Gaze-Augmented Drone Navigation

Kayla Pineda
Bhanuka Mahanama
Vikas G. Ashok
Sampath Jayarathna
kpine003@odu.edu
bhanuka@cs.odu.edu
vganjigu@cs.odu.edu
sampath@cs.odu.edu
Department of Computer Science, Old Dominion University
Norfolk, VA, USA

ABSTRACT

The use of unmanned aerial vehicles (UAVs) or drones, has significantly increased over the past few years. There is a growing demand in the drone industry, creating new workforce opportunities such as package delivery, search and rescue, real estate, transportation, agriculture, infrastructure inspection, and many others, signifying the importance of effective and efficient control techniques. We propose a scheme for controlling a drone through gaze extracted from eye-trackers, enabling an operator to navigate through a series of way points. Then we demonstrate and test the utility of our approach through a pilot study against traditional controls. Our results indicate gaze as a promising control technique for navigating drones revealing novel research avenues.

CCS CONCEPTS

• **Human-centered computing** → *User studies.*

KEYWORDS

drones, navigation, gaze-based control

ACM Reference Format:
Kayla Pineda, Bhanuka Mahanama, Vikas G. Ashok, and Sampath Jayarathna. 2023. Gaze-Augmented Drone Navigation. In *Augmented Humans Conference (AHs '23), March 12–14, 2023, Glasgow, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3582700.3583702

1 INTRODUCTION

Unmanned aerial vehicles (UAVs) or drones have been widely used in various domains, such as the military and healthcare. Practical applications of UAVs include but are not limited to, search and rescue, transportation of goods, farming, and building inspection. It is anticipated that the UAV market will exceed $92 billion, surpassing the 2020 value of $9.5 billion [4].

As drones become increasingly popular for everyday use, the number of user interactive methods increases with different types of control methods, such as hand gestures, voice control, and even brain control [26]. A type of control method that is still in development is gaze-augmented control. Gaze-based interactions have a variety of uses such as aiding those with disabilities [17, 29], search and rescue [20], driving [25], programming [27], gaming [16], and simulation [15, 24]. Unlike traditional control mechanisms that use handheld controls, gaze-augmented navigation offers users additional mobility. Moreover, combining autonomous control can prevent potential user errors such as overshooting and undershooting, often associated with traditional controllers.

This unique contribution presents opportunities to combine the wealth of existing infrastructure such as real-time data analytics using eye-tracking measures [9–11], gaze detection [18, 23], object detection and filtering of eye movements in dynamic area-of-interest (AOI) [12, 13], and for advanced eye movement analysis [14, 19, 21]. This will be critical in our future studies towards understanding "Trust", a construct that can determine secure and successful human-automation interactions in domains that the gaze-contingent control technique is beneficial.

2 RELATED WORK

In the context of gaze-based drone navigation, eye-tracking is commonly used as a companion input to another input method or requires additional work to navigate the drone [22]. Gaze-based control follows the same principle of looking in the direction of movement [15]. This creates problems as the user is essentially looking at two places at once; the user must look at the area to navigate to while making sure the device in flight remains stable. One solution is to filter eye movements as input [12, 13]. Using eye movements as input has been compared to the Midas Touch: all of the user's gaze is taken as valid input [8]. As a starting point, some type of on/ off switch can be implemented, requiring some additional input systems or methods. In [7] authors present keyboard controls as the companion input to eye-tracking using a desktop eye-tracker and keyboard input to control four degrees of freedom of the drone: rotation, speed, altitude, and translation. They found that the best control mode was using eye-tracking to control the rotation and speed of the drone and using the keyboard to control the translation and altitude of the drone. For methods that use only eye-tracking as its input, the user is required to do extra steps, such as following specific patterns with their eyes [5]. In [28], authors present single-stroke gaze gestures to navigate a drone through

---

AHs '23, March 12–14, 2023, Glasgow, United Kingdom

ACM ISBN 978-1-4503-9984-5/23/03.
https://doi.org/10.1145/3582700.3583702

a path. They found that control using gaze gestures did perform slower compared to keyboard, joystick, and dwell time controls, but participants reported gaze gestures required a lower mental workload compared to the other methods. In GazeGuide [3], authors present AR-based drone navigation using eye tracking optics and markers. The work is limited to the maneuvering of the camera with the UAV fixed in a predefined direction.

