Semi-Autonomous Domestic Service Robots: Evaluation of a User Interface for Remote Manipulation and Navigation With Focus on Effects of Stereoscopic Display

Marcus Mast, Zdeněk Materna, Michal Španěl, Florian Weisshardt, Georg Arbeiter, Michael Burmester, Pavel Smrž & Birgit Graf

International Journal of Social Robotics

ISSN 1875-4791

Int J of Soc Robotics
DOI 10.1007/s12369-014-0266-7

Int J of Soc Robotics
DOI 10.1007/s12369-014-0266-7

Semi-Autonomous Domestic Service Robots: Evaluation of a User Interface for Remote Manipulation and Navigation With Focus on Effects of Stereoscopic Display

Marcus Mast · Zdeněk Materna · Michal Španěl · Florian Weisshardt · Georg Arbeiter · Michael Burmester · Pavel Smrž · Birgit Graf

Accepted: 25 October 2014

Abstract In this article, we evaluate a novel type of user interface for remotely resolving challenging situations for service robots in domestic environments. Our focus is on potential advantages of stereoscopic display. The user interface is based on a control architecture that allows involvement of a remote human operator when the robot encounters a problem. It offers semi-autonomous remote manipulation and navigation with low-cost interaction devices, incorporates global 3D environment mapping, and follows an ecological visualization approach that integrates 2D laser data, 3D depth camera data, RGB data, a robot model, constantly updated global 2D and 3D environment maps, and indicators into a single 3D scene with user-adjustable viewpoints and optional viewpoint-based control. We carried out an experiment with 28 participants in a home-like environment investigating the utility of stereoscopic display for three types of task: defining the shape of an unknown or unrecognized object to be grasped, positioning the gripper for semi-autonomous reaching and grasping, and navigating the robot around obstacles. Participants were able to successfully complete all tasks and highly approved the user interface in both monoscopic and stereoscopic display modes. They were significantly faster under stereoscopic display in positioning the gripper. For the other two task types, there was a tendency for faster task completion in stereo mode that would need to be verified in further studies. We did not find significant differences in perceived workload between display types for any type of task. We conclude that stereoscopic display seems to be a useful optional display mode for this type of user interface but that its utility may vary depending on the task.

Keywords Human-robot interaction · User interfaces · Semi-autonomy · Telemanipulation · Teleoperation

1 Introduction

The prospect of a robot helping at home with daily chores is alluring. Service robots may, for example, enable older adults with physical restrictions to reduce their dependence on caregivers and to remain living in their familiar environment. However, domestic environments pose immense challenges for robots. The environments are heterogeneous, unstructured, and keep changing. They contain objects difficult to recognize with reflective surfaces or covered by other objects, varying illumination, objects difficult to handle, and moving objects like people or pets. Robots need to successfully judge the environment and make complex decisions under uncertainty. Due to these and other challenges, sophisticated service robots with manipulation capabilities that would reliably and autonomously cover a wide range of tasks are unlikely to be realized in the near future.

A pragmatic approach for a shorter-term deployment of multi-purpose service robots in homes is the concept of semi-autonomy [1,2]. If a robot is unable to carry out a task or fails during execution, a human operator may take control remotely, solve the problem, and hand back control to the robot [3–7]. The system may further assist the opera-

M. Mast (✉) · M. Burmester
Stuttgart Media University, Stuttgart, Germany
e-mail: mast@hdm-stuttgart.de

M. Mast
Linköping University, Linköping, Sweden

Z. Materna · M. Španěl · P. Smrž
Brno University of Technology, Brno, Czech Republic

F. Weisshardt · G. Arbeiter · B. Graf
Fraunhofer Institute for Manufacturing Engineering and Automation (IPA), Stuttgart, Germany

tor during teleoperation. Introducing a human in the loop can improve a system's real-world fitness and expand its application range. With continuing progress of technology and augmented by human teaching and machine learning, a robot's autonomy can increase over time, gradually reducing the need for human involvement [8–11].

In this article, we evaluate a user interface for resolving challenging situations for domestic service robots through semi-autonomous remote manipulation and navigation. We particularly focus on potential benefits of stereoscopic display. Studies have suggested that stereo display can be useful in video-based telerobotics (see Sect. 1.4). However, to our knowledge, its effects have not previously been investigated for more advanced user interfaces as the one evaluated, which incorporates semi-autonomous control and an ecological visualization approach that integrates data from various sensors in a 3D scene with user-adjustable viewpoints.

1.1 User Interface Design Challenges

To successfully solve problems remotely with a semi-autonomous system, adequate user interfaces are crucial. For effective and efficient task performance [12], operators should be aware of the remote situation [13–15], be able to build an accurate mental model of the remote environment [15,16], and should not experience high cognitive load [15,17]. For the interface to appeal to users, the user experience [18] should also be engaging and stimulating [19,20].

Achieving these design goals is a major challenge as is reflected in the heterogeneity of interaction approaches in the literature and as has been shown in many user studies [21]. For example, user interfaces for remote navigation relying merely on a video image and a 2D map or displaying information in numerous screen regions have shown to provide insufficient situation awareness [13,22] and lead to high cognitive load [23], especially when using direct teleoperation without any system autonomy [13,22]. To provide a more complete representation of the remote environment, employing multiple and panospheric cameras has been explored [24,25] but this imposes a high load on network bandwidth and does not allow assessing depth in the scene. User performance has shown to improve when enriching 2D map and video with schematic 3D environment models [23,26]. However, this requires manual modeling of each apartment and can be misleading when environment features have changed.

Remote manipulation tends to be even more challenging due to requirements on dexterity, degrees of freedom, and precision as well as increased importance of judging three-dimensional spatial relations. For example, fixed viewpoints on the remote scene [27,28], misalignments between the user's viewpoint and the coordinate system for controlling the end effector [29], and restrictions on the degrees of freedom (DOF) for controlling the end effector [30] have shown to be detrimental for task performance. Direct unassisted telemanipulation is often associated with problems like environment collisions due to transmission delay [31,32] or the need for sophisticated and expensive interaction devices to mirror the robot's dexterous abilities [33–35].

1.2 Promising Approaches

Due to the difficulties with direct teleoperation it is promising to introduce, or retain, some robot autonomy during the teleoperation task at lower and more reliable levels of sensing and actuation. This way a human's sophisticated high-level interpretation skills and world knowledge can be utilized for solving the main problem (often requiring environment interpretation), while relieving them from the load and difficulties associated with low-level control. Promising results have recently been achieved with semi-autonomous, system-assisted remote manipulation [36–38] and navigation [6,16,23,27,39]. As a higher level of robot autonomy may come at the cost of reduced controllability, it can make sense to offer several levels of autonomy [23,26,36,40–42].

Further, as precise depth sensors have become very affordable recently, it seems beneficial to incorporate the display of 3D point clouds and global 3D maps, based on these sensors' data, in teleoperation user interfaces. This should allow users to accurately assess a remote scene in all dimensions during manipulation and navigation tasks, especially when combined with a freely adjustable viewpoint [27,28]. Except for a few applications, e.g., in underwater [43] or indoor manipulation [36,44], the potential of these sensors has not yet been widely leveraged in teleoperation user interfaces. As an application example in remote navigation, we have used 3D mapping techniques [45,46] to create global environment maps from the point cloud sensor data while the robot moves around. We found that displaying such 360° 3D environment representations in the user interface was advantageous for resolving problematic search and navigation situations [16].

1.3 Innovations of the Evaluated User Interface

The user interface evaluated in this article is part of a wider usage concept for assisting elderly people at home [2,47] (see Sect. 2.2). It incorporates several innovative approaches (see also Sect. 2.3):

- An underlying control architecture that allows recognition of robot failure states and smooth handover of control between robot and human remote operators [5]
- System-assisted, semi-autonomous telemanipulation with offline trajectory execution and editing of 3D collision

maps by the user for high controllability of the telemanipulation result yet safe, collision-free operation
- Remote navigation with several control modes of different autonomy, with and without collision avoidance
- An ecological visualization approach that integrates data from various sensors as well as other data like a robot model and collision indicators into a single 3D representation of the remote environment [16]
- Visualization of colored point clouds enhanced by global 3D environment mapping [16,45] with a freely adjustable viewpoint for high situation awareness
- Reliance on low-cost interaction devices in form of a standard mouse for viewpoint adjustments and a 3D mouse for precise and intuitive 6-DOF gripper positioning and 3-DOF robot navigation
- Avoidance of the need for human compensation of display-control misalignment by the ability to control robot and gripper in the coordinate system of the remote operator's chosen viewpoint on the 3D scene
- Optional stereoscopic display mode using contemporary active shutter technology

While most of these approaches have been used previously in an isolated form in various domains (e.g., [3,4,6,22,27, 29,32,36–39,46]), our user interface integrates them into a coherent experience, highly focused on real-world applications.

1.4 Investigating Stereoscopic Display

As operators need to perceive and act in a remote three-dimensional environment, e.g., judging distances and object sizes, depth perception has always been an important issue in telerobotics [21,32,48–50]. There are a variety of cues facilitating depth perception, such as motion parallax (when the observer or scene moves, closer objects move faster), the kinetic depth effect (depth perception from rotational motion of an object), occlusion (an object covering another is perceived as closer), linear perspective (depth perception from judging the angle at which parallel lines converge in the distance), and binocular disparity (the difference in images on the left and right retinae) [51,52].

Binocular disparity, not available when using standard displays and realized through stereoscopic technology, can be a powerful cue to depth. Compared to monoscopic display, stereoscopic display has shown to improve users' task performance and accuracy of environment judgments in a number of domains related to the present user interface like reaching for objects [53], positioning and resizing objects [54], and path tracing [55–57] in virtual environments, video-based robotic telemanipulation [50,58,59], video-based remote robot navigation [49,60,61], and remote robot navigation based on egocentric 3D visualizations extrapolated from 2D laser scanner data [62].

On the other hand, while there is a general tendency that with more available depth cues an observer can better assess depth [63,64], there are experimental results suggestive of no or little added value of stereoscopy if certain other depth cues are present. For example, participants' performance in pick-and-place tasks was nearly equivalent under mono and stereo display conditions when occlusion cues [48] or perspective cues (a grid or reference lines) [65] were present. In a study on remote video-based robot navigation, stereo display lead to significantly fewer collisions than mono display only when haptic feedback cues were not available [49]. A study on object positioning and resizing showed considerable improvements in users' positioning accuracy in the presence of object shadows in mono but not in stereo viewing mode [54].

Some experimental results have suggested diminishing relative merits of stereo display in the presence of motion-based cues (through head-coupled perspective or object rotation), e.g., when disambiguating wireframe images [66] or tracing winding 3D paths [67–69]. However, other studies found substantial benefits of stereo display even when motion-based cues were available, in path tracing [55,56] or object matching tasks [54]. Some studies employed motion through user-controllable scene rotation around one [54,55] or two axes [56,69] but results on the relative merits of stereo display varied with three of these studies [54–56] suggesting benefits but one not [69].

The different results in studies with similar objectives suggest that many factors affect whether and to what extent stereo display improves depth judgments and user performance. In addition to the presence of other depth cues, the value of stereoscopic display has shown to depend on such factors as task [50,53,71], spatial frequency [72,73], image resolution [74,75], distance of the perceived objects [53,71,76,77], and orientation of the scene (or viewer's perspective) [78,79].

