ARTICLE TEMPLATE

HandBrush for Efficient Object Grouping in Virtual Environment with Bare-Hand

Sichun Huang[a], Jian wu[a], Runze Fan[a], Sio Kei Im [c] and Lili Wang[a ,b]

[a]State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China;
[b]Peng Cheng Laboratory, Shengzhen, China;
[c]Faculty of Applied Sciences, Macao Polytechnic University, Macau, China.

ARTICLE HISTORY
Compiled October 29, 2024

ABSTRACT
Object grouping task is an important research direction for fast manipulation of a large number of objects. It can help users to improve the efficiency of multi-object manipulation. However, the current research on this aspect is still immature. For this task, in this paper, based on the brush metaphor, we propose a method for grouping objects based on bare hands in virtual reality scenes. We design a number of interactions to facilitate the user's grouping of objects in the three-dimensional virtual space. The results of the empirical study show that our method has better performance in accomplishing the object grouping task compared to the Ray method, Screen method and Cone method. Object grouping in virtual reality could encompass two subtasks: group generation and group modification. The emphasis of these tasks varies, with the former focusing on creating groups from ungrouped objects and the latter focusing on modifying group members once they are generated. In the group generation task, we have faster speed, higher operation success rate, and shorter hand travel distance. In the group modification task, we also have faster speed and shorter hand travel distance.

KEYWORDS
Human-Computer Interaction; Virtual reality; Object grouping; Bare-hand;

1. Introduction

The emergence of virtual reality (VR) technology has greatly enriched our interaction with the digital world, revolutionizing various fields by providing immersive experiences and intuitive user interfaces. With the widespread application of VR in industries such as entertainment, education, design, and even healthcare, there is a growing demand for precise manipulation and effective management of objects in virtual environments, such as room decoration and design, architectural design, etc. in VR environments. For example, in 3D scene layout, grouping objects can help users achieve effective management of multiple objects. Users can attach the same attributes to multiple objects at the same time. Users can also perform operations on multiple objects

Lili Wang is corresponding author: wanglily@buaa.edu.cn, ZY2306429@buaa.edu.cn

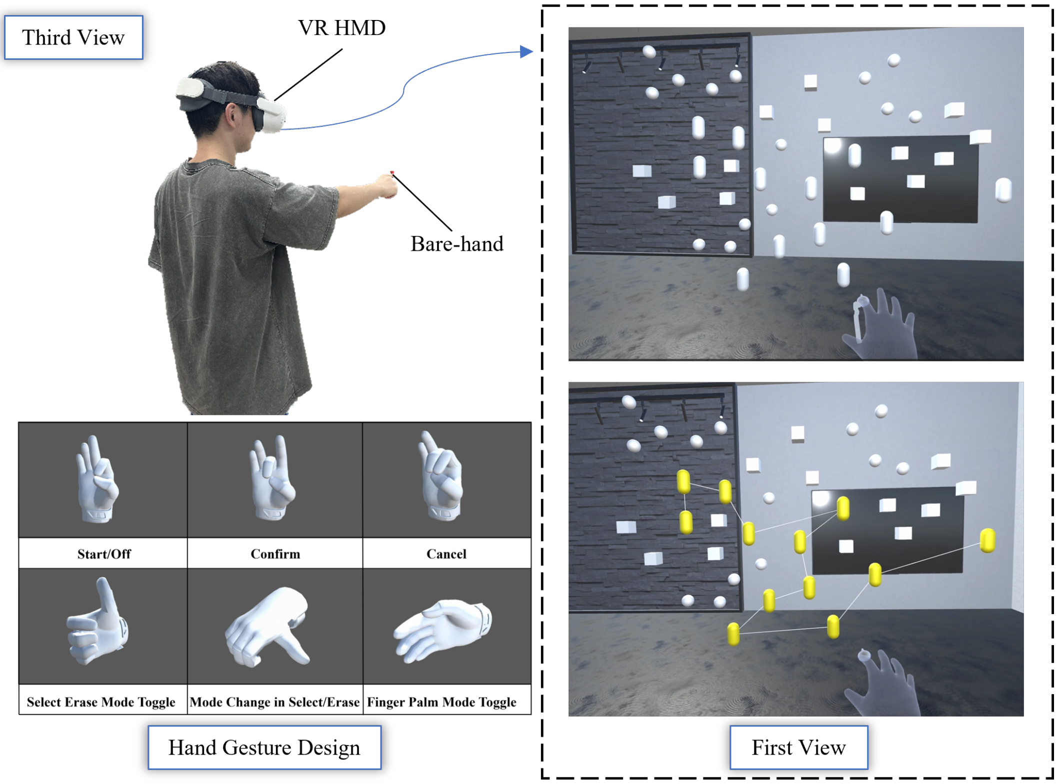

Figure 1. A user performs object grouping tasks using HandBrush. The two images on the left show the third-person perspective view and our gesture design respectively. The right two images show the first-person perspective view in VR. The user is grouping objects in the scene using HandBrush's Line mode.

at the same time, such as moving, rotating, etc. This allows the user to complete the scene layout more efficiently, avoiding tedious and identical operations. In such a context, users need a method that enables convenient selection, grouping, modification, and organization of objects.

Currently, there is limited research on group-based operations in VR. Shi, Zhang, Stuerzlinger, & Liang (2022) devised four interaction techniques for performing a three-degree-of-freedom group-based object alignment task. Oh, Stuerzlinger, & Dadgari (2006) presented a system for manipulating object groups through a gravity hierarchy. However, this approach method limits the freedom of object grouping and affects the user experience. Object grouping in VR is not a well-studied task and should encompass two subtasks: group generation and group modification. The emphasis of these tasks varies, with the former focusing on creating groups from ungrouped objects and the latter focusing on modifying group members once they are generated.

As hand-tracking technology advances, bare-hand control becomes more accurate and cost-effective. Some of the most advanced VR headsets (e.g., Meta Quest 3 and PICO 4) are equipped with inside-out cameras that enable bare-handed aerial input without the need for additional hand-tracking devices. Users can use real-world gestures to interact with objects in the virtual environment, eliminating the burden of carrying and charging controllers while providing a more fluid and immersive user experience.

In this paper, we present HandBrush, a method for bare-hand object grouping in a virtual environment. First, we introduce the design rationales of HandBrush. Then, we introduce the various modes and features that HandBrush has. Finally, we designed two user studies to evaluate the performance of HandBrush under two subtasks: group generation and group modification. As shown in Figure 1, the user is wearing a VR device to group objects in VR. The objects in the scene are not grouped at first, and the user is using Line mode of HandBrush to group them. After grouping, the objects will be strung together with lines. We compare our method with Ray method, Screen method and Cone method (Shi, Zhang, Yue, Yu, & Liang, 2023). Compared to the comparison methods, our HandBrush shows significant improvements in task completion time, operation success rate, and fatigue reduction, demonstrating the potential of HandBrush to improve the efficiency of object manipulation and user satisfaction in VR environments. In addition, we will explore the potential applications and future directions of HandBrush in VR interaction design. To see more examples of HandBrush, please visit: `https://www.youtube.com/watch?v=DHdkH52rnzg`.

In summary, the main contributions of our work include:

- We present HandBrush, a bare-hand object grouping method which incorporates three modes (Line mode, Block mode and Erase mode) and two features (Split feature and Connect feature) for fast and efficient group selection and group modification by analogizing the hand to a paintbrush.
- We conducted two empirical user studies to evaluate the performance of our HandBrush.

2. Related Work

In this section, we review the current state of research on multiple objects selection, object grouping and bare hand based VR interaction.

2.1. Multiple Objects Selection

Multiple objects selection tasks focus on how to quickly select multiple objects in a scene. Traditional approaches to multi-object selection in 3D virtual worlds mainly include techniques based on mouse drag-and-drop box selection, rotating viewpoint selection, etc. These methods usually rely on interactive tools in the user interface (UI), such as selection boxes or lasso tools. For example, ”cursor selection” methods (Li, Sarcar, Kim, Tu, & Ren, 2022; Long, Li, Yu, & Gu, 2011; Moscovich & Hughes, 2006; Sears, Lin, & Karimullah, 2002; Ware & Lowther, 1997) in VR use a joystick or cursor to point at a target object, which is then selected by clicking or dragging. These methods are intuitive but inefficient when dealing with complex scenes and large numbers of objects.

Lucas (2005) et al. proposed a Selection Box and Tablet Free-hand Lasso to help users select multiple targets in 3D space. Stenholt (2012) proposed 3D Spherical Selector (the technique is called 3D Spherical Brush), 9Dof lasso in 3D space, and an algorithm based on seed objects and Gestalt proximity law (the technique is called Magic Wand) to solve the problem of selecting large-scale objects (hundreds of objects) in a scene. These works are based on devices such as a stylus or mouse to make selections of objects in a 3-dimensional scene in a computer and do not involve VR technology. With the development of VR technology, there are more and more researches on selections in VR scenarios. Montano-Murillo et al. (2020) proposed a hybrid selection technique called ”slice volume” for 3D selection in dense virtual reality environments (e.g., point clouds). Benavides, Khadka, & Banic (2019) allowed users to choose the intended pile of rendered spatial data points with bare-hand mid-air gestures and movements. Li et al. (2022) proposed a sequential target selection method called Sewing to improve the consistency of target selection. Shi et al. (2023) proposed Cone selection for the simultaneous selection of multiple objects. Wu, Yu, & Goncalves (2023) proposed a novel group selection method based on the touch approach. Shi et al. (2024) present an empirical comparison of six freehand techniques derived from prior work in object selection. Compared with other methods for multi-object selection in VR environments, our approach offers greater degrees of freedom. Many existing multi-object selection techniques have certain limitations and inconveniences when handling single-object selection tasks. Our method comprehensively considers various scenarios, designing distinct modes to accommodate different user requirements.

Multi-object selection techniques in 3D virtual worlds have applications in many fields. In VR and augmented reality (AR), multi-object selection is used for scene editing, virtual modeling, and interactive game development. In computer-aided design (CAD), multi-object selection tools help designers quickly select and edit complex models.

Despite significant progress in multi-object selection techniques, several challenges remain. First, many methods have high computational overheads when dealing with complex scenes and large-scale data. Second, how to maintain high-precision multi-object selection in dynamic scenes is also an urgent problem.

2.2. Object Grouping

The study of object manipulation interaction in 3D scenes is one of the main research problems in VR today (Bergström, Dalsgaard, Alexander, & Hornbæk, 2021; Hancock, Carpendale, & Cockburn, 2007; Mendes, Caputo, Giachetti, Ferreira, & Jorge, 2019; Poupyrev, Billinghurst, Weghorst, & Ichikawa, 1996; Ruddle, 2005). Group manipulation is one of the main research focuses. Many interaction techniques have been proposed for manipulating a single object at a time. While repeating the same operation for multiple objects is possible, it can also be more time-consuming and tedious than first grouping objects and then manipulating them together, especially when there are many objects to be manipulated. Compared to multi-object selection, object grouping tasks will have modification and deletion operations for groups, as well as manipulation of objects within groups. Users will have additional interactions with the generated groups.

On grouping objects in the 2D plane, there are some work has been done by researchers. Dehmeshki & Stuerzlinger (2009) proposed an enhanced form of Lasso selection in 2D space for object grouping. Strothoff, Stuerzlinger, & Hinrichs (2015) applied visual tag markers to objects to visualize their group association. Dehmeshki & Stuerzlinger (2010) presented a domain-independent perceptual-based selection technique that allow selecting arbitrary groups. On grouping objects in the 3D scene, Shi et al. (2022) presented a variety of methods for implementing group-based object alignment in VR and compared their performance and effectiveness. The application of dual constraints (Stuerzlinger & Smith, 2002) enabled the rapid formation of object clusters and facilitated their direct manipulation within the desktop virtual environment. By applying dual constraints to adjacent objects, it became possible to move, rotate, and reorganize the entire group cohesively (Shi, Zhu, Liang, & Zhao, 2021; Stuerzlinger & Smith, 2002). When working with virtual objects that are somehow related to each other, the hierarchical relationship of groups implicitly defines constraints so that users can interact with groups of objects or their subgroups (Oh et al., 2006; Wolfartsberger, Zenisek, Sievi, & Silmbroth, 2017). In the Beyond Snapping technique (Ciolfi Felice, Maudet, Mackay, & Beaudouin-Lafon, 2016), users are able to align target graphical objects onto pre-established StickyLines, subsequently applying alignment or distribution operations to these objects. In a similar fashion, a handlebar mechanism allows for the attachment and coordinated manipulation of virtual objects (Song, Goh, Hutama, Fu, & Liu, 2012). Both of these approaches achieve group formation and manipulation through the explicit creation of line constraints. Utilizing Plane, Ray, and Point tools, users can establish a plane constraint to align objects within immersive VR settings (Hayatpur, Heo, Xia, Stuerzlinger, & Wigdor, 2019). With these methods, it is necessary for users to first define the constraints explicitly. These constraints then serve as the basis for group formation and subsequent manipulation activities.

However, none of this work deals with grouping objects in 3D scene according to the user's needs. In our work, we focus on conveniently grouping objects and modifying them in VR according to the user's wishes.

2.3. Bare Hand Based VR Interaction.

Bare-hand interaction allows users to interact with their hands in virtual environments without the use of physical controllers or peripheral devices, and thus, it has attracted significant attention in VR. This section reviews the existing literature and research work related to bare-hand interaction in VR.

Von Hardenberg & Bérard (2001) developed and evaluated a finger recognition and gesture recognition algorithm and developed three example applications. Finger tracking and gesture recognition were used to draw on a virtual wall, to control a presentation through gestures, and to move virtual objects on the wall during brainstorming sessions. Fernandes & Fernández (2009) presented a system that tracks the 2D position of the user's hands on a tabletop surface, allowing the user to move, rotate and resize the virtual objects over this surface. Ong & Wang (2011) presented 3D bare-hand interaction in an augmented assembly environment to manipulate and assemble virtual components. Using the leap motion controller, Nabiyouni, Laha, & Bowman (2014) designed travel techniques with bare-hand interaction. Wang, Ong, & Nee (2016) proposed a novel human Cognition-based interactive Augmented Reality Assembly Guidance System (CARAGS) to investigate how AR can provide various modalities of guidance to assembly operators for different phases of user cognition process during assembly tasks. Zhao, Ong, & Nee (2016) presented a low-cost and multimodal residential-based AR-assisted therapeutic healthcare exercise system. Fang & Hong (2022) proposed a bare-hand occlusion-aware interactive AR assembly method based on the monocular image, and a lightweight deep neural network is established to infer the depth relationship between the 3D virtual model and real scene including gesture manipulation, and thus the ambiguous AR instruction by inaccurate occlusion deduction would be prevented, leading to more realistic bare-hand interactive AR guidance for manual assembly. Schäfer, Reis, & Stricker (2022) explored different techniques for bare-hand locomotion since it could offer a promising alternative to existing freehand techniques. All of the above research is based on existing bare hand tracking techniques that are used in VR and AR to design the way the user interacts with objects in the scene. Our approach is similar in that we use the bare hand tracking technology that comes with VR headsets as a way to design the way users interact with objects in the VR world.

