International Core Journal of Engineering Volume 12 Issue 4, 2026
ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

Robust Multi-Sensor Fusion for Dynamic Robotic Grasping Using Fuzzy Adaptive Extended Kalman Filter

Daohu Zhang*

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

*Email: 232260500@st.usst.edu.cn

Abstract

Precise target localization and tracking are fundamental prerequisites for robotic arm grasping tasks in dynamic environments. Traditional single-sensor systems, such as depth cameras, are susceptible to environmental interference, particularly under varying illumination and complex backgrounds. To solve this problem, this paper proposes a novel lightweight multi-sensor fusion framework combining 2D LiDAR and depth camera data using a Fuzzy Adaptive Extended Kalman Filter (FA-EKF). Unlike standard EKF methods that use a fixed observation noise covariance matrix R, the proposed FA-EKF utilizes a fuzzy logic controller to dynamically adjust the noise covariance based on real-time measurement innovations. This allows the system to robustly handle non-linearities, sensor degradation, and time-varying noise. In the visual perception module, an improved YOLOv8 equipped with a Convolutional Block Attention Module (CBAM) is utilized to enhance feature extraction. Meanwhile, the LiDAR module extracts cylindrical targets using a robust RANSAC-based geometric fitting algorithm rather than traditional least-squares. Experimental validations performed on an AUBO-i5 robotic arm tracking a hovering coaxial drone demonstrate that the proposed FA-EKF method reduces the Root Mean Square Error (RMSE) to 0.0038m, and achieves a grasping success rate of 96% even under low-light conditions. The proposed system offers a real-time and highly robust solution for dynamic robotic grasping.

Keywords

Multi-Sensor Fusion; Fuzzy Logic; Extended Kalman Filter; Robotic Grasping; 2D LiDAR; RANSAC.

1. Introduction

With the continuous advancement of robotics, robotic arms are increasingly deployed in agriculture, medical care, and automated manufacturing. Precise spatial perception and dynamic target tracking are the foundations for a successful robotic grasp. Currently, most traditional grasping systems rely heavily on visual sensors, such as RGB-D cameras. Although depth cameras provide rich semantic information and dense 3D point clouds, their depth measurement accuracy is highly sensitive to illumination changes, resulting in significant errors when facing transparent or highly reflective targets. On the other hand, 2D LiDAR provides robust and highly accurate distance measurements independent of lighting conditions, but lacks the necessary semantic context for complex object recognition.

To overcome the limitations of single sensors, fusing camera and LiDAR data has become a popular research direction. Sensor fusion generally falls into two categories: feature-level fusion and data-

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

level fusion. Feature-level fusion relies on complex deep neural networks to extract and fuse high-dimensional representations. While accurate, these methods are computationally heavy, require extensive training datasets, and are unsuitable for edge devices with strict real-time constraints. Data-level fusion, typically using Bayesian filters like the Extended Kalman Filter (EKF)[1], offers a lightweight alternative. However, standard EKF algorithms use fixed process and measurement noise covariance matrices, which perform poorly in real-world scenarios where sensor noise is non-Gaussian, unpredictable, and time-varying.

To address these challenges, this paper proposes a Fuzzy Adaptive Extended Kalman Filter (FA-EKF). By introducing a fuzzy inference system, the proposed method evaluates the real-time innovation sequence and dynamically adjusts the measurement noise covariance.

The main contributions of this paper are summarized as follows:

We propose a FA-EKF multi-sensor fusion algorithm that uses a fuzzy logic controller to dynamically adjust the sensor weights and noise covariances, significantly enhancing system stability in unknown or high-noise environments.

We improve the visual and LiDAR perception pipelines by integrating the CBAM attention mechanism into YOLOv8 and adopting the RANSAC algorithm for robust point cloud geometric fitting.

The proposed framework is deployed on an actual AUBO-i5 robotic arm. Real-world experiments demonstrate that the system achieves high-precision dynamic grasping without requiring massive computational resources.

2. Related Work

2.1 Single-Modality Target Detection

Single-modality target detection in robotic grasping primarily relies on either vision-based sensing or LiDAR-based sensing.

LiDAR sensors are widely adopted in robotic perception because of their high ranging accuracy and robustness to illumination variations. Traditional LiDAR-based methods usually perform object localization through point cloud clustering, contour extraction, or geometric model fitting. These methods are computationally efficient and suitable for real-time applications, but their performance often degrades when the point cloud is sparse, partially occluded, or lacks sufficient geometric structure. To improve representation capability, deep learning-based point cloud methods such as PointNet [2], PointNet++ [3], and graph-based models such as Point-GNN [4] have been proposed. These approaches significantly enhance 3D feature extraction and object recognition performance, but they also introduce higher computational complexity and memory overhead, which may limit deployment on resource-constrained industrial robotic platforms.

