Research Article

Hobbit: Providing Fall Detection and Prevention for the Elderly in the Real World

Markus Bajones [ID],[1] David Fischinger,[1] Astrid Weiss,[1] Daniel Wolf,[1] Markus Vincze,[1] Paloma de la Puente,[2] Tobias Körtner,[3] Markus Weninger,[3] Konstantinos Papoutsakis,[4] Damien Michel,[4] Ammar Qammaz,[4] Paschalis Panteleris,[4] Michalis Foukarakis,[4] Ilia Adami,[4] Danai Ioannidi,[4] Asterios Leonidis,[4] Margherita Antona,[4] Antonis Argyros,[4] Peter Mayer,[5] Paul Panek,[5] Håkan Eftring,[6] and Susanne Frennert[6]

[1]*Automation and Control Institute (ACIN), TU Wien, Vienna, Austria*
[2]*Universidad Politécnica de Madrid, Madrid, Spain*
[3]*Akademie für Altersforschung am Haus der Barmherzigkeit, Vienna, Austria*
[4]*Institute of Computer Science, FORTH, Heraklion, Greece*
[5]*Institute for Design and Assessment of Technology, TU Wien, Vienna, Austria*
[6]*Department of Design Sciences, Lund University, Lund, Sweden*

Correspondence should be addressed to Markus Bajones; markus.bajones@tuwien.ac.at

We present the robot developed within the Hobbit project, a socially assistive service robot aiming at the challenge of enabling prolonged independent living of elderly people in their own homes. We present the second prototype (Hobbit PT2) in terms of hardware and functionality improvements following first user studies. Our main contribution lies within the description of all components developed within the Hobbit project, leading to autonomous operation of 371 days during field trials in Austria, Greece, and Sweden. In these field trials, we studied how 18 elderly users (aged 75 years and older) lived with the autonomously interacting service robot over multiple weeks. To the best of our knowledge, this is the first time a multifunctional, low-cost service robot equipped with a manipulator was studied and evaluated for several weeks under real-world conditions. We show that Hobbit's adaptive approach towards the user increasingly eased the interaction between the users and Hobbit. We provide lessons learned regarding the need for adaptive behavior coordination, support during emergency situations, and clear communication of robotic actions and their consequences for fellow researchers who are developing an autonomous, low-cost service robot designed to interact with their users in domestic contexts. Our trials show the necessity to move out into actual user homes, as only there can we encounter issues such as misinterpretation of actions during unscripted human-robot interaction.

1. Introduction

While socially assistive robots are considered to be potentially useful for society, they can provide the highest value to older adults and homebound people. As reported in [1], future robot companions are expected to be

(1) strong machines that can take over burdensome tasks for the user,

(2) graceful and soft machines that will move smoothly and express immediate responses to their users,

(3) sentient machines that offer multimodal communication channels and are context-aware and trustable.

More and more companies and research teams present service robots with the aim of assisting older adults (e.g., Giraff (http://www.giraff.org), Care-O-Bot (http://www.care-o-bot.de), and Kompai (https://kompai.com)) with services such as entertainment, medicine reminders, and video telephony. Requirement studies on needs and expectations of older adults towards socially assistive robots [2] indicate that they expect them to help with household chores (e.g.,

cleaning the kitchen, bath, and toilet), lifting heavy objects, reaching for and picking up objects, delivering objects, and so forth. However, most of these tasks cannot satisfyingly be performed by state-of-the-art robotic platforms; hardly any companion robot fulfills the requirements mentioned above and only very few robots entered private homes of older adults so far. One of the biggest challenges is offering sufficient useful and social functionalities in an autonomous and safe manner to achieve the ultimate goal of prolonging independent living at home. The ability of a robot to interact autonomously with a human requires sophisticated cognitive abilities including perception, navigation, decision-making, and learning. However, research on planners and cognitive architectures still faces the challenge of enabling flexibility and adaptation towards different users, situations, and environments while simultaneously being safe and robust. To our conviction, for successful long-term human-robot interaction with people in their private homes, robotic behavior needs to be above all safe, stable, and predictable. During our field trials, this became increasingly evident, as the users failed to understand the robot's behavior during some interaction scenarios.

In this article, we present the Hobbit PT2 platform, referred to in the remainder of this article as Hobbit. A former version of Hobbit has been presented in detail in [3]. Hobbit is a socially assistive robot that offers useful personal and social functionalities to enable independent living at home for seniors. To the best of our knowledge, the Hobbit trials mark the first time a social service robot offering multifunctional services was placed in users' homes, operated autonomously and whose usage was not restricted by a schedule or any other means. The main contribution of this paper is twofold. First, we give a description of the hardware that is based on improvements derived from the first user trials on the previous version of Hobbit. Second, we describe the implemented functionality and its integration into the behavior coordination system. The building blocks of the behavior coordination system are based on a set of hierarchical state-machines implemented using the SMACH framework [4]. Each behavior was built upon simpler building blocks, each responsible for one specific task (e.g., speech and text output, arm movements, and navigation) to add up to the complex functionalities presented in Sections 3.3 and 4. Finally, we present the lessons learned from the field trials in order to support fellow researchers in their developments of autonomous service robots for the domestic environment. We evaluated Hobbit during 371 days of field trials with five platforms with older adults in their private homes in Austria, Greece, and Sweden. However, details on the field trials will be published elsewhere.

The paper proceeds as follows. Section 2 reflects on relevant related work on behavior coordination for social service robots and on studies of such robots outside of the laboratory environment. In Section 3, we give an overview on the project vision for Hobbit and its historical development up to the Hobbit PT2 platform, followed by a detailed description of its hardware and interaction modalities. Section 4 presents the behavior coordination system. We outline how we developed the interaction scenarios and transferred them into an implementable behavior concept. Section 5 presents an overview on the field trials. Lessons learned from the development and testing of Hobbit and a summary and conclusions are provided in Sections 6 and 7.

2. Related Work

Moving towards autonomous service robots, behavior coordination systems constitute an important building block to fulfill the requirements of action planning, safe task execution, and integration of human-robot interaction. HAMMER from Demiris and Khadhouri [5] is built upon the concept of using multiple forward/backward control loops, which can be used to predict the outcome of an action and compare this against the actual result of the action. Through this design, it is possible to choose the action with the highest probability of reaching the desired outcome, which has successfully been used in a collaboratively controlled wheelchair system [6], in order to correct the user's input to avoid an erroneous situation. Cashmore et al. [7] introduced ROSPlan, a framework that uses a temporal planning strategy for planning and dispatching robotic actions. Depending on the needs, a cost function can be optimized for planning in a certain manner (e.g., time- or energy-optimized). However, the constructed plan is up until now only available as a sequence of executed actions and observed events, but no direct focus is put on the human, besides modeling the user as means to acquire some event (e.g., moving an object from one location to another). Mansouri and Pecora [8] incorporate temporal and spatial reasoning in a robot tasked with pick and place in environments suited for users. In the context of ALIAS, Goetze et al. [9] designed their *dialogue manager* for the tasks of emergency call, a game, e-ticket event booking, and the navigation as state-machines. However, there are still significant research challenges regarding how to incorporate humans into the planning stages and decide when the robot needs to adapt to the user instead of staying with the planned task.

Most of those behavior coordination and planning systems treat the human as an essential part of the system [6] (e.g., for command input) and rely on the user to execute actions planned by the coordination system [10]. Such systems only work under the precondition that the robot will execute a given task for the user independently of the user input [8]. A crucial aspect, however, to successfully integrate a multifunctional service robot into a domestic environment is that it needs not only to react to user commands but also to proactively offer interaction and adapt to user needs (e.g., the user wanting a break from the robot or a proactive suggestion for an activity they could perform together). Our proposed solution is based on state-machines, which reflect turn-taking in the interaction, providing adaptations within certain states (e.g., voice dialogues) or situations (e.g., user approach). We integrated the possibility not only to handle robot-driven actions on a purely scheduled basis but also to adapt this scheduling and actions based on the user's commands.

2.1. State of the Art: Robotic Platforms. According to a study conducted by Georgia Tech's Healthcare Robotics Lab, people

with motor impairment drop items on average 5.5 times a day. Their small tele-operated Dusty (http://pwp.gatech.edu/hrl/project_dusty/) robots are developed for that purpose: picking up objects from the floor, which they achieve with a scoop-like manipulator. Cody, a robotic nurse assistant, can autonomously perform bed (sponge) baths. Current work focuses on GATSBII (http://www.robotics.gatech.edu), a willow Garage PR2, as a generic aid for older adults at home. The Care-O-Bot research platforms developed at the Fraunhofer Institute (IPA) are designed as general purpose robotic butlers, with a repertoire from fetching items to detecting emergency situations, such as a fallen person. Also from Fraunhofer is Mobina (https://www.ipa.fraunhofer.de/de/referenzprojekte/MobiNa.html), a small (vacuum-sized) robot specifically performing fallen person detection and video calls in emergency. Carnegie Mellon University's HERB (https://personalrobotics.ri.cmu.edu/) is another general-purpose robotic butler. It serves as the main research platform at the Personal Robotics Lab, which is part of the Quality of Life Technology (QoLT) Center. KAIST in Korea has been developing their Intelligent Sweet Home (ISH) smart home technology including intelligent wheelchairs, intelligent beds, and robotic hoists [11]. Their system also employs the bimanual mobile robot Joy to act as an intermediary between these systems and the end user. Robotdalen (http://www.robotdalen.se), a Swedish public-private consortium, has funded the development of needed robotic products such as Bestic (http://www.camanio.com/en/products/bestic/), an eating device for those who cannot feed themselves; Giraff, a remote-controlled mobile robot with a camera and monitor providing remote assistance and security; or TrainiTest, a rehabilitation robot that measures and evaluates the capacity of muscles and then sets the resistance in the robot to adapt to the users' individual training needs. Remote presence robots have recently turned up in a variety of forms, from simple Skype video chats on a mobility platform (Double Robotics (https://www.doublerobotics.com/)) to serious medical assistance remote presence robots such as those provided by the partnership between iRobot and InTouch Health (https://www.intouch-health.com/about/press-room/2012/InTouch-Health-and-iRobot-to-Unveil-the-RP-VITA-Telemedicine-Robot.html), Giraff, and VGo Communications' postop pediatric at-home robots (http://www.vgocom.com/) for communication with parents, nurses, doctors, and patients.

