Person Tracking Control of Mobile Robots Using A Lightweight Object Detection and Tracking System

Yu-Chen Chiu, Huan-Wei Hsu, and Chi-Yi Tsai*
Department of Electrical and Computer Engineering
Tamkang University
151 Ying-chuan Road, Tamshui District, New Taipei City 251, Taiwan R.O.C.
*Email: chiyi_tsai@mail.tku.edu.tw

2024 8th International Conference on Robotics and Automation Sciences (ICRAS) | 979-8-3503-5250-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRAS62427.2024.10654470

Abstract—**This paper aims to develop a lightweight person tracking control system to facilitate real-time pedestrian tracking tasks for mobile robots. The proposed system utilizes deep learning technology and lightweight network architecture to perform pedestrian detection tasks to achieve powerful pedestrian detection capabilities. Subsequently, the integration of multi-object tracking technology enables the system to effectively track identified pedestrian targets, thereby enabling real-time pedestrian tracking of mobile robots. Experimental results show that the proposed lightweight person tracking control system exhibits robust pedestrian following performance, stability, and effectiveness under various scenarios and environmental conditions. These properties increase the potential of the proposed system in many practical applications.**

Keywords—Multi-object tracking, pedestrian detection, deep learning, person tracking control, mobile robots

I. Introduction

With the advent of the era of artificial intelligence, countries around the world are placing increasing emphasis on the development of the artificial intelligence and robotics industry. The current applications of robots can be broadly categorized into two main types: industrial robots [1] and service robots [2]. The primary purpose of industrial robots is to replace humans in tasks that involve high risk, precision, or repetition. In recent years, advanced countries have been experiencing an ongoing phenomenon of aging populations and declining birth rates. As a result, service robots continue to be developed and emphasized, and with the advancement of technology, they have gradually been applied to daily life.

In the past, industrial robots were primarily applied through the use of Automated Guided Vehicles (AGVs) in material handling applications [3]. However, it requires pre-established mobile tracks for the vehicle to follow a planned path. This approach lacks flexibility and cannot adapt to unforeseen circumstances [4]. With technological advancements, many factories are accelerating the level of factory automation [5]. AGVs have also evolved into forms that are flexible, agile, and capable of collaborating with human operators [6], addressing the practicalities of the factory environment. Furthermore, to enhance the tracking control performance of AGVs, the control systems of these robots often integrate a tracking algorithm to estimate the target's motion trajectory. Therefore, developing a real-time object detection and tracking system for mobile robots has become an important development task.

One of the primary challenges in current visual detection and tracking frameworks is how to match target detection results with their corresponding trajectories. In [7], the authors proposed a method to handle unreliable detection by collecting candidates from both detection and tracking outputs. The intuition behind generating redundant candidates is that detection and tracking can complement each other in different scenarios. High-confidence detection results prevent long-term tracking drifts, while track predictions can handle noisy detection caused by occlusion. In [8], the authors performed robust detection and tracking based on laser data. The method is effective in a variety of indoor and outdoor environments and on different robotic platforms (Smart Electric Wheelchair and Clearpath Husky). In [9], the authors utilized AI-driven depth information for indoor positioning and real-time human tracking, significantly increasing its applicability and flexibility. In [10], the authors used attention modules, masking-sensitive hard sample mining, and masking-sensitive loss to address detection errors in highly occluded pedestrians, assigning higher weights for such cases.

This paper proposes a real-time lightweight pedestrian tracking network based on multi-object tracking (MOT). The proposed approach utilizes images and depth information obtained from an RGBD camera for object detection and visual tracking. Moreover, we draw insights from literature [7-10] regarding human tracking methods and improved occlusion strategies by employing lightweight detectors [11] and efficient trackers [12]. By incorporating MOT [13] and a proportional controller [14], we achieve visual motion tracking capabilities, developing a visually guided mobile robot applicable in various domains. The contributions of this paper include: (1) the design of a visual system for mobile robots to handle person tracking tasks, (2) the use of the MOT algorithm to enhance the robustness of the tracking algorithm, capable of functioning in low-light conditions, and (3) the use of the lightweight detector for faster computational speed.

