MarkAirs: Around-Device Interactions with Tablets Using Fiducial Markers – An Evaluation of Precision Tasks

Fernando Garcia-Sanjuan
ISSI/DSIC, Universitat Politècnica de València, Valencia (UPV), Spain
fegarcia@dsic.upv.es

Javier Jaen
ISSI/DSIC, Universitat Politècnica de València, Valencia (UPV), Spain
fjaen@upv.es

Geraldine Fitzpatrick
HCI Group, University of Technology (TU Wien), Vienna, Austria
geraldine.fitzpatrick@tuwien.ac.at

Alejandro Catala
LoUISE/I3A, University of Castilla-La Mancha (UCLM), Albacete, Spain
alejandro.catala@uclm.es

CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA
ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2892486

Abstract
This paper evaluates MarkAirs, an interaction technique that uses fiducial markers to perform mid-air interactions. MarkAirs offers several advantages: the proposed technique does not require any tracking external hardware other than the front camera of a mobile device; it is robust even when the markers are partially occluded; and it enables precise 2D manipulations (translation, rotation and scaling). An evaluation study points to the feasibility and precision of the proposed technique and the perceived usability and subjective workload impressions of the participants.

Author Keywords
Around-Device Interactions (ADI); mid-air interactions; fiducial markers; tablets; evaluation; precision

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User Interfaces – Input Devices and Strategies, Interaction Styles

Introduction
Handheld devices are becoming widespread and offer several advantages with respect to tabletops in terms of cost, scalability, and mobility to build collaboration spaces [5,13]. In these scenarios, mid-air or Around-

Figure 1. Interaction with MarkAirs to control the position of an element. On the left, the fiducial marker used.

Device Interactions (ADIs) could be used to enable manipulations in the available 3D space above the surfaces. However, when ADIs have been implemented in the past they have required pre-installed specialized hardware that limited the choice of where these interactions could take place, have not been accurate enough to perform high precision manipulations of objects on the devices, or have relied on hand and finger gestures which do not allow the identification of the users.

In this work, we present MarkAirs (see Figure 1), an interaction technique that makes use of the built-in front camera of handheld devices and fiducial markers to enable precise ADIs. We do not only show that interactions above handhelds are realizable without any external specialized hardware, but also evaluate the precision of basic 2D manipulations (translations, rotations, and scaling) on digital objects and the perceived subjective usability and workload of the proposed technique.

Related Work

Previous works have explored ADIs with tabletops. These exploit the space above the surface either to explore 3D virtual spaces (e.g., [16]) or to reach and manipulate distant elements (e.g., [1]). These high-precision tasks, however, often rely on complex hardware settings composed of several external cameras, projectors, or gloves with reflective markers like the ones used in motion capture (e.g., [12]).

Other approaches rely on embedded cameras that are installed within tabletops to recognize aerial interactions. Hilliges et al. [8], for example, use an “enhanced” tabletop and make use of hands to perform gestures in the air to manipulate 3D digital objects rendered on a tabletop. However, even though hands enable natural interactions, they do not allow the identification of the user who makes them. To solve this problem, Gallardo and Jordà [4] propose the use of fiducial markers for such interactions. However, their approach requires expensive specially designed hardware to control the degree of transparency of the surface, so that image projection and gesture detection are interleaved. Moreover, their markers cannot be occluded for the system to work properly.

Several works have explored ADIs for handheld devices using external cameras or depth sensors with the purpose of performing 3D rotations using hand gestures around the handheld (e.g., [11]), reducing the occlusion produced by touch contacts on the screen (e.g., [9]), or mitigating the cluttering of the surface by storing/retrieving digital elements to/from the space around the device (e.g., [7]). These approaches allow performing manipulations with a certain precision but, again, require careful installation of external hardware. Others reduce the hardware complexity by making use of the built-in sensors in the tablets. Ketabdar et al. [10], for instance, exploit the magnetic (compass) sensor of the device and enable in-air interactions using magnets. Unlike optical approaches, this solution is more robust to occlusion, but it is less precise and the system is not capable of differentiating between different magnets since they do not have an encoded ID. In contrast, others rely on the built-in camera of the devices to detect hand gestures (e.g., [15]) or fiducial markers (e.g., [5]). Whereas the former avoids any additional equipment, the use of markers enables the identification of the user as well as detecting as many actions as markers available, which are considerably higher than the gestures that can be achieved with hands only. However, the use of frame markers in [5] is not robust to the occlusion of the