3 METHODOLOGY

We design our application architecture based on two tasks involved in controlling a drone using eye-tracking; 1) Identify and locate the user's AOI based on gaze, and 2) Search and navigate a drone to a given AOI. Based on this architecture, we start by composing processes for each task and defining AOIs using ArUco markers [6] for the simplicity of the application.

Eye movements were captured using the PupilLabs Core eye-tracking headset with a 200 Hz sampling rate. We used a DJI Tello drone, manufactured by Ryze Robotics. This lightweight drone is equipped with an HD camera. Joystick control is done using the Tello app, and the movement for the drone was programmed using the DJI Tello Python library. The checkpoints were designated by ArUco markers: synthetic square markers with a wide black border and an inner binary matrix that determines its identifier. In our experiment, we used four ID size 5x5 markers that measure 175x175 mm.

Gaze Tracking Process: The gaze tracking process starts by sampling the gaze positions from the eye-tracker of the user for a predefined period. We sample gaze positions along with the field of view (FOV) for 5 seconds. During the period, we ignore samples that correspond to blinks, missing data points, and low-confidence gaze estimates. Then, we scan for markers at each FOV and compute the distances from the gaze location to each marker in the FOV. Finally, we obtain the average distance of each marker and determine the marker corresponding to the gaze location by considering the least distance.

Drone Navigation Process: Our algorithm for drone navigation comprises two steps: 1) Scan for the marker, and 2) navigate to the located marker. During the scanning process, the drone iterates through a pre-defined set of relative angles for which the drone will rotate and scan for the selected marker. Detection of the marker during a scanning step causes the application to start navigating the drone to the marker. Should the selected marker not be detected at all, the user must reselect the marker or choose the next marker in the sequence to be scanned.

For the navigation task, we consider a coordinate system passing through the drone: the x-axis passing through the front to back of the drone, the y-axis passing through the sides of the drone, and the z-axis passing through the top of the drone. From this, we consider three main types of motions for navigating the drone: 1) Horizontal motion (along x-axis), 2) Vertical motion (along z-axis), and 3) Yaw rotation (about z-axis). For the horizontal motion, we use a piecewise function based on the area of the ArUco marker as observed by the drone camera. For the vertical motion and yaw rotation, we use a modified Proportional Integral Derivative (PID) controller based on the work of related setups [1, 2]. To successfully track the markers, these three motions must be in specified ranges.

Since the drone's rotation during the initial scanning phase and some sudden movements destabilizes the camera feed, the implementation includes short delays while waiting for camera feed stabilization. Moreover, to counter possible decode errors during transmission, we add retries where we try to read frames from the drone camera.

3.1 Study Tasks

To evaluate the proposed system, we conducted a pilot study with five participants. Based on within-subject task allocation, each participant ran three trials using the eye-tracking control and three trials using the joystick control. In total, there were 15 trials conducted using the eye-tracking control, and 15 with the joystick control. We used an indoor environment comprising four ArUco markers on alternating sides. The task was to navigate the drone in a given sequence of markers. The experiment consisted of two tasks: the Baseline task and the Gaze-augmented task (see Figure 1). In both tasks, the time to track individual markers and the total time to complete the course was measured.

3.1.1 Baseline Task. During the baseline task, participants were instructed to navigate through a series of markers using virtual joystick controls on a mobile application (see Figure 1 (a)). After a briefing and training session, the proctor revealed the navigation sequence. For each marker, the proctor manually checked the approximate distance to ensure consistency.

