IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 November 2023, accepted 26 November 2023, date of publication 28 November 2023,
date of current version 1 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3337372

RESEARCH ARTICLE

GazeHand: A Gaze-Driven Virtual Hand Interface

JAEJOON JEONG[1], (Student Member, IEEE), **SOO-HYUNG KIM**[2], **HYUNG-JEONG YANG**[3], **GUN A. LEE**[4], AND **SEUNGWON KIM**[2]

[1]Department of Software Engineering, Chonnam National University, Gwangju 61186, South Korea
[2]School of Artificial Intelligence, Chonnam National University, Gwangju 61186, South Korea
[3]Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju 61186, South Korea
[4]University of South Australia, Adelaide, SA 5001, Australia

Corresponding author: Seungwon Kim (Seungwon.Kim@jnu.ac.kr)

This work was supported in part by the Ministry of Science and Information and Communications Technology (ICT) (MSIT) and the National Research Foundation of Korea (NRF), through the Innovative Human Resource Development for Local Intellectualization Support Program, under Grant IITP-2023-RS-2022-00156287; in part by the Artificial Intelligence Convergence Innovation Human Resources Development under Grant IITP-2023-RS-2023-00256629; in part by the Korea Government (MSIT) under Grant RS-2023-00219107; and in part by the ICT Research and Development Project through the Institute for Information Communication Technology Planning and Evaluation (IITP) under Grant 2022-0-00874.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Chonnam National University Institutional Review Board under Application No. 1040198-221121-HR-145-01.

ABSTRACT In this paper, we present a novel interface named GazeHand which is designed for distant object interaction in virtual reality. The GazeHand interface translates the virtual hands near to a distant object that the user is looking at and allows direct hand gesture interaction for manipulating an object. Either eye-gaze or head-gaze can be applied to the GazeHand interface to decide the translated positions of the hands. In a user study, we compared these two variants of the GazeHand interface, i.e., Eye-GazeHand and Head-GazeHand, with a baseline condition of a modified Go-Go interface. The results showed that both GazeHand interfaces support the high-level usability of the direct hand operation for distant object interaction while providing the benefits of the gaze interaction: speed and reachability. The two variants of GazeHand interfaces showed better performance than the modified Go-Go interface and the effects were more prominent as the difficulty level of the task increased. The most preferred interface was the Head-GazeHand which took the benefit of a stable head-gaze, while the Eye-GazeHand was less stable using eye-gaze. Meanwhile, the Eye-GazeHand interface showed its advantage over the Head-GazeHand when the task required much gaze movement as it used faster eye-gaze without requiring head movement.

INDEX TERMS Distant object interaction, pivot point, virtual hand, virtual reality.

I. INTRODUCTION

Interacting with a virtual object at a distance has been one of the common interests in 3D Virtual Reality (VR) interface design [1], [2], [3], [4]. Among many prior works, Go-Go [1] and HOMER [2] (Hand-centered Object Manipulation Extending Ray-casting) interfaces are the most well-known solutions, and they adopt hand gesture control because it is natural and intuitive to interact with an object. The Go-Go interface extends the virtual arm to reach the distant object and the HOMER interface uses the ray casting technology for selecting a distant object and positioning the hands at it for manipulation. However, the Go-Go interface is difficult to control the length of the arm and has imprecision of hand control with amplified movement in proportion to the ratio of arm extension. Additionally, the HOMER interface has a limitation in the hand movement range as the hand cannot move outside the arm-reachable area after being placed at the selected object.

Other researchers exploited the gaze technique [5] for distant object interaction to overcome the limited range of hand gesture. Since the object selection includes both indication and confirmation, gaze-based selection used the simple looking activity for the indication and several confirmation

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

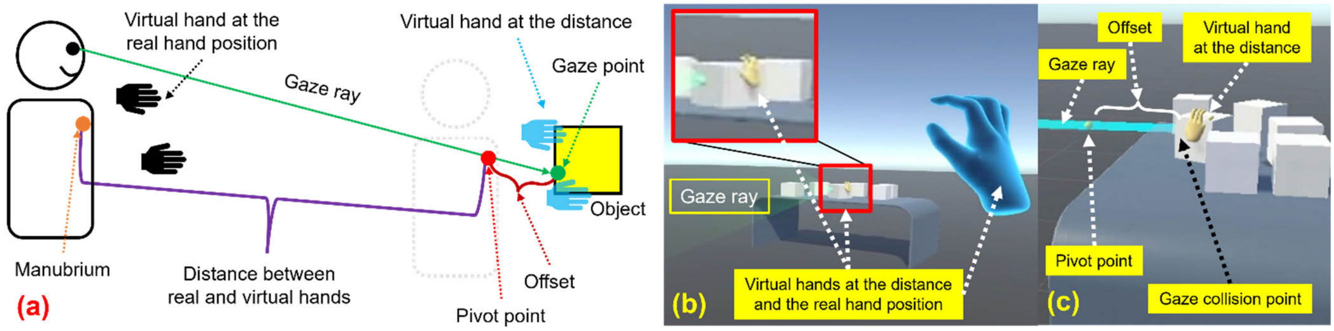
Virtual hand at the real hand position
Virtual hand at the distance
Gaze ray
Gaze point
Object
Manubrium
Distance between real and virtual hands
Offset
Pivot point
(a)
Gaze ray
Virtual hands at the distance and the real hand position
(b)
Offset
Virtual hand at the distance
Gaze ray
Pivot point
Gaze collision point
(c)

FIGURE 1. A concept image for the GazeHand. A user's virtual hands are positioned near a distant object that the user is looking at, based on the pivot and gaze points. The user can perform natural hand interaction with the distant object like in the real world: (a) schematics for translating virtual hand at the distance based on gaze; (b) user's view; (c) side view.

methods such as dwell time interaction [6], [7], [8], button press [9], [10] and hand pinch [11], [12]. The dwell time confirmation, however, usually hinders the task performance with them requiring 0.3∼1.3 seconds of gaze dwelling time, and the button press interaction requires additional gadgets such as a controller then reduces naturalness. Prior work found the gaze indication with the hand pinch confirmation was intuitive because hand pinch confirmation was easily extended to hand object manipulation [11], [12]. However, recent works on gaze and hand combinations still have issues of limited hand operation area by abandoning gaze interaction after distant-object selection [13] and reduced usability of natural hand control by abandoning one-hand object rotation [12].

In this paper, we introduce a novel interaction technique, named GazeHand (see Fig. 1) in which the bare and direct hand gestures are used for natural object interaction while the gaze translates the virtual hands near to a distant object. In this way, a user can perform natural hand-object interaction as in the real world and still take the benefits of the gaze interaction which supports fast movements and is able to reach every part of the visible task space. Moreover, it supports a two-hand operation simultaneously and requires less hand movement because major translation is done by the gaze interaction.

In order to demonstrate and validate the benefit of the GazeHand interface, we present the design and implementation details and also report on a user experiment evaluating the proposed method. To the best of our knowledge, this research is the first to introduce and evaluate the interface that uses the gaze only for translating virtual hands and combines it with natural hand gestures for object selection and manipulation.

II. RELATED WORK

In this section, we describe previous works using gaze, hand, and both for distant object interaction, to identify their limitations.

A. GAZE SELECTION AND MANIPULATION

Gaze has the benefit of being faster than other selection interfaces such as mouse and hand selection [14], [15], [16]. It also allows people to select unreachable objects [17] and express their point of interest to a system or other people [18], [19], [20].

Despite the benefits, gaze selection has a limitation of the Midas Touch Problem [3], [21] where users unintentionally select objects by looking at them. To solve the problem, several researchers [9], [12], [22] divided the gaze selection into two steps of interactions: indication by looking at activity and confirmation by several different methods. Dwell time interaction [6], [7] is one of the famous confirmation methods for gaze selection by calculating the time of gaze hovering at an object and confirming the selection if the time is over the threshold dwell time. Besides, eye blinking is also introduced as another confirmation method for gaze selection [23], [24].

Others have used a controller or hand gesture for confirmation in selection. A well-known example of using a button clicking for confirmation [25] is Microsoft HoloLens. A user looks at an object for indication while wearing the Microsoft HoloLens and clicks a button on the handheld controller for confirmation. Pfeuffer et al. [12] introduced an interface that confirmed a selection by hand pinch interaction that is also supported on HoloLens.

Prior research also used gaze for object manipulation [6], [26], [27] but their methods were alternative when a proper hand manipulation is not available because hand manipulation is more natural and effective than gaze manipulation. Additionally, since the gaze interface provided only point information, several researchers tried to extend its use by splitting complex interaction into a sequence of multiple steps (e.g., using dwell time interaction twice for object manipulation [6]).

Some researchers reported the difficulty of using the eye-gaze interaction with the accuracy level [28] and suggested using the head-gaze as an alternative [29], [30]. The head-gaze is defined with the head-mounted display (HMD) forwarding direction with a center point of the HMD view [31], [32]. Previous works [33], [34] reported that head-gaze interaction is preferred to eye-gaze interaction because it showed more stable performance than the eye-gaze interaction. Some others reported that eye-gaze interaction usually has less head movement than head-gaze [35].

Some works combined eye-gaze and head-gaze for selection. One study [36] used the head-gaze and eye-gaze to confirm selection when both gaze pointers are matched. Another study [37] explored both head-gaze and eye-gaze to predict a user's desired target object.

B. HAND SELECTION AND MANIPULATION

Humans use hands for holding (selecting) and manipulating objects in real life, so many researchers adopted hand gestures for direct object manipulation [38], [39] to establish natural interaction. However, it is difficult to interact with distant objects out of reach.

