Jurnal
Teknologi

Full Paper

A Review of 3D Gesture Interaction for Handheld Augmented Reality

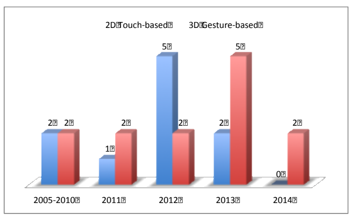
Article history
Received
15 June 2015
Received in revised form
1 October 2015
Accepted
13 October 2015

Cik Suhaimi Yusof[a*], Huidong Bai[b], Mark Billinghurst[b], Mohd Shahrizal Sunar[a]

[a]UTM-IRDA Digital Media Centre, Universiti Teknologi Malaysia, Johor, Malaysia
[b]HITLabNZ, University of Canterbury, Christchurch, New Zealand

*Corresponding author
suhaimi@utm.my

Graphical abstract



Abstract

Interaction for Handheld Augmented Reality (HAR) is a challenging research topic because of the small screen display and limited input options. Although 2D touch screen input is widely used, 3D gesture interaction is a suggested alternative input method. Recent 3D gesture interaction research mainly focuses on using RGB-Depth cameras to detect the spatial position and pose of fingers, using this data for virtual object manipulations in the AR scene. In this paper we review previous 3D gesture research on handheld interaction metaphors for HAR. We present their novelties as well as limitations, and discuss future research directions of 3D gesture interaction for HAR. Our results indicate that 3D gesture input on HAR is a potential interaction method for assisting a user in many tasks such as in education, urban simulation and 3D games.

Keywords: Gesture Interaction, Handheld, Augmented Reality

Abstrak

Interaksi untuk *Handheld Augmented Reality* (HAR) adalah satu topik penyelidikan mencabar disebabkan paparan skrinnya yang kecil dan pilihan input terhad. Walaupun input skrin sentuh 2D digunakan secara meluas, interaksi isyarat 3D adalah kaedah input alternatif yang disyorkan. Baru-baru ini kajian interaksi isyarat 3D terutamanya memberi tumpuan kepada menggunakan kamera *RGB-Depth* untuk mengesan kedudukan ruang dan jari, dengan menggunakan data ini untuk manipulasi objek maya di persekitaran AR. Dalam kertas kerja ini kami mengkaji semula penyelidikan isyarat 3D terdahulu pada metafora interaksi alatan tangan untuk HAR. Kami paparkan keaslian penyelidikan mereka berserta batasan, dan membincangkan arah penyelidikan masa depan 3D interaksi isyarat untuk HAR. Keputusan kami menunjukkan bahawa input isyarat 3D pada HAR adalah kaedah interaksi yang berpotensi untuk membantu pengguna dalam banyak penggunaan seperti dalam bidang pendidikan, simulasi bandar dan permainan 3D.

Kata kunci: Gesture Interaction, *Handheld*, *Augmented Reality*

1.0 INTRODUCTION

Augmented Reality (AR) is technology that provides a new way to interact between the physical and virtual world, and is a very important area for future research [1]. However, based on the discussion by Bowman [2], there are a number of interaction research issues that need to be addressed to create a seamless 3D AR User Interface.

The increasing performance of computational and graphics hardware on mobile devices makes the AR feasible and popular on handheld devices. Drastically improved components like processors, displays, cameras, and various sensors, enables researchers to explore the full potential of Handheld Augmented Reality (HAR).

Zhou *et al*. [3] indicated that handheld displays are a promising platform for HAR applications because they are minimally intrusive, social acceptable, readily available and highly mobile. Hurst and Wezel [4] found that interacting with HAR applications is often limited to pure 2D pointing and clicking on the devices touch-screen. They used 3D gesture interaction by using finger tracking to resolve the issues that come from 2D touch interaction such as screen occlusion, limited screen size and using 2D input for 3D interaction.

The advantages of using handheld devices for Augmented Reality include not having to wear the device like a head mounted display (HMD), facilitating easy swapping to real world in case of a collaborative interface, supporting an immersive viewing mode, being able to easily share the hardware, and having display and input combined in one device. However, handheld device also has drawbacks such as requiring the user to hold the device with one hand while stretching out the other hand for interaction. This removes the capability of performing free bimanual interaction [5], and also may cause fatigue after a long period of holding the device. As more handheld devices are used to deliver

AR experiences, natural and intuitive interaction techniques are required by users such as spatial gesture input, instead of simple pointing and clicking, and manipulating 3D content from a 2D surface.

