Enhancing Automotive Safety through Advanced Human Action Recognition Techniques

Aravinda C.V[1,*,†], Sannidhan M S[2,†], Soumya Aswath[3,†], Jyothi Shetty[4,†] and Arjun B.C[5,†]

[1]*Nitte, deemed to be University, N.M.A.M Institute of Technology, Department of Computer Science and Engineering, Nitte, karkala, INDIA, 574110*
[2]*Nitte, deemed to be University, N.M.A.M Institute of Technology, Nitte, Department of Computer Science and Engineering, karkala, INDIA, 574110*
[3]*Nitte, deemed to be University, N.M.A.M Institute of Technology, Department of Computer Science and Engineering, Nitte, karkala, INDIA, 574110*
[4]*Nitte, deemed to be University, N.M.A.M Institute of Technology, Department of Computer Science and Engineering, Nitte, karkala, INDIA, 574110*
[5]*Rajeev Institute of Technology, Department of Information Science and Engineering, HASSAN, Karnataka, INDIA*

Abstract

This paper presents the development and implementation of an innovative rover, designed for versatile terrestrial navigation and controlled through a custom mobile application. The rover's design is inspired by the Rocker Bogie Mechanism, a robust suspension system renowned for its effectiveness in NASA's Mars exploration rovers. This mechanism ensures adaptability to varied terrains, enhancing the rover's operational flexibility. The core of the system is powered by a Raspberry Pi 4, serving as the central hub for integrating various hardware components and enabling seamless communication between the rover and its controlling mobile application, developed using MIT App Inventor.

A significant aspect of this project is the incorporation of Human Action Recognition (HAR) capabilities, achieved through the implementation of Deep Convolutional Neural Networks (CNNs). This feature introduces a novel approach to rover control and interaction, expanding its potential applications in remote exploration and monitoring tasks. Furthermore, the system boasts live camera streaming functionality, utilizing Flask for server-side operations and NGROK for secure port forwarding. This allows for real-time video feed access over the internet, thus facilitating global operational capabilities.

The integration of these technologies into a single coherent system not only demonstrates the feasibility of advanced control mechanisms in unmanned ground vehicles, but also sets a precedent for future innovations in remote exploration and surveillance. The potential applications of this technology span a wide range of fields, from environmental monitoring to search and rescue operations, underscoring its importance in the advancement of autonomous vehicle technologies.

Keywords

Human Action Recognition, Flask, Rocker Bogie Mechanism, Internet of Things, NGROK

The 6th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2024), August-19-22, 2024., Kusatsu, Japan
*Corresponding author.
†All authors contributed equally.
aravinda.cv@nitte.edu.in (A. C.V); sannidhan@nitte.edu.in (S. M. S); soumya.ashwath@nitte.edu.in (S. Aswath); $jyothi_shetty@nitte.edu.in$ (J. Shetty); $bc.arjun@gmail.com$ (A. B.C)
0000-0003-4588-4105 (A. C.V)

1. Introduction

In the rapidly evolving digital age, video surveillance has emerged as an indispensable tool for businesses worldwide. Historically conceived for security purposes, the scope and functionality of video surveillance systems have significantly expanded, offering unparalleled benefits in security surveillance, production monitoring, and deterrence of undesirable behaviors. The integration of advanced technologies has further enhanced the efficacy and application of video surveillance systems, making them a cornerstone in the operational infrastructure of both large corporations and small enterprises.

The concept of video surveillance is no longer limited to passive monitoring. It has evolved into a dynamic and interactive system that not only records events but also provides actionable insights to improve security, operational efficiency, and decision-making processes. The advent of Internet Security Systems, or IP cameras, marks a significant milestone in this evolution. Unlike their analog counterparts, IP cameras utilize the internet to transmit and receive data, thereby facilitating real-time monitoring and analysis of video feeds. This capability has transformed the landscape of video surveillance, offering businesses the ability to remotely monitor their operations, assets, and personnel with unprecedented ease and flexibility.

