Intuitive Human-Swarm Interaction with Gesture Recognition and Machine Learning

Golan Khen
University of Maryland - College Park
College of Computer, Mathematical, and Natural Sciences
College Park, Maryland, USA
gkhen@umd.edu

Detim Zhao
Arizona State University - Polytechnic Campus
Ira A. Fulton Schools of Engineering
Mesa, Arizona, USA
dzhao44@asu.edu

José Baca
Texas A&M University - Corpus Christi
Department of Engineering
Corpus Christi, Texas, USA
jose.baca@tamucc.edu

ABSTRACT

Unmanned Aerial Systems (UAS), commonly known as drones, have revolutionized various industries with their diverse applications. As the demand for seamless and intuitive drone control grows, researchers are exploring innovative approaches to improve human-swarm interaction. This paper presents a novel method for operating a swarm of drones in real time using wearable technology and machine learning. Through the integration of motion capture data and classification algorithms, we strive to achieve an intuitive level of control that is accessible to users with varying skill levels. While the full realization of this approach remains a work in progress, our research lays the groundwork for future endeavors in this domain. In this paper, we discuss the limitations of existing control methods and present our methodology for data preprocessing, model training and testing, and result analysis. Our findings indicate the potential of this approach and open avenues for refining the interaction between humans and drone swarms.

CCS CONCEPTS

• **Computing methodologies → Supervised learning by classification**; *Motion capture*; • **Human-centered computing →** *Gestural input*.

KEYWORDS

Unmanned Aerial Systems (UAS), drones, human-swarm interaction, wearable technology, machine learning, motion capture, classification algorithms, intuitive control, real-time control, gesture recognition.

ACM Reference Format:
Golan Khen, Detim Zhao, and José Baca. 2023. Intuitive Human-Swarm Interaction with Gesture Recognition and Machine Learning. In *The Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '23), October 23–26, 2023, Washington, DC, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3565287.3617621

MobiHoc '23, October 23–26, 2023, Washington, DC, USA

ACM ISBN 978-1-4503-9926-5/23/10…$15.00
https://doi.org/10.1145/3565287.3617621

1 INTRODUCTION

Unmanned Aerial Systems (UAS), also known as drones, have been transforming our world. Hence, methods of controlling them have evolved and developed over the years, with recent studies aiming to find more intuitive and seamless ways to control, as the standard remote controller requires lots of skill, practice and precision to operate effectively. These methods have mainly included the use of Electroencephalography (EEG), drone cameras, and wearable technologies. In this study, we aim to develop an intuitive and seamless method of operating a swarm of drones in real-time. It will be easy and simple for any individual to control a swarm or singular drone, even without prior knowledge or experience. While we were not able to completely accomplish our goal, we have laid the foundation for future work to do so. In the rest of the paper, we will discuss in detail other current methods of controlling drones, our methodology, our conclusions, and future discussion.

2 LITERATURE REVIEW

In this section, we will go in-depth into the several methods being researched to control drones.

2.1 EEGs

EEGs have enabled drone control through Brain-Computer Interfaces (BCIs). BCIs interpret brain signals into commands, enabling drone manipulation in simulations and the real world [2, 8, 10, 20]. However, complexity is constrained; studies have rudimentary commands and can only manage a singular drone [2, 20] or the entire swarm [10].

2.2 Drone Camera

Cameras have been a popular method to control drones, particularly with machine learning (ML). The user's hand gestures prompt the computer to capture the hand image and subsequently identify and respond accordingly [18]. Another use of ML is Deep Learning and Neural Networks. In [18] the authors utilized the YOLOv3 algorithm and Deep Neural Networks (DNN) with computer vision to recognize hand gestures, and then to control the drone. Although the use of cameras is successful in controlling the drone, it has several constraints that hinder its real-world practicality, such as the need for a clear line of sight between the operator and reliable lighting.

2.3 Wearable Technologies

There are many kinds of wearable technologies that have been used to control drones. Some use glove(s) or a sleeve with sensors, and others use different motion controllers such as smart watches.

