Bridging Distance with a Collaborative Telepresence Robot for Older Adults – Report on Progress in the CO-HUMANICS Project

Söhnke Benedikt Fischedick, Kay Richter, Tim Wengefeld, Daniel Seichter, Andrea Scheidig,
Nicola Döring, Wolfgang Broll, Stephan Werner, Alexander Raake, and Horst-Michael Gross
Ilmenau University of Technology, 98684 Ilmenau, Germany

Abstract

In an aging society, the social needs of older adults, such as regular interactions and independent living, are crucial for their quality of life. However, due to spatial separation from their family and friends, it is difficult to maintain social relationships. Our multidisciplinary project, CO-HUMANICS, aims to meet these needs, even over long distances, through the utilization of innovative technologies, including a robot-based system. This paper presents the first prototype of our system, designed to connect family members or friends virtually present through a mobile robot with an older adult. The system incorporates bi-directional video telephony, remote control capabilities, and enhanced visualization methods. A comparison is made with other state-of-the-art robotic approaches, focusing on remote control capabilities. We provide details about the hard- and software components, e.g., a projector-based pointing unit for collaborative telepresence to assist in everyday tasks. Our comprehensive scene representation is discussed, which utilizes 3D NDT maps, enabling advanced remote navigation features, such as autonomously driving to a specific object. Finally, insights about past and concepts for future evaluation are provided to assess the developed system.

I. Introduction

The ability to take part in regular social interactions is a fundamental aspect of independent living for older adults. However, achieving this can be challenging due to factors such as geographic distance from family and friends or limited mobility. To address these issues, the CO-HUMANICS project (Co-Presence of Humans and Interactive Companions for Seniors) aims to develop innovative approaches to enhance co-presence. Co-presence refers to the concept that individuals have a sense of perceiving others and being actively perceived by them [13]. The project's primary focus lies in utilizing augmented and mixed reality solutions, specifically through head-mounted displays and smart glasses, and robot-based systems as alternative technologies for co-presence. In this paper, we focus on the robot companion, which will be located at the apartment of an older adult (local host) and can be accessed and controlled remotely by family and friends (virtual guests). Although co-presence systems for robotics have been extensively studied for applications such as care for older adults, there remains a lack of systems that facilitate natural social interaction. Most existing systems focus on video-based co-presence with limited robot control features. To address this, we explore how a robot-based system can enhance co-presence while emphasizing natural interaction. Our goal is to develop an easy-to-use and reliable system that not only

This work has received funding from Carl-Zeiss-Stiftung to the project *Co-Presence of Humans and Interactive Companions for Seniors* (CO-HUMANICS) running 6/21-5/26. http://www.co-humanics.de

Fig. 1. The CO-HUMANICS robot companion enables a virtual guest to remotely assist an older adult in a repair task by highlighting a pipe in the video call with the projector.

enables remote control of the robot but also offers the ability to highlight objects in the apartment, providing a more immersive and collaborative experience. This functionality can be helpful for various scenarios, such as assisting an older adult in repairing objects as shown in Fig. 1, browsing through a photo album of past shared memories or participating in other collaborative tasks. To accomplish these scenarios, the robot needs to have basic skills, such as localization and navigating in the environment, tracking and following the local host to maintain focus during video calls, and even adapting its behavior to suit their preferences and needs.

In this paper, we summarize related work in the field of robotic systems with co-presence features and give a report on the progress in the CO-HUMANICS project. We present the first prototype of our robot system developed for the CO-HUMANICS project, focusing on the hardware and software specifications and the vision of features. Finally, we will introduce a concept for the planed evaluation and discuss potential future work in this area.

II. Related Work

One way to establish co-presence is by deploying telepresence technologies, which refer to a set of technologies that enable individuals to feel present in distant locations [14]. This can be achieved by deploying a robot at a remote location that senses the environment and can be controlled remotely. Our project not only aims to make the virtual guest feel present with the local host, but also to make the local host feel present with the virtual guest, to enable collaborative activities. However, while a wide range of recent robotic projects have video call capabilities [1]–[12] (see Tab. I), just a few provide advanced features for telepresence, such as remote navigation. By contrast, their main focus is to deploy services that are autonomously provided by the robot itself. This includes high-level applications, such as object transportation, the delivery of daily reminders, and user-centered entertainment services. These applications require basic robotic skills, such as autonomous navigation, person perception and recognition, as well as human-robot-interaction (HRI) skills, such as speech and gesture recognition. In the following, we give a brief overview of recent real-world robotic projects with advanced telepresence features, i.e., video calls with the opportunity to remotely control the robot. SERROGA [1] is a companion robot that can be controlled remotely via an Android device. Besides simple drive commands, the virtual guest is also able to click in the video to let the robot drive to selected positions. The MoveCare project [8], [9] provides video call and remote-control features to a caregiver in case of a detected emergency. However, no specific description of the provided remote-control capabilities is given in the publications. The MORPHIA project [12] focuses on a social network between older adults and their relatives, realizing telepresence with up to three virtual guests in parallel. Simple drive commands, i.e., drive forward or rotate left/right, can be used to control the robot. To facilitate controlling, the camera's angle can be altered with a simple tilt mechanism. Additionally, the virtual guest is able to activate person follow or search behaviors if the local host changes positions during a conversation. Moreover, predefined positions can be selected to let the robot drive autonomously to these locations in the apartment. Similar behaviors for person search and follow as well as driving to predefined positions are implemented for the SAM robot [11]. Alternatively, the virtual guest can control the robot by a gamepad.

