POLITECNICO DI TORINO
Repository ISTITUZIONALE

Human-Centered Navigation and Person-Following with Omnidirectional Robot for Indoor Assistance and Monitoring

Original
Human-Centered Navigation and Person-Following with Omnidirectional Robot for Indoor Assistance and Monitoring / Eirale, A., Martini, M., Chiaberge, M.. - In: ROBOTICS. - ISSN 2218-6581. - 11:5(2022), pp. 108-124. [10.3390/robotics11050108]

Availability:
This version is available at: 11583/2972211 since: 2022-10-11T08:35:49Z

Publisher:
MDPI

Published
DOI:10.3390/robotics11050108

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

06 November 2026

robotics

MDPI

Article

Human-Centered Navigation and Person-Following with Omnidirectional Robot for Indoor Assistance and Monitoring

Andrea Eirale [1,2], Mauro Martini [1,2] and Marcello Chiaberge [1,2,*]

[1] Department of Electronics and Telecommunications (DET), Politecnico di Torino, 10129 Torino, Italy
[2] PIC4SeR PoliTO Interdepartmental Center for Service Robotics, Politecnico di Torino, 10129 Torino, Italy
* Correspondence: marcello.chiaberge@polito.it; Tel.: +39-331-671-4686

Abstract: Robot assistants and service robots are rapidly spreading out as cutting-edge automation solutions to support people in their everyday life in workplaces, health centers, and domestic environments. Moreover, the COVID-19 pandemic drastically increased the need for service technology to help medical personnel in critical conditions in hospitals and domestic scenarios. The first requirement for an assistive robot is to navigate and follow the user in dynamic environments in complete autonomy. However, these advanced multitask behaviors require flexible mobility of the platform to accurately avoid obstacles in cluttered spaces while tracking the user. This paper presents a novel human-centered navigation system that successfully combines a real-time visual perception system with the mobility advantages provided by an omnidirectional robotic platform to precisely adjust the robot orientation and monitor a person while navigating. Our extensive experimentation conducted in a representative indoor scenario demonstrates that our solution offers efficient and safe motion planning for person-following and, more generally, for human-centered navigation tasks.

Keywords: autonomous mobile robots; indoor service robotics; person-following; human-centered navigation; omnidirectional robot navigation

check for updates

Citation: Eirale, A.; Martini, M.; Chiaberge, M. Human-Centered Navigation and Person-Following with Omnidirectional Robot for Indoor Assistance and Monitoring. *Robotics* **2022**, *11*, 108. https://doi.org/10.3390/robotics11050108

Academic Editor: Marco Ceccarelli

Received: 9 September 2022
Accepted: 5 October 2022
Published: 10 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Robot assistants have recently emerged as a promising solution for elderly care and monitoring in the indoor domestic environment. The increasing demand for service robotic platforms for indoor assistance has paved the way for the development of diverse robotic solutions, especially devoted to elderly care [1,2]. According to the World Population Prospects (2019) provided by the United Nations [3], life expectancy reached 72.6 years, with a future expectation of 77.1 in 2050. Furthermore, projections reveal that there will be more people aged 65 years or over than young aged 15 to 24 years by 2050 [3]. Population ageing dramatically impacts our society's organization, exacerbating delicate issues, such as the isolation of numerous vulnerable subjects and elderly people in their homes for most of the day. Moreover, the recent emergency related to the COVID-19 outbreak has further increased the need for a reliable and automatic assistance tool in both hospitals and patients' residential environments. In this scenario, robots were demonstrated to be a key technological ally in fighting the pandemic and its dramatic social effects, such as isolation [4,5]. Indeed, they can offer support to both medical staff and families whenever the services of dedicated assistive operators or volunteers are not available due to the intensive demand generated by the pandemic.

Robotic solutions are often focused on the interactive social aspects with the user [6,7], or, conversely, they try to continuously check the health status of the patient [8,9]. However, a reliable and effective navigation algorithmic stack is a necessary condition to realistically deploy a robotic platform in a cluttered environment for humans. The most recent advances in human-aware robot navigation [10] show how planning and control algorithms can be successfully adapted to social circumstances.

For the specific case of a robotic assistant that aims at constantly monitoring, accompanying, and supporting the user within the domestic or medical environment, the ability to follow the person is crucial. Indeed, person-following [11,12] is the primary challenging task to enable any visual or vocal interaction with the robot, while the user is moving around. On the other hand, reduced mobility subjects may also need the robot to accomplish desired services in the room, moving around towards different destinations. Keeping an eye on the users during the execution of such secondary functions constitutes a huge benefit for the robot assistant's main goal: monitoring the person's condition. Person-following and goal-based navigation probably represent an indoor robot assistant's two most common navigation behaviors. However, monitoring while navigating might raise some serious difficulties in the case of conventional differential drive platforms, which do not have the possibility to describe a curved motion without a change in orientation. This limitation often leads differential drive robots to lose the human target while avoiding obstacles or following an occluded path. The same argument does not apply to an omnidirectional platform: in this case, the robot can handle its motion along all horizontal plane directions without changing its orientation.

In this work, we focus our research on the development of a human-centered autonomous navigation system for a robotic assistant, which aims at fulfilling the user assistance requirement in both the described scenarios: goal-based navigation (Figure 1) and person-following (Figure 2). Hence, we decided to adopt a tiny-size omnidirectional robotic base platform, to fully exploit its kinematic advantages and propose an optimized person-following methodology, always guaranteeing collision-free trajectory planning combined with continuous visual tracking of the user. Nonetheless, our solution also enables the robot assistant to move toward the desired destination while adjusting the orientation of the platform to keep active visual contact with the user. This results in increased reliability of the robotic assistant, which is able to perform different tasks while continuously checking the status of the person and calling for help in case dangerous situations are detected.

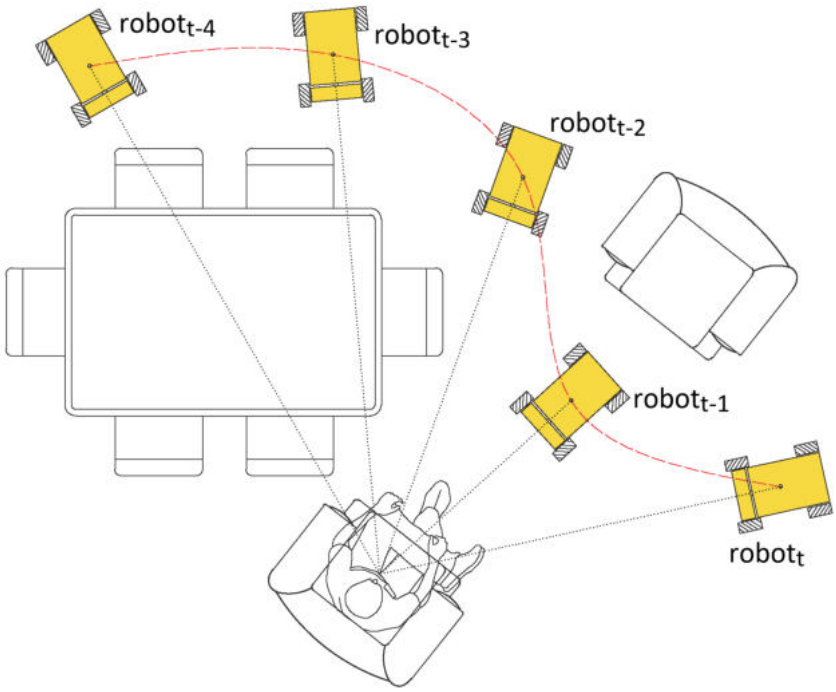
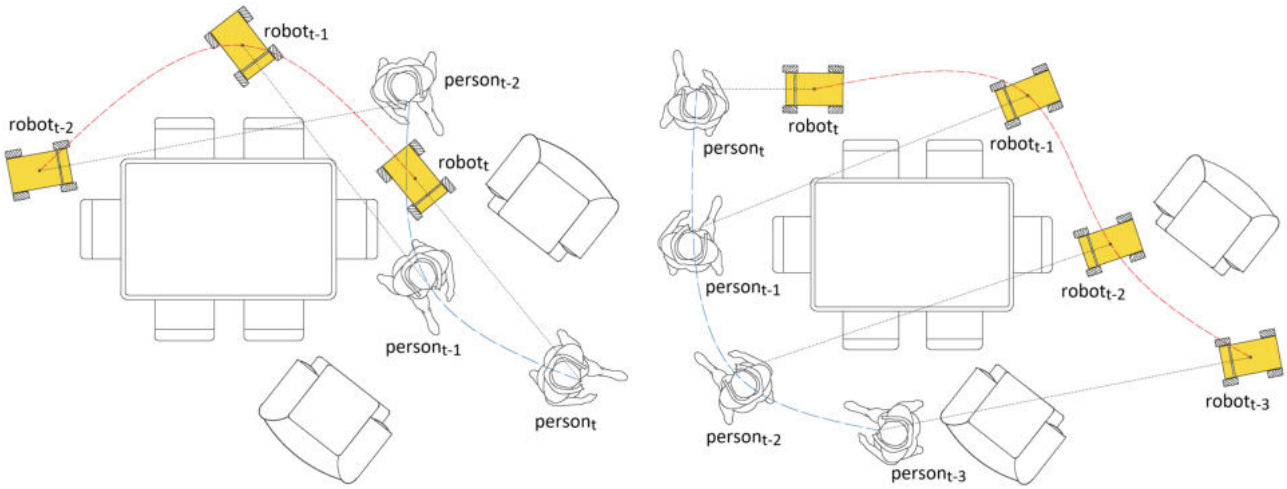


