

# Intuitive Bimanual Telemanipulation under Communication Restrictions by Immersive 3D Visualization and Motion Tracking

Tobias Rodehutsors, Max Schwarz, and Sven Behnke

**Abstract**—Robots which solve complex tasks in environments too dangerous for humans to enter are desperately needed, e.g. for search and rescue applications. As fully autonomous robots are not yet capable of operating in highly unstructured real-world scenarios, teleoperation is often used to embed the cognitive capabilities of human operators into the robotic system. The many degrees of freedom of anthropomorphic robots and communication restrictions pose challenges to the design of teleoperation interfaces, though. In this work, we propose to combine immersive 3D visualization and tracking of operator head and hand motions to an intuitive interface for bimanual teleoperation. 3D point clouds acquired from the robot are visualized together with a 3D robot model and camera images using a tracked 3D head-mounted display. 6D magnetic trackers capture the operator hand motions which are mapped to the grippers of our two-armed robot Momaro. The proposed user interface allows for solving complex manipulation tasks over degraded communication links, as demonstrated at the DARPA Robotics Challenge Finals and in lab experiments.

## I. INTRODUCTION

Disaster scenarios like the Fukushima nuclear accident clearly reveal the need for robots which are capable to meet the requirements which arise during operation in real-world, highly unstructured and unpredictable situations, where human workers cannot be deployed due to radiation, danger of collapse or toxic contamination. As a consequence of the incident in Fukushima, the Defense Advanced Research Projects Agency (DARPA) hold the DARPA Robotics Challenge<sup>1</sup> (DRC) to foster the development of robots capable of solving tasks which are required to relief catastrophic situations and to benchmark these robots in a competition. During the DRC, the robots had to solve eight tasks within one hour: 1. Drive a vehicle to the disaster site, 2. Egress from the vehicle, 3. Open a door, 4. Turn a valve, 5. Cut a hole into a piece of drywall, 6. Solve a surprise manipulation task, 7. Overcome rough terrain or a field of debris, and 8. Climb some stairs. The tasks 4 to 7 needed to be solved inside a simulated building, where the communication between the operators and the robot was limited.

To participate in the DRC, we constructed the mobile manipulation robot Momaro and an operator station for it. The DRC requirements are beyond the state of the art of autonomous robotics. As fully autonomous systems which



Fig. 1: Intuitive teleoperation. Left: Upper (front) and lower body (back) operator. Right: Momaro opening the door.

work in these complex environments are not feasible yet, often human intelligence is embedded into the robot through teleoperation to improve the overall performance of the system. Human operators can easily react to unforeseen events, but require awareness of the situation. To this end, we equipped our robot with a 3D laser scanner and multiple cameras.

In this work, we are addressing two challenges of solving complex bimanual telemanipulation tasks at the DRC. The first challenge was that DARPA degraded the communication between the operators and the robot, and data transmission had to be carefully managed. To address these communication restrictions, we propose strategies for combining a low-latency low-bandwidth channel with a high-latency high-bandwidth channel. The second challenge is posed by the many DoFs of our robot. To successfully solve complex bimanual manipulation tasks, means are needed which enable the operator to control the robot in an intuitive way and relief him of the burden of translating control commands to resulting actions of the robot. In this work, we propose a teleoperation interface consisting of a stereoscopic head-mounted display (HMD) and two 6D magnetic trackers for the hands of the operator. The operator head motions are tracked to render views based on the available 3D point clouds for the HMD, which fit to his motions and therefore increase his feeling of immersion. The position and orientation of the magnetic trackers are mapped to the end-effectors of our robot using inverse kinematics with redundancy resolution to calculate positional control commands for Momaro's anthropomorphic arms. We use the Oculus Rift as HMD and the Razer Hydra as 6D hand trackers.

This work was supported by the European Union's Horizon 2020 Programme under Grant Agreement 644839 (CENTAURO).

All authors are with Rheinische Friedrich-Wilhelms-Universität Bonn, Computer Science Institute VI, Autonomous Intelligent Systems, Friedrich-Ebert-Allee 144, 53113 Bonn, rodehutsors@ais.uni-bonn.de, max.schwarz@uni-bonn.de, behnke@cs.uni-bonn.de

<sup>1</sup><http://www.theboticschallenge.org/>

Both are consumer-grade products and therefore available at a low price. This paper mainly focuses on the upper body manipulation capabilities of our system. Robot locomotion is described in a first report [1].

The contributions of this work are:

- 1) The design of an intuitive bimanual telemanipulation interface using low cost hardware.
- 2) The development of a communication strategy for the combination of a low-latency low-bandwidth channel with a high-latency high-bandwidth channel, which allows for intuitive telemanipulation under degenerated communication.
- 3) The integration and evaluation of our system during the DRC Finals and in lab experiments.

## II. RELATED WORK

Telemanipulation has been investigated by many groups. Here we focus on approaches for two-armed robots.

