Comparing Spatial Understanding Between Touch-Based and AR-Style Interaction

Jason Wither*
Nokia Research Center

Sean White
Nokia Research Center

Ronald Azuma
Nokia Research Center

ABSTRACT

There are currently two primary ways of viewing location specific information in-situ on hand-held mobile device screens: using a see-through augmented reality interface and using a touch-based interface with panoramas. The two approaches use fundamentally different interaction metaphors: an AR-style of interacting where the user holds up the device and physically moves it to change views of the world, and a touch-based technique where panorama navigation is independent of the physical world. We have investigated how this difference in interaction technique impacts a user's spatial understanding of the mixed reality world. Our study found that AR-style interaction provided better spatial understanding overall, while touch-based interaction changed the experience to have more similar characteristics to interaction in a separate virtual environment.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input Devices and Strategies

1 MOTIVATION AND CONTRIBUTION

Until recently, traditional live-camera-view AR was the only way to view virtual information in-situ across whole cities. However, in the last few years, there has been an explosion of street level panoramic imagery captured from vehicles by companies like NAVTEQ and Google. Presenting virtual information on this imagery in-situ (where the user stands at, or near, the location the panorama was taken from) is now a viable option for replacing the live camera view traditionally used in AR. Using panoramas with touch-based input in place of traditional AR could be advantageous because users would not be required to hold up the phone to look at the world around them, an interaction that can be socially awkward, and physically taxing.

With this in mind, we were curious to look for general, non-application specific differences between the two methods of interaction aside from the physical affordances. In a perfect world, AR would have some obvious advantages over panorama based interaction, such as the application being based on a live camera view rather than a previously captured panorama. This advantage is currently severely lessened because of non-perfect tracking, though. As we have shown in previous work [4], providing a tracked window into a panorama is often preferred to a live view because it can improve the perceived tracking when used with AR-style interaction.

In this work, we looked specifically for differences in spatial understanding between the physical world and on-device representation of the world when using AR-style interaction and touch-based interaction. To eliminate a confounding factor, we used a panoramic image as the background for both interaction techniques.

---

*e-mail: jason.wither@nokia.com

Figure 1: The indoor and outdoor locations for our study with users demonstrating the touch-based and AR-style interaction techniques.

2 RELATED WORK

Loomis and Knapp [2] have written a very relevant overview of work related to distance and orientation estimation. Of particular interest to our work is how spatial perception transfers from a virtual environment to the physical environment, since in some ways interacting with a touch-based panorama is like interacting in a virtual environment. A study by Bakker et al. [1] is probably most similar to ours in that they attempted to measure the effect of proprioceptive feedback on users' ability to estimate changes in orientation in a VE with no known landmarks. They found far more underestimation from users who could only use visual stimuli (in a virtual environment with no known landmarks) to estimate their rotation compared to those who physically turned. Our study differs from this in part because in a physical environment users are surrounded by distinct visual landmarks in all conditions.

3 STUDY DESCRIPTION

The primary goal of the study was to examine how the difference in interaction technique affected a user's ability to understand the spatial layout of the on-device representation of the world, and how that on-device world related to the physical world. We compared AR-style interaction where users looked directly at a tracked view of the world [4] to touch-based interaction, where users manipulated the image representing the world with touch. In both cases, the horizontal field of view of the virtual camera was $54°$, and all interaction was done on a Nokia N900 with an InterSense InertiaCube3 attached. To compare the two techniques we asked users to report angular measurements between locations shown on screen. These locations were shown by having a cursor move in a circle around the user (on the device), stopping at the given locations. From the initial calibration location the cursor first moved to a random starting location, stopped and turned green. From this location the cursor would turn red, move to another location, stop, turn green, and then repeat this process two more times. Users were expected to follow the cursor, using the prescribed interaction technique, and report the three angles between the four locations where the cursor stopped and turned green. From the random starting location, the target moved randomly either left or right 45, 90, 135 or 180 degrees for each step. Discrete angles were chosen so analysis could be done on user performance at different angles. Users were

not told that the movements were restricted to these discrete angles, so that user responses covered a continuous range.

Users were shown sets of five tests in a row, alternating sets between conditions (order alternated between users), where each test contained the previously mentioned three recall angles. The first set with each condition was used as a practice set, after which users completed two sets with each condition. All users first participated in the test in a space well-known to them, a large room in our lab space which had a large number of prominent visual features. Users were then taken to an outdoor, less well known, location in the center of a park with fewer prominent landmarks around them and asked to complete two more sets of tests with each technique. Both locations can be seen in Figure 1. We chose to test all users indoors first for convenience. Because there was no feedback process, the order of location should not have affected outcomes, provided users were given sufficient time for training.

The study was completed with 9 users, all Nokia Research Center employees between the ages of 25 and 37 who were all male. Due to the lack of female participants in our study we cannot claim that our results also extend to them, since it has been shown that their spatial reasoning differs significantly from men's [3].

4 RESULTS AND DISCUSSION

When conducting the study, we did not provide users with any instructions on how they were supposed to perform the task in order to avoid influencing their spatial memory choices. Because of this, people ended up using very different strategies, which we have grouped into two categories based on post-study interviews. The first group employed what we call a visual strategy. These users would visually remember landmarks where the previous stop of the cursor was and calculate the traversed angle on the fly, rarely if ever looking away from the device while performing the task. In essence, they performed the entire task within the world of the device rather than in the physical world around them.

