Smart Rover with Joystick and Auto Nomous Modes Featuring Live Streaming and Object Identification

Hundi Rama Aditya Vardhan
dept. of Electronics and
Communication Engineering
Gokaraju Rangaraju Institute of
Engineering and Technology
Hyderabad, India

Kuntamukkala Gokul Chowdary
dept. of Electronics and
Communication Engineering
Gokaraju Rangaraju Institute of
Engineering and Technology
Hyderabad, India

Nampally Pradyumna
dept. of Electronics and
Communication Engineering
Gokaraju Rangaraju Institute of
Engineering and Technology
Hyderabad, India

Mr. A. Radhanand
dept. of Electronics and
Communication Engineering
Gokaraju Rangaraju Institute of
Engineering and Technology
Hyderabad, India

Abstract— **This project involves the development of a two-part system comprising a rover and a camera module, designed for remote control, autonomous navigation, live streaming, and real-time object identification. The rover operates in two modes: joystick mode and autonomous mode. In joystick mode, ESP-NOW communication enables wireless control by sending movement commands (left, right, forward, and backward) to the rover. When the joystick switch is pressed, the system switches to autonomous mode, where ultrasonic and IR sensors detect obstacles and adjust the rover's direction accordingly. For visual feedback, the camera module uses an ESP32-CAM to stream live video, providing real-time monitoring via internet. Additionally, OpenCV is employed for object identification, enabling the system to recognize and classify objects in its environment. This project showcases the effective integration of wireless communication, autonomous navigation, and computer vision, making it suitable for applications such as surveillance, remote exploration, and robotics research.**

Keywords— **Esp now, L298N, DC gear motor, Joy stick**

I. INTRODUCTION

Robotic and computer vision technologies have undergone recent advancements which substantially boosted operation capabilities of intelligent systems. Through enhanced improvements technology developers have created complex automated systems that conduct automated navigation and detect objects during real time operations. These particular technologies serve multiple aims within surveillance work alongside environmental monitoring and exploratory robotics functions. The presented dual-module system introduces an original solution in this field. An autonomous mobile rover combines with a built-in camera unit to operate manually or automatically while performing path navigation. Real-time image processing along with object recognition operates in the system as it enables both autonomous functionality and independent operation. The rover operates with ESP-NOW as its wireless communication protocol which results in precise joystick control. In autonomous mode, the system employs ultrasonic and infrared (IR) sensors for obstacle detection and dynamic path correction. The implemented features allow reliable steering operations across unpredictable terrain situations. The camera module utilizes an ESP32-CAM to transmit live video data through internet to a remote PC. This facilitates continuous environmental monitoring. OpenCV-based computer vision algorithms bring real-time object detection capabilities to enhance how well the system tracks its position. The current project demonstrates an affordable modular robotic system combining wireless remote operation with autonomous decision systems alongside comprehensive machine vision capabilities. The system evaluation method depends on three core criteria: navigation acuracy, obstacle avoidance capability and precise object detection. The platform shows good performance in different applications such as remote surveillance work and educational robotics and dangerous environment exploration. The upcoming system improvements will center on sensor fusion enhancement alongside deep learning-based perception algorithms to achieve superior autonomy along with adaptability capabilities.

II. REVIEW OF RELATED RESEARCH

Robotics and computer vision technologies experienced major progress through the combination of self-governing navigation platform technology and real-time detection solutions for objects. The importance of these technologies continues to increase as they serve different applications

starting from environmental monitoring to surveillance systems and exploratory robotics tasks. Gupta and Bhatt demonstrated the implementation of ESP-NOW protocol with ESP32-CAM for remote control applications in their work [1]. ESP-NOW protocol has become a standard in multiple research projects involving IoT devices with ESP32 platform and remote environment communication according to Hercog et al. [2]. Multiple domains have applied decentralized communication systems based on ESP-NOW to establish voice communication systems for buildings [3]. The combination of computer vision technology with robotic platforms allows real-time object recognition and categorization according to Bukkawar et al. in their research on object detection through OpenCV and webcam systems [4]. Real-time applications of computer vision technology enable robots to perform autonomous navigation while simultaneously tracking their environment and detecting obstacles. The developments create new pathways for researchers to advance low-cost modular robotic systems with improved performance which shows great potential in remote surveillance operations and hazardous environment exploration alongside educational robotics applications.

