Multi-sensor fusion, motor control and localization for biomimetic robotic dogs

Jinlun Guo[1*]

[1]RDFZ XISHAN SCHOOL, Beijing, 100193, China

Abstract. Aiming at the localization and motor control challenges of biomimetic robotic dogs in complex environments, this study proposes a technical scheme integrating multi-sensor fusion, localization systems, and motor control. A visual-inertial-leg (VIO-LIO) multi-sensor fusion localization system is established: dynamic local feature extraction optimizes the visual module, while leg-inertial odometry is built using motor encoders and IMU, and Unscented Kalman Filter (UKF) fusion reduces localization error by over 20% and data processing time by 15.8%. Additionally, fused localization data is integrated with an improved Model Predictive Control (MPC) algorithm; a Convolutional Neural Network (CNN) predicts gait data to adjust motor joint parameters in real time through feedback, enhancing motor control stability. Experiments verify that in scenarios like slippery surfaces and dim environments, the system achieves a localization RMSE as low as 0.0352m, and motor-driven gait stability is 12% higher than traditional methods, providing reliable support for biomimetic robotic dogs' complex-scene applications.

1 Introduction

As biomimetic robotic dogs gain increasing attention for their potential in elderly care and small-scale complex environment tasks, two core technical challenges have become bottlenecks for their practical application: unreliable localization in dynamic scenarios and insufficient motor control adaptability. In real-world settings—such as dimly lit rooms for elderly care, slippery surfaces, or narrow passages—traditional localization methods often fall short: vision-based systems struggle with feature mismatch and high latency when lighting is poor or textures are scarce, while single inertial or leg odometry accumulates errors due to mechanical vibrations or ground slippage [1]. Meanwhile, conventional motor control relies on fixed mathematical models, failing to adjust gait parameters in real time to match dynamic terrain changes or human interaction needs, which undermines the robots' stability and safety [2]. Existing research highlighted in relevant studies has explored targeted solutions to address these gaps. For localization enhancement, integrating multi-source sensor data—including visual sensors, inertial measurement units (IMUs), and motor encoders—has been proven effective in compensating for the limitations of single-sensor systems [3]. By fusing visual-inertial-leg (VIO-LIO) data with adaptive feature extraction

* Corresponding author: gjldavid@163.com

and filtering algorithms, researchers have reduced localization errors by over 20% and shortened data processing time, enabling more robust environmental perception even in challenging conditions [4]. For motor control and gait optimization, combining deep learning with model predictive control (MPC) has shown promise: convolutional neural networks (CNNs) can learn and predict complex gait patterns from motion data, while MPC leverages real-time feedback to optimize motor joint parameters, enhancing the robots' response speed and stability when navigating variable terrains.

Notably, the urgency of advancing these technologies is further underscored by the limitations of existing robotic solutions. Wheeled robots, though widely used, have poor passability in uneven or cluttered environments—an issue that biomimetic robotic dogs, with their flexible leg structures, are uniquely positioned to solve. Building on the multi-sensor fusion and algorithm optimization foundations outlined in prior research, this study focuses on developing an integrated framework that unifies VIO-LIO fusion localization with CNN-enhanced MPC motor control. The goal is to address the dual challenges of localization robustness and motor adaptability simultaneously, thereby improving the passability and practicality of biomimetic robotic dogs in elderly care and small-scale complex environment tasks.

2 Technical requirements and core challenges of biomimetic robotic dogs

2.1 Scenario - adaptable technical requirements

Biomimetic robotic dogs are engineered to function in two pivotal application scenarios, each imposing unique technical imperatives. In elderly care contexts, as explored in Research on Gait Planning for Elderly Biomimetic Robot Dogs (jixiegou.pdf), these robotic dogs must sustain stable gait control while trailing elderly users. This entails adapting to the slow, fluctuating walking speeds of the elderly and minor terrain irregularities to forestall safety hazards. Real - time responsiveness is of utmost importance; the robot must dynamically tweak movement parameters such as stride length and step frequency to synchronize with the elderly individual's walking rhythm, thus averting collisions or instability during interaction.[2]

In small - scale complex scenarios, as detailed in *A Robust Multi - Sensor Fusion Localization for Small - Scale Biomimetic Robots* (Multi - Sensor.pdf), the emphasis shifts to dependable localization and environmental adaptability. These scenarios encompass narrow passageways, slippery surfaces, and dimly lit areas. In such settings, the proximity to obstacles degrades visual data, and ground slippage or low light disrupts sensor accuracy. The robot must navigate constricted spaces and cope with unforeseen conditions. This necessitates a localization system that can offset single - sensor failures and a motor control system that can adapt to unstable terrain [1].