The second group of users used variations of what we have termed physical strategies. These strategies were more varied than those used by the visual group, but in all cases users tried estimate all angles in the physical space around them while using both the AR and touch-based interaction styles. One approach to this was for users to remember the physical locations the cursor stopped at, rather than the angle traversed, and then look around in the physical world after the cursor had stopped and calculate the three angles using their memory of those locations. Overall five users chose to use a visual strategy, while four used a physical strategy. We performed a between subjects ANOVA on user strategy, which showed a significant difference regardless of interaction technique ($F(1,1078) = 7.01$, $p = 0.008$) with visual strategy users averaging $14.9°$ of absolute error while physical strategy users averaged $18.2°$.

4.1 Hypotheses and Results

Hypothesis: *Users will make more accurate angle estimations with AR-style interaction than with touch-based interaction.*

Between all users we found a significant difference in interaction technique (mixed design ANOVA $F(1,7) = 6.06$, $p = 0.043$). This result was also impacted by individual user's choice of strategy. Users of a physical strategy showed no significant differences in absolute error between interaction techniques (within subjects ANOVA $F(1,3) = 0.20$, $p = 0.69$), while users of a visual strategy showed a much greater difference between interaction techniques (AR-style = $12.06°$ of error, touch-based = $17.70°$). This difference was nearly significant ($F(1,4) = 6.95$, $p = 0.058$), and possibly would have been significant with a larger user population. The effect was significant when looking just at the indoor case ($F(1,4) = 9.00$, $p = 0.040$), and was also significant across both locations ($F(1,4) = 7.86$, $p = 0.049$) when looking at signed error instead of unsigned error.
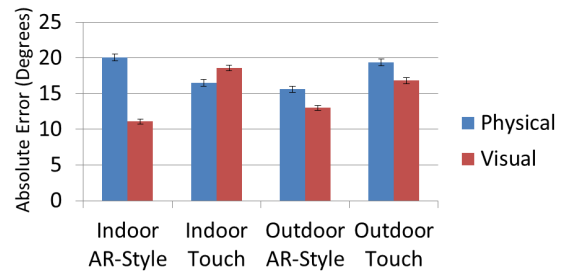

Figure 2: Overall absolute error rates are shown grouped by location and interaction style. Differences between user strategies are shown for each category.

Hypothesis: *The outdoor location will reduce user accuracy.*

We thought this would be the case because the outdoor surrounding environment had fewer visual features. In spite of the small difference near the border of significance in the above case there were no overall significant differences between the indoor and outdoor locations ($F(1,1078) = 0.24$, $p = 0.63$). The overall means were $16.66°$ of absolute error indoors and $16.05°$ outdoors. There were also no significant interactions between location and either strategy or interaction style.

Hypothesis: *Touch-based interaction will lead to greater underestimation of angle values, particularly for larger angles.*

We thought this would be the case because participants were interacting more with a co-located virtual space than the real surrounding environment, and underestimation has previously been shown in virtual environments [5]. When looking at signed error, some amount of underestimation occurred in nearly all conditions. There was a significant difference between angles ($F(3,5) = 14.12$, $p = 0.007$) with underestimation peaking at $135°$. However, users with a physical strategy showed no significant differences in error between the various angles ($F(3,9) = 1.12$, $p = 0.39$), likely because the calculations they were doing were external to the rotation itself, instead being done statically in the physical environment. Users with a visual strategy, on the other hand, had significantly more error with larger angles ($F(3,12) = 5.50$, $p = 0.013$), and had significantly less underestimation when using AR-style interaction than they did when using touch-based interaction ($F(1,592) = 13.72$, $p = 0.0002$). This would suggest that there was a much stronger connection to the physical world for those users when using AR-style interaction. The touch-based interaction on the other hand had characteristics suggestive of interaction in a colocated virtual world.

5 CONCLUSIONS

The evaluation presented in this paper suggests that AR-style interaction has some benefits over touch-based interaction when used in-situ with panoramas. These benefits include increased spatial understanding and a more connected, less virtual MR world.

REFERENCES

[1] N. H. Bakker, P. J. Werkhoven, and P. O. Passenier. Aiding orientation performance in virtual environments with proprioceptive feedback. In *VRAIS '98: Proceedings of the Virtual Reality Annual International Symposium*, 1998.

[2] J. M. Loomis and J. M. Knapp. Visual perception of egocentric distance in real and virtual environments. *Virtual and Adaptive Environments*, pages 21–46, 2003.

[3] M. Masters and B. Sanders. Is the gender difference in mental rotation disappearing? *Behavior Genetics*, 23:337–341, 1993.

[4] J. Wither, Y. Tsai, and R. Azuma. Indirect augmented reality. *Computers & Graphics*, In Press, Accepted Manuscript, 2011.

[5] B. G. Witmer and P. B. Kline. Judging perceived and traversed distance in virtual environments. *Presence: Teleoperators and Virtual Environments*, 7(2):144–167, 1998.