Figure 1. Visualization of the human-centered navigation service task for domestic robot assistance: the rover has to reach different goals while keep monitoring the user.



(**a**) (**b**)

Figure 2. Visualization of human-centered person-following task for domestic assistance: the omnidirectional capability allows the rover to follow the user maintaining its orientation towards them while avoiding obstacles. (**a**) The robot can follow the same path of the person while avoiding obstacles. (**b**) The robot must follow a different path keeping active the monitoring of the person.

We first set up a real-time perception pipeline to identify and track the person's pose (x_P, y_P). This position is exploited in the case of person-following, where it constitutes the dynamic goal of navigation. Differently, in the case of a goal-based navigation task, the goal is represented by the desired coordinates (x_G, y_G). A local planner generates a collision-free trajectory, handling the linear velocity commands v_x and v_y, while an additional module tunes the control of the angular yaw velocity ω, in order to constantly maintain the orientation towards the person (Figure 1).

The contribution of this work is threefold:

- We identify an omnidirectional motion planning approach as a robust, effective solution to boost the mobility of a robotic assistant during its principal navigation activities (person-following and goal-based navigation);
- We set up a real-time, cost-effective perception pipeline to extract the coordinate of the person and visually track its pose;
- We effectively integrate a navigation algorithmic stack that separately handles trajectory generation for obstacle avoidance and orientation control for person monitoring.

Nonetheless , compared to most previous works, we carried out extensive experimentation for both person-following and static goal navigation with the robot. To this end, we set up an innovative experimental framework based on an ultra-wideband (UWB) anchors system to localize both the person and the robot while moving and measure their relative distance and orientation. Our results validate the performance of our solution and show the competitive advantage and robustness it can provide in visually monitoring the user while avoiding obstacles in a cluttered indoor environment, such as a domestic one.

The article is organized as follows. In Section 2, we discuss related works presented in literature. In Section 3, we first introduce the human-centered navigation tasks, then we discuss the core methods of our solution, describing the perception and the omnidirectional navigation algorithms. In Section 4, the experimental settings and validation scenarios for both person-following and goal-based navigation are thoroughly presented, discussing the relevance of the obtained results. Section 5 conclude the article and propose possible future works.

2. Related Works

Similar works have been proposed in literature exclusively treating the person-following task. Ref. [11] propose a thorough categorization of recent person-following systems based

on five major features: the medium of operation, the choice of sensors, the mode of interaction, the granularity, and the degree of autonomy. However, very few of them propose a complete framework for person-following, devoting a great focus to both person identification and tracking, and effective navigation planning for obstacle avoidance. A more detailed overview of the most related following navigation works in literature is proposed in Section 2.2.

The omnidirectional service platform adopted for this study has been recently presented in [13], where details related to the design and the capabilities of the platform can be found. Nonetheless, the human-centered navigation stack proposed in this work can be fully replicated and applied with a generic omnidirectional platform.

2.1. Person Identification and Tracking

Most studies concentrated only on the person identification and tracking problem with different sensor strategies. Identification systems often aim at recognizing a person from legs patterns in the collection range of a LiDAR laser scan [14–16], or Time-Of-Flight (TOF) 3D points clouds [17,18]. However, in the last decade, computer-vision systems have been the preferred solution, demonstrating to be the most efficient, reliable, and cost-effective. Deep Neural Networks (DNN) have largely demonstrated to be a meaningful answer to a wide variety of visual perception tasks, such as real-time object and person detection [19], semantic segmentation [20], or pose estimation [21]. Some works proposed a quite complex visual tracking system, which has been more recently replaced by simpler tracking algorithms based on the Kalman filter [22]. The perception system used by the recent work [23] combined OpenPose and Kalman filter to identify and track the person using a monocamera. Moreover, [23–25] aim at recognizing or re-identifying a specific person. Alternative approaches use sensor fusion, for example, to mix images with ultrasonar data for 3D person tracking [26], while [27] exploits the gait to recognize the user. Another solution implemented to keep track of the user during navigation consists in the adoption of an omnidirectional camera for a 360° Field-Of-View (FOV), [28] or also a rotating camera, such as gimbal systems, typical of Unmanned Aerial Vehicles (UAVs) following [29].

2.2. Navigation and Obstacle Avoidance

Person-following systems are often based on naive visual-control strategy, directly coupling the generation of heuristic commands for the robot with the person coordinate in the image [30,31]. A simple PID (proportional–integral–derivative) controller is alternatively used by [32] with the assumption that replicating the estimated trajectory of the target person can lead the robot to no collisions. However, this simple idea is not reliable when the robot must deal with challenging environments like domestic ones. Narrow passages and obstacles of diverse shapes can occlude the sensor's field of view (FOV) and forbid the identification of the person or, diversely, the robot may have to choose between losing the tracking of the target and avoiding a collision. [33] proposes an obstacle avoidance system devoted to person-following with a dynamic window approach, although using only 2D LiDAR points for both person detection and navigation could easily lead to target loss due to obstacles occlusion. Thus, obstacle avoidance and person tracking are often conflicting objectives and are not usually jointly tackled in literature. Indeed, the integration of a suitable trajectory planner with person tracking is often neglected. To this end, omnidirectional platforms can provide significant advantages. To our knowledge, few attempts have been carried out to use omnidirectional motion planning for person-following, poorly presenting and correlating methodology and experiments [34,35].

3. Human-Centered Autonomous Navigation

We define human-centered navigation as the service robotic task of autonomously navigating within a domestic environment while maintaining constant track of the subject of interest. On this basis, we propose a novel system to handle human-centered autonomous

navigation in cluttered and unstructured environments, using an omnidirectional robotic platform (Figure 3). We define two different use cases: in the first, the rover has to move towards a series of specific destinations, keeping visual contact with the user during the whole operation. In the second, the rover performs a person-following task where the position of the subject, extracted from the perception system, is used as a dynamic goal for navigation. According to this concept, the autonomous platform should always be aware of the subject's position during its navigation, which means keeping its orientation towards the person and maintaining them in the camera's field of view.

Figure 3. The omnidirectional platform we set up for experimentation and validation of our novel methodology. The vertical shaft allows the camera to be elevated over most indoor environment obstacles.

