IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Mobile Robotic Arm System With Intelligent Gesture And Voice Control

[1]Eshwari H N, [2]G Udhbav, [3]Lakshmi C Koujalagi, [4]Pruthvi Narayana Reddy, [5]Dr.M.A.Dorairangawamy

[1-4]Final year Students, [5]Senior Professor,

[1-5] Department of AIML, [1-5]Bangalore Technological Institute.

Abstract

This paper presents the design and development of a CNN-based smart robotic arm that mimics human hand gestures for intuitive control. [1] Unlike traditional robotic arms that rely on joysticks or pre-programmed instructions, the proposed system enables natural human–robot interaction through gesture recognition, voice commands, and vision-based object classification, with the robotic arm mounted on a wireless RC car to allow mobility in dynamic environments. The system integrates multiple modules, including gesture recognition using deep learning (CNN/LSTM) models trained on the ASL Alphabet Dataset for static gestures and the ChaLearn LAP IsoGD dataset for dynamic gestures [2]; robotic arm control through ESP32 and servo motors for grip, lift, and rotation; RC car navigation controlled via voice commands captured by a ReSpeaker 4-Mic Array [4]; and a computer vision module using a Raspberry Pi camera for colour and shape detection to assist in pick-and-place tasks [3]. This multimodal approach ensures intuitive, flexible, and robust control, addressing real-world challenges such as gesture variability, multimodal fusion, and real-time deployment. Applications of the proposed work include assistive robotics, warehouse logistics, industrial automation, and smart surveillance.

Index Terms - CNN, Gesture Recognition, Robotic Arm, ESP32, Raspberry Pi, Voice Commands, Object Detection, Shape Recognition, Colour Classification, Human–Robot Interaction, Multimodal Control.

1. Introduction

1.1 Overview

In recent years, human–robot interaction has gained significant attention as robotics moves beyond industrial automation into healthcare, education, surveillance, and assistive technologies. Traditional control interfaces such as joysticks, remotes, and scripted instructions limit adaptability and require technical expertise. **Gesture-based control** provides a natural and contactless way to interact with robots, while **voice commands** [4] and **computer vision** add further flexibility.

The proposed project, **A Mobile Robotic Arm System with Intelligent Gesture and Voice Control**, integrates **gesture recognition, robotic arm control, RC car mobility, and vision-based object detection** into a unified system. This ensures mobility, precision, and adaptability, making the robot capable of operating in semi-structured environments.

1.2 Problem Statement

Gesture-controlled robotic systems, despite advances in robotics and AI, face key challenges that limit their adoption. Recognition accuracy is affected by **gesture variability** across individuals, while reliance on **single-modality control** restricts adaptability. Running deep learning models on embedded devices such as Raspberry Pi also introduces **real-time performance constraints**. In addition, the absence of **safety mechanisms** like emergency stops or confidence thresholds can lead to unintended actions, and the lack of **object classification capabilities** (colour and shape) limits intelligent automation. This project overcomes these limitations by developing a **multimodal robotic platform** that combines static and dynamic gesture recognition, voice-controlled mobility, and vision-based classification, deployed on Raspberry Pi [3] and ESP32 for efficient and safe real-time interaction.

1.3 Objectives

The key objectives are:

- **Gesture Recognition**: Build CNN/LSTM models for static and dynamic gestures using ASL Alphabet and ChaLearn LAP IsoGD datasets [1].
- **Robotic Arm Control**: Enable the arm to mimic gestures such as grip, lift, rotate, and release using servo motors [3].
- **RC Car Mobility**: Use ESP32 for wireless control of RC car movements (forward, backward, left, right, stop) via voice commands [5].
- **Vision-Based Classification**: Implement colour and shape recognition to assist in intelligent pick-and-place.
- **Hybrid Multimodal System**: Integrate gesture, voice, and vision to create a robust real-time robotic system.

1.4 Motivation

The motivation is to develop a **low-cost, efficient, and multimodal robotic system** that can serve multiple real-world applications. Traditional arms are either too rigid (pre-programmed) or too expensive for general use. By leveraging **deep learning, Raspberry Pi, and ESP32**, this project demonstrates how AI and affordable hardware can be combined to create **intelligent assistive robotics**.

2. Aim

The aim of this project is to design and implement a CNN-based robotic arm system that can mimic human hand gestures, operate as a mobile robot via voice commands, and perform object classification tasks using vision. The ultimate goal is to build a multimodal robotic system that demonstrates intelligent, flexible, and real-time human–robot interaction.

