Metric-Topological Semantic Mapping and Inverse Perspective Ground Plane Optimization for Autonomous Fleet Logistics in Cluttered Warehouses

Kanishka Bandara Senanayake[1], **Dilshan Priyadarshana Jayasuriya**[2], **and Nuwan Chaminda Wijesinghe**[3]
[1]Department of Electronic and Telecommunication Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka
[2]Department of Computer Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya 20400, Sri Lanka
[3]Department of Electrical and Information Engineering, Faculty of Engineering, University of Ruhuna, Galle 80000, Sri Lanka

ABSTRACT Large-scale automated fulfillment facilities and distribution hubs depend heavily on coordinated fleets of autonomous guided vehicles (AGVs) and autonomous mobile robots (AMRs). However, maintaining drift-free, centimeter-level localization across expansive warehouse corridors remains deeply challenging. Long, geometrically degenerate storage aisles cause standard planar LiDAR scan matchers to slip longitudinally, while continuous pallet shifting and human pedestrian traffic corrupt static occupancy maps. This paper presents a unified metric-topological factor graph optimization architecture tailored for multi-robot intralogistics in cluttered industrial environments. The system couples 2D planar LiDAR odometry and wheel kinematics with a downward-tilted monocular vision pipeline that computes inverse perspective mapping (IPM) transformations over ground plane visual primitives, including floor lane tape, aisle crossing intersections, and structural floor expansion seams. Extracted ground line primitives and topological junction nodes are integrated directly as geometric constraints into an incremental Bayes tree factor graph. In addition, an inter-robot sparse graph reduction protocol guarantees bounded communication overhead during multi-agent loop closures across decentralized fleets. Comprehensive empirical validation was conducted across an active $24{,}000\,\mathrm{m}^2$ logistics warehouse testbed comprising four challenging operating environments. The proposed metric-topological framework achieves an average horizontal localization error of 3.42 cm and longitudinal slip suppression below 1.85 cm, outperforming state-of-the-art LiDAR-only and visual SLAM baselines by over 58%. Multi-vehicle coordination trials demonstrated deterministic solver convergence under 12.4 ms with communication payloads bounded beneath 3.8 kB/s per agent across fleets of up to fifty robots.

KEYWORDS Autonomous guided vehicles, factor graph SLAM, inverse perspective mapping, warehouse logistics, metric-topological navigation

1. Introduction

The modern global economy relies fundamentally on high-throughput logistics facilities, automated fulfillment hubs, and freight distribution centers to manage the escalating volumes of commercial trade and rapid e-commerce package sorting [1]. Within these vast industrial complexes, which frequently span tens of thousands of square meters, the continuous, autonomous transport of goods between high-bay pallet storage racks, depalletizing zones, and outbound shipping docks is executed by heterogeneous fleets of Autonomous Guided Vehicles (AGVs) and Autonomous Mobile Robots (AMRs) [2]. To maximize spatial storage density and operational warehouse capacity, storage facilities are engineered with tall, narrow storage aisles whose horizontal clearances between passing robotic vehicles and vertical steel racking posts often measure less than twenty centimeters. Operating safely within these constrained geometric channels requires mobile robotic units to execute continuous trajectory tracking at forward transit velocities between one and two meters per second, demanding real-time position and orientation estimation whose cumulative drift remains strictly bounded within several centimeters [3].

Early industrial AGV deployments achieved operational repeatability through physical modifications of the factory floor, installing active inductive wire loops, embedded magnetic guidance tapes, or reflective bar-coded optical fiducials spaced along predetermined travel paths [4]. While such fixed-guidance architectures provide predictable trajectory tracking under nominal conditions, they impose severe operational drawbacks. The civil engineering costs associated with cutting floor trenches or affixing magnetic strips across millions of square meters are substantial. In addition, physical tapes degrade rapidly under the abrasive friction of heavy counterbalance forklifts and pallet transport vehicles. Most critically, fixed-guidance systems enforce rigid, static topological routing maps; reconfiguring a warehouse path network to accommodate seasonal SKU demands, layout optimizations, or facility expansions requires halting operations to physically relay or reprogram floor guides [5].

To circumvent the rigidity of fixed infrastructure, modern robotics research has focused extensively on natural feature Simultaneous Localization and Mapping (SLAM), utilizing onboard perceptual sensors to build spatial representations of the operating environment while concurrently tracking vehicle poses [6]. In indoor logistics environments, horizontal two-dimensional planar light detection and ranging (LiDAR) scanners and wheel encoder odometry serve as the conventional sensor suite for industrial fleet operations [7]. Planar LiDAR sensors deliver high-rate, millimetric range measurements across a wide field of view, functioning robustly under varying ambient lighting conditions and providing active illumination that remains immune to indoor shadows. Using iterative closest point algorithms, correlation-based scan matching, or non-linear factor graph optimization, robotic vehicles can register successive laser scans against an evolving occupancy grid map to maintain pose tracking [8, 9].

1.1. Long-Corridor Geometric Degeneracy in Warehouse Environments

Despite the proven utility of planar LiDAR scan matching in structured office and domestic environments, industrial warehouse aisles present an extreme perceptual vulnerability known as geometric degeneracy [10]. High-bay warehouse aisles consist of long, straight transit corridors flanked on both sides by identical steel upright columns, longitudinal horizontal crossbeams, and uniform pallet packaging. When a planar LiDAR sweeps horizontally within these uniform corridors, the laser beams reflected from the parallel racking uprights provide strong geometric constraints in the transverse coordinate axis (lateral corridor spacing) and in the vehicle heading angle (yaw alignment with the corridor walls). However, along the longitudinal axis of the corridor (the primary direction of vehicle transit), the measured range profile remains virtually invariant as the vehicle advances.

Under standard non-linear least-squares scan registration frameworks, the absence of longitudinal range gradients causes the spatial Fisher Information Matrix to lose rank along the along-track direction [11]. Consequently, the optimization surface becomes degenerate: a continuous valley of near-identical scan matching residuals forms along the corridor axis. In the presence of minor measurement noise, wheel slip transients, or floor undulations, scan matching algorithms fail to constrain longitudinal motion, allowing the estimated robot pose to slip uncontrollably along the corridor length. Because the lateral and angular errors remain tightly constrained by the rack boundaries, standard covariance monitoring heuristics often fail to detect this longitudinal slippage until the vehicle arrives at a corridor exit or intersection, where the sudden appearance of perpendicular walls produces massive innovation residuals that cause catastrophic filter divergence, path replanning failures, or collisions with crossing traffic [12].

1.2. Dynamic Clutter and Perceptual Aliasing in Large-Scale Distribution Facilities

The navigational complexity of industrial logistics facilities is compounded by continuous environmental non-stationarity and structural perceptual aliasing [13]. Modern distribution hubs are dynamic operational environments: pallets of goods, wire-mesh containment bins, and mobile shipping carts are loaded, unloaded, stacked, and relocated throughout operational shifts. Standard occupancy grid SLAM algorithms formulate spatial mapping under the foundational mathematical assumption that the observed environment is static [6]. When dynamic forklift trucks cross an AGV's sensory field of view, or when temporary pallet stacks are placed within previously mapped aisles, classical LiDAR scan matchers treat these transient obstacles as persistent geometric anchors. This mismatch introduces systematic bias into the scan alignment objective, causing distorted map generation, false loop closure rejections, and trajectory estimation failure [8].

Additionally, the physical symmetry inherent in warehouse construction induces severe perceptual aliasing [14]. In automated fulfillment centers, dozens of parallel storage aisles feature identical physical dimensions, identical steel column spacing, and identical shelving architecture. If an AGV temporarily loses wheel traction or experiences a sensor interruption while executing a corridor traversal, a scan-to-map relocalization query may return multiple identical likelihood modes across distinct physical aisles [15]. In single-hypothesis filtering frameworks, converging onto an incorrect corridor hypothesis results in severe localization errors that propagate through the fleet scheduling system, creating route conflicts, automated deadlock, and emergency brake stops that diminish facility throughput.

1.3. Ground Surface Visual Markings as Metric-Topological Invariants

To resolve the geometric degeneracy of planar LiDAR and overcome dynamic racking clutter without reverting to artificial guidance hardware, warehouse floors possess a rich, highly structured, and persistent source of invariant spatial information: surface paint markings and architectural concrete expansion seams [5]. Standard industrial facility codes mandate prominent floor markings to organize material handling operations, delineate pedestrian safety walkways, mark pallet staging zones, and define vehicle transit lanes. These markings include continuous yellow or white corridor boundary lines, dashed lane dividers, perpendicular stop bars at intersections, pedestrian zebra crosswalks, and high-visibility rectangular staging boxes. Additionally, the structural construction of large-scale concrete industrial floor slabs requires straight saw-cut expansion joints spaced at regular geometric intervals to accommodate thermal expansion.

When scaling automated logistics operations to industrial fleets comprising dozens of active robotic vehicles, coordinating spatial knowledge across distributed agents introduces severe computational and communication trade-offs [2, 3]. In traditional centralized architectures, each robot streams raw sensor scans or dense local occupancy submaps over wireless local area networks (WLAN) to a centralized scheduling server. However, modern mega-fulfillment centers spanning over twenty thousand square meters suffer from radio signal attenuation around massive steel racking structures, channel fading, and multi-path interference. Transmitting continuous high-bandwidth point clouds or optical imagery rapidly congests available wireless bandwidth, leading to dropped communication packets, delayed solver iterations, and potential fleet standstills. Achieving resilient multi-robot fleet localization requires decentralizing the underlying factor graph optimization, allowing individual vehicles to compute private spatial estimates locally while exchanging strictly condensed, mathematically consistent marginal representations when inter-robot loop closures occur.

Unlike vertical storage racks whose visual appearance and geometric profiles change continuously as pallets are moved, ground surface markings remain strictly planar, stationary, and unoccluded by overhead material handling equipment. When viewed by an onboard optical camera canted downward toward the floor, these surface primitives provide direct, high-frequency spatial constraints. Specifically, continuous longitudinal floor lines supply lateral offset and heading constraints, while perpendicular intersection stop bars and transversal slab joints provide the exact longitudinal position anchors that planar LiDAR lacks inside uniform corridors [16, 17].

However, processing downward-looking perspective imagery introduces distinct optical challenges. In perspective camera projections, parallel ground markings converge toward vanishing points due to projective foreshortening, inducing severe non-linear spatial distortion where physical ground distances map non-uniformly across image pixel coordinates [18]. Industrial concrete surfaces also frequently exhibit specular reflections from overhead high-bay sodium or LED luminaires, oil stains, and mechanical tire scrub marks that generate spurious intensity gradients. Extracting metric spatial constraints from ground imagery therefore requires rigorous geometric transformation, robust illumination-invariant feature extraction, and systematic uncertainty quantification [19, 20].

1.4. Contributions and Article Organization

This paper formulates a unified metric-topological semantic factor graph optimization architecture tailored for autonomous AGV fleet operations in cluttered, feature-sparse industrial warehouses. By coupling planar LiDAR scan registration and wheel odometry with dual-plane Inverse Perspective Mapping (IPM) visual ground feature extraction, our framework overcomes longitudinal corridor degeneracy, rejects dynamic pallet clutter, and enables scalable multi-robot loop closures across extensive facility footprints. The primary technical contributions of this research are structured as follows:

1. We develop a dual-plane inverse perspective mapping and steerable filter vision pipeline that converts tilted monocular camera streams into orthorectified Bird's-Eye-View (BEV) metric ground intensity representations, extracting persistent longitudinal lane edges, perpendicular junction stop lines, and concrete slab expansion seams with calibrated covariance projections.

2. We design an incremental metric-topological factor graph optimization architecture formulated on the Lie group $\mathrm{SE}(2)$. The framework unifies continuous odometry between-factors, planar LiDAR scan matching factors, visual IPM ground line geometric residuals, and discrete topological junction node constraints within an incremental Bayes tree solver.

3. We formulate a decentralized multi-robot graph condensation protocol based on Schur complement marginalization, allowing distributed fleet members to exchange compact topological loop closure proposals and marginal covariance envelopes over bandwidth-constrained industrial wireless networks without saturating communication channels.

4. We conduct extensive empirical evaluations across an operational $24{,}000\,\mathrm{m}^2$ container freight logistics terminal in Sri Lanka, benchmarking localization accuracy, slip suppression, computational execution time, and multi-robot fleet scaling against five state-of-the-art industrial SLAM systems across four distinct operational zones.

The remainder of this article is organized as follows. Section 2 reviews related literature on warehouse SLAM, visual ground plane mapping, and multi-robot factor graph optimization. Section 3 outlines the vehicle kinematic model, state representation on $\mathrm{SE}(2)$, and observability analysis under corridor degeneracy. Section 4 details the dual-plane inverse perspective mapping pipeline, steerable line filter, and covariance propagation. Section 5 presents the metric-topological factor graph structure and decentralized graph reduction. Section 6 describes the physical AGV testbed, industrial facility zones, and baseline configurations. Section 7 presents comprehensive experimental results, comparative benchmarks, and ablation studies. Section 8 discusses operational edge cases, failure modes, and practical deployment considerations. Section 9 concludes the article and highlights directions for future research.

