International Conference on Intelligent Robots and Systems (IROS), Macau, China, November 2019.

A VR System for Immersive Teleoperation and Live Exploration with a Mobile Robot

Patrick Stotko[1], Stefan Krumpen[1], Max Schwarz[2], Christian Lenz[2], Sven Behnke[2], Reinhard Klein[1], and Michael Weinmann[1]

***Abstract*— Applications like disaster management and industrial inspection often require experts to enter contaminated places. To circumvent the need for physical presence, it is desirable to generate a fully immersive individual live teleoperation experience. However, standard video-based approaches suffer from a limited degree of immersion and situation awareness due to the restriction to the camera view, which impacts the navigation. In this paper, we present a novel VR-based practical system for immersive robot teleoperation and scene exploration. While being operated through the scene, a robot captures RGB-D data that is streamed to a SLAM-based live multi-client telepresence system. Here, a global 3D model of the already captured scene parts is reconstructed and streamed to the individual remote user clients where the rendering for e.g. head-mounted display devices (HMDs) is performed. We introduce a novel lightweight robot client component which transmits robot-specific data and enables a quick integration into existing robotic systems. This way, in contrast to first-person exploration systems, the operators can explore and navigate in the remote site completely independent of the current position and view of the capturing robot, complementing traditional input devices for teleoperation. We provide a proof-of-concept implementation and demonstrate the capabilities as well as the performance of our system regarding interactive object measurements and bandwidth-efficient data streaming and visualization. Furthermore, we show its benefits over purely video-based teleoperation in a user study revealing a higher degree of situation awareness and a more precise navigation in challenging environments.**

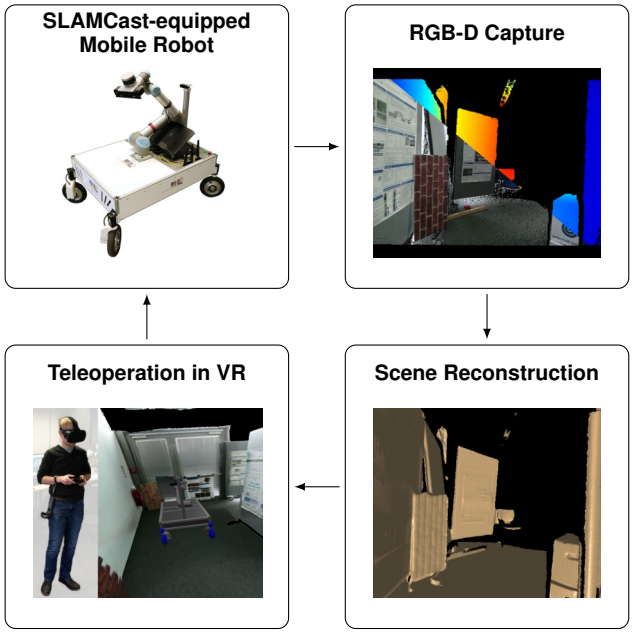

Fig. 1. High-level overview of our novel immersive robot teleoperation and scene exploration system where an operator controls a robot using a live captured and reconstructed 3D model of the environment.

I. Introduction

Due to the significant progress in VR displays in recent years, the immersive exploration of scenes based on virtual reality systems has gained a lot of attention with diverse applications in entertainment, teleconferencing [1], remote collaboration [2], medical rehabilitation and education. The quality of immersive experience of places, while being physically located in another environment, opens new opportunities for robotic teleoperation scenarios. Here, the major challenges include aspects such as resolution and frame rates of the involved display devices or the presentation and consistency of the respective data that increase the awareness of being immersed into the respective scene [3], [4], [5]. Another key challenge is the preservation of a high degree of situation awareness regarding the teleoperated robot's pose within its physical environment to allow precise navigation.

Purely video-based robot teleoperation and scene exploration is rather limited in the sense that the view is directly coupled to the area observed by the camera. This affects/impacts both the degree of immersion and the degree of situation awareness as remotely maneuvering a robot without having a complete overview regarding its current local environment is challenging, especially in case of narrow doors or corridors. Furthermore, remembering the locations of relevant scene entities is also complicated for video-only teleoperation which impacts independent visual navigation to already previously observed scene parts outside the current camera view. In contrast, transmitting the scene in terms of a reconstructed 3D model and immersing the teleoperator into this virtual scene is a promising approach to overcome these problems. Highly efficient real-time 3D reconstruction and real-time data transmission recently have been proven to be the key drivers to high-quality tele-conferencing within room-scale environments [1] or for immersive telepresence

[1]P. Stotko, S. Krumpen, R. Klein, and M. Weinmann are with the Institute of Computer Science II – Computer Graphics, University of Bonn, Germany {stotko,krumpen,rk,mw}@cs.uni-bonn.de

[2]M. Schwarz, C. Lenz and S. Behnke are with the Institute of Computer Science VI – Autonomous Intelligent Systems, University of Bonn, Germany {schwarz,lenz}@ais.uni-bonn.de, behnke@cs.uni-bonn.de

This work was supported by the DFG projects KL 1142/11-1 and BE 2556/16-1 (DFG Research Unit FOR 2535 Anticipating Human Behavior) as well as KL 1142/9-2 and BE 2556/7-2 (DFG Research Unit FOR 1505 Mapping on Demand).

based remote collaboration tasks in large-scale environments [2]. The benefit regarding situation awareness can still be preserved in case of network interruptions as the remote user remains immersed into the so far reconstructed scene and, after re-connection, newly arriving data can directly be integrated into the already existing scene model. However, a manual capturing process as used by Stotko et al. [2] is not possible within contaminated places. To the best of our knowledge, these kind of systems have not been adapted to the constraints of robot teleoperation – in our opinion, because the quality and scalability of 3D reconstruction methods has been too low until recently.

