International Journal of Engineering Research and Science & Technology

IOT ENABLED AI HUMANOID ROBOT USING RASPBERRY PI 5 FOR HUMAN FOLLOWING AND VOICE COMMUNICATION

[1]Mr. S. BALAKRISHNA,Assistant Professor

[2]GUNI SAI TEJA

[3]H. HARI PRIYA

[4]K. LOKESH

[5]K. JASHWANTH

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

TKR COLLEGE OF ENGINEERING & TECHNOLOGY

(AUTONOMOUS)

(Accredited by NBA and NAAC with 'A+' Grade)

Medbowli, Meerpet, Saroornagar, Hyderabad-500097

ABSTRACT

This project presents the design and development of an **IoT-enabled AI humanoid robot** powered by Raspberry Pi 5, integrating advanced technologies such as computer vision, natural language processing, and real-time IoT communication. The proposed system is capable of performing autonomous human-following and interactive voice communication, enabling effective human-robot interaction in real-world environments. The robot utilizes a deep learning-based vision system (YOLOv8) for accurate human detection and tracking, combined with PID-based motor control for smooth navigation and obstacle avoidance. Additionally, a voice interaction module incorporating speech recognition (Whisper) and text-to-speech synthesis (Piper) enables natural conversational capabilities.

The system architecture is structured into multiple layers, including sensing, actuation, communication, and power management, all coordinated through a modular software framework running on Raspberry Pi OS. IoT integration is

achieved using the MQTT protocol, allowing real-time monitoring, remote control, and cloud-based data visualization through a dashboard interface. Experimental results demonstrate high accuracy in human detection, efficient real-time response, and reliable system performance with low latency.

The developed humanoid robot offers a cost-effective and scalable solution for applications in healthcare, education, service industries, and smart environments. By combining embedded AI with IoT connectivity, the system highlights the potential of intelligent, interactive robotic platforms for future human-centric automation.

Keywords

IoT, Humanoid Robot, Raspberry Pi 5, Computer Vision, YOLOv8, Human Detection and Tracking, Natural Language Processing, Speech Recognition, Text-to-Speech, MQTT Protocol, Embedded Systems, Human-Robot Interaction (HRI), Autonomous Navigation, PID Control, Smart Robotics

I. INTRODUCTION

The rapid advancement of artificial intelligence (AI), embedded systems, and Internet of Things (IoT) technologies has significantly accelerated the development of intelligent robotic systems. Among these, humanoid robots have gained considerable attention due to their ability to interact naturally with humans and operate effectively in human-centric environments. These robots are increasingly being deployed in domains such as healthcare, education, service industries, and smart homes, where intuitive human-robot interaction (HRI) is essential for efficient task execution.

Traditional robotic systems, however, have often been limited by high cost, bulky hardware, and lack of intelligent interaction capabilities. With the emergence of compact and powerful embedded platforms like the Raspberry Pi 5, it has become feasible to design cost-effective humanoid robots capable of performing complex AI-driven tasks. The integration of modern deep learning techniques, particularly in computer vision and natural language processing, further enhances the robot's ability to perceive and interact with its environment in real time.

A key requirement for effective human-robot interaction is the ability of the robot to both physically engage with humans and communicate naturally. This includes capabilities such as human detection and

tracking, autonomous navigation, and voice-based interaction. Recent advancements in object detection models like YOLOv8 have enabled accurate and efficient real-time human tracking even on embedded systems. Simultaneously, improvements in speech technologies, including automatic speech recognition (ASR) and text-to-speech (TTS), have enabled more natural and context-aware communication between humans and machines.

In addition to intelligence and interaction, connectivity plays a crucial role in modern robotic systems. IoT integration allows robots to communicate with cloud platforms, enabling remote monitoring, control, and data analysis. Lightweight communication protocols such as MQTT facilitate efficient real-time data exchange, making them suitable for resource-constrained devices. This connectivity enhances scalability and enables the deployment of robotic systems in distributed and smart environments.

This project focuses on the design and implementation of an IoT-enabled AI humanoid robot using Raspberry Pi 5, capable of autonomous human following and real-time voice communication. The system integrates computer vision, speech processing, IoT communication, and motor control into a unified architecture. By combining affordability, intelligence, and connectivity, the proposed system aims to provide a practical and scalable solution for next-generation interactive robotic applications.

II. LITERATURE REVIEW

Humanoid robotics has evolved significantly over the past decades, transitioning from simple mechanical systems to intelligent machines capable of complex perception and interaction. Early developments such as WABOT-1 demonstrated basic human-like movement, while later systems like Honda's ASIMO introduced stable locomotion and task execution capabilities. More recent robots, including advanced platforms developed by research institutions and industries, have incorporated artificial intelligence to enhance adaptability and autonomy. Despite these advancements, many high-end humanoid robots remain expensive and rely on specialized hardware, limiting their accessibility. In contrast, modern approaches emphasize cost-effective solutions using embedded platforms like Raspberry Pi, enabling wider adoption in research and real-world applications.