3.1.2 Gaze-Augmented Task. The gaze-augmented task uses the same sequence using the eye-tracking control as described earlier. In the experimental setup, we used audio feedback to note key events in the system: start of gaze sampling, end of gaze sampling, arrival at a marker, and completion of marker tracking.

4 RESULTS

Participants favored the eye-tracking controls over the typical joystick control. The common feedback was the joystick was too sensitive, making it easy to overshoot and difficult to position the drone in the appropriate location. With eye-tracking, little was required from the participant, making the eye-tracking method the more favorable control method between the two. Next, we measured different times based on events in the navigation process. To compare the two approaches, we calculate *Marker Tracking Time* and the *Total Time* during each trial.

Marker Tracking Time. We define the marker tracking time as the time between proctor instruction and the proctor acknowledgment of success during the baseline task. For the gaze-controlled task, we define it as the time from the drone scanning and reaching the target marker. Then we obtained the mean marker tracking times for each marker for evaluations across all trials (see Table 1a).

Total Time. We define the total time as the time between the drone initialization and the marker tracking success acknowledgment (proctor or audio feedback) of the final marker. We report the average time taken by each user to complete the task using both control mechanisms (see Table 1b).

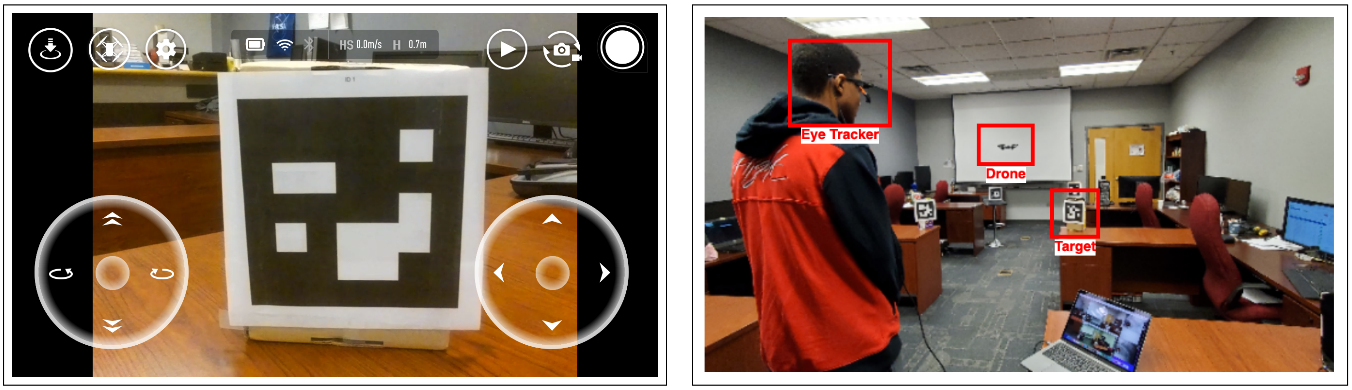

Figure 1: Experimental Tasks (a) Navigating drone with virtual controls on the mobile application, and (b) Navigating drone using gaze.

Table 1: Quantitative Data: (a) Average Individual Tracking Times (sec), and (b) Average Total Times (sec)

Marker ID	Joystick	Eye-tracking
0	21.73	33.01
1	19.46	33.50
2	25.31	29.65
3	20.27	27.37
Average	21.69	30.88

Participant	Joystick	Eye-tracking
1	112.67	121.60
2	107.83	128.11
3	102.48	124.35
4	88.01	140.18
5	114.92	109.02
Average	103.68	123.53

5 CONCLUSION

In this feasibility study, we presented a navigation mechanism using eye-tracking. We demonstrated the utility of our approach through a pilot study and compared the performance against a traditional controller (joystick control). Despite our approach being more time-consuming than traditional controls, we identified potential approaches to improve performance, utility, and robustness. In the context of drone navigation, our findings indicate gaze as a feasible control method. In the future, we are planning to integrate object tracking, and eye tracking enabled mixed reality (MR) (e.g., Microsoft HoloLens2) to support rich and immersive gaze-based interactions for drone navigation.