2.3.1 Gloves and Sleeves. Using gloves and sleeves has been a common method for controlling drones. These are equipped with sensors that process movement data and interpret certain motions as certain commands. They all follow the general framework of receiving raw data from sensors on the glove(s) or sleeve, then data is processed to identify which command should be performed [3, 7, 9, 11, 13, 15–17]. Many of the gloves utilize flex sensors to better detect hand movements [7, 9, 16, 17], and ML is often incorporated in these methods in order to more accurately identify the correct command to perform. Some common ML algorithms used are Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Naive Bayes (NB) [11, 17]. Out of these four algorithms, SVM and KNN performed the best, with SVM having a slight edge (around 2%-3% higher in accuracy, precision, recall, and F-score) over KNN [11, 17]. However, these studies all have very specific predefined gestures or motions for every command, making operating the drones more of a memory exercise than an intuitive one.

2.3.2 Smaller Devices. Another popular method is to utilize motion controllers with sensors in some way to control drones [1, 4, 6, 12, 14, 19]. Usually, these devices, such as smartwatches, control the drone by recognizing the user's motions [1, 6, 12]. The wearable device obtains the user's data, then utilizes a computer or algorithm to process and identify the command. A commonly-used sensor is the Inertial Measurement Unit (IMU), as it allows the computer to accurately obtain movement data, including rotation [3, 6, 11, 15]. The Leap Motion Controller is another commonly-used controller that uses its sensors to track the hand [19]. These motion controllers are advantageous in their compact and convenient form factor, but they are limited to fewer and more precise motions. These studies were tested in simulated environments and real life. However, most lacked the intuitive experience that we are aiming for. We are going to utilize ML classification algorithms in order to develop an intuitive and seamless experience to operate a swarm of drones using wearable technology.

3 METHODOLOGY

In this section, we will go into detail about the methodology followed in this project.

3.1 Preprocess Data

We utilized two existing .csv files containing motion capture data, both recorded using a Vicon optical motion capture camera system. The first dataset was provided by Dr. José Baca and the other was acquired from an online repository [5]. The first dataset consisted of 8 gesture classes: Dab Left, Dab Right, Go Forward, Go Left, Go Right, Land, Stop, and Up. The second dataset consisted of 5 gesture classes: Go Forward, Land, Point, Stop, and Up. The first dataset contained data from a single person, whereas the second dataset contained data from 14 individuals who performed every gesture. Additionally, the first dataset had 5,096 instances whereas the second dataset had 78,095 instances. For both datasets, we dropped irrelevant and missing values, ignored irrelevant nodes (e.g. feet, ankles, knees, etc.), and split the data into training and testing groups.

3.2 Train and Test Classification Models

For the first dataset, we split the data frame-by-frame for each gesture and organized it into training and testing sets, ensuring effective model training and testing. With the second dataset, the data was partitioned based on each user. To have a fair and effective split, we chose to include nine users in the training set and the remaining five in the testing set. We trained several different machine-learning models: KNN, SVM, Random Forest (RF), Naive Bayes, and an Ensemble. Using the scikit-learn Python library, we implemented, tested, and obtained metrics and confusion matrices for each model.

3.3 Analyze Test Results

For each model, we generated a confusion matrix alongside a classification report of the precision, recall, and f1 score for each gesture, and overall accuracy for the model. From these results, we fine-tuned the parameters for each model. For KNN, we found the optimal number of neighbors for each dataset. This value was eight for the first dataset and four for the second dataset. For SVM, we found the optimal kernel for each dataset to be the Radial Basis Function. For Random Forest, we found the optimal number of trees in the forest of each dataset to be 100. Finally, for Naive Bayes, we found the optimal variant of the algorithm for each dataset to be the Gaussian variant.

4 RESULTS

After separately applying each dataset to each of the classification models, the model attempts to classify the gesture based on the dataset's labels. In this section, we will show our finalized confusion matrices and metrics for each model of each dataset. For the first dataset, Naive Bayes had the highest accuracy of 84% compared to the other models. The f1 score for the Dab Right class in each classification model was 0%, possibly indicating the model's overfitting due to consistent confusion of Dab Right with Dab Left. Go Forward and Land gestures occurred to be the most classified in a majority of the classification models of the first dataset. For the second dataset, the Ensemble model had the highest accuracy of 74% compared to the other models. The Up gesture occurred to be the most classified in a majority of the classification models of the second dataset. All results of the first dataset can be seen in Figure 1, and all results of the second dataset can be seen in Figure 2.