To solve the issue, several works such as Go-Go [1] and HOMER [2] brought hands to the target object by either extending an arm or using ray-casting. However, they still have limitations. Since the Go-Go interface [1] extends arms, the hands at the end of the arm could move too much with a small physical arm movement at the rate of the extension, and it makes it difficult to conduct precise hand interaction. Additionally, the arm extension is limited with their extension formula, so their hand interaction has limited operation area. To overcome such limitation the HOMER interface [2] uses the ray-casting for selection and then positions the virtual hand to the target object for manipulation. It still has the issue of the hand movement range being limited to the arm-reachable area after being placed at the selected object.

C. HAND INTERACTION WITH GAZE INPUT

Pfeuffer et al. [17], [40] developed a system allowing users to perform the selection and manipulation anywhere in the 2D screen. The position of the gaze point is the area that the user's hand screen interaction is applied to, so it could reduce hand movement. Additionally, several researchers used hand and gaze for menu selection [41], [42], [43], [44], text entry [45], or 2D region selection [46].

Recently, researchers started to introduce interfaces that combine hand and gaze interaction for 3D VR selection and manipulation with an HMD. Pfeuffer et al. [12] introduced the 'Gaze+Pinch' interface which allows users to select an object by gaze indication and hand pinch confirmation. After selection, the interface uses the relative position of the user's physical hand in interacting with objects. A user can start hand translation at any hand position with a pinch gesture and translate an object according to the relative position of the current hand to the initial hand pinch pose. Additionally, the interface allows two hand gestures for scaling and rotating objects. However, they exploited the relative position of two hands for rotating objects (according to the rotation of the line between the two hands) and it is not natural compared to real-world hand manipulation. Additionally, the 'Gaze+Pinch' interaction applied a formula (Movement $_{\text{Object}}$ = Movement $_{\text{Hand}}$ * Distance $_{\text{Object\_to\_user}}$) that amplifies grabbing-hand movement at the rate of the distance to the grabbed object, so precise manipulation of the distant object might be difficult.

Ryu et al. [13] adopted the hand and gaze selection for a distant object in a dense environment with many objects. The gaze refers to an area, and the objects in the gazed area become candidates for selection. Among the candidates, the user chooses one with a grasping gesture to indicate the width of the target object by the distance between the thumb and other fingers. After selection, users can manipulate it with natural hand gesture. The range of the manipulation, however, is still limited to the arm-reachable area. Additionally, there is an ambiguity on which object to select if there are multiple objects having the same width.

Yu et al. [9] introduced two interactions: 'ImplicitGaze' and '3DMagicGaze' using gaze and hands for object selection and manipulation. In the '3DMagicGaze', a user indicates an object with a gaze and confirms the selection with a controller button press. Then, the system creates a spherical area (with the radius at a tangent 10 degrees of the length between the user and the selected object) where the controller-hand object manipulation is supported. If the gaze goes beyond the spherical area, the object snaps to the gaze point direction (without changing its depth to the user) and the system creates another spherical area at the following gaze fixation then the snapped object can be manipulated by the controller hand in the spherical area. The 'ImplicitGaze' interface is the same as the '3DMagicGaze' interface but has a dynamic spherical area that becomes larger when the user keeps looking at it. As a result, Yu developed an interface that uses both gaze and controller-based hand interactions not only for selection but also for manipulation. They used the spherical area as a solution for unstable gaze movement and then allowed controller-hand manipulation in the sphere.

D. LIMITATIONS OF PREVIOUS WORK

In previous studies, we found that the gaze interface provided point information and it can move fast and reach any point regardless of the distance (as far as it is visible) [14], [15]. However, interacting with objects solely relying on the gaze-point information was too complex and may reduce the intuitiveness and easiness for the use.

The hand interaction is the most natural method for object selection (by grabbing) and manipulation because it is the way people do in the real world. However, it had mainly two issues for interacting with a distant object: limited hand interaction area [2], [13] and difficulty in conducting precise hand interaction with the amplified hand movement (such as in the Go-Go interface [1]).

Several researchers [9], [12], [13] have tried to solve the limitation of gaze and hand interactions by combining them together, but their approaches reduced the usability of natural hand manipulation (e.g., using relative position for object interaction and one hand object rotation is not supported [12], or using a controller rather than bare hand interaction [9]), and still did not properly adopt the benefit of the gaze interaction (e.g., hand-operation limited to arm reachable area [13]). Additionally, while they used gaze for object indication in the process of object selection, our approach does not include any gaze-object interaction.

III. THE GazeHand INTERFACE

In this section, we describe how we designed and developed the GazeHand interface, how it operates, and its two variants: Eye-GazeHand and Head-GazeHand.

A. ROLES OF GAZE AND HAND INTERACTION

Considering the pros and cons of the gaze and hand interaction (described in the section II.D), we combine them with the roles of gaze and hand interaction for the GazeHand interface as below:

- The hand interaction performs object selection and manipulation because it is more natural than the gaze interaction.
- The gaze interaction is used for translating hands to accelerate hand movement, so it solves the issue of slow movement of hands.
- The gaze translates the hands to every part of the visible task space regardless of its distance for resolving the issue of hand operation being limited to arm reachable area.

With the roles, the GazeHand interface takes the benefits of the natural and intuitive bare hand control for distant object interaction but adopts the two major benefits of the gaze interaction: fast-moving and being able to reach distant task space. The virtual hands are synchronized with the real-world hands at a one-to-one mapping rate, so their behavior is natural and similar to the real world.

B. TRANSLATING HANDS BY GAZE

To implant the benefits of the gaze interaction onto the hands, we translate the hands to the gaze point, so the GazeHand interface can move fast and go every part in the given space like the gaze point. However, translating hands simply to the gaze point has an issue in using hands. In the real world, there is a room between hands and an object when having the hand interaction. Since the gaze point is the collision point between the gaze ray and the object, there is no room for hand interaction against the object.

This can be solved by generating a pivot point in front of the collision gaze point and attaching the hands to it, so there is a room where the hands can move around (see Fig. 1). We position the pivot point on a gaze ray and the distance between pivot point and gaze point is two-thirds of average arm length [47] (Female: 72.20cm, Male: 78.65cm) by following Poupyrev's Go-Go interaction (which reports this as comfortable distance). We call this distance, an offset. Additionally, we match the pivot point to the manubrium (see Fig. 1) because the manubrium is placed at the top center of the front side of a user's body, and their arm's height is similar to the manubrium's height.

C. STABILITY OF THE PIVOT POINT

During object manipulation, the depths of the gaze and pivot points can change as the object is rotated (see Fig. 2).

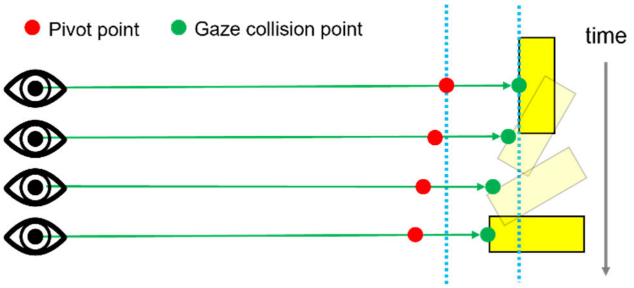

FIGURE 2. Issue of changing depths of gaze collision and pivot points when the object is rotated.

Algorithm: The GazeHand Interface

```
// initialization of variables and constants
Set the offset as the two-thirds of the arm length from a user
Set the delta as (the position of the HMD - the position of the user's
      manubrium)
Set the max_depth as 10 // can be changed depending on the context
Set the initial last_collider as null
Set the initial last_manipulated_object as null
Set the initial gaze_point as the position of the user's manubrium
Set the initial pivot_point as the position of the user's manubrium
Set the initial hand_distance as 0
Set the initial depth as 0

while true do
    // Calculate the pivot point
    Get ray information (origin and direction) from the HMD
    according to the type of gazes
    Generate a ray
    Raycast along the ray
    if the ray collides with any object then
        if the object != last_collider || the object !=
        last_manipulated_object then
            Update the depth as the magnitude of
            (position of collided object – origin)
        end if
        Update the current gaze_point as (origin + depth *
        direction)
        if the object != last_manipulated_object then
            Update the current pivot_point as (collision
            point of collided object – offset * direction)
        end if
        Render the ray with its length set as depth
    else
        Update the current gaze_point as (origin + max_depth *
        direction)
        if last_manipulated_object != null then
            Update the current pivot_point as (gaze_point
            – offset * direction)
        Render the ray with its length set as max_depth
    end if
    if the object != null then
        Update the last_collider as the object
    end if
    if a user is grabbing the object then
        Update the last_manipulated_object as the object
    end if
    Render the pivot point and gaze point on the ray
    // Position the virtual hand
    Set the hand_distance as max(0, depth – offset)
    Set the position of hands at the distance as (the position of the
    user's hand + delta + hand_distance * direction)
```

FIGURE 3. The GazeHand Interface algorithm that creates and updates a pivot point and moves the hand at the distant target object. Variables are in italic type. Vectors or points are underlined. * indicates scalar multiplication.

