Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37219

# *Multimodal Perception Fusion for Cooperative Multi-UAV Networks*

**Haoyuan Zhu**

*Faculty of Science and Engineering, University of Nottingham Ningbo, Ningbo, China*
*ssyhz29@nottingham.edu.cn*

***Abstract.*** Cooperative multi-UAV networks extend airborne perception beyond the observations of a single platform to build a networked sensing system composed of multiple nodes, multiple modalities, and mission-oriented operation. Therefore, the performance of the system will also be affected by which fusion algorithm is selected, how sensor payloads are distributed, how information is transmitted in a dynamic aerial network, and how the fusion results are used for cooperative mission planning. The three problems in this paper are sensor resource allocation, multi-source information sharing, and cooperative mission planning. Summarise and compare research on centralised, distributed, and hybrid fusion architectures. Based on the above evidence, centralised architectures are suitable for small teams with stable, high-capacity links; distributed architectures are more robust in environments with weak communication or high dynamism of the mission; and hybrid designs often offer the best engineering trade-off. Simplified simulations are used to show the trends of latency scalability, information gain from sensor allocation, and fusion error under packet loss to illustrate the above trade-offs. The paper proposes the following directions for future research in the end: value-aware communication, semantic-level fusion, online resource reconfiguration, and reproducible experimental benchmarks.

***Keywords:*** multi-UAV cooperation, multimodal perception fusion, sensor resource allocation, distributed estimation

## 1. Introduction

Multi-UAV systems are changing from groups of aircraft that simply fly together at the same time to cooperative perception systems where heterogeneous UAVs are connected by a communication network. For missions such as target search, disaster relief, border patrol, and battlefield reconnaissance, a single UAV is constrained by payload capacity, endurance, field of view, and communication range, etc., and thus cannot offer full-coverage operation for an extended period. A team of UAVs with different sensors can share local observations and cooperate to plan trajectories for large-area coverage efficiently, localisation geometry, and all-weather fault tolerance [1, 2].

Multimodal perception fusion of multiple UAVs is not merely the aggregation of various data streams. Visible-light cameras, infrared sensors, Lidar, and millimeter-wave radar have different output formats, noise characteristics, and spatiotemporal resolutions. Inter-UAV links are also highly mobile and suffer from bandwidth fluctuation, packet loss, and rapid topology change [2]. If all

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37219

platforms need to continuously upload raw data, the central node will become a bottleneck; otherwise, only exchanging low-dimensional decisions will lose the evidence that is later required.

Therefore, the problem is not due to a single deficient algorithm. It is relatively close to the arrangement of sensors, network communication, and mission planning. Therefore, this paper considers multimodal fusion in a multi-UAV system as a closed-loop cooperation problem and investigates the following three research questions. First, the problem of sensor resource allocation: which sensors should be carried by which UAVs under payload, power, bandwidth, and mission-value constraints? The second is multi-source information sharing: should the vehicles transmit raw data, features, states, or semantic summaries, and how to handle asynchronous observations and packet loss? The third is cooperative mission planning: how to feed back the fused estimates into task allocation and trajectory planning?

## 2. Literature sources and analytical framework

The literature search used IEEE Xplore, Web of Science, and Scopus as its principal databases and covered 2013-2026. The main search terms were multi-UAV, cooperative perception, sensor fusion, distributed estimation, task allocation, and ad hoc network. Three Screening Rules Were Used. First, the focus was on reviews and representative methods directly related to multi-UAV communication networks and task allocation and distributed estimation [2-6]. Second, the foundational reviews and representative systems in multimodal fusion and collaborative SLAM were retained [1, 7-9]. Finally, work on path planning and cooperative coverage was added only when it directly supported the analysis of cooperation [6, 10].

The two types of problems are analysis problems and architecture problems. The problem dimensions are sensor resource allocation, multi-source information sharing, and cooperative mission planning; the architectural dimensions are centralised, distributed, and hybrid systems. A 2D graph can show the progress of the individual problems in the different architectures more clearly and identify the technical bottlenecks. It does not suffer from the typical defect of chronological arrangement of studies; that is, it does not list them solely by publication time.

## 3. Advances in sensor resource allocation, information sharing, and cooperative planning

Early multi-UAV missions usually fixed the payload first and planned trajectories around the resulting platform capabilities. Heterogeneous systems have matured to some extent and are now mission-driven. Different platforms can now carry different sensors and adjust their sensing roles during the mission [10].