IP cameras represent a quantum leap in surveillance technology, offering features such as high-definition video quality, wide-angle coverage, night vision, and motion detection. Moreover, the advent of cloud storage solutions has alleviated concerns regarding data storage and retrieval, ensuring that high volumes of video data can be securely stored, accessed, and analyzed at any time. The ease of installation and the user-friendly nature of IP camera systems have democratized access to advanced surveillance technologies, enabling businesses of all sizes to fortify their security measures and operational oversight.

Despite these advancements, traditional video surveillance systems are not without their limitations. The static nature of most security cameras, for instance, poses significant challenges in surveilling large or complex premises. Fixed cameras can only monitor the areas within their field of view, leaving blind spots that can be exploited for unauthorized activities. Additionally, the installation of a comprehensive network of static cameras to cover every potential angle can be prohibitively expensive and aesthetically intrusive, especially in environments that prioritize visual appeal or where structural limitations exist.

The recognition of these limitations has spurred innovation in the realm of video surveillance, giving rise to mobile surveillance solutions that promise to address the challenges posed by static camera systems as refered in the Figure 1. These mobile solutions, ranging from unmanned aerial vehicles (UAVs) to robotic ground units equipped with video capture and transmission technology, offer the flexibility to move and adapt to changing surveillance needs. This mobility not only enhances the coverage and effectiveness of surveillance efforts but also introduces a new dimension of interactivity, where surveillance can be dynamically adjusted in response to specific incidents or threats.

2. Literature Survey

The Internet of Things (IoT) stands as a transformative force in the landscape of Human Activity Recognition (HAR), leveraging the nuanced capabilities of Channel State Information (CSI) derived from WiFi signals to discern distinct human activities. This novel application of CSI for HAR underscores the burgeoning potential of IoT devices, projected to surpass 50 billion units, in addressing complex challenges within our digital society [1]. Our research capitalizes on this potential by employing a Raspberry Pi 4 to meticulously collect and convert CSI data into images for seven daily human activities, thereby augmenting the granularity of activity recognition [2].

Video transmission in IoT environments, particularly on devices constrained by limited hardware resources like the Raspberry Pi, presents significant challenges. Our investigations reveal that video coding, rather than distributed communication frameworks, constitutes the primary bottleneck in high-definition video transfer [3]. To circumvent these limitations, we have innovatively harnessed a Raspberry Pi equipped with a night-vision camera, leveraging the Python programming language for system development. VLC media player facilitates live streaming to a host device, with VNC server and viewer enabling robust remote connections[4].

In the broader context of IoT, the synergy among diverse devices via data sharing is pivotal for advanced surveillance and monitoring applications. Our system, powered by the Raspberry Pi, Amazon Web Services, and Google Drive, exemplifies this innovative integration, offering a scalable solution for surveillance needs [5]. Furthermore, the critical role of HAR within computer vision (CV) for video surveillance applications is increasingly acknowledged, with our project contributing to this active research domain through practical implementations [6].

The design of mobile rovers, capable of navigating challenging terrains, benefits significantly from the Rocker-Bogie suspension system. This system, favored for space exploration vehicles, ensures high mobility and reliability by minimizing thermal variation impacts on motor function [7] [8]. Our project aims to refine this design for enhanced performance, demonstrating the system's robustness in handling uneven terrains by distributing payloads evenly across six wheels [9] [10].

The advent of sophisticated mobility systems in robotic vehicles has led to significant advancements in traversing challenging terrains. A prime example of such innovation is the Rocker-Bogie Mobility System, designed for slow-speed operations and remarkable obstacle navigation capabilities. Its design allows it to overcome obstacles approximately the size of its wheels. During the navigation of sizable obstacles, the system momentarily halts the vehicle's movement, enabling the front wheel to climb effectively. This mechanism ensures reliable and efficient traversal over rough terrains, highlighting the system's engineering ingenuity and its applicability in extraterrestrial exploration vehicles [11].

