

Gesture Based Control Technology for Drones : Challenges and Solutions

Jiabao Wang

Ulink College of Shanghai, Shanghai, CN216, China

jiabao.wang@ulink.cn

Abstract. Gesture-based control systems have become an innovative approach in human-computer interaction, enabling users to operate devices like UAVs using hand movements. These systems generally consist of three main components: sensors to capture movements, gesture recognition algorithms to interpret them, and control interfaces to execute actions. However, several challenges still exist that limit the performance and practicality of these systems. Key problems include accuracy in varying environmental conditions, latency issues caused by sensor overload, and user ergonomics, which can affect usability over time. For instance, sensors like optical cameras and IMUs may struggle in low-light or complex environments, leading to errors in gesture interpretation. Moreover, real-time processing is essential for seamless control, but the high computational demands of machine learning algorithms often lead to delays. This paper explores these challenges in detail and presents solutions, such as using multi-sensor setups to improve data accuracy, implementing adaptive algorithms that can learn and adjust to user behavior, and refining control interfaces with real-time feedback mechanisms to enhance the user experience. Addressing these issues will make gesture-based control systems more efficient, reliable, and user-friendly, enabling wider application across industries like virtual reality, robotics, and military operations.

Keywords: Gesture recognition, drones, human-computer interaction, sensors, real-time processing.

1. Introduction

In recent years, the advancement of unmanned aerial vehicles (UAVs), commonly known as drones, has opened up a multitude of probabilities across various industries, ranging from aerial photography to search and rescue operations. However, conventional control methods such as joysticks or mobile applications can be restrictive and cumbersome in dynamic environments. With the introduction of “gesture-control” technology, which enables users to operate machines using natural hand movements, an intuitive interface between humans and machines is established. This not only facilitates convenience through uncomplicated gestures but also encourages novice users by eliminating complex maneuvers. What’s more, in the military field, certain tasks can be expedited with simplified actions: raising hands for ascent and lowering hands for descent; moving left hand to the left and right hand to the right.

However, while promising significant advancements, the development of gesture-controlled drone systems presents several challenges. This paper provides a comprehensive overview of key technologies involved in implementing gesture-based control for drones while addressing associated challenges and

proposing solutions to overcome these obstacles. It explores the principles behind gesture recognition systems, identifies the primary problems associated with each, and purposes the solution to overcome these obstacles. This paper may offer valuable references for future studies about the gesture-based control.

2. Techniques and tools for implementation

Gesture-based control systems have revolutionized the way users interact with drones. By monitoring and interpreting specific gestures, these systems provide a seamless and intuitive experience for drone operators. This innovative technology typically relies on a combination of hardware and software technologies to capture, process, and analyze gestures accurately. At the heart of a complete gesture-based control system are sensors, gesture recognition and the control interface.

2.1. Sensors

In gesture-based drone systems, sensors are vital as they capture and process the movements of the user's body, typically hands, and translate them into control commands. Sensors can be placed on the drone or external devices, and they detect gestures through various means. The following are some common sensors for this system:

Microsoft Kinect: a RGB-Depth sensor applied for human computer interface. The Kinetic contains an RGB camera and two 3D depth sensors. One of them is the emitter of infrared light, and the other one is the receiver, which detects the position of the operator [1].

The IMUs (Inertial Measurement Units) is another significant sensor used in this system. It consists of accelerometers and gyroscopes, which are often used to detect movements and gestures; they can offer the position and orientation of your motions.

Optical Camera: Traditional optical cameras rely on 2D imagery; however, when coupled with advanced computer vision algorithms, they can identify and track the hand positions and gestures of users in real-time. An example is in Utsumi's paper, which uses multiple cameras to track the motion and position of hands [2].

Wearable Devices: Devices with sensors that are used to detect motion and muscle activity, making them ideal for wearable gesture recognition systems. These sensors are often mounted on smart gloves or wristbands, enabling users to control machines by simply moving their hands or fingers.

2.2. Gesture recognition algorithm

As mentioned above, the model Kinetic can be used as a sensor for capturing the gesture. It can represent the body part as skeletons, as shown in Figure 1.

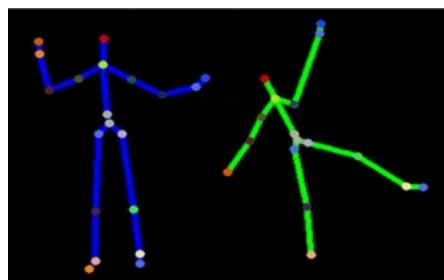


Figure 1. The skeletons [3]

Obviously, the nodes can be separated into multiple parts. The left and right elbow, wrist, and shoulder. They all have their own coordinate positions. Thus, we can design some specific motions for controlling the drones based on it [4]. Once the gesture data is collected, it should be processed by a sophisticated algorithm that is responsible for identifying and interpreting. These algorithms employ some machine learning models, for example, convolutional neural networks (CNNs), which can recognize complex hand gestures with high precision. For example, in “3D-CNN based dynamic gesture

recognition for Indian sign language modeling,” a 3d gesture recognition system was built using the CNN machine learning method [5]. Template matching algorithms are another type of algorithm that works by comparing the captured gesture to a set of pre-defined templates stored in the system. If the captured gesture matches one of the templates, the system triggers the corresponding action [6].

