Security System with Autonomous Anomaly Detection and Response Robot

PROJECT REPORT

Submitted by

TEAM C09
I.CHARUHAS REDDY - (CB.SC.U4AIE23233)
CB HARSHA VARDHAN - (CB.SC.U4AIE23220)
C VISHNUVARDHAN CHOWDARY (CB.SC.U4AIE23219)
Y BHASITH SAI REDDY- (CB.SC.U4AIE23264)

in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE ENGINEERING - ARTIFICIAL INTELLIGENCE

COMPUTER SCIENCE ENGINEERING - ARTIFICIAL INTELLIGENCE

AMRITA SCHOOL OF ARTIFICIAL INTELLIGENCE

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112 (INDIA)

APRIL - 2025

COMPUTER SCIENCE ENGINEERING - ARTIFICIAL INTELLIGENCE

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112

AMRITA VISHWA VIDYAPEETHAM

श्रद्धावान् लभते ज्ञानम्

DEEMED TO BE UNIVERSITY UNDER SECTION 3 OF THE UGC ACT, 1956

BONAFIDE CERTIFICATE

This is to certify that the thesis entitled **'Security System with Autonomous Anomaly Detection and Response Robot"** submitted by **I CHARUHAS REDDY- CB.SC.U4AIE23233), CB HARSHAVARDHAN - (CB.SC.U4AIE23220), C VISHNUVARDHAN CHOWDARY - (CB.SC.U4AIE23219), Y BHASITH SAI REDDY- (CB.SC.U4AIE23264)** for the award of the **Degree of Bachelor of Technology** in the **'COMPUTER SCIENCE ENGINEERING - ARTIFICIAL INTELLIGENCE"** is a bonafide record of the work carried out by them under my guidance and supervision at Amrita School of Artificial Intelligence, Coimbatore.

Dr. D. Palmani
Project Guide
Senior Grade Assistant Professor

Dr. Milton Mondal
Project Guide
Assistant Professor

Submitted for the university examination held on 16/04/2025

INTERNAL EXAMINER **EXTERNAL EXAMINER**

AMRITA SCHOOL OF ARTIFICIAL INTELLIGENCE

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112

DECLARATION

We, **I CHARUHAS REDDY- (CB.SC.U4AIE23233), CB HARSHAVARDHAN - (CB.SC.U4AIE23220), C VISHNUVARDHAN CHOWDARY - (CB.SC.U4AIE23219), Y BHASITH SAI REDDY- (CB.SC.U4AIE23264)** hereby declare that this thesis entitled **"Security System with Autonomous Anomaly Detection and Response Robot"**, is the record of the original work done by us under the guidance of **Dr. D. Palmani** and **Dr. Milton Mondal**, Senior Grade Assistant Professor, Amrita School of Artificial Intelligence, Coimbatore. To the best of my knowledge this work has not formed the basis for the award of any degree/diploma/ associateship/fellowship/or a similar award to any candidate in any University.

Place: **Signature of the Student**

Date:

COUNTERSIGNED

Dr. K.P.Soman
Professor and Dean
Amrita Vishwa Vidyapeetham

Contents

Acknowledgement

We would like to express our sincere appreciation to all those who have helped and guided us step by step throughout this project. Most importantly, we are significantly grateful to our project guide, **Dr. D. Palmani** and **Dr. Milton Mondal** for their constant encouragement, expert advice, and constructive criticism. His guidance played a pivotal part in the direction and success of the project. We would also like to appreciate the staff and faculty at Amrita School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore, for providing us with the facilities, materials, and learning environment to deliver our research went without a hitch to completion. Second, we appreciate our peers and teammates for their cooperation, motivation, and zeal. Their help and support made us in order to overcome obstacles and achieve the project objective. Lastly, we would like to thank our families for supporting us throughout, for being patient with us, and for believing in us along the way during our learning process. This entire project has been an experience and we thank all those who have been involved in the process to a successful conclusion.