II. The Proposed Person Tracking Control System

In this section, we introduce the proposed person tracking control system based on a lightweight object detector and a multi-object tracker.

A. Design of Lightweight Vision System for Mobile Robots

Fig. 1 shows the block diagram of the proposed system, which primarily uses an RGBD camera to obtain two signals,

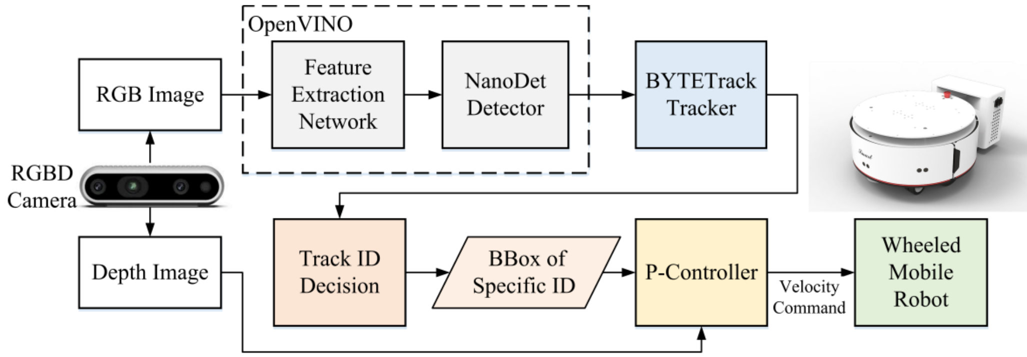

Fig. 1. Block diagram of the proposed lightweight object detection and tracking system for visual tracking control of wheeled mobile robots.

namely the RGB Image and Depth Image, for detecting target objects in visual tracking tasks. The entire system process can be divided into four stages; each stage is introduced step by step below.

First, the RGB Image captured by the camera is input into the Feature Extraction Network. The role of this network is to extract features of pedestrians from the image, which will be used for object detection in subsequent steps. Meanwhile, the Depth Image contains depth information, which helps to more accurately understand distance information in the scene.

Secondly, the processed image is used for object detection through a lightweight detector, and OpenVINO is used for acceleration to improve detection efficiency. After the detection is completed, we obtain the bounding box (BBox) of the target object, which is a rectangular box surrounding the detected object.

To achieve powerful tracking performance, we integrated the MOT tracker into the control system. The MOT tracker is based on the BBox of the target object in the previous frame and the current frame. However, if there are too many objects in the frame during tracking, the tracker may cause tracking errors. Therefore, we designed a Track ID Decision module to focus on a specific tracking target. Through this module, we can select a specific target ID to ensure that the tracking system focuses on tracking a specific target.

After locking the target ID that needs to be tracked, the proposed system uses the P-Controller module to perform following control on the mobile robot to ensure that the tracked target continues to be in the center of the image. Finally, the proposed system outputs linear velocity and angular velocity commands to control the mobile robot to achieve target following actions.

The entire system architecture integrates deep learning, object detection, object tracking, and robot control techniques, allowing us to effectively implement the person tracking function of mobile robots. This provides a solid foundation for the application of autonomous tracking control systems in mobile robots.

B. *NanoDet Detector*

NanoDet [11] is an ultra-fast and lightweight anchor-free target detection model designed for mobile devices. Its lightweight design enables it to run on resource-constrained devices such as smart cameras, mobile devices, or embedded systems. It achieves faster detection speeds compared to other detectors. This algorithm supports various target detection tasks, including object detection, pedestrian detection, vehicle detection, and can identify multiple categories of objects. Since the proposed person tracking control system needs to be executed on a platform without GPU accelerator and meet the requirements of fast response, we chose to use NanoDet as the default detector.

C. *BYTETrack Tracker*

BYTETrack [12] is an efficient MOT algorithm. Its core idea is to use the confidence of each detection BBox to establish association instead of using the traditional association strategy. This allows BYTETrack to better handle situations such as object occlusion and appearance changes. The architecture of BYTETrack consists of three components: the detection head, associator, and tracker. In this study, we set the detection head as NanoDet, the associator uses the Hungarian algorithm for association, and the tracker employs the Kalman filter for tracking. Compared to other trackers, it is more lightweight, making tracking processing of moving targets faster.

