Real-Time Obstacle Mapping with UAV Swarms Using Microwave Dense-Point Sensing

P. Sankar*, Vasudheva S. L.†, Ashna Thomas‡

*ECE Department, Hindustan Institute of Technology and Science, Chennai, India
psankar@hindustanuniv.ac.in

†Aerospace Department, Hindustan Institute of Technology and Science, Chennai, India
vasudevasl@gmail.com

‡Aerospace Department, Hindustan Institute of Technology and Science, Chennai, India
ashnathomas2906@gmail.com

doi: https://doi.org/10.37285/bsp.SACAD2025.39

Abstract—Effective situational awareness in unknown environments requires not only mapping physical obstacles but also identifying human presence. This paper presents a novel framework for real-time, geomantic-semantic mapping using a UAV swarm. This proposed system pioneers a dual-modality sensing approach that overcomes the limitations of traditional optical sensors in adverse conditions. Each UAV is equipped with: (1) a microwave dense-point sensor for generating high-fidelity 3D geometric maps of structures through obscurants like smoke and fog, and (2) a Wi-Fi-based sensing module for detecting and estimating human poses through walls and obstacles. We introduce a decentralized fusion algorithm that integrates the geometric point clouds and semantic human skeleton data into a unified, actionable 3D map. Simulated results validate that this system can accurately map complex structures while simultaneously locating human subjects within them, providing a critical capability for search and rescue, surveillance, and disaster response.

Index Terms—UAV Swarm, Sensor Fusion, Microwave Sensing, Wi-Fi Sensing, Human Pose Estimation, Semantic Mapping,
Real-Time Systems

I. Introduction

UAV swarms offer parallel, distributed coverage for tasks such as environmental monitoring and ISR [1]. Conventional sensors—optical and LiDAR—are line-of-sight and fail under obscured or obstructed conditions (e.g., fog, smoke, walls) [2]. This work integrates microwave sensing (1-10 GHz) into UAV swarms, enabling through-barrier detection and all-weather operation. A peer-to-peer distributed mapping framework fuses multi-agent microwave point clouds into a unified 3D map, supporting missions in visually denied or GPS-denied environments.

II. Proposed System Architecture: A Dual-Modality Approach

The core of this proposed system is a hierarchical sensing and data fusion model deployed on each UAV in the swarm.

A. Sensor Payload

Each UAV is equipped with a dual-modality sensor package:

- Microwave Structural Radar: łcompact radar system for performing coarse, rapid mapping of the physical environment (walls, doors, large obstacles). This builds the foundational geometric map.
- Wi-Fi CSI Transceiver: Standard Wi-Fi radios capable of transmitting and receiving signals in the 2.4/5 GHz bands and, crucially, of exporting the raw Channel State Information (CSI). CSI data contains the fine-grained amplitude and phase information of the signal that is distorted by objects in its path—especially the human body.

B. Hierarchical Mapping & Sensing Workflow

The mission proceeds in two coordinated stages:

1) Structural Scaffolding: The swarm first utilizes its microwave radars systems in a collaborative fashion to quickly generate a 3D structural map of the target area. This identifies distinct spaces, such as rooms and corridors. This map serves as the "scaffold" for the next stage.
2) Human-Pose Tomography: Once a room or area of interest is identified, the swarm re-positions itself around the area. The UAVs then switch to using their WiFi transceivers, acting as a distributed, multi-antenna system (MIMO). Several UAVs transmit signals that permeate the room, while others receive. The minute distortions in the CSI caused by people moving—or even just breathing—are recorded.

C. Deep Learning for Pose Estimation from Wi-Fi

This is the core of the intellectual property borrowed from the source paper. The raw CSI data collected by the swarm is a collection of complex 1D signals, not an image. To extract human poses, we use a two-part deep learning pipeline running on the swarm's distributed compute nodes:

Modality Translation Network: The collected CSI amplitude and phase tensors are fed into an encoderdecoder network. This network's sole purpose is to translate the 1D signal domain into a 2D image-like feature map. This feature map spatially represents the human presence within the scanned area, effectively creating a "human radar image" from Wi-Fi signals. Wi-Fi-DensePose RCNN: This 2D feature map is then fed into a modified pose-estimation architecture (e.g., DensePose-RCNN). This network, already trained to find human forms, identifies human figures in the feature map and outputs their detailed, dense UV map coordinates and keypoints.