Due to the many factors involved and complex interactions between them [52,53,68], it is difficult to predict the effects of introducing stereoscopy in a particular applied context. Further, it should be noted that stereoscopic display is usually associated with drawbacks such as the need to wear glasses, to dim lights, or potential discomfort like eyestrain or dizziness after prolonged use [80,81]. It is therefore important to verify benefits of stereoscopic display in a particular usage scenario before relying on it. The telerobotics-related studies mentioned above investigated egocentric video displays and virtual environments. We are not aware of previous studies investigating the effects of stereoscopic display under parameters similar to the present user interface, namely robotic remote manipulation and navigation tasks with point cloud visualization of real indoor scenes and a fully user-adjustable viewpoint including rotation, panning, and zoom-

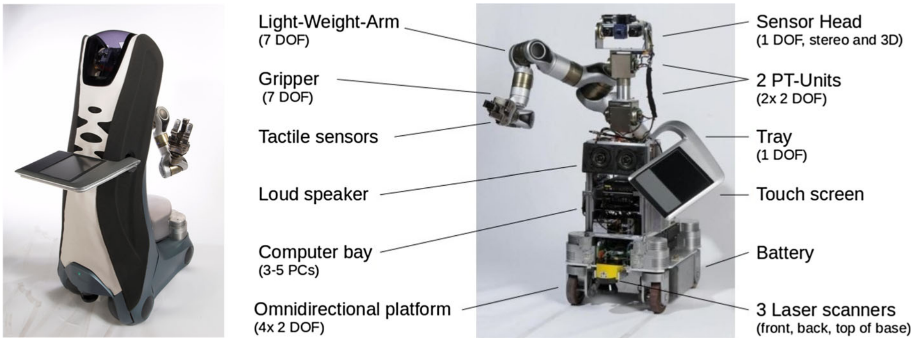


Fig. 1 Care-O-bot 3 and its hardware components

ing. We were therefore interested in investigating effects of stereoscopic display under these conditions.

2 Robot and User Interface

2.1 Robotic Platform

Our implementation is based on the robotic platform Care-O-bot 3 from Fraunhofer IPA (Fig. 1) [82], running ROS (Robot Operating System [83]). The service robot has a mobile base with an omnidirectional drive, an arm with seven degrees of freedom (DOF), a 3-finger gripper, a retractable tray for object handover between robot and people, three 2D laser scanners mounted on the base for 360° registration of the environment at a height close to the floor, a sensor head with an RGB camera, an infrared 3D sensor (Kinect), and a stereo RGB camera. The robot is further equipped with microphones, speakers, and colored LEDs.

For autonomous navigation, the robot builds a 2D environment map from laser scanner data and uses it to localize itself and to plan collision-free paths [82]. The robot can learn and later recognize and localize objects based on textures [84]. It can identify suitable grasp configurations and plan arm trajectories to an object to be manipulated [85,86].

2.2 Overall Usage Concept

The user interface evaluated in this article was developed iteratively in a human-centered design process [18] as part of a broader robotic usage concept for assisting elderly people at home [2,47] in the European project SRS [87]. Over the course of 3 years and 3 months, a total of 430 prospective users were involved in this process (40 % over 65).

In the first phase, we carried out user requirements studies. A focus group study [88] with 59 participants (elderly people living at home, informal caregivers like family or friends, and formal caregivers) determined needs and difficulties of potential user groups. In a subsequent survey, 64 elderly people and 19 informal caregivers rated the usefulness of various robot services [2]. In an ethnographic study, we visited 15 elderly people in their homes, recording the living conditions and determining challenges for service robots. We also visited telemedical and home emergency teleassistance centers to analyze the workplaces, tasks, routines, and artifacts in use [2]. One result of that study was that we consider the skills (e.g., computer proficiency) and educational background (often in nursing or care) of present staff as suitable for assuming the role of robot teleoperators after some training.

In the second phase we carried out technical assessments [2]: an analysis determining frequent instances of robot failure and required teleoperator tasks, an interaction analysis to determine with what kind of interactions teleoperators would provide remote assistance, and an assessment of the suitability of various interaction devices.

In the third phase, we iteratively developed a usage concept and user interfaces for three different user groups. The initial usage concept was evaluated with 30 elderly people, 23 informal caregivers, and 5 professional caregivers in an acceptability study [2] and improved based on the results. The usage concept specifies the following primary user groups:

- Elderly people living at home and in need of moderate forms of assistance with activities of daily living
- Relatives or friends willing to provide assistance but currently not able to or only to a limited extent because they do not live on site or are often away, e.g., due to work
- Professional teleoperators available 24 hours in a call center

As a result of the user requirements studies [2,88], we focused on fetch and carry services (which are the foundation for many other user-requested services), on assistance in cases of emergency (e.g., assessing the situation or opening a door for the ambulance), and on assistance with reaching objects in places problematic for elderly people (high or low) [2,47].

The concept of the user interfaces is shown in Fig. 2. Its foundation is a control architecture that enables the robot to recognize failure states, contact human remote operators for assistance, and reassume control afterwards [5]. Local elderly people are equipped with a portable, smartphone-

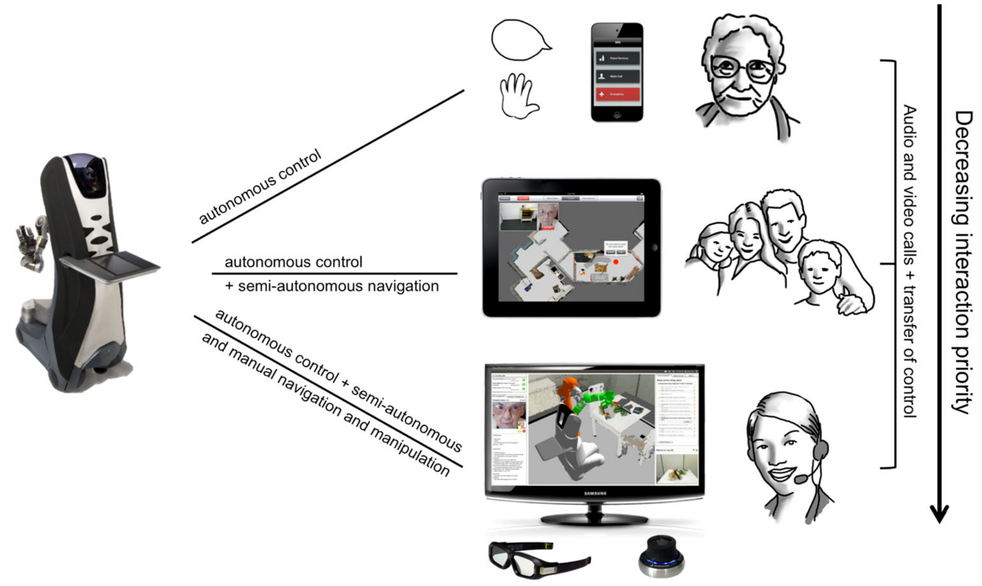


Fig. 2 User interface concept with three user groups: elderly people at home, remote relatives, and professional teleassistants

sized multi-touch device so the robot can be commanded from any position in the apartment. The user interface, designed specifically for the needs of elderly people [2,47], provides access to autonomous robot functions only. If the robot encounters a problem during task execution, by default it first contacts relatives for assistance. They have a tablet computer with extended functionality such as map-based semi-autonomous navigation. In case relatives are unavailable or the functionality of their user interface is not sufficient for solving the robot's problem, the robot contacts the 24-hour teleassistance center. Their user interface is based on a personal computer, a 3D mouse, and optional stereoscopic display. It is the most feature-rich of the three user interfaces and the one evaluated in the present article. Beyond the capabilities of the interface for relatives it most importantly allows semi-autonomous manipulation.

A teleoperation session is only initiated after confirmation of the local elderly user and during teleoperation, an audio or video call remains open. This is to address ethical concerns regarding privacy and loss of control [89].

The three user interfaces were evaluated in ten usability studies [90] at different points in time with a total of 86 participants from all three user groups [2,47]. Evaluations were initially carried out at Stuttgart Media University's User Experience Research Lab using horizontal prototypes (screen sequences simulating interaction) and the Gazebo robot simulator [91], later in a model kitchen at Fraunhofer IPA using implemented user interfaces, and finally in real and model apartments in Germany and Italy. Figure 3 shows impressions of the evaluations. Overall, we exposed 138 usability problems of varying severity in the three user interfaces over the course of the project and addressed them in subsequent development [2,47].

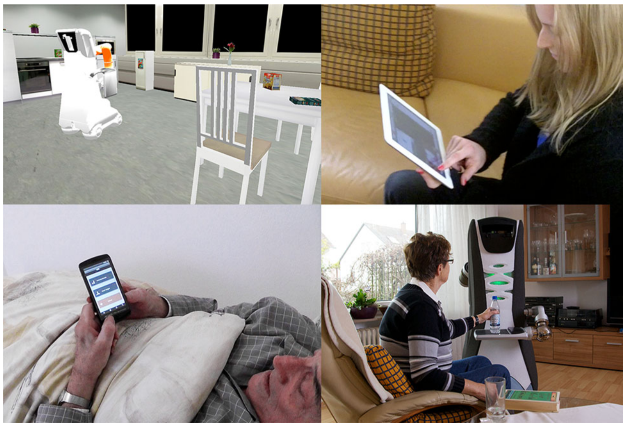


Fig. 3 Usability studies of the user interfaces. *Top left* Gazebo simulator used in early lab-based studies; *top right* evaluation of user interface for relatives; *bottom* evaluation of user interface for elderly people

In the final phase, our evaluation focus turned to specific innovative aspects of the user interface for professional teleassistants. A previously published experiment investigated the usefulness of global 3D environment maps for solving navigation problems remotely [16]. The present article describes our second experiment on this user interface, which investigated both, remote navigation and manipulation under monoscopic versus stereoscopic display.

2.3 Evaluated User Interface

As the last instance in the concept of user interfaces (Fig. 2), the evaluated user interface for professional teleassistants addresses a wide range of robots' possible failures and insufficiencies in reasoning about the environment, and in problem solving, preventing autonomous manipulation or navigation.

Table 1 Examples of robot failures and corresponding teleoperator interventions with the user interface

Robot failure or insufficiency	Reason as assessed by the teleoperator	Possible teleoperator intervention
Unable to plan or execute navigation path to destination	Narrow passage	Remotely navigate the robot through passage with or without autonomous obstacle avoidance
	Obstacles blocking path	Remove obstacles by pushing with robot body or using arm and gripper
Unable to plan arm trajectory to pre-grasp position for recognized object to be fetched	Collision map contains noise or soft objects that may be touched by the arm	Delete noise or passable space in 3D collision map using collision map editing tool and have robot re-plan trajectory
	Robot position not ideal	Navigate robot to better position and re-plan trajectory
	There is a suitable pre-grasp position but robot could not find it	Specify gripper pre-grasp position manually with 3D mouse (e.g., from the top)
	Occluding objects prevent arm trajectory planning	Remove occluding objects using autonomous or semi-autonomous manipulation and re-plan trajectory
Object to be grasped could not be recognized	Object is in a position and orientation fine for detection but recognition still failed	Adjust 3D shape from object database over the object to inform robot of its location and orientation, then have robot plan arm trajectory and grasping
	Object is in the room but robot did not find it	Move robot to proper position and restart detection
	Illumination is insufficient for object recognition	Turn on light in the room or open blinds using remote manipulation, then restart detection
	Occluding objects prevent recognition (cluttered scene)	Remove occluding objects using remote manipulation, then restart detection
Object to be fetched is unknown (not in database)	No assessment necessary	Use remote navigation and manipulation to fetch object manually; optionally, teach object by adjusting a 3D shape to fit the object
Arm is in collision	Moving the robot should be safe	Move robot and recover arm
	Moving the robot may be unsafe	Ask local user to remove objects, then move robot and recover arm

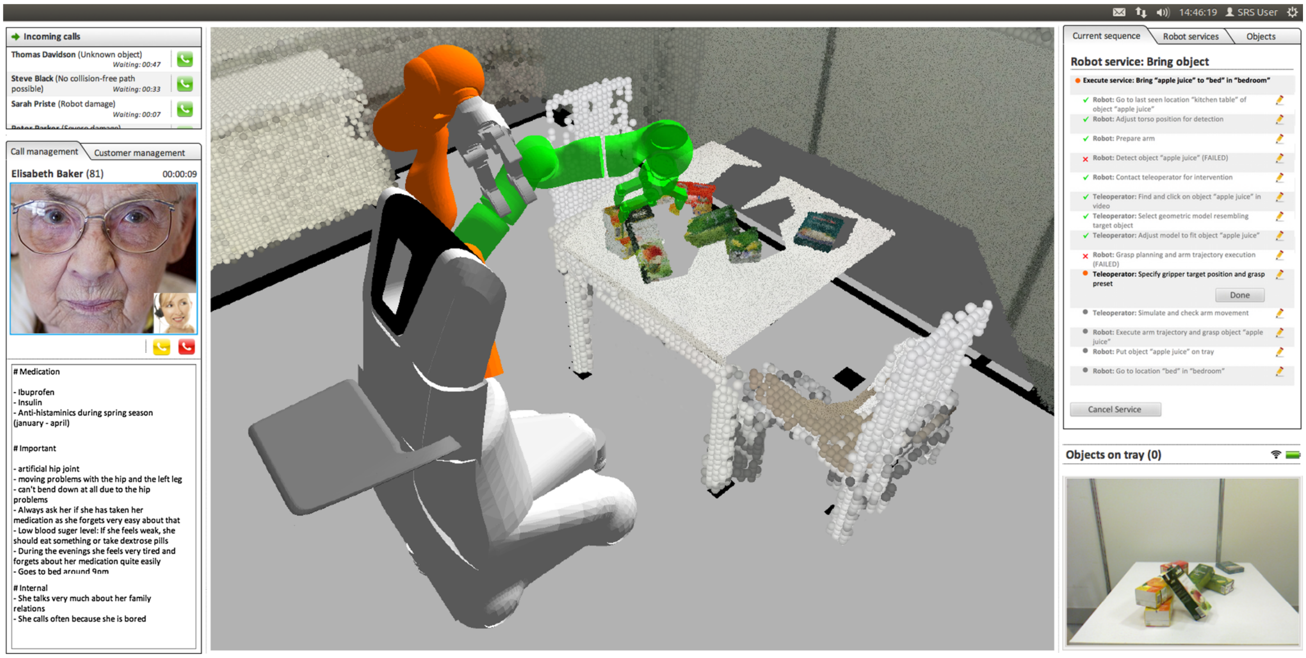
Table 1 shows examples of problems determined in our study of robot failures [2] along with corresponding possible teleoperator interventions.

