IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 17, 2019, accepted August 14, 2019, date of publication August 29, 2019, date of current version September 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938366

Grasping Objects From the Floor in Assistive Robotics: Real World Implications and Lessons Learned

PALOMA DE LA PUENTE[1,4], (Member, IEEE), DAVID FISCHINGER[2,4], MARKUS BAJONES[2,4], (Member, IEEE), DANIEL WOLF[3,4], (Member, IEEE), AND MARKUS VINCZE[4], (Member, IEEE)

[1]Centre for Automation and Robotics, UPM-CSIC, Universidad Politécnica de Madrid, 28006 Madrid, Spain
[2]Aeolus Robotics, Inc., 1040 Vienna, Austria
[3]DAQRI Holographics Research Center, 1010 Vienna, Austria
[4]Vienna University of Technology (TU Wien), 1040 Vienna, Austria

Corresponding author: Paloma de la Puente (paloma.delapuente@upm.es)

This work was supported in part by the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant 288146, HOBBIT, in part by the Spanish Ministry of Science, Innovation and Universities through the project COGDRIVE - Artificial Intelligence Techniques and Assistance to Autonomous Navigation under Grant DPI2017-86915-C3-3-R, in part by the RoboCity2030-DIH-CM, Madrid Robotics Digital Innovation Hub, under Grant S2018/NMT-4331, in part by the Programas de Actividades I+D en la Comunidad de Madrid, and in part by the Structural Funds of the EU.

ABSTRACT This paper presents a system enabling a mobile robot to autonomously pick-up objects a human is pointing at from the floor. The system does not require object models and is designed to grasp unknown objects. The robot decides by itself if an object is suitable for grasping by considering measures of size, position and the environment suitability. The implementation is built on the second prototype of the home care robot Hobbit, thereby verifying that complex robotic manipulation tasks can be performed with economical hardware. The presented system was already tested in real apartments with elderly people. We highlight this by discussing the additional complexity for complete autonomous behavior in apartments compared with tests in labs.

INDEX TERMS Autonomous systems, grasping, human-robot interaction, intelligent robots, mobile robots, service robots, system integration.

I. INTRODUCTION

Robots have been envisioned as helpers at home for a really long time. In this direction, great advancements are achieved every year but, if we look closer, a general multi-purpose autonomous robotic butler is still far away: it is indeed an ambitious goal [1]–[5]. An important aspect of service robots is their potential to support independent living at home for the elderly, which is becoming a challenge due to the aging of society. The greatest danger is undoubtedly the risk of falling [6], [7].

Consequently, the Hobbit robot was developed so as to increase the old users feeling of safety [8]. In addition to providing support and calling for help if something happens, it also incorporates preventive measures: by picking up objects from the floor the robot can reduce the user's chances of tripping or losing balance.

In order to get closer to the real application, we must leave the typically fixed experimental setups and start focusing on real homes. As recently highlighted, it is indeed very challenging to transfer manipulation abilities from controlled situations to dynamic and unstructured environments that require mobility and a seamless integration of different technologies [9].

We move on and do not let the robot stand in front of an empty area to grab an object from a nice position, in a simplified scenario. Instead, we address the whole process: starting the pick up command via pointing gestures, integrating autonomous navigation to the selected area, incorporating methods for the robot to find the object and including a checking stage to assess whether the object size, properties and position are adequate for a grasping task. The arm will start to move, in order to pick up the object, if and only if it

The associate editor coordinating the review of this manuscript and approving it for publication was Hui Xie.

is possible and safe. Height Accumulated Features (HAF) are used for grasping previously unknown objects [6].

In the rest of the paper we will discuss and illustrate why this is significantly more difficult than what other grasping systems have attempted thus far. Our contribution is a comprehensive approach for autonomous object grasping at home. The main novelty is the integration of the whole pipeline for picking up desired objects in real world circumstances.

The additional required components include: a mobile head which must be able to perceive in different directions, calibration for different distances, recognition of gestures, long distance detection and abstraction of objects as well as the robot movement towards a suitable position close to the object, avoiding obstacles if necessary. Moreover, a proper behavior coordination to handle different possible situations is needed. When you really go into the actual homes of users, it is very important that the robot does not break itself and only intervenes if there is no significant risk. The robot should also be able to recover from mistakes, retry the grasping if necessary or at least should be able to tell whether it succeeded or not.

Assistive robotics is a complex and wide field, and grasping objects from the floor from a mobile platform is one of the biggest challenges. Even though the scope of all of them is very different, a number of previous publications are related to this article, so we will briefly mention the ones that best help understand all the context. The first contributions for grasping unknown objects in cluttered controlled scenarios by means of machine learning techniques and specifically extracted features such as HAF were presented in [10]–[12]. The interested reader is referred to the extended detailed version with more experiments [6]. The first prototype version of the Hobbit robot for multiple tasks and results from laboratory trials were described in [13]. This version included a preliminary set of functions to check whether an object should be grasped or not but the limited arm kinematics and a slow method to decide if the task was successful or not, together with limited navigation capabilities, resulted in a not very impressive overall performance. The sensor setup configuration for the multi purpose assistive robot was discussed in [14]. The whole robotic system for the elderly based on the second prototype was already presented in [8]. This work included a high level general overview of all components and functionalities of the system. It also presented the lessons learned from field trials conducted with real users in their own homes for up to three weeks, mostly from a Human-Robot-Interaction perspective. Other papers explain certain parts of the system and their own challenges, such as RGB-D navigation in real homes [4], [15]. This article focuses on the advanced capabilities for autonomously picking up selected objects from the floor. We discuss the motivation and difficulties, provide a literature review and then explain the main requirements and detailed components as well as how we integrated them, finally showing our most relevant and innovative results.

The paper is organized as follows. In the next section we discuss related work. In Section III we describe the hardware and discuss contradictory design requests for a multi-functional assistive robot. Section V presents the components for the picking up scenario: i.e. navigation, fine positioning, grasping and failure detection. In Section VI we present our test results. Thereby we also focus on trial runs in which the robot should decide not to grasp, since safety for the user and the robot itself is a main issue for robots autonomously operating in real human environments.