In the context of the DRC, O’Flaherty et al. [2] used 6 DoF magnetic trackers to directly control the end-effectors of their Hubo robot. Their work did not address situation awareness of the operator.

Kron et al. [3] designed a bimanual haptic telepresence system for use in explosive ordnance disposal following a master-slave approach. Their robot is equipped with two 4 DoF manipulators and jaw grippers as end-effectors. A fixed stereo camera pair is used to transmit a video stream from the remote environment to the operator who is wearing a HMD. The operator can control the manipulators using two PHANToM devices, which act as master and track motions in 6 DoF. The PHANToM devices as well as the finger gripping devices, which the operator uses to close the grippers, can display force feedback from the slave. Our robot can locomote by itself, has anthropomorphic arms with 7 DoF, is equipped with a 3D sensor and our HMD enables the user to freely look around in the remote scene.

Martins and Ventura [4] showed that the performance of users in a teleoperated search and rescue scenario significantly increased when using a HMD with an integrated head-tracker. Their robot was able to follow the operator’s head movements and thus improved the depth perception and the situation awareness of the user.

Telemanipulation is also used for minimally invasive surgery [5]. Hagn et al. [6], for example, developed the DLR MiroSurge setup, which allows for bimanual operation using the Omega7 haptic hand controllers. These controllers measure 6 DoF for the hand motions of the surgeon, provide an additional DoF for grasping, and render force feedback to the surgeon’s hands. The system uses robot arms which resemble the kinematic configuration of the human arm to ensure predictability, like we do in our Momaro system. The sensors are limited due to the spatial constraints of the field of application. An endoscopic stereo video camera is used and its images are displayed on a 3D screen.

The idea of using consumer-grade equipment for robotic applications is not new. Kot and Novák [7] used the Oculus

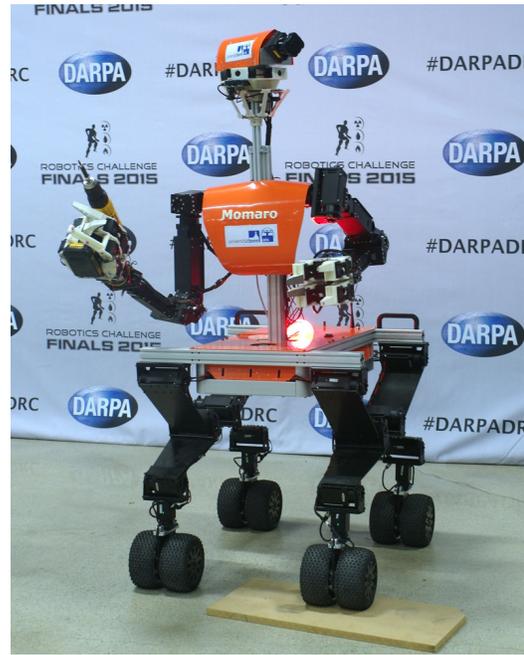


Fig. 2: Mobile manipulation robot Momaro.

Rift as well in their mobile manipulation setup using a four-wheeled robot with a 3 DoF arm.

Similarly, Smith and Christensen used the low-priced Wiimote game controller with an additional IR camera to track the position and orientation of the operator hand [8]. They use a minimum jerk human motion model to improve the precision of the tracking and achieved good results for minimally instructed users in a simple manipulation task. In contrast to the Wiimote, which can only measure linear accelerations, the Razer Hydra is able to determine absolute positions using a magnetic field.

There are also other groups which use the Razer Hydra and Oculus Rift in their robotic applications. SRI has designed the Taurus Dexterous Telepresence Manipulation System which is based on the daVinci technology and is intended for explosive ordnance disposal. It can be operated using the Razer Hydra controllers<sup>2</sup>. The UMass Lowell Robotics Lab demonstrated the control of a Baxter robot using the Razer Hydra and Oculus Rift as well<sup>3</sup>. The same was done by Willow Garage using their PR2 robot platform<sup>4</sup>. To the best of our knowledge, these groups have not published their results.

## III. ROBOT HARDWARE

Our mobile manipulation robot Momaro, shown in Fig. 2 was specifically designed for the requirements of the DRC. Since state of the art approaches for bipedal locomotion are prone to falls and current generation robots are mostly not able to recover after these falls by themselves, we decided to equip Momaro with a total of four legs to minimize the

<sup>2</sup><https://youtu.be/cqBm97jBvuY>

<sup>3</sup><https://youtu.be/JHIZ-Y5qCmY>

<sup>4</sup><https://youtu.be/H0EoEyvTmiY>

probability of a falling. As robot locomotion using stepping is comparably slow, the legs end in pairs of steerable wheels. This allows the robot to omnidirectionally drive over flat terrain and to execute steps only if they are necessary to overcome larger obstacles.