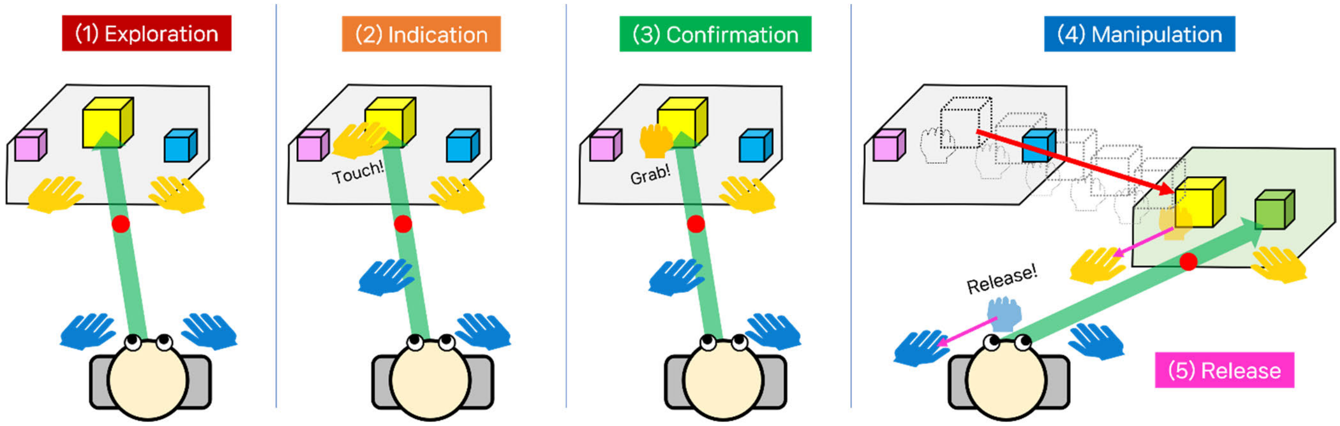

FIGURE 4. Continuous five steps of object manipulation in GazeHand. Black hands are a user's actual hands, and blue hands are their virtual hands for object manipulation. A green arrow indicates a user's gaze ray, a red arrow shows the path the object follows, and two pink arrows describe hand movement after releasing the hand. Each pink arrow has the same length. A red dot indicates a pivot point.

For example, if the object rotates and has a different depth of gaze point as shown in Fig. 2, then the pivot point can be moved back and forth. This unstable movement makes it difficult to control the hand interaction. To solve the issue, we make rules for calculating the depth of the pivot point. The overall GazeHand algorithm, including pivot point calculation, is described in Fig. 3.

- When grabbing an object, the system does not update the depths of gaze and pivot points with the collision between the grabbed object and the gaze ray.

With this rule, the hand interaction can be stable while grabbing an object, but another issue occurs when continuing the last manipulation with the same object once after releasing it. The system updates the depths when releasing the object even though a user tries to continue the next manipulation in a series by the previous one. To solve it, we add the following rule.

- If keeping the gaze on the last manipulated object, the system does not update the depths.

This rule was inspired by the fact that a user keeps looking at the manipulated object to check and continue the manipulation. With these rules, the depth of the pivot point remains unchanged until completing the manipulation of an object, so would establish a high level of stability.

In other cases where there is a collision between the gaze ray and another object that is not grabbed and not continuing manipulation, the GazeHand interface calculates the pivot point with the collision object, so the hands and grabbed object move to nearby it. These may reduce the depth control interaction by automatically moving the grabbed object to a target object. Additionally, we implemented another mode that does not update depth when grabbing objects.

D. GazeHand OPERATION

When a user puts down his/her hands, our interface only displays a gaze pointer. When hands are held up and in a tracked area, the system displays a half-transparent gaze ray and a pivot point with virtual hands. When using the GazeHand, there are five steps to interact with objects: exploration, object indication, select confirmation, manipulation, and release. More detailed explanations are described below and in Fig. 4.

- Exploration: The state when a user looks up the target object. If the user holds her hand up, the virtual hands appear near the position where she is looking at.
- Indication: The state when the user moves the virtual hands, and they reach the target object and collide with them.
- Confirmation: The state when a user grabs the target after the indication. The target object is in the control of hand and gaze interaction.
- Manipulation: The state when a user manipulates the object with the hand and gaze interaction. The gaze movement makes a fast and large translation of the hands and the grabbed object, and the hand control makes precise translation and orientation.
- Release: The state when a user releases the target object with ungrabbing hand interaction.

E. TWO GAZE MODES

There are two types of gazes: Eye-gaze and head-gaze, and our GazeHand interface has two variants with two gaze types.

Eye-GazeHand uses the eye-gaze ray and pointer with tracking information. The eye tracker takes a live pupil image and decides the direction of where the user is looking. With the gaze direction, the system creates a gaze ray from the pupil and finds a collision point with an object to place the gaze pointer. With the gaze ray and the gaze pointer, the system creates the pivot point as described in the previous sections and positions the virtual hands at it, then forms the Eye-GazeHand interface. It is controlled not only by the pupil movement but also by the head movement because the head movement is always applied to the pupil movement.

Head-GazeHand is similar to the Eye-GazeHand except it does not include eye tracking information. To create a head-gaze ray, it simply uses the front direction of the HMD that is mostly matched with the front head direction, and the starting position of the ray is the center of the viewpoint as several previous studies did [24], [25]. Thus, the Head-GazeHand is controlled by the user's head movement and orientation.

F. RESEARCH QUESTIONS

To evaluate our new combination of gaze and hand interactions whether supports the high level of usability of the hand-object interaction while adopting the benefits of the gaze interaction, we defined following research questions:

- RQ1: Is the GazeHand interface effective for distant object interaction with a high level of usability?
- RQ2: Does the GazeHand accelerate the hand interaction with the gaze operation, leading to improved task performance?

Additionally, considering that some researchers [31], [32], [33], [34] suggested using the head-gaze instead of the eye-gaze because of the unintended gaze movement and eye tracking accuracy, we have another question:

- RQ3: Does the Head-GazeHand have better performance and usability than the Eye-GazeHand?

IV. USER STUDY

We conducted a within-subject user study and compared three conditions (the modified Go-Go, Head-GazeHand, and Eye-GazeHand) with three different task types (selection, translation, and complex task). We did not directly compare our Gaze-Hand interface with previous gaze-based interfaces [9], [12], [13] because the GazeHand interface does not include any gaze-object interaction while they do, and our study focuses on evaluating a new interface supporting natural hand-distance object interaction while adopting the benefits of gaze interaction: moving fast and being able to reach every part of the task space regardless of the distance. Thus, we compared our two GazeHand interfaces with the one supporting natural hand interaction but having a different method of positioning hands to a distant object, to evaluate how well the GazeHand interface adopts the benefits of gaze interaction.

A. CONDITIONS

We developed Eye-GazeHand and Head-GazeHand conditions as described in the previous section and the modified Go-Go condition based on the Go-Go [1]. We implemented bare-hand and gaze interaction with dedicated hand and eye tracking systems provided by HTC Vive Pro Eye VR HMD.

When a user holds up their hands, all three conditions calculate the distance between the user's head and hands. If the distance is more than one-third of the average arm length [47] (Female: 72.20cm, Male: 78.65cm), the system starts operating the conditions. With the two GazeHand conditions, the system calculates the pivot point and positions the hands near the object that the user looks at. With the modified Go-Go condition, the system starts moving the virtual hand in the direction of the line from the shoulder (we manually measure the relative position of the shoulder from the user's head before starting the user study with a tracker - controller) to the hand. This forward-moving virtual hand behavior is the difference between it and the two GazeHand conditions. We adjusted the modified Go-Go condition to let the virtual hand reach the target object when the distance between the shoulder and the hands is more than two-thirds of the average arm length. This requires the same amount of arm stretch to reach the target object among the three conditions because the pivot point of the two GazeHand conditions is the same distance away from the target object. If a user holds up their real hand with the stretched arm having more than two-thirds of the average arm length, there is no virtual-hand forwarding animation, but the virtual hand is instantly positioned at the target object identified with the collision point of the ray at the direction of the line from the shoulder to the real hand (Fig. 5). If there is no collision with the ray, the condition takes the previous depth length. The modified Go-Go condition also creates a pivot point according to the position of the previously animated hand, so the other hand's interaction can start near the already positioned hand with the pivot point.

- Eye-GazeHand: The user's eye-gaze translates the virtual hands by placing them near the object that the eye-gaze points at and can grab and manipulate the objects with the direct hand gesture interaction (Fig. 4).
- Head-GazeHand: Works as same as the Eye-GazeHand except using the head-gaze to translate the virtual hands (Fig. 4).
- Modified Go-Go: supporting direct hand-object interaction like the two GazeHand interfaces but using hands themselves to control the positioning of virtual hands close to a distant object (Fig. 5).

B. HARDWARE AND SOFTWARE SETUP

The experimental system was developed with Unity game engine 2021.3.8f1. The system runs on a PC (AMD Ryzen 7 5800X 8-Core Processor 3.80GHz CPU, 32 GB RAM, and NVIDIA GeForce RTX 3060 graphics card) and the HMD was HTC Vive Pro Eye embedded with a Tobii eye tracker. To connect the HMD with the Unity engine, SteamVR Plugin 2.7.3 was used. To get eye and hand tracking information, Vive Eye & Facial Tracking SDK 1.3.3.0, Vive Hand Tracking SDK 1.0.0, and TobiiXR SDK 3.0.1 were used.

C. TASKS

We prepared three different tasks to have different difficulty levels. All tasks are distant object selection and/or manipulation (see Fig. 6). We prepared three types of tasks: selection, translation, and complex task including orientation, and they were designed according to Bowman's taxonomy of interaction in VR environment [37]: selection, translation, and orientation.