Compared with LiDAR, vision-based methods provide richer semantic and texture information and have become a mainstream solution for object detection and pose estimation. In particular, one-stage detectors represented by YOLO (You Only Look Once) [5] have achieved an excellent balance between speed and accuracy, making them highly suitable for real-time robotic applications. Subsequent improvements, such as lightweight architectural optimization and attention mechanisms, further enhance robustness in cluttered scenes and under partial occlusion. For example, CBAM (Convolutional Block Attention Module) [6] improves feature discrimination by adaptively emphasizing informative spatial and channel responses. Nevertheless, monocular or RGB-D vision systems remain sensitive to illumination changes, reflective surfaces, and depth measurement instability, especially in complex industrial environments. Therefore, relying on a single sensing modality is often insufficient to guarantee the precision and robustness required for high-accuracy robotic grasping.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

2.2 Multi-Sensor Fusion

To overcome the limitations of single-modality perception, multi-sensor fusion has been extensively studied as an effective strategy for combining the complementary strengths of LiDAR and vision. In general, LiDAR provides accurate geometric distance measurements, while cameras offer dense semantic and appearance information. Recent studies have explored deep learning-based feature-level fusion frameworks that map image and point cloud data into a shared representation space. Representative methods include BEVFusion [7], which unifies multimodal features in a bird's-eye-view representation, and TransFusion [8], which employs transformer-based cross-modal interaction to improve robustness under calibration errors and degraded visual conditions. These methods achieve state-of-the-art performance in 3D perception tasks, but they usually require powerful GPUs and large-scale training datasets, making them less practical for industrial edge deployment and real-time robotic manipulation systems.

In contrast, data-level fusion methods directly combine low-level sensor measurements and are often more computationally efficient. Among them, Kalman-filter-based approaches have been widely adopted for target tracking and state estimation because of their recursive structure and real-time capability. The Extended Kalman Filter (EKF) is particularly suitable for nonlinear systems and has been used in many robotic perception and localization tasks. However, the standard EKF assumes relatively accurate prior noise statistics and is sensitive to abnormal observations, model mismatch, and time-varying measurement uncertainty. In practical robotic grasping scenarios, sensor noise often changes dynamically due to illumination variation, partial occlusion, reflective interference, or target motion, which can significantly degrade the performance of a conventional EKF.

To address these limitations, recent research has focused on robust and adaptive Kalman filtering strategies, such as outlier-resistant filtering and adaptive covariance adjustment [9], [10]. Inspired by this line of work, the proposed Fuzzy Adaptive Extended Kalman Filter (FA-EKF) introduces an adaptive noise regulation mechanism that dynamically adjusts measurement confidence according to observation consistency. This design preserves the computational efficiency of data-level fusion while improving robustness against sensor uncertainty and outlier disturbances, making it more suitable for real-time robotic grasping in complex industrial environments.

3. Methodology

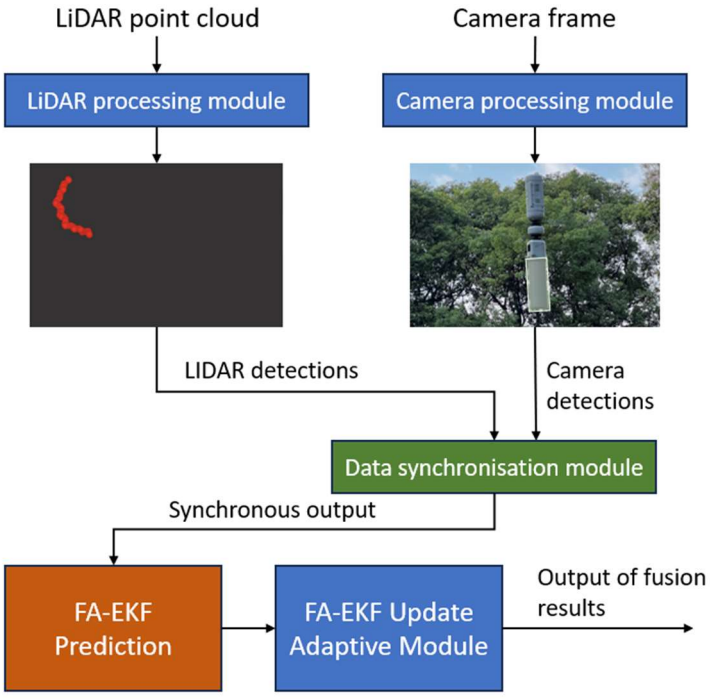

Figure 1. The overall framework of the proposed multi-sensor fusion system.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

