choice for visual systems, while the factor graph emerged as a unifying representation in which variables (poses, landmarks, biases) and the constraints relating them are made explicit [12]. Pose-graph optimization, a reduced form retaining only poses and their relative constraints, is widely used where a full map does not need to be optimized [13, 14]. Practical efficiency derives from several devices: keyframe selection to bound problem size, sliding-window or fixed-lag optimization with marginalization of old states, and incremental solvers that reuse computation as the graph grows [12]. Filtering persists mainly in tightly-coupled visual–inertial estimators where computational and latency budgets are stringent [15]. Whether a method filters or optimizes is, alongside its sensing configuration, one of the two principal axes along which the systems in each section differ.

2.4 Loop closure and place recognition

Loop closure is the mechanism that converts odometry into SLAM. When the platform re-observes a previously mapped place, a constraint is added linking the current pose to the earlier one. Global optimization then redistributes the accumulated error across the intervening trajectory, bounding drift and rendering the map globally consistent. Recognizing such revisits is the task of place recognition. In vision, place recognition is approached through appearance-based methods such as bag-of-visual-words models [16, 17] and, increasingly, learned global descriptors. In ranging modalities the approach is through geometric or structural descriptors that summarize a scan independently of viewpoint [18]. Place recognition is markedly harder for low-resolution, noise-dominated modalities such as sonar and radar. Because a single incorrect loop closure can corrupt the entire estimate, back-ends incorporate outlier-robust mechanisms. These include robust cost kernels [19], switchable constraints [20], and geometric verification of candidate matches [21]. In GPS-denied settings the relevance of this component is twofold: it is the primary means of achieving global consistency without external positioning, yet in long, featureless, or non-revisiting traversals loop-closure opportunities are scarce, so drift mitigation falls increasingly on the front-end and on inertial aiding.

2.5 Inertial fusion and IMU preintegration

An Inertial Measurement Unit (IMU) provides high-rate measurements of angular velocity and linear acceleration that are independent of environmental appearance or structure, making it robust precisely where exteroceptive sensors fail. However, used alone, it drifts rapidly, as pose is recovered through double integration in the presence of sensor bias. Its complementarity with cameras and ranging sensors explains its presence in four of the six configurations surveyed here. Fusion is termed loosely coupled when independent pose estimates are combined and tightly coupled when

raw inertial and exteroceptive measurements are jointly estimated [22], the latter generally yielding higher accuracy and robustness [23]. The key enabling technique for optimization-based tightly-coupled fusion is IMU preintegration [24], which summarizes the many inertial samples between two keyframes into a single relative-motion constraint expressed on the manifold, so that the back-end need not re-integrate the inertial stream at every solver iteration. In the filtering paradigm, the multi-state constraint Kalman filter plays the analogous role [15]. For GPS-denied operation, inertial aiding is what sustains a usable estimate through transient failures of the primary sensor, namely darkness, dust, or texture-poor passages and through the aggressive motion typical of agile platforms.

2.6 Map representations

The map is simultaneously an output of the system and, increasingly, a participant in the estimation itself, and the choice of representation has expanded considerably over the period this survey covers. The earliest and most economical representation is a sparse landmark map of discrete feature points, sufficient for localization but not for dense reconstruction. Dense and volumetric representations followed: occupancy grids and their hierarchical, memory-efficient octree variants for navigation [25], truncated signed-distance fields and volumetric fusion for surface reconstruction [26], and Euclidean signed-distance fields tailored to motion planning [27]. Surfel maps, which model surfaces as oriented discs, offer an efficient middle ground for dense mapping [28]. More recently, implicit neural representations have encoded the scene as a continuous function realized by a neural network, yielding compact, complete, and photorealistic maps at substantial computational cost [29]. Since 2023, 3D Gaussian splatting has introduced an explicit yet fully differentiable representation supporting real-time photorealistic rendering [30], and has been adopted rapidly within SLAM (Figure 2). The trajectory of this evolution, from geometry-only maps toward dense, differentiable, photorealistic representations co-estimated with the trajectory, is a theme that recurs across several modalities and constitutes a major frontier direction discussed in Section 11.

2.7 Observability, degeneracy and robustness

Because GPS-denied estimation is unaided by any global fix, the conditioning of the underlying optimization is a first-order concern, and certain environments and motions render the state partially unobservable. Ranging-based systems suffer geometric degeneracy in self-similar or structure-poor surroundings such as long corridors, tunnels, open spaces, where scan registration is under-constrained along one or more directions. Monocular visual systems face scale unobservability and degrade under low texture, and acoustic and radar returns are sparse and noise-dominated by nature. Two strategies recur throughout the survey in response. The first is degeneracy detection, in which the ill-conditioned directions of the estimation problem are identified and the update is constrained accordingly [31]. The second, and the more general, is multi-sensor complementarity: fusing modalities whose failure modes are statistically uncorrelated, so that one sensor constrains the state where another cannot [32].

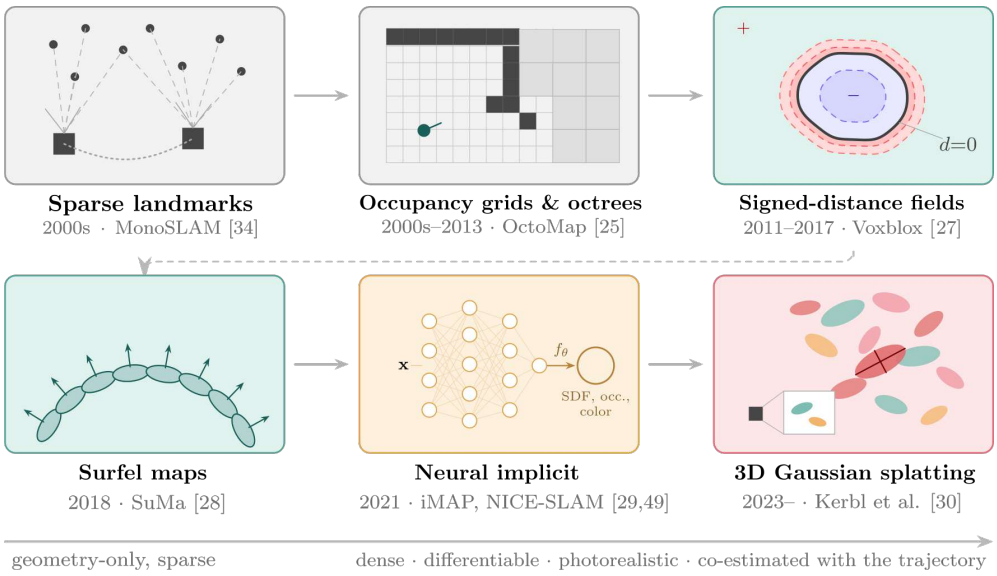


Fig. 2 Evolution of map representations in SLAM, from sparse geometric maps toward dense, differentiable, photorealistic representations that are co-estimated with the trajectory. Representative systems and years are indicated below each representation.

This principle is the central motivation for the fused configurations of visual–inertial, LiDAR–inertial, and LiDAR–visual–inertial, that dominate the later sections and the current state of the art in robust GPS-denied operation [33].

3 Scope and Review Methodology

This section defines the boundaries of the review and the procedure by which methods were selected and organized. We survey the literature along two orthogonal axes: the sensing configuration a method requires, and its maturity within its own subfield. Thus, we make the selection criteria explicit so that inclusion is traceable and the organization reproducible.

3.1 Scope and inclusion criteria

We define GPS-denied operation as any setting in which a global satellite navigation fix is persistently unavailable, unreliable, or too inaccurate to serve as a localization reference. Although GNSS-denied is the technically precise umbrella term, encompassing GPS, GLONASS, Galileo, and BeiDou alike, we retain GPS-denied throughout as the term entrenched in the robotics literature, and use it interchangeably with GNSS-denied. These settings include subterranean, indoor, and underwater environments, as well as densely vegetated, obstructed, or adversarial areas. As established in Section 2.1, odometry and SLAM occupy a continuum rather than forming disjoint classes, and we include both: full SLAM systems that enforce global consistency

through loop closure, and odometry-only pipelines that estimate motion incrementally. Under GPS-denied operation this inclusiveness is necessary rather than merely permissive, since no external fix exists to reset accumulated error and the quality of the odometry estimate is frequently the dominant factor in overall localization performance.

We exclude two classes of method. The first comprises approaches that fuse GNSS or other global positioning signals, as these presuppose the very reference our setting assumes absent. If the estimator of such a method remains well defined once the global-positioning factor is removed, as in several loosely-coupled LiDAR–inertial–GNSS systems, we consider the GNSS-free variant. The second comprises methods that depend on pre-installed external infrastructure, such as fixed motion-capture rigs, ultra-wideband beacon networks, or engineered fiducial-marker fields, since these substitute one form of external positioning for another and are inapplicable in the unprepared environments of interest. The review spans methods published from the foundational works of the early 2000s up to the time of writing.

3.2 Taxonomy by sensing configuration

Because the choice and combination of sensors is the primary determinant of how a system behaves under the degraded conditions characteristic of GPS-denied environments (Section 2.7), we adopt a hardware-centric taxonomy and group methods into six categories by their required sensing configuration:

1. Visual: relying on one or more cameras,
2. Visual–Inertial: augmenting cameras with an inertial measurement unit (IMU),
3. LiDAR: using a ranging laser scanner alone,
4. LiDAR–Inertial: fusing LiDAR with an IMU,
5. LiDAR–Visual–Inertial: jointly exploiting all three,
6. A residual category: encompassing the remaining modalities relevant to GPS-denied operation, principally acoustic (sonar), radar, and thermal sensing.

The categories are ordered by increasing sensory redundancy, from single-sensor visual estimation to fully fused LiDAR–visual–inertial systems, with the residual category placed last. An inertial unit appears in four of the six configurations, reflecting its role (Section 2.5) as a low-cost, environment-independent complement whose failure modes are uncorrelated with those of exteroceptive sensors. Many systems are modular and admit optional sensors. We classify each by the sensors its core estimator requires rather than by those it can optionally incorporate, so that, for example, a LiDAR–inertial system offering an optional camera module is reviewed under LiDAR–inertial. Figure 3 presents the resulting taxonomy.

3.3 Temporal organization and selection criteria

Within each modality we organize methods chronologically, but not as a flat ranking by citation count. Citation accrual naturally lags publication; therefore, raw citation counts inherently favor older methods, and even a citations-per-year normalization systematically undercredits the most recent work, whose citing literature does not

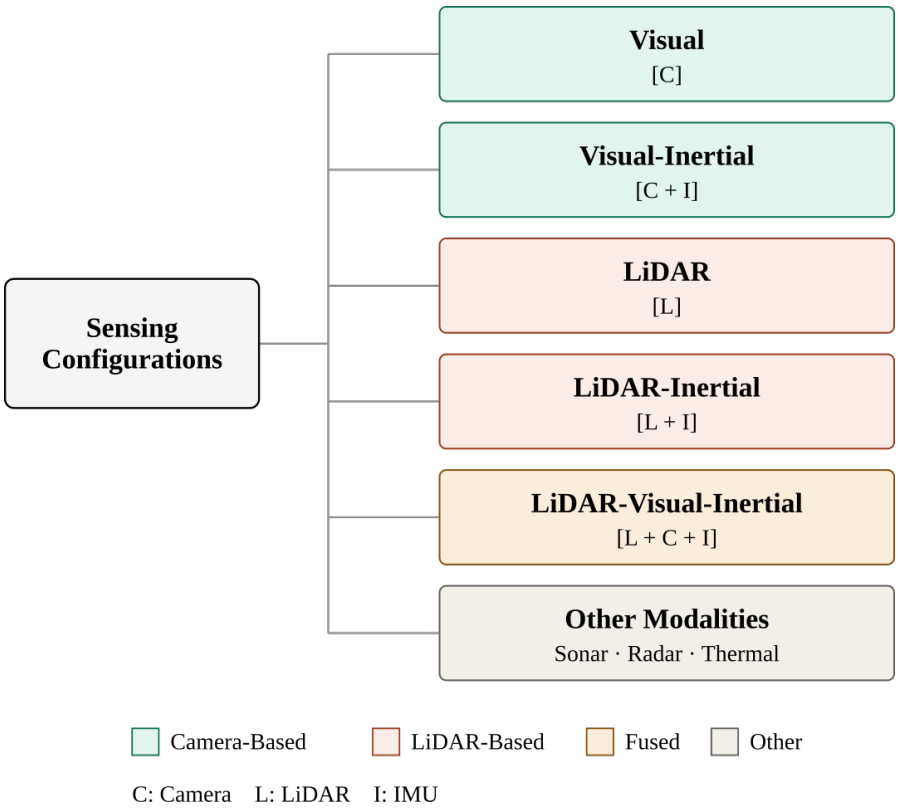


Fig. 3 Taxonomy of the surveyed sensing configurations

yet exist. No single bibliometric score can fairly rank field-defining classics alongside methods only months old. We therefore group methods into four maturity tiers: foundational, established, recent, and frontier. Each tier is defined by the selection criterion appropriate to its maturity rather than by a fixed publication window. Foundational methods are the field-defining works of a modality, included by community consensus, with citation counts serving only to confirm canonical status. Established methods are mature, widely adopted systems whose citation curves have stabilized, for which absolute citations and citations-per-year are reliable signals. Recent methods, whose citation curves are still rising, are selected by citations-per-year supplemented by open-source adoption and reported benchmark performance. Frontier methods are too recent for citation data to be informative and are selected instead by publication venue, open-source adoption, standing on public benchmarks, and, where necessary, curatorial judgment. Figure 4 summarizes the selection procedure, from candidate identification through tier-appropriate selection to the per-modality boundary calibration detailed in Section 3.4.

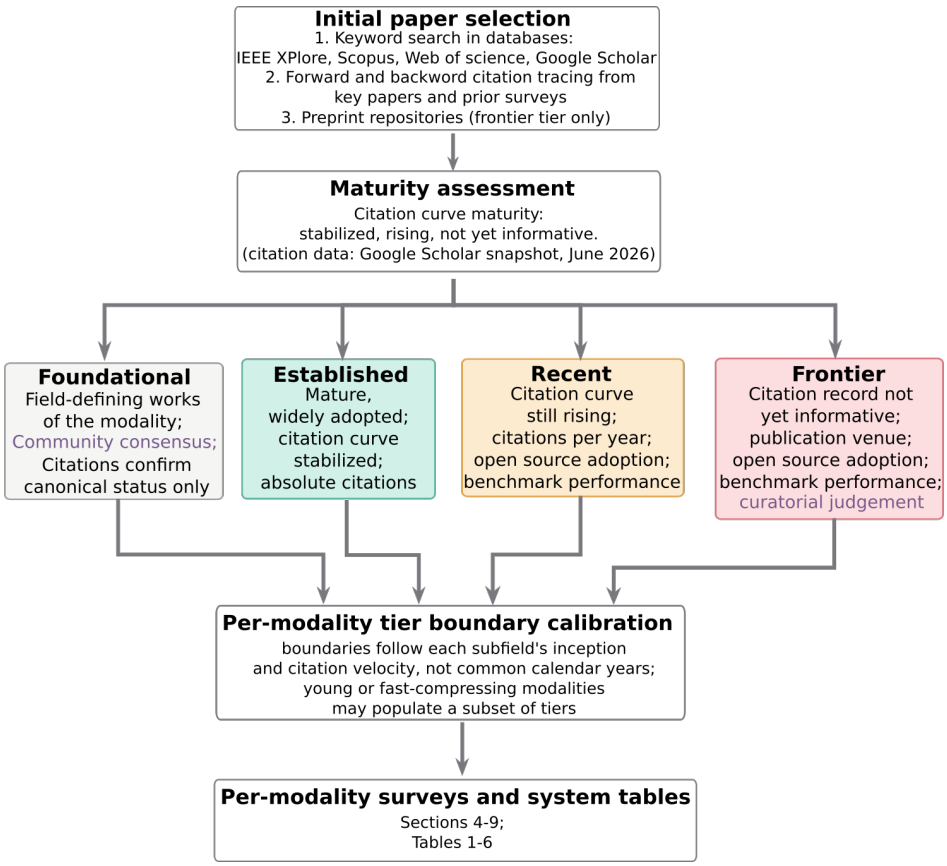


Fig. 4 Selection workflow of the review methodology (Sections 3.3-3.5). Candidates identified from the listed sources are assigned to maturity tiers according to the state of their citation curve, and each tier applies the selection criterion appropriate to its maturity. Criteria involving editorial judgment are indicated in purple. Tier boundaries are subsequently calibrated per modality (Figure 6). Citation data reflect a Google Scholar snapshot of June 2026.

3.4 Per-modality boundary calibration

Because subfields emerged at different times and advance at different rates, these tier boundaries are not aligned to common calendar years across modalities. The foundational/established boundary follows the inception and maturation of each subfield. The foundational period for visual SLAM, for instance, precedes that of LiDAR–inertial or 4D-radar SLAM by over a decade, while the recent/frontier boundary follows the citation velocity of each subfield, and is correspondingly narrower for fast-moving areas and wider for those that publish slowly. The same tier label therefore denotes different time spans across modalities, and emerging modalities may populate only a subset of the four tiers. Thermal SLAM, for example, currently presents foundational, established, and frontier tiers but no settled recent tier, since post-established work is too recent to have separated into distinct recent and frontier stages. Conversely,

radar's four tiers are split across two sensor generations, the scanning-radar line supplying the foundational and established tiers and the 4D line the recent and frontier. The converse case also arises: where a subfield has compressed its maturation into a few years, adjacent tiers may not yet be separable. In LiDAR–visual–inertial SLAM (Section 8), the widely adopted baselines of the modality and their immediate successors appeared within a span of roughly three years, too short an interval for citation velocity to distinguish established from recent work without chronological inconsistency. We therefore treat the two as a single consolidated tier for that modality, and it presents foundational, established, and frontier tiers only. Figure 6 illustrates how the boundaries fall for representative modalities. We state these criteria explicitly so that the inclusion of each method remains traceable and the procedure reproducible.

3.5 Literature sources and selection procedure

This is a curated, critical review so the foundational and established tiers approximate community consensus and the recent tier rests on quantitative signals. However, selection at the frontier necessarily involves editorial judgment, which we have sought to make transparent through the criteria above. Candidate works were identified through IEEE Xplore, Scopus, Web of Science, and Google Scholar, supplemented by forward and backward citation tracing from key papers and prior surveys and by the proceedings of the principal robotics and vision venues such as IEEE Transactions on Robotics, the International Journal of Robotics Research, IEEE Robotics and Automation Letters, ICRA, IROS, RSS, CoRL, and CVPR/ICCV/ECCV, together with the domain-specific venues relevant to the acoustic and radar modalities. Figure 5 depicts the resulting selection procedure, from database query through tier-appropriate ranking to inclusion. For the frontier tier we additionally consulted preprint repositories. Citation counts and citations-per-year were obtained from Google Scholar as of June 2026 and should be read as a snapshot, since they evolve continuously. Where citation data were uninformative, we relied on publication venue, open-source adoption gauged from public code repository activity, and standing on community benchmarks. The benchmarks themselves are surveyed in Section 10. Figure 7 situates the reviewed corpus against the growth of its subfields: panel (a) traces the publication volume of each sensing configuration, and panel (b) shows the temporal and modal composition of the 135 papers selected by the procedure above. Throughout, we count as a system a complete localization pipeline assigned to a maturity tier. Component methods such as loop-closure modules, registration front-ends, and map post-processing, marked ‡ in Tables 3,4 and 6, and the two-sensor precursors marked † in Table 5, are discussed where they shaped the development of a modality, included in the paper count, but not assigned system tiers.