On top of its flexible base, Momaro has an anthropomorphic upper body consisting of two adult-sized, 7DoF arms and a sensor head. The upper body of the robot is connected to the base by a torso yaw joint that increases the workspace of the end-effectors and allows the system to execute more tasks without the use of locomotion. Each arm is equipped with a custom hand with four fingers with 2DoF each. While the proximal segment of each finger is rigid, Festo FinGrippers are used as distal segments. These grippers deform if force is applied to them to better enclose a grasped object by enlarging the contact surface between object and gripper. The position of the finger tips on each finger can manually be reconfigured to allow pinch grips as well as cylindrical grasps. All joints of the robot are Robotix Dynamixel actuators.

Our robot is equipped with a variety of different sensors. First of all, a rotating 3D laser scanner is mounted on top of the sensor head providing a spherical field-of-view. Three full HD color cameras are attached to the sensor head for a panoramic view of the environment in front of the robot and a top-down wide angle camera is used to observe the movement of the arms of the robot and its interaction with the environment. Each hand is equipped with a camera which is located between its fingers. These cameras can be used to visually verify the correct grasp of objects. Furthermore, since these cameras are mounted at the end-effectors of the robot and can therefore be moved, they can be used to extend the view of the operators, for example, to view a scene from another perspective if the view from the head mounted top-down camera is occluded. The right hand of the robot is also equipped with a microphone to give the operators auditory feedback. In addition, the robot can measure joint torque and is equipped with an inertial measurement unit (IMU). The robot weighs about 58 kg including batteries for approximately 1.5 hours of operation.

#### IV. COMMUNICATION MANAGEMENT

One constraint during the DRC was the limited communication between the operator station and the robot, which was enforced to simulate degenerated communication as may occur in a real-world mission. The uplink from the operator station to the robot was limited to 9600 bit/s all the time. The downlink from the robot to the operator station was limited to 300 Mbit/s outside of the building during the driving tasks, the door task, and the stairs task. As usual, the wireless communication link does not guarantee packet delivery, so robotic systems have to deal with packet loss. Inside the building, the downlink was limited to 9600 bit/s, interleaved with one second long bursts of 300 Mbit/s bandwidth. These burst became more frequent during the run and the interruptions vanished completely after 45 minutes into the run.

To cope with this degraded communication, sensor information cannot be transferred unselected and uncompressed. The main idea of our communication system is to transfer stationary information about the environment over the high-latency high bandwidth channel, while we use the low-latency low bandwidth channel to transfer frequently changing data. Both are then combined on the operator station to render immersive 3D visualizations with low latency for the operators.

The point clouds generated by the 3D laser scanner are not transmitted over the low bandwidth link. Instead, a local multi-resolution map is generated on the robot, which is maintained by aggregating the measurements of the 3D laser scanner [9]. This map is transmitted during a burst to the operator station. To be able to render the movement of the robot in the environment, we transfer odometry data based on the measurements of the IMU and wheel odometry with 1 Hz over the low-latency communication channel. Similarly, we transfer the joint positions of the robot with 1 Hz and a resolution of 16 bits over the low-latency link, except for the most distal joints in the kinematic tree, which are sent with 8 bit resolution. This makes it possible to give the operators fast feedback for transmitted motion commands by means of a rendered robot model in the environment.

Since visual information is of crucial importance to human operators, we also transmit a low resolution video stream. As Momaro is equipped with a variety of cameras, an operator needs to select the camera whose output should be sent over the low bandwidth link. The selection of the camera depends on the currently executed task and is also often changed during a task. The video stream has an update rate of 1 Hz, too. The transferred images are downsampled to a resolution of  $160 \times 120$  pixels and compressed using H.264. To cope with loss of frames, the encoder is used in the periodic intra refresh mode, which can guarantee recovery from frame loss after a number of frames, without the need to periodically transmit keyframes. During a communication burst, all camera images are transferred in high quality using JPEG compression. All images depicted in this paper which are taken from camera feeds of the robot are shown in high quality, regardless if the system was currently in low bandwidth communication or not, as we recorded video from all cameras during the runs on the robot.

As the Dynamixel actuators can easily overheat, it is quite important to observe their temperature during operation to be able to take corrective actions. Therefore, we transfer the current temperature and current torque of each joint during a burst. Furthermore, the temperature and torque of the hottest actuator is also transferred every second when only limited bandwidth is available.

#### V. TELEOPERATION INTERFACE

The control of the robot is divided among several operators. To ensure distinctness and ease the communication between the operators, the fingers of the robot model in our 3D visualisations are colored. In addition, the inner side and outer side of the robot links are colored differently to enable



Fig. 3: Upper-body operator interfaces: Left: Oculus Rift DK2 HMD. Right: Razer Hydra 6DoF controllers.

the operators to easily perceive the orientation of the links (see top of Fig. 4).

