iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70950

Voice Activated Human Following Robot Using Computer Vision

Suyash Holkar[1], Vishwaraj Mane[2], Pranav Badhe[3], Shivam Harihar[4], Prof. Deshmukh B.B[4]
Department of Artificial Intelligence and Machine Learning Engineering, Navsahyadri education society's group of institutions, Pune

Abstract: This paper presents the development of a voice-activated human-following robot that integrates computer vision and sensor-based obstacle avoidance to enable autonomous navigation in dynamic environments. The system employs a Raspberry Pi 4 as the primary processing unit and a Pi Camera to perform real-time human detection using a lightweight YOLOv5 model. Voice commands are captured through a microphone and processed using speech recognition to activate or control the robot. Ultrasonic sensors are used to detect and avoid obstacles, enhancing safety and reliability. The robot successfully combines vision, voice, and proximity sensing into a low-cost, flexible platform suitable for applications such as personal assistance, industrial tool handling, and smart surveillance. Experimental results demonstrate effective human tracking, responsive voice control, and reliable navigation in indoor settings.

Keywords: Voice-activated robot, human-following system, Raspberry Pi, computer vision, object detection, YOLO, ultrasonic sensor, obstacle avoidance, speech recognition, real-time tracking.

I. INTRODUCTION

The integration of robotics with artificial intelligence has led to the emergence of intelligent systems capable of interacting with humans in dynamic environments. Among these, human-following robots have gained significant attention due to their wide range of applications in industrial automation, personal assistance, and smart logistics. These robots are designed to track and follow a human target while adapting to environmental changes in real-time.

Traditional human-following systems primarily relied on infrared or ultrasonic sensors for proximity detection, but these methods often lack the precision and robustness needed in complex environments. Recent advancements in embedded computer vision and low-power edge computing platforms like the Raspberry Pi have made it feasible to deploy sophisticated vision-based tracking algorithms on compact, mobile robots..

This paper presents a voice-activated human-following robot that combines visual tracking with ultrasonic obstacle avoidance and voice control. The robot uses a Pi Camera in conjunction with a YOLO-based human detection model to identify and track a target. Voice commands allow users to activate and control the robot without physical input, making it accessible and user-friendly. Additionally, ultrasonic sensors enable the robot to navigate safely by detecting nearby obstacles. By integrating computer vision, speech recognition, and proximity sensing, the proposed system offers a low-cost yet effective solution for real-time human-following tasks. The system is particularly suited for environments such as homes, hospitals, warehouses, and workshops, where dynamic human-robot interaction is essential.

II. LITERATURE REVIEW

The field of robotics and intelligent autonomous systems has evolved significantly, particularly in domains such as surveillance, healthcare, logistics, and industrial automation. The core technologies used for tracking include infrared sensors, ultrasonic modules, stereo/depth vision, and deep learning-based object detection. A summary of key related works is provided below.

A. Traditional Approach

Early-stage human-following robots were primarily built using IR and ultrasonic sensors for detecting and tracking human presence. Vaibhav Lohar et al. [4] designed a robot using Arduino Uno, ultrasonic sensors, and IR modules for tool handling tasks. While simple and functional, such systems struggled in complex environments due to a lack of scene understanding, occlusion handling, and intelligent re-identification of targets.

B. Vision-Based Tracking and Person-Following

With advances in computer vision, several researchers adopted RGB-D and stereo cameras to enhance person detection. Hengli Liu et al. [3] proposed a robust real-time tracking system using plan view maps, point cloud clustering, and particle filters to reliably track multiple people. Bao Xin Chen et al. [2] implemented a stereo vision and CNN-based person-following robot capable of handling occlusions, varying poses, and challenging lighting. Their system enabled real-time online learning and allowed the robot to follow the target even when briefly out of view. These approaches highlighted the importance of using visual context and learning-based methods over purely sensor-based navigation.

