IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Brain-Inspired Visual Topometric Localization via Roadnetwork-Constraint Hidden Markov Model

Jinyu Li[1], Taiping Zeng[2] and Bailu Si[1]

***Abstract*—Accurate localization in GPS-denied environments remains a critical challenge for autonomous robot navigation. Animals exhibit remarkable navigational abilities in complex, dynamic environments by relying on mental cognitive maps. Inspired by neural representations such as head direction cells and grid cells, numerous robotic cognitive mapping systems can efficiently cover large areas; however, they often lack the precise metric information required for accurate localization. To address this challenge, we propose a neurodynamically driven monocular visual topometric localization approach based on road network constraints. We introduce the Roadnetwork-Constraint Hidden Markov Model (RC-HMM) to enhance the semi-metric map by incorporating road network constraints, forming a coherent topometric map that maintains vertex relationships and improves localization accuracy. Experimental results in the CARLA Town07 environment demonstrate the remarkable efficiency of our topometric cognitive map. Compared to the semi-metric map, our approach achieves a 95% reduction in Absolute Pose Error (APE) and an 81% reduction in Relative Pose Error (RPE). Compared to binocular ORB-SLAM3, our monocular approach reduces CPU usage by 96.7% and map storage by 77.7%, with an APE of 3.6 m and RPE of 1.4 m — closely matching ORB-SLAM3's 3.86 m APE and 0.96 m RPE. Furthermore, by leveraging neurodynamics of grid cells and head direction cells, our monocular topometric localization robustly delivers the localization accuracy of 3.86 meters, comparable to binocular ORB-SLAM3. This approach integrates road network metrics into topological maps, enhancing brain-inspired navigation with topometric maps in complex environments. A project webpage is available at https://brain-inspired-navigation.github.io/topometric-loc/.**

***Index Terms*—Biologically-Inspired Robots, Localization, Cognitive Map, SLAM, Neurorobotics**

I. Introduction

ROBOTS are evolving from research lab prototypes to practical real-world applications, where accurate localization is critical for reliable navigation. While GPS is commonly used due to its simplicity and low cost, it often proves unreliable in large-scale, dynamic, unpredictable environments such as urban areas, forests, or tunnels, where signal loss or inaccuracies are frequent. Consequently, developing robust GPS-free localization systems has become crucial for robotic tasks, such as autonomous driving in urban environments, search and rescue in disaster zones, and navigation in GPS-denied areas like warehouses, forests, and underwater [1].

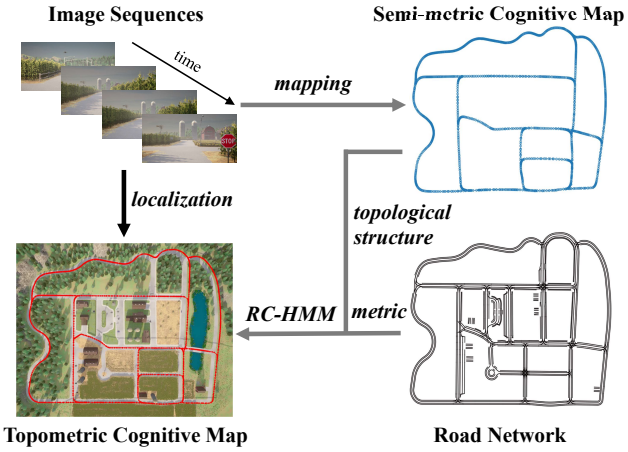

Fig. 1. **Overview of the Topometric Mapping Process.** Continuous image sequences and self-motion are input to build a semi-metric cognitive map. The RC-HMM method then integrates road network data to enhance the map's metric accuracy.

However, animals show amazing navigation ability [2], [3], [4], [5], efficiently navigating large-scale, dynamic environments for a very long time due to sophisticated cognitive and neural mechanisms. Cognitive maps of the environment were proposed to describe the internal representation of spatial environments in animal brains that support their superior navigation abilities [6]. Hippocampal place cells [7], head direction cells [8], and entorhinal grid cells [9] provide a neural basis for cognitive maps, facilitating spatial navigation and environmental representation [10], [11], [12].

The neural mechanisms underlying mammalian navigation have long inspired robot navigation systems. Brain-inspired cognitive models of navigation can be broadly categorized into two groups, each striking a different balance between biological realism and practical performance. One group consists of models designed to test and validate theories of biological navigation mechanisms [13], [14], [15]. The other, much smaller group of models employed neural systems to construct semi-metric cognitive maps of the environment [16], [17], [18], [19]. However, these semi-metric cognitive maps do not incorporate enough metric information, which limits the accuracy of robot localization and poses challenges for tasks that demand accurate localization.

To enhance the accuracy of semi-metric mapping and

Manuscript received: March, 3, 2025; Revised June, 3, 2025; Accepted August, 26, 2025.

This paper was recommended for publication by Editor Tetsuya Ogata upon evaluation of the Associate Editor and Reviewers' comments. *(Corresponding authors: Taiping Zeng)*

[1]Jinyu Li and Bailu Si are with School of Systems Science, Beijing Normal University, Beijing 100091, China jinyuli@mail.bnu.edu.cn, bailusi@bnu.edu.cn

[2]Taiping Zeng is with Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 201203, China zengtaiping@fudan.edu.cn

Digital Object Identifier (DOI): see top of this page.

localization in brain-inspired navigation systems, we propose the Roadnetwork-Constraint Hidden Markov Model (RC-HMM) for constructing topometric maps (Fig. 1). This method integrates semi-metric map and road network, effectively transforming a semi-metric cognitive map into a topometric cognitive map. The sequential visual inputs stimulate local view cells, which activate the neurodynamics of grid cells and head direction cells, thus facilitating precise localization. In the CARLA Town07 environment, our monocular system demonstrates comparable performance to binocular ORB-SLAM3 in both mapping and localization accuracy. Our contributions are summarized as follows:

- We propose a brain-inspired monocular visual localization method for GPS-denied environments, using neural dynamics of grid and head-direction cells to build a coherent topometric cognitive map that integrates topological and metric information for precise robot localization.
- We introduce RC-HMM incorporates road network metric into a semi-metric topological map. By jointly optimizing global vertex projections and local distance constraints, our method constructs a consistent and metrically accurate topometric map that supports robust localization by neurodynamics.
- We validate our topometric approach in Carla Town07, reducing APE by 95% and RPE by 81% compared to semi-metric maps. Against binocular ORB-SLAM3, our monocular map achieves an APE of 3.6 m and RPE of 1.4 m (vs. 3.86 m and 0.96 m), while cutting CPU usage by 96.7%, map storage by 77.7%, and localization accuracy remains comparable (3.86 m vs. 4.63 m) with a 99% reduction of initial localization time.