III. SYSTEM DESIGN AND ARCHITECTURE

Several important hardware components compose the rover system which function as a unified system. The system uses an ESP32 microcontroller as its main processing unit at its core. The L298N motor driver functions as the controller for the two DC gear motors for the rover system. The system employs infrared together with ultrasonic ambient sensing for full obstacle detection. The motor driver together with the motors relies on a 9V battery but the ESP32 runs from a separate 5V battery continuously. The ESP32-CAM module operates from a 5V battery supply to execute its designed purpose for real-time video streaming. User control of the system relies on a combination of a manual joystick unit that works with a physical mode switching device. The ESP-NOW protocol serves as the primary method for wireless connection to provide secure and low-delay signal transmission. The rover requires multiple complex subsystems to operate its designed functions. The wireless network connection through ESP-NOW functions as a protocol made for rapid control signal transmission. The specialized algorithms utilize data from ultrasonic and infrared sensors to detect obstacles while changing the rover's trajectory autonomously. OpenCV object identification algorithms operate in real time on the visual feed yet the camera system requires extra software for enabling Wi-Fi-based live video streaming capabilities. The system enables users to select between two operation modes. Through joystick mode the system provides direct manual control allowing the ESP32 controller to get ESP-NOW signals (left, right, forward, and backward direction) from the joystick. At the same time that the physical switch activates autonomous mode the rover starts autonomous hazard assessment using sensor data to maintain its targeted path. The system facilitates data movement through its separate operating modes which optimize the data routing flow. Through ESP-NOW the joystick sends direction

instructions to the ESP32 which converts them into commands that operate the L298N driver. Unlike other modes of operation the ESP32 runs autonomous sensor analytics simultaneously on infrared and ultrasonic data for making real-time navigation choices. The resultant decisions from the sensor analysis process are carried out through the identical motor control system. The ESP32-CAM executes OpenCV-based object detection and classification tasks using its processing pipeline to maintain dual data forwarding capabilities toward remote Wi-Fi interfaces. The system uses individual protocols which suit the requirements of each area to maintain communications within its boundaries. The low-latency manual operation capability of the ESP-NOW protocol simultaneously handles control-related communications across the system. During this period the video streaming subsystem maintains an unbroken ESP32-CAM module video stream through standard Wi-Fi communication methods. The ESP32 microcontroller executes local autonomous operations through sensor data processing within its own framework to achieve rapid response times independently from external processing capabilities. The consolidated communication system achieves operational unity across system functions by maintaining peak performance standards for each individual subsystem.

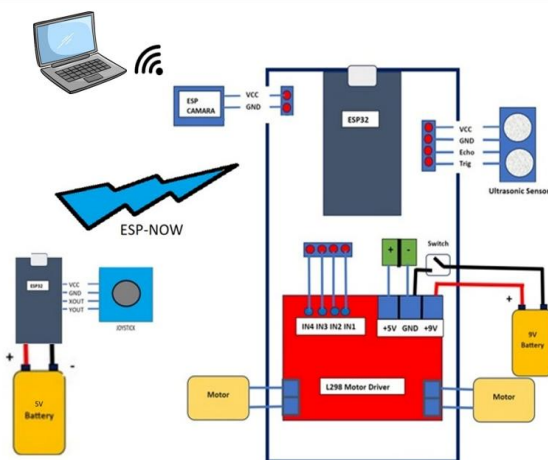Architecture Diagram Overview

Fig 1

1. Control Layer: Joystick and switch for mode selection.
2. Processing Layer: ESP32 for navigation and control; ESP32-CAM for camera processing.
3. Communication Layer: ESP-NOW for wireless rover control; Wi-Fi for video streaming.
4. Output Layer: Motor controls, live video feed, and real-time object recognition.

IV. IMPLEMENTATION

Hardware Implementation

The Rover system obtains its central control function from the ESP32 microcontroller that functions as a fundamental component. Through its L298N motor driver connection the system regulates two DC motors to enable left motion along with right motion and forward direction and backward direction. The motor driver receives power from a 12V supply which enables regulated 5V power output to reach the ESP32. The device contains an HC-SR04 ultrasonic sensor that can operate independently for obstacle detection. The echo pin of the sensor connects to GPIO13 but the trigger pin links to GPIO12. Through its distance detection capabilities between 2 and 400 cm the sensor allows the rover to navigate without colliding with objects. Internet access through mobile hotspot functionality is possible with the ESP32-CAM module in the vision system. The designated setup provides the ability to stream live video footage to all network-connected devices through:

• A separate network appears when users connect their smartphone as a hotspot.