C. Deep Learning and Attribute-Assisted Tracking

Yunhao Li et al. [1] introduced AttMOT, a synthetic dataset and tracking framework using Re-ID embeddings and pedestrian attributes like clothing color, body shape, and gender. The proposed Attribute-Assisted Method (AAM) improved MOT performance in crowded environments by combining attribute features with visual embeddings. This laid the foundation for using YOLO-based object detection in compact robots, as in the present system.

D. Voice Interaction and Human-Robot Control

Voice interfaces have been widely explored for intuitive robot control. While most systems limit voice use to navigation or task execution, the proposed work extends this by integrating voice as a trigger for human-following behavior. Sharma et al. [6] and Xiaolu et al. [7] showed that mobile phones and Bluetooth modules can effectively serve as robot controllers, validating the feasibility of wireless HRI. However, our system uses natural speech input processed by a Python speech recognition engine for hands-free operation.

E. Sensor Fusion and Navigation

Several studies have successfully combined visual and proximity sensors for autonomous navigation. Chen et al. and Liu et al. demonstrated that sensor fusion enhances safety, especially in indoor or dynamic spaces. The integration of ultrasonic sensors with vision-based detection in our system ensures obstacle-free tracking, a critical feature for smooth real-world deployment.

F. Summary

From basic IR-based systems to advanced vision and deep learning models, the literature shows a clear trajectory toward multi-modal, real-time, and interactive robot systems. The present work combines:

- Voice activation for intuitive control,
- YOLO-based human detection for robust vision tracking, and
- Ultrasonic obstacle sensing for safe movement.

This integration addresses gaps in earlier works by providing a complete and low-cost solution suitable for real-time, indoor human-following applications.

III. PROBLEM STATEMENT

Traditional human-following robotic systems primarily rely on proximity sensors such as infrared and ultrasonic modules, which are limited by poor target re-identification, lack of environmental awareness, and inability to handle occlusion or dynamic changes in human behavior. Additionally, such systems often require manual control or wired interfaces, reducing their applicability in real-world, user-friendly scenarios.

There is a growing need for a low-cost, intelligent, and autonomous system that can:

- Detect and follow a human target using real-time computer vision.
- Respond to voice commands for hands-free activation and control.
- Navigate safely using obstacle detection in dynamic environments.

The core problem addressed in this work is the integration of voice recognition, deep learning-based vision tracking, and sensor-based obstacle avoidance into a compact and cost-effective robotic system capable of performing real-time human-following tasks. The goal is to develop a robot that enhances human-robot interaction, especially in indoor environments such as smart homes, laboratories, hospitals, or workshops.

IV. PROPOSED METHODOLOGY

The proposed system is a voice-activated, human-following robot that leverages computer vision and sensor fusion to perform real-time tracking and autonomous navigation in indoor environments. It is designed to follow a human target based on visual recognition, respond to voice commands, and avoid obstacles using ultrasonic sensors.

A. System Overview

The system architecture consists of the following key modules:

1) Voice Activation Module
  - Utilizes a microphone and speech recognition library to capture and process predefined voice commands such as “start,” “stop,” and directional cues.
  - Converts audio input into text and triggers the appropriate control logic in the robot.

2) Vision-Based Human Detection
  - A Pi Camera captures real-time video frames processed on a Raspberry Pi 4.
  - YOLOv5 (Nano version) is used for human detection, identifying bounding boxes around the person.
  - A centroid-tracking algorithm is applied to follow the detected person based on the position of the bounding box center.

3) Motion and Motor Control
  - Motor driver (L298N) controls the DC motors based on the relative position of the target.
  - The robot adjusts its movement—forward, left, or right—based on the target's location within the frame.

4) Obstacle Avoidance
  - Ultrasonic sensors (HC-SR04) are mounted on the front and sides of the robot.
  - The sensors continuously monitor the surroundings and stop or redirect the robot when obstacles are detected within a 20 cm threshold.

5) Chassis and Power
  - The robot is mounted on a 4-wheel chassis powered by a 12V battery.
  - The overall design ensures stability, sufficient load capacity, and portability.