D. *Track ID Decision*

In the Track ID Decision step, the system first counts the number of occurrences of tracked users. The user must remain within a specific range of the image until the count reaches a preset threshold, at which point the system resets the tracker and assigns the user an ID of 1. Here, the default threshold is set to 10. Subsequently, the visual tracking system continues to send the BBox information of the user with ID 1 to the P-Controller for following control.

This step ensures that the system focuses on a specific user and achieves effective identification and tracking of targets through a defined counting mechanism and ID reset. Such a system design not only enhances the stability of tracking, but also ensures tracking accuracy in complex environments, providing a reliable foundation for the following control of mobile robots.

E. *P-Controller of Wheeled Mobile Robots*

Fig. 2 shows the block diagram of the proposed P-Controller module for visual tracking control of wheeled mobile robots. Let x and d denote the x coordinate and distance value of the center point of the BBox of the target

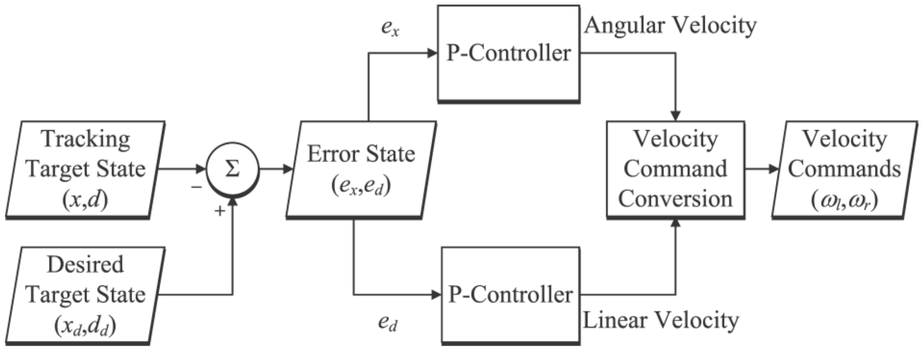

Fig. 2. Block diagram of the proposed P-Controller module for visual tracking control of wheeled mobile robots.

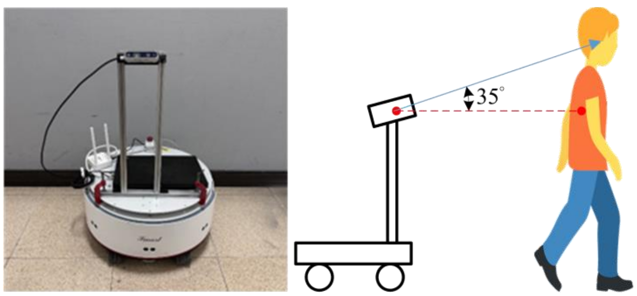

Fig. 3. The EAIBOT SMART mobile robot used in this study. An RGBD camera is installed on the platform for person detection and tracking.

person, respectively. Define (x,d) and (x_d,d_d) as the tracking target state and desired target state of the control system, respectively. Then, the error state of the tracking system is given by

$$e(t) = \begin{bmatrix} e_x(t) \\ e_d(t) \end{bmatrix} = \begin{bmatrix} x_d \\ d_d \end{bmatrix} - \begin{bmatrix} x(t) \\ d(t) \end{bmatrix}. \quad (1)$$

