

Comparison of Interaction Modalities for Mobile Indoor Robot Guidance: Direct Physical Interaction, Person Following and Pointing Control

Aleksandar Jevtić, *Member, IEEE*, Guillaume Doisy, *Student Member, IEEE*, Yisrael Parmet, Yael Edan, *Member, IEEE*

Abstract—Three advanced natural interaction modalities for mobile robot guidance in an indoor environment were developed and compared using two tasks and quantitative metrics to measure performance and workload. The first interaction modality is based on direct physical interaction requiring the human user to push the robot in order to displace it. The second and third interaction modalities exploit a 3D vision-based human-skeleton tracking allowing the user to guide the robot by either walking in front of it or by pointing towards a desired location. In the first task, the participants were asked to guide the robot between different rooms in a simulated physical apartment requiring rough movement of the robot through designated areas. The second task evaluated robot guidance in the same environment through a set of waypoints, which required accurate movements. The three interaction modalities were implemented on a generic differential drive mobile platform equipped with a pan-tilt system and a Kinect camera. Task completion time and accuracy were used as metrics to assess the users' performance, while the NASA-TLX questionnaire was used to evaluate the users' workload. A study with 24 participants indicated that choice of interaction modality had significant effect on completion time ($F(2,61)=84.874$, $p<0.001$), accuracy ($F(2,29)=4.937$, $p=0.016$), and workload ($F(2,68)=11.948$, $p<0.001$). The direct physical interaction required less time, provided more accuracy and less workload than the two contactless interaction modalities. Between the two contactless interaction modalities, the person-following interaction modality was systematically better than the pointing-control one: the participants completed the tasks faster with less workload.

This research was supported by the EU-funded Initial Training Network (ITN) in the Marie-Curie People Programme (FP7): INTRO (INteractive RObotics research network), grant agreement no. 238486, and partially supported by the Helmsley Charitable Trust through the Agricultural, Biological and Cognitive Robotics Initiative and by the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering, both at Ben-Gurion University of the Negev, and the Beatriu de Pinós grant no. 2013 BP-B 00239 of the Catalan Government and the EU-funded Marie Curie Actions COFUND programme.

A. Jevtić was with Robosoft, Bidart 64210, France. He is now with the Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona 08028, Spain, and the Institute of Advanced Industrial Technologies (ESTIA), Bidart 64210, France (email: ajevtic@iri.upc.edu).

G. Doisy, Y. Parmet and Y. Edan are with the Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer Sheva 84110, Israel (e-mail: doisyg@post.bgu.ac.il; iparmet@bgu.ac.il; yael@bgu.ac.il).

Index Terms—Human-robot interaction, mobile robots, direct physical interaction, person tracking, person following, gesture recognition

I. INTRODUCTION

ALTHOUGH mobile indoor robots perform meaningful tasks in several application domains, very few can be found in real commercial settings. One of the reasons for this limited use is the complexity of the human-robot interaction (HRI) for non-expert users [1]. A mobile home-assistant robot, such as the one used in this work, is required to move inside the user's living space and often the user must guide the robot to a desired location. Several modalities are commonly applied to motion control of mobile robots [2] ranging from traditional methods such as using personal digital assistants and joysticks to more advanced methods such as using haptic drivers, speech and hand gestures [3], [4], [5], [6], [7], [8], [9], [10], [11]. While some of the modalities such as speech recognition have matured enough for commercial use, others have not, often due to the novelty of the underlying technologies they use. In several cases multi-modal interactions were developed combining the advantages of each modality and switching between them [12]. Natural interfaces aim to create a seamless interaction between the human and the machine [13]. Several advanced natural interfaces for control of robot's motions along a designated path using physical interaction [14], person-following [15] and pointing [16], which represent different interaction modalities have never been evaluated and compared in user studies in the context of robot guidance. It is important to evaluate each interaction modality within the context of its uses and provide quantitative results in the comparison [2].

In this paper, three interaction modalities that allow users to intuitively guide the robot inside a test apartment are evaluated, namely direct physical interaction (DPI), person-following and pointing-control. They were chosen and designed with the goal to make personal robots simple to use in a home environment for untrained users. Extensions of [14], [15], [16] include new developments. While DPI modality investigated a tactile human-robot interaction without the use of additional sensors, person-following and pointing-control

modalities exploited a powerful human tracking ability of a recently released depth camera.

A. Direct Physical Interaction

Direct physical interaction (DPI), also known as physical human-robot interaction (pHRI), allows the user to influence the robot behavior through physical contact with it [17]. The primary application of physical human-robot interaction was safety, in order for a robot and a human to share the same workspace without the risk of traumatic injury. For example, comparison of the force generated by the robot's actuators with the values predicted by a dynamic model of the robot allows detection of the force created from physical contact (i.e. a disturbance in the model), as shown by a case of interaction with a robotic arm in [18], [19]. Also, in [20], [21] a robotic arm is capable of detecting a collision and stopping its motion in order to prevent an injury to the human operator or a damage to the robot. This method can be applied on the same hardware to prevent soft tissue injuries such as human skin cuts caused by a knife or a similar sharp tool, which is manipulated by the robotic arm [22]. Furthermore, DPI has also been used in manipulation of robotic arms, generally through impedance control [23], which enabled the exploration of new natural techniques for human-robot contact [24], human-robot cooperation [25], [26], object transfer [27], teleoperation [28], or kinesthetic learning by imitation [29].

The use of DPI is not limited to robotics arms and its benefits have been explored in manipulation and guidance tasks of mobile robots proving it to be more intuitive than classical gamepad interfaces [30]. State-of-the-art methods propose the use of external force sensors or torque sensors mounted on the robot [31], or compliant joints located on the robot upper part in order to measure the contact forces, as with the robots like Cody [17], [30], [32], PR2 [33], Justin [34] and IRL-1 [35]. However, these methods limit the surface on the robot body where physical interaction can take place. Doisy [14] has presented a proof of concept that an indoor mobile robot can be controlled by applying force to any part of the robot body without the need of external sensors. Based on this concept, in this paper we developed a new DPI interaction modality that does not rely on external sensors. This was achieved by implementing a friction compensation method, which is explained in Section II.A.

B. Person Following

The idea of a person-following robot is not new and it has been applied in areas including robot companions, smart shopping carts, transporters, and walking assistants. Two challenging tasks constitute the person-following behavior, namely robot navigation and person tracking. While the robot navigation has been investigated [24], the lack of affordable and powerful sensors and efficient person-tracking techniques limited its application to person-following.

While some rely on smart environments [36], most object and person detection and tracking methods use vision-based techniques [37], [38]; however, they are sensitive to illumination changes that can degrade segmentation results

[39]. Laser rangefinders (LRFs) provide accurate distance measurements and they are generally used to detect the legs of the person [40], [41]. Legs can easily be confused with tables and chairs, so they must be filtered out by mapping the environment. Some authors propose filtering [42], [43] or sensor fusion techniques [44], [45], [46], [47], [48] in order to improve tracking performance. The use of stereo vision cameras for person tracking has also been reported [49], in combination with LRF [50], or LRF and color-image segmentation [51], [52].

Affordable depth sensors for indoor applications such as Microsoft Kinect has led to development of new algorithms for human body segmentation [53], [54]. The Kinect can provide depth images at the rate of 30 fps allowing real-time object segmentation, which is based on distance gradient and insensitive to variable lighting conditions. This technology has influenced a wide range of application domains such as object detection [55], person tracking [56], [57], [58], SLAM [59], 3D surface reconstruction [60], and human gesture and action recognition [61], [62].

The depth sensors simplify the problem of indoor person tracking and allow development of more efficient person-following robots [15], [63], [64]. Various person-following strategies using a Kinect sensor and a pan-tilt mechanism mounted on an indoor mobile robot were compared [15]. Although we proved that uninterrupted person tracking and following can be performed with the proposed combination of hardware and algorithms [15], this was not evaluated. In the current paper, the proposed person-following interaction modality which combines a Kinect sensor mounted on a pan tilt mechanism and a unique algorithm is implemented and evaluated in a comparative user study with two other interaction modalities.