II. RELATED WORK

The problem of picking up objects lying on a plane was addressed by Xu *et al.* [16]. In this work, the authors used an iRobot Create with an additional 1-DOF arm and a compliant finger to sweep an object onto a flat surface mounted on the robot and hold it there. For this approach a non-prehensile manipulator was employed. No sensor data were considered for controlling the predefined manipulation sequence, the robot was assumed to be already placed at a position from where the grasping script would succeed. A success rate of 94.71% in 680 trials -combining 4 floor types with 34 objects of particular relevance to assistive applications in 5 different poses each ($4 \times 34 \times 5 = 680$)- showed very good results. Even for objects like coins, credit cards, keys, a dollar bill and a pill the system had an impressive performance. In [17] an improved version of the robot which was able to lift objects for delivery was presented and evaluated with 20 people suffering from ALS. For these trials the robot was controlled by the test users, who decided where to place to robot and which pick up step should be started each time.

In [18] an intelligent assistive robotic 6-DOF manipulator was introduced for grasping objects for people in wheelchairs. In this work no systematic evaluation of the grasping process is documented. A number of other publications also present wheelchair-mounted robot arms (WMRAs) (e.g. [19]–[22]) to assist users in performing various tasks including object pick up and object release.

Grasp evaluations on a PR2 robot for known objects and unknown objects [10]–[12] in a predefined setup achieve good success rates, but these works are assuming special conditions and do not deal with a number of issues that make the picking up task presented in this paper more complex: navigation (including localization and obstacle avoidance), accurate fine positioning relative to the object, changing position of the robot camera and limited perception quality due to a higher distance between the camera and the object.

Several autonomous mobile robots have been developed to fetch and deliver objects to people [23]–[26]. None of these publications evaluates the robot to grasp objects from the floor, and none of these publications evaluates the process of approaching an object and grasping it as a combined action.

Recent work by Levesque *et al.* focused on grasping difficult unknown objects -including thin objects- with a commercial arm and a commercial gripper [27]. A controlled scenario was established, with no navigation and no sophisticated perception required. The width and thickness of the detected objects was checked to decide the most robust method in each case. Kasaei *et al.* highlighted the importance of tightly integrating perception and manipulation for enhanced robustness in service robotics [28]. They developed a learning based framework, exploiting a Working Memory and Perceptual Memory system and incorporating human instruction capabilities. A commercial robotic arm and an external RGB-D sensor were used for the evaluation. No navigation was involved in this work.

Another approach [29] uses an in-hand camera to detect three objects at fixed positions, evaluating 3 different interface-methods to chose the object to be grasped. The procedure works as follows: the user selects an object, the robot moves there, the robot places the arm above the object, the object is segmented from the in-hand camera image, the gripper is positioned and it goes down until tactile sensors touch the object, then the gripper is closed. A success rate of 94.8% was achieved, given that the robot could repeat the grasp trial up to 4 times when it failed. A total of 134 trials were performed with 8 users. The interesting follow up work presented in [30] proposed a highly specialized system for the purpose of picking up. This new version included laser based object pointing, navigation using odometry and collision checks, with very promising results. Using odometry presents well known limitations that map based navigation overcomes.

This article presents a robot that is able to provide a number of additional functions using cheap components. The ability to provide many functions with contradictory requirements for the hardware design creates demanding challenges on its own. To the best knowledge of the authors we are the first ones to present a realistic autonomous pickup evaluation on a sophisticated multi-functional robot that is additionally low priced in production (final total price of the whole system around 15000 Euro).

III. MULTI-FUNCTIONAL LOW COST ROBOT HARDWARE

Multi-functionality and low cost production go hand in hand with suboptimal conditions for specific grasping tasks. Furthermore, the required autonomy of an unsupervised service robot for private homes demands a very robust hardware system that is not easily damaged. As a result, the design should include a low number of potentially breakable sensors, motors and other devices. An overcautious behavior is preferred. An achievement of the presented work is to cope with unavoidable compromise conditions regarding a grasping task. In this section we introduce the robot components, explain the shortcomings regarding a pickup task and mention possible improvements and their implications regarding low cost, multi-functionality or system complexity requirements. A total price below 15000 Euro was achieved for the whole robot.

A. MOBILE PLATFORM

The second prototype of Hobbit is based on a differential drive system with two driving wheels on the front side (main driving direction) and a castor wheel close to the back side. A holonomic robot could compensate fine positioning errors of the robot and hence dynamically compensate shortcomings of the manipulator kinematics. However, using an omnidirectional platform would significantly increase the robot cost, making it difficult to satisfy affordability requirements.

B. PERCEPTION (RGB-D CAMERAS)

Hobbit PT2 uses 2 Asus Xtion Pro cameras. One camera is mounted in the front area of the robot's base, 35 cm high, and is used for localization. The second camera is positioned at the head, 125 cm above the ground, and is used for obstacle avoidance, object learning, user and gesture detection and for picking up objects. The position of the head camera is suboptimal for picking up objects from the floor regarding:

- Height, since there is reduced resolution due to the sensor distance to the floor
- Field of view, since there are self occlusions and there is little overlap between the graspable area of the manipulator and the detection area of the sensor

The low cost depth camera imposes a minimal distance for user detection and for object learning requiring manipulation, and also for obstacle avoidance, so the sensors configuration was chosen as a compromise solution to achieve the desired multi-functionality [14]. Additional cameras, such as a camera for visual servoing in the robot hand, or technical add-ons, to get more degrees of freedom for positioning the cameras, would create a conflict with the low cost, low complexity and high robustness demands.

Localization can be crucial for a complex pickup task. It is needed for the robot to decide which objects lay in a graspable area (e.g. not too close to walls) and to plan a collision free path towards the object. The base camera used for localization has a field of view of about 60°, which is rather small compared to laser scanners with at least 180° field of view, and it presents a significantly shorter maximum range, which may hinder localization. However, the cost is significantly lower and we found this kind of domestic navigation not perfect but feasible [4]. Details, limitations and advantages of our sensor setup for navigation are included in [4], [14].

C. HEAD AND NECK

The head, designed and built by Blue Danube Robotics, is equipped with the top Asus Xtion Pro sensor mentioned above, an infrared camera, two Raspberry Pis with displays to show emotions and two speakers. Two servos enable horizontal and vertical head movement. The head is still a prototype and considerable force is needed for active head movement. In certain positions, the head load results in inaccurate servo feedback. Small deviations of the servo feedback result in arm to head calibration errors which can lead to failed grasps. Again, a trade-off between cost and functionality led to the chosen hardware setup.

