Proceeding

Offline ISSN 2799-7316
Online ISSN 2287-433X

SMA 2026

The 15th International Conference on Smart Media & Applications

August 10th ~ 12th, 2026

National University of Mongolia, Ulaanbaatar, Mongolia



Smart Media
KOREAN INSTITUTE OF SMART MEDIA

NATIONAL UNIVERSITY OF MONGOLIA

Proceeding

Offline ISSN 2799-7316
Online ISSN 2287-433X

SMA 2026

The 15th International Conference on Smart Media & Applications

August 10th ~ 12th, 2026

National University of Mongolia, Ulaanbaatar, Mongolia

Ikh Surguuliin Gudamj-1, 14201 Ulaanbaatar, Mongolia



Smart Media
KOREAN INSTITUTE OF SMART MEDIA

NATIONAL UNIVERSITY OF MONGOLIA

Oral Session 5

Time	August 11 Tuesday (15:30 ~ 17:00)
Room	E-Library of NUM Room 202
Session Chair	Young Chul Kim, Hongik University

			Page
01	Title	**Analyzing Prefix-Nested Training Schedules for Rank-Heterogeneous Federated LoRA**	110
	Authors	*Anam Nasir, Shivani Kolekar, Dere Roshidat Oluwabukola, Kyungbaek Kim (Chonnam National University)*	
02	Title	**A Study on the Design of Opt-Out Criteria through Analysis of Personal Information Exposure Characteristics in Speech Data**	115
	Authors	*Seo Gene Roh, Jin Gyu Lee, Ja Won Kim (Daejin University)*	
03	Title	**INPUT-OUTPUT NETWORK ANALYSIS OF THE MONGOLIAN ECONOMY: A SECTORAL CENTRALITY APPROACH**	120
	Authors	*GantumurTserenpagma, PurevsurenDalaijargal (National University of Mongolia)*	
04	Title	**Domestic AI Semiconductor-Based Cooperative Multi-Drone Semantic Mapping Framework for Real-Time Reconnaissance Systems**	126
	Authors	*Chanyul Yang (DAON I&C), Janghwan Kim, Chaeyun Seo, R. Young Chul Kim (Hongik University)*	
05	Title	**Mongolian Sentence-Pair Semantic Matching Using Machine-Translated Corpora: A Cross-Lingual Study Based on PAWS and SNLI**	131
	Authors	*Tuvshinbaatar Munkh-Orshikh, Chagnaa Altangerel (National University of Mongolia)*	

Domestic AI Semiconductor-Based Cooperative Multi-Drone Semantic Mapping Framework for Real-Time Reconnaissance Systems

Chanyul Yang
DAON I&C, Republic of Korea
yangcy2000@da-on.com

Janghwan Kim
SE Lab.,Hongik University
Sejong, Republic of Korea
lentoconstante@hongik.ac.kr

Chaeyun Seo
SE Lab., Hongik University
Sejong, Republic of Korea
chaeyun@hongik.ac.kr

R. Young Chul Kim
SE Lab., Hongik University
Sejong, Republic of Korea
bob@hongik.ac.kr

ABSTRACT

Although multi-drone reconnaissance can reduce mission time and expand spatial coverage, conventional systems often rely on high-power GPU platforms or centralized video transmission, which limits power efficiency, heat dissipation, bandwidth, and operational duration. We propose a cooperative multi-drone semantic mapping framework for real-time reconnaissance systems using a domestic AI semiconductor. The proposed framework distributes object detection to each drone via an edge AI board based on a domestic neural processing unit. Then it combines detected classes, confidence scores, timestamps, and GPS- and IMU-derived positions into semantic object observations. A ground control system fuses multi-drone observations, removes duplicated detections, updates object states, and generates an integrated semantic map for situational awareness. The framework is designed around five functional modules: on-board perception, georeferencing, wireless data exchange, collaborative fusion, and semantic map visualization. A benchmark discussion based on a Korean AI semiconductor application demonstration report indicates that a domestic AI board can deliver real-time object detection performance while offering improved power efficiency compared with a compact foreign edge-AI baseline. The proposed approach can be extended to defense reconnaissance, disaster response, public safety monitoring, and infrastructure inspection.

