IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 25, 2019, accepted March 10, 2019, date of publication March 20, 2019, date of current version April 9, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2906394*

# 3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review

**EG SU GOH[1,2,3], MOHD SHAHRIZAL SUNAR[1,2], (Member, IEEE), AND AJUNE WANIS ISMAIL[1,2]**
[1]Faculty of Engineering, School of Computing, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[2]Media and Game Innovation Centre of Excellence, Institute of Human Centred Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[3]Tanjung Piai Community College, Pontian 82000, Malaysia

Corresponding author: Eg Su Goh (egsu_goh@hotmail.com)

The work of E. S. Goh was supported by the Ministry of Education (MOE), Malaysia, during her study leave. The work of M. S. Sunar was supported in part by the Fundamental Research Grant Scheme under Grant R.J130000.7809.5F093.

**ABSTRACT** Handheld mobile devices referring to mobile phones and tablets become the major output medium for augmented reality (AR), which have seen significant growth in popularity and usage among the public due to the growing release of consumer-oriented communication products nowadays, especially touchscreen smartphones. Unlike traditional desktop or tabletop AR (large display-based) and head-mounted display-based AR systems, small display-based AR (handheld mobile display) requires different interaction techniques that mostly utilize single-hand interaction as well as the limitation of small screen display and limited activity time due to the battery operation hour of handheld mobile devices. However, in handheld mobile AR, research is still lacking, especially research that focuses on 3D interaction in handheld mobile AR for virtual object manipulation. Thus, this paper provides an overview of 3D interaction techniques in handheld mobile AR with critical analysis. First, we describe three main interaction technique categories that applicable in handheld mobile AR, which is touch-based interaction, mid-air gestures-based interaction, and device-based interaction techniques, of their basic concepts on 3D object manipulation. Then, we classify and systematize the highlighted techniques and discuss the advantages and drawbacks of each. Previous studies for widely used techniques have been studied comprehensively. We then draw up a comparison among the different techniques based on the important elements considered in handheld mobile AR. The aim of this paper is to provide researchers with background information on AR and those who are interested in the field of 3D interaction, realizing each technique category.

**INDEX TERMS** 3D object manipulation, device-based interaction technique, mid-air gestures-based interaction technique, handheld mobile AR, touch-based interaction technique.

## I. INTRODUCTION

According to Milgram and Kishino [1], augmented reality (AR) refers to all cases in which the display of an otherwise real environment is augmented by means of virtual (computer graphic) objects. Extended from this definition, Azuma [2] defined that AR is the variation of virtual environments or virtual reality; however, unlike virtual reality that immerses a user inside a synthetic environment and the user cannot see the real world around him/her, AR allows the user to see the real world, with virtual objects composited with the real world. Beyond that, some researchers have done a number of surveys or reviews about AR. For example, Krevelen and Poelman [3], drawing up conclusions from several previous works, defined the AR system as: 1) combining real and virtual objects in a real environment; 2) registering real and virtual objects with each other and 3) running interactively, in 3D and in real-time. From another point of view, Mekni and Lemieux [4] proposed to define AR as systems that have the following characteristics: 1) combining real and virtual content; 2) interactive in real time and 3) registered in 3D.

Thus, conclusions from the above statements provide a commonly accepted definition of AR as a technology which: 1) combines real and virtual imagery; 2) is interactive in real-time and 3) registers the virtual imagery with the real world

The associate editor coordinating the review of this manuscript and approving it for publication was Waleed Alsabhan.

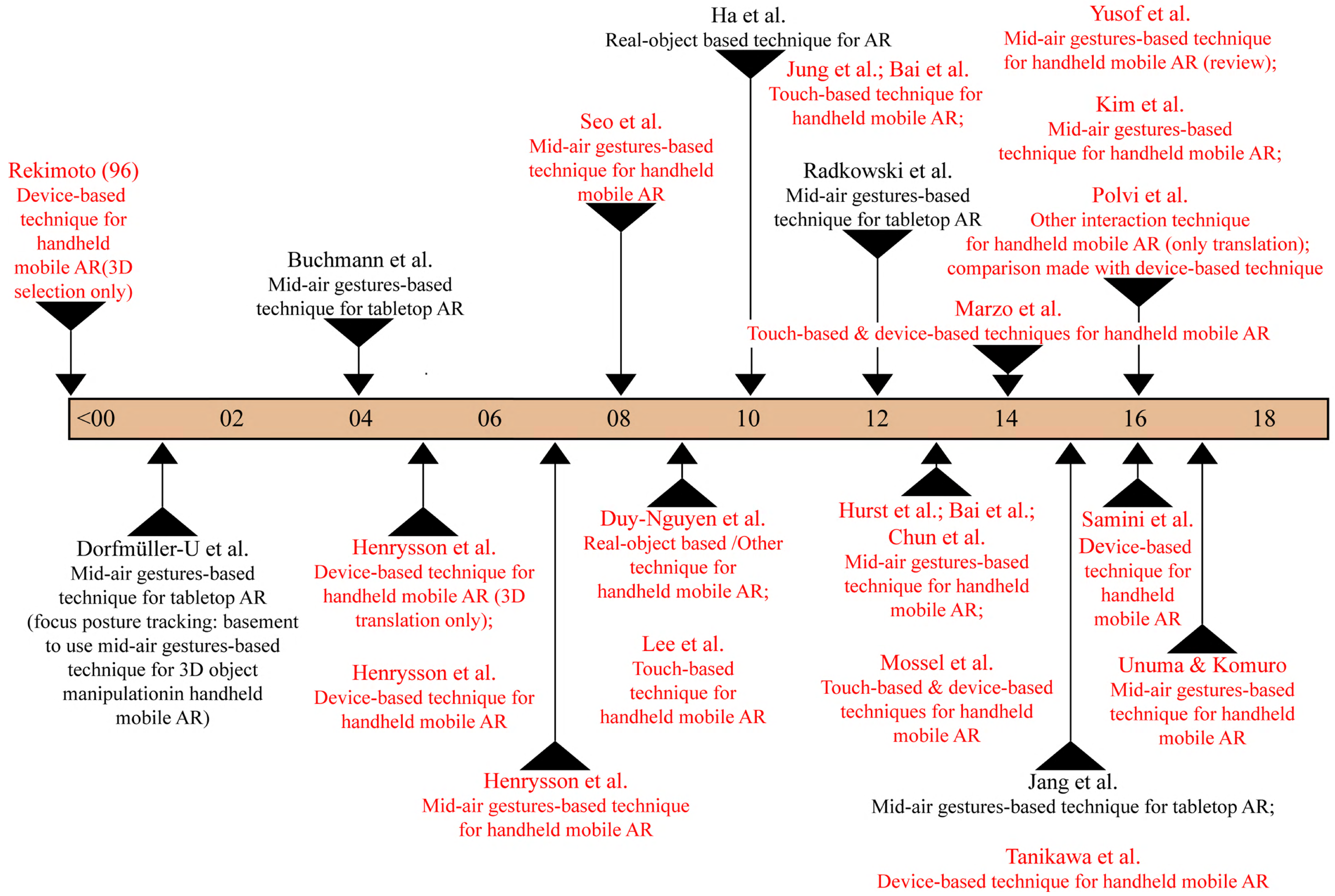


**FIGURE 1.** **Timeline of the researches related to interaction techniques for 3D object manipulation in the AR system. Researches related to handheld mobile AR that stated with their related technique categories are colored in red.**

in 3 dimensions (3D). However, do not limit the potential of AR technology in different aspects, the current definition may be modified based on different cases and different conditions; for example, AR should not be limited to particular display technologies which only focus on our sense of sight, and AR can be potentially applicable to all senses, including hearing, touch and even smell [5]. Although there are several definitions of AR, interaction is still the keyword and the main concept in this context.

### *A. MOTIVATION*

For over 20 years, researchers have proposed various techniques to manipulate 3D objects either in virtual or augmented environments that provide the basement and improvement space towards 3D object manipulation in handheld mobile AR. However, to date, there has not been a comprehensive overview of interaction techniques for 3D object manipulation in handheld mobile AR and their relevant technical requirements and the attached issues. Fig. 1 shows a timeline of researches conducted related to the 3D interaction technique for 3D object manipulation in handheld mobile AR. From this timeline, it is showing the rising trend of researches related with interaction techniques for 3D object manipulation in handheld mobile AR [6] while there are no related survey or review papers recently in regards to the different types of interaction techniques for 3D object manipulation, while researches into interaction techniques in handheld mobile AR have increased over the years [6].

There have been several surveys or review papers, but their focuses are not on interaction techniques for 3D object manipulation in handheld mobile AR [2]–[4], [6]–[8]. Some papers have reviewed only a single interaction technique [9], [10]. Hence, this paper reviews the field of primary interaction techniques for 3D object manipulation in handheld mobile AR which can be categorized as touch-based, mid-air gestures-based, and device-based interaction techniques. Specifically, it provides insights into the fundamentals of those techniques and presents an overview of their phylogeny. Finally, it discusses the differences including the strengths and limitations of each technique, as well as the common problems attached to each of them besides identifying opportunities for future research.

## II. HANDHELD MOBILE AR INTERFACE

Handheld mobile interaction has been upheld as the new direction in human-computer interaction (HCI) field to enhance human ability in HCI using AR [11]. Due to this concern, interaction in AR becomes the main topic and research interest among researchers. One of the important aspects is creating appropriate interaction techniques for AR applications that allow end users to interact with virtual content in an intuitive way [12]. In AR, there are several categories of interaction technique and basically, interaction can be divided into three major parts: 1) tangible user interaction, 2) multimodal input [13] and 3) mobile interaction [14]. Since interaction is the wide topic, this paper focuses on handheld mobile interaction and entail 3D interaction in AR due to the rapid advent of the application on cell phones, and now smartphones and tablets, bringing AR almost to the mainstream [15].

In this paper, handheld mobile AR is our main focus. Handheld mobile AR represents the AR system that based on handheld mobile devices especially smartphone or tablet which are the popular and daily communication devices nowadays. Recent available handheld mobile AR applications promote advanced tools such as Wikitude [16], Layar [17] and Junaio [18] that show the potential of handheld mobile AR interface in consumer markets. In several studies reviewed in this paper, researchers using the word ''mobile'' AR to represent mobile/smartphone or tablet-based AR in their studies while some other studies using the word ''handheld'' AR to represent mobile/smartphone or tablet-based AR system. In order to avoid any confusion of the main concern of this paper, we combining the term ''handheld'' and ''mobile'' as ''handheld mobile'' AR to represent mobile/smartphone or tablet based AR differentiate from ''mobile'' AR that not only represents mobile/smartphone based AR but also meaning head-mounted displays (HMD) based AR system such as Google glasses [19], [20] which not the main scope of this paper. We also avoiding for using ''handheld'' AR which not only represents mobile/smartphone-based AR since handheld devices also bringing the meaning of handheld entertaining devices such as Vive controller used in gamification field which is outside the scope of our discussion.

In handheld mobile AR, there are a number of important differences between using a handheld mobile AR interface and a traditional desktop/tabletop AR interface, including: 1) different input options (no mouse or keyboard) [19], 2) limited screen size [19], and 3) limited activity time due to the battery operation hour of handheld mobile devices. Meanwhile, compared to the HMDs-based AR system, in an AR application on a phone/tablet, the display is handheld rather than head-worn, and the display and input components are connected within a single device mean that interface metaphors developed for desktop and HMDs-based systems may not be appropriate for those handheld mobile AR [19]. Therefore, this paper devoted to handheld mobile devices due to their robustness than HMDs to enter the mass market. Furthermore, based on the analysis done by [6], for the papers published within years 2005 to 2014, the numbers of researches using HMDs and desktop/tabletop started to decrease replaced by the researches done using handheld mobile displays because of the popularity and ubiquity of smartphones. Consequently, this trend leads to researches and studies being conducted to explore the full potential of handheld mobile AR.

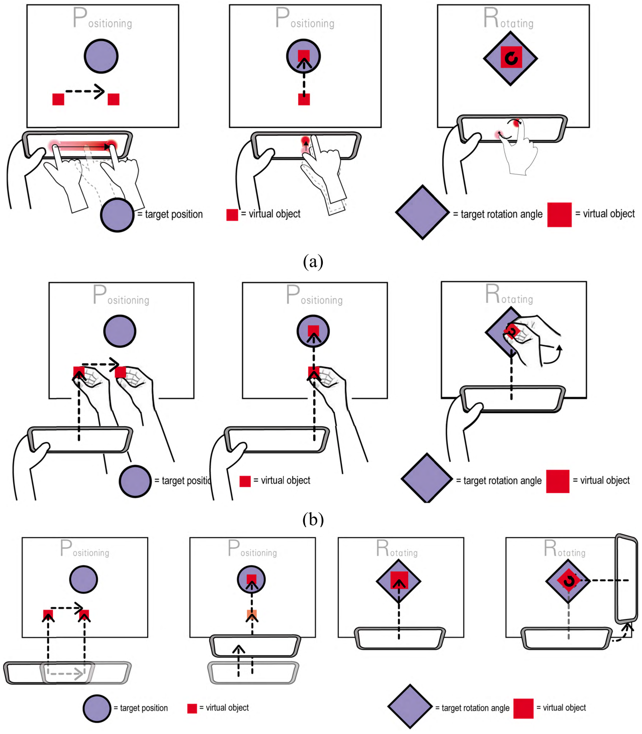
On the basis of these concepts, many techniques have been introduced based on handheld mobile AR interaction that allows the user to interact with the virtual object in real environments. In handheld mobile AR, the 3D interaction task was discussed in many previous works [21]–[23] that are based on fundamental 3D object manipulation tasks including 3D object selection, translation, and rotation [24]. As stated in [19], most of the earlier AR systems were used to view virtual models in application domains [25], [26] while only little support for creating or modifying the AR content.

As stated in the AR definition, interaction in 3D is one of the basic elements of the AR system and interaction in 3D are always involving 3D object manipulation tasks due to its practicability and functionality potential. Due to the huge market potential for handheld mobile devices, handheld mobile AR has played a predominant role in present researches [6], [12]. It has found application in tourist industries [27], maintenance [28] and task support [29], education [30], and medical practices [31]–[33] respectively. An essential part of AR applications is the spatially consistent alignment of virtual and real objects [5] that required accurate and efficient 3D object manipulation. It is crucially important when AR potentially has been explored to support the real-world assembly tasks such as [34]–[38].

