COBEM 2021 26th International Congress of Mechanical Engineering

ABCM Associação Brasileira de Engenharia e Ciências Mecânicas
November 22-26, 2021, Virtual Congress, Brazil

COB-2021-1187
PEOPLE FOLLOWING SYSTEM FOR HOLONOMIC ROBOTS USING AN RGB-D SENSOR

Gabriel Fischer Abati
João Carlos Virgolino Soares
Pontifícia Universidade Católica do Rio de Janeiro, Department of Mechanical Engineering. R. Marquês de São Vicente 225, Gávea, Rio de Janeiro, RJ - Brazil - 22451-900
fischerabati@gmail.com
virgolinosoares@gmail.com

Marcelo Gattass
Pontifícia Universidade Católica do Rio de Janeiro, Department of Informatics. R. Marquês de São Vicente 225, Gávea, Rio de Janeiro, RJ - Brazil - 22451-900
mgattass@tecgraf.puc-rio.br

Marco Antonio Meggiolaro
Pontifícia Universidade Católica do Rio de Janeiro, Department of Mechanical Engineering. R. Marquês de São Vicente 225, Gávea, Rio de Janeiro, RJ - Brazil - 22451-900
meggi@puc-rio.br

Abstract.
People following is an important task in mobile robotics, with applications in several areas, such as industry, hospitals and home services. Recent advances on deep learning techniques and the availability of RGB-D sensors resulted in a higher robustness in identifying the target during these tasks. This paper presents a system for people following using an RGB-D sensor. The proposed method uses object tracking, combined with a vision-based controller that exploits the advantages of the high maneuverability of holonomic robots. Experimental tests are performed with a mecanum-wheeled robot equipped with an RGB-D camera to evaluate the effectiveness of the proposed approach.

Keywords: *People following, Human detection and tracking, Holonomic robots, RGB-D Sensor*

1. INTRODUCTION

Collaborative tasks between humans and robots have benefited from recent developments on systems and techniques that assure safety and robustness in their interaction. Autonomous robots capable of following a person are beneficial in industry, health, elderly care, military, and film making. Obstacle avoidance, person tracking in crowded environments, and re-identification of a lost target person are examples of challenges that an autonomous robot must surpass to follow its human companion.

The choice of the sensor of a person-following robot has a high influence on overcoming those challenges. There are several types of sensors that can be used in the people-following task. An RGB-D camera (Depth Sensor) is a good choice due to its low cost and richness of information. A single RGB-D camera can provide both geometric and semantic data about the environment. The use of cameras allows performing high-level tasks such as Object Detection and Tracking. These tasks are very useful for people-following, especially with the recent advances of convolutional neural networks (CNN), which improved both the speed and precision of its results. Moreover, the choice of the locomotion type can considerably improve the performance of people-following. Holonomic robots are well suited for this application due to their ability to move without changing their orientation and working in tight spaces.

Thus, this paper presents a complete people-following system for holonomic robots with an RGB-D camera. The system uses a YOLOv3-based (Redmon and Farhadi, 2018) object detection model, and SORT (Bewley *et al.*, 2016), a tracking method that combines a Kalman filter (Kalman, 1960) and the Hungarian algorithm (Kuhn, 1955) for movement prediction and data association, respectively. Furthermore, a vision-based controller is developed to drive a mecanum-wheeled robot to follow a target person, using the object tracking results combined with the RGB-D camera's depth map. The system is tested in a robot with four mecanum wheels and a Kinect v2 sensor, following a person in an indoor environment. The main contribution of our work is the low-cost of the methodology, both in financial and computational aspects, as it only needs a single RGB-D camera to operate, and achieves a robust performance in real-time.

This paper is organized as follows. Section 2 shows related works, Section 3 describes the proposed methodology, Section 4 shows the experiments and results, and Section 5 presents the conclusion and future works.

2. RELATED WORK

People following robots can be useful in several tasks and scenarios, such as for medical purposes (Ilias *et al.*, 2014), (Engelberger, 1993), supermarket assistance (Sales *et al.*, 2016) and guiding people through museums and shopping malls (Thrun *et al.*, 1999). Different techniques can be used for people following tasks, as well as different sensors, such as cameras or Laser Range Finders (LRF).

