DRONE CONTROL BY CHATBOT WITH ARTIFICIAL INTELLIGENCE AND VISUAL ANALYSIS

Murilo de Lima Barros[1], Henrique Antônio Guanais Corneau[1], Ramon Vinícius Ferreira de Souza[1], Vinícius Wallace Rebouças Silva[1], Bruno Silva[1], Helton Maia[*1]

[1]Universidade Federal do Rio Grande do Norte, Natal, Brasil (helton.maia@ufrn.br)

Abstract: This paper presents a proposal for controlling an autonomous Unmanned Aerial Vehicle (UAV) operated by Artificial Intelligence. A chatbot, integrated with Gemini AI, processes textual commands from the user and an image captured by the DJI Tello drone to generate and execute flight instructions, aiming at an intuitive and contextual interaction. This system allows the control of the drone through natural language, with the AI analyzing the scene for assisted and safe navigation.

Keywords: UAV; Artificial Intelligence; Chatbot; Natural Language; Flight Controls.

INTRODUCTION

The growing dissemination of Unmanned Aerial Vehicles (UAVs) has revolutionized several applications, from precision agriculture to security and monitoring operations, as discussed by Floreano and Wood (2015) in their study on the scientific and technological advances that drive the future of autonomous drones. In parallel with this expansion, there has been a significant advance towards the autonomy of these systems, seeking to reduce the need for direct human intervention and optimize the execution of complex tasks. This context drives the exploration of new control interfaces that allow a more intuitive and efficient interaction with autonomous drones. Recent studies highlight the integration of Large Language Models (LLMs) as a key pathway to achieving this, leveraging their advanced capabilities to enhance data analysis and decision-making in autonomous systems (Javaid et al., 2024).

This pursuit of greater autonomy has spurred extensive research lines focusing on challenges like ensuring safe landings and aiding in search and rescue operations. In this scenario, proposals emerge that integrate onboard sensors, computer vision and local decision-making algorithms, allowing small UAVs to operate autonomously even in challenging environments, with all processing performed on board (Tomic et al., 2012).

This study is part of this theme, proposing the development and evaluation of a control system for an autonomous UAV that uses artificial intelligence (AI) as the main interface between the user and the aircraft. The AI is responsible for interpreting natural language commands provided by the user and analyzing the image captured by the drone's camera. This contextual understanding is then translated into specific flight commands.

The main objectives of this study are to develop a chatbot capable of processing natural language commands and visual data to control a drone, implement its integration with the image processing AI and UAV control systems, and evaluate the effectiveness and intuitiveness of the human-AI-drone interaction in simulated or controlled scenarios.

MATERIALS AND METHODS

The drone control system operates through a continuous interaction flow between the user interface, the artificial intelligence process and the communication mechanisms with the aircraft, as illustrated in Figure 1.

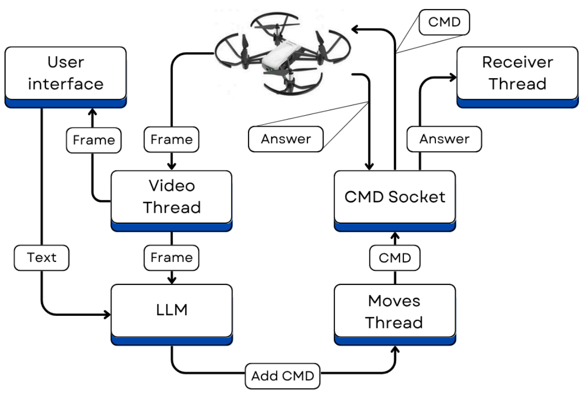

Figure 1 - Diagram of the general operation of the control.

Initially, the drone transmits the frame captured by its camera. This image is received and processed by the

Video Thread, which makes it available simultaneously to two components: the user interface, allowing real-time visual monitoring of the operation, and the Large Language Model (LLM), which uses it for contextual analysis.

The process of interaction between the user and Artificial Intelligence (AI) for generating drone control commands is detailed in Figure 2. This flow is initiated when the user submits a command in text format through the graphical interface and triggers the sending button. Immediately, the system captures the most recent video frame, which is continuously provided by the Video Thread.