Thus, this paper focusing on 3D object manipulation techniques on handheld mobile AR while HMDs-based AR and desktop/tabletop AR are included only as background discussion since many of the researches in this context are related with the phylogeny of 3D object manipulation techniques within handheld mobile AR.

Based on 3D object manipulation, the interaction technique in handheld mobile AR can be categorized into 1) touch-based interaction, 2) mid-air gestures-based interaction and 3) device-based interaction [23], [39], [40] as shown in Fig. 2. Each type of interaction has its own limitations and strengths. The implementation of the 3D object manipulation of virtual objects in handheld mobile AR has been complicated by the fact that almost all handheld mobile AR systems are implemented on smartphones and tablet by means of the tracker (used to track the AR marker) is the built-in camera which also has been used as the rendering system associated with the display screen, all together on the same platform(smartphone or tablet).

As a rough definition of each technique category, touch-based interaction is an interaction category that involves all uses of on-screen touch inputs that are performed by fingertips for the manipulation of 3D objects. Mid-air gestures-based interaction category includes all techniques by means of tracking the bare hands or finger gestures as inputs that allow the manipulation of 3D objects. Device-based interaction category includes techniques that enable the handled mobile device's physical attributes to be tracked for 3D object manipulation (see Fig. 2). Users rotate, tilt, skew and move



(a)

(b)

(c)

**FIGURE 2.** **Examples of the concepts of three (3) main interaction techniques for 3D object manipulation in handheld mobile AR. (a) Touch-based. (b) Mid-air gestures-based. (c) Device-based.**

the mobile device itself as inputs by which to manipulate the 3D object.

In the next section, we describe the three (3) main interaction technique categories with their phylogeny comprehensively and summarize the issues and problems consisted within each category in an integrated table. In order to describe each category in more affluent manner due to their phylogeny, the relevant techniques in virtual environment and tabletop displays also been discussed since many of the interaction techniques proposed and used in large-display based AR systems with assistance of head-worn devices until handheld mobile AR (small displays) have formerly come from the techniques used previously in those systems. These techniques had been modified, improved or even applied directly into the AR system based on their suitability and applicability. Therefore, we firstly discussed each category from a macroscopic perspective (including techniques used in the virtual environment and large-display AR systems) and narrowing it into handheld mobile AR based on 3D object manipulation techniques in a synchronized way.

## III. TOUCH-BASED INTERACTION

Touch-based interaction becomes the main research focus recently after the introduction of touchscreen handheld mobile devices [41]. In the earlier approach on the mobile touchscreen, it is designed to use fingers to navigate on the screen which alters the interaction experience with mobile applications [42] so users can directly interact with the on-screen objects through their fingers [41].

In AR, 3D interaction is one of the main concepts defined in its definition [2]–[4]. 3D object manipulation is a challenge for interface designer since it involves the control of six degrees of freedom (6DOF) which include 3DOF for object translation (x, y, and z-axes) and 3DOF for object rotation (x, y, and z-axes), so does AR. On account that 3D object manipulation is essential in AR interfaces, touch-based interaction provides another option for users to perform 3D object manipulation.

The touch-based concept is utilized by using touch input through multiple contact points instead of using the traditional keyboard or mouse inputs that have already been verified much slower than real object manipulation [43]. Besides, by performing touch-based interaction, users can perform direct manipulation by touching the 3D data directly on screen [44]. Finger gestures that include multi-contact points especially for object rotation are used to manipulate on-screen objects as proposed by Hancock *et al.* [21], which is an example of touch-based interaction technique used to manipulate 2D data on touchscreen tabletops using RNT (Rotate N'Translate) mechanism.

However, touch-based interaction has only been widely explored to do 2D object manipulation that includes x and y-axes only while depth manipulation (z-axis) is still unexplored. This arises due to the difficulty of mapping 2D touch points to 3D attributes to perform the complete 3D object manipulation while the 3D object manipulations consist of 6DOF [22] are mostly executed significantly in virtual environments. Thus, to understand touch-based interaction better, firstly we need to learn more about its basic concepts that rely early on the multi-touch tabletop display.

Based on several previous studies, additional hardware is needed to provide in-depth information to enable manipulation along z-axis integrally. For example, Wilson *et al.* [45] claimed that new range-sensing camera is needed to provide in-depth information to construct a rich 3D model, assisting the user when lifting a 3D object up or down along the z-axis to place it on or below another object. The requirement of new hardware assistance may cause difficulty on both compatibility and cost-effectiveness [22], [45]. Due to the similar issue of in-depth information, Martinet *et al.* [46] introduced the Z-technique to provide a more immersive experience to the user while only a single viewport is available on multi-touch tabletop display and the user can perform 3D object positioning using two (2) finger touches. The movement of the second finger backward or forward provides information to control the depth position. Their study focused on the 3D positioning task and has not been extended to full 3D object manipulation.

The effort to perform full 3D object manipulation was then realized by applying touch control separately instead to perform integral 3D object manipulation as stated previously. In this context, Hancock *et al.* [21] introduced the use of three-touches interaction called Shallow-depth to perform the

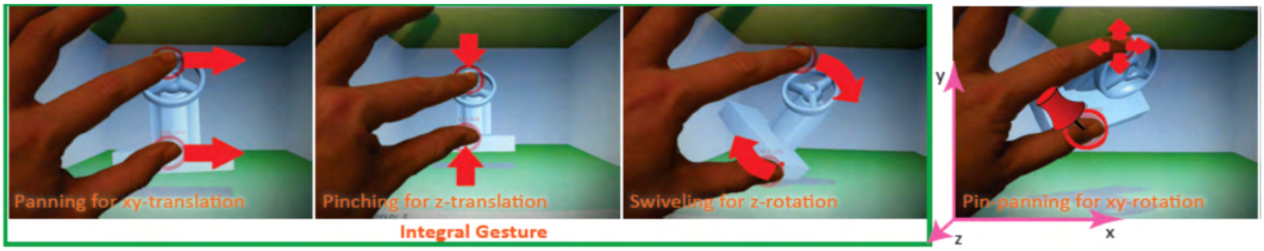


**FIGURE 3.** Two-finger gestures for 3D object manipulation on small touch-based display suggested by [52].

5DOF of translating and rotating 3D objects on the multi-touch tabletop display. This technique requires utilizing both hands of the user to map in-depth information and all the axes. By calculating the distance between two fingers, the user can perform 3D object manipulation along the z-axis. Separating the inputs used for different tasks allows the user to perform 3D object manipulation. Their research found that their technique is faster and more precise, also preferred by their users. Extended from the previous work, Hancock *et al.* [47] introduced sticky tools to consummate the previous technique to perform the complete 6DOF manipulation by extending the standard RNT mechanism into 3D using the screen-space approach. Oppositely, Reisman *et al.* [48] suggested several interaction techniques based on the same control structure mapped along x, y, and z-axes in order to provide integral 3D object rotation.

Based on both researches done by Hancock *et al.* [47] and Reisman *et al.* [48], a new touch-based interaction technique, tBox [49] was then proposed. tBox widget appeared as a wireframe box to ease the user when performing 3D object manipulation, where a single finger was used to do 3D object translation and rotation for direct access while scaling was done by applying two fingers. By using tBox, the user understands accordingly which axis is involved when a single manipulation task is activated through separating the DOF of 3D object translation, rotation, and scaling.

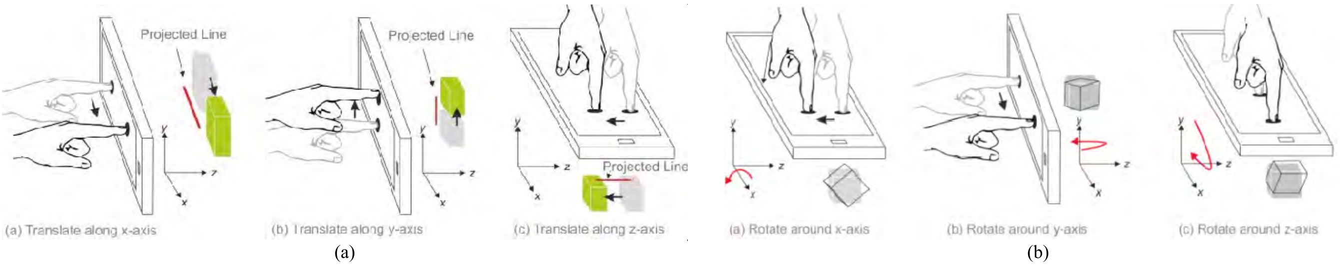
Since DOF integration and separation are two aspects affect the performances in touch-based interaction discussed widely in [47] and [48]. Martinet *et al.* [22] provided a comparison between the touch-based interaction techniques which are sticky tools [47], screen-space [48] and DS3 [22] while DS3 separates the DOF of 3D object translation and rotation. DS3 combines the z-technique and screen-space technique to perform the 3D object rotation. These three (3) different control structures were compared towards the 3D object manipulation task designed. The results showed that the separation of 3D object translation and rotation increases the performance in task completion time compared with just performing both 3D object translation and rotation integrally (all 6DOF integration) or separately (all 6DOF separation).

These studies [22], [47], [48] provide a certain degree of inspiration for the followers, as an example, the effectiveness of the integration of DOF and the total DOF that can be controlled by the fingers of a single hand were evaluated by Brouet *et al.* [50], and the results showed that the restrictions between each finger such as the motion and their interferences cause difficulty in using all fingers in controlling 3D object manipulation. Besides that, users like to limit themselves to using one to a maximum of three (3) fingers per hand to complete complex interactions such as lifting or scaling and also manipulations that consist of all x, y, and z-axes for 3D object translation and rotation. Due to the difficulty in performing full 6DOF of 3D object manipulation, Guo *et al.* [51] proposed another touch-based technique that combines multi-touch gestures with an additional assistant axis performed on touch-based tabletops. They claimed that their technique can perform 7DOF of 3D object manipulation including uniform zooming instead of 3D object translation and rotation. However, their technique requires the operation of both hands with indirect contact points.

### A. 3D OBJECT MANIPULATION FOR HANDHELD MOBILE DISPLAYS

Instead of large-scale multi-touch display common in tabletops as discussed above, the touch-based interaction technique has also been investigated of its possibility after the obtainability of touchscreen handheld mobile devices commonly referred to as smartphones due to its popularity and mobility nowadays. In the early stage, Liu *et al.* [52] conducted a study relating to touch-based interaction that considers small touch-based display. In their research, they proposed a two-finger gesture for full 6DOF 3D object manipulation on a portable small touch-based display. They had improved the work done by Hancock *et al.* [47] in which two moving fingers on a single hand are used for 3D object translation and rotation on one axis, totally consisting of 4DOF of 3D object manipulation, while another 2DOF (2 more axes for 3D object rotation) are done by using one moving finger with another fixed finger (1f-1m) as shown in Fig. 3. Their technique focuses on small touch-based display but the gesture to rotate a 3D object using the 1f- 1m technique is sometimes difficult since the action itself may cause the difficulty in making sure that a finger is really fixed on a position. Slice motion occurs and causes the inaccuracy of the 3D object manipulation.

In the same context, Telkenaroglu and Capin [53] also did a similar study that focuses on limited touch display due to the small screen size. They evaluated their technique with several



(a) (b)

**FIGURE 4.** The 3DTouch technique proposed by Mossel *et al.* [61] to perform 3D object manipulation separately within handheld mobile AR interface. (a) 3D object translation using 3DTouch. (b) 3D object rotation using 3DTouch.

well-known 3D interaction techniques that are not completely touch-based. For 3D object selection, they suggested combining the touch-based technique with ray casting to improve the accuracy in selecting single objects within multiple-object environments. As for 3D object manipulation, the same control structure suggested by Liu *et al*. [52] was implemented. Apart from the difficulty in using 2-finger touch on small touch display for 3D object rotation especially, occlusion is another highlighted issue [54].

Compared with 3D object translation, 3D object rotation is a crucial problem in touch-based interaction as discussed previously. Therefore, Scheurich and Stuerzlinger [55] did similar research focusing on 3D object rotation by applying z-technique [22] on a small touchscreen display. Besides, Rousset *et al.* [56] also came out with a relevant study by introducing surjection as the main criterion instead of integration for 3D object rotation. By improving the existing 3D object rotation technique produced by [56] considering surjection, they claimed that it would provide better results.

### B. 3D OBJECT MANIPULATION FOR HANDHELD MOBILE AR

Touch-based interaction is one of the popular interaction categories based on the suitability to be implemented in AR. In the early stage, the potential of using touch-based interaction was studied by Xin *et al.* [57], Lee [58], and Guven *et al.* [59] but not focusing on 3D object manipulation. Existing touch-based interactions using single fingers and two fingers were then implemented in handheld mobile AR environments studied earlier by Jung *et al.* [60]. The hybrid technique was then introduced afterward by Mossel *et al.* [61] and Marzo *et al.* [62]. They conducted similar works by introducing two types of 3D object manipulation techniques in AR environments. Both researches suggested to use touch-based interaction for 3D object rotation instead of translation to reduce the number of fingers used in 3D object manipulation, and mode switching method is required to separate 3D object translation and rotation (shown in Fig. 4), while the well-known Arcball concept had been studied and improved to perform 3D object rotation with one finger. However, if the user wants to perform 3D object manipulation integrally, at least two fingers are required. The existing issues include the difficulty to control the movements of each contact point mapped with each finger beside the occlusion problem.

After the wide discussion about the touch-based interaction, the important researches had been compressed in a systematic summarized table (Table 1) by keywords either it becomes the basement to AR research works, about 3D object manipulation, or some relevant issues based on touch-based techniques.