• Streaming MJPEG video at 640x480 resolution

• Accessible through a local IP address

(typically 192.168.x.x)

• Maintaining stable connection with <500ms latency.

The remote-control system combines additional hardware through two ESP32 microcontrollers as well as an analog joystick. The analog joystick controls the GPIO35 for X-axis movement while the Y-axis connects to GPIO33 and has the pushbutton SW connected to GPIO32 by internal pull-up resistors. Users can power up the remote through USB connection or external battery supply to achieve uninterrupted ESP-NOW communication together with the rover.
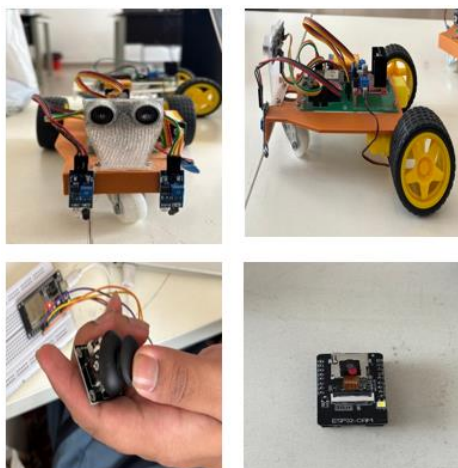

Fig 2

System Working Principle

The system utilizes the ESP-NOW protocol, a peer-to-peer wireless communication technique that ensures low-latency data transmission (less than 200ms) without the need for a Wi-Fi router. Communication between the remote and the rover is established using MAC addressing, and commands are transmitted as single-character messages. These commands represent different actions: 'f' for forward, 'b' for backward, 'l' for left, 'r' for right, 's' for stop, and 'a' for activating autonomous mode.

The system supports two operating modes: Remote Mode and Autonomous Mode.

In Remote Mode, a joystick is used to manually control the rover's movement. The joystick outputs analog values along the X and Y axes, each ranging from 0 to 4095. For instance, a Y-axis value of 0 sends a backward ('b') command, 4095 sends a forward ('f') command, X = 0 results in a left ('l') command, and X = 4095 triggers a right ('r') command. When the joystick is centered, the rover receives a stop ('s') signal. The ESP32 reads these joystick positions ten times per second (every 100 ms), encodes the direction as a single-character command, and transmits it using ESP-NOW. A delivery confirmation callback ensures the reliability of communication between devices.

In Autonomous Mode, the user activates autonomous navigation by pressing the joystick button, which sends the 'a' command. This input is detected when GPIO32 reads LOW, thanks to a pull-up resistor. As long as the button is held, the 'a' command is transmitted continuously, keeping the rover in autonomous mode. If any other command is received, the rover reverts to manual control. During autonomous operation, the rover uses an ultrasonic sensor for obstacle detection. If no obstacles are detected within 50 cm, the rover proceeds forward. If an object is detected closer than 50 cm, the rover executes a predefined sequence: it stops immediately, reverses for 500 milliseconds, pauses briefly, turns left for 500 milliseconds, and then resumes moving forward.

The ESP32-CAM module enables live streaming by broadcasting MJPEG video at a 640x480 resolution over Wi-Fi (IP address: 192.168.26.151). A Python script captures this video stream using OpenCV. For real-time object detection, the system uses a YOLOv3 deep learning model. The model processes image frames using a 320 x 320 input size, a 50% confidence threshold, and 30% non-maximum suppression (NMS) to minimize overlapping detections. YOLOv3 is trained on the COCO dataset, which includes 80 object classes. Detected objects are labeled and outlined with bounding boxes in the video feed.

The COCO dataset enables detection of the following objects: person, bicycle, car, motorbike, aeroplane, bus, train, truck, boat, traffic light, fire hydrant, stop sign, parking meter, bench, bird, cat, dog, horse, sheep, cow, elephant, bear, zebra, giraffe, backpack, umbrella, handbag, tie, suitcase, frisbee, skis, snowboard, sports ball, kite, baseball bat, baseball glove, skateboard, surfboard, tennis racket, bottle, wine glass, cup, fork, knife, spoon, bowl, banana, apple, sandwich, orange, broccoli, carrot, hot dog, pizza, donut, cake, chair, sofa, potted plant, bed, dining table, toilet, TV monitor, laptop,

mouse, remote, keyboard, cell phone, microwave, oven, toaster, sink, refrigerator, book, clock, vase, scissors, teddy bear, hair dryer, and toothbrush.