C. Pointing Control

Pointing is recognized as one of the most intuitive gestures for indicating a location or an object of interest [65]. The idea of using this gesture for robot control appeared early [66], and also proved to be an accepted way of interaction for a specific category of lay users such as the older adults [67]. Various combinations of sensors and algorithms can be used to track the pointing motion. Smart devices that are manipulated by the user proved to provide accurate tracking results, as shown in the case of laser pointers [67], [68], [69], [70], a mobile phone [71] or other devices that were specifically developed for this application such as the XWand [72], the WorldCursor [73], or the Google Glass [74]. However, the need to hold or wear a device to perform pointing is neither practical nor intuitive.

Research in image and video processing for pointing gesture recognition has been conducted with systems comprising of one or more cameras [75], [76], [77], [78], [79], [80], [81]. The proposed techniques lack accuracy and fixed-camera systems spatial constraints limit the application in a mobile-robot scenario. To improve the accuracy some authors proposed to limit the number of "points of interest" and providing a prior knowledge about their location [82]. As for the person-following interaction modality, the pointing

interaction modality can benefit from the skeleton-tracking ability of the depth sensors. The novel algorithms for human-body segmentation [53], [54] provide improved speed and accuracy and have shown benefits for hand tracking, arm tracking, posture recognition [83] and pointing recognition [84].

The pointing target is derived from the position of the arm and hand joints using their 3D location obtained from a depth sensor input. Various combinations of joints provide good results in pointing target recognition such as the hand and the elbow, the hand and the shoulder, or the hand and the head [85]. Alternatively, in [86] we proposed to allow the robot to learn the relation between pointing gestures and control commands. In the current work, the pointing-control interaction modality uses the location of the hand and the elbow joints to calculate the pointing target, as in [16]. Here, the concept is implemented on a different platform and evaluated in a comparative user study with two other interaction modalities.

II. ROBOT INTERACTION MODALITIES

This work focuses on the evaluation of advanced interaction modalities for guiding a mobile robot. The aim was to evaluate the interaction modalities in a study with untrained users. The study was performed for two guidance tasks that differed in the required accuracy of movement. The three interaction modalities developed and evaluated in this paper were based on [14], [15], [16] implementing necessary insights and modifications to further advance these interaction modalities. The interaction modalities described in this section were compared in a study in an apartment facility with a total of 24 users from different backgrounds. Quantitative metrics were used for the comparison of performance and workload. The user study is described in section III.

A. Direct Physical Interaction

The DPI interaction modality allows users to guide the robot by pushing it around the apartment. One of the advantages of this interaction modality is that it allowed users to push the robot at any point of its body; however, the mobile base generally allowed the easiest transfer of user-generated force. When the users push the robot, they experience only the resistance caused by the robot's inertia. As a result of the proposed friction compensation the robot manipulation requires very little effort. The controller compensates the friction forces induced by the motors, the wheels and the ground; little friction remains uncompensated to ensure stability, e.g. prevent the robot from constantly accelerating. The overall experience when manipulating the robot is as if it is on ice and pushing it with a tip of one's finger is enough to make it move. Although this interaction modality was shown in [16] it has not been formally described. It differs from the concept previously described in [14] in the sense that a whole range of speeds from 0 to 1m/s are directly accessible to the user, whereas in [14] the user could only initiate or stop one of the four predefined robot movement: moving forward or backward at 0.5 m/s, and rotating in place clockwise and

counter-clockwise at 120 deg/s.

In designing the friction-compensation controller, the total friction torque generated by the friction forces on each motor was measured as a function of speed. In a set of experiments the robot was driven at constant wheel speeds from 0 to 13.3 rad/s (corresponding to ground speeds from 0 to 1m/s) with 0.5 rad/s increment and the corresponding measured torques were noted. For absolute angular speeds from 0 to $\hat{\theta}_l = 2.4$ rad/s (corresponding to a ground speed of 0.18 m/s), the friction is approximately proportional to the wheel angular speed. From $\hat{\theta}_l$ it does not increase with speed anymore and stays approximately constant.

The obtained friction torque measurements that were measured from the current drawn from the electrical motors were approximated by the following formulas:

$$\begin{aligned} \tau_f &= \frac{C_f}{\hat{\theta}_l} * -\hat{\theta} \quad \text{when} \quad \|\hat{\theta}\| \ll \hat{\theta}_l \\ \tau_f &= \frac{C_f * -\hat{\theta}}{\|\hat{\theta}\|} \quad \text{when} \quad \|\hat{\theta}\| > \hat{\theta}_l \end{aligned} \quad (1)$$

where τ_f is the friction torque in Nm, C_f is the friction coefficient determined experimentally of 0.85 Nm, and $\hat{\theta}$ is the wheel angular speed in rad/s.

Then, having an approximation of the friction torque for each motor, the friction was compensated with the following control law:

$$\tau_R = -\tau_{fR} * 0.8 \quad (2)$$

$$\tau_L = -\tau_{fL} * 0.8 \quad (3)$$

where τ_R , τ_L , τ_{fR} , and τ_{fL} are right and left wheels command torques and friction torques. The control law loop runs at 200 Hz. The 0.8 factor is present to ensure stability and to keep the virtual friction non-zero.

B. Person Following

In the person-following robot interaction modality, the user walks in front of the robot and the robot follows it; the user leads it to the desired location in the apartment. The robot smoothly follows the user at a safe distance so any physical contact between the robot and the user is prevented. The user could stop the robot at any time by raising his or her left hand above the level of the left elbow, and could restart the robot motion by putting the left hand back to the position below the level of the left elbow.

Person-following in this paper is achieved through uninterrupted user tracking. The integration of the pan-tilt mechanism on top of which the Kinect sensor is mounted enables decoupled motion of the sensor and the robot and extends the Kinect's horizontal detection range. The position of the user obtained from the Kinect is fed to the visual control module of the pan-tilt mechanism, which then ensures that the sensor always faces the user. The position of the pan-tilt mechanism and the position of the robot on the map are used

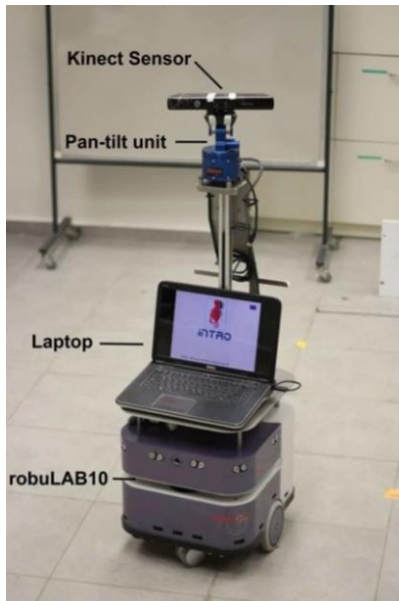


Fig. 1. Robot platform used in the user study.

to compute the position of the user on the map. This position is fed back to the robot navigation module that executes the person-following behavior. Detailed description of the tracking, control and adaptive person-following algorithms is given in [15] and was modified in this paper for a new platform to enable evaluation.

C. Pointing Control

In the pointing interaction modality employed in this paper, the user guides the robot by pointing with the right hand at a desired location on the floor. The robot is stopped by raising the left hand above the level of the left elbow. By lowering the left hand back to the position below the left elbow the robot is restarted and continues to move. During the demo, the participants were instructed to either point to the desired target location or farther from it, but then stop the robot by raising their left hand at the moment the robot crosses over the target location. The participants were told that pointing-control has an inherent limited accuracy which could introduce an error as a result of imperfect tracking. This measure was taken in order to minimize the influence of human-tracking inaccuracy on users' performance.