Another class of robots aims more specifically at well-being of older adults. The recently completed FP7 project M obiserv (https://cordis.europa.eu/project/rcn/93537_en.html) aimed to develop solutions to support independent living of older adults as long as possible, in their home or in various degrees of institutionalization, with a focus on health, nutrition, well-being, and safety. These solutions encompass smart clothes for monitoring vital signs, a smart home environment to monitor behavioral patterns (e.g., eating) and detect dangerous events, and a companion robot. The robot's main role is to generally activate, stimulate, and offer structure during the day. It also reminds its user of meals, medication, and appointments and encourages social contacts via video calls. The US NSF is currently running the Socially Assistive Robotics project (https://www.nsf.gov/awardsearch/showAward?AWD_ID=1139078) with partners Yale, University of Southern California, MIT, Stanford, Tufts, and Willow Garage. Their focus is on robots that encourage social, emotional, and cognitive growth in children, including those with social or cognitive deficits. The elder care robot Sil-Bot (http://www.roboticstoday.com/robots/sil-bot) developed at the Center for Intelligent Robotics (CIR) in Korea is devised mainly as an entertainment robot to offer interactive games that have been codeveloped with Seoul National University Medical Center specifically to help prevent Alzheimer's disease and dementia. Sil-Bot is meant to be a companion that helps encourage an active, healthy body and mind. Its short flipper-like arms do not allow for actual manipulation. Another public-private partnership is the EC-funded CompanionAble project (http://www.companionable.net/), which created a robotic assistant for the elderly called Hector. The project integrates Hector to work collaboratively with a smart home and remote control center to provide the most comprehensive and cost-efficient support for older people living at home.

Hoaloha Robotics (http://www.hoaloharobotics.com/) in the United States are planning to bring their elder care robot to market soon. Based on a fairly standard mobile platform offering safety and entertainment, they focus on an application framework that will provide integration of discrete technological solutions like biometric devices, remote doctor visits, monitoring and emergency call services, medication dispensers, online services, and the increasing number of other products and applications already emerging for the assistive care market. Japan started a national initiative in 2013 to foster development of nursing care robots and to spread their use. The program supports 24 companies in developing and marketing their elderly care technologies, such as the 40 cm tall PALRO conversation robot (https://palro.jp/) that offers recreation services by playing games, singing, and dancing together with residents of a care facility. Another example is the helper robot by Toyota, which is mostly remotely controlled from a tablet PC. Going specifically beyond entertainment capabilities, Waseda University's Twendy One (http://www.twendyone.com) is a sophisticated bimanual robot that provides human safety assistance, dexterous manipulation, and human-friendly communication. It can also support a human to lift themselves from a bed or chair. Going even further, the RIBA-II robot (http://rtc.nagoya.riken.jp/RIBA/index-e.html) by RIKEN-TRI Collaboration Center for Human-Interactive Robot Research (RTC) can lift patients of up to 80 kg from a bed to a wheelchair and back. The Pepper robot (https://www.ald.softbankrobotics.com/en/robots/pepper) from Softbank Robotics (Aldebaran) is used in a growing number of projects focusing on human-robot interaction scenarios. Some ADL (activities of daily living) tasks are directly addressed by walking aids, for example [12], and cognitive manipulation training, for example, using exoskeletons [13, 14].

The short overview indicates that individually many ADL tasks are approached. However, they all require different types

of robots. The goal of grasping objects from the floor, while at the same time keeping the robot affordable, has led us to design and build the custom Hobbit platform. Moreover, the robot should offer everyday life suitable tasks in a socially interactive manner to be sustainably used by the older adults.

3. The Hobbit Robot

Hobbit is able to provide a number of safety and entertainment functions with low-cost components. The ability to provide many functions with sometimes contradictory requirements for the hardware design creates a demanding challenge on its own. To the best of our knowledge, we are the first to present a robot that operates in users' homes in a fully autonomous fashion for a duration of 21 days per user, while providing an extensive set of functionalities like manipulation of objects with an included arm.

3.1. General Vision. The motivation for Hobbit's development was to create a low-cost, social robot to enable older adults to independently live longer in their own homes. One reason for the elderly to move into care facilities is the risk of falling and eventually inflicted injuries. To reduce this risk, the "must-haves" for the Hobbit robot are *emergency detection* (the robot patrolling autonomously through the flat after three hours without any user activity and checking if the user is well and did not suffer a fall), *emergency handling* (automatic calls to relatives or emergency services), and *fall prevention* (searching and bringing known objects to the user and picking up objects from the floor pointed to by the user and a basic fitness program to enhance the user's overall fitness). Hobbit also provides a *safety check* feature that informs the user about possible risks in specific rooms (e.g., wet floor in the bathroom and slippery carpets on wooden floors) and explains how to reduce such risks.

In science fiction, social robots are often depicted as a butler, a fact that guides the expectations towards such robots. However, as state-of-the-art technology is not yet able to fulfill these expectations, Hobbit was designed to incorporate the Mutual Care interaction paradigm [15] to overcome the robot's downfalls by creating an emotional bond between the users and the robot. The *Mutual Care* concept envisioned that the user and the robot provide help in a reciprocal manner to each other, therefore creating an emotional bond between them, so that the robot not only provides useful assistance but also acts as a companion. The resulting system complexity based on the multifunctionality was considered as acceptable to fulfill the main criteria (*emergency detection and handling*, *fall prevention*, and *providing a feeling of safety*).

3.2. Mutual Care as Underlying Interaction Paradigm. The *Mutual Care* concept was implemented through two different social roles, one that enforces this concept and one that does not. Hobbit started in the Mutual Care-disabled mode during the field trials and changed after 11 days to the Mutual Care mode. The differences between these two modes or *social roles* of the robot were mainly in its dialogues, proactivity, and the proximity in which the robot would remain when the user stops interacting with the robot. In more detail,

the main characteristics of the *Mutual Care* mode were the following: (1) return of favor: Hobbit asked if it could return the favor after situations where the user had helped Hobbit to carry out a task, (2) communication style: Hobbit used the user's name in the dialogue and was more human-like such as responding to a reward from the user by saying *You are welcome* instead of *Reward has been received*, (3) proactivity: Hobbit was more proactive and initiated interactions with the user, and (4) presence: Hobbit stayed in the room where the last interaction has taken place for at least 30 minutes instead of heading directly back to the charging station. In order to avoid potential biases, users were not told about the behavioral change of the robot beforehand.

3.3. Development Steps Leading to Hobbit. To gain insight into the needs of elderly living alone, we invited primary users (PU), aged 75 years and older and living alone, and secondary users (SU), who are in regular contact with the primary users, to workshops in Austria (8 PU and 10 SU) and Sweden (25 PU). A questionnaire survey with 113 PU in Austria, Greece, and Sweden and qualitative interviews with 38 PU and 18 SU were conducted. This iterative process [16] not only resulted in the user requirements but also influenced the design and material decisions, which were incorporated into the development of the Hobbit robots as seen in Figure 1. Based on these requirements and laboratory studies with the PT1 platform [17] with 49 users (Austria, Greece, and Sweden), the following main functionalities for Hobbit were selected:

(1) *Call Hobbit*: summon the robot to a position linked to battery-less call buttons

(2) Emergency: call relatives or an ambulance service. This can be triggered by the user from emergency buttons and gesture commands or by the robot during patrolling

(3) *Safety check*: guide the user through a list of common risk sources and provide information on how to reduce them

(4) *Pick up objects*: objects lying on the floor are picked up by the robot with no distinction between known or unknown objects

(5) *Learn and bring objects*: visual learning of user's objects to enable the robot to search and find them within the environment

(6) *Reminders*: deliver reminders for drinking water and appointments directly to the user

(7) *Transport objects*: reduce the physical stress on the user by placing objects on to the robot and letting it transport them to a commanded location

(8) *Go recharging*: autonomously, or by a user command, move to the charging station for recharging

(9) *Break*: put the robot on break when the user leaves the flat or when the user takes a nap

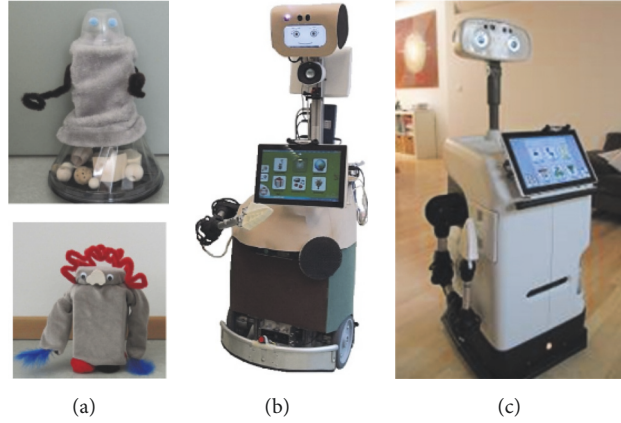(10) *Fitness*: guided exercises that increase the overall fitness of the user

(a) (b) (c)

FIGURE 1: (a–c) First mock-ups designed by secondary users: the first (PT1) and second generation of Hobbit as used during the field trials.

(11) *Entertainment*: brain training games, e-books, and music

3.4. Robot Platform and Sensor Setup.

The mobile platform of the Hobbit robot has been developed and built by MetraLabs (http://www.metralabs.com). It moves using a two-wheeled differential drive, mounted close to the front side in driving direction. For stability, an additional castor wheel is located close to the back. To fit all the built-in system components, the robot has a rectangular footprint with a width of 48 cm and a length of 55 cm. For safety reasons, a bumper sensor surrounds the base plate, protecting the hull and blocking the motors when being pressed. This ensures that the robot stops immediately if navigation fails and an obstacle is hit. An additional bumper sensor is mounted below the tablet PC, which provides the graphical user interface. During situations in which the user might not be able to reach the tablet PC (e.g. the person has fallen), a hardware emergency button is located on the bottom front side.

On its right side, the robot is equipped with a 6-DoF arm with a two-finger fin-ray gripper, such that objects lying on the floor can be picked up and placed in a tray on top of the robot's body. Furthermore, the arm can grasp a small turntable stored on the right side of the body, which is used to teach the robot unknown objects.

The robot's head, together with the neck joint with motors for pan and tilt movements, has been developed by Blue Danube Robotics (http://www.bluedanuberobotics.com). It contains two speakers for audio output, two Raspberry Pis with one display each for the eyes of the robot, a temperature sensor, and a RGB-D sensor. This sensor, referred to in the remainder of the paper as *head camera*, is used for obstacle avoidance, for object and gesture recognition, and—in conjunction with the temperature sensor—for user and fall detection. Similar to the previous prototype of the robot [3, 18], the visual sensor setup is completed by a second RGB-D sensor, mounted in the robot's body at a height of 35 cm facing forward. This sensor, referred to in the remainder of the paper as *bottom camera*, is used for localization, mapping, and user following. Figure 2 shows an overview of the Hobbit hardware; a more detailed explanation of the single components is given in the following sections.