II. RELATED WORKS

Simultaneous Localization and Mapping (SLAM) plays a crucial role in enabling robots to map their environment and determine their location in real-time [20]. SLAM algorithms are generally classified into two categories: metric-based SLAM and topological-based SLAM, with each offering distinct advantages depending on the application's demands for precision and scalability.

1) Metric SLAM: Metric SLAM, encompassing both LiDAR and visual methods, is commonly used for accurate localization. LiDAR-based approaches like Cartographer and LOAM combine IMU data with laser scans for real-time mapping [21], [22]. To reduce drift, LIO-SAM [23] improves trajectory estimation with a nonlinear motion model, while LEGO-LOAM [24] segments point clouds for better efficiency on low-power systems. Visual SLAM methods like ORB-SLAM series [25], [26], [27], as well as approaches incorporating map constraints [28], rely on monocular, stereo, and RGB-D cameras for feature-based localization and image alignment. However, metric SLAM requires high computational and storage resources, limiting its scalability and real-time performance in large-scale environments.

2) Topological SLAM: Topological SLAM offers a scalable solution by representing environments through key landmarks (vertices) and their spatial relationships (edges). A notable example is the brain-inspired RatSLAM [16], [17], which integrates odometry and landmarks via a competitive attractor network, enabling the creation of semi-metric cognitive maps in large, real-world environments. Yu et al. [19] extended RatSLAM [16], [17] to 3D environments and validated it in real-world scenarios. In contrast, our proposed NeuroBayesSLAM [18], [29] uses a Bayesian attractor network integrating visual and vestibular cues by parametric probabilistic distributions, which reduces computational cost and resource usage while maintaining biological plausibility. It encodes orientation and position through head direction and grid cell networks, offering robust performance in large-scale environments. However, while these methods are scalable and resilient, they typically provide only approximate position information and lack the fine-grained precision of metric localization. Furthermore, most existing brain-inspired SLAM methods primarily focus on mapping quality while overlooking map reusability.

Current topometric mapping integrates metric data from diverse sources such as GPS [30], visual SLAM [31], or dense LiDAR [32]. Badino et al. [33] combine GPS with visual and 3D features for real-time localization. Topomap [31] transforms sparse visual SLAM maps into topological graphs by clustering free space and extracting occupancy data. Similarly, the hybrid framework [32] builds dense indoor maps via autonomous exploration, maintaining submaps to support efficient planning. However, these methods are dependent on high-precision GPS data or rely heavily on dense fine-grained metric information from LiDAR or vision sensors, resulting in increased computational and storage demands.

In this paper, we present the RC-HMM for constructing topometric cognitive maps in GPS-denied situation, leveraging ProtoBuf for efficient map reuse. This approach facilitates precise localization while minimizing computational requirements.

III. TOPOMETRIC COGNITIVE MAPPING AND VISUAL LOCALIZATION

We present a brain-inspired visual localization method based on a topometric cognitive map, constructed using the proposed RC-HMM. RC-HMM integrates a semi-metric topological map – generated by our previous NeuroBayesSLAM [18] – with metric information extracted from road networks via Hidden Markov Model (HMM) map matching [34]. By leveraging the neurodynamics of grid cells and head direction cells, precise localization is achieved within the topometric cognitive map.

A. Semi-metric Cognitive Mapping

1) Bayesian Attractor Model: The NeuroBayesSLAM framework [18] integrates visual and vestibular inputs using a Bayesian attractor network model, facilitating robust pose estimation. This framework employs two subnetworks to represent the robot's pose: the head direction network encodes angular information, while the grid cell network encodes positional information.