The goal of the selection task is to select ten cubes 4 meters away from participants. There are five small cubes (in 20cm width) and five large cubes (in 30cm width) that are placed with an interval of one meter. Five of them are at the one-meter height, and the others are at the two-meter height. Once the task starts, one of the cubes will change its color to yellow

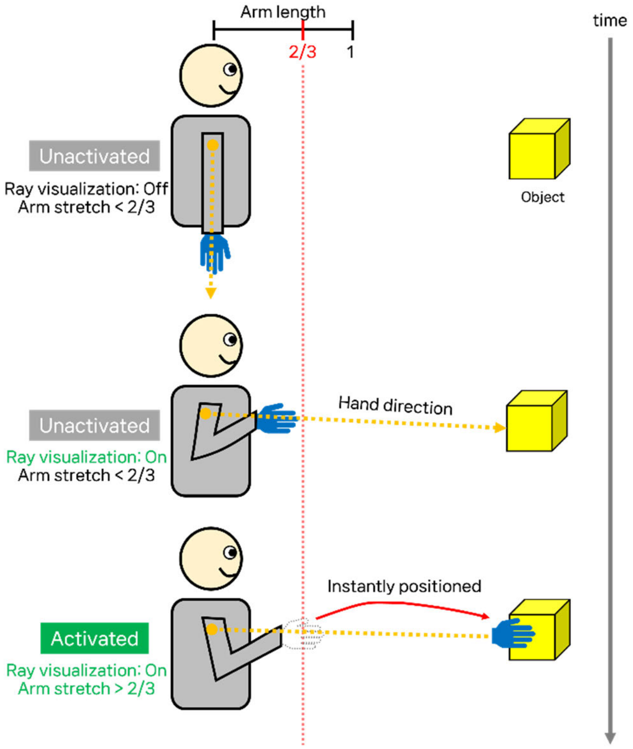

FIGURE 5. **Object selection with the modified Go-Go. If a user stretches an arm more than two-thirds of the arm length, the virtual hand instantly reaches at the target object at the direction of the line from the shoulder to the hand.**

and the participant should select it by grabbing interaction. The system counts an error if the grabbing gesture is made but does not hold the yellow cube. After selecting the cube, it returns to white, and another cube will be in yellow. The task continues until all ten cubes are properly selected and the order of the yellow highlighted cube is random.

In the translation task, there are five planes in four cardinal directions (east, west, south, and north) and at the center, and the task is to translate a yellow cube (in 30cm width) from the center white plane to one of the planes in four-directions. All planes are 2 meters by 2 meters square and there is a 2 meters interval between the center and each directional plane. The participant is 2 meters south from the south plane and 3 meters above the floor. The participant should translate the cube eight times as translation in each direction is required twice. The instruction showing the direction to move is displayed with an arrow on a blackboard and is 5 meters above the white plane. If dropping a cube or placing it at any wrong plane, the cube returns to the initial position, and it is counted as an error. The order of the directional translation is random.

The complex task includes precise translation and rotation of a virtual object, a teddy bear. There are three teddy bears on a table (1m × 2m × 0.6m) with a 60cm interval, and each teddy bear is 35cm in width, 35cm in depth, and 50cm in height. The goal of the task is to position and orientate the three teddy bears to match the half-transparent targets on a 4-meter front shelf. Two of them are rotated 90 degrees at each X and Y axis, and another one is facing toward the participant. We set thresholds as 20cm for position and 40 degrees for orientation (sum of difference at three axes) for success. There are three indicators above the shelf to indicate the success of three trials by changing its color from red to yellow.

D. PROCEDURE & DATA COLLECTION

On arrival, we explained the purpose and procedure of the study to the participants. Once they agreed with the study, they filled out a consent form and demographic questionnaire asking about gender, age, and VR experience. The participant then wore the HTC Vive Pro Eye HMD and calibrated their eye-gaze, and the experimenter measured the relative position of both shoulders to the head using the tracked controller, for the use of the modified Go-Go interface.

Before the experiment, the participants had time to practice with the interfaces of three conditions and got acquainted with the object interaction task with seven white cubes (in 20cm width) and two tables. During the practice, the experimenter verbally gave an instruction on using three interfaces and asked them to perform object selection, translation, and rotation until they felt familiar with it.

After the practice, the participant performed three tasks (in the order of selection, translation, and complex tasks) with the interface of the given condition. The order of three conditions was counter-balanced with a Balanced Latin square design. While experimenting, we collected some data to answer our RQ1 (usability) and RQ2 (task performance). After completing the tasks with each condition, the participants answered the subjective questionnaires: System Usability Scale (SUS) [48], [49], Subjective Mental Effort Questionnaire (SMEQ) [50], and NASA Task Load Index (NASA-TLX) [51], [52]. During the experiment, the experimenter collected system logs of the five data types: task completion time, grab time (sum of time taken for grabbing an object), hand movement (sum of hand movement in meters), head rotation (sum of head rotation in degrees), and error count (as described in the previous section) within the given task and condition. After experiencing all three conditions, we asked the participant to rank the three conditions as 1st (Most preferred), 2nd (Normal), and 3rd (Least preferred).

After completing tasks with a given condition, the experimenter conducted an interview asking the pros and cons of the three conditions and suggestions to improve the GazeHand interface. The overall experiment took about 70 minutes per participant. They received a gift certificate worth ten dollars as a reward.

We summarized data collection to answer our research questions in Table 1.

E. PARTICIPANTS

We initially recruited 24 participants who were local undergraduate and graduate students, and office staffs, but three of them could not complete the user study. Two of them wore thick and dark glasses which made the eye tracker fail to track.

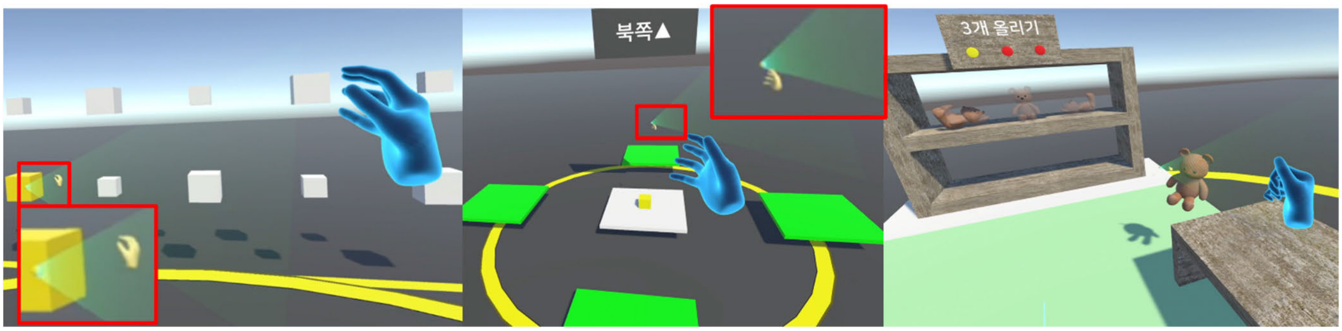

FIGURE 6. Three tasks from our user study. Left: the selection task. Center: the translation task. Right: the complex task.

TABLE 1. Data collection to answer the research questions.

Research Question	Measurements
RQ1. Is the GazeHand interface effective for distant object interaction with a high level of usability?	SUS, SMEQ, and Raw NASA-TLX
RQ2. Does the GazeHand accelerate the hand interaction with the gaze operation, leading to improved task performance?	Task Completion Time, Hand Movement, Head Rotation, Grab Time, and Error Count
RQ3. Does the Head-GazeHand have better performance and usability than the Eye-GazeHand?	All measurements above

Another participant withdrew due to poor eye tracking quality. Eventually, 21 participants completed the study and their ages ranged from 19 to 28 years with an average of 23.05 (SD = 2.96). They were 9 males and 12 females, and nine of them wore glasses. Most of them had little experience in using VR interfaces (less than once a month: n = 12; never: n = 9).

V. RESULTS

We present the results of the user study with objective data, subjective data, and user feedback. According to Shapiro-Wilk test results, our objective data (task completion time, grab time, hand movement, head rotation, and error count) are not normally distributed for each condition, and subjective questionnaire data (SUS, SMEQ, NASA-TLX, and preference) are in ordinal scales. We thus ran the Friedman test ($\alpha = .05$), and for those results showing a significant difference between the three conditions, we conducted pairwise comparisons for further investigation. Given the measures being nonparametric, and the study design in within-subject design, we used Wilcoxon signed-rank tests for pairwise comparisons while adjusting the alpha level with Bonferroni correction ($\alpha = .0167$). Given there are three pairwise comparisons, we divided the significance level by 3 (i.e., .05 / 3 = .0167) according to the Bonferroni correction.

In this section, we use the abbreviations: GH, head-GH, eye-GH, and Go-Go, for the GazeHand, Head-GazeHand, Eye-GazeHand, and modified Go-Go conditions, respectively.

A. OBJECTIVE DATA

Results from objective data are summarized in Fig. 7. Friedman tests found significant differences between the three conditions in all measurements with three task types (all p-values are less than .001 except the error count with the selection task, which is .019).

1) TASK COMPLETION TIME

In the selection task, both head and eye GHs completed the task significantly faster than the Go-Go ($Z = -4.015$, $p < .001$; $Z = -3.180$, $p = .001$, respectively). There was no significant difference between the eye and head GHs ($Z = -1.408$, $p = .159$). Similar results were shown in the translation task. The head and eye GHs completed the task significantly faster than the Go-Go ($Z = -3.806$, $p < .001$; $Z = -3.980$, $p < .001$, respectively), but there was no significant difference between the two GHs ($Z = -0.226$, $p = .821$). In the complex task, the two GHs were also faster than the Go-Go (head-GH: $Z = -3.945$, $p < .001$; eye-GH: $Z = -3.319$, $p = .001$). Interestingly, there was a significant difference between the two GHs ($Z = -2.416$, $p = .0157$) as the head-GH (M = 64.31 seconds, SD = 23.89) was faster than the eye-GH (M = 87.95 seconds, SD = 49.80).