The desired destination is computed from the intersection of the ground floor plane with a line passing through the right hand and the right elbow joints, whose locations are obtained from the Kinect sensor. The joints locations are transformed beforehand from the Kinect frame of reference to the map

frame of reference, using the same method as in the previous section. Similarly, the user-tracking algorithm is identical as for the person-following.

The elbow-hand pair of joints was chosen for calculation of the pointing target as the result of a set of preliminary tests performed on the following combination of joints: wrist-hand, elbow-hand, shoulder-hand, head-hand, and the mean value of all four joint-pairs. Although some studies suggest that direct calculation of the pointing target from the joints locations is not the most accurate approach, these findings were limited to a specific setting in which the interaction took place such as a table top [87]. In [85] the authors suggest that the pointing target derived from a combination of the joints' locations provides a sufficiently good approximation if it is applied to a limited range where the interaction takes place. Our preliminary tests performed during the design of the interaction modality showed that the wrist-hand combination was less precise than the other combinations of joints due to the proximity of these two joints; i.e. even a small movement of one of the joints would cause large, hard-to-control, displacement of the pointing target. The pointing-control using the shoulder-hand and the head-hand combinations of joints was often inaccurate as result of the bending of the user's elbow, which is commonly done when pointing in close interaction. Similar inaccuracy was experienced when the pointing target was calculated as the mean value of the pointing targets from all four pairs of joints. These preliminary experiments provided empirical evidence that the elbow-hand combination of joints was the best choice for the pointing-control in the proposed robot guidance.

III. METHODS

A. Apparatus

The three interaction modalities were implemented on a customized Robosoft's Kompai robot shown in Fig. 1. The robot's base is a robuLAB10, a generic differential drive mobile platform with two propulsive wheels and two castor wheels that comes with basic navigation functions. A rigid structure was added on top of the platform, including three tubes and a tray for a laptop PC. On the top of this structure a TRAC Labs Biclops pan-tilt mechanism and a Kinect sensor were added, making a total height of 1.3 m from the ground to the top of the Kinect sensor. For navigation purposes, the base is equipped with a SICK S300 laser-range finder (LRF), which is positioned at the height of 0.24m and provides distance measurements of up to 30m with an angular field of view of 270°.



Fig. 2. Test environment: the apartment at Robosoft premises.

The pan-tilt mechanism has a tilt range of 120° and a pan range of 360° with a maximum angular velocity of $170^\circ/\text{s}$ and a maximum angular acceleration of $3000^\circ/\text{s}^2$. The precision of the angular position measurements is 0.01° . The mechanism can support a maximum payload of 4kg which is more than the weight of the Kinect sensor. In this work, the tilt value was set to 0° and person tracking was performed in the horizontal plane, using only the pan axis. The communication between the laptop PC and the pan-tilt mechanism is maintained via a USB port with a data transfer rate of up to 416kbps.

The Kinect sensor provides depth measurement from 0.8m to 4m with a vertical viewing angle of 43° and the horizontal viewing angle of 57° . It provides depth images at the resolution of 640×480 pixels and at the maximal frame rate of 30 fps. The Microsoft Kinect SDK provides person detection and skeleton joints tracking features.

The laptop PC used in this work is powered by an Intel quad-core i7 Q740 CPU with 4 GB of RAM.

B. Testing Environment

The user study was performed in a simulated physical apartment at Robosoft premises, in Bidart, France in June 2013. The apartment was fully furnished to have the functionality of a common home environment, as shown in Fig. 2. The apartment consists of a foyer with a bathroom, one large room and a kitchen which is separated from the room by a bar table. The floor is uniformly covered with a carpet allowing the robot to smoothly displace itself around the apartment.

C. Robot Guidance Tasks

In both tasks the robot starting location was the same. In the first task, the area-guidance task, the participants were instructed to guide the robot through three different areas that were clearly marked on the apartment floor: the square in the center of the living room, the foyer, and the kitchen. The aim was to evaluate the general ability to guide the robot inside the apartment with no need for fine accuracy of the robot's motions. In the second task, the waypoint-guidance task, the goal was to guide the robot through a set of three waypoints, marked as exact locations on the apartment floor, in order to evaluate the users' accuracy in guiding the robot. The two

tasks differed in the level of accuracy required from the user in the control of the robot's movements: for the area-guidance task the robot could be sent anywhere within a few squared meters area, whereas for the waypoint-guidance task it should be sent as close as possible to a single point.

The Kompai robot is capable of performing very accurate maneuvers, such as driving along the wall or through a narrow passage between two obstacles. This allows creating tasks of various levels of difficulty. The waypoints were chosen to provide a reasonable level of difficulty while minimizing the occurrence of loss-of-tracking events.

A robot-generated map of the apartment with the resolution of 1.67cm/pixel is shown in Fig. 3. The size of the apartment was 6.55m x 5.20m. The Karto library that implements the Monte Carlo Localization algorithm [88] was used to generate the map from the LRF readings, but also to provide robot localization and path planning with both static and dynamic obstacle avoidance. The algorithm generates the optimal path between the robot's current location and the target location on the map taking into account the preset minimal distance between the obstacles and the path points. If the robot encounters a non-mapped obstacle on the generated path, the

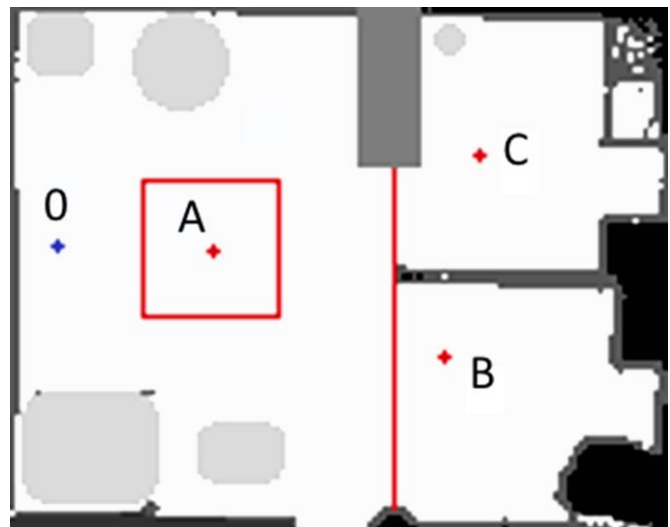


Fig. 3. Karto-generated map of the apartment with marked target areas and waypoints.



Fig. 4. Screenshots of three interaction modalities: person following (left); pointing control (center); direct physical interaction (right).

algorithm recalculates a new path taking into account the obstacle's dimensions, which are estimated from the LRF readings. If a solution for the new path cannot be found, the robot will stop and wait for a new target location from the user. On the map (see Fig. 3), the placement of the furniture is displayed in gray, the starting robot location is displayed as the blue cross with the label "0", and the target areas and waypoints are shown as red lines and crosses, respectively, with labels "A", "B" and "C" representing the square in the center of the living room, the foyer and the kitchen, respectively. The waypoints' coordinates are given in centimeters assuming "0"-waypoint as the origin of the coordinate system: A (31, 0), B (385, -110), C (420, 90). The Euclidean distances between the subsequent waypoints are: $0A = 155\text{cm}$, $AB = 252.83\text{cm}$, $BC = 203.5\text{cm}$, and $CA = 279.85\text{cm}$.

1) Area guidance task

The area guidance task was used to evaluate the difficulty in using each of the three interaction modalities for robot guidance in the apartment. In each trial, the robot was placed at the starting point "0" and guided by the participant through a set of areas in the following order: 0-A-B-C-A. The robot was considered to be inside an area once its central axis crossed the area's border line (red lines on the map in Fig. 3).

