Physical Telepresence: Shape Capture and Display for Embodied, Computer-mediated Remote Collaboration

Citation	Leithinger, Daniel, Sean Follmer, Alex Olwal, and Hiroshi Ishii. "Physical Telepresence." Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology - UIST '14 (2014).
As Published	http://dx.doi.org/10.1145/2642918.2647377
Publisher	Association for Computing Machinery (ACM)
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/109197
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/4.0/

Physical Telepresence: Shape Capture and Display for Embodied, Computer-mediated Remote Collaboration

Daniel Leithinger* **Sean Follmer*** **Alex Olwal**[†] **Hiroshi Ishii**
MIT Media Lab
Cambridge, MA, USA
{daniell, sean, olwal, ishii}@media.mit.edu

ABSTRACT

We propose a new approach to Physical Telepresence, based on shared workspaces with the ability to capture and remotely render the shapes of people and objects. In this paper, we describe the concept of shape transmission, and propose interaction techniques to manipulate remote physical objects and physical renderings of shared digital content. We investigate how the representation of user's body parts can be altered to amplify their capabilities for teleoperation. We also describe the details of building and testing prototype Physical Telepresence workspaces based on shape displays. A preliminary evaluation shows how users are able to manipulate remote objects, and we report on our observations of several different manipulation techniques that highlight the expressive nature of our system.

Categories and Subject Descriptors: H.5.2 User Interfaces: Haptic I/O, Interaction Styles

General Terms: Design, Human Factors

Keywords: Shape-Changing User Interfaces; Shape Displays; Actuated Tangible Interfaces; Teleoperation; Physical Telepresence

INTRODUCTION

As the world becomes increasingly connected, collaborative work is often distributed across multiple locations. A large number of commercial products are available to support video-mediated communication and shared digital documents, but many limitations for telepresence and telecollaboration remain. Much co-located collaborative work focuses on physical objects, surfaces, spaces and interpersonal relationships that rely on physical interaction with other participants, such as a handshake. The affordances of the physical environment and presence of co-located collaborators are often missing in screen-based remote collaboration.

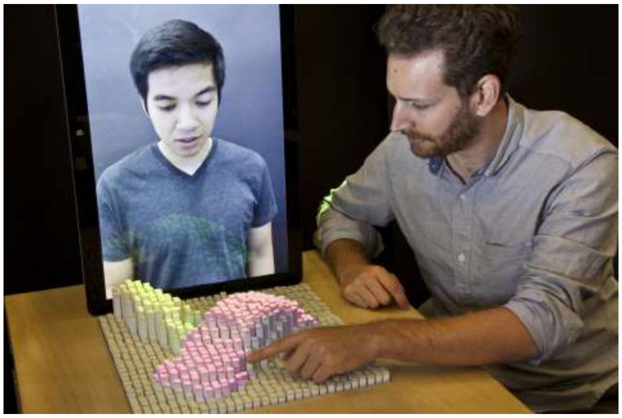

Figure 1. Physical Telepresence provides physical embodiment, remote manipulation and new capabilities through computer-mediated teleoperation. Here, local and remote users physically interact with a 3D car model.

Researchers in distributed computer-supported collaborative work (CSCW) propose improving video-mediated collaboration by situating it in the physical world through the use of shared media spaces [30] and projected Augmented Reality (AR) [33]. However, these interfaces still lack many physical aspects of collaboration, as remote participants are only visually present on the screen, which limits their ability to collaborate through physical objects. To overcome these challenges, telepresence robots have been proposed to embody remote participants for social presence and to manipulate the physical world from a distance [13].

A different approach to physical remote collaboration is presented by remote Tangible User Interfaces (TUIs), which focus on *synchronized distributed physical objects* [2]. These physical objects, or tokens, are synchronized with a remote counterpart to represent shared content, rather than embodying collaborators.

By introducing *Physical Telepresence*, one of our goals is to extend the physical embodiment of remote participants, which is common in telepresence robotics, and combine it with the physical embodiment of shared content, common in remote TUIs. An example of such a system is shown in Figure 1, where the hands of a remote collaborator along with a shared digital model are materialized on a shape display.

In this paper, we investigate different ways in which shape displays [23, 7] can physically embody remote people and

---

* The first two authors contributed equally to this work.

[†] Secondary affiliations: KTH – Royal Institute of Technology (Stockholm, Sweden); Google [x] (Mountain View, CA)

UIST '14, October 05 - 08 2014, Honolulu, HI, USA.

ACM 978-1-4503-3069-5/14/10$15.00.
http://dx.doi.org/10.1145/2642918.2647377

objects to enable communication and collaboration. The flexibility of shape rendering also allows us to loosen the direct 1:1 link between remote shapes, both in terms of rendering users and data. We introduce computer-mediation for physical telepresence and explore how shape displays enable new forms of interactions that overcome the user's physical limitations. We can, for example, replace a remote user's arm with custom end effectors that are rendered on demand, as opposed to being limited to the static, prebuilt end effectors of a telepresence robot. We also propose interaction techniques that allow users to manipulate remote objects through gestures, mediating objects, tools, and direct touch. Finally, we report on a preliminary evaluation to explore potential use patterns in telemanipulation scenarios.

CONTRIBUTIONS

- An exploration of physical telepresence for shared workspaces, using shape capture and rendering for rich, real-time physical embodiment and manipulation.
- Interaction techniques for manipulation of remote physical objects and shared, physically rendered, digital models.
- Interaction techniques that leverage physical computer-mediation to amplify user capabilities in remote operation.
- A technical software and hardware platform, which enables synchronous physical interaction for remote users, physical objects, and shape deformation.

RELATED WORK

Video-Mediated Communication

Several early research projects investigate video-mediated collaboration for shared workspaces [31, 30, 11]. More recently, these techniques have been applied to applications, such as collaborative website development [4], remote board games [33] and family communications [34].

Mixed Reality

Collaborative tools can also leverage more spatial interaction and mixed reality displays. TelePointer [17] allows an expert to point in a remote user's real environment through a user-worn laser pointer, while more advanced Spatial AR techniques have been proposed for future collaborative workspaces [24, 1]. Handheld AR can support mobile scenarios [28] whereas larger, situated displays have other benefits, such as immersion [9].