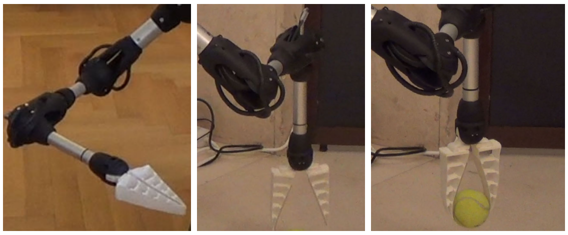


FIGURE 1. The robot gripper. Left: while the arm is moving in order to grasp an object. Middle: before grasping an object. Right: after grasping an object.

D. ARM

As manipulator, we used a 6-DoF IGUS arm with stepper motors to move the joints via Bowden cables made of Dyneema. The best grasping results can in general be achieved if the gripper is approaching a grasping position for an object in a straight trajectory, without changing the gripper orientation. This way pre-mature contact between gripper and object due to calibration errors or not perceived object data can be minimized. For the 6-DoF arm, the area where inverse kinematic solutions for straight gripper trajectories with fixed gripper orientations can be found is strongly limited, even for short (10 cm approx.) trajectories. Using a 7-DoF arm to solve this issue would conflict with system robustness requirements, specifically due to the Bowden cable construction. Cost requirements would not be respected either if making such a decision. The arm is non-compliant, so a cautious behaviour is required for non supervised actions.

E. GRIPPER

Hobbit is equipped with a Festo fin-ray gripper with one degree of freedom developed by Hella Automation (see Fig. 1). The design and construction of the plastic fingers is suitable for form-adaptable grasps. Even though there are shortcomings compared to more advanced grippers that offer less gripper force and maximum payload, given that the price of advanced grippers [31] exceeds the overall price of the Hobbit robot, the fin-ray gripper was deemed a reasonable choice.

IV. OBJECT AND ENVIRONMENT SPECIFICATIONS FOR GRASPING

We specify what type of objects can be grasped and which environmental conditions are admitted.

- Object size: the smallest side of the object is at least of 20 mm and not larger than 100 mm, the opening width of the gripper. Objects should be shorter than 30 cm to fit into the tray of the robot.
- Object weight is up to 500 g for the fully extended arm.
- The surface properties of the objects are opaque and matte. Glossy parts of the surface may contain reflections that do not allow object recognition. Object surface properties should be different to the background surface properties, for example with different colours.
- Soft deformable objects such as a kitchen cloth or foam pieces are admitted
- Environmental conditions are such that illumination should correspond to an indoor lighting situation with no direct sun light and no light from electric bulbs with more than 500 Lux (equivalent to cloudy daylight). This condition is also acceptable if there is no ambient illumination, for example, at night.
- Object situation in the environment: The robot must be able to approach the location of the object (enough space to come close enough to the object). There needs to be free space between the robot and the object so that the robot arm can freely reach the object.

Most other works existing in the literature also present similar limitations and restrictions but do not always explicitly mention them. Jain and Kemp [30], for instance, check whether the object to be grasped fits or not between the fingers of the gripper, discarding objects larger than 12 cm along the direction of minimum variation. Gualtieri *et al.* [22] highlight limitations related to the point of view and distance from the camera to the object in order to plan a trajectory for information gathering.

V. COMPONENTS FOR THE PICKING UP SCENARIO

For a whole picking up scenario we propose several components, considering a number of possible situations that are usually encountered in the real world. The implemented state machine is depicted in Figure 2 for illustration purposes, to show how all these components were integrated within our system. Please note that a controlled individual pickup task would require just a minimum subset of this system diagram.

The overall process works as follows. In the first place, the pick up function may be triggered by pressing a button from the tablet user interface mounted on the robot or by the voice command ‘‘Hobbit, Pick up’’. Then the robot asks the user to point at an object. If the gesture is properly recognized, the robot starts to navigate towards a pose from which the selected object can be perceived. Once there, the robot head moves so as to look at the approximate position of the object. If the object is detected, simple discrete motion commands are applied for a fine positioning of the robot relative to the object, in order to reach a final pose from which grasping is possible. If grasping the object is deemed safe, the robot executes the arm trajectory and subsequently checks whether the object was successfully picked up. If not, it will try again.

A. PERCEPTION OF POINTING GESTURE

For detection of the user and recognition of a pointing gesture we use the work presented by Michel *et al.* in [32] combined with a full body skeleton tracking solution. This method is based on the detection and tracking of body parts across RGB-D frames, using a layered representation of a hand model and comparing a set of possible candidates in terms of geometric shape and trajectory properties. For the picking up application, the user's arm which is further away from the center of the human body is always the one selected.

