

# exTouch: Spatially-Aware Embodied Manipulation of Actuated Objects Mediated by Augmented Reality

**Shunichi Kasahara**

Sony Corporation  
MIT Media Lab  
Cambridge, MA, USA  
Shunichi.Kasahara@jp.sony.com

**Ryuma Niiyama**

MIT Media Lab  
Cambridge, MA, USA  
ryuma@media.mit.edu

**Valentin Heun**

MIT Media Lab  
Cambridge, MA, USA  
heun@media.mit.edu

**Hiroshi Ishii**

MIT Media Lab  
Cambridge, MA, USA  
ishii@media.mit.edu

## ABSTRACT

As domestic robots and smart appliances become increasingly common, they require a simple, universal interface to control their motion. Such an interface must support a simple selection of a connected device, highlight its capabilities and allow for an intuitive manipulation. We propose "exTouch", an embodied spatially-aware approach to touch and control devices through an augmented reality mediated mobile interface. The "exTouch" system extends the users touchscreen interactions into the real world by enabling spatial control over the actuated object. When users touch a device shown in live video on the screen, they can change its position and orientation through multi-touch gestures or by physically moving the screen in relation to the controlled object. We demonstrate that the system can be used for applications such as an omnidirectional vehicle, a drone, and moving furniture for reconfigurable room.

## Author Keywords

Tangible interfaces, Direct Manipulation, Embodied Interactions, Augmented Reality, Spatially Aware Interface

## ACM Classification Keywords

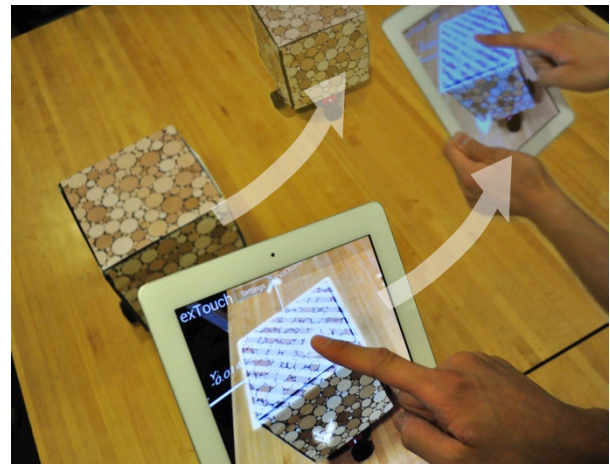
H.5.2 [Information interfaces and presentation]: User Interfaces: Input devices and strategies; H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities; I.3.6 [Methodology and techniques]: Interaction techniques.

## General Terms

Design, Human Factors

## INTRODUCTION

Increasingly, smart appliances are widely used in domestic settings, for instance, networked TV, motorized window shades and digitally controlled lights. Meanwhile, users have to adapt to their complicated operation. Moreover, the growth of robotic technology provides new kinds of products for the living space such as pan/tilt security cameras, robotic vacuum cleaners, and robotic lawn mowers. These products require not only simple on/off



**Figure 1. exTouch, User can simply touch and drag the target device through live video on the screen**

control but also complicated 3D motion control in space. These situations led to invent a "universal remote controller".

A universal controller using a portable computer (often a smart phone or tablet) has benefits compared to conventional switch/button interface. Specifically, it provides unlimited interaction design by software, control of unreachable objects, and secure remote operation. On the other hand, existing universal controllers require undesirable effort for selecting a target device from multiple devices and are not well suited for advanced motion control.

Recent approaches for direct selection of the target device are based on hand gestures, and laser pointers [1]. Another approach for device selection is using a touch screen with camera captured image [2][4]. To provide assistive guidance for the target device with complicated action and function, overlay image based on augmented reality approach is a standard technique [5]. In addition, spatial interaction with virtual objects overlaid on live video is a highly discussed issue [6]. However, developing intuitive interfaces for three-dimensional motion control of physical objects in real life has been a challenge that has not yet been solved elegantly from user's point of view.

To address these issues, we propose the "exTouch" system (Fig.1). Users can control the physical motion of a target

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

TEI 2013, Feb 10-13, 2013, Barcelona, Spain.