Telerobotics

A body of research focuses on representing remote people with telepresence robots (e.g., [32] [21]). The Personal Roving Presence concept [21] was an early exploration into tele-embodiment using, e.g., screen-based robots and flying blimps. Research has also explored such devices' role in the workplace, and their influence on the sense of social presence [13]. Our approach of rendering captured geometry without semantic knowledge of user input, is quite different from common telepresence robots. Our interactions focus on leveraging the shape display for a rich, shared workspace, with less emphasis on mobility.

Another related domain of telerobotics is telemanipulation. Early master-slave manipulators used mechanical linkages, and now robotics arms and end effectors, to remotely handle hazardous materials or other objects at a distance [20]. A more specific example is that of telesurgery robots, which not only allow for remote operation, but give surgeons more precise control and remove tremors [29]. In contrast to prior work on telemanipulation, we focus on supporting collaborative work and the ability to switch between arbitrary end effectors on demand.

Tangible Remote Collaboration

TUIs can also be beneficial for remote collaboration. PsyBench [2] actuated a tangible object to mirror the movement of a remote physical token. Actuated Workbench [19] added more advanced control over multiple tokens, and represented remote user's presence through projected digital shadows and haptic feedback. Researchers have also explored small tabletop robots for remote TUI collaboration [26, 27]. Video portals have been used to play games over a distance, where physical pucks appearing from underneath the video create the illusion of a single shared object [18]. TUIs for remote collaboration often utilize actuated physical objects to represent content, rather than collaborators [2, 19, 26, 27]. However, as remote actors themselves are not physically embodied, object movement can result in a disconnected experience, since graphics can only partially simulate the presence of a co-located collaborator.

Shape Displays for Remote Collaboration

Since Project Feelex [12], various form factors for 2.5D shape displays have been developed for co-located collaboration and interaction [15, 23, 7, 22]. While Lumen discussed the potential for remote presence and demonstrates an example application [23], remote collaboration has been less of a focus for past shape display research [25].

PHYSICAL TELEPRESENCE FOR SHARED SPACES

We introduce the concept of Physical Telepresence for shared workspaces, where capturing and rendering shapes has the potential to increase the sense of presence, expand the interaction bandwidth, and to extend human capabilities with computer-mediated interaction techniques.

Physical Telepresence is especially relevant for domains where physical presence, spatial information display and rich interaction is required. We primarily emphasize the potential in interactions with shared digital models of arbitrary shapes, linked physical objects, and manipulation of remote objects through direct gesture.

We explore these interactions through three implemented systems:

- Bi-directional interaction through deformation on two linked shape displays (see Figure 15).
- Bi-directional interaction through shape capture and rendering: split-view on a single shape display (see Figure 3).
- Asymmetric teleoperation: A shape display linked to a video-mediated environment with shape capture (see Figure 14).

Figure 2. A user manipulating a sphere through his remote physical embodiment.

Example Scenario: Remote 3D design collaboration

Bill and Jean are two designers working on a new car body design at different locations using their respective Physical Telepresence systems. Bill sees Jean's face on a vertical screen and her hands physically rendered on a horizontal shape display; Jean has a similar view of Bill. As Jean moves her hands, Bill's shape display adapts to update the hands' physical form.

Bill opens their shared car design model, which appears as a physical rendering. To bring the headlights in view, Jean gestures above it on her shape display, rotating and scaling the car, which updates the car model in real-time for Bill. When Bill deforms the headlights with his hands, these changes in the shared model are propagated to Jean's display.

After discussing the changes, Bill moves the car to the side, and reaches for a headlamp that he 3D printed earlier. When placing it on the surface, a physical rendering appears on Jean's display. Jean rotates the rendering, which also rotates the 3D print on Bill's side. They make some digital annotations, which are projected over the physical part and the physical rendering. After reviewing necessary changes, they part ways, and Bill 3D prints the updated model.

Remote Physical Embodiment

The ability to provide relevant and adequate representations of remote participants and environments, which otherwise would have been co-located, is one of the grand challenges of telepresence systems. This embodiment has a great influence on realism, interaction capabilities and sense of presence.

In a shape-display-mediated telepresence scenario, we identify four primary types of content to represent:

Users. Tracked remote and local users, with their representation transmitted to support interaction and telepresence.

Environment. The static or dynamic physical environment at each location, where the interaction takes place. We introduce a number of novel possibilities for computational control and actuation of the local and remote environment.

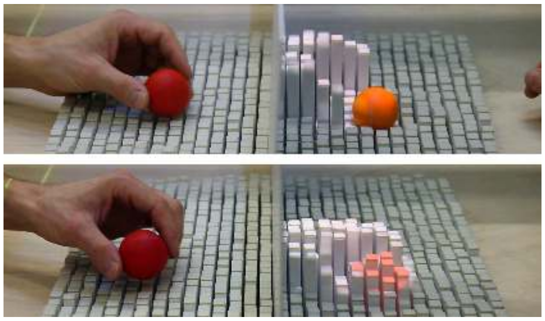

Figure 3. Top: Two linked tangible tokens at two locations, with synchronized position. Bottom: A tangible token rendered with the shape display at a remote location.

Objects and Tools. Tangible artifacts can be sensed in the environment, allowing them both to be manipulated by the system, or by a remote user through the shape display, in addition to direct manipulation by a local user.

Shared Digital Models. We can use a shape display to represent shared objects in multiple locations. Changes are updated in the computational model and propagated to remote locations. Shared Digital Models may have different viewports in different locations.

Representations: Physical and Virtual

Various techniques exist to capture the appearance, geometry and interactions from local participants and objects, which the system can use to form a representation that can be transmitted and materialized at a remote location.

Physical Shape Rendering. The physical shape of remote users, objects, and environment can be dynamically rendered through actuation on a shape display.

Linked Tangible Objects. Passive tangible objects, or tokens, placed on the surface of the table can represent remote objects. These passive objects can be manipulated through shape actuation [7]. This makes it possible to go beyond the limited degrees-of-freedom of the shape display by using external objects on its surface (see Figure 3, top, right).