2.3. Control interface

After all gestures are interpreted, the recognized commands should be translated into actions that the drone can execute; the drone’s feedback is essential. There can be techniques to provide feedback. The integration of AR and VR in drone control systems is gaining popularity due to the immersive experience these technologies offer. AR headsets or VR goggles allow users to interact with virtual elements of the drone’s environment in real-time. This is particularly useful in applications such as monitoring or remote inspection, where the operator can manipulate the drone through virtual overlays rather than looking down at a physical controller. For example, in Konstantoudakis’s research, an application is created on the Android platform. It acts as a control interface between drones and other system architectures that collect the feedback and readings from drones [7]. The feedback mechanism is another type of interface that is used to confirm successful gesture recognition or guide users to perform the correct gesture in the control interface.

3. Problems in gesture-based drone control system

As the author mentioned in the intro part, this technology is still immature and in the early stages of development; thus, there are still plenty of obstacles that need to be overcome. This part separates the concepts of problems into three parts, respectively.

3.1. Sensor related problems

To make a robust system of gesture control, the researcher is supposed to collect precise data or images in order to proceed; however, there exist various factors that will affect the data accuracy. The complicated background can be an enormous influence factor because the depth and optical cameras are highly susceptible to environmental changes. The lighting condition will vary the data collected by sensors like Kinetic; it struggles with insufficient light or those reflective surface. The mistaken data from non-gesture moving objects will also be an issue, as they cannot avoid controlling drones in a crowded area. Those sensors are highly sensible and will capture the motion of every visible creature [8]. What’s more, latency and processing overload would be a problem that occurs at a high frequency in systems that have sensors: IMUs can generate a huge amount of data, and when combined with high-volume optical or depth sensor inputs, the processing requirements become overwhelming. The delays can result in a lag, which can be dangerous or even lethal in practical applications such as drone-based rescue operations.

3.2. Gesture recognition algorithm issues

Developing gesture recognition algorithms poses several challenges. Real-time processing is essential for applications such as virtual reality, where any delay in gesture recognition can impact the user experience. Additionally, algorithms must be able to handle variability in gesture execution, as different users may perform the same gesture in slightly different ways; thus, gesture recognition algorithms must ensure that they work across different users and in varied environments, for instance, variations in movement speed, and hand sizes. Moreover, machine learning models for gesture recognition typically require high processing power, which limits their use in real-time applications. In addition, the need for large training datasets makes it hard to deploy these models in real-world scenarios where gesture types may not be well documented or consistent.

3.3. Control interface challenges

The real-time feedback in the control interface part played a decisive role; the lack of it leaves the user uncertain about whether their gestures have been properly interpreted or not. This can frustrate the users

and reduce the overall usability of the system. Further more, the use of gesture-based control systems can impose a high cognitive load on the user, particularly in complex environments where multiple gestures need to be performed in quick succession, such as the missions of Attack UAVs' operators in the war. This can result in increasing human errors.

4. Solutions

To address the challenges outlined above, several solutions will be proposed, ranging from advancements in sensor technology to improvements in algorithm design and control interface.

4.1. Improvements in sensors

One approach to improving the accuracy of gesture recognition systems is monitoring through several sensors. By combining data from multiple sensors, the system can compensate for the limitations of individual sensors. For example, a depth camera may struggle to recognize gestures in low-light conditions, but combining it with an accelerometer can provide additional data on the user's hand movements, improving overall accuracy. Multi-sensors can also enhance the system's ability to detect subtle gestures that may be missed by a single sensor, for instance, the wearable sensors can capture hand movements, while Kinetic can track the body movements, providing a more comprehensive view of the user's gestures. Another solution to sensor-related challenges is the development of more adaptive sensors. Creating those sensors that can adjust their sensitivity or processing methods based on different environmental conditions can overcome some obstacles faced by current systems. As mentioned above, the noisy data may cause latency and damage the user experience. Therefore, it is available to introduce a technique called 'Smoothing Filter-Based Kalman Predictor' [9]. Obviously, it makes real-time gesture recognition more feasible. The predictor can reduce the overlay and improve responsiveness while operating.

4.2. Advances in gesture recognition algorithms

Machine learning techniques offer a promising solution to the challenge of user variability in gesture recognition. By training the algorithm on a large dataset containing gestures performed by different users under different conditions, the system can learn to generalize variations in gesture execution. A specific learning system can personalize the algorithm for individual operators. These systems continuously learn from the user's behavior, adapting to their unique gesture style and preferences. To deal with the issue of computational load, edge computing and cloud computing provide the solution [9]. Edge computing reduces the need for cloud-based computation, minimizing latency and power consumption.