2) Waypoint-guidance task

The waypoint-guidance task was used to evaluate the difficulty in using each of the three interaction modalities for accurate robot guidance in the apartment. As in the Task 1, in each trial, the robot was placed at the starting point "0" where it was further guided by the participants, but this time the users had to direct the robot through a set of waypoints in the following order: 0-A-B-C-A. The participants could stop the robot at any distance from a waypoint, and this distance was later used to evaluate the guidance accuracy.

D. Performance and Workload Measures

The following metrics were used to assess the performance and the workload of the users.

1) Completion time

The completion time is the total time used by the participant to guide the robot through the set of waypoints or areas. Lower completion times were indicators of a better performance.

2) Accuracy

Accuracy was measured as a function of distance between the waypoint and the location where the participant stopped the robot, which was obtained from the robot localization data. This metric was used to evaluate the users' performance in accurately controlling the robot and it was applied only to the waypoint-guidance task. Lower distances were indicators of better performance.

3) Raw NASA-TLX questionnaire

Participants completed a computerized version of the questionnaire after each trial. The raw NASA-TLX enables the collection of six dimensions of workload ranging from 0 to 100 [89], and was used to assess the participant workload when guiding the robot, similarly as in [90].

E. Participants

A mixed between- and within-subject design was used in the user study. 24 participants, 9 females and 15 males, aged from 22 to 58 years (average 37.2, $SD=11.3$) were divided in two groups. The users had a variety of backgrounds (5 administrative personnel, 2 teachers, 2 students, 6 technical personnel, 5 unemployed, 5 miscellaneous: physical worker, salesman, environmentalist, salesman, pharmacist). None had previous experience with the developed interaction modalities.

F. Procedure

The researcher conducting the study was present in the apartment, but did not interfere during the task execution. Before each trial, the participants were given a short presentation about the robot features and abilities. They were told that the robot can be guided using three modalities, namely pushing, following and pointing (as shown in Fig. 4), and that they will perform three trials, one with each interaction modality. They were also assured that the robot speed is limited and that no harm will happen either to them or to the robot in case of collision.

The order of the interaction modalities was permuted between the participants to avoid any learning bias. Before each trial, the participant was informed about the procedure that consisted of: 1) a demo of the interaction modality by the researcher, 2) a test trial by the participant, 3) the trials performed by the participant, and 4) a questionnaire about the performed trials. The goal of the study in terms of the speed and accuracy, depending on the task at hand, was described to

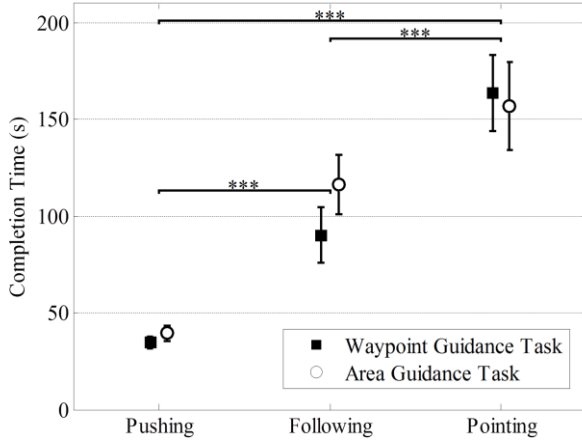


Fig. 5. Effect of the interaction modality and task type on the task completion time. Significant effects are marked with stars: *, ** and *** respectively represent the significance level of 0.05, 0.01 and 0.001. Error bars represent the standard error of the mean.

the participant and presented as a competition with other participants in order to motivate them to perform at their best abilities. The test trial was limited to only one room and lasted up to several minutes. Participants did not receive any feedback on their performance before starting the trials. However, after all the trials were finished, they were informed about their ranking relative to the others.

At the beginning of each trial, the robot was placed at its starting position (point 0 at Fig. 3). For the person-following and the pointing-control modalities the robot was started by the operator from the GUI on the robot-mounted laptop PC. With the DPI interaction modality the participants had access to the GUI on the laptop PC and they started the robot themselves. During the task execution, the participants were instructed to stop the robot at the target areas (task 1) or waypoints (task 2). For the pointing and the following interaction modalities, the participants could stop the robot by raising their left hand above the level of their left elbow; this action was detected by the Kinect sensor and it would store the robot location in the apartment. Lowering the hand back to the position below the level of the elbow would restart the robot and the trial could continue. For the DPI interaction modality, the participants would physically stop the robot and then click on the GUI to store its location. An occasional loss of tracking would activate the recovery procedure that displaces the robot to the starting point and position. The participants could interrupt this procedure at any time by standing in front of the robot within the detection range of the Kinect sensor; this would restore tracking and allow the participants to continue with the trial.

G. User Study Design

The guidance task type was the between-subject variable: waypoint-guidance task (group A) and area guidance task (group B). The interaction modality was the within-subject variable: each participant completed the guidance task using each of the three interaction modalities once. The possible order effect was counter-balanced by permuting the order of

the interaction modalities used between the participants.

The specific technical features of the Kompai robot may influence the overall user experience. Only the DPI modality required physical contact with the robot. Two procedures were included in the user study to minimize the undesirable effect of influence of the specific system. First, the researcher conducting the study described the robot features to each participant before starting the trials. Second, the researcher performed a demo and allowed each participant to test the robot and become familiar with its navigating abilities by using all three interaction modalities.

H. Data Analysis

Since task completion times and accuracy (distance between the waypoints and the locations where the participant stopped the robot) have a skewed distribution they were log-transformed to achieve model assumptions of normality and homoscedasticity. Therefore, on the original scale we do not assume homoscedasticity but a positive relationship between the mean and the variance, for both accuracy and completion time data. Then, a Linear Mixed Model (LMM) analysis [91] was conducted on all the metrics with the interaction modality (pointing-control, person-following and DPI) as the within-group fixed effect and the task type (waypoint and area guidance task) as the between-group fixed effect. Accuracy was analyzed using a LMM with only one factor: the interaction modality. Participants were included as a random effect to account for individual differences among participants and the correlations among repeated measures within participants. LMM analysis was employed rather than ordinary ANOVA with repeated measures due to the fact that there were missing values and in order to utilize the information of those observations without the need for supplementary data.

When necessary, post-hoc pairwise comparisons were conducted using the Least Significant Difference method. We considered the results as significant below an alpha of 0.05.

IV. RESULTS

Out of the 72 trials conducted by the 24 participants, 6 were not completed by 5 different participants. These failures occurred in 3 pointing-control trials, 1 person-following trial and 2 direct physical interaction trials.

A. Completion Time

The analysis conducted on the log of the completion times reveals that there is no significant effect of the task type on the completion time. However, the effect of the interaction modality on completion time was significant, $F(2,61)=84.874$, $p<0.001$. Participants completed the tasks faster when using the DPI interaction modality (37 seconds, $SD=11.8$), slower when using the pointing interaction modality (160 seconds, $SD=68$), and had intermediate completion times when using the person-following interaction modality (103 seconds, $SD=52$). Post-hoc pairwise comparisons confirm that the difference between each interaction modality is significant, with $p<0.001$ for the three comparisons. Completion times are

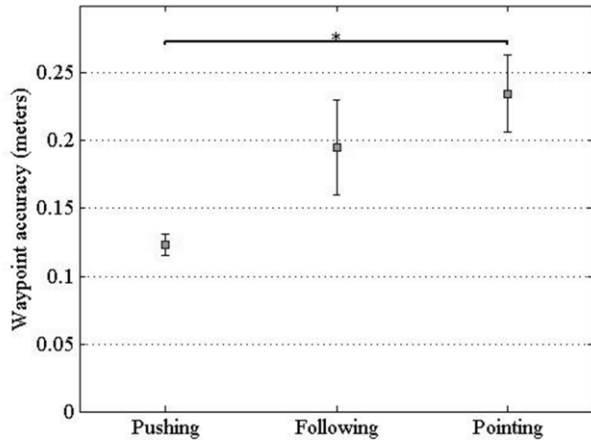


Fig. 6. Effect of the interaction modality and task type on the waypoint accuracy. Significant effects are marked with stars: *, ** and *** respectively represent the significance level of 0.05, 0.01 and 0.001. Error bars represent the standard error of the mean.

shown in Fig. 5.