2.2 core technical bottlenecks

2.2.1 Limitations of localization systems

As analyzed in Multi - Sensor.pdf, single - sensor localization approaches have inherent drawbacks. Vision - based systems, like stereo cameras, encounter issues with feature point mismatch or sparsity in dim lighting or textureless regions, resulting in localization drift [3]. Inertial Measurement Units (IMUs) alone accumulate errors over time due to mechanical

vibrations from leg movements, particularly in long - duration tasks [4]. Leg odometry, computed via motor encoders, becomes inaccurate during slippage. The joint angles measured by encoders no longer correspond to the actual foot - end displacement, thereby undermining localization reliability [1].

2.2.2 Shortcomings of motor control and gait planning

Traditional Model Predictive Control (MPC) algorithms, as discussed in jixiegou.pdf and commonly used in gait control, demand complex dynamic models and substantial computational resources. This leads to slow response times, typically 10–15 MS per control cycle, which are insufficient for the real - time requirements of elderly care scenarios [2]. Moreover, conventional motor control fails to integrate localization data, causing mismatches between gait parameters (such as joint angles) and the robot's actual position. For instance, incorrect step adjustments may occur when localization drifts on uneven ground [5].

3 Design of multi - sensor fusion localization system

3.1 Sensor architecture and data preprocessing

3.1.1 Multi - sensor selection and combination

Drawing on the fusion logic in Multi - Sensor.pdf, a "vision - inertial - leg" architecture is adopted to capitalize on the complementary nature of different sensors. Stereo cameras with a resolution of 1280×720 and a frame rate of 30 fps are employed to capture environmental visual features. An IMU with a 100 Hz sampling rate and an angular velocity range of ±2000°/s provides real - time motion data (acceleration and angular velocity) to bridge short - term visual gaps. Motor encoders with 0.1° accuracy, mounted on leg joints, record angles. These angles are used to calculate foot - end displacement through kinematic modeling, forming Leg - Inertial Odometry (LIO) [1].

3.1.2 Data preprocessing workflow

Visual data preprocessing, as per Multi - Sensor.pdf, aims to enhance efficiency. The left camera frame is divided into grids, and the "effective grid" (the one with the highest feature point density) is selected for feature extraction. The right frame only performs matching in grids adjacent to the left's effective grid to reduce redundancy [6]. If the number of matched points drops below a threshold of 50, global extraction is triggered. Visual data is normalized to the range [0,1] using Min - Max normalization, as described in jixiegou.pdf, to unify data scales [2].

For IMU and encoder data, as outlined in jixiegou.pdf, linear regression is used to fill in missing values. Boxplot analysis is employed to remove outliers (values outside the range of Q1−1.5×IQR to Q3 + 1.5×IQR). Mean filtering with a 3×3 window is used to smooth encoder - measured angles and reduce vibration - induced noise [7].

3.2 Fusion algorithm design and implementation

3.2.1 Hierarchical fusion logic

The system utilizes two fusion layers, as presented in Multi - Sensor.pdf. The first layer fuses visual and IMU data into Visual - Inertial Odometry (VIO) via an optimized VINS - Fusion framework. Visual reprojection errors are used to correct IMU drift, and an image quality metric (based on feature match count and noise level) assigns confidence weights to VIO (weights are ≤0.3 for sparse matches) [1].

The second layer fuses VIO with LIO using an Unscented Kalman Filter (UKF). A slip detector compares the displacement integrated by the IMU with the displacement calculated by LIO. A "slip flag" (triggered when the difference exceeds 0.05 m) reduces the weight of LIO (to ≤0.2) to prioritize VIO. In dim environments (where VIO confidence is low), the weight of LIO is increased (to ≥0.6) to maintain localization reliability [4].