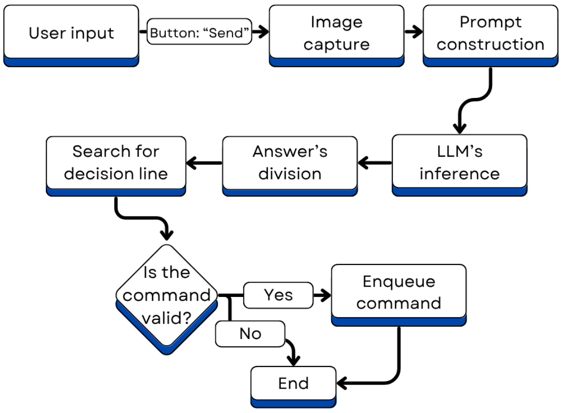

Figure 2 - AI command processing diagram.

With the user's text and the video frame in hand, the prompt construction step is performed. This prompt is carefully crafted to provide the LLM with the necessary context, including the user's input, the image of the current scene, the list of commands available to the drone, and the history of recent commands.

Once constructed, the prompt is sent to the AI, which performs inference to generate a multimodal response that is displayed in the interface. This textual response from the LLM, which includes an analysis of the scene, the command decision, and a justification, goes through a process of syntactical division and analysis. The system identifies the specific line in the LLM response that contains the proposed technical command.

After extraction, the proposed command is subjected to a validation step, where it is verified whether it belongs to the list of allowed commands. If the command is considered valid, it is queued to be processed by the Moves Thread and sent to the drone. If the command is invalid, the flow is interrupted for that specific interaction.

The Moves Thread is responsible for transmitting the command to the drone using a dedicated UDP socket (CMD Socket). After the command is received and processed by the aircraft, the drone sends a response message, typically "ok" or "error". This response is captured by the Receiver Thread, which monitors the same CMD Socket, thus completing the communication and control cycle.

RESULTS AND DISCUSSION

To evaluate the performance of the chatbot, a series of flight tests were conducted in a controlled indoor setting using a Tello drone. The evaluation focused on the system's effectiveness in processing natural language commands and its capacity for safe navigation based on visual feedback. The results indicate a high degree of reliability and precision in these tasks.

The AI proved to be efficient in navigating the drone indoors, since the commands are short and one-dimensional, making decision-making easier. During the tests, there were no accidents involving the drone falling or hitting obstacles visible to the drone's camera, demonstrating that the LLM is capable of processing and interpreting the image and evaluating the next movements in a manner consistent with the objectives instructed by the user. These observations suggest that AI control in indoor environments is promising.

The user interface displaying the drone's video feed can be seen in Figure 3. The text box is located just below the image, next to it are the drone's parameters.

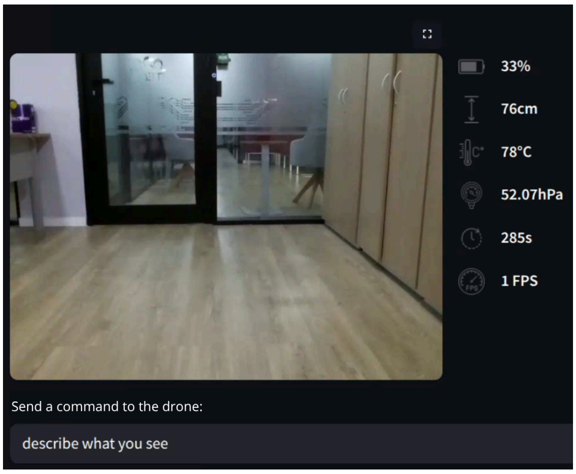

Figure 3 - User interface displaying image acquired by the drone.

In addition to the decision made by the AI, an analysis of the image and a justification for the decision are also displayed, as explained previously. An example of a response from the LLM can be seen in Figure 4, where the input text sent was "describe what you see", and the image is the frame from Figure 3.

