An Evaluation of Mobile Phone Pointing in Spatial Augmented Reality

Jeremy Hartmann
School of Computer Science
University of Waterloo, Canada
j3hartma@uwaterloo.ca

Daniel Vogel
School of Computer Science
University of Waterloo, Canada
dvogel@uwaterloo.ca

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada
ACM 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188535

Abstract
We investigate mobile phone pointing in Spatial Augmented Reality (SAR). Three pointing methods are compared, ray-casting, viewport, and tangible (i.e. direct contact), using a five-projector "full" SAR environment with targets distributed on varying surfaces. Participants were permitted free movement in the environment to create realistic variations in target occlusion and target incident angle. Our results show raycast is fastest for high and distant targets, tangible is fastest for targets in close proximity to the user, and viewport performance is in between.

Author Keywords
pointing; mobile interaction; spatial augmented reality; XR

ACM Classification Keywords
H.5.2 [User Interfaces]: Interaction Techniques

Introduction
Spatial Augmented Reality (SAR) [6] places digital content directly onto objects in a real physical environment. This is currently achieved with projection mapping, where one or more digital projectors are calibrated to environment geometry. An advantage SAR has over other augmented reality methods is that content is viewable without wearing a head-mounted display (HMD), however the best way to interact in SAR content remains an open problem.

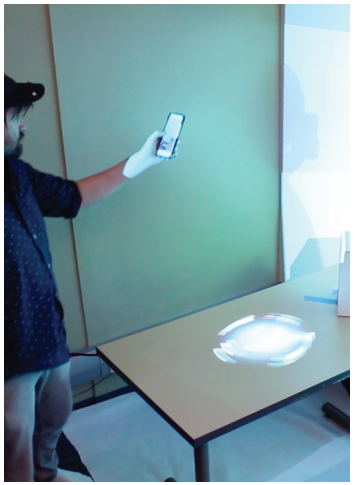

Figure 1: Mobile phone pointing in SAR: (top) viewport distant pointing technique; (bottom) tangible pointing technique.

Mobile phones may offer a solution. They are ubiquitous and techniques already exist to track their 3D position using on-board cameras and sensors. Previous work has investigated mobile phone pointing for large displays [9], multi-display environments [2], and viewport AR [7]. To our knowledge, no previous work has compared mobile phone pointing in "full" SAR, where users have free movement and targets are distributed on surfaces that differ in orientation and scale.

Mobile Phone Pointing in SAR

We briefly describe the technical infrastructure to create our SAR environment and provide details for the three mobile phone pointing techniques compared in the experiment.

SAR Environment

Our system is based on SAR projection mapping [6], and it shares many traits with other room-scale projection systems like RoomAlive [5]. However, our system also integrates accurate mobile phone tracking and interaction techniques that we evaluate in a complex environment with variation in planar geometry orientation, scale, and occlusion.

Our SAR environment is situated in a corner occupying approximately 4×4 meters of floor space. Placed around the environment are five digital projectors, six Kinect cameras (each connected to an IntelNUC PC), and a ten-camera Vicon motion tracking system for real time tracking of a mobile phone and a person's head (both using 6.4mm markers). Projectors and Kinect cameras are calibrated using the RoomAlive toolkit [5]. Our software uses C# and Unity3D for the rendering back-end based on the 3D reconstruction of the room from RoomAlive. The mobile phone is a Google Pixel (5.0 inch display, dimension with case $149 \times 74 \times 11$mm) running Android Nougat. A custom Android app communicates with the server to render a simple interface for experiment control and pointing techniques. Using this system, we created three mobile phone pointing techniques suitable for SAR (see Figure 1).

Viewport Pointing

Using the phone's camera as a viewport to interact with remote targets has been used in many contexts. Some use a fixed cross-hair at the center of the screen [7], we use a more versatile method where targets can be selected anywhere with a tap [1]. For experimental control, the view is a 3D rendering of the room using a virtual camera that matches the real mobile phone camera. To use the technique, the user moves the phone to roughly frame the desired target and taps to select it (Figure 2a).