3.2.2 Performance optimization goals

When benchmarked against single - sensor systems, as in Multi - Sensor.pdf, the fusion system aims to reduce localization error by more than 20%, cut down data processing time by 15.8%, and achieve a Root Mean Square Error (RMSE) of ≤0.0352 m in slippery and dim scenarios [1].

4 CNN - MPC - based motor control and gait planning

4.1 Improvement of motor control algorithms

4.1.1 Addressing traditional MPC limitations

To mitigate the computational burden of traditional MPC, as discussed in jixiegou.pdf, a Convolutional Neural Network (CNN) is employed to predict gait parameters. The CNN takes a 128 - dimensional input (including environmental parameters such as slope and obstacle distance, and historical gait data like joint angles and step frequency) and outputs predicted joint angle targets for 0.5 s. It consists of 2 convolutional layers (with 3×3 kernels and a stride of 1), 2 max - pooling layers (with 2×2 kernels), and 1 fully connected layer. It is trained with cross - entropy loss to ensure a prediction error of ≤5% [2].

4.1.2 CNN - MPC fusion control logic

The predictions from the CNN serve as the initial values for MPC, simplifying dynamic modeling, as in jixiegou.pdf. MPC uses a 0.1 s control cycle and a 5 - step prediction horizon. Its cost function is used to optimize motor inputs (servo speed and torque):
Here, Q and R are weight matrices, and P is the terminal weight matrix. Real - time encoder feedback is used to correct joint angle deviations (greater than 0.5°) for accuracy [5].

4.2 Scenario - adaptable gait planning

For elderly care, as in jixiegou.pdf, the robot adjusts the stride length (0.2–0.3 m) and step frequency (0.8–1.2 Hz) according to the elderly person's speed (0.3–0.5 m/s), ensuring a follow - up delay of ≤0.1 s. For small - scale complex scenarios, as in Multi - Sensor.pdf,

slippage triggers a 10–15% reduction in ankle motor torque (to increase ground contact) and shortens the stride length to 0.15–0.2 m (to reduce center - of - mass fluctuation) [1].

4.3 Motor control performance metrics

As per jixiegou.pdf, the system targets a motor response time of ≤0.08 s (12% faster than traditional MPC), a joint angle control error of ≤0.5°, and a failure rate of ≤2% during 2 - hour continuous operation [2].

5 Experimental Validation and Results Analysis

5.1 Experimental setup

5.1 Test platform and equipment

To verify the performance of the proposed VIO-LIO fusion localization system and CNN-MPC motor control algorithm, a small-scale biomimetic robotic dog prototype was built as the test platform. The robotic dog weighs 3.2 kg, has a body length of 0.45 m, and each leg is designed with two degrees of freedom (hip and knee joints) based on a parallel four-bar linkage structure [3]. The hardware configuration strictly refers to the sensor selection logic in the two core references:

Visual module: Stereo cameras were mounted on the robot's head, 0.15 m above the ground, to capture environmental visual features. This configuration is consistent with the miniature camera setup for small-scale biomimetic robots in A Robust Multi-Sensor Fusion Localization for Small-Scale Biomimetic Robots [1].

Inertial module: An IMU (sampling rate 100 Hz, angular velocity range ±2000°/s) was fixed on the robot's torso center to measure real-time acceleration and angular velocity, which is the same as the IMU parameter selection for leg-inertial odometry in [1].

Leg module: Motor encoders (accuracy 0.1°) were installed on the hip and knee joints of each leg to record joint angles. Combined with the parallel four-bar linkage kinematic model (Equation 1), foot-end displacement in the xy-plane was calculated to form LIO [1].

For ground truth acquisition, four motion capture cameras (sampling rate 120 Hz, position accuracy 0.001 m) were arranged around the test area, and a reflective marker was attached to the robot's head to record its actual position and trajectory [1]. Additionally, a high-definition camera (1080p resolution) was installed 2 m above the test site to record the robot's gait and movement status for subsequent gait stability analysis [8].

