Camera-Assisted Drone OCC Ranging System in Low Resolution Point Clouds

Muhammad Alfi Aldolio, Moh Moh Thet Aung, May Thu, Muhammad Fairuz Mummtaz, and Yeong Min Jang
Department of Electronics Engineering, Kookmin University Seoul 02707, Korea
Email: alfialdo@kookmin.ac.kr, mohmohthetaung@kookmin.ac.kr,
maythu@kookmin.ac.kr, muhammadfairuz@kookmin.ac.kr, yjang@kookmin.ac.kr

***Abstract*—Accurate inter-drone ranging is critical for collision avoidance and swarm coordination but remains challenging on lightweight platforms due to the limitations of monocular scale ambiguity and LiDAR sparsity. Traditional sensor fusion methods rely on bounding box detection, which frequently introduces significant error by averaging background points into the depth estimation, particularly when targets are small or distant. To address this, we propose a Camera-Assisted Drone Ranging System that augments low-resolution LiDAR with visual semantic precision. Our framework integrates a robust targetless extrinsic calibration pipeline with an instance segmentation model that acts as a precise spatial filter, isolating foreground drone points from background clutter. Experimental validation demonstrates that this mask-based approach solves the bimodal depth distribution issue inherent in bounding box methods. While standard averaging estimators degrade to errors exceeding 5 meters at a 15-meter range, our proposed median-based segmentation framework maintains a consistent error rate of less than 0.5 meters. These results confirm that high-fidelity ranging is achievable with low-cost, low-resolution sensors through improved semantic filtering.**

***Index Terms*—Drone, LiDAR, Camera, Segmentation, Ranging, 3D Projection, Sensor Fusion**

I. Introduction

As drones become more common in industrial and commercial airspace, the ability for a drone to detect and measure accurate distance to a nearby drone is beneficial for cooperative tasks such as collision avoidance, formation flying, and swarm coordination [1]. A schematic of this inter-drone relative ranging in a multi-UAV scenario is illustrated in Fig. 1, where each drone determines the pairwise distance (d_{ij}) to its neighbors. However, achieving this utilizing only a single sensor is challenging, especially in environments with complex texture or low visibility. Monocular cameras are lightweight and excellent at identifying objects but struggle to estimate the real distance accurately due to scale ambiguity [2]. On the other hand, 3D LiDARs provide low-error distance measurements but often heavy or expensive for drones [3]. Another issue is that lightweight LiDARs have low resolution, leaving large gaps between laser beams that can easily miss a small, floating target like a drone.

The fundamental issue of LiDAR ranging system is the foreground and background ambiguity caused by low resolution point cloud. Integrating camera with LiDAR able to add semantic understanding to the sparse point cloud, usually by performing object detection on the top of the captured image to provide bounding box around the target, and all point clouds inside that box are then averaged to find the distance. However, this approach is problematic for drones due to their irregular shape and the empty space between frame, propellers, etc. The rectangular bounding box inevitably includes background points. Given the sparsity of LiDAR data, the system might capture ten points on the wall and only two on the drone resulting inaccurate distance estimation.

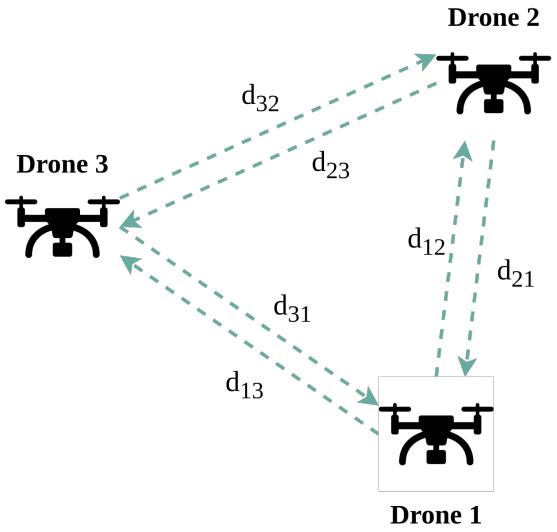

Fig. 1. Inter-drone Relative Ranging System in Multiple Drone Scheme

To address these challenges, this paper proposes a Camera-Assisted Drone Ranging System that leverages visual semantic precision to overcome LiDAR sparsity. This paper present a comprehensive, end-to-end framework to find and tuning extrinsic calibration parameters between LiDAR and camera, performing drone localization in 3D space, projecting 3D point cloud to the 2D image, and finally extracting more accurate distance measurement from the filtered point cloud. Unlike the traditional bounding box detection, The framework incorporate instance segmentation model to generate a pixel-level mask of the target drone as the foreground and background filter mechanism. This ensures even if only a few laser beams hit the target, they are still isolated from the background noise. The successful implementation of this system demonstrates

that robust sensor fusion is achievable using low-resolution LiDAR when augmented by monocular vision. Consequently, the key contributions of this paper are summarized as follows:

1) **Robust Calibration Pipeline**: Outline the necessary steps to perform targetless sensor calibration and ensure accurate projection.
2) **Mask-based Point Cloud Selection**: Demonstrate how replacing bounding boxes with segmentation masks able to improves the performance.
3) **Experimental Validation**: Evaluate the system in real-world setups, providing clear metrics on how this method precisely selecting the corresponding point clouds and extracting the distance values through low-resolution point clouds.

