

Intuitive Gesture Control Interface for Swarm of Aerial Robots

Utkarsh Gupta*

FEAS

Cranfield University

Bedfordshire, United Kingdom

utkarsh.gupta@cranfield.ac.uk

Gilbert Tang*

FEAS

Cranfield University

Bedfordshire, United Kingdom

g.tang@cranfield.ac.uk

Phil Webb

FEAS

Cranfield University

Bedfordshire, United Kingdom

p.f.webb@cranfield.ac.uk

*Corresponding author

*Corresponding author

Abstract—In recent years, research in Unmanned Aerial Vehicles (UAV) swarm technology has seen accelerated progress across various fields, including disaster response, environmental monitoring, precision agriculture, surveillance, military, last-mile delivery, and search-and-rescue operations, due to their promising potential in solving challenging problems. The ability to rapidly access locations that are difficult to reach for humans and other robots has made UAVs unique. While certain tasks can be accomplished by a single drone, a fleet of drones can perform the same tasks more efficiently, reducing task completion time and with enhanced capabilities. Human Swarm Interaction (HSI) is an important consideration in the system deployment of drone swarms, although it is needed primarily for information exchange and control for some applications, the controllability and intuitiveness of drone swarms command and control are crucial for any applications that involve human intervention. This paper explores human swarm interactions through the development of a novel gesture-based control framework, and its demonstration for efficient aerial swarm management. A custom Py-thon-based framework is designed and developed, leveraging Tensor-Flow Lite and MediaPipe for real-time hand gesture recognition, enabling precise and efficient drone swarm control.

Keywords— *Human Swarm Interaction, Gesture Control, Human Robot Interface.*

I. INTRODUCTION

Unmanned Aerial Vehicles (UAV) are widely applied across different domains including disaster response, environmental monitoring, precision agriculture, surveillance, military, last-mile delivery, and search-and-rescue operations, due to their promising potential in solving challenging problems. With increasing demand, extensive research is being conducted, particularly in aerial swarming, where multiple drones operate simultaneously. Aerial swarms offer advantages in large-area coverage, making them valuable in emergency response and surveillance. However, controlling a swarm is far more complex than operating a single drone, as no standardized control method exists.

This research presents an intuitive, gesture-based control framework for drone swarms, minimizing hardware dependencies while enhancing user accessibility. By leveraging

hand gestures which is a natural form of human communication, the system enables intuitive swarm control without requiring extensive technical expertise. The framework integrates a graphical user interface (GUI) with a customizable gesture recognition module, allowing adaptability to different user needs. Additionally, it provides a streamlined approach for connecting multiple drones, ensuring scalability and seamless operation. To validate its effectiveness, the system undergoes rigorous testing, culminating in a physical demonstration of stable and efficient swarm control. The framework is designed to support the simultaneous control of multiple drones and has been successfully tested with up to five DJI Tello drones via a private network. Built using open-source libraries such as TensorFlow Lite and MediaPipe, it ensures robust gesture recognition. The system is scalable, allowing future expansions to larger drone fleets with minimal modifications. Custom hand gestures were defined and trained using a neural network-based model, ensuring accurate and responsive control.

II. RELATED WORK

Human-Swarm Interaction (HSI) has emerged as a critical research area, particularly in developing intuitive and efficient control mechanisms for aerial robot swarms. Traditional methods rely on centralized control systems, predefined flight paths, or joystick-based interfaces, which often lack real-time adaptability and scalability. To address these challenges, gesture-based control has gained significant attention due to its potential to provide a more natural and intuitive way of interacting with drone swarms.

Several studies have explored gesture-based HSI systems. Podevijn et al. (2014) investigated direct human control of robot swarms using gesturing at subswarms, demonstrating the feasibility of high-level command inputs through gestures. Similarly, Nagi et al. (2014) proposed a system where spatial gestures enabled operators to communicate with drone swarms, improving responsiveness in dynamic environments. However, these studies primarily focused on proof-of-concept implementations with limited scalability and hardware constraints. [[3], [4]]

More recent advancements in gesture-based control have introduced deep-learning-driven gesture recognition. Serpiva et

al. (2021) presented DronePaint, a swarm light painting system utilizing deep neural network (DNN)-based gesture recognition for artistic and interactive applications. This study highlighted the potential of gesture-based HSI in creative domains but did not focus on swarm-level coordination and command precision. Muzzinoglu and Karakose (2020) proposed a wearable glove-based system for human-UAV interaction, which improved accuracy in gesture recognition. However, reliance on specialized wearable devices limited its accessibility and practicality for general users. [[5], [6]]