In this study, we set the desired target state as $(x_d,d_d)=(W/2, 0.8)$, where W is the width of the image. Accordingly, the P-Controller calculates control commands based on the current error state so that the system state can stably reach the desired state. Let v and ω be the linear velocity and angular velocity output by the controller. Then, the P-Controller can be expressed by the following formula

$$\begin{bmatrix} \omega(t) \\ v(t) \end{bmatrix} = \begin{bmatrix} k_x & 0 \\ 0 & k_d \end{bmatrix} e(t), \quad (2)$$

where k_x and k_d are the proportional gains of the x and d states, respectively. Finally, a velocity conversion is required to convert the linear and angular velocities to the left and right wheels' angular velocities (ω_l,ω_r) such that

$$\begin{bmatrix} \omega_l(t) \\ \omega_r(t) \end{bmatrix} = \begin{bmatrix} -D/2R & 1/R \\ D/2R & 1/R \end{bmatrix} \begin{bmatrix} \omega(t) \\ v(t) \end{bmatrix}, \quad (3)$$

where R and D denote, respectively, the radius of both the robot's wheels and the distance between the center of the two wheels. The above control process ensures that the mobile robot can flexibly and accurately follow the tracking target, enhancing the overall performance of the visual tracking control system.

III. Experimental Results

This section presents the results of the proposed visual tracking control system for the person tracking task on a wheeled mobile robot equipped with an RGBD camera. We implemented the proposed tracking control system on a laptop equipped with i7-6700HQ CPU, and the input image size is 320x320. The test results show that the overall system processing speed is about 18-21 FPS, effectively achieving real-time detection and tracking performance.

A. The Mobile Robot Platform Used in the Experiments

In order to validate the tracking control performance of the proposed method, we implemented the proposed control system on a mobile robot platform for experimental testing. Fig. 3 shows the mobile robot platform used in the experiments. We used the EAIBOT SMART mobile robot equipped with a RealSense D435i camera as the test platform for this study. In order to reduce the possibility of capturing other pedestrians, the RGBD camera was installed with a fixed elevation angle of 35 degrees.

B. Results of Person Tracking Control

Fig. 4 shows the experimental results of the person tracking control of the proposed method. During the following process, although multiple pedestrians appeared in the scene, the mobile robot still successfully followed the target person with ID 1. In addition, observing the corresponding linear velocity and angular velocity commands, it can be found that the output control command changes very smoothly. The output linear velocity is stably maintained at the desired linear velocity, and the output angular velocity also correctly responds to the moving direction of the person in front. This implies that the mobile robot moves smoothly and continuously while following the target person.

On the other hand, it can be observed from Fig. 5 that each pedestrian will be clearly assigned a different ID. Since the proposed tracking control system only tracks the target with ID 1, the mobile robot still focuses on tracking the specific person with ID 1 in a multi-pedestrian scenario.

C. Robustness of the Proposed Tracking Control System

The experimental results in Fig. 6 show the robustness of the proposed tracking control system to ambient light changes. It can be observed from Fig. 6 that during the following control process, the mobile robot can still stably follow the specific person in front even under non-uniform background light