3.4.1. Visual Perception System Using RGB-D Cameras.

For the visual perception system, Hobbit is equipped with two Asus Xtion Pro RGB-D sensors. The head camera is mounted inside the head and used for obstacle avoidance, object learning and recognition, user detection, and gesture recognition and to detect objects to pick up. Since the head can perform pan and tilt movements, the viewing angle of this camera can be dynamically adapted to a particular task at hand. In contrast, the bottom camera, used for localization, mapping, and user following, is mounted at a fixed position at a height of 35 cm in the front of the robot's body, facing forward. This setup is a trade-off between the cost of the sensor setup (in terms of computational power and money) and the necessary data for safe usage and feature completeness, which we found to be most suitable for the variety of different tasks that require visual perception.

The cameras, which only cost a fraction of laser range sensors commonly used for navigation in robotics, offer a resolution of 640×480 pixels of RGB-D data and deliver useful data in a range of approximately 50 cm to 400 cm. Therefore, our system has to be able to cope with a blind spot in front of the robot. Furthermore, the quality of data acquired with the head camera from an observed object varies depending on the task. For example, in the learning task, an object that is placed on the robot's turntable is very close to the head camera, just above the lower range limit. In the pickup task, on the contrary, the object detection method needs to be able to detect objects at the upper range limit of the camera, where data points are already severely influenced by noise.

Because two of the main goals for the final system were affordability and robustness, we avoided incorporating additional cameras, for example, for visual servoing with the robot's hand. For further details and advantages of our sensor setup for navigation, we refer the reader to [18].

3.4.2. Head and Neck.

Besides the head camera, the head contains an infrared camera for distance temperature measurement, two speakers for audio output, and two Raspberry Pis with displays showing the robot's eyes. Through its eyes,

Temperature sensor — Head
— Head camera
Speakers —
Eyes showing emotions — Neck joint (pan/tilt)
— Neck
Tray for personal belongings
Tray where Hobbit puts objects — Tablet PC with graphical UI
— Bumper sensor
Stored turntable —
Gripper — Bottom camera
6-DoF arm — Water bottle holder
— Emergency help button
Bumper sensor —

FIGURE 2: Hardware setup of the Hobbit platform.

Very tired   Sleeping   Happy   Very happy

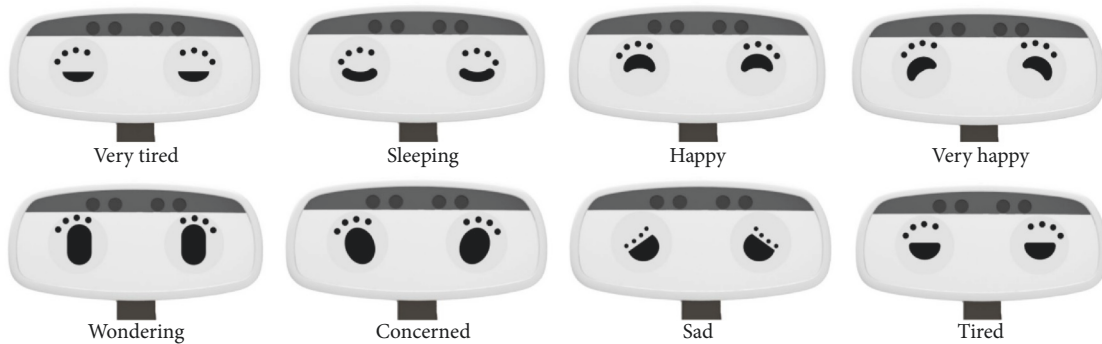Wondering   Concerned   Sad   Tired

FIGURE 3: List of emotions shown by Hobbit's eyes.

the robot is able to communicate a set of different emotions to the user, which are shown in Figure 3. The neck joint contains two servo motors, controlling the horizontal and vertical movement of the head.

3.4.3. Arm and Gripper. To be able to pick up objects from the floor or to grab its built-in turntable, Hobbit is equipped with a 6-DoF IGUS arm and a two-finger fin-ray gripper. As a cost-effective solution, the arm joints are moved by stepper motors via Bowden cables; the used fin-ray gripper offers one DoF and is designed to allow form-adaptable grasps. While an additional DoF would increase flexibility and lower the need for accurate self-positioning to successfully grasp objects, for the sake of overall system robustness and low hardware costs, the 6-DoF version was the model of choice for the arm. The arm is not compliant; therefore, cautious behavior implementation with reduced velocities for

unsupervised actions was required to minimize the risk of breakage.

4. Behavior Coordination

As Hobbit's goal directly called for an autonomous system running for several weeks, providing interactions on an irregular schedule and on-demand basis, the behavior coordination of the Hobbit robots was designed and implemented in a multistage development process. Based on the workshops with PU and SU and the user study with Hobbit PT1, elderly care specialists designed the specific scenarios. They designed detailed scripts for the 11 scenarios (see Section 3.3) the robot had to perform. Those 11 scenarios were subsequently planned in a flowchart-like fashion, which eased the transition from the design process to the implementation stage.
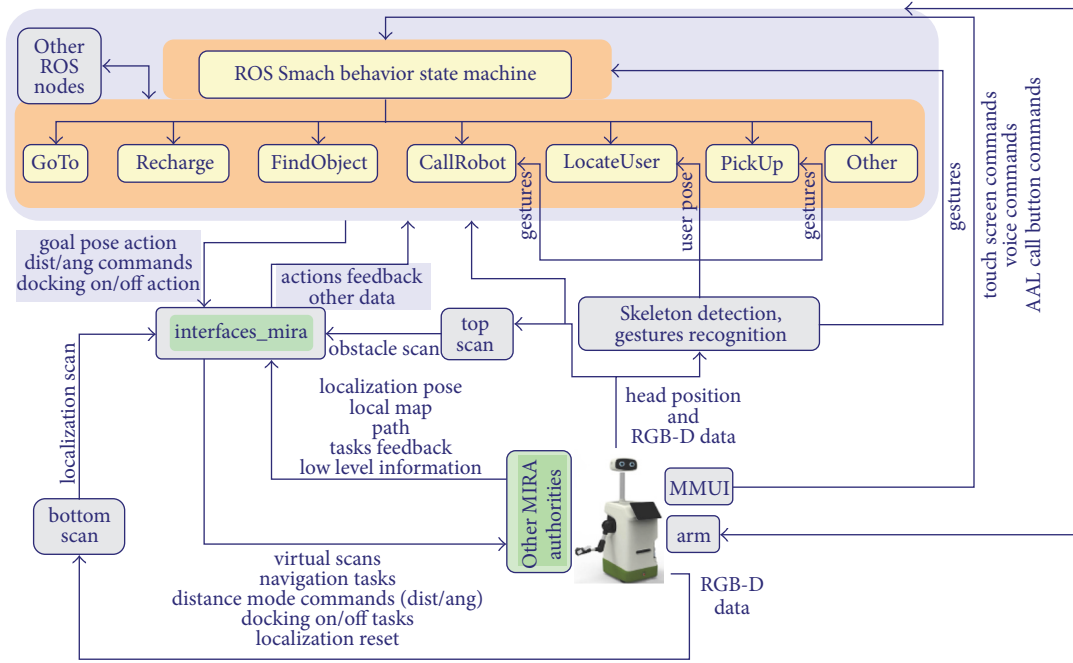
FIGURE 4: Hobbit behavior architecture.

In the following, we discuss the overall behavior coordination architecture and how the Mutual Care concept was implemented and go into detail of some of the building blocks necessary to construct the 11 scenarios. We further present the methods we developed to realize the goals of the project while respecting the limits set by the low-cost approach of our robots.

4.1. Behavior Coordination Architecture. Following the scenario descriptions, as defined by our specialists in elderly care, their implementation and the execution followed a script-based approach. A state-machine framework, SMACH (http://wiki.ros.org/smach), was therefore chosen to handle the behavior execution for all high-level codes.

An overview of the implemented architecture is shown in Figure 4. The top structure in this architecture is the PuppetMaster, which handles the decision-making outside of any scenario execution, where it can start, preempt, and restart any sub-state-machines. For this, it collects the input from those ROS nodes that handle gesture and speech recognition, text input via touchscreen, emergency detection (fallen and falling person detection, emergency button on the robot itself, and emergency gesture), and scheduled commands that need to be executed at a specific time of the day. The PuppetMaster delegates the actual scenario behavior execution to the sub-state-machines, which only rely on the input data needed for the current scenario. Each of these sub-state-machines corresponds to one of the scenarios designed to assist the users in their daily lives. As we needed to deal with many different commands with different execution priorities, it was necessary to ensure that every part of the execution of the state-machines can safely be interrupted without the risk of lingering in an undefined state. Particularly in situations when the arm of the robot was moving, it was necessary to

be able to bring it into a position in which it would be safe to perform other tasks. The movement of the robot within the environment would have been unsafe if the arm would still stick out of the footprint of the robot itself. The priorities of the commands were defined with respect to the safety of the user, so that emergency situations can always preempt a possibly running state-machine, regardless of the state the system is currently in.

4.2. RGB-D Based Navigation in Home Environments. Autonomous navigation in user's homes, especially with low-cost RGB-D sensors, is a challenging aspect of care mobile robots. These RGB-D sensors pose additional challenges for safe navigation [18, 20–22]. The reduced field of view, the blind detection area, and the short maximum range of this kind of sensors provides limited information about the robot's surroundings. If the robot, for example, turns around in a narrow corridor, it might happen that the walls are already too close to be observed while turning, leading to increased localization uncertainty. In order to prevent such cases, we defined *no-go* areas around walls in narrow passages, preventing the robot from navigating too close to walls in the first place. For obstacle avoidance, the head is tilted down during navigation, so that the head camera partially compensates for the blind spot of the bottom camera. If obstacles are detected, they are remembered for a certain time in the robot's local map. However, a suitable trade-off had to be found for the decay rate. On one hand, the robot must be able to avoid persisting obstacles, but, on the other hand, it should not be blocked for too long when an obstacle in front of it (e.g., a walking person) is removed.

While localization methods generally assume that features of the environment can be detected, this assumption

FIGURE 5: Risky areas to be avoided. Obstacles like high shelves or stairs may not be perceived by Hobbit's sensor setup.

FIGURE 6: Examples of installed ramps to overcome door thresholds.

does not hold for the used RGB-D cameras with limited range and long corridors. In this situation, according to the detected features, the robot could be anywhere along the parallel walls, which can cause problems in cases where the robot should enter a room after driving along in such a corridor. When entering a room, it is especially important that the robot be correctly localized in the transversal direction to the doorway and that the doorway be approached from the front, so accurately driving through doors located on one side of a corridor is much more difficult than through doors located at the beginning or at the end of a corridor. In order to approach doors from the front, avoiding getting too close to the corner sides, a useful strategy for wide enough places is adding *no-go* areas at sides of a doorway entrance or at sharp corners. This way, it is possible to have safer navigation behavior in wide areas while keeping the ability to go through narrower areas. This provides more flexibility than methods with fixed security margins for the whole operational area.