3.1. Perception and Tracking

In this work, we developed a deep learning perception pipeline that allows the robot to visually track the person. The scheme presented in Figure 4 describes the complete perception pipeline used to extract, at each time instant, the coordinate of the person in the robot reference frame from RGB-D images. A RealSense D435i Depth Camera, mounted on the rover at a human height, is used to collect color images of the environment. In a first step, the person's presence is detected through PoseNet [36], a lightweight deep neural network that estimates the pose of humans in images and videos. For each person present in the scene, the network outputs the position of 17 key joints (such as elbows, shoulders, or feet). In our implementation, PoseNet runs on the Google Coral Edge TPU device https://coral.ai (accessed on 11 July 2020) at 30 frame-per-second (FPS), which corresponds to the maximum frame rate supported by the RealSense D435i camera. The key-points predicted by PoseNet are then translated into a bounding box that localizes the person within the image. The resulting bounding box is tracked with SORT [37], a very simple online and real-time tracking algorithm based on the Kalman filter. SORT also keeps track of the subject when they leave the frame for a few moments, and associates an ID to each person in the image. This ID is maintained as long as the person does not leave the frame for several time instants.

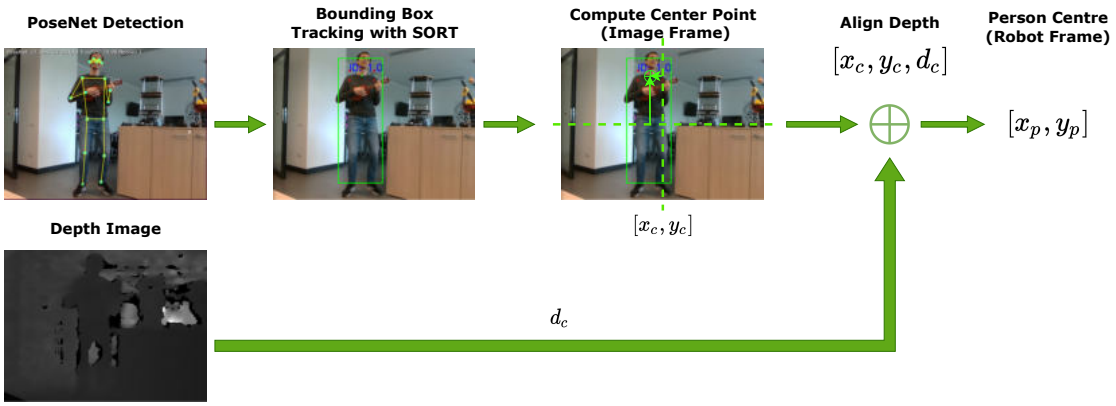


Figure 4. Real-time visual perception pipeline for person identification, tracking, and coordinate extraction. $[x_C, y_C]$ are the coordinate of the person's center in the image frame, while $[x_p, y_p]$ indicates it in the robot reference frame. The person's pose is continuously estimated by PoseNet at 30 fps and tracked with SORT, then a set of reliable pose key-points are used to extract the center coordinates.

At this point, a depth image extracted from the RealSense camera, aligned with the RGB image, can be used to extrapolate the relative position of the detected individual in the robot reference frame (x_p, y_p). To do so, it is necessary to identify a precise area, or better, a specific point of the image where we can confidently expect to find the person. For this purpose, the output key-joints of PoseNet represent particularly suitable information: in comparison, a conventional person detection approach can only localize the person in an approximate bounding box area. This information is inadequate for the person position tracking task, since the bounding box contains points associated with the person and points belonging to the background. The risk is that the system could treat a point of the image belonging to the background as a point belonging to the human body, causing an error in the correct evaluation of the subject position. A set of particularly reliable key points of the estimated pose is selected to find the person's center point C on the color image. When the neural network identifies both the shoulders of a person with high confidence, the point C is selected as the average of these joints. If shoulders are not recognized, but the hips are, then the selected point becomes the one between the two hips. If neither shoulders nor hips are recognized with a certain degree of confidence, then the detection of the person is considered invalid. This structure guarantees reliable esteem of the person's position in the environment to be fully usable by the robot navigation system, avoiding the risk of misleading target estimates and, consequently, inaccurate motion planning. The distance of the person from the robot d_C is then extracted from the depth frame as the value corresponding to the point C. At each time instant, the complete information contained in the resulting array (x_C, y_C, d_C) is translated into the person's position in the robot's reference frame (x_P, y_P), with basic reference frame transformations. This position will be used by the navigation control stack described in the following section.

As an interesting point of discussion, we found that the detection of people present in the image could not be sufficient to efficiently track a specific human subject. In particular, two well-known problems could arise:

- Especially in crowded environments, where multiple people are present in every frame, the subject could be mistaken for another person in the image (or vice-versa);
- Without a component capable of tracking observations at previous time instants, it could be very difficult to guarantee real-time performances if the detection of the subject is lost for a few consecutive frames. This problem can be particularly critical in all those situations with an occluded view of the subject due to obstacles or other people present in the scene.

Although a person re-identification algorithm could mitigate the first problem allowing to recognize a specific person, at the cost of additional computation, dealing with the

second can be much more arduous without a proper component specifically designed for tracking. Aiming to solve both the problems at the most convenient computational cost, we decided to adopt SORT in the person detection pipeline to exploit also future estimates of the person pose and allow the rover to keep tracking the desired subject as long as necessary, discriminating from other people in the scene. Nonetheless, a Re-Identification neural network could be easily integrated as the first stage of our pipeline if strictly requested by the particular case study.

3.2. Omnidirectional Motion Planner and Obstacle Avoidance

Typically, a navigation system requires some fundamental components. The first necessity is to localize the robot in the operating environment. In order to compute the trajectory towards a goal, the system needs to acquire the pose (position and orientation) of the rover with respect to a fixed reference frame. This piece of information needs to be retrieved with a certain frequency to ensure real-time performances, since we need to maintain track of the position of the rover during time as it moves towards a different location. Obviously, the required frequency of localization acquisition increases as the speeds assumed by the platform grow. In our implementation, we exploited a RealSense T265 Tracking Camera to obtain information about the rover's pose. This camera employs state-of-the-art Visual Inertial Odometry (VIO) algorithms, which use visual information to provide odometry data at a frequency of roughly 200 fps, more than sufficient for any indoor autonomous platform.

Second, we need a path planner. If a map of the operative scenario is provided from the beginning, it is possible to compute an optimal trajectory knowing a priori the location of each obstacle (global planner). However, in the majority of service robotics navigation cases, a global planner is not sufficient, since a map is not always available. Moreover, if the map is also given, real-life domestic environments are highly dynamic environments, where obstacles' position could be changed over time (chairs, bins) or they can move on their own (people, animals, other autonomous platforms). In these cases, a real-time perception system together with a local planner is necessary to dynamically re-plan the upcoming commands on the base of the last perceived data, and perform an effective obstacle avoidance. The visual perception pipeline described in Section 3.1 is uniquely used to extract the coordinates of the person (x_P, y_P) in the scene. Differently, we use an RPLiDAR A1 to retrieve 2D laser scan distance measurements of the obstacles around the robot at each time instant, which are subsequently used to feed a local path planner.

Information regarding the rover's and obstacles' position is passed to the navigation system, which we developed tailoring the Navigation2 navigation stack (https://navigation.ros.org/ (accessed on 13 July 2022)) for the specific use case of assistance and person monitoring. Nav2 is a highly modular navigation system based on behavior trees, which allows integration with custom plugins adapted for any specific application. It provides default modules for converting laser scan data into cost-map representation, planning a path towards a goal, and controlling the rover along it. Although Nav2 is a very complete system for a conventional navigation application, we needed to modify it extensively to customize the overall algorithmic stack to handle both person-following and goal-based navigation with a unique solution for person-monitoring, integrating new plugins and behavior tree entries.