B. Accuracy

Analysis shows, as seen in Fig. 6, that when using the DPI interaction modality participants were significantly more accurate than when using the pointing-control interaction modality ($p=0.016$). No significant difference was found with the person-following interaction modality.

C. Raw NASA-TLX Questionnaire

1) Overall Workload

The overall workload is the average of the six dimensions of workload measured with the NASA-TLX questionnaire.

Analysis reveals, as shown in Fig. 7 that the interaction modality had significant influence on the overall workload, $F(2,68)=11.948$, $p<0.001$. However, there is no significant effect of the task type on the overall workload.

These overall workload results are coherent with the completion time results. The DPI interaction modality appears to be the easiest to use, the pointing interaction modality the hardest, and the person-following interaction modality induced an intermediate workload compared to the two other interaction modalities. Post-hoc pairwise confirms that the difference between each interaction modality is significant.

2) Detailed workload dimensions

The interaction modality had significant effect on the four workload dimensions, as shown in Fig. 8: *Mental Demand* ($F(2,68)=18.642$, $p<0.001$), *Performance* ($F(2,68)=8.324$, $p<0.001$), *Effort* ($F(2,68)=7.274$, $p<0.001$), and *Frustration* ($F(2,68)=13.117$, $p<0.001$). The effect of the interaction modality was the same on these four dimensions: the DPI interaction modality appears to be the least demanding, whereas the pointing interaction modality is the most demanding and the person-following interaction modality workload dimensions score in between. Pairwise comparisons differences were checked and they are significant on all pairs apart between DPI and person-following for *Effort* and *Frustration*.

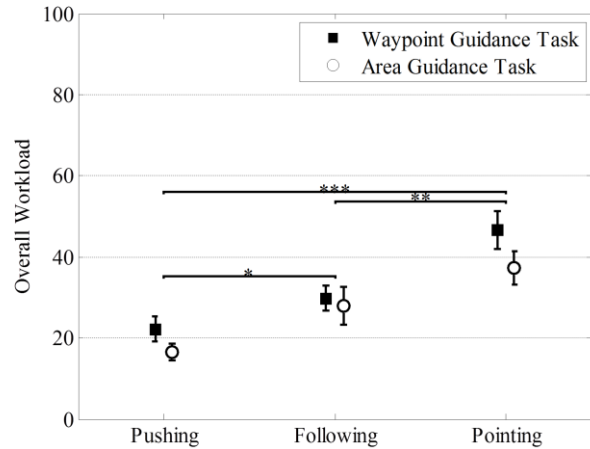


Fig. 7. Effect of the interaction modality and task type on the overall workload. Significant effects are marked with stars: *, ** and *** respectively represent the significance level of 0.05, 0.01 and 0.001. Error bars represent the standard error of the mean.

Two workload dimensions (Physical Demand and Temporal Demand), were not affected by the choice of the interaction modality. Out of the six workload dimensions, none were significantly impacted by the task type.

V. DISCUSSION

Three advanced interaction modalities for guiding an indoor mobile robot in two different guidance tasks were compared. The effect of the interaction modality was statistically significant for almost all the variables measured, and remarkably consistent. The DPI is systematically better than the two indirect interaction modalities. Participants completed the tasks faster, with more accuracy, less mental demand, less effort, less frustration and had the feeling they performed better. This result highlights the advantage of robot physical control in terms of performance and workload compared to contactless interaction modalities.

In this study, between the two contactless interaction modalities tested, the person-following interaction modality appears to be systematically better: participants completed the task faster and more accurately, with less mental demand, less effort, less frustration and had the feeling they performed better. The variance in the accuracy results is higher for the person-following and the pointing-control interaction modalities than in the case of the DPI interaction modality. This may be explained by the fact that, unlike pushing, person following and pointing do not give direct control over the robot in real time; the user commands are introduced with a delay which adds some noise in the movements of the robot. It also confirms the log-normal assumption that a higher mean value produces a higher variance. Moreover, the person-following interaction modality uses the head-joint location as input, and while walking the head performs lateral movement perpendicular to the walking direction, which is an additional source of noise. The advantage of the person-following over

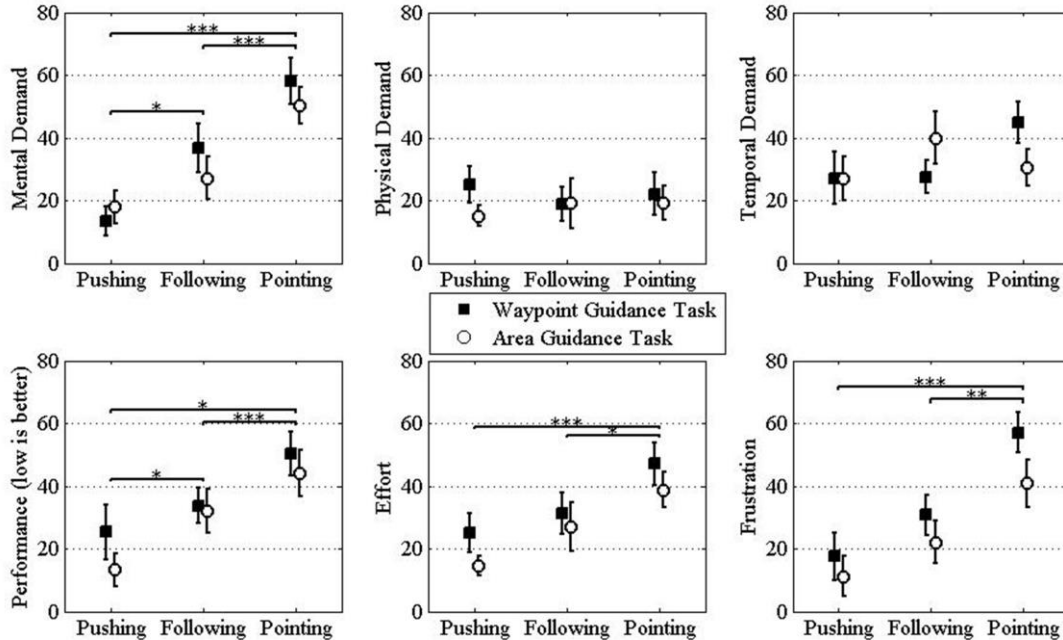


Fig. 8. Effect of the interaction modality and task type on the 6 dimensions of workload. Significant effects are marked with stars: *, ** and *** respectively represent the significance level of 0.05, 0.01 and 0.001. Error bars represent the standard error of the mean.

pointing-control is in the fact that once started, the participants do not need to actively guide the robot, it is a “start and forget” technique, they just have to walk, knowing that the robot is following their steps. Whereas, with the pointing-control, the participants had to constantly take care of the robot guidance requiring more effort. Additionally, pointing-control sometimes required the user to walk while maintaining a pointing direction with their arm. This may have disturbed the gait, which would explain the longer completion time of this modality.

Our results of workload are consistent with our objective measures, suggesting that the most efficient ones are also the ones the users felt more comfortable with to some extent. Direct physical interaction was the best interaction modality in the performed study. If contactless guidance is needed, for instance when the hands of the operator are busy, the person-following interaction modality is the best. Yet one could argue that this result could vary depending on the robustness of the person-following algorithm and the environmental situation. For instance, in a complex and dynamic environment, it is more likely that the robot will lose track of the followed operator and therefore active pointing-control would be preferred. Still, person-following algorithm robustness is a technological issue, and when properly working like in this study, it presents definitive advantages over other interaction modalities. We believe that in terms of workload and in the context of service robots, it is better than conventional robot interaction modalities.