4 Configuration: Visual SLAM

Visual SLAM relies on one or more cameras as its sole exteroceptive sensor. Cameras are passive, inexpensive, lightweight, and information-rich, which makes them attractive for the size, weight, and power-constrained platforms common in GPS-denied

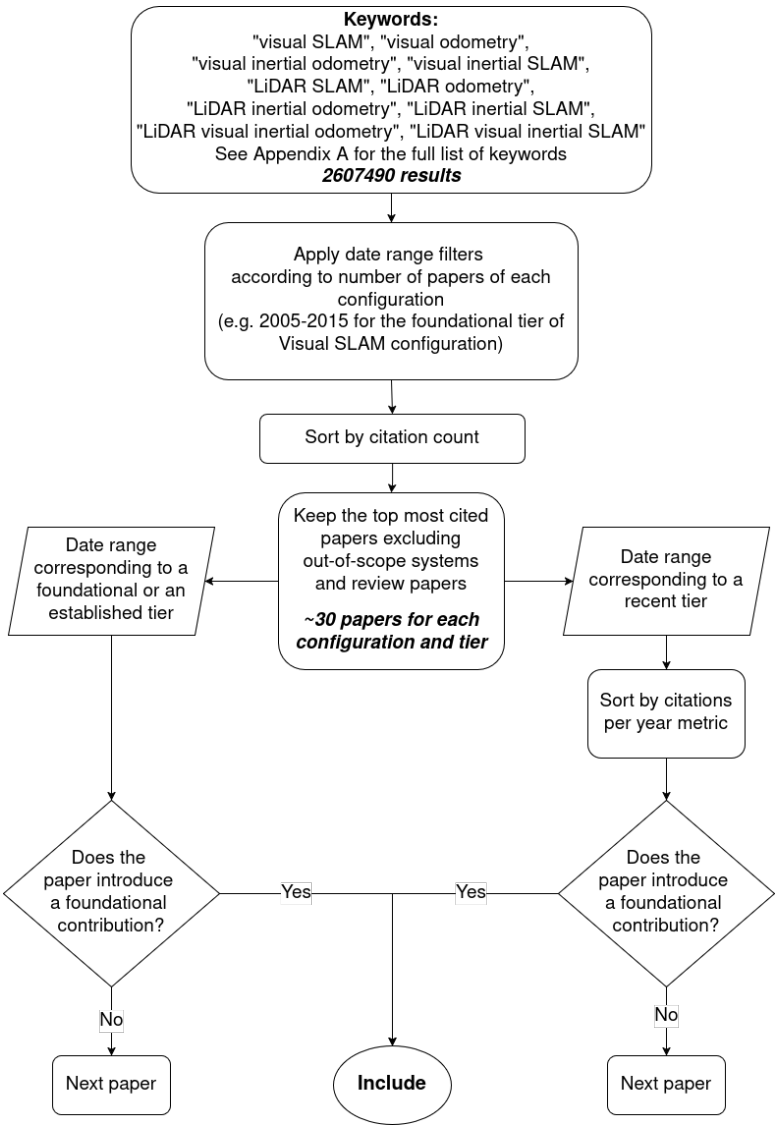


Fig. 5 Procedure by which candidate methods were identified and selected. For each sensing configuration, database queries were issued using the keyword sets shown and the results partitioned by the date ranges corresponding to that modality's maturity tiers (Section 3.4). Candidates were then ranked by the bibliometric signal appropriate to the tier (absolute citation count for the foundational and established tiers, citations per year for the recent tier) after excluding review articles and systems outside the scope. Highly ranked candidates were retained only where they introduced a distinct methodological contribution rather than an incremental variant of an approach already represented. Consistent with the curated rather than systematic character of this review, frontier-tier works, whose citation records are not yet informative (Section 3.3), were selected instead by publication venue, open-source adoption, and benchmark standing, and are not depicted here. Reported result counts are Google Scholar snapshots as of June 2026.

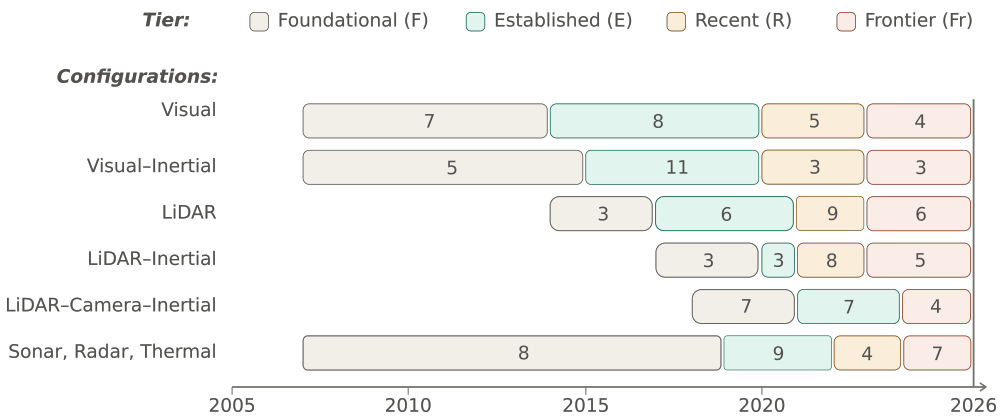


Fig. 6 Tier boundaries are calibrated per modality rather than to common calendar years. A subfield's foundational period begins at its inception, while the recent/frontier boundary tracks its citation velocity. Young modalities populate only a subset of the tiers. The number on each bar represents the number of papers that passed the exclusion criteria.

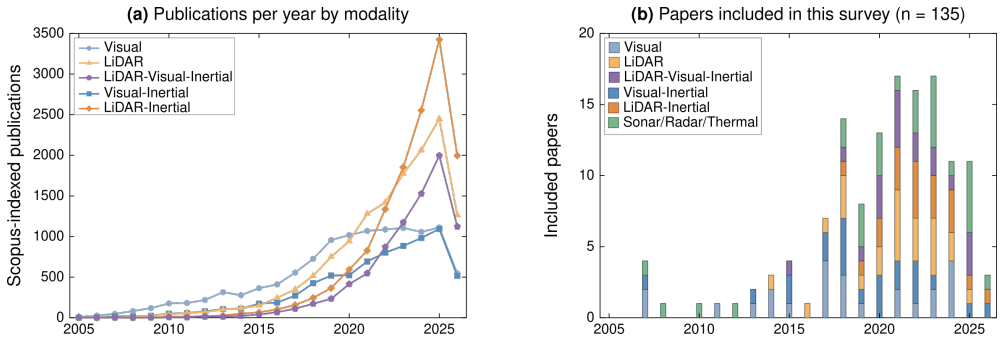


Fig. 7 Growth of the surveyed subfields and composition of the reviewed corpus. (**a**) Scopus indexed publications per year for each sensing configuration, retrieved June 2026 using the query strings listed in Appendix A. Counts are indicative, as queries overlap for multi-sensor systems and keyword conventions drift over time. The residual sonar/radar/thermal category is omitted from (a) as its heterogeneous terminology admits no comparable query. (**b**) The 135 papers included in this survey by publication year and sensing configuration (Tables 1-6). 2026 is partial, covering publications through June.

operation, and the image stream supports not only geometry but also appearance-based place recognition (see Section 2.4). These same properties bound the modality's robustness. A monocular camera cannot observe metric scale, and all cameras degrade under low illumination, weak or repetitive texture, motion blur, and abrupt changes in dynamic range, precisely the conditions encountered in tunnels, at night, or amid dust and smoke. Visual SLAM is therefore at once the foundation of the field and, in its pure form, among the most fragile configurations under the degraded conditions that define GPS-denied environments. Much of the subsequent development surveyed in this paper can be read as a response to these limitations. The evolution of the modality traces a clear arc: from sparse, feature-based estimation, through direct and dense

formulations, to learned front-ends and, most recently, neural and point-based differentiable map representations co-estimated with the trajectory. Table 1 summarizes the visual SLAM systems discussed below across the four maturity tiers.

4.1 Foundational methods

MonoSLAM (Davison et al., 2007) [34] established that real-time monocular SLAM was feasible, maintaining a probabilistic extended-Kalman-filter estimate over the camera pose and a sparse set of mapped features. The quadratic growth of the covariance of the filter with map size confined it to small workspaces. PTAM (Klein and Murray, 2007) [4] introduced the architectural idea that still underlies most systems: separating per-frame tracking from keyframe-based bundle adjustment and running them as parallel threads, as explained in Section 2.2. This set the grounds that made repeated optimization tractable in real time. The same period saw the emergence of direct methods, which operate on pixel intensities rather than extracted features. DTAM (Newcombe et al., 2011) [35] demonstrated dense, whole-image tracking against a per-pixel depth map on the GPU, an approach extended to RGB-D sensors by direct frame-to-keyframe alignment over photometric and depth residuals (Kerl et al., 2013) [36]. Later, LSD-SLAM (Engel et al., 2014) [37] achieved large-scale, CPU-real-time monocular operation by reconstructing only high-gradient, semi-dense regions and maintaining a pose graph of keyframes linked by similarity transforms to absorb the scale drift inherent to monocular estimation. SVO (Forster et al., 2014) [38] and its extension (Forster et al., 2017) [39] occupied a middle ground, combining direct alignment for speed with feature-based refinement for accuracy. As an odometry pipeline without loop closure (Section 3.1), it was designed for the fast, downward-facing camera configurations of micro aerial vehicles. Together these systems established the two axes, feature-based versus direct, sparse versus dense, along which visual SLAM continues to be organized.

4.2 Established methods

The feature-based line matured into ORB-SLAM (Mur-Artal et al., 2015) [40], which combined ORB features with three parallel threads for tracking, local mapping, and loop closing, a covisibility graph for scalable local optimization, and bag-of-words place recognition for relocalization and loop closure. Its robustness and completeness made it the de facto reference monocular SLAM system, and ORB-SLAM2 (Mur-Artal and Tardós, 2017) [41] extended it to stereo and RGB-D input with full bundle adjustment.

A recurring weakness of point-feature methods is their dependence on well-textured scenes, which leaves them short of correspondences in the low-texture but geometrically structured environments, such as corridors, tunnels, man-made interiors, that are common in GPS-denied operation. PL-SLAM (Gomez-Ojeda et al., 2019) [42] addressed this by fusing point and line-segment features in a single stereo SLAM pipeline, exploiting the structural regularity of such scenes to sustain tracking where points alone are too sparse. Line features thereby became a standard device for hardening visual estimation against texture-poor conditions.

A second limitation of the systems above is their static-world assumption: moving objects introduce spurious correspondences that corrupt both pose and map. DS-SLAM (Yu et al., 2018) [43] paired a semantic-segmentation network with a moving-consistency check to identify and discard features on dynamic objects, constructing a static-only map atop ORB-SLAM2, while the contemporaneous DynaSLAM (Bescos et al., 2018) [44] combined segmentation with multi-view geometry and inpainted the occluded background. These are early representatives of dynamic-environment handling, a concern that recurs across the modalities surveyed here.

On the direct side, DSO (Direct Sparse Odometry, Engel et al., 2018) [45] jointly optimized camera poses, inverse depths, and photometric calibration parameters over a sliding window, achieving high accuracy from raw intensities without feature extraction. As an odometry method it omitted loop closure, though later extensions reintroduced it. This tier marks the point at which monocular and RGB-D visual estimation became accurate and robust enough for routine use in well-conditioned environments, and its systems remain widely used baselines.

4.3 Recent methods

Two distinct developments mark the recent tier. The first is the entry of learning into the estimation core. Learned components had been explored earlier. CNN-SLAM (Tateno et al., 2017) [46] fused CNN-predicted dense depth into a keyframe-based monocular system to recover the metric scale unobservable to a single camera, and DeepVO (Wang et al., 2017) [47] regressed camera motion end-to-end with a recurrent convolutional network. These early efforts, though influential, were not yet accurate enough to displace geometric methods. DROID-SLAM (Teed and Deng, 2021) [48] marked the turn: it replaced hand-engineered tracking with a learned update operator that iteratively refines camera poses and per-pixel depth, coupled to a differentiable dense bundle-adjustment layer. Trained once, it generalized across monocular, stereo, and RGB-D input and proved markedly more robust than classical systems on difficult sequences, establishing learned dense optimization as a serious alternative to feature- and intensity-based pipelines. The second is the adoption of neural implicit representations as the map itself. iMAP (Sucar et al., 2021) [29] first showed that a single multilayer perceptron, trained continually online, could serve as the sole scene representation for real-time RGB-D SLAM, yielding complete, watertight reconstructions of small scenes. NICE-SLAM (Zhu et al., 2022) [49] introduced hierarchical feature grids with pretrained decoders to scale the approach to larger environments. Subsequent work pushed efficiency and fidelity further, with ESLAM (Johari et al., 2023) [50] employing multi-scale feature planes and Co-SLAM (Wang et al., 2023) [51] combining coordinate and sparse parametric encodings with real-time bundle adjustment over keyframes. These neural systems prioritized reconstruction quality and completeness and were largely demonstrated on RGB-D input at room scale, rather than in the large or degraded environments typical of GPS-denied fieldwork.

4.4 Frontier methods

The current frontier is defined by the rapid adoption of 3D Gaussian splatting, a representation introduced for novel-view synthesis (Kerbl et al., 2023) [30]. This representation models a scene as a set of anisotropic 3D Gaussians rendered by differentiable rasterization, offering an explicit, fast, and fully differentiable alternative to implicit neural fields. Within a year it was incorporated into SLAM: SplaTAM (Keetha et al., 2024) [52] performs dense RGB-D tracking and mapping against a Gaussian map with silhouette-guided densification. Gaussian Splatting SLAM (Matsuki et al., 2024) [53] demonstrated incremental monocular and RGB-D operation with direct optimization of camera pose against the Gaussian map using analytic Jacobians. GS-SLAM (Yan et al., 2024) [54] and Photo-SLAM (Huang et al., 2024) [55] further developed adaptive map expansion and the pairing of explicit geometry with photorealistic appearance. Because citation records for these systems are not yet informative (Section 3.3), their selection here rests on publication venue, open-source adoption, and benchmark performance. Their shared thrust is a map that is at once a geometric estimate and a photorealistic, differentiable model co-optimized with the trajectory.

4.5 Trends and outlook

The trajectory of visual SLAM points toward differentiable, photorealistic maps and learned or foundation-model front-ends, which promise richer scene understanding and stronger generalization across environments. Yet the constraints that have always bounded the modality persist: metric scale remains unobservable from a single camera, and the appearance and texture dependence of all visual estimation leaves it vulnerable in exactly the low-light, low-texture, and visually degraded conditions that characterize many GPS-denied settings. Additionally, highly dynamic and crowded scenes, partially resolved by the semantic and geometric rejection strategies of Section 4.2, and the computational cost of dense and neural maps on embedded hardware, remain open problems. These limitations explain why, for robust GPS-denied operation, visual estimation is most often not deployed alone but fused with an inertial sensor whose complementary, environment-independent measurements resolve scale and bridge transient visual failures. The Visual-Inertial configuration is the one which we address next.

5 Configuration: Visual-Inertial SLAM

Visual–inertial (VI) methods augment one or more cameras with an inertial measurement unit (IMU), and the pairing is strongly complementary. The IMU supplies the metric scale that is unobservable to a single camera, an absolute gravity direction, and high-rate measurements of ego-motion that remain reliable through fast rotation and motion blur, all independent of scene appearance. The camera, in turn, supplies the exteroceptive constraints that bound the rapid drift of inertial integration (Section 2.5). The result is a metric, drift-bounded estimate that tolerates brief visual failures such as a few blurred or low-texture frames, that would defeat a camera alone. This complementarity, achieved with a minimal and lightweight sensor pair,

has made visual–inertial estimation the most widely deployed configuration for GPS-denied operation, and the default choice on the size, weight, and power constrained aerial platforms where it first matured. Its central limitation is equally important and is taken up in Section 5.5: inertial aiding bridges only transient visual degradation, and sustained loss of visual information eventually exceeds the horizon over which the IMU can be trusted. The development of the modality follows two estimation paradigms that emerged together, recursive filtering and sliding-window optimization, before converging on tightly-coupled formulations and, more recently, learned front-ends and novel sensing. Table 2 summarizes the visual-inertial SLAM systems discussed below across the four maturity tiers.

Table 1 Summary of *Configuration: Visual SLAM* systems surveyed in Section 4, by maturity tier (F = foundational, E = established, R = recent, Fr = frontier).
Input: M = monocular, S = stereo, D = RGB-D. LC = loop closure.

System	Tier	Input	Approach	Map representation	LC
MonoSLAM (2007) [34]	F	M	Feature, EKF filter	Sparse landmarks	–
PTAM (2007) [4]	F	M	Feature, keyframe BA	Sparse points	–
DTAM (2011) [35]	F	M	Direct, dense (GPU)	Per-pixel depth	–
Kerl et al. (2013) [36]	F	D	Direct	Keyframe depth maps	–
LSD-SLAM (2014) [37]	F	M	Direct, semi-dense	Semi-dense depth, pose graph	✓
SVO (2014/2017) [38, 39]	F	M	Semi-direct	Sparse points	–
ORB-SLAM (2015) [40]	E	M	Feature (ORB)	Sparse, covisibility graph	✓
ORB-SLAM2 (2017) [41]	E	M/S/D	Feature (ORB)	Sparse	✓
CNN-SLAM (2017) [46]	E	M	Hybrid (CNN depth)	Keyframe dense depth	✓
DeepVO (2017) [47]	E	M	Learned end-to-end	–	–
DS-SLAM (2018) [43]	E	D	Feature + semantics	Static semantic map	✓
DynaSLAM (2018) [44]	E	M/S/D	Feature + sem./geom.	Static map (inpainted)	✓
DSO (2018) [45]	E	M	Direct, sparse	Sparse inverse-depth	–
PL-SLAM (2019) [42]	E	S	Feature (points+lines)	Sparse points and lines	✓
DROID-SLAM (2021) [48]	R	M/S/D	Learned dense BA	Dense per-pixel depth	✓
iMAP (2021) [29]	R	D	Neural implicit	Single MLP	–
NICE-SLAM (2022) [49]	R	D	Neural implicit	Hierarchical feature grids	–
ESLAM (2023) [50]	R	D	Neural implicit	Multi-scale feature planes	–
Co-SLAM (2023) [51]	R	D	Neural implicit	Joint coord./param. encodings	–
SplaTAM (2024) [52]	Fr	D	3D Gaussian splatting	3D Gaussians	–
Gauss. Splatting SLAM (2024) [53]	Fr	M/D	3D Gaussian splatting	3D Gaussians	–
GS-SLAM (2024) [54]	Fr	D	3D Gaussian splatting	3D Gaussians (adaptive)	–
Photo-SLAM (2024) [55]	Fr	M/S/D	Geometric + 3DGS	Points + Gaussians	✓

5.1 Foundational methods

The filtering paradigm was established by the Multi-State Constraint Kalman Filter (MSCKF, Mourikis and Roumeliotis, 2007) [15], which avoided the unbounded state growth of naive EKF-SLAM by maintaining a sliding window of past camera poses rather than the landmarks themselves, expressing each feature as a geometric constraint across the poses that observed it and then marginalizing it. This kept computation linear in the number of features and made real-time vision-aided inertial navigation feasible. ROVIO (Bloesch et al., 2015, 2017) [56, 57] extended the filtering line with a direct, photometric front-end, tracking image patches and including their intensities directly in the EKF state, which yielded a compact and robust monocular estimator. The optimization paradigm was established by OKVIS (Leutenegger et al., 2013, 2015) [58, 59], which posed VI estimation as the joint nonlinear minimization of camera reprojection errors and inertial error terms over a bounded window of keyframes, with older states removed by marginalization to preserve real-time operation. Two organizing distinctions were settled in this period: filtering versus sliding-window optimization, and loosely versus tightly-coupled fusion, with the tightly-coupled formulations that jointly estimate raw measurements proving the more accurate and robust (Section 2.5) and becoming dominant in all subsequent work.