List of Figures

List of Tables

List of Abbreviations

Table 1: List of Abbreviations

Abbreviation	**Full Form**
AI	Artificial Intelligence
NLP	Natural Language Processing
API	Application Programming Interface
CNN	Convolutional Neural Network
CSV	Comma-Separated Values
DOF	Degrees of Freedom
FPS	Frames Per Second
YOLO	You Only Look Once
LLM	Large Language Model
SORT	Simple Online and Realtime Tracking
MiDaS	Monocular Depth Estimation with Small Model
LLM	Large Language Model
ROS	Robot Operating System

Abstract

Human-monitor-based security systems usually have issues with sluggish threat detection, lack of proper monitoring due to fatigue, and insufficient capacity to respond in evolving situations. This project addresses such problems through the development of a novel security system operating independently to identify and respond to anomalous events, based on a miniaturized computing platform. The system monitors real-time video got from the multiple cameras to detect suspicious actions such as lingering too long or carrying something threatening and dispatches a robot to investigate. The robot determines how people are feeling, whether they look frightened or anxious, and speaks to them in an intelligent, responsive manner to determine what is happening. This combined strategy accurately monitors people and objects, knows their actions and emotions, and talks effectively. Through the application of tracking, behavior analysis, emotion detection and interaction, it offers a versatile configuration for automated security. This project demonstrates it's competent at detecting issues, monitoring things, and knowing people, proving it gets the job done. This project establishes a good foundation for security systems, leaving room to expand in robot autonomy and usage.

Chapter 1

Introduction

This project creates an autonomous security system that combines real-time anomaly detection with a robotic response mechanism, deployed on the NVIDIA Jetson Nano platform. The system utilizes a high-level pipeline of artificial intelligence technologies, such as a fine-tuned YOLOv8 model for object detection, fine-tuned DeepSort for robust multi-object tracking, MediaPipe Pose for behavioral analysis, and MiDaS for depth-based spatial positioning. These parts co-exist to detect abnormalities in a live feed such as prolonged loitering or weapon presence and activate a robot to engage people using the natural language processing (NLP) module. The integration seeks to create an active, autonomous system that redefines contemporary security solutions using the computational power of an edge device such as the Jetson Nano.

1.1 Problem statement

Traditional surveillance systems heavily depend on human operators, leading to significant inefficiencies that compromise performance. These systems suffer from delayed anomaly detection due to slow human response times, inconsistent decision-making from fatigue during prolonged monitoring, and limited range of observation constrained by human perception, often resulting in missed threats. In addition, manual control and passive setups lack the ability to actively respond to detected anomalies, causing a disconnect between identification and action. This project tackles these challenges by developing an autonomous security robot system that is human-in-the-loop minimized, using a central framework with an array of cameras as feedback to monitor large areas, incorporating detection, tracking, and interaction capabilities for an active and effective response.

1.2 Literature Review

Table 1.1: Summary of Related Work

Ref.	Authors	Contribution Summary
[1]	Salna Joy et al. (2023)	This paper presents a Raspberry Pi-based patrolling robot that enhances security by detecting unusual sounds with sensors, capturing live video, and sending GPS coordinates to users via email.
[2]	Jun Zhang et al. (2011)	This paper introduces an indoor security system featuring a jumping robot that overcomes obstacles like stairs, using PIR sensors and a ZigBee network to detect intruders and send photos to homeowners via the Internet.
[3]	P. R. Sriram et al. (2023)	This paper proposes an autonomous quadcopter drone using a PixHawk controller for defense machinery maintenance and surveillance, integrating YOLOv3 for object detection and cloud storage via Amazon S3
[4]	Ruifeng Li et al. (2007)	This paper presents a general-purpose indoor security robot that uses an approximate arc motion model and Hough transform for localization and mapping, equipped with multi-sensor units for environmental monitoring.
[5]	Juan Bravo-Arrabal et al. (2025)	This paper explores enhancing multi-robot systems for search and rescue (SAR) by integrating advanced robotics with 6G communication technologies, improving coordination and efficiency.
[6]	Yungeun Choe et al. (2012)	This paper proposes a security robot system with a software architecture split into autonomous navigation (SLAM) and semantic perception for urban surveillance. It outlines a three-step patrol scenario: collecting environmental data, comparing it with prior maps, and alarming supervisors about anomalies.
[7]	Ho-chul Shin et al. (2023)	This paper presents an anomaly detection algorithm for a security robot, generating elevation and thermal maps using LIDAR and thermal imaging, with a two-stage classification process (auto-encoder and CNN) to identify abnormal situations.

