3D Coordinate Detection System Using YOLO and Ultrasonic Sensing

[1]Milind Lohar , [2]Niranjan Pagere , [3]Ayush Kirte, [4]Aryan Paparkar, [5]Rahul Waikar
[1-5]Department of Mechanical Engineering,
Vishwakarma Institute of Technology, Pune, India
[1] milind.lohar23@vit.edu, [2] niranjan.pagere22@vit.edu, [4]Ayush.kirte23@vit.edu

[4]aryan.paparkar22@vit.edu, [5]rahul.waikar@vit.edu

Abstract: **Accurate localization of objects in three-dimensional (3D) space is essential for numerousapplications in robotics, automation, and intelligent systems. This project presents a low-cost, real-time 3D coordinate detection system that combines computer vision and range sensing techniques. To find the X and Y coordinates of an object in the image frame, the suggested system combines a webcam with an Open CV-based YOLO (You Only Look Once) object detection model. Concurrently, the Z coordinate is obtained by measuring the distance between the sensor and the object using an ultrasonic sensor. The system can estimate the object's XYZ position in space by combining data from both modules. The method is suitable for a variety of embedded and robotic applications since it provides a useful and effective way for 3D localization with little hardware and open-source tools.**

Key words: **- Ultrasonic sensor, Yolo, 3D Coordinate**

I. INTRODUCTION

In many contemporary applications, such as robotics, automation, surveillance, and smart systems, the ability to detect an object's position in three dimensions is crucial. Although they are frequently expensive and complicated, traditional techniques like LiDAR and stereo vision provide excellent precision. The goal of this project is to combine computer vision and ultrasonic sensing to create an affordable and effective 3D coordinate detection system. The system detects the object and determines its X and Y coordinates in the image frame using a webcam and an Open CV YOLO (You Only Look Once) model. The Z coordinate is obtained by measuring the distance to the object using an ultrasonic sensor. When combined, these sensors enable the XYZ position of the object to be estimated in real time.

Application**:** This solution is lightweight, cost-effective, and suitable for various practical applications such as- robotic arm positioning, object tracking in automation systems, collision avoidance, and smart inventory management, etc.

II. LITERATURE REVIEW

3D object localization is an active area of research in computer vision and robotics. [4] Traditional approaches rely on depth sensors such as stereo cameras, LiDAR, or structured light systems to determine the position of objects in three-dimensional space. Despite their high accuracy, these techniques are less appropriate for small-scale or cost-sensitive applications because they are frequently costly and require complicated calibration.

[1] Strong object detection algorithms like YOLO (You Only Look Once), which provide real-time performance and high accuracy, have been made possible by recent developments in deep learning. YOLO models are widely used in many vision-based systems and are especially good at detecting 2D objects in video streams. Nevertheless, conventional YOLO implementations do not offer depth information

[2] To bridge this gap, sensor fusion techniques are increasingly being explored. Combining vision-based detection with Range-finding sensors such as ultrasonic, infrared, or time-of-flight sensors allows for a more completes patial understanding of the environment. Ultrasonic sensors, in particular, are widely used in embedded and robotic systems due to their simplicity, low cost, and effective short-range depth estimation.

[3] In this context, integrating YOLO-based object detection with ultra-sonic sensing

III. **METHODOLOGY**

The proposed system estimates the 3D coordinates of an object by combining 2D visual data from a camera with depth data from an ultrasonic sensor. The methodology is divided into two main subsystems: vision-based object detection and distance measurement.

A. The Material used for frame is Aluminum.

A.Vision-Based Object Detection (X, Y Coordinates)

A standard USB webcam is used to capture live video frames of the environment. These frames are processed using a pre-trained YOLO (You Only Look once) object detection model integrated with the Open CV library in Python. YOLO detects the object of interest in the frame and returns a bounding box around it. From this bounding box, the center point is calculated, which represents the X and Y coordinates of the object within the image plane. These coordinates are relative and represent the object's position from the camera's point of view.