The goal of this paper is to review previous research on 3D gesture interaction techniques for AR on handheld. This paper is organized in four subtopics, starting with investigating the previous study of 2D and 3D interaction methods on HAR. We explore ten years of previous work on 3D gesture input compared with 2D touch interaction. Next we discuss details of 3D interaction issues as well as their limitation. Lastly we conclude and list future direction for 3D gesture interaction research on HAR devices.

2.0 PREVIOUS STUDY OF 2D AND 3D HAR INTERACTION

The first handheld AR experiences used LCD displays that were tethered to desktop computers. For example, Rekimotos transvision project allowed to users to sit across a table and see the same AR content on a handheld displays with a camera and tracker attached [6]. The AR pad [7] added game controller input to allow for richer interaction in a collaborative HAR. In the MagicBook project [8] a handheld device provided a display with a handle that a user could view and manipulate AR content. However these applications were not mobile, requiring users to stay close to the tethered PC.

Moehring *et al*. [9] presented the first self-contained AR system running on a cell phone. It supported optical tracking of passive paper markers and correct integration of 2D/3D graphics into the live video-stream at interactive rates. However, the system was limited because of low video stream resolution, simple graphics and memory capabilities, and a slow processor. Wagner *et al*. [10] ported the ARToolKit [11] tracking library into a mobile system to run HAR applications on a self-contained PDA. They also developed a HAR collaboration application that using touch input with a stylus to control a virtual train application [12].

Most of early interaction techniques were designed from a device-centric perspective such as using 2D touch screen input as well as a stylus pen, keypad, keyboard and device sensors to interact with the virtual objects [21]. In contrast, Mossel *et al*. [22] proposed a new 3D Touch manipulation technique that offered 6DOF manipulation called HOMER-S. By combining 2D touch information with the device orientation, manipulation of the remaining 3DOFs can be achieved for scaling and rotation tasks. Manipulation is designed intuitively and the interaction space for transformation control is not limited to the physical size of the mobile screen, while still supporting interaction with the 2D screen display.

Baldauf *et al*. [15] created a gesture engine to detect finger pinch markerlessly by using skin-color segmentation techniques, and used this to manipulate virtual objects. The earliest 3D gesture interaction was developed by Henrysson *et al*. [13], using a fiducial marker attached to the index fingertip and tracked in the front of mobile phone. This was used to control a 3D painting application.

There are various previous works on HAR interaction metaphors that have been implemented by researchers. These can be divided into 2 categories; (1) 2D Touch-based interaction and (2) 3D Gesture-based interaction. Table 1 and Figure 1 show all of the papers published in these categories over the last 10 years. The number of 3D Gesture papers is increasing, reflecting the needs of the field. In contrast, publications using 2D Touch-based input are decreasing over time. This result shows that 3D Gesture-based interaction is becoming more widely used than before. Overall, the total number of related 3D gesture-based interaction research publication is 13 percentages higher than the traditional 2D touch-based methods in 2014.

Table 1 Paper of interaction technique on HAR

YEAR	2D Touch-based	3D Gesture-based
2014	0	[19], [18]
2013	[22], [23]	[20] [17] [24] [25] [14]
2012	[26] [27] [28] [29] [30]	[26], [31]
2011	[4]	[4], [15]
2005-2010	[32], [33]	[13], [16]

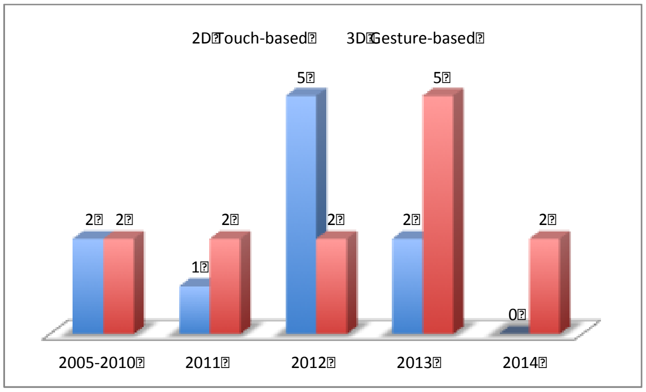


Figure 1 Graph number of paper of iinteraction techniques in HAR

3.0 3D GESTURE INTERACTION IN HAR

Gesture interaction for handheld AR is an example of a user-centric interaction that directly utilizes the users natural behaviors. This is different from device-centric interaction techniques that rely on the device as an in-between, mediating between the user and the content they are interacting with. In most cases, when a person uses a handheld device they typically hold the device with one hand while the other hand is available for gesture interaction. Based on this, researchers have begun to study one-handed gesture-based interactions using finger [13] or hand tracking [16].