Other research efforts have explored advanced control modalities, such as brain-computer interfaces (BCIs). Kim et al. (2021) developed a system for controlling drone swarm formations using electroencephalogram (EEG) signals, enabling mind-controlled interactions. While innovative, this approach required extensive user training and expensive hardware, making it less feasible for widespread adoption. [7]

A major limitation in existing gesture-controlled swarm systems is their dependence on specific hardware. Tang and Webb (2018) introduced an ergonomic, contact-less gesture control system for industrial robots using Leap Motion sensors, improving usability but restricting adaptability to drones. In a related study, Tang, Asif, and Webb (2015) combined static pose recognition with dynamic hand motion tracking, providing precise command execution. However, both approaches were tailored for industrial robotic arms rather than aerial swarms. [[1],[2]]

Recent studies have emphasized the need for scalable and hardware-independent gesture control systems. Liu et al. (2024) introduced HGIC, a hand gesture-based interactive control system designed for multi-UAV operations, achieving high accuracy without external wearables. Suresh and Martinez (2018) explored formation control using gesture interpreters, laying the foundation for gesture-driven swarm formations. These advancements underscore the growing demand for robust,

real-time gesture control frameworks that reduce sensor dependencies, enhance scalability, and improve usability. [[9], [10]].

III. FRAMEWORK OVERVIEW

The proposed framework integrates hardware and software components to enable seamless gesture-based control of a drone swarm. It consists of a hand tracking and gesture recognition module, a graphical user interface (GUI), and a communication mechanism for swarm control, which is illustrated below in Figure 1 as architecture diagram of the framework.

For hardware, a fleet of five DJI Tello Edu drones was used, selected for their vision-based positioning system and Python-compatible SDK. A TP-Link Archer C50 Wi-Fi router established a dedicated network in Access Point mode, ensuring stable drone communication. The processing system ran on a high-performance laptop with a multi-core processor, dedicated GPU, and sufficient RAM to ensure real-time processing and seamless gesture recognition.

A custom desktop application was developed to offer an intuitive interface for swarm control. The GUI features real-time video streaming, interactive buttons, a gestures guide, recognized gesture and derived drone command, allowing users to configure parameters, monitor gesture recognition, and execute commands. Screen-shots provided below in Figure 2 and Figure 3, illustrates the GUI screens of the desktop application.

