Voice Activated Human Following Robot Using Computer Vision

Prof. Yuga Shedge,
Vishvaraj Mane, Pranav Badhe, Suyash Holkar, Shivam Harriar,
Department of AI-ML Engineering,
Navsahyadri Groups of Institute Faculty Engineering

KEYWORDS

Kinect sensor, human-robot interaction, autonomous following, localization, mapping.

ABSTRACT

Abstract- The Kinect sensor captures the 3-dimensional information of the surroundings and recognizes the human body by retrieving the depth information that does not require wearing any kind of intrusive sensors. Firstly, this robotic system follows an individual by detection of torso point that is required for steering and maintaining a fixed safe distance for localization and mapping and provides a robust and reliable system. The proposed system can be utilized in a wide variety of applications such as work assistants, luggage carrying carts.This robotic system is reliable and strong since it can identify a person's torso point and keeps a safe distance for mapping and localization.Work aid and luggage carrying carts are only two of the many applications for the suggested technology.

1. INTRODUCTION

The growing population is contributing to significant challenges in the care of elderly individuals. Recent advancements in robotics and automation technology have occurred at a rapid pace, largely due to their applications in industrial, military, and medical fields. However, these technologies have not yet found widespread application in daily life. This situation presents an opportunity to harness technology to support the elderly. In this regard, this paper introduces a robotic system designed to assist with various daily needs. The human-following robot functions as a personal assistant, accompanying users as they move throughout their environment. This system allows for the easy transportation of luggage from one place to another, effectively acting as a robotic trolley. Its important applications include transporting items in malls, airports, hospitals, and other contexts that require hauling assistance.Numerous studies have been conducted on "Assisting Robots," which typically use sensory equipment to identify and locate their target people. Even while techniques using sensors like infrared and ultrasonic have been developed, they frequently lack the precision required for efficient tracking and do not make it easier to identify human targets specifically.[1] The three basic color components—red, green, and blue—are captured by RGB cameras, which are widely utilized in image processing applications for tracking and object detection. However, because of their dependability and depth perception, 3D motion sensors—also known as depth sensors—have become the go-to option for human monitoring. Consequently, focus has turned to visual sensors that don't require the usage of integrating devices, such smartphone apps.

Much study has been done in order to address the issues that are currently being faced. As far as we know, vision-based approaches and multi-sensor fusion techniques, such as laser range finders, sonar, and Lidar, can be used to detect and track humans. In order to identify and track people, vision-based detection and tracking systems usually use certain traits like skin detection, facial recognition, upper body or torso identification, or even full-body tracking. As an alternative, different visual cues could be combined to make a tracking system that is more flexible and robust. Additionally, multi-sensor human tracking and detection systems use the advantages of several sensors to improve accuracy and dependability.

2. Research Elaborations

The first person-following robot research was conducted in 1998, using color and contour information for tracking. Similar color-based tracking methods, such as the usage of H-S Histograms in the hue-saturation-value (HSV) color space, were used in later studies. These approaches, however, found it difficult to control occlusions and variations in appearance. The necessity that the target's motion differ from the background's limited the potential of early optical flow works, including the ones mentioned. [2]Yoshimi et al. presented simpler feature-based methods that included edges, corners, and more color and texture information. Numerous studies also looked into pre-trained appearance models. Lucas-Kanade features, SIFT features, HOG features, and a combination of height and gait with appearance-based features are some noteworthy feature-based techniques.

The use of Selected Online Ada-Boosting for online learning, which refines target search parameters using depth information, was a relatively recent development in 2017[3]. Researchers have also looked into a variety of sensor technologies, such as RGBD cameras like Kinect and laser-based systems, for person-following robots.[4] Kinect can only be used indoors, but laser systems might not be able to operate in places like hospitals and shopping malls. On the other hand, our method uses a stereo camera, which can be used both indoors and outdoors.