Since domestic scenarios fall within unstructured environments, for which a map is rarely provided, we decided to focus on a local planner. This option allows the rover to be deployed in unknown scenarios without the need for preliminary information, since the system plans its navigation paths depending only on real-time spatial data deriving from the LiDAR sensor. The resulting navigation system consists of a DWB local planner and controller, able to generate an obstacle-free trajectory towards the goal and drive the rover along it. To detach the control of linear and angular velocities, we decided to forbid the DWB to include the yaw velocity in the dynamic path planning, forcing it, instead, to plan a safe trajectory and control the rover using only the two linear velocities

$[v_x, v_y]$, along x and y axes of the horizontal plane. The goal of the navigation task (x_G, y_G) coincides with the person's position (x_P, y_P) in the specific case of the person-following, diversely it is a separate target point to be reached while monitoring the person in the service navigation scenario.

3.3. Person-Focused Orientation Control

The angular velocity ω is provided by another system node, which at any instant computes the angular difference $\Delta\theta$ between the orientation of the rover and the orientation of the vector connecting the rover's center of rotation with the person position, retrieved from the perception module:

$$\Delta\theta = \arctan(y_P, x_P)$$

The exact yaw velocity is then calculated as follow:

$$\omega = sign(\Delta\theta) \cdot min(\|k \cdot \Delta\theta\|, \omega_{max})$$

where

k is a parameter used to linearly increase ω as $\Delta\theta$ grows

ω_{max} is another parameter used to limit the maximum value assumed by ω

After some tests on our indoor application, we found optimal values of these parameters, respectively, at 1.3 and 1.5 rad/s, but they can be changed depending on the specific operating scenario.

Figure 5 resumes the complete proposed human-centered navigation system. The upper blue section of the scheme contains the extraction of the person position (x_P, y_P) in the robot reference frame through the visual perception pipeline (presented in Section 3.1). The yaw controller then processes this position to obtain the angular velocity command ω needed to keep the platform oriented towards the person. On the lower red section of the scheme, the DWB local planner receives the LiDAR range points and the goal coordinate (x_G, y_G) to produce a collision-free trajectory and provide linear velocities $[v_x, v_y]$. The full velocity command for the robot is, therefore, obtained by combining linear and angular velocities in the vector $[v_x, v_y, \omega]$. Obviously, the view of the subject can be occluded by physical obstacles, but if the RGB camera is mounted on the robot at a height greater than most objects in the operating environment (such as tables, chairs, sofas, desks), as we did on our platform, the rover can navigate through cluttered spaces still maintaining its sight centered on the user.

This intuition was initially conceived for an autonomous indoor assistant, addressed to elderly or disabled users who need constant monitoring, even when the platform needs to move to another place of the room to carry out a different task, but can be adapted to many different applications, for example, in all the situations in which the platform has to perform a specific operation while constantly focusing on another human operator, for monitoring purposes or to receive new instruction through visual inputs. This application can be particularly useful in the fight against COVID-19 and future pandemics for assisting patients in hospitals and their houses. The rover can replace medical personnel's intervention, greatly reducing the risk of contagion and spread of the virus, continue to monitor the patient, and eventually request human help in case of abnormal situations.

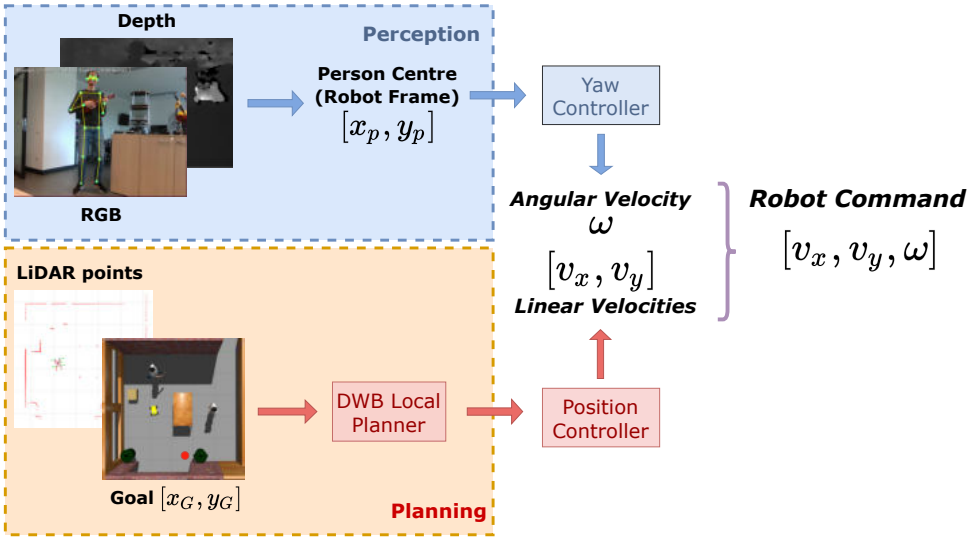


Figure 5. Human-centered navigation methodology pipeline scheme. Linear and angular velocity $[v_x, v_y, \omega]$ are generated separately to successfully carry out obstacle avoidance through local trajectory planning together with person monitoring through yaw control.

4. Experiments and Results

For our experimentation, we used a cheap omnidirectional robotic platform with four mecanum wheels, presented in [13]. The whole software system is executed on a single Intel NUC11TNHv5 PC, directly integrated within the rover. As stated before, the platform mounts an RPLiDAR A1 sensor, a RealSense D435i camera for person detection, and a RealSense T265 camera for visual odometry. Overall, the platform presents a very basic configuration, easily replicable with simple commercial components on a generic omnidirectional platform. In this sense, our solution is cost-effective, avoiding the necessity of more complex and expensive sensors and systems for person tracking, such as active gimbals or 360-degree cameras. Furthermore, the software system is lightweight enough to run on integrated hardware at the edge and reach real-time performances.

All the software components and technologies needed to perceive and navigate the environment have to be merged into a single organic system, in order to fulfill the different tasks. The most widespread solution in literature requires using a Middleware [38], an abstraction layer that resides between the operating system and software applications. In this work, we decided to adopt the Robot Operating System 2 (ROS2) https://docs.ros.org/en/foxy/index.html (accessed on 18 July 2022), due to the variety of compatible algorithms and the very active community supporting it. It provides several advantages and improvements compared to the original ROS https://www.ros.org/ (accessed on 28 July 2022) since it is more suitable for real-time systems and it has access to more advanced applications [39]. ROS2 is based on a Data Distribution Service (DDS) structure, with nodes that publish and subscribe to different topics.

Two different kinds of experiments are conducted:

1. The first experimental stage aims at demonstrating the efficiency of the person-centered navigation task for monitoring purposes, where the rover has to navigate from a point A to a target point B of coordinate (x_G, y_G), maintaining its focus on the subject located in (x_P, y_P);
2. The second series of experiments take into consideration the person-following task, where (x_G, y_G) and (x_P, y_P) coincide and represent the dynamic goal obtained from the visual perception pipeline, which identifies and tracks the person of interest.

For these tests, the system has been integrated with additional functionalities to refine the platform's behavior and further increase the person's awareness during the navigation.

Safety distance module During the rover operation, the user's safety should always be ensured, even if this leads to the failure of the requested task. For this reason, a module able to truncate the navigation path of the rover is inserted, which guarantees a minimum distance of one meter always to be maintained from any person.

Recovery policy for person tracking During the navigation towards a specified goal, the rover may lose track of the person. In case the track is not resumed within a certain time interval, a specific module we added sends a command to the rover to interrupt the navigation and to start rotating towards the direction the person was last perceived in an attempt to regain visual contact with the user.

Recovery policy for person-following The same problem described above can occur during the person-following task but, in this case, consequences could be even worse since the knowledge of the person's position affects not only the yaw but also the linear directions of the navigation. To re-establish track with the person, first of all the rover heads towards the last known position of the user, maintaining its orientation towards that location. This decision compensate for all those cases in which the person takes a turn behind an obstacle, such as a wall, and simply moving towards the corner of the curve where the user was last seen is enough to regain visual contact. If this should reveal not sufficient, once the robot has reached the last known position, it starts rotating as described before.