The user interface was designed to be well usable by current staff of home emergency and telemedical service centers, i.e. users with general computer expertise but with no expert knowledge on robotics [2]. We aimed to achieve this by (1) applying established design principles for user- friendly screen design [92–94] (2) iterative improvement based on usability testing results [2,18,47], (3) giving the robot as much autonomy as possible, with the teleoperator only solving the core problem and then handing back control to the robot, (4) integrative, fused visualization of sensor data (see Sect. 2.3.1), (5) global 3D environment mapping to give operators a complete picture of the situation around the robot (see Sect. 2.3.2; [16]), and (6) intuitive robot navigation and arm control using a 3D mouse and avoiding display-control misalignments (see Sects. 2.3.3 and 2.3.4).

2.3.1 *Functional Areas and Visualization Approach*

An overview of the user interface is shown in Fig. 4. The left pane contains information and functionality related to the elderly customers: incoming teleoperation requests, active audio or video calls, customer management, and a text field for entering individual comments on the customer such as required medication or health problems. The latter functionality was adopted from current systems in use at home emergency teleassistance centers as a result of our ethnographic study [2].

The right pane contains information and functionality related to the robot. The tab “Current Sequence” shows at



Thomas Davidson (Unknown object) Waiting: 00:47
Steve Black (No collision-free path possible) Waiting: 00:33
Sarah Priste (Robot damage) Waiting: 00:07
Peter Parker (Severe damage)

Elisabeth Baker (81) 00:00:09

Medication

- Ibuprofen
- Insulin
- Anti-histaminics during spring season (january - april)

Important

- artificial hip joint
- moving problems with the hip and the left leg
- can't bend down at all due to the hip problems
- Always ask her if she has taken her medication as she forgets very easy about that
- Low blood suger level: If she feels weak, she should eat something or take dextrose pills
- During the evenings she feels very tired and forgets about her medication quite easily
- Goes to bed around 9pm

Internal
- She talks very much about her family relations
- She calls often because she is bored

Robot service: Bring object

● Execute service: Bring "apple juice" to "bed" in "bedroom"
- ✓ **Robot:** Go to last seen location "kitchen table" of object "apple juice"
- ✓ **Robot:** Adjust torso position for detection
- ✓ **Robot:** Prepare arm
- ✗ **Robot:** Detect object "apple juice" (FAILED)
- ✓ **Robot:** Contact teleoperator for intervention
- ✓ **Teleoperator:** Find and click on object "apple juice" in video
- ✓ **Teleoperator:** Select geometric model resembling target object
- ✓ **Teleoperator:** Adjust model to fit object "apple juice"
- ✗ **Robot:** Grasp planning and arm trajectory execution (FAILED)
- ● **Teleoperator:** Specify gripper target position and grasp preset
- ● **Teleoperator:** Simulate and check arm movement
- ● **Robot:** Execute arm trajectory and grasp object "apple juice"
- ● **Robot:** Put object "apple juice" on tray
- ● **Robot:** Go to location "bed" in "bedroom"

Objects on tray (0)

Fig. 4 User interface during semi-autonomous manipulation

which point in the robot's autonomous action sequence the problem occurred and what the teleoperator can do to resolve it. The tab "Robot Services" provides access to autonomous services as in the user interfaces for elderly people and relatives. The tab "Objects" provides access to the robot's database of known objects with editing and teaching functionalities. A resizable and movable video image of the robot's current view is shown in the lower part of the right pane. Features in the left and right panes have only been partially implemented at this point, as our development focused on the core functionality for solving problems, accomplished with the central area.

The visualization in the central area of the user interface is based on the concept of ecological interfaces [22,93,95,96], where affordances of the environment [95,96], i.e., what it offers (e.g., regions the robot can pass or objects that can be grasped), can be directly inferred from the scene without further cognitive effort. The visualization is realized using RViz [97] and employs sensor fusion [98,99] to integrate data from various sensors into a single 3D representation of the remote environment. The approach addresses a number of problems with conventional video and 2D map-based interfaces found in the literature, such as limited field of view [22,26,100], unawareness of the robot's surroundings [13,14,100], and divided attention due to task-relevant information in different areas of the screen [14,22,23].

The viewpoint on the 3D scene is freely adjustable by the user through mouse manipulations, including rotation (using the two-axis valuator technique, which has shown to be better usable than other techniques [101]), panning, and zooming with the wheel. This allows choosing robot-centric or exocentric perspectives with zoom levels appropriate for problem at hand. User studies on remote navigation and manipulation have suggested that an adjustable viewpoint is advantageous [27,28].

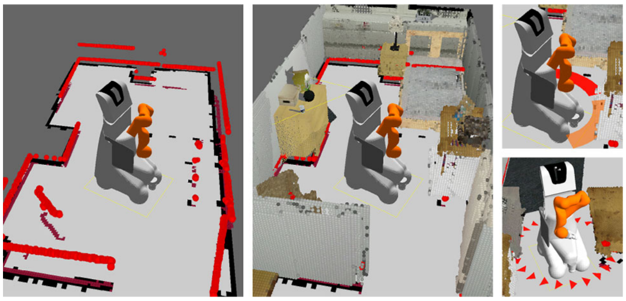


Fig. 5 Visualization elements of the 3D scene

The visualization elements of the 3D scene are shown in Fig. 5. Elements can be turned on and off individually for specific situations but all are visualized by default. The left shows: (1) robot with realistic shape and dimensions, (2) safety-relevant area around robot ("footprint"; yellow rectangle), (3) 2D floor map based on laser data (grey and black), (4) live 2D laser data from all three scanners (red), (5) obstacle map (recent laser data; purple). In the middle of Fig. 5, further elements are added: (6) colored point cloud of fused depth and RGB data in the robot camera's current field of view (640 × 480 pixels), (7) field of view boundaries (two yellow lines), (8) global 3D environment map outside the robot's field of view. The right of Fig. 5 shows (9) in-scene

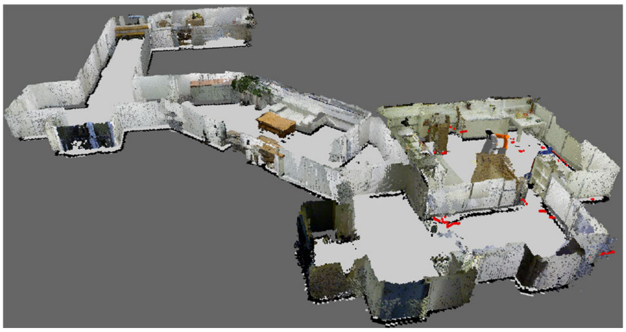


Fig. 6 User interface showing multi-room environment mapped using voxel-based global 3D environment mapping technique

collision indicators for translational (top) and rotational (bottom) movements appearing when navigating the robot toward an obstacle (in collision avoidance mode, this is associated with the robot slowing down and eventually coming to a stop before the obstacle).

2.3.2 Global 3D Environment Mapping

A central innovation of the user interface is its 360° 3D environment representation. While the robot moves around, it generates and continuously updates a global 3D environment map, combining depth and RGB sensor data. Two techniques are available: A voxel-based technique uses octrees to represent 3D occupancy grids and is based on the OctoMap framework [46]. This technique was used in the present evaluation. A fully mapped environment is shown in Fig. 6. As an alternative consuming less network bandwidth, a geometric mapping technique is available [16,45]. 3D mapping is a significant improvement over previous approaches that relied on manual modeling or extrapolation from 2D data [22,26,62] due to high realism, detail, currency, and automated generation. It allows assessing what is behind, right, and left of the robot, which has been a major problem in robot teleoperation [13,14,100].

2.3.3 Robot Navigation

Semi-autonomy in remote navigation is often found in the form of a human operator steering the robot while the robot avoids obstacles [6,23,27,39]. Lower and higher levels of autonomy have been explored too [23,26]. In our user interface, four control modes with different robot autonomy can be chosen according to the current situation:

1. Have the robot plan an autonomous path by clicking on a destination point in the 3D scene with the mouse (high robot autonomy; collision avoidance based on 2D laser data active)
2. Dragging and holding with the mouse an in-scene disc that the robot attempts to follow autonomously (medium robot autonomy; collision avoidance active)
3. Using in-scene navigation controls around the robot with the mouse—arrows for translational and a ring for rotational movement (low robot autonomy; collision avoidance optional)
4. Using the SpaceNavigator 3D mouse [102] (low robot autonomy; collision avoidance optional): The user can choose between control in the coordinate system of the robot (i.e., pressing left moves to robot to the robot's left) versus the user's currently chosen viewpoint on the 3D scene (i.e., pressing left moves the robot to the remote user's left). The viewpoint mode was designed to avoid cognitive effort for mental transformations due to display-control misalignments [29] and was specified as the default mode based on user observation and interviews in pre-tests.

Mode 4 with 2D collision avoidance and viewpoint-based control was used in the present evaluation.

2.3.4 Manipulation

Another central innovation of the user interface is the approach for semi-autonomous remote manipulation. The approach is based on collision-free arm trajectory planning [86] using a global environment map [16,46] and offline motion execution, which makes the user interface fully robust against transmission delays that have been a long-standing problem in telemanipulation [31,32,103]. Despite autonomous execution, the user still has a high level of control over the process through various possible interventions.

Figure 7 shows the main interaction steps and resulting robot actions for reaching and grasping an untrained object or a trained object in the database that could not be recognized (e.g., due to a difficult pose, clutter, or low illumination). The scenario may be that the local elderly user commanded the robot to fetch the apple juice box but the robot failed to recognize it, stopped, and called a teleoperator for help (Fig. 7(1)). The teleoperator first chooses from a library a predefined basic 3D shape (e.g., box or cylinder) resembling the target object and fits it over the object (Fig. 7(2)–(4)) by using in-scene arrows for moving and resizing and in-scene rotation wheels with the mouse. This informs the collision avoidance algorithm that the defined space will later, when grasping and moving the object, be passable for the arm and gripper and also that this space near the gripper needs to be taken into account for avoiding collisions of the object with the robot or environment. In case the object was recognized (but the robot still could not find a suitable grasping configuration or plan an arm trajectory), steps 2 to 4 are not necessary.