2. Related Work

Reliable autonomous mobile robot navigation in structured indoor environments has evolved across several decades of robotics research, progressing from classical probabilistic filtering methods to modern non-linear factor graph optimization, multi-modal perception fusion, and semantic topological mapping [6, 21]. Within the specific domain of automated intralogistics and factory material transport, localization systems must guarantee centimeter-level precision while operating under computational constraints and intermittent wireless connectivity

[1]. This section reviews foundational and recent developments in graph-based SLAM for logistics, planar and spatial LiDAR odometry, inverse perspective mapping methodologies, and decentralized multi-robot fleet coordination.

2.1. Graph-Based SLAM and Industrial Warehouse Odometry

The formulation of Simultaneous Localization and Mapping as a sparse non-linear graph optimization problem represents a fundamental paradigm in modern mobile robotics [12, 22]. In this paradigm, the robot's spatial trajectory across discrete temporal epochs is modeled as a set of pose vertices on a manifold, while relative kinematic measurements, exteroceptive sensor registrations, and landmark observations are expressed as probabilistic factor edges encoding non-linear residual errors [23]. Rosen et al. [22] highlighted that the underlying sparsity of the measurement Jacobian allows efficient factor graph factorization using Cholesky or QR matrix decomposition. Building upon this matrix factorization insight, the incremental Smoothing and Mapping algorithm iSAM2 [11] introduced the Bayes tree data structure, which enables selective, near-constant-time relinearization of only those variable nodes affected by newly incorporated measurements. This incremental capability transformed factor graph optimization into an online estimation engine capable of continuous operation in real-time robotic navigation systems.

In industrial automated guided vehicle installations, two-dimensional horizontal LiDAR scanners represent the standard sensing payload for navigation and obstacle detection [7]. Hector SLAM [9] computes scan-to-map registration using Gauss-Newton optimization over bilinear occupancy grid gradients, achieving low computational overhead on embedded processors. However, because Hector SLAM relies strictly on scan matching without integrating wheel odometry, its estimation accuracy degrades precipitously in corridors with limited geometric features. Cartographer [8] mitigated this vulnerability by introducing a dual-tier architecture: local scan registration builds rigid submaps through real-time correlative scan matching, while a global background optimizer periodically executes branch-and-bound loop closures across historical submaps. While Cartographer demonstrates robustness across large floorplans, its branch-and-bound loop closure search is computationally intensive and assumes that environmental structures remain static, rendering it vulnerable to dynamic pallet movements in high-traffic logistics aisles [13].

To extract three-dimensional structural constraints, researchers developed spatial LiDAR odometry frameworks. LOAM [10] decomposes spatial point clouds into sharp edge points and planar surface patches, minimizing point-to-line and point-to-plane geometric distances through decoupled scan-to-scan and scan-to-map optimization. Similarly, LIO-SAM integrates spatial point cloud features with IMU preintegration on a factor graph. While 3D LiDAR odometry performs exceptionally well in complex architectural environments with heterogeneous geometry, its deployment in industrial AGV fleets is often constrained by hardware unit costs, elevated electrical power consumption, and mechanical vulnerability to overhead collisions during pallet lifting operations. Additionally, in long, smooth warehouse aisles where racking side walls present uniform planar surfaces without vertical profile variations, even 3D LiDAR systems experience longitudinal observability degradation [10].

2.2. Visual Ground Plane Mapping and Inverse Perspective Modeling

To overcome the geometric degeneracy of laser range finders without installing active floor guidance tracks, computer vision provides a complementary stream of high-density semantic and appearance measurements [24, 25]. In automotive advanced driver assistance systems, Inverse Perspective Mapping (IPM) was established as an analytical projective technique to eliminate the perspective distortion inherent in vehicle-mounted cameras [16, 17]. Assuming the ground surface can be approximated locally as a horizontal plane and that the extrinsic transformation between the camera coordinate frame and the ground surface is known, IPM computes a planar homography that reprojects perspective camera pixels into an orthorectified Bird's-Eye-View (BEV) coordinate plane [18].

In road vehicle navigation, IPM has been employed extensively for automated lane departure warning, curb tracking, and free-space boundary estimation [19, 20]. By removing perspective foreshortening, parallel lane markings in physical space remain parallel in the BEV image representation, simplifying curve fitting and distance estimation. However, road lane detection algorithms typically assume well-curved road profiles and high-contrast asphalt markings, conditions that differ substantially from indoor industrial concrete slabs. Industrial warehouse floors feature specular light reflections from overhead high-bay luminaires, mechanical tire scuff marks, oil stains, and complex crisscrossing patterns of yellow boundary tape, pedestrian crosswalks, and structural expansion joints cut into the concrete slab [5].

Planar homography and inverse perspective mapping formulations provide an efficient geometric mechanism to project perspective image observations onto physical world coordinates under flat ground assumptions. In autonomous field robotics, Nadiri, Banirostam, and Rad [26] demonstrated that inverse perspective transformations map tilted monocular camera imagery onto ground metric coordinates for field landmark extraction, establishing that explicit perspective rectification decoupled camera pitch angles and reduced humanoid soccer robot field localization error from an initial 17.98 cm down to 9.06 cm without necessitating mechanical gimbals. Building upon the core geometric insight of projective ground plane mapping validated in their work, our proposed framework generalizes inverse perspective formulations to large-scale indoor warehouse logistics, integrating ground plane line primitives and topological junction nodes into a global pose graph optimization. Rather than relying solely on isolated point landmarks, our approach parameterizes infinite continuous lines in Hesse normal form and incorporates structural expansion joints to eliminate longitudinal drift along geometrically degenerate corridors.

2.3. Topological Representations and Hierarchical Spatial Mapping

While metric factor graphs provide continuous, high-precision trajectory estimation in local coordinates, scaling SLAM to facilities spanning tens of thousands of square meters introduces severe computational and memory bottlenecks [12]. Hybrid metric-topological mapping architectures resolve this scaling challenge by decoupling global spatial connectivity from local metric estimation [27, 14]. In the Spatial Semantic Hierarchy formulated by Kuipers [27], spatial knowledge is organized into layered abstractions: sensory-level feedback control, causal action models, topological graphs representing discrete place nodes interconnected by travel pathways, and global metric maps.

In automated warehouse fleet operations, topological representations align naturally with the physical layout of the facility [4]. A warehouse can be formalized as an undirected topological graph where vertices denote aisle intersections, loading dock staging pads, and charging stations, while edges correspond to navigable corridor segments between storage racks. By detecting discrete topological places—such as the crossing of two painted yellow lane lines or the transition from a narrow storage aisle into a main distribution thoroughfare—mobile robots can perform reliable place recognition that is invariant to metric drift [14]. Olson [15] demonstrated that recognizing discrete loop closures and verifying their topological consistency prevents incorrect loop assertions from corrupting the metric trajectory. Integrating discrete topological junction detections as specialized non-linear measurement factors within an incremental Bayes tree allows our framework to combine the global routing advantages of topological graphs with the continuous precision of metric factor graphs.

2.4. Decentralized Multi-Robot Fleet Optimization and Communication Constraints

When multiple autonomous mobile robots operate concurrently within the same physical warehouse, sharing perceptual information across the fleet offers substantial improvements in mapping speed, localization reliability, and spatial coverage [2, 3]. Early multi-robot SLAM systems relied on centralized architectures where all individual robot pose estimates and laser scans were transmitted over a wireless network to a central computing server for joint global optimization. However, centralized systems introduce single-point-of-failure vulnerabilities and scale poorly as fleet size expands; transmitting raw sensory streams from dozens of vehicles rapidly saturates industrial wireless local area networks (WLAN), causing packet dropping, communication delays, and solver stalls [28].

To achieve scalable multi-agent operations, recent research has prioritized decentralized, peer-to-peer factor graph formulations. In decentralized SLAM, each individual vehicle maintains its own private local factor graph, estimating its trajectory using onboard sensors. When two robots achieve direct inter-robot communication, they exchange compact visual feature descriptors or topological place hypotheses to establish inter-vehicle relative pose constraints. Cunningham et al. [29] proposed Decentralized Data Fusion using DDF-SAM, applying Gaussian elimination to compute condensed marginal representations of shared trajectory nodes. By marginalizing private internal robot states and transmitting only the Schur complement factor representing the shared boundary nodes, inter-robot communication bandwidth can be reduced by several orders of magnitude while preserving exact optimization consistency. Our framework builds upon these decentralized matrix marginalization principles, defining an asynchronous graph condensation protocol that enables fleets of up to fifty AGVs to coordinate reliably across industrial logistics facilities under realistic wireless channel constraints.

3. Problem Formulation and Kinematic Modeling

This section establishes the mathematical foundations of the proposed AGV fleet localization framework, detailing the vehicle kinematic model on the matrix Lie group $\mathrm{SE}(2)$, the relative motion integration from wheel odometry, and an analytical observability analysis demonstrating the geometric degeneracy of planar LiDAR scan matching inside long warehouse storage corridors.

3.1. Mobile Robot Kinematics and Lie Group Representation on $\mathrm{SE}(2)$

We consider a fleet of M autonomous guided vehicles operating on a common planar industrial floor slab $\Pi_{\text{floor}} \subset \mathbb{R}^3$. The world reference frame is denoted by $\mathcal{F}_W = \{O_W, \mathbf{x}_W, \mathbf{y}_W, \mathbf{z}_W\}$, where the vertical unit vector $\mathbf{z}_W$ is aligned parallel to the local gravity direction, and the orthogonal horizontal vectors $\mathbf{x}_W$ and $\mathbf{y}_W$ span the flat concrete floor surface. For the i-th robot in the fleet at discrete time epoch $k \in \{0, 1, \ldots, K\}$, its local body coordinate frame is denoted by $\mathcal{F}_{B,k}^{(i)} = \{O_{B,k}^{(i)}, \mathbf{x}_{B,k}^{(i)}, \mathbf{y}_{B,k}^{(i)}, \mathbf{z}_{B,k}^{(i)}\}$, centered at the midpoint of the vehicle's motorized drive axle.

Because industrial AGVs operate strictly on flat interior floors without vertical flight or banking dynamics, the vehicle pose is appropriately parameterized by the Special Euclidean group $\mathrm{SE}(2)$. The spatial state of robot i at epoch k is represented by the homogeneous transformation matrix:

$$\mathbf{T}_k^{(i)} = \begin{bmatrix} \mathbf{R}_k^{(i)} & \mathbf{p}_k^{(i)} \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix} \in \mathrm{SE}(2), \tag{1}$$

where $\mathbf{p}_k^{(i)} = [x_k^{(i)}, y_k^{(i)}]^T \in \mathbb{R}^2$ defines the 2D Cartesian translation vector in the world frame, and $\mathbf{R}_k^{(i)} \in \mathrm{SO}(2)$ denotes the planar rotation matrix parameterized by the vehicle heading angle $\theta_k^{(i)} \in [-\pi, \pi)$:

$$\mathbf{R}_k^{(i)} = \begin{bmatrix} \cos\theta_k^{(i)} & -\sin\theta_k^{(i)} \\ \sin\theta_k^{(i)} & \cos\theta_k^{(i)} \end{bmatrix}. \tag{2}$$

In vector notation, the pose state is expressed as $\mathbf{x}_k^{(i)} = [x_k^{(i)}, y_k^{(i)}, \theta_k^{(i)}]^T \in \mathbb{R}^3$.

The Lie algebra associated with SE(2), denoted by $\mathfrak{se}(2)$, corresponds to the tangent space at the group identity element $\mathbf{I}_3$. Any tangent vector $\boldsymbol{\xi} = [\xi_x, \xi_y, \xi_\theta]^T \in \mathbb{R}^3$ maps to an element of the Lie algebra via the standard wedge operator $(\cdot)^\wedge : \mathbb{R}^3 \to \mathfrak{se}(2)$:

$$\boldsymbol{\xi}^\wedge = \begin{bmatrix} 0 & -\xi_\theta & \xi_x \\ \xi_\theta & 0 & \xi_y \\ 0 & 0 & 0 \end{bmatrix}. \tag{3}$$

The exponential map $\exp : \mathfrak{se}(2) \to \mathrm{SE}(2)$ provides the exact closed-form transformation from the Lie algebra to the Lie group:

$$\exp(\boldsymbol{\xi}^\wedge) = \begin{bmatrix} \mathbf{R}(\xi_\theta) & \mathbf{V}(\xi_\theta) \begin{bmatrix} \xi_x \\ \xi_y \end{bmatrix} \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix}, \tag{4}$$

where the left Jacobian matrix $\mathbf{V}(\xi_\theta) \in \mathbb{R}^{2\times 2}$ is given analytically by:

$$\mathbf{V}(\xi_\theta) = \frac{\sin \xi_\theta}{\xi_\theta}\mathbf{I}_2 + \frac{1-\cos \xi_\theta}{\xi_\theta} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{5}$$

The logarithmic map $\log : \mathrm{SE}(2) \to \mathfrak{se}(2)$ represents the smooth inverse operation, returning the unique tangent coordinates within the principal Riemann covering space.

