HAL
open science

Combining Multi-touch Input and Device Movement for 3D Manipulations in Mobile Augmented Reality Environments

Asier Marzo, Benoît Bossavit, Martin Hachet

To cite this version:

Asier Marzo, Benoît Bossavit, Martin Hachet. Combining Multi-touch Input and Device Movement for 3D Manipulations in Mobile Augmented Reality Environments. ACM Symposium on Spatial User Interaction [short paper], Oct 2014, Honolulu, United States. hal-01056226

HAL Id: hal-01056226

https://inria.hal.science/hal-01056226

Submitted on 18 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Combining Multi-touch Input and Device Movement for 3D Manipulations in Mobile Augmented Reality Environments

Asier Marzo
Public University of Navarre
Pamplona, Spain.
asier.marzo@unavarra.es

Benoît Bossavit
Public University of Navarre
Pamplona, Spain.
benoit.bossavit@unavarra.es

Martin Hachet
INRIA Bordeaux
Talence, France.
Martin.hachet@inria.fr

ABSTRACT
Nowadays, handheld devices are capable of displaying augmented environments in which virtual content overlaps reality. To interact with these environments it is necessary to use a manipulation technique. The objective of a manipulation technique is to define how the input data modify the properties of the virtual objects. Current devices have multi-touch screens that can serve as input. Additionally, the position and rotation of the device can also be used as input creating both an opportunity and a design challenge. In this paper we compared three manipulation techniques which namely employ multi-touch, device position and a combination of both. A user evaluation on a docking task revealed that combining multi-touch and device movement yields the best task completion time and efficiency. Nevertheless, using only the device movement and orientation is more intuitive and performs worse only in large rotations.

Author Keywords
Mobile Augmented Reality; manipulation; multi-touch.

ACM Classification Keywords
H.5.1. Information interfaces and presentation: Artificial, augmented, and virtual realities.

INTRODUCTION
Mobile Augmented Reality (MAR) is a visualization technique which superimposes virtual content over the real environment on a mobile device. Currently, it has reached a mature state as it is supported by most handheld devices and numerous MAR apps are available. This was made possible by the advances in mobile technologies as well as by the progresses in computer vision and graphics techniques. Nonetheless, the way of interacting with these augmented environments is to some point still unclear and thus constitutes a bottleneck for the appearance of highly interactive MAR applications.

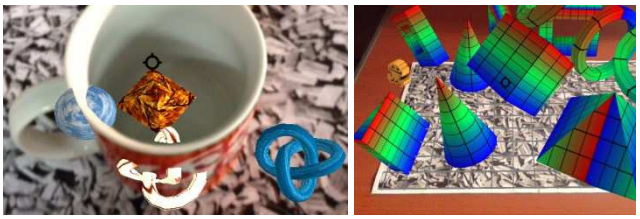

Figure 1. Manipulating virtual objects in a MAR environment. Left) Virtual objects positioned around a real object. Right) Positioning obstacles in a space game.

To interact with a virtual environment, one of the most important tasks is the manipulation of virtual objects. For example, in MAR the task of manipulation is required as soon as the user needs to position and orientate 3D virtual objects in the real environment, as illustrated on Figure 1.

At this time, mobile devices offer multi-touch screens as a standard input. Additionally, in MAR applications, the position and orientation of the device in the real world can be known. Therefore, the challenge is to determine which of these input modalities are better suited for MAR 3D manipulation and if it is beneficial to combine them.

RELATED WORK
In this study, we are focused on manipulation techniques that can be implemented in normal handheld devices and that may achieve efficient interactions. Therefore, we excluded input modalities such as finger or hand tracking as they are not stable on current mobile devices. Similarly, although some studies have used additional devices in combination with the phone, we preferred to concentrate on the input space offered by a standard mobile.