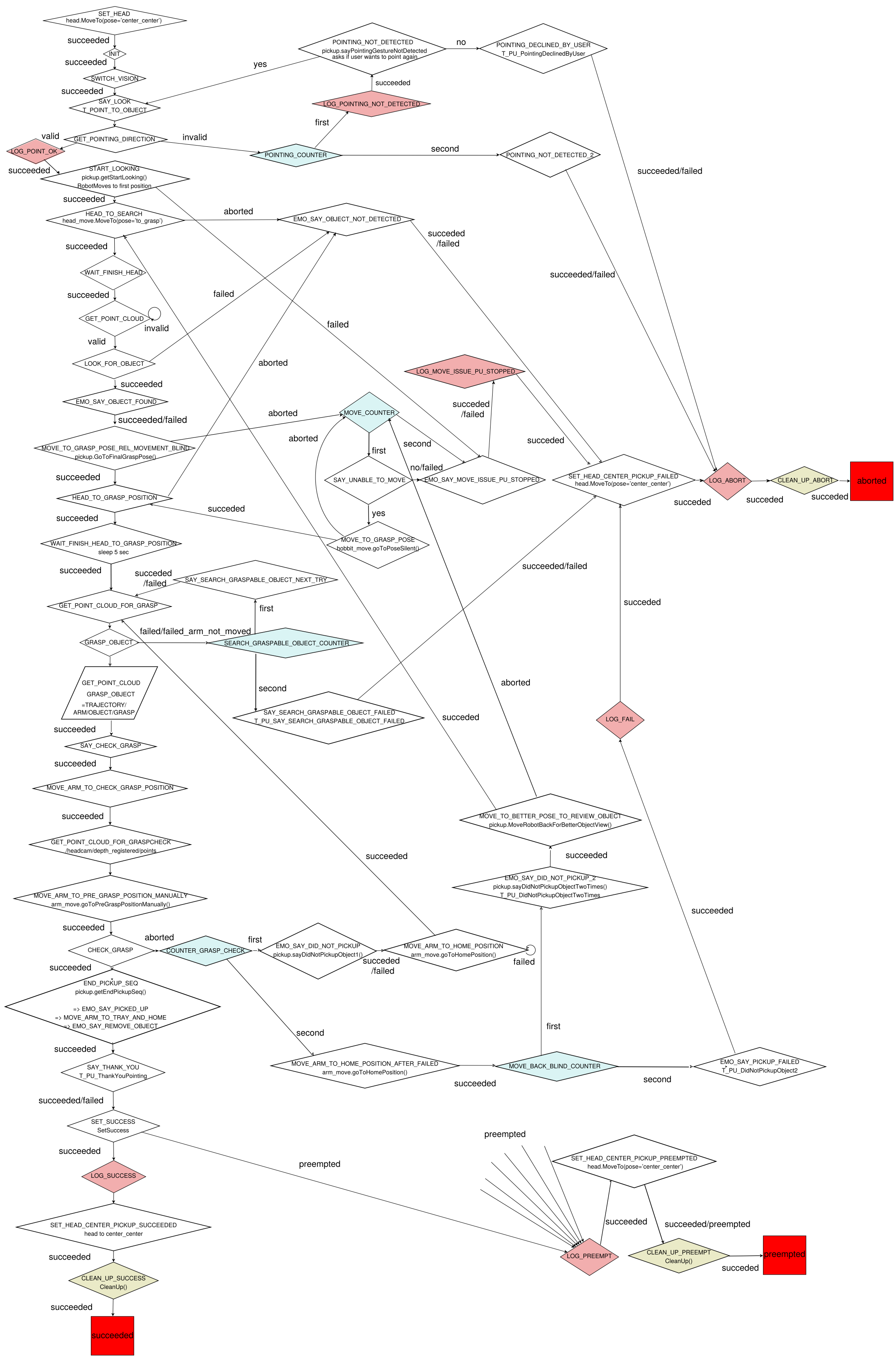


FIGURE 2. Picking up scenario SMACH state machine.

To rule out unintentional or wrong pointing gestures and enhance the accuracy of the detections, several checks are performed. In the first place, a plausibility check is applied to determine whether the pointing gesture is pointing towards the floor. Then a pointing gesture is only accepted as valid if the pointing direction is similar enough to the previous one. Pointing gestures are calculated at a rate of 15 frames per second. This provides a 2D goal to be reached in a safe manner by the map based navigation system, as described below.

B. NAVIGATION

As outlined before, autonomous navigation in real user homes using RGB-D perception presents numerous challenges. We integrated MIRA [33] navigation methods into our ROS based framework and extended it for several functionalities. GMapping [34] was used for building occupancy grid maps of the user homes in an initial setup phase. New detected obstacles not included in the map are remembered for a while to overcome the short range blind detection area of the sensor. Our developments, adopted solutions and identified issues are analyzed in more detail in [4], [15].

C. FINE POSITIONING

To guarantee an exact position of the robot to bring the arm in a position where the gripper could 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 global navigation. This is crucial since the overlapping region where the head camera can perceive objects and the 6-DoF arm can generally execute straight movements towards the floor without changing the orientation of the gripper is limited to a more or less 15 cm × 10 cm area.

D. GRASP POINT DETECTION

For calculating grasps, we use the method of Height Accumulated Features ([6]). Height Accumulated Features provide a compact representation of local shapes that reduces the complexity of a perceived point cloud input, increases the value of the given information and hence enables the use of machine learning for grasp detection of unknown objects in cluttered and non cluttered scenes. Basically, the best position where to place a 2-finger gripper given an arbitrary surface of an object is learned.

E. GRASP PLANNING

For the calculation of a grasp trajectory the simulation environment OpenRAVE [35] is used. The perceived floor plane is detected and removed and the remaining object data is segmented and added to the simulation environment for collision checking and hence to calculate how close the gripper can get to the object. To optimize the grasp success rate, experience shows that a straight gripper movement perpendicular to the closing direction of a two finger gripper gives the best results and can best compensate small errors due to not perceived object surfaces or inaccurate data (e.g due to camera calibration or erroneous angles of pitch-yaw servos). Experiments in OpenRAVE showed that straight movements without changing the gripper orientation from a height around 20cm down to the floor are strongly limited for our 6-DoF robot arm. The tested area is reduced to the visible area of the head camera. The Bi-directional Rapidly-Exploring Random Trees (BiRRT) [36] method used for arm path planning in OpenRAVE achieves good results for a 10 cm × 15 cm area. This algorithm combines two RRTs connected at the start and goal configurations and was originally designed for 7 DOF arms motion planning. Due to the randomized nature of this state-of-the-art method, valid trajectories are not guaranteed. To enhance the probability of getting valid trajectories the gripper approach angle is modified after 40 attempts to find a suitable trajectory. After 200 tries the system notifies that no suitable trajectory was found. Due to accurate fine positioning and the variation of the gripper approach angle, trajectories are generally found by our system. These results were hard to achieve by our previous version based on global navigation only.

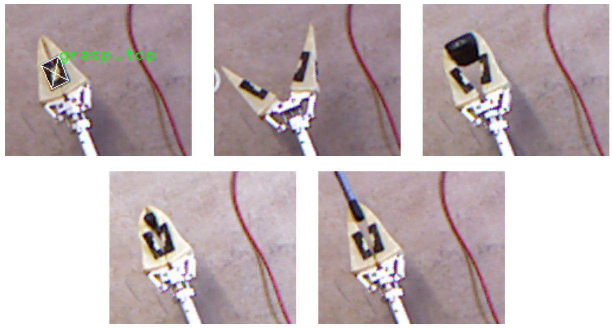


FIGURE 3. To decide if an object was grasped, we check whether a pattern can be detected on the gripper. The marker will only be detected if the gripper is closed without an object in between the fingers (see the first picture). In any other case the marker cannot be detected (see pictures 2-5).

F. VISUAL GRASP SUCCESS EVALUATION

We developed three methods for grasp success evaluation based on vision. For the first method an easy-to-detect marker was cut and fixed on an elastic overlay specifically developed for the fin-ray fingers of our robot [37]. The marker parts were placed so that the pattern can be observed only when the gripper is closed (see Fig. 3). If an object is grasped there must be a gap between the fingers and hence the marker will not be detected, which is a first indication of a successful grasp.

