GG Interaction: a gaze–grasp pose interaction for 3D virtual object selection

Kunhee Ryu¹ · Joong-Jae Lee² · Jung-Min Park³

Received: 15 February 2017 / Accepted: 10 July 2019 / Published online: 19 July 2019
© The Author(s) 2019

Abstract

During the last two decades, development of 3D object selection techniques has been widely studied because it is critical for providing an interactive virtual environment to users. Previous techniques encounter difficulties with selecting small or distant objects, as well as naturalness and physical fatigue. Although eye-hand based interaction techniques have been promoted as the ideal solution to these problems, research on eye-hand based spatial interaction techniques in 3D virtual spaces has progressed very slowly. We propose a natural and efficient spatial interaction technique for object selection, which is motivated by understanding the human grasp. The proposed technique, gaze–grasp pose interaction (GG Interaction), has many advantages, such as quick and easy selection of small or distant objects, less physical fatigue, and elimination of eye-hand visibility mismatch. Additionally, even if an object is partially overlapped by other objects, GG Interaction enables a user to select the target object easily. We compare GG Interaction with a standard ray-casting technique through a formal user study (participants = 20) across two scenarios. The results of the study confirm that GG Interaction provides natural, quick and easy selection for users.

Keywords Human–computer interaction · Virtual reality · 3D object selection technique · Natural user interaction

1 Introduction

Selection and manipulation of a virtual object are essential features for interacting with a virtual environment. Methods for 3D object selection in virtual environments have been widely studied, [11,23,28,38]. Additionally, immersive 3D virtual environments have recently gained attention as next generation technologies due to their applicability in VR gaming, fully immersive movie theaters, VR medical operating

rooms, and VR social networks. In a virtual environment, selection is one of the most fundamental interaction features [3]. To provide users with a more immersive virtual environment, it is important to develop an efficient, natural, and intuitive selection technique for 3D virtual objects.

1.1 Selection techniques and design factors for a 3D virtual environment

Ray-casting is one of the most well known pointing-based selection techniques [11,21]. Ray-casting is widely used because it is convenient and intuitive. It is similar to selecting an object with a laser pointer. Kopper et al. [19], and Steed and Parker [32] noted that ray-casting is slow and error-prone when the visual scale of a target is small due to the object size, occlusion, or distance from the user. Particularly, as the distance from the origin (hand or device) to a point along a ray increases, a small movement of a user's hand is mapped to an increasingly large movement of the point. This makes it difficult for a user to select faraway objects. These drawbacks become more evident in a dense 3D virtual environment.

Forsberg et al. [14] proposed the aperture technique, which is a modification of the flashlight technique [10]. The tech-

✉ Jung-Min Park
pjm@kist.re.kr
Kunhee Ryu
ryuhhh@gmail.com
Joong-Jae Lee
arbitlee@gmail.com

¹ School of Robotics, Kwangwoon University, 60, Kwangwoon-ro, 1-gil, Nowon-gu, Seoul, Republic of Korea

² Center of Human-Centered Interaction for Coexistence, 5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul, Republic of Korea

³ Center for Intelligent & Interactive Robotics, Robot and Media Institute, Korea Institute of Science and Technology, 5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul, Republic of Korea

nique provides a fixed spread angle cone for selection and a user selects a virtual object by including it in the cone. With the aperture technique, a user can control the spread angle of the selection cone. Even though this technique provides a user with a method of reducing the ambiguity problem, it is still not completely freed from ambiguity when objects are aligned along the center line of the selection cone. In these cases, the closest object to the selection device is selected. To enable selection of an object overlapped by others, Bacim et al. [5] introduced the SQUAD technique, which is based on progressive refinement. A user first selects a group of objects, then recursively narrows ambiguity by selecting sub-groups until the desired object is selected. This approach improves accuracy as long as a user makes no mistakes. SQUAD requires several steps for selection. Performing several steps to make a selection hinders a user from feeling immersion, even though the technique conceptually guarantees accurate selection in extremely dense environments. It is important that selection not only be accurate, but also fast and natural to provide immersive and seamless interaction to a user in a 3D virtual environment.

Naturalness is an essential part of the design of interaction techniques. Strong semantic mapping between a virtual selection technique and a real-world action gives a user sense of naturalness. Many researchers agree that ‘naturalness’ means representing natural real-world behavior [7,24,34,37]. To provide users with a sense of naturalness, researchers proposed selection techniques with novel metaphors. Benko and Feiner [8] proposed the Balloon Selection method, which selects an object by controlling a balloon. In this technique, a user generates a balloon attached to a string by controlling his/her fingers, and then selects a 3D virtual object by correctly positioning their fingers. Song et al. [31] proposed a selection and manipulation technique using a handle bar metaphor. To select an object, a user generates a virtual handle bar through a bimanual gesture, ‘Point’, and then selects a 3D virtual object with another bimanual gesture, ‘Close’. Despite the novelty of these techniques, we do not use a balloon or handle bar to select objects in real life. Mimicking real-world behavior is likely a better approach for giving users a sense of naturalness. In the real world, we select objects in several different ways, such as grasping, pointing, looking, speaking to a listener. Among these, grasping is perhaps the most familiar action for selecting objects. If a grasping motion can be used for 3D object selection, it could give users a good sense of naturalness.

