3D Object Selection for Hand-held Auto-stereoscopic Display

Euijai Ahn
Digital Experience Laboratory
Korea University, Seoul, Korea
saintpio@korea.ac.kr
+82-2-3290-3579

Hyunseok Yang
Digital Experience Laboratory
Korea University, Seoul, Korea
baroeye@korea.ac.kr
+82-2-3290-3579

Gerard J. Kim
Digital Experience Laboratory
Korea University, Seoul, Korea
gjkim@korea.ac.kr
+82-2-3290-3196

ABSTRACT
Interacting in a small (mobile) auto-stereoscopic display can be difficult because of the lack of accurate tracking of an interaction proxy, and having to maintain a fixed viewpoint and adapt to a different level of depth perception sensitivity. In this paper, we first propose to modify a standard stylus into a mechanical chain with joint sensors for 3D tracking. We also investigate a way to assist the user in selecting an object in the small phone space through supplementary multimodal feedback, such as sound and tactility. We have carried out an experiment comparing the effects of various combinations of multimodal feedback to object selection performance.

Author Keywords
Auto-stereoscopy, Depth perception, Mobile interaction, Multimodal interaction, Selection, 3D tracking.

ACM Classification Keywords
H.5.m [Information Interfaces and Presentation]: Miscellaneous.

General Terms
Experimentation, Human Factors.

INTRODUCTION
With the continuing technological innovation and recent keen public interests, stereoscopic displays are becoming more commonplace these days. They are being adopted for TV's, desktop displays and finally smart phones and hand-held devices [1, 2, 3, 4], but in most cases, still used for viewing only. An exception is for when used for special purpose virtual reality (VR) based interactive applications. In fact, 3D interaction techniques for VR (with relatively large-sized stereoscopic display) have been studied considerably [5]. However, not much attention has been paid to the problem of interacting in a relatively “small-sized” hand-held stereoscopic display such as that of a smart phone.

One of the difficulties arise from the lack of accurate and robust method for 3D tracking within the small “phone” space (e.g. small rectangular volume right above the phone display, also see Figure 2a). Another possible source of complications is the fact that small mobile 3D displays are invariably auto-stereoscopic (e.g. parallax barrier type), limiting the user to fix and maintain one’s view point to feel the 3D effect. This is undoubtedly a more difficult task with hand-held devices than with large fixed displays. In addition, it is plausible to expect some differences in workings of the human’s depth perception in the significantly “small” phone space compared to the nominally sized space (e.g. human scale).

Figure 1. 3D interaction on a stereoscopic phone using an articulated stylus with joint sensors (for 3D tracking). The stylus can be folded for easy stow away (left).

Our paper thus starts with a proposal for a practical solution to 3D tracking for a mobile phone, using an articulated stylus with joint sensors (see Figure 1). Then we also propose to assist the user in selecting an object in the small phone space through supplementary multimodal feedback, such as sound and tactility to overcome the aforementioned projected difficulties. We have carried out an experiment comparing the effects of various combinations of multimodal feedback to object selection performance.

Our paper is organized as follows. First we provide a review of previous and related research. Then we describe the 3D auto-stereoscopic smart phone and newly proposed tracking system, used for the following experiment. Section 4 gives details of the experiment and results. Finally we discuss and summarize the findings from our experiments to conclude the paper.
RELATED WORK
3D interaction techniques have been studied much in depth mainly in the virtual reality community. An excellent review of the various techniques and their taxonomy are given in [5]. However, subtle difference in their performance or usability according to different types of 3D displays (e.g. auto-stereoscopic, active or passive type, head mounted display) has not been looked at much, especially for small displays [6]. One noteworthy work by [7] studied interaction for small hand-held stereoscopic (passive chromatic anaglyph) display. In this work, the interaction was indirect or gesture based realized by tracking the user’s fingertip on the other side of the display using the back facing camera. To our knowledge, there has not been a research study for directly interacting with stereoscopically rendered object in 3D. This is partly due to the difficult problem of accurate tracking in the “phone” space in a self-contained way (i.e. without any third party sensor). The most prevalent approach is to use the phone camera, however, due to its limited field of view, it is not feasible for the tracking volume to cover the entire “phone” space, especially near the display surface.

