User-defined Surface+Motion Gestures for 3D Manipulation of Objects at a Distance through a Mobile Device

Hai-Ning Liang[1,2], Cary Williams[2], Myron Semegen[3], Wolfgang Stuerzlinger[4], Pourang Irani[2]

[1] Dept. of Computer Science and Software Engineering
Xi'an Jiatong-Liverpool University, Suzhou, China
haining.liang@xjtlu.edu.cn

[2] Dept. of Computer Science
University of Manitoba, Winnipeg, Canada
umwill22@cc.umanitoba.ca, {haining, irani}@cs.umanitoba.ca

[3] Virtual Reality Centre
Industrial Technology Centre, Winnipeg, Canada
msemegen@itc.mb.ca

[4] Department of Computer Science and Engineering
York University, Toronto, Canada
wolfgang@cse.yorku.ca

ABSTRACT

One form of input for interacting with large shared surfaces is through mobile devices. These personal devices provide interactive displays as well as numerous sensors to effectuate gestures for input. We examine the possibility of using surface and motion gestures on mobile devices for interacting with 3D objects on large surfaces. If effective use of such devices is possible over large displays, then users can collaborate and carry out complex 3D manipulation tasks, which are not trivial to do. In an attempt to generate design guidelines for this type of interaction, we conducted a guessability study with a dual-surface concept device, which provides users access to information through both its front and back. We elicited a set of end-user surface- and motion-based gestures. Based on our results, we demonstrate reasonably good agreement between gestures for choice of sensory (i.e. tilt), multi-touch and dual-surface input. In this paper we report the results of the guessability study and the design of the gesture-based interface for 3D manipulation.

Author Keywords

Motion gestures; surface gestures; input devices; interaction techniques; multi-display environments; mobile devices; 3D visualizations; collaboration interfaces.

ACM Classification Keywords

H.5.2. Information and interfaces and presentation: User Interfaces. Input devices and strategies.

INTRODUCTION

Large displays are becoming more widespread and frequently used in the collaborative analysis and exploration of 3D visualizations. Manipulating 3D visualizations on large displays is not trivial, but present many challenges to designers [5,6,7,18,32]. Researchers have investigated the use of mobile devices to interact with objects located on distant-shared displays [2,19,20]. However, there is little research on how mobile devices can be used to carry out 3D interactions with objects at a distance. Malik et al. [17] suggest that interacting at a distance using mouse-based input is inefficient when compared to gestural interaction. Aside from being more natural, gesture-based interactions can be learned by observing other users.

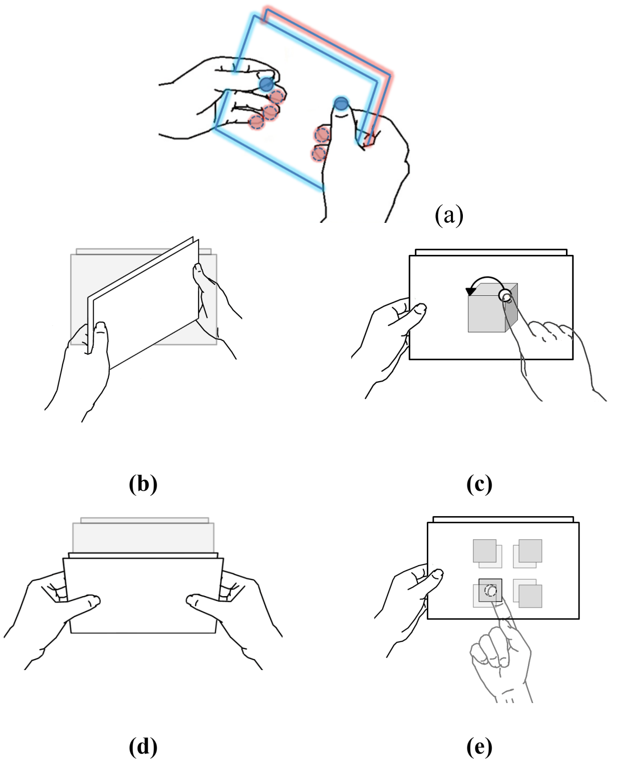


Figure 1. The dual-surface bimanual touch- and motion-enabled concept device (a); different ways of making gestures with the device: (b) rotating along the y-axis (motion-based gesture); (c) rotating along the x-axis (motion-based gesture); (d) rotating along the z-axis through the front-side (surface-based gesture); (e) interacting with a occluded objects through the back-side (surface-based gesture)

Most mobile devices now come with a touch-enabled display which can detect gestures on its surface (i.e., *surface*

APCHI '12, August 28–31, 2012, Matsue-city, Shimane, Japan.

gestures); furthermore, these devices usually incorporate highly sophisticated sensors (e.g., accelerometers, gyroscope, and orientation registers) which can recognize a variety of motions (i.e., *motion gestures*). The combination of these input capabilities enables users to express a rich set of gestural language for enhanced interaction with the mobile devices themselves [26] but also with other types of systems, such as a tabletop or wall display [2,19].

In this work, we develop a set of gestures that are easy to learn and use for 3D manipulations of distant objects via a mobile device. Gestures can be surface-based (e.g., sliding of a finger on the touch-sensitive display) and/or motion-based (e.g., shaking the device). Wobbrock et al. [39], proposed a set of *surface gestures* for tabletop systems, using a participatory approach to elicit a set of user-defined gestures. They subsequently showed that the user-specified set was easier for users to master [22]. Ruiz et al. [26] followed Wobbrock and Morris' approach and developed a user-defined set of *motion gestures* to operate mobile phones (e.g., answering a call, hanging up, etc.).

