The 11th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI 2014)
Nov. 12-15, 2014 at Double Tree Hotel by Hilton, Kuala Lumpur, Malaysia

Task-oriented Teleoperation through Natural 3D User Interaction

Hyoungnyoun Kim, Jun-Sik Kim, Kwanghyun Ryu, Seyoung Cheon, Yonghwan Oh, and Ji-Hyung Park
Human-centered Interaction & Robotics Research Center,
Korea Institute of Science and Technology, Seoul, 136-791, Republic of Korea
(Tel : +82-2-958-5632; E-mail: {nyoun, junsik.kim, 022301, 023111, oyh, jhpark}@kist.re.kr)

Abstract - Teleoperation provides a technical means to perform desired tasks in remote environments. Teleoperation using direct control and haptic-mediated interactions requires significant effort for unskilled users to understand how to operate the robot. Task-oriented teleoperation supports human-level understanding to directly transfer the meaning of tasks from the user to the robot. In this paper, we propose a natural 3D interface to transfer task knowledge to a remote robot. We design a 3D manipulation system in a mixed reality environment that combines human hand gestures and a 3D view of the remote robot. We demonstrate that the remote robot successfully executes the ordered task even in dynamic environments.

Keywords - Teleoperation, natural user interface, 3D manipulation, mixed reality.

1. Introduction

Teleoperation is commonly used in robotics to refer to robots that are operated by a user from a distance [1]. Many types of teleoperation have been introduced and they are often associated with different levels of robotic automation [2]. Direct control connects the user action to the robot motion, whereas mediated teleoperation interprets the coordination of the user action into a robotic actuator action. If dynamic autonomy is applied to a robot in a remote space, the robot completes tasks and informs the user of the results. Regardless of the diversity of the types of teleoperation, the user accomplishes a specific task in an unreachable space through telerobotics.

However, the teleoperation system is difficult to control for an unskilled user. Even in emergency situations, only experts can command robots to conduct some tasks. In order to support unskilled users to control remote robots via teleoperation, the system should provide a simple and comprehensive interface. Rather than including all details of the operation, using high-level controls can enable easier control of the remote robot. Therefore, the required system is based on mediated teleoperation and the delivered information includes high-level meanings.

Task-oriented information is proposed for use as high level interpreted information. A task is a combination of separate actions and each action represents the motion data that a robot must conduct [3].

Even for the simple delivery of a task list, a natural interface must be provided. In terms of the human-computer interaction, there have been numerous suggestions of providing users with immersive and natural experiences using particular interactive platforms [4-6]. In this paper, we propose a task-oriented robotic teleoperation based on a natural 3D user interface. Through the applied 3D interaction method, the user can easily send orders to the remote robot for specific tasks, e.g. grasping and releasing objects, even if the user is considered to be unskilled. Moreover, task-oriented teleoperation enables the robot to adapt to dynamic environmental changes because the detail of the robotic motion is interpreted into new situations. We introduce a 3D interaction platform and then describe how the interface is applied to a robotic teleoperation system.

2. Natural User Interface for 3D Manipulation

In a teleoperation system, a user sends a task order to a robot in a remote space. The user view is restricted to an image stream from the remote space. When the images are displayed in 2D, the user has little freedom to manipulate and understand the visual content. Therefore, in the proposed system, 3D images that are composed of point clouds are used and these can be more easily manipulated.

The 3D image stream is displayed in real time. Regarding the 3D images, the focus is on the target object to be grasped and placed in a certain position. If the image contains numerous objects to grasp, the virtual models of the objects are augmented for the detected position in the 3D images. Thus, target objects can be manipulated.

The primary problem from the user side is how to intuitively manipulate the remote space and objects. In 3D interaction platforms in mixed reality, several studies have targeted the representation of physical spaces in 3D virtual information using glass, mirrors, and transparent displays. Holodesk provides 3D interaction with a see-through display that combined a mirror and a glass [4]. Lee et al. proposed a 3D interaction method based on a transparent display [5]. MirageTable is a mixed reality platform that provides a continuous space between the display and interaction area [6]. These platforms enlarge the opportunity to interact with the digital world in a physical space.