The phenomenon of altered depth perception with the use of stereoscopic display also has been reported in [8, 9]. For example, humans tend to underestimate depth when head mounted display (HMD) is used [8]. Not much is known about the dynamics of depth perception for auto-stereoscopic displays that use parallax barriers or lenticular sheets, not to mention for small-sized ones. Yang et al. compensated for the depth underestimation in HMDs by manipulating its geometric field of view providing additional multimodal feedback [10]. Similarly, multimodal feedback has been regarded one way to improve 3D task performance (which must be related to depth perception) [11, 12, 13, 14, 15, 16, 17]. For example, Stephen et al. [14] used multimodal (visual and aural) feedback to help users perceive depth more accurately and carry out 3D spatial tasks. However, to date, the results are not consistent in terms of which modality combination is most helpful due to differences in the task and experimental conditions.

EXPERIMENTAL PLATFORM
3D Tracking: Articulated Stylus
In order to interact directly with 3D objects, 3D tracking is required. Our proposal is to use an articulated stylus with joint sensors as shown in Figure 1. We believe such a device can be designed almost as compact as the conventional stylus with miniaturized yet highly accurate joint encoding sensors. Such a design consideration is necessary to keep the mobile phone light and “handy” to use. The sensors and feedback devices would be directly connected into the smart phone for end-effector coordinate computation and feedback control. Note that the articulated stylus is not a haptic device, but merely a tracking one.

When not in use, it can be simply be folded and stowed away.

Our actual “lab” implementation was bigger in size than the envisioned version, with four degrees of freedom articulation, and using analog potentiometers at the joints (rather than more accurate high resolution digital encoders). Additional circuitry were needed for digital conversion and interfacing into the smart phone using Bluetooth. Lego pieces were used for the links (see Figure 2). A small vibrator and button were attached for the interaction purpose.

Figure 2. Actual “lab” implementation of the articulated stylus using lego pieces, potentiometers and associated circuitries. (a) button, (b) vibrator, (c) stylus tip.

As for the associated circuitry, the Arduino board [18] using the Atmega328 MCU was used and the on-board 10 bit A/D converter was used for converting the potentiometer joint angles into digital values. Based on a standard forward kinematic formulation (which we omit the details in this paper) [19], the positional coordinate of the stylus tip is easily computed. For stable output, basic low pass filtering was applied. The orientation (of the tip) was not computed nor used in this work. All the computations (for now) were carried out on the MCU and transmitted to the smart phone at a rate of approximately 35 Hz. The vibrator motor was controlled by the pulse width modulation signal output from the same board. We re-emphasize that, if professionally built with the state-of-the-art components, the stylus can be as compact and accurate as originally proposed.

Tracking Accuracy and Calibration
To measure the accuracy of our device, we built small 3D structures with Lego blocks (see Figure 3) and compared the computed coordinates of the stylus tip and the ground
truth of various points in the structures. Figure 4 illustrates
the accuracy of the articulated stylus in the x-y plane
(ground truth: red circles, measured and computed: blue
diamonds). In all three directions, the errors were on the
average within about 2mm. As our focus was more with
deriving an effective 3D object selection technique, no
further significant effort was made to improve the accuracy.
However, due to the personal variations in depth perception,
we asked each user, during the experiment, to designate
several 3D landmark points (similarly to calibrating a touch
screen) to calibrate them against the corresponding ones in
the virtual space (see Figure 5).

**Auto-stereoscopic smart phone**
The auto-stereoscopic phone used in our experiment was an
LG Optimus 3D [4] with a 4.3 inch 3D (parallax barrier)
LCD display (480 x 800 pixels, 16M colors). Parallax
barrier technology refers to creating the 3D effect by using
a barrier (layer of material with a series of precision slits)
placed on the image source (e.g. LCD) such that each sees
the respective right or left image (without the need to wear
special glasses, see Figure 6) [20]. A disadvantage of the
technology is that the user must be positioned in a well-
defined spot to experience the 3D effect. The exact spot
depends on the inter-ocular distance of the user, but for this
phone model, it was approximately 30cm perpendicularly
above from the center of the screen. Another disadvantage
is that the effective horizontal pixel count viewable for each
eye is reduced by one half. The typical operational phone
space was assumed be shaped as a rectangular volume with
the physical dimensions of 56mm x 93mm x 40mm, as
viewed from the sweet spot. Note that above figures are
nominal values only; both the proper viewing position and
perceived size of the phone space would be slightly
different for different users.