HAR interaction can benefit from a user-centric natural input design. In Table 1 we summarize most of the previous research in natural gesture input for HAR. From this table, we can see that gesture interaction in HAR typically involves mid-air gestures instead of any direct touch on the mobile device. We listed the 3D gesture interaction work for HAR systems based on the improvement from simple fingertip-based solution to more complete skeleton-based one. In the table, we describe the implementation and interface design at the same time.

The user can directly manipulate virtual objects in AR environments with most interfaces, performing 6DOF operations using hand gestures in midair. Previous gesture-based interaction techniques in HAR include marker-based input detection and now are increasingly focusing more on marker-less solutions.

It is noticeable that use of external controllers such as RGB-Depth camera devices [19], [18], [20], [17] is preferred to marker-less gesture detection and estimation, offering more intuitive user experience. Finger tracking can be achieved by using these depth devices. Some of these works are also using a client-server framework to receive depth data from the device. In this case the handheld device is set as that client and that PC with the depth camera attached is the server.

4.0 DISCUSSION

For HAR interaction research, we have to solve two significant problems to build a complete system: (1) AR target tracking and (2) interaction gesture detection.

Based on the systems shown in Table 2, we can observe that HAR interaction methods are evolving over time from simple marker tracking to marker-less tracking.

Marker-based interaction was mainly used in early research work like [22], in which the interface required the user to attach a black-white square marker to their fingertip. The phones front camera then tracked the marker and calculated its spatial position and orientation, which was used as the input value. Moving the tracked fingertip in space, combined with keypad pressing, would allow various manipulations to be done to the virtual object such as selection, translation and rotation.

In this implementation, the rear camera was used for AR target tracking while the front camera was used for fingertip detection. The front camera based method was used because the square marker tracking fails once the hand blocks part of the marker, and the field of view for the rear mobile camera was very narrow.

The gesture engine [22] uses the same AR tracking technique but detects the fingertips based on the skin color segmentation via the rear mobile camera. We can tell from the paper that the field of view is much wider than the old phones because of hardware updates. However, the tracking issue caused by the occlusion still exists.

Two approaches are commonly used in practice to overcome the tracking occlusion problem. The first way is to use sensor-based AR tracking without any image tracking involved [4]. In this case, the rear camera will only track the hand without any conflict with the AR tracking. Another solution uses natural feature tracking to keep the vision-based AR tracking available, overcoming the occlusion failure [5], [15], [18], [19].

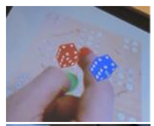
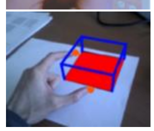
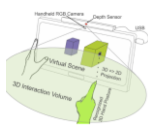
This marker-less tracking improvement provides a more comfortable and convenient way for users to interaction with the HAR scene stretching out one arm and placing the hand behind the mobile camera for gesture detection and AR object manipulation while keeping the AR tracking alive.

We only discuss the gesture input identified by computer vision algorithms, but from the Table 2 we can also observe that HAR interaction methods are evolving from single 2D fingertip interaction to full hand 3D gesture input.

Attaching a visual marker to the fingertip for hand gesture input is one of the most straightforward methods that was initially used for HAR interaction. The color-based marker requires very cheap calculation but only offers a very rough depth value of the fingertip relative to the mobile camera [4], [24]. As we mentioned before, a square marker was used on the fingertip for identifying its spatial position and pose to a certain degree of accuracy [22]. However, the marker-based method may be uncomfortable for the user because of the extra setup on the hand, and it may cause visual confusion or distraction during the interaction.

Fingertip and palm detection methods based on skin color segmentation have been developed to overcome the negative factors described above.

