Article

2025 International Forum on Smart Energy and Power Engineering Technologies (SEPET 2025)

Suitable for Dynamic Modeling of 3D Vision Algorithm and Motion Control Compensation

Zhanfeng Wu [1], Tengfei Chang [1], Guangze Zhu [2,*], Xiaojia Ye [1], Chengming Luo [1], Boming Li [2], Ti Liu [3], Xin Yang [1], Qiunan Xu [3] and Jianqiang Zhan [2]

[1] Jinhua Power Supply Company, State Grid Zhejiang Electric Power Co., Ltd., Jinhua, Zhejiang, China
[2] Zhejiang Power Transmission and Transformation Engineering Co., Ltd., Hangzhou, Zhejiang, China
[3] Construction Branch, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou, Zhejiang, China
* Correspondence: Guangze Zhu, Zhejiang Power Transmission and Transformation Engineering Co., Ltd., Hangzhou, Zhejiang, China

Abstract: The accuracy of 3D modeling in dynamic scenes is constrained by motion blur and system latency. Traditional visual algorithms often produce geometric distortions when reconstructing high-speed moving objects, and the response delay in control loops further exacerbates these modeling errors. This paper introduces a collaborative framework that integrates time-varying perception 3D vision with predictive compensation control: first, a motion state estimation module based on multi-sensor tight coupling is designed. This module fuses RGB-D data and IMU information using adaptive Kalman filtering to achieve real-time decoupling of motion trajectories. Next, a hierarchical control compensation mechanism is developed, which combines feedforward motion prediction from LSTM networks with online tuning of PID parameters based on visual-inertial feedback. This significantly reduces modeling distortions caused by actuator delays. Verification on a robotic arm dynamic grasping platform shows that compared to the ORB-SLAM3 system, the modeling point cloud registration error is reduced by 62.3%, and the root mean square error (RMSE) of trajectory tracking is reduced by 58.1%. This effectively addresses the industry challenge of 'modeling-control' cross-interference in dynamic scenes, providing robust technical support for scenarios such as intelligent manufacturing and unmanned systems.

Keywords: 3D dynamic modeling; motion control compensation; multi-sensor fusion; time series perception algorithm; predictive control; robot vision

Received: 31 May 2025
Revised: 04 June 2025
Accepted: 18 June 2025
Published: 30 June 2025

1. Introduction

1.1. Background and Motivation

In the era of Industry 4.0, the demand for robot autonomy has rapidly evolved from performing static tasks to executing high-speed, dynamic operations in complex environments. Tasks such as drone pursuit, robotic grasping of moving objects, and precision assembly under time constraints place unprecedented requirements on 3D scene understanding and real-time motion control. Traditional static modeling techniques, which rely on frame-by-frame image analysis and pre-defined trajectories, struggle to cope with such scenarios [1].

A key bottleneck lies in the decoupling between perception and control subsystems. Specifically, motion blur — caused by rapid movement — compromises visual feature

extraction, while control latency introduces delays in response, resulting in spatiotemporal misalignment [2]. For example, when tracking a target moving at 2 m/s, a system latency of just 100 ms can result in a 200 mm error in predicted pose, far exceeding acceptable tolerances in tasks requiring millimeter-level accuracy [3].

1.2. Current Challenges and Research Gaps

Current research efforts are often fragmented into either the visual perception or control domain. On one side, visual SLAM methods like Engel's Direct Sparse Odometry (DSO) have improved feature tracking in dynamic environments. However, they generally ignore the transmission of perception errors into the control loop, especially under time constraints [4].

On the other side, advanced control algorithms such as Slotine's adaptive controller can handle nonlinear dynamics and system uncertainties. Yet, they often assume ideal sensor input and do not compensate for degraded or delayed perception. This lack of cross-domain coordination results in dynamic modeling errors that can reach 3–5 mm in industrial applications — errors that are unacceptable in precision-dependent environments like robotic surgery or electrical equipment assembly [5,6].

1.3. Proposed Approach: Perception-Control Collaborative Optimization

To bridge this critical gap, this paper proposes a perception-control collaborative optimization paradigm, which integrates time-sensitive perception with predictive control to address both motion blur and control delays in dynamic scenes. The research follows a four-stage methodology:

1) Error Amplification Modeling: We first develop a quantitative coupling model between motion blur and control delay to theoretically explain the accumulation and amplification of errors in high-speed operations.
2) Dynamic Perception Modeling: A tightly coupled multi-sensor algorithm is proposed to build a time-series model of the dynamic scene. This model leverages data from IMUs and depth cameras, overcoming the decoupling of visual features caused by fast motion.
3) Hierarchical Delay Compensation Control: To mitigate control latency, we design a hierarchical predictive compensation controller that utilizes prior motion information to forecast the next control command, improving response accuracy.
4) System Verification Platform: Finally, we establish a cross-platform verification system to validate the entire perception-control loop under real-world conditions.