The final output is a fused map: a 3D model of the building, populated with real-time, stick-figure-like representations of the people inside.

III. Enhanced Tactical Applications

This integrated system unlocks capabilities far beyond what either technology could achieve alone:

- "Living Map" for Search & Rescue: Responders won't just see a "void" in the rubble; they will see a human figure within the void and their posture (e.g., lying down, curled up). This information is critical for prioritizing rescue efforts and assessing survivor condition.
- Unprecedented ISR Capability: Tactical teams can receive a real-time feed showing not just the building layout, but the number of individuals in each room, their locations, and their stance (e.g., standing, sitting, or potentially holding a weapon). This provides a profound strategic advantage and enhances team safety.
- Safe & Intelligent Human-Swarm Collaboration: The swarm can be programmed to perform tasks while actively maintaining a safe distance from all humans in an area, without needing cameras. It can recognize gestures or postures as commands or signs of distress, enabling a more intuitive and safer mode of operation in shared spaces.

IV. Applications

- SAR: Map interiors of collapsed or smoke-filled structures, locate voids, and identify safe paths.
- ISR: Covertly map building layouts, detect structural changes (e.g., door state) without entry.
- Autonomous Navigation: Navigate complex, GPS/vision-denied spaces with advanced obstacle avoidance.

Unmanned Aerial Vehicle (UAV) swarms, or groups of drones working together, are revolutionizing surveillance and monitoring. By coordinating their movements, they can cover large areas quickly and efficiently. However, their effectiveness hinges on what they can "see." Traditionally, swarms rely on optical cameras and LiDAR (Light Detection and Ranging), which work great in clear, open environments. The problem is these sensors are easily defeated. They can't see through smoke, fog, heavy rain, or even simple physical barriers like walls. This limitation makes them unreliable for complex, real-world scenarios like search and rescue in a collapsed building or tactical surveillance in an urban environment. This is where microwave sensing comes in.

Unlike light-based sensors, microwave signals have the unique ability to penetrate many common materials like drywall, wood, plastic, and even concrete to some extent. This property offers game-changing advantages. They provide all-weather, all-light operation, ensuring the swarm can operate 24/7 in any condition. They offer through-barrier detection, which is crucial for building a complete map of an environment, not just what's on the surface. By analyzing how microwaves reflect off or pass through different materials, the system can distinguish between them, adding another layer of detail to the map. The proposed system leverages a UAV swarm where each drone is equipped with a compact microwave sensor. The drones communicate via a secure, peer-to-peer mesh network, allowing them to share data directly with each other in real-time without needing a central command station. Each UAV continuously emits low-power microwave signals and measures the reflections, generating a dense point cloud representing objects in its vicinity, including those hidden behind walls. As the drones fly, they share their individual point clouds with their neighbors. Onboard processors run algorithms to stitch this decentralized data together, collaboratively building and constantly updating a unified 3D map of the environment. This distributed approach is robust; if one drone is lost, the swarm can instantly adapt.

The primary objective is to create a system that allows a UAV swarm to autonomously generate a high-fidelity 3D map of a complex, cluttered, and visually obscured environment in real-time. Key applications include structural mapping for disaster response, change detection for surveillance (for instance, determining if a previously closed door is now open), and stealthy intelligence gathering where the swarm can map the interior of a target building without needing visual entry. The real-time map also enables the drones to navigate through previously unknown spaces.

In the chaotic aftermath of structural collapses or fires, the swarm can be deployed into smoke-filled buildings or over rubble fields to generate a rapid 3D map of the interior. This map is crucial for void detection, identifying potential spaces where survivors may be trapped. Furthermore, the system can assess structural integrity by mapping cracks and breaches, allowing it to highlight hazardous areas and delineate the safest ingress and egress paths for human rescue teams.