5.2 Established methods

The optimization line matured into VINS-Mono (Qin et al., 2018) [60], a complete tightly-coupled monocular system that paired IMU preintegration with a robust initialization procedure recovering scale, gravity, velocity, and biases from motion, and added loop closure through bag-of-words place recognition and four degree-of-freedom pose-graph optimization. Its robustness and open-source availability made it, alongside OKVIS, a primary reference, and VINS-Fusion (Qin et al., 2019) [61] extended it to stereo and globally-referenced sensors. On the filtering side, OpenVINS (Geneva et al., 2020) [62] developed the MSCKF into a clean and extensible platform with online intrinsic and extrinsic calibration and long-term SLAM landmarks, becoming a widely used research base, while S-MSCKF (Sun et al., 2018) [63] showed that the low computational footprint of a stereo MSCKF could sustain robust estimation aboard fast, weight-constrained aerial vehicles. Several systems broadened the modality's outputs: BASALT (Usenko et al., 2020) [64] coupled real-time VI odometry with non-linear factor recovery for accurate mapping. Kimera (Rosinol et al., 2020) [65] combined a VI front-end with a robust pose-graph back-end, lightweight 3D mesh reconstruction, and semantic labeling, bringing metric-semantic mapping into the VI setting. ORB-SLAM3 (Campos et al., 2021) [66], building on the earlier visual–inertial extension of ORB-SLAM with map reuse (Mur-Artal and Tardós, 2017) [67], carried the line to tightly-coupled VI operation with a maximum-a-posteriori IMU initialization and a multi-map mechanism for relocalization and multi-session use, recording state-of-the-art accuracy on standard benchmarks. The structured-scene robustness afforded by line features, introduced for the visual case in Section 4.2, was carried into the inertial setting by point-and-line VIO (He et al., 2018) [68], sustaining

tracking in the low-texture but geometrically regular environments common to GPS-denied operation. This tier represents the point at which visual–inertial estimation became accurate, robust, and complete, with loop closure, mapping, semantics, and multi-session capability, and its systems remain the community's standard baselines.

5.3 Recent methods

The recent tier extends the established lines and, as in the visual case (Section 4.3), sees learning enter the pipeline. Two developments continue the geometric tradition. DM-VIO (von Stumberg and Cremers, 2022) [69] carried the direct, photometric approach of DSO into the inertial setting, introducing a delayed-marginalization scheme and pose-graph bundle adjustment for IMU initialization that allowed a monocular–inertial system to match or surpass stereo–inertial competitors. OKVIS2 (Leutenegger, 2022) [70] completed the original OKVIS by adding full loop closure and global optimization, converting a sliding-window odometry estimator into a scalable SLAM system. In parallel, learning-based approaches also began to emerge in visual-inertial estimation. Mirroring the trajectory of purely visual methods, early end-to-end learned approaches initially lagged behind their geometric counterparts in accuracy. Beginning with VINet (Clark et al., 2017) [71], which formulated visual–inertial odometry as a sequence-to-sequence learning problem, subsequent works introduced learned selective-fusion mechanisms that adaptively weighted visual and inertial measurements to improve robustness under sensor degradation [72]. Despite these advances, their accuracy and reliability remained inferior to those of contemporary geometric methods. More enduring contributions focused on the inertial channel itself. Although outside the scope of visual-inertial SLAM, TLIO (Liu et al., 2020) [73] demonstrated that learning statistical displacement priors directly from raw IMU measurements can effectively regularize inertial integration. Robustness to non-static scenes was also carried into the VI setting by systems such as DynaVINS (Song et al., 2022) [74], which rejected features inconsistent with the inertial motion prior, paralleling the dynamic-environment thread introduced in Section 4.2.

5.4 Frontier methods

The current frontier is more diffuse than the single dominant direction of the visual case, and spans three threads, none yet supported by informative citation records. The first is the move toward unified, sensor-agnostic estimators: OKVIS2-X (Boche et al., 2025) [75] generalizes the OKVIS architecture into a single framework configurable with dense depth, LiDAR, or globally-referenced measurements, anticipating the fused configurations of Sections 7 and 8. The second is the maturation of learned front-ends into standard components, with deep local features and learned matching (DeTone et al., 2018) [76] increasingly replacing hand-engineered detectors, for instance augmenting the ORB-SLAM3 pipeline with learned keypoints and loop-closure descriptors (Syed et al., 2026) [77]. The third is the adoption of novel sensors whose characteristics complement the camera–IMU pair under difficult conditions, most prominently the event camera, whose microsecond latency and high dynamic range have motivated a growing body of event–inertial odometry (Vidal et al., 2018, Chen et al., 2023) [78, 79].

Selection across all three threads rests on publication venue, open-source adoption, and benchmark standing rather than citation count.

Table 2 Summary of *Configuration: Visual-Inertial SLAM* systems surveyed in Section 5. All systems are tightly coupled unless noted.
Camera: M = monocular, S = stereo, D = RGB-D. LC = loop closure.

System	Tier	Camera	Paradigm	Distinctive contribution	LC
MSCKF (2007) [15]	F	M	Filter (EKF)	Sliding window of poses; cost linear in features	–
OKVIS (2013/2015) [58, 59]	F	M/S	Optimization	Keyframe sliding-window VI optimization	–
ROVIO (2015/2017) [56, 57]	F	M	Filter (iEKF)	Direct photometric patches in the EKF state	–
VINet (2017) [71]	E	M	Learned	First sequence-to-sequence learned VIO	–
VINS-Mono (2018) [60]	E	M	Optimization	Robust initialization; 4-DoF pose graph	✓
S-MSCKF (2018) [63]	E	S	Filter (MSCKF)	Low-cost stereo VIO for agile MAVs	–
PL-VIO (2018) [68]	E	M	Optimization	Point + line features for low-texture scenes	–
Ultimate SLAM (2018) [78]	E	Event+frame	Optimization	Event-inertial fusion for HDR/high speed	–
VINS-Fusion (2019) [61]	E	M/S	Optimization	Generalization to multiple sensor suites	✓
OpenVINS (2020) [62]	E	M/S	Filter (MSCKF)	Extensible platform; online calibration	–
BASALT (2020) [64]	E	S	Optimization	Non-linear factor recovery for VI mapping	✓
Kimera (2020) [65]	E	S	Optimization	Robust PGO; mesh and metric-semantic mapping	✓
ORB-SLAM3 (2021) [66, 67]	E	M/S/D	Optimization	MAP IMU init.; multi-map, multi-session	✓
DM-VIO (2022) [69]	R	M	Optim. (direct)	Delayed marginalization; PG bundle adjustment	–
OKVIS2 (2022) [70]	R	M/S	Optimization	Full loop closure and global optimization	✓
DynaVINS (2022) [74]	R	M/S	Optimization	Dynamic-feature rejection via inertial prior	✓
ESVIO (2023) [79]	Fr	Event stereo	Optimization	Event-based stereo VIO	–
OKVIS2-X (2025) [75]	Fr	Configurable	Optimization	Unified estimator: dense depth / LiDAR / GNSS	✓
SuperPoint-SLAM3 (2026) [77]	Fr	M/S	Hybrid	Learned features + LC descriptors in ORB-SLAM3	✓

5.5 Trends and outlook

Visual–inertial estimation is the workhorse of GPS-denied localization, and its trajectory points toward learned and foundation-model front-ends for stronger generalization, unified estimators that treat additional sensors as optional inputs, and novel

sensing such as event cameras for high-speed and high-dynamic-range operation. The defining constraint of the modality nonetheless remains: the IMU bounds drift only over the interval its bias and noise characteristics permit, so inertial aiding rescues the estimate through transient visual failures but not through sustained degradation such as extended darkness, dense airborne obscurants, or long featureless passages that outlast the inertial horizon. Furthermore, sensitivity to IMU initialization and calibration persists alongside the perennial cost of dense and learned maps on embedded hardware. Where visual information is unreliable not momentarily but continuously, a sensor that perceives geometry independently of ambient light and surface texture becomes necessary, which motivates the ranging-based configurations of the sections that follow.

6 Configuration: LiDAR SLAM

LiDAR-based methods use a ranging laser scanner as their sole exteroceptive sensor, and their characteristics are in many respects the complement of the visual modality. A LiDAR measures scene geometry directly, accurately, and at metric scale, independent of ambient illumination and surface texture. It operates in total darkness and requires no visual features, which makes it the natural primary sensor for a large class of GPS-denied environments, like subterranean, indoor, and nocturnal settings. These strengths are offset by distinct costs: the sensor is active, heavier, and more power-hungry than a camera. Additionally, its returns are vertically sparse, and, consequently, it degrades in geometrically self-similar or structure-poor surroundings such as long corridors, tunnels, tree farms, and open halls, where scan registration becomes under-constrained along one or more directions (Section 2.7). The development of the modality proceeds from feature-based scan matching, through dense and efficiency-oriented refinements, to a recent bifurcation between a deliberately minimal LiDAR-only line and the field's broader migration toward inertial fusion, and finally to neural and differentiable map representations. Table 3 summarizes the LiDAR SLAM systems discussed below across the four maturity tiers.

6.1 Foundational methods

LOAM (Zhang and Singh, 2014, 2017) [80, 81] established the template for real-time LiDAR odometry and mapping by extracting edge and planar feature points from each scan and splitting the computation into a high-rate, lower-accuracy odometry thread and a lower-rate, higher-accuracy mapping thread, achieving low-drift motion estimation without an inertial sensor or a prior map. It became the reference against which subsequent LiDAR systems were measured and the ancestor of a large family of follow-on work. Hess et al., 2016 [82] established the complementary grid-based line, performing real-time scan-to-submap matching with a pose-graph back-end and branch-and-bound loop closure over occupancy grids, and supporting both 2D and 3D operation. Robust loop closing made it a widely deployed mapping system. Together these fixed the two dominant back-end styles — sparse feature registration and dense grid or submap alignment — that organize the modality.

6.2 Established methods

The LOAM lineage matured into systems tuned for particular platforms and for greater efficiency and robustness. LeGO-LOAM (Shan and Englot, 2018) [83] added ground-plane segmentation and a two-stage optimization that lightened computation and improved reliability for ground vehicles on variable terrain, while lighter-weight registration variants further reduced its cost (Wang et al., 2021) [84]. Loam-Livox (Lin and Zhang, 2020) [85] adapted LOAM's feature extraction and registration to the non-repetitive scanning pattern and narrow field of view of low-cost solid-state LiDARs, a sensor class that would go on to anchor much of the LiDAR–inertial (Section 7) and LiDAR–visual–inertial (Section 8) work. On the dense side, SuMa (Behley and Stachniss, 2018) [28] represented the map as a set of surfels rather than feature points (Section 2.6), performing projective scan-to-model registration and detecting loop closures directly in the surfel map, and IMLS-SLAM (Deschaud, 2018) [86] matched each scan against an implicit moving-least-squares surface model for high-accuracy scan-to-model odometry. This tier marks the point at which geometric LiDAR SLAM became accurate, real-time, and deployable across ground and aerial platforms, and its systems, LOAM and its descendants above all, remain the standard baselines and the foundation on which the inertial-fusion systems of Section 7 were built.

6.3 Recent methods

The recent period is shaped by a migration that partly empties it. As inertial fusion proved to resolve LiDAR's motion-distortion and degeneracy weaknesses, the field's most active development moved to the tightly-coupled LiDAR–inertial systems surveyed in Section 7 and the pure-LiDAR recent literature is correspondingly thin. This migration itself illustrates the evolution that the per-modality tier calibration of Section 3.4 is designed to capture. Notably, progress within LiDAR-only localization was achieved not through increasing algorithmic complexity, but by simplifying or fundamentally rethinking existing approaches. CT-ICP (Dellenbach et al., 2022) [87] introduced a continuous-time formulation with an elastic, intra-scan-deformable trajectory and integrated loop closure. KISS-ICP (Vizzo et al., 2023) [88] demonstrated that a carefully engineered point-to-point ICP with adaptive thresholding and motion compensation, but no feature extraction, no learned components, and no inertial sensor, matches or surpasses far more elaborate pipelines across a wide range of platforms and environments, a deliberate counterpoint to the prevailing trend toward heavier systems. Other LiDAR-only systems of the period reached comparable performance through different registration strategies (MULLS, DLO [89, 90]). In parallel, a learning-based line explored replacing geometric registration with networks that regress scan-to-scan motion: end-to-end architectures beginning with LO-Net (Li et al., 2019) [91] and extended by the hierarchical PWCLO-Net (Wang et al., 2021) [92], alongside self-supervised formulations such as DeLORA (Nubert et al., 2021) [93] that removed the need for ground-truth poses, though none of these displaced geometric methods at deployment. Learning was applied to loop closure as well as to odometry: LCDNet (Cattaneo et al., 2022) [94] jointly detected revisits and estimated the relative transform between scans through learned global descriptors and point-cloud

registration, a learned complement to the handcrafted place-recognition descriptors of Section 2.4. A further thread addressed the static-world assumption implicit in all of the above: moving objects leave spurious traces in the accumulated map, and dynamic-point-removal methods such as Kim and Kim, 2020 [95] and ERASOR (Lim et al., 2021) [96] reconstructed clean static maps by identifying and discarding those returns, making them the LiDAR counterpart to the dynamic-environment handling introduced for the visual case in Section 4.2.

6.4 Frontier methods

The frontier mirrors those of the visual and visual–inertial modalities (Sections 4.4, 5.4): the replacement of explicit geometric maps by neural and differentiable representations co-estimated with the trajectory, none yet supported by informative citation records. PIN-SLAM (Pan et al., 2024) [97] is the central instance, representing the map as a point-based implicit neural field and exploiting the elasticity of that representation to maintain global consistency under loop closure while remaining LiDAR-only. It builds on a lineage of implicit-neural LiDAR mapping that emerged in parallel, including large-scale hierarchical approaches such as SHINE-Mapping (Zhong et al., 2023) [98], online neural-only systems such as LONER (Isaacson et al., 2023) [99], and incremental neural odometry-and-mapping systems such as NeRF-LOAM (Deng et al., 2023) [100]. This neural line has begun to absorb the dynamic-scene concern as well, with 4D implicit representations such as Zhong et al., 2024 [101] modeling moving objects directly rather than removing them. Most recently, the 3D Gaussian-splatting representation has reached the LiDAR setting, with PINGS (Pan et al., 2025) [102] unifying Gaussian splatting and neural distance fields within a single point-based implicit map. Selection across these systems rests on publication venue, open-source adoption, and benchmark standing rather than citation count.

6.5 Trends and outlook

LiDAR is the backbone of geometric localization in GPS-denied environments, and its trajectory follows the field-wide movement toward neural, differentiable, and photorealistic maps that are optimized jointly with the trajectory and support reconstruction and rendering as well as localization. The modality's defining limitation is nonetheless structural and unshared by the sensors it is often paired with: in geometrically degenerate surroundings (uniform tunnels, featureless corridors, and open spaces that recur precisely in GPS-denied fieldwork) scan registration is under-constrained and the pose estimate drifts or fails, and no amount of improved point-cloud processing removes an ambiguity that is inherent in the geometry the sensor observes. Sparsity, sensor cost and power, and artifacts from dynamic objects compound the difficulty. Because the degenerate directions of a LiDAR-only estimate are typically well constrained by sensors with uncorrelated failure modes such as an inertial unit that observes motion through the ambiguous interval, or a camera that supplies texture where geometry is uninformative, the dominant response has been to fuse LiDAR with exactly

Table 3 Summary of *Configuration: LiDAR SLAM* systems surveyed in Section 6. Entries marked ‡ are modules or map-processing methods rather than standalone localization systems. LC = loop closure.

System	Tier	Registration	Distinctive contribution	LC
LOAM (2014/2017) [80, 81]	F	Edge/planar features	Split odometry/mapping threads; low drift	–
Cartographer (2016) [82]	F	Scan-to-submap (grids)	Branch-and-bound LC; 2D and 3D	✓
LeGO-LOAM (2018) [83]	E	Segmented features	Ground-optimized, lightweight for UGVs	✓
SuMa (2018) [28]	E	Projective scan-to-model	Surfel map; LC in surfel space	✓
IMLS-SLAM (2018) [86]	E	Scan-to-implicit-model	Moving-least-squares surface registration	–
LO-Net (2019) [91]	E	Learned	End-to-end LiDAR odometry	–
Loam-Livox (2020) [85]	E	Features (solid-state)	Small-FoV, non-repetitive-scan LiDARs	–
F-LOAM (2021) [84]	E	Features	Lightweight LOAM-family registration	–
Removert (2020)‡ [95]	–	(map post-proc.)	Multi-resolution dynamic point removal	n/a
ERASOR (2021)‡ [96]	–	(map post-proc.)	Pseudo-occupancy dynamic object removal	n/a
MULLS (2021) [89]	R	Multi-metric ICP	Versatile linear least-squares registration	✓
PWCLO-Net (2021) [92]	R	Learned	Hierarchical embedding mask optimization	–
DeLORA (2021) [93]	R	Learned (self-sup.)	No ground-truth poses needed	–
DLO (2022) [90]	R	Direct (dense points)	Fast localization with dense clouds	–
CT-ICP (2022) [87]	R	Continuous-time ICP	Elastic intra-scan trajectory	✓
LCDNet (2022)‡ [94]	–	Learned (LC module)	Joint revisit detection + registration	✓
KISS-ICP (2023) [88]	R	Point-to-point ICP	Minimal, robust; no features or learning	–
SHINE-Mapping (2023)‡ [98]	–	(mapping)	Sparse hierarchical implicit neural map	–
LONER (2023) [99]	Fr	Neural implicit	Online LiDAR-only neural SLAM	–
NeRF-LOAM (2023) [100]	Fr	Neural implicit	Incremental large-scale neural odometry/mapping	–
PIN-SLAM (2024) [97]	Fr	Point-based implicit	Elastic neural map; global consistency	✓
Zhong et al. (2024) [101]	Fr	4D neural implicit	Dynamic scenes modeled, not removed	–
PINGS (2025) [102]	Fr	Gaussians + dist. field	3DGS unified with neural distance field	✓

these complements, which motivates the LiDAR–inertial and LiDAR–visual–inertial configurations of the following sections.

7 Configuration: LiDAR-Inertial SLAM

LiDAR–inertial (LI) methods fuse a ranging laser scanner with an inertial measurement unit, and the pairing is even more tightly motivated than in the visual case. A LiDAR does not capture a scan instantaneously, but the points are sampled sequentially as the sensor sweeps, so platform motion during the sweep distorts the

point cloud and breaks the rigid-body assumption that registration relies on. The IMU's high-rate motion estimate corrects this distortion directly. It constrains the under-determined directions of the state when scan registration becomes geometrically degenerate (Section 2.7). It carries the estimate through brief intervals in which registration fails and it supplies gravity and metric scale. The LiDAR, in return, bounds the rapid drift of inertial integration. Because this fusion resolves the two failure modes, motion distortion and degeneracy, that most limit a LiDAR alone, LiDAR–inertial estimation has become the dominant configuration for GPS-denied localization on ground and aerial robots, and the workhorse of subterranean and industrial deployment. Its development, like the visual–inertial case, is organized by the same two estimation paradigms — factor-graph smoothing and recursive filtering, that this survey has traced throughout (Section 5). Table 4 summarizes the LiDAR-inertial SLAM systems discussed below across the four maturity tiers.

7.1 Foundational methods

Early LiDAR systems used inertial data only loosely, integrating IMU measurements to de-skew the point cloud before an otherwise LiDAR-only registration (as in the LOAM lineage of Section 6). The foundational LiDAR–inertial work replaced this with tight coupling, jointly estimating motion from raw LiDAR and inertial measurements. An early instance, LIPS (Geneva et al., 2018) [103], formulated tightly-coupled LiDAR–inertial estimation over a graph of 3D-plane measurements. LIO-mapping (Ye et al., 2019) [104] adapted the sliding-window optimization of VINS-Mono to the LiDAR setting, jointly minimizing LiDAR feature and IMU preintegration residuals for accurate but computationally heavy estimation. LINS (Qin et al., 2020) [105] established the filtering counterpart, formulating tightly-coupled LiDAR–inertial estimation as an iterated extended Kalman filter for robust and efficient navigation, making it the direct precursor of the filter-based systems that would follow. These systems settled the tightly-coupled formulation on which the entire modality now rests.

7.2 Established methods

Two systems, one from each paradigm, became the field's universal baselines. LIO-SAM (Shan et al., 2020) [106] built a tightly-coupled factor graph in the smoothing framework, combining a LeGO-LOAM-style front-end with IMU preintegration factors and optional loop-closure and GPS factors, all optimized incrementally. Its accuracy and clean open-source implementation made it the reference optimization-based system. FAST-LIO (Xu and Zhang, 2021) [107] established the filtering line with a tightly-coupled iterated EKF built around a novel, computationally efficient Kalman-gain formulation. It targeted the non-repetitive scanning and narrow field of view of the small solid-state LiDARs introduced in Section 6.2, extending LiDAR–inertial estimation to that fast-growing, low-cost sensor class. The same sensor class was taken up on the optimization side by keyframe-based systems such as LiLi-OM (Li et al., 2021) [108]. These two systems, representing smoothing and filtering respectively, are to LiDAR–inertial SLAM what VINS-Mono and OpenVINS are to the visual–inertial case, and remain the baselines against which new methods are measured.

