

Mobile manipulation, tool use, and intuitive interaction for cognitive service robot Cosero

Jörg Stückler, Max Schwarz, Sven Behnke

Angaben zur Veröffentlichung / Publication details:

Stückler, Jörg, Max Schwarz, and Sven Behnke. 2016. "Mobile manipulation, tool use, and intuitive interaction for cognitive service robot Cosero." *Frontiers in Robotics and AI* 3: 58.
<https://doi.org/10.3389/frobt.2016.00058>.



Mobile Manipulation, Tool Use, and Intuitive Interaction for Cognitive Service Robot Cosero

Jörg Stückler, Max Schwarz and Sven Behnke*

Institute for Computer Science VI, Autonomous Intelligent Systems, University of Bonn, Bonn, Germany

Cognitive service robots that shall assist persons in need of performing their activities of daily living have recently received much attention in robotics research. Such robots require a vast set of control and perception capabilities to provide useful assistance through mobile manipulation and human–robot interaction. In this article, we present hardware design, perception, and control methods for our cognitive service robot Cosero. We complement autonomous capabilities with handheld teleoperation interfaces on three levels of autonomy. The robot demonstrated various advanced skills, including the use of tools. With our robot, we participated in the annual international RoboCup@Home competitions, winning them three times in a row.

OPEN ACCESS

Edited by:

Jochen J. Steil,
Bielefeld University, Germany

Reviewed by:

Nikolaus Vahrenkamp,
Karlsruhe Institute of Technology,
Germany
Feng Wu,
University of Science and Technology
of China, China

*Correspondence:

Sven Behnke
behnke@cs.uni-bonn.de

Specialty section:

This article was submitted to
Humanoid Robotics,
a section of the journal
Frontiers in Robotics and AI

Received: 29 May 2016

Accepted: 20 September 2016

Published: 07 November 2016

Citation:

Stückler J, Schwarz M and Behnke S
(2016) Mobile Manipulation, Tool
Use, and Intuitive Interaction for
Cognitive Service Robot Cosero.
Front. Robot. AI 3:58.
doi: 10.3389/frobt.2016.00058

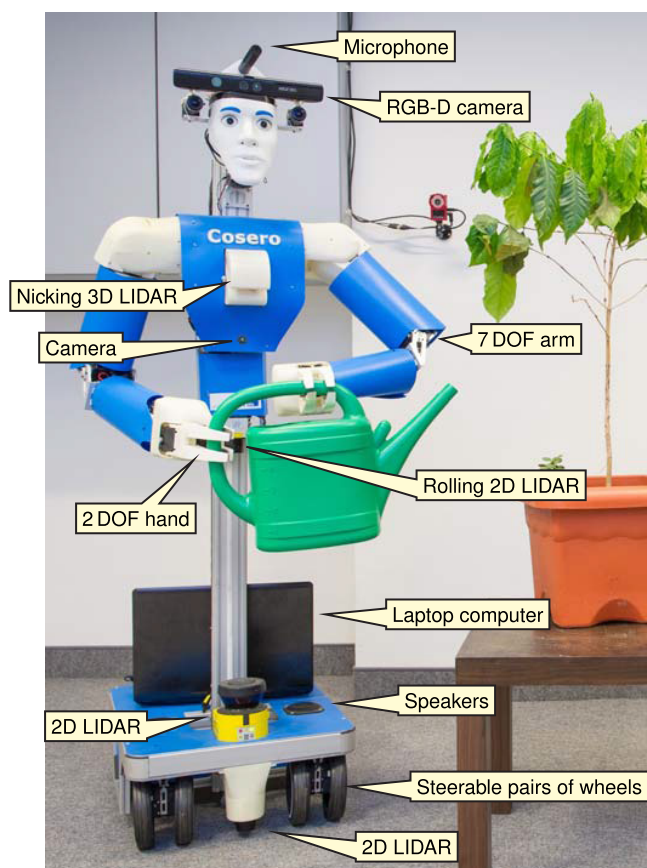
Keywords: cognitive service robots, mobile manipulation, object perception, human–robot interaction, shared autonomy, tool use

1. INTRODUCTION

In recent years, personal service robots that shall assist, e.g., handicapped or elderly persons in their activities of daily living have attracted increasing attention in robotics research. The everyday tasks that we perform, for instance, in our households, are highly challenging to achieve with a robotic system, because the environment is complex, dynamic, and structured for human needs. Autonomous service robots require versatile mobile manipulation and human–robot interaction skills in order to really become useful. For example, they should fetch objects, serve drinks and meals, and help with cleaning. Many capabilities that would be required for a truly useful household robot are still beyond the state-of-the-art in autonomous service robotics. Complementing autonomous capabilities of the robot with user interfaces for teleoperation enables the use of human cognitive abilities whenever autonomy reaches its limits and, thus, could bring such robots faster toward real-world applications.

We develop cognitive service robots since 2009, according to the requirements of the annual international RoboCup@Home competitions (Wisspeintner et al., 2009). These competitions benchmark integrated robot systems in predefined test procedures and in open demonstrations in which teams can show the best of their research. Benchmarked skills comprise mobility in dynamic indoor environments, object retrieval and placement, person perception, complex speech understanding, and gesture recognition.

In previous work, we developed the communication robot Robotinho (Nieuwenhuisen and Behnke, 2013) and our first-generation domestic service robot Dynamaid (Stückler and Behnke, 2011). In this article, we focus on the recent developments targeted at our second-generation cognitive service robot Cosero, shown in **Figure 1**. Our robot is mobile through an omnidirectional wheeled drive. It is equipped with an anthropomorphic upper body with 7 degree of freedom (DoF) arms that have human-like reach and features a communication head with an RGB-D camera and a directed microphone.



Environment Perception

- 2D laser SLAM
- 3D navigation (Sec. 5)
- 3D object segmentation (Sec. 6.1.1)
- 3D object recognition (Sec. 6.1.2)
- 3D object pose estimation (Sec. 6.1.3)
- 3D object tracking (Sec. 6.1.5)

Manipulation

- Task-space compliance (Sec. 4.2)
- Motion primitives
- Grasp planning (Sec. 6.2)
- Bin picking (Sec. 6.3)
- Tool use (Sec. 6.4)

Human-Robot-Interaction

- Person tracking (Sec. 7.1.1)
- Gesture recognition (Sec. 7.1.3)
- Speech commands (Sec. 7.2.2)
- Physical interaction and cooperation (Sec. 7.3.3)
- Teleoperation interfaces (Sec. 7.4)

FIGURE 1 | Overview of the cognitive service robot Cosero. The right column lists main capabilities of the system and references their corresponding sections in this article.

An overview of Cosero's capabilities is given in **Figure 1**. We discuss related work in the next section. Section 3 details Cosero's mechatronic design. The overall software architecture, low-level control for omnidirectional driving, and compliant manipulation are presented in Section 4. Section 5 covers mapping, localization, and navigation in dynamic indoor environments. The perception of objects and different manipulation skills, including tool use and skill transfer to novel instances of a known object category, are presented in Section 6. Different human-robot interaction modules, including the recognition of gestures and the understanding of speech commands, are discussed in Section 7. We report results obtained in the RoboCup@Home competitions 2011–2014 in Section 8.

2. RELATED WORK

An increasing number of research groups worldwide are working on complex robots for domestic service applications. The Armar III robot developed at KIT (Asfour et al., 2006) is an early prominent example. Further robots with a humanoid upper body design are the robots Twendy-One (Iwata and Sugano, 2009) developed

at Waseda University, and the CIROS robots developed at KIST (Sang, 2011). The Personal Robot 2 [PR2 (Meeussen et al., 2010)], developed by Willow Garage, was adopted by multiple research groups and led to wide-spread use of the ROS middleware (Quigley et al., 2009). The robot has two 7 DOF compliant arms on a liftable torso. Its omnidirectional drive has four individually steerable wheels, similar to our robot. PR2 is equipped with a variety of sensors such as a 2D laser scanner on the mobile base, a 3D laser scanner in the neck, and a structured light stereo camera rig in the head. Bohren et al. (2011) demonstrated an application in which a PR2 fetched drinks from a refrigerator and delivered them to human users. Both the drink order and the location at which it had to be delivered were specified by the user in a web form. Beetz et al. (2011) used a PR2 and a custom-built robot to cooperatively prepare pancakes. Nguyen et al. (2013) proposed a user interface for flexible behavior generation within the ROS framework for the PR2. The interface allows for configuring behavior in hierarchical state machines and for specifying goals using interactive markers in a 3D visualization of the robot and its internal world representation.

An impressive piece of engineering is the robot Rollin' Justin (Borst et al., 2009), developed at DLR, Germany. Justin is equipped

with larger-than-human compliantly controlled light weight arms and two four-finger hands. The upper body is supported by a four-wheeled mobile platform with individually steerable wheels, similar to our design. The robot demonstrated several dexterous manipulation skills such as pouring of a drink into a glass or cleaning windows. It also prepared coffee in a pad machine (Bäumel et al., 2011). For this, the robot grasped coffee pads and inserted them into the coffee machine, which involved opening and closing the pad drawer. For cleaning windows, Leidner et al. (2014) proposed an object-centered hybrid reasoning approach that parametrizes tool-use skills for the situation at hand.

The robot Herb (Srinivasa et al., 2010), jointly developed by Intel Research Labs and Carnegie Mellon University, manipulated doors and cabinets. It is equipped with a single arm and uses a Segway platform as drive. In the healthcare domain, Jain and Kemp (2010) present EL-E, a mobile manipulator that assists motor-impaired patients by performing pick and place operations to retrieve objects. The Care-O-Bot 3 (Parlitz et al., 2008) is a domestic service robot developed at Fraunhofer IPA. The robot is equipped with four individually steerable wheels, a 7 DOF lightweight manipulator, and a tray for interaction with persons. Objects are not directly passed from the robot to persons but placed on the tray. This concept was recently abandoned in favor of a two-armed, more anthropomorphic design in its successor Care-O-Bot 4 (Kittmann et al., 2015). The robot HoLLie (Hermann et al., 2013) developed at the FZI Karlsruhe also has an omnidirectional drive and is equipped with two 6 DoF arms with anthropomorphic hands. A bendable trunk allows HoLLie to pick objects from the floor.

While the above systems showed impressive demonstrations, the research groups frequently focus on particular aspects and neglect others. Despite many efforts, so far, no domestic service robot has been developed that fully addresses the functionalities needed to be useful in everyday environments. To benchmark progress and to facilitate research and development, robot competitions gained popularity in recent years (Behnke, 2006).

For service robots, the RoboCup Federations hold annual competitions in its @Home league (Iocchi et al., 2015). Systems competing in the most recent competition, which was held 2015 in Suzhou, China, are described, e.g., by Chen et al. (2015), Seib et al. (2015), zu Borgsen et al. (2015), and Lunenburg et al. (2015). Most of these custom-designed robots consist of a wheeled mobile base with laser and RGB-D sensors and a single manipulator arm.

Competitions in different domains include RoboCup Humanoid Soccer (Gerndt et al., 2015), the DARPA Robotics Challenge (Guizzo and Ackerman, 2015), and the DLR SpaceBot Cup (Stückler et al., 2016). Recently, Atlas – an impressive hydraulic humanoid robot developed by Boston Dynamics for the DARPA Robotics Challenge – has demonstrated some household chores, which were programed by Team IHMC (Ackerman, 2016).

3. MECHATRONIC DESIGN

For the requirements of domestic service applications, and in particular, the tasks of the RoboCup@Home competitions, we developed a robot system that balances the aspects of robust mobility, human-like manipulation, and intuitive human-robot interaction. Extending the functionality of its predecessor Dynamaid (Stückler and Behnke, 2011), we designed our cognitive service robot Cosero (see Figure 1) to cover a wide range of tasks in everyday indoor environments.

Figure 2 gives an overview of the mechatronic components of Cosero and shows their connectivity. The robot is equipped with an anthropomorphic torso and two 7 DoF arms that provide adult-like reach. Compared to Dynamaid, the arms are twice as strong and support a payload of 1.5 kg each. The grippers have twice as many fingers. They consist of two pairs of Festo FinGripper fingers – made from lightweight, deformable plastics material – on rotary joints. When a gripper is closed on an object, the bionic fin ray structure adapts the finger shape to the object surface. By this, the contact surface between fingers and object increases significantly compared to a rigid mechanical structure.

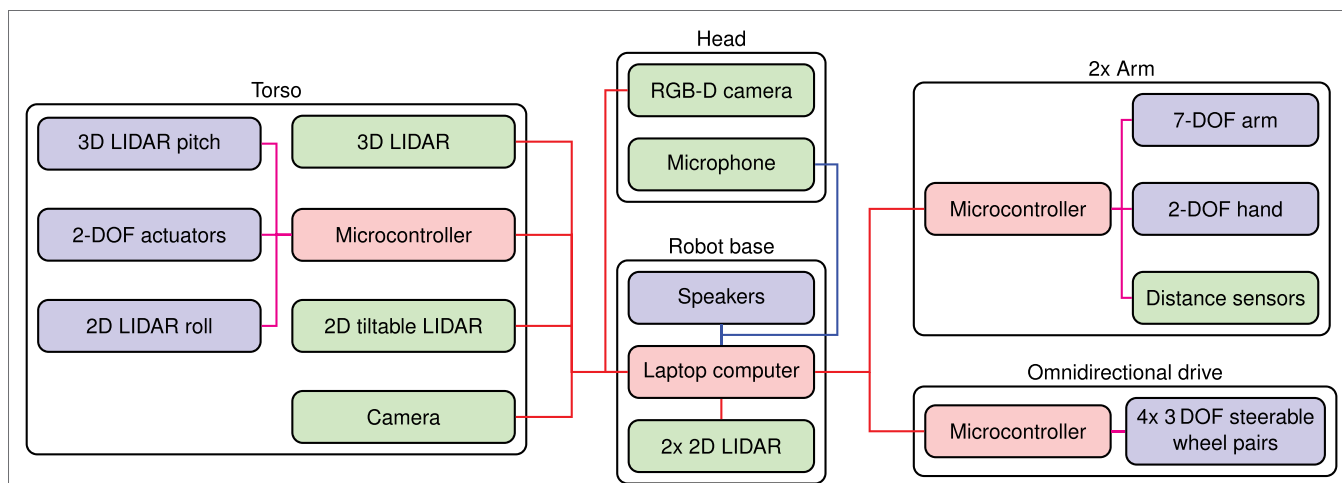


FIGURE 2 | Overview of the mechatronic design of Cosero. Sensors are colored green, actuators blue, and other components red. USB connections are shown in red, analog audio links are shown in blue, and the low-level servo connections are shown in magenta.