The proposed system architecture consists of three main modules: visual perception, LiDAR data processing, and FA-EKF multi-sensor fusion. As shown in Figure 1, the depth camera provides semantic and spatial cues, while the 2D LiDAR supplies robust geometric distance measurements. These complementary observations are then fused by the proposed FA-EKF to achieve stable and accurate target localization for dynamic robotic grasping.

3.1 Vision Processing Module

To improve the robustness of visual perception in complex environments, the Convolutional Block Attention Module (CBAM)[6] is integrated into the backbone network of YOLOv8[11]. CBAM enhances intermediate feature representations by sequentially modeling channel-wise and spatial-wise attention, enabling the detector to focus more effectively on salient target regions while suppressing irrelevant background interference.

Given an intermediate feature map $\mathbf{F}$, the channel attention refinement is first applied as

$$\mathbf{F}^{c} = M_{c}(\mathbf{F}) \odot \mathbf{F} \tag{1}$$

where $M_{c}(\mathbf{F})$ denotes the channel attention map and $\odot$ represents element-wise multiplication. Subsequently, the spatial attention module further refines the channel-enhanced feature map:

$$\mathbf{F}^{cs} = M_{s}(\mathbf{F}^{c}) \odot \mathbf{F}^{c} \tag{2}$$

where $M_{s}(\mathbf{F}^{c})$ denotes the spatial attention map, $\mathbf{F}^{c}$ is the feature map after channel attention refinement, and $\mathbf{F}^{cs}$ is the final refined feature map after sequential channel–spatial attention processing.

This two-stage attention mechanism is particularly suitable for the proposed dynamic grasping scenario, where the hovering target occupies a relatively small region in the image and is easily affected by illumination variations and cluttered laboratory backgrounds. Consequently, the CBAM-enhanced YOLOv8 detector can effectively suppress background noise (e.g., laboratory equipment and surrounding interference) and improve bounding box localization accuracy under challenging lighting conditions. As illustrated in Figure 2, the improved YOLOv8 model can reliably detect the hovering target and maintain stable bounding box localization even under complex background interference.

Figure 2. Target detection results of the vision processing module using improved YOLOv8.

3.2 LiDAR Processing Module

Given that the hovering drone exhibits an approximately cylindrical profile, the 2D LiDAR scans typically form a semicircular point distribution in the sensing plane. Accordingly, the LiDAR-based

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

target localization process consists of three main steps: coordinate transformation, noise filtering, and robust geometric fitting.

First, the raw LiDAR measurements in polar coordinates (d_m, θ) are transformed into Cartesian coordinates in the LiDAR plane as:

$$x_m = d_m \sin\theta \tag{3}$$

$$z_m = d_m \cos\theta \tag{4}$$

where d_m denotes the measured distance and θ represents the corresponding scan angle. The transformed point set (x_m, z_m) describes the target contour in the 2D sensing plane.

Next, to reduce the influence of sparse noise and isolated measurement disturbances, a Statistical Outlier Removal (SOR)[12] filter is applied to the point cloud. This step effectively removes scattered outlier points caused by sensor noise or unstable reflections, thereby improving the structural consistency of the observed target contour.

Finally, instead of adopting the conventional Least Squares fitting method, the Random Sample Consensus (RANSAC) algorithm is employed for robust geometric fitting. RANSAC iteratively selects random subsets of the filtered point cloud to estimate the circular parameters of the target profile, including the circle center (x_c, z_c) and radius r. By maximizing the number of inlier points that satisfy the fitted model, RANSAC significantly improves robustness against outliers introduced by the rotating propellers of the drone or environmental interference. As a result, the LiDAR subsystem can stably estimate the geometric center of the hovering target, providing reliable lateral and depth constraints for subsequent multi-sensor fusion and grasp planning.

Figure 3. Point cloud extraction and robust RANSAC circle fitting results.