No-go areas were also useful to avoid potentially dangerous and restricted areas and rooms. A few examples are shown in Figure 5. Areas with cables and thin obstacles on the floor and very narrow rooms (usually kitchens), where a nonholonomic robot as Hobbit cannot maneuver, were also avoided. However, it is worth noting that *no-go* areas are

only useful as long as overall localization is precise enough. Other challenging situations were caused by thresholds and bumps on the floor and carpets. To overcome thresholds, we tested commercial and homemade ramps (Figure 6). After testing different configurations and finding proper incline limits, the robot was usually able to pass thresholds. Problems with standard planning methods, for example, when a new plan caused the robot to turn while driving on a ramp, were observed. A situation-dependent direct motion control instead of a plan-based approach can reduce the risk during such situations.

In order to facilitate the tasks to be carried out in the home environment, the concept of using rooms and labeled places inside the rooms (locations) was applied. The rooms are manually defined, such that spatial ambiguity is not a problem. Also, the geometry of the defined rooms does not have to be very precise with respect to the map, as long as the rooms contain all the places of interest that the user wants to label. Places are learned by tele-operating the robot to specific locations and the subsequent association of places to rooms operates automatically, based on the crossing number algorithm to detect whether a point lies inside a generic polygon [23]. Figure 7 shows several examples of rooms and places defined in the user trials for different tasks.
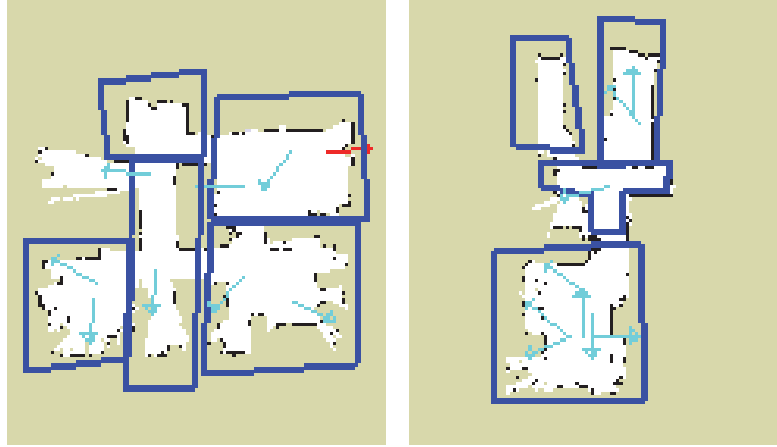
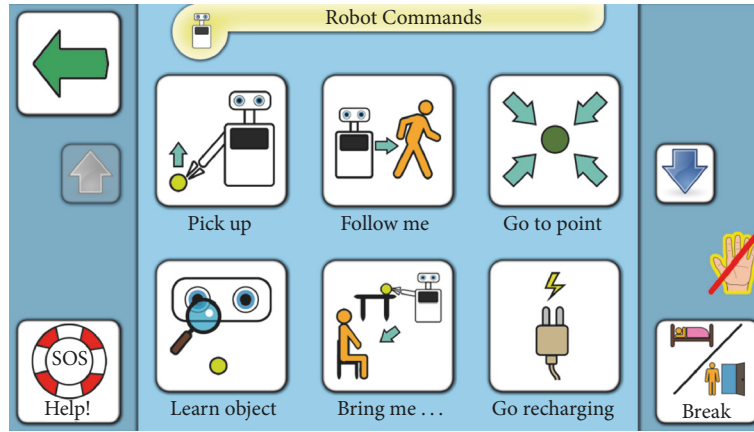FIGURE 7: Rooms and places defined in two real apartments in Vienna.

FIGURE 8: GUI of Hobbit showing one of the menu pages for robot commands. The strike-through hand on the right side indicates that the gesture input modality is disabled currently. A similar indicator was used for the speech input.

4.3. Multimodal Interaction Between the User and the Robot. The Hobbit robot deploys an improved version of the multimodal user interface (MMUI) used on Hobbit PT1. Generally speaking, the MMUI is a framework containing the following main building blocks: a Graphical User Interface (GUI) with touch, Automatic Speech Recognition (ASR), Text to Speech (TTS), and Gesture Recognition Interface (GRI). The MMUI provides emergency call features, web services (e.g., weather, news, RSS feed, and social media), control of robotic functions, and entertainment features. Compared to PT1, the graphical design of the GUI (Figure 8) was modified to better meet the user's needs. Graphical indicators on the GUI for showing current availability of GRI and ASR were iteratively improved.

During PT1 trials, we found that most of the users did not use the option of extending the MMUI to a comfortable ergonomic position for them. Therefore the mounting of the touchscreen was changed to a fixed position on Hobbit. Additionally, while the PT1 robot approached the user from the front, the Hobbit robot approaches the user from the right or left side while seated, which is more positively experienced by the user [24]. This offers the additional advantage that

the robot is close enough for the user to interact via the touchscreen, while at the same time does not invade the personal space of the user (limiting her/his movement space or restricting other activities such as watching TV). Hobbit makes use of the MMUI to combine the advantages of the various user interaction modalities [25]. The touchscreen has strengths such as intuitiveness, reliability, and flexibility for multiple users in different sitting positions but requires a rather narrow distance between user and robot (Figure 9). ASR allows a larger distance and can also be used when no free hands are available, but it has the disadvantage of being influenced by the ambient noise level, which may reduce recognition performance significantly. GRI allows a wider distance between the robot and user and also works in noisy environments, but it only succeeds when the user is in the field of view of the robot. The interaction with Hobbit always depends on the distance between the user and Hobbit. It can be done through a wireless call button (far from other rooms), ASR and GRI (2 m to 3 m), and touchscreen (arm length, see Figure 9).

The ASR of Hobbit is speaker-independent, continuous, and available in four languages: English, German, Swedish,

FIGURE 9: Different interaction distances between user and Hobbit seen from a ceiling camera. Short range: touch; middle range: speech and gesture; long range: wireless call button.

and Greek. Contemporary ASR systems work well for different applications, as long as the microphone is not moved far from the speaker's mouth. The latter case is called distant or far-field ASR and shows a significant drop in performance, which is mainly due to three different types of distortion [26]: (a) background noise, (b) echo and reverberation, and (c) other types of distortions, for example, room modes or the orientation of the speaker's head. For distant ASR, currently no off-the-shelf solution exists, but acceptable error rates can be achieved for distances up to 3 m by careful tuning of the audio components and the ASR engine [27]. An interface to a cloud based calendar was introduced, allowing PU and SU of Hobbit to access and partly to also edit events and reminders.

Despite the known difficulties with speech recognition in the far field and the local dialects of the users, the ASR of Hobbit worked as expected. The ASR was activated all over the Hobbit user trials, but the performance rate was commented on by users as necessary to be improved. The same was observed for the GRI. Eventually, the touchscreen as input modality was used most often by the majority of users, followed by speech and gesture. Touch was used more than twice as often as it was the case with ASR. Additionally, many users did not wait until the robot had completed its own speech output before starting to give a speech command which reduced the recognition rate. Considering these lessons learned, the aims for future work on the ASR are twofold: improving the performance of the ASR and providing better indication when the MMUI is listening to spoken commands and when it is not. The aspect of using two different variants for text messages from the robot to the user was taken over from Hobbit PT1. Based on other researches, it can be concluded that using different text variants does have an influence, for example, by increasing users' impression of interacting with a (more) vivid system. Some users demanded additional ASR commands, for example, *right, left, forward, reverse*, and *stop* in addition to *come closer*, as they would like to position (move) the robot with the help of voice commands or a remote control.

4.4. Person Detection and Tracking. To serve as building block for components like activity recognition [28] and natural human-robot communication [19, 29] as well as specialized functions like the fitness application [30], we developed a human body detection and tracking solution. Person detection and tracking in home environments is a challenging problem because of its high dimensionality

and the appearance variability of the tracked person. A challenging aspect of the problem in Hobbit-related scenarios is that elderly users spend a considerable amount of time sitting in various types of chairs or couches. Therefore, human detection and tracking should consider human body figures that do not stand out from their background. On the contrary, they may interact with cluttered scenes, exhibiting severe partial occlusions. Additionally, the method needs to be capable of detecting a user's body while standing or walking based on frontal, back, or side views.

The adopted solution [31] enables 3D part-based, full/upper body detection and tracking of multiple humans based on the depth data acquired by the RGB-D sensor. The 3D positions and orientations for all joints of the skeletal model (full or upper body) relative to the depth sensor are computed for each time stamp. A conventional face detection algorithm [32] is also integrated using the color data stream of the sensor to facilitate human detection in case the face of the user is visible by the sensor. The proposed method has a number of beneficial properties that are summarized as follows: (1) performs accurate markerless 3D tracking of the human body that requires no training data, (2) requires simple inexpensive sensory apparatus (RGB-D camera), (3) exhibits robustness in a number of challenging conditions (illumination changes, environment clutter, camera motion, etc.), (4) has a high tolerance with respect to variations in human body dimensions, clothing, and so forth, (5) performs automatic human detection and automatic tracking initialization, thus recovering easily from possible tracking failures, (6) handles self-occlusions among body parts or occlusions due to obstacles/furniture and so forth, and (7) achieves real-time performance on a conventional computer. Indicative results of the method are illustrated in Figure 10.

4.5. Gesture Recognition. A vision-based gestural interface was developed to enrich the multimodal user interface of Hobbit in addition to speech and touch modalities. This enables natural interaction between the user and the robot by recognizing a predefined set of gestures performed by the user using her/his hands and arms. Gestures can be of varying complexity and their recognition is also affected by the scene context, actions that are performed in the foreground or the background at the same time, and by preceding and/or following actions. Moreover, gestures are often culture-specific, providing additional evidence to substantiate the interesting as well as challenging nature of the problem.