In rare situations the deformable fingers kept a slightly curved shape even without holding an object, which made us switch to an even more reliable grasp detection method. This second method is based on the detection of the gripper from visual RGB-D cues, directly checking if an object is detected between the gripper fingers [38].

A third method relies on moving the last link of the arm to check if there is still an object at the place where it was before (see Fig. 4). For the experiments in Section VI we used this

(a) Grasping the object (b) Lifting it up (c) Moving the object

FIGURE 4. Checking whether grasping was successful or not. (a) Grasping attempt. (b) The object is lifted. (c) The object is moved forward to check if something has changed at the previous position of the object on the floor.

last method only, since it is completely sufficient for the test scenarios with single objects on the floor and it showed the highest reliability.

VI. TESTS AND EXPERIMENTS

The system was tested in the lab, in several event demos and in real home user trials. Different challenges were encountered and methods to deal with them were designed, implemented, adjusted and tested.

A. LAB EXPERIMENTS

In the lab environment we executed a number of experiments to evaluate different aspects of our system.

1) EXPERIMENT 1

In the first experiment we mainly tested the global navigation capabilities, conducting ten tests with the robot. The object to be grasped was a piece of foam with dimensions 4 cm × 5 cm × 12 cm which was placed exactly 2m in front of the robot. A perfect gesture command was sent to the robot (no real detection), simulating a user who would exactly point at the position of the object. In all the experiments the robot was able to follow a path to the object, i.e. to a position from where it could find and properly segment the object. However, the final position reached by the robot by means of global navigation was not accurate enough with respect to the object for the limited workspace described in V-E. In eight out of the ten test runs the robot had to be manually turned a bit by the experimenter until the object could be found in the area where grasping was possible. Then the object was autonomously grasped by the robot in all the tests. In one out of the ten trials the first grasp failed (probably because the scene was perceived while the robot was still being manually rotated). Since the checking procedure described in V-F detected the failure, the grasp was repeated and the second attempt was successful.

2) EXPERIMENT 2

In a second experiment we calculated a suitable position for grasping relative to a found object. The robot can then reach this position from the global navigation goal by rotating the given angle and moving forward the given distance (as long as no obstacles are in the way). A final rotation may also be applied if necessary.

Ten tests were again conducted for the same object as in the first experiment. The position of the object and the gesture were fixed and provided likewise. In nine out of the ten runs the whole process was successfully completed in a fully autonomous manner: the object was reached, found, accurately approached for grasping, accepted, grasped, verified and moved to the robot tray. In one of these nine trials the robot had to grasp a second time, after the gripper failed to move towards the floor (probably due to a hardware issue). In the tenth run, the experimenter was forced to turn to robot a bit, until it could see the object. After this help, the robot proceeded autonomously and successfully completed the task. A subsequent analysis of this single failure showed that a rotation angle of $-9.48°$ was calculated for the fine positioning movement. It turned out that the minimal degree for discrete motion rotations in MIRA was set to 10°, so the robot did not turn at all and slightly missed a suitable position for grasping. For further test runs this limitation could be easily solved.

3) EXPERIMENT 3

In the third experiment the object to grasp was a 3 cm × 5 cm × 9 cm Aspirin box lying on the floor. This time we varied the position of the obstacle around 50cm in each direction (except the z-axis) and let the experimenter point to the object, so the system had to detect the pointing gesture itself. We conducted ten test runs. From the first five runs four were successful and one time the arm did not react when it should go down, which caused the termination of this run without any more attempts.

For the last five runs the gripper touched the object before trying to grasp it. Collisions between the gripper and the object would not be possible if the object point cloud data were complete and accurate, since then the simulation would stop the arm before a collision could take place. Besides camera calibration errors (especially related to head camera pitch angle servo values), an important cause for premature collisions between gripper and object was the acquisition of incomplete object surface data. Fig. 5 shows an example of a bad quality data point cloud that was perceived from the Aspirin box in these tests. This experiment was conducted in a corridor with direct sun incidence and we concluded that the lightning conditions were the main reason for these problematic perception results.

4) EXPERIMENT 4

For autonomous robots safety is crucial. To decide and evaluate when it is safe to avoid grasping or robot movement is even more important than successful grasps, since incautious robot movements can cause material damage of household items or the robot hardware itself, and in the worst case can injure the target users, who are physically not in the best conditions.

In experiment number four we tested if objects are graspable. For grasping from the floor the system first detects

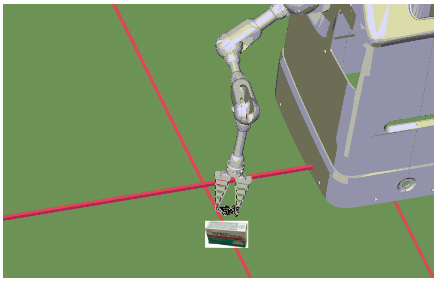


FIGURE 5. Perceived object mesh for the Aspirin box. Next to it we added the approximate orientation and size of the real box. Due to the missing surface a suboptimal grasp was detected, which led to premature collisions between the gripper and the object.

FIGURE 6. Objects which are too close to a wall or a fixed obstacle should not be grasped.

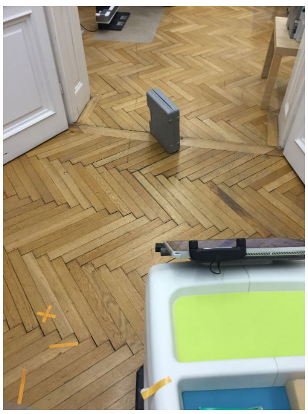
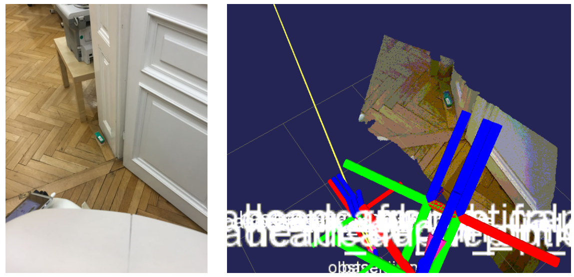


FIGURE 7. The used suitcase is too large for grasping and is properly excluded by the feasibility check.



(a) Object to grasp (b) Scene from head view

FIGURE 8. The foot of the small table was partially perceived. Only by checking if the object is near a wall or a fixed obstacle the system can rule the object out as feasible for grasping (after the corresponding approximation to the object).

(a) A chair blocking the object. (b) A user blocking the robot arm.

