"Take a seat, please": Approaching and Recognition of Seated Persons by a Mobile Robot *

Thanh Q. Trinh, Tim Wengefeld, Steffen Müller, Alexander Vorndran, Michael Volkhardt, Andrea Scheidig, and Horst-Michael Gross[a]
[a] *Neuroinformatics and Cognitive Robotics Lab*, *Ilmenau University of Technology*, PF 100565, 98684 Ilmenau, Germany, e-mail: quang-thanh.trinh@tu-ilmenau.de

Abstract

As a long-term goal, robots should be able to interact with humans in multiple forms in order to assist them in everyday life. Over the course of several projects with public operational environments, for our mobile robots equipped with touch displays as primary input/output device, we found an intuitive way to interact with the robot is in a seated position and being approached by the robot, so that the user can physically operate the robot. Sitting down is also a signal for interaction intention which is easily conveyed and even visible from afar, as opposed to speech recognition systems. To realize this behavior, we propose a succinct yet effective method of recognizing a seated person by utilizing the height estimation of a person detector, as well as a method of finding an interaction pose to approach the user and navigating to that pose even in dynamic unstructured environments. The proposed approaches are evaluated in an experimental setup similar to the dynamic hospital environment of our project ROGER where a robotic gait training coach for orthopedic rehabilitation of patient with hip prosthesis is developed.

1 Introduction

In our ongoing project ROGER[1], we aim at developing a robotic rehabilitation assistant for patients with newly implanted hip prosthesis. The robot is used as personal trainer and helps to regain the patient's normal gait pattern. A training consists of analyzing and correcting the patients gait while guiding her/him through the hospital hallways. For optimal analyses the robot needs to keep the patient at a distance of 2 - 3 m to completely perceive the body. Despite this distance, the patient has to be able to signal her/his intention to interact with the robot's primary input/output device - a touch display mounted at the front of the robot. Commonly, a training is prescribed one or two days after the implantation. To avoid overstrain and possible injuries, the patients are required to use crutches. In addition, chairs are setup along the hallways providing resting places. Since both the patient's hands are occupied, and the hospital can be very noisy at rush hours, remote controllers or speech recognition systems are not reliable options for the patient to signal interaction intentions. An intuitive way for signaling such intentions is the transition from standing to seated. After the recognition of this transition, the robot would start approaching the patient, i.e. closing the distance to her/him, to create the necessary distance to physically engage with the robot's touch display.

*This work has received funding from the Thuringian Ministry for Economic Affairs, Science and Digital Society (TMWWDG) within the project ROGER (grant agreement no. 2015 FE 9088) and the German Federal Ministry of Education and Research (BMBF) to the project 3D-PersA2 (grant agreement no. 03ZZ0460) in the program Zwanzig20 Partnership for Innovation as part of the research alliance 3Dsensation.

[1] http://bit.ly/ROGERRehab (in German)

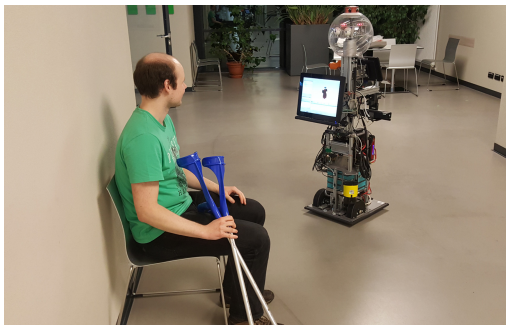


Figure 1 A typical situation of a seated patient before being approached by his training robot.

When working with recently operated hip patients, unnecessary movements have to be avoided for the patients. To present the display to a seated patient, the robot have to choose a pose as close as possible while heading towards the patient, so that s/he can comfortably reach and look at the touch display without repositioning herself/himself. This can be difficult for highly dynamic environments such as hospitals. Patients and hospital staff are moving in the corridors with beds, supply and cleaning carts, or wheelchairs occupying the hallways. This results in more or less restricted space conditions. Furthermore, the crutches putted aside after the patient sat down to operate the display, pose a permanent obstacle for the robot.

To realize the human-robot interaction behavior of recognizing seated persons and consecutively approaching them, we propose a succinct yet effective method of detecting the seated state by utilizing a person detector which

Int. Symposium on Robotics (ISR), Munich, Germany, pp. 240-247, VDE Verlag 2018

is able to estimate a person's height, as well as a technique to approach the person by solving two optimization problems: First, the best interaction pose is found given the current environmental state. Second, to navigate to the found interaction pose, motion commands hast to be determined which avoid obstacles and bystanders in a socially acceptable manner.