Bare-hand interaction has been widely used in various virtual reality applications such as education, gaming, medicine, and engineering. Researchers have developed various educational applications based on bare-hand interaction, such as virtual laboratories and interactive learning environments, to facilitate learning and knowledge transfer. In addition, in the field of gaming and entertainment, bare-hand interaction provides users with a more immersive and interactive experience.

Despite significant progress, challenges remain in the development and application of bare-hand interaction in VR. Technical challenges include improving the accuracy of hand tracking, robustness of occlusion, and reducing latency to enhance responsiveness. Additionally, addressing ergonomic issues to ensure adaptation to users with different hand sizes and abilities is an important consideration in advancing bare-hand interaction technology in VR.

3. HandBrush: Bare-hand Object Grouping Method

Inspired by the works done by Strothoff et al. (2015), Dehmeshki & Stuerzlinger (2009) and Dehmeshki & Stuerzlinger (2010) in the 2D plane, we present HandBrush, a method for bare-hand object grouping in a virtual environment. In this section, we'll go into more detail about our HandBrush method.

3.1. Design Rationale

During our design process, we want our approach to be more immersive, easier to learn, and more efficient in accomplishing tasks. From these three points, we designed the HandBrush method. In terms of better immersion, since in the physical world we interact with objects directly through our hands, this also means that the use of facilities such as joysticks may have some impact on user immersion. That's why our new approach will be based on designing with bare hands. In terms of ease of learning, we referenced the concept of transfer learning when designing the HandBrush methods. We drew on the characteristics of brushes and how they are used to make it easier for users to master our methods. In addition, in terms of efficiency, we considered scenarios with different levels of complexity to ensure that our approach is as efficient as possible to help users accomplish their tasks in any scenario.

3.2. Design Metaphor

The brush, as an essential tool for artistic creation, has provided humanity with a rich means of artistic expression. Its use has promoted the development of art forms such as painting and calligraphy, becoming a vital medium for cultural transmission. The design of the HandBrush system draws on the brush's intuitiveness and expressiveness in artistic creation. We simulate the user's hand movements as a brush, allowing for the selection and manipulation of objects in virtual space through these actions. This design aims to translate the natural motions of users in the real world into intuitive interactions within the virtual environment, thereby enhancing user immersion and reducing the learning curve.

As shown in Figure 2, we define HandBrush as $B(s, c)$, and the parameters s and c are the parameters that determine the running state of our brush. Where s represents the size of our brush and c represents the dye that our brush is dipped in.

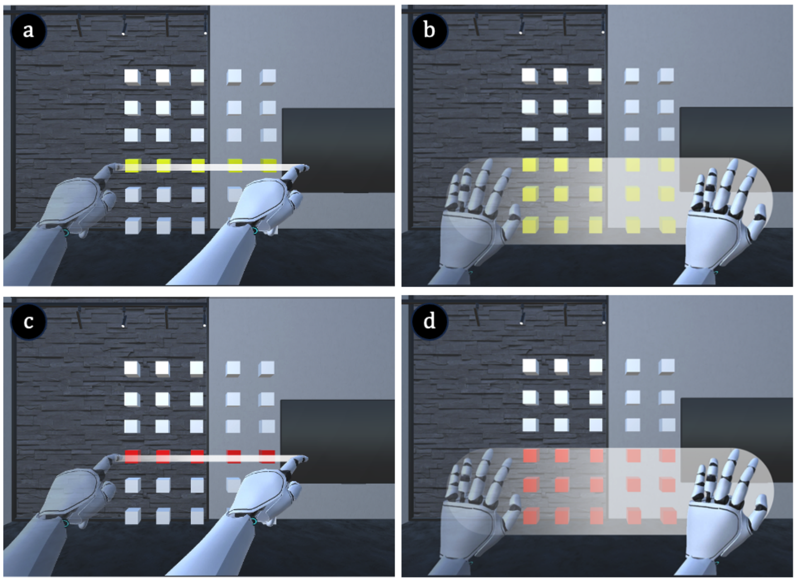

Figure 2. Schematic visualisation of the HandBrush method under different s and c conditions. (a) $s = 1, c = 1$, means that the user is selecting an object in 3D space using the small brush. (b) $s = 2, c = 1$, means that the user is deleting an object in 3D space using the large brush. (c) $s = 1, c = 2$, means that the object is deleting a group or an object within a group that exists in 3D space using the small brush. (d) $s = 2, c = 2$, means that the object is deleting a group or an object within a group that exists in 3D space using the big brush.

3.2.1. Size of Brush

In traditional art creation, the size of a brush directly influences the coverage area and detail representation during painting. Large-sized brushes can quickly cover large areas, suitable for broad strokes and rapid color layout, while small-sized brushes are suitable for precise line drawing and detail handling. This relationship between size and function inspired the design of the HandBrush system.

The HandBrush system simulates this brush characteristic by using different parts of the user's hand to mimic brushes of varying thickness, thereby achieving different selection and operation effects in the VR environment. HandBrush offers two sizes.

When the value of s (the parameter representing the brush size) is 1, the user's index fingertip is simulated as a fine brush. This mode is suitable for fine manipulation, such as selecting small objects or performing precise modification tasks. The fine contact area of the index fingertip allows the user to cover a smaller area per unit movement path for finer selections.

In contrast, with the value of s is 2, the user's entire hand is simulated as a wide brush. This mode is suitable for fast and large selection tasks, such as selecting multiple objects at once or performing operations over large areas. The large contact area of the palm allows the user to cover a larger area per unit movement path, improving selection efficiency.

By offering these two modes, which correspond to different brush sizes, the HandBrush system provides users with flexibility in selecting and operating objects in the VR environment. Users can choose the size of brush that best suits their specific task requirements and preferences.

3.2.2. Paint Color

In the art of painting, the choice of brush color is not only a matter of aesthetics but is also often used to convey specific emotions and information. Color, as a powerful visual language, can quickly communicate the state and purpose of a work to the observer. We use color changes to intuitively indicate the current operation and method to the user. This design not only enhances the visual experience but also improves the user's cognitive efficiency regarding different modes.

When the value of c (the parameter representing the paint color) is 1, the dye used by the brush is yellow, and HandBrush is in Line mode and Block mode to select objects to generate a group. When the value of c is 2, the dye used for the brush is red, and HandBrush is in Erase mode for deleting objects in a group or deleting a group.

By employing color coding, the HandBrush system provides users with a quick and intuitive way to recognize and switch between different modes. This design reduces the cognitive load associated with mode recognition, allowing users to focus more on the task at hand, thereby enhancing overall interaction efficiency.

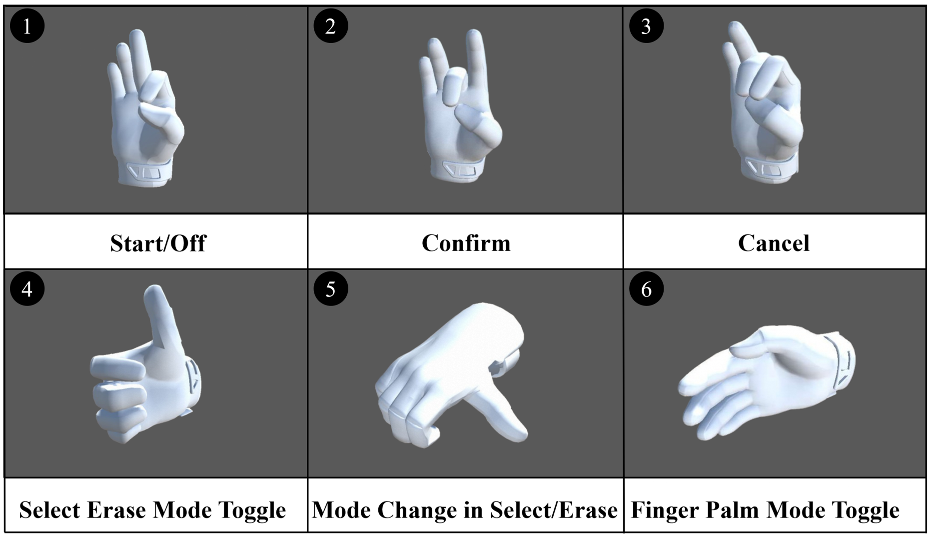

Figure 3. Gesture *1* controls the opening and closing of our selection state. Gesture *2* is a confirmation gesture of the selection result. Gesture *3* is a gesture to undo the current selection. Gesture *4* is the mode-switching gesture of Select and Erase. Gesture *5* is the switching of Line mode and Block mode or Erase-object and Erase-group in Erase mode. Gesture *6* is the switch between Line-finger and Line-palm in Line mode.

3.3. Interaction Design

Artistic creation is a crucial component of human culture. The brush, as a widely used tool, provides artists with a rich means of expression. With the continuous development of VR technology, bringing this artistic creation experience into virtual environments has become feasible. HandBrush is an object-grouping interaction method for VR environments designed to mimic the intuitiveness and fluidity of using a paintbrush in the real world.

In order to achieve smooth control of HandBrush and a more natural gesture-switching process, as shown in Figure 3, we defined six different gestures and designed a complete interaction control process so that the user can use our HandBrush to select, group and modificate objects in the virtual scene. When designing gestures, our

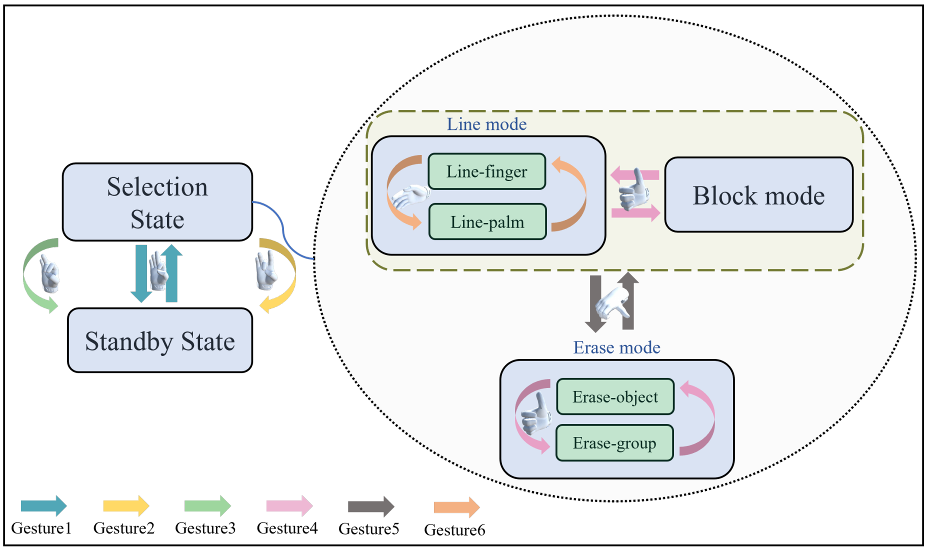

Figure 4. The state transition diagram of the HandBrush method.

first consideration was simplicity. We wanted all the operations to be done with one hand, so as not to put too much physical burden on the user. We borrowed common gestures from other commercially available headsets, such as the finger pinch gesture and thumb cocking gesture, designed a variety of gestures in conjunction with the Pico neo3. During our gesture design process, we discovered that the gesture recognition of Pico neo3 is not very sensitive to certain gestures. So we chose these 6 gestures that Pico neo3 recognizes more accurately after several attempts. The state transition diagram of the HandBrush is shown in Figure 4. We draw the associated state transfer diagrams, which allow users to realize silky smooth transitions before different modes through gestures.

3.3.1. Line Mode

The core of this mode lies in selecting objects by drawing lines. In this process, lines serve not only as selection tools but also as visual links that connect and organize the selected objects. Specifically, we connect objects intended to be grouped together using lines, thereby forming a unified collection both visually and logically, as shown in Figure 5.

To accommodate different usage scenarios and user preferences, Line mode is further divided into two operational types: Line-finger and Line-palm. In Line-finger, the user's fingertip acts as a brush, allowing the user to draw lines by moving their fingertip to select objects. This mode is suitable for precise operations, enabling users to select with high accuracy. Conversely, in Line-palm, the user's entire palm acts as a larger brush, selecting objects by moving the palm. This mode is ideal for quick, large-scale selection tasks, especially when multiple objects need to be selected simultaneously.

In both modes, the space swept by the defined brush is considered the selection area, and all objects within this area are deemed selected. The system uses lines to string objects together in the order in which we swept them. The trajectory of the strokes is intended to show the order in which objects are selected. This selection mechanism is not only intuitive and easy to understand but also simple to operate, significantly enhancing the user experience. Once the selection is completed and the user makes a confirmation gesture (gesture 2 in Figure 3), all selected objects are automatically grouped by the system. To further reinforce this grouping relationship, the system connects these objects with a line, creating a clear visual and logical link.

Through this design, Line mode not only improves selection efficiency but also enhances the user's sense of control over the selection process. Additionally, connecting objects with lines facilitates subsequent object manipulation and management, such as splitting and merging groups. Users can more intuitively identify and operate on the formed object collections, with these functions to be introduced in the following sections.

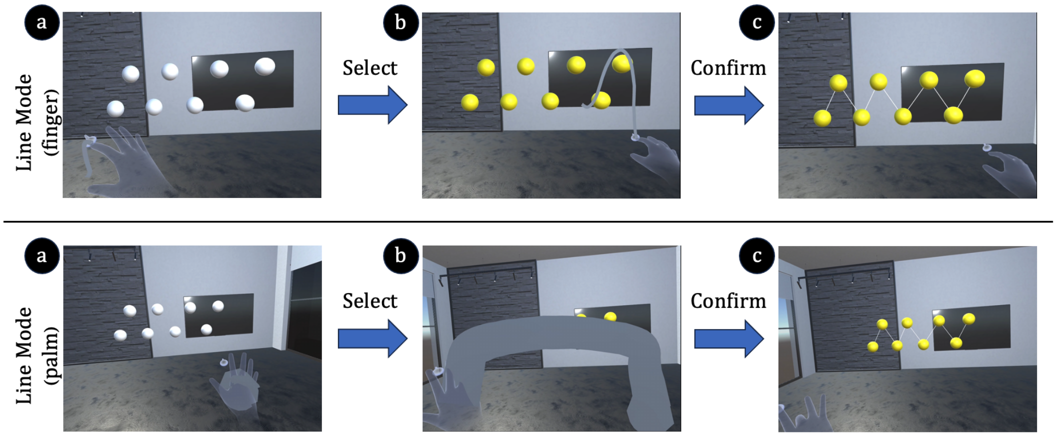

Figure 5. Screenshots of the Line mode.