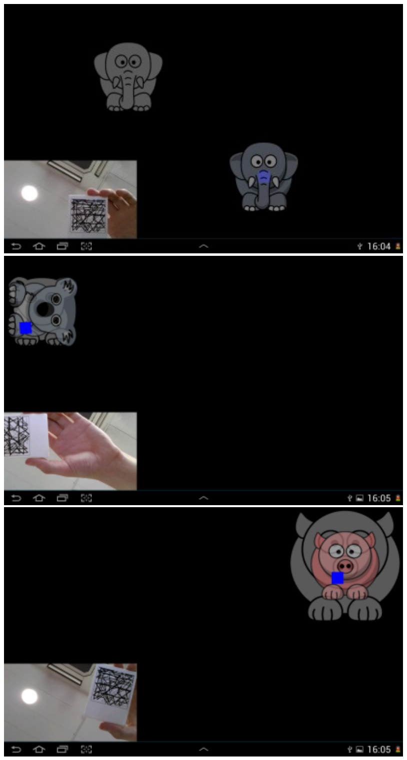

Figure 2. Manipulations evaluated: translation (top), rotation (center), and scaling (bottom).

marker and the precision of the technique to perform fine-grained manipulations is not evaluated.

Interacting with MarkAirs

As explained in [6], MarkAirs is an ADI involving fiducial markers that are recognized by a tablet's front camera (see Figure 1 and Figure 2). These markers can be attached to physical cards or objects or even displayed on other digital devices, and consist of arbitrary drawings that are tracked by Vuforia™'s natural feature tracking algorithms. They present several advantages with respect to other more classical kind of markers (e.g., the ones used in [4,5]): a) they can be tracked even if they are partially occluded (see the bottom-left corner of the tablet in Figure 1, which shows the image captured by the device's camera); b) one can have a virtually infinite number of markers; and c) a marker can be a meaningful photograph, which can be related to the digital information content that is associated with the card (e.g., if we want a card to contain a game character, the marker can be a picture of the character itself).

In this work, the marker's 3D position and orientation are mapped to 2D manipulations as follows: translation (marker's position in the XY axes), scaling up/down (marker's position in the Z axis), and rotation (marker's yaw angle).

Evaluation

The following reports on an evaluation study of the use and usability of MarkAirs as a potential ADI technique. This study had three goals: First, to determine the level of precision that can be achieved both in the termination phase of the proposed 2D manipulations (when reaching a final desired position/orientation/size), and in the execution phase (when translating/rotating/scaling at a given pace or following a specific trajectory). Second, to assess whether there are differences in the precision achieved when these manipulations are performed in isolation (first translation, then rotation, and finally scaling) and when all three are performed conjointly in the same task. And, finally, to obtain insights into the perceived subjective usability and workload of the proposed interaction.

Participants and Apparatus

Twenty-four volunteers participated in this study, fifteen of whom were males, with ages ranging from 14 to 60 ($M = 34.54$, $SD = 12.30$). The experiment was conducted with a 10.1" Android tablet having a resolution of 1280x800 pixels, and running Vuforia™'s computer vision algorithms. A 5x5cm marker was attached to a 6x9cm piece of cardboard, as depicted in Figure 1 (left).

Tasks

Eight different tasks were performed (see supplementary video[1]), three to evaluate the precision of each 2D manipulation in the termination phase, three to evaluate the precision in the execution phase, and two to evaluate the precision of all three manipulations being performed conjointly in the termination and the execution phase, respectively. In all tasks there is a target 2D object (O_T) to be manipulated by means of ADIs and a reference 2D object (O_R) whose random position, orientation or size must be reached (see Figure 2).

The termination-precision tasks are carried out in three steps. First is an *acquisition* step in which the

[1] https://www.dropbox.com/s/0ewgy3xw8ri1u1f/MarkAirs.avi

participant must use the marker to control a small squared cursor on the screen and acquire O_T. Once acquired, the task moves on to a *manipulation* step, when the corresponding manipulation must be performed to make O_T reach O_R's position, orientation, and/or size (depending on each task). Once the system determines that O_T and O_R match within a certain error margin, i.e., at least 90% overlap (position/size) and/or their respective orientations differ less than 10º, the task moves on to a *precision measuring* step. In this step the participant must maintain O_T as close as possible to O_R (which remains static) for 5 seconds while the system measures precision (which will be described below).

