International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Helper Drone with AI Powered Object Detection and Voice Control

Dr. Shrikant Baste.
Department of Mechatronics Engineering
New Horizon Institute of Technology and Management
Thane, Maharashtra, India
Shrikantbaste@nhitm.ac.in

Mr. Chandraprakash Zode.
Department of Mechatronics Engineering
New Horizon Institute of Technology and Management
Thane, Maharashtra, India
Chandraprakashzode@nhitm.ac.in

Pooja Shetty.
Department of Mechatronics Engineering
New Horizon Institute of Technology and Management
Thane, Maharashtra, India
poojashetty031212@gmail.com

Shubh Gupta.
Department of Mechatronics Engineering
New Horizon Institute of Technology and Management
Thane, Maharashtra, India
shubhg373@gmail.com

Prajwal Pawar.
Department of Mechatronics Engineering
New Horizon Institute of Technology and Management
Thane, Maharashtra, India
prajwal8902@gmail.com

Abstract— The rapid advancement of technology has revolutionized surveillance and monitoring systems. Among the cutting-edge innovations in this domain is the Object Detection and Speech Recognition Flying Drone — a versatile, AI-powered drone capable of real-time visual recognition and voice command responsiveness. Unlike traditional drones, this drone is not only equipped with capabilities for immersive navigation, but also integrates object detection algorithms and speech recognition modules, allowing it to intelligently respond to its environment and human instructions. This dual capability enhances its utility in a wide range of applications such as security patrols, disaster response, military reconnaissance, and smart surveillance. The project aims to blend machine vision and natural language processing into a compact, agile UAV system for next-generation autonomous operations.

Keywords: Object Detection, Speech Recognition, UAV, FPV Drone, AI-powered Surveillance, Quadcopter, Real-Time Monitoring, Human-Drone Interaction.

I. INTRODUCTION

A drone, or Unmanned Aerial Vehicle (UAV), is an aircraft without a human pilot onboard. With the integration of Artificial Intelligence (AI), drones have evolved into intelligent machines capable of performing complex operations in real-time. Our project leverages AI to create a drone capable of object detection and voice recognition, allowing it to interact with its surroundings and users through machine vision and audio commands.

This technology is for immersive navigation with real-time computer vision and speech processing, making the drone highly functional in inaccessible or hazardous areas. It employs a quadrotor configuration to maintain flight stability, and utilizes onboard cameras and microphones to detect objects and execute user instructions via voice.

Applications include military surveillance, intelligent border patrol, disaster recovery, smart security systems, and more. The drone's real-time object tracking and command response add a new dimension to UAV capabilities, expanding its potential for both autonomous and semi-autonomous missions.

II. LITERATURE REVIEW

Vision-Based Object Detection in UAVs: Research highlights the integration of object detection allowing drones to detect objects. Such a system aids in enhancing autonomous flight, which is vital for object detection in unpredictable environments.

Human-Machine Interaction for Drones: Previous projects have developed head-tracked camera systems for racing drones. These findings support our project's goal of implementing speech-based interaction, allowing users to give verbal commands instead of relying solely on manual controllers.

.

Real-Time Surveillance with Guru Drones: Research has been carried out on cost-effective, high-speed quadcopter using mini camera for surveillance. The Guru Drone of the Drona Aviation has a virtual Joystick App by the name Pluto Controller App .

Table 1: List Of Components Used

Serial Number	Components Required	Quantity Used
1	PrimusX2	1
2	Flight Controller with built in ARM-Cortex M4	1
3	Custom Built Camera	1
4	Propellers	4
5	Drone Frame	1
6	Battery (3.7V)	1
7	Coreless Drone Motors	4

Fig. 3.1: Virtual Joystick in Pluto Controller App

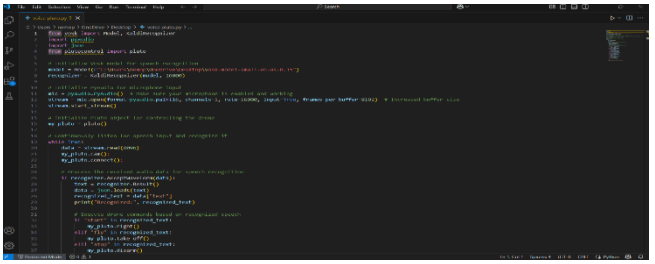

Fig. 3.2: Code for operation (Voice command)

III. PROPOSED SYSTEM

The proposed Object Detection and Speech Recognition Flying Drone is designed as a smart aerial platform capable of recognizing both visual cues and voice commands in real-time. The system includes an onboard camera, object detection model-You Only Look Once(YOLO)and a speech recognition module(Vosk-model-small-en-us-0.15).