7.3 Recent methods

The recent period is one of rapid refinement along both lineages, driven largely by the data structures used to store and query the map and by the handling of within-scan time. FAST-LIO2 (Xu et al., 2022) [109] removed feature extraction entirely, registering raw points directly against the map, as also pursued by [110], and introduced an incremental k-d tree (ikd-Tree) that supports efficient insertion, deletion, and query as the map grows. This combination set a new standard for accuracy and speed and anchored a large family of successors. Faster-LIO (Bai et al., 2022) [111] pursued the same efficiency goal with a sparse incremental voxel (iVox) structure and parallelization. A second thread addressed aggressive motion: Point-LIO (He et al., 2023) [112] processed measurements point by point at their individual sampling instants rather than aggregated into a single scan frame, yielding a high-bandwidth estimator able to track through motion violent enough to saturate the IMU. A third thread revisited within-scan time explicitly through continuous-time formulations, with DLIO (Chen et al., 2023) [113], the LiDAR–inertial successor to the DLO odometry of Section 6.3, using a continuous-time motion model for accurate point-cloud deskewing. In the same thread, SLICT (Nguyen et al., 2023) [114] combines surfel maps with continuous-time B-spline trajectories and Wildcat (Ramezani et al., 2022) [115] applied a continuous-time formulation at scale, later deployed for decentralized multi-robot exploration of GPS-denied subterranean environments. A fourth thread carried dynamic-environment handling into the LiDAR–inertial setting, with RF-LIO (Qian et al., 2021) [116] adopting a removal-first strategy that excises dynamic points before registration to preserve accuracy in high-dynamic scenes. Across these threads the estimators grew markedly more efficient, accurate, and robust to motion while remaining, at their core, refinements of the two established paradigms.

7.4 Frontier methods

The current frontier is dominated not by the neural-map direction of the earlier modalities but by a thread of direct GPS-denied consequence: degeneracy and localizability awareness. Building on early observability-aware navigators for perceptually-degraded settings, X-ICP (Tuna et al., 2024) [117] made the localizability of a LiDAR registration explicit, detecting the ill-conditioned directions of the scan-matching problem and constraining the update along exactly those directions to prevent the silent drift that degeneracy otherwise induces. This provides a robust, in-estimator solution to the geometric ambiguities discussed in Section 6. This has become an active line, with systems such as LODESTAR (Lee et al., 2026) [118] embedding degeneracy handling in an adaptive Schmidt–Kalman filter and COIN-LIO (Pfreundschuh et al., 2024) [119] augmenting geometric registration with LiDAR-intensity images to recover constraint in geometrically featureless but visually textured passages such as tunnels. An earlier response to the same problem reaches outside the LiDAR–IMU pair for an additional proprioceptive cue: EKF-LOAM (da Cruz Júnior et al., 2022) [120] adaptively fuses wheel odometry with LiDAR and inertial data to preserve localization in confined spaces with few geometric features, at the cost of requiring a wheel-equipped ground platform. A second thread continues the map-structure refinement of Section

7.3 toward probabilistic voxel representations, with Gaussian-voxel maps (as in LIO-GVM [121]) modeling each cell as a Gaussian for a favorable accuracy–efficiency trade-off. The neural and differentiable-map direction that defines the frontier of the visual modalities appears only tentatively here, in geometry-focused systems such as Splat-LOAM [122]. Its full, photorealistic form requires the appearance information a camera provides, and therefore emerges in the multimodal setting of Section 8. Selection across these threads rests on publication venue, open-source adoption, and benchmark standing rather than citation count.

Table 4 Summary of *Configuration: LiDAR-Inertial SLAM* systems surveyed in Section 7. All are tightly coupled unless noted.
Entries marked ‡ are modules rather than standalone localization systems. CT = continuous-time. LC = loop closure.

System	Tier	Paradigm	Distinctive contribution	LC
LIPS (2018) [103]	F	Optimization (graph)	3D-plane primitive LiDAR-inertial graph	–
LIO-mapping (2019) [104]	F	Optimization	VINS-Mono-style sliding window for LiDAR	–
LINS (2020) [105]	F	Filter (iEKF)	First iterated-EKF tight LiDAR-inertial fusion	–
LIO-SAM (2020) [106]	E	Optim. (factor graph)	Preintegration + optional LC/GPS factors	✓
FAST-LIO (2021) [107]	E	Filter (iEKF)	Efficient Kalman gain; solid-state LiDARs	–
LiLi-OM (2021) [108]	E	Optim. (keyframe)	Solid-state LiDAR keyframe optimization	✓
RF-LIO (2021) [116]	R	Optimization	Removal-first dynamic point handling	–
FAST-LIO2 (2022) [109]	R	Filter (iEKF)	Direct raw-point registration; ikd-Tree	–
Faster-LIO (2022) [111]	R	Filter (iEKF)	Sparse incremental voxels (iVox); parallelism	–
Wildcat (2022) [115]	R	Optimization (CT)	CT at scale; multi-robot subterranean	✓
EKF-LOAM (2022) [120]	R	Filter (adaptive)	Adds wheel odometry for feature-poor spaces	✓
Point-LIO (2023) [112]	R	Filter (point-wise)	Per-point updates; survives IMU saturation	–
DLIO (2023) [113]	R	Filter + CT model	Continuous-time deskewing	–
SLICT (2023) [114]	R	Optimization (CT)	Surfels + B-spline continuous-time trajectory	–
X-ICP (2024)‡ [117]	–	(registration module)	Localizability detection; constrained updates	–
COIN-LIO (2024) [119]	Fr	Filter	Intensity images recover constraint in tunnels	–
LIO-GVM (2024) [121]	Fr	Filter	Gaussian voxel map; accuracy–efficiency trade-off	–
Splat-LOAM (2025) [122]	Fr	Gaussian splatting	Geometry-focused 3DGS odometry/mapping	–
LODESTAR (2026) [118]	Fr	Filter (Schmidt-Kalman)	Degeneracy-aware adaptive filtering	–

7.5 Trends and outlook

LiDAR–inertial estimation is the most mature and widely deployed configuration for GPS-denied operation, and its trajectory points toward resilience in perceptual degradation, ever-more-efficient probabilistic map structures, and, where a camera is available, differentiable and photorealistic maps. Its limitations mark precisely the boundary of the following section. Degeneracy-aware methods mitigate but do not eliminate the failure that arises when the geometry itself is uninformative and the inertial estimate has drifted beyond its reliable horizon: in a long, straight, featureless corridor, neither sensor constrains motion along the corridor axis. The modality also yields geometry without appearance (no texture, color, or semantics) limiting its maps for tasks beyond localization, and it remains sensitive to dynamic objects. Each of these gaps is addressed by the same additional sensor: a camera, whose texture disambiguates geometrically degenerate scenes, whose appearance enriches the map, and whose failure modes are uncorrelated with those of the LiDAR–IMU pair. The tightly-coupled fusion of all three sensors is the subject of Section 8.

8 Configuration: LiDAR-Visual-Inertial SLAM

The configurations surveyed so far each leave a residual weakness that a single additional sensor would resolve, and LiDAR–visual–inertial (LVIO) fusion is the point at which all three primary sensors are brought together so that the strengths of each cover the failures of the others. This configuration exemplifies the complementarity principle introduced in Section 2.7 in its most comprehensive form. A LiDAR measures accurate, illumination-independent geometry but becomes unobservable in structure-poor scenes and carries no appearance. A camera supplies the dense texture and appearance that disambiguate exactly those geometrically degenerate scenes, but fails under low light, weak texture, and high dynamic range. An IMU measures motion independently of the environment and bridges the transient failures of both, but drifts without correction. Fused tightly, the three constrain the state across a far wider range of conditions than any pair, which makes LVIO the most robust configuration for GPS-denied operation and the one validated in the field, most visibly in the DARPA Subterranean Challenge. Two architectural philosophies organize the modality: tightly-coupled joint estimation, in which all measurements enter a single estimator, and complementary or loosely-coupled designs, in which semi-independent LiDAR–inertial and visual–inertial subsystems cross-constrain one another or hand off as each degrades. The development traces from loosely-coupled origins, through both architectures maturing, to efficient direct fusion, field-validated resilience, and finally the photorealistic neural maps deferred from the preceding sections. Table 5 summarizes the LiDAR-visual-inertial SLAM systems discussed below across their maturity tiers.

8.1 Foundational methods

Early laser–visual–inertial systems adopted loosely coupled, sequential architectures. V-LOAM (Zhang and Singh, 2015) [123] refined a high-rate visual–inertial motion estimate with lower-rate LiDAR scan matching, achieving low-drift, robust odometry by cascading the two rather than jointly estimating them, and established that combining the modalities substantially outperformed either alone. Subsequent work moved toward tightly coupled estimation. LIC-Fusion (Zuo et al., 2019) [124] fused LiDAR edge and plane features, sparse visual features, and inertial measurements in a single MSCKF-based estimator with online spatiotemporal calibration, and its successor LIC-Fusion 2.0 (Zuo et al., 2020) [125], added sliding-window plane-feature tracking for more robust LiDAR registration. Together, these systems established the feasibility of tightly coupled LiDAR–visual–IMU fusion and defined the two principal architectural paradigms that continue to characterize the field: cascaded refinement and joint estimation. In parallel, a separate line of research investigated LiDAR–camera fusion without inertial sensing. These systems exploited LiDAR measurements to provide the sparse depth required to recover the metric scale that a monocular camera alone cannot observe, following the same feature-based, direct, and semi-direct registration paradigms used in visual odometry. Representative examples include the feature-based LIMO (Graeter et al., 2018) [126], the direct DVL-SLAM (Shin et al., 2020) [127] and the semi-direct SDV-LOAM (Yuan et al., 2023) [128]. Low-cost implementations such as CamVox (Zhu et al., 2021) [129] were built on the solid-state LiDARs discussed in Section 6.2. Although these two-sensor systems lie outside the LiDAR–camera–IMU configuration considered in this section, and are accordingly excluded from its maturity tiers, they established the LiDAR-to-vision depth association that underpins subsequent three-sensor fusion methods.

8.2 Established methods

As anticipated in Section 3.4, the interval separating this modality's baseline systems from their successors is too short for the citation-maturity signals of Section 3.3 to draw a stable boundary between established and recent tiers, and we treat them as one. Two systems fixed the architectural templates that most subsequent work follows. LVI-SAM (Shan et al., 2021) [130] realized the complementary philosophy within a factor-graph smoothing framework, coupling a LiDAR–inertial subsystem derived from LIO-SAM with a visual–inertial subsystem derived from VINS-Mono such that each could operate independently and supply the other when it failed. The visual subsystem carries the estimate through LiDAR-degenerate scenes and the LiDAR subsystem through visually degraded ones. R2LIVE (Lin et al., 2021) [131] realized the tightly-coupled philosophy, fusing all measurements in an iterated-EKF estimator refined by factor-graph optimization for high robustness in real time. Its widely used successor R3LIVE (Lin and Zhang, 2022) [132] extended it into a mapping system that reconstructed dense, RGB-colored radiance maps online, the LiDAR–inertial subsystem recovering geometry while the visual subsystem rendered per-point colour. These systems established both the graceful-degradation and joint-estimation architectures

as robust, real-time, and openly available, and R3LIVE in particular anticipated the appearance-rich mapping that would define the frontier.

The tier's later development brought a decisive gain in efficiency and the field validation of the modality's robustness claims. FAST-LIVO (Zheng et al., 2022) [133] fused two tightly-coupled direct subsystems, a sparse-direct LIO and a direct VIO that registered image patches without feature extraction, into an efficient system well suited to resource-constrained platforms, carrying the direct-method line of FAST-LIO2 into the multimodal setting. In parallel, the DARPA Subterranean Challenge drove the development of LVIO systems engineered for sustained perceptual degradation in real GPS-denied environments, and validated the complementarity argument empirically: Super Odometry (Zhao et al., 2021) [134] placed the IMU at the centre of a factor graph and treated LiDAR and visual estimates as constraints upon it, delivering robust odometry through darkness, dust, and geometric self-similarity. CompSLAM (Khattak et al., 2020) [32] fused visual, thermal, LiDAR, and inertial estimates in a hierarchical, complementary architecture that degraded gracefully as individual modalities failed and formed the core of the CERBERUS system that won the Challenge finals. A further direction augmented the fused estimator with proprioceptive kinematics for a specific platform class: VILENS (Wisth et al., 2023) [135] combined visual, inertial, LiDAR, and leg-odometry measurements in a single factor graph for robust state estimation on all-terrain legged robots, at the cost of requiring a legged platform with joint sensing. Together these systems demonstrated, under field and competition conditions, that multi-modal fusion is what sustains localization where any single modality is intermittently blinded, the central premise of this review's attention to GPS-denied operation.

8.3 Frontier methods

The frontier joins the modality's two trajectories, ever-more-efficient fusion and ever-richer maps, and is where the neural, differentiable map deferred from Sections 4 through 7 finally appears in full. FAST-LIVO2 (Zheng et al., 2025) [136] advanced the efficient-fusion line with a unified voxel map and a direct estimator that achieves state-of-the-art accuracy and efficiency, and has become a common odometry backbone for the mapping systems that follow. On the map-representation side, 3D Gaussian splatting has been adopted rapidly, and its photorealistic form emerges specifically here, where the camera supplies the appearance the representation renders: Gaussian-LIC (Lang et al., 2025) [137] and LIV-GaussMap (Hong et al., 2024) [138] co-estimate a Gaussian radiance-field map alongside the LiDAR–inertial–visual trajectory for real-time photorealistic reconstruction. GS-LIVO (Hong et al., 2025) [139] demonstrated such a system running in real time on embedded hardware. Selection across these systems rests on publication venue, open-source adoption, and benchmark standing rather than citation count.

8.4 Trends and outlook

LiDAR–visual–inertial fusion is the most robust and complete configuration surveyed. It is the current endpoint of the sensor-complementarity argument, and its trajectory

Table 5 Summary of *Configuration: LiDAR-Visual-Inertial SLAM* systems surveyed in Section 8. Systems marked † are LiDAR-camera only (no IMU), included as precursors of three-sensor fusion. As they lie outside the modality, they are not assigned a maturity tier.
LC = loop closure.

System	Tier	Architecture	Distinctive contribution	LC
V-LOAM (2015) [123]	F	Loose (cascaded)	VI motion prior refined by LiDAR matching	–
LIMO (2018)† [126]	–	Feature-based	LiDAR depth for monocular scale	–
LIC-Fusion (2019) [124]	F	Tight (MSCKF)	Joint L/V/I fusion with online calibration	–
LIC-Fusion 2.0 (2020) [125]	F	Tight (MSCKF)	Sliding-window plane-feature tracking	–
DVL-SLAM (2020)† [127]	–	Direct	Sparse-depth direct visual-LiDAR SLAM	✓
CamVox (2021)† [129]	–	Feature-based	Low-cost solid-state LiDAR-assisted visual SLAM	✓
SDV-LOAM (2023)† [128]	–	Semi-direct	Semi-direct visual-LiDAR odometry	–
CompSLAM (2020) [32]	E	Hierarchical complem.	V/T/L/I fusion; core of SubT-winning system	–
LVI-SAM (2021) [130]	E	Complementary subsystems	LIO-SAM + VINS-Mono; graceful degradation	✓
Super Odometry (2021) [134]	E	IMU-centric factor graph	Robust through darkness, dust, self-similarity	–
R2LIVE (2021) [131]	E	Tight (iEKF + graph)	Joint estimation, robust in real time	–
R3LIVE (2022) [132]	E	Tight (two subsystems)	Online dense RGB-colored radiance maps	–
FAST-LIVO (2022) [133]	E	Tight (direct)	Sparse-direct LIO + direct VIO; efficiency	–
VILENS (2023) [135]	E	Tight (factor graph)	Adds leg kinematics for legged platforms	–
LIV-GaussMap (2024) [138]	Fr	Tight + 3DGS map	Real-time radiance-field map rendering	–
FAST-LIVO2 (2025) [136]	Fr	Tight (direct)	Unified voxel map; SOTA accuracy/efficiency	–
Gaussian-LIC (2025) [137]	Fr	Tight + 3DGS map	Photorealistic Gaussian map co-estimation	–
GS-LIVO (2025) [139]	Fr	Tight + 3DGS map	Real-time Gaussian mapping on embedded HW	–

points toward unified estimators that treat each sensor as an optional input and toward dense, photorealistic, differentiable maps that serve reconstruction and rendering as well as localization. Its costs are equally real and set practical limits: fusing three sensors demands accurate spatiotemporal calibration and synchronization across them. Additionally, it imposes a computational and size-weight-power burden that stands in direct tension with the minimal visual–inertial pairing of Section 5, and adds little where the environment defeats all of its constituent sensors at once. That last case is not hypothetical: underwater, optical cameras and LiDAR are both attenuated and no amount of fusing them helps. In dense smoke, dust, or adverse weather, both the visual and the LiDAR channels degrade together. These regimes lie outside the reach of even the fullest visual–geometric fusion and are the province of the sensing modalities — acoustic, radar, and thermal — that perceive where light and conventional ranging cannot, which are the subject of Section 9.

9 Configuration: Sonar, Radar and Thermal SLAM

The five configurations of Sections 4 through 8 are built on cameras and LiDAR, and they share a boundary: all of them fail where light is absent or attenuated and where laser ranging cannot penetrate. This section surveys the modalities that perceive precisely in those regimes. Sonar is the natural recourse underwater, the archetypal intrinsically GPS-denied environment, in which optical cameras are range-limited and attenuated by turbidity while satellite positioning is entirely unavailable. Radar occupies a complementary niche above water: its millimeter wavelengths propagate through the rain, snow, fog, dust, and smoke that scatter both visible light and LiDAR pulses, and two sensor families organize its literature, mechanically scanning radars that rotate to capture dense 360° scans, and solid-state single-chip 4D millimeter-wave radars that return sparse point clouds annotated with per-point Doppler velocity. Thermal-infrared cameras image emitted heat rather than reflected light and therefore see in total darkness and penetrate the smoke and dust that blind visible-spectrum cameras, at the cost of imagery that is low-contrast, weakly textured, and subject to sensor drift and periodic non-uniformity correction. This is the heterogeneous residual category anticipated in Section 3.4: three distinct sensing principles with separate research communities, hardware, and historical trajectories. We retain the four maturity tiers used throughout, tracing each modality's line of development within them; because those lines began at different times and advance at different rates, the tier boundaries fall differently for each (Section 3.4). What unites the three is that each extends SLAM into an environment that is GPS-denied not incidentally but intrinsically, and that is closed to the sensors surveyed so far. Consistent with the scope of this review, each is also affected by the field-wide movements already traced: the shift from filtering to optimization, and the recent arrival of learned and neural-map methods. Table 6 consolidates the systems discussed above, making the sparse tier structure of the residual modalities visible at a glance.

9.1 Foundational methods

Foundational work in each of the three modalities addressed the same problem: extracting usable geometry from returns markedly noisier and sparser than those of a camera or a LiDAR.

Acoustic SLAM established the two devices that recur throughout, profiling and imaging sonar for ranging, together with the probabilistic back-ends needed to fuse their returns, and did so across several distinct paradigms. Dense volumetric mapping using octree-based evidence grids was applied to the exploration of underwater tunnel environments (Fairfield et al., 2007) [140]. Extended Kalman filter (EKF) SLAM, driven by mechanically scanned imaging sonar, was employed in structured man-made environments such as marinas and harbours (Ribas et al., 2008) [141]. Pose-graph formulations incorporating forward-looking sonar constraints followed, carrying the smoothing paradigm into the acoustic setting (Li et al., 2018) [142].