3.3.2. Block Mode

In Block mode, we use the tip of the user's index finger as a brush to draw a closed ring of any shape in the user's view. When strokes cross, it means the user has drawn a ring. The objects in the ring from the user's view will be selected, and then this selection ends, exiting Block mode. This design grants users a high degree of freedom, allowing them to select specific sets of objects as needed without being restricted to preset shapes. This intuitive operation not only enhances the flexibility of selection but also allows users to have more precise control over the selection process.

Once the closed loop is drawn, all objects within the loop in the user's view are considered selected, as shown in Figure 6. This selection mechanism, based on spatial positioning, ensures that users can quickly and accurately select the objects they need. After completing the selection and making a confirmation gesture, the system will automatically group the selected objects for subsequent manipulation and management.

To further distinguish different object groups, the system uses blocks matching the shape of the closed loops to enclose the selected objects. These blocks visually indicate the selected area, and each group's corresponding block has a distinct color. This color-coding mechanism provides users with a simple and effective way to identify and differentiate between various object groups, thereby improving efficiency and accuracy during complex operations.

3.3.3. Erase Mode

To further enhance user experience and provide more flexible editing functionalities, the HandBrush system introduces the Erase mode, allowing users to finely edit grouped objects. In this mode, the system offers two types: Erase-object and Erase-group, to

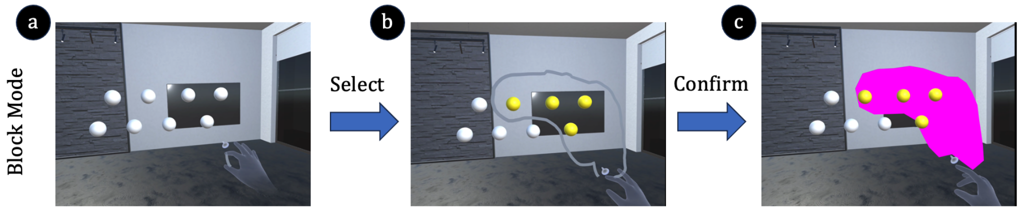

Figure 6. Screenshots of the Block mode

accommodate different editing needs.

- **Erase-object**: The Erase-object focuses on deleting certain objects in a group. For example, if a user mistakenly includes an unnecessary object during the initial selection process or decides to remove an object during subsequent editing, the Erase-object provides a direct and straightforward solution. HandBrush allows users to select and erase specific objects within a group. In this mode, when the brush passes over an object, it turns red. Users can perform consecutive operations to select multiple objects for batch editing. Once the user indicates completion of the selection through a confirmation gesture, the selected objects are removed from their respective groups, allowing for precise control and editing of group members, as shown in Figure 7.
- **Erase-group**: Unlike the Erase-object, the Erase-group allows users to remove entire groups of objects. When a user needs to delete an entire set of objects or when the entire group no longer meets the user's needs, the Erase-group provides a quick way to handle this. In this mode, when a user uses the HandBrush to pass over an object, the entire group to which the object belongs is selected. At this point, the lines connecting the object or the blocks encircling the object turn red, clearly indicating the selected group. Users can continue to select multiple groups for editing. When the user gives the confirmation gesture, the selected groups will be removed, and all objects within the groups will return to an ungrouped state, allowing users to reorganize or perform other operations, as shown in Figure 7.

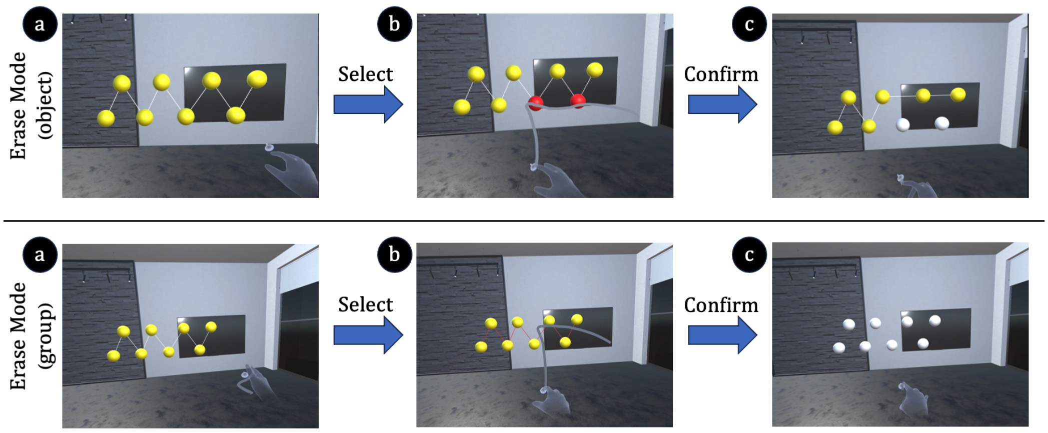

Figure 7. Screenshots of the Erase mode

3.3.4. Split Feature

The Split feature in the HandBrush system is designed for both Line and Block mode, providing users with an innovative means of group editing. By leveraging the segmentation ability of lines and blocks, the Split feature allows users to divide existing groups through intuitive gesture operations, thereby creating new groups or further refining existing ones.

For groups determined through Line mode, users can use the HandBrush to split connected lines as shown in Figure 8. When users wish to separate a group into two independent ones, they can swipe the brush at the desired splitting position. This action divides the original line into two separate lines, each connecting a group of objects. As a result, the original group is explicitly divided into two distinct groups, each with its own line to identify and connect the objects within.

For groups determined through Block mode, the Split feature allows users to split closed-loop block shapes that enclose objects as shown in Figure 8. Users can use the HandBrush to split the block at any position, dividing it into two parts. Each newly formed part encloses a portion of objects, constituting a new group.

Specifically, the system keeps track of the objects that the user's finger has scratched. If the number of objects crossed by the user is 1, and the type of the object is the "line" we generated in the line mode, the system will consider that the operation is the split feature of Line. If the number of objects crossed is 1, and the object type is "block" generated in Block mode, then the system will consider this operation as a split feature of Block.

The Split feature is not only simple and intuitive but also allows users to easily adjust the grouping when needed without having to redo the entire grouping process.

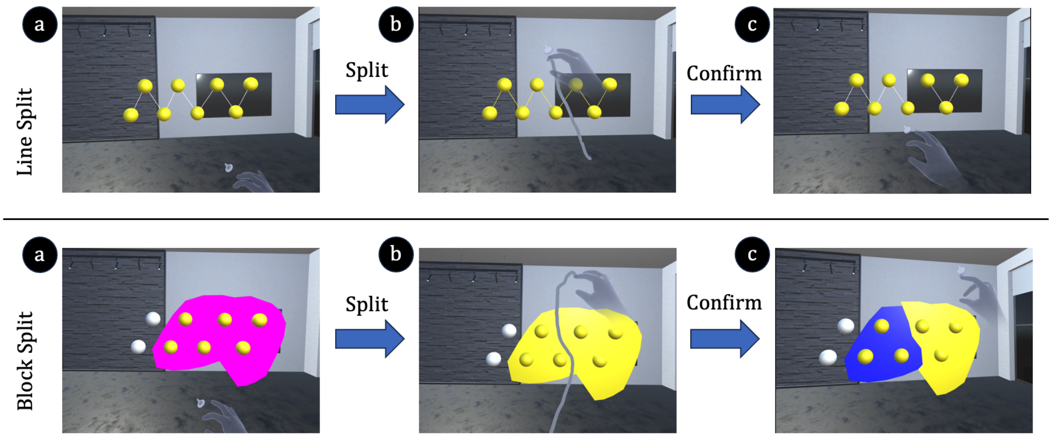

Figure 8. Screenshots of the Split feature

3.3.5. Connect Feature

The Connect feature in the HandBrush system is an efficient interaction design that leverages the joinable properties of lines (generated by Line mode) and blocks (generated by Block mode), allowing users to merge different groups into a unified whole through simple gestures. The user selects a group of lines or blocks in the view as the starting point and draws a line along the expected path, which serves as a merging medium to connect the lines or blocks that pass through it. Specifically, if our stroke trajectory crosses multiple "lines" and "blocks", the system considers the operation to be a connect operation between groups, i.e. the Connect feature. All groups that are connected merge into a new group. The original lines will extend and form a larger line structure as Figure 9 shows.

This method, through visualized connection operations, allows users to intuitively see which groups will be merged. Users are free to choose the groups they want to merge, without being constrained by fixed patterns. It simplifies the steps for merging between groups, thereby enhancing overall work efficiency. The design of the Connect feature reflects the HandBrush system's profound understanding of user habits and its emphasis on interaction experience. It provides users with a simple, intuitive, and powerful group management tool, greatly improving users' work efficiency and flexibility when performing complex tasks.

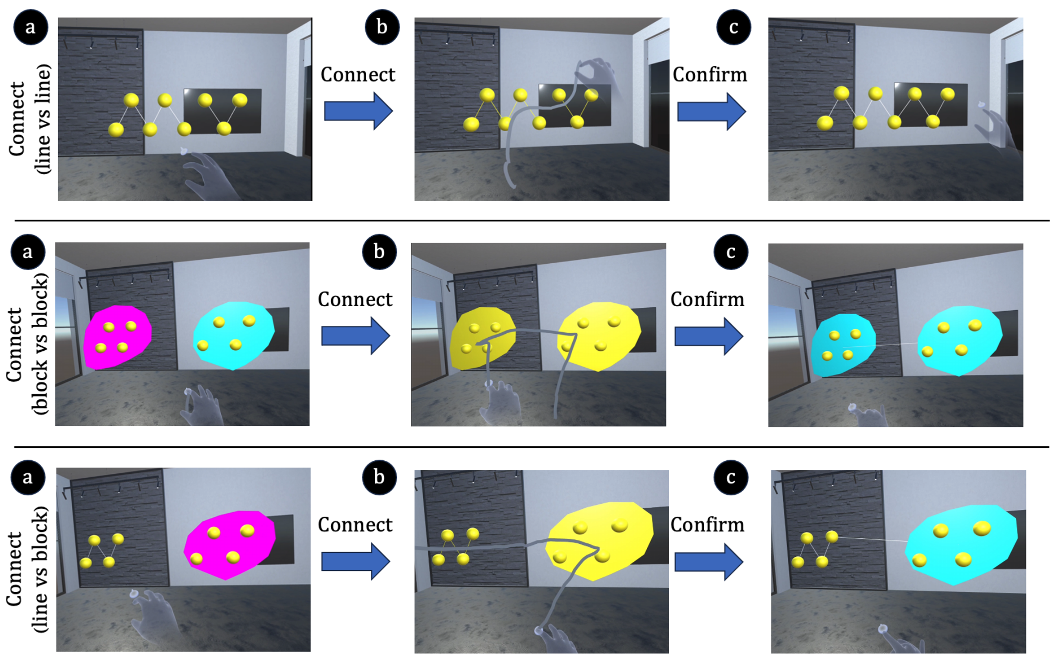

Figure 9. Screenshots of the Connect feature

3.4. Similarities and Difference with Comparison Methods

In this article, we compared our HandBrush method with the Ray method (a ray is shot from the user's hand, and objects that the ray collides with are considered to be selected), Screen method (the user uses a pinch gesture to draw a rectangular box in the user's view. Objects in the box are considered to be selected), and Cone method proposed by Shi et al. (2023). In this section, we will compare the similarities and differences between the HandBrush method and the comparison method.

In our group generation task, the HandBrush method and the comparison methods share a common goal: to provide a selection mechanism that allows users to conveniently select and manage objects in a virtual environment. In terms of selection mechanisms, HandBrush's Line mode and the Ray method from the comparison methods both adopt a sequential selection approach, offering a continuous selection solution. HandBrush's Block mode, like the Screen and Cone method from the comparison methods, selects objects by defining an area.

At the same time, there are differences between our method and the comparison methods. In terms of selection mechanism, Line mode, compared to the Ray method, strings objects together through a brush-like motion, emphasizing the continuity and intuitiveness of group selection. The Ray method, on the other hand, resembles using a beam to cut through objects, focusing more on the precise selection of individual objects. Block mode allows users to customize the shape of the closed loop, providing greater flexibility to enclose any shape of object collection according to user needs. In contrast, the Screen and Cone method offer fixed shapes for the selection area, such as rectangular or conical regions, limiting the ability to customize the selection shape.

Both our method and the comparison methods offer group deletion functionality in the group modification task. We can complete this task by deleting existing groups and regrouping. However, our method differs by providing additional Split and Connect features. These features help modify existing groups, offering users more freedom.

4. Empirical Study 1

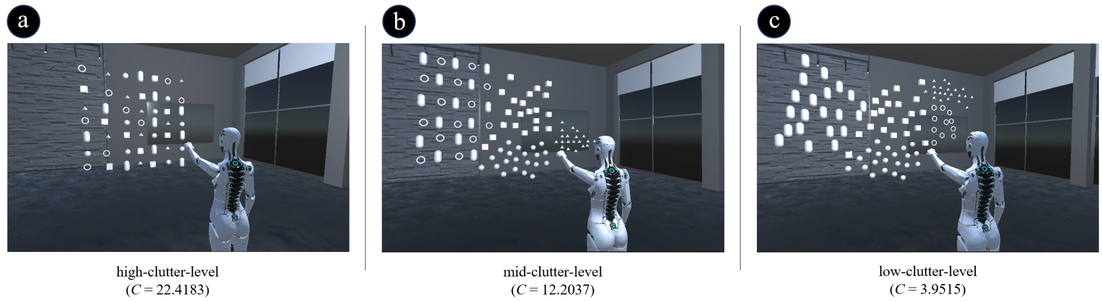

Figure 10. Group Generation of Experimental Scenarios. Participants need to group objects according to shape. Objects of the same shape are grouped together. (a) is the scene of high clutter level. (b) is the scene of mixed clutter level. (c) is the scene of low clutter level.

In this section, we designed a user study to evaluate the performance of our HandBrush in group generation task. We compared our HandBrush method with the Ray method (a ray is shot from the user's hand, and objects that the ray collides with are considered to be selected), Screen method (the user uses a pinch gesture to draw a rectangular box in the user's view. Objects in the box are considered to be selected), Cone method proposed by Shi et al. (2023). A schematic of the different methods to accomplish this task is shown in Figure 11

In this study, we investigated the effects of scene clutter and technique on group generation performance. Based on the results of this study, we tested the following hypotheses:

H1: HandBrush method can accomplish the task of grouping objects significantly faster than comparison methods

H2: Compared to comparison methods, HandBrush method can accomplish tasks with higher success rate.