A thin layer of anti-skidding material on the fingers establishes a robust grip on objects. Having two fingers on each side of the gripper supports grasps stable for torques in the direction of the fingers and for forces in the direction between opponent fingers.

Cosero's torso can be twisted around the vertical axis to extend its work space. A linear actuator moves the whole upper body up and down, allowing the robot to grasp objects from a wide range of heights – even from the floor. Its anthropomorphic upper body is mounted on a base with narrow footprint and omnidirectional driving capabilities. By this, the robot can maneuver through narrow passages that are typically found in indoor environments, and it is not limited in its mobile manipulation capabilities by non-holonomic constraints. Many parts of the upper body, e.g., shoulders and wrist, are covered by 3D-printed shells. Together with human-like proportions and a friendly face, this contributes to the human-like appearance of our robot, which facilitates intuitive interaction of human users with the robot.

For perceiving its environment, we equipped the robot with multimodal sensors. Four laser range scanners on the ground, on top of the mobile base, and in the torso (rollable and pitchable) measure distances to objects, persons, or obstacles for navigation purposes. The head is mounted on a pan-tilt joint and features a Microsoft Kinect RGB-D camera for object and person perception in 3D and a directed microphone for speech recognition. A camera in the torso provides a lateral view onto objects in typical manipulation height.

A high-performance Intel Core-i7 quad-core notebook that is located on the rear part of the base is the main computer of the robot. Cosero is powered by a rechargeable 5-cell 12 Ah LiPo battery.

4. SOFTWARE AND CONTROL ARCHITECTURE

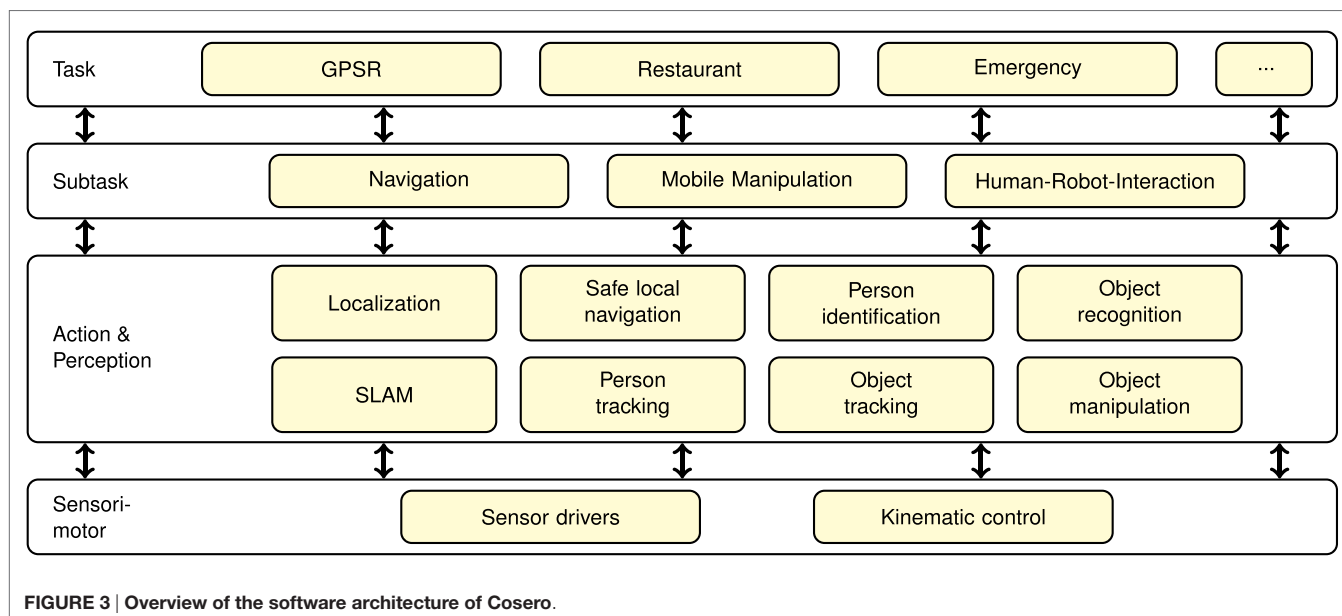
Cosero's autonomous behavior is generated in the modular multithreaded control framework ROS (Quigley et al., 2009). The key

motivation for using ROS is its large community, which continuously develops software modules and widens the scope of ROS. In this manner, ROS has become a standard for sharing research work related to robotics. The software modules for perception and control are organized in four layers, as shown in **Figure 3**. On the sensorimotor layer, data is acquired from the sensors and position targets are generated and sent to the actuators. The action and perception layer contains modules for person and object perception, safe local navigation, localization, and mapping. These modules process sensory information to estimate the state of the environment and generate reactive action. Modules on the subtask layer coordinate sensorimotor skills to achieve higher-level actions like mobile manipulation, navigation, and human-robot interaction. For example, the mobile manipulation module combines safe omnidirectional driving with object detection, recognition, and pose estimation, and with motion primitives for grasping, carrying, and placing objects. Finally, at the task layer, the subtasks are further combined to solve complex tasks that require navigation, mobile manipulation, and human-robot interaction.

This modular architecture reduces complexity by decomposing domestic service tasks into less complex modules. One organizing principle used is the successive perceptual abstraction when going up the hierarchy. On the other hand, higher-layer modules configure lower-layer modules to make abstract action decisions more concrete. Lower-layer modules are executed more frequently than modules on higher layers to generate real-time control commands for the actuators.

4.1. Omnidirectional Driving

As its predecessor Dynamaid (Stückler and Behnke, 2011), Cosero supports omnidirectional driving for flexible navigation in restricted spaces. The linear (\dot{x}, \dot{y}) and angular $\dot{\theta}$ velocities of the base can be set independently and can be changed continuously. This target velocity is mapped to steering angles and velocities of the individual pairs of wheels, so that the wheels are always rolling



into the local driving direction, tangential to the instantaneous center of rotation of the base.

Collision detection and avoidance using laser range sensors in the robot enables safe omnidirectional driving. Obstacle avoidance in 3D is performed by continuously tilting the laser scanner in the robot's chest. During safe collision-aware driving, linear and rotational velocities are limited when necessary to avoid the closest obstacles to the robot.

4.2. Compliant Arm Motion Control

For the anthropomorphic arms, we implemented differential inverse kinematics with redundancy resolution and compliant control in task space. We limit the torque of the joints according to how much they contribute to the achievement of the motion in task space. Our approach not only allows adjusting compliance in the null-space of the motion, but also in the individual dimensions in task space.

Our method determines a compliance $c \in [0,1]^n$ in linear dependency of the deviation of the actual state from the target state in task space, such that the compliance is one for small displacements and zero for large ones. For each task dimension, the motion can be set compliant in the positive and the

negative direction separately, allowing, e.g., for being compliant in upward direction, but stiff downwards. If a task dimension is not set compliant, we wish to use high holding torques to position control this dimension. If a task dimension is set compliant, the maximal holding torque interpolates between a minimal value for full compliance and a maximum torque for zero compliance. We then measure the responsibility of each joint for the task-space motion through the inverse of the forward kinematics Jacobian and determine a joint activation matrix from it. **Figure 4** shows an example matrix. The torque limits are distributed to the joints according to this activation matrix. Further details on our method can be found in Stückler and Behnke (2012).

We applied compliant control to the opening and closing of doors that can be moved without the handling of an unlocking mechanism. Refrigerators or cabinets are commonly equipped with magnetically locked doors that can be pulled open without special manipulation of the handle. To open a door, our robot approaches the handle, and grasps it. The drive moves backward while the gripper moves to a position on the side of the robot in which the opening angle of the door is sufficiently large to approach the open fridge or cabinet. The gripper follows the

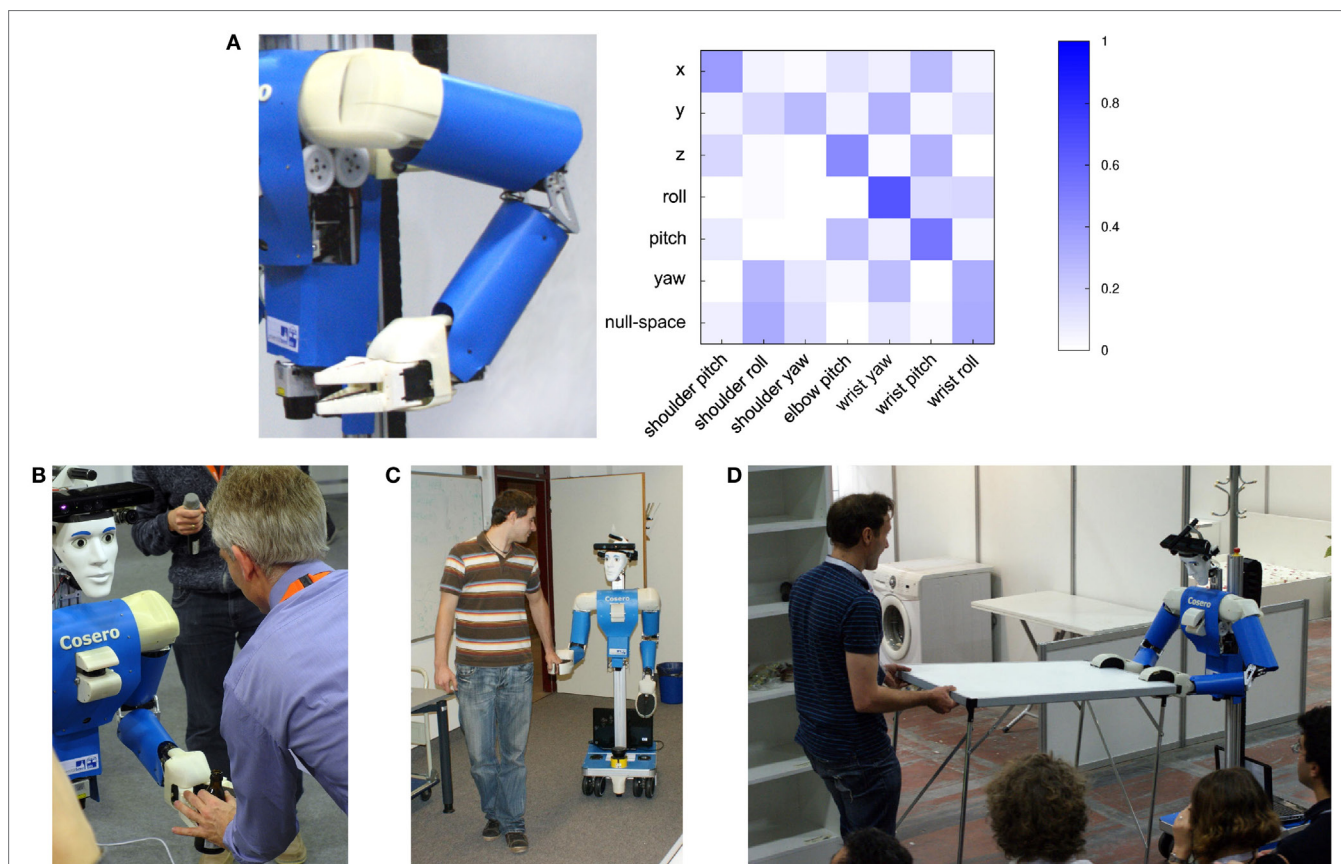


FIGURE 4 | Compliant arm control (Stückler and Behnke, 2012). (A) Activation matrix for compliant control in an example arm pose. The task-space dimensions correspond to forward/backward (x), lateral (y), vertical (z), and rotations around the x-axis (roll), y-axis (pitch), and z-axis (yaw); Examples for use of compliance in physical human–robot interaction: (B) object hand over from robot to user signaled by the user pulling on the object. (C) Cosero is guided by a person. (D) Cosero follows the motion of a person to cooperatively carry a table.

motion of the door handle through compliance in the lateral and the yaw directions. The robot moves backward until the gripper reaches its target position. For closing a door, the robot has to approach the open door leaf, grasp the handle, and move forward while it holds the handle at its initial grasping pose relative to the robot. Since the door motion is constrained by the hinge, the gripper will be pulled sideways. The drive corrects for this motion to keep the handle at its initial pose relative to the robot and thus follows the circular trajectory implicitly. The closing of the door can be detected when the arm is pushed back toward the robot.

5. NAVIGATION

Our robot navigates in indoor environments on horizontal surfaces. The 2D laser scanner on the mobile base is used as the main sensor for navigation. 2D occupancy maps of the environment are acquired using simultaneous localization and mapping [gMapping (Grisetti et al., 2007)]. The robot localizes in these maps using Monte Carlo localization (Fox, 2001) and navigates to goal poses by planning obstacle-free paths in the environment map, extracting waypoints, and following them. Obstacle-free local driving commands are derived from paths that are planned toward the next waypoint in local collision maps acquired with the robot's 3D laser scanners.