2) HAND MOVEMENT

In the selection task, the two GHs showed less amount of hand movement than the Go-Go (head-GH: $Z = -4.015$, $p < .001$; eye-GH: $Z = -4.015$, $p < .001$). This trend was also present in both translation (head-GH: $Z = -3.980$, $p < .001$; eye-GH: $Z = -4.015$, $p < .001$) and complex tasks (head-GH: $Z = -4.015$, $p < .001$; eye-GH: $Z = -3.841$, $p < .001$). Interestingly, we found that the eye-GH (M = 4.62m, SD = 3.00) required less amount of hand movement than the head-GH (M = 7.55m, SD = 4.52) in the translation task ($Z = -2.833$, $p = .005$), but no significant difference between them on the other tasks (selection task: $Z = -1.755$, $p = .079$; complex task: $Z = -0.747$, $p = .455$).

3) HEAD ROTATION

In the selection task, both head-GH ($Z = -3.180$, $p = .001$) and eye-GH ($Z = -3.945$, $p < .001$) required less

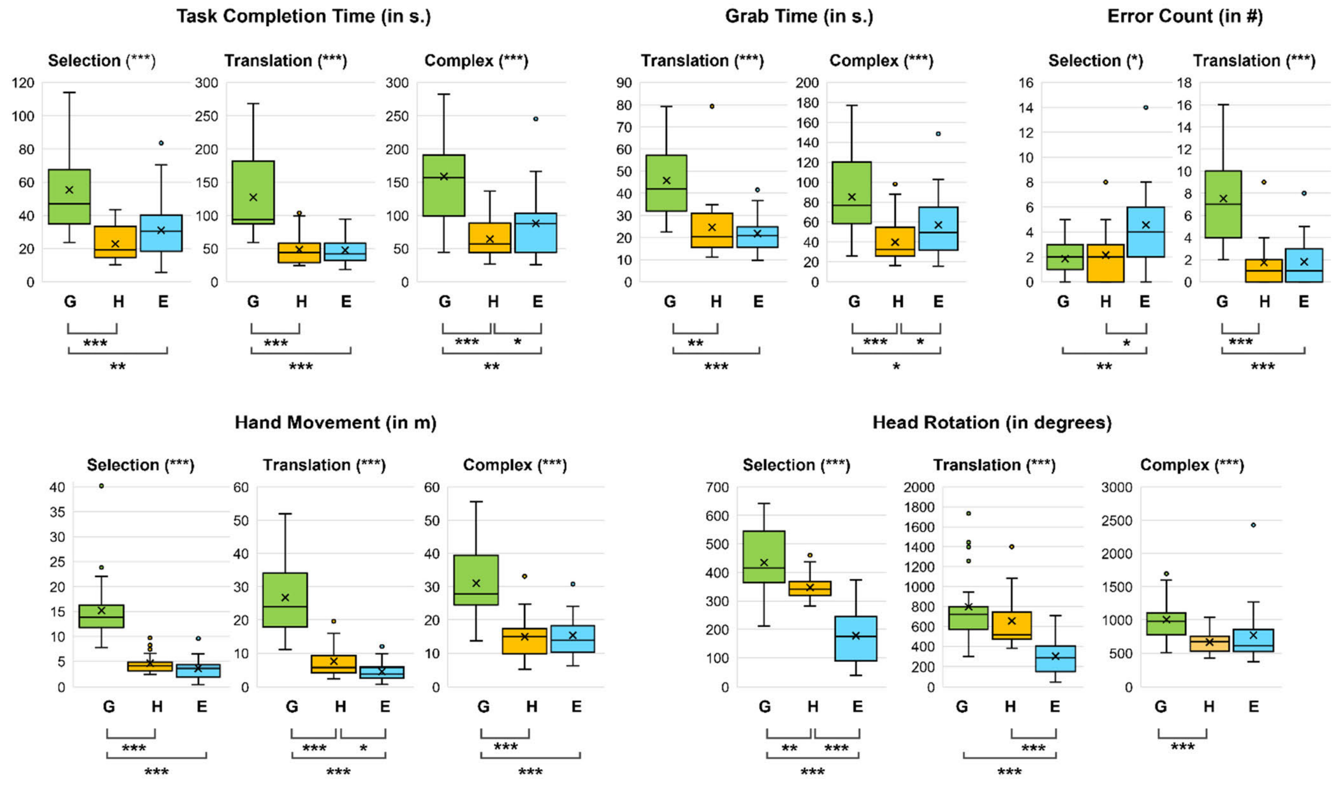

FIGURE 7. Results of objective data: task completion time, grab time, amount of the hand movement, head rotation, and error counts (G: modified Go-Go, H: Head-GazeHand, E: Eye-GazeHand, X: mean, o: outliers, * = p < .05 for Friedman test and p < .0167 for Wilcoxon signed rank test with Bonferroni Correction, ** = p < .005, and *** = p < .001).

amount of head rotation than the Go-Go. Also, the eye-GH (M = 177.64°, SD = 102.00) required less amount of the head rotation ($Z = -3.945$, $p < .001$) than the head-GH (M = 346.87°, SD = 43.27). However, the translation task showed different results. There was no difference between the Go-Go and head-GH ($Z = -1.651$, $p = .099$), but comparisons between the eye-GH and the Go-Go ($Z = -3.945$, $p < .001$) and between the eye-GH and the head-GH ($Z = -3.736$, $p < .001$) showed significant differences.

Interestingly, the results with the complex task showed a different trend compared to the one with the translation task. There was a significant difference between the Go-Go and the head-GH ($Z = -3.771$, $p < .001$). Meanwhile, no significant difference was found between eye-GH and Go-Go ($Z = -2.207$, $p = .027$) and between eye-GH and head-GH ($Z = -0.713$, $p = .476$).

4) GRAB TIME

We did not measure the time of grabbing an object with the selection task, as we compared three conditions in grab time only for translation and complex tasks. In the translation task, the participants took significantly more time to grab objects with the Go-Go than with the head-GH ($Z = -3.424$, $p = .001$) and the eye-GH ($Z = -3.841$, $p < .001$). There was no significant difference between the two GHs ($Z = -0.295$, $p = .768$). In the complex task, the participants took significantly more time to grab objects with the Go-Go than with the head-GH ($Z = -3.806$, $p < .001$) and the eye-GH ($Z = -2.694$, $p = .007$). However, the head-GH (M = 39.90 seconds, SD = 22.20) spent less grab time ($Z = -2.555$, $p = .011$) than the eye-GH (M = 57.06 seconds, SD = 31.51).

5) ERROR COUNT

We counted errors in the selection and translation tasks because all trials for the complex task were completed without errors but took more time than other tasks. Interestingly, the participants made more errors with the eye-GH than both Go-Go ($Z = -2.873$, $p = .004$) and head-GH ($Z = -2.440$, $p = .015$) in the selection task. There was no significant difference between Go-Go and head-GH ($Z = -0.498$, $p = .619$). Meanwhile, with the translation task, the participants had a greater number of errors with the Go-Go than the head-GH ($Z = -3.691$, $p < .001$) and the eye-GH ($Z = -3.774$, $p < .001$). There was no significant difference between the two GHs ($Z = -0.416$, $p = .677$).

B. SUBJECTIVE DATA

Overall questionnaire results with statistical values are described in Table 2 and 3. All questionnaire results from the Friedman test showed significance among three conditions (all *p*-values are less than .001 except the SUS usability test with the selection task, which is .006).

TABLE 2. Subjective results of the study.

Measurement	Task	Mean and standard deviation: ***M* (*SD*)**			Friedman test: $\chi^2(2)$ **(*p*)**	Wilcoxon signed-rank test: ***Z* (*p*)**		
		Go-Go	head-GH	eye-GH		Go-Go VS. head-GH	Go-Go VS. eye-GH	eye-GH VS. head-GH
SUS	Selection	72.857 (22.154)	84.643 (9.024)	73.452 (16.535)	10.243 (.006*)	-2.964 (.003**)	-0.187 (.852)	-2.854 (.004**)
	Translation	58.690 (21.471)	82.262 (11.723)	77.857 (16.494)	16.683 (<.001***)	-3.550 (<.001***)	-3.225 (<.001***)	-0.733 (.463)
	Complex	55.476 (26.500)	83.810 (11.198)	70.238 (17.228)	26.000 (<.001***)	-3.725 (<.001***)	-2.094 (.036)	-3.712 (<.001***)
Raw NASA-TLX	Selection	31.984 (19.928)	15.238 (10.779)	29.762 (18.873)	18.099 (<.001***)	-3.099 (.002**)	-0.825 (.409)	-3.528 (<.001***)
	Translation	45.000 (23.256)	19.921 (10.833)	18.810 (16.391)	15.537 (<.001***)	-3.715 (<.001***)	-3.216 (.001**)	-0.336 (.737)
	Complex	51.190 (21.208)	17.460 (13.214)	32.222 (15.806)	29.810 (<.001***)	-4.015 (<.001***)	-2.920 (.003**)	-3.859 (<.001***)
SMEQ	Selection	19.190 (16.345)	8.190 (7.393)	21.667 (15.104)	20.027 (<.001***)	-2.978 (.003**)	-0.332 (.740)	-3.848 (<.001***)
	Translation	36.238 (24.962)	13.476 (11.418)	15.143 (13.868)	13.169 (.001**)	-3.180 (.001**)	-2.839 (.005*)	-0.202 (.840)
	Complex	51.571 (30.938)	13.190 (14.236)	27.333 (16.286)	23.455 (<.001***)	-3.885 (<.001***)	-2.679 (.007*)	-3.661 (<.001***)

The table shows SUS, SMEQ, and raw NASA-TLX results of the three tasks with the three conditions (Go-Go: modified Go-Go condition, head-GH: the Head-GazeHand condition, eye-GH: the Eye-GazeHand condition). SUS and raw NASA-TLX range from 0 to 100. Significant results are in gray with the annotation (* = p < .05 for the Friedman test and p < .0167 for the Wilcoxon signed-rank test with Bonferroni correction, ** = p < .005, and *** = p < .001).

