Environmental Mapping for Rat Robots Using Hierarchical ORB-SLAM

Yuhong Huang[1], Han Li[2], Boyang Li[2], Zitao Zhang[1], Zhenshan Bing[3], Kai Huang[2], and Alois Knoll[1]

***Abstract*— Simultaneous Localization and Mapping (SLAM) on small quadruped robots, such as rat robots, is challenged by body vibrations, limited onboard computation, and low-texture environments. These factors cause sparse features, unstable viewpoints, and frame loss, which hinder conventional SLAM. To address these challenges, we redesign the rat robot with an external antenna to improve communication and a monocular camera to reduce bandwidth, while running monocular ORB-SLAM on a remote computer. In addition, we introduce a hierarchical SLAM framework that switches between coarse frame matching under sparse features and fine keyframe matching when features are sufficient. Experiments in a low-texture drainage channel show that our system maintains continuous mapping despite frequent interruptions, demonstrating feasibility for real-world deployment of rat-sized quadruped robots. A demo of the experiments can be found at https://youtu.be/2vSgOohzkUA.**

I. Introduction

Quadruped robots have gained attention for their adaptability in unstructured environments [1]. To repeatedly perform diverse tasks such as field inspection, they must accurately perceive and understand their surroundings. Integrating Simultaneous Localization and Mapping (SLAM) offers a promising solution, enabling these robots to build environmental awareness and operate effectively across different scenarios [2].

SLAM integrates data from LiDAR, cameras, and inertial systems to help autonomous vehicles map unknown environments and localize accurately during revisits [3], [4]. However, quadruped robots face challenges distinct from wheeled vehicles, which operate on flat surfaces. To traverse rough terrain, they rely on dynamic foothold adjustments that introduce body pose variations and reduce stability, particularly during fast trotting or bounding gaits. Such instability generates sensor disturbances, including vibrations that impair visual and LiDAR front-ends and rapid drift in IMU biases [5]. The problem is more severe in small quadruped robots due to limited stability and hardware constraints [6]. For example, rat robot often use low-torque motors that struggle to maintain balance, while their compact design restricts power supply and sensor payload [7]. Consequently, conventional SLAM sensors like LiDAR or depth cameras are difficult to integrate. To address this, Shi et al. equipped a ratquadruped with a miniature binocular camera for visual SLAM (V-SLAM) [8]. They further reduced sensor disturbances by using a walking gait with three legs in ground contact, enabling stable though slow navigation.

Most SLAM methods assume stable robot motion and reliable sensing. As a result, they are typically designed for wheeled robots with planar motion or quadrupeds using highly stable gaits [9], [10], while overlooking the unique challenges of small quadruped robots. These challenges include significant body oscillations, limited onboard computation, and operation in low-texture environments. First, low-texture scenes provide sparse visual features within a frame, reducing inter-frame correspondences and hindering reliable feature matching (Fig. 1(a)). Second, body oscillations during locomotion introduce divergence in feature extractionFig. 1(b)). For example, in fast gaits such as trotting, only two legs contact the ground at once, producing large body motions, unstable viewpoints, and perceptual errors such as misidentifying feature points or mistaking static textures for dynamic ones. Third, limited onboard computation restricts image capture frequency and may cause signal loss (Fig. 1(c)). Rat robot cannot carry high-performance processors, requiring computation to be uploaded to remote devices. This uploading is further constrained by camera sampling rates, communication bandwidth, and latency. In addition, the robot's small antenna weakens wireless transmission, leading to intermittent frame drops that break continuous frame matching in SLAM.

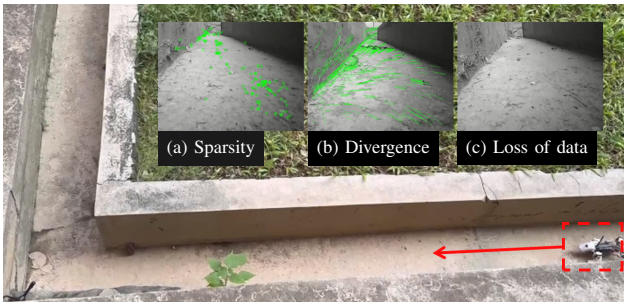


Fig. 1. Challenges of ORB-SLAM in the rat robot. (a) illustrates the sparsity of image features captured in low-texture environments, (b) demonstrates the potential divergence of feature points caused by motion-induced instability, and (c) depicts frame loss resulting from the limitations of the rat robot's onboard system.