5 CONCLUSIONS

Drones, or UAS, have become an integral part of various industries, and finding intuitive ways to control them is essential for their widespread adoption. Existing methods of humans controlling drones include EEG, drone cameras, and wearable technologies. However, each method comes with its limitations as discussed. To address these limitations, we introduced a novel approach utilizing machine learning classification algorithms with motion capture data. Although we were not able to complete our objective, we have provided the foundational work to approach the ongoing efforts

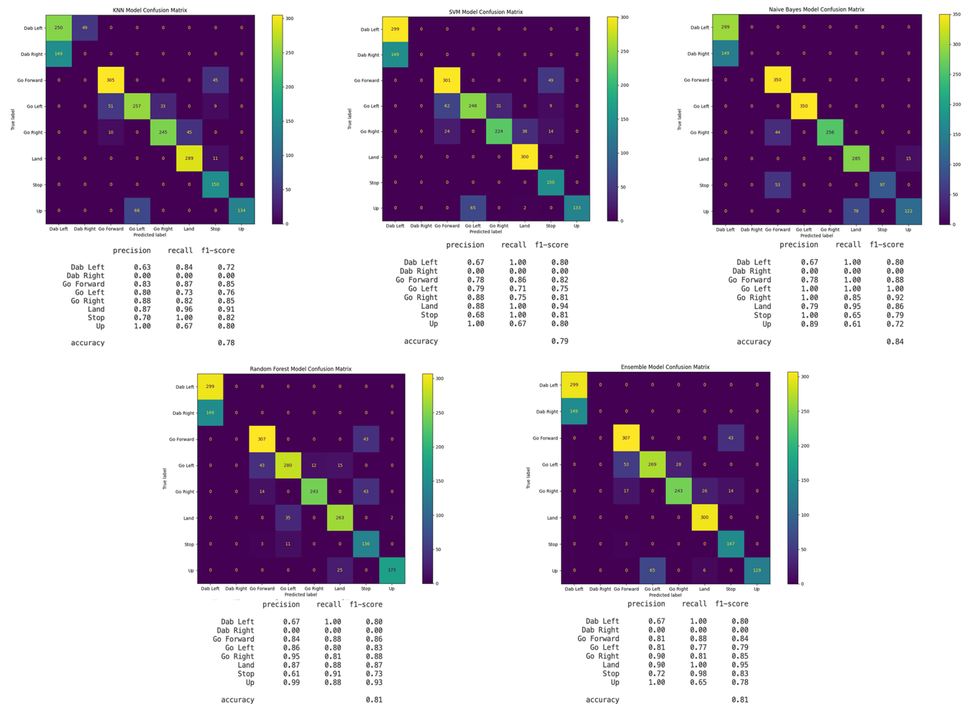

Figure 1: Confusion matrices and metrics of the machine-learning models for the first dataset

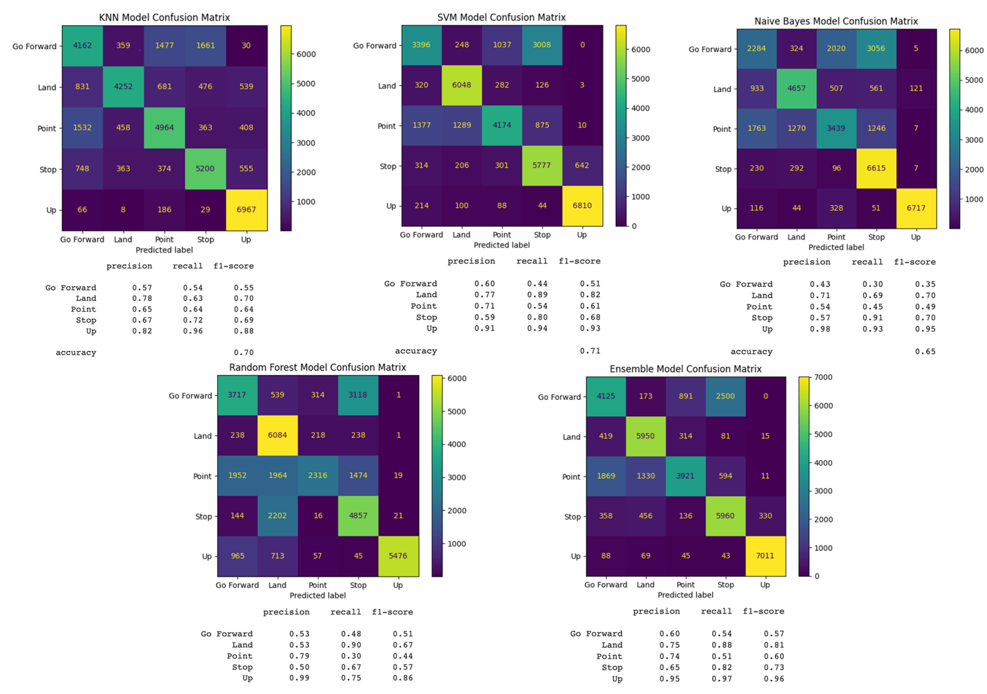

Figure 2: Confusion matrices and metrics of the machine-learning models for the second dataset