In this paper, we tackle the aforementioned challenges based on a novel system for immersive robot teleoperation and scene exploration within live-captured environments for remote users based on virtual reality and real-time 3D scene capture (see Fig. 1). This creation of an immersive teleoperation experience implies that the aforementioned conditions are met under strong time constraints to allow an immersive live teleoperation of the robot within the considered scenes and, hence, relies on on-the-fly scene reconstruction, immediate data transmission and visualization of the models to remote-connected users. For this purpose, our system involves a robot which is teleoperated through a respective scenario while capturing RGB-D data. To provide an as-complete-as-possible scene reconstruction for the teleoperation, the involved RGB-D camera can be moved via a manipulator, if existing on the robot. The captured data is sent to a reconstruction client component, that performs real-time dense volumetric Simultaneous Localization And Mapping (SLAM) based on voxel block hashing, and the current 3D model is managed on the server based on an efficient hash map data structure. Finally, the current model is streamed to the remote exploration client based on a low-bandwidth representation. Our approach allows a re-thinking regarding current exploration scenarios as encountered in e.g. disaster management, so that, on the long term, humans do not have to be exposed to e.g. contaminated environments but still can interact with the environment. It is furthermore desirable to add the functionality offered by the proposed framework to existing robotic systems. Therefore, we impose no requirements on the robotic platform: The robot-side system, consisting of an RGB-D camera and a notebook, is entirely self-contained. Optional interfaces allow tighter integration with the robot. Besides an evaluation of the performance of our system in terms of bandwidth requirements, visual quality and overall lag, we additionally provide the results of a psychophysical study that indicates the benefit of immersive VR based teleoperation in comparison to purely video-based teleoperation. Finally, we also show several example applications by demonstrating how the remote users can interact with both the robot and the scene.

In summary, the main contributions of this work are:

- The development of a novel system for immersive robot teleoperation and scene exploration within live-captured environments for remote users based on virtual reality and fast 3D scene capture – as needed e.g. for the inspection of contaminated scenes that cannot directly be accessed by humans,
- the implementation of the aforementioned system in terms of hardware and software,
- the evaluation of the benefits offered by this kind of immersive VR-based robot teleoperation over purely video-based teleoperation in the scope of a respective psychophysical study, and
- the evaluation of the system within proof-of-concept experiments regarding the robotic application of remote live site exploration.

II. Related Work

In this section, we review the progress made in telepresence systems with a particular focus on their application for teleoperation and remote collaboration involving robots.

Telepresence Systems: The key to success for the generation of an immersive and interactive telepresence experience is the real-time 3D reconstruction of the scene of interest. In particular due to the high computational burden and the huge memory requirements required to process and store large scenes, seminal work on multi-camera telepresence systems [6], [7], [8], [9], [10], [11] with less powerful hardware available at that time faced limitations regarding the capability to capture high-quality 3D models in real-time and to immediately transmit them to remote users. More recently, the emerging progress towards affordable commodity depth sensors including e.g. the Microsoft Kinect has successfully been exploited for the development of 3D reconstruction approaches working at room scale [12], [13], [14], [15]. Yet the step towards high-quality reconstructions remained highly challenging due to the high sensor noise as well as temporal inconsistency in the reconstructed data.

Recently, a huge step towards an immersive teleconferencing experience has been achieved with the development of the Holoportation system [1]. This system has been implemented based on the Fusion4D framework [16] that allows an accurate 3D reconstruction at real-time rates, as well real-time data transmission and the coupling to AR/VR technology. However, real-time performance is achieved based on massive hardware requirements involving several high-end GPUs running on multiple desktop computers and most of the hardware components have to be installed at the local user's side. Furthermore, only an area of limited size that is surrounded by the involved static cameras can be captured which allows the application of this framework for teleconferencing but prevents it from being used for interactive remote exploration of larger live-captured scenes.

Towards the goal of exploring larger environments as related to the exploration of contaminated scenes envisioned in this work, Mossel and Kröter [17] presented a system that allows interactive VR-based exploration of the captured scene by a single exploration client. Their system benefits from the real-time reconstruction based on current voxel block hashing techniques [18], however, it only allows scene exploration by one single exploration client, and, yet, the bandwidth requirements of this approach have been reported

to be up to 175 MBit/s. Furthermore, the system relies on the direct transmission of the captured data to the rendering client, which is not designed to handle network interruptions that force the exploration client to reconnect to the reconstruction client and, consequently, scene parts that have been reconstructed during network outage will be lost.

The recent approach by Stotko et al. [2] overcomes these problems and allows the on-the-fly scene inspection and interaction by an arbitrary number of exploration clients, and, hence, represents a practical framework for interactive collaboration purposes. Most notably, the system is based on a novel compact Marching Cubes (MC) based voxel block representation maintained on a server. Efficient streaming at low-bandwidth requirements is achieved by transmitting MC indices and reconstructing and storing the models explored by individual exploration clients directly on their hardware. This makes the approach both scalable to many-client-exploration and robust to network interruptions as the consistent model is generated on the server and the updates are streamed once the connection is re-established.