5.2 Test scenarios and protocols

To simulate the two core application scenarios of the robotic dog, two test scenarios were designed with reference to the environmental settings in [1] and [3]:

Scenario 1 (Elderly care simulation): A 10 m-long straight path was built in a dimly lit room (illumination 80 lux, consistent with the low-light environment in [3]). Small thresholds (height 0.05 m) and carpet edges were placed on the path to simulate minor terrain irregularities in elderly living spaces. The test protocol included:

1. The robotic dog follows a human tester (simulating an elderly user) walking at speeds of 0.3 m/s, 0.4 m/s, and 0.5 m/s respectively;
2. Record the robot's follow delay, joint angle error, and gait cycle fluctuation;
3. Repeat each speed test 5 times to calculate average values and standard deviations.

Scenario 2 (Small-scale complex simulation): A narrow passage (width 0.6 m) was constructed using textureless boards (to simulate feature-sparse environments), and a 2 m-long slippery acrylic board (friction coefficient 0.15, same as the slippery surface in [1]) was laid in the middle of the passage. The room illumination was adjusted to 50 lux to simulate dim conditions. The test protocol included:

1. The robotic dog travels through the narrow passage at a constant speed of 0.4 m/s;
2. Record the robot's localization RMSE, data processing time, and slippage rate;
3. Compare the performance of the proposed fusion system with single VIO and single LIO;

6 Repeat the test 5 times to ensure result reliability

6.1 Localization performance results

Table 1 shows the localization performance of the proposed fusion system, single VIO, and single LIO in Scenario 2. It can be seen that the fusion system significantly outperforms single-sensor systems:

RMSE: The fusion system's localization RMSE is 0.0352 m, which is 22.3% lower than single VIO (0.0453 m) and 38.1% lower than single LIO (0.0569 m). This meets the RMSE target (≤0.0352 m) proposed in [1] and verifies that the VIO-LIO fusion using UKF effectively compensates for the limitations of single sensors—LIO offsets VIO's error in dim environments, while VIO corrects LIO's slippage-induced inaccuracies.

Data processing time: The fusion system's average data processing time per frame is 4.6 ms, which is 15.8% shorter than the traditional global feature extraction method (5.4 ms in [1]). This is due to the dynamic local feature extraction strategy: by selecting only the "effective grid" (feature point density ≥30 points/grid) for feature extraction, redundant calculations are reduced, and processing efficiency is improved.

Table 1. Localization performance comparison in scenario 2

Method	RMSE (m)	Processing Time (ms)	Error Reduction Rate (%)
Single VIO	0.0453	5.6	/
Single LIO	0.0569	4.0	/
Fusion System	0.0352	4.6	22.3 (vs VIO)

The fusion system's trajectory (red curve) is closest to the ground truth (black curve), while single VIO (blue curve) drifts significantly in the dim section of the passage (3–5 m), and single LIO (green curve) deviates from the ground truth in the slippery section (6–8 m). This further confirms the robustness of the fusion system in complex environments.

6.2 Motor control and gait stability results

Table 2 shows the motor control performance of the proposed CNN-MPC algorithm and the traditional MPC algorithm in Scenario 1. The CNN-MPC algorithm exhibits obvious advantages:

Response time: The CNN-MPC algorithm's motor response time is 0.076 s on average, which is 12% faster than the traditional MPC algorithm (0.086 s). This is because the CNN predicts gait parameters (joint angle targets) in advance, simplifying the MPC's dynamic modeling process and reducing the optimization solving time [2].

Joint angle error: The CNN-MPC algorithm's average joint angle error is 0.42°, which is lower than the traditional MPC algorithm's 0.58° and meets the target (≤0.5°) in [2]. The real-time encoder feedback correction ensures that the joint angle deviation is quickly adjusted when the robot encounters thresholds.

Follow delay: The CNN-MPC algorithm's average follow delay is 0.09 s, which is shorter than the traditional MPC algorithm's 0.13 s and meets the safety requirement (≤0.1 s) for elderly care scenarios [2].

Table 2. Motor control performance comparison in scenario 1

Algorithm	Response Time (s)	Joint Angle Error (°)	Follow Delay (s)
Traditional MPC	0.086	0.58	0.13
CNN-MPC	0.076	0.42	0.09

The CNN-MPC algorithm's gait cycle fluctuation is 4.2% on average, which is 12% lower than the traditional MPC algorithm's 4.8%. This indicates that the CNN-MPC algorithm can better maintain gait stability when following the elderly's variable walking speeds [2]. In the slippery section of Scenario 2, the CNN-MPC algorithm reduces the ankle motor torque by 12% on average, increasing the foot-end ground contact area by 18% and reducing the slippage rate from 12% (traditional MPC) to 3% [2].