KEYWORDS : Domestic AI semiconductor, multi-drone, semantic mapping, edge AI, real-time reconnaissance, object detection

1 INTRODUCTION

Uncrewed aerial vehicles (UAVs) have become important platforms for reconnaissance, disaster response, smart-city monitoring, and large-area inspection. In time-critical missions, a single drone is limited by battery capacity, field of view, and communication range. A cooperative multi-drone system can divide the search area, observe the same target from different viewpoints, and maintain mission reliability even when one platform is partially occluded or experiences temporary communication loss [1]. However, many existing systems still transmit raw video to a central server or depend on high-performance GPU modules on board. Such approaches increase bandwidth consumption, heat generation, and power demand. They also tend to aggregate detections at the object-list level, without constructing a map that explicitly links object semantics to physical coordinates. Semantic visual SLAM and multi-map SLAM provide useful foundations for robot perception [2,3]. Still, real-time cooperative semantic mapping among multiple small drones remains challenging because each drone has different localization noise, viewing angles, and update rates. Recent edge AI accelerators and neural processing units (NPUs) offer an alternative. Lightweight object detectors such as YOLOv9 can provide real-time detection under constrained computational budgets by improving the flow of gradient information and enabling efficient layer aggregation [4]. Edge-device studies in UAV-swarm environments also emphasize that on-board detection must balance accuracy, throughput, latency, and energy efficiency, rather than maximizing model accuracy alone [5].

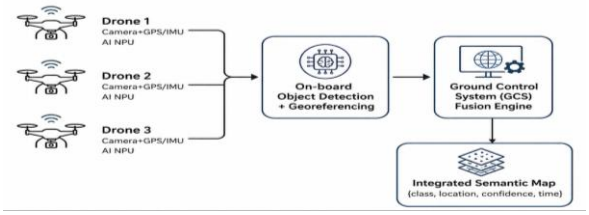


Figure 1: Overall architecture of the proposed domestic AI semiconductor-based cooperative multi-drone semantic mapping framework.

This paper presents a framework that enables each drone to detect objects locally, convert detections into georeferenced semantic observations, and transmit compact object-level messages to a ground control system (GCS). The GCS performs duplicate removal and confidence-aware fusion to create a shared semantic map. The main contributions are: (1) an end-to-end architecture for domestic AI semiconductor-based real-time multi-drone reconnaissance, (2) a semantic object observation model for map-level fusion, and (3) a benchmark-oriented discussion using application demonstration data from a domestic AI semiconductor drone project [6].

2 RELATED WORKS

2.1 Multi-Drone Reconnaissance and Mapping

Multi-robot navigation research has shown that cooperative platforms can improve coverage, resilience, and task allocation in large environments [1]. In UAV reconnaissance, collaboration is especially valuable because aerial platforms must simultaneously handle sensing, localization, communication, and flight control. Nevertheless, many approaches focus on path planning or formation control, while the higher-level problem of creating a shared object-level semantic map is less mature.

2.2 Semantic SLAM and Visual Understanding

ORB-SLAM3 demonstrates robust visual, visual-inertial, and multi-map SLAM across monocular, stereo, and RGB-D configurations [2]. Semantic visual SLAM extends geometric mapping by adding object categories and attributes, enabling robots to reason about the meaning of the environment [3]. Foundation models such as Segment Anything further suggest that segmentation-driven map construction can become more flexible in unseen scenes [7]. For small drones, however, full semantic SLAM is often too computationally expensive, motivating object-level semantic mapping as a practical edge-AI compromise.