Inspired by the work of Wobbrock et al. and Ruiz et al., we developed a user-defined gesture set. We targeted 3D manipulations performed at a distance and integrated both surface and motion types of gestures—an area with little development. In this work, we addressed two research questions: (1) if users have access to more input degrees-of-freedom (multi-touch, dual-touch, and tilt), will they actually make use and benefit from them?; (2) do users have consensus as to what kinds of surface and motion gestures are natural for 3D manipulations via a mobile device? To answer these questions, we developed an experimental prototype (see Figure 1) which enables surface gestures through both the back and front sides of a tablet and can sense multiple, simultaneous finger movements. The device also detects changes in orientation, allowing users to express commands using motion. The combination of dual-surface input with simultaneous motion input can allow users varied ways of expressing gestures.

In the following sections we describe in more detail the background of our work, our experimental setup and our findings. We also elaborate on a design and a preliminary study of a potential interface for 3D manipulation.

RELATED WORK

Our work builds upon prior research on back- and front-side two-handed (or dual-surface bimanual) interaction, user-defined gestures, interaction at a distance with mobile devices, multi-display environments, and 3D interaction.

Dual-surface and bimanual interaction

The prototype (see Figure 1) used to elicit user-preferred gestures was influenced by research on back- and front-of-device, two-handed (bimanual), and dual-surface interaction.

Back-of-device interaction has been explored for mobile devices, particularly for mobile phones [1,28,29,36,41]. This type of interaction allows users to use the back-side of a device as an additional input space. RearType [28], for example, enable users to perform text-entry activities by placing a key pad on the back. HybridTouch [31] and Yang et al's prototype [41] have a trackpad mounted on the back of the PDA to enable gesture-based commands for tasks such as scrolling and steering, while Wobbrock et al. [40] suggest that such a trackpad will let users perform gestures to input unistroke alphabet letters.

Some back-of-device input enabled prototypes emphasize the use of one hand, while others require users to use both hands—i.e., in a bimanual mode. One of the benefits of bimanual interaction is the division of labor to perform simultaneous tasks. For example, Silfverberg et al.'s prototype [29] has two trackpads on the back, one for each hand, so that one hand can be delegated to zooming and the other hand to panning actions. Similarly, users need two hands to input text from a keyboard placed on the back-side in RearType [28].

Bimanual interaction is also common when interacting through the front of touch-enabled mobile devices. Touch Projector [2], a system that enables users to interact with remote screens through their mobile devices, require users to employ both hands, one for aiming at and selecting a distant device (e.g., a wall display or tabletop) and the other for manipulating objects. Researchers have claimed that two-handed interaction is more efficient, cognitively less demanding, and more aligned with natural practices than its one-handed counterpart [2,13,36].

Researchers have experimented using both sides of a device to enable input—hence, *dual-surface* input [36,41]. Yang et al. [41] have showed that *one-handed* operations can be enhanced with synchronized interactions using the back and front of a mobile device in target selection and steering tasks. Similarly, for *bimanual* operations, Wigdor et al. [36], using their *LucidTouch* dual-surface prototype, have demonstrated that users found favorable the additional dimension of back-side input because, among other things, it enabled them to interact using all of their fingers. Our dual-surface prototype was inspired by such systems that enable back-side input.

Surface and motion gestures

Aside from touch-enabled displays, current mobile devices come with other sensors which can detect motion and orientation changes. Given these capabilities, Ruiz et al. [26] have categorized gestures that these mobile devices can perform into two groups: (1) *surface* gestures and (2) *motion* gestures.

Surface gestures are carried out on the touch-enabled screen and are primarily two-dimensional. These gestures have frequently been studied in multi-touch tabletop systems (e.g., [8,10,22,39]). Morris et al. [20], from an evaluation of a multi-user photo application, have identified a classification, or 'design space,' for collaborative gestures with seven axes: symmetry, parallelism, proxemics distance, additivity, identity-awareness, number of users, and number of devices. For single tabletop users, Wobbrock et al. [38] present a taxonomy of gestures and a set of user-specified gestures derived from observing how 20 users would perform gestures for varied tasks. Surface gestures on mobile devices have also been a theme of intense study. Bragdon et al. [3] have found that, in the presence of distractors, gestures offer better performance and also reduced attentional load. Techniques, such as Gesture Avatar [16] and Gesture Search [14], show that gestures can support fast, easy target selection and data access. Gestures can also increase the usability and accessibility of mobile devices to blind people [12].

Motion gestures, on the other hand, are performed by translating or rotating the device in 3D space. These gestures have been studied for different tasks, such as to input text [10,23,34], to validate users' identity [15] and to navigate an information space [25]. Because of its wide availability, tilt has been often explored more than other types of motions. Current mobile devices allow for a rich set of motions. Ruiz et al. [26] provide a taxonomy of motion gestures, which has two main dimensions: *gesture mapping* and *physical characteristics*. Gesture mapping refers to the manner by which users map gestures to device commands and depends on the *nature*, *context* and *temporal* aspects of the motion. Physical characteristics, on the other hand, deal with the nature of the gestures themselves: the *kinetic impulse* of the motion, along what *dimension* or axes the motion occurs, and how *complex* the motion is. Ruiz et al.'s taxonomy was formulated based on a guessability study, similar to Wobbrock et al.'s study [39]. From the study, they also developed a user-inspired set of motion gestures.

To the best of our knowledge, there has not been any published research examining surface and motion gestures for dual-surface mobile devices in the content of manipulating 3D objects from a distance.