to enhance human-drone interaction. By fine-tuning parameters and evaluating the performance metrics of each model, we aimed to find the most practical and accurate classification model. Our results indicate the potential of our approach, as demonstrated by the confusion matrices and metrics obtained for different models and datasets. From the first dataset, Naive Bayes had the highest accuracy of 84%. For the second dataset, the Ensemble model had the highest accuracy of 74%. Because the first dataset was from only one person and is much smaller in size, it is likely that the models overfit to the data. This suggests that the second dataset and its models are more reliable and practical, despite their lower overall accuracy. In conclusion, our study contributes to the ongoing efforts to improve human-drone interaction by offering a method that is intuitive, seamless, and accessible to a wide range of users.

6 FUTURE WORK

Future work should focus on incorporating more complex gestures to achieve a higher level of control, expanding the datasets, and refining the classification models.

Table 1: Commands and Corresponding Gestures

Command	Gestures
Select individual drone(s)	Point
Select all drones in a group	Big circle motion
Split drones into two groups	Chop down and to the left or right
Follow me	Wave hand towards body
Go to this place	Both hands point to destination
Takeoff	Both hands raise up
Land	Both hands raise down
Ascend	Single arm raise up
Descend	Single arm raise down
Rotate	Single hand move with wrist left/right
Move left	Single arm moves left entirely
Move Right	Single arm moves right entirely
Stop	Elbow bent, palm facing out
Forward	Single arm jab/move outwards
Backward	Single arm jab/move inward

We have compiled gesture-command pairs for future data collection (Table 1), prioritizing an intuitive experience. To expand the datasets, we will collect our own data, aiming to involve 10 individuals performing each gesture 50 times. Ensuring consistent time intervals or frames during data collection will enable effective machine-learning training through windowing techniques, preventing overfitting. A real-time simulation in Gazebo would follow, utilizing Motion Capture suit data for real-time processing by the machine-learning model. Each model undergoes separate testing within the simulation. Accuracy, precision, recall, f1 score, and gesture recognition time will be calculated per model, supported by confusion matrices. These metrics will assist us in determining the practicality of each model for real-world application.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2150351. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We would also like to give a special thanks to Dr. Dulal Kar, Sathwika Dasu, and Kimberly Lopez.

REFERENCES

[1] Dario Ortega Anderez, Luis Pedro Dos Santos, Ahmad Lotfi, and Salisu Wada Yahaya. 2019. Accelerometer-based Hand Gesture Recognition for Human-Robot Interaction. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. 1402–1406. https://doi.org/10.1109/SSCI44817.2019.9003136

[2] Kshitij Chhabra, Pranay Mathur, and Veeky Baths. 2020. BCI Controlled Quadcopter Using SVM and Recursive LSE Implemented on ROS. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 4250–4255. https://doi.org/10.1109/SMC42975.2020.9282898

[3] Joseph DelPreto and Daniela Rus. 2020. Plug-and-Play Gesture Control Using Muscle and Motion Sensors. In *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 439–448.

[4] Aleksey Fedoseev, Ahmed Baza, Ayush Gupta, Ekaterina Dorzhieva, Riya Neelesh Gujarathi, and Dzmitry Tsetserukou. 2022. DandelionTouch: High Fidelity Haptic Rendering of Soft Objects in VR by a Swarm of Drones. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 1078–1083. https://doi.org/10.1109/SMC53654.2022.9945390

[5] Selmic R. Gardner, A. and J. Kanno. 2017. Motion Capture Hand Postures. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5TG86.