In the realm of smart home health monitoring, there exists a delicate balance between the intrusiveness of monitoring methods and user acceptance. Ambient, non-intrusive monitoring techniques, although limited in their data collection capabilities, are often preferred by residents due to their minimal impact on daily life. Conversely, more intrusive methods such as video surveillance and wearable devices can provide richer data sets for analysis but may face resistance due to privacy concerns. A promising solution lies in the utilization of radio frequency-based approaches, such as Channel State Information (CSI), which leverage low-cost, off-the-shelf

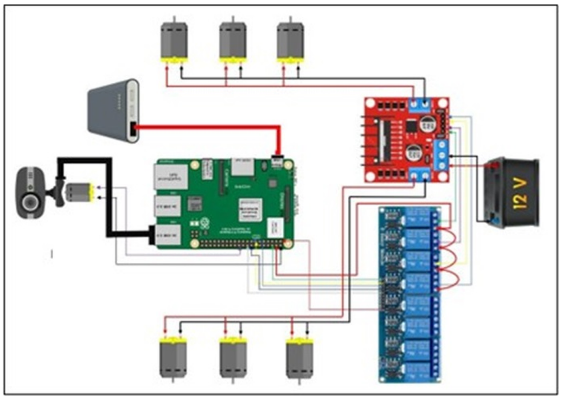

Figure 1: Architecture of the Vehicle

WiFi hardware to monitor human activities without the need for direct physical interaction or surveillance [12]

The MIT App Inventor emerges as a revolutionary online platform, democratizing the development of mobile applications by emphasizing computational thinking and user-friendly design principles. It enables users, regardless of their programming expertise, to create functional applications by visually assembling components. This approach not only simplifies the app development process but also aligns with users' mental models, facilitating a deeper understanding of computational concepts and fostering a culture of rapid, iterative design. The platform exemplifies how abstraction and user-centered design can accelerate learning and innovation in the digital age [13].

Furthermore, the field of computer vision offers transformative potential in enhancing our understanding and interaction with the digital world. By enabling computers to interpret visual information as humans do, it opens up a myriad of applications from automated surveillance to interactive interfaces. This paper showcases the practical implementation of face detection technology using OpenCV, a popular open-source library for computer vision tasks. By integrating this technology into a web application via Flask, it demonstrates the accessibility and versatility of computer vision techniques, making advanced digital interactions more achievable for developers and end-users alike [14].

This literature underscores the interdisciplinary nature of our research, spanning IoT, HAR, and mobile robotic systems, and highlights our contributions to the field through innovative system design and practical implementations.

3. Problem Statement

Let $A = \{a_1, a_2, \ldots, a_n\}$ represent the set of human activities to be recognized by the HAR-V. The goal is to develop a function f that maps an input dataset D of observed behaviors and environmental factors to the set of activities A. The images of the Rasberry-pi and Driver is shown in the Figure 2 and Figure 3, Mathematically, this can be expressed as:

$$f : D \to A$$

Variables and Data Representation:

- D: Dataset of observed behaviors and environmental factors, where $D = \{d_1, d_2, \ldots, d_m\}$ and each d_i is a vector of observed features at time i.
- A: The set of human activities to be recognized.
- V: A set representing the vehicle's state and controls, where $V = \{v_1, v_2, \ldots, v_k\}$ encapsulates parameters such as location, speed, and camera orientation.