For instance, Jung *et al.* (2012) and Lee *et al.* (2006) proposed a people following system based only on 2D LRF measurements, usually identifying body parts, such as torso or legs. Kobilarov *et al.* (2006), Bellotto and Hu (2006) and Itoh *et al.* (2006) used a combination of an RGB camera and a LRF sensor, that resulted in a better following method, since visual sensors acquire richer information allowing a full body detection.

However, a single RGB-D camera can provide both geometric and semantic data about the environment. Thus, currently it is the most used sensor for people following systems. For example, Babaians *et al.* (2015) combined a skeleton tracker with an OpenTLD visual tracker using a Kalman filter in a mecanum robot to follow a target.

Mi *et al.* (2016) also utilizes the skeleton approach with a human walking model, in order to predict the next step of a target person. This method was applied in a non-holonomic robot, and the results can lead to non-smooth trajectories, depending on the target's behaviour. The skeleton tracking method was also applied by Chi *et al.* (2018) to create a dataset for human gait recognition for a service robot. Despite having satisfactory results in terms of accuracy, the system was not tested in real time.

Yang and Song (2019) developed a human-following system with a combination of the Inception deep network (Szegedy *et al.*, 2016) and an Extended Kalman Filter (EKF) for robust people detection. The Inception network, however, has a lower accuracy for human activity recognition than other networks, such as YOLOv3, according to Mustafa *et al.* (2020).

Algabri and Choi (2020) described a deep learning-based person following system for a differential mobile robot. In 2021 they focused on a recovery system and person trajectory prediction. Even though their framework achieved real-time performance using CPU, their approach of extracting clothes color from the target person led to failures when there were drastic illumination changes and more people with similar clothing.

3. METHODOLOGY

Figure 1 shows the flowchart of the proposed methodology. The system receives as input the data provided by the RGB-D sensor and outputs commands to the motors. We propose a system that combines the semantic information of RGB images to localize the target person, and the geometric information of the depth images to control the distance of the robot from the person. We use a double-proportional controller based on the distance between the image centroid and the bounding box centroid.

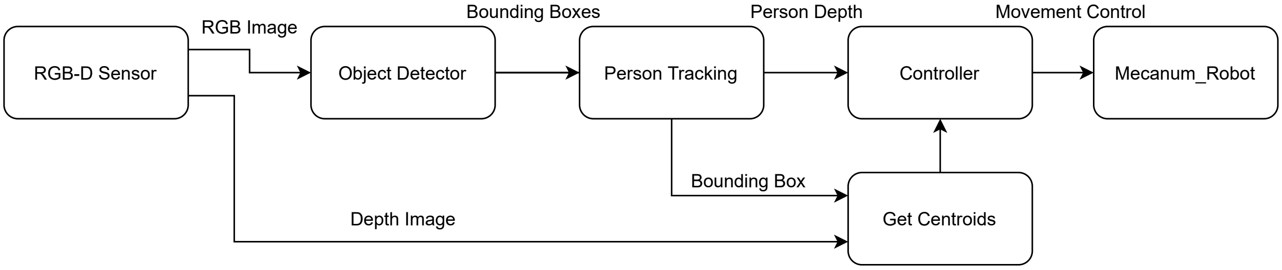


Figure 1: System Flowchart

3.1 Robot Model

Holonomic robots have the controllable degrees of freedom equal to the total degrees of freedom (Siegwart *et al.*, 2004). Thus, they can move in any direction while changing its orientation. Mecanum-wheeled robots are a specific type of holonomic robots with four mecanum wheels. Our method is tested in a mecanum-wheeled robot, but the general formulation can be applied for any type of holonomic robot. The model of the robot in shown in Fig. 2, where ϕ_n is the rotation of the wheel n. Equation 1 describes the forward kinematics of the system, where $\dot{\phi}_n$ is the angular velocity of the wheel n, and $\dot{X}_r$ and $\dot{Y}_r$ are the velocities in the X_r and Y_r directions, respectively.