The remainder of the paper is organized as follows. Sec. 2 addresses related work regarding the seated estimation and methods for approaching persons. Sec. 3 gives a brief overview of the robot platform of project ROGER used in this work. In Sec. 4, our approach for recognizing seated persons is presented. Sec. 5 presents our technique of approaching a person. In Sec. 6, the performance of our proposed approaches are evaluated in experiments. Sec. 7 concludes the paper.

2 Related Work

2.1 Seated Estimation

In medical applications, a wide variety of approaches use intrusive sensors, like accelerometers, to classify the event of sitting down. The overview given in [1] states that with an increasing amount of sensors, detection rates up to 100% can be achieved. Nevertheless, these highly precise sensor setups come with the price of becoming uncomfortable for the wearer. Since our task is to analyse unaffected gait pattern, these approaches using intrusive sensors seem to be impracticable for our scenario. Approaches, which make use of data from external sensors, like RGB [2] and depth images [3], rely on learned background models. This facilitates the feature extraction from data generated by persons and improves the classification output. However, because the dynamics of the scenery are changing very quickly, such background models can not be generated on a mobile platform. Other approaches [4, 5] use estimated skeletons of the user to estimate the state of seated. The drawback of these methods are, that they rely on the Kinect SDK, which was developed for gaming purpose in a static sensor setup. Unfortunately, our tests revealed that even small vibrations cause the skeleton estimation to stop. At this time, such approaches are unlikely to work on a mobile platform and further investigations of the Kinect sensor are needed. In this paper, we will present a simple approach which is based on a time series of height estimations generated from RGB images. It is able to estimate the seated state of a person on a mobile robot and can be applied on every platform with a minimalistic sensor setup which provides images from at least one camera.

2.2 Approaching Persons

The publications regarding approaching a person can roughly be divided in investigative work analyzing how to best approach a person by using Wizard-of-Oz-like experiments and more technical work presenting how to realize an approach behavior.

The approach parameters primarily considered in investigative publications are the relative position and distance to the human partner, as well as the driving speed. To the best of our knowledge, no work has been published yet investigating the approaching parameters for presenting a display to a person. The scenario most similar to ours is a handing over task performed by a robot with arms [6]. In these works, the participants preferred to be approached from the front with a stopping distance of about 0.5 m. However, these parameters can not be generally transferred to different settings, since they are strongly dependent on the robots appearance, behavior and the interaction task [7]. We thus propose a parametric approach able to adapt to different requirements of relative position, stopping distance and speed.

Most technical publications consider the scenario of approaching moving persons to intercept them and start a conversation. In [8], an approaching pose is inferred from a graph-based human movement prediction model to intercept the moving person. The model was learned from real trajectories. Instead of a learned model, [9] used a linear movement model to find the best approaching pose in an outdoor scenario. The robot motion planning is formulated as optimal control problem and solved approximately. In contrast to our scenario, where the robot has to get close enough to allow a seated person to access the touch display, moving person scenarios have a weaker restriction on the reached distance, since a person already in motion is also able to close the distance to the robot by her/himself.

In [10], a framework for approaching persons is described. As motion planer the Dynamic Window Approach [11] is used. Among other scenarios a seated person is approached. However, the focus lies in finding poses which do not violate the personal space. In [12], another work focusing on the personal space but approaching standing persons is presented. Since our goal is making the robot's display reachable, it is inevitable but acceptable to enter the personal space in this situation.

Recently, learning from expert demonstration is used to teach a robot to approach a person. In [13], Inverse Reinforcement Learning [14] is utilized to learn a cost function to guide the robot motion planner. This work show promising results but was only evaluated in simulations or restricted environments leaving the generalization ability unanswered. Furthermore, training data need to be generated by an expert which can be a cumbersome work for real-world scenarios.

3 Robot Training Platform ROGER

Our robot is a customized SCITOS[2] platform with a relative small footprint of 45 x 55 cm and a height of 1.5 m (Fig. 2). Its differential drive is balanced by a caster and can reach driving speeds of up to 1.4 m/s. For perception of the environment and communication, the robot is equipped with various sensors and two touch displays. The displays are mounted at different heights allowing the patient to operate while seated or standing. The sensors consist of two SICK laser range finders covering a 360° field of view, two Asus RGB-D cameras directed forward and a panoramic

[2] http://bit.ly/MLSCITOS

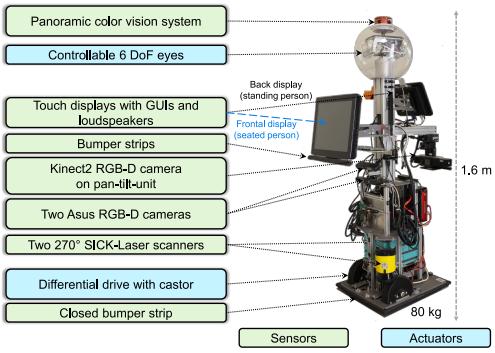


Figure 2 Sensors and actors of our robot training platform.