One additional factor to consider when designing 3D selection techniques is physical fatigue. If the selection technique causes lots of physical fatigue, selection becomes increasingly time consuming and inaccurate, which causes inconvenience to the user. Argelaguet and Andujar [2], as well as Argelaguet et al. [4] discussed a problem in hand-rooted pointing techniques called *eye-hand visibility*

mismatch. The hand-rooted pointing technique is a generic term for pointing techniques where the origin of the ray is the user’s hand. Due to occlusions, the set of objects visible to user’s eyes might be different from the objects visible from the hand position. For example, when a relatively small object such as a dice is stacked on top of a wide object such as a plate, a user with a hand-rooted pointing technique will be unable to select the dice from the bottom because the plate is occluding the dice. Unless the user aligns their hand to the viewing direction, this problem will require physical effort to select the virtual object from an uncomfortable position. Using gaze information is one way to reduce arm fatigue and overcome the *eye-hand visibility mismatch*. We propose a natural selection technique that combines gaze and hand motion, which is motivated by human grasping behavior, to select a 3D virtual object.

1.2 Eye-hand based selection techniques

Following the work of Hutchinson et al. [15] and Jacob [17] concerning gaze interaction, several studies have been performed. According to Bonino et al., using gaze information for 3D interaction has several advantages [9]. First, it is faster than other input modalities [34]. Second, it is easy to operate because a user does not need any particular training to simply look at an object. Third, it reduces physical fatigue caused by arm and hand movements. Finally, gaze information contains clues about the user’s areas of interest.

Chatterjee et al. presented a set of interaction techniques combining gaze and free-space hand gestures [12]. The gaze-hand based interactions are complementary, mitigating the issues of imprecision and limited expressivity found in gaze-alone techniques. Results showed that gaze–gesture combinations can outperform systems that use gaze or gesture alone.

Pfeuffer et al. introduced gaze-shifting as a new mechanism for switching between input modes based on the alignment of manual input and a user’s visual attention [25]. Even though gaze-shifting uses a pen as the primary input device, it employs the user’s gaze for supplementary input and support of other modalities.

Zhang et al. investigated the potential of integrating gaze with hand gestures for remote interaction with a large display, focusing on user experience and preference [39]. They conducted a lab study with a photo-sorting task and compared two different interaction methods: gesture only and a combination of gaze and gesture. The results showed that a combination of gaze and gesture input leads to significantly faster selection, reduced hand fatigue, and increased ease of use compared to using only hand gestures.

Each of these studies shows that multimodal interaction techniques using a combination of gaze and gesture are beneficial in terms of user experience and preference. However,

Table 1 Summary of the eye-hand based selection techniques for a virtual object

	Dim.	Selection		Gestures	Feature
		Pointing	Confirmation		
Chatterjee et al. [12]	2D	Gaze-ray	Gesture	Grasp/shake	Select objects of various sizes
Pfeuffer et al. [25]	2D	Gaze-ray	Pen-based touch	–	Accurate pointing required
Pfeuffer et al. [26]	3D	Gaze-ray	Gesture	Pinch	Uni-/bi-manual selection
Pouke et al. [27]	3D	Gaze-ray	Gesture	Jerk/shake/tilt	Accurate pointing required
Yoo et al. [36]	3D	Face orientation	Gesture	Pull/push	Accurate pointing required

the aforementioned studies only covered interaction in 2D virtual spaces. Unlike in 2D virtual space, spatial interaction techniques in 3D virtual space have made little progress.

Yoo et al. presented an interaction technique that combines gaze and hand gestures for interaction with a large-scale display [36]. The proposed 3D interaction technique enables a user to select, browse, and shuffle 3D objects using hand movements. It is motivated by human behaviors such as pulling a lever or pushing a button on a machine. The results showed that users prefer the interaction method that combines gaze and hand gestures, and the authors determined that the reason for this is because the combined method is more attentive and immersive than a conventional UI.

Pouke et al. proposed a gaze and non-touch gesture based interaction technique for mobile 3D virtual spaces on tablet devices [27]. Users can select objects with gaze, as well as grab and manipulate objects using non-touch gestures. The gestures set consists of *Grab/Switch*, *Tilt*, *Shake*, and *Throw*. *Grab/Switch* is a fast downward jerk used for selecting objects and switching between interaction modes (movement and rotation). *Tilt* is used for performing movement and rotation of an object. Users can release a grabbed object with *Shake*, which is performed by quickly turning the hand left and right as if turning a doorknob.

Pfeuffer et al. proposed *gaze+pinch interaction* [26] which combines user's gaze and gesture for the selection of an object in 3D virtual space. The method provides interaction capabilities on targets at any distance without relying on an extra controller device. However, the pinch gesture is additional motion required to select a virtual object and not natural because the users do not use the pinch to select the actual object. They proposed 'flick away' to refine selection for overlapping objects, potentially offset gaze estimation can still lead to a false positive.