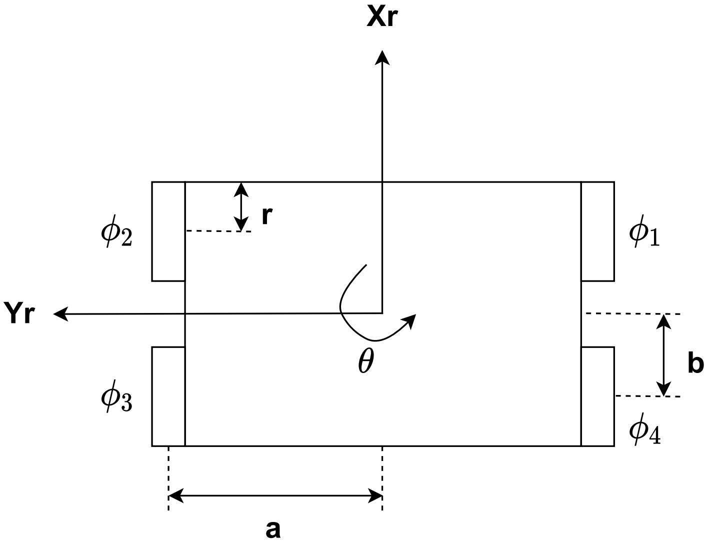


Figure 2: Robot Model

$$\begin{bmatrix} \dot{X}_r \\ \dot{Y}_r \\ \dot{\theta} \end{bmatrix} = \frac{r}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ \frac{1}{a+b} & -\frac{1}{a+b} & -\frac{1}{a+b} & \frac{1}{a+b} \end{bmatrix} \begin{bmatrix} \dot{\varphi}_1 \\ \dot{\varphi}_2 \\ \dot{\varphi}_3 \\ \dot{\varphi}_4 \end{bmatrix} \tag{1}$$

3.2 Object Detection

Object Detection is the task of identifying and localizing objects in an image. This task used to be performed with classic computer vision techniques. However, recent CNN architectures such as YOLO (Redmon and Farhadi, 2018) and R-CNN (Girshick *et al.*, 2014) considerably improved the results of this task. Our method uses CYTi (Soares *et al.*, 2021) for Object Detection, a network based on YOLOv3-tiny that is specifically trained to work in human environments. It has an accuracy similar to YOLOv3, but with the efficiency of the YOLOv3-tiny, running in 50 FPS using only CPU. It receives as input an RGB image and outputs the bounding box size, position, class and a confidence score. Figure 3 shows an example of the output of CYTi.

3.3 Object Tracking

The people bounding box is tracked over time using SORT (Bewley *et al.*, 2016), an efficient algorithm for 2D tracking. SORT uses a Kalman filter (Kalman, 1960) to predict the position of the bounding box in the next frame. The state of each target is defined by Eq. 2, where u_o and v_o are the pixel coordinates of the center of the target, s_o represents the scale, and r_o is the aspect ratio of the bounding box. SORT also uses the Hungarian algorithm (Kuhn, 1955) to perform data association in an efficient manner, using the intersection over union distance between detected and predicted boxes as a metric. Figure 4 shows a person being tracked, with its respective ID. This method allows the system to deal with more than one person in the image. If a second person appears in the scene, the robot continues to follow the initial person using its ID. However, this method cannot deal with long-term occlusion and re-identification.

$$\mathbf{x}_o = [u_o, v_o, s_o, r_o, \dot{u}_o, \dot{v}_o, \dot{s}_o] \tag{2}$$

3.4 Controller

The controller is designed to simultaneously maintain the robot in a safe distance from the target person, and maintain the vision of the robot aligned with the center of the tracked bounding box. Figure 4 show the block diagram of the described control design. The distance error is calculated using a determined threshold and the depth measurement of the center of the tracked person. The alignment error is calculated by the difference between the RGB image centroid pixel coordinate and the bounding box centroid pixel coordinate. Both error values are then multiplied by a proportional gains, tuned experimentally. To keep the same scale between the two controller sections, the alignment error is converted from pixel to millimeter using Eqs. 3-5, where c_x and c_y compose the optical center in pixels, and f_x and f_y compose the focal length in pixels, obtained through camera calibration. Finally, the two error values are added and sent to the motors as velocity commands using Eq. 1.