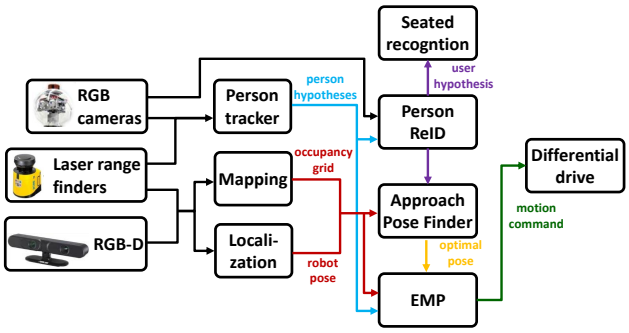


Figure 3 Schematic overview of the processing pipeline of our approach.

color vision system on the top of the head. Additionally, a Kinect2 facing backward is mounted on a pan-tilt unit allowing to freely move the line of vision.

For person perception, a multi-hypothesis and multi-cue tracking system is used. The tracker is based on a 7D Kalman filter and is able to estimate the positions and velocities from different detection cues (Sec. 4.1). To guarantee a gait training, tracking alone is insufficient. The patient also has to be re-identified among all tracked persons. For re-identification, a metric-learning approach with color and texture features is used [15].

To safely navigate in dynamic environments, the subtasks of localization, obstacle detection and motion planning need to be addressed. At their cores, our localization and mapping system are an adaptive Monte Carlo approach respectively an occupancy grid mapping approach, but are generically designed to process both 2D laser scans and 3D informations [16, 17]. For generating motion commands, an objective-based motion planner using evolutionary optimization is utilized (Sec. 5.2).

The complete robotic system was developed with MIRA [18]. A detailed overview of all implemented interaction and navigation skills, along with behaviors of our robot coach is given in [19].

4 Recognition of Seated Persons

4.1 Person Tracking and Reidentification

The presented approach to detect the seated state of a person is integrated in the person tracking- and re-identification framework described in [20] (see Fig. 3). Data from multiple sensors are processed by different detection cues. We use the 2D laser-based detector from [21] to detect persons using walking aids, which is a crucial precondition in the target scenario. Additionally, the part-based approach from [22] is used for detections from the panorama camera system. Each detection from the different cues is transformed into global coordinates for further processing in an consistent reference frame. A Kalman-filter based tracker assigns detections to already existing person hypotheses or create new ones if a certain distance to these hypotheses is exceeded (0.8m in our case). Afterwards, a person re-identification module determines which one of the hypotheses corresponds to the user who is interacting with the robot. This user hypothesis is than used by the seated recognition module, to assign detections for the estimation of the seated-/ standing state of the user, which is described in the following.

4.2 Height-Based Seated Estimation

4.2.1 Height Estimation

As mentioned, the estimation of a person seated in the surroundings of the robot, can be done with any person detector, which is able to estimate the height of a person. In our case, the detector from [22] delivered the best rates in the target scenario for the person detection task. This detector delivers rectangles around the person in a given color image. From this rectangle, a pinhole camera model is used, to create a 3D detection in the image coordinate system. The distance D of the detected person to the camera is estimated by the intercept theorem (see eq. 1), where O_w is the estimated real object width, R_w is the width of the detection rectangle and f is the focal length of the camera model. The resulting scalar value D is then multiplied with the a vector from the origin of the coordinate system to the center of the detection rectangle in the image plane. This yields a 3D point which corresponds to the center of the person in the camera coordinate system. This point is transformed into global coordinates using the extrinsic parameters of the camera. Afterwards, an offset is added to the z-coordinate, to align this 3D detection to the head of the person. This z-coordinate represents the current estimation of the persons height and is further processed to learn this height while standing and to determine the height when a person is seated.

$$D = \frac{O_w \cdot f}{R_w} \tag{1}$$

4.2.2 Standing Height Learning

To classify the seated state of a person using the current height estimation, first of all the person's height while standing has to be learned. For this task we use the assumption, that when a person is moving the current height represents the standing height. Therefore, when the current user hypothesis' movement speed is larger than $0.5\mathrm{m/s}$, we collect the z-coordinates of the incoming detections. To

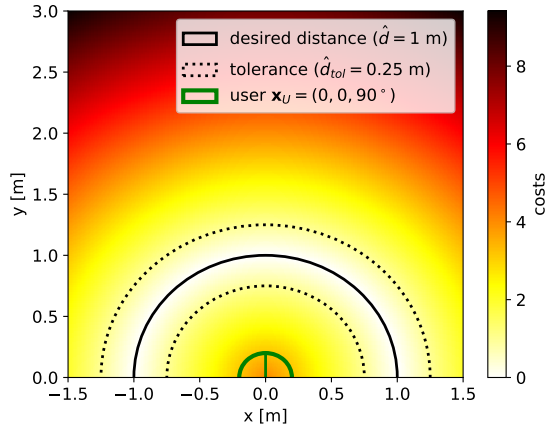


Figure 4 Visualization of the desired distance cost by keeping the user pose $\mathbf{x}_U$ constant and varying the robot pose $\mathbf{x}_R$. If the distance of a robot pose to the user pose equals the desired distance, the resulting costs will be zero. Deviations from the desired distance result in increasing costs where poses outside of the tolerance will take cost value greater than 1.

