

Experiences with the Impact of Tracking Technology in Mobile Augmented Reality Evaluations

Alessandro Mulloni^{1,2}, Jens Grubert¹, Hartmut Seichter^{1,2}, Tobias Langlotz^{1,2}, Raphael Grasset^{1,2},
Gerhard Reitmayr¹, Dieter Schmalstieg^{1,2}

¹Institute for Computer Graphics and Vision, Graz University of Technology

²Christian Doppler Laboratory for Handheld Augmented Reality

[mulloni | grubert | seichter | langlotz | grasset | reitmayr | schmalstieg] @ icg.tugraz.at

ABSTRACT

In this paper, we discuss the impact of tracking technology on user studies of mobile augmented reality applications. We present findings from several of our previous publications in the field, discussing how tracking technology can impact, influence and compromise experimental results. Lessons learned from our experience show that suitable tracking technology is a key requirement and a fundamental factor in the user experience of the application. Tracking technology should therefore be considered not only during implementation but also as a factor in the design and evaluation phases.

Author Keywords

tracking, mobile augmented reality, evaluation

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. H.5.2 [Interfaces and Presentation]: User Interfaces - Evaluation/methodology

General Terms

Performance, Design, Experimentation, Human Factors

INTRODUCTION

In the last decade, the tracking technology used for mobile Augmented Reality (AR) applications became increasingly viable and robust. In particular, vision-based tracking offers accurate real-time registration in AR applications. Wagner et al. [11] were the first to show accurate real-time natural-feature tracking of planar targets on mobile phones, in 2008. Just a few years later, various commercial libraries have brought this technology outside the research labs, such as Qualcomm's Vuforia¹. Such libraries enable developers, who do not have advanced computer-vision skills, to rapidly develop advanced AR applications with robust state-of-the-art tracking.

In contrast to vision-based tracking, the sensors available on mobile devices do not provide a level of accuracy suitable for AR applications. Yet, many AR applications on mobile phones still employ sensor-based tracking, and typically produce inaccurate and unstable AR. This is particularly true for applications that operate in large-scale environments, such as AR browsers or navigation systems, because current vision-based tracking technology cannot support continuous tracking in such scenarios.

The gap between state-of-the-art vision-based tracking used in research labs and the much more inaccurate sensor-based tracking adopted in many commercial projects is a theme of frequent discussions within the AR research community. Understanding that a reliable tracking implementation is fundamental for making convincing AR is an established consideration that has a large support within the AR community. Tracking accuracy has been evaluated from technological [6] and psychological viewpoints [7]. Only little work takes into account the factors of the tracking technology in the interface design [5]. To the best of our knowledge, none considers the impact of tracking on user studies.

In this paper, we argue for the importance of *considering tracking at all stages of the design process of a handheld AR system*. Besides considerations on the implementation, we stress to not disregard tracking inaccuracies in the interface design, as well as a factor in user evaluations. We support our claim with a number of case studies from several of our previous publications in the field.

CASE STUDIES

Our research group focuses on mobile AR, bridging technical advancements with user-centered design of a variety of applications. For this goal, we require a sufficient level of robustness that allows for real-world evaluations of AR applications with end users. Within the scope of such real-world evaluations, it is unreasonable to ignore how tracking can be a confounding factor or how the interface design supports users if tracking fails. In the past years, we thus faced the challenges of implementing novel robust tracking technologies for mobile devices.

In the following, we give an overview of some of our user studies, for which tracking had an impact. We first report on evaluations of our own prototypes for wayfinding, augmented maps and augmented games. We then report on a survey of user feedback on existing AR browser products.

Wayfinding

In [10], we conducted an exploratory evaluation on user adoption of map and AR, for outdoor wayfinding. We implemented a multimodal navigation system (Figure 1, left): we provided a *forward-up map*, highlighting the user's position and the path to be followed, and hints as *glyphs* and as *audio instructions*, to support eye-free usage. We integrated an on-demand *AR interface*, augmenting the environment with virtual arrows indicating the direction to follow. Similarly to most commercial systems, we tracked

¹ Qualcomm's Vuforia: <http://www.qualcomm.com/ar>



Figure 1. Our prototypes for outdoor wayfinding (left) [10] and indoor wayfinding (right) [9].

users’ position with GPS, and orientation with compass and magnetometer. We also used vision-based tracking to stabilize the augmentations if users stood still.

The goal of our evaluation was to see *where* and *how* people exploit AR during outdoor wayfinding, and when in contrast the map is preferred to AR. Our results show that users exploit AR mostly in proximity of road intersections: these are therefore the most important locations to support with AR. However, we saw a low adoption level for AR, justified by participants both because the map was sufficient, more familiar and gave a better overview of the path, but also because *the AR visualization was not sufficiently stable*. Overall, we observed that inaccurate tracking caused a loss of trust of the users in the interface, and this likely impacted on the level of adoption of the interface.