1.3 Organization of Report

The report is structured into the following chapters:

- Introduction

  With increasing worries about safety in urban and industrial areas, conventional surveillance systems are beginning to look their age. Most of them still rely on human observation, are unable to adapt in real time, and tend to overlook subtle or sophisticated anomalies. This project introduces a wiser solution: an intelligent, autonomous security robot that can overcome these shortcomings. By integrating cutting-edge robotics, multi-sensor information, and artificial intelligence, the system proactively observes the environment, recognizes abnormal behavior, and moves through its surroundings with accuracy. The outcome is a forward-thinking, efficient, and scalable solution designed to address the changing needs of this century security.

- Background

  Develop an intelligent and adaptive security system. SLAM (Simultaneous Localization and Mapping) is employed by it to autonomously navigate the robot, including in unknown areas. Deep learning drives its object recognition and detecting suspicious activity capability, while numerous sensors—such as LiDAR combine to provide the system with a comprehensive understanding of the environment.

- Proposed Work

  Describes the complete system architecture including model inference pipeline.

- Conclusion

  Summarizes the key contributions of the system, discusses its performance and outlines possible directions for future enhancements.

1.4 Objectives

- To create an integrated, autonomous security system blending real-time anomaly detection, spatial location, and robotic interaction, optimised for running on the NVIDIA Jetson Nano platform.

- To train and tune the YOLOv8m object detection model using a Suspicious Detection Dataset from Roboflow.

- To use DeepSort and MediaPipe Pose to provide solid multi-object tracking and behavioral analysis, identifying the left baggage, flagging anomalies like prolonged loitering

- To deploy a convolutional neural network (CNN) for real-time emotion recognition, classifying facial expressions (e.g., fear, tension) to evaluate people's emotional states.

- To deploy a natural language processing (NLP) module, allowing adaptive, context-dialogue with people to gauge intent.

Chapter 2

Background

2.1 Introduction

This chapter introduces the core ideas and technologies behind our autonomous security system. This project uses a strong set of technologies to create an intelligent autonomous security robot system that surveillance using a centralized multi-camera system. The Robot Operating System (ROS) offers an platform, making path-planning easy and coordination of the robot possible to autonomously patrol large regions with real-time responsiveness. Object Detection and Anomaly Recognition: YOLO powers fast anomaly detection across multi-camera feeds, and MiDaS calculates depth from 2D images to improve spatial awareness of threats like intrusions or obstacles. Multi-Camera Fusion and Tracking: MediaPipe enables pose estimation and object detection in camera feeds, coupled with DeepSORT for multi-object tracking, allowing comprehensive coverage and extended monitoring across wide areas. These integrated technologies transcend conventional surveillance limitations by rapid threat detection, expanding the range of observation, and bridging the detection-action gap.

Chapter 3

Proposed Work

3.1 Overview

The autonomous security system that we are suggesting is designed to identify, monitor, and react to suspicious behavior in real time with a mobile robot. It uses emotion recognition and conversation to not only sense its environment but to understand human intent as well. With the NVIDIA Jetson Nano power, the solution combines state-of-the-art computer vision, smart navigation, and natural language processing to provide a smart, intuitive security solution that embodies physical action and cognitive awareness.