For all control interaction modalities, surprisingly, no significant effect of the task type was found on any of the metrics measured, for both objective and subjective metrics.

One can argue that the two guidance tasks tested, waypoint guidance and area guidance, were similar, but the waypoint guidance required more accurate control of the robot. When completing the guidance task with the added constraint of passing through a waypoint, it was expected that the participants’ workload would increase. But this was not the case, and for none of the six measured dimensions significant difference was noted. This result shows that accurate control of the robot is not more costly for the three interaction modalities tested here.

The study compared the performance and the workload of the three interaction modalities on a group of untrained users for two different guidance tasks. Further research may show that the performance and workload of these interaction modalities could vary for example, when the users after training become more familiar with the robot. Hence, the margin of progression of these interaction modalities in terms of performance and workload is certainly a subject to investigate in the future. Moreover, the performance and the workload associated with each interaction modality may not characterize their usability in a strict sense; however, they can be considered as a good first approximation and base of comparison. The choice of the interaction modality to use may also be affected by the target user groups that can have different requirements and abilities, e.g. disabled people, older adults, young children, and so forth. Future applications should take into account both user and task requirements. Additionally, the different characteristics of the robot platform such as its power, size, form factor and sensors can also have an effect and these conclusions need to be tested on different hardware to be generalized. The specific combination of

hardware and software and the unique testing environment used in the study make the comparison with other research results difficult; nevertheless, the results that we obtained could be used by other researchers who are interested in setting up their experiments in a similar fashion. Despite the noted limitations, this research sheds light on the different interaction modalities for robot guidance along paths or between areas making it easier to cope with the challenges of developing and operating such modalities.

REFERENCES

- [1] P. Rouanet, P.-Y. Oudeyer, F. Danieau, and D. Filliat, "The impact of human-robot interfaces on the learning of visual objects," *IEEE Trans. Robot.*, vol. 29, no. 2, pp. 525–541, Apr. 2013.
- [2] A. Eliav, T. Lavie, Y. Parmet, H. Stern, and Y. Edan, "Advanced methods for displays and remote control of robots," *Appl. Ergon.*, vol. 42, no. 6, pp. 820–9, Nov. 2011.
- [3] T. Fong, F. Conti, S. Grange, and C. Baur, "Novel interfaces for remote driving : gesture, haptic and PDA," *Proc. SPIE, Int. Soc. Opt. Eng.*, vol. 4195, pp. 300–311, 2000.
- [4] T. Fong, C. Thorpe, and C. Baur, "Collaborative control: a robot-centric model for vehicle teleoperation," Ph.D. Dissertation, Robotics Institute, Carnegie Mellon University, USA, 2001.
- [5] T. Fong, C. Thorpe, and C. Baur, "A safeguarded teleoperation controller," in *IEEE International Conference on Advanced Robotics (ICAR)*, Budapest, Hungary, 2001.
- [6] J. D. Lee and K. A. See, "Trust in automation: designing for appropriate reliance," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 46, no. 1, pp. 50–80, Jan. 2004.
- [7] S. Lee, G. Sukhatme, G. J. Kim, and C.-M. Park, "Haptic teleoperation of a mobile robot: a user study," *Presence Teleoperators Virtual Environ.*, vol. 14, no. 3, pp. 345–365, Jun. 2005.
- [8] N. M. Sgouros and S. Gerogiannakis, "Integrating WAP-based wireless devices in robot teleoperation environments," in *Proc. 2002 IEEE International Conference on Robotics and Automation (ICRA 2002)*, 2002, vol. 2, pp. 1191–1196.
- [9] A. Chaudhary, J. L. Raheja, K. Das, and S. Raheja, "Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey," *Int. J. Comput. Sci. Eng. Surv.*, vol. 2, no. 1, pp. 122–133, Mar. 2011.
- [10] M. Forbes, R. P. N. Rao, L. Zettlemoyer and M. Cakmak, "Robot programming by demonstration with situated spatial language understanding," in *2015 IEEE International Conference on Robotics and Automation (ICRA 2015)*, Seattle, Washington, USA, May 2015, pp. 2014–2020.
- [11] L. Chan, F. Naghdy and D. Stirling, "Application of adaptive controllers in teleoperation systems: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 3, pp. 337–352, Feb. 2014.
- [12] A. Jaimes and N. Sebe, "Multimodal human-computer interaction: A survey," *Comput. Vis. Image Underst.*, vol. 108, no. 1–2, pp. 116–134, Oct. 2007.
- [13] D. A. Norman, "The way I see it: Natural user interfaces are not natural," *Interactions*, vol. 17, no. 3, p. 6, May 2010.
- [14] G. Doisy, "Sensorless collision detection and control by physical interaction for wheeled mobile robots," in *Proc. 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2012)*, 2012.
- [15] G. Doisy, A. Jevtić, E. Lucet, and Y. Edan, "Adaptive person-following algorithm based on depth images and mapping," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, Workshop on Robot Motion Planning, 2012, pp. 43–48.
- [16] A. Jevtić, G. Doisy, S. Bodiřoza, Y. Edan, and V. V. Hafner, "Human-robot interaction through 3D vision and force control," in *Proc. 2014 ACM/IEEE international conference on Human-robot interaction (HRI '14)*, 2014, pp. 102–102.
- [17] T. L. Chen and C. C. Kemp, "Lead me by the hand: evaluation of a direct physical interface for nursing assistant robots," in *Proc. 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI '10)*, 2010, pp. 367–374.
- [18] A. De Luca and R. Mattone, "Sensorless robot collision detection and hybrid force/motion control," in *Proc. 2005 IEEE Int. Conf. Robot. Autom.*, no. April, pp. 999–1004, 2005.
- [19] A. De Luca and R. Mattone, "An adapt-and-detect actuator FDI scheme for robot manipulators," in *Proc. IEEE International Conference on Robotics and Automation (ICRA 2004)*, 2004, vol. 5, pp. 4975–4980.
- [20] S. Haddadin, A. Albu-Schaffer, A. De Luca, and G. Hirzinger, "Collision detection and reaction: A contribution to safe physical Human-Robot Interaction," in *Proc. 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2008)*, 2008, pp. 3356–3363.
- [21] A. De Luca, A. Albu-Schaffer, S. Haddadin, and G. Hirzinger, "Collision detection and safe reaction with the DLR-III lightweight manipulator arm," in *Proc. 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, 2006, pp. 1623–1630.
- [22] S. Haddadin, A. Albu-Schaffer, and G. Hirzinger, "Soft-tissue injury in robotics," in *Proc. 2010 IEEE International Conference on Robotics and Automation (ICRA 2010)*, 2010, pp. 3426–3433.
- [23] N. Hogan, "Impedance control: an approach to manipulation," in *American Control Conference*, 1984, pp. 304–313.
- [24] J. G. Hale and F. E. Pollock, "'Sticky hands': learning and generalization for cooperative physical interactions with a humanoid robot," *IEEE Trans. Syst. Man Cybern. Part C*, vol. 35, no. 4, pp. 512–521, Nov. 2005.
- [25] R. Ikeura and H. Inooka, "Variable impedance control of a robot for cooperation with a human," in *Proc. 1995 IEEE International Conference on Robotics and Automation (ICRA 1995)*, 1995, vol. 3, pp. 3097–3102.
- [26] B. Whitsell and P. Artemiadis, "On the role duality and switching in human-robot cooperation: An adaptive approach," in *Proc. 2015 IEEE International Conference on Robotics and Automation (ICRA 2015)*, Seattle, Washington, USA, May 2015, pp. 3770–3775.
- [27] A. Edsinger and C. C. Kemp, "Human-robot interaction for cooperative manipulation: handing objects to one another," in *Proc. 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2007)*, 2007, pp. 1167–1172.
- [28] L. J. Love and W. J. Book, "Force Reflecting Teleoperation With Adaptive Impedance Control," *IEEE Trans. Syst. Man Cybern. Part B*, vol. 34, no. 1, pp. 159–165, Feb. 2004.
- [29] K. Mulling, J. Kober, O. Kroemer, and J. Peters, "Learning to select and generalize striking movements in robot table tennis," *Int. J. Rob. Res.*, vol. 32, no. 3, pp. 263–279, Jan. 2013.
- [30] T. L. Chen and C. C. Kemp, "A direct physical interface for navigation and positioning of a robotic nursing assistant," *Adv. Robot.*, vol. 25, no. 5, pp. 605–627, Jan. 2011.
- [31] T. Takeda, Y. Hirata, and K. Kosuge, "Dance step estimation method based on hmm for dance partner robot," *IEEE Trans. Ind. Electron.*, vol. 54, no. 2, pp. 699–706, Apr. 2007.
- [32] A. Jain and C. C. Kemp, "Pulling open novel doors and drawers with equilibrium point control," in *Proc. 9th IEEE-RAS International Conference on Humanoid Robots*, 2009, pp. 498–505.
- [33] K. A. Wyrobek, E. H. Berger, H. F. M. Van der Loos, and J. K. Salisbury, "Towards a personal robotics development platform: Rationale and design of an intrinsically safe personal robot," in *Proc. 2008 IEEE International Conference on Robotics and Automation (ICRA 2008)*, 2008, pp. 2165–2170.
- [34] M. Fuchs, C. Borst, P. R. Giordano, A. Baumann, E. Kraemer, J. Langwald, R. Gruber, N. Seitz, G. Plank, K. Kunze, R. Burger, F. Schmidt, T. Wimboeck, and G. Hirzinger, "Rollin' Justin - Design considerations and realization of a mobile platform for a humanoid upper body," in *Proc. 2009 IEEE International Conference on Robotics and Automation (ICRA 2009)*, 2009, pp. 4131–4137.
- [35] F. Ferland, A. Aumont, D. Letourneau, and F. Michaud, "Taking your robot for a walk: Force-guiding a mobile robot using compliant arms," in *Proc. 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2013)*, 2013, pp. 309–316.
- [36] N. Najmaei and M. R. Kermani, "Applications of artificial intelligence in safe human-robot interactions," *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, vol. 41, no. 2, pp. 448–59, Apr. 2011.
- [37] Z. Jia, A. Balasuriya, and S. Challa, "Vision based target tracking for autonomous land vehicle navigation: a brief survey," *Recent Patents Comput. Sci.*, vol. 2, no. 1, pp. 32–42, Jan. 2009.
- [38] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Image Underst.*, vol. 104, no. 2–3, pp. 90–126, Nov. 2006.