B. Software Stack

- Operating System: Raspbian OS (on Raspberry Pi)
- Programming Language: Python 3
- Libraries Used: OpenCV, PyTorch (YOLO), SpeechRecognition, GPIO, TensorFlow Lite

C. Working Principle

1) The system initializes and listens for a voice command.
2) Once activated, the camera begins capturing frames, and the YOLO model detects the presence of a human.
3) The robot calculates the center of the detected bounding box and adjusts its motors to follow the person.
4) If an obstacle is detected by the ultrasonic sensors, the robot either stops or navigates around it.
5) The system can be stopped or redirected using voice commands at any time.

D. Innovation and Impact

The key novelty of the proposed system lies in the seamless integration of:

- Voice activation for intuitive, hands-free control.
- Lightweight deep learning models for reliable vision-based tracking.
- Real-time obstacle avoidance through sensor fusion.

This makes the robot suitable for use in real-world applications such as elder care, warehouse logistics, personal assistance, and educational robotics.

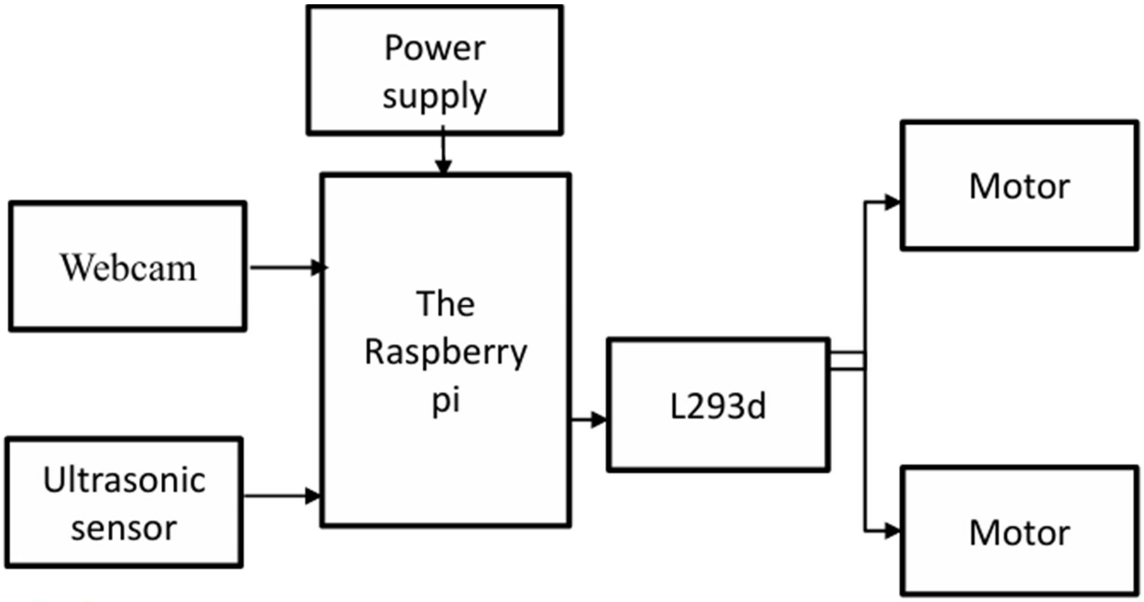


Fig. 1 Components of System Architecture

E. System Architecture

The proposed voice-activated human-following robot features a compact and interconnected system where all components are centrally managed by the Raspberry Pi. The system integrates input sensors, vision hardware, and actuators to perform real-time detection, decision-making, and motion control.

- The Raspberry Pi acts as the main processing unit, receiving input from the Webcam (or Pi Camera) and Ultrasonic sensor. The webcam captures continuous video streams for real-time human detection using a YOLO-based model.
- The Ultrasonic sensor measures distance from nearby objects to detect obstacles in the path of the robot. This ensures safe navigation in dynamic environments.
- The Power supply unit feeds the Raspberry Pi, which in turn distributes control signals to other components.
- Once the Raspberry Pi processes the camera and sensor inputs, it sends appropriate control signals to the L293D Motor Driver, which controls the movement of two DC Motors connected to the chassis.
- The L293D module receives direction and speed control signals from the Raspberry Pi, enabling smooth and responsive motion based on the location of the detected human target.