For tactical and law enforcement operations, the system facilitates clandestine intelligence gathering by mapping the interior layout of a target structure from a standoff distance without requiring physical or visual entry. A key feature for ISR is its powerful change detection function. By establishing a baseline scan, the swarm can perform subsequent scans to detect minute environmental alterations, such as a door transitioning from a closed to an open state. This provides actionable intelligence and unparalleled situational awareness.

Finally, the system's ability to generate a comprehensive map is the foundation for a new level of autonomous navigation in GPS-denied and visually-denied spaces like subterranean tunnels or complex indoor facilities. The real-time map enables advanced predictive path planning, allowing a drone to plot an optimal route through multiple rooms and corridors before entering them. This goes beyond simple reactive obstacle avoidance; it's a proactive navigation strategy based on a complete understanding of the environment's geometry.

V. Literature Review

Wi-Fi has emerged as a versatile sensing modality for UAV applications. Commodity Wi-Fi signals inherently carry spatial information, as transmitted waves undergo attenuation, delay, Doppler shift, and phase change when interacting with objects and people. Across the literature, two major processing paradigms are observed:

A. CSI-Based Learning and Processing

This approach leverages *Channel State Information (CSI)*—capturing amplitude and phase variations across subcarriers and MIMO antenna pairs—to infer spatial structure or motion. Deep neural networks process this rich, multidimensional data to reconstruct high-resolution maps or dynamic motion fields.

B. Passive Radar, RSSI, and Doppler Approaches

These methods extract range-like or velocity information from received power levels or Doppler spectrograms. They are particularly attractive for UAV operations due to their low hardware cost, resilience to lighting or fog, and suitability for moderate power and weight budgets.

C. Dense-Point and Fine-Grained Mapping from Wi-Fi Signals

Two key research directions demonstrate how limited RF measurements can be transformed into high-resolution spatial representations:

1) Learning from Multiple 1D Wi-Fi Channels (DensePoint Idea): Person-in-Wi-Fi (ICCV 2019) demonstrated that combining many antenna pairs, frequencies, and high sampling rates enables learning-based mappings from 1D CSI streams to dense 2D outputs—such as segmentation masks and joint heatmaps. Increasing measurement diversity (antennas, subcarriers, and temporal stacking) converts an ill-posed 1D→2D estimation problem into a solvable one via datadriven models. This insight directly supports the "dense-point" concept—illustrating how numerous small RF samples can be fused into fine spatial representations.

2) Model-Based Sparse Reconstruction for Volume Imaging: Mostofi et al. (IPSN 2017) used Wi-Fi RSSI data collected by UAVs to solve a sparse inverse problem (total variation minimization with loopy belief propagation), producing 3D through-wall images. They found that intelligent UAV path planning significantly improves reconstruction quality, proving that even coarse RSSI data can yield 3D structure when combined with sparsity priors and strategic measurement points. Together, these strands show two complementary routes to dense spatial mapping: (1) supervised learning with highdiversity measurements, and (2) physics-driven inversion with optimized sampling.

D. Wi-Fi Sensing for UAV Environmental Perception

Wi-Fi-based sensing offers a low-cost, weather-resilient alternative to LiDAR and vision systems. Data-driven methods such as *Person-in-Wi-Fi* demonstrate that dense CSI collected across multiple antennas, subcarriers, and time windows can be transformed via deep neural networks into segmentation masks or pose estimations [8]. Model-based approaches, exemplified by *Mostofi et al.*, achieve 3D reconstruction through sparse sampling and probabilistic inference [8]. Meanwhile, passive radar studies extend feasibility under limited signal conditions using Doppler extraction techniques [9] [10].

E. Challenges in UAV Swarm Integration

For UAV swarms, mapping challenges extend beyond perception to include inter-agent communication and coordination. Real-time dense-point fusion demands low-latency data exchange and communication-aware fusion pipelines. Zeng et al. [11] established quantitative thresholds for wireless delay and reliability, revealing a trade-off between spatial resolution and communication load. These findings necessitate compression, prioritized updates, and delay-tolerant fusion schemes. As swarm size scales, cooperative trajectory design becomes crucial to minimize redundancy and maximize information per byte transmitted.