- [39] X. Liu and K. Fujimura, "Hand gesture recognition using depth data," in *Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 529–534.
- [40] J. M. Martínez-Otzeta, A. Ibarra, A. Ansuategi, and L. Susperregi, "Laser based people following behaviour in an emergency environment," in *Proc. 2nd International Conference on Intelligent Robotics and Applications (ICIRA '09)*, 2009, pp. 33–42.
- [41] V. Alvarez-Santos, X. M. Pardo, R. Iglesias, a. Canedo-Rodriguez, and C. V. Regueiro, "Feature analysis for human recognition and discrimination: Application to a person-following behaviour in a mobile robot," *Rob. Auton. Syst.*, vol. 60, no. 8, pp. 1021–1036, Aug. 2012.
- [42] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, "People tracking with mobile robots using sample-based joint probabilistic data association filters," *Int. J. Rob. Res.*, vol. 22, no. 2, pp. 99–116, Feb. 2003.
- [43] Y. Gu and M. Veloso, "Effective multi-model motion tracking using action models," *Int. J. Rob. Res.*, vol. 28, no. 1, pp. 3–19, Jan. 2009.
- [44] M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia, "People tracking and following with mobile robot using an omnidirectional camera and a laser," in *Proc. 2006 IEEE International Conference on Robotics and Automation (ICRA 2006)*, 2006, no. May, pp. 557–562.
- [45] N. Bellotto and H. Hu, "Multisensor-based human detection and tracking for mobile service robots," *IEEE Trans. Syst. Man. Cybern. Part B.*, vol. 39, no. 1, pp. 167–81, Feb. 2009.
- [46] L. Spinello, R. Triebel, and R. Siegwart, "Multiclass multimodal detection and tracking in urban environments," *Int. J. Rob. Res.*, vol. 29, no. 12, pp. 1498–1515, Oct. 2010.
- [47] Y. Motai, S. Kumar Jha, and D. Kruse, "Human tracking from a mobile agent: Optical flow and Kalman filter arbitration," *Signal Process. Image Commun.*, vol. 27, no. 1, pp. 83–95, Jan. 2012.
- [48] P. Stein, A. Spalanzani, V. Santos and C. Laugier, "On leader following and classification," in *Proc. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)*, Chicago, IL, USA, Sept. 2014, pp. 3135–3140.
- [49] J. Satake and J. Miura, "Stereo-based multi-person tracking using overlapping silhouette templates," in *Proc. 20th International Conference on Pattern Recognition*, 2010, pp. 4304–4307.
- [50] J. M. Martínez-Otzeta, A. Ibarra, A. Ansuategi, C. Tubío, and J. Aristondo, "People following behaviour in an industrial environment using laser and stereo camera," *Trends Appl. Intell. Syst.*, vol. 6098, pp. 508–517, 2010.
- [51] T. Yoshimi, M. Nishiyama, T. Sonoura, H. Nakamoto, S. Tokura, H. Sato, F. Ozaki, N. Matsuhira, and H. Mizoguchi, "Development of a Person Following Robot with Vision Based Target Detection," in *Proc. 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, 2006, pp. 5286–5291.
- [52] D. Calisi, L. Iocchi, and R. Leone, "Person following through appearance models and stereo vision using a mobile robot," in *Proc. International Workshop on Robot Vision*, 2007, pp. 46–56.
- [53] L. A. Schwarz, A. Mkhitarian, D. Mateus, and N. Navab, "Human skeleton tracking from depth data using geodesic distances and optical flow," *Image Vis. Comput.*, vol. 30, no. 3, pp. 217–226, Dec. 2012.
- [54] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Proc. 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, 2011, pp. 1297–1304.
- [55] M. Camplani and L. Salgado, "Efficient Spatio-Temporal Hole Filling Strategy for Kinect Depth Maps," in *IS&T/SPIE Int. Conf. on 3D Image Processing (3DIP) and Applications*, 2012.
- [56] M. Luber, G. Diego Tipaldi, and K. O. Arras, "Place-dependent people tracking," *Int. J. Rob. Res.*, vol. 30, no. 3, pp. 280–293, Jan. 2011.
- [57] G. Doisy, A. Jevtić, and S. Bodoiroža, "Spatially unconstrained, gesture-based human-robot interaction," in *Proc. 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2013)*, 2013, pp. 117–118.
- [58] D. Brščić and T. Kanda, "Changes in usage of an indoor public space: Analysis of one year of person tracking," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 2, pp. 2168–2291, Apr. 2015.
- [59] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard, "An evaluation of the RGB-D SLAM system," in *Proc. 2012 IEEE International Conference on Robotics and Automation (ICRA 2012)*, 2012, vol. 3, no. c.
- [60] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera," in *Proc. 24th annual ACM symposium on User interface software and technology (UIST '11)*, 2011, pp. 559–568.
- [61] S. Bodoiroža, G. Doisy, and V. V. Hafner, "Position-invariant, real-time gesture recognition based on dynamic time warping," in *Proc. 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2013)*, 2013, pp. 87–88.
- [62] C. Chen, R. Jafari y N. Kehtarnavaz, "Improving human action recognition using fusion of depth camera and inertial sensors," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 1, pp. 51–61, Feb. 2015.
- [63] N. Pradhan, "Mobile Robot Navigation for Person Following in Indoor Environments," Ph.D. Dissertation, Clemson University, Electrical Engineering, 2013.
- [64] D. Pucci, L. Marchetti, and P. Morin, "Nonlinear control of unicycle-like robots for person following," in *Proc. 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2013)*, 2013, pp. 3406–3411.
- [65] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan, "Vision-based hand-gesture applications," *Commun. ACM*, vol. 54, no. 2, p. 60, Feb. 2011.
- [66] R. Cipolla and N. J. Hollinghurst, "Human-robot interface by pointing with uncalibrated stereo vision," *Image Vis. Comput.*, vol. 14, no. 3, pp. 171–178, Apr. 1996.
- [67] J. M. Beer, A. Prakash, C.-A. Smarr, T. L. Mitzner, C. C. Kemp, and W. A. Rogers, "Commanding your robot: older adults' preferences for methods of robot control," in *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 56, no. 1, pp. 1263–1267, Oct. 2012.
- [68] C. C. Kemp, C. D. Anderson, H. Nguyen, A. J. Trevor, and Zhe Xu, "A point-and-click interface for the real world: Laser designation of objects for mobile manipulation," in *Proc. ACM/IEEE International Conference on Human-Robot Interaction (HRI 2013)*, 2013, pp. 241–248, 2008.
- [69] T. Suzuki, A. Ohya, and S. Yuta, "Operation direction to a mobile robot by projection lights," in *IEEE Workshop on Advanced Robotics and its Social Impacts*, 2005., 2005, pp. 160–165.
- [70] S. Shibata, T. Yamamoto, and M. Jindai, "Human-robot interface with instruction of neck movement using laser pointer," in *Proc. 2011 IEEE/SICE International Symposium on System Integration (SII)*, 2011, pp. 1226–1231.
- [71] S. Koceski, N. Koceska, and I. Koccev, "Design and evaluation of cell phone pointing interface for robot control," *Int. J. Adv. Robot. Syst.*, vol. 9, p. 1, 2012.
- [72] A. Wilson and S. Shafer, "XWand," in *Proc. Conference on Human Factors in Computing Systems (CHI '03)*, 2003, p. 545.
- [73] H. P. Andrew Wilson, "Pointing in intelligent environments with the World Cursor," in *Proc. International Conference on Human-Computer Interaction (INTERACT 2003)*, 2003, pp. 495–502.
- [74] D. A. Lazewatsky and W. D. Smart, "Evaluation of Interfaces for 3d Pointing," in *Proc. ACM/IEEE International Conference on Human-Robot Interaction (HRI 2015)*, Portland, USA, Mar. 2015, pp. 25–26.
- [75] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.
- [76] A. Azarbayejani and A. Pentland, "Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features," in *Proc. 13th International Conference on Pattern Recognition*, 1996, vol. 3, pp. 627–632 vol.3.
- [77] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, "Integrated Person Tracking Using Stereo, Color, and Pattern Detection," *Int. J. Comput. Vis.*, vol. 37, no. 2, pp. 175–185, Jun. 2000.
- [78] N. Jovic, B. Brumitt, B. Meyers, S. Harris, and T. Huang, "Detection and estimation of pointing gestures in dense disparity maps," in *Proc. Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 468–475.
- [79] R. Cipolla, P. Hadfield, and N. Hollinghurst, "Uncalibrated stereo vision with pointing for a man-machine interface," in *Proc. IAPR Workshop on Machine Vision Applications (MVA '94)*, International Association for Pattern Recognition, 1994, pp. 163–166.
- [80] K. Nickel and R. Stiefelhagen, "Visual recognition of pointing gestures for human-robot interaction," *Image Vis. Comput.*, vol. 25, no. 12, pp. 1875–1884, Dec. 2007.
- [81] P. Kondaxakis, J. Pajarinen and V. Kyrki, "Real-time recognition of pointing gestures for robot to robot interaction," in *Proc. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)*, Chicago, IL, USA, Sept. 2014, pp. 2621–2626.
- [82] M. Pateraki, H. Baltzakis and P. Trahanias, "Visual estimation of pointed targets for robot guidance via fusion of face pose and hand