Interaction at a distance

Interaction at a distance occurs due to the unavailability of touch and unreachability of certain regions of a display. Large displays are affected by these issues, as users and the display could be separated at various distances [34]. One solution that has been proposed is to bring the content closer to the user by coupling a hand-held mobile device to the large display [2,19,20,]. Given that mobile devices also have a display, they can show a scaled-down version of the complete version shown in the large display, or as Stoakley et al. [30] would have called it a '*world in miniature*'. The coupling between the two displays can bring several benefits. It allows users to be more mobile, especially in the case of tabletops, because they do not need to touch the table surface during interaction. In addition, it supports *direct and indirect input*. Users can manipulate the content by interacting through the small device and see the effects on the large display (i.e., *indirect input*) or they can interact with the small device and observe what happens to the content on the small device itself (i.e., *direct input*). Furthermore, the small device can provide some '*personal*' or '*private*' viewing and input space only to one user—something often not available or not possible to have on large displays.

3D manipulation on 2D surfaces

Manipulating 3D object on multi-touch surfaces is non-trivial and different solutions have been proposed [4,7,8,9,18,33,37]. Davidson and Han [4] have suggested that objects' movement in the *z*-axis could be achieved using pressure. With Hancock et al.'s technique, Shallow-Depth [7], users can perform rotation and translation movements with a single finger, but 3D operations (such as rolling and pitch) will require two different touches, one for selecting the object and the other for gesturing. Another technique, Sticky Tools [8,33], need the users to first define a rotational axis using two fingers and then using a third finger to do rotation motions. The movement along the *z*-axis in both Shallow-Depth and Sticky Tools involves using a pinching gesture. Studies show that both techniques could be learned; however, they cannot be considered '*natural*' [9]. Hilliges et al. [9] and Reisman et al. [24]suggest that a more natural way of manipulating 3D objects in multi-touch surfaces is to simulate how people interact with physical objects—for example, by allowing these objects to be 'picked up' off the surface. However, understanding how the technique works is not easy because of ambiguity issues.

These proposed solutions could be categorized into two groups. The first concerns providing users with more degrees-of-freedom (e.g., [8]), while the second with offering users interactions that are natural (e.g., [9,24]). Our work is inspired by both these groups. We use a prototype which allows for a large number of degrees-of-freedom and types of input mechanisms so that we can assess whether and how they are used; and, we also develop natural interactions through a user-elicitation study with our prototype.

User-elicitation studies

A common approach to conceptualizing new interaction techniques is through *user-elicitation*, an important component of *participatory* design [27]. User-elicitation or *guessability* studies have been used by Wobbrock et al. [39] to develop their set of *surface gestures for tabletops* and by Ruiz et al. [26] to inform the design of their set of *motion gestures for smartphones*. The idea underlying a guessability study [38] is to observe what actions users will follow given the effect of a gesture (i.e., asking users to provide the *cause* for the *effect*); then, from observations across a group of users, find whether there are patterns and consensus about how a gesture is performed. In line with Wobbrock et al. and Ruiz et al., we have also developed a user-defined surface and motion gesture set by employing a user-elicitation guessability study, which we describe next.

DEVELOPING A USER-DEFINED GESTURE SET FOR A DUAL-SURFACE AND MOTION INPUT DEVICE

Our primary goal was to elicit user-defined gestures using our bimanual dual-surface tablet device (see Figure 1). The secondary goal was to identify which of the following sensory input users would employ most: (1) *Front-side* multi-touch surface; (2) *Back-side* multi-touch surface; (3) *Gyroscope* (for orientation); and/or (4) *Accelerometer* (for tilt).

Participants and apparatus

We recruited 12 participants (10 male) from a local university between the ages of 22 to 35. All participants had some experience with touch-based mobile devices.

Our experimental prototype was a dual-surface device created by putting back-to-back two Acer Iconia tablets running Android OS. The prototype had a 10.1" multi-touch surface on the front and back and connected through a wireless network. Each tablet supported up to ten touches simultaneously and came with an accelerometer and gyroscope. Users could perform *surface* gestures by moving (sliding) one finger or a set of fingers; whereas *motion* gestures were performed through rotating (rolling, pitching, or yawing) the dual-surface device. The device allowed immediate visual feedback of all users' touches on the front surface of the device

Task

Participants were asked to design and perform a gesture (surface, motion, or a hybrid of the two) via the dual-surface device (a *cause*) that they could potential use to carry out the task (an *effect*). There were 14 different tasks (see Table 1). We asked participants to do a gesture twice and explain why they chose the gesture. Participants were not told of the difference between surface and motion gesture, but only asked to perform a gesture that they feel comfortable doing.

Procedure

Each participant was asked to define a set of gestures for the above listed 14 different 3D manipulations using the dual-surface device. Participants were then handed the device so that they could get a feel for it; they began the experiment when ready.

The 14 manipulations were graphically demonstrated via 3D animations on the front display of the device. After an animation was run once, the researcher would explain the task for clarity. The animation could be replayed as many times as needed. The participant was then asked to create a gesture to effectuate the effect seen in the animation. This could be with whichever sensory input they wanted and in whatever manner they wished. While creating their gesture, the participant was asked to think aloud. Afterward s/he was asked to sketch or write a short description of the gesture on paper. This process was repeated for all 14 manipulation animations.