B. Ultrasonic Sensing (Z Coordinate)

An ultrasonic sensor is used to measure the distance from the sensor to the object along the Z-axis. The sensor works by emitting ultrasonic waves and calculating the time taken for the echo to return after hitting the object. It Provides a practical and accessible method for low-cost 3D localization. This hybrid approach leverages the strengths of both technologies: YOLO's ability to recognize and locate objects in 2D space, and the ultrasonic sensor's capability to measure distance along the Z-axis. Such a system can serve as a foundational framework for various applications in automation, robotics, and human-computer interaction. object. The distance is then calculated using the formula :

$$\text{Distance (Z)} = \frac{\text{Speed of Sound} \times \text{Time}}{2}$$

This distance represents how far the object is from the sensor, providing the Z coordinate in real-world units (typically centimeters).

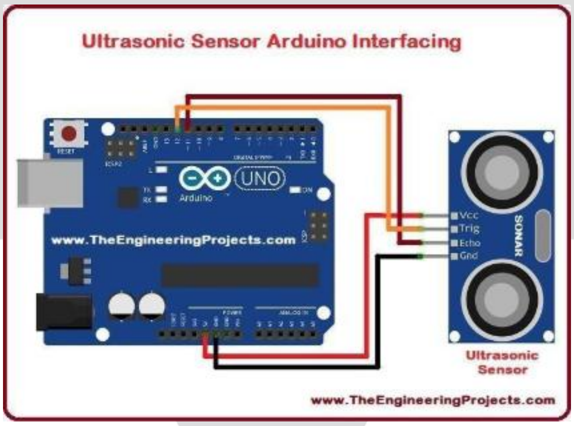

Figure 1. Ultrasonic Sensor Arduino Interfacing

C. Coordinate Fusion

The data from both sub systems are processed independently and the combined. The X and Y coordinates from the YOLO detection are used to determine the object's horizontal and vertical position, while the Z coordinate from the ultrasonic sensor provides the depth. The fusion of these three values gives a complete (X, Y, Z) coordinate set representing the object's position in 3D space.

ultrasonic

```
#define TRIG_PIN 6
#define ECHO_PIN 9

void setup() {
    Serial.begin(9600);
    pinMode(TRIG_PIN, OUTPUT);
    pinMode(ECHO_PIN, INPUT);
}

void loop() {
    long duration;
    float distance;


    digitalWrite(TRIG_PIN, LOW);
    delayMicroseconds(2);
    digitalWrite(TRIG_PIN, HIGH);
    delayMicroseconds(10);
    digitalWrite(TRIG_PIN, LOW);


    duration = pulseIn(ECHO_PIN, HIGH);

    distance = (duration * 0.0343) / 2;

    Serial.print("Distance: ");
    Serial.print(distance);
    Serial.println(" cm");

    delay(100);
}
```

Figure 2. Code for Ultrasonic

```python
import torch
import cv2
import numpy as np

# Load YOLOv5s model from Ultralytics (auto-downloads if not present)
model = torch.hub.load('ultralytics/yolov5', 'yolov5s', pretrained=True)
model.eval()

# Open webcam
cap = cv2.VideoCapture(0)  # 0 = default webcam

if not cap.isOpened():
    print("Error: Cannot open webcam.")
    exit()

print("Press 'q' to quit.")

while True:
    ret, frame = cap.read()
    if not ret:
        break

    # Convert frame to RGB
    img_rgb = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)

    # Inference
    results = model(img_rgb)

    # Extract detections
    for *xyxy, conf, cls in results.xyxy[0]:
        x1, y1, x2, y2 = map(int, xyxy)
        x_center = int((x1 + x2) / 2)
        y_center = int((y1 + y2) / 2)

        # Draw bounding box and center point
        cv2.rectangle(frame, (x1, y1), (x2, y2), (0, 255, 0), 2)
        cv2.circle(frame, (x_center, y_center), 5, (0, 0, 255), -1)
```