The first study using a mainstream mobile device to manipulate virtual objects in MAR employed different input modalities to separately translate or rotate an object [4]. According to the authors, the best strategy was to map the physical translations of the device to the virtual object translations, and to use the keypads to rotate the object. In later studies keypad, tilt and finger tracking were compared as an input method to complement device movement [5]. Despite finger tracking being promising, keypad input for rotations outperformed the rest of the methods. Multi-touch screens were not tested as they were not widespread yet.

More recently, a user study compared the performance of touch gestures and finger tracking [1]. The former input modality could be used while pointing with the device to

the virtual object (free-touch) or in a freeze-touch mode. The freeze-touch mode consists in manipulating the object from the point of view of a selected still frame; thus it is not necessary to point to the object during the process. Finger tracking performed two times worse in terms of task completion time (TCT) and accuracy; freeze-touch was slightly better than free-touch but not significantly. The techniques described above were designed to manipulate only 2DOF and none of them employed the device movement alone or in combination with other techniques.

A current study aimed at comparing touch modality with device movement [8]. They concluded that moving the device was better than using the touch screen for translating or rotating virtual objects. Nonetheless, the employed techniques used only one-hand interactions and no hybrid approach was tested.

In a previous preliminary study we observed that combining multi-touch and device movement could be beneficial [6]. In this paper, we analyse more thoroughly the usage of each modality and also report qualitative feedback.

INTERACTION TECHNIQUES

Three interaction techniques to manipulate virtual objects in MAR environments are described in this section.

Multi-touch

Various multi-touch techniques have been developed to manipulate objects with 6DOF [3][7]. However, most of them do not adapt well to small handheld devices. For instance, with small screens, it is difficult to use more than two fingers simultaneously. Similarly, although widgets are a current trend in manipulations for touchscreens [2], they are not adequate for small screens due to cluttering issues and fat finger problems. We conducted a pilot study with 6 subjects to determine which technique was better suited for 6DOF manipulations on handheld devices. This led us to propose a variation of the DS3 technique [7].

The resulting manipulation technique uses one finger to move the object along a plane parallel to the device that pass through the object. The interaction with two fingers serves various purposes. Firstly, the distance between the two fingers modifies the position across the line that joins the object and the device; therefore, a zoom gesture will bring the object closer and a pinch gesture will move it away (Z-Dist). Secondly, the angle variation of the vector that joins the two fingers is transferred to the rotation of the object across the view axis (Z-Rot). Lastly, the displacement of the middle point between the two fingers controls an Arcball rotation [9]. This last part of the technique differs from the original DS3 technique that uses a constraint solver. This solver tries to preserve contact points between the fingers and the object, which is not appropriate on small screens. To maximize the interaction surface, the Arcball rotation of this technique employs the entire screen as the virtual ball, instead of using only the object.

Device Position and Rotation

The most straightforward approach to manipulate a virtual object with the device is to transfer the translation and rotation applied to the device (input) to the translation and rotation of the object (output). This technique may work for translations. Nonetheless, orienting the device to rotate the object may be problematic, as the user will lose sight of the manipulated object in the screen.

Previous approaches [4][5][8] employed the grasping technique in which the position and orientation of the object remain constant relatively to the device. In the pilot study, this technique appeared as the most suitable one for manipulating virtual objects using only device movements. It keeps the object inside the virtual field of view and the metaphor of grasping the objects helps the users to understand how the objects will behave. The users can clutch the manipulation by touching or not the screen.

Hybrid

This third approach mixes multi-touch screen input and device movements. Previous work [4] suggests that for object translation, using the device movement is the most efficient way; and that for rotation, using the keypad is the best option. We took into account these suggestions by combining the grasping technique for translation with a touch control for rotation. The most used technique to perform general rotations is Arcball [9]. Therefore, since one-finger touch input is not dedicated to translations anymore, it is used to perform Arcball rotations in view space. The Hybrid technique also supports the use of two fingers to supplement the Arcball rotation (Z-Rot and Z-Dist gestures).