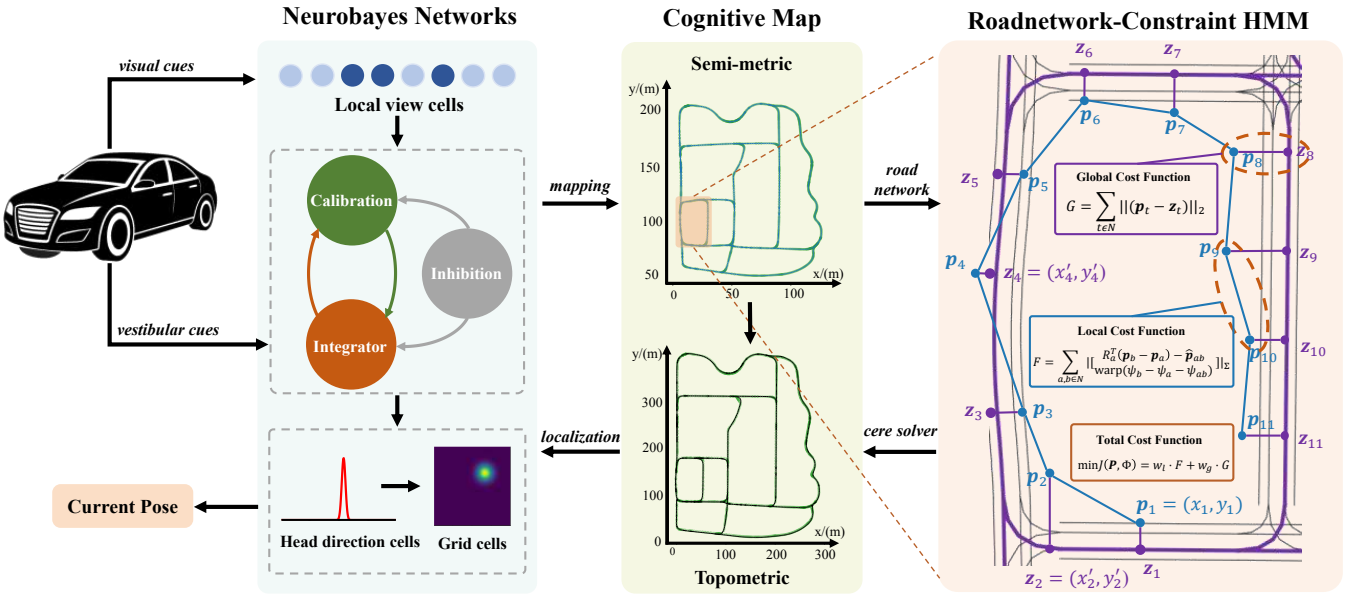

Fig. 2. **Framework for Topometric Cognitive Mapping.** Visual and vestibular cues are integrated to construct a semi-metric cognitive map. Local view cells encode visual features, while integration and calibration cells support the formation of head direction and grid cell models for orientation and spatial representation. RC-HMM incorporates road networks, refining the map with global and local constraints. The neurodynamics of grid and head direction cells enable visual localization within the topometric map.

The head direction cell network is modeled through a Bayesian inference process involving a population of integrator cells for vestibular cues integration, as well as a population of calibration cells for visual cue calibration. The two populations are interconnected through mutual inhibition and global inhibition. The belief of each population is defined as:

$$p(\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-|\theta-\mu|^2/2\sigma^2} \tag{1}$$

Where $\theta \in [0, 2\pi)$ is the label of a cell. $|\cdot|$ takes the difference of two angles on a circle. The mean μ is the maximum likelihood estimation of the head direction given by the population, and the variance σ^2 is the uncertainty of the belief. The reliability of the head direction cell derived from the combined contributions of integrator and calibration cells, which are uniform without prior knowledge [18].

Similar to the head direction cell network, the grid cells in the model form 2D torus attractors with single-peaked activity profiles.

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-|x-\mu_x|^2/2\sigma_x^2 + |y-\mu_y|^2/2\sigma_y^2} \tag{2}$$

Where $x, y \in [0, 2\pi)$ are the coordinates of the cells in the respective 2D neural manifold. (μ_x, μ_y) is the spatial phase encoded by the cells. $1/\sigma_x^2$ and $1/\sigma_y^2$ are the reliabilities of spatial phase estimation in each dimension.

2) Attractor dynamics: Taking the head direction cell network as an example, conflicts between integrator and calibration cells are resolved through global and mutual inhibition, enabling attractor dynamics.

Global inhibition maintains a constant network energy and is defined as:

$$\frac{1}{(\sigma_{\text{inte}}^t)^2} = \frac{E}{W} \frac{1}{(\sigma_{\text{inte}}^{t-1})^2} \tag{3}$$

$$\frac{1}{(\sigma_{\text{cali}}^t)^2} = \frac{E}{W} \frac{1}{(\sigma_{\text{cali}}^{t-1})^2} \tag{4}$$

where $(\sigma_{\text{inte}}^{t-1})^2$ and $(\sigma_{\text{cali}}^{t-1})^2$ represent the uncertainty in the beliefs of integrator and calibration cells before updating. W is the total Fisher information that acts as a normalization factor.

$$W = \frac{1}{(\sigma_{\text{inte}}^{t-1})^2} + \frac{1}{(\sigma_{\text{cali}}^{t-1})^2} \tag{5}$$

E is a predefined constant representing the total Fisher information for head direction. Mutual inhibition ensures a single stable activity peak over time, defined as:

$$\frac{1}{(\sigma_{\text{inte}}^t)^2} = \frac{1}{(\sigma_{\text{inte}}^{t-1})^2} - \Delta_{\text{inte}} \frac{1}{(\sigma_{\text{cali}}^{t-1})^2} \tag{6}$$

$$\frac{1}{(\sigma_{\text{cali}}^t)^2} = \frac{1}{(\sigma_{\text{cali}}^{t-1})^2} - \Delta_{\text{cali}} \frac{1}{(\sigma_{\text{inte}}^{t-1})^2} \tag{7}$$

Here, Δ_{inte} and Δ_{cali} denote the inhibition strength on integrator and calibration cells, respectively.

The competition between these two populations allows the network to temporarily accommodate different beliefs while gradually integrating information to form a coherent head direction representation.

3) Construction And Serialization: By modeling head direction and grid cell networks and utilizing a Bayesian attractor neural dynamics mechanism, we build a semi-metric cognitive map [18] to store spatial positions and transformations derived from visual information. To ensure efficient storage and facilitate accessibility, we employ Protocol Buffers (protobuf) [35], a structured data serialization framework. This approach enables us to store data as protobuf stream files (pbstream), thereby maintaining the compactness and

accessibility of the semi-metric cognitive map. The structure of this map is organized into head direction and grid cells, local view cells, and the experiences map.

B. RoadNetwork-Constraint Hidden Markov Model (RC-HMM)

Metric SLAM achieves high-precision mapping but is computationally intensive, while topological SLAM is more efficient but less accurate. To balance this trade-off, we propose the RC-HMM that incorporates road network information to refine semi-metric maps into topometric representations, enhancing accuracy while maintaining efficiency.

1) Deserialization of the Topological Structure: To extract the topological structure of the semi-metric cognitive map, we employ protobuf [35] deserialization to convert the experience map into a topological graph comprising vertex sequences and connectivity relationships, known as the semi-metric topological map.

2) Formulation of the Global Cost Function: To incorporate metric information, we apply the HMM map matching algorithm [34] to project the topological vertices onto the road network. The resulting matched positions serve as global constraints for constructing the global cost function in RC-HMM.