For Hobbit, existing upper body gestures/postures as used on PT1 had to be replaced with more intuitive hand/finger-based gestures that can be performed more easily by elderly users while sitting or standing. We redesigned the gestural vocabulary for Hobbit that now consists of six hand gestures that convey messages of fundamental importance in the context of human-robot dialogue. Aiming at natural, easy-to-memorize means of interaction, users have identified gestures consisting of both static and dynamic hand configurations that involve different scales of observation (from arms to fingers) and exhibit intrinsic ambiguities. Recognition needs to be performed in continuous video streams containing
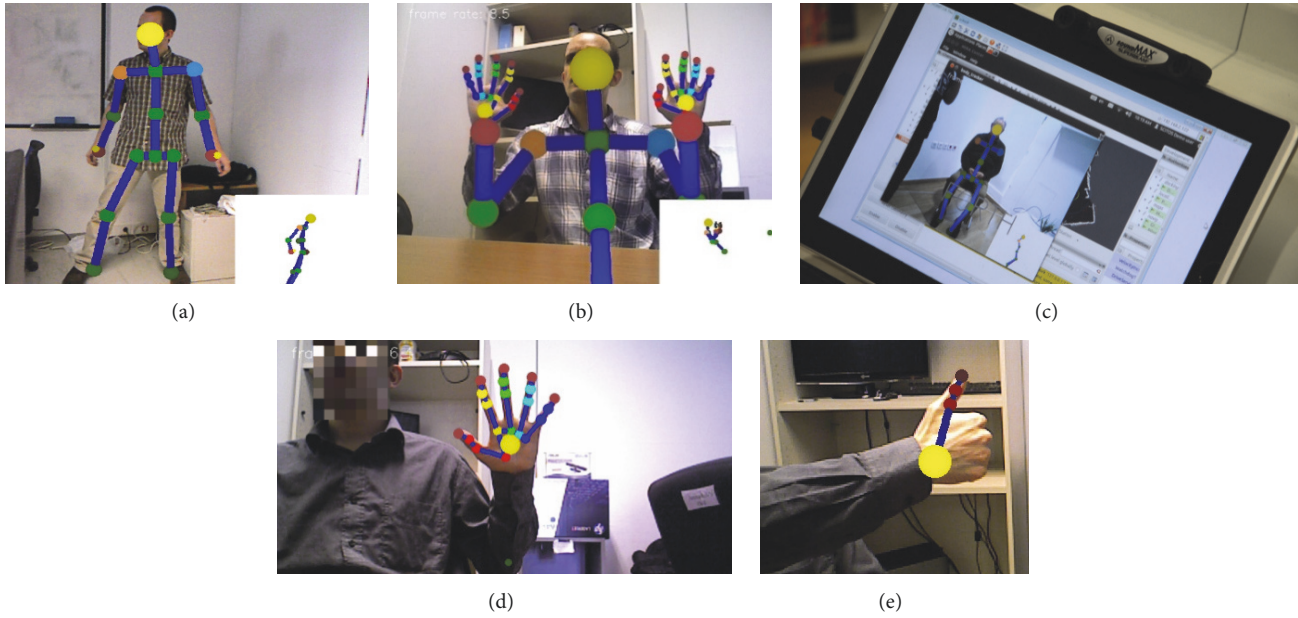
FIGURE 10: Qualitative results of the 3D skeletal model-based person detection and tracking method. (a) Full model of a standing user. (b) Upper body (including hands and fingers) of a sitting user. (c) Full model of a sitting user. ((d) and (e)) Hand and finger detection supporting the gesture recognition framework (see Section 4.5).

other irrelevant actions. All the above need to be achieved by analyzing information acquired by a possibly moving RGB-D camera in cluttered environments with considerable light variations.

The proposed framework for gesture recognition [19, 29] consists of a complete system that detects and tracks arms, hands, and fingers and performs spatiotemporal segmentation and recognition of the set of predefined gestures, based on data acquired by the head camera of the robot. Thus, the gesture recognition component is integrated with the human detection and tracking module (see Section 4.4). At a higher level, hand posture models are defined and serve as building blocks to recognize gestures based on the temporal evolution of the detected postures. The 3D detection and tracking of hands and fingers relies on depth data acquired by the head camera of Hobbit, geometrical primitives, and minimum spanning tree features of the observed structure of the scene in order to classify foreground and background and further discriminate between hand and nonhand structures in the foreground. Upon detection of the hand (palm and fingers), the trajectories of their 3D positions across time are analyzed to achieve recognition of hand postures and gestures (Table 1). The last column describes the assignment of the chosen physical movements to robot commands. The performance of the developed method has been tested not only by users acquainted with technology but also by elderly users [19] (see Figure 11). Those tests formed a very good basis for fine-tuning several algorithmic details towards delivering a robust and efficient hand gesture recognition component. The performance of the final component was tested during field trials achieving high performance according to the evaluation results.

4.6. Fall Detection. According to the assessed user needs and the results of PT1 laboratory studies [17], a top-priority and prominent functionality of Hobbit regards fall prevention and fall detection. We hereby describe a relevant vision-based component that enables a patrolling robot to (a) perform fall detection and (b) detect a user lying on the floor. We focused mostly on the second scenario, as observing a user falling in the field of view of an autonomous assistive robot is of very low probability. The proposed vision-based emergency detection mechanism consists of three modes, each of which initiates an emergency handling routine upon successful recognition of the emergency situation:

(1) Detection of a falling user in case the fall occurs while the body is observable by the head camera of the robot

(2) Detection of a fallen user who is lying on the floor while the robot is navigating/patrolling

(3) Recognition of the emergency (help) gesture that can be performed by a sitting or standing user via the gesture recognition interface of Hobbit (see Figure 11, middle)

The methodology for (1) regards a simple classifier trained on the statistics of the 3D position and velocity of the observed human body joints acquired by the person detection and tracking component. For (2), once the general assumption, the fact that the human's head is above the rest of the body, does no longer hold true, an alternative, simple, yet effective approach to the problem has been adopted. This capitalizes on calibrated depth and thermal visual data acquired from two different sensors that are available on the head of Hobbit. More specifically, depth data from both cameras of the robot

Table 1: Set of hand/arm postures/gestures considered for the gestural interface of Hobbit.

User command	Upper body gesture/posture	Robot command	Related scenarios/tasks
Yes	Thumb up-palm closed	Positive response to confirmation dialogues. YES gesture	All (1 m to 2 m distance to robot)
No	Close palm, waving with index finger up	Negative response to confirmation dialogues	All (1 m to 2 m distance to robot)
Come closer	Bend the elbow of one arm repeatedly towards the platform and the body	Reposition the platform closer to the sitting user	All (1 m to 2 m distance to robot)
Cancel task	Both open palms towards the robot	Terminate an on-going robot behavior/task	All
Pointing	Extend one arm and point in 3D space towards an object (lying on the floor)	Detect and grasp the object of interest towards the pointing 3D direction	Pick up an (unknown) object from the floor
Reward	Open palm facing towards the robot and circular movement (at least one complete circle is needed)	Rewards the robot for an accomplished action/task	Approach the user
Emergency	Cross hands pose (normal-range interaction)	Emergency detection, initiated by the user	Emergency detection

Figure 11: Snapshots of Hobbit users performing gestures during lab trials. The recognition results are superimposed as text and a circle on the images indicating the location and the name of the recognized gesture (taken from [19]).

(head and base) are acquired and analyzed while observing the floor area in front of the robot. Figure 12 illustrates sample results of the fallen user detection component. In Figure 12(a), the upper part illustrates the color frame captured by the head camera of the robot that is titled down towards the floor, while navigating. In the bottom image, the viewpoint of the bottom camera is illustrated, after the estimation of the 3D floor plane has been performed.

The methodology for vision-based emergency detection of case (3) refers to successful recognition of the emergency "Help me," based on the gesture and posture recognition module, as described in Section 4.5. The developed component is constantly running in the background within the robot's behavior coordination framework, while the robot is active during all robot tasks, except from object detection and recognition tasks.

4.7. Approaching the User. Specific behavior coordination was developed so that the robot could approach the user in a more flexible and effective way compared to standard existing methods. Using fixed predefined positions can be sufficient in certain scenarios, but it often presents limitations in real-world conditions [22]. The approach we developed incorporates user detection and interaction (Section 4.4), remembered obstacles and discrete motion for coming closer to the user with better, and adaptive positioning.

First, a safe position to move to is obtained from the local map and the robot moves there. Secondly, the user communicates to the robot whether it should move even closer or not in any of the three available modes (speech, touch, or gesture). Finally, the robot moves closer by a fixed distance of 0.15 m for a maximum of three times if the user wishes. This gives the users more control over final distance adjustments. A more detailed description of this novel approach will be published elsewhere.

4.8. User Following. As the head camera is not available for observing the full body of a user during navigation (obstacle detection), we designed a new approach [33] to localize a user by observing its lower body part, mainly the legs, based on RGB-D sensory data acquired by the bottom camera of the platform.

The proposed method is able to track moving objects such as humans, estimate camera ego-motion, and perform map construction based on visual input provided by a single RGB-D camera that is rigidly attached to a moving platform. The moving objects in the environment are assumed to move on a planar floor. The first step is to segment the static background from the moving foreground by selecting a small number of points of interest whose 3D positions are estimated directly from the sensory information. The camera motion is computed by fitting those points to a progressively built model of the environment. A 3D point may not match the current version of the map either because it is a noise contaminated observation or because it belongs to a moving object or because it belongs to a structure
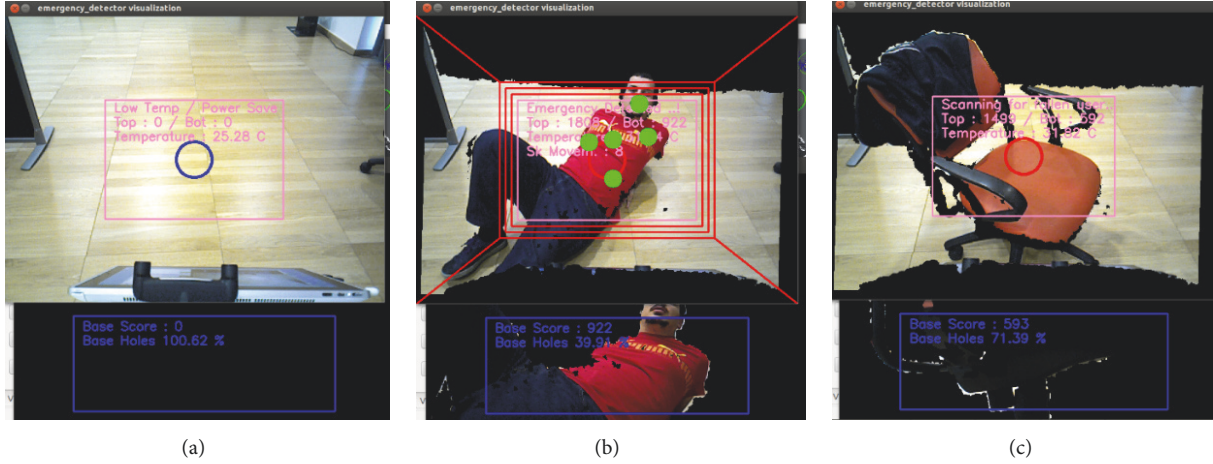
FIGURE 12: Vision-based emergency detection of a fallen user lying on the floor. The upper and lower middle images show the captured frame from the head and bottom cameras, respectively. The green dots mark a found skeleton within the search area (green and blue rectangles). (a–c) No human, no detection; person lying on the floor, correct detection; volumetric data from the head's depth and temperature sensor are in conflict with the volumetric date provided by the bottom depth sensor.
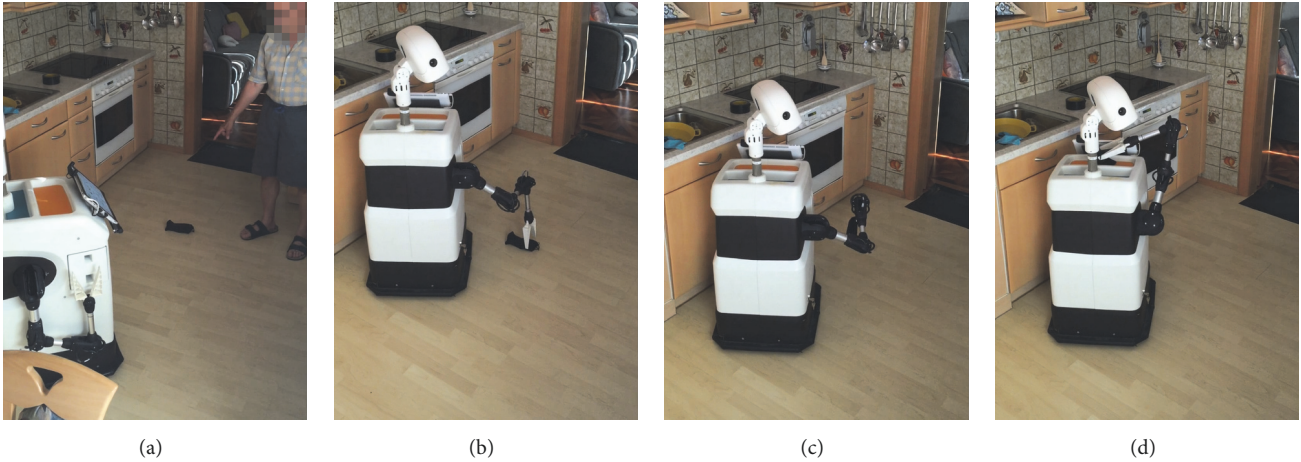