H3: Compared to comparison methods, HandBrush method can accomplish tasks with less hand movement.

H4: HandBrush method brings less fatigue compared to comparison methods.

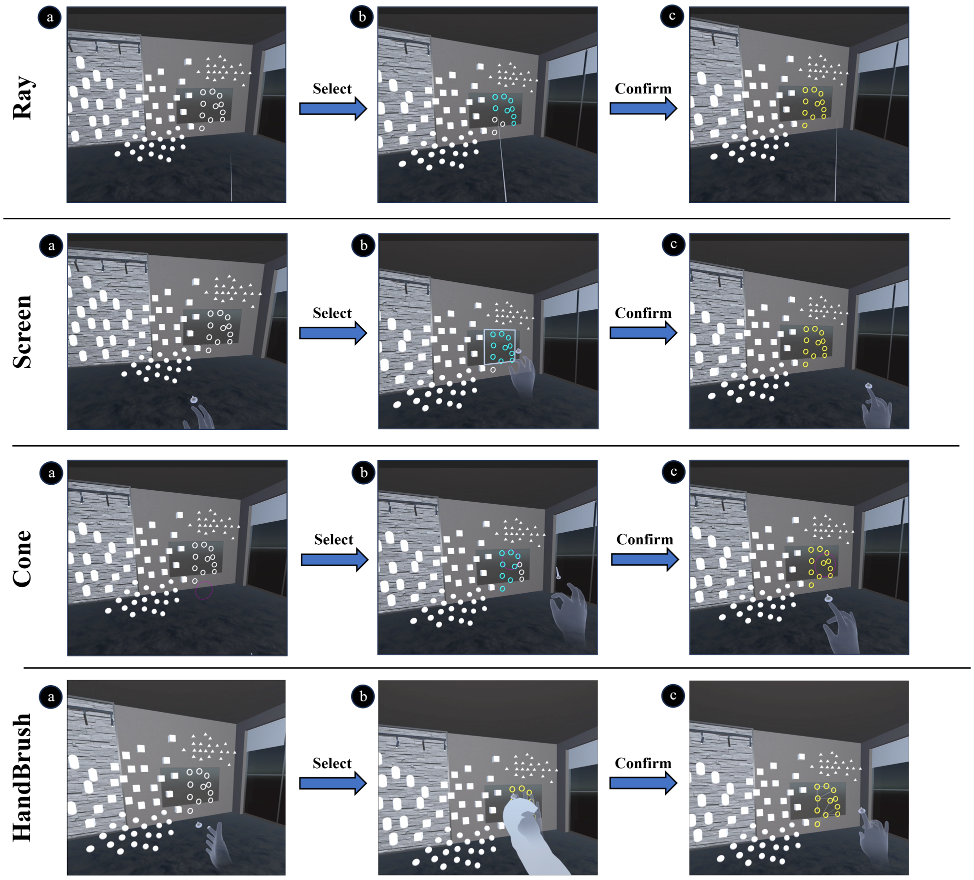

Figure 11. Screenshots of 4 different methods in this task. From top to bottom are Ray method, Screen method, Cone method and HandBrush method.

4.1. Participants and Apparatus

Sixteen participants (10 males, 6 females), aged 21 to 26 (mean 22.88, variance 1.36), took part in this experiment. None of them had experience with VR equipment. All participants had normal vision (or were corrected to normal vision by wearing glasses). The experimental system used a Unity 2022 and a Pico Neo3 headset. The system ran on the Pico Neo3.

4.2. Experimental Design

We used a 4×3 within-subjects design in which technique (Ray method vs Screen method vs Cone method vs HandBrush) and clutter level (low, medium, high) were used as two independent variables.

We randomly placed items of various shapes in a rectangular space located 10 meters away from the participants. Due to the absence of occlusion considerations, our items were arranged on a single plane. We considered items of the same shape to belong to the same group. We designed three scenes with different levels of clutter for users to complete the tasks.

The degree of clutter is defined as the level of complexity and confusion in a scene or environment. Here, we use spatial information entropy (Rényi, 1961) to calculate our degree of clutter. We divide the scene into n discrete small grids of size $5 \times 5 \times 5$. Next, we compute the distribution probability of a certain kind of object in each grid, i.e., we count the number of a certain kind of object in each grid as a proportion of the total number of objects in the grid. Next, we calculate the amount of partial entropy for each kind of object in each grid. Finally, we superimpose the amount of partial entropy for each object in each grid to obtain the scene clutter information entropy of our scene.

$$C(X) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p(x_{ij}) \log_2 p(x_{ij}) \tag{1}$$

where m represents the number of scene grids and n represents the number of object types in the scene. $p(x_{ij})$ represents the probability of occurrence of item j in the i-th grid.

4.3. Task and Procedure

We set up 3 scenarios with different levels of clutter, as shown in Figure 10. The three metric values for scene clutter level are 22.4183, 12.2037 and 3.9515, representing high scene clutter, medium scene clutter and low scene clutter, respectively. The user's task in this experiment is to group the objects in the scene according to their shapes under three different complexity scenarios using the comparison method and our method respectively, with objects of the same shape as a group. This task is complete only when all objects in the scene are correctly grouped.

Each participant spent approximately 90 minutes on the entire experiment. First, they completed a questionnaire about their personal information and prior experience with virtual reality head-mounted displays (VR HMDs). Then, they were introduced to the VR equipment, the experimental design, and the tasks. Next, they put on the head-mounted display (HMD) and began the experiment. Participants performed the tasks while standing. Before the formal experiment under each technique condition, participants had a fixed 10-minute training session to familiarize themselves with the technique. They will be asked to complete tasks using different technologies in specific training scenarios to achieve a certain level of proficiency in each technology.

Participants then started the experiment. They were required to group objects in the scene as quickly and accurately as possible, with objects of the same shape forming a group. We used a Latin-square design to balance the order of technique conditions and randomized the sequence of clutter levels within each technique condition. Participants took a 2-minute break after completing the group generation task using one technique in each scenario. On average, each task took 2 minutes to complete. In total, we collected 768 data points (16 participants * 4 techniques * 3 clutter levels * 4 repetitions).

After each technique condition, participants were asked to complete the Usability questionnaire (Kim, Lee, & Billinghurst, 2015) and the NASA-TLX questionnaire (Hart, 2006) and take a short break. At the end of the experiment, a brief interview was conducted to gather their subjective impressions. We will ask users how they feel about using various techniques, and how they feel about and prefer different modes and features of HandBrush in different scenarios.

4.4. Metrics

In this study, we established four objective metrics and two subjective metrics to measure the performance of different techniques in scenarios with different levels of clutter.

Generation time. In this study, generation time refers to the duration from when the user initiates a task until all objects are correctly grouped. This metric is crucial for evaluating the performance of different techniques in immersive environments.

Operation success rate. In this experiment, we consider a successful operation when the user selects all objects with the same shape as a group. When the user mistakenly selects objects of different shapes as a group or omits to select objects of the same shape when dividing the group, we consider it a selection error. The operation success rate is the number of successful operations divided by the total number of operations.

Hand travel distance. In this study, we introduced the concept of Hand travel distance to quantify the spatial distance traveled by the virtual hand of participants while performing tasks. This metric was chosen as a key parameter to measure the impact of different technologies on user fatigue. Specifically, hand travel distance reflects the total distance that the user moves his/her hand while performing operations in the virtual environment, which is directly related to the physical exertion and fatigue felt by the user while completing the task.

Usability. A usability questionnaire (Kim et al., 2015) to evaluate users' perceptions of technology usability, focusing on intuition, efficiency, accuracy, naturalness, satisfaction, and ease of use, scored from 1 to 10. The six questions are: is this method intuitive (Q1), is the method efficient (Q2), is the method accurate (Q3), is the method natural (Q4), is the method satisfied (Q5), is the method easy to use (Q6).

NASA-TLX. The NASA-TLX questionnaire (Hart, 2006), assessing workload across six dimensions (mental, physical, time, performance, effort, and frustration) with a 1-20 scoring scale. The scores of the six dimensions were weighted and averaged to obtain a composite load score.

4.5. Results

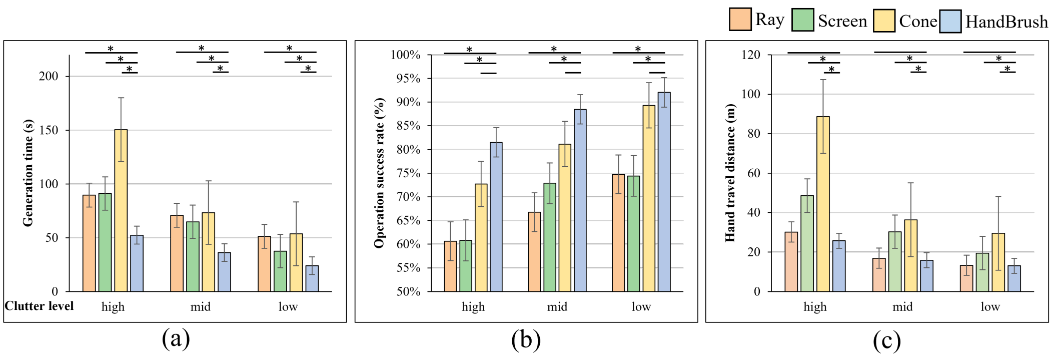

Figure 12. Plots (a) to (c) show three objective metrics results. The horizontal coordinates indicate different levels of scene clutter. The four methods of comparison are distinguished by different colors. Error lines indicate standard deviations, and asterisks indicate significant differences in the results of the comparison between the two methods.

4.5.1. Objective Measurements

For each method and confusion condition, outlier data points (± 3 standard deviations) were first filtered out. In total, we excluded 10 data points (1.30%). Then we calculated each user's average score for each metric. In the next analyses, we used a two-factor multivariate analysis of variance (MANOVA) to analyze the effects of technique and clutter level on the performance of the three objective metrics described above. Figure 12 gives the results for each objective metric (a is generation time, b is operation success rate, c is hand travel distance). In each subfigure, the results of our method are compared with those of the comparison methods. Results marked with an asterisk indicate that the difference is statistically significant.

The assumptions of the methodology were tested before proceeding with the analysis. Scatter plots showed a linear relationship between the dependent variables in each set of independent variables. The Pearson correlation test found no multicollinearity ($|r| < 0.9$) between the two dependent variables. Box plots did not find unidirectional outliers and Mahalanobis distance did not find multivariate outliers.

The Shapiro-Wilk test showed that the two dependent variables (generation time, hand travel distance) obeyed a normal distribution ($p > 0.05$), whereas the operation success rate variable did not obey a normal distribution under certain conditions ($p = 0.001$). However, because MANOVA is robust to deviations from normal distribution, especially when the sample sizes of the groups are equal or nearly equal, non-normal distribution does not significantly increase the likelihood of Type I errors. Therefore, we continued the test without additional adjustments.

Box's M test showed that the variance/covariance matrices of the three dependent variables within each group of the independent variables were not equal ($p < 0.001$). So, we used Pillai's criterion statistic because it is more robust to unequal covariance matrices. Levene's test showed that the dependent variables within each group of the independent variables were equal in variance ($p > 0.05$).

We also quantified effect sizes using Cohen's d. Cohen's d values (Sawilowsky, 2009) were transformed into qualitative estimates of effect sizes including huge (d > 2.0), very large (2.0 > d > 1.2), large (1.2 > d > 0.8), medium (0.8 > d > 0.5), small (0.5 > d > 0.2), and very small (0.2 > d > 0.01).

Interaction Effect There was a statistically significant effect of the interaction of technique and clutter level on the dependent variable ($F_{18,540} = 10.285, p < 0.001^{*}, \Lambda_{\text{Pillai}} = 0.766, \eta_p^2 = 0.453$), meaning that the effect of scene clutter level on the dependent variable differed across the four methods. MANOVA tests showed that the interaction between technique and scene clutter level was statistically significant in terms of generation time ($F_{6,180} = 24.496, p < 0.001^{*}, \eta_p^2 = 0.450$) and hand travel distance ($F_{6,180} = 70.143, p < 0.001^{*}, \eta_p^2 = 0.700$). However, the effect on success rate was not significant ($F_{6,180} = 0.374, p = 0.895, \eta_p^2 = 0.012$).

Simple Main Effect Because of the statistically significant interaction effects between technique and scene clutter level, we will report the simple main effects of each factor separately here.

Technique. Simple main effects analyses showed that the differences between the four technologies were statistically significant at all clutter levels in terms of generation time ($p < 0.001$), operation success rate ($p < 0.001$) and hand travel distance ($p < 0.001$).

For each dependent variable and for different levels of scene clutter, the Tukey post-hoc pairwise comparisons of the four techniques were made.

For generation time, there is a significant difference between HandBrush method and Ray method ($p < 0.001$), HandBrush method and Screen method ($p < 0.001$), HandBrush method and Cone method ($p < 0.001$), Cone method and Ray method ($p < 0.001$), Cone method and Screen method ($p < 0.001$) at high clutter level, while Ray method and Screen method are not statistically significant ($p = 0.999$). At medium clutter level, HandBrush method and Ray method, Screen method and Cone method are all significantly different ($p < 0.001$), but Ray method and Screen method ($p = 0.999$), Ray method and Cone method ($p = 0.999$), Screen method and Cone method ($p = 0.470$) are not significantly different. At low clutter level, HandBrush method and Ray method ($p < 0.001$), HandBrush method and Screen method ($p = 0.034$), HandBrush method and Cone method ($p < 0.001$), Screen method and Cone method ($p = 0.007$), Screen method and Ray method ($p = 0.034$) are all significantly different. But Cone method and Ray method do not have significant difference ($p = 0.999$). Table 1 shows the part of the pairwise comparison between HandBrush and the comparison method on the generation time metrics.