Table 2 Implementation and interface of 3D gesture-based interaction in HAR^

Interface	Demonstration	Implementation
Marker- based tracking and single fingertip [13]		A 3D paint application that drops virtual cubes to create a simple virtual sculpture. Front camera gesture input is used to interact with the index fingertip with an attached ARToolKit Marker
Color Marker- tracking based with single fingertip interaction [4]		Virtual object selection by using single color (green) marker-based tracking of the index fingertip. There are three object selection tasks that have been explored: test case (single object), easy (multiple non-overlapping objects) and hard (multiple overlapping objects).
Color marker- based tracking with single/two fingertips interaction [14]		A green marker and a red marker are attached to the thumb and index fingertip respectively to track fingers. Using circular movement to rotate the virtual object.
Marker-less tracking with one/two fingertip interaction [15]		Uses an engine for fingertip detection in real-time, which is implemented on a HTC Desire phone. The project manipulates virtual 3D objects by using skin color segmentation.
Marker-less tracking with full hand interaction [16]		3D gesture interaction with a virtual object is implemented on a Sony UMPC. The interaction occurs when the hand is opened. For example, a flower is opened and a bee comes out and buzzes around it.
Marker-less tracking with one fingertip interaction using depth camera interaction [17]		A client-server framework in a small workspace. The server uses the Kinect camera to detect the depth data of the index fingertip and to share into the handheld application.
Marker-less tracking with fingertip and hand interaction [18]		A Kinect camera is connected to the Asus tablet to detect user fingertips. There are two different hand postures recognized to pro- vide natural 3D interaction with a mobile VR/AR scene.
Marker-less 3D hand skeleton- based tracking with fingertip pinch interaction [19]		User study comparing 3D gesture input methods with traditional touch-based techniques by using canonical manipulation. Implementing a Client-Server framework in which Primesense camera is attached into the server to detect fingertip position. There are three technologies involved, which are 3Gear System, Vuforia and Alljoyn wireless data communication.
Marker-less 3D hand skeleton- based tracking with hand interaction [20]		Users can perform 6DOF manipulation of virtual objects in AR environments using their bare hands in midair. This Free-hand interaction is implemented on a Samsung tablet with a Softkinetic camera attached.

This method does not require any extra setup on the hand for gesture detection [21], [30], and offers a more natural experience. However there are also obvious shortcomings. If the background color in the camera image is similar to the hand color, then the hand region will not be well detected and incorrect fingertip recognition may occur. Moreover, all results will be still extracted from the 2D camera frame, which has no depth information, so it cannot provide spatial input and has fewer degree of manipulation freedom.

To overcome these limitations, researchers have begun to investigate the usage of RGB-Depth cameras for 3D gesture interaction in HAR applications. In a HAR application, the hand is often the closest object in front of the mobile camera, so the depth camera will segment the foreground area directly based on the distance, and then easily obtain the hand region using both color and depth map information.

The depth sensor was not available for the mobile phone until recently, so a few client-server frameworks using depth sensors were presented. By using the depth camera connected to a desktop computer as the server for 3D fingertip [18] or hand skeleton [19] detection, the mobile client will wirelessly receive the gesture data in real time and perform the coordinate transformation to project the gesture input data into the AR tracking system for virtual object control.

In recent studies, a few tablets do natively support short-range (15cm 100cm) depth cameras and provide a feasible platform for self-contained 3D gesture interaction in HAR environments [19].

Current research is moving towards self-contained full-fledged 3D gesture interaction for mobile AR applications, trying to offer 6DOF input to manipulate the AR content with customized hand gestures, although currently pointing and pinching are widely used.

5.0 CONCLUSION

In this paper, we have reviewed 3D gesture-based techniques for handheld Augmented Reality. We provided a detailed discussion on the 3D gesture implementations that have been done as well as limitations for each project. The implementation of 3D gesture interaction is going to be better by using depth sensor to markerlessly track the users finger in the scene. The main limitation with this is typically the small volume space that the virtual object can be interacted within.

In the future, 3D gesture interaction techniques developed for handheld devices with built-in depth sensors could use all the fingers to naturally manipulate the AR content, offering more mobility for different scenarios.