The execution-precision tasks are on the other hand conducted in two steps. First is an *acquisition* step, as described above. Second, once O_T is acquired, there is a *manipulation* step in which O_R continuously and dynamically changes its position, orientation, and/or size (depending on the task) for 40 seconds. The participant must continuously perform the manipulation to maintain O_T as close as possible to O_R while the system measures precision.

Procedure and Experimental Design

The participants stand in front of the tablet placed on a table. The first stage of the experiment is to explain the interactions to the users and give them some time to train in order to minimize learning effects. Once they feel familiar with the interaction, the experiment begins. Each user performs the three manipulations (translation, rotation, and scaling) with two different precision measurement situations or phases (termination and execution), first each manipulation separately, and then combined. In order to avoid order and carryover effects, the isolated manipulations are presented following a 3x3 balanced Latin Squares design. For each manipulation, the termination-precision task is presented first, followed by the execution-precision one.

The manipulation error during the *precision measuring* step is measured for each task as follows: the distance in pixels between the centers of O_T and O_R for the translation tasks, the difference between their orientation angles in degrees for rotation tasks, and the difference between their scaling factors in percentage for scaling tasks. This error is calculated as the average of a number of samples obtained continuously during the *precision measuring* step. When all three manipulations are performed conjointly, all three errors are measured. In addition, the subjective workload of each task is measured via the NASA-RTLX questionnaire [3]. Finally, a SUS questionnaire is administered post-tasks to evaluate the overall usability of the technique.

	Trans-lation error (px.)	**Rotat-ion error (deg.)**	**Scal-ing error (%)**
Termination phase	11.92 (4.45)	3.09 (2.84)	2.91 (1.68)
Execution phase	15.89 (5.82)	6.11 (4.49)	5.44 (3.32)

Table 1. Average (and standard deviation) errors for each manipulation performed separately (in columns) and phase (in rows).

Results

PRECISION ERRORS

For each 2D manipulation executed separately, the users performed a total of 72 (24 users x 3 manipulations) interactions in the termination-precision task and 72 in the execution-precision task, so that each 2D manipulation was performed 144 times. Table 1 shows the average (and standard deviation) error measures for each type of manipulation and phase. Additionally, a repeated measures ANOVA ($\alpha = 0.05$) showed a significant effect of the phase on the precision error for translation manipulations ($F_{1,71} = 51.029$, $p < 0.001$), rotations ($F_{1,71} = 23.055$, $p < 0.001$), and scaling ($F_{1,70} = 37.453$, $p < 0.001$), with the termination phase outperforming the execution one in terms of precision.

	Translation error (px.)	Rotation error (deg.)	Scaling error (%)
Termination phase	11.93 (5.53)	2.05 (2.92)	3.51 (2.24)
Execution phase	31.11 (12.75)	6.94 (3.71)	9.94 (5.12)

Table 2. Average (and standard deviation) errors for each manipulation performed conjointly with the other two in the same task (in columns) and phase (in rows).

Table 2 shows the average (and standard deviation) errors for each manipulation in the task where all three are performed conjointly, also both in the termination and execution phases. A repeated measures ANOVA ($\alpha = 0.05$) also showed a significant effect of the phase on the precision error for translation manipulations ($F_{1,71} = 155.358$, $p < 0.001$), rotations ($F_{1,71} = 83.857$, $p < 0.001$), and scaling ($F_{1,71}$ = , $p < 0.001$), with the termination phase outperforming the execution one.

Additionally, pairwise within-subject t-tests ($\alpha = 0.05$) were conducted to assess whether the errors obtained performing each manipulation individually differed from when all three were combined in the same task, showing only significant differences for the translation in the execution phase ($p < 0.001$), the rotation in the termination phase ($p = 0.036$), and the scaling in the execution phase ($p < 0.001$).

Qualitative Measures

NASA-RTLX scores for workload assessment were analyzed for the different manipulations both individually and all combined. In general, the users reported relatively low workload, rating all dimensions on average between 13.13 and 26.88 (0 being "very low" and 100 "very high") when each manipulation was performed individually. When performed conjointly, the workload slightly increased across all dimensions on average (rating between 27.92 and 37.40). Besides, differences were found between some of the manipulations; pairwise comparisons using Wilcoxon signed-rank tests with a Bonferroni adjustment ($\alpha = 0.05/C(4,2) = 0.05/6 = 0.008$) revealed that scaling manipulations were considered significantly less time-demanding than translations ($p = 0.001$), and also received significantly better performance scores ($p = 0.003$), and lower values for effort ($p = 0.001$) and frustration ($p = 0.003$). Not surprisingly, performing all three manipulations at once received worse workload assessments overall, except in two cases: no significant differences were found between only translating and performing all three manipulations in terms of performance ($p = 0.021$) and frustration ($p = 0.124$).