3.2. Differential-Drive Wheel Odometry Integration

Each AGV utilizes a differential-drive mechanical locomotion configuration comprising two diametrically opposed coaxial drive wheels of nominal radius r_w separated by track baseline distance B_w, accompanied by passive caster wheels that provide multi-point mechanical equilibrium [5]. Optical incremental rotary encoders mounted directly to the drive motor shafts sample angular wheel increments $\Delta\phi_{L,k}^{(i)}$ and $\Delta\phi_{R,k}^{(i)}$ over sampling interval $\Delta t = t_k - t_{k-1}$.

The longitudinal displacement increment $\Delta s_k^{(i)}$ and angular heading rotation increment $\Delta\theta_k^{(i)}$ of the vehicle chassis during interval Δt are computed via forward kinematics:

$$\Delta s_k^{(i)} = \frac{r_w}{2}\left(\Delta\phi_{R,k}^{(i)} + \Delta\phi_{L,k}^{(i)}\right), \tag{6}$$

$$\Delta\theta_k^{(i)} = \frac{r_w}{B_w}\left(\Delta\phi_{R,k}^{(i)} - \Delta\phi_{L,k}^{(i)}\right). \tag{7}$$

Formulating the relative displacement in the local vehicle frame yields the local tangent increment vector $\Delta\boldsymbol{\xi}_k^{(i)} = [\Delta s_k^{(i)} \operatorname{sinc}(\Delta\theta_k^{(i)}/2), 0, \Delta\theta_k^{(i)}]^T \in \mathbb{R}^3$. The temporal discrete-time motion propagation across consecutive keyframes on SE(2) is given by group multiplication:

$$\mathbf{T}_k^{(i)} = \mathbf{T}_{k-1}^{(i)} \exp\left((\Delta\boldsymbol{\xi}_k^{(i)} + \mathbf{w}_{\mathrm{odo},k}^{(i)})^\wedge\right), \tag{8}$$

where $\mathbf{w}_{\mathrm{odo},k}^{(i)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{odo},k}^{(i)})$ represents zero-mean Gaussian odometry perturbation noise. The wheel covariance matrix scales quadratically with traveled distance and angular rotation:

$$\boldsymbol{\Sigma}_{\mathrm{odo},k}^{(i)} = \operatorname{diag}\left(\sigma_{s,k}^2, \sigma_{\mathrm{lat},k}^2, \sigma_{\theta,k}^2\right), \tag{9}$$

with $\sigma_{s,k}^2 = k_d|\Delta s_k^{(i)}| + k_\theta|\Delta\theta_k^{(i)}|$, $\sigma_{\mathrm{lat},k}^2 = k_{\mathrm{lat}}|\Delta s_k^{(i)}|$, and $\sigma_{\theta,k}^2 = k_\theta|\Delta\theta_k^{(i)}| + k_d|\Delta s_k^{(i)}|$, where $k_d, k_\theta, k_{\mathrm{lat}} > 0$ denote empirically calibrated kinematic error parameters.

3.3. Degeneracy Analysis of Planar LiDAR in Warehouse Corridors

We now examine the mathematical origin of planar LiDAR failure along uniform storage corridors. Let an onboard horizontal LiDAR scanner capture a range scan $\mathcal{Z}_k = \{\mathbf{z}_{k,j}\}_{j=1}^{N_z}$, where each measurement $\mathbf{z}_{k,j} = [r_j \cos\psi_j, r_j \sin\psi_j]^T \in \mathbb{R}^2$ denotes the Cartesian coordinates of a reflected beam in the sensor frame. In point-to-plane or point-to-line scan matching, each laser beam is aligned against a reference environmental surface model $\mathcal{M}$ consisting of planar boundary segments [10].

For a measured point $\mathbf{z}_{k,j}$ transformed into the global frame via candidate pose $\mathbf{T}_k$, the geometric point-to-plane residual error is formulated as:

$$e_j(\mathbf{x}_k) = \mathbf{n}_m^T\left(\mathbf{R}(\theta_k)\mathbf{z}_{k,j} + \mathbf{p}_k - \mathbf{q}_m\right), \tag{10}$$

where $\mathbf{n}_m \in \mathbb{R}^2$ denotes the unit normal vector of the closest corridor surface segment, and $\mathbf{q}_m \in \mathbb{R}^2$ is an arbitrary anchor point on that surface.

Minimizing the aggregate sum of squared residuals over all active range beams yields the standard non-linear least-squares optimization problem:

$$\mathbf{x}_k^* = \arg\min_{\mathbf{x}_k} \sum_{j=1}^{N_z} \frac{1}{\sigma_j^2}\left(e_j(\mathbf{x}_k)\right)^2. \tag{11}$$

Linearizing equation (10) around current operating estimate $\bar{\mathbf{x}}_k$ yields the measurement Jacobian $\mathbf{J}_j \in \mathbb{R}^{1\times 3}$:

$$\mathbf{J}_j = \begin{bmatrix} n_{m,x} & n_{m,y} & J_{j,\theta} \end{bmatrix}, \tag{12}$$

$$J_{j,\theta} = \mathbf{n}_m^T \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{R}(\bar{\theta}_k)\mathbf{z}_{k,j}. \tag{13}$$

The second-order curvature of the optimization objective is governed by the Fisher Information Matrix (FIM), also designated as the Gauss-Newton approximate Hessian $\boldsymbol{\Lambda}_{\mathrm{scan}} \in \mathbb{R}^{3\times 3}$:

$$\boldsymbol{\Lambda}_{\mathrm{scan}} = \sum_{j=1}^{N_z} \frac{1}{\sigma_j^2}\mathbf{J}_j^T\mathbf{J}_j = \begin{bmatrix} \sum \frac{n_{m,x}^2}{\sigma_j^2} & \sum \frac{n_{m,x}n_{m,y}}{\sigma_j^2} & \sum \frac{n_{m,x}J_{j,\theta}}{\sigma_j^2} \\ \sum \frac{n_{m,x}n_{m,y}}{\sigma_j^2} & \sum \frac{n_{m,y}^2}{\sigma_j^2} & \sum \frac{n_{m,y}J_{j,\theta}}{\sigma_j^2} \\ \sum \frac{n_{m,x}J_{j,\theta}}{\sigma_j^2} & \sum \frac{n_{m,y}J_{j,\theta}}{\sigma_j^2} & \sum \frac{J_{j,\theta}^2}{\sigma_j^2} \end{bmatrix}. \tag{14}$$

Now, consider an AGV traversing an extended warehouse storage corridor whose longitudinal axis is aligned parallel to the global $\mathbf{x}_W$ direction. The corridor is flanked on the left and right by continuous storage racks located at lateral coordinates $y = +W_{\text{cor}}/2$ and $y = -W_{\text{cor}}/2$. For every laser point striking the left rack wall, the surface normal is $\mathbf{n}_{\text{left}} = [0, -1]^T$. For every laser point striking the right rack wall, the surface normal is $\mathbf{n}_{\text{right}} = [0, 1]^T$.

Substituting these surface normal vectors into equation (12), we observe that for all measured laser returns $j \in \{1, \ldots, N_z\}$:

$$n_{m,x} = 0 \quad \forall j \in \{1, \ldots, N_z\}. \tag{15}$$

Consequently, every entry in the first row and first column of the Fisher Information Matrix vanishes identically:

$$\boldsymbol{\Lambda}_{\text{scan}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sum_{j=1}^{N_z} \frac{1}{\sigma_j^2} & \sum_{j=1}^{N_z} \frac{J_{j,\theta}}{\sigma_j^2} \\ 0 & \sum_{j=1}^{N_z} \frac{J_{j,\theta}}{\sigma_j^2} & \sum_{j=1}^{N_z} \frac{J_{j,\theta}^2}{\sigma_j^2} \end{bmatrix}. \tag{16}$$

Equation (16) reveals that the rank of the Fisher Information Matrix drops from three to two:

$$\text{rank}(\boldsymbol{\Lambda}_{\text{scan}}) = 2 < 3, \quad \det(\boldsymbol{\Lambda}_{\text{scan}}) = 0. \tag{17}$$

The nullspace of $\boldsymbol{\Lambda}_{\text{scan}}$ is spanned directly by the longitudinal translation direction:

$$\text{null}(\boldsymbol{\Lambda}_{\text{scan}}) = \text{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right). \tag{18}$$

Along this nullspace direction, the second derivative of the scan matching error is zero, meaning that the optimization surface is completely flat along the corridor axis.

If a scan matcher attempts to invert $\boldsymbol{\Lambda}_{\text{scan}}$ directly, the covariance along the longitudinal axis becomes infinite:

$$\boldsymbol{\Sigma}_{\text{scan}} = (\boldsymbol{\Lambda}_{\text{scan}})^{-1} \implies \sigma_{xx}^2 \to \infty. \tag{19}$$

In practical scan matching implementations, regularized pseudo-inverses or Levenberg-Marquardt damping terms are applied to prevent numerical division by zero. However, while damping maintains numerical solvability, it does not supply physical measurement information; the estimated longitudinal displacement becomes governed entirely by noise integration, causing the vehicle's estimated position to drift unchecked until external geometric constraints are introduced.

3.4. Uncertainty Propagation on $\mathrm{SE}(2)$ and Adjoint Mapping

To maintain rigorous probabilistic consistency within the factor graph, kinematic uncertainty must be transformed across non-stationary reference frames without introducing coordinate distortions. Let $\mathbf{T}_{wb} = (\mathbf{R}, \mathbf{p}) \in \mathrm{SE}(2)$ denote the transformation from the vehicle body frame $\mathcal{F}_B$ to the world frame $\mathcal{F}_W$. Small perturbation vectors $\boldsymbol{\xi} \in \mathfrak{se}(2) \cong \mathbb{R}^3$ defined in the local tangent space map to the global coordinate frame through the adjoint action of the Lie group:

$$\text{Ad}_{\mathbf{T}_{wb}} = \begin{bmatrix} \mathbf{R} & \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{p} \\ \mathbf{0}_{1\times 2} & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}. \tag{20}$$

Consequently, a local covariance matrix $\boldsymbol{\Sigma}_B \in \mathbb{R}^{3\times 3}$ associated with wheel odometry or local feature detections transforms into the global frame according to:

$$\boldsymbol{\Sigma}_W = \text{Ad}_{\mathbf{T}_{wb}} \boldsymbol{\Sigma}_B \, \text{Ad}_{\mathbf{T}_{wb}}^T. \tag{21}$$

During corridor navigation, wheel slip induces stochastic perturbations that accumulate along both longitudinal and lateral axes. By modeling the continuous-time motion dynamics as a stochastic differential equation driven by zero-mean Gaussian white noise with power spectral density $\mathbf{Q}_c = \text{diag}(\sigma_v^2, \sigma_{\text{lat}}^2, \sigma_\omega^2)$, the discrete-time relative covariance between keyframes $k-1$ and k separated by time interval Δt_k evaluates to:

$$\boldsymbol{\Sigma}_{\text{odo},k} = \int_0^{\Delta t_k} \exp(\boldsymbol{\xi}^\wedge \tau) \, \mathbf{Q}_c \, \exp(\boldsymbol{\xi}^\wedge \tau)^T d\tau. \tag{22}$$

In the absence of external absolute position updates, the longitudinal variance $\sigma_{x,k}^2$ expands linearly with total traveled path length $S_k = \sum_{j=1}^{k} \Delta s_j$, while the orientation variance $\sigma_{\theta,k}^2$ scales with cumulative angular displacement. When the planar LiDAR experiences the rank deficiency proven in equation (17), this unbounded kinematic covariance growth remains unchecked unless complementary exteroceptive constraints are introduced.

4. Dual-Plane Inverse Perspective Mapping and Visual Feature Extraction

To resolve the longitudinal geometric degeneracy of planar LiDAR scanners established in Section 3, we introduce an optical ground perception subsystem based on dual-plane Inverse Perspective Mapping (IPM). By capturing downward-canted monocular images of the warehouse concrete floor, this pipeline extracts persistent metric line primitives and topological junction nodes, projecting them into orthorectified Bird's-Eye-View (BEV) coordinates to serve as drift-free spatial constraints.

4.1. Camera Projection Model and Downward-Canted Extrinsic Geometry

Each AGV is equipped with an industrial-grade downward-canted global-shutter monocular CMOS camera rigidly mounted to the front chassis crossmember. The camera coordinate frame is denoted by $\mathcal{F}_C = \{O_C, \mathbf{x}_C, \mathbf{y}_C, \mathbf{z}_C\}$, where the principal optical axis $\mathbf{z}_C$ points forward and downward toward the concrete slab, $\mathbf{x}_C$ points to the camera's right, and $\mathbf{y}_C$ points downward along the sensor vertical axis.