To address these challenges, we propose a hierarchical SLAM framework that enables rat robots to perform mapping at both coarse and fine levels of feature points. Our main contributions are as follows:

- To mitigate unstable image sampling caused by onboard limitations, we improved the robot design from our prior work [11]. The updated system uses an external antenna to enhance transmission and a monocular camera to reduce bandwidth, with monocular ORB-SLAM executed on a remote computer.
- To handle interruptions from insufficient feature correspondences (Fig. 1), we introduce a hierarchical framework that applies coarse-grained frame matching when

[1] Authors from the Technical University of Munich, Munich, German yuhong.huang@tum.de
[2] Authors from the Sun Yat-Sen University, Guangdong China
[3] Authors from the Nanjing University, Nanjing, China
Corresponding author: Zhenshan Bing

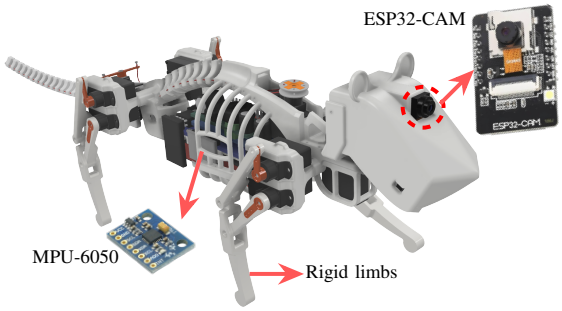


Fig. 2. Main components of the rat robot for ORB-SLAM. This robot employs ESP32-CAM as the central control unit, integrates IMU data from MPU6050, and utilizes rigid-leg locomotion to support ORB-SLAM-based mapping.

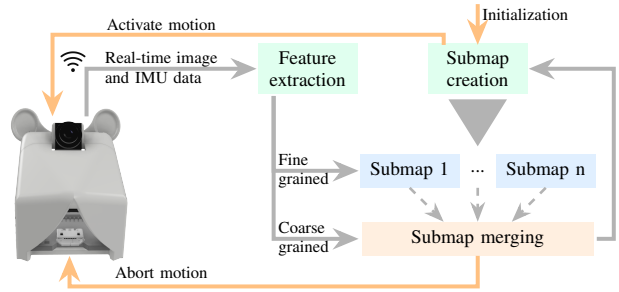


Fig. 3. Hierarchical framework of ORB-SLAM in rat robot applications. The figure illustrates the operational logic and data flow. The locomotion states are triggered or suspended according to mapping process. The density of extracted feature points determines which mapping layer is activated.

features are sparse and fine-grained keyframe matching when features are sufficient.

- We evaluated our approach in a real-world, low-texture drainage channel. Results show that the hierarchical ORB-SLAM framework supports reliable mapping for rat robots in outdoor environments, maintaining continuous performance even under frequent SLAM interruptions.

II. Foundation of ORB-SALM

ORB-SLAM relies on robust feature extraction and matching as the foundation for map reconstruction [12]. In each incoming frame, ORB features are detected and described, forming a set $\mathcal{F}_t = \{(\mathbf{u}_i^t, \mathbf{d}_i^t)\}$. To establish temporal correspondences, descriptors are compared across frames by minimizing the Hamming distance,

$$m_{ij} = \arg\min_j \|\mathbf{d}_i^t \oplus \mathbf{d}_j^{t-1}\|_H, \tag{1}$$

which yields matched pairs of image points. With sufficient correspondences between multiple views, triangulation is performed to generate new 3D landmarks. These landmarks, together with previously established ones, are incrementally refined through bundle adjustment,

$$\{\mathbf{X}\}^* = \arg\min_{\mathbf{X}} \sum_{i,j} \left\|\mathbf{u}_{ij} - \pi(\mathbf{T}_j \mathbf{X}_i)\right\|^2, \tag{2}$$

where $\pi(\cdot)$ denotes the projection model. This iterative cycle of feature matching, triangulation, and optimization allows ORB-SLAM to construct a consistent and scalable representation of the environment while supporting localization and loop closure.