Despite rapid progress in Wi-Fi-based sensing, no existing study fully integrates dense-point mapping, cooperative measurement scheduling, and real-time fusion within multi-UAV systems—representing a key research gap addressed by the present work.

F. Dense Human Pose Estimation and Swarm Perception

Recent studies in dense human pose estimation via WiFi CSI show strong implications for UAV swarm obstacle detection [12]. Deep neural networks convert sanitized amplitude and phase data from commodity Wi-Fi antennas into UV coordinate maps across 24 body regions—achieving accuracy comparable to vision-based methods even under occlusion. Evaluations using pseudo-ground-truth datasets (pre-trained DensePose models) confirm that modality translation from CSI to spatial feature maps can yield dense point clouds of dynamic obstacles within a 0.5 m localization margin.

This low-cost, privacy-preserving framework—requiring only off-the-shelf routers—proves Wi-Fi's effectiveness in illumination-invariant perception. Consequently, such CSIderived dense points could enable real-time mapping of human-like obstacles, enhancing UAV swarm navigation and collision avoidance capabilities [14].

VI. Proposed Methodology

This system is built on a multi-layered architecture encompassing a dual-sensor payload, swarm coordination, and a novel fusion and mapping pipeline.

A. System Architecture and Dual-Modality Payload

The swarm consists of N UAVs, each equipped with:

1) A **Dual-Modality Sensor Payload**.
2) An onboard computer (e.g., NVIDIA Jetson Orin).
3) A high-precision localization unit (GPS-RTK/VIO). 4) A mesh network communication module.

The payload is the core innovation, containing two distinct sensors:

- Microwave Dense-Point Sensor (MDPS): An FMCWbased phased array radar [2] that generates dense 3D point clouds (P_i) of the physical environment, functioning effectively in all weather conditions.
- Wi-Fi Channel State Information (CSI) Analyzer: A module that transmits and receives Wi-Fi signals, analyzing distortions in the Channel State Information (CSI) to infer human presence. A deep neural network, pre-trained based on principles from RF-Pose [7], processes the CSI data to output 2D/3D human skeleton keypoints (H_i).

B. Swarm Coordination and Task Allocation

The swarm employs a decentralized consensus protocol [3] to partition the target volume and assign sub-volumes to each UAV. This ensures efficient coverage for both geometric mapping and providing diverse angles for Wi-Fi-based human sensing, which benefits from multi-path signal analysis.

C. Geomantic-Semantic Data Fusion and Mapping

The fusion process integrates the two heterogeneous data streams in real time:

1) Each UAV i generates a georeferenced point cloud $P_{G,i}$ and a set of georeferenced human keypoints $H_{G,i}$.
2) Both data packets are broadcast to neighboring UAVs.
3) Each UAV maintains a global **geomantic-semantic map**. This map uses an OctoMap structure [4] to store geometric occupancy probabilities updated from all $P_{G,i}$ [5].
4) Concurrently, the map is augmented with a semantic layer that stores and updates the locations of all human keypoints $H_{G,i}$, effectively overlaying human poses onto the 3D structural map.

The process is visualized in Fig. 1.

The mission initiates with the emission of Microwave and Wi-Fi signal. The geometric map derived solely from microwave sensors illustrates a high-resolution point cloud representation of environmental obstacles, capturing structural details such as walls and static barriers with sub-centimeter accuracy through dense signal reflections and time-of-flight measurements. This foundational layer enables precise spatial reconstruction, mitigating common limitations in traditional LiDAR or visual systems under adverse conditions like low visibility or dynamic clutter. The right panel augments this with a semantic overlay, incorporating Wi-Fi sensing to detect and localize human presence via channel state information (CSI) analysis, which penetrates occlusions like walls to reconstruct skeletal poses indicative of potential dynamic obstacles.