Resource allocation can be expressed as a combinatorial optimisation problem under platform constraints. Maximize the amount of mission-related information under constraints on energy consumption, bandwidth, and risk. For a fleet of UAVs and a set of sensor types, the allocation should maximise the net team benefit - total information gain minus communication overhead, energy consumption, and safety cost - subject to payload limits, link-capacity constraints, and requirements such as ensuring that every task is covered by at least one vehicle. In other words, a large number of sensors are not necessarily effective; rather, it is necessary to balance the value of the information with the cost of acquiring and transmitting it.

Information Gain is a typical measure, but for a fleet, the concept of complementarity is more appropriate than that for a single sensor. The two sensors observing the same state from near locations will likely have reduced marginal benefits. A sensor providing the semantics of the target and another measuring range or velocity can be used together effectively. Therefore, recent work has

DOI: 10.54254/2755-2721/2026.37219

shifted from selecting the best sensor for each individual node to optimising the observation structure of the entire team [2, 10].

Resource Allocation has gone from greedy heuristics to auction-based algorithms and, more recently, to reinforcement learning. Greedy schemes allocate sensors sequentially according to local information gain. It is computationally simple but may be prone to getting stuck in a local optimum. A Consensus-Based Bundle Algorithm (CBBA) for auctions integrates bidding and consensus to achieve near-optimal allocation through local communication only [3]. Sensor scheduling has also begun to employ multi-agent reinforcement learning. The agent adapts to changes in the mission and communication environment by interacting with the environment through learning. Training convergence and weak interpretability are still serious problems.

### 3.1. Multisource information sharing

The first problem of multisource information sharing is what to share. Existing approaches can be grouped into four levels. Raw-data sharing retains the most information but has a relatively high communication cost. Feature Sharing is the sharing of information and bandwidth. State sharing is suitable for distributed estimation and cooperative tracking. Semantic summaries need to have low bandwidth and be consistent in calibrated confidence and semantics across platforms. Multimodal machine-learning research identifies representation, alignment, fusion, translation, and co-learning as the core technical problems [7], and communication constraints make each of them more difficult in a multi-UAV scenario.

Research on flying ad hoc networks (FANETs) shows why aerial networks cannot be considered typical terrestrial MANETs. Their nodes move rapidly in three dimensions, their topology changes frequently, and their links are strongly affected by separation, attitude, and occlusion [2]. Therefore, the Design of the information-sharing policy should consider the actual capacity of the network. A typical indicator of the value of a message is the expected estimation benefit minus the communication cost; otherwise, no transmission of the message will take place. Therefore, it will no longer be transmitted every time but only when the information is worth the cost.

The first problem of distributed fusion is unknown correlation. If two UAV estimates already contain some of the same evidence, treating them as independent during a later fusion step produces an overconfident covariance. Covariance intersection is a relatively conservative approach that avoids divergence when the correlation is unknown [5]. It is less precise but more consistent, and thus suitable for multi-UAV systems with intermittent links [11]. Recently, data-provenance tracking, factor-graph summaries, and event-triggered exchange have been combined to achieve a finer balance between consistency and communication efficiency [4].

### 3.2. Cooperative mission planning

Cooperative Mission Planning for Multiple UAVs includes target allocation, area partitioning, obstacle avoidance, link maintenance, and dynamic re-planning. However, the goal of multimodal cooperative perception is not just to reach a place. The team needs to move and use its resources efficiently to obtain good information under the constraints of operation [6, 10].

A considerable body of research has been conducted on multi-robot task allocation and multi-UAV planning [3]. CBBA stably distributes tasks through bidding and consensus based on local communication and partial information. In a multi-UAV system, the bid can be extended to represent the sensing value provided by a particular payload. Recent studies have also shown that task allocation in UAV networks needs to consider dynamically arriving tasks, heterogeneous platform

DOI: 10.54254/2755-2721/2026.37219

capabilities, and communication constraints simultaneously [2]. Coverage, search, tracking, and replanning are also moving away from isolated single-objective formulations towards coupled multi-objective optimisation [6].

A small, well-communicating team can use central planning to optimise both the goals and collision avoidance for the mission. The other is the high cost of frequent replanning. Distributed approaches are more adaptable to changes or communication failures in a dynamic environment and are relatively resilient; however, they may not reach the global optimum. Hybrid Strategies address the above deficiencies by dividing the work hierarchically; a central node or cluster head divides the operating area and manages inter-cluster cooperation, while vehicles in each cluster independently plan paths and schedule sensors [12].

## 4. Comparative assessment of three fusion architectures

Centralised, distributed, and hybrid multimodal fusion architectures differ significantly in data flow, computational load, communication overhead, robustness, fusion accuracy, scalability, and suitable operating conditions. Table 1 shows the three architectures for these seven attributes.