Solely relying on a 2D map for localization and path planning, however, has several limitations. One problem of such 2D maps occurs in path planning, if non-traversable obstacles cannot be perceived on the height of a horizontal laser scanner. Localization with 2D lasers imposes further restrictions if dynamic objects occur, or the environment changes in the scan plane of the laser. Then, localization may fail since large parts of the measurements are not explained by the map. To address such shortcomings, we realized localization and navigation in 3D maps of the environment.

We choose to represent the map in a 3D surfel grid, which the robot acquires from multiple 3D scans of the environment. **Figure 5** demonstrates an example map generated with our approach. From the 3D maps, we extract 2D navigation maps by exploring the traversability of surfels. We check height changes between surfels, which exceed a threshold of few centimeters, defining traversability of the ground, and for obstacles within the robot height.

For localization in 3D surfel grid maps, we developed an efficient Monte Carlo method that can incorporate full 3D scans and 2D scan lines. When used with 3D scans, we extract surfels from the scans and evaluate the observation likelihood. From 2D scans, we extract line segments and align them with surfels in the map. Localization in 3D maps is specifically useful in crowded environments. The robot can then focus on measurements above the height of people to localize at the static parts of the environment. More general, by representing planar surface elements in the map, we can also rely for localization mainly on planar structures, as they more likely occur in static environment parts. For further details, please refer to Kläss et al. (2012).

6. MANIPULATION

Service tasks often involve the manipulation of objects and the use of tools. In this section, we present perceptual and action modules that we realized for Cosero. Several of these modules have been combined with the control strategies in the previous sections to implemented tool-use skills and other domestic service demonstrations.

6.1. Object Perception

When attempting manipulation, our robot captures the scene geometry and appearance with its RGB-D camera. In many situations, objects are located well separated on horizontal support surfaces, such as tables, shelves, or the floor. To ensure good visibility, the camera is placed at an appropriate height above and distance from the surface, pointing downwards with an angle of approximately 45°. To this end, the robot aligns itself with tables or shelves using the rollable laser scanner in its hip in its vertical scan plane position. **Figure 6A** shows a resulting screen capture.

6.1.1. Object Segmentation

An initial step for the perception of objects in these simple scenes is to segment the captured RGB-D images into support planes and objects on these surfaces. Our plane segmentation algorithm rapidly estimates normals from the depth images of the RGB-D camera and fits a horizontal plane through the points with roughly vertical normals by RANSAC (Stückler et al., 2013b). The points above the detected support plane are grouped to object candidates based on Euclidean distance. All points within a range threshold form a segment that is analyzed separately. In **Figure 6A**, the

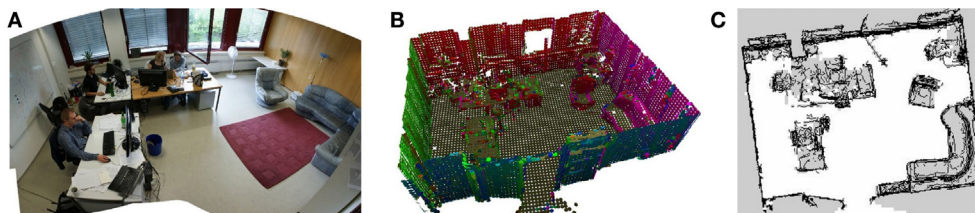
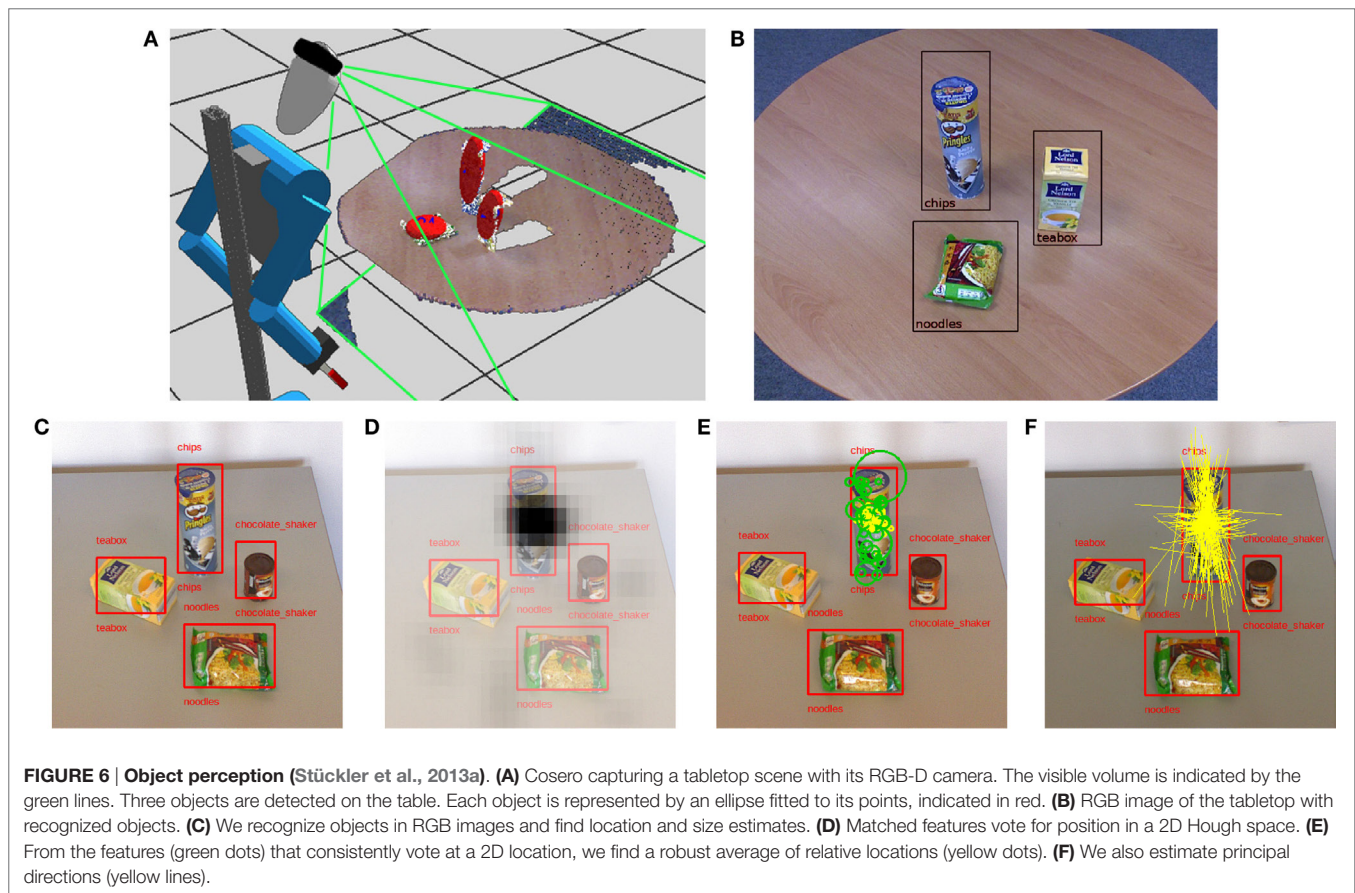


FIGURE 5 | 3D Surfel grid maps for navigation (Kläss et al., 2012). (A) Panorama image of an office. **(B)** 3D surfel map learned with our approach (surfel orientation coded by color). **(C)** Navigation map derived from 3D surfel map.



detected segments are visualized using an ellipse fitted to their 3D points.

6.1.2. Object Recognition

Cosero recognizes objects by matching SURF interest points (Bay et al., 2008) in RGB images to an object model database (Stückler et al., 2013a). An example recognition result is shown in **Figure 6B**. We store in an object model the SURF feature descriptors along with their scale, orientation, relative location of the object center, and orientation and length of principal axes. The bottom row of **Figure 6** illustrates the recognition process. We efficiently match features between the current image and the object database according to the descriptor using kd-trees. Each matched feature then casts a vote to the relative location, orientation, and size of the object. We consider the relation between the feature scales and orientation of the features to achieve scale- and rotation-invariant voting. Hence, our approach also considers geometric consistency between features. When unlabeled object detections are available through planar RGB-D segmentation (see **Figure 6C**), we project the detections into the image and determine the identity of the object in these regions of interest.

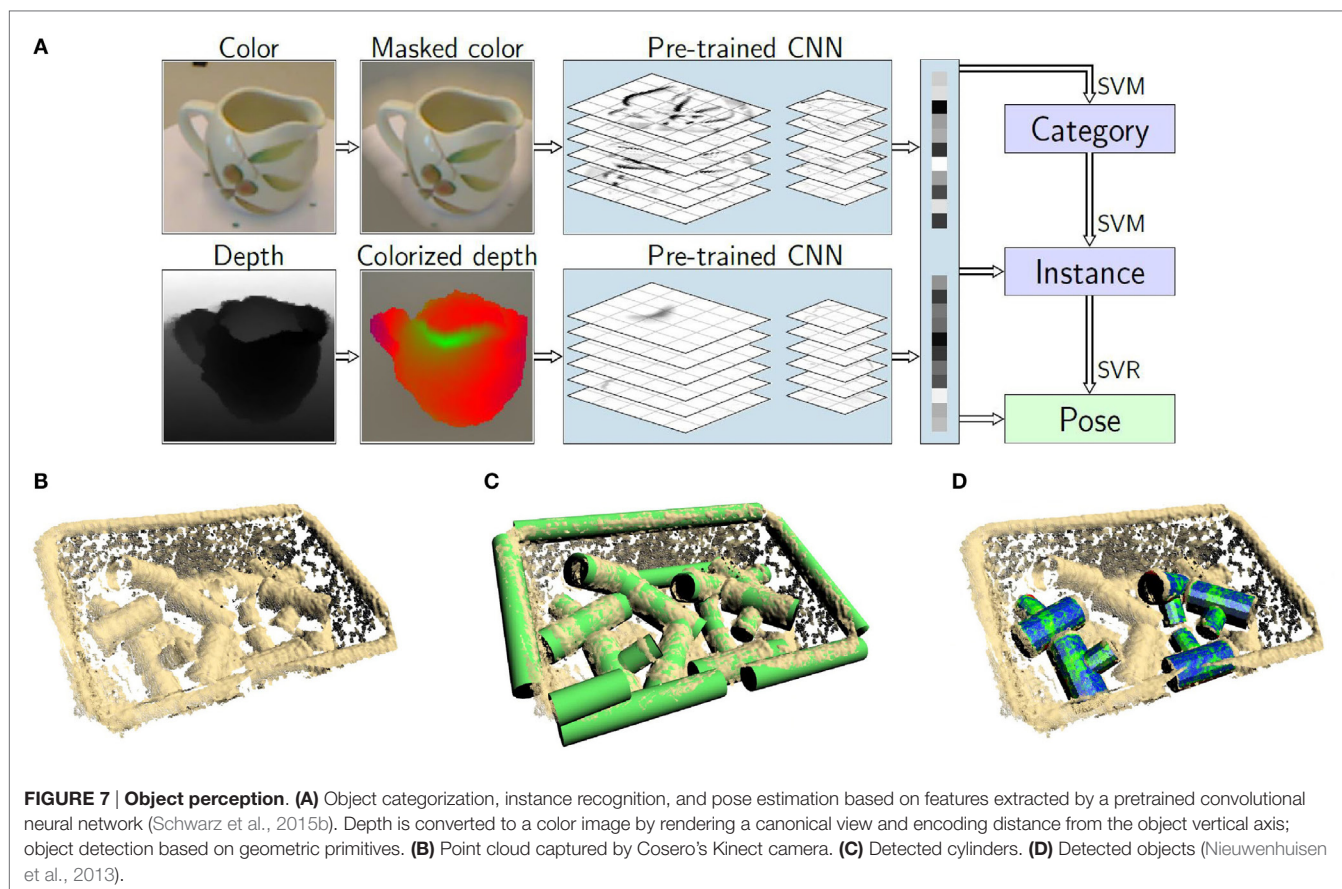
6.1.3. Object Categorization and Pose Estimation

For categorizing objects, recognizing known instances, and estimating object pose, we developed an approach that analyzes

an object, which has been isolated using tabletop segmentation. The RGB-D region of interest is preprocessed by fading out the background of the RGB image (see **Figure 7A**, top left). The depth measurements are converted to an RGB image as well by rendering a view from a canonical elevation and encoding distance from the estimated object vertical axis by color, as shown in **Figure 7A** bottom left. Both RGB images are presented to a convolutional neural network, which has been pretrained on the ImageNet data set for categorization of natural images. This produces semantic higher-layer features, which are concatenated and used to recognize object category, object instance, and to estimate the azimuth viewing angle onto the object using support vector machines and support vector regression, respectively. This transfer learning approach has been evaluated on the popular Washington RGB-D Object data set and improved the state-of-the-art (Schwarz et al., 2015b).

6.1.4. Primitive-based Object Detection

Objects are not always located on horizontal support surfaces. For a bin picking demonstration, we developed an approach to detect known objects, which are on top of a pile, in an arbitrary pose in transport boxes. The objects are described by a graph of shape primitives. The bottom row of **Figure 7** illustrates the object detection process. First, individual primitives, like cylinders of appropriate diameter are detected using RANSAC. The relations



between these are checked. If they match, the graph describing the object model, an object instance is instantiated, verified, and registered to the supporting 3D points. This yields object pose estimates in 6D. Further details are described in Nieuwenhuisen et al. (2013), who demonstrated mobile bin picking with Cosero. The method has been extended to the detection of object models that combine 2D and 3D shape primitives by Berner et al. (2013).

6.1.5. Object Tracking

Cosero tracks the pose of known objects using models represented as multiresolution surfel maps [MRSMs (Stückler and Behnke, 2014c)], which we learn from moving an RGB-D sensor around the object and performing SLAM. Our method estimates the camera poses by efficiently registering RGB-D key frames. After loop closing and globally minimizing the registration error, the RGB-D measurements are represented in a multiresolution surfel grid, stored as an octree. Each volume element represents the local shape of its points as well as their color distribution by a Gaussian. Our MRSMs also come with an efficient RGB-D registration method, which we use for tracking the pose of objects in RGB-D images. The object pose can be initialized using our planar segmentation approach. Figures 8A,B illustrates the tracking with an example. To handle difficult situations, like occlusions, McElhone et al. (2013) extended this approach to joint detection and tracking of objects modeled as MRSMs using a particle filter (see Figure 8C).