Acknowledgement

We would like to express our appreciation to the Human Interface Technology Laboratory New Zealand (HITLabNZ) at the University of Canterbury. This work was funded by the Ministry of Science, Technology and Innovation through the eScieneFund Grant 01-01-06-SF1188, Intuitive Mobile-based Tracking for Urban Simulation in Augmented Reality Environment.

References

[1] Azuma, R. T. *et al.* 1997. A Survey Of Augmented Reality. Presence. 6(4): 355-385.

[2] Bowman, D. A., Kruijff, E., La Viola Jr, J. J. and Poupyrev, I. 2004. *3D User Interfaces: Theory And Practice.* Addison-Wesley.

[3] Zhou, F., Duh, H. B. L., and Billinghurst, M. 2008. Trends in Augmented Reality Tracking, Interaction And Display: A Review Of Ten Years Of Ismar. *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, ser. ISMAR '08*. Washington, DC, USA: IEEE Computer Society, 2008. 193-202.

[4] Hurst W. and van Wezel, C. 2011. Multimodal Interaction Concepts For Mobile Augmented Reality Applications. In *Proceedings of the 17th International Conference on Advances in Multimedia Modeling - Volume Part II, ser. MMM'11.* Berlin, Heidelberg: Springer-Verlag. 157-167.

[5] Hinckley, K., Pausch, R., Proffitt, D., Patten, J. and Kassell, N. 1997. Cooperative Bimanual Action. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '97*. New York, NY, USA: ACM. 27-34.

[6] Rekimoto, J. 1996. Transvision: A Hand-Held Augmented Reality System For Collaborative Design. *Proc. Virtual Systems and Multimedia.* 85-90.

[7] Mogilev, D., Kiyokawa, K., Billinghurst, M. and Pair, J. 2002. Ar pad: An Interface For Face-To-Face Ar Collaboration. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems, ser. CHI EA '02*. New York, NY, USA: ACM. 654-655.

[8] Billinghurst, M., Kato, H. and Poupyrev, I. 2001. The Magicbook: A Transitional Ar Interface. *Computers and Graphics*. 25: 745-753.

[9] Moehring, M., Lessig, C. and Bimber, O. 2004. Optical Tracking And Video See-Through Ar On Consumer Cell Phones. Workshop On Virtual And Augmented Reality Of The GI-Fachgruppe AR/VR. *Citeseer.* 193-204.

[10] Wagner, D. and Schmalstieg, D. 2003. Artoolkit On The Pocketpc Platform. *IEEE International Augmented Reality Toolkit Workshop. Darmstadt, Gernany.* 14-15.

[11] Kato, I. P. H., Billinghurst, M. and Poupyrev, I. 2000. Artoolkit User Manual, Version 2.33. Human Interface Technology Lab, University of Washington. vol. 2.

[12] Wagner, D., Pintaric, T. and Schmalstieg, D. 2004. The Invisible Train: A Collaborative Handheld Augmented Reality Demonstrator. ACM SIGGRAPH 2004 Emerging Technologies. ACM. 12.

[13] Henrysson, A., Marshall, J. and Billinghurst, M. 2007. Experiments In 3d Interaction For Mobile Phone AR. *Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques* in Australia and Southeast Asia, ser. GRAPHITE '07. New York, NY, USA: ACM. 187-194.

[14] Hurst, W. and van Wezel, C. 2013. Gesture-based Interaction Via Finger Tracking for Mobile Augmented Reality. *Multimedia Tools and Applications*. 62(1): 233-258.

[15] Baldauf, M., Zambanini, S., Fro ̈hlich, P. and Reichl, P. 2011. Markerless Visual Fingertip Detection For Natural Mobile Device Interaction. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, ser*. MobileHCI '11. New York, NY, USA: ACM. 539-544.

[16] Seo, B.-K., Choi, J., Han, J.-H., Park, H. and Park, J.-I. 2008. One-handed Interaction With Augmented Virtual Objects On Mobile Devices. *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, ser*. VRCAI '08. New York, NY, USA: ACM. 8: 1-6.

[17] Huidong, B., Lei, G. and Billinghurst, M. 2013. Poster: Markerless Fingertip- Based 3D Interaction For Handheld Augmented Reality In A Small Workspace. 3D User Interfaces (3DUI*). 2013 IEEE Symposium on, 2013*. 129-130.