Robot-based Remote Telepresence: The benefits of an immersive telepresence experience have also been investigated in robotic applications. Communication via telepresence robots (e.g. [19], [20], [21]) is typically achieved based on a video/audio communication unit on the robot. More closely related to our approach are the developments regarding teleoperation in the context of exploring scenes. Here, remote users usually observe the video stream acquired by the cameras of the involved exploration robots to perform e.g. the navigation of the robot though a scene as well as the inspection of certain objects or areas. The visualization can be performed based on projecting live imagery onto large screens [22], walls [23], monitors [24], [25], [26], [27] or based on head-mounted display (HMD) devices [28], [29], [30], [31], [32], [33], [34]. Some of this work [30], [32], [33], [34] additionally coupled the interactions recorded by the HMD device to perform a VR-based teleoperation. However, the dependency on the current view of the used cameras does not allow an independent exploration of the scene required e.g. when remote users with different expertise have to focus on their individual tasks. Most closely related to our work is the approach of Bruder et al. [35], where a point cloud based 3D model of the environment is captured by a mobile robot and displayed by a VR-HMD. As discussed by the authors, the sparsity of the point cloud leads to the impression that objects or walls only appear solid when being observed from a sufficient distance and dissolve when being approached. This distance, in turn, also depends on the density of the point cloud. Furthermore, common operations including selection, manipulation, or deformation have to be adapted as ray-based approaches cannot be applied. Our approach overcomes these problems by capturing a surface-based 3D mesh model that can be immersively explored via live-telepresence based on HMDs.

Robot Platform: In Schwarz et al. [36], the rescue robot Momaro is described, which is equipped with interfaces for immersive teleoperation using an HMD device and 6D trackers. The immersive display greatly benefited the operators by increasing situational awareness. However, visualization was limited to registered 3D point clouds, which carry no color information. As a result, additional 2D camera images were displayed to the operator to visualize texture. Momaro served as a precursor to the Centauro robot [37], which extends the Momaro system in several directions, including immersive display of RGB-D data. However, the system is currently limited to displaying live data without aggregation.

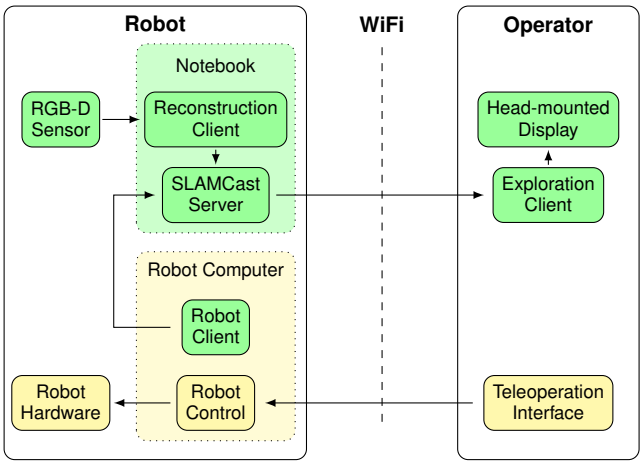

Fig. 2. Implementation of our immersive teleoperation system. The system allows the operator to immerse into the reconstructed scene to gain a third-person overview, while teleoperating the robot using existing teleoperation devices (e.g. a gamepad). Components in green are part of the SLAMCast framework; yellow boxes correspond to existing parts of the robotic system.

III. Overview

The main goal of this work is the design and implementation of a practical system for immersive robot teleoperation and scene exploration within live-captured environments for remote users based on virtual reality and real-time 3D scene capture (see Fig. 2). For this purpose, our proposed system involves (1) a robotic platform moving through the scene and performing scene capture, (2) an optional robot client that provides information about the current robot posture, (3) a reconstruction client that takes the captured data and computes a 3D model of the already observed scene parts, (4) a server that maintains the model and controls the streaming to the individual exploration clients, and (5) the connected exploration clients that perform the rendering e.g. on HMDs and can be used for teleoperation. By design, our system offers the benefits of allowing a large number of exploration clients, where, in addition to the teleoperator maneuvering the robot, several remote users may independently inspect the reconstructed scene and communicate with each other, e.g. for disaster management purposes.

In the following, we provide more details regarding the implementation of the involved components.

IV. Robot-based Scene Scanning

Mobile scene scanning was performed using the ground robot Mario (see Fig. 3), a robot with steerable wheels

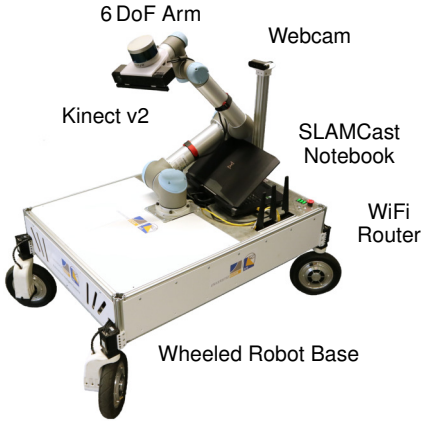

Fig. 3. The Mario robot is the exemplary target platform of our work. It has been equipped with an additional Kinect v2 RGB-D sensor and a notebook for processing and streaming of the reconstructed scene.

capable of omnidirectional locomotion. Mario won the Mohammed bin Zayed International Robotics Challenge 2017 (MBZIRC)[1] both in the UGV task and the Grand Challenge. For details on Mario, we refer to the work of Schwarz et al. [38]. Important for this work, Mario offers a large footprint, which yields high stability and few high-frequency movements of the camera. On the other hand, Mario can be difficult to maneuver in tight spaces, since it is designed for high-speed outdoor usage. Mario can be operated remotely using a WiFi link based on various sensors on the robot.