The above interaction techniques are certainly novel interaction in virtual space, but there is room for improvement in intuitiveness or naturalness when compared to selection in the real world. The techniques use specially coded gestures to select and manipulate objects. It may be easy to memorize the actions, but the actions and outcomes are not directly related. In the system proposed by Yoo et al., users perform

the action of a mid-air hand press to select an object [36]. On the other hand, users of the system proposed by Pouke et al. must perform the jerk action [27]. Both gestures are unlikely to be associated with the action of selection in the real world, and are more similar to a mouse click. While these gestures can be useful in certain scenarios, it is difficult to ensure that they will retain that usefulness when applied to a virtual space mimicking the real world. Furthermore, users must inevitably learn and adapt to the meaning of each gesture. Additionally, these methods require accurate pointing for the desired object, as they do not provide a method for the selection of objects that are partially overlapped with others in a dense environment. Table 1 presents the related works on selection method using gaze (or face orientation) and hand input.

In our research, the proposed gaze–grasp pose interaction (GG Interaction) technique is designed to achieve the following goals:

- Fast and easy selection for small or distant objects.
- Fast and easy selection for an object partially overlapped by others.
- High resemblance to human grasping.
- Low physical fatigue.
- Elimination of the eye-hand visibility mismatch.
- Smooth transition from selection to 6DOF manipulation.

2 Gaze–grasp pose interaction

2.1 Overview

When we want to grasp an object in the real world, we begin by looking at the object. This is a searching step, which is a prerequisite for selecting an object. Next, we actually grasp the object. We expand this simple behavior to the realm of 3D virtual object selection. Figure 1 is an illustration of GG Interaction. A user can select an object by looking at it and performing a grasping action. In Fig. 1, the user is selecting the red cylindrical object. GG Interaction consists of two stages: *Generating a candidate group* and *Picking out a target object*.

Fig. 1 An overview of GG Interaction

Generating a candidate group—A candidate group is defined as the group of objects which fall within an arbitrary threshold distance from the line-of-sight. The user does not need to point exactly at a target object with his/her eyes. The circle in Fig. 1 represents a candidate group and the red line represents the line-of-sight of the user. The candidate group in Fig. 1 contains four objects based on the definition of a candidate group.

Picking out a target object—This step is the procedure for picking out the target among the objects in a candidate group. A candidate group can contain the target object along with several other objects as shown in Fig. 1. The *Picking out* procedure is only performed on a candidate group. To pick out the target object, GG Interaction uses hand gestures. As shown in Fig. 1, the user selects the target object by making a motion such as a ‘grasp’. The technique picks out the target object by comparing selection costs. For the object i , the selection cost e_{sel}^i consists of the gaze cost e_{gaze}^i and the grasp pose cost e_{grasp}^i , and the detailed definitions for the costs will be provided in Sect. 2.2. Note that a candidate group is continuously regenerated in each frame based on the user’s line-of-sight. Thus, when the moves their hands for grasping, they can select a target object instantly, reducing overall selection time.

GG Interaction uses gaze and hand information simultaneously. Generally, gaze information is highly sensitive to the sensor noise and hard to control accurately. In addition, it is hard to select an object that is placed behind other objects. This will likely cause undesired selections. Likewise, using hand information only is problematic when there are many objects of the same size in the scene. GG Interaction uses both the gaze and the hand information to identify the object that the user selects. This approach, which uses both complementary modalities, is less error-prone than unimodal interactions and more useful in implementing an immersive virtual environment [18].

2.2 Implementation

We describe the two stages of GG Interaction in detail in this section. Let the group of all objects and the candidate group be denoted by $\mathbb{G}$ and $\mathbb{C} \subset \mathbb{G}$, respectively.

Generating a candidate group—A candidate group is generated by calculating the gaze cost of each of the i th object, e_{gaze}^i , which is an evaluation of how close an object is to the user’s line-of-sight. To find the elements in $\mathbb{C}$, the system tracks a user’s gaze ray and calculates gaze cost for each object. We assume that user’s eye point, p , is fixed and known. Using a gaze tracker, we obtain a directional vector, u , and calculate the equation of a straight line parameterized by t when $l(t) = p + tu$. The gaze cost for the i th object is defined as follows.

$$e_{gaze}^i = \|o_i q_i\|, \quad \text{for } i \in \mathbb{G} \quad (1)$$

where o_i is the spatial position of the i th object and q_i is one foot of perpendicular distance from l to o_i . Whether or not the i th object is an element of $\mathbb{C}$, is determined by the following decision rule:

$$\text{Decision rule 1, } \begin{cases} i \in \mathbb{C}, & \text{if } e_{gaze}^i < c_1 \\ i \notin \mathbb{C}, & \text{otherwise} \end{cases}$$

where c_1 is a positive threshold value. The candidate group is re-generated on a frame-by-frame basis. As you can see in Fig. 1, the candidate group may contain several objects when the target object is overlapped by others.