6.1.6. Non-rigid Object Registration

To be able to manipulate not only known objects, but also objects of the same category that differ in shape and appearance, we extended the coherent point drift [CPD (Myronenko and Song, 2010)] method to efficiently perform deformable registration between dense RGB-D point clouds (see Figure 8D). Instead of processing the dense point clouds of the RGB-D images directly with CPD, we utilize MRSMs to perform deformable registration on a compressed measurement representation. The method recovers a smooth displacement field, which maps the surface points between both point clouds. It can be used to establish shape correspondences between a partial view on an object in a current image and a MRSM object model. From the displacement field, the local frame transformation (i.e., 6D rotation and translation) at a point on the deformed surface can be estimated. By this, we can determine how poses such as grasps or tool end effectors change by the deformation between objects (Figure 8E). Further details on our deformable registration method can be found in Stückler and Behnke (2014b).

6.1.7. Object Localization using Bluetooth Low Energy Tags

As an alternative to tedious visual search for localizing objects, we developed a method for localizing objects equipped with low-cost Bluetooth low energy tags. Figure 9A shows examples.

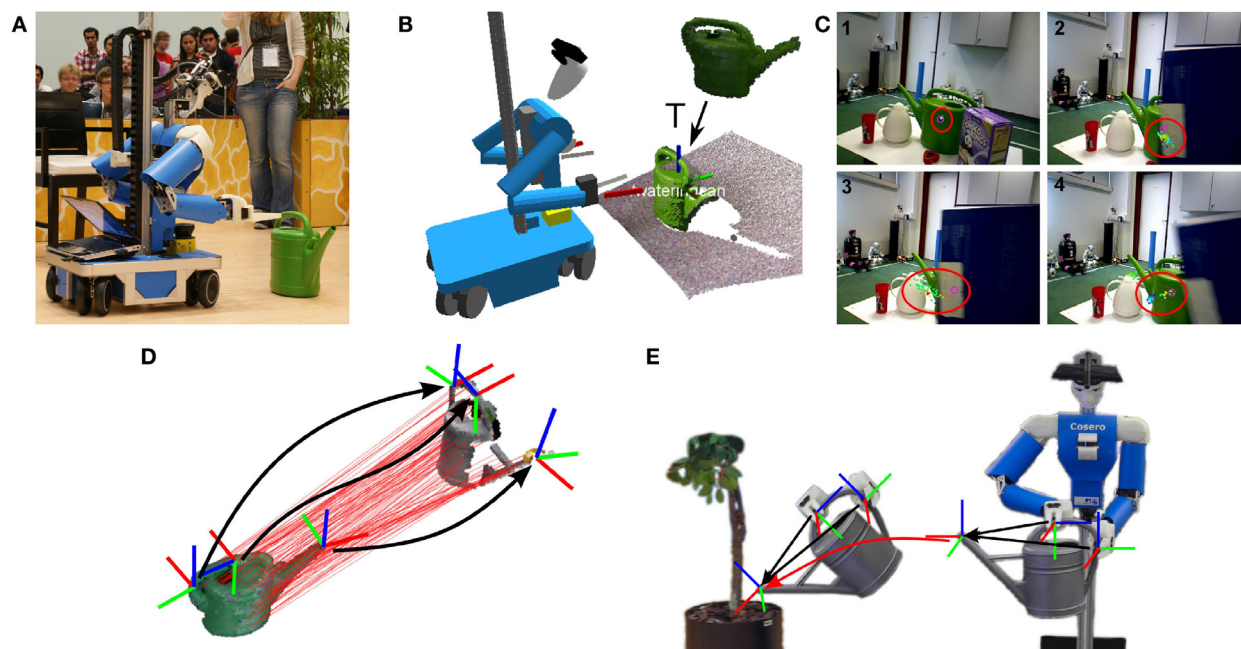


FIGURE 8 | Object tracking using registration of object models. (A) Cosero approaching a watering can. (B) We train multiview 3D models of objects using multiresolution surfel maps. The model is shown in green in the upper part. The model is registered with the current RGB-D frame to estimate its relative pose T , which is used to approach and grasp the watering can (Stückler et al., 2013a). (C) Joint object detection and tracking using a particle filter, despite occlusion (McElhone et al., 2013). Object manipulation skill transfer (Stückler and Behnke, 2014b): (D) an object manipulation skill is described by grasp poses and motions of the tool tip relative to the affected object. (E) Once these poses are known for a new instance of the tool, the skill can be transferred.

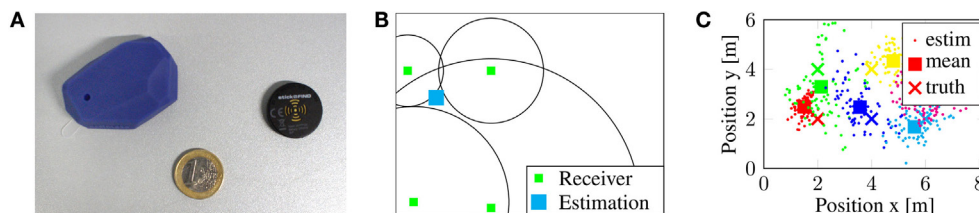


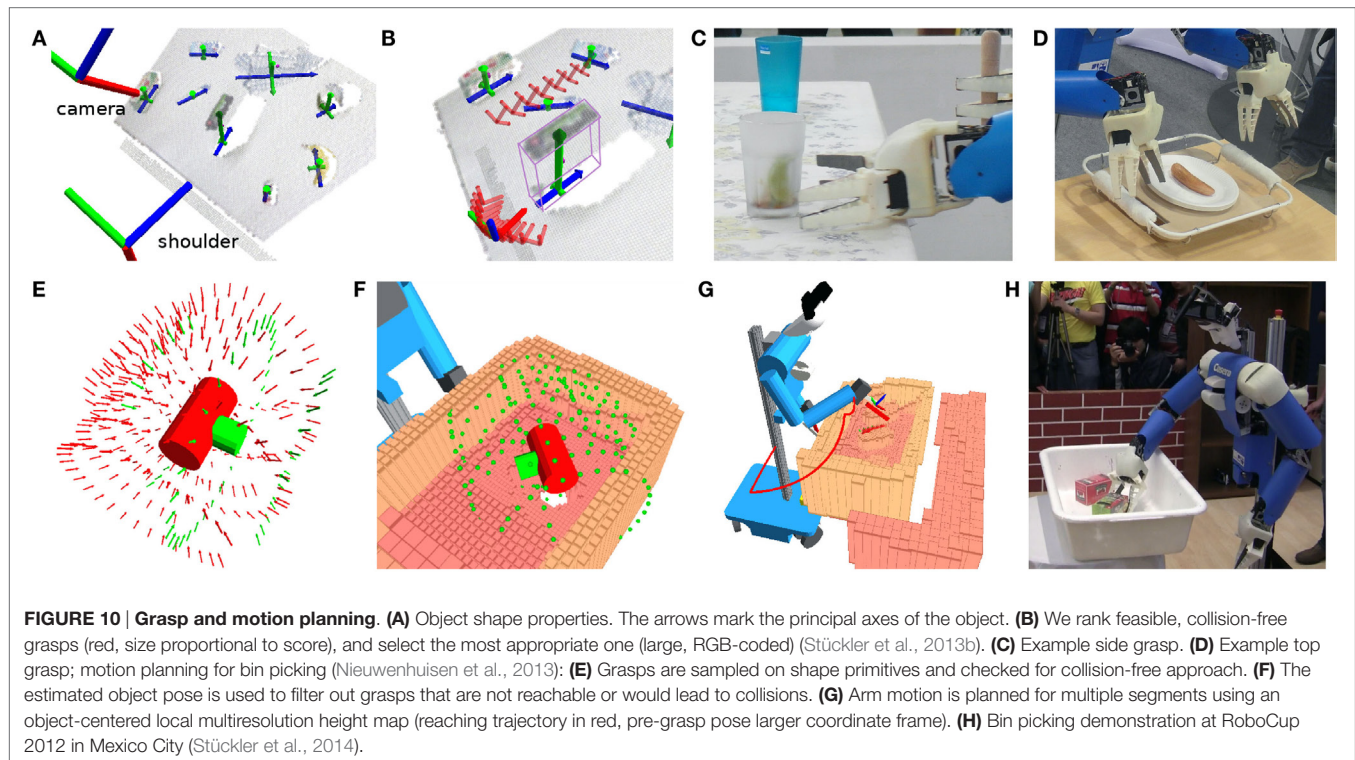
FIGURE 9 | Bluetooth low energy tag localization (Schwarz et al., 2015a). (A) Used Bluetooth tags (Estimate and StickNFind). (B) Signal strength-based tag localization by lateration. (C) Localization experiment with multiple tags.

Other Bluetooth devices, such as smart phones and tablets can also be localized. Our method requires the instrumentation of the environment with static Bluetooth receivers. The receivers report an RSSI (Received Signal Strength Indication) value, which decreases with the distance of the tag from the receiver.

For position estimation, an over-constrained lateration problem needs to be solved (see Figure 9B). The receivers report relative distances to the tag. We simultaneously optimize for the tag position and a common scaling factor for signal strength. As shown in Figure 9C, the resulting position estimates are still noisy, so we smooth them with a windowed mean over 30 s. Such coarse position estimates hint to typical placement locations in the environment, from which our robot can retrieve the objects. Further details of the method are reported by Schwarz et al. (2015a).

6.2. Object Grasping and Placement

Grasping objects from support surfaces is a fundamental capability. For objects segmented above horizontal surfaces as described in Section 6.1.1, we developed an efficient approach that is illustrated in Figure 10. We consider two kinds of grasps on these objects: top grasps that approach low objects from above and side grasps that are suitable for vertically elongated objects such as bottles or cans. We plan grasps by first computing grasp candidates on the raw object point cloud as perceived with the RGB-D camera. Our approach extracts the object principle axes in the horizontal plane and its height. We sample pre-grasp postures for top and side grasps and evaluate the grasps for feasibility under kinematic and collision constraints. The remaining grasps are ranked according to efficiency and robustness criteria. The best grasp is selected and finally executed with a parametrized motion



primitive. For collision detection, we take a conservative but efficient approach that checks simplified geometric constraints. Further details are provided by Stückler et al. (2013b).

For 3D modeled objects, which are tracked using the method described in Section 6.1.5, we define reference poses for grasping them with one or two hands for their task-specific usage (see Figure 8D, bottom). The estimated 6D object pose is used to transform the reference poses to the robot frame, parameterizing motion primitives for task-specific object manipulation like watering plants, pushing chairs, turning on devices, etc.

Our robot also supports the placement of objects on planar surfaces and the throwing of objects into trash bins.

6.3. Bin Picking

Objects are not always placed well-separated on horizontal support surfaces but also come densely packed in transport containers. To show the utility of our robot in such scenarios, we developed a mobile bin picking demonstration. Cosero navigates to an allocentric pose in front of the bin. It aligns precisely with the bin by perceiving it using its hip laser scanner in horizontal pose.

Objects in the bin are detected using the shape primitive-based method described in Section 6.1.4. We plan grasps in an efficient multistage process that successively prunes infeasible grasps using tests with increasing complexity. In an initial offline stage, we find collision-free grasps on the object, irrespective of object pose and not considering its scene context (Figure 10E). We sample grasp poses on the shape primitives and retain the ones, which allow for a collision-free gripper motion from pre-grasp to grasp pose.

During online planning, we transform the remaining grasp poses using the estimated object pose and check that a collision-free solution of the inverse kinematics in the current situation exists (Figure 10F). We allow collisions of the fingers with other parts in the transport box in the direct vicinity of the object to grasp, because the shape of the fingers allows for pushing them into narrow gaps between objects. The feasible grasps are ranked according to a score that incorporates efficiency and stability criteria.

The final step is to identify the best-ranked grasp that is reachable from the current posture of the robot arm. To this end, we successively plan reaching motions for the found grasps (Figure 10G). We test the grasps in descending order of their score. For motion planning, we employ LBKPIECE (Sucan and Kavraki, 2008). We split the reaching motion into multiple segments. This allows for a quick evaluation if a valid reaching motion can be found by planning in the descending order of the probability that planning for a segment will fail. Planning in the vicinity of the object needs a more exact environment representation as planning farther away from it. This is accomplished by centering a local multiresolution height map at the object to grasp, which is used for collision checking. This approach also leads to larger safety margins with increasing distance to the object. The object removal motion is planned with the robot extended by the grasped object. Further details are provided in Nieuwenhuisen et al. (2013).

6.4. Tool Use

Service tasks often involve the use of specialized tools. For a firm grip on such tools, we designed 3D-printed tool adapters

matching the four-finger grippers of Cosero. When the gripper closes on the adapter, the fingers bend around the shape of the adapter and establish form closure. The ridge on the center of the adapter fits into the space between the fingers. It fixates the adapter for exerting torques in pitch direction. For some tools such as pairs of tongs, the opening of the gripper is also used to operate the tool. To create form closure with the fingers at various opening angles, the adapters have flat springs for each finger.

6.4.1. Using a Pair of Tongs

When grasping sausages from a barbecue, the robot should not directly grasp with its grippers. Instead, it uses a pair of tongs to keep the food clean and to keep the grippers clear of the hot barbecue (see **Figure 11A**).

We segment the sausages from a plate or from the barbecue using plane segmentation (Section 6.1.1) and adapt the grasping motion to the position and orientation of the sausages. We exploit that the height of the barbecue or the plates on the plane is known and discard points of these support objects. The remaining points are clustered by Euclidean distance. We then estimate the principal axes of the segments and compare length (first principal axis) and width (second principal axis) with the expected size of the sausages. If these measures are within nominal ranges, the segment is classified as a sausage.