Table 1. Comparison of three fusion architectures

| Comparison criterion | Centralized | Distributed | Hybrid |
|---|---|---|---|
| Data flow | Raw data uploaded to the central node | Feature/state exchange among neighbors | Fine exchange within clusters + cross-cluster summaries |
| Computational load | Concentrated at the central node | Shared across nodes | Cluster-head aggregation + central support on demand |
| Communication overhead | Increases rapidly with the number of nodes | Neighbor-only; grows slowly | Intermediate |
| Single-point failure | Center failure causes system-wide failure | No center; highly robust | Switchable cluster heads; partially robust |
| Fusion accuracy | Globally optimal in a rich-communication environment | Locally consistent or near-optimal | Depends on Summary Granularity |
| Scalability | Poor (central bottleneck) | Good (inherently scalable) | Fair (scales by cluster) |
| Suitable setting | Small scale; strong links; strict supervision | Large scale; weak links; highly dynamic missions | Practical Compromise for Most Field Missions |

Centralised Fusion is the most mature of the three architectures. A central node performs data association and state estimation consistently, and generally has high accuracy for small-scale missions. Raw data upload will quickly exhaust the available bandwidth. Distributed fusion is more suitable for the highly dynamic environment of a FANET, and each UAV independently estimates and decides based on its own observations and information from neighbours. As a result, the tolerance for a single-node failure is relatively large, but managing correlations among estimates,

DOI: 10.54254/2755-2721/2026.37219

asynchronous observations, and global optimality is more challenging [4, 5]. Hybrid fusion is often a more feasible engineering compromise. Typical implementations are cluster-based fusion, leader-vehicle fusion, and event-triggered central coordination. The first two problems in their design are how to form clusters and which cluster heads to select.

The evidence does not indicate a simple migration from centralised to distributed fusion. A more probable direction is a hybrid design applied differently at the mission, communication, and fusion layers. The system will collect the initial evidence of the target centrally, distribute the search area to the vehicles, and switch to local autonomy after losing the connection. An architecture that can switch modes, degrade gracefully, and explain its decisions is more suitable for field operations than any fixed design.

## 5. Simplified simulations and trend analysis

Three simplified simulations are used to differentiate the three architectures. Their plots are derived from trend models based on stated assumptions, and thus they show architectural trade-offs rather than measured results for any particular device or algorithm. A Random Geometric Graph is used, and the mean node degree is set to 4. The fleet increases linearly from 3 to 20 UAVs. Each observation packet is 512 bytes; the link bandwidth is 1 Mbps; and the packet loss rate is between 0% and 50%. A weighted-mean model represents fusion. Centralised operation is assumed to have a linearly increasing latency; distributed operation is judged by the number of consensus rounds required.

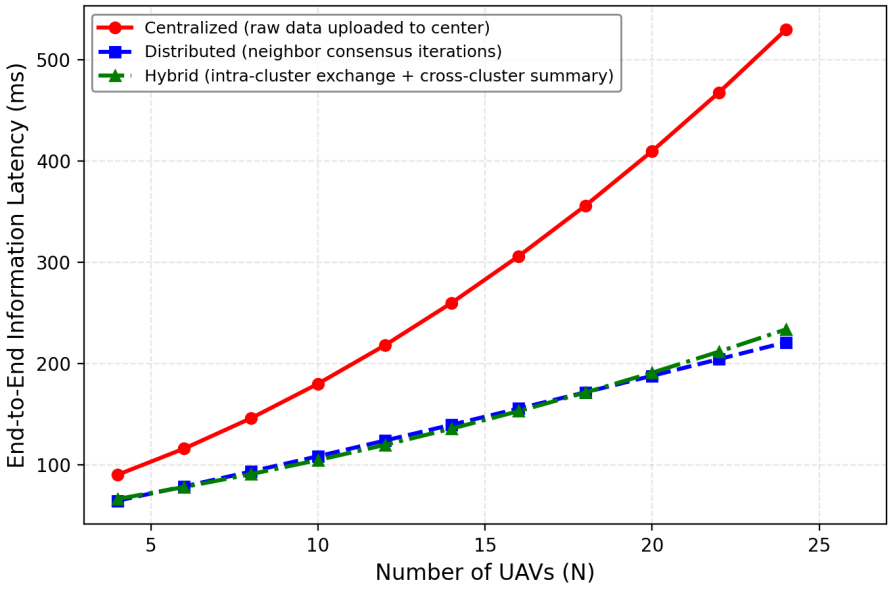



Figure 1. End-to-end information latency versus number of UAVs (photo/picture credit: original)

Figure 1 shows end-to-end latency with increasing fleet size. Centralised fusion has a bottleneck; distributed fusion scales more gently but requires repeated consensus. Hybrid Performance Relies on Clustering.