Table 1. Generation Time (s)

Clutter level	Technique	Avg ± std. dev.	$(CC_i\text{-}EC)$ / CC_i	p	Cohen's d	Effect size
High	$HandBrush(EC)$	54.30 ± 12.19				
	$Ray(CC_1)$	89.67 ± 13.45	39.44%	$< 0.001^*$	2.75	Huge
	$Screen(CC_2)$	91.06 ± 18.25	40.36%	$< 0.001^*$	2.36	Huge
	$Cone(CC_3)$	150.46 ± 14.01	63.91%	$< 0.001^*$	7.32	Huge
Mid	$HandBrush(EC)$	36.14 ± 8.24				
	$Ray(CC_1)$	70.87 ± 10.43	36.57%	$< 0.001^*$	3.69	Huge
	$Screen(CC_2)$	64.75 ± 17.76	31.82%	$< 0.001^*$	2.06	Huge
	$Cone(CC_3)$	73.31 ± 13.66	41.48%	$< 0.001^*$	3.29	Huge
Low	$HandBrush(EC)$	24.06 ± 6.46				
	$Ray(CC_1)$	51.13 ± 16.51	57.55%	$< 0.001^*$	2.15	Huge
	$Screen(CC_2)$	37.61 ± 10.75	40.50%	0.034^*	1.52	Very Large
	$Cone(CC_3)$	53.58 ± 16.54	60.27%	$< 0.001^*$	2.35	Very Large

For the operation success rate, HandBrush method and Ray method ($p < 0.001$), HandBrush method and Screen method ($p < 0.001$) are significantly different at high clutter level, but HandBrush method and Cone method ($p = 0.351$), Screen method and Ray method ($p = 0.999$), Screen method and Cone method ($p = 0.063$), Ray method and Cone method ($p = 0.056$) do not have a significant differences. At medium level of clutter, HandBrush method and Ray method ($p < 0.001$), HandBrush method and Screen method ($p = 0.005$), Cone method and Ray method ($p = 0.013$) are significantly different, but Cone method and Screen method ($p = 0.449$), HandBrush method and Cone method ($p = 0.679$), Screen method and Ray method ($p = 0.999$) are not significantly different. At low clutter level, HandBrush method and Ray method ($p < 0.001$), HandBrush method and Screen method ($p = 0.001$), Cone method and Ray method ($p = 0.011$), Cone method and Screen method ($p = 0.009$) are significantly different, but HandBrush method and Cone method ($p = 0.999$), Screen method and Ray method ($p = 0.999$) are not significantly different. Table 2 shows the part of the pairwise comparison between HandBrush and the comparison method on the operation success rate metrics.

For hand travel distances, there are significant differences between HandBrush method and Screen method ($p < 0.001$), HandBrush method and Cone method ($p < 0.001$), Cone method and Ray method ($p < 0.001$), Cone method and Screen method ($p < 0.001$), Screen method and Ray method ($p < 0.001$) at high clutter level. But HandBrush method and Ray method ($p = 0.169$) are not statistically significant. At medium clutter level, HandBrush method and Screen method ($p < 0.001$),

Table 2. Operation Success Rate (%)

Clutter level	Technique	Avg $\pm$ std. dev.	$(EC\text{-}CC_i)$ / CC_i	p	Cohen's d	Effect size
High	$HandBrush(EC)$	$81.51\% \pm 13.63\%$				
	$Ray(CC_1)$	$60.60\% \pm 5.36\%$	34.50%	$< 0.001^*$	2.01	Huge
	$Screen(CC_2)$	$60.80\% \pm 7.31\%$	34.06%	$< 0.001^*$	1.89	Very Large
	$Cone(CC_3)$	$72.72\% \pm 13.30\%$	12.09%	0.08	0.65	Small
Mid	$HandBrush(EC)$	$88.48\% \pm 10.62\%$				
	$Ray(CC_1)$	$66.74\% \pm 17.30\%$	32.57%	$< 0.001^*$	1.51	Very Large
	$Screen(CC_2)$	$72.86\% \pm 16.98\%$	21.44%	0.005^*	1.10	Large
	$Cone(CC_3)$	$81.13\% \pm 14.79\%$	9.06%	0.679	0.58	Medium
Low	$HandBrush(EC)$	$92.06\% \pm 11.41\%$				
	$Ray(CC_1)$	$74.73\% \pm 11.68\%$	23.19%	$< 0.001^*$	1.50	Very Large
	$Screen(CC_2)$	$74.41\% \pm 13.91\%$	23.72%	$< 0.001^*$	1.39	Very Large
	$Cone(CC_3)$	$89.31\% \pm 14.79\%$	3.08%	0.999	0.21	Small

HandBrush method and Cone method ($p < 0.001$), Cone method and Ray method ($p < 0.001$), Cone method and Screen method ($p = 0.018$), Screen method and Ray method ($p < 0.001$) are significantly different. But HandBrush method and Ray method ($p = 0.999$) are not statistically significant. At low clutter level, HandBrush method and Screen method ($p < 0.001$), HandBrush method and Cone method ($p < 0.001$), Cone method and Ray method ($p < 0.001$), Cone method and Screen method ($p < 0.001$), Screen method and Ray method ($p = 0.013$) are significantly different. But HandBrush method and Ray method ($p = 0.999$) are not statistically significant. Table 3 shows the part of the pairwise comparison between HandBrush and the comparison method on the hand travel distance metrics.

Table 3. Hand Travel Distance (m)

Clutter level	Technique	Avg $\pm$ std. dev.	$(EC\text{-}CC_i)$ / CC_i	p	Cohen's d	Effect size
High	$HandBrush(EC)$	25.72 ± 6.50				
	$Ray(CC_1)$	30.14 ± 5.96	16.86%	0.169	0.70	Medium
	$Screen(CC_2)$	48.55 ± 7.67	44.72%	$< 0.001^*$	3.21	Huge
	$Cone(CC_3)$	88.72 ± 6.52	69.74%	$< 0.001^*$	9.68	Huge
Mid	$HandBrush(EC)$	15.86 ± 5.75				
	$Ray(CC_1)$	16.91 ± 5.07	29.704%	0.999	0.19	Very Small
	$Screen(CC_2)$	30.32 ± 5.94	20.157%	$< 0.001^*$	2.47	Huge
	$Cone(CC_3)$	36.34 ± 5.51	10.449%	$< 0.001^*$	3.63	Huge
Low	$HandBrush(EC)$	12.06 ± 3.54				
	$Ray(CC_1)$	13.26 ± 4.37	0.74%	0.999	0.30	Small
	$Screen(CC_2)$	19.49 ± 4.59	38.96%	$< 0.001*$	1.81	Very Large
	$Cone(CC_3)$	29.47 ± 5.14	62.98%	$< 0.001^*$	3.94	Huge

Clutter level. The simple main effect analysis revealed that the difference between the three clutter levels was statistically significant for all technologies on generation time ($p < 0.001$) and hand travel distance ($p < 0.001$). On the operation success rate, the difference among the three clutter levels was statistically significant for Ray method ($p = 0.010$), Screen method ($p = 0.006$) and Cone method ($p = 0.002$) but not for Handbrush method ($p = 0.070$).

For each dependent variable and for different techniques, the Turkey post-hoc pairwise comparisons of the three clutter level were made.

For generation time, the differences between pairwise comparisons at each scene clutter level were statistically significant ($p < 0.05$) under all methods.

For the operation success rate, there is a significant difference between high scene clutter level and low scene clutter level ($p = 0.008$) when using the Ray method, but the differences comparing low scene clutter level and medium scene clutter level ($p = 0.255$), medium scene clutter level and high scene clutter level ($p = 0.555$) are not statistically significant. When using the Screen method, there is a significant difference between high scene clutter level and low scene clutter level ($p = 0.011$), between high scene clutter level and medium scene clutter level ($p = 0.029$). However, there is no statistically significant difference in comparing low scene clutter level with medium scene clutter level ($p = 0.999$). When using the Cone method, there is a significant difference between high scene clutter and low scene clutter level ($p = 0.001$), but there is no statistically significant difference comparing low scene clutter level and medium scene clutter level ($p = 0.234$) as well as medium scene clutter level and high scene clutter level ($p = 0.210$). When using the HandBrush method, there is no statistically significant difference in comparing high scene clutter level and low scene clutter level ($p = 0.070$), low scene clutter level and medium scene clutter level ($p = 0.999$), medium scene clutter level and high scene clutter level ($p = 0.070$).

For hand travel distances, there were significant differences between high and mid scene clutter levels ($p < 0.001$), high and low scene clutter levels ($p < 0.001$) when using the Ray method, but the differences between pairwise comparisons of low and mid scene clutter levels ($p = 0.209$) were not statistically significant. Differences between pairwise comparisons of each clutter level were statistically significant ($p < 0.05$) when using the Screen and Cone methods. When using the HandBrush method, there was a significant difference between high and medium scene clutter level ($p < 0.001$), between high and low scene clutter level ($p < 0.001$), but the difference between the low and medium scene clutter level ($p = 0.486$) was not statistically significant.

Main effect Below we report the main effects of each factor.

Technique. Multivariate analyses revealed a statistically significant main effect of the object grouping method on the dependent variable ($F_{9,540} = 55.598, p < 0.001^*, \Lambda_{\text{Pillai}} = 1.443, \eta_p^2 = 0.481$). Univariate main effects tests revealed statistical significance of the grouping method on generation time ($F_{3,180} = 128.132, p < 0.001^*, \eta_p^2 = 0.681$), operation success ($F_{3,180} = 25.631, p < 0.001^*, \eta_p^2 = 0.299$) and hand travel distance ($F_{3,180} = 485.426, p < 0.001^*, \eta_p^2 = 0.244$).

Since technology is a four-categorical variable, we report the results of multiple comparisons for each dependent variable. For generation time, HandBrush method and Ray method ($p < 0.001$), HandBrush method and Screen method ($p < 0.001$), HandBrush method and Cone method ($p < 0.001$), Cone method and Screen method ($p < 0.001$), Cone method and Ray method ($p < 0.001$) were significantly different. But Ray method and Screen method ($p = 0.113$) were not significantly different. For operation success rate, HandBrush method and Ray method ($p < 0.001$), HandBrush method and Screen method ($p < 0.001$), Cone method and Screen method ($p < 0.001$), Cone method and Ray method ($p < 0.001$) have significant difference. But HandBrush method and Cone method ($p = 0.088$), Ray method and Screen method ($p = 0.876$) do not have significant difference. For hand travel distance, HandBrush method and Screen method ($p < 0.001$), HandBrush method and Cone method ($p < 0.001$), Cone method and Screen method ($p < 0.001$), Cone method and Ray method ($p < 0.001$), Ray method and Screen method ($p < 0.001$) have significant differences, but Ray method and HandBrush method ($p = 0.360$) do not have significant differences.

Clutter level. Multivariate analyses revealed a statistically significant main effect of scene clutter on the dependent variable ($F_{6,358} = 52.889, p < 0.001^*, \Lambda_{\text{Pillai}} = 0.940, \eta_p^2 = 0.470$). Univariate main effects tests revealed statistical significance of scene clutter on generation time ($F_{2,180} = 263.626, p < 0.001^*, \eta_p^2 = 0.745$), operation success rate ($F_{2,180} = 17.979, p < 0.001^*, \eta_p^2 = 0.167$) and hand travel distance ($F_{2,180} = 485.426, p < 0.001^*, \eta_p^2 = 0.844$).

Since scene clutter level is a tri-categorical variable, we report the results of multiple comparisons for each dependent variable. For task completion time, high scene clutter level and medium scene clutter level ($p < 0.001$), high scene clutter level and low scene clutter level ($p < 0.001$), medium scene clutter level and low scene clutter level ($p < 0.001$) were significantly different. For success rate, there was a significant difference between high scene clutter level and medium scene clutter level ($p = 0.001$), high scene clutter level and low scene clutter level ($p < 0.001$), but not between medium scene clutter level and low scene clutter level ($p = 0.057$). For hand travel distance, there is a significant difference between high scene clutter level and medium scene clutter level ($p < 0.001$), high scene clutter level and low scene clutter level ($p < 0.001$), medium scene clutter level and low scene clutter level ($p < 0.001$).

4.5.2. Subjective Measurements

We performed a Friedman test on the subjective measures. Technique was the only independent variable. We also performed pairwise comparisons with Bonferroni correction.

Usability scores. Friedman's test showed that technique had a significant main effect on intuition ($\chi_2^2 = 34.713, p < 0.001^*$), efficiency ($\chi_2^2 = 41.535, p < 0.001^*$), accuracy ($\chi_2^2 = 33.243, p < 0.001^*$), naturalness ($\chi_2^2 = 31.510, p < 0.001^*$), satisfaction ($\chi_2^2 = 33.801, p < 0.001^*$) and ease of use ($\chi_2^2 = 29.286, p < 0.001^*$). Significant differences identified from pairwise tests were summarized in Figure 13. Based on the figs we can see that HandBrush received significantly better ratings from the experiment participants on all metrics compared to Ray method, Screen method and Cone method. This represents the fact that users perceived our method as a better method in terms of performance and experience when solving the task.

NASA-TLX workload. Friedman's test showed that technique had significant main effects on mental ($\chi_2^2 = 37.181, p < 0.001^*$), physical ($\chi_2^2 = 39.382, p < 0.001^*$), time ($\chi_2^2 = 29.983, p < 0.001^*$), performance ($\chi_2^2 = 44.032, p < 0.001^*$), effort ($\chi_2^2 = 36.058, p < 0.001^*$), frustration ($\chi_2^2 = 37.196, p < 0.001^*$), and overall ($\chi_2^2 = 45.196, p < 0.001^*$). The results of the pairwise tests are shown in Figure 14. According to the figure we can see that HandBrush scores significantly better in all metrics compared to the comparison methods. It shows that our method puts significantly less workload on the user in completing the group generation task than the other methods.

4.6. Discussion

We designed empirical studies to compare the effectiveness of four different object grouping methods for performing group generation tasks in three VR scenes of varying complexity. The results clearly show that our HandBrush method significantly improves performance. Based on the above assumptions, we have conducted the following discussion.

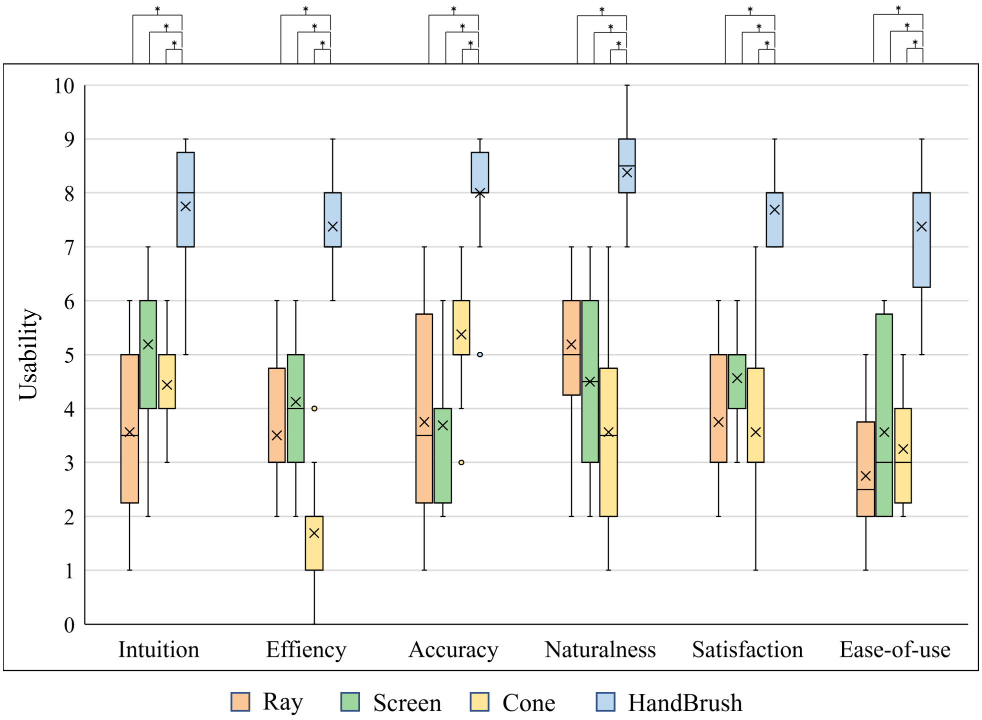

Figure 13. Usability scores for individual questions in empirical study 1. Significant difference are denoted with the asterisk and line.