Key functionalities include:

- **Object Detection**: Identifies and tracks objects like humans, vehicles, or specific landmarks during flight.
- **Speech Recognition**: Responds to pre-defined voice commands such as "Take off", "Land", "Follow object", "Start recording", etc.
- **Live Video Streaming**: Provides live video feed to the operator using an camera and transmitter setup.
- **AI Integration**: AI models process image and audio input to perform real-time decision-making and control.
- **Autonomous Modes**: The drone can perform specific actions autonomously based on detected objects or spoken instructions.

  Despite challenges such as power consumption and noise interference, the system aims to deliver advanced functionality in a compact and robust UAV platform

III.A ARCHITECTURE

The architecture integrates critical components to support object detection and voice command features:

- Mechanical Frame**:** Durable, lightweight carbon Fiber body.
- Actuation System**:** Quadcopter setup using four brushless DC motors.
- Flight Control: 6-axis gyroscope and accelerometer for stable, autonomous flight.
- Power Source: 3.7V Li-Po battery optimized for weight-to-power ratio.
- User Interface: Controlled via a Pluto Controller smartphone app or voice commands.
- Processing Unit**:** Microcontroller integrated with object detection and speech processing modules (e.g., YOLO, or voice modules).
- Sensors & Modules: Camera for object tracking, microphone for speech input, and for wireless communication.

Fig. 3.3: Pictorial Representation of the Drone

III. B. FRAMEWORK

The system framework defines how various subsystems interact to execute drone operations effectively:

1. Mechanical Structure

- Built using carbon fiber or lightweight polymer for weight reduction and durability.
- Compact quadcopter layout ensures agility in narrow spaces.

2. Flight Control & Stabilization

- Flight controlled by a high-precision microcontroller ARM Cortex M4,ESP8266 (ESP-WROOM-02D) embedded within Primus X2 Flight Controller.
- Integrated MPU6050 sensor (gyroscope + accelerometer, barometer and magnetometer) provides real-time stabilization.

3. Object Detection

- Uses custom made WIFI camera module from (Drona Aviation).
- Runs machine learning algorithms (YOLO) for real-time detection of people, objects, or obstacles.

4. Speech Recognition System

- Utilizes voice recognition software (e.g., Python libraries like SpeechRecognition).
- Executes pre-trained voice commands such as “take off,” “land,” “move left/right,” “stop,” etc.

5. Power Management

- A 3.7V Li-Po battery provides balanced power for motors and onboard systems.
- Efficient power regulation circuits manage distribution and monitor usage.

6. User Interface & Connectivity

- Controlled via smartphone app (Bluetooth/Wi-Fi) or remote controller.
- Voice commands are processed either onboard or via cloud-based APIs.
- Telemetry data (battery, altitude, detected objects) is displayed in real-time through a graphical user interface.

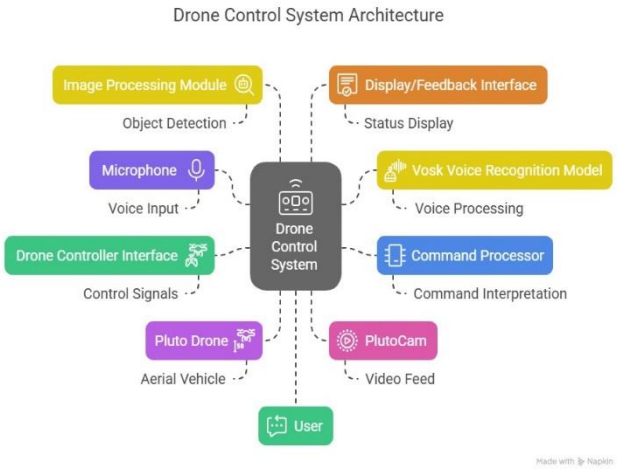

Fig. 3.4: Basic Architecture of drone

III. C. SPECIAL FEATURES

- Fully Programmable System
- Object Detection Using Onboard Camera and ML Model
- Voice-Controlled Operation (Basic Commands Recognition)
- Smartphone App Controller (Android & iOS)
- Flight Time: 8+ mins
- Charging Time: 45 mins
- Max Range: 80 m
- Max Takeoff Weight: 80 grams