III. Implementation of ORB-SLAM on Rat Robot

This section first describes the components of our rat robot, followed by an explanation of the proposed hierarchical ORB-SLAM architecture and its mapping process.

A. Components of the Rat Robot

Inspired by rat locomotion, we previously developed the NeRmo robot with a bioinspired skeletal design [11]. However, severe size constraints prevent onboard execution of computationally intensive SLAM, requiring real-time image transmission to a remote computer. This approach is limited by the small antenna, which cannot support the bandwidth needed for real-time SLAM. To overcome these limitations, we reconstructed the control system using an ESP32-CAM, a compact image acquisition board with an external antenna, as the central processor. To reduce bandwidth demands, the platform employs a monocular camera, with visual data fused with IMU measurements for SLAM. In addition, the flexible legs of the original NeRmo design were replaced with rigid link structures to reduce body oscillations and improve stability. As illustrated in Fig. 2, the redesigned robot supports real-time monocular SLAM while executing a trotting gait.

B. Structure of the Hierarchical ORB-SLAM for Rat Robot

In outdoor environments, the rat robot encounters multiple challenges (Fig. 1) that prevent the remote computer from extracting sufficient feature points to form keyframes for ORB-SLAM, causing ORB-SLAM to stall even as images are continuously captured. Typically, ORB-SLAM resumes mapping by matching newly generated keyframes with historical ones after a stall. However, locomotion instability, the restricted field of view resulting from the robot's low height and small camera, and the limited resolution of the onboard sensor all reduce the similarity between new and historical keyframes. Consequently, keyframe matching frequently fails, causing ORB-SLAM to break and interrupting continuous mapping. To address this limitation, we propose a hierarchical mapping framework capable of recovering from SLAM failures by merging historical data and supporting system reinitialization, as illustrated in Fig. 3.

As illustrated in Fig. 3, the rat robot communicates with a remote computer via Wi-Fi through an external gain antenna. Firstly, a submap construction process is created simultaneously with the activation of the robot's motion. During trotting, the robot continuously collects image and IMU data, which are transmitted to the remote computer for processing. The computer extracts feature points from the images and evaluates their density. If sufficient features are available to form a keyframe, the SLAM system operates in a fine-grained state; otherwise, it enters a coarse-grained state, signaling a potential SLAM stall. In the fine-grained state, the framework follows the standard SLAM pipeline, generating keyframes and matching them with historical ones within

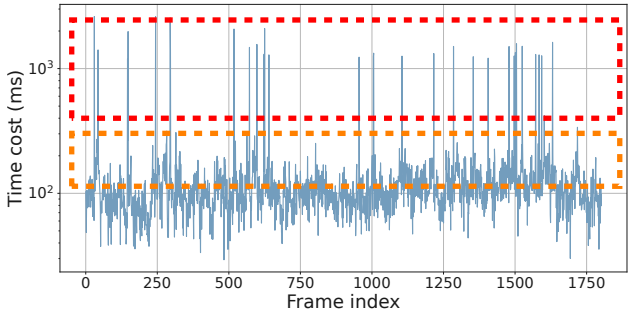


Fig. 4. Time cost of image collection on the remote computer. The red box marks high latency, while the orange box shows fluctuating latency that disrupts image acquisition and mapping.

the current submap to maintain continuous mapping. In the coarse-grained state, submap construction is suspended, and the most recent submap is merged with previous ones to preserve global consistency. This merging process leverages kinematic information and submap keyframe coordinates rather than feature correspondences. Because locomotion introduces pose variations and coordinate shifts, the robot temporarily halts during merging. Once merging is complete, both submap generation and locomotion are reinitialized.