Radar SLAM comprises two subfields at markedly different maturities: a mature scanning-radar line that supplies the foundational and established tiers, and an emerging 4D single-chip radar line that appears only in the recent and frontier tiers.

Foundational radar odometry established scan registration for the scanning type, from Fourier-Mellin scan matching (Checchin et al., 2010) [143] to direct phase-correlation methods such as PhaRaO (Park et al., 2020) [144].

Thermal SLAM began with handheld monocular demonstrations (Vidas and Sridharan, 2012) [145], and its decisive foundational contribution for GPS-denied robotics was keyframe-based thermal-inertial odometry (Khattak et al., 2019) [146], which operated directly on raw 16-bit radiometric data to enable aerial navigation in environments that are simultaneously GPS-denied and visually degraded (dark, textureless, and filled with dust or smoke). Contemporaneous work by Delaune et al. (2019) [147] achieved day-and-night thermal-inertial estimation on embedded hardware.

The differing spans are informative in themselves: the foundational period runs from 2007 to 2018 for sonar, from 2010 to 2020 for radar, and from 2012 to 2019 for thermal, so that a single tier label here denotes three different stretches of work.

9.2 Established methods

In each modality the established tier is where a reference system emerged against which subsequent work is measured.

For sonar this is SVIn2 (Rahman et al., 2018, 2019, 2022) [148–150], which fused imaging sonar, camera, IMU, and water-pressure depth in a pose-graph optimization with loop closure, achieving robust operation in the wrecks, caves, and underwater structures where any single sensor is insufficient. The distributed multi-robot setting was addressed by DRACo-SLAM (McConnell et al., 2022) [151] for teams of imaging-sonar-equipped vehicles.

For radar, the scanning line matured into robust large-scale SLAM. RadarSLAM (Hong et al., 2020) [152] demonstrated all-weather operation, and CFEAR (Adolfsson et al., 2021, 2023) [153, 154] achieved efficient, accurate odometry, generalizing across environments from urban to woodland to underground mines without retuning.

For thermal, learning entered early and out of necessity, since low-contrast imagery defeats hand-engineered detectors. Deep thermal feature extractors such as TP-TIO (Zhao et al., 2020) [155] were followed by the first end-to-end trainable thermal-inertial odometry (Saputra et al., 2020) [156] and subsequently by the first complete thermal-inertial SLAM system, combining a neural front-end with graph optimization (Saputra et al., 2022) [157].

9.3 Recent methods

In sonar, neural representations reached the modality, with neural implicit surface reconstruction from imaging sonar (Qadri et al., 2023) [158].

In radar, attention shifted to the compact 4D single-chip sensors, where the Doppler channel aids both ego-motion estimation and dynamic-object rejection. 4DRadarSLAM (Zhang et al., 2023) [159] performs pose-graph SLAM directly on the sparse 4D point clouds, while 4D iRIOM (Zhuang et al., 2023) [160] tightly fuses the radar returns with inertial data for odometry and mapping. A distinct and less mainstream radar type was also explored for vision-denied indoor operation: ultra-wideband (UWB) impulse radar, whose long wavelengths penetrate smoke and are

unaffected by the reflective surfaces that defeat LiDAR. It was used for anchorless, infrastructure-free SLAM by detecting environmental features directly rather than ranging to deployed beacons (Premachandra et al., 2023) [161].

9.4 Frontier methods

At the frontier the three modalities converge, for the first time, on the same directions as the mainstream configurations, and none of this work is yet supported by informative citation records.

Sonar has moved toward tighter multi-sensor coupling and explicit robustness to visual failure. AQUA-SLAM (Xu et al., 2025) [162] tightly couples acoustic, visual, and inertial measurements with dedicated sensor calibration. Additionally, RUSSO (Pan et al., 2025) [163] leans on sonar optimization to hold the estimate as vision degrades.

In radar, the Doppler measurement has increasingly become an organizing principle in its own right, as in Doppler-SLAM (Wang et al., 2025) [164], which fuses 4D radar and FMCW LiDAR with inertial data through a dedicated Doppler-based scan-matching front-end. The measurement noise of radar has motivated the same probabilistic and Gaussian representations seen elsewhere, as radar odometry adopts the summarized scene models now common across modalities [165].

In thermal, post-established work has not yet separated into distinct recent and frontier stages, forming instead a single emerging frontier. Current research focuses on enabling sustained long-term operation despite sensor drift and the periodic non-uniformity corrections that destabilize thermal features (Keil et al., 2024) [166], as well as on developing self-supervised feature learning to reduce reliance on labeled thermal datasets (Self-TIO, Lee et al., 2025) [167]. As in the visual and LiDAR domains, recent work has also begun extending Gaussian-splatting-based mapping to thermal sensing by accounting for the motion blur, rolling-shutter distortion, and fixed-pattern noise characteristic of uncooled thermal cameras [168].

Selection across these threads rests on publication venue, open-source adoption, and benchmark standing rather than citation count.

9.5 Trends and outlook

Sonar, radar, and thermal sensing extend SLAM into the three regimes that defeat the camera-LiDAR configurations of the preceding sections, underwater, adverse weather, and darkness or smoke. They share a common profile: each returns noisier and sparser measurements than an optical sensor, each is supported by smaller datasets and communities, and each is therefore both harder to work with and less mature. Two trajectories are nonetheless clear. First, these modalities are rarely used alone. Their dominant deployment is in fusion with an inertial sensor and, increasingly, with the very cameras and LiDARs whose gaps they fill, so that the complementarity principle of Section 2.7 extends naturally to a fourth and fifth sensor. Thermal sensing accordingly appears in this survey not only standalone but fused, as one channel of the visual-thermal-LiDAR-inertial CompSLAM system that anchored the field-robustness discussion of Section 8.2. Second, the field-wide movements traced throughout this review are reaching them too, with learned front-ends already standard in thermal

Table 6 Summary of *Configuration: Sonar, Radar and Thermal SLAM* systems and methods for the residual modalities surveyed in Section 9.
Entries marked ‡ are modules rather than standalone localization systems.

System	Tier	Sensors	Distinctive contribution
Fairfield et al. (2007) [140]	F	Profiling sonar	Octree evidence grids; underwater tunnels
Ribas et al. (2008) [141]	F	MSIS	EKF-SLAM in structured marinas/harbours
Li et al. (2018) [142]	F	Forward-looking sonar	Pose-graph SLAM with FLS constraints
Checchin et al. (2010) [143]	F	Scanning radar	Fourier–Mellin scan matching
PhaRaO (2020) [144]	F	Scanning radar	Direct phase-correlation odometry
Vidas & Sridharan (2012) [145]	F	Thermal camera	Handheld monocular thermal SLAM
Khattak et al. (2019) [146]	F	Thermal + IMU	Keyframe direct TIO on raw 16-bit radiometry
Delaune et al. (2019) [147]	F	Thermal + IMU	Day/night estimation on embedded hardware
SVIn2 (2018–2022) [148–150]	E	Sonar + camera + IMU + depth	Reference multi-sensor underwater SLAM
DRACo-SLAM (2022) [151]	E	Imaging sonar (teams)	Distributed, communication-efficient SLAM
RadarSLAM (2020) [152]	E	Scanning radar	Large-scale all-weather SLAM
CFEAR (2021, 2023) [153, 154]	E	Scanning radar	Efficient odometry; generalizes without retuning
TP-TIO (2020) [155]	E	Thermal + IMU	Deep thermal feature extraction
Saputra et al. (2022) [157]	E	Thermal + IMU	First complete thermal-inertial SLAM
Qadri et al. (2023) [158]	R	Imaging sonar	Neural implicit surface reconstruction
4DRadarSLAM (2023) [159]	R	4D mmWave radar	Pose-graph SLAM for sparse 4D point clouds
4D iRIOM (2023) [160]	R	4D radar + IMU	4D radar-inertial odometry and mapping
UWB radar SLAM (2023) [161]	R	UWB impulse radar	Anchorless, smoke-penetrating indoor SLAM
Keil et al. (2024) [166]	Fr	Thermal	Long-term operation under thermal drift
AQUA-SLAM (2025) [162]	Fr	Sonar + camera + IMU	Tight coupling with dedicated calibration
RUSSO (2025) [163]	Fr	Sonar + camera	Sonar optimization against visual degradation
Doppler-SLAM (2025) [164]	Fr	4D radar + FMCW LiDAR + IMU	Doppler-based scan-matching front-end
RadarSplat (2025)‡ [165]	–	Scanning radar	Gaussian splatting for radar scene reconstruction
Self-TIO (2025) [167]	Fr	Thermal + IMU	Self-supervised 16-bit feature learning
TRGS-SLAM (2026) [168]	Fr	Thermal + IMU	Gaussian splatting for noisy thermal images

odometry and neural and Gaussian map representations now appearing across all three. Their heterogeneity, and the fact that they are best assessed on different datasets and against different baselines than the mainstream modalities, is one motivation for the cross-cutting treatment of datasets and evaluation that follows.

10 Benchmarking Datasets and Evaluation Metrics

The datasets surveyed in this section were identified through the same procedure used for method selection (Section 3.5). Candidate datasets were gathered by three complementary routes: the benchmarks on which the surveyed systems reported results, traced backward from the method papers of Sections 4-9; dedicated dataset and benchmark papers retrieved from IEEE Xplore, Scopus, and Google Scholar using queries combining each sensing configuration with terms such as *dataset*, *benchmark*, and *ground truth*; and the data-availability statements and evaluation sections of the prior surveys consulted throughout this review. From this pool datasets meeting three criteria were included: public availability to the research community, provision of quantitative ground-truth trajectories or maps suitable for the metrics of Section 10.2, and relevance to at least one of the six sensing configurations under a GPS-denied operating assumption. Consistent with the curated rather than systematic character of this review (Section 3.1), the resulting list is representative rather than exhaustive: it favours the benchmarks most widely adopted within each modality and, for the residual acoustic, radar, and thermal modalities, the small number of datasets that exist at all. The uneven coverage that results is reported as a finding in its own right rather than smoothed over.

The preceding six sections surveyed methods within each sensing configuration. This section reads across them, through the datasets on which the field measures progress, the metrics by which it defines accuracy, and the comparative picture that emerges when the six configurations are placed side by side. A recurring theme organizes the discussion: evaluation in SLAM is fragmented along the same modality lines as the methods, and this fragmentation is itself both a practical obstacle to cross-modal comparison and a symptom of the field's uneven maturity.

10.1 Datasets and benchmarks

The visual and visual–inertial modalities are served by the field's most mature and widely adopted benchmarks. KITTI (Geiger et al., 2012) [169] established outdoor autonomous-driving evaluation with stereo imagery, LiDAR, and GPS/INS ground truth, and remains the reference for both visual and LiDAR odometry. EuRoC MAV (Burri et al., 2016) [170] provides the standard visual–inertial benchmark, with stereo and IMU data from a micro-aerial vehicle and millimeter-accurate motion-capture and laser-tracker ground truth. The TUM RGB-D and TUM VI benchmarks (Sturm et al., 2012, Schubert et al., 2018) [171, 172] anchor RGB-D and hand-held visual–inertial evaluation in indoor settings. For LiDAR and LiDAR–inertial systems, KITTI is complemented by benchmarks with higher-grade ground truth and greater sensor diversity. The Newer College Dataset (Ramezani et al., 2020) [173] pairs handheld LiDAR, inertial, and vision with survey-grade reference maps, and the Hilti SLAM Challenge (Helmberger et al., 2022) [174] targets the tight accuracy demands of construction environments. Broader multi-sensor ground-robot collections such as M2DGR (Yin et al., 2021) [175] extend coverage across scenarios.

For the GPS-denied environments that are the focus of this review, the most comprehensive benchmark is SubT-MRS (Zhao et al., 2024) [176], which was assembled

largely from data collected during the DARPA Subterranean Challenge. The dataset provides time-synchronized LiDAR, RGB camera, thermal camera, and inertial measurements acquired from aerial, legged, and wheeled robotic platforms operating in tunnels, caves, and other confined environments. Importantly, it includes systematically controlled multi-sensor degradation scenarios, ranging from structure-poor corridors to smoke- and dust-filled environments, together with sub-centimeter-accuracy ground truth. As such, SubT-MRS is the closest available benchmark to a truly multimodal, degradation-aware evaluation framework. Furthermore, as discussed below, it introduces a robustness-oriented evaluation methodology that is absent from earlier benchmark suites. In contrast, the remaining sensing modalities discussed in Section 9 are supported primarily by smaller, more recent datasets rather than widely adopted community benchmarks. Examples include SNAIL Radar (Huai et al., 2025) [177] and NTU4DRadLM (Zhang et al., 2023) [178], both developed for 4D radar SLAM. The absence of mature, standardized benchmarks comparable to those established for visual and LiDAR-based localization reflects the comparatively early stage of development of these sensing modalities.

Table 7 Datasets and benchmarks discussed in Section 10.1.
GT = ground truth. Degradation = systematically controlled sensor-degradation scenarios.

Dataset	**Year**	**Platform**	**Sensors**	**Environment / ground truth**	**Degradation**
KITTI [169]	2012	Car	Stereo, LiDAR, GPS/INS	Outdoor driving; GPS/INS GT	–
TUM RGB-D [171]	2012	Handheld/robot	RGB-D	Indoor; motion-capture GT	–
EuRoC MAV [170]	2016	MAV	Stereo, IMU	Indoor; mm-accurate mocap/laser tracker	–
TUM VI [172]	2018	Handheld	Stereo, IMU	Indoor/outdoor; mocap GT (partial)	–
Newer College [173]	2020	Handheld	LiDAR, IMU, vision	Campus; survey-grade reference maps	–
M2DGR [175]	2021	Ground robot	Multi-sensor	Diverse indoor/outdoor scenarios	–
Hilti Challenge [174]	2022	Handheld/robot	Multi-sensor	Construction sites; high-accuracy GT	–
NTU4DRadLM [178]	2023	Ground robots	4D-radar-centric	Urban/campus routes	–
SubT-MRS [176]	2024	Aerial/legged/wheeled	LiDAR, RGB, thermal, IMU	Tunnels, caves, confined; sub-cm GT	✓
SNAIL Radar [177]	2025	Vehicles	4D radar + multi-modal	Large-scale, diverse conditions	✓

10.2 Evaluation metrics

Trajectory accuracy is assessed almost universally by two complementary measures. Absolute Trajectory Error (ATE) measures the root-mean-square translational deviation between the estimated trajectory and the ground truth after alignment by a single rigid or similarity transformation, thereby quantifying global trajectory consistency. Relative Pose Error (RPE) measures the accumulated error over fixed-length trajectory segments, providing an estimate of local drift that is independent of global alignment. For odometry systems, performance is also commonly reported using translational drift as a percentage of distance travelled and rotational drift per unit distance, following the KITTI odometry benchmark convention. Collectively, these metrics assess the accuracy of a localization system under the assumption that tracking is successfully maintained throughout the trajectory.

A note on the scope of this comparison is warranted. This review deliberately refrains from tabulating the trajectory-accuracy figures reported by the surveyed systems, for two reasons. The first is commensurability: published results are obtained under heterogeneous evaluation protocols, such as differing alignment conventions (rigid versus similarity), sequence subsets, loop-closure settings, parameter-tuning effort, and computational platforms, such that a table of self-reported figures would imply a rigor of comparison that the underlying evaluations do not support. The second follows from the argument developed throughout this review: the benchmarks on which such figures are predominantly reported contain few or no controlled degradation scenarios, so the accuracy they measure characterizes benign-condition operation, precisely the regime in which the choice of sensing configuration matters least. For GPS-denied deployment, where the governing question is whether localization is maintained at all under perceptual degradation, nominal accuracy rankings on benign benchmarks are a weak and potentially misleading guide to modality choice. We therefore direct the reader to the original publications and to the community leaderboards of the benchmarks in Table 7 for current quantitative standings, and confine the cross-modal synthesis that follows to the qualitative failure-regime and trade-off analysis that the available evidence can support.

Two developments, however, complicate this evaluation framework. First, the neural and Gaussian-splatting scene representations discussed throughout this review are evaluated not only in terms of localization accuracy but also with respect to map quality and rendering performance. Common measures include reconstruction accuracy and completeness against reference models, together with photometric metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) for novel-view synthesis. These metrics characterize different aspects of system performance than trajectory accuracy. A system may produce highly realistic renderings despite poor localization, or conversely maintain accurate localization while generating lower-quality reconstructions. Consequently, rendering and reconstruction metrics cannot be regarded as substitutes for localization metrics, and the relative ranking of methods depends on the evaluation objective. Second, and more importantly for GPS-denied operation, conventional trajectory accuracy metrics provide little insight into robustness. They quantify performance only for successful runs and do not capture whether, when, or how a system

fails under perceptual degradation. This limitation has motivated the development of field-oriented benchmarks that explicitly evaluate robustness in addition to accuracy. SubT-MRS, for example, introduces dedicated robustness evaluation protocols alongside conventional trajectory metrics, reflecting the practical reality that, in the degraded environments considered throughout this review, maintaining localization is often as important as minimizing localization error.

10.3 Cross-modal synthesis

A comparison of the six localization configurations indicates that no single sensing configuration is universally superior. Rather, each is better suited to particular environmental conditions and robotic platforms. Table 8 summarizes this comparison, and Figure 8 renders the same assessment graphically as per-configuration profiles. Each sensing modality exhibits a characteristic failure regime, and these regimes are largely complementary. Visual estimation deteriorates under low illumination, texture-poor scenes, and high dynamic range. LiDAR-based estimation degrades in geometrically repetitive or structure-deficient environments. Visual–Inertial Odometry (VIO) extends the operational envelope of purely visual methods by exploiting inertial constraints, but its performance ultimately deteriorates when prolonged visual degradation exceeds the temporal support provided by inertial integration. Similarly, LiDAR–Inertial Odometry (LIO) compensates for motion-induced distortion and reduces geometric degeneracy, yet remains fundamentally limited in environments where geometric information is insufficient in all directions. Acoustic, radar, and thermal sensing modalities address operating regimes in which optical and laser-based sensors become ineffective, including underwater environments, adverse weather, complete darkness, and smoke-filled spaces. The largely disjoint nature of these failure modes provides the empirical basis for sensor complementarity. Evidence from the DARPA Subterranean Challenge further demonstrates that fusing sensing modalities with uncorrelated failure characteristics is a key factor in maintaining robust localization under severe perceptual degradation.

Robustness, however, comes at a cost and forms the second axis of comparison. Sensor configurations trade localization accuracy and resilience against computational complexity, calibration requirements, and constraints on size, weight, and power. At one end of the spectrum, a Visual–Inertial configuration is lightweight, energy-efficient, and well suited to resource-constrained aerial platforms, but remains vulnerable to prolonged visual degradation. At the other end, a LiDAR–Visual–Inertial configuration provides the highest level of robustness and environmental observability, although at the expense of increased payload, power consumption, computational demand, and calibration complexity. Consequently, the optimal sensing configuration is application dependent. It should be selected according to the platform's constraints and the dominant sources of perceptual degradation expected in the operating environment, rather than according to any universal ranking of sensor configurations.

Two cross-cutting observations complete this comparison. First, the emergence of neural, differentiable, and photorealistic scene representations, discussed from Section 4 through Section 9, represents a paradigm shift that extends across sensing modalities. However, these approaches are typically evaluated using metrics such as

Table 8 Qualitative comparison of the six sensing configurations for GPS-denied operation, synthesizing Sections 4-9 and 10.3.
SWaP = Size, Weight, and Power.