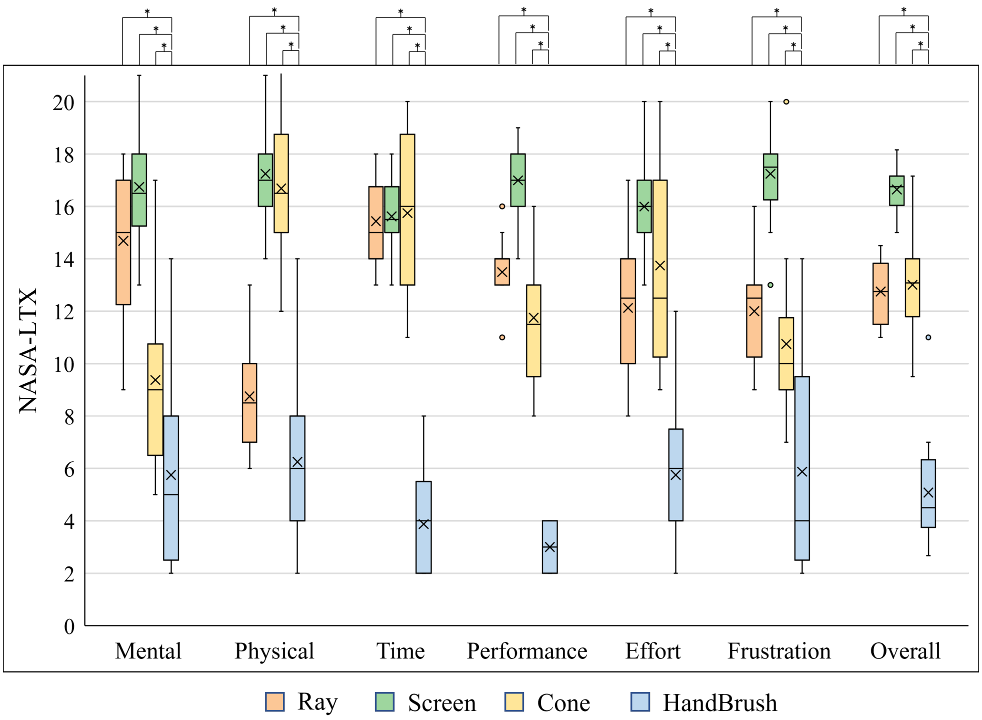

Figure 14. NASA-TLX scores for individual questions in empirical study 1. Significant difference are denoted with the asterisk and line.

4.6.1. Methods Comparison

Overall, our method performed significant improvement compared to all comparison methods. On the metric of generation time, the comparison method operates in much greater time than the HandBrush method, regardless of any level of scene clutter as shown in Table 1. Hypothesis **H1** is supported. In terms of operation success rate as shown in Table 2, all comparison methods except the Cone method have a significantly lower operation success rate than the HandBrush method. There is not much difference between the operation success rate of the Cone method and the HandBrush method, probably because the Cone method is designed to focus more on operation accuracy at the expense of operation speed. Hypothesis **H2** is supported. In terms of hand travel distance as shown in Table 3, all the compared methods do not perform as well as the HandBrush method. The hand travel distance of the Screen method and the Cone method is significantly larger than that of the HandBrush method, while the hand travel distance of the Ray method is not much different from that of the HandBrush method. This may be due to the intrinsic properties of the Ray method. The Ray method allows the user to select objects by means of rays emanating forward from the hand. Small movements of the hand result in large movements of the far end of the ray, allowing the user to select objects in a larger spatial area with fewer hand movements. Hypothesis **H3** is supported. The results of the subjective metrics show that our method has lower task load and higher system availability for users to other compared methods as shown in Figure 13 and Figure 14. Combined with the result of hand travel distance, hypothesis **H4** is proved.

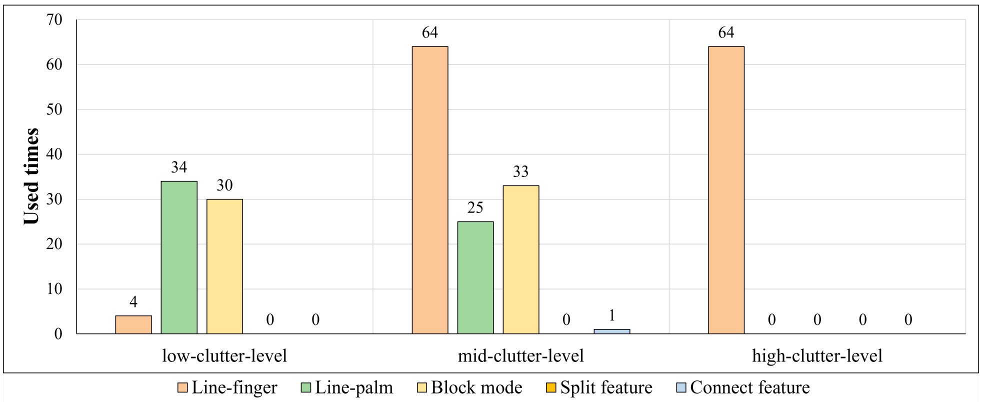

Figure 15. Feature usage statistics, which record the number of times different modes and features were used in the task.

4.6.2. Feature Usage

We recorded the number of times each mode and feature of HandBrush was used at different scene clutter level. The statistical results are shown in Figure 15, revealing the users' preference for each HandBrush feature in different scenarios.

The experimental data showed that the number of users utilizing the Block mode was inversely related to the clutter level of the scene. Specifically, in low-clutter scenarios where objects were orderly arranged and similar items were clustered together, Block mode was favored by more users. This could be attributed to users' ease in identifying clustered areas of objects in such scenarios, leading them to prefer selecting entire regions of objects at once.

Conversely, the Line-finger under the Line mode was widely adopted in medium to high clutter scenarios. This might be because users preferred precise selections to avoid mistakenly selecting non-target objects in cluttered and chaotic environments. The Line-finger offers higher selection accuracy, making it suitable for fine operations in complex scenes. More people choose to use Line-palm mode when the scene is moderately chaotic or less chaotic. This phenomenon may reflect users' preference for more efficient selection when faced with selection tasks in less cluttered scenarios. The Line-palm allows users to select objects through a wider range of choices, which is suitable for situations where objects are more dispersed and easy to recognize. In the group generation task, we observed relatively low willingness among users to use the Split and Connect features. This might be related to users' familiarity with these features, or users might not have encountered scenarios requiring the use of these features during the experiment. Additionally, users may prefer completing the selection in one go rather than engaging in complex editing operations afterward.

4.6.3. Subjective Feedback

After participants completed the subjective questionnaire, we asked them about their experiences with the four methods.

The vast majority of users think that the HandBrush method is a better method, and the modes it provides, such as Line-finger, Line-palm, Block mode, etc. can help them accomplish the grouping task very well. However, some users think that after the objects are grouped, the lines connecting these objects will be intertwined, resulting in a cluttered scene. In the actual selection process, the visual characteristics of these lines are not fully utilized, but rather aggravate the user's visual burden. Based on this feedback, we suggest that if line patterns are to be applied in actual system design, users should be provided with customizable options that allow them to choose whether or not to display lines according to their preferences and task requirements. Such flexibility can help users to minimize visual distractions when needed, thereby improving operational clarity and efficiency.

Users also reported that they didn't need the Split and Connect features as much in this task. They preferred to select all objects at once rather than merge groups later.

By asking users about their experiences with the three comparison methods, we found that the main issue users reacted to with the Ray method was the tendency to mistakenly select other objects, which is related to both the bare-hand recognition accuracy of the device and the principle of ray selection. Since the rays are emitted by our hands, some behaviors (e.g., finger taps used for confirmation) may result in small changes in orientation that can affect pointing accuracy. This variation is amplified with the length of the ray (distance to the target) and can lead to selection errors or omissions. This is well-known Heisenberg effect (Wolf, Gugenheimer, Combosch, & Rukzio, 2020). The main problems users reported with the Screen method were that it was too inconvenient for selecting scenes with a high degree of clutter and that the shape of the range was fixed for each selection compared to our Block mode. This resulted in the need to make multiple selections for irregularly arranged objects, which greatly increased hand fatigue. For Cone mode, participants mainly reported that dragging our Cone to move to select each time brought great fatigue to our hands, which was not a good experience.

5. Empirical Study 2

In this section, we designed a user study to evaluate the performance of our HandBrush in the group modification task. Based on the results of this study, we tested the following hypotheses:

$H1$: Compared to traditional methods, HandBrush can significantly reduce the time

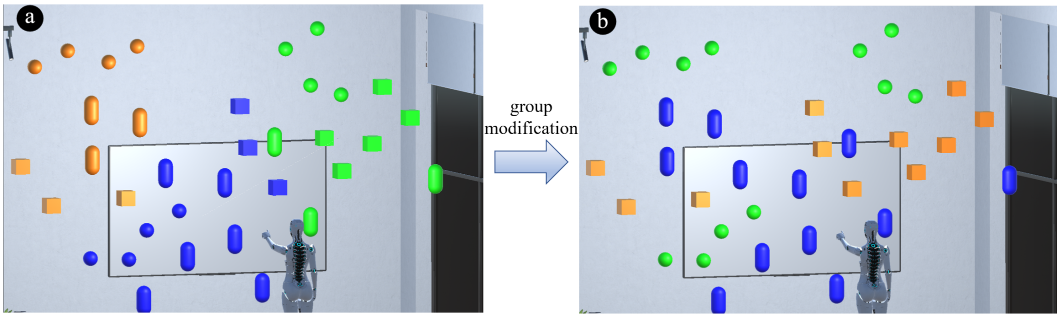

Figure 16. Group Modification Experiment Scenario. At the beginning of the experiment, the objects in the scene are randomly divided into three groups based on their spatial relationships. Participants are required to regroup the objects in the scene according to their shapes, with objects of the same shape being grouped together.

required to complete the group modification task.

$H2$: Compared to traditional methods, HandBrush reduces hand travel distance.

$H3$: HandBrush method brings less fatigue compared to traditional methods.

5.1. Participants and Apparatus

Sixteen participants (11 males, 5 females), aged between 21 and 25(mean 22.56, variance 1.26), took part in the experiment. Some of them had prior experience with VR devices. All participants had normal vision (or corrected-to-normal vision with glasses). The experimental system was built using Unity 2022 and deployed on the Pico neo3 headset. The system ran smoothly on the Pico neo3 platform.

5.2. Experimental Design

We randomly placed objects of different shapes within a rectangular space 10 meters away from the participants (since occlusion is not considered, the objects are placed on a single plane). Initially, we divided these objects into three groups based on their spatial relationships.

We evaluated five methods: Ray method, Screen method, Cone method (Shi et al., 2023), LineBlock method (HandBrush without Split and Connect features), and HandBrush method (with Split and Connect features). For those comparison methods, the user is required to delete the existing groups in the scene by comparing them to the delete gesture and then re-group the objects in the scene. A schematic of the different methods to accomplish this task is shown in Figure 17

5.3. Task and Procedure

There are three different shapes of objects in the user's task scene, which have been divided into three groups according to their positional relationships. The user's task is to modify an existing group in the scene so that objects of the same shape in the modified scene are a group.

The entire experiment lasted approximately 60 minutes per participant. Participants first completed a questionnaire regarding their personal information and previous experience with VR HMDs.

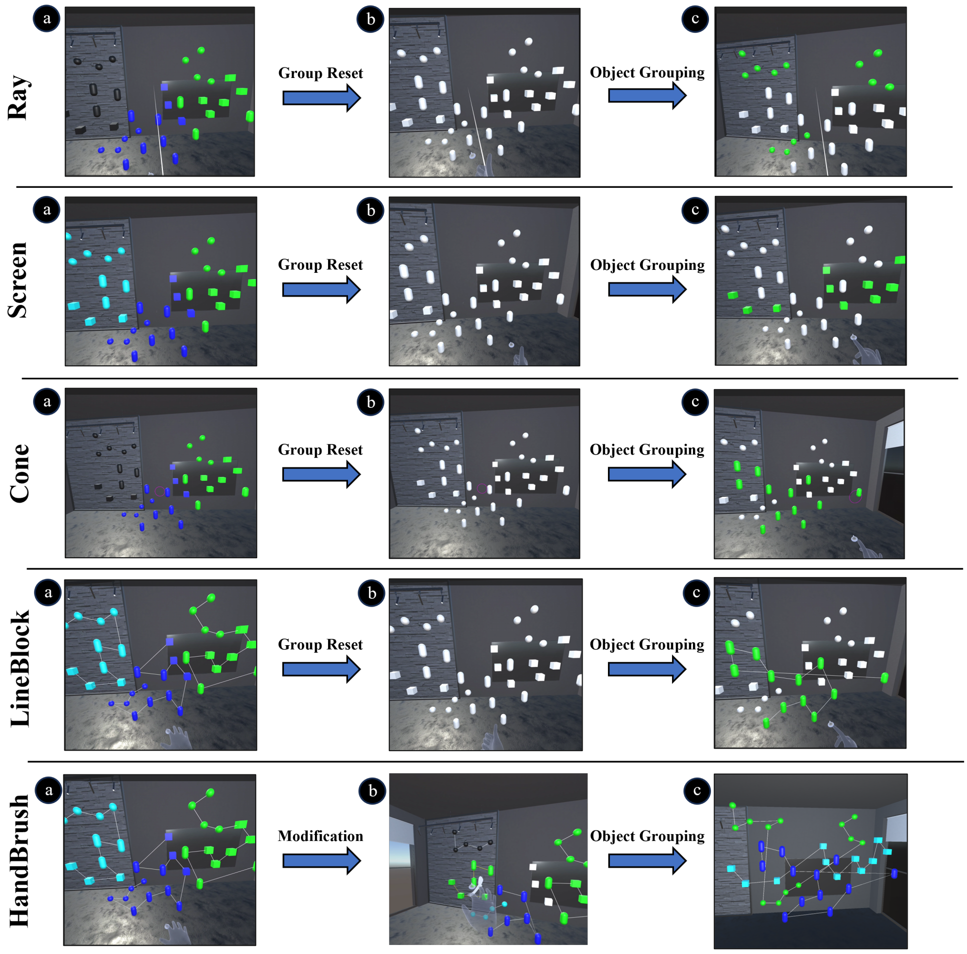

Figure 17. Screenshots of 5 different methods in this task. From top to bottom are Ray method, Screen method, Cone method, LineBlock method and HandBrush method.