**Overall interaction architecture**
Figure 7 illustrates the overall architecture combining the
tracking device to the smart phone. The computed stylus
tip coordinates are relayed to the smart phone which
visually renders the virtual world in stereo, and other modal
feedback (e.g. aural and tactile).

**Figure 3.** One of the 3D structures used for accuracy
measurement.

**Figure 4.** Accuracy in the x-y plane (units in mm). Ground
truth are marked in red circles, and measured and computed in
blue diamonds.

**Figure 5.** The “Phone” space and the five designated points for
calibration to the virtual volume.

**Figure 6.** The parallax barrier technology used in the LG auto-
stereoscopic phone.

**Figure 7.** Overall interaction architecture: tracking module on
the left and smart phone on the right.
EXPERIMENT: OBJECT SELECTION TECHNIQUE

Experiment Design
With the experimental platform in place, rudimentary object selection in the 3D phone space has become possible. Nevertheless due to the factors mentioned previously (e.g. fixed spot viewing, reduced resolution, unknown dynamics of the depth perception in small sized volume, etc.), we expect some difficulties in fluid interaction. As such, we propose to take advantage of supplementary modal feedback, namely, aural and tactile, and carry out an experiment to explore the feedback design space. We compare four different feedback conditions as an aid to making object selection. They are (1) visual only (V, the reference), (2) visual and aural (VA), (3) visual and tactile (VT), and (4) visual, aural and tactile (VAT).

Since the correct depth perception is the matter of importance in this work, as for the experimental task, we presented two objects (cubes) of similar depth and asked the user to disambiguate the depth between them. More experimental details follow in the subsequent subsections. In summary, the experiment had one factor, the type of multimodal feedback, with four levels (1 x 4 within subject repeated measure) and the task performance and usability were measured as major dependent variables. Our main hypothesis was that higher degree of multimodal feedback would generally improve the object selection or depth perception performance.

Multimodal Feedback
The visual feedback merely consisted of rendering of the cubes. To remove any external bias, we rendered the cubes with orthogonal projection and minimal lighting effects (Figure 8). It is well known that perspective projection alone is a very strong psychological depth cue. We eliminated all depth cues except for the binocular disparity.

Similarly, the amplitude/frequency of the vibratory tactile feedback was inversely proportional to the depth (from the user) of the object as well. The vibration motor, when controlled by the PWM signal, varies the vibration frequency and its amplitude at the same time. Ideally, only the amplitude would be varied with a fixed vibration frequency, supposedly most perceptible by humans, for instance at around 250 Hz [21]. In this experiment, both feedback lasted for one second when generated. Table 1 shows the stimulation parameters used for the respective modal feedback.

<table>
<thead>
<tr>
<th>Depth from user (cm)</th>
<th>Depth from screen (cm)</th>
<th>Tone Frequency (Aural Feedback)</th>
<th>Vibration Duty rate (Tactile Feedback)</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.0</td>
<td>1.0</td>
<td>730Hz</td>
<td>31%</td>
</tr>
<tr>
<td>28.5</td>
<td>1.5</td>
<td>1360Hz</td>
<td>43%</td>
</tr>
<tr>
<td>28.0</td>
<td>2.0</td>
<td>1990Hz</td>
<td>54%</td>
</tr>
<tr>
<td>27.5</td>
<td>2.5</td>
<td>2610Hz</td>
<td>66%</td>
</tr>
<tr>
<td>27.0</td>
<td>3.0</td>
<td>3240Hz</td>
<td>77%</td>
</tr>
<tr>
<td>26.5</td>
<td>3.5</td>
<td>3870Hz</td>
<td>89%</td>
</tr>
<tr>
<td>26.0</td>
<td>4.0</td>
<td>4500Hz</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1. Stimulation parameter values used for aural and vibratory tactile feedback.