The most important operators are the upper body operator and the lower body operator. Furthermore, there are several support operators for special tasks. The lower body operator is responsible for controlling the locomotion of the robot by either omnidirectionally driving the robot by means of a joystick or through the execution of stepping motions. The work station of the lower body operator is equipped with several monitors which display the different camera views and a 3D visualisation of the robot in its environment as is shown on the left of Fig. 1.

The upper body operator is responsible for controlling the arms of the robot using the Razer Hydra<sup>5</sup> controllers. To give the operator an immersive feeling of the robot in its environment, he is wearing an Oculus Rift<sup>6</sup>. Both devices are shown in Fig. 3.

The Oculus Rift is a HMD which displays stereoscopic images and tracks the movement of the operator head in 6DoF. It uses a combination of a 3 axes gyrometer and acceleration sensors to estimate the rotation of the head and an additional camera-based tracking unit to determine the head position. It displays an egocentric view from the perspective of the robot which is based on the generated local multi-resolution map. The tracked head movements of the operator are used to update the stereoscopic view and allow the operator to freely look around in the current scene. In addition, the transferred 2D camera images can be displayed in the view of the upper body operator to give him additional clues about the current situation as can be seen in the upper part of Fig. 4. The selection and positioning of these views are performed by an additional support operator using a graphical user interface (see bottom of Fig. 4).

The Razer Hydra hand-held controllers use a weak magnetic field to sense the position and orientation of the hands of the operator with an accuracy of 1 mm and 1°. The controllers have several buttons, an analog stick and a trigger. These controls map to different actions which the upper body operator can trigger. The measured position and orientation of the operator hands are mapped to the position and orientation of the respective robot gripper to allow the operator to intuitively control them. We do not aim for a one-to-one mapping between the workspace of the robot

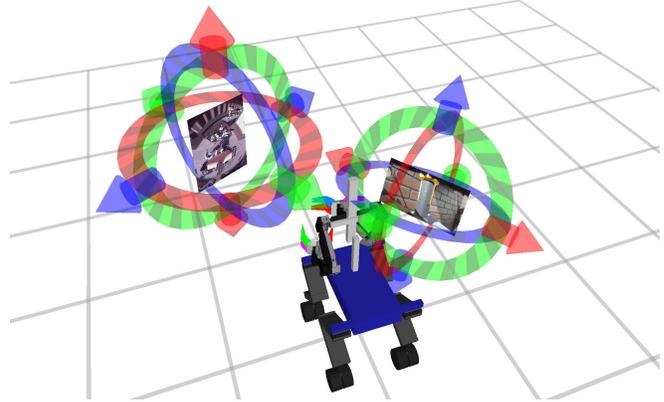
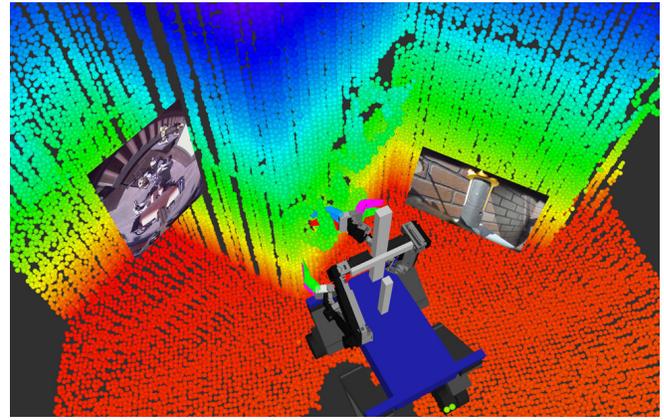


Fig. 4: Top: 3rd person view of the upper body operator view. Bottom: Same scene as seen by a support operator.

and the reachable space of the magnetic trackers. Instead, differential commands are sent to the robot. Therefore, the operator has to pull the trigger on the respective controller if he wants to control the right or the left arm. Vice versa, the upper body operator needs to release the trigger to give up the control. This indexing technique enables the operator to move the robot grippers to the boundaries of the workspace in a comfortable way. Due to the limitation of the bandwidth, we send the currently desired 6D poses of the end-effectors only with a rate of 5 Hz to the robot.

When such a task space command reaches the robot, it does not plan its motion towards the desired position but instead uses the Reflexes library [10] to interpolate between the current and desired end-effector position. For each intermediate 6D pose, we calculate the inverse kinematics with redundancy resolution using the selectively damped least squares (SDLS) approach [11]. SDLS is an iterative method based on the singular value decomposition of the Jacobian of the current robot configuration. It applies a damping factor for each singular value based on the difficulty of reaching the target position. Furthermore, SDLS sets the target position closer to the current end-effector position if the target position is too far away from the current position. SDLS effectively computes target position as close as possible to 6D poses if they are not within the reachable workspace of the end-effector. Furthermore, we combined

<sup>5</sup><http://sixense.com/razerhydra>

<sup>6</sup><https://www.oculus.com/en-us/rift/>



Fig. 5: Left: Inserting the plug as seen from the right hand camera. Middle: Momaro turns the valve. Right: Flipping the switch as seen from the top-down camera.