II. Methodology

A. Sensors Integration & Calibration

Fig. 2. Camera and LiDAR sensors mounting.

To ensure optimal field-of-view (FOV) overlap, the monocular camera is rigidly mounted directly on top of LiDAR using a custom 3D-printed bracket as shown in Fig. 2. This vertical displacement minimizes horizontal parallax, ensuring that objects visible to the LiDAR are centrally located in the captured frame. Since the camera and LiDAR operate at different frequencies, synchronization performed through Robot Operating System (ROS) node, matching the closest timestamps between the LiDAR scanned point clouds and the camera frames as shown in Fig. 3(a). This step is critical to prevent data mismatch and errors during the calibration data collection.

The targetless calibration pipeline [4] started with data preprocessing to enhance the distinctiveness. Histogram equalization is applied to both the grayscale camera images and the LiDAR intensity points. This normalization mitigates the impactof varying lightning conditions and ensures edges or textures are statiscally comparable across both modalities. Following the preprocessing step, initial alignment performed by identifying the corresponding features in the 2D image and the 3D point cloud utilizing Super Glue and optimized using Perspective-n-Point (PnP) solver wrapped in a RANSAC loop.

For the final refinement, the Normalized Information Distance (NID) method employed to maximizes the mutual information between the camera and LiDAR. The algorithm voxelizes the point cloud and iteratively adjusts the extrinsic parameters (rotation R and translation t) to minimize the NID metric. This cross-modal optimization allows the system to fine-tune the alignment even in environments with repetitive textures, resulting in a robust extrinsic matrix T_{ext} that accurately maps 3D LiDAR points to the 2D image plane.

B. Visual 2D Drone Segmentation

The visual preception serves as the primary guidance system for the proposed drone ranging system. In the framework specifically Fig. 3(b), this stage processes the capture RGB frames in ROS segmentation node from the synchronization node. The framework integrate segmentation model to generate instance level pixel mask, localizing drone spatial occupany within the 2D image plane, distinguishing the target drone from complex and cluttered environment backgrounds. To achieve this, we utilize lightweight instance segmentation network YOLOv8-seg [5] fine-tuned with custom drone dataset.

The core of this module is the generation of a pixel-level binary mask, denoted as M, rather than a standard bounding box. As depicted in Fig 3(b), the segmentation model classifies every pixel $p_{u,v}$ as either foreground/drone or background. This distinction will assist low-resolution LiDAR fusion by creating a tight silhouette that adheres strictly to the drone's morphology, the segmentation mask effectively removes this negatives space ensuring that the fusion pipeline ignores the background noise that typically confuses standard detection algorithms.

C. 3D LiDAR Point Cloud Projection

In parallel with the segmentatino processs, geometric transformation perfromed to map the 3D spatial information from synchronized LiDAR data onto 2D image plane. As illustrated in Fig. 3(c), the raw point cloud generated by the LiDAR is initially defined in its own local coordinate system, L. To align this data with the visual feed, each point $P_L = [x_l, y_l, z_l, 1]^T$ is first transformed into the camera coordinate system, C, using the extrinsic calibration matrix T_{ext}. This transformation accounts for the rotation (R) and translation (t) offsets between the vertically stacked sensors, mathematically expressed as:

$$P_C = \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = T_{ext} \cdot P_L = \begin{bmatrix} R_{3\times3} & t_{3\times1} \\ 0_{1\times3} & 1 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} \quad (1)$$

Once transformed into the camera frame, the 3D points are projected onto the 2D image plane using the pinhole camera model. This step utilizes the camera's intrinsic matrix K,