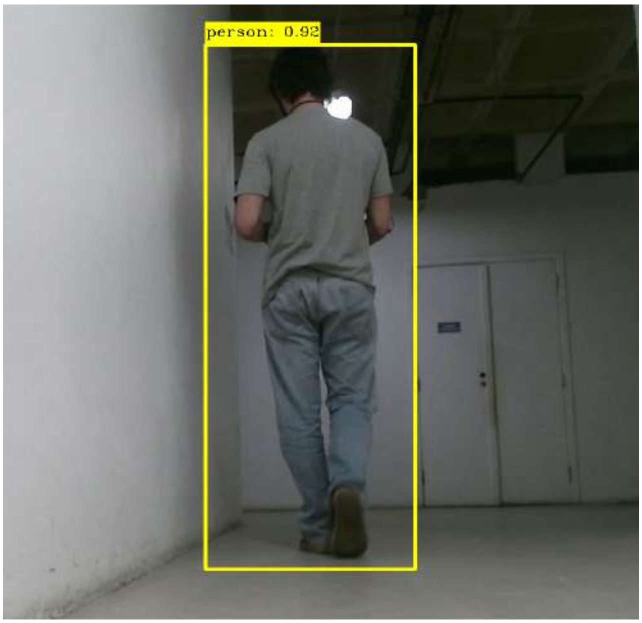


Figure 3: People Detection with CYTi

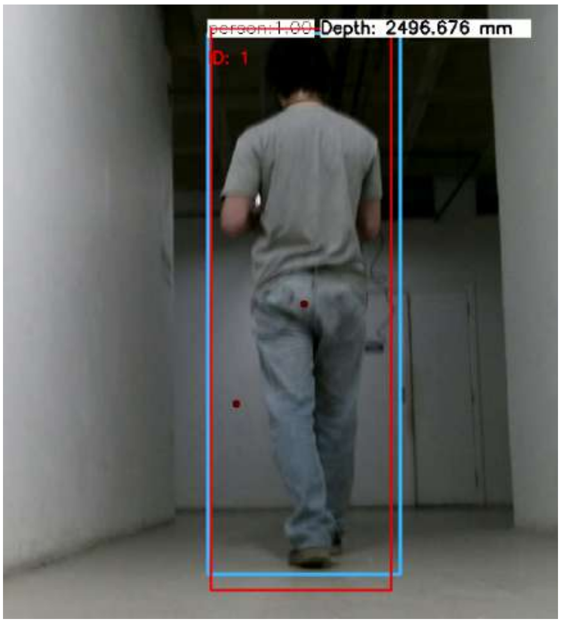


Figure 4: People Tracking

$$X = \frac{(u - c_x)}{f_x} Z \tag{3}$$

$$Y = \frac{(v - c_y)}{f_y} Z \tag{4}$$

$$Z = depth\_image(u, v) \tag{5}$$

The direction which the robot must move is determined by the position of the target person centroid inside the frame. The pixels from the image are divided in a "X" shape into four regions, as shown in Fig. 6. If the object centroid is inside the upper or lower region, then the robot moves forward. If the object centroid is located in the left or right regions, then the robot moves in the diagonal left or right direction, respectively.