The key features of the robotic capturing system are:

Driving Unit: Based on the assumption of mostly flat terrain, we used a four-wheel-based robot system to allow a stable operation. In particular, we use an omnidirectional base due to its benefits regarding the precise positioning of the robot and the avoidance of complicated maneuvering for small adjustments as required in our envisioned contaminated site exploration scenario. Driven by the requirements of MBZIRC, the direct-drive brushless DC hub motors inside each steerable wheel allow reaching velocities of up to 4 m/s. In the indoor exploration scenario considered here, we limit the velocity to 0.15 m/s.

Robot Arm: Mario is equipped with a Universal Robots UR5, an off-the-shelf 6 DoF arm which offers more than sufficient working range to pan and tilt the endeffector-mounted camera sensor in order to increase the captured scene area. During scene exploration, the camera is automatically moved along Z-shaped trajectories to increase the field of view and thus the completeness of the captured model.

RGB-D Sensor: We extended the arm with the Microsoft Kinect v2, an off-the-shelf RGB-D sensor. This camera provides RGB-D data with a resolution of 512×424 pixels at 30 Hz. Note that RGB-D sensors in smartphones like the ASUS Zenfone AR sensor could also be used. Although these have a lower resolution and frame rate, they still allow for a sufficient reconstruction as shown by Stotko et al. [2].

[1] http://www.mbzirc.com

Electrical System: To meet the high voltage requirements imposed by the brushless wheel motors, the robot is powered by an eight-cell LiPo battery with 16 Ah and 29.6 V nominal voltage which allows operation times of up to 1 h depending on the task intensity. The UR5 arm is also run directly from the battery.

Data Transmission: The system is equipped with a Netgear Nighthawk AC1900 router that allows remotely monitoring the system as well as transmission of the scene data to clients. Additionally, the robot is equipped with a Velodyne VLP-16 3D LiDAR as well as a wide-angle Logitech webcam (that can be used for teleoperation). To keep requirements minimal, we did not integrate the LiDAR into our system, although this is a possible extension point. During the experiments, the robot is teleoperated through an existing wireless gamepad interface, which controls the omnidirectional velocity (2D translation and rotation around the vertical axis). We do not impose any requirements on the teleoperation method besides that it is compatible with third-person control, i.e. that it is usable while standing next to the robot (in reality or in VR).

V. Live Teleoperation and Exploration System

The aforementioned robotic capturing system is used in combination with an efficient teleoperation system consisting of the following components:

A. Reconstruction Client

RGB-D data captured by the robot are transmitted to the reconstruction client component, where a dense virtual 3D model is reconstructed in real-time using volumetric fusion into a sparse set of spatially-hashed voxel blocks based on implicit truncated signed distance fields (TSDFs) [39], [18]. Fully reconstructed voxel blocks, i.e. blocks that fall outside the current camera frustum, are queued for transmission to the central server component. Furthermore, the set of actively reconstructed visible voxel blocks is also added to the set of to-be-streamed blocks when the robot stops moving as well as at the end of the session [2]. Subsets of these blocks are then progressively fetched, compressed using lossless real-time compression [40], and streamed to the server. In addition, the reconstruction client transmits the current estimated camera pose to the server which is broadcasted to the exploration clients and used for the visualization of the camera's view frustum and the robot within the scene.

B. Robot Client

We introduce a novel component in the SLAMCast framework that allows the efficient and modular extension to a robot-based live telepresence and teleoperation system. This component is required if the camera is actuated on the robot – in this case, the pose of the robot components cannot be computed from the camera pose alone. The robot client solves this problem by providing the SLAMCast system with the poses of all robot links (in our exemplary case with Mario the posture of the 6 DoF arm as well as the wheel orientations). This information is transmitted to the

SLAMCast server and then broadcasted to the exploration clients. In combination with the estimated camera pose, this enables an immersive visualization of the robot within the scene. Note that the interface to the robotic system could be extended by streaming additional sensor data (e.g. LiDAR data) to the server. However, this work focuses on a minimally-invasive solution for immersive teleoperation and such extensions are thus out of scope.

C. Server

The server component manages the global model as well as the stream states of each connected exploration client, i.e. the set of updated voxel blocks that need to be streamed to the individual client. For efficient streaming to the clients, the received TSDF voxel blocks are converted to the bandwidth-efficient MC voxel block representation [2] and then added to the stream sets of each connected exploration client. Here, we used a simplified version of the Marching Cubes (MC) technique [41] where the weights have been discarded. In case a client re-connects to the server, the complete list of voxel blocks is added to its stream set in case the previously streamed parts are lost caused by e.g. accidentally closing the client by the user.

D. Exploration Client

At the remote expert's site, the exploration client requests updated scene parts either based on its current viewing pose, i.e. the parts that the user is currently exploring and interested in, in the order of the reconstruction, which resembles the movement of the robot, or in an arbitrary order which can be used to prefetch the remaining parts of the model outside the current view. Once the requested compressed MC voxel data arrived, they are uncompressed and passed to a reconstruction thread which generates a triangle mesh using Marching Cubes [41] as well as three additional levels of detail for efficient rendering. Furthermore, a virtual model of the robot is visualized within the scene using the estimated camera pose as well as the poses of the robot components. Since the estimated robot position might be affected by jittering effects due to imperfect camera poses, we apply a temporal low-pass filter on the robot's base pose. This ensures a smooth and immersive teleoperation experience.

In addition, our system can handle changes in the scene over time as e.g. occurring when doors have been opened or objects/obstacles have been removed. This is achieved by a reset function with which the exploration client may request scene updates for selected regions. In this case, the already reconstructed parts of the 3D model of the scene that are currently visible are deleted and the respective list of blocks is propagated to the server and exploration clients.

VI. Experimental Results

After evaluating our VR-based teleoperation system in the scope of a user study, we provide a brief performance evaluation of the proposed approach as well as some proof-of-concept applications regarding how a remote user can interact with the scene. A subset of this functionality is also demonstrated in the supplemental video.