Emission Probability Modeling The vertex of the semi-metric topological map contains positional and orientation information and can be expressed as $(\boldsymbol{p_t}, \psi_t)$. Here, $\boldsymbol{p_t}$ represents the position of vertex as a two-dimensional position vector, while ψ_t denotes the orientation.

Consider $\boldsymbol{p}_t$ as a measurement, and define a candidate road segment area with a radius r around $\boldsymbol{p}_t$. Road segments within this area are considered potential matches, denoted as r_i. The emission probability $p(\boldsymbol{p}_t|r_i)$ represents the likelihood of observing $\boldsymbol{p}_t$ if the agent is located on road segment r_i. The emission probability is calculated using (8), where $\boldsymbol{z}_{t,i}$ represents the closest point on road segment r_i to the measurement $\boldsymbol{p}_t$, and $||\boldsymbol{p}_t - \boldsymbol{z}_{t,i}||$ is the shortest distance between them.

$$p(\boldsymbol{p}_t|r_i) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-0.5(\frac{||\boldsymbol{p}_t - \boldsymbol{z}_{t,i}||}{\sigma_z})^2} \tag{8}$$

Where $\sigma_z = 1.4826\text{median}_t(||\boldsymbol{p}_t - \boldsymbol{z}_{t,i^*}||)$ can be estimated by [34], i^* indicate the correct matched road.

Transition Probability Estimation Each measurement $\boldsymbol{p}_t$ has a set of potential road segment matches, and the next measurement $\boldsymbol{p}_{t+1}$ has its own candidates. Transition probabilities estimate the likelihood of the agent moving between these road segments. The driving distance between two points on candidate road segments is called the route distance, $||\boldsymbol{z}_{t,i} - \boldsymbol{z}_{t+1,j}||_{route}$, while the distance between measurements $\boldsymbol{p}_t$ and $\boldsymbol{p}_{t+1}$ is the measurement distance. According to [34], the absolute difference between these distances for correct matches follows an exponential probability distribution, as outlined in Equation (9):

$$p(d_t) = \frac{1}{\beta} e^{-d_t/\beta} \tag{9}$$

here, $d_t = |||\boldsymbol{p}_t - \boldsymbol{p}_{t+1}|| - ||\boldsymbol{z}_{t,i^*} - \boldsymbol{z}_{t+1,j^*}||_{route}|$, i^* and j^* indicate the ground truth road segments of route. We estimate the value of β with a robust estimator suggested by [36] as $\beta = 1/ln(2)\text{median}_t(d_t)$.

Optimal Path Computation Then we employ the Viterbi algorithm to determine the most likely path and the closest point to the measurement on that path.

The global cost function G, as defined in Equation (10), is designed to minimize the distance between the vertices on the semi-metric topological map and their corresponding matched points in the road network.

$$G = \sum_{t \in N} ||(\boldsymbol{p}_t - \boldsymbol{z}_t)||_2 \tag{10}$$

where N indicates the number of vertices on semi-metric topological map, $\boldsymbol{z}_t$ denotes the projection of $\boldsymbol{p}_t$ onto the correctly matched road.

3) Formulation of the Local Cost Function: The connections between topological map vertices serve as local constraints in constructing the local cost function. The local cost function, denoted as F, described in Equation (11), consists of two main components: positional and orientation differences.

$$F = \sum_{a,b \in N} ||\begin{bmatrix} R_a^T(\boldsymbol{p}_b - \boldsymbol{p}_a) - \hat{\boldsymbol{p}}_{ab} \\ \text{warp}(\psi_b - \psi_a - \hat{\psi}_{ab}) \end{bmatrix}||_\Sigma \tag{11}$$

The function $|\cdot|_\Sigma$ is the Mahalanobis distance with respect to covariance Σ, a measure that accounts for the correlation between distances of connected vertices. And the warp$(\cdot)$ function normalizes angle ranges, constraining direction values within $[-\pi, \pi]$. Set $\hat{\boldsymbol{p}}_{ab}$ and $\hat{\psi}_{ab}$ as a measure of the relative transformation between connected vertices a and b, capturing changes in both relative position and direction. R_a is used for coordinate transformation.

4) Optimization of the Cost Function: Combining global and local constraint, we define the total cost function as:

$$\min J(\boldsymbol{P}, \Phi) = w_l \cdot F + w_g \cdot G \tag{12}$$

where $\boldsymbol{P}$ and Φ represent the positions and orientations of vertices on topometric map, while w_l and w_g denote the weights assigned to local constraints and global constraints, respectively.

We formulated the least squares problem based on (12) and optimized it using the Ceres solver. Common approaches for solving such problems include gradient descent, the Gauss-Newton method, and the Levenberg-Marquardt algorithm. In this paper, we adopt the Gauss-Newton method with Cholesky decomposition to construct the topometric map, which integrates both topological structure and metric information. Following the same serialization process as the semi-metric cognitive map, we use protobuf [35] to store vertices and the relationships between vertices in the topometric map as a metric experience map, which, along with local view cells and grid cells, ultimately forms the topometric cognitive map.

C. Visual Localization with Neurodynamics

Visual localization is performed within the topometric cognitive map, utilizing visual images as input and leveraging the neural dynamics of head direction and grid cells.