Human detection and tracking have been central to improving human-robot

interaction. Early systems relied on sensor-based approaches such as laser scanners for detecting human movement, which were effective but limited to controlled environments. Subsequently, vision-based techniques using color segmentation were introduced, offering improved flexibility but suffering from sensitivity to lighting variations. With the advent of deep learning, object detection models such as YOLO (You Only Look Once) have revolutionized this domain by enabling accurate, real-time detection of humans in dynamic environments. Recent versions like YOLOv8 provide enhanced performance and efficiency, making them suitable for deployment on embedded systems. Additionally, the integration of depth-sensing technologies and tracking algorithms has further improved robustness in complex scenarios.

Voice communication has also undergone substantial advancements, evolving from basic command-based systems to sophisticated conversational interfaces. Traditional speech recognition systems, such as CMU Sphinx, provided offline capabilities but were limited in accuracy. The introduction of cloud-based speech APIs significantly improved recognition performance and language support. More recently, lightweight deep learning models like Whisper have enabled high-accuracy speech recognition on edge devices, while neural text-to-speech systems such as Piper provide natural and real-time voice output. Furthermore, the integration of large language models (LLMs) has enhanced the ability of robots to generate context-aware and human-like responses, thereby improving user engagement and interaction quality.

The integration of IoT technologies has further expanded the capabilities of robotic systems by enabling remote connectivity, monitoring, and control. Communication protocols such as MQTT have gained popularity due to their lightweight design and efficiency in handling real-time data exchange. Compared to traditional HTTP-based systems, MQTT offers lower latency and better performance for resource-constrained devices. The combination of IoT platforms with robotics allows for scalable deployments, cloud-based analytics, and seamless system updates, making it a critical component of modern intelligent systems.

The Raspberry Pi platform has emerged as a widely used solution in robotics due to its affordability, compact size, and strong community support. Earlier versions demonstrated the feasibility of performing real-time computer vision tasks on low-cost hardware, while the latest Raspberry

Pi 5 offers enhanced processing power and improved I/O capabilities. These advancements make it suitable for running AI models and managing multiple subsystems simultaneously. As a result, it serves as an effective foundation for developing intelligent, IoT-enabled humanoid robots.

III. METHODOLOGY

The proposed IoT-enabled AI humanoid robot is developed using a modular and layered approach that integrates hardware and software components to achieve autonomous human following and real-time voice interaction. The methodology is structured into system design, hardware implementation, software architecture, and system integration phases to ensure reliability, scalability, and efficient performance.

The system architecture is organized into five functional layers: computing, sensing, actuation, communication, and power management. The Raspberry Pi 5 serves as the central processing unit, handling all computational tasks including vision processing, speech recognition, decision-making, and IoT communication. The sensing layer consists of a high-resolution camera for visual perception, ultrasonic sensors for obstacle detection, an IMU for orientation sensing, and a microphone array for voice input. The actuation layer includes DC motors and servo motors controlled via motor drivers and PWM controllers, enabling movement and physical interaction. Communication is established through Wi-Fi using the MQTT protocol, while the power management system ensures stable operation through regulated power distribution.

The human-following mechanism is implemented using a combination of computer vision and control algorithms. A deep learning-based object detection model (YOLOv8) is used to identify humans in real time from the camera feed. The detected person is tracked across frames using a tracking algorithm, ensuring consistent identification even under partial occlusions. A priority-based selection mechanism determines the target individual based on proximity, position, and size. The robot's movement is controlled using PID-based motor control, where lateral and longitudinal errors are computed to adjust direction and speed. Differential drive kinematics is applied to convert control signals into wheel movements, enabling smooth and accurate navigation.

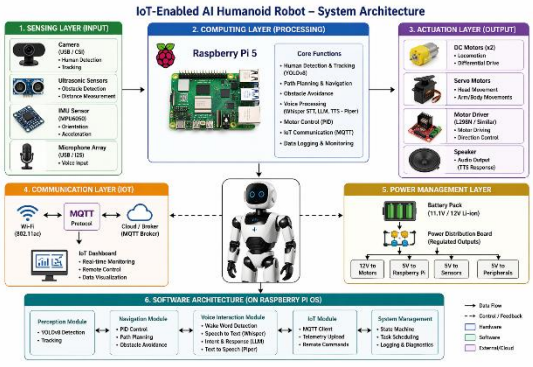
Obstacle avoidance is implemented as a safety layer using ultrasonic sensors. Distance measurements are continuously monitored, and predefined thresholds are used to trigger actions such as stopping, turning, or reversing. An emergency stop mechanism is incorporated to immediately halt the robot when obstacles are detected within critical distance, ensuring safe operation in dynamic environments.

The voice communication system follows a multi-stage pipeline consisting of wake word detection, speech recognition, intent classification, response generation, and speech synthesis. The system detects a predefined wake word to activate listening mode, after which speech input is processed using a speech-to-text model. The extracted text is analyzed using a rule-based and AI-driven approach to determine user intent. For predefined commands, direct responses are generated, while complex queries are handled using a language model. The final response is converted into speech using a text-to-speech engine and played through the speaker, enabling natural interaction.