1) SUS (USABILITY)

In the selection task, the head-GH showed the highest usability level, which is significantly higher than the usability levels of the Go-Go ($Z = -2.964$, $p = .003$) and the eye-GH ($Z = -2.854$, $p = .004$). No significant difference between the Go-Go and eye-GH ($Z = -0.187$, $p = .852$) was shown. In the translation task, there was no significant difference between the two GHs ($Z = -0.733$, $p = .463$), but the Go-Go showed a significantly lower level of usability than both head-GH ($Z = -3.550$, $p < .001$) and eye-GH ($Z = -3.225$, $p < .001$). Surprisingly, the results of the complex task showed a similar trend to the results of the selection task rather than those of the translation task. Thus, the head-GH showed a higher usability level than the Go-Go ($Z = -3.725$, $p < .001$) and eye-GH ($Z = -3.712$, $p < .001$). There was no significant difference between the Go-Go and eye-GH ($Z = -2.094$, $p = .036$).

2) NASA-TLX (TASK LOAD) AND SMEQ (MENTAL EFFORT)

The results of the task load and the required mental effort showed the same trend, and it was similar to the results of the usability. In the selection task, using the head-GH required significantly lower task load and mental effort than when using the Go-Go (task load: $Z = -3.099$, $p = .002$; mental effort: $Z = -2.978$, $p = .003$) and the eye-GH (task load: $Z = -3.528$, $p < .001$; mental effort: $Z = -3.848$, $p < .001$). There was no significant difference between the Go-Go and the eye-GH (task load: $Z = -0.825$, $p = .409$; mental effort: $Z = -0.332$, $p = .740$). In the translation task, however, the Go-Go required a significantly higher level of task load and mental effort compared to not only the head-GH (task load: $Z = -3.715$, $p < .001$; mental effort: $Z = -3.180$, $p = .001$) but also the eye-GH (task load: $Z = -3.216$, $p = .001$; mental effort: $Z = -2.839$, $p = .005$). There was no significant difference between the two GHs (task load: $Z = -0.336$, $p = .737$; mental effort: $Z = -0.202$, $p = .840$). In the complex task, using the Go-Go needed a higher level of task load and mental effort than using both head-GH (task load: $Z = -4.015$, $p < .001$; mental effort: $Z = -3.885$, $p < .001$) and eye-GH (task load: $Z = -2.920$, $p = .003$; mental effort: $Z = -2.679$, $p = .007$). Interestingly, the participants felt that they spent more task load and mental effort when using the eye-GH than when using the head-GH (task load: $Z = -3.859$, $p < .001$; mental effort: $Z = -3.661$, $p < .001$).

3) PREFERENCE

The participants ranked the conditions according to their preferences. The results of the user preference (see Table 3) showed the same trend with the one of the usability, which implies that participants preferred conditions that have high usability.

In the selection task, participants chose the head-GH as the best-preferred condition than the eye-GH ($Z = -2.901$, $p = .004$) and the Go-Go ($Z = -3.334$, $p = .001$). No significant difference was found between the eye-GH and the Go-Go ($Z = -0.037$, $p = .971$). In the translation task, the participants preferred both GH conditions compared to the Go-Go condition (eye-GH: $Z = -2.776$, $p = .006$; head-GH: $Z = -3.620$, $p < .001$). In the complex task, the head-GH was significantly preferred more than the Go-Go

TABLE 3. Preference results of the study.

Task	Preference				Friedman test		Wilcoxon signed-rank test			
	Rank	G	H	E				G - H	G - E	H - E
Selection	1st	2	15	4	$\chi^2(2)$	14.000	*Z*	-3.334	-0.037	-2.901
	2nd	10	5	6	*p*	.001**	*p*	.001**	.971	.004**
	3rd	9	1	11						
Translation	1st	1	9	11	$\chi^2(2)$	14.100	*Z*	-3.620	-2.776	-0.423
	2nd	5	11	5	*p*	.001**	*p*	<.001***	.006*	.673
	3rd	15	1	5						
Complex	1st	0	18	3	$\chi^2(2)$	26.570	*Z*	-4.144	-2.201	-3.334
	2nd	6	3	12	*p*	<.001***	*p*	<.001***	.028	.001**
	3rd	15	0	6						

The table shows user preference among the three conditions in the three task types (G: the modified Go-Go, H: the Head-GazeHand, E: the Eye-GazeHand). The numbers in the preference columns are the number of participants choosing the condition. Significant results are in gray with annotation (* = $p < .05$ for the Friedman test and $p < .0167$ for the Wilcoxon signed-rank test with Bonferroni correction, ** = $p < .005$, and *** = $p < .001$).

($Z = -4.144$, $p < .001$) and the eye-GH ($Z = -3.334$, $p = .001$). There was no significant difference between Go-Go and eye-GH ($Z = -2.201$, $p = .028$).

C. INTERVIEW

Three participants (Participants 5 – P5 afterward, P10, and P19) reported arm fatigue when holding the hand up with the Go-Go. Other participants commented on the difficulty of controlling virtual hands. P21 stated, *"When I released an object, my hand trembled and the position of the object (that I released) was unintentionally changed."* This issue of releasing an object with the Go-Go may result in a large number of errors in the translation task while it does not affect the error count in the selection task because the selection task does not include releasing object activity.

Some participants compared the two GHs. Eight participants stated that eye-GH required less physical motion than head-GH. P5 said, *"The head-gaze condition required head movement, while it can be done simple and fast eye movement with the eye-gaze condition"*. Many participants, however, felt some inconvenience in using the eye-GH with its drawbacks. Three participants felt that eye-GH distracted looking-around activity because the grabbed object always followed their gaze. P19 stated, *"The eye-gaze point distracted me while glancing at another place*."

Four participants also mentioned instability of eye-GH (*"the hands sometimes unintentionally moved with the eye condition"*). As a solution to the issues, three participants (P2, P15, and P20) thought that users may need practice to use eye-GH and need time to be familiar with it. Additionally, P4 reported an optimal usage of the two GHs ("*The eye one is fast while doing simple task, and the head one is nice for the task which requires preciseness.*").

Additionally, P12 suggested a function for helping limited hand rotation capacity (*"keep rotating object when a user keeps the pose of completely rotated hand after the rotating object interaction, so reduce the number of same rotating interaction until completing it"*).

D. OBSERVATION

There were five interesting findings from our observation. First, the participants tried to hold their virtual hand in the line of the gaze ray, trying to establish an easy combination of the gaze and hand interaction. Second, the participants performed the grabbing-object hand interaction almost together with the gaze fixation at the final position in the selection task for the fast task completion time. Third, they translated objects more with eye movement rather than with head movement when playing the easy selection and translation tasks, but it was not obvious with the complex tasks. Thus, as revealed in head rotation data collection, there was less difference in the amount of head rotation between the eye and head GHs for the complex task than for the selection and translation tasks. Fourth, there was some small unintended hand movement when releasing an object and it affected the use of the Go-Go because the small unintended hand movement resulted in a big change of the distant virtual hand and grabbed object positions.

VI. DISCUSSION

A. BENEFITS OF GazeHand

The two GHs generally had a better usability level with lower levels of task load and mental effort compared to the Go-Go (see Table 2). Therefore, we can answer RQ1 (Is the GazeHand interface effective for distant object interaction with a high level of usability?) as 'yes'. This may be because the two GHs kept the high-level usability of the direct hand gesture interaction with easy gaze interaction translating the hands to a distant object, while the Go-Go did not as users' hand gesture was not only for hand-object interaction but also for moving virtual hands to a distant object, suffering with the issues described in the previous section.

Interestingly, the benefits of the eye-GH compared to the Go-Go are revealed only with the translation and complex tasks but not with the selection task in the questionnaire results. Regarding this result, we contemplate the number of errors with the conditions (see Fig. 7 – error count).

The Go-Go mostly had an issue when releasing an object, so it had fewer errors with the selection task that did not include releasing interaction. The participants tried quick and instant motion for the selection task, and this behavior was not in harmony with the eye-GH that requires time for eye tracking results to be stable, otherwise making errors.

Both GHs required less task completion time and less amount of hand movement and grab time compared to when using the Go-Go (see Table 2), because of the help of the fast gaze operation for moving hands near the target object. Therefore, the answer for the RQ2 (Does the GazeHand accelerate the hand interaction with the gaze operation, leading to improved task performance?) can be agreeable.

B. Head-GazeHand VS Eye-GazeHand

There was a noticeable difference between the two GHs while both kept the benefit of gaze and hand operations as described in the previous section. The main differences were stability and using eye movement or not. The results showed that the eye-GH had the issue of stability because of the gaze tracking accuracy while the head-GH did not. Hence, it led to lower usability and more errors compared to the head-GH in the selection and complex tasks (see Table 2 and Fig. 7). Interestingly, this trend was not obvious with the translation task because the translation task did not require precise eye control when placing a cube on the large target plane (2m × 2m). In short, the eye-GH had lower usability and more error for the tasks requiring precise operation because of the issue of gaze tracking accuracy, but it might be less affected with the task requiring less precise operation (i.e., translation task in our user study).