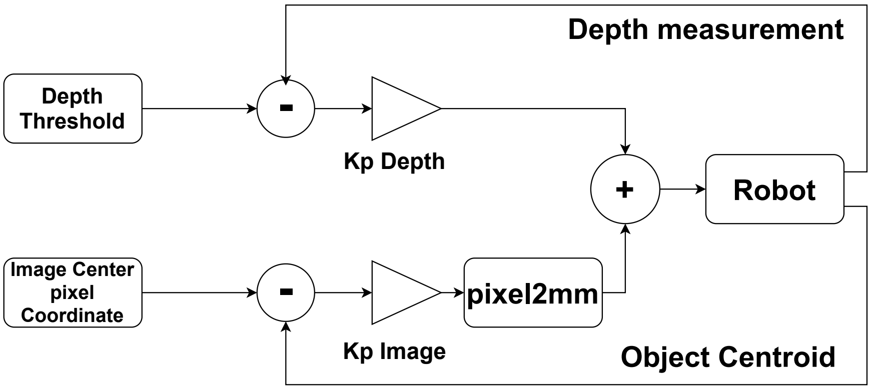


Figure 5: Controller block diagram

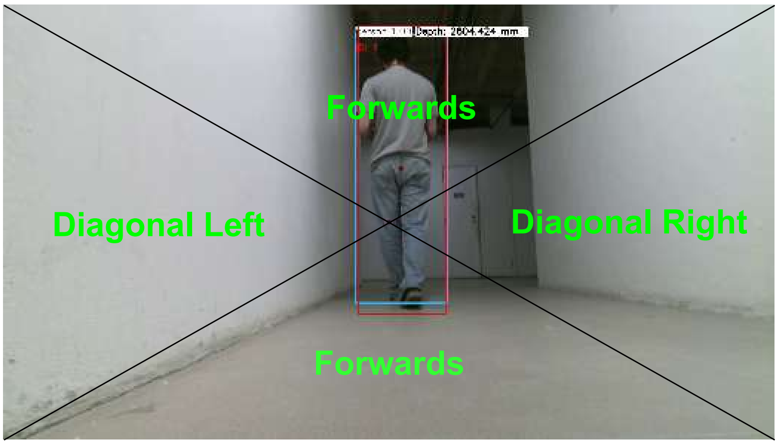


Figure 6: Robot direction decision

3.5 Implementation Details

The methodology was implemented in C++, using the OpenCV library for CYTi object detection. The Libfreenect2 driver (Xiang *et al.*, 2016) is used to obtain data from the Kinect V2 sensor. The high-level system communicates with the mecanum base via Serial messages. The system can run in real time using only CPU in a standard notebook.

4. RESULTS

The tests were made in the Robotics Laboratory of PUC-Rio using a mecanum-wheeled robot developed in our previous work (Soares *et al.*, 2019). Figure 7 shows the robot, composed of four independent DC motors attached to mecanum wheels, the associated control eletronics, a Kinect V2 camera, and a notebook with an Intel Core i7 2.60 GHz and 16 GB of RAM running Ubuntu Linux 18.04 LTS.

The robot had to follow a person through a corridor. Figure 8 shows a sequence of images taken from the robot's perspective. At the start of the experiments, a single person stays in front of the robot and becomes the target person. The target person starts to move forward, and the robot follows the person with $ID = 1$. After a few seconds, a second person appears and moves to the opposite direction. Figure 9 presents the robot's trajectory. The video footage from the experiment was saved and applied to Crowd-SLAM (Soares *et al.*, 2021) in order to estimate the poses of the robot, represented by the blue rectangles in the image. The robot does not change its trajectory when the second person appears, since the person has an $ID \neq 1$.