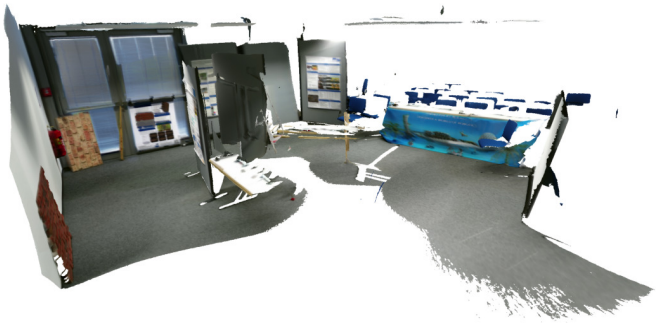

Fig. 4. Reconstructed 3D model of the teleoperation scene.

Fig. 5. Teleoperation experiment. Left: Baseline experiment with wide-angle camera feed. Right: Teleoperation using the proposed VR system.

A. Implementation

To implement the live teleoperation system, we use a laptop running the reconstruction client as well as the server component and a desktop computer that acts as the exploration client. The laptop and the desktop computer have been equipped with an Intel Core i7-8700K CPU (laptop) and Intel Core i7-4930K CPU (desktop), 32 GB RAM as well as a NVIDIA GTX 1080 GPU with 8 GB VRAM. Note that the system also allows additional exploration clients to be added if desired. Additionally, the visualization of the data for the exploration client users is performed using an HTC Vive HMD device that has a native resolution of 1080×1200 pixels per eye. Due to the lens distortion applied by the HTC Vive system, the rendering resolution is 1512×1680 pixels per eye as reported by the VR driver resulting in a total resolution of 3024×1680 pixels. Throughout all experiments, both computers were connected via WiFi. Furthermore, we used a voxel resolution of 5 mm and a truncation region of 60 mm – common choices for voxel-based 3D reconstruction.

B. Evaluation of User Experience

To assess the benefit of our immersive VR-based teleoperation system, we conducted a user study where we asked the participants to maneuver a robot through an elaborate course with challenges of different difficulties (see Fig. 6). A reconstructed 3D model of the course is shown in Fig. 4.

Participants: In total, 20 participants voluntarily took part in the experiment (2 females and 18 males between 22 and 56 years, mean age 29.25 years). All the participants were naïve to the goals of the experiment, provided informed consent, reported normal or corrected-to normal visual and hearing acuity. Before conducting the experiments, the users

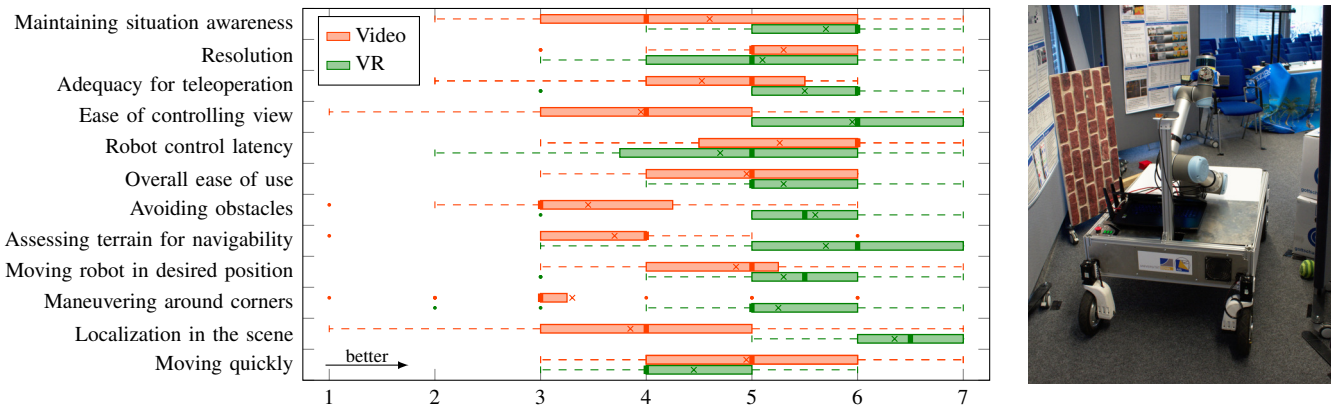

Fig. 6. User study. Left: Statistical results, i.e. median, lower and upper quartile (includes interquartile range), lower and upper fence, outliers (marked with •) as well as the average value (marked with ×), for each aspect as recorded in our questionnaire. For most aspects, our VR-based system achieved higher ratings on the 7-point Likert scale than a video-based approach. Right: Our robot Mario in the most difficult part of the course.

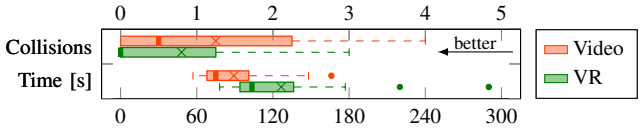

Fig. 7. User study: Statistical results for the number of collisions between robot and environment, and time needed for completing the course.

got a brief training regarding the control instructions and a short practical training for all involved conditions.