3.3 Fuzzy Adaptive EKF (FA-EKF) Fusion

To achieve robust and accurate dynamic target tracking under complex sensing conditions, a Fuzzy Adaptive Extended Kalman Filter (FA-EKF) is proposed for multi-sensor fusion. The proposed method integrates the depth camera and 2D LiDAR measurements within a unified state estimation framework while adaptively adjusting the measurement noise covariance according to the real-time observation consistency. This design effectively improves the robustness of the filter against illumination changes, partial occlusions, and LiDAR measurement disturbances.

In the FA-EKF framework, the state vector of the dynamic target is defined as:

$$\mathrm{X}_k = [x_k \quad y_k \quad z_k \quad v_{x,k} \quad v_{y,k} \quad v_{z,k} \quad a_{x,k} \quad a_{y,k} \quad a_{z,k}]^T \tag{5}$$

where (x_k, y_k, z_k) denote the 3D position of the target at time step k, $(v_{x,k}, v_{y,k}, v_{z,k})$ denote the corresponding velocities, and $(a_{x,k}, a_{y,k}, a_{z,k})$ denote the accelerations along the three axes.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

3.3.1 Standard EKF Process

Assuming a constant-acceleration motion model, the prediction step of the EKF is formulated as:

$$\hat{\mathrm{X}}_{k|k-1} = \mathrm{F}\hat{\mathrm{X}}_{k-1|k-1} \tag{6}$$

$$\mathrm{P}_{k|k-1} = \mathrm{F}\mathrm{P}_{k-1|k-1}\mathrm{F}^{T} + \mathrm{Q} \tag{7}$$

where $\hat{\mathrm{X}}_{k|k-1}$is the predicted state vector, $\mathrm{P}_{k|k-1}$is the predicted state covariance matrix, Fis the state transition matrix, and Qdenotes the process noise covariance matrix.

The measurement innovation (residual) is computed as:

$$\mathrm{e}_k = \mathrm{Z}_k - h\left(\hat{\mathrm{X}}_{k|k-1}\right) \tag{8}$$

where $\mathrm{Z}_k$denotes the fused sensor observation at time step k, and $h(\cdot)$is the nonlinear measurement function that maps the predicted state to the observation space.

The corresponding innovation covariance is given by:

$$\mathrm{S}_k = \mathrm{H}_k\mathrm{P}_{k|k-1}\mathrm{H}_k^{T} + \mathrm{R}_k \tag{9}$$

where $\mathrm{H}_k$is the Jacobian matrix of the measurement function and $\mathrm{R}_k$denotes the measurement noise covariance matrix.

Based on the innovation covariance, the Kalman gain is calculated as:

$$\mathrm{K}_k = \mathrm{P}_{k|k-1}\mathrm{H}_k^{T}(\mathrm{H}_k\mathrm{P}_{k|k-1}\mathrm{H}_k^{T}+\mathrm{R}_k)^{-1} \tag{10}$$

and the state and covariance update equations are expressed as:

$$\hat{\mathrm{X}}_{k|k} = \hat{\mathrm{X}}_{k|k-1} + \mathrm{K}_k\mathrm{e}_k \tag{11}$$

$$\mathrm{P}_{k|k} = (\mathrm{I}-\mathrm{K}_k\mathrm{H}_k)\mathrm{P}_{k|k-1} \tag{12}$$

where $\hat{\mathrm{X}}_{k|k}$and $\mathrm{P}_{k|k}$denote the updated state estimate and covariance matrix, respectively.

3.3.2 Fuzzy Logic Controller Design

In practical dynamic grasping scenarios, the measurement quality of the depth camera and 2D LiDAR may vary significantly due to illumination fluctuations, background clutter, reflective interference, or temporary target occlusion. Under such conditions, a fixed measurement noise covariance matrix $\mathrm{R}_k$may cause the filter to over-trust degraded observations, which can lead to estimation bias or even filter divergence. To address this issue, a fuzzy logic controller (FLC) is introduced to adaptively adjust the measurement noise covariance in real time.

To quantify the consistency between the predicted state and the current observation, the normalized innovation metric is defined as:

$$\gamma_k = \mathrm{e}_k^{T}\mathrm{S}_k^{-1}\mathrm{e}_k \tag{13}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

where e_k is the innovation vector defined in (9), and S_k is the corresponding innovation covariance defined in (10). The scalar γ_k reflects the mismatch between the actual observation and the theoretical prediction. A larger value of γ_k indicates that the current measurement is less consistent with the predicted state, implying a higher likelihood of measurement degradation or abnormal disturbance.