Activity Recognition Function:
The function f utilizes deep learning techniques, particularly convolutional neural networks (CNNs), to recognize activities. Given an input $x_i \in D$, the function f outputs a prediction $a_j \in A$. This can be mathematically represented as:

$$f(x_i) = a_j$$

Vehicle Control Function:
Let g be the function that maps the recognized activity a_j to the vehicle's control actions v_k. This mapping ensures the vehicle's mobility and functionality in response to the recognized activities. Mathematically, this can be expressed as:

$$g(a_j) = v_k$$

Internet Connectivity and Control:
The vehicle's ability to be controlled from anywhere through the Internet can be modeled by defining a set C of control signals received over the Internet. Let h be the function that maps these control signals to the vehicle's state and controls:

$$h : C \to V$$

Overall System Function:
The overall functionality of the HAR-V can be encapsulated by combining the functions f, g, and h. This composite function takes inputs from the dataset D and control signals C, processes these through the activity recognition and control mapping functions, and outputs the vehicle's control actions V:

$$F(D, C) = g(f(D)) + h(C) = V$$

Optimization and Learning:
The parameters of f and g are optimized through a learning process, often involving back-propagation and gradient descent, to minimize the difference between the predicted activities and the true activities, as well as to optimize the vehicle's responses to these activities.

4. Methodology

The proliferation of accessible computing platforms such as Raspberry Pi has enabled innovative applications in AI and machine learning. This guide focuses on setting up a HAR system that uses video surveillance to identify human activities through advanced neural network technologies.

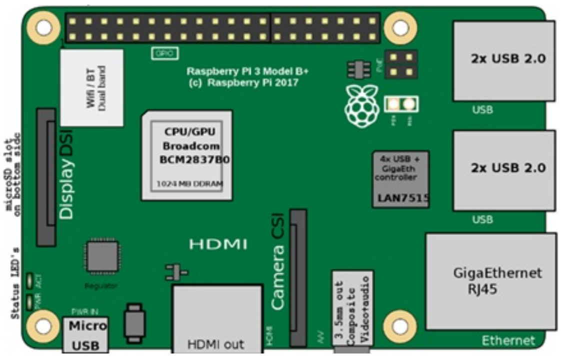

Figure 2: Raspberry-pi-v4

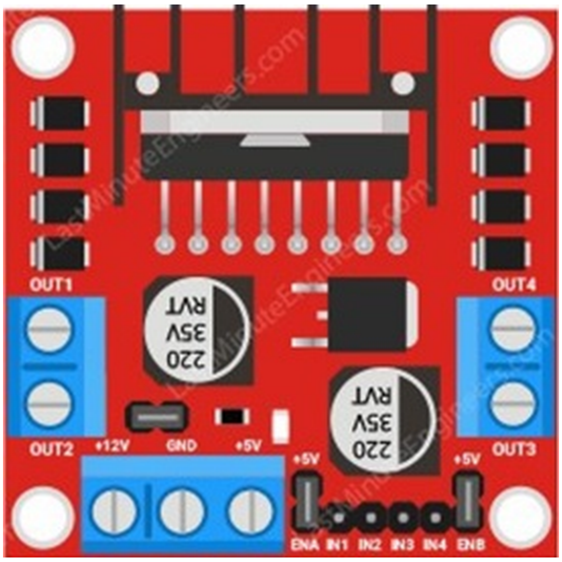

Figure 3: Driver-Module

4.1. System Setup

4.1.1. Required Downloads

- **Win32 Disk Imager**: Essential for writing Raspbian images to the SD card. https://sourceforge.net/projects/win32diskimager/
- **SD Card Formatter**: Formats the SD card optimally. https://www.sdcard.org/downloads/formatter_4/
- **Raspbian OS**: The operating system for Raspberry Pi. https://www.raspberrypi.org/downloads/raspbian/

Algorithm 1 Activity Recognition Function

Input: Dataset D of observed behaviors
Output: Recognized activity a
procedure RecognizeActivity(D)
    for each d_i in D **do**
        Extract features from d_i
        $a \leftarrow$ classify features using CNN
        return a
    end for
end procedure

Algorithm 2 Vehicle Control Function based on Recognized Activity

Input: Activity a, Control Parameters V
procedure ControlVehicle(a, V)
    if $a =$ Activity1 **then**
        Perform control action 1
    else if $a =$ Activity2 **then**
        Perform control action 2
    else
        Perform default action
    end if
end procedure

Algorithm 3 Remote Control via Internet

Input: Control signals C received over the Internet
procedure RemoteControl(C)
    for each signal c_i in C **do**
        Decode c_i to corresponding control action v
        Execute control action v on vehicle
    end for
end procedure

4.1.2. SD Card Preparation

Secure a minimum of a 32GB SD card, and after formatting it using SD Card Formatter, write the Raspbian OS image using Win32 Disk Imager.