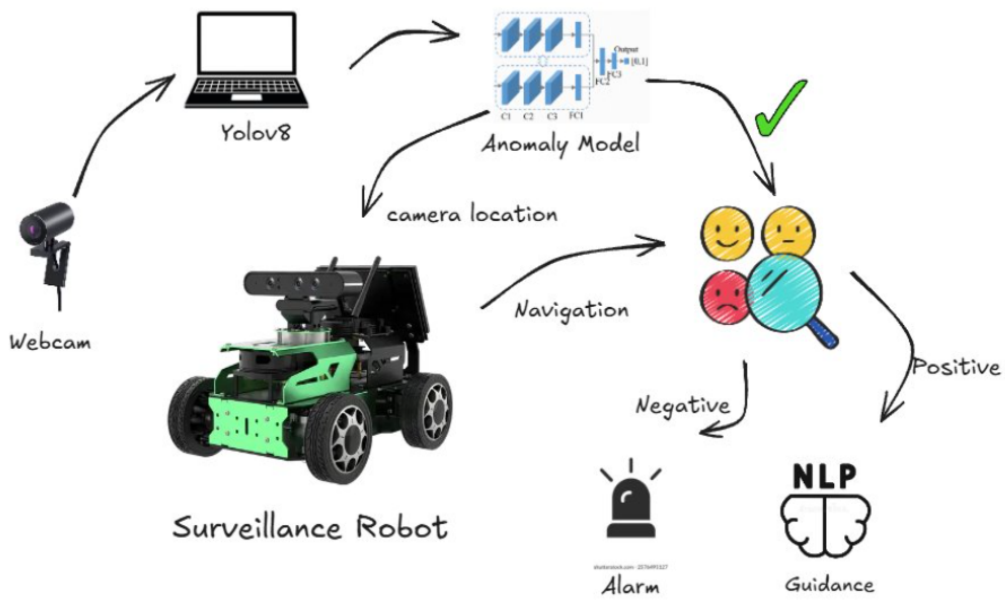

Figure 3.1: System architecture of the surveillance robot with YOLOv8, anomaly detection, navigation, and NLP response.

The system begins by taking live video streams from a few strategically placed cameras that cover the entire area of surveillance. In this way, all activity within its line of sight is closely monitored. The system, which is meant for dynamic scenes, processes one frame at a time to give real-time performance even on a lowly powered edge device such as the Jetson Nano. All the video frames are passed through a highly optimized YOLOv8m model trained on a personal Suspicious Detection Dataset from Roboflow. In comparison to the previous versions of YOLO that emphasized speed above all else, YOLOv8m gives a perfect balance of speed and accuracy with better feature extraction. This gives it enough reliability in the detection of large objects—such as knives, weapons, backpacks, and people even in challenging situations such as low light or when some objects are being partially occluded. After detection, each object is labeled as categories such as "weapons," "bags," or "people," and its location is followed by bounding box coordinates for further processing.

As soon as an object is identified, the system uses DeepSort to monitor its trajectory by allocating an individual ID for each object. Kalman filtering and deep learning-based association measurements power this monitoring so that the system can smoothly track objects throughout video frames. Concurrently, MediaPipe Pose examines body landmarks to detect aberrant behavior. For instance, it detects loitering when a person remains at same spot (inside a tiny 50x50 pixel box) for over 300 frames approximately 10 seconds at 30 frames per second. To know how close objects or individuals are to the camera, the MiDaS depth estimation model produces a depth map. It converts 2D pixel locations to actual world 3D coordinates (X, Y, Z), employing a pre-calibrated system that projects image data onto physical space. It makes certain that the robot has a clear idea of where to move when something unusual occurs.