We also observed that participants dealt with inaccuracies by choosing more “robust” and therefore more trusted interfaces. This was the case of a map, in the case of outdoor wayfinding, but we noticed a similar effect in another experiment we conducted on indoor wayfinding [9], where we also evaluated a prototype that combined maps with AR arrows (Figure 1, right). In both cases, we observed that participants coped with tracking errors by choosing more informative interfaces – this is in line with previous findings of Butz et al. [1] and Hallaway et al. [5], who also suggest increasing the informativeness of the interface when the system has a higher tracking uncertainty.

Due to this confounding factor, it is not possible to clearly understand if a low adoption level of AR is an actual effect of the better suitability of another interface for the task, or an effect of the particular tracking implementation used in the experiment. The scope of validity of such results is thus bound to the technology used (for example, map vs. *sensor-based* AR) and researchers should be careful in generalizing them to higher-level questions (for example, map vs. AR).

Finally, participants interpreted misplacements of the AR arrows due to tracking inaccuracy as intentional wayfinding instructions. For example, one participant interpreted the positional error of a left-turn arrow as an instruction to first cross the street, and only then turn left onto the opposite pavement. Another participant interpreted orientation errors of a straight-forward arrow (Figure 1, third image) as instructions to leave the pavement and walk in the middle of the street, or to walk back onto the pavement. The affor-

dance of the arrow caused expectations of a highly accurate visualization: our naïve design did not communicate tracking uncertainty, misleading users to the point of convincing them to walk in the middle of the street, rather than on the pavement. *Visualizations should communicate tracking uncertainty more clearly*, as done for example by Google Maps for GPS (using variable-radius circles) or by Coelho et al. [2] in the context of AR. Overall, one should consider that user errors might depend on a particular visualization choice rather than being generally due to the AR interface.

Exploring augmented maps

In [8], we presented MapLens (Figure 2), a prototype that uses our natural-feature tracking technology [11] to augment paper maps with digital content retrieved from an online source. We conducted two exploratory evaluations with 74 distinct participants (37 for each study) divided into 3-person teams. In the context of our experiments, the digital content consisted of game clues necessary for solving game riddles. During the two experiments, we observed how people operate the augmented map in teams to play the game.

In the first experiment, tracking operated at 5–12 frames per second, at a distance of 15–40 cm between map and device and for tilt angles within $\pm 30^\circ$ from the perpendicular view. We observed how the map fostered *place making*: the physical map acted as a place to discuss and reach joint understanding on game strategies. Yet, we also observed that tracking technology constrained the possibilities for place making: participants needed to stabilize the physical map and the device to be able to use MapLens robustly, and therefore favored places where they could lay down the map, for example on a table or on the floor.

For the second experiment we improved the performance of tracking in term of computation time and accuracy: it operated at 16–20 frames per second, at a distance between 10 cm and 2 m and for tilt angles up to almost 90° . We observed again a form of place making, but the new tracker allowed for more agile and spontaneous behaviors: we observed a more agile place making – stopping briefly to check a detail before moving on – emerging aside standing for longer periods of time or setting down the map. Indeed, many participants never set down the map but always used MapLens while on the move.

Overall, in the sequence of two experiments we observed that *different tracking implementations have an influence*



Figure 2. Our prototype for augmented maps (left), presented in [8], and in use by some of our study participants (right).

on how people operate our AR interface. People adapt to operating MapLens in ways that make the technology work robustly. When conducting experiments that look at how people use an AR system, it is therefore important to consider how different ways of using the interface are influenced, or even imposed, by the chosen tracking technology.

Gaming with a Magic Lens Interface

In [4], we wanted to observe how people employ AR and static peephole² interfaces for games in public spaces, and if they can solve game tasks with either interface. To identify factors specific to the public setting, a control group conducted the tasks in a lab. We designed a find-and-select task with background music, audio and graphical effects to create a game experience lasting approximately 1 minute per level. Participants played 8 levels: levels did not increase in difficulty, and showed similar views on the game to lower the mental gap when switching between them. Participants were free to use any of the interfaces for playing the game. Participants completed a learning phase ahead of the actual task and were asked to practice avoiding tracking errors and recovering from them. They could practice until they felt comfortable with the interface. We report on pooled results of the public and lab setting for 16 participants.