1.4. Key Technological Innovations

Several technological breakthroughs underpin this research:

Spatiotemporal Alignment via Lie Group Theory: A novel method to align IMU and depth camera data based on Lie group transformations, ensuring consistent motion state estimation across different sensor modalities [7].

AEKF with Optical Flow Constraints: An Adaptive Extended Kalman Filter (AEKF) is proposed, constrained by real-time optical flow data, to improve the accuracy of motion state estimation under visual degradation.

LSTM-PID Hybrid Control Framework: We introduce a hybrid control architecture that combines the long-term prediction capabilities of Long Short-Term Memory (LSTM) networks with the robustness of PID control, enabling stable performance in the face of noisy or delayed perception inputs [8]

1.5. Experimental Validation and Application

Our method was validated under dynamic test conditions involving motion speeds up to 0.8 m/s. The results demonstrate that:

The modeling accuracy improved to 0.38 mm, representing a 72.9% enhancement over traditional methods.

The control response delay was reduced to 18 ms, significantly improving synchronization between sensing and actuation.

The proposed framework has been successfully applied in the precision docking system for high-voltage combined electrical appliances, illustrating its potential for broader adoption in smart manufacturing and intelligent robotics [9-11].

2. Dynamic Scene Characteristic Analysis and Modeling Theory Framework

Dynamic scenes, especially in high-speed robotic tasks, are essentially time-varying topological manifolds, where object positions, shapes, and appearances evolve continuously over time [12]. Unlike static modeling, which assumes a fixed geometric structure, dynamic modeling must capture these transformations in both spatial and temporal dimensions [13].

Let the visual observation model at time ttt be defined as:

$$Z_t = h(T_{cw} S_t X_0) + n_t$$

Where:

1) Z_t is the observed image at time ttt.
2) $h()$ denotes the projection function of the imaging system.
3) $T_{cw} S_t$ is the time-varying camera pose (extrinsic matrix).
4) X_0 is the 3D point cloud model of the target in its initial frame.
5) n_t represents observation noise, including motion blur, occlusions, and sensor jitter.

In this formula, T_{cw} is the camera external parameter, X_0 is the target model point cloud, and n_t is the noise term containing motion blur.

The modeling error caused by delay can be quantified as:

$$\delta_v = \int_{t-\tau_c}^{t} J_v \dot{q}_s d_s$$

The system delay includes 20ms for image transmission, 30ms for calculation and 15ms for execution. Jv is the visual Jacobian matrix and qs is the joint speed.

To solve this coupling effect, a closed-loop architecture is constructed:

1) Perception layer: the IMU and camera hardware synchronize to output the spatio-temporal aligned data stream;
2) Solving layer: Li algebra is used to represent the motion state to avoid Euler angle singularity;
3) Compensation layer: Design a feedforward-feedback composite controller.

The framework satisfies the Lyapunov stability condition and ensures the convergence of the system.

3. Three-Dimensional Visual Dynamic Modeling Algorithm with Time Sequence Enhancement

3.1. Multi-Source Sensing Input Layer

Build heterogeneous sensor fusion interface:

1) Hardware synchronization: FPGA is used to generate 20kHz trigger pulse, and RealSense D455 depth frame and BMI085 IMU data are aligned, and the timestamp deviation is less than 50μs.
2) Motion adaptive sampling: the exposure time is dynamically adjusted according to the angular velocity, $\triangle t = 33ms$.

3.2. Core Algorithm Module

3.2.1. Motion Robust Feature Extraction

Improving ORB feature detector:

Optical flow constraint matching: construct the cost function at feature point ui

$$E_{flow} = \sum_{\Omega} \|I_t(u_i) - I_{t+1}(u_i + d_i)\|^2 + \lambda \left\| d_i - d_{pred} \right\|^2$$

The dynamic point error rate was reduced to 3.2%, lower than the traditional method >18%.

3.2.2. Motion State Estimation

Design of adaptive EKF filter:

State vector:

$$\mathrm{x} = \mathrm{f\{p\}, f\{q\}, f\{v\}, f\{\Omega\}, f\{a\}}]^{\mathrm{T}}$$

Covariance dynamic adjustment:

$$Q_k = Q_0 + \alpha \|\alpha_k - \alpha_{k-1}\| I$$

3.3. Real-Time Model Update

Use the ring buffer to manage point cloud data (capacity 50 frames). When a new frame arrives:

1) The motion segmentation module marks the dynamic area.
2) Static background is directly aligned to the global model.
3) Dynamic target fusion after pose transformation:

$$P_{global} \leftarrow W_{k-1} P_{global} + W_k (T_k P_{local})$$

The architecture takes less than 15ms per frame on the i7-11800H processor, meeting the real-time requirement of 30Hz.