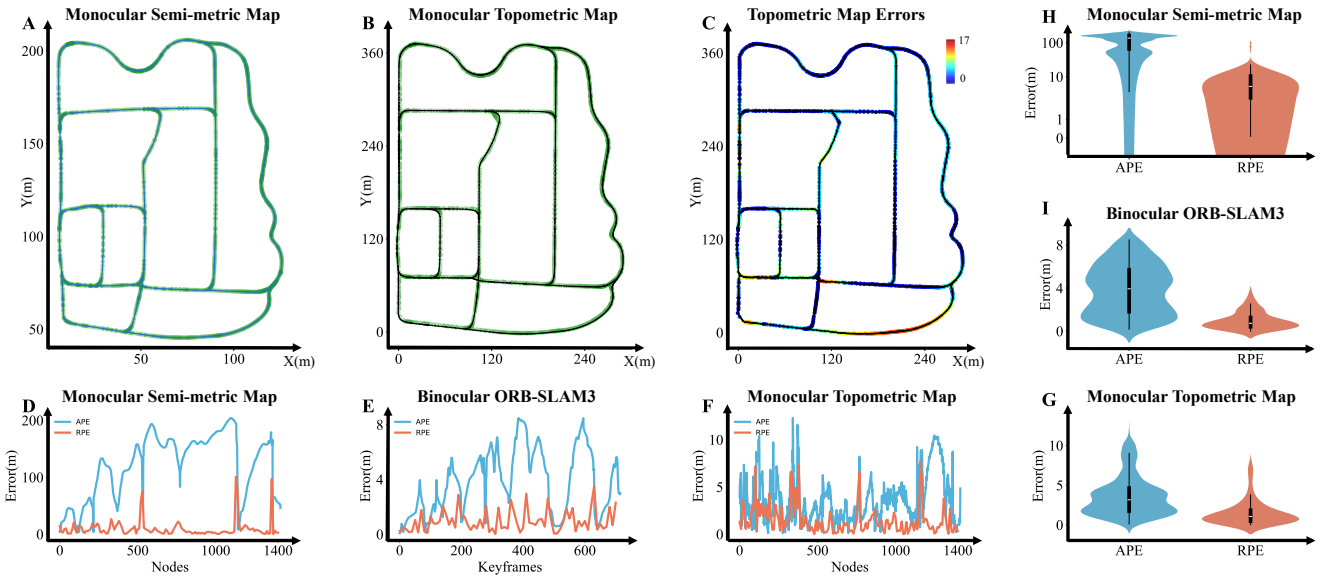

Fig. 3. **The topometric cognitive mapping experiments.** (A) semi-metric topological map of Town07. (B) Topometric map overlaid with ground truth for accuracy comparison. (C) Distribution of APE for the topometric map relative to ground truth. (D, E, F) Comparative analysis of APE and RPE for semi-metric map, binocular ORB-SLAM3 keyframes and topometric maps, highlighting a nearly tenfold difference. (H, I, G) Statistical analysis showing APE and RPE across three map types.

A new visual input is compared with the visual templates stored in the local view cells. If the input corresponds to a previously encountered scene, the associated local view cell is reactivated and injects corrective energy into the head direction and grid cell networks. Specifically, for the head direction network, the activity of calibration cells can be expressed as:

$$\frac{1}{(\sigma^t_{\text{cali}})^2} = \frac{1}{(\sigma^{t-1}_{\text{cali}})^2} + \frac{1}{(\sigma^t_{\text{inject}})^2} \tag{13}$$

$$\mu^t_{\text{cali}} = \frac{(\sigma^t_{\text{cali}})^2}{(\sigma^{t-1}_{\text{cali}})^2}\mu^{t-1}_{\text{cali}} + \frac{(\sigma^t_{\text{cali}})^2}{(\sigma^t_{\text{inject}})^2}\mu^t_{\text{inject}} \tag{14}$$

where $\frac{1}{(\sigma^t_{\text{inject}})^2}$ is the reliability of the visual cue, μ^t_{inject} is the location where the current is injected to the one dimensional neural manifold of head direction cells.

The current head direction is then estimated by integrating information from both the integrator and calibration cells:

$$\frac{1}{(\sigma^t_{\text{f}})^2} = \frac{1}{(\sigma^t_{\text{inte}})^2} + \frac{1}{(\sigma^t_{\text{cali}})^2} \tag{15}$$

$$\mu^t_{\text{f}} = \frac{(\sigma^t_{\text{f}})^2}{(\sigma^t_{\text{inte}})^2}\mu^t_{\text{inte}} + \frac{(\sigma^t_{\text{f}})^2}{(\sigma^t_{\text{cali}})^2}\mu^t_{\text{cali}} \tag{16}$$

where $\frac{1}{(\sigma^t_{\text{f}})^2}$ is the reliability and μ^t_{f} the center of the belief.

The integrator and calibration cells in grid cell network are updated in the same way as those in the head direction cell network to encode position of robot.

IV. EXPERIMENTAL SETUP

A. Simulation Platform and Datasets

CARLA, an open-source simulator widely used in autonomous driving research, provided the testing environment for this study. We conducted experiments in the Carla Town07 scenario, a rural setting with diverse road structures. We recorded the dataset in ROS bag format, capturing monocular and stereo image sequences over 1,315 s, totaling 14,717 frames. Monocular data was used to evaluate semi-metric cognitive map and our topometric cognitive map, while stereo data supported ORB-SLAM3 assessment.

To ensure a fair comparison, we selected binocular ORB-SLAM3 as the baseline over monocular ORB-SLAM3 with IMU. The latter relies on high-frame-rate IMUs and requires an initialization step involving acceleration and sensor tilting relative to gravity [27]. However, mobile robots often experience limited tilt variations, leading to frequent initialization failures. In contrast, binocular ORB-SLAM3 directly estimates scale from visual disparity, offering greater stability and precision.

All algorithms were executed on a unified platform running Ubuntu 20.04, powered by an Intel® Xeon® W-2195 CPU at 2.30 GHz. This ensured consistent computational performance across all experiments.

B. Mapping Evaluation

Performance Evaluation We conducted mapping experiments using monocular semi-metric cognitive map, our monocular topometric cognitive map and binocular ORB-SLAM3 on the same rosbag. To evaluate mapping performance, we measured CPU utilization, keyframe count, map storage size, vertex and edge numbers during execution.