When the system recognizes a potential threat such as an individual with a weapon or an individual hanging around for too long or individuals leaving their luggage in public areas it deploys the robot. Powered by the Robot Operating System (ROS) and LiDAR sensors, the robot begins processing its environment. ROS nodes utilize SLAM (Simultaneous Localization and Mapping) to combine with LiDAR sensor data with the depth estimates of MiDaS, for creating a real-time 3D map of the environment. Based on the map, the system determines a safe path and directs the robot to the point of concern directly using ROS's Navigation Stack.

When the system identifies a potential threat such as a person with a weapon or loitering, it deploys the robot. With LiDAR sensors and driven by the Robot Operating System (ROS), the robot begins to explore its environment. ROS nodes employ SLAM (Simultaneous Localization and Mapping) to fuse LiDAR data with depth estimates from MiDaS and construct a real-time 3D map of the world. The system uses this map to calculate a safe path and drive the robot straight to the area of concern using ROS's Navigation Stack.

At the same time, the system keeps an elaborate record of each anomaly that it encounters specifying the object type, where it was encountered, and the outcome of any interaction. For example, if the object encountered is a knife and the subject is clearly frightened, the robot is able to classify the incident as a threat and sound an alarm. On the other hand, if the person clearly states that it's a work tool calmly, the system may decide that there is no threat and back down. This feature is accomplished by combining both LLM and emotion detection model. These interactions are meant to be quick, relevant, and context-aware.

With the integration of smart perception, robot navigation, and human interaction, the system is a step above traditional surveillance. Object in the real world, unusual activity, emotional cues - whatever it is, the system can react in real time. Its flexibility makes it particularly well-suited to a broad range of environments, from residences and schools to public areas, with a scalable and self-sustaining method towards new security.

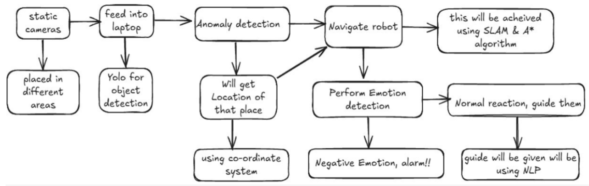

Figure 3.2: Flow diagram of the proposed autonomous security system.

3.2 YOLOv8

YOLOv8 is a speed and accurate optimized object detection model that detects objects directly from video frames by predicting class probabilities and bounding boxes in real-time. It is particularly suited for real-time surveillance purposes. In this project, we fine-tuned the medium model YOLOv8m with a Suspicious Detection Dataset provided by Roboflow. This model more efficient at recognizing security-specific items, like weapons like knives and guns. To enhance readability in monitoring, we also used color-coded bounding boxes: weapons in red, people in green, common objects in white, and unknown in yellow.

3.3 DeepSort for Multi-Object Tracking

DeepSort enhances the Simple Online and Realtime Tracking (SORT) algorithm with deep learning. It enhances multi-object tracking by predicting movement, but also identifying how objects appear. This renders it far more reliable in difficult scenarios such as when individuals or objects intersect, turn sharply, or become out of sight for a very brief period of time.

In our system, DeepSort employs Kalman filtering to predict where an object will probably go next, based on its past movement. This provides the system with a clever, probabilistic prediction of where each object will be in the subsequent frames. By integrating this motion prediction with appearance-based recognition, DeepSort can assign and hold a distinct ID for each person or object and track them even in challenging situations such as crowd scenes or temporary occlusion.

3.4 MediaPipe Pose for Behavioral Analysis

MediaPipe Pose is a lightweight, high-speed machine learning-based solution to estimate real-time human pose. It inspects video frames for significant body points such as nose, shoulders, elbows, hips, and knees and provides a close-up look at an individual's posture and movement.

MediaPipe Pose is also used here to monitor behavior and detect suspicious behavior. It's an additional level of intelligence over basic object detection because it monitors how individuals move or don't move. One very important behavior that it tries to detect is loitering, or where an individual is in a given location or does not move whatsoever for an extended amount of time.