3D Manipulation Tasks	
Manipulation	Animation Descriptions
Rotation	
About *X* Axis	Rotate the cube so that the **top face** is facing **forward**
About *Y* Axis	Rotate the cube so that the **left face** is facing **forward**
About *Z* Axis	Rotate the cube so that the **top-right** corner becomes the **top-left** corner
Translation	
Along *X* Axis	Move the **red** cube **beside** the **blue** cube (i.e., red cube left side of blue cube)
Along *Y* Axis	Move the **red** cube **on top** of the **blue** cube
Along *Z* Axis	Move or push the **red** cube **back** towards the **blue** cube
Stretch	
Along *X* Axis	Stretch the cube **horizontally** to the **right**
Along *Y* Axis	Stretch the cube **vertically up**
Along *Z* Axis	Stretch the cube by **pulling** the cube **forwards**
Plane Slicing	
XZ plane	Cut the cube into an **upper and lower** portion
YZ plane	Cut the cube into an **left and right** portion
XY pane	Cut the cube into an **front and back** portion
Selection	
*2*D	Select the cube in the **top-left** corner
*3*D	Select the cube in the **back bottom-left** corner, hidden behind the front bottom left cube

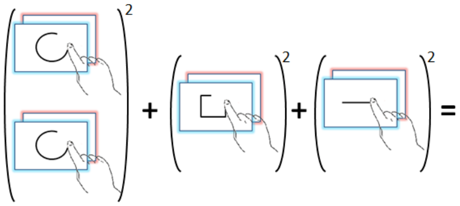
Table 1. The 3D tasks given to participants by category.

Results

From the collected gestures, we were able to create a set of gestures that seemed natural to users. We grouped identical gestures for each task, and the largest group was chosen as the user-defined gesture for the task. The set composed of the largest group for each task represents the *user-defined gesture set*. We then calculated an agreement score [38,39,26] for each task using the group size. The score reflects in one number the degree of consensus among participants. The formula for calculating the agreement scores is:

$$A_t = \sum_{P_i} \left(\frac{P_i}{P_t}\right)^2$$

where t is a task in the set for all tasks T; P_t is the set of proposed gestures for t; and P_i is a subset of identical gestures from P_t. the range for A_t is between 0 and 1 inclusive. As an example let us assume that for a task, four participants gave each a gesture, but only two are very similar. Then, the agreement score for that task would be calculated according to Figure 2.



$$A_t = (2/4)^2 + (1/4)^2 + (1/4)^2 = 0.375$$

Figure 2. Example of an agreement score calculation for a task.

Figure 3 shows the agreement scores for the gesture set, ordered in descending order. The highlighted square shows the gestures with relatively high agreement scores. The scores involving the Z Axis are located at the lower end, indicating a lower consensus. Figure 4 (next page) shows the resulting 3D gestures from the user study and obtained from the agreement scores.

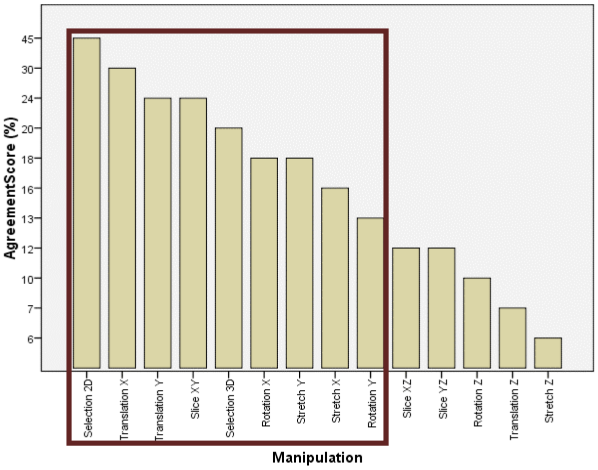
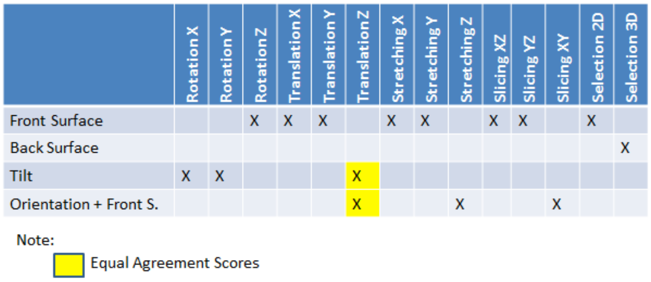


Figure 3. Agreement scores for all tasks sorted in descending order.

Figure 5 shows the user-defined gestures grouped by the sensory input used. Participants were allowed to use compound gestures. For example, to move an object along the Z Axis, some participants asked if they could rotate the entire scene and then perform a gesture along the X or Y axes to obtain the same result. The yellow cells correspond to interactions with equal agreement scores for a given input method. The front-side surface seems to be most frequently used input modality, followed by both tilt and orientation+front surface, and finally by back-side surface.



	Rotation X	Rotation Y	Rotation Z	Translation X	Translation Y	Translation Z	Stretching X	Stretching Y	Stretching Z	Slicing XZ	Slicing YZ	Slicing XY	Selection 2D	Selection 3D
Front Surface			X	X	X		X	X		X	X		X	
Back Surface														X
Tilt	X	X				🟨 X								
Orientation + Front S.						🟨 X			X			X		

Note:
🟨 Equal Agreement Scores

Figure 5. Gestures grouped by sensory input.

Discussion

From Figure 3, we observe that the agreement scores are high for tasks related to the X and Y axes, unlike the scores for tasks in the Z Axis. This shows that gestures along the Z Axis are difficult to perform. We observed that if a participant could not think of a gesture for manipulating the 3D object along the Z Axis, they would ask if the scene could be rotated in order to perform the manipulation using a gesture along the X and Y axes.