Copyright 2013 ACM 978-1-4503-1898-3/13/02....\$15.00.

device by simply touching and dragging through first-person live view video on the screen. The system is based on a portable computer with built-in camera, and does not require external equipment like large-scale motion capture system or ceiling cameras. We demonstrate that the system is applicable for smart devices such as an omnidirectional vehicle, a drone, and moving furniture.

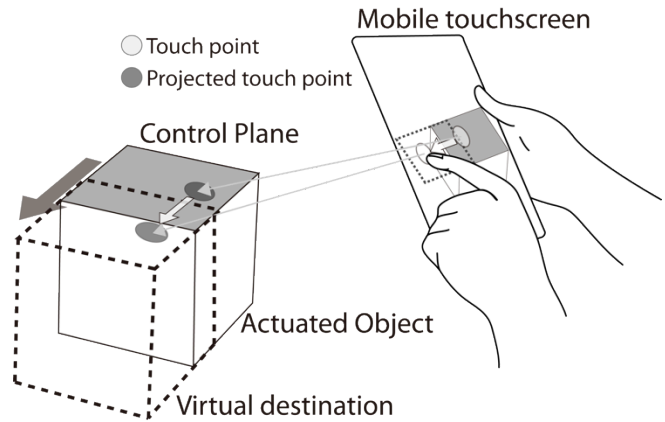
## RELATED WORK

There have been several studies for intuitive interface to control objects in living space. Laser Gesture [1] is a system for operating mobile robots by using a laser pointer and ceiling camera for detecting pointer. It is convenient but not applicable for complex motion and also faces laser safety issues.

Another promising approach uses camera captured images on a touch screen. The idea of controlling devices through a video image was introduced by Tani et al. [2] in the project Object Oriented Video. Users can manipulate digital functions through the captured video. CRISTAL [4] is a system for manipulating home electronics in the room via tabletop multi-touch interface that uses perspective image of the room. For physical motion control, TouchMe [3] is a system for tele-operating a mobile robot through a touch panel. These approaches provide real-time visual feedback and intuitive touch interaction. However, they require the user to consider the spatial relation between miniature world in the monitor and the real world as reading a scale map.

## SPATIALLY-AWARE INTERACTION

We present a concept of "spatially-aware interaction" to transfer the benefits of direct manipulation to the touch screen interface. The spatially-aware interaction satisfies the following features; 1) user-egocentric manipulation, and 2) spatially consistent relation between virtual world and



**Figure 2. Spatial projection of the user touch input into the target motion**

real world. The user-centered manipulation means that the action of target object is always described in the user's coordinate system independent from the object posture. In contrast, conventional gamepad/joystick sends a command based on object's coordinate system. Spatial consistency represents that a camera captured image represents on the control interface always corresponds to the real world, in terms of position, orientation, and scale. For example, when a tourist reads a scale map and walks, the tourist has to consider the conversion of direction, standing, and scale. In the spatially-aware interaction, the virtual world on the screen emulates user's egocentric view in real time.

## EXTOUCH

### User interaction

The exTouch is a system for motion control of actuated objects based on spatially-aware interaction mediated by augmented reality. A user holds the mobile device with a touch screen and back-mounted camera, and points it at the target object. Once the object comes into the range of the

Motion Control					
	Planar Translation	Vertical Translation	Roll rotation	Yaw rotation	Pitch rotation
Screen Drag					
	One finger drag	Pinch in / Out	Two finger rotation	Two finger drag	Two finger drag
Physical Drag					
	Translational shift	Vertical shift	Roll rotation	Yaw rotation	Pitch rotation

**Table 1. exTouch Interaction for motion control. Users can change the position and rotation through multi-touch gestures or by physically moving the mobile device spatially.**

camera, superimposed virtual graphics are rendered on the captured image based on AR approach. Users can control the target objects by dragging the superimposed graphics with multi-touch gestures on the screen or by physically moving the device. Then the physical object moves to match the location of the on-screen graphics. User can see the results through the live video on the screen.

### Implementation overview

The exTouch system employs augmented reality recognition technology to obtain a three-dimensional pose matrix (homogeneous transformation matrix), which describes the position and orientation of the target object in relation to the screen. The system allows spatial projection of the user input into the target motion by using egocentric live image from back-mounted camera and touch screen (Fig. 2). The system performs visual servo control to make the physical object closer to a desired position by using the difference between the recognized object position and dragged virtual graphics position.