Picking out a target object—To pick out a target object, GG Interaction compares user grasping size, d , with the width of each object, w^i , in a candidate group. The technique then picks out the object with the minimum cost and sets it as the target object by using the results from this comparison. Grasping size, d is defined as the minimum distance from the thumb tip to other fingertips of a user. The system first finds the finger which has the shortest distance from the thumb tip, and uses that distance as d . Thus, we obtain d as follows:

$$d = \min\{\|p_1 p_i\|\} \quad \text{for } i = 2, \dots, 5. \quad (2)$$

where p_1 to p_5 are the spatial position vectors from the thumb to each fingertip. The grasp pose cost for the i th object, e_{grasp}^i , is calculated by the following equation:

$$e_{grasp}^i = |w^i - d|, \quad \text{for } i \in \mathbb{C} \quad (3)$$

where w^i is the width of the i th object. Note that i in Eq. (3) is an element of $\mathbb{C}$. Calculating the grasp pose cost is only performed on elements of $\mathbb{C}$. The selection cost for each

Algorithm 1 GG Interaction

Require: No selected object

```

1: // generating a candidate group
2: for  $i \in \mathbb{G}$  do
3:   Calculate  $e_{gaze}^i$ 
4:   if  $e_{gaze}^i < c_1$  then
5:      $i \in \mathbb{C}$ 
6:   else
7:      $i \notin \mathbb{C}$ 
8:   end if
9: end for
10:
11: // picking out a target object
12: Calculate  $d$ 
13: for  $i \in \mathbb{C}$  do
14:   Calculate  $e_{grasp}^i$  and  $e_{sel}^i$ 
15: end for
16: Find  $\bar{i}$ 
17:
18: if  $e_{sel}^{\bar{i}} < c_2$  then
19:    $\bar{i}$ th object is 'selected'
20: end if

```

object, e_{sel}^i is calculated as follows:

$$e_{sel}^i := \alpha^T e^i = [\alpha_1 \quad \alpha_2] \begin{bmatrix} e_{gaze}^i \\ e_{grasp}^i \end{bmatrix}, \text{ for } i \in \mathbb{C} \quad (4)$$

where α_1 and α_2 are weight values for the contribution of the selection cost and $\|\alpha\| = 1$. The system then finds the object with the minimum e_{sel}^i among all $i \in \mathbb{C}$. Let the object with the minimum e_{sel}^i be $\bar{i}$, and, the system will pick out the target object based on the following decision rule:

$$\text{Decision rule 2, } \begin{cases} \bar{i} \text{ is 'Selected',} & \text{if } e_{sel}^{\bar{i}} < c_2 \\ \text{'None',} & \text{otherwise} \end{cases}$$

where c_2 is a positive threshold value for picking out the target object from the candidate group. The algorithm for implementation of GG Interaction is shown in Algorithm 1. Lines 1 through 9 in Algorithm 1 are associated with *generating a candidate group*, and lines 11 through 20 are associated with *picking out a target object*.

2.3 Characteristics

Figure 2 illustrates the procedure of selecting a target object using GG Interaction. GG Interaction uses gaze information to specify a ROI (region of interest). In this case, the user is not required to look exactly at the target object. This relaxed requirement reduces user eye fatigue that is generated by voluntary control, stemming from attempts to accurately pinpoint the target object with user's gaze. It also reduces

errors from gaze jittering during the selection task. The picking out procedure uses fingertip information, which helps a user to feel naturalness due to a close resemblances to real-world behavior of grasping. 'Grasp' is one of the most well mapped behavior for 'select'. Additionally, selecting with a hand gesture such as grasping an object enables the user to feel a seamless transition from the selection task to the positioning task [10]. Once the target is selected, a user can manipulate the target in 6DOF using their hand. GG Interaction reduces arm fatigue that stems from the user moving their hand or arm to a specific position to select an object. The user can position their hand anywhere that feels comfortable, because hand position does not affect selection. GG Interaction uses grasping size to pick out the target object. In the case where the target object is overlapped by others with different widths, a user can select the target object by using proper grasping size. Furthermore, selecting a small or distant target with GG Interaction is easy because the user is not required to gaze exactly at it and they can use their own previous experience about the width of various objects. If a user wants to select a book, placed at a distance, it would be demanding work using pointing techniques, because the object appears small to the user. GG Interaction however, uses the real size of the object for selection. In other words, when the book is placed at a distance, the user can select it by looking at it and forming their hand into the grasp pose with a grasping size similar to the size of the book regardless of the distance. Finally, eye-hand visibility mismatch does not occur when using GG Interaction, because the system picks out a target object from the candidate group generated from the user's gaze.

3 User study

We conducted within-subjects experiments to compare GG Interaction with a standard ray-casting technique. This method has a smaller sample size than a between-subjects design and can detect differences between design metrics. It has the disadvantage in that a learning effect can occur. To counteract carryover effects, we employ counterbalancing.