We extend the robot kinematic chain with the grasped tool. A parametrized motion primitive uses position and orientation of the closest sausage to pick it up with the tongs. The robot holds the tool above the objects on the table at all times during the demonstrations, so that collisions with these objects are avoided.

6.4.2. Bottle Opening

Opening a capped bottle with a bottle-opening tool is challenging, since the tool must be accurately placed onto the cap (see **Figure 11B**). The robot first grasps the bottle with its left hand and the tool with its right hand. It holds both objects close to each other above a horizontal surface. In order to stabilize the bottle, it rests it on the horizontal surface, still holding it in the hand. To execute the opening motion precisely, the robot must compensate for several sources of inaccuracy. First, an exact calibration between the robot sensors and end effector may not be known. Also, the pose of the tool in the gripper or the manipulated object cannot be assumed to be known precisely. We therefore implemented perception of the tips of the tool and the manipulated object using the head-mounted RGB-D camera. During manipulation, our robot looks at the tool and bottle,

segments the objects from the surrounding using our efficient segmentation method (see Section 6.1.1), and detects the endings of the objects in the segments.

We detect the tip of the opening tool in-hand by segmenting points in the depth image from the planar background. We select the segment closest to the position of the robot gripper and search for the farthest position from the gripper along its forward direction. The cap of the bottle in the other gripper is found in a similar way: Within the segment closest to the gripper, we search for the highest point. Since we know the size of the opening tool and the bottle, we can verify the found positions by comparing them to nominal positions. The bottle opening motion primitive is parameterized using the perceived cap and tool tip positions. The robot verifies the success of the bottle-opening through the lowering of the estimated bottle top position.

6.4.3. Watering Plants and Object Shape-Based Skill Transfer

For watering a plant with a watering can, our robot uses both arms (see **Figure 11C**). For grasping a previously known watering can, the robot approaches the can using 6D object tracking (Section 6.1.5; **Figures 8A,B**) and grasps the can at predefined reference poses. It navigates to an allocentric pose in front of the plant and positions itself relative to the plant pot that is perceived in its horizontal laser scanner. Water is poured into the pot by moving the can spout in a predetermined way through synchronized motion of both arms.

Preprogramming such a tool-use skill for every shape variation of watering cans is not desirable. We use the deformable registration method described in Section 6.1.6 to transfer the skill from the known can instance to a novel can. The skill of using the watering can is described by grasp poses relative to the object surface and motion trajectories of the can spout (see **Figures 8D,E**). To transfer this skill to a new variant of cans, we segment the new can from its support plane and establish shape correspondences to the object model of the known can. We estimate local frame transformations of the grasp poses and the tool end effector of the known can toward the observed can. The robot executes the transformed grasps to pick up the new can. For watering a plant, the robot moves the can end-effector frame relative to the plant in the same way as for the modeled can. This constrains the motion of the arms to keep the relative position of the transformed grasp poses to the transformed tool end effector pose. Further details of our adaptive tool-use methods can be found in Stückler and Behnke (2014a).



FIGURE 11 | Tool use. (A) Grasping sausages with a pair of tongs. Cosero perceives position and orientation of the sausages in RGB-D images (Stückler and Behnke, 2014a). **(B)** Bottle opening. **(C)** Plant watering skill transfer to unknown watering can (Stückler and Behnke, 2014b; cf. to **Figure 8E**).

7. INTUITIVE INTERACTION

In addition to the mobile manipulation capabilities described so far, intuitive user interfaces are key for making service robots useful. Speech, gestures, and body language are key modalities for human–human interaction. Employing them for face-to-face human–robot interaction is a natural way to configure the high-level autonomous behaviors of our robot when the user is in its vicinity. In addition, we also developed a handheld teleoperation interface, which is particularly useful for immobile users. The handheld interface gives the user the ability to control the robot on three levels of robot autonomy. Besides direct control on the body level to move the base and the arms or to adjust the gaze direction, this interface also allows for executing navigation or manipulation skills, and for sequencing skills in prolonged tasks.

7.1. Perception of Persons

A key prerequisite for a robot that engages in human–robot interaction is the perception of persons in its surrounding. This includes the detection and tracking of people, the identification of persons, and the interpretation of their gestures.

7.1.1. Person Detection and Tracking

We combine complementary information from laser range scanners and camera images to continuously detect and keep track of humans. In horizontal laser scans, the measurable features of persons like the shape of legs are not very distinctive, such that parts of the environment may cause false detections. However, laser scanners can be used to detect person candidates, to localize them, and to keep track of them in a wide field-of-view at high rates. In camera images, we verify that a track belongs to a person by detecting more distinctive human features: faces and upper bodies.

One horizontal laser scanner mounted 30 cm above the ground detects legs of people. We additionally detect torsos of people with the laser scanner in the robot's torso in horizontal scan position. In a multi-hypothesis tracker, we fuse both kinds of detections to tracks (see **Figure 12A**). Position and velocity of each track are estimated by a Kalman filter (KF). In the KF prediction step, we use odometry information to compensate for the motion of the

robot. After data association, the tracks are corrected with the observations of legs and torsos. We use the Hungarian method (Kuhn, 1955) to associate each torso detection in a scan uniquely with existing hypotheses. In contrast, as both legs of a person may be detected in a scan, we allow multiple leg detections to be assigned to a hypothesis. Only unassociated torso detections are used to initialize new hypotheses. A new hypothesis is considered a person candidate until it is verified as a person through vision. For this, our robot actively looks at a hypothesis using its pan-tilt Kinect camera and employs the face detector of Viola and Jones (2001) and HoG upper body detection Dalal and Triggs (2005). Spurious tracks with low detection rates are removed.

7.1.2. Person Identification

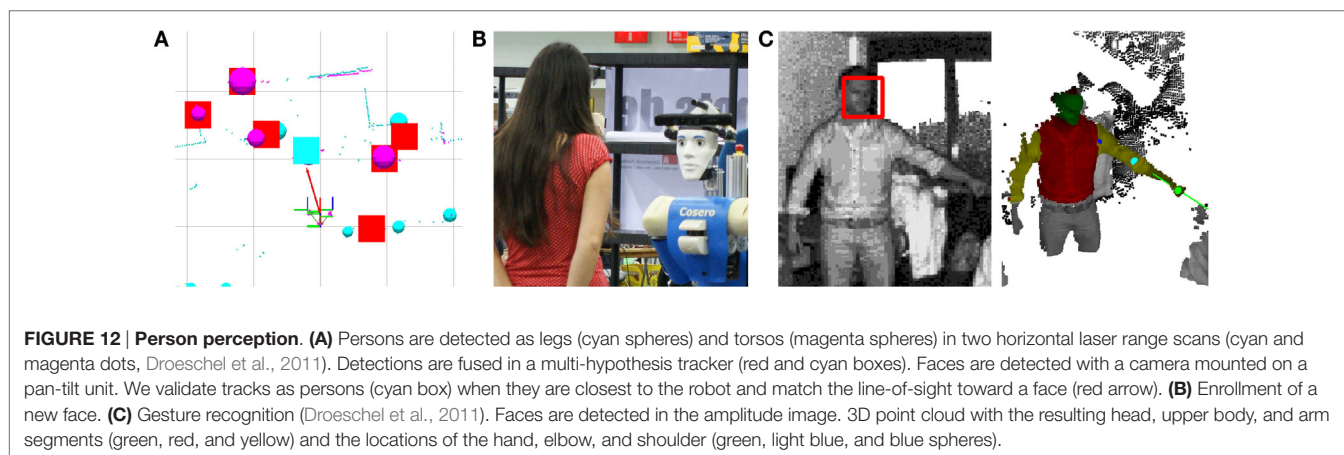
To determine the identity of the tracked persons, we implemented a face enrollment and identification system using the VeriLook SDK.¹ In the enrollment phase, our robot approaches detected persons, looks at their face with its camera, and asks them to look into the camera (**Figure 12B**). Face detection is done using the Viola and Jones' (2001) algorithm. The cut-out faces are passed to VeriLook, which extracts face descriptors that are stored in a repository. If the robot wants to identify a person later, it approaches the person, looks at their face, and compares the new descriptor to the stored ones.

7.1.3. Gesture Recognition

Gestures are a natural way of communication in human–robot interaction. A pointing gesture, for example, can be used to draw the robot's attention to a certain object or location in the environment. We implemented the recognition of pointing gestures, the showing of objects, and stop gestures. The primary sensor in our system for perceiving a gesture is the depth camera mounted on the robot's pan-tilt unit. Starting from faces detected in the amplitude image, we segment the person, its trunk, and arms in the depth image and determine the position of the head, hand, shoulder, and elbow. This is illustrated in **Figure 12C**.

We detect pointing gestures when the arm is extended, and the hand is held at a fixed location for a short time interval. To

¹<http://www.neurotechnology.com/verilook.html>.



compensate for the narrow field-of-view of the ToF camera, the robot adjusts its gaze to keep the hand as well as the head of the person in the image. After a pointing gesture has been detected, we infer its intended pointing target. Especially for distant targets, the line through eyes and hand yields a good approximation to the line toward the target. Showing of objects and stop gestures are detected when the arm of the human extends toward the robot. Further details of the method can be found in Droeschel et al. (2011). Droeschel and Behnke (2011) also developed an adaptive method for tracking an articulated 3D person model from the perspective of our robot.

7.2. Multimodal Dialog System

The perception of persons is the basis for intuitive multimodal interaction with users by means of speech, gestures, and body language. To interact with a person, our robot approaches a person track and points its RGB-D camera and directed microphone toward the user's head. This not only allows for the capturing of high-resolution images and depth measurements of the user's upper body and for recording of user utterances with a good signal-to-noise ratio but also signals to the user that our robot is in a state where it is ready to receive user input.

7.2.1. Speech Recognition and Synthesis

For speech recognition and synthesis, we use the Loquendo SDK.² Its speech synthesis supports colorful intonation and sounds natural. The robot generates speech depending on the task state, e.g., to inform the user about its current intent or to request user input. Loquendo's speech recognition is speaker independent and is based on predefined grammars that we attribute with semantic tags for natural language understanding. Again, the task state determines, which user utterances are understood.

7.2.2. Interpretation of Dialogs

On the task level, our robot supports spoken dialogs for specifying complex commands that sequence multiple skills. The ability to understand complex speech commands, to execute them, to detect failures, and to plan alternative actions in case of failures is assessed in the General Purpose Service Robot test in the RoboCup@Home competition. We describe the capabilities of the robot by a set of primitive skills that can be parameterized by specifying an object and/or a location. For instance, the skill `navigate_to_location` depends on a goal location while `fetch_object_from_location` requires a target object and an object location.

The robot knows a set of specific objects that are referenced by the object name in spoken commands. These specific objects are included in the visual object recognition database (Section 6.1.2). It is also possible to define an unspecific object using labels such as “unknown,” “some object,” or the label of an object category (e.g., “tool”). If multiple skills with object references are chained, the reflexive pronoun “it” refers to the last object that occurred in the task command. Hence, objects are referred to by labels and may have the additional attributes of being specific, unspecific, and reflexive.

Persons are handled in a similar way, but the notion of a person category is not included in our system. Our robots can enroll new persons and link their identity with their face appearance in the database of known persons (Section 7.1.2).

Specific locations, location categories, unspecific locations, or location-specific adjectives (like “back”) can be indicated by the user as well. We provide sets of navigation goal poses for specific locations as well as location categories. Different lists of poses are used for the purposes of object search, exploration for persons, or simple presence at a spot.

We utilize semantic tags in Loquendo's grammar specification to implement action, object, and location semantics in speech recognition. We appropriately designed the grammar so that recognition provides its semantic parse tree as a list of actions with attributed objects and locations.

Alternatively, for specific tasks, it is also possible to reference objects or locations by pointing gestures, and to recognize objects that the user shows the robot by holding them toward the robot.

Behavior control interprets the recognized semantic parse tree and sequences actions in a finite state machine. The robot executes this state machine and reports progress through speech synthesis. In case of a failure (e.g., desired object not found), the failure is recorded, the robot returns to the user, and reports the error through speech. Note that our behavior control does not involve a planning system. We observed that quite complex tasks can be communicated to the robot as a spoken sequence of actions, including unspecific objects or locations and reflexive pronouns, which can be translated into finite state machine behavior.

7.2.3. Synthesis of Body Language and Gestures

By moving in the environment, turning toward persons or toward manipulation locations, etc., our robot generates body language. Its anthropomorphic upper body makes it easy to interpret the robot actions. We also paid attention to make the robot motion look human-like. For example, the robot orients its head, upper body, and base into the driving direction to measure potential obstacles but also to indicate where it intends to go. According to the rotated masses, different time scales are used for this turning (Faber et al., 2008). Similarly, our robot directs its head toward the object that it wants to grasp, which not only provides good RGB-D measurements of the object but also makes manipulation actions predictable.

As part of its multimodal dialog system, our robot not only recognizes gestures but also performs gestures such as pointing to a location or waving toward a user using parametrized motion primitives.

7.3. Physical Human–Robot Interaction

Our robot does not only approach persons to communicate with them using speech and vision but also interacts with users in a physical way. Physical interaction occurs, for example, when handing objects over or when collaboratively working with objects. A key prerequisite for this direct human–robot interaction is compliant control of the arms (see Section 4.2). This, and also the lightweight robot construction, low torques of its actuators, and the friendly anthropomorphic design make a safe interaction without fear possible.

²<http://www.nuance.com/support/loquendo>.