To monitor this, the system maintains a dictionary named loitering time that tracks when an individual enters each small part of the video frame (named grid cells) for the first time. When a person stays at one place for a very long time, the system identifies it as loitering and marks the precise frame when the person is at the place at that time. This enables the robot to see not only what the people are carrying but also how people walk.

3.5 MiDaS for Depth Estimation

Monocular Depth Estimation with Small Model (MiDaS) estimates the relative depth of every pixel from the camera through deep neural networks. MiDaS, in this case, is used to produce 3-dimensional coordinates so that a human or an object can be positioned accurately within the monitoring range. MiDaS is used to scan every video frame to produce a depth map, which is converted into actionable X, Y, Z coordinates.

$$X = \frac{(u - c_x) \times Z}{f_x}, \quad Y = \frac{(v - c_y) \times Z}{f_y}$$

Figure 3.3: Conversion from image coordinates to 3D coordinates using intrinsic parameters

Where:

- f_x, f_y: Focal lengths in pixels (x and y directions)
- c_x, c_y: Principal point (image center) in pixels
- Z: Depth (distance along the camera's optical axis)

3.6 Convolutional Neural Networks (CNNs) for Emotion Detection

Convolutional Neural Networks (CNNs) are used to learn spatial features and execute classification tasks, especially in application cases such as facial emotion recognition. CNN is trained to specialize in this project using FER-2013 Dataset of 28,709 samples and the public test set has 3,589 samples. The data contains Angry (958 samples), Disgust (111 samples), Fear (1024 samples), Happy (1774 samples), Neutral (1233 samples), Sad (1247 samples), surprise (831 samples). It makes strong behavioral observations that improve the system's capability to identify anomalies. ReLU activation function is applied in the CNN layers to add non-linearity, improving the model's capability to identify subtle emotional clues. Processing emotional clues after detecting anomalies, it adds an additional layer of behavioral context that cannot be offered by physical detection.

3.6.1 Convolutional Layer

The convolution layer is defined as

$$O(i,j) = f\left(\sum_{m=0}^{2}\sum_{n=0}^{2} I(i+m, j+n) \cdot K(m,n) + b\right)$$

Where,

$O(i,j)$	: The output value at position (i,j) in the feature map, capture a detected feature.
$f(\cdot)$	: The ReLU activation function, defined as $f(x) = \max(0,x)$.
$\sum_{m=0}^{2}\sum_{n=0}^{2}$	: Double summation over a 3×3 filter.
$I(i+m, j+n)$	: The input image pixel value at shifted position $(i+m, j+n)$.
$K(m,n)$	: The filter (kernel) weight at position (m,n) in the 3×3 kernel.
b	: The bias term, a scalar added to the convolution output.

3.6.2 ReLU Activation Function

The ReLU (Rectified Linear Unit) function is defined as:

$$\phi(z) = \max(0, z)$$

Its derivative, used during backpropagation, is:

$$\phi'(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases}$$

3.6.3 Forward Propagation

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}$$

$$\mathbf{a}^{(l)} = \phi(\mathbf{z}^{(l)})$$

with the input layer defined as:

$$\mathbf{a}^{(0)} = \mathbf{x}$$

where,

- $\mathbf{x} \in \mathbb{R}^d$: input vector
- L: number of layers
- $\mathbf{W}^{(l)}$: weight matrix at layer l
- $\mathbf{b}^{(l)}$: bias vector at layer l
- $\mathbf{z}^{(l)}$: linear combination at layer l
- $\mathbf{a}^{(l)}$: output of layer l after ReLU