Configuration	SWaP/cost	Visual degrad.	Geom. degen.	Obscu-rants	Characteristic failure regime
Visual	Low	Low	Moderate	Low	Low light, weak texture, HDR, motion blur; monocular scale unobservable
Visual-inertial	Low	Moderate	Moderate	Low	Sustained visual degradation beyond the inertial horizon
LiDAR	Mod.-high	High	Low	Low	Self-similar or structure-poor geometry; dust, fog, smoke
LiDAR-inertial	Mod.-high	High	Moderate	Low	Prolonged geometric degeneracy exceeding inertial horizon
LiDAR-visual-inertial	High	High	High	Low-mod.	Simultaneous degradation of all channels; calibration/sync burden
Sonar/radar/thermal	Varies	High	Moderate	High	Sparse, noisy returns; limited datasets and maturity

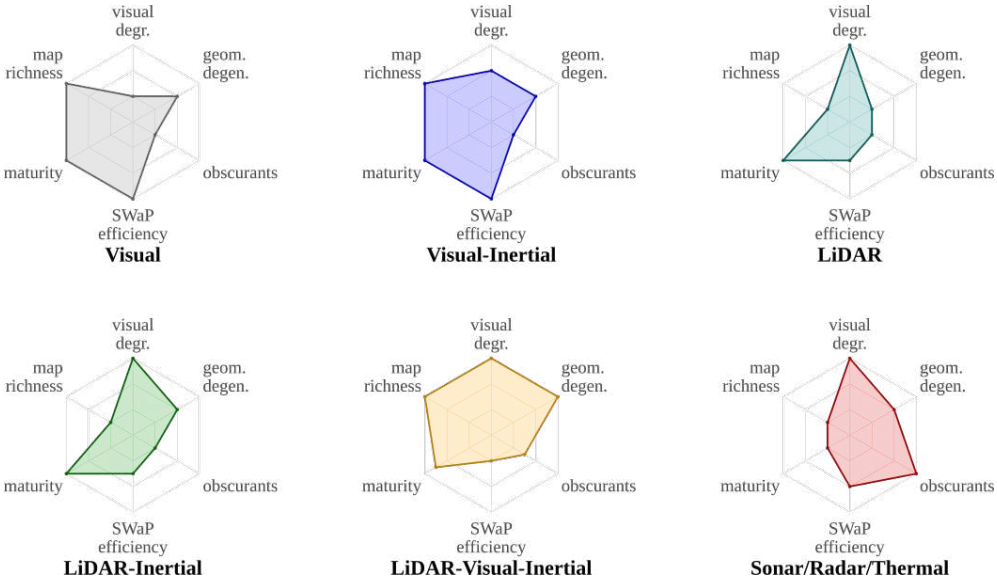


Fig. 8 Qualitative comparison of the six sensing configurations for GPS-denied operation, mirroring Table 8. The first three axes denote robustness to visual degradation, geometric degeneracy, and airborne obscurants or adverse weather. SWaP efficiency denotes a low size, weight, and power burden. Maturity denotes dataset, benchmark, and community support. Map richness denotes the appearance information contained in the produced map. Concentric rings denote low, moderate, and high. Ratings are qualitative syntheses of Sections 4-9 rather than quantitative benchmark results.

rendering fidelity and reconstruction quality, which are only partially aligned with the localization accuracy and robustness criteria used throughout this review. This

divergence in evaluation objectives further complicates the development of a unified framework for comparing localization methods across modalities. Second, the fragmented state of experimental evaluation remains a major methodological challenge. While mature, standardized benchmarks exist for visual and LiDAR-based systems, datasets for other sensing modalities are comparatively limited, isolated, and less established. Moreover, few benchmark suites simultaneously span multiple sensing modalities while systematically varying environmental degradations. As a result, rigorous empirical comparison across modalities remains difficult. The development of unified, multimodal, degradation-aware benchmarks incorporating standardized robustness metrics, of which SubT-MRS is an early example, represents the most promising path toward a comprehensive and empirically grounded evaluation of localization systems.

11 Open Challenges and Future Directions

The six sensing configurations surveyed above are converging toward tightly-coupled fusion, learned front-ends, and neural map representations. Their convergence brings a common set of open problems into focus. This section draws these together, ordered so that the challenges most consequential for GPS-denied operation come first, and emphasizing throughout that they are interdependent rather than separable.

11.1 Robustness under perceptual degradation

The defining unsolved challenge in GPS-denied localization is maintaining reliable state estimation under sustained, simultaneous degradation of all available sensing modalities. As argued throughout this review, sensor fusion derives its strength from combining modalities with complementary and largely uncorrelated failure modes. However, even the most comprehensive LiDAR–Visual–Inertial configurations ultimately fail when every constituent sensor is simultaneously deprived of informative measurements, for example, in long, featureless, smoke-filled corridors under complete darkness, where neither geometric nor visual information is sufficient to constrain the estimation problem.

Recent advances in degeneracy and localizability aware LiDAR–Inertial methods (Section 7.4) represent an important step toward addressing this challenge. By identifying poorly constrained directions in the estimation problem, these methods can detect conditions under which conventional estimators are prone to silent drift. Detection alone, however, does not resolve the underlying lack of observability. A system that recognizes its estimates are becoming unreliable does not thereby recover the missing information required for accurate localization.

A promising direction for future research is therefore the development of localization systems that explicitly reason about their own reliability. Such systems should quantify uncertainty on a per-sensor and per-direction basis, detect the onset of perceptual degradation, adaptively re-weight or switch between sensing modalities, as demonstrated by the complementary and switching architectures discussed in Section 8, and recover gracefully following tracking failure. Although substantial progress has been made in improving localization accuracy, reliable uncertainty

quantification, failure prediction, and recovery remain comparatively underexplored. Advances in these capabilities, rather than incremental improvements in nominal accuracy, are likely to be the key enablers of robust long-term autonomy in GPS-denied environments.

11.2 Neural and differentiable map representations

One of the most significant cross-cutting trends identified in this review is the transition, across all sensing modalities from explicit geometric map representations toward neural implicit and 3D Gaussian-splatting scene representations that are jointly optimized with the estimated trajectory. These approaches offer the prospect of dense, photorealistic, and differentiable scene representations capable of supporting localization, reconstruction, and novel-view synthesis within a unified framework. However, this promise is accompanied by substantial practical and methodological challenges.

First, neural map representations remain computationally and memory intensive, making them difficult to deploy on the size, weight, and power constrained platforms that dominate GPS-denied field robotics. Although systems such as GS-LIVO (Section 8.3) demonstrate that embedded real-time operation is achievable, such implementations remain the exception rather than the norm. Second, most existing methods have been validated only at room or building scale, leaving questions of scalability to large and complex environments, as well as long-term online map maintenance and updating, largely unresolved. Finally, as discussed in Section 10.2, these representations are evaluated not only by localization accuracy but also by rendering fidelity and reconstruction quality. Because these metrics characterize different aspects of system performance, disentangling improvements in mapping quality from improvements in localization remains an open challenge. Consequently, the development of efficient, scalable, and incrementally updatable differentiable scene representations that preserve localization accuracy while remaining compatible with real-time, resource-constrained operation constitutes one of the most important research directions for GPS-denied localization in the coming years.

11.3 Learning, generalization, and foundation models

Another prominent trend identified throughout this review is the increasing integration of learning-based methods into the core of the localization pipeline. Learning is no longer confined to auxiliary perception tasks but now contributes directly to state estimation through learned dense bundle adjustment, learned local feature extraction and matching, learned inertial priors, and learned LiDAR odometry. Despite this progress, a consistent conclusion emerges across all sensing modalities: geometric methods remain the standard for real-world deployment. This reflects, in part, the fact that the generalization performance and failure characteristics of learned components outside their training distribution are not yet sufficiently predictable for safety-critical robotic applications. Consequently, the dominant paradigm remains hybrid rather than fully learned, combining learning-based front ends with geometrically grounded back-end optimization. This architecture leverages the perceptual capabilities of learned models while retaining the interpretability, consistency, and reliability of model-based estimation.

Several research directions naturally follow from this trend. First, improving the generalization of learned components across diverse environments, sensing conditions, and robotic platforms remains a prerequisite for dependable field deployment. Second, self-supervised learning and simulation-based training offer promising approaches to alleviating the scarcity of annotated data, a limitation that is particularly acute for sonar, radar, and thermal sensing modalities (Section 9). Finally, the emergence of large pretrained foundation models for vision and geometric understanding raises the prospect of general-purpose perceptual front ends for SLAM. Whether such models can deliver the reliability, computational efficiency, and real-time performance required for resource-constrained robotic platforms remains an open research question.

11.4 Toward lifelong, semantic, and collaborative operation

Beyond improvements in single-session localization accuracy, three closely related research directions are expanding the scope of SLAM toward persistent, semantic, and distributed spatial intelligence. The first is lifelong SLAM, in which maps are maintained and refined across multiple sessions, seasons, and changing environmental conditions rather than reconstructed from scratch for each deployment. RTAB-Map (Labbé and Michaud, 2019) [179] pioneered scalable memory-management strategies that enable long-term online mapping by limiting the active working map while preserving the ability to perform loop closure. Nevertheless, robust map maintenance under both appearance and structural change, together with principled mechanisms for forgetting obsolete information, remain challenging open problems and are only sparsely represented in the benchmark suites discussed in Section 10. The second direction is semantic and object-level mapping, reflecting a transition from purely geometric representations to maps that encode the semantic structure of the environment. This progression spans from metric-semantic mapping capabilities to hierarchical scene graph representations, with future developments expected to incorporate language-grounded, task-aware, and interactive scene representations capable of supporting higher-level robotic reasoning. The third direction is collaborative multi-robot SLAM. Distributed localization and mapping offer particular advantages in GPS-denied environments, where multiple robots can increase spatial coverage, provide complementary observations, and improve resilience through cooperative state estimation. Achieving communication-efficient, scalable, and robust distributed estimation under realistic bandwidth, latency, and reliability constraints remains an important area for future research.

11.5 Evaluation and benchmarking

Underpinning all of the preceding research directions is the methodological challenge established in Section 10: experimental evaluation remains fragmented along the same modality lines as the methods themselves, and few benchmark suites span multiple sensing modalities while systematically varying environmental degradation. The consequence for the field's research agenda is twofold. First, unified, multi-modal, degradation-aware benchmark suites incorporating standardized measures of localization accuracy, robustness, map quality, computational efficiency, and resource

consumption are needed to enable rigorous and reproducible comparison across sensing configurations. Second, progress on every direction identified in this section depends on such frameworks existing: robustness under degradation (Section 11.1) cannot be demonstrated without controlled degradation scenarios. Additionally, the localization–reconstruction trade-offs of differentiable maps (Section 11.2) cannot be disentangled without metrics that separate them, and the generalization of learned components (Section 11.3) cannot be assessed without evaluation across diverse conditions. Without such frameworks, advances across these directions cannot be measured in a consistent and objective manner, and the identification of genuinely robust localization solutions for GPS-denied environments remains out of reach.

12 Conclusion

This review has surveyed SLAM and odometry for GPS-denied environments through a sensor-centric lens, organizing the literature into six sensing configurations and, within each, into maturity tiers with explicit, per-modality selection criteria. The organizing choice reflects the central finding of the survey: under GPS-denied operation, where no external fix exists to reset accumulated error, the sensors a platform carries determine the environments in which reliable localization is possible, and therefore define the design space of localization systems.

The cross-modal comparison supports a consistent conclusion: no sensing configuration is universally optimal. Each modality fails in its own characteristic regime, and these regimes are largely complementary: vision under low illumination and weak texture, LiDAR under geometric degeneracy, both under obscurants, where acoustic, radar, and thermal sensing take over. Robust GPS-denied localization therefore depends not on maximizing the performance of any single modality, but on fusing sensors with uncorrelated failure modes and on estimators capable of recognizing and responding to their own uncertainty. Reliable uncertainty quantification, degeneracy awareness, adaptive fusion, and graceful recovery from tracking failure are likely to prove more consequential for future GPS-denied autonomy than incremental improvements in nominal accuracy.

As robotic systems are increasingly deployed in subterranean, underwater, indoor, and disaster environments, robust localization from onboard sensing alone becomes a fundamental requirement for autonomy. The next generation of GPS-denied localization systems will ultimately be distinguished not only by how accurately they estimate a trajectory under favorable conditions, but by how reliably they maintain situational awareness when sensing itself becomes uncertain.

Declarations

Funding. This research received no external funding.

Conflict of interest. The authors declare no conflict of interest.

Ethics approval and consent to participate. Not applicable.

Data availability. No new data were created in this study.

Consent for publication. All authors read and approved the final manuscript.

Author contribution. A.G. and P.P. wrote the main manuscript text. A.G. prepared the Tables and Figures. P.P. is A.G.'s advisor. He reviewed and updated the manuscript.

Appendix A Publication-Trend Query Strings

The per-year publication counts shown in Figure 7a were obtained from Scopus in June 2026, using one title-abstract-keyword query per sensing configuration, restricted to the years 2005-2026 and to the subject areas of Engineering and Computer Science. The queries are reproduced verbatim below. Counts should be read as indicative of subfield growth rather than as exact population sizes: the queries overlap, since a publication describing a multi-sensor system may match several of them. Keyword conventions drift over time, so early work is undercounted and database coverage itself evolves. The residual sonar, radar, and thermal category of Section 9 is omitted from Figure 7a because its three heterogeneous sub-modalities do not admit a single query comparable to those below.

- **Visual:** `TITLE-ABS-KEY(("visual SLAM" OR "visual odometry" OR "camera SLAM" OR "monocular SLAM" OR "stereo SLAM") AND NOT ("visual inertial"))`
- **Visual-Inertial:** `TITLE-ABS-KEY("visual inertial odometry" OR "visual inertial SLAM" OR "camera inertial SLAM" OR "VIO")`
- **LiDAR:** `TITLE-ABS-KEY(("LiDAR SLAM" OR "LiDAR odometry" OR "laser SLAM") AND NOT ("LiDAR inertial"))`
- **LiDAR-Inertial:** `TITLE-ABS-KEY("LiDAR inertial odometry" OR "LiDAR inertial SLAM")`
- **LiDAR-Visual-Inertial:** `TITLE-ABS-KEY( "LiDAR visual inertial odometry" OR "LiDAR visual inertial SLAM" OR "LiDAR camera inertial odometry" OR "LiDAR camera inertial SLAM")`

References

[1] Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. MIT press, Cambridge, MA, USA (2005)

[2] Durrant-Whyte, H., Bailey, T.: Simultaneous localization and mapping: part I. IEEE Robotics & Automation Magazine **13**(2), 99–110 (2006) https://doi.org/10.1109/MRA.2006.1638022

[3] Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J.: Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. IEEE Transactions on Robotics **32**(6), 1309–1332 (2016) https://doi.org/10.1109/TRO.2016.2624754

[4] Klein, G., Murray, D.: Parallel Tracking and Mapping for Small AR Workspaces. In: 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 225–234 (2007). https://doi.org/10.1109/ISMAR.2007.4538852

[5] Smith, R., Self, M., Cheeseman, P.: Estimating uncertain spatial relationships in robotics. In: Proceedings. 1987 IEEE International Conference on Robotics and Automation, vol. 4, pp. 850–850 (1987). https://doi.org/10.1109/ROBOT.1987.1087846

[6] Doucet, A., Freitas, N., Murphy, K., Russell, S.: Rao-blackwellised particle filtering for dynamic Bayesian networks. In: Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence. UAI'00, pp. 176–183 (2000)

[7] Montemerlo, M., Thrun, S.: Simultaneous localization and mapping with unknown data association using FastSLAM. In: 2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422), vol. 2, pp. 1985–19912 (2003). https://doi.org/10.1109/ROBOT.2003.1241885

[8] Grisetti, G., Stachniss, C., Burgard, W.: Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters. IEEE Transactions on Robotics **23**(1), 34–46 (2007) https://doi.org/10.1109/TRO.2006.889486

[9] Bailey, T., Nieto, J., Nebot, E.: Consistency of the FastSLAM algorithm. In: Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006., pp. 424–429 (2006). https://doi.org/10.1109/ROBOT.2006.1641748

[10] Dellaert, F.: Square root SAM. In: Robotics: Science and Systems, pp. 177–184 (2005). https://doi.org/10.15607/RSS.2005.I.024

[11] Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle Adjustment — A Modern Synthesis. In: Vision Algorithms: Theory and Practice, pp. 298–372. Springer, Berlin, Heidelberg (2000)

[12] Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J., Dellaert, F.: iSAM2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering. In: 2011 IEEE International Conference on Robotics and Automation, pp. 3281–3288 (2011). https://doi.org/10.1109/ICRA.2011.5979641

[13] Grisetti, G., Kümmerle, R., Stachniss, C., Burgard, W.: A Tutorial on Graph-Based SLAM. IEEE Intelligent Transportation Systems Magazine **2**(4), 31–43 (2010) https://doi.org/10.1109/MITS.2010.939925

[14] Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: G2o: A general framework for graph optimization. In: 2011 IEEE International Conference on Robotics and Automation, pp. 3607–3613 (2011). https://doi.org/10.

1109/ICRA.2011.5979949

[15] Mourikis, A.I., Roumeliotis, S.I.: A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation. In: Proceedings 2007 IEEE International Conference on Robotics and Automation, pp. 3565–3572 (2007). https://doi.org/10.1109/ROBOT.2007.364024

[16] Cummins, M.J., Newman, P.: FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. The International Journal of Robotics Research **27**, 647–665 (2008)

[17] Gálvez-López, D., Tardós, J.D.: Bags of Binary Words for Fast Place Recognition in Image Sequences. IEEE Transactions on Robotics **28**(5), 1188–1197 (2012) https://doi.org/10.1109/TRO.2012.2197158

[18] Kim, G., Kim, A.: Scan Context: Egocentric Spatial Descriptor for Place Recognition Within 3D Point Cloud Map. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4802–4809. IEEE Press, Madrid, Spain (2018). https://doi.org/10.1109/IROS.2018.8593953 . https://doi.org/10.1109/IROS.2018.8593953

[19] Agarwal, P., Tipaldi, G.D., Spinello, L., Stachniss, C., Burgard, W.: Robust map optimization using dynamic covariance scaling. In: 2013 IEEE International Conference on Robotics and Automation, pp. 62–69 (2013). https://doi.org/10.1109/ICRA.2013.6630557

[20] Sünderhauf, N., Protzel, P.: Switchable constraints for robust pose graph SLAM. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1879–1884 (2012). https://doi.org/10.1109/IROS.2012.6385590

[21] Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007). https://doi.org/10.1109/CVPR.2007.383172

[22] Huang, G.: Visual-Inertial Navigation: A Concise Review. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 9572–9582 (2019). https://doi.org/10.1109/ICRA.2019.8793604

[23] Delmerico, J., Scaramuzza, D.: A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 2502–2509. IEEE Press, Brisbane, Australia (2018). https://doi.org/10.1109/ICRA.2018.8460664 . https://doi.org/10.1109/ICRA.2018.8460664

[24] Forster, C., Carlone, L., Dellaert, F., Scaramuzza, D.: On-Manifold Preintegration for Real-Time Visual–Inertial Odometry. IEEE Transactions on Robotics

33(1), 1–21 (2017) https://doi.org/10.1109/TRO.2016.2597321

[25] Hornung, A., Wurm, K.M., Bennewitz, M., Stachniss, C., Burgard, W.: OctoMap: an efficient probabilistic 3D mapping framework based on octrees. Auton. Robots **34**(3), 189–206 (2013) https://doi.org/10.1007/s10514-012-9321-0

[26] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A.: KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. UIST '11, pp. 559–568. Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/2047196.2047270 . https://doi.org/10.1145/2047196.2047270

[27] Oleynikova, H., Taylor, Z., Fehr, M., Siegwart, R., Nieto, J.: Voxblox: Incremental 3D euclidean signed distance fields for on-board MAV planning. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1366–1373 (2017). https://doi.org/10.1109/IROS.2017.8202315

[28] Behley, J., Stachniss, C.: Efficient Surfel-Based SLAM using 3D Laser Range Data in Urban Environments. In: Proc. of Robotics: Science and Systems (RSS) (2018)

[29] Sucar, E., Liu, S., Ortiz, J., Davison, A.J.: iMAP: Implicit Mapping and Positioning in Real-Time. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6209–6218 (2021). https://doi.org/10.1109/ICCV48922.2021.00617

[30] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Transactions on Graphics **42**(4) (2023)

[31] Zhang, J., Kaess, M., Singh, S.: On degeneracy of optimization-based state estimation problems. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 809–816 (2016). https://doi.org/10.1109/ICRA.2016.7487211

[32] Khattak, S., Nguyen, H., Mascarich, F., Dang, T., Alexis, K.: Complementary Multi–Modal Sensor Fusion for Resilient Robot Pose Estimation in Subterranean Environments. In: 2020 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 1024–1029 (2020). https://doi.org/10.1109/ICUAS48674.2020.9213865