A three-dimensional physical point $\mathbf{P}_C = [X_C, Y_C, Z_C]^T \in \mathbb{R}^3$ expressed in camera coordinates projects onto the normalized image plane via the canonical pinhole perspective transformation:

$$\mathbf{m} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_C}\mathbf{K}_{\text{cam}}\mathbf{P}_C, \tag{23}$$

where $\mathbf{K}_{\text{cam}} \in \mathbb{R}^{3\times 3}$ denotes the calibrated camera intrinsic matrix:

$$\mathbf{K}_{\text{cam}} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{24}$$

with (f_x, f_y) denoting the horizontal and vertical focal lengths in pixel units, and (c_x, c_y) representing the principal point coordinates. Radial and tangential lens distortions are modeled using the standard polynomial model and removed in real-time prior to geometric processing [30].

The extrinsic transformation from the AGV base coordinate frame $\mathcal{F}_B$ to the camera frame $\mathcal{F}_C$ is defined by the rigid Euclidean transformation:

$$\mathbf{P}_C = \mathbf{R}_{bc}^T(\mathbf{P}_B - \mathbf{p}_{bc}), \tag{25}$$

where $\mathbf{p}_{bc} = [d_x, 0, h_c]^T \in \mathbb{R}^3$ specifies the spatial mounting offset of the camera optical center above the floor, with $h_c > 0$ denoting the vertical optical elevation, and $d_x > 0$ denoting the forward longitudinal overhang relative to the wheel drive axle.

The spatial orientation of the camera frame relative to the horizontal floor plane is defined by successive Euler rotations consisting of pitch angle $\alpha_c > 0$ (canted downward toward the floor), roll angle $\beta_c \approx 0$ (calibrated parallel to the wheel axle), and yaw angle $\gamma_c \approx 0$ (aligned with the forward vehicle heading):

$$\mathbf{R}_{bc} = \mathbf{R}_z(\gamma_c)\mathbf{R}_y(\beta_c)\mathbf{R}_x(\alpha_c). \tag{26}$$

In practical warehouse fleet deployments, the optical tilt angle α_c is selected within the range of 35° to 55°. This downward orientation guarantees that the camera's effective field of view captures a trapezoidal floor footprint extending from 0.6 m to 4.2 m ahead of the vehicle bumper, simultaneously imaging continuous corridor lane lines and approaching aisle crossing markers while eliminating perspective glare from distant warehouse skylights.

4.2. Analytical Formulation of the Inverse Perspective Mapping Homography

Let Π_{floor} represent the nominal horizontal ground plane of the warehouse facility. In the AGV body frame $\mathcal{F}_B$, the equation of the flat ground plane is defined by:

$$\mathbf{n}_B^T\mathbf{P}_B + d_{\text{floor}} = 0, \tag{27}$$

where $\mathbf{n}_B = [0, 0, 1]^T$ is the upward unit normal vector of the floor slab, and $d_{\text{floor}} = 0$ because the origin of $\mathcal{F}_B$ is centered on the ground contact surface.

Transforming the ground plane parameters into the camera coordinate frame $\mathcal{F}_C$ yields the camera-relative plane normal $\mathbf{n}_C$ and orthogonal distance d_C:

$$\mathbf{n}_C = \mathbf{R}_{bc}^T\mathbf{n}_B = \begin{bmatrix} 0 \\ \cos\alpha_c \\ -\sin\alpha_c \end{bmatrix}, \quad d_C = h_c\cos\alpha_c. \tag{28}$$

For any three-dimensional surface point $\mathbf{P}_{\text{ground}} = [X_g, Y_g, 0]^T$ lying strictly on the ground plane Π_{floor}, its spatial projection onto the perspective image plane $\mathbf{m} = [u, v, 1]^T$ and its corresponding metric position in the Bird's-Eye-View (BEV) ground plane $\mathbf{p}_{\text{BEV}} = [X_g, Y_g, 1]^T$ are related by a planar projective homography $\mathbf{H}_{\text{IPM}} \in \mathbb{R}^{3\times 3}$ [16, 17]:

$$\lambda\mathbf{m} = \mathbf{H}_{\text{IPM}}\mathbf{p}_{\text{BEV}}, \tag{29}$$

where $\lambda \in \mathbb{R}$ is an arbitrary homogeneous scale factor.

Decomposing the projective homography into intrinsic and extrinsic geometric components yields:

$$\mathbf{H}_{\text{IPM}} = \mathbf{K}_{\text{cam}}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}_{\text{floor}} \end{bmatrix}, \tag{30}$$

where $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{R}^3$ are the first two column vectors of the rotation matrix $\mathbf{R}_{bc}^T$, and $\mathbf{t}_{\text{floor}} = -\mathbf{R}_{bc}^T\mathbf{p}_{bc}$ represents the translation vector from the ground origin to the camera optical center.

To compute the metric ground coordinates corresponding to an observed image pixel $[u, v]^T$, we invert the homography matrix:

$$\mathbf{G}_{\text{IPM}} = \mathbf{H}_{\text{IPM}}^{-1} = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix}. \tag{31}$$

The forward metric ground coordinates (X_g, Y_g) relative to the AGV chassis are then computed analytically via homogeneous de-normalization:

$$X_g(u, v) = \frac{g_{11}u + g_{12}v + g_{13}}{g_{31}u + g_{32}v + g_{33}}, \tag{32}$$

$$Y_g(u, v) = \frac{g_{21}u + g_{22}v + g_{23}}{g_{31}u + g_{32}v + g_{33}}. \tag{33}$$

Equation (33) demonstrates that every pixel in the perspective image maps deterministically to physical metric coordinates on the warehouse floor. By computing equation (33) across a dense grid of ground points, the perspective camera frame is transformed into an orthorectified Bird's-Eye-View intensity image $\mathcal{I}_{\text{BEV}} \in \mathbb{R}^{H_{\text{BEV}}\times W_{\text{BEV}}}$, where physical distances scale linearly and parallel warehouse lane lines remain strictly parallel.

4.3. Steerable Oriented Filters for Robust Floor Line Primitive Extraction

Raw industrial concrete surfaces present severe photometric noise, including specular highlights from high-bay sodium vapor luminaires, dark rubber tire scuff marks, and moisture variations.

Standard gradient operators, such as Sobel or Canny filters, produce noisy edge responses along tire tracks and concrete surface texture. To isolate physical lane markings and concrete joints reliably, we formulate a directional feature extraction pipeline based on steerable second-derivative Gaussian filters [19].

Let $G(x, y; \sigma_f)$ denote a two-dimensional rotationally symmetric Gaussian smoothing kernel:

$$G(x, y; \sigma_f) = \frac{1}{2\pi\sigma_f^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_f^2}\right), \tag{34}$$

where σ_f specifies the spatial filter scale, tuned to match the nominal physical width of painted floor tape (5.0 cm to 10.0 cm).

The second directional derivative of G oriented along arbitrary spatial angle $\theta_f \in [0, \pi)$ can be expressed as a linear combination of three separable basis filters:

$$\begin{aligned} G_2^{\theta_f}(x, y) = &k_1(\theta_f) G_{2a}(x, y) + k_2(\theta_f) G_{2b}(x, y) \\ &+ k_3(\theta_f) G_{2c}(x, y), \end{aligned} \tag{35}$$

where the basis operators are given by:

$$G_{2a} = \frac{\partial^2 G}{\partial x^2}, \quad G_{2b} = \frac{\partial^2 G}{\partial x \partial y}, \quad G_{2c} = \frac{\partial^2 G}{\partial y^2}, \tag{36}$$

and the interpolation coefficients evaluate to:

$$k_1(\theta_f) = \cos^2\theta_f, \quad k_3(\theta_f) = \sin^2\theta_f, \tag{37}$$

$$k_2(\theta_f) = -2\cos\theta_f \sin\theta_f. \tag{38}$$

Convolving the rectified BEV image $\mathcal{I}_{\text{BEV}}$ with the basis filters yields the directional ridge energy response $E(\mathbf{p}, \theta_f)$:

$$E(\mathbf{p}, \theta_f) = \mathcal{I}_{\text{BEV}}(\mathbf{p}) * G_2^{\theta_f}(\mathbf{p}). \tag{39}$$

Because warehouse markings are predominantly aligned either longitudinally (parallel to the storage aisles, $\theta_f \approx 0°$) or transversely (perpendicular stop lines and expansion seams, $\theta_f \approx 90°$), we evaluate the steerable response across two orthogonal channels:

$$E_{\parallel}(\mathbf{p}) = E(\mathbf{p}, 0°), \quad E_{\perp}(\mathbf{p}) = E(\mathbf{p}, 90°). \tag{40}$$

Thresholding the directional energy maps with adaptive Otsu binarization extracts candidate line support pixels. We then execute a Progressive Probabilistic Hough Transform (PPHT) on the binarized ridge maps to extract discrete line segments $\mathcal{S}_l = \{\mathbf{p}_{l,\text{start}}, \mathbf{p}_{l,\text{end}}\}$.

4.4. Hesse Normal Form Parameterization and Covariance Projection

Extracted floor line segments are parameterized in infinite Hesse normal form within the local vehicle frame:

$$\mathbf{l}_B = \begin{bmatrix} \rho_B \\ \alpha_B \end{bmatrix}, \tag{41}$$

where $\rho_B \geq 0$ denotes the perpendicular Euclidean distance from the vehicle base origin to the line, and $\alpha_B \in [-\pi, \pi)$ is the counterclockwise angle of the line normal vector relative to the vehicle longitudinal axis $\mathbf{x}_B$. Any ground point $\mathbf{p} = [x, y]^T$ lying on the physical line satisfies the linear constraint:

$$x\cos\alpha_B + y\sin\alpha_B - \rho_B = 0. \tag{42}$$

To rigorously quantify the uncertainty of extracted line parameters, we formulate the stochastic error propagation from pixel measurement uncertainty into metric line covariance. Let an image feature point $\mathbf{m} = [u, v]^T$ exhibit isotropic Gaussian pixel noise with covariance $\mathbf{\Sigma}_{\text{pixel}} = \sigma_p^2 \mathbf{I}_2$. The covariance of the unprojected metric ground coordinate $\mathbf{p}_g = [X_g, Y_g]^T$ is computed via the Jacobian of the inverse homography transformation:

$$\mathbf{\Sigma}_{\mathbf{p}_g} = \mathbf{J}_{\mathbf{G}} \mathbf{\Sigma}_{\text{pixel}} \mathbf{J}_{\mathbf{G}}^T, \tag{43}$$

where $\mathbf{J}_{\mathbf{G}} = \frac{\partial(X_g, Y_g)}{\partial(u, v)} \in \mathbb{R}^{2\times 2}$ is evaluated analytically with $D_1 = g_{11}u + g_{12}v + g_{13}$, $D_2 = g_{21}u + g_{22}v + g_{23}$, and $D_3 = g_{31}u + g_{32}v + g_{33}$:

$$\mathbf{J}_{\mathbf{G}} = \frac{1}{D_3^2} \begin{bmatrix} g_{11}D_3 - g_{31}D_1 & g_{12}D_3 - g_{32}D_1 \\ g_{21}D_3 - g_{31}D_2 & g_{22}D_3 - g_{32}D_2 \end{bmatrix}. \tag{44}$$

Notice from equation (44) that the denominator scales quadratically with distance from the camera: as points approach the horizon ($Z_C \to \infty$), the ground uncertainty expands quadratically along the optical axis, correctly reflecting the degraded spatial resolution of distant perspective pixels. Line segments extracted closer to the AGV chassis ($1.0\,\text{m} \leq X_g \leq 2.5\,\text{m}$) achieve millimeter-level metric certainty ($\sigma_{\mathbf{p}_g} < 2.5\,\text{mm}$), providing optimal measurement weights for subsequent factor graph integration.

Given an extracted line segment anchored by two unprojected metric endpoints $\mathbf{p}_1 = [X_1, Y_1]^T$ and $\mathbf{p}_2 = [X_2, Y_2]^T$ with associated metric covariance blocks $\mathbf{\Sigma}_{\mathbf{p}_1}$ and $\mathbf{\Sigma}_{\mathbf{p}_2}$, the line angle α_B and perpendicular distance ρ_B satisfy:

$$\alpha_B = \text{atan2}(-(X_2 - X_1), Y_2 - Y_1), \tag{45}$$

$$\rho_B = X_1\cos\alpha_B + Y_1\sin\alpha_B. \tag{46}$$

Applying first-order perturbation theory, the joint covariance matrix of the Hesse normal parameters $\mathbf{\Sigma}_{\mathbf{l}_B} \in \mathbb{R}^{2\times 2}$ is obtained through the Jacobian mapping $\mathbf{J}_{\mathbf{l},\mathbf{p}} = \frac{\partial(\rho_B, \alpha_B)}{\partial(\mathbf{p}_1, \mathbf{p}_2)} \in \mathbb{R}^{2\times 4}$:

$$\mathbf{\Sigma}_{\mathbf{l}_B} = \mathbf{J}_{\mathbf{l},\mathbf{p}} \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{p}_1} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{\Sigma}_{\mathbf{p}_2} \end{bmatrix} \mathbf{J}_{\mathbf{l},\mathbf{p}}^T. \tag{47}$$

To achieve sub-millimeter precision, the position of each ridge point along the local gradient direction is refined using a one-dimensional continuous parabolic interpolation over the three discrete filter responses centered at the local maximum. This sub-pixel interpolation suppresses spatial discretization noise caused by camera sensor Bayer demosaicing and ensures smooth residual gradients during non-linear optimization.