SDLS with a nullspace optimization based on the projection of a cost function gradient to the nullspace [12]. The used cost function is a sum of three different components:

- 1) Joint angles near the limits of the respective joint are punished to avoid joint limits if possible.
- 2) The difference between the robot's last and newly calculated configuration is penalized to avoid jumps during a motion.
- 3) The difference from a user-specified convenient configuration and the newly calculated configuration is punished to reward this specific arm position. We chose this convenient configuration to position the elbow of each arm next to the body as seen in the top of Fig. 4.

To better control the end-effectors, the upper body operator can switch between a precision mode and the regular mode. In precision mode, motion is scaled down, such that large movements of the controllers result in smaller movements of the robot arms, thus enabling the operator to perform tasks with higher accuracy. The upper body operator also has the ability to rotate the torso around the yaw axis using the analog stick on the left hand-held controller. The upper body operator can close or open the robot grippers with a button push. Since the number of buttons on the Razer Hydra controllers is limited and the system has several different predefined grasps, a support operator can trigger these grasps using a simple graphical user interface.

In addition, the upper body operator has the ability to move its point of view freely in the  $xy$  plane out of the egocentric view using the analog stick of the right Razer Hydra controller and he can also flip the perspective by  $180^\circ$  at the push of a button. Both allows the operator to inspect the current scene from another perspective.

The control system checks for self-collisions and displays the links which are nearly in collision color-coded to the operators. The system stops the execution of motion commands if the operator moves the robot further into nearly self-collision. We do not check collisions with the environment, as they are necessary to perform manipulation tasks.

## VI. EVALUATION IN THE DRC FINALS

Our system was evaluated in the DRC Finals. In the following, we describe the tackled tasks and our approach for solving them in detail.

### A. Task Descriptions

Despite the fact that most tasks require a coordinated approach of locomotion and manipulation, four of the eight tasks were mainly concerned with manipulation. This paper focuses on the description of these manipulation-related tasks: Opening a door, turning a valve, cutting drywall, and two surprise manipulation tasks. Each team in the DRC Finals had two independent runs. Seven of the eight tasks stayed fixed for both runs and were known prior to the challenge. The surprise manipulation task, however, was subject of change and was announced to the teams the evening before the respective run.

1) *Opening a Door:* The first task which must be completed after egressing from the vehicle is to open a closed door. The handle of the door is located on the left-hand side and the door opens inwards, away from the robot. The door opens either by pressing the door handle down from above or up from below. First, the lower body operator centers the robot manually in front of the door in a way that it can directly pass through the door as soon as the door is opened. Since the finger tips of Momaro are flexible and can break easily if too much force is applied to them, a support operator triggers a motion primitive which folds the finger tips of the left hand aside and allows the robot to press the door handle with the joint servos instead. The upper body operator now uses the Razer Hydra controller to position the left hand below the door handle. Thereupon, the lower body operator increases the height of the base of the robot by extending its legs. As soon as the door handle is pushed upwards, the lower body operator drives the robot forwards to open the door. Only minor force is required to open the door. As soon as the door is fully opened, it is designed to stay open. The point for the completion of this task is given when the robot has passed completely through the door. Inside the building, degenerated communication kicks in.

2) *Turning a Valve:* This task requires the robot to open a valve by rotating it counter-clockwise by  $360^\circ$ . The exact diameter of the valve is not known prior to the run, but it is between 10cm and 40cm. The lower body operator positions the robot roughly in front of the valve. Then, a support operator marks the position and orientation of the valve for the robot using an 6D interactive marker [13] in

a 3D graphical user interface. After the valve is marked, a series of parameterized motion primitives, which use the marked position and orientation, are executed by the support operator to fulfill the task. First, the right hand is opened widely and the right arm moves the hand in front of the valve. The correct alignment of the hand and the valve is verified using the camera in the right hand and the position of the hand is corrected if the alignment is not yet good enough. Next, the flexible finger tips close around the outer part of the valve to get a firm grasp of the valve. Then, the hand is rotated counter-clockwise by  $180^\circ$ . After that, the hand opens again and the sequence is repeated until the valve is fully opened. The upper body operator is not involved in this task.

3) *Cutting Drywall*: The cutting task requires the robot to grasp one of two different supplied drill tools and use the tool to remove a marked circle from a piece of drywall by cutting around it. We decided to use the tool which needs to be switched on only once, instead of the tool which needs to be triggered constantly to keep working. To switch on the tool, one finger of the right hand of the robot is equipped with an additional bump to improve access to the trigger of the tool. If the tool is grasped correctly, this bump can be used to push the trigger of the tool to switch it on. The robot is not able to switch the tool off. After five minutes, the tool switches off automatically. The tool is grasped by the upper body operator using the Razer Hydra controller by moving the gripper in front of the tool and triggering a predefined grasp. The arm is then retracted by the upper body operator and a support operator triggers a motion primitive which rotates the hand by  $180^\circ$ . As the first grasp does not close the hand fully, the tool can now slip into a desired predefined position. A support operator now executes a grasp closure motion to switch the tool on. After the tool is switched on, the upper body operator positions the right hand with the tool in front of the drywall. A parameterized motion primitive is then used to cut an approximately circular hole into the drywall. When the task is completed, the upper body operator puts the tool down on the floor.