ACKNOWLEDGMENTS

This work was supported in part by NSF CAREER 2045523, NSF REU Supplement 2214096, and Virginia Space Grant Consortium Undergraduate STEM Research Scholarship for Kayla Pineda.

REFERENCES

- [1] Mohamed Attigui. 2021. Autonomous drone Programming Feature detection, tracking, and obstacle avoidance. (09 2021). https://doi.org/10.13140/RG.2.2.29978.16329
- [2] V. Madhu Babu, Kaushik Das, and Swagat Kumar. 2017. Designing of self tuning PID controller for AR drone quadrotor. In *2017 18th International Conference on Advanced Robotics (ICAR)*. 167–172. https://doi.org/10.1109/ICAR.2017.8023513
- [3] Pavan Kumar BN, Adithya Balasubramanyam, Ashok Kumar Patil, and Young Ho Chai. 2020. GazeGuide: An eye-gaze-guided active immersive UAV camera. *Applied Sciences* 10, 5 (2020), 1668.
- [4] David Daly. 2020. Drone industry outlook for the United States: 2020-2030. *Consortiq* (Nov 2020). https://consortiq.com/uas-resources/drone-industry-outlook-us-2020-2030
- [5] Heiko Drewes, Alexander De Luca, and Albrecht Schmidt. 2007. Eye-Gaze Interaction for Mobile Phones. In *Proceedings of the 4th International Conference on Mobile Technology, Applications, and Systems and the 1st International Symposium on Computer Human Interaction in Mobile Technology* (Singapore) *(Mobility '07)*. Association for Computing Machinery, New York, NY, USA, 364–371. https://doi.org/10.1145/1378063.1378122
- [6] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Rafael Medina-Carnicer. 2016. Generation of fiducial marker dictionaries using mixed integer linear programming. *Pattern recognition* 51 (2016), 481–491.
- [7] John Hansen, Alexandre Alapetite, I Mackenzie, and Emilie Møllenbach. 2014. The Use of Gaze to Control Drones. *Eye Tracking Research and Applications Symposium (ETRA)*. https://doi.org/10.1145/2578153.2578156
- [8] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (apr 1991), 152–169. https://doi.org/10.1145/123078.128728
- [9] Yasith Jayawardana, Vikas G Ashok, and Sampath Jayarathna. 2022. StreamingHub: interactive stream analysis workflows. In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries*. 1–10.
- [10] Yasith Jayawardana and Sampath Jayarathna. 2020. Streaming Analytics and Workflow Automation for DFS. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. 513–514.
- [11] Yasith Jayawardana, Gavindya Jayawardena, Andrew T Duchowski, and Sampath Jayarathna. 2021. Metadata-driven eye tracking for real-time applications. In *Proceedings of the 21st ACM Symposium on Document Engineering*. 1–4.
- [12] Gavindya Jayawardena and Sampath Jayarathna. 2020. Automated Filtering of Eye Gaze Metrics from Dynamic Areas of Interest. In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE, 67–74.
- [13] Gavindya Jayawardena and Sampath Jayarathna. 2021. Automated Filtering of Eye Movements Using Dynamic AOI in Multiple Granularity Levels. *International Journal of Multimedia Data Engineering and Management (IJMDEM)* 12, 1 (2021), 49–64.
- [14] Gavindya Jayawardena, Anne Michalek, and Sampath Jayarathna. 2019. Eye tracking area of interest in the context of working memory capacity tasks. In *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE, 208–215.
- [15] Juno Kim, Stephen Palmisano, April Ash, and Robert Allison. 2010. Pilot Gaze and Glideslope Control. *TAP* 7 (06 2010). https://doi.org/10.1145/1773965.1773969

[16] Krzysztof Krejtz, Cezary Biele, Dominik Chrzastowski, Agata Kopacz, Anna Niedzielska, Piotr Toczyski, and Andrew Duchowski. 2014. Gaze-Controlled Gaming: Immersive and Difficult but Not Cognitively Overloading. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1123–1129. https://doi.org/10.1145/2638728.2641690