To further characterize the temporal variation of the innovation consistency, the change rate of the mismatch parameter is defined as:

$$\Delta\gamma_k = \gamma_k - \gamma_{k-1} \tag{14}$$

The fuzzy controller takes γ_k and $\Delta\gamma_k$ as two input variables and outputs an adaptive scaling factor α_k. For both input variables and the output variable, three linguistic fuzzy subsets are defined as $\{S$ (Small)$, M$ (Medium)$, L$ (Large)$\}$. Based on the observation consistency and its variation trend, the fuzzy rule base is constructed as shown in Table 1.

Table 1. Fuzzy rule base for adaptive scaling factor $\boldsymbol{\alpha_k}$.

$\boldsymbol{\gamma_k \backslash \Delta\gamma_k}$	S (Small)	M (Medium)	L (Large)
S (Small)	S	S	M
M (Medium)	S	M	L
L (Large)	M	L	L

Using the centroid defuzzification method, the fuzzy inference system produces a crisp adaptive factor α_k, which is then used to update the measurement noise covariance matrix as:

$$R_k^* = \alpha_k R_k \tag{15}$$

The adapted covariance matrix R_k^* is subsequently used in place of the nominal R_k during the measurement update stage. Specifically, the adaptive innovation covariance and Kalman gain can be rewritten as:

$$S_k^* = H_k P_{k|k-1} H_k^T + R_k^* \tag{16}$$

$$K_k^* = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k^*)^{-1} \tag{17}$$

Through this adaptive mechanism, the proposed FA-EKF can automatically reduce the influence of unreliable sensor measurements when observation quality deteriorates, while maintaining high confidence in stable observations under normal conditions. As a result, the filter achieves more robust and accurate target state estimation in dynamic grasping scenarios, thereby providing reliable 3D position inputs for downstream robotic grasp planning and control.

4. Experimental Results and Analysis

4.1 Experimental Setup

To verify the effectiveness of the proposed FA-EKF method, the system was deployed on an AUBO-i5 6-DOF robotic arm. A ZED2i depth camera and a SLAMTEC S2E 2D LiDAR were mounted on the end-effector. The target object was a stable hovering coaxial drone, and the fusion algorithm ran

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

on an industrial PC using ROS Melodic. As shown in Figure 4, both the ZED2i depth camera and the SLAMTEC S2E 2D LiDAR were mounted on the end-effector of the AUBO-i5 robotic arm to ensure synchronized perception during the dynamic grasping process.

Figure 4. Experimental setup showing the AUBO-i5 robotic arm, ZED2i camera, and 2D LiDAR.

4.2 Target Localization Accuracy

We compared the localization error of our FA-EKF algorithm with single-sensor methods and the standard EKF. As shown in Table 2, single sensors exhibit higher errors along certain axes (e.g., the vision sensor has a high Z-axis error of 0.0168m). While the standard EKF improves accuracy, the proposed FA-EKF achieves the lowest Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) across all axes.

Table 2. Localization Error Comparison Under Different Modes

Metric	Axis	Single Camera	Single LiDAR	Standard EKF	Proposed FA-EKF
RMSE (m)	X	0.0113	0.0061	0.0052	**0.0038**
	Y	0.0095	N/A	0.0091	**0.0075**
	Z	0.0168	0.0053	0.0049	**0.0036**
MAE (m)	X	0.0092	0.0050	0.0046	**0.0035**
	Y	0.0084	N/A	0.0075	**0.0068**
	Z	0.0121	0.0045	0.0039	**0.0031**

To test the adaptability of the algorithm to different geometric features, we tested various targets. As shown in Table 3, the RANSAC-based geometric processing combined with FA-EKF performs excellently across different shapes.

Table 3. Detection and Fusion Performance on Different Geometric Targets

Target Type	Grasping Success Rate (%)	RMSE (m)	Processing Delay (ms)
Cylinder (Drone)	**98**	**0.0042**	**8.5**
Cube	93	0.0088	9.2
Polyhedron	87	0.0145	10.1

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

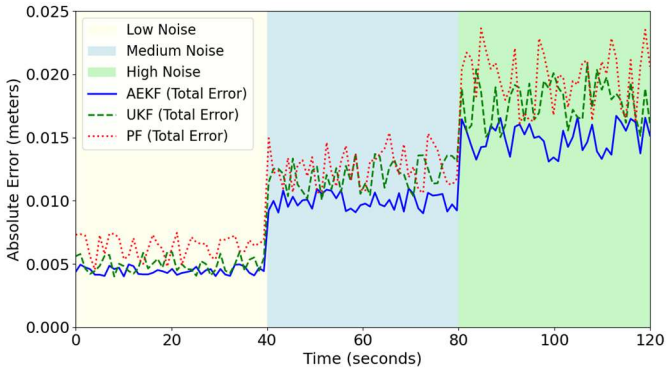

Figure 5. Error convergence and stability comparison under varying noise levels.