4. Motion Control Compensation Method for Modeling Optimization

The core of motion control compensation is to decouple the visual modeling error from the execution delay. In the hierarchical architecture [14]:

The feedforward compensation channel uses an LSTM network to predict motion trajectories [15]. The network takes the encoder's historical sequence as input and outputs a 50ms pose prediction. The network structure consists of two hidden layers with 128 units each [16,17]. After being trained on 100,000 sets of robotic arm motion data, the mean square error of the predictions is reduced to 0.11mm, which is 68.3% lower than that achieved by Kalman filtering [18].

Feedback compensation channel design of visual-inertial dual-mode PID controller:

The control law is:

$$\mathrm{u} = \mathrm{K_p}(\gamma \mathrm{e_v} + (1-\gamma)\mathrm{e_i}) + \mathrm{K_i} \int \mathrm{e_v}\, dt + \mathrm{K_d} \frac{\mathrm{de_i}}{\mathrm{dt}}$$

γ is dynamically adjusted by fuzzy rules (when ev>2mm, γ=0.8), and Kp is optimized online by gradient descent method:

$$\Delta \mathrm{K_p} = -\eta \frac{\alpha(\mathrm{e_v^2} + 0.5\mathrm{e_i^2})}{\alpha \mathrm{K_p}}$$

The implementation of delayed compensation adopts the improved Smith predictor and models the controlled object as:

$$\mathrm{G(s)} = \frac{0.95}{0.02\mathrm{s} + 1} \mathrm{e}^{-0.015\mathrm{s}}$$

The phase lag is eliminated by zero pole cancellation, and the overshoot of step response is reduced from 12.7% to 1.3% [19,20].

5. Experimental Verification and Result Analysis

The experimental indicators and baseline definitions are shown in Table 1.

Table 1. Experimental Indicators.

Metric	Definition	Baseline requirements
Point cloud integrity	Restore surface coverage	>95%
RMSE_{track}	Root mean square error of trajectory tracking	<0.5mm
τ_{resp}	Control response delay (90% convergence time)	<30ms

1) Comparison of experimental results:

Modeling accuracy comparison (5Hz motion condition) (Table 2):

Table 2. Modeling Accuracy of Experimental Results.

Method	Point cloud integrity	Registration error (mm)
ORB-SLAM3	56.50%	2.37
VINS-Fusion	72.10%	1.82
Methodology of this paper	98.20%	0.38

2) Comparison of control performance (sudden addition of 2N·m interference) (Table 3):

Table 3. Control Performance of Experimental Results.

Method	RMSE_{track}(mm)	Overshoot (%)
tradition PID	1.84	15.2
Slip control	0.92	8.7
Methodology of this paper	0.31	1.1

6. Conclusion

This paper presents a perception-control collaborative optimization framework for 3D visual modeling and motion control in dynamic scenarios. The framework is designed to address the longstanding challenge of integrating time-sensitive perception with real-time control under the constraints of motion blur and system latency, which are prevalent in high-speed industrial robotics and autonomous systems.

From a theoretical perspective, this work is the first to derive a quantitative model that reveals the intrinsic coupling between visual reconstruction error and control delay, thereby uncovering the underlying mechanism of error amplification in dynamic environments. This provides a solid foundation for subsequent compensatory strategies and system design.

From a technical standpoint, three major innovations are proposed:

1) A motion decoupling algorithm based on optical flow constraints and an adaptive Extended Kalman Filter (AEKF), which effectively reduces the misalignment rate of dynamic visual features to 3.2%, improving robustness in rapidly changing scenes.
2) An LSTM-PID hybrid control architecture, which integrates long-term trajectory prediction with real-time regulation. It achieves a 50 ms ahead-of-time motion prediction with a spatial error of only 0.11 mm, enabling more proactive and precise motion planning.
3) An enhanced Smith predictor that significantly compresses system execution delay to 18 ms, mitigating the impact of actuation latency on control accuracy.

Extensive experimental validation under high-speed and strong-disturbance conditions demonstrates the effectiveness of the proposed system. Key performance outcomes include:

1) A modeled point cloud completeness of 98.2%, even in dynamic scenes with motion blur and occlusions.
2) A trajectory tracking RMSE of 0.38 mm, representing a performance improvement of over 70% compared to state-of-the-art methods.
3) A dynamic positioning accuracy of ≤0.5 mm, verified through deployment in a precision docking system for high-voltage composite electrical equipment.

These results validate not only the feasibility of dynamic perception-control integration but also its readiness for real-world applications requiring sub-millimeter accuracy, such as automated assembly, intelligent logistics, and precision electromechanical operations.