- orientation,” *Computer Vision and Image Understanding*, vol. 120, pp. 1–13, Mar. 2014
- [83] J. A. Diego-Mas and J. Alcaide-Marzal, “Using KinectTM sensor in observational methods for assessing postures at work,” *Appl. Ergon.*, vol. 45, no. 4, pp. 976–85, Jul. 2014.
- [84] J. Suarez and R. R. Murphy, “Hand gesture recognition with depth images: A review,” in *Proc. 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2012)*, 2012, pp. 411–417.
- [85] D. Droschel, J. Stückler, and S. Behnke, “Learning to interpret pointing gestures with a time-of-flight camera,” in *Proc. 6th international conference on Human-robot interaction (HRI '11)*, 2011, p. 481.
- [86] S. Bodiroža, A. Jevtić, B. Lara, and V. V. Hafner, “Relation of motion control and gestures through self-exploration,” in *Proc. 1st Workshop on Robotics Challenges and Vision, Robotics: Science and Systems Conference*, 2013, pp. 21–24.
- [87] P. McGuire, J. Fritsch, J. J. Steil, F. Rothling, G. A. Fink, S. Wachsmuth, G. Sagerer, and H. Ritter, “Multi-modal human-machine communication for instructing robot grasping tasks,” in *IEEE/RSJ International Conference on Intelligent Robots and System (IROS 2002)*, 2002, vol. 2, pp. 1082–1088.
- [88] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, Cambridge, USA, 2005.
- [89] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” in *Human Mental Workload*, vol. 52, no. null, P. A. Hancock and N. Meshkati, Eds. Elsevier, 1988, pp. 139–183.
- [90] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks, “Ecological interfaces for improving mobile robot teleoperation,” *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 927–941, Oct. 2007.
- [91] C. E. McCulloch and S. R. Searle, *Generalized, Linear, and Mixed Models*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2000.



Aleksandar Jevtić (S’10–M’11) received his B.S. degree in Electrical Engineering from the University of Belgrade, Serbia in 2005, and his M.S. and Ph.D. degrees in Computer Science from the Technical University of Madrid (UPM), Spain in 2007 and 2011, respectively. Since 2014 he works as a Beatriu de Pinós fellow at IRI institute, Spain and holds a position of

a visiting researcher at ESTIA institute, France. His research interests focus on multi-modal human-robot interaction.



Guillaume Doisy (M’11) received his Mechatronics Engineering degree from the Université de Technologie de Compiègne (UTC), France and his M.Sc in Autonomous Vehicle Dynamics and Control from Cranfield University, UK, in 2010. Since 2010 he works as a Marie Curie Fellow and pursues his Ph.D. thesis

on Intelligent Interface Design at the Ben Gurion University of the Negev (BGU), Israel, while being a scientific volunteer for the French embassy in Israel.



Yisrael Parmet is a faculty member in the Department of IEM. He holds a B.A. in Economics and Statistics and M.Sc. and Ph.D. degrees in Statistics from Tel-Aviv University. He specializes in areas of design of experiments and statistical modeling. During his studies for a master

and doctorate degrees he served as research assistant at the statistical laboratory at Department of Statistics and OR, Tel Aviv University which granted him knowledgeableness in practical data analysis.



Yael Edan (M’88) received a B.Sc in Computer Engineering, Technion, M.Sc in Agricultural Engineering, Technion and Ph.D. in Engineering, Purdue University. She is a Full Professor in Industrial Engineering at Ben-Gurion University of the Negev (BGU). She has performed research in robotics, sensors, simulation,

and decision-making systems. In addition, she has made major contributions in the introduction and application of intelligent automation and robotic systems to the field of agriculture with several patents.