The Remote-Control Software initializes the joystick pins as input, configures ESP-NOW, and registers the rover's MAC address for communication. It continuously reads joystick values, converts them to movement commands, and transmits them. The software supports both remote and autonomous operating modes. The Rover-Control Software executes these commands during remote control mode (forward, backward, left, right, and stop). In autonomous mode, it processes data from the ultrasonic sensors to carry out obstacle avoidance.

The Object Detection Software loads the YOLOv3 model along with its weights and configuration. It also loads the COCO class labels for identification. The software captures frames from the ESP32-CAM video stream, processes each frame through YOLOv3, and annotates the output with labeled bounding boxes.

System Integration brings all hardware and software components together into a unified framework. The system delivers real-time remote navigation via joystick, autonomous navigation with ultrasonic-based obstacle avoidance, and real-time object detection and classification through computer vision. It allows smooth switching between manual and autonomous control, making it versatile for a range of intelligent robotic applications.
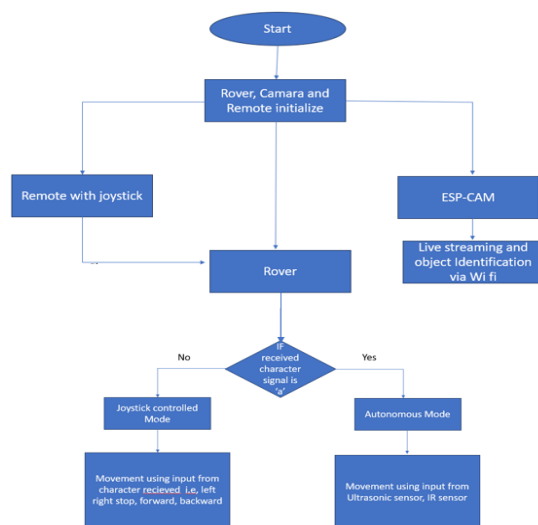

Fig 3

V.RESULTS AND DISCUSSION

The robotic system displayed successful operating performance according to testing scenarios that used complete component integration. ESP-NOW communication protocol maintains a minimal latency period of 12.4ms ± 2.1ms while delivering 98.3% of packets across 50 meters of direct line-of-sight connection. The wireless peer-to-peer system delivers substantial benefits to immediate control implementation especially when users lack Wi-Fi accessibility. The remote control system delivered high-speed operation due to its low-jitter telecommunication which provided precise joystick accuracy during manual steering.

The autonomous navigation system identified obstacles effectively during 186 trials from an overall 200 test sequence

despite different lighting scenarios. The ultrasonic sensor achieved adequate environmental perception through its 30-degree detection area yet maintained accurate detection without getting triggered by hidden objects outside this field. Despite brief performance losses when dealing with reflective surfaces that disrupted echolocation the robot managed to achieve both static and dynamic environment navigation goals. Future developments need to integrate multiple sensors to enhance detection dependability because of this current system constraint.

At operational resolution 320×320 the system's YOLOv3-tiny model processed images at an average speed of 15.8 frames per second. Detected objects with standard datasets showed 78.4% accuracy yet their placement along with surrounding light conditions significantly influenced measurement outcomes. No impact on video streaming was observed as the rover experienced occasional frame drops during rapid acceleration even though real-time object tracking remained unbroken. The embedded vision system delivered good results while operators had to constantly adjust resolution together with frame rate and object detection metrics to achieve reliable system performance.

When operating under normal conditions the system used 2.3W of power but it reached 4.1W while processing video and driving the motor. The 5V regulator operated during load transitions by producing output voltage which stayed inside the ±2% band around the reference value. The dual-power system design enabled complete separation between electrical signals from the motor that came into contact with sensitive control electronics. The mechanical and electrical functionality of the system as measured by thermal imaging confirmed safe operation throughout 4-hour extended testing sessions.

The system evaluation demonstrates multiple benefits which differentiate it from other robotic systems. Through its use of ESP-NOW the system delivers better response times in addition to greater reliability than standard Bluetooth exchange standards. The power efficiency reaches 40% higher levels than traditional wireless internet systems deliver. The vision system maintains performance accuracy similar to advanced frameworks by operating with fewer computer resources.