IoT integration is achieved using a publish–subscribe communication model based on the MQTT protocol. The robot continuously transmits telemetry data such as position, battery status, sensor readings, and system state to a cloud-based dashboard. It also receives remote commands for control and monitoring. Secure communication is ensured through encryption protocols, and over-the-air updates enable remote system maintenance.

The software system is designed using a modular architecture where individual components such as vision processing, motor control, voice interaction, and IoT communication operate as independent services. These modules communicate through a lightweight message-passing mechanism, ensuring flexibility and fault tolerance. A finite state machine governs the overall behavior of the robot, managing states such as idle, searching, following, talking, and emergency stop.

IV. SYSTEM ARCHITECTURE



V. RESULTS & DISCUSSION

The developed IoT-enabled AI humanoid robot was evaluated through a series of experiments to analyze its performance in

terms of vision accuracy, human-following capability, voice interaction, IoT communication, and overall system reliability. The results demonstrate that the system achieves efficient real-time operation with a good balance between accuracy and computational cost on the Raspberry Pi 5 platform.

The vision system, based on the YOLOv8 model, showed high detection accuracy under various environmental conditions. In well-lit environments, the system achieved detection rates above 99% with high confidence scores, while maintaining satisfactory performance even under moderate lighting and partial occlusion scenarios. The model operated at an average speed of over 20 frames per second, enabling real-time human detection and tracking. However, performance degradation was observed in low-light and highly occluded conditions, indicating a limitation of vision-based systems in challenging environments.

The human-following mechanism demonstrated stable and reliable performance. The robot was able to acquire a target within a short time and maintain a consistent following distance with minimal error. The integration of PID control ensured smooth navigation and reduced oscillations during movement. The system also showed effective re-acquisition capability when the target was temporarily lost. Obstacle avoidance mechanisms worked efficiently, allowing the robot to safely navigate around obstacles and perform emergency stops when required.

The voice communication system provided natural and responsive interaction. Speech recognition accuracy was high in quiet environments, with acceptable performance in moderate noise conditions. The end-to-end response time for voice interaction was approximately two seconds, which is suitable for real-time applications. The use of text-to-speech synthesis produced clear and understandable responses, enhancing user experience. However, increased noise levels affected recognition accuracy, highlighting the need for improved noise handling techniques in future enhancements.

IoT connectivity tests indicated reliable and low-latency communication between the robot and the cloud dashboard. MQTT-based communication achieved minimal delay in local networks and acceptable latency in cloud-based scenarios. The system successfully transmitted real-time telemetry data and responded to remote commands efficiently. Additionally, the

system demonstrated stable performance during prolonged operation, with minimal connection drops and effective recovery mechanisms.

Power and thermal analysis showed that the robot operates within safe limits under typical workloads. The system achieved a battery life of approximately 4–5 hours, making it suitable for practical applications. Temperature levels remained within acceptable thresholds, ensuring stable operation without overheating. These results confirm the feasibility of deploying the system in real-world environments with sustained performance.

VI. CONCLUSION

This project successfully presents the design and implementation of an IoT-enabled AI humanoid robot using Raspberry Pi 5, integrating computer vision, voice interaction, and real-time IoT communication into a unified system. The developed robot demonstrates the ability to autonomously detect and follow humans while maintaining safe navigation through obstacle avoidance mechanisms. Additionally, the incorporation of speech recognition and text-to-speech technologies enables natural and interactive communication, enhancing the overall human-robot interaction experience.

The modular hardware and software architecture ensures flexibility, scalability, and efficient system performance. The use of deep learning models for perception and lightweight communication protocols for IoT connectivity allows the system to operate effectively on a compact and cost-efficient embedded platform. Experimental results confirm that the robot achieves high accuracy in human detection, stable following behavior, reliable voice response, and low-latency communication, making it suitable for real-world applications.

Despite its effectiveness, the system has certain limitations, such as reduced performance in low-light environments and sensitivity to background noise during voice interaction. These challenges highlight opportunities for future improvements, including the integration of advanced sensors, improved noise filtering techniques, and more robust AI models. Additionally, incorporating advanced navigation techniques and expanding conversational intelligence can further enhance system capabilities.

In conclusion, the proposed system demonstrates a practical and scalable

approach to developing intelligent humanoid robots by combining AI and IoT technologies. It provides a strong foundation for future research and development in service robotics, smart environments, and human-centric automation, contributing to the advancement of next-generation interactive robotic systems.

REFERENCES

[1] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot Modeling and Control*. New York, NY, USA: Wiley, 2006.

[2] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Berlin, Germany: Springer, 2016.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779–788.

[4] G. Jocher*et al.*, "Ultralytics YOLOv8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[5] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017.

[6] A. Radford *et al.*, "Whisper: Robust speech recognition via large-scale weak supervision," OpenAI, 2022.

[7] A. Banks and R. Gupta, "MQTT version 3.1.1," OASIS Standard, 2014.

[8] Raspberry Pi Foundation, "Raspberry Pi 5 Documentation," 2024. [Online]. Available: https://www.raspberrypi.org

[9] G. Bradski, "The OpenCV library," *Dr. Dobb's Journal of Software Tools*, 2000.

[10] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ, USA: Prentice Hall, 2021.