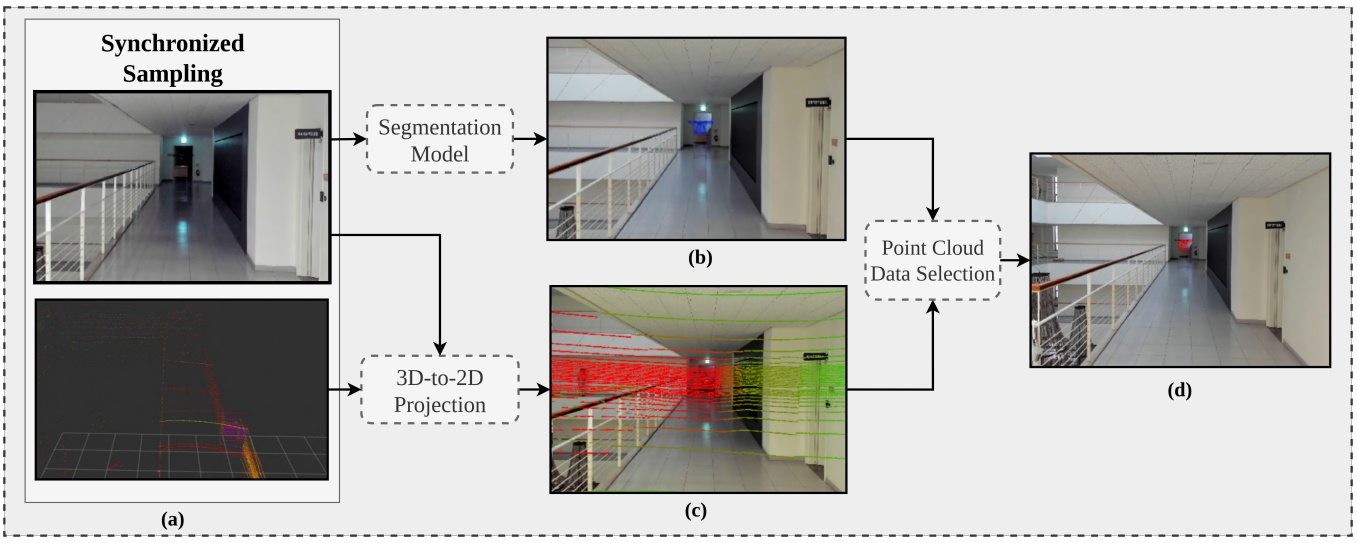

Fig. 3. Proposed framework for Drone Ranging System. (a) Camera frame and LiDAR data sampling synchronization. (b) Drone pixels segmentation in 2D frame. (c) Point cloud projection from 3D-to-2D space. (d) Projected point clouds selection through drone mask.

which encapsulates the focal lengths (f_x, f_y) and the principal point (c_x, c_y). The relationship between a point in the camera frame P_C and its corresponding pixel coordinates (u, v) is given by:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot P_C = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{2}$$

The result of this projection is visualized in Fig. 3(c), where the LiDAR scan lines are overlaid onto the RGB frame. Points that outside the camera's FOV are mathematically clipped and excluded. The remaining points create a sparse depth map registered to the visual features. In the visualization, these points are color-coded by distance (heatmap), illustrating how the 3D structure of the environment such as the building, walls, and the target drone is directly mapped onto the 2D pixel grid.

D. Drone Point Cloud Depth Extraction

The final process are extract projected points data and estimate a robust relative distance to the target drone. Intially, the projection overlays the entire LiDAR scan onto the image, creating a depth map that includes not only the target drone but also the backround and other objects. To isolate the target drone, a logical conjunction performed between the projected points and the binary segmentation mask M generated from the AI model. Let the set of all projected points be denoted as $\mathcal{P}_{proj} = \{(u_i, v_i, z_i)\}_{i=1}^{N}$, where (u_i, v_i) are the pixel coordinates and z_i is the depth. The subset of valid drone points, $\mathcal{Z}_{target}$, is defined as the collection of depth values z_i for which the corresponding pixel coordinates fall within the foreground of the mask:

$$\mathcal{Z}_{target} = \{z_i \mid M(u_i, v_i) = 1\} \tag{3}$$

This masking process transforms the noisy scene from Fig. 3(c) into the filtered and target-specific cluster of point cloud shown in Fig. 3(c). By strictly enforcing this semantic constraint, effectively filtering scanned point cloud from LiDAR. However, due to the low resolution of the LiDAR and the complex geometry of the drone, the set $\mathcal{Z}_{target}$ may still contain outliers especially from slight calibration misalignments [6] that allow background pixels to bleed into the mask edges. Thus, to derive an accurate scalar distance d_{est}, The framework estimate the distance by statiscally extracting the median value of corresponding points in the target drone:

$$d_{est} = \text{median}(\mathcal{Z}_{target}) \tag{4}$$

This statistical filtering ensures that the reported distance remains stable and accurate, effectively representing the drone's visible surface based on point clouds in Fig. 3(d).

III. Experiment and Results

A. Experiment Setup

The experiment and framework evaluation are deployed on Lattepanda Sigma connected with logitech webcam and Velodyne-32 LiDAR on the top of Ubuntu OS. The system mainly utilize ROS environment for integrating sensors, collecting the synchronized data, data processing, and visualization. It also incorporate Pocket AI RTX A500 GPU to process intensive computation such as segmentation model inference, point cloud matrix multiplication and projection, etc. The experiment of drone ranging system then conducted with real hexacopter drone as the target object in the indoor environment.