Stimuli: Robot teleoperation was performed in two different modes (see Fig. 5). In *VR mode*, the robot navigation was performed based on immersing the user into the remote location of the robot via standard VR devices (in this case, the HTC Vive) and were able to follow the robot in terms of walking behind or, in case of larger distances, teleporting to the desired positions in the scene. Here, the scene depicted in the HMD corresponds to the 3D model of the already reconstructed scene parts, which can be explored independently from the current view of the camera. The rationale behind this experiment are the expected higher degrees of immersion and situation awareness as users get a better impression regarding distances in the scene as well as occurring obstacles. Note that automatically following the robot instead is highly susceptible to motion sickness as it may not fit to the motion inherent to human behavior. In *video mode*, the users had to steer the robot through the same scenario purely based on video data depicting the current view of the camera on the robot arm. Hereby, the flexibility of getting information outside the current camera view is lost. As a consequence, we expect a lower situation awareness due to a more difficult perception of distances between objects in the scene as well as occurring obstacles. Each participant performed the task once in VR and once in video mode. We varied the order of these stimuli over the participants to avoid possibly occurring systematic bias due to training effects. Since further multi-modal feedback is rather suited for attention purposes and less for accuracy of control, we left the integration and analysis of this aspect for future work.

Performance measures: In addition to gathering individual ratings for certain properties on a 7-point Likert scale, we also analyze the number of errors (collisions with the environment) made in the different modes and the total execution time required to navigate from the starting point to the target location.

Discussion: In Fig. 6, we show the statistical results obtained from the ratings provided by the participants for both VR-based and video-based robot teleoperation. The main benefits of our VR system can be seen in the ratings regarding self-localization in the scene, maneuvering around narrow corners, avoiding obstacles, the assessment of the terrain for navigability as well as the ease of controlling the view. For these aspects, the boxes defined by the medians and interquartile ranges do not overlap indicating a significant difference in favor of the VR-based teleoperation. Furthermore, there is evidence that the VR mode is rated to be well-suited for teleoperation and that the robot can be easier moved to target positions. These facts also support the general impression of the participants regarding a higher degree of situation awareness with the VR teleoperation, thereby following our expectations stated above.

On the other hand, it is likely that the higher degree of immersion also leads to closer, more-time consuming inspection, thus, limiting the speed of robot motion. Furthermore, the perceived latency was rated slightly better for the video-based mode. The time until the scene data are streamed from the reconstruction client to the server, i.e. the time until it is fully reconstructed or prefetched, depends on the camera movement and is within a few seconds. A further slight deviation of the ratings in favor of the video-based mode can be seen regarding the resolution – which is, in the case of the VR-based system, limited to the voxel resolution. While the SLAMCast system supports on-demand local texture mapping of the current camera image onto the reconstructed 3D model, further advances towards the enhancement of texture resolution could help to bridge this last gap.

Fig. 7 shows the statistical results for the number of collisions and time needed to complete the course with both modes. The participants completed the course faster using

Fig. 8. Completion of scene model during capturing process: The images depict the scene model at different time steps. Depending on the regions that have been captured by the robot while moving through the scene, the captured 3D model of the environment gets more complete.

video mode since more time was used in VR mode for inspecting the situation (e.g. by walking around the robot in VR). Teleportation inside the VR environment generally took some time, especially for participants without VR experience. This could be improved by creating even more intuitive user interfaces for movement in VR and issuing navigation goals. However, due to the improved situation awareness, more collisions could be avoided in VR mode.

C. Performance Evaluation

For performance evaluation, we first provide an overview of the bandwidth requirements as well as a visual validation of the completeness of the virtual 3D model generated over time of the proposed system. For this purpose, we acquired two datasets based on the robotic platform and performed the reconstruction of the 3D models on the reconstruction client which are streamed to the server (first computer). A benchmark client (second computer) requests voxel block data with a package size of 512 blocks at a fixed frame rate of 100 Hz. To avoid overheads that may bias the benchmark, we directly discard the received data.

We observed a mean bandwidth required for streaming the data from the server to the benchmark client of 14 MBit/s and a maximum bandwidth of 25 MBit/s, which is well within the typical limits of a standard Internet connection. In Fig. 8, we demonstrate the completeness of the generated 3D model over time. While at the beginning only a small area of the scene is visible to the exploration client, the remaining missing parts of the scene are progressively scanned by the robot, transmitted, and integrated in the client's local model. In contrast to point cloud based techniques [35], a closed-surface representation preserves the impression that objects or walls appear solid when viewed from varying distances.

D. Interaction of Remote Users with the Scene

Managing contaminated site exploration or evacuation scenarios often involves the measurement of distances such as door widths in order to select and guide required equipment to the respective location. For this purpose, we implemented operations for measuring 3D distances based on the controllers of the HMD device to allow user-scene interaction. This can be useful in order to determine whether a different robot or the required equipment would fit through a narrow space, for example a door as shown in Fig. 9. The measurement accuracy is determined by the voxel resolution, which is chosen according to the noise of the RGB-D camera as well as the tracking accuracy of the 3D reconstruction algorithm. Considering the height and width of the doors measured in the corridor (see Fig. 9), we observed errors of up to 1 cm which is sufficient for rescue management.

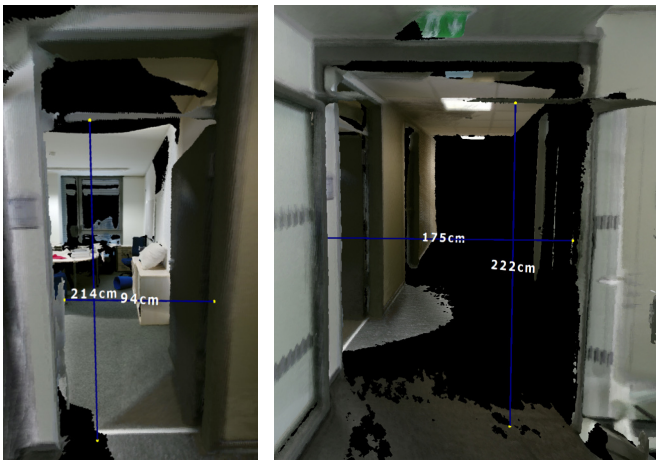

Fig. 9. Examples of interactively taken measurements of heights and widths of a corridor as well as door widths taken to guide the further management process. The real sizes of the doors (i.e. the ground truth values) are 95 cm×215 cm (left) and 174 cm×222 cm (right).