3. System Modules

The proposed robotic system is organized into ten functional modules that collectively enable multimodal interaction and control.

The **Data Acquisition and Preprocessing Module** captures gestures, voice commands, and object images using a Raspberry Pi camera and ReSpeaker microphone array, followed by normalization, landmark extraction, and noise reduction. The **Gesture Recognition Module** [1] employs CNN models for static gestures and LSTM/GRU networks for dynamic gestures, trained on the ASL Alphabet and ChaLearn IsoGD datasets. The **Voice Command Module** processes a limited vocabulary of navigation commands using a lightweight speech-to-text engine for reliable RC car control.

Recognized inputs are mapped through the **Command Translation Module**, which converts predictions into high-level robotic instructions. These are executed by the **Robotic Arm Control Module**, which actuates servos for grip, lift, and rotation, and the **RC Car Mobility Module**, which drives DC motors for navigation. The **Vision Module** enhances perception by classifying object colours and shapes using OpenCV-based HSV segmentation and contour approximation.

Wireless connectivity is managed by the **Communication Module**, with ESP32 providing low-latency links between Raspberry Pi and hardware components. Additional features include a **Feedback and Monitoring Module** that displays live status through a GUI, and a **Power Management Module** that stabilizes supply to Raspberry Pi, ESP32, motors, and sensors. Together, these modules establish a **scalable, safe, and multimodal robotic system** capable of gesture-controlled manipulation, voice-guided mobility, and intelligent object recognition.

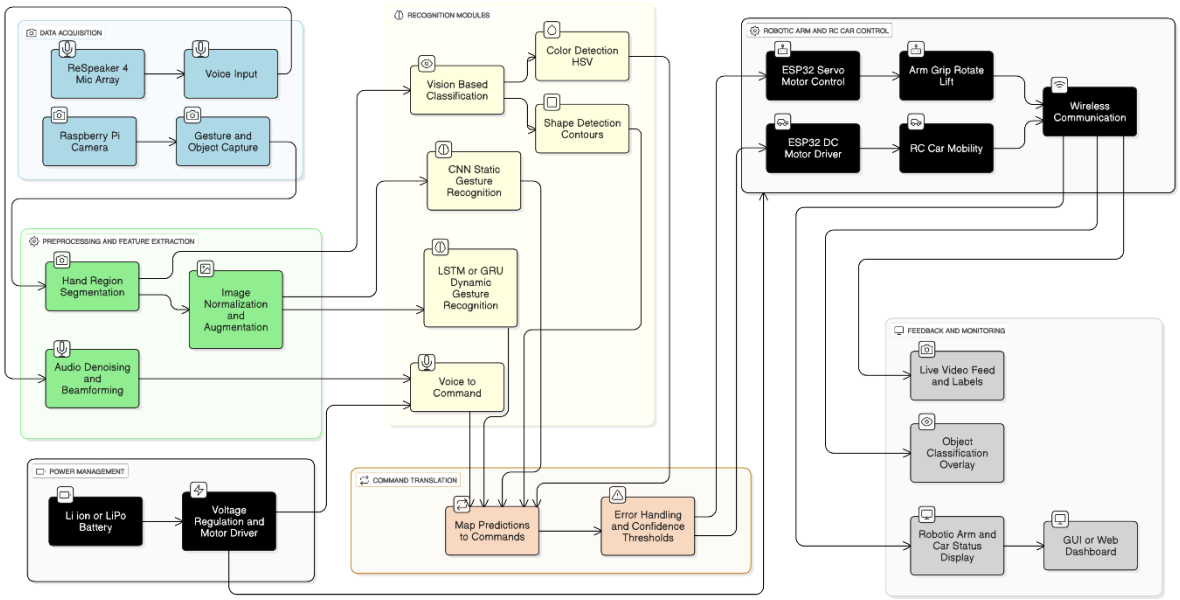

Fig 3.1. System Architecture

4. Methodology

The methodology of the proposed system follows a structured workflow that integrates gesture recognition, voice command processing, vision-based classification, and robotic control into a unified multimodal framework.

Dataset Selection and Preprocessing: The system leverages two benchmark datasets: the **ASL Alphabet Dataset** for static hand gestures and the **ChaLearn LAP IsoGD dataset** for dynamic gestures. Images are resized, normalized, and hand landmarks are extracted using Mediapipe. Dynamic gestures are converted into frame sequences for temporal modelling. Voice data are denoised using the ReSpeaker's beamforming features, while visual input is processed in HSV colour space for colour recognition and contour analysis for shape detection.