### Motion Control Interaction

The fundamental metaphor used in “exTouch” for remote motion control of physical objects is “touch and drag.” Users can change the position and rotation of the target object through two types of touch and drag interaction. For convenience, we call these “screen drag” and “physical drag” to describe these gestural grammar (Table 1).

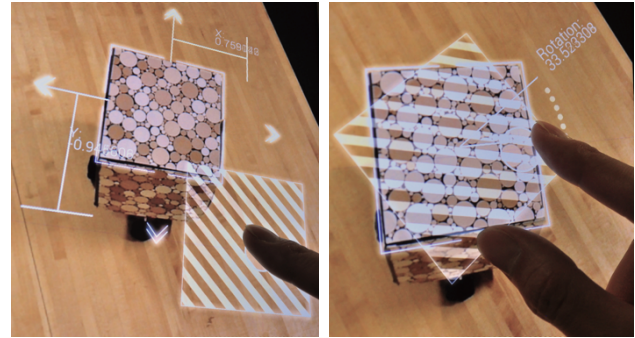
When a user touches the screen, a control plane corresponding to the object movements is determined based on the touch points. Once the control plane is determined, touch points are projected into the recognized coordinate system based on the pose matrix of the control plane and the camera parameters. Finger motion on the touch screen and physical movement of the device while touching both turn out to be a motion of projected point on the control plane. Therefore, both screen drag and physical drag interactions can be used simultaneously.

#### Screen Drag

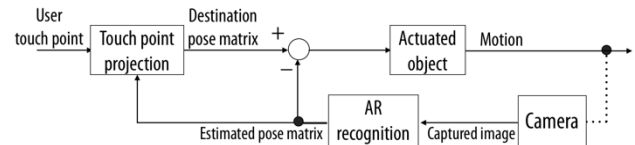
Users can use gestures with one or two fingers on the touch screen. For translational motion, user can control planar motion by one-finger drag, which implies an impression that the finger is generating a force to move the object. For vertical motion user can use two-finger pinch in/out gesture which is based on the metaphor that the object looks larger when it’s closer. For rotation, user can control roll axis by two-finger rotation gesture, which is well known multi-touch gesture in touch screen device, and user can control pitch/yaw axes by two-finger drag gesture. This design resolves the ambitiousness of translation motion and rotation motion by assigning one-finger and two-finger gesture.

#### Physical Drag

Physical movement of the device while touching the screen also results in movements of the projective point on the control plane as well as the screen drag. This feature allows user to control translational and rotational motion by



**Figure 3. Virtual graphics show the virtual destination subject to motion constraints. Left; planar translational motion, Right; one axis rotation.**



**Figure 4. Visual servoing process. Motion control signal for the actuated object is generated from the difference between the recognized object position and dragged virtual graphics position**

physically moving the mobile device while touching the object on the screen. This interaction provides the metaphor that a finger fixes the spatial relationship between the object and the finger.

### Graphic motion affordance

When objects are recognized from the camera image, prompting graphics are overlaid on the camera-captured image in real time. Moreover, once the specific control plane is determined, assistive graphics are rendered to indicate the motion constraints of the object (Fig. 3).

These guide graphics also show the virtual destination of the object in real-time. This allows users to confirm where the target object is going to move in the real world. In our current implementation, when user releases his/her finger, the virtual destination moves back toward the current position and orientation of the object with animated transition.

### Visual Servoing

To achieve a desired position and orientation designated by user, the system performs visual servo control (Fig. 4). The pose difference is calculated from the current pose matrix of the object and the target pose matrix. The pose difference is mapped into control command and sent to the selected object.

### APPLICATION

We implemented the exTouch system with several different actuated objects to understand its real world performance. In order to explore the effectiveness of the exTouch’s interaction design, devices with various degrees of freedom of movements were chosen, omnidirectional vehicle, a drone, and moving furniture.



We employ image based AR recognition technology (Vuforia™ [7]) which recognizes a natural image as a target planar object with registered dictionary data. Our prototype system composed of the mobile application running on an Apple iPad 2, and actuated objects connected via wireless network.

#### Omni directional robot

Wheeled mobile robots are the most popular form of robots in office and home use. We built an omnidirectional mobile robot to simulate practical robots such as cleaner robot, and tele-presence robot. The robot has two degrees of freedoms (DoF) for planar translation and 1 DoF for rotation about a vertical axis (Fig 1).