Accuracy Evaluation For accuracy assessment, we utilized odometry-derived ground truth and employed Absolute Pose Error (APE) and Relative Pose Error (RPE) as evaluation metrics. APE quantifies the absolute deviation of each mapped vertex from its corresponding ground truth position, while

TABLE I
COMPARISON OF SEMI-METRIC MAP, BINOCULAR ORB-SLAM3 KEYFRAMES AND TOPOMETRIC MAP

	CPU Usage (%)	Map Size (M)	CameraFrame	KeyFrame	Vertex Nums	Edge Nums	APE (m)	RPE (m)
Semi-metric map (Mono)	1.3	56.1	-	-	1417	1567	121.2	9.4
ORB-SLAM3 (Binocular)	40.57	259	14717	715	-	-	3.86	**0.96**
Topometric map (Mono, ours)	**1.3**	**56.1**	-	-	1417	1567	**3.6**	1.4

RPE evaluates spatial consistency by analyzing the relative positional errors between consecutive vertices.

C. Localization Evaluation

Performance Evaluation Similarly, we conducted multiple localization experiments with our monocular topometric map and binocular ORB-SLAM3, recording initial localization time and trajectory length as performance indicators.

Accuracy Evaluation For localization accuracy evaluation, we again used odometry-derived ground truth. The localization error was defined as the deviation between the estimated position using grid cell representations and head-direction cell neurodynamics and the ground truth trajectory.

V. RESULTS

A. Topometric Cognitive Mapping

We conducted experiments in Town07, as shown in Fig. 3. Initially, we used visual and vestibular cues to construct a semi-metric map using NeuroBayesSLAM, as depicted in Fig. 3A. We then used RC-HMM to integrate road networks into the semi-metric map, forming the topometric map shown in Fig. 3B, with green vertices as map points and the black line as ground truth.

To evaluate mapping performance, we compared our monocular topometric cognitive map with a semi-metric map and binocular ORB-SLAM3 using the same rosbag. Key metrics, including CPU usage, map type, and storage size, are summarized in Table I. Due to an efficient data structure, our map uses 1,416 vertices and 1,567 edges, reducing storage by 77% and CPU usage by 96.7% compared to ORB-SLAM3's 14,717 trajectory points and 694 keyframes, enhancing computational and storage efficiency for mapping.

To assess performance, we evaluated the APE and RPE of three approaches: monocular semi-metric map, binocular ORB-SLAM3 keyframes, and our monocular topometric map, with results shown in Fig. 3D–F. The semi-metric map showed the largest errors (mean APE: 121 m; RPE: 9.6 m), highlighting its limitations. ORB-SLAM3 achieved greater precision (mean APE: 3.86 m; RPE: 0.96 m). Fig. 3C illustrates the APE distribution on our topometric map, which delivered the best overall performance, striking an optimal balance between precision (mean APE: 3.6 m; RPE: 1.4 m) and computational efficiency, such as reduced processing time compared to ORB-SLAM3.

Statistical analysis of APE and RPE across monocular semi-metric map, binocular ORB-SLAM3 keyframes, and monocular topometric map, as presented in Fig. 3G–I. These results underscore the monocular topometric map's advantage: it combines the high computational efficiency of semi-metric maps with the precision of metric maps, offering a versatile solution for spatial mapping. While Fig. 3C shows higher APEs resulting from map alignment and the sparsity of topometric map. In future work, we aim to develop an efficient navigation system that leverages topometric maps for global planning and orientation, better supporting large-scale navigation tasks.

B. Topometric Localization

To explain the neurodynamic process of visual localization, we refer to Fig. 4A through Fig. 4D. Fig. 4A displays the visual input over time, with shorter intervals in the initial phase and longer intervals after successful localization. Fig. 4B and C represent the dynamic activity of grid and head direction cells, corresponding to the visual input in Fig. 4A. The localization error throughout the process is visualized in Fig. 4D, where error magnitude is represented by color-coded vertices. The initial point at $T = 0$, shown in Fig. 4D, corresponds to the starting positions depicted in Fig. 4A–C and matches the initial map position, marked as a dark gray dot in Fig. 4E. As visual input is processed, grid and head direction cells stabilize through neurodynamic interactions, leading to a significant reduction in localization error and enabling successful localization. This trajectory is highlighted by a pentagram in Fig. 4E, where the color of the vertices represents the magnitude of the localization error, consistent with the color-coding scheme in Fig. 4D.

To evaluate the localization performance of our algorithm, we selected 10 image sequences of varying trajectory lengths and conducted comparative tests using both our monocular topometric map and binocular ORB-SLAM3. We record the trajectory lengths, initial localization time, and localization error for both algorithms, as summarized in Table II. Compared to ORB-SLAM3, our monocular topometric map significantly improves localization speed while maintaining comparable localization accuracy.

Then further testing was conducted on a 1,816 meter trajectory, highlighted by a quadrilateral star in Fig. 4E, where vertices colors reflect the magnitude of localization error of our monocular topometric map. The localization accuracy results for our monocular topometric map and binocular ORB-SLAM3 are presented in Fig. 4F and G, respectively. Both algorithms exhibit initial localization errors of approximately 120 m. Our monocular topometric map (Fig. 4F) achieves