2D Graphics. Different displays can be used to add graphics to the physical output. Previous work has explored integrating LED lights for color and illumination into shape display pins [23]. To overcome the low resolution of shape displays, high-resolution visual content can also be projected onto the shape rendering. This makes it possible to synchronize the 3D capture of geometry with a video stream that is projection-mapped onto a dynamic, physical surface topology.

Augmented Reality. With Sublimate, we showed how AR can add registered virtual content that extends beyond the range of the shape display [14]. This allows content that is floating in mid-air, has variable opacity, uses complex shapes, or has other properties that are not physically feasible or practical using current shape display technology.

Complimentary Displays. Other communication tools can be added to complement shape rendering, such as a vertical screen that give users face-to-face video. The video can, e.g.,

Figure 4. Improvising with physical objects: Using a basket to scoop up remote balls.

be aligned with the shape rendering to provide the illusion that the physical hands are coming out of the screen.

Switching between Representations. It may be advantageous to switch between these representations. A tangible token, e.g., could be represented by a remote tangible object, but if the remote user does not have enough tokens a physical rendering can be used instead (see Figure 3). Hands can be represented as physical renderings when manipulating an object, or as a virtual rendering when pointing or annotating (see Figure 7). These state transitions are in part inspired by our Sublimate concept [14].

PHYSICALLY MANIPULATING REMOTE OBJECTS

Prior work [7, 5] has shown how shape displays can be used to manipulate physical objects, e.g., using translation and rotation. Here, we explore different control schemes to allow a user to manipulate remote objects. Using shape capture and display, users can reach through the network and pick up a remote physical object. These interaction techniques were developed through iterative prototyping with our inFORM system. Hundreds of visitors to our lab tried different techniques to manipulate objects remotely, and their comments and our observations lead us to develop and improve the following techniques.

Gesture Control

Direct Gesture Control allows a user to interact directly with a remote tangible object through transmitted physical embodiment, which is rendered on the remote shape display. The rendered shape of the user directly applies a force to objects placed on the surface. For example, if the user forms a cup with the hands and raises them, this will be rendered and the pins could move a ball upwards. By observing the reaction of the physical object to their transmitted gestures, users can improvise to expressively manipulate a variety of objects. While our shape displays currently apply vertical forces, objects can still be translated laterally by tilting and sliding [7]. In addition to users' hands and arms, any object that is placed in the shape capture area can be transmitted and used to manipulate remote objects (see Figure 4). Because our shape display, based on inFORM [7], can only render shapes 0-100 mm in height, there is a question of how to render the remote environment, which most likely extends beyond 100 mm in height. We explored two mappings: Scaled and 1:1 with gesture zones. The scaled mapping takes all depth data and maps its height values to the shape display's maximum extent. The second mapping, 1:1 with gesture zones, take some portion of the input space that is the same height as the maximum height travel and renders it on the shape display. This can be directly above the surface, or mid air, allowing users to touch the pin array without manipulating remote objects. In our teleoperation scenario we use a 1:1 mapping, with the gesture zone of physical rendering starting right above the horizontal screen.

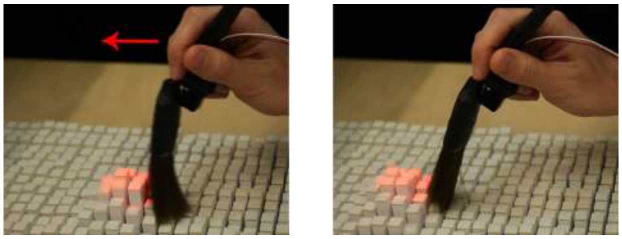

Figure 5. Pushing against the side of a shape-rendered objects with the brush tool moves it.

Mediated Gesture Control exploits that the remote object's pose is tracked and can be updated and moved to keep it synchronized with its underlying digital model [7]. *Physics-based Gestures* detect the user's collisions with the model, to update and move it, using a physics library. The updated model then causes the remote object to be physically moved. This is similar to the proxy-based approach in HoloDesk [10], but with actuated physical output. *Iconic Gestures* provide users with more abstract control. The user can pinch over the representation of the remote object to grab it, move their arm to another location, and open the pinch gesture to release it. The remote object is then actuated by the system to move to that location.

Interface Elements

Interface elements, such as virtual menus, can be projected around the rendering of a remote object to provide access to different operations. *Dynamically rendered physical affordances* [7], such as buttons or handles, that appear around the remote objects can be used for control. The user could press, push or pull such affordances to move the object.

Tangibles and Physical Tools

Tangible Tokens can be used to control a remote object. As the user moves the token, the model of the remote object is updated, and the remote object is moved to reflect the changed state. Two tangible tokens can be linked such that moving one causes the other to move, and vice versa, allowing for bi-directional control [19].

Tools allow users to manipulate remote objects by interacting with the local physical rendering. Tools can provide additional degrees of freedom (DOFs) of input, when interacting with the rendering. Our brush tool, for example, allows users to push remote objects (see Figure 5). The bristles of the brush serve two purposes. First, they decrease the friction between the tool and the pins, which may get stuck when a lateral force is applied. Second, they smooth the haptic feedback resulting from the discrete, jerky motion when a physically rendered object is translated on a limited resolution shape display. To determine the direction in which to move the object, the brush tool's pose is tracked by an overhead camera, while a mechanical switch senses pressure.

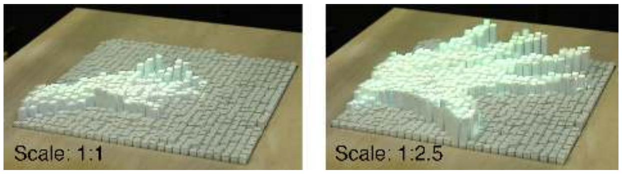

Figure 6. Scaling a users hand to interact with larger objects.

Figure 7. Copying and rendering multiple hands simultaneously (left) and switching to a virtual, non-physical, representation (right).

COMPUTER-MEDIATED SHAPE I/O

While shape capture and output literally adds a physical dimension to telecommunication, we find it interesting to go beyond representations that are symmetric in time and space. How can we create interfaces that enable remote participants to go beyond physically being there? The relaxation of 1:1 capture/output mapping opens up significant potential for new interaction techniques. Consider, e.g., an audio system with real-time translation of a speaker's voice to the listener's language. Here, the mediation emphasizes the relevant and simplifies the interaction. The introduction of shape output allows us to apply similar transformation concepts to the physical domain. We explore shape mediation, such as transformation of physical form, altered representation, data filtering and replication, changing motion dynamics and time-domain manipulations.