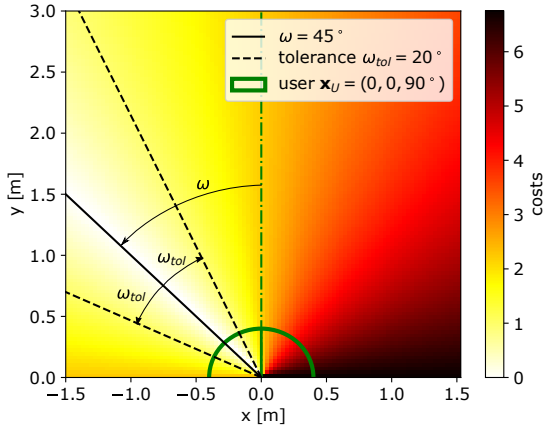


Figure 5 Visualization of the relative position cost by keeping the user pose $\mathbf{x}_U$ constant and varying the robot pose $\mathbf{x}_R$. The relative position cost defines a circular sector where the robot pose preferably should be. This visualization is also valid for the heading cost, since the Eq. 6 of the heading cost can be rearranged from Eq. 4 by reversing the role of the user pose and the robot pose, and setting $\omega = 0$, $\omega_{tol} = \lambda_{tol}$.

filter outliers, the median of these z-coordinates is used to estimate the standing height H_s of the user.

4.2.3 Seated Hysteresis

$$S_t = \begin{cases} UNKNOWN & \text{if } t = 0 \\ STANDING & \text{if } H_c > T_{standing} \\ SEATED & \text{if } H_c < T_{seated} \\ S_{t-1} & \text{otherwise} \end{cases} \tag{2}$$

The current standing/seated state is estimated using the current height H_c of the user. It is calculated like the standing height, but without the restriction of a moving hypothesis. Since the seated state estimation has to be available as soon as possible, a smaller filter horizon for outlier reduction is applied. Therefore, the mean of a sliding window of the last ten height detections is used to create the current height estimation H_c. Since outliers may be correlated, H_c is then fed into a dual threshold hysteresis S_t at time t (see Eq. 2) . The outcome of this hysteresis is the current seated/standing state of the user, where the thresholds T are determined by the estimated standing height H_s, where $T_{standing} = H_s - 10$ cm and $T_{seated} = H_s - 30$ cm. In Fig. 7, a visualization of the hysteresis behavior can be seen.

5 Approaching Seated Persons

In this section, our technique for approaching a seated person is presented. To accomplish this task, the problem is decomposed into finding an appropriate interaction pose and navigating to the found pose. It is noted that the proposed technique is also applicable for standing persons.

5.1 Finding Approriate Approaching Poses

As stated in Sec. 2.2, relevant parameters for determining good approaching poses are the relative position and the distance to the interaction partner. To present the display, the robot's heading direction also has to be considered. Furthermore, in a dynamic environment, obstacles might stand in the vicinity of the user, restricting the possible approaching poses. To find poses respecting these restrictions, we formulate a multi-objective optimization problem over the search space of robot poses $\mathbf{x} = (x, y, \phi)$ in a planar world. Each objective is a cost function representing one of the mentioned restrictions. Given the current obstacle configuration as grid map $\mathbf{m}$, the robot's current pose $\mathbf{x}_R$ and the user's pose $\mathbf{x}_U$ to be approached, the global cost function