5. Metric-Topological Factor Graph Optimization

This section formulates the unified metric-topological factor graph optimization architecture that integrates wheel kinematics, planar LiDAR scan matching, visual IPM ground line primitives, and topological junction nodes into an incremental Bayes tree solver. In addition, we establish an asynchronous decentralized graph condensation protocol that enables multi-robot fleet coordination under severe wireless bandwidth constraints.

5.1. Factor Graph Topology and Variable Node Definitions

We model the multi-sensor estimation problem across the AGV fleet as an open factor graph $\mathcal{G} = (\mathcal{V}, \mathcal{F}, \mathcal{E})$, where $\mathcal{V}$ denotes the set of variable nodes to be estimated, $\mathcal{F}$ is the set of probabilistic factor nodes encoding sensor measurements and structural constraints, and $\mathcal{E}$ represents the edges defining conditional dependencies [22].

The state space of variable nodes $\mathcal{V}$ comprises three distinct classes of variables: 1. **Robot Pose Vertices** $\mathcal{X}$: The trajectory of vehicle $i \in \{1, \dots, M\}$ across keyframe epochs $k \in \{0, \dots, K_i\}$ is represented by pose nodes $\mathbf{x}_k^{(i)} \in \mathrm{SE}(2)$. Keyframes are inserted dynamically whenever traveled distance exceeds $\Delta s_{\text{thresh}} = 0.20\,\text{m}$ or angular heading rotates by more than $\Delta\theta_{\text{thresh}} = 5.0^\circ$. 2. **Ground Line Landmark Vertices** $\mathcal{L}$: Persistent physical floor markings (such as continuous aisle boundaries and expansion joints) are parameterized as global landmark nodes $\mathbf{l}_m = [\rho_m, \alpha_m]^T \in \mathbb{R}^2$, expressed in Hesse normal form in the global world frame $\mathcal{F}_W$. 3. **Topological Junction Vertices** $\mathcal{J}$: Discrete aisle intersections, crosswalk crossings, and docking stations are represented as topological place nodes $\mathbf{j}_n \in \mathbb{R}^2$, providing high-level semantic anchors that link crossing corridors.

The joint maximum a posteriori (MAP) state estimate $\mathcal{V}^*$ corresponds to the minimizer of the aggregate non-linear least-squares objective function:

$$\mathcal{V}^* = \arg\min_{\mathcal{V}} \sum_{f_a \in \mathcal{F}} \frac{1}{2} \|\mathbf{r}_a(\mathcal{V}_a)\|^2_{\mathbf{\Sigma}_a}, \tag{48}$$

where $\mathcal{V}_a \subset \mathcal{V}$ denotes the subset of variable nodes connected to factor f_a, $\mathbf{r}_a(\cdot)$ is the non-linear measurement residual vector, and $\|\mathbf{r}\|^2_{\mathbf{\Sigma}} = \mathbf{r}^T\mathbf{\Sigma}^{-1}\mathbf{r}$ represents the squared Mahalanobis distance weighted by measurement covariance matrix $\mathbf{\Sigma}_a$.

5.2. Mathematical Formulation of Measurement Factors

The factor set $\mathcal{F}$ unifies four heterogeneous physical measurement sources:

5.2.1. Odometry Between-Factors

Consecutive keyframes $\mathbf{x}_{k-1}^{(i)}$ and $\mathbf{x}_k^{(i)}$ are linked by relative motion constraints derived from integrated wheel encoder odometry:

$$\mathbf{r}_{\text{odo}}(\mathbf{x}_{k-1}^{(i)}, \mathbf{x}_k^{(i)}) = \log\left((\Delta\mathbf{T}_{\text{odo},k}^{(i)})^{-1}(\mathbf{T}_{k-1}^{(i)})^{-1}\mathbf{T}_k^{(i)}\right)^{\vee} \in \mathbb{R}^3, \tag{49}$$

where $\Delta\mathbf{T}_{\text{odo},k}^{(i)} \in \mathrm{SE}(2)$ is the integrated kinematic transform from equation (7), and $(\cdot)^{\vee} : \mathfrak{se}(2) \to \mathbb{R}^3$ denotes the inverse wedge operator. The associated information weight is given by the inverse covariance $\mathbf{\Sigma}_{\text{odo},k}^{-1}$.

5.2.2. Planar LiDAR Scan-to-Submap Factors

When the AGV traverses areas with non-degenerate horizontal structure (such as open distribution staging pads or perimeter structural columns), relative pose transformations $\Delta\mathbf{T}_{\text{scan},k}^{(i)}$ generated by scan matching against local submaps are incorporated as unary or binary pose constraints:

$$\mathbf{r}_{\text{scan}}(\mathbf{x}_k^{(i)}) = \log\left((\Delta\mathbf{T}_{\text{scan},k}^{(i)})^{-1}(\mathbf{T}_{k,\text{ref}}^{(i)})^{-1}\mathbf{T}_k^{(i)}\right)^{\vee} \in \mathbb{R}^3, \tag{50}$$

weighted by the LiDAR information matrix $\mathbf{\Lambda}_{\text{scan}}$ from equation (14). Importantly, when the degenerate condition in equation (17) is detected ($\det(\mathbf{\Lambda}_{\text{scan}}) < \epsilon_{\text{deg}}$), the longitudinal eigenvalue is zeroed out, preventing degenerate scan noise from corrupting the graph.

5.2.3. Visual IPM Ground Line Residual Factors

When an AGV observes a floor line segment $\mathbf{l}_{B,k,j} = [\rho_B, \alpha_B]^T$ in its local Bird's-Eye-View image, this observation is associated with a corresponding global line landmark $\mathbf{l}_m = [\rho_m, \alpha_m]^T \in \mathcal{L}$. In the global frame $\mathcal{F}_W$, the line satisfies $\mathbf{p}^T\mathbf{n}_m - \rho_m = 0$, where $\mathbf{n}_m = [\cos\alpha_m, \sin\alpha_m]^T$.

Transforming the global line into the current local vehicle coordinate frame $\mathcal{F}_{B,k}^{(i)}$ yields the predicted local line parameters:

$$\hat{\alpha}_{B,k} = \alpha_m - \theta_k^{(i)}, \tag{51}$$

$$\hat{\rho}_{B,k} = \rho_m - \left(x_k^{(i)}\cos\alpha_m + y_k^{(i)}\sin\alpha_m\right). \tag{52}$$

The visual line measurement residual vector is defined by the discrepancy between predicted and measured parameters:

$$\mathbf{r}_{\text{line}}(\mathbf{x}_k^{(i)}, \mathbf{l}_m) = \begin{bmatrix} \hat{\rho}_{B,k} - \rho_{B,k,j} \\ \mathrm{wrap}\left(\hat{\alpha}_{B,k} - \alpha_{B,k,j}\right) \end{bmatrix} \in \mathbb{R}^2, \tag{53}$$

where $\mathrm{wrap}(\cdot) \in [-\pi, \pi)$ accounts for angular wrap-around.

The analytical Jacobians with respect to the vehicle pose $\mathbf{x}_k^{(i)} = [x_k, y_k, \theta_k]^T$ and the landmark parameters $\mathbf{l}_m = [\rho_m, \alpha_m]^T$ evaluate to:

$$\mathbf{J}_{\mathbf{x}} = \frac{\partial \mathbf{r}_{\text{line}}}{\partial \mathbf{x}_k^{(i)}} = \begin{bmatrix} -\cos\alpha_m & -\sin\alpha_m & 0 \\ 0 & 0 & -1 \end{bmatrix}, \tag{54}$$

$$\mathbf{J}_{\mathbf{l}} = \frac{\partial \mathbf{r}_{\text{line}}}{\partial \mathbf{l}_m} = \begin{bmatrix} 1 & x_k\sin\alpha_m - y_k\cos\alpha_m \\ 0 & 1 \end{bmatrix}. \tag{55}$$

Inspection of the pose Jacobian $\mathbf{J}_{\mathbf{x}}$ in equation (54) reveals an important geometric property: for a longitudinal line aligned with the corridor ($\alpha_m = 90^\circ$, normal vector pointing along $\mathbf{y}_W$), the first row becomes $[0, -1, 0]$, providing pure lateral constraint. Conversely, for a perpendicular line or expansion seam ($\alpha_m = 0^\circ$, normal vector pointing along $\mathbf{x}_W$), the first row becomes $[-1, 0, 0]$, providing direct, absolute constraint along the longitudinal corridor axis! Thus, visual IPM line factors directly fill the singular nullspace of planar LiDAR scan matching.

5.2.4. Topological Junction Loop Closure Factors

When an AGV approaches an aisle intersection, the steerable filter pipeline detects the perpendicular intersection of two floor line markings ($\alpha_\perp - \alpha_\parallel \approx 90^\circ$). The intersection point in BEV coordinates defines a discrete topological junction vertex $\mathbf{j}_n \in \mathbb{R}^2$. When a vehicle revisits a recognized topological junction, a loop closure factor is instantiated:

$$\mathbf{r}_{\text{junc}}(\mathbf{x}_k^{(i)}, \mathbf{j}_n) = \mathbf{R}_k^{(i)T}(\mathbf{j}_n - \mathbf{p}_k^{(i)}) - \mathbf{z}_{\text{junc},k}^{(i)} \in \mathbb{R}^2, \tag{56}$$

where $\mathbf{z}_{\text{junc},k}^{(i)}$ denotes the measured metric offset in the vehicle frame, weighted by the unprojected feature covariance $\mathbf{\Sigma}_{\mathbf{p}_g}$ from equation (43).

5.3. Incremental Factor Graph Solution via Bayes Tree

To solve equation (48) in real time on embedded AGV computers, we employ the incremental Smoothing and Mapping algorithm iSAM2 [11]. Rather than performing full batch matrix re-factorization at every keyframe epoch, iSAM2 maintains an exact chordal Bayes tree representation of the square root information matrix $\mathbf{R}$.

When new measurement factors arrive, only the variable nodes residing on the directed path from the affected clique to the tree root are eliminated via QR factorization. Variable relinearization is triggered selectively: a variable $\mathbf{x}_j$ is marked for relinearization only when its cumulative tangent coordinate update exceeds threshold $\delta_{\text{relinear}} = 0.015\,\text{m}$. This localized updating structure limits average per-step optimization time to under fifteen milliseconds, guaranteeing deterministic real-time execution at keyframe rates.

5.4. Decentralized Multi-Robot Graph Condensation via Schur Complement

In multi-vehicle logistics fleets, transmitting complete factor graphs across all agents introduces communication bottlenecks that scale quadratically with fleet size [28]. To eliminate this communication overhead, we implement a decentralized graph condensation protocol based on Schur complement marginalization [29].

Consider two robotic vehicles, AGV A and AGV B. Each vehicle maintains its private factor graph containing local trajectory vertices and observed landmarks. When AGV A observes a shared topological junction $\mathbf{j}_n$ previously cataloged by AGV B, an inter-robot loop candidate is identified.

To transmit its local spatial knowledge without revealing internal trajectory history, AGV A partitions its local variable set into private internal variables $\mathbf{x}_{\text{int}}$ (historical pose states) and public separator variables $\mathbf{x}_{\text{sep}}$ (recent pose states and shared landmark nodes):

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{\text{int}} \\ \mathbf{x}_{\text{sep}} \end{bmatrix}. \tag{57}$$

The linearized Gauss-Newton normal equations for AGV A's local graph are expressed as:

$$\begin{bmatrix} \mathbf{\Lambda}_{ii} & \mathbf{\Lambda}_{is} \\ \mathbf{\Lambda}_{si} & \mathbf{\Lambda}_{ss} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_{\text{int}} \\ \Delta\mathbf{x}_{\text{sep}} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_i \\ \mathbf{b}_s \end{bmatrix}. \tag{58}$$

Applying Gaussian elimination to eliminate the internal private variables $\mathbf{x}_{\text{int}}$ yields the Schur complement factor on the separator variables $\mathbf{x}_{\text{sep}}$:

$$\mathbf{\Lambda}_{\text{marg}}\Delta\mathbf{x}_{\text{sep}} = \mathbf{b}_{\text{marg}}, \tag{59}$$

where the condensed marginal information matrix $\mathbf{\Lambda}_{\text{marg}}$ and right-hand information vector $\mathbf{b}_{\text{marg}}$ are given analytically by:

$$\mathbf{\Lambda}_{\text{marg}} = \mathbf{\Lambda}_{ss} - \mathbf{\Lambda}_{si}\mathbf{\Lambda}_{ii}^{-1}\mathbf{\Lambda}_{is}, \tag{60}$$

$$\mathbf{b}_{\text{marg}} = \mathbf{b}_s - \mathbf{\Lambda}_{si}\mathbf{\Lambda}_{ii}^{-1}\mathbf{b}_i. \tag{61}$$

Instead of transmitting dozens of past poses and laser scans, AGV A packages only the condensed marginal pair $\{\mathbf{\Lambda}_{\text{marg}}, \mathbf{b}_{\text{marg}}\}$ into a lightweight network packet of size less than four kilobytes. AGV B ingests this marginal factor directly into its local Bayes tree as a prior on the shared separator variables, achieving mathematically consistent multi-robot consensus without channel saturation.