Figure 5 appears to suggest that participants preferred using *surface* gestures over *motion* gestures. However, Figure 4 indicates that participants also made use of motion gestures, especially for rotation tasks and tasks dealing with the Z Axis. During the study, we observed that most participants did not like to make large movements with the dual-surface device to create gestures. This shows that, although participants can make use of motion gestures, there seemed to be some hesitation, perhaps due to their unfamiliarity with motion gestures or maybe because the relatively large size of the device made it more difficult to perform motions with it.

From figures 4 and 5, we can see that most gestures were carried out on the front-side of the dual-surface device. That is, the front-side was the *main* input space. Figure 5 shows that the back-side was not used frequently. The few gestures that were performed on the back were unique among participants, and they therefore produced low agreement scores (see Figure 3).

There is one observation that the figures 3-5 do not show and that is that participants would touch (or begin to make a gesture from certain regions on or around the object (in our case a cube) to perform interactions. For example, to stretch along the X Axis, many participants would usually begin by touching the midpoint of the object's left and right edges. The same pattern was found for other tasks, especially those with high agreement scores (see Figure 6 for other tasks).

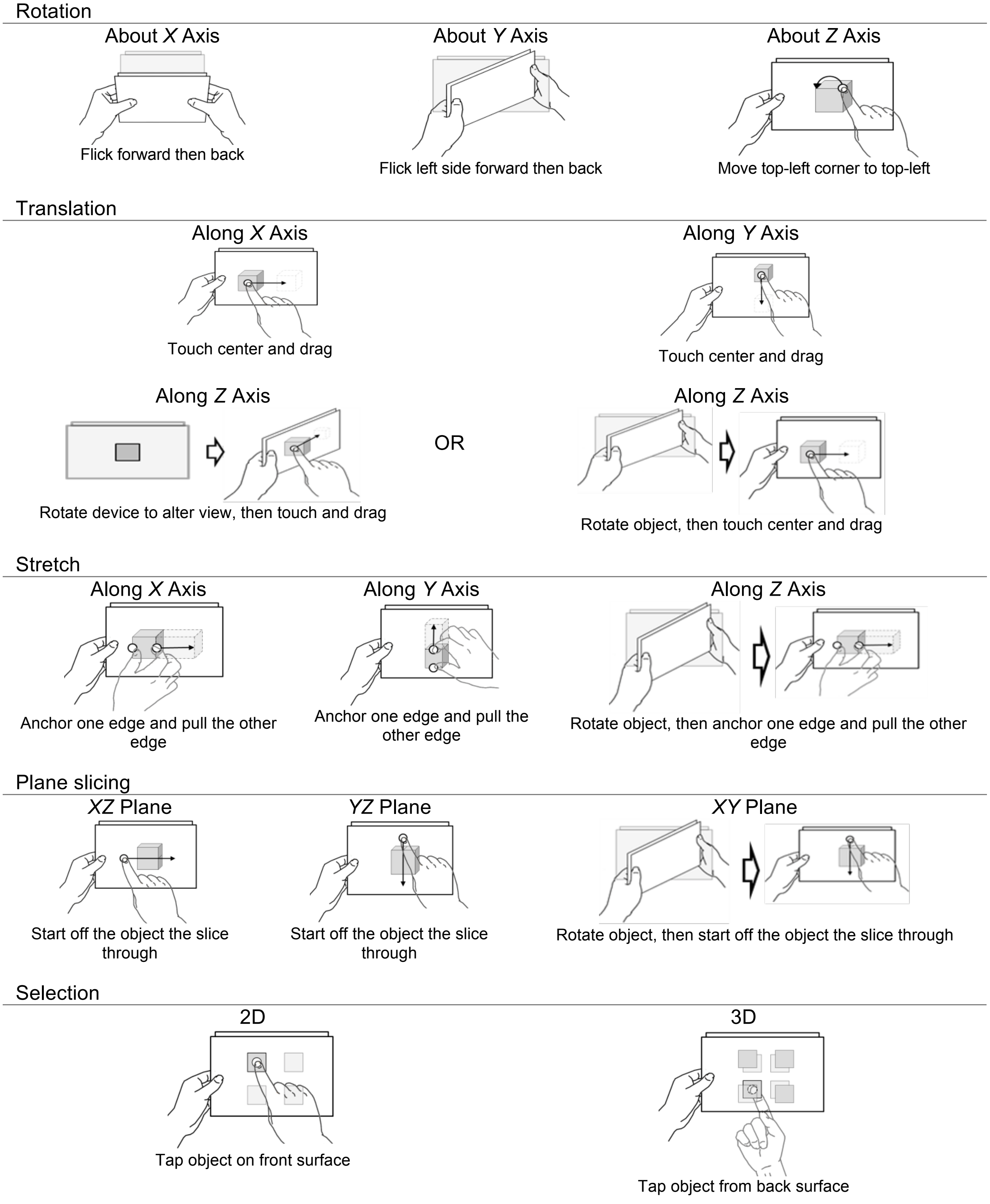


Figure 4. Resulting user-defined gesture set.

(a) Translation **(b) Rotation**

(c) Scaling **(d) Slicing**

Figure 6. Specific manipulation regions for four tasks.

SQUAREGRIDS: AN INTERFACE FOR 3D MANIPULATION THROUGH GESTURES

From the above experiment, we observed that (1) participants preferred to perform actions on the *front-side* of the dual-surface device; (2) they preferred to enact *surface* gestures along the *X* and *Y* axes; and (3) they touched specific regions (or 'hotspots') on virtual objects when performing gestures. These findings led us to modify our experimental device and design a new interface for 3D manipulation, SquareGrids (Figure 7).