Besides the aforementioned projects that evaluate in real-world application, some other work has been done to improve telepresence features in experimental setups. This is done by augmenting the data presented to the virtual guest. An enhanced representation of laser data is used in [15] to augment a camera image with additional information about obstacles in the surroundings of the robot. In [16], [17], a rendered 3D scene from laser and RGB-D data is presented to the virtual guest. Similarly, [18] renders the environment from an artificially generated 3D model instead of sensor data. [19] examines different remote-control visualizations, including a camera view, a top-down view of the laser sensor data and a top-down view of a simplified map with the current localization of the robot. While these works target the data

TABLE I

Implemented robotic skills in comparison to the CO-HUMANICS project. Classification to the best of our knowledge: ✓ – skill is used during user trial; (✓) – skill is mentioned in the system architecture but was deactivated during user trial; ✗ – skill is not mentioned and therefore assumed to be not available.

	[1] **SERROGA**	[2] **DOMEO**	[3] **SYMPARTNER**	[4], [5] **SocialRobot**	[6] **MobiKa**	[7] **Hobbit**	[8], [9] **MoveCare**	[10] **ENRICHME**	[11] **SAM**	[12] **MORPHIA**	**CO-HUMANICS**
Autonomous Navigation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Telepresence Navigation	✓	✗	(✓)	✗	✗	✗	✓	✗	✓	✓	✓
Person Detect.	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Person Re-Identification	✗	✗	(✓)	✓	✗	✗	✓	✓	✗	✗	✗
Following	✓	✗	(✓)	✗	✗	✓	✗	✗	✓	✓	✓
Transportation	✗	✓	✗	✗	✓	✓	✓	✗	✗	✓	✗
Object Interaction	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Speech Recognition	✗	✓	✗	✓	✗	✓	✓	✗	✗	✗	✗
Gesture Recognition	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Emotion Recognition	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗
Video Calling	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Entertainment	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗
Reminders	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗

presentation aspect, the implementation of the control part is rather simple, mostly utilizing real [16] or virtual [18], [19] joystick or gamepads [17] for steering commands.

In conclusion, as shown in Tab. I, only a fraction of the approaches deployed to real-world applications support telepresence navigation. If supported, the realized telepresence system is rather simple without any enhanced visualization of the surroundings presented to the virtual guest. Moreover, the systems often lack suitable actuators, and thus the ability to interact with the surroundings that is crucial for enabling collaborative activities.

For the CO-HUMANICS robot companion, we target collaborative activities between the virtual guest and the local host by utilizing telepresence technologies. We refer to this as collaborative telepresence. This includes advanced methods for data visualization, robot control, and the support of highlighting objects in the surroundings of the robot with a projector. Additionally, we aim to incorporate advanced methods for scene analysis [20]–[22] to get a better understanding of the surroundings in order to simplify and improve usability of the robot control. The overall goal is to enable even inexperienced users a better collaborative telepresence experience with focus on social aspects.

III. Robot Platform

The CO-HUMANICS robot companion is shown in Fig. 2. It is based on the TORY platform of the German robotics company MetraLabs. The TORY platform provides safety, stability, and adaptability. This design offers several advantages as it simplifies the navigation in confined spaces, enables a rotation without translation of the platform, and could fit the results from the requirement analysis [23]. It features a circular footprint with a diameter of $50\,\mathrm{cm}$ and rotation symmetry by placing the differential drive on the middle axis. The default setup of the platform features a 2D Sick S300 laser range scanner, with an opening angle of $270°$, enabling efficient mapping and obstacle detection. It is mounted $10\,\mathrm{cm}$ above the ground plane and automatically limits the speed of the robot

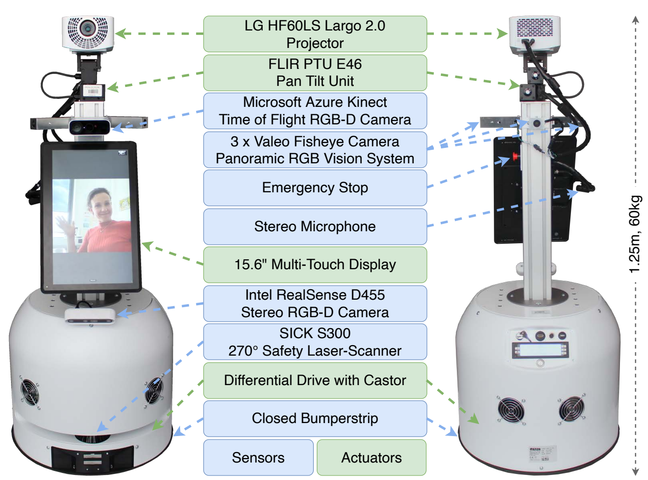

Fig. 2. The CO-HUMANICS robot companion with its sensors (blue) and actuators (green).

platform to $0.3\,\mathrm{m/s}$ when obstacles are nearby as a safety feature. An additional bumper ring located at the bottom of the robot automatically stops its movement when triggered, further ensuring safety during operation. The $40\,\mathrm{Ah}$ LFMP battery enables an operation time of up to five hours, depending on the used features.