4.3. Enhanced control interfaces and user experience

Enhancing control interfaces with real-time feedback mechanisms can significantly improve the user experience in gesture recognition systems. By providing users with immediate feedback on whether their gestures were recognized correctly, the system can guide users to perform gestures more accurately. For example, a visual cue, such as a highlighted menu item or an on-screen icon, can indicate that the system has recognized a gesture, while auditory or haptic feedback can confirm the action. Real-time feedback can also help users correct their gestures before they are fully executed, reducing the likelihood of errors. For example, a VR system might display a visual guide to help the user align their hand correctly for a specific gesture, ensuring more accurate recognition. Finally, improving the design of gestures themselves can enhance the overall user experience and reduce physical strain. User-centered design approaches can help developers create gestures that feel natural and simple to users. For example, incorporating feedback from users during the design process can lead to the development of gestures that are easier to perform and more comfortable over extended periods. Additionally, ergonomically designed gestures can reduce the risk of fatigue or an uncomfortable experience, making the system more accessible and enjoyable to use. For example, minimizing the need for large, repetitive gestures or designing gestures that require minimal hand movement can improve the system's usability.

5. Conclusion

Gesture recognition systems hold immense potential to revolutionize human interactions with computers by providing users with an intuitive and natural way to control devices like drones. This technology offers a promising alternative to traditional control methods, such as joysticks or mobile apps, by enabling users to interact with machines through simple hand gestures. Despite the many advantages, significant challenges still need to be addressed before gesture-based control systems can achieve their full potential, especially in applications where precision and reliability are critical, such as military operations, healthcare, or rescue missions. One of the main challenges is sensor-related accuracy. Environmental factors such as lighting, background noise, and the presence of other moving objects can interfere with sensor performance, leading to errors in gesture detection.

To overcome these issues, future systems could rely on multi-sensor setups, combining data from different types of sensors (e.g., depth cameras, IMUs, and wearable devices) to improve accuracy and reduce the impact of environmental factors. Moreover, advancements in sensor technology, such as adaptive sensors that adjust their sensitivity in response to environmental changes, will also play a crucial role in improving system reliability. Another major challenge is the computational complexity of gesture recognition algorithms. Machine learning models, such as convolutional neural networks (CNNs), offer high precision but require substantial processing power, which can result in delays, making real-time control difficult. To mitigate this issue, the use of edge computing and cloud-based solutions can help offload computational tasks, reducing latency and improving responsiveness. Additionally, adaptive learning systems can be designed to continuously adjust to users' unique gesture styles, increasing the system's ability to generalize across different users and usage scenarios. User experience is another crucial aspect that must be addressed. Enhancing the control interface by incorporating real-time feedback mechanisms, such as visual or auditory cues, can guide users in performing gestures more accurately and help them understand whether their actions have been recognized correctly. Furthermore, designing more ergonomic and user-friendly gestures can reduce physical strain, especially in applications requiring prolonged usage, making the system more accessible and enjoyable to use.

In conclusion, while challenges remain, the potential of gesture-based control systems is vast. By addressing issues related to sensors, algorithms, and user interfaces, these systems can become more robust, accurate, and user-friendly. As technology advances, gesture-based control will likely play an increasingly important role across industries, from virtual reality and gaming to healthcare and military operations.

References

- [1] Ruan, X., & Tian, C. (2015, August). Dynamic gesture recognition based on improved DTW algorithm. In 2015 IEEE International Conference on Mechatronics and Automation (ICMA) (pp. 2134-2138). IEEE.
- [2] Utsumi, A., & Ohya, J. (1999, June). Multiple-hand-gesture tracking using multiple cameras. In Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149) (Vol. 1, pp. 473-478). IEEE.
- [3] Microsoft Central Engineering (2011). "Kinect for Windows SDK Beta", [Online]. <https://www.microsoft.com/en-us/research/uploads/prod/2011/04/kinect.png>
- [4] Ma, L., & Cheng, L. L. (2016, April). Studies of AR drone on gesture control. In 2016 3rd International Conference on Materials Engineering, Manufacturing Technology and Control (pp. 1864-1868). Atlantis Press.
- [5] Singh, D. K. (2021). 3d-cnn based dynamic gesture recognition for indian sign language modeling. *Procedia Computer Science*, 189, 76-83.
- [6] Yun, L., Lifeng, Z., & Shujun, Z. (2012). A hand gesture recognition method based on multi-feature fusion and template matching. *Procedia Engineering*, 29, 1678-1684.

- [7] Konstantoudakis, K., Christaki, K., Tsiakmakis, D., Sainidis, D., Albanis, G., Dimou, A., & Daras, P. (2022). Drone control in AR: an intuitive system for single-handed gesture control, drone tracking, and contextualized camera feed visualization in augmented reality. *Drones*, 6(2), 43.
- [8] Pabendon, E., Nugroho, H., Suheryadi, A., & Yunanto, P. E. (2017, November). Hand gesture recognition system under complex background using spatio temporal analysis. In *2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)* (pp. 261-265). IEEE.
- [9] Li, X., Ou, X., Li, Z., Wei, H., Zhou, W., & Duan, Z. (2018). On-line temperature estimation for noisy thermal sensors using a smoothing filter-based Kalman predictor. *Sensors*, 18(2), 433.