C. Hierarchical Mapping Inspection and Submap Merging

One of the core components of the framework illustrated in Fig. 3 is the determination of the SLAM mapping state based on the density of feature points. At time t, let n_t denote the density of detected feature points. If $n_t \geq \eta$, the system is considered to be in a fine-grained feature state, where keyframe can be constructed to support map building. Accordingly, the fine-grained ORB-SLAM mapping process can be expressed as

$$\mathcal{M}_k^t = \mathcal{M}_k^{t-1} \bigcup K_t, \quad K_t = \begin{cases} (\mathbf{T}_t, I_t), & \text{if } n_t \geq \eta, \\ \emptyset, & \text{otherwise.} \end{cases} \tag{3}$$

In (3), $\mathcal{M}_k^t$ represents the mapping result of the k-th submap at time t. The keyframe K_t consists of the robot's pose $\mathbf{T}_t$, obtained from kinematics, and the image I_t, captured by the onboard camera. Thus, $\mathcal{M}_k^t$ is obtained by merging $\mathcal{M}_k^{t-1}$ with the feature correspondences from K_t, expressed using a union operator. If the conditions shown in Fig. 1 occur, too few feature points are extracted to form a keyframe, causing $K_t = \emptyset$ and $\mathcal{M}_k^t = \mathcal{M}_k^{t-1}$. In this context, to detect when SLAM is approaching a stall, we introduce a counter defined as

$$C_t = \sum_{\tau=t-\lambda+1}^{t} [\, \mathcal{M}_k^{\tau} = \mathcal{M}_k^{\tau-1} \,]. \tag{4}$$

Here, λ denotes the observation window length, and $[\, \mathcal{M}_k^{\tau} = \mathcal{M}_k^{\tau-1} \,]$ indicates whether the submap has changed. If C_t falls below a predefined threshold, SLAM is regarded as stalled and enters the coarse-grained state. In this state, submap construction is suspended, and the k-th submap $\mathcal{M}_k$ is merged into the global map $\mathcal{M}^{\mathcal{W}}$ as

$$\mathcal{M}^{\mathcal{W}} = \mathcal{M}^{\mathcal{W}} \bigcup \mathcal{M}_k^{\mathcal{W}}, \quad \mathcal{M}_k^{\mathcal{W}} = \mathcal{M}_k \cdot \Delta\mathbf{T}_{k \to k-1}, \tag{5}$$

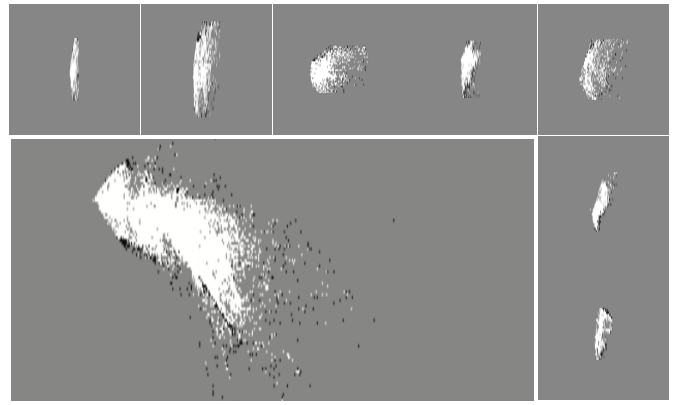


Fig. 5. Global map constructed from submap merging at different stages. The smaller images illustrate the submap generated before each SLAM stalls, while the larger image in the lower left corner shows the final global map obtained through submap merging.

where $\mathcal{M}_k \cdot \Delta\mathbf{T}_{k \to k-1}$ denotes mapping the submap $\mathcal{M}_k$ into the world coordinate system, and $\Delta\mathbf{T}_{k-1 \to k}$ represents the relative pose change from submap $k-1$ to k, which can be computed from the recorded poses $\mathbf{T}_t$ in (3). Once the merging of $\mathcal{M}_k$ is completed, ORB-SLAM and the robot's locomotion are reinitialized to ensure continuous mapping.