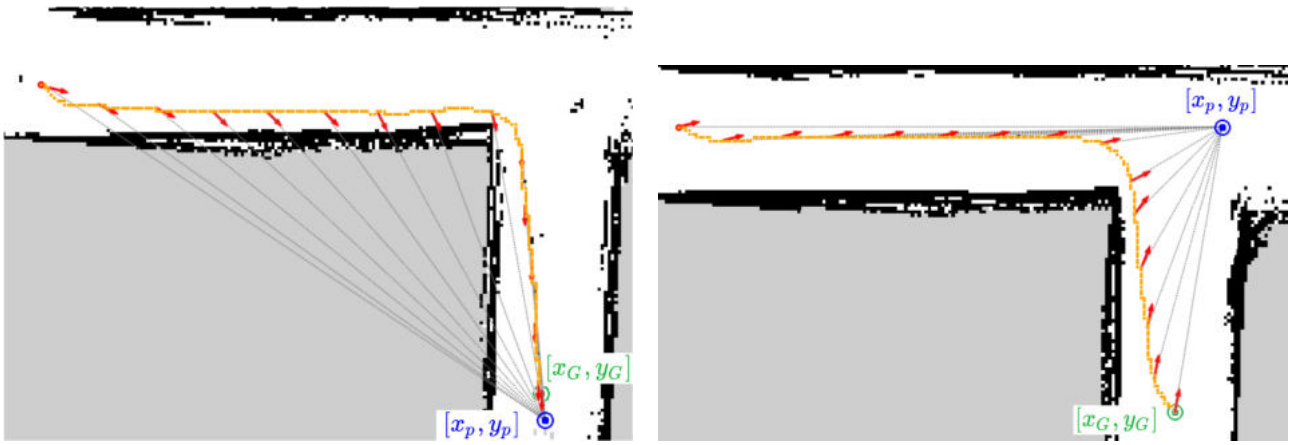
For each tested scenario, tests are performed with the same omnidirectional rover in two different configurations. In the first configuration, the rover adopts our novel navigation methodology: it plans collision-free trajectories fully exploiting its omnidirectional kinematics, combining both the two linear velocities $[v_x, v_y]$. The angular yaw velocity ω is controlled by the person tracking module to always maintain visual contact with the followed person. In the second configuration, the rover behaves like a differential platform. This means it can only exploit velocity v_x, while control of velocity v_y is denied, and the angular yaw velocity ω is solely dedicated to navigation purposes. This procedure allows the comparison of performances between our solution and a generic differential platform in tracking the user.

4.1. Person-Centered Navigation

Tests are performed in two different scenarios, depicting a 90° hallway characterized by low walls, which represent any potential obstacle present in a realistic domestic scene. The rover camera can see over walls, but the platform is forced to avoid them in order to reach its goal. The starting point and the destination (x_G, y_G) are the same in the two cases. What changes is the position of the person (x_P, y_P): near the destination point in the first scenario (Figure 6a), and in the corner of the hallway in the second (Figure 6b). In these preliminary trials, the person maintains their position during the whole extent of the test. The rover odometry data are acquired with a frequency of 5 Hz.

Seven tests are performed for each scenario and both configurations, omnidirectional and differential. The error term is represented by the angular difference $\Delta\theta$ between the orientation of the rover and the orientation of the vector connecting the rover's center of rotation with the person's position. The horizontal FOV of the RealSense D435i (RGB stream) is equal to 69°. The angular difference $\Delta\theta$ should never be higher than half this angle, approximately 34.5°, to constantly keep track of the person's position.

Considered metrics for each test are the average angular difference $\Delta\theta$ with its standard deviation, the root mean square error (RMSE), and the mean absolute error (MAE) maintained along the whole path, considering $\Delta\theta = 0$ as the optimal value. In Table 1 are reported, for each scenario and each metric, the average value computed over all the different tests, and the percentage of improvement introduced by our methodology.



(**a**) (**b**)

Figure 6. Omnidirectional person-centered navigation results in two scenarios with the person in different positions: in (**a**) the person is close to the navigation goal, in (**b**) the person is at the corner of the hallway. Red arrows indicate position and orientation of the rover at different time instants, the blue point is the person's position, while the orange spline represents the path crossed by the rover.

Table 1. Results obtained from the person-centered navigation test are expressed in terms of mean angular difference $\Delta\theta$, its standard deviation, root mean square error (RMSE), and mean absolute error (MAE) considering $\Delta\theta = 0$ as the optimal value. The person is located close to the destination point in the first scenario (Figure 6a) and in the corner of the hallway in the second (Figure 6b). Contrary to the differential configuration, omnidirectional motion drastically reduce the maximum error $\Delta\theta$ between the orientation of the rover and the person during the navigation.

$\Delta\theta$ **Error**	**Mean**	**Std.Dev.**	**RMSE**	**MAE**
First Scenario				
Omnidir.	−2.88	4.63	5.47	4.32
Differential	−32.75	28.71	43.55	33.94
Improvement	**91.21%**	**83.87%**	**87.44%**	**87.27%**
Second Scenario				
Omnidir.	−2.23	3.98	4.58	2.51
Differential	−75.08	79.88	109.62	75.08
Improvement	**97.03%**	**95.02%**	**95.82%**	**96.66%**

As seen from the results and Figure 6, the omnidirectional system is able to efficiently navigate towards the goal, constantly maintaining its orientation towards the person. The $\Delta\theta$ angular error is kept at extremely low average values equal to −2.88 and −2.23, respectively, in the two scenarios. Furthermore, the maximum recorded value of $\Delta\theta$ does not exceed 17°, which is well below the limit of 34° imposed by the camera's FOV. This means the system can keep tracking the person for the whole extent of the navigation. Moreover, from data collected during the experimentation, the perception and tracking system described at Section 3.1 was able to correctly recognize and localize the followed person within the environment on average 29 times per second. On the other hand, velocity commands are provided with frequencies over 15 fps at any time.

For comparison, we also added the results obtained with the differential drive configuration. However, this comparison is uneven: as explained before, a differential platform has to choose whether to navigate towards the goal or remain orientated towards the person. This is particularly evident in the second scenario, where the person and the goal have two completely different positions.

4.2. Person Following

For the person-following task, tests are performed in four different scenarios. The geometric configuration can be seen in Figure 7. Similar to the previous test stage, obstacles are constituted by low walls, except for the fourth, where they are full-height walls. Contrary to the previous case, the person to be followed moves for the whole extent of the test. The rover has to follow the person, using the position (x_P, y_P) extracted from the visual perception pipeline as a dynamic goal of the navigation. For this reason, to ensure an accurate ground truth data collection, we set up a localization system based on four ultra-wideband anchors placed in the testing area. One additional anchor is placed upon the rover, and the followed person holds a second one. The rover's orientation is also aligned with the one used by the ultra-wideband system. In such a way, it is possible to know the actual relative position between rover and followed person. This allows us to correctly compute the angular difference $\Delta\theta$ at any time instant. To our knowledge, this experimental setting is the first attempt in the literature to quantitatively measure a person's quality following system performance, going beyond the typical qualitative evaluation.