Several customizations to the default setup have been made to account for the specific needs in the CO-HUMANICS project. An Intel RealSense D455 RGB-D stereo camera is used to monitor the floor in front of the robot to detect obstacles outside the laser plane. To enable mounting further sensors and actuators easily, a mounting profile is attached to the top of the robot, leading to a total height of $1.25\,\mathrm{m}$. A Microsoft Azure Kinect, based on time-of-flight (ToF) technology, is used as a second RGB-D camera to observe the local host and to obtain 3D information about the environment. The local host can interact with the robot from the same side through a 15.6" multi-touch display. To detect and track the local host as well as the environment in all directions, three RGB Valeo fisheye cameras with a horizontal opening angle of $195°$ are mounted to the mounting profile with an $120°$ offset. For enhancing the virtual guest's ability to communicate and enabling a collaborative telepresence with a local host, a top-mounted LG HF60LS Largo 2.0 projector is connected to a FLIR PTU E46 pan tilt unit (PTU) on the robot. With this feature, the virtual guest can highlight objects in the environment, allowing for a collaborative telepresence.

To minimize latency and maintain stable connections, all required computing capabilities are held onboard. An Intel NUC featuring an Intel i7 1165G7 processor, 64 GB RAM, and an NVIDIA GeForce RTX 2060 GPU is the primary onboard computer for running the robot's software and processing sensor data. To offload some of the GPU intensive computations, an NVIDIA Jetson AGX Orin 32GB is also included in the robot. This dual-computer configuration ensures that the robot can efficiently process a variety without offloading tasks to an external server. This approach ensures that most data remains within the robot itself, addressing data privacy concerns [23], while enabling effective task processing and maintaining a responsive user experience.

In the future, we plan to use a microphone array to provide directional recording of the local host as well as noise cancellation and dereverberation for an improved audio quality during video calls.

IV. Software Architecture

In this section, we describe the software architecture of the CO-HUMANICS robot companion. An overview of the software architecture is shown in Fig 3. Since the robot companion is still in development, the depicted architecture includes both already existing modules and some modules that are still under development. Our architecture is based on a modular layered approach first presented in [24], where each layer builds on top of the modules of the layer below. Next, we give a brief overview over the modules in each layer.

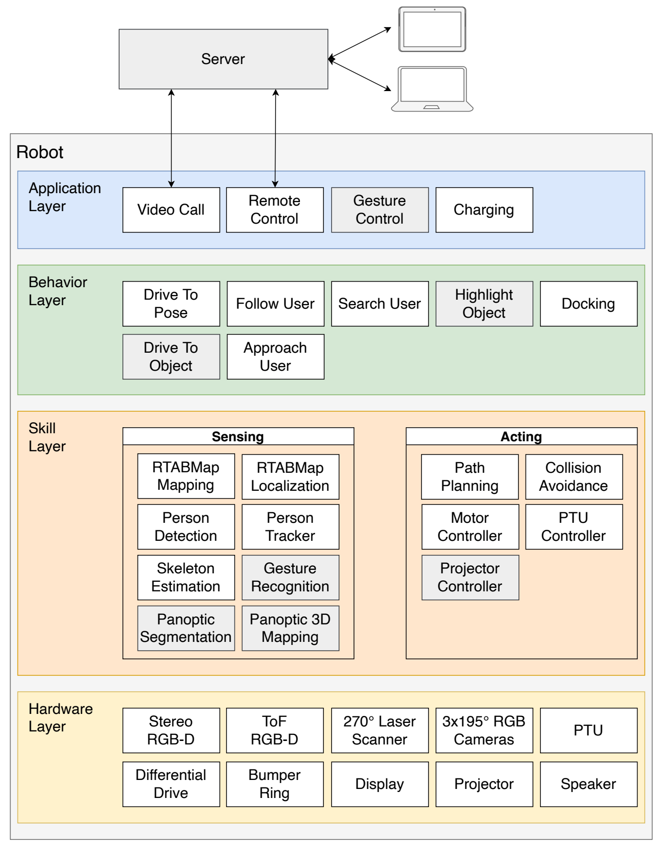

Fig. 3. Software architecture for the CO-HUMANICS robot companion. Modules in gray boxes are in development and not yet fully integrated.

A. Skill Layer

The skill layer makes use of the hardware and implements basics skills for processing sensor readings. The goal of the modules in the skill layer is to extract all the information required for the subsequent behavior layer.