As stated in VR field that also is the issue in AR, the occlusion problem, which may cause the fat finger effect, the main issue faced when using the touch-based interaction technique, which is already widely discussed above. Several suggestions and solutions have been introduced based on this basis. For example, the Shift technique proposed by Vogel and Baudisch [63] improved the well-known offset pointer technique proposed at the early stage [64] focusing on small screen touch-based interfaces. This technique reveals the occluded screen content in a callout displayed near the action touch by the user's finger. It provides a potential solution to make sure the user can still fine-tune and see the target content on the screen although it still consists of many shortcomings such as the portion of on-screen content overlaid by the callout cannot be seen by the user, including the part of the content that is occluded by the finger below the touch point. Furthermore, Paudisch and Chu [65] suggested to use the back-side touch to solve the occlusion problem; however, this technique still needs more investigation since the new design of the current handheld mobile device is required to enable such technique.

Due to the difficulties in interacting with virtual objects in handheld mobile AR scenes when the user needs to hold the handheld mobile device with one hand and touch the screen with the other, freeze view [58], [66] becomes a potential solution. According to [59], the freeze-frame technique can be used to allow the user to snapshot the environment to work on it and map the results back to the physical world. Their works had been extended by Bai *et al.* [66] to ease 3D object manipulation in AR interfaces while users can freeze the view to stop the AR marker tracking process so that users can focus on the 3D object manipulation task. Although the freeze view touch performed fastest and most accurately, it mitigated the user's AR experiences since the static view reduces real-time engagement and so was not attractive enough for them [66].

**TABLE 1.** **Summary on relevant researches on touch-based interaction techniques for 3D object manipulation.**

| Researchers/ Year | Display | | 3D Object Manipulation | | Solutions on Specific Issues | |
|---|---|---|---|---|---|---|
| | Tabletop | Handheld Mobile | Translation | Rotation | Occlusion | Multi-touch Difficulties |
| Hancock *et al.* /2006 | √ | | √ (2D) | √ (2D) | | |
| Hancock *et al.* /2007<br>Wilson *et al.* /2008;<br>Reisman *et al.* /2009;<br>Hancock *et al.* /2009<br>Cohé *et al.* /2011;<br>Martinet *et al.* /2012;<br>Brouet *et al.* /2013;<br>Guo *et al.* /2017 | √ | | √ | √ | | |
| Vogel & Baudisch /2007 | | √ | | | √ | |
| Lee *et al.* /2009 | | √ | | | | √ |
| Baudisch & Chu /2009 | | √ | √ | √ | √ | |
| Martinet *et al.* /2010 | √ | | √ | | | |
| Liu *et al.* /2012;<br>Jung *et al.*/2012;<br>Mossel *et al* /2013;<br>Telkenaroglu & Capin /2013<br>Marzo *et al.* /2014 | | √ | √ | √ | | |
| Bai *et al.* /2012 | | √ | √ | √ | | √ |
| Scheurich & Stuerzlinger /2013;<br>Rousset *et al.* /2014 | √ | | | √ | | |

Note: √ represents the relevant keywords in the highlighted research

Furthermore, these kinds of interactions without real-time engagement have already violated the concept of AR universally defined and accepted by researchers in this field [2]–[4].

## IV. MID-AIR GESTURES-BASED INTERACTION

Mid-air gestures-based interaction is powerful in providing a more natural and intuitive interaction mechanism for the user. One of the earlier researches that investigates the potential to model human hands and limbs for interaction purpose was conducted by Rehg and Kanade [67]. In their study, human hands and limbs were modeled as articulated mechanisms and formulated as a real-time visual tracking called DigitEyes. Since their modeling approach was based on 3D kinematics, this made it possible to track any subset of hand or body states with the same basic algorithm and it was possible to track in total 27 DOF of hand model from a two-camera image sequence under orthographic projection. However, DigitEyes could not be applied in complicated backgrounds and suffered an occlusion problem.

Jennings [68] is another researcher focusing on mid-air gestures tracking system. Rehg and Kanade [67] focused on finger tracking. He combined the stereo-range images, color segmentation and shape information to track the highly overconstrained models of fingers assisted with multiple enclosed cameras. It can deal with complex backgrounds and can recover quickly when tracking is lost in the real-time system. However, the combination of too much information leads to slow real-time response. His research and then is followed by Letessier and Bérard [69] that focused on bare-finger tracking, claiming that their system can provide a more robust interaction that only requires a low standard camera and without a large interactive surface through image differencing segmentation and novel shape filtering algorithm although their works were limited within controlled environments and had not yet been studied in providing more information such as finger orientations for more complicated tasks especially for 3D object manipulation.

There had been a wide exploration of hand pose tracking and estimation researches until Erol *et al.* [9] did a solid analysis of the existing studies. Based on their study, hand and finger motions can be captured based on two (2) types of system; a) data glove sensing system and b) computer vision-based sensing system. Computer vision-based sensing has become a promising alternative to data glove sensing due to its naturality, unencumbered and non-contact interaction. In their study, they focused on the difficulties of the vision-based hand pose estimation system which consists of the high-dimensional problem, self-occlusions, processing speed, uncontrolled environments, and rapid hand motion. They also studied the hand modeling-related issues and concluded that vision-based hand tracking system can achieve robustness, accuracy and high processing speed.

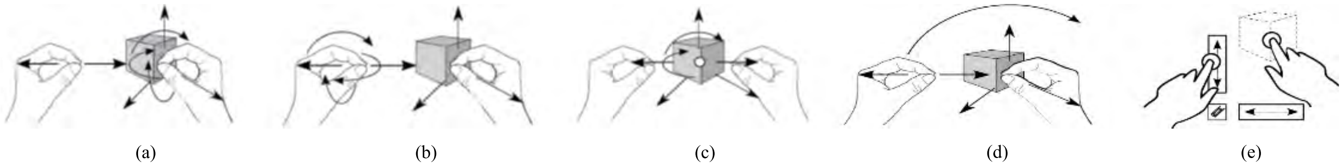


**FIGURE 5.** The five (5) 3D object manipulation techniques (four mid-air techniques with one touch-based technique) in [79]. (a) Thee 6DOF hand technique. (b) The 3DOF hand technique. (c) Handler bar technique. (d) The air translate-rotate-scale technique. (e) Touch-based technique.

Although there have been many researches relating to hand and finger tracking, but studies that focus on 3D object selection and manipulation can be found in Benko and Feiner [70], which is an earlier research relating to 3D object selection. In their study, a mid-air multi-finger technique called balloon selection was proposed to perform 3D object selection. By closing two hands, the user could move the balloon up while moving two hands apart bringing the balloon downwards on a stereoscopic tabletop. While Hilliges *et al.* [71] explored the possibility to perform 3D object manipulation on interactive tabletops. In their study, two prototypes were proposed with different setups, whereby the first prototype was set up without additional hardware by applying shadow mapping, using the shadows to provide the partial depth information and allow the user to perform 4DOF manipulation while another prototype assisted by an in-depth camera called ZSense camera enables higher DOFs. Both techniques represented in virtual environments on tabletop display require new hardware assistance which might cause difficulty on both compatibility and cost-effectiveness [71]. Strothoff *et al.* [72] then proposed another potential mid-air gestures-based interaction technique that uses a triangle cursor to allow the specification of a 3D object position and yaw rotation above the interactive tabletop. The user can use the thumb and index finger in mid-air to control the triangle cursor for 3D object manipulation tasks and move their hand up and down to translate the 3D object.

Due to the potential of mid-air gestures-based interaction in providing an intuitive and natural interaction, the concept to use sensing glove was proposed by Wong and Popovic [73]. Color gloves had been used in [73] to ease hand pose estimation and tracking in order to provide a foundation in developing new interaction techniques in several platforms such as AR and animation modeling for 3D object manipulation purpose, whereas the technique called the continuous interaction space [74] combined the multi-touch function with mid-air gestures-based tracking to enable 6DOF 3D object manipulation although it is still constrained by the person's reach of the hand, and the rotation is limited by the movement around the wrist joint.

Afterward, the Kinect depth sensor by Microsoft was launched in November 2010. Due to its robustness to provide a heuristic hand tracker, quite a large number of researches have been carried out based on Kinect sensor. For example, Frati and Prattichizzo [75] suggested to combine wearable haptic devices with Kinect to solve the poor position sensing issue faced by the usage of wearable haptic devices while Kulshreshth *et al.* [76] proposed a Kinect-based finger detection technique that focused on the finger pose recognition and claimed that their technique was more accurate in tracking the number of fingers raised up compared with the existing technique that uses the K-curvature algorithm. Besides, Araùjo *et al.* [77] also introduced a 3D modeling technique above tabletops supported with a Kinect tracker named Mockup builder by using both hands supported with several widgets to perform 3D object manipulation. Also, Song *et al.* [78] proposed the handler bar metaphor to perform 3D object manipulation. With the support of the Kinect sensor, users can perform 3D object translation and rotation in mid-air with both hands.

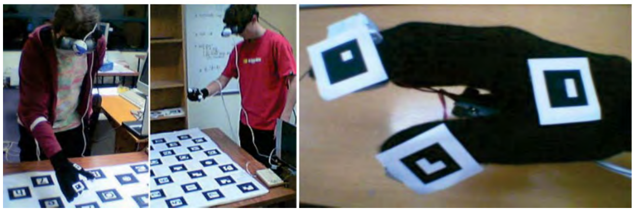
Since there had been many different mid-air interaction techniques proposed previously [70], [72], [77], [79], comparisons were made by Mendes *et al.* [79]. For example, to compare four mid-air interaction techniques with one touch-based interaction technique as the baseline (Fig. 5). Results showed that self-occlusion is still the main issue in mid-air gestures-based interaction while the handler bar technique [78] can spatially solve this problem. The quality of depth camera used might also affect the hand tracking accuracy while new generations of hardware such as Kinect for examples can be explored in this topic.

Presently, Mendes *et al.* [80] again conducted a study to separate the DOF of the mid-air interaction technique due to the results gained by Martinet *et al.* [22] showing that separating DOF can increase the performance either accuracy or task completion time. They suggested to use custom transformation axes and developed MAiOR system to evaluate their works. However, their results showed that DOF separation was still weak in achieving high precision when performing complex tasks compared with touch-based interaction.

### A. 3D OBJECT MANIPULATION FOR HANDHELD MOBILE DISPLAYS

Instead of both hands interaction as discussed above, the mid-air gestures-based interaction technique and then been investigated of its possibility towards using single hand after the obtainability of handheld mobile devices commonly referred to as smartphones due to its popularity and mobility nowadays that require the user to hold the device with one hand. Therefore, apart from the mid-air gestures-based interaction utilizes both hands, interaction in mid-air using single hand also had been studied by Segen and Kumar [81], O'Hagan and Zelinsky [82], and Baldauf *et al.* [83]. These earlier researches showed that one-hand can be used for mid-air interaction by applying the finger pose and motion estimation. Since the 3D object rotation seem much difficult compared with 3D object translation due to the self-occlusion [9], [80], Kratz *et al.* [84] proposed PalmSpace that used the user's palm to perform 3D object rotation. PalmSpace had been compared with the multi-touch technique in 3D object rotation task. They claimed that their prototype performed well and could be extended to 3D object translation. However, due to the constraint of the wrist joint, only small-range rotation can be performed.

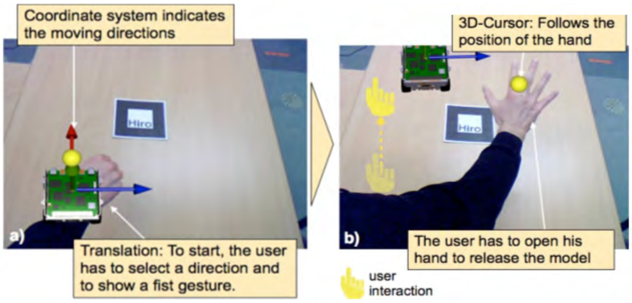
Difference mid-air gestures used to perform 3D object manipulation causing postures recognition became crucial important, the evaluation had been conducted by Fritz *et al.* [85] lately. Three (3) types of hand or finger postures (five (5) fingers outstretched; pointer finger outstretched and hand palm) had been utilized to form difference control structures to do 3D interaction. RGB-D (depth) data had been used to improve the detection's accuracy of hand palm postures. Results from their pose recognition test showed that RGB+D hand posture detection methods can improve the hand postures recognition within a complicated background (cluttered).



i- The urban planning demonstration setup

ii- The glove, showing the AR markers used for hand tracking

(a)



(b)

**FIGURE 6.** Using mid-air gestures-based interaction in an augmented environment [87], [88]. (a) Using FingARtips for urban planning workplace [87]. (b) 3D translation [88].

### B. 3D OBJECT MANIPULATION FOR HANDHELD MOBILE AR

Due to the naturality and the ability to provide intuitive and seamless interaction experience, mid-air gestures-based interaction then had been extended from the previous works that mostly done in stereoscopic interactive tabletop or wall towards the AR environment. The earlier research related to mid-air gestures-based interaction in AR environment done by Dorfmüller- Ulhaas and Schmalstieg [86], Buchmann *et al.* [87], and Radkowski and Stritzke [88]. In [86], the user's own hand had been used to play AR chess. To allow relatively unrestrained environment, a marked glove was used for real-time separation of the finger from the background to provide robust interaction technique to select, translate and rotate the chess piece.

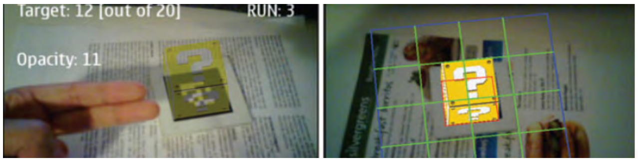
While in [87], FingARtips had been introduced by marking the user's fingertips as fiducial markers to track the user's finger gestures to manipulate a virtual object in the augmented environment. The prototype had been developed to enable FingARtips for urban planning workplace showed in Fig. 6a. Four different types of gesture interaction were implemented: grabbing, pointing, navigating and command gestures. Followed by [88], hand pose had been used to perform 3D object manipulation, the user moved his/her hand after grasping a moving direction to translate the 3D object and open his/her hand to release the object (showed in Fig. 6b).

The mobility and popularity of handheld mobile devices especially smartphones lead the research trend from tabletops (large displays) to handheld mobile AR (small displays). Thus, in the last five years, researches relating to large display-based AR had reduced and switched to handheld mobile AR. At the early stage, Henrysson *et al.* [89] did a comparison between several interaction techniques for 3D object manipulation in handheld mobile AR (cell phone). One of the techniques compared was mid-air gestures-based interaction that uses the user's own fingers. In their study, cell phones with built-in camera were used as primary handheld mobile devices that are limited to low resolution video frame, whereas a simple frame-differencing tracking method was used to enable 2D manipulation while 3D object manipulation could be activated through switching modality to access the third rotation axis but was not available for cell phones at that time due to the slow interaction process.