[18] Fritz, D., Mossel, A. and Kaufmann, H. 2014. Evaluating Rgb+D Hand Posture Detection Methods For Mobile 3d Interaction. *Proceedings of the 2014 Virtual Reality International Conference, ser*. VRIC '14. New York, NY, USA: ACM. 27: 1-4.

[19] Bai, H., Lee, G., Ramakrishnan, M. and Billinghurst, M. 2014. One-handed Interaction With Augmented Virtual Objects On Mobile Devices. *Proceedings of the 7th ACM SIGGRAPH Asia 2014 Symposium on Mobile Graphics and Interactive Applications, ser*. SA MGIA '14. New York, NY, USA: ACM.

[20] Bai, H., Gao, L., El-Sana, J. and Billinghurst, M. 2013. Free-hand Interaction For Handheld Augmented Reality Using An Rgb-Depth Camera. *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications, ser.* SA '13. New York, NY, USA: ACM. 22: 1-4.

[21] Wagner, D. and Schmalstieg, D. 2006. Handheld Augmented Reality Displays. *Virtual Reality Conference, 2006. IEEE*. 321-321.

[22] Mossel, A., Venditti, B. and Kaufmann, H. 2013. 3dtouch And Homer-S: Intuitive Manipulation Techniques For One-Handed Handheld Augmented Reality. *Proceedings of the Virtual Reality International Conference: Laval Virtual.* ACM. 12.

[23] Vincent, T., Nigay, L. and Kurata, T. 2013. Precise Pointing Techniques For Handheld Augmented Reality. *Human-Computer Interaction– INTERACT 2013.* Springer, 2013. 122-139.

[24] Bai, H., Gao, L., El-Sana, J. and Billinghurst, M. 2013. Markerless 3d Gesture-Based Interaction For Handheld Augmented Reality Interfaces. *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications, ser.* SA '13. New York, NY, USA: ACM. 22: 1.

[25] Chun, W. H. and Ho ̈llerer, T. 2013. Real-time Hand Interaction For Augmented Reality On Mobile Phones. *Proceedings of the 2013 International Conference on Intelligent User Interfaces, ser.* IUI '13. New York, NY, USA: ACM. 307-314.

[26] Bai, H., Lee, G. A. and Billinghurst, M. 2012. Freeze View Touch And Finger Gesture Based Interaction Methods For Handheld Augmented Reality Interfaces. *Proceedings of the 27th Conference on Image and Vision Computing New Zealand, ser. IVCNZ '12.* New York, NY, USA: ACM, 2012. 126-131.

[27] Lu, H. and Li, Y. 2012. Gesture Coder: A Tool For Programming Multi-Touch Gestures By Demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2012. 2875-2884.

[28] Telkenaroglu, C. and Capin, T. 2013. Dual-finger 3d Interaction Techniques for Mobile Devices. *Personal Ubiquitous Comput*. 17(7): 1551-1572.

[29] Lindeman, R., Lee, G., Beattie, L., Gamper, H., Krishnan Pathinarupothi, R. and Akhilesh, A. 2012. GeoBoids: A Mobile AR Application for Exergaming. *11th IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2012)-Arts, Media, and Humanities Proceedings*, Atlanta, Georgia, USA, 2012. 93-94.

[30] Sukan, M., Feiner, S., Tversky, B. and Energin, S. 2012. Quick Viewpoint Switching For Manipulating Virtual Objects In Hand-Held Augmented Reality Using Stored Snapshots. Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on. IEEE, 2012. 217-226.

[31] Kratz, S., Rohs, M., Guse, D., Mu ̈ller, J., Bailly, G. and Nischt, M. 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, ser.* AVI '12. New York, NY, USA: ACM, 2012. 181-188.

[32] Lee, G. A., Yang, U., Kim, Y., Jo, D., Kim, K.-H., Kim, J. H. and Choi, J. S. 2009. Freeze-set-go Interaction Method For Handheld Mobile Augmented Reality Environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology, se*r. VRST '09. New York, NY, USA: ACM, 2009. 143-146.

[33] Roudaut, A., Huot, S. and Lecolinet, E. 2008. Taptap and magstick: Improving One-Handed Target Acquisition On Small Touch-Screens. *Proceedings of the Working Conference on Advanced Visual Interfaces, ser. AVI '08.* New York, NY, USA: ACM, 2008. 146-153.