Subsequently, we introduced the experiment to the participants and guided them to stand at the center of the experimental area while wearing the HMD. We allotted sufficient time for all participants to familiarize themselves with the virtual reality system and to practice grouping objects in our training scenarios using the five methods described above. During the practice session, adjustments were made to the height of all targets and the distance between participants and the targets (participants could slightly adjust their standing position within a 20 cm radius). Participants were instructed to use all methods to modify object groups.

Once the formal experiment commenced, we initially randomized the objects into three groups based on their spatial relationships. Participants' task was then to modify these groups based on the shapes of the objects, with the ultimate goal of regrouping the objects in the scene according to their shapes.

In order to reduce the influence of order effects on the experimental results, we used a Latin-square design to balance the order of the technique conditions, which ensured the fairness and objectivity of the experiment. In total, we collected 320 data points (16 participants * 5 techniques * 4 repetitions).

After each technical condition, participants were asked to complete the Usability questionnaire and the NASA-TLX questionnaire and to take a short break. At the end of the experiment, a brief interview was conducted to gather their subjective impressions.

5.4. Metrics

In this study, we established three objective metrics and two subjective metrics to measure the performance of different techniques in scenarios with different levels of clutter.

Modification time. In this study, modification time specifically refers to the time taken to complete the entire group modification task. To assess the impact of different techniques on user efficiency in performing group modification tasks

Hand travel distance. In this study, hand travel distance specifically refers to the spatial distance traveled by the virtual hand of participants while performing group modification tasks. This metric was chosen as a key parameter to measure the impact of different technologies on user fatigue. Specifically, hand travel distance reflects the total distance that the user moves his/her hand while performing operations in the virtual environment, which is directly related to the physical exertion and fatigue felt by the user while completing the task.

Usability. A usability questionnaire to evaluate users' perceptions of technology usability, focusing on intuition, efficiency, accuracy, naturalness, satisfaction, and ease of use, scored from 1 to 10. The six questions are: is this method intuitive (Q1), is the method efficient (Q2), is the method accurate (Q3), is the method natural (Q4), is the method satisfied (Q5), is the method easy to use (Q6).

NASA-TLX. The NASA-TLX questionnaire, assessing workload across six dimensions (mental, physical, time, performance, effort, and frustration) with a 1-20 scoring scale. The scores of the six dimensions were weighted and averaged to obtain a composite load score.

5.5. Results

5.5.1. Objective Measurement

We first identified and removed outliers for each condition where the selection time exceeded $M \pm 3 \times SD$. In total, we excluded 6 data points (1.88%). Next, we calculated each user's average score for each metric.

Then, we used a one-factor multivariate analysis of variance (MANOVA) to analyze the effects of technique on the performance of the two objective metrics described above.

The assumptions of the methodology were tested before proceeding with the analysis. Scatter plots showed a linear relationship between the dependent variables in each set of independent variables. The Pearson correlation test found no multicollinearity ($|r| < 0.9$) between the two dependent variables. Box plots did not find unidirectional outliers and Mahalanobis distance did not find multivariate outliers.

The Shapiro-Wilk test showed that the two dependent variables (modification time, hand travel distance) obeyed a normal distribution ($p > 0.05$).

Box's M test showed that the variance/covariance matrices of the two dependent variables within each group of the independent variables were equal ($p = 0.066$). Levene's test showed that the dependent variables within each group of the independent variables were equal in variance ($p > 0.05$).

The results shows there is a statistically significant effect of technology on the dependent variable ($F(8, 148) = 69.782, p < 0,0001^*, \eta^2 = 0.790$). Technology has a significant effect on modification time ($F(4, 75) = 51.475, p < 0,0001^*, \eta^2 = 0.733$), hand travel distance ($F(4, 75) = 160.60, p < 0,0001^*, \eta^2 = 0.895$).

Modification time. For modification time, the Tukey post-hoc test shows that HandBrush method and Ray method ($p < 0.001$), HandBrush method and Screen method ($p < 0.001$), HandBrush method and Cone method ($p < 0.001$), HandBrush method and LineBlock method ($p = 0.024$), LineBlock method and Ray method ($p < 0.001$), LineBlock method and Screen method ($p = 0.001$), LineBlock method and Cone method ($p < 0.001$), Cone method and Ray method ($p < 0.001$), Cone method and Screen method ($p < 0.001$) have significant differences. But Screen method and Ray method ($p = 0.332$) do not have significant difference. Table 4 shows the part of the pairwise comparison between HandBrush and the comparison method on the modification time metrics.

Table 4. Modification Time (s)

Technique	Avg $\pm$ std. dev.	$(CC_i\text{-}EC)$ / CC_i	p	Cohen's d	Effect size
$HandBrush(EC)$	50.21 ± 8.56				
$Ray(CC_1)$	96.71 ± 17.24	48.08%	$< 0.001^*$	3.41	Huge
$Screen(CC_2)$	87.18 ± 14.23	42.40%	$< 0.001^*$	3.15	Huge
$Cone(CC_3)$	120.41 ± 24.22	56.47%	$< 0.001^*$	4.85	Huge
$LineBlock(CC_4)$	65.70 ± 12.82	23.57%	0.024^*	1.42	Large

Hand travel distance. The Tukey post-hoc test shows that there is a significant difference between the HandBrush method and Screen method ($p < 0.001$), HandBrush method and Cone method ($p < 0.001$), HandBrush method and LineBlock method ($p = 0.043$), LineBlock method and Screen method ($p < 0.001$), LineBlock method and Cone method ($p < 0.001$), Cone method and Ray method ($p < 0.001$), Cone method and Screen method ($p < 0.001$), Screen method and Ray method ($p < 0.001$). But HandBrush method and Ray method ($p = 0.986$), LineBlock method and Ray method ($p = 0.144$) do not have significant differences. Table 5 shows the part of the pairwise comparison between HandBrush and the comparison method on the hand travel distance metrics.

Table 5. Hand Travel Distance (m)

Technique	Avg $\pm$ std. dev.	$(CC_i\text{-}EC)$ / CC_i	p	Cohen's d	Effect size
$HandBrush(EC)$	25.38 ± 3.68				
$Ray(CC_1)$	26.39 ± 6.81	3.82%	0.986	0.18	Small
$Screen(CC_2)$	42.76 ± 5.02	40.65%	$< 0.001^*$	3.94	Huge
$Cone(CC_3)$	67.81 ± 7.55	61.61%	$< 0.001^*$	5.92	Huge
$LineBlock(CC_4)$	31.01 ± 3.67	18.15%	0.026^*	1.53	Very Large

5.5.2. Subjective Measurement

We performed a Friedman test on the subjective measures. Technique was the only independent variable. We also performed pairwise comparisons with Bonferroni correction.

Usability scores. Friedman's test showed that technique had a significant main effect on intuition ($\chi^2_2 = 50.475, p < 0.001^*$), efficiency ($\chi^2_2 = 55.174, p < 0.001^*$), accuracy ($\chi^2_2 = 51.741, p < 0.001^*$) naturalness ($\chi^2_2 = 51.355, p < 0.001^*$), satisfaction ($\chi^2_2 = 56.091, p < 0.001^*$) and ease of use ($\chi^2_2 = 56.839, p < 0.001^*$). Significant differences identified from pairwise tests were summarized in Figure 18. Based on the results we can see that compared to Ray method, Screen method, Cone method and

LineBlock method, HandBrush received better scores from users on all metrics. The data shows how users perceived our approach as a better method in terms of both performance and experience when solving the task.

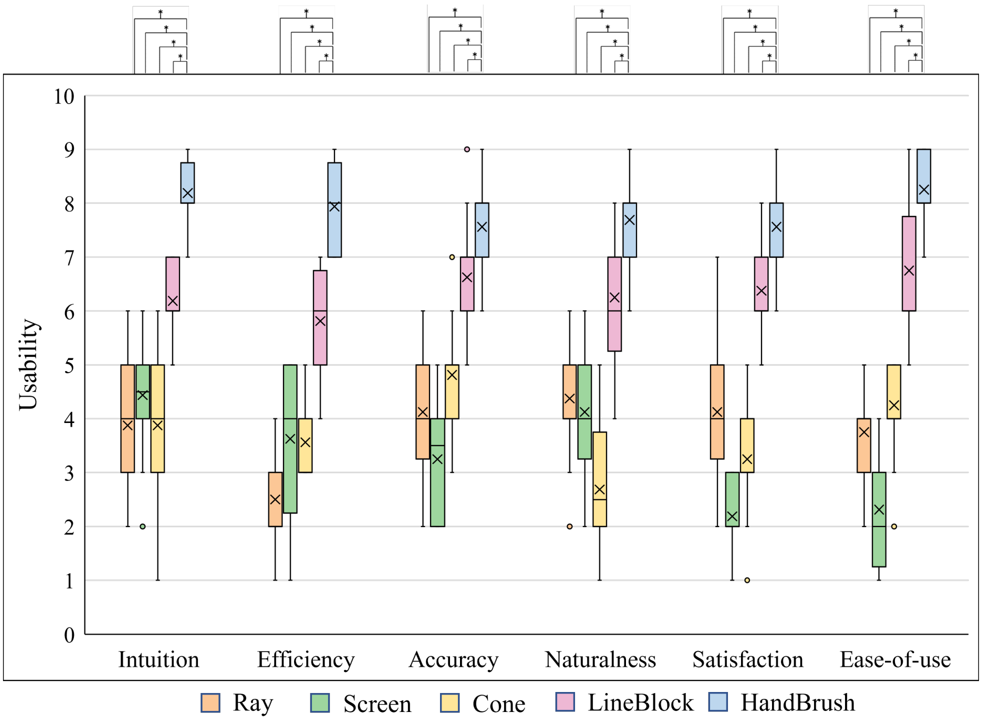

Figure 18. Usability scores for individual questions in empirical study 2. Significant difference are denoted with the asterisk and line.

NASA-TLX workload. Friedman's test showed that technique had significant main effects on mental ($\chi^2_2 = 50.584, p < 0.001^*$), physical ($\chi^2_2 = 55.494, p < 0.001^*$), time ($\chi^2_2 = 50.804, p < 0.001^*$), performance ($\chi^2_2 = 43.449, p < 0.001^*$), effort ($\chi^2_2 = 55.456, p < 0.001^*$), frustration ($\chi^2_2 = 48.673, p < 0.001^*$), and overall ($\chi^2_2 = 54.943, p < 0.001^*$). Moreover, based on the collected data, our method significantly outperforms Ray method, Screen method, Cone method and LineBlock method across all different metrics. The results of the pairwise comparisons are shown in Figure 19. From the figure, we can see that compared to the comparison methods, users believe that we are significantly better than them in each indicator. It can be seen that in the task of completing group modifications, our HandBrush with Split and Connect features brings less workload of use than the comparison method.

5.6. Discussion

5.6.1. Comparison Methods

Overall, the comparison methods do not perform as well as the HandBrush methods on the group modification task. In the modification time metric as shown in Table 4, the modification time of all the comparison methods is significantly larger than that of the HandBrush method, which represents that HandBrush has a better efficiency performance in this task. Hypothesis **H1** is supported. In the hand travel distance metric as shown in Table 5, the hand travel distance of all compared methods is larger than that of the HandBrush method. The Screen method, the Cone method

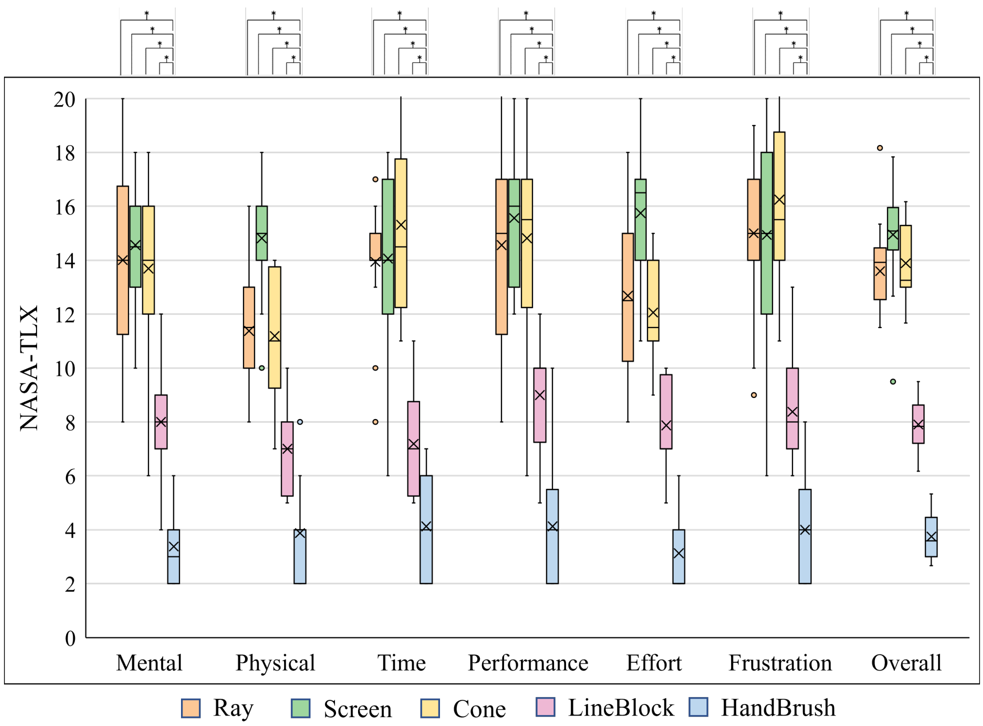

Figure 19. NASA-TLX scores for individual questions in empirical study 2. Significant difference are denoted with the asterisk and line.

and the LineBlock method (HandBrush method without Split and Connect features) are significantly larger than the HandBrush method. The difference between the Ray method and the HandBrush method in terms of hand travel distance is not significant, but the Ray method takes much longer than the HandBrush method. Hypothesis **H2** is supported. The results of the subjective metrics show that our method has lower task load and higher system availability for users compared to other compared method. Combining objective indicators, Hypotheses **H3** is supported.