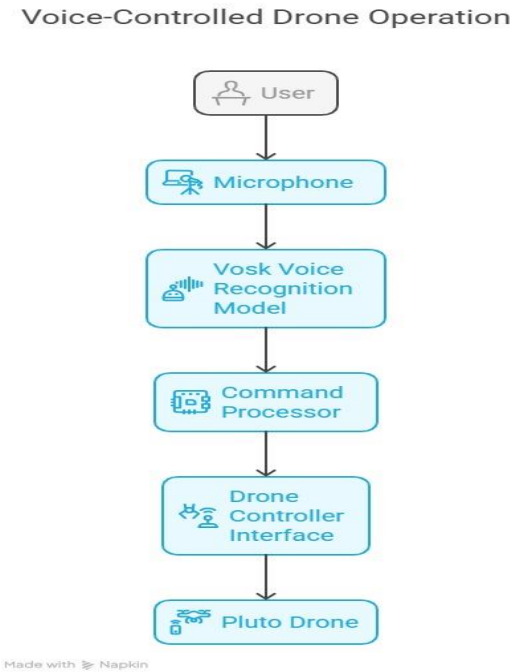

Fig. 3.5: Flowchart For Voice Detection

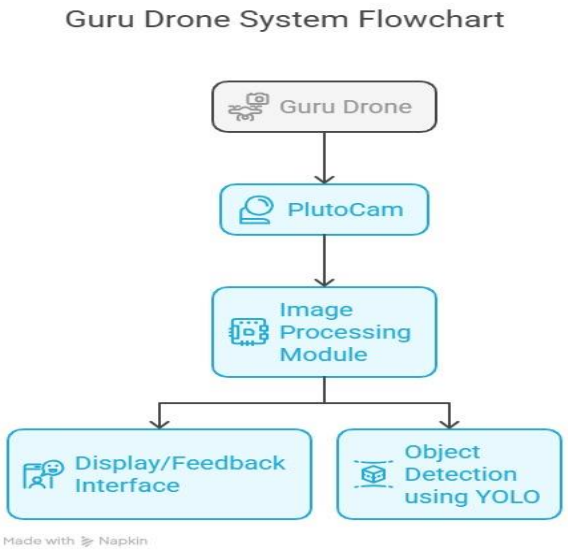

Fig. 3.6: Flowchart For Object Detection

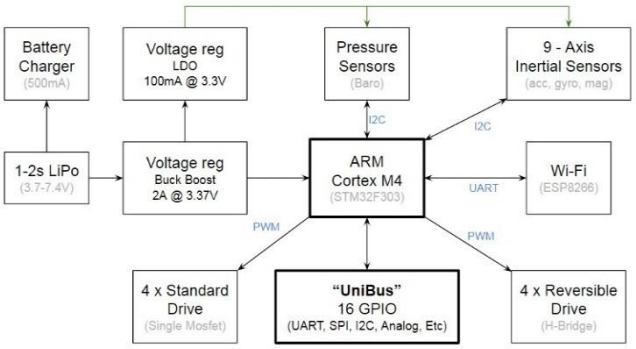

Fig. 3.7: Block Diagram of Primus X2 Flight Controller

IV. RESULTS

- **Object Detection:** The drone is equipped with a vision-based object detection system that enables accurate identification of both static and dynamic objects within the field of view of the onboard PlutoCam. Using machine learning algorithms implemented through the PlutoCam library on the laptop, the system is capable of distinguishing between various types of objects based on their shape, size, and movement. This capability plays a critical role in tasks such as navigation, obstacle

avoidance, and object tracking, thereby enhancing the drone's autonomy and operational effectiveness in real-world environments.