Furthermore, Seo *et al.* [90] proposed to use the user's palm for registering 3D objects while the palm was treated as a marker, and through the palm pose estimation process, the 3D object could be augmented precisely on the user's palm. 3D object manipulation could be performed when users move or rotate his/her palm while additional vibration sensors were added to enhance the tactile experience. The increasing number of researches relating to 3D object manipulation in handheld mobile AR then induced Bai *et al.* [66] to do the comparison between touch-based interaction and mid-air gestures-based interaction. In their study, the most prominent fingertip would be identified and marked by a small white circle in the ''Translation'' and ''Rotation'' modes, while a second fingertip would be recognized and marked as a small red circle in the ''Scaling'' mode. Results showed that the midair technique implemented in their study appeared to have less usability compared with touch-based technique due to the immatureness of their gesture recognition software.

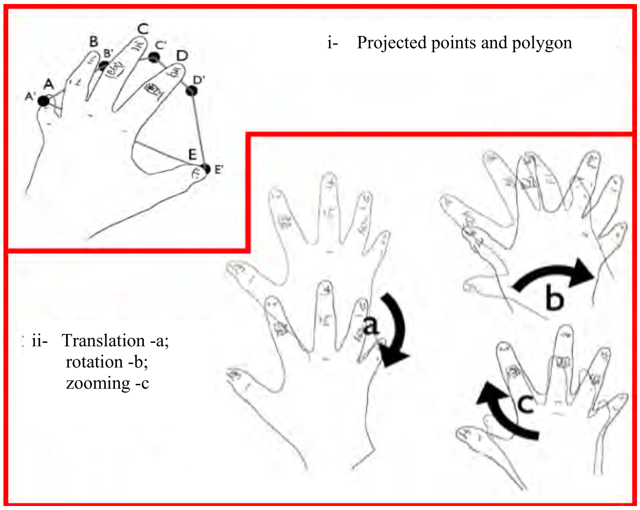
Moreover, Hurst and Wezel [91] did a comprehensive study to utilize the mid-air gestures-based interaction technique towards 3D object selection and manipulation. Results from their user studies showed that participants felt fun and were willing to use their technique because of its intuitiveness and naturality. However, this technique was limited to the translation distance since it was constrained within the arm length and the self-occlusion problem was instead less accurate compared with touch-based interaction. Bai *et al.* [92], [93] then suggested the markerless fingertip-based 3D interaction consisting of seven (7) components which are 1) fingertip detection; 2) fingertip depth acquisition; 3) marker tracking; 4) fingertip coordinate transformation; 5) data communication; 6) fingertip-based interaction and 7) virtual content, rendering in solving the depth dimension problem. Their system needed additional hardware such as Kinect sensor and desktop computer as a server to offer complete 3D object manipulation.

i- Translation ii- Scaling



iii- Continuous value adjustment for box transparency iv- 4x4 grid used to track the hand

(a)



i- Projected points and polygon

ii- Translation -a; rotation -b; zooming -c

(b)

**FIGURE 7.** **Processes involved in [94] and [95]. (a) Processes involved in [94]. (b) Finger tracking in [95].**

Their works and then were followed by Chun and Höllerer [94] that did a user study to measure their interaction technique towards 3D translation and scaling tasks (Fig. 7a). Their technique can successfully address the problem of potential camera motion during gesture detection by using projective texture mapping and background subtraction in maintaining a good frame rate. Moreover, their gesture recognition can avoid self-occlusion and can detect finger gestures accurately.

Extended from [94], Bai *et al.* [95] applied free-hand interaction using an RGB-depth camera for 3D object translation, rotation, and zooming. By mapping the fingertips to project points and polygons (Fig. 7b-i), users can perform 6DOF 3D object translation and rotation (Fig. 7b-ii) although their technique still suffered self-occlusion and accuracy limitations.

Furthermore, pinch gestures used in [70] and [79] were studied and evaluated in handheld mobile AR environments by Bai *et al.* [96] to investigate the usefulness of mid-air gestures-based interaction in 3D object manipulation tasks. Convincing results gained in their study encouraged mid-air gestures-based interaction to be studied in more details to become an intuitive interaction solution in AR environments. Afterward, Kim and Lee [97] did a similar study by evaluating the mid-air gestures-based interaction technique based on richer postures such as pinch, grab and open hand postures and the results were also convincing in terms of the intuitiveness of mid-air gestures-based interaction. Presently in 2017, Unuma and Komuro [98] proposed to align the user view in allowing the user to manipulate 3D object more accurate using the mid-air gestures-based technique. However, additional depth camera and lens to recognize the user's face were required in order to adjust the view. The prototype proposed still immature and additional evaluation was required to measure its performance.

Since AR had been claimed can provides intuitive interaction experiences, mid-air gestures-based interaction technique which is the intuitive interaction way had been widely explored in AR field including handheld mobile AR interface discussed above and so Yusof *et al.* [10] produced a short review about mid-air gestures-based interaction technique in handheld mobile AR due to the active responses of many researchers towards this interaction. They summarized some of the related studies and discussed the existing limitations of the mid-air gestures-based interaction such as the quality of depth sensors nowadays directly affect the recognition of the hand or finger postures, self-occlusion and lack accuracy still the main issues to be cured although there have some researches that provide potential solutions towards these issues such as 3D finger CAPE that proposed by Jang *et al.* [99] against occlusion estimated the 3D finger joint and 3D fingertip locations to form fingertip detector based on depth images. Their framework had been evaluated focusing on occlusion-invariant fingertip position estimation and the results showed that 3D finger CAPE can enable 3D object manipulation intuitively within the AR environment. However, the utilized hand tracking method sometimes fails to track when most of the fingers are completely occluded and it is expensive due to the calculation of different postures

**TABLE 2. Summary on relevant researches on mid-air gestures-based interaction techniques for 3D object manipulation.**

| Researchers/ Year | Display | | 3D Object Manipulation | | Solutions on Specific Issues | | |
|---|---|---|---|---|---|---|---|
| | Tabletop | Handheld Mobile | Translation | Rotation | Posture Tracking/ Recognition | Occlusion | Accuracy |
| Rehg & Kanade /1994;<br>Jennings /1999;<br>Dorfmüller-Ulhaas & Schmalstieg /2001<br>Letessier & Bérard /2004<br>Wong & Popovic /2009;<br>Kulshreshth *et al.* /2013 | √ | | | | √ | | |
| Hagan & Zelinsky /2002 | √ | | √ | √ | √ | | |
| Buchmann *et al.* /2004;<br>Radkowski & Stritzke /2012 | √ | | √ | √ | | | √ |
| Erol *et al.* /2007 | √ | √ | | | √ | √ | √ |
| Benko & Feiner /2007<br>Hilliges *et al.* /2009 | √ | | √ | | | | |
| Henrysson *et al.* /2007<br>Seo *et al.* /2008<br>Hurst & Wezel /2013;<br>Bai *et al.* /2013;<br>Kim & Lee /2016 | | √ | √ | √ | | | |
| Strothoff & Hinrichs /2011;<br>Song et al; Araùjo et al. /2012<br>Mendes *et al.* /2014 & /2017 | √ | | √ | √ | | | |
| Baldauf *et al.* /2011 | | √ | √ | | √ | | |
| Kratz *et al.* /2012 | | √ | | √ | | | |
| Chun & Höllerer /2013 | | √ | √ | | | | |
| Fritz *et al.* /2014 | | √ | | | √ | | |
| Jang *et al.* /2015 | √ | | | | √ | √ | |
| Yusof *et al.* /2016 | | √ | √ | √ | √ | | |
| Unuma & Komuro /2017 | | √ | √ | | | | √ |

Note: √ represents the relevant keywords in the highlighted research

based on the occluded data and still limited towards the different combinations of the finger actions.

Due to the complicated computational tracking information (in-depth data required) and the essential of additional depth devices (such as in-depth camera) may exhaust the battery capacity of a handheld mobile device [6], [96] also the inaccuracy of the finger tracking (caused by occlusion) prolong the interaction period and limit the mid-air gestures-based 3D object manipulation techniques to be used within handheld mobile AR.

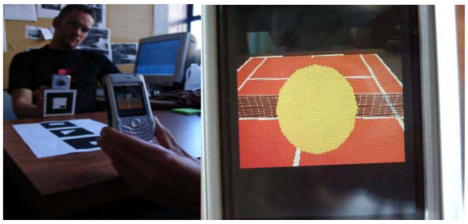
After the wide discussion about the mid-air gestures-based interaction, the important researches had been compressed in a systematic summarized table (Table 2) by keywords either it becomes the basement to AR research works, about 3D object manipulation, or some relevant issues.

## V. DEVICE-BASED INTERACTION

Device-based interaction technique is the technique when the device's own attributes are being utilized and users use the device as the controller for interaction purpose. The earlier device-based interaction technique proposed by Rekimoto [100] used the tilt input of a small screen device. In their study, a palmtop device was used as the tilting interface to select menu items through the device's tilting actions. It was followed by Cho *et al.* [101] that used the tilt input to browse the photos stored in cell phones with additional accelerometer added. Users could either tilt the phone left or right to browse the photos.

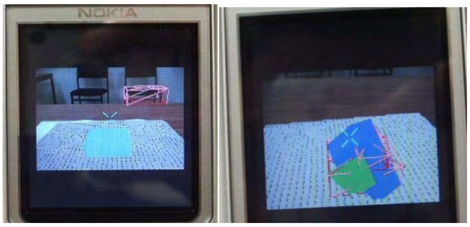
### A. 3D OBJECT MANIPULATION FOR HANDHELD MOBILE AR

In early 2005, the device-based interaction technique was firstly implemented in handheld mobile AR interfaces as a 3D object manipulation tool [102]–[104]. For example, by using cell phones as tennis racquets in [102] as shown in Fig. 8a and using the cell phone to translate and rotate 3D objects in an augmented environment [103], [104] as shown in Fig. 8b. Comparisons have been made towards traditional keypad and



ii- AR tennis game i- Hitting the ball over the net

(a)



iii- A virtual block and translation iv- A virtual block and rotation target

(b)

**FIGURE 8.** **Using the cell phones as tennis racquets in AR tennis game and mapping the position of cell phone's built-in camera to translate and rotate a 3D object. (a) AR tennis [102]. (b) 3D object manipulation [103], [104].**

Arcball techniques with device-based techniques and results showed that the device-based interaction technique is effective in 3D object translation.

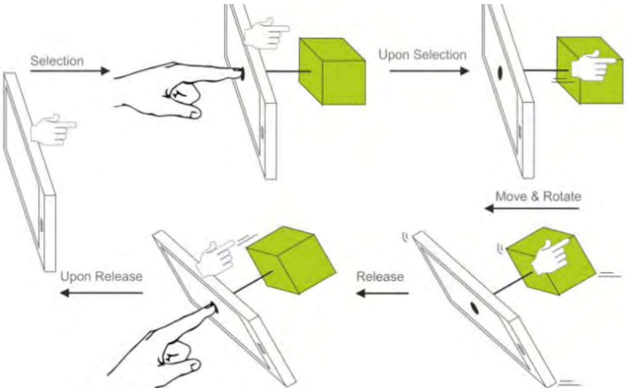
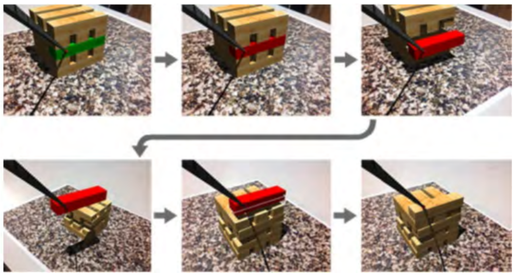


**FIGURE 9.** **Homer-S in [61] for 3D object manipulation within a handheld mobile AR.**

These early researches become the basement for recent studies while the device-based interaction technique becomes a potential interaction solution for 3D object manipulation in handheld mobile AR due to its robustness such as the ability to hold the handheld mobile device with both hands and be free from occlusion [39], [40] until several studies were carried out within the last few years [39], [40], [61], [62]. For examples in [61] and [62], device-based interaction was used for 3D object manipulation (Fig. 9) in AR interfaces and results showed that it is a natural and intuitive interaction technique and can translate 3D objects faster than touch-based interaction although the user cannot complete large-range 3D rotation tasks. By mapping the position of the handheld mobile device's built-in camera with the 3D object registered on the AR marker, the user can use the handheld mobile device as a moving and rotating tool for manipulation. Since the same input (device movement) has been used in both 3D translation and rotation, position deviation occurs causing slow and less accurate positioning in performing 3D object manipulation tasks integrally.

Unlike mid-air gestures-based techniques that require the calculation towards depth information (translation of z-axis for example), device-based techniques are highly suitable to be applied in handheld mobile AR because handheld mobile device movement in depth axis is mapped with the 3D object, when the device is lifted up, the 3D object also be lifted up by using device-based technique thus, additional calculation (on z-axis) that may exhaust the battery capacity can be reduced.

(a)



(b)

**FIGURE 10.** **Device-based interaction technique used in AR Jenga game. (a) AR Jenga game [39]. (b) 3D object manipulation in AR Jenga game.**

Presently, Tanikawa *et al.* [39] applied the device-based interaction technique in AR Jenga game (shown in Fig. 10) while Samini and Palmerius [40] focusing on 3D object rotation issue that stated in [39], [61], and [62] for device-based interaction. In [40], the hold function was proposed to assist the device-based interaction technique that utilizes the device's movement to move and rotate 3D objects in AR interfaces. By presenting the hold and release action, the user can lock the 3D object on different surfaces to complete large-range 3D object rotation (Fig. 11). However, their solution still consists of several drawbacks such as the user might over-rotate the handheld mobile device causing 3D object registration error and AR marker tracking error when the camera sensor goes beyond the viewable and trackable range instead of slow rotation due to the frequent hold and release actions.