B. Performance Analysis

The framework performance then evaluated through 2 key metrics: error rate of ranging system and uniformity of filtered

points, with variety of drone positioned at 5m, 10m, and 15m. The error rate metric used to validate overall accuracy of the system and empirically demonstrate the chosen statistical approach is superior compared to other aggregation method. Meanwhile, the uniformity of the final projected points in 2D plane is performed to validate that using a segmentation mask creates a more stable point coulds distribution over bounding boxes.

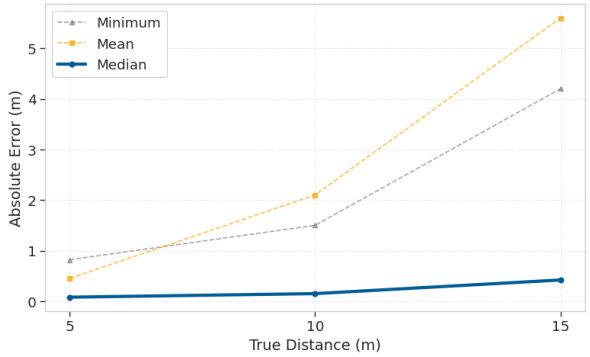

Fig. 4. Comparison of distance absolute error from different aggregation method.

The Fig. 4 presents a quantitative comparison of absolute ranging error across statistical approach (mean, minimum, median) with drone target distances of 5m, 10m, and 15m. The data reveals that the mean estimator degrades severely as the distance increases, reaching an error of over 5 meters at the 15m mark. Similarly, the minimum estimator shows a steady increase in error, likely susceptible(grey line) to sensor noise or misalignments. In contrast, the proposed median estimator demonstrates robustness, maintaining a consistently low error rate (¡0.5m) across all tested distances.

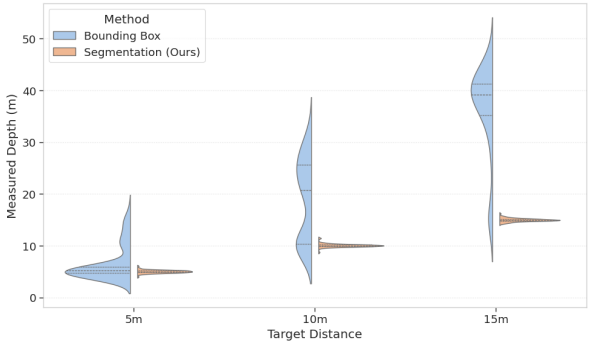

Fig. 5. Uniformity of filtered points based on bounding box and segmentation.

Then Fig. 5 illustrates depth distribution uniformity of the filtered point clouds, contrasting the traditional bounding box method with the proposed segmentation mask. As the target distance increases to 10m and 15m, the Bounding Box method yields a highly dispersed, bimodal distribution where the majority of points align with the distant background rather than the drone. Conversely, the proposed segmentation method consistently produces a tight, unimodal cluster centered precisely at the ground truth distance. This confirms that the semantic mask effectively acts as a spatial filter, isolating valid drone surface.

Conclusion

By replacing traditional bounding box detection with pixel-level instance segmentation, we successfully eliminated the background noise that compromises distance estimation in low-resolution point clouds. Our experimental results confirm that the proposed mask-based spatial filtering, combined with a robust median estimator, ensures high depth uniformity and resilience against sensor outliers. The system achieved better ranging performance (error < 0.5 m) at distances up to 15 meters, significantly outperforming baseline methods which exhibited errors over 5 meters due to background interference. This work demonstrates that semantic-aware sensor fusion can effectively compensate for hardware limitations, offering a scalable solution for precise relative localization in multi-drones networks.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by Korean Government (MSIT) under Grant 2022R1A2C1007884.

References

[1] M. R. Rezaee, N. A. W. A. Hamid, M. Hussin, and Z. A. Zukarnain, "Comprehensive review of drones collision avoidance schemes: Challenges and open issues," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 7, pp. 6397–6426, 2024.

[2] L. Wang, Y. Wang, L. Wang, Y. Zhan, Y. Wang, and H. Lu, "Can scale-consistent monocular depth be learned in a self-supervised scale-invariant manner?" in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 12 707–12 716.

[3] L. Qingqing, Y. Xianjia, J. P. Queralta, and T. Westerlund, "Multi-modal lidar dataset for benchmarking general-purpose localization and mapping algorithms," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 3837–3844.

[4] K. Koide, S. Oishi, M. Yokozuka, and A. Banno, "General, single-shot, target-less, and automatic lidar-camera extrinsic calibration toolbox," 2023.

[5] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[6] Z.-X. Xia, S. Fadadu, Y. Shi, and L. Foucard, "Robust long-range perception against sensor misalignment in autonomous vehicles," 2025.