4) *Flipping a Switch*: This task was the surprise task for the first run. The task is to flip a big switch from its on-position into its off-position. After the robot was driven in front of the switch, the upper body operator solves this task on his own. He closes the fingers of the right hand half way using a predefined motion and then moves the hand towards the lever of the switch. As soon as the hand encloses the lever, the operator moves the hand downwards to flip the switch into its off-position.

5) *Plug*: This task was the surprise task for the second run. The task was to pull a plug from a socket and plug it into a different socket which was located 0.5 m horizontally away from the first plug. For this task, we added additional fingers to the left hand of the robot to increase the surface area which has contact with the plug. During this task, a support operator controls the left gripper using a 6D interactive marker. The interactive marker allows to move the gripper in individual Cartesian directions, which is difficult using the hand-held



Fig. 6: Top Left: Grasping the cutting tool as seen from the right hand camera. Right: Same scene as seen from the top-down camera. Middle Left: Grasp used to switch on the tool. Bottom Left: Momaro cutting the drywall as seen from a sensor head camera.

controllers.

## B. Results

We publicly demonstrated our telemanipulation approach during the DRC Finals in 2015 in Pomona, USA. Each team participating in the challenge had two independent runs to demonstrate the capabilities of their robotic system. During our runs, we split all operation between the two main operators and a total of seven support operators. One support operator assisted the upper body operator by modifying his view on his commands. Two operators were responsible for clearing the generated local multi-resolution maps from undesirable artifacts. Another support operator monitored the hardware and its temperature during the runs. Two more operators assisted the upper body operator by triggering additional predefined parameterized motions and grasps and were able to control the arms and grippers in joint space as well as in task space using a graphical user interface if necessary.

After we successfully demonstrated driving the vehicle and egress from the vehicle in our first run, we tried to open the door. On our first attempt, we missed the door handle, as the robot was too far away from the door. We corrected this and succeeded on the second attempt. The elapsed time for this task as well as all other attempted manipulation tasks are displayed in Table I. During our second run, the egress from the vehicle failed and we requested a reset. The robot was positioned directly in front of the door during the reset which decreases the time consumed by the door task in our second run. We demonstrated the turning of the valve successfully in both runs. During the first run, one finger tip of the right gripper slipped into the valve and was damaged when we retracted the end-effector from the valve. We continued the run, as this was only a minor damage. In both runs, we chose

TABLE I: Results obtained during the DRC Finals.

Task	Success	Time [min:s]	
		1st run	2nd run
Door	2/2	2:25	0:27
Valve	2/2	3:13	3:27
Cutting	1/1	12:23	-
Switch	1/1	4:38	-
Plug	1/1	-	9:58

The listed times are calculated based on a recorded video feed. All attempted manipulation task were successfully solved. The listed times include the time for the locomotion from the previous task to the current task.

to skip the cutting task first and attempt it as the last indoor task, as it had proven to be the most time consuming task during practicing in our lab. Since we did not have a mockup of the switch, we were not able to train this task prior to the run. Nevertheless, we succeeded at our first attempt. In our second run, it took us several attempts to solve the plug task. We used the camera in the right hand to verify that we successfully inserted the plug into the socket as can be seen in Fig. 5. After the surprise task, we solved the debris task in our first run by driving through it. Then, we attempted the previously skipped cutting task. We grasped the tool and rotated it upside down (see Fig. 6). Some manual adaptation of the gripper in joint space was necessary since the tool was not grasped as desired and was therefore unable to slip into its designated position which would prohibit us from switching it on. We used auditory feedback of the right hand microphone to verify that we switched the tool on. As we tried to cut the drywall, we became aware that the cutting tool was not correctly assembled. Therefore, our first run was paused by the DARPA officials and the cutting tool was replaced. The lost time was credited to us. During our second cutting attempt, our parameterized cutting motion primitive was not executed correctly as the robot was not properly aligned to the drywall. Therefore, the automated cutting motion did not remove all the designated material. We noticed this error during the execution of the motion and a support operator moved the right arm manually upwards with a lot of force. This broke the drywall and a point was awarded to us for the fulfillment of this task. Unfortunately, our robot became stuck during the traversal of the debris during our second run. Therefore, we were not able to execute this task on the second day.

Overall, our system was able to solve all attempted manipulation tasks. In our first run, we solved seven out of eight tasks in only 34 minutes. The stair climbing task was not attempted. This resulted in a 4th place in the final ranking. A compacted video<sup>7</sup> of our first run is available online.