4.1.3. Software and Device Configuration

Upon preparing the SD card and booting the Raspberry Pi with the camera module set up, deploy a Python Flask application for live video streaming, accessible over NGROK.

4.2. Implementation of Human Action Recognition

4.2.1. Data Preparation

Utilize the UCF50 - Action Recognition Dataset for the model, processing the videos by resizing frames and normalizing pixel values.

4.2.2. Model Training and Evaluation

Employ a ConvLSTM-based model using Keras for the HAR system. After training, apply the model to new video data from the live streaming service, evaluating the model's performance through confidence scores.

5. Experimental Setup and Implementation

The following improvements and configurations were systematically applied to the rover and its associated mobile application, enhancing its operational efficiency and user interface:

- Wheels were upgraded from 65mm to 160mm in diameter to improve mobility.
- The rover's structure was reinforced by adding axles to all pairs of wheels.
- A new camera frame was constructed to elevate the camera, enhancing the field of vision.
- A cover design was developed to conceal all internal components, improving aesthetics and protecting the electronics.
- Wiring was redesigned to eliminate protrusions, ensuring a cleaner and safer setup.
- **Putty** and **WinSCP** were utilized for remote control of the Raspberry Pi and for downloading captured data, respectively.

Getting Started with Ngrok

The integration of ngrok provides a secure method to access local services from any location, following these steps:

1. **Local Web Service:** Pre-established by hosting our Python script using Flask on port 5000.
2. **Install the ngrok Agent:** For Linux, the agent can be installed using the following Apt commands:

```
curl -s https://ngrok-agent.s3.amazonaws.com/ngrok.asc | \
sudo tee /etc/apt/trusted.gpg.d/ngrok.asc > /dev/null && \
echo "deb https://ngrok-agent.s3.amazonaws.com buster main" | \
sudo tee /etc/apt/sources.list.d/ngrok.list && \
sudo apt update && sudo apt install ngrok
```

3. **Connect Your Agent to Your Ngrok Account:** Obtain your Authtoken from the ngrok dashboard and link it using:

```
ngrok config add-authtoken TOKEN
```

4. **Start Ngrok:** Initiate ngrok with the command:

   ```
   ngrok http 5000
   ```

   This step securely exposes the Flask application running on port 5000 to the internet.

6. Result Discussion

6.1. Current Achievements

1. Speed of the Vehicle: The vehicle, traveling at a speed of 3-5 KMPH, can recognize a person at a distance of 10-12 meters.
2. Mobile Application and Rover Synchronization: The project has successfully achieved a seamless integration between the rover and its controlling mobile application. This synchronization facilitates direct command and control over the rover, with a noted minor delay in live streaming that is slated for future improvement.
3. Camera Functionality: The implemented camera system, capable of precise rotation as required, has met the project's initial objectives. This functionality enhances the rover's ability to survey its surroundings effectively.
4. Responsive Flask Web Pages: The Flask-based web interfaces developed for this project have demonstrated high responsiveness, ensuring user-friendly interaction and control over the rover's operations.
5. Human Activity Recognition (HAR): Preliminary tests of the HAR system have been conducted on a limited dataset, successfully classifying basic human movements such as running, walking, and jumping. This initial success lays the groundwork for more extensive application and refinement as shwon in Figure 4, Figure 5, Figure 6 respectively.
6. Object Detection: The project has incorporated advanced object detection algorithms, specifically Deep Convolution Network . These models have proven effective in identifying a wide range of objects, including people, furniture, and personal items, showcasing the system's versatility.
7. Accuracy: The recognition rate achieves an accuracy of approximately 85% to 90%.