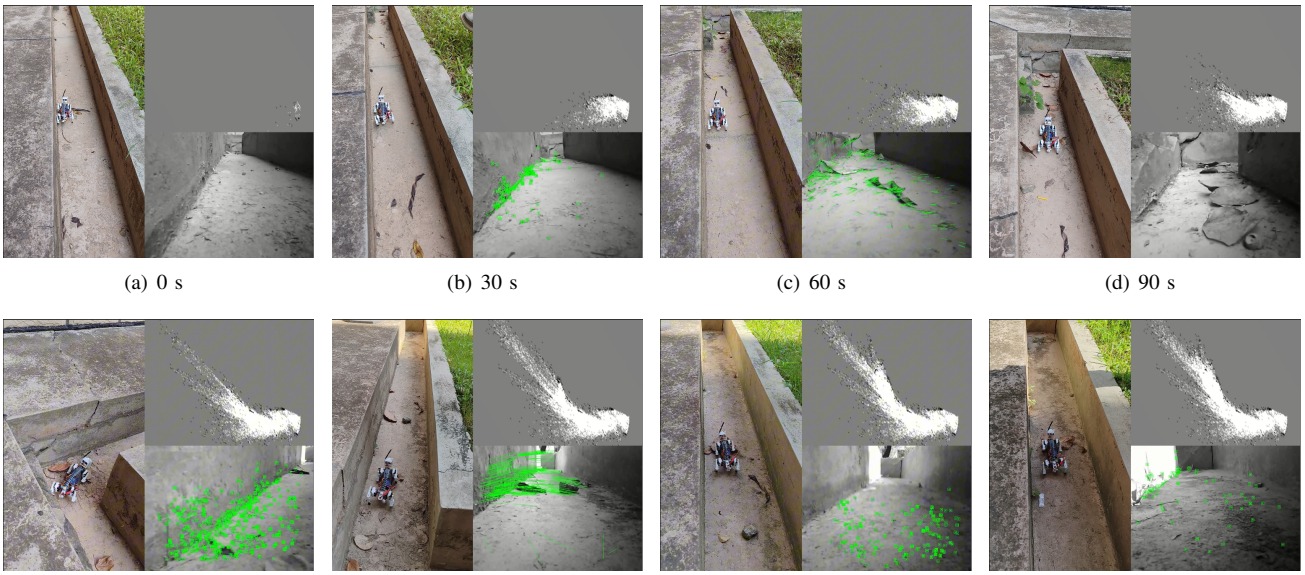
IV. Real-world Experiments

In this section, we deploy the rat robot in the drainage channel illustrated in Fig. 1 to conduct mapping experiments. In this scenario, the robot is required to navigate a 25 cm-wide corridor containing a right-angle turn. We then analyze the interruptions encountered during the mapping process as well as the corresponding results of continuous mapping.

A. Constraints in the Mapping Process

In real-world deployments of ORB-SLAM on the rat robot, image acquisition on the remote computer is constrained not only by the camera frame rate but also by long-range wireless transmission. Fig. 4 shows the time cost for each frame when the robot navigates the corridor in Fig. 1. The red dashed box highlights high-latency periods, during which the remote computer struggles to receive images in real time, reflecting the frame loss problem described in Fig. 1(c). The orange dashed box indicates fluctuating latency, where transmission drops below 10 Hz. This delay may introduce image blur and further aggravate sparse or divergent feature extraction. These hardware-induced transmission constraints, combined with locomotion instability and low-texture environments, frequently cause ORB-SLAM interruptions during real-world operation.