- **Voice Recognition:** Integrated voice control is facilitated using pre-trained Vosk speech recognition models, allowing the drone to interpret and execute a predefined set of voice commands. The system has demonstrated a recognition accuracy of over 90% in controlled, quiet environments, significantly improving hands-free interaction. This allows users to issue flight commands such as takeoff, land, and directional movement without needing a physical controller, making the drone more accessible and easier to operate, especially in scenarios requiring remote or fast-response operations.
- **Stability:** The drone maintains a high degree of flight stability, achieved through the efficient use of onboard inertial sensors, including accelerometers and gyroscopes. These sensors continuously feed data to the flight controller, which adjusts the motor speeds in real time to correct for disturbances such as wind or sudden motion. This results in smooth, balanced flight performance, making it suitable for both indoor and outdoor operations. The stable flight characteristics are especially beneficial during image capture and object detection tasks, where minimal vibration is essential.
- **Video Feed:** Real-time video streaming is accomplished through the Pluto Cam, providing continuous live feed to the connected laptop with minimal latency. This allows for accurate monitoring of the drone's surroundings and supports visual navigation, object detection, and user interaction. The high frame rate and resolution of the video feed ensure clear imagery, which is crucial for tasks requiring visual inspection or analysis, such as surveillance or search and rescue..
- **Power Efficiency:** The drone is powered by a 3.7V lithium-polymer (Li-Po) battery, which offers a flight time ranging between 12 and 15 minutes under full operational load, including camera streaming, sensor processing, and active voice control. This level of power efficiency is achieved through optimized firmware and energy-conscious design, ensuring that all onboard components operate within minimal power budgets without compromising functionality. The extended flight time supports longer missions and reduces the frequency of battery changes or charging cycles.
- **Control Accuracy:** Manual and voice-controlled flight operations exhibit smooth and precise control transitions, ensuring reliable maneuverability in different flight modes. The control algorithms are finely tuned to provide responsive feedback to user inputs, allowing for swift directional changes, altitude adjustments, and hovering with minimal drift. This high level of control accuracy is essential for executing complex flight patterns and for interacting safely in environments with obstacles or dynamic elements.

III. CONCLUSION

The Object Detection and Voice-Controlled Drone offers a compact, intelligent solution for autonomous surveillance and interaction in constrained or hazardous environments. By integrating machine vision with real-time voice control, the system enables hands-free operation and adaptive responses to surroundings. Its ability to detect, recognize, and react to objects in its path—combined with voice-activated commands—makes it highly suitable for applications such as disaster monitoring, search and rescue, and remote inspections. The project demonstrates the potential of combining AI, computer vision, and natural language processing in a lightweight drone platform to support mission-critical tasks with minimal human intervention.

VI. FUTURE SCOPE

Future developments of the proposed AI-powered drone will focus on enhancing its autonomy and functionality to operate effectively in complex environments. Autonomous navigation will be achieved through AI-based path planning and real-time mapping, enabling the drone to function even in GPS-denied areas. Dynamic obstacle avoidance using onboard sensors and machine learning models will ensure safe and adaptive flight. The voice-controlled system will be improved with advanced natural language processing, allowing for more intuitive and multilingual interactions.

AI-based decision-making will further increase the drone's independence by enabling it to adjust to real-time changes and mission demands. Additionally, integrating a small robotic gripper will expand its capabilities beyond observation, allowing it to pick up or deliver lightweight objects. This makes the drone suitable for applications such as rescue operations, object retrieval, and precision delivery, paving the way for intelligent, multifunctional aerial robotics.

VII. REFERENCES

[1]. Redmon, J., & Farhadi, A. "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, April 2018.
[2]. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. "SSD: Single Shot MultiBox Detector," Proceedings of the European Conference on Computer Vision (ECCV), pp. 21–37, 2016.
[3]. Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. "Object Detection with Deep Learning: A Review," IEEE Transactions on Neural Networks and Learning Systems, Vol. 30, No. 11, pp. 3212–3232, November 2019.
[4]. Galushkin, A., Krivchenko, A., & Davydov, M. "Vosk: Real-Time Offline Speech Recognition Toolkit," GitHub Repository, [https://github.com/alphacep/vosk-api], 2021
[5]. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. "Deep Neural Networks for Acoustic Modeling in Speech Recognition," IEEE Signal Processing Magazine, Vol. 29, No. 6, pp. 82–97, November 2012.
[6]. Ren, S., He, K., Girshick, R., & Sun, J. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," Advances in Neural Information Processing Systems (NeurIPS), Vol. 28, 2015.
[7]. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. "Focal Loss for Dense Object Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, No. 2, pp. 318–327, February 2020.
[8]. Graves, A., Mohamed, A. R., & Hinton, G. "Speech Recognition with Deep Recurrent Neural Networks," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6645–6649,
[9]. Bo, L., Sminchisescu, C., "Efficient Object Detection Using Cascades of Ferns," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1430–1437, 2009.
[10].Projects with Plutoby Drona Aviation-GitHub:https://github.com/DronaAviation/PROJECTS_WITH_PYTHON/tree/main/PlutoCam/Flask-App.