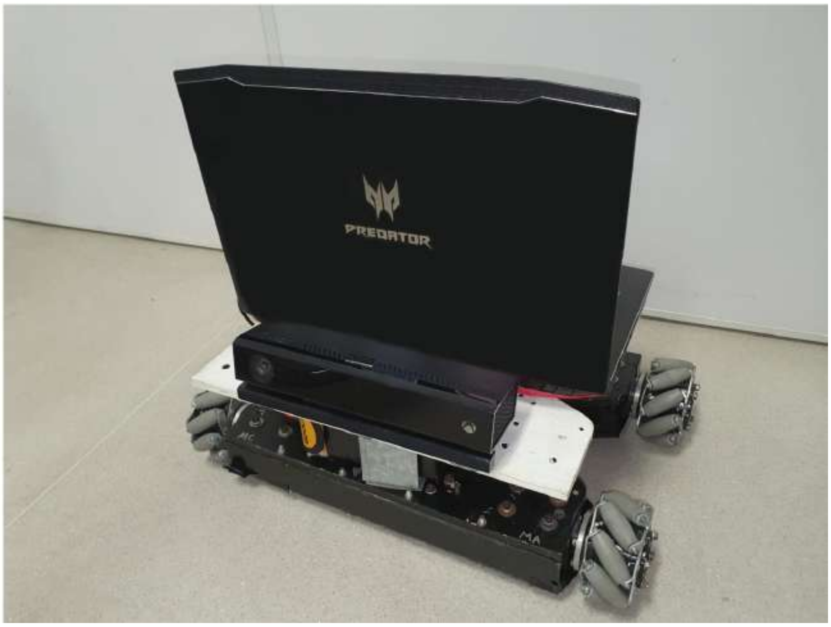
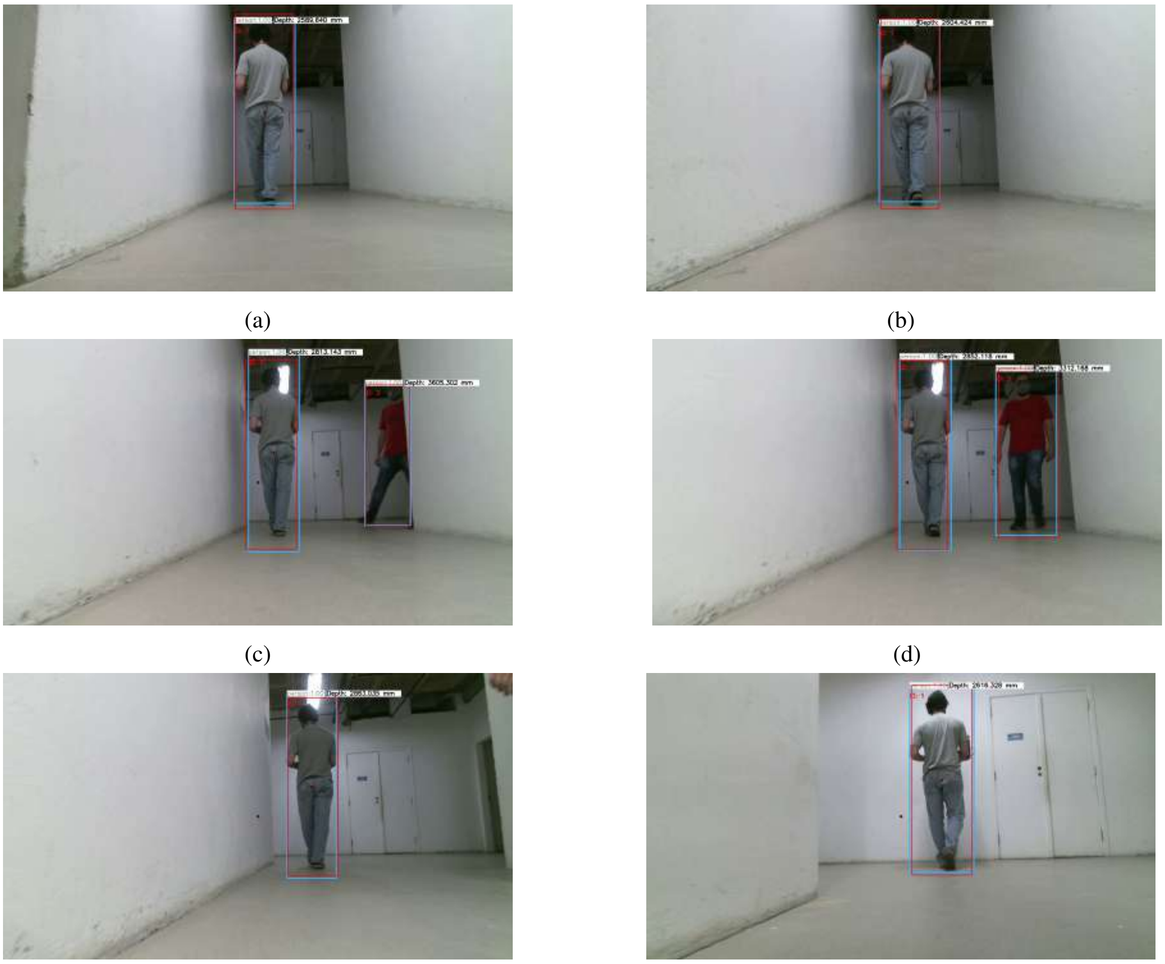