Using range data from stereo vision, Bajracharya et al.[5] developed an integrated system for human identification, localization, and tracking from a moving vehicle. In order to categorize pedestrians, the scene was divided into various sections of interest, and shape elements were taken out. To generate the likelihood map for person candidates, two distinct plan view maps—the occupancy map and the height map—were used.[6] To ascertain which of the human candidates identified match the individual being tracked, we used Kalman filtering and an MLE. Mun ̧oz-Salinas et al.22 also used a height map and an occupancy map to record the items' height and volume. Furthermore, a confidence map was used to combine the data from several cameras. The tracking of individuals in the fused plan view map was then suggested using a particle filter method.

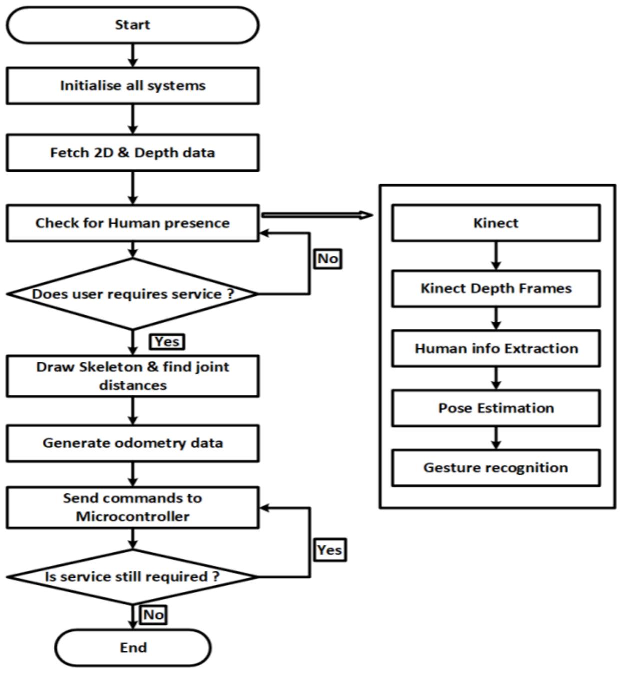
- **Sensor Fusion:**
  - **Depth Sensors (like Kinect):** These sensors capture 3D information about the environment, allowing the robot to accurately perceive the human's position and distance.
  - **Cameras:** Cameras provide visual data, enabling the robot to recognize and track the human's features.
  - **Lidar:** Lidar sensors offer precise distance measurements, helping the robot navigate obstacles and maintain a safe distance.
- **Object Detection and Tracking:**
  - **Computer Vision Algorithms:** These algorithms process the sensor data to detect and identify the human target.
  - **Tracking Algorithms:** Once the human is detected, tracking algorithms continuously monitor their movement and adjust the robot's path accordingly.
- **Motion Planning and Control:**
  - **Path Planning Algorithms:** These algorithms generate optimal paths for the robot to follow the human, considering obstacles and dynamic environments.
  - **Control Systems:** The robot's actuators (motors) are controlled to execute the planned movements, ensuring smooth and precise following behavior.

3. Results and Findings

The human-machine interaction often needs another object such as an RF tag. To overcome this limitation, this project presents a mobile robot car using the Microsoft Kinect sensor as its visual tool. The process flow for the working of the developed system [7]. The methodology used in this project can be categorized into three sub-sections.

A. Human Gesture Recognition Gesture recognition is the study of interpreting the gestures of an individual. Human-machine interaction is possible with gesture [8]recognition, thus controlling a robot can be done without the requirement of any other device. Useful poses of the human body can be made by the user which involves tracking the hands or other joints of the user

B. Motion Tracking Depth data from the inbuilt depth sensor of Kinect and the skeletal joints distance data are used for locating the person in the vicinity of the Kinect. This depth information is then used by the robot to identify the human target.[9] The human has to lock onto the robot by performing a particular locking gesture. Once a user is locked, the depth data is used to follow the user, while maintaining a safe distance from the target. The user can also terminate the service as desired[10].



Candidate segmentation is approached through the identification of probable intervals of human or object instances within the depth distribution of a 3D point cloud. A human candidate is detected by locating a local maximum in this depth distribution and selecting a depth interval centered around that maximum. This is based on the observation that any object represented in a point cloud consists of a collection of points that share similar depth values or are organized in spatially adjacent clusters. To facilitate candidate detection, our methodology simplifies these spatially adjacent clusters into a two-dimensional (2D) ground plane.

c. Components

Kinect Sensor- The Kinect sensor, developed by Microsoft, is a revolutionary device that has significantly impacted the field of human-computer interaction (HCI). It's a depth camera that captures 3D information about the environment, enabling a wide range of applications, from gaming to robotics.