One another interesting point is that the benefit of the quick gaze movement was revealed with the task requiring less precise operation (i.e., translation task) because of the reduced effect from the issue of eye tracking accuracy. In the translation task, the user view at a proper angle could cover all task space without moving the head, and simple eye movement was enough to complete the cube translation task with the large target plane (see the center figure of Fig. 6). This reduced the head rotation with the eye-GH while the head-GH required many head rotation movement to bring the cube to the large plane.

Therefore, the answer for the RQ3 (Does the Head-GazeHand have better performance and usability than the Eye-GazeHand?) is not verified, but we report that the head-GH is better for the complex task requiring precise interaction and the eye-GH is better for simple translation if the task does not require heavy head and hand movement.

C. Interaction with Relative and Absolute Position

The 'Gaze + Pinch' interface [12] exploits relative position in interacting with objects. Thus, a user can grab an object and start hand translation at any hand position with a pinch gesture and translate an object according to the relative position of the current hand to the hand position when pinching pose. This may help users do convenient hand positioning but is less realistic. One of the big issues in using relative position for hand-object interaction occurs when rotating an object. The Gaze + Pinch uses a relative position of two hands for object rotation by mapping the rotation of the line between two hands to the object. This may require a high level of physical effort when many rotations are required. Our GHs facilitate the absolute position, so a participant should position the hand at the proper location for interacting with an object. Thus, a user may have inconveniences with hand pose compared to the relative-position-based hand-object interaction. However, it is more realistic, and a user can perform rotation with one hand rather than using two hands.

D. Limitations

Limited depth control is a major limitation of our GH. Our GH, however, still supports depth control within arm length and supports positioning objects with a depth of gaze point on another object. Since the majority of objects in the real world and virtual world are on another object (e.g., a vase on a desk), so limitation of the depth control may not be significant in the VR environment imitating the real world. Besides, the depth of a grabbed object is automatically calculated when placed on another object, so it reduces the user's effort to control depth.

We also note that the quality of eye tracking [28] could affect user experiences as shown in our results: lower usability with the eye-GH and head-GH. Since eye movement is faster than head movement, the results comparing eye-GH and head-GH would be different with ours if the stable eye-tracking system is supported.

Our user study did not investigate close object interaction, nor other general dependent factors in VR such as presence and embodiment. Additionally, the participants in our study cannot represent diverse VR users. They were under 30 years old and most of them did not have much VR experience. However, we believe that our user study results are generally accepted and can be applied to other VR applications because its design is as simple as using gaze to move hands to a distance object and hands to interact with objects.

VII. Conclusion and Future Work

In this paper, we introduced a novel interface, named GazeHand which combines the gaze and hand interactions for distant object selection and manipulation. The GazeHand interface enables a user to position the virtual hands near the distant object that he/she is looking at, so the direct hand gesture interaction can be used for the selection and manipulation of objects even at a distance. The GazeHand interface can use either eye- or head-gaze, so we implemented both variants of the GazeHands interface and compared them to the modified Go-Go interface in a user study. Since both GazeHand interfaces adopt the fast gaze movement, they completed the selection and manipulation tasks faster than the modified Go-Go interface with a higher level of usability and a lower level of mental and task load. The participants preferred the Head-GazeHand most because of its benefits of using a fast and stable head-gaze while the advantage of the Eye-GazeHand interface was prominent when the task required many gaze movements.

Although this study explored the benefits of combining gaze and hand, we did not compare our GazeHand with interfaces which use gaze-based selection [9], [12], [13]. To further explore our interfaces, we have a plan to conduct another comparison study with other gaze- and hand-based interfaces in the future. We will also extend our study for multi-user remote collaboration; we can explore how the visibility of virtual hands influences the communication between the collaborators under a large task space (i.e., a factory with a digital twin system), which requires distant object interaction, by comparing our interfaces with indirect object manipulation interfaces.

REFERENCES

[1] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa, "The go-go interaction technique: Non-linear mapping for direct manipulation in VR," in *Proc. 9th Annu. ACM Symp. User Interface Softw. Technol. (UIST)*, Seattle, WA, USA, 1996, pp. 79–80, doi: 10.1145/237091.237102.

[2] D. A. Bowman and L. F. Hodges, "An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments," in *Proc. Symp. Interact. 3D Graph. (SID)*, Providence, RI, USA, 1997, pp. 35–38, doi: 10.1145/253284.253301.

[3] P. Mohan, W. B. Goh, C.-W. Fu, and S.-K. Yeung, "DualGaze: Addressing the midas touch problem in gaze mediated VR interaction," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, Munich, Germany, Oct. 2018, pp. 79–84, doi: 10.1109/ismar-adjunct.2018.00039.

[4] S. Kim, A. Jing, H. Park, G. A. Lee, W. Huang, and M. Billinghurst, "Hand-in-air (HiA) and hand-on-target (HoT) style gesture cues for mixed reality collaboration," *IEEE Access*, vol. 8, pp. 224145–224161, 2020, doi: 10.1109/ACCESS.2020.3043783.

[5] A. T. Duchowski, "Gaze-based interaction: A 30 year retrospective," *Comput. Graph.*, vol. 73, pp. 59–69, Jun. 2018, doi: 10.1016/j.cag.2018.04.002.

[6] C. Liu, A. Plopski, and J. Orlosky, "OrthoGaze: Gaze-based three-dimensional object manipulation using orthogonal planes," *Comput. Graph.*, vol. 89, pp. 1–10, Jun. 2020, doi: 10.1016/j.cag.2020.04.005.

[7] A. Steed, "Towards a general model for selection in virtual environments," in *Proc. 3D User Interfaces (3DUI)*, Alexandria, VA, USA, 2006, pp. 103–110, doi: 10.1109/vr.2006.134.

[8] J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard, "A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display," in *Proc. Workshop Commun. Gaze Interact.*, Warsaw, Poland, Jun. 2018, pp. 1–5, doi: 10.1145/3206343.3206344.

[9] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves, "Gaze-supported 3D object manipulation in virtual reality," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Yokohama, Japan, May 2021, pp. 1–13, doi: 10.1145/3411764.3445343.

[10] Y. Y. Qian and R. J. Teather, "The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality," in *Proc. 5th Symp. Spatial User Interact.*, Brighton, U.K., Oct. 2017, pp. 91–98, doi: 10.1145/3131277.3132182.

[11] I. Chatterjee, R. Xiao, and C. Harrison, "Gaze+gesture: Expressive, precise and targeted free-space interactions," in *Proc. ACM Int. Conf. Multimodal Interact.*, Seattle, WA, USA, Nov. 2015, pp. 131–138, doi: 10.1145/2818346.2820752.

[12] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen, "Gaze + pinch interaction in virtual reality," in *Proc. 5th Symp. Spatial User Interact.*, Brighton, U.K., Oct. 2017, pp. 99–108, doi: 10.1145/3131277.3132180.

[13] K. Ryu, J.-J. Lee, and J.-M. Park, "GG interaction: A gaze–grasp pose interaction for 3D virtual object selection," *J. Multimodal User Interfaces*, vol. 13, no. 4, pp. 383–393, Dec. 2019, doi: 10.1007/s12193-019-00305-y.

[14] V. Tanriverdi and R. J. K. Jacob, "Interacting with eye movements in virtual environments," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, The Hague, The Netherlands, Apr. 2000, pp. 265–272, doi: 10.1145/332040.332443.

[15] L. E. Sibert and R. J. K. Jacob, "Evaluation of eye gaze interaction," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, The Hague, The Netherlands, Apr. 2000, pp. 281–288, doi: 10.1145/332040.332445.

[16] C. Ware and H. H. Mikaelian, "An evaluation of an eye tracker as a device for computer input2," *ACM SIGCHI Bull.*, vol. 17, no. SI, pp. 183–188, May 1986, doi: 10.1145/30851.275627.

[17] K. Pfeuffer, J. Alexander, M. K. Chong, and H. Gellersen, "Gaze-touch: Combining gaze with multi-touch for interaction on the same surface," in *Proc. UIST*, Honolulu, HI, USA, 2014, pp. 509–518, doi: 10.1145/2642918.2647397.

[18] H. Kuzuoka, G. Ishimoda, Y. Nishimura, R. Suzuki, and K. Kondo, "Can the gesturecam be a surrogate?" in *Proc. ECSCW*, Stockholm, Sweden, 1995, pp. 181–196, doi: 10.1007/978-94-011-0349-7_12.

[19] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, and M. Billinghurst, "Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Scotland, U.K., May 2019, pp. 1–13, doi: 10.1145/3290605.3300403.

[20] K. Higuch, R. Yonetani, and Y. Sato, "Can eye help you: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, San Jose, CA, USA, May 2016, pp. 5180–5190, doi: 10.1145/2858036.2858438.

[21] R. J. K. Jacob, "What you look at is what you get: Eye movement-based interaction techniques," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. Empowering People (CHI)*, Seattle, WA, USA, 1990, pp. 11–18, doi: 10.1145/97243.97246.

[22] D. A. Bowman, D. B. Johnson, and L. F. Hodges, "Testbed evaluation of virtual environment interaction techniques," *Presence*, vol. 10, no. 1, pp. 75–95, Feb. 2001, doi: 10.1162/105474601750182333.

[23] F. Lu, S. Davari, and D. Bowman, "Exploration of techniques for rapid activation of glanceable information in head-worn augmented reality," in *Proc. SUI*, USA, 2021, pp. 1–11, doi: 10.1145/3485279.3485286.

[24] X. Lu, D. Yu, H.-N. Liang, W. Xu, Y. Chen, X. Li, and K. Hasan, "Exploration of hands-free text entry techniques for virtual reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Porto de Galinhas, Brazil, Nov. 2020, pp. 344–349, doi: 10.1109/ISMAR50242.2020.00061.