Figure 7: Mecanum robot with an RGB-D sensor



(a)

(b)

(c)

(d)

(e)

(f)

Figure 8: Images taken from the robot's perspective during the task

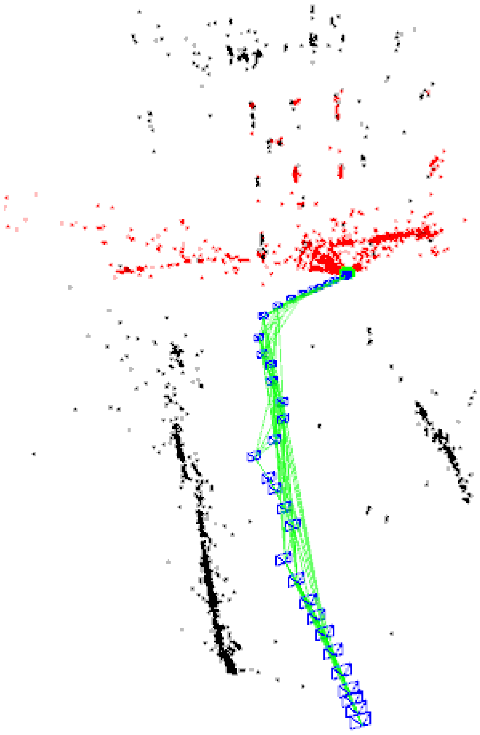


Figure 9: Trajectory of the robot

5. CONCLUSIONS

This paper proposed a people following system for holonomic robots using an RGB-D sensor. The framework consists of an efficient Object Detection module, a robust people tracking algorithm, and a vision-based controller designed for holonomic robots. Experiments were conducted in a mecanum-wheeled robot with a Kinect V2 following a person in an indoor environment. The results showed that the robot successfully followed the target person in real time, even when there was another person in the scene. Potential applications of the proposed method include object transportation for airports and industrial sites. Future works include the improvement of the tracking system using an Extended Kalman filter, and the development of a re-identification system that deals with long-term occlusions and lost tracks.

6. REFERENCES

Algabri, R. and Choi, M., 2020. "Deep-learning-based indoor human following of mobile robot using color feature". *Sensors*, Vol. 20, No. 9.

Algabri, R. and Choi, M., 2021. "Target recovery for robust deep learning-based person following in mobile robots: Online trajectory prediction". *Applied Sciences*, Vol. 11, No. 9.

Babaians, E., Korghond, N.K., Ahmadi, A., Karimi, M. and Ghidary, S.S., 2015. "Skeleton and visual tracking fusion for human following task of service robots". In *3rd RSI International Conference on Robotics and Mechatronics*.

Bellotto, N. and Hu, H., 2006. "Vision and laser data fusion for tracking people with a mobile robot". In *IEEE Internacional Conference on Robotics and Biomimetics*.

Bewley, A., Ge, Z., Ott, L., Ramos, F. and Upcorft, B., 2016. "Simple online and realtime tracking". In *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP)*.

Chi, W., Wang, J. and Meng, M.Q., 2018. "A gait recognition method for human following in service robots". In *IEEE Transactions on systems, Man, and Cybernetics: Systems*.

Engelberger, J.F., 1993. "Health-care robotics goes commercial: the helpmate experience". In *Robotica*.

Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. "Rich feature hierarchies for accurate object detection and semantic segmentation". In *Conference on Computer Vision and Pattern Recognition*.

Ilias, B., Shukor, S.A.A., Yaacob, S., Adom, A. and Razali, M.H.M., 2014. "A nurse following robot with high speed kinect sensor". In *Journal of Engineering and Applied Science)*.