The phase in which the manipulation occurred (termination vs. execution) was also analyzed using Friedman's χ^2 tests for ranks (df = 1, $\alpha = 0.05$), and the results showed the participants perceiving manipulating at the termination phase with a static object less demanding in terms of workload in all NASA-RTLX dimensions ($p < 0.001$).

The general usability of MarkAirs, assessed via a SUS questionnaire, shows a total average score (calculated as in [2]) of 83.44 (SD = 13.31), which, according to Sauro [14] is above average (68). In his letter-grade system (from A+ to F), the overall usability of MarkAirs would receive an A.

Discussion

In terms of the precision of the proposed aerial interactions in the termination phase, the implemented technique offers acceptable levels for all three manipulations (translation, rotation and scaling) and it allows users to perform fine-grained interactions with average errors of 11.92 pixels, 3.09º, and 2.91%, respectively when only one manipulation is being carried out at a time. When users have to perform all three manipulations conjointly to move, rotate, and resize an element, the translation and scaling average errors increase slightly but insignificantly, and the rotation error in this case even decreases significantly to 2.05º (probably due to a learning effect since the task combining all three manipulations was performed last).

It is also interesting to note that the precision of the proposed interactions worsens in the execution phase, when the users are forced to adapt their behavior to a given dynamic trajectory, rotation, or size of the reference object. Moreover, when all manipulations are performed together, the translation and scaling operations are significantly less precise than when they are performed individually (average errors in the worst case of 31.11 pixels, 6.94º, and 9.94%). This means that the proposed technique, although supporting these additional dynamic execution requirements, would not be adequate for very high-precision 2D manipulations.

In terms of subjective perception, MarkAirs showed a very good SUS usability score (an A grade) and low workload levels, in general. Moreover, the scaling tasks received better subjective impressions with respect to time demands, performance, effort, and frustration, in comparison with the other two manipulations. This could be explained because scaling by means of a vertical arm movement is perceived as requiring little ergonomic effort. In addition, as it was expected, the users clearly expressed better subjective workload impressions of the tasks where the reference object was static. They found the precision requirement less demanding in the termination phase than in the execution phase.

Finally, it is also worth mentioning that, during the course of the experimental sessions, the users expressed very positive comments on the interaction technique. Specifically, they referred to it as being "very fluid", "relaxing", "fun", "entertaining" and "astonishing". Indeed, it was found to cause amusement, and even in the more challenging execution-precision tasks, the subjects expressed positive reactions when the reference object changed its position/rotation/size dynamically.

Conclusions and Future Work

This paper presented MarkAirs, an interaction technique that uses the built-in front camera of a tablet/smartphone and fiducial markers to perform translation, rotation and scaling ADIs. The precision of the proposed technique was evaluated in both the execution and termination phases. The results show that good precision levels can be achieved in all situations. The analysis of the participants' responses to SUS and NASA-RTLX questionnaires reveals that they consider it highly usable and undemanding in terms of workload. Observational data also revealed that MarkAirs was an engaging and entertaining experience for first-time users. In future work, we will use the 6-DOF information obtained by the tracking algorithms to explore new ways of making translations by performing roll and pitch rotations on the marker and "project" the cursor onto the screen, thus eliminating the need for arm movements in these tasks. We consider this interaction useful for multi-user environments since it could enable richer interactions using the 3D space above the surface as well as provide identification for the users' actions and also help reduce the clutter on screen by using the cards as containers of digital elements. Another future line of research will be to study MarkAirs in a collaborative context in a multi-display environment.

Acknowledgements

This work is funded by the European Development Regional Fund (EDRF-FEDER) and supported by Spanish MINECO with Project TIN2014-60077-R, by UPV under Project UPV-FE-2014-24, and by fellowship ACIF/2014/214 within the VALi+d program from GVA.