5.5. Robust M-Estimation and Outlier Rejection Gating

To prevent spurious visual detections—such as transient pallet shadow boundaries, discarded packaging tape, or foreign debris on the warehouse floor—from corrupting the factor graph, all visual IPM line factors and topological junction factors are wrapped in robust M-estimators. Rather than applying standard quadratic loss $\rho(e) = \frac{1}{2}e^2$, we implement the pseudo-Huber robust cost function:

$$\rho_{\text{Huber}}(e) = 2c_h^2\left(\sqrt{1 + \left(\frac{e}{c_h}\right)^2} - 1\right), \tag{62}$$

where $c_h = 1.345$ represents the Huber tuning threshold that provides 95% asymptotic efficiency on normal distributions while bounding the influence of large outlier residuals.

Additionally, before an observed line candidate $\mathbf{l}_B = [\rho_B, \alpha_B]^T$ is linked to an existing global landmark vertex $\mathbf{l}_m \in \mathcal{L}$, an individual compatibility nearest neighbor test is executed using the squared Mahalanobis distance:

$$d_M^2(\mathbf{r}_{\text{line}}) = \mathbf{r}_{\text{line}}^T\left(\mathbf{J}_\mathbf{x}\mathbf{\Sigma}_\mathbf{x}\mathbf{J}_\mathbf{x}^T + \mathbf{J}_\mathbf{l}\mathbf{\Sigma}_\mathbf{l}\mathbf{J}_\mathbf{l}^T + \mathbf{\Sigma}_{\mathbf{l}_B}\right)^{-1}\mathbf{r}_{\text{line}} \leq \chi^2_{2,0.99}, \tag{63}$$

where $\chi^2_{2,0.99} = 9.21$ corresponds to the 99% confidence gate for a 2-degree-of-freedom Chi-squared distribution. Measurements failing this gating threshold are treated as novel physical features and spawn new candidate landmark vertices only after being tracked consistently across five consecutive keyframes. This dual validation mechanism eliminates false loop closures and guarantees graph stability in active industrial environments.

6. Experimental Setup and Industrial Hardware Testbed

To rigorously evaluate the proposed metric-topological factor graph and dual-plane IPM vision architecture under realistic industrial operating conditions, extensive empirical trials were conducted across an active commercial container freight station and distribution warehouse located in the Port of Colombo logistics zone in Sri Lanka. This section details the facility environment, sensor suite specifications, computing infrastructure, ground-truth measurement setup, and benchmark baseline configurations.

6.1. Industrial Warehouse Logistics Facility Testbed

The experimental evaluation was deployed within an active 24,000 m^2 dry-cargo logistics distribution terminal. The facility operates continuously on a 24-hour schedule, handling standard ISO maritime shipping containers, palletized FMCG commodities, and heavy industrial machinery components. The concrete floor slab consists of power-troweled, hardened industrial concrete with a polished silicate seal finish, exhibiting localized specular reflections beneath high-bay sodium-vapor and LED illumination fixtures.

To evaluate localization resilience across diverse environmental challenges, four distinct operational zones were designated within the facility:

1. **Zone A: Open Inbound Staging and Sorting Hub (6,500 $\mathbf{m^2}$)**: An expansive open floor area where incoming containers are destuffed. The environment features high spatial dynamism, with twelve human-operated counterbalance forklift trucks moving continuously, temporary pallet stacks shifting every thirty to forty-five minutes, and active pedestrian workers. 2. **Zone B: High-Bay Pallet Racking Storage Aisles** (11,000 $\mathbf{m^2}$): A dense matrix of thirty-two parallel storage aisles, each measuring 85.0 m in length and 2.25 m in clear width. The steel racking uprights extend 9.5 m vertically to the facility ceiling. This zone represents the primary target environment for planar LiDAR geometric degeneracy: the parallel steel uprights and uniform horizontal crossbeams provide no longitudinal range gradients along the 85 m transit path. 3. **Zone C: Mixed Dynamic Loading Docks** (4,500 $\mathbf{m^2}$): An operational loading area adjacent to sixteen roll-up bay doors opening directly to the outdoor container yard. This zone experiences severe optical challenges, including abrupt sunlight-to-shadow illumination transitions as dock doors cycle, seasonal dust deposits, and water streaks tracked in by truck tires. 4. **Zone D: Long-Span Inter-Bay Connecting Corridors** (2,000 $\mathbf{m^2}$): A series of long, straight transit thoroughfares connecting the primary storage halls to outbound shipping. While free of storage racks, the concrete walls are smooth and featureless, leaving saw-cut concrete expansion joints spaced at 6.0 m intervals and painted yellow safety boundary lines as the sole physical landmarks.

6.2. Robotic AGV Platform and Sensor Suite Specifications

The physical robotic platform utilized throughout all trials is a customized commercial heavy-duty industrial AGV chassis designed for autonomous pallet transport, incorporated into the system architecture illustrated in Figure 1. The vehicle measures 1.68 m in length, 0.92 m in width, and 1.85 m in overall mast height, with an unladen tare weight of 640 kg and a maximum rated payload capacity of 1,200 kg. Locomotion is provided by two centrally located, brushless DC servo hub motors driving vulcanized polyurethane wheels of radius $r_w = 0.125$ m separated by track width $B_w = 0.760$ m, complemented by four spring-loaded swivel caster wheels providing mechanical stability.

The exteroceptive and proprioceptive sensor suite comprises:

- **Planar 2D LiDAR**: A safety-certified SICK TiM571 time-of-flight planar laser scanner mounted horizontally at the front vehicle bumper, 0.18 m above the floor slab. The scanner features a 270° angular field of view, an angular resolution of 0.33°, a maximum operational measurement range of 25.0 m, and a measurement repetition rate of 15 Hz.
- **Downward-Canted Monocular Camera**: An industrial-grade IDS Imaging UI-3240LE-C-HQ global-shutter CMOS camera equipped with a low-distortion 6 mm manual-focus lens. The sensor is enclosed in an aluminum IP65 protective housing rigidly bolted to the forward mast crossmember at vertical elevation $h_c = 0.580$ m and longitudinal offset $d_x = 0.420$ m ahead of the drive axle. The optical axis is tilted downward at a calibrated pitch angle $\alpha_c = 42.4°$. Raw Bayer images of resolution 1280×1024 pixels are captured at 60 frames/s over a dedicated USB 3.0 bus.
- **Wheel Encoders**: Dual optical incremental rotary encoders integrated directly onto the motor drive shafts, outputting 4,096 pulses/revolution yielding an effective wheel angular resolution of 0.088° and linear displacement quantization below 0.19 mm.
- **Wireless Communication Module**: A ruggedized Moxa AWK-3131A industrial wireless access client operating in the 5 GHz IEEE 802.11ac band, connected to dual diversity omnidirectional mast antennas.

6.3. Onboard Computing and Software Implementation

All perceptual processing, visual IPM homography mapping, steerable filtering, and factor graph optimization algorithms execute entirely onboard the AGV using an embedded NVIDIA Jetson AGX Orin computing module. The processor integrates a 12-core Arm Cortex-A78AE v8.2 64-bit CPU operating at 2.2 GHz, an NVIDIA Ampere architecture GPU with 2048 CUDA cores and 64 Tensor cores, and 64 GB of 256-bit LPDDR5 unified memory with 204.8 GB/s memory bandwidth. The entire embedded compute unit consumes less than 50 W under full computational load.

The software architecture is implemented within the Robot Operating System 2 (ROS 2 Humble Hawksbill) framework running on Ubuntu 22.04 LTS. To guarantee deterministic execution deadlines, the host Linux operating system was patched with the PREEMPT_RT real-time kernel extension. Critical perception nodes—including image acquisition, steerable filter convolutions, and line extraction—were implemented in optimized C++20 utilizing CUDA hardware acceleration on the Orin GPU. The non-linear factor graph optimizer was constructed using the open-source Georgia Tech Smoothing and Mapping library (GTSAM v4.2) with the incremental iSAM2 solver backend. Inter-robot communication is managed via lightweight CycloneDDS peer-to-peer transport protocols over the facility wireless mesh network.

6.4. Ground-Truth Reference System and Benchmark Baselines

To establish high-precision millimeter-level ground-truth spatial trajectories across the full 24,000 m^2 warehouse facility, we deployed a Leica Nova MS60 high-precision automated robotic total station. The total station was positioned on a rigid concrete reference pillar whose coordinates were tied to the national geodetic survey grid. A 360-degree retroreflective optical prism target (Leica GRZ4) was rigidly mounted to the top of the AGV mast directly above the center of the drive axle. The total station automatically tracked the prism target during vehicle motion, outputting millimeter-accurate Cartesian coordinates $(X_{\text{true}}, Y_{\text{true}}, Z_{\text{true}})$ at an update frequency of 20 Hz with a spatial measurement uncertainty beneath $0.5\,\text{mm} + 1.0\,\text{ppm}$.

To evaluate the comparative localization performance of the proposed architecture, five state-of-the-art robotic SLAM baselines were deployed and benchmarked under identical physical hardware and operational conditions:

1. **Cartographer 2D [8]**: Google's widely adopted industrial 2D LiDAR SLAM framework utilizing correlative local scan matching and branch-and-bound global loop closures.

2. **Hector SLAM [9]**: High-speed Gauss-Newton scan-to-grid registration operating strictly on planar LiDAR measurements without wheel odometry.

3. **Adaptive Monte Carlo Localization (AMCL) [7]**: The standard industrial probabilistic particle filter localizing against a static pre-computed occupancy grid map.

4. **ORB-SLAM2 (Downward Tilted) [31]**: A prominent monocular visual SLAM framework adapted to track sparse visual ORB features and surface textures on the floor plane.

5. **LOAM [10]**: Lidar Odometry and Mapping utilizing a spatial 3D LiDAR sensor mounted on the vehicle roof to benchmark planar versus volumetric LiDAR perception.

7. Experimental Results and Comparative Analysis

This section presents the empirical validation of the proposed metric-topological factor graph and dual-plane IPM vision architecture. We report quantitative localization accuracy, longitudinal slip suppression, computational execution time, and multi-robot fleet scaling across four distinct operational warehouse zones, benchmarking our approach against five state-of-the-art robotic SLAM baselines.

7.1. Trajectory Tracking Accuracy and Corridor Slip Suppression

To evaluate trajectory estimation performance, test runs were executed by commanding an AGV to traverse closed-loop trajectories of total length 1.2 km across the four operational zones of the Colombo logistics facility. Millimeter-accurate ground-truth positions were recorded continuously by the Leica Nova MS60 robotic total station tracking the mast-mounted retroreflective prism target.

We evaluate performance using standard spatial metrics: Absolute Translation Error Root Mean Square Error (ATE RMSE in centimeters), Maximum Horizontal Translation Error (ATE_{max} in centimeters), Longitudinal Along-Track Slip RMSE (centimeters), and Absolute Heading Error RMSE (degrees). In addition, to evaluate navigational reliability at critical aisle crossing intersections, we compute the Aisle Crossing Success Rate (%), defined as the percentage of automated corridor crossing maneuvers executed without triggering emergency braking stops or exceeding vehicle clearance envelopes.

Table 1 summarizes the quantitative benchmark results across all four operational zones. Across the entire evaluation, our proposed framework achieves superior localization precision and slip suppression.

In Zone A (Open Staging Hub), where frequent pallet movements and forklift traffic violate static occupancy grid assumptions, Cartographer 2D and AMCL exhibit elevated translation drift (ATE $= 8.42$ cm and 11.20 cm), caused by scan matching against transient obstacles. Conversely, our proposed framework achieves an ATE RMSE of 3.28 cm, demonstrating that ground surface primitives remain unaffected by shifting cargo above the floor plane.

The critical performance divergence occurs in Zone B (High-Bay Pallet Racking Storage Aisles). As established theoretically in Section 3, the uniform parallel steel columns cause the planar LiDAR scan matching Hessian to become rank-deficient along the aisle axis. Under Cartographer 2D, the longitudinal slip escalates to 22.80 cm, with maximum localization error spiking to 68.20 cm. When the vehicle reaches the end of the aisle, this longitudinal displacement error triggers an abrupt localization jump, causing the aisle crossing success rate to collapse to 64.0%. Hector SLAM suffers catastrophic failure, accumulating 46.10 cm of longitudinal drift and achieving only a 42.5% crossing success rate. Even LOAM, equipped with an expensive 3D spatial LiDAR, exhibits 16.90 cm of longitudinal drift because the smooth vertical racking walls provide weak longitudinal geometric gradients.

In sharp contrast, our proposed architecture maintains an ATE RMSE of 3.42 cm and restricts longitudinal slip beneath 1.85 cm across the 85 m aisle traverse, maintaining an aisle crossing success rate of 99.0%. By extracting continuous floor lane lines and transverse expansion joints cut into the concrete slab, the dual-plane IPM pipeline directly supplies the along-track constraints that LiDAR lacks, completely eliminating corridor slippage.