We applied the *DuplicateSpace* platform that supports 3D direct manipulation in mixed reality [7]. *DuplicateSpace* is a 3D interaction platform that enables direct manipulation using two-handed interactions in a virtual environment. The approach synchronizes the virtual space to the physical space so that the user understands the virtual space through navigating the physical space. Furthermore, space registration [8] and

978-1-4799-5333-2/14/$31.00 ©2014 IEEE

Mobiature [9] enable intuitive manipulation of 3D virtual objects.

At first, the space of the 3D images is reproduced on a physical space in front of the user. The user assigns a cube to duplicate the displayed 3D space. Then, the user adjusts the cube size in order to easily navigate the physical space. In this interaction platform, we used bare-handed interactions. The user's hands in the user-created space are also reproduced in the virtual space. Without focusing on the position of their hands, the user becomes aware of the interpreted space.

Users must see other viewpoints from the robotic view. Even though the camera view is fixed on a certain point, each 3D point in the point cloud can be changed using a rotation matrix. The user's left hand could change the viewpoint. In the physical user space, the camera view changes according to the movements of the left hand. In this case, the exact position of the user's hand is matched to the position of the camera. If the user raises their hand, the camera view zooms out from the space. In order to zoom into the 3D images, the user lowers their hand.

After the 3D image space is synchronized with the user-centered space, the following request manipulates the objects. Unlike *DuplicateSpace*, the user's bare right hand is used to grasp and release target objects. The gestures of clenching the user's fist and unfolding their palm are related to grasping and releasing an object, respectively. On the display, a virtual hand navigates using the position of the user's right hand. At the object, a grasping gesture changes the navigating virtual hands into the target object. Then, according to the hand movement, the target object changes its position until the releasing gesture is made. Through using bimanual bare-handed gestures, the user can manipulate the objects in the 3D images. Figure 1 illustrates the manipulation of object and space through the 3D interaction platform.

The hand gestures are recognized using a depth camera installed on the upper side of the display. Each hand is tracked using depth information. The hand images are separated from the background subtraction over all depth images. Then, the segmented hand image goes through a thinning process. If there is only one end point in the hand image, the hand is in a folding state. If there are five end points, this indicates that the hand is in an unfolded state.

3. Robotic Teleoperation System

In this study, we constructed a robotic teleoperation system through applying the proposed 3D user interface. In order to implement the teleoperation, we prepared two spaces that were separated: one space for the user and another space for the robot (Fig. 2). It was assumed that the user was not an expert in controlling and ordering the robot to undertake tasks and that the robot had no prior knowledge for doing the ordered tasks. However, the robot could recognize objects and grasp objects from previously learned examples.

In the robot space (i.e. the robot side), the robot was a one-armed robot and the arm had seven degrees of freedom. The robot is capable of grasping and releasing objects at specific positions [10]. The primary task of the robot is to pick up and place objects on the table at certain positions that the user determines. The robot has 3D vision using a Kinect camera. The objects can be recognized from the previously trained object models that were constructed using point clouds [11].

In the user space (i.e. the user side), the user sits in front of a large display that views the remote robot. The user can observe the table through the robot's camera and see the robot's perspective. Through the proposed interface for 3D manipulation, the user conducts a task that is composed of several grasping and releasing actions. After finishing the task, the user transmits the task information to the robot. Then, the robot receives the task order and executes the actions step-by-step. The overall process is depicted in Fig. 3.

The task protocol to transfer the task information from the user to the robot is composed of action lists. From the user side, a user-created task is transferred to the robot with in the form of an action list. An action contains an

Fig. 1. 3D interaction platform for the manipulation of 3D images and augmented virtual objects. The left hand is used to navigate the camera view in the 3D point cloud image. The augmented virtual objects are grasped, moved, and released by the user's right hand.

Fig. 2. System configuration for robotic teleoperation: The left image depicts the one-armed robot that is grasping an object using a previously trained object model. The right image presents the user interface. After transmitting the task information, the user observes the robot execution in real time with 3D images.

object identity and six dimensional positions for the current location and destination. Figure 4 illustrates the detail of the task protocol.