[6] Boris Gromov, Jérôme Guzzi, Luca M. Gambardella, and Alessandro Giusti. 2020. Intuitive 3D Control of a Quadrotor in User Proximity with Pointing Gestures. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 5964–5971. https://doi.org/10.1109/ICRA40945.2020.9196654

[7] Kianoush Haratiannejadi, Neshat Elhami Fard, and Rastko R. Selmic. 2019. Smart Glove and Hand Gesture-based Control Interface For Multi-rotor Aerial Vehicles. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. 1956–1962. https://doi.org/10.1109/SMC.2019.8914464

[8] Zhen Huo, Tao Deng, Zhiyan Dong, Lihua Zhang, and Lan Niu. 2022. A BCI-Based Motion Control System for Heterogeneous Robot Swarm. In *8th IEEE International Conference on Cloud Computing and Intelligent Systems, CCIS 2022, Chengdu, China, November 26-28, 2022*. IEEE, 261–266. https://doi.org/10.1109/CCIS57298.2022.10016398

[9] Roman Ibrahimov, Evgeny Tsykunov, Vladimir Shirokun, Andrey Somov, and Dzmitry Tsetserukou. 2019. DronePick: Object Picking and Delivery Teleoperation with the Drone Controlled by a Wearable Tactile Display. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 1–6. https://doi.org/10.1109/RO-MAN46459.2019.8956344

[10] Ji-Hoon Jeong, Dae-Hyeok Lee, Hyung-Ju Ahn, and Seong-Whan Lee. 2020. Towards Brain-Computer Interfaces for Drone Swarm Control. In *2020 8th International Winter Conference on Brain-Computer Interface (BCI)*. 1–4. https://doi.org/10.1109/BCI48061.2020.9061646

[11] Ji-Won Lee and Kee-Ho Yu. 2023. Wearable Drone Controller: Machine Learning-Based Hand Gesture Recognition and Vibrotactile Feedback. *Sensors* 23, 5 (feb 2023), 2666. https://doi.org/10.3390/s23052666

[12] Yijia Lu, Fei Han, Lei Xie, Yafeng Yin, Congcong Shi, and Sanglu Lu. 2017. I Am the UAV: A Wearable Approach for Manipulation of Unmanned Aerial Vehicle. In *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*. 1–3. https://doi.org/10.1109/SMARTCOMP.2017.7947014

[13] Matteo Macchini, Thomas Havy, Antoine Weber, Fabrizio Schiano, and Dario Floreano. 2020. Hand-worn Haptic Interface for Drone Teleoperation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 10212–10218. https://doi.org/10.1109/ICRA40945.2020.9196664

[14] Matteo Macchini, Fabrizio Schiano, and Dario Floreano. 2020. Personalized Telerobotics by Fast Machine Learning of Body-Machine Interfaces. *IEEE Robotics and Automation Letters* 5, 1 (2020), 179–186. https://doi.org/10.1109/LRA.2019.2950816

[15] A. Menshchikov, D. Ermilov, I. Dranitsky, L. Kupchenko, M. Panov, M. Fedorov, and A. Somov. 2019. Data-Driven Body-Machine Interface for Drone Intuitive Control through Voice and Gestures. In *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, Vol. 1. 5602–5609. https://doi.org/10.1109/IECON.2019.8926635

[16] Abdullah Mughees, Syed Zulqadar Hassan, Mobeen Yaseen, Neelam Mughees, Anam Mughees, Ali Abbas, Tariq Kamal, Zain Iqbal, and Muhammad Abbas Khan. 2020. Gesture Controlled Hexacopter using MEMS Technology. In *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*. 1–5. https://doi.org/10.1109/ICETST49965.2020.9080741

[17] Taha Müezzinoğlu and Mehmet Karaköse. 2021. An Intelligent Human–Unmanned Aerial Vehicle Interaction Approach in Real Time Based on Machine Learning Using Wearable Gloves. *Sensors* 21, 5 (mar 2021), 1766. https://doi.org/10.3390/s21051766

[18] Gabriel Alexis Guijarro Reyes, Juan Martinez, Luis Rodolfo Garcia Carrillo, Ignacio Rubio Scola, José Baca, and Scott A. King. 2020. Deep Neural Network-Inspired Approach for Human Gesture-Triggered Control Actions Applied to Unmanned Aircraft Systems. In *Advances in Automation and Robotics Research*. Springer International Publishing, 92–111. https://doi.org/10.1007/978-3-030-40309-6_10

[19] Sang-Yun Shin, Yong-Won Kang, and Yong-Guk Kim. 2019. Hand Gesture-based Wearable Human-Drone Interface for Intuitive Movement Control. In *2019 IEEE International Conference on Consumer Electronics (ICCE)*. 1–6. https://doi.org/10.1109/ICCE.2019.8662106

[20] Shouheng Sun and Jingnan Ma. 2019. Brain Wave Control Drone. In *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*. IEEE. https://doi.org/10.1109/aiam48774.2019.00066