As for the aural and tactile feedback, they were generated when the 3D cursor (drawn at the tip of the articulated stylus) came into some proximal distance with a nearby object (Figure 9).

A sine tone was generated as aural feedback and the frequency was determined by the depth (or discrete depth level) of the object to whom the 3D cursor is proximal. The higher the object was (i.e. distant from the screen, closer to the user), likewise the tone frequency. A reasonable audible frequency range was mapped to the depth range of the stereoscopic display.

Experimental Task
The experimental task involved the user to determine the relative depth between two cubes. The subject was to carry out a series of these depth determination tasks as fast and correctly as possible. Two cubes with different (randomly chosen) depth values appeared in the 3D phone space (but with an equal planar distance), and the user was to choose the deeper object using the articulated stylus under different treatment or feedback conditions (see Figure 10).
Experimental Procedure
Twenty one paid subjects (15 men and 6 women) participated in the experiment with the mean age of 24.5. After collecting one's basic background information, the subject was briefed about the purpose of the experiment and instructions for the experimental task. A short training (15 to 20 minutes) was given for the subject to get familiarized to the experimental process and using the stylus. The subject tried out each treatment combination in a balanced order. For each treatment, the depth test was conducted five times. The task completion time and correctness data were captured and after all the treatments were tried, a general usability questionnaire was filled out (answered in 7 scale Likert scale).

RESULTS
Task Performance
ANOVA has reaffirmed the effect of the multimodal feedback. VAT exhibited the fastest task performance but with no statistical difference from VA. In addition, VT was neither differentiated from V. Thus, in our experiment, only the aural feedback was meaningfully effective (Figure 11).

Usability
The usability questionnaire asked the subject to comparatively rate the four selection (or feedback) techniques in terms of ease of use, degree to which feedback was helpful in recognizing the depth, ease of learning, interaction naturalness and the level of fatigue. Figure 13 illustrates the results.

Figure 10. Experimental task: depth determination.
Figure 11. Task completion times for the four feedback conditions.
Figure 12. Number of incorrect responses for the four feedback conditions.
The usability results are quite consistent with that of the quantitative performance results. For instance, with multimodal feedback, the user felt the task was generally easier, and the easiest for VA and VAT, which were statistically not different, again showing the reduced role of the tactile feedback (Figure 13a). Users also responded that the aural feedback was the most helpful, and less so when only tactile feedback was present or mixed (Figure 13b). Figure 13c shows that, as we have hypothesized, that it was difficult for the users to determine depth solely from visual feedback. Again, the subjects felt the selection technique was easiest to learn, most natural and least tiring with the aural feedback only. We observe in general that when aural and tactile feedback are both presented, the usability and task performance was lower than when only aural feedback is presented. Thus there seems to be an interaction among these two elements. In fact, it is reported that redundant feedback may degrade task performance [22] and this result is also consistent with cases when object selection is carried out in larger interaction space [13]. However, it is also quite possible the vibratory tactile feedback we devised was not ideally designed to human perception.

**DISCUSSION AND CONCLUSION**

Based on our experiment, we reaffirmed that multimodal feedback helped users select objects better. Post-briefing with the subjects further confirmed this deduction. Many complained of dizziness and blurred imagery in trying to perceive 3D. This was more apparent with the higher degree of negative parallax (object being felt to hover further out of the screen). Despite the possibility of non-ideally designed vibratory tactile feedback method, we converge to a conclusion that only one supplementary and aural feedback was the most effective object selection method. Many subjects indicated that they were confused when both aural and tactile feedback were given and preferred only one of the two. They also reported the difficulty to sense the depth with the “vibrating” stylus and due to the low disambiguating power (relative to the depth range) of the tactile feedback itself. Some even consciously tried to block tactile feedback when presented together with the aural feedback.

Our future work includes investigating in other interaction techniques such as object manipulation and menu selection. Other forms of aural and tactile feedback

**ACKNOWLEDGMENTS**

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MEST) (No. 2011-0030079).

**REFERENCES**