Model Development: For static gestures, a **CNN-based classifier** is trained on the ASL dataset. For dynamic gestures, an **LSTM/GRU network** is trained on sequential frames from the IsoGD dataset. Models are optimized for real-time inference on Raspberry Pi using TensorFlow Lite. Voice commands are processed with a lightweight speech-to-text engine restricted to a limited vocabulary, improving reliability in embedded environments. Object recognition is achieved using OpenCV pipelines for detecting both **colours** and **shapes**.

Hardware Integration: The **Raspberry Pi** [3] serves as the main processing unit, executing gesture, voice, and vision recognition tasks. Processed commands are transmitted via **ESP32**, which controls servo motors for the robotic arm and DC motors for the RC car [5]. The robotic arm mimics human hand gestures for manipulation, while the RC car provides mobility based on voice commands [4].

System Integration and Safety: A **command decision module** fuses multimodal inputs, applies confidence thresholds, and ensures safe execution. Emergency stop functionality is implemented using both voice and gesture triggers. The modular architecture allows scalability, with optional GUI-based monitoring for real-time visualization and logging.

5. Results and Discussion

The proposed multimodal system demonstrated high performance across all modules. **Gesture recognition** achieved ~95% accuracy for static gestures (ASL dataset) and ~88–90% for dynamic gestures (ChaLearn IsoGD), with inference latencies of 40–150 ms on Raspberry Pi. **Voice commands** attained >96% accuracy using a limited vocabulary with the ReSpeaker 4-Mic Array, ensuring reliable RC car navigation, while **vision-based classification** reached ~92% accuracy in colour detection and ~90% in shape recognition.

End-to-end latency ranged between 200–300 ms, suitable for real-time operation, and safety mechanisms such as confidence thresholds and emergency stops (<100 ms) enhanced reliability. Compared to unimodal systems, the proposed work offers ~90–95% overall accuracy with improved flexibility, mobility, and adaptability through integrated gesture, voice, and vision control.

6. Applications

- **Assistive Robotics**: Helping differently-abled individuals with daily tasks.
- **Smart Warehousing**: Gesture-controlled pick-and-place robots for logistics.
- **Industrial Automation**: Contactless control in hazardous environments.
- **Surveillance**: Mobile robots with gesture/voice control for security tasks.
- **Education & Research**: Low-cost testbed for human–robot interaction studies.

7. Conclusion and Future Scope

The proposed system successfully integrates **gesture recognition, voice commands, and vision-based classification** into a mobile robotic arm platform using **Raspberry Pi and ESP32**, achieving ~90–95% accuracy with low-latency performance. Safety mechanisms such as emergency stops and confidence thresholds enhance reliability, while multimodal control makes the system suitable for applications in **assistive robotics, logistics, and automation**. Future improvements include adopting **advanced deep learning models**, expanding gesture and voice vocabularies, and incorporating **sensor fusion** for richer perception, enabling the system to evolve into a more **scalable and intelligent robotic assistant**.

Reference

[1] S. C. Sethuraman, G. Reddy Tadkapally, A. Kiran, S. P. Mohanty and A. Subramanian, "SimplyMime: A Dynamic Gesture Recognition and Authentication System for Smart Remote Control," in IEEE Sensors Journal, vol. 24, no. 24, pp. 42472-42483, 15 Dec.15, 2024, doi: 10.1109/JSEN.2024.3487070.

[2] A. Dahiya, D. Wadhwa, R. Katti and L. G. Occhipinti, "Efficient Hand Gesture Recognition Using Artificial Intelligence and IMU-Based Wearable Device," in IEEE Sensors Letters, vol. 8, no. 12, pp. 1-4, Dec. 2024, Art no. 6015604, doi: 10.1109/LSENS.2024.3501586.

[3] G. Li et al., "A System for Automatic Recognition of Robotic Hand Gesture Based on Raspberry Pi and Convolutional Neural Network by Using Specklegram," in IEEE Sensors Journal, vol. 25, no. 7, pp. 11839-11846, 1 April1, 2025, doi: 10.1109/JSEN.2025.3537703.

[4] N. J. Kumar, A. M. Ali, T. B. B. R, N. Partheeban and V. Nagaraju, "A Novel Voice Assisted Internet of Things based Residential Automation Scheme with Learning Support," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-6, doi: 10.1109/ACCAI53970.2022.9752496.

[5] R. Khande and S. Rajapurkar, "Smart Voice and Gesture Controlled Wheel Chair," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 413-417, doi: 10.1109/ICOEI53556.2022.9777223.