[33] Ebadi, K., Bernreiter, L., Biggie, H., Catt, G., Chang, Y., Chatterjee, A., Denniston, C.E., Deschênes, S.-P., Harlow, K., Khattak, S., Nogueira, L., Palieri, M., Petráček, P., Petrlík, M., Reinke, A., Krátký, V., Zhao, S., Agha-mohammadi,

A.-a., Alexis, K., Heckman, C., Khosoussi, K., Kottege, N., Morrell, B., Hutter, M., Pauling, F., Pomerleau, F., Saska, M., Scherer, S., Siegwart, R., Williams, J.L., Carlone, L.: Present and Future of SLAM in Extreme Environments: The DARPA SubT Challenge. IEEE Transactions on Robotics **40**, 936–959 (2024) https://doi.org/10.1109/TRO.2023.3323938

[34] Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: MonoSLAM: Real-Time Single Camera SLAM. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(6), 1052–1067 (2007) https://doi.org/10.1109/TPAMI.2007.1049

[35] Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: Dense tracking and mapping in real-time. In: 2011 International Conference on Computer Vision, pp. 2320–2327 (2011). https://doi.org/10.1109/ICCV.2011.6126513

[36] Kerl, C., Sturm, J., Cremers, D.: Dense visual SLAM for RGB-D cameras. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2100–2106 (2013). https://doi.org/10.1109/IROS.2013.6696650

[37] Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: Large-Scale Direct Monocular SLAM. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision - ECCV 2014, pp. 834–849. Springer, Cham (2014)

[38] Forster, C., Pizzoli, M., Scaramuzza, D.: SVO: Fast Semi-Direct Monocular Visual Odometry. In: IEEE International Conference on Robotics and Automation, pp. 15–22 (2014). https://doi.org/10.1109/ICRA.2014.6906584

[39] Forster, C., Zhang, Z., Gassner, M., Werlberger, M., Scaramuzza, D.: SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems. IEEE Transactions on Robotics **33**(2), 249–265 (2017) https://doi.org/10.1109/TRO.2016.2623335

[40] Mur-Artal, R., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM: A Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics **31**(5), 1147–1163 (2015) https://doi.org/10.1109/TRO.2015.2463671

[41] Mur-Artal, R., Tardós, J.D.: ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. IEEE Transactions on Robotics **33**(5), 1255–1262 (2017) https://doi.org/10.1109/TRO.2017.2705103

[42] Gomez-Ojeda, R., Moreno, F.-A., Zuñiga-Noël, D., Scaramuzza, D., Gonzalez-Jimenez, J.: PL-SLAM: A Stereo SLAM System Through the Combination of Points and Line Segments. IEEE Transactions on Robotics **35**(3), 734–746 (2019) https://doi.org/10.1109/TRO.2019.2899783

[43] Yu, C., Liu, Z., Liu, X.-J., Xie, F., Yang, Y., Wei, Q., Fei, Q.: DS-SLAM: A Semantic Visual SLAM towards Dynamic Environments. In: 2018 IEEE/RSJ

International Conference on Intelligent Robots and Systems (IROS), pp. 1168–1174 (2018). https://doi.org/10.1109/IROS.2018.8593691

[44] Bescos, B., Fácil, J.M., Civera, J., Neira, J.: DynaSLAM: Tracking, Mapping, and Inpainting in Dynamic Scenes. IEEE Robotics and Automation Letters **3**(4), 4076–4083 (2018) https://doi.org/10.1109/LRA.2018.2860039

[45] Engel, J., Koltun, V., Cremers, D.: Direct Sparse Odometry. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(3), 611–625 (2018) https://doi.org/10.1109/TPAMI.2017.2658577

[46] Tateno, K., Tombari, F., Laina, I., Navab, N.: CNN-SLAM: Real-Time Dense Monocular SLAM with Learned Depth Prediction. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6565–6574 (2017). https://doi.org/10.1109/CVPR.2017.695

[47] Wang, S., Clark, R., Wen, H., Trigoni, N.: DeepVO: Towards end-to-end visual odometry with deep Recurrent Convolutional Neural Networks. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 2043–2050 (2017). https://doi.org/10.1109/ICRA.2017.7989236

[48] Teed, Z., Deng, J.: DROID-SLAM: Deep Visual SLAM for Monocular, Stereo, and RGB-D Cameras. Advances in neural information processing systems (2021)

[49] Zhu, Z., Peng, S., Larsson, V., Xu, W., Bao, H., Cui, Z., Oswald, M.R., Pollefeys, M.: NICE-SLAM: Neural Implicit Scalable Encoding for SLAM. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12776–12786 (2022). https://doi.org/10.1109/CVPR52688.2022.01245

[50] Johari, M.M., Carta, C., Fleuret, F.: ESLAM: Efficient Dense SLAM System Based on Hybrid Representation of Signed Distance Fields. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 17408–17419 (2023). https://doi.org/10.1109/CVPR52729.2023.01670

[51] Wang, H., Wang, J., Agapito, L.: Co-SLAM: Joint Coordinate and Sparse Parametric Encodings for Neural Real-Time SLAM. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13293–13302 (2023). https://doi.org/10.1109/CVPR52729.2023.01277

[52] Keetha, N., Karhade, J., Jatavallabhula, K.M., Yang, G., Scherer, S., Ramanan, D., Luiten, J.: SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21357–21366 (2024). https://doi.org/10.1109/CVPR52733.2024.02018

[53] Matsuki, H., Murai, R., Kelly, P.H.J., Davison, A.J.: Gaussian Splatting SLAM. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition

(CVPR), pp. 18039–18048 (2024). https://doi.org/10.1109/CVPR52733.2024.01708

[54] Yan, C., Qu, D., Xu, D., Zhao, B., Wang, Z., Wang, D., Li, X.: GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 19595–19604 (2024). https://doi.org/10.1109/CVPR52733.2024.01853

[55] Huang, H., Li, L., Cheng, H., Yeung, S.-K.: Photo-SLAM: Real-Time Simultaneous Localization and Photorealistic Mapping for Monocular, Stereo, and RGB-D Cameras. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 21584–21593 (2024). https://doi.org/10.1109/CVPR52733.2024.02039

[56] Bloesch, M., Omari, S., Hutter, M., Siegwart, R.: Robust visual inertial odometry using a direct EKF-based approach. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 298–304 (2015). https://doi.org/10.1109/IROS.2015.7353389

[57] Bloesch, M., Burri, M., Omari, S., Hutter, M., Siegwart, R.: Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback. Int. J. Rob. Res. **36**(10), 1053–1072 (2017) https://doi.org/10.1177/0278364917728574

[58] Leutenegger, S., Furgale, P.T., Rabaud, V., Chli, M., Konolige, K., Siegwart, R.Y.: Keyframe-Based Visual-Inertial SLAM using Nonlinear Optimization. In: Robotics: Science and Systems Conference (2013). https://api.semanticscholar.org/CorpusID:13984777

[59] Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., Furgale, P.T.: Keyframe-based visual-inertial odometry using nonlinear optimization. The International Journal of Robotics Research **34**(3), 314–334 (2015)

[60] Qin, T., Li, P., Shen, S.: VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. IEEE Transactions on Robotics **34**(4), 1004–1020 (2018) https://doi.org/10.1109/TRO.2018.2853729

[61] Qin, T., Cao, S., Pan, J., Shen, S.: A General Optimization-based Framework for Global Pose Estimation with Multiple Sensors (2019). https://arxiv.org/abs/1901.03642

[62] Geneva, P., Eckenhoff, K., Lee, W., Yang, Y., Huang, G.: OpenVINS: A Research Platform for Visual-Inertial Estimation. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 4666–4672 (2020). https://doi.org/10.1109/ICRA40945.2020.9196524

[63] Sun, K., Mohta, K., Pfrommer, B., Watterson, M., Liu, S., Mulgaonkar, Y.,

Taylor, C.J., Kumar, V.: Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight. IEEE Robotics and Automation Letters **3**(2), 965–972 (2018) https://doi.org/10.1109/LRA.2018.2793349

[64] Usenko, V., Demmel, N., Schubert, D., Stückler, J., Cremers, D.: Visual-Inertial Mapping With Non-Linear Factor Recovery. IEEE Robotics and Automation Letters **5**(2), 422–429 (2020) https://doi.org/10.1109/LRA.2019.2961227

[65] Rosinol, A., Abate, M., Chang, Y., Carlone, L.: Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 1689–1696 (2020). https://doi.org/10.1109/ICRA40945.2020.9196885

[66] Campos, C., Elvira, R., Rodríguez, J.J.G., M. Montiel, J.M., D. Tardós, J.: ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. IEEE Transactions on Robotics **37**(6), 1874–1890 (2021) https://doi.org/10.1109/TRO.2021.3075644

[67] Mur-Artal, R., Tardós, J.D.: Visual-Inertial Monocular SLAM With Map Reuse. IEEE Robotics and Automation Letters **2**(2), 796–803 (2017) https://doi.org/10.1109/LRA.2017.2653359

[68] He, Y., Zhao, J., Guo, Y., He, W., Yuan, K.: PL-VIO: Tightly-Coupled Monocular Visual–Inertial Odometry Using Point and Line Features. Sensors **18**(4) (2018) https://doi.org/10.3390/s18041159

[69] Stumberg, L.v., Cremers, D.: DM-VIO: Delayed Marginalization Visual-Inertial Odometry. IEEE Robotics and Automation Letters **7**(2), 1408–1415 (2022) https://doi.org/10.1109/LRA.2021.3140129

[70] Leutenegger, S.: OKVIS2: Realtime Scalable Visual-Inertial SLAM with Loop Closure (2022). https://arxiv.org/abs/2202.09199

[71] Clark, R., Wang, S., Wen, H., Markham, A., Trigoni, N.: VINet: visual-inertial odometry as a sequence-to-sequence learning problem. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI'17, pp. 3995–4001. AAAI Press, San Francisco, California, USA (2017)

[72] Chen, C., Rosa, S., Miao, Y., Lu, C.X., Wu, W., Markham, A., Trigoni, N.: Selective Sensor Fusion for Neural Visual-Inertial Odometry. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10534–10543 (2019). https://doi.org/10.1109/CVPR.2019.01079

[73] Liu, W., Caruso, D., Ilg, E., Dong, J., Mourikis, A.I., Daniilidis, K., Kumar, V., Engel, J.: TLIO: Tight Learned Inertial Odometry. IEEE Robotics and Automation Letters **5**(4), 5653–5660 (2020) https://doi.org/10.1109/LRA.2020.3007421

[74] Song, S., Lim, H., Lee, A.J., Myung, H.: DynaVINS: A Visual-Inertial SLAM for Dynamic Environments. IEEE Robotics and Automation Letters **7**(4), 11523–11530 (2022) https://doi.org/10.1109/LRA.2022.3203231

[75] Boche, S., Jung, J., Laina, S.B., Leutenegger, S.: OKVIS2-X: Open Keyframe-Based Visual-Inertial SLAM Configurable With Dense Depth or LiDAR, and GNSS. IEEE Transactions on Robotics **41**, 6064–6083 (2025) https://doi.org/10.1109/TRO.2025.3619051

[76] DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperPoint: Self-Supervised Interest Point Detection and Description. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 337–33712 (2018). https://doi.org/10.1109/CVPRW.2018.00060

[77] Syed, S.N., Roongta, I., Ravie, K., Nageswar, G.: SuperPoint-SLAM3: Augmenting ORB-SLAM3 with Deep Features, Adaptive NMS, and Learning-Based Loop Closure (2026). https://arxiv.org/abs/2506.13089

[78] Vidal, A.R., Rebecq, H., Horstschaefer, T., Scaramuzza, D.: Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High-Speed Scenarios. IEEE Robotics and Automation Letters **3**(2), 994–1001 (2018) https://doi.org/10.1109/LRA.2018.2793357

[79] Chen, P., Guan, W., Lu, P.: ESVIO: Event-Based Stereo Visual Inertial Odometry. IEEE Robotics and Automation Letters **8**(6), 3661–3668 (2023) https://doi.org/10.1109/LRA.2023.3269950

[80] Zhang, J., Singh, S.: LOAM: Lidar Odometry and Mapping in Real-time. In: Robotics: Science and Systems, pp. 1–9 (2014)

[81] Zhang, J., Singh, S.: Low-drift and real-time lidar odometry and mapping. Autonomous Robots **41**(2), 393–404 (2017)

[82] Hess, W., Kohler, D., Rapp, H., Andor, D.: Real-time loop closure in 2D LIDAR SLAM. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 1271–1278 (2016). https://doi.org/10.1109/ICRA.2016.7487258

[83] Shan, T., Englot, B.: LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4758–4765 (2018). https://doi.org/10.1109/IROS.2018.8594299

[84] Wang, H., Wang, C., Chen, C.-L., Xie, L.: F-LOAM : Fast LiDAR Odometry and Mapping. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4390–4396 (2021). https://doi.org/10.1109/IROS51168.2021.9636655

[85] Lin, J., Zhang, F.: Loam livox: A fast, robust, high-precision LiDAR odometry and mapping package for LiDARs of small FoV. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 3126–3131 (2020). https://doi.org/10.1109/ICRA40945.2020.9197440

[86] Deschaud, J.-E.: IMLS-SLAM: Scan-to-Model Matching Based on 3D Data. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 2480–2485 (2018). https://doi.org/10.1109/ICRA.2018.8460653

[87] Dellenbach, P., Deschaud, J.-E., Jacquet, B., Goulette, F.: CT-ICP: Real-time Elastic LiDAR Odometry with Loop Closure. In: 2022 International Conference on Robotics and Automation (ICRA), pp. 5580–5586 (2022). https://doi.org/10.1109/ICRA46639.2022.9811849

[88] Vizzo, I., Guadagnino, T., Mersch, B., Wiesmann, L., Behley, J., Stachniss, C.: KISS-ICP: In Defense of Point-to-Point ICP – Simple, Accurate, and Robust Registration If Done the Right Way. IEEE Robotics and Automation Letters **8**(2), 1029–1036 (2023) https://doi.org/10.1109/LRA.2023.3236571

[89] Pan, Y., Xiao, P., He, Y., Shao, Z., Li, Z.: MULLS: Versatile LiDAR SLAM via Multi-metric Linear Least Square. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 11633–11640 (2021). https://doi.org/10.1109/ICRA48506.2021.9561364

[90] Chen, K., Lopez, B.T., Agha-mohammadi, A.-a., Mehta, A.: Direct LiDAR Odometry: Fast Localization With Dense Point Clouds. IEEE Robotics and Automation Letters **7**(2), 2000–2007 (2022) https://doi.org/10.1109/LRA.2022.3142739

[91] Li, Q., Chen, S., Wang, C., Li, X., Wen, C., Cheng, M., Li, J.: LO-Net: Deep Real-Time Lidar Odometry. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8465–8474 (2019). https://doi.org/10.1109/CVPR.2019.00867

[92] Wang, G., Wu, X., Liu, Z., Wang, H.: PWCLO-Net: Deep LiDAR Odometry in 3D Point Clouds Using Hierarchical Embedding Mask Optimization. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15905–15914 (2021). https://doi.org/10.1109/CVPR46437.2021.01565

[93] Nubert, J., Khattak, S., Hutter, M.: Self-supervised Learning of LiDAR Odometry for Robotic Applications. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 9601–9607 (2021). https://doi.org/10.1109/ICRA48506.2021.9561063

[94] Cattaneo, D., Vaghi, M., Valada, A.: LCDNet: Deep Loop Closure Detection and Point Cloud Registration for LiDAR SLAM. IEEE Transactions on Robotics **38**(4), 2074–2093 (2022) https://doi.org/10.1109/TRO.2022.3150683

[95] Kim, G., Kim, A.: Remove, then Revert: Static Point cloud Map Construction using Multiresolution Range Images. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 10758–10765 (2020). https://doi.org/10.1109/IROS45743.2020.9340856

[96] Lim, H., Hwang, S., Myung, H.: ERASOR: Egocentric Ratio of Pseudo Occupancy-Based Dynamic Object Removal for Static 3D Point Cloud Map Building. IEEE Robotics and Automation Letters **6**(2), 2272–2279 (2021) https://doi.org/10.1109/LRA.2021.3061363

[97] Pan, Y., Zhong, X., Wiesmann, L., Posewsky, T., Behley, J., Stachniss, C.: PIN-SLAM: LiDAR SLAM Using a Point-Based Implicit Neural Representation for Achieving Global Map Consistency. IEEE Transactions on Robotics **40**, 4045–4064 (2024) https://doi.org/10.1109/tro.2024.3422055

[98] Zhong, X., Pan, Y., Behley, J., Stachniss, C.: SHINE-Mapping: Large-Scale 3D Mapping Using Sparse Hierarchical Implicit Neural Representations. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 8371–8377 (2023). https://doi.org/10.1109/ICRA48891.2023.10160907

[99] Isaacson, S., Kung, P.-C., Ramanagopal, M., Vasudevan, R., Skinner, K.A.: Loner: Lidar only neural representations for real-time slam. IEEE Robotics and Automation Letters **8**(12), 8042–8049 (2023)

[100] Deng, J., Wu, Q., Chen, X., Xia, S., Sun, Z., Liu, G., Yu, W., Pei, L.: NeRF-LOAM: Neural Implicit Representation for Large-Scale Incremental LiDAR Odometry and Mapping. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 8184–8193 (2023). https://doi.org/10.1109/ICCV51070.2023.00755

[101] Zhong, X., Pan, Y., Stachniss, C., Behley, J.: 3D LiDAR Mapping in Dynamic Environments Using a 4D Implicit Neural Representation. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15417–15427 (2024). https://doi.org/10.1109/CVPR52733.2024.01460

[102] Pan, Y., Zhong, X., Jin, L., Wiesmann, L., Popović, M., Behley, J., Stachniss, C.: PINGS: Gaussian Splatting Meets Distance Fields within a Point-Based Implicit Neural Map (2025). https://arxiv.org/abs/2502.05752

[103] Geneva, P., Eckenhoff, K., Yang, Y., Huang, G.: LIPS: LiDAR-Inertial 3D Plane SLAM. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 123–130 (2018). https://doi.org/10.1109/IROS.2018.8594463

[104] Ye, H., Chen, Y., Liu, M.: Tightly Coupled 3D Lidar Inertial Odometry and Mapping. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 3144–3150 (2019). https://doi.org/10.1109/ICRA.2019.8793511

[105] Qin, C., Ye, H., Pranata, C.E., Han, J., Zhang, S., Liu, M.: LINS: A Lidar-Inertial State Estimator for Robust and Efficient Navigation. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 8899–8906 (2020). https://doi.org/10.1109/ICRA40945.2020.9197567

[106] Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Rus, D.: LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5135–5142 (2020). https://doi.org/10.1109/IROS45743.2020.9341176

[107] Xu, W., Zhang, F.: FAST-LIO: A Fast, Robust LiDAR-Inertial Odometry Package by Tightly-Coupled Iterated Kalman Filter. IEEE Robotics and Automation Letters **6**(2), 3317–3324 (2021) https://doi.org/10.1109/LRA.2021.3064227

[108] Li, K., Li, M., Hanebeck, U.D.: Towards High-Performance Solid-State-LiDAR-Inertial Odometry and Mapping. IEEE Robotics and Automation Letters **6**(3), 5167–5174 (2021) https://doi.org/10.1109/LRA.2021.3070251

[109] Xu, W., Cai, Y., He, D., Lin, J., Zhang, F.: FAST-LIO2: Fast Direct LiDAR-Inertial Odometry. IEEE Transactions on Robotics **38**(4), 2053–2073 (2022) https://doi.org/10.1109/TRO.2022.3141876

[110] Wang, Z., Zhang, L., Shen, Y., Zhou, Y.: D-LIOM: Tightly-Coupled Direct LiDAR-Inertial Odometry and Mapping. IEEE Transactions on Multimedia **25**, 3905–3920 (2023) https://doi.org/10.1109/TMM.2022.3168423

[111] Bai, C., Xiao, T., Chen, Y., Wang, H., Zhang, F., Gao, X.: Faster-LIO: Lightweight Tightly Coupled Lidar-Inertial Odometry Using Parallel Sparse Incremental Voxels. IEEE Robotics and Automation Letters **7**(2), 4861–4868 (2022) https://doi.org/10.1109/LRA.2022.3152830