[25] D. D. Salvucci and J. R. Anderson, "Intelligent gaze-added interfaces," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, The Hague, The Netherlands, Apr. 2000, pp. 273–280, doi: 10.1145/332040.332444.

[26] A. L. Simeone, A. Bulling, J. Alexander, and H. Gellersen, "Three-point interaction: Combining bi-manual direct touch with gaze," in *Proc. Int. Work. Conf. Adv. Vis. Interfaces*, Bari, Italy, Jun. 2016, pp. 168–175, doi: 10.1145/2909132.2909251.

[27] J. Turner, J. Alexander, A. Bulling, and H. Gellersen, "Gaze+RST: Integrating gaze and multitouch for remote rotate-scale-translate tasks," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Seoul, South Korea, Apr. 2015, pp. 4179–4188, doi: 10.1145/2702123.2702355.

[28] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris, "Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Denver, CO, USA, May 2017, pp. 1118–1130, doi: 10.1145/3025453.3025599.

[29] N. Murray and D. Roberts, "Comparison of head gaze and head and eye gaze within an immersive environment," in *Proc. 10th IEEE Int. Symp. Distrib. Simulation Real-Time Appl.*, Malaga, Spain, 2006, pp. 70–76, doi: 10.1109/ds-rt.2006.13.

[30] O. Špakov, H. Istance, K.-J. Räihä, T. Viitanen, and H. Siirtola, "Eye gaze and head gaze in collaborative games," in *Proc. 11th ACM Symp. Eye Tracking Res. Appl.*, Jun. 2019, pp. 1–9, doi: 10.1145/3317959.3321489.

[31] R. Atienza, R. Blonna, M. I. Saludares, J. Casimiro, and V. Fuentes, "Interaction techniques using head gaze for virtual reality," in *Proc. IEEE Region Symp. (TENSYMP)*, Bali, Indonesia, May 2016, pp. 110–114, doi: 10.1109/TENCONSpring.2016.7519387.

[32] R. C. Zeleznik, A. S. Forsberg, and J. P. Schulze, "Look-that-there: Exploiting gaze in virtual reality interactions," Brown Univ., Providence, RI, USA, Tech. Rep. CS-05, 2005, pp. 1–7.

[33] S. Jalaliniya, D. Mardanbeigi, T. Pederson, and D. W. Hansen, "Head and eye movement as pointing modalities for eyewear computers," in *Proc. 11th Int. Conf. Wearable Implant. Body Sensor Netw. Workshops*, Zurich, Switzerland, Jun. 2014, pp. 50–53, doi: 10.1109/BSN.Workshops.2014.14.

[34] N. Pathmanathan, M. Becher, N. Rodrigues, G. Reina, T. Ertl, D. Weiskopf, and M. Sedlmair, "Eye vs. head: Comparing gaze methods for interaction in augmented reality," in *Proc. ETRA*, Stuttgart, Germany, 2020, pp. 1–5, doi: 10.1145/3379156.3391829.

[35] J. Blattgerste, P. Renner, and T. Pfeiffer, "Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views," in *Proc. Workshop Commun. Gaze Interact.*, Warsaw, Poland, Jun. 2018, pp. 1–9, doi: 10.1145/3206343.3206349.

[36] L. Sidenmark and H. Gellersen, "Eye&head: Synergetic eye and head movement for gaze pointing and selection," in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol.*, New Orleans, LA, USA, Oct. 2019, pp. 1161–1174, doi: 10.1145/3332165.3347921.

[37] Y. Wei, R. Shi, D. Yu, Y. Wang, Y. Li, L. Yu, and H.-N. Liang, "Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Hamburg, Germany, Apr. 2023, pp. 1–14, doi: 10.1145/3544548.3581042.

[38] D. A. Bowman, E. Kruijff, J. J. LaViola Jr., and I. P. Poupyrev, *3D User Interfaces: Theory and Practice*. Boston, MA, USA: Addison-Wesley, 2017, pp. 251–316.

[39] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst, "Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques," *Comput. Graph. Forum*, vol. 17, no. 3, pp. 41–52, Aug. 1998, doi: 10.1111/1467-8659.00252.

[40] K. Pfeuffer, J. Alexander, M. K. Chong, Y. Zhang, and H. Gellersen, "Gaze-shifting: Direct-indirect input with pen and touch modulated by gaze," in *Proc. 28th Annu. ACM Symp. User Interface Softw. Technol.*, Charlotte, NC, USA, Nov. 2015, pp. 373–383, doi: 10.1145/2807442.2807460.

[41] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. S. Grønbæk, K. Pfeuffer, and H. Gellersen, "A Fitts' law study of gaze-hand alignment for selection in 3D user interfaces," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Hamburg, Germany, Apr. 2023, pp. 1–15, doi: 10.1145/3544548.3581423.

[42] K. Reiter, K. Pfeuffer, A. Esteves, T. Mittermeier, and F. Alt, "Look & turn: One-handed and expressive menu interaction by gaze and arm turns in VR," in *Proc. Symp. Eye Tracking Res. Appl.*, Seattle, WA, USA, Jun. 2022, pp. 1–7, doi: 10.1145/3517031.3529233.

[43] K. Pfeuffer, L. Mecke, S. D. Rodriguez, M. Hassib, H. Maier, and F. Alt, "Empirical evaluation of gaze-enhanced menus in virtual reality," in *Proc. 26th ACM Symp. Virtual Reality Softw. Technol.*, Canada, Nov. 2020, pp. 1–11, doi: 10.1145/3385956.3418962.

[44] M. N. Lystbæk, P. Rosenberg, K. Pfeuffer, J. E. Grønbæk, and H. Gellersen, "Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, pp. 1–18, May 2022, doi: 10.1145/3530886.

[45] M. N. Lystbæk, K. Pfeuffer, J. E. S. Grønbæk, and H. Gellersen, "Exploring gaze for assisting freehand selection-based text entry in AR," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, pp. 1–16, May 2022, doi: 10.1145/3530882.

[46] R. Shi, Y. Wei, X. Qin, P. Hui, and H.-N. Liang, "Exploring gaze-assisted and hand-based region selection in augmented reality," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, pp. 1–19, May 2023, doi: 10.1145/3591129.

[47] C. C. Gordon, C. L. Blackwell, B. Bradtmiller, J. L. Parham, P. Barrientos, S. P. Paquette, B. D. Corner, J. M. Carson, J. C. Venezia, B. M. Rockwell, M. Mucher, and S. Kristensen, "2012 Anthropometric survey of U.S. army personnel: Methods and summary statistics," U.S. Army Natic Soldier Res., Develop. Eng. Center, Natick, MA, USA, Tech. Rep., 2014, p. 240. [Online]. Available: https://apps.dtic.mil/sti/citations/ADA611869

[48] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability Evaluation in Industry*, 1st ed. Boca Raton, FL, USA: CRC Press, 1996, pp. 189–194, doi: 10.1201/9781498710411.

[49] M. Gao, P. Kortum, and F. L. Oswald, "Multi-language toolkit for the system usability scale," *Int. J. Hum.-Comput. Interact.*, vol. 36, no. 20, pp. 1883–1901, Aug. 2020, doi: 10.1080/10447318.2020.1801173.

[50] R. H. Zijlstra, "Efficiency in work behaviour: A design approach for modern tools," Ph.D. dissertation, Dept. Ind. Des. Eng., Delft Univ. Tech., Delft, The Netherlands, 1993.

[51] G. Sandra Hart and E. LowellStaveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, Apr. 1988, doi: 10.1016/S0166-4115(08)62386-9.

[52] S. G. Hart, "Nasa-task load index (NASA-TLX); 20 years later," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006, doi: 10.1177/154193120605000909.

JAEJOON JEONG (Student Member, IEEE) is currently pursuing the B.S. and M.S. degrees in computer science and software engineering with Chonnam National University, Gwangju, Republic of Korea. His research interest includes multimodal object interaction methods in virtual reality and augmented reality environments.

SOO-HYUNG KIM received the B.S. degree in computer engineering from Seoul National University, in 1986, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology, in 1988 and 1993, respectively. Since 1997, he has been a Professor with the Department of Artificial Intelligence Convergence, Chonnam National University, South Korea. His research interests include pattern recognition, document image processing, medical image processing, and ubiquitous computing.

HYUNG-JEONG YANG received the B.S., M.S., and Ph.D. degrees from Chonbuk National University, South Korea. She is currently a Professor with the Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju, Republic of Korea. Her main research interests include multimedia data mining, medical data analysis, social network service data mining, and video data understanding.

GUN A. LEE received the Ph.D. degree in computer science and engineering from POSTECH, in 2009, with a focus on immersive authoring methods for creating VR and AR content, developing the first system in the world that allowed people to create AR applications within the AR environment. From 2005 to 2011, he was a Senior Researcher with ETRI, where he developed VR and AR systems for industrial applications. He is currently a Senior Research Scientist with the Empathic Computing Laboratory, University of South Australia, researching interaction and visualization methods for mobile and wearable augmented reality (AR) and virtual reality (VR) systems. He has produced more than 60 publications in the areas, such as augmented reality and virtual reality.

SEUNGWON KIM received the B.S. and M.S. degrees from the University of Tasmania, in 2008 and 2010, respectively, and the Ph.D. degree from the HIT Laboratory NZ, New Zealand, in 2016, under the supervision of Prof. M. Billinghurst. During the Ph.D. study, he developed a remote collaboration system supporting stabilized sketch and pointer cues. He is currently a Professor with the Department of Artificial Intelligence Convergence, Chonnam National University, South Korea. His research interests include remote collaboration using augmented virtual communication cues and sharing experiences between distance users.