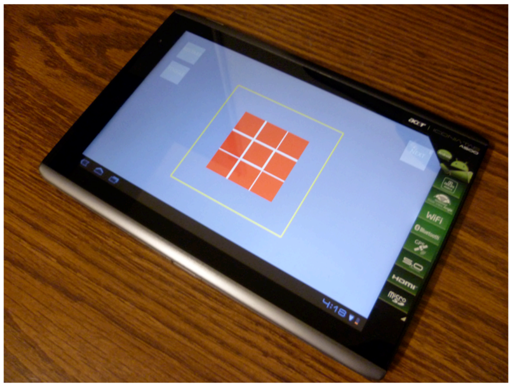
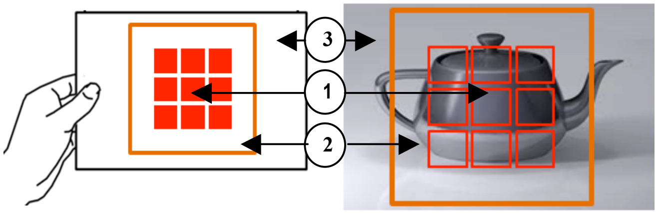


Figure 7. SquareGrids: A potential interface for 3D manipulations of distant objects.

SquareGrids used a single-sided multi-touch tablet with a accelerometer and gyroscope. Based on the gesture-input mappings obtained from the first experiment, the touch surface and the accelerometer were used as the primary input mechanisms. In addition, a new graphical interface was developed for the tablet based on the hotspots touched by users when manipulating objects.

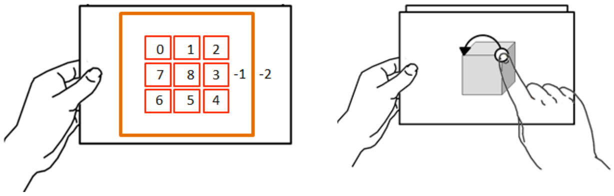
The interface was partitioned into 3 major regions, *on-object*, *off-object* and *environment* manipulations (Figure 8). The center of the interface consisted of a 3×3 grid representing the nine regions (or hotspots) that map to the 3D object designated for *on-object* interactions (Region 1 in Figure 9). The middle region (Region 2; area contained within the orange box but outside of the 3×3 grid) was an area designated for *off-object* interactions. A combination of *off-* and *on-object* interactions could be defined. For instance, most participants preferred to start the plane slicing gesture just outside the 3D object's boundaries then slice through the object (see Figure 4 plane slicing). Outside the orange box was a region for *environment* interactions (Region 3). If gestures were performed in this region, a user can manipulate the entire 3D scene (e.g., changing the camera's point of view).



(a) view from tablet device **(b) view from the large display (without the squares)**

Figure 8. Mapping of the three main regions (for (1) *on-object*, (2) *off-object*, (3) *environment* manipulations) of the SqureGrids interface (a) to a 3D object displayed on other screen (b).

Each region and their subdivision were assigned an ID (see Figure 9a). As users drag their fingers across the regions of the interface to perform a gesture, a sequence of numbers would be generated. For instance, the gesture in Figure 9b would generate the number sequences 2, -1, 0. As the gesture is being performed, the gesture recognition engine then checks the number sequence against a set of predefined gesture sequences. Once the engine recognizes the gesture, the correct 3D transformation is invoked. The gesture would continue until the user stopped the gesture motion.



(a) **(b)**

Figure 9. (a) assignment of ID numbers to reach region; (b) a user performing a gesture with sequence 2, -1, 0.

User evaluation of SquareGrids

A preliminary usability study was conducted to assess the performance of new interface against the traditional mouse for 3D manipulations.

Participants, apparatus, and task

Six male participants between the ages of 23 and 35 were recruited from a local university to participate in this study.

All participants used computers on a daily basis and are familiar with touch-based mobile devices.

To conduct this experiment we used a desktop computer (with 1.86 GHz Core 2 Duo running Windows XP) with a regular USB mouse and connected to a 24” LCD monitor. In addition, we had a laptop (with 2.0 GHz Dual Core and an Intel GMA running Windows XP) connected to another 24” LCD monitor which was linked to the mobile device prototype via a wireless network.

The task was to manipulate a solid red block by rotating, scaling and/or slicing it so that it would match in size a semi-transparent block and then dock the solid red block inside the semi-transparent block (Figure 10).

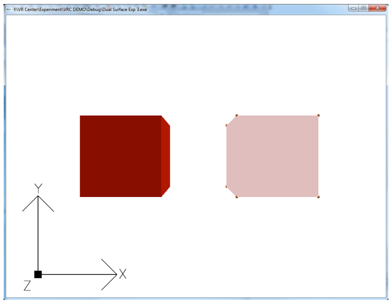


Figure 10. The 3D manipulation task: (1) match the left solid to the right solid in terms of size; (2) move the left solid inside the right solid

Conditions and procedure

This study compared two interfaces: *Mouse* (GUI-based interactions) and *Tablet* (with *SquareGrids*). In the Mouse condition participants interacted with a toolbar to select the manipulation mode and handles on the 3D object to interact with it. In the Tablet condition, participants interact with the 3D object via SquareGrids.

Each trial consisted of these tasks: *Rotation*, *Scaling* or *Plane Slicing* of the 3D object, followed by *Translation* of the object to dock it inside the semitransparent solid.