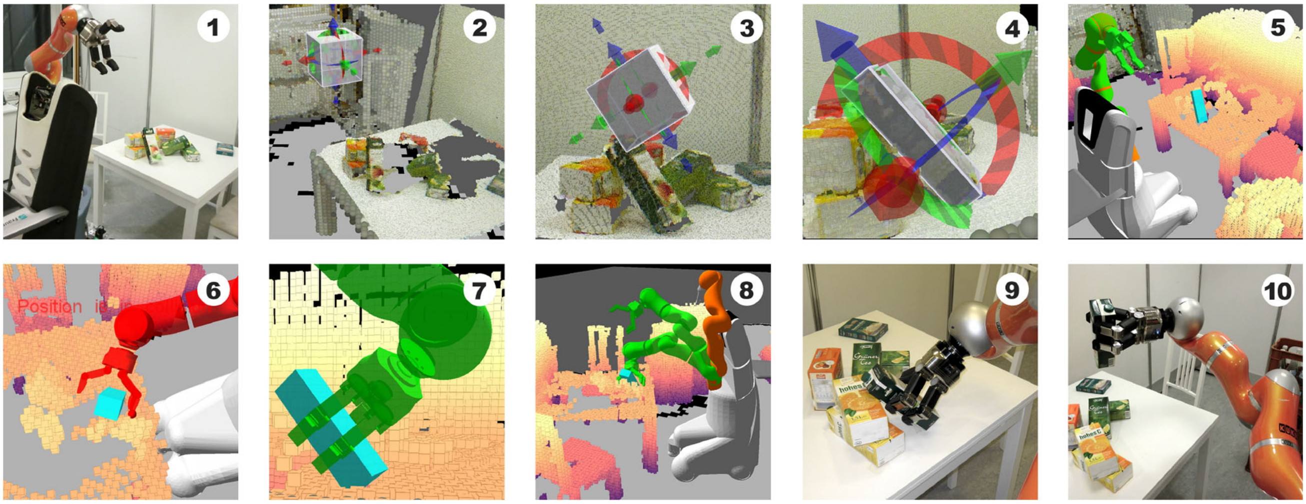


Fig. 7 Main interaction steps and robot actions for semi-autonomous reaching and grasping of an untrained or unrecognized object

In steps 5 to 7 in Fig. 7, the teleoperator specifies a suitable gripper target position and orientation. This is performed with the 3D mouse. As is default for 3D mouse based navigation in our user interface (see Sect. 2.3.3), gripper control too is performed in the viewpoint coordinate system. The environment visualization changes from colored point cloud to the more task-relevant collision map display at this point. The 360° 3D collision map is based on the data of the constantly updated global model. The arm turns red if the target position is unreachable or in collision with the environment (Fig. 7(6), red ball indicates point of collision).

When a suitable gripper target position has been specified (Fig. 7(7)), the teleoperator is shown an animation of the planned arm trajectory (Fig. 7(8)). The operator checks from various angles if it would be safe to execute or if the arm may come too close to delicate objects or to people. When satisfied with the trajectory, the operator executes it and watches the arm move in real-time in the 3D scene and RGB video stream. For grasping, the operator chooses a grasp strategy preset (e.g., for soft or for hard objects) and executes the grasp (Fig. 7(9)). Subsequently, unless there is a further problem, control is transferred back to the autonomous system for finishing the service, e.g., placing the object on the tray and delivering it to another room (Fig. 7(10)).

As a further means of intervention, the operator can edit the 3D collision map with an approach similar to steps 2 to 4 in Fig. 7 to remove noise or to specify regions the arm may touch (e.g., uncritical soft objects) while still avoiding all real obstacles.

2.3.5 Stereoscopic Display

The optional stereoscopic display mode investigated in this article was realized using Nvidia 3D Vision 2 wireless shutter glasses and an Asus VG278H 27” 1,920 × 1,080 LCD monitor with 2ms response time, Nvidia LightBoost, and 120 Hz for flicker-free stereoscopic display. We modified RViz [97] and the underlying Ogre3D library [104] to support this type of stereoscopic display. The stereo mode displays the 3D scene stereoscopically with 2D RGB video.

3 Research Questions

The main question we investigated was how stereoscopic display affects users’ task performance and cognitive workload in manipulation and navigation tasks with the present novel type of user interface. As laid out in Sect. 1.4, some previous studies have suggested benefits of stereo display while others have not, which is likely due to a range of differences between those studies, most notably in availability of other depth cues, in tasks, in distances of objects, and in viewpoints on the scene. Compared to previous studies (see Sect. 1.4), distinct characteristics of our user interface, potentially affecting whether or not stereoscopic display will be beneficial, are:

- *Fully user-adjustable viewpoints on the 3D scene, including rotation, translation, and zooming:* This feature introduces particular uncertainty as it is a priori unknown how often users will change the viewpoint (affecting availability of motion-based depth cues, e.g., during scene rotation), which viewpoints they will adopt (affecting the viewpoint factor), and how far zoomed in they will work (affecting distances of objects).
- *Realistic sensor-based 3D representation of the remote environment with display of colored point clouds enhanced by global 3D environment mapping:* This means that rich environmental cues are available, such as textures and shadows. As users can rely on those for assessing

depth, this could be regarded an indication against finding benefits of stereo display. On the other hand, such cues are also available in video-based displays, where studies have suggested benefits of stereo display in several settings [49,50,58–60].

- *Various, partially novel, types of task for semi-autonomous manipulation and navigation:* As task is an influential factor [50,53,71], effects may vary according to the task. Likely, users' viewpoint adjustment behavior also depends on their current task. While robot navigation as implemented in the present user interface can be considered a fairly common type of task, target object shape definition by fitting a 3D primitive and specifying the gripper target position in a 3D environment representation represent more uncommon and novel types of task from an interaction perspective.

We are not aware of previous studies on effects of stereoscopic display with a user interface with similar characteristics. Our primary goal thus was to investigate effects of stereo display on users' task performance and cognitive workload for this contemporary type of user interface. We were further interested in secondary aspects such as the magnitude of effects if present, potential discomfort due to stereo display (e.g., eyestrain or dizziness, see Sect. 1.4), and a potential correlation between collisions and task completion time in remote navigation (e.g., stereo display may lead to fewer collisions but at the cost of longer task completion).

Apart from the focus on stereoscopic display, we also aimed to obtain an overall assessment of the suitability of the user interface and system with its various integrated novel approaches. This included ratings of usability, user experience, quality of users' spatial mental models of the remote situation, and system usefulness. While effects of display type cannot be ruled out for these variables, our main focus here was not on differences between display modes. As part of the overall evaluation focus, we were also interested in technical system performance.

4 Method

4.1 Experimental Design

Our experimental design employed display type (mono, stereo) as a between-subjects variable. Effects of display type were investigated for three types of task (within-subjects): target object shape definition (manipulation subtask 1), positioning of the gripper at its target position (manipulation subtask 2), and robot navigation around obstacles. Manipulation with its two subtasks was performed in two environments (two rooms with different objects).

4.2 Participants

Twenty-eight people participated in the experiment (14 in each group). Participants' age ranged from 20 to 37 years (mono: $M = 25.7$, $SD = 4.9$; stereo: $M = 26.5$, $SD = 4.5$). All participants were male to avoid a confounding effect due to known gender differences in spatial problem solving [105]. Participants had no prior experience with robots, teleoperation user interfaces, or 3D mice. Participants' mean weekly computer usage ranged from 10 to 60 hours (mono: $M = 30.6$, $SD = 16.4$; stereo: $M = 32.4$, $SD = 12.3$). Participants' cumulated mean monoscopic professional 3D application usage and monoscopic 3D gaming was between 0 and 3 hours a week (mono: $M = 0.93$, $SD = 1.20$; stereo: $M = 0.80$, $SD = 1.00$). All participants stated to be engaged with stereoscopic 3D applications and 3D games 0 hours per week. All participants had university degrees or were currently pursuing such. Studied disciplines varied (e.g., information science, aeronautics, business administration) with a tendency toward more technical subjects. Participants received a compensation of €40. Participants requiring vision aids when working at a computer screen wore them during the experiment.

Prior to the experiment, participants underwent several tests and were assigned to experimental conditions based on achieved scores in a balanced way, particularly ensuring an equal balance of outliers: (1) Lang I test for stereopsis (mono: $M = 2.93$ out of 3, $SD = 0.27$; stereo: $M = 3.00$, $SD = 0.00$), (2) abbreviated Vandenberg Mental Rotations Test for spatial ability [106] (mono: $M = 4.00$ out of 6, $SD = 1.62$; stereo: $M = 4.29$, $SD = 1.73$), (3) Snellen visual acuity test (all scores 5 or higher; mono: $M = 6.57$, $SD = 1.16$; stereo: $M = 6.43$, $SD = 1.02$), (4) a brief color vision test where all participants but one (assigned to the mono condition) achieved the full score.

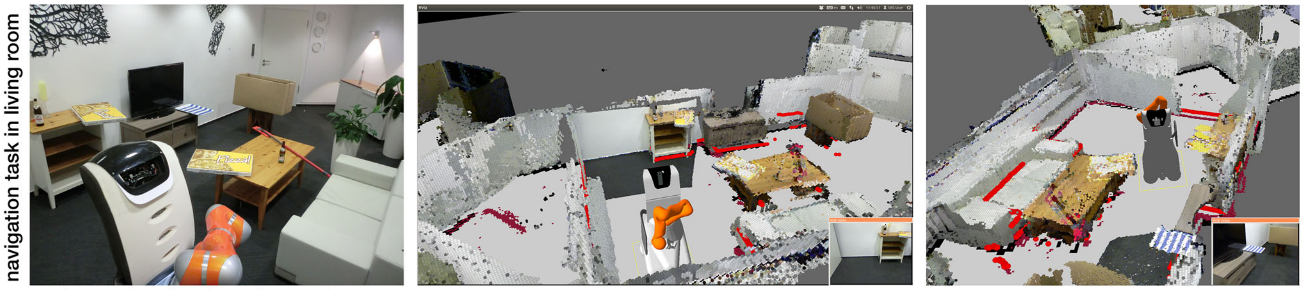
4.3 Site, Equipment, and User Interface Settings

The robot's environment consisted of three rooms connected by corridors: a kitchen, a living room, and a bedroom (see Fig. 6 for an overview and Fig. 8 for pictures). The rooms are offices and a model kitchen under normal conditions and were largely cleared and equipped with 80 household furniture items and objects to simulate a realistic home environment. To control for illumination, we covered all windows and used interior light only. The only robot-specific modification we made to the environment was to attach tape strips at the bottom of furniture items in larger spaces without legs so the robot's laser-based 2D collision avoidance for objects approximately at floor height would work reliably.

task overview (goal: grasp tilted apple juice box)

whole user interface at beginning of subtask 1 (target object shape definition)

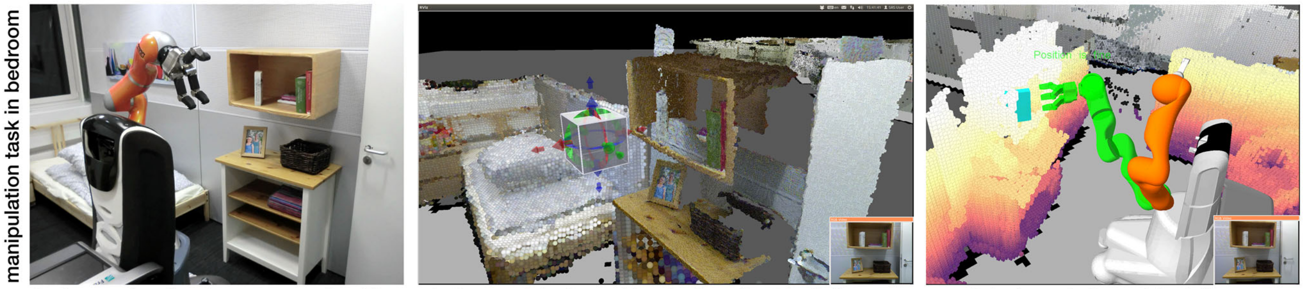
user interface after arm trajectory execution (post-subtask 2)



task overview (goal: navigate robot through the room without colliding)

whole user interface at beginning of task

user interface during task



task overview (goal: grasp grey book)

whole user interface at beginning of subtask 1 (target object shape definition)

user interface during subtask 2 (gripper positioning)