Raspberry Pi -A Raspberry Pi is a small, affordable computer that you can connect to a keyboard, mouse, and monitor. It's incredibly versatile and can be used for a wide range of projects, from simple tasks like web browsing to complex projects like building robots and home automation systems.

Ultrasonic Sensor-An ultrasonic sensor is a device that measures distance by emitting ultrasonic sound waves and then listening for the echo that bounces back. By measuring the time it takes for the sound wave to travel to the object and return, the sensor can calculate the distance.

A. Task 1: System operation in the presence of single human in the field of vision The first task of the robot is to detect the human presence (if any) in the field of vision. The system checks, whether any of the humans want to avail of the service of the robot, locks the user if specific locking actions such as raising left hand, raising right hand, raising both hands, etc. are performed, these actions are predefined for availing the services of the robot. There is a two-level locking mechanism used for better control i.e. two different poses are used sequentially to lock the user. shows an individual raising his right hand for locking. B.

B. Task 2: Obstacle avoidance is a fundamental capability for autonomous robots, enabling them to navigate safely in dynamic environments. It involves detecting obstacles in the robot's path and taking appropriate actions to avoid them

4. CONCLUSION

The aim of enhancing human-machine interaction is realized through the application of gesture recognition methodologies and Kinect sensor technology. The Microsoft Kinect functions as the vision sensor for the robot, facilitating human detection and tracking. This proposed system is characterized by a reduced number of components in comparison to traditional stereo vision approaches, while still demonstrating effective performance in scenarios involving both single and multiple individuals. Additionally, the system can incorporate audio-based user control and authentication, leveraging the Kinect's built-in microphone for audio input, which can be processed by the system's processor to execute the appropriate actions. Future enhancements may include the implementation of collision avoidance techniques to bolster the system's robustness and versatility.

.

5. ACKNOWLEDGMENT

I would like to express my sincere gratitude to Shedge Y. D, Voice Activated Human Following Robot, for their invaluable guidance, support, and encouragement throughout this research. Their expertise and insightful feedback were instrumental in shaping this work.

I am also deeply indebted to [Name of Collaborators/Researchers], for their valuable contributions, suggestions, and assistance. Their collaboration significantly enhanced the quality of this research.

I would like to thank [Name of Institution/Department] for providing the necessary resources and facilities to carry out this research.

Finally, I would like to express my heartfelt thanks to my family and friends for their unwavering support and encouragement. Their belief in me has been a constant source of motivation.

6. REFERENCES

[1] [1] 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.

[2] Working of Kinect Sensor, available [online]: http://pages.cs.wisc.edu/ ahmad/Kinect.pdf , Accessed on Oct 07, 2017.

[3] 4. Calisi, D., Iocchi, L., Leone, R.: Person following through appearance models and stereo vision using a mobile robot. In: VISApp Workshop on Robot Vision, pp. 46–56 (2007)

[4] 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.

[5] Mun ˜oz-Salinas R, Aguirre E, Garca-Silvente M, et al. Multi agent system for people detection and tracking using stereo vision in mobile robots. Robotica 2009; 27: 715–727.

[6] Burke, M & Brink, W(2010) Estimating Target Orientation With A Single Camera For use in a Human Following Robot, South Africa

[7] Gu, J, Ding ,X ,Wang, S, wu : Action Recognition from Recovered 3-d Human Joints, systems, Man, and Cybernetics(2010)

[8] Kinect Processing Library, available [online]:https://github.com/technicolorenvy/Processing-Libraries/tree/master/SimpleOpenNI , Accessed on Nov 10, 201

[9] Zhang, L., van der Maaten, L.: Structure preserving object tracking. In: Proceed ings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1838–1845 (2013)

[10] Zhang, K., Zhang, L., Yang, M.H.: Real-time object tracking via online discrimi native feature selection. IEEE Trans. Image Process. 22(12), 4664–4677 (2013)