## VII. EVALUATION OF A BIMANUAL TASK

During the DRC Finals, we rarely used more than one end-effector at a time. During the plug task, for example, we used the right end-effector camera to observe the motions of the left gripper. To evaluate the bimanual teleoperation

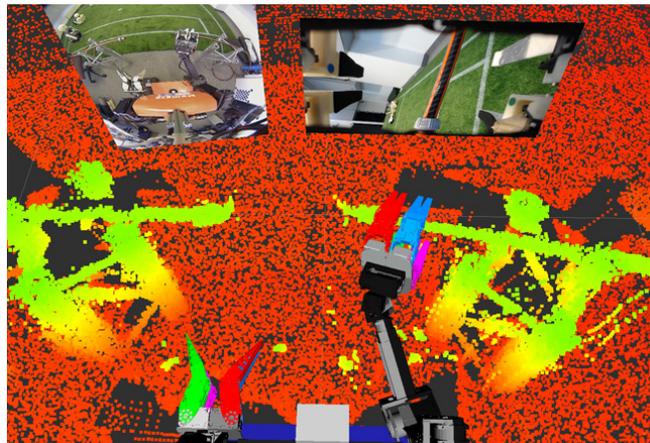


Fig. 7: Upper body operator view during the hose task.

capabilities of our system, we designed an additional task, which exceeds the requirements of the DRC Finals.

The task is to connect two flexible and unmodified water hoses (see Fig. 8). No locomotion is needed during this task, as the hoses are placed within the reachable workspace of the robot arms. The ends of the hoses, which need to be connected are not placed on the floor. Instead, traverses are used as support for the hoses to ease grasping. This task requires bimanual teleoperation as the hoses are flexible and not attached to a stable base. Therefore, the operator has to grasp both hoses with the left and right gripper, respectively. To establish the connection between both hoses, the extension adapter attached to the first hose must be inserted into the connector of the second hose and both hoses must be pushed together in the correct angle.

One support operator assisted the trained upper body operator during the task by controlling the camera images which are displayed in his HMD and by triggering grasps. A monoscopic view from the perspective of the upper body operator can be seen in Fig. 7. The hoses as well as the support traverses are clearly visible in the 3D point cloud, which gives the operator a good overview over the current situation. 2D camera images are displayed to aid the operator with additional visual clues. Self-collision detection was switched off, as it might prevent close proximity of the gripper fingers, which can be necessary to fulfill the hose



Fig. 8: Momaro connecting two hoses.

<sup>7</sup><https://youtu.be/NJHSFeIPsGc>

TABLE II: Execution times for the hose task (10 trials).

Task	Time [min:s]				
	Avg.	Median	Min.	Max.	Std. Dev.
Left grasp	0:44	0:38	0:27	1:20	0:16
Right grasp	0:45	0:40	0:34	1:04	0:10
Connect	1:36	1:32	1:07	2:04	0:21
Total	3:04	2:57	2:21	3:51	0:28

task. The operators were in a different room than the robot during the experiments and received information over the state of the robot and its environment only from the robot sensors. The communication bandwidth was not limited.

We performed the hose task 11 times in our lab. The execution of one trial was stopped, as the upper body operator moved the right arm into the base of the robot as he was grasping for the right hose. The results of the remaining 10 executions of the experiments are shown in Table II. The task consists of three parts which are separately listed:

- 1) Grab the left hose with the left gripper,
- 2) Grab the right hose with the right gripper, and
- 3) Connect both hoses.

On average, a little bit more than three minutes were needed to complete the whole task. The hardest part of the task was to establish the actual connection between both hoses, which accounted on average for more than half of the total elapsed time, as the upper body operator needed always more than one attempt to connect both hoses.

## VIII. CONCLUSION

In this paper, we gave a detailed description of our approach to intuitive bimanual telemanipulation under constrained communication. The upper body operator is equipped with a HMD showing a stereoscopic view from the perspective of its tracked head pose, which is roughly placed at the robot head pose. Together with the rendering of a 3D animated robot model, this gives him an immersive feeling of being inside the environment of the robot and directly controlling its arm and hand motions. The position and orientation of the operator hands are tracked and mapped to motions of the robot's anthropomorphic arms, thus enabling simple and intuitive use of the end-effectors. Operator assistance functions such as self-collision detection and redundancy resolution are applied to ensure the practicability of the system. We successfully integrated our telemanipulation method with our mobile manipulation robot Momaro and demonstrated its performance during the DRC. Additionally, we conducted lab experiments to evaluate the bimanual teleoperation capabilities of our system.

To solve complex manipulation tasks, our operators currently rely on 3D point clouds, visual and auditory feedback, and joint sensors from the robot. Additional touch and force-torque sensing in combination with a force feedback system for the upper body operator could potentially improve the manipulation capabilities of the human-robot system. This could, for example, be beneficial for peg-in-hole tasks such as the plug task during the DRC or the hose task, which require nimble manipulation skills.