Raycast Pointing

Raycast methods may be implemented with physical lasers [9], but we used 3D tracking to create a virtual ray. To use the technique, the user holds the mobile phone with either hand, points the front end towards the currently active target, and taps a button on the screen to make a selection (Figure 2b).

Tangible Pointing

Tangible interaction, where a target is selected by directly touching it with the mobile phone, has primarily been used with tabletops [8] and large displays [3]. We use the tracked position of the phone to test if the corresponding virtual mobile phone bounding box intersects with room geometry. To use the technique, the user holds the mobile phone with either hand and physically taps the currently active target with a corner, side, or face (Figure 2c).

Related Work

Our work relates to previous evaluations of mobile phone pointing, including similar mid-air hand-held devices like WiiMotes or laser pointers. For the most part, these evalu-

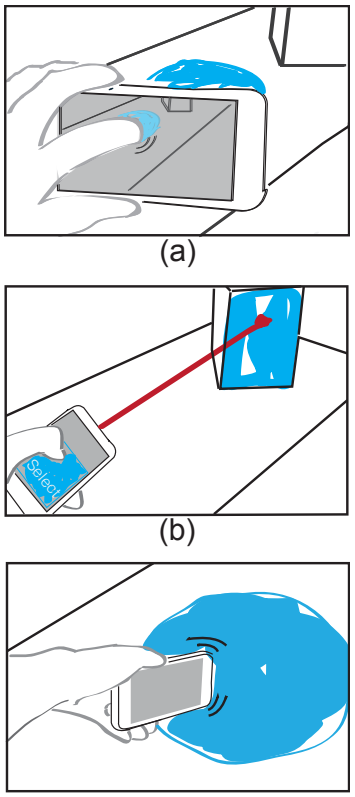

Figure 2: Mobile phone pointing techniques: (a) viewport, where targets are captured by a camera-like view; (b) raycasting, where targets are pointed at; (c) tangible, where targets are directly contacted.

ations have been conducted in environments different than full SAR: large displays, multi-display environments, tabletops, and viewport AR with near-planar scenes.

Handheld devices have been proposed for AR environments that use pointing methods similar to raycasting. XWand [10] is a multi-modal device combining AR pointing with gestures and speech. Results indicated that wand pointing accuracy can be improved through auditory feedback. GyroWand [4] utilizes an inertial measurement unit (IMU) as a means to perform 3-D pointing tasks for VR presented in an HMD.

Rhos and Oulasvirta [7] proposed and evaluated a predictive model for AR selection times using viewport mobile phone pointing, but the scene used for pointing was essentially near-planar. Boring et al. [1] explored viewport interaction through a live camera feed by utilizing touch interaction with digital content captured within the frustum of the mobile phone's camera.

Experiment

This within-subjects experiment compares the three mobile phone pointing techniques in a realistic full SAR environment. For each technique, participants are free to move around as they select a series of 2-D targets mapped to different surfaces in the environment. We log key metrics to identify how factors like target occlusion, projector occlusion, participant distance from target, and target visual angle may affect performance.

Participants

We recruited 18 participants (ages 21 to 50; 5 female; 17 right-handed). Most actively used a mobile device an average of 3.1 hours a day. Participants received $10, and a research ethics board approved the study.

Task

The task was to select two targets in sequence as quickly and accurately as possible. The first target was a circular *Start Target* ($r = 18$cm). It was always placed at the center of a table. The second target could be either a *circle* ($r = 13$cm) or a *rectangle* of varying dimensions. There were 19 targets grouped into five types, HIGH, MID, LOW, TABLE, and LARGE (see Figure 4).

Design and Protocol

The primary independent variables are TECHNIQUE (3 levels: VIEWPORT, RAYCAST, and TANGIBLE) and TARGET (19 different targets, grouped into five categories: HIGH, MID, LOW, TABLE, and LARGE). The ordering of TECHNIQUE for each participant was counter balanced using a Latin square. For each TECHNIQUE, the participant completed 5 BLOCKS of 19 TARGET selection pairs presented in random order. Auditory feedback was given for successful and unsuccessful selections.