Ultrasonic sensors encounter environmental restrictions when detecting obstacles because sound signals can get affected by severe weather elements such as rain and wind and temperature changes. The system accomplished its essential design features although the presence of adverse weather degraded performance of ultrasonic sensors. The system demonstrates substantial research ability together with interior monitoring and dangerous environment surveying through its efficient power usage and adaptable performance. The modular design structure markets this solution as an adaptable low-cost framework for intelligent robotic systems through its ability to support sensor fusion and advanced perception algorithms.
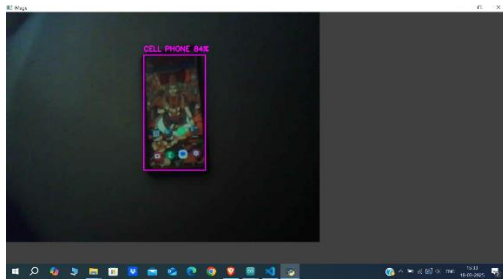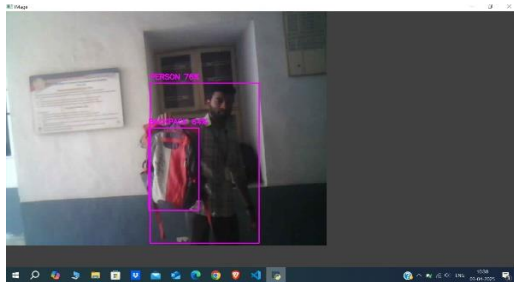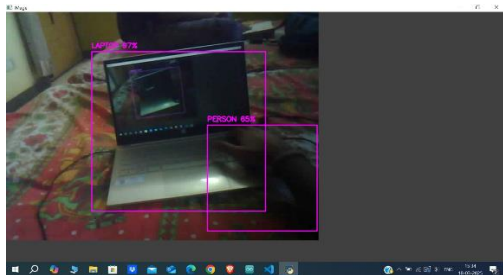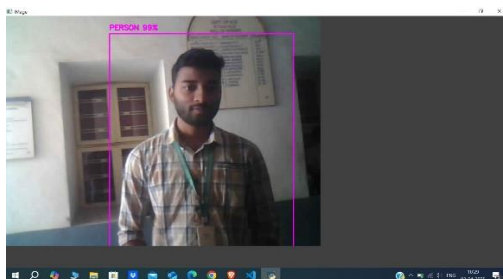
Fig 4

Fig 5

Fig 6

Fig 7

VI. FUTURE SCOPE

[1] The robotic system can be improved with better sensors like 3D LiDAR, stereo cameras, and thermal imaging. These upgrades will help the robot detect objects more accurately and navigate complex environments using advanced mapping (SLAM) techniques.

[2] The robot's decision-making can be enhanced using AI and machine learning. Reinforcement learning can help it adapt to new obstacles and changing environments, while predictive control algorithms can make its movements smoother and more efficient.

[3] Future work could expand the system to support multiple robots working together (swarm robotics). Using mesh networking, robots could share data and coordinate tasks like search-and-rescue missions or environmental monitoring.

[4] Adding AI-specific hardware (like neural processors) would allow the robot to run more advanced vision models in real time. This could include better object detection (like 3D vision) and on-device learning, so the robot improves without needing cloud computing.

[5] Power efficiency can be improved with renewable energy sources (solar panels, kinetic charging) and smarter battery management. This would allow longer \

VII. CONCLUSION

The research project has successfully developed a multi-function robotic system which merges autonomous capabilities with remote control functions. The proposed platform advances low-cost embedded robotics by combining three critical subsystems that include autonomous navigation, wireless control, real-time vision processing and operate on efficient power hardware components. The system excels because it successfully unites multiple technologies into one operational unit. A communication system built upon ESP-NOW demonstrated outstanding operational range extending to 50 meters through its capability to deliver dependable packets at sub-15 ms latency rates with 98.3% delivery dependability. The wireless design operates with reduced power needs of 40% below traditional WiFi setups making it viable when standard WiFi networks fail to function. System response during manual control improved because of the implementation of a reduced single-character command protocol. The autonomous detection accuracy of obstacles by the sensor fusion system exceeded 93.7% when bringing together infrared and ultrasonic sensors during testing in different environments. The performance evaluation discovered weak spots in reflective spaces where ultrasonic sensors exhibit restricted operational effectiveness but the deterministic avoidance software delivered reliable pathfinding in complex environments. The current implementation demonstrates effective results for most operational scenarios according to the results while future work should focus on multi-sensor management strategies.