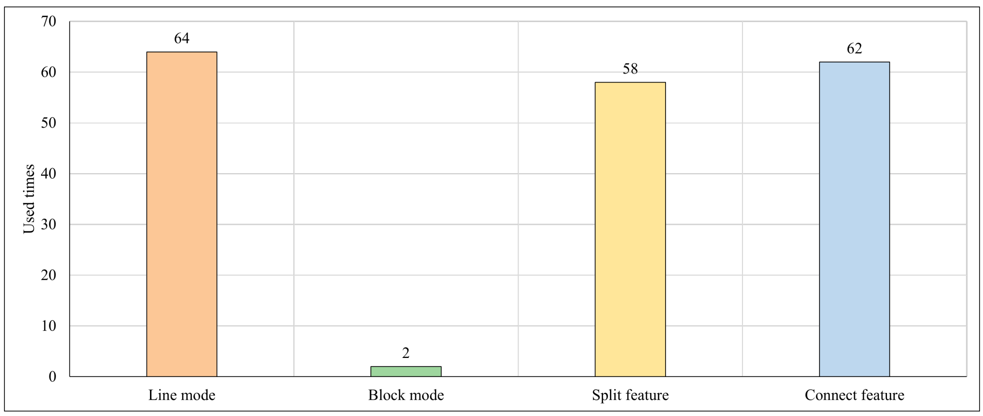

Figure 20. Feature usage statistics, which record the number of times different modes and features were used in the task.

5.6.2. Feature usage

We record the number of times different modes and features were used in the task. Through experimental results (as shown in Figure 20), we were able to intuitively observe user preferences and usage patterns for the different features of the HandBrush system during group modification tasks. The experimental data revealed a clear trend: during group modification tasks, the vast majority of users tended to use the split and connect features provided by the HandBrush system. This indicates that these features are widely welcomed by users due to their ability to provide more efficient operational workflows. Specifically, the Split feature allows users to decompose complex object groups through intuitive gesture operations, while the Connect feature enables users to easily reassemble these decomposed objects into new groups. The combined use of these two features greatly enhances the flexibility and efficiency of users in performing group editing tasks. At the same time, the experimental results also showed that most users would use Line Mode to group objects during our group editing tasks. This may be related to the design of our modification tasks; when facing overly complex group editing tasks, users may choose to directly delete the previously grouped objects and then re-group the objects in the scene. This choice reflects users' adaptive strategies when faced with complex tasks, as they tend to adopt more direct and faster methods to achieve their goals. It is worth noting that Block mode was only used twice in this mission. This is closely related to the design of our task scenario. In this task, objects that need to be grouped may be arranged in a crossed manner, and using the Block Mode method may significantly increase the difficulty of user selection.

5.6.3. Subjective Feedback

After participants completed the subjective questionnaire, we asked them about their experiences with the four methods.

More than half of the users believe that the Split and Connect features significantly facilitate the execution of Group modification tasks. Users generally perceive that these features reduce the need for repetitive selection operations, thereby improving editing efficiency. Particularly, when adjustments or rearrangements to complex group structures are necessary, the Split and Connect features allow users to edit in a more intuitive and flexible manner without having to reselect all objects from scratch.

Although the Split and Connect features have been widely welcomed, some users expressed different usage preferences. In cases where object grouping involves intertwined lines or chaotic scenes, or when facing complex group editing tasks, these users tend to prefer the method of deleting previously partitioned groups and then re-grouping the objects in the scene. They believe that while this method may appear time-consuming, it actually avoids precise cutting and splicing operations in chaotic lines, thereby reducing operational complexity.

6. Conclusions, Limitations, and Future Work

In this paper, we delve into the technique of bare-hand object grouping in VR environments and propose an innovative gesture-adaptive technique called HandBrush. We compare it with existing techniques, including the Ray method, Screen method, and Cone method (Shi et al., 2023), through two user studies for comparison and evaluation. Our results validate the advantages of HandBrush in both group generation and group modification tasks. In the group generation task, HandBrush showed significant improvements in task completion time, operation success rate, and distance traveled in hand space compared to the comparison method. In the group modification task, HandBrush has a significant improvement in completion time and hand travel distance compared to the comparison method. Moreover, the Split and Connect features of HandBrush provide us with significant convenience in group modification tasks.

As an initial exploration of bare-hand object grouping techniques in VR head-mounted displays (HMDs), further work is needed to fully explore and understand the potential of HandBrush. Firstly, the Ray method was devised and implemented for controller selection, whereas we utilize the built-in cameras of HMDs to track hands as a bare-hand approach, which may lead to slightly different outcomes due to tracking method differences. Secondly, in 3D virtual environments, there may be variations in target density and size, posing challenges during the use of HandBrush. Therefore, as part of future work, research will be conducted on target density and size to explore their impact on HandBrush performance. Additionally, in real-world scenarios, object placement will be more complex. Objects may be placed on different planes, and there may be occlusions between 3D targets. These complexities necessitate further exploration of visualization techniques during the object grouping process, as well as the impact of different visual feedback methods on users' selection of desired targets. Finally, regarding gesture selection in bare hand interaction, a large part of our gesture design also depends on the ability of our headset to recognize gestures. The current gesture design may not be optimal and has potential limitations. Some gestures, although not used in this method, may be more appropriate for object grouping.

A series of studies will be conducted to clarify and guide users on how to more effectively complete bare-hand object grouping tasks.

7. Acknowledgements

This work is supported by the National Natural Science Foundation of China through Project [61932003, 62372026], by Beijing Science and Technology Plan Project [Z221100007722004], and by National Key R&D plan [2019YFC1521102].

References

Benavides, A., Khadka, R., & Banic, A. (2019). Physically-based bimanual volumetric selection for immersive visualizations. In *Virtual, augmented and mixed reality. multimodal interaction: 11th international conference, vamr 2019, held as part of the 21st hci international conference, hcii 2019, orlando, fl, usa, july 26–31, 2019, proceedings, part i 21* (pp. 183–195).

Bergström, J., Dalsgaard, T.-S., Alexander, J., & Hornbæk, K. (2021). How to evaluate object selection and manipulation in vr? guidelines from 20 years of studies. In *proceedings of the 2021 chi conference on human factors in computing systems* (pp. 1–20).

Ciolfi Felice, M., Maudet, N., Mackay, W. E., & Beaudouin-Lafon, M. (2016). Beyond snapping: Persistent, tweakable alignment and distribution with stickylines. In *Proceedings of the 29th annual symposium on user interface software and technology* (pp. 133–144).

Dehmeshki, H., & Stuerzlinger, W. (2009). Ice-lasso: An enhanced form of lasso selection. In *2009 ieee toronto international conference science and technology for humanity (tic-sth)* (pp. 630–635).

Dehmeshki, H., & Stuerzlinger, W. (2010). Design and evaluation of a perceptual-based object group selection technique. In *Proceedings of hci 2010.*

Fang, W., & Hong, J. (2022). Bare-hand gesture occlusion-aware interactive augmented reality assembly. *Journal of Manufacturing Systems*, *65*, 169–179.

Fernandes, B., & Fernández, J. (2009). Bare hand interaction in tabletop augmented reality. In *Siggraph'09: Posters* (pp. 1–1).

Hancock, M., Carpendale, S., & Cockburn, A. (2007). Shallow-depth 3d interaction: design and evaluation of one-, two-and three-touch techniques. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1147–1156).

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, pp. 904–908).

Hayatpur, D., Heo, S., Xia, H., Stuerzlinger, W., & Wigdor, D. (2019). Plane, ray, and point: Enabling precise spatial manipulations with shape constraints. In *Proceedings of the 32nd annual acm symposium on user interface software and technology* (pp. 1185–1195).

Kim, H., Lee, G., & Billinghurst, M. (2015). A non-linear mapping technique for bare-hand interaction in large virtual environments. In *Proceedings of the annual meeting of the australian special interest group for computer human interaction* (pp. 53–61).

Li, Y., Sarcar, S., Kim, K., Tu, H., & Ren, X. (2022). Designing successive target selection in virtual reality via penetrating the intangible interface with handheld controllers. *International Journal of Human-Computer Studies*, *165*, 102835.

Long, J., Li, Y., Yu, T., & Gu, Z. (2011). Target selection with hybrid feature for bci-based 2-d cursor control. *IEEE Transactions on biomedical engineering*, *59*(1), 132–140.

Lucas, J. F. (2005). *Design and evaluation of 3d multiple object selection techniques* (Unpublished doctoral dissertation). Virginia Tech.

Mendes, D., Caputo, F. M., Giachetti, A., Ferreira, A., & Jorge, J. (2019). A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum* (Vol. 38, pp. 21–45).

Montano-Murillo, R. A., Nguyen, C., Kazi, R. H., Subramanian, S., DiVerdi, S., & Martinez-Plasencia, D. (2020). Slicing-volume: Hybrid 3d/2d multi-target selection technique for dense virtual environments. In *2020 ieee conference on virtual reality and 3d user interfaces (vr)* (pp. 53–62).

Moscovich, T., & Hughes, J. F. (2006). Multi-finger cursor techniques. In *Proceedings of graphics interface 2006* (pp. 1–7).

Nabiyouni, M., Laha, B., & Bowman, D. A. (2014). Poster: Designing effective travel techniques with bare-hand interaction. In *2014 ieee symposium on 3d user interfaces (3dui)* (pp. 139–140).

Oh, J.-Y., Stuerzlinger, W., & Dadgari, D. (2006). Group selection techniques for efficient 3d modeling. In *3d user interfaces (3dui'06)* (pp. 95–102).

Ong, S., & Wang, Z. (2011). Augmented assembly technologies based on 3d bare-hand interaction. *CIRP annals*, *60*(1), 1–4.

Poupyrev, I., Billinghurst, M., Weghorst, S., & Ichikawa, T. (1996). The go-go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th annual acm symposium on user interface software and technology* (pp. 79–80).

Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the fourth berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics* (Vol. 4, pp. 547–562).

Ruddle, R. (2005). *3d user interfaces: Theory and practice.* MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ....

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of modern applied statistical methods*, *8*, 597–599.

Schäfer, A., Reis, G., & Stricker, D. (2022). Controlling continuous locomotion in virtual reality with bare hands using hand gestures. In *International conference on virtual reality and mixed reality* (pp. 191–205).

Sears, A., Lin, M., & Karimullah, A. S. (2002). Speech-based cursor control: understanding the effects of target size, cursor speed, and command selection. *Universal Access in the Information Society*, *2*, 30–43.

Shi, R., Wei, Y., Hu, X., Liu, Y., Yue, Y., Yu, L., & Liang, H.-N. (2024). Experimental analysis of freehand multi-object selection techniques in virtual reality head-mounted displays. *arXiv preprint arXiv:2409.00982*.

Shi, R., Zhang, J., Stuerzlinger, W., & Liang, H.-N. (2022). Group-based object alignment

in virtual reality environments. In *Proceedings of the 2022 acm symposium on spatial user interaction* (pp. 1–11).
Shi, R., Zhang, J., Yue, Y., Yu, L., & Liang, H.-N. (2023). Exploration of bare-hand mid-air pointing selection techniques for dense virtual reality environments. In *Extended abstracts of the 2023 chi conference on human factors in computing systems* (pp. 1–7).
Shi, R., Zhu, N., Liang, H.-N., & Zhao, S. (2021). Exploring head-based mode-switching in virtual reality. In *2021 ieee international symposium on mixed and augmented reality (ismar)* (pp. 118–127).
Song, P., Goh, W. B., Hutama, W., Fu, C.-W., & 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* (pp. 1297–1306).
Stenholt, R. (2012). Efficient selection of multiple objects on a large scale. In *Proceedings of the 18th acm symposium on virtual reality software and technology* (pp. 105–112).
Strothoff, S., Stuerzlinger, W., & Hinrichs, K. (2015). Pins'n'touches: an interface for tagging and editing complex groups. In *Proceedings of the 2015 international conference on interactive tabletops & surfaces* (pp. 191–200).
Stuerzlinger, W., & Smith, G. (2002). Efficient manipulation of object groups in virtual environments. In *Proceedings ieee virtual reality 2002* (pp. 251–258).
Von Hardenberg, C., & Bérard, F. (2001). Bare-hand human-computer interaction. In *Proceedings of the 2001 workshop on perceptive user interfaces* (pp. 1–8).
Wang, X., Ong, S., & Nee, A. Y.-C. (2016). Multi-modal augmented-reality assembly guidance based on bare-hand interface. *Advanced Engineering Informatics*, *30*(3), 406–421.
Ware, C., & Lowther, K. (1997). Selection using a one-eyed cursor in a fish tank vr environment. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *4*(4), 309–322.
Wolf, D., Gugenheimer, J., Combosch, M., & Rukzio, E. (2020). Understanding the heisenberg effect of spatial interaction: A selection induced error for spatially tracked input devices. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–10).
Wolfartsberger, J., Zenisek, J., Sievi, C., & Silmbroth, M. (2017). A virtual reality supported 3d environment for engineering design review. In *2017 23rd international conference on virtual system & multimedia (vsmm)* (pp. 1–8).
Wu, Z., Yu, D., & Goncalves, J. (2023). Point-and volume-based multi-object acquisition in vr. In *Ifip conference on human-computer interaction* (pp. 20–42).
Zhao, M., Ong, S.-K., & Nee, A. Y. (2016). An augmented reality-assisted therapeutic healthcare exercise system based on bare-hand interaction. *International Journal of Human–Computer Interaction*, *32*(9), 708–721.

About the Authors

Sichun Huang is a master student in the School of Computer Science and Engineering of Beihang University, China. His current research focuses on virtual reality, augmented reality, and HCI.

Jian Wu received his Ph.D. degree from the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China. He is currently an assistant professor at the School of Computer Science and Engineering, Beihang University, Beijing, China. His current research focuses on virtual reality, augmented reality, human-computer interaction and visualization.

Runze Fan is a Ph.D student at the School of Computer Science and Engineering, Beihang University, China. His current research focuses on virtual reality and augmented reality.

Sio Kei Im received his Ph.D. degree from Queen Mary University of London, United Kingdom. He is a professor at the Faculty of Applied Sciences, Macau Polytechnic University. His research interests include video coding, image processing, machine learning.

Lili Wang received her Ph.D. degree from the Beihang University, Beijing, China. She is a professor with the School of Computer Science and Engineering of Beihang University, and a researcher with the State Key Laboratory of Virtual Reality Technology and Systems. Her interests include virtual reality, augmented reality, mixed reality, real-time rendering and realistic rendering.