Results

Repeated measures ANOVA and pairwise t-tests with Holm correction were used for all measures[1]. Trials were aggregated by participant and the factors being analyzed. All time data was aggregated using the median to account for a skewed distribution. To remove outliers, target times more than 3 standard deviations from the mean were excluded. These accounted for 1.5% of the total data. Error rate is calculated as the ratio between the number of trials with one or more errors to the total number of trials. There was no significant effect of BLOCK on movement time across all techniques ($p > .89$). With no strong learning effect present, all blocks were retained in the analysis.

[1]When the assumption of sphericity was violated, we corrected the degrees of freedom using Greenhouse-Geisser (Greenhouse-Geisser's $\varepsilon < 0.75$) or Huynh-Feldt (Greenhouse-Geisser's $\varepsilon \geq 0.75$).

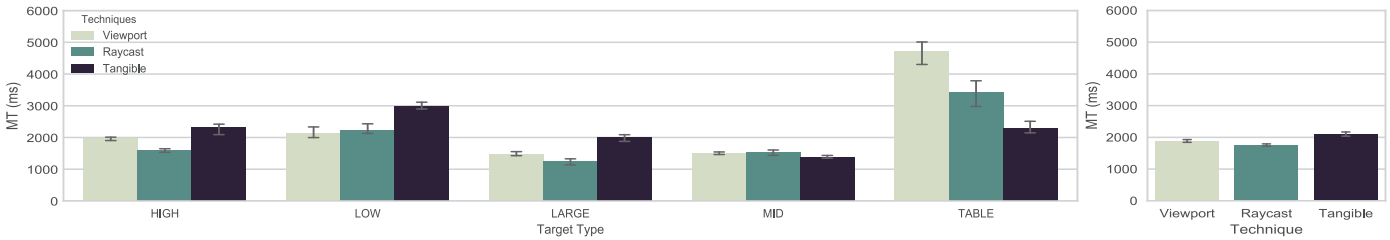

Figure 3: MT for TECHNIQUE and TARGET types (left). MT for TECHNIQUE on all combined types (right).

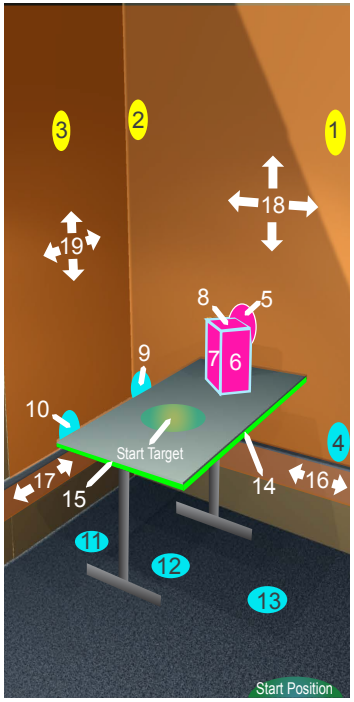

Figure 4: The SAR environment showing all targets. Where yellow is HIGH, orange is LARGE, fuchsia is MID, neon-green is TABLE, and light-blue is LOW. The start target is the yellow-green oval on the table, and the start position is the blue-green oval on the floor.

Movement Time

Movement time (MT) is the time between selecting the start target and the experimentally manipulated measurement target. There is a significant main effect for TECHNIQUE on MT ($F_{2,34} = 7.39$, $p < .002$). Post-hoc tests show all techniques are significantly different ($p < .001$): RAYCAST (1754ms) is slightly faster then both VIEWPORT (1885ms) and TANGIBLE (2104ms) overall (Figure 3 right).

The significant interaction between TECHNIQUE and TARGET type on MT ($F_{8,136} = 69.59$, $p < .0001$) reveals consistent patterns. Post-hoc tests verify differences between all techniques and target types ($p < .03$) except LOW which has no effect between VIEWPORT and RAYCAST ($p > .41$). Highlighting the most salient results: RAYCAST was fastest for HIGH (1588ms) and LARGE (1238ms), but both RAYCAST (2236ms) and VIEWPORT (2112ms) are essentially tied for LOW; TANGIBLE has the fastest MT for both MID (1383ms) and TABLE (2293ms) (see Figure 3 left). Our results show that raycasting and tangible may be better suited for different types of targets, but viewport can be an equivalent choice in some cases.