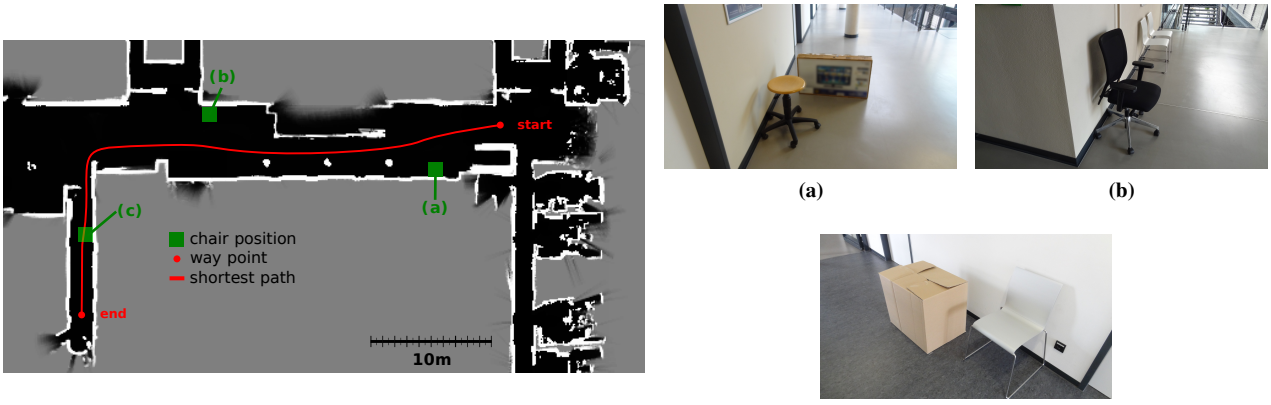
$$f(\mathbf{x}|\mathbf{m}, \mathbf{x}_R, \mathbf{x}_U) = w_1 \cdot f_{position} + w_2 \cdot f_{dist} + w_3 \cdot f_{heading} + w_4 \cdot f_{obstacles} \tag{3}$$

is defined as the weighted sum of these objectives. The given inputs $\mathbf{m}$, $\mathbf{x}_R$ and $\mathbf{x}_U$ are generated by the mapping, localization respectively person re-identifiction module (Fig. 3). Since the objectives are not differentiable, Particle Swarm Optimization [23] is utilized to find near-optimal solutions. In the following, the objectives are explained in more detail.

Relative Position

$$f_{position}(\mathbf{x}|\mathbf{x}_U, \omega, \omega_{tol}) = \frac{|\alpha_{\mathbf{x} \to \mathbf{x}_U^{\omega}}|}{\omega_{tol}} \tag{4}$$

defines the circular sector where the robot should preferable reside. To freely control the direction of this circular sector, $\mathbf{x}_U^{\omega}$ is determined by changing the orientation of $\mathbf{x}_U$



(a) (b) (c)

Figure 6 The test track at our lab's building and three chairs (stool, office chair, normal chair) set up along the track. The height of the chair seats was kept constant at 0.45 m. The shortest path from the starting to the end was approx. 75 m long. Each trial took about 15 minutes to finish.

by ω. $\alpha_{\mathbf{x}\to\mathbf{x}_U^\omega}$ can now be defined as the angular coordinate of $\mathbf{x}$ in the polar coordinate system centered at $\mathbf{x}_U^\omega$. ω_{tol} is a parameter controlling the tolerance for deviating from the circular sector. For a visualization of this cost, see Fig. 5.

Desired Distance

$$f_{dist}(\mathbf{x}|\mathbf{x}_U,\hat{d},\hat{d}_{tol}) = \frac{|\hat{d}-\delta(\mathbf{x},\mathbf{x}_U)|}{\hat{d}_{tol}} \tag{5}$$

determines at what distance relative to the user the robot should stop approaching. In this cost, $\delta(\mathbf{x},\mathbf{x}_U)$ defines the euclidean distance of the planar position of $\mathbf{x}$ and $\mathbf{x}_U$, whereas $\hat{d}$ and $\hat{d}_{tol}$ are parameters controlling the desired distance respectively the error tolerance. See Fig. 4 for a visualization of the desired distance cost.

Heading

$$f_{heading}(\mathbf{x}|\mathbf{x}_U,\lambda_{tol}) = \frac{|\alpha_{\mathbf{x}_U\to\mathbf{x}}|}{\lambda_{tol}} \tag{6}$$

defines a cost forcing the robot to orientate its front side along the line of sight to the user. Here, $\alpha_{\mathbf{x}_U\to\mathbf{x}}$ is the angular coordinate of the user's pose $\mathbf{x}_U$ in the polar coordinate system centered at $\mathbf{x}$, and λ_{tol} a parameter controlling the tolerance.

Obstacles

$$f_{obstacles}(\mathbf{x}|\mathbf{m},\mathbf{x}_R,d_{max}) = \frac{\Delta(\mathbf{x}|\mathbf{m},\mathbf{x}_R)}{d_{max}} \tag{7}$$

determines the reachability of the pose $\mathbf{x}$ from the current robot pose $\mathbf{x}_R$ given the obstacle configuration $\mathbf{m}$ by using the distance function $\Delta(\mathbf{x}|\mathbf{m},\mathbf{x}_R)$ calculated with Dijkstra's algorithm [24] and normalized by the parameter d_{max}. The distance function is calculated with the current robot pose as the source and obstacles being marked as unreachable. This allows the robot to find the closest position minimizing the driving time.