2.3 Edge AI Object Detection and AI Semiconductors

YOLO-family detectors are widely used for real-time vision because they jointly predict class and bounding-box information in a single inference pipeline. YOLOv9 introduces programmable gradient information and the GELAN architecture to improve information preservation and parameter utilization [4]. MLPerf Inference provides a general benchmark methodology for comparing inference systems across different hardware and software stacks [8]. In the uploaded demonstration report, a domestic AI semiconductor was applied to the SWARM-X drone platform for edge-AI-based real-time mapping and reconnaissance; the report describes the use of TACHY-BS, a comparison with the Google Edge TPU, and a first-year service focused on multi-drone real-time mapping and object/position recognition [6].

3 PROPOSED FRAMEWORKS

3.1 System Configuration

The proposed framework consists of three layers: the drone layer, the communication layer, and the GCS layer. Each drone carries a camera, flight controller, GPS/IMU sensors, wireless communication module, and domestic AI semiconductor-based edge board. The drone layer performs local inference and generates compact semantic observations. The communication layer exchanges only object-level messages instead of raw video streams. The GCS layer manages duplicate removal, state update, visualization, and operator decision support.

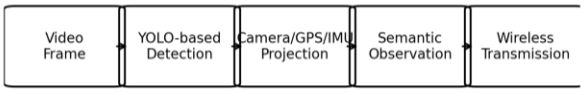


Figure 2: On-board edge AI processing flow from video frames to transmitted semantic observations.

3.2 On-board Perception and Georeferencing

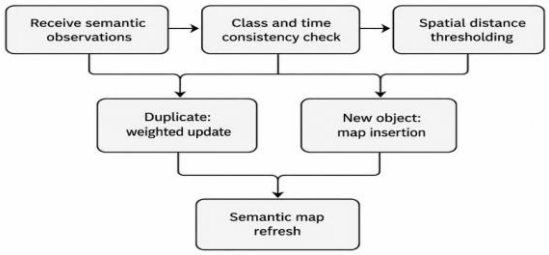
For each input frame, the on-board object detector outputs a class label, bounding box, and confidence score. The drone then combines this detection with GPS, altitude, IMU attitude, camera intrinsic parameters, and timestamp information. The resulting georeferenced observation is intentionally lightweight so that it can be sent through limited-bandwidth wireless links in real time. In the demonstration report, the service scenario includes object search for persons, vehicles, and armored targets, and the first-year objective includes verification of prototype power efficiency, benchmark performance, multi-drone search capability, and detection accuracy [6].

Object ID	temporary or fused ID
Class	person / vehicle / armored object
Position	latitude, longitude, altitude
Confidence	detection confidence score
Timestamp	observation time
Source	drone ID and sensor metadata

Figure 3: Semantic object observation schema used for cooperative map fusion.

3.3 Collaborative Fusion Algorithm

The GCS receives observations from all drones and compares each new observation with existing map entities. Two observations are considered candidates for the same object when their classes match, their timestamps are within a predefined temporal window, and their geodesic distance is below a spatial threshold. If the observation matches an existing entity, the object state is updated using a confidence-weighted position estimate. Otherwise, the observation is inserted as a new entity. This



method reduces duplicated detections and compensates for single-drone blind spots. The association step can also draw on tracking-by-detection methods that combine motion and appearance information to achieve robust object association [9].

Figure 4: Collaborative fusion and duplicate-removal procedure in the ground control system.

Let an observation be represented as o = {c, p, s, t, d}, where c is the class, p is the georeferenced position, s is the confidence score, t is the timestamp, and d is the source drone. A map entity is updated when c matches and distance(p, p_hat) < tau_d within a temporal window tau_t. The updated position is calculated as a weighted average of current and incoming observations, allowing more reliable detections to influence the map more strongly.