7.3.1. Object Hand Over

Object hand over from the robot to a human could be implemented with several strategies. For instance, object release could be triggered by speech input or by specialized sensory input such as distance or touch sensors. We establish a very natural way of hand over by simply releasing the object when the human pulls on the object. More in detail, the skill is executed in the following way:

- The robot approaches the user and holds the object toward the person while uttering an object hand-over request,
- The user intuitively understands the hand-over offer and pulls on the object,
- We control the motion of the end-effector compliant in forward and upward direction as well as in pitch rotation. Our robot releases the object when it detects a significant displacement of its end effector.

Figure 4B shows an example of such a hand over.

When the user shall hand an object to the robot, the robot offers its open hand to the user. This signals the user to insert the object into the gap between the robot fingers. The object insertion is detected using in-hand distance sensors, which triggers closing of the fingers to firmly grasp the object. We observed that users intuitively understand the hand-over request and leave space for the closing fingers.

7.3.2. Guiding the Robot at Its Hand

A second example of physical human–robot interaction is taking the robot by its hand and guiding it. This is a simple and intuitive way to communicate locomotion intents to the robot (see **Figure 4C**). We combine person perception with compliant control to implement this behavior using the following procedure:

- The robot extends one of its hands forward and waits for the user,
- As soon as the user appears in front of the robot and exerts forces on the hand, the robot starts to follow the motion of the hand by translational motion through its drive.
- The robot avoids the user in a local potential field. It rotates with the drive toward the user to keep the guide at a constant relative angle (e.g., at 45°).

7.3.3. Cooperatively Carrying a Table

The third example of physical human–robot interaction is the task of cooperatively carrying a table (see **Figure 4D**). It combines object perception, person awareness, and compliant control, and consists of the following key steps:

- The task starts when the human user appears in front of the robot,
- The robot approaches the table, grasps it, and waits for the person to lift it,
- When the robot detects the lifting of the table, it also lifts the table and starts to follow the motion of the human,
- The human user ends the carrying of the table by lowering the table.

We apply our object registration and tracking method (Section 6.1.5) to find the initial pose of the table toward the robot. The robot then keeps track of the object while it drives toward a predefined approach pose, relative to the table. It grasps the table and waits, until the person lifts the table, which is indicated by a significant pitch rotation (0.02 rad) of the table.

As soon as the lifting is detected, the robot also lifts the table. It sets the motion of the grippers compliant in the sagittal and lateral direction and in yaw orientation. By this, the robot complies when the human pulls and pushes the table. The robot follows the motion of the human by controlling its omnidirectional base to realign the hands to the initial grasping pose with respect to the robot. During that, it keeps track of the table using MRSMAP registration. When the user puts the table down, it detects a significant pitch of the table, stops, and also lowers the table.

7.4. Handheld Teleoperation Interfaces

Direct interaction of the user with the robot is not always feasible or desirable. In particular, if the user is immobile or at a remote location, means for controlling the robot from a distance are needed. Such teleoperation interfaces must give the user good situation awareness through the display of robot sensor information and must provide intuitive ways to specify robot tasks.

Some teleoperation interfaces require special hardware, such as head-mounted displays or motion trackers (e.g., Rodehutsors et al. (2015)), but the use of such complex interfaces is not feasible in a domestic service setting. Because modern handheld computers such as smart phones and tablets are already widely used and provide display and input modalities, we implemented teleoperation with a handheld computer on three levels of autonomy (Schwarz et al., 2014): (I) the user directly controls body parts such as the end effectors, the gaze direction, or the omnidirectional drive on the *body level*. (II) On the *skill level*, the user controls robot skills, e.g., by setting navigation goals or commanding objects to be grasped. (III) On the *task level*, the user configures autonomous high-level behaviors that sequence skills.

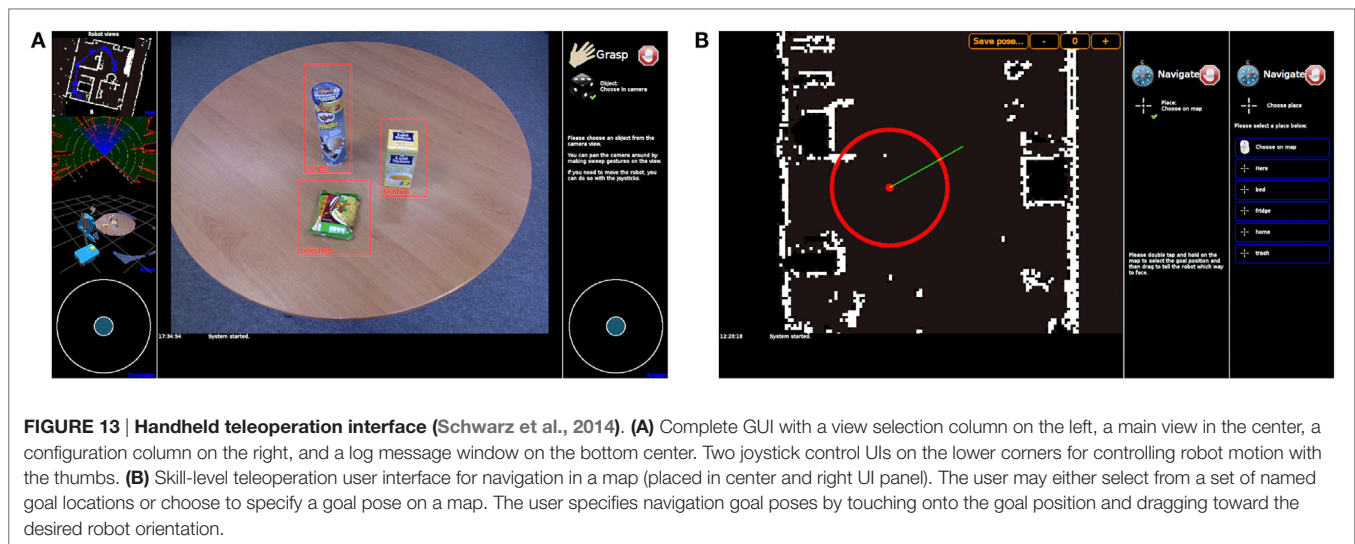
Our goal is to design a user interface in which the workload of the operator decreases with the level of robot autonomy. The operator selects the level of autonomy that is appropriate for the current situation. If the autonomous execution of a task or a skill fails, the user can select a lower level – down to direct body control – to solve the task.

7.4.1. Main User Interface Design

The main user interface is split into a main interactive view in its center and two configuration columns on the left and right side (see **Figure 13A**). In the left column, further scaled-down views are displayed that can be dragged into the main view. The dragged view then switches positions with the current main view. Below the main view in the center, a log screen displays status information in textual form.

7.4.2. Body-Level Teleoperation

The user has full control of the omnidirectional drive through two virtual joysticks in the lower left and right corners of the GUI. We intend the user to hold the mobile device in landscape orientation with two hands at its left and right side and to control



the UI elements with the left and right thumb. Obstacle avoidance decelerates the robot when the user drives onto an obstacle, but body-level controls do not support autonomous driving around the obstacle. In Muszynski et al. (2012), we also evaluated the use of the two virtual joysticks for end-effector control. In this mode, a centered slider at the bottom lets the user adjust the hand closure.

We support swipe gestures on the camera image for changing the gaze direction. By using swipe gestures instead of the joystick panels, one can control gaze and drive concurrently without switching controls.

7.4.3. Skill-Level Teleoperation

The skill-level user interfaces configure robot skills that require the execution of a sequence of body motions (see **Figure 13B**). The robot controls these body motions autonomously. By that, the workload on the user is reduced. While the robot executes the skill, the user can supervise its progress. Compared to body-level control, the skill-level UI does require less communication bandwidth, since images and control commands have to be transmitted with less frequency. Hence, this mode is less affected by low quality or low bandwidth communication.

On the skill level, the user has access to the following autonomous robot skills: navigation to goal poses in a map, navigation to semantic goal locations, grasping objects in the view of the robot, grasping semantically specific objects, receiving objects from a user, handing objects over to a user, and throwing objects into a trash bin. Execution failures are reported in the log so that the user can respond appropriately.

7.4.4. Task-Level Teleoperation

The task-level teleoperation UI is intended to provide the high-level behavior capabilities of the robot. These behaviors sequence multiple skills in a finite state machine. The user can compose actions, objects, and locations similar to the speech-based implementation of the parsing of complex speech commands described in Section 7.2.2.

This module allows the user to compose a sequence of skills in a two-stage user interface. On the top-level UI, the user adds and removes skills from the sequence. Skills can be added from a displayed list. Once a skill is selected, the user specifies location and object for the skill on a second-level UI. Finally, the user can start the execution of the task by touching a button on the bottom of the UI. A monitoring UI lets the user keep track of the robot's execution status, but the user can watch the progress also in the body-level and skill-level visualizations. Detected failures can be instantly reported to the user on the handheld, instead of physically returning to the user and reporting failures by speech. In case a failure occurs, the current skill is stopped and the execution of the task is interrupted so that the user can take over control.

8. RESULTS

Competitions and challenges have become important means in robotics to benchmark complex robot systems (Behnke, 2006; Gerndt et al., 2015; Guizzo and Ackerman, 2015). Since 2009, we compete with our cognitive service robots in the RoboCup@Home league (Wisspeintner et al., 2009; Iocchi et al., 2015), which is the top venue for benchmarking domestic service robots. In this annual international competition, robots have to demonstrate human–robot interaction and mobile manipulation in an apartment-like and in other domestic environments. The competition consists of several tests with predefined tasks, procedures, and performance measures that benchmark service robot skills in integrated systems. In addition, open demonstrations allow teams to show the best of their research. For our competition entries, we balanced mobile manipulation and human–robot interaction aspects. In the following, we report results of the RoboCup competitions in the years 2011–2014, where Cosero was used.

8.1. Mobile Manipulation and Tool Use

Several predefined tests in RoboCup@Home include object retrieval and placement. We often used open challenges to demonstrate further object manipulation capabilities such as tool use.

At RoboCup 2011 in the *Go Get It!* test, Cosero found a user-specified object and delivered it. In the *Shopping Mall* test, it learned a map of a previously unknown shopping mall and navigated to a shown location. In the *Demo Challenge*, the robot cleaned the apartment. It was instructed by gestures (Section 7.1.3) where to stow different kinds of laundry, picked white laundry from the floor, and put it into a basket. It then grasped carrots and tea boxes from a table. In the finals, our robot demonstrated a cooking task. It moved to a cooking plate to switch it on. For this, we applied our real-time object tracking method (Section 6.1.5) in order to approach the cooking plate and to estimate the switch grasping pose. Then, Cosero drove to the location of the dough and grasped it. Back at the cooking plate, it opened the bottle by unscrewing its lid and poured its contents into the pan. Meanwhile, our second robot Dynamaid opened a refrigerator (Section 4.2), picked a bottle of orange juice out of it, and placed the bottle on the breakfast table.

At RoboCup 2012 in the *Clean Up* test, our robot Cosero had to find objects that were distributed in the apartment, recognize them, and bring them to their place. Our robot detected three objects, from which two were correctly recognized as unknown objects. It grasped all three objects and deposited them in the trash bin. In the *Open Challenge*, we demonstrated a housekeeping scenario. Cosero took over an empty cup from a person and threw it into the trash bin. Afterward, it approached a watering can and watered a plant. In the *Restaurant* test, our robot Cosero was guided through a previously unknown bar. The guide showed the robot where the shelves with items and the individual tables were. Our robot built a map of this environment, took an order, and navigated to the food shelf to search for requested snacks. In the final, Cosero demonstrated the approaching, bimanual grasping, and moving of a chair to a target pose. It also approached and grasped a watering can with both hands and watered a plant. For this, approaching and bimanual grasping of the chair and the watering can was realized through registration of learned 3D models of the objects (Section 6.1.5). After this demonstration, our second robot Dynamaid fetched a drink and delivered it to the jury. In the meantime, Cosero approached a transport box, analyzed its contents (Section 6.1.4), and grasped a perceived object using grasp and motion planning described in Section 6.3 (Figure 10H).

At RoboCup 2013, Cosero found in the *Clean Up* test two objects and brought one of it to its correct place. In the *Restaurant* test, Cosero was shown the environment and the location of food and drinks, which it later found again. In the *Demo Challenge*, we demonstrated a care scenario in which the robot extended the mobility of a user with its mobile manipulation capabilities through the teleoperation interface described in Section 7.4. Cosero also moved a chair to its location.

In the *Open Challenge*, Cosero demonstrated tool-use skill transfer based on our deformable registration method (Section 6.1.6). The jury chose one of two unknown cans. The watering skill was trained for a third instance of cans before. Cosero successfully transferred the tool-use skill and executed it (Figure 11C).

In the final, Cosero demonstrated grasping of sausages with a pair of tongs. The robot received the tongs through object hand over from a team member. It coarsely drove behind the barbecue

that was placed on a table by navigating in the environment map and tracked the 6-DoF pose of the barbecue using MRSMaps (Section 6.1.5) to accurately position itself relative to the barbecue. It picked one of two raw sausages from a plate next to the barbecue with the tongs (Section 6.4.1) and placed it on the barbecue. While the sausage was grilled, Cosero handed the tongs back to a human and went to fetch and open a beer. It picked the bottle opener from a shelf and the beer bottle with its other hand from a table. Then it executed the bottle opening skill described in Section 6.4.2. After our robot placed the bottle opener on the table, it delivered the beer to a jury member. Then it received the tongs again and returned to the barbecue to grasp the sausage and to place it on a clean plate.

In the finals of German Open 2014, Cosero demonstrated again the use of the tongs and the bottle opener (Figures 11A,B). This time, the sausage was placed on the grill in advance. Accordingly, the task of Cosero was to pick it from the barbecue and place it on a plate, which was located on a tray. Our robot grasped the tray with both hands and delivered the sausage to a jury member (Figure 10D).