Both selection techniques utilize dwell time (700ms) to select an object without the Midas Touch problem [17]. The experiments consist of objective and subjective components. For objective component, we compute a *selection time* value for both tests in the following manner. Let t_1 be the time when the target is indicated using a visual cue (changing color and drawing a box), and t_2 be the time when the target is successfully selected. Then, *selection time* = $t_2 - t_1$. Note that *selection time* contains both user reaction time (recognizing a target object) and dwell time. Additionally, we

Fig. 2 A user is selecting a distant, overlapped object [29]. The target object is the larger box in the red circle. The blue bar indicates a section of the gaze ray. **a** User is looking a target object for selection. The user's line-of-sight is passing through two boxes. In this case, the candidate group contains both boxes. **b** To select the target object, the user performs a gesture like grasping. GG Interaction performs *picking*

out a target object according to the user's hand information. **c** Target object (the larger box) is selected. Now the user manipulates the selected object. **d** Once an object is selected, GG Interaction only uses hand information (position and orientation) to move the selected object according to the user's hand movements

record a *misselection* value for both tests as the number of misselections (selection of a non-target object) between t_1 and t_2 per trial. For the subjective component, subjects filled out a questionnaire that rates mental effort, physical effort, general comfort, ease of selection, naturalness, intuitiveness, and adaptability with both techniques. All subjective questions were composed based on [30,35] and the scores were rated with five-point Likert scales. Feedback mechanism for each technique is as follows. The feedback for ray-casting is in the form of a ray emitted from the device [20]. Feedback for GG Interaction is in the form of a ray projected from the user's eye. Prior to the experiments, all users went through a calibration procedure for Leonar3Do, as well as the gaze and hand tracker. For both techniques, the graphical feedback on the object is the brightening of the object to be selected.

3.1 Participants

Twenty unpaid participants (six females, fourteen males), aged from 22 to 40 years (mean age = 28.9, SD = 3.5), took part in our user study. They were all right-handed and reported previous exposure to 3D VR systems, such as playing 3D video games, using a head-mounted display (HMD) or watching 3D movies.

3.2 System setup

The display used was a 40" 3D monitor with a resolution of 1920×1080 pixels. The distance from the display to a user was approximately 70 cm, and all users wore 3D polarized glasses during the experiments. For the ray-casting technique, we used Leonar3Do [20], a commercial input device. For GG Interaction, we used the Tobii Rex [33], gaze tracker

Fig. 3 System setup for GG Interaction. An RGBD sensor and a gaze tracker are used. A user performs the *toy block test* with GG Interaction

Fig. 4 A user performs the *3D reciprocal tapping test* with ray-casting

for gathering gaze data. For gathering hand information, we used PrimeSense Carmine 1.09, an RGBD sensor, and 3Gear Nimble SDK [1]. The experimental program was executed on a desktop PC with an Intel i7-4790 CPU, 8 GB RAM, an NVIDIA GeForce GTX780, and Microsoft Windows 8.1. Figures 3 and 4 illustrate the overall system setup for both techniques.

3.3 Two scenarios

We designed two experimental scenarios: a *Toy block test* and a *3D Reciprocal tapping test*. Subjects were asked to perform both scenarios with GG Interaction and ray-casting. Before beginning both scenario, subjects were given 3 min to practice with both techniques. The total number of trials is 1440.

3.3.1 Toy block test

The *Toy block test* is a simple object manipulation scenario. The setup is shown in Fig. 3. Blocks have different shapes such as cube, triangular prism, and cylinder. In this scenario, a trial is defined in the following manner. Subjects were asked to select the target object indicated by a bright box. Once the target object is selected, the user must move the object

to the goal position. After the user releases the target near the goal position, a new target object will be designated. No overlapped object exists in this scenario. Subjects completed this scenario using two interaction techniques: ray-casting and GG Interaction. Each user performed three attempts and each attempts consisted of six trials. Each user performed a total of 36 trials in this scenario across both interaction techniques. In total, 720 results were recorded for the 20 participants.

3.3.2 3D reciprocal tapping test

The *3D reciprocal tapping test* is a 3D version of the Reciprocal Tapping Task and Dragging Test [16,22]. The *Dragging Test* and *Reciprocal Tapping Task* are tests for performance of non-keyboard input in 2D space. We expanded the tests to a 3D virtual space as shown in Fig. 4. Dice with three different sizes (small = 60 mm, medium = 90 mm and large = 120 mm) are radially positioned, and in some cases overlap with other dice. In this scenario, a trial is defined in the following manner. Subjects were asked to select the target object indicated by a green color. After the target object is successfully selected, the user must move the object to the home position which is in the center of the 3D virtual space (black dice). If the user positions the target object near the home position (within 30 mm), a green cube appears around the home position. After the user releases the target object near the home position, a new target object will be designated by changing its color to green. In this scenario, the total number of dice is 16 (8 red dice, 4 white dice, and 4 sky blue dice). The diagonal red, white and sky blue dice with respect to the center dice (dark die) are partially overlapped by others in 3D space as shown in Fig. 4. Subjects completed this scenario using two interaction techniques: ray-casting and GG Interaction. Each user performed three attempts and each attempts consisted of six trials. Each user performed 36 total trials in this scenario across both interaction techniques. In total, 720 results were recorded for the 20 participants.

3.4 Results

In this section, we discuss the results of the user study. We begin by nothing that there was no interaction effect between the two scenarios. Additionally, we divided the 18 trials for each selection technique into three attempts. Thus, six trials were performed per attempt for both GG Interaction and ray-casting. Selection time is the time from when the target object is assigned to selected. Error rate is a mis-selection rate. For instance, if there were three misselections before the selection for the target object, the error rate would be 75%.