Transformation of Physical Form: Bending Body Limits

In our system, users can apply transformations to their activity in the remote environment, for example with scaling, translation, rotation, shearing, stretching and other distortions. Translation offsets geometry and can extend reach, with potential ergonomic benefits. Scaling can make a hand larger or smaller for manipulation of objects of varying sizes (see Figure 6). A small hand could avoid undesirable collisions in dense topologies, while an enlarged hand could carry multiple items. The transformations allow continuous real-time changes during the interaction, e.g., enabling smooth changes in size or position while holding an object. Examples of other transformations include replication or mirroring, e.g., to approach objects from multiple angles.

Altered Representation: Becoming Something Else

With user and object tracking, there are benefits to switching representation for new capabilities. A system that captures geometry does not need to propagate all of it. It may be useful to just send a user's hand and not the arm. As we are only limited by what the shape display can render, we can also morph into other tools that are optimal for the task, while controlled by the user. Examples include grippers, bowls, ramps, and claws — tools with specific properties that facilitate or constrain the interactions (see Figure 8 and 9). The tools could also be animated or semi-autonomously use the sensed geometry on the remote side to influence their behavior. Switching to a purely graphical representation to avoid collisions, is another example (see Figure 7).

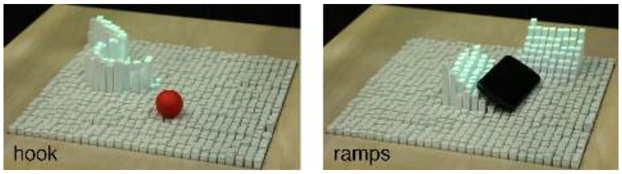

Figure 8. Replacing hands with a hook to reach or ramps to slide objects.

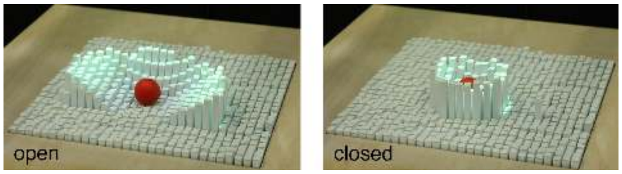

Figure 9. The claw tool open and closed to enclose and move an object.

Filtering: Adding and Removing Motion

Signal processing can be applied to refine propagated motion, e.g., using smoothing or low-pass/high-pass filters. Such approaches are in use, e.g., in surgical robotics where hand tremors can be suppressed. Our system could also prevent fast motion or access to protected areas to avoid involuntary movements. In addition to reducing human noise, it may also alleviate system limitations, such as sampling resolution, speed, range and vibrations.

Motion Dynamics: Warping Time

Non-linear mapping of the propagated motion is interesting for many interactions. The properties of certain remote artifacts might, e.g., require slower or faster mapped motion, or require brief freezing or slow-down to emphasize an effect or make it legible. Such manipulations of time need, however, to be designed with great care, as they break the temporal link between the remote locations.

PROTOTYPE APPLICATIONS

Telepresence Workspace

When discussing a physical design over distance, it is important for both parties to have an understanding of a shared model. We propose to render physical models on shape displays during remote collaboration meetings. The shape output is combined with video for viewing the upper body and face of remote collaborators. By aligning the tabletop shape display with the vertical screen, the two collaborators perceive a shared physical workspace, where the remote person can reach out of the vertical screen to physically point at a model. We support collaboration through shape rendering in several ways:

Shared Digital Model. The model is mirrored on the remote shape displays and provides a shared frame of reference.

Transmitted Physical Model. When a user places a physical model onto the surface, its shape and texture is transmitted to the remote site.

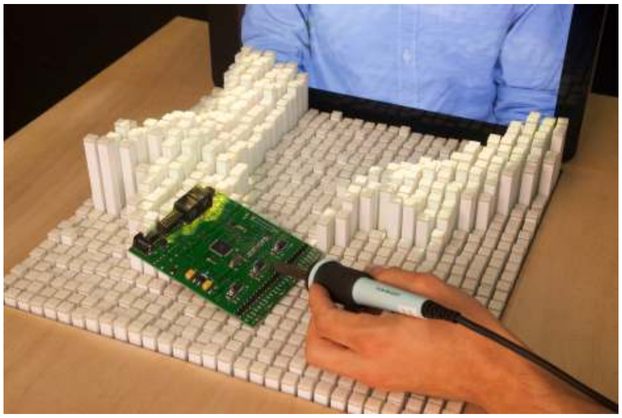

Figure 10. Remote assistance: A remote expert helps a local user by supporting the workpiece with one hand and pointing with the other.

Physically-rendered Remote Collaborators. When a user reaches out and places a hand on the table, its shape is physically rendered at the remote site. This conveys presence and enables remote pointing and manipulation. Figure 10 shows an example application, where a remote expert supports a local user.

Unidirectional Shape Output

The shape output link described in the above scenarios can be uni- or bidirectional. An ideal bidirectional configuration, where both sites have a shape display, may not always be feasible due to size, cost and required infrastructure. A unidirectional link is still advantageous, as it allows remote participants to be more present with physically rendered body representations. It makes it possible to also capture and transmit physical objects to the remote site.

Collaborative 3D Modeling

A specific example of a telepresence workspace with a shared digital model is collaborative 3D modeling. In this case, users can deform the model, with any change reflected through the model on the connected site (see Figure 11). However, rendering the remote person's hands may obstruct the changes to the model. Therefore, our application allows switching to rendering the model geometry without the user's hands.

Another consideration is the potential conflict that may occur when two users try to simultaneously deform a model. We therefore use a turn-taking protocol for manipulation. When the system detects deformation, control is passed to the active user. The changes will be reflected on the remote connected model, while concurrent remote input is ignored. When the user stops modifying the model, the collaborator at the remote site may take control by starting a new deformation. Passing control between users can be applied to the whole model or to each pin individually.