In particular, we enabled the robot to adapt to dynamic environments. While the user is demonstrating a task, the initial environment could be changed. A new object could appear or the object positions could change. In order to adjust to the dynamic environment, the robot recognizes the identity and position of the objects whenever it executes an action. Because the task protocol interprets the destination position in relation to the objects, the robot grasps the target object from the detected position and releases the object in the new destination. Figure 5 presents a demonstration of the task execution in a dynamic environment.

4. Discussion and Conclusion

In this paper, we presented a task-oriented teleoperation system that is mediated by bare-handed interactions using both hands. The proposed teleoperation system has two key contributions. First, the system supports unskilled users, who do not have a detailed understanding of the mechanics and program that controls the robot, to order a remote robot to undertake a task. Users can easily order the remote robot to execute a specific task based on the direct manipulation-based 3D interface. Second, the robot receives orders for human-level tasks that it can adapt to environmental changes during the task execution. These advantages are possible because the teleoperation system

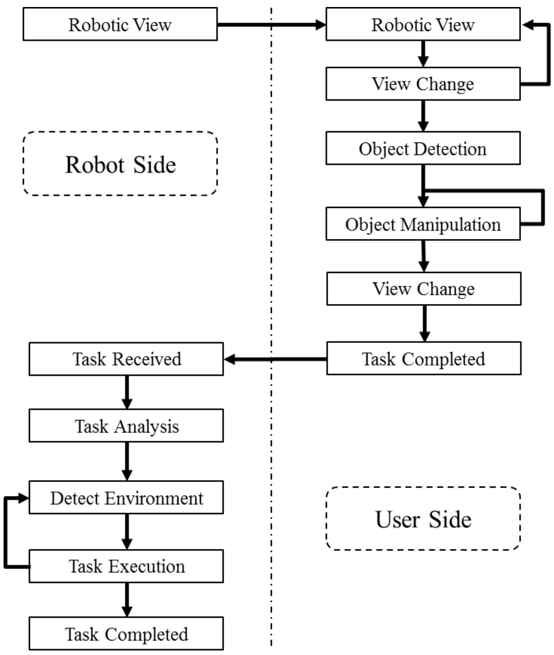

Fig. 3. Process flow of the teleoperation: on the robotic side, the robot provides 3D images in real time and receives the task information to execute according to the user's action list. Both the robot and user sides are asynchronously undertaken with time delays.

Action 1	Target ID	Cur. Pos.	Cur. Rot.	Dest. ID	Dest. Pos.	Dest. Rot.
Action 2	Target ID	Cur. Pos.	Cur. Rot.	Dest. ID	Dest. Pos.	Dest. Rot.
			...			
Action n	Target ID	Cur. Pos.	Cur. Rot.	Dest. ID	Dest. Pos.	Dest. Rot.

Fig. 4. Task protocol: A task is composed of actions, which include the position of the current object and destination. Because the destination contains the identity of the object, pick-and-place tasks can be adapted in dynamic conditions.

is based on task-oriented communication.

Furthermore, the proposed robotic teleoperation is based on asynchronous communication. That is, there is a time interval between the user execution of the task and the robot execution of the task. Unlike haptic-mediated teleoperation, the robot waits until the user sends all task information and then begins the task execution. This prevents the safety concerns that direct control methods cause.

However, the proposed system requires more development. Currently, the demonstrated tasks are limited to pick-and-place tasks because a one-armed robot with a gripper was used. If the robot has advanced hardware, e.g. multiple fingers for object manipulation, the tasks can be extended and become more complicated.

There is also a limitation related to the coverage of task logs. The tasks in the task protocol are composed of basic information such as identity and position. Here, the details of the human demonstration such as the action speed and object condition are not considered. Sometimes, some objects must be moved via a specific path. Some objects with certain conditions should be moved carefully using slow movements. Even though a user may carefully execute a certain task, the robot did not receive information regarding the conditions of the action; thus, the task could be interpreted as wrong. Furthermore, on the user side, the feedback from the remote robot side is constrained to visual information only. However, haptic and auditory feedback could assist the user to understand the current robot status and conditions.