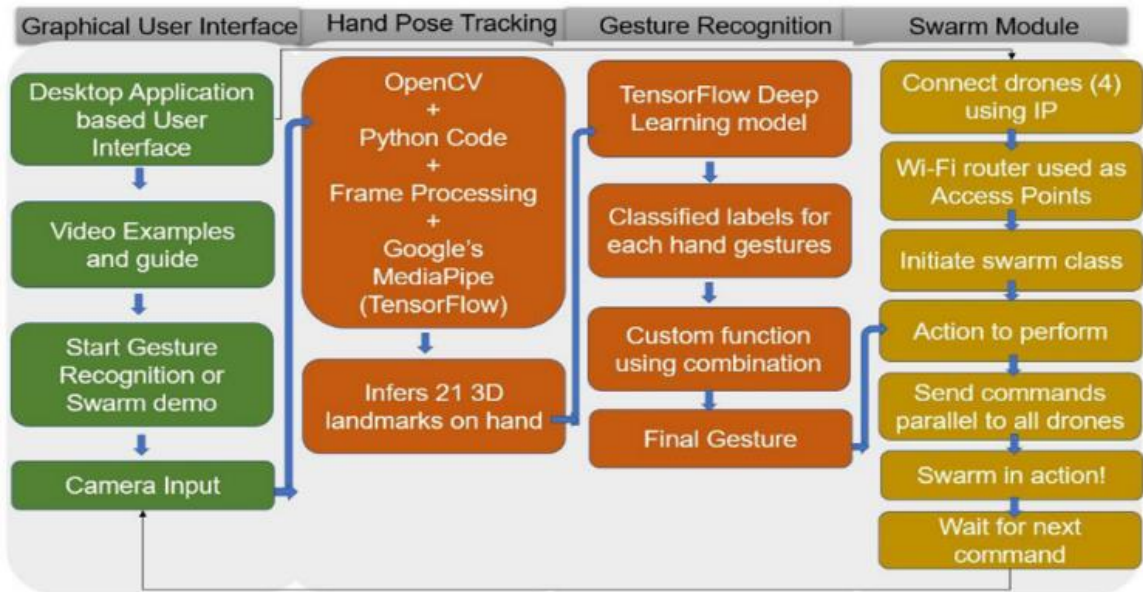


Figure 1: Framework Architecture

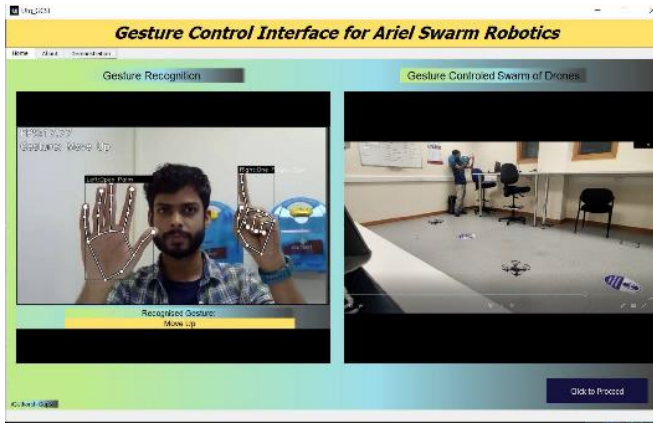


Figure 2: Desktop Application Home Page

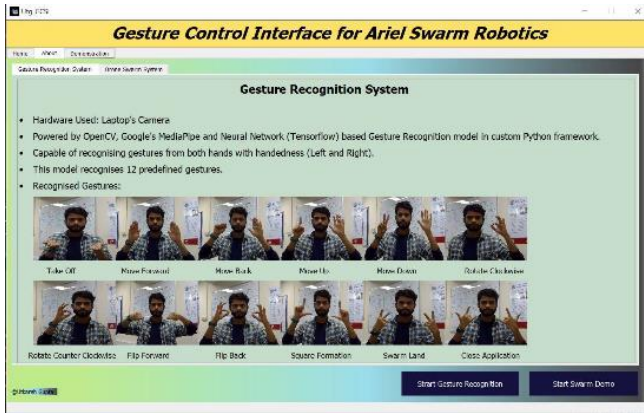


Figure 3: About Page - Gesture Recognition Tab

Hand pose tracking was implemented using MediaPipe Hands, an open-source library by Google that detects 21 three-dimensional (3D) landmarks on the hand. The system utilizes a two-stage machine learning pipeline: a Palm Detection Model, which identifies hand presence, and a Hand Landmark Model, which extracts high-precision 3D key points for motion tracking. Real-time feedback ensures smooth interaction, though lighting conditions and occlusions may impact detection accuracy as depicted in Figure 4 and Figure 5.



Figure 4: Both hands are not detected as one hand is occluded by other hand.



Figure 5: Both hands are detected properly, and landmarks can be seen accordingly.

The gesture recognition module processes the detected hand landmarks and classifies them into eight predefined gestures (e.g., Open Palm, Flat Palm, Close Fist, Two Finger Open). To extend capabilities, a custom function interprets twelve composite gestures using left and right-hand combinations (e.g., Take Off, Move Forward, Flip Backward, Swarm Land). These commands are then transmitted to the drone swarm for execution.

Swarm communication is managed over a Wi-Fi-based network, with each drone operating in Access Point mode. Commands are sent using the User Datagram Protocol (UDP), and the laptop retrieves drone IP addresses to establish a connection. Network ping tests are used to verify connectivity before swarm execution.

During operation, the swarm operator faces the camera while ensuring a stable connection between the laptop and drones. Recognized gestures are processed and transmitted either concurrently or sequentially, allowing synchronized swarm movement. The system demonstrates the feasibility of real-time, gesture-based drone swarm control, offering a scalable approach for future enhancements.

IV. EXPERIMENTATION, RESULTS AND DISCUSSIONS

To evaluate the effectiveness and reliability of the proposed gesture control interface for swarm drones, two key experiments one to assess systems scalability by varying the numbers of drones and the other one to test the accuracy of recognized gestures were conducted.

The results on the impact of swarm size and task execution time indicate that in-creasing the number of drones increases latency in connecting to the swarm but does not significantly affect flight control commands. The TABLE I. illustrates the variations in swarm size and time taken to complete a task, showing that the system re-mains efficient and responsive as swarm size grows.

TABLE I. DATA COMPARISON FOR SWARM DRONES' VARIATIONS

Time consumed for given task	Number of drones		
	1 drone	2 drones	3 drones
connect to swarm	0.11149	0.126029	0.251055
take off	6.237072	8.744017	8.177699
move forward	2.394539	2.646319	2.630836
move up	2.524	2.618543	2.746213
move down	2.850156	2.841621	2.508138
move back	2.510539	2.399204	3.398748
rotate clockwise	2.196545	2.283308	2.195876
rotate counterclockwise	2.08133	2.401639	2.945644
swarm land	3.852303	3.820147	3.724363
swarm end	0	0	0.000805
full test	24.868221	28.022858	28.688841

At the core of the framework is a gesture recognition system employing a land-mark-first approach. Hand pose data, represented by 21 key landmarks extracted via the MediaPipe library, is processed using a lightweight neural network built with TensorFlow Lite and Keras. This approach significantly reduces computational complexity compared to pixel-based methods and mitigates issues stemming from variable lighting, camera quality, and background noise. The model, trained on 6,268 gesture sample results in approximately 96-100% accuracy across gestures combinations under different conditions as shown in the TABLE II. below. These results confirm the robustness of the hand-tracking model, ensuring reliable and precise control of the swarm.