In addition, we also allow the remote user to label areas as interesting, suspicious or incomplete which is integrated into the overall map and the capturing robot may return to complete or refine the scan. Since the SLAMCast system supports multi-client telepresence, a further remote user may perform this task while the other one is teleoperating the robot. This enrichment of the captured 3D map with possibly annotated scene parts that have to be completed or refined can also directly be provided to further robots or the already used capturing robot. Thereby the respective interactions of these robots with the scene can be guided (scan completion or refinement, transport of equipment). So far, we did not include this functionality but leave it for future developments.

VII. Conclusion

We presented a novel robot-based live immersive and teleoperation system for exploring contaminated places that are not accessible by humans. For this purpose, we used a state-of-the-art robotic system which captures the environment with an RGB-D camera moved by its arm and transmits these data to a reconstruction and telepresence platform. We demonstrated that our system allows interactive immersive

scene exploration at acceptable bandwidth requirements as well as an immersive teleoperation experience. Based on the implementation of several example operations, we also show the benefit of our proposed setup regarding the improvement of the degree of immersion and situation awareness for the precise navigation of the robot as well as the interactive measurement of objects within the scene. In contrast, this level of immersion and interaction cannot be reached with video-only systems.

REFERENCES

[1] S. Orts-Escolano *et al.*, "Holoportation: Virtual 3D Teleportation in Real-time," in *Proc. of the Annual Symp. on User Interface Software and Technology*, 2016, pp. 741–754.

[2] P. Stotko, S. Krumpen, M. B. Hullin, M. Weinmann, and R. Klein, "SLAMCast: Large-Scale, Real-Time 3D Reconstruction and Streaming for Immersive Multi-Client Live Telepresence," *IEEE Trans. on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 2102–2112, 2019.

[3] G. Fontaine, "The Experience of a Sense of Presence in Intercultural and Int. Encounters," *Presence: Teleoper. Virtual Environ.*, vol. 1, no. 4, pp. 482–490, 1992.

[4] R. M. Held and N. I. Durlach, "Telepresence," *Presence: Teleoper. Virtual Environ.*, vol. 1, no. 1, pp. 109–112, 1992.

[5] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," *Presence: Teleoper. Virtual Environ.*, vol. 7, no. 3, pp. 225–240, 1998.

[6] H. Fuchs, G. Bishop, K. Arthur, L. McMillan, R. Bajcsy, S. Lee, H. Farid, and T. Kanade, "Virtual Space Teleconferencing Using a Sea of Cameras," in *Proc. of the Int. Conf. on Medical Robotics and Computer Assisted Surgery*, 1994, pp. 161 – 167.

[7] T. Kanade, P. Rander, and P. J. Narayanan, "Virtualized reality: constructing virtual worlds from real scenes," *IEEE MultiMedia*, vol. 4, no. 1, pp. 34–47, 1997.

[8] J. Mulligan and K. Daniilidis, "View-independent scene acquisition for tele-presence," in *Proc. IEEE and ACM Int. Symp. on Augmented Reality*, 2000, pp. 105–108.

[9] H. Towles *et al.*, "3D Tele-Collaboration Over Internet2," in *Proc. of the Int. Workshop on Immersive Telepresence*, 2002.

[10] T. Tanikawa, Y. Suzuki, K. Hirota, and M. Hirose, "Real World Video Avatar: Real-time and Real-size Transmission and Presentation of Human Figure," in *Proc. of the Int. Conf. on Augmented Tele-existence*, 2005, pp. 112–118.

[11] G. Kurillo, R. Bajcsy, K. Nahrsted, and O. Kreylos, "Immersive 3D Environment for Remote Collaboration and Training of Physical Activities," in *IEEE Virtual Reality Conference*, 2008, pp. 269–270.

[12] A. Maimone, J. Bidwell, K. Peng, and H. Fuchs, "Enhanced personal autostereoscopic telepresence system using commodity depth cameras," *Computers & Graphics*, vol. 36, no. 7, pp. 791 – 807, 2012.

[13] A. Maimone and H. Fuchs, "Real-time volumetric 3D capture of room-sized scenes for telepresence," in *Proc. of the 3DTV-Conference*, 2012.

[14] D. Molyneaux, S. Izadi, D. Kim, O. Hilliges, S. Hodges, X. Cao, A. Butler, and H. Gellersen, "Interactive Environment-Aware Handheld Projectors for Pervasive Computing Spaces," in *Proc. of the Int. Conf. on Pervasive Computing*, 2012, pp. 197–215.

[15] B. Jones *et al.*, "RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-camera Units," in *Proc. of the Annual Symp. on User Interface Software and Technology*, 2014, pp. 637–644.

[16] M. Dou *et al.*, "Fusion4D: Real-time Performance Capture of Challenging Scenes," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 114:1–114:13, 2016.

[17] A. Mossel and M. Kröter, "Streaming and Exploration of Dynamically Changing Dense 3D Reconstructions in Immersive Virtual Reality," in *Proc. of IEEE Int. Symp. on Mixed and Augmented Reality*, 2016, pp. 43–48.