For creating a 2D occupancy grid map and for localizing within this map, readings from the Azure Kinect RGB-D camera, the wheel odometry, and the laser scanner are processed by the SLAM approach RTABMap [25]. The resulting map is used for global path planning by utilizing the A* or E* algorithm. Readings from the RealSense RGB-D camera are used to build a local 3D occupancy normal distribution transform (NDT) map [26], which is then projected to the ground plane in order to get a local 2D occupancy grid map. Together with data of the laser scanner, the local planner, i.e., a dynamic window approach, can avoid dynamic obstacles.

To analyze the environment even more thoroughly, semantic 3D NDT maps [27] are built from readings from the Azure Kinect and the predicted semantic segmentation of ESANet [20]. This skill is currently extended to accomplish 3D panoptic mapping [28] by utilizing the semantic and instance predictions of EMSANet [21] or EMSAFormer [22]. To provide even more information to the subsequent behavior layer, the remaining outputs of EMSANet or EMSAFormer for instance orientation estimation as well as scene classification are additionally incorporated in the panoptic NDT maps, realizing a comprehensive scene understanding. This allows for even more complex drive-to behaviors like "drive to the chair in the living room while respecting its orientation", as described later in Sec. IV-B.

In order to achieve robust person detection across all RGB cameras, OpenPifPaf [29] and YOLOv3 [30] are used in parallel, reducing false negatives. For extracting 3D skeleton hypothesis, the Azure Kinect SDK is used as an additional detector. All these detections are used as inputs to our multi-modal person tracker [31]. Furthermore, we plan to use the skeletons for gesture recognition, as it would offer another option to interact with the robot.

Apart from that, the skill layer also includes more modules for acting, such as simple controllers for the motor of the differential drive or the PTU. Additionally, a skill is currently in development to control the projector, for switching it on and off as well as showing any image or video stream. The example in Fig. 4 shows how the combination of the projector and the PTU controller can be used to highlight objects by simply showing a green dot. As we plan to integrate a microphone array into the robot companion in the future, there is potential for integrating additional skills to extract room acoustic information, leveraging the insights from [32], [33], such as performing a direction-dependent analysis of the room acoustics. For this, approaches of autonomous and mobile room acoustics measurement robots can be adapted and extended [34], [35]. This integration could not only further enhance audio quality during video calls but also offer valuable insights into acoustic properties, which can be integrated into the existing maps. Furthermore, this integration has the potential to serve as an additional method for localizing the robot within the environment [33].

B. Behavior Layer

In the behavior layer, various skills are combined to create more complex behaviors. For example, the drive-to-pose behavior combines the path planner, the motor controller, and the collision avoidance skills to safely navigate the robot to a given pose. While this behavior enables driving to goals based on

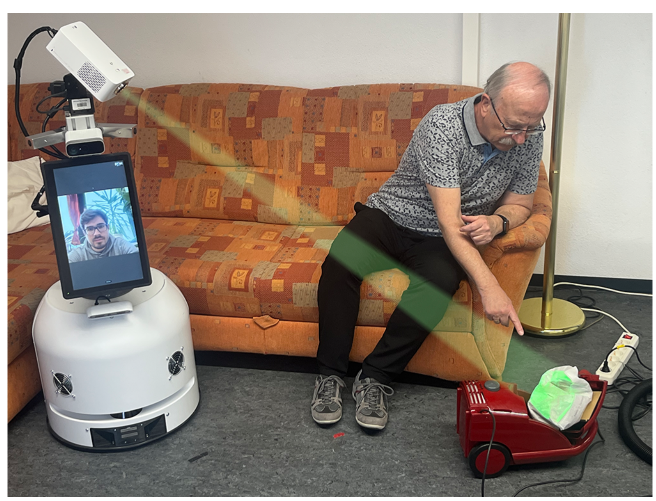

Fig. 4. Example scenario, where the virtual guest helps the local host by highlighting the problem.

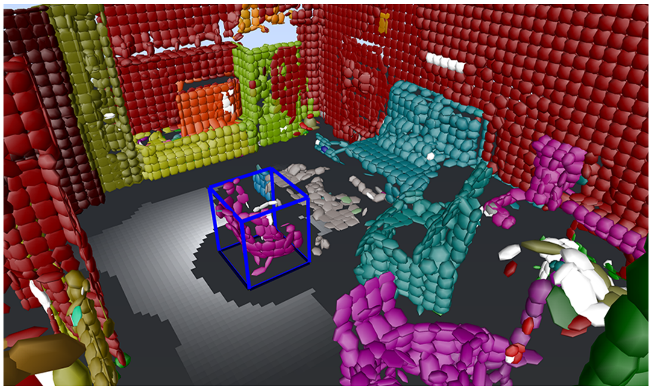

Fig. 5. Panoptic NDT map of a real apartment with a 3D bounding box highlighting a chair selected as target for robot navigation. The robot respects the chair's orientation as visualized by the cost function (ground plane): white indicates optimal while black indicates least optimal. Panoptic labels are visualized by small color differences based on the semantic colors: red - wall, yellow - cabinet, purple - chair, cyan - sofa, gray - table, orange - tv.

simple poses, the more intelligent drive-to-object is intended to enable driving to known objects in the environment, as shown in Fig. 5. Another important behavior is the docking behavior, which drives the robot to a docking station to enable recharging. Additional behaviors, such as follow-user, search-user, or approach-user are implemented using the person tracker and motor controller skills. These behaviors are useful for keeping the local host in view during a video call or for searching for her or him when a new call arrives. To further improve the user search behavior in regard to searching time, we are currently extending the behavior to also make use of the panoptic NDT maps [28]. The idea is to search for the local host based on intelligent priors, e.g., in the kitchen or dining area at lunch time. The highlight-object behavior utilizes the projector controller skill to adjust the projector for pointing or highlighting, as shown in Fig. 4.