**TABLE 3.** Summary on relevant researches on device-based interaction techniques for 3D object manipulation.

| Researchers/ Year | Display | | 3D Object Manipulation | | Solutions on Specific Issues |
|---|---|---|---|---|---|
| | Tabletop | Handheld Mobile | Translation | Rotation | 3D Object Rotation |
| Rekimoto /1996 | √ | | | | |
| Henrysson *et al.* /2005a;<br>Polvi *et al.* /2016 | | √ | √ | | |
| Henrysson *et al.* /2005b;<br>Mossel *et al.* /2013<br>Marzo *et al.* /2014 | | √ | √ | √ | √ |
| Tanikawa *et al.* /2015 | | √ | √ | √ | |
| Samini & Palmerius /2016 | | √ | | √ | √ |

Note: √ represents the relevant keywords in the highlighted research

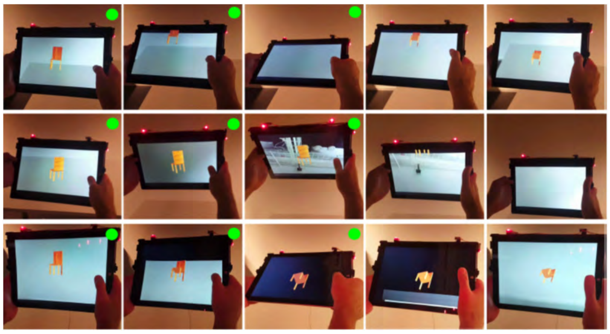


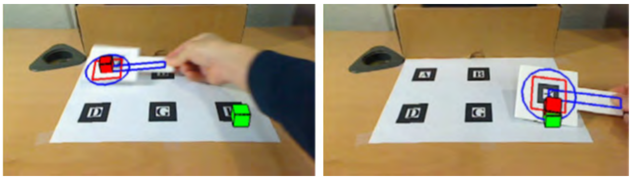
**FIGURE 11.** Device-based interaction technique used in [40] assisted with the hold function.

Due to the effectiveness of the device-based interaction technique in 3D object translation, Polvi *et al.* [23] did a comparison on their 3D object positioning techniques named SlidAR that used ray casting assisted with epipolar geometry and the existing device-based positioning technique that utilized the device movement through mapping the built-in camera position with the 3D object. Although the results showed that SlidAR is faster than the device-based technique, but it is not applicable to 3D object rotation. Furthermore, touch input is needed in their technique, which is included in the different interaction technique categories.

After the wide discussion about the device-based interaction, the important researches had been compressed in a systematic summarized table (Table 3) by keywords either it becomes the basement to AR research works, about 3D object manipulation, or some relevant issues.

Comprehensively, those researches stated in Table 1, 2 and 3 and then had been filtered to refine the particular researches that consisting 3D object translation and rotation in handheld mobile AR and compressed in a systematic summarized table (Table 4) categorized on each of their key findings, limitations and strengths in chronological order based on different technique categories.

(a)



(b)

**FIGURE 12.** Some examples of real object-based interaction techniques in [106]. (a) 2D cup technique for selection and translation. (b) 2D paddle technique for selection and translation.

## VI. REMAINING ISSUES IN 3D OBJECT MANIPULATION WITHIN HANDHELD MOBILE AR

Interaction techniques for 3D object manipulation can be roughly divided into three (3) main categories as stated in the previous section which are the touch-based interaction techniques, mid-air gestures-based interaction techniques and device-based interaction techniques although there are some other interaction techniques such as the real object-based interaction techniques which have been studied in [105] and [106] for handheld mobile AR. In real object-based interaction, the virtual object is coordinated with the real object marked by AR as the controller to translate and rotate the virtual object, but the position and orientation of virtual objects are restricted by those of the corresponding real objects (Fig. 12), where virtual objects are unable to retain the same posture and height as the real objects. Furthermore, the user needs to make sure all the corresponding real objects are within the trackable range to avoid any AR marker tracking failure which limits its potential such as size and translation distance. Since there

**TABLE 4.** Systematic summary of existing important researches on 3D object manipulation in handheld mobile AR interface using touch-based (TBI), mid-air gestures-based (MBI) or Device-based (DBI) techniques.

| Year | Research Title (Researchers) | Interaction Technique Categories (TBI/ MBI/ DBI) | Key Findings | Primary Strengths | Primary Drawbacks |
|---|---|---|---|---|---|
| 2005 | Virtual object manipulation using a mobile phone (Henrysson *et al.)* | DBI | First research about DBI focusing on 3D object manipulation | -Claimed that DBI technique performed better in 3D object translation compared with keypad technique<br>-User study was carried out as proofs | -Slow 3D object rotation<br>-Cell phones used in this study were not smartphones that commonly being used nowadays |
| 2005 | Face to face collaborative AR on mobile phones (Henrysson *et al.*) | DBI | First research about DBI utilizing built-in camera | -Mapping camera position with 3D pad to play AR tennis game<br>-User study was carried out as proofs | -Not being applied in 3D object manipulation<br>-This study utilized cell phone, not smartphone |
| 2007 | Experiment in 3D interaction for mobile phone AR (Henrysson *et al.)* | MBI | The MBI technique is lack in accuracy after comparing with several other categories | -Compared different interaction techniques and verified each performance<br>-User study was carried out as proofs | Focusing on existing interaction techniques for 3D object manipulation, not introducing new techniques |
| 2008 | One-handed interaction with augmented virtual objects on mobile devices (Seo *et al.*) | MBI | Proposed single-hand posture tracking for handheld mobile AR interaction | -Tactile sensor was added to provide more realistic AR experiences<br>-Hand palm had been treated as an AR marker to avoid occlusion | -3D objects cannot retain their actual sizes since hand palm structures are different for each user<br>-Constraints on 3D object rotation due to the wrist rotation limitation |
| 2009 | Free-set-go interaction method for handheld mobile AR environments (Lee *et al.*) | TBI | Proposed to use freeze view to ease interaction within handheld mobile AR interface | -Easing user to manipulate AR scene<br>-Potential to solve the difficulty when using dual-finger touches to manipulate 3D content [45] | -No implementation on 3D object manipulation; one of the elements that form AR definition [2-4]<br>-User's AR experience degraded [66]<br>-Violated real-time engagement stated in AR definition |
| 2012 | Freeze view touch and finger gesture-based interaction methods for handheld AR interfaces (Bai *et al.)* | TBI | Proposed to use freeze view to ease interactions using TBI for 3D object manipulation tasks | -Focusing on 3D object manipulation<br>-Potential to solve the difficulty when using dual-finger touches technique to manipulate 3D content [60]<br>-User study was carried out as proofs | -Feedback from participants claimed that their AR experience were degraded<br>-Violated real-time engagement stated in AR definition [2-4] |
| 2012 | Smartphone as an AR authoring tool via multi-touch-based 3D interaction method (Jung *et al.*) | TBI | Implemented the TBI technique for 3D object manipulation learned from several previous control structures used on tabletop display | -The early implementation of TBI technique for 3D object manipulation in handheld mobile AR<br>-First explored the possibility to use multi-touch control structure in handheld mobile AR interface | -Instability when using multiple finger actions causing failure in AR marker' tracking process [66]<br>-Occlusion [66]<br>-Instability of touch inputs lead to deviation problem [55]<br>-Objects near the screen corners and edges are difficult to be selected and manipulated [56] |
| 2013 | 3DTouch and HOMER-S: Intuitive manipulation techniques for one-handed AR (Mossel *et al.)* | TBI & DBI | -Proposed 3DTouch (TBI) used for 3D object manipulation<br>-First research that proposed DBI technique (Homer-S) in smart phone AR | -Only one-finger touch action supported with task widget needed to separate the DOF of 3D object translation & rotation, avoiding problems discussed in [58] and [66]<br>-Occlusion problem mitigated (TBI)<br>-Homer-S performed better in 3D object translation while 3DTouch performed better in 3D object rotation<br>-User study was carried out as proofs | -Cannot perform 3D object manipulation integrally with single interaction category<br>-Occlusion problem still retain for TBI technique<br>-3D object rotation issue in DBI technique<br>-Position deviation problem in DBI technique when user performed 3D object translation and rotation integrally |
| 2013 | Gesture-based interaction via finger tracking for mobile AR (Hurst & Wezel)<br>Markerless 3D gesture-based interaction for handheld AR interfaces (Bai *et al.*) | MBI | Perform bare-finger tracking for 3D object manipulation | -Performed complete 3D object manipulation<br>-User study was carried out as proofs | -Lack of accuracy<br>-Occlusion problem<br>-Additional depth sensor was needed |

have been no recent studies about the real object-based interaction technique thus it is not be discussed further in this paper.

After a detailed investigation towards the three (3) main interaction categories, several primary issues are highlighted as shown in Table 5. Based on Table 5, it shows the striving

**TABLE 4.** ***(Continued.)*** **Systematic summary of existing important researches on 3D object manipulation in handheld mobile AR interface using touch-based (TBI), mid-air gestures-based (MBI) or Device-based (DBI) techniques.**

| | | | | | |
|---|---|---|---|---|---|
| 2013 | Real-time hand interaction for AR on mobile phones (Chun & Höllerer) | MBI | Performed bare-hand and finger tracking for 3D object selection, translation, and scaling | -Included the evaluation of several fingers' detection tasks<br>-User study was carried out as proofs | -3D object rotation not included |
| 2014 | Combining multi-touch input & device movement for 3D object manipulations in mobile AR environments (Marzo *et al.*) | TBI & DBI | Proposed to combine multi-touch input with DBI techniques for complete 6DOF 3D object manipulation | -Using hybrid interaction technique to improve the performance of TBI for 3D object manipulation task [61] | -Occlusion problem still retain (TBI)<br>-Cannot provide a seamless interaction experience (TBI & DBI techniques contained in different categories)<br>-No user study conducted |
| Year | Research Title (Researchers) | Interaction Technique Categories (TBI/ MBI/ DBI) | Key Findings | Primary Strengths | Primary Drawbacks |
| 2015 | Integrated view-input AR interaction for virtual object manipulation using tablets and smartphones (Tanikawa *et al.)* | DBI | Added visual rod to hold and select 3D object in DBI | -Enhanced realistic and intuitive 3D object manipulation<br>-User study was carried out as proofs | -No investigation to solve 3D object rotation problem in DBI and deviation problem when used to perform 3D object translation and rotation integrally |
| 2016 | Touch and hand gesture-based interactions directly manipulating 3D virtual objects in mobile AR (Kim & Lee) | MBI | Applied hand gestures: pinch, grab and open to do 3D object manipulation | -Can performed complete 6DOF 3D object manipulation<br>-User study was carried out as proofs | -Occlusion problem [91]<br>-Lack accurate tracking process [91]<br>-Additional sensor such as Leap motion or Kinect is needed |
| 2016 | A review of 3D gestures interaction for handheld AR (Yusof *et al.)* | MBI | Reviewed several researches relating to MBI until year 2014 | -Studied the current trend of MBI techniques<br>-Discussed the current strengths and drawbacks in MBI | -Only discussed one interaction category<br>-Not provide comprehensive review |
| 2016 | SlidAR: A 3D positioning method for SLAM-based handheld AR (Polvi *et al.)* | DBI (as comparator) | Proposed ray casting with epipolar geometry to improve the current 3D object positioning method that using DBI | -SlidAR performed better in 3D object positioning tasks compared with current DBI technique<br>-User study was carried out as proofs | -Not applicable in 3D object rotation |
| 2016 | A study on improving close & distant device movement pose manipulation for handheld AR (Samini & Palmerius) | DBI | Proposed solving 3D object rotation issue in DBI through the holding stage | -Applied hold and release actions to repose handheld mobile device to solve large-range 3D object rotation problem in DBI<br>-User study was carried out as proofs | -Frequent reposition actions of handheld mobile device slowed down 3D object rotation speed<br>-Over-rotation of handheld mobile device for large-range 3D object rotation still may occur and causes 3D object registration error<br>-Over-rotation of handheld mobile device also may cause user cannot view 3D object on the display screen |
| 2017 | 3D Interaction with virtual objects in a precisely aligned view using a see-through mobile AR system (Unuma & Komuro) | MBI | Improve the performance of 3D object manipulation on task completion time using MBI through aligning the user' view | -Provide solution to improve the current MBI technique for 3D object manipulation with faster task completion through aligning user' view to ensure the view see through the screen and outside the screen were consistent<br>-User study was carried out as proofs | -Prototype developed still immature and additional evaluation is required.<br>-The solution proposed not significantly improve the performance of MBI technique on 3D object manipulation task<br>-Additional hardware like depth camera is needed |

**TABLE 5.** Comparison between the touch-based (TBI), mid-air. Gestures-based (MBI) and device-based interaction (DBI) techniques for 3D object manipulation in handheld mobile AR.

| Issues | Interaction Technique | | |
|---|---|---|---|
| | TBI | MBI | DBI |
| Occlusion problem [60-63, 65, 66, 95, 96, 99] | Yes | Yes | No |
| Fatigue phenomenon [10, 21, 70, 107, 109] | Serious | Serious | Slight |
| Prior knowledge needed | Yes | No | No |
| Intuitiveness/ naturality [23, 39, 91, 95] | Low | High | High |
| Low precision [65, 91] | No | Yes | No |
| Position mismatch in 3D object translation task [46, 48, 61] | No | Yes | No |
| Orientation mismatch in 3D object rotation task [40, 46, 48, 61] | No | Yes | No |
| Position and orientation deviations in 3D object manipulation task [40, 53, 61, 62, 91] | Yes | Yes | Yes |
| 3DOF 3D object translation | Applicable | Applicable | Applicable |
| 3DOF 3D object rotation | Applicable | Applicable | Applicable |
| Large-range 3D object translation [74] | Applicable | Limited | Applicable |
| Large-range3D object rotation [40, 61, 62, 74] | Applicable | Limited | Limited |

levels (Serious or Slight) and the potential issues that may arise (Yes or No; Low or High; Applicable or Limited) among the three main interaction technique categories.