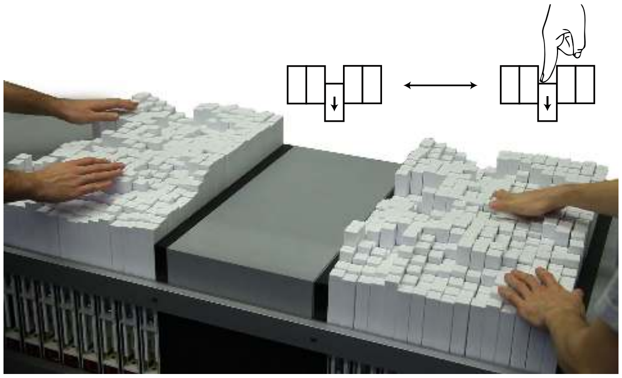

Figure 11. Collaborative 3D modeling of two linked landscape models.

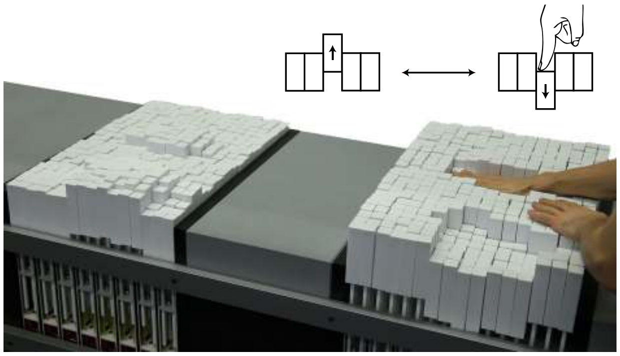

Figure 12. Connected Membrane: When a user pushes into the membrane on one site, the inverse shape appears at the other.

Shared Presence: Connected Membranes

This application conveys the presence of remote participants through a shared digital membrane that uses two connected shape displays, each rendering the inverse shape of the other (see Figure 12). When a users pushes down a pin, the corresponding remote pin pops out. This is conceptually similar to a physical link between the pins. We implemented two modes:

Pin screen. Pins remain in their position after users push or pull them. This leaves behind a trace of past deformations, until it is erased.

Elastic. Pins always spring back to a neutral flat state when they are not being deformed. This mode conveys whether someone on the other end is pushing at that very moment.

TECHNICAL IMPLEMENTATION

Shape Output

Our proposed systems utilize 2.5D shape displays, which consist of arrays of linear actuators.

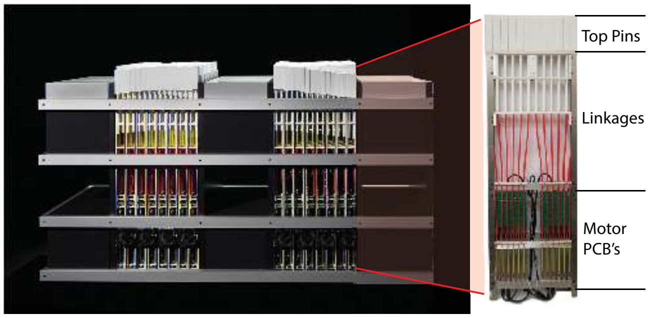

Figure 13. Side view of the TRANSFORM system with a single 12x2 actuation module.

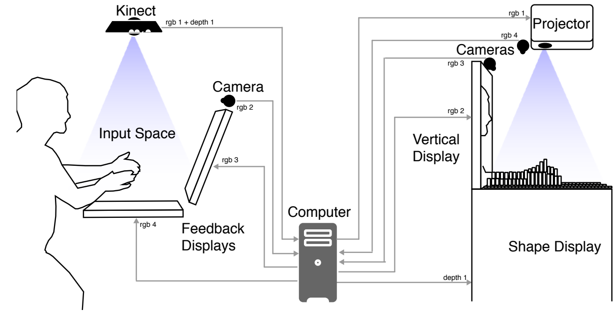

Figure 14. System diagram of the telemanipulation system using shape capture and remore shape display.

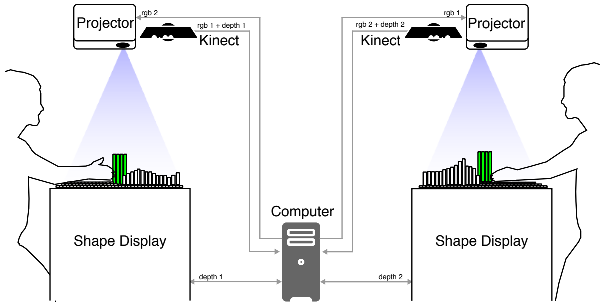

Figure 15. System diagram of connected workspaces with bidirectional shape capture and shape rendering.

The *inFORM* platform [7] consists of 30×30 individually actuated, plastic pins, built into a tabletop. The pins cover an area of 381×381 mm with each pin measuring 9.5×9.5 mm with 3.175 mm spacing between pins, and they can extend up to 100 mm in height. Actuation speed is 0.644 m/s and each pin can exert up to 1.08 Newtons.

TRANSFORM is a platform based on actuator modules with 2×12 styrene pins on an area of 305×50.8 mm, which extend up to 100 mm from the surface (see Figure 13). The individual actuators and control boards are the same as used by inFORM. The modules can be seamlessly combined to form a larger shape display surface. We created two shape displays, 16×24 pins each, covering an area of 406.4×610 mm.

The shape display hardware uses custom Arduino boards that run a PID controller to sense and move the positions of 6 connected styrene pins through motorized slide potentiometers. Our applications are written in C++/OpenFrameworks and support 3D models (OBJ) and grayscale images as contents. The application renders a depth image, which is sent to the shape display over USB to RS485. Graphics from OpenGL are projected onto the white shape display pins through a calibrated ceiling-mounted projector (see [7] for details).

Sensing Shape Deformation Input

The pin positions that are reported from the shape display to the computer can be used to detect when a person deforms the surface. By comparing the shape image sent to the display and the position image received from the display, any deformation will be visible as a discrepancy. As the physical shape rendering contains some noise and the motors require a certain time to reach their target position, our detection algorithm thresholds the deformation image and compensates for the time delay. This filtering is critical when sharing the depth information with a connected system, as the network delay and added noise can result in a feedback loop that results in false deformation detection.