FIGURE 9. Situations in which additional obstacles or a user make grasping unsafe.

and removes points from the floor plane and segments the remaining data. The system decides whether the segmented objects are suitable for grasping by criteria such as position, maximal height, minimal and maximal number of received points per segmented object, distance to walls or other fixed obstacles. For this experiment we placed one or two objects near a door or a wall (see Fig. 6) and used a heavy big suitcase (see Fig. 7) to test the grasping feasibility decision making of the robot.

In Fig. 8 we show one reason why it is so important to exclude objects near fixed obstacles such as a wall, and thereby define regions for grasping. We can see the foot of a small table that is partially perceived by the robot. The perceived part has a valid height and size due to fact that the rest of the small table was not perceived because it was not in the field of view.

In all ten test runs the system decided not to grasp an object, due to its size or distance to the wall. In two cases the object was placed directly at the wall and was therefore not segmented and hence not eligible for grasping. The experiment also showed that objects that are not intended to be grasped are often segmented (in the experiments up to 6 objects were segmented). These objects can often be rejected due to the position where they are found, since when the robot starts to search for the object it already knows roughly where it is supposed to be.

5) EXPERIMENT 5

In the last experiment we tested if the robot avoids grasping when an additional obstacle such as a chair (9(a)) or a user (9(b)) is blocking the space needed for grasping.

Ten tests were performed and in all ten test cases the arm movement was not executed due to the new obstacles located inside the working space of the robot arm.

Table 1 presents a summary of the previously described results for these experiments conducted in the lab environment. Hardware problems include critical and not critical ones, i.e. those that could not be overcome in a second attempt and those which were followed by another try. Successful

TABLE 1. Summary of lab experiments. More details and explanations can be found in the text.

	Description	Hardware problems	Human assistance needed	First attempt grasping failures	Successful cases
1	Global navigation, perfect gesture sent	0	8 (small rotation)	1	10
2	Global navigation + fine positioning, perfect gesture sent	1	1 (small rotation, fixed)	1	10
3	Global navigation + fine positioning, real gesture	6	0	0	4
4	Grasping avoidance due to distance or size	0	0	-	10
5	Grasping avoidance due to obstacles	0	0	-	10

(a) Grasping command (b) Pointing gesture (c) Starting to move (d) Avoiding obstacle

(e) Finding the object (f) After fine positioning (g) Grasping the object (h) Releasing the object

FIGURE 10. Grasping an object initially out of the field of view. Check the multimedia material for the whole process video.

cases include those that required human assistance as well as those that required a second grasping attempt, provided that they were considered good in the end.

B. GRASPING AN OBJECT INITIALLY OUT OF THE FIELD OF VIEW

Our approach for fully autonomous behaviour allows the robot to grasp objects which are not initially in the camera's field of view. This is an especially challenging task which requires proper integration of perception capabilities and grasping capabilities with navigation capabilities, including obstacle avoidance. An example of the whole process in this situation is depicted in Fig. 10. More information can be found in the multimedia material.

C. USER TRIALS

User trials with real users in their own apartments were conducted for up to three weeks per user in Austria, Sweden and Greece, including a total of eighteen users. For details regarding the whole system and the user trials information the reader is referred to [8].

Some of the user comments regarding grasping failures were: "I tried a couple of times but Hobbit never saw the object", "Hobbit is stupid, it can't see the object even though Hobbit turns its head in the right direction", "Why can't Hobbit see the object?", in accordance to the fact that most of the problems were indeed related to RGB-D perception and head calibration limitations.

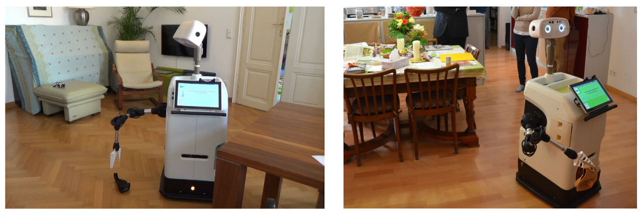


FIGURE 11. Examples of pick up tasks during pilot trials and review meetings.

Reasons why a whole pickup task was aborted may be divided into two main groups, either the user stopped the task or the robot stopped the task. Regarding the former, often the users activated other tasks or simply pressed the "Cancel" o "Back" button on the tablet. Regarding the latter, sometimes no valid pointing gesture was detected, sometimes no valid object was detected and sometimes a higher priority task was initiated.

Prior to conducting the user trials, several pilot studies and review meetings took place. Testing the system in real domestic environments in a supervised manner gave us insightful experience to deploy, check and evaluate the proposed integrated methods. Fig. 11 shows a couple of examples of successful picking up tasks in this kind of setting. The supplemental video shows an example of performance at a real user's home and Fig. 12 shows a few additional images of this and other pick up tasks in real elderly home environments.

Table 2 shows illustrative time measurements of different functions involved in a pickup task in real homes.

TABLE 2. Approximate time results for real experiments.

	Autonomous global navigation (s)	Fine positioning (s)	Grasp execution (s)	Success evaluation (s)	Release on tray (s)
1	11	10	50	15	16
2	15	11	46	15	18
3	8	-	38	15	15
4	12	11	40	15	15
5	8	-	37	14	15

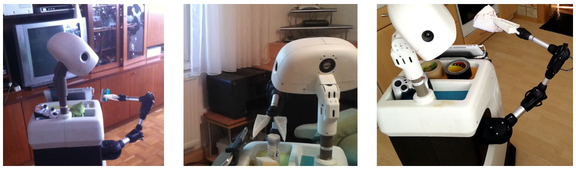


FIGURE 12. Examples of pick up tasks at actual users' homes.

The autonomous global navigation time provides an estimate of how far or hidden was the object to be grasped. The absence of data for fine positioning in cases three and five indicates that no further movement relative to the object was needed, even though some time was indeed spent in moving the head and deciding this. The grasp execution time is the time since the arm begins to move until it stops for the validation check. The time spent in success evaluation is the time required to move the arm and check if the object is still on the floor or not. The last column of the table shows how long it took to bring the object to the robot tray. Please note that head movements and transition times between functions are not included here.

VII. CONCLUSIONS AND FUTURE WORK

This paper addressed the problem of grasping objects from the floor in real world conditions, focusing on issues related to fully autonomous behaviour. The robot hardware and all the required components for a completely autonomous grasping scenario were described, together with the proposed architecture. The developed components include: pointing gestures perception, navigation including obstacle avoidance capabilities, fine positioning with respect to the object, grasp point detection, grasp planning and visual-based grasping success evaluation. Additional methods to check whether an object should be grasped or not in realistic applications were conceived, incorporated and evaluated in the lab environment. Examples of challenging situations and grasping tasks performed in real user homes were also shown.