DOI: 10.54254/2755-2721/2026.37219

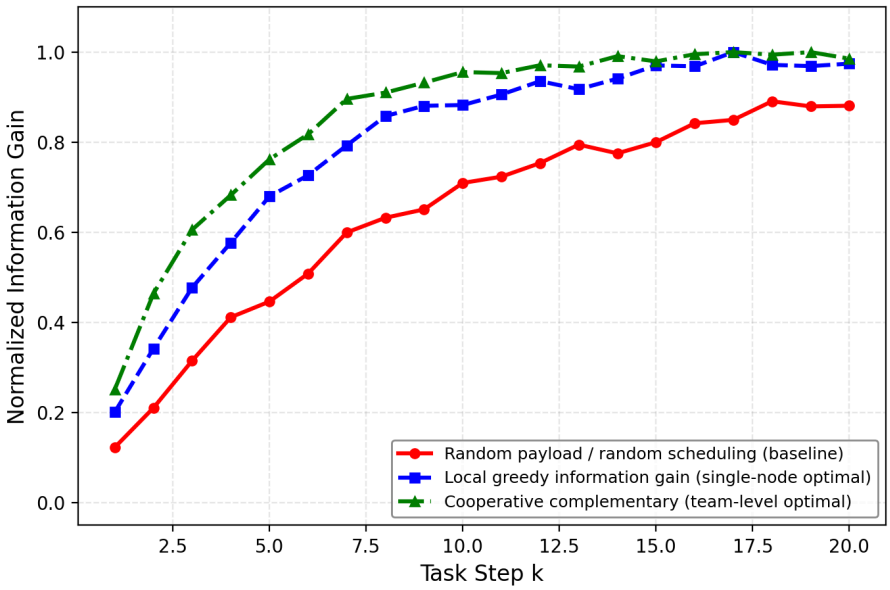



Figure 2. Information gain under three sensor-allocation strategies (photo/picture credit: original)

Figure 2 compares normalized information gain. Random Allocation saturates early; local greediness rapidly reduces uncertainty but may select redundant modalities. Complementarity across platforms offers the highest team-level gain under a cooperative strategy for accounting.

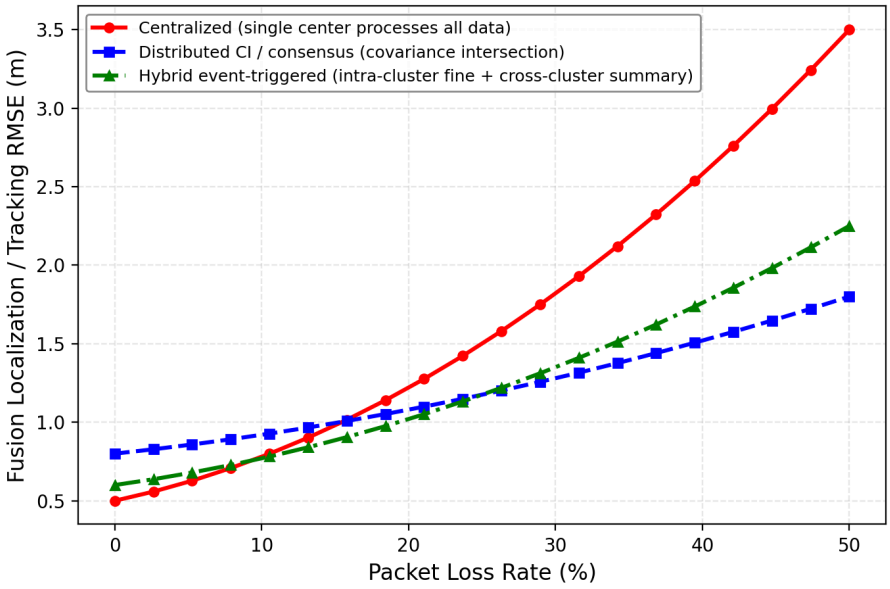



Figure 3. Effect of packet loss on fusion localization/tracking RMSE (photo/picture credit: original)

Figure 3 shows the impact of packet loss on localisation and tracking RMSE after fusion. Centralised fusion has a low loss rate but is very sensitive to degradation of the link with the centre. Distributed fusion deteriorates more gradually as loss increases, while hybrid fusion lies between the two.