Occlusion (Fig. 13) is one of the common issues that has been mentioned frequently in the previous researches [60]–[62], [91], [94]–[96]. Occlusion always exists in the touch-based interaction techniques when users directly touch the 3D object on the screen while the 3D object touched may be occluded. On the other hand, self-occlusion occurs in mid-air gestures-based interaction when the user's own hand or finger causes the 3D object being occluded while bare-hand or fingers actions behind the camera scene but in front of the AR marker may also cause occlusion on the AR marker and then leads to the 3D object registration error.

Due to the occlusion problem, several solutions had been suggested such as [63], [65], and [99]. In [63], callout display method is used to display the occluded area near the finger touch but the entire finger below the touch still occluded while Paudisch and Chu [65] suggested new hardware design to perform touch input behind the touchscreen display that not yet been launched in the consumer market. In [99], self-occlusion had been spatially solved by estimating the occluded hand and finger motions and gestures, but it is expensive due to the huge calculation requirement and also the limitation of the numbers of dataset according to each of the potential hand and fingers motions. In this context, device-based interaction technique category is the only occlusion-free category showed in Fig. 14.

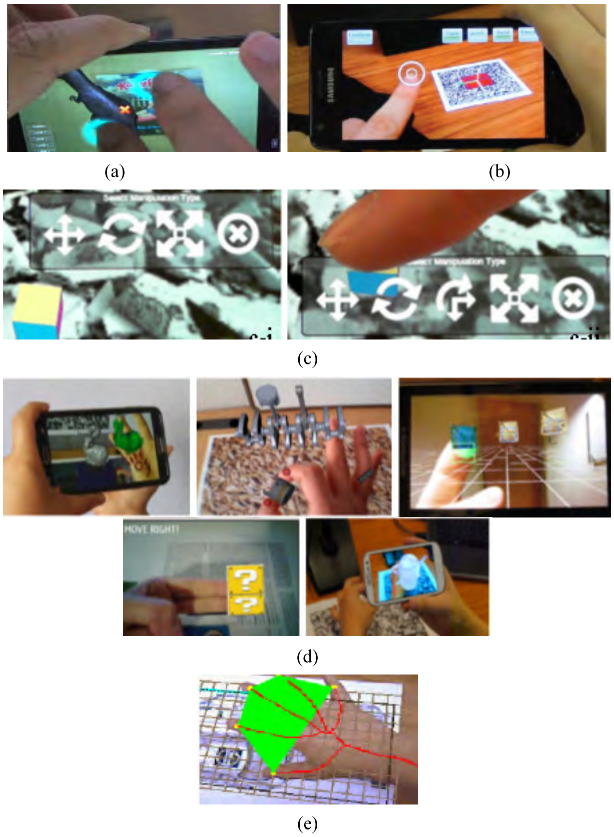


**FIGURE 13.** Some examples of occlusion cases happened in the touch-based and mid-air gestures-based interaction techniques; a virtual object occluded by the user's fingers [60], b indirect touch to avoid occlusion [66], c i-virtual object on screen ii-virtual object occluded by finger [61], d Self-occlusions [91], [94], [96], [97], e Major features of AR marker had been occluded [95].

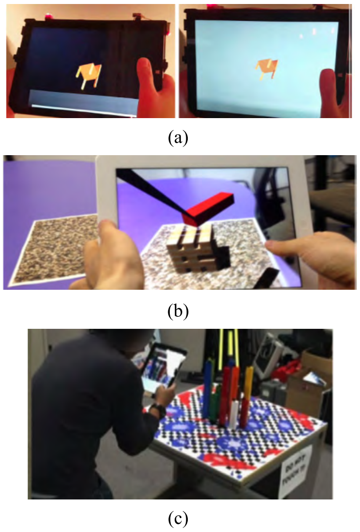


**FIGURE 14.** Occlusion had been excluded in the device-based interaction techniques. (a) In [40]. (b) In [39]. (c) In [23].

For fatigue phenomenon, it is linked with the ability for the user to hold the handheld mobile device with both hands when interacting. In [10], fatigue phenomenon has been highlighted and supported with [70] that holding device with one hand and using another hand for interaction might cause fatigue to the user. Based on [70], hand tremor and arm fatigue may occur when the user cannot rest their arms and hands. This effect is obvious in mid-air gestures-based interaction when a user needs to stretch his/her hand to perform long distance 3D object translation [108] or spin the wrist when performing 3D object rotation. In touch-based interaction, the user needs to hold the handheld mobile device and free up one hand to perform finger touch actions on the display screen. When multi-touch actions are required such as for 3D object rotation, fatigue for fingers occur [109]. Also, interactions that utilize both hands can effectively reduce fatigue compared with single-hand interaction [70]. Since fatigue phenomenon may affect the user's AR experience that directly determines the effective interaction time in prolonging the activity period. It is one of the problems that require serious attention. Compared with touch-based and mid-air gestures-based interaction techniques, the device-based interaction technique category seems more robust in this context because the user can interact using both hands in handheld mobile AR environments.

Prior knowledge is needed when the user uses multi-touch actions to perform 3D object manipulation in handheld mobile AR interfaces. Twist, drag, unhinged dual fingers are some of the control structures in the touch-based interaction techniques. Besides the difficulty when using multi-touch actions on the touchscreen display, the user also needs to know in advance the control structure implemented for a certain touch-based interaction technique. However, the mid-air gestures-based interaction techniques have been proposed due to its intuitiveness and naturality. Users utilize his/her hand and fingers to manipulate 3D objects as he/she does in the real world thus no prior knowledge is needed. In the device-based interaction, handheld mobile devices have been treated as sticky/holding tools to translate or rotate 3D object simulating the real case in our daily life such as using our own hand/fingers to pick an object to move to another position or rotate it. In the real world, our own hand and fingers are the sticky/holding tools while in the device-based interaction technique category, it is changed to our handheld mobile device. Thus, both mid-air gesture-based and device-based interaction techniques are intuitive and natural as stated in [23], [39], [91], and [95].

Position mismatch and low precision problem are another two (2) aspects in our investigation. These two (2) problems exist in the mid-air gesture-based interaction techniques when independently compared with the touch-based and device-based interaction technique categories for 3D object manipulation. Because of the occlusion, the position deviation problem might occur although users thought that they move the 3D object to the exact position [65]. There is also low precision in mid-air gestures-based interaction as stated previously in [65] and [91]. Although inaccuracy may also be present in touch-based interaction when multi-touch inputs are used causing instability and failure of the AR marker's tracking stage [91] while objects near the screen corners and edges are difficult to be selected and manipulated [91], solutions had been given by [61] to solve this problem through one-finger touch and drag separating DOF of 3D object translation and rotation to relief the difficulty of dual-finger action. Oppositely, device-based techniques have been proved more accurate and even faster compared with touch-based techniques when users perform 3D translation [61], [62]. Even so, position deviation may occur when the same input set is used for both 3D object translation and rotation [40], [46], [48], [61] in touch-based and device-based interaction techniques, and this position deviation problem appears more in mid-air gestures-based interaction since it happens not only for 3D object manipulation (both translation and rotation performed integrally) but also for 3D object translation and rotation independently.

For 3D fundamental tasks, 3D object translation and rotation are the basis. Both 3D object translation and rotation can be done through all the three main interaction techniques whereas users may feel difficult to perform large-range translation using the mid-air gestures-based interaction technique because of the restriction by the person's reach of hand [74]. Moreover, constraints towards large-range rotation also occur when using the mid-air gestures-based interaction techniques, limited by the movement around the wrist joint [74], which also happens within device-based interaction [40], [61], [62] because 3D object registration error occurs when the user rotates the handheld mobile device outside the viewable range or AR marker's trackable range. In this context, the touch-based interaction techniques seem to be more robust without limitation towards large-range 3D object translation and rotation.

## VII. CONCLUSION AND FUTURE DIRECTIONS

Currently, many of the existing handheld mobile AR applications are not considered very practical due to insufficient functionality and they do not fully answer to the needs of the users [17], [110]. Many design and technical challenges still remain and 3D object manipulation (including 3D object translation and rotation) is one of them. In order for handheld mobile AR to become widely accepted, the users must be able to create AR contents by positioning and rotating virtual objects in the real environment [111], [112].

After discussed the remaining issues related with 3D object manipulation techniques in handheld mobile AR and also their substantial effects in many fields, future directions for research in interaction techniques for 3D object manipulation in handheld mobile AR can be identified in the areas of occlusion, position deviation, accuracy, speed, intuitiveness and large-range manipulation that go beyond hybrid interaction and realistic AR experience as well as the wireless networking. Speed and accuracy rate are two indicators in most AR-related user studies [6], thus also have been included as

**TABLE 6.** Potential research directions in touch-based (TBI), mid-air gestures-based (MBI) and device-based (DBI) interaction techniques for 3D object manipulation in handheld mobile AR.

| Issues | Interaction Technique | | |
|---|---|---|---|
| | TBI | MBI | DBI |
| Occlusion problem [60-63, 65, 66, 95, 96, 99] | * | * | |
| Fatigue phenomenon [10, 21, 70, 107, 109] | * | * | |
| Intuitiveness/ naturality [23, 39, 91, 95] | * | * | * |
| Accuracy [6] | * | * | * |
| Speed [6] | * | * | * |
| Position mismatch [46, 48, 61] | | * | |
| Orientation mismatch [40, 46, 48, 61] | | * | |
| Position and orientation deviations (3D object manipulation) [40, 53, 61, 62, 91] | * | * | * |
| Difficulty of dual fingers action [58, 95] | * | | |
| Hand/fingers tracking and recognition [9, 10] | | * | |
| 3D object rotation (large-range) [40, 61, 62, 74] | | * | * |
| 3D object translation (large-range) [74] | | * | |
| Realistic AR experience | * | * | * |
| Slow or/and unstable wireless networking [7, 8] | * | * | * |

Note: * represents the highlighted issue that is applicable for the technique stated in line

one of the future directions while solutions can be focused on speeding up the 3D object manipulation with greater accuracy rate in handheld mobile AR.

Realistic AR experience also had been added as one of the future directions while solutions can be focused on improving the AR experience. Haptic aspect [39], [113], [114] for example, which is importance when AR system is applied in the real-world tasks such as assembly-related tasks while physic simulation that applicable on handheld mobile AR like force and deflection feedback can be visually done to improve the user's AR experience like he/she is manipulating the real object.

Wireless networking is one more issue to be focused on. Wireless networking is needed to communicate with other people and computers while on the run. Dynamic and flexible handheld mobile AR will rely on up-to-the-second information that cannot possibly be stored on the computing device before application run-time [7], [8].

In AR, wireless networking is crucially important in handheld tour guide and applications. Location-based AR applications such as AR outdoor navigation system always needs wireless networking to access to a remote server to update the current information [12]. Even a tracking process that requires GPS also needs the wireless network to function properly. For 3D object manipulation, the wireless network is essential, when the 3D content registered are stored in the remote server or involving multiple users.

The main issue related to wireless networking in handheld mobile AR may point to its stability and speed especially in rural area. The researches and studies in this area still lack and more effort should be put into this particular area.

Based on the remaining issue we have discovered together with the specific issues stated previously in Table 1, 2 and 3 and other aspects stated as speed, accuracy, realistic AR experience, and wireless networking, we categorized potential research directions followed by each interaction category in Table 6 as a rough guide for readers and researchers.

This paper reviewed the field of interaction techniques for 3D object manipulation in handheld mobile AR. Specifically, the three (3) primary interaction technique categories accessible until 2018 are touch-based interaction, mid-air gestures-based interaction and device-based interaction techniques. It provides insights into the fundamentals of interaction techniques for 3D object manipulation, as well as issues relating to each technique category. These interaction techniques focus on achieving 3D object manipulation tasks, that is, the complete 6DOF of 3D object manipulation of virtual content in handheld mobile AR environments. Besides 3D object manipulation, further discussion in addition to improving the user's AR experiences, including occlusion, hand and finger tracking or recognition, and other specific issues relating to each technique should be considered to achieve a quality of interaction experience as high as possible.

## ACKNOWLEDGMENT

Thanks go to other members of the Media and Game Innovation Centre of Excellence (MaGICX), Universiti Teknologi Malaysia for their time, effort, and enthusiasm. Gratitude is expressed to Malaysia Ministry of Education for kindly sponsoring and supporting this research. This study was undertaken during the study leave of the main author under the 2014 Federal Training (HLP) scholarship scheme and supported by Fundamental Research Grant Scheme (R.J130000.7809.5F093).

## REFERENCES

[1] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Trans. Inf. Syst.*, vol. E77-D, no. 2, pp. 1321–1329, Dec. 1994.

[2] R. T. Azuma, "A survey of augmented reality," *Presence*, vol. 6, no. 4, pp. 355–385, 1997.

[3] D. W. F. van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations," *Int. J. Virtual Reality*, vol. 9, no. 2, pp. 1–20, Jun. 2010.

[4] M. Mekni and A. Lemieux, "Augmented reality: Applications, challenges and future trends," in *Proc. 13th Int. Conf. Appl. Comput. Appl. Comput. Sci. (ACACOS)*, Kuala Lumpur, Malaysia, 2014, pp. 205–214.

[5] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47, Nov. 2001.

[6] A. Dey, M. Billinghurst, R. W. Lindeman, and J. Swan, "A systematic review of 10 years of augmented reality usability studies: 2005 to 2014," *Frontiers Robot. AI*, vol. 5, p. 37, Apr. 2018.

[7] P. Fraga-Lamas, T. M. Fernández-Caramés, O. Blanco-Novoa, and M. A. Vilar-Montesinos, "A review on industrial augmented reality systems for the industry 4.0 shipyard," *IEEE Access*, vol. 6, pp. 13358–13375, 2018.

[8] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui, "Mobile augmented reality survey: From where we are to where we go," *IEEE Access*, vol. 5, pp. 6917–6950, Apr. 2017.

[9] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Comput. Vis. Image Understand.*, vol. 108, nos. 1–2, pp. 52–73, Oct./Nov. 2007.

[10] C. F. Yusof, H. Bai, M. Billinghurst, and M. S. Sunar, "A review of 3D gesture interaction for handheld augmented reality," *J. Teknol. Sci. Eng.*, vol. 78, no. 2, pp. 15–20, Dec. 2016.

[11] J. Rekimoto, "A new you: From augmented reality to augmented human," in *Proc. 9th ACM Int. Conf. Interact. Tabletops Surf.*, Dresden, German, 2014, pp. 1–2.