6.2. Future work to be carried out

1. Live Streaming Optimization: Efforts will be concentrated on reducing the delay in live streaming, aiming for real-time performance to enhance the system's responsiveness and operational efficiency.
2. GPS Tracking Module: The integration of a GPS tracking module is anticipated, which will enable precise location tracking of the rover. This addition will significantly enhance the system's utility for outdoor navigation and surveillance.
3. Suspension System Enhancement: To improve the rover's mobility and adaptability to varied terrains, the introduction of a more advanced suspension system is planned. This upgrade will ensure greater flexibility and durability in the rover's operational capabilities.

Figure 4: Running Image is recognized

Figure 5: Bending of the human recognized

Figure 6: Action of the human hands recongized

7. Conclusion

The work has laid a solid foundation in the realms of remote-operated surveillance and interaction technologies, marked by the successful deployment of HAR and object detection functionalities. With identified pathways for future enhancements, including live streaming optimization, GPS tracking, and suspension system improvements, the project is set to evolve into a more robust and versatile system, promising significant contributions to the field of robotics and remote surveillance.

References

[1] Real-time medicine packet recognition system in dispensing medicines for the elderly. Available online: https://doi.org/10.1016/j.measen.2021.100072 (accessed on 18 March 2024).

[2] Low Delay Video Streaming on the Internet of Things Using Raspberry Pi. Available online: https://www.mdpi.com/2079-9292/5/3/60 (accessed on 18 March 2024).

[3] Live Video Streaming from Remote Location Using Raspberry Pi. Available online: https://indjst.org/articles/live-video-streaming-from-remote-location-using-raspberry-pi (accessed on 18 March 2024).

[4] IoT Enabled Video Surveillance System Using Raspberry Pi. Available online: https://www.researchgate.net/publication/327323813_IoT_Enabled_Video_Surveillance_System_Usin (accessed on 18 March 2024).

[5] Human Action Recognition: A Paradigm of Best Deep Learning Features Selection and Serial Based Extended Fusion. Available online: https://www.mdpi.com/1424-8220/21/23/7941/htm# (accessed on 18 March 2024).

[6] Design of Rocker Bogie Mechanism. Available online: https://www.researchgate.net/publication/313403109_Design_of_Rocker_Bogie_Mechanism (accessed on 18 March 2024).

[7] Design analysis of Rocker Bogie Suspension System and Access the possibility to implement in Front Loading Vehicles. Available online: https://www.researchgate.net/publication/330854620_Design_analysis_of_Rocker_Bogie_Suspension (accessed on 18 March 2024).

[8] Design of All Terrain Vehicle Using Rocker Bogie Mechanism. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3451713 (accessed on 18 March 2024).

[9] Remote Controlled Rover Using Rocker Bogie Mechanism. Available online: https://www.ijraset.com/research-paper/remote-controlled-rover-using-rocker-bogie-mechanism (accessed on 18 March 2024).

[10] A study on Rocker-Bogie Suspension for a Planetary Rover Prototype. Available online: https://www.ijraset.com/research-paper/remote-controlled-rover-using-rocker-bogie-mechanism (accessed on [access date]).

[11] Rocker Bogie. Available online: https://www.ijraset.com/research-paper/rocker-bogie (accessed on [access date]).

[12] WiFi-based Human Activity Recognition using Raspberry Pi. Available online: https://ieeexplore.ieee.org/document/9288199 (accessed on [access date]).

[13] MIT App Inventor: Objectives, Design, and Development. Available online: https://www.researchgate.net/publication/337361811_MIT_App_Inventor_

Objectives_Design_and_Development (accessed on [access date]).
[14] How to Display Video Streaming From A Webcam Using Flask. Available online: https://towardsdatascience.com/how-to-display-video-streaming-from-a-webcam-using-flask-7a15e26fbab8 (accessed on [access date]).