C. Application Layer

The application layer implements the overall application logic as well as the interface to both the local host and the virtual guest. It is built as a state machine and is composed of behaviors and simple rules to switch between them. To effectively enable a collaborative telepresence, our application is tightly coupled with a server that hosts the simple web-based application. Fig. 6 shows the prototype of the web interface.

Currently, Jitsi Meet is used to establish the connection for the video call, enabling the virtual guest to use its notebook or tablet, while the local host can see her or him on the 15.6" display of the robot. We build upon Jitsi Meet as it can be self-hosted and is open source and thus is well suited to be integrated in our application. Unfortunately, it is based on virtual rooms the users can connect to, and thus there are no call notifications. By contrast, the local host and the virtual guest need to agree to meet at a certain time beforehand. We plan to incorporate an additional external service to overcome this limitation and to deliver call notifications to the local host by the search behavior as realized in [12].

Once the video call is established, the virtual guest is allowed to remotely control the robot. However, the local host is still able to command the robot, and retains the ability to terminate the connection at any point, effectively revoking control from the virtual guest. To address the concern of intuitive control, identified in [23], we provide a variety of control options for the robot. The virtual guest is able to access the cameras (see Fig. 6a) and can control the robot through a virtual joystick (see Fig. 6b). However, the field of view is limited, and remote control of the robot can be difficult, especially for inexperienced virtual guests. To mitigate this problem, we display an additional top-down view with laser information of the local surroundings of the robot (see Fig. 6c). In this view, critical obstacles are displayed that may not be visible in the camera images. Together with rendered spatial relations of the robot, this view facilitates the control in unknown environments. Alternative control options are to use predefined navigation points (Fig. 6d) or to set a goal by clicking on the floor seen in the camera image (Fig. 6a). When a click is made, a ray gets projected from the camera's center to the ground plane to retrieve the target position, which is then used as an input to the drive-to behavior, navigating the robot autonomously to the selected location. This method results in two advantages. First, it reuses the already implemented collision avoidance, without distracting the virtual guest with additional information. Second, this method works even with a high-latency connection or even occasional connection losses. However, while already improving the remote control to inexperienced virtual guests, this solution does not work when objects or walls are selected. To overcome this limitation, we intend to project the clicked target position into the semantic or panoptic 3D NDT map as

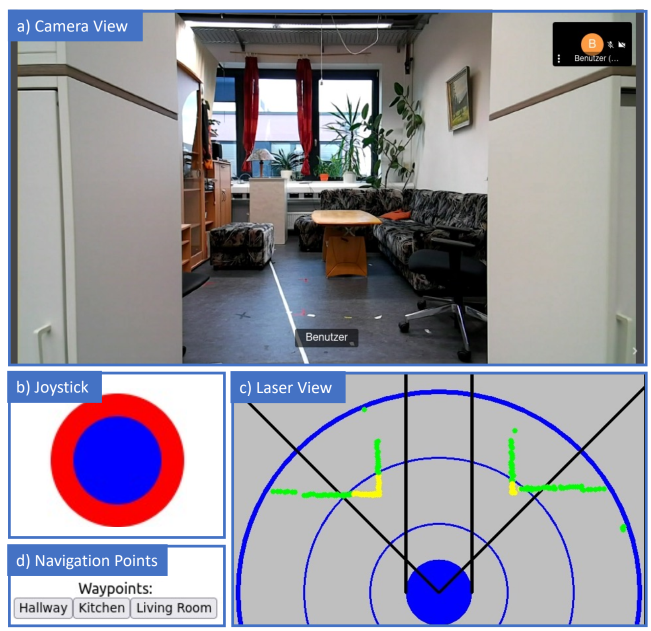

Fig. 6. Current prototype of the user interface of the virtual guest: a) current camera view, b) virtual joystick to send drive commands, c) corresponding top view of local laser measurements (yellow and green depending on distance) with range circles and dimension of the robot and d) predefined positions to let the robot navigate to.

well. The selected cell can then be used for further analysis in order to derive a valid target pose close by. A valid pose can be determined by considering 3D information only or even smarter by incorporating additional knowledge from the semantic, panoptic, or orientation information. For example, when clicking on the chair in Fig. 5, the robot could use the estimated orientation of the chair to find a non-obstructive pose to look at the seat. If there are multiple options, the virtual guest could select which position should be used as the target.