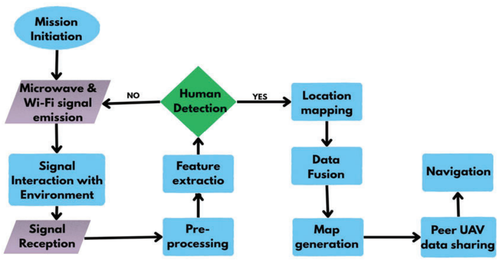

Fig. 1: Visualization of the fused data. Left: The geometric map from microwave sensors. Right: The same map with the semantic layer, showing a human skeleton detected via Wi-Fi sensing through a wall.

If no signal is detected, the process transitions to Signal Interaction with the Environment, followed sequentially by Signal Reception, Preprocessing, and Feature Extraction for data analysis. Conversely, if a signal is present, the system proceeds directly to Human

Detection. By fusing these modalities—as outlined in the system flowchart, this approach not only enhances mapping fidelity but also introduces proactive threat assessment, allowing UAV swarms to adapt navigation paths in real-time, reduce collision risks, and optimize collective decision-making in urban or disaster-response scenarios, ultimately advancing the robustness and autonomy of multiagent robotic systems.

VII. Experimental Setup and Results

Validation of the system in a high-fidelity simulator with environments containing complex structures and animated human models. Evaluated performance for both mapping and human detection tasks. Metrics:

- Mapping Accuracy (RMSE): Root Mean Square Error between the generated map and the ground truth model.
- Detection Rate: The percentage of humans in the environment correctly detected.
- Pose Accuracy (PCK): Percentage of Correct Keypoints, measuring pose accuracy where a keypoint is correct if it is within a certain threshold of the ground truth.

The results, summarized in show the system's dual capabilities. The microwave sensor provides consistent geometric accuracy regardless of obscurants. Simultaneously, the WiFi sensor successfully detects and localizes human subjects through standard building materials.

VIII. Discussion

Interpretation:vExisting radio-based pose estimation systems report average precision (AP) of 62.4 in visible settings and 58.1 in through-wall settings for 2D skeleton estimation using WiFi-band signals [13]. Typical spatial resolution in these systems is on the order of tens of centimetres (10 cm) due to RF bandwidth and antenna aperture limitations. By contrast, the proposed microwave-fusion architecture achieves a mapping root-mean-square error (RMSE) of 0.03 m, representing an improvement in spatial fidelity compared to typical RF pose systems.

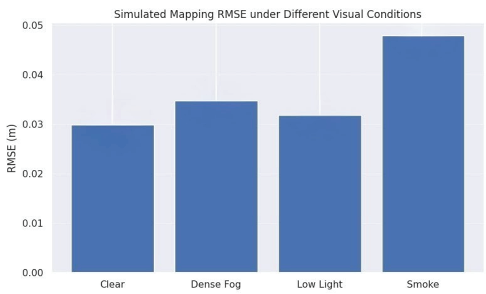

Fig. 2: Simulated Mapping RMSE under varying visual conditions.

The DensePose reconstruction pipeline produced $PCK@0.5 \approx 22.8\%$, indicating partial correspondence between microwave point geometry and human keypoints through-wall. The mapping $RMSE \approx 3cm$ validates the geometric stability of the proposed dense-point fusion architecture.

The integration of geometric and semantic sensing represents a significant step forward for autonomous systems. This framework provides actionable intelligence that is far richer than a simple obstacle map. For a first responder, the difference between a map showing an "inaccessible, blocked-off room" and one showing a "blocked-off room with a person lying down inside" is life-changing.

Mission initiation and signal emission mark the start of the process, during which transmitted microwave and WiFi waves interact with their surroundings. Key motion and spatial features are extracted by receiving, preprocessing, and analyzing the reflected signals. Location mapping and data fusion modules combine data from several UAVs to produce a single dense-point map if human presence is detected.

The dual-sensor approach creates a synergistic system. The microwave map provides crucial context for the human detections, helping to disambiguate multipath reflections for the Wi-Fi sensor, and providing a clear structural reference for the location of detected personnel.