4 IMPLEMENTATION AND DISCUSSION

4.1 Dataset and Edge-AI Deployment

The uploaded AI semiconductor application demonstration report states that the project-built hardware and airframes for operating domestic AI semiconductors, completed dataset construction, and prepared a total of 20,600 data items, consisting of 10,300 original images and 10,300 augmented images [6]. The same report identifies the domestic chip as the Tachy BlackSwan 402. It describes it as a low-power SoC-type AI semiconductor with sensor interfaces and real-time image deep learning acceleration features [6]. These characteristics make the platform suitable for drones, as the perception module must continuously process video under strict power constraints.

4.2 Benchmark-Oriented Evaluation

The benchmark table in the report compares a Google Edge TPU baseline and a Deeper-I LeptonTau domestic AI board for real-time multi-drone mapping object detection. The final measurements include 29.50 FPS processing speed, 34.12 ms latency, mAP@0.75 of 0.8685 on the self-collected dataset, mAP@0.5 of 0.5772 on the MS COCO 2017 dataset, and 9.30 FPS/W power efficiency for the domestic AI board. These values indicate that the domestic edge-AI platform can support real-time object detection while improving energy efficiency over the compact foreign baseline in the reported configuration [6].

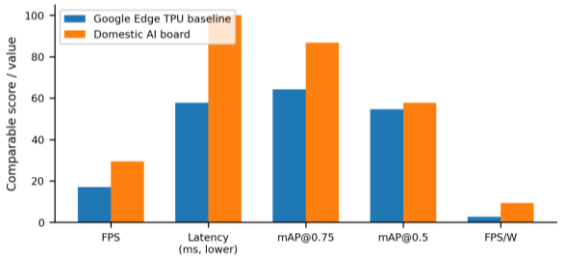


Figure 5: Application benchmark summary based on the reported first-year multi-drone real-time mapping object detection results [6].

4.3 Application Scenarios

The proposed framework can be applied to defense reconnaissance, disaster-response search, wildfire monitoring, infrastructure inspection, and smart-city surveillance. In defense and public-safety environments, the semantic map allows an operator to view detected objects as map entities rather than isolated bounding boxes. In disaster response, object-level map updates can prioritize search zones and reduce the need for repeated inspections. In industrial inspection, the same architecture can detect abnormal objects or vehicles and record their timestamps and spatial coordinates.

4.4 Limitations

Several limitations remain. First, GPS/IMU-based object projection can be inaccurate at low altitudes, in strong winds, or due to poor camera calibration. Second, wireless packet loss may delay map updates. Third, object-level fusion is simpler than full semantic SLAM and may not capture fine-grained geometry. Finally, model performance may vary when the target domain differs from the training data, so continuous dataset expansion and validation are required.

4.5 Mission Scenario and Data Flow

A typical reconnaissance mission begins with area partitioning at the GCS. The operator defines the region of interest and assigns subregions to individual drones. Each drone executes its local flight path while continuously processing camera frames on the domestic AI semiconductor board. When a target candidate is detected, the drone creates a semantic observation and transmits the object-level message. The GCS then compares the incoming observation with the current semantic map and determines whether to create a new entity or update an existing one. This design avoids continuous raw video streaming and keeps communication demand proportional to the number of detected objects rather than the frame rate.

The semantic map can be updated at different temporal resolutions depending on mission constraints. In a high-speed reconnaissance case, the GCS may favor recent observations and rapidly discard stale detections. In a disaster-response case, the system may retain older observations longer because victims, vehicles, and hazards may remain in place. Thus, the same data structure can support both dynamic and semi-static missions by adjusting temporal thresholds and confidence decay functions.

4.6 Localization Reliability

Object localization is one of the most important factors in cooperative semantic mapping. A detection with high visual confidence may still be misleading if the drone attitude, altitude, or camera calibration is inaccurate. The framework therefore separates visual confidence from spatial confidence. The detector produces visual confidence, whereas spatial confidence is estimated from GPS quality, IMU stability, camera angle, altitude, and distance from the optical axis. The fusion module can use both scores to decide whether an observation should update the map or remain as a temporary candidate.