Occluded Targets

To measure target occlusion, the system rendered the participant's view of the target from a virtual camera positioned at their head with both the environment geometry and without. This enabled us to identify whole or partially occluded targets by calculating the ratio between visible and non-visible portions. Using this metric, we found that in 20.3% of trials, the measurement target had at least some occlusion.

The significant interaction between TECHNIQUE and OCCLUSION on MT explains how each technique handled initial occlusion ($F_{2,34} = 12.19$, $p < .0001$). Post-hoc tests show VIEWPORT suffered the most, with initially occluded targets taking an additional 2115ms. This is compared with RAYCAST in which an additional 1683ms was needed and TANGIBLE an additional 891ms. It is interesting how relatively robust TANGIBLE is to occlusion when compared with the other techniques (see Figure 5).

An interaction between TECHNIQUE, TARGET type, and OCCLUSION is also significant for MT ($F_{8,5022} = 13.98$, $p < .0001$). Post-hoc tests show techniques behaved differently for most target types when occluded ($p < .001$), with the exception of RAYCAST and TABLE, and TANGIBLE and MID ($p > .065$). Looking at the largest differences in MT for each technique, VIEWPORT was 2595ms slower for LOW, followed by RAYCAST being 1835ms slower for LARGE, and finally TANGIBLE was 1323ms slower for LOW. Unsurprisingly, the overall trend between MT and occlusion resulted in longer selection times when targets

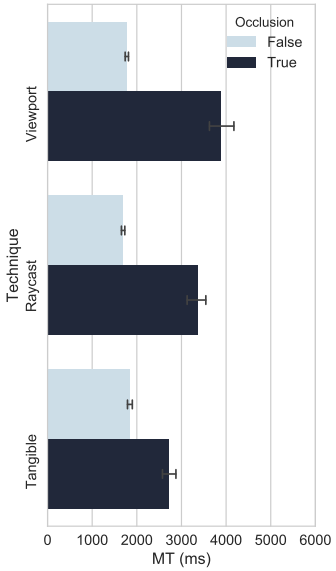

Figure 5: MT for occluded and non-occluded targets by TECHNIQUE.

were occluded; this is reported across TECHNIQUE and TARGET type. There was no significant interaction between TECHNIQUE and OCCLUSION on error rate ($F_{2,34} = 1.77$, $p = 0.18$).

Discussion

We discuss our findings and effects of self occlusion.

Prefer Raycast or Tangible

In the discussion of our experiment results, we were cautious to recommend viewport, except for some subjective preference of certain target types, there was no clear reason to choose it over raycast or tangible for a given target context. In general, participants appeared to be reluctant to adjust their physical proximity in relation to the currently active target. Rather, they would stand away from the target and continuously attempt at selection through rapid touch repetition. Even though participants had the the opportunity move closer and expand the targets relative size within the frustum of the camera, they tended to resort to brute force selection. In a post-experiment survey, only 7 out of 18 participants claimed to prefer viewport over the others, and 5 explicitly stated they disliked it. Overall, it is more likely that each technique has advantages and disadvantage when used in a dynamic and large SAR environment.

Projector Self Occlusion

An interesting consequence of our projector-based SAR environment are situations were a person self-occluded the projection of a target with their body. This primarily occurred on a small subset of the low target types. From our observations, this seemed to be an issue for the raycast and tangible techniques exclusively. Over time, when participants would suspect that self-occlusion was occurring, they would sway back and forth as a means to glimpse at the occluded target underneath. Some others would try to completely avoid self-occlusion, even standing in awkward or inefficient viewing locations to prevent it. However, our observations of participants using viewport suggested it is not limited by self occlusion. Most participants engaged in what Rohs and Oulasvirta call virtual pointing [7]. Since all targets were visible through the viewport of the mobile phone, self occlusion was a non-issue. Only when the user was unable to immediately find the target through the viewport did they resort to physically scanning the room.