#### Flying drone control

Control of an unmanned aerial vehicle (UAV), commonly known as a drone, is an emerging field. The hovering flight of a quadcopter drone is suited to carrying instrument such as flying video/photo shooting. Our mini quadcopter has 3DoF for translation and 1DoF for yaw rotation (Fig 5-a).

#### Transformable furniture control

Transformable furniture is a novel approach to utilize robotics technology in future homes. We used motorized 1DoF motorized wall/rack for the test (Fig 5-b).

#### DISCUSSION

The drag metaphor is used to command an absolute target position through the touch screen. While the physical object cannot pursue the target as quick as user gesture due to the limitation of motion control, the superimposed virtual graphic greatly helps to buffer this issue.

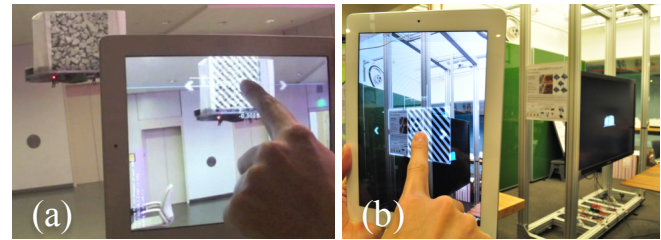
In our system, the control signal is determined based on visual servo control so that the target object approaches an absolute target position from the user's point of view. Thus the user does not need to consider the position difference or control command.

One of advantage of the exTouch gesture design is that screen drag and physical drag are not exclusive, and can be performed simultaneously. This enables continuous interaction with limited size screen of mobile devices.

The prototype application suggests the necessity of further evaluation of the user's cognitive load regarding spatial perception. Especially, the manipulation from free angle by using proposed system should be assessed as compared with fixed camera system. Validating usability of the system at sight with inexperienced users is another issue.

#### CONCLUSION

We introduced "exTouch", a spatially-aware approach to control devices through an augmented reality mediated mobile interface. The system allows spatial projection of the user input into the target motion by using egocentric camera captured image and multi-touch gesture. User can perform both screen drags (move finger on the screen) and



**Figure 5. Application example. Flying drone (a), Transformable furniture (b).**

physical drags (fix finger and move the screen physically) simultaneously. This feature allows embodied continual interaction not limited by the screen size. The preliminary exploration on the control of omnidirectional vehicle, a drone, and motorized furniture show the potential effectiveness on various types of actuated objects.

We achieved spatial extension of touch screen interaction through spatially-aware approach by augmented reality. We envision that proposed spatially-aware interaction provide further enhancement of human physical ability through spatial extension of user interaction.

#### ACKNOWLEDGMENTS

We thank Kent Larson and Hasier Larrea for their help with moving wall system and discussions about human home interaction.

#### REFERENCES

1. Ishii, K., Zhao, S., Inami, M., Igarashi, T., and Imai, M., Designing Laser Gesture Interface for Robot Control. In *Proc INTERACT 2009*, Part 2, pp.479-492.
2. Tani, M., Yamaashi, K., Tanikoshi, K., Futakawa, M., and Tanifuji, S. Object-oriented video: interaction with real-world objects through live video. In *Proc. CHI 1992*, pp. 593-598.
3. Hashimoto, S., Ishida, A., Inami, M., and Igarashi, T. TouchMe: An Augmented Reality Based Remote Robot Manipulation. The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011.
4. Seifried, T., Haller, M., Scott, S., Perteneder, F., Rendl, C., Sakamoto, D., and Inami, M. CRISTAL: Design and Implementation of a Remote Control System Based on a Multi-touch Display. In *Proc. ITS2009*, pp. 37-44.
5. D. W. F. van Krevelen, R. Poelman. A Survey of Augmented Reality Technologies, Applications and Limitations. The International Journal of Virtual Reality, 2010 Vol. 9, No. 2., pp. 1-20.
6. Bimber, O., Raskar, R. Spatial Augmented Reality: Merging Real and Virtual Worlds. A K Peters, Ltd, Wellesley, Massachusetts. Retrieved December 1, 2010.
7. Qualcomm , Vuforia™ : <http://www.qualcomm.com/solutions/augmented-reality>