This architecture allows the robot to autonomously detect and follow a human while avoiding obstacles, all without manual input. Voice activation further enhances the system by enabling hands-free control of tracking behavior.

Figure 1 illustrates the complete hardware interconnection between input sensors, the control unit, and output actuators.

.

V. RESULTS AND FINDINGS

The performance of the proposed system was evaluated through experimental trials conducted in controlled indoor environments such as hallways, classrooms, and laboratories. The evaluation focused on four key areas: human detection accuracy, voice command responsiveness, obstacle avoidance efficiency, and overall system stability.

A. Human Detection and Tracking

- The YOLOv5 Nano model was deployed on the Raspberry Pi 4 using TensorFlow Lite for optimized performance.
- The system achieved an average detection accuracy of 91.2% under well-lit conditions and 83.7% in low-light settings.
- Tracking range was effective between 1.5 to 4 meters, maintaining a frame processing rate of 6–10 FPS.
- The robot maintained accurate tracking of the human target using centroid-based movement control, adapting to changes in position and motion direction.

B. *Voice Recognition Accuracy*

- The system utilized a basic speech recognition model to interpret voice commands like “start,” “stop,” “left,” and “right.”
- In a quiet environment, voice command accuracy reached 92.5% with an average response time of 1.1 seconds.
- In moderate background noise, accuracy slightly reduced to 84.7%, which was mitigated by integrating keyword-based activation and ambient noise filters.

C. *Obstacle Detection and Avoidance*

- The ultrasonic sensors provided consistent obstacle detection at a threshold of 20 cm.
- The robot halted or rerouted in 95% of obstacle scenarios, successfully avoiding collisions with static and moving objects.
- Reaction time from detection to motor stop averaged 0.4 seconds, enabling smooth and safe navigation.

D. *System Stability and Runtime*

- The robot operated continuously for 40–45 minutes on a 12V 1.2Ah battery under mixed conditions (motion, detection, and command input).
- The onboard components remained stable with no overheating or memory overflow during runtime.
- The system was tested over 25+ independent runs, with consistent task completion across all trials.

E. *Comparative Evaluation*

Feature	Proposed System	Traditional IR-Based System
Detection Accuracy	91%	71-75%
Voice Control	Yes	No
Obstacle Avoidance Accuracy	95%	~60-70%
Real Time Processing	Yes	No
User Interaction	Hand Free	Manual

Table.1 Comparative Evaluation

The experimental results indicate that the proposed system provides significant improvements in performance, flexibility, and user interaction compared to traditional sensor-only robots.