[18] O. Kähler, V. A. Prisacariu, C. Y. Ren, X. Sun, P. Torr, and D. Murray, "Very High Frame Rate Volumetric Integration of Depth Images on Mobile Devices," *IEEE Trans. on Visualization and Computer Graphics*, vol. 21, no. 11, pp. 1241–1250, 2015.

[19] A. Kristoffersson, S. Coradeschi, and A. Loutfi, "A Review of Mobile Robotic Telepresence," *Adv. in Hum.-Comp. Int.*, vol. 2013, pp. 3:3–3:3, 2013.

[20] I. Rae, B. Mutlu, and L. Takayama, "Bodies in motion: Mobility, presence, and task awareness in telepresence," in *SIGCHI Conf. on Human Factors in Computing Systems*, 2014, pp. 2153–2162.

[21] L. Yang, C. Neustaedter, and T. Schiphorst, "Communicating Through A Telepresence Robot: A Study of Long Distance Relationships," in *CHI Conf. Extended Abstracts on Human Factors in Computing Systems*, 2017, pp. 3027–3033.

[22] D. Wettergreen, D. Bapna, M. Maimone, and G. Thomas, "Developing Nomad for robotic exploration of the Atacama Desert," *Robotics and Autonomous Systems*, vol. 26, no. 2, pp. 127–148, 1999.

[23] D. J. Roberts, A. S. Garcia, J. Dodiya, R. Wolff, A. J. Fairchild, and T. Fernando, "Collaborative telepresence workspaces for space operation and science," in *IEEE Virtual Reality*, 2015, pp. 275–276.

[24] G. Podnar, J. M. Dolan, A. Elfes, M. Bergerman, H. B. Brown, and A. D. Guisewite, "Human Telesupervision of a Fleet of Autonomous Robots for Safe and Efficient Space Exploration," in *Annual Conf. on Human-Robot Interaction*, 2006.

[25] I. Rekleitis, G. Dudek, Y. Schoueri, P. Giguere, and J. Sattar, "Telepresence across the Ocean," in *Canadian Conf. on Computer and Robot Vision*, 2010, pp. 261–268.

[26] D. G. Macharet and D. A. Florencio, "A collaborative control system for telepresence robots," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012, pp. 5105–5111.

[27] V. Kaptelinin, P. Björnfot, K. Danielsson, and M. Wiberg, "Mobile Remote Presence Enhanced with Contactless Object Manipulation: An Exploratory Study," in *CHI Conf. on Human Factors in Computing Systems, Extended Abstracts*, 2017, pp. 2690–2697.

[28] B. P. Hine III *et al.*, "The Application of Telepresence and Virtual Reality to Subsea Exploration," in *Proc. of the 2nd Workshop on Mobile Robots for Subsea Environments*, 1994.

[29] A. J. Elliott, C. Jansen, E. S. Redden, and R. A. Pettitt, "Robotic Telepresence: Perception, Performance, and User Experience," Unites States Army Research Laboratory, Tech. Rep., 2012.

[30] U. Martinez-Hernandez, L. W. Boorman, and T. J. Prescott, "Telepresence: Immersion with the iCub Humanoid Robot and the Oculus Rift," in *Biomimetic and Biohybrid Systems*, 2015, pp. 461–464.

[31] ——, "Multisensory Wearable Interface for Immersion and Telepresence in Robotics," *IEEE Sensors Journal*, vol. 17, no. 8, pp. 2534–2541, 2017.

[32] L. Peppoloni, F. Brizzi, E. Ruffaldi, and C. A. Avizzano, "Augmented reality-aided tele-presence system for robot manipulation in industrial manufacturing," in *ACM Symp. on Virtual Reality Software and Technology*, 2015, pp. 237–240.

[33] P. Kurup and K. Liu, "Telepresence Robot with Autonomous Navigation and Virtual Reality: Demo Abstract," in *ACM Conf. on Embedded Network Sensor Systems*, 2016, pp. 316–317.

[34] J. I. Lipton, A. J. Fay, and D. Rus, "Baxter's Homunculus: Virtual Reality Spaces for Teleoperation in Manufacturing," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 179–186, 2018.

[35] G. Bruder, F. Steinicke, and A. Nüchter, "Poster: Immersive point cloud virtual environments," in *IEEE Symp. on 3D User Interfaces*, 2014, pp. 161–162.

[36] M. Schwarz *et al.*, "DRC Team NimbRo Rescue: Perception and Control for Centaur-like Mobile Manipulation Robot Momaro," in *The DARPA Robotics Challenge Finals: Humanoid Robots To The Rescue*. Springer, 2018, pp. 145–190.

[37] T. Klamt, D. Rodriguez, M. Schwarz, C. Lenz, D. Pavlichenko, D. Droeschel, and S. Behnke, "Supervised Autonomous Locomotion and Manipulation for Disaster Response with a Centaur-like Robot," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2018.

[38] M. Schwarz *et al.*, "Team NimbRo at MBZIRC 2017: Autonomous Valve Stem Turning using a Wrench," *Journal of Field Robotics*, 2018.

[39] M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger, "Real-time 3D Reconstruction at Scale Using Voxel Hashing," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 169:1–169:11, 2013.

[40] Y. Collet and C. Turner, "Smaller and faster data compression with Zstandard," https://code.facebook.com/posts/1658392934479273/smaller-and-faster-data-compression-with-zstandard, 2016, accessed: 2019-03-01.

[41] W. E. Lorensen and H. E. Cline, "Marching Cubes: A High Resolution 3D Surface Construction Algorithm," in *Proc. of the 14th Annual Conf. on Computer Graphics and Interactive Techniques*, 1987, pp. 163–169.