Fig. 5 Selection time with SEMs of different object sizes for both interaction techniques

- **Selection time**—The results for selection time for various object sizes are presented in Fig. 5. For selection time, we performed three-way repeated measures using ANOVA with three independent variables: selection technique, object size, and attempt. Reported p -values and *post-hoc* include Bonferroni correction. There was a statistically significant effect from the selection technique ($F(1, 19) = 43.986, p < 0.001$), object size ($F(2, 38) = 173.225, p < 0.001$), and attempt ($F(2, 38) = 6.464, p < 0.005$). There was also statistically significant interaction in technique-size ($F(2, 38) = 11.704, p < 0.001$), and technique-attempt ($F(2, 38) = 4.452, p < 0.05$). Other interactions were not statistically significant.

Post-hoc—Mean selection time was 3.30 ± 2.08 s with GG Interaction, and 6.21 ± 4.50 s with ray-casting. Mean selection time for small, medium, and large objects with ray-casting were 8.68 ± 5.86 s, 4.44 ± 1.46 s, and 5.50 ± 3.84 s respectively. Mean selection time for small, medium, and large objects with GG Interaction were 3.42 ± 2.43 s, 2.92 ± 1.42 s, and 3.56 ± 2.22 s respectively. Mean selection time for the first, second, and third attempts with ray-casting were 7.33 ± 5.45 s, 5.82 ± 3.87 s, and 5.47 ± 3.81 s respectively. Mean selection time for the first, second, and third attempts with GG Interaction were 3.45 ± 1.95 s, 3.52 ± 2.36 s, and 2.94 ± 1.88 s, respectively.

- **Error rate**—The results for error rate for various object sizes are presented in Fig. 6. For error rate, we performed three-way repeated measures using ANOVA with three independent variables: selection technique, object size, and attempt. There was a statistically significant effect from the selection technique ($F(1, 19) = 5.123, p <$

Fig. 6 Error rate with SEMs of different object sizes for both interaction techniques

0.05), object size ($F(2, 38) = 10.660, p < 0.001$), and attempt ($F(2, 38) = 3.423, p < 0.05$). There was also a statistically significant interaction in technique-size ($F(2, 38) = 6.733, p < 0.005$). Other interactions were not statistically significant.

Post-hoc—Mean error rate was $21 \pm 43\%$ with GG Interaction, and $32 \pm 55\%$ with ray-casting. Mean error rate for small, medium, and large objects with ray-casting were $61 \pm 77\%$, $10 \pm 11\%$, and $25 \pm 39\%$, respectively. Mean error rate for small, medium and large objects with GG Interaction were $25 \pm 52\%$, $12 \pm 17\%$, and $25 \pm 49\%$, respectively. Mean error rate for the first, second, and third attempts with ray-casting were $45 \pm 63\%$, $27 \pm 42\%$, and $24 \pm 54\%$, respectively. Mean error rate for the first, second, and third attempts with GG Interaction were $23 \pm 44\%$, $18 \pm 42\%$, and $21 \pm 43\%$ respectively.

- **Subjective rating questionnaire**—Fig. 7 displays the mean rating for each of the seven questionnaire topics. A Friedman test revealed that there were significant differences in the ratings for general comfort ($\chi^2(1) = 4.765, p < 0.05$), naturalness ($\chi^2(1) = 9.941, p < 0.005$), and adaptability ($\chi^2(1) = 4.571, p < 0.05$) between the two techniques. For the mental and physical effort, a lower score is favored and the opposite is true for the other cases.

4 Discussion

From the results, one can see that GG Interaction provides better performance than standard ray-casting in terms of mean selection time.

The mean selection time of GG Interaction was 47% shorter than that of ray-casting on average. The mean selec-

Fig. 7 Results of subjective rating for each of the selection techniques with SEMs

Fig. 8 Mean selection time and mean error rate with SEMs for overlapped and non-overlapped cases for both interaction techniques

tion time for both techniques contains reaction time and dwell time. This could be one reason why overall mean selection time is larger than in the results of previous studies. In Fig. 5, it can be seen that the mean selection time for GG Interaction for various object sizes is relatively even, while the mean selection time for ray-casting for small objects is relatively large compared to other object sizes. This reflects the chronic problem of difficulty in selecting small objects. Thus, GG Interaction is relatively more robust than ray-casting in terms of selection time for various object sizes.

The mean error rate with regards to object size for GG Interaction is relatively low (21%) compared to that of ray-casting (32%). Specifically, this difference comes from cases of small objects (mean error rate for small objects with SEMs: GG Interaction = $25 \pm 6\%$ and ray-casting = $61 \pm 10\%$). These results for ray-casting contain relatively long selection time and high error rate when compared to other studies [11,13].

This is because our user study scenarios contain overlapping object cases. Figure 8 shows the mean selection time and the error rate for cases when the target object is over-

lapped (visually screened) by others and not overlapped by others. In terms of selection time, the difference in performance between the two techniques is bigger in overlapped cases than non-overlapped (visually fully open) cases. When it comes to mean error rate, ray-casting provided better performance (1.7%) in non-overlapped cases. GG Interaction, however, provided better performance in overlapped cases. These results may imply that GG Interaction could provide better performance than ray-casting in practical scenarios, which contain many objects.