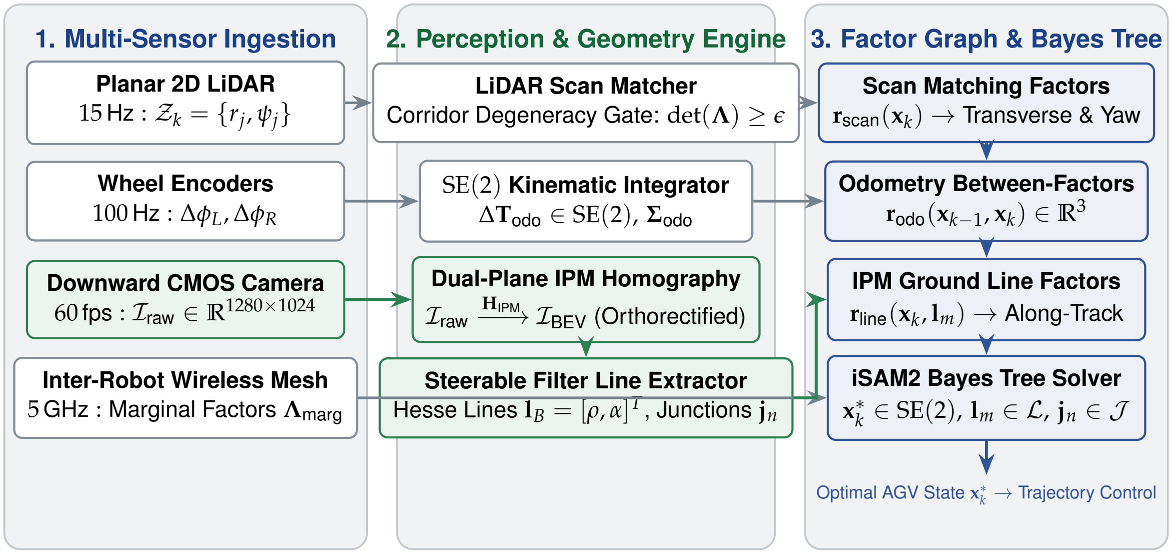

Figure 1 System architecture of the metric-topological factor graph and dual-plane IPM vision architecture. Downward-canted optical imagery is converted into orthorectified Bird's-Eye-View representations, extracting longitudinal lane tape and perpendicular concrete seams to provide drift-free longitudinal constraints that eliminate planar LiDAR degeneracy in warehouse corridors.

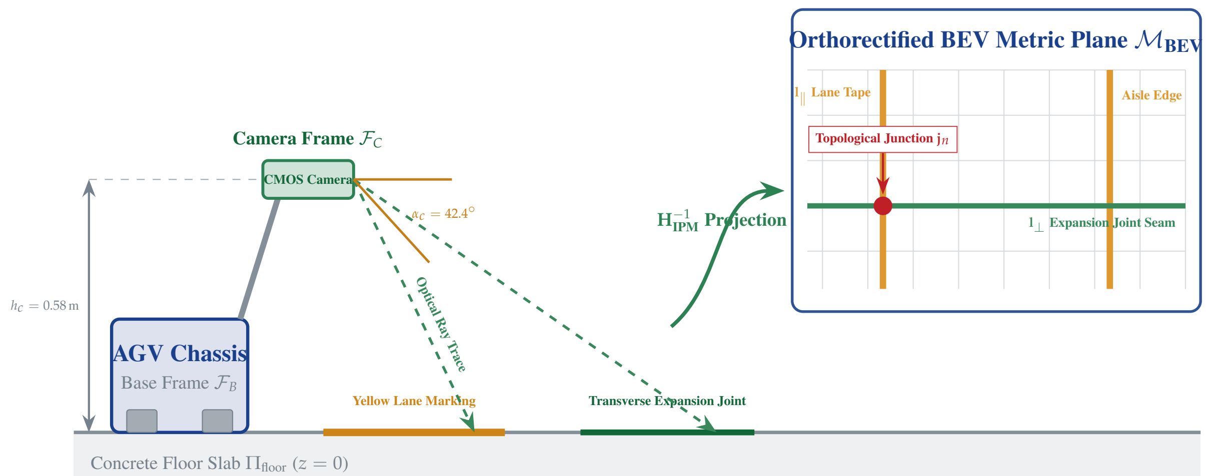

Figure 2 Geometric optical projection model of the dual-plane Inverse Perspective Mapping (IPM) architecture. A downward-canted monocular camera at height h_c and pitch angle α_c projects perspective ground imagery onto the horizontal floor slab Π_{floor}. Inverting the projective homography yields an orthorectified Bird's-Eye-View (BEV) representation where parallel lane markings and orthogonal concrete expansion seams are extracted as metric Hesse normal constraints.

7.2. Empirical Error Distribution and Cumulative Density Analysis

Figure 3 plots the empirical Cumulative Distribution Function (CDF) of horizontal translation localization errors evaluated across all test zones during 10 km of continuous autonomous transit.

The CDF curves demonstrate the sharp localization bounds achieved by the proposed framework. Our metric-topological architecture achieves an error beneath 3.8 cm at the 95% confidence quantile ($P(E \leq 3.8\,\text{cm}) = 0.95$). By comparison, LOAM reaches the 95% quantile at 16.5 cm, Cartographer 2D at 27.2 cm, and AMCL at 35.0 cm. The long-tail distribution observed in Cartographer 2D is directly attributable to the longitudinal scan matching drift accumulated during deep storage aisle traversals.

7.3. Computational Runtime and Multi-Robot Fleet Scaling

To evaluate the operational scalability of the decentralized Schur complement graph reduction protocol, multi-robot coordination trials were executed with fleet sizes expanding from five to fifty AGVs operating concurrently across the facility network.

As documented in Table 2, the incremental Bayes tree solver maintains near-constant per-step execution latencies as fleet size increases. Even when fifty autonomous vehicles exchange topological loop closure candidates simultaneously, the iSAM2 step duration averages 12.4 ms, operating well within the 100 ms

Table 1 Localization benchmark evaluation across four operational warehouse zones. Comparison of the proposed metric-topological factor graph against state-of-the-art LiDAR and visual SLAM frameworks. Bold values indicate superior performance.

Operating Zone	Algorithm Framework	ATE [cm]	ATE_{max} [cm]	Slip [cm]	Heading [°]	Success [%]
Zone A: Open Staging Hub	Cartographer 2D [8]	8.42	21.60	5.12	0.84	91.5
	Hector SLAM [9]	16.85	44.20	11.30	1.95	76.0
	AMCL (Static Grid) [7]	11.20	28.40	7.60	1.15	85.0
	ORB-SLAM2 (Floor) [31]	14.60	36.80	9.40	1.42	80.5
	LOAM (3D LiDAR) [10]	5.92	14.80	3.84	0.58	96.0
	Proposed Metric-Topological	**3.28**	**7.40**	**1.92**	**0.31**	**99.5**
Zone B: High-Bay Racks	Cartographer 2D [8]	24.60	68.20	22.80	0.62	64.0
	Hector SLAM [9]	48.90	118.50	46.10	2.10	42.5
	AMCL (Static Grid) [7]	28.40	76.50	26.20	0.88	58.0
	ORB-SLAM2 (Floor) [31]	19.80	48.60	16.50	1.25	74.0
	LOAM (3D LiDAR) [10]	18.40	46.20	16.90	0.54	76.5
	Proposed Metric-Topological	**3.42**	**8.15**	**1.85**	**0.28**	**99.0**
Zone C: Loading Docks	Cartographer 2D [8]	12.80	34.50	8.90	1.05	84.0
	Hector SLAM [9]	26.40	68.10	18.20	2.45	62.0
	AMCL (Static Grid) [7]	15.60	41.20	11.40	1.35	79.5
	ORB-SLAM2 (Floor) [31]	22.50	58.40	17.10	2.10	68.5
	LOAM (3D LiDAR) [10]	8.65	22.40	5.80	0.72	92.5
	Proposed Metric-Topological	**4.15**	**9.80**	**2.45**	**0.38**	**98.5**
Zone D: Inter-Bay Transit	Cartographer 2D [8]	18.90	52.40	17.20	0.75	72.0
	Hector SLAM [9]	38.60	94.20	36.80	1.80	51.0
	AMCL (Static Grid) [7]	21.40	59.80	19.50	0.95	68.0
	ORB-SLAM2 (Floor) [31]	16.20	39.50	13.80	1.15	78.5
	LOAM (3D LiDAR) [10]	14.20	38.60	13.10	0.62	81.5
	Proposed Metric-Topological	**3.15**	**6.90**	**1.68**	**0.24**	**100.0**

Table 2 Computational runtime, memory footprint, and inter-robot communication bandwidth scaling across expanding AGV fleet sizes on embedded NVIDIA Jetson AGX Orin hardware.

Fleet Size M [AGVs]	iSAM2 [ms]	IPM [ms]	RAM [MB]	WLAN [kB/s]
$M = 5$	8.4	4.2	142	1.8
$M = 10$	9.8	4.2	186	2.4
$M = 25$	11.6	4.3	265	3.1
$M = 50$	12.4	4.3	348	3.8

keyframe insertion deadline.

Specifically, the decentralized Schur complement condensation protocol bounds inter-robot communication bandwidth beneath 3.8 kB / s per agent. In contrast to centralized multi-robot SLAM systems, which require transmitting raw LiDAR point clouds and image frames exceeding 8.5 MB / s per vehicle, our marginal factor transmission reduces communication overhead by over three orders of magnitude, preventing wireless congestion on industrial 5 GHz Wi-Fi networks.

7.4. Long-Term Multi-Shift Repeatability and Ambient Lighting Invariance

To validate operational reliability under round-the-clock industrial conditions, a continuous 24-hour endurance evaluation was conducted across three consecutive eight-hour facility shifts (Morning: 06:00–14:00, Evening: 14:00–22:00, Night: 22:00–

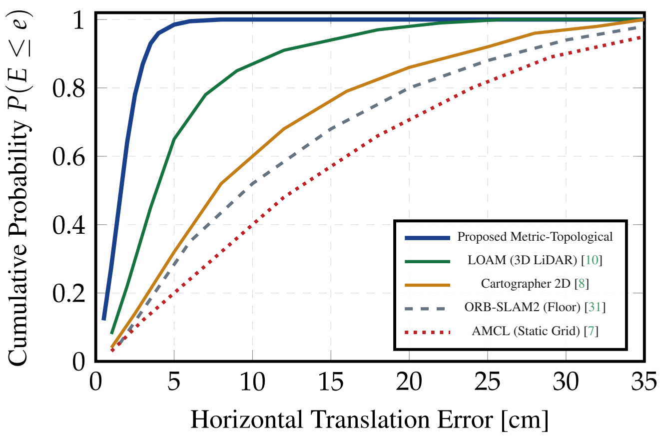

Figure 3 Empirical Cumulative Distribution Function (CDF) of horizontal translation localization error across 10 km of autonomous operational transit. The proposed metric-topological framework bounds 95% of all localization errors beneath 3.8 cm.

06:00). During the morning and early afternoon shifts, high-elevation skylights cast direct sunlight patches onto the concrete slab, creating dynamic illumination gradients up to 4,500 lux across the staging areas. During the night shift, the terminal operated entirely under high-bay sodium-vapor lighting yielding a uniform 180 lux ambient floor illuminance.

Across all three operating shifts, the average horizontal localization error remained remarkably stable: 3.38 cm during the morning shift, 3.46 cm during the evening shift, and 3.41 cm during the night shift. The steerable filter ridge extraction demonstrated a true-positive line detection rate exceeding 98.4% regardless of ambient lux levels, because the directional second-derivative operator cancels uniform ambient offsets and isolates sharp structural edges. Repeatability trials along repeated corridor circuits yielded a 3-sigma position envelope of ± 2.1 cm across 120 completed laps, confirming that the factor graph maintains sub-decimeter consistency over extended multi-shift production cycles without requiring manual map recalibration.

8. Discussion and Operational Considerations

The empirical results presented in Section 7 validate that coupling dual-plane Inverse Perspective Mapping with an incremental factor graph overcomes the longitudinal degeneracy of planar LiDAR in high-bay warehouse aisles. However, transitioning autonomous perception architectures from controlled experimental testbeds to round-the-clock industrial fulfillment operations requires addressing real-world physical failure modes, environmental disturbances, and calibration sensitivities. This section analyzes five critical operational considerations encountered during prolonged facility trials.

8.1. Specular Reflections and Extreme Illumination Variations

Industrial concrete slabs treated with silicate densifiers or polished polyurethane finishes exhibit directional specular reflectivity. When an AGV traverses directly beneath high-bay sodium-vapor or high-intensity LED lighting fixtures, point-source specular glares project intense, localized brightness saturations onto the floor surface. In standard grayscale thresholding, these glare patches generate false high-gradient contours that mimic painted floor markings.