The framework also supports multi-view correction. When two or more drones detect the same object from different directions, the GCS can reduce uncertainty by combining their projected positions. If detections disagree beyond a threshold, the map entity can be marked as unstable and requested for re-observation. This mechanism is useful in urban canyons, forested areas, disaster sites, and military training grounds where occlusion and GPS drift are frequent.

4.7 Communication and Computation Trade-Offs

The main advantage of on-board inference is that each drone can convert raw video into compact semantic messages before transmission. A raw video stream requires continuous bandwidth, whereas an object-level message includes only class, position, confidence, timestamp, and metadata. This can significantly reduce network load when the number of objects is smaller than the number of video frames. However, the approach requires stable on-board inference and careful synchronization between the flight controller, camera, and AI board.

The proposed architecture therefore adopts a hybrid strategy. Mission-critical detections are transmitted immediately, while low-confidence observations can be batched or sent at lower priority. The GCS may request additional evidence from a drone when the map contains conflicting observations. This feedback loop allows the system to allocate communication resources to uncertain regions rather than uniformly transmitting all sensor data.

4.8 Practical Deployment Considerations

For practical deployment, the AI model should be optimized together with the mission profile. A model tuned for high-altitude wide-area surveillance may require stronger small-object detection, while a low-altitude inspection mission may prioritize reduced latency and stable tracking. The domestic AI semiconductor board should also be evaluated under vibration, temperature variations, and sustained flight conditions, as real drones face harsher physical constraints than laboratory edge devices. Operational safety is another key requirement. Even when the system is used only for reconnaissance, automatic detection results should remain decision-support information rather than direct autonomous action. Human operators must be able to inspect map updates, verify uncertain detections, and override drone tasking. Logging of observations, model versions, timestamps, and operator actions is also important for post-mission analysis and accountability.

4.9 Evaluation Metrics

A complete evaluation should include both component-level and mission-level metrics. Component-level metrics include inference frame rate, end-to-end latency, detector mAP, energy efficiency, temperature stability, and wireless packet loss. Mission-level metrics include area coverage ratio, time to first detection, object localization error, duplicate detection rate, semantic map update delay, and operator confirmation time. These metrics are important because a model that performs well in an offline dataset may still fail to improve mission effectiveness if it increases power consumption or delays map updates.

For object localization, the mean geodesic error between the fused object position and the ground-truth position can be used as a metric. For map freshness, the update delay is measured as the time from image capture to visualization on GCS. For cooperative fusion, the duplicate-removal rate can be calculated by comparing the number of repeated observations before and after fusion. These metrics enable the proposed framework to be compared not only with single-drone baselines but also with centralized video analysis and cloud-assisted alternatives.

4.10 Experimental Protocol

A recommended field experiment can be organized in three phases. In the first phase, each drone is evaluated independently to confirm on-board inference speed, communication stability, and object localization accuracy. In the second phase, three drones operate as a single reconnaissance team over a shared area with partially overlapping fields of view. The GCS records the semantic observations, fused map entities, and operator interactions. In the third phase, environmental stress conditions such as low illumination, partial occlusion, increased altitude, and communication loss are introduced to evaluate robustness. Ground truth can be constructed using known target positions, GPS reference points, or manually annotated aerial imagery. The experiment should report not only average accuracy but also failure cases, because reconnaissance systems are sensitive to rare errors. For example, a duplicated target may cause the operator to overestimate the number of objects, while a missing target may reduce mission safety. Therefore, qualitative inspection of the semantic map should be combined with quantitative metrics.

4.11 Security and Data Governance

Because reconnaissance data may include sensitive locations and objects, the system should incorporate basic security and governance mechanisms. Communication between drones and the GCS should be authenticated and, when possible, encrypted. The system should also store model version, dataset version, inference configuration, and map update logs. These records are necessary for reproducibility, post-mission review, and accountability when the system is deployed in public-safety or defense-related contexts.