Figure 3. Coding of YOLO Module (part 1)

```python

        # Print coordinates
        print(f"Object: Center = ({x_center}, {y_center})")

    # Show the frame
    cv2.imshow('YOLOv5 Real-Time Detection', frame)

    # Exit on pressing 'q'
    if cv2.waitKey(1) & 0xFF == ord('q'):
        break

# Cleanup
cap.release()
cv2.destroyAllWindows()
```

Figure 4. Coding of YOLO Module (part 2)

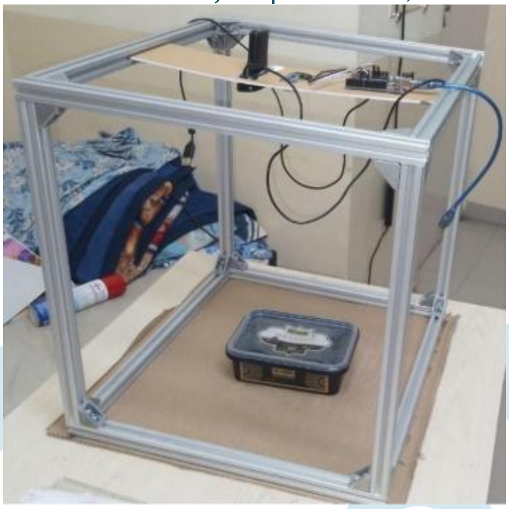

Figure 5. Aluminium Frame Setup

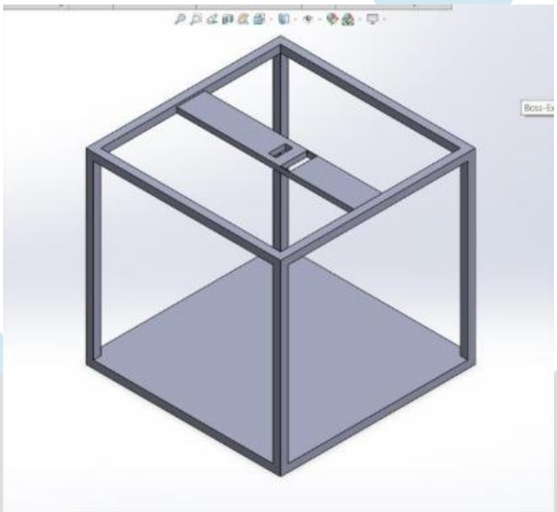

Figure 6. CAD Model

Figure 7. Final Output

IV. CONCLUSION

This project successfully demonstrates a low-cost and efficient method for detecting the 3D coordinates of an object using a combination of compute revision and ultrasonic sensing. By integrating a YOLO based object detection model with OpenCV for 2D (X, Y) localization and an HC-SR04 ultrasonic sensor for Z-axis depth measurement, the system provides a practical solution for basic 3D positioning tasks.

The approach minimizes hardware complexity while maintaining functional Accuracy, making it suitable for a wide range of applications such as robotic navigation, object tracking, and intelligent automation systems. The modular nature of the system also allows for future enhancements, such as integrating more precise sensors or real-world calibration to map image coordinates to actual physical dimensions.

Overall, this project highlights the potential of combining AI-based vision with traditional sensors to build intelligent mechatronic systems for real-world use.

V. REFERENCES

[1] S. H. Venugopala, "Comparative study of 3D object detection frameworks based on LiDAR data and sensor fusion techniques," arXivpreprintarXiv:2202.02521, Feb.2022.

[2] Aniobi, A. O. (2024). Sensor Fusion for Real-Time Object Detection and Spatial Positioning in Unmanned Vehicles Using YOLOv8 and ESP32-Cam. Preprints, 202411011.

[3] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv:1804.0277.

[4] R. Murendeni, A. Mwanza, and I. C. Obagbuwa, "Using a YOLO Deep Learning Algorithm to Improve the Accuracy of 3D Object Detection by Autonomous Vehicles," World Electr. Veh. J., vol. 1, no. 1, p. 9, 2025, doi:10.3390/wevj1010009