3.6.4 Loss Function

Assuming a regression problem with target $\mathbf{y}$ and prediction $\hat{\mathbf{y}} = \mathbf{a}^{(L)}$, the Mean Squared Error (MSE) loss is:

$$J = \frac{1}{2}\|\hat{\mathbf{y}} - \mathbf{y}\|^2$$

3.6.5 Backpropagation with ReLU

To train the network, we compute the gradient :

$$\delta^{(L)} = \nabla_{\mathbf{a}^{(L)}} J \odot \phi'(\mathbf{z}^{(L)})$$

$$\delta^{(l)} = \left((\mathbf{W}^{(l+1)})^T \delta^{(l+1)}\right) \odot \phi'(\mathbf{z}^{(l)})$$

$$\frac{\partial J}{\partial \mathbf{W}^{(l)}} = \delta^{(l)} (\mathbf{a}^{(l-1)})^T \qquad \frac{\partial J}{\partial \mathbf{b}^{(l)}} = \delta^{(l)}$$

3.6.6 Weight Update Rule

Using gradient descent:

$$\mathbf{W}^{(l)} \leftarrow \mathbf{W}^{(l)} - \eta \frac{\partial J}{\partial \mathbf{W}^{(l)}} \qquad \mathbf{b}^{(l)} \leftarrow \mathbf{b}^{(l)} - \eta \frac{\partial J}{\partial \mathbf{b}^{(l)}}$$

3.7 Natural Language Processing (LLM Integration)

Natural Language Processing (NLP) in the project allow robot to communicate with the people and to identify their intention. The Large Language Model that we have used is Llama 3.1 of 8 billion parameters. The system is combined with speech to text and text to speech feature. So that the user can communicate via voice and can also receive via voice. The model is fine tuned in a manner in which the questions are asked dynamically by the answers provided by the individuals. The provided answers are analyzed further and identified as suspicious or routine behavior

3.8 Simultaneous Localization and Mapping (SLAM)

SLAM (Simultaneous Localization and Mapping) enables a robot to create and maintain a map of a foreign environment as well as determine its own location in that space. In the project, SLAM is applied to assist the robot in traversing complicated, unstructured outdoor spaces effectively. It is integral to the system's real-time reaction to abnormalities such as noticing someone standing around or brandishing a weapon identified through the video pipeline analysis using YOLOv8 and MediaPipe Pose.

When an anomaly is detected, the system inserts itself into the video frame. SLAM then translates such visual information into world coordinates by merging robot LiDAR sensor data and MiDaS depth estimation. The LiDAR shoots laser pulses to scan the world, measuring distances and sensing nearby objects. SLAM merges this with depth data to build a 3D world model and accurately localize the robot within it. In this project, SLAM has been optimized to be executed on the Jetson Nano in an efficient way, and all real-time mapping and navigation operations are performed with minimal hardware overhead.

3.9 Robot Operating System (ROS)

Robot Operating System (ROS) is an framework used in robot application development. ROS offers a rich collection of tools, libraries, and communication protocols such as hardware abstraction and a publish-subscribe pattern to support the integration of different components of a robotic system, i.e., perception, navigation, and control.

ROS (Robot Operating System) to function seamlessly on the NVIDIA Jetson Nano. It is the bridge that gets SLAM-based navigation, object detection, and the natural language processing (NLP) module to function in one system. This close integration enables the robot to react intelligently and effectively in real-time.

Chapter 4

Results and Discussion

article graphicx caption

The surveillance vision module YOLOv8m model, which is highly optimized and has a mean average precision (mAP) of 80%, validates its efficiency in object and anomaly detection activities across various environments. DeepSORT adds another layer of tracking by associating unique IDs with individuals and their luggage (i.e., their bags). The system tracks the distance between a person and the corresponding baggage ID. When the distance exceeds a certain threshold, the baggage is marked as abandoned, triggering a possible security alarm. This combination provides robust anomaly detection and tracking while moving.

Figure 4.1: Train and Loss Validation for the CNN

For enhancing the human-robot interaction, emotion detection is incorporated through a Convolutional Neural Network (CNN) model that is trained on grayscale image datasets. The system is 85% accurate, enabling it to identify emotional cues and respond accordingly to improve the human-robot interaction experience.

The use of MediaPipe provides real-time pose detection and accurate interpretation of human posture. It enhances contextual awareness, allowing the robot to differentiate between suspicious and non-suspicious patterns of behavior.