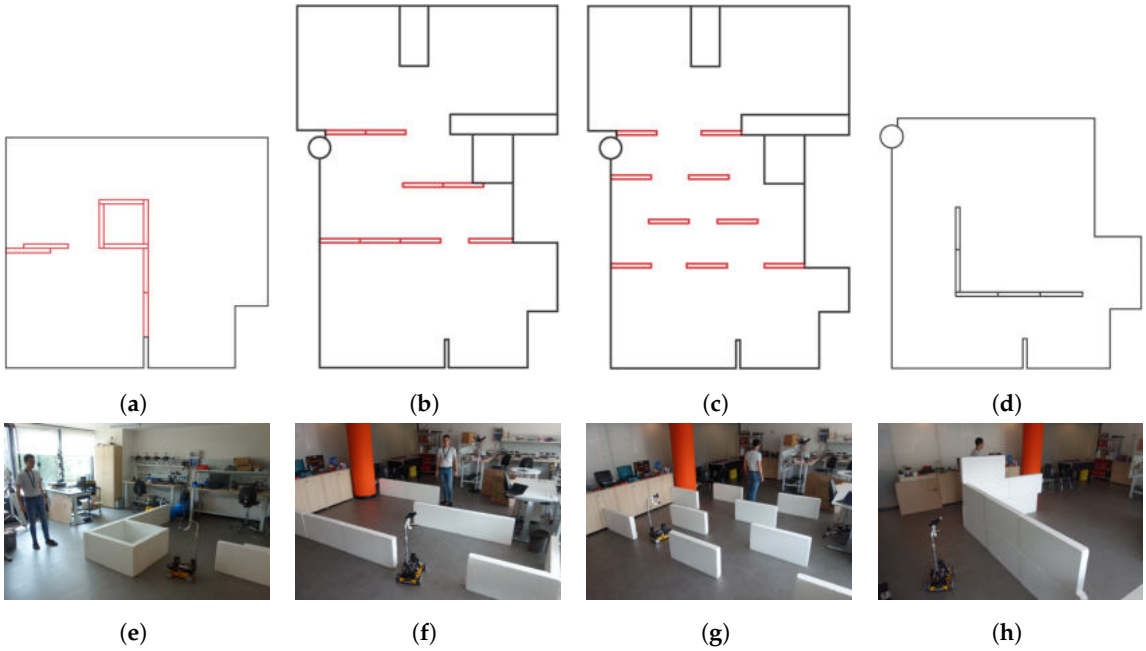


Figure 7. Qualitative visualization of the four scenarios set up for the person-following test. In the upper row, a schematic representation is shown, where red objects represent low-height obstacles over which the robot's camera can see. In the lower row, the real testing area with the robot is shown.

As already completed for the first test, seven validation runs are performed for each rover configuration in every scenario. The same error term $\Delta\theta$ and metrics discussed in the previous section are used to evaluate the person-following performances. Results can be consulted in Table 2. Furthermore, in Figures 8 and 9, for each scenario and each configuration, a visualization of the performed test is reported. The gridmaps reported in the figure are directly obtained from the rover during the navigation, while rover and person poses are obtained from the ultra-wideband system. As can be seen, our methodology proves to robustly track the followed person more effectively than a traditional differential drive navigation in all the considered scenarios. In the omnidirectional configuration (Figures 8a,c and 9a,c), the rover manages to always maintain the user within the camera's view, contrary to the differential drive case, where the visual contact is instead lost several times. This generally leads to higher performance in following the user, with the rover planning more optimal collision-free trajectories, fully satisfying also the person monitoring requirement. The obtained values of $\Delta\theta$ clearly show the performance gap in all scenarios, demonstrating the successful behavior in monitoring the person provided by our solution.

Additionally, in the fourth scenario (Figure 9c,d), where after the curve the wall obstructs the rover's view of the user, it appears clear that the ability to remain facing the human dynamic goal allows for a more accurate re-acquisition of tracking as soon as the obstacle is passed. In this last scenario, the differential drive system registers the highest orientation error, with a substantial $\Delta\theta$ average gap from our solution.

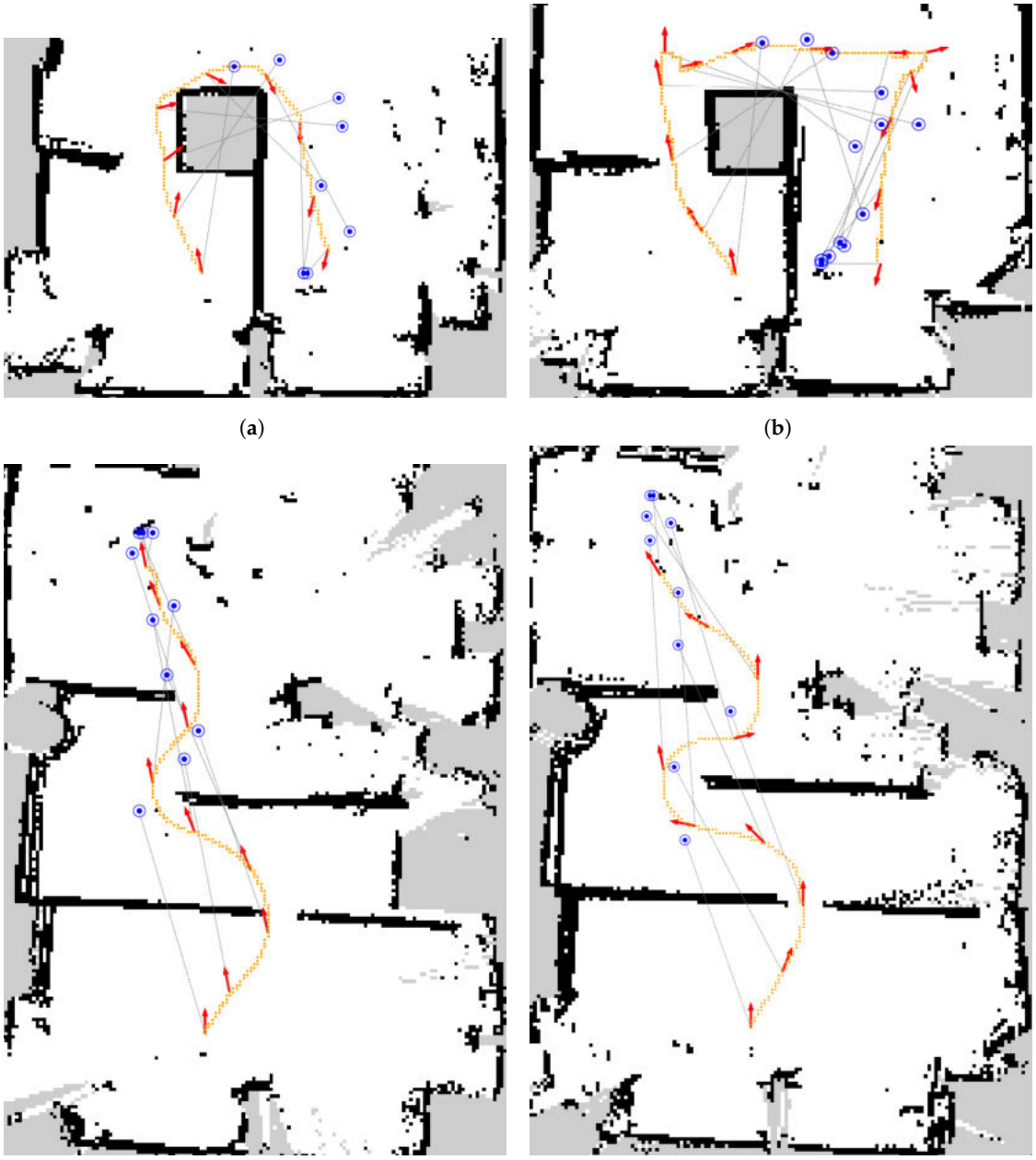
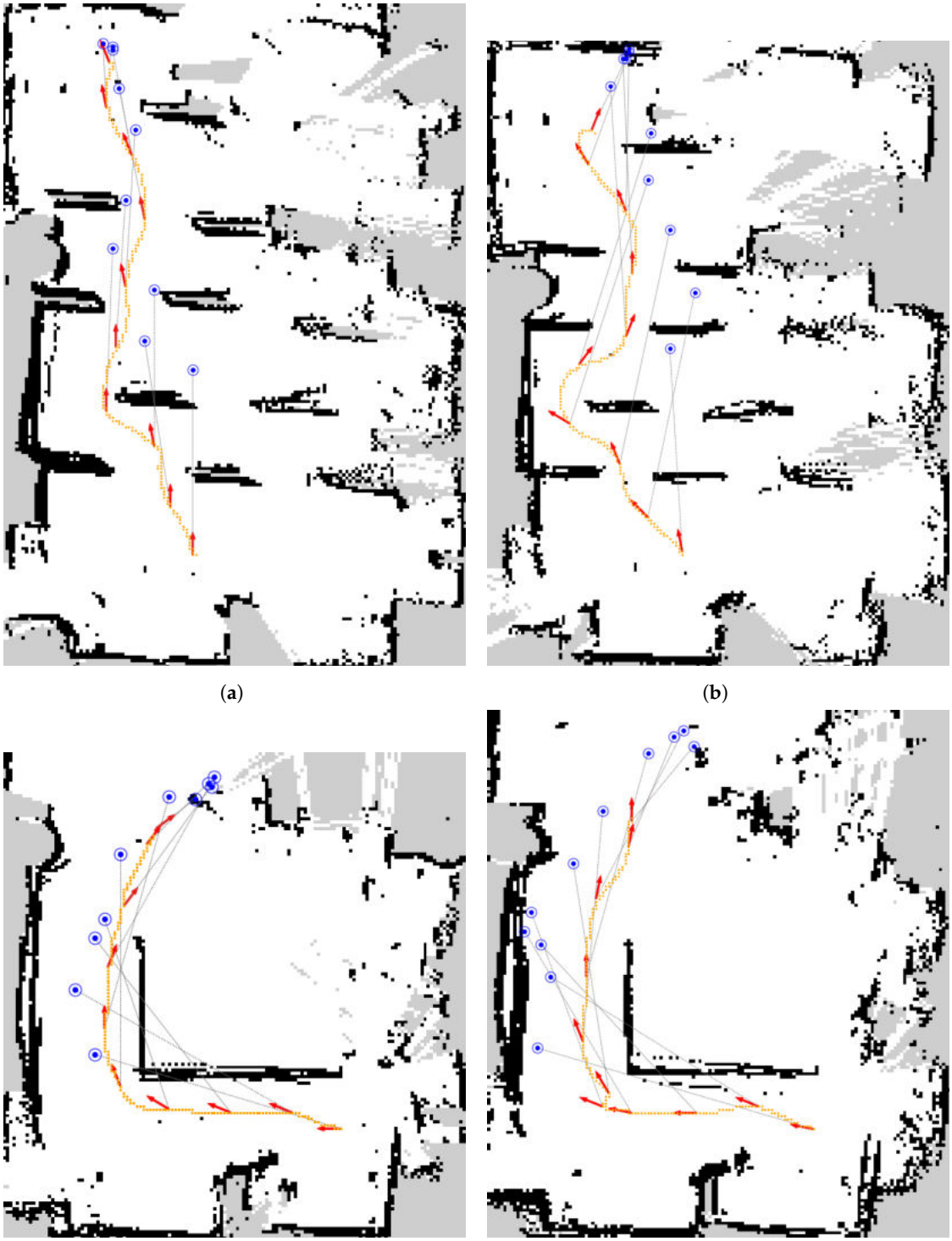