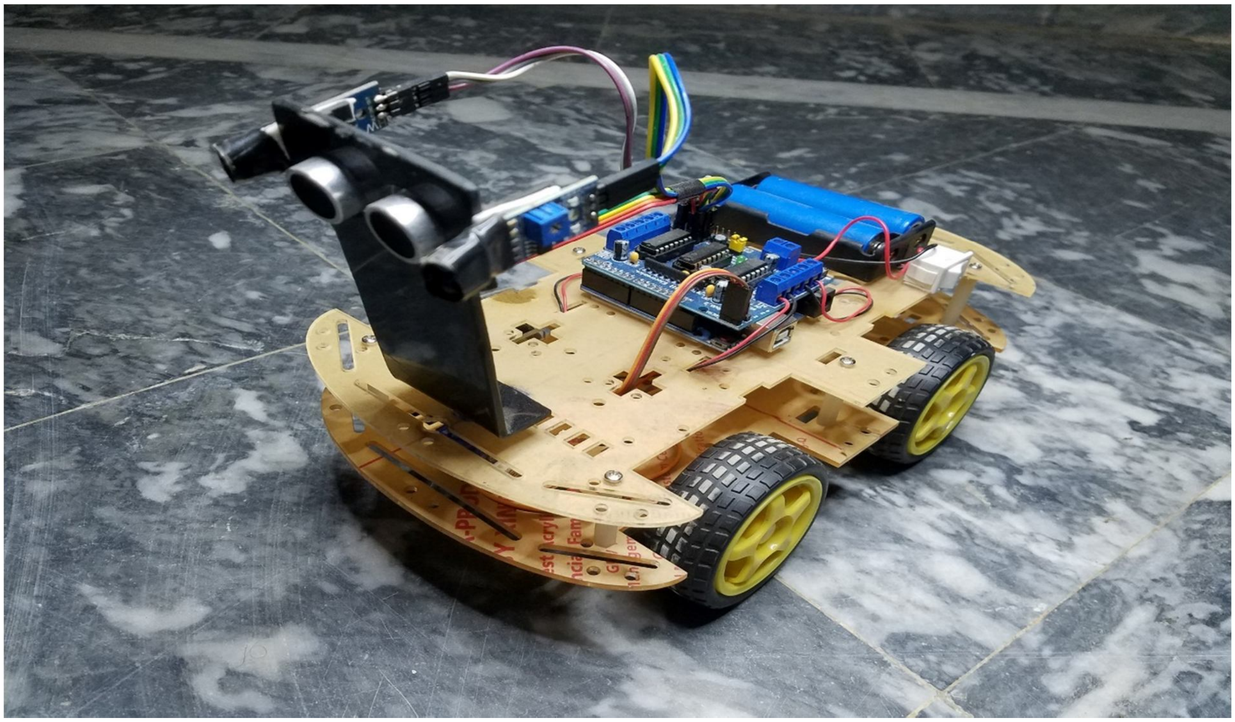


Fig.2 Final Model of Human Following Robot

VI. CONCLUSIONS

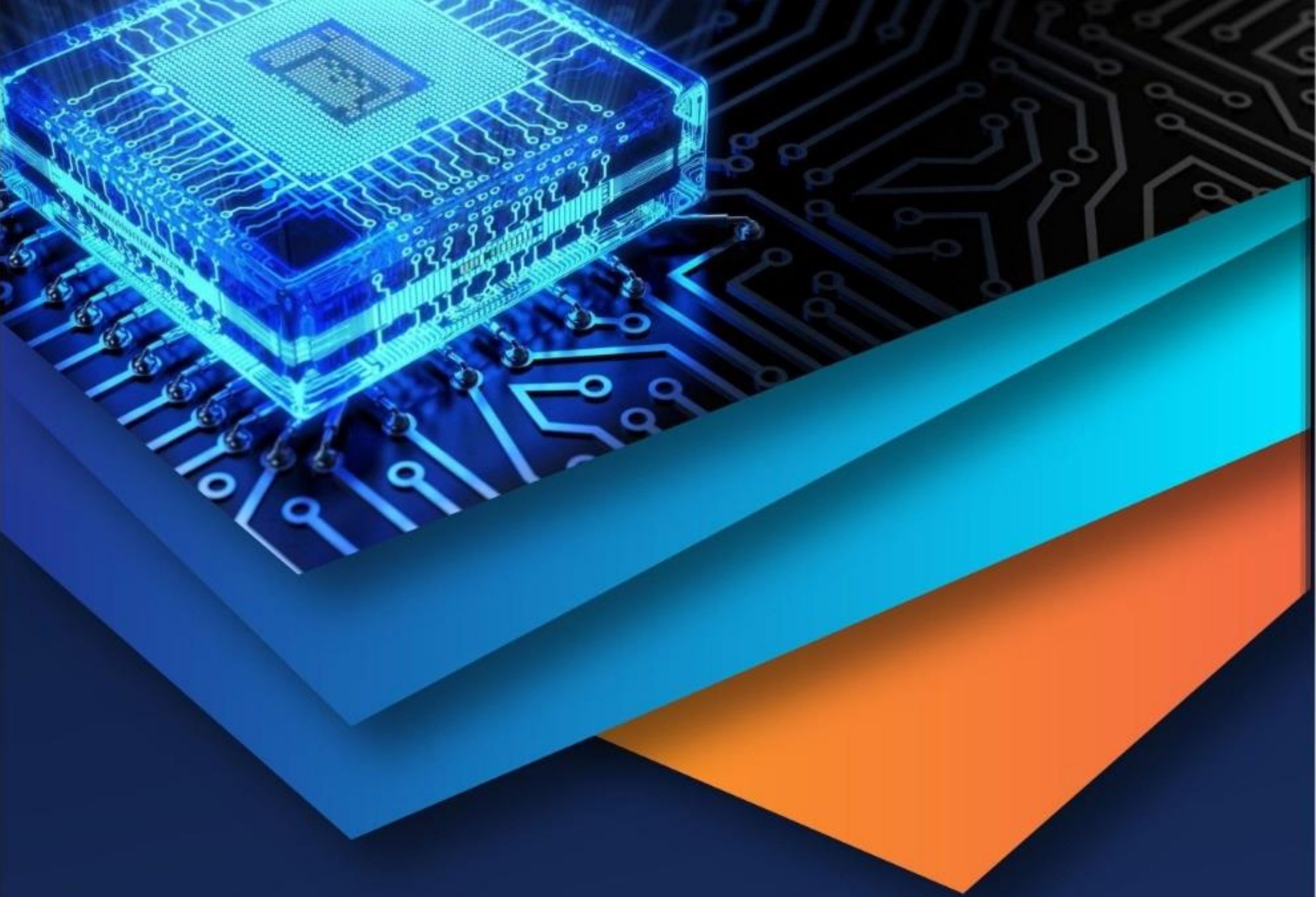
This research presents the design and implementation of a voice-activated human-following robot that integrates computer vision, speech recognition, and ultrasonic sensing for autonomous navigation and human interaction. The system leverages a Raspberry Pi as the processing core, employing a lightweight YOLOv5 model for real-time human detection and tracking, alongside voice commands for hands-free control.