We first explain how each of the two interfaces would work and then gave participants practice trials (3 for Translation; 3 for Rotate+Translate; 3 for Scale+Translate; and 3 for Slice+Translate). In the actual experiment, participant repeated the same type of tasks, but these were slightly more complicated. The experiment lasted an hour.

We used a within-subject design. The independent variables were *Interface* (*Mouse* and *Tablet*) and *Task Type*. The order of the presentation of the interface was counterbalanced using a Latin Square design.

Results

Results indicated that participants completed the manipulation and docking tasks faster with the traditional mouse and GUI. These results could be partly due to the fact that most users were familiar with this type of interface because of frequent use.

However, participants commented that they enjoyed using the tablet interface more than the mouse interface and could see themselves using the interface in future applications. One interesting observation was that participants only needed to look at the user-defined gesture set (from Experiment 1) once or twice at the initial stages of the study. That is, the Tablet interface was easy to learn and use. This was supported by participants’ comments (e.g., “the interface was intuitive to use.”).

DISCUSSION

Implications for user interfaces

A few implications can be derived from this work. First, more modalities may not be better. As our study results suggest, despite the availability of sensors which can detect motion (both tilt and rotation), users have difficulty performing these types of actions. The size of our device could have affected users in making motion-based gestures, and a smaller device (e.g., a smartphone) would perhaps lend itself better in supporting motions. Therefore, when dealing with tablets of 10.1” or greater in size, designers should minimize the use of motion gestures. Second, although research has shown that the back-side could enrich users’ interactive experiences, our results show that users, given the choice of using the front-side, will try to minimize their use of the back-side. Such is the case despite the fact that the back-side would have enabled them to use several fingers simultaneously, potentially facilitating concurrent operations. As such, designers should perhaps maximize the use of the front-side. Third, we observed that even using the front-side, users would barely rely on multiple fingers to issue gestures. This observation indicates that users may have difficulty employing multiple touches at once, and therefore designers should be careful when designing gestures based on multi-finger operations using a 10.1” handheld tablet.

Limitations and future work

We conducted our guessability study with a mobile device of one size only. This may have influenced the types of gestures participants would make. A future line of exploration is to assess whether we can obtain the same or similar set of gestures with devices of smaller sizes, perhaps between 3.5 to 5” (the range of sizes of smartphones).

In addition, our guessability study was performed mainly with one object being displayed. We cannot be certain that we will obtain the same results if we have more than one 3D object on the screen. For instance, if objects are dense or the view shows 2 objects side-by-side, a swipe may instead affect more than one object, an operation which may not be desired. Only further research can help us come to a more definite conclusion.

Finally, related to the previous point, selection of an occluded object required participants to know in advance where the object was hidden and that there was only one object hidden by the occluding object. If there were more than one hidden object, we may not have arrived at such high agreement scores for selection operations in 3D selection. However, only further research will be able to tell us how different the gestures across users could be for these selection tasks.

SUMMARY

In this paper, we describe a guessability study to elicit a set of user-defined surface and motion gestures for a mobile device to support 3D manipulations of distant objects. The results show that there is a broad agreement in user gestures to carry out actions dealing with the *X* and *Y* axes, whereas there is a wide disagreement of those actions concerning the *Z* Axis. In addition, our observations indicate that users would likely prefer to use the front-side of a device than its back-side to perform gestures. Furthermore, observations suggest that users may be more readily able to use surface gestures than motion gestures. Finally, we provide a potential interface derived from our observations and describe a user study with the device. Our results suggest that the interface could be easy to learn and use and enables the performance of 3D tasks with a simple interface.

ACKNOWLEDGMENTS

We thank the participants for their time. We would also like to thank the reviewers for their comments and suggestions which have helped to improve the quality of the paper. We acknowledge NSERC and the Virtual Reality Centre for partially funding this project.