To provide a similar intuitive way for determining navigation poses to the local host, we plan to integrate a gesture control. To avoid misinterpreted or ambiguous gestures, the local host is supposed to use one distinct gesture to gain the robots attention, similar to the wake word of voice assistants. Afterward, the local host could, for example, indicate the robot to come closer or send it to a goal by pointing to a target.

To enable easier collaboration, we offer the virtual guest options to highlight objects in the apartment of the local host. This will be enabled through a click into the camera image. Highlighting can be done by displaying a dot on the projector (see Fig. 4). We further plan to implement a more intelligent highlighting by additionally incorporating information from the 3D panoptic maps. This way, highlighting only one specific object through the projector can be enabled. As an additional use case for the projector, we consider the possibility of using the 3D NDT map to find a free area on the wall to project any visual media into the local host's apartment. Some examples could be the projection of the virtual guest, movies, or internet content. This feature has the potential to enhance social connection and facilitate shared activities, such as watching a movie together or sharing a photo album.

After the video call, the robot automatically navigates to its docking station if recharging is required. This ensures availability and long operational times.

In [23], concerns were raised regarding the security of the robot as well as data privacy. To address these concerns, access to the application is restricted to trusted networks. By utilizing a self-hosted Jitsi server for video conferencing, we eliminate the need for third-party servers, thereby increasing data privacy. Additionally, all required computations are performed on the robot itself, reducing data transfer and further enhancing privacy by keeping most data on the robot. Furthermore, private data, such as video streams, are not stored persistently. All these measures contribute to improved data privacy.

V. Concept for Evaluation

To assess the effectiveness of our CO-HUMANICS robot companion in improving the social interactions of older adults, we plan to conduct a series of experiments from both the robotics and social perspectives. These experiments will help to assess the impact on co-presence and collaboration. In this section, we highlight relevant prior work and outline potential ideas for these experiments while discussing suitable metrics.

A. Technical Evaluation

We will conduct a series of experiments for evaluating the application as well as the underlying behaviors. One experiment will involve evaluating the search-user behavior, which is responsible for finding the local host within the apartment. This can be done by measuring the success rate and average search time, similar to [36]. Another experiment will evaluate the success rate of the follow-user behavior by measuring how often the person is lost on average over a given distance, how close the robot can maintain the desired following distance, and how centered the person is in the camera view, similar to [37]. For object highlighting, an experiment will evaluate the accuracy of the projection. To assess the quality, we plan to compare the highlighted and the desired area by means of the intersection over union. We further plan to repeat this experiment over various distances. Another experiment will evaluate the remote-control features by driving through an obstacle course while virtual guests are just able to control the robot with one of the implemented features, i.e., using the virtual joystick or selecting targets in the camera image. The frequency of near collisions, actual collisions, and time to course completion will be measured to compare the features.

B. Social-Science Studies

Several studies have been conducted within the project to explore the social aspect of robots for co-presence. This includes research on the frequency of use and perceived quality of established media, as well as the potential use of innovative media such as the presented robot [38]. Additionally, a scoping review of reviews has provided an overview of the current state of using communication technologies to counteract loneliness in older adults, examining theoretical frameworks, study designs, and positive effects of the technology [39]. Research on the requirements of the target group has also been conducted, guiding the development process and following a human-centered design approach [23]. The robot platform enables future in-depth social-science studies to be conducted. These studies might evaluate a variety of social factors, such as the impact of the robot companion on social interactions, the effectiveness of different control features, or the level of perceived co-presence during video calls [40], [41]. Experiments might also explore factors such as recall of information, proxemics, and other social phenomena similar as in [19], [40], [42], [43]. By conducting experiments from both the perspective of the local host and the virtual guest, we might gain insights into how the robot facilitates social interactions for both parties.

VI. Conclusion

In this paper, we have summarized related work in the field of robotic systems with telepresence features and have given a report on the progress in the CO-HUMANICS project. We have presented the first prototype of the CO-HUMANICS robot companion, focusing on hardware and software specifications and the vision of features. For our system we aim for natural interaction, for which we intend to provide remote control ability as well as features to highlight objects in the apartment, enabling an immersive and collaborative experience. Finally, we have introduced a concept for a future evaluation.

References

[1] H.-M. Gross, St. Mueller, Ch. Schroeter, M. Volkhardt, A. Scheidig, K. Debes, K. Richter, and N. Doering, "Robot Companion for Domestic Health Assistance: Implementation, Test and Case Study under Everyday Conditions in Private Apartments," in *Proc. of IROS*. IEEE, 2015, pp. 5992–5999.

[2] K. Zsiga, A. Tóth, T. Pilissy, O. Péter, Z. Dénes, and G. Fazekas, "Evaluation of a companion robot based on field tests with single older adults in their homes," *Assistive Technology*, vol. 30, no. 5, pp. 259–266, 2018.

[3] H.-M. Gross, A. Scheidig, St. Mueller, B. Schuetz, C. Fricke, and S. Meyer, "Living with a Mobile Companion Robot in your Own Apartment - Final Implementation and Results of a 20-Weeks Field Study with 20 Seniors," in *Proc. of ICRA*. IEEE, 2019, pp. 2253–2259.