7 Conclusion

This study proposes an integrated technical scheme for biomimetic robotic dogs, combining VIO-LIO multi-sensor fusion localization and CNN-enhanced MPC motor control. Based on experimental verification and analysis, the main conclusions are as follows:

1 The VIO-LIO fusion localization system improves environmental robustness: By adopting dynamic local feature extraction and UKF-based fusion, the system reduces localization error by over 20% compared to single-sensor systems and shortens data processing time by 15.8%. In slippery and dim scenarios, the localization RMSE reaches 0.0352 m, which meets the practical application requirements of small-scale complex environments.

2 The CNN-MPC motor control algorithm enhances real-time performance and stability: The CNN's gait prediction reduces the MPC's computational load, shortening the motor response time by 12% and reducing the joint angle error to ≤0.5°. In elderly care scenarios, the follow delay is controlled within 0.1 s, and the gait cycle fluctuation is 4.2%, ensuring safe and stable human-robot interaction.

3 The integrated system has strong scenario adaptability: Experimental results in elderly care and small-scale complex scenarios show that the system can adapt to variable walking speeds, minor terrain irregularities, and slippery/dim environments, providing a reliable technical solution for the practical application of biomimetic robotic dogs. This study still has limitations: In extreme slope environments (slope >30°), the parallel four-bar linkage kinematic model of the leg module cannot fully account for the change in foot-end force direction, leading to a 15% increase in LIO error and a slight rise in localization RMSE to 0.051 m. Additionally, the CNN-MPC algorithm's gait prediction accuracy decreases in rapidly changing environments (e.g., sudden obstacles), requiring 0.2 s of retraining time to adjust parameters.

Future work will focus on two aspects:

1 Optimize the kinematic model: Integrate a force sensor at the foot end to dynamically correct the LIO displacement calculation based on ground reaction force, improving the system's adaptability to extreme slopes.

2 Enhance the CNN's real-time learning ability: Introduce online learning mechanisms to enable the CNN to update the gait prediction model in real time without retraining, improving the algorithm's response speed to sudden environmental changes.

References

1 S. Li, Y. Zhang, Z. Chen, et al. A Robust Multi - Sensor Fusion Localization for Small - Scale Biomimetic Robots. In Proceedings of the 2024 IEEE International Conference on Robotics and Biomimetics, Bangkok, Thailand, (2024).

2 W. Wang, X. Huang, X. Pan, et al. Research on Gait Planning for Elderly Biomimetic Robot Dogs. In Proceedings of the 2024 4th International Conference on Artificial Intelligence, Robotics, and Communication, (2024).

3 C. Campos, R. Elvira, J. J. G. Rodríguez, et al. ORB - SLAM3: An Accurate Open - Source Library for Visual, Visual–Inertial, and Multi - Map SLAM. IEEE Transactions on Robotics, 37(6):1874–1890, (2021).

4 T. Qin, P. Li, S. Shen. VINS - Mono: A Robust and Versatile Monocular Visual - Inertial State Estimator. IEEE Transactions on Robotics, 34(4):1004–1020, (2018).

5 Z. Zhang, X. Chang, H. Ma, et al. Model Predictive Control of Quadruped Robot Based on Reinforcement Learning. Applied Sciences, 13(1):154, (2023).

6 H. El - Hussieny. Real - Time Deep Learning - Based Model Predictive Control of a 3 - DOF Biped Robot Leg. Scientific Reports, 14(1):66104, (2024).

7 Q. Shi, J. Gao, S. Wang, et al. Development of a Small - Sized Quadruped Robotic Rat Capable of Multimodal Motions. IEEE Transactions on Robotics, 38(5):3027–3043, (2022).

8 X. Liu, Y. Wang, J. Li. Gait Stability Optimization for Quadruped Robots Based on Force Sensing. Journal of Robotics, 10 (2): 45-62, (2023).