Figure 8. Person-following results in the first two scenarios: scenario 1 is composed of a wide U-shaped path, while scenario 2 presents narrow passages through obstacles. Red arrows indicate position and orientation of the rover associated with the person's position (blue point) at the same instant. The orange spline represents the path crossed by the rover. (**a**) Scenario 1—Omnidirectional configuration; (**b**) Scenario 1—Differential configuration; (**c**) Scenario 2—Omnidirectional configuration; (**d**) Scenario 2—Differential configuration.



(**a**) (**b**)

(**c**) (**d**)

Figure 9. Person-following results in the third and fourth scenario: scenario 3 presents a high number of obstacles and possible paths, while scenario 4 is composed of a high 90° wall to be circumnavigated. Red arrows indicate position and orientation of the rover associated with the person's position (blue point) at the same instant. The orange spline represents the path crossed by the rover. (**a**) Scenario 3—Omnidirectional configuration; (**b**) Scenario 3—Differential configuration; (**c**) Scenario 4—Omnidirectional configuration; (**d**) Scenario 4—Differential configuration.

Table 2. Results obtained from the person-following test in four different scenarios are expressed in terms of mean angular difference $\Delta\theta$, its standard deviation, root mean square error (RMSE), and mean absolute error (MAE) considering $\Delta\theta = 0$ as the optimal value. Our omnidirectional planning and control system clearly demonstrates a performance gap in keeping the tracking of the person while following its motion: the $\Delta\theta$ error is drastically reduced in comparison with a differential drive navigation.

$\Delta\theta$ **Error**	**Mean**	**Std.Dev.**	**RMSE**	**MAE**
Scenario A				
Omnidir.	2.99	10.54	11.00	8.98
Differential	16.00	63.41	68.31	57.20
Improvement	**81.31%**	**83.38%**	**83.90%**	**84.30%**
Scenario B				
Omnidir.	−4.09	8.75	9.93	8.19
Differential	−15.67	53.99	58.48	50.11
Improvement	**73.90%**	**83.79%**	**83.02%**	**83.66%**
Scenario C				
Omnidir.	0.31	8.28	8.81	6.93
Differential	12.34	42.19	45.05	37.38
Improvement	**97.49%**	**80.37%**	**80.44%**	**81.46%**
Scenario D				
Omnidir.	4.46	11.84	12.84	10.10
Differential	27.66	20.95	35.07	29.19
Improvement	**83.88%**	**43.48%**	**63.39%**	**65.40%**

5. Conclusions and Future Works

In this work, we propose a novel, cost-effective approach for human-centered autonomous navigation in the context of domestic robotic assistance. In particular, we devote a great focus on developing a robust solution to visually monitor the user in two different case studies, which we consider the most relevant and common for a robot assistant: person monitoring during navigation to a target goal and person-following. Differently from previous works, the core of our robot assistive solution relies on the idea that keeping the platform oriented towards the subject permits us to continuously check their status, also when the robot is moving and avoiding obstacles typically present in a realistic indoor environment. To this end, we first set up a real-time visual perception pipeline that reliably provides the coordinate of the person in the robot reference frame using a cheap RGB-D camera. Then, adopting a generic omnidirectional platform, we propose a navigation system that separately treats orientation control and dynamic trajectory planning to fulfill both the monitoring and the obstacle avoidance objectives of the robotic assistive task. Our extensive experimentation conducted for both the considered use cases in realistic settings demonstrates the competitive advantages and the robustness of our solution compared to a common differential drive navigation. Moreover, it also advances the typical experimental framework for person-following, quantitatively evaluating the physical tracking of the person with an ultra-wideband localization system. To our knowledge, this is the first study to investigate omnidirectional capability of a robotic platform to enable true human-centered navigation, where the care and attention for the user's health are considered the main focus of the navigation task. Future works may investigate the integration of a

person re-identification deep neural network in the visual perception pipeline to recognize a specific user, which will contribute significantly to a real application.

Author Contributions: Conceptualization, A.E. and M.M.; methodology, A.E. and M.M.; software, A.E. and M.M.; validation, A.E. and M.M.; formal analysis, A.E. and M.M.; investigation, A.E. and M.M.; data curation, A.E.; writing—original draft preparation, A.E. and M.M.; writing—review and editing, A.E. and M.M.; visualization, A.E. and M.M.; supervision, M.C.; project administration, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The work presented in this paper has born from the collaboration between the PIC4SeR Centre for Service Robotics at Politecnico di Torino and Edison S.p.A. In particular, we sincerely thank Riccardo Silvestri and Stefano Ginocchio, as well as the entire team from Officine Edison Milano.