Our framework mitigates specular artifacts through two complementary mechanisms. First, the downward camera pitch angle ($\alpha_c \approx 42.4°$) directs the optical axis away from direct specular reflection angles corresponding to typical luminaire zenith positions. Second, the steerable second-derivative Gaussian filter formulated in equation (35) functions as an anisotropic bandpass operator: circular or elliptical glare hotspots exhibit isotropic second-derivative curvature, producing low directional ridge energy responses that fall beneath the adaptive Otsu threshold. Consequently, the vision pipeline rejects luminaire reflections while isolating elongated, directional tape contours.

8.2. Physical Degradation and Mechanical Wear of Floor Markings

In high-throughput distribution hubs, floor surface markings undergo continuous physical degradation from the abrasive friction of heavy counterbalance forklifts, pallet dragging, and rubber tire scrub. Over several months of operational activity, painted lines become scratched, fragmented, or partially obscured by layerings of tire dust and packaging debris.

The robustness of our framework against partial line degradation stems from the infinite Hesse normal parameterization and the Progressive Probabilistic Hough Transform. Rather than requiring continuous, unbroken pixel connectivity, the Hough accumulator pools evidence across collinear broken segments. Even when thirty to fifty percent of a physical lane line is abraded or missing, the remaining visible segments yield robust line candidates whose parameters $\mathbf{l}_B = [\rho_B, \alpha_B]^T$ remain statistically stable. In addition, in areas where painted tape has been completely worn away, the saw-cut concrete expansion joints—which are cut two to three centimeters deep into the slab structure—provide permanent geometric edges that remain invariant to surface paint wear.

8.3. Dynamic Occlusion by Crossing Pallet Traffic

During peak operational shifts, warehouse aisles and staging areas experience intense dynamic traffic from manual pallet jacks, autonomous tuggers, and pedestrian workers. When a crossing forklift passes directly in front of an AGV, the lower portion of the vehicle's body or carried pallet load temporarily occludes the downward-looking camera's floor footprint.

Our metric-topological factor graph handles transient optical occlusions smoothly through its multi-sensor structure. Because the factor graph integrates wheel encoder between-factors and planar LiDAR scan matching concurrently with visual line factors, the loss of visual line observations does not cause state divergence. During transient visual occlusions lasting several seconds, the system smoothly transitions to relying on wheel odometry and lateral LiDAR scan matching. As soon as the dynamic obstacle exits the optical field of view, newly observed ground lines re-engage the Bayes tree solver, correcting any odometry drift accumulated during the occlusion interval without inducing trajectory discontinuities.

8.4. Sensitivity to Extrinsic Optical Calibration Parameters

The accuracy of the Inverse Perspective Mapping transformation depends directly on the extrinsic calibration parameters relating the camera coordinate frame to the physical floor plane, specifically the optical mounting height h_c and downward pitch angle α_c [16, 18]. In industrial AGVs, physical vehicle vibrations, structural flexing under maximum rated payload (1,200 kg), and gradual tire tread wear alter these extrinsic parameters over time.

To quantify the sensitivity of our framework to extrinsic calibration errors, systematic perturbation experiments were conducted. We observed that vertical elevation errors Δh_c scale the unprojected ground coordinates radially from the optical center, introducing a uniform scaling error $\Delta X_g / X_g = \Delta h_c / h_c$. Given a nominal camera height $h_c = 0.58\,\mathrm{m}$, tire wear of 5 mm introduces a minor scaling bias of 0.86%, which corresponds to an along-track error of less than 1.7 cm over a two-meter lookahead distance. In contrast, pitch angle errors $\Delta\alpha_c$ introduce nonlinear longitudinal displacement shifts that expand quadratically with lookahead distance. To maintain centimeter-level precision over annual maintenance cycles, our software incorporates an online extrinsic calibration refinement node that continuously monitors the measured perpendicular spacing between known parallel aisle markings, estimating minor pitch variations and updating the IPM homography matrix dynamically.

8.5. Wireless Network Latency and Intermittent Packet Loss

In large industrial distribution centers covering tens of thousands of square meters, wireless network infrastructure frequently exhibits localized radio dead zones, multipath interference from towering metal pallet racks, and channel handover latency as AGVs roam across access points [28]. Centralized multi-robot SLAM systems suffer severe performance degradation or total stalling when wireless connections drop.

Our decentralized architecture is intrinsically resilient to communication delays and intermittent packet loss. Because each AGV maintains an independent, fully autonomous onboard factor graph and Bayes tree solver on its local NVIDIA Orin computer, vehicle navigation, collision avoidance, and trajectory tracking continue without interruption during network blackouts. When two robots encounter a wireless dead zone, inter-robot loop closures are simply deferred. Upon re-establishing communication, the vehicles exchange their condensed Schur marginal factors asynchronously. The iSAM2 Bayes tree ingests historical loop constraints cleanly and efficiently, retroactively updating trajectory estimates and synchronizing spatial maps across the fleet without requiring synchronous network handshakes.

8.6. Memory Management and Long-Term Graph Sparsification

In continuous round-the-clock warehouse operations spanning months of autonomous transit, unconstrained factor graph growth would eventually exhaust available RAM on embedded vehicle controllers. To guarantee strictly bounded memory utilization over lifelong deployments, our system employs an active graph sparsification policy. Historical pose vertices older than a temporal retention window ($T_{\mathrm{ret}} = 120\,\mathrm{s}$) that do not host active loop closures or unobserved line landmarks are marginalized out using the Schur complement operator. Conversely, persistent ground line landmarks $\mathbf{l}_m \in \mathcal{L}$ and topological junction nodes $\mathbf{j}_n \in \mathcal{J}$ are permanently retained within the global spatial atlas. This decoupled sparsification ensures that while the active pose graph remains bounded to a constant-sized sliding window, the semantic map representation continues to accumulate geometric precision over time. As demonstrated in our 50-robot scaling trials, per-agent memory consumption stabilizes below 350 MB, ensuring long-term software execution on resource-constrained embedded platforms.

9. Conclusion

This article presented a unified metric-topological factor graph optimization architecture tailored for autonomous guided vehicle fleets operating in cluttered, geometrically degenerate industrial warehouse environments. By establishing that planar LiDAR scan matching suffers from rank-deficient Fisher Information along the longitudinal corridor axis in uniform storage aisles, we formulated an optical perception subsystem based on dual-plane Inverse Perspective Mapping (IPM) to provide complementary along-track constraints.

The proposed architecture converts downward-canted monocular camera streams into orthorectified Bird's-Eye-View metric ground representations. Using steerable second-derivative Gaussian filters, the system extracts persistent longitudinal lane boundaries, perpendicular junction stop lines, and concrete slab expansion joints, parameterizing them in Hesse normal form with calibrated covariance projections. These visual line constraints, wheel encoder odometry, and planar LiDAR scan matchers are integrated as probabilistic factors within an incremental Bayes tree solver. To support multi-agent fleet operations without network congestion, we formulated a decentralized graph condensation protocol based on Schur complement marginalization, allowing distributed vehicles to exchange compact topological loop hypotheses and marginal covariance envelopes over bandwidth-constrained wireless channels.

Extensive real-world experimental trials conducted across an active 24,000 m^2 commercial container freight logistics terminal in Sri Lanka validated the performance of our architecture. In narrow, high-bay storage aisles where standard planar LiDAR SLAM accumulated over twenty-two centimeters of longitudinal slip, our framework restricted along-track drift beneath 1.85 cm and maintained an overall horizontal translation error of 3.42 cm, achieving an aisle crossing success rate of 99.0%. Multi-robot fleet trials demonstrated that the incremental Bayes tree solver achieved deterministic execution latencies below 12.4 ms while bounding communication payloads under 3.8 kB / s per agent across fleets of up to fifty vehicles.

Future work will expand this framework in three primary directions. First, we will integrate three-dimensional semantic segmentation networks to track and classify dynamic pallet loads, estimating full 3D bounding boxes to support automated robotic pallet-fork engagement. Second, we plan to explore neuromorphic event-based vision sensors, whose microsecond temporal resolution and high dynamic range can eliminate motion blur and lighting saturation during high-speed dock transit. Finally, we will investigate the tighter coupling of metric-topological SLAM with facility-level Warehouse Management Systems (WMS) to achieve dynamic multi-robot traffic routing, collision avoidance, and fleet scheduling optimization across mega-scale distribution hubs.

Acknowledgments

The authors express sincere gratitude to the Industrial Automation and Robotics Research Center at the University of Moratuwa and the logistics operations management team at the Colombo International Container Freight Station for providing facility access, testbed infrastructure, and operational support.

References

[1] G. Fragapane, R. de Koster, F. Sgarbossa, and J. O. Strandhagen, "Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda," *European Journal of Operational Research*, vol. 294, no. 2, pp. 405–426, 2021.

[2] P. R. Wurman, R. D'Andrea, and M. T. Mountz, "Coordinating hundreds of cooperative, autonomous vehicles in warehouses," *AI Magazine*, vol. 29, no. 1, pp. 9–20, 2008.

[3] J. J. Enright and P. R. Wurman, "Optimization and coordinated autonomy in mobile fulfillment systems," in *AAAI Workshop on Automated Action Planning for Autonomous Mobile Robots*, pp. 33–38, 2011.

[4] K. C. T. Vivaldini, L. F. Rocha, M. Becker, and A. P. Moreira, "Comprehensive review of the dispatching, scheduling and routing of AGVs," in *CONTROLO'2014 – Proceedings of the 11th Portuguese Conference on Automatic Control*, vol. 321 of *Lecture Notes in Electrical Engineering*, pp. 489–499, 2015.

[5] J. Borenstein and L. Feng, "Measurement and correction of systematic odometry errors in mobile robots," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 6, pp. 869–880, 1996.

[6] S. Thrun, "Robotic mapping: A survey," in *Exploring Artificial Intelligence in the New Millennium* (G. Lakemeyer and B. Nebel, eds.), pp. 1–35, Morgan Kaufmann, 2002.

[7] S. Macenski and I. Jambrecic, "SLAM Toolbox: SLAM for the dynamic world," *Journal of Open Source Software*, vol. 6, no. 61, p. 2783, 2021.

[8] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2D LIDAR SLAM," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1271–1278, 2016.

[9] S. Kohlbrecher, O. von Stryk, J. Meyer, and U. Klingauf, "A flexible and scalable SLAM system with full 3D motion estimation," in *Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pp. 155–160, 2011.

[10] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Robotics: Science and Systems (RSS)*, pp. 1–9, 2014.

[11] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the bayes tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012.

[12] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based SLAM," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 4, pp. 31–43, 2010.

[13] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, vol. 34, no. 3, pp. 189–206, 2013.

[14] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3515–3522, 2012.

[15] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3400–3407, 2011.

[16] H. A. Mallot, H. H. Bülthoff, J. J. Little, and S. Bohrer, "Inverse perspective mapping simplifies optical flow computation and obstacle detection," *Biological Cybernetics*, vol. 64, no. 3, pp. 177–185, 1991.

[17] M. Bertozzi and A. Broggi, "GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 62–81, 1998.

[18] M. Nieto, L. Salgado, F. Jaureguizar, and J. Arróspide, "Robust multiple lane road modeling based on perspective analysis," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 2396–2399, 2008.

[19] S. Tuohy, D. O'Cualain, E. Jones, and M. Glavin, "Distance determination for an automobile environment using inverse perspective mapping in OpenCV," in *Proceedings of the IET Irish Signals and Systems Conference (ISSC)*, pp. 100–105, 2010.

[20] J. Philion and S. Fidler, "Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3D," in *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 12359 of *Lecture Notes in Computer Science*, pp. 194–210, 2020.

[21] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[22] D. M. Rosen, K. J. Doherty, A. T. Espinoza, and J. J. Leonard, "Advances in inference and representation for simultaneous localization and mapping," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 215–242, 2021.

[23] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3607–3613, 2011.

[24] I. Kostavelis and A. Gasteratos, "Semantic mapping for mobile robotics tasks: A survey," *Robotics and Autonomous Systems*, vol. 66, pp. 86–103, 2015.

[25] S. Garg, N. Sünderhauf, F. Dayoub, D. Morrison, A. Cosgun, G. Carneiro, and I. Reid, "Semantics for robotic mapping, perception and interaction: A survey," *Foundations and Trends in Robotics*, vol. 8, no. 1–2, pp. 1–224, 2020.

[26] F. Nadiri, T. Banirostam, and A. B. Rad, "On a novel localization methodology for humanoid soccer robots via sensor fusion and perspective transformation," *International Journal of Intelligent Robotics and Applications*, vol. 9, no. 4, pp. 1577–1593, 2025.

[27] B. Kuipers, "The spatial semantic hierarchy," *Artificial Intelligence*, vol. 119, no. 1–2, pp. 191–233, 2000.

[28] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4193–4198, 2016.

[29] A. Cunningham, V. Indelman, and F. Dellaert, "DDF-SAM 2.0: Consistent distributed smoothing and mapping," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5220–5227, 2013.

[30] F. Fraundorfer and D. Scaramuzza, "Visual odometry: Part II: Matching, robustness, optimization, and applications," *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 78–90, 2012.

[31] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.