In the subjective evaluation, subjects indicated that GG Interaction was more comfortable than the ray-casting technique. This is because in the *3D Reciprocal Tapping Task*, users had to bring the hand-held device (Leonar3Do) close to their eyes in order to resolve eye-hand mismatch. In terms of naturalness, the mean score for GG Interaction was higher than ray-casting. Some users commented that it would be very helpful to add kinesthetic or haptic feedback, particularly for GG Interaction. Some users were confused concerning recognition of the size of objects, which was reflected in the error rate.

Although GG Interaction provides better performance than ray-casting, there are some potential limitations. Because GG Interaction calculates cost values using the width of objects, an algorithm that defines object width is necessary particularly for objects with complex shapes. One approach to solving this problem is to use a minimum bounding box [6]. For GG Interaction, a minimum bounding box can be defined as the smallest box containing all parts of an object. Additionally, current GG Interaction only supports one handed interaction. This means that a user cannot select a large object, such as a desk or a bed, which is impossible to grasp with one hand. This limitation can be overcome by expanding GG Interaction to use both hands. Another limitation is observed when many objects with the same width overlap along the user's line of sight. Assuming that the cost of each

object is exactly the same as all other objects, and less than c_2 , GG Interaction considers the closest object to the user to be the selected object. This may differ from the user's intention. Furthermore, the threshold value c_1 for generating a candidate group is a design parameter. Although we used a fixed threshold ($c_1 = 100$ mm) in this study, more optimized thresholds should be considered to improve the performance of GG Interaction.

5 Conclusion

A natural 3D selection technique, GG Interaction, is proposed. It has several advantages including easy selection of small, distant, or overlapping objects, less arm fatigue, and a high resemblance to a human grasping motion. GG Interaction utilizes gaze-hand information. Gaze information is used for generating a candidate group, and hand information is used for picking out a target object from the candidate group. Therefore, the users are not required to look at the target exactly, which minimizes eye fatigue. Furthermore, there is no eye-hand mismatch because the system picks out the target object from a candidate group generated based on a user's gaze. GG Interaction's performance and advantages are demonstrated through a formal user study, where it is compared to a standard ray-casting technique. GG Interaction provides better performance than ray-casting in cases with overlapping objects. Additionally, fluctuation in object sizes has a smaller impact on GG Interaction than on ray-casting in terms of selection time and error rate. Finally, users indicated that GG Interaction is more natural and easy to use in their subjective rating questionnaires. For future work, we plan to investigate how selection time is impacted by various feedback methods such as sound, haptic, and kinesthetic feedback.

Funding This research was supported by the Global Frontier R&D Program on "Human-centered Interaction for Coexistence" funded by the National Research Foundation of Korea grant funded by the Korean Government (MSIP) (2011-0031425).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- 3Gear Nimble SDK (2014) <http://nimblevr.com/> Accessed 1 Feb 2017
- Argelaguet F, Andujar C (2009) Efficient 3d pointing selection in cluttered virtual environments. *IEEE Comput Graph Appl* 29(6):34–43
- Argelaguet F, Andujar C (2013) A survey of 3d object selection techniques for virtual environments. *Comput Graph* 37(3):121–136
- Argelaguet F, Andujar C, Trueba R (2008) Overcoming eye-hand visibility mismatch in 3d pointing selection. In: *Proceedings of the 2008 ACM symposium on virtual reality software and technology*, ACM, New York, NY, USA, VRST'08, pp 43–46
- Bacim F, Kopper R, Bowman DA (2013) Design and evaluation of 3d selection techniques based on progressive refinement. *Int J Hum Comput Stud* 71(7):785–802
- Barequet G, Har-Peled S (2001) Efficiently approximating the minimum-volume bounding box of a point set in three dimensions. *J Algorithms* 38(1):91–109
- Barfield W, Hendrix C, Bystrom K (1997) Visualizing the structure of virtual objects using head tracked stereoscopic displays. *Virtual reality annual international symposium*. IEEE 1997:114–120
- Benko H, Feiner S (2007) Balloon selection: a multi-finger technique for accurate low-fatigue 3d selection. In: *2007 IEEE symposium on 3D user interfaces*, pp 79–86
- Bonino D, Castellina E, Corno F, Russis LD (2011) Dogeye: controlling your home with eye interaction. *Interact Comput* 23(5):484–498
- Bowman D, Kruijff E, LaViola J, Poupyrev I (2004) 3D user interfaces: theory and practice. CourseSmart eTextbook, Pearson Education, London
- Bowman DA, Hodges LF (1997) An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: *Proceedings of the 1997 symposium on interactive 3D graphics*, ACM, New York, NY, USA, I3D'97, pp 35–38
- Chatterjee I, Xiao R, Harrison C (2015) Gaze+ gesture: expressive, precise and targeted free-space interactions. In: *Proceedings of the 2015 ACM on international conference on multimodal interaction*, ACM, pp 131–138
- Cournia N, Smith JD, Duchowski AT (2003) Gaze- vs. hand-based pointing in virtual environments. In: *CHI'03 extended abstracts on human factors in computing systems*, ACM, New York, NY, USA, CHI EA'03, pp 772–773
- Forsberg A, Herndon K, Zeleznik R (1996) Aperture based selection for immersive virtual environments. In: *Proceedings of the 9th annual ACM symposium on user interface software and technology*, ACM, New York, NY, USA, UIST'96, pp 95–96
- Hutchinson TE, White KP, Martin WN, Reichert KC, Frey LA (1989) Human-computer interaction using eye-gaze input. *IEEE Trans Syst Man Cybern* 19(6):1527–1534
- ISO/DIS 9241-9 (2000) Ergonomic requirements for office work with visual display terminals (vdts)—part 9: requirements for non-keyboard input devices. Iso, International Organization for Standardization, Geneva, Switzerland
- Jacob RJK (1991) The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans Inf Syst* 9(2):152–169
- Kaiser E, Olwal A, McGee D, Benko H, Corradini A, Li X, Cohen P, Feiner S (2003) Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In: *Proceedings of the 5th international conference on multimodal interfaces*, ACM, New York, NY, USA, ICMI'03, pp 12–19
- Kopper R, Bacim F, Bowman DA (2011) Rapid and accurate 3d selection by progressive refinement. In: *2011 IEEE symposium on 3D user interfaces (3DUI)*, pp 67–74
- Leonar3Do (2012) <http://leonar3do.com> Accessed 15 July 2017
- Liang J, Green M (1994) JDCAD: a highly interactive 3d modeling system. *Comput Graph* 18(4):499–506
- MacKenzie IS, Sellen A, Buxton WAS (1991) A comparison of input devices in element pointing and dragging tasks. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, New York, NY, USA, CHI'91, pp 161–166