Figure 4.2: Robot navigating to the provided coordinates

Experimental results (Figure 4.2) depict that the autonomous surveillance robot moves smoothly towards target coordinates using a ROS-based path-planning module. Camera-based live feedback enables real-time path updates. Depth estimation, assisted by the MiDaS model, facilitates effective collision avoidance.

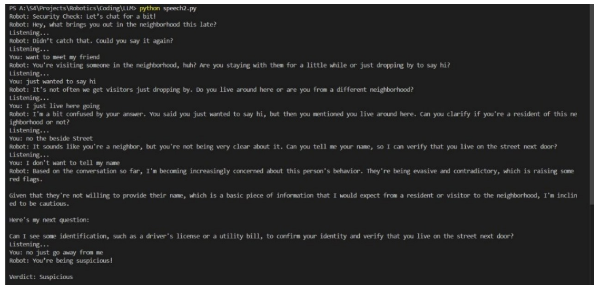

Figure 4.3: Robot conducting interrogation based on LLaMA integration

Figure 4.3 shows the robot engaging in an interrogation to evaluate suspicious activity. The robot responds to speech inputs guided by security logic and reaches a "Suspicious" verdict when sufficient identity verification is not provided.

Chapter 5

Conclusion

This project successfully demonstrated an intelligent autonomous security robot system that features navigation using LIDAR sensors, SLAM, ROS technologies. Multi-tasking performance via a hybrid control architecture, the system is much better than current security solutions to offer greater safety, efficiency, and scalability. Future research can involve further improving multi-robot coordination for larger coverage areas, further enhancing real-time adaptability in high-density obstacle environments using 6G communication, and incorporating other modalities like gesture-based human-robot interaction to further improve the operational performance.

References

[1] S. Joy, R. L. Paulraj, P. M, S. M, S. Goudar, and R. S, "A Raspberry Pi based Smart Security Patrol Robot," in Proceedings of the 7th International Conference on Computing Methodologies and Communication (ICCMC-2023), Bengaluru, India, 2023, pp. 1140-1143, IEEE Xplore Part Number: CFP23K25-ART, ISBN: 978-1-6654-6408-6.

[2] J. Zhang, G. Song, G. Qiao, T. Meng, and H. Sun, "An Indoor Security System with a Jumping Robot as the Surveillance Terminal," IEEE Transactions on Consumer Electronics, vol. 57, no. 4, pp. 1774-1781, Nov. 2011

[3] P. R. Sriram, S. K. Ramani, R. V. Shrivatsav, M. M. Manikandan, and N. Ayyappaa, "Autonomous Drone for Defence Machinery Maintenance and Surveillance," *Vision4th International Confer ence on Computer and Information Sciences (ICCOINS)* , Sriperumbudur, India, 2023 , pp. 1-5.

[4] R. Li, L. Zhao, L. Ge, L. Sun, and T. Gao, "The Development of a General Type of Security Robot," Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics (ROBIO), Sanya, China, Dec. 2007, pp. 47-52, doi:

10.1109/ROBIO.2007.4522130.

[5] J. Bravo-Arrabal, R. Vázquez-Martín, J. J. Fernández-Lozano, and A. García-Cerezo, "Strengthening Multi-Robot Systems for SAR: Co-Designing Robotics and Communication Towards 6G," IEEE Communications Magazine, 2025, doi: 10.1109.

[6] Y. Choe and M. J. Chung, "System and Software Architecture for Autonomous Surveillance Robots in Urban Environments," Proceedings of the 9th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Daejeon, Korea, Nov. 2012, pp. 536-537, doi: 10.1109/URAI.2012.6463082

[7] H. Shin, "Anomaly Detection using Elevation and Thermal Map for Security Robot," in *Proceedings of an IEEE Conference*, Daejeon, Rep. of Korea, 2023, pp. 1760-1762