[12] F. Zhou, H. B. L. Duh, and M. Billinghurst, "Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR," in *Proc. Int. Conf. Mixed Augmented Reality (ISMAR)*, Sep. 2008, pp. 193–202.

[13] A. W. Ismail, M. Billinghurst, M. S. Sunar, and C. S. Yusof, "Designing an augmented reality multimodal interface for 6DOF manipulation techniques," in *Intelligent Systems and Applications* (Advances in Intelligent Systems and Computing), vol. 868, K. Arai, S. Kapoor, and R. Bhatia, Eds. Cham, Switzerland: Springer, 2018, pp. 309–322.

[14] M. Billinghurst, H. Kato, and I. Poupyrev, "The MagicBook: A transitional AR interface," *Comput. Graph.*, vol. 25, no. 5, pp. 745–753, Oct. 2001.

[15] E. Kruijff, J. E. Swan, II, and S. Feiner, "Perceptual issues in augmented reality revisited," in *Proc. Int. Conf. Mixed Augmented Reality (ISMAR)*, Seoul, South Korea, Oct. 2010, pp. 3–12.

[16] *Wikitude*. Accessed: Oct. 31, 2018. [Online]. Available: https://www.wikitude.com/products/wikitude-sdk/

[17] T. Olsson, E. Lagerstam, T. Kärkkäinen, and K. Väänänen-Vainio-Mattila, "Expected user experience of mobile augmented reality services: A user study in the context of shopping centres," *Pers. Ubiquitous Comput.*, vol. 17, no. 2, pp. 287–304, Feb. 2013.

[18] L. Madden, *Professional Augmented Reality Browsers for Smartphones: Programming for Junaio, Layar and Wikitude*. London, U.K.: Wiley, 2011.

[19] M. Billinghurst, H. Kato, and S. Myojin, "Advanced interaction techniques for augmented reality applications," in *Virtual and Mixed Reality*, Prlando, FL, USA, R. Shumaker, Ed. Berlin, Germany: Springer-Verlag, 2009, pp. 13–22.

[20] T. H. Höllerer and S. Feiner, "Mobile augmented reality," in *Telegeoinformatics: Location-Based Computing and Services*. London, U.K.: Taylor & Francis, 2004, pp. 221–260.

[21] M. S. Hancock, S. Carpendale, F. D. Vernier, D. Wigdor, and C. Shen, "Rotation and translation mechanisms for tabletop interaction," in *Proc. 1st IEEE Int. Workshop Horizontal Interact. Hum.-Comput. Syst.*, San Jose, CA, USA, Jan. 2007, pp. 1147–1156.

[22] A. Martinet, G. Casiez, and L. Grisoni, "Integrality and separability of multitouch interaction techniques in 3D manipulation tasks," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 3, pp. 369–380, Mar. 2012.

[23] J. Polvi, T. Taketomi, G. Yamamoto, A. Dey, C. Sandor, and H. Kato, "SlidAR: A 3D positioning method for SLAM-based handheld augmented reality," *Comput. Graph.*, vol. 55, pp. 33–43, Apr. 2016.

[24] D. A. Bowman, E. Kruijff, J. LaVoila, and I. P. Poupyrev, *3D User Interfaces: Theory and Practice*. Reading, MA, USA: Addison-Wesley, 2005.

[25] M. Bajura, H. Fuchs, and R. Ohbuchi, "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient," in *Proc. 19th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, New York, NY, USA, 1992, pp. 203–210.

[26] S. Feiner, B. MacIntyre, and D. Seligmann, "Knowledge-based augmented reality," *Commun. ACM*, vol. 36, no. 7, pp. 53–62, Jul. 1993.

[27] W. Chen, "Historical Oslo on a handheld device—A mobile augmented reality application," *Procedia Comput. Sci.*, vol. 35, pp. 979–985, Jan. 2014.

[28] D. Mourtzisa, V. Zogopoulosa, and E. Vlachoua, "Augmented reality application to support remote maintenance as a service in the Robotics industry," *Procedia CIRP*, vol. 63, pp. 46–51, Jan. 2017.

[29] J. Polvi *et al.*, "Handheld guides in inspection tasks: Augmented reality versus picture," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 7, pp. 2118–2128, Jul. 2018.

[30] W. J. Rezende, E. S. Albuquerque, and A. P. Ambrosio, "Use of augmented reality to support education-creating a mobile E-learning tool and using it with an inquiry-based approach," in *Proc. 9th Int. Conf. Comput. Supported Educ. (CSEDU)*, Porto, Portugal, vol. 1, 2017, pp. 100–107.

[31] J. H. Shuhaiber, "Augmented reality in surgery," *Arch. Surg.*, vol. 139, no. 2, pp. 170–174, Feb. 2004.

[32] P. Vávra *et al.*, "Recent development of augmented reality in surgery: A review," *J. Healthcare Eng.*, vol. 2017, Aug. 2017, Art. no. 4574172.

[33] T. Kilgus *et al.*, "Mobile markerless augmented reality and its application in forensic medicine," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 5, pp. 573–586, May 2015.

[34] A. C. Boud, D. J. Haniff, C. Baber, and S. J. Steiner, "Virtual reality and augmented reality as a training tool for assembly tasks," in *Proc. Int. Conf. Inf. Vis.*, London, U.K., Jul. 1999, pp. 32–36.

[35] S. Webel, U. Bockholt, T. Engelke, and M. Peveri, "Augmented reality training for assembly and maintenance skills," *Robot. Auto. Syst.*, vol. 61, pp. 398–403, Apr. 2013.

[36] R. Radkowski, J. Herrema, and J. Oliver, "Augmented reality-based manual assembly support with visual features for different degrees of difficulty," *Int. J. Hum.-Comput. Interact.*, vol. 31, pp. 337–349, May 2015.

[37] M. Funk, T. Kosch, S. W. Greenwald, and A. Schmidt, "A benchmark for interactive augmented reality instructions for assembly tasks," in *Proc. Int. Conf. Mobile Ubiquitous Multimedia*, Linz, Austria, 2015, pp. 253–257.

[38] G. Dini and M. D. Mura, "Application of augmented reality techniques in through-life engineering services," *Procedia CIRP*, vol. 38, pp. 14–23, Jul. 2015.

[39] T. Tanikawa, H. Uzuka, T. Narumi, and M. Hirose, "Integrated view-input AR interaction for virtual object manipulation using tablets and smartphones," in *Proc. Int. Conf. Adv. Comput. Entertainment Technol. (ACE)*, Iskandar, Malaysia, 2015, Art. no. 7.

[40] A. Samini and K. L. Palmerius, "A study on improving close and distant device movement pose manipulation for hand-held augmented reality," in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol.*, Munich, Germany, 2016, pp. 121–128.

[41] J. Morris, *Android User Interface Development: Beginner's Guide*. Birmingham, U.K.: Packt, 2011.

[42] G. Waloszek, *Interaction Design Guide for Touchscreen Applications*. Newtown Square, PA, USA: SAP Design, 2008.

[43] Y. Wang, C. L. MacKenzie, V. A. Summers, and K. S. Booth, "The structure of object transportation and orientation in human-computer interaction," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Los Angeles, CA, USA, 1998, pp. 312–319.

[44] J. Rekimoto, "SmartSkin: An infrastructure for freehand manipulation on interactive surfaces," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Pittsburgh, PA, USA, 2002, pp. 113–120.

[45] A. D. Wilson, S. Izadi, O. Hilliges, A. Garcia-Mendoza, and D. Kirk, "Bringing physics to the surface," in *Proc. 1st Ann. ACM Symp. User Interface Softw. Technol.*, Monterey, CA, USA, 2008, pp. 67–76.

[46] A. Martinet, G. Casiez, and L. Grisoni, "The design and evaluation of 3D positioning techniques for multi-touch displays," in *Proc. IEEE Symp. 3D User Interfaces*, Waltham, MA, USA, Mar. 2012, pp. 115–118.

[47] M. Hancock, T. T. Cate, and S. Carpendale, "Sticky tools: Full 6DOF force-based interaction for multi-touch tables," in *Proc. ACM Int. Conf. Interact. Tabletops Surf.*, Banff, AB, Canada, 2009, pp. 133–140.

[48] J. L. Reisman, P. L. Davidson, and J. Y. Han, "A screen-space formulation for 2D and 3D direct manipulation," in *Proc. 22nd Annu. ACM Symp. User Interface Softw. Technol. (UIST)*, Victoria, BC, Canada, 2009, pp. 69–78.

[49] A. Cohé, F. Decle, and M. Hachet, "tBox: A 3D transformation widget designed for touch-screens," in *Proc. ACM CHI Conf. Hum. Factors Comput. Syst.*, Vancouver, BC, Canada, 2011, pp. 3005–3008.

[50] C. R. Brouet, R. Blanch, and M. P. Cani, "Understanding hand degrees of freedom and natural gestures for 3D interaction on tabletop," in *Proc. Int. Conf. INTERACT*, Cape Town, South Africa, 2013, pp. 297–314.

[51] J. Guo, Y. Wang, P. Du, and L. Yu, "A novel multi-touch approach for 3D object free manipulation," in *Proc. 3rd Int. Workshop Next Gener. Comput. Animation Techn. (AniNex)*, Bournemouth, U.K., 2017, pp. 159–172.

[52] J. Liu, O. K.-C. Au, H. Fu, and C.-L. Tai, "Two-finger gestures for 6DOF manipulation of 3D objects," *Comput. Graph.*, vol. 31, pp. 2047–2055, Sep. 2012.

[53] C. Telkenaroglu and T. Capin, "Dual-finger 3d interaction techniques for mobile devices," *Pers. Ubiquitous Comput.*, vol. 17, no. 7, pp. 1551–1572, Oct. 2013.

[54] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine, "Image plane interaction techniques in 3D immersive environments," in *Proc. Symp. Interact. 3D Graph. (I3D)*, Providence, RI, USA, 1997, pp. 39–43.

[55] D. Scheurich and W. Stuerzlinger, "A one-handed multi-touch mating method for 3D rotations," in *Proc. Int. Conf. Comput. Hum. Interact. (CHI)*, Paris, France, 2013, pp. 1623–1628.

[56] E. Rousset, F. Bérard, and M. Ortega, "Two-finger 3D rotations for novice users: Surjective and integral interactions," in *Proc. 12th Int. Work. Conf. Adv. Vis. Interfaces (AVI)*, Como, Italy, 2014, pp. 217–224.

[57] M. Xin, E. Sharlin, and M. Sousa, "Napkin sketch: Handheld mixed reality 3D sketching," in *Proc. ACM Symp. Virtual Reality Softw. Technol.*, Bordeaux, France, 2008, pp. 223–226.

[58] G. Lee *et al.*, "Freeze-Set-Go interaction method for handheld mobile augmented reality environments," in *Proc. 16th ACM Symp. Virtual Reality Softw. Technol. (VRST)*, Kyoto, Japan, 2009, pp. 143–146.

[59] S. Guven, S. Feiner, and O. Oda, "Mobile augmented reality interaction techniques for authoring situated media on-site," in *Proc. Int. Symp. Mixed Augmented Reality (ISMAR)*, Santa Barbard, CA, USA, Oct. 2006, pp. 235–236.

[60] J. Jung, J. Hong, S. Park, and H. S. Yang, "Smartphone as an augmented reality authoring tool via multi-touch based 3D interaction method," in *Proc. Int. Conf. Virtual-Reality Continuum Appl. Ind. (VRCAI)*, Singapore, 2012, pp. 17–20.

[61] A. Mossel, B. Venditti, and H. Kaufmann, "3DTouch and HOMER-S: Intuitive manipulation techniques for one-handed handheld augmented reality," in *Proc. Virtual Reality Int. Conf. Laval Virtual (VRIC)*, 2013, Art. no. 12.

[62] A. Marzo, B. Bossavit, and M. Hachet, "Combining multi-touch input and device movement for 3D manipulations in mobile augmented reality environments," in *Proc. Int. Conf. Spatial User Interaction (SUI)*, Honolulu, HI, USA, 2014, pp. 13–16.

[63] D. Vogel and P. Baudisch, "Shift: A technique for operating pen-based interfaces using touch," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, San Jose, CA, USA, 2007, pp. 657–666.

[64] A. Sears and B. Shneiderman, "High precision touchscreens: Design strategies and comparisons with a mouse," *Int. J. Man-Mach. Stud.*, vol. 34, no. 4, pp. 593–613, Apr. 1991.

[65] P. Paudisch and G. Chu, "Back-of-device interaction allows creating very small touch devices," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Boston, MA, USA, 2009, pp. 1923–1932.

[66] H. Bai, G. A. Lee, and M. Billinghurst, "Freeze view touch and finger gesture based interaction methods for handheld augmented reality interfaces," in *Proc. Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)*, Dunedin, New Zealand, 2012, pp. 126–131.

[67] J. M. Rehg and T. Kanade, "DigitEyes: Vision-based hand tracking for human-computer interaction," in *Proc. IEEE Workshop Motion Non-Rigid Agricult. Objects*, Austin, TX, USA, 1994, pp. 16–22.

[68] C. Jennings, "Robust finger tracking with multiple cameras," in *Proc. Int. Workshop Recognit., Anal., Tracking Faces Gestures Real-Time Syst.*, Corfu, Greece, 1999, pp. 152–160.

[69] J. Letessier and F. Bérard, "Visual tracking of bare fingers for interactive surfaces," in *Proc. 17th Annu. ACM Symp. User Interface Softw. Technol. (UIST)*, Santa Fe, NM, USA, 2004, pp. 119–122.

[70] H. Benko and S. Feiner, "Balloon selection: A multi-finger technique for accurate low-fatigue 3D selection," in *Proc. IEEE Symp. 3D User Interfaces*, Charlotte, NC, USA, Mar. 2007, pp. 22–29.

[71] O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz, "Interactions in the air: Adding further depth to interactive tabletops," in *Proc. 22nd Ann. ACM Symp. User Interface Softw. Technol.*, Victoria, BC, Canada, 2009, pp. 139–148.