TABLE II. DATA FOR DIFFERENT GESTURE ACCURACIES

Gestures	Accuracy for Left Hand	Accuracy Right Hand
Flat palm	71	88
Open palm	99	98
1 finger open	99	99
2 fingers open	99	99
Pinch	98	100
Close fist	99	99
Thumb and index	98	98
Rock and roll	99	96

The experimental results substantiate that the custom gesture control interface for aerial robots is scalable, robust, and computationally efficient. This section examines the system's performance, emphasizing key insights and challenges. The inter-face effectively coordinates multiple drones with minimal latency and achieves a gesture recognition accuracy exceeding 96% under diverse conditions.

In another comparative experiment which was conducted to evaluate voice input vs. gesture control. While voice commands worked well in a quiet setting, they failed in real-world conditions due to drone motor noise interfering with speech recognition. In contrast, gesture-based control remained highly effective regardless of environmental noise, reinforcing its suitability for swarm operation.

V. CONCLUSION

This research presents a novel, fully customizable gesture control framework for aerial robots, developed through extensive research and the integration of open-source libraries. The design emphasizes a user-friendly graphical interface and an efficient gesture recognition system, reducing the dependency on highly skilled operators and enabling smooth, flexible control of drone swarms.

A comprehensive physical demonstration was conducted in front of over 30 participants, validating the system's real-time performance, robustness, and ease of use under practical conditions. This demonstration not only showcased the framework's ability to deliver precise swarm control but also underscored its potential to outperform traditional input methods such as voice control.

REFERENCES

- [1] Tang, G., & Webb, P. (2018), "The design and evaluation of an ergonomic contactless gesture control system for industrial robots", *Journal of Robotics*, 2018, Article ID 9791286. <https://doi.org/10.1155/2018/9791286>
- [2] Tang, G., Asif, S., & Webb, P. (2015), "The integration of contactless static pose recognition and dynamic hand motion tracking control system for industrial human and robot collaboration", *Industrial Robot: An International Journal*, 42(5), 416-428. <https://doi.org/10.1108/IR-03-2015-0059>
- [3] PODEVIJN, G., O'Grady, R., Nashed, Y. S. G., & Dorigo, M. (2014), "Gesturing at subswarms: Towards direct human control of robot swarms", *Lecture Notes in Computer Science*, 2014.
- [4] Nagi, J., Giusti, A., Gambardella, L. M., & Di Caro, G. A. (2014), "Human-swarm interaction using spatial gestures", *IEEE International Conference on Intelligent Robots and Systems*, 2014.
- [5] Serpiva, V., Karmanova, E., Fedoseev, A., Perminov, S., & Tsetserukou, D. (2021), "DronePaint: Swarm light painting with DNN-based gesture recognition", *SIGGRAPH 2021*.
- [6] Muzzinoglu, T., & Karakose, M. (2020), "Wearable glove-based approach for human-UAV interaction", *IEEE International Symposium on Systems Engineering*.
- [7] Kim, S. J., Kwon, B. H., & Jeong, J. H. (2021), "Intuitive visual imagery decoding for drone swarm formation control from EEG signals", *IEEE International Winter Conference on Brain-Computer Interface (BCI)*, 2021.
- [8] Macchini, M., De Matteis, L., Schiano, F., & Floreano, D. (2021), "Personalized human-swarm interaction through hand motion", *IEEE Robotics and Automation Letters*, 2021.
- [9] Liu, R., et al. (2024), "HGIC: A hand gesture-based interactive control system for efficient and scalable multi-UAV operations", *arXiv preprint arXiv:2403.05478*.
- [10] Suresh, A., & Martinez, S. (2018), "Gesture-based human-swarm interactions for formation control using interpreters", *arXiv preprint arXiv:1804.08676*.



DSPACE

<https://dspace.org/>

Intuitive gesture control interface for swarm of aerial robots

Gupta, Utkarsh; Tang, Gilbert; Webb, Phil

2025-12-04

Institute of Electrical and Electronics Engineers (IEEE)

<https://doi.org/10.1109/iccr67607.2025.11372070>

Gupta U, Tang G, Webb P. (2025) Intuitive gesture control interface for swarm of aerial robots. In: Proceedings of the 2025 7th International Conference on Control and Robotics (ICCR), 4-6 Dec 2025, Kyoto, Japan, pp. 292-295