References

1. Amartya Banerjee, Jesse Burstyn, Audrey Girouard, and Roel Vertegaal. 2011. Pointable: An in-air pointing technique to manipulate out-of-reach targets on tabletops. *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ACM, 11–20. http://doi.org/10.1145/2076354.2076357
2. John Brooke. 1996. SUS: A "quick and dirty" usability scale. In *Usability Evaluation in Industry*, Patrick W. Jordan, B. Thomas, Ian Lyall McClelland and Bernard Weerdmeester (eds.). Taylor and Francis, London, UK, 189–194.
3. James C. Byers, Alvah C. Bittner Jr., and Susan G. Hill. 1989. Traditional and raw Task Load Index (TLX) correlations: Are paired comparisons necessary? In *Advances in Industrial Ergonomics and Safety*, Anil Mital (ed.). Taylor and Francis, London, UK, 481–485.
4. Daniel Gallardo and Sergi Jordà. 2013. SLFiducials: 6DoF markers for tabletop interaction. *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, ACM, 401–404. http://doi.org/10.1145/2512349.2514914
5. Fernando Garcia-Sanjuan, Alejandro Catala, Geraldine Fitzpatrick, and Javier Jaen. 2015. Around-Device Interactions: A Usability Study of Frame Markers in Acquisition Tasks. *Proceedings of the 15th IFIP TC.13 International Conference on Human-Computer Interaction*, Springer International Publishing, 195–202. http://doi.org/10.1007/978-3-319-22668-2_16
6. Fernando Garcia-Sanjuan, Javier Jaen, Alejandro Catala, and Geraldine Fitzpatrick. 2015. Airsteroids: Re-designing the Arcade Game Using MarkAirs. *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, ACM Press, 413–416. http://doi.org/10.1145/2817721.2823480
7. Khalad Hasan, David Ahlström, and Pourang Irani. 2013. Ad-binning: Leveraging Around Device Space for Storing, Browsing and Retrieving Mobile Device Content. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 899–908. http://doi.org/10.1145/2470654.2466115
8. Otmar Hilliges, Shahram Izadi, Andrew D. Wilson, Steve Hodges, Armando Garcia-Mendoza, and Andreas Butz. 2009. Interactions in the air: adding further depth to interactive tabletops. *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, ACM, 139–148. http://doi.org/10.1145/1622176.1622203
9. Brett Jones, Rajinder Sodhi, David Forsyth, Brian Bailey, and Giuliano Maciocci. 2012. Around device interaction for multiscale navigation. *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, ACM, 83–92. http://doi.org/10.1145/2371574.2371589
10. Hamed Ketabdar, Kamer Ali Yüksel, and Mehran Roshandel. 2010. MagiTact: Interaction with mobile devices based on compass (magnetic) sensor. *Proceedings of the 15th International Conference on Intelligent User Interfaces*, ACM, 413–414. http://doi.org/10.1145/1719970.1720048
11. Sven Kratz, Michael Rohs, Dennis Guse, Jörg Müller, Gilles Bailly, and Michael Nischt. 2012. PalmSpace: Continuous around-device gestures vs. multitouch for 3D rotation tasks on mobile devices. *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ACM, 181–188. http://doi.org/10.1145/2254556.2254590
12. Nicolai Marquardt, Ricardo Jota, Saul Greenberg, and Joaquim A. Jorge. 2011. The continuous interaction space: Interaction techniques unifying touch and gesture on and above a digital surface. *Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction*, Springer Berlin Heidelberg, 461–476. http://doi.org/10.1007/978-3-642-23765-2_32
13. Roman Rädle, Hans-Christian Jetter, Nicolai Marquardt, Harald Reiterer, and Yvonne Rogers. 2014. HuddleLamp: Spatially-aware mobile displays for ad-hoc around-the-table collaboration. *Proceedings of the 9th ACM International Conference on Interactive Tabletops and Surfaces*, ACM, 45–54. http://doi.org/10.1145/2669485.2669500

14. Jeff Sauro. 2011. Measuring usability with the System Usability Scale (SUS). Retrieved from http://www.measuringusability.com/sus.php
15. Jie Song, Gábor Sörös, Fabrizio Pece, et al. 2014. In-air gestures around unmodified mobile devices. *Proceedings of the 27th annual ACM symposium on User interface software and technology*, ACM Press, 319–329. http://doi.org/10.1145/2642918.2647373
16. Martin Spindler, Sophie Stellmach, and Raimund Dachselt. 2009. PaperLens: Advanced magic lens interaction above the tabletop. *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ACM, 69–76. http://doi.org/10.1145/1731903.1731920