Fig. 8 The experiment's three tasks, each illustrated with overview picture (*left*), screenshot of full-screen user interface at beginning of task (*middle*), and user interface detail during or after task (*right*)

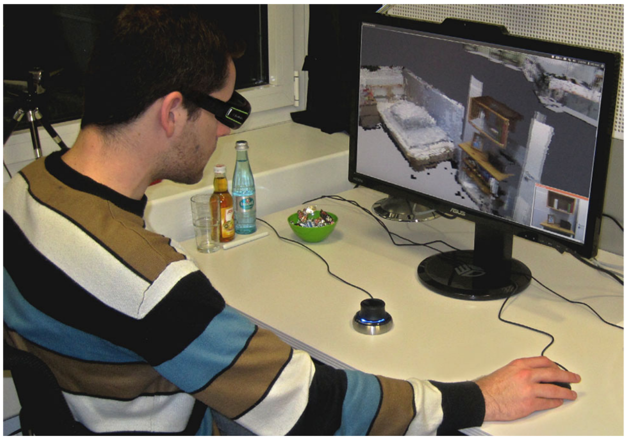


Fig. 9 Participant in front of user interface with stereo glasses, standard mouse, and SpaceNavigator 3D mouse

Participants operated the robot from a separate fourth room (Fig. 9) and did not see the robot or its environment until after the trials. They used a standard PC with mouse, SpaceNavigator 3D mouse for robot navigation and arm control (fixed to the desk; also see Sects. 2.3.3 and 2.3.4), a 27" LCD, and in the stereo condition Nvidia Vision 2 shutter glasses (also see Sect. 2.3.5). The user interface was displayed with the central part (3D scene) occupying the full screen and a small RGB video window overlaid in the lower right (as shown in Fig. 8, middle column). Voxel-based environment mapping (see Sect. 2.3.2) with a spatial resolution of 2.5 cm and viewpoint-based control (see Sects. 2.3.3 and 2.3.4) were active for all tasks. We set display brightness to a subjectively similar value and used dimmed room illumination in both conditions.

4.4 Procedure

Participants first underwent vision and spatial ability tests (see Sect. 4.2) and were assigned to an experimental group based on their scores. They then received a 1-hour training on robot hardware and sensors, usage concept and user interfaces, 3D mouse usage, and remote manipulation and navigation with the user interface connected to the Gazebo simulator (display type according to the participants' assigned condition). Participants then carried out three tasks (in fixed order due to organizational constraints):

1. Manipulation task in the kitchen (Figs. 7 and 8, upper row; including both subtasks as described in Sect. 2.3.4): Participants were asked to make the robot grasp a tilted apple juice box in a cluttered scene on a table.
2. Navigation task in the living room (Fig. 8, middle row): Participants were asked to navigate the robot to the other end of the room, through a course with five elevated, protruding obstacles. Participants used navigation mode 4 (see Sect. 2.3.3) with 2D collision avoidance (only avoiding objects close to the floor—not the elevated ones)
3. Manipulation task in the bedroom (Fig. 8, lower row; both subtasks): Participants were asked to make the robot grasp a book in a shelf, representing a potentially problematic situation for the robot's object recognition with one object being contained in another.

Before each task, the evaluator described a realistic problem of the robot and the task's goal. In the navigation task, the goal was to pass a finish line near the end of the room, ideally without hitting an obstacle. In the manipulation tasks, the goal was to reach a specified degree of precision that should allow for successful manipulation (subtask 1: less than 0.5 cm divergence of the 3D primitive on any side of the target object; subtask 2: gripper approximately centered at a previously specified side of the object and enclosing it as much as possible). Participants were asked to solve each task quickly but without becoming stressed or sacrificing accuracy. The user interface was minimized at the beginning of each (sub-) task. When participants were ready to start, they maximized it.

The navigation task was completed when the robot passed the finish line. For both manipulation subtask types, participants were instructed to minimize the user interface as soon as they thought they had reached the success criterion. The evaluator then objectively assessed whether or not the success criterion was reached. If it was not, participants were asked to improve accuracy. This process was repeated until the success criterion was reached. After completing the second manipulation subtask (specifying the gripper position), the arm trajectory was first simulated and then executed; then grasping was executed. To allow assessing the grasp, the robot lifted the object a few centimeters.

After each (sub-) task, participants rated their perceived cognitive load (see Sect. 4.5). After the last task, there was a 5-minute recapitulating qualitative interview on participants' positive and negative impressions. Lastly, participants filled in a post-experiment questionnaire. Sessions were recorded with video cameras and screen capturing.

4.5 Measures

To assess participants' performance, we measured task completion time for all (sub-) tasks. For the manipulation tasks, only the time to reach the success criterion was considered. Periods between attempts as well as trajectory simulation, execution, and grasping were not included in task completion time. For the navigation task, we further measured the number of collisions with objects. We further assessed for each (sub-) task the perceived cognitive workload with the NASA Task Load Index (NASA-TLX) [107] in its unweighted (raw) variant.

In a post-experiment questionnaire participants rated their level of discomfort on a 7-point scale ("no discomfort at all" to "very strong discomfort") and described the discomfort, if any. They further rated perceived pragmatic (usability) and hedonic (stimulation, identification) qualities of the user experience with the AttrakDiff mini instrument [108,109]. The perceived quality of participants' spatial mental model of the remote environment was assessed with the subscale "spatial situation model" (SSM) from the Spatial Presence Questionnaire MEC-SPQ [110,111]. We further included three questions on the suitability of control modes for the three task types and one on the usefulness of the system as a whole (i.e. including robot and overall usage concept).

To assess system stability, we recorded task abortions due to technical problems. To evaluate the system's executional accuracy in manipulation, two robotics experts rated the quality of each robot movement (trajectory in, grasp, trajectory out) in percent (0 % meaning absolute fail and 100 % meaning fully adequate).

4.6 Data Analysis

We analyzed effects of display type separately for each type of task. For each of the two manipulation subtask types, we carried out analyses of variance (ANOVA) for task completion time and for perceived workload, with display type (mono, stereo) as between-subject factor and environment type (kitchen, bedroom) as within-subject factor. We were not interested in potential main effects of environment type or in interaction effects of environment type and display type as we did not counterbalance environments due to organizational constraints and thus there was a confounding factor due to potential order effects. For the navigation task, we carried out a Wald test assuming Poisson distribution for collisions and two-tailed independent-samples t tests for task completion time and perceived workload. We adjusted these seven main tests for multiple testing with Bonferroni, leading to a significance level adjusted from $\alpha = .05$ to $\alpha = .007$.

To estimate the magnitude of effects, we calculated a confidence interval for difference in means whenever a significant main effect of display type was found. To examine a potential relationship between time to complete the navigation task and collisions, we calculated a two-tailed Kendall's tau correlation. To investigate effects of display type for the discomfort question and for overall system evaluation questions we used two-tailed independent-samples t tests. We adjusted the group of eight system evaluation questions with Bonferroni from $\alpha = .05$ to $\alpha = .006$.

5 Results

5.1 Task Performance and Cognitive Workload

5.1.1 Manipulation Subtask Type 1: Target Object Shape Definition

At the adjusted significance level of $\alpha = .007$, we found no significant main effect of display type on task completion time ($F(1, 26) = 4.44$, $p = .045$, $\eta_p^2 = .146$). We also found no significant main effect of display type on perceived cognitive load ($F(1, 26) = 2.26$, $p = .145$, $\eta_p^2 = .080$). Table 2 shows means, standard deviations, confidence intervals, and p values.

5.1.2 Manipulation Subtask Type 2: Gripper Positioning

At the adjusted significance level of $\alpha = .007$, there was a significant main effect of display type on task completion time ($F(1, 26) = 13.79$, $p = .001$, $\eta_p^2 = .347$). Task completion time was significantly lower in the stereo condition ($M = 172$ s, $SD_p = 44.8$ s) than in the mono condition ($M = 239$ s, $SD_p = 72.2$ s). The 95 % confidence interval for difference between means revealed an expectable advantage of stereo display in the range of 30 to 104 seconds (or 12.5–43.6 %). We did not find a significant main effect of display type on perceived cognitive load ($F(1, 26) = .357$, $p = .555$, $\eta_p^2 = .014$). Table 3 shows means, standard deviations, confidence intervals, and p values.

5.1.3 Robot Navigation Task

At the adjusted significance level of $\alpha = .007$, we did not find significant differences between display conditions for task completion time ($t(26) = 2.37$, $p = .025$), number of collisions ($\chi^2(1, N = 28) = .991$, $p = .319$), or perceived cognitive workload ($t(26) = 0.40$, $p = .969$). Table 4 shows means, standard deviations, confidence intervals, and p values.

Table 2 Target object shape definition: means, pooled standard deviations, 95 % confidence intervals based on estimated model parameters, p values

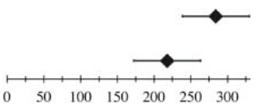
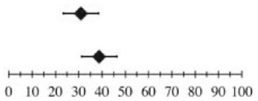
Measure		M	SD_p	M and 95 % CI	p
Task completion time in sec	Mono	285	110.1		.045
	Stereo	219	79.0		
Cognitive workload (0–100)	Mono	31.0	15.7		.145
	Stereo	38.8	14.3		

Table 3 Gripper positioning: means, pooled standard deviations, 95 % confidence intervals based on estimated model parameters, p values

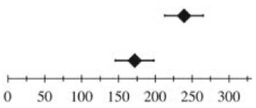
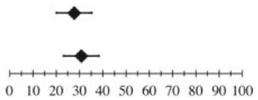
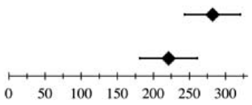
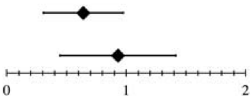
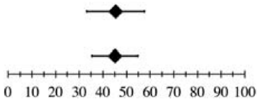
Measure		M	SD_p	M and 95 % CI	p
Task completion time in sec	Mono	239	72.2		.001
	Stereo	172	44.8		
Cognitive workload (0–100)	Mono	27.7	14.5		.555
	Stereo	30.8	14.9		

Table 4 Robot navigation: means, standard deviations, 95 % confidence intervals, p values

Measure		M	SD	M and 95 % CI	p
Task completion time in sec	Mono	282	67.3		.025
	Stereo	221	69.0		
Number of collisions	Mono	0.64	0.63		.319
	Stereo	0.93	0.92		
Cognitive workload (0–100)	Mono	45.4	22.0		.969
	Stereo	45.1	17.5		

Exploratory analysis showed no significant correlation between number of collisions and task completion time ($\tau_b = .000$, $p = 1.000$).

5.2 Discomfort

Mean discomfort ratings were not significantly different between mono ($M = 1.43$, $SD = 1.09$, Max = 5) and stereo ($M = 1.71$, $SD = 0.91$, Max = 3) conditions, $t(26) = -0.75$, $p = .459$. In the mono group 3 participants (21 %) reported some discomfort, in the stereo group 6 participants (43 %). Among participants who reported discomfort, responses in the mono group included fatigue and dry eyes. Responses in the stereo group included fatigue, eyestrain, slight headache, and slight dizziness.

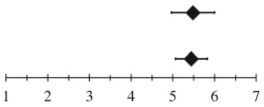
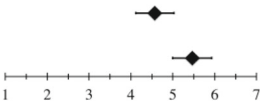
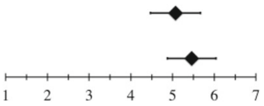
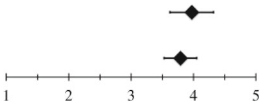
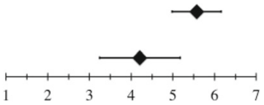
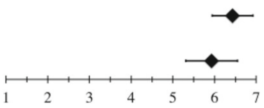
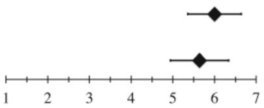
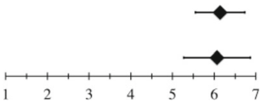
5.3 Technical Performance

All participants were able to complete all tasks without technical system failures. Trajectories planned by the robot were always considered executable without re-planning, however we considered them to sometimes use too much space around the robot. The arm or gripper never collided with the environment, although there were a few close calls. Robotics experts' ratings of executional manipulation accuracy were overall high ($M = 94.8$ %, $SD_p = 8.41$ pp, Min = 40 %, Max = 100 %). In the cases with ratings below 100 %, the arm had sometimes not reached the target position with sufficient precision, came too close to other objects, or the grasped object was in an unstable position in the gripper (e.g., tilted).