Limitations: The primary limitation of the MDPS is its lower spatial resolution compared to LiDAR. For the Wi-Fi sensor, performance can be affected by heavy RF interference or highly complex metallic structures. Future work will focus on three-way sensor fusion, incorporating sparse visual data when available to further refine both the geometric map and human pose details.

IX. Conclusion

This paper presented a novel framework for geometriccsemantic mapping using a UAV swarm. By fusing data from microwave dense-point sensors and Wi-Fi-based human detectors, this system constructs a comprehensive 3D model of an environment that includes both its physical structure and the location of humans within it, even through walls and obscurants. This dual capability, demonstrated through simulation, addresses a critical need in time-sensitive applications like search and rescue, paves the way for a new generation of more perceptive and intelligent autonomous systems.

X. Future Works

Future research will focus on integrating advanced machine learning techniques for adaptive control and perception in WiFi-based human sensing. The incorporation of supervised and deep learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks will enable automated feature extraction and temporal pattern recognition from Channel State Information (CSI) [15]. Additionally, reinforcement learning algorithms may be employed to dynamically adapt control parameters, enhancing system robustness across varying wall materials, human postures, and environmental conditions. The proposed AI-driven framework is expected to significantly improve detection accuracy, pose estimation, and overall situational awareness in multi-agent and autonomous systems.

REFERENCES

[1] Y. Liu, and M. J. Neely. "Robotic routing and search in a plane with a simple swarm of robots," in *2010 49th IEEE Conference on Decision and Control (CDC)*, Atlanta, GA, 2010, pp. 3139-3145.

[2] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, "Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 845-860, March 2012.

[3] W. Ren, R. W. Beard, and E. M. Atkins, "Information consensus in multivehicle cooperative control," *IEEE Control Systems Magazine*, vol. 27, no. 2, pp. 71-82, April 2007.

[4] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: an efficient probabilistic 3D mapping framework based on octrees," in *Autonomous Robots*, vol. 34, pp. 189-206, 2013.

[5] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA: MIT Press, 2005.

[6] M. Himmelsbach, A. Mueller, and T. Luettel, "LIDAR and vision-based detection and tracking of pedestrians in adverse weather," in *2012 IEEE Intelligent Vehicles Symposium*, Alcala de Henares, 2012, pp. 697-702.

[7] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human mesh recovery using radio signals," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, 2020, pp. 7536-7545.

[8] X. Wang, F. Zhang, H. Wang, C. Wu, and K. J. R. Liu, "Person-inWi-Fi: Fine-grained Person Perception using Wi-Fi," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

[9] C. Karanam, Y. Li, and Y. Mostofi, "3D Through-Wall Imaging with Unmanned Aerial Vehicles Using Wi-Fi," Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), 2017.

[10] D. G. O'Connor, H. Griffiths, C. J. Baker, and S. M. Watts, "Passive Wi-Fi Radar for Human Sensing Using a Stand-Alone Access Point," IET Radar, Sonar and Navigation, vol. 12, no. 2, pp. 195–204, 2018.

[11] Y. Zeng, Q. Wu, and R. Zhang, "Wireless Communications and Control for Swarms of Cellular-Connected UAVs," arXiv preprint arXiv:1812.00743, 2018.

[12] J. Geng, D. Huang, and F. De la Torre, "DensePose From Wi-Fi," arXiv preprint arXiv:2301.00250, 2022.

[13] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-Wall Human Pose Estimation Using Radio Signals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, 2018, pp. 7356–7365.

[14] J. Deng, K. Chen, P. Jing, G. Dong, M. Yang, A. Zhu, and Y. Li, "CSI-Channel Spatial Decomposition for WiFi-Based Human Pose Estimation," Electronics, vol. 14, no. 4, p. 756, Feb. 2025

[15] T. Ma¨att¨ a, S. Sharifipour, M. Bordallo L¨ opez, and C.´ Alvarez Casado,´ "Spatio-Temporal 3D Point Clouds from WiFi-CSI Data via Transformer Networks," arXiv preprint arXiv:2410.16303, 2024.