Tracking People and Objects With a Depth Camera

We track objects and gestures with a depth camera (first-generation Microsoft Kinect, 640×480@30 Hz). The 8-bit depth image is cropped, scaled and thresholded to fit to the size of the interaction area. The values are then normalized to match the range of the shape display. Our algorithm analyses the depth image to differentiate between the users' hands, shared tangible objects and arbitrary geometry. Hands are detected contours that touch the image boundary. After detecting the tip of the hand, its shape can be replaced by a tool. Shared tangible objects are colored spheres, which are tracked through their size and color.

Latency and Manipulating Objects

The latency of our telemanipulation system, see Figure 14, is between 150 and 200 ms, depending on the amount and direction of travel for the pins. This is inline with latency used for other types of telemanipulation systems [16] and its impact is lessened by the lack of haptic feedback to the remote user. The latency is caused by a number of factors, such as the latency of the depth camera, depth image processing, communication bus, motor speed, and the camera and video latency for displaying to the remote user. Latency causes problems for keeping remote connected objects in sync, as shown in Figure 3. In order to adress this, we track the current position of the local object and move it at a controlled rate to the updated target position (dictated by the remote object).

PRELIMINARY EVALUATION

Demonstrations at our lab of the teleoperation system to hundreds of users provided us with many interesting informal insights. Although the limited degrees of freedom of the system do not perfectly map to operations such as real-world grasping, lifting, and moving of objects, it does still provide meaningful interaction due to its real-world physics. We wished to examine this further in a controlled, qualitative study, to get a more nuanced picture of remote object manipulation strategies and interaction techniques.

In this preliminary evaluation we focus on teleoperation, since we believe that our techniques for remote physical manipulation introduce and enable significant new capabilities. We were particularely interested in the gestures that participants would use.

Apparatus

We used our telemanipulation setup (see Figure 14), where participants were in the same room as the shape display, but visually separated by a divider (see Figure 16). The remote environment is shown with an orthographic 1:1 top-down view on a horizontal display, and a perspective view on a

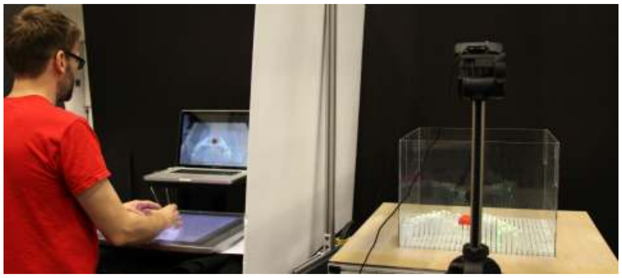

Figure 16. Setup of the telemanipulation study. Left: The participant's gestures are captured and two screens provide visual feedback. Right: Shape output moves the target ball.

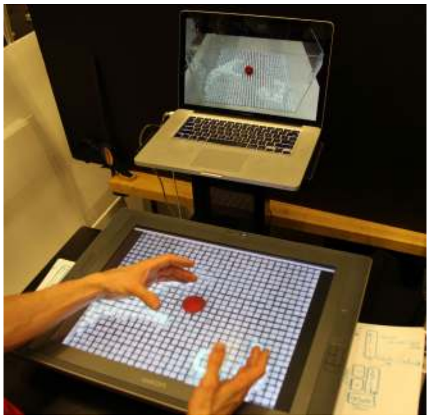

Figure 17. Participant's view of the remote scene, with a top down 1:1 view on the horizontal screen, and a perspective view on the vertical.

vertical display. An overhead, depth-sensing camera tracks gestures above the horizontal display (see Figure 17).

Participants

We recruited 8 participants (2 female), ages 21–37 years old from our institution, who were compensated with 15 USD. All were daily computer users, 5 were monthly users of depth-sensing devices (e.g., Microsoft Kinect). Sessions lasted 45–60 min.

Task, Conditions and Procedure

The three tasks required the participant to move a remote wooden sphere (⌀4 cm) to a target location (⌀5 cm). Users were not required to hold the ball in the target, only to touch (overlap greater than 50%) or cross it.

1. *2D docking*. Randomized target position projected on the remote surface, and 30 cm away from previous target for consistent task difficulty.
2. *3D docking*. Randomized target position as before, but at a height between 40–80 mm.
3. *Obstacle*. Participants were instructed to not knock over the obstacle (water bottle, ⌀7 cm, 20 cm in height) that was placed at the center of the surface, in a variation of the 2D docking task.

There were four interaction technique conditions:

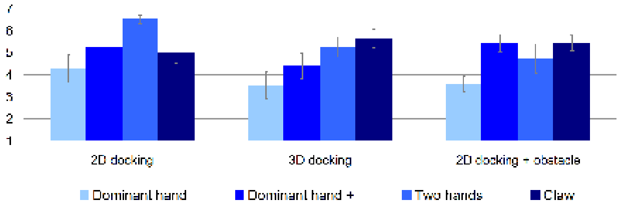

Figure 18. Averages of users' ratings of task difficulty for 4 different input conditions in each task. (7-point likert scale, 7 is best/easiest). Error Bars are standard error of the mean (SEM).

1. *Dominant hand*.
2. *Dominant hand+*, which could be scaled, rotated or switched to virtual-only, by button press with the non-dominant hand.
3. *Two hands*.
4. *Claw* representation of dominant hand. Non-dominant hand could press a button to open and close the claw.

Participants had 5–10 minutes to familiarize themselves with the 4 different interaction techniques. Conditions were randomized and counterbalanced across participants. After each task, participants rated task difficulty in each condition and provided explanations. A post-test questionnaire had the participants provide additional comments. Video was recorded and reviewed for interaction patterns.

Results and Discussion

All users successfully completed all tasks. While some users noticed the system latency of 200 ms, they did not comment that it impeded them. However, we observed that all users adjusted their gesture speed when moving objects to compensate for latency effects. Because of our interest in emerging use patterns and a small sample size, we focus on reporting qualitative data and feedback.

The ratings (1–7, 7 is best/easiest) indicate that different techniques were preferred in different scenarios (Figure 18). Participants preferred using two hands in 2D docking, using the claw for 3D docking, and using one hand that could be transformed or the claw in the third, obstacle-avoiding task. The scores matched our observations and post-test interviews.