We used Qualcomm’s Vuforia for tracking. Tracking failed a number of times during the game: in total, 267 tracking errors occurred (6 outliers removed). Tracking errors lasted 9% of the overall time that participants spent in AR mode (which was used 72% of the gaming time). The median duration of tracking errors was 1.8 seconds (1st quartile: 0.8, 3rd quartile: 2.6). Figure 3 shows the total number of errors per level. Despite the learning phase, the number of tracking errors per level decreased over the course of playing the game. The strong negative correlation between level and number of tracking errors (Kendall’s $\tau = -0.82$) suggests that participants were still learning to cope with tracking errors over time, and the learning phase might not have been sufficient to learn how to avoid tracking errors.

Most tracking errors were resolved in a short time directly in the AR mode. However, in 34 occasions participants tilted the phone downwards and upwards at least once, to re-initialize the tracking (a technique shown by the instructor). Furthermore, 12 participants changed their hand poses over time. While there might be several causes for changing hand poses (e.g., fatigue), when interviewed 3 participants mentioned explicitly to have changed their hand poses in order to stabilize tracking (see Figure 4, left and middle). We also observed that 4 out of the 6 participants who had more than 20 tracking errors stabilized the phone only at its bottom resulting in increased camera shake when touching its surface (see Figure 4, right). Twelve out of all the 16 participants mentioned issues with the tracking robustness when asked about their usage and preferences of the inter-

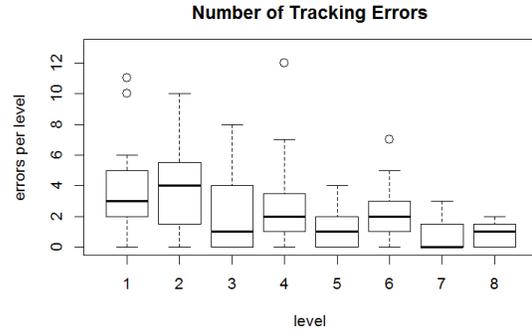


Figure 3. Number of tracking errors per level.



Figure 4: Participant changes hand pose to stabilize tracking (left/middle). Participant with many tracking errors holds the phone at its bottom (right), causing camera shake on touching.

faces. User commented that “if you have to look for [re-initializing the] tracking again and again it is not as much fun as if it is stable”, “I noticed that the image was lost from time to time”, “I moved slower than I would like to”, “it was more shaky”. One participant explicitly mentioned that “the tracking errors did not distract me much”. Several participants used the static-peephole view as a fallback solution if tracking did not work. One participant said “In the live [AR] mode it was hard to hit the monsters at the bottom [of the poster]. Then I realized that I can just tilt the phone down to catch them in the other mode”. However, despite the tracking error occurrence participants still used the AR view significantly longer than the static-peephole view.

The study did not explicitly focus on how tracking influenced user adoption. Hence, it was difficult to distinguish all causes that eventually led to changing user behavior. For example, we could not always distinguish if users employed the static-peephole view or changed hand poses due to tracking errors, due to fatigue effects, or because of other reasons. Furthermore, the analysis of the tracking data showed that *learning how to cope with tracking errors happened throughout the study*, leading to a decreasing number of tracking errors as the game progressed.

AR Browser User Feedback

In [3], we conducted an online survey and app-store analysis to study real-world adoption of AR browsers. In the survey we asked 77 end-users to provide reasons for dropping their usage of AR browsers, if they did so. We also analyzed 1135 comments on the Apple and Android app stores. Eight survey participants (10%) mentioned tracking issues as a reason to stop using AR browsers, while 18 participants mentioned tracking as a feature to improve in the future. In the app stores, 573 (50.5%) comments had a nega-

² A static peephole is an interface, in which the view can be panned and scrolled manually by the user.

tive connotation, 97 of which (8.5%) touched tracking issues. Our survey and app-store analysis indicate that about 10% real-world AR-browser users stop using the application due to bad experiences related to tracking errors.

Few comments in the survey related to the technology, e.g. “not so reliable. Often the compass and the GPS doesn’t work”. Most comments rather referred to the consequent inaccuracy in the placement of the augmentations. Survey participants commented about the “lack of relevance to physical surroundings” of information, which was “not useful as it was not spatially accurate”. Requests for improvement were also largely in this direction, asking for “better location accuracy, robust POI display” or “[...] better overlay on real world objects”. This is also backed by the comments in the app stores such as “you expect to place a 3D dinosaur in a car parking spot, and the best you get is a floating icon wafting around in the general area.” Finally, some survey comments were about instability and jitter in the visualization, e.g. “find a way to calm down the jumpiness [...]”. Similar comments occur also in the app stores: “how do you read the text when it’s jumping all over the screen [...]”, or “[...] pictures float around with no real indication as to what to do with the info. Might as well just use Google maps.” Overall, the comments suggest that *bad tracking can compromise both the usability of the application and the user experience through ill-placed information and jitter.*