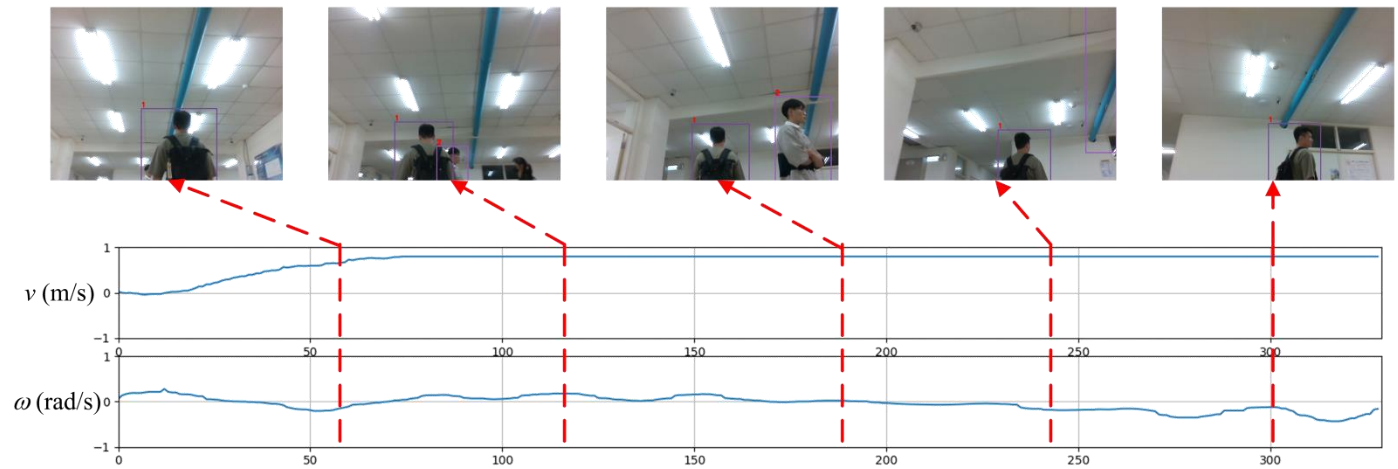

Fig. 4. Experimental results of person tracking control and the corresponding linear and angular velocity commands of the mobile robot.

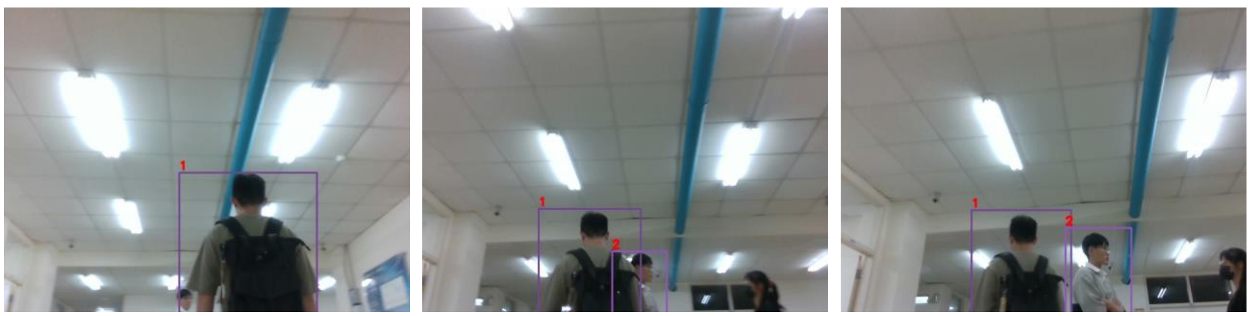

Fig. 5. The proposed tracking control system can continuously lock the specific person with ID 1 and continue tracking in a multi-pedestrian environment.

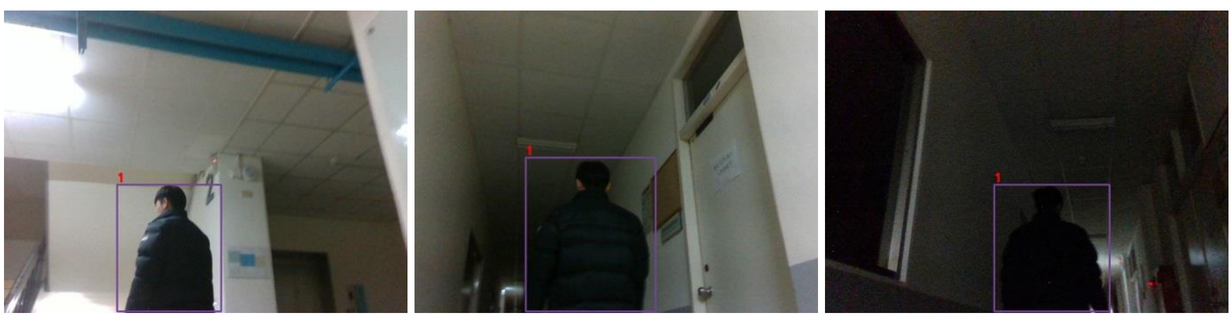

Fig. 6. Robustness of the proposed method to non-uniform ambient light conditions.

conditions. In addition, the mobile robot can maintain accurate tracking of people under very low lighting conditions. The above experimental results effectively validate the tracking performance and robustness of the proposed tracking control system. A video clip of the experimental results is available online at [15].

IV. Conclusion

In this paper, we develop a real-time lightweight object detection and tracking system based on deep learning and MOT technology. The advantage of using MOT technology is that it can track multiple people in the scene at the same time and distinguish the tracking target person from other pedestrians. When combined with the proposed Track ID Decision module, the ID number of a specific person can be further reset to improve the identification of the target being followed. In addition, the proposed control system operates with an RGBD camera, which helps to obtain distance information stably and output stable linear velocity commands.