Itoh, K., Kikuchi, T., Takemura, H. and Mizoguchi, H., 2006. "Development of a person following mobile robot in complicated background by using distance and color information". In *32nd Annual Conference on IEEE Industrial Eletronics*.

Jung, E., Yi, B. and Yuta, S., 2012. "Control algorithms for a mobile robot tracking a human in front". In *International Conference on Intelligent Robots and Systems*.

Kalman, R.E., 1960. "A new approach to linear filtering and prediction problems". *Journal of Basic Engineering*, Vol. 82, No. 1, pp. 35–45.

Kobilarov, M., Sukhatme, G., Hyams, J. and Batavia, P., 2006. "People tracking and following with mobile robot using an omnidirectional camera and a laser". In *IEEE International Conference on Robotics and Automation (ICRA)*.

Kuhn, H.W., 1955. "The hungarian method for the assignment problem". *Naval Research Logistics Quarterly*, Vol. 2, No. 1-2, pp. 83–97.

Lee, J.H., Tsubouchi, T., Yamamoto, K. and Egawa, S., 2006. "People tracking using a robot in motion with laser range finder". In *International Conference on Intelligent Robots and Systems*.

Mi, W., Wang, X., Ren, P. and Hou, C., 2016. "A system for an anticipative front human following robot". In *International Conference on Artificial Intelligence and Robotics and the Internactional Conference on Automation, Control and Robotics Engineering*.

Mustafa, T., Dhavale, S. and Kuber, M.M., 2020. "Performance analysis of inception-v2 and yolov3-based human activity recognition in videos". *SN Computer Science*, Vol. 1, No. 3.

Redmon, J. and Farhadi, A., 2018. "Yolov3: An incremental improvement". In *Computer Vision and Pattern Recognition*.

Sales, J., Marti, J.V., Marin, R., Cervera, E. and Sanz, P.J., 2016. "Comparob: The shopping cart assistance robot". In *Internacional Journal of Distributed Sensor Networks*.

Siegwart, R., Nourbakhsh, I.R. and Scaramuzza, D., 2004. *Introduction to Autonomous Mobile Robotics, Second Edition)*. MIT Press.

Soares, J.C.V., Abati, G.F., Lima, G.D., de Souza Junior, C.M. and Meggiolaro, M.A., 2019. "Project and development of a mecanum-wheeled robot for autonomous navigation tasks". In *Proceedings of the XVIII International Symposium on Dynamic Problems of Mechanics (DINAME)*.

Soares, J.C.V., Gattass, M. and Meggiolaro, M.A., 2021. "Crowd-SLAM: Visual SLAM towards crowded environments using object detection". *Journal of Intelligent & Robotic Systems*, Vol. 102, No. 50.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. "Rethinking the inception architecture for computer vision". In *Computer Vision and Pattern Recognition*.

Thrun, S., Bennewitz, M., Burgard, W., Cremers, A.B., Dellaert, F., Fox, D., Hahnel, D., Rosenberg, C., Roy, N., Schulte, J. and Schulz, D., 1999. "Minerva: A second-generation museum tour-guide robot". In *International Conference on Robotics & Automation (ICRA)*.

Xiang, L., Echtler, F., Kerl, C., Wiedemeyer, T., Lars, hanyazou, Gordon, R., Facioni, F., laborer2008, Wareham, R., Goldhoorn, M., alberth, gaborpapp, Fuchs, S., jmtatsch, Blake, J., Federico, Jungkurth, H., Mingze, Y., vinouz, Coleman, D., Burns, B., Rawat, R., Mokhov, S., Reynolds, P., Viau, P., Fraissinet-Tachet, M., Ludique, Billingham, J. and Alistair, 2016. "libfreenect2: Release 0.2". doi:10.5281/zenodo.50641. URL `https://doi.org/10.5281/zenodo.50641`.

Yang, C. and Song, K., 2019. "Control design for robotic human-following and obstacle avoidance using an rgb-d camera". In *19th International Conference on Control, Automation and Systems*.

7. RESPONSIBILITY NOTICE

The authors are solely responsible for the printed material included in this paper.