5.2 Evolutionary Motion Planning

After an approaching pose was found, the Evolutionary Motion Planner (EMP), developed at our lab, is used for navigation. EMP is a versatile motion planner enabling the optimization of motion command sequences under various combinations of criteria. Since we do not restrict these criterion nor the motion sequences, the optimization is high-dimensional and generally not solvable in closed form. Therefore, the EMP relies on an evolutionary algorithm for optimization.

Evolutionary algorithms regard solutions as individuals. In our case, an individual equals a motion command sequence. A sequence can be expressed as a vector $(c^{(1)},\ldots,c^{(T)})$ with T being the planning horizon and $c^{(t)}$ the command specific to the robot's drive. Starting with an initial population of command sequences, the population is improved over multiple iterations with operations inspired by biological evolution mechanisms. An iteration consists of, first, assigning a fitness value to each individual of the current population according to the given objective function, second, in a reproduction process two parent individuals are selected and combined to create a new individual, whereby individuals with better fitness have a higher chance to be selected. These new individuals form a new population and replace the current population for the next iteration.

For creating new individuals, the ideas of genetic recombination and mutations are used. A new individual is recombined by copying the motion command for each time step from one of the selected parent. In each time step, the chosen parent is switched with a predefined probability. For more diversity, the new individual is further mutated by systematically perturbing its elements with a normal distributed value. More details on EMP can be found in [25].

The robot's navigation behavior is controlled by the given objective function. Similar to finding approaching poses, we decompose the objective function into multiple sub-objectives. For approaching a user, we utilize the objectives presented in [25], which realize a goal oriented

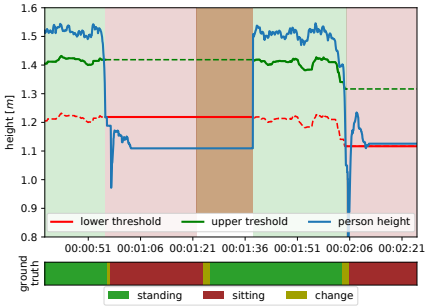


Figure 7 Time profile of height estimation (ground-to-nose) and classifications results for a 1.6 m tall test subject. The ground truth conforms with our results except for the transition from standing to seated since during the brown marked time the subject left the sensor's field of view.

movement while avoiding obstacles and keeping a social acceptable space to detected bystander. Furthermore, since the driving speed is a relevant parameter when approaching (Sec. 2.2), we extend the base objectives with a speed objective limiting the robot's driving speed. Overall, the used objectives comprises of

1. a *path and heading objective* responsible for the movement to the goal by approaching the minimum in a globally planned navigation function (using E* planner [26]) and turning the robot towards a given goal orientation when in proximity to the goal position,
2. a *direction objective* preferring forward motion accounting for the motor's limitation with a slower speed when driving backward,
3. a *distance objective* for avoiding collisions with static and dynamic obstacles,
4. a *personal space objective* to keep distance to people in the close proximity of the robot by predicting their movements with a linear model,
5. a *speed objective* denying motion sequences which exceed a given speed limit.

6 Experimental Evaluations

6.1 Experimental Setup

To evaluate the performance of our approach, five subjects were guided by the robot through a test track installed on the hallways of our lab's building (Fig. 6). Similar to the hospital environment, chairs were set up at walls along the track. In total, three chairs of different types were used. Around two chairs, additional obstacles were placed to mimic the confined hospital space. During a trial, the subjects were asked to use crutches. At each chair, the subject had to sit down and place the walking aid aside, so that the robot could approach to present the front touch display. The place where the crutches were placed could freely be chosen by the subject. The state transition between seated and standing was manually triggered by a staff member from a distance, which was also recorded as ground truth to evaluate our seated estimation. After the robot finished approaching, the poses relative to the chair have been manually recorded, and the test subject was asked if s/he could operate the touch screen. Each subject was guided from the start to end of the track and back. Overall, 30 interaction attempts (10 for each chair) consisting of sitting down and approaching were recorded.

Relative Position	$\omega = 0$ ($\omega = 40°$), $\omega_{tol} = 60°$
Desired Distance	$\hat{d} = 0.4$ m, $\hat{d}_{tol} = 1.0$ m
Heading	$\lambda_{tol} = 60°$
Obstacles	$d_{max} = 5.0$ m
Weights	$w_1 = w_3 = 5.0$, $w_2 = 15$, $w_4 = 10$

Table 1 The used parameters of the approach cost function.