Qualitative user feedback

Two hands for the *2D task* was intuitive and similar to how one would move a real ball. *"Moving the ball with two hands was a much more natural interaction for me"* Participants also used hand gestures similar to how they would in real life. *"When I had to claw the ball, I could psychologically and physiologically feel the pressure of grabbing it, although it was virtual."*

Using two hands was the most effective strategy in 2D, since participants could better enclose the ball, stop it, or move it. There were, however, other strategies: *"The biggest challenge was whether the cupping object (hand or claw) would sufficiently enclose the ball to prevent it from rolling away while moving it. The claw, scaled hand, and two hands were all large enough to easily move the ball, but the scale control was the only mode that allowed for tuning the size."*

In the *3D docking task*, raising the ball without having it fall out of the remote hands, was a challenge for many participants. The claw tool was more popular here, as it constrained the ball well during 3D movement.

Real-world collaborative workspaces tend to have many objects in use. We were interested in how users would handle additional obstacles, without haptic feedback. As expected, participants found the *obstacle task* the most challenging, while also the most realistic. *"Two hands were surprisingly frustrating in comparison to the prior tasks because they take up more space"*. But participants were able to use a variety of different representations to more easily complete the tasks.

The claw's small footprint simplified navigation around objects. "*It enabled me to move the ball without having my arm get in the way. Rotating the view while using one hand also worked fairly well because it enabled me to adjust my trajectory to avoid the obstacle"*

Other participants preferred the invisible hand for obstacle avoidance. *"...most useful to go around the obstacles."* But in general this technique was harder for people to discover. *"It took me a while to get a hang of the transparent feature, but it was good because it made my interaction faster, and give it a mouse like feel where I could hover and not click as when you use a mouse."* Participants also commented on picking up objects: *"I also found myself wanting a way to come from below an object – and didnt realize until much later that this is what the ghost/invisible mode enables."*

Observed patterns of use

A number of techniques were used to manipulate the remote object and we outline the most prevalent. To avoid bias, we had not given any instructions on gestures.

- *Push:* Hand (sideways like paddle), back of the hand, index finger, back of the arm, V-shape with fingers (thumb + index).
- *Hook:* Hook shape with hand (bring object closer).
- *Flicking:* Flick object with index finger.
- *Ghost:* Switch to virtual, move inside object, switch to shape to pick up.
- Scoop: Cup hands, scoop and move around. *"Two hands were easier to use when I had my palms close together. "*
- Capture: Approach with two opposing V-shaped hands.

Our overall observation on the gestures emerging during the study was that participants adapted quickly to the degrees of freedom supported by the system and did not try to grasp the object. Instead, everyone interacted as if the remote sphere was a slippery object; pushing it sideways to translate on the surface, and scooping it with cupped hands to move in 3D.

LIMITATIONS AND FUTURE WORK

Many of the limitations of Physical Telepresence are related to the hardware used to implement it.

2.5D rendering using physical pins limits our system's DOFs to rendering reliefs and prevents overhangs. This is especially important for telemanipulation, since it only allows application of vertical forces to surface objects. The system can thus not grasp objects [3], only lift them or translate them by tilting and sliding. Robotic grippers could, however, be combined with shape displays to provide more dexterous manipulation. Shape displays with more DOFs of output can also be explored to allow lateral forces with more intricate control. Latency, as discussed, is another parameter that limits remote manipulation, and further improvements in hardware and interaction techniques could address this.

Limited resolution of current shape displays affect what content can be represented. Due to sampling requirements we need to have a resolution twice that of the smallest feature to be able to clearly display it. We observed this issue when rendering fingers (2 cm wide) on our shape display with 2.54 cm spacing. Increasing the resolution and pin travel will allow for more complex content and more realistic representation of remote users.

Collisions between remote and local objects can affect the possible physical rendering. We implement techniques to address this, such as not physically rendering geometry where users' hands are on the local surface. More compliant shape displays could be built using soft actuators [6, 8]. In addition, users cannot reach inside a rendered object, as is possible with virtual graphics [10]. Our previously introduced Sublimate system, however, provides an approach that combines AR and shape rendering [14].

Network latency was not investigated with our current systems, as these were not deployed as distributed setups. However, the effect of latency is a critical factor for effective remote manipulation, and further studies are required to investigate how it will affect operator performance.

In addition to the limitations of shape display hardware, further work and evaluation is needed to explore new types of interaction techniques to control these different renderings of remote participants.

Beyond these improvements to the system, we would also like to further explore other application domains, such as education, medical, and industrial. We envision that teachers could show remote students how to play instruments, and slow down the playback of their hands so pupils could see more clearly, while doctors could use Physical Telepresence for telepalpation of remote patients.

CONCLUSIONS

We have introduced the space of Physical Telepresence enabled by shape capture and shape display in a shared workspace context. Physical Telepresence allows for the physical embodiment and manipulation of remote users, environments, objects, and shared digital models. By loosening the 1:1 link in shape capture and remote shape output, we have begun to explore how remote users can go beyond physically being there. In addition, we highlighted a number of application domains and example applications that show how Physical Telepresence can be put to use for shared work on 3D models, remote assistance, or communication.

ACKNOWLEDGMENTS

We thank Philipp Schoessler, Mark Feldmeier, James Patten, Jen Lusker, David Doan, Melinda Cross, Clara Rhee and the members of the Tangible Media Group. This research was supported in part by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374 and the Swedish Research Council.