[72] S. Strothoff, D. Valkov, and K. Hinrichs, "Triangle cursor: Interactions with objects above the tabletop," in *Proc. ACM Int. Conf. Interact. Tabletops Surf.*, Kobe, Japan, 2011, pp. 111–119.

[73] R. Y. Wong and J. Popović, "Real-time hand-tracking with a color glove," *ACM Trans. Graph.*, vol. 28, no. 3, Jan. 2009, Art. no. 63.

[74] N. Marquardt, R. Jota, S. Greenber, and J. A. Jorge, "The continuous interaction space: Interaction techniques unifying touch and gesture on and above a digital surface," in *Proc. 13th IFIP TC13 Conf. Hum. Comput. Interact. (INTERACT)*, Lisbon, Portugal, 2011, p. 16.

[75] V. Frati and D. Prattichizzo, "Using Kinect for hand tracking and rendering in wearable haptics," in *Proc. IEEE World Haptics Conf.*, Istanbul, Turkey, Jun. 2011, pp. 317–321.

[76] A. Kulshreshth, C. Zorn, and J. J. LaViola, "Poster: Real-time markerless Kinect based finger tracking and hand gesture recognition for HCI," in *Proc. IEEE Symp. 3D User Interface*, Orlando, FL, USA, Mar. 2013, pp. 187–188.

[77] B. R. Araùjo, G. Casiez, and J. A. Jorge, "Mockup builder: Direct 3D modeling on and above the surface in a continuous Interaction space," in *Proc. Graph. Interface*, Toronto, ON, Canada, 2012, pp. 173–180.

[78] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu, "A handle bar metaphor for virtual object manipulation with mid-air interaction," in *Proc. Int. Conf. Comput. Hum. Interact. (CHI)*, Austin, TX, USA, 2012, pp. 1297–1306.

[79] D. Mendes, F. Fonseca, B. Araùjo, A. Ferreira, and J. Jorge, "Mid-air interactions above stereoscopic interactive tables," in *Proc. IEEE Symp. 3D User Interfaces*, Minneapolis, MN, USA, 2014, pp. 3–10.

[80] D. Mendes, M. Sousa, R. Lorena, A. Ferreira, and J. Jorge, "Using custom transformation axes for mid-air manipulation of 3D virtual objects," in *Proc. 17th ACM Symp. Virtual Reality Softw. Technol. (VRST)*, Gothenburg, Sweden, 2017, Art. no. 27.

[81] J. Segen and S. Kumar, "Gesture VR: Vision-based 3D hand interface for spatial interaction," in *Proc. 6th Int. Multimedia Conf.*, Bristol, U.K., 1998, pp. 455–464.

[82] R. G. O'Hagan, A. Zelinsky, and A. Rougeaux, "Visual gesture interfaces for virtual environments," *Interact. Comput.*, vol. 14, no. 3, pp. 231–250, Apr. 2002.

[83] M. Baldauf, S. Zambanini, P. Frölich, and P. Reichl, "Markerless visual fingertip detection for natural mobile device interaction," in *Proc. 13th Int. Conf. Hum. Comput. Interact. Handheld Mobile Devices Services (MobileHCI)*, Stockholm, Sweden, 2011, pp. 539–544.

[84] S. Kratz, M. Rohs, D. Guse, J. Müller, G. Bailly, and M. Nischt, "PalmSpace: continuous around-device gestures vs. multitouch for 3D rotation tasks on mobile devices," in *Proc. Int. Work. Conf. Adv. Vis. Interfaces (AVI)*, New York, NY, USA, 2012, pp. 181–188.

[85] D. Fritz, A. Mossel, and H. Kaufmann, "Evaluating RGB+D hand posture detection methods for mobile 3D interaction," in *Proc. Virtual Reality Int. Conf.*, Laval, France, 2014, Art. no. 27.

[86] K. Dorfmüller-Ulhaas and D. Schmalstieg, "Finger tracking for interaction in augmented environments," in *Proc. IEEE ACM Int. Symp. Augmented Reality*, Washington, DC, USA, Oct. 2001, pp. 55–64.

[87] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn, "FingARtips: Gesture based direct manipulation in augmented reality," in *Proc. 2nd Int. Conf. Comput. Graph. Interact. Techn. Australas. South East Asia*, Singapore, 2004, pp. 212–221.

[88] R. Radkowski and C. Stritzke, "Interactive hand gesture-based assembly for augmented reality applications," in *Proc. 5th Int. Conf. Adv. Comput.-Hum. Interact. (ACHI)*, Valencia, Spain, 2012, pp. 303–308.

[89] A. Henrysson, J. Marshall, and M. Billinghurst, "Experiments in 3D interaction for mobile phone AR," in *Proc. 5th Int. Conf. Comput. Graph. Interact. Techn. Australia Southeast Asia (GRAPHITE)*, Perth, WA, Australia, 2007, pp. 187–194.

[90] B. K. Seo, J. Choi, J. H. Han, H. Park, and J. Park, "One-handed interaction with augmented virtual objects on mobile devices," in *Proc. 7th ACM SIGGRAPH Int. Conf. Virtual-Reality Continuum Appl. Ind. (VRCAI)*, Singapore, 2008, Art. no. 8.

[91] W. Hürst and C. van Wezel, "Gesture-based interaction via finger tracking for mobile augmented reality," *Multimedia Tools Appl.*, vol. 62, no. 1, pp. 233–258, Jan. 2013.

[92] H. Bai, L. Gao, and M. Billinghurst, "Poster: Markerless fingertip-based 3D interaction for handheld augmented reality in a small workspace," in *Proc. IEEE Symp. 3D User Interfaces*, Orlando, FL, USA, Mar. 2013, pp. 30–129.

[93] H. Bai, L. Gao, J. El-Sana, and M. Billinghurst, "Markerless 3D gesture-based interaction for handheld augmented reality interfaces," in *Proc. IEEE Int. Symp. Mixed Augmented Reality, Sci. Technol.*, Adelaide, SA, Australia, Oct. 2013, Art. no. 22.

[94] W. H. Chun and T. Höllerer, "Real-time hand interaction for augmented reality on mobile phones," in *Proc. Int. Conf. Intell. User Interfaces*, Santa Monica, CA, USA, 2013, pp. 307–314.

[95] H. Bai, L. Gao, J. El-Sana, and M. Billinghurst, "Free-hand interaction for handheld augmented reality using an RGB-depth camera," in *Proc. Conf. Exhib. Comput. Graph. Interact. Techn. Asia (SIGGRAPH Asia)*, Hong Kong, 2013, Art. no. 22.

[96] H. Bai, G. A. Lee, M. Ramakrishnan, and M. Billinghurst, "3D gesture interaction for handheld augmented reality," in *Proc. Conf. Exhib. Comput. Graph. Interact. Techn. Asia (SIGGRAPH Asia)*, Shenzhen, China, 2014, Art. no. 7.

[97] M. Kim and J. Y. Lee, "Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality," *Multimedia Tools Appl*, vol. 75, no. 23, pp. 16529–16550, Dec. 2016.
[98] Y. Unuma and T. Komuro, "3D interaction with virtual objects in a precisely-aligned view using a see-through mobile AR system," *ITE Trans. Media Technol. Appl.*, vol. 5, no. 2, pp. 49–56, Dec. 2017.
[99] Y. Jang, S. T. Noh, H. J. Chang, T. K. Kim, and W. Woo, "3D finger CAPE: Clicking action and position estimation under self-occlusions in egocentric viewpoint," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 4, pp. 501–510, Apr. 2015.
[100] J. Rekimoto, "Tilting operations for small screen interfaces," in *Proc. 9th Annu. ACM Symp. User Interface*, Washington, DC, USA, 1996, pp. 167–168.
[101] S.-J. Cho, Y. Sung, R. Murray-Smith, K. Lee, C. Choi, and Y.-B. Kim, "Dynamics of tilt-based browsing on mobile devices," in *Proc. Extended Abstr. Hum. Factors Comput. Syst. (CHI)*, San Jose, CA, USA, 2007, pp. 1947–1952.
[102] A. Henrysson, M. Billinghurst, and M. Ollila, "Face to face collaborative AR on mobile phones," in *Proc. 4th IEEE ACM Int. Symp. Mixed Augmented Reality*, Vienna, Austria, Oct. 2005, pp. 80–89.
[103] A. Henrysson, M. Billinghurst, and M. Ollila, "Virtual object manipulation using a mobile phone," in *Proc. Int. Conf. Augmented Tele-Existence Christchurch (ICAT)*, New York, NY, USA, 2005, pp. 164–171.
[104] A. Henrysson, M. Billinghurst, and M. Ollila, "Mobile phone based AR scene assembly," in *Proc. 4th Int. Conf. Mobile Ubiquitous Multimedia*, Christchurch, New Zealand, 2005, pp. 95–102.
[105] D. T. Huynh, K. Raveendran, Y. Xu, K. Spreen, and B. MacIntyre, "Art of defense: A collaborative handheld augmented reality board game," in *Proc. Int. Conf. Special Interest Group Comput. Graph. Interact. Techn. (SIGGRAPH)*, Orleans, LA, USA, 2009, pp. 135–142.
[106] T. Ha and W. Woo, "An empirical evaluation of virtual hand techniques for 3D object manipulation in a tangible augmented reality environment," in *Proc. Int. Conf. 3D User Interfaces (3DUI)*, Washington, DC, USA, Mar. 2010, pp. 91–98.
[107] Y. Xu *et al.*, "BragFish: Exploring physical and social interaction in co-located handheld augmented reality games," in *Proc. Int. Conf. Adv. Comput. Entertainment Technol.*, Yokohama, Japan, 2008, pp. 276–283.
[108] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, "Consumed endurance: A metric to quantify arm fatigue of mid-air interactions," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Toronto, ON, Canada, 2014, pp. 1063–1072.
[109] S. Boring, D. Ledo, X. Chen, N. Marquardt, A. Tang, and S. Greenberg, "The fat thumb: Using the thumb's contact size for single-handed mobile interaction," in *Proc. 14th Int. Conf. Hum.-Comput. Interaction Mobile Devices Services*, San Francisco, CA, USA, 2012, pp. 207–216.
[110] J. Grubert, T. Langlotz, and R. Grasset, "Augmented reality browser survey," Univ. Technol. Graz, Graz, Austria, Tech. Rep. ICG-TR-1101, Dec. 2011.
[111] S. Kurkovsky, R. Koshy, V. Novak, and P. Szul, "Current issues in handheld augmented reality," in *Proc. IEEE Int. Conf., Commun. Inf. Technol. (ICCIT)*, Hammamet, Tunisia, Jun. 2012, pp. 68–72.
[112] T. Langlotz, "AR 2.0: Social media in mobile augmented reality," Ph.D. dissertation, Univ. Technol. Graz, Graz, Austria, 2013.
[113] Y. Kurita, A. Ikeda, T. Tamaki, T. Okasawara, and K. Nagata, "Haptic augmented reality interface using the real force response of an object," in *Proc. 16th ACM Symp. Virtual Reality Softw. Technol. (VRST)*, Kyoto, Japan, 2009, pp. 83–86.
[114] A. Lécuyer, "Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback," *Presence, Teleoperators Virtual Environ.*, vol. 18, no. 1, pp. 39–53, Feb. 2009.

**EG SU GOH** received the B.S. degree in information technology and the M.S. degree in technical and vocational education from Tun Hussein Onn University, Malaysia, in 2006 and 2008, respectively. She is currently pursuing the Ph.D. degree with the Faculty of Engineering, School of Computing, UTM, under the supervision of Prof. M. S. Sunar and Dr. A. W. Ismail. She was a Lecturer with the Community College Tanjung Piai, Pontian, Malaysia. She is currently a Scholar and a Researcher with the Media and Game Innovation Centre of Excellence (MaGICX), UTM. She has published some articles in national conferences. Her research interests include computer graphics and information technology.

**MOHD SHAHRIZAL SUNAR** received the B.S. degree in computer science majoring in computer graphics from Universiti Teknologi Malaysia (UTM), the M.S. degree in computer graphics and virtual environment from The University of Hull, U.K., and the Ph.D. degree from the National University of Malaysia, in 2008. He was also mentoring technology startup companies and entrepreneurs in his research related field. His major field of study is real-time and interactive computer graphics and virtual environment. He is currently the Director of the Institute of Human Centered Engineering (iHumEn) and the Founding Director of the Media and Game Innovation Centre of Excellence (MaGICX), UTM. MaGICX becomes a subsidiary company of UTM, a joint-venture with Iskandar Regional Development Authority (IRDA), where he is currently the Chief Executive Officer. He is one of the CEO of Ministry of Higher Education of the Faculty Programme 2.0 Fellow, who attached with Petronas. Since 2009, he has been given responsibility to become the Head of the Department and leading the Virtual, Visualization, and Vision Research Group (ViCubeLab), UTM. He has been serving as an Academic Member of the Computer Graphics and Multimedia Department, Faculty of Computer Science and Information System, UTM, since 1999. He has published numerous articles in high-impact refereed journals, conference proceedings, and technical papers, including the article in magazines. He is an active Professional Member of ACM SIGGRAPH and the IEEE Computer Society. He is also a member of the Malaysian Society of Mathematics and Science (PERSAMA) and an Exco Member of the Malaysian Muslim Scientist Association (PERINTIS). He has been serving the government of the Malaysia Qualification Agency (MQA) Panel, as an Assessor, in the field of computing and multimedia, since 2005. The current research program that led by him are augmented reality, virtual reality, gamification, user interaction, and creative content technology. He has invented a number of innovative products from his research that are to be commercialized.

**AJUNE WANIS ISMAIL** received the B.S. degree in computer graphics and computer vision, and the M.S. and Ph.D. degrees from UTM, in 2016. She has started research on augmented reality, which is currently her main research area. She has completed the M.S. research mode degree in a short period of three semesters, and her research was vision-based tracking in augmented reality. In 2013, she joined the Human Interface Technology Laboratory New Zealand (HITLabNZ), University of Canterbury, as a Researcher, where she has completed her research in the Ph.D. degree in three years. She is currently the Head of Mixed and Virtual Environment Research Lab (Mivielab), UTM. She is currently a Senior Lecturer with the Universiti Teknologi Malaysia (UTM), Johore, Malaysia. Her research interests include augmented reality and mixed reality environments.

• • •