DISCUSSION AND CONCLUSION

The examples presented in this paper show that tracking technology has an impact on user adoption of mobile AR interfaces. One must carefully analyze the different possible causes for user behavior and usage patterns, including those bound to the specific characteristics of the adopted tracking technology. Consequently, study results about user adoption of mobile (sensor- or vision-based) AR systems can not be easily generalized to the whole field of mobile AR. This is particularly important when comparing AR with other interfaces which use a different tracking technology, or provide a visualization that is more robust to small tracking inaccuracies (such as a 2D map). Borrowing the clear words from a peer review we received, if we do not consider the impact on the study results of the specific tracking technology we employ, “*the results of such studies will most probably not live longer than the current generation of mobile phones.*”

Summing up, we can reflect on a number of lessons learned from our experiences with mobile AR evaluations.

Consider tracking as a factor in mobile AR studies, already in the design phase of the experiment. Our original research questions for the presented studies did not specifically focus on the influence of tracking errors on user behavior. This might be one reason why we could not distinguish if certain user behaviors were due to tracking or other causes. Our experiments show that tracking can change usage patterns, or even cause users to adopt a different interface than AR.

Reflect tracking accuracy in the interface: this is in line with the generic HCI principle of communicating the state

of the application clearly in the interface. Also *provide a fallback solution* if tracking is inaccurate or does not work at all. While one might limit the influence of tracking errors in lab environments, it is much more challenging to control the tracking accuracy in mobile AR interfaces in the field. This is particularly true for systems deployed “in the wild”, such as commercial AR applications. Ultimately, enforcing AR interfaces despite insufficient tracking quality will cause users to distrust it.

Short learning phases might not be sufficient for users to learn how to cope with tracking errors. Learning effects related to the usage of tracking technology should be considered when analyzing experimental results. The learning curve should be also considered when deploying AR applications.

Overall, our experience shows that tracking issues can be a confounding factor when evaluating mobile AR interfaces, and a disruptive factor when deploying them to end users. While it is worthwhile to improve the accuracy and stability of tracking systems for future interfaces, one should also try to account for the influence of tracking errors on interface design, experimental results, as well as on users’ adoptions of mobile AR applications.

ACKNOWLEDGMENTS

This work was supported by the Christian Doppler Laboratory for Handheld AR, the EU funded projects IPCity (FP6-IST) and MARCUS (FP7-PEOPLE-IRSES), and by the Austrian National Research Funding Agency (FFG) in the SmartReality project.

REFERENCES

1. Butz, A., Baus, J., Krüger, A., and Lohse, M. A hybrid indoor navigation system. *Proceedings of IUI* (2001), 25–32.
2. Coelho, E.M., MacIntyre, B., and Julier, S.J. Supporting interaction in augmented reality in the presence of uncertain spatial knowledge. *Proceedings of UIST* (2005), 111–114.
3. Grubert, J., Langlotz, T., and Grasset, R. Augmented reality browser survey. *Technical report 1101, ICG, University of Technology Graz, Austria*, (2011).
4. Grubert, J., Morrison, A., Munz, H., and Reitmayr, G. Playing it Real: Magic Lens and Static Peephole Interfaces for Games in a Public Space. *Proceedings of MobileHCI*, (2012).
5. Hallaway, D., Feiner, S., and Höllerer, T. Bridging the gaps: hybrid tracking for adaptive mobile augmented reality. *Applied Artificial Intelligence* 18, 6 (2004), 477–500.
6. Lieberknecht, S., Benhimane, S., Meier, P., and Navab, N. A dataset and evaluation methodology for template-based tracking algorithms. *ISMAR* (2009), 145–151.
7. Livingston, M.A. and Ai, Z. The effect of registration error on tracking distant augmented objects. *ISMAR* (2008), 77–86.
8. Morrison, A., Mulloni, A., Lemmelä, S., et al. Collaborative use of mobile augmented reality with paper maps. *Computers & Graphics* 35, 4 (2011), 789–799.
9. Mulloni, A., Seichter, H., and Schmalstieg, D. Handheld Augmented Reality Indoor Navigation with Activity-Based Instructions. *Proceedings of MobileHCI*, (2011), 212–220.
10. Mulloni, A., Seichter, H., and Schmalstieg, Dieter. Enhancing Handheld Navigation Systems with Augmented Reality. *Proceedings of MobileHCI, Workshop on Mobile AR*, (2011).
11. Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., and Schmalstieg, D. Pose tracking from natural features on mobile phones. *ISMAR* (2008), 125–134