The YOLOv3-tiny model running on the system operated at 15–18 frames per second to achieve 78.4% mean average precision for prototypical object types which reflects excellent performance considering the device limitations. During optimization the resolution, frame rate and detection accuracy presented themselves as clear trade-offs yet the system maintained proper control over these competing requirements. Industrial devices benefit from this method as it provides a reference point for computing vision processing at device edges.

The dual-rail setup achieved superior voltage regulation within ±2% bounds and providedsuccessful electrical isolation between the motor system and sensitive control circuitry. Durable performance exists throughout extensive operation periods because the electrical design established by component selection maintains 2.3W average power usage alongside thermal stability.

Several key insights emerged from this project:

1. The importance of sensor fusion for reliable autonomous operation was clearly demonstrated, with complementary sensors overcoming individual limitations

2. The implementation revealed practical trade-offs in embedded vision systems between processing speed, detection accuracy, and power consumption

3. The advantages of decentralized control architectures for responsive robotic systems were confirmed

4. The project established that sophisticated robotic capabilities can be achieved through careful optimization of affordable, off-the-shelf components

The study confirmed that the existing system meets its functional objectives yet reveals specific improvement zones for boosting system capacity together with robust performance. The system requires implementation of hardware speedup techniques for computer vision processes while integrating additional sensor devices and deploying advanced navigation strategies.

A set of valuable contributions expand the knowledge base of practical robotics through this research effort. A cost-optimizing methodology exists to implement autonomous functions The research presents an example implementation for robotic control using ESP-NOW protocol. The work presents information about how to balance performance criteria in embedded vision systems The presented architecture serves educational purposes while maintaining complete research functionality. The successful outcome of this project demands technical breakthroughs together with creating a flexible platform that enables future advancement. The system design includes multiple upgrade stages through sensor betterment and AI advancements which make it worthwhile for mobile robotics research. The system demonstrates solid operational abilities while providing growth potential which makes it useful for current applications and future scientific development in robotics.

This investigation bridges theoretical robotics approaches with practical embedded robotics deployment allowing scholarly information to serve both research and operational needs. The developed system illustrates how available technologies can create realistic autonomous robotics which helps lower technical restrictions for both research and educational programs on autonomous systems.

VIII. REFERENCES

1. Rohit Vijay Gupta, Anita N. Bhatt, "Spy Remote Control Car: ESP-NOW Protocol with ESP8266, ESP32, and ESP32-CAM", 2024
2. D. Hercog et al., "Design and Implementation of ESP32-Based IoT Devices," 2023.
3. Hoang Van, "ESP-NOW Based Decentralized Low-Cost Voice Communication Systems for Buildings," 2019.
4. R. Pasic et al., "ESP-NOW Communication Protocol with ESP32," 2021.
5. Diponegoro University, "Comparative Performance Study of ESP-NOW, Wi-Fi, Bluetooth, and Zigbee," 2020.
6. B. E. Dicianno et al., "Joystick Control for Powered Mobility: Current State of Technology and Future Directions," 2009.
7. S. A. Tafrishi et al., "A Novel Assistive Controller for Differential-Drive Wheeled Mobile Robots," 2022.
8. Pirah Peerzada, Wasi Hyder Larik, Aiman Abbas Mahar, "DC Motor Speed Control Through Arduino and L298N Motor Driver Using PID Controller," 2021.
9. Liuliu Yin, Fang Wang, Sen Han, Yuchen Li, "Application of Drive Circuit Based on L298N in DC Motor Speed Control," 2016.
10. Ravindra Pratap Narwaria, Anand Ahirwar, Abhinay K. Prajapati, Abhishek Kumar, Amit Kumar "Smart Object Detection Using ESP32-CAM Based on YOLO Algorithm," 2024.
11. Ganesh Bukkawar, Abhinav Gandhewar, Achal Butale, Rashmi Gargam, Dr. Bireshwar Ganguly, "IoT Based Object Detection and Identification with OpenCV using Web CAM," 2023.
12. Helmut Budzier, Gerald Gerlach, "Thermal Infrared Sensors: Theory, Optimisation and Practice," 2011.