[4] D. Portugal, L. Santos, P. Alvito, J. Dias, G. Samaras, and E. Christodoulou, "SocialRobot: An interactive mobile robot for elderly home care," in *2015 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2015, pp. 811–816.

[5] D. Portugal, P. Alvito, E. Christodoulou, G. Samaras, and J. Dias, "A Study on the Deployment of a Service Robot in an Elderly Care Center," *International Journal of Social Robotics*, vol. 11, no. 2, pp. 317–341, 2019.

[6] F. Graf, Ç. Odabaşi, T. Jacobs, B. Graf, and T. Födisch, "MobiKa - Low-Cost Mobile Robot for Human-Robot Interaction," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2019, pp. 1–6.

[7] M. Bajones, D. Fischinger, A. Weiss, P. D. L. Puente, D. Wolf, M. Vincze, T. Körtner, M. Weninger, K. Papoutsakis, D. Michel *et al.*, "Results of Field Trials with a Mobile Service Robot for Older Adults in 16 Private Households," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 9, no. 2, pp. 1–27, 2019.

[8] M. Luperto, J. Monroy, J. R. Ruiz-Sarmiento, F.-A. Moreno, N. Basilico, J. Gonzalez-Jimenez, and N. A. Borghese, "Towards Long-Term Deployment of a Mobile Robot for at-Home Ambient Assisted Living of the Elderly," in *2019 European Conference on Mobile Robots (ECMR)*. IEEE, 2019, pp. 1–6.

[9] M. Luperto, M. Romeo, J. Monroy, A. Vuono, N. Basilico, J. Gonzalez-Jimenez, and N. A. Borghese, "What is my Robot Doing? Remote Supervision to Support Robots for Older Adults Independent Living: a Field Study," in *2021 European Conference on Mobile Robots (ECMR)*. IEEE, 2021, pp. 1–7.

[10] S. Coşar, M. Fernandez-Carmona, R. Agrigoroaie, F. Ferland, F. Zhao, S. Yue, N. Bellotto, A. Tapus *et al.*, "ENRICHME: Perception and Interaction of an Assistive Robot for the Elderly at Home," *International Journal of Social Robotics*, vol. 12, no. 3, pp. 779–805, 2020.

[11] S. Laniel, D. Létourneau, F. Grondin, M. Labbé, F. Ferland, and F. Michaud, "Toward enhancing the autonomy of a telepresence mobile robot for remote home care assistance," *Paladyn, Journal of Behavioral Robotics*, vol. 12, no. 1, pp. 214–237, 2021.

[12] T. Wengefeld, B. Schuetz, G. Girdziunaite, A. Scheidig, and H.-M. Gross, "The MORPHIA Project: First Results of a Long-Term User Study in an Elderly Care Scenario from Robotic Point of View," in *Proc. of ISR*, 2022.

[13] K. Nowak, "Defining and differentiating copresence, social presence and presence as transportation," in *presence 2001 conference, Philadelphia, PA*, vol. 2. Citeseer, 2001, pp. 686–710.

[14] M. Minsky, "Telepresence," *Omni Magazine*, p. 45–51, 1980.

[15] B. Ricks, C. W. Nielsen, and M. A. Goodrich, "Ecological displays for robot interaction: A new perspective," in *Proc. of IROS*, vol. 3. IEEE, 2004, pp. 2855–2860.

[16] F. Ferland, F. Pomerleau, C. T. Le Dinh, and F. Michaud, "Egocentric and exocentric teleoperation interface using real-time, 3d video projection," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 2009, pp. 37–44.

[17] A. Reveleau, F. Ferland, M. Labbé, D. Létourneau, and F. Michaud, "Visual representation of interaction force and sound source in a teleoperation user interface for a mobile robot," *Journal of Human-Robot Interaction*, vol. 4, no. 2, pp. 1–23, 2015.

[18] D. Labonte, P. Boissy, and F. Michaud, "Comparative analysis of 3-d robot teleoperation interfaces with novice users," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 5, pp. 1331–1342, 2010.

[19] F. Michaud, P. Boissy, D. Labonté, S. Briere, K. Perreault, H. Corriveau, A. Grant, M. Lauria, R. Cloutier, M.-A. Roux *et al.*, "Exploratory design and evaluation of a homecare teleassistive mobile robotic system," *Mechatronics*, vol. 20, no. 7, pp. 751–766, 2010.

[20] D. Seichter, M. Köhler, B. Lewandowski, T. Wengefeld, and H.-M. Gross, "Efficient RGB-D Semantic Segmentation for Indoor Scene Analysis," in *Proc. of ICRA*, 2021, pp. 13 525–13 531.

[21] D. Seichter, S. Fischedick, M. Köhler, and H.-M. Gross, "Efficient Multi-Task RGB-D Scene Analysis for Indoor Environments," in *Proc. of IJCNN*, 2022, pp. 1–10.

[22] S. B. Fischedick, D. Seichter, R. Schmidt, L. Rabes, and H.-M. Gross, "Efficient Multi-Task Scene Analysis with RGB-D Transformers," in *Proc. of IJCNN*, 2023.