5.4 User Experience, Spatial Mental Model, and System Usefulness

Table 5 shows the results of participants' ratings on user experience (AttrakDiff mini instrument [108,109]), perceived quality of spatial situation models (MEC-SPQ SSM [110,111]), suitability of control modes for manipulation subtask 1 (standard mouse), subtask 2 (3D mouse) and navigation (3D mouse), as well as on whole-system usefulness. At the adjusted significance level of $\alpha = .006$, significant differences between means of mono and stereo conditions were found only for AttrakDiff hedonic quality ratings ($t(26) = -2.96$, $p = .006$).

Table 5 Post-experiment ratings: means, standard deviations, 95 % confidence intervals, p values

Measure		M	SD	M and 95% CI	p
AttrakDiff pragmatic quality (1-7)	Mono	5.48	0.90		.905
	Stereo	5.45	0.66		
AttrakDiff hedonic quality (1-7)	Mono	4.57	0.79		.005
	Stereo	5.46	0.80		
AttrakDiff attractiveness (1-7)	Mono	5.07	1.04		.319
	Stereo	5.46	1.01		
MEC-SPQ spatial model (1-5)	Mono	3.97	0.60		.361
	Stereo	3.79	0.45		
Control suitability shape def. (1-7)	Mono	5.57	1.02		.017
	Stereo	4.21	1.67		
Control suitability gripper positioning (1-7)	Mono	6.43	0.85		.183
	Stereo	5.93	1.07		
Control suitability navigation (1-7)	Mono	6.00	1.11		.424
	Stereo	5.64	1.22		
Whole-system usefulness (1-7)	Mono	6.14	1.03		.878
	Stereo	6.07	1.39		

5.5 Recapitulating Interview and Observation of Usage Behavior

Qualitative data from the recapitulating interview and observation of users' behavior provided insights into participants' impressions and strategies. Overall participants got along well with the user interface and described it as, e.g., "well usable", "easy", or "intuitive". A higher resolution of the 3D scene was often suggested by participants as a way to improve the user interface. The control modes for navigation and gripper positioning, both relying on the 3D mouse, were highly approved by participants. Some usability problems occurred with the control mode for target object shape definition (see Fig. 7(2)–(4)), which relies on a standard mouse. As the Nvidia stereo technology did not support drawing of the cursor and a special 3D cursor would have introduced other problems, the mouse pointer was drawn as a 2D overlay. This meant that, depending on the depth of the scene, there often appeared to be two mouse pointers in stereo mode. We taught strategies for coping with this in the training, e.g., hovering with the mouse pointer over the arrow of the 3D primitive and paying attention to when it lights up. As the arrows were quite large and due to the highlighting on mouseover, most participants were able to cope with this deficiency but it still likely affected task completion times in the stereo group to some degree. Further, sometimes environment features covered arrows of the 3D primitive but as arrows were on both sides there was usually a straightforward solution to this.

We could not identify a homogenous strategy of how participants achieved the tasks' goals. For example, some participants first adjusted the 3D primitive's size and then its position while others did it vice versa. With regard to viewpoints, some participants first rotated and zoomed the scene to get an overview, then chose a preferred viewpoint and often adjusted it. Others worked from the initial viewpoint for a long time and rarely changed it even if it meant they sometimes worked from non-ideal viewpoints. Participants tended to change viewpoints less often when positioning the gripper, compared to fitting the 3D primitive. In the navigation task, there seemed to be a tendency to adopt a bird's eye view and work more zoomed out than in manipulation tasks.

Regarding benefits of stereoscopic display, in the mono group we observed instances where participants, after changing the viewpoint, were surprised about how unsuitably they had positioned the 3D primitive or gripper using the previous viewpoint. We did not observe such instances in the stereo group and attribute it to a lack of depth information.

6 Discussion

6.1 Effects of Display Type on Task Performance and Cognitive Workload

Regarding effects of display type on task performance, the most prominent finding of this study is the highly significant temporal advantage of stereo display for the task of positioning the gripper for grasping. We regard the estimated temporal advantage between 30 and 104 s relevant for practical applications. This shows that even for contemporary types of user interface relying on semi-autonomy, point cloud display, and free viewpoint adjustment, stereoscopy can bring substantial benefits. This is noteworthy because previous studies on robotic telemanipulation suggesting benefits of stereo display have widely relied on video-based displays with a fixed viewpoint and no or little system autonomy [50,58,59].

It may seem contradictory that studies on pick-and-place tasks have suggested no or little added value of stereo display in the presence of occlusion cues [48] or perspective cues [65], which were also available in our study. This highlights the difficulties associated with predicting effects of stereoscopy as their occurrence also depends on a variety of further factors (see Sect. 1.4). Differences in tasks, environments, or viewing angles may account for the different results.

Ambiguous results have been obtained in previous studies regarding the question of whether or not stereo display is advantageous when motion-based depth cues are present [54–56,66–69]. The particular case in our study was that these were only present when users changed the viewpoint. It is noteworthy that stereo display was still beneficial for the gripper-positioning task, despite users' ability to obtain motion-based depth cues whenever needed through viewpoint changes.

The longer task completion times in the monoscopic condition for the gripper-positioning task may primarily be due to depth misjudgments and necessary corrections of the gripper position, as we observed in some instances (see Sect. 5.5). They may also in part be explained by additional time taken for viewpoint changes in order to obtain a depth percept through motion of the scene. This would be in agreement with a study on path tracing with user-controllable scene rotation where it was found that rotation was used for longer periods under mono display than under stereo display [55].

For the two other investigated task types, i.e. target object shape definition for semi-autonomous manipulation of unknown or unrecognized objects and navigating the robot, there was no statistically significant difference between display conditions after correction for multiple testing. However, it should be noted that, without multiplicity correction, there was a nominally significant temporal advantage of stereo display for both task types. This can be interpreted as an indication for potential advantages that would need to be confirmed in further studies.

As for target object shape definition, the user interface's lack of stereoscopic adaptation of the mouse pointer (see Sect. 5.5; [69,112]) likely contributed to some extent to longer task completion times under stereo display. However, this aspect should not be overrated as most participants coped well with this restriction. Maybe more importantly, this task involves judging the fit of a 3D primitive from all sides to precisely adjust it over a target object. Irrespective of the display mode, this necessitates changing the viewpoint relatively often (especially compared to the gripper-positioning task), in which case motion-based depth cues become available. A number of studies have suggested no or little merit of stereo display when motion-based depth cues were present [57,66–69]. The unclear stereoscopy advantage may thus be explained by a relatively strong presence of motion-based depth cues. A future in-depth analysis of viewpoint changes may help clarifying the feasibility of this account.

Navigation tasks generally involve less uncertainty in depth judgment as the robot moves along a visible 2D plane and control is reduced from six to three degrees of freedom. This could be a reason why effects of stereoscopy were not clear for this type of task. Also, obstacles were on all sides of the robot so adopting a wider, more zoomed-out perspective was a viable approach for accomplishing this task. We observed that users often chose a bird's eye perspective. Reduced perspective cues and reduced binocular disparity due to the farther viewpoint may thus be further reasons for less merit of stereo display. While we found a tendency for faster task completion, we did not find any indication for fewer collisions under stereoscopic presentation nor a correlation between task completion time and number of collisions. Therefore, in case the temporal advantage should hold in future studies, it may not come at the cost of more collisions. Compared to previous studies that found advantages of stereoscopic display in remote robot navigation based on video displays [49,60,61] or egocentric 3D visualizations [62], the point cloud display enhanced by global environment mapping [16] as provided in our user interface may suffice for effective and efficient task completion even without stereoscopic display.

The fact that we did not find any evidence for differences in perceived cognitive workload for any type of task is in agreement with other studies investigating workload under mono and stereo display [20,61]. The result could be considered surprising given that task performance tended to be better with stereo display and that it can generally be expected that when solving a task becomes easier workload will also be reduced. We cannot exclude a counter effect of the stereoscopic display technology we employed leading to increased cognitive load. However, this seems unlikely in view of the interview results and discomfort ratings, where we could not identify differences either. For us, the most important meaning of this result is that, despite the tendency of users to be slower under mono display, we found no evidence that accomplishing tasks would be more demanding.

6.2 Overall Evaluation of the User Interface and System

Technical performance can be regarded very satisfactory with all participants able to complete the tasks successfully without technical failures. We stabilized the system prior to the experiment in an extended period of feature freeze and bug fixing. Occasional inaccuracies in manipulation task execution can be attributed to sensor inaccuracies and cumulated errors of the multiple components involved. Arm trajectory planning should be improved to use as little space outside the camera's field of view as possible as this can be a safety risk in real-world applications, e.g., when people come close to the robot after the global 3D environment map was created.

The fact that the system worked reliably and all users were well able to accomplish the tasks suggests to us that the user interface might well be suited for resolving more difficult situations than the ones evaluated. For example, we achieved good results in informal tests with grasping objects from the top in highly cluttered scenes.

The results for perceived usability, user experience, and system usefulness show that participants highly approved the user interface and system (see Table 5). The pragmatic quality dimension of the AttrakDiff instrument [108] measures perceived usability, which can be considered high with average ratings of 5.5 out of 7 for both display types. AttrakDiff hedonic quality relates to stimulation and identification. This is the only measure with a significant difference between display types after multiplicity correction. Stereo display achieved higher mean scores (5.5 vs. 4.6). Higher hedonic quality ratings of stereo display were also found in other studies [20]. It should however be noted that hedonic ratings are highly influenced by novelty and surprise [19], so over time a harmonization with the mono scores may occur. The AttrakDiff attractiveness dimension relates to overall attractiveness of the user interface. Mean scores are fairly high for both display types.

Participants were further subjectively well able to build a mental spatial representation of the remote environment, as indicated by the SSM subscale of the MEC-SPQ questionnaire [110,111]. These results substantiate similar findings of a previous study on the same user interface, focusing on remote navigation only [16], which suggested that global 3D environment mapping plays a crucial role in this. The ratings of control modes are very high for the 3D mouse-based control modes used for navigation and gripper positioning. The control mode for shape definition with a standard mouse achieved lower scores, particularly in stereo mode, which probably reflects the previously discussed lack of mouse pointer adaptation. Lastly, participants found the whole system, including the robot and usage concept, highly useful, with mean scores of 6.1 out of 7.

Looking at average task completion times overall in a range of roughly three to five minutes, compared to a human who needs to travel to the house, there is a substantial temporal advantage. Compared to a human caregiver on site, task completion times are fairly long. Operators may get faster with more experience, as suggested by expert operators among the researchers who were substantially faster. From our experience, for elderly or handicapped people it is primarily relevant whether or not a task can be accomplished and not that it gets done quickly. However, a judgment on the adequateness of the length of task completion largely depends on the situation and urgency. It should also be considered that in our usage concept, teleoperation is an exceptional usage mode only relied upon when autonomous operation fails. Still, the completion times leave room for improvement. We discuss some ideas in Sect. 8.

7 Conclusion

We evaluated a novel type of semi-autonomous user interface for remotely resolving challenging situations of robots in domestic environments and focused on effects of stereoscopic display. Participants were well able to accomplish all tasks in monoscopic as well as stereoscopic display modes and highly approved the user interface's usability and the system's usefulness. Stereoscopic display clearly offered a temporal advantage for positioning the gripper. For the tasks of defining an object unknown to or unrecognized by the robot and remotely navigating the robot around obstacles, the tendency for faster task completion in stereo mode is notable but would need to be confirmed in further studies. As the use of stereoscopic displays is associated with drawbacks such as the need to wear glasses and to dim the light for ideal display, based on our results, we consider stereoscopy not essential but a useful optional display mode, at the very least for gripper positioning. It could be used in time-critical situations, for example in cases of emergency, or for solving particularly complex problems.