DOI: 10.54254/2755-2721/2026.37219

## 6. Research outlook

Several important problems in multimodal fusion for multi-UAV systems remain unresolved. A useful research agenda can be organised progressively, starting with shared experimental foundations, then communication and fusion algorithms, and finally closed-loop system planning. There are no standardised reference values at the experimental level yet. Studies use different simulators, sensor combinations, link models, and mission metrics, so their results are not directly comparable. Future datasets will include raw sensor data, extracted features, log links, vehicle status, and mission labels. They should also provide standardised indicators, such as the mission success rate, information gain per unit bandwidth, estimation accuracy, and replanning delay. At the level of algorithms, the problem is no longer how to transmit more data, but rather what information is useful for the specific purpose. Event-triggered communication, semantic summaries, and age-of-information-aware scheduling are particularly promising. A practical system should vary the granularity of what it shares as the mission progresses: low-cost coverage summaries during search, high-confidence features or selected raw observations during confirmation, and uncertainty and link-state information during replanning. Distributed fusion also needs more work. Covariance intersection handles unknown correlations but is too conservative and loses some information. One approach is to use provenance tracking, factor-graph summaries, local-map boundaries, and message deduplication to ensure that the estimates are consistent in the presence of incomplete communication. Fusion and mission planning should be designed together; otherwise, even if the estimation metric is improved, the mission will not be realised. Many studies at the system level still consider the sensor configuration to be fixed before deployment, even though the value of a sensor changes according to the environment, target state, and communication conditions. Online reconfiguration will be given more weight. When a target moves out of range of search, a link is lost, or a vehicle runs out of power, that vehicle should be reallocated from sensing duties. Cooperative Planners should optimise for other factors than path length or coverage area. They need to consider the uncertainty of the goal, observation geometry, network connectivity, sensor power consumption, and collision risk simultaneously. Cluster formation and cluster-head selection in a hybrid architecture should be planning parameters, not fixed network settings. For high-demand missions, integrated communication, computation, perception, and control will be more necessary than individual enhancements to any single algorithm.

## 7. Conclusion

Multimodal perception fusion in cooperative multi-UAV networks is no longer a problem of data fusion. It is now a large-scale system problem that connects sensor resources, network communication, and mission planning. Sensor allocation is moving away from a fixed-payload model to an online schedule based on information gain and mission value. Multisource sharing is also gradually moving away from free raw-data sharing to value-aware exchange. Cooperative planning now includes setting paths and tasks, and has formed a closed loop jointly constrained by perception uncertainty and communication connectivity. Centralized, distributed, and hybrid architectures therefore suit different operating conditions. The future development will be an integrated communication-perception-planning framework that supports interpretability, graceful degradation, and rigorous validation.

DOI: 10.54254/2755-2721/2026.37219

## References


[1] Khaleghi, B., Khamis, A., Karray, F. O., et al. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information Fusion, 14*(1), 28–44.

[2] Poudel, S., & Moh, S. (2022). Task assignment algorithms for unmanned aerial vehicle networks: A comprehensive survey. *Vehicular Communications, 35*, 100469.

[3] Choi, H. L., Brunet, L., & How, J. P. (2009). Consensus-based decentralized auctions for robust task allocation. *IEEE Transactions on Robotics, 25*(4), 912–926.

[4] Olfati-Saber, R., Fax, J. A., & Murray, R. M. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE, 95*(1), 215–233.

[5] Julier, S. J., & Uhlmann, J. K. (1997). A non-divergent estimation algorithm in the presence of unknown correlations. In *Proceedings of the 1997 American Control Conference* (pp. 2369–2373).

[6] Song, J., Zhao, K., & Liu, Y. (2023). Survey on mission planning of multiple unmanned aerial vehicles. *Aerospace, 10*(3), 208.

[7] Baltrusaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443.

[8] Lajoie, P. Y., Ramtoula, B., Wu, F., et al. (2022). Towards collaborative simultaneous localization and mapping: A survey of the current research landscape. *Field Robotics, 2*, 971–1000.

[9] Lajoie, P. Y., & Beltrame, G. (2024). Swarm-SLAM: Sparse decentralized collaborative simultaneous localization and mapping framework for multi-robot systems. *IEEE Robotics and Automation Letters, 9*(1), 475–482.

[10] Zhang, M., Wu, X., & Li, J. (2023). Integrated design of cooperative area coverage and target tracking with multi-UAV system. *Journal of Intelligent & Robotic Systems, 108*, 35.

[11] Cao, Z., Zhao, Y., & Wu, Z. (2026). Quality-aware distributed state estimation for multi-UAV cooperative localization under communication and navigation constraints. *Drones, 10*(6), 439.

[12] Zhou, X., Wen, X., Wang, Z., et al. (2022). Swarm of micro flying robots in the wild. *Science Robotics, 7*(66), eabm5954.