REFERENCES

1. Benko, H., Jota, R., and Wilson, A. MirageTable: freehand interaction on a projected augmented reality tabletop. In *ACM CHI ’12* (2012), 199–208.
2. Brave, S., Ishii, H., and Dahley, A. Tangible interfaces for remote collaboration and communication. In *ACM CSCW ’98* (Nov. 1998), 169–178.
3. Cutkosky, M., and Howe, R. Human grasp choice and robotic grasp analysis. In *Dextrous Robot Hands*, S. Venkataraman and T. Iberall, Eds. Springer New York, 1990, 5–31.
4. Everitt, K. M., Klemmer, S. R., Lee, R., and Landay, J. A. Two worlds apart: bridging the gap between physical and virtual media for distributed design collaboration. In *ACM CHI ’03* (Apr. 2003), 553–560.
5. Festo. WaveHandling:Transporting and sorting in one. In *http://www.festo.com/cms/en_corp/13136.htm* (2013).
6. Follmer, S., Leithinger, D., Olwal, A., Cheng, N., and Ishii, H. Jamming user interfaces: programmable particle stiffness and sensing for malleable and shape-changing devices. In *ACM UIST ’12* (Oct. 2012), 519–528.
7. Follmer, S., Leithinger, D., Olwal, A., Hogge, A., and Ishii, H. inFORM: dynamic physical affordances and constraints through shape and object actuation. In *ACM UIST ’13* (Oct. 2013), 417–426.
8. Genecov, A. M., Stanley, A. A., and Okamura, A. M. Perception of a Haptic Jamming display: Just noticeable differences in stiffness and geometry. In *2014 IEEE Haptics Symposium (HAPTICS)*, IEEE (Feb. 2014), 333–338.
9. Gross, M., Lang, S., Strehlke, K., Moere, A. V., Staadt, O., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., Koller-Meier, E., Svoboda, T., and Van Gool, L. blue-c: a spatially immersive display and 3D video portal for telepresence. *ACM Transactions on Graphics 22*, 3 (July 2003), 819–827.
10. Hilliges, O., Kim, D., Izadi, S., Weiss, M., and Wilson, A. HoloDesk: direct 3d interactions with a situated see-through display. In *ACM CHI ’12* (May 2012), 2421–2430.
11. Ishii, H., and Kobayashi, M. ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In *ACM CHI ’92* (June 1992), 525–532.
12. Iwata, H., Yano, H., Nakaizumi, F., and Kawamura, R. Project FEELEX: adding haptic surface to graphics. In *ACM SIGGRAPH ’01* (Aug. 2001), 469–476.
13. Lee, M. K., and Takayama, L. ”Now, i have a body”: uses and social norms for mobile remote presence in the workplace. In *ACM CHI ’11* (May 2011), 33–42.
14. Leithinger, D., Follmer, S., Olwal, A., Luescher, S., Hogge, A., Lee, J., and Ishii, H. Sublimate: state-changing virtual and physical rendering to augment interaction with shape displays. In *ACM CHI ’13* (Apr. 2013), 1441–1450.
15. Leithinger, D., Lakatos, D., DeVincenzi, A., Blackshaw, M., and Ishii, H. Direct and gestural interaction with relief: a 2.5D shape display. In *ACM UIST ’11* (Oct. 2011), 541–548.
16. Lum, M. J., Rosen, J., Lendvay, T. S., Sinanan, M. N., and Hannaford, B. Effect of time delay on telesurgical performance. In *IEEE ICRA’09* (2009), 4246–4252.
17. Mann, S. Telepointer: Hands-Free Completely Self Contained Wearable Visual Augmented Reality without Headwear and without any Infrastructural Reliance. In *2012 16th International Symposium on Wearable Computers* (Oct. 2000), 177–178.
18. Mueller, F. F., Cole, L., O’Brien, S., and Walmink, W. Airhockey over a distance. In *CHI ’06 extended abstracts* (Apr. 2006), 1133–1138.
19. Pangaro, G., Maynes-Aminzade, D., and Ishii, H. The actuated workbench: computer-controlled actuation in tabletop tangible interfaces. In *ACM UIST ’02* (Oct. 2002), 181–190.
20. Paul, R. P. *Robot manipulators: mathematics, programming, and control: the computer control of robot manipulators*. Richard Paul, 1981.
21. Paulos, E., and Canny, J. PRoP: personal roving presence. In *ACM CHI ’98* (Jan. 1998), 296–303.
22. Piper, B., and Ishii, H. PegBlocks: a learning aid for the elementary classroom. In *ACM CHI ’02 extended abstracts* (Apr. 2002), 686–687.
23. Poupyrev, I., Nashida, T., and Okabe, M. Actuation and Tangible User Interfaces: the Vaucanson Duck, Robots, and Shape Displays. In *ACM TEI ’07* (Feb. 2007), 205–212.
24. Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. The office of the future: a unified approach to image-based modeling and spatially immersive displays. In *ACM SIGGRAPH ’98* (July 1998), 179–188.
25. Rasmussen, M. K., Pedersen, E. W., Petersen, M. G., and Hornbæ k, K. Shape-changing interfaces: a review of the design space and open research questions. In *ACM CHI ’12* (May 2012), 735–744.
26. Richter, J., Thomas, B. H., Sugimoto, M., and Inami, M. Remote active tangible interactions. In *ACM TEI ’07* (Feb. 2007), 39–42.
27. Riedenklau, E., Hermann, T., and Ritter, H. An integrated multi-modal actuated tangible user interface for distributed collaborative planning. In *ACM TEI ’12* (Feb. 2012), 169–174.
28. Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P., and Maciocci, G. BeThere: 3D mobile collaboration with spatial input. In *ACM CHI ’13* (Apr. 2013), 179–188.
29. Staub, C., Ono, K., Mayer, H., Knoll, A., Ulbrich, H., and Bauernschmitt, R. Remote minimally invasive surgery –haptic feedback and selective automation in medical robotics. *Applied Bionics and Biomechanics 8*, 2 (Apr. 2011), 221–236.
30. Tang, J. C., and Minneman, S. VideoWhiteboard: video shadows to support remote collaboration. In *ACM CHI ’91* (Mar. 1991), 315–322.
31. Tang, J. C., and Minneman, S. L. Videodraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems 9*, 2 (Apr. 1991), 170–184.
32. Tsui, K. M., Desai, M., Yanco, H. A., and Uhlik, C. Exploring use cases for telepresence robots. In *ACM/IEEE HRI ’11* (Mar. 2011), 11–18.
33. Wilson, A. D., and Robbins, D. C. Playtogether: Playing games across multiple interactive tabletops. In *IUI Workshop on Tangible Play: Research and Design for Tangible and Tabletop Games* (2007).
34. Yarosh, S., Tang, A., Mokashi, S., and Abowd, G. D. ”almost touching”: parent-child remote communication using the sharetable system. In *ACM CSCW ’13* (Feb. 2013), 181–192.