Future research lines identified after conducting this work are related to further increasing the level of robustness and to making the system responses faster, always keeping safety as the main requirement.

REFERENCES

[1] A. Bilyea, N. Seth, S. Nesathurai, and H. A. Abdullah, "Robotic assistants in personal care: A scoping review," *Med. Eng. Phys.*, vol. 49, pp. 1–6, Nov. 2017.

[2] K. Zsiga, A. Tóth, T. Pilissy, O. Péter, Z. Dénes, and G. Fazekas, "Evaluation of a companion robot based on field tests with single older adults in their homes," *Assistive Technol.*, vol. 30, no. 5, pp. 259–266, 2017.

[3] C. Mucchiani, S. Sharma, M. Johnson, J. Sefcik, N. Vivio, J. Huang, P. Cacchione, M. Johnson, R. Rai, A. Canoso, T. Lau, and M. Yim, "Evaluating older adults' interaction with a mobile assistive robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 840–847.

[4] P. de la Puente, M. Bajones, C. Reuther, D. Wolf, D. Fischinger, and M. Vincze, "Robot navigation in domestic environments: Experiences using RGB-D sensors in real homes," *J. Intell. Robot. Syst.*, vol. 94, no. 2, pp. 455–470, 2018.

[5] N. Koceska, S. Koceski, P. B. Zobel, V. Trajkovik, and N. Garcia, "A telemedicine robot system for assisted and independent living," *Sensors*, vol. 19, no. 4, p. 834, 2019.

[6] D. Fischinger, A. Weiss, and M. Vincze, "Learning grasps with topographic features," *Int. J. Robot. Res.*, vol. 34, no. 9, pp. 1167–1194, 2015.

[7] K. de Miguel, A. Brunete, M. Hernando, and E. Gambao, "Home camera-based fall detection system for the elderly," *Sensors*, vol. 17, no. 12, p. 2864, 2017.

[8] M. Bajones *et al.*, "Hobbit: Providing fall detection and prevention for the elderly in the real world," *J. Robot.*, vol. 2018, Jun. 2018, Art. no. 1754657.

[9] T. Sandy, A. Gawel, B. Sankaran, S. Maggs, and J. Nieto, "ICRA 2019 workshop on high accuracy mobile manipulation in challenging environments," 2019. [Online]. Available: https://rsl.ethz.ch/scientific-events/workshops/ICRA-2019/hammworkshop-icra19.html

[10] D. Fischinger and M. Vincze, "Empty the basket—A shape based learning approach for grasping piles of unknown objects," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2012, pp. 2051–2057.

[11] D. Fischinger and M. Vincze, "Shape based learning for grasping novel objects in cluttered scenes," in *Proc. 10th IFAC Symp. Robot Control (SYROCO)*, Sep. 2012, pp. 787–792.

[12] D. Fischinger, M. Vincze, and Y. Jiang, "Learning grasps for unknown objects in cluttered scenes," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2013, pp. 609–616.

[13] D. Fischinger, P. Einramhof, K. Papoutsakis, W. Wohlkinger, P. Mayer, P. Panek, S. Hofmann, T. Koertner, A. Weiss, A. Argyros, and M. Vincze, "Hobbit, a care robot supporting independent living at home: First prototype and lessons learned," *Robot. Auton. Syst.*, vol. 75, pp. 60–78, Jan. 2016.

[14] 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 *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2014, pp. 2587–2594.

[15] P. de la Puente, M. Bajones, C. Reuther, D. Fischinger, D. Wolf, and M. Vincze, "Experiences with RGB-D based navigation in real home robotic trials," in *Proc. OAGM-ARW Joint Workshop Comput. Vis. Robot.*, 2016, pp. 153–165.

[16] Z. Xu, T. Deyle, and C. C. Kemp, "1000 Trials: An empirically validated end effector that robustly grasps objects from the floor," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2009, pp. 2160–2167.

[17] C.-H. King, T. L. Chen, Z. Fan, J. D. Glass, and C. C. Kemp, "Dusty: An assistive mobile manipulator that retrieves dropped objects for people with motor impairments," *Disab. Rehabil., Assistive Technol.*, vol. 7, no. 2, pp. 168–179, 2012.

[18] F. Farahmand, M. T. Pourazad, and Z. Moussavi, "An intelligent assistive robotic manipulator," in *Proc. 27th Annu. Conf. IEEE Eng. Med. Biol.*, vol. 5, Jan. 2006, pp. 5028–5031.

[19] M. Palankar, K. J. De Laurentis, R. Alqasemi, E. Veras, R. Dubey, Y. Arbel, and E. Donchin, "Control of a 9-DoF wheelchair-mounted robotic arm system using a P300 brain computer interface: Initial experiments," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Feb. 2009, pp. 348–353.

[20] P. Schrock, F. Farelo, R. Alqasemi, and R. Dubey, "Design, simulation and testing of a new modular wheelchair mounted robotic arm to perform activities of daily living," in *Proc. IEEE Int. Conf. Rehabil. Robot. (ICORR)*, Jun. 2009, pp. 518–523.

[21] J. Xu, G. G. Grindle, B. Salatin, J. J. Vazquez, H. Wang, D. Ding, and R. A. Cooper, "Enhanced bimanual manipulation assistance with the personal mobility and manipulation appliance (PerMMA)," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2010, pp. 5042–5047.

[22] M. Gualtieri, J. Kuczynski, A. M. Shultz, A. T. Pas, R. Platt, and H. Yanco, "Open world assistive grasping using laser selection," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/Jun. 2017, pp. 4052–4057.

[23] A. Remazeilles, C. Leroux, and G. Chalubert, "SAM: A robotic butler for handicapped people," in *Proc. 17th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2008, pp. 315–321.

[24] S. S. Srinivasa, D. Ferguson, C. J. Helfrich, D. Berenson, A. Collet, R. Diankov, G. Gallagher, G. Hollinger, J. Kuffner, and M. V. Weghe, "HERB: A home exploring robotic butler," *Auto. Robots*, vol. 28, pp. 5–20, Jan. 2010.