Data minimization is another benefit of the proposed architecture. Since each drone transmits semantic observations instead of continuous raw video, the amount of potentially sensitive imagery stored or transmitted can be reduced. However, raw video snippets may still be needed to verify important events, so the framework should support selective evidence captured only when the confidence score, an operator request, or a mission rule requires it.

4.12 Comparative Discussion

Compared with a single-drone detector, the proposed framework improves coverage and robustness through distributed observation. Compared with a centralized video analysis system, it reduces bandwidth usage by performing inference at the edge. Compared with a full multi-agent semantic SLAM system, it is simpler and more feasible for small drones because it focuses on object-level semantic entities rather than dense semantic reconstruction. This design choice makes the system suitable for near-term field demonstrations where power, weight, network reliability, and operational simplicity are critical. The uploaded demonstration report provides a useful basis for this design, as it already includes a domestic AI semiconductor, a multi-drone mapping service, a benchmark comparison, and a drone-oriented object detection scenario. The proposed paper's contribution is to organize these components into a general research framework that can be evaluated using clear metrics: detection accuracy, inference speed, latency, power efficiency, duplicate removal rate, object localization error, semantic map update delay, and mission-level coverage.

5 CONCLUSIONS AND FUTURE WORK

This paper proposed a domestic, AI-semiconductor-based cooperative multi-drone semantic mapping framework for real-time reconnaissance. The framework combines on-board object detection, georeferencing, compact semantic communication, collaborative fusion, and integrated map visualization. By processing video at the edge and transmitting semantic observations rather than raw video, the system can reduce communication burden while maintaining operator-level situational awareness. The discussion, based on the domestic AI semiconductor application demonstration report, suggests that domestic edge-AI boards can deliver real-time performance and improved power efficiency in multi-drone mapping scenarios.

Future work will extend the system in four directions. First, outdoor experiments should be conducted under diverse illumination, weather, altitude, and communication conditions. Second, multi-view triangulation and terrain-aware projection should be integrated to improve the accuracy of object localization. Third, distributed fusion among drones should be studied to reduce dependence on the GCS. Fourth, vision-language models and generative AI can be connected to the semantic map for mission summarization, risk explanation, and automatic route recommendation.

ACKNOWLEDGMENTS

This study is the result of a project by the 6th cohort trainees of the Metaverse Convergence SW Academy at Hongik University (Project Number: 202301830004).

REFERENCES

[1] W. Chen, W. Chi, S. Ji, H. Ye, J. Liu, Y. Jia, J. Yu, and J. Cheng, “A survey of autonomous robots and multi-robot navigation: Perception, planning and collaboration. Biomimetic Intelligence and Robotics 5, 2, 100203, 2025

[2] C. Campos, R. Elvira, J. J. Gómez Rodríguez, J. M. M. Montiel, and J. D. Tardos,”ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multi-Map SLAM”, IEEE Transactions on Robotics 37, 6, 1874-1890. 2021.

[3] K. Chen, S. Chen, et al, “Semantic Visual Simultaneous Localization and Mapping: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2025.

[4] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “ YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information”, In Proceedings of ECCV 2024.

[5] D. Meimetis, P. D. Karkazis, and M. S. Papadopoulos, “Comparative Analysis of Object Detection Models for Edge Devices in UAV Swarm Environments”, Machines 13, 8, 684. 2025.

[6] Daon I&C, “First-Year Result Report for the AI Semiconductor Application Demonstration Support Project: Real-Time Multi-Mapping and VTOL Drone Application Demonstration Using a Domestic AI Semiconductor” ,Technical report. 2025.

[7] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollar, and R. Girshick, “Segment Anything in Proceedings of ICCV 2023”, 4015-4026. 2023.

[8] V. J. Reddi, C. Cheng, D. Kanter, P. Mattson, G. Schmuelling, C.-J. Wu, et al. MLPerf Inference Benchmark. In Proceedings of ISCA 2020, 446-459. 2020.

[9] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, “BoT-SORT: Robust Associations Multi-Pedestrian Tracking”, arXiv:2206.14651. 2022.