FIGURE 13: (a–d) User points to an object on the floor and Hobbit drives to a point from where it can be picked up and moves the arm to a position to grasp it. The object is lifted and the check if grasp was successful is performed: the object is moved forward to check if something has changed at the previous position of the object on the floor. If successful, the object is placed on the tray on top of the robot.

attached to the static environment that is observed for the first time. A classification mechanism is used to perform this disambiguation. Additionally, the method estimates the camera (ego) motion and the motion of the tracked objects in a coordinate system that is attached to the static environment (robotic platform). In essence, our hypothesis is that a pair of segmented and tracked objects of specific size/width that move independently side-by-side at the same distance and direction in the field of view of a moving RGB-D camera correspond to user's legs being followed by the robot with high probability. The method provides the 3D position of user's legs with respect to the moving or static robotic platform. Other moving objects in the environment are filtered out or can be provided to an obstacle avoidance mechanism as moving obstacles, thus facilitating safe navigation of the robot.

4.9. Pick Up Objects from the Floor. To reduce the risk of falling, Hobbit was designed to be able to pick up unknown objects from the floor. Figure 13 shows the steps of the "Pick up object" task. The user starts the command and points at the object on the floor. If the pointing gesture is recognized, the robot navigates to a position from where it could observe the object. At this position, the robot looks at the approximate position of the object. Hobbit then makes fine adjustments to position itself at a location from where grasping is possible. If it is safe to grasp the object, the robot executes the arm trajectory and subsequently checks if the grasp was successful and will try to do so a second time if it was not.

Several autonomous mobile robots have been developed to fetch and deliver objects to people [34–38]. None of these publications evaluate their robot grasping from floor, and none evaluate the process of approaching an object and

| Trainer | You | | Trainer | You |

We will start with your right arm. Follow me.
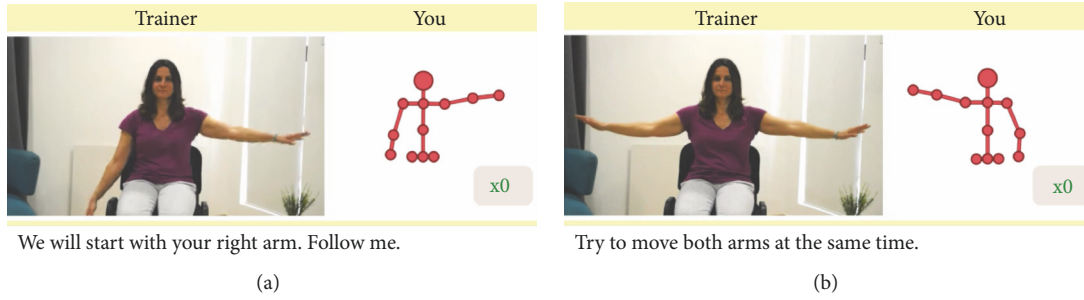
Try to move both arms at the same time.

(a)

(b)

FIGURE 14: (a) Avatar mirroring the trainer's movement proved easier for users to follow. (b) Correction suggested by the system to the user.

grasping it as a combined action. Detection of the user and recognition of a pointing gesture were performed using the work presented in [19, 31]. Checks are performed to rule out unintentional or wrong pointing gestures and to enhance the accuracy of the detected pointing gesture.

A plausibility check tests if the pointing gesture is pointing towards the floor. To guarantee an exact position of the robot to bring the arm in a position where the gripper can approach the object in a straight line before closing, the accurate movement to the grasping position can be done as a relative movement to the object instead of using the global navigation. This is a crucial step as the region, in which the head camera is able to perceive objects and where the 6-DoF arm is able to perform a movement straight down to the floor without changing gripper orientation, is limited to $15 \times 10\,cm^2$. For calculating grasps, we use the method of Height Accumulated Features [39]. These features reduce the complexity of a perceived point cloud input, increase the value of given information, and hence enable the use of machine learning for grasp detection of unknown objects in cluttered and noncluttered scenes.

4.10. Fitness Application. The fitness application was introduced as a feature to the Hobbit robot after the PT1 trials and was made available during the PT2 trials for evaluation. The motivation behind this application comes from the fact that physical activity can have a significant positive impact on the maintenance or even on the improvement of motor skills, balance, and general physical well-being of elderly people, which in turn can lower the risk of falls in the long run. Based on feedback from the Community and Active Ageing Center of the municipality of Heraklion, Greece, the following requirements were produced. The exercises must (1) be easy to learn, (2) target different joints and muscles, (3) provide appropriate feedback to the user, (4) keep the user engaged while providing enough breaks, and (5) be designed to be performed from a seated position.

Based on these requirements and feedback from test users, we developed an application including three difficulty levels and seven different exercises. The user interface consisted of a split view of a video recording of the actual trainer performing each exercise on the left side and an avatar figure depicting the user's movement while executing the instructed exercise on the right side as shown in Figure 14. This side-to-side viewing setup allowed the user to compare his or her movements to those of the trainer. The bottom part of the

interface was allocated for the instructions at the beginning of each exercise and also for any feedback and guidance to the user when needed. The design and development of the fitness application are described in more detail in [30]. The fitness application was explained to the participants of the trials by the facilitator at the initial introduction of the system during the installation day. The participants could access the application if desired at any time. Almost all users tried the fitness application at least once with some using it multiple times during the three-week evaluation period. From the comments received during the mid-term and end-of-trial interviews, it can be concluded that the overall concept of having the fitness program as a feature of the robot received positive marks by many of users as far as its usefulness and importance are concerned. However, most users who tried it out said that they would have liked it to be more challenging and to offer a larger variety of exercise routines with various challenging levels to choose from.

5. Field Trials

We conducted field trials in the households of 18 PU with 5 Hobbit robots in Austria, Greece, and Sweden. The trials lasted ~21 days for each household, resulting in a total of 371 days. During this time, the robots were placed in the homes of 18 older adults living on their own, where users could use and explore the robot on a 24/7 basis. Detailed results of the trials will be published elsewhere; preliminary results can be found in [40] (a first analysis only of the robot log data without any cross-analysis to the other data collected) and in [41] (a first overview on the methodological challenges faced during the field trials).

The trial sample consisted of 16 female and 2 male PU; their age ranged from 75 to 90 years ($M = 79.67$). All PU were living alone, either in flats (13 participants) or in houses. In adherence with inclusion criteria set by the research consortium, all participants had fallen in the last two years or were worried about falling and had moderate impairments in at least one of the areas of mobility, vision, and hearing. 15 PU had some form of multiple impairments. Furthermore, all participants had sufficient mental capacity to understand the project and give consent. In terms of technology experience, 50.0% of the PU stated that they were using a computer every day, 44.45% stated that they were never using a computer or used it less than once a week, and only one participant used a computer two to three times a week.

Before the actual trials, the PU were surveyed to make sure that they matched the criteria for inclusion and to discuss possible necessary changes to their home environments for the trials (e.g., removing carpets and covering mirrors). After an informed consent was signed, the robot was brought into the home and the technical setup took place. After this setup, a representative from the elderly care facility explained the study procedure and the robot functionalities to the PU in an individual open-ended manner. Afterwards, a manual was left within the household in case participants wanted to look up a functionality during the 21 days. All users experienced two behavioral roles of the robot. The robot was set to *device-mode* until day 11 when it was switched to *companion-mode* (i.e., Mutual Care). The real-world environment in which the field tests took place bears certain challenges, such as unforeseen changes in the environment and uncontrollable settings. Assessment by means of qualitative interviews and questionnaires took place at four stages of each trial: before trial, midterm, end of trial, and after trial (i.e., one week after the trial had ended). Moreover, log data was automatically recorded by the robot during the whole trial duration. The field trial methodology is comparable to similar studies (e.g. [42]).

The field trials revealed that several functions of the robot lack stability over time. Those technical issues certainly influenced the evaluation of the system because a reliable working technical system is a prerequisite for positive user experience. We tried to minimize potential negative feelings due to potential malfunctioning by informing our users that a prototype of a robot is a very complex technical system that might malfunction. Additionally, they were given the phone number of the facilitator who was available for them around the clock, 7 days per week, for immediate support. However, malfunctions certainly had an influence on subjects' answers during the assessments and may have attracted attention with the result that the subtle behavioral changes introduced by the switch from *device-mode* to *companion-mode* may have been shifted out of the attentional focus. Availability of commands was equally distributed across the two phases of *Mutual Care* as can be seen in Table 2. Please note that unavailability or malfunctioning of functions in one but not the other mode (unequal distribution of functionality) would have led to a bias within the evaluation. Table 2 gives an overview of the functional status across all PU during the field trials. It is based on the combination of (i) a check of the robot's features by the facilitator during the preassessment, midterm assessment, and end-of-trial assessments, (ii) protocols of the calls of the users because they had a problem with the robot, and (iii) analysis of the log data by technical partners.