USER STUDY

We conducted a user study to compare the three modalities associated with the interaction techniques described above. The study involved 12 people (2 female and 10 male). Participants' age ranged from 24 to 40 (M=27.3, SD=8.9). Participants had medium experience with 3D and three users were left-handed. The experiment was performed using a 3rd generation iPod. The software ran under iOS 6, was developed using C++ and used Vuforia SDK for the tracking. The participants were seated in front of a table that reached below their chests. A paper marker is needed to achieve the tracking. It was printed on a 49x34 cm paper sheet and was placed on the table.

The task to perform was a 3D docking similar to Zhai's et al. [10]. Participants were asked to manipulate a bright chair (cursor) in order to overlap it with a static dark chair (target). The target chair was always static, located at the centre of the scene, whereas the cursor was located at different positions and orientations according to the trial. The experiment was divided into 4 blocks of 5 trials each. The order of the trials was: *Only translation*, *Small Simple Rotation*, *Small Complex Rotation (x2)* and *Large Complex Rotation.* All the trials implied translations.

Small angles ranged from 30 to 60 degrees and large ones from 110 to 180 degrees. *Simple rotations* implied rotations around one of the primary axes whereas *complex* ones were performed around a random axis. The distance between the target and the cursor varied from 7 to 15 centimetres. The size of the chairs was 4x4x6 cm. The error tolerance was 1 cm for translation and 12 degrees for rotation. If a participant could not complete the trial in 40 seconds, the trial ended in failure. Selection methods for the objects were removed from the evaluation and the cursor object was always preselected.

Each participant spent approximately 40 minutes for completing the entire experiment. For each condition, the participant had a short training during few minutes. Then, they performed the docking task. Finally, they filled in the NASA TLX questionnaire and a custom questionnaire. The order of the modalities was determined according to a Latin Square. The transformations of the chairs were randomly generated but they were the same for each participant. The timer was stopped when the marker needed to track the device was not visible. This tracking method is a current limitation that will be overcome in the future. For similar reasons, subjects were not allowed to manipulate the marker.

To summarize, the experiment consisted of: 12 participants x 3 techniques x 4 blocks x 5 trials = 720 docking tasks.

RESULTS

Data were analyzed using RM-ANOVA to detect significant effects of technique; T-pair tests with Bonferroni adjustment were used as post-hoc tests. Greenhouse-Geisser correction was applied when sphericity was violated. Only completed tasks were included in the analysis. Success rate was close to 100% across all participants and techniques.

Task Completion Time

Task Completion Time (TCT) represents the average time measured in seconds that the participants needed to complete each trial. TCT split by type of trial and technique is shown on Figure 2, left.

Inefficiency

Inefficiency is the ratio between the length of the real path and the length of the optimal path [10]. Inefficiency is shown on Figure 2, right.

Device Movement and Position

Device movement and speed were measured. They are shown averaged per trial on Figure 3. All the values are reported using the OpenGL coordinate system.

Multi-touch Gestures

We examined the percentage of time that a certain number of fingers were used to interact (Figure 4, left). Specific gestures usage is shown on Figure 4, right. Move and Z-Dist gestures are expressed in the displacement of fingers relatively to the height of the screen (5.5cm). Z-Rot and Arcball are reported in the amount of radians that were applied. All the data are averaged per trial.

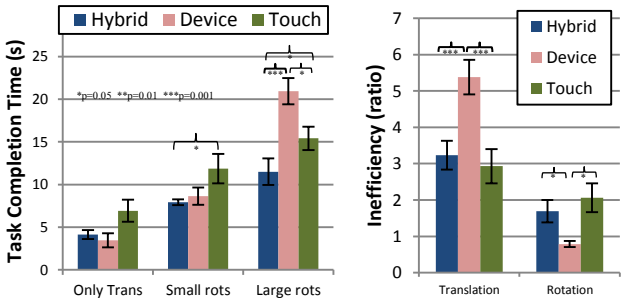

Figure 2. Left) task completion time grouped by technique and type of trial. Error bars represent standard deviation. Right) Inefficiency in translation and rotation.