[25] B. Graf, U. Reiser, M. Hägele, K. Mauz, and P. Klein, "Robotic home assistant Care-O-bot3—Product vision and innovation platform," in *Proc. IEEE Workshop Adv. Robot. Social Impacts (ARSO)*, Nov. 2009, pp. 139–144.

[26] GarageWillow. (2011). *Beer Me, Robot*. [Online]. Available: http://www.willowgarage.com/blog/2010/07/06/beer-me-robot

[27] F. Lévesque, B. Sauvet, P. Cardou, and C. Gosselin, "A model-based scooping grasp for the autonomous picking of unknown objects with a two-fingered gripper," *Robot. Auton. Syst.*, vol. 106, pp. 14–25, Aug. 2018.

[28] S. H. Kasaei, M. Oliveira, G. H. Lim, L. S. Lopes, and A. M. Tomé, "Towards lifelong assistive robotics: A tight coupling between object perception and manipulation," *Neurocomputing*, vol. 291, pp. 151–166, May 2018.

[29] Y. S. Choi, C. D. Anderson, J. D. Glass, and C. C. Kemp, "Laser pointers and a touch screen: Intuitive interfaces for autonomous mobile manipulation for the motor impaired," in *Proc. 10th Int. ACM SIGACCESS Conf. Comput. Accessibility*, 2008, pp. 225–232. [Online]. Available: http://portal.acm.org/citation.cfm?id=1414471.1414512

[30] A. Jain and C. C. Kemp, "EL-E: An assistive mobile manipulator that autonomously fetches objects from flat surfaces," *Auton. Robots*, vol. 28, pp. 45–64, Jan. 2010.

[31] K. Tai, A.-R. El-Sayed, M. Shahriari, M. Biglarbegian, and S. Mahmud, "State of the art robotic grippers and applications," *Robotics*, vol. 5, no. 2, p. 11, 2016.

[32] D. Michel, K. Papoutsakis, and A. A. Argyros, "Gesture recognition supporting the interaction of humans with socially assistive robots," in *Advances in Visual Computing*. Cham, Switzerland: Springer, 2014, pp. 793–804.

[33] E. Einhorn, T. Langner, R. Stricker, C. Martin, and H.-M. Gross, "MIRA—middleware for robotic applications," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2012, pp. 2591–2598.

[34] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with Rao–Blackwellized particle filters," *IEEE Trans. Robot.*, vol. 23, no. 1, pp. 34–46, Feb. 2007.

[35] R. Diankov and J. Kuffner, "OpenRAVE : A planning architecture for autonomous robotics," Robotics Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-08-34, Jul. 2008.

[36] J. J. Kuffner and S. LaValle, "RRT-connect: An efficient approach to single-query path planning," in *Proc. Millennium Conf. IEEE Int. Conf. Robot. Automat. Symposia (ICRA)*, vol. 2, Apr. 2000, pp. 995–1001.

[37] C. Künzel, "Visuelle Erkennung erfolgreicher Greifversuche für die Mobile Robotik mittels Pattern-Matching-Algorithmus," M.S. thesis, Vienna Univ. Technol., Vienna, Austria, 2016.

[38] E. Martinez-Martin, D. Fischinger, M. Vincze, and A. P. del Pobil, "An RGB-D visual application for error detection in robot grasping tasks," in *Intelligent Autonomous Systems 14*. Cham, Switzerland: Springer, 2017, pp. 243–254.

PALOMA DE LA PUENTE received the engineering degree in automatic control and electronics, the M.Sc., and the Ph.D. degrees in robotics and automation from the Universidad Politécnica de Madrid (UPM), in 2007, 2008, and 2012, respectively. She was at Caltech for a research stay as a Predoctoral Visitor for 2+3 months. She was a Postdoctoral Researcher at DISAM-UPM and at the ACIN Institute of Automation and Control-Vienna University of Technology (TUW). She also has professional experience with Ixion Industry and Aerospace. She is currently an Assistant Professor with UPM. Her current research interests include mobile robots navigation, mapping, SLAM, spatial cognition, sensor data processing, human-robot interaction for service robotics, and systems engineering. She has participated in several national and European projects and also in international robotics competitions.

DAVID FISCHINGER received the bachelor's degree in technical mathematics from the Vienna University of Technology, in 2005, and the master's degrees (Hons.) in computer science management and in computational intelligence (informatics), in 2007, and the Ph.D. degree from the Vision for Robotics (V4R) Laboratory, Vienna University of Technology, under the supervision of Prof. M. Vincze, in 2014. From 2007 to 2010, he was a Senior Programmer at the management and consulting company Accenture in Munich, Vienna, and Hyderabad (India). In 2017, he started as a Manager of research and development at Aeolus Robotics, Inc. and setup the research location in Austria with a focus on robotic grasping. His current research interests include robotic grasping, machine learning, and vision for robotics.

MARKUS BAJONES received the degree in electrical engineering from the Vienna University of Technology, in 2013, and the Ph.D. degree from the Vision for Robotics (V4R) Laboratory, Vienna University of Technology, under the supervision of Prof. M. Vincze's. In 2018, he continued his work on human-robot interaction at Aeolus Robotics, Inc. His current research interests include human-robot interaction in elderly care, behavior coordination for social robots, and machine learning.

DANIEL WOLF received the diploma degree in electrical engineering and information technology from the Technical University of Munich (TUM), in 2012, and the Ph.D. degree from TU Wien (TUW), specializing on semantic segmentation and 3D scene understanding for indoor robotics, in 2017. He is currently with the DAQRI Holographics Research Center, Vienna.

MARKUS VINCZE received the diploma degree in mechanical engineering from Technical University Wien (TUW), in 1988, and the M.Sc. degree from the Rensselaer Polytechnic Institute, USA, in 1990, and the Ph.D. degree from TUW, in 1993, and the habilitation degree in robotics, in 2004. With a grant from the Austrian Academy of Sciences, he worked at HelpMate Robotics Inc., and at the Vision Laboratory of Gregory Hager at Yale University. He currently leads the Vision for Robotics (V4R) Team at TUW with the vision to make robots see. V4R regularly coordinates EU and national research projects and contributes to research and innovation projects. With Gregory Hager, he edited a book on *Robust Vision* (IEEE). He is the (co)author of 58 peer reviewed journal articles and over 400 reviewed other publications. His special interest includes cognitive computer vision techniques for robotics solutions situated in real-world environments and, especially, homes. He was the Program Chair of ICRA 2013 in Karlsruhe and organized HRI 2017 in Vienna.