[112] He, D., Xu, W., Chen, N., Kong, F., Yuan, C., Zhang, F.: Point-LIO: Robust High-Bandwidth Light Detection and Ranging Inertial Odometry. Advanced Intelligent Systems **5**(7), 2200459 (2023) https://doi.org/10.1002/aisy.202200459

[113] Chen, K., Nemiroff, R., Lopez, B.T.: Direct LiDAR-Inertial Odometry: Lightweight LIO with Continuous-Time Motion Correction. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 3983–3989 (2023). https://doi.org/10.1109/ICRA48891.2023.10160508

[114] Nguyen, T.-M., Duberg, D., Jensfelt, P., Yuan, S., Xie, L.: SLICT: Multi-Input Multi-Scale Surfel-Based Lidar-Inertial Continuous-Time Odometry and Mapping. IEEE Robotics and Automation Letters **8**(4), 2102–2109 (2023) https://doi.org/10.1109/LRA.2023.3246390

[115] Ramezani, M., Khosoussi, K., Catt, G., Moghadam, P., Williams, J., Borges, P.,

Pauling, F., Kottege, N.: Wildcat: Online Continuous-Time 3D Lidar-Inertial SLAM (2022). https://arxiv.org/abs/2205.12595

[116] Qian, C., Xiang, Z., Wu, Z., Sun, H.: RF-LIO: Removal-First Tightly-coupled Lidar Inertial Odometry in High Dynamic Environments. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4421–4428 (2021). https://doi.org/10.1109/IROS51168.2021.9636624

[117] Tuna, T., Nubert, J., Nava, Y., Khattak, S., Hutter, M.: X-ICP: Localizability-Aware LiDAR Registration for Robust Localization in Extreme Environments. IEEE Transactions on Robotics **40**, 452–471 (2024) https://doi.org/10.1109/TRO.2023.3335691

[118] Lee, E.M., Marsim, K.C., Myung, H.: LODESTAR: Degeneracy-Aware LiDAR-Inertial Odometry With Adaptive Schmidt-Kalman Filter and Data Exploitation. IEEE Robotics and Automation Letters **11**(1), 922–929 (2026) https://doi.org/10.1109/LRA.2025.3634906

[119] Pfreundschuh, P., Oleynikova, H., Cadena, C., Siegwart, R., Andersson, O.: COIN-LIO: Complementary Intensity-Augmented LiDAR Inertial Odometry. In: 2024 IEEE International Conference on Robotics and Automation (ICRA), pp. 1730–1737 (2024). https://doi.org/10.1109/ICRA57147.2024.10610938

[120] Júnior, G.P.C., Rezende, A.M.C., Miranda, V.R.F., Fernandes, R., Azpúrua, H., Neto, A.A., Pessin, G., Freitas, G.M.: EKF-LOAM: An Adaptive Fusion of LiDAR SLAM With Wheel Odometry and Inertial Data for Confined Spaces With Few Geometric Features. IEEE Transactions on Automation Science and Engineering **19**(3), 1458–1471 (2022) https://doi.org/10.1109/TASE.2022.3169442

[121] Ji, X., Yuan, S., Yin, P., Xie, L.: LIO-GVM: An Accurate, Tightly-Coupled Lidar-Inertial Odometry With Gaussian Voxel Map. IEEE Robotics and Automation Letters **9**(3), 2200–2207 (2024) https://doi.org/10.1109/LRA.2024.3354616

[122] Giacomini, E., Giammarino, L.D., Rebotti, L.D., Grisetti, G., Oswald, M.R.: Splat-LOAM: Gaussian Splatting LiDAR Odometry and Mapping (2025). https://arxiv.org/abs/2503.17491

[123] Zhang, J., Singh, S.: Visual-lidar odometry and mapping: low-drift, robust, and fast. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 2174–2181 (2015). https://doi.org/10.1109/ICRA.2015.7139486

[124] Zuo, X., Geneva, P., Lee, W., Liu, Y., Huang, G.: LIC-Fusion: LiDAR-Inertial-Camera Odometry. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5848–5854. IEEE Press, Macau, China (2019). https://doi.org/10.1109/IROS40897.2019.8967746 .

https://doi.org/10.1109/IROS40897.2019.8967746

[125] Zuo, X., Yang, Y., Geneva, P., Lv, J., Liu, Y., Huang, G., Pollefeys, M.: LIC-Fusion 2.0: LiDAR-Inertial-Camera Odometry with Sliding-Window Plane-Feature Tracking. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5112–5119 (2020). https://doi.org/10.1109/IROS45743.2020.9340704

[126] Graeter, J., Wilczynski, A., Lauer, M.: LIMO: Lidar-Monocular Visual Odometry. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 7872–7879 (2018). https://doi.org/10.1109/IROS.2018.8594394

[127] Shin, Y.-S., Park, Y.S., Kim, A.: DVL-SLAM: sparse depth enhanced direct visual-LiDAR SLAM. Auton. Robots **44**(2), 115–130 (2020) https://doi.org/10.1007/s10514-019-09881-0

[128] Yuan, Z., Wang, Q., Cheng, K., Hao, T., Yang, X.: SDV-LOAM: Semi-Direct Visual–LiDAR Odometry and Mapping. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(9), 11203–11220 (2023) https://doi.org/10.1109/TPAMI.2023.3262817

[129] Zhu, Y., Zheng, C., Yuan, C., Huang, X., Hong, X.: CamVox: A Low-cost and Accurate Lidar-assisted Visual SLAM System. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 5049–5055 (2021). https://doi.org/10.1109/ICRA48506.2021.9561149

[130] Shan, T., Englot, B., Ratti, C., Rus, D.: LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 5692–5698 (2021). https://doi.org/10.1109/ICRA48506.2021.9561996

[131] Lin, J., Zheng, C., Xu, W., Zhang, F.: R ^{2} LIVE: A Robust, Real-Time, LiDAR-Inertial-Visual Tightly-Coupled State Estimator and Mapping. IEEE Robotics and Automation Letters **6**(4), 7469–7476 (2021) https://doi.org/10.1109/LRA.2021.3095515

[132] Lin, J., Zhang, F.: R3LIVE: A Robust, Real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state Estimation and mapping package. In: 2022 International Conference on Robotics and Automation (ICRA), pp. 10672–10678 (2022). https://doi.org/10.1109/ICRA46639.2022.9811935

[133] Zheng, C., Zhu, Q., Xu, W., Liu, X., Guo, Q., Zhang, F.: FAST-LIVO: Fast and Tightly-coupled Sparse-Direct LiDAR-Inertial-Visual Odometry. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4003–4009 (2022). https://doi.org/10.1109/IROS47612.2022.9981107

[134] Zhao, S., Zhang, H., Wang, P., Nogueira, L., Scherer, S.: Super Odometry: IMU-centric LiDAR-Visual-Inertial Estimator for Challenging Environments. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 8729–8736 (2021). https://doi.org/10.1109/IROS51168.2021.9635862

[135] Wisth, D., Camurri, M., Fallon, M.: VILENS: Visual, Inertial, Lidar, and Leg Odometry for All-Terrain Legged Robots. IEEE Transactions on Robotics **39**(1), 309–326 (2023) https://doi.org/10.1109/TRO.2022.3193788

[136] Zheng, C., Xu, W., Zou, Z., Hua, T., Yuan, C., He, D., Zhou, B., Liu, Z., Lin, J., Zhu, F., Ren, Y., Wang, R., Meng, F., Zhang, F.: FAST-LIVO2: Fast, Direct LiDAR–Inertial–Visual Odometry. IEEE Transactions on Robotics **41**, 326–346 (2025) https://doi.org/10.1109/TRO.2024.3502198

[137] Lang, X., Li, L., Wu, C., Zhao, C., Liu, L., Liu, Y., Lv, J., Zuo, X.: Gaussian-LIC: Real-Time Photo-Realistic SLAM with Gaussian Splatting and LiDAR-Inertial-Camera Fusion. In: 2025 IEEE International Conference on Robotics and Automation (ICRA), pp. 8500–8507 (2025). https://doi.org/10.1109/ICRA55743.2025.11128712

[138] Hong, S., He, J., Zheng, X., Zheng, C.: LIV-GaussMap: LiDAR-Inertial-Visual Fusion for Real-Time 3D Radiance Field Map Rendering. IEEE Robotics and Automation Letters **9**(11), 9765–9772 (2024) https://doi.org/10.1109/LRA.2024.3400149

[139] Hong, S., Zheng, C., Shen, Y., Li, C., Zhang, F., Qin, T., Shen, S.: GS-LIVO: Real-Time LiDAR, Inertial, and Visual Multisensor Fused Odometry With Gaussian Mapping. IEEE Transactions on Robotics **41**, 4253–4268 (2025) https://doi.org/10.1109/TRO.2025.3582809

[140] Fairfield, N., Kantor, G., Wettergreen, D.: Real-Time SLAM with Octree Evidence Grids for Exploration in Underwater Tunnels. Journal of Field Robotics **24**(1-2), 03–21 (2007) https://doi.org/10.1002/rob.20165

[141] Ribas, D., Ridao, P., Tardós, J.D., Neira, J.: Underwater SLAM in man-made structured environments. Journal of Field Robotics **25**(11-12), 898–921 (2008) https://doi.org/10.1002/rob.20249

[142] Li, J., Kaess, M., Eustice, R.M., Johnson-Roberson, M.: Pose-Graph SLAM Using Forward-Looking Sonar. IEEE Robotics and Automation Letters **3**(3), 2330–2337 (2018) https://doi.org/10.1109/LRA.2018.2809510

[143] Checchin, P., Gérossier, F., Blanc, C., Chapuis, R., Trassoudaine, L.: Radar Scan Matching SLAM Using the Fourier-Mellin Transform. In: Howard, A., Iagnemma, K., Kelly, A. (eds.) Field and Service Robotics, pp. 151–161. Springer, Berlin, Heidelberg (2010)

[144] Park, Y.S., Shin, Y.-S., Kim, A.: PhaRaO: Direct Radar Odometry using Phase Correlation. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 2617–2623 (2020). https://doi.org/10.1109/ICRA40945.2020.9197231

[145] Vidas, S., Sridharan, S.: Hand-held monocular SLAM in thermal-infrared. In: 2012 12th International Conference on Control Automation Robotics & Vision (ICARCV), pp. 859–864 (2012). https://doi.org/10.1109/ICARCV.2012.6485270

[146] Khattak, S., Papachristos, C., Alexis, K.: Keyframe-based Direct Thermal–Inertial Odometry. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 3563–3569 (2019). https://doi.org/10.1109/ICRA.2019.8793927

[147] Delaune, J., Hewitt, R., Lytle, L., Sorice, C., Thakker, R., Matthies, L.: Thermal-Inertial Odometry for Autonomous Flight Throughout the Night. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1122–1128 (2019). https://doi.org/10.1109/IROS40897.2019.8968238

[148] Rahman, S., Li, A.Q., Rekleitis, I.: Sonar Visual Inertial SLAM of Underwater Structures. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 5190–5196 (2018). https://doi.org/10.1109/ICRA.2018.8460545

[149] Rahman, S., Li, A.Q., Rekleitis, I.: SVIn2: An Underwater SLAM System using Sonar, Visual, Inertial, and Depth Sensor. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1861–1868 (2019). https://doi.org/10.1109/IROS40897.2019.8967703

[150] Rahman, S., Li, A.Q., Rekleitis, I.: SVIn2: A multi-sensor fusion-based underwater SLAM system. The International Journal of Robotics Research **41**(11-12), 1022–1042 (2022) https://doi.org/10.1177/02783649221110259

[151] McConnell, J., Huang, Y., Szenher, P., Collado-Gonzalez, I., Englot, B.: DRACo-SLAM: Distributed Robust Acoustic Communication-efficient SLAM for Imaging Sonar Equipped Underwater Robot Teams. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 8457–8464 (2022). https://doi.org/10.1109/IROS47612.2022.9981822

[152] Hong, Z., Petillot, Y., Wang, S.: RadarSLAM: Radar based Large-Scale SLAM in All Weathers. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5164–5170 (2020). https://doi.org/10.1109/IROS45743.2020.9341287

[153] Adolfsson, D., Magnusson, M., Alhashimi, A., Lilienthal, A.J., Andreasson, H.: Cfear radarodometry - conservative filtering for efficient and accurate radar odometry. In: 2021 IEEE/RSJ International Conference on Intelligent

Robots and Systems (IROS), pp. 5462–5469 (2021). https://doi.org/10.1109/IROS51168.2021.9636253

[154] Adolfsson, D., Magnusson, M., Alhashimi, A., Lilienthal, A.J., Andreasson, H.: Lidar-Level Localization With Radar? The CFEAR Approach to Accurate, Fast, and Robust Large-Scale Radar Odometry in Diverse Environments. IEEE Transactions on Robotics **39**(2), 1476–1495 (2023) https://doi.org/10.1109/TRO.2022.3221302

[155] Zhao, S., Wang, P., Zhang, H., Fang, Z., Scherer, S.: TP-TIO: A Robust Thermal-Inertial Odometry with Deep ThermalPoint. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4505–4512 (2020). https://doi.org/10.1109/IROS45743.2020.9341716

[156] Saputra, M.R.U., Gusmao, P.P.B., Lu, C.X., Almalioglu, Y., Rosa, S., Chen, C., Wahlström, J., Wang, W., Markham, A., Trigoni, N.: Deeptio: A deep thermal-inertial odometry with visual hallucination. IEEE Robotics and Automation Letters **5**(2), 1672–1679 (2020) https://doi.org/10.1109/LRA.2020.2969170

[157] Saputra, M.R.U., Lu, C.X., Gusmao, P.P.B., Wang, B., Markham, A., Trigoni, N.: Graph-Based Thermal–Inertial SLAM With Probabilistic Neural Networks. IEEE Transactions on Robotics **38**(3), 1875–1893 (2022) https://doi.org/10.1109/TRO.2021.3120036

[158] Qadri, M., Kaess, M., Gkioulekas, I.: Neural Implicit Surface Reconstruction using Imaging Sonar. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 1040–1047 (2023). https://doi.org/10.1109/ICRA48891.2023.10161206

[159] Zhang, J., Zhuge, H., Wu, Z., Peng, G., Wen, M., Liu, Y., Wang, D.: 4DRadarSLAM: A 4D Imaging Radar SLAM System for Large-scale Environments based on Pose Graph Optimization. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 8333–8340 (2023). https://doi.org/10.1109/ICRA48891.2023.10160670

[160] Zhuang, Y., Wang, B., Huai, J., Li, M.: 4D iRIOM: 4D Imaging Radar Inertial Odometry and Mapping. IEEE Robotics and Automation Letters **8**(6), 3246–3253 (2023) https://doi.org/10.1109/LRA.2023.3266669

[161] Premachandra, H.A.G.C., Liu, R., Yuen, C., Tan, U.-X.: UWB Radar SLAM: An Anchorless Approach in Vision Denied Indoor Environments. IEEE Robotics and Automation Letters **8**(9), 5299–5306 (2023) https://doi.org/10.1109/LRA.2023.3293354

[162] Xu, S., Zhang, K., Wang, S.: AQUA-SLAM: Tightly Coupled Underwater Acoustic-Visual-Inertial SLAM With Sensor Calibration. IEEE Transactions on Robotics **41**, 2785–2803 (2025) https://doi.org/10.1109/TRO.2025.3554396

[163] Pan, S., Hong, Z., Hu, Z., Xu, X., Lu, W., Hu, L.: RUSSO: Robust Underwater SLAM With Sonar Optimization Against Visual Degradation. IEEE/ASME Transactions on Mechatronics **30**(6), 5456–5467 (2025) https://doi.org/10.1109/TMECH.2025.3550730

[164] Wang, D., Haag, H., Herraez, D.C., May, S., Stachniss, C., Nüchter, A.: Doppler-SLAM: Doppler-Aided Radar-Inertial and LiDAR-Inertial Simultaneous Localization and Mapping. IEEE Robotics and Automation Letters **10**(9), 9438–9445 (2025) https://doi.org/10.1109/LRA.2025.3595071

[165] Kung, P.-C., Harisha, S., Vasudevan, R., Eid, A., Skinner, K.A.: Radarsplat: Radar gaussian splatting for high-fidelity data synthesis and 3d reconstruction of autonomous driving scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 27596–27606 (2025)

[166] Keil, C., Gupta, A., Kaveti, P., Singh, H.: Towards Long Term SLAM on Thermal Imagery. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 10174–10181 (2024). https://doi.org/10.1109/IROS58592.2024.10802139

[167] Lee, J., Ando, T., Shinozaki, M., Kitajima, T., An, Q., Yamashita, A.: Self-TIO: Thermal-Inertial Odometry via Self-Supervised 16-Bit Feature Extractor and Tracker. IEEE Robotics and Automation Letters **10**(2), 1003–1010 (2025) https://doi.org/10.1109/LRA.2024.3518303

[168] Carmichael, S., Skinner, K.A.: TRGS-SLAM: IMU-Aided Gaussian Splatting SLAM for Blurry, Rolling Shutter, and Noisy Thermal Images (2026). https://arxiv.org/abs/2603.20443

[169] Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354–3361 (2012). https://doi.org/10.1109/CVPR.2012.6248074

[170] Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R.: The EuRoC micro aerial vehicle datasets. The International Journal of Robotics Research **35**(10), 1157–1163 (2016) https://doi.org/10.1177/0278364915620033

[171] Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of RGB-D SLAM systems. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 573–580 (2012). https://doi.org/10.1109/IROS.2012.6385773

[172] Schubert, D., Goll, T., Demmel, N., Usenko, V., Stückler, J., Cremers, D.: The TUM VI Benchmark for Evaluating Visual-Inertial Odometry. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),

pp. 1680–1687 (2018). https://doi.org/10.1109/IROS.2018.8593419

[173] Ramezani, M., Wang, Y., Camurri, M., Wisth, D., Mattamala, M., Fallon, M.: The Newer College Dataset: Handheld LiDAR, Inertial and Vision with Ground Truth. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4353–4360 (2020). https://doi.org/10.1109/IROS45743.2020.9340849

[174] Helmberger, M., Morin, K., Berner, B., Kumar, N., Cioffi, G., Scaramuzza, D.: The Hilti SLAM Challenge Dataset. IEEE Robotics and Automation Letters **7**(3), 7518–7525 (2022) https://doi.org/10.1109/LRA.2022.3183759

[175] Yin, J., Li, A., Li, T., Yu, W., Zou, D.: M2dgr: A multi-sensor and multi-scenario slam dataset for ground robots. IEEE Robotics and Automation Letters **7**(2), 2266–2273 (2021)

[176] Zhao, S., Gao, Y., Wu, T., Singh, D., Jiang, R., Sun, H., Sarawata, M., Qiu, Y., Whittaker, W., Higgins, I., Du, Y., Su, S., Xu, C., Keller, J., Karhade, J., Nogueira, L., Saha, S., Zhang, J., Wang, W., Wang, C., Scherer, S.: SubT-MRS Dataset: Pushing SLAM Towards All-weather Environments. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22647–22657 (2024). https://doi.org/10.1109/CVPR52733.2024.02137

[177] Huai, J., Wang, B., Zhuang, Y., Chen, Y., Li, Q., Han, Y.: SNAIL radar: A large-scale diverse benchmark for evaluating 4D-radar-based SLAM. The International Journal of Robotics Research **44**(12), 1941–1958 (2025) https://doi.org/10.1177/02783649251329048

[178] Zhang, J., Zhuge, H., Liu, Y., Peng, G., Wu, Z., Zhang, H., Lyu, Q., Li, H., Zhao, C., Kircali, D., Mharolkar, S., Yang, X., Yi, S., Wang, Y., Wang, D.: NTU4DRadLM: 4D Radar-Centric Multi-Modal Dataset for Localization and Mapping. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), pp. 4291–4296 (2023). https://doi.org/10.1109/ITSC57777.2023.10422606

[179] Labbé, M., Michaud, F.: RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. Journal of Field Robotics **36**(2), 416–446 (2019) https://doi.org/10.1002/rob.21831