Fig. 5 presents the mapping results of the robot navigating through the initial straight section of the corridor. Each submap in Fig. 5 corresponds to one interruption in the mapping process. Using conventional ORB-SLAM, which relies on feature-based keyframe matching, the robot often builds only the first submap before crashing, as new keyframes fail to match historical ones. In contrast, our proposed framework explicitly accounts for such failures by restarting mapping independently after each crash. Although this approach sacrifices some accuracy by not enforcing matches with prior



(a) 0 s (b) 30 s (c) 60 s (d) 90 s

(e) 120 s (f) 150 s (g) 180 s (h) 210 s

Fig. 6. Montage of hierarchical ORB-SLAM mapping performed by the rat robot in a right-angled corridor. Each subfigure includes the locomotion following view (left), the mapping results over time (top right), and the real-time camera view (bottom right).

keyframes, it produces a set of submaps aligned to the global coordinate system with pose information. These submaps are then merged based on pose constraints to form the global map. The larger image in the lower left of Fig. 5 shows the mapping result of the straight corridor: despite multiple SLAM restarts, our framework successfully reconstructs the approximate corridor structure through submap stitching.

B. Mapping under Sparse Feature Constraints

In Fig. 6, the rat robot navigates a right-angled drainage channel (Fig. 1) to construct its two-dimensional outline. The task is challenged by low-texture scenes, limited remote sampling frequency, and unstable locomotion causing large viewpoint shifts. As shown in Fig. 6, the mapping process encounters multiple difficulties: sparse feature extraction at 30 s and 210 s, divergent features at 60 s and 150 s, and severe communication delays at 90 s that prevent timely image updates. Fine-grained mapping is achieved only at 120 s and 180 s, when sufficient features are successfully extracted. Although these challenges repeatedly force the system into coarse-grained states and trigger several SLAM restarts, the proposed framework still enables the rat robot to produce a coherent global outline of the corridor, demonstrating its feasibility under real-world constraints.

V. Conclusions

We present a hierarchical ORB-SLAM framework for rat-sized quadruped robots that addresses instability of locomotion, low-texture scenes, and communication limits. By switching between fine-grained and coarse-grained mapping and merging submaps using pose information, the system maintains mapping continuity under frequent interruptions. Experiments in drainage channels show that, despite sparse features, viewpoint shifts, and delays, the robot can construct a coherent global outline, confirming the feasibility of reliable mapping in constrained environments.

References

[1] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science Robotics*, vol. 5, no. 47, p. eabc5986, Oct. 2020.

[2] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part i," *IEEE robotics & automation magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[3] 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, 2017.

[4] 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.

[5] C. Zhang, Z. Yang, Q. Fang, C. Xu, H. Xu, X. Xu, and J. Zhang, "Frl-slam: A fast, robust and lightweight slam system for quadruped robot navigation," in *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2021, pp. 1165–1170.

[6] C. Li, Y. Zhang, Z. Yu, X. Liu, and Q. Shi, "A robust visual slam system for small-scale quadruped robots in dynamic environments," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, pp. 321–326.

[7] Z. Zhang, Y. Huang, Z. Zhao, Z. Bing, A. Knoll, and K. Huang, "A hierarchical reinforcement learning approach for adaptive quadruped locomotion of a rat robot," in *2023 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2023, pp. 1–6.

[8] Y. Zhang, Z. Chen, C. Li, Z. Yu, S. Li, and Q. Shi, "Depth-predictable vslam for a small-scale robotic rat in dynamic environments," *Journal of Field Robotics*, 2024.

[9] B. Al-Tawil, T. Hempel, A. Abdelrahman, and A. Al-Hamadi, "A review of visual slam for robotics: evolution, properties, and future applications," *Frontiers in Robotics and AI*, vol. 11, 2024.

[10] P. Arm, G. Waibel, J. Preisig, T. Tuna, R. Zhou, V. Bickel, G. Ligeza, T. Miki, F. Kehl, H. Kolvenbach, and M. Hutter, "Scientific exploration of challenging planetary analog environments with a team of legged robots," *Science Robotics*, vol. 8, no. 80, p. eade9548, 2023.

[11] Z. Bing, A. Rohregger, F. Walter, Y. Huang, P. Lucas, F. O. Morin, K. Huang, and A. Knoll, "Lateral flexion of a compliant spine improves motor performance in a bioinspired mouse robot," *Science Robotics*, vol. 8, no. 85, p. eadg7165, 2023.

[12] 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.