As depicted in Figure 5, under high-noise environments, traditional filtering methods experience significant fluctuations, whereas the FA-EKF maintains stable error margins due to its fuzzy adaptation mechanism.

4.3 Robotic Grasping Performance in Complex Environments

To further validate practical applicability, we conducted task-level grasping experiments under Normal Light and Low Light scenarios. We performed 20 grasping attempts for each condition. As shown in Figure 6, the proposed system can successfully track and grasp the hovering target under both normal-light and low-light environments, demonstrating strong practical robustness in real-world task execution.

Figure 6. Successful grasping sequences under varying lighting conditions.

Table 4. Real-world Grasping Success Rates and Real-Time Performance

Method	Grasping Success Rate under Normal Lighting (%)	Grasping Success Rate under Low-Light Conditions (%)	Computing Platform	Average Processing Delay (ms)
DenseFusion (Deep Learning-Based)	95	90	High-End GPU	49.8
Standard EKF	92	87	Embedded CPU	22.0
Proposed FA-EKF	**98**	**96**	**Embedded CPU**	**26.5**

As shown in Table 4, deep learning methods (like DenseFusion) provide high accuracy but suffer from severe delays (~50ms) and require expensive GPUs. The proposed FA-EKF method achieves a superior grasping success rate (96% even in low light) while running efficiently on an embedded CPU with only a 26.5ms delay.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0031

5. Conclusion

In this paper, a Fuzzy Adaptive Extended Kalman Filter (FA-EKF) is proposed for multi-sensor fusion in robotic grasping tasks. By integrating 2D LiDAR data processed via RANSAC with depth camera data enhanced by a CBAM-YOLOv8 network, the system captures highly accurate spatial and semantic information. The designed fuzzy logic controller continuously monitors the innovation sequence to dynamically scale the measurement noise covariance, providing strong robustness against sensor degradation and nonlinear disturbances. Real-world experiments on an AUBO-i5 robotic arm tracking a dynamic drone validate that the proposed framework achieves an RMSE of 0.0038m and a grasping success rate of 96% under complex lighting conditions. The lightweight nature of FA-EKF makes it highly suitable for industrial applications without requiring high-end computing hardware. Future work will extend this framework to multi-agent robotic systems and 3D LiDAR fusion.

References

[1] M. Lin and B. Kim, "Extended particle-aided unscented Kalman filter based on self-driving car localization," Applied Sciences, vol. 10, no. 15, p. 5045, 2020.

[2] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 652-660.

[3] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," Advances in neural information processing systems, vol. 30, 2017.

[4] W.-H. Liao, C.-C. Wang, and W.-C. Lin, "Gnn-based point cloud maps feature extraction and residual feature fusion for 3d object detection," in 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023: IEEE, pp. 7010-7016.

[5] M. Hussain, "Yolov1 to v8: Unveiling each variant–a comprehensive review of yolo," IEEE access, vol. 12, pp. 42816-42833, 2024.

[6] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3-19.

[7] T. Liang et al., "Bevfusion: A simple and robust lidar-camera fusion framework," Advances in Neural Information Processing Systems, vol. 35, pp. 10421-10434, 2022.

[8] X. Bai et al., "Transfusion: Robust lidar-camera fusion for 3d object detection with transformers," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 1090-1099.

[9] H. Fang, M. A. Haile, and Y. Wang, "Robust extended Kalman filtering for systems with measurement outliers," IEEE Transactions on Control Systems Technology, vol. 30, no. 2, pp. 795-802, 2021.

[10] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "An outlier-robust Kalman filter," in 2011 IEEE international conference on robotics and automation, 2011: IEEE, pp. 1551-1558.

[11] M. Sohan, T. Sai Ram, and C. V. Rami Reddy, "A review on yolov8 and its advancements," in International Conference on Data Intelligence and Cognitive Informatics, 2024: Springer, pp. 529-545.

[12] J. Guo, W. Feng, T. Hao, P. Wang, S. Xia, and H. Mao, "Denoising of a multi-station point cloud and 3D modeling accuracy for substation equipment based on statistical outlier removal," in 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), 2020: IEEE, pp. 2793-2797.