REFERENCES

1. Baudisch, P. and Chu, G. (2009). Back-of-device Interaction allows creating very small touch devices. CHI'09. pp. 1923 - 1932.
2. Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. (2010). Touch Projector: Mobile Interaction Through Video. CHI'10, pp. 2287-2296.
3. Bragdon, A., Nelson, E., Li, Y., and Hinckley, K. (2011). Experimental Analysis of Touch-Screen Gesture Designs in Mobile Environments. CHI'11, pp. 403-412.
4. Davidson, P. L. and Han, J. Y. (2008). Extending 2D object arrangement with pressure-sensitive layering cues. *UIST'08*, pp. 87-90.
5. Grossman, T., Balakrishnan, R., Kurtenbach, G., Fitzmaurice, G.W., Khan, A., and Buxton, W. (2001). Interaction techniques for 3D modeling on large displays. I3DG'01, pp. 17-23.
6. Grossman, T. and Wigdor, D. (2007). Going Deeper: a Taxonomy of 3D on the Tabletop. Proc. IEEE International Workshop on Horizontal Interactive Human-Computer Systems, pp. 137-144.
7. Hancock, M., Carpendale, S., and Cockburn, A. (2007). Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. CHI'07, pp. 1147-1156.
8. Hancock, M., ten Cate, T. and Carpendale, S. (2009). Sticky Tools: Full 6DOF Force-Based Interaction for Multi-Touch Tables. ITS'09, pp. 145-152.
9. Hilliges, O., Izadi, S., Wilson, A. D., Hodges, S., Garcia-Mendoza, A., and But, A. (2009). Interactions in the Air: Adding Further Depth to Interactive Tabletops. UIST'09, pp. 139-147.
10. Hinrichs, U. and Carpendale, S. (2011). Gestures in the Wild: Studying Multi-Touch GestureSequences on Interactive Tabletop Exhibits. CHI '11, pp. 3023-3032.
11. Jones, E., Alexander, J., Andreou, A., Irani, P., and Subramanian, S. (2010). GesText: Accelerometer-Based Gestural Text-Entry Systems. CHI '10, pp. 2173-2182.
12. Kane, S.K., Bigham, J.P. and Wobbrock, J.O. (2008). Slide Rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques. ASSETS '08, pp. 73-80.
13. Kin, K., Hartmann, B., and Agrawala, M. (2011). Two-handed marking menus for multitouch devices. ACM Trans on CHI, 18(3), 1-23.
14. Li, Y. (2010). Gesture Search: A Tool for Fast Mobile Data Access. UIST'10, pp. 87-96.
15. Liu, J., Zhong, L., Wickramasuriya, J., and Vasudevan, V. (2009). User evaluation of lightweight user authentication with a single tri-axis accelerometer. MobileHCI '09, pp. 1-10.
16. Lu, H. and Li, Y. (2011). Gesture Avatar: A Technique for Operating Mobile User Interfaces Using Gestures. CHI'11, pp. 207-216.
17. Malik, S., Ranjan, A., and Balakrishnan, R. (2005). Interacting with large displays from a distance with vision-tracked multi-finger gestural input. UIST'05 pp. 43-52.
18. Martinet, A., Casiez, G., and Grisoni, L. (2010). The effect of DOF separation in 3D manipulation tasks with multi-touch displays. VRST '10, pp. 111-118.
19. McAdam, C & Brewster, S. (2011). Using mobile phones to interact with tabletop computers. ITS'11, p. 232-241.
20. McCallum, D.C. and Irani, P. (2009). ARC-Pad: absolute+relative cursor positioning for large displays with a mobile touchscreen. UIST '09, pp. 153-156.
21. Morris, M.R., Huang, A., Paepcke, A., and Winograd, T. (2006). Cooperative Gestures: Multi-User Gestural Interactions for Co-located Groupware. CHI'06, pp. 1201-1210.

22. Morris, M.R., Wobbrock, J., and Wilson, A. (2010). Understanding Users' Preferences for Surface Gestures. GI'10, pp. 261-268.
23. Partridge, K., Chatterjee, S., Sazawal, V., Borriello, G., and Want, R. (2002). TiltType: accelerometer-supported text entry for very small devices. UIST '02, pp. 201-204.
24. Reisman, J.L., Davidson, P.L., and Han, J.Y. (2009). A screen-space formulation for 2D and 3D direct manipulation. UIST '09, pp. 69-78.
25. Rekimoto, J. (1996) Tilting operations for small screen interfaces. UIST '96, pp. 167-168.
26. Ruiz, J., Li, Y., and Lank. E. (2011). User-Defined Motion Gestures for Mobile Interaction. CHI'11, pp. 197-206.
27. Schuler, D. (1993). Participatory design: principles and practices. L. Erlbaum Associates, Hillsdale N.J., 1993
28. Scott, J., Izadi, S., Rezai, L. S., Ruszkowski, D., Bi, X., and Balakrishnan, R. (2010). RearType: Text Entry Using Keys on the Back of a Device. MobileHCI'10. pp. 171-180.
29. Silfverberg, M., Korhonen, P., and MacKenzie,I.S. (2006). Zoomig and panning content on a display screen. United States patent 7075513, July 11, 2006.
30. Stoakley, R., Conway, M.J., and Pausch, R. (1995). Virtual reality on a WIM: interactive worlds in miniature. CHI'95, pp. 265-272.
31. Sugimoto, M. and Hiroki, K. (2006). HybridTouch: an intuitive manipulation technique for pdas using their front and rear surfaces. Extended Abstracts of MobileHCI'06. pp. 137-140.
32. Valkov, D., Steinicke, F., Bruder, G., and Hinrichs, K.H. (2011). 2D Touching of 3D Stereoscopic Objects. CHI'11, pp. 1353-1362.
33. Vlaming, L., Collins, C., Hancock, M., Nacenta, M., Isenberg, T. and Carpendale, S. (2012). Integrating 2D mouse emulation with 3D manipulation for visualizations on a multi-touch table. ITS'10, pp. 221-230.
34. Vogel, D. and Balakrishnan, R. (2004). Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. UIST'04 , pp. 137-146.
35. Wigdor, D. and Balakrishnan, R. (2003). TiltText: Using tilt for text input to mobile phones. UIST '03, pp. 81-90.
36. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., and Shen, C. (2007). LucidTouch: A See-through Mobile Device. UIST'07. pp. 269-278.
37. Wilson, A., Izadi, S., Hilliges, O., Garcia-Mendoza, A., and Kirk, D. (2008). Bringing physics to the surface. UIST'08, pp. 67-76.
38. Wobbrock, J.O., Aung, H.H., Rothrock, B., and Myers, B.A. (2005). Maximizing the guessability of symbolic input. CHI '05 Extended Abstracts, 1869.
39. Wobbrock, J.O., Morris, M.R. and Wilson, A.D. (2009). User-defined gestures for surface computing. CHI'09, pp. 1083-1092.
40. Wobbrock, J. O., Myers, B. A., and Aung, H. H. (2008). The performance of hand postures in front- and back-of-device interaction for mobile computing. IJHCS. v66 (12), pp. 857-875.
41. Yang, X. D., Mak, E., Irani, P., and Bischof, W. F. (2009). Dual-Surface Input: Augmenting One-Handed Interaction with Coordinated Front and Behind-the-Screen Input. MobileHCI'09. pp. 1-10.