23. Mine MR (1995) TR95-018 virtual environment interaction techniques. Technical report, Department of Computer Science University of North Carolina at Chapel Hill
24. Petersen N, Stricker D (2009) Continuous natural user interface: reducing the gap between real and digital world. In: Proceedings of the 2009 8th IEEE international symposium on mixed and augmented reality, IEEE Computer Society, Washington, DC, USA, ISMAR'09, pp 23–26
25. Pfeuffer K, Alexander J, Chong MK, Zhang Y, Gellersen H (2015) Gaze-shifting: direct–indirect input with pen and touch modulated by gaze. In: Proceedings of the 28th annual ACM symposium on user interface software & technology, ACM, pp 373–383
26. Pfeuffer K, Mayer B, Mardanbegi D, Gellersen H (2017) Gaze+Pinch interaction in virtual reality. In: Proceedings of the 5th symposium on spatial user interaction, ACM, pp 99–108
27. Pouke M, Karhu A, Hickey S, Arhippainen L (2012) Gaze tracking and non-touch gesture based interaction method for mobile 3d virtual spaces. In: Proceedings of the 24th Australian computer–human interaction conference, ACM, pp 505–512
28. Poupyrev I, Weghorst S, Billinghurst M, Ichikawa T (1997) A framework and testbed for studying manipulation techniques for immersive vr. In: Proceedings of the ACM symposium on virtual reality software and technology, ACM, New York, NY, USA, VRST'97, pp 21–28
29. Ryu K, Hwang W, Lee J, Kim J, Park J (2015) Distant 3D object grasping with gaze-supported selection In: Proceedings of 12th international conference on ubiquitous robots and ambient intelligence, IEEE, pp 28–30
30. Sears A, Jacko J (2009) Chapter 4: survey design and implementation in HCI. CRC Press, Boca Raton, Human factors and ergonomics
31. Song P, Goh WB, Hutama W, Fu CW, Liu X (2012) A handle bar metaphor for virtual object manipulation with mid-air interaction. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, NY, USA, CHI'12, pp 1297–1306
32. Steed A, Parker C (2004) 3d selection strategies for head tracked and non-head tracked operation of spatially immersive displays. In: 8th international immersive projection technology workshop, pp 13–14
33. Tobii Rex Deveolpment Kit (2001) <http://www.tobii.com> Accessed 15 July 2017
34. Ware C, Mikaelian HH (1987) An evaluation of an eye tracker as a device for computer input2. In: Proceedings of the SIGCHI/GI conference on human factors in computing systems and graphics interface, ACM, New York, NY, USA, CHI'87, pp 183–188
35. Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. *Presence* 7(3):225–240
36. Yoo B, Han JJ, Choi C, Yi K, Suh S, Park D, Kim C (2010) 3d user interface combining gaze and hand gestures for large-scale display. In: CHI'10 extended abstracts on human factors in computing systems, ACM, pp 3709–3714
37. Zhai S, Milgram P (1993) Human performance evaluation of manipulation schemes in virtual environments. In: Virtual reality annual international symposium. 1993 IEEE, pp 155–161
38. Zhai S, Buxton W, Milgram P (1994) The “silk cursor”: investigating transparency for 3d target acquisition. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, NY, USA, CHI'94, pp 459–464
39. Zhang Y, Stellmach S, Sellen A, Blake A (2015) The costs and benefits of combining gaze and hand gestures for remote interaction. In: Human–computer interaction. Springer, pp 570–577

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.