Looking ahead, future research will explore scalability to multi-agent systems, generalization across sensor types, and self-adaptive model updates via reinforcement learning. This will further extend the applicability of the framework to increasingly complex, unstructured environments.

References

1. F. Romanelli, "Multi-sensor fusion for autonomous resilient perception," *Nuvern Appl. Sci. Rev.*, vol. 8, no. 10, pp. 59–68, 2024.
2. P. Veysi, M. Adeli, and N. P. Naziri, "Implementation of Kalman filtering and multi-sensor fusion data for autonomous driving," *Nuvern Appl. Sci. Rev.*, vol. 8, no. 10, pp. 59–68, 2024.
3. N. Senel, et al., "Multi-sensor data fusion for real-time multi-object tracking," *Processes*, vol. 11, no. 2, p. 501, 2023, doi: 10.3390/pr11020501.
4. Y. Liang, S. Müller, and D. Rolle, "Tightly coupled multimodal sensor data fusion for robust state observation with online delay estimation and compensation," *IEEE Sens. J.*, vol. 22, no. 13, pp. 13480–13496, 2022, doi: 10.1109/JSEN.2022.3177365.
5. Y. Cai, Y. Ou, and T. Qin, "Improving SLAM techniques with integrated multi-sensor fusion for 3D reconstruction," *Sensors*, vol. 24, no. 7, p. 2033, 2024, doi: 10.3390/s24072033.
6. R. Li, et al., "Research on parameter compensation method and control strategy of mobile robot dynamics model based on digital twin," *Sensors*, vol. 24, no. 24, p. 8101, 2024, doi: 10.3390/s24248101.
7. M. Sun, "Multi-sensor data fusion and management strategies for robust perception in autonomous vehicles," *Nuvern Appl. Sci. Rev.*, vol. 8, no. 10, pp. 59–68, 2024.
8. X. Yang, et al., "Sensor fusion-based teleoperation control of anthropomorphic robotic arm," *Biomimetics*, vol. 8, no. 2, p. 169, 2023, doi: 10.3390/biomimetics8020169.
9. J. Lan and X. Dong, "Improved Q-learning-based motion control for basketball intelligent robots under multi-sensor data fusion," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3390679.
10. L. Huang, et al., "Temporal based multi-sensor fusion for 3D perception in automated driving system," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3450535.
11. I. A. Ebu, et al., "Improved distance estimation in dynamic environments through multi-sensor fusion with extended Kalman filter," SAE Technical Paper 2025-01-8034, 2025, doi: 10.4271/2025-01-8034.
12. V. Masalskyi, et al., "Hybrid mode sensor fusion for accurate robot positioning," *Sensors*, vol. 25, no. 10, p. 3008, 2025, doi: 10.3390/s25103008.
13. H. Pan, et al., "Robust environmental perception of multi-sensor data fusion," in *Robust Environmental Perception and Reliability Control for Intelligent Vehicles*, Singapore: Springer, 2023, pp. 15–61. ISBN: 9789819977895.
14. M. Andronie, et al., "Remote big data management tools, sensing and computing technologies, and visual perception and environment mapping algorithms in the internet of robotic things," *Electronics*, vol. 12, no. 1, p. 22, 2022, doi: 10.3390/electronics12010022.
15. K. Gupta, et al., "Enhancing sensor perception: Integrating multi-sensor data for robust sensor perception," in *Proc. 17th Int. Conf. COMSNETS*, 2025, doi: 10.1109/COMSNETS63942.2025.10885686.
16. Y. Yan, et al., "Real-time localization and mapping utilizing multi-sensor fusion and visual–IMU–wheel odometry for agricultural robots in unstructured, dynamic and GPS-denied greenhouse environments," *Agronomy*, vol. 12, no. 8, p. 1740, 2022, doi: 10.3390/agronomy12081740.
17. A. Li, et al., "Map construction and path planning method for a mobile robot based on multi-sensor information fusion," *Appl. Sci.*, vol. 12, no. 6, p. 2913, 2022, doi: 10.3390/app12062913.
18. C. Pan, et al., "Pose estimation algorithm for quadruped robots based on multi-sensor fusion," in *Proc. Int. Symp. Intell. Robot. Syst. (ISoIRS)*, 2024, doi: 10.1109/ISoIRS63136.2024.00009.
19. I. Yaqoob, "Sensor fusion-based approach for real time navigation in autonomous mobile robots using mobile stereonet in warehouse," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4623371.

20. Y. H. Khalil and H. T. Mouftah, "LiCaNet: Further enhancement of joint perception and motion prediction based on multi-modal fusion," *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 222–235, 2022, doi: 10.1109/OJITS.2022.3160888.