8 Future Work

In future studies it would be interesting to systematically investigate the relationship between the duration (or frequency) of viewpoint changes (e.g., as a function of task, training, or individual differences) and the utility of stereoscopic display. This may lead to better predictability of when stereoscopic display will be advantageous. For example, tasks entailing more viewpoint changes may benefit less from stereoscopic display, which might explain our different results for the different task types.

We consider several improvements to extend the application range of the user interface and improve task completion times. For defining the target object's shape, we have already implemented a video-based object pre-selection tool to speed up the process of placing the geometric primitive close to the target object. In place of the current control relying on a standard mouse, control with a 3D mouse might be more suitable. We also consider an alternative approach where the user defines the object's outer corners, based on which a shape is constructed. This would allow defining more complex shapes.

For specifying the gripper position faster, we have already implemented user-selectable presets for the gripper starting position close to the target object. Spatial constraints will be used in trajectory planning to improve safety by restricting the allowable space used by the arm. We also consider gripper orientation constraints for orientation-sensitive objects like a glass with liquid. To improve remote navigation, we consider more autonomous assistance in situations where the robot is stuck between objects. Also, work on collision avoidance based on 3D sensors for remote navigation is in progress.

Looking further into the future, coordinated simultaneous navigation and manipulation (e.g., to open a door) in semi-autonomous mode should be realized, as well as support for robots with two arms, and more detailed control over the grasping process to accomplish complex grasps.

Acknowledgments This research was supported by the European Commission, FP7, project "SRS", Grant Agreement No. 247772. We would like to thank Thiago de Freitas Oliveira Araújo, Ali Shuja Siddiqui, Markus Noack, Anne Reibke, Bianca Bannert, and Monika Heinzel-Gutenbrunner for supporting work.

References

1. Parasuraman R, Sheridan T, Wickens CA (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern A 30:286–297
2. Mast M, Burmester M, Krüger K, Fatikow S, Arbeiter G, Graf B et al (2012) User-centered design of a dynamic-autonomy remote interaction concept for manipulation-capable robots to assist elderly people in the home. J Hum Robot Interact 1:96–118
3. Martens C, Prenzel O, Gräser A (2007) The rehabilitation robots FRIEND-I & II: daily life independency through semi-autonomous task-execution. In: Kummo SS (ed) Rehabilitation robotics. Itech, Vienna, pp 137–162
4. Durand B, Godary-Dejean K, Lapierre L et al. (2010) Fault tolerance enhancement using autonomy adaptation for autonomous mobile robots. In: Proc Conf Control Fault Toler Syst, pp 24–29
5. Qiu R, Ji Z, Noyvirt A et al. (2012) Towards robust personal assistant robots: experience gained in the SRS project. In: Proc IEEE/RSJ Int Conf Intell Robot Syst (IROS), pp 1651–1657
6. Doroodgar B, Ficocelli M, Mobedi B, Nejat G (2010) The search for survivors: cooperative human-robot interaction in search and rescue environments using semi-autonomous robots. In: Proc IEEE Int Conf Robot Autom (ICRA), pp 2858–2863
7. Shiomi M, Sakamoto D, Kanda T et al (2011) Field trial of a networked robot at a train station. Int J Soc Robot 3:27–40
8. Mason M, Lopes M (2011) Robot self-initiative and personalization by learning through repeated interactions. In: Proc Int Conf Hum Robot Interact (HRI), pp 433–440
9. Campbell CL, Peters RA, Bodenheimer RE, Bluethmann WJ, Huber E, Ambrose RO (2006) Superpositioning of behaviors learned through teleoperation. IEEE Trans Robot 22:79–91
10. Jenkins OC, Peters RA, Bodenheimer RE (2006) Uncovering success in manipulation. In: Proc RSS workshop manipulation for human environments, Philadelphia
11. Kemp CC, Edsinger A, Torres-Jara E (2007) Challenges for robot manipulation in human environments. IEEE Robot Autom Mag 14:20–29
12. ISO 9241–11 (1998) Ergonomic requirements for office work with visual display terminals. Part 11: Guidance on usability
13. Drury JL, Scholtz J, Yanco HA (2003) Awareness in human-robot interactions. In: Proc IEEE Int Conf Syst Man Cybern, pp 912–918
14. Yanco HA, Drury J (2004) "Where am I" Acquiring situation awareness using a remote robot platform. In: Proc IEEE Int Conf Syst Man Cybern, pp 2835–2840
15. Steinfeld A, Fong T, Kaber D et al. (2006) Common metrics for human-robot interaction. In: Proc Int Conf Hum Robot Interact (HRI), pp 33–40
16. Mast M, Španěl M, Arbeiter G et al (2013) Teleoperation of domestic service robots: effects of global 3D environment maps in the user interface on operators' cognitive and performance metrics. In: Hermann G et al (eds) Int Conf Soc Robot (ICSR) LNAI. Springer, Cham, pp 392–401
17. Kaber D, Onal E, Endsley M (2000) Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. Hum Factors Ergon Manuf 10:409–430

18. ISO 9241–210 (2010) Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems
19. Hassenzahl M (2001) The effect of perceived hedonic quality on product appealingness. Int J Hum Comput Interact 13:481–499
20. Broy N, André E, Schmidt A (2012) Is stereoscopic 3D a better choice for information representation in the car? In: Proc Automot UI, pp 93–100
21. Chen JYC, Haas EC, Barnes MJ (2007) Human performance issues and user interface design for teleoperated robots. IEEE Trans Syst Man Cybern C 37:1231–1245
22. Nielsen CW, Goodrich MA, Ricks RW (2007) Ecological interfaces for improving mobile robot teleoperation. IEEE Trans Robot 23:927–941
23. Bruemmer DJ, Few DA, Boring RL, Marble JL et al (2005) Shared understanding for collaborative control. IEEE Trans Syst Man Cybern A 35:494–504
24. Keyes B, Casey R, Yanco HA et al (2006) Camera placement and multi-camera fusion for remote robot operation. In: Proc IEEE Int Workshop Safety, Security, Rescue robot
25. Fiala M (2005) Pano-presence for teleoperation. In: Proc IEEE/RSJ Int Conf Intell Robot Syst (IROS), pp 3798–3802
26. Labonté D, Boissy P, Michaud F (2010) Comparative analysis of 3-D robot teleoperation interfaces with novice users. IEEE Trans Syst Man Cybern B 40:1331–1342
27. Michaud F, Boissy P, Labonté D et al (2010) Exploratory design and evaluation of a homecare teleassistive mobile robotic system. Mechatronics 20:751–766
28. Das H, Sheridan TB, Slotine J (1989) Kinematic control and visual display of redundant teleoperators. In: Proc IEEE Int Conf Syst Man Cybern, pp 1072–1077
29. Chintamani K, Cao A, Ellis RD, Pandya AK (2010) Improved telemanipulator navigation during display-control misalignments using augmented reality cues. IEEE Trans Syst Man Cybern A 40:29–39
30. Pongrac H, Peer A, Färber B, Buss M (2008) Effects of varied human movement control on task performance and feeling of telepresence. In: Ferre M (ed) Haptics: perception, devices and scenarios. EuroHaptics 2008. LNCS, vol 5024. Springer, New York, pp 755–765
31. Sheridan TB, Ferrell WR (1963) Remote manipulative control with transmission delay. IEEE Trans Hum Fact Electron 4:25–29
32. Stassen HG, Smets GJF (1997) Telemanipulation and telepresence. Control Eng Practice 5:363–374
33. Buss M, Peer A, Schauß T, Stefanov N, Unterhinninghofen U (2010) Development of a multi-modal multi-user telepresence and teleaction system. Int J Robot Res 29:1298–1316
34. Honda M, Miyoshi T, Imamura T et al (2011) Tele-operation between USA and Japan using humanoid robot hand/arm. In: Proc Int Conf Hum Robot Interact (HRI), pp 151–152
35. Mavridis N, Giakoumidis N (2012) A novel evaluation framework for teleoperation and a case study on natural human-arm-imitation through motion capture. Int J Soc Robot 4:5–18
36. Leeper A, Hsiao K, Ciocarlie M, Takayama L, Gossow D (2012) Strategies for human-in-the-loop robotic grasping. In: Proc Int Conf Hum Robot Interact (HRI), pp 1–8
37. You E, Hauser K (2012) Assisted teleoperation strategies for aggressively controlling a robot arm with 2D input. In: Durrant-Whyte H et al (eds) Robotics: science and systems, vol VII. MIT Press, USA, pp 354–361
38. Griffin WB, Provancher WR, Cutkosky MR (2005) Feedback strategies for telemanipulation with shared control of object handling forces. Presence 14:720–731
39. Nieto J, Slawiñski E, Mut V, Wagner B (2011) Toward safe and stable time-delayed mobile robot teleoperation through sampling-based path planning. Robotica 30:351–361
40. Bradshaw JM, Feltovich PJ, Jung H et al (2004) Dimensions of adjustable autonomy and mixed-initiative interaction. In: Nickles N et al (eds) Agents and Computational autonomy, AUTONOMY 2003. LNCS. Springer, New York, pp 17–39
41. Schermerhorn P, Scheutz M (2009) Dynamic robot autonomy: investigating the effects of robot decision-making in a human-robot team task. In: Proc ICMI-MLMI, pp 63–70
42. Brookshire J, Singh S, Simmons R (2004) Preliminary results in sliding autonomy for assembly by coordinated teams. In: Proc IEEE/RSJ Int Conf Intell Robot Syst (IROS), pp 706–711
43. Prats M, Fernández JJ, Sanz PJ (2012) An approach for semi-autonomous recovery of unknown objects in underwater environments. In: Proc OPTIM, pp 1452–1457
44. Atherthon JA, Goodrich MA (2009) Supporting remote manipulation with an ecological augmented virtuality interface. In: Proc AISB, pp 16–23
45. Arbeiter G, Bormann R, Fischer J et al (2012) Towards geometric mapping for semi-autonomous robots. In: Stachniss C et al (eds) Spatial Cognition, LNAI, vol 7463. Springer, New York, pp 114–127
46. Hornung A, Wurm KM, Bennewitz M et al (2013) OctoMap: an efficient probabilistic 3D mapping framework based on octrees. Auton Robot 34:189–206
47. Mast M, Burmester M, Graf B et al. (in press) Design of the human-robot interaction for a semi-autonomous service robot to assist elderly people. In: Wichert R, Klausing H (eds) Advanced technologies and societal change: ambient assisted living. 7. AAL-Kongress. Springer, New York
48. Liu A, Tharp G, French L, Lai S, Stark L (1993) Some of what one needs to know about using head-mounted displays to improve teleoperator performance. IEEE Trans Robot Autom 5:638–648
49. Lee S, Kim GJ (2008) Effects of haptic feedback, stereoscopy, and image resolution on performance and presence in remote navigation. Int J Hum Comput Stud 66:701–717
50. Edmondson R, Light K, Bodenhamer A et al. (2012) Enhanced operator perception through 3D vision and haptic feedback. In: Proc SPIE, vol 8387
51. Howard IP, Rogers BJ (2012) Perceiving in depth, vol 2. Oxford University Press, New York
52. Howard IP (2012) Perceiving in depth, vol 3. Oxford University Press, New York
53. Arsenault R, Ware C (2004) The importance of stereo and eye-coupled perspective for eye-hand coordination in fish tank VR. Presence 13:549–559
54. Hubona GS, Shirah GW, Jennings DK (2004) The effects of cast shadows and stereopsis on performing computer-generated spatial tasks. IEEE Trans Syst Man Cybern A 34:483–493
55. Sollenberger RL, Milgram P (1993) Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. Hum Factors 35:483–499
56. Ware C, Franck G (1996) Evaluating stereo and motion cues for visualizing information nets in three dimensions. ACM Trans Gr 15:121–140
57. Barfield W, Hendrix C, Bystrom K (1999) Effects of stereopsis and head tracking on performance using desktop virtual environment displays. Presence 8:237–240
58. Drascic D (1991) Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing. In: Proc Hum Fact Ergon Soc Annu Meet, pp 1367–1371
59. Hutto CJ, Vincenzi DA, Hall S, Gangadharan S (2004) The effects of viewing medium on depth perception in human performance of a telerobotics manipulation task. In: Vincenzi DA et al (eds) Proc HPSAA II. Lawrence Erlbaum, Mahwah, pp 112–117