6.2 Seated Estimation

To evaluate the seated estimation in an objective manner, we counted the flanks of the seated hysteresis output as they would trigger the application. Therefore, an event of sitting down is counted as correctly classified, if the output of the hysteresis threshold jumps from standing to seated in a range of $\pm$1s from the labeled ground truth. Otherwise, this event counts as not detected. A flank from standing to seated which happens outside this range is counted as false alarm. During the experiments a total of 30 events of sitting down occurred. A true positive rate of 70% was achieved, where 21 of the events were classified correctly while 9 were missed. Most of the missed events were caused by the absence of detections in hard to classify person appearances, especially when the person was sitting down while the angle to the camera center was relatively large. In contrast, 9 false alarms happened within 60 minutes, which corresponds to one false alarm every 7 minutes approximately. These false alarms mainly occurred when the head of the person was not included in the detection rectangle of the person detector. This situation is likely to happen when the user is close to the robot, so the sensor's field of view could not cover the full person appearance.

6.3 Approaching Performance

For the experiments. we set the parameters of the cost function for finding approaching poses (Sec. 5.1) to the values given in Table 1. To test the ability for approaching persons from different sides, ω was set for the midway chair (Fig. 6b) to 40°. Furthermore, we limited the driving speed during an approach to 0.6 m/s, and the distance and angular tolerance for reaching the approaching pose was set to 0.25 m respectively 10°.

The manually recorded approaching poses are depicted in Fig. 8. In 23 out of 30 approaches (76 % success rate), the test subjects rated the robot's stopping pose after the approaching finished to be sufficient in distance and heading orientation for operating the front touch display. Only a slight forward leaning was sometimes necessary to reach the display. On average, a distance to the perceived person position of 0.47 m ($\pm$0.11) and an angular difference to the optimal heading direction of 10.26° ($\pm$6.97) was reached. In the 7 unsuccessful cases (marked as red dots), the partic-

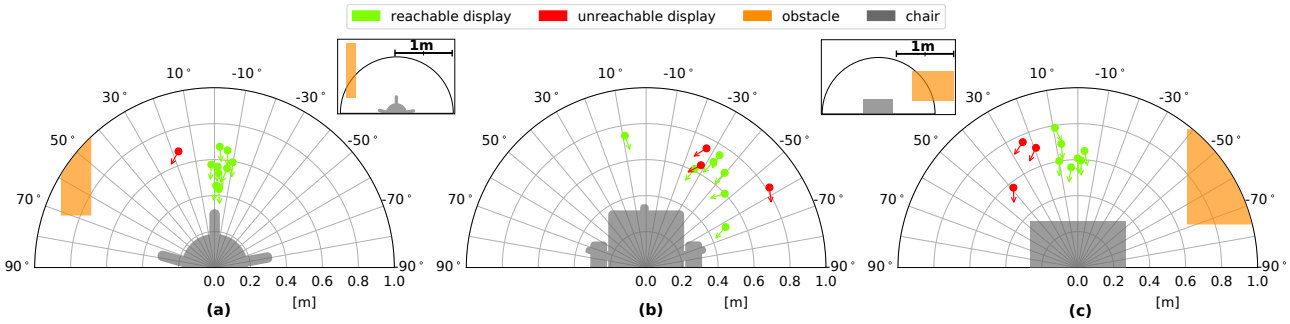


Figure 8 The manually recorded stopping poses (green and red dots with arrows, see above) of the robot after approaching each chair of the experimental setup (Fig. 6). The position (dot) of a pose marks the center point of the front display projected on the ground while the direction (arrow) is perpendicular to the front display. For the chairs (a) and (b) a bigger overview of the vicinity, containing the whole obstacles, is depicted in a mini-map.

ipants rated the distance as sufficient, but the heading orientation as unsuitable to operate the display. In these cases, the average distance to the perceived person position was 0.59 m (± 0.09), whereas the average angular difference to the optimal heading direction was 48.27° (± 15.33). Further investigation showed, that in these cases the robot was either incorrectly localized (position and orientation) or could not turn to the proposed heading orientation without colliding with an obstacle. The latter case occurs, when the robot already reached the proposed position, but was standing to close to the seated person.

During all the approaching attempts, the robot's driving speed did not exceeded the defined value of 0.6 m/s. The average speed was 0.45 m/s (± 0.1). The end poses in Fig. 8b indicate that most of them resides in the given angular limit. Only one pose deviated and appeared in front of the chair. This was caused by a test subject placing his foot as an obstacle in the optimal angular sector. Thus, the optimal region could not be reached, and the robot found the next best pose.

7 Conclusion

In this paper, we have presented an approach applicable on any mobile platform equipped with cameras to detect the seated state of a person interacting with a rehabilitation robot using a learned height of a person over several time steps. Furthermore, a method for approaching a seated person to make the robot's display operable is presented. Approaching is realized by heuristically solving two optimization for finding good approaching poses and motion commands for driving to the approaching pose.