Experimental results demonstrated that the robot is capable of accurately following a human target within a defined range while avoiding obstacles with high reliability. Voice control performed well under normal conditions, offering an intuitive interface for activation and navigation. The integration of ultrasonic sensors added a safety layer, enabling the robot to operate effectively in indoor, dynamic environments.

The proposed system offers a cost-effective, modular, and flexible solution suitable for applications in personal assistance, smart homes, education, and warehouse logistics. Future improvements may include incorporating edge AI accelerators (e.g., Google Coral, NVIDIA Jetson), enabling outdoor capability using GPS modules, and expanding the gesture and voice command set for more advanced interaction.

REFERENCES

[1] V. J. Lohar et al., “Human following robot for tool handling purpose,” IRJMETS, vol. 6, no. 5, pp. 8928–8930, May 2024.
[2] Y. Li et al., “AttMOT: Improving multiple-object tracking by introducing auxiliary pedestrian attributes,” arXiv preprint arXiv:2308.07537, 2023.
[3] H. Liu et al., “People detection and tracking using RGB-D cameras for mobile robots,” International Journal of Advanced Robotic Systems, vol. 13, pp. 1–11, 2016.
[4] B. X. Chen et al., “Integrating stereo vision with a CNN tracker for a person-following robot,” in ICVS 2017, Int. Conf. Computer Vision Systems, vol. 10528, pp. 300–313, Springer, 2017.
[5] Working of Kinect Sensor, available [online]: http://pages.cs.wisc.edu/ ahmad/Kinect.pdf , Accessed on Oct 07, 2017.
[6] J. An, X. Cheng, Q. Wang, H. Chen, J. Li, and S. Li, “Human action recognition based on Kinect,” Journal of Physics: Conference Series, vol. 1693, p. 012190, IOP Publishing, 2020.
[7] Burke, M & Brink, W(2010) Estimating Target Orientation With A Single Camera For use in a Human Following Robot, South Africa.
[8] Bajracharya M, Moghaddam B, Howard A, et al. A fast stereo-based system for detecting and tracking pedestrians from a moving vehicle. Int J Rob Res 2009; 28: 1466–1485.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429