Conclusion and Future Work

Our results show how effective raycast is at pointing, and how versatile a simple method like tangible can be.

The targets in our experiment were chosen to replicate realistic scenarios that may be encountered in future SAR environments, and we intentionally permitted free movement in the space. With many targets and a random order, we believe our results generalize, but we plan to further test these techniques using a more classically controlled target positions and sizes. The challenge is that the ISO 9241-411 task does not translate into SAR, so alternative methods need to be discovered. We are also investigating a predictive model for mobile phone pointing in SAR. Our results suggest target occlusion could be an important parameter.

A hybrid raycast and tangible pointing technique could be created where the user can either tap targets directly with the mobile phone or use the mobile phone to point at them from a distance. Pointing is only one type of SAR interaction. We also plan to investigate rotate-scale-translate (RST) input using a mobile phone in SAR. Ultimately, our goal is to combine the best pointing and RST techniques into a mobile phone-SAR system where users can place and adjust digital content literally anywhere.

REFERENCES

1. Sebastian Boring, Dominikus Baur, Andreas Butz, Sean Gustafson, and Patrick Baudisch. 2010. Touch projector: mobile interaction through video. In *SIGCHI Conference on Human Factors in Computing Systems (CHI'10)*. ACM Press, New York, New York, USA, 2287–2296. DOI: http://dx.doi.org/10.1145/1753326.1753671

2. Andrew Bragdon, Rob DeLine, Ken Hinckley, and Meredith Ringel Morris. 2011. Code space: touch + air gesture hybrid interactions for supporting developer meetings. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '11*, Vol. 16. ACM Press, New York, New York, USA, 212. DOI: http://dx.doi.org/10.1145/2076354.2076393

3. Robert Hardy and Enrico Rukzio. 2008. Touch & interact. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services - MobileHCI '08*. ACM Press, New York, New York, USA, 245. DOI: http://dx.doi.org/10.1145/1409240.1409267

4. Juan David Hincapié-Ramos, Kasim Ozacar, Pourang P. Irani, and Yoshifumi Kitamura. 2015. GyroWand: IMU-based Raycasting for Augmented Reality Head-Mounted Displays. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction - SUI '15*. ACM Press, New York, New York, USA, 89–98. DOI:http://dx.doi.org/10.1145/2788940.2788947

5. Brett Jones, Lior Shapira, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, and Nikunj Raghuvanshi. 2014. RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-camera Units. *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14* (2014), 637–644. DOI: http://dx.doi.org/10.1145/2642918.2647383

6. Ramesh Raskar, Greg Welch, and Henry Fuchs. 1998. *Spatially Augmented Reality*. A. K. Peters, Ltd., Natick, MA, USA. 63–72 pages. http://dl.acm.org/citation.cfm?id=322690.322696

7. Michael Rohs and Antti Oulasvirta. 2008. Target acquisition with camera phones when used as magic lenses. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press, New York, New York, USA, 1409. DOI:http://dx.doi.org/10.1145/1357054.1357275

8. Dominik Schmidt, Julian Seifert, Enrico Rukzio, and Hans Gellersen. 2012. A cross-device interaction style for mobiles and surfaces. In *Proceedings of the Designing Interactive Systems Conference on - DIS '12*. ACM Press, New York, New York, USA, 318. DOI: http://dx.doi.org/10.1145/2317956.2318005

9. J Seifert, A Bayer, and E Rukzio. 2013. PointerPhone: Using mobile phones for direct pointing interactions with remote displays. (2013). DOI: http://dx.doi.org/10.1007/978-3-642-40477-1_2

10. Andrew Wilson and Steven Shafer. 2003. XWand: UI for intelligent spaces. In *Proceedings of the conference on Human factors in computing systems - CHI '03*. ACM Press, New York, New York, USA, 545. DOI: http://dx.doi.org/10.1145/642611.642706