On the other hand, the proposed system uses a lightweight detector to meet the needs of real-time computing and has tracking robustness to non-uniform ambient light conditions. Experimental results show that the tracking control performance of the proposed tracking control system is quite robust and is suitable for person tracking control applications of mobile robots, and provides important insights for the development of tracking control systems for mobile robots.

Acknowledgment

This research was supported by the National Science and Technology Council of Taiwan, R.O.C., under grant NSTC 112-2221-E-032-036-MY2.

REFERENCES

[1] T. Ribeiro, I. Garcia, D. Pereira, J. Ribeiro, G. Lopes and A. F. Ribeiro, "Development of a prototype robot for transportation within industrial environments," *2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, Coimbra, Portugal, 2017.

[2] Z. A. barakeh, S. alkork, A. S. Karar, S. Said and T. Beyrouthy, "Pepper Humanoid Robot as a Service Robot: a Customer Approach," *2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART)*, Paris, France, 2019.

[3] V. T and N. U. H D, "Vision-Based Automated Guided Vehicle," *2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)*, Tumakuru, India, 2023.

[4] De Ath, George, and Richard Everson, "Visual object tracking: The initialisation problem," *2018 15th Conference on Computer and Robot Vision (CRV)*. IEEE, 2018.

[5] W. Xia, J. Goh, C. A. Cortes, Y. Lu and X. Xu, "Decentralized coordination of autonomous AGVs for flexible factory automation in the context of Industry 4.0," *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, Hong Kong, China, 2020.

[6] J.-I. Ham, I.-C. Ryu, J.-O. Park, J.-W. Joeng, S.-C. Kim and H.-S. Ahn, "Indoor Pedestrian-Following System by a Drone with Edge Computing and Neural Networks: Part 2 - Development of Tracking System and Monocular Depth Estimation," *2023 23rd International Conference on Control, Automation and Systems (ICCAS)*, Yeosu, Korea, Republic of, 2023.

[7] L. Chen, H. Ai, Z. Zhuang, C. Shang, "Real-time multiple people tracking with deeply learned candidate selection and person re-identification," *2018 IEEE international conference on multimedia and expo (ICME)*. IEEE, 2018.

[8] A. Leigh, J. Pineau, N. Olmedo and H. Zhang, "Person tracking and following with 2D laser scanners," *2015 IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, WA, USA,pp.726-733, 2015.

[9] J. Xie, Y. Pang, M. H. Khan, R. M. Anwer, F. S. Khan and L. Shao, "Mask-Guided Attention Network and Occlusion-Sensitive Hard Example Mining for Occluded Pedestrian Detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 3872-3884, 2021.

[10] Jung-Il Ham, In-Chan Ryu, Jun-Oh Park, Jae-Woo Joeng, Sung-Chang Kim, Hyo-Sung Ahn, "Indoor Pedestrian-Following System by a Drone with Edge Computing and Neural Networks: Part 2 - Development of Tracking System and Monocular Depth Estimation," *23rd International Conference on Control, Automation and Systems (ICCAS)*, Yeosu, Republic of Korea, 2023.

[11] R. Lyu, "NanoDet-Plus: Super fast and high accuracy lightweight anchor-free object detection mode," aviailable online: https://github.com/RangiLyu/nanodet, 2021.

[12] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, X. Wang. "Bytetrack: Multi-object tracking by associating every detection box," *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022.

[13] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, T.-K. Kim, "Multiple object tracking: A literature review," *Artificial intelligence* 293,2021.

[14] Mohd Saifizi Saidonr, Hazry Desa, Md Noor Rudzuan, "A differential steering control with proportional controller for an autonomous mobile robot," *IEEE 7th International Colloquium on Signal Processing and its Applications*, Penang, Malaysia, 2011.

[15] A demo video of the proposed person tracking control system. Available online: https://www.youtube.com/watch?v=QEv8kB3UvLE