[23] M. Conde, V. Mikhailova, and N. Döring, "Towards augmented reality-based and social robot-based social integration of older adults: A user requirements analysis," in *HCI International 2023 Posters*, C. Stephanidis, M. Antona, S. Ntoa, and G. Salvendy, Eds. Cham: Springer Nature Switzerland, 2023, pp. 426–432.

[24] H.-M. Gross, Ch. Schroeter, St. Mueller, M. Volkhardt, E. Einhorn, A. Bley, Ch. Martin, T. Langner, and M. Merten, "Progress in Developing a Socially Assistive Mobile Home Robot Companion for the Elderly with Mild Cognitive Impairment," in *Proc. of IROS*. IEEE, 2011, pp. 2430–2437.

[25] M. Labbé and F. Michaud, "RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of Field Robotics*, vol. 36, no. 2, pp. 416–446, 2019.

[26] E. Einhorn and H.-M. Gross, "Generic 2D/3D SLAM with NDT maps for lifelong application," in *Proc. of ECMR*, 2013.

[27] D. Seichter, P. Langer, T. Wengefeld, B. Lewandowski, D. Höchemer, and H.-M. Gross, "Efficient and Robust Semantic Mapping for Indoor Environments," in *Proc. of ICRA*, 2022.

[28] D. Seichter, B. Stephan, S. B. Fischedick, S. Mueller, L. Rabes, and H.-M. Gross, "PanopticNDT: Efficient and Robust Panoptic Mapping," in *Proc. of IROS*, 2023.

[29] S. Kreiss, L. Bertoni, and A. Alahi, "OpenPifPaf: Composite Fields for Semantic Keypoint Detection and Spatio-Temporal Association," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13 498–13 511, 2022.

[30] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[31] T. Wengefeld, St. Mueller, B. Lewandowski, and H. M. Gross, "A Multi Modal People Tracker for Real Time Human Robot Interaction," in *IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN), New Delhi, India*. IEEE, 2019, p. 8 pages.

[32] L. Treybig, S. Saini, S. Werner, U. Sloma, and J. Peissig, "Room acoustic analysis and BRIR matching based on room acoustic measurements," *Audio Engineering Society Conference on Audio for Virtual and Augmented Reality, Redmond, USA*, 2022.

[33] L. Treybig, S. Werner, F. Klein, and S. V. Gari, Amengual Garí, "Robust reverberation time estimation for audio augmented reality applications," *154. Convention of the Audio Engineering Society, Espoo, Finnland*, 2023.

[34] G. Stolz, S. Werner, F. Klein, L. Treybig, A. Bley, and C. Martin, "Autonomous robotic platform to measure spatial room impulse responses," *49. Annual Conference on Acoustics, Hamburg, Germany*, 2023.

[35] G. Götz, I. Ananthabhotla, S. V. A. Garí, and P. Calamia, "Autonomous room acoustic measurements using rapidly-exploring random trees and gaussian processes," *10. Convention of the European Acoustics Association Forum Acusticum*, 2023.

[36] M. Volkhardt and H.-M. Gross, "Finding People in Home Environments with a Mobile Robot," in *Europ. Conf. on Mobile Robots (ECMR)*, 2013.

[37] A. Vorndran, T. Q. Trinh, St. Mueller, A. Scheidig, and H. M. Gross, "How to always keep an eye on the user with a mobile robot?" in *Proc. of ISR*. VDE Verlag, 2018, pp. 219–225.

[38] N. Döring, V. Mikhailova, K. Brandenburg, W. Broll, H.-M. Gross, S. Werner, and A. Raake, "Digital media in intergenerational communication: Status quo and future scenarios for the grandparent–grandchild relationship," *Universal Access in the Information Society*, Dec 2022.

[39] N. Döring, M. Conde, K. Brandenburg, W. Broll, H.-M. Gross, S. Werner, and A. Raake, "Can communication technologies reduce loneliness and social isolation in older people? a scoping review of reviews," *International Journal of Environmental Research and Public Health*, vol. 19, no. 18, 2022.

[40] M. L. Walters, K. Dautenhahn, R. Te Boekhorst, K. L. Koay, D. S. Syrdal, and C. L. Nehaniv, “An Empirical Framework for Human-Robot Proxemics,” *Procs of New Frontiers in Human–Robot Interaction*, 2009.
[41] A. Kristoffersson, S. Coradeschi, K. Severinson Eklundh, and A. Loutfi, “Sense of Presence in a Robotic Telepresence Domain,” in *Universal Access in Human-Computer Interaction. Users Diversity*, C. Stephanidis, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 479–487.
[42] M. Lombard, T. B. Ditton, and L. Weinstein, “Measuring presence: the temple presence inventory,” in *Proceedings of the 12th annual international workshop on presence*, 2009, pp. 1–15.
[43] R. Mead and M. J. Matarić, “Robots Have Needs Too: How and Why People Adapt Their Proxemic Behavior to Improve Robot Social Signal Understanding,” *J. Hum.-Robot Interact.*, vol. 5, no. 2, p. 48–68, sep 2016.