The Hobbit field trials marked the first time an autonomous, multifunctional service robot, able to manipulate objects, was put into the domestic environment of older adults for a duration of multiple weeks. Our field trials provided insight into how the elderly used the Hobbit robot and which functionalities they deemed useful for themselves and how the robot influenced their daily life. Furthermore, we could show that it is in principal feasible to support elderly with a low-cost, autonomous service robot controlled by a rather simple behavior coordination system.

6. Lessons Learned

Based on all the insights gained from developing and testing Hobbit in the field, we can summarize the following recommendations for fellow researchers in the area of socially assistive robots for enabling independent living for older adults in domestic environments.

6.1. Robot Behavior Coordination. The developed behavior control based on a state-machine proved to be very useful and allowed us to implement many extensions in a short time. A close interconnection with the user was therefore helpful. In the following, we present our main lessons learned regarding the implementation of the robot behavior.

6.1.1. Transparency. Actions and their effects need to be communicated in a clear fashion so that the robot's presented functionality can be fully understood by the user. Users reported missing or nonworking functionality (e.g., reminders not being delivered to them and patrol not being executed). Most of these reported issues were caused by the fact that the users did not understand the technical interdependencies between robot functions. For example, if a command was not available due to a certain internal state of the robot, the user was not aware of this and did not understand the shown behavior of the robot. These functional relations need to be made explicit and stated more clearly to the users.

6.1.2. Legibility. The log data and conversations with participants revealed that the robot needs to communicate its intentions. For instance, when the robot proactively moved out of its charging station, the user was not always aware what was going to happen next. When they did not understand what the robot was doing, they canceled the robot's action, effectively stopping part of the robot's benefit to them. To work around this, a robot needs to clearly state the reason of its action and which goal it is trying to achieve when performing an autonomously started task.

6.1.3. Contradictory Commands. Log data presented an interesting effect while interacting with the touchscreen. When moving the hand towards the touchscreen on the robot, the gesture recognition system detected the movement of the hand as the *come closer* gesture, shortly followed by a command from the touch input on the GUI. We could replicate this behavior later on in our internal tests in the lab. A simple solution for such contradictions of commands is to simply wait for a short period of time (less than 0.2 seconds) before a gesture close to the robot is processed by the behavior coordination system to wait for a possibly following touch input.

6.1.4. Transparency of Task Interdependencies. The interviews revealed that the interdependencies between the tasks were not clear to the user; the best example was the learn-and-bring-object task. As described, for the bring-object task, the object first had to be learned so that it can be found in the apartment. However, this fact needs to be remembered by

TABLE 2: System reliability across 18 PU.

(a)

muc mode	Statistics	Call Hobbit	Come closer	Stop Hobbit	Emergency	Pick up object	Teach a new object	Bring object to user	Calendar reminders	Follow me	Move to location
	Days total	226	226	226	226	226	226	226	226	226	226
	Days of introduction	31	31	31	31	31	31	31	31	31	31
	Days switched off	55	55	55	55	55	55	55	55	55	55
Device	Days in use	140	140	140	140	140	140	140	140	140	140
	Days when feature was not working	14	13	13	19	84	12	79	49	105	15
	Days when feature was only partially working	23	20	22	44	47	116	62	83	13	32
	Days total	148	148	148	148	148	148	148	148	148	148
	Days switched off	20	20	20	20	20	20	20	20	20	20
	Days in use	128	128	128	128	128	128	128	128	128	128
Companion	Days when feature was not working	14	10	8	12	83	17	85	54	92	9
	Days when feature was only partially working	25	16	17	38	40	95	43	64	18	29
Device	Working over days in use	81.79%	83.57%	82.86%	70.71%	23.21%	50.00%	21.43%	35.36%	20.36%	77.86%
Companion	Working over days in use	79.30%	85.94%	87.11%	75.78%	19.53%	49.61%	16.80%	32.81%	21.09%	81.64%

(b)

muc mode	Statistics	Go recharge	Take a break	Telephone	Information	Surprise me	Entertainment audio	Entertainment games	Entertainment fitness	Reward
	Days in total	226	226	226	226	226	226	226	226	226
	Days of introduction	31	31	31	31	31	31	31	31	31
	Days feature was disabled	55	55	55	55	55	55	55	55	55
Device	Days of feature in use	140	140	140	140	140	140	140	140	140
	Days when feature was not working	19	16	19	11	11	11	23	20	11
	Days when feature was only partially working	20	6	22	9	23	7	20	27	6
	Days in total	148	148	148	148	148	148	148	148	148
	Days feature was disabled	20	20	20	20	20	20	20	20	20
	Days of feature in use	128	128	128	128	128	128	128	128	128
Companion	Days when feature was not working	10	8	22	8	8	7	26	19	14
	Days when feature was only partially working	33	9	16	7	23	8	11	23	7
Device	Working over days in use	79.29%	86.43%	78.57%	88.93%	83.93%	89.64%	76.43%	76.07%	90.00%
Companion	Working over days in use	79.30%	90.23%	76.56%	91.02%	84.77%	91.41%	75.39%	76.17%	86.33%

the user and as this is often not the case, users wanted to ask Hobbit to bring them an object even though it had not learned any objects before. In this specific case, the problem could be easily fixed by only offering the task "bring object" when an object was actually learned beforehand (e.g., the task could be greyed out in the MMUI).

6.1.5. Full Integration without External Programs. The handling of user input and output must be fully integrated with the rest of the robot's software architecture to be able to handle interruptions and continuations of interaction between the user and the robot. The user interface on the tablet computer (MMUI) incorporated multiple external programs (e.g., Flash games, speech recognition, and the fitness functionality). As those were not directly integrated, the behavior coordination was not aware about their current state, leading to multiple interaction issues with users. For example, a game would be exiting when a command with higher priority (e.g., emergency from fall detection) would start the emergency scenario. External programs need to be included in a way that makes it possible to suspend and resume their execution at any time.

6.1.6. Avoiding Loops. Reviewing the log data revealed that the behavior coordination system could be trapped in a loop without a way to continue the desired behavior execution. The behavior coordination needs to provide a fallback solution in case of a seemingly endless loop in any part of the behavior. The behavior coordination communicates with the MMUI in a way that does not provide immediate feedback over the same channels of communication. Due to timing issues, it occurred that a reply was lost between the communicating partners (i.e., the fact that the robot stopped speech output). From there on, the behavior coordination was in a state that should not be reached and was unable to continue program execution in the desired manner. Thus, the communication structures should always have a fallback solution to continue execution as well as the feedback data on the same channels to prevent such a stop in a scenario.

6.2. Human-Robot Interaction with the MMUI. The interaction with the user was based on a multimodal user interface that was perceived as easy to use during our field trials. While touch input turned out to be the most reliable modality, speech and gesture interaction was highly welcome. Many of the entertainment functions of the MMUI relied on Internet connectivity. Many users either were not interested in some UI features which therefore should be removed or asked for special configuration of the preferred features (e.g., selection of entertainment). The main way the user was able to communicate remotely with Hobbit was with the use of physical switches (call buttons) placed at several fixed places inside the house of the user. The user had to physically go to the designated switch spot and press the switch for the robot to approach her/him. A smartphone/tablet application could be developed to allow a better remote communication experience with the robot.

6.2.1. Internet Connectivity. Internet connectivity was not reliable depending on location and time. While in most countries Internet (line-based or mobile) coverage is no problem in general, local availability and quality vary significantly, which makes Internet-based services difficult to implement for technically unaware users. The integration of rich Internet-based content into the interaction therefore lacks usability in case of intermittent connectivity.

6.2.2. Graphical User Interface. The GUI could be personalized by the user for increased comfort during interaction. This, however, shows the need for localized content to be available. As the setup phase during the trials showed that PU are likely not aware what content is available, some (remote) support and knowledge from SU are necessary for the configuration of the user interface.

6.2.3. Speech Recognition. Field trials showed that speech recognition is still not working well for many users. Despite the overall acceptable recognition rate that varies largely from user to user and from language to language and that is based on the environment and distance, users often do not support the needs of current ASR technology for clearly expressed and separated commands in normal voice. The Sweet-Home project once more emphasizes the findings from the DiRHA 2 project that practical speech recognition for old people in the home environment is still a major challenge by itself [43]. However, our ASR provided a positively experienced natural input channel when used in a multimodal HRI, where the touchscreen with its GUI provides a reliably working base.

6.2.4. Smarthome Integration. The setup phase during the field trials showed that the integration into smarthome environments can be beneficial. Field trials showed that context awareness and adaptations highly impact the acceptance of the robot. Imagined features could be automatic on/off of the light or the stove or adjusting the proactively level of the robot based on the user's mood.

6.2.5. Remote End User Control. Reflecting on the field trial indicates that a potential valuable extension of the interaction modalities would be a remote control of the robot, for instance, on a smartphone enabling PU but also maybe SU to control the robot from outside the home. Potential useful scenarios could be to send the robot to the docking station or to patrol the flat and search for an object or the PU or the SU video calling the PU.

6.3. Implementation of Mutual Care Behavior. In the beginning of the trials, we implemented Mutual Care in such a fashion that in the companion mode the robot offers to return the favor after every interaction with the user. This was done in order to guarantee that the users would notice the difference between the modes during the interaction. The positive fact was that users noticed the changes. However, they were soon very annoyed by the robot. Consequently, we changed this implementation during the running trials. The return of favor frequency was reduced; it was no longer offered after the commands *Recharge batteries*, *Go to*, *Call*

button, and *Surprise*. Further feedback from the second and third Austrian and the second and third Swedish users led to further reduction of the return a favor frequency to offering it only after the following three commands:

(1) Pick up command (favor: Hobbit offers music: *I'd like to return the favor. I like music. Shall I play some music for you?*)

(2) Learn object command (favor: Hobbit offers to play a game (suitable because the user is already sitting down): *I'd like to return the favor. Do you want to play a game?*)

(3) Reward command (favor: Hobbit offers to surprise the user: *I'd like to return the favor. I like surprises. Do you want a surprise?*)

However, as the interviews showed, these behavioral changes were no longer recognized by the users. Similarly, the differences in proactivity and presence were not reflectively noticed by the users, but the changes in dialogue were noticed.

6.3.1. Help Situations. For the development of Mutual Care behavior in completely autonomous scenarios, which helping situations the robot can really identify in order to ask for help and how the robot can notice that it actively recovered through the help have to be considered.

6.3.2. Design of Neediness. In the interviews, PU reflected that they did not really recognize that the robot needed their input to continue its task. For Mutual Care, the *need of help* seems to be essential. For future version of the robot, how to design the *neediness* needs to be considered. This could be achieved with facial expressions, sounds, or movements. Also for behaviors such as presence and proactivity, the robot could say after an interaction: *"I would prefer staying with you in your room"* or proactivity (e.g., *"I would like to spend more time with you"* before offering an activity). This would give a better explanation of the robot's behavior to the user and an expected raise of acceptance.