Conflicts of Interest: The authors declare that they have no know competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Martinez-Martin, E.; del Pobil, A.P. Personal robot assistants for elderly care: An overview. In *Personal Assistants: Emerging Computational Technologies*; Springer: Cham, Switzerland, 2018; pp. 77–91.
2. Vercelli, A.; Rainero, I.; Ciferri, L.; Boido, M.; Pirri, F. Robots in elderly care. *Digit.-Sci. J. Digit. Cult.* **2018**, *2*, 37–50.
3. United-Nations. *Shifting Demographics*; United-Nations: New York, NY, USA, 2019.
4. Novak, L.L.; Sebastian, J.G.; Lustig, T.A. The World Has Changed: Emerging Challenges for Health Care Research to Reduce Social Isolation and Loneliness Related to COVID-19. *NAM Perspect.* **2020**, *2020*. https://doi.org/10.31478/202009b.
5. Shen, Y.; Guo, D.; Long, F.; Mateos, L.A.; Ding, H.; Xiu, Z.; Hellman, R.B.; King, A.; Chen, S.; Zhang, C.; et al. Robots under COVID-19 pandemic: A comprehensive survey. *IEEE Access* **2020**, *9*, 1590–1615.
6. Abdi, J.; Al-Hindawi, A.; Ng, T.; Vizcaychipi, M.P. Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open* **2018**, *8*, e018815.
7. Góngora Alonso, S.; Hamrioui, S.; de la Torre Díez, I.; Motta Cruz, E.; López-Coronado, M.; Franco, M. Social robots for people with aging and dementia: A systematic review of literature. *Telemed. E-Health* **2019**, *25*, 533–540.
8. Gasteiger, N.; Loveys, K.; Law, M.; Broadbent, E. Friends from the Future: A Scoping Review of Research into Robots and Computer Agents to Combat Loneliness in Older People. *Clin. Interv. Aging* **2021**, *16*, 941.
9. Yatsuda, A.; Haramaki, T.; Nishino, H. A Study on Robot Motions Inducing Awareness for Elderly Care. In Proceedings of the 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Taichung, Taiwan, 19–21 May 2018; pp. 1–2. https://doi.org/10.1109/ICCE-China.2018.8448739.
10. Möller, R.; Furnari, A.; Battiato, S.; Härmä, A.; Farinella, G.M. A survey on human-aware robot navigation. *Robot. Auton. Syst.* **2021**, *145*, 103837.
11. Islam, M.J.; Hong, J.; Sattar, J. Person-following by autonomous robots: A categorical overview. *Int. J. Robot. Res.* **2019**, *38*, 1581–1618.
12. Honig, S.S.; Oron-Gilad, T.; Zaichyk, H.; Sarne-Fleischmann, V.; Olatunji, S.; Edan, Y. Toward socially aware person-following robots. *IEEE Trans. Cogn. Dev. Syst.* **2018**, *10*, 936–954.
13. Eirale, A.; Martini, M.; Tagliavini, L.; Gandini, D.; Chiaberge, M.; Quaglia, G. Marvin: An Innovative Omni-Directional Robotic Assistant for Domestic Environments. *Sensors* **2022**, *22*, 5261.
14. Jia, D.; Hermans, A.; Leibe, B. DR-SPAAM: A Spatial-Attention and Auto-regressive Model for Person Detection in 2D Range Data. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; pp. 10270–10277. https://doi.org/10.1109/IROS45743.2020.9341689.
15. Cha, D.; Chung, W. Human-Leg Detection in 3D Feature Space for a Person-Following Mobile Robot Using 2D LiDARs. *Int. J. Precis. Eng. Manuf.* **2020**, *21*, 1299–1307. https://doi.org/10.1007/s12541-020-00343-7.
16. Guerrero-Higueras, Á.M.; Álvarez-Aparicio, C.; Olivera, M.C.; Rodríguez-Lera, F.J.; Fernández-Llamas, C.; Rico, F.M.; Matellán, V. Tracking People in a Mobile Robot From 2D LIDAR Scans Using Full Convolutional Neural Networks for Security in Cluttered Environments. *Front. Neurorobotics* **2019**, *12*, 85. https://doi.org/10.3389/fnbot.2018.00085.
17. Wang, W.; Liu, P.; Ying, R.; Wang, J.; Qian, J.; Jia, J.; Gao, J. A High-Computational Efficiency Human Detection and Flow Estimation Method Based on TOF Measurements. *Sensors* **2019**, *19*, 729. https://doi.org/10.3390/s19030729.

18. Zoghlami, F.; Sen, O.K.; Heinrich, H.; Schneider, G.; Ercelik, E.; Knoll, A.; Villmann, T. ToF/Radar early feature-based fusion system for human detection and tracking. In Proceedings of the 2021 22nd IEEE International Conference on Industrial Technology (ICIT), Valencia, Spain, 10–12 March 2021; Volume 1, pp. 942–949. https://doi.org/10.1109/ICIT46573.2021.9453703.
19. Zhao, Z.Q. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232.
20. Zhang, X.; Chen, Z.; Wu, Q.J.; Cai, L.; Lu, D.; Li, X. Fast semantic segmentation for scene perception. *IEEE Trans. Ind. Inf.* **2018**, *15*, 1183–1192.
21. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 172–186.
22. Gupta, M.; Kumar, S.; Behera, L.; Subramanian, V.K. A novel vision-based tracking algorithm for a human-following mobile robot. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *47*, 1415–1427.
23. Koide, K.; Miura, J.; Menegatti, E. Monocular person tracking and identification with on-line deep feature selection for person following robots. *Robot. Auton. Syst.* **2020**, *124*, 103348.
24. Koide, K.; Miura, J. Identification of a specific person using color, height, and gait features for a person following robot. *Robot. Auton. Syst.* **2016**, *84*, 76–87.
25. Eisenbach, M.; Vorndran, A.; Sorge, S.; Gross, H.M. User recognition for guiding and following people with a mobile robot in a clinical environment. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 3600–3607.
26. Wang, M.; Liu, Y.; Su, D.; Liao, Y.; Shi, L.; Xu, J.; Miro, J.V. Accurate and real-time 3-D tracking for the following robots by fusing vision and ultrasonar information. *IEEE/ASME Trans. Mechatron.* **2018**, *23*, 997–1006.
27. Chi, W.; Wang, J.; Meng, M.Q.H. A gait recognition method for human following in service robots. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *48*, 1429–1440.
28. Kobilarov, M.; Sukhatme, G.; Hyams, J.; Batavia, P. People tracking and following with mobile robot using an omnidirectional camera and a laser. In Proceedings of the 2006 IEEE International Conference on Robotics and Automation, ICRA 2006, Orlando, FL, USA, 15–19 May 2006; pp. 557–562.
29. Huh, S.; Shim, D.H.; Kim, J. Integrated navigation system using camera and gimbaled laser scanner for indoor and outdoor autonomous flight of UAVs. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 3158–3163.
30. Boschi, A.; Salvetti, F.; Mazzia, V.; Chiaberge, M. A cost-effective person-following system for assistive unmanned vehicles with deep learning at the edge. *Machines* **2020**, *8*, 49.
31. Pang, L.; Zhang, Y.; Coleman, S.; Cao, H. Efficient hybrid-supervised deep reinforcement learning for person following robot. *J. Intell. Robot. Syst.* **2020**, *97*, 299–312.
32. Chen, B.X.; Sahdev, R.; Tsotsos, J.K. Integrating stereo vision with a CNN tracker for a person-following robot. In Proceedings of the International Conference on Computer Vision Systems, Shenzhen, China, 10–13 July 2017; pp. 300–313.
33. Cen, M.; Huang, Y.; Zhong, X.; Peng, X.; Zou, C. Real-time Obstacle Avoidance and Person Following Based on Adaptive Window Approach. In Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 4–7 August 2019; pp. 64–69.
34. Zhang, K.; Zhang, L. Autonomous following indoor omnidirectional mobile robot. In Proceedings of the 2018 Chinese Control and Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 461–466.
35. Chen, C.W.; Tseng, S.P.; Hsu, Y.T.; Wang, J.F. Design and implementation of human following for separable omnidirectional mobile system of smart home robot. In Proceedings of the 2017 International Conference on Orange Technologies (ICOT), Singapore, 8–10 December 2017; pp. 210–213.
36. Papandreou, G. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
37. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple Online and Realtime Tracking. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016.
38. Saha, O.; Dasgupta, P. A Comprehensive Survey of Recent Trends in Cloud Robotics Architectures and Applications. *Robotics* **2018**, *7*, 47. https://doi.org/10.3390/robotics7030047.
39. Maruyama, Y.; Kato, S.; Azumi, T. *Exploring the Performance of ROS2*; EMSOFT '16; Association for Computing Machinery: New York, NY, USA, 2016. https://doi.org/10.1145/2968478.2968502.