Therefore, we will build task knowledge from the user demonstration via the user interface in our future research. Furthermore, repeated tasks do not need to be directly operated by a user. Instead, the tasks are learned in the task model; then, the robot executes the task from their knowledge model. The task data is recorded from user demonstrations through the proposed interface.

Acknowledgement

This work was supported by a research grant from Korea Institute of Science and Technology (2E24800)

References

[1] T. B. Sheridan, "Teleoperation, telerobotics and telepresence: A progress report," Control Engineering Practice Vol. 3, No. 2, pp. 205-214, 1995.

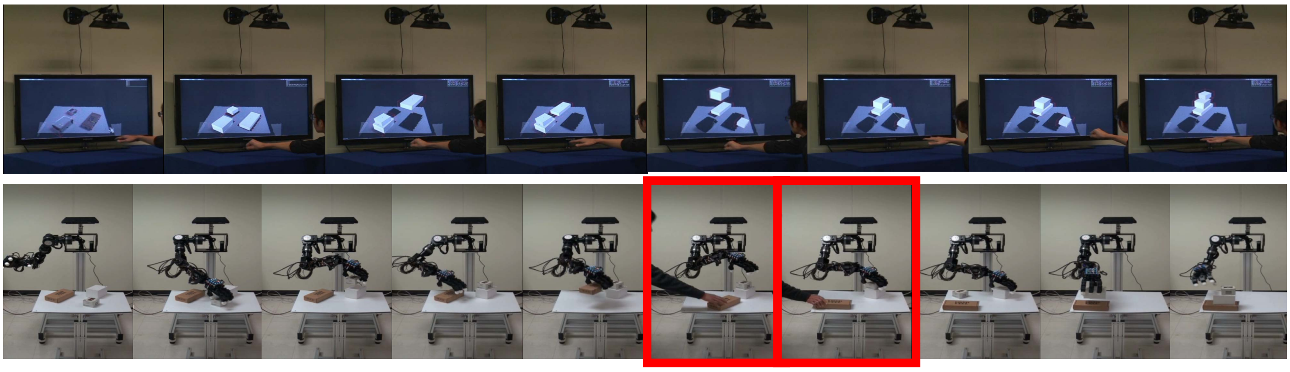

Fig. 5. Robotic teleoperation in dynamic conditions: The upper eight images represent a user-centered task, which is to build the top of three boxes using a 3D interaction platform. The lower ten images illustrate the robot execution. During the task, external interruptions occurred and the destination position changed; this occurred in the steps indicated by red rectangles. In that condition, the robot accomplished the task of building the tops of the blocks.

[2] M. A. Goodrich and A. C. Schultz. "Human-robot interaction: a survey," Foundations and trends in human-computer interaction Vol. 1, No. 3, pp. 203-275, 2007.

[3] S. J. Payne and T. R. Green, "Task-action grammars: A model of the mental representation of task languages," Human-computer interaction, Vol. 2, No. 2, pp. 93-133, 1986.

[4] O. Hilliges, D. Kim, S. Izadi, M. Weiss and A. Wilson, "Holodesk: direct 3d interactions with a situated see-through display," In Proc. of CHI, 2012.

[5] J. Lee, and C. Boulanger, "Direct, spatial, and dexterous interaction with see-through 3d desktop," In Proc. of SIGGRAPH, 2012.

[6] H. Benko, R. Jota and A. Wilson, "Miragetable: freehand interaction on a projected augmented reality tabletop," In Proc. of CHI, 2012.

[7] H. Kim and J. H. Park, "DuplicateSpace: Enhancing Operability of Virtual 3D Objects by Asymmetric Bimanual Interaction," URAI, 2014.

[8] H. Kim, D. Yoon and J. H. Park, "User-centered space registration for 3D remote interaction," In Proc. of APCHI, 2012.

[9] D. Yoon, J. Lee, K. Yeom and J. H. Park, "Mobiature: 3d model manipulation technique for large displays using mobile devices," In Proc. of ICCE, 2011.

[10] S. Cheon, K. Ryu and Y. Oh, "Object Manipulation using Robot Arm-Hand System," In Proc. of URAI, 2013.

[11] J. S. Kim, "Object Detection using RGBD Data for Interactive Robotic Manipulation," URAI, 2014.