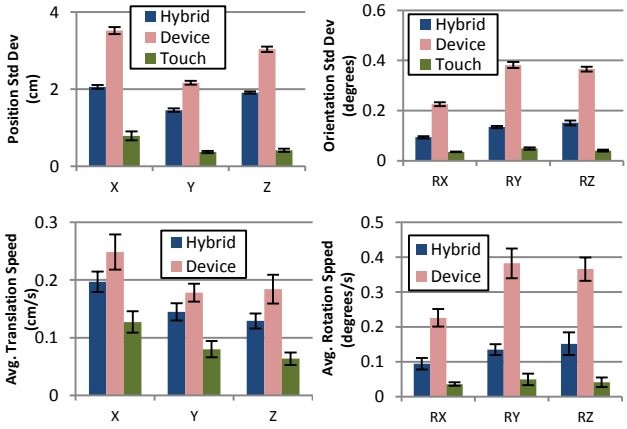

Figure 3. Top) Standard dev. in position (left) and in rotation (right). Bottom) Speed for position (left) and rotation (right).

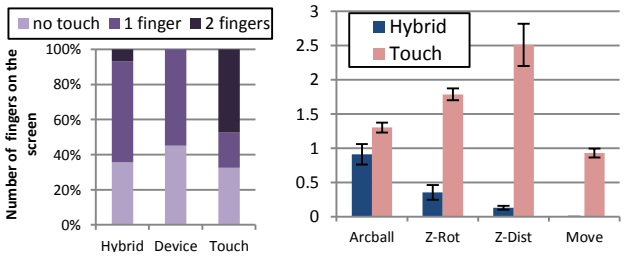

Figure 4. Left) percentage of time spent with a certain number of fingers on the screen. Right) specific gestures usage.

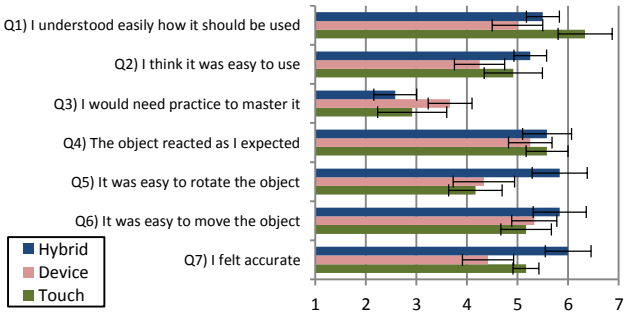

Figure 5. Subjective Custom Questionnaire.

Subjective Ratings

The score of the NASA TLX ranges from 0 (low cognitive load) to 100 (high cognitive load). In our case, this test showed a score of 25.23 (SD=18.41) for *Hybrid*, 31.25 (SD=20.89) for *Device* and 40.05 (SD=19.59) for *Touch*. The results of the custom questionnaire are shown on Figure 5.

DISCUSSION

In average, *Hybrid* had the lowest TCT. For *only translation*, *Hybrid* and *Device* had similar TCT as they both use the grasping technique; *Touch* was the worst but not significantly. For small rotations, *Hybrid* and *Device* performed similarly because the users always employed the device in *Hybrid*. *Touch* performed significantly worse than the other techniques. It seems that for small rotations using the device is better than using multi-touch gestures. For large rotations, *Hybrid* was the best. *Device* was the worst in large rotations as the object had to be rotated using successive manoeuvres. This fact is reflected in the high inefficiency that *Device* had in translation because while rotating the object it was also translated purposelessly. The other two techniques were more efficient in translation but *Device* was the most efficient in rotation.