[17] Vinay Krishna Sharma, Kamalpreet Saluja, Vimal Mollyn, and Pradipta Biswas. 2020. Eye Gaze Controlled Robotic Arm for Persons with Severe Speech and Motor Impairment. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 12, 9 pages. https://doi.org/10.1145/3379155.3391324

[18] Bhanuka Mahanama, Yasith Jayawardana, and Sampath Jayarathna. 2020. GazeNet: Appearance-based gaze estimation using capsule networks. In *Proceedings of the 11th augmented human international conference*. 1–4.

[19] Bhanuka Mahanama, Yasith Jayawardana, Sundararaman Rengarajan, Gavindya Jayawardena, Leanne Chukoskie, Joseph Snider, and Sampath Jayarathna. 2022. Eye movement and pupil measures: A review. *Frontiers in Computer Science* 3 (2022), 733531.

[20] James Mardell, Mark Witkowski, and Robert Spence. 2011. Gaze-Contingent Enhancements for a Visual Search and Rescue Task. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (Toulouse, France) *(APGV '11)*. Association for Computing Machinery, New York, NY, USA, 109. https://doi.org/10.1145/2077451.2077472

[21] Primesh Pathirana, Shashimal Senarath, Dulani Meedeniya, and Sampath Jayarathna. 2022. Eye gaze estimation: A survey on deep learning-based approaches. *Expert Systems with Applications* 199 (2022), 116894.

[22] Christian Pfeiffer, Simon Wengeler, Antonio Loquercio, and Davide Scaramuzza. 2022. Visual attention prediction improves performance of autonomous drone racing agents. *Plos one* 17, 3 (2022), e0264471.

[23] Shashimal Senarath, Primesh Pathirana, Dulani Meedeniya, and Sampath Jayarathna. 2022. Customer gaze estimation in retail using deep learning. *IEEE Access* 10 (2022), 64904–64919.

[24] Nikolai Smolyanskiy and Mar Gonzalez-Franco. 2017. Stereoscopic First Person View System for Drone Navigation. *Frontiers in Robotics and AI* 4 (2017). https://doi.org/10.3389/frobt.2017.00011

[25] Martin Tall, Alexandre Alapetite, Javier San Agustin, Henrik H.T Skovsgaard, John Paulin Hansen, Dan Witzner Hansen, and Emilie Møllenbach. 2009. Gaze-Controlled Driving. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI EA '09)*. Association for Computing Machinery, New York, NY, USA, 4387–4392. https://doi.org/10.1145/1520340.1520671

[26] Dante Tezza and Marvin Andujar. 2019. The State-of-the-Art of Human–Drone Interaction: A Survey. *IEEE Access* 7 (2019), 167438–167454. https://doi.org/10.1109/ACCESS.2019.2953900

[27] Astrid Thomschke, Daniel Stolpe, Marcel Taeumel, and Robert Hirschfeld. 2016. Towards Gaze Control in Programming Environments. In *Proceedings of the Programming Experience 2016 (PX/16) Workshop* (Rome, Italy) *(PX/16)*. Association for Computing Machinery, New York, NY, USA, 27–32. https://doi.org/10.1145/2984380.2984384

[28] Mingxin Yu, Yingzi Lin, David Schmidt, Xiangzhou Wang, and Yu Wang. 2014. Human-Robot Interaction Based on Gaze Gestures for the Drone Teleoperation. *Journal of Eye Movement Research* 7 (09 2014), 1–14. https://doi.org/10.16910/jemr.7.4.4

[29] Guangtao Zhang, John Paulin Hansen, Katsumi Minakata, Alexandre Alapetite, and Zhongyu Wang. 2019. Eye-Gaze-Controlled Telepresence Robots for People with Motor Disabilities. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (Daegu, Republic of Korea) *(HRI '19)*. IEEE Press, 574–575.