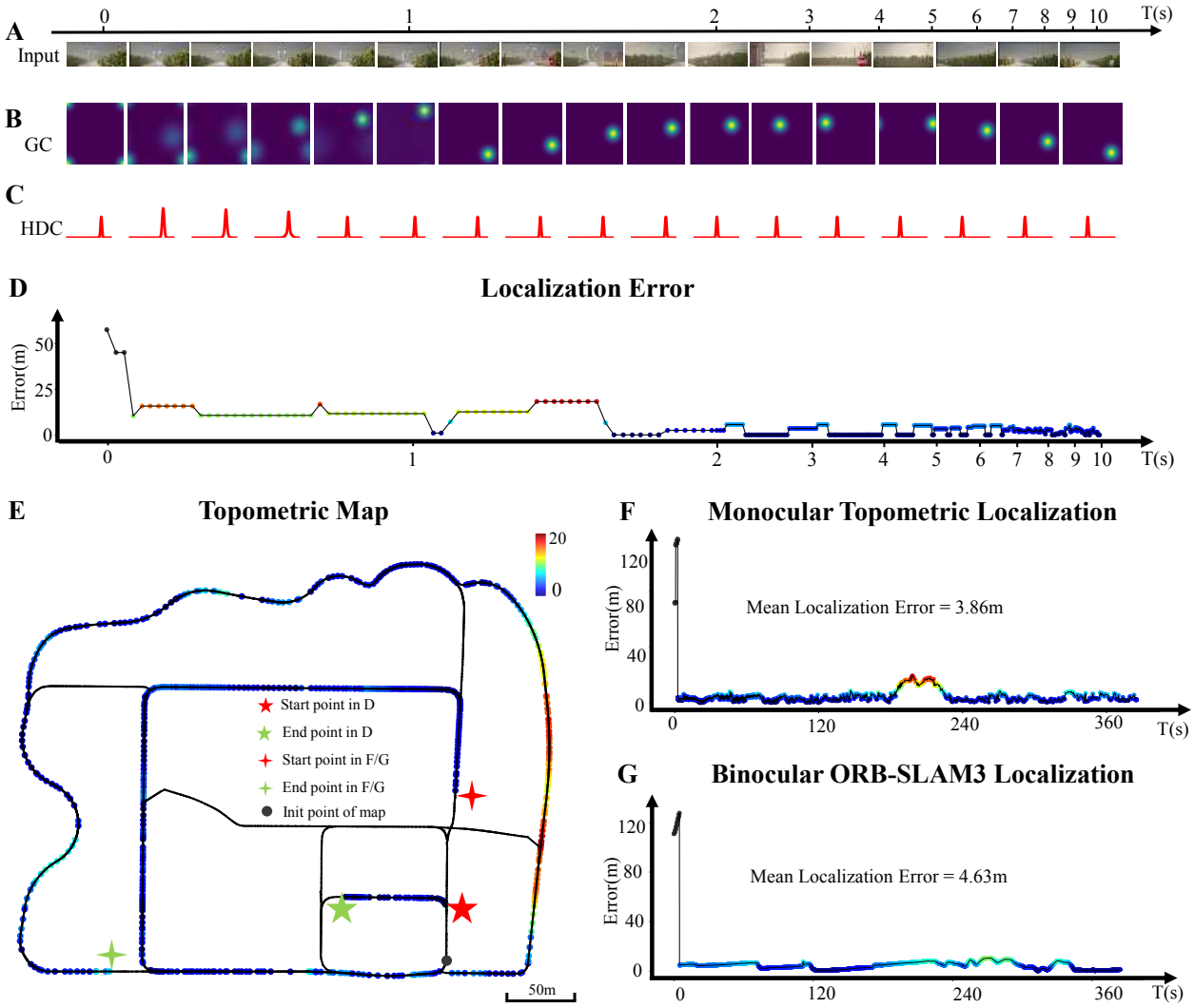

Fig. 4. **Visual Topometric Localization Experiments.** (A–D) Neurodynamics of localization with a logarithmic time scale. (A) Visual input; neurodynamics of (B) grid cells and (C) head direction cells firing patterns; (D) localization errors. The start and end points of the test segment in (D) are marked by pentagram, while dark gray points indicate the initial position. (F, G) The localization accuracy of both our monocular topometric map and binocular ORB-SLAM3, corresponding to the start and end points of the quadrilateral stars in the test trajectory shown in (D).

TABLE II
LOCALIZATION PERFORMANCE RESULTS

Method	Metric	Seq 1	Seq 2	Seq 3	Seq 4	Seq 5	Seq 6	Seq 7	Seq 8	Seq 9	Seq 10
ORB-SLAM3 (Binocular)	Trajectory Length (m)	372.7	804.3	170	732.6	812.7	1485.3	559	294.2	1423.9	830.9
	Initial Localization Time (ms)	794	1058	1077	978	1104	1171	1176	894	993	1064
	Localization Error(m)	**2.82**	4.77	**3.69**	4.03	4.85	**5.22**	6.29	4.86	5.29	4.5
Topometric map (Mono, ours)	Initial Localization Time (ms)	**1.54**	**0.13**	**0.65**	**0.15**	**2.02**	**0.28**	**0.12**	**0.25**	**0.21**	**0.18**
	Localization Error(m)	4.03	**3.74**	5.19	**3.59**	**3.56**	6.33	**3.65**	**4.8**	**4.63**	**3.89**

fast initial localization (0.56 ms) and stabilizes at an average error of 3.86 m. In comparison, binocular ORB-SLAM3 (Fig. 4G) localizes in 780 ms with an average error of 4.63 m. Throughout testing, our monocular topometric map maintains a lower average error while substantially reducing initial localization time compared to binocular ORB-SLAM3. However, the accuracy of RC-HMM is ultimately constrained by the sparsity of the topometric map, whereas ORB-SLAM3 benefits from the availability of abundant and high-quality feature points.

VI. CONCLUSIONS

In this paper, we introduced a brain-inspired visual topometric localization approach aimed at enhancing the practical ap-

plications of robotic navigation in complex environments. The proposed RC-HMM integrates semi-metric cognitive maps from NeuroBayesSLAM [18] with metric constraints using the HMM map matching [34], thus improving metric accuracy. Meanwhile, the incorporation of neurodynamic models inspired by mammalian brains ensures accurate localization.

Our method was evaluated in the CARLA Town07 virtual environment, demonstrating substantial improvements in computational efficiency. Our monocular topometric cognitive map significantly reduced CPU usage and map storage compared to binocular ORB-SLAM3 while maintaining comparable mapping accuracy. Furthermore, leveraging the neurodynamic mechanisms of grid cells and head direction cells, our monocular topometric cognitive map drastically reduced initial localization time without compromising precision. These results highlight the potential of neuro-inspired topometric mapping for achieving both efficiency and accuracy in robot navigation. Future work will focus on extending evaluations to large-scale real-world environments and exploring long-term operational scenarios to further validate the approach.