Our telemanipulation system has currently only a low degree of autonomy and instead requires multiple human operators to control it. This allows our team to easily react to unforeseen events. However, the number of operators needed is quite high and so many trained operators are not always available. Therefore, it is necessary to add more autonomous monitoring and operator assistance functions to make the system manageable by fewer operators. Furthermore, the load on the operators could be reduced by carrying out more tasks autonomously. To this end, we plan to extend our methods for autonomous navigation, object manipulation and tool use that we developed for our cognitive service robot Cosero [14], [15], [16], [17] and our exploration and mobile manipulation robot Explorer [18].

## REFERENCES

- [1] M. Schwarz and S. Behnke, "Semi-autonomous concurrent driving-stepping locomotion," in *Late Breaking Result Abstracts at IEEE International Conference on Robotics and Automation (ICRA)*, 2015.
- [2] R. O'Flaherty, P. Vieira, M. Grey, P. Oh, A. Bobick, M. Egerstedt, and M. Stilman, "Humanoid robot teleoperation for tasks with power tools," in *Techn. for Practical Robot Applications (TePRA)*, 2013.
- [3] A. Kron, G. Schmidt, B. Petzold, M. Zäh, P. Hinterseer, E. Steinbach, et al., "Disposal of explosive ordnances by use of a bimanual haptic telepresence system," in *Robotics and Automation (ICRA)*, 2004.
- [4] H. Martins and R. Ventura, "Immersive 3-D teleoperation of a search and rescue robot using a head-mounted display," in *Proc. of Emerging Technologies & Factory Automation (ETFA)*, 2009.
- [5] G. H. Ballantyne and F. Moll, "The da Vinci telerobotic surgical system: The virtual operative field and telepresence surgery," *Surgical Clinics of North America*, vol. 83, no. 6, pp. 1293–1304, 2003.
- [6] U. Hagn, R. Konietzschke, A. Tobergte, M. Nickl, S. Jörg, B. Kübler, G. Passig, M. Gröger, F. Fröhlich, U. Seibold, et al., "DLR MiroSurge: a versatile system for research in endoscopic telesurgery," *Computer Assisted Radiology and Surgery*, vol. 5, no. 2, pp. 183–193, 2010.
- [7] T. Kot and P. Novák, "Utilization of the Oculus Rift HMD in mobile robot teleoperation," *Applied Mechanics and Materials*, vol. 555, pp. 199–208, 2014.
- [8] C. Smith, H. Christensen, et al., "Wiimote robot control using human motion models," in *Intelligent Robots and Systems (IROS)*, 2009.
- [9] D. Droschel, J. Stückler, and S. Behnke, "Local multi-resolution representation for 6D motion estimation and mapping with a continuously rotating 3d laser scanner," in *Robotics and Automation (ICRA)*, 2014.
- [10] T. Kröger, "Opening the door to new sensor-based robot applications: Reflexes motion libraries," in *Robotics and Automation (ICRA)*, 2011.
- [11] S. R. Buss and J.-S. Kim, "Selectively damped least squares for inverse kinematics," *Journal of Graphics, GPU, and Game Tools*, vol. 10, no. 3, pp. 37–49, 2005.
- [12] A. Liegeois, "Automatic supervisory control of the configuration and behavior of multibody mechanisms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 7, no. 12, pp. 868–871, 1977.
- [13] D. Gossow, A. Leeper, D. Hersberger, and M. Ciocarlie, "Interactive markers: 3-D user interfaces for ros applications," *IEEE Robotics & Automation Magazine*, vol. 4, no. 18, pp. 14–15, 2011.
- [14] M. Schwarz, J. Stückler, and S. Behnke, "Mobile teleoperation interfaces with adjustable autonomy for personal service robots," in *Human-Robot Interaction (HRI)*, 2014.
- [15] J. Stückler, D. Droschel, K. Gräve, D. Holz, M. Schreiber, A. Topalidou-Kyniazopoulou, M. Schwarz, and S. Behnke, "Increasing flexibility of mobile manipulation and intuitive human-robot interaction in RoboCup@Home," in *RoboCup 2013: Robot World Cup XVII*, ser. LNCS, vol. 8371. Springer, 2014, pp. 135–146.
- [16] J. Stückler and S. Behnke, "Adaptive tool-use strategies for anthropomorphic service robots," in *Humanoid Robots (Humanoids)*, 2014.
- [17] J. Stückler, R. Steffens, D. Holz, and S. Behnke, "Efficient 3D object perception and grasp planning for mobile manipulation in domestic environments," *Robotics and Autonomous Systems*, 2013.
- [18] J. Stückler, M. Schwarz, M. Schadler, A. Topalidou-Kyniazopoulou, and S. Behnke, "NimbRo Explorer: Semi-autonomous exploration and mobile manipulation in rough terrain," *Journal of Field Robotics*, 2015.