At RoboCup 2014 in Brazil, Cosero demonstrated in the *Basic Functionality* test object recognition and grasping as well as navigation in the arena where an additional obstacle was placed and a door was closed. It demonstrated opening a bottle in the *Open Challenge* (Figure 14A). In the *Enduring General Purpose Service Robot* test, our robot recognized two complex speech commands (Section 7.2.2) and carried out the requested actions. In the final, Cosero demonstrated the use of tools. It grasped a dustpan and a swab in order to clean some dirt from the floor (Figure 14B). Although the dirt detection failed, our robot executed the cleaning motion and continued the demo by pouring out the contents of the dustpan into the dustbin. It placed the tools back on a table and started to make caipirinha. For this, it used a muddler to muddle lime pieces (Figure 14C).

8.2. Human–Robot Interaction

Person detection (Section 7.1.1) followed by face enrollment and later identification (Section 7.1.2) has been demonstrated by Cosero at multiple occasions during RoboCup@Home competitions throughout the years 2011–2014. At the 2011 RoboCup in the *Follow Me* test, Cosero met a previously unknown person and followed him reliably through an unknown environment. Cosero showed that it can distinguish the person from others and that it recognizes stop gestures (Section 7.1.3). In 2012, the *Follow Me* test was made more difficult. Cosero learned the face of the guide and was not disturbed later by another person blocking the line-of-sight. It followed the guide into an elevator and left it on another floor.

In *Who Is Who*, two previously unknown persons introduced themselves to Cosero. Later in the test, our robot found one of the previously unknown persons, two team members, and one unknown person and recognized their identity correctly. In the 2012 *Who Is Who* test, Cosero learned the faces of three persons, took an order, fetched three drinks into a basket and with each of its grippers, and successfully delivered two of them within the time limit of 5 min. In 2013, face recognition has been embedded in the *Cocktail Party* test, where the robot took drink orders,

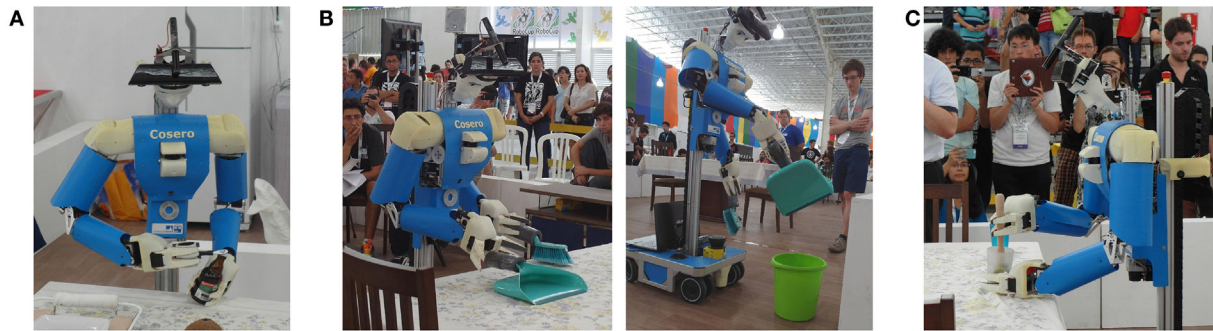


FIGURE 14 | Tool use at RoboCup 2014 in João Pessoa, Brazil. (A) Bottle opener, (B) Dustpan and swab, and (C) Muddler.

fetches the drinks, and delivered these to persons identified after they moved to a different room.

The recognition of pointing gestures (Section 7.1.3) has been demonstrated in several tests, e.g., in the German Open 2011 final, where a jury member showed our robot the exit door, in the RoboCup 2011 *Demo Challenge*, where our robot tidied up the apartment by moving objects into shelves indicated by pointing gestures, in the RoboCup 2011 finals, where a user showed our robot where it finds a bottle of dough to make an omelet, in the German Open 2012 *Demo Challenge*, where a human showed the robot in which baskets to put colored and white laundry, and in the RoboCup 2012 *Demo Challenge*, where the robot picked up an object referenced by pointing from the floor. In the 2012 *Open Challenge*, Cosero demonstrated that it could recognize a waving person. It took over an empty cup from this person and threw it into the trash bin.

At RoboCup 2013, the *Emergency Situation* test was introduced. Here, Cosero found a standing person, asked the person if he required help, and guided him to the exit.

Cooperative carrying of a table by Cosero and a human (Section 7.3.3) was demonstrated in the RoboCup 2011 final (Figure 4D) and in the German Open 2012 *Open Challenge*. In this test, also guiding the robot by taking its hand (Section 7.3.2) was demonstrated (Figure 4C). Human–robot object hand over in both directions was demonstrated very often, e.g., at the German Open 2014 (Figure 4C).

Task-level behavior generation according to complex speech commands as described in Section 7.2.2 is tested in the *General Purpose Service Robot* test of the RoboCup@Home competition. At RoboCup 2012 in Mexico, Cosero recognized speech commands from two out of three categories. It recognized a complex speech command consisting of three skills. It also understood a speech command with unspecific information and posed adequate questions to retrieve missing information. In 2013, the *Enduring General Purpose Service Robot* test was introduced, where three robots were tested in a round-robin procedure for up to 40 min. Again, Cosero performed well in this test, understanding two commands in two command categories. In the first trial, Cosero understood a complex command composed of three skills. The second complex command was sequencing navigation skills and was solved by Cosero easily. It then received a command with unspecific information where it also asked questions to make the

task specific. It now should grasp from the armrest of a couch, but could not find the object. Cosero detected this error, returned to the user, and reported the problem.

At RoboCup 2012 and German Open 2013, we also demonstrated teleoperation using a handheld device (Section 7.4). In the *Demo Challenge* at RoboCup 2012, we showed an elderly care scenario in which a user commanded the robot to fetch a drink from another room. At first, the person let the robot fetch a specific beverage. The robot drove to the assumed location of the drink, but since it was not available, the remote user had to take a different choice. The user switched to the skill-level control user interface and selected one of the other beverages that were perceived by the robot on the table and displayed in live images on the handheld screen. Finally, the robot grasped the selected drink, brought it to the user, and handed it over. At German Open 2013, we extended the *Demo Challenge* with receiving objects from users and putting the object in a waste bin.

We demonstrated our signal strength-based object localization approach (Section 6.1.7) publicly during the *Demo Challenge* at the 2014 RoboCup German Open competition in Magdeburg. A user asked Cosero to retrieve his medicine that he could not find. The medicine had been placed at one of two locations, which was chosen by a jury member. A Bluetooth tag had been attached to the medicine, which was localized coarsely using signal strength-based localization from four receivers in the room corners. Cosero drove to the table close to the estimated medicine position, searched, detected, and grasped the medicine, and brought it to the user. In a second run, the robot localized and retrieved the object from the other location.

8.3. Competition Results

With Cosero, we participated in four international RoboCup@Home and four RoboCup German Open @Home competitions in the years 2011–2014. Our robot systems performed consistently well in the predefined tests and our open demonstrations convinced the juries, which consisted of team leaders, members of the executive committee of the league, and representatives of the media, science, and industry.

Our team NimbRo@Home won the international RoboCup@Home competitions in 2011 [Istanbul (Stückler et al., 2012)], 2012 [Mexico City (Stückler et al., 2013a)], and 2013 [Eindhoven (Stückler et al., 2014)]. Our team also continuously achieved 1st

TABLE 1 | RoboCup@Home competition results 2011–2014.

| @Home competition | Winner (normal score) | Second place (normal score) | Third place (normal score) |
|-------------------|-----------------------|------------------------------|-----------------------------------|
| 2014, João Pessoa | Wright Eagle, China | Tech United Eindhoven | NimbRo@Home |
| German Open 2014 | NimbRo@Home (100) | Tech United Eindhoven (59) | ToBi Bielefeld (50) |
| 2013, Eindhoven | NimbRo@Home (99) | Wright Eagle, China (86) | Tech United Eindhoven (73) |
| German Open 2013 | NimbRo@Home (100) | SmartBots@Ulm (67) | Homer, University of Koblenz (61) |
| 2012, Mexico city | NimbRo@Home (100) | eR@sers, Japan (74) | ToBi Bielefeld (64) |
| German Open 2012 | NimbRo@Home (100) | b-it-bots, St. Augustin (56) | Golem, Mexico (39) |
| 2011, Istanbul | NimbRo@Home (100) | Wright Eagle, China | b-it-bots, Sankt Augustin |
| German Open 2011 | NimbRo@Home | SmartBots, Ulm | b-it-bots, Sankt Augustin |

place in the RoboCup German Open competitions of the league from 2011 to 2014. **Table 1** summarizes these results. Detailed competition reports, including pictures and videos, can be found on our website.³

8.4. Lessons Learned

Our continuous success in international RoboCup@Home competitions demonstrates our achievements in designing and integrating a domestic service robot system that balances mobile manipulation and human–robot interaction capabilities. Currently, our system is limited to short task demonstrations (ca. 10 min) in partially controlled competition environments. The development of the system gave us many insights into future research steps that are necessary to potentially scale domestic robot systems further toward real application scenarios.

- A soft and compliant mechatronic design would increase the inherent safety of our robot,
- We designed special tool handles to overcome limitations of our current gripper design. Dexterous human-like hands with delicate tactile sensing would allow for more complex manipulation skills without such special tool handles,
- Locomotion with a wheeled base is limited to flat ground with small slopes and steps. A future direction could be to combine wheeled with legged locomotion to also pass over steps or stairs,
- Our navigation system currently models the environment in static maps. Changes in the environment are handled using probabilistic measurement models and probabilistic state estimation. A dynamic environment representation could handle changes more flexibly and could allow for keeping track of the moving objects,
- Object recognition and handling is mostly limited to small-scale predefined sets of objects. We explored first steps toward

scaling the system to a larger variety of unknown objects with our shape-based skill transfer approach,

- Object perception in our system is currently focused on the robot on-board sensory percepts. External sensors such as Bluetooth tags give an important further cue for the perception of the state of objects. It could be a viable option to instrument the environment with various further sensors to increase the awareness on the objects in the environment.
- Our robot perceives person through detection and facial identification. It can also interpret a set of short gestures. The observation of prolonged human actions and behavior, the understanding of user intents, and the predictions of future actions would allow our system to achieve increased levels of human–robot interaction.

9. CONCLUSION

In this paper, we detailed our approaches to realize mobile manipulation, tool use, and intuitive human–robot interaction with our cognitive service robot Cosero.

We equipped our robot with an anthropomorphic upper body and an omnidirectional drive to perform tasks in typical household scenarios. Through compliant control of the arms, our robot interacts physically with humans and manipulates objects such as doors without accurate models. We proposed several object perception methods to implement the variety of manipulation skills of our robots. We segment scenes at high frame rate into support surfaces and objects. In order to align to objects for grasping, we register RGB-D measurements on the object with a 3D model using multiresolution surfel maps (MRSMaps). Through deformable registration of MRSMaps, we transfer object manipulation skills to differently shaped instances of the same object category. Tool use is one of the most complex manipulation skills for humans and robots in daily life. We implemented several tool-use strategies using our perception and control methods.

For human–robot interaction, communication partners are perceived using laser range sensors and vision. Our robot can recognize and synthesize speech and several gestures. It can parse the semantics of complex speech commands and generate behavior accordingly. To control the robot on three levels of autonomy, we developed teleoperation user interfaces for handheld devices.

The outstanding results achieved at multiple national and international RoboCup@Home competitions clearly demonstrate our success in designing a balanced system that incorporates mobile manipulation and intuitive human–robot interfaces. The development and benchmarking of the system gave us many insights into the requirements for complex personal service robots in scenarios such as cleaning the home or assisting the elderly. Challenges like RoboCup@Home show that a successful system not only consists of valid solutions to isolated action and perception problems. The proper integration of the overall system is equally important.

Despite a large number of successful demonstrations, our system is currently limited to short tasks in partially controlled

³<http://www.ais.uni-bonn.de/nimbRo@Home>.

environments. In order to scale toward real application in domestic service scenarios, we need to address open issues. On the mechatronic side, a soft and compliant robot design would further increase the inherent safety of the robot, and more dexterous hands would enable more complex manipulation skills and reduce the need for special tool handles. Object recognition and handling that scales to the large variety of objects in our daily homes is still an open research problem. Significant progress has been made, e.g., through deep learning methods, but occlusions and material properties like transparency or highly reflective surfaces make it still challenging to analyze typical household scenes. Similarly, perceiving people and understanding their actions in the many situations possible in everyday environments is a challenge. One promising approach to address these challenges is to instrument the environment with a multitude of sensors in order to track all objects continuously with high accuracy (Fox, 2016).