While the proposed approach for seated estimation is theoretically sound, the experiments showed that it is strongly dependent on the sensor setup and the performance of the used person detector. The results show an acceptably high true positive rate but with too many false alarms. These false alarms were mainly caused by the small vertical field of view of the used camera, so that persons standing near the robot could not be perceived completely. To be applicable in a real-world setup, better person detectors and cameras with a wider field of view need to be used. Furthermore, independent from the Kinect SDK we plan to use the high-resolution color image of the Kinect2 to extract skeletal data to improve the seated estimation.

The results of the proposed method for approaching persons show a good success rate. The combination of particle swarm optimization with the described cost function were able to find appropriate approaching poses. However, in some cases, shortcomings in the localization module and driving behavior led to an incorrect alignment of the robot to the found pose. For an autonomous application, further improvements are needed. To improve the localization accuracy, additional cues from color images should be used. The mentioned disadvantageous driving behavior originates from the used sub-objectives rather than the EMP. Thus, aside from the maximum speed the driving behavior should be further adapted to the approaching task. To this end, a driving direction to the approaching pose could be imposed by a new sub-objective. This would avoid the turning behavior near obstacles.

Literature

[1] F. Attal *et al.*, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31 314–31 338, 2015.

[2] A. Dubois and F. Charpillet, "Human activities recognition with rgb-depth camera using hmm," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2013, pp. 4666–4669.

[3] E. Akagündüz *et al.*, "Silhouette orientation volumes for efficient fall detection in depth videos," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 756–763, May 2017.

[4] T. H. Nguyen and H. A. Trinh, "Pca-svm algorithm for classification of skeletal data-based eigen postures," in *American Journal of Biomedical Engineering*, vol. 6, no. 5, 2016.

[5] L. Xia *et al.*, "View invariant human action recog-

nition using histograms of 3d joints," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 20–27.

[6] K. L. Koay *et al.*, "Social roles and baseline proxemic preferences for a domestic service robot," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 469–488, Nov 2014.

[7] P. A. M. Ruijten and R. H. Cuijpers, "Stopping distance for a robot approaching two conversating persons," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2017, pp. 224–229.

[8] D. Brščić *et al.*, "Do you need help? a robot providing information to people who behave atypically," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 500–506, April 2017.

[9] D. Carton *et al.*, "Proactively approaching pedestrians with an autonomous mobile robot in urban environments," in *Experimental Robotics: The 13th International Symposium on Experimental Robotics*, 2013, pp. 199–214.

[10] X. T. Truong and T. D. Ngo, ""To Approach Humans?": A unified framework for approaching pose prediction and socially aware robot navigation," *IEEE Transactions on Cognitive and Developmental Systems*, vol. PP, no. 99, pp. 1–1, 2017.

[11] D. Fox *et al.*, "The dynamic window approach to collision avoidance," *IEEE Robotics Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.

[12] J. Kessler *et al.*, "Approaching a person in a socially acceptable manner using a fast marching planner," in *Intelligent Robotics and Applications*, 2011, pp. 368–377.

[13] O. A. I. Ramírez *et al.*, "Robots learning how and where to approach people," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2016, pp. 347–353.

[14] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the Twenty-first International Conference on Machine Learning*, ser. ICML '04. New York, NY, USA: ACM, 2004, pp. 1–.

[15] M. Eisenbach *et al.*, "User recognition for guiding and following people with a mobile robot in a clinical environment," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 3600–3607.

[16] E. Einhorn and H.-M. Gross, "Generic 2d/3d slam with ndt maps for lifelong application," in *European Conference on Mobile Robots (ECMR)*, 2013, pp. 240–247.

[17] Th. Schmiedel *et al.*, "Iron: A fast interest point descriptor for robust ndt-map matching and its application to robot localization," in *IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 3144–3151.

[18] E. Einhorn *et al.*, "MIRA - middleware for robotic applications," in *IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012, pp. 2591–2598.

[19] H.-M. Gross *et al.*, "Mobile robot companion for walking training of stroke patients in clinical post-stroke rehabilitation," in *Proc. ICRA*, 2017, pp. 1028–1035.

[20] T. Wengefeld *et al.*, "May i be your personal coach? bringing together person tracking and visual re-identification on a mobile robot," in *Int. Symposium on Robotics (ISR)*. VDE, 2016, pp. 141–148.

[21] Ch. Weinrich *et al.*, "People detection and distinction of their walking aids in 2D laser range data based on generic distance-invariant features," in *IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2014, pp. 767–773.

[22] P. F. Felzenszwalb *et al.*, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, Sept 2010.

[23] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. ICNN*, vol. 4, 1995, pp. 1942–1948.

[24] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[25] St. Mueller *et al.*, "Local real-time motion planning using evolutionary optimization," in *Proc. TAROS*, vol. 10454, 2017, pp. 211–221.

[26] R. Philippsen and R. Siegwart, "An interpolated dynamic navigation function," in *Proc. IEEE ICRA*, 2005, pp. 3782–3789.