ACKNOWLEDGMENT

This work was supported by the National Science and Technology Innovation 2030 Major Program of China (Grant No. 2022ZD0205005).

REFERENCES

[1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[2] M. J. Milford, J. Wiles, and G. F. Wyeth, "Solving navigational uncertainty using grid cells on robots," *PLoS computational biology*, vol. 6, no. 11, p. e1000995, 2010.

[3] M. Geva-Sagiv, L. Las, Y. Yovel, and N. Ulanovsky, "Spatial cognition in bats and rats: from sensory acquisition to multiscale maps and navigation," *Nature Reviews Neuroscience*, vol. 16, no. 2, pp. 94–108, 2015.

[4] M. Parrini, G. Tricot, P. Caroni, and M. Spolidoro, "Circuit mechanisms of navigation strategy learning in mice," *Current Biology*, vol. 34, no. 1, pp. 79–91, 2024.

[5] E. A. Maguire, N. Burgess, J. G. Donnett, R. S. Frackowiak, C. D. Frith, and J. O'Keefe, "Knowing where and getting there: a human navigation network," *Science*, vol. 280, no. 5365, pp. 921–924, 1998.

[6] E. C. Tolman, "Cognitive maps in rats and men." *Psychological review*, vol. 55, no. 4, p. 189, 1948.

[7] K. J. Jeffery, M. I. Anderson, R. Hayman, and S. Chakraborty, "A proposed architecture for the neural representation of spatial context," *Neuroscience & Biobehavioral Reviews*, vol. 28, no. 2, pp. 201–218, 2004.

[8] J. Rank, "Head-direction cells in the deep layers of dorsal presubiculum of freely moving rats," in *Soc. Neuroscience Abstr.*, vol. 10, 1984, p. 599.

[9] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, "Microstructure of a spatial map in the entorhinal cortex," *Nature*, vol. 436, no. 7052, pp. 801–806, 2005.

[10] N. Burgess, "The 2014 nobel prize in physiology or medicine: a spatial model for cognitive neuroscience," *Neuron*, vol. 84, no. 6, pp. 1120–1125, 2014.

[11] E. I. Moser, E. Kropff, and M.-B. Moser, "Place cells, grid cells, and the brain's spatial representation system," *Annu. Rev. Neurosci.*, vol. 31, no. 1, pp. 69–89, 2008.

[12] A. Mathis, A. V. Herz, and M. Stemmler, "Optimal population codes for space: grid cells outperform place cells," *Neural computation*, vol. 24, no. 9, pp. 2280–2317, 2012.

[13] A. Arleo, "Spatial learning and navigation in neuro-mimetic systems: modeling the rat hippocampus," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, Dec. 2000.

[14] J. L. Krichmar, D. A. Nitz, J. A. Gally, and G. M. Edelman, "Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task," *Proceedings of the National Academy of Sciences*, vol. 102, no. 6, pp. 2111–2116, 2005.

[15] N. Burgess, J. G. Donnett, K. J. Jeffery, and J. O-keefe, "Robotic and neuronal simulation of the hippocampus and rat navigation," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 352, no. 1360, pp. 1535–1543, 1997.

[16] M. J. Milford, G. F. Wyeth, and D. Prasser, "Ratslam: a hippocampal model for simultaneous localization and mapping," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, vol. 1. IEEE, 2004, pp. 403–408.

[17] D. Ball, S. Heath, J. Wiles, G. Wyeth, P. Corke, and M. Milford, "Openratslam: an open source brain-based slam system," *Autonomous Robots*, vol. 34, pp. 149–176, 2013.

[18] T. Zeng, F. Tang, D. Ji, and B. Si, "Neurobayesslam: Neurobiologically inspired bayesian integration of multisensory information for robot navigation," *Neural Networks*, vol. 126, pp. 21–35, 2020.

[19] F. Yu, J. Shang, Y. Hu, and M. Milford, "Neuroslam: A brain-inspired slam system for 3d environments," *Biological cybernetics*, vol. 113, no. 5, pp. 515–545, 2019.

[20] 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.

[21] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2d lidar slam," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 1271–1278.

[22] J. Zhang, S. Singh *et al.*, "Loam: Lidar odometry and mapping in real-time." in *Robotics: Science and systems*, vol. 2, no. 9. Berkeley, CA, 2014, pp. 1–9.

[23] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2020, pp. 5135–5142.

[24] T. Shan and B. Englot, "Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4758–4765.

[25] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[26] 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.

[27] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[28] C. Bamann and P. Henkel, "Visual-inertial odometry with sparse map constraints for planetary swarm exploration," in *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*. IEEE, 2019, pp. 290–295.

[29] T. Zeng and B. Si, "A brain-inspired compact cognitive mapping system," *Cognitive Neurodynamics*, vol. 15, no. 1, pp. 91–101, 2021.

[30] H. Badino, D. Huber, and T. Kanade, "Real-time topometric localization," in *2012 IEEE International conference on robotics and automation*. IEEE, 2012, pp. 1635–1642.

[31] F. Blochliger, M. Fehr, M. Dymczyk, T. Schneider, and R. Siegwart, "Topomap: Topological mapping and navigation based on visual slam maps," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 3818–3825.

[32] C. Gomez, M. Fehr, A. Millane, A. C. Hernandez, J. Nieto, R. Barber, and R. Siegwart, "Hybrid topological and 3d dense mapping through autonomous exploration for large indoor environments," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9673–9679.

[33] H. Badino, D. Huber, and T. Kanade, "Visual topometric localization," in *2011 IEEE Intelligent vehicles symposium (IV)*. IEEE, 2011, pp. 794–799.

[34] P. Newson and J. Krumm, "Hidden markov map matching through noise and sparseness," in *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, 2009, pp. 336–343.

[35] Google, "Protocol buffers," 2023. [Online]. Available: https://developers.google.com/protocol-buffers

[36] U. Gather and V. Schultze, "Robust estimation of scale of an exponential distribution," *Statistica Neerlandica*, vol. 53, no. 3, pp. 327–341, 1999.