REFERENCES

- Ackerman, E. (2016). "IHMC's ATLAS robot learning to do some chores," in *Automaton, IEEE Spectrum Blog*. Available at: <http://spectrum.ieee.org/automaton/robotics/humanoids/atlas-does-some-chores>
- Asfour, T., Regenstein, K., Azad, P., Schröder, J., Bierbaum, A., Vahrenkamp, N., et al. (2006). "ARMAR-III: an integrated humanoid platform for sensory-motor control," in *Proc. IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (Genova: IEEE), 169–175.
- Bäumel, B., Schmidt, F., Wimböck, T., Birbach, O., Dietrich, A., Fuchs, M., et al. (2011). "Catching flying balls and preparing coffee: Humanoid Rollin'Justin performs dynamic and sensitive tasks," in *IEEE International Conference on Robotics and Automation (ICRA)* (Shanghai, China), 3443–3444.
- Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (SURF). *Comput. Vision Image Understand.* 110, 346–359. doi:10.1016/j.cviu.2007.09.014
- Beetz, M., Klank, U., Kresse, I., Maldonado, A., Mösenlechner, L., Pangercic, D., et al. (2011). "Robotic roommates making pancakes," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (Bled: IEEE), 529–536.
- Behnke, S. (2006). "Robot competitions – ideal benchmarks for robotics research," in *IROS Workshop on Benchmarks in Robotics Research*, Beijing.
- Berner, A., Li, J., Holz, D., Stückler, J., Behnke, S., and Klein, R. (2013). "Combining contour and shape primitives for object detection and pose estimation of pre-fabricated parts," in *IEEE International Conference on Image Processing (ICIP)* (Melbourne: IEEE), 3326–3330.
- Bohren, J., Rusu, R., Jones, E., Marder-Eppstein, E., Pantofaru, C., Wise, M., et al. (2011). "Towards autonomous robotic butlers: lessons learned with the PR2," in *IEEE International Conference on Robotics and Automation (ICRA)* (Shanghai: IEEE), 5568–5575.
- Borst, C., Wimböck, T., Schmidt, F., Fuchs, M., Brunner, B., Zacharias, F., et al. (2009). "Rollin'Justin–Mobile platform with variable base," in *IEEE International Conference on Robotics and Automation (ICRA)* (Kobe: IEEE), 1597–1598.
- Chen, X., Shuai, W., Liu, J., Liu, S., Wang, X., Wang, N., et al. (2015). "KeJia: the intelligent domestic robot for RoboCup@Home 2015," in *RoboCup@Home League Team Descriptions for the Competition in Hefei, China*. Available at: http://robocup2015.oss-cn-shenzhen.aliyuncs.com/TeamDescriptionPapers/RoboCup@Home/RoboCup_Symposium_2015_submission_124.pdf
- Dalal, N., and Triggs, B. (2005). "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision and Pattern Recognition (CVPR)* (San Diego: IEEE), 886–893.
- Droeschel, D., and Behnke, S. (2011). "3D body pose estimation using an adaptive person model for articulated ICP," in *4th International Conference on Intelligent Robotics and Applications (ICIRA)* (Aachen: IEEE), 157–167.
- Droeschel, D., Stückler, J., and Behnke, S. (2011). "Learning to interpret pointing gestures with a time-of-flight camera," in *6th ACM International Conference on Human-Robot Interaction (HRI)* (Lausanne: ACM/IEEE), 481–488.

AUTHOR CONTRIBUTIONS

JS served as Team Leader of the NimbRo@Home team 2009–2013. He contributed most of the developed methods. MS was team member since 2009 and team leader since 2014. He contributed object perception using pretrained neural networks and Bluetooth tags. SB initiated and managed the project and compiled the paper.

ACKNOWLEDGMENTS

The authors thank the members of team NimbRo@Home for their support in preparing and running the RoboCup competition systems: Nikita Araslanov, Ishrat Badami, David Droschel, Kathrin Gräve, Dirk Holz, Jochen Kläss, Manus McElhone, Matthias Nieuwenhuisen, Michael Schreiber, David Schwarz, Ricarda Steffens, and Angeliki Topalidou-Kyniazopoulou.

- Faber, F., Bennewitz, M., and Behnke, S. (2008). "Controlling the gaze direction of a humanoid robot with redundant joints," in *17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Munich: IEEE), 413–418.
- Fox, D. (2001). "KLD-sampling: adaptive particle filters and mobile robot localization," in *Advances in Neural Information Processing Systems (NIPS)* (Vancouver: MIT Press), 26–32.
- Fox, D. (2016). "The 100-100 tracking challenge," in *Keynote at IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm.
- Gerndt, R., Seifert, D., Baltes, J., Sadeghnejad, S., and Behnke, S. (2015). Humanoid robots in soccer—robots versus humans in RoboCup 2050. *IEEE Robot. Autom. Mag.* 22, 147–154. doi:10.1109/MRA.2015.2448811
- Grisetti, G., Stachniss, C., and Burgard, W. (2007). Improved techniques for grid mapping with Rao-Blackwellized particle filters. *IEEE Trans. Rob.* 23, 34–46. doi:10.1109/TRO.2006.889486
- Guizzo, E., and Ackerman, E. (2015). The hard lessons of DARPA's robotics challenge. *Spectrum* 52, 11–13. doi:10.1109/MSPEC.2015.7164385
- Hermann, A., Sun, J., Xue, Z., Rhl, S. W., Oberländer, J., Roennau, A., et al. (2013). "Hardware and software architecture of the bimanual mobile manipulation robot HoLLiE and its actuated upper body," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)* (Wollongong: IEEE), 286–292.
- Iocchi, L., Holz, D., Ruiz-del Solar, J., Sugiura, K., and van der Zant, T. (2015). RoboCup@Home: analysis and results of evolving competitions for domestic and service robots. *Artif. Intell.* 229, 258–281. doi:10.1016/j.artint.2015.08.002
- Iwata, H., and Sugano, S. (2009). "Design of human symbiotic robot TWENDY-ONE," in *IEEE International Conference on Robotics and Automation (ICRA)* (Kobe: IEEE), 580–586.
- Jain, A., and Kemp, C. C. (2010). EL-E: an assistive mobile manipulator that autonomously fetches objects from flat surfaces. *Auton. Robots* 28, 45–64. doi:10.1007/s10514-009-9148-5
- Kittmann, R., Fröhlich, T., Schäfer, J., Reiser, U., Weißhardt, F., and Haug, A. (2015). "Let me introduce myself: i am Care-O-bot 4, a gentleman robot," in *Mensch und Computer 2015 – Proceedings* (Berlin: DeGruyter Oldenbourg), 223–232.
- Kläss, J., Stückler, J., and Behnke, S. (2012). "Efficient mobile robot navigation using 3D surfel grid maps," in *7th German Conference on Robotics (ROBOTIK)*, Munich.
- Kuhn, H. (1955). The hungarian method for the assignment problem. *Naval Res. Logist. Q.* 2, 83–97. doi:10.1002/nav.3800020109
- Leidner, D., Dietrich, A., Schmidt, F., Borst, C., and Albu-Schäffer, A. (2014). "Object-centered hybrid reasoning for whole-body mobile manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong: IEEE), 1828–1835.
- Lunenburg, J., van den Dries, S., Ferreira, L. B., and van de Molengraft, R. (2015). "Tech United Eindhoven @Home 2015 team description paper," in

- RoboCup@Home League Team Descriptions for the Competition in Hefei, China*. Available at: http://robocup2015.oss-cn-shenzhen.aliyuncs.com/TeamDescriptionPapers/RoboCup@Home/RoboCup_Symposium_2015_submission_147.pdf
- McElhone, M., Stückler, J., and Behnke, S. (2013). "Joint detection and pose tracking of multi-resolution surfel models in RGB-D," in *European Conference on Mobile Robots (ECMR)* (Barcelona: IEEE), 131–137.
- Meeussen, W., Wise, M., Glaser, S., Chitta, S., McGann, C., Mihelich, P., et al. (2010). "Autonomous door opening and plugging in with a personal robot," in *IEEE International Conference on Robotics and Automation (ICRA)* (Anchorage: IEEE), 729–736.
- Muszynski, S., Stückler, J., and Behnke, S. (2012). "Adjustable autonomy for mobile teleoperation of personal service robots," in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (France: IEEE), 933–940.
- Myronenko, A., and Song, X. (2010). Point set registration: coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 2262–2275. doi:10.1109/TPAMI.2010.46
- Nguyen, H., Ciocarlie, M., Hsiao, K., and Kemp, C. (2013). "ROS commander: flexible behavior creation for home robots," in *IEEE International Conference on Robotics and Automation (ICRA)* (Karlsruhe: IEEE), 467–474.
- Nieuwenhuisen, M., and Behnke, S. (2013). Human-like interaction skills for the mobile communication robot Robotinho. *Int. J. Soc. Rob.* 5, 549–561. doi:10.1007/s12369-013-0206-y
- Nieuwenhuisen, M., Droschel, D., Holz, D., Stückler, J., Berner, A., Li, J., et al. (2013). "Mobile bin picking with an anthropomorphic service robot," in *IEEE International Conference on Robotics and Automation (ICRA)* (Karlsruhe: IEEE), 2327–2334.
- Parlitz, C., Hägele, M., Klein, P., Seifert, J., and Dautenhahn, K. (2008). "Care-obot 3 – rationale for human-robot interaction design," in *39th International Symposium on Robotics (ISR)* (Seoul: IEEE), 275–280.
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., et al. (2009). "ROS: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, Kobe.
- Rodehutsors, T., Schwarz, M., and Behnke, S. (2015). "Intuitive bimanual telemanipulation under communication restrictions by immersive 3D visualization and motion tracking," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (Seoul: IEEE), 276–283.
- Sang, H. (2011). *Introduction of CIROS*. Available at: http://www.irobotics.re.kr/eng_sub1_3
- Schwarz, D., Schwarz, M., Stückler, J., and Behnke, S. (2015a). "Cosero, find my keys! object localization and retrieval using Bluetooth low energy tags," in *RoboCup 2014: Robot Soccer World Cup XVIII, Volume 8992 of Lecture Notes in Computer Science* (Springer), 195–206.
- Schwarz, M., Schulz, H., and Behnke, S. (2015b). "RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features," in *IEEE International Conference on Robotics and Automation (ICRA)* (Seattle: IEEE), 1329–1335.
- Schwarz, M., Stückler, J., and Behnke, S. (2014). "Mobile teleoperation interfaces with adjustable autonomy for personal service robots," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction (HRI)* (Bielefeld: ACM), 288–289.
- Seib, V., Manthe, S., Holzmann, J., Memmesheimer, R., Peters, A., Bonse, M., et al. (2015). "RoboCup 2015 – homer@UniKoblenz (Germany)," in *RoboCup@Home League Team Descriptions for the Competition in Hefei, China*. Available at: http://robocup2015.oss-cn-shenzhen.aliyuncs.com/TeamDescriptionPapers/RoboCup@Home/RoboCup_Symposium_2015_submission_131.pdf
- Srinivasa, S., Ferguson, D., Helfrich, C., Berenson, D., Romea, A. C., Diankov, R., et al. (2010). Herb: a home exploring robotic butler. *Auton. Robots* 28, 5–20. doi:10.1007/s10514-009-9160-9
- Stückler, J., Badami, I., Droschel, D., Gräve, K., Holz, D., McElhone, M., et al. (2013a). "NimbRo@Home: winning team of the RoboCup@Home competition 2012," in *RoboCup 2012: Robot Soccer World Cup XVI, Volume 7500 of Lecture Notes in Computer Science*, eds X. Chen, P. Stone, L. E. Sucar, and T. van der Zant (Springer), 94–105.
- Stückler, J., Steffens, R., Holz, D., and Behnke, S. (2013b). Efficient 3D object perception and grasp planning for mobile manipulation in domestic environments. *Rob. Auton. Syst.* 61, 1106–1115. doi:10.1016/j.robot.2012.08.003
- Stückler, J., and Behnke, S. (2011). Dynamaid, an anthropomorphic robot for research on domestic service applications. *Automatika J Control Meas. Electron. Comput. Commun.* 52, 233–243.
- Stückler, J., and Behnke, S. (2012). "Compliant task-space control with back-drivable servo actuators," in *RoboCup 2011: Robot Soccer World Cup XV, Volume 7416 of Lecture Notes in Computer Science* (Springer), 78–89.
- Stückler, J., and Behnke, S. (2014a). "Adaptive tool-use strategies for anthropomorphic service robots," in *14th IEEE/RAS International Conference of Humanoids Robotics (Humanoids)* (Madrid: IEEE), 755–760.
- Stückler, J., and Behnke, S. (2014b). "Efficient deformable registration of multi-resolution surfel maps for object manipulation skill transfer," in *IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong: IEEE), 994–1001.
- Stückler, J., and Behnke, S. (2014c). Multi-resolution surfel maps for efficient dense 3D modeling and tracking. *J. Visual Commun. Image Represent.* 25, 137–147. doi:10.1016/j.jvcir.2013.02.008
- Stückler, J., Droschel, D., Gräve, K., Holz, D., Kläß, J., Schreiber, M., et al. (2012). "Towards robust mobility, flexible object manipulation, and intuitive multimodal interaction for domestic service robots," in *RoboCup 2011: Robot Soccer World Cup XV, Volume 7416 of Lecture Notes in Computer Science*, eds Th. Roefer, N. M. Mayer, J. Savage, and U. Saranlı (Springer), 51–62.
- Stückler, J., Droschel, D., Gräve, K., Holz, D., Schreiber, M., Topalidou-Kyniazopoulou, A., et al. (2014). "Increasing flexibility of mobile manipulation and intuitive human-robot interaction in RoboCup@Home," in *RoboCup 2013: Robot Soccer World Cup XVII, Volume 8371 of Lecture Notes in Computer Science*, eds S. Behnke, M. M. Veloso, A. Visser, and R. Xiong (Springer), 135–146.
- Stückler, J., Schwarz, M., Schädler, M., Topalidou-Kyniazopoulou, A., and Behnke, S. (2016). NimbRo explorer: semiautonomous exploration and mobile manipulation in rough terrain. *J. Field Rob.* 33, 411–430. doi:10.1002/rob.21592
- Sucar, I. A., and Kavraki, L. E. (2008). "Kinodynamic motion planning by interior-exterior cell exploration," in *Algorithmic Foundation of Robotics VIII (Workshop Proceedings)*, Guanajuato.
- Viola, P., and Jones, M. (2001). "Rapid object detection using a boosted cascade of simple features," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (Kauai: IEEE), 511–518.
- Wisspeintner, T., Van Der Zant, T., Iocchi, L., and Schiffer, S. (2009). RoboCup@Home: scientific competition and benchmarking for domestic service robots. *Interact. Stud.* 10, 392–426. doi:10.1075/is.10.3.06wis
- zu Borgsen, S. M., Korthals, T., Ziegler, L., and Wachsmuth, S. (2015). "ToBI – team of bielefeld: the human-robot interaction system for RoboCup@Home 2015," in *RoboCup@Home League Team Descriptions for the Competition in Hefei, China*. Available at: http://robocup2015.oss-cn-shenzhen.aliyuncs.com/TeamDescriptionPapers/RoboCup@Home/RoboCup_Symposium_2015_submission_96.pdf

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Stückler, Schwarz and Behnke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.