60. Livatino S, Muscato G, Sessa S et al (2008) Mobile robotic teleguide based on video images. IEEE Robot Autom Mag 15(4):58–67
61. Chen JYC, Oden RVN, Merritt JO (2014) Utility of stereoscopic displays for indirect-vision driving and robot teleoperation. Ergonomics 57:12–22
62. Livatino S, Muscato G, Sessa S, Neri V (2010) Depth-enhanced mobile robot teleguide based on laser images. Mechatronics 20:739–750
63. Jameson D, Hurvich LM (1959) Note on the factors influencing the relation between stereoscopic acuity and observation distance. J Opt Soc Am 49:639
64. Mather G, Smith DRR (2004) Combining depth cues: effects upon accuracy and speed of performance in a depth-ordering task. Vis Res 44:557–562
65. Kim WS, Tendick F, Stark LW (1987) Visual enhancements in pick-and-place tasks: human operators controlling a simulated cylindrical manipulator. IEEE J Robot Autom 3:418–425
66. Dosher BA, Sterling G, Wurst SA (1986) Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. Vis Res 26:973–990
67. Ware C, Arthur K, Booth KS (1993) Fish tank virtual reality. In: Proc Comput Hum Interact (CHI), pp 37–42
68. Naepflin U, Menozzi M (2001) Can movement parallax compensate lacking stereopsis in spatial explorative search tasks? Displays 22:157–164
69. van Schooten BW, van Dijk EMAG, Zudilova-Seinstra E et al (2010) The effect of stereoscopy and motion cues on 3D interpretation task performance. In: Proc AVI, pp 167–170
70. Hubona GS, Shirah GW, Fout DG (1997) The effects of motion and stereopsis on three-dimensional visualization. Int J Hum Comput Stud 47:609–627
71. Bradshaw MF, Parton AD, Glennerster A (2000) The task-dependent use of binocular disparity and motion parallax information. Vis Res 40:3725–3734
72. Schor CM, Wood I (1983) Disparity range for local stereopsis as a function of luminance spatial frequency. Vis Res 23:1649–1654
73. Siderov J, Harwerth RS (1995) Stereopsis, spatial frequency, and retinal eccentricity. Vis Res 16:2329–2337
74. Lloyd CJ (2012) Effects of spatial resolution and antialiasing on stereoacuity and comfort. In: Proc AIAA Model Simul Conf, pp 1–13
75. Jää-Aro K, Kjelldahl L (1997) Effects of image resolution on depth perception in stereo and non-stereo images. Proc SPIE 3012:319–326
76. Bradshaw MF, Glennerster A (2006) Stereoscopic acuity and observation distance. Spat Vis 19:21–36
77. Surdick RT, Davis ET, King RA, Hodges LF (1997) The perception of distance in simulated visual displays: a comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. Presence 6:513–531
78. Kim WS, Ellis SR, Tyler ME, Hannaford B, Stark LW (1987) Quantitative evaluation of perspective and stereoscopic displays in three-axis manual tracking tasks. IEEE Trans Syst Man Cybern 17:61–72
79. Fujiwara T, Kamegawa T, Gofuku A (2011) Stereoscopic presentation of 3D scan data obtained by mobile robot. In: Proc IEEE Int Symp Saf Secur Rescue Robot, pp 178–183
80. Lambooij M, Ijsselsteijn W, Fortuin M, Heynderickx I (2009) Visual discomfort and visual fatigue of stereoscopic displays: a review. J Imaging Sci Technol 53:030201-1–030201-14
81. Camposeco F, Avilés C, Careaga B et al. (2011) Constraints on human stereo vision for tele-operation. In: Proc LARC, pp 1–6
82. Reiser U, Connette C, Fischer J et al. (2009) Care-O-bot 3 - creating a product vision for service robot applications by integrating design and technology. In: Proc IEEE/RSJ Int Conf Intell Robot Syst (IROS), pp 1992–1998
83. ROS documentation. http://www.ros.org/wiki/. Accessed 11 Nov 2014
84. Fischer J, Arbeiter G, Bormann R, Verl A (2012) A framework for object training and 6DoF pose estimation. In: Proc ROBOTIK, pp 513–518
85. Kunz T, Reiser U, Stilman M, Verl A (2010) Real-time path planning for a robot arm in changing environments. In: Proc IEEE/RSJ Int Conf Intell Robot Syst (IROS), pp 5906–5911
86. ROS documentation: arm navigation. http://wiki.ros.org/arm_navigation. Accessed 11 Nov 2014
87. SRS: multi-role shadow robotic system for independent living. srs-project.eu. Accessed 11 Nov 2014
88. Mast M, Burmester M, Berner E et al. (2010) Semi-autonomous teleoperated learning in-home service robots for elderly care: a qualitative study on needs and perceptions of elderly people, family caregivers, and professional caregivers. In: Proc Int Conf Robot Mechatron, pp 1–6
89. Sharkey A, Sharkey N (2012) Granny and the robots: ethical issues in robot care for the elderly. Ethics Inf Technol 14:27–40
90. Dumas JC, Fox JE (2008) Usability testing: current practice and future directions. In: Jacko JA, Sears A (eds) The handbook of human-computer interaction, 2nd edn. Erlbaum, Mahwah, pp 1129–1149
91. Koenig N, Howard A (2004) Design and use paradigms for gazebo, an open-source multi-robot simulator. In: Proc IEEE/RSJ Int Conf Intell Robot Syst (IROS), pp 2149–2154
92. ISO 9241–110 (2006) Ergonomics of human-system interaction. Part 110: dialogue principles
93. Goodrich MA, Olsen DR Jr (2003) Seven principles of efficient human robot interaction. In: Proc IEEE Int Conf Syst Man Cybern, pp 3942–3948
94. Steinfeld A (2004) Interface lessons for fully and semi-autonomous mobile robots. In: Proc IEEE Int Conf Robot Autom (ICRA), pp 2752–2757
95. Gibson JJ (1979) The ecological approach to visual perception. Houghton Mifflin, Boston
96. Norman D (1988) The design of everyday things. Basic Books, New York
97. ROS documentation: RViz, http://wiki.ros.org/rviz. Accessed 11 Nov 2014
98. Fong T, Thorpe C (2001) Vehicle teleoperation interfaces. Auton Robot 11:9–18
99. Yanco HA, Drury JL, Scholtz J (2004) Beyond usability evaluation: analysis of human-robot interaction at a major robotics competition. Hum Comput Interact 19:117–149
100. Voshell M, Woods DD, Phillips F (2005) Overcoming the keyhole in human-robot coordination: simulation and evaluation. In: Proc Hum Fact Ergon Soc Annu Meet, pp 442–446
101. Bade R, Ritter F, Preim B (2005) Usability comparison of mouse-based interaction techniques for predictable 3d rotation. In: Butz A et al (eds) Smart graphics, SG 2005, LNCS. Springer, New York, pp 138–150
102. 3Dconnexion SpaceNavigator. http://www.3dconnexion.com/products/spacenavigator.html. Accessed 11 Nov 2014
103. Franken M, Stramigioli S, Misra S et al (2011) Bilateral telemanipulation with time delays: a two-layer approach combining passivity and transparency. IEEE Trans Robot 27:741–756
104. Ogre3D. http://www.ogre3d.org. Accessed 11 Nov 2014
105. Jones CM, Healy SD (2006) Differences in cue use and spatial memory in men and women. Proc Royal Soc B 273:2241–2247
106. Peters M, Laeng B, Latham K et al (1995) A redrawn Vandenberg and Kuse mental rotations test: different versions and factors that affect performance. Brain Cogn 28:39–58

107. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Human mental workload. North Holland, Amsterdam, pp 139–183
108. Hassenzahl M, Burmester M, Koller F (2003) AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Szwillus G, Ziegler J (eds) Mensch und Computer. Teubner, Stuttgart, pp 187–196
109. Hassenzahl M, Monk A (2010) The inference of perceived usability from beauty. Hum Comput Interact 25:235–260
110. Wirth W, Hartmann T, Böcking S et al (2007) A process model of the formation of spatial presence experiences. Media Psychol 9:493–525
111. Wirth W, Schramm H, Böcking S et al (2008) Entwicklung und Validierung eines Fragebogens zur Entstehung von räumlichem Präsenzerleben. In: Matthes J et al (eds) Die Brücke zwischen Theorie und Empirie: Operationalisierung. Messung und Validierung in der Kommunikationswissenschaft, Halem Verlag, Köln, pp 70–95
112. Hill A, Johnson A (2008) Withindows: a framework for transitional desktop and immersive user interfaces. In: Proc IEEE Symp 3D User Interfaces, pp 3–10

Marcus Mast has focused his research in the field of human-robot interaction on user experience and interaction design for semi-autonomous service robots. In the wider context of human-machine interaction, he has worked on such issues as developing user experience evaluation methods and investigating spatiotemporal patterns in users' visual attention. He received his MSc in Information Management from Stuttgart Media University, Germany, and his PhD in Cognitive Science from Linköping University, Sweden, in cooperation with Stuttgart Media University.

Zdeněk Materna received his Bachelor's degree in Computer Systems from the College of Polytechnics Jihlava, and his Master's degree in Cybernetics, Control, and Measurement from Brno University of Technology. Currently he is a PhD student at Brno University of Technology. His research is focused on human-robot interaction, teleoperation, and telemanipulation.

Michal Španěl is an Assistant Professor at the Department of Computer Graphics and Multimedia at Brno University of Technology, Czech Republic. He received his MSc degree in Computer Science and Engineering in 2003 and his PhD degree in Information Technology in 2011, both from Brno University of Technology. His current research interests include the development of machine vision algorithms for dynamic 3D environment perception for robotics, human-robot interaction, and 3D environment visualization techniques for teleoperated robots.

Florian Weisshardt completed his studies of Mechatronics and Automation with focus on robot systems at the University of Stuttgart in 2009. Since then he has been working as a research associate and project manager at the Service Robotics Group at Fraunhofer IPA. He leads several nationally and internationally funded projects and coordinates the open source software development for the Care-O-bot service robot and ROS-Industrial. He is working on a PhD in the field of system engineering and software development for service robots.

Georg Arbeiter received his degree in Mechatronics from University of Erlangen. He received his PhD at Fraunhofer Institute for Manufacturing Engineering and Automation in the field of 3D perception for mobile robots. In his current position he works as an algorithm research engineer with Continental AG.

Michael Burmester is a Professor for Ergonomics and Usability at Stuttgart Media University in Germany. He further is the Vice Dean for Research, the Representative of the Information Experience and Design Research Group, and the Director of the User Experience Research Lab at Stuttgart Media University. After studying psychology in Regensburg he completed his PhD in psychology on user interfaces for elderly people at the University of Wuppertal. He worked as a researcher at the Fraunhofer Institute for Industrial Engineering and at Siemens Corporate Technology, and was manager of the usability engineering department of a user interface design consultancy. His current research interests include the development of technology concepts inducing positive usage experiences as well as the development of methods and design strategies for a positive user experience.

Pavel Smrž is an Associate Professor and a project leader at the Faculty of Information Technology at Brno University of Technology, Czech Republic. His research interests include advanced autonomous systems, human-machine interaction, knowledge extraction, multimedia semantic processing, and machine learning. He has served as a principal investigator in various European as well as national projects and authored more than 70 papers in scientific journals and conference proceedings.

Birgit Graf is Manager of the Domestic and Personal Robots Group at Fraunhofer IPA and has been working in mobile robotics for more than 15 years. The technological focus of her group is on environment modeling, object detection, and adaptive motion planning for complex mobile service robots with manipulators. She received her degree in Computer Science from Stuttgart University in 1999 and completed her PhD on mobile robot navigation in 2008. Birgit Graf was involved in the development of several generations of the robotic home assistant Care-O-bot and other robotic solutions to support elderly people, people in need of care, and caregivers. She also has been leading several service robot developments for industrial clients.