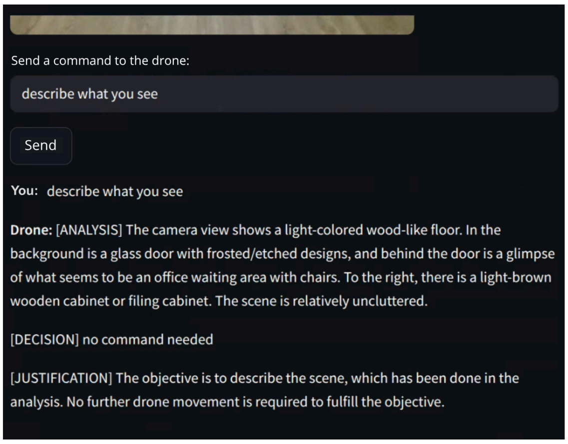

Figure 4 - Example of AI output in testing describing the image.

Figure 5 shows an example of the AI output for a movement command. A user can input an instruction like rotate 90 degrees, specifying the direction (clockwise or counter-clockwise), for example. In response, the LLM is instructed to provide a scene analysis, the specific flight command the drone will execute, and a rationale for this action based on the user's input.

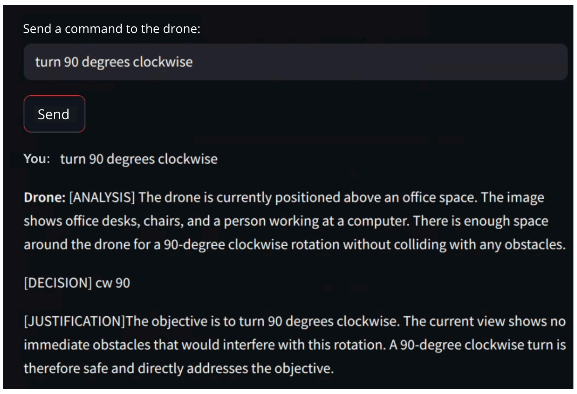

Figure 5 - Example of response to flight commands.

In initial tests, a performance issue was observed in the system with a low frame rate per second (FPS) in the user interface and an apparent slowness in responding to commands. The introduction of parallel processing with the implementation of the Moves Thread (responsible for sending commands) significantly improved the fluidity of image display.

The drone only executes the movement after the command has been effectively received and processed. Without a dedicated thread for sending commands, the main application loop would have to wait for each of these command sending and processing steps to complete before proceeding. During this waiting period, the capture and rendering of new video frames would be interrupted or severely delayed, resulting in a drastic drop in FPS.

CONCLUSION

This work demonstrates the potential of multimodal artificial intelligence-based drone control, integrating natural language processing with visual analysis and recognition. The proposal presents an innovative approach by integrating a chatbot with LLM as a means of control, providing an intuitive interface for interaction between the user and the robotic system.

The results obtained from the tests confirm that the system is capable of interpreting text commands and the environment around it, generating efficient flight commands. Although the control cannot be considered to be in real time, as there is a delay between sending the command and executing the action, the system is fluid, maintaining stable behavior when viewing the images transmitted by the drone and when making decisions.

Future work aims to use other forms of control, such as voice or hand gestures, as well as integrate new sensors to expand AI perception and explore the use of more specialized models for tasks such as object detection or trajectory planning. In short, this study highlights the important role of artificial intelligence in creating more accessible control systems for drones or UAVs of any type.

ACKNOWLEDGMENTS

We would like to thank the Escola de Ciências e Tecnologia (ECT), the Departamento de Engenharia de Computação e Automação (DCA) and the Núcleo de Pesquisa e Inovação em Tecnologia da Informação (NPITI) at UFRN, for all their support.

REFERENCES

CASTELLANO, Giovanna; MENCAR, Corrado; VESSIO, Gennaro. *Tecniche di Computer Vision per applicazioni di IA sostenibile mediante droni*. Bari: Dipartimento di Informatica, Università degli Studi di Bari “Aldo Moro”, 2022.

FLOREANO, Dario; WOOD, Robert J. *Science, technology and the future of small autonomous drones*. Nature, [s. l.], v. 521, n. 7553, p. 460–466, 2015.

JAVAID, S.; FAHIM, H.; HE, B.; SAEED, N. Large language models for UAVs: current state and pathways to the future. IEEE Open Journal of Vehicular Technology, v. 5, p. 1166-1192, 2024. DOI: 10.1109/OJVT.2024.3446799.

TOMIC, T. et al. Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue. *Robotics and Autonomous Systems*, v. 62, n. 6, p. 574–585, 2014. DOI: https://doi.org/10.1016/j.robot.2013.12.004.