The amount of movement in position and rotation of the device is always in the order *Device>Hybrid>Touch*. In position, there is more movement on the X axis and less in the Y axis. The *Hybrid* technique reduces the amount of position movement significantly but not for the Y axis since it is the only way of moving the object up. For *Touch*, position movement was low for all the axes but not for X. This could be caused by the users' necessity of adopting different points of view. In rotation, movement was smaller for X in *Device* probably because it was hard to point upwards without losing the target chair or the marker. *Hybrid* mitigates this problem as the finger can be used to rotate around the X axis. Furthermore, *Hybrid* reduces the amount of rotation in all the axes. Speeds follow the same pattern as the amount of movement.

Users spent a similar percentage of time without touching the screen in all techniques. In *Touch*, the amount of time with one and two fingers is proportional to the DOF that can be controlled with them. The *Hybrid* technique used the Arcball gesture slightly less than *Touch*, possibly because the device was used for the remaining rotation. The Z-Dist gesture in *Touch* is unusually high; we assume that this is caused by small involuntary changes of the distance between the two fingers when the user is performing other two-finger gestures (Arcball or Z-Rot).

NASA TLX showed a tendency of *Touch* inducing the highest cognitive load and *Hybrid* the lowest one. In the custom questionnaire, preference was *Hybrid >Device>Touch*; however, it was easier to move (Q6) and rotate (Q5) with *Touch* than with *Device*. *Device* appeared as the most intuitive (Q1) and *Hybrid* as the most accurate (Q7). Nonetheless, results were not significant.

A previous study [8] stated that *Device* was better than *Touch*. Differently, our study reveals that they are similar for translation; *Device* is better for small rotations and *Touch* for large ones. This difference could be due to the solely use of one-hand techniques. Another study [1] revealed that freeze-touch is better for DOF-constrained manipulations. For constrained manipulations, it could be enough to use *Touch* or *Device*. However, none of these studies evaluated a hybrid approach.

CONCLUSION

Three techniques to manipulate virtual objects with 6 DOF have been compared in a MAR environment. Our user study suggests that combining device movement with multi-touch input offers the overall best results. Nonetheless, using only the device performs worse exclusively on large rotations. Using only the device is interesting as it appeared intuitive and removes the necessity of doing touch gestures on the screen.

REFERENCES

1. Bai, H., Lee, G. A. and Billinghurst, M. Freeze view touch and finger gesture based interaction methods for handheld augmented reality interfaces. IVCNZ. ACM (2012).
2. Cohé, A., Dècle, F. and Hachet, M. tBox: a 3d transformation widget designed for touch-screens. CHI. ACM (2011).
3. Hancock, M., Carpendale, S. and Cockburn, A. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. CHI. ACM (2007).
4. Henrysson, A., Billinghurst, M. and Ollila, M. Virtual object manipulation using a mobile phone. ICAT. ACM (2005).
5. Henrysson, A., Marshall, J. and Billinghurst, M. Experiments in 3D interaction for mobile phone AR. GRAPHITE. ACM (2007).
6. Marzo A., Bossavit B. and Hachet M. [Poster] Combining Multi-touch and Device Movement in Mobile Augmented Reality Manipulations. ISMAR. IEEE 2014.
7. Martinet, A., Casiez, G. and Grisoni, L. Integrality and separability of multitouch interaction techniques in 3D manipulation tasks. IEEE Transactions on Visualization and Computer Graphics, 18(3), 369-380. (2012).
8. Mossel, A., Venditti, B. and Kaufmann, H. 3DTouch and HOMER-S: intuitive manipulation techniques for one-handed handheld augmented reality. Virtual Reality International Conference: Laval Virtual. ACM (2013).
9. Shoemake, K. ARCBALL: a user interface for specifying three-dimensional orientation using a mouse. Graphics Interface (Vol. 92, pp. 151-156). (1992).
10. Zhai, S. and Milgram, P. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. CHI. ACM (1998).