7. Conclusions

In this article, we presented the second prototypical implementation of the Hobbit robot, a socially assistive service robot. We presented the main functionality it provided, as well as the behavior coordination that enabled autonomous interaction with the robot in real private homes. Hobbit is designed especially for fall detection and prevention, providing various tasks (e.g., picking up objects from the floor, patrolling through the flat, and employing reminder functionalities), and supports multimodal interaction for different impairment levels. We focused on the development of a service robot for older adults, which has the potential to promote aging in the home and to postpone the need to move to a care facility. Within the field trials, we reached the desirable long-term goal that a mobile service robot with manipulation capabilities enters real homes of older adults and showed its usefulness and potential to support independent living for elderly users.

To conclude, we believe that methods, results, and lessons learned presented in this article constitute valuable knowledge for fellow researchers in the field of assistive service robotics and serve as a stepping stone towards developing affordable care robots for the aging population.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] P. Dario, P. F. M. J. Verschure, T. Prescott et al., "Robot companions for citizens," *Procedia Computer Science*, vol. 7, pp. 47–51, 2011.

[2] J. M. Beer, C.-A. Smarr, T. L. Chen et al., "The domesticated robot: Design guidelines for assisting older adults to age in place," in *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction, (HRI '12)*, pp. 335–342, USA, March 2012.

[3] D. Fischinger, P. Einramhof, K. Papoutsakis et al., "Hobbit - The Mutual Care Robot," in *Assistance and Service Robotics in a Human Environment Workshop in conjunction with IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.

[4] J. Bohren, R. B. Rusu, E. G. Jones et al., "Towards autonomous robotic butlers: Lessons learned with the PR2," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation, (ICRA '11)*, pp. 5568–5575, China, May 2011.

[5] Y. Demiris and B. Khadhouri, "Hierarchical attentive multiple models for execution and recognition of actions," *Robotics and Autonomous Systems*, vol. 54, no. 5, pp. 361–369, 2006.

[6] T. Carlson and Y. Demiris, "Collaborative control for a robotic wheelchair: evaluation of performance, attention, and workload," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 3, pp. 876–888, 2012.

[7] M. Cashmore, M. Fox, D. Long et al., "Rosplan: Planning in the robot operating system," in *Proceedings of the 25th International Conference on Automated Planning and Scheduling, (ICAPS '15)*, pp. 333–341, June 2015.

[8] M. Mansouri and F. Pecora, "More knowledge on the table: Planning with space, time and resources for robots," in *Proceedings of the 2014 IEEE International Conference on Robotics and Automation, (ICRA '14)*, pp. 647–654, China, June 2014.

[9] S. Goetze, S. Fischer, N. Moritz, J. E. Appell, and F. Wallhoff, "Multimodal Human-Machine Interaction for Service Robots in Home-Care Environments," in *Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments*, 2012.

[10] H. S. Koppula, A. Jain, and A. Saxena, "Anticipatory planning for human-robot teams," *Springer Tracts in Advanced Robotics*, vol. 109, pp. 453–470, 2016.

[11] P. Kwang-Hyun and Z. Zenn Bien, "Intelligent sweet home for assisting the elderly and the handicapped," in *Independent Living for Persons with Disabilities and Elderly People: ICOST*, p. 151, 2003.

[12] T. Fukuda, P. Di, F. Chen et al., "Advanced service robotics for human assistance and support," in *Proceedings of the International Conference on Advanced Computer Science and Information Systems, (ICACSIS '11)*, pp. 25–30, December 2011.

[13] D.-J. Kim, R. Lovelett, and A. Behal, "An empirical study with simulated ADL tasks using a vision-guided assistive robot arm," in *Proceedings of the 2009 IEEE International Conference on Rehabilitation Robotics, (ICORR '09)*, pp. 504–509, Japan, June 2009.

[14] M. B. Hong, S. J. Kim, T. Um, and K. Kim, "KULEX: An ADL power-assistance demonstration," in *Proceedings of the 10th International Conference on Ubiquitous obots and Ambient Intelligence, (URAI '13)*, pp. 542–544, November 2013.

[15] L. Lammer, A. Huber, A. Weiss, and M. Vincze, "Mutual Care: How older adults react when they should help their care Robot," in *Proceedings of the 3rd international symposim on New Frontiers in Human-Robot interaction*, April 2014.

[16] T. Körtner, A. Schmid, D. Batko-Klein et al., "How social robots make older users really feel well - A method to assess users' concepts of a social robotic assistant," in *Proceedings of the International Conference on Social Robotics*, vol. 7621, pp. 138–147, Springer, 2012.

[17] D. Fischinger, P. Einramhof, K. Papoutsakis et al., "Hobbit, a care robot supporting independent living at home: First prototype and lessons learned," *Robotics and Autonomous Systems*, vol. 75, pp. 60–78, 2016.

[18] P. De La Puente, M. Bajones, P. Einramhof, D. Wolf, D. Fischinger, and M. Vincze, "RGB-D sensor setup for multiple tasks of home robots and experimental results," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS '14)*, pp. 2587–2594, USA, September 2014.

[19] D. Michel, K. Papoutsakis, and A. A. Argyros, "Gesture recognition supporting the interaction of humans with socially assistive robots," in *Advances in Visual Computing*, vol. 8887 of *Lecture Notes in Computer Science*, pp. 793–804, Springer International Publishing, Berlin, Germany, 2014.

[20] J. González-Jiménez, J. Ruiz-Sarmiento, and C. Galindo, "Improving 2D Reactive Navigators with Kinect," in *Proceeding of the 10th International Conference on Informatics in Control, Automation and Robotics (ICINCO '13)*, 2013.

[21] M. Jalobeanu, G. Shirakyan, G. Parent, H. Kikkeri, B. Peasley, and A. Feniello, "Reliable kinect-based navigation in large indoor environments," in *Proceedings of the 2015 IEEE International Conference on Robotics and Automation, (ICRA '15)*, pp. 495–502, USA, May 2015.

[22] P. De La Puente, M. Bajones, Reuther. C., D. Fischinger, Wolf. D., and M. Vincze, "Experiences with rgb-d based navigation in real home robotic trials," in *Austrian Robotics Workshop (ARW)*, 2016.

[23] M. Shimrat, "Algorithm 112: Position of point relative to polygon," *Communications of the ACM*, vol. 5, no. 8, p. 434, 1962.

[24] M. L. Walters, K. Dautenhahn, S. N. Woods, and K. L. Koay, "Robotic etiquette: Results from user studies involving a fetch and carry task," in *Proceedings of the HRI 2007: 2007 ACM/IEEE Conference on Human-Robot Interaction - Robot as Team Member*, pp. 317–324, USA, March 2007.

[25] P. Mayer, C. Beck, and P. Panek, "Examples of multimodal user interfaces for socially assistive robots in ambient assisted living environments," in *Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications, (CogInfoCom '12)*, pp. 401–406, Slovakia, December 2012.

[26] M. Wölfel and J. McDonough, *Distant Speech Recognition*, John Wiley & Sons, 2009.

[27] P. Panek and P. Mayer, "Challenges in adopting speech control for assistive robots," in *Ambient Assisted Living*, Advanced Technologies and Societal Change, pp. 3–14, Springer International Publishing, Berlin, Germany, 2015.

[28] D. I. Kosmopoulos, K. Papoutsakis, and A. A. Argyros, "Online segmentation and classification of modeled actions performed in the context of unmodeled ones," in *Proceedings of the 25th British Machine Vision Conference, (BMVC '14)*, pp. 1–12, September 2014.

[29] D. Michel and K. Papoutsakis, "Gesture Recognition Apparatuses," *Methods and Systems for Human-Machine Interaction*, 2016.

[30] M. Foukarakis, I. Adami, D. Ioannidi et al., "A robot-based application for physical exercise training," in *Proceedings of the 2nd International Conference on Information and Communication Technologies for Ageing Well and e-Health, (ICT4AWE '16)*, pp. 45–52, April 2016.

[31] D. Michel and A. Argyros, "Apparatuses, methods and systems for recovering a 3-dimensional skeletal model of the human body, 2016".

[32] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[33] P. Panteleris and A. A. Argyros, "Vision-based SLAM and moving objects tracking for the perceptual support of a smart walker platform," *Computer Vision-ECCV 2014 Workshops*, vol. 8927, pp. 407–423, 2014.

[34] A. Remazeilles, C. Leroux, and G. Chalubert, "SAM: A robotic butler for handicapped people," in *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, pp. 315–321, Germany, August 2008.

[35] S. S. Srinivasa, D. Ferguson, C. J. Helfrich et al., "HERB: A home exploring robotic butler," *Autonomous Robots*, vol. 28, no. 1, pp. 5–20, 2010.

[36] B. Graf, U. Reiser, M. Hägele, K. Mauz, and P. Klein, "Robotic home assistant care-O-bot 3—product vision and innovation platform," in *Proceedings of the IEEE Workshop on Advanced Robotics and Its Social Impacts (ARSO '09)*, pp. 139–144, Tokyo, Japan, November 2009.

[37] D. Fischinger and M. Vincze, "Empty the basket - A shape based learning approach for grasping piles of unknown objects," in *Proceedings of the 25th IEEE/RSJ International Conference on Robotics and Intelligent Systems, (IROS '12)*, pp. 2051–2057, Portugal, October 2012.

[38] D. Fischinger and M. Vincze, "Shape based learning for grasping novel objects in cluttered scenes," in *Proceedings of the 10th IFAC Symposium on Robot Control, (SYROCO '12)*, pp. 787–792, Croatia, September 2012.

[39] D. Fischinger, A. Weiss, and M. Vincze, "Learning grasps with topographic features," *International Journal of Robotics Research*, vol. 34, no. 9, pp. 1167–1194, 2015.

[40] M. Bajones, A. Weiss, and M. Vincze, "Log data analysis of long-term household trials: Lessons learned and pitfalls. 2016. Workshop on Challenges and best practices to study HRI in natural interaction settings," in *Proceeding of the International Conference on Human-Robot Interaction*, 2016.

[41] J. Pripfl, T. Körtner, D. Batko-Klein et al., "Results of a real world trial with a mobile social service robot for older adults," in *Proceedings of the 11th Annual ACM/IEEE International*

Conference on Human-Robot Interaction, (HRI '16), pp. 497-498, IEEE Press, New Zealand, March 2016.

[42] F. Palumbo, D. La Rosa, E. Ferro et al., "Reliability and human factors in ambient assisted living environments," *Journal of Reliable Intelligent Environments*, vol. 3, no. 3, pp. 139–157, 2017.

[43] M. Vacher, F. Portet, A. Fleury, and N. Noury, "Development of audio sensing technology for ambient assisted living: